A *Papaver somniferum* 10-Gene Cluster for Synthesis of the Anticancer Alkaloid Noscapine

Thilo Winzer et al.

*Science* **336**, 1704 (2012);
DOI: 10.1126/science.1220757

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by clicking here.

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines here.

The following resources related to this article are available online at [www.sciencemag.org](http://www.sciencemag.org) (this information is current as of August 19, 2012):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:
http://www.sciencemag.org/content/336/6089/1704.full.html

Supporting Online Material can be found at:
http://www.sciencemag.org/content/suppl/2012/05/30/science.1220757.DC1.html

A list of selected additional articles on the Science Web sites related to this article can be found at:
http://www.sciencemag.org/content/336/6089/1704.full.html#related

This article cites 45 articles, 18 of which can be accessed free:
http://www.sciencemag.org/content/336/6089/1704.full.html#ref-list-1

This article has been cited by 2 articles hosted by HighWire Press; see:
http://www.sciencemag.org/content/336/6089/1704.full.html#related-urls

This article appears in the following subject collections:
Botany
http://www.sciencemag.org/cgi/collection/botany
**A *Papaver somniferum* 10-Gene Cluster for Synthesis of the Anticancer Alkaloid Noscapine**

Thilo Winzer,1 Valeria Gazda,1 Zhesi He,2 Filip Kaminski,1 Marcelo Kern,1 Tony R. Larson,1 Yi Li,1 Fergus Meade,1 Roxana Teodor,1 Fabián E. Vaistij,1 Carol Walker,2 Tim A. Bowser,2 Ian A. Graham1

Noscapine is an antitumor alkaloid from opium poppy that binds tubulin, arrests metaphase, and induces apoptosis in dividing human cells. Elucidation of the biosynthetic pathway will enable improvement in the commercial production of noscapine and related bioactive molecules. Transcriptomic analysis revealed the exclusive expression of 10 genes encoding five distinct enzyme classes in a high noscapine–producing poppy variety, HN1. Analysis of an F₂ mapping population indicated that these genes are tightly linked in HN1, and bacterial artificial chromosome sequencing confirmed that they exist as a complex gene cluster for plant alkaloids. Virus-induced gene silencing resulted in accumulation of pathway intermediates, allowing gene function to be linked to noscapine synthesis and a novel biosynthetic pathway to be proposed.

Noscapine was first characterized from opium poppy, *Papaver somniferum*, by the distinguished French chemist Pierre Jean Robiquet in 1817 (1) but unlike codeine (which he also discovered) and several other opiates, noscapine is neither painkilling nor addictive.

Noscapine has been used as a human cough suppressant for decades; its effect on the cough reflex and bronchial muscles was reported as early as 1954 (2), and the bioavailability and pharmacokinetics of its orally administered form were established in 1990 (3). The demonstration in 1998 that noscapine acts as a potent antitumor agent that binds to tubulin and affects its polymerization, thereby arresting cell division and inducing apoptosis (4), was followed by confirmation of its antitumor activity in various forms of cancer (5–8). Because of its history of safe use as an antitussive, rapid absorption after oral administration, and apoptosis-inducing effect on a number of cancer cell lines, noscapine may have an advantage over other tubulin-binding anticancer natural products such as the well-established taxanes (9).

Noscapine belongs to the phthalideisoquinoline subclass of the structurally diverse isoquinoline alkaloids, whereas codeine, morphine, thebaine, and oripavine belong to the morphinan subclass (10). Although the biosynthesis of morphinans has been elucidated at the molecular level (11–16), our knowledge of noscapine biosynthesis has not substantially advanced since the 1960s, when isotopic-labeling experiments showed that it is derived from scoulerine (17). Understanding the biochemical genetics that underlie noscapine biosynthesis should enable improved production of this important pharmaceutical and related molecules in both poppy and other systems.

We created metabolite profiles of capsule extracts from three opium poppy varieties developed in Tasmania for alkaloid production, which we designate as high morphine 1 (HM1), high thebaine 1 (HT1), and high noscapine 1 (HN1) on the basis of the most abundant alkaloid in each case (Fig. 1A). Noscapine as well as 53 other

