Cloning and Characterization of the Biosynthetic Gene Cluster for Tomaymycin, an SJG-136 Monomeric Analog

Wei Li, ShenChieh Chou, Ankush Khullar, and Barbara Gerratana*
Department of Chemistry and Biochemistry, University of Maryland, College Park, Maryland 20742

Received 9 October 2008/Accepted 25 February 2009

Tomaymycin produced by *Streptomyces achromogenes* is a naturally produced pyrrolobenzodiazepine (PBD). The biosynthetic gene cluster for tomaymycin was identified and sequenced. The gene cluster analysis reveals a novel biosynthetic pathway for the anthranilate moiety of PBDs. Gene replacement and chemical complementation studies were used to confirm the proposed biosynthetic pathway.

Naturally produced pyrrolobenzodiazepines (PBDs) are monomeric compounds with weak antibiotic properties that specifically alkylate the minor groove of DNA at a 5'-Pu-G-Pu base sequence (27). Chemical diversity among the PBDs is provided by different degrees and types of substituents at the A and C rings (Fig. 1A). Ring A can be substituted at different positions with methyl, hydroxy, and methoxy groups or with sibirosamine moiety (Fig. 1A). However, two major groups of naturally produced PBDs, those with a C-9 hydroxyl substituent and those without, can be identified (Fig. 1A). Ring C can be fully saturated, unsaturated at the C-2–C-3 bond, or exocyclically unsaturated at C-2 as in neothramycin, sibiromycin, and tomaymycin, respectively (Fig. 1A). The chemical liability of the imine bond complicates the synthesis of PBDs, limiting the complexity of PBDs compounds synthetically accessible (2, 16, 17, 29). Interest in the PBDs is driven by the remarkable antitumor properties of these compounds such as sibiromycin and tomaymycin (28).

Structural activity studies (17) led to the synthesis of SJG-136, a PBD dimer (Fig. 1B) which has been shown to reduce drastically the mass of 10 different xenografts (1). Phase I clinical trials of SJG-136, recently completed, showed promising results in the treatment of ovarian cancer (http://dtp.nci.nih.gov/timeline/notflash/success_stories/S7_SJG-136.htm). Tomaymycin produced by *Streptomyces achromogenes* is similar to the monomeric unit of SJG-136 (Fig. 1B). Both tomaymycin and the monomeric unit of SJG-136 are substituted at C-7 with a methoxide and are O hydroxylated at C-8, and both contain a C-2-α-unsaturated tetrahydropyrrole moiety. Tomaymycin is also significantly cytotoxic against leukemia, plasmacytoma, and ovarian cancer cell lines with 50% inhibitory concentrations of 3.7, 1.8, and 0.13 nM, respectively (28).

To clone the tomaymycin gene cluster, we assumed that the diazepine ring formation is catalyzed by nonribosomal peptide synthetase (NRPS) enzymes as in anthramycin biosynthesis (9), and we designed primers specific for two consensus regions (A3 and A8) of the adenylation domain of NRPS (21). The DNA amplified from *S. achromogenes* genomic DNA encoded a protein with an amino acid sequence 52% identical to ORF21 of the anthramycin biosynthesis (9) and was used to screen a cosmid library of *S. achromogenes* genomic DNA. A positive clone was sequenced by a shotgun approach and subsequent primer walking to cover the gaps to give a continuous region of 44.1 kb for 35 open reading frames (ORFs) by bioinformatic analysis. Seventeen of these ORFs were assigned to the tomaymycin biosynthetic gene cluster (nucleotide sequence accession number FJ768957) (Table 1) based on the presence of similar sequences in the genomic sequence of S. achromogenes (28).

* Corresponding author. Mailing address: Department of Chemistry and Biochemistry, Bldg. 091, University of Maryland, College Park, MD 20742. Phone: (301) 405-1541. Fax: (301) 314-9121. E-mail: bgerrata@umd.edu.

