Isolation and Characterization of cDNA Clones for Rat Liver 10-Formyltetrahydrofolate Dehydrogenase*

(Received for publication, August 22, 1990)

Robert J. Cook‡§, R. Stephen Lloyd†‡, and Conrad Wagner‡¶

From the §Department of Biochemistry and the ¶Center in Molecular Toxicology, Vanderbilt University School of Medicine, Nashville, Tennessee 37232 and the ‡Veterans Administration Medical Center, Nashville, Tennessee 37212

We have isolated and characterized cDNA clones encoding rat liver cytosol 10-formyltetrahydrofolate dehydrogenase (EC 1.5.1.6). An open reading frame of 2706 base pairs encodes for 902 amino acids of $M_r$ 99,015. The deduced amino acid sequence contains exact matches to the NH$_2$-terminal sequence (28 residues) and the sequences of five peptides derived from cyanogen bromide cleavage of the purified protein. The amino acid sequence of 10-formyltetrahydrofolate dehydrogenase has three putative domains. The NH$_2$-terminal sequence (residues 1-203) is 24-30% identical to phosphoribosylglycinamide formyltransferase (EC 2.1.2.2) from Bacillus subtilis (36%), Escherichia coli (24%), Drosophila melanogaster (24%), and human hepatoma HepG2 (27%). Residues 204-416 show no extensive homology to any known protein sequence. Sequence 417-900 is 46% (mean) identical to the sequences of a series of aldehyde dehydrogenase (NADP$^+$) (EC 1.2.1.1). Intact 10-formyltetrahydrofolate dehydrogenase exhibits NADP$^+$-dependent aldehyde dehydrogenase activity. The sequence identity to phosphoribosylglycinamide formyltransferase is discussed, and a binding region for 10-formyltetrahydrofolate is proposed.

The enzyme 10-formyltetrahydrofolate dehydrogenase (EC 1.5.1.6) catalyzes the NADP$^+$-dependent oxidation of 10-HCO-H$_4$PteGlu to H$_4$PteGlu and CO$_2$. The enzyme was first described in detail by Kutzbach and Stokstad (1) in pig liver, and it has subsequently been purified to apparent homogeneity from pig liver (2) and rat liver (3). In each case, purified 10-HCO-P$_4$teGlu dehydrogenase has also been shown to perform an NADP$^+$-independent hydrolysis of 10-HCO-H$_4$PteGlu to yield H$_4$PteGlu and formate, leading to the assumption that this enzyme, like other folate-dependent enzymes, has more than one activity (1-3).

Cook and Wagner (4) have purified and described a rat liver cytosol folate-binding protein denoted FBP-CI. This protein has a subunit $M_r$ of 100,000 and contains tightly bound H$_4$PteGlu. No enzymatic function requiring H$_4$PteGlu could be ascribed to this protein. Recently, it was shown by Min et al. (5) that FBP-CI is 10-HCO-H$_4$PteGlu dehydrogenase, although these authors did not purify the protein to homogeneity. We have since confirmed this observation by purifying FBP-CI from rat liver cytosol to apparent homogeneity and showing concomitant purification of H$_4$PteGlu, rebinding and 10-HCO-H$_4$PteGlu dehydrogenase and hydrolysis activities (52). Case et al. (6) recently reported that the dehydrogenase and hydrolyase activities reported above (1-3) actually reside in two separate proteins that are compartmentalized. Their experiments using rat liver showed that 10-HCO-H$_4$PteGlu dehydrogenase is present exclusively in cytosol, whereas the 10-HCO-H$_4$PteGlu hydrolyase is exclusive to the mitochondria. The original procedure for purification of 10-HCO-H$_4$PteGlu dehydrogenase used liver that had been homogenized under conditions which did not protect the integrity of the mitochondria (1-3), whereas Case et al. (6) used liver that had been homogenized with a Teflon pestle in isotonic buffered mannitol/sucrose/EDTA. Case et al. (6) claim that the dehydrogenase and hydrolyase enzymes are so closely related that they will copurify under conditions where mitochondria are damaged by the homogenization procedure. These results are clearly contrary to those of Wagner et al. (52) and those of Barlowe and Appling (7), who showed clearly that there is 10-HCO-H$_4$PteGlu dehydrogenase activity in rat liver mitochondria and that this activity is predominantly confined to the mitochondrial matrix.

