JOURNAL OF BACTERIOLOGY, Feb. 1993, p. 982–992
0021-9193/93/040982-11$02.00/0

Vol. 175, No. 4

Structural Genes for Thiamine Biosynthetic Enzymes (*thiCEFGH*) in *Escherichia coli* K-12

PETER B. VANDER HORN,[1] ALLYSON D. BACKSTROM,[1] VALLEY STEWART,[2] AND TADHG P. BEGLEY[1]*

Department of Chemistry[1] and Sections of Microbiology and Genetics & Development,[2] Cornell University, Ithaca, New York 14853

Received 21 January 1992/Accepted 7 December 1992

***Escherichia coli* K-12 synthesizes thiamine pyrophosphate (vitamin B_1) de novo. Two precursors [4-methyl-5-(β-hydroxyethyl)thiazole monophosphate and 4-amino-5-hydroxymethyl-2-methylpyrimidine pyrophosphate] are coupled to form thiamine monophosphate, which is then phosphorylated to make thiamine pyrophosphate. Previous studies have identified two classes of *thi* mutations, clustered at 90 min on the genetic map, which result in requirements for the thiazole or the hydroxymethylpyrimidine. We report here our initial molecular genetic analysis of the *thi* cluster. We cloned the *thi* cluster genes and examined their organization, structure, and function by a combination of phenotypic testing, complementation analysis, polypeptide expression, and DNA sequencing. We found five tightly linked genes, designated *thiCEFGH*. The *thiC* gene product is required for the synthesis of the hydroxymethylpyrimidine. The *thiE*, *thiF*, *thiG*, and *thiH* gene products are required for synthesis of the thiazole. These mutants did not respond to 1-deoxy-D-*threo*-2-pentulose, indicating that they are blocked in the conversion of this precursor compound to the thiazole itself.**

Thiamine pyrophosphate (THI-PP) is a cofactor for several enzymes, including transketolase, pyruvate dehydrogenase, and α-ketoglutarate dehydrogenase (17). Despite the wealth of information on THI-PP-dependent enzymes, the synthesis of THI-PP in *Escherichia coli* and *Salmonella typhimurium* is poorly understood (Fig. 1; reviewed in references 5 and 47). Thiamine monophosphate (THI-P) is formed by the coupling of two precursors, 4-methyl-5-(β-hydroxyethyl)thiazole monophosphate (THZ-P) and 4-amino-5-hydroxymethyl-2-methylpyrimidine pyrophosphate (HMP-PP). 4-Methyl-5-(β-hydroxyethyl)thiazole (THZ) is synthesized from cysteine, tyrosine, and 1-deoxy-D-*threo*-2-pentulose (8, 10, 13). 4-Amino-5-hydroxymethyl-2-methylpyrimidine phosphate (HMP-P) is synthesized from 5-aminoimidazole ribonucleotide, an intermediate in purine biosynthesis (12, 14). In *E. coli*, five kinases involved in THI-PP biosynthesis have been identified (5, 19, 20, 22, 24, 29–31): HMP kinase (*thiN* gene product), HMP-P kinase (*thiD* gene product), THZ kinase (*thiM* gene product), thiamine kinase (*thiK* gene product), and THI-P kinase (*thiL* gene product). In addition, the coupling enzyme, thiamine phosphate pyrophosphorylase, has been purified from *Saccharomyces cerevisiae* and partially purified from *E. coli* (25, 26).

The genes involved in THI-PP synthesis are spread throughout the *E. coli* genome, and their regulation is poorly understood. The kinase-encoding genes, *thiL* (10 min), *thiK* (25 min), *thiM*, *thiN*, and *thiD* (approximately 46 min), have been located on the *E. coli* genetic map (19, 20, 29, 30). The *nuvC* gene, which maps in the vicinity of 42 to 46 min, is required for synthesis of 4-thiouridine in tRNA (37). Either *nuvC* or a closely linked gene (4, 37) is also required for THZ synthesis. All other known mutations leading to thiamine auxotrophy (*thi*) map at the *thi* cluster, located at 90 min and loosely linked to *argG* (2). All of the kinase structural genes have been cloned (15, 30).

Three phenotypic classes of *thi* cluster mutants have been described: THZ responsive (Thz^- phenotype), HMP responsive (Hmp^- phenotype), and THI requiring (Thi^- phenotype). Analysis of constitutive mutants indicates that the *thi* cluster genes are coordinately regulated. This observation, coupled with their tight linkage, has suggested that the *thi* cluster genes are organized in an operon (22, 24).

We wish to further understand the enzymology and genetics of THI-PP biosynthesis in *E. coli*. We report here our initial molecular genetic analysis of the *thi* cluster. We isolated a 6.6-kb molecular clone that complemented all available *thi* cluster mutations. Our combined complementation, protein expression, and DNA sequence analysis revealed five genes, which we designate *thiCEFGH*. The *thiE*, *thiF*, *thiG*, and *thiH* mutants were Thz^-, and the *thiC* mutants were Hmp^-.

MATERIALS AND METHODS

Nomenclature. Previously, *thi* cluster mutations have been designated *thiA* (Thz^-), *thiB* (Thi^-), and *thiC* (Hmp^-). We found four genes in this cluster whose products are required for THZ synthesis, so we have abandoned the name *thiA* and designate these genes *thiEFGH*.

Bacterial strains and plasmids. *E. coli* K-12 strains and plasmids used in this study are listed in Table 1. *thi* mutations were transduced into strain VJS1390 by selection for $argE^+$, followed by scoring for the Thi phenotype. *arg* and *thi* were approximately 5% linked. The *thiF451*::Km and *thiG455*::Km mutations were constructed by inserting kanamycin cassettes (1) into the *Eco*RV and *Sal*I sites, respectively, of *thi* clones. These insertions were crossed to the *E. coli* chromosome by recombination in the *recBC sbcBC* strain VJS2889, as described previously (42, 45). Genetic crosses were performed via bacteriophage P1 *kc*-mediated transduction (28).

Media. Defined and complex media for genetic manipulations were used as described previously (9, 28). Defined media were supplemented with 0.2% glucose and 100 μM

* Corresponding author.

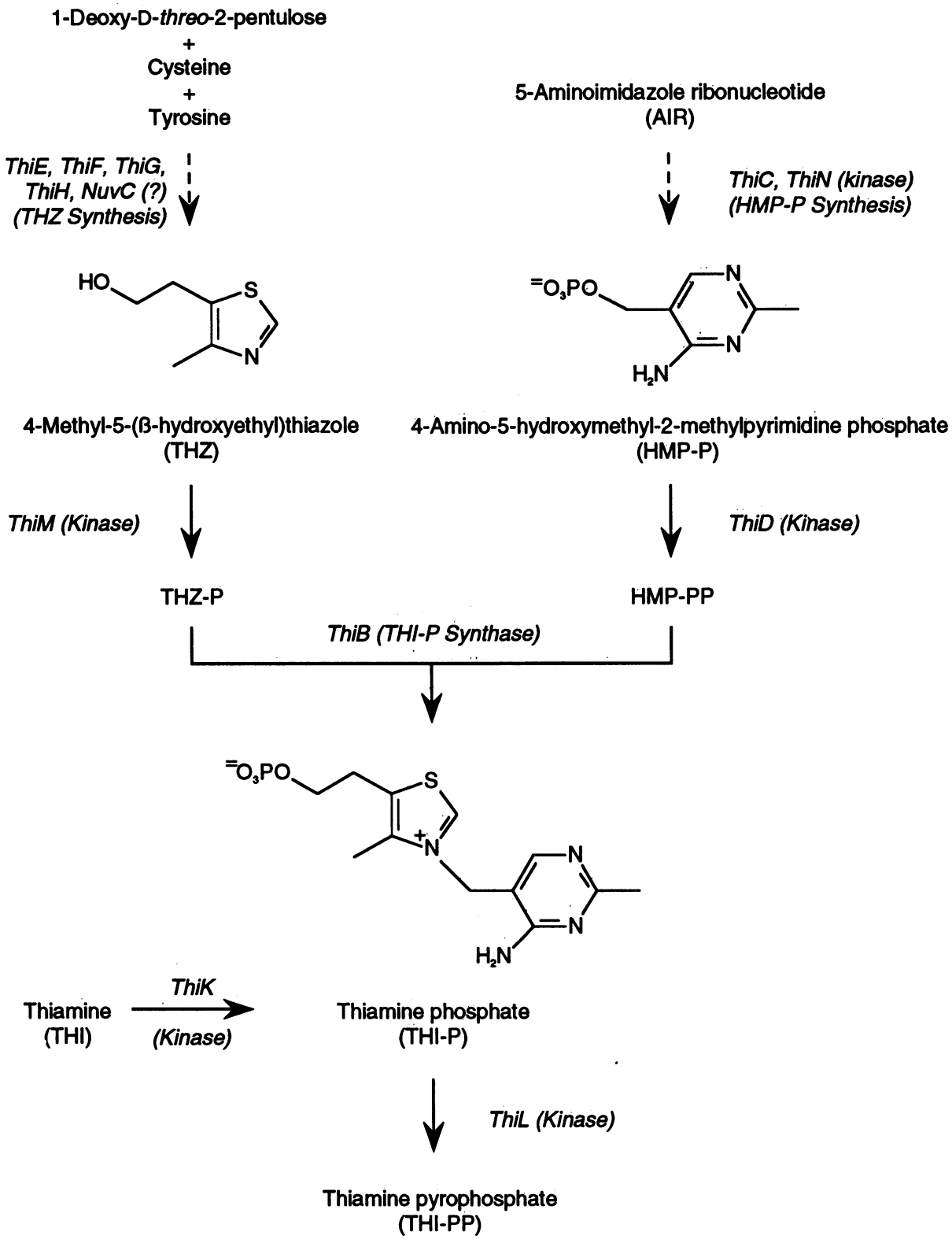

FIG. 1. Pathway for THI-PP biosynthesis in *E. coli*.

