# Molecular Characterization of the Genes Encoding Pyruvate Formate-Lyase and Its Activating Enzyme of *Clostridium pasteurianum*

GERHARD WEIDNER[1] AND GARY SAWERS[1,2]*

*Department of Microbiology and Genetics, University of Munich, D-80638 Munich, Germany,[1] and
Nitrogen Fixation Laboratory, John Innes Centre, Norwich Research Park,
Norwich NR4 7UH, United Kingdom[2]*

**Formate is the major source of C1 units in many species of the genus *Clostridium*. In this study we have cloned and characterized the genes encoding pyruvate formate-lyase and its activating enzyme of *Clostridium pasteurianum*. The genetic and transcriptional organizations of the genes and the high level of homology exhibited by the respective gene products to their *Escherichia coli* counterparts indicate strong evolutionary conservation of these enzymes.**

Pyruvate formate-lyase (Pfl) catalyzes the reversible, coenzyme A-dependent, nonoxidative dissimilation of pyruvate to acetyl-coenzyme A and formate and forms the major route of pyruvate catabolism in anaerobically growing *Escherichia coli* cells (10). The enzyme is a homodimer composed of 85-kDa subunits (4). The elegant studies of Knappe and his colleagues have elucidated the catalytic reaction sequence involved in the homolytic cleavage of the α-oxocarboxylic C-C bond of pyruvate (9, 16, 32). The cleavage of the C-C bond is catalyzed by an organic free radical which is located on a glycyl residue at amino acid position 734 on the polypeptide chain (33). The extreme oxygen sensitivity of the free radical is characterized by the fragmentation in vitro of the polypeptide backbone between serine-733 and glycine-734, delivering 82- and 3-kDa products (33). The anaerobic introduction of the radical into the polypeptide chain is catalyzed by a 28-kDa iron-dependent Pfl-activating enzyme (Act) (10). The removal of the radical is catalyzed by the trifunctional *adhE* gene product (8).

Pfl enzyme activity has been detected in a number of anaerobic bacteria, including *Enterococcus faecalis* (12), *Streptococcus mutans* (28), *Clostridium kluyveri* (30), and *Clostridium butyricum* (29). Evidence has also been presented which indicates that, as with the *E. coli* enzyme, Pfl is interconvertible between inactive and active species in clostridiae (34). In contrast to its catabolic function in *E. coli*, *E. faecalis*, and *S. mutans*, it has been proposed that the clostridial enzyme has an anabolic function in furnishing the cells with C1 units (29, 30). Until now, Pfl could not be detected in *Clostridium pasteurianum* and it had been proposed that formate is provided solely by a $CO_2$ reductase (29). This study presents the cloning and characterization of the *pfl* and *act* genes of *C. pasteurianum*.

**Cross-reaction between anti-*E. coli* Pfl antiserum and an 85-kDa polypeptide in extracts of *C. pasteurianum*.** *E. coli* strains (Table 1) were grown in Luria-Bertani medium (18). The *Clostridium* species were grown in the modified minimal medium described by O'Brien and Morris (15). The *Strepto-*

*coccus* and *Enterococcus* species were grown in Todd-Hewitt broth (Oxoid). After anaerobic growth under a nitrogen atmosphere (2) or, in the case of *Streptococcus* and *Enterococcus* species, after semianaerobic growth, cell extracts from several gram-positive bacteria were prepared, and after the separation of their polypeptides by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (11) and their transfer to nitrocellulose (31), the cell extracts were challenged with antiserum raised against Pfl from *E. coli*. A strong cross-reaction was observed with polypeptides in extracts from *C. butyricum*, *Streptococcus salivarius*, and *E. faecalis*, while a weaker, but nevertheless clear, cross-reaction was detectable with a *C. pasteurianum* extract (Fig. 1). The extract from *Streptococcus sanguis* also exhibited a very weak cross-reacting species. These cross-reacting polypeptides migrated with molecular masses of approximately 80 to 85 kDa, which is characteristic of the Pfl of *E. coli* (33) (Fig. 1, lane 1), and they were the only polypeptides in these extracts that exhibited any cross-reaction with the Pfl antibodies. The extracts from *C. butyricum* and *S. salivarius* possessed two closely migrating polypeptides which exhibited cross-reaction (Fig. 1, lanes 4 and 6). Such a doublet is frequently observed with Pfl from *E. coli* and results from a specific fragmentation of the polypeptide chain when the radical-bearing enzyme species is exposed to dioxygen (33).

***C. pasteurianum* has homologs of *E. coli* Pfl and its activating enzyme.** The immunological study described above strongly suggests that *C. pasteurianum* has a Pfl enzyme. Since the cross-reacting polypeptides are often observed as a doublet (see below), this, together with their clear interaction with anti-Pfl antibodies, suggested that Pfl from *C. pasteurianum* and that from *E. coli* are similar. We therefore designed two degenerate oligonucleotides (PCR2-act, 5′-CCCCGGGCA
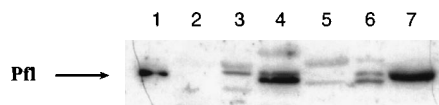


FIG. 1. Western immunoblot of cell extracts from different gram-positive bacteria challenged with anti-*E. coli* Pfl polyclonal antibodies. Extracts (100 μg of protein [13], except for *E. coli* FM420, of which only 10 μg was applied) were separated by SDS-PAGE in a gel containing 8% acrylamide (11). Lane 1, *E. coli* FM420 (wild type); lane 2, *E. coli* RM202 (with *pfl* deleted); lane 3, *C. pasteurianum*; lane 4, *C. butyricum*; lane 5, *S. sanguis*; lane 6, *S. salivarius*; lane 7, *E. faecalis*.

