ELSEVIER

Biochimica et Biophysica Acta 1188 (1994) 432–438

BBA Biochimica et Biophysica Acta

Short Sequence-Paper

DNA sequence of the *cut* A, B and C genes, encoding the molybdenum containing hydroxylase carbon monoxide dehydrogenase, from *Pseudomonas thermocarboxydovorans* strain C2

Danita M. Pearson, Catherine O'Reilly *, John Colby, Gary W. Black [1]

School of Health Sciences, University of Sunderland, Sunderland SR1 3SD, UK

Received 22 June 1994; revised 30 August 1994

Abstract

Pseudomonas thermocarboxydovorans strain C2 is capable of using carbon monoxide as the sole source of carbon and energy. The key enzyme for CO utilisation is the molybdenum containing iron-flavoprotein carbon monoxide dehydrogenase (CODH). This paper reports the DNA sequencing of a 4.7 kb region of the C2 genome which appears to encode the CODH enzyme. The genes for the three subunits of CODH, which we have named *cut* A, B and C, have been identified and they appear to form an operon. The predicted protein sequences of the three subunits have homology to the structurally related protein, xanthine dehydrogenase, from *Drosophila melanogaster*. By comparison with xanthine dehydogenase it can be predicted that the molybdenum cofactor binds to the large subunit of CODH, the small subunit of CODH contains the iron-sulphur centers and the medium subunit binds FAD/NAD + .

Keywords: DNA sequence; Molybdenum; Carbon monoxide; (*P. thermocarboxydovorans*)

Pseudomonas thermocarboxydovorans strain C2 is capable ofusing carbon monoxide as the sole source of carbon and energy [1,2]. The key enzyme involved in the utilisation of CO is carbon monoxide dehydrogenase EC 1.2.2.4 (CODH) [2]. The enzyme is a molybdenum iron-sulphur flavoprotein which has been classified as a molybdenum containing hydroxylase with similarities to a number of eukaryotic enzymes of this class [3]. Most information on molybdenum containing hydroxylases comes from studies done on xanthine dehydrogenase and xanthine oxidase from a variety of eukaryotic sources, but work done on xanthine dehydrogenase from *Drosophila melanogaster* has been particularly useful in identifying the regions associated with the iron-sulphur centres, FAD and the molybdopterin cofactor [3,4]. The CODH enzyme responsible for aerobic utilisation of CO has been studied in a number of prokaryotic systems and the enzyme appears very similar in all the organisms studied [5]. The genes encoding the enzyme may be either plasmid borne as is the case in *P. carboxydovorans* [6,7] or chromosomally borne as is the case with *P. thermocarboxydovorans* [8]. The enzyme from *P. thermocarboxydovorans* contains three different subunits with approximate molecular masses of 87 kDa for the large subunit (L), 30 kDa for the medium subunit (M) and 17 kDa for the small subunit (S) [2,8]. The active enzyme contains two of each subunit. A region of approximately 11 kb of the *P. thermocarboxydovorans* genome has been cloned and has previously been shown to encode at least the large and small subunits of CODH [8,9] with the region encoding the two subunits, the *cut* A and *cut* C genes, identified as being fully contained within a 4 kb *Eco*RI fragment. A 6 kb region of the original bacteriophage lambda clone, P22, has been subcloned to give pDP5 and a restriction map of this is shown in Fig. 1. The complete DNA sequence of a region of 4723 bp from this clone (indicated as a solid line in Fig. 1) has been determined (Fig. 2). The region includes the 4 kb *Eco*RI fragment which had been previously shown to

* Corresponding author. Fax: +44 91 5152502.

The nucleotide sequence reported in this paper has been submitted to the EMBL Data Bank with Accession number X77931.

[1] Present address: Dept. of Agricultural Biochemistry, University of Newcastle-upon-Tyne, Newcastle-upon-Tyne, UK.

SSDI 0005-2728(94)00152-9

encode at least the large and the small subunits of CODH [8]. Previous expression analysis indicated that *cut* A is contained within the 2.7 kb *Sph*I fragment subcloned in pGB 2 [8]. Preliminary identification of the *cut* A, B and C genes was done on the size of the protein products of the ORFs and homology between the predicted N-terminal sequence and the published N-terminal sequence of the enzyme from *Pseudomonas carboxydovorans* OM5 and *Pseudomonas carboxydoflava* [5]. Analysis of the sequenced region reveals four open reading frames and each of these will be discussed separately.

The largest ORF is ORF 3 which spans 2529 bp from 1612 to 4140 bp and the predicted product is a protein of 843 aa with a molecular mass of 91.8 kDa. This is in good agreement with the size of 87 kDa determined for the large CODH subunit. The ORF is completely contained within the *Sph*I fragment (1417–4185 bp) which had been previously shown to encode the large subunit. Comparison of the predicted N-terminal sequence to the limited available N-terminal sequences of the enzyme from other bacteria [5] is shown in Fig. 3. The homology is greatest to the large subunit of the enzyme from *P. carboxydoflava* with 6/9 residues identical. It is therefore concluded that ORF 3 encodes the CODH large subunit and is therefore designated *cut* A.

