**Citation**

**As Published**
http://dx.doi.org/10.1073/pnas.1009003107

**Publisher**
National Academy of Sciences (U.S.)

**Version**
Final published version

**Accessed**
Sat Jun 17 18:15:13 EDT 2017

**Citable Link**
http://hdl.handle.net/1721.1/64435

**Terms of Use**
Article is made available in accordance with the publisher’s policy and may be subject to US copyright law. Please refer to the publisher’s site for terms of use.

**Detailed Terms**

---

A homolog of tocopherol C-methyltransferases catalyzes N-methylation in anticancer alkaloid biosynthesis

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.
Homolog of tocopherol C methyltransferases catalyzes N methylation in anticancer alkaloid biosynthesis

David K. Liscombe, Aimee R. Usera, and Sarah E. O’Connor

Department of Chemistry, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139

Edited* by Jerrold Meinwald, Cornell University, Ithaca, NY, and approved September 17, 2010 (received for review June 23, 2010)

Madagascar periwinkle (Catharanthus roseus) is the sole source of the anticancer drugs vinblastine and vincristine, bisindole alkaloids derived from the dimerization of the terpenoid indole alkaloids vindoline and catharanthine. Full elucidation of the biosynthetic pathways of these compounds is a prerequisite for metabolic engineering efforts that will improve production of these costly molecules. However, despite the medical and commercial importance of these natural products, the biosynthetic pathways remain poorly understood. Here we report the identification and characterization of a C. roseus cDNA encoding an S-adenosyl-L-methionine-dependent N methyltransferase that catalyzes a nitrogen methylation involved in vindoline biosynthesis. Recombinant enzyme produced in Escherichia coli is highly substrate specific, displaying a strict requirement for a 2,3-dihydro bond in the aspidosperma skeleton. The corresponding gene transcript is induced in methyl jasmonate-elicited seedlings, along with the other known vindoline biosynthetic transcripts. Intriguingly, this unique N methyltransferase is most similar at the amino acid level to the plastidic γ-tocopherol C methyltransferases of vitamin E biosynthesis, suggesting an evolutionary link between these two functionally disparate methyltransferases.

Madagascar periwinkle (Catharanthus roseus) plants produce over 100 monoterpenoid indole alkaloids derived from tryptamine and the iridoid terpene secologanin (1). These products include the bisindole alkaloids vinblastine 9a and vincristine 9b, which are derived through the dimerization of vindoline 7a, an aspidosperma-type alkaloid, and catharanthine 8 monomers (Fig. 1). Vinblastine 9a and vincristine 9b are microtubule disruptors that have been used extensively in the treatment of several types of cancer, including leukemia and lymphoma (2). Although yields of vincristine 9a and vinblastine 9b are approximately 0.0003% and 0.01%, respectively, from dried leaves of C. roseus (3), this plant nevertheless remains the only commercial source for these drugs. No other known plant species produces these compounds, and total syntheses of vinblastine 9a and vincristine 9b (for representative examples, see refs. 4 and 5) have not yet been adopted as industrial production methods.

Despite the medicinal and commercial importance of these compounds, several aspects of bisindole alkaloid biosynthesis are still not understood. Identification of the biosynthetic genes of this pathway is required for rational metabolic engineering efforts to improve production levels in the plant, explore heterologous production platforms, and to reprogram the pathway to make useful alkaloid derivatives. Although a number of the genes encoding vindoline 7a biosynthetic enzymes have been cloned and characterized (6–9), genes encoding the hydrating enzyme or N methyltransferase (NMT) involved in converting tabersonine 1 to vindoline 7a have not been identified (Fig. 1). Many of the validated biosynthetic enzymes are represented in the collection of approximately 20,000 ESTs for C. roseus available in GenBank (10, 11). Additional genes corresponding to unknown steps might also be represented in this dataset. Pioneering work by De Luca and colleagues demonstrated that the NMT of vindoline 7a biosynthesis is associated with thylakoid membranes of chloroplasts (12–14). Therefore, when we identified a number of candidate cDNAs that encode putative methyltransferases, we paid particular attention to a subset of these genes that also displayed sequence similarity to plastidial γ-tocopherol methyltransferases (γ-TMTs) (15–17). Functional expression of these...
candidate cDNAs led to the identification of a \( \gamma \)-TMT homolog that methylates the aromatic nitrogen of a model aspidosperma substrate and can selectively methylate proposed native substrates present in seedling extracts.