* Corresponding author. Mailing address: Nitrogen Fixation Laboratory, John Innes Centre, Norwich Research Park, Colney Ln., Norwich NR4 7UH, United Kingdom. Phone: 44 1603 456900, ext. 2750. Fax: 44 1603 454970. Electronic mail address: gary.sawers@bbsrc.ac.uk.

```
CCTATAGCTATATTCATATAGCTTAATCTAAATGAGTAATATGTTATATTAAAATATTTACAATATTCCAAGTAA   75

GATTTATTTTTGTTTTTGAAGTAAAATTTTTACAAAACATAAAAAAATAAATACT[TTTAAT]TATAAA[TAGAGA]TA  150

TGA[TATAAT][GTGAAAT]ATGAAAAGTCTAACTTATGTTAATTAAGATATTATTTATATTGTTGAAATAAGATGATT  225

AATTAGTTATAATATCTATATAATAAAAAATTATTAATTAGTAACAATTAAAAACAATTTCATTATAAATTAATTA  300

                                  pfl
GCATAGAATAAGGAGATGATATT TTG TTT AAA CAA TGG GAA GGC TTT CAA GAT GGA GAA TGG      362
         SD              M   F   K   Q   W   E   G   F   Q   D   G   E   W       13

ACT AAT GAT GTA AAT GTA AGA GAC TTT ATC CAA AAA AAT TAT AAA GAG TAT ACT GGA GAT  422
 T   N   D   V   N   V   R   D   F   I   Q   K   N   Y   K   E   Y   T   G   D    33

AAA AGC TTT TTA AAG GGT CCT ACA GAA AAA ACT AAA AAA GTT TGG GAT AAA GCA GTT TCA  482
 K   S   F   L   K   G   P   T   E   K   T   K   K   V   W   D   K   A   V   S    53

CTT ATT TTA GAG GAA TTA AAA AAG GGA ATA CTT GAT GTA GAT ACA GGA ACA ATC TCT GGA  542
 L   I   L   E   E   L   K   K   G   I   L   D   V   D   T   E   T   I   S   G    73

ATA AAT AGC TTT AAA CCT GGA TAT TTG GAT AAA GAT AAT GAA GTT ATA GTA GGT TTT CAG  602
 I   N   S   F   K   P   G   Y   L   D   K   D   N   E   V   I   V   G   F   Q    93

ACG GAT GCA CCA TTA AAG AGA ATA ACT AAT CCT TTT GGT GGA ATT AGA ATG GCA GAA CAA  662
 T   D   A   P   L   K   R   I   T   N   P   F   G   G   I   R   M   A   E   Q   113

TCA CTT AAA GAA TAT GGC TTT AAA ATT AGT GAT GAA ATG CAC AAT ATA TTT ACA AAT TAT  722
 S   L   K   E   Y   G   F   K   I   S   D   E   M   H   N   I   F   T   N   Y   133

AGA AAA ACA CAC AAT CAA GGT GTA TTT GAT GCT TAT TCA GAA GAA AGG ATA GCT AGG      782
 R   K   T   H   N   Q   G   V   F   D   A   Y   S   E   E   R   I   A   R       153

TCT GCT GGA GTA TTA ACA GGA CTT CCA GAT GCC TAT GGA AGA GGA AGA ATA ATA GGA GAT  842
 S   A   G   V   L   T   G   L   P   D   A   Y   G   R   G   R   I   I   G   D   173

TAT AGA AGA GTA GCC CTG TAC GGA ATA GAT TTC TTA ATA CAA GAG AAA AAG AAA GAT TTA  902
 Y   R   R   V   A   L   Y   G   I   D   F   L   I   Q   E   K   K   K   D   L   193

AGT AAT CTT AAA GGT GAT ATG CTG GAT GAG CTT ATA AGA CTT AGA GAA GTA TCG GAA      962
 S   N   L   K   G   D   M   L   D   E   L   I   R   L   R   E   E   V   S   E   213

CAA ATA AGA GCT CTT GAT GAG ATT AAA AAA ATG GCT CTT TCT TAT GGT GTG GAT ATA TCA 1022
 Q   I   R   A   L   D   E   I   K   K   M   A   L   S   Y   G   V   D   I   S   233

AGG CCA GCT GTT AAT GCT AAA GAA GCA GCT CAG TTT TTG TAT TTT GGG TAT TTA GCA GGT 1082
 R   P   A   V   N   A   K   E   A   A   Q   F   L   Y   F   G   Y   L   A   G   253

GTA AAG GAA AAT AAT GGT GCG GCT ATG TCA CTT GGA AGG ACA AGT ACG TTC TTA GAT ATT 1142
 V   K   E   N   N   G   A   A   M   S   L   G   R   T   S   T   F   L   D   I   273

TAT ATT GAA AGA GAT TTA GAG CAA GGA CTT ATT ACA GAA GAC GAG GCA CAA GAA GTA ATA 1202
 Y   I   E   R   D   L   E   Q   G   L   I   T   E   D   E   A   Q   E   V   I   293

GAT CAA TTT ATT ATA AAA CTT AGA CTG GTA AGA CAT TTA AGA ACA CCA GAA TAT AAT GAA 1262