The sequenced region is known from expression studies to encode the small subunit of CODH. ORF 2 and ORF 4 both encode proteins with a predicted molecular mass of approximately 17 kDa. Comparison of the N-terminal sequences of these ORFs to the published N-terminal sequence of the small subunit of CODH from *P. carboxydovorans* and *P. carboxydoflava* indicates that the protein encoded by ORF 2 has significant homology (11/21 identical to *P. carboxydovorans.*) (Fig. 3c) while no apparent homology is detectable with ORF 4. It can therefore be concluded that ORF2 is *cut* C. To confirm this conclusion two subclones of pGB1, pDP1 and pDP2, were tested for expression of the CODH subunits. pDP2 which has ORF 4 intact does not produce any small subunit while pDP1 which has ORF 2 intact does produce the small subunit, thus confirming the conclusion that ORF 2 is *cut* C (Fig. 4 A).

The plasmid pGB1 does not express the medium subunit of CODH. Initial sequencing of the 4 kb *Eco*RI fragment of pGB1 indicated the position of the *cut* A and C genes as discussed above and also that the *cut* C gene was preceded by an incomplete ORF which started at the beginning of the pGB1 insert and ended 23 bp before the ATG of *cut* C. In order to obtain the complete sequence of this ORF a further 700 bp of sequence of the region 5′ of the *Eco*RI site was determined and revealed ORF 1 as indicated in Fig. 1. ORF 1 encodes a protein of predicted molecular mass 29.98 kDa which is in good agreement with the determined molecular mass of the medium subunit of CODH, and the N-terminal sequence shows good homology with the N-terminal sequences of the medium subunits of CODH from *P. carboxydovorans* (11/19 residues identical) and *P. carboxydoflava* (8/14 residues identical) (Fig. 3b). However, when a plasmid, pDP5, was constructed which contained a 6.0 kb fragment covering all of the sequenced region plus approximately 1.3 kb of 5′ sequence, no expression of any of the subunits was seen in *E. coli*. Apparently linking this 2.0 kb upstream region to the 4 kb *Eco*RI fragment is inhibiting expression of the two genes *cut* A and C which had been previously expressed. Analysis of the 5′ upstream region of each of the ORFs indicates that at least ORF 1, 2 and 3 form an operon as the distance between the genes is small and each of the ORFs are preceded by an identical 9 bp sequence, GAGAGGAAC, which has the appropriate sequence and is at the correct distance from each ORF to act as a ribosome binding site [10]. While the distance between ORF3 and 4 is also short (47 bp), there does not appear to be an RBS associated with ORF 4 although the codon usage of the predicted protein of ORF 4 is biased in the same way as in the other three ORFs (see below). A promoter has not been identified 5′ to ORF 1 but there is a sequence, TTGCA, at position 153 which is identical to the consensus sequence for the −10 region of σ^{54} type promoters [11], although no homology to the −26 region is obvious.

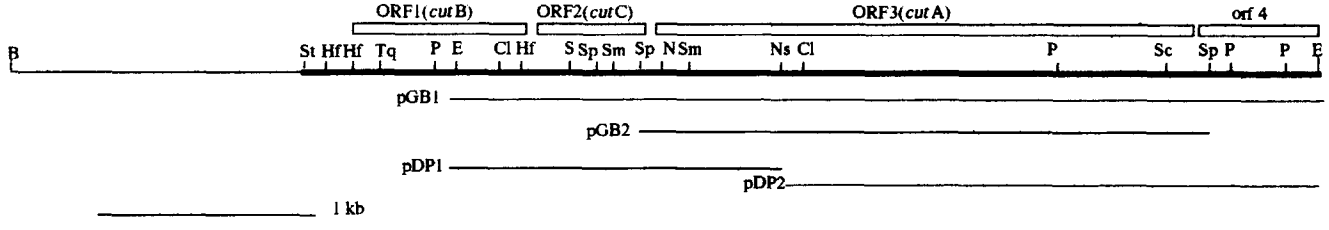

Fig. 1. Restriction map of the Bam-Eco-Eco insert of pDP5 showing key restriction sites only. The heavy line indicates the region sequenced and the position of each of the open reading frames is indicated. The structure of the other plasmids used in this study is shown. The restriction enzymes are: St, *Stu*I; Hf, *Hin*fI; Tq, *Taq*I; P, *Pst*I; E, *Eco*RI; Cl, *Cla*I; S, *Sal*I; Sp, *Sph*I; Sm, *Sma*I; N, *Nru*I; Ns, *Nsi*I; Sc, *Sac*I.

