Protein Science (1998), 7:1839–1842. Cambridge University Press. Printed in the USA.

FOR THE RECORD

Overexpression of recombinant proteins with a C-terminal thiocarboxylate: Implications for protein semisynthesis and thiamin biosynthesis

CYNTHIA KINSLAND, SEAN V. TAYLOR, NEIL L. KELLEHER, FRED W. McLAFFERTY, AND TADHG P. BEGLEY

Department of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853

(RECEIVED March 26, 1998; ACCEPTED April 21, 1998)

Abstract: A facile and rapid method for the production of protein C-terminal thiocarboxylates on DNA-encoded polypeptides is described. This method, which relies on the mechanism of the cleavage reaction of intein-containing fusion proteins, can produce multimilligram quantities of protein C-terminal thiocarboxylate quickly and inexpensively. The utility of this method for protein semisynthesis and implications for studies on the biosynthesis of thiamin are discussed.

Keywords: C-terminal thiocarboxylate; protein semisynthesis; recombinant proteins; thiamin biosynthesis

Protein thiocarboxylates, prepared by solid-phase synthesis, are useful intermediates in the synthesis of proteins. For example, reaction between a protein thiocarboxylate and a peptide functionalized with bromoacetamide at the amino terminus gives a peptide in which the glycine amide has been replaced with a thioester. Additionally, methods for coupling an unprotected polypeptide containing a C-terminal thiocarboxylate with an unprotected polypeptide containing either histidine, cysteine, or bromoalanine at the N-terminus have been described (Fig. 1) (Tam et al., 1995; Liu et al., 1996; Muir et al., 1997; Zhang & Tam, 1997). An in vivo route to protein thiocarboxylates should greatly simplify this useful protein synthesis strategy.

In *Escherichia coli*, a protein thiocarboxylate, produced by the post-translational modification of a 66 amino acid protein (ThiS–COOH, 7,310 Da), functions as the immediate sulfur donor for the biosynthesis of the thiazole moiety of thiamin (Fig. 2) (Taylor et al., 1998).

As part of our research on the mechanistic enzymology of thiamin biosynthesis (Begley, 1996), we needed a reliable source of multi-milligram quantities of ThiS–COSH. We decided to explore the possibility of using the recently discovered intein processing chemistry for this synthesis because protein thiocarboxylate synthesis using solid-phase methods is both time consuming and expensive (Barany & Merrifield, 1979) compared to in vivo production.

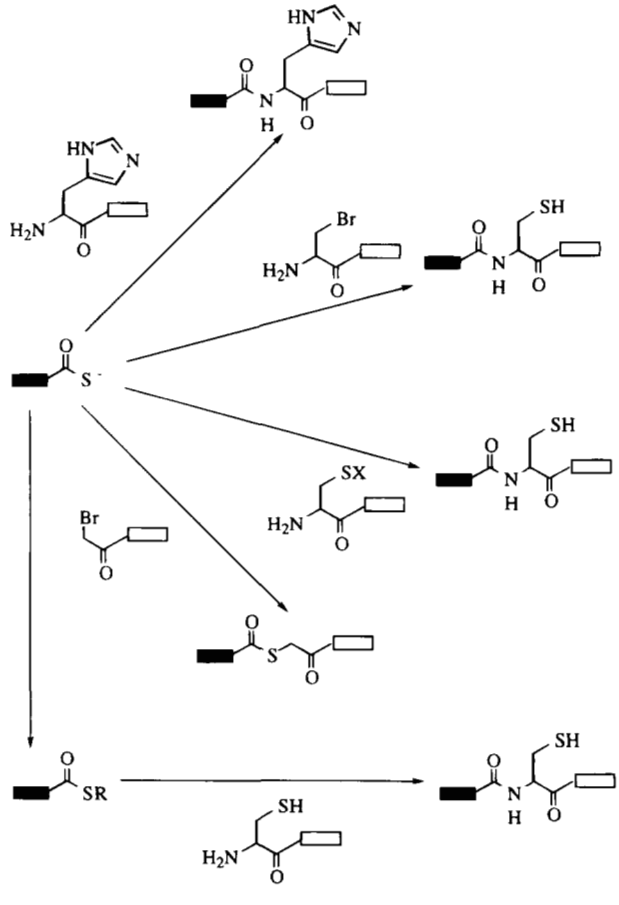

Fig. 1. Utility of C-terminal protein thiocarboxylate for peptide coupling. Coupling reagents are not shown.

