Volume 14 Number 13 1986 Nucleic Acids Research

Nucleotide sequence and deduced amino acid sequence of *Escherichia coli* pyruvate oxidase, a lipid-activated flavoprotein

Charlotte Grabau and John E.Cronan, Jr

University of Illinois, Department of Microbiology, 131 Burrill Hall, 407 South Goodwin, Urbana, IL 61801, USA

Received 27 March 1986; Revised and Accepted 9 June 1986

ABSTRACT

The entire nucleotide sequence of the poxB (pyruvate oxidase) gene of Escherichia coli K-12 has been determined by the dideoxynucleotide (Sanger) sequencing of fragments of the gene cloned into a phage M13 vector. The gene is 1716 nucleotides in length and has an open reading frame which encodes a protein of Mr 62,018. This open reading frame was shown to encode pyruvate oxidase by alignment of the amino acid sequences deduced for the amino and carboxy termini and several internal segments of the mature protein with sequences obtained by amino acid sequence analysis. The deduced amino acid sequence of the oxidase was not unusually rich in hydrophobic sequences despite the peripheral membrane location and lipid binding properties of the protein. The codon usage of the oxidase gene was typical of a moderately expressed protein. The deduced amino acid sequence shares homology with the large subunits of the acetohydroxy acid synthase isozymes I, II,and III, encoded by the ilvB, ilvG, and ilvI genes of E. coli.

INTRODUCTION

Pyruvate oxidase of Escherichia coli is a peripheral membrane flavoenzyme, encoded by the poxB gene, which catalyzes the conversion of pyruvate to acetate and CO_2 (1, 2). This homotetrameric enzyme (Mr 62,000 subunit) has been well characterized, particularly with respect to its interactions with lipids. In the presence of the substrate pyruvate, and cofactor thiamine pyrophosphate, the addition of any of a wide variety of lipids or detergents results in a 20-25 fold increase in the specific activity of pyruvate oxidase and a 10-fold decrease in the Km for pyruvate (3-5). This activation is accompanied by the tight association of the activator with the enzyme. A similar extent of activation can be attained by limited proteolysis of the protein (in the presence of substrate and cofactor) during which a small Mr ca 3,000 peptide is released from the carboxyl terminus of the protein (6, 7). Following proteolytic activation, the enzyme can no longer bind nor be activated by lipid. Likewise, lipid activation protects against proteolytic cleavage (5). This observation suggests that the region involved in lipid binding includes the carboxyl terminal portion of the protein.

We (8) and Chang and Cronan (9, 10) have reported the isolation of several mutants deficient in pyruvate oxidase activity, three of which produce enzymes unable to function *in vivo* due to alterations in their interaction with lipid. As a necessary step in the understanding of the interaction of pyruvate oxidase with lipids we have determined the amino acid sequence of the wild type protein by deduction from the nucleotide sequence of the structural gene.

EXPERIMENTAL PROCEDURES

Bacterial strains, chemicals, and enzymes

Strain JM103 (Δ*lacpro*, *supE*, *thi*, *strA*, *sbcB15*, *endA*, *hsdR4*/F' *traD36*, *proAB*, *lacI*^{q}, *lacZ*ΔM15) has been previously described (11). Bacteriophage M13 vector mp11 (11), deoxynucleoside triphosphates and dideoxynucleoside triphosphates were purchased from PL Biochemicals. [α-^{32}P]dATP (800 Ci/mmol specific activity) was purchased from Amersham. Restriction endonucleases and T4-DNA ligase were from commerical sources. DNA polymerase I (large fragment) was obtained from Boehringer-Mannheim. The 15-base sequencing primer was from New England Biolabs. Oligonucleotides were synthesized using an Applied Biosystems 380 DNA Synthesizer in the laboratory of R. Gesteland, University of Utah.

DNA sequencing

Growth and transformation of strain JM103 was performed as described by Messing (11). Single stranded templates for sequencing were isolated from the supernatants of infected cultures of JM103 by precipitation of phage overnight at 4° in 2% polyethylene glycol 8000, 0.5 M NaCl, 3 mM EDTA, and 30 μg/ml RNAse. After centrifugation the pellet was resuspended in 10 mM Tris-HCl (pH 7.8), 1 mM EDTA, 0.2% Sarkosyl, containing 50 μg/ml proteinase K, and incubated 20 min at 55°. NaCl was added at a concentration of 0.4 M and the mixture was extracted once with phenol-chloroform, then twice with ether, and ethanol precipitated. Sequencing was carried out by the chain termination method described by Sanger *et al.* (12, 13).

Computer-assisted Sequence Analysis

Sequence analyses were performed utilizing University of Minnesota Apple II Sequence Analysis Programs, Version 2.1 (Larson and Messing) and Bionet software.

