Volume 317, number 1,2, 44–48 FEBS 12063 February 1993

Cloning and sequencing of glutamate mutase component E from *Clostridium tetanomorphum*

Organization of the *mut* genes

Daniel E. Holloway and E. Neil G. Marsh

Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge, CB2 1QW, UK

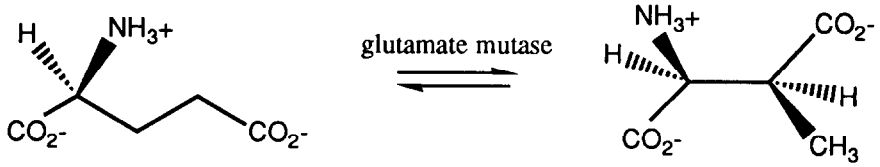
Received 1 December 1992; revised version received 16 December 1992

The gene encoding component E, the large subunit, of adenosylcobalamin (coenzyme B_{12})-dependent glutamate mutase from *Clostridium tetanomorphum* has been cloned and sequenced. The *mutE* gene encodes a protein of 485 amino acid residues, with M_r 53,708. The *mutE* gene is situated some 1,400 bp downstream of the *mutS* gene, which encodes the small subunit of glutamate mutase. Between the two is an open reading frame encoding a protein of 462 amino acids, with M_r 50,171, and of unknown function. All three genes appear to be transcribed as an operon and lie immediately upstream of the gene encoding β-methylaspartase, the next enzyme in the pathway of glutamate fermentation. Local homology exists between mutE and a region of β-methylaspartase which contains an active-site serine residue.

Glutamate mutase; Cobalamin-dependent enzyme; Sequence homology

1. INTRODUCTION

Glutamate mutase is one of several adenosylcobalamin (AdoCbl)-dependent enzymes which catalyze unusual isomerizations of carbon skeletons [1,2]. In this case, L-glutamate is converted to *threo*-β-methyl-L-aspartate in the first step of glutamate fermentation by *Clostridium tetanomorphum* [3].



Scheme 1.

The stereochemical course of the glutamate mutase-catalyzed reaction is known in some detail [4] and the role of AdoCbl as the intermediate hydrogen carrier in the rearrangement has been demonstrated [5]. However, further progress in understanding the mechanism will require detailed information on the structure of the enzyme itself.

Glutamate mutase comprises two subunits, originally designated component E and component S [6], both of which have been purified to homogeneity ([7], E.N.G. Marsh, unpublished results). Component E is a homodimer of subunit $M_r \approx 50{,}000$ which binds both AdoCbl and substrate but is completely inactive without component S present [7,8]. Component S is a monomeric protein of $M_r \approx 15{,}000$ which binds neither substrate nor AdoCbl and which contains reactive thiol groups that must be reduced for the protein to be active [9]. We have recently cloned and sequenced the gene encoding S (*mutS*) and identified sequence homologies between it and the C-terminus of AdoCbl-dependent methylmalonyl-CoA mutase and the cobalamin-binding region of methionine synthase [10]. This suggests that component S may interact with AdoCbl bound to component E, possibly through the dimethylbenzimidazole ribofuranosyl portion of the coenzyme, which acts as the axial ligand to cobalt. This is consistent with the observation that increasing the molar ratio of S to E decreases the apparent K_m for AdoCbl [9].

We now describe the cloning and sequencing of *mutE*, the gene which encodes component E of glutamate mutase from *Cl. tetanomorphum*. Unusually for genes which encode subunits of the same enzyme, *mutE* is not

Correspondence address: E.N.G. Marsh, Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge, CB2 1QW, UK. Fax: (44) (223) 333 345.

 Published by Elsevier Science Publishers B V

adjacent to *mutS*. Instead a third gene, the sequence of which we report here, was found between *musS* and *mutE*. The function of this gene is unknown, and we have designated it *mutL*.

2. MATERIALS AND METHODS

2.1. *Construction and screening of genomic DNA library*

Genomic DNA was purified from *Cl. tetanomorphum* (NCIMB no. 11547) by standard methods and used to construct a library in the λ vector, EMBL3 [11]. The pUC-based plasmid, pNM1 [10], which contains a 1,900 bp fragment of DNA immediately upstream from *mutE*, was used to generate a probe for the gene. The insert of pNM1 was excised from the vector by restriction with *Eco*RI endonuclease and purified by agarose gel electrophoresis. The purified insert was then used as a template to produce a ^{32}P-labelled random-primed probe [11]. Approximately 60,000 plaques were screened using random-primed probe according to published protocols [11].