TABLE 1. Strains, plasmids, and phages

Strain, plasmid, or phage	Genotype	Source or reference
E. coli strains[a]		
CAG12185	*argE86*::Tn*10*	40
CAG18500	*thiC39*::Tn*10*	40
GW4212	As JC7623, but *recA938*::*cat*	45
H642	*thiE1 gal-3 lacY1 malT1 mtl-2 purH47 rpsL117 supE44 xyl-7*	CGSC[b]
JC7623	*recB21 recC22 sbcB15 sbcC201 argE3 his-4 leuB6 proA-2 thr-1 ara-14 galK2 lacY1 mtl-1 xyl-5 thiE1 rpsL31 supE44 tsx-33*	45
JM83	*ara* Δ(*lac-proAB*)*X111 rpsL* ϕ80d Δ(*lacZ*)*M15*	46
JM109	*endA1 gyrA96 hsdR17* Δ(*lac-proAB*)*X111 recA1 relA1 supE44 thiE1* F′ *lacI*^{q} Δ(*lacZ*)*M15 proAB*$^+$ *traD36*	46
KG33	Hfr P4X *thiH32 relA1 spoT1*	CGSC
KG1673	Hfr P4X *thi-33 relA1 spoT1*	CGSC
KG6593	Hfr P4X *thiC34 relA1 spoT1*	CGSC
OMP-1	*thiC43 argG6 hisG1 leuB6 metB1 trp-31 gal-6 galP63 lacY1 malT1 mtl-2 xyl-7 cpxB11 fhuA2 rfbD1 rpsL104 supE44*	CGSC
RK4353	*araD139* Δ(*argF-lac*)*U169 flhD5301 gyrA219 non-9 rpsL150 ptsF25 relA1 deoC1*	Stewart collection
VJS482	Δ(*argF-lac*)*U169 galK2 galT22 hsdR514 metB1 recA56 supE44 supF58 trpR55*	42
VJS533	As JM83, but *recA56*	42
VJS773	*mal*::Mu cts *recA938*::*cat*/pEG5005	Stewart collection
VJS803	As JC7623, but Δ(*argF-lac*)*U169* Δ(*trpEA*)2	42
VJS1390	As RK4353, but *argE68*::Tn*10*	CAG12185 × RK4353
VJS1391	As RK4353, but *thiE1*	H642 × VJS1390
VJS1392	As RK4353, but *thiH32*	KG33 × VJS1390
VJS1393	As RK4353, but *thiC39*::Tn*10*	CAG18500 × RK4353
VJS1400	As RK4353, but *thiC39*::Tn*10 recA938*::*cat*	GW4212 × VJS1393
VJS1737	As RK4353, but *thi-33*	KG1673 × VJS1390
VJS1740	As RK4353, but *thi-33 recA938*::*cat*	GW4212 × VJS1737
VJS1742	As RK4353, but *thiE1 recA938*::*cat*	GW4212 × VJS1391
VJS2801	As RK4353, but *thiH32 recA938*::*cat*	GW4212 × VJS1392
VJS2889	As VJS803, but *thi*$^+$ *arg*$^+$	This work
VJS2895	As RK4353, but *thiF451*::Km	This work
VJS2897	As RK4353, but *thiG455*::Km	This work
Plasmids		
pEG5005	Ap^r Km^r	16
pGEM-3Zf(+)	Ap^r, T7 ϕ10 promoter	Promega
pGP-1	Km^r, T7 gene *1* (RNA polymerase)	43
pTHI4	Ap^r, *thiE*$^+$*F*$^+$*G*$^+$ (*Nru*I subclone in pUC19; expressed orientation)[c]	This work
pUC19	Ap^r	46
pVJS709	Ap^r, *thiC*$^+$*E*$^+$*F*$^+$*G*$^+$*H*$^+$ in pEG5005	This work
pVJS715	Ap^r, *thiC*$^+$*E*$^+$*F*$^+$*G*$^+$*H*$^+$ in *Pst*I site of pGEM-3Zf(+) (nonexpressed orientation)[d]	This work
pVJS716	Ap^r, *thiC*$^+$*E*$^+$*F*$^+$*G*$^+$*H*$^+$ in *Pst*I site of pGEM-3Zf(+) (expressed orientation)[d]	This work
pVJS717	As pVJS716, but Δ(*Hin*dIII-*Pst*I)	This work
pVJS719	As pVJS717, but Δ*thiC* (*Bgl*II reduction)	This work
pVJS720	As pVJS716, but Δ*thiC* (*Eco*RI reduction)	This work
pVJS722	As pVJS716, but Δ*thiFGH* (*Eco*RV reduction)	This work
pVJS723	As pVJS715, but Δ*thiFGH* (*Eco*RV reduction)	This work
pVJS725	As pVJS715, but Δ*thiCEFGH* (*Bgl*II reduction)	This work
pVJS730	As pVJS717, but Δ*thiCEFG* (*Sal*I reduction)	This work
pVJS737	As pVJS719, but Δ*thiGH* (*Sal*I reduction)	This work
pVJS750	As pVJS719, but Δ*thiF* (*Eco*RV reduction)	This work
pVJS751	As pVJS737, but Δ*thiEF* (*Eco*RV reduction)	This work
pVJS765	As pVJS717, but Δ*thiCEF.* (*Eco*RV reduction)	This work
pVJS769	As pVJS719, but *thiG455*::Km (insertion at *Sal*I)	This work
Phages		
M13mp19		46
M13mp19.S4	As M13mp19; contains leftward *Pst*I-*Sal*I fragment (noncoding strand)	This work
M13mp19.S6	As M13mp19; contains leftward *Pst*I-*Sal*I fragment (coding strand)	This work
M13mp19.B8	As M13mp19; contains leftward *Pst*I-*Bgl*II fragment (coding strand)	This work

[a] All strains are λ$^-$ and F$^-$ except as indicated.
[b] CGSC, Coli Genetic Stock Center, courtesy of B. J. Bachmann.
[c] "Expressed" orientation of the insert with respect to the *lac* promoter.
[d] "Expressed" and "nonexpressed" orientation of the insert with respect to the T7 RNA polymerase promoter.

THI, THZ, or HMP, as indicated. Ampicillin, chloramphenicol, kanamycin, and tetracycline were used at 200, 75, 200, and 20 μg/ml, respectively. Cultures were aerated at 37°C except as otherwise indicated. Agar and dehydrated media were from Difco Laboratories (Detroit, Mich.). HMP was a gift from Hoffmann-La Roche (Basel, Switzerland). 1-Deoxy-D-*threo*-2-pentulose was synthesized in this laboratory (3) by a method different from that of David et al. (8). Other components were from Sigma Chemical Co. (St. Louis, Mo.).

Phenotypic analysis. All *thi* mutants were tested for growth with 1-deoxy-D-*threo*-2-pentulose, THZ, HMP, and THI. Strains were grown in THI-limiting defined medium and plated as a lawn on defined medium. Substrates (1 μl of a 100 μM solution) were spotted on the plates. Growth around or between the test compounds was determined after overnight incubation.

Molecular cloning. Standard methods were used for DNA restriction endonuclease digestion, ligation, and transformation of DNA (9, 27). Single-stranded plasmid DNA was isolated by the helper phage method specified by Promega Corp. (Madison, Wis.). Restriction enzymes and T4 DNA ligase were from New England BioLabs, Inc. (Beverly, Mass.) and United States Biochemical Corp. (Cleveland, Ohio). Strains JM109 and VJS533 were used as recipients for transformation during plasmid constructions.

