Chemistry & Biology *13*, 1183–1191, November 2006 DOI 10.1016/j.chembiol.2006.09.012

Enzymatic Generation of the Antimetabolite γ,γ-Dichloroaminobutyrate by NRPS and Mononuclear Iron Halogenase Action in a Streptomycete

Masashi Ueki,[1,2,3] Danica P. Galonić,[2,3] Frédéric H. Vaillancourt,[2,3,4] Sylvie Garneau-Tsodikova,[2,5] Ellen Yeh,[2] David A. Vosburg,[2,6] Frank C. Schroeder,[2] Hiroyuki Osada,[1] and Christopher T. Walsh[2,*]

[1] Antibiotics Laboratory
RIKEN Institute
2-1 Hirosawa
Wako
Saitama 351-0198
Japan

[2] Department of Biological Chemistry
and Molecular Pharmacology
Harvard Medical School
240 Longwood Avenue
Boston, Massachusetts 02115

Summary

Four adjacent open reading frames, *cytC1–C4*, were cloned from a cytotrienin-producing strain of a *Streptomyces* sp. by using primers derived from the conserved region of a gene encoding a nonheme iron halogenase, CmaB, in coronamic acid biosynthesis. CytC1–3 were active after expression in *Escherichia coli*, and CytC4 was active after expression in *Pseudomonas putida*. CytC1, a relatively promiscuous adenylation enzyme, installs the aminoacyl moieties on the phosphopantetheinyl arm of the holo carrier protein CytC2. CytC3 is a nonheme iron halogenase that will generate both γ-chloro- and γ,γ-dichloroaminobutyryl-*S*-CytC2 from aminobutyryl-*S*-CytC2. CytC4, a thioesterase, hydrolytically releases the dichloroaminobutyrate, a known streptomycete antibiotic. Thus, this short four-protein pathway is likely the biosynthetic source of this amino acid antimetabolite. This four-enzyme system analogously converts the *proS*-methyl group of valine to the dichloromethyl product regio- and stereospecifically.

Introduction

In recent studies on the biosynthesis of the coronamic acid moiety, 1-amino-1-carboxy-2-ethyl cyclopropane, of the pseudomonal phytotoxin coronatine, we have determined that the cyclopropane ring is constructed by a cryptic chlorination pathway [1]. Among the novel enzymatic features are a nonheme mononuclear iron halogenase (CmaB) that acts at the unactivated CH_3 of an L-*allo*-Ile moiety while it is tethered in thioester linkage to the pantetheinyl arm of a carrier protein, CmaD [1]. The γ-chloro-aminoacyl thioester then is acted on by CmaC, which generates an α-carbanion equivalent for intramolecular displacement of the γ-Cl, producing the cyclopropyl aminoacyl scaffold still tethered to the prosthetic arm of the carrier protein. The last step is hydrolytic release of the coronamic acid, mediated by the thioesterase CmaT [2].

The ansa-bridged cytotrienin (Figure 1) [3], produced by a soil bacteria *Streptomyces* sp. strain, contains a 1-aminocyclopropane-1-carboxylate (ACC) moiety, reminiscent of coronamic acid. Feeding studies show that the ACC moiety in this producer is derived from the aminobutyryl moiety of L-methionine [4]. In plants, ACC is biosynthesized from *S*-adenosylmethionine by the pyridoxal phosphate (PLP)-dependent ACC synthase [5], and this product serves as a precursor to the plant hormone ethylene [6]. However, no such pathway has been detected in prokaryotic microorganisms [4]. To evaluate whether similar logic and enzymatic machinery to that of coronamic acid biosynthesis is in play for ACC biogenesis in the cytotrienin producer, we have used PCR primers homologous to the conserved sequences in the CmaB halogenase gene to look for a comparable ORF. Herein, we report the cloning and sequencing of four adjacent ORFs, named *cytC1–C4*, as well as their heterologous expression and functional characterization. CytC1 is a free-standing adenylation enzyme for amino acids, including aminobutyrate. CytC2 is a 10 kDa peptidyl carrier protein that can be primed with phosphopantetheine and loaded with various amino acids. CytC3 is a mononuclear nonheme iron halogenase that can doubly chlorinate aminobutyryl-*S*-CytC2 as well as valyl-*S*-CytC2 at the unactivated γ-carbon. CytC4 is a thioesterase and cleaves the product amino acid from the pantetheine tether. The net action of these four tandemly encoded proteins is the generation of the known *Streptomyces* antimetabolite γ,γ-dichoroaminobutyrate [7]. Despite additional sequencing and cloning efforts, no cyclopropane-forming homolog has yet been detected in the cytotrienin producer.

