Thiamin Biosynthesis in Escherichia coli

IDENTIFICATION OF ThiS THIOCARBOXYLATE AS THE IMMEDIATE SULFUR DONOR IN THE THIAZOLE FORMATION*

(Received for publication, July 7, 1997, and in revised form, March 18, 1998)

Sean V. Taylor, Neil L. Kelleher, Cynthia Kinsland, Hsiu-Ju Chiu, Colleen A. Costello, Allyson D. Backstrom, Fred W. McLafferty, and Tadhg P. Begley‡

From the Department of Chemistry, Cornell University, Baker Laboratory, Ithaca, New York 14853

ThiFS and ThiI are required for the biosynthesis of the thiazole moiety of thiamin in Escherichia coli. The overproduction, purification, and characterization of ThiFS and the identification of two of the early steps in the biosynthesis of the thiazole moiety of thiamin are described here. ThiS isolated from E. coli thiS* is post-translationally modified by converting the carboxylic acid group of the carboxyl-terminal glycine into a thio-carboxylate. The thiI gene plays an essential role in the formation of the thiocarboxylate because ThiS isolated from a thiI- strain does not contain this modification. ThiF catalyzes the adenylation by ATP of the carboxyl-terminal glycine of ThiS. This reaction is likely to be involved in the activation of ThiS for sulfur transfer from cysteine or from a cysteine-derived sulfur donor.

The thiamin biosynthetic pathway in Escherichia coli is outlined in Fig. 1 (1–3). This pathway involves the separate synthesis of the thiazole (5-methyl-4-(β-hydroxyethyl)thiazole phosphate) and the pyrimidine (4-amino-5-hydroxymethylpyrimidine pyrophosphate) moieties, which are then coupled to give thiamin phosphate. The pyrimidine is derived from 5-aminoimidazole ribotide (4–7). The thiazole is derived from tyrosine (8–10), cysteine (11, 12), and 1-deoxy-o-xylulose-5-phosphate (13, 14). The mechanistic enzymology of the thiazole and the pyrimidine formation is still poorly understood. A five gene operon (thiCDEFGH) involved in thiamin biosynthesis has been cloned and characterized (GenBank accession number U00006; Ref. 15).

The thiC gene complements pyrimidine (4-amino-5-hydroxymethylpyrimidine) requiring mutants, the thiFGH genes complement thiazole-requiring mutants, and thiE codes for thiamin phosphate synthase (16). An additional gene (thiI) required for thiazole biosynthesis in Salmonella typhimurium has recently been identified (17). This gene maps at 9.5 min on the E. coli chromosome and complements a nucC mutation.2,3 NuvC is required for the biosynthesis of thiouridine in tRNA (18) as well as the thiazole moiety of thiamin (19) (Fig. 2). This suggests that ThiI may play a role in the sulfur transfer chemistry involved in the thiazole biosynthesis.

When the gene product, originally assigned as ThiG, was purified from an overexpression strain, mass spectrometry and Edman sequencing both demonstrated that the protein consisted of two subunits. One of the subunits, which we will call ThiS, has a mass of 7310.74 Da. The second subunit, which will retain the ThiO name, has a mass of 26896.5 Da.4 Although the anomalous mass of ThiG was previously noted (15), ThiS was missed in the original reading frame assignment because of errors in both published sequences (15, 21) and also because of the difficulty of detecting ThiS, which stains poorly with Coomassie Blue and migrates at the dye front during normal SDS-PAGE analysis. The carboxyl terminus of ThiS has the -Gly-Gly sequence found at the carboxyl terminus of human erythrocyte ubiquitin (GenBank accession number 1070588). This sequence similarity was striking because ThiG shows significant sequence similarity to the ubiquitin-activating enzyme (GenBank accession number 731039) including the ATP-binding site. This suggested that ThiF might catalyze the adenylation of ThiS and that ThiS-COAMP might react with cysteine (or a cysteine-derived sulfur donor) to give ThiS-COSH (Fig. 3). In addition, the dual role of Thi in thiazole and in 4-thiouridine biosynthesis suggested that ThiI might play a role in this sulfur transfer reaction. In this paper, we describe experiments to test these hypotheses.