† Supplemental material for this article may be found at http://aem.asm.org/.

‡ Published ahead of print on 6 March 2009.
of NRPS enzymes with high sequence identity with the NRPS enzymes in the anthramycin gene cluster (9). To confirm the assignment of the tomaymycin gene cluster, we inactivated \textit{tomA}, which encodes one of the NRPS enzymes, by gene replacement using REDIRECT technology (7). Inactivation of this gene abolished tomaymycin production (Fig. 2). Tomaymycin production in this mutant was rescued by a complementation with pUWL201b/\textit{tomA}, ruling out any polar effect, and verified by bioassay and high-pressure liquid chromatography–electrospray ionization (HPLC-ESI) analyses (Fig. 2).

Gene inactivation of \textit{orfX1}, the ORF upstream to \textit{tomA}, resulted in a strain capable of producing tomaymycin similar to the wild-type producer, as verified by bioassay and HPLC-ESI analyses. This gene encodes a putative MarR family transcriptional regulator. Because members of this family often act as repressor proteins, inactivation of \textit{orfX1} might yield to an increase in tomaymycin production (4). However, tomaymycin production did not increase in an \textit{ΔorfX1} strain (see Fig. S1 in the supplemental material). Therefore, we have assigned \textit{tomA} as the first gene in the tomaymycin gene cluster. Downstream of the gene cluster, \textit{orfX2} encodes a putative transposase, and it is unlikely to be involved in the biosynthesis of tomaymycin. Gene inactivation experiments, bioassays, and HPLC-ESI analyses confirmed this assignment. Upstream of \textit{orfX}, \textit{tomQ} encodes a putative flavin-dependent oxidoreductase with high similarity with ORF24 of the anthramycin gene cluster (see the accompanying study [18] and Table 1). Gene inactivation of \textit{tomQ} did not abolish tomaymycin production, as judged by HPLC-ESI analysis, suggesting that this enzyme is not essential for tomaymycin biosynthesis. The exact role, if any, of TomQ in tomaymycin biosynthesis, as well as of ORF24, in the anthramycin biosynthesis (18) remains unclear. TomP is a second copy of anthranilate synthetase, and it is likely involved in biosynthesis of the anthranilate moiety of tomaymycin (see below).

The tomaymycin gene cluster identified includes only one gene, \textit{tomM}, involved in resistance and no gene involved in regulation. \textit{tomM} encodes a putative resistance protein homologous to proteins identified in the sibiromycin (18) and anthramycin (9) gene clusters. Gene inactivation experiments, bioassays, and HPLC-ESI analyses confirmed this assignment. Upstream of \textit{tomM}, \textit{tomO} encodes a flavin-containing amine oxidase, which is likely involved in the biosynthesis of the flavin moiety of tomaymycin (see below).