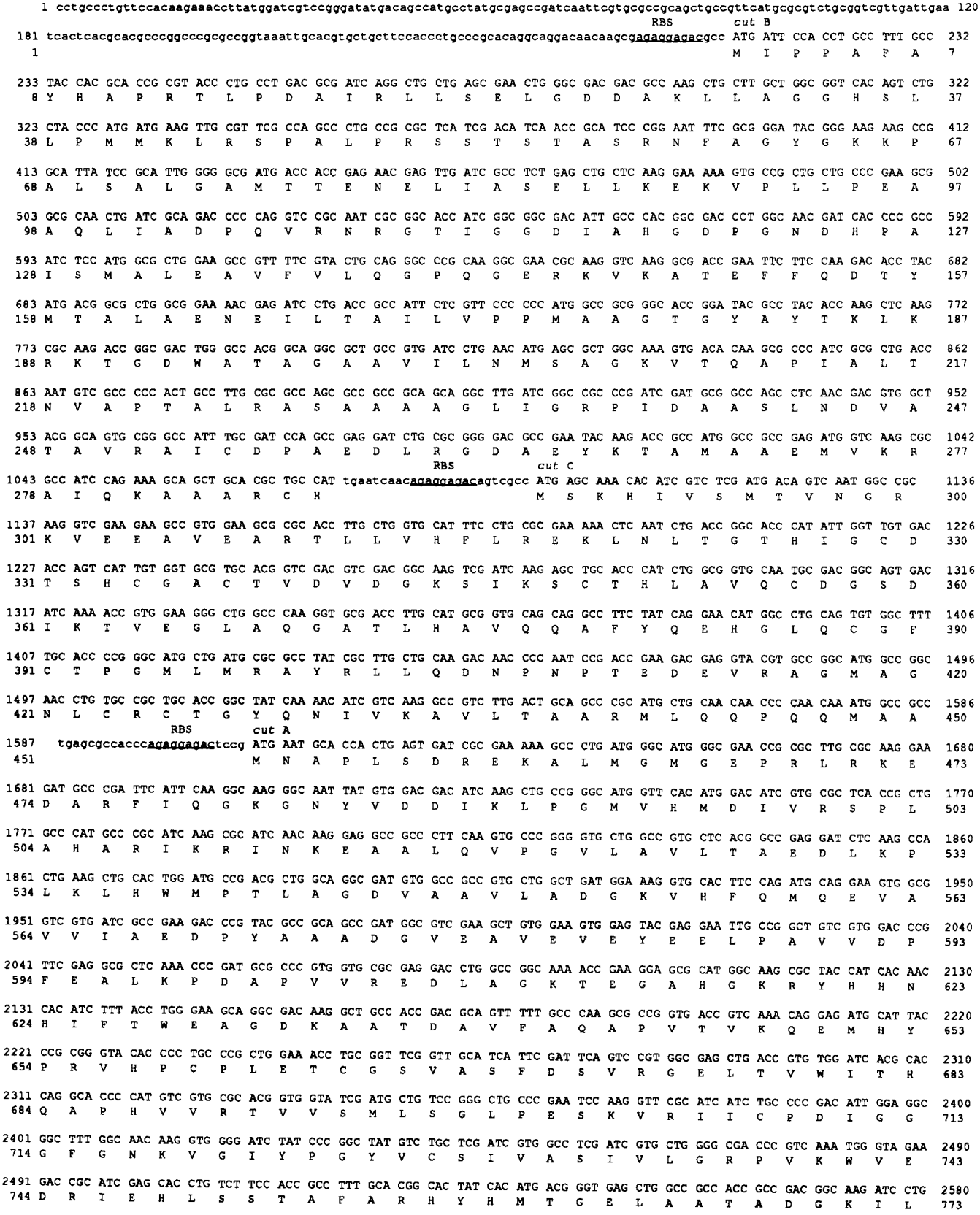

Fig. 2. See opposite.

```
2581 GCG CTG CGT GCC AAT GTG GTG GCC GAC CAC GGC GCC TTC GAC GCC TGC GCC GAC CCG AGC AAA TTC CCA GCC GGC CTG TTT CAC ATC TGC 2670
 774 A   L   R   A   N   V   V   A   D   H   G   A   F   D   A   C   A   D   P   S   K   F   P   A   G   L   F   H   I   C  803

2671 ACG GGC AGC TAC GAC ATA CCT ACT GCT TAC TGC CGG GTC GAT GGG GTC TAT ACC AAC AAG GCC CCC GGC GGC GTT GCC TAT CGC TGC TCG 2760
 804 T   G   S   Y   D   I   P   T   A   Y   C   R   V   D   G   V   Y   T   N   K   A   P   G   G   V   A   Y   R   C   S  833

2761 TTC CGC GTC ACC GAA GCC GTG TAT CTG ATC GAG CGC ATG GTG GAT GTC CTG GCG CAG AAG CTC AAC ATC GAC AAA GCC GAG ATT CGA GCC 2850
 834 F   R   V   T   E   A   V   Y   L   I   E   R   M   V   D   V   L   A   Q   K   L   N   I   D   K   A   E   I   R   A  863

2851 AAA AAT TTC ATT CGT AAG GAA CAG TTT CCC TAC CCG CAG GCA TTC GGA TTC GAA TAC GAC TCG GGT GAC TAT CAC ACC GCT CTC CAA AAA 2940
 864 K   N   F   I   R   K   E   Q   F   P   Y   P   Q   A   F   G   F   E   Y   D   S   G   D   Y   H   T   A   L   Q   K  893

2941 GTA CTC GAA GCC GTC GAT TAC AAA GGC TTG CGC GAA GAG CAG GCA CGC AAG CGT GCC GAT CCG AAC TGC CCG ACC CTG ATG GGC ATC GGC 3030
 894 V   L   E   A   V   D   Y   K   G   L   R   E   E   Q   A   R   K   R   A   D   P   N   C   P   T   L   M   G   I   G  923

3031 CTG GTC ACC TTC ACC GAA GTG GTC GGT GCC GGC CCT ACG AAG GTG TGC GAC ATC CTG GGT GTC GGC ATG TTC GAC TCC TGC GAA ATC CGC 3120
 924 L   V   T   F   T   E   V   V   G   A   G   P   T   K   V   C   D   I   L   G   V   G   M   F   D   S   C   E   I   R  953

3121 GTC CAT CCG ACC GGC AGC GCG ATT GCC CGT ATG GGA ACG ATC ACG CAA GGC CAG GGC CAT CAG ACC ACC TAT GCG CAA ATC ATC GCC ACC 3210
 954 V   H   P   T   G   S   A   I   A   R   M   G   T   I   T   Q   G   Q   G   H   Q   T   T   Y   A   Q   I   I   A   T  983

3211 GAA CTG GGG ATA CCC AGC GAC TTG ATC CAG GTG GAA GAG GGC GAC ACC GCC ACC GCC CCG TAC GGC TTG GGC ACG TAC GGC TCG CGC TCG 3300
 984 E   L   G   I   P   S   D   L   I   Q   V   E   E   G   D   T   A   T   A   P   Y   G   L   G   T   Y   G   S   R   S  1013

