Nucleotide Sequence of a cDNA for Branched Chain Acyltransferase with Analysis of the Deduced Protein Structure*

Kimberly B. Hummel, Stuart Litwer, Andrew P. Bradford, Alastair Aitken, Dean J. Danner, and Stephen J. Yeaman

From the Division of Medical Genetics, Department of Pediatrics, Emory University, Atlanta, Georgia 30322, the Department of Biochemistry, University of Newcastle upon Tyne, Newcastle upon Tyne NE1 7RU, United Kingdom, and the National Institute for Medical Research, Mill Hill, London NW7 1AA, United Kingdom

Nucleotide sequence was determined for a 1.6-kilobase human cDNA putative for the branched chain acyltransferase protein of the branched chain α-ketoacid dehydrogenase complex. Translation of the sequence reveals an open reading frame encoding a 315-amino acid protein of molecular weight 35,759 followed by 560 bases of 3' untranslated sequence. Three repeats of the polyadenylation signal hexamerATTAAA are present prior to the polyadenylate tail. Within the open reading frame is a 10-amino acid fragment which matches exactly the amino acid sequence around the lipoyl-lysine residue in bovine kidney branched chain acyltransferase, thus confirming the identity of the cDNA. Analysis of the deduced protein structure for the human branched chain acyltransferase revealed an organization into domains similar to that reported for the acyltransferase proteins of the pyruvate and α-ketoglutarate dehydrogenase complexes. This similarity in organization suggests that a more detailed analysis of the proteins will be required to explain the individual substrate and multi-enzyme complex specificity shown by these acyltransferases.

Branch chain acyltransferase (BCKD-E2) forms the core unit of the mitochondrial multienzyme complex, branched chain α-ketoacid dehydrogenase (BCKD) (1-3). BCKD-E2 is nuclear coded, and the mature form of the protein migrates with an apparent Mr value of 52,000 (4). A covalently bound lipoate is positioned near the N terminus of the protein (2, 5). Although the catalytic properties of BCKD-E2 are well characterized, details of the structural properties are limited. Amino acid sequence for the protein has not been determined since it has been difficult to purify BCKD-E2 free from the associated subunits. To circumvent this problem, amino acid sequence was deduced from the nucleic acid sequence of a cDNA encoding BCKD-E2.

We have previously reported the isolation of a human liver cDNA clone which, based on antibody selection, putatively coded for the human branched chain acyltransferase (6). Here we report the complete nucleotide sequence for this cDNA with the deduced amino acid sequence. We confirm the identity of this cDNA by matching a portion of the amino acid sequence to that of a peptide fragment isolated from bovine kidney BCKD-E2 (7). The entire deduced protein structure for human BCKD-E2 is then analyzed by comparison to other acyltransferase proteins.

MATERIALS AND METHODS

The 1.6-kilobase (kb) cDNA, putatively BCKD-E2, was cloned into M13 mp18 and mp19. Nucleic acid sequencing was done according to protocols given in Ref. 22 using [α-35S]ATP as the labeled nucleotide. Oligonucleotide primers (17 bases) synthesized in the Microchemistry Center, Emory University, were used to minimize subcloning. Fragments for sequence analysis were also produced and subcloned using a Cyclone Kit from International Biotechnologies, Inc. according to their instructions. Genepro software programs (Riverside Scientific Enterprises) were used to translate nucleotide sequences and for comparison analyses.

BCKD was acylated using [U-14C]-α-ketoisocaproate in the presence of N-ethylmaleimide. Peptic fragments were isolated and purified by high voltage electrophoresis and high pressure liquid chromatography. Amino acid sequence of these fragments was determined with an Applied Biosystems 477 pulsed liquid Sequencer as previously described (8).

RESULTS

A restriction enzyme cleavage map for the 1.6-kb cDNA is shown in Fig. 1A. Based on the strategy shown in Fig. 1B, the nucleotide sequence for this cDNA was established. Predicted enzyme cleavage sites were confirmed by sequence analysis. The complete nucleic acid sequence and deduced amino acid sequence are shown in Fig. 2. The encoded protein has a molecular weight of 35,759 based on amino acid content in the longest open reading frame.

Using acylated lipoate as a marker, a peptide fragment containing this lipoate binding site was isolated from bovine kidney BCKD. The determined amino acid sequence for this 10-residue peptide is Glu-Val-Gln-Ser-Asp-Lys-Ala-Ser-Val-Thr. As can be seen in Fig. 2, amino acid residues 95-104 for human BCKD-E2 exactly match this bovine amino acid sequence. Furthermore, the amino acid sequence flanking the lipoate binding site in human liver BCKD-E2 was compared to several other proteins (Table I). As is observed, both the critical lysine and a glycine, 11 residues C-terminal to the lysine, are conserved in all proteins.

