JOURNAL OF BACTERIOLOGY, May 1997, p. 3030–3035 Vol. 179, No. 9
0021-9193/97/$04.00+0

Characterization of the *Bacillus subtilis thiC* Operon Involved in Thiamine Biosynthesis

YI ZHANG, SEAN V. TAYLOR, HSIU-JU CHIU, AND TADHG P. BEGLEY*

Department of Chemistry, Cornell University, Ithaca, New York 14853

Received 17 December 1996/Accepted 25 February 1997

The characterization of a three-gene operon (the *thiC* operon) at 331 min, which is involved in thiamine biosynthesis in *Bacillus subtilis*, is described. The first gene in the operon is homologous to transcription activators in the *lysR* family. The second and third genes (*thiK* and *thiC*) have been subcloned and overexpressed in *Escherichia coli*. ThiK (30 kDa) catalyzes the phosphorylation of 4-methyl-5-(β-hydroxyethyl)thiazole. ThiC (27 kDa) catalyzes the substitution of the pyrophosphate of 2-methyl-4-amino-5-hydroxymethylpyrimidine pyrophosphate by 4-methyl-5-(β-hydroxyethyl)thiazole phosphate to yield thiamine phosphate. Transcription of the *thiC* operon is not regulated by thiamine or 2-methyl-4-amino-5-hydroxymethylpyrimidine and is only slightly repressed by 4-methyl-5-(β-hydroxyethyl)thiazole.

Thiamine (THI) pyrophosphate (vitamin B_1) is the cofactor required for transketolase, α-keto acid decarboxylase, and α-keto acid oxidase, enzymes that are important in carbohydrate metabolism (8). In contrast to the situation in *Escherichia coli*, *Salmonella typhimurium*, and *Saccharomyces* species, little is known about the THI biosynthetic pathway in *Bacillus subtilis* (5). The cofactor is assumed to be formed by linking thiazole (THZ) phosphate (THZ-P) and pyrimidine (HMP) pyrophosphate (HMP-PP) (5) (Fig. 1). These phosphorylated precursors can be formed in vivo from the corresponding alcohols THZ and HMP. Labeling studies have demonstrated that THZ is derived from glycine (23) and HMP is derived from formate (27). This suggests that THZ biosynthesis is similar to that in yeast and HMP biosynthesis is similar to that in *E. coli*. This is tentative, however, until the origin of all the other atoms of THI has been determined.

Three genetic loci involved in the biosynthesis of THI in *B. subtilis* have been identified: *thiA* (70 min), *thiB* (105 min), and *thiC* (331 min) (11, 17, 25, 29). In feeding experiments, a *thiA* mutant is HMP requiring, a *thiB* mutant is THZ requiring, and a *thiC* mutant is THI requiring. An HMP biosynthetic gene (*thiA*) has recently been cloned and sequenced and shows high sequence similarity to the *E. coli thiC* gene (30). None of the other THI biosynthetic genes have been characterized, and the regulation of the pathway has not been studied. As part of the *B. subtilis* genome sequencing project, a cluster of three genes in the min 331 region of the genome has been cloned and identified as being involved in THI biosynthesis by complementation studies (7). In this paper, we use this sequence information to further characterize this putative three-gene operon.

MATERIALS AND METHODS

Bacterial strains, growth media, and chemicals. *B. subtilis* wild-type strain CU1065 was a gift from S. Zahler. Minimal growth medium for *B. subtilis* (CMP) was Spizizen's minimal medium (22) containing 1% (wt/vol) vitamin-free casein hydrolysate (Nutritional Biochemicals Corp., Cleveland, Ohio). Defined medium consisted of CMP supplemented with 10 or 50 μM THI, 20 μM 4-methyl-5-(β-hydroxyethyl)thiazole (THZ), 20 μM 4-amino-5-hydroxymethyl-2-methylpyrimidine (HMP), or 20 μM HMP plus 20 μM THZ as indicated in Results and Discussion.

E. coli was grown in Luria-Bertani broth or in 2× YT broth (14). Antibiotics were used at the following concentrations: ampicillin, 200 μg/ml; kanamycin, 25 μg/ml; and chloramphenicol, 15 μg/ml. THZ-P was synthesized in two steps from THZ by the method of Backstrom and Begley (3). HMP was a gift from Hoffmann-La Roche (Basel, Switzerland). Type IV-S potato acid phosphatase, THZ, THI, and ATP were from Sigma Chemical Co. Potassium ferricyanide, sodium hydroxide, hydrogen peroxide, potassium dihydrogen phosphate, trichloroacetic acid, spectroscopic-grade isobutanol, and potassium acetate were from Fisher Scientific.

DNA methodology. Standard techniques were used for plasmid isolation, restriction endonuclease digestion, isolation of DNA fragments from gels, and ligations (2, 20). Linearized plasmids were purified with Geneclean II (Bio 101, Inc., La Jolla, Calif.). PCR was carried out as described by Innes et al. (10). DNA sequence determination was performed with an ABI 373A automatic DNA sequencer, with a *Taq* DyeDeoxy Terminator Cycle Sequencing Kit (Perkin-Elmer, Applied Biosystems Division, Foster City, Calif.), at the Cornell University DNA Sequencing Facility. Partial DNA sequencing of double-stranded DNA was performed with the Sequenase sequencing kit (U.S. Biochemical Corp.). The DNA in the partial DNA sequencing was radiolabeled with [γ-^{32}P]ATP (Amersham).

Computer-aided analysis. DNA sequences were assembled and analyzed with DNA Inspector IIe (Textco, West Lebanon, N.H.) and DNA Strider (Institute de Recherche Fondamentale, Commissariat à l'Energie Atomique, Paris, France). The nucleotide and deduced amino acid sequences were compared with available sequences in the GenBank database with the NCBI BLAST program mail server (1). Peptide sequence alignment was done with both BLAST and the Genetics Computer Group Package from the University of Wisconsin. Enzyme kinetic data was analyzed with EnzymeKinetics, version 1.4 (Trinity Software, Campton, N.H.).