Detection of plasmid-encoded proteins. Plasmid-encoded proteins were detected in strain VJS482 by using an in vivo T7 expression system (43). This method involved cloning genes of interest downstream from the phage T7 φ10 promoter in pGEM-3Zf(+) and then transforming them into strain VJS482 containing pGP1-2. Plasmid pGP1-2 carries the gene for T7-RNA polymerase (gene *1*) under the control of a heat-inducible promoter and is compatible with pGEM-3Zf(+). Cultures were grown at 30°C with shaking in TY medium containing ampicillin and kanamycin and then were starved for methionine by washing cells twice in defined medium. After heat induction of T7 RNA polymerase at 42°C, rifampin was added to inhibit host RNA polymerase. Subsequent addition of ^{35}S-methionine resulted in selective labeling of plasmid-encoded gene products. Samples and prestained molecular weight markers were electrophoresed on Laemmli gels (18), fixed, treated with sodium salicylate (6), dried, and autoradiographed.

DNA sequencing. DNA sequences were determined from single-stranded templates by the dideoxynucleotide chain termination method (38) with modified T7 DNA polymerase (44) and [α-^{32}P]dATP labelling. T7 DNA polymerase (Sequenase) and [α-^{32}P]dATP (3,000 Ci mmol^{-1}) were purchased from United States Biochemical Corp. and Amersham Corp. (Arlington Heights, Ill.), respectively. Oligonucleotide primers were synthesized at the Cornell Biotechnology Analytical and Synthetic Facility. All sequencing reactions were conducted as specified by the manufacturer except for two modifications: the Sequenase enzyme was diluted to a final activity of 0.2 U ml^{-1} in the labeling reaction, and the termination reactions were carried out at 42°C. The nucleotide dITP was used in place of dGTP to resolve band compressions. Both DNA strands were sequenced in their entirety.

Computer-assisted sequence analysis. The program DNA INSPECTOR II (Textco, West Lebanon, N.H.) was used to compile and analyze DNA sequence data. Further analysis employed the Genetics Computer Group program suite (11) running on the Cornell University Biotechnology Program's BIOVAX computer. The program FASTA (35) was used to compare the *thi* sequence with sequences in the Genbank data base. The algorithm of Needleman and Wunsch (32) was used to align the ThiF and MoeB (ChlN) polypeptides; gap weight was 3.0, and length weight was 0.1.

Nucleotide sequence accession number. The GenBank nucleotide sequence accession number for the *E. coli thiC*, *thiE*, *thiF*, *thiG*, and *thiH* genes is M88701.

TABLE 2. Phenotypes of *thi* mutants

Mutant	Growth[a] in response to:					Phenotype
	Pentulose	THZ	HMP	THZ + HMP	THI	
thiC34	−	−	++	++	++	Hmp$^-$
thiC39::Tn*10*	−	−	±	++	++	Hmp$^-$
thiC34	−	−	++	++	++	Hmp$^-$
thiE1	−	++	−	++	++	Thz$^-$
thiF451::Km	−	++	−	++	++	Thz$^-$
thiG455::Km	−	++	−	++	++	Thz$^-$
thiH32	−	++	−	++	++	Thz$^-$
thi-33	−	±	−	++	++	Thz$^-$

[a] Growth on minimal medium in response to the indicated compound, determined as described in Materials and Methods. −, no growth; ++, robust growth; ±, weak growth.

RESULTS

Phenotypes of *thi* mutants. We determined the growth response of *thi* mutants to THI and its precursors as described in Materials and Methods. Results are summarized in Table 2. None of the strains responded to 1-deoxy-D-*threo*-2-pentulose, a precursor of THZ (Fig. 1), indicating that none of the mutations affect synthesis of this compound. It has been established that *E. coli* is permeable to this pentulose, as it has been used in isotopic labelling studies of THZ biosynthesis (8). The *thiE1*, *thiF451*::Km, *thiG455*::Km, *thiH32*, and *thi-33* mutants were Thz$^-$, and the *thiC34* and *thiC43* mutants were Hmp$^-$. The *thi-33* mutation had been described previously as Thi$^-$ (23). The *thiC39*::Tn*10* strain responded relatively poorly to HMP, but grew well with a combination of HMP and THZ. Complementation analysis (see below) suggests that this insertion is in *thiC*, the first gene in the *thi* cluster. We believe that the partial THZ requirement exhibited by the *thiC39*::Tn*10* strains may be caused by decreased expression of the downstream THZ biosynthetic genes due to polarity. Thus, we classify the phenotype caused by this insertion as Hmp$^-$.

Molecular cloning of the *thi* cluster. We used the in vivo cloning vector pEG5005 (16) to isolate several clones that complemented the Thi$^-$ phenotype of a *thiE1* strain. DNA fragments were subcloned into pGEM3Zf(+). One pair of subclones, pVJS715 and pVJS716, carried an 8-kb *Pst*I fragment in both orientations with respect to the vector (Fig. 2). These subclones complemented all available *thi* mutations and were used for all subsequent manipulations. The left and right ends of the insert (as drawn in Fig. 2) include 1.5 and 0.1 kb of bacteriophage Mu DNA derived from the pEG5005 vector. Thus, these clones contain approximately 6.6 kb of *E. coli* DNA. Several insertion and deletion subclones of both pVJS715 and pVJS716 were constructed for complementation and protein expression studies. The most relevant of these clones are described in Table 1 and diagrammed in Fig. 2.

Complementation analysis. Deletion and insertion subclones were transformed into *recA* derivatives of several *thi* mutants, and the resulting transformants were tested for the