Another issue concerning 10-HCO-H$_4$PteGlu dehydrogenase is its physiological function. Johlin et al. (8) suggest that it functions in the metabolism of formate, which involves the conversion of formate to 10-HCO-H$_4$PteGlu by the action of 10-formyltetrahydrofolate synthetase (EC 6.3.4.3) and the subsequent oxidation of this product by 10-HCO-H$_4$PteGlu dehydrogenase to CO$_2$ and H$_4$PteGlu. In primates, this is the exclusive pathway for metabolizing formate (9), and the investigation of 10-HCO-H$_4$PteGlu dehydrogenase is of prime importance in understanding the mechanism of methanol poisoning in humans (8). Krebs et al. (10) proposed that 10-HCO-H$_4$PteGlu dehydrogenase is the major route for the disposal of excess one-carbon units produced in folate-mediated amino acid metabolism. The physiological significance of the hydrolyase activity is not clear. It can recycle 10-HCO-H$_4$PteGlu back to H$_4$PteGlu, but the formate produced would
be available for reincorporation into 10-HCO-H4PteGlu in a potentially " futile cycle" (2). With the pig liver enzyme, both dehydrogenase and hydrolase activities have been shown to occur simultaneously, suggesting two active sites (2).

10-HCO-H4PteGlu dehydrogenase is an abundant enzyme in rat liver cytosol, where it is present at 0.5-1.0% of the cytosolic protein and binds a substantial proportion of the H4PteGlu pool (4). This observation and the fact that H4PteGlu, the product of both the dehydrogenase and hydrolase reactions, is a very potent inhibitor of these reactions with a Ki value approximately equal to the Km value for the substrate, 10-HCO-H4PteGlu (1-3), suggest that this enzyme may have another as yet, unknown function. As an initial step in resolving the questions of bifunctionality and physiological function of 10-HCO-H4PteGlu dehydrogenase, we have isolated and characterized cDNA clones of this enzyme. The translated 902-amino acid open reading frame reveals a protein of three distinct domains. Two of these domains show extensive identity to phosphoribosylglycinamidine formyltransferase (EC 2.1.2.2) and aldehyde dehydrogenase (NADP+)-dependent alcohol dehydrogenase (EC 1.2.1.3), whereas the third domain shows no homology to any known protein sequence.

**EXPERIMENTAL PROCEDURES**

**Materials**—All common chemicals were of the highest grade available. Restriction enzymes and T4 polynucleotide kinase were obtained from New England Biolabs, Inc. HPLC solvents and water were supplied by Burdick & Jackson Laboratories Inc. Trifluoroacetic acid was purchased from Pierce Chemical Co. CNBr was obtained from Fisher. Nitrocellulose and NA45 membrane were obtained from Schleicher & Schuell. Custom synthetic oligonucleotides for cDNA library screening and nucleotide sequencing were supplied by Research Genetics (Huntsville, AL).

**Bacterial Strains**—Escherichia coli strains BB4 and XL1-Blue were supplied by Stratagene and were grown and used as described by the supplier.

**Purification of Rat Liver Cytosol 10-HCO-H4PteGlu Dehydrogenase**—10-HCO-H4PteGlu dehydrogenase was purified as described by Cook and Wagner (4, 11) for FBP-CI, with the exception that dehydrogenase and hydrolase activities were assayed by the method of Kutzbach and Stokstad (1), in addition to the measurement of binding of [H]H4PteGlu. The purification involved the isolation of rat liver cytosol by centrifugation (100,000 × g for 60 min) after gentle homogenization in 0.25 M sucrose, 10 mM potassium phosphate (pH 7.0), 0.1 mM 2-mercaptoethanol, 1 mM sodium azide, and 0.1 M phenylmethylsulfonyl fluoride. This was followed by Sephadex G-150 gel chromatography, DEAE-cellulose chromatography, affinity chromatography using 5-HCO-H4PteGlu linked to aminoethyl-Sepharose 4B (4) and finally Mono Q (50-500 ng; calmodin exchange; Pharmacia LKB Biotechnology Inc.) column chromatography. The protein was judged pure by virtue of the presence of a single band on SDS-polyacrylamide gel electrophoresis, nondenaturing gel electrophoresis, and chromatofocusing. The specific activity of pure dehydrogenase was 0.19 µmol of NADPH min-1 mg-1 measured at room temperature (22 °C).

**Cyanogen Bromide Cleavage of 10-HCO-H4PteGlu Dehydrogenase**—Pure protein (1 mg) was lyophilized and dissolved in 100 µl of 70% (v/v) formic acid, Cyanogen bromide (1 mg) was added; and the tube was flushed with nitrogen, sealed, and incubated at room temperature for 24 h in the dark. The reaction mixture was evaporated to dryness under a stream of nitrogen, dissolved in 0.1% trifluoroacetic acid, and stored at -20 °C.

**Purification of CNBr Peptides**—Peptides were separated and purified by reverse-phase HPLC using a SynChropak RP-P column (4.1 x 250 mm; SynChrom, Lindert, IN) as described by Cook et al. (12). Digests dissolved in 0.1% trifluoroacetic acid were applied to the column, and peptides were eluted either with acetonitrile/ water (80:20, v/v) containing 0.1% trifluoroacetic acid or with isopropl alcohol: water (80:20, v/v) containing 0.1% trifluoroacetic acid. Peptides were detected at 214 nm. Peaks were collected and lyophilized.