3301 ACA CCG GTG GCC GGA GCC GCC ATT GCC ATG GCG GCG CGC AAG ATC CAC GCC AAG GCC AGA AAG ATC GCC GCA CAC CTG CTG GAG GTC AGC 3390
1014 T   P   V   A   G   A   A   I   A   M   A   A   R   K   I   H   A   K   A   R   K   I   A   A   H   L   L   E   V   S  1043

3391 GAA GCC GAT CTC GAA TGG GAG ATC GAC CGC TTC AAG GTC AAA GGC CGC GAT GAC AAG TTC AAA ACC ATG AAA GAC ATT GCC TGG GCG GCC 3480
1044 E   A   D   L   E   W   E   I   D   R   F   K   V   K   G   R   D   D   K   F   K   T   M   K   D   I   A   W   A   A  1073

3481 TAC CAC CAG CCG CCT GCA GGT CTG GAG CCG GGG CTG GAA GCC GTG CAC TAT TAC GAC CCG CCG AAT TTC ACT TAT CCC TTC GGC GTC TAT 3570
1074 Y   H   Q   P   P   A   G   L   E   P   G   L   E   A   V   H   Y   Y   D   P   P   N   F   T   Y   P   F   G   V   Y  1103

3571 CTG TGC GTG GTG GAC ATC GAC AAA GGC ACG GGT GAG ACC AAG ATC CGC CGC TTC TAT GCG CTG GAC GAC TGC GGC ACC CGC ATC AAC CCG 3660
1104 L   C   V   V   D   I   D   K   G   T   G   E   T   K   I   R   R   F   Y   A   L   D   D   C   G   T   R   I   N   P  1133

3661 ATG ATC ATC GAA GGT CAG ATC CAC GGT GGA CTG ACG GAG GGC TTT GCC GTG GCC ATG GGG CAG CTG CTG TCC TTC GAC AAG CAG GGC AAC 3750
1134 M   I   I   E   G   Q   I   H   G   G   L   T   E   G   F   A   V   A   M   G   Q   L   L   S   F   D   K   Q   G   N  1163

3751 ATC CAG GGC AAC TCC TTC ATG GAC TAC TTC ATC CCG ACG GCG GTG GAA ACC CCG AAA TGG GAA ACC GAC TAC ACC GTA ACC CCT TCG CCC 3840
1164 I   Q   G   N   S   F   M   D   Y   F   I   P   T   A   V   E   T   P   K   W   E   T   D   Y   T   V   T   P   S   P  1193

3841 CAT CAC CCC ATC GGC GCC AAA GGC GTG GCT GAA TCG CCC CCA CGT GGG CAG CAT CCC AAC CTT CAC CAA CGC CAT CCG TCG ATG CCT TTG 3930
1194 H   H   P   I   G   A   K   G   V   A   E   S   P   P   R   G   Q   H   P   N   L   H   Q   R   H   P   S   M   P   L  1223

3931 CCA TCT CGG CGT AAC GCA CAT CAA CAT GCC CCA TAC GGC TTG GCG GGT GTG GCA AGA GCT CAA AAA GAA CGG GGT AGC CAC CAG CTG ACC 4020
1224 P   S   R   R   N   A   H   Q   H   A   P   Y   G   L   A   G   V   A   R   A   Q   K   E   R   G   S   H   Q   L   T  1253