Sequence Strategy

We previously cloned the structural gene for pyruvate oxidase, *poxB*, on a 3.2 kilobase pair (kbp) DNA fragment into the *PstI* site of the plasmid pBR322

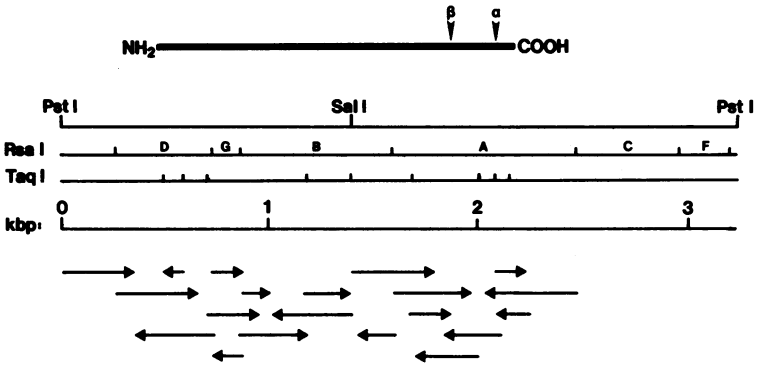

Figure 1. Restriction endonuclease sites and strategy of sequencing the poxB gene. The dark line at the top of the figure represents the protein in relation to the poxB gene on a 3.2 kbp PstI fragment in plasmid pCG7 (see text). α and β refer to the sites for α-chymotrypsin cleavage of the protein in the presence (α) and absence (β) of substrate and cofactor. The letters (A-G) correspond to the fragments generated by digestion of the plasmid by RsaI. Only the TaqI sites within the poxB gene are shown. The arrows denote sequencing direction and the length of the arrows are proportional to the number of nucleotides determined.

(14). The resulting plasmid, pCG7, was digested with restriction endonuclease RsaI, electrophoresed on a 1.8% agarose gel and the DNA fragments corresponding to insert were recovered from the gel using the DEAE cellulose paper method of Dretzen et al. (15). Each of the RsaI fragments (see Fig. 1) was then ligated into the HincII site of bacteriophage M13 vector mp11 and used to transform strain JM103. DNA containing the insert in both orientations was isolated from the phage and sequenced using a 15 base sequencing primer which hybridizes to a site upstream from the multiple cloning site.

The 3.2 kbp fragment carrying the complete poxB gene was also cloned into mp11w at the PstI site in both orientations and sequenced to give information on both ends of the 3.2 kbp fragment. The insert contains a single SalI site

Table 1. Oligonucleotides synthesized as sequencing primers.

Oligonucleotides Synthesized	Position #
5'TCAACGGCCTGTTCGATTGC 3'	245-264 (423-442)
5'GAAATAGCCGCTGCCAATTT 3'	317 ← 336 (495 ← 514)
5'CTGGTTTCCAGCCCGGAGCA 3'	388-407 (566-585)
5'CGCCAGTTCGATCACTTCAT 3'	1679 ← 1698 (1857 ← 1876)