Reprint requests to: Tadhg P. Begley, Baker Laboratory, Cornell University, Ithaca, New York 14853; e-mail: tpb2@cornell.edu.

Abbreviations: CBD, chitin binding domain; ESI/FTMS, electrospray ionization Fourier-transform mass spectrometry; Tris, trizma base; DTT, dithiothreitol; EDTA, ethylenediaminetetraacetic acid; IPTG, isopropyl-β-D-thiogalactopyranoside; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; $(NH_4)_2S$, ammonium sulfide; PCR, polymerase chain reaction; TE, 10 mM Tris, 1 mM EDTA, pH 7.5.

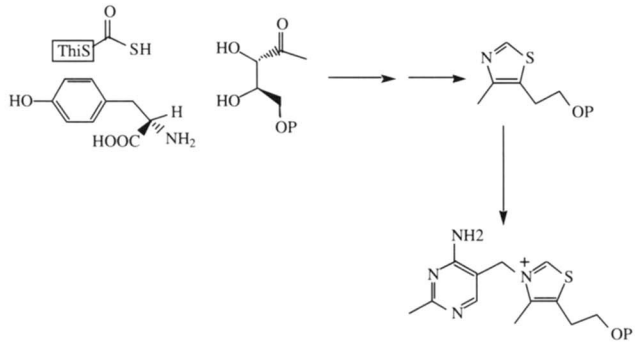

Fig. 2. The role of ThiS thiocarboxylate as the sulfur carrier in thiamin biosynthesis. The N and C_2 of thiazole are derived from the C_α and N of tyrosine; the rest of the carbon backbone is derived from 1-deoxy-D-xylulose-5-phosphate.

The intriguing sequence of reactions involved in intein excision (Xu et al., 1994; Shao et al., 1995; Perler et al., 1997; Shao & Paulus, 1997) has been used in the design of a novel overexpression system (the IMPACT system, New England Biolabs, Fig. 3), which consists of a fusion between the target for overexpression, the intein, and a chitin binding domain (CBD) (Chong et al., 1997). The latter facilitates the purification of the fusion protein by chromatography on a chitin column. The IMPACT system has been used for the expression and purification of a variety of proteins (Chong et al., 1997). During the cleavage reaction intein catalyzed N-S acyl shift at the intein target junction, followed by transesterification of the resulting thioester with DTT results in the release of the target protein as a DTT thioester that is readily hydrolyzed under mild conditions. We reasoned that the corresponding cleavage of the target from the column using ammonium sulfide would give the protein thiocarboxylate. Here we describe the successful implementation of this strategy.

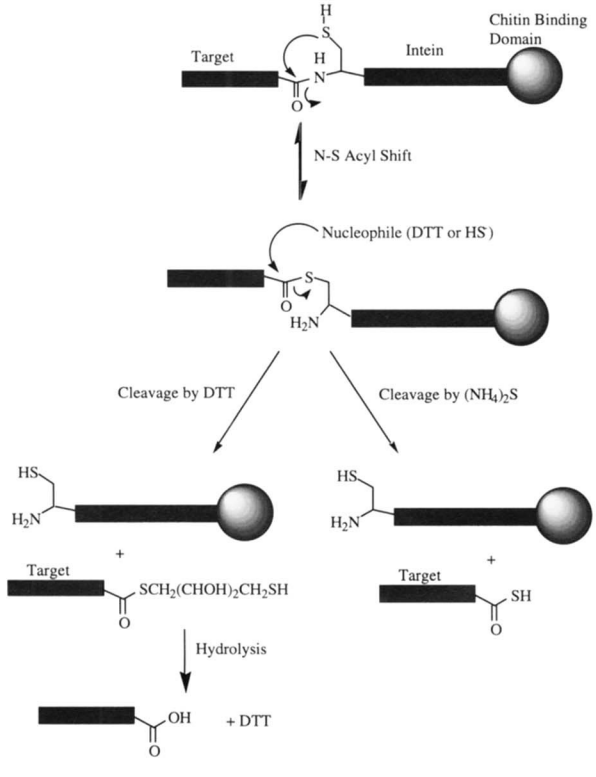

Fig. 3. Purification of target protein from the IMPACT system as either the C-terminal carboxylate or the C-terminal thiocarboxylate.