EXPERIMENTAL PROCEDURES

LB broth was purchased in dehydrated form from Life Technologies, Inc. Tryptose blood agar base was purchased from Difco (Detroit, MI). Ampicillin and isopropyl-β-D-thiogalactopyranoside were from Jersey Lab and Glove Supply (Livingston, NJ). Tris, DTT, EDTA, (NH₄)₂SO₄, and ATP were from Sigma. Sodium chloride was from Fisher (Pittsburgh, PA). Acetic acid and MeOH were from Aldrich. Acrlyamide/Bis (37.5:1) was purchased from Bio-Rad. Dialysis membrane was from Spectrum (Houston, TX). All buffers were prepared from distilled, deionized water and were filtered through Millipore type HA 0.45-μm filters before use. Wizard® phospho-certification kit was purchased from Promega (Madison, WI). BL21(DE3) and the âDE3 Lysogenization kit were purchased from Novagen (Madison, WI).

Protein purifications were performed on a Waters 650 chromatography instrument (Milford, MA). Gel filtration column (Superdex 75) was from Amersham Pharmacia Biotech. Reverse-phase mass spectrometric preparations were on Microm BioResources (Auburn, CA) reverse-phase peptide traps. Protein concentration was assayed using Coomassie

* This work was supported by Grants DK44083 (to T. P. B.) and GM16609 (to F. W. M.) from the National Institutes of Health, a fellowship from the ACS Analytical Chemistry Division sponsored by the Perkin-Elmer Corp. (to N. L. K.), a National Institutes of Health Training Grant (to S. V. T. and N. L. K.), and a National Science Foundation Fellowship (to C. L. K.). 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.

‡ To whom correspondence should be addressed.

1 This compound has been referred to as 1-deoxy-o-threo-2-pentulose-5-phosphate.

2 C. Kinsland and T. P. Begley, unpublished results.

3 Mueller, E. G., Buck, C. J., Palencchar, P. M., Barnhart, L. E., and Paulson, J. L. (1998) Nucleic Acids Res., in press.

4 Kelleher, N. L., Taylor, S. V., Grannis, D., Kinsland, C., Chiu, H.-J., Begley, T. P., and McLafferty, F. W. (1998) Protein Sci., in press.

5 The abbreviations used are: PAGE, polyacrylamide gel electrophoresis; ES/FTMS, electrospray ionization Fourier transform mass spectrometry; DTT, dithiothreitol; SWIFT, stored waveform inverse Fourier transform; SORI, sustained off-resonance irradiation; PCR, polymerase chain reaction; ThiS-COAMP, C-terminal carboxy adenylated ThiS.
showed the lowest level of basal expression and a high level of isopropyl-β-D-thiogalactopyranoside-induced expression of T7 RNA polymerase was chosen for use as an overexpression strain.

Construction of the ThiFS Overexpression Vector—The 3 base pairs immediately upstream of the Thi start codon on plasmid pET-22b (15) were mutated to insert a SalI site (CATATA) into the plasmid. The 5′ primer was 5′-CCG GGA ATT AAT GTG AGT TGA AGT ATG ACC GTG ACT TTA-3′ and the 3′ primer was 5′-CCA GAT AGC CAC TGG CGG-3′. The desired 1.8-kilobase pair fragment was purified using Wizard™ PCR prep (Promega) and the NdeI/SalI fragment cloned into the corresponding sites on pET-22b (Novagen) to yield plasmid pG201. The thiI/SalI fragment (549 base pairs) was exchanged with the corresponding DNA from pVJS720 (15). A representative plasmid was named pCAC111. The remaining PCR-derived DNA (520 base pairs) was sequenced, and no mutations were observed.