Subcloning of *thiC* and *thiK*. Plasmid pλthiC2 (obtained from P. Glasser, Pasteur Institute), which contained the *thiC* gene cluster on a 10-kb insert, was propagated by transforming *E. coli* XL1-Blue (Stratagene) and selecting for Cm^r.

For the subcloning of *thiC* (*ipa-26d*), the 1,600-bp *Hin*dIII fragment from pλthiC2 was ligated into the *Hin*dIII site of pGEM-3zf(+) (Promega), yielding plasmid pYZCB4829. This was then digested with *Nla*III, and the resulting 822-bp fragment was ligated into the *Sph*I site of the pQE-32 expression vector (18) to yield plasmid pYZC6927. This encodes a recombinant protein containing the *thiC* gene product with a six-histidine tag on its N terminus. Plasmid pYZC6927 was partially sequenced to confirm the sequence of *thiC* and to confirm that *thiC* had been inserted in the correct reading frame on the expression vector.

The *thiK* (*ipa-25d*) gene and a substantial amount of downstream DNA were amplified by PCR using pλthiC2 as the template and oligonucleotides 5′GGG GATCCATGGATGCACAATCAGCAGC3′ and 5′GGCTGCAGACGGAGT CATGAGACAGTCAC3′ as the primer pair. The PCR product was digested with *Bam*HI and *Pst*I and cloned into the expression vector pQE-30 (18), yielding plasmid pYZK2. The *Nae*I-*Pst*I fragment of pYZK2 was then exchanged with the *Nae*I-*Pst*I fragment from pYZK1, yielding plasmid pYZK3. This encodes a recombinant protein containing the *thiK* gene product with a six-histidine tag on its N terminus. [Plasmid pYZK1 was constructed by subcloning the 2,205-bp *Sac*I-*Bam*HI fragment from pλthiC2 into pGEM-3zf(+).] Plasmid pYZK3 was partially sequenced to confirm the sequence of the PCR-derived part of *thiK* and to

* Corresponding author. Mailing address: Department of Chemistry, 120 Baker Laboratory, Cornell University, Ithaca, NY 14853-1301. Phone: (607) 255-7133. Fax: (607) 255-4137. E-mail: tpb2@cornell.edu.

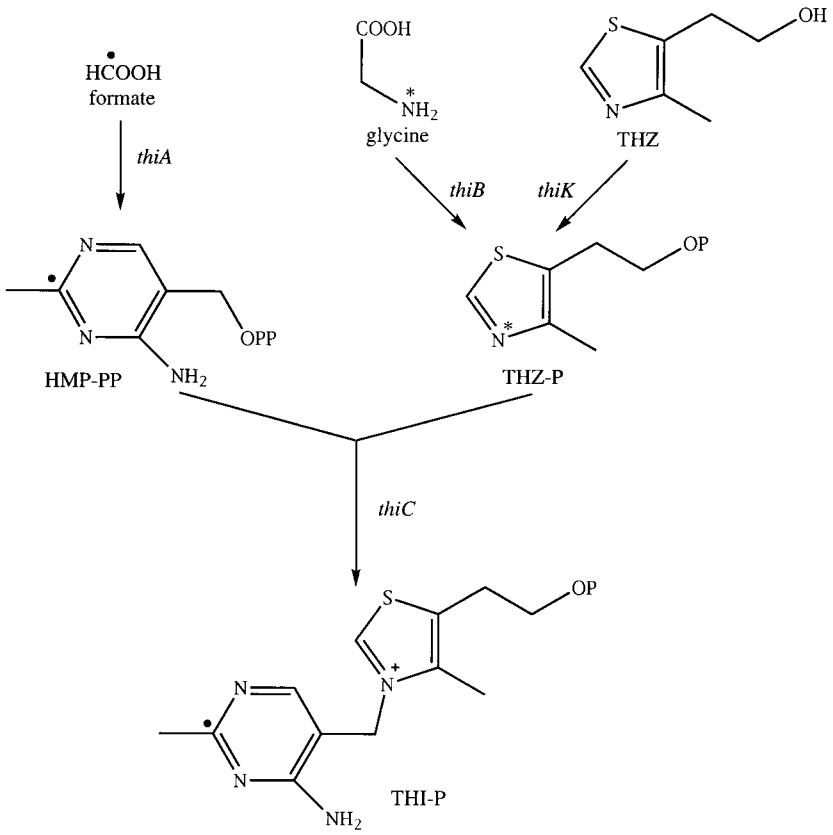

FIG. 1. THI biosynthesis in *B. subtilis*. THI phosphate (THI-P) is formed by linking THZ-P and HMP-PP. The corresponding alcohols (THZ, HMP, and THI) can be phosphorylated in vivo and incorporated into the cofactor. The THZ nitrogen is derived from glycine, and the C-2 of HMP is derived from formate. Three genetic loci involved in the biosynthesis of THI have been identified. In feeding experiments, a *thiA* mutant is HMP requiring, a *thiB* mutant is THZ requiring, and a *thiC* mutant is THI requiring.

confirm that *thiK* had been inserted in the correct reading frame on the expression vector.

