THE JOURNAL OF BIOLOGICAL CHEMISTRY

Vol. 262, No. 1, Issue of January 5, pp. 97-102, 1987
Printed in U.S.A.

Biosynthetic Thiolase from *Zoogloea ramigera*

III. ISOLATION AND CHARACTERIZATION OF THE STRUCTURAL GENE*

(Received for publication, January 31, 1986)

Oliver P. Peoples‡§, Satoru Masamune¶, Christopher T. Walsh¶, and Anthony J. Sinskey‡

From the Departments of ‡Applied Biological Sciences, Biology, and ¶Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

The gene coding for the biosynthetic thiolase from *Zoogloea ramigera* has been isolated by using antibody screening methods to detect its expression in *Escherichia coli* under the transcriptional control of the *lac* promoter. We have located and determined the nucleotide sequence of the gene. The structural gene is 1173 nucleotides long and codes for a polypeptide of 391 amino acids; 282 nucleotides 5′ and 58 nucleotides 3′ to the coding sequence are also reported. By comparing the amino acid sequence data predicted from the gene with data determined experimentally, we have derived the complete primary structure of thiolase. A catalytically essential cysteine is located at residue 89. The DNA sequence presented has a very high G/C content, 66.2%, typical of the *Z. ramigera* genome. In the coding region, this increases to 68.2% and is strongly reflected in the codon usage which demonstrates a strong preference for G or C in the third position. Examination of the 5′-flanking sequence establishes that the NH_2-terminal methionine is specified by an ATG codon, 7 nucleotides downstream from a Shine-Dalgarno sequence.

Zoogloea ramigera I-16-M is a floc-forming bacteria isolated from activated sludge (1). This organism plays an important role in waste water treatment both by its ability to lower the biological oxygen demand and by inducing the formation of flocculated masses which settle to the bottom as sludge deposits (see Ref. 1). As with many other microorganisms, *Z. ramigera* accumulates poly(3-hydroxybutyrate) as its principal intracellular lipid reserve. Many of the enzymes involved in the metabolism of this polymer have been isolated and characterized (2). The first step in the biosynthesis of poly-β-hydroxybutyrate involves the condensation of two acetyl-CoA moieties by the enzyme thiolase (3). The acetoacetyl-CoA formed is then reduced by an NADP-linked acetoacetyl-CoA reductase and polymerized by poly-β-hydroxybutyrate synthase (4, 5). A second thiolase is thought to be involved in fatty acid degradation.

The focus of our studies is 2-fold. First, we are examining the structure/function relationship of the biosynthetic thiolase-enzyme with regard to C–C bond formation (see accompanying papers, Refs. 39 and 40). Second, we are studying the organization and regulation of the genes involved in the metabolism of poly-3-hydroxybutyrate at the genetic level. Multiple forms of thiolases (acetyl-CoA:acetyl-CoA-*C*-acetyltransferase, EC 2.3.1.9) have been isolated from a number of sources where they are involved in the β-oxidation of fatty acids, ketogenesis, and cholesterogenesis. In particular, the enzymes from rat liver (6), pig heart (7), *Clostridium* (8, 9), and *Escherichia coli* (10) have been well studied. To date, no complete protein sequence data or crystal structure has been reported for any of these enzymes. Furthermore, the only thiolase which has been characterized at the genetic level is the *E. coli* thiolase 1. This enzyme forms part of the fatty acid oxidation multienzyme complex coded by the *fad*AB operon which has recently been cloned (11).

In this paper, we describe the use of antibodies raised against the *Z. ramigera* biosynthetic thiolase to screen a λgt11 expression library for recombinant clones producing this protein in *E. coli*. Using this approach, we have isolated a recombinant clone containing the complete thiolase gene sequence. We have analyzed the genomic organization of the *Z. ramigera* DNA insert in this clone and sequenced the entire thiolase gene-coding region. By using protein sequence data for the NH_2-terminal 25 amino acids and a 10-amino acid tryptic peptide from the active site of the protein (39), we have determined the correct reading frame for translation and hence the complete primary amino acid sequence of the protein.

MATERIALS AND METHODS

Strains—Z. ramigera strain I-16-M (ATCC 19623) was maintained on agar slants (trypticase soy broth 1.5% w/v, bactoagar 1% w/v) and subcultured every 2 weeks. Liquid cultures were grown at 30 °C in zoo broth (casamino acids 0.5% w/v; yeast extract 0.5% w/v; K_2HPO_4 0.2% w/v; KH_2PO_4 0.1% w/v). *E. coli* strains Y1088, Y1089, and Y1090 were as described by Young and Davis (12), BNN97/λgt11 (13).

Enzymes and Radiochemicals—Restriction endonucleases were purchased from IBI and used under the manufacturers conditions. *Eco*RI methylase and DNA polymerase 1 were obtained from New England Biolabs. Calf intestinal alkaline phosphatase was purchased from Boehringer Mannheim. The Klenow fragment of DNA polymerase and [α-^{35}S]dATP and [α-^{32}P]dATP were obtained from Amersham Corp.

Isolation of DNA—Z. ramigera DNA was purified from 200-ml mid-log phase cultures as follows: cells were harvested by centrifugation, washed in 20 mM Tris-HCl, pH 8.2, and resuspended in 10 ml of this solution. The cells were then spheroplasted by the addition of 10 ml of polyethylene glycol 8000 (24% w/v) and 2 ml of lysozyme (25 mg/ml) followed by incubation at 37 °C for 30 min. Spheroplasts were subsequently harvested by centrifugation, resuspended in 5 ml of TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA), 300 μl of SDS[1]

* This work was supported in part by United States Public Health Service Grant 5 R01 GM33039 (to S. M.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) J02636.

§ To whom correspondence should be sent.

[1] The abbreviations used are: SDS, sodium dodecyl sulfate; kb, kilobase pairs; IPTG, isopropyl-1-thio-β-D-galactopyranoside; bp, base pairs.

