The Primary Structure of the Subunits of Carbon Monoxide Dehydrogenase/Acetyl-CoA Synthase from *Clostridium thermoaceticum*

(Received for publication, November 2, 1991)

Thomas A. Morton‡, Jennifer A. Runquist§§, Stephen W. Ragsdale§§§, T. Shanmugasundaram****, Harland G. Wood*** ‡‡†, and Lars G. Lundahl‡‡‡ §§§

*From the ‡Department of Biochemistry and Center for Biological Resource Recovery, University of Georgia, Athens, Georgia 30602, the §Department of Chemistry, University of Wisconsin, Milwaukee, Wisconsin 53201, and the ****Department of Biochemistry, Case Western Reserve University, Cleveland, Ohio 44106*

CO dehydrogenase/acetyl-coenzyme A synthase (CODH) is the central enzyme in the pathway of acetyl-coenzyme A biosynthesis in *Clostridium thermoacetica*. It catalyzes the interconversion of CO and CO$_2$ and the synthesis of acetyl-coenzyme A from the methylyated corrinoid/iron sulfur protein, CO$_2$ and coenzyme A. It is a nickel-iron-sulfur protein and contains two subunits in the form $\alpha\beta$. Reported here is the cloning and sequencing of the genes for both subunits of CODH. The gene for the $\alpha$ subunit codes for a protein with 729 amino acids and a molecular weight of 81,730, and the $\beta$ gene for a protein with 674 amino acids and a molecular weight of 72,928. The $\alpha$ subunit follows the $\beta$ subunit by 23 bases and the genes for both subunits are preceded by a sequence which is similar to the Shine-Dalgarno sequence of *Escherichia coli*. No significant amino acid sequence homology has been found to any known sequence. Labeling CODH with 2,4-dinitrophenylsulfenyl chloride and isolating labeled peptide fragments demonstrated that a tryptophan, residue 418 of the $\alpha$ subunit, is protected by coenzyme A and thus may be considered a potential part of the coenzyme A site.

There is a large group of anaerobic bacteria that grow autotrophically with CO or CO$_2$ and H$_2$ as the source of carbon and energy. They produce acetate and use the acetate as an anabolic source of carbon, and play a significant role in the carbon cycle (1–3). The acetate is produced by the acetyl-CoA pathway of CO$_2$ fixation. The methyl group of acetyl-CoA is formed from CO$_2$ by a series of reactions involving formate dehydrogenase and tetrahydrololate dependent enzymes. It is then transferred to a cobamide bound to a protein designated the corrinoid iron-sulfur protein (C/Fe-SP) (4). The 3.6-kDa subunit of acetyl-CoA is formed by reduction of CO$_2$ to CO. This is catalyzed by CO dehydrogenase/acetyl-CoA synthase (CODH) which also catalyzes the final step in acetyl-CoA biosynthesis by condensing the methyl, CO, and CO$_2$ to yield acetyl-CoA (5).

CODH from *Clostridium thermoaceticum* is a hexamer with a molecular weight of 440,000 consisting of two subunits in the form $\alpha\beta$. It is one of the few naturally occurring nickel proteins, each $\alpha\beta$-dimer containing two nickel, 11–13 iron, 14 inorganic sulfur, and one zinc (6). This central enzyme of the acetyl-CoA pathway, in addition to catalyzing acetyl-CoA synthesis, catalyzes exchange reactions between acetyl-CoA and CO (5), acetyl-CoA and CoA (7, 8, 9), and acetyl-CoA and methyl groups bound to either CODH or C/Fe-SP (10). It must, therefore, have separate binding sites for the methyl, carboxyl, and CoA groups designated X, Y, and Z sites, respectively. Recent studies indicate the methyl and CO sites may be different coordination sites of a Ni-Fe center (9), and evidence has been presented that tryptophan is at or near the CO site (11). But nothing is known about the binding of at least four different metal centers (12), including the Ni-Fe center and other ligands. The elucidation of the binding of the ligands will require sequence information of the subunits of CODH.