The position number refers to the corresponding base number in the poxB gene. The numbers in parentheses correspond to the base number used in Fig. 2. The 2nd and 4th oligonucleotides are complimentary to the non-coding strand of the gene in the opposite orientation (arrows).

```
ACTGGCCTGCTCCTGCCATCCGCAGGGGGATTTGGTTCTCGCATAATCGCCTTATGCC
                                                                                          58
CGATGATATTCCTTTCATCGGGCTATTTAACCGTTAGTGCCTCCTTTCTCTCCCATCCCTTCCCCCTCCGTCAGATGAACTAAACTTGTTACCGTTATCACATTCAGGAGATGGAGAACC
                                                                                         178
1
MET Lys Gln Thr Val Ala Ala Tyr Ile Ala Lys Thr Leu Glu Ser Ala Gly Val Lys Arg Ile Trp Gly Val Thr Gly Asp Ser Leu Asn
ATG AAA CAA ACG GTT GCA GCT TAT ATC GCC AAA ACA CTC GAA TCG GCA GGG GTG AAA CGC ATC TGG GGA GTC ACA GGC GAC TCT CTG AAC
                                                                                         268
31
Gly Leu Ser Asp Ser Leu Asn Arg Met Gly Thr Ile Glu Trp Met Ser Thr Arg His Glu Glu Val Ala Ala Phe Ala Ala Gly Ala Glu
GGT CTT AGT GAC AGT CTT AAT CGC ATG GGC ACC ATC GAG TGG ATG TCC ACC CGC CAC GAA GAA GTG GCG GCC TTT GCC GCT GGC GCT GAA
                                                                                         358
61
Ala Gln Leu Ser Gly Glu Leu Ala Val Cys Ala Gly Ser Cys Gly Pro Gly Asn Leu His Leu Ile Asn Gly Leu Phe Asp Cys His Arg
GCA CAA CTT AGC GGA GAA CTG GCG GTC TGC GCC GGA TCG TGC GGC CCC GGC AAC CTG CAC TTA ATC AAC GGC CTG TTC GAT TGC CAC CGC
                                                                                         448
91
Asn His Val Pro Val Leu Ala Ile Ala Ala His Ile Pro Ser Ser Glu Ile Gly Ser Gly Tyr Phe Gln Glu Thr His Pro Gln Glu Leu
AAT CAC GTT CCG GTA CTG GCG ATT GCC GCT CAT ATT CCC TCC AGC GAA ATT GGC AGC GGC TAT TTC CAG GAA ACC CAC CCA CAA GAG CTA
                                                                                         538
121
Phe Arg Glu Cys Ser His Tyr Cys Glu Leu Val Ser Ser Pro Glu Gln Ile Pro Gln Val Leu Ala Ile Ala Met Arg Lys Ala Val Leu
TTC CGC GAA TGT AGT CAC TAT TGC GAG CTG GTT TCC AGC CCG GAG CAG ATC CCA CAA GTA CTG GCG ATT GCC ATG CGC AAA GCG GTG CTT
                                                                                         628
151
Asn Arg Gly Val Ser Val Val Val Leu Pro Gly Asp Val Ala Leu Lys Pro Ala Pro Glu Gly Ala Thr Met His Trp Tyr His Ala Pro
AAC CGT GGC GTT TCG GTT GTC GTG TTA CCA GGC GAC GTG GCG TTA AAA CCT GCG CCA GAA GGG GCA ACC ATG CAC TGG TAT CAT GCG CCA
                                                                                         718
181
Gln Pro Val Val Thr Pro Glu Glu Glu Glu Leu Arg Lys Leu Ala Gln Leu Leu Arg Tyr Ser Ser Asn Ile Ala Leu Met Cys Gly Ser
CAA CCA GTC GTG ACG CCG GAA GAA GAA GAG TTA CGC AAA CTG GCG CAA CTG CTG CGT TAT TCC AGC AAT ATC GCC CTG ATG TGT GGC AGC
                                                                                         808
211
Gly Cys Ala Gly Ala His Lys Glu Leu Val Glu Phe Ala Gly Lys Ile Lys Ala Pro Ile Val His Ala Leu Arg Gly Lys Glu His Val
GGC TGC GCG GGG GCG CAT AAA GAG TTA GTT GAG TTT GCC GGG AAA ATT AAA GCG CCT ATT GTT CAT GCC CTG CGC GGT AAA GAA CAT GTC
                                                                                         898
241
Glu Tyr Asp Asn Pro Tyr Asp Val Gly Met Thr Gly Leu Ile Gly Phe Ser Ser Gly Phe His Thr Met Met Asn Ala Asp Thr Leu Val
GAA TAC GAT AAT CCG TAT GAT GTT GGA ATG ACC GGG TTA ATC GGC TTC TCG TCA GGT TTC CAT ACC ATG ATG AAC GCC GAC ACG TTA GTG
                                                                                         988
271
Leu Leu Gly Thr Gln Phe Pro Tyr Arg Ala Phe Tyr Pro Thr Asp Ala Lys Ile Ile Gln Ile Asp Ile Asn Pro Ala Ser Ile Gly Ala
CTA CTC GGC ACG CAA TTT CCC TAC CGC GCC TTC TAC CCG ACC GAT GCC AAA ATC ATT CAG ATT GAT ATC AAC CCA GCC AGC ATC GGC GCT
                                                                                        1078
301
His Ser Lys Val Asp Met Ala Leu Val Gly Asp Ile Lys Ser Thr Leu Arg Ala Leu Leu Pro Leu Val Glu Glu Lys Ala Asp Arg Lys
CAC AGC AAG GTG GAT ATG GCA CTG GTC GGC GAT ATC AAG TCG ACT CTG CGT GCA TTG CTT CCA TTG GTG GAA GAA AAA GCC GAT CGC AAG
                                                                                        1168
331
Phe Leu Asp Lys Ala Leu Glu Asp Tyr Arg Asp Ala Arg Lys Gly Leu Asp Asp Leu Ala Lys Pro Ser Glu Lys Ala Ile His Pro Gln
TTT CTG GAT AAA GCG CTG GAA GAT TAC CGC GAC GCC CGC AAA GGG CTG GAC GAT TTA GCT AAA CCG AGC GAG AAA GCC ATT CAC CCG CAA
                                                                                        1258
361
Tyr Leu Ala Gln Gln Ile Ser His Phe Ala Ala Asp Asp Ala Ile Phe Thr Cys Asp Val Gly Thr Pro Thr Val Trp Ala Ala Arg Tyr
TAT CTG GCG CAG CAA ATT AGT CAT TTT GCC GCC GAT GAC GCT ATT TTC ACC TGT GAC GTT GGT ACG CCA ACG GTG TGG GCG GCA CGT TAT
                                                                                        1348
391
Leu Lys Met Asn Gly Lys Arg Arg Leu Leu Gly Ser Phe Asn His Gly Ser Met Ala Asn Ala Met Pro Gln Ala Leu Gly Ala Gln Ala
CTA AAA ATG AAC GGC AAG CGT CGC CTG TTA GGT TCG TTT AAC CAC GGT TCG ATG GCT AAC GCC ATG CCG CAG GCG CTG GGT GCG CAG GCG
                                                                                        1438
421
Thr Glu Pro Glu Arg Gln Val Val Ala Met Cys Gly Asp Gly Gly Phe Ser Met Leu Met Gly Asp Phe Leu Ser Val Val Gln Met Lys
ACA GAG CCA GAA CGT CAG GTG GTC GCC ATG TGC GGC GAT GGC GGT TTT AGC ATG TTG ATG GGC GAT TTC CTC TCA GTA GTG CAG ATG AAA
                                                                        β               1528
451                                                                     v
Leu Pro Val Lys Ile Val Val Phe Asn Asn Ser Val Leu Gly Phe Val Ala Met Glu Met Lys Ala Gly Gly Tyr Leu Thr Asp Gly Thr
CTG CCA GTG AAA ATT GTC GTC TTT AAC AAC AGC GTG CTG GGC TTT GTG GCG ATG GAG ATG AAA GCT GGT GGC TAT TTG ACT GAC GGC ACC
                                                                                        1618
481
Glu Leu His Asp Thr Asn Phe Ala Arg Ile Ala Glu Ala Cys Gly Ile Thr Gly Ile Arg Val Glu Lys Ala Ser Glu Val Asp Glu Ala
GAA CTA CAC GAC ACA AAC TTT GCC CGC ATT GCC GAA GCG TGC GGC ATT ACG GGT ATC CGT GTA GAA AAA GCG TCT GAA GTT GAT GAA GCC
                                                                                        1708
511
Leu Gln Arg Ala Phe Ser Ile Asp Gly Pro Val Leu Val Asp Val Val Val Ala Lys Glu Glu Leu Ala Ile Pro Pro Gln Ile Lys Leu
CTG CAA CGC GCC TTC TCC ATC GAC GGT CCG GTG TTG GTG GAT GTG GTG GTC GCC AAA GAA GAG TTA GCC ATT CCA CCG CAG ATC AAA CTC
                                    α                                                   1798
541                                 v
Glu Gln Ala Lys Gly Phe Ser Leu Tyr Met Leu Arg Ala Ile Ile Ser Gly Arg Gly Asp Glu Val Ile Glu Leu Ala Lys Thr Asn Trp
GAA CAG GCC AAA GGT TTC AGC CTG TAT ATG CTG CGC GCA ATC ATC AGC GGA CGC GGT GAT GAA GTG ATC GAA CTG GCG AAA ACA AAC TGG
                                                                                        1888
571
Leu Arg OC
CTA AGG TAA AAAGGGTGGCATTTCCCGTCATAATAAGGACATGCCATGATTGATTTACGCAGTGATACCGTTACCCGACCAAGCCG
                                                                                1974
```

