

ARTICLE

Received 30 Dec 2015 | Accepted 31 May 2016 | Published 15 Jul 2016

DOI: 10.1038/ncomms12116

OPEN

# Structural investigation of heteroyohimbine alkaloid synthesis reveals active site elements that control stereoselectivity

Anna Stavrinides<sup>1,\*</sup>, Evangelos C. Tatsis<sup>1,\*</sup>, Lorenzo Caputi<sup>1</sup>, Emilien Foureau<sup>2</sup>, Clare E.M. Stevenson<sup>1</sup>, David M. Lawson<sup>1</sup>, Vincent Courdavault<sup>2</sup> & Sarah E. O'Connor<sup>1</sup>

Plants produce an enormous array of biologically active metabolites, often with stereochemical variations on the same molecular scaffold. These changes in stereochemistry dramatically impact biological activity. Notably, the stereoisomers of the heteroyohimbine alkaloids show diverse pharmacological activities. We reported a medium chain dehydrogenase/reductase (MDR) from *Catharanthus roseus* that catalyses formation of a heteroyohimbine isomer. Here we report the discovery of additional heteroyohimbine synthases (HYSs), one of which produces a mixture of diastereomers. The crystal structures for three HYSs have been solved, providing insight into the mechanism of reactivity and stereoselectivity, with mutation of one loop transforming product specificity. Localization and gene silencing experiments provide a basis for understanding the function of these enzymes *in vivo*. This work sets the stage to explore how MDRs evolved to generate structural and biological diversity in specialized plant metabolism and opens the possibility for metabolic engineering of new compounds based on this scaffold.

<sup>1</sup>The John Innes Centre, Department of Biological Chemistry, Norwich NR4 7UH, UK. <sup>2</sup>Université François-Rabelais de Tours, EA2106 'Biomolécules et Biotechnologies Végétales', Tours 37200, France. \* These authors contributed equally to this work. Correspondence and requests for materials should be addressed to V.C. (email: vincent.courdavault@univ-tours.fr) or to S.E.O.C. (email: sarah.oconnor@jic.ac.uk).

**H**eteroyohimbines are a prevalent subclass of the monoterpene indole alkaloids (Corynanthe type skeleton), having been isolated from many plant species, primarily from the Apocynaceae and Rubiaceae families<sup>1</sup>. These alkaloids exhibit numerous biological activities: ajmalicine is an  $\alpha$ 1-adrenergic receptor antagonist<sup>2–5</sup>, and mayumbine (19-epi-ajmalicine) is a ligand for the benzodiazepine receptor (Fig. 1)<sup>6</sup>. Oxidized beta-carboline heteroyohimbines also exhibit potent pharmacological activity: serpentine has shown topoisomerase inhibition activity<sup>7</sup> and alstonine has been shown to interact with 5-HT<sub>2A/C</sub> receptors and shows promise as an anti-psychotic agent<sup>8–13</sup>. In addition, heteroyohimbines are biosynthetic precursors of many oxindole alkaloids, which also display a wide range of biological activities<sup>14</sup>. Although a total of 16 heteroyohimbine stereoisomers are possible, only 8 are reported to be found in nature, at stereocentres C3, C19, C20 (Fig. 1)<sup>14–20</sup>. How and why the stereoselectivity is controlled in the biosynthesis of these alkaloids remains unclear.

The medicinal plant *Catharanthus roseus* produces three of these isomers, ajmalicine (raubasine), tetrahydroalstonine and 19-epi-ajmalicine (mayumbine) (Fig. 1)<sup>21</sup>. These heteroyohimbines, along with the majority of monoterpene indole alkaloids, are derived from deglycosylated strictosidine (strictosidine aglycone)<sup>22</sup>. The removal of a glucose unit from strictosidine by strictosidine glucosidase (SGD) forms a reactive dialdehyde intermediate that can rearrange to form numerous isomers<sup>23</sup>. The stabilization of these isomers by enzyme-catalyzed reduction is hypothesized to be the stepping stone for the extensive chemical diversity observed in the monoterpene indole alkaloids (Fig. 1)<sup>21,22</sup>. We recently reported the first cloning of a biosynthetic gene encoding an enzyme that acts on strictosidine aglycone. This zinc-dependent medium chain dehydrogenase/reductase (MDR), named tetrahydroalstonine synthase (THAS), produces the heteroyohimbine tetrahydroalstonine (Fig. 1)<sup>24</sup>. Although these studies demonstrated that THAS is a key enzyme in heteroyohimbine biosynthesis, the mechanism by which this enzyme controls the stereoselectivity of the reduction remained unknown. Moreover, it is important to note that strictosidine aglycone serves as the precursor for many alkaloid scaffolds, and therefore represents a central branch point in the monoterpene indole alkaloid biosynthetic pathway. Therefore, we set out to identify additional heteroyohimbine synthases (HYSs) with different stereochemical product profiles that would more clearly define how structural diversity, in this case the formation of different stereoisomers, is controlled in this system.