2.2. *Subcloning and sequence analysis*

Positively hybridizing λ clones were analyzed by restriction mapping and appropriate restriction fragments subcloned into pUC119 using *E coli* TG1 [12] carrying the *recO* mutation [13] as the host (the generous gift of Dr. P. Oliver, Department of Genetics, University of Cambridge, UK). Randomly overlapping fragments of the appropriate restriction fragment were generated by sonication, and 'shotgun' cloned into M13mp18 [14]. The DNA sequence was determined by sequencing these M13mp18 subclones using 'universal' primer [14] and by double-stranded sequencing [15] of the parent plasmid using specifically synthesized oligonucleotide primers. The sequence data was compiled and analyzed using the Staden [16] and University of Wisconsin Genetics Computer Group programs [17].

3. RESULTS AND DISCUSSION

The plasmid pNM1 was used to probe a library of *Cl. tetanomorphum* genomic DNA for the *mut* genes. The insert of pNM1 comprises a 1,900 bp PCR product obtained using primers designed from the N-terminal sequences of components S and E [10]. Out of approximately 60,000 plaques screened only two contained DNA which hybridized to the insert of pNM1. From one of these clones, λGM1, a 5.2 kbp *Sal*I fragment which hybridized with the probe was subcloned into pUC119 to the construct, pGM1. The nucleotide sequence of *mutS* has previously been determined by partially sequencing this plasmid [10].

The sequence downstream of *mutS* was determined from pGM1, and this established that only the first 350 nucleotides of the *mutE* gene were present on this plasmid. The insert of pGM1 mapped to the extreme end of the parent λ clone such that the remainder of the *mutE* gene was not encompassed by λGM1 either (Fig. 1). It was necessary, therefore, to subclone a complete copy of *mutE* from the other λ clone, λGM2. An 8.5 kbp *Sal*I–*Bam*HI fragment was identified by hybridization and attempts made to subclone this fragment into pUC119. A recombinant colony was obtained which contained a plasmid, pGM2, that hybridized to the insert of pNM1. However, the *Sal*I–*Bam*HI fragment could not be excised from the plasmid by digestion with the appropriate restriction enzymes. During growth, cultures containing pGM2 spontaneously deleted it to give a plasmid of less than 2 kbp that was refractory to digestion by restriction enymes which cut in the polylinker of pUC119. To maintain pGM2 the culture had to be streaked out and non-deleted clones identified before each large scale growth of *E. coli* TG1*recO* cells containing the plasmid. Even this approach was only partially successful in overcoming the problem. It was,

