Localisation of a strongly conserved section of coding sequence in glutamate dehydrogenase genes

Iain W. Mattaj, Michael J. McPherson and John C. Wootton*

Department of Genetics, University of Leeds, Leeds LS2 9JT, England

Received 10 August 1982

Glutamate dehydrogenase gene *Cooling sequence* (Escherichia coli) *Sequence conservation* (Neurospora crassa) *DNA sequence*

1. INTRODUCTION

NADP-specific glutamate dehydrogenase (NADP-GDH; EC 1.4.1.4) is a major enzyme of ammonium assimilation in many prokaryotic and eukaryotic microorganisms. Many enzymological and physiological studies have been made (review [1]) but little is known about the organisation and regulation of genes determining glutamate dehydrogenases (GDHs). The *gdhA* gene of *Escherichia coli* K12, which encodes NADP-GDH, has been cloned in several laboratories ([2–4] and this work) by selection of recombinant plasmids that complement a glutamate auxotroph of *E. coli* lacking both NADP-GDH and glutamate synthase. We have determined most of the DNA sequence of the *E. coli gdhA* gene, and report here a section of coding sequence that is very strongly conserved in prokaryotic and eukaryotic NADP-GDHs. The mammalian GDHs, which can utilise either NADP or NAD (EC 1.4.1.3), and a fungal NAD-specific GDH (EC 1.4.1.2), possess corresponding sequences that are less strongly homologous. This sequence comparison suggests that genes encoding these 3 classes of GDHs, of different coenzyme specificity, diverged at a very early stage of evolution before the divergence of a separate eukaryotic line. We also report the map positions of the chromosomal *gdhA1* point mutation and some restriction sites around the *gdhA* region of *E. coli*.

*To whom correspondence should be addressed

2. MATERIALS AND METHODS

2.1. *Bacterial strains and plasmids*

Escherichia coli K12 glutamate auxotroph, strain CLR207 (F^-; *thi-1 leu-6 his-1 argH1 gltB31 gdhA1 lacY1 gal-6 xyl-7 mtl-2 rpsL9 trpR tonA2 tsx-68 supE44 hsdR*$_K^-$ *hsdM*$_K^+$; λ^-), was constructed by Claudia L. Riordan in this laboratory from strain CB100 [5] by transduction with P1 propagated on strain 803 (obtained from W. Brammar). Strain CLR207 *recA* was derived from CLR207 by a further transduction with P1 propagated on a *recA srl*::*Tn10*(Tet^R) strain originating from N. Kleckner's laboratory. The inserts in different recombinant plasmids carrying the *gdhA* gene are defined in the legend to fig.1. pSG1 was obtained from J.D. Windass (named pACYC184 *gdh* [2]) and was also selected independently in this laboratory by complementation in CLR207*recA* from fragments of a *Sal*I digest of 803 DNA ligated into the *Sal*I site of pACYC184[6]. pBG1 and pBG3 were similarly selected from a ligation of a *Bam*HI digest of 803 DNA into the *Bam*HI site of pACYC184, and a plasmid apparently identical to pBG1 was also obtained from J.D. Windass. pG14F2 and pG12C6 are hybrid cosmids selected by complementation in CLR207*recA* following an in vitro packaging procedure [7] using a partial *Sau*3A digest of 803 DNA ligated into the *Bam*HI site of cosmid pHC79 [8].

2.2. *Construction of deleted derivatives of pBG1 and pSG1*

The extents of deletions in 6 derivatives are shown in fig.1(b). These plasmids were obtained