Overproduction and purification of ThiC and ThiK. The standard Qiagen procedure for gene expression and protein purification (18) was used with minor modifications. Plasmids pYZC6927 and pYZK3 were used to transform *E. coli* SG13009(pREP4). Transformants were selected on ampicillin-kanamycin plates. A single colony of *E. coli* containing pYZC6927 or pYZK3 was grown to mid-log phase at 37°C in 3 ml of 2× YT broth supplemented with ampicillin and kanamycin. The culture was then diluted to 1 liter with the same medium and grown to an optical density at 600 nm of approximately 1.0. Expression was induced by adding isopropyl-β-D-thiogalactopyranoside (IPTG; 2 mM, final concentration), and growth was continued at 37°C for 5 to 6 h. The cells were then harvested by centrifugation (30 min, 5,000 × *g*) and disrupted by treatment with lysozyme followed by sonication. After centrifugation (15 min, 16,000 × *g*), the crude cell extract was loaded onto a column containing 1 ml of Ni^{2+}-nitrilotriacetic acid (NTA)-agarose resin. The column was then washed with 20 to 25 column volumes of SB buffer (50 mM Na-phosphate [pH 7.8], 300 mM NaCl buffer) and 20 to 25 column volumes of WB buffer (50 mM Na-phosphate, 300 mM NaCl, 10% glycerol, pH 6.0) containing 1 mM imidazole. The six-His-tagged proteins were eluted with WB buffer containing 80 mM imidazole. Protein-containing fractions were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). ThiC- or ThiK-containing fractions were pooled, concentrated in an Amicon Ultrafiltration cell with a YM-10 membrane, dialyzed against 50 mM Tris-HCl (pH 7.5), and stored at −70°C in 10% glycerol. The purified enzymes were >85% homogeneous (see Fig. 6), and the yield was ≈25 mg/liter for both proteins.

RNA methodology. Total cellular RNA was prepared from mid-exponential-phase cells of strain CU1065 by the method of Maes and Messens (12). The 5′ end of the in vivo transcript was mapped by primer extension analysis according to the method described by Sambrook et al. (20). The synthetic oligonucleotide 5′GATATCGTCGGCTGAGAAACA3′ (PPE24d) was radiolabeled with T4 polynucleotide kinase (Promega) and [γ-^{32}P]ATP (Amersham). Ten micrograms of RNA, 10 U of avian myeloblastosis virus reverse transcriptase (Promega), and the buffer supplied by the manufacturer were used in the primer extension reaction. A portion of the reaction mixture was loaded onto a 6% sequencing gel next to a sequencing ladder generated with plasmid pλthiC2 and primer PPE24d. A no-DNA control was included in the primer extension analysis.

The Northern (RNA) blot was produced according to the method described by Sambrook et al. (20). Ten micrograms of cellular RNA and RNA size markers (G3191; Promega) was separated on a 1% agarose-formaldehyde gel and transferred to a positively charged nylon membrane (Boehringer Mannheim). Bands on the gel were visualized by adding ethidium bromide to the sample buffer. Probes were generated by PCR with primers designed to generate DNA fragments that were equivalent to the coding regions of *ipa-23r*, *ipa-24d*, *ipa-25d*, *ipa-26d*, and *ipa-27d*. The template DNA for the PCR was plasmid pλthiC2 (for the *ipa-23r*, *ipa-24d*, *ipa-25d*, and *ipa-26d* probes) and pλthiC7 (for the *ipa-27d* probe). The *thiA* probe was also generated by PCR with template pYZ3 (28) (GenBank accession number U26178) and primers CPA6 (5′TCAGTGCAGC AAGCCAACAACAT3′) and L3 (5′ACGGGCTTTCGATAACGC3′). Probe DNAs were labeled with [α-^{32}P]dATP (Amersham). For each blot, 10^6 cpm of the probe was used. Prehybridization and hybridization were carried out at 65°C in Church and Gilbert solution (1 M Na-phosphate buffer, 20% SDS, 1 mg of bovine serum albumin per ml, 0.5 M EDTA), and washings were carried out as described by Sambrook et al. (20). Northern blots were visualized by X-ray autoradiography.

Enzyme assay for ThiC. The thiochrome assay for THI was carried out as described by Wyatt et al. using potassium ferricyanide instead of cyanogen bromide as the oxidant (26). The reaction mixture (250 μl [total volume]) contained 50 mM Tris (pH 7.5), 6 mM $MgSO_4$, 2 μg of purified ThiC protein, 0 to 50 μM THZ-P, and 0 to 50 μM HMP-PP. Aliquots (60 μl) were removed at 30-s intervals over a 2-min period and quenched with 60 μl of 10% trichloroacetic acid. These mixtures were centrifuged to remove protein, and 100 μl of each clarified solution was added to 200 μl of 4 M potassium acetate. Potato acid phosphatase (type IV-S) was then added (100 μl of a 2-mg/ml solution), and the mixture was incubated for 20 h at 37°C. The resulting THI was oxidized to thiochrome by the addition of 100 μl of 3.8 mM $K_3Fe(CN)_6$ in 7 M NaOH, immediately followed by thorough mixing. The reaction was quenched after 30 s by the addition of 100 μl of 0.06% H_2O_2 in 5.5 M KH_2PO_4, diluted with 400 μl of water, and then extracted with 1.4 ml of isobutanol. After phase separation, 1.2 ml of the organic phase was transferred to a cuvette and the thiochrome