---

**Fig. 1.** Identification of genes exclusively present in the genome of a noscapine-producing poppy variety, HN1. (A) Relative abundance of the major alkaloids extracted from the capsules of three commercial varieties of poppy: HM1, HT1, and HN1. M, morphine; C, codeine; T, thebaine; O, oripavine; N, noscapine; rt, retention time in seconds. (B) EST libraries from stem and capsule were generated by pyrosequencing and unique contiguous sequences assembled as described (18). The first 10 genes are represented only in EST libraries from the HN1 variety; the last five genes are present in EST libraries from all three varieties. All genes are represented at consistently higher levels in stem than in capsule, as shown in color code. PCR on genomic DNA from all three varieties revealed that the 10 HN1-specific genes are absent from the genomes of the HM1 and HT1 varieties, whereas the five other functionally characterized genes are present in all three varieties (fig. S6).
low-abundance compounds, some of which are candidate intermediates in the noscapine biosynthetic pathway, are produced in HN1; none of these compounds appear in either HM1 or HT1 (table S1). We performed Roche 454 pyrosequencing on complementary DNA libraries derived from stem and capsule tissue from all three varieties. Analysis of expressed sequence tag (EST) abundance led to the discovery of a number of previously uncharacterized genes that are expressed in the HN1 variety but are completely absent from the HM1 and HT1 EST libraries (Fig. 1B). We putatively identified the corresponding genes as three O-methyltransferases (PSMT1, PSMT2, PSMT3; fig. S1), four cytochrome P450s (CYP82X1, CYP82X2, CYP82Y1, and CYP719A21; fig. S2), an acetyltansferase (PSAT1; fig. S3), a carboxylesterase (PSCXE1; fig. S4) and a short-chain dehydrogenase/reductase (PSSDR1; fig. S5). In contrast, a number of other functionally characterized genes associated with benzylisoquinoline alkaloid synthesis—including Berberine Bridge Enzyme (BBE), Tetrahydropseudoberberine cis-N-Methyltransferase (TNMT), Salutaridine Reductase (SalR), Salutaridinol 7-O-Acetyltransferase (SalAT), and Thebaine 6-O-demethylase (T6ODM)—were expressed in all three varieties (Fig. 1B). Polymerase chain reaction (PCR) analysis revealed that the genes exclusively expressed in HN1 are present as expected in the genome of HN1 but are absent from the genomes of HM1 and HT1 (Fig. 1B and fig. S6).

Using HN1 and HM1 as parents, we generated an F2 mapping population of 271 individuals. Genotyping of the field-grown F2 population revealed that the HN1-specific genes are tightly linked and associated with the presence of noscapine, which suggests that they occur as a gene cluster involved in noscapine biosynthesis (Fig. 2). Analysis of noscapine levels in field-grown F2 capsules revealed that individuals containing this putative gene cluster fell into two classes. The first, containing 150 individuals, had relatively low levels of noscapine; the second, containing 63 individuals, exhibited the high-noscapine trait of the parental HN1 variety. The 58 F2 individuals that lacked the putative gene cluster contained undetectable levels of noscapine (Fig. 2B). F3 family analysis confirmed that F2 individuals exhibiting the high-noscapine trait were homozygous for the gene cluster, whereas those exhibiting the

**Fig. 2.** Segregation analysis of noscapine content in an F2 mapping population demonstrates requirement for the noscapine gene cluster. (A) Box plot depiction of noscapine levels as percentage dry weight (DW) in glasshouse-grown parental lines HN1 and HM1 and the F1 generation. (B) The F2 generation segregates into three classes of zero, low, and high noscapine (N); F2 GC- and F2 GC+ indicate the absence and presence, respectively, of the noscapine gene cluster. Numbers in parentheses indicate number of individuals in each class. (C) Total major morphinans (T/O/C/M) of the parental HN1 variety. The 58 F2 individuals containing this putative gene cluster fell into two classes. The first, containing 150 individuals, had relatively low levels of noscapine; the second, containing 63 individuals, exhibited the high-noscapine trait of the parental HN1 variety. The 58 F2 individuals that lacked the putative gene cluster contained undetectable levels of noscapine (Fig. 2B). F3 family analysis confirmed that F2 individuals exhibiting the high-noscapine trait were homozygous for the gene cluster, whereas those exhibiting the