4021 CCC ACC GGC GCG CAG GCG TCT CGC GCC TGC GCG ACT TCG ACT TCT ACA TCC TAC AAC ATC CGA GTA CAT ATT CAT GGA AGT CGT CAT CGA 4110
1254 P   T   G   A   Q   A   S   R   A   C   A   T   S   T   S   T   S   Y   N   I   R   V   H   I   H   G   S   R   H   R  1283
                                                                                                orf 4
4111 CAA GCA ATA TCC CGT GGC CGC CGG CCT tgatgccgcctgggccgtgttatccaacatcaacgagctggccacctgc ATG CCC GGT GCA TCG ATC ACC GAG 4210
1284 Q   A   I   S   R   G   R   R   P                                                    M   P   G   A   S   I   T   E  1300

4211 CAG CTG GAC GAG CGC CAC TAC AAA GGC CAG GTC CGC GTT AAG GTG GGT CCG GCC GTG GCC GCT TTT GCC GGC AGC ATC GAA GTG CTG CAG 4300
1301 Q   L   D   E   R   H   Y   K   G   Q   V   R   V   K   V   G   P   A   V   A   A   F   A   G   S   I   E   V   L   Q  1330

4301 CTT GAT GCC GCT CGC CGC AGC CTC AAG ATG GTG GGC AAA GGG GCA GAC AAG GCG GGT TCT TCT GCC TCC ATG GAA TTG GAA GCC GTG CTT 4390
1331 L   D   A   A   R   R   S   L   K   M   V   G   K   G   A   D   K   A   G   S   S   A   S   M   E   L   E   A   V   L  1360

4391 TTG CCC GCC GAA GGC GGC CGC TGC ACA CTG CAA GGC CAG GCT CGG GTG ATC GTC AGC GGC AAA TTT GCG CAG TTC GGC GGC CGC ATG ATG 4480
1361 L   P   A   E   G   G   R   C   T   L   Q   G   Q   A   R   V   I   V   S   G   K   F   A   Q   F   G   G   R   M   M  1390

4481 ACC TCG GTC TCC GAC ATG ATC CTG TCC CAG TTC GCC GAA ACC TTT TCG CAA AAA GCA CAG GCC CTG CAG GGC ACG GCT TCA GTC GCC GAC 4570
1391 T   S   V   S   D   M   I   L   S   Q   F   A   E   T   F   S   Q   K   A   Q   A   L   Q   G   T   A   S   V   A   D  1420

4571 ACG TCT GGC GCG CAG GCC AGC CCC GCC ACC ACA GCT GCG CCC GCT GCC GCC AAA GAA CTG AAC GCA CTC GGC CTG CTT TGG GCC ATG GTC 4660
1421 T   S   G   A   Q   A   S   P   A   T   T   A   A   P   A   A   A   K   E   L   N   A   L   G   L   L   W   A   M   V  1450

4661 AGA AAC TTC TTT GCC GGC TTG TTC GGC AAG AAA AAG GCC tgatacagccaagccgatgaattc                                          4723
1451 R   N   F   F   A   G   L   F   G   K   K   K   A                                                                  1463
```