### TABLE 1. Deduced functions of ORFs in the tomaymycin biosynthetic gene cluster

<table>
<thead>
<tr>
<th>Gene</th>
<th>Protein size (aa)</th>
<th>Putative function</th>
<th>NCBI accession no. of protein homolog</th>
<th>% Identity/similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{orfX1}</td>
<td>154</td>
<td>MarR transcriptional regulator family</td>
<td>YP_001339142</td>
<td>50/61</td>
</tr>
<tr>
<td>\textit{tomA}</td>
<td>614</td>
<td>Nonribosomal peptide synthetase</td>
<td>ABW71852 (ORF21)</td>
<td>39/52</td>
</tr>
<tr>
<td>\textit{tomB}</td>
<td>1,542</td>
<td>Nonribosomal peptide synthetase</td>
<td>ABW71853 (ORF22)</td>
<td>37/49</td>
</tr>
<tr>
<td>\textit{tomC}</td>
<td>406</td>
<td>Putative 3-deoxy-D-arabinose-heptulosonic 7-phosphate (DHAP) synthase</td>
<td>CAL34108</td>
<td>47/58</td>
</tr>
<tr>
<td>\textit{tomD}</td>
<td>677</td>
<td>Phenazine biosynthesis protein PhzE</td>
<td>YP_001348741</td>
<td>44/59</td>
</tr>
<tr>
<td>\textit{tomE}</td>
<td>206</td>
<td>Phenol-2-monoxygenase reductase component</td>
<td>YP_702781</td>
<td>36/46</td>
</tr>
<tr>
<td>\textit{tomF}</td>
<td>553</td>
<td>Phenol-2-monoxygenase oxygenase component</td>
<td>YP_702343</td>
<td>63/74</td>
</tr>
<tr>
<td>\textit{tomG}</td>
<td>234</td>
<td>O-Methyltransferase</td>
<td>YP_001644688</td>
<td>47/60</td>
</tr>
<tr>
<td>\textit{tomH}</td>
<td>145</td>
<td>1-DOPA-2,3-dioxygenase</td>
<td>CAA55747 (LmbB1)</td>
<td>54/59</td>
</tr>
<tr>
<td>\textit{tomI}</td>
<td>318</td>
<td>Tyrosine hydroxylase</td>
<td>ABX00599 (LmbB2)</td>
<td>48/61</td>
</tr>
<tr>
<td>\textit{tomJ}</td>
<td>319</td>
<td>F-420-dependent reductase</td>
<td>ABX00622 (LmbY)</td>
<td>50/63</td>
</tr>
<tr>
<td>\textit{tomK}</td>
<td>287</td>
<td>Unknown</td>
<td>CAA55771 (LmbX)</td>
<td>45/56</td>
</tr>
<tr>
<td>\textit{tomL}</td>
<td>577</td>
<td>γ-Glutamyltransferase</td>
<td>ABX080597 (LmbA)</td>
<td>50/59</td>
</tr>
<tr>
<td>\textit{tomM}</td>
<td>783</td>
<td>UvrA drug resistance pump</td>
<td>ABW71839 (ORF8)</td>
<td>66/78</td>
</tr>
<tr>
<td>\textit{tomN}</td>
<td>66</td>
<td>4-Oxalocrotonate tautomerase</td>
<td>ZP_01514863</td>
<td>44/63</td>
</tr>
<tr>
<td>\textit{tomO}</td>
<td>394</td>
<td>Salietyl-CoA hydroxylase (NADH-dependent flavin oxidoreductase)</td>
<td>YP_117454</td>
<td>25/32</td>
</tr>
<tr>
<td>\textit{tomP}</td>
<td>637</td>
<td>Anthranilate synthase</td>
<td>ABW71837 (ORF6)</td>
<td>49/58</td>
</tr>
<tr>
<td>\textit{tomQ}</td>
<td>482</td>
<td>Flavin-containing amine oxidase</td>
<td>ABW71855 (ORF24)</td>
<td>57/71</td>
</tr>
<tr>
<td>\textit{orfX2}</td>
<td>373</td>
<td>IS\textit{\textit{4}} family transposase</td>
<td>NP_821294</td>
<td>89/91</td>
</tr>
</tbody>
</table>

\textsuperscript{a} aa, amino acids.

\textsuperscript{b} The proteins encoded by the lincomycin A and anthramycin gene clusters are indicated in parentheses.

\textsuperscript{c} See the accompanying study (18) for proteins encoded by the sibiromycin gene cluster.
mycin (9) biosynthetic gene clusters, suggesting a common resistance mechanism to PBDs. In addition to tomaymycin, the wild-type strain produces a second PBD compound, oxotomaymycin, which is bio-inactive (13) (Fig. 2). Transformation of tomaymycin to oxotomaymycin was proposed to be a resistance mechanism used by the cell (11). No clear candidate encoding the functionality for this transformation was identified in the characterized gene cluster, leaving open the possibility that this gene is located outside the gene cluster.