 D   Q   F   I   I   K   L   R   L   V   R   H   L   R   T   P   E   Y   N   E   313

TTG TTT GCA GGA GAT CCA ACT TGG GTT ACT GAA TCT ATA GCT GGA GTT GGA ATA GAT GGA 1322
 L   F   A   G   D   P   T   W   V   T   E   S   I   A   G   V   G   I   D   G   333

AGA AGC CTT GTT ACA AAG AAT TCT TTT AGA TAT CTT CAT ACA TTA ATA AAT CTT GGA TCA 1382
 R   S   L   V   T   K   N   S   F   R   Y   L   H   T   L   I   N   L   G   S   353

GCA CCT GAA CCT AAT ATG ACA GTA TTG TGG TCA GAA AAT TTA CCA GAA AGC TTT AAA AAA 1442
 A   P   E   P   N   M   T   V   L   W   S   E   N   L   P   E   S   F   K   K   373

TTC TGT GCG GAA ATG TCT ATT TTA ACT GAT TCA ATT CAG TAT GAA AAT GAT GAT ATT ATG 1502
 F   C   A   E   M   S   I   L   T   D   S   I   Q   Y   E   N   D   D   I   M   393

AGA CCT ATA TAT GGA GAT GAT TAT GCA ATT GCC TGC TGT GTA TCT GCT ATG AGA GTA GGA 1562
 R   P   I   Y   G   D   D  _Y___A___I___A___C___C___V_  S   A   M   R   V   G   413

AAA GAC ATG CAA TTC TTT GGA GCT AGA TGT AAT CTT GCT AAA TGT CTC CTT TTA GCA ATA 1622
 K   D   M   Q   F   F   G   A   R   C   N   L   A   K   C   L   L   L   A   I   433

AAT GGT GGT GTA GAT GAA AAG AAA GGT ATA AAA GTT GTT CCT GAT ATT GAG CCT ATA ACA 1682
 N   G   G   V   D   E   K   K   G   I   K   V   V   P   D   I   E   P   I   T   453

GAT GAA GTA TTA GAT TAT GAA AAA GTA AAG GAA AAT TAT TTT AAA GTT CTT GAA TAT ATG 1742
 D   E   V   L   D   Y   E   K   V   K   E   N   Y   F   K   V   L   E   Y   M   473

GCA GGA CTC TAT GTT AAT ACT ATG AAT ATA ATA CAC TTT ATG CAT GAT AAA TAT GCC TAT 1802
 A   G   L   Y   V   N   T   M   N   I   I   H   F   M   H   D   K   Y   A   Y   493

GAA GCT AGT CAA ATG GCA CTT CAT GAT ACC AAG GTA GGA AGA CTT ATG GCT TTT GGT ATT 1862
 E   A   S   Q   M   A   L   H   D   T   K   V   G   R   L   M   A   F   G   I   513

GCA GGA TTT TCT GTA GCA GCT GAT TCC CTA AGT GCA ATA CGA TAT GCT AAA GTA AAA CCA 1922
 A   G   F   S   V   A   A   D   S   L   S   A   I   R   Y   A   K   V   K   P   533

ATA AGA GAA AAT GGT ATA ACT GTA GAT TTT GTT AAA GAG GGA GAT TTC CCT AAG TAT GGT 1982
 I   R   E   N   G   I   T   V   D   F   V   K   E   G   D   F   P   K   Y   G   553

AAT GAC GAT GAT AGA GTA GAT AGT ATA GCT GTA GAA ATT GTT GAA AAA TTT TCT GAT GAA 2042
 N   D   D   D   R   V   D   S   I   A   V   E   I   V   E   K   F   S   D   E   573

TTA AAG AAG CAT CCA ACT TAT AGG AAT GCT AAA CAT ACG CTT TCA GTG CTT ACT ATT ACA 2102
 L   K   K   H   P   T   Y   R   N   A   K   H   T   L   S   V   L   T   I   T   593

TCA AAT GTA ATG TAC GGT AAA AAG ACG GGT ACA ACA CCA GAT GGA AGA AAG GTT GGA GAA 2162
 S   N   V   M   Y   G   K   K   T   G   T   T   P   D   G   R   K   V   G   E   613

CCA TTA GCA CCT GGG GCT AAT CCA ATG CAT GGA AGA GAT ATG GAA GGA GCA TTA GCA TCC 2222
 P   L   A   P   G   A   N   P   M   H   G   R   D   M   E   G   A   L   A   S   633

TTA AAC TCA GTT GCT AAA GTG CCA TAT GTA TGC TGT GAA GAT GGA GTT TCA AAT ACA TTT 2282
 L   N   S   V   A   K   V   P   Y   V   C   C   E   D   G   V   S   N   T   F   653

TCT ATA GTA CCA GAT GCA TTA GGA AAT GAT CAT GAT GTA AGA ATA AAT AAT TTA GTT AGT 2342
 S   I   V   P   D   A   L   G   N   D   H   D   V   R   I   N   N   L   V   S   673

ATT ATG GGT GGA TAT TTT GGG CAA GGA GCT CAT CAC TTG AAT GTT AAC GTG CTA AAT AGA 2402
 I   M   G   G   Y   F   G   Q   G   A   H   H   L   N   V   N   V   L   N   R   693