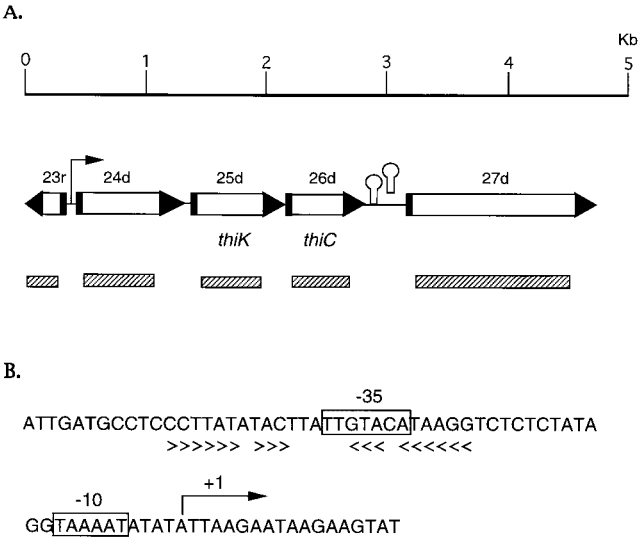

FIG. 2. (A) Schematic representation of the *B. subtilis thiC* operon and flanking DNA. The genes previously determined as part of the *B. subtilis* genome sequencing project are shown at the top of the figure (*ipa-23r*, *-24d*, *-25d*, *-26d*, and *-27d*). The hatched boxes indicate the regions corresponding to the DNA probes designed to hybridize to RNAs in the Northern blot analysis shown in Fig. 3 and 7. (B) The *thiC* promoter-operator region. The sequence of the nontranscribed strand is shown. Symbols: □▶, ORF with orientation; ▮, *Bacillus* ribosome binding site; ♀, putative rho-independent transcription termination site; ⌐→, transcription start site determined by primer extension analysis; >>>> <<<<, putative operator region with dyad symmetry structure.

fluorescence was recorded on a Turner model 450 fluorometer containing a narrow-band 360-nm excitation filter and a sharp-cutoff 430-nm emission filter. A calibration curve was prepared by converting standard THI phosphate solutions to thiochrome in an identical manner. Initial rates were collected at five different substrate concentrations, and the kinetic parameters were calculated with the program EnzymeKinetics, version 1.4.

Enzyme assay of ThiK. A 1.5-ml assay solution containing 14 mM THZ and 14 mM ATP was prepared in 20 mM Na_2HPO_4–10 mM $MgSO_4$–5 mM dithiothreitol, pH 7.5. ThiK (1.3 mg) was added to the assay mixture. After incubation for 2.5 h at 37°C the reaction mixture and a control lacking ThiK were lyophilized for 12 h. Each sample was resuspended in 500 μl of D_2O. ^1H nuclear magnetic resonance (NMR) spectra were recorded on a Varian 400-MHz NMR spectrometer.

RESULTS AND DISCUSSION

Identification of the *thiC* operon. A 97-kb region, from 325° to 333°, has previously been cloned and sequenced as part of the *B. subtilis* genome sequencing project (7). It was proposed that three adjacent genes, *ipa-24d*, *ipa-25d*, and *ipa-26d* (in the 331° region), formed an operon. The orientation of each open reading frame (ORF) was the same, and a putative transcription terminator was located at the end of the three ORFs (Fig. 2A). In addition, an *ipa-24–ipa-26d*-containing plasmid complemented the *thiC* mutant (THI requiring) of *B. subtilis*, and disruption of *ipa-24–ipa-26d* caused THI auxotrophy. However, it was not clear which gene of this putative operon corresponded to *thiC* or whether *ipa-24–ipa-26d* were regulatory genes or structural genes (7).

The genes contained in the *thiC* operon were experimentally identified by Northern blotting. Plasmid pλthiC2, which contains *ipa-24–ipa-26d* on a 10-kb insert, was used as the PCR template to generate probe DNAs. The probes were equivalent to the length of each gene in the putative operon: *ipa-24d*, *ipa-25d*, and *ipa-26d*, as well as the upstream and downstream flanking genes, *ipa-23r* and *ipa-27d* (Fig. 2A). When wild-type strain CU1065 RNA was probed with *ipa-24d*, *ipa-25d*, and *ipa-26d* DNA fragments, an ~2.4-kb band was observed on the Northern blot in each case. The size of this RNA coincided with the calculated size of *ipa-24–ipa-26d* from the DNA sequence (2.409 kb). The smears located above and below the 2.4-kb band may be due to contamination with genomic DNA and RNA degradation, respectively (Fig. 3). This RNA was strongly expressed in cells grown in THI-free medium. These results suggest that the *ipa-24d*, *-25d*, and *-26d* genes are transcribed in one mRNA molecule and belong to one operon (the *thiC* operon) that is involved in THI biosynthesis.

When wild-type strain CU1065 RNA was probed with *ipa-23r*, a faint 1,300-nucleotide (nt) band was observed (Fig. 3), indicating that the *ipa-23r* gene is transcribed divergently and does not belong to the *thiC* operon. When the same RNA was probed with *ipa-27d*, a faint 1,300- to 3,400-nt smear (and possibly an ~3,400-nt band) was observed (Fig. 3). It is unlikely that *ipa-27d* is cotranscribed with *ipa-24d–ipa-25d*. Because the calculated size of *ipa-24d–ipa-27d* is 3,800 nt, any band resulting from cotranscription of *ipa-24d–ipa-27d* should have migrated above the 3,638-nt molecular size marker. The lack of a band above the 3,638-nt molecular size marker implies that *ipa-27d* is not cotranscribed with *ipa-24d–ipa-26d*. Since *ipa-23r* and *ipa-27d* are expressed at a very low level in THI-free medium and are not cotranscribed with *ipa-24d–ipa-26d*, they are probably not involved in THI biosynthesis as structural genes.

Organization of the *thiC* operon. The start site for transcription of the *thiC* operon was determined by primer extension analysis. An oligonucleotide complementary to the *ipa-24d* sequence at positions 24537 to 24553 in the published *B. subtilis* genomic DNA sequence (7) was labeled with [γ-^{32}P]ATP and hybridized to the RNA extracted from strain CU1065 (wild type) grown in THI-free medium. A single start site was mapped 47 bp upstream of the proposed start codon of the *ipa-24d* ORF (Fig. 2B and 4). An oligonucleotide complementary to the *ipa-25d* sequence was also used as the primer in the primer extension analysis for the transcription start site, but no band was detected (data not shown).