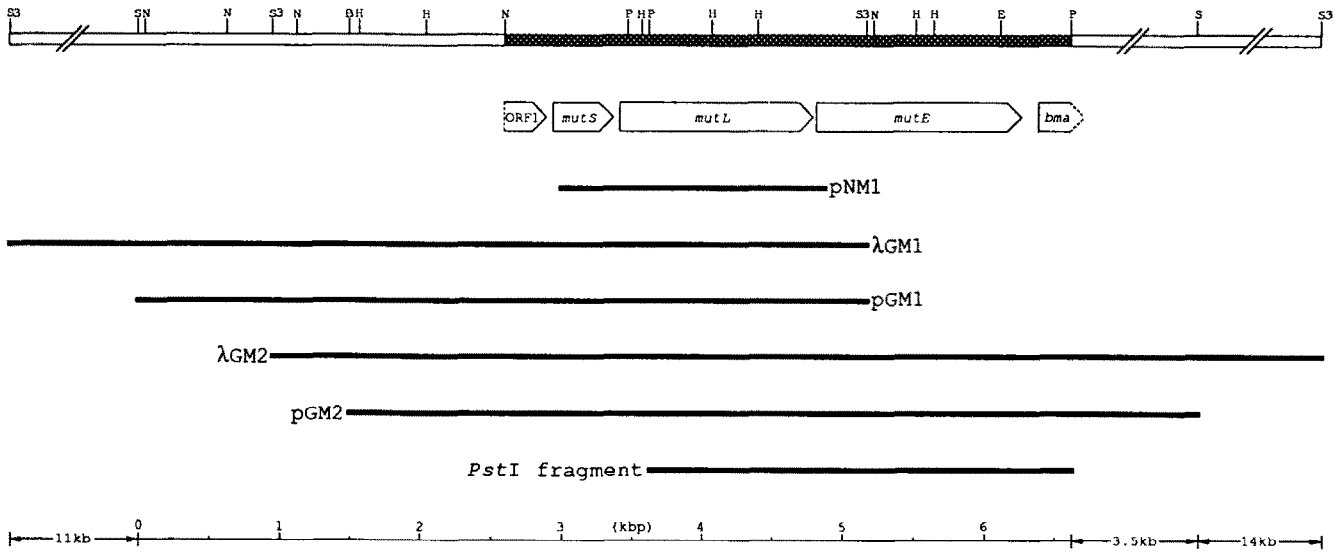
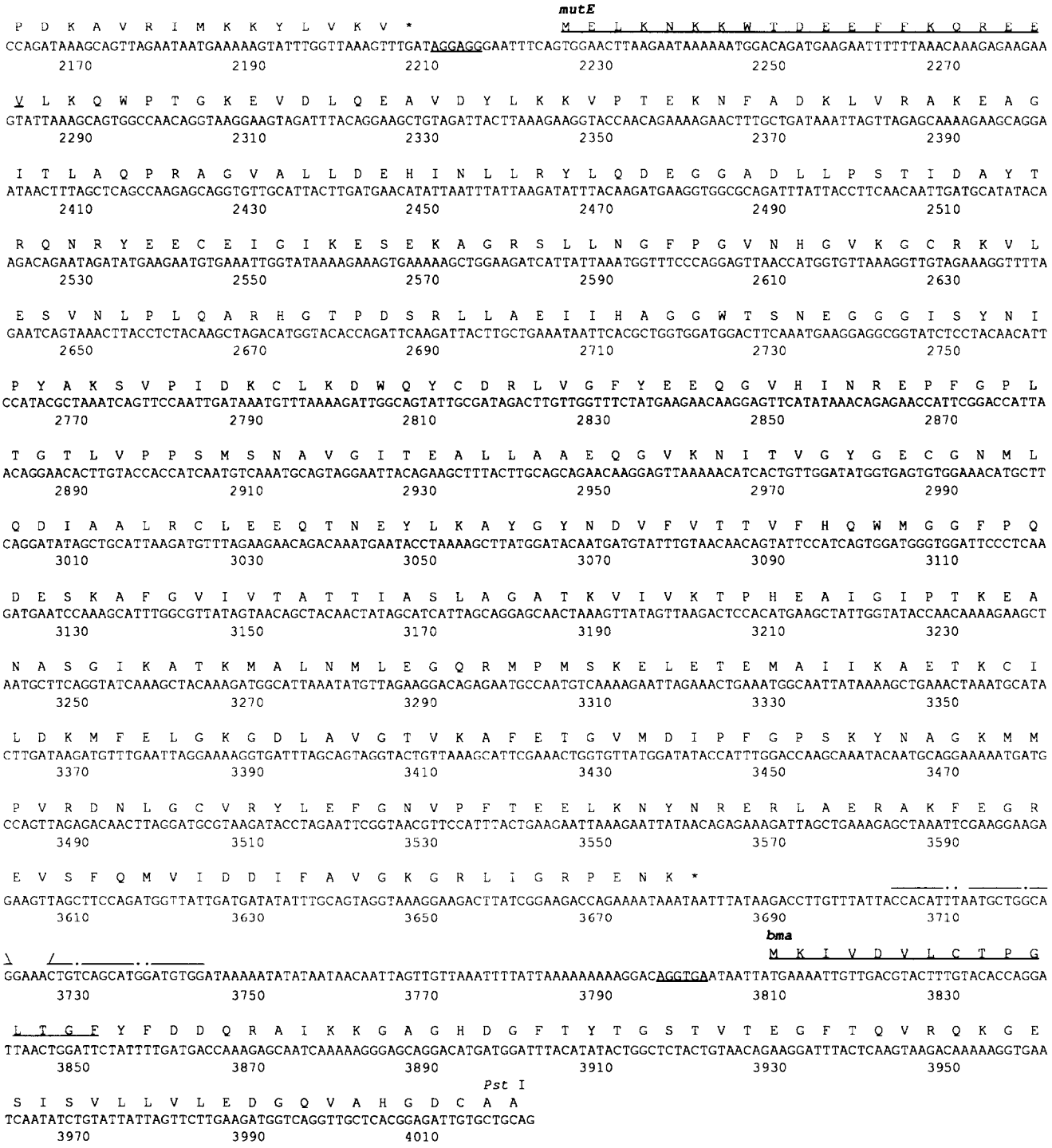