Overexpression and Purification of ThiFS from E. coli BL21(DE3) (thiI)—Plasmid pCAC111 was transformed into E. coli strain BL21(DE3), and transformants were selected on tryptose blood agar plates containing ampicillin (200 µg/ml). A single colony of E. coli containing pCAC111 was grown at 37 °C for 12 h in 3 ml of LB broth supplemented with ampicillin (200 µg/ml); 1 ml of this starter culture was diluted into 1 liter of the same medium and grown to an A600 of approximately 0.6. Expression was induced by adding isopropyl-β-D-thiogalactopyranoside to a final concentration of 0.3 mM, and growth was continued at 37 °C for 12 h. Cells were harvested by centrifugation (8000 × g, 30 min) and stored at −70 °C. A frozen cell pellet from 1 liter of cell culture was thawed and resuspended in 25 ml of loading buffer (50 mM Tris, 2 mM DTT, 2 mM EDTA, pH 7.5), treated with lysozyme (5 mg, 40 min, 4°C), sonicated at 4°C for 4 min, and centrifuged (27000 × g, 15 min). The resulting cell free lysate was taken to 20% (NH₄)₂SO₄ saturation at 4°C by slow addition of solid (NH₄)₂SO₄ and gently stirring over 45 min. This mixture was centrifuged (27000 × g, 15 min), and the supernatant was taken to 50% (NH₄)₂SO₄ saturation at 4°C by slow addition of solid (NH₄)₂SO₄ with gentle stirring over 45 min. This mixture was centrifuged as before, and the resulting pellet was stored at −70°C. The 20–50% (NH₄)₂SO₄ pellet was thawed, resuspended in 10 ml of loading buffer, and dialyzed (12000–14000 molecular weight cut-off) for 12 h at 4°C against 4 liters of loading buffer to exchange out (NH₄)₂SO₄ from the solution. This was then passed through a 0.45-µm filter and loaded onto a Protein-Pak DEAE HR 8-µm 1000 Å column (20 × 100 mm) (Waters). The ThiF/ThiS protein mixture was purified by gradient elution (flow rate was 4 ml/min). The column with bound protein was washed with loading buffer for 10 min and then taken from 0% elution buffer (50 mM Tris, 2 mM DTT, 2 mM EDTA, 1 mM NaCl, pH 7.5) and ThiS eluted together after elution for 30 min. The ThiF/ThiS eluted at 15% elution buffer, suggesting that ThiS is bound to ThiF. The ThiF-containing fractions were identified by SDS-PAGE analysis, pooled, concentrated, exchanged into gel filtration buffer (50 mM Tris, 2 mM DTT, 2 mM EDTA, 20 mM NaCl) using a YM10 Amicon ultrafiltration membrane, and loaded onto a Superdex 75 gel filtration column (Amersham Pharmacia Biotech). At a flow rate of 1.2 ml/min, ThiF and ThiS again eluted together after elution for 1.5 h. The ThiFS-containing fractions were identified by SDS-PAGE analysis and pooled, concentrated, and exchanged into 50 mM Tris, 2 mM DTT, pH 7.5, using a YM10 Amicon ultrafiltration membrane (Millipore). The protein solution was diluted by adding glycerol to 10% and frozen at a final concentration of 16 mg/ml until use. SDS-PAGE analysis of the purification is shown in Fig. 4A.

Overexpression and Purification of ThiFS from E. coli VJS8890(DE3) (thiI)—The purification of ThiFS from E. coli VJS8890(DE3) (thiI) was identical to the purification of ThiFS from E. coli BL21(DE3) (thiI). SDS-PAGE analysis of the purification is shown in Fig. 4B.

Phosphoryse Assay—Phosphoryse was assayed using the Enz-Chek phosphoryse assay kit (27). The assay volume was 1 ml. The reaction buffer was 20 mM Tris-HCl, 1 mM MgCl₂, pH 7.5. 75 µl of ThiFS isolated from BL21(DE3) (thiI) (approximate concentration 16 mg/ml; the molar ratio of ThiF to ThiS was not determined) was added to initiate the reaction. ATP was added to a concentration of 2.5 mM. Assays were run for 1 min.

Mass Spectrometric Detection of a Covalent ThiS-AMP Adduct—Recombinant ThiF (10 µg/ml) was incubated with 10 mM ATP, 20 mM Tris-HCl, 1 mM MgCl₂, pH 7.5, in a total volume of 25 µl, then acidified to pH 3 by addition of glacial acetic acid (2 µl). The assay mixture was diluted with water (100 µl) and loaded onto a reverse-phase peptide trap (Michrom BioResources, Inc.). The trap was washed with 1 ml of 98% H₂O/1% MeOH/1% acetic acid, and the protein was eluted with 150 µl of 80% methanol/20% acetic acid (the first 50 µl was
discarded). 2 µl of this eluent was loaded into a borosilicate glass capillary (1.5 mm inner diameter) pulled to a 2-µm tip at one end (28).