**Results and Discussion**

**Identification of \( N \)-Methyltransferase Candidates.** We relied on two distinguishing features of the NMT of vindoline 7a biosynthesis to guide the gene discovery process. First, this enzyme is associated with thylakoid membranes of chloroplasts (12, 13, 18), and, thus, we expected the translation product of the NMT gene to possess a transit peptide, or to resemble known plastid-localized proteins. It is also well established that vindoline 7a biosynthesis occurs primarily in the leaves (7, 11), so we focused on those candidates that were expressed in this biosynthetically active tissue. *C. roseus* EST assemblies on the PlantGDB server (http://plantgdb.org/cgi-bin/blast/PlantGDBblast) were mined using tBLASTn, and all annotated plant NMT sequences in GenBank were used as queries. This initial list of NMT candidates derived from BLAST results was manually edited, where contigs that were identical or nearly identical to methyltransferases of known function (i.e., protein methyltransferases) were removed, to provide the master candidate list presented in Table S1. A phylogenetic tree (Fig. 2) illustrates the relationship between *C. roseus* NMT candidates and selected, functionally characterized *S*-adenosyl-L-methionine (SAM)-dependent enzymes from plants and bacteria. Predicted translation products of Cr5804, Cr6424, Cr8458, and Cr3910022 were too short to be included in the tree. We noted that five of our candidates (Cr1196, Cr2270, Cr7756, Cr3910017, and Cr3910022) were homologous to \( \gamma \)-TMTs, enzymes that are known to reside in plastids, where they participate in tocopherol biosynthesis (15, 17). Phylogenetic analysis places these candidates within a distinct clade that includes functionally characterized plant \( \gamma \)-TMT enzymes (Fig. 2) (15–17). The other analyzed candidates fall within a clade representing functionally diverse type-I and -II methyltransferases (Fig. 2) (19), including tabersonine 16-\( O \) methyltransferase, \( O \) methyltransferases of benzyloquinoline and ipecac alkaloid biosynthesis, and an anthranilate \( N \) methyltransferase. The apparent full-length cDNAs of the latter candidate group do not encode putative chloroplast transit peptides, according to ChloroP (20), and, therefore, the corresponding polypeptides are not expected to be plastidic. The \( \gamma \)-TMT homologs Cr2270 and Cr7756 appeared to code for full-length proteins, lacking only the 5′ region that would be expected to encode transit peptides (Fig. S1). However, Cr2270 was the only apparent full-length \( \gamma \)-TMT homolog that is expressed in leaves (Table S1), and we were unable to amplify the Cr7756 ORF from leaf or seedling cDNA. ORFs encoding Cr2270, and Cr1196 (both lacking putative transit peptide region), as well as Cr6996, and Cr7789, were amplified from elicited seedling cDNA and expressed in *E. coli*. Recombinant Cr2270, Cr1196, and Cr6996 proteins were purified to homogeneity and used for NMT assays (Fig. S2). As previously reported (21), we also observed that recombinant Cr7789 was insoluble and could not be purified.

**Recombinant Cr2270 Catalyzes Aromatic Nitrogen Methylation of 2,3-Dihydrotabersonine.** The known enzymes of vindoline 7a biosynthesis are highly substrate specific (7, 22), which presumably ensures that the enzymes act in a strictly defined sequence. De Luca and coworkers previously reported that the NMT of vindoline 7a biosynthesis, partially purified from *C. roseus* leaves, only accepted aspidosperma substrates in which the 2,3 alkene (see Fig. 1 for numbering) is reduced and the 6,7 alkene is intact (13, 18). The model substrate (2R,3S)-2,3-dihydrotabersonine (DHT) 10 was among the few substrates readily turned over by this enzyme (13, 18), and because this compound is synthetically accessible through sodium cyanoborohydride reduction of tabersonine 1 (Fig. 3A), we used this as a test substrate to screen for the desired NMT activity. The structure of the synthetic DHT 10 substrate was confirmed by \( ^1 \)H-NMR (Fig. 3B; full spectrum in Fig. S3A), \( ^13 \)C-NMR, and COSY-NMR (Fig. S3 B and C). Analysis of enzyme assays by liquid chromatography-mass spectrometry (LC-MS) revealed that in the presence of SAM, Cr2270 converted DHT 10 to a more hydrophobic compound.