Published by Elsevier Biomedical Press
00145793/82/0000–0000/$2.75

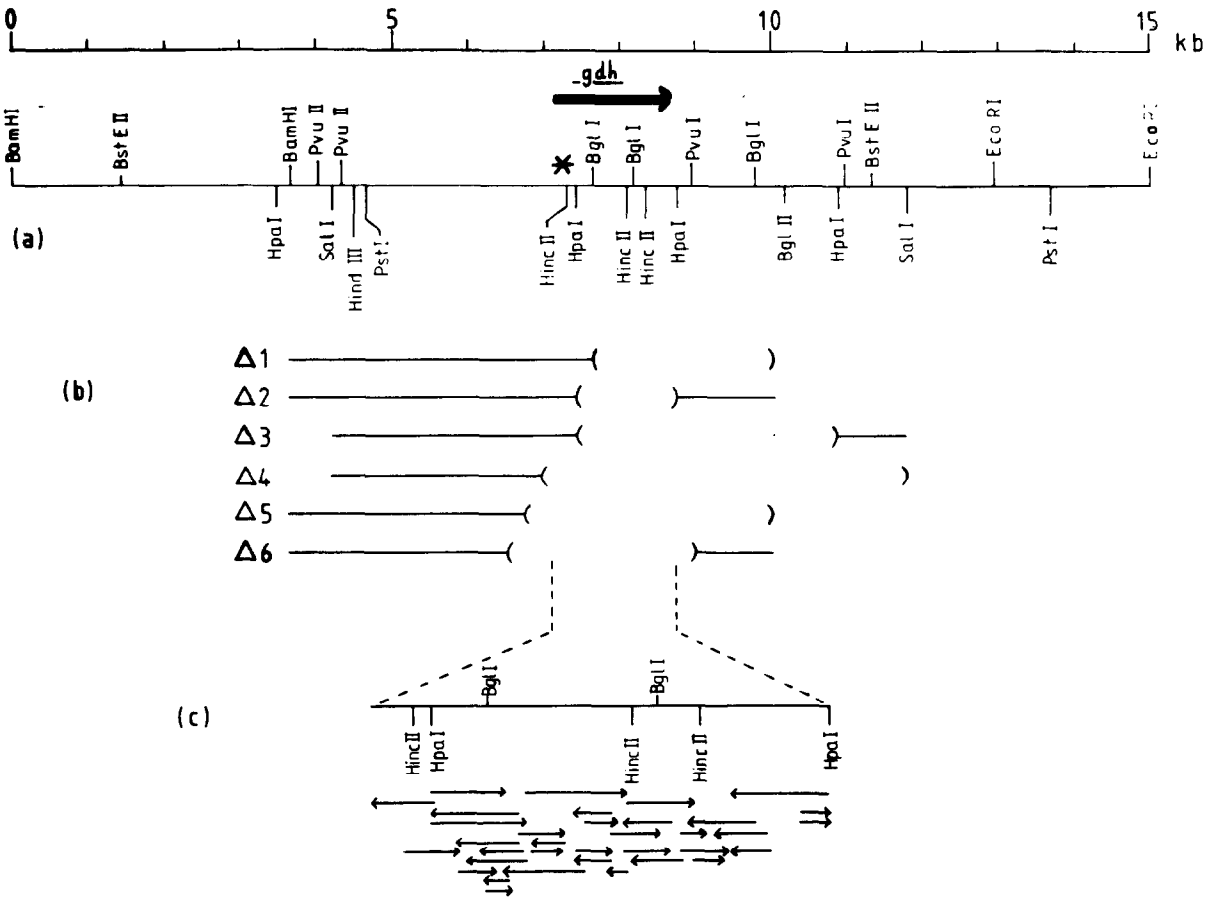

Fig.1. (a) Restriction map of 15 kb of *E. coli* DNA including the *gdhA* gene. The extent of the gdhA coding sequence and direction of transcription are shown by the rightward-pointing arrow. Cleavage sites for the enzymes shown were mapped from fragment sizes generated by single and double, complete and partial digests using the following plasmids: pSG1 (insert coordinates 4.2–11.8, between *Sal*I sites, in *Sal*I site of pACYC184); pBG1 (insert coordinates 3.7 (*Bam*HI site)–10.2, in *Bam*HI site of pACYC184); pBG3 (as pBG1 but insert coordinates 3.7–13.5); pG14F2 (a hybrid cosmid with insert extending 35 kb leftwards from coordinate 10.1, vector pHC79); pG12C6 (a hybrid cosmid with insert including the complete 15 kb shown, vector pHC79). Precision of mapping (~95% confidence limits) was ± 0.05 kb for intervals <0.8 kb (measured relative to φX174 *Hae*III fragments using 3.5% acrylamide/0.5% agarose gel electrophoresis) and ± 0.2 kb for intervals >0.8 kb (measured relative to λ *Hin*dIII fragments using 1% agarose gel electrophoresis). Colinearity of the restriction map with *E. coli* chromosomal DNA was shown by hybridisation of Southern blots of strain 5K DNA using nick-translated *Pst*I–*Pvu*I fragment (coordinates 4.7–8.9) as probe, which showed fragment sizes as follows: *Bst*EII, 9.8 kb; *Hpa*I, 3.9, 2.1 and 1.3 kb; *Sal*I, 7.6 kb; *Pst*I, 9.0 kb; *Bgl*II, 12.0 kb; *Bam*HI, ~25 kb; *Eco*RI, ~23 kb; *Hin*dIII, 14 kb; (*) position of chromosomal *gdhA* mutation in CLR207, mapped by marker rescue. (b) Deleted derivatives of pBG1 (Δ1, Δ2, Δ5, Δ6) and pSG1 (Δ3, Δ4) used in marker rescue experiments: (—) DNA present in each derivative; ( ) deleted segments. (c) Clones used for sequencing DNA of the *gdhA* gene: (⇌) extent and direction of sequences obtained from different clones in M13mp7.

by re-ligation of partial digests of pBG1 (Δ1, Δ2, Δ5, Δ6) or pSG1 (Δ3, Δ4) digested with *Bgl*I (Δ1, Δ4, Δ5) or *Hpa*I (Δ2, Δ3, Δ6), using a high concentration of T4 DNA ligase and low concentration of plasmid DNA to encourage circularisation by ligation of blunt ends (*Hpa*I) or incompletely complementary tails (*Bgl*I). Ligation mixtures were transformed into CLR207 and chloramphenicol-

