

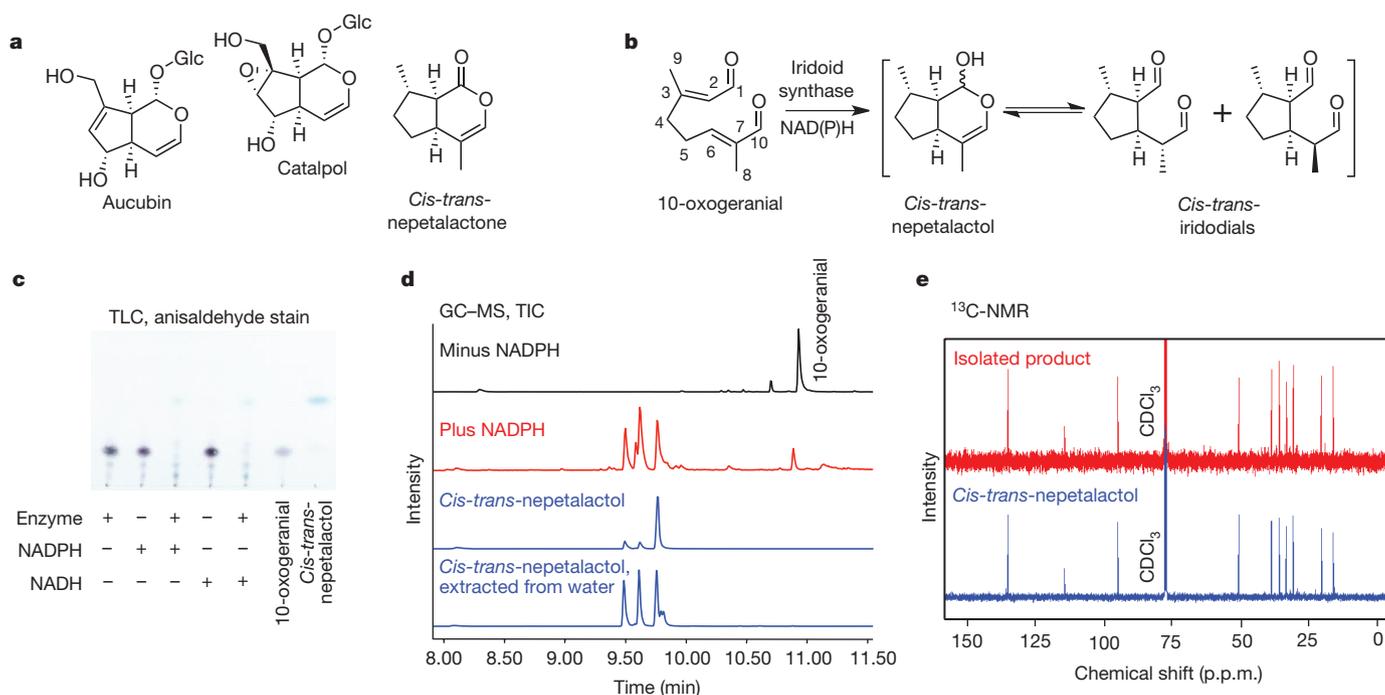
# An alternative route to cyclic terpenes by reductive cyclization in iridoid biosynthesis

Fernando Geu-Flores<sup>1</sup>, Nathaniel H. Sherden<sup>1</sup>, Vincent Courdavault<sup>2</sup>, Vincent Burlat<sup>3,4</sup>, Weslee S. Glenn<sup>1,5</sup>, Cen Wu<sup>6</sup>, Ezekiel Nims<sup>5,†</sup>, Yuehua Cui<sup>6</sup> & Sarah E. O'Connor<sup>1,7</sup>

The iridoids comprise a large family of distinctive bicyclic monoterpenes that possess a wide range of pharmacological activities, including anticancer, anti-inflammatory, antifungal and antibacterial activities<sup>1–4</sup>. Additionally, certain iridoids are used as sex pheromones in agriculturally important species of aphids, a fact that has underpinned innovative and integrated pest management strategies<sup>5</sup>. To harness the biotechnological potential of this natural product class, the enzymes involved in the biosynthetic pathway must be elucidated. Here we report the discovery of iridoid synthase, a plant-derived enzyme that generates the iridoid ring scaffold, as evidenced by biochemical assays, gene silencing, co-expression analysis and localization studies. In contrast to all known monoterpene cyclases, which use geranyl diphosphate as substrate and invoke a cationic intermediate, iridoid synthase uses the linear monoterpene 10-oxogeranial as substrate and probably couples an initial NAD(P)H-dependent reduction step with a subsequent cyclization

step via a Diels–Alder cycloaddition or a Michael addition. Our results illustrate how a short-chain reductase was recruited as cyclase for the production of iridoids in medicinal plants. Furthermore, we highlight the prospects of using unrelated reductases to generate artificial cyclic scaffolds. Beyond the recognition of an alternative biochemical mechanism for the biosynthesis of cyclic terpenes, we anticipate that our work will enable the large-scale heterologous production of iridoids in plants and microorganisms for agricultural<sup>5–8</sup> and pharmaceutical<sup>1–4,9</sup> applications.

The crucial step in the biosynthesis of all cyclic terpenes is the cyclization step that gives rise to the individual ring systems. In all kingdoms of life, this step has been shown to be catalysed by terpene cyclases, which are terpene synthases that use polyprenyl diphosphates such as geranyl diphosphate to produce a cationic species that is subsequently cyclized and rearranged to form one of hundreds of possible ring structures<sup>10,11</sup>. However, the cyclization step that leads to the



**Figure 1 | Identification of the iridoid synthase from *C. roseus*.** **a**, Chemical structure of three plant iridoids. Glc,  $\beta$ -glucopyranosyl. **b**, The iridoid synthase reaction in *C. roseus* (products may possess alternative stereochemistries in other organisms). **c**, Analysis of the reaction by TLC.  $\text{CH}_2\text{Cl}_2$  extracts of different reaction mixtures were run in parallel with authentic samples of the substrate and expected product. **d**, Analysis of the reaction product by GC–MS. Total ion

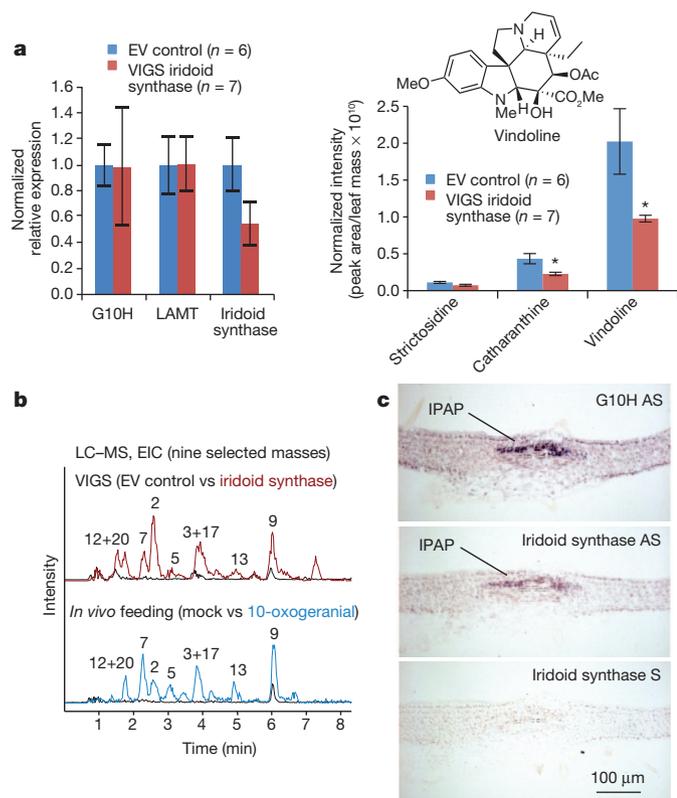
chromatograms (TICs) of  $\text{CH}_2\text{Cl}_2$  extracts of the iridoid synthase reaction ('plus NADPH') and of a negative control ('minus NADPH') are compared to the TICs of freshly synthesized *cis-trans*-nepetalactol or *cis-trans*-nepetalactol that had been dissolved in water and subsequently extracted with  $\text{CH}_2\text{Cl}_2$ . **e**, Comparison of the <sup>13</sup>C-NMR spectra of the purified major product from a milligram-scale enzymatic reaction and the synthetic *cis-trans*-nepetalactol.