Fig. 1. Restriction map showing constructs employed in the cloning and sequencing of the *mut* genes. Only restriction sites employed in the generation and/or mapping of clones are shown. Restriction enzymes are B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; N, *Nco*I; P, *Pst*I; S, *Sal*I; S3, *Sau*3AI. The nucleotide sequence of the shaded region is shown in Fig. 2. For details refer to the text.

Fig. 2. The nucleotide and deduced protein sequence of the *mut* genes. Potential ribosome binding sites are underlined in the DNA sequence. The opposing arrows above the nucleotide sequence indicate a region of dyad symmetry which could potentially function as a transcription terminator; the gap denotes a mismatch and the dots represent neutral G·T pairings. Stop codons are not denoted by *. The underlined protein sequences were previously determined by automated sequencing [21] of enzymes purified from *Cl. tetanomorphum* ([10], E.N.G. Marsh, unpublished results). *Bma* encodes β-methylaspartase.

```
Nco I          ORF1
   M  E  T  R  F  S  F  I  C  I  S  E  E  F  K  F  K  V  R  D  A  L  E  S  A  G  L  G  N  I  I  I  T  Y  T  N  S  S  D
CCATGGAAACTAGATTTTCTTTCATCTGTATATCAGAAGAATTTAAATTTAAAGTAAGGGATGCATTAGAGAGTGCTGGATTAGGAAATATAATTATTACCTATACCAATTCTTCAGAT
      10                  30                  50                  70                  90                  110

 R  E  E  L  M  E  V  I  E  N  S  D  V  I  I  T  S  P  G  R  Y  K  E  L  Y  E  I  N  N  G  R  R  Q  I  I  N  F  L  Y  S
AGAGAAGAGTTAATGGAAGTAATAGAAAATTCTGATGTAATCATAACATCACCAGGAAGATATAAAGAACTTTATGAGATTAATAATGGTAGAAGACAAATTATAAATTTTCTTTACAGT
      130                 150                 170                 190                 210                 230

 L  D  D  G  S  V  K  A  L  K  S  K  L  L  E  I  K  Y  S  K  *
TTAGATGATGGTTCTGTAAAAGCTCTAAAGTCTAAATTGTTGGAGATTAAATATTCAAAATAAATTTCATGCATAAATATATAAATTATATATATATAAATTATCTAGGAGGTTTTATAA
      250                 270                 290                 310                 330                 350
mutS
 M  E  K  K  T  I  V  L  G  V  I  G  S  D  C  H  A  V  G  N  K  I  L  D  H  S  F  T  N  A  G  F  N  V  V  N  I  G  V  L
ATGGAGAAAAAGACTATTGTTCTTGGAGTTATTGGTTCAGACTGTCATGCAGTTGGTAACAAAATATTAGACCACTCATTTACAAATGCAGGCTTCAATGTTGTTAACATAGGAGTTTTA
      370                 390                 410                 430                 450                 470

 S  S  Q  E  D  F  I  N  A  A  I  E  T  K  A  D  L  I  C  V  S  S  L  Y  G  Q  G  E  I  D  C  K  G  L  R  E  K  C  D  E
TCATCACAGGAAGATTTTATAAATGCAGCTATAGAAACTAAAGCAGACCTTATATGTGTTTCTTCATTATATGGACAGGGAGAAATTGACTGTAAAGGATTAAGAGAAAAGTGTGATGAA
      490                 510                 530                 550                 570                 590

 A  G  L  K  G  I  K  L  F  V  G  G  N  I  V  V  G  K  Q  N  W  P  D  V  E  Q  R  F  K  A  M  G  F  D  R  V  Y  P  P  G
GCAGGACTTAAAGGAATAAAATTATTTGTTGGCGGAAACATTGTTGTTGGTAAACAAAACTGGCCAGATGTTGAACAGAGATTTAAAGCAATGGGATTTGATAGAGTATATCCACCAGGA
      610                 630                 650                 670                 690                 710
                                                                                                    mutL
 T  S  P  E  T  T  I  A  D  M  K  E  V  L  G  V  E  *                                          M  D  A  Y  L  L  L
ACATCTCCAGAAACAACAATAGCTGATATGAAAGAAGTTTTAGGAGTAGAATAATTTAACTAAAGGCTACTGTATTGCCTAAATTAGGAGTTGATTAAGGTGGATGCTTATTTACTTTTA
      730                 750                 770                 790                 810                 830

 D  F  G  S  T  Y  T  K  L  T  A  V  D  I  E  N  E  G  I  L  A  T  A  K  D  I  T  T  I  E  S  D  I  M  V  G  F  N  K  A