Fig. 2. DNA sequence and predicted protein products of the *Stu*I-Eco-Eco region of pDP5 as indicated in Fig. 1. The ORFs are in capital letters. The 9 bp sequence which includes the ribosome binding site (RBS), and precedes ORFs 1, 2 and 3, is underlined. *DNA sequencing*: Manual and automated sequencing was carried out. Manual sequencing of double stranded plasmid DNA was by the dideoxy chain termination method of Sanger et al. [19] adapted for double stranded DNA by Chen and Seeburg [4]. Manual sequencing was carried out using either the Klenow polymerase or Sequenase. Non-standard primers were synthesised on an ABI 381A DNA synthesiser. Automated sequencing was carried out on an ABI 373A DNA sequencer. The complete sequence of both strands was determined and all junctions were sequenced over.

To determine if the protein product of ORF 1 is in fact the CODH medium subunit, ORF 1 was inserted as a transcriptional fusion into the *Nco*I site of the expression vector pKK 233–2 [12] to give pDP17. Analysis by Western blotting of proteins produced by cells containing this plasmid indicate that ORF 1 is *cut* B, as pDP17 leads to the production of an immunoreactive protein identical in size to the CODH medium

```
a)P.carboxydovorans          M N I Q T T V E P T A E R
  P.thermocarboxydovorans    M N A P L S D R E K A L M
  P.carboxydoflava           M N A P V Q D A E

b)P carboxydovorans          M M I P G S F D Y H R P K S I A D A V
  P.thermocarboxydovorans      M I P P A F A Y H R P R T L P D A I
  P.carboxydoflava           M M I P G   F E Y H A P K H V

c)P.carboxydovorans          M A K A H I   E L T I N G H P V E S L V E P
  P.thermocarboxydovorans    M S K   H I V S M T V N G R K V E E A V E A
  P carboxydoflava           M A K K I I   T V N V   G K A Q E K A V E P
```

Fig. 3. Comparison of the published N-terminal sequences of the (a) large, (b) medium and (c) small subunits of CODH from *P. carboxydovorans* and *P. carboxydoflava* [5] with the predicted N-terminal sequence of the enzyme from *P. thermocarboxydovorans*.

subunit (Fig. 4b). Insertion of ORF 1 in the opposite orientation to the *tac* 1 promoter of pKK 233–2 leads to no novel protein production. A plasmid, pDP 21, was constructed from pDP17 which has all three *cut* genes linked to the *tac* 1 promoter and cells containing this plasmid produce all three subunits of CODH (Fig. 4B), but no enzyme activity is detectable. The lack of enzyme activity is not surprising due to the complex nature of the active enzyme, and at present efforts are being made to reintroduce the cloned genes into a *cut* A::TN5 mutant strain of *P. thermocarboxydovorans* on a broad host range plasmid to look for CODH enzyme activity.

When the sequences of the predicted protein products of the *cut* A, B and C genes were compared to those in the NBRF protein data base, each subunit showed greatest homology to the xanthine dehydrogenase enzyme of *Drosophila melanogaster* [13,14]. The predicted sequence of the CODH small subunit showed greatest homology with 40% identity over the region aligned (Fig. 5), while the medium subunit is 24%

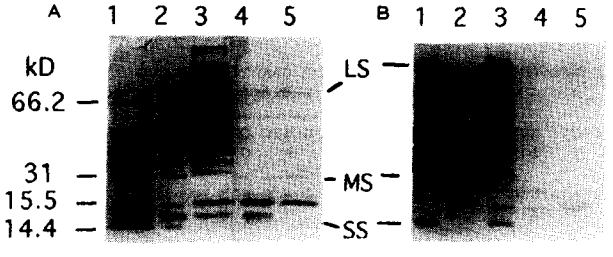

Fig. 4. Western blots, with CODH antiserum, of crude cell extracts of *P. thermocarboxydovorans* strain C2 (indicated as C2) and plasmid containing strains of *E. coli* (indicated by the plasmid name). (Panel A) tracks: 1, M_r standard; 2, C2; 3, pGB1; 4, pDP1; 5, pDP2. (Panel B) tracks: 1, C2; 2, pDP17; 3, pDP21; 4, pDP5; 5, pKK233–2. The positions of the large (LS), medium (MS) and small (SS) subunits of CODH are indicated. A large amount of degradation of the LS appears to take place as when the large subunit is not been expressed, as for example in tracks 4 and 5 of panel A, the large amount of immunoreactive material, seen in all cases where the large subunit is present, disappears. All attempts to reduce this degradation have failed. Crude cell extracts of *P. thermocarboxydovorans* and *E. coli* for immunoblotting were prepared as described previously [3,5]. SDS-PAGE of proteins for Western blotting was carried out using a BioRad mini-protean II cell. BioRad low range biotinylated molecular mass standards were used. Western blotting was done using standard procedures as described in Harlow and Lane (16).