**Amino Acid Sequence Analysis**—Automated amino acid sequence analysis of CNBr peptides or complete protein was performed on either an Applied Biosystems Model 470A Gas-Phase Sequencer or a Model 475 Pulse-Liquid Sequencer using the manufacturer’s protocol. Samples contained a minimum of 25 pmol.

**Construction of Rat Liver cDNA Library**—Rat liver poly(A)+ RNA (250-4000 bp) was obtained from Clontech. The cDNA library was generated using a Pharmacia LKB Biotechnology cDNA synthesis kit following the manufacturer’s protocols. The cDNA with cohesive EcoRI linkers was ligated into XZAP (Stratagene) and packaged using Gigapack Plus kit (Stratagene) using the host strain according to the manufacturer’s protocols. The library was put through one round of amplification using BB4 as the host strain according to Stratagene protocols.

**Isolation of Rat 10-HCO-H4PteGlu Dehydrogenase cDNA Clones**—The rat liver cDNA library in XZAP was grown in the E. coli host strain BB4 according to the manufacturer’s protocols. Bacterial colonies were screened using a double-lift procedure, where plaques were transferred to nitrocellulose filters (80 mm) in the first lift for 1 min and in the second lift for 5 min. The nitrocellulose filters were processed for hybridization as described by Maniatis et al. (13). The filters were prehybridized for 2 h at 25 °C in 20% formamide, 5 x SSPE (1 x SSPE = 0.15 M NaCl, 10 mM NaH2PO4, 1.3 mM EDTA (pH 7.4), 2.5 x Denhardt’s solution, 0.1% SDS, and 62.5 µg/ml denatured calf thymus DNA. Positive plaques were identified by hybridization with 5’-32P-labeled best-guess oligonucleotides added at 5 x 10⁶ cpm/ml of hybridization fluid. Hybridization was performed at 25 °C for 16 h. The filters were washed at 38 °C with 5 x SSPE at 15 min in 2 x SSPE. Filters screened with oligonucleotide 5 were then washed in 4 x SSPE at 30 °C for 30 min and then at 34 °C in fresh 4 x SSPE. Filters screened with oligonucleotide 6 were washed under the same conditions and then at 38 and 42 °C in 2 x SSPE. The filters were monitored at all stages of the washing procedure for background radioactivity. The filters were then exposed for 16 h at -70 °C with intensifying screens. Selected double positive clones were coded into sterile SM media (13) and rescreened through two more rounds until plaque-pure.

**Southern Analysis**—Southern analysis combined with restriction enzyme digestion was used to characterize positive clones. Digested DNA was subjected to agarose gel electrophoresis and then transferred to nitrocellulose bidirectionally using the method of Smith and Summers (14). The filters were hybridized to either 5’-end-labeled oligonucleotide or nick-translated cDNA. The hybridization fluid was as described above, whereas hybridization temperatures and washing conditions are given in the appropriate figure legends.

**Northern Analysis**—Total RNA from rat liver was extracted by the method of Chirgwin et al. (15) and was kindly donated by Dr. K. E. Hill (Division of Gastroenterology, Vanderbilt University School of Medicine). RNA was separated by electrophoresis on 1.2% formaldehyde-agarose gels as described by Maniatis et al. (13) and then transferred to GeneScreen Plus using the manufacturer’s protocol. Prehybridization was carried out at 42 °C for 2 h in 50% formamide, 10% dextran sulfate, 1% SDS, 1 mM NaCl, 5 x Denhardt’s solution, and 100 µg/ml denatured calf thymus DNA. Hybridization was performed for 16 h at 42 °C with denatured nick-translated 3.2-kb cDNA fragments (2 x 10⁶ cpm/ml). After washing the filters with 2 x SSC (1 x SSC = 0.15 M NaCl, 15 mM sodium citrate) at room temperature for 5 min and then at 65 °C for 30 min in 2 x SSC + 0.1% SDS and was finally rinsed in 0.2 x SSC + 0.1% SDS prior to exposure for 2 h at room temperature.

**Production of 32P-Labeled DNA Probes**—Purified cDNA fragments (50-500 ng) were nick-translated using a DuPont-New England Nuclear nick translation kit according to the supplied protocol. Labeled fragments were separated from unincorporated [α-32P]CTP by the use of a Sephadex G-50 spin column procedure (13). Oligonucleotides (100 ng) were 5’-end-labeled with [γ-32P]ATP using T4 polynucleotide kinase and purified on a Sephadex G-25 spin column as described by Maniatis et al. (13).