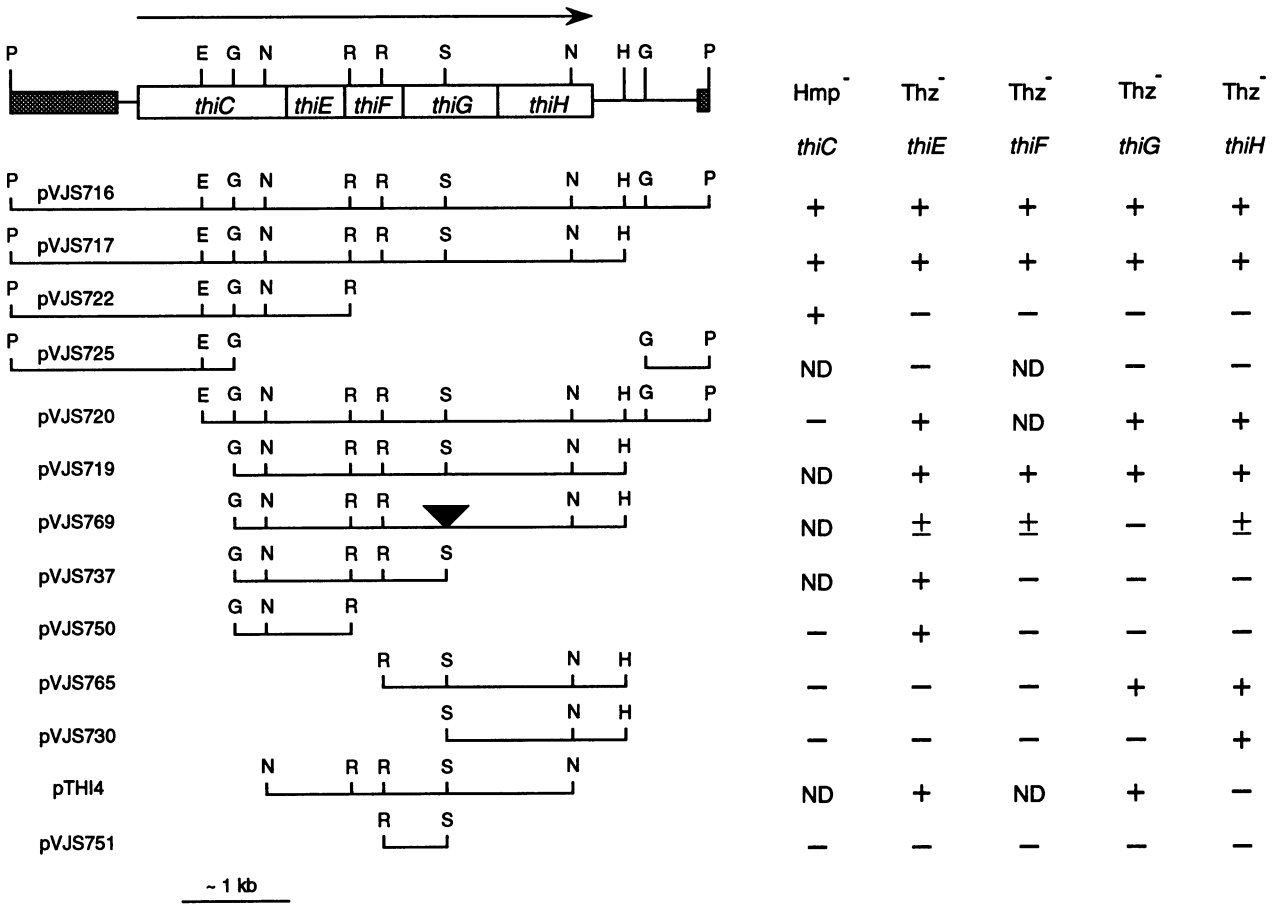

FIG. 2. Complementation of *thi* mutations. The restriction map of the *thi* cluster is shown at top left, approximately to scale. Coding sequences are indicated by open boxes, and DNA derived from plasmid pEG5005 is indicated by cross-hatched boxes. Restriction endonuclease sites are as follows: E, *Eco*RI; G, *Bgl*II; H, *Hin*dIII; N, *Nru*I; P, *Pst*I; R, *Eco*RV; S, *Sal*I. The rightward-pointing arrow indicates the direction of transcription. Subclones are indicated below the restriction map; vector sequences are not shown. The complementation pattern of each subclone is shown on the right. The *thiC39*::Tn*10* mutation was complemented only by plasmids pVJS715, pVJS716, and pVJS717. The data shown for *thiC* reflect complementation of *thiC34* and *thiC43*. The data shown for *thiG* reflect complementation of *thi-33* except for pVJS765, which complemented *thiG455*::Km but not *thi-33*. ND, not determined.

ability to grow on defined medium in the absence of THI. Results are summarized in Fig. 2. The *thiC39*::Tn*10* mutant was complemented only by the relatively large clones, pVJS715, pVJS716, and pVJS717. This complementation pattern is consistent with the idea that *thiC39*::Tn*10* is an insertion in the most proximal gene of the (presumed) *thi* operon. The *thiC34* and *thiC43* mutants (Hmp$^-$) were complemented by pVJS722, indicating that *thiC* does not extend past the leftmost *Eco*RV site. They were not complemented by pVJS720, indicating that the *Eco*RI site is within *thiC*.

The *thiE1* mutant (Thz$^-$) was complemented by pVJS750 and pTHI4, which places *thiE* between the leftmost *Nru*I site and the leftmost *Eco*RV site. This also indicates that *thiC* is upstream of *thiE*. Confoundingly, pVJS722 did not complement *thiE1*, even though pVJS750 contains a smaller subfragment. We do not understand the reason for this, but we believe that the results with pVJS719, pVJS720, pVJS737, pVJS750, and pTHI4 establish the location of *thiE*.

By definition, *thiF* includes the two *Eco*RV sites, because we constructed the *thiF451*::Km mutant (Thz$^-$) by allele exchange. This insertion is apparently polar, because the *thiF451*::Km mutation was only complemented by plasmids carrying *thiF*$^+$, *thiG*$^+$, and *thiH*$^+$ (pVJS717 and pVJS719, but not pTHI4).

By definition, *thiG* spans the *Sal*I site, because we constructed the *thiG455*::Km mutant (Thi$^-$) by allele exchange. Consistent with this, *thiG455*::Km was complemented by pVJS765, but not by pVJS730. Interestingly, pTHI4 complemented *thiG455*::Km, even though it does not contain the entire *thiH* gene. This indicates that *thiG455*::Km-containing strains, which carry a (presumably) polar insertion upstream of *thiH*, still express sufficient *thiH* to synthesize thiazole. The *thi-33* mutant (Thz$^-$) was not complemented by pVJS765, but otherwise shared the same complementation pattern as *thiG455*::Km. We suspect that *thi-33* may be a deletion or other rearrangement that also affects *thiF*, but we have not yet directly tested this idea.

The *thiH32* mutant (Thz$^-$) was complemented by pVJS730, but not by pTHI4. This establishes that the rightmost *Nru*I site is within *thiH*. Taken together, these complementation data establish the gene order as *thiCEFGH*.

***thiCEFGH*-encoded polypeptides.** We used an in vivo T7 expression system to detect the proteins encoded by *thiCE FGH*, as described in Materials and Methods. Figure 3

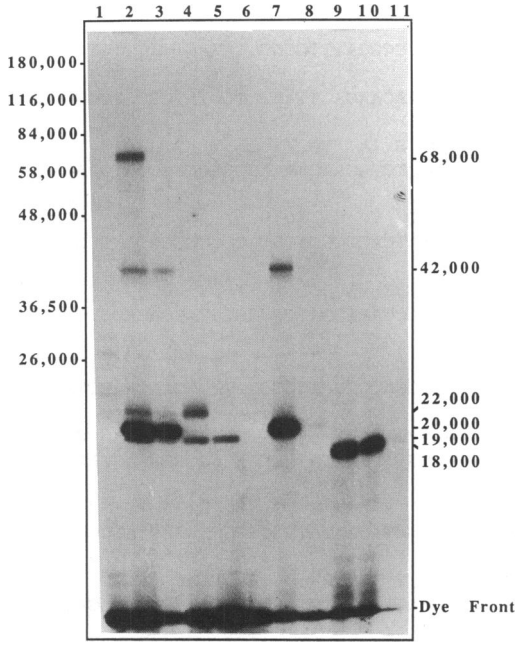

FIG. 3. Polypeptides encoded by *thiCEFGH*. Polypeptides were detected with a T7 expression system (see Materials and Methods). Samples were electrophoresed on a 10% polyacrylamide Laemmli gel. Lanes: 1, pGEM-3Zf(+) (vector); 2, pVJS716; 3, pVJS719; 4, pVJS737; 5, pVJS750; 6, pVJS751; 7, pVJS765; 8, pVJS730; 9, pVJS715; 10, pVJS725; 11, pVJS723. See the legend to Fig. 2 for descriptions of plasmids. Molecular mass markers were triosephosphate isomerase (26 kDa), lactic hydrogenase (36.5 kDa), fumarase (48.5 kDa), pyruvate kinase (58 kDa), fructose-6-phosphate kinase (84 kDa), β-galactosidase (116 kDa), and α_2-macroglobulin (180 kDa).

shows the proteins encoded by pVJS715 and pVJS716. Plasmid pVJS716 encoded five proteins with estimated molecular masses of 68, 42, 21, 20, and 19 kDa. The insert in pVJS716 is oriented such that the T7 promoter initiates transcription from the left end, as drawn in Fig. 2. Plasmid pVJS715, which has the insert in the opposite direction relative to the T7 promoter, produced a single polypeptide with an estimated molecular mass of 18 kDa.

To determine the order and location of the five genes in the *thiCEFGH* region, we examined the polypeptides expressed from various subclones of pVJS716. Plasmid pVJS720 (Δ*thiC*) did not express the 68-kDa polypeptide, demonstrating that ThiC has an approximate molecular mass of 68 kDa. Plasmid pVJS750 (*thiE*$^+$) expressed the 19-kDa polypeptide, demonstrating that ThiE has an approximate molecular mass of 19 kDa. Plasmid pVJS737 (*thiE*$^+$ *thiF*$^+$) expressed both the 21- and the 19-kDa polypeptides, demonstrating that ThiF has an approximate molecular mass of 21 kDa. Plasmid pVJS751, in which both *thiE* and *thiG* are truncated, did not express any polypeptides.

Plasmid pVJS765 (*thiG*$^+$ *thiH*$^+$) expressed 42- and 20-kDa polypeptides. In contrast, pVJS730 (*thiH*$^+$) did not express either the 42- or the 20-kDa polypeptide. Expression of the downstream *thiH* gene product required an intact *thiG* gene in this experiment. We do not know the reason for this, but one possibility is that these genes are translationally coupled (34). We have not tested this idea experimentally. Overexposing the autoradiogram of polypeptides expressed from pVJS769 (*thiG*::Km *thiH*$^+$) revealed a band at 42 kDa but not one at 20 kDa (data not shown). This suggests that ThiG and ThiH have approximate molecular masses of 20 and 42 kDa, respectively.

To determine the location of the gene that encodes the 18-kDa polypeptide, expressed in the opposite direction, deletion derivatives of plasmid pVJS715 were examined. Plasmid pVJS725 encoded the 18-kDa polypeptide, whereas pVJS723 (with the same insert as pVJS722) did not. This places the coding region for the 18-kDa polypeptide between the rightmost *Bgl*II and *Pst*I sites (Fig. 2). This gene will be discussed later. It is not necessary for complementation of any of the *thi* mutants tested (Fig. 2, see pVJS717).

DNA sequence analysis. The sequence of the *thi*-complementing DNA is shown in Fig. 4. The DNA sequence predicts five open reading frames corresponding to *thiCE FGH*. With the exception of ThiF and ThiG, there is good agreement between the polypeptide masses predicted by the DNA sequence and the masses determined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis: ThiC, 70 (sequence) versus 68 (observed) kDa; ThiE, 23 (sequence) versus 19 (observed) kDa; ThiF, 27 (sequence) versus 21 (observed) kDa; ThiG, 34 (sequence) versus 20 (observed) kDa; and ThiH, 43 (sequence) versus 41 (observed) kDa. At this time we have no explanation for the differences between predicted and observed masses of ThiF and ThiG.

