Using genomic information to investigate the function of ThiI, an enzyme shared between thiamin and 4-thiouridine biosynthesis

EUGENE G. MUELLER AND PETER M. PALENCHAR

Department of Chemistry & Biochemistry, University of Delaware, Newark, Delaware 19716

(RECEIVED June 2, 1999; ACCEPTED July 23, 1999)

Abstract

The gene *thiI* encodes a protein (ThiI) that plays a role in the transfer of sulfur from cysteine to both thiamin and 4-thiouridine, but the reaction catalyzed by ThiI remains undetermined. Based upon sequence alignments, ThiI shares a unique “P-loop” motif with the PP_i synthetase family, four enzymes that catalyze adenylation and subsequent substitution of carbonyl oxygens. To test whether or not this motif is critical for ThiI function, the Asp in the motif was converted to Ala (D189A), and a screen for in vivo 4-thiouridine production revealed the altered enzyme to be inactive. Further scrutiny of sequence data and the crystal structures of two members of the PP_i synthetase family implicated Lys321 in the proposed adenylation function of ThiI, and the critical nature of Lys321 has been demonstrated by site-directed mutagenesis and genetic screening. Our results, then, indicate that ThiI catalyzes the adenylation of a substrate at the expense of ATP, a narrowing of possible reactions that provides a strong new basis for deducing the early steps in the transfer of sulfur from cysteine to both thiamin and 4-thiouridine.

Keywords: adenylation; 4-thiouridine; nuv; PP_i synthetase; RNA modification; sulfur incorporation; thiamin; ThiI

The pathways of sulfur incorporation into thiamin (vitamin B₁) and 4-thiouridine (s⁴U) in tRNA remain incompletely elucidated, though in both cases the sulfur originates in cysteine. Evidence that the two pathways share common steps was first provided by Lipsett and coworkers, who discovered that some mutant *Escherichia coli* that fail to make s⁴U are also thiamin auxotrophs (Ryals et al., 1982). One gene shared between both of these sulfur incorporation pathways is *thiI*. The group of Downs identified *thiI* as essential for thiazole biosynthesis (Webb et al., 1997), and we subsequently demonstrated a role for *thiI* in s⁴U biosynthesis (Mueller et al., 1998). Neither group, however, identified the reaction catalyzed by ThiI, the gene product of *thiI*. We now report that ThiI very likely catalyzes an adenylation, probably of a substrate bearing a carbonyl group. Several lines of evidence were examined along the way to this conclusion.

First, the reactions that flank it in metabolism could be used to deduce the conversion catalyzed by ThiI, but too much remains unknown about the passage of sulfur from cysteine to thiamin and s⁴U. Begley and coworkers have elegantly demonstrated that a thiocarboxylate at the C-terminal glycine of the small protein ThiS is an intermediate in thiazole biosynthesis (Taylor et al., 1998). In a *thiI* mutant strain of *E. coli*, no ThiS is detected in the thiocar-

boxylate form, which represents a sizable fraction of ThiS in wild-type *E. coli* (Taylor et al., 1998). This result narrows the possible reactions catalyzed by ThiI to those lying between cysteine and thiocarboxylate formation, but the number and identities of sulfur-containing intermediates remain to be determined. Despite extensive surveying, we have been unable to detect any reaction of cysteine catalyzed by ThiI in the presence or absence of ATP, pyridoxal phosphate, tRNA transcript, a mixture of reduced and oxidized glutathione, or combinations of these reagents.

Second, the operon containing *thiI* might provide clues to the precise chemical function of *thiI*. Because bacterial genes involved in the same pathway are frequently clustered behind one promoter, the identity of other genes in the operon may suggest a chemical role for *thiI*. Because *thiI* lies in a monocistronic operon (Webb et al., 1997), no such information is available. Similarly, the response of expression to inducers or the –35 sequence of the promoter can be used to group the expression of a given operon with that of other operons. The operon containing *thiI*, however, appears to be constitutive (Webb et al., 1997), and we cannot discern a –35 sequence.