Results

The CytC Gene Cluster

A gene fragment homologous to *cmaB* [1, 8] was used to clone out genes involved in the loading and halogenation of amino acids in this producer. This cloning strategy identified a DNA fragment containing four tandemly arranged genes, termed *cytC1*, *cytC2*, *cytC3* (the gene homologous to *cmaB*), and *cytC4* (Table 1). All of the protein products represent different nonribosomal peptide synthetase (NRPS) domains. CytC1 is predicted to function as an adenylation (A) domain, CytC2 as a phosphopantetheine (PP)-binding (PCP) domain, CytC3 as an aliphatic halogenase, and CytC4 as a thioesterase (TE) domain (Figure 2A). CytC3 is homologous to the

*Correspondence: christopher_walsh@hms.harvard.edu
[3] These authors contributed equally to this work.
[4] Present address: Department of Biological Sciences, Research and Development, Boehringer Ingelheim (Canada) Ltd., Laval, Quebec H7S 2G5, Canada.
[5] Present address: Department of Medicinal Chemistry in the College of Pharmacy and Life Sciences Institute, University of Michigan, Ann Arbor, Michigan 48109.
[6] Present address: Department of Chemistry, Harvey Mudd College, Claremont, California 91711.

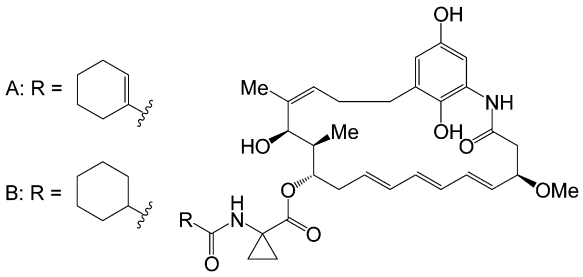

Figure 1. Cytotrienin A and Cytotrienin B

small but growing class of nonheme iron halogenases [1, 9–12]. It possesses the key active site residues characteristic of this class of enzyme (Figure 2B). The typical carboxylate ligand of nonheme iron hydroxylases is replaced by an alanine, the key feature of the nonheme iron halogenases. The replacement of the carboxylate ligand by an alanine allows for the binding of exogenous chloride to the iron, as shown in the crystal structure of the syringomycin halogenase SyrB2 [12, 13].

Substrate Specificity of CytC1

Most A domains in NRPS modules are highly specific for one amino acid. For example, CmaA, which is an A and T didomain protein involved in coronamic acid biosynthesis, has high specificity for L-*allo*-Ile and accepts other aliphatic amino acids with less than 8% efficiency when compared to L-*allo*-Ile [1]. Several A domains that activate aliphatic amino acid substrates have a slightly broader specificity. For example, LicC from lichenysin synthetase accepts not only L-Ile, but also L-Leu and L-Val [14]. It is known that the substrate specificity of NRPS A domains is determined by the 8 residues around the substrate-binding pocket [15]. No direct similarity of these key residues in CytC1 with other A domains was observed. However, this set is most similar to residues in those A domains that activate L-Val, L-Ile, and L-Leu. CytC1 might therefore prefer aliphatic amino acids.