<sup>1</sup>Department of Biological Chemistry, John Innes Centre, Norwich NR4 7UH, UK. <sup>2</sup>Université François-Rabelais de Tours, EA2106, Biomolécules et Biotechnologies Végétales, 37200 Tours, France.

<sup>3</sup>Université de Toulouse, UPS, UMR 5546, Laboratoire de Recherche en Sciences Végétales, BP 42617 Auzeville, F-31326 Castanet-Tolosan, France. <sup>4</sup>CNRS, UMR 5546, BP 42617, F-31326 Castanet-Tolosan, France. <sup>5</sup>Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. <sup>6</sup>Department of Statistics and Probability, Michigan State University, East Lansing, Michigan 48824, USA. <sup>7</sup>School of Chemistry, University of East Anglia, Norwich NR4 7TJ, UK. †Present address: Ra Pharmaceuticals, Inc., Cambridge, Massachusetts 02139, USA.

characteristic 5–6 bicyclic scaffold of iridoids (Fig. 1a) is markedly different. Feeding studies have suggested that the direct precursor of all iridoids in plants is not geranyl diphosphate, but the linear monoterpene 10-oxogeranial<sup>12–15</sup>. These studies also suggest that the bicyclic compound nepetalactol is the general precursor of the more derivatized iridoids. Assays using crude plant extracts have indicated that 10-oxogeranial can be converted to nepetalactol in a NADH/NADPH-dependent manner (Fig. 1b)<sup>16</sup>. Nevertheless, the molecular identity of the enzyme catalysing this reaction, which we here call iridoid synthase, has hitherto remained unknown.

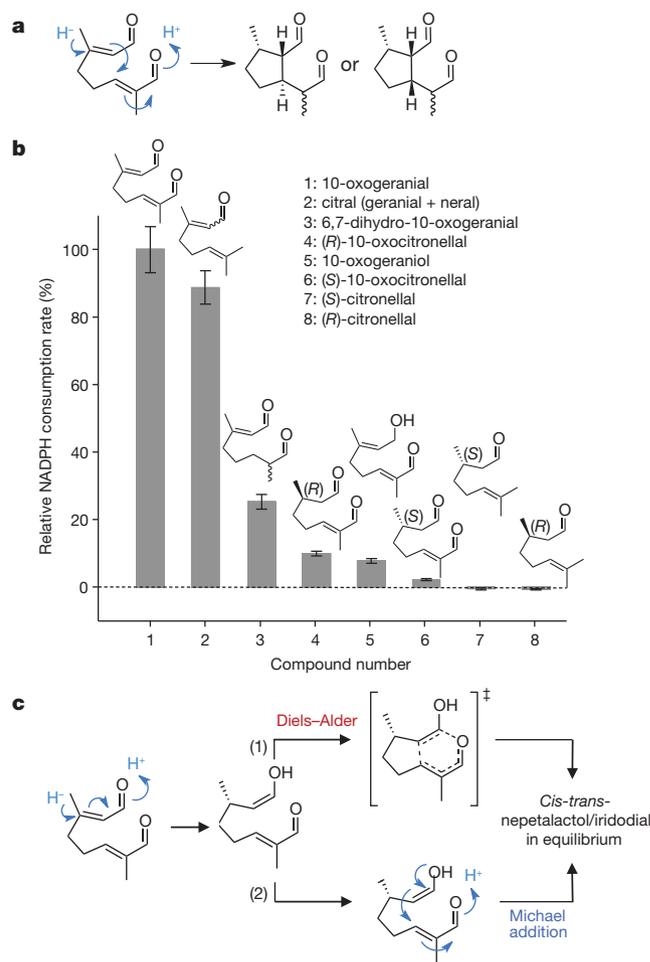
To identify iridoid synthase, we used recently available transcriptomic data from *Catharanthus roseus* (Apocynaceae)<sup>17</sup>, a medicinal plant that produces a variety of iridoid-derived monoterpene indole alkaloids (MIAs), including the anticancer alkaloid vinblastine. Because hundreds of NAD(P)H-dependent enzymes are encoded in the *C. roseus* transcriptome, we first applied a co-regulation criterion to reduce the number of candidates, reasoning that the iridoid synthase would probably have an expression profile similar to geraniol 10-hydroxylase (G10H), the closest characterized enzyme upstream of the cyclization step. In addition to a number of transcripts likely to be involved in the biosynthesis of MIAs, we observed two transcripts coding for NADPH-using enzymes among the 20 best co-regulated transcripts (Supplementary Table 1). The higher-ranking of these two



**Figure 2 | *In vivo* studies of iridoid synthase.** **a**, Transcript downregulation by VIGS. Left, relative transcript quantification by real-time PCR, including data for two known iridoid-related enzymes, G10H and loganic acid methyltransferase (LAMT). Right, quantification of the three major monoterpene indole alkaloids (MIAs) by LC-MS. EV, empty vector control. Asterisks denote statistical significance ( $P < 0.05$  on unpaired Student's *t*-tests). Error bars represent standard errors. **b**, LC-MS analysis of compounds accumulating in both VIGS experiments (upper traces) and 10-oxogeranial feeding experiments (lower traces), including their respective controls (black traces). Traces are extracted ion chromatograms (EICs). Peak numbers refer to Supplementary Table 2. **c**, Cellular localization by RNA *in situ* hybridization on serial longitudinal sections of developing *C. roseus* leaves. Sense ('Iridoid synthase S'), antisense ('Iridoid synthase AS') and G10H probes ('G10H AS') are shown. IPAP, internal phloem-associated parenchyma cells.

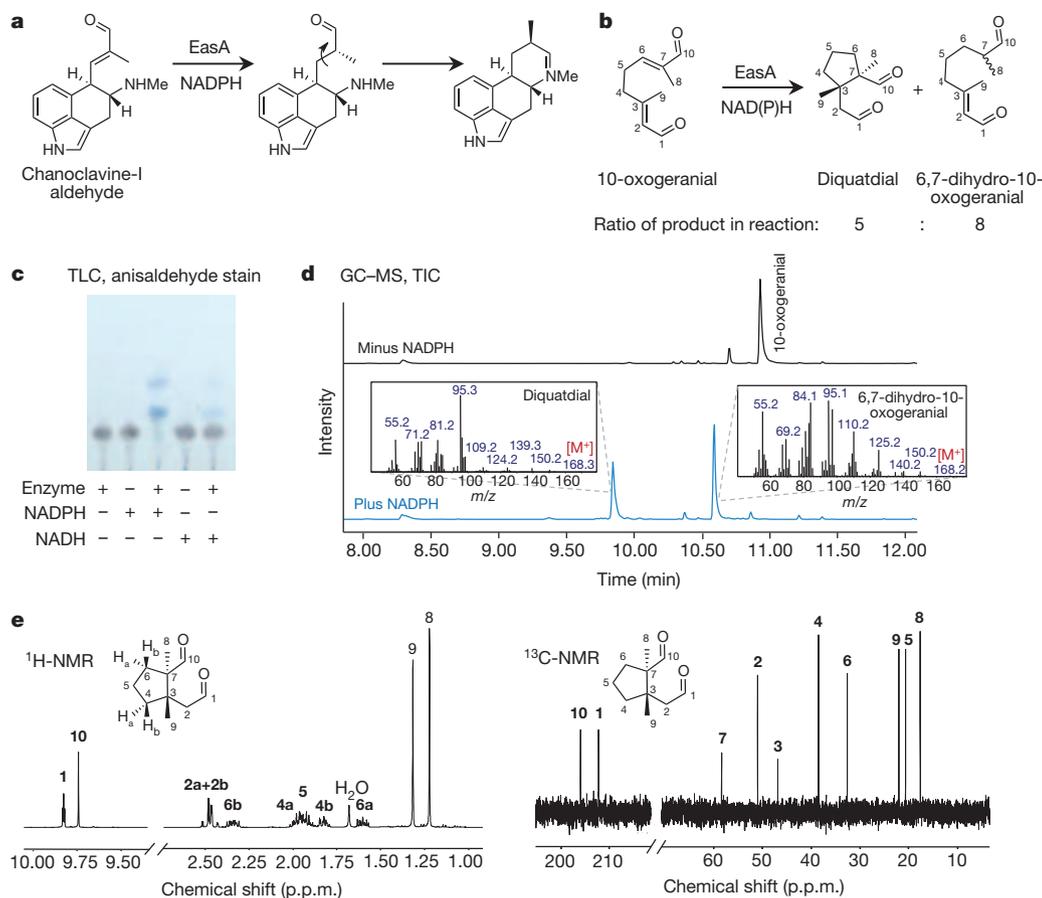
NADPH-using enzymes showed high similarity to progesterone-5 $\beta$ -reductase (P5 $\beta$ R) (67% amino acid identity compared to *Digitalis purpurea* P5 $\beta$ R2) (Supplementary Fig. 1a–c), an enzyme that reduces the C–C double bond of the  $\alpha,\beta$ -unsaturated carbonyl of progesterone in cardenolide biosynthesis<sup>18</sup>. Interestingly, many plant species, including *C. roseus*, possess P5 $\beta$ R homologues but do not produce cardenolides, indicating that numerous genes annotated as P5 $\beta$ R have some other physiological function<sup>19</sup>.