**Fig. 3.** The HN1 gene cluster. The structure and position of the 10 HN1-specific genes expressed in stems and capsule tissues are shown above the central black line, which represents 401 kb of genomic sequence. Exons are represented by solid gray boxes and introns by fine black lines. Arrows indicate the 5′ to 3′ orientation of each gene. Additional ORFs depicted below the central black line are as defined by the key. None of these ORFs are represented in the stem and capsule EST libraries. The location and annotation of all ORFs in the 401-kb sequence are detailed in table S5.
Fig. 4. Functional characterization using virus-induced gene silencing of six genes from the HN1 gene cluster. (A to F) Results from both leaf latex (left panels) and capsules (right panels) are consistent with each of these genes encoding enzymes involved in noscapine biosynthesis. All compounds that accumulate, apart from scoulerine, have been putatively identified on the basis of mass spectra (fig. S10). The mass-to-charge (m/z) value (M) followed by retention time in seconds (T) is shown for each compound on the horizontal axis. Metabolites showing a positive change by more than a factor of 2 (VIGS versus controls) and >0.05% total alkaloid profile are shown as percentage total metabolites. A complete list of all metabolites that were significantly changed in the VIGS experiments is shown in table S7. (G) Proposed pathway for noscapine biosynthesis based on VIGS data. Solid arrows depict steps supported by VIGS data; dashed arrows depict additional proposed steps. Italics depict those reactions that await gene assignment. The noscapine structure is numbered according to the IUPAC convention. Labeling of the 3-OH secoberbine intermediates is based on the numbering of the noscapine structure. For the secoberbine intermediates that accumulate in VIGS experiments, R1 = H or OH, R2 = H or OH, and R3 = CH2OH, CHO, or COOH (fig. S10). The proposed pathway assumes R1 = H, R2 = H, and R3 = CHO in the secoberbine intermediates. The crossed-out PSMT2 depicts silencing of this gene product resulting in narcotoline accumulation, as described in the text.
low-noscapine trait were heterozygous (table S2). Noscapine levels in both the glasshouse-grown F1 population and the heterozygous F2 class were much lower than the intermediate levels expected for a semidominant trait, suggesting the involvement of some form of repression. A number of other HN1-specific metabolites were similarly decreased (table S3), suggesting global down-regulation of this branch of alkaloid metabolism when the gene cluster is in the heterozygous state. However, quantitative reverse transcription PCR analysis did not detect any transcriptional repression in the heterozygous state on PSMT1 and PSMT2, which we determine below to encode the first step and an intermediate step, respectively, in the noscapine pathway (fig. S7).

The morphinan branch of alkaloid metabolism, on the other hand, remained largely unaffected in F1 (table S4) and heterozygous F2 material, showing only a decrease in capsules producing high levels of noscapine, presumably due to substrate competition (Fig. 2C). The large step change to high noscapine in the homozygous F2 class suggests that this trait is linked to the gene cluster locus rather than spread quantitatively among other loci.

To further characterize the putative noscapine gene cluster, we prepared a bacterial artificial chromosome (BAC) library from genomic DNA isolated from HN1 and identified six overlapping BACs containing genes from the cluster. Next-generation and Sanger sequencing were used to generate a high-quality assembly of 401 kb, confirming the arrangement of the 10 genes in a cluster spanning 221 kb (Fig. 3 and table S5) (18). The homology and intron-exon structure of the CYP82 and PSMT genes suggest tandem gene duplication after genome reorganization of the progenitor genes. A similar case can be made for PSCEX1 because a second homolog, PSCEX2, is present in the region flanking the gene cluster (Fig. 3 and fig. S4), but this gene is not represented in any of our EST libraries. However, on only one occasion are members of the same gene family adjacent in the cluster, when CYP82X1 and CYP82X2 are inverted with respect to each other. CYP82Y1 is separated from CYP82X2 by a gap of 45 kb containing PSAT1 and PSMT2. PSMT3 and PSMT2 are separated by a gap of 73 kb containing CYP82Y1. Interspersed among the 10 genes are both retrotransposon and DNA transposable element (TE) sequences (Fig. 3 and table S5), which may have some function in gene rearrangement for cluster formation, as is thought to be the case for the thalianol and mameral clusters from Arabidopsis thaliana (19). A search of the PLACE database of plant cis-acting DNA elements (20) revealed a number of short motifs (four or five bases) present in the 1-kb predicted promoter regions upstream of the open reading frames (ORFs) of the 10 genes, among which the WRKY elements are noteworthy (21, 22) (table S6).

To functionally characterize the genes in the HN1 cluster, we performed virus-induced gene silencing (VIGS) on poppy seedlings by established methods (18). VIGS in poppy seedlings persists through to mature plant stages (23), and we therefore routinely assayed both leaf latex and capsule extract (Fig. 4). We managed to silence six of the eight genes we tested by VIGS, as determined by mRNA abundance in infected leaf tissue (fig. S8). Silencing PSMT1, which shows high homology with scoulerine-9-O-methyltransferase from Coptis japonica (24) (fig. S1A), resulted in accumulation of scoulerine in both latex and capsules, as well as low levels of reticuline in latex (Fig. 4A). We expressed the PSMT1 gene product in Saccharomyces cerevisiae and, consistent with the VIGS data, this converted scoulerine to tetrahydrocolumbamine at high efficiency (fig. S9). We therefore conclude that PSMT1 is responsible for the first committed step in the pathway to noscapine synthesis.