Figure 2. Nucleotide sequence and derived amino acid sequence of the poxB gene. Deduced amino acids that agree with known amino acid sequences are overlined. A putative Shine-Delgarno sequence is underlined at position 164-168. α and β refer to the site of α-chymotrypsin cleavage of the protein in the presence (α) and absence (β) of substrate and cofactor. Regions of dyad symmetry are overlined with arrows. The amino acid sequence data are from references 6, 17, and 18.

which divides the insert into two fragments of 1.4 and 1.8 kbp. The vector mp11w also carries a SalI site between the PstI site (into which the poxB fragment was inserted) and the site to which the primer anneals (11). Digestion of the replicative form of the phage DNA carrying the insert in either of the two orientations followed by religation resulted in the deletion of the segment of DNA between the two SalI sites. Sequencing using the universal primer yielded the DNA sequence to either side of the internal SalI site.

Due to the large size of some of the RsaI fragments, sequencing in from the ends did not yield overlapping sequences in the middle of the fragment. These fragments (RsaIA and RsaIB, Fig. 1) were digested with TaqI to generate smaller fragments which were then cloned into M13mp11 and sequenced completely.

Oligonucleotides (Table 1) corresponding to the end of fragments RsaI D, both ends of RsaI G and to an internal region of RsaI A were synthesized and used as primers for sequencing to eliminate any gaps that existed between the fragments. This allowed the alignment (by matching the overlapping sequences) of all of the RsaI fragments.

RESULTS AND DISCUSSION

Criteria for Reading Frame Choice

We deduced the amino acid sequence of E. coli pyruvate oxidase from the nucleotide sequence of the poxB gene (Fig. 2). The translational reading frame of the oxidase was established by the following criteria: (i) Only one reading frame could encode a protein of the size of the oxidase. This reading frame predicted a protein of Mr 62,018. Since the Mr of the monomer has been estimated to be 60,000 by SDS-polyacrylamide gel electrophoresis (5-7) and 63,000 by hydrodynamic studies of the tetramer (16), our deduced Mr is in excellent agreement. (ii) The sequences of five segments of the protein have been obtained by direct amino acid sequence analyses; these include the N-terminus (6) and C-terminus (17), the N-terminal sequences of the α (6) and β (18) peptides released by limited cleavage with α-chymotrypsin, and a peptide released by treatment with thermolysin (18). The amino acid sequences and relative positions of all of these segments correspond exactly to the deduced sequence (Fig. 2). (iii) The total amino acid composition of both the intact protein and of the β-peptide agree very closely (especially the latter) with the compositions obtained from the deduced sequences (Table 2). (iv) The codon usage of the gene (Table 3) is consistent with the level of expression observed in vivo indicating the proper choice of reading frame had been made

Table 2. Amino acid composition of E. coli pyruvate oxidase.