We expressed the iridoid synthase candidate in *Escherichia coli*, and although the heterologous enzyme did not metabolize progesterone, analysis by thin-layer chromatography (TLC) showed that it metabolized 10-oxogeranial in the presence of either NADH or NADPH (Fig. 1c). The identity of the product was confirmed by gas chromatography–mass spectrometry (GC–MS) analysis, which also evidenced the expected equilibrium between nepetalactol and its open dialdehyde forms (iridodials) in water (Fig. 1d). A milligram-scale enzyme reaction allowed the NMR-based assignment of the purified product as *cis-trans*-nepetalactol (Fig. 1e, see nomenclature in ref. 20), which is the expected isomer. The steady-state kinetic constants of the reaction



**Figure 3 | Substrate specificity and mechanistic considerations.**

**a**, Hypothetical fully concerted mechanism, which yields products with different stereochemistry than that of the *cis-trans* product observed experimentally (compare to Fig. 1a). Only one set of possible enantiomers is depicted. **b**, Substrate specificity. Relative activities for the different test substrates are expressed as percentages of the NADPH consumption rate for 10-oxogeranial. Product elucidation by GC–MS is presented in Supplementary Fig. 6 for substrates 2, 3, 4 and 6. Error bars represent standard errors. **c**, Two proposed mechanisms for the reductive cyclization of 10-oxogeranial. After 1,4-addition of NADPH, the enol intermediate can either undergo an inverse electron demand hetero-Diels–Alder (route (1)) or an intramolecular Michael addition (route (2)). The proposed reduction intermediate could also be an enolate.



**Figure 4** | An alternative cyclization catalysed by the fungal reductase EasA to generate the new artificial cyclopentanoid, named ‘diquatdial’ here. **a**, The physiological reaction carried out by EasA in *Aspergillus fumigatus*. **b**, The reaction catalysed by EasA when using 10-oxogeranial as substrate. **c**, Analysis of reaction mixtures by TLC ( $\text{CH}_2\text{Cl}_2$  extracts). **d**, Analysis of the reaction by GC-MS. Total ion chromatograms (TICs) of  $\text{CH}_2\text{Cl}_2$  extracts from the iridoid synthase reaction (‘plus NADPH’) are compared to a negative control (‘minus NADPH’). **e**, Assigned  $^1\text{H}$ - and  $^{13}\text{C}$ -NMR spectra of the purified cyclized product from a milligram-scale enzymatic reaction (absolute stereochemistry still uncharacterized).

were determined through spectrophotometric NADPH consumption assays (for 10-oxogeranial,  $K_M = 4.5 \pm 0.2 \mu\text{M}$ ,  $k_{\text{cat}} = 1.6 \pm 0.1 \text{ s}^{-1}$ ; for NADPH,  $K_M = 4.7 \pm 0.7 \mu\text{M}$ ,  $k_{\text{cat}} = 2.2 \pm 0.2 \text{ s}^{-1}$ , all data mean  $\pm$  s.e.) (Supplementary Fig. 2).

To verify that iridoid synthase catalysed a similar reaction *in vivo*, we used virus-induced gene silencing (VIGS) to downregulate the iridoid synthase in young *C. roseus* leaves (Fig. 2a) and observed a decreased accumulation of the two major downstream MIA metabolites, vindoline and catharanthine (Fig. 2a). This decrease was not accompanied by a build-up of 10-oxogeranial, a reactive dialdehyde, but was accompanied instead by the accumulation of 22 unidentified compounds, as shown by untargeted liquid chromatography–mass spectrometry (LC-MS) analysis (Supplementary Table 2). Feeding millimolar amounts of 10-oxogeranial to young *C. roseus* leaves caused the accumulation of 10 out of the 22 compounds observed in the VIGS experiment, thus linking those 10 compounds to the presence of 10-oxogeranial in leaves (Fig. 2b and Supplementary Table 2).

Further *in vivo* evidence was obtained by studies on cellular and sub-cellular localization. The long MIA pathway in *C. roseus* is known for the disparate cellular localization of its constituent pathway branches<sup>21</sup>. Using RNA *in situ* hybridization in developing *C. roseus* leaves, we localized the iridoid synthase transcripts to internal phloem-associated parenchyma (IPAP) cells, which are the same cells to which G10H localizes<sup>22</sup> (Fig. 2c). Four putative enzymes of the non-mevalonate pathway also localize in these cells<sup>22,23</sup>, suggesting that in *C. roseus*, all of nepetalactol biosynthesis occurs in one cell type. To probe sub-cellular localization, we carried out analyses using fluorescent protein fusions as well as bimolecular fluorescence complementation assays, and these evidenced the formation of dimers or higher-order structures of exclusive cytosolic localization (Supplementary Figs 3 and 4). This is also consistent with the localization of G10H, which is an endoplasmic reticulum-anchored protein with a cytosolic active site<sup>24</sup>.

We next examined the chemistry of this unusual reductive cyclization. A concerted mechanism would invoke hydride donation at carbon 3 of 10-oxogeranial, the concomitant generation of a C–C bond between carbons 2 and 6, and further  $\pi$ -electron translocation towards the terminal oxygen at carbon 10, where a proton could be donated to give iridodial/nepetalactol (Fig. 3a, leftmost structure). This mechanism is unlikely because, for steric reasons, the hydride delivery event and the creation of the new C–C bond must occur from opposite sides of the  $\pi$  nodal plane of carbons 2 and 3, thereby yielding iridodial/nepetalactol with incorrect stereochemistry (Fig. 3a and Supplementary Fig. 5). Instead, we propose a two-step mechanism in which iridoid synthase first catalyses a ‘stand-alone’ 1,4-reduction similar to the one catalysed by P5 $\beta$ R, and then a cyclization step follows.

To test directly the ability of the iridoid synthase to perform stand-alone 1,4-reductions, we performed assays with a suite of 10-oxogeranial-related substrates that included two compounds—citral (mixture of geranial and neral) and 6,7-dihydro-10-oxogeranial—with intact  $\alpha,\beta$ -unsaturated carbonyls at the desired position, but that were unlikely to cyclize upon reduction. Both citral and 6,7-dihydro-10-oxogeranial were converted efficiently to their respective 1,4-reduction products (Fig. 3b and Supplementary Fig. 6), thus supporting the feasibility of a multi-stage mechanism beginning with a 1,4-reduction.