Results and discussion: As indicated by SDS-PAGE analysis (Fig. 4), the ThiS–Intein–CBD fusion is overexpressed to a high level and is soluble (lanes 1 and 2). Purification of the target protein relies on the tight binding between the CBD of the fusion protein and the chitin resin. The tightness of this binding is demonstrated by the lack of fusion protein in the column flow through (lane 3). Cleavage of the target protein from the fusion is induced by washing the chitin column with the cleaving reagent and allowing it to react overnight. The target protein is then eluted from the column, leaving the intein and CBD bound to the resin (lane 4). The eluted ThiS–COSH is only slightly contaminated and is easily purified to near homogeneity by one gel filtration step (lane 5). A yield of 5 mg of ThiS–COSH per liter of cell culture was typically obtained.

The ESI-FTMS analysis of the purified ThiS from the ammonium sulfide-induced cleavage clearly demonstrates that ThiS cleaved from the fusion as the thiocarboxylate. The peak at m/z of

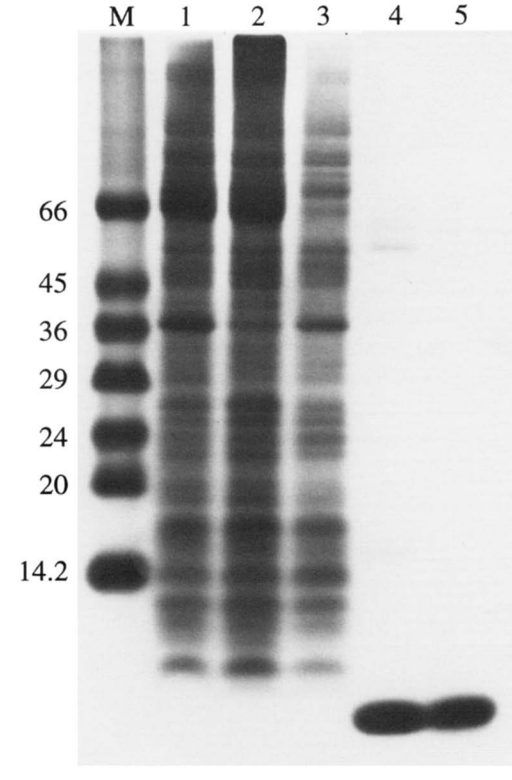

Fig. 4. SDS-PAGE analysis of the purification of ThiS–COSH. M: Molecular weight markers (bovine albumin, 66 kDa; egg albumin, 45 kDa; glyceraldehyde-3-P-dehydrogenase, 36 kDa; bovine carbonic anhydrase, 29 kDa; bovine pancreas trypsinogin, 24 kDa; soybean trypsin inhibitor, 20 kDa; bovine milk α-lactalbumin, 14.2 kDa). 1: Total cellular protein after induction of ThiS–intein–CBD fusion. 2: Soluble protein fraction. 3: Flow-through of the chitin column. 4: ThiS–COSH as eluted from the chitin column. 5: ThiS–COSH after gel filtration chromatography.

1,222.0 Da corresponds to the +6 charge state of ThiS–COSH (7,326.7 Da) (Fig. 5a). The sample was only slightly contaminated with ThiS–COOH (7,310.7 Da, m/z of the +6 charge state = 1,219.5). For the further characterization of ThiS–COSH and as a simple model reaction for one of the protein semisynthesis reactions, we have demonstrated that ThiS–COSH was cleanly monoalkylated by iodoacetic acid. We have previously demonstrated that this alkylation occurs specifically at the thiocarboxylate group when ThiFS is treated under similar conditions (Taylor et al., 1998). The mass of the alkylated C-terminal thiocarboxylate of ThiS is 7,384.7 Da (7,325.7 Da for ThiS–COS^- plus 59.0 Da for the $-CH_2COOH$ group). In the ESI-FTMS of the iodoacetic acid-treated ThiS–COSH, the major component is at m/z of 1,231.9, which corresponds to the +6 charge state of ThiS–$COSCH_2COOH$ (Fig. 5b).

Using the IMPACT system to overexpress and purify proteins as their C-terminal thiocarboxylates has many advantages over solid-phase synthesis. By producing the protein in vivo, much longer sequences are accessible. Additionally, once the DNA sequence coding for the desired peptide is cloned, a virtually limitless supply of that peptide in the thiocarboxylate form is available at a very low cost using a rapid and simple purification.