All of the cistrons contain overlapping start and stop codons with the exception of the *thiCE* intercistronic region. Each open reading frame begins with an AUG start codon preceded by a Shine-Dalgarno sequence (39). Plasmids pVJS715 and pVJS716 both complemented all *thi* mutations tested, suggesting that the cloned DNA includes a *thi* promoter upstream of *thiC*.

Data base searches revealed that ThiF shares sequence similarity with the MoeB (ChlN) polypeptide of *E. coli* (accession number M21151). MoeB is a protein involved in molybdopterin synthesis (21, 33). The MoeB sequence is 44% identical to ThiF (Fig. 5). Identity was distributed throughout the entire protein sequence. ThiC, ThiE, ThiG, and ThiH were not found to be significantly similar to other sequences in the data bases.

Physical map location of *thiCEFGH*. The *thi* cluster maps at 90 min on the *E. coli* genetic map, between *rpoBC* and *purHD* (2). We used the program FASTA to search the GenBank data base for sequences similar to the region downstream of *thiH* (Fig. 4). The clone contains a 822-bp overlap with the 3′ end of the *rpoBC* operon (accession number V00339) which encodes the β and β′ subunits of RNA polymerase (41). Figure 4 shows 155 bp of the *rpoBC* sequence overlap which ends at the underlined *Hin*dIII site. This establishes the location of *thiCEFGH* as immediately adjacent to *rpoBC* and transcribed in the opposite orientation, counterclockwise with respect to the genetic map.

DISCUSSION

Previous studies identified at least two *thi* genes clustered at 90 min on the genetic map, mutations in which led to requirements for THZ ("*thiA*") and HMP (*thiC* [22–24]). These genes are tightly linked, coordinately expressed, and probably organized as an operon (22, 24). However, the fine structure and regulation of the *thi* genes have not been explored. We report here our initial molecular examination of the *thi* cluster. A combination of phenotypic testing, complementation analysis, polypeptide expression, and DNA sequencing revealed five genes, *thiCEFGH* (Fig. 2).

Four genes in this cluster, *thiE*, *thiF*, *thiG*, and *thiH*, were

```
      taaaaagcagtaattaatacatctgtttcatttgaagcgcgaaagctaaagttttcgcatttatcgtgaaacgctttcgcgtttttcgtgcgccgcttca

   1  GCCGCCCGCGTCAAACATCCTGCTTGAGTTCTGCGCTGTTAACGCGTAATTTACATTCAATGCCCCATTTGCGGGGCTAATTTCTTGTCGGAGTGCCTTA

 101  ACTGGCTGAGACCGTTTATTCGGGATCCGCGGAACCTGATCAGGCTAATACCTGCGAAGGGAACAAGAGTTAATCTGCTATCGCATCGCCCCTGCGGCGA

                                                                thiC M  S  A  T  K  L  T  R  R  E  Q  R  A  R
 201  TCGTCTCTTGCTTCATCCGTCGTCTGACAAGCCACGTCCTTAACTTTTTGGAATGAGCTATGTCTGCAACAAAACTGACCCGCCGCGAACAACGCGCCCG

       A  Q  H  F  I  D  T  L  E  G  T  A  F  P  N  S  K  R  I  Y  I  T  G  T  H  P  G  V  R  V  P  M  R
 301  GGCCCAACATTTTATCGACACCCTGGAAGGCACCGCCTTTCCCAACTCAAAACGCATTTATATCACTGGCACACACCCCGGCGTGCGCGTGCCGATGCGT

       E  I  Q  L  S  P  T  L  I  G  G  S  K  E  Q  P  Q  Y  E  E  N  E  A  I  P  V  Y  D  T  S  G  P  Y  G
 401  GAGATCCAGCTTAGCCCGACGCTAATTGGCGGTAGCAAAGAACAGCCGCAGTACGAAGAAAACGAAGCGATTCCGGTCTACGACACCTCCGGCCCGTATG

         D  P  Q  I  A  I  N  V  Q  Q  G  L  A  K  L  R  Q  P  W  I  D  A  R  G  D  T  E  E  L  T  V  R  S
 501  GTGATCCGCAGATTGCCATTAACGTGCAGCAAGGGCTGGCAAAACTACGCCAGCCGTGGATCGATGCGCGCGGCGATACCGAAGAACTTACCGTGCGCAG

       S  D  Y  T  K  A  R  L  A  D  D  G  L  D  E  L  R  F  S  G  V  L  T  P  K  R  A  K  A  G  R  R  V
 601  TTCCGATTACACTAAAGCGCGGCTGGCAGATGATGGCCTCGACGAACTGCGTTTTAGCGGCGTACTAACACCAAAACGCGCCAAAGCAGGACGCCGTGTC

       T  Q  L  H  Y  A  R  Q  G  I  I  T  P  E  M  E  F  I  A  I  R  E  N  M  G  R  E  R  I  R  S  E  V  L
 701  ACCCAACTGCACTACGCCCGCCAGGGCATCATCACGCCGGAAATGGAATTCATCGCCATCCGCGAGAATATGGGCCGCGAGCGCATCCGTAGCGAGGTTT

         R  H  Q  H  P  G  M  S  F  G  A  H  L  P  E  N  I  T  A  E  F  V  R  D  E  V  A  A  G  R  A  I  I
 801  TACGCCACCAGCATCCGGGAATGAGCTTTGGCGCACATCTGCCGGAAAATATCACTGCGGAATTTGTCCGTGATGAAGTTGCTGCCGGACGTGCGATTAT

       P  A  N  I  N  H  P  E  S  E  P  M  I  I  G  R  N  F  L  V  K  V  N  A  N  I  G  N  S  A  V  T  S
 901  CCCGGCCAACATTAATCATCCGGAATCGGAGCCGATGATTATTGGTCGCAATTTCCTGGTAAAAGTTAACGCCAATATCGGCAACTCGGCGGTCACCTCT

       S  I  E  E  E  V  E  K  L  V  W  S  T  R  W  G  A  D  T  V  M  D  L  S  T  G  R  Y  I  H  E  T  R  E
1001  TCCATCGAAGAAGAAGTGGAAAAGCTGGTATGGTCCACGCGCTGGGGAGCGGATACGGTGATGGATCTCTCCACCGGTCGCTATATTCACGAAACCCGCG

         W  I  L  R  N  S  P  V  P  I  G  T  V  P  I  Y  Q  A  L  E  K  V  N  G  I  A  E  D  L  T  W  E  A
1101  AGTGGATTTTGCGTAACAGCCCGGTGCCGATCGGTACAGTGCCGATCTACCAGGCGCTGGAGAAGGTTAACGGGATCGCCGAAGATCTTACCTGGGAAGC

       F  R  D  T  L  L  E  Q  A  E  Q  G  V  D  Y  F  T  I  H  A  G  V  L  L  R  Y  V  P  M  T  A  K  R
1201  GTTCCGCGACACGCTGCTGGAACAGGCCGAGCAAGGTGTGGATTACTTCACTATCCATGCGGGCGTACTGCTGCGCTATGTGCCGATGACCGCGAAACGC

       L  T  G  I  V  S  R  G  G  S  I  M  A  K  W  C  L  S  H  H  Q  E  N  F  L  Y  Q  H  F  R  E  I  C  E
1301  CTGACCGGTATCGTCTCTCGCGGCGGTTCGATTATGGCGAAATGGTGCCTCTCCCATCATCAGGAAAATTTCCTCTATCAACACTTCCGCGAAATTTGTG

         I  C  A  A  Y  D  V  S  L  S  L  G  D  G  L  R  P  G  S  I  Q  D  A  N  D  E  A  S  F  A  E  L  H
1401  AAATCTGTGCCGCTTATGACGTTTCGCTGTCGCTGGGCGACGGTCTGCGCCCCGGTTCTATTCAGGACGCCAACGATGAAGCGCAGTTTGCCGAGCTGCA

       T  L  G  E  L  T  K  I  A  W  E  Y  D  V  Q  V  M  I  E  G  P  G  H  V  P  M  Q  M  I  R  R  N  M
1501  TACGCTGGGCGAACTGACCAAAATTGCCTGGGAATATGACGTGCAGGTGATGATTGAAGGCCCAGGCCACGTGCCGATGCAGATGATCCGCCGCAATATG

       T  E  E  L  E  H  C  H  E  A  P  F  Y  T  L  G  P  L  T  T  D  I  A  P  G  Y  D  H  F  T  S  G  I  G
1601  ACCGAGGAGTTAGAGCACTGCCACGAAGCGCCGTTTTACACTCTGGGGCCGCTAACTACCGATATTGCGCCGGGCTATGACCACTTCACGTCGGGGATTG

         A  A  M  I  G  W  F  G  C  A  M  L  C  Y  V  T  P  K  E  H  L  G  L  P  N  K  E  D  V  K  Q  G  L
1701  GTGCGGCGATGATTGGCTGGTTTGGCTGCGCGATGCTCTGTTACGTAACGCCAAAAGAGCATCTGGGTCTGCCCAATAAAGAAGATGTTAAGCAGGGGCT

       I  T  Y  K  I  A  A  H  A  A  D  L  A  K  G  H  P  G  A  Q  I  R  D  N  A  M  S  K  A  R  F  E  F
1801  TATCACCTATAAGATTGCTGCCCACGCCGCTGACCTGGCGAAAGGGCATCCGGGCGCGCAAATTCGCGATAACGCCATGTCGAAAGCCCGCTTCGAATTT

       R  W  E  D  Q  F  N  L  A  L  D  P  F  T  A  R  A  Y  H  D  E  T  L  P  Q  E  S  G  K  V  A  H  F  C
1901  CGCTGGGAAGACCAGTTTAATCTGGCCCTCGACCCGTTTACCGCCCGCGCTTATCACGATGAAACCCTGCCGCAAGAGTCAGGTAAAGTCGCCCATTTTT

         S  M  C  G  P  K  F  C  S  M  K  I  S  Q  E  V  R  D  Y  A  A  T  Q  T  I  E  M  G  M  A  D  M  S
2001  GCTCCATGTGTGGGCCGAAATTCTGCTCGATGAAAATCAGCCAGGAAGTGCGTGATTACGCCGCCACGCAAACTATTGAAATGGGAATGGCGGATATGTC

       E  N  F  R  A  R  G  G  R  N  L  P  A  *           thiE M  Y  Q  P  D  F  P  P  V  P  F  R  S  G  L
2101  GGAGAACTTCCGTGCCAGAGGCGGGAGAAATCTACCTGCGTAAGGAGGAAGCGTGATGTATCAGCCTGATTTTCCTCCTGTACCTTTTCGTTCAGGACTG

       Y  P  V  V  D  S  V  Q  W  I  E  R  L  L  D  A  G  V  R  T  L  Q  L  R  I  K  D  R  R  D  E  E  V  E
2201  TACCCGGTGGTGGACAGCGTACAGTGGATCGAACGTCTGTTGGATGCAGGCGTACGTACTCTCCAGCTACGCATCAAAGATCGGCGCGATGAAGAGGTGG
```