Third, sequence information can be used to relate ThiI to enzymes of known function. When the sequence of ThiI is compared in its entirety to other proteins in the genomic database using BLASTP (Altschul et al., 1990), only orthologs are returned, which does not illuminate function. However, the more refined PSI-BLAST algorithm uses artificial intelligence methods to search for

Reprint requests to: Eugene G. Mueller, Department of Chemistry & Biochemistry, University of Delaware, Newark, Delaware 19716; e-mail: emueller@udel.edu.

patterns of amino acids (Altschul et al., 1997). Querying the database with the sequence of *ThiI*, four enzymes emerged as the best matches after five iterations of PSI-BLAST. The common motif is SGGXD(S/T), where X denotes any amino acid and (S/T) denotes the presence of either serine or threonine (Fig. 1). *ThiI* shares this motif with GMP synthetases, NAD⁺ synthetases, asparagine synthetases, and argininosuccinate synthetases (Fig. 2), which have been previously grouped as the “PP_i synthetase family” (Tesmer et al., 1996). Furthermore, the SGGXD(S/T) sequence had been identified as a “P-loop” motif specific for this family of enzymes on the basis of sequence analysis (Bork & Koonin, 1994; Tesmer et al., 1996), which was confirmed crystallographically (Rizzi et al., 1996; Tesmer et al., 1996).

Because the four grouped enzymes catalyze the adenylation of an oxygen in either a urea or carboxylate to activate it for substitution with an amine (Fig. 2), it seemed highly unlikely that the match with *ThiI* arose by chance. Also arguing against a chance grouping, other enzymes that adenylate molecules for subsequent nucleophilic substitution (e.g., the tRNA aminoacyl synthetases) do *not* share this motif, suggesting that it has been refined for very particular catalytic properties. Finally, the structures of GMP synthetase (Tesmer et al., 1996) and NAD⁺ synthetase (Rizzi et al., 1996) with products bound show that the SGGXD(S/T) motif (P-loop) provides contacts with ATP in the active site. These lines of reasoning, however, are highly suggestive rather than definitive regarding the role of the identified motif in *ThiI*. Fortunately, feasible experiments were available to address that role.

While ignorance of the reaction catalyzed by *ThiI* precludes *in vitro* assay, the *in vivo* function of s⁴U as a sensor of near-UV light provides a powerful screen for *ThiI* function. The s⁴U at position 8 of bacterial tRNA cross-links to the juxtaposed cytidine at position 13 when the bacteria are exposed to near-ultraviolet (UV) light, which activates a 2+2 cycloaddition between the two bases (Favre et al., 1969; Bergstrom & Leonard, 1971; Favre et al., 1971). The cross-linked tRNAs are poor aminoacylation substrates (Carre et al., 1974), so uncharged tRNAs accumulate and induce the stringent response in the bacteria, which enter growth arrest (Ramabhadran & Jagger, 1976). Mutants in *thiI*, however, do not generate s⁴U in tRNA and therefore continue to grow when exposed to near-UV light (the sensor is absent). This phenomenon provides a powerful screen: mutant bacteria can be transformed with plasmids bearing *thiI*, and when functional *ThiI* is expressed, the mutant bacteria have s⁴U in tRNA and regain sensitivity to near-UV light (are complemented). Using this screen, we identified *thiI* as a critical gene for s⁴U generation (Mueller et al., 1998), and we now report the use of the same methods to assess whether or not site-directed mutants of *ThiI* are functional *in vivo*.

The first test of the functional significance of the SGGXD(S/T) motif was the mutation of the Asp of the motif to Ala. This D189A

<i>ThiI</i>	180	VLSLI SGG DS GVSSYML	197
GMP synthetase	230	VILGL SGG DS SVTAMLL	247
NAD ⁺ synthetase	28	VLGIS SGG DS TALAGKD	45
argininosuccinate synthetase	13	IGIAF SGGL DS SAALLWM	30
asparagine synthetase	230	YGVLL SGGL DS SLLSAIT	247

Fig. 1. Alignment of *ThiI* and members of the PP_i synthetase family (all from *E. coli*). The signature amino acids of this “P-loop” motif are denoted by bold type, and the numbers denote the position of the first and last amino acids listed.

Fig. 2. The reactions catalyzed by the PP_i synthetase family.