The His-tagged CytC1 was overproduced in *Escherichia coli* as a soluble protein and was purified by using the Ni-NTA resin. The first half-reaction of CytC1, the reversible formation of aminoacyl-AMP, was followed by amino acid-dependent exchange of $^{32}PP_i$ into ATP (Table 2). CytC1 activated different amino acids in the following order of specificity: L-Val > ACC > D-Aba > D-Val > 4-Cl-L-Aba > L-Aba > L-*allo*-Ile ~ D-Cys > L-Cys > L-Ile ~ L-Leu. Only two other proteinogenic amino acids, L-Thr and L-Ala, were slightly activated by CytC1 (23% and 5% of the L-Aba rate, respectively). All other proteinogenic amino acids were tested and showed rates of 0.1%–0.8% when compared to L-Aba. The surprising ability of both D-Val and D-Aba to be activated to the aminoacyl-AMP as well or even better than the L-isomer does have a precedent in the activation of both L- and D-Phe by the first module of tyrocidine synthetase [15, 16]. ACC with Cα substituents mimicking both an D and an L center has a rate of ~25% of that of the best substrate, L-Val, in this first half of the CytC1 reaction.

Loading of Amino Acids onto CytC2

The second half-reaction of CytC1 is the transfer of the activated amino acid (aminoacyl-AMP) to the phosphopantetheinyl moiety of the T domain CytC2. N-terminally His-tagged CytC2 overproduced in *E. coli* was obtained as a soluble protein and was purified by using the Ni-NTA resin. The holo (phosphopantetheinylated) form of the T domain was generated in situ via the action of phosphopantetheinyl transferase Sfp [17] on purified CytC2. Using universally labeled ^{3}H-coenzyme A, incorporation of the radiolabel into CytC2 was monitored, and it was demonstrated to reach a plateau within 5 min

Table 1. Designations and Functions of the Genes Inserted in pSPHE02

Predicted Polypeptide	Amino Acids/MW (kDa)	Predicted Function	Most Homologous Gene (Organism)	E Value	Homologous Genes in *S. avermitilis*[a]	Homologous genes in *S. coelicolor*[a]
Orf1	708/75.4	Nitrate reductase catalytic subunit	Nitrate reductase catalytic subunit (*Amycolatopsis mediterranei*)	0	SAV2330	SCO2473
Orf2	462/48.4	Nitrate extrusion protein	Putative nitrate extrusion protein (*Streptomyces avermitilis*)	e^{-109}	SAV5119	SCO2959
Orf3	329/33.9	Unknown	Hypothetical transport protein yyaM (*Bacillus subtilis*)	$3e^{-18}$	NE	NE
CytC1	524/56.8	NRPS adenylation domain	—[b]	—	—	—
CytC2	87/9.42	NRPS PP-binding domain	—[b]	—	—	—
CytC3	319/36.4	Nonheme iron halogenase	Nonheme iron halogenase SyrB2 (*Pseudomonas syringae* pv. *Syringae*)	e^{-100}	NE	NE
CytC4	244/25.9	NRPS thioesterase domain	Putative thioesterase SAV3201 (*Streptomyces avermitilis*)	$2e^{-16}$	SAV3201	SCO6287
Orf4	315/33.6	ABC-type sugar transporter	Putative transport system integral membrane protein SAV2145 (*Streptomyces avermitilis*)	e^{-114}	SAV2145	SCO6086
Orf5	284/30.6	ABC-type sugar transporter	Putative transport system integral membrane protein SCO6087 (*Streptomyces coelicolor* A3(2))	$5e^{-99}$	SAV2144	SCO6087

Based on predicted amino acid sequences, the putative functions of the predicted proteins are shown with homologous genes. NE: no homologous gene with an E value less than e^{-10}.
[a] Most homologous genes of *Streptomyces avermitilis* MA-4680 and *S. coelicolor* A3(2).
[b] Homologous genes are not listed because too many genes are found in the database.