![Fig. 2.](image)

Unrooted, amino acid similarity tree of plant and bacterial S-adenosyl-L-methionine dependent methyltransferases constructed using the Neighbor-Joining method. Bootstrap frequencies (as percentage of 1,000 iterations) are shown for each clade. Protein names and accession numbers are provided in Table S3.
that displayed the expected molecular weight of N-methyl-DHT 11 (Fig. 3C). Furthermore, an authentic standard of N-methyl-DHT 11 that was synthesized by treatment of DHT 10 with sodium hydride and iodomethane (Fig. 3A and B and Fig. S3 D–E) coeluted with the enzymatic product (Fig. 3C). In contrast, recombinant Cr1196 or Cr6996 was not capable of methylating DHT 10 when SAM was included in the assays. Steady-state kinetic analyses for Cr2270 showed that DHT 10 evidently serves as an excellent model substrate for Cr2270, exhibiting a $K_m$ of 8.8 μM and a $k_{cat}$ of 2.4 s$^{-1}$ (Table 1 and Fig. S4 A and B). As expected, Cr2270 is sensitive to inhibition by S-adenosylhomocysteine (SAH) ($K_i = 7.3$ μM) (Fig. S4).

Consistent with the previously reported substrate specificity for the NMT of vindoline 7a biosynthesis, Cr2270 does not methylate tabersonine 1 (13, 18). A variety of other alkaloids were assayed with recombinant Cr2270, but none were converted to any product as evidenced by monitoring the disappearance of starting material and formation of any product by LC-MS (Fig. S5). The failure of Cr2270 to turn over tabersonine 1 indicates that the 2,3-single bond of the aspidosperma skeleton is a critical recognition element for the enzyme. Without this functionality, the substrate is not recognized by the enzyme in vitro, presumably due to conformational differences between the substrates, or because the nitrogen of DHT 10 is more electron rich than that of tabersonine 1. The known alkaloid products of C. roseus—specifically the lack of N-methylated tabersonine—suggest that the substrate specificity is the same in vivo. In contrast, the 16-methoxy substitution on the indole ring does not appear to qualitatively affect the action of this NMT.

The substrate specificity of this NMT undoubtedly plays a role in determining the order of enzymatic transformations and defining the complement of alkaloids produced by C. roseus plants. The substrate specificity of subsequent steps is also consistent with the order of vindoline 7a biosynthesis (Fig. 1): D4H activity relies on the presence of the N-methyl group and reduced 2,3-bond (22), and deacetylvindoline acetyltransferase (DAT) clearly requires the 4-hydroxy group for acetylation. Although N desformylvincrivestine (N desmethylvinblastine) has been isolated, it is likely that this compound arises from deacylation of vincristine 9b in planta or during the isolation process (23), or via demethylation of vinblastine. The inherent substrate specificity of the biosynthetic enzymes has obvious metabolic engineering implications. Currently, all known N-methylated aspidosperma alkaloids in C. roseus lack the 2,3-double bond; understanding the molecular basis of substrate recognition by this enzyme may allow N methylation of a wider range of alkaloid intermediates for production of many previously undescribed alkaloid structures.

Recombinant Cr2270 Methylates the Proposed Native Substrates in Seedling Extracts. The precursors to vindoline 7a and vindorosine 7b are proposed to serve as native substrates for this NMT (Fig. 1), and the known alkaloid products of C. roseus seedlings increased over the course of a 10-d methyl jasmonate treatment (22), and deacetylvindoline acetyltransferase (DAT) clearly requires the 4-hydroxy group for acetylation. Although N desformylvincrivestine (N desmethylvinblastine) has been isolated, it is likely that this compound arises from deacylation of vincristine 9b in planta or during the isolation process (23), or via demethylation of vinblastine. The inherent substrate specificity of the biosynthetic enzymes has obvious metabolic engineering implications. Currently, all known N-methylated aspidosperma alkaloids in C. roseus lack the 2,3-double bond; understanding the molecular basis of substrate recognition by this enzyme may allow N methylation of a wider range of alkaloid intermediates for production of many previously undescribed alkaloid structures.