**DNA Sequencing**—Plaque-pure positive clones for 10-HCO-H4PteGlu dehydrogenase in XZAP were excised and recircularized into pBluescript using the helper phage R408 as described by Strano et al. (13).
tagene protocols. The rescued phagemid was then selected by growth in E. coli XL1-Blue on LB-ampicillin plates. Rescued XL1-Blue colonies were maintained on LB-ampicillin, and large amounts of phagemid DNA were produced from these colonies by growth of XL1-Blue in liquid culture. Restriction mapping showed the presence of many of the unique Bluescript restriction sites within the cDNA inserts, which eliminated the use of nested deletions as a method of sequencing. All sequences were derived from dideoxy sequencing (16) using [α-32P]dATP with Sequenase (U. S. Biochemical Corp.) and base-denatured double-stranded template following the manufacturer's protocol. Both strands were sequenced using specifically synthesized oligonucleotide primers.

Sequence Compilation and Analysis—Nucleotide sequences were compiled and analyzed using DNA Strider 1.0. Data base searches and sequence alignments were performed either by FASTA (17) or by FASTDB (copyrighted software product of IntelliGenetics Inc.). Multiple sequence alignments were performed by the Needleman-Wunsch algorithm (18) GENALIGN (a copyrighted software product of IntelliGenetics Inc.; the program was developed by Dr. Hugo Martinez (University of California, San Francisco)).

Aldehyde Dehydrogenase Assay—Aldehyde dehydrogenase activity was assayed according to the method of Lindahl and Evces (19). The aldehydes (50 mM) used were formaldehyde, acetaldehyde, propionaldehyde, and benzaldehyde. NAD or NADP was present at a final concentration of 10 mM.

RESULTS

Amino Acid Sequencing and Best-guess Oligonucleotide Design—Pure 10-HCO-H,PteGlu dehydrogenase from two different preparations was subjected to automated Edman degradation using both a Gas-Phase Sequencer and a Pulse Liquid Sequencer. The derived sequence (MKIAVIGQSLFGQEVYXQLRKEGHEVVG) agrees with that from the first residue is identified as methionine and that arginine is confirmed at position 20. The reason for the failure of Johlin et al. (8) to identify the NH₂-terminal amino acid was ascribed to the NH₂-terminal methionine residue is not apparent, except that their purification procedure results in very low yields of pure 10-HCO-H,PteGlu dehydrogenase.

Peptides generated by CNBr digestion of 10-HCO-H,PteGlu dehydrogenase were separated and purified to apparent homogeneity by reverse-phase HPLC (see "Experimental Procedures"). Five peptides, 9, 15, 17A, 17B, and 23, were selected and sequenced. (These sequences are underlined in Fig. 3.) Best-guess sense oligonucleotides were designed to peptide 15 (MYIAKKE), ATGTGACACTCGCCAAGGGAGGAG (oligonucleotide 5), and peptide 23 (MKIGNPLE), ATGAA-GATCGGGAACCCTCCTGAGG (oligonucleotide 6), using mammalian cDNA sequence data (20, 21). A third, best-guess antisense oligonucleotide was designed to the NH₂-terminal sequence (MKIAVIGQSLFG), CCGACAGGCGACTGCGCCWGTCCCTCAT (oligonucleotide 3).

Isolation, Purification, and Characterization of cDNA Clones 10-HCO-H,PteGlu Dehydrogenase—An initial screen of 5 × 10⁶ plaques gave 17 positive clones to oligonucleotide 5 and 81 positive clones to oligonucleotide 6. Comparison of the autoradiographs showed that 23 were double positive clones, of which 12 were cored. A second round of plaque purification reduced the number of double positive clones to eight. These were screened through a third and final round of plaque purification. Southern analysis of EcoRI or ClaI-digested DNA from these clones with 5⁻²³P-labeled best-guess oligonucleotide 5 revealed insert sizes which ranged from 2.0 to 4.5 kb (data not shown). The eight plaque-pure clones were excised from XZAP and recircularized into the Bluescript phagemid (see "Experimental Procedures"). DNA was purified from the clones, digested with EcoRI, and electrophoresed on an agarose gel (Fig. 1A). Four clones contained 3.2-kb inserts; these three also contained smaller fragments. The