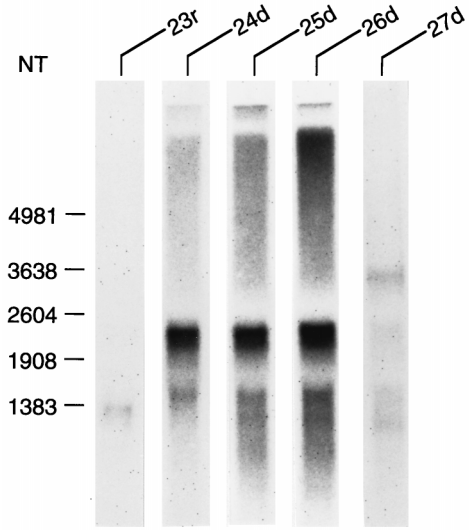

FIG. 3. Northern blot analysis of the gene components of the *thiC* operon. RNA loaded on each lane (12 μg) was extracted from exponentially growing strain CU1065 (wild type) in minimal medium in the absence of THI and its precursors. The probes used to hybridize to the RNAs were DNA fragments corresponding to the ORFs *ipa-23r*, *ipa-24d*, *ipa-25d*, *ipa-26d*, and *ipa-27d*. RNA size markers are indicated on the left.

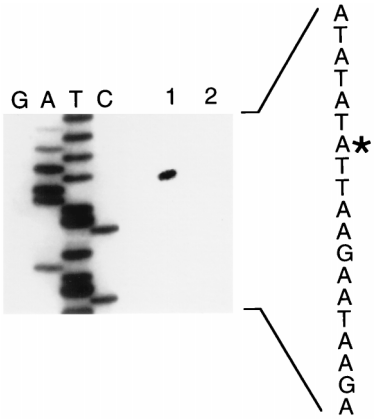

FIG. 4. Determination of the transcription start site of the *thiC* operon by primer extension analysis. RNA was purified from strain CU1065 (wild type). An oligonucleotide complementary to positions 24537 to 24553 in the published *B. subtilis* genomic DNA sequence was labeled with [γ-^{32}P]ATP and hybridized to the RNA. Reference DNA sequencing reactions (lanes C, A, T, and G) were initiated with the same primer. The primer extension product, derived from 10 μg of RNA, was loaded in lane 1. The primer extension reaction mixture in lane 2 contained no RNA. The DNA sequence is the complement of the sequence that can be read from the sequencing ladder and spans positions 24404 to 24425. The asterisk marks the start site of transcription.

The promoter region of the *thiC* operon was identified by sequence analysis. The best match to the σ^A factor -10 consensus sequence (TATAAT) is TAAAAT, and the best match to the -35 consensus sequence (TTGACA) is TTGTAC (15). There are 18 bp between the -10 and -35 sequences. A dyad symmetry structure with the sequence CCTTATATACTTAT TGTACATAAGG is located in this putative -35 promoter region and is likely to be the operator of the operon (Fig. 2B).

No putative promoter region was found for the second and third genes in the proposed *thiC* operon. The *ipa-25d* ORF is 7 bp downstream from the *ipa-24d* ORF, and the *ipa-25d* and *-26d* ORFs overlap by 4 bp.

A putative rho-independent transcription terminator has been located downstream of the *ipa-26d* ORF (GCTGTTCT GTAAAAGGACAGC). This overlaps the stop codon and is followed by six T's.

Promoter-like structures, similar to the σ^A factor consensus sequences, as well as Shine-Dalgarno sequences (21) were found for the upstream (*ipa-23r*) and the downstream (*ipa-27d*) genes (data not shown). *ipa-23r* is transcribed from a putative promoter that is divergent from but overlapping with the *thiC* promoter. A dyad symmetry structure is located 65 bp downstream of the putative *ipa-26d* transcription terminator. Since this is located between the start codon of *ipa-27d* and its putative promoter, we propose that it is a regulatory site for *ipa-27d* rather than the transcription terminator for the *thiC* operon.

Sequence homology search for the genes in the *thiC* operon. To determine the function of the proteins encoded by the *thiC* operon and the flanking genes, the deduced amino acid sequences of the *ipa-23r*, *-24d*, *-25d*, *-26d*, and *-27d* products were compared with available sequences in the GenBank database.

ipa-23r and *ipa-27d*, the upstream and downstream genes of the *thiC* operon, exhibited no significant similarity to any DNA or protein sequence found in the GenBank database.

Ipa-24d showed significant homology to regulatory proteins of the *lysR* family (9). Two *B. subtilis* transcription activators, GltC in glutamate biosynthesis (6) and AlsR in acetoin biosynthesis (19), show significant homology to Ipa-24d. For these three proteins, ~30% of the residues are identical and ~30% are conservative changes, giving an overall similarity of ~60% (data not shown). When *gltC* and *alsR* are disrupted, the glutamate and acetoin biosynthesis enzymes are not expressed. Unlike the roles of *gltC* and *alsR*, however, insertional mutagenesis (28) of *ipa-24d* in *B. subtilis* did not generate THI auxotrophy, though it did impair the growth rate of the cell (29). The role of *ipa-24d* is unclear at this stage, and further investigation is needed.

Ipa-26d showed significant homology with *E. coli* THI phosphate synthase. This enzyme catalyzes the substitution of the pyrophosphate of HMP-PP with THZ-P to give THI phosphate (24). The database search also revealed that Ipa-25d and Ipa-26d were significantly homologous to the C-terminal region and to the N-terminal region, respectively, of Thi6 from *Saccharomyces cerevisiae* (16) and Thi4 from *Schizosaccharomyces pombe* (31). Ipa-26d and Ipa-25d together have 21% identical residues and 13% conservative substitutions compared to Thi6 and Thi4, giving an overall 34% homology for the total 494-amino-acid sequence of Ipa-26d–Ipa-25d (Fig. 5). Thi6 is a bifunctional protein that catalyzes the ATP-dependent phosphorylation of THZ and the substitution of the pyrophosphate of HMP-PP with THZ-P to give THI phosphate (16). The kinase activity is associated with the C-terminal region, and the THI phosphate synthase activity is associated with the N-terminal region (16). This indicates that Ipa-26d (which we will now call ThiC) could be THI phosphate synthase and Ipa-25d (which we will now call ThiK) could be 4-methyl-5-(β-hydroxyethyl)thiazole kinase. To test this proposal, ThiC and ThiK were overexpressed in *E. coli* as fusion proteins and were tested for enzymatic activity in vitro.