ThiI was made using the QuikChangeTM protocol (Stratagene, La Jolla, California) starting with pBH114, in which *thiI* with a Shine-Dalgarno sequence lies behind the T7 promoter of pGEM-T (Promega, Madison, Wisconsin). The near-UV resistant mutant VJS2890(DE3) is complemented by pBH114 (wild-type *ThiI*), but the plasmid (pBH125) bearing the gene for the D189A *ThiI* fails to complement (Fig. 3). Analysis of tRNA isolated from VJS2890 (DE3) harboring pBH125 confirmed the absence of s⁴U, while the same analysis revealed s⁴U in tRNA isolated from the complemented mutant (Fig. 4). To assess whether or not the D189A *ThiI* was properly folded, the gene for the mutant enzyme was ligated into pET15b (Novagen, Madison, Wisconsin) for overexpression and rapid purification with an N-terminal His₆-Tag.¹ The D189A *ThiI* behaves identically to wild-type *ThiI* with respect to overexpression and storage, and the CD spectra of the purified enzymes are qualitatively identical, indicating that the substitution of Asp189 does not cause gross structural alteration.

Additional scrutiny with PSI-BLAST revealed a weakly aligned motif (Bork & Koonin, 1994) that is shared between *ThiI*, the GMP synthetases, the NAD⁺ synthetases, and the argininosuccinate synthetases. In the crystal structures of GMP synthetase (Tesmer et al., 1996) and NAD⁺ synthetase (Rizzi et al., 1996), the Lys contained in this motif lies very near the Asp of the P-loop and is positioned to interact with P_γ of ATP. Interestingly, Arg is found in place of Lys in the argininosuccinate synthetases. To test whether

¹ Wild-type *thiI* in the analogous construct complements VJS2890(DE3) (Mueller et al., 1998).

Fig. 3. Complementation analysis comparing wild-type and D189A ThiI. The plates are masked so that half of the plate is exposed to cycling near-UV light during growth. The mutant VJS2890(DE3) is *thiI*⁻, so it fails to enter growth arrest upon exposure to near-UV light. **A:** pBH114 (expressing wild-type ThiI) complements VJS2890(DE3), and no growth is observed on the side exposed to near-UV light. **B:** pBH125 (expressing D189A ThiI) fails to complement VJS2890(DE3), and growth is observed on both sides of the plate.

ThiI catalyzes adenylation, we mutated the Lys (K321) in this second motif to both Met and Arg. We predicted that the loss of a cationic residue in K321M ThiI should disrupt function while the cationic side chain of Arg in K321R would suffice for function. The results nicely support our predictions.

The plasmid bearing the gene for K321M ThiI fails to complement VJS2890(DE3), and isolated tRNA shows no s⁴U (Fig. 4). Mutagenesis to produce the K321R ThiI yielded genes containing either an AGA or a CGA at codon 321 (both code for Arg). For

Fig. 4. UV/visible spectra of tRNA isolated from VJS2890(DE3) transformed with pBH114 (wild-type ThiI), pBH125 (D189A ThiI), pBH126 (K321M ThiI), pBH127 (K321R ThiI, AGA codon), and pBH128 (K321R ThiI, CGA codon). For s⁴U, λ_{max} = 334 nm.

the first time, an intermediate phenotype was observed: when VJS2890(DE3) bearing these two plasmids was analyzed, the side of the plate exposed to near-UV light bore only 25–75% of the number of colonies observed on the masked side of the plate. In all other selections, either equal numbers (within 10%) of colonies were observed on both sides of the plate, or *no* colonies were observed on the near-UV side. A somewhat greater degree of complementation was exhibited by *thiI* with AGA (vs. CGA) at codon 321. The level of s⁴U in tRNA isolated from these cells (Fig. 4) tracks with the complementation data: more s⁴U is apparent with AGA than with CGA at codon 321 (the latter displays only a hint of s⁴U). The observed phenotype for cells expressing K321R ThiI is consistent with an impaired but functional enzyme.² The genes for the K321M and K321R ThiI were placed into pET15b for overexpression and rapid purification. A wild-type fold for both mutant proteins was suggested by CD spectroscopy and by overexpression and storage behavior.

In summary, genomic database searches with the PSI-BLAST algorithm revealed that ThiI shares a sequence motif with four other enzymes, each of which adenylates a carbonyl oxygen, activating it for nucleophilic substitution. The crystal structures for two of this group of enzymes place the motif at the active site for adenylation (Rizzi et al., 1996; Tesmer et al., 1996), and together the structural and sequence data imply critical roles for Asp189 and Lys321 of ThiI. Elimination of chemical functionality at those positions by site-directed mutagenesis (the D189A and K321M ThiI) eliminates *in vivo* function. Structural and sequence data also

²The electronic supplementary material contains a more extensive discussion of the intermediate phenotype and supporting experiments suggested by the findings of Siegle and Hu (1997) regarding differential expression within a monoclonal culture.

imply that Arg will effectively substitute for Lys321, and K321R ThiI does function in vivo at least part of the time.