identical and the large subunit is 26% identical. Xanthine dehydrogenase consists of a single polypeptide of 1319 amino acids and has three functional sections binding Fe-S, NAD + /FAD and the molybdenum cofactor [15,3]. The N-terminal segment of the enzyme contains the iron-sulphur, the middle segment the flavin and NAD + + binding sites and the C-terminal segment the molybdenum cofactor. The alignment of the CODH subunits would suggest that the small subunit binds iron-sulphur, the medium subunit binds FAD/NAD + and the large subunit binds the molybdenum cofactor. It is interesting to note that the order of the *cut* genes is B-C-A for the M, S, and L subunits, respectively, is different from the order of sections in eukaryotic xanthine dehydrogenases and it is therefore unlikely that they have evolved from a common ancestor, and the similarities seen must reflect functional similarities in the enzymes. The similarities seen allow the identification of residues potentially important for activity which will be analysed by site-directed mutagenesis. The conservation of cysteine residues in the small subunit of CODH is particularly striking. The *rosy* gene of *Drosophila melanogaster*, which encodes xanthine dehydrogenase, has been widely studied for many years. The effect of a number of *rosy* point mutations on the activity of xanthine dehydrogenase has been studied in detail [4]. Of the point mutants analysed in the study, the residue which is changed in the mutants is conserved between CODH and xanthine dehydrogenase in 7/11 cases (Fig. 5) and in the four others the amino acid change is to a similar amino acid. These residues will also be studied by site directed mutageneis. Another feature of interest in the sequence is the GGGFG sequence (underlined in Fig. 5) which is identical to the XDH sequence and is also conserved in rat XDH [15]. Mutation in *Drosophila* XDH leading to the loss of the first G of this sequence inactivates XDH (Fig. 5) [8]. It has been suggested [2] that the sequence, GGGFGG, in rat liver xanthine dehydrogenase may be involved in dinucleotide binding. The molydenum cofactor of CODH from *P. carboxydovorans* is molybopterin cytosine dinucleotide (MCD) [5] and while the cofactor from the *P. thermocarboxydovorans* has not been isolated, it is reasonable to suggest that it also will be MCD and that the sequence GGGFG may be involved in the binding of the cofactor through the pyrophosphate moiety. The consensus sequence for dinucleotide binding is GXGXXG [2] and while the sequence found in CODH lacks the last G residue, the amino acid at this position is a conservative change to asparagine. The medium subunit should, by analogy to XDH, bind FAD and NAD + ; however while there are 38 glycine residues in the sequence of the medium subunit, no sequence fitting the consensus sequence can be found. Eight of the glycine residues are conserved between *Drosophila*