In this study, we assayed 14 MDR homologues identified from the *C. roseus* transcriptome<sup>25,26</sup> that have homology to THAS (Cr\_024553). This screen revealed three additional enzymes with THAS activity (Cr\_010119, Cr\_021691, Cr\_032583a), and, importantly, an enzyme that produced a mixture of heteroyohimbine diastereomers (Cr\_032583b). Crystal structures of THAS (here referred to as THAS1), a second representative THAS (Cr\_021691, THAS2) and the structure of the promiscuous homologue (Cr\_032583b, HYS) were solved and mutants revealed key residues that control the stereochemistry of the product profiles. Notably, analysis of the subcellular localization of some of these HYSs indicates an unusual nuclear localization pattern and an interaction with the previous enzyme, SGD. These discoveries provide insight into the mechanism and evolution of a crucial branch point in a specialized metabolic pathway with both pharmacological and evolutionary importance.

## Results

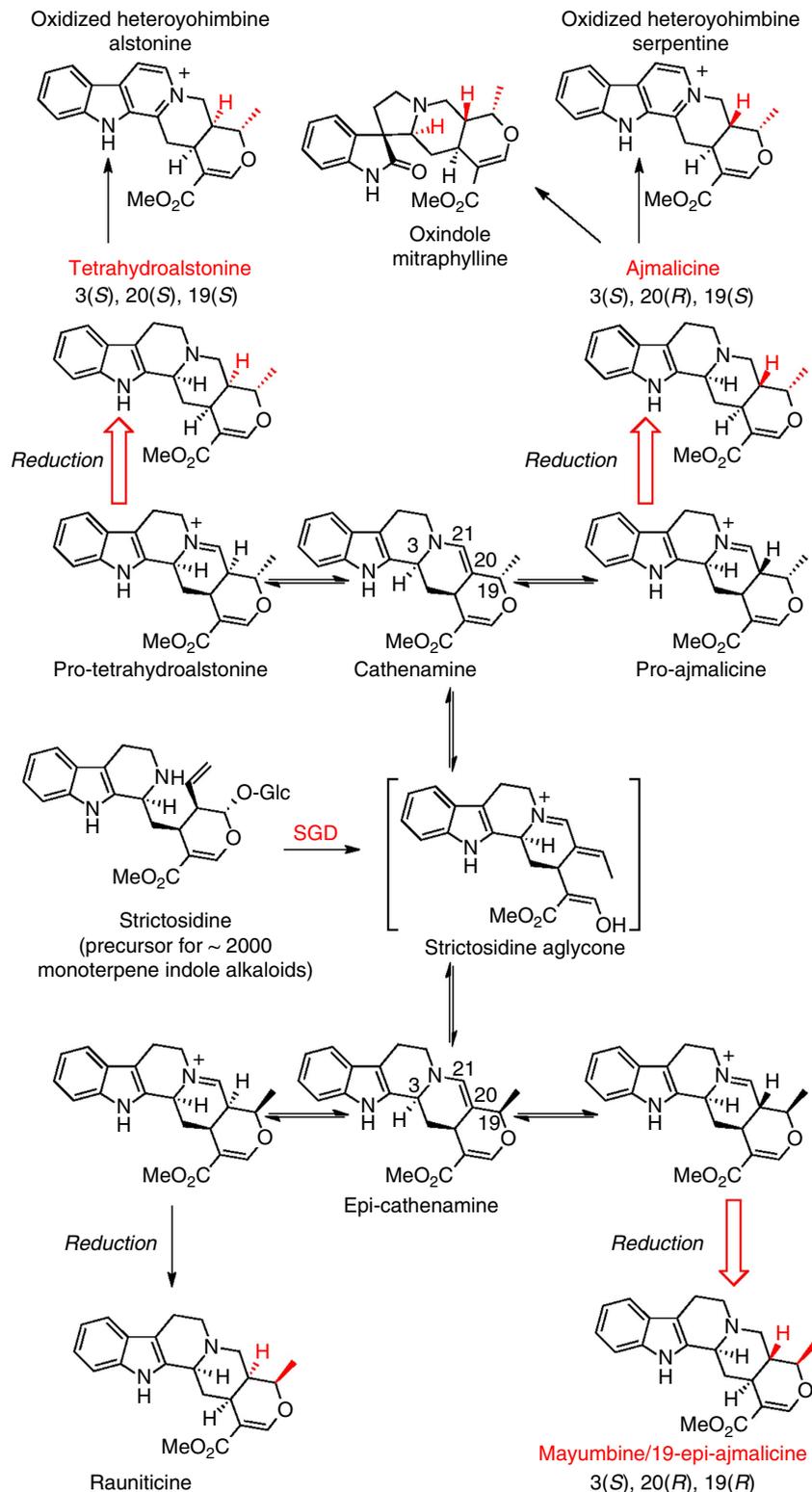
**Discovery of HYSs.** Guided by our initial discovery of THAS1 (ref. 24) we identified candidates from the MDR protein family in