We conclude, therefore, that ThiI very likely catalyzes the adenylation (or perhaps cytidylation, uridylation, or guanylation) of a substrate, probably a molecule bearing a carbonyl oxygen. This finding provides a new basis for deducing the chemistry of the early steps in the transfer of sulfur from cysteine to thiamin and 4-thiouridine, and we hope that our results stimulate the formation and testing of hypotheses by others interested in this sulfur traffic. Our search for the other ThiI substrate (and the nucleophile that eventually attacks it) continues.

Supplementary material in Electronic Appendix

Experimental procedures, SDS-PAGE gels showing protein induction and purification, CD spectra for wild-type and each mutant, an analysis of the intermediate phenotype with K321R ThiI. The sequences of all plasmids described in this communication can be found at <http://www.udel.edu/chem/mueller>.

Acknowledgments

This work was supported in part by a grant from the Exxon Education Foundation. P.M.P. was supported by USPHS grant T32 GM08550.

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410.
- Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucl Acids Res* 25:3389–3402.
- Bergstrom DE, Leonard NJ. 1971. Photoreaction of 4-thiouracil with cytosine. Relation to photoreactions in *Escherichia coli* transfer ribonucleic acids. *Biochemistry* 11:1–9.
- Bork P, Koonin EV. 1994. A P-loop-like motif in a widespread ATP pyrophosphatase domain—Implications for the evolution of sequence motifs and enzyme-activity. *Proteins Struct Funct Genet* 20:347–355.
- Carre DS, Thomas G, Favre A. 1974. Conformation and functioning of transfer RNAs—Crosslinked transfer RNAs as substrates for transfer RNA nucleotidyl-transferase and aminoacyl synthetases. *Biochimie* 56:1089–1101.
- Favre A, Michelson AM, Yaniv M. 1971. Photochemistry of 4-thiouridine in *Escherichia coli* transfer RNA^{Val}. *J Mol Biol* 58:367–379.
- Favre A, Yaniv M, Michelson AM. 1969. The photochemistry of 4-thiouridine in *Escherichia coli* tRNA^{Val}. *Biochem Biophys Res Commun* 37:266–271.
- Mueller EG, Buck CJ, Palenchar PM, Barnhart LE, Paulson JL. 1998. Identification of a gene involved in the generation of 4-thiouridine in tRNA. *Nucl Acids Res* 26:2606.
- Ramabhadran TV, Jagger J. 1976. Mechanism of growth delay induced in *Escherichia coli* by near ultraviolet-radiation. *Proc Natl Acad Sci USA* 73:59–63.
- Rizzi M, Nessi C, Mattevi A, Coda A, Bolognesi M, Galizzi A. 1996. Crystal structure of NH₃-dependent NAD⁺ synthetase from *Bacillus subtilis*. *EMBO J* 15:5125.
- Ryals J, Hsu R-Y, Lipsett MN, Bremer H. 1982. Isolation of single-site *Escherichia coli* mutants deficient in thiamine and 4-thiouridine syntheses—Identification of a *muvC* mutant. *J Bacteriol* 151:899.
- Siegele DA, Hu JC. 1997. Gene expression from plasmids containing the *araBAD* promoter at subsaturating inducer concentrations represents mixed populations. *Proc Natl Acad Sci USA* 94:8168–8172.
- Taylor SV, Kelleher NL, Kinsland C, Chiu HJ, Costello CK, Backstrom AD, McLafferty FW, Begley TP. 1998. Thiamin biosynthesis in *Escherichia coli*—Identification of this thiocarboxylate as the immediate sulfur donor in the thiazole formation. *J Biol Chem* 273:16555.
- Tesmer JG, Klem TJ, Deras ML, Davisson VJ, Smith JL. 1996. The crystal structure of GMP synthetase reveals a novel catalytic triad and is a structural paradigm for two enzyme families. *Nat Struct Biol* 3:74–86.
- Webb E, Claas K, Downs DM. 1997. Characterization of *thiI* a new gene involved in thiazole biosynthesis in *Salmonella typhimurium*. *J Bacteriol* 179:4399.