Expression of Cr2270 Transcripts Is Coordinateley Regulated with Known Alkaloid Biosynthetic Genes. C. roseus seedlings produce alkaloids as part of their normal developmental program, and this accumulation can be enhanced by methyl jasmonate (25). As previously demonstrated, we noted the amount of vindoline 7a found in seedlings increased over the course of a 10-d methyl jasmonate

Table 1. Apparent kinetic parameters for recombinant Cr2270

<table>
<thead>
<tr>
<th>Substrates</th>
<th>DHT</th>
<th>SAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_m$ (μM)</td>
<td>8.8 ± 1.0</td>
<td>22.0 ± 1.8</td>
</tr>
<tr>
<td>$V_{max}$ (pmol s$^{-1}$ mg$^{-1}$ protein)</td>
<td>67.0 ± 1.8</td>
<td>151.6 ± 3.9</td>
</tr>
<tr>
<td>$k_{cat}$ (s$^{-1}$)</td>
<td>2.4 ± 0.1</td>
<td>5.4 ± 0.1</td>
</tr>
<tr>
<td>$k_{cat}/K_m$ (s$^{-1}$ M$^{-1}$)</td>
<td>267,744</td>
<td>243,053</td>
</tr>
</tbody>
</table>

Liscombe et al.
elicitation when compared to nonelicited controls (Fig. 5A). We therefore used qPCR to track the expression levels of Cr2270, in addition to the four previously characterized vindoline biosynthetic genes. We noted that all known vindoline biosynthetic genes and the NMT Cr2270 gene were up-regulated relative to the nonelicited seedlings (Fig. S6). The expression profile for Cr2270 exhibits the same pattern as 16OMT, suggesting that expression of these genes is coordinately regulated (Fig. 5B).

In contrast, expression of the putative methytransferase gene Cr1196, which appears to have no role in alkaloid biosynthesis, was not up-regulated by methyl jasmonate (Fig. 5B). The coordinate regulation of Cr2270 expression with other vindoline biosynthetic genes in response to methyl jasmonate, along with the lack of transcriptional induction of a highly similar methytransferase that is not involved in alkaloid biosynthesis, provides further support for the role of Cr2270 as the 16-methoxy-2,3-dihydro-3-hydroxytabersonine N-methyltransferase in vindoline biosynthesis.

**On the Evolution of N-Methyltransferase Function.** Sequence analysis of Cr2270 and related proteins (Fig. 2 and Fig. S1) suggest that Cr2270 represents a previously undescribed class of NMT that likely arose from γ-TMTs. *Arabidopsis thaliana* possesses a single gene that encodes a functional γ-TMT (17), whereas *Brassica napus* (26) and *C. roseus* both have at least four γ-TMT-like genes. In the latter case, we now know that members of this gene family encode enzymes that are functionally distinct. We speculate that Cr2270 arose from γ-TMT gene duplication events, which would explain why these gene products reside in the same subcellular compartment. Moreover, although DHT methylation by Cr2270 is inhibited by γ-tocopherol (Fig. S7A), γ-tocopherol is not turned over by this enzyme (Fig. S7B). Perhaps the evolution of substrate specificity was facilitated by the presence of nitrogen-containing γ-tocopherol derivatives, such as 5-nitro-γ-tocopherol, which is generated in diverse plant systems (27). Rubisco large subunit methyltransferase (RubLSMT) is known to modulate γ-TMT activity through methylation of a C-terminal lysine residue (28). Interestingly, the RubLSMT consensus motif (28) is conserved in γ-TMTs, and Cr1196, but has been lost in Cr2270, Cr3710017, and Cr7756 (Fig. S1).