the *C. roseus* transcriptome<sup>25,26</sup> based on amino acid similarity to this enzyme (Supplementary Table 1). Each of these candidates was cloned from *C. roseus* cDNA and expressed in *Escherichia coli*, with the exception of Cr\_017994, which could not be expressed and was not considered further. The remaining candidates were assayed with the substrate strictosidine aglycone, and product formation was monitored by liquid chromatography mass spectrometry (LC-MS). Of these, four (Cr\_010119, Cr\_021691, Cr\_032583a, Cr\_032583b) reduced strictosidine aglycone to a product corresponding to one of the heteroyohimbines (Fig. 2, Supplementary Fig. 1). The products of the enzymatic reactions were identified based on LC-MS data and comparison to authentic standards (Supplementary Fig. 2). Enzymes that failed to produce a heteroyohimbine product were not studied further (Supplementary Fig. 1). Three of the enzymes (Cr\_021691, THAS2; Cr\_010119, THAS3; Cr\_032583a, THAS4) produced tetrahydroalstonine in ~85% yield, with small amounts of 19-epi-ajmalicine (mayumbine) (<15%) also observed in these reactions, similar to the previously reported THAS1. Notably, one enzyme (Cr\_032583b, HYS) produced a dramatically different product profile consisting of a mixture of ajmalicine/tetrahydroalstonine/mayumbine (55:27:15, at pH 6) (Fig. 2). The discovery of this enzyme, HYS, now provides a molecular basis to understand the generation of stereochemical diversity in this alkaloid family.

**Crystallography of three HYSs.** To understand the mechanism of stereochemical control at this crucial biosynthetic branch point, we crystallized three HYSs. THAS1 and THAS2, which produce predominantly tetrahydroalstonine, were both crystallized, since their amino acid sequence identity is relatively low (55%) and the predicted active sites have numerous differences (Fig. 3). HYS, which has a distinctly different product profile, was also crystallized to explore the structural basis behind this distinct stereochemical outcome. Structures (Supplementary Table 3) were obtained for THAS1 and THAS2 with NADP<sup>+</sup> bound (THAS1, 1.05 Å resolution (Fig. 4a,b; Supplementary Figs 3–5); THAS2, 2.10 Å resolution (Fig. 4d, Supplementary Fig. 4)) and in apo form (THAS1, 2.25 Å resolution (Fig. 4c, Supplementary Fig. 5); THAS2, 2.05 Å resolution (Fig. 4c)), while HYS could only be crystallized in the apo form (2.25 Å resolution, Fig. 4c).

**Structural features of HYS active sites.** The five HYS structures described here are similar to sinapyl alcohol dehydrogenase (SAD; PDB accession codes 1YQX and 1YQD) and the SAD homologue cinnamyl alcohol dehydrogenase (CAD; PDB accession codes 2CF5 and 2CF6)<sup>27–29</sup>, which reduce the aldehyde moiety of lignin precursors. Indeed, pairwise superpositions of subunits from these structures gave RMSD values of <2 Å (Supplementary Table 4). The biological unit is an elongated homodimer, with each subunit divided into a substrate and cofactor-binding domain; the latter also being responsible for forming the dimer interface (Supplementary Fig. 3). The overall structure of THAS1, with active site, cofactor and loops highlighted, is shown in Supplementary Fig. 3.