![Fig. 1. Agarose gel electrophoresis and Southern blot analysis of 10-HCO-H,PteGlu dehydrogenase clones. A, agarose gel electrophoresis of cDNA from clones in pBluescript. DNA (2 μg) was digested with EcoRI for 16 h at 37 °C and run on a 1% agarose gel. Lanes 1 and 10, DNA standards of λ HindIII and αX174 digested with HaeIII; lane 2, clone 13C; lane 3, clone 15C; lane 4, clone 17C; lane 5, clone 5-1C; lane 6, clone 14-1C; lane 7, clone 14-2C; lane 8, clone 16C; lane 9, clone 18C. B, bidirectional Southern blot of A probed with nick-translated 3.2-kb insert of clone 15C (see "Experimental Procedures"). The filter was prehybridized for 3 h at 42 °C and hybridized for 16 h at 42 °C with 5 × 10⁶ cpm/ml 32P-probe as described under "Experimental Procedures." The filter was rinsed in 4 × SSPE at room temperature and then washed in 4 × SSPE for 30 min each at 42, 50, 65, and 70 °C and finally rinsed in 0.1 × SSPE at room temperature. The filter was then exposed for 3 h at room temperature. C, other Southern blot of A probed with 5⁻²³P-oligonucleotide 3. The filter was prehybridized at 30 °C for 3 h and hybridized for 16 h at 30 °C with 5 × 10⁶ cpm/ml 32P-probe as described under "Experimental Procedures." The filter was then rinsed in 4 × SSPE at room temperature and then washed in 4 × SSPE for 30 min at 35 and 40 °C and finally rinsed in 0.1% SSPE at room temperature. The filter was exposed for 3 h at ~70 °C with intensifying screens.

other four clones contained inserts of 2.7 or 2.2 kb. The gel was bidirectionally blotted onto nitrocellulose and probed either with a nick-translated 3.2-kb insert purified from clone 15 (Fig. 1B) or with 5⁻²³P-labeled oligonucleotide 3, a best-guess antisense oligonucleotide to the NH₂-terminal amino acid sequence (Fig. 1C). The nick-translated 3.2-kb probe hybridized to the 3.2-, 2.7-, and 2.2-kb fragments (Fig. 1B) under very stringent washing conditions, indicating that these sequences were highly related. The end-labeled NH₂-terminal best-guess oligonucleotide only hybridized to the 3.2-kb inserts (Fig. 1C), which suggests that these inserts contained a complete open reading frame for 10-HCO-H,PteGlu dehydrogenase, whereas the other shorter clones were truncated at the 5'-end.

Sequencing of cDNA Clones—Three clones were chosen for sequencing, 17C (3.2-kb insert), 14-1C (2.7-kb insert), and 5-1C (2.2-kb insert). These three clones contained only one insert (Fig. 1A) and were suitable for sequencing a double-stranded template in both directions. Initially, clone 17C (3.2 kb) was sequenced with two primers, 7pBS and M13 universal primer, which read from the pBluescript vector into both ends of the insert, and with the two best-guess oligonucleotides, 5 and 6, used to isolate the clones. The primer 7pBS showed an ATG start codon 163 bp downstream of the left EcoRI linkage site. Translation of the nucleotide sequence revealed an amino acid sequence identical to that of the NH₂ terminus obtained by automated Edman degradation. The results from the M13 universal primer (which primes in the reverse direction) and oligonucleotides 5 and 6 revealed three overlapping sequences
The first domain extends from the NH2 terminus to amino acid 203 and shows identity to GAR transformylase. To facilitate easy discussion of these 12 sequences, a putative TGA stop codon; an AATAAA polyadenylation signal; and, 23 bp downstream, a short poly(A) tail. These results clearly show that clone 17C encodes the entire open reading frame for 10-HCO-H’PteGlu dehydrogenase.

The sequencing strategy is shown in Fig. 2. Both strands of each of the three clones, 17C, 14-1C, and 5-1C, were sequenced. The complete nucleotide sequence of clone 17C (3109 bp) and the deduced amino acid sequence are shown in Fig. 3. Only one extended open reading frame of 2706 bp coding for 902 amino acids was found. This deduced amino acid sequence contained the NH2-terminal sequence and all five of the sequences derived from the CNBr peptides. The deduced amino acid sequence has a calculated molecular weight of 100,015, which agrees with the SDS-polyacrylamide electrophoresis estimate of 100,000 (4).

All three clones had identical sequences, with the exception of the 3'-untranslated region. Clone 14-1C ended at bp 3058 (Fig. 3), whereas clone 5-1C did not contain bases 3093–3103 (Fig. 3); but it did have an extended 29-bp poly(A) tail.

Northern Analysis—Northern analysis of rat liver total RNA with a nick-translated full-length cDNA insert revealed two bands, one minor band just below the 28 S RNA marker at ~5.2 kb (equivalent in size to clone 17C and sufficient for the open reading frame) and a larger, major band just above the 28 S marker at ~3.7 kb (Fig. 4). The abundance of the 3.7-kb message suggests that it is the mature transcript and the 3.2-kb message may be a degradation product.

All three clones had identical sequences, with the exception of the 3'-untranslated region. Clone 14-1C ended at bp 3058 (Fig. 3), whereas clone 5-1C did not contain bases 3093–3103 (Fig. 3); but it did have an extended 29-bp poly(A) tail.