GAA ACA TTA ATT GAT GCT ATG AAT AAT CCT GAT AAG TAT CCA ACG CTT ACT ATA AGA GTT 2462
 E   T   L   I   D   A   M   N   N   P   D   K   Y   P   T   L   T  _I___R___V_   713

TCA GGA TAT GCT GTC AAT TTT AAT AGA CTT TCA AAG GAT CAT CAA AAA GAA GTT ATA AGT 2522
_S___G___Y___A___V_  N   F   N   R   L   S   K   D   H   Q   K   E   V   I   S   733

AGA ACT TTT CAT GAA AAA TTA TAATAAATAGAGGCTGTCTCACAATTTATAAATTGTGAGACAGCTTT    2590
 R   T   F   H   E   K   L                                                     740

TTTGATATTTTACTATTAAAATGGTTATTAAAATCTAATTTTAAATAATTAAATCAGTACATTTAATTCCTTTTG   2665

AGATAACTTATTTGTTATATTAGTATAAGTATGTATCTAACGATTTAGATAATAATAGTAATAAATTATTAATTA   2740

AATATTAGTTATTGAATAAAATATAAAACTAACTAGAACTAATTCAATGTATGGAAATTTATAAAATTATAT[TTA   2815

TCAA]ATAATACTTAAAAA[TATGAT]ATACTAAATTGTATATATTTTTTTAACAATACTAATAATTTGATTATTTAT   2890

TAAATTGTAGGTAAGTTTAAAATAATATATAAAAGTATTATTAAGTTAATTAACATTTATAAATGGCAGGGAGAA   2965
                                                                       SD
                    act
ATAACT ATG GTA ATG GGA AGA ATT CAT TCA ATA GAA AGT ATG GGG CTT GTA GAT GGT CCA 3025
        M   V   M   G   R   I   H   S   I   E   S   M   G   L   V   D   G   P    18

GGT ATT AGA ACA GTA GTA TTT TTT CAA GGC TGT GGA TTA AGA TGT TCA TAT TGC CAT AAT 3085
 G   I   R   T   V   V   F   F   Q   G   C   G   L   R   C   S   Y   C   H   N   38

CCT GAT ACA TGG AAT ATG GCA GGC GGG AAA GAG TTA GCA GCA GAA GAG TTG TTG AAA AAG 3145
 P   D   T   W   N   M   A   G   G   K   E   L   T   A   E   E   L   L   K   K   58

TTA TTG AGA TTT AAG CCT TAT TTT GAT AGA TCA GGA GGT GGG GTA ACT TTT TCT GGC GGG 3205
 L   L   R   F   K   P   Y   F   D   R   S   G   G   G   V   T   F   S   G   G   78

GAA GTA CTT TTA CAA CCT GAA TTT TTA ATA GAC ATA TTA AAA TTG TGC AAA GAA CAG GGT 3265
 E   V   L   L   Q   P   E   F   L   I   D   I   L   K   L   C   K   E   Q   G   98

ATT CAC ACA GCT ATT GAT ACA GCT GGG TAT GGA TAT GGA AAT TAT GAA GAA ATT TTA AAA 3325
 I   H   T   A   I   D   T   A   G   Y   G   Y   G   N   Y   E   E   I   L   K   118

CAT ACA GAT TTG GTA TTA TTA GAT ATA AAA CAT GTA GAT GAT GGA TAC AAA TGT ATT      3385
 H   T   D   L   V   L   L   D   I   K   H   V   D   D   D   G   Y   K   C   I   138

ACT GGA AAG GGT AAA AGA GGT TTT GAT GAT TTC TTA AAG GCA GTA GAA AAT ATA GGT GTA 3445
 T   G   K   G   K   R   G   F   D   D   F   L   K   A   V   E   N   I   G   V   158

AAG GTA TGG ATA AGA CAT GTA ATA GTT CCT ACA TTA ACT GAT TCT AAA GAA AAT ATA AGA 3505
 K   V   W   I   R   H   V   I   V   P   T   L   T   D   S   K   E   N   I   R   178

AAA CTG GCT AAT ATC ATA AAG AAT ATA AGA AAT GTA GAA AAA GTA GAG TTA TTG CCT TAT 3565
 K   L   A   N   I   I   K   N   I   R   N   V   E   K   V   E   L   L   P   Y   198

CAT ACT TTA GGG ATA AAT AAA TAT GAA AAA TTA AAT TTA GAT TAT AAA CTT AGA GAT ATA 3625
 H   T   L   G   I   N   K   Y   E   K   L   N   L   D   Y   K   L   R   D   I   218

GAA GCA ATG GAT AAA GAA AAG CGC AAA AAG CTT GAG AAA TAT TTA AAA GAA TTA TTA GAA 3685
 E   A   M   D   K   E   K   R   K   K   L   E   K   Y   L   K   E   L   L   E   238

TAGATTATATTGATTTTAGAGGTGATTTATAGAAATTTATTGTTATATGTTTTATAACGAATGAATAATACATTA   3760

ATAATTAAATGAATATAACATTAGATGGTGGTGCAAAGATGAGATTAATTCCCATACAATGCGTAAAAGAAAATA   3835

CTAGATTGGCAAAGTCTATATATGAT                                                    3861
```