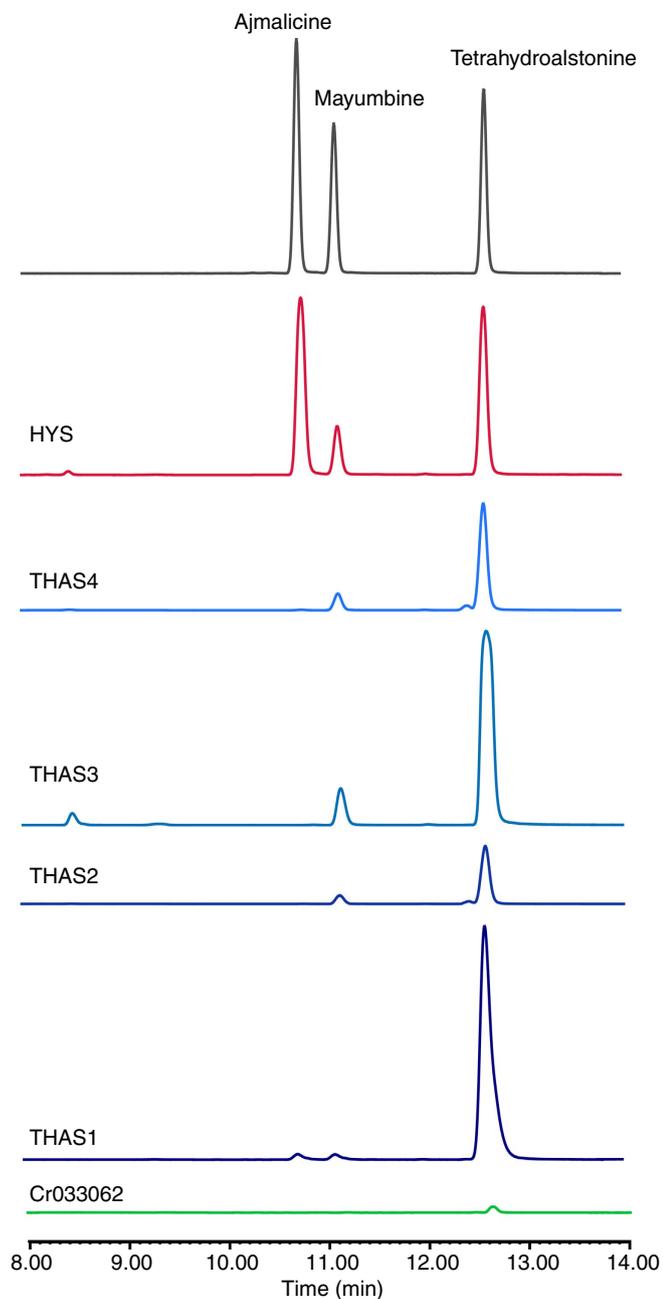
The active site cavities of the HYSs are framed by helix  $\alpha$ 2, the catalytic zinc coordination sphere, and loops 1 and 2, with the NADP(H) co-substrate binding at the base of the active site (Figs 3 and 4a,b,d, , Supplementary Fig. 4). Loop 2, which is positioned above the active site, is highly variable in length and sequence (Figs 3 and 4d). In both THAS1 and THAS2, a network of amino acids holds NADP<sup>+</sup> in place. Most notably, Glu59 of THAS1 anchors NADP(H) through a bidentate interaction with both ribose hydroxyls, with His59 playing a comparable role in SAD, although here the interaction is with the 3' OH only (Fig. 4b).



**Figure 1 | Heteroyohimbine alkaloid biosynthesis.** Heteroyohimbines with 3(S) stereochemistry derive from strictosidine aglycone. The three diastereomers found in *Catharanthus roseus*, are highlighted with red arrows. Alkaloids derived from heteroyohimbines are also shown.

Glu59 is conserved in HYS, but an aspartate residue (THAS3 and THAS2) or a tyrosine residue (THAS4) serves this role in other homologues. MDRs usually contain two zinc ions<sup>30</sup>, a distal ‘structural’ zinc ion, which in this case is coordinated by four cysteine residues, and a proximal ‘catalytic’ zinc ion near the active site, which is coordinated by two cysteines, one histidine and one glutamate residue (Figs 3 and 4b)<sup>27,31</sup>. The proximal zinc of

THAS1 is ~2 Å further away from the cofactor relative to SAD and thus may play no direct role in catalysis (Supplementary Fig. 4). However, it may have a function in maintaining the tertiary structure since three of the liganding residues are in the substrate-binding domain and the fourth is in the cofactor-binding domain. SAD/CAD utilize an active site serine that protonates the alkoxide that results from reduction of the aldehyde substrate<sup>27,29</sup>; this



**Figure 2 | LC-MS analysis of active MDR candidates against strictosidine aglycone.** Cr033062 exhibited only trace activity. See Supplementary Fig. 1 for chromatograms of assays with inactive enzymes and negative controls.

serine is replaced with a tyrosine residue in THAS1 (Tyr56) and HYS (Tyr53) (Fig. 3). In THAS2, this tyrosine on helix  $\alpha 2$  is replaced with a tryptophan residue, but a tyrosine at position 120 points into the active site. Interestingly, a non-proline *cis*-peptide is present in the THAS1-NADP<sup>+</sup> and HYS apo structures (Supplementary Fig. 5). Closer inspection of this region in THAS1 shows that when this bond is in the *trans* conformation, the side chain of Asp340 is projected into the cofactor-binding site such that it would prevent NADPH binding.