The availability of the immediate sulfur source for the biosynthesis of the thiazole moiety of thiamin, in multi-milligram quantities, solves a major problem in the reconstitution of the thiazole-forming reactions. Studies on these reactions using overexpressed ThiGH are currently underway.

Materials and methods: A dehydrated form of LB broth was purchased from Gibco/BRL (Gaithersburg, Maryland). Ampicillin and IPTG were from Jersey Lab and Glove Supply (Livingston, New Jersey). Tris, agarose, EDTA, agar, DTT, SDS-PAGE molecular weight markers, and Taq DNA polymerase were from Sigma (St. Louis, Missouri). Sodium chloride was from Fisher (Pittsburgh, Pennsylvania). Acetic acid, methanol, and $(NH_4)_2S$ were from Aldrich Chemical Co. (St. Louis, Missouri). Acrylamide/Bis (37.5:1) was purchased from BioRad Laboratories (Hercules, California). Serva BlueG dye was from Serva (Hauppauge, New York).

The IMPACT system, including the plasmid pCYB1, was obtained from New England Biolabs (Beverly, Massachusetts). The T7 promoter-based overexpression plasmid pET-22b(+) was purchased from Novagen (Madison, Wisconsin), and pGEM3zf(+) was from Promega (Madison, Wisconsin).

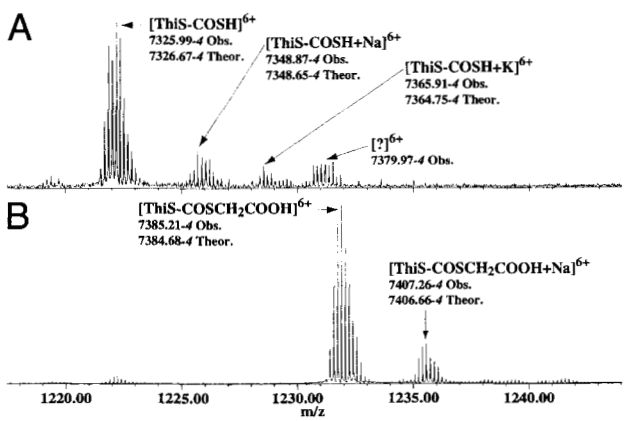

Fig. 5. A: ESI-FTMS of ThiS–COSH. **B:** ESI-FTMS of iodoacetic acid derivatized ThiS–COSH.

Molecular cloning: Standard methods were used for DNA restriction endonuclease digestion, ligation, and transformation of DNA (Ausubel et al., 1987; Sambrook et al., 1989). Plasmid DNA was purified by the Wizard™ Plus DNA miniprep kit (Promega, Madison, Wisconsin). DNA fragments were separated by agarose gel electrophoresis, excised and purified by the GeneClean III kit (Bio101, Vista, California). *E. coli* strain DH5α was used as a recipient for transformations during plasmid construction and for plasmid propagation and storage. All restriction endonucleases and T4 DNA ligase were purchased from New England Biolabs (Beverly, Massachusetts). Perkin–Elmer GeneAmp PCR System 2400 was used for PCR.

Construction of the T7-based IMPACT ThiS overexpression system : The *Nde*I/*Sap*I fragment containing the ThiS coding region from the PCR reaction with pVJS716 (Vander Horn et al., 1993) as the template (upstream mutagenesis primer: 5′-GCG GAG GAA CAT ATG CAG ATC CTG TTT AAC GAT CAA GCG-3′; downstream mutagenesis primer: 5′-GCA ATA GCT CTT CCG CAA CCC CCT GCA ATA ACC TG-3′) was cloned into the corresponding sites of pCYB1 to give pCLK410. The *Sal*I/*Pst*I fragment of pCYB1 containing the intein and chitin binding domain was ligated into the matching sites on pGEM3zf(+). The resulting vector, pCLK901, was digested with *Sal*I and partially digested with *Hind*III. The fragment containing the intact intein and CBD was ligated into *Sal*I/*Hind*III digested pET-22b(+) to yield pCLK902. ThiS was inserted into this vector by ligating the *Nde*I/*BamH*I fragment from pCLK410 containing ThiS and the intein into similarly digested pCLK902 to yield pCLK413 (Fig. 6). (The T7-based system was constructed because the Ptac promoter of pCLK410 gave very low levels of expressed protein. New England Biolabs now markets a T7 promoter-based IMPACT system.)