**Conclusion**

The discovery of this NMT further closes the gap in the tabersonine to vindoline branch of monoterpene indole alkaloid biosynthesis, with only one uncharacterized step remaining in this pathway. This discovery will facilitate metabolic engineering efforts for both enhanced production of bisindole alkaloids in plants or in heterologous expression systems. Moreover, these data will allow a detailed study of the substrate specificity of this enzyme, which controls the production of the distinct alkaloid structures produced naturally. Furthermore, the amino acid sequence of this enzyme provides a suggestion as to how one
piece of this specialized metabolic pathway evolved from an entirely distinct metabolic pathway.

Materials and Methods

Plant Growth. Catharanthus roseus cv. ‘Little Bright Eyes’ seeds were purchased from B & T World Seeds (http://b-t-worldseeds.com). Sterile seedlings were cultured and treated with methyl jasmonate according to Aerts et al. (25).

Chemicals. Tabersonine was generously provided by Viresh Rawal (University of Chicago, Chicago). Unless otherwise noted, all other chemicals were obtained from Sigma-Aldrich.

Synthesis of (2R,3S)-2,3-Dihydrotabersonine. Excess sodium cyanoborohydride (0.027 g, 0.430 mmol) was added to a solution of tabersonine (0.024 g, 0.071 mmol) in 2 mL of methanol-water (0.5:1) acidified with acetic acid (pH 4). This mixture was stirred under argon for 16 h at room temperature. Additional sodium cyanoborohydride (0.008 mg, 0.017 mmol) was then added, and the reaction was allowed to stir at room temperature for another 4 h. The mixture was then evaporated to dryness, resuspended in 500 μL of water, and then extracted 4 times with ethyl acetate (500 μL). The organic fractions were combined, evaporated to dryness, and purified by flash silica chromatography (100 % MeCN in quantitative yield to give (2R,3S)-2,3-dihydrotabersonine (0.024 g, 0.071 mmol). 1H-NMR (CD3OD, 400 MHz): δ 0.72 (d, 1H, J = 7.5 Hz), 6.95 (dd, 1H, J = 8.0, 7.5, 1.0 Hz), 6.60 (dd, 1H, J = 7.5, 7.0, 1.0 Hz), 6.51 (1H, J = 8.0 Hz), 5.79–5.75 (m, 1H), 5.37–5.35 (m, 1H), 4.02 (d, 1H, J = 2.5 Hz), 3.77 (2H, q, J = 6.34–6.30 (m, 1H), 3.34 (1H, J = 1.5 Hz), 3.22 (dt, 1H, J = 9.0, 3.0 Hz), 2.62 (1H, J = 15.5 Hz), 2.40–2.34 (m, 1H), 2.30 (s, 3H, CH3) ppm. The solution was treated with 1 M NaOH. Reaction mixtures were then centrifugated at 17,000 x g for 10 min and pelleted precipitated protein. The supernatant was subjected to DNA sequencing.

Alkaloid Extraction. Aseptically grown C. roseus seedlings (~0.1 g) were ground to a fine powder under liquid N2 and extracted in 1 mL of methanol, containing amylaline (10 μM) as an internal standard. The solution was centrifuged, filtered through a 2-μm disposable filter, and then subjected to analysis by LC-MS.

Race to Obtain 3’ Ends of Cr1196 and Cr2251. Total RNA was isolated from elicited seedlings using a Plant RNeasy kit (Qiagen), according to the manufacturer’s protocol. First strand cDNA synthesis was performed using M-MuLV Reverse Transcriptase (New England Biolabs), and cDNA cloning primer (Table S2) according to manufacturer’s instructions. The missing regions were amplified by PCR using primers listed (Table S2). Amplicons were cloned into pGEM-T Easy and subjected to DNA sequencing.

Construction of Expression Plasmids. cDNA was obtained from seedlings as described above, except that an oligo d(T) primer was used in place of the CDNA cloning primer. Open reading frames of candidate cDNAs were amplified by PCR using a sense and antisense primer pairs (Table S2), and High Fidelity Platinum Taq Polymerase (Invitrogen). PCR products were cloned into pGEM-T Easy (Promega). The ORFs were then subcloned into multiple cloning sites of pET28a (EMD) or pRSSET (Invitrogen) using restriction sites engineered into the primers (Table S2).