**Strictosidine aglycone binding.** Despite extensive efforts, both product and substrate failed to co-crystallize with any of the enzymes. Therefore, molecular docking was used to visualize the

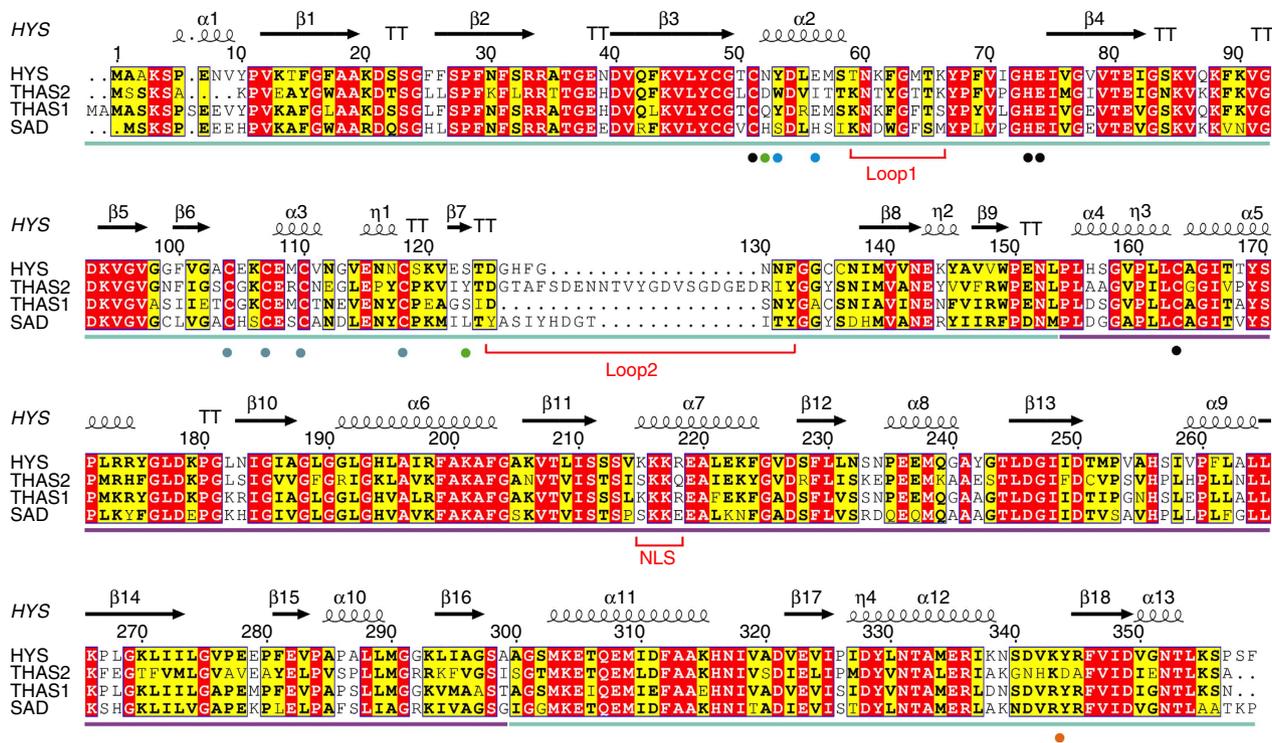
position of strictosidine aglycone in THAS1 (Fig. 4b). To ensure that the correct substrate tautomer was used for docking, we identified the most predominant strictosidine aglycone isomer that forms in solution. Although product precipitation prevented monitoring the SGD reaction *in situ* under aqueous conditions (Methods), <sup>1</sup>H,<sup>15</sup>N-HMBC NMR showed that an enamine species was the predominant product in aqueous methanolic solution (Supplementary Fig. 6). This is consistent with literature reports that cathenamine is the major product of SGD, and is the proposed precursor of ajmalicine and tetrahydroalstonine (Fig. 1)<sup>21,23</sup>. *In silico* docking with THAS1 positions cathenamine between the nicotinamide of the NADP<sup>+</sup> and Tyr56, which is located on helix  $\alpha 2$  (Fig. 4b). THAS1 loop 1 contains Phe65 that also projects into the active site and may interact with the aromatic cathenamine substrate.

**Mechanism of reduction and heteroyohimbine formation.** In tetrahydroalstonine biosynthesis, we hypothesize that cathenamine tautomerizes to the iminium form by protonation at C20, followed by addition of the hydride at C21. Protonation at C20 must occur from the bottom face to yield the *S* stereochemistry observed at this position (Fig. 5a). While there does not appear to be an appropriately positioned active site residue to perform this role, the crystal structures of these enzymes reveal the presence of numerous water molecules in the active site that could potentially protonate this carbon (Fig. 4a). <sup>1</sup>H,<sup>15</sup>N-HMBC measurements of strictosidine aglycone at different pH values show formation of the iminium species in solution when the pH was reduced to ~3.5, indicating that this tautomer can readily form in the presence of an acidic moiety (Supplementary Fig. 6).

To elucidate the stereo and regioselectivity of reduction by NADPH, we isolated tetrahydroalstonine from reactions using THAS1 and pro-*R*-NADPD. Analysis by <sup>1</sup>H-NMR showed that tetrahydroalstonine is labelled with deuterium in the pro-*R* position at C21, consistent with previously reported experiments performed in crude cell extracts (Fig. 6a)<sup>32</sup>. It is possible that THAS1 could reduce the enamine directly, in which case hydride addition would occur at C21, followed by protonation at C20 by a water molecule as described above. The presence of mayumbine/19-*epi*-ajmalicine in some of the enzymatic reactions suggests that small amounts of cathenamine can open and form 19-*epi*-cathenamine, either in solution or in the active site.