Enzymatic assays of ThiC. *thiC* was cloned into the expression vector PQE-32 and overexpressed in *E. coli*. The overexpressed protein has an apparent mass of 26 kDa by SDS-PAGE (23.7 kDa calculated) and was soluble. ThiC was purified (>85%) by affinity chromatography on a Ni^{2+}-NTA-agarose resin (Fig. 6).

We confirmed the proposed THI phosphate synthase activity for ThiC using the thiochrome assay (26). This assay measures the amount of THI formed from THZ-P and HMP-PP by converting it to the highly fluorescent thiochrome. Mg^{2+} is required for activity. The kinetic parameters are as follows: K_m(HMP-PP) = 0.7 ± 0.2 μM, K_m(THZ-P) = 1.2 ± 0.5 μM, V_{max} = 0.7 ± 0.2 μmol mg^{-1} min^{-1}, and k_{cat} = 0.0014 ± 0.0005 s^{-1}. These parameters are similar to those previously reported for the *E. coli* THI phosphate synthase (4), suggesting that the amino-terminal six-His fusion has little or no effect on the catalytic activity of ThiC.

Enzymatic assays of ThiK. *thiK* was cloned into the expression vector PQE-30 and overexpressed in *E. coli*. The overexpressed protein has an apparent mass of 30 kDa by SDS-PAGE (27.8 kDa calculated) and was soluble. ThiK was purified (>85%) by affinity chromatography on a Ni^{2+}-NTA-agarose resin (Fig. 6).

We confirmed the proposed THZ kinase activity of ThiK by ^{1}H NMR analysis of the THZ-ATP-ThiK reaction mixture. The spectrum measured in D_2O clearly demonstrated the formation of THZ-P from THZ and ATP only when in the presence of ThiK and Mg^{2+}. Peaks attributable to THZ-P (δ 2.22 [s, 3H], 2.96 [t, J = 6.4 Hz, 2H], 3.84 [dt, J = 6.4 Hz, $J_{P\text{-}CH}$ = 6.4 Hz, 2H], 5.98 [s, 1H]) were easily identified and distinct from peaks attributable to THZ (δ 2.20 [s, 3H], 2.87 [t, J = 6 Hz, 2H], 3.65 [t, J = 6 Hz, 2H], 5.98 [s, 1H]). We were unable to determine the kinetic parameters for THZ kinase. A coupled assay using THI phosphate synthase to convert THZ-P to THI followed by thiochrome formation gave poor-quality ki-

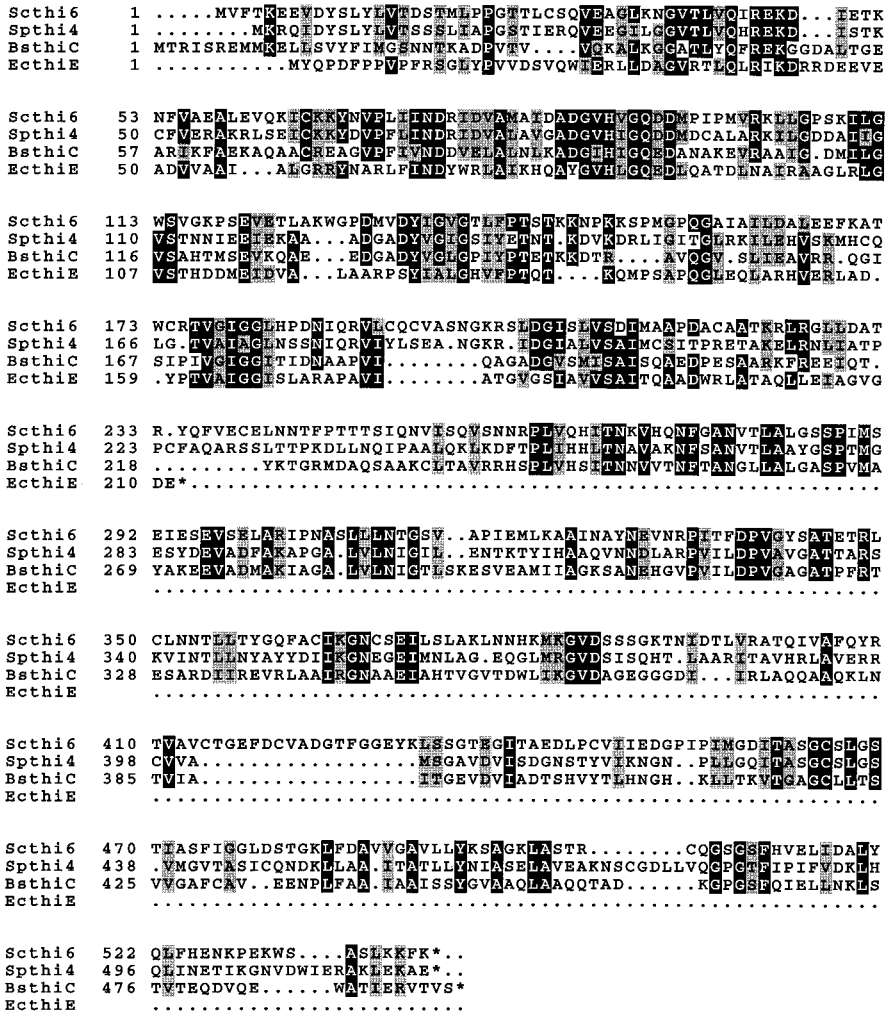

FIG. 5. Alignment of the deduced amino acid sequence of *B. subtilis* Ipa25d-Ipa26d (BsthiC) with the sequences of *E. coli* ThiE (EcthiE) (GenBank accession no. M88710), *S. cerevisiae* Thi6 (Scthi6) (GenBank accession no. D31908), and *S. pombe* Thi4 (Spthi4) (GenBank accession no. X78824). For BsthiC, amino acids 1 to 222 are from Ipa-26d and amino acids 223 to 495 are from Ipa-25d. Black boxes indicate complete identity, and shaded boxes indicate conservative amino acid substitutions.

netic data. Analysis of the time course of the reaction by NMR using the integrated rate expression also gave poor-quality kinetic data.