In this investigation, the genes of the $\alpha$ and $\beta$ subunits have been sequenced using clones obtained from a genomic library constructed at the University of Georgia and from one established by Roberts et al. (13) who showed that a gene cluster contained the genes for the $\alpha$ and $\beta$ subunits of CODH, methyltransferase, and the corrinoid (C/Fe-SP) enzymes. In addition, the primary structure derived from the DNA has been confirmed by sequencing peptides derived from the $\alpha$ and $\beta$ subunits, one of which may be a component of the CoA site of CODH. Thus far, combination of the expressed gene products has been confirmed. The abbreviations used are: C/Fe-SP, corrinoid iron-sulfur protein; CODH, CO dehydrogenase/acetyl-CoA synthase; LB, Luria-Bertani; TY, tryptone-yeast extract; SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis; DNPS-Cl, 2,4-dinitrophenylsulfenyl chloride; TFA, trifluoroactic acid; kb, kilobase(s); HPLC, high performance liquid chromatography.
and β subunits has not led to formation of an active CODH, perhaps because the cloned subunits do not contain competent Ni-Fe-sulfur or other Fe-sulfur centers.

EXPERIMENTAL PROCEDURES

RESULTS AND DISCUSSION

Cloning and Mapping the CODH Genes—C. thermoacetum

Genomic DNA was digested with EcoRI and hybridized to an oligonucleotide probe for the β subunit of CODH. A single 3.7-kb band was observed (data not shown). A partial library using EcoRI fragments ranging in size from 2.3 to 4.3 kb was constructed in pBR322 and screened using the CODH β subunit probe. All selected positives contained a 3.7-kb insert which was mapped using three restriction enzymes (see Fig. 1, miniprint).

To test for the presence of multiple copies of the genes, genomic DNA was digested with two restriction enzymes and probed with the EcoRI fragment containing the CODH genes. Only bands with molecular weights expected for the cloned fragment were observed, consistent with the presence of only a single copy of the CODH genes in the C. thermoacetum genome.

DNA Sequencing and Deduced Amino Acid Sequence—For sequencing, the 3.7-kb EcoRI fragment was cloned into M13mp18 in both directions, and subclones were generated by single stranded deletions. Fig. 1 (miniprint) shows the sizes and orientations of the subclones which were sequenced. Because of the difficulty in accurately sizing the single stranded subclones there was considerable duplication in some regions. If the sequence from a subclone was entirely contained within another subclone, it was not included in the figure. The EcoRI fragment contained the complete gene of the β subunit, but only 69% of the α subunit. The remainder of the sequence, from 3680 to 4592, where the gene for the C/Fe-SP begins, was obtained from a 10-kb Sau3A insert in pUC9 (pCt946A, ref. 13). The subclones in this region, also shown in Fig. 1 (miniprint), were generated by double stranded deletions and sequenced. Gaps not covered by deletions were filled using sequence-specific primers.

The DNA sequence from the start of the EcoRI fragment to the first codon of the 55-kDa subunit of the C/Fe-SP is shown in Fig. 2, along with the derived amino acid sequence. As indicated in Fig. 1, all of the sequence shown has been determined in both directions. The gene for the β subunit begins at base 270 and ends at base 2294. The α gene follows the β gene by 23 bases. It begins at base 2315 and ends at base 4504. There is a space of 84 bp between the termination codon for the α subunit gene and the initiation codon for the 55-kDa subunit of the C/Fe-SP. Based on the DNA sequence, the gene for the β subunit encodes a protein with 674 amino acids and an M, of 72,928, and the α gene encodes a protein with 729 amino acids and an M, of 81,730. These values are similar to the molecular weights of 71,000 and 78,000 determined by SDS-PAGE (6). The underlined amino acid sequences were confirmed by N-terminal amino acid sequencing of both subunits and tryptic fragments, obtained directly from CODH (light underlining of Fig. 2; the residues Cys-11 of the β subunit and His-51 and Thr-58 of the α subunit were not underlined since they were not conclusively determined during amino acid sequencing). In addition, tryptic fragments obtained after the enzyme had been reacted with DNPS-Cl to identify peptides containing tryptophan residues were sequenced (heavy underlining of Fig. 2 which are numbered as in Fig. 4 of miniprint). In addition, the amino acid compositions for the two subunits predicted by the DNA sequence were found to be in close accord to those determined by acid hydrolysis of the protein and of the isolated subunits (data not shown).