Amino acid	Residues/PoxB subunit DNA	amino acid analysis		Residues/β peptide DNA	amino acid analysis
lys	29	29[a]	29[b]	5	5
his	18	16	18	1	1
arg	25	23	-	6	6
cys	10	11	10	1	1
asp/asn	28/17	47	44	6/2	8
thr	23	25	18	5	5
ser	33	30	14	4	4
glu/gln	38/20	56	59	10/3	12
pro	27	23	29	3	2
gly	48	44	49	9	9
ala	66	60	65	12	12
val	44	41	41	8	8
met	20	16	18	1	1
ile	32	30	32	9	8
leu	55	54	53	10	10
tyr	14	13	15	2	2
phe	20	20	19	3	3
trp	5	6	-	1	-

The amino acid composition values are those of Russell[a] (17) or Williams et al.[b] (2) for the E. coli W enzyme. The β peptide data are those of Recny (33). The dash indicates that no value has been reported.

(19). It should be noted that the amino acid compositions of the E. coli K-12 and E. coli W oxidases are essentially identical, in agreement with previous observations (biochemical and kinetic parameters, 2-D gel electrophoresis,

Table 3. Codon Usage of the poxB gene.

TTT-Phe 10	TCT-Ser 2	TAT-Tyr 10	TGT-Cys 3
TTC-Phe 10	TCC-Ser 5	TAC-Tyr 4	TGC-Cys 7
TTA-Leu 10	TCA-Ser 2	TAA- 1	TGA- 0
TTG-Leu 5	TCG-Ser 7	TAG- 0	TGG-Trp 5
CTT-Leu 5	CCT-Pro 2	CAT-His 7	CGT-Arg 7
CTC-Leu 4	CCC-Pro 3	CAC-His 11	CGC-Arg 17
CTA-Leu 5	CCA-Pro 12	CAA-Gln 10	CGA-Arg 0
CTG-Leu 26	CCG-Pro 10	CAG-Gln 10	CGG-Arg 0
ATT-Ile 15	ACT-Thr 2	AAT-Asn 4	AGT-Ser 4
ATC-Ile 17	ACC-Thr 9	AAC-Asn 13	AGC-Ser 13
ATA-Ile 0	ACA-Thr 5	AAA-Lys 25	AGA-Arg 0
ATG-Met 20	ACG-Thr 7	AAG-Lys 4	AGG-Arg 1
GTT-Val 10	GCT-Ala 9	GAT-Asp 17	GGT-Gly 13
GTC-Val 10	GCC-Ala 27	GAC-Asp 11	GGC-Gly 24
GTA-Val 4	GCA-Ala 8	GAA-Glu 27	GGA-Gly 5
GTG-Val 20	GCG-Ala 22	GAG-Glu 11	GGG-Gly 6

The number of times each codon is used in the coding sequence is given.

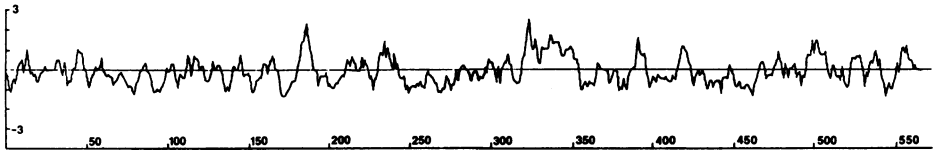

Figure 3. Hydrophilicity profile of the deduced amino acid sequence of pyruvate oxidase. The hydrophilicity values (averaged over 6 amino acids) are plotted against the sequence positions by method of Hopp and Woods (20). Hydrophilic values are negative.

spectral profiles, immunological assays, and protease cleavage patterns) that the oxidases from the two E. coli strains are indistinguishable.

Analysis of the Overall Structure

We have applied the predictive algorithms of Hopp and Woods (20) (Fig. 3) and of Chou and Fasman (21) (Fig. 4) to the deduced amino acid sequence. By the Hopp-Woods analysis the protein was not found to be unusually hydrophobic despite its ability to associate through hydrophobic interactions (22) with the surface of the E. coli cytoplasmic membrane and with synthetic lipid and detergent vesicles (Fig. 3). This finding is consistent with the numerous indications (16, 23, 24) that lipid binding is the property of a small segment of the oxidase.

The Chou-Fasman predictive rules suggest a protein containing 45% α-helical and 29% β structures (see Fig. 4), values somewhat greater than those obtained by circular dichroism [28% α-helix, 14% β (23)]. However, circular dichroism fails to detect short α-helices and β-sheets (21). It should be noted that the Chou-Fasman algorithm and two other methods were previously used to predict that amino acid residues 559-568 of the α-peptide region are structured into an amphipathic helix (6).

Possible Coenzyme Binding Sites

Porter and Kasper (25) have reported the conservation of amino acid sequences among several flavin binding proteins including E. coli fumarate reductase. We found no evidence for the presence of these conserved sequences in E. coli pyruvate oxidase. Hearst and Sauer (26) have noted the conservation of a Met-His sequence in several ubiquinone binding proteins. Pyruvate oxidase is a ubiquinone reductase (27) and has a Met-His sequence at residues 174-175 (Fig. 2). It is not known if these residues are involved with the interaction of pyruvate oxidase with ubiquinone.