Northern Analysis—Northern analysis of rat liver total RNA with a nick-translated full-length cDNA insert revealed two bands, one minor band just below the 28 S RNA marker at ~5.2 kb (equivalent in size to clone 17C and sufficient for the open reading frame) and a larger, major band just above the 28 S marker at ~3.7 kb (Fig. 4). The abundance of the 3.7-kb message suggests that it is the mature transcript and the 3.2-kb band may be a degradation product.

### Discussion

We have isolated a full-length cDNA clone for rat liver 10-HCO-H’PteGlu dehydrogenase. The open reading frame codes for 902 amino acids of M, 99,015, which is in agreement with the predicted subunit size of 100,000 (4, 9). The NH2-terminal amino acid sequence and sequences of CNBr peptides derived from pure 10-HCO-H’PteGlu dehydrogenase are present in the deduced amino acid sequence.

A search of the two major protein sequence data bases (PIR, release 23; and SWISS-PROT, release 13) using FASTA (17) or FASTDB (IntelliGenetics Inc.) revealed that 10-HCO-H’PteGlu dehydrogenase has three putative domains (Fig. 5).

The first domain extends from the NH2 terminus to amino acid 203 and shows identity to GAR transformylase from Bacillus subtilis (22), E. coli (23), Drosophila (24, 25), and human hepatoma HepG2 (26) (Fig. 5). No identity to the yeast GAR transformylase (27) was seen on the initial data base search. Subsequently, an alignment was revealed using the Needleman-Wunsch algorithm (18) (data not shown). The extent of sequence identity increases from B. subtilis (99 amino acids) to the human enzyme, where it covers the entire length of the GAR transformylase sequence. The GAR transformylases from B. subtilis (22) and E. coli (23) are single polypeptides, whereas those from Drosophila (24, 25), human (26), and yeast (27) are the COOH-terminal domain in a trifunctional enzyme. The other two activities associated with this purine biosynthetic complex are 5’-phosphosulfobisglycinamidase synthetase (EC 6.3.4.13) and 5’-phosphoribosylpyrimidinase synthetase-GAR synthetase (EC 6.3.3.1), which form the sequence H2N-GAR synthetase-5’-phosphoribosylaminomimidazole synthetase-GAR transformylase-COOH (25). The role of GAR transformylase is to transfer a one-carbon unit from 10-HCO-H’PteGlu to 5’-phosphoribosylglycinamidase, the product of GAR synthetase, to yield 5’-phosphoribosyl-N-formylglycinamidase (28). These observations suggest that the NH2-terminal domain of 10-HCO-H’PteGlu dehydrogenase may bind 10-HCO-H’PteGlu.

The sequence HPSLLP (residues 106–111) and a glycine (residue 115) 4 residues downstream (Fig. 6) are strictly conserved in enzymes that use 10-HCO-H’PteGlu as a substrate, except in the yeast GAR transformylase (27), which has a short sequence identity to GAR transformylase from other species. This strictly conserved sequence does not occur in the trifunctional C1-THF synthase4 which synthesizes 10-HCO-H’PteGlu (29–32). However, visual inspection of the 10-HCO-THF synthetase domain (residues 305–935) (31) of human C1-THF synthase reveals a similar sequence (QPQQGPTFG). This sequence is conserved in other C1-THF synthases (29–32) and in 10-HCO-THF synthetase from Clostridium acidurici (33) and Clostridium thermocaceticum (34). By aligning all 12 sequences and introducing a gap before the final glycine in the 10-HCO-THF synthetase sequences, there is a strict conservation of 4 amino acids in this 10-residue sequence (Fig. 6), with the obvious exception of yeast GAR transformylase. To facilitate easy discussion of these 12 sequences, the residues have been numbered 1–10 in Fig. 6. The differences at positions 1 (histidine to glutamine or glutamate) and 5 (leucine to glycine) are split between the top six enzymes.
The primary structure of rat 10-HCO-H₄PteGlu dehydrogenase is depicted in the figure. The nucleotide sequence is shown with amino acid sequences derived from it. The figure includes a legend indicating the positions of stop codons, polyadenylation splice sites, and other important features. The molecular weight of the enzyme is provided as 4969.

**Primary Structure of Rat 10-HCO-H₄PteGlu Dehydrogenase**

The figure shows the amino acid sequence derived from the nucleotide sequence, indicating the primary structure of the enzyme. The figure includes a boxed section highlighting specific regions of interest, such as potential cleavage sites and functional domains.