The codon usage of the two CODH genes was used to calculate the coding probability of each of the three reading frames of the sequence by the method of Staden and McLachlan (14). In support of the accuracy of the DNA sequence, the highest coding probability was found in the open reading frames used to deduce the protein sequence.

The sequence AGGAGG (underlined in Fig. 2) was found 8 bases in front of the genes for both subunits of CODH. This is similar to the consensus E. coli ribosome binding site (15) and presumably has the same function in C. thermoacetum. Also shown in Fig. 2 is a similar sequence (AGGAGT) preceding the gene for the 55-kDa subunit of the C/Fe-SP.

Protein Sequence Analysis—The sequences of the two subunits of CODH have been compared to a translation of GenBank (release 61.0) using the sequence comparison program FASTDB (Intelligenetics Inc.). No significant homology was found to any known sequence by this method. Recently, Eggen et al. (16) have determined the amino acid sequence of a CODH from Methanothermus soehngenii which consists of two large subunits of 79.4 kDa and two small subunits of 19 kDa. A comparison of the sequences reveals that no extensive regions of homology exist between the enzymes. However, residues 495–500, VVATGC, and 548–551, GSCV, of the β subunit of the C. thermoacetum enzyme are identical with residues 455–550, 583–586, respectively, of the large subunit of M. soehngenii enzyme.

Spectrometric studies of CODH from C. thermoacetum (12, 17–21) have demonstrated the presence of at least two [Fe3S4] like clusters, one of which apparently interacts with nickel in the presence of CO. In addition, the enzyme may have a 2Fe complex and a FeS4 species. A total of 31 cysteines are present in the CODH, but arrangements of these as found in iron-sulfur proteins, including ferredoxins, are not evident in the primary structure (22). However, the α and β subunits each contain two cysteine-rich regions that include residues 506, 509, 518, 528, and 583, 595, 597, 608 of the α subunit and residues 59, 67, 68, 71, 76, 90, and 316, 317, 342, 350, 355, 366 of the β subunit. These are potential sites for binding metal clusters. Of interest are the two pairs of vicinal cysteines (residues 67, 68 and 316, 317 of the β subunit). Poston et al. (23) noticed that acetate synthesis in cell-free extracts of C. thermoacetum from methyl-B12 and pyruvate is inhibited by arsenite and cadmium chloride, reagents which react with vicinal sulfhydryls. Locating the amino acid residues responsible for ligating the Fe-S centers will require further study.

Evidence has been presented that the CoA-binding site is near the Ni-Fe-C center formed when CO reacts with CODH (17, 24) and involves tryptophan and arginine residues (11, 24, 25). To identify this site, CODH was labeled with DNPS-Cl, a reagent specific for tryptophan residues. In a time course study of the reaction of CODH with DNPS-Cl, it was found that 8.2 mol of DNPS-Cl were incorporated per mol of dimer during a 6-h period. The exchange activity of CO with the carbonyl group of acetyl-CoA was completely abolished (Fig. 3, miniprint) and no further incorporation of DNPS-Cl was observed up to a period of 24 h. The DNPS-Cl modification of CODH was also carried out in the presence of 100 µM CoA (Fig. 3, miniprint). Under this condition, only 6.8 mol of CODH has been modified.

* Portions of this paper ("Experimental Procedures" and Figs. 1, 3, and 4) are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are included in the microfilm edition of the Journal that is available from Waverly Press.
DNPS-CI was incorporated per mol of dimer after 6 h without loss of enzymatic activity. This experiment suggests that at least one tryptophan is protected by CoA and may be at the CoA site. The kinetics of the loss of activity upon reaction with DNPS-CI appears to be zero order (Fig. 3, miniprint), whereas normally it might be expected to be first order. This indicates a more complicated reaction pattern than just a reaction with a specific tryptophan and may involve a conformational change. It should be noted that in reacting CODH with N-bromosuccinimide, another tryptophan reagent, the reaction was exponential and that the loss of activity was prevented by CoA (11).