resistant transformants screened for glutamate auxotrophy. Extents of deletions in purified *gdhA*$^-$ derivative plasmids were mapped as described for the original *gdhA*+ plasmids (legend to fig.1) and confirmed by heteroduplex mapping. In Δ1 a *Bgl*I site was reconstituted between vector and insert DNA, and in Δ2 and Δ3 new *Hpa*I sites were reconstituted within insert DNA. Further deletions occurred in Δ4, Δ5 and Δ6 before ligation.

2.3. *DNA sequencing*

The 1.3 kilobase (kb) *Hpa*I fragment of pBG1, containing most of the *gdhA* gene (coordinates 7.4 – 8.7, fig.1(a)), was recovered from a 1% low melting point agarose gel and cloned into *Hin*cII-cleaved M13mp7. This *Hpa*I fragment was also subdigested with *Hae*III, *Hpa*II, *Taq*I or *Sau*3A and these digests cloned into M13mp7 cleaved with, respectively, *Hin*cII, *Acc*I, *Acc*I and *Bam*HI. A clone extending the sequence leftwards from the *Hpa*I site (at 7.4., fig.1(a)) was obtained by ligating the larger *Hpa*I fragment of pBG1 into *Hin*cII-cleaved M13mp7, and a clone overlapping the same *Hpa*I site was isolated from *Hae*III-cleaved pBG1 cloned into *Hin*cII-cleaved M13mp7. Methods for manipulating M13mp7 and *E. coli* strain JM101 (both obtained from the MRC Laboratory, Cambridge) were as described in [9–11]. DNA sequences were determined by the dideoxy method [11] using synthetic 17-mer primer prepared at the MRC Laboratory, Cambridge [12].

3. RESULTS

3.1. *Map of restriction sites and the* gdhA1 *mutational site*

Fig.1(a) shows the restriction map of 15 kb of the *E. coli* chromosome, including the *gdhA* gene, as deduced from pSG1, pBG1, pBG3, pG14F2, pG12C6, and blot hybridisations of restriction digests of *E. coli* chromosomal DNA (see fig.1 legend). This map agrees, within error, with less extensive published maps [2,4]. The location and orientation of the *gdhA* coding sequence (fig.1(a)) is deduced from the translation of the DNA sequence, and demonstrates that a previous mapping based on minicell experiments [4] was essentially correct.