FIG. 4. Nucleotide sequence of the *thiCEFGH*. The nucleotide sequence of the noncoding strand is shown. Sequences derived from the bacteriophage Mu cloning vector (pEG5005) are displayed in lowercase letters, and *E. coli* chromosomal sequences are displayed in capital letters. Potential Shine-Dalgarno regions and initiator codons are underlined. Derived amino acid sequences are shown in standard single-letter code. The underlined *Hin*dIII site indicates where the sequence overlap between the *thi* cluster and the 3′ end of *rboBC* begins.

```
        A D V V A A I A L G R R Y N A R L F I N D Y W R L A I K H Q A Y G
2301    AAGCCGATGTCGTGGCGGCAATTGCGCTGGGCCGCCGCTATAACGCGCGATTGTTTATCAACGATTACTGGCGGCTGGCGATCAAGCATCAGGCGTATGG

        V H L G Q E D L Q A T D L N A I R A A G L R L G V S T H D D M E I
2401    CGTCCATTTGGGGCAGGAAGATTTGCAAGCCACCGATCTCAATGCCATCCGCGCGGCAGGCCTGCGGCTGGGCGTTTCGACACATGACGATATGGAAATC

        D V A L A A R P S Y I A L G H V F P T Q T K Q M P S A P Q G L E Q L
2501    GACGTCGCGCTGGCAGCACGCCCCTCTTATATCGCGCTGGGACATGTGTTCCCGACGCAAACCAAACAGATGCCTTCTGCACCGCAGGGGCTGGAACAGC

        A R H V E R L A D Y P T V A I G G I S L A R A P A V I A T G V G S
2601    TGGCACGGCATGTTGAGCGACTGGCGGATTATCCCACCGTGGCGATTGGCGGTATCAGTCTGGCACGCGCGCCTGCGGTGATAGCAACGGGTGTCGGCAG

                                                              thiF M N D R D F
        I A V V S A I T Q A A D W R L A T A Q L L E I A G V G D E *
2701    TATCGCCGTCGTCAGCGCCATTACTCAAGCCGCAGACTGGCGTTTGGCAACGGCACAGTTGCTGGAAATTGCAGGAGTTGGCGATGAATGACCGTGACTT

        M R Y S R Q I L L D D I A L D G Q Q K L L D S Q V L I I G L G G L
2801    TATGCGTTATAGCCGCCAAATCCTGCTCGACGATATCGCTCTGGACGGGCAGCAAAAACTGCTCGACAGCCAGGTGCTGATTATCGGTCTGGGCGGGCTG

        G T P A A L Y L A G A G V G T L V L A D D D D V H L S N L Q R Q I L
2901    GGTACACCTGCTGCGCTGTACCTGGCGGGCGCTGGCGTCGGGACGCTGGTACTGGCAGATGACGACGATGTGCATTTAAGCAATCTGCAACGACAAATCC

        F T T E D I D R P K*S Q V S Q Q R L T Q L N P D I Q L T A L Q Q R
3001    TCTTTACCACTGAAGATATCGATCGCCCGAAATCGCAGGTCAGCCAACAGCGACTGACACAGTTGAATCCCGACATTCAACTGACAGCATTACAACAACG

        L T G E A L K D A V A R A D V V L D C T D N M A T R Q E I N A A C
3101    GTTAACGGGTGAGGCGTTAAAAGATGCGGTTGCACGGGCCGATGTGGTGCTCGACTGTACCGACAATATGGCGACTCGCCAGGAGATTAATGCCGCCTGC

        V A L N T P L I T A S A V G F G G Q L M V L T P P W E Q G C Y R C L
3201    GTGGCACTCAACACGCCGCTTATCACCGCCAGCGCGGTCGGATTTGGCGGTCAGTTGATGGTACTGACGCCGCCCTGGGAGCAGGGGTGTTACCGCTGCC

        W P D N Q E P E R N C R T A G V V G P V V G V M G T L Q A L E A I
3301    TGTGGCCAGATAACCAGGAGCCAGAACGCAACTGCCGCACGGCGGGCGTGGTTGGCCCGGTGGTCGGGGTTATGGGCACTTTGCAGGCACTGGAAGCCAT

        K L L S G I E T P A G E L R L F D G K S S Q C S L A L R R A S G C
3401    TAAGTTATTAAGCGGTATAGAGACACCTGCGGGAGAACTCCGACTGTTCGACGGTAAATCGAGCCAGTGCAGCCTGGCGTTGCGCCGCGCCAGTGGTTGC

                thiG M Q I L F N D Q A M Q C A G R A N C S R T T G A T R P
        P V C G G S N A D P V *
3501    CCGGTATGCGGAGGAAGCAATGCAGATCCTGTTTAACGATCAAGCGATGCAGTGCGGCCGCCGGGCAAACTGTTCACGAACTACTGGAGCAACTCGACCA

        T T S G R G L W R L I S K S S R V S S G R N I S C R M A T R S C F F
3601    ACGACAAGCGGGCGCGGCCTCTGGCGATTAATCAGCAAATCGTCCCGCGTGAGCAGTGGGCGCAACATATCGTGCAGGATGGCGACCAGATCCTGCTTTT

        R L L Q G V E M L R I A D K T F D S H L F T G T G K F A S S Q L M
3701    TCAGGTTATTGCAGGGGGTTGAAATGTTACGTATTGCGGACAAAACGTTTGATTCACATCTGTTTACCGGCACAGGGAAATTCGCTTCTTCACAACTGAT

        V E A I R A S G S Q L V T L A M K R V D L R Q H N D A I L E P L I
3801    GGTGGAGGCGATCCGCGCTTCCGGCAGCCAGCTGGTGACACTGGCGATGAAACGTGTCGACTTGCGCCAGCACAACGACGCTATCCTCGAACCGCTTATC

        A A G V T L L P N T S G A K T A E E A I F A A H L A R E A L G T N W
3901    GCGGCGGGTGTGACCCTGCTGCCAAATACATCCGGGGCGAAAACAGCGGAAGAAGCCATTTTCGCCGCCCATCTGGCTCGTGAAGCGTTAGGCACAAACT

        L K L E I H P D A R W L L P D P I E T L K A A E T L V Q Q G F V V
4001    GGTTAAAATTAGAGATTCACCCTGACGCCCGCTGGCTGTTGCCCGATCCCATCGAAACCCTGAAAGCCGCCGAAACGCTGGTACAACAGGGATTTGTCGT

        L P Y C G A D P V L C K R L E E V G C A A V M P L G A P I G S N Q
4101    GCTGCCTTACTGCGGGGCCGATCCGGTATTGTGTAAACGTCTGGAAGAAGTCGGCTGTGCAGCGGTGATGCCGCTCGGCGCGCCGATTGGCTCGAATCAG

        G L E T R A M L E I I I Q Q A T V P V V V D A G I G V P S H A A Q A
4201    GGACTGGAAACCCGCGCCATGCTGGAGATTATTATCCAGCAGGCCACAGTGCCGGTGGTTGTCGATGCTGGCATCGGCGTTCCCAGCCATGCCGCGCAGG

        L E M G A D A V L V N T A I A V A D D P V N M A K A F R L A V E A
4301    CGCTGGAAATGGGGGCCGACGCGGTGTTAGTGAATACGGCGATTGCCGTCGCGGACGATCCCGTCAACATGGCGAAGGCATTTCGTCTGGCGGTAGAAGC
```