FIG. 2. Complete nucleotide and amino acid sequences of the *pfl* and *act* genes of *C. pasteurianum*. The putative Shine-Dalgarno (SD) sequences are underlined, the −10 and −35 consensus promoter sequences are boxed, and the transcription initiation sites are marked with angled arrows. The location of a potential rho-independent terminator structure is marked with converging arrows above the sequence. Broken underlining indicates peptide sequences used to derive the degenerate oligonucleotides.

TABLE 1. Bacteria and plasmids used in this study

| Organism | Relevant characteristic(s) | Source or reference |
|---|---|---|
| Bacteria | | |
| *C. butyricum* DSM552 | Wild type | DSM[a] |
| *C. pasteurianum* DSM525 | Wild type | DSM |
| *S. sanguis* DSM20066 | Wild type | DSM |
| *S. salivarius* ATCC 13419 | Wild type | R. Wirth |
| *E. faecalis* | Wild type | 6 |
| *E. coli* | | |
| FM420 | F⁻ *araD139* Δ(*argF-lac*)*U169 ptsF25 deoC1 relA1 flbB530 rpsL150 recA56* λ⁻ | 37 |
| RM202 | FM420 Δ*pfl25* Ω(*pfl*::Cam^r) | 22 |
| JM109 | *recA1 endA1 gyrA96 thi hsdR17 supE44 relA1* F(*lac-proAB*) | 35 |
| K38 | F⁻ HfrC *phoA4 pit-10 tonA22 ompF627 relA1* λ⁺ | 14 |
| Plasmids | | |
| pBR322 | Ap^r Tc^r | 3 |
| pUC19 | Ap^r *lacIPOZ*⁺ | 35 |
| pT7-4 | Ap^r T7φ10 | 27 |
| pT7-5 | Ap^r T7φ10 | 27 |
| pT7-6 | Ap^r T7φ10 | 27 |
| pBCP-1 | pBR322 carrying 3.3-kb DNA insert from *C. pasteurianum* chromosome | This work |
| pBCP-2 | pBR322 carrying 3.2-kb DNA insert from *C. pasteurianum* chromosome | This work |
| pBCP-4 | pBR322 carrying 3.5-kb DNA insert from *C. pasteurianum* chromosome | This work |
| pUACT-1 | pUC19 Ap^r *act*⁺ | This work |
| pTACT-1 | pT7-5 Ap^r *act*⁺ | This work |
| pTPFL-1 | pT7-6 Ap^r *pfl*⁺ | This work |

*[a]* Deutsche Sammlung von Mikroorganismen.

ATAAATTTCTATAAATCACCTC-3′, and PCR3-act, 5′-GG GATCCAGGGAGAAATAACTATGGTAATGGG-3′) on the basis of amino acid sequence motifs which are essential for the activity of the *E. coli* Pfl enzyme. The first motif (YAIACCV) includes the neighboring cysteinyl residues at amino acid positions 418 and 419 on the polypeptide chain (17), and the second motif (TIRVSGY) includes Gly-734, which carries the free radical in the active Pfl enzyme (33). Performing a PCR with these degenerate oligonucleotides and *C. pasteurianum* chromosomal DNA delivered a single amplified product of 970 bp, which correlates precisely with that predicted from the amino acid sequence of the *E. coli* enzyme (data not shown). This PCR product was then purified and used to probe a *C. pasteurianum* gene library. The gene library was constructed with chromosomal DNA fragments generated by partial digestion with the restriction enzyme *Sau*3A. The DNA (~25 μg) was fractionated in an agarose gel, and fragments in the size range of 3 to 10 kb were eluted and ligated into the *Bam*HI site of pBR322 (18). From a total of 10,000 colonies screened, three overlapping clones (pBCP-1, pBCP-2, and pBCP-4; Table 1) which showed a strong hybridization with the PCR-generated *pfl* probe were isolated. Nucleotide sequence analysis (19) revealed that the inserts of the clones overlapped and that they included portions of a gene whose 740-amino-acid product (molecular weight, 83,104) was 60% identical (65% overall similarity) to Pfl from *E. coli*. Figure 2 shows the complete nucleotide sequence of a 4-kb portion of the *C. pasteurianum* chromosome which includes the *pfl* gene. The initiation codon of the gene is UUG, which is used as the initia-

tion codon in an estimated 1% of genes (5). The gene is preceded by a Shine-Dalgarno sequence which matches the consensus sequence in five of six crucial nucleotides (24). No open reading frame could be detected in the 300 bp of DNA that extended upstream of the *pfl* gene. Directly 3′ to the *pfl* coding sequence is a potential rho-independent terminator structure. These data suggest that the *pfl* gene does not form part of an operon and that it is monocistronic. The GC content of the DNA was approximately 28%, which is in the range characteristic of genes from *C. pasteurianum* (36). The codon usage also showed a strong bias toward the use of A or U in the wobble position, again a characteristic feature of clostridiae (36). The two amino acid sequence motifs (YAIACCV and TIRVSGY) are completely conserved between the *E. coli* and *C. pasteurianum* enzymes (Fig. 2), which is consistent with the PCR amplification with degenerate oligonucleotides of the 970-bp DNA fragment from the *C. pasteurianum* chromosome.