GATTTTGGTAGCACCTATACAAAACTTACTGCAGTAGATATAGAAAATGAAGGGATATTAGCTACAGCAAAAGATATAACAACTATAGAAAGCGATATAATGGTAGGGTTTAACAAAGCC
      850                 870                 890                 910                 930                 950

 Y  E  K  L  T  E  Q  L  E  G  K  E  V  N  F  V  K  K  L  A  C  S  S  A  A  G  G  L  K  M  I  A  I  G  L  V  P  E  L  T
TACGAAAAGCTTACTGAACAGTTAGAAGGAAAAGAAGTGAATTTTGTTAAAAAGTTAGCATGTTCATCTGCAGCAGGTGGACTAAAGATGATAGCCATCGGTCTTGTACCAGAACTTACA
      970                 990                 1010                1030                1050                1070

 A  E  A  A  K  R  A  A  L  G  A  G  A  R  V  L  N  V  Y  S  Y  D  L  T  N  K  E  V  E  E  I  K  N  S  N  L  D  I  I  L
GCAGAAGCTGCAAAGAGAGCTGCTCTTGGAGCAGGAGCTAGGGTACTAAATGTATATAGTTACGATTTGACTAATAAAGAAGTTGAAGAAATAAAAAATTCTAACTTGGACATAATACTT
      1090                1110                1130                1150                1170                1190

 L  A  G  G  T  D  G  G  N  K  E  C  M  I  H  N  A  K  M  L  A  E  H  G  V  K  L  P  I  V  V  A  G  N  K  V  V  S  D  E
TTAGCAGGTGGTACAGATGGTGGAAATAAAGAATGTATGATTCATAATGCAAAAATGCTAGCTGAGCATGGAGTTAAACTTCCAATAGTTGTAGCAGGAAATAAAGTAGTTAGTGATGAA
      1210                1230                1250                1270                1290                1310

 V  S  E  I  F  D  K  A  G  I  F  Y  R  V  T  E  N  V  M  P  K  L  N  T  L  N  V  E  P  A  R  E  E  I  R  Q  I  F  M  K
GTATCAGAAATATTTGATAAAGCAGGTATATTTTATAGAGTTACTGAAAATGTAATGCCCAAGTTAAATACATTAAATGTAGAACCTGCAAGAGAAGAAATAAGACAAATATTTATGAAA
      1330                1350                1370                1390                1410                1430

 K  I  V  E  A  K  G  M  S  N  A  E  S  F  I  N  G  I  L  M  P  T  P  A  A  V  L  K  A  A  R  V  L  A  E  G  T  D  K  E
AAAATTGTAGAAGCAAAAGGAATGTCAAATGCTGAAAGCTTTATTAATGGAATACTTATGCCTACTCCAGCAGCAGTTTTAAAGGCAGCTAGAGTATTAGCAGAAGGAACCGATAAAGAA
      1450                1470                1490                1510                1530                1550

 D  G  I  G  D  L  I  V  V  D  I  G  G  A  T  T  D  V  H  S  L  A  D  G  E  P  S  K  P  G  V  T  L  R  G  L  E  E  P  F
GATGGTATTGGCGATTTAATAGTAGTAGATATTGGTGGTGCTACTACAGATGTACACTCACTTGCAGATGGCGAGCCATCTAAGCCAGGAGTTACATTAAGAGGGCTTGAAGAACCTTTT
      1570                1590                1610                1630                1650                1670

 A  K  R  T  V  E  G  D  L  G  M  R  Y  S  A  I  S  L  W  E  A  S  G  T  R  K  L  Q  K  Y  L  C  D  N  T  V  D  V  E  A
GCAAAAAGAACTGTTGAAGGCGACCTAGGAATGAGATATTCTGCCATTTCCTTATGGGAGGCATCAGGAACAAGAAAACTTCAAAAATATCTATGTGATAATACTGTAGATGTAGAAGCT
      1690                1710                1730                1750                1770                1790

 C  C  K  Y  R  A  E  H  I  K  M  V  P  E  T  E  E  E  I  K  F  D  E  A  M  A  K  V  A  T  D  M  A  M  E  R  H  V  G  V
TGTTGTAAATATAGAGCTGAACATATAAAAATGGTACCCGAGACAGAAGAAGAAATTAAGTTTGATGAAGCTATGGCTAAGGTTGCAACTGATATGGCTATGGAAAGACATGTAGGAGTT
      1810                1830                1850                1870                1890                1910

 I  E  S  M  Y  T  P  M  G  V  I  Y  S  Q  I  G  K  D  L  L  N  V  K  C  V  I  G  T  G  G  V  L  V  H  S  K  N  P  G  E
ATAGAAAGTATGTATACTCCTATGGGAGTCATATATAGTCAGATAGGTAAAGACTTGTTAAACGTAAAGTGTGTAATAGGAACTGGAGGAGTACTAGTTCATAGTAAGAATCCAGGTGAA
      1930                1950                1970                1990                2010                2030

 I  L  K  A  G  S  F  D  M  A  D  A  T  H  L  K  P  Q  H  P  E  Y  Y  I  D  K  T  Y  I  L  S  A  M  G  L  L  A  E  D  L
ATACTAAAAGCAGGGTCATTTGATATGGCCGATGCTACTCATTTGAAACCACAGCATCCAGAATATTATATAGATAAAACATATATATTATCTGCAATGGGGCTTCTAGCAGAAGATCTT
      2050                2070                2090                2110                2130                2150
```