The structures of three minor side-products isolated from the large-scale enzymatic reaction with 10-oxogeranial were also consistent with the ability of iridoid synthase to catalyse a stand-alone reduction reaction. Combined NMR and GC-MS analyses against authentic standards led to identification of these three compounds as (*S*)-10-oxocitronellal, 10-oxogeraniol and 10-oxolinalool (Supplementary Fig. 7). The (*S*)-10-oxocitronellal side-product probably resulted from a reaction similar to the reduction of progesterone by P5 $\beta$ R, a 1,4-addition mechanism<sup>25</sup> (Supplementary Fig. 8). P5 $\beta$ R itself belongs to a super-family of NAD(P)H-dependent reductases specialized in

1,2-addition reductions<sup>25</sup>, and the 10-oxogeraniol product could have arisen via this mechanism. In turn, 10-oxolinalool may have been produced from 10-oxogeraniol by rearrangement after the 1,2-reduction. It seems that the three non-cyclic side-products are derived from vestigial 'reduction only' activities encompassing both 1,2- and 1,4-addition mechanisms.

Therefore, we propose that iridoid synthase first catalyses a reaction that parallels the 1,4-reduction of progesterone by P5 $\beta$ R (Supplementary Fig. 8). The analogous 1,4-reduction of 10-oxogeraniol would yield an enol species that is poised for an inverse electron demand oxo-Diels–Alder reaction (Fig. 3c, route (1)), a rare occurrence in enzymology<sup>26</sup>. One of the iridoid synthase by-products, (S)-10-oxocitronellal, could be formed when the enol tautomerizes to the aldehyde, preventing a Diels–Alder reaction. The inability of the iridoid synthase to cyclize (S)-10-oxocitronellal (Supplementary Fig. 6) is not at odds with the Diels–Alder proposal, because the cycloaddition would only be possible with the enol form, which is greatly disfavoured under the assay conditions and might only exist in the enzyme's active centre following 1,4-reduction. Alternatively, however, an intramolecular Michael addition of the enol reduction product could also lead to the expected product (Fig. 3c, route (2)). More detailed mechanistic, structural and modelling studies to definitively establish whether the concerted Diels–Alder reaction occurs are ongoing. Whatever the mechanistic details, this reaction is attributable to both the enzyme and the chemical structure of 10-oxogeraniol, which is primed for cyclization. Notably, after incubating 10-oxogeraniol with a fungal reductase<sup>27</sup> unrelated to iridoid synthase, a previously unreported cyclized compound was formed (Fig. 4). This implies that a wealth of structurally diverse products can be obtained by fine-tuning different reductases to act on 10-oxogeraniol.

The identification of the iridoid synthase highlights the power of unbiased bioinformatic strategies to rapidly mine the extraordinary chemistry harboured within medicinal plants. We anticipate that our molecular identification of this key enzyme will underpin ongoing efforts aimed at modifying and reconstituting iridoid and MIA pathways in crop plants or microbes for agricultural and pharmaceutical purposes.

## METHODS SUMMARY

The iridoid synthase was identified from a transcriptome assembly of *C. roseus* complementary DNA through mutual-rank co-expression analysis using G10H as bait. The His-tagged protein was expressed in *E. coli* and purified by affinity chromatography. The substrate, 10-oxogeraniol, was synthesized from geraniol by protective acetylation, oxidation with SeO<sub>2</sub>/tert-butyl hydroperoxide (TBHP), basic hydrolysis of the acetate ester, and Swern oxidation. The predicted product, *cis-trans*-nepetalactol, was synthesized from catnip oil as described<sup>20</sup>. Reaction mixtures were qualitatively analysed by TLC (anisaldehyde stain) and GC–MS. The main reaction product and three minor by-products were purified from a large-scale enzymatic reaction (including an NADPH regeneration system) using flash chromatography and the purified compounds were analysed by NMR. The (S)-10-oxocitronellal by-product was further analysed by chiral GC–MS. Kinetic studies were performed by measuring NADPH consumption spectrophotometrically. To conduct substrate specificity assays, different 10-oxogeraniol-related compounds were synthesized using published techniques based on SeO<sub>2</sub>/TBHP oxidations<sup>22</sup>. For the synthesis of 6,7-dihydro-10-oxogeraniol and 2,3,6,7-tetrahydro-10-oxogeraniol, large-scale enzymatic reactions with the fungal reductase EasA were used (instead of chemical synthesis). Relative activities towards different substrates were measured spectrophotometrically and products were identified by GC–MS analysis. The large-scale reaction producing 6,7-dihydro-10-oxogeraniol from 10-oxogeraniol also produced a new cyclopentanoid (diquatdial), which was purified and analysed by NMR. The iridoid synthase transcript was downregulated in young *C. roseus* leaves using TRV-based VIGS as described<sup>28</sup>. Relative transcript quantification was carried out using real-time PCR, and untargeted metabolite analysis was performed by LC–MS. The results of the VIGS experiments were compared to the results of feeding millimolar amounts of 10-oxogeraniol to detached *C. roseus* leaves. The cellular localization of the iridoid synthase transcript was investigated in *C. roseus* leaves using RNA *in situ* hybridization with digoxigenin-labelled probes. The respective sub-cellular localization was probed in transformed *C. roseus* cells using fluorescent protein fusions.

**Full Methods** and any associated references are available in the online version of the paper.

Received 19 August; accepted 19 October 2012.

Published online 21 November; corrected online 5 December 2012 (see full-text HTML version for details).