The product of PSMT1, tetrahydrocolumbamine, accumulated in latex and capsules that were silenced for CYP719A21, indicating that this gene is responsible for the second step in the pathway (Fig. 4B and fig. S10). CYP719A21 shows high homology to cytochrome P450 oxidases that act as methylendioxy bridge-forming enzymes, and we therefore propose that CYP719A21 encodes a cannabidiyn synthase (25, 26) (fig. S2B). We propose that cannabidiyn is methylated to form N-methylanacrine, which in turn is converted to secoberbine intermediates (Fig. 4G). Consistent with this, cannabidiyn and N-methylanacrine are HN1-specific metabolites (table S1). The product of TNMT has previously been shown to specifically N-methylate protoberberine alkaloids, including cannabidiyn (27). TNMT is present and expressed in HN1, HM1, and HT1 (Fig. 1B) and does not appear to be associated with the HN1 gene cluster (Fig. 3). Three other ORFs with TNMT homology are present in the flanking region of the HN1 gene cluster, but these are not expressed in stems or capsules, further implicating TNMT as having a role in the pathway.

Silencing of a second cytochrome P450 gene, CYP82X2, resulted in accumulation of several secoberbine intermediates, some of which may represent side products to the main synthetic pathway (Fig. 4C). The fragmentation pattern of intermediate 1 (fig. S10) is consistent with the compound being narcotolinol (R1 = OH, R2 = H, and R3 = CH2OH), implying CYP82X2 hydroxylates at the R2 position. The production of these secoberbine intermediates from N-methylanacrine requires breakage of the berberine bridge and ring opening, as depicted in Fig. 4G. Silencing of the carboxylesterase gene PSCEX1 resulted in accumulation of up to 20% total alkaloid content of the acetylated compound papaveroxine (Fig. 4D and fig. S10). Synthesis of papaveroxine from secoberbine intermediates requires hydroxylation and methylation at the position equivalent to the C4´ position of noscapine as well as acetylation of the hydroxyl group at the C3 position (Fig. 4G). The accumulation of papaveroxine in material silenced for PSCEX1 implies that the corresponding enzyme removes an acetyl group from this compound to produce narcotinehemiacetal. That narcotinehemiacetal is an intermediate in the pathway to noscapine synthesis is substantiated by the fact that it accumulates upon silencing of the short-chain dehydrogenase/reductase gene PSSDR1 (Fig. 4E and fig. S10). Conversion of narcotinehemiacetal to noscapine requires dehydrogenation, and we therefore conclude that PSSDR1 is involved in this final synthetic step (Fig. 4G).

The VIGS data for PSCEX1 and PSSDR1 therefore support a biosynthetic route to noscapine that involves O-methylation of secoberbine intermediates at the position equivalent to the C4´ hydroxyl group of noscapine (Fig. 4G). However, silencing PSMT2 did not show any impact on secoberbine intermediates but instead led to accumulation of narcotoline at up to 20% total alkaloids (Fig. 4F). These results suggest that narcotoline is an end product of a desmethyl pathway that accumulates when PSMT2-mediated methylation at the 4´OH group of secoberbine intermediates is compromised. As for the noscapine pathway, the production of narcotoline via a desmethyl pathway is expected to require acetyltransferase, carboxylesterase, and dehydrogenase activities.

Our findings provide evidence for the involvement of 6 of the 10 genes from the HN1 gene cluster in noscapine biosynthesis. The remaining oxidation and acetyltransferase steps in the proposed pathway remain unaccounted for; they could be encoded by the CYP82X1, CYP82Y1, and PSAT1 genes, which remain to be characterized.