Overexpression of ThiS–Intein–CBD fusion: The overexpression strain pCLK413/BL21(DE3) was grown in LB supplemented with 200 μg/mL ampicillin at 37 °C in a shaker at 200 rpm until the absorbance at 600 nm was approximately 0.4. The temperature was then adjusted to 15 °C. Once the absorbance at 600 nm was 0.6, overexpression was induced by adding IPTG to 0.5 mM and growth was continued for 12 to 15 h. Cells were harvested by centrifugation (6,100 × g, 30 min), resuspended in 20 mL of column buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 0.1 mM EDTA, 0.1% Triton X-100) per 250 mL of culture, lysed by sonication (Heat Systems Ultrasonics model W-385 sonicator equipped with a 0.5 in. tip on a 5 s cycle, 50% duty for 3 min) and the lysate was cleared by centrifugation (27,000 × g, 30 min).

Purification of ThiS carboxylate and thiocarboxylate: The protocol outlined in the IMPACT manual was followed for the purification of ThiS. All procedures were carried out at 4 °C. For the purification of protein from a 250 mL culture, a 10 mL bed volume of chitin beads was equilibrated with 100 mL of column buffer. The cleared cell lysate was loaded onto the chitin resin at a flow rate of 0.25–0.5 mL/min. The column was washed with 200 mL of column buffer at a flow rate of 1–1.5 mL/min followed by 40 mL of column buffer lacking Triton X-100 at the same flow rate. Cleavage of the target protein (ThiS) from the intein-CBD was then induced by rapidly washing the column with 30 mL of cleavage buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 0.1 mM

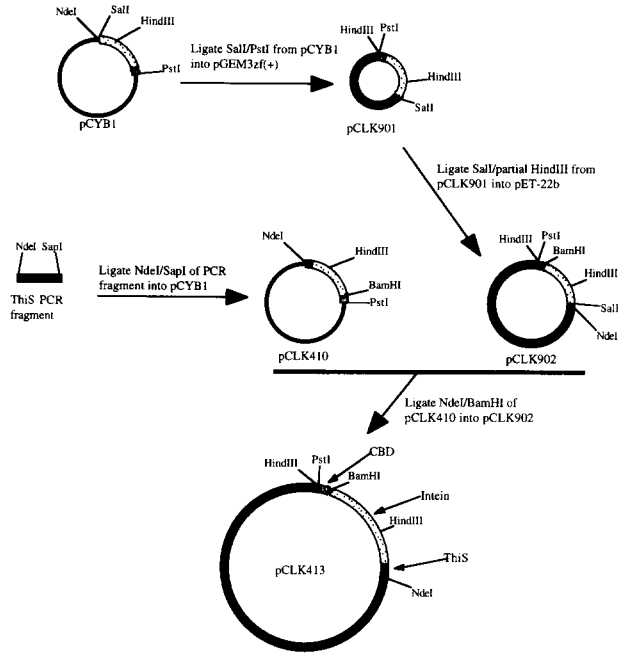

Fig. 6. Construction of the ThiS fusion overexpression vector, pCLK413.

EDTA, 30 mM cleavage reagent) and then capping the column and allowing it to stand for 14–16 h at 4 °C. For purification of ThiS–COOH, DTT was used as the cleavage reagent. For purification of ThiS–COSH, ammonium sulfide was used to induce cleavage. Cleaved target protein was eluted with 30 mL of column buffer and collected in 5 mL fractions. Fractions containing protein were concentrated in an Amicon ultrafiltration cell using a YM-3 membrane. The target protein was purified to near homogeneity by gel filtration through a 30 mL bed volume of Sephadex G-50 Medium equilibrated with TE buffer. SDS-PAGE analysis of the purification (Fig. 4) was performed according to the method of Schägger and von Jagow (Schägger & von Jagow, 1987).

Mass spectrometric analysis of ThiS–COSH: ThiS–COSH (15 μL, 7 mg/mL) was thawed, diluted with water (200 μL), and acidified to pH 3 by the addition of glacial acetic acid (2 μL). The mixture was loaded onto a reverse-phase peptide trap (Michrom BioResources, Inc., Auburn, California). The trap was washed with 1 mL of 98% H_2O/1% MeOH/1% acetic acid, and the protein was eluted with 150 μL 80% methanol/20% acetic acid (the first 50 μL was discarded). Two microliters of this eluent was loaded into a borosilicate glass capillary (1.5 mm i.d.) pulled to a 2 μm tip at one end (Wilm & Mann, 1996). 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 6 T Fourier-transform mass spectrometer as described previously (McLafferty, 1994). 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.