(10% w/v) added, and the cells lysed by incubating at 55 °C for 10 min. An additional 10 ml of TE was added and the lysate incubated with RNaseA (50 μg/ml) and proteinase K (30 μg/ml) for 1 h at 37 °C. The DNA was then purified by CsCl gradient centrifugation.

Plasmid DNA preparations were carried out using the method of Birnboim and Dolly (14) as described by Ish-Horowicz and Burke (15). λ DNA was prepared by standard procedures as described in Maniatis *et al.* (16).

Preparation of the λgt11 Expression Library—A recombinant library of randon *Z. ramigera* DNA fragments was constructed using the λgt11 expression vector described by Young and Davis (12). *Z. ramigera* DNA was first methylated using *Eco*RI methylase and then partially digested with DNaseI in the presence of Mn^{2+} (17). The ends of the partially digested DNA fragments were repaired using Klenow polymerase, *Eco*RI linkers added, and the DNA digested to completion with an excess of *Eco*RI. Fragments of 2–8 kb were size-selected on a 1.5% agarose gel, purified by electroelution, and ligated with *Eco*RI-digested, phosphatased λgt11 DNA. Ligations were carried out for 18 h at 4 °C using 2 μg of λgt11 DNA and 1 μg of target DNA in a total volume of 10 μl. The complete ligation reactions were packaged *in vitro* using λ extracts prepared from *E. coli* strains BHB2688 and BHB2690 (18) as described by Maniatis *et al.* (16). Packaged phage were plated out and amplified on *E. coli* Y1088.

Preparation of Antibodies—Thiolase antiserum was prepared in New Zealand White female rabbits, using purified thiolase protein, by standard procedures. Antibody titer was estimated by the double-diffusion assay (19). Purified antibody was prepared from the serum by chromatography on protein A agarose (20).

Immunological Screening of the λgt11 Expression Library—The screening of the λgt11 expression library was carried out using rabbit anti-thiolase antibodies and a modification of the procedure described by Young and Davis (12). Approximately 4×10^4 recombinant phage adsorbed to *E. coli* Y1090 were plated out on 15-cm LB-agar plates and incubated at 42 °C for 3 h. The plates were then overlayed with nitrocellulose filters (Schleicher & Schüll, BA85), which had previously been saturated in 10 mM IPTG, and incubated a further 4 h at 37 °C. Filters were removed, washed for 10 min in TBST (50 mM Tris-HCl, pH 7.9, 150 mM NaCl, 0.05% Tween-20), incubated in TBST plus fetal calf serum (20% v/v) for 30 min, and rinsed in TBST. First antibody was bound by incubating the filters in TBST (10 ml) plus purified anti-thiolase antibody (10 μl) for 1 h at room temperature. The filters were subsequently washed in three changes of TBST for 5 min each time. Bound first antibody was detected using a biotin-avidin horseradish peroxidase detection system (Clontech Laboratories) and horseradish peroxidase color development reagent (Bio-Rad).

Western Blotting and Immunodetection—Proteins were separated by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (21) and electrophoretically transferred to nitrocellulose filters (Schleicher & Schüll BA85) essentially as described by Burnette (22). Following transfer, overnight at 30 V, filters were rinsed in TBS (TBST without Tween-20) and incubated in TBS plus 5% bovine serum albumin. Proteins reacting with anti-thiolase serum were detected by incubating the filters in 100 ml of TBS, 1% gelatin containing 2 ml of anti-thiolase serum for 1–2 h. Bound first antibody was subsequently detected using goat anti-rabbit IgG horseradish peroxidase conjugate and horseradish peroxidase color development reagent (Bio-Rad).

DNA Blotting and Hybridization Analysis—DNA blots were prepared using DNA fragments separated on agarose gels by the sandwich blot method (23) based on the technique developed by Southern (24). Filters were hybridized with DNA probes labeled to a high specific activity ($0.1–1 \times 10^8$ cpm/μg of DNA) with [α-^{32}P]dATP, by nick translation (25). Prehybridizations and hybridizations were carried out at 65 °C in sealed polythene bags. The prehybridization/hybridization solution contained 5 × SSCP (1 × SSCP contains 0.15 M NaCl, 0.15 M sodium citrate, 10 mM Na_2HPO_4, 10 mM NaH_2PO_4), 5 × Denhardt's solution, 0.1% (w/v) SDS, 10 mM EDTA, and 100 μg/ml sonicated denatured salmon DNA. Filters were prehybridized for 8–18 h and hybridized for 16–18 h using 10^7 cpm of labeled DNA probe per filter.

Preparation of Lysogens and Protein Analysis—Lysogens of λgt11 recombinant clones were prepared in *E. coli* Y1089 as described by Young and Davis (12). For the preparation and analysis of λ-coded proteins, lysogens were grown at 30 °C in LB (100 ml) until they reached an OD_{600} of 0.5. The prophage was induced by a 20-min incubation of 45 °C, IPTG added to 5 mM and the induced lysogens incubated at 37 °C for 1 h. Cells were harvested, resuspended in assay buffer (0.1 M Tris-HCl, pH 7.5, 5 mM β-mercaptoethanol, 5% (v/v) glycerol), lysed by sonication, cell debris pelleted by centrifugation, and the cell extracts stored at −20 °C. The protein concentrations of bacterial lysates were assayed by the method of Bradford (26), using bovine serum albumin as a standard. Thiolase-enzyme assays were performed as described by Nishimura *et al.* (3).

DNA Sequencing—DNA fragments were cloned into the M13 vectors mp10 and mp11 (27) and sequenced using the dideoxy chain-termination method of Sanger *et al.* (28). The M13 sequencing primer, and other reagents were purchased from Amersham Corp. G/C rich regions were resequenced using dITP in place of dGTP as described by Mills and Kramer (29). Computer-assisted sequence analysis was accomplished using the Staden programs (30).