---

**Legend:**
- **FITC:** Fluorescein isothiocyanate
- **GFP:** Green fluorescent protein
- **mCherry:** Monarch red fluorescent protein

*Fig. 3—continued*
this sequence is that it occurs in the same relative position in each of the proteins. In the GAR transformylase sequences and rat 10-HCO-H,PteGlu dehydrogenase, the histidine occurs at positions 106–116, with the exception of the yeast sequence, which is displaced to residue 125. The clostridial sequences start with a glutamate at position 94 or 98, whereas the C. THF synthase sequences start with a glutamine at positions 100–108, relative to the beginning of the 10-HCO-THF synthetase domain. The beginning of the 10-HCO-THF synthetase domain is set at Ile<sup>305</sup> in the human C, THF synthase (31), which is conserved in all the reported sequences (29–32). The strict conservation of proline and serine at positions 2 and 3, respectively, proline at position 6 (except for yeast GAR transformylase), and glycine at position 10 may represent the boundaries of a secondary structural feature that is involved in the binding of 10-HCO-H,PteGlu by enzymes that both synthesize and utilize 10-HCO-H,PteGlu.

The extent of this domain may now be extended. Inglese et al. (35) recently reported that E. coli GAR transformylase can be inactivated by using the affinity label N<sup>3</sup>-(bromoacetyl)-5,8-dideazafolate. It was subsequently shown by Inglese et al. (36) that the dideazafolate label is covalently attached by an ester linkage to Asp<sup>34</sup>. This aspartate residue is conserved in all known GAR transformylase sequences including yeast (Asp<sup>167</sup>). Site-directed mutagenesis of Asp<sup>167</sup> to Asn<sup>167</sup>. This histidine residue is strictly conserved in Dro-
Table I

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Identity*</th>
<th>Extent*</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human cytosol</td>
<td>Thr121-Thr300</td>
<td>39, 40</td>
<td></td>
</tr>
<tr>
<td>Human mitochondria</td>
<td>Gin67-Thr43</td>
<td>41, 42</td>
<td></td>
</tr>
<tr>
<td>Horse cytosol</td>
<td>Leu1-Thr41</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>Horse mitochondria</td>
<td>Gin232-Phe44</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td>Rat mitochondria</td>
<td>Gin243-Phe44</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>Aspergillus nidulans</td>
<td>Pro63-Phe46</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>P. elodea</td>
<td>Thr18-Gly38</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>NAD⁺-dependent aldehyde dehydrogenase</td>
<td>Val106-Thr307</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>Rat tumor-specific</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Percent of residues which are exactly matched.

** Extent refers to the region of identity in 10-HCO-H₄PteGlu dehydrogenase.

Fig. 6. Putative 10-HCO-H₄PteGlu-binding site. The sequences were taken from the references indicated. The number next to the first residue of each sequence is the position of that residue in the protein sequence. The exceptions to this are underlined and represent the position of the residue within the 10-HCO-THF synthetase (10-FTHFS) domain of the Cl-THF synthase (Cl-SYNT). The reference point for the start of the 10-HCO-THF synthetase domain is listed for the human Cl-THF synthase (31). This residue is conserved in all four Cl-THF synthase sequences listed. Identical residues are indicated by vertical lines, whereas conservative changes are indicated by colons. GART, glycaminamide ribonucleotide transformylase; 10-FTHFDH, 10-HCO-H₄PteGlu dehydrogenase.

Fig. 7. Alignment of putative 10-HCO-H₄PteGlu-binding site with active-site Asp of E. coli GAR transformylase. The GAR transformylase (GART) (22-26) and rat 10-HCO-H₄PteGlu dehydrogenase (10-FTHFDH) sequences were aligned by GENALIGN (IntelliGenetics Inc.; see “Experimental Procedures”) using the Needleman-Wunsch algorithm (18). The Cl-THF synthase (Cl-SYNT) sequences (29, 31) are included to show the putative 10-FTHFDH, 10-HCO-H₄PteGlu dehydrogenase.

Fig. 8. Alignment between 10-HCO-H₄PteGlu dehydrogenase and human liver cytosol aldehyde dehydrogenase. The sequences were aligned by FASTA (17). Identical residues are identified by vertical lines, whereas conservative changes are identified by colons. The conserved active-site Cys residue is in boldface type. Arg-g', identified by vertical lines, whereas conservative changes are indicated by colons. The conserved active-site Cys residue is in boldface type. Arg-g', identified by colons. The conserved active-site Cys residue is in boldface type. Arg-g', identified by colons. The conserved active-site Cys residue is in boldface type. Arg-g', identified by colons.