A platinum wire inserted into the distal end of the glass capillary made contact with the solution and a voltage of 0.7–1.4 kV on the wire determined the flow rate (25–75 nl/min). Droplets formed by ESI were sampled by the heated metal capillary (110 °C) of a 6T Fourier transform mass spectrometer described previously (29, 30). Briefly, ions are guided through five stages of differential pumping by a skimmer and three quadrupole ion guides into the magnet bore held at 10⁻⁹ torr. Ions of interest were isolated in the trapped ion cell by the stored wave form inverse Fourier transform (SWIFT) technique (31) and collisionally activated by sustained off-resonance irradiation (SORI) (32) in the presence of 10⁻⁶ torr N₂.

Mass Spectrometric Detection of ThiS Thiocarboxylate—ThiFS (25 µl, 16 mg/ml) isolated from E. coli BL21(DE3) (thiI⁺) was thawed and acidified to pH 3 by addition of glacial acetic acid (2 µl). This mixture was diluted with water (100 µl) and loaded onto a reverse-phase peptide trap (Michrom BioResources, Inc.). The trap was washed with 1 ml of 98% H₂O/1% MeOH/1% acetic acid, and the protein was eluted with 150 µl of 80% methanol/20% acetic acid (the first 50 µl was discarded). 2 µl of this eluent was loaded into a borosilicate glass capillary and electrosprayed as described above.

**RESULTS**

**Purification of ThiFS**—The purification of ThiFS is summarized in Fig. 4. Both of the ThiFS overexpression strains provided high levels of soluble protein. The average yield of purified (>90%) ThiFS was 5 mg/liter of cell culture. ThiF and ThiS copurify through several steps, demonstrating that ThiF and ThiS form a stable complex. ThiS stains poorly with Coomassie Blue and is not visible on the gel. The mass spectra discussed below clearly demonstrate its presence in all ThiFS samples examined.

**Pyrophosphate Production by ThiFS Isolated from E. coli BL21(DE3) (thiI⁺)**—ThiFS catalyzed the formation of pyrophosphate from ATP (Fig. 5). This reaction required Mg²⁺ as a cofactor. Pyrophosphate production reached a plateau, presum-
Thiamin Biosynthesis in E. coli

Mass Spectrometric Detection of a Covalent ThiS-AMP Adduct—The ESI/FTMS of ATP-treated ThiFS isolated from E. coli VJS2890(DE3) (thiI') is shown in Fig. 6. In addition to showing ions that can be assigned to [ThiS]+ (m/z = 1463.15), the spectrum shows ions at m/z = 1528.97 that can be assigned to [ThiS-COAMP]+.

Mass Spectrometric Characterization of ThiS Isolated from E. coli BL21(DE3) (thiI') and from E. coli VJS2890(DE3) (thiI')—ESI/FTMS on ThiFS isolated from E. coli BL21(DE3) (thiI') demonstrated that the predominant ThiS ion had an additional mass of 16 Da compared with ThiS isolated from E. coli VJS2890(DE3) (thiI') (Fig. 7, A and B, respectively). This is consistent with the replacement of one oxygen with sulfur on the protein. The ratio of ThiS to ThiS+16 in the (thiI') sample was approximately 1:9. SWIFT isolation and SORI activation of the [ThiS+16]+ ions gave 10 fragment ions; two of these had masses of 7235.64 and 7178.69 and were assigned as b65 and b64 ions (fragments that contain the NH2 terminus of the protein) (33). These represent the loss of 91 Da (COOH-terminal glycine + 16 Da) and 148 Da (COOH-terminal glycine + 16 Da + internal glycine), respectively, from the COOH terminus of the ThiS+16 adduct and strongly suggested that the COOH-terminal glycine carboxylate of ThiS has been converted to a thiocarboxylate in the (thiI') strain (Fig. 8).