- Dinda, B., Chowdhury, D. R. & Mohanta, B. C. Naturally occurring iridoids, secoiridoids and their bioactivity. An updated review, part 3. *Chem. Pharm. Bull. (Tokyo)* **57**, 765–796 (2009).
- Dinda, B., Debnath, S. & Banik, R. Naturally occurring iridoids and secoiridoids. An updated review, part 4. *Chem. Pharm. Bull. (Tokyo)* **59**, 803–833 (2011).
- Dinda, B., Debnath, S. & Harigaya, Y. Naturally occurring secoiridoids and bioactivity of naturally occurring iridoids and secoiridoids. A review, part 2. *Chem. Pharm. Bull. (Tokyo)* **55**, 689–728 (2007).
- Tundis, R., Loizzo, M. R., Menichini, F. & Statti, G. A. Biological and pharmacological activities of iridoids: recent developments. *Mini Rev. Med. Chem.* **8**, 399–420 (2008).
- Dewhurst, S. Y., Pickett, J. A. & Hardie, J. In *Vitamins and Hormones* Vol. 83 (ed. Litwack Gerald) 551–574 (Academic Press, 2010).
- Dobler, S., Petschenka, G. & Pankoke, H. Coping with toxic plant compounds—the insect's perspective on iridoid glycosides and cardenolides. *Phytochemistry* **72**, 1593–1604 (2011).
- McElvain, S. M., Walters, P. M. & Bright, R. D. The constituents of the volatile oil of catnip. II. The neutral components. Nepetalic anhydride. *J. Am. Chem. Soc.* **64**, 1828–1831 (1942).
- Søe, A. R. B., Bartram, S., Gatto, N. & Boland, W. Are iridoids in leaf beetle larvae synthesized *de novo* or derived from plant precursors? A methodological approach. *Isotopes Environ. Health Stud.* **40**, 175–180 (2004).
- van der Heijden, R., Jacobs, D. I., Snoeijer, W., Hallard, D. & Verpoorte, R. The *Catharanthus* alkaloids: pharmacognosy and biotechnology. *Curr. Med. Chem.* **11**, 607–628 (2004).
- Chen, F., Tholl, D., Bohlmann, J. & Pichersky, E. The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *Plant J.* **66**, 212–229 (2011).
- Degenhardt, J., Köllner, T. G. & Gershenzon, J. Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry* **70**, 1621–1637 (2009).
- Inouye, H. Neuere Ergebnisse über die Biosynthese der Glucoside der Iridoidreihe. *Planta Med.* **33**, 193–216 (1978).
- Uesato, S., Matsuda, S., Iida, A., Inouye, H. & Zenk, M. H. Intermediacy of 10-hydroxygeraniol, 10-hydroxynerol and iridodial in the biosynthesis of ajmaline and vomilenine in *Rauwolfia serpentina* suspension cultures. *Chem. Pharm. Bull. (Tokyo)* **32**, 3764–3767 (1984).
- Uesato, S., Matsuda, S. & Inouye, H. Mechanism of iridane skeleton formation from acyclic monoterpenes in the biosynthesis of secologanin and vindoline in *Catharanthus roseus* and *Lonicera morrowii*. *Chem. Pharm. Bull. (Tokyo)* **32**, 1671–1674 (1984).
- Uesato, S., Ueda, S., Kobayashi, K. & Inouye, H. Mechanism of iridane skeleton formation in the biosynthesis of iridoid glucosides in *Gardenia jasminoides* cell cultures. *Chem. Pharm. Bull. (Tokyo)* **31**, 4185–4188 (1983).
- Uesato, S., Ikeda, H., Fujita, T., Inouye, H. & Zenk, M. H. Elucidation of iridodial formation mechanism - partial purification and characterization of the novel monoterpene cyclase from *Rauwolfia serpentina* cell suspension cultures. *Tetrahedr. Lett.* **28**, 4431–4434 (1987).
- Giddings, L. A. et al. A stereoselective hydroxylation step of alkaloid biosynthesis by a unique cytochrome P450 in *Catharanthus roseus*. *J. Biol. Chem.* **286**, 16751–16757 (2011).
- Gavidia, I., Tarrío, R., Rodríguez-Trelles, F., Pérez-Bermudez, P. & Seitz, H. U. Plant progesterone 5 $\beta$ -reductase is not homologous to the animal enzyme. Molecular evolutionary characterization of P5 $\beta$ R from *Digitalis purpurea*. *Phytochemistry* **68**, 853–864 (2007).
- Bauer, P. et al. Highly conserved progesterone 5 $\beta$ -reductase genes (P5 $\beta$ R) from 5 $\beta$ -cardenolide-free and 5 $\beta$ -cardenolide-producing angiosperms. *Phytochemistry* **71**, 1495–1505 (2010).
- Liblikas, I. et al. Simplified isolation procedure and interconversion of the diastereomers of nepetalactone and nepetalactol. *J. Nat. Prod.* **68**, 886–890 (2005).
- Ziegler, J. & Facchini, P. J. Alkaloid biosynthesis: metabolism and trafficking. *Annu. Rev. Plant Biol.* **59**, 735–769 (2008).
- Burlat, V., Oudin, A., Courtois, M., Rideau, M. & St-Pierre, B. Co-expression of three MEP pathway genes and geraniol 10-hydroxylase in internal phloem parenchyma of *Catharanthus roseus* implicates multicellular translocation of intermediates during the biosynthesis of monoterpene indole alkaloids and isoprenoid-derived primary metabolites. *Plant J.* **38**, 131–141 (2004).
- Oudin, A. et al. Spatial distribution and hormonal regulation of gene products from methyl erythritol phosphate and monoterpene-secoiridoid pathways in *Catharanthus roseus*. *Plant Mol. Biol.* **65**, 13–30 (2007).
- Guirmand, G. et al. Optimization of the transient transformation of *Catharanthus roseus* cells by particle bombardment and its application to the subcellular localization of hydroxymethylbutenyl 4-diphosphate synthase and geraniol 10-hydroxylase. *Plant Cell Rep.* **28**, 1215–1234 (2009).
- Thorn, A. et al. The crystal structure of progesterone 5 $\beta$ -reductase from *Digitalis lanata* defines a novel class of short chain dehydrogenases/reductases. *J. Biol. Chem.* **283**, 17260–17269 (2008).

26. Kim, H. J., Rusczycki, M. W. & Liu, H.-W. Current developments and challenges in the search for a naturally selected Diels-Alderase. *Curr. Opin. Chem. Biol.* **16**, 124–131 (2012).
27. Cheng, J. Z., Coyle, C. M., Panaccione, D. M. & O'Connor, S. E. A role for old yellow enzyme in ergot alkaloid biosynthesis. *J. Am. Chem. Soc.* **132**, 1776–1777 (2010).
28. Liscombe, D. K. & O'Connor, S. E. A virus-induced gene silencing approach to understanding alkaloid metabolism in *Catharanthus roseus*. *Phytochemistry* **72**, 1969–1977 (2011).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** The present study was funded by the Danish Council for Independent Research (Natural Sciences) through Postdoctoral Fellowship 10-082858 granted to F.G.-F. Additional funding was received from the NIH (GM074820), the j004561SRC (BB/J004561/1 and BB/J009091/1) and the John Innes Foundation. This work was conducted under Defra Plant Health Licence PHSI 449/6612(07/2011). W.S.G. gratefully acknowledges a National Science Foundation Predoctoral Fellowship. Y.C. would like to thank the NSF for grant DMS-1209112. K. R. Chauhan is acknowledged for providing initial nepetalactol samples. We thank S. Fairhurst, L. Hill and A. Jones for NMR, LC-MS and GC-MS assistance, respectively. J. Ward and N. Hawkins are acknowledged for the GC-MS-based high-resolution mass

spectra (HRMS) of synthesized compounds. R. Buell and E. Góngora-Castillo provided many helpful discussions on the transcriptome data. The purified EasA protein was provided by J. Z. Cheng, to whom we are deeply indebted. S.E.O. receives salary support through a synergy initiative between the John Innes Centre and the University of East Anglia.

**Author Contributions** F.G.-F. and S.E.O. conceived the project; C.W., Y.C. and F.G.-F. constructed the G10H mutual ranking list; E.N. cloned the iridoid synthase from cDNA; F.G.-F. and W.S.G. expressed the recombinant protein; N.H.S. performed all chemical synthesis; F.G.-F. carried out all enzyme assays, enzyme-based synthesis, VIGS experiments, and feeding experiments; V.C. performed the sub-cellular localization studies; V.B. carried out the cellular localization studies; F.G.-F. and S.E.O. drafted the manuscript; and S.E.O. acted as overall study director.

**Author Information** The sequence of iridoid synthase has been deposited in GenBank/EMBL/DDBJ with accession no. JX974564 and the short sequence reads are available in the NCBI SRA database under accession number SRP005953. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Sarah E. O'Connor ([sarah.o'connor@jic.ac.uk](mailto:sarah.o'connor@jic.ac.uk)).

## METHODS

**Assembly of the *C. roseus* transcriptome and co-expression analysis.** A *C. roseus* transcriptome assembly was generated using paired-end sequences (76 base pairs) from a normalized cDNA library constructed by pooling equimolar RNA isolated from mature leaf, immature leaf, flower, upper stem and entire roots of a single *C. roseus* reference plant as described previously<sup>29</sup>. The assembly is composed of 41,875 total transcripts with an N50 size of 982 bp. Expression abundances for representative transcripts were determined as described<sup>29</sup> from single end-reads from RNA-seq libraries constructed from 17 development and elicitor treated tissues. The transcriptome assembly (*cra\_v1\_33kmer\_transcripts.fa.txt*) and expression matrix (*Catharanthus\_matrix\_v1\_33kmer.xls*) are available at [ftp://ftp.plantbiology.msu.edu/pub/data/MPGR/Catharanthus\\_roseus/](ftp://ftp.plantbiology.msu.edu/pub/data/MPGR/Catharanthus_roseus/). The reads are available in the NCBI SRA under accession number SRP005953.