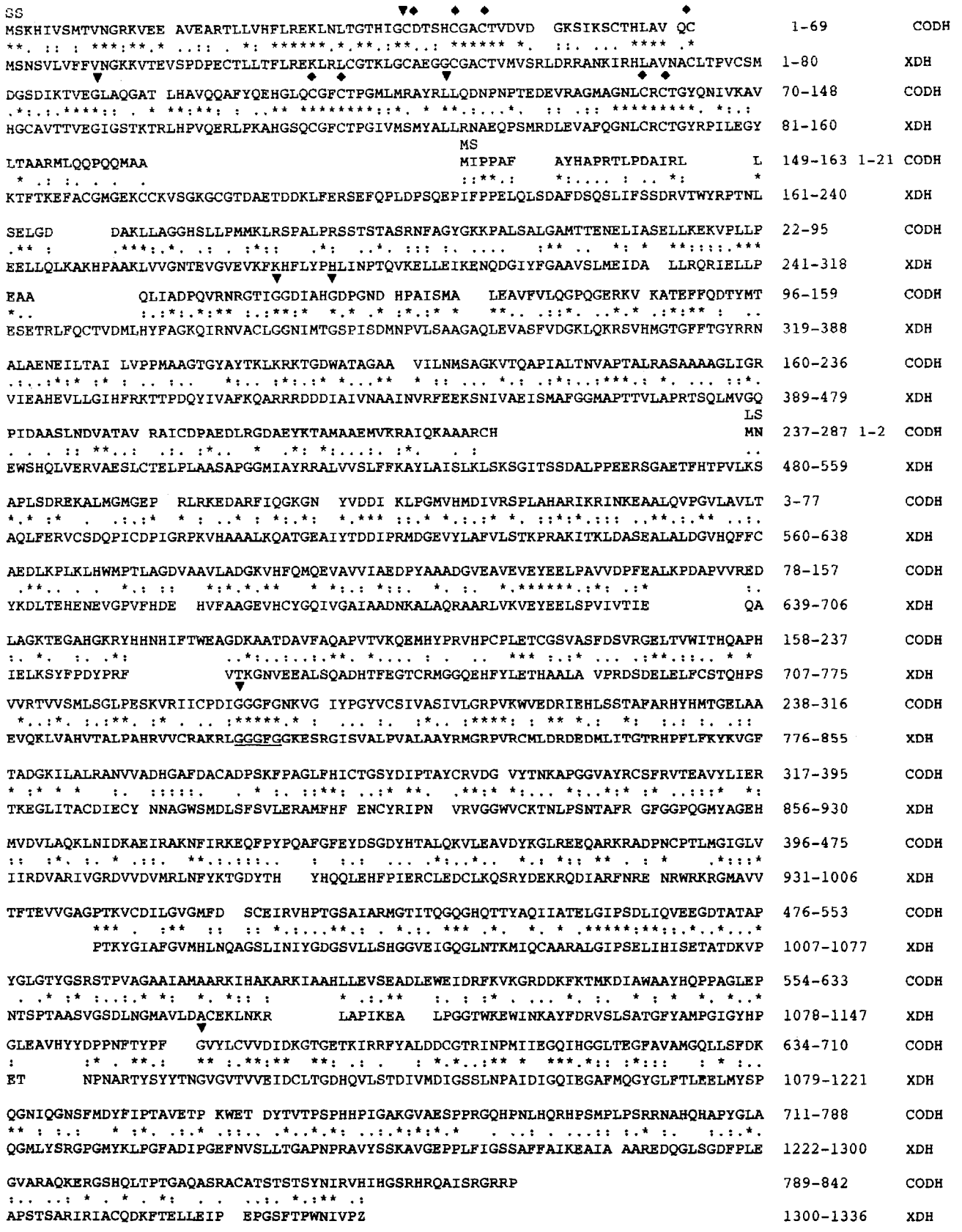

Fig. 5. Alignment of the predicted amino acid sequence of the three subunits of CODH with predicted sequence of XDH from *Drosophila melanogaster* [13,14]. The positions of the three subunits are indicated by SS, MS and LS over the initial methionine residue of the corresponding subunits. Symbols: ▼, residues conserved in CODH and XDH, at which mutation in the *Drosophila rosy* gene leads to a well defined change in XDH activity, as decribed by Hughes et al. [8]; ♦, cysteine residues conserved between the small subunit of CODH and the iron-sulphur binding domain of XDH. Symbols between the sequences are (*), identity and (.) and (:), conservative changes with (:) being to more similar amino acids than (.). The alignment was done using the Clustal program on the SERC Daresbury Laboratory computer system.

XDH and the CODH medium subunit, including two whose loss by mutation has been shown to inactivate XDH (Fig. 5) [4].

The work presented here is the first published sequence of a prokaryotic molybdenum-containing hydroxylase, and it indicates that these enzymes have more in common with eukaryotic molybdenum-containing hydroxylases than with prokaryotic molybdenum containing enzymes as was originally suggested by the work of Wootton et al. [3].

References

[1] Lyons, C.M., Justin, P., Colby, J. and Williams, E. (1984) J. Gen. Microbiol. 130, 1097–1105.
[2] Lyons, C.M, (1987) Ph.D. Thesis, University of Newcastle-upon-Tyne.
[3] Wooton, J.C., Nicolson, R.E., Cock, J.M., Walters, D.E., Burke, J.F., Doyle, W.A. and Bray, R.C. (1991) Biochim. Biophys. Acta 1057, 157–185.
[4] Hughes, R.K., Doyle, W.A., Chovnick, A., Whittle, J.R.S., Burke, J.F. and Bray, R.C. (1992) Biochem. J. 285, 507–513.
[5] Meyer, O., Frunzke, K. and Morsdorf, G. (1993) in Microbial Growth on C_1 Compounds (Murrell, J.C. and Kelly, D.P., eds.), Intercept, Andover, UK.
[6] Kraut, M. and Meyer, O. (1988) Arch. Microbiol. 149, 540–546.
[7] Kraut, M., Hugendieck, I., Herwig, S. and Meyer, O. (1989) Arch. Microbiol. 152, 335–341.
[8] Black, G.W., Lyons, C.M., Williams, E., Colby, J., Kehoe, M. and O'Reilly, C. (1990) FEMS Microbiol. Lett. 70, 249–254.
[9] O'Reilly, C., Colby, J., Pearson, D.M. and Black, G.W. (1993) in Microbial Growth on C_1 Compounds (Murrell, J.C. and Kelly, D.P., eds.), Intercept, Andover, UK.
[10] Gold, L. and Stormo, G. (1987) in *Escherichia coli* and *Salmonellla typhimurium*. Cellular and Molecular Biology Vol. 2 (Neihardt, F.C., ed.), pp. 1302–1307, ASM, Washington DC.
[11] Thomas, C.M. and Franklin, F.C.H. (1989) in Promiscuous Plasmids of Gram-negative Bacteria, Academic Press, New York.
[12] Amann, E. and Brosius, J. (1985) Gene 40, 183–190.
[13] Keith, T.P., Riley, M.A., Kreitman, M., Lewontin, R.C., Curtis, D. and Chambers, G. (1987) Genetics 116, 67–73.
[14] Lee, C.S., Curtis, D., McCarron, M., Love, C., Gray, M., Bender, W. and Chovnick, A. (1987) Genetics 116, 55–66.
[15] Amaya, Y., Yamazaki, K., Sato, M., Noda, K., Nishino, T. and Nishino, T. (1990) J. Biol. Chem. 265, 14170–14175.
[16] Harlow, E. and Lane, D. (1988) Antibodies: A laboratory Manual, Cold Spring Harbor Press, Cold Spring Harbor, NY.