In the case of HYS, which produces both ajmalicine (*R* C20) and tetrahydroalstonine (*S* C20), protonation must also occur from the opposite face to yield *R* stereochemistry at C20 (Fig. 5b). Products of HYS generated with pro-*R*-NADPD were also isolated and analysed by <sup>1</sup>H-NMR, and, as for tetrahydroalstonine from THAS1 in each case showed deuterium labelling in the pro-*R* position at C21 (Fig. 6a–c). Therefore, the stereochemical course of hydride addition is not altered in HYS compared with THAS1.

The major difference between HYS and THAS1/THAS2 appears to be the extended loop over the HYS active site (D125-GHFGNN-F132 in HYS and D128-SN-Y131 in THAS1, loop 2 in Fig. 3). The histidine residue in HYS loop 2 (His127) appears to be positioned appropriately to provide an alternative proton source for the opposite ('top') face of the substrate, which could explain the appearance of ajmalicine in the product profile of HYS. Reactions with THAS1 and HYS performed at different pH conditions (5–8) revealed that while changes in pH did not substantially impact the product profile of THAS1, HYS produced increased amounts of ajmalicine relative to tetrahydroalstonine at pH 6 compared with higher pH values (Supplementary Fig. 7). The increased level of ajmalicine in HYS at lower pH values is consistent with the pK<sub>a</sub> value of the



**Figure 3 | Sequence alignment of *Catharanthus roseus* HYS enzymes and *Populus tremuloides* SAD** Numbering corresponds to HYS. Identical and similar amino acids are highlighted in red and yellow, respectively. Secondary structure elements of the HYS apo crystal structure are displayed. THAS1- and HYS-active site amino acids (Y56/53 and E59/56) are indicated by blue dots, and THAS2 active site amino acids (Y120 and D49) are indicated by green dots. Ligands for catalytic and structural zinc ions are highlighted by black and grey dots, respectively. The nuclear localization signal of (THAS1 and HYS) and loops 1 and 2, respectively, are indicated in red. A non-proline *cis*-peptide bond that is observed in THAS1 holo, in one subunit of THAS1 apo (Supplementary Fig. 5), in HYS apo, and not at all in THAS2 is indicated with an orange dot. The substrate-binding domain and the cofactor-binding domain are indicated by blue and purple bars, respectively.

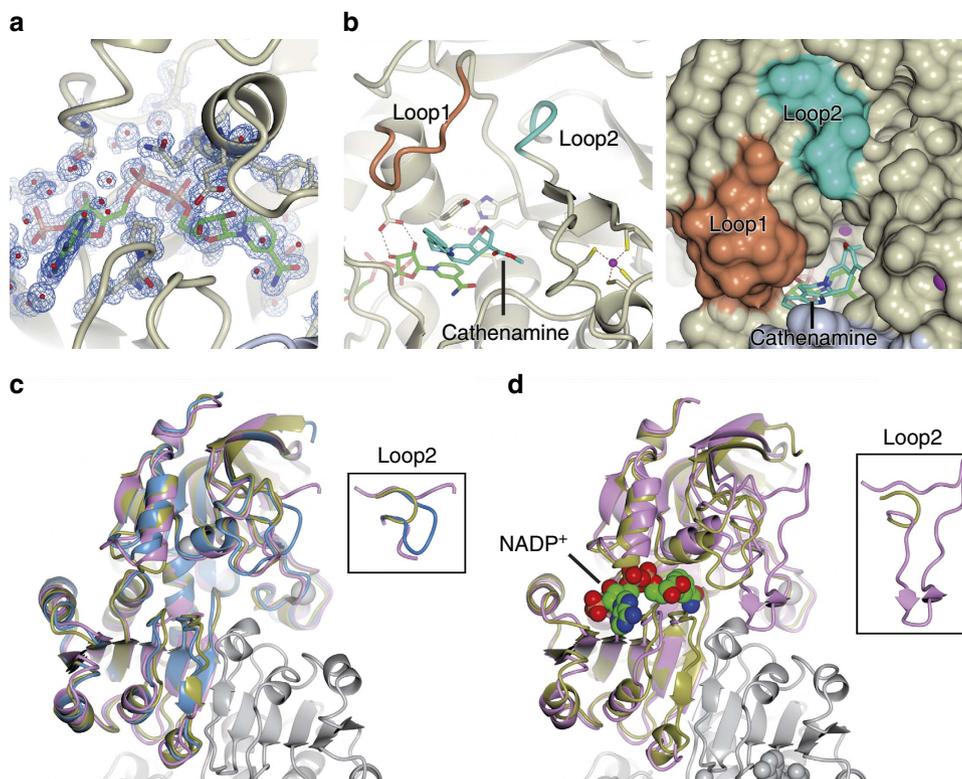
histidine side chain and supports the role of this histidine in ajmalicine biosynthesis, though attributing pH dependence to specific residues must be approached with caution<sup>33</sup>.