Mass spectrometric analysis of iodoacetic acid-derivatized ThiS–COSH: ThiS–COSH (15 μL, 7 mg/mL, 14 nmol in TE buffer) and iodoacetic acid (15 μL, 6.67 mM, 100 nmol in water) were mixed and incubated for 15 min at 25 °C, then diluted with water (200 μL) and acidified to pH 3 by addition of glacial acetic acid (2 μL). The mixture was loaded onto a reverse-phase peptide trap (Michrom BioResources, Inc., Auburn, California). The trap was washed with 1 mL of 98% H_2O/1% MeOH/1% acetic acid, and the protein was eluted with 150 μL 80% methanol/20% acetic acid (the first 50 μL was discarded). Two microliters of this eluent was loaded into a borosilicate glass capillary and electrosprayed as described above.

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

References

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K, eds. 1987. *Current protocols in molecular biology*. New York: John Wiley & Sons.

Barany G, Merrifield RB. 1979. Solid-phase peptide synthesis. In: Gross E, Meienhofer J, eds. *The peptides*, vol. 2. New York: Academic Press. pp 1–284.

Begley TP. 1996. The biosynthesis and degradation of thiamin (vitamin B1) (January 1986 to January 1996). *Natural Product Rep 13*:177–186.

Chong S, Mersha FB, Comb DG, Scott ME, Landry D, Vence LM, Perler FB, Benner J, Kucera RB, Hirvonen CA, Pelletier JJ, Paulus H, Xu MQ. 1997. Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene (Amsterdam) 192*:271–281.

Liu C-F, Rao C, Tam JP. 1996. Acyl disulfide-mediated intramolecular acylation for orthogonal coupling between unprotected peptide segments. Mechanism and application. *Tetrahedron Lett 37*:933–936.

McLafferty FW. 1994. High-resolution tandem FT mass spectrometry above 10 kDa. *Acc Chem Res 27*:379–386.

Muir TW, Dawson PE, Kent SBH. 1997. Protein synthesis by chemical ligation of unprotected peptides in aqueous solution. In: Fields GB, ed. *Methods in enzymology*, vol. 289. San Diego: Academic Press, Inc. pp 266–298.

Perler F, Xu MQ, Paulus H. 1997. Protein splicing and autoproteolysis mechanisms. *Curr Opin Chem Biol 1*:292–299.

Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning: A laboratory manual*. Plainview, NY: Cold Spring Harbor Laboratory Press.

Schägger H, von Jagow G. 1987. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal Biochem 166*:368–379.

Shao Y, Paulus H. 1997. Protein splicing: Estimation of the rate of O—N and S—N acyl rearrangements, the last step of the splicing process. *J Peptide Res 50*:193–198.

Shao Y, Xu MQ, Paulus H. 1995. Protein splicing: Characterization of the aminosuccinimide residue at the carboxyl terminus of the excised intervening sequence. *Biochemistry 34*:10844–10850.

Tam JP, Lu Y-A, Liu C-F, Shao J. 1995. Peptide synthesis using unprotected peptides through orthogonal coupling methods. *Proc Natl Acad Sci USA 92*:12485–12489.

Taylor SV, Kelleher NL, Kinsland C, Chia H-J, Costello CA, Backstrom AD, McLafferty FW, Begley TP. 1998. Identification of ThiS thiocarboxylate as the immediate sulfur donor in the thiazole formation. *J Biol Chem 273*:16555–16560.

Vander Horn PB, Backstrom AD, Stewart V, Begley TP. 1993. Structural genes for thiamine biosynthetic enzymes (thiCEFGH) in *Escherichia coli* K-12. *J Bacteriol 175*:982–992.

Wilm M, Mann M. 1996. Analytical properties of the nanoelectrospray ion source. *Anal Chem 68*:1–8.

Xu MQ, Comb DG, Paulus H, Noren CJ, Shao Y, Perler FB. 1994. Protein splicing: An analysis of the branched intermediate and its resolution by succinimide formation. *EMBO J 13*:5517–5522.

Zhang L, Tam JP. 1997. Orthogonal coupling of unprotected peptide segments through histidyl amino terminus. *Tetrahedron Lett 38*:3–6.