The CODH modified with DNPS-CI was digested with trypsin and the resulting peptides separated using HPLC. The elution profile of the peptides is shown in Fig. 4 (miniprint). Seven major peptides, along with a few minor peptides, were labeled by DNPS-CI. The major peptides, together accounting for more than 60% of DNPS-CI incorporated into CODH were repurified using reversed-phase HPLC.

The sequences were determined of all of the peptides that were labeled with DNPS-CI in the experiment of Fig. 4A (peptides 1–7). The N termini of the peptides of Fig. 4B were determined. All N termini of the peptides of Fig. 4B corresponded to those of the tryptophan-labeled peptides of Fig. 4A except peptide 4 was missing which has an IHDFI amino terminus. These facts show that peptide 4 of Fig. 4A was protected by CoA and thus did not become labeled with DNPS-CI in the experiment of Fig. 4B. This peptide, identified with residues 407–423 in the α subunit of CODH, contains the tryptophan residue 418 which apparently is protected by CoA and may be a portion of the CoA binding site. Since the enzyme is not expressed in the active form, mutation of Trp-418 could not be carried out in the present study to determine if this mutation causes inactivation of the enzyme. Other portions of the enzyme may also be involved in the CoA site that do not contain tryptophan.

All of the tryptic fragments were found in the CODH sequence with the exception of peptide 6, which did not correspond to the sequence in any of three reading frames of the CODH genes. This peptide was compared to the Protein Identification Resource (PIR) protein sequence data bank and was found to be very similar to a fragment of the elongation factor TU from E. coli (26) as follows:

Elongation factor TU

| Residues | Sequence |
|----------|----------|
| 334      | VVTYGTTELPEGVMVMPDN1K |

Pep tide 6

7TWDGVLNLPFGVMVMPGWDR
The CODH used in the reaction with DNPS-C1 had a CO dehydrogenase activity of 500 units (µmol·min⁻¹) per mg and CO acetyl-CoA exchange activity of 150 milliunits per mg (Fig. 3) indicating a purity of 90% or higher. This was confirmed by gel electrophoresis which, however, also revealed a contaminant (about 10% protein having an Mₐ of about 50,000). We feel peptide 6 to be derived from this protein. We feel peptide 6 to be derived from this protein.

A search for a binding site was done by comparing the amino acid sequences of the CoA binding site of other enzymes. A consensus sequence for a CoA binding site has been proposed (28) based on the crystal structure of citrate synthase (29) and a comparison of the amino acid sequences of citrate synthase from three different organisms (30). This consensus sequence was not found in either subunit of CODH.

Acknowledgments—We thank Dr. Mary Graves, Warren McComas, and Richard Motyka of Hoffman-LaRoche for providing a number of sequencing primers, F. Carl Haase of Rohm and Haas Company for carrying out some preliminary amino acid sequencing, W. M. Lindahl, P. A., Ragsdale, S. W., and Wood, H. G., and L. G. Ragsdale, S. W., Lindahl, P. A., and Munck, E. for advice in editing the manuscript.