**Switching stereoselectivity of HYSs.** Since the major sequence and structural difference between HYS and THAS1 is the extended loop over the HYS active site, loop 2 (Fig. 3), we swapped these loop regions in THAS1 and HYS to determine whether the stereochemical product profile could be switched (Fig. 7a). Loop 1, which is near the active site, was also swapped (Fig. 7a). While the THAS1 mutant containing the swaps displayed reduced activity rather than an altered product profile (Fig. 7c), the HYS mutant containing the shorter THAS1 loop 2 resulted in a product profile similar to that of THAS1 (Fig. 7b, Supplementary Fig. 8). Since His127 is the only ionizable residue in this loop, we hypothesized that this residue protonates C20, as discussed above. Therefore, we mutated this histidine to alanine or asparagine in HYS. Both of these mutants gave the same THAS-like profile, suggesting that His127 is required for producing the ajmalicine (*R* C20) stereochemistry (Supplementary Fig. 9). Mutation of other conserved ionizable residues in the THAS1 active site (Tyr56, Ser102 and Thr166) did not result in substantial changes in the distribution of products (Supplementary Table 6, Supplementary Figs 10 and 11). Mutations to Glu59, which anchors the NADPH cofactor, resulted in a slight increase in product promiscuity (Supplementary Fig. 11), perhaps by causing a shift in the cofactor position. The reactivity of the substrate<sup>34</sup> and precipitation at high concentrations during the assay makes obtaining accurate kinetic constants challenging<sup>24</sup>, so end-point assays were used to assess

stereoselectivity and relative activity of the mutant enzymes (Supplementary Table 6). However, the  $k_{cat}$  (observed) values could be measured for THAS1 ( $1.518 \pm 0.059 \text{ s}^{-1}$ ), THAS2 ( $0.033 \pm 0.001 \text{ s}^{-1}$ ), THAS3 ( $0.102 \pm 0.004 \text{ s}^{-1}$ ), THAS4 ( $0.044 \pm 0.006 \text{ s}^{-1}$ ), HYS ( $0.083 \pm 0.005 \text{ s}^{-1}$ ), HYS\_loop2 swap ( $1.970 \pm 0.153 \text{ s}^{-1}$ ), THAS1 Y56S ( $0.118 \pm 0.005 \text{ s}^{-1}$ ) and THAS1 E59A ( $0.061 \pm 0.005 \text{ s}^{-1}$ ).

**In planta localization of HYSs.** Plants use spatial organization on the organ, tissue and intracellular levels to control product distribution (Supplementary Fig. 12). At the subcellular level, we previously showed that THAS1 has an unusual nuclear localization pattern<sup>24</sup>, which is also where SGD, the enzyme that synthesizes strictosidine aglycone, is localized<sup>34</sup>. Physical interactions using Bimolecular Fluorescence Complementation (BiFC) were observed between these two enzymes<sup>24</sup>. THAS2 and HYS localization were similarly investigated by expressing yellow fluorescent protein (YFP) fusions in *C. roseus* cells. Microscopy of transiently transformed cells revealed that THAS2-YFP (Fig. 8a) was located in both the cytosol and the nucleus while HYS-YFP (Fig. 8e), similar to THAS1, displayed a preferential nuclear localization (Fig. 8 and Supplementary Fig. 13). As reported for THAS1, this localization relies on the presence of a class V nuclear localization sequence in HYS (215-KKKR-218) that is absent from THAS2 (Fig. 3). BiFC assays revealed that both THAS2 and HYS are capable of self-interactions (Supplementary Fig. 14).

BiFC assays were used to determine whether THAS2 and HYS also interact with SGD (Fig. 9). C-terminal split-YFP fragment fusions of both enzymes (THAS2-YFPC and HYS-YFPC) were