Amino acid sequences deduced from cDNAs for the acyl-
transf erase proteins from Escherichia coli, pyruvate dehydrogenase complex acetyltransferase (PDH-E2) (EC 2.3.1.12), and α-ketoglutarate dehydrogenase succinyltransferase (OGDH-E2) (EC 2.3.1.61) have been reported and structural domains defined (9). Recently the cDNA for rat liver PDH-E2 was published, and we include this data in our comparison (10). Using the lipoate-lysine to align these proteins with human BCKD-E2 reveals a conservation of structural and functional domains (Fig. 3). One domain in particular among the four proteins exhibits a highly conserved 32-amino acid fragment (Fig. 4).

The first 25 amino acids of human BCKD-E2 contain no acidic residues and 3 arginine residues separated by 3-8 amino acids (Figs. 2 and 3). Such unusual properties are hallmarks of leader sequences which direct nuclear coded proteins to the mitochondria (11). The other acyltransferase proteins we compared with human BCKD-E2 lack this region.

**DISCUSSION**

Antibodies specific for the BCKD proteins were used to isolate the 1.6-kb cDNA clone from human liver (6). The longest open reading frame within this cDNA begins at the first methionine, located 12 bases after the 5'- EcoRI cloning site, and translates into a 315-amino acid protein of 35,759 daltons (Fig. 2). Within the 3' untranslated end there are three repeats of the polyadenylation signal hexamer, ATTAAA, followed by a poly(A) tail (12). A 10-amino acid sequence (95-104) within the open reading frame exactly matches the lipoate-lysine fragment of bovine kidney BCKD-E2 and confirms that this cDNA encodes the human liver BCKD-E2.

BCKD-E2 isolated from mitochondria migrates in sodium dodecyl sulfate-polyacrylamide gels (SDS-PAGE) with an M, of 52,000. The disparity in M, value and the ~36-kDa size, predicted from amino acid content of the translated cDNA, must be explained. Size discrepancies of this type are also reported for both PDH-E2 and OGDH-E2 acyltransferase proteins from two analogous mitochondrial multienzyme complexes (3). Abrerrantly high M, values in these proteins are attributed to an extended region of helical structure due to proline-alanine-rich domains (13).

Since the three multienzyme complexes (BCKD, PDH, and OGDH) function in an analogous manner with the acyltransferase as the core component (3), we asked if the size difference for human BCKD-E2 could also be explained by those properties of PDH-E2 and OGDH-E2 which lead to disparate molecular weight values. To approach this question, we aligned the structures of these three proteins using the lipoate-lysine as the focal point (Fig. 3). Escherichia coli PDH-E2 contains three lipoate-lysine domains. Studies using recombinant proteins showed that a constructed protein, containing only the third lipoate-lysine domain, conferred full activity to a reconstituted PDH complex (13). Hence this third lipoate-lysine site was used to align with the other proteins. Four structural domains are defined for OGDH-E2 and PDH-E2.

**TABLE 1**

| Source        | Enzyme | Sequence |
|--------------|--------|----------|
| Human liver  | BCKD-E2| EYVQSDKASVTLSLDVYDG*  |
| Bovine kidney| BCKD-E2| EYVQSDKASVT |
| Bovine kidney| OGDH-E2| IEDRTKSIVQPSPEG*    |
| E. coli      | OGDH-E2| IEDRTKVVLPSADGG*    |
| Rat liver    | PDH-E2 | IEDRTKSIVQPSADGG*   |
| Rat heart    | PDH-E2 | IEDRTKSIVQPSADGG*   |
| Chicken liver| Glycine cleavage H protein | IEDRTKSIVQPSADGG* |

*Sequence determined in this work
K, lipoate-lysine; G, conserved glycine.

**FIG. 2.** Nucleotide sequence and deduced amino acid sequence for 1.6-kb cDNA encoding human BCKD-E2. Circles indicate arginine residues in the putative leader sequence. The lipoate-lysine binding domain is underlined, and the E3 binding domain is contained in brackets. Polyadenylation nucleotide signal hexamers are overlaid.

**FIG. 3.** Alignment of BCKD-E2 from chicken liver (EVGATCAAMCCTCCGGTATTCACAGGCAAMGACAAMCAGAACCCATAMAGGCTTTCAM), human liver (EVQSDIJASVTSLDVYDG), and bovine kidney (EVQSDIJASVT) lipoate-lysine binding domains. The E3 binding domain is underlined, and the E3 binding domain is contained in brackets.
The fourth domain, termed the catalytic domain, is located in the C-terminal portion of the proteins. This domain confers substrate specificity to the protein and may contain the binding site for the E1 subunit of the complex. Specificity of the different acyltransferases for the homologous E1 subunit is high. This specificity was suggested from experiments with isolated subunits from the BCKD and PDH complexes (17). When subunits were mixed to give chimeric complexes, little complex activity was noted (17). This low activity could be due to the substrate specificity of the acyltransferases or result from improper binding between the mixed subunits. Since antibodies specific for BCKD-E2 do not recognize either OGDH-E2 or PDH-E2 (18), the antigenic specificity may also reside in this C-terminal end. Again additional studies are necessary to confirm or disprove this hypothesis. Since multiple E2 subunits interact in the formation of the active complex (3), binding sites for E2 to itself may also be present in this domain. Direct answers to some of our initial questions regarding human BCKD-E2 are not possible from these studies. However, these data confirm that the acyltransferase proteins are highly conserved in structure and function. These observations provide a basis for further analysis of BCKD complex assembly and subunit interaction. As clones for the other subunits of BCKD become available (19), in vitro synthesis of these proteins can be used to study these processes.