Examination of the DNA sequence downstream of the *pfl* gene revealed an intergenic region of 424 bp, which has no coding capacity, followed by an open reading frame that has the potential to code for a protein with a molecular weight of 27,116 (Fig. 2). The initiation codon is AUG. The product of the downstream gene shows 46% identity (52% similarity) with the Pfl-activating enzyme of *E. coli*. As with the Pfl enzyme, the clostridial Act enzyme is marginally (3%) smaller than its *E. coli* homolog. These findings indicate that the chromosomal organizations of the *pfl* and *act* genes of *E. coli* and *C. pasteurianum* are identical.

**Specific synthesis of the *C. pasteurianum* Pfl and Act polypeptides in *E. coli*.** The complete *pfl* and *act* genes were cloned into the T7 promoter-polymerase vectors (27), and the synthesis of their respective gene products was examined by SDS-PAGE after specific labelling with [³⁵S]methionine (25) (Fig. 3). Pfl migrated as an 85-kDa protein, which is in good agreement with its deduced molecular mass. The Act protein migrated as a 28-kDa polypeptide, and this is also in good agreement with the predicted molecular mass of the protein (Fig. 3).

**Transcript mapping of the *pfl* and *act* genes.** Primer extension (1, 23) with oligonucleotides that hybridized in the 5′ portion of the *pfl* and *act* genes and total RNA isolated from an exponentially growing culture of *C. pasteurianum* was performed (Fig. 4). Two major transcription start sites which were separated from each other by 11 bp were identified for the *pfl* gene; a weak signal was also observed between these two ini-
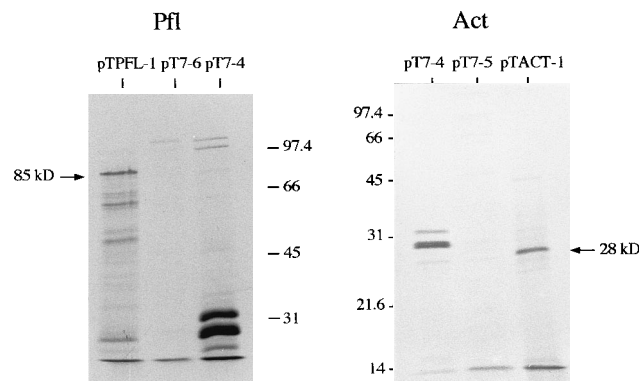


FIG. 3. Specific [³⁵S]methionine labelling of the Pfl and Act proteins of *C. pasteurianum* in the *E. coli* T7 promoter-polymerase system. The extract containing Pfl (labelled 85 kD) was separated by SDS–8% PAGE, while the extract containing Act (labelled 28 kD) was separated by SDS–12% PAGE. The *pfl* gene was cloned in the vector pT7-6, and the *act* gene was cloned in pT7-5. Molecular mass markers (in kilodaltons) are shown at the side of each gel.
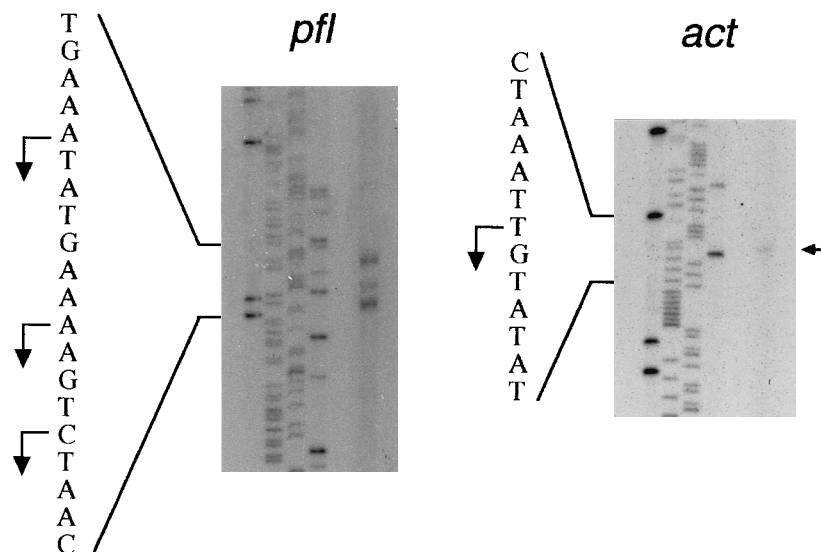
FIG. 4. Determination of the transcription initiation sites of the *C. pasteurianum pfl* and *act* genes. The angled arrows designate the transcription start sites. RNA was isolated from exponentially growing minimal medium cultures. Reaction products were separated in denaturing 8% polyacrylamide gels (18). The horizontal arrow designates the location of the *act* gene primer extension product.

tiation sites (Fig. 4). The sites were located approximately 150 bp upstream of the translation initiation codon of the *pfl* gene (Fig. 2). A perfect *E. coli* −10 box sequence (TATAAT) was centered at −8 relative to the upper start site which initiated at an A residue. There was also a putative *E. coli* −35 box sequence, but this was conserved less well than the −10 box (Fig. 3). The lower start site had poorly conserved −35 and −10 sequences and may result from the processing of the major transcript or from reverse transcriptase stalling.