To reduce the transcriptomic data set to a workable size, we filtered out all genes exhibiting low expression in leaves (around 5,000 genes with fragments per kilobase of exon per million fragments mapped (FPKM) values of less than 2), and then filtered out genes that were not induced upon treatment with methyl jasmonate, a known elicitor of monoterpenoid indole alkaloids (around 13,000 genes with FPKM ratios of less than 0.9 when comparing uninduced with induced seedlings, 12 days after induction). All known MIA-related enzyme-coding transcripts remained in the data set after these filtering steps. The transcripts were then ordered by the mutual ranking analysis in terms of the similarity of their expression patterns to that of gene G10H. In step 1, we calculated the Pearson correlation coefficient (PCC) of G10H with respect to each of the rest of the contigs. The contigs were then ranked according to their corresponding PCC in descending order, where the contig with the highest correlation with gene G10H was ranked in the top. This is the rank for contig *i* in the rank list of G10H (denoted as 'Fwd rank' in Supplementary Table 1). In step 2, we picked the top 200 best-correlated contigs from the list, and computed the PCC of contig *i* (*i* = 1, 2, ..., 200) with respect to the rest of the contigs in the top 200 list. Following the same ranking procedure, we obtained a ranking list for each contig *i* (*i* = 1, 2, ..., 200), and recorded the rank of gene G10H in each contig's ranking list (denoted as 'Rev rank' in Supplementary Table 1). In step 3, the mutual rank value of contig *i* (*i* = 1, 2, ..., 200) was calculated by  $\sqrt{\text{Fwd rank} \times \text{Rev rank}}$ . We then ranked the 200 mutual rank values in ascending order to generate the final mutual rank list. The 20 best co-regulated transcripts on the mutual rank list are shown in Supplementary Table 1 of Supplementary Materials.

**Cloning of the iridoid synthase from cDNA.** The full-length sequence of iridoid synthase was assembled *in silico* using contig 729 from the *C. roseus* transcriptome assembly (see above) and EST 8191 from the Plant Genomic Database (<http://www.plantdb.org/>). For the molecular cloning, total RNA was isolated from seedlings using an RNeasy Plant Mini Kit (Qiagen), and cDNA synthesis was performed using SuperScript III Reverse Transcriptase (Invitrogen) and oligo-dT primers. The entire coding sequence was amplified using primers ATGAG TTGGTGGTGAAGAGGT and CTAAGGAATAAACCTATAATCCCTCAT CTTATCAAT, and cloned into the Gateway compatible vector pCR8-GW-TOPO-T/A (Invitrogen) to give pCR8-*ISY*. Subcloning into the expression vector pDEST17 using Gateway cloning yielded the *E. coli* expression construct.

**Heterologous protein expression and purification.** Rosetta 2 cells (Merck Millipore) harbouring the desired plasmid were grown at 37 °C, induced with isopropyl β-D-1-thiogalactopyranoside (1 mM) at a  $D_{600\text{ nm}}$  of 1, and then cultured at 15 °C for 12 h. The cells were lysed by sonication and the soluble portion was equilibrated with Ni-NTA agarose resin (Qiagen). The enzyme was then eluted via a stepwise imidazole gradient (25 mM, 50 mM, 100 mM, 150 mM and 300 mM), where the enzyme of interest eluted in the 100 mM, 150 mM and 300 mM imidazole fractions. These fractions were then pooled, concentrated and buffer-exchanged into a storage buffer of 20 mM MOPS.

**Enzymatic assays.** All enzyme assays were carried out using 20 mM MOPS pH 7.0 as buffer. Monoterpene substrates were kept as 50 mM stocks in tetrahydrofuran (THF) at -20 °C. During the preparation of master mixes, care was taken not to exceed TFH concentrations higher than 0.5% in the presence of enzyme, as concentrations above 1% TFH were found to affect activity adversely. The substrate, 10-oxogeranial, was synthesized from geraniol by protective acetylation, oxidation with SeO<sub>2</sub>/TBHP, basic hydrolysis of the acetate ester and Swern oxidation. The predicted product, *cis-trans*-nepetalactol, was synthesized from catnip oil as described<sup>20</sup>. To conduct substrate specificity assays, different 10-oxogeranial-related compounds were synthesized using published techniques based on SeO<sub>2</sub>/TBHP oxidations<sup>22</sup>. For the synthesis of 6,7-dihydro-10-oxogeranial and 2,3,6,7-tetrahydro-10-oxogeranial, large-scale enzymatic reactions with the fungal reductase EasA were used (instead of chemical synthesis). Detailed information on all synthetic procedures, including spectral characterization of novel chemical entities, can be found under Supplementary Materials. Individual characteristics of the different enzymatic assays performed are presented below.

**TLC- and GC-MS-based assays (initial studies and substrate specificity studies).** Reactions (100 μl) were set up in glass vials using 200 μM monoterpene, 400 μM NADH/NADPH and 2.2 μg of purified protein, and were terminated after 1 h by adding 400 μl CH<sub>2</sub>Cl<sub>2</sub>. The organic phase was used directly for GC-MS analysis. For analysis by TLC, 300 μl of the organic phase was vacuum-concentrated to approximately 10 μl, spotted onto normal-phase TLC plates, run using 3:7 ethyl acetate:hexanes, and visualized with anisaldehyde stain.

**Spectrophotometry-based assays (kinetic and substrate specificity studies).** For kinetic studies, the absorbance at 340 nm of 200-μl assays with 80 ng of purified protein was measured using a 96-well plate reader. See the legend to Supplementary Fig. 2 for further experimental details. For substrate specificity studies, the absorbance at 340 nm of 200-μl assays with 200 μM monoterpene, 200 μM of NADPH and 0.41 μg of purified protein was measured in the same manner. Data were collected for 3–5 min, with individual measurements being taken for a single sample every 9 s when using full plate. The measured absorbances were plotted manually on Excel and the initial delta absorption/delta time values were calculated for the linear part of each reaction, with 12 data points having been the minimum number of data points considered for an individual slope calculation. NADPH consumption rates were calculated from these delta absorption/delta time values considering background NADPH decay and an extinction-coefficient-like value (dependent on assay volume) calculated separately.