Regulation of Gene Expression

E. coli pyruvate oxidase is a moderately expressed protein (about 1000

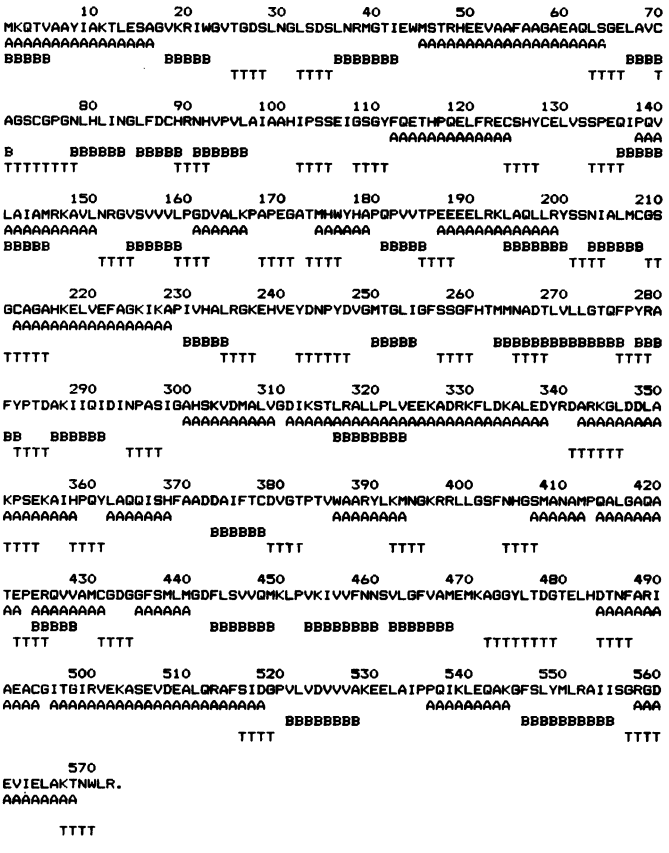

Figure 4. Secondary structure analysis of pyruvate oxidase by the method of Chou and Fasman (21). A = alpha helix, B = beta sheet, T = turn. Sequences with an average conformational parameter $P_\alpha \geq 1.03 > P_\beta$ are assigned an α-helix structure, those with $P_\beta \geq 1.05 > P_\alpha$ are labeled as a β-sheet (21). In some regions of the protein α helix and β-sheet were predicted for overlapping sequences of the peptide chain with essentially identical probabilities. This is depicted by the presence of both symbols.

molecules/cell) as calculated from the radioimmunoassay data of Chang and Cronan (28) and its pattern of codon usage was indistinguishable from that of proteins with similar levels of expression (29). An appropriately spaced region with a good match to the consensus translational initiation (Shine-Delgarno) sequence (30) was present at nucleotides 164-168 (Fig. 2) but no sequences with plausible matches to the consensus sequences (31) of E. coli promotors or factor independent terminators (32) were found in the DNA sequence. Thus, the poxB gene may either be in an operon or be dependent on

protein factors in addition to the RNA polymerase for transcription initiation and termination. It should be noted that the PstI insert of pCG7 apparently contains a promotor, since plasmids harboring the fragment in the opposite orientation are poxB$^+$ (14) and also give overexpression of the poxB gene product.

Homology to the large subunits of acetohydroxy acid synthases I, II, and III

The National Biomedical Research Foundation Protein Sequence database was searched for proteins homologous to pyruvate oxidase. The only proteins which exhibited significant homology were the large subunits of acetohydroxy acid synthases (AHAS) I, II, and III (enzymes respectively encoded by the ilvB, ilvG and ilvI genes which are involved in the pathways of branched chain amino acid synthesis). Recently Wek et al. (34) and Friden et al. (35) have reported extensive homology among the proteins encoded by the ilvB, ilvG, and ilvI genes of E. coli. Figure 5 shows the alignment of pyruvate oxidase with these three polypeptides. Pyruvate oxidase shares 28.9%, 29.6%, and 29.4% homology with the IlvB, IlvG, and IlvI proteins, respectively. As aligned in Figure 5, over 40% of the amino acids of pyruvate oxidase are homologous to at least one of the three AHAS proteins. Moreover, the regions of homology between PoxB and the AHAS proteins are generally located in the regions of homology conserved among the three AHAS polypeptides (see references 34 and 35 for details on homology among ilvG, ilvB, and ilvI).

The sequence data strongly suggest that pyruvate oxidase and the AHAS proteins are related enzymes. The four proteins show other striking similarities: all are flavoproteins (with the possible exception of AHAS III) which use α ketoacids as substrates in the presence of thiamine pyrophosphate and Mg^{++} (36, 37). The gene products encoded by these genes are also very similar in size [IlvB, 562 amino acids; IlvG, 548 amino acids; IlvI, 566 amino acids (34); PoxB, 572 amino acids]. Further studies of the relatedness of these enzymes are in progress.