Expression of the *thiC* operon. The effects of THI and its biosynthetic precursors (THZ and HMP) on the expression of the *thiC* operon were examined by Northern blot analysis. RNA was isolated from wild-type strain CU1065 grown in minimal medium (CMP) supplemented with THI (10 or 50 μM), HMP (20 μM), or THZ (20 μM). The probe was a *thiC*-specific DNA fragment generated by PCR using *ipa-26d* as the template. The results (Fig. 7A) indicated that the expression of the *thiC* operon was weakly repressed by THZ and not at all by THI. This is in contrast to the expression of another *B. subtilis* gene, *thiA*, which is located in the min 70 region of the genome and is involved in the biosynthesis of the HMP moiety. *thiA* was strongly repressed by HMP (its putative biosynthetic product), THZ plus HMP, and THI (Fig. 7B). This pattern of regulation is different from the pattern found in *S. pombe* and *S. cerevisiae*. In *S. pombe*, the THI phosphate synthase gene is significantly repressed by as little as 1 μM THI and slightly repressed by THZ (31), and the gene involved in HMP formation is strongly repressed by THI (13). The THI

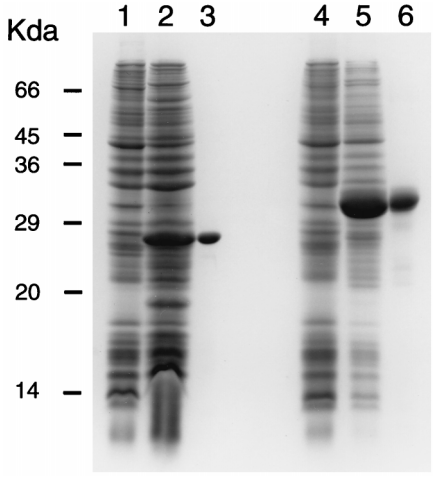

FIG. 6. Coomassie blue-stained SDS-PAGE gel demonstrating expression and purification of His-tagged ThiC (lanes 1 to 3) and ThiK (lanes 4 to 6). Lanes 1 and 4, whole-cell extract before induction by IPTG; lanes 2 and 5, whole-cell extract 6 h after induction; lanes 3 and 6, ThiC and ThiK proteins eluted with WB buffer containing 80 mM imidazole. Molecular mass standards are given on the left.

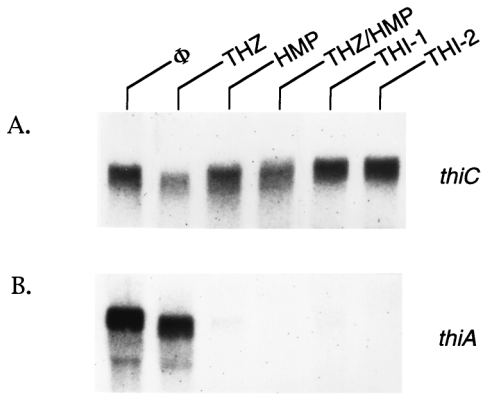

FIG. 7. Northern blot analysis of the transcription of the *thiC* operon and the *thiA* gene. RNA loaded on each lane (12 μg) was extracted from exponentially growing strain CU1065 (wild type) in the presence and absence of THI and its precursors. Lanes: Φ, minimal medium; THZ, 20 μM THZ; HMP, 20 μM HMP; THZ/HMP, 20 μM THZ and 20 μM HMP; THI-1, 10 μM THI; THI-2, 50 μM THI. (A) RNA was probed with a *thiC* DNA fragment; (B) as control, the same loading of RNA samples was probed with a *thiA* DNA fragment.

phosphate synthase gene in *S. cerevisiae* is also strongly repressed by THI (16). Clearly, the regulation of the THI biosynthetic pathway is complex. In the organisms looked at so far, the biosynthesis genes are scattered throughout the chromosome. Even where the pathways appear to be similar, e.g., *E. coli* and *B. subtilis*, the gene organization is different. The logic of the regulation of THI biosynthesis is not understood in any system and will probably not be clear until all of the gene clusters in a single organism have been characterized.

ACKNOWLEDGMENTS

We thank Philippe Glasser for providing plasmids pλthiC2 and pλthiC7. We thank Stanley Zahler for providing *B. subtilis* CU1065 and for helpful advice. We thank Peter Vander Horn, Allyson Backstrom, Dolf van Loon, Peter Hohmann, Knut Eichler, and John Helmann for helpful advice.

This research was supported by grants from Hoffmann-La Roche and the National Institutes of Health (DK44083). S.V.T. was supported in part by an NIH training grant.