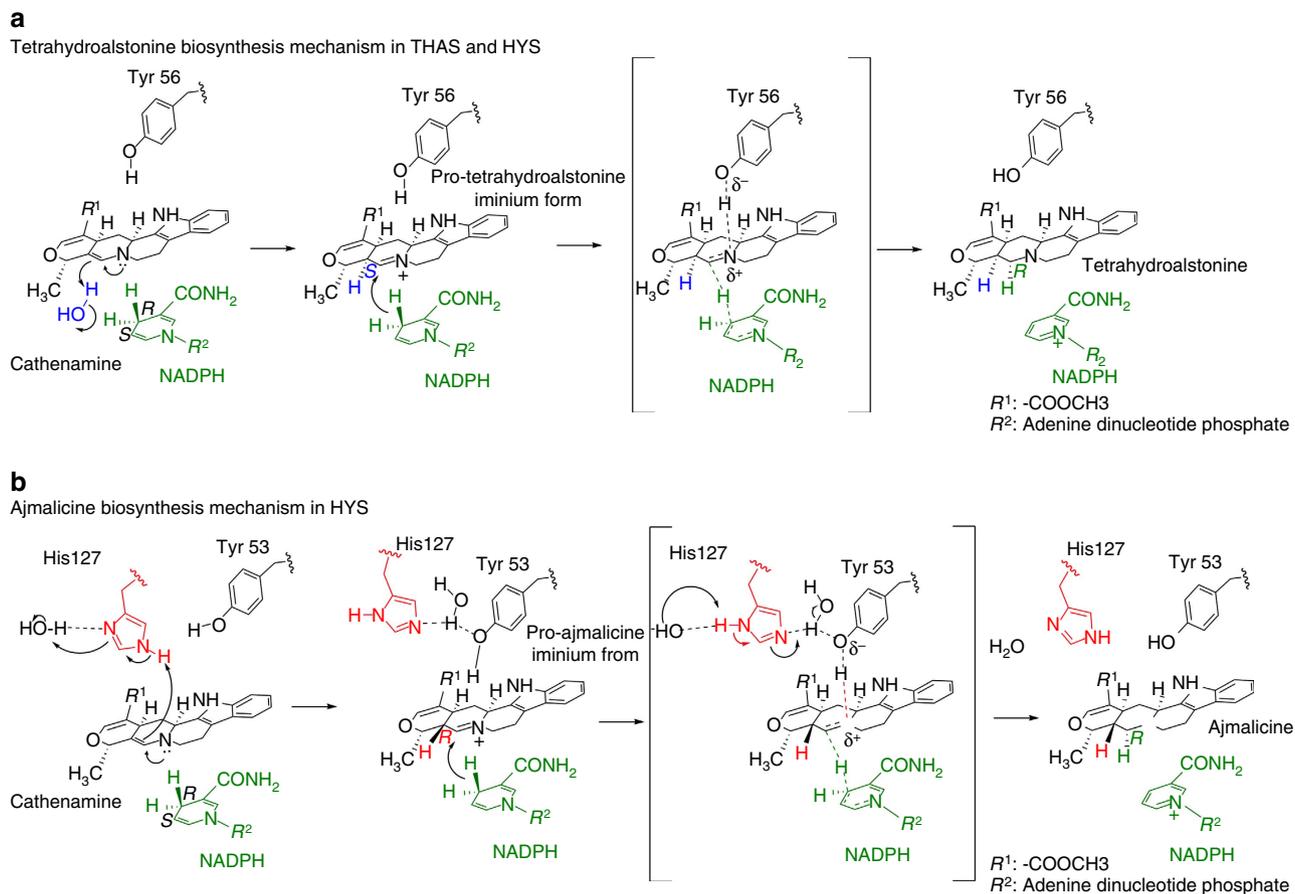
**Figure 4 | Crystal structures of heteroyohimbine synthases THAS1, THAS2 and HYS.** (a) Sample of automatically derived experimentally phased electron density from THAS1 (at 1.12 Å resolution) superimposed on the final model showing the active site region with the NAD<sup>+</sup> cofactor (green carbons) together with neighbouring residues (magnolia carbons) and water molecules (small red spheres). (b) THAS1 docked with cathenamine (pale blue carbons) with the protein shown in both cartoon (left) and space filling (right) modes. The NAD<sup>+</sup> cofactor is shown with green carbons; loop 1 is in orange and loop 2 is in cyan. Zinc ions are displayed as magenta spheres. The active site is largely contained within a single subunit (magnolia surface), although the mouth of the channel leading to the active site is partially bounded by the second subunit of the biological dimer (grey surface). (c) Superposition of the apo structures of THAS1 (gold), THAS2 (pink) and HYS (blue). (d) Superposition of the holo (NAD<sup>+</sup> containing) structures of THAS1 (gold) and THAS2 (pink), with the cofactor of THAS1 shown as van der Waals spheres for emphasis. For (c) and (d), the structures were superposed onto the THAS1 structure, based on the upper subunit alone; only part of the lower subunit of the THAS1 structure is shown in grey for reference (see Supplementary Fig. 3 for images of the full THAS1 dimer). The insets emphasize the differing lengths of loop 2 between the various structures; the central portion of loop 2 in apo THAS2 was disordered.

co-transformed with SGD that was fused to a N-terminal split-YFP fragment (YFPN-SGD). The formation of a nuclear BiFC complex suggests that both of these enzymes interact with SGD in the nucleus (Fig. 9a–d). Interestingly, the emitted