RESULTS

Construction and Screening of the Z. ramigera/λgt11 Expression Library—A recombinant library of random *Z. ramigera* DNA fragments was constructed in the *E. coli* expression vector λgt11 as described under "Materials and Methods." Approximately 2×10^5 recombinants were obtained, from 1 μg of purified target DNA, and amplified in *E. coli* Y1088. In a series of experiments, a total of 10^5 amplified phage were screened using purified rabbit anti-thiolase antibodies. The initial screening identified 10 potentially positive clones (LDBK1–LDBK10). These clones were picked and rescreened until a pure phage stock of each clone was obtained. At this stage, phage DNA was isolated from clones LDBK1–LDBK10 and analyzed by restriction digestion (data not shown). From the results of these experiments, it became clear that clones LDBK2–10 were identical. Clones LDBK1 and LDBK2 were selected for further study. LDBK1 has an insert composed of 2 *Eco*RI fragments of 3.6 kb and 0.75 kb. LDBK2 has an insert composed of 2 *Eco*RI fragments of 1.65 kb and 1.08 kb.

Analysis of the Proteins Coded by the LDBK1 and LDBK2 Insert Sequences—The proteins coded for by the LDBK1 and LDBK2 insert sequences were analyzed both for thiolase-enzyme activity and for cross-reaction to rabbit anti-thiolase serum. Lysogenic strains of *E. coli* Y1089 containing LDBK1 and LDBK2 phage DNA were prepared as described under Materials and Methods. Several lysogens were obtained for each clone and two of these, Y1089/LDBK1 and Y1089/LDBK2, were used for subsequent studies. A lysogen of the λgt11 vector, BNN97/λgt11 was used as a control. The results of the thiolase-enzyme assays (Fig. 1*a*) clearly indicate that the proteins from Y1089/LDBK1 contain a substantial amount of thiolase activity. Furthermore, the thiolase activity is inducible, greater than 5-fold, by the addition of IPTG. This shows that expression of the thiolase-coding sequences is under the transcriptional control of the *lac* promoter contained in the λgt11 vector. Neither the Y1089/LDBK2 or the BNN97/λgt11 protein lysates demonstrate any significant thiolase-enzyme activity even when induced with IPTG (Fig. 1*a*).

The size of the proteins responsible for the initial positive reaction to rabbit anti-thiolase antibodies was investigated by Western blot experiments. Protein lysates were separated by SDS-polyacrylamide gel electrophoresis, transferred to nitrocellulose filters, and screened with rabbit anti-thiolase serum. The results presented in Fig. 1*b* show the presence of an immunoreactive 40 K_d protein in both the IPTG-induced (Fig. 1*b*, *lane 3*) and non-IPTG-induced lysate (Fig. 1*b*, *lane 4*) of Y1089/LDBK1. These bands are identical in size to the purified *Z. ramigera* thiolase protein run in lanes 2 and 9 as a positive control. No immunoreactive band of this size is present in lanes 5 and 6 which contain the Y1089/LDBK2 proteins or in lanes 7 and 8 which contain the BNN97/λgt11 lysates as a negative control. Furthermore, examination of lanes 5 and 6 for a possible *lac*Z-thiolase fusion protein reveals

(a)

Lysate	Thiolase Activity (units/mg protein)		
	Y1089/LDBK1	Y1089/LDBK2	BNN97/λgt11
+IPTG	23.5	2.6	0.9
-IPTG	4.1	2.1	0.7

(b)

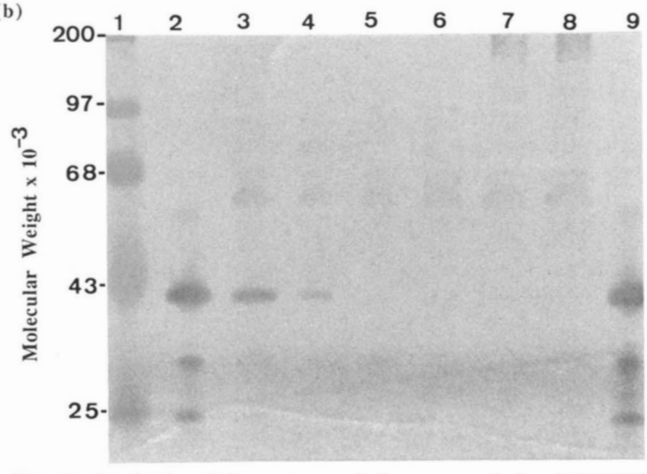

FIG. 1. **Analysis of lysogen proteins.** *a*, analysis of Y1089/LDBK1 and Y1089/LDBK2 *E. coli* lysogen proteins for thiolase activity. BNN97/λgt11 lysogen was used as a control. Results shown are the average of duplicate samples, one unit being defined as the amount of enzyme required to convert 1 μmol of acetoacetyl-CoA to acetyl-CoA in 1 min. Protein lysates were prepared as described under "Materials and Methods." Once the cultures reached an OD_{600} of 0.5 the lysogen was switched on by incubating at 45 °C for 20 min, the culture divided into two equal lots, and IPTG added to one lot at a final concentration of 2 mM. All cultures were then incubated for 1 h at 37 °C. *b*, Western blot analysis of Y1089/LDBK1 and Y1089/LDBK2 protein lysates. The filter was prepared and screened as described under "Materials and Methods." Samples from the same lysates assayed in *a* were used for this analysis. Lanes contained the following samples: *lane 1*, protein size standards: myosin (H-chain), phosphorylase b, bovine serum albumin, ovalbumin, α-chymotrypsinogen; *lanes 2* and *9*, purified *Z. ramigera* thiolase, 0.2 μg/lane; *lanes 3* and *4*, IPTG-induced and noninduced Y1089/LDBK1 proteins, 10 μg/lane; *lanes 5* and *6*, IPTG-induced and noninduced Y1089/LDBK2 proteins, 10 μg/lane; and *lanes 7* and *8*, IPTG-induced and noninduced BNN97/λgt11 proteins, 10 μg/lane.