FIG. 4—*Continued.*

required for THZ synthesis. The molecular masses of their products, predicted from the DNA sequence, are 23, 27, 34, and 43 kDa, respectively. An unlinked gene, *nuvC* (or a closely linked gene [4, 37]) is also required for THZ synthesis. None of these mutants grew with 1-deoxy-D-*threo*-2-pentulose, a precursor of THZ (Fig. 1) (3), indicating that their products are involved in THZ synthesis subsequent to the pentulose biosynthesis. It is not known how many steps

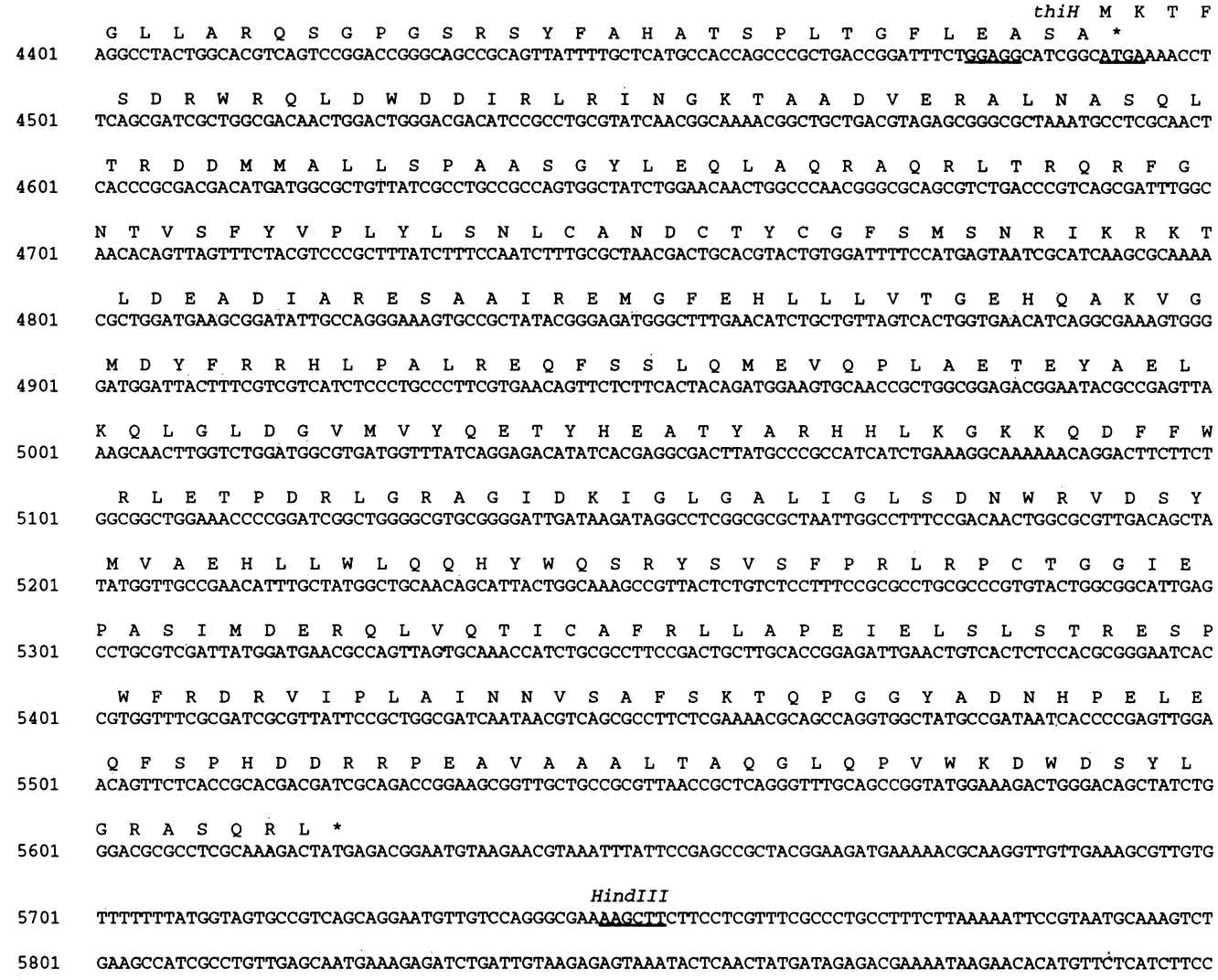

FIG. 4—*Continued.*

are involved in THZ synthesis (Fig. 2). The predicted ThiF polypeptide shares substantial sequence identity with MoeB (ChlN), an enzyme involved in molybdopterin synthesis (Fig. 5) (21). The exact role of MoeB in molybdopterin synthesis has not been established, but it is thought to be involved in the insertion of sulfur into the cofactor (36). One of the genes cloned, *thiC*, was required for HMP-P synthesis. The predicted molecular mass of the ThiC polypeptide is 70 kDa.

The *thi-33* allele (termed *thiB*) was originally reported to confer a Thi$^-$ phenotype and was presumed to define the structural gene for THI-P synthase (23). In our studies,

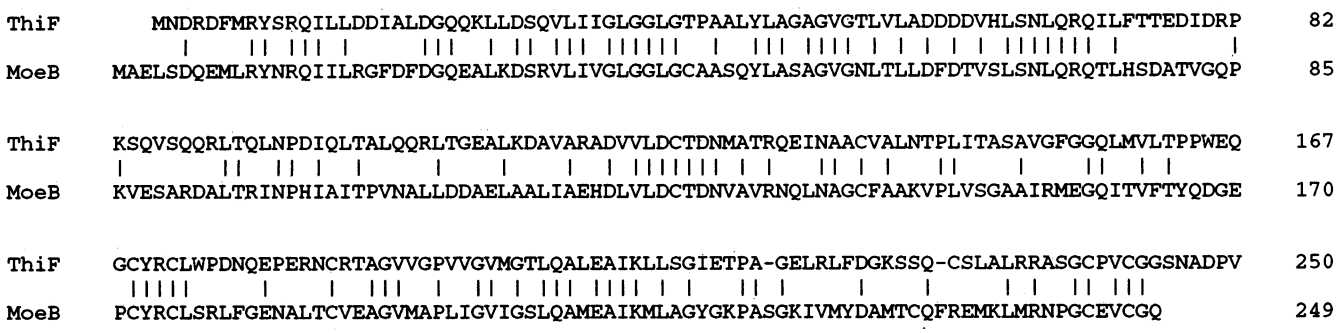

FIG. 5. Sequence similarity between the entire ThiF and MoeB (ChlN) polypeptides. Sequence identities are indicated by vertical bars.

thi-33 strains were THZ auxotrophs and were complemented by clones carrying both *thiF* and *thiG*. Thus, the structural gene for THI-P synthase awaits identification.

Do genes *thiCEFGH* form an operon? The genes are tightly linked and in some cases actually overlap. However, we have hints that the transcriptional organization of this cluster may be complex. Two of the three insertions studied (*thiC39*::Tn*10* and *thiG455*::Km) appear to be incompletely polar as determined by phenotypic testing and by complementation analysis, respectively. Although insertion mutations can be incompletely polar in special circumstances (7), these observations may reflect the presence of internal promoters within the *thi* cluster. Our long-term goals are to understand the mechanistic enzymology and genetic regulation of the THI-PP biosynthetic pathway.

ACKNOWLEDGMENTS

P.B.V. and A.D.B. contributed equally to the experiments described in this paper.

We are grateful to Barbara Bachmann, Hans Bremer, and Carol Gross for providing strains and to Hoffman-LaRoche for a gift of HMP. A.D.B. thanks the members of the Stewart laboratory for their help and hospitality during her stay.