Fig. 2 continued.



nevertheless, possible to excise from pGM2 a 3.0 kbp *Pst*I fragment which overlapped pGM1 (Fig. 1) and the nucleotide sequence of *mutE* was obtained by sequencing this fragment.

The nucleotide sequence and deduced protein sequence of the *mut* genes are shown in Fig. 2. The genes are flanked upstream by an open reading frame (ORF1) which showed no significant homology to other protein sequences in the database. Downstream, *mutE* is followed by the gene encoding β-methylaspartase (*bma* in Fig. 2) [18] and E.N.G. Marsh, unpublished results) which is the next enzyme in the pathway of glutamate fermentation. The 57 bp untranslated region upstream of *mutS* contains an AT-rich tract of DNA with se-

```
                170  *    180        190        200        210
bma     GAEINAVPVFAQSGDDRYDNVDKMIIKEADVLPHALINNVEE-KLGLKG-EKLLEYVK
        ||:: : :: | : ::||:: : : |||::   ::|:|:  : : |:||  |:|| |:
mutE    GADLLPSTIDAYTRQNRYEECE-IGIKESEKAGRSLLNGFPGVNHGVKGCRKVLESVN
          90        100        110        120        130        140
```

Fig. 3. Alignment of the deduced amino acid sequence of mutE and β-methylaspartase (bma). The aligned regions are G87 to N143 of mutE and G161 to K216 of β-methylaspartase. Identical residues are denoted by '|' and conserved residues by: ':'. The active-site serine of β-methylaspartase is identified by *.

quence motifs corresponding to the *E. coli* TATAAT promoter element. Downstream of the *mutE* gene is a region of dyad symmetry which is characteristic of ρ-independent transcription termination signals in clostridia [19]. The intergenic regions between *musS*, *mutL* and *mutE* are relatively short and do not contain recognizable clostridial promoter and termination motifs [19]. Thus it appears that all three genes are probably transcribed as an operon, but, in the absence of RNA transcript mapping, it is unclear whether the operon extends to include adjacent genes.

The *musS* gene has been described previously [10]. The next gene, *mutL*, encodes a protein of 462 amino acid residues with M_r 50,171. The function of this protein is unknown; it showed no significant homology to mutS, mutE, β-methylaspartase nor any proteins in the SWISSPROT database. The *mutE* gene encodes a protein of 485 amino acid residues with M_r 53,708, which is in good agreement with the subunit molecular weight for component E determined by SDS-PAGE and gel filtration ([17] and E.N.G. Marsh, unpublished results). Comparison of the deduced N-terminal sequence of mutE with that obtained previously by protein sequencing [10] revealed four discrepancies. The misassignment of W8 as L, E12 as G, R17 as A and E19 as G can all be attributed to contamination of the sequencing sample with mutS which masked the mutE sequence at these positions.