no bands in the correct size range (>115 K_d). We have therefore concluded that this clone, one of nine identical clones picked from the initial screening, is a "false" positive, and the reason for the initial positive reaction during the plaque-screening remains unclear. However, it should be noted that the occurrence of "false positives" is a common problem with antibody screening methods (see, for example, Voordouw *et al.* (31)). The additional 60-K_d band present in lanes 3–8 is due to cross-reaction of the anti-thiolase serum with an unidentified *E. coli* protein. From the results presented in Fig. 1, it is obvious that clone LDBK1 codes for a functional thiolase of the correct size and therefore contains the complete gene sequence. The levels of both the thiolase enzyme activity (Fig. 1*a*) and immunoreactive protein (Fig. 1*b*, *lanes 3* and *4*) are clearly induced by the addition of IPTG, and hence the expression of the thiolase gene is under the control of the λgt11 *lac* promoter.

Characterization of the DNA Insert of LDBK1—A restriction map of the LDBK1 insert, indicating the orientation with respect to the λgt11 *lac*Z gene is shown in Fig. 2*a*. The large 3.6-kb *Eco*RI fragment, which we have demonstrated to contain the complete thiolase gene region (data not shown) was subcloned into the plasmid vector pUC8 for ease of manipulation. Restriction analysis of one of the subclones obtained, pUCDKB1, confirmed the restriction map of this fragment in LDBK1 (compare Fig. 2*a* and *b*). pUCDBK1 DNA was labeled to a high specific activity with ^{32}P and hybridized to nitrocellulose filters containing *Z. ramigera* chromosomal DNA digested with the same enzymes used to restriction map pUCDBK1 (Fig. 2*b*). Examination of the results, shown in Fig. 3, confirm that the pUCDBK1 insert has not undergone any detectable rearrangements during its cloning and propagation in *E. coli*. For example, the following internal restriction fragments predicted from the restriction map (Fig. 2*b*) are clearly evident as bands of hybridization on the result of the Southern blot shown in Fig. 3: 2.4-kb *Bam*HI-*Eco*RI fragment (*lane 4*); 1.05-kb *Sal*I doublet (*lane 5*); the 790-bp *Sma*I fragment (*lanes 6* and *7*); and the 1.6-kb *Sma*I-*Bam*HI fragment (*lane 7*). The additional hybridization bands in each lane can be accounted for by restriction fragments extending from within the pUCDBK1 insert to the next site in the flanking genomic DNA. There are, however, two additional *Sal*I bands of 5.4 and 4.3 kb (*lane 5*) instead of the expected single band. A possible explanation for this is that the more intense 5.4-kb fragment results from incomplete *Sal*I digestion at the second *Sal*I site upstream from the *Bam*HI site (Fig. 2*b*) to give a 1.05-kb fragment and the 4.3-kb fragment. Subsequent Southern hybridization experiments confirmed that the 5.4-kb genomic fragment hybridizes to both the 1.45-kb *Sal*I-*Eco*RI and one of the 1.05-kb *Sal*I fragments from pUCDBK1, indicating that our interpretation is correct. Repeated attempts to cleave the 5.4-kb genomic fragment with a vast excess of *Sal*I have proved unsuccessful, and we assume that this site is resistant to *Sal*I digestion, possibly as a consequence of DNA methylation. From the result of the Southern hybridization experiment (Fig. 3), we conclude that the cloned pUCDBK1 insert is represented in the *Z. ramigera* genome only once. It is worth noting, however, that longer exposures of these filters reveal additional weak bands of hybridization in each digest (data not shown) indicating the presence of other *Z. ramigera* DNA fragments homologous to the pUCDBK1 insert sequence.

DNA Sequence of the Biosynthetic Thiolase Gene—DNA sequence analysis of the pUCDBK1 insert was carried out using the M13/Sanger dideoxy chain termination method. To locate the gene-coding region, individual DNA sequences were scanned in all six reading frames for homology to the NH_2-terminal amino acid sequence (39). By using this approach, we identified the gene-coding region within the 1.46-kb *Eco*RI-*Sal*I fragment. Fig. 4 (Appendix) illustrates the strategy used to complete the sequence analysis of this DNA fragment. The complete nucleotide sequence of the plus strand of the gene is shown in Fig. 5 (Appendix). 290 bp downstream from the *Eco*RI site lies the start of the thiolase structural gene, assigned by comparing the DNA sequence to the NH_2-terminal amino acid sequence (39). The NH_2-terminal sequence lies in the single long open reading frame which extends from position −89 to the stop codon (TAG) at nucleotide 1174. Beginning with a serine and extending for 25 residues, the experimentally determined NH_2-terminal sequence aligns identically with residues 2 through 26 of the translated DNA sequence. Translation of the DNA sequence was then used to deduce the remaining amino acid sequence

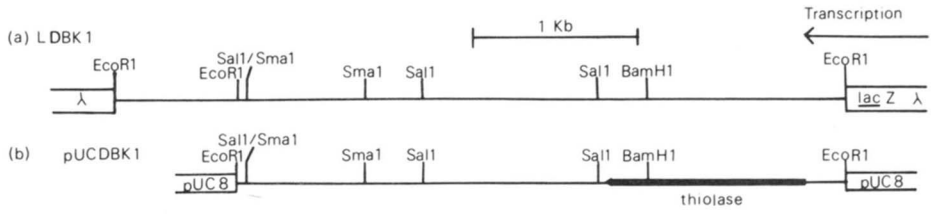

FIG. 2. *a*, a restriction map of the LDBK1 insert showing the orientation with respect to the λgt11 *lac*Z gene. The direction of transcription from the *lac* promoter is indicated by the *arrow*. *b*, a restriction map of the LDBK1 subclone insert sequence of pUCDBK1.