This study was supported by Public Health Service grants DK44083 from the National Institute of Diabetes and Digestive and Kidney Diseases (T.P.B.) and GM36877 from the National Institute of General Medical Sciences (V.S.).

REFERENCES

1. **Arps, P. J., and M. E. Winkler.** 1987. Structural analysis of the *Escherichia coli hisT* operon by using a kanamycin resistance cassette. J. Bacteriol. **169:**1061–1070.
2. **Bachmann, B. J.** 1990. Linkage map of *Escherichia coli* K-12, edition 8. Microbiol. Rev. **54:**130–197.
3. **Backstrom, A. D., R. A. S. McMordie, and T. P. Begley.** Unpublished data.
4. **Backstrom, A. D., V. Stewart, and T. P. Begley.** Unpublished data.
5. **Brown, G. M., and J. M. Williamson.** 1987. Biosynthesis of folic acid, riboflavin, thiamine, and pantothenic acid, p. 521–538. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*. Cellular and molecular biology, vol. 1. American Society for Microbiology, Washington, D.C.
6. **Chamberlin, J. P.** 1979. Flourographic detection of radioactivity in polyacrylamide gels with the water-soluble fluor, sodium salycilate. Anal. Biochem. **98:**132–135.
7. **Ciampi, M. S., and J. R. Roth.** 1988. Polarity effects in the *hisG* gene of Salmonella require a site within the coding sequence. Genetics **118:**193–202.
8. **David, S., B. Estramareix, J.-C. Fischer, and M. Therisod.** 1982. The biosynthesis of thiamine. Synthesis of [1,1,1,5-$^2H_4$]-1-deoxy-D-*threo*-2-pentulose and incorporation of this sugar in biosynthesis of thiazole by *Escherichia coli* cells. J. Chem. Soc. Perkin Trans. **1:**2131–2137.
9. **Davis, R. W., D. Botstein, and J. R. Roth.** 1980. Advanced bacterial genetics. A manual for genetic engineering. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
10. **DeMoll, E., and W. Shive.** 1985. Determination of the metabolic origin of the sulfur atom in thiamine of *Escherichia coli* by mass spectrometry. Biochem. Biophys. Res. Commun. **132:**217–222.
11. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:**387–395.
12. **Estramareix, B., and S. David.** 1990. Conversion of 5-aminoimidazole ribotide to the pyrimidine of thiamin in enterobacteria: study of the pathway with specifically labeled samples of riboside. Biochim. Biophys. Acta **1035:**154–160.
13. **Estramareix, B., and M. Therisod.** 1972. Tyrosine as a factor in biosynthesis of the thiazole moiety of thiamine in *Escherichia coli*. Biochim. Biophys. Acta **273:**275–282.
14. **Estramareix, B., and M. Therisod.** 1984. Biosynthesis of thiamine: 5-aminoimidazole ribotide as the precursor of all the carbon atoms of the pyrimidine moiety. J. Am. Chem. Soc. **106:**3857–3860.
15. **Fujio, T., M. Hayashi, A. Iida, T. Nishi, and T. Hagihara.** November 1990. European patent 90309617.0.
16. **Groisman, E. A., and M. J. Casadaban.** 1986. Mini-Mu bacteriophage with plasmid replicons for in vivo cloning and *lac* gene fusing. J. Bacteriol. **168:**357–364.
17. **Haake, P.** 1987. Thiamine dependent enzymes, p. 390–403. *In* M. I. Page and A. Williams (ed.), Enzyme mechanisms. Royal Society of Chemistry, London.
18. **Hames, B. D., and D. Rickwood.** 1981. Gel electrophoresis of proteins: a practical approach. IRL Press, London.
19. **Imamura, N., and H. Nakayama.** 1981. *thiD* locus of *Escherichia coli*. Experientia **37:**1265–1266.
20. **Imamura, N., and H. Nakayama.** 1982. *thiK* and *thiL* loci of *Escherichia coli*. J. Bacteriol. **151:**708–717.
21. **Johnson, J. L., M. M. Wuebbens, and K. V. Rajagopalan.** 1989. The structure of a molybdopterin precursor. J. Biol. Chem. **264:**13440–13447.
22. **Kawasaki, T., A. Iwashima, and Y. Nose.** 1969. Regulation of thiamine biosynthesis in *Escherichia coli*. J. Biochem. **65:**407–416.
23. **Kawasaki, T., T. Nakata, and Y. Nose.** 1968. Genetic mapping with a thiamine-requiring auxotroph of *Escherichia coli* K-12 defective in thiamine phosphate phosphorylase. J. Bacteriol. **95:**1483–1485.
24. **Kawasaki, T., and Y. Nose.** 1969. Thiamine regulatory mutants in *Escherichia coli*. J. Biochem. **65:**417–425.
25. **Kayama, Y., and T. Kayasaki.** 1973. Purification and properties of thiaminephosphate pyrophosphorylase from *Escherichia coli*. Arch. Biochem. Biophys. **158:**242–248.
26. **Leder, I. G.** 1970. Thiamine monophosphate pyrophosphorylase (crystalline) (2-methyl-4-amino-5-hydroxymethylpyrimidine-pyrophospate: 4-methyl-5-(2′-phosphoethyl)-thiazole-2-methyl-4-aminopyrimidine-5-methenyltransferase, EC 2.5.1.3. Methods Enzymol. **18A:**207–212.
27. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
28. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
29. **Mizote, T., and H. Nakayama.** 1989. The *thiM* locus and its relation to phosphorylation of hydroxyethylthiazole in *Escherichia coli*. J. Bacteriol. **171:**3228–3232.
30. **Nakayama, H.** 1990. Genetic analysis of thiamin pyrophosphate biosynthesis in Escherichia coli. Vitamins **64:**619–632.
31. **Nakayama, H., and R. Hayashi.** 1972. Biosynthetic pathway of thiamine pyrophosphate: a special reference to the thiamine monophosphate-requiring mutant and the thiamine pyrophosphate-requiring mutant of *Escherichia coli*. J. Bacteriol. **112:**1118–1126.
32. **Needleman, S. B., and C. D. Wunsch.** 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. **48:**443–453.
33. **Nohno, T., Y. Kasai, and T. Saito.** 1988. Cloning and sequencing of the *Escherichia coli chlEN* operon involved in molybdopterin biosynthesis. J. Bacteriol. **170:**4097–4102.
34. **Oppenheim, D., and C. Yanofsky.** 1980. Translational coupling during expression of the tryptophan operon of *Escherichia coli*. Genetics **95:**785–795.
35. **Pearson, W. R., and D. J. Lipman.** 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85:**2444–2448.
36. **Rajagopalan, K. V., and J. L. Johnson.** 1992. The pterin molybdenum cofactors. J. Biol. Chem. **267:**10199–10202.
37. **Ryals, J., R.-Y. Hsu, M. N. Lipsett, and H. Bremer.** 1982. Isolation of single-site *Escherichia coli* mutants deficient in thiamine and 4-thiouridine synthesis: identification of a *nuvC* mutant. J. Bacteriol. **151:**899–904.
38. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequenc-

ing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
39. **Shine, J., and L. Dalgarno.** 1975. Determinants of cistron specificity in bacterial ribosomes. Nature (London) **254**:34–38.
40. **Singer, M., T. A. Baker, G. Schnitzler, S. M. Deischel, M. Goel, W. Dove, K. J. Jaacks, A. D. Grossman, J. W. Erickson, and C. A. Gross.** 1989. Collection of strains containing genetically linked alternative antibiotic resistance elements for genetic mapping of *Escherichia coli*. Microbiol. Rev. **53**:1–24.
41. **Squires, C., A. Krainer, G. Barry, W.-F. Shen, and C. L. Squires.** 1981. Nucleotide sequence at the end of the gene for the RNA polymerase β′ subunit (*rpoC*). Nucleic Acids Res. **9**:6827–6839.
42. **Stewart, V., and J. Parales, Jr.** 1988. Identification and expression of genes *narL* and *narX* of the *nar* (nitrate reductase) locus of *Escherichia coli* K-12. J. Bacteriol. **170**:1589–1597.
43. **Tabor, S., and C. C. Richardson.** 1985. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. Proc. Natl. Acad. Sci. USA **82**:1074–1078.
44. **Tabor, S., and C. C. Richardson.** 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA **84**:4767–4771.
45. **Winans, S. C., S. J. Elledge, J. H. Krueger, and G. C. Walker.** 1985. Site-directed insertion and deletion mutagenesis with cloned fragments in *Escherichia coli*. J. Bacteriol. **161**:1219–1221.
46. **Yanisch-Perron, C., J. Viera, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33**:103–119.
47. **Young, D. W.** 1986. The biosynthesis of the vitamins thiamin, riboflavin, and folic acid. Nat. Prod. Rep. **3**:395–419.