The site of the *gdhA1* mutation was mapped by marker rescue experiments in which the deleted plasmids, Δ1–Δ6 (fig.1(b)), were tested for their ability to correct by recombination the *gdhA1* lesion in CLR207. Transformants of CLR207 carrying plasmids Δ1, Δ2 or Δ3 segregated Gdh+ recombinants at a frequency ($\sim 10^{-3}$) typical of recombinational events between episomes and the *E. coli* chromosome. Transformants of CLR207 carrying Δ4, Δ5 or Δ6 gave no Gdh+ recombinants. None of these 6 plasmids gave Gdh+ recombinants following transformation into CLR207*recA*, showing that correction of *gdhA1* required the *recA*+-dependent homologous recombination pathway. This demonstrates that the mutational site lies between the *Hpa*I site and the deletion endpoint in Δ4 (coordinates 7.4 and 7.1, fig.1(a)). It is a revertible ($\sim 10^{-8}$) point mutation, probably causing an amino acid change within the N-terminal 80 residues of the NADP-GDH polypeptide, consistent with the observation [4] that strain CB100 produces an inactive protein of the same M_r-value as NADP-GDH which is precipitated by anti-NADP-GDH antibodies.

3.2. *GDH sequence comparisons*

Fig.2 shows 354 nucleotides of coding sequence of the *E. coli gdhA* gene, and the deduced amino acid sequence which is compared with the corresponding parts of published amino acid sequences of *Neurospora crassa* NADP-GDH [13] and bovine liver GDH [14]. The bacterial and fungal NADP-GDHs are very strongly homologous showing 79% amino acid identities (93 out of 118 amino acids compared). Bovine GDH is considerably less strongly homologous to these NADP-GDHs than are the NADP-GDHs to each other. It shows 28% (33/120 amino acid identities) and 31% (37/120 identities) homology to respectively *E. coli* and *N. crassa* NADP-GDHs. The corresponding section of the amino acid sequence of the distinct NAD-specific GDH of *N. crassa* [15] is more distantly related to all 3 of these GDHs, showing ~17%, 18% and 25% amino acid identities (over this 118-residue section) to, respectively, *E. coli* and *N. crassa* NADP-GDHs and the bovine enzyme (based on an alignment [16], not shown in fig.2, which is arbitrary in parts although supported by secondary structure predictions [16]).