The first 25 amino acids of BCKD-E2 are unique due to the absence of acidic residues and the enrichment in arginine. These characteristics define other known leader sequences which direct nuclear encoded proteins to the mitochondria (11). The N-terminal amino acid of mature BCKD-E2 is not known so the exact length of the leader sequence cannot be determined from this data. (Attempts at end terminal amino acid sequence analysis suggest N-terminal blockage.) In comparison with the other acyltransferases, only the rat liver PDH-E2 should contain a leader sequence since the other two proteins are of prokaryotic origin. Since data on the cDNA for the rat liver enzyme used in this report are derived from a literature report (10), we do not know if the clone encodes a full-length protein. Experiments are now being done with the in vitro synthesized protein from this human BCKD-E2 cDNA to determine if this 5'-end does indeed encode a leader sequence.

Similarities between the acyltransferases suggest a common ancestor gene for these proteins. The mitochondrial location of these proteins in mammals is consistent with a prokaryote parent gene. BCKD-E2 does not occur normally in E. coli; however, a prokaryote form has been reported in Pseudomonas (20). Until the cDNA and protein sequences are known for these prokaryote BCKD components, a complete comparison and evolutionary prediction would be premature.

S. J. Yeaman, unpublished observations.
REFERENCES
1. Pettit, F. H., Yeaman, S. J., and Reed, L. J. (1978) Proc. Natl. Acad. Sci. U. S. A. 75, 4881–4885
2. Heffelfinger, S. C., Sewell, E. T., and Danner, D. J. (1983) Biochemistry 22, 5519–5522
3. Yeaman, S. J. (1986) Trends Biochem. Sci. 11, 293–296
4. Danner, D. J., Armstrong, N., Heffelfinger, S. C., Sewell, E. T., Priest, J. H., and Elsas, L. J. (1985) J. Clin. Invest. 75, 858–860
5. Hu, C-W. C., Griffin, T. A., Lau, K. S., Cox, R. P., and Chuang, D. T. (1986) J. Biol. Chem. 261, 343–349
6. Litwer, S., and Danner, D. J. (1985) Biochem. Biophys. Res. Commun. 131, 961–967
7. Lawson, R., Cook, K. G., and Yeaman, S. J. (1983) FEBS Lett. 157, 54–58
8. Bradford, A. P., Howell, S., Aitken, A., James, L. A., and Yeaman, S. J. (1987) Biochem. J. 245, 919–922
9. Packman, L. C., and Perham, R. N. (1986) FEBS Lett. 206, 193–198
10. Gershwin, M. E., Mackay, I. R., Sturgess, A., and Coppel, R. L. (1987) J. Immunol. 138, 3525–3531
11. Douglas, M. G., McCammon, M. T., and Vassarotti, A. (1986) Microbiol. Rev. 50, 166–178
12. Birnstiel, M. L., Bisslinger, M., and Strub, K. (1985) Cell 41, 349–359
13. Guest, J. R., Lewis, H. M., Graham, L. D., Packman, L. C., and Perham, R. N. (1985) J. Mol. Biol. 185, 743–754
14. Spencer, M. E., Darlison, M. G., Stephens, P. E., Duckenfield, I. K., and Guest, J. R. (1984) Eur. J. Biochem. 141, 361–374
15. Packman, L. C., and Perham, R. N. (1987) Biochem. J. 242, 531–538
16. Deleted in proof
17. Cook, K. G., Bradford, A. P., and Yeaman, S. J. (1985) Biochem. J. 225, 731–735
18. Heffelfinger, S. C., Sewell, E. T., and Danner, D. J. (1983) Biochem. J. 213, 339–344
19. Zhang, B., Kuntz, M. J., Goodwin, G. W., Harris, R. A., and Crabb, D. W. (1987) J. Biol. Chem. 262, 15220–15224
20. McCully, V., Burns, G., and Sokatch, J. R. (1986) Biochem. J. 233, 737–743
21. Bradford, A. P., Aitken, A., Beg, F., Cook, K. G., and Yeaman, S. J. (1987) FEBS Lett. 222, 211–214
22. M13 Cloning and Sequencing Handbook (1985) pp. 27–35, American Corp., Arlington Heights, IL