To confirm this, ThiFS from E. coli BL21(DE3) (thiI') was alkylated with iodoacetic acid. Analysis of the reaction mixture by ESI/FTMS demonstrated the formation of a new species with mass of 7384.70, corresponding to [ThiS+16-
H+CH₂COOH (Fig. 9A). SWIFT isolation of the 6+ charge state of the sodium adduct at m = 7406.7 (formed in a separate sample ionization), followed by SORI activation gave a series of fragments (Fig. 9B). All of these fragments were assigned either as b ions or y ions (fragments that contain the COOH terminus of the protein) (33). All of the observed b ions through b₆₃ had a mass identical to that predicted from the unmodified ThiS sequence, and the y₆₂, y₆₃, and y₆₄ ions (sodium adducts) all contained the modified amino acid. This localized the site of modification to the carboxyl-terminal Ala₆₄, Gly₆₅, and Gly₆₆ residues (Fig. 9C). Because the only functional group in these three residues is the Gly₆₆ carboxylate, ThiS isolated from E. coli BL21(DE3) (thiI¹) must contain a thiocarboxylate at the COOH terminus.

**DISCUSSION**

The sequence similarity between ThiF/ThiS and the ubiquitin-activating enzyme E1/ubiquitin suggested that ThiF might catalyze the adenylation of ThiS and that ThiS might be the sulfur carrier in thiamin biosynthesis. In addition, the dual role of ThiI in both thiazole biosynthesis and in sulfur transfer in 4-thiouridine biosynthesis in tRNA suggested that ThiI may play a role in the sulfur transfer to ThiS. The high level over-expression of soluble ThiFS from E. coli BL21(DE3) (thiI¹) and
from *E. coli* VJS2890(DE3) (thiI') and the partial purification of these proteins (>90%) has made it possible to test these hypotheses (Fig. 3).

When ThiFS isolated from *E. coli* BL21(DE3) (thiI') was incubated with ATP, a rapid initial burst of pyrophosphate formation was observed (Fig. 5). The reaction then stops presumably because of the complete conversion of ThiS to ThiS-COAMP. It was possible to directly detect this species using ESI/FTMS (Fig. 6). This demonstrates that ThiF catalyzes the adenylation of ThiS on an oxygen nucleophile.

Analysis of ThiFS isolated from *E. coli* BL21(DE3) (thiI') by ESI/FTMS demonstrated the presence of a ThiS+16 adduct as the major form of isolated ThiS (Fig. 7A). This could be because of methionine oxidation during isolation or to the replacement of an oxygen by sulfur. Partial sequencing of ThiS+16 (Fig. 8) and the product resulting from alkylation of ThiS+16 by iodoacetate (Fig. 9A) unambiguously identified the +16 adduct as an oxygen to sulfur replacement and uniquely localized the site of modification to the carboxyl-terminal glycine residue.

MS analysis of ThiS isolated from *E. coli* VJS2890(DE3) (thiI') demonstrated that the ThiS-COSH was not formed in this strain (Fig. 7B). This suggests that ThiI plays an essential role in the conversion of ThiS-COAMP to ThiS-COSH.

The detection of ThiS-COSH was surprising. None of the previous mechanistic proposals for the formation of the thiazole moiety of thiamin predicted the existence of such an intermediate (1).

The mechanistic enzymology of the sulfur transfer chemistry involved in the biosynthesis of thiamin, molybdopterin, biotin, and lipoic acid is still poorly understood. The immediate sulfur source for the formation of biotin and lipoic acid is unknown, and the mechanism of the sulfur insertion chemistry is not yet clear (34–38). The sulfur transfer chemistry involved in thiamin and molybdopterin biosynthesis is likely to be similar. A three-enzyme system (MoeB, MoaD, and MoaE) for the sulfur transfer to precursor Z in molybdopterin biosynthesis has been reconstituted (39). ThiF and MoeB (GenBank™ accession number 127233) show high sequence similarity, and both MoeD (GenBank™ accession number 2507065) and ThiS contain the carboxyl-terminal GG residues and bind sulfur as a thioethercarboxylate (20, 40). With the identification of the immediate sulfur donor in thiamin biosynthesis and the development of an overexpression strain yielding milligram quantities of ThiS-COSH, it should now be possible to reconstitute the biosynthesis of the thiazole moiety. Efforts along these lines are in progress.

---

6 Kinsland, C., Taylor, S. V., Kelleher, N. L., McLafferty, F. W., and Begley, T. P. (1998) *Protein Sci.*, in press.