Hurley et al. have shown that the biosynthetic precursor of the hydropyrrole moiety of lincomycin A (Fig. 1B), sibiromycin, anthramycin, and tomaymycin is L-tyrosine (3, 10). The first two steps of this transformation have been biochemically characterized in the lincomycin A biosynthesis (10, 25, 26). However, the assignment of the subsequent steps remained difficult due to the lack of significant similarity of the ORFs identified with functionally known enzymes. With the identification of the anthramycin (9), sibiromycin (see accompanying paper) and tomaymycin a comparative analysis of the gene clusters that takes into consideration chemical similarities and differences in the hydropyrrole moiety allowed us to assign the biosynthetic steps with a certain degree of confidence (Fig. 3A). We proposed based on this analysis and labeling experiments (10) that the PBDs and lincomycin A biosyntheses proceed to a common intermediate 4-vinyl-2,3-dihydropyrrole-2-carboxylic acid (Fig. 3A). Homologous enzymes are present in all gene clusters sequenced including the tomaymycin gene cluster (TomH, TomI, and TomK) (Table 1). The first two enzymes are similar to the tyrosine hydroxylase and L-DOPA 2,3-dioxygenase of the lincomycin A biosynthesis (24, 25) (Table 1). A BLAST search of TomK shows no similar enzyme except for LmbX, ORF15, and SibS (Table 1) that was proposed to catalyze the unusual C-C hydrolysis of 4-(3-carboxy-3-oxo-propenyl)-2,3-dihydropyrrole-2-carboxylic acid (8, 18).

The tetrahydropyrrole moiety of tomaymycin differs from those of lincomycin, sibiromycin, and anthramycin by the presence of an ethylidene side chain instead of a propylidene side chain. The terminal carbon of the propylidene side chain shown by labeling experiments to derive by L-methionine is likely added by an S-adenosylmethionine-dependent methyltransferase. In the accompanying study (18) we propose that SibZ from the sibiromycin gene cluster catalyzes a methyl transfer coupled to tautomerization. This assignment is further confirmed by the absence of a homologous enzyme in the tomaymycin gene cluster, which does not require this activity. In order to obtain the imine that will undergo reduction by a F-420 reductase (TomJ) (Table 1) in tomaymycin biosynthesis, we predicted the presence of a tautomerase. The identification of an enzyme, TomN, highly similar to tautomeras (44% identity) proved our hypothesis correct and validates the proposed biosynthetic pathway (Fig. 3A).

A common biosynthetic pathway starting from L-tryptophan for the anthranilate moiety of all PBDs has been previously proposed (10, 12). However, feeding experiments with labeled anthranilic acid (14) showed the incorporation of labeled atoms only in tomaymycin and not in sibiromycin, whose producer can significantly take up anthranilic acid (14). The difference in the incorporation of labeled anthranilic acid in tomaymycin and sibiromycin suggested to us the presence of different biosynthetic pathways for the anthranilate moiety.
Specifically, we reasoned that two different biosynthetic pathways are used for the biosynthesis of the PBD ring system. The first for the biosynthesis of PBDs with a C-9 hydroxyl group such as anthramycin and sibiromycin (Fig. 1) has adopted the L-tryptophan degradation biosynthetic machinery from primary metabolism for the formation of the intermediate 3-hydroxyanthranilic acid whose hydroxyl group ends up at C-9 in the PBD ring system (18). The second is used for the biosynthesis of PBDs not substituted at C-9, such as tomaymycin (Fig. 1). The tomaymycin gene cluster here reported confirms the presence of a second biosynthetic pathway for the PBD ring system in which the anthranilate moiety derives from chorismate, the product of the shikimate pathway, which is then transformed by anthranilate synthetase to anthranilic acid (30) (Table 1 and Fig. 3B). Proteins similar to anthranilate synthetase, TomD and TomP, and to 3-deoxyarabino-heptulosonate 7-phosphate (DHAP) synthase, TomC, are encoded by the tomaymycin gene cluster. DHAP synthase catalyzes the first and key regulatory step of the shikimate pathway. Thus, it is not surprising that, to produce more chorismate and bypass the regulation of the primary DHAP synthase, the tomaymycin gene cluster contains a copy of DHAP synthase. Endogenous enzymes of the shikimate pathway are proposed to catalyze the remaining steps (Fig. 3B). Similarly, the phenazine biosynthetic pathway from Pseudomonas fluorescens contains only DHAP synthase and anthranilate synthetase and utilizes the primary enzymes for the remaining steps of the shikimate pathway (22).