In summary, the DNA sequence of poxB has been established. The knowledge of the sequence will allow further study of the protein, particularly with respect to its interaction with lipid. The sequence established the size and amino acid sequence of both the intact protein (Mr 62,018) and the peptides released upon α-chymotrypsin treatment of the enzyme (in the presence of substrate and cofactor: Mr 59,400 + 2,600; in the absence of substrate and cofactor: Mr 51,000 +11,000). Since the small Mr 2,600 "α" peptide plays an important role in lipid binding, knowledge of the amino acid sequence allows for hypotheses concerning the mode of interaction of this region of pyruvate

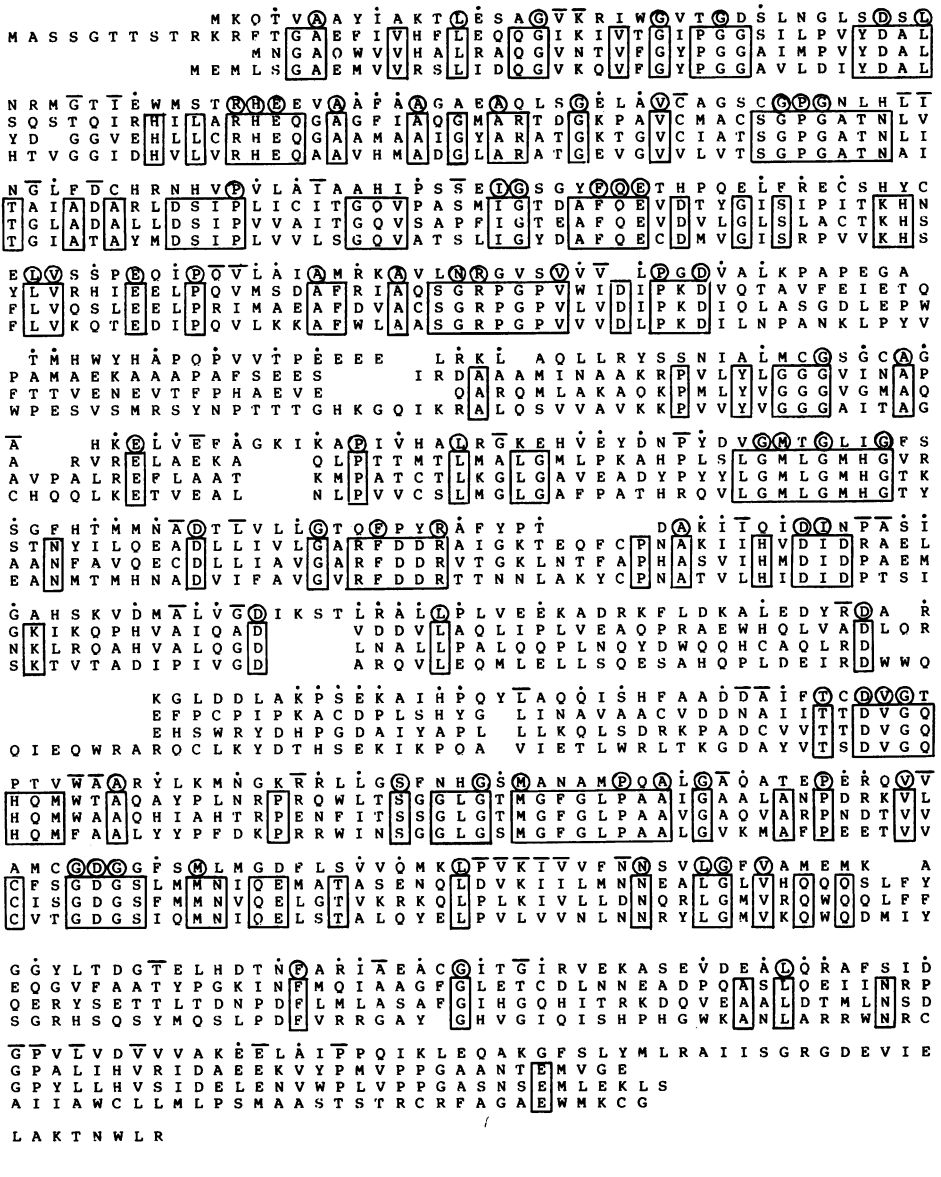

Figure 5. Amino acid sequence comparison of the polypeptides encoded by the poxB and ilvB, ilvG, and ilvI genes (first, second, third, and fourth lines, respectively). The sequences and alignment of the IlvB, IlvG, and IlvI proteins were taken from Wek et al. (34). Boxed regions indicate homology between IlvB, IlvG, and IlvI. Circles correspond to those PoxB amino acids homologous to all three Ilv polypeptides. Overlined positions indicate homology between PoxB and 2 of the 3 Ilv proteins. A dot indicates positions at which PoxB shares homology with one of the other polypeptides.

oxidase with the membrane. Indeed, analysis of the sequence predicts that amino acids 559-568 form an amphipathic α helical structure (6). The sequence has also allowed the use of site-directed mutagenesis of the C-terminal region to ascertain what portions of this region are important in lipid binding (Grabau and Cronan, unpublished results).