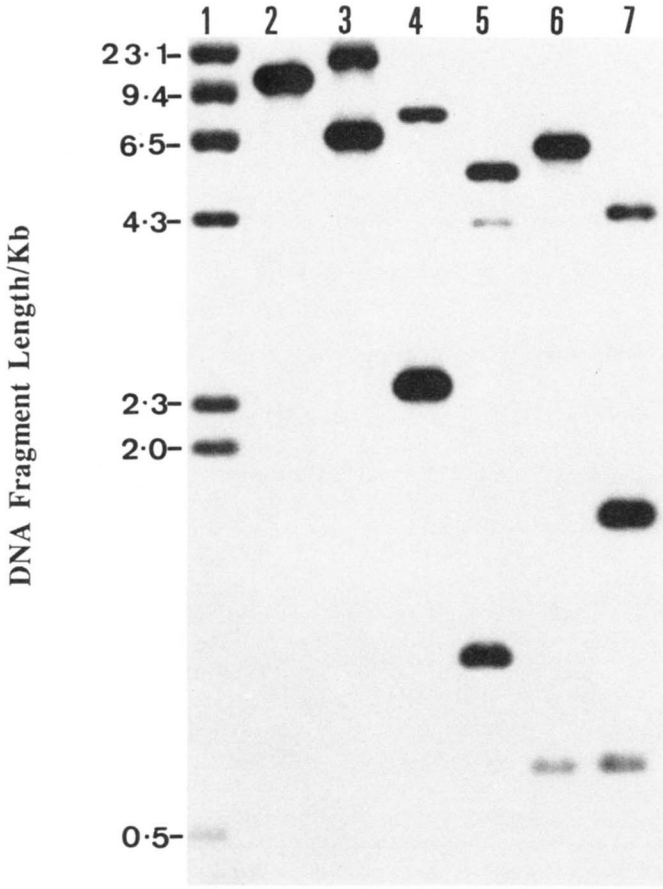

FIG. 3. **Southern blot hybridization analysis of restriction fragments of *Z. ramigera* genomic DNA.** ^{32}P-labeled pUCDBK1 DNA was used as a probe. Lanes contained the following samples: *lane 1*, ^{32}P-labeled λ/*Hin*dIII restriction fragments as size standards; *lanes 2–7*, 1 μg of *Z. ramigera* DNA digested with the following enzymes, *Eco*RI (*lane 2*), *Bam*HI (*lane 3*), *Bam*HI/*Eco*RI (*lane 4*), *Sal*I (*lane 5*), *Sma*I (*lane 6*), and *Sma*I/*Bam*HI (*lane 7*). DNA fragments were transferred from a 1% agarose gel and hybridized as described under "Materials and Methods." A final wash in 2 × SSC, 0.1% (w/v) SDS for 30 min at 65 °C was used.

TABLE I

Amino acid composition of thiolase

Protein data were obtained from Ref. 39.

aa	Amino acid composition from the gene sequence	Amino acid analysis of protein
Cys	5	3
Asp/Asn	20/14	33–34
Thr	18	16–17
Ser	20	20
Glu/Gln	20/14	36
Pro	14	14
Gly	48	50
Ala	64	62–63
Val	27	24–25
Ile	24	20–21
Leu	26	26
Tyr	3	3
Phe	10	11–12
His	6	6
Lys	18	20–21
Arg	18	18–19
Met	17	13–14
Tryp	5	5

TABLE II

Codon usage data for the thiolase structural gene region

Phe	TTT	0	Ser	TCT	0	Tyr	TAT	1	Cys	TGT	0
	TTC	10		TCC	11		TAC	2		TGC	5
Leu	TTA	0		TCA	0	och	TAA	0	umb	TGA	0
	TTG	0		TCG	4	amb	TAG	1	Trp	TGG	5
Leu	CTT	3	Pro	CTT	0	His	CAT	3	Arg	CGT	1
	CTC	18		CCC	3		CAC	3		CGC	17
	CTA	0		CCA	0	Gln	CAA	0		CGA	0
	CTG	5		CCG	11		CAG	14		CGG	0
Ile	ATT	2	Thr	ACT	0	Asn	AAT	2	Ser	AGT	0
	ATC	22		ACC	7		AAC	12		AGC	5
	ATA	0		ACA	0	Lys	AAA	0	Arg	AGA	0
Met	ATG	17		ACG	11		AAG	18		AGG	0
Val	GTT	4	Ala	GCT	1	Asp	GAT	7	Gly	GGT	3
	GTC	14		GCC	41		GAC	13		GGC	44
	GTA	0		GCA	0	Glu	GAA	9		GGA	0
	GTG	9		GCG	22		GAG	11		GGG	1

from residue 27 to 391 (nucleotide 79–1173) shown in Fig. 5. Hence, translation of the DNA sequence from nucleotide 1 to 1174 (in this reading frame) encodes a 391-amino acid polypeptide with a calculated M_r of 40,598. This value is in very good agreement with that of M_r = 42,000 determined by SDS-polyacrylamide gel electrophoresis (Fig. 1; 39). Two additional pieces of evidence confirm that this translation product is the correct amino acid sequence of thiolase. First, a search of the predicted amino acid sequence for the active site peptide (NH_2-Gly-Met-Asn-Gln-Leu-Cys-Gly-Ser-Gly-COOH; 39) located this peptide at residues 84–92. Finally, the predicted amino acid composition from the translation product and that determined experimentally are in excellent agreement (Table I).