Hydroxylation and O methylation of anthranilic acid are likely catalyzed by the remaining ORFs identified in the tomaymycin gene clusters tomE, tomF, tomG, and tomO. TomO shares sequence similarity (25% identity) with salicylyl-coenzyme A (CoA) 5-hydroxylase shown to catalyze the hydroxylation of salicylyl-CoA to gentisyl-CoA (15). This reaction is identical to the 5-hydroxylation of anthranilic acid with the exception that the carboxylate of anthranilic acid is esterified by CoA (15). Therefore, TomO is proposed to catalyze the 5-hydroxylation of anthranilic acid in the tomaymycin gene cluster. Comparative analysis with the sibiromycin gene cluster supported this assignment that shows the presence of a TomO homolog, SibG (40% identity; Table 1) shown to catalyze the hydroxylation at C-7 of sibiromycin (see the accompanying study [18]). To confirm this assignment, we constructed a ΔtomO strain and analyzed its fermentation products by HPLC-ESI. This mutant strain is unable to produce tomaymycin, but it produces a dideoxy analog of tomaymycin (Fig. 4A), eluting at 18.2 min with the characteristic [M+H]⁺ at m/z = 227.1 for the imine form, m/z = 245.1 for the carbinolamine form, and m/z = for the 259.1 for the carbinolamine methyl ether form consistent with the molecular formula C₁₄H₁₄N₂O₁ (calculated, 226.1), C₁₄H₁₆N₂O₂ (calculated, 244.1), and C₁₅H₁₈N₂O₂ (calculated, 258.1), respectively (Fig. 4B). In ad-
dition, we have also detected the dideoxy analog of oxotomycin, suggesting that the enzyme involved in the resistance mechanism can accommodate different substrates. The production of tomaymycin was rescued by chemical complementation with 5-hydroxyanthranilic acid but not with 4-hydroxyanthranilic acid (synthesized as described in reference 23) (Fig. 4A). These results strongly implicate TomO in the 5-hydroxylation of anthranilic acid that precedes C-4 hydroxylation (Fig. 3B). Therefore, we propose that TomE and TomF, similar to the reductase and oxygenase components, respectively, of phenol-2-monoxygenase (Table 1), catalyze the C-4 hydroxylation of 5-hydroxyanthranilic acid. There is only one gene, tomG, in the tomaymycin gene cluster that encodes a protein with similarity with a methyltransferase. Whether TomG methylates the hydroxyl group before or after formation of the pyrrolobenzodiazepine ring remains to be determined (Fig. 3B).

We are grateful to B. Gust (University of Tubingen) for Escherichia coli ET12567 cells and plasmids pIJ790, pIJ773, and pUWL201.

This study was supported by the Marlene and Stewart Greenebaum Cancer Center’s American Cancer Society Institutional Research Grant (01-4-30170), by the National Institutes of Health (GM084473), and by startup funds from the University of Maryland, College Park.

REFERENCES


5. Reference deleted.

6. Reference deleted.


FIG. 4. (A) HPLC and bioassay analyses of the secondary metabolites produced by wild-type and ΔtomO strains and by chemical complementation of the ΔtomO strain with 5-hydroxyanthranilic acid. (B) ESI spectra of the peaks labeled 1 and 2 in the HPLC chromatogram from panel A. The dideoxy analog of tomaymycin has much weaker antibiotic properties than tomaymycin.


19. Reference deleted.

20. Reference deleted.