REFERENCES

1. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
2. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. D. Serdman, J. A. Smith, and K. Struhl (ed.).** 1987. Current protocols in molecular biology. Greene Publishing Associates and Wiley Interscience, New York, N.Y.
3. **Backstrom, A. D., and T. P. Begley.** 1993. Unpublished data.
4. **Backstrom, A. D., R. Austin, S. McMordie, and T. P. Begley.** 1995. Biosynthesis of thiamin. I. The function of the *thiE* gene product. J. Am. Chem. Soc. **117:**2351–2352.
5. **Begley, T. P.** 1996. The biosynthesis and degradation of thiamin (vitamin B_1). Nat. Prod. Rep. **13:**177–185.
6. **Bohannon, D. E., and A. L. Sonenshein.** 1989. Positive regulation of glutamate biosynthesis in *Bacillus subtilis*. J. Bacteriol. **171:**4718–4727.
7. **Glasser, P., F. Kunst, M. Arnaud, M.-P. Coudart, W. Gonzales, M.-F. Hullo, M. Lonescu, B. Lubochinsky, L. Marcelino, I. Moszer, E. Presecan, M. Santana, E. Schneider, J. Schweizer, A. Vertes, G. Rapoport, and A. Danchin.** 1993. Bacillus subtilis genome project: cloning and sequencing of the 97 kb region from 325° to 333°. Mol. Microbiol. **10:**371–384.
8. **Haake, P.** 1987. Thiamin dependent enzymes, p. 390–403. *In* M. I. Page and A. Williams (ed.), Enzyme mechanisms. Royal Society of Chemistry, London, United Kingdom.
9. **Henikoff, S., G. W. Haughn, J. M. Calvo, and J. C. Wallace.** 1988. A large family of bacterial activator proteins. Proc. Natl. Acad. Sci. USA **85:**6602–6606.
10. **Innes, M. A., D. H. Gelfand, J. J. Sninsky, and T. J. White.** 1990. PCR protocols: a guide to methods and applications. Academic Press, New York, N.Y.
11. **Kelly, M. S.** 1967. Physical mapping properties of distant linkages between genetic markers in transformation of *Bacillus subtilis*. Mol. Gen. Genet. **99:**333–349.
12. **Maes, M., and E. Messens.** 1992. Phenol as grinding material in RNA preparations. Nucleic Acids Res. **20:**4374.
13. **Maundrell, K.** 1990. *nmt 1* of fission yeast. J. Biol. Chem. **265:**10857–10864.
14. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
15. **Moran, C. P., Jr.** 1989. Sigma factors and the regulation of transcription, p. 167–184. *In* I. Smith, R. A. Slepecky, and P. Setlow (ed.), Regulation of procaryotic development: structure and functional analysis of bacterial sporulation and germination. American Society for Microbiology, Washington, D.C.
16. **Nosaka, K., H. Nishimura, Y. Kawasaki, T. Tsujihara, and A. Iwashima.** 1994. Isolation and characterization of the *thi6* gene encoding a bifunctional thiamin-phosphate pyrophosphorylase/hydroxethylthiazole kinase from *Saccharomyces cerevisiae*. J. Biol. Chem. **269:**30510–30516.
17. **Pigott, P. J., M. Amjad, J.-J. Wu, H. Sandoval, and J. Castro.** 1990. Genetic and physical maps of *Bacillus subtilis* 168, p. 493–543. *In* C. R. Harwood and S. M. Cutting (ed.), Molecular biological methods for *Bacillus*. John Wiley and Sons, Chichester, United Kingdom.
18. **Qiagen Inc.** 1992. The QIAExpressionist: the high level expression and protein purification system. Qiagen Inc., Chatsworth, Calif.
19. **Renna, M. C., N. Najimudin, L. R. Winik, and S. A. Zahler.** 1993. Regulation of the *Bacillus subtilis alsS*, *alsD*, and *alsR* genes involved in post-exponential-phase production of acetoin. J. Bacteriol. **175:**3863–3875.
20. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
21. **Shine, J., and L. Dalgarno.** 1975. Determinants of cistron specificity in bacterial ribosomes. Nature (London) **254:**34–38.
22. **Spizizen, J.** 1958. Transformation of biochemically deficient strains of *Bacillus subtilis* by deoxyribonucleate. Proc. Natl. Acad. Sci. USA **44:**407–408.
23. **Tazuya, K., M. Morisaki, K. Yamada, H. Kumaoka, and K. Saiki.** 1987. Biosynthesis of thiamin. Different biosynthetic routes of the thiazole moiety of thiamin in aerobic organisms and anaerobic organisms. Biochem. Int. **14:**153–160.
24. **Vander Horn, P. B., A. D. Backstrom, V. Stewart, and T. P. Begley.** 1992. Structural genes for thiamine biosynthetic enzymes (*thiCEFGH*) in *Escherichia coli* K-12. J. Bacteriol. **175:**982–992.
25. **Vandeyar, M. A., and S. T. Zahler.** 1986. Chromosomal insertions of Tn*917* in *Bacillus subtilis*. J. Bacteriol. **167:**530–534.
26. **Wyatt, D. T., M. Lee, and R. E. Hillman.** 1989. Factors affecting a cyanogen bromide-based assay of thiamin. Clin. Chem. **35:**2173–2178.
27. **Yamada, K., M. Morisaki, and H. Kumaoka.** 1983. Different biosynthetic pathways of the pyrimidine moiety of thiamin in prokaryotes and eukaryotes. Biochim. Biophys. Acta **756:**41–48.
28. **Youngman, P.** 1990. Use of transposons and integrational vectors for mutagenesis and construction of gene fusion in *Bacillus* species, p. 221–266. *In* C. R. Harwood and S. M. Cutting (ed.), Molecular biological methods for *Bacillus*. John Wiley and Sons, Chichester, United Kingdom.
29. **Zahler, S. A. (Cornell University).** 1994. Personal communication.
30. **Zhang, Y.** 1995. Unpublished data.
31. **Zurlinden, A., and M. E. Schweingruber.** 1994. Cloning, nucleotide sequence, and regulation of *Schizosaccharomyces pombe thi4*, a thiamine biosynthetic gene. J. Bacteriol. **176:**6631–6635.