REFERENCES

1. Wood, H. O., and Ljungdahl, L. G. (1991) in Variations in Autotrophic Life (Shively, J. M., and Barton, L. L., eds) pp. 201–260, Academic Press, London
2. Wood, H. G., Ragsdale, S. W., and Pezacka, E. (1986) Trends Biochem Sci. 11, 14–18
3. Fuchs, G. (1988) FEMS Microbiol. Rev. 39, 181–213
4. Ragsdale, S. W., Lindahl, P. A., and Münck, E. (1987) J. Biol. Chem. 262, 14289–14297
5. Ragsdale, S. W., and Wood, H. G. (1988) J. Biol. Chem. 263, 3970–3977
6. Ragsdale, S. W., Clark, J. E., Ljungdahl, L. G., Lundeis, L. L., and Drake, H. L. (1983) J. Biol. Chem. 258, 2364–2369
7. Pezacka, E., and Wood, H. G. (1986) J. Biol. Chem. 261, 1609–1615
8. Ramer, S. E., Raybuck, S. A., Orme-Johnson, W. H., and Walsh, C. T. (1989) Biochemistry 28, 4675–4680
9. Ljungdahl, L. G., Ragsdale, S. W. (1991) J. Biol. Chem. 266, 3554–3564
10. Lu, W.-P., and Ragsdale, S. W. (1990) J. Biol. Chem. 265, 3124–3133
11. Shannugasundaram, T., Kumar, G. K., and Wood, H. G. (1988) Biochemistry 27, 6489–6503
12. Lindahl, P. A., Ragsdale, S. W., and Münck, E. (1990) J. Biol. Chem. 265, 3880–3888
13. Roberts, D. L., James-Hagstrom, J. E., Garvin, D. K., Gorst, C. M., Runquist, J. A., Baur, J. R., Haase, F. C., and Ragsdale, S. W. (1989) Proc. Natl. Acad. Sci. U. S. A. 86, 32–36
14. Staden, R., and McLachlen, A. D. (1982) Nucleic Acids Res. 10, 141–157
15. Neidhardt, F. C. (1987) Escherichia coli and Salmonella typhimurium. Cellular and Molecular Biology, Vol. 2, pp. 1231–1240, American Society of Microbiology, Wash. D. C.
16. Eggens, R. I., Geering, A. C. M., Jetten, M. S. M., and de Vos, W. M. (1991) J. Biol. Chem. 266, 6883–6887
17. Ragsdale, S. W., Wood, H. G., and Antholine, W. E. (1985) Proc. Natl. Acad. Sci. U. S. A. 82, 6811–6814
18. Lindahl, P. A., Münck, E., and Ragsdale, S. W. (1990) J. Biol. Chem. 265, 3873–3879
19. Cramer, S. P., Eldasses, M. K., Pan, W.-H., Morton, T. A., Ragsdale, S. W., DerVartanian, D. V., Ljungdahl, L. G., and Scott, R. A. (1987) Inorg. Chem. 26, 2477–2479
20. Bastain, N. E., Diekert, G., Niederhofer, E. C., Teo, B. K., Walsh, C. T., and Orme-Johnson, W. H. (1988) J. Am. Chem. Soc. 110, 5581–5582
21. Fan, C., Gorst, C. M., Ragsdale, S. W., and Hoffman, B. M. (1991) Biochemistry 30, 431–435
22. Elliott, J., Yang, S.-S., Ljungdahl, L. G., Travis, J., and Reilly, C. F. (1982) Biochemistry 21, 3294–3298
23. Poston, J. M., Kuratomi, K., and Stadtmann, E. R. (1966) J. Biol. Chem. 241, 4209–4216
24. Shannugasundaram, T., Ragsdale, S. W., and Wood, H. G. (1988) Biofactors 1, 147–152
25. Shannugasundaram, T., Kumar, B. C., and Wood, H. G. (1989) Biochemistry 28, 7212–7216
26. Laursen, R. A., L’Italien, J. J., Nagarkatti, S., and Miller, D. L. (1981) J. Biol. Chem. 256, 8102–8109
27. Yang, F., Demina, M., Warren, V., Bharawadarghane, S., and Condeelis, J. (1990) Nature 345, 494–496
28. Buck, D., Spencer, M. E., and Guest, J. R. (1988) Biochemistry 27, 6245–6252
29. Remington, S., Wiegand, G., and Huber, R. (1982) J. Mol. Biol. 158, 111–152
30. Sussia, M., Suda, K., and Schatz, G. (1984) EMBO J. 3, 1773–1781
31. Ljungdahl, L. G., and Andreesen, J. R. (1978) Methods Enzymol. 53, 360–372
32. Markovits, T., Fritsch, E. F., and Szafrbrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
33. Padjak, W. (1973) Clin. Chim. Acta 48, 113–115
34. Sober, H. A. (ed) (1970) CRC Handbook of Biochemistry, p. C-281, Chemical Rubber Co., Cleveland, OH
35. Hirs, C. H. W. (1967) Methods Enzymol. 11, 59–62
36. Scoffen, E., Fontana, A., and Roocchi, R. (1968) Biochemistry 7, 971–979
37. Burton, C. J., and Hartley, B. S. (1970) J. Mol. Biol. 52, 165–178
38. Marmur, J. (1961) J. Mol. Biol. 3, 208–215
39. Wood, W. I., Gitasheir, J., Lasky, L. A., and Lawn, R. M. (1985) Proc. Natl. Acad. Sci. U. S. A. 82, 1585–1588
40. Kushner, S. R. (1978) in Genetic Engineering (Beyer, H. W., and Wicosia, S., eds) pp. 17–23, Elsevier, Amsterdarn
41. Finkel, A. P., and Vogelstein, B. (1983) Annu. Rev. Biochem. 52, 63–13
42. Dale, R. M. K., McClure, B. A., and Houchins, J. P. (1985) Plasmid 13, 31–40
43. Ansonge, W., and Labeit, S. (1984) J. Biochem. Biophys. Methods 10, 237–243
44. Devereux, J., Haeberli, P., and Smithies, O. (1984) Nucleic Acids Res. 12, 387–396
**Sequence of CODH/Acetyl-CoA Synthase from C. thermoaceticum**