As this is the first *Z. ramigera* nucleotide sequence reported, several additional features of the DNA sequence were determined. The G/C content of the 1.46-kb *Eco*RI-*Sal*I fragment is high, 66.2%. When considered separately, the 5′-flanking 290 bp has a G/C content of 57.4% and the structural gene region 68.4%. The codon usage data for the structural gene region shown in Table II reflects this high G/C content. Seven nucleotides upstream from the ATG start codon is a potential ribosome-binding site, 5′-CTGAGGA-3′ identified by homology to the *E. coli* sequence. Additional start codons including two GTGs which can initiate translation in some bacterial genes are located further upstream. Examination of the 5′-flanking region for homology to the "−10" and "−35" *E. coli* promoter elements (38), identified a potential "−35 region" at residues −122 to −116. However, the corresponding "−10 region" 5′-TATAAT-3′ is not obvious around position −95. We note the presence of a poly(T) tract at position −255–

266. With respect to the absence of a *lacZ* thiolase fusion protein, the coding sequences for the two genes are in different reading frames. Furthermore, translation of *lacZ* should extent into the *Z. ramigera* sequence only as far as the in frame TGA stop codon at position −92.

DISCUSSION

The biosynthetic thiolase from *Z. ramigera* I-16-M catalyzes the condensation of two acetyl-CoA units to form acetoacetyl-CoA, the first step of the PHB biosynthetic pathway (3). In the absence of a gene transfer system for this organism, we have used immunological screening methods to identify a clone, LDBK1, which expresses the thiolase gene in *E. coli*. This clone encodes a protein which has both thiolase-enzyme activity (Fig. 1*a*) and comigrates with the native enzyme in Western blot experiments (Fig. 1*b*). Analysis of the genomic organization of the large *Eco*RI fragment from LDBK1, (Fig. 3) confirms that it is a *Z. ramigera* DNA sequence, that it has not undergone any rearrangements, and that it is present as a single copy in the *Z. ramigera* genome. The complete nucleotide sequence of the thiolase gene-coding region of this clone has been determined (Fig. 5). From the nucleotide sequence we have predicted the complete amino acid sequence of the thiolase protein and confirmed this by comparison with the known protein data.

The amino acid sequence confirms that the *Z. ramigera* thiolase, like the biosynthetic thiolase from pig heart mitochondria (32) contains 5 cysteine residues. In both these enzymes, a cysteine is essential for catalytic activity, and we have located the *Z. ramigera* active site cysteine at residue Cys-89. Additional cysteines which may be involved in inter- or intradisulphide bonds are Cys-125, Cys-323, Cys-377, and Cys-387. NH_2-terminal sequence analysis indicated a serine at position 1. It is clear that the only post-translational processing of thiolase is the removal of the *N*-formylmethionine residue, as alternate start codons, ATG or GTG, are either out of frame or have an in-frame stop codon before the start of the structural gene.

The value of 66.2% for the G/C content of the complete sequence shown in Fig. 5 is in good agreement with that calculated for the *Z. ramigera* genome (64.5%; 33). A preference for either G or C in the wobble position for codons is fairly typical of bacteria with a high G/C content (*e.g. Streptomyces*, 34; and *Pseudomonas*, 35). Similarly, for the thiolase gene 91% of the wobble bases are either a G or C. A particularly interesting feature of the remaining codons is that only 2.3% have an A in the third position, and all of these codons are for glutamate. This may be a reflection of the level of specific tRNA molecules in *Z. ramigera*. A correlation between codon usage and the level of specific tRNAs has previously been described for *E. coli* and proposed as a translation regulation mechanism (36, 37).

The 5′-flanking DNA of the thiolase gene is also relatively high in G/C content, 57.4%. A potential Shine-Dalgarno sequence located at position −7 to −13 is indicative that the RNA start site and promoter regions could be located further upstream. Although a region homologous to the *E. coli* −35 has been identified, we still require to identify the promoter region of this gene. We are currently carrying out S1-nuclease protection studies to identify the messenger RNA start site and Northern blot analysis to determine the size of the RNA. It will be of particular interest to identify the promoter elements for this gene and examine its regulation with respect to the biosynthesis and degradation of poly-β-hydroxybutyrate. We are currently working on the overproduction of the *Z. ramigera* biosynthetic thiolase in *E. coli* and large scale purification to facilitate x-ray crystallography studies.

Acknowledgments—We would like to thank Stephen Picataggio for useful discussions during the course of this work, Jerrie Gavalchin for purifying the thiolase antibodies, Keith Vass for assistance with the computer analysis, and Ginger Burr for preparing this manuscript.