The *act* gene had a single transcription start site located 123 bp upstream of the translation initiation codon of the *act* gene (Fig. 4). The putative −35 box sequence was TTATCA, and the −10 box sequence was TATGAT (Fig. 3). The results demonstrate that the *C. pasteurianum pfl* and *act* genes have transcriptional organizations similar to those in *E. coli*.

**Conclusions.** The primary structures of the Pfl and Act enzymes are highly conserved between *C. pasteurianum* and *E. coli*. Of particular relevance are the identities of the Pfl amino acid sequences around the active-site cysteinyl residues and the glycyl radical (16, 33). These findings suggest very similar catalytic mechanisms for both enzymes.

The genetic organizations of the *pfl* locus in *C. pasteurianum* and *E. coli* are similar; in both organisms the *pfl* and *act* genes are adjacent on the chromosome, but they are transcribed independently. In *E. coli* the genes are differentially regulated (20–22). Future studies will determine whether this is also the case for *C. pasteurianum*. From an evolutionary viewpoint, the physical proximity of the genes is advantageous to any organism that acquires them by horizontal gene transfer, because Pfl is without function in the absence of the activating enzyme. The major differences between these loci are that in *E. coli* the *pfl* gene forms an operon with the *focA* gene and that the expression of this operon is controlled by multiple promoters (21, 23, 26). Because the role of FocA is the maintenance of pH homeostasis (formate export) in *E. coli* (26), it is undesirable for *C. pasteurianum* to have a protein that exports a valuable intermediate for anabolism.

The reason that Pfl in *C. pasteurianum* went undetected in earlier studies may be due to the fact that the initiation codon of the *pfl* gene is UUG, which has a reduced translation initi-

ation efficiency compared with that of AUG codons (5). These studies, however, clearly demonstrate that *C. pasteurianum* has two routes of formate synthesis: through the actions of a $CO_2$ reductase (7, 29) and Pfl. Detailed physiological and biochemical characterizations of these two pathways will be the focus of future studies.

**Nucleotide sequence accession number.** The nucleotide sequence reported in this paper has been submitted to the GenBank and EMBL databases under accession number X93463.

**REFERENCES**

1. **Ausubel, F. M., R. Brent, R. E. Kingston, D. E. Moore, J. G. Seidman, J. A. Smith, and K. Struhl.** 1991. Current protocols in molecular biology. John Wiley & Sons, New York.
2. **Balch, W. E., and R. S. Wolfe.** 1976. New approaches to the cultivation of methanogenic bacteria: 2-mercaptoethane sulfonic acid (HS-CoM)-dependent growth of *Methanobacterium ruminantium* in a pressurized atmosphere. Appl. Environ. Microbiol. **32:**781–791.
3. **Bolivar, F., R. L. Rodriguez, P. J. Greene, M. C. Betlach, H. L. Heyneker, H. W. Boyer, J. H. Crosa, and S. Falkow.** 1977. Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. Gene **2:**95–113.
4. **Conradt, H., M. Hohmann-Berger, H. P. Hohmann, H. P. Blaschkowski, and J. Knappe.** 1984. Pyruvate formate-lyase (inactive form) and pyruvate-formate-lyase-activating enzyme of *Escherichia coli*: isolation and structural properties. Arch. Biochem. Biophys. **228:**133–142.
5. **Gold, L., and G. Stormo.** 1987. Translation initiation, p. 1302–1307. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
6. **Ike, Y., R. A. Craig, B. A. White, Y. Yagi, and D. B. Clewell.** 1983. Modification of *Streptococcus faecalis* sex pheromones after acquisition of plasmid DNA. Proc. Natl. Acad. Sci. USA **80:**5369–5373.
7. **Jungermann, K., H. Kirschniawy, and R. K. Thauer.** 1970. Ferredoxin dependent $CO_2$ reduction to formate in *Clostridium pasteurianum*. Biochem. Biophys. Res. Commun. **41:**682–689.
8. **Kessler, D., W. Herth, and J. Knappe.** 1992. Ultrastructure and pyruvate-formate-lyase quenching property of the multienzymic AdhE protein of *Escherichia coli*. J. Biol. Chem. **267:**18073–18079.

9. **Knappe, J., H. P. Blaschkowski, P. Gröbner, and T. Schmitt.** 1974. Pyruvate formate-lyase of *Escherichia coli*: the acetyl enzyme intermediate. Eur. J. Biochem. **50:**253–263.

10. **Knappe, J., and G. Sawers.** 1990. A radical-chemical route to acetyl-CoA: the anaerobically induced pyruvate formate-lyase system of *Escherichia coli*. FEMS Microbiol. Rev. **75:**383–398.

11. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227:**680–685.