Thomas A. Morton, Jocelyn A. Kang, Stephen W. Ragdale, T. Stanislavsky, Harold M. Wood and Lars C. Gjermoe

**SUPPLEMENTAL MATERIAL TO**

**THE PRIMARY STRUCTURE OF THE SUBUNITS OF CARBON MONOXIDE \textit{DEHYDROGENASE}/\textit{ACETYL-COA SYNTHASE FROM \textit{CLOSTRIDIUM THERMOSACETICUM}}**

Fig. 1. Restriction enzyme map, coding regions and sequencing strategy of the genes for CODH.

The large arrows represent the open reading frames encoding the two subunits. The circle and the dotted line of the small arrows indicate the sequence information obtained from each subunit. The restriction enzyme map is shown at the top of the figure as B: HindIII, S: SalI and H: HindII.

Fig. 2. Time course of inactivation of CODH by DNPS-CI in the presence and absence of CoA. CODH (25 mg, 8.8 nmol) with specific activities in the CO hydrogense reaction of 190 10^-3 U/mg and in the exchange reaction between the carbon group of acetyl-CoA and Co2+ of 190 10^-3 mg/mole was incubated with DNPS-CI (40 μg) in the presence (A) and absence (B) of 100 μM Cys in 50 mM sodium buffer, pH 7.0 containing 0.1% DTE. At the indicated times, aliquots were removed for measurement of residual activity between CO2 and acetyl-CoA.

Fig. 3. HPLC profiles of CODH modified by DNPS-CI as described in Fig. 3 in the absence (A) and presence (B) of CoA. The native enzyme was applied to a 4 X 1 cm column and run with a linear gradient generated from 61% TFA/50% H2O down to 0.1% TFA/water. All samples were passed through a column of 61% TFA/50% H2O before HPLC analysis. The HPLC solvent was filtered through a 0.2 μm filter and equilibrated with the mobile phase prior to injection. The flow rate was 1 ml/min.

The arrows and sequences of the major labeled peptides were identified in the DNA cloned sequence (see Fig. 2) as follows: Peptide 1: α-subunit 424-429; Peptide 2: α-subunit 667-672; peptide T: a-subunit 48-63; peptide 4: α-subunit 642-652; peptide 5: a-subunit 177-179; and peptide 7: a-subunit 627-631. The peaks A and B of Fig. 4B did not originate any peaks as evidenced by the peptide analysis suggesting that they are artifacts formed during the reaction between CODH (native) and DNPS-CI.