Comparisons of mutE with other proteins in the databases failed to reveal any significant homologies, nor were any noteworthy similarities found when mutE was aligned with other cobalamin-binding proteins. However, comparison of mutE with β-methylaspartase using the FASTA program [17] did identify a region of significant local similarity between the two (Fig. 3). Interestingly, the alignment includes Ser-173 of β-methylaspartase, which has been implicated in the mechanism of deamination [18,20]. This suggests that these partially conserved sequences may be involved in substrate binding.

A surprising result of our work to clone and sequence the genes encoding glutamate mutase has been the identification of a third gene, *mutL*, between *mutS* and *mutE*. Usually bacterial genes encoding subunits of the same enzyme lie adjacent to each other on the chromosome and are transcribed as an operon. It is probable that all three *mut* genes are transcribed together, which strongly hints at mutL playing some role in the glutamate fermentation pathway. However, mutL is clearly not essential to the glutamate mutase-catalyzed reaction and its function, if any, in glutamate metabolism in *Cl. tetanomorphum* is at present unknown.

Acknowledgements. We thank Dr. Jesus Cortes for his generous gifts of vectors and strains and especially for his help and advice throughout. Automated protein sequence determinations were carried out by Dr. James McCormick and oligonucleotide synthesis undertaken by Mr. Michael Weldon, both of the SERC/Wellcome-funded Protein and Nucleic Acid Chemistry Facility (Department of Biochemistry, University of Cambridge). Financial support was provided by SERC and the Royal Society. E.N.G.M. is a member of the Cambridge Centre for Molecular Recognition.

REFERENCES

[1] Switzer, R.L. (1982) in: B_{12}, vol. 2. (Dolphin, D., ed.) pp. 289–306, Wiley, New York.
[2] Halpern, J. (1985) Science 227, 869–875.
[3] Barker, H.A., Weissbach, H. and Smyth, R.D. (1958) Proc. Natl. Acad. Sci. USA 44, 1993–1997.
[4] Hartrampf, G. and Buckel, W. (1984) FEBS Lett. 171, 73–78.
[5] Switzer, R.L., Baltimore, B.G. and Barker, H.A. (1969) J. Biol. Chem. 244, 5363–5268.
[6] Barker, H.A., Roove, V., Suzuki, F. and Iodice, A.A. (1964) J. Biol. Chem. 239, 3260–3266.
[7] Leutbecher, U., Bocher, R., Dietman, L. and Buckel, W. (1992) Eur. J. Biochem. 205, 759–765.
[8] Suzuki, F. and Barker, H.A. (1966) J. Biol. Chem. 241, 878–888.
[9] Switzer, R.L and Barker, H.A. (1967) J Biol. Chem. 242, 2658–2674.
[10] Marsh, E.N.G. and Holloway, D.E. (1992) FEBS Lett. 310, 167–170.
[11] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[12] Gibson, T.J. (1984) Ph.D. Thesis, University of Cambridge, UK.
[13] Morrison, P.T., Lovett, S., Gilson, L.E. and Kolonder, R.C. (1989) J. Bacteriol. 171, 3641–3649.
[14] Bankier, A.T., Weston, K.M. and Barrell, B.G. (1987) Methods Enzymol. 155, 51–93.
[15] Murphy, G. and Ward, E.S. (1989) in: Nucleic Acids Sequencing: A Practical Approach (Howe, C.J. and Ward, E.S., eds.) pp. 99–115, IRL Press, Oxford.
[16] Staden, R. (1986) Nucleic Acids Res. 14, 217–238.
[17] Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387–395.
[18] Goda, S.K., Minton, N.P., Botting, N.P. and Gani, D. (1992) Biochemistry 31, 10747–10756.
[19] Young, M., Minton, N.P. and Stadenbauer, W.L. (1989) FEMS Microbiol. Rev. 63, 301–326.
[20] Wu, W.T. and Williams, V.R. (1968) J. Biol. Chem. 243, 5644–5650.
[21] Hunkapiller, M.W., Hewich, R.M., Dreyer, W.J. and Hood, L.E. (1983) Methods Enzymol. 91, 399–412.