12. **Lindmark, D. G., P. Paolella, and N. P. Wood.** 1969. The pyruvate formate-lyase system of *Streptococcus faecalis*: purification of the formate-pyruvate exchange enzyme. J. Biol. Chem. **244:**3606–3612.

13. **Lowry, O. H., N. J. Rosebrough, A. L. Farr, and R. J. Randall.** 1951. Protein measurement with the Folin phenol reagent. J. Biol. Chem. **193:**265–275.

14. **Lyons, L. B., and N. D. Zinder.** 1972. The genetic map of the filamentous phage f1. Virology **49:**45–60.

15. **O'Brien, R. W., and J. G. Morris.** 1971. Oxygen and the growth and metabolism of *Clostridium acetobutylicum*. J. Gen. Microbiol. **68:**307–318.

16. **Plaga, W., R. Frank, and J. Knappe.** 1988. Catalytic-site mapping of pyruvate formate-lyase. Hypophosphite reaction on the acetyl-enzyme intermediate affords carbon-phosphorus bond synthesis (1-hydroxyethylphosphonate). Eur. J. Biochem. **178:**445–450.

17. **Rödel, W., W. Plaga, R. Frank, and J. Knappe.** 1988. Primary structures of *Escherichia coli* pyruvate formate-lyase and pyruvate-formate-lyase-activating enzyme deduced from the DNA nucleotide sequences. Eur. J. Biochem. **177:**153–158.

18. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

19. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.

20. **Sauter, M., and R. G. Sawers.** 1990. Transcriptional analysis of the gene encoding pyruvate formate-lyase-activating enzyme of *Escherichia coli*. Mol. Microbiol. **4:**355–363.

21. **Sawers, G.** 1993. Specific transcriptional requirements for positive regulation of the anaerobically inducible *pfl* operon by ArcA and FNR. Mol. Microbiol. **10:**737–747.

22. **Sawers, G., and A. Böck.** 1988. Anaerobic regulation of pyruvate formate-lyase from *Escherichia coli* K-12. J. Bacteriol. **170:**5330–5336.

23. **Sawers, G., and A. Böck.** 1989. Novel transcriptional control of the pyruvate formate-lyase gene: upstream regulatory sequences and multiple promoters regulate anaerobic expression. J. Bacteriol. **171:**2485–2498.

24. **Shine, J., and L. Dalgarno.** 1974. The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome

25. **Studier, F. W., and B. A. Moffat.** 1986. Use of bacteriophage T7 RNA polymerase to direct high-level expression of cloned genes. J. Mol. Biol. **189:**113–130.

26. **Suppmann, B., and G. Sawers.** 1994. Isolation and characterization of hypophosphite-resistant mutants of *Escherichia coli*: identification of the FocA protein, encoded by the *pfl* operon, as a putative formate transporter. Mol. Microbiol. **11:**965–982.

27. **Tabor, S., and C. C. Richardson.** 1985. A bacteriophage T7 polymerase/promoter system for controlled exclusive expression of specific genes. Proc. Natl. Acad. Sci. USA **82:**1074–1078.

28. **Takahashi, S., K. Abbe, and T. Yamada.** 1982. Purification of pyruvate formate-lyase from *Streptococcus mutans* and its regulatory properties. J. Bacteriol. **149:**1034–1040.

29. **Thauer, R. K., F. H. Kirschniawy, and K. A. Jungermann.** 1972. Properties and function of the pyruvate formate-lyase reaction in clostridiae. Eur. J. Biochem. **27:**282–290.

30. **Thauer, R. K., E. Rupprecht, and K. Jungermann.** 1970. The synthesis of one-carbon units from $CO_2$ via a new ferredoxin dependent monocarboxylic acid cycle. FEBS Lett. **8:**304–307.

31. **Towbin, H., T. Staehelin, and J. Gordon.** 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci. USA **76:**4350–4354.

32. **Unkrig, V., F. A. Neugebauer, and J. Knappe.** 1989. The free radical of pyruvate formate-lyase. Characterization by EPR spectroscopy and involvement in catalysis as studied with the substrate-analogue hypophosphite. Eur. J. Biochem. **184:**723–728.

33. **Wagner, A. F. V., M. Frey, F. A. Neugebauer, W. Schäffer, and J. Knappe.** 1992. The free radical in pyruvate formate-lyase is located on glycine-734. Proc. Natl. Acad. Sci. USA **89:**996–1000.

34. **Wood, N. P., and K. Jungermann.** 1972. Inactivation of the pyruvate formate-lyase of *Clostridium butyricum*. FEBS Lett. **27:**49–52.

35. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33:**103–119.

36. **Young, M., W. L. Staudenbauer, and N. P. Minton.** 1989. Genetics of *Clostridium*, p. 63–103. *In* N. P. Minton and D. J. Clarke (ed.), Clostridia. Plenum Press, New York.

37. **Zinoni, F., A. Birkmann, T. C. Stadtman, and A. Böck.** 1986. Nucleotide sequence and expression of the selenocysteine-containing polypeptide of formate dehydrogenase (formate hydrogen-lyase-linked) from *Escherichia coli*. Proc. Natl. Acad. Sci. USA **83:**4650–4654.

binding sites. Proc. Natl. Acad. Sci. USA **71:**1342–1346.