This discovery extends the involvement of gene clusters to the alkaloid class of secondary metabolites in higher plants. Noscapine has been reported in a number of Papaver species, and it will be interesting to establish whether it has evolved as a single event prior to speciation of P. somniferum or independently multiple times, as recently reported for the glucosinolate gene cluster in Lotus japonicus (28). As with the other plant gene clusters reported to date (19, 28–31), donor sequences could be recruited from genes encoding related plant enzymes in a process involving gene duplication and neo-functionalization. The arrangement of the HN1 cluster suggests that genome reorganization is an ongoing process, occurring in some cases before duplication, as evidenced by the small gene families (PSCEX, PSCY82, and PSMT), or after duplication, as evidenced by the single-copy genes (PSSDR1, PSAT1, and CYP719A21). The selective advantage to drive cluster evolution in this way could come from co-inheritance of favorable combinations of alleles and coordinate regulation of gene expression at the level of chromatin (32). This work provides a platform for the improved production of noscapine and related bioactive molecules through the molecular breeding of commercial poppy varieties or engineering of new production systems.
Acyl amido synthetases of the GH3 family act as critical prereceptor modulators of plant hormone conjugation of chemically diverse compounds, stability in all organisms. In plants, amino acid molecules by conjugation reactions is a general strategy for the conjugation of amino acids to diverse acyl acid substrates through a two-step mechanism involving adenylation and transference activities (6–8, 14–17) (fig. S1C). GH3 proteins are widespread in plants, and multiple mutant phenotypes connect these enzymes to auxin, auxin, and SA responses in Arabidopsis, rice, and other species (6–8, 14, 15, 18–20). Two of the best-studied Arabidopsis GH3 proteins are AtGH3.11 (JAR1) (14, 20) and AtGH3.12 (PBS3) (18, 19). The Arabidopsis jar1 mutant is defective in AtGH3.11, which prevents formation of JA-Ile and blocks JA-mediated responses (2, 11). Similarly, the Arabidopsis SA-respnsue mutant pbs3 disrupts AtGH3.12, resulting in enhanced susceptibility to Pseudomonas syringae infection due to SA signaling defects (18). Surprisingly, SA is an extremely poor substrate of AtGH3.12.

Structural Basis for Prereceptor Modulation of Plant Hormones by GH3 Proteins

Corey S. Westfall,1* Chloe Zubieta,2* Jonathan Herrmann,1 Ulrike Kapp,2 Max H. Nanno,3,4 Joseph M. Jez1†

Acyl amido synthetases of the GH3 family act as critical prereceptor modulators of plant hormone action; however, the molecular basis for their hormone selectivity is unclear. Here, we report the crystal structures of benzoate-specific Arabidopsis thaliana AtGH3.12/PBS3 and jasmonic acid–specific AtGH3.11/JAR1. These structures, combined with biochemical analysis, define features for the conjugation of amino acids to diverse acyl acid substrates and highlight the importance of conformational changes in the carboxyl-terminal domain for catalysis. We also identify residues forming the acyl acid binding site across the GH3 family and residues critical for amino acid recognition. Our results demonstrate how a highly adaptable three-dimensional scaffold is used for the evolution of promiscuous activity across an enzyme family for modulation of plant signaling molecules.

Plants produce a variety of bioactive signaling compounds, including jasmonates, auxins, and benzoates, in response to intrinsic and environmental cues (1–4). Modification of these molecules and other small signaling molecules by conjugation reactions is a general mechanism for regulation of their activity and stability in all organisms. In plants, amino acid conjugation of chemically diverse compounds, including jasmonic acid (JA), the auxin indole acetic acid (IAA), and salicylic acid (SA), alters the cellular concentrations of their bioactive forms to control plant growth, developmental processes, and defense responses (5–8). For example, formation of amino acid conjugates of JA, IAA, or SA differentially affects the potency of each molecule (fig. S1A). Conjugation of isolectin to JA yields jasmonyl-isolectin (JA-Ile), the active jasmonate hormone that binds to the F-box protein COI1 to trigger JA-mediated degradation of jasmonate ZIM domain proteins and subsequent hormone responses (fig. S1B) (9–11). Formation of IAA-Asp and IAA-Glu targets auxin for degradation and leads to attenuation of auxin signaling. In contrast, conjugation of IAA with either alanine or leucine results in hydrolysable auxin storage forms (3, 12). Similarly, modification of SA can alter its biological activity to modulate defense responses (4, 13). In each system, the GH3 family of acyl acid amido synthetases contributes to the active levels of critical plant hormones necessary for regulating distinct physiological responses.

1Department of Biology, Washington University, St. Louis, MO 63130, USA. 2European Synchrotron Radiation Facility, 38000 Grenoble, France. 3European Molecular Biology Laboratory, Grenoble, France. 4Unit of Virus Host-Cell Interactions, Université Joseph Fourier–European Molecular Biology Laboratory–CNRS, Grenoble, France.

*These authors contributed equally to this work. †To whom correspondence should be addressed. E-mail: jjez@biology2.wustl.edu

See supplementary materials on Science Online.

References and Notes
1. P. J. Robiquet, Annales de Chimie et de Physique 5, 275 (1817).

Fig. 1. Overall structure of AtGH3.12. Ribbon diagram showing the N- and C-terminal domains with α helices and β strands colored gold and blue, respectively. AMP in the active site is shown as a space-filling model. The hinge loop that switches conformation during the reaction sequence is also indicated.