REFERENCES

1. Crabtree, K., McCoy, E., Boyle, W. C., and Rohlich, G. A. (1965) *Appl. Microbiol.* **13,** 218–226
2. Tomita, K., Saito, T., and Fukui, T. (1983) in *Biochemistry of Organic Processes* (Lennon, D. L. F., Stratman, F. W., and Zahlten, R. N., eds) pp. 353–366, Elsevier Scientific Publishing Co., Amsterdam
3. Nishimura, T., Saito, T., and Tomita, K. (1978) *Arch. Microbiol.* **116,** 21–27
4. Saito, T., Fukui, T., Ikeda, F., Tanaka, Y., and Tomita, K. (1977) *Arch. Microbiol.* **114,** 211–217
5. Fukui, T., Yoshimoto, A., Matsumoto, M., Hosokawa, S., Saito, T., Nishikawa, H., and Tomita, K. (1976) *Arch. Microbiol.* **110,** 149–156
6. Middleton, B. (1973) *Biochem. J.* **132,** 717–730
7. Staack, H., Binstock, J. F., and Schulz, H. (1978) *J. Biol. Chem.* **253,** 1827–1831
8. Berndt, H., and Schlegel, H. G. (1975) *Arch. Microbiol.* **103,** 21–30
9. Hartmanis, M. G. N., and Stadtman, T. C. (1982) *Proc. Natl. Acad. Sci. U. S. A.* **79,** 4912–4916
10. Feigenbaum, J., and Schulz, H. (1975) *J. Bacteriol.* **122,** 407–411
11. Spratt, S. K., Black, P. N., Ragozzino, M. M., and Nunn, W. D. (1984) *J. Bacteriol.* **158,** 535–542
12. Young, R. A., and Davis, R. W. (1983a) *Science* **222,** 778–782
13. Young, R. A., and Davis, R. W. (1983b) *Proc. Natl. Acad. Sci. U. S. A.* **80,** 1194–1198
14. Birnboim, H. C., and Doly, J. (1979) *Nucleic Acids Res.* **7,** 1513–1523
15. Ish-Horowicz, D., and Burke, J. F. (1981) *Nucleic Acids Res.* **9,** 2989–2998
16. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
17. Anderson, S. (1981) *Nucleic Acids Res.* **9,** 3015–3026
18. Hohn, B., and Murray, K. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 3259–3263
19. Ouchterlony, O. (1949) *Acta Pathol. Microbiol. Scand.* **26,** 507–515
20. Bigbee, W. L., Vanderlaan, M., Fong, S. S. N., and Jensen, R. H. (1983) *Mol. Immunol.* **20,** 1353–1357
21. Laemmli, U. K. (1970) *Nature* **227,** 680–685
22. Burnette, W. N. (1981) *Anal. Biochem.* **112,** 195–203
23. Smith, G. E., and Summers, M. D. (1980) *Anal. Biochem.* **109,** 123–129
24. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517
25. Rigby, P. W. J., Dieckmann, M., Rhodes, C., and Berg, P. (1977) *J. Mol. Biol.* **113,** 237–251
26. Bradford, M. M. (1976) *Anal. Biochem.* **72,** 248–254
27. Messing, J., and Vieira, J. (1982) *Gene* **19,** 269–276
28. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–5467
29. Mills, D. R., and Kramer, F. R. (1979) *Proc. Natl. Acad. Sci. U. S. A.* **76,** 2232–2235
30. Staden, R., and McLachlan, A. D. (1982) *Nucleic Acids Res.* **10,** 141–158
31. Voordouw, G., Walker, J. E., and Brenner, S. (1985) *Eur. J. Biochem.* **148,** 509–514
32. Izbicka-Dimitrijevic, E., and Gilbert, H. F. (1982) *Biochemistry* **21,** 6112–6118
33. Dugan, P. R. (1981) in *The Prokaryotes* (Starr, M. P., Stolp, H., Trufer, H. G., Balows, A., and Schlegel, H. G. eds) pp. 764–770, Springer-Verlag, New York
34. Thompson, C. J., and Gray, C. S. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80,** 5190–5194
35. Brown, N. L., Ford, S. J., Pridmore, R. D., and Fritzinger, D. C. (1983) *Biochemistry* **22,** 4089–4095
36. Ikemura, T. (1981) *J. Mol. Biol.* **151,** 389–409
37. Gouy, M., and Gautier, C. (1982) *Nucleic Acids Res.* **10,** 7055–7074

38. Rosenberg, M., and Court, D. (1979) *Annu. Rev. Genet.* **13**, 319–353
39. Davis, J. T., Moore, R. N., Imperali, B., Pratt, A., Kobayashi, K., Masamune, S., Sinskey, A. J., and Walsh, C. T. (1987) *J. Biol. Chem.* **262**, 82–89
40. Davis, J. T., Chen, H.-H., Moore, R., Nishitani, Y., Masamune, S., Sinskey, A. J., and Walsh, C. T. (1987) *J. Biol. Chem.* **262**, 90–96

APPENDIX

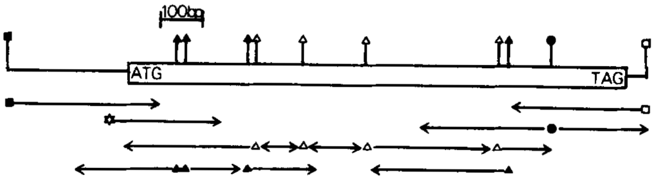

FIG. 4. **DNA sequencing strategy.** Only those restriction sites from which sequence data was obtained are shown. The extent of the sequence determined from each site is indicated by the arrows. Restriction sites are presented as follows: ■, *Eco*RI; □, *Sal*I; ●, *Bam*HI; △, *Sau*3AI; ▲, *Ava*I; ✡, *Bal*31 deletion. The *boxed portion* of the *Eco*RI-*Sal*I fragment is the thiolase structural gene.