Production and Purification of Recombinant Protein. Escherichia coli Rosetta 2 (DE3) pLysS cells (EMD) harboring the pET28-Cr2270, pET28-Cr6996, or pRSSET-Cr1196 expression constructs were grown in 1 L of Luria–Bertani medium supplemented with 50 μg mL−1 kanamycin (pET28) or 100 μg mL−1 ampicillin (pRSSET), 34 μg mL−1 chloramphenicol (and 3% (vol/vol) ethyl for Cr1196) at 37 °C with shaking (220 rpm) to an OD600 of 0.6, and then induced with 1 mM IPTG. After 3 h of growth at 37 °C, cells were collected by centrifugation at 5,000 × g, flash-frozen in liquid N2, and stored at −80 °C until further use. Pelleted cells were resuspended in buffer A [100 mM Tris-HCl (pH 7.5), 100 mM KCl, 10% glycerol, 20 mM β-mercaptoethanol] (14) and lysed by sonication. For Cr2270 and Cr1196, Tergitol NP-10 (0.1%, w/v) was added to the supernatant, which was subsequently incubated for 2 h at 4 °C with gentle agitation. Cell debris was removed by centrifugation (10,000 × g) for 30 min at 4 °C. The supernatant was bound to Talo cobalt affinity resin (Clontech), which was then washed twice with 4 mL of buffer A and eluted in a stepwise manner with 750 μL of buffer A containing increasing concentrations (10, 50, 100, and 200 mM) of imidazole. To remove imidazole, these fractions were exchanged into buffer B [100 mM NaPO4 (pH 7.4), 25% glycerol, 20 mM β-mercaptoethanol] using a Zeba desalt spin column (Pierce).

Enzyme Assays. NAMT assays contained 1 μg of purified protein, 0.78–200 μM 2,3-dihydrotabersonine, 1.13–1000 μM S-adenosyl-L-methionine, and 10 μM ajamiline (internal standard), in buffer C (100 mM NaPO4, pH 7.4; 20 mM β-mercaptoethanol, 1 mM EDTA), in a total volume of 50 μL. Assays were incubated at 30 °C for 30 or 60 min and quenched with the addition of 50 μL of 1 N HCl. Aliquots of the reactions were then centrifuged at 17,000 × g for 10 min to pellet precipitated protein. Assays were further diluted 1:150, and 7 μL were analyzed by LC-MS.

Kinetic Analysis. Kinetic assays were performed under optimal conditions (30 °C, pH 7.4). For DHT and SAH kinetins, SAM was fixed at 200 μM, and for SAM kinetins DHT was fixed at 100 μM. Kinetic data were fitted by non-linear regression and analyzed using GraphPad Prism version 4 for Macintosh (GraphPad Software).

Liquid Chromatography-Mass Spectrometry. Ultra performance liquid chromatography analysis was performed using an Acquity Ultra Performance BEH C18 column with a 1.7-μm particle size, 2.1 × 100 mm dimension, and a flow rate of 0.5 mL min−1. The column elution was coupled to MS analysis carried out using a Micromass LCT Premier TOF Mass Spectrometer with an electrospray ionization source. Both modules are from Waters Corporation. The capillary and sample cone voltages were 3,000 V and 30 V, respectively. The desolvation and source temperatures were 300 and 100 °C, respectively. The cone and desolvation gas flow rates were 60 and 400 L h−1. Alkaloid methyltransferase assays were analyzed in E5+ mode using the following gradient: 10:90 to 50:50 acetonitrile-formic acid water over 5 min. Tocopherol methylation was analyzed in E5− mode using the following gradient: 10:90 to 50:50 acetonitrile-formic acid water 5 min. 5-acetoxy-7-deoxyacetanilide was used as an internal standard. Both modules are from Waters Corporation.

Quantitative Real-Time PCR (qPCR). Gene expression analysis was performed as described by Rungpan et al. (29) using the qPCR primers listed in Table S2. In addition to C. roseus rps9, EF1α was used as a second endogenous reference transcript (30). The 2−ΔΔCT method was used for relative gene expression analysis (31).

Phylogenetic Analysis. Amino acid alignments were performed using ClustalX (32). The neighbor-joining phylogeny was generated, and bootstrap analysis was performed with TREECON (33). GenBank accession numbers for the sequences used are provided in Table S3.


Liscombe et al.