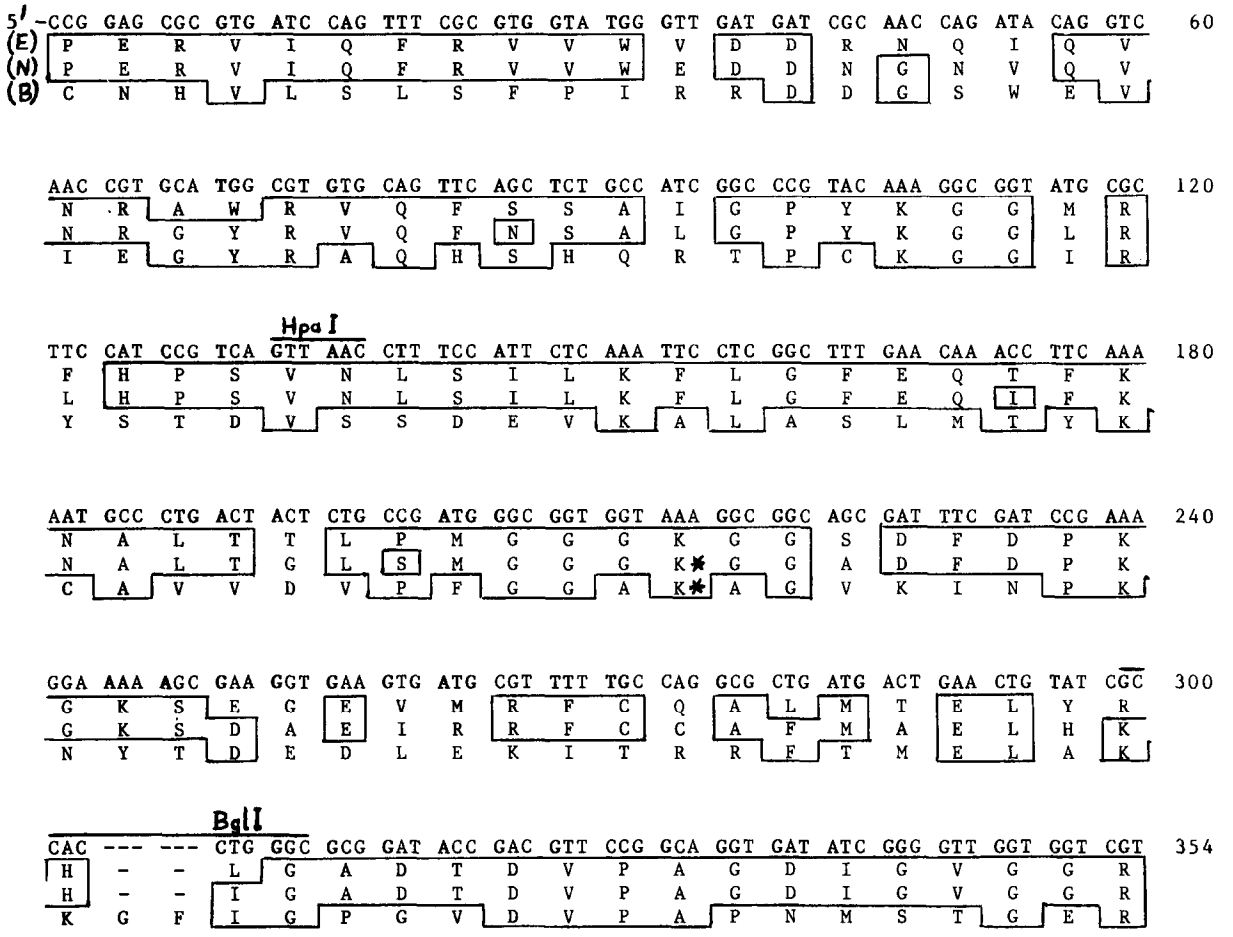

Fig.2. A strongly-conserved section of *gdhA* coding sequence. The DNA sequence was determined from 16 independent overlapping clones in M13mp7 (fig.1(c)) covering both strands and leaving no ambiguities. *Hpa*I and *Bgl*I sites (coordinates 7.4 and 7.6, fig.1(a)) are shown as landmarks. The deduced amino acid sequence of the corresponding part of *E. coli* NADP-GDH (E) is shown below the DNA sequence, aligned with the homologous sections of *N. crassa* NADP-GDH (N, amino acids 42–159) and bovine liver GDH (B, amino acids 55–174, including two residues, 156–157, that have no equivalents in the other GDH sequences). (*) Reactive lysine (−113 in *N. crassa*, −126 in bovine GDHs); (□) enclose identical amino acid sequences.

4. DISCUSSION

GDHs show an unusually strong degree of sequence conservation, as demonstrated by comparisons of human, bovine and chicken liver enzymes [14] and the remarkable similarity of parts of the prokaryotic (*E. coli*) and eukaryotic (*N. crassa*) NADP-GDHs reported here (fig.2). Recent experiments (J.C.W., B. Hohn unpublished) have shown that labelled probes made from sections of the *E. coli gdhA* coding sequence cross hybridise at low stringency with DNA sequences in Southern blots of genomic DNA or cloned genomic DNA libraries prepared from several distantly related prokaryotes, fungi and plants. This suggests that strongly-conserved genes, presumably encoding GDHs, are widespread in these diverse classes of organisms.