Fig. 9. Alignment of rat liver 10-HCO-H₄PteGlu dehydrogenase, human cytosol aldehyde dehydrogenase, and yeast Δ-1-pyrroline-5-carboxylate dehydrogenase. The alignment was achieved with GENALIGN (IntelliGenetics Inc.) using the Needleman-Wunsch algorithm (18). Residues which are identical are indicated by vertical lines. Spaces (indicated by dashes) were introduced to improve the alignment. 10-FTHFDH, 10-HCO-H₄PteGlu dehydrogenase; CYTO ALDH, human cytosol aldehyde dehydrogenase (39); Δ1PSC-DH, yeast Δ-1-pyrroline-5-carboxylate dehydrogenase (51).
show that out of 484 amino acid residues, 226 were identical (47% identity), whereas 165 residues (34%) were judged to be conservative changes, which would give an 81% overall identity. A lower identity (28%) was shown to an NAD$^+$-dependent aldehyde dehydrogenase from Pseudomonas oleovorans (47) and to a rat tumor-specific NADP$^+$-dependent aldehyde dehydrogenase (25% identity) (48). The extent of identity is essentially along the entire length of the aldehyde dehydrogenase sequences.

The degree of conservation between 10-HCO-H$_4$PteGlu dehydrogenase and NAD$^+$-dependent aldehyde dehydrogenase prompted us to check for this activity. Using the method of Lindahl and Evces (19), which is designed for the assay of rat liver aldehyde dehydrogenase, no activity was found when NAD$^+$ was used. However, with added NADP$^+$, activity was seen for formaldehyde (0.008 $\mu$mol of NADPH min$^{-1}$ mg$^{-1}$), acetaldehyde (0.008 $\mu$mol of NADPH min$^{-1}$ mg$^{-1}$), propionaldehyde (0.049 $\mu$mol of NADPH min$^{-1}$ mg$^{-1}$), and benzaldehyde (0.005 $\mu$mol of NADPH min$^{-1}$ mg$^{-1}$). Rat liver aldehyde dehydrogenase has not been purified, so a direct comparison between these rates is not possible. However, the specific activity of the pure human cytosol NAD$^+$-dependent aldehyde dehydrogenase assayed under similar conditions (1.3 $\mu$mol of NADPH min$^{-1}$ mg$^{-1}$) (49) is ~25-fold higher than the NADP$^+$-dependent aldehyde dehydrogenase activity of 10-HCO-H$_4$PteGlu dehydrogenase with propionaldehyde as the substrate. An oriental variant of human mitochondrial aldehyde dehydrogenase, where glutamate 502, 13 residues from H$_4$PteGlu dehydrogenase is arginine 894, a basic residue, substrate. An oriental variant of human mitochondrial aldehyde dehydrogenase and NAD$^+$-dependent aldehyde dehydrogenase (0.005 $\mu$mol of NADPH min$^{-1}$ mg$^{-1}$), Rat liver aldehyde dehydrogenase (0.049 $\mu$mol of NADPH min$^{-1}$ mg$^{-1}$), and benzaldehyde (0.005 $\mu$mol of NADPH min$^{-1}$ mg$^{-1}$). The effect of disulfide bond 4 on 10-HCO-H$_4$PteGlu dehydrogenase has not been determined.

The active site of human liver cytosol aldehyde dehydrogenase contains a disulfiram-sensitive thiol that has been identified as cysteine 302 (39). This cysteine residue is conserved in all the enzymes listed in Table I and in 10-HCO-H$_4$PteGlu dehydrogenase (Cys$^{302}$) (Fig. 8). The effect of disulfiram on 10-HCO-H$_4$PteGlu dehydrogenase has not been determined.

Another enzyme that shows high identity to a significant region of the COOH terminus of 10-HCO-H$_4$PteGlu dehydrogenase is the yeast mitochondrial enzyme D-1-pyrroline-5-carboxylate dehydrogenase (EC 1.5.1.12), which converts D-1-pyrroline-5-carboxylate to glutamate in an NAD$^+$-dependent pyrrole ring-opening oxidation (51). Alignments of 10-HCO-H$_4$PteGlu dehydrogenase, human cytosol aldehyde dehydrogenase, and D-1-pyrroline-5-carboxylate dehydrogenase reveal a high level of identity (29%) among all three sequences in a region close to the putative active-site cysteine 302 of aldehyde dehydrogenase (Fig. 9).

The revelation that 10-HCO-H$_4$PteGlu dehydrogenase has three putative domains suggests that it may have evolved from a triple-gene fusion. The presence of the GAR transformylase domain and the aldehyde dehydrogenase domain now allows us to specifically inhibit or inactivate these domains and to investigate the folate-binding region(s), the mechanism of action, and the physiological role of 10-HCO-H$_4$PteGlu dehydrogenase.

Acknowledgments—We wish to thank William T. Briggs for excellent technical assistance. We also would like to thank T. Porter for performing the automated peptide sequence analysis and Dr. Marion L. Dodson for assistance in the implementation of the FASTA and IntelliGenetics software.
Primary Structure of Rat 10-HCO-H₄PteGlu Dehydrogenase