FIG. 5. **DNA sequence of the thiolase gene.** The DNA sequence was determined by the method of Sanger *et al.* (28). The sequence was obtained for both strands for the entire structural gene region and from multiple sequencing runs. Due to the high G/C content, each sequence run was repeated using dITP. The first 26 amino acids and the 10-amino acid sequence of the active site peptide are boxed. A potential Shine-Dalgarno sequence is *underlined, S.D.*

```
                                          -270            -250            -230
                     EcoRI
                     GAATTCCTCA TCCCCGTGTT CCTCTTTTCT TTTTTGAGCC TTGCGAGTAT CCGCAAAAAG ACGCTTCCTG
  -210         -190              -170            -150            -130

ACAATCGAAT TGTGCGATGC GGCGAGGGAT GCTGCGCAAC ATGGATTTTC ATGCTGCGCG GCCCCCCGCA TGAAATTTTA GCCCCGCGGG ATATCCCGTT TGACTCTCCT
-110          -90              -70             -50             -30             -10

TCTCCGCTTT GCCAGTTTTG ATTTCGGGAC AACCGCCGGG CTCGCGCCGT TCGTGCCCCG GACCCGGTTC CAAGGCCGGC CGCAAGGCTG CCAGAACTGA GGAAGCACAT
    10                   30                   50                   70                  SD    90

ATG AGC ACC CCG TCC ATC GTC ATC GCC AGC GCC CGC ACC GCG GTC GGT TCC TTC AAC GGC GCT TTC GCC AAC ACG CCC GCC CAT GAA CTC
Met Ser Thr Pro Ser Ile Val Ile Ala Ser Ala Arg Thr Ala Val Gly Ser Phe Asn Gly Ala Phe Ala Asn Thr Pro Ala His Glu Leu
            110                 130                 150                 170

GGC GCG ACC GTG ATT TCG GCG GTT CTC GAG CGC GCG GGC GTT GCG GCG GGC GAG GTG AAC GAG GTG ATT CTC GGC CAG GTG CTG CCG GCC
Gly Ala Thr Val Ile Ser Ala Val Leu Glu Arg Ala Gly Val Ala Ala Gly Glu Val Asn Glu Val Ile Leu Gly Gln Val Leu Pro Ala
        190                 210                 230                 250                 270

GGC GAA GGC CAG AAC CCG GCC CGC CAG GCC GCC ATG AAG GCC GGC GTG CCG CAG GAG GCG ACC GCC TGG GGC ATG AAC CAG CTT TGC GGC
Gly Glu Gly Gln Asn Pro Ala Arg Gln Ala Ala Met Lys Ala Gly Val Pro Gln Glu Ala Thr Ala Trp Gly Met Asn Gln Leu Cys Gly
                        290                 310                 330                 350

TCG GGC CTG CGC GCC GTC GCG CTC GGC ATG CAG CAG ATC GCC ACG GGC GAT GCG AGC ATC ATC GTC GCC GGC GGC ATG GAA TCC ATG TCC
Ser Gly Leu Arg Ala Val Ala Leu Gly Met Gln Gln Ile Ala Thr Gly Asp Ala Ser Ile Ile Val Ala Gly Gly Met Glu Ser Met Ser
        370                 390                 410                 430                 450

ATG GCC CCG CAT TGC GCG CAT CTG GCC GGC GTG AAG ATG GGC GAT TTC AAG ATG ATC GAC ACG ATG ATC AAG GAC GGC CTG ACC GAC GCC
Met Ala Pro His Cys Ala His Leu Ala Gly Val Lys Met Gly Asp Phe Lys Met Ile Asp Thr Met Ile Lys Asp Gly Leu Thr Asp Ala
                        470                 490                 510                 530

TTC TAC GGC TAC CAC ATG GGC ACG ACC GCC GAG AAT GTC GCC AAG CAG TGG CAG CTT TCC CGC GAC GAG CAG GAC GCC TTC GCC GTC GCC
Phe Tyr Gly Tyr His Met Gly Thr Thr Ala Glu Asn Val Ala Lys Gln Trp Gln Leu Ser Arg Asp Glu Gln Asp Ala Phe Ala Val Ala
        550                 570                 590                 610                 630

TCG CAG AAC AAG GCC GAG GCC GCC CAG AAG GAC GGC CGC TTC AAG GAC GAG ATC GTT CCC TTC ATC GTC AAG GGC CGC AAG GGC GAC ATC
Ser Gln Asn Lys Ala Glu Ala Ala Gln Lys Asp Gly Arg Phe Lys Asp Glu Ile Val Pro Phe Ile Val Lys Gly Arg Lys Gly Asp Ile
                        650                 670                 690                 710

ACC GTC GAT GCC GAC GAA TAT ATC CGC CAC GGC GCG ACG CTC GAT TCC ATG GCG AAG CTC CGC CCG GCC TTC GAC AAG GAA GGC ACG GTG
Thr Val Asp Ala Asp Glu Tyr Ile Arg His Gly Ala Thr Leu Asp Ser Met Ala Lys Leu Arg Pro Ala Phe Asp Lys Glu Gly Thr Val
        730                 750                 770                 790                 810

ACG GCC GGC AAC GCC TCC GGC CTC AAT GAC GGC GCG GCC GCG GCC CTC CTG ATG AGC GAA GCG GAA GCC TCG CGC CGC GGC ATC CAG CCG
Thr Ala Gly Asn Ala Ser Gly Leu Asn Asp Gly Ala Ala Ala Ala Leu Leu Met Ser Glu Ala Glu Ala Ser Arg Arg Gly Ile Gln Pro
                        830                 850                 870                 890

CTC GGC CGC ATC GTT TCC TGG GCG ACG GTC GGC GTC GAT CCC AAG GTC ATG GGC ACC GGC CCG ATC CCG GCC TCC CGC AAG GCG CTC GAG
Leu Gly Arg Ile Val Ser Trp Ala Thr Val Gly Val Asp Pro Lys Val Met Gly Thr Gly Pro Ile Pro Ala Ser Arg Lys Ala Leu Glu
        910                 930                 950                 970                 990

CGC GCC GGC TGG AAG ATC GGC GAT CTC GAC CTC GTG GAA GCC AAC GAA GCC TTC GCG GCG CAG GCC TGC GCG GTC AAC AAG GAC CTC GGC
Arg Ala Gly Trp Lys Ile Gly Asp Leu Asp Leu Val Glu Ala Asn Glu Ala Phe Ala Ala Gln Ala Cys Ala Val Asn Lys Asp Leu Gly
                        1010                1030                1050                1070

TGG GAT CCG TCC ATC GTC AAC GTC AAC GGC GGT GCC ATC GCC ATC GGC CAC CCG ATC GGC GCG TCC GGC GCC CGC ATC CTC AAC ACG CTC
Trp Asp Pro Ser Ile Val Asn Val Asn Gly Gly Ala Ile Ala Ile Gly His Pro Ile Gly Ala Ser Gly Ala Arg Ile Leu Asn Thr Leu
        1090                1110                1130                1150                1170

CTC TTC GAG ATG AAG CGT CGC GGC GCC CGC AAG GGT CTC GCC ACG CTC TGC ATC GGC GGC GGC ATG GGC GTG GCG ATG TGC ATC GAG AGC
Leu Phe Glu Met Lys Arg Arg Gly Ala Arg Lys Gly Leu Ala Thr Leu Cys Ile Gly Gly Gly Met Gly Val Ala Met Cys Ile Glu Ser
                        1190                1210                1230

CTT TAG GCG TCA GCT TAT CCA AAA CTT TGC CAT GAC CTG CCG CCG GCG AGG GCG ACA GGT CGA C
Leu End                                                                     SalI
```