The sequence comparisons support the idea that the 3 distinct classes of GDH of known sequence diverged from each other at a very early stage of cellular evolution, before the existence of a sepa-

rate eukaryotic line. Probably the class of homotetrameric NAD-specific GDHs diverged first from a common ancestral GDH line, and a more recent divergence occurred between the homohexameric NADP-GDHs (as found in bacteria and fungi) and the homohexameric GDHs of dual coenzyme specificity (as found in vertebrate livers). This interpretation assumes that the strong homology of bacterial and fungal NADP-GDHs does not reflect inter-kingdom transfer of *gdh* genes between prokaryotes and eukaryotes in relatively recent evolutionary time.

The section of GDH polypeptide shown in fig.2, which is generally the most conserved part of the sequence [13–17], probably includes the residues involved in binding of dicarboxylate substrates, in catalytic activity and in interactions affecting an allosteric conformational equilibrium. These functions are supported (in the absence of a crystallographic structure determination of any GDH) by chemical modification studies (summarised in [14–17]) and the properties of mutational variants [18,19]. A coenzyme-binding domain has been identified [16,17] in the less conserved C-terminal half of the molecule.

ACKNOWLEDGEMENTS

We thank John Windass, Claudia Riordan, Bill Brammar and Barbara Hohn for bacterial strains and plasmids, and members of the MRC Laboratory, Cambridge, for advice on DNA sequencing and gifts of primer and strains. Financial support was from an MRC project grant (J.C.W. and I.W.M.), an SRC studentship (M.J.M), and in part from an EMBO fellowship supporting visiting work (J.C.W.) in Barbara Hohn's Laboratory, Friedrich Miescher Institut, Basel. We thank Nick Munton for technical assistance.

REFERENCES

[1] Tyler, B. (1978) Annu. Rev. Biochem. 47, 1127–1162.
[2] Windass, J.D., Worsey, M.J., Pioli, E.M., Pioli, D., Barth, P.T., Atherton, K.T., Dart, E.C., Byrom, D., Powell, K. and Senior, P.J. (1980) Nature 287, 396–401.
[3] Covarrubias, A.A., Sánchez-Pescador, R., Osorio, A., Bolivar, F. and Bastarrachea, F. (1980) Plasmid 3, 150–164.
[4] Sánchez-Pescador, R., Sanvicente, E., Valle, F. and Bolivar, F. (1982) Gene 17, 1–8.
[5] Berberich, M.A. (1972) Biochem. Biophys. Res. Commun. 47, 1498–1503.
[6] Chang, A.C.Y. and Cohen, S.N. (1978) J. Bacteriol. 134, 1141–1156.
[7] Hohn, B. (1980) Methods Enzymol. 68, 299–309.
[8] Hohn, B. and Collins, J. (1980) Gene 11, 291–298.
[9] Winter, G. and Fields, S. (1980) Nucleic Acids Res. 8, 1965–1974.
[10] Winter, G., Fields, S., Gait, M.J. and Brownlee, G.G. (1981) Nucleic Acids Res. 9, 237–245.
[11] Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.H.J. and Roe, B.A. (1980) J. Mol. Biol. 142, 161–178.
[12] Duckworth, M.L., Gait, M.J., Goelet, P., Hong, G.F., Singh, M. and Titmas, R.C. (1981) Nucleic Acids Res. 9, 1691–1706.
[13] Holder, A.A., Wootton, J.C., Baron, A.J., Chambers, G.K. and Fincham, J.R.S. (1975) Biochem. J. 149, 757–773.
[14] Julliard, J.H. and Smith, E.L. (1979) J. Biol. Chem. 254, 3427–3428.
[15] Haberland, M.E. and Smith, E.L. (1980) J. Biol. Chem. 255, 7984–7992.
[16] Austen, B.M., Haberland, M.E. and Smith, E.L. (1980) J. Biol. Chem. 255, 8001–8004.
[17] Wootton, J.C. (1974) Nature 252, 542–546.
[18] Brett, M., Chambers, G.K., Holder, A.A., Fincham, J.R.S. and Wootton, J.C. (1976) J. Mol. Biol. 106, 1–22.
[19] Ashby, B., Nixon, J.S. and Wootton, J.C. (1981) J. Mol. Biol. 149, 521–540.