**Milligram-scale (large-scale) assays.** Both milligram-scale assays (with the iridoid synthase and with EasA) were carried out using an NADPH-generation/regeneration system consisting of glucose-6-phosphate (G6P), glucose-6-phosphate dehydrogenase (G6PDH) and NADP<sup>+</sup>. The reaction with the iridoid synthase was set up in 447 ml of 50 mM MOPS pH 7.0 using 833 μg of purified enzyme, 15.7 mg of 10-oxogeranial, 85 mg of G6P, 211 U of G6PDH and 17.1 mg of NADP<sup>+</sup>. The reaction was monitored continuously by TLC using a short-wave ultraviolet lamp for detection, and went to completion after 1 h. The products were extracted twice with an equal volume of CH<sub>2</sub>Cl<sub>2</sub>, 1 h later. The 13.2 mg of raw product mix were separated using silica-gel-based chromatography with ether/pentanes 3:7 in a Pasteur pipette. A final fraction was eluted in ethyl acetate, vacuum-concentrated, and used for a second flash chromatography column, this time with ether/pentanes 6:4. All products were characterized using a combination of GC-MS and NMR in comparison to synthesized standards as shown in Supplementary Fig. 7. The (S)-10-oxocitronellal by-product was further analysed by chiral GC-MS. The reaction with EasA was set up in 447 ml of 50 mM MOPS pH 7.0 using 491 μg of purified holoenzyme kindly provided by J. Cheng, 16.3 mg of 10-oxogeranial, 111 mg of G6P, 125 U of G6PDH and 17.1 mg of NADP<sup>+</sup>. The reaction was monitored continuously by TLC-UV, and the 10-oxogeranial substrate was nearly completely consumed after 6.75 h. The products were extracted twice with an equal volume of CH<sub>2</sub>Cl<sub>2</sub>. The 12.9 mg of raw product mix were separated using silica-gel-based chromatography with ether/pentanes 3:7 on a Pasteur pipette. The first eluting product, diquatdial, was assigned by correlation spectroscopy (COSY), nuclear Overhauser effect spectroscopy (NOESY), edited heteronuclear single-quantum correlation spectroscopy (HSQC), and heteronuclear multiple-bond correlation spectroscopy (HMBC) two-dimensional NMR experiments. Given additional reaction time, EasA begins to reduce the 6,7-dihydro-10-oxogeranial to give the di-reduction product, which is difficult to separate from diquatdial by TLC. <sup>1</sup>H-NMR (diquatdial, CDCl<sub>3</sub>): δ 1.06 (s, 3H, H<sub>8</sub>), 1.16 (s, 3H, H<sub>9</sub>), 1.43–1.52 (m, 1H, H<sub>6a</sub>), 1.68–1.77 (m, 1H, H<sub>4b</sub>), 1.77–1.95 (m, 3H, H<sub>4a</sub> and H<sub>5</sub>), 2.25–2.35 (m, 1H, H<sub>6b</sub>), 2.42 (dd, <sup>3</sup>J<sub>2a,2b</sub> = 15.3 Hz, <sup>3</sup>J<sub>2a,1</sub> = 2.3 Hz, 1H, H<sub>2a</sub>), 2.47 (dd, <sup>3</sup>J<sub>2b,2a</sub> = 15.3 Hz, <sup>3</sup>J<sub>2b,1</sub> = 2.8 Hz, 1H, H<sub>2b</sub>), 9.71 (s, 1H, H<sub>10</sub>), 9.80 (dd, <sup>3</sup>J<sub>1,2a</sub> ≈ <sup>3</sup>J<sub>1,2b</sub> ≈ 2.6 Hz, 1H, H<sub>1</sub>). <sup>13</sup>C-NMR (CDCl<sub>3</sub>): δ 17.44 (C<sub>8</sub>), 20.49 (C<sub>5</sub>), 21.86 (C<sub>5</sub>), 32.48 (C<sub>6</sub>), 38.38 (C<sub>4</sub>), 46.76 (C<sub>3</sub>), 50.89 (C<sub>2</sub>), 58.31 (C<sub>7</sub>), 202.20 (C<sub>1</sub>), 205.97 (C<sub>10</sub>). HRMS (EI<sup>+</sup>) *m/z* calculated for C<sub>10</sub>H<sub>16</sub>O<sub>2</sub> [M]<sup>+</sup>: 168.1150, found 168.1150. The assignment of the product eluting second, 6,7-dihydro-10-oxogeranial, is presented under Supplementary Information (Synthesis of Chemicals), section E. Note: diquatdial underwent stepwise air oxidation in CDCl<sub>3</sub> at room temperature after 7 days. The resulting diacid was characterized spectroscopically using COSY and edited HSQC two-dimensional NMR experiments. <sup>1</sup>H-NMR (CDCl<sub>3</sub>): δ 1.11 (s, 3H, H<sub>9</sub>), 1.22 (s, 3H, H<sub>8</sub>), 1.52–1.63 (m, 1H, H<sub>6a/b</sub>), 1.62–1.84 (m, 2H, H<sub>5</sub>), 1.71–1.94 (m, 2H, H<sub>4</sub>), 2.42 (d, <sup>3</sup>J<sub>2a/b,2a/b</sub> = 14.0 Hz, 1H, H<sub>2a/b</sub>), 2.39–2.54 (m, 1H, H<sub>6a/b</sub>), 2.76 (d, <sup>3</sup>J<sub>2a/b,2a/b</sub> = 13.9 Hz, 1H, H<sub>2a/b</sub>). <sup>13</sup>C-NMR (CDCl<sub>3</sub>): δ 19.49 (C<sub>5</sub>), 21.24 (C<sub>8</sub>), 22.32 (C<sub>9</sub>), 33.91 (C<sub>6</sub>), 36.96 (C<sub>4</sub>), 41.37 (C<sub>2</sub>), 45.66 (C<sub>3</sub>), 55.62 (C<sub>7</sub>), 178.16 (C<sub>1</sub>), 181.57 (C<sub>10</sub>). HRMS (EI<sup>+</sup>) *m/z* calculated for C<sub>10</sub>H<sub>12</sub>O<sub>2</sub> [M - 2H<sub>2</sub>O]<sup>+</sup>: 164.0837, found 164.0838.

**Plant growth conditions.** Plants were grown in a walk-in growth chamber at 25 °C under 12-h days using the John Innes compost mix No. 2 (peat-based).

**Virus-induced gene silencing (VIGS).** To expedite the assembly of VIGS constructs, we designed and constructed a USER-compatible VIGS vector based on pTRV2. A DNA cassette consisting of oligonucleotides ATTCGCTGAGGCGCA TCGCAACCTCAGCG and AATTCGCTGAGGTTGCGATCGCGCCTCAGCG

was cloned into pTRV2-MCS (The *Arabidopsis* Biological Resource Center, ABRC) using EcoRI. The resulting vector, pTRV2u, was prepared for USER cloning by digesting with AsiSI and Nt.BbvCI. An approximately 500-nt fragment of the *C. roseus* iridoid synthase gene was USER-cloned<sup>30</sup> from pCR8-*ISY* (see earlier section on cloning of the iridoid synthase) into digested pTRV2u using primers GGC GCGAUGGTGGTGAAGAGGTCATTG and GGTTCGCAUCCTCGGTAAAGCGGAATCATGTG. The resulting pTRV2u-*ISY* construct was used to silence the iridoid synthase in *C. roseus* seedlings essentially as described in ref. 28. Briefly, 1-month-old plants were pinched once on their stems with fine forceps that had been dipped in a mixture of *Agrobacterium tumefaciens* strains harbouring the vector of interest and the pTRV1 vector (ABRC). The empty pTRV2u was used as a negative control, and pTRV2-ChlH<sup>28</sup> was used as visual marker for the timing of the silencing response. When pTRV2-ChlH infiltrated plants showed fully bleached leaves 1–3-cm long, a pair of leaves of similar size (equivalent leaves) was harvested from plants infiltrated with the constructs of interest. For each plant, one of these equivalent leaves was destined for LC-MS analysis and the other for quantitative real-time PCR. One out of the eight plants infiltrated with the pTRV2-*ISY* construct presented iridoid synthase expression levels above the empty vector control average and was therefore considered non-silenced and taken out of both final qPCR and LC-MS data pools.

**In vivo feeding experiment of 10-oxogeraniol to *C. roseus*.** Six 2–3-cm long *C. roseus* leaves were fed 100 µl of 5 mM 10-oxogeraniol in 1.5% THF through their cut petioles. Six equivalent leaves were fed 1.5% THF as a mock control. After uptake of the initial 100 µl of liquid, which took between 1 and 6 h, 300 µl of water was added, and the leaves were left overnight under constant light conditions. Twenty-four hours after the start of the experiment, the leaves were collected for analysis by LC-MS.

**Subcellular localization studies.** The subcellular localization of the iridoid synthase was studied by creating fluorescent fusion proteins using the pSCA-cassette YFPi plasmid<sup>24</sup>. The full-length ORF of the iridoid synthase was amplified from pCR8-*ISY* (see section *Cloning of the iridoid synthase from cDNA*) using the specific primers CTGAGATCTAGAATGAGTTGGTGGTGAAGAGGTC and CTGAGATCTAGAAGGAATAAACCTTATCCCTCATCTTATCAATA, which were designed to introduce XbaI and SpeI restriction sites. The PCR product was cloned in-frame either at the 5' end of the yellow fluorescent protein (YFP) coding sequence to generate the iridoid synthase-YFP fusion or at the 3' end to generate the YFP-iridoid synthase fusion. The self-interaction ability of the iridoid synthase was characterized by bimolecular fluorescence complementation (BiFC) assays using the pSCA-SPYNE173 and pSCA-SPYCE(M) plasmids<sup>31</sup>, which allow expressing proteins fused to the amino-terminal extremity of the two split-YFP fragments (YFP<sup>N</sup>, residues 1–173 and YFP<sup>C</sup>, residues 156–239). The previously amplified PCR product was cloned via SpeI in pSCA-SPYNE173 for expression of *ISY*-YFP<sup>N</sup> or in pSCA-SPYCE(M) to produce *ISY*-YFP<sup>C</sup>. Transient transformation of *C. roseus* cells by particle bombardment and fluorescence imaging were performed following the procedures described in refs 24 and 31. Briefly, *C. roseus* plated cells were bombarded with DNA-coated gold particles (1 µm) and 1,100 p.s.i. rupture disc at a stopping-screen-to-target distance of 6 cm, using the Bio-Rad PDS1000/He system. Cells were cultivated for 14 h to 38 h before being collected, treated and observed. The subcellular localization was determined using an Olympus BX-51 epifluorescence microscope equipped with an Olympus DP-71 digital camera and a combination of YFP and cyan fluorescent protein (CFP) filters. The pattern of localization presented in this work is representative of approximately 50 observed cells. The cytosolic localization of the different fusion proteins was confirmed by co-transformation experiments using the cytosolic CFP-β-glucuronidase (GUS) marker<sup>32</sup>. Such plasmid co-transformations were performed using 400 ng of each plasmid or 100 ng for BiFC assays.