ACKNOWLEDGEMENTS

We thank J. Mankovich for helpful advice on DNA sequencing, M. Recny, S. Hamilton, and L. Hager for sharing their amino acid sequence data before publication, Y.-Y. Chang for helpful discussions concerning the protein homology, R. Gesteland for the use of his laboratory facilities for the synthesis of oligonucleotides, and R. S. Wolfe for the use of his computer during part of this work. This work was supported by National Science Foundation Grant PCM 8308778 and a National Institute of Health Traineeship (GM 07283). We acknowledge BIONET National Computer Resource for Molecular Biology, whose funding is provided by the Biomedical Research Technology Program Division of Research Resources NIH Grant # 1U41RR01685-02.

REFERENCES

1. Hager, L. P. (1957) J. Biol. Chem. 229, 251-263.
2. Williams, F. R. and Hager, L. P. (1966) Arch. Biochem. Biophys. 116, 168-176.
3. Cunningham, C. C. and Hager, L. P. (1971) J. Biol. Chem. 246, 1575-1582.
4. Cunningham, C. C. and Hager, L. P. (1971) J. Biol. Chem. 246, 1583-1589.
5. Russell, P., Schrock, H. L. and Gennis, R. B. (1977) J. Biol. Chem. 252, 7883-7887.
6. Recny, M. A., Grabau, C., Cronan, J. E., Jr. and Hager, L. P. (1985) J. Biol. Chem. 260, 14287-14291.
7. Recny, M. A. and Hager, L. P. (1983) J. Biol. Chem. 258, 5189-5195.
8. Grabau, C. and Cronan, J. E., Jr. (1986) Biochemistry. In Press.
9. Chang, Y.-Y. and Cronan, J. E., Jr. (1986) J. Bacteriol. In Press.
10. Chang, Y.-Y. and Cronan, J. E., Jr. (1984) Proc. Natl. Acad. Sci. U.S.A. 81, 4348-4352.
11. Messing, J. (1983) Methods Enzymol. 101, 20-78.
12. Sanger, F., Nicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 5463-5467.
13. Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. and Roe, B. A. (1980) J. Mol. Biol. 143, 161-178.
14. Grabau, C. and Cronan, J. E., Jr. (1984) J. Bacteriol. 160, 1088-1092.
15. Dretzen, G., Bellard, M., Sassone-Corsi, P. and Chambon, P. (1981) Anal. Biochem. 112, 295-298.
16. Raj, T., Russell, P., Flygare, W. H. and Gennis, R. B. (1977) Biochim. Biophys. Acta 481, 42-49.
17. Russell, P. (1978) Ph.D. Thesis, University of Illinois.
18. Hamilton, S., Recny, M. A. and Hager, L. P. (1985) Personal communication.

19. Gribskov, M., Devereux, J. and Burgess, R. R. (1984) Nuc. Acids Res. 12, 539-549.
20. Hopp, T. P. and Woods, K. R. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 3824-3828.
21. Chou, P. Y. and Fasman, G. D. (1978) Adv. Enzymol. 47, 45-148.
22. Blake, R., II, Hager, L. P. and Gennis, R. B. (1978) J. Biol. Chem. 253, 1963-1971.
23. O'Brien, T. A., Shelton, E., Mather, M. and Gennis, R. B. (1982) Biochim. Biophys. Acta 705, 321-329.
24. Stevens, D. J. and Gennis, R. B. (1980) J. Biol. Chem. 255, 379-383.
25. Porter, T. D. and Kasper, C. B. (1985) Proc. Natl. Acad. Sci. U.S.A. 82, 973-977.
26. Hearst, J. E. and Sauer, K. (1984) Z. Naturforsch. 39C, 421-424.
27. Koland, J. G., Miller, M. J. and Gennis, R. B. (1984) Biochemistry 23, 445-453.
28. Chang, Y.-Y. and Cronan, J. E., Jr. (1983) J. Bacteriol. 154, 756-762.
29. Grosjean, H. and Fiers, W. (1982) Gene 199-209.
30. Stormo, G. D., Schneider, T. D. and Gold, L. M. (1982) Nuc. Acids Res. 10, 2971-2996.
31. Hawley, D. K. and McClure, W. R. (1983) Nuc. Acids Res. 11, 2237-2255.
32. Rosenberg, M. and Court, D. (1979) Ann. Rev. Genet. 13, 319-353.
33. Recny, M. A. (1983) Ph.D. Thesis. University of Illinois.
34. Wek, R. C., Hauser, C. A. and Hatfield, G. W. (1985) Nucl. Acids Res. 13, 3995-4010.
35. Friden, P., Donegan, J., Mullen, J., Tsui, P. and Freundlich, M. (1985) Nucl. Acids Res. 13, 3979-3993.
36. De Felice, M., Squires, C. and Levinthal, M. (1978) Biochim. Biophys. Acta 541, 9-17.
37. De Felice, M., Levinthal, M., Iaccarino, M. and Guardiola, J. (1979) Microbiol. Rev. 43, 42-58.