**Cellular localization studies.** The PCR product described earlier in section Subcellular localization studies) was cloned into the pGEM-T easy vector (Promega) and the resulting vector was used for the synthesis of sense and anti-sense RNA probes as previously described<sup>33</sup>. For G10H, a previously described plasmid was used for the riboprobe transcription<sup>22</sup>. Paraffin-embedded serial longitudinal sections of young leaves were hybridized with digoxigenin-labelled transcripts and localized with anti-digoxigenin-alkaline phosphatase conjugates according to ref. 33.

**Quantitative real-time PCR.** Leaves collected were frozen in liquid nitrogen, powdered using a pre-chilled mortar and pestle, and subjected to RNA extraction using an RNeasy Plant Mini Kit (Qiagen). RNA (1 µg) was used to synthesize cDNA in 20-µl reactions using the iScript cDNA Synthesis Kit (Bio-Rad). The cDNA served as template for quantitative PCR performed using the CFX96 Real

Time PCR Detection System (Bio-Rad) and the SSO Advanced SYBR Green Supermix (Bio-Rad). As reference genes, we used both the *Elongation Factor 1 (EF1α)* and the *40S Ribosomal Protein S9 (rps9)*, with the respective primer pairs TCAGGAGGCTCTTCTGGTGA/AGCTCCCTGGCAGGGTCAT and TTGAGCCGTATCAGAAATGC/CCCTCATCAAGCAGACCATA. For G10H, LAMT, and the iridoid synthase genes, the primer pairs used were CATTATTAGCGCACCAACC/GAAGTCTTTTCGCCATTGTT, GAGTAATGTATGCAGC CAAG/TTGATTGGATCAAAGATTGG and CCTAGGCTAAATGTCCCAA/GTCTATGGACAGACCATTGTT, respectively. All primer pair efficiencies were between 98% and 105%, and the individual efficiency values were considered in the calculation of normalized relative expression, which was performed using the Gene Study feature of CFX Manager Software. All biological samples were measured in technical duplicates.

**GC-MS.** GC-MS analyses were carried out in an Agilent 6890N GC system coupled to an Agilent 5973 MS detector. All non-chiral separations were performed with a Zebron ZB-5 HT column (30 m × 0.25 mm × 0.10 µm) using helium as carrier gas at a 1 ml min<sup>-1</sup> (linear velocity of 37 cm s<sup>-1</sup>) and with an injector temperature of 220 °C. The program used was the following: 5 min isothermal at 60 °C, 20 °C min<sup>-1</sup> gradient up to 150 °C, 45 °C min<sup>-1</sup> gradient up to 280 °C, 4 min isothermal at 280 °C (run time = 16.39 min). The chiral analysis of 10-oxocitronellal was achieved with a Supelco β-DEX 225 column (30 m × 0.25 mm, 0.25 µm) using hydrogen as carrier gas at a 2.5 ml min<sup>-1</sup> flow (linear velocity of 86 cm s<sup>-1</sup>) and an injector temperature of 220 °C. The program for this chiral separations was the following: 2 min isothermal at 60 °C, 30 °C min<sup>-1</sup> gradient up to 110 °C, 60 min isothermal at 110 °C, 0.5 °C min<sup>-1</sup> gradient up to 111 °C, 30 min isothermal at 111 °C, 0.5 °C min<sup>-1</sup> gradient up to 112 °C, 15 min isothermal at 112, 0.5 °C min<sup>-1</sup> gradient up to 113 °C, 7 min isothermal at 113 °C, 0.5 °C min<sup>-1</sup> gradient up to 114 °C, 3.5 min isothermal at 114 °C, 30 °C min<sup>-1</sup> gradient up to 180 °C, 4 min isothermal at 180 °C (run time = 133.97 min). The long program reflects the on-column instability of the compound above 115 °C and its high on-column retention below 110 °C. Although marginal, the separation of the 10-oxocitronellal isomers was reproducible under these conditions. The analysis presented in Supplementary Fig. 7i was repeated three times with similar results.

High-resolution mass spectra were obtained using a Waters GCT system consisting of an Agilent 6890 fitted with a cool-on-column injector and coupled to a Waters GCT Classic mass spectrometer. The column used was a Phenomenex Zebron HT-5ms (30 m × 0.25 mm, 0.25 µm) with a 5-m guard column, and helium was used as carrier gas at 1 ml min<sup>-1</sup> flow. To avoid thermal degradation of the diacid derived from diquadtal, a program with slower temperature gradient was used: 5 °C min<sup>-1</sup> gradient from 35 to 280 °C with initial and final hold times of 5 min. The mass axis was calibrated from *m/z* 49–285 using 2,4,6-tris(trifluoromethyl)-1,3,5-triazine at before the analysis of each sample. Data were acquired using background reference ions as lock masses—either column bleed ion 207.0329 (C<sub>5</sub>H<sub>15</sub>O<sub>3</sub>Si<sub>3</sub>) or 129.9144 (<sup>35</sup>Cl isotope of trichloroethylene).

**LC-MS.** Leaves were weighed and collected into a fixed volume of methanol (200–600 µl) and incubated at 56 °C for 45 min. After a 30-min centrifugation step at 5,000g, an aliquot of the supernatant (50 µl) was mixed with an equal volume of water and analysed on a Thermo-Finnigan instrument equipped with a Deca XP ion trap detector. The column used was a Phenomenex Luna 3µ C18 (2) (100 × 2.00 mm, 3 µm), and the binary solvent system consisted of acetonitrile (ACN) and 0.1% formic acid in water. The elution program was the following: 1 min isocratic at 15% ACN, 5 min gradient up to 25% ACN, 2 min gradient up to 80%, 3 min isocratic at 80%, 1 min gradient down to 15%, and 4 min isocratic at 15%. Peak areas were calculated using the ICIS algorithm in Finnigan's Xcalibur software and normalized by leaf mass (fresh weight).

29. Góngora-Castillo, E. *et al.* Genomic approaches for interrogating the biochemistry of medicinal plant species. *Methods Enzymol.* **517**, 139–159 (2012).
30. Nour-Eldin, H. H., Geu-Flores, F. & Halkier, B. A. USER cloning and USER fusion: the ideal cloning techniques for small and big laboratories. *Methods Mol. Biol.* **643**, 185–200 (2010).
31. Guirimand, G. *et al.* Strictosidine activation in Apocynaceae: towards a “nuclear time bomb”? *BMC Plant Biol.* **10**, 182 (2010).
32. Guirimand, G. *et al.* Spatial organization of the vindoline biosynthetic pathway in *Catharanthus roseus*. *J. Plant Physiol.* **168**, 549–557 (2011).
33. Mahroug, S., Courdavault, V., Thiersault, M., St-Pierre, B. & Burlat, V. Epidermis is a pivotal site of at least four secondary metabolic pathways in *Catharanthus roseus* aerial organs. *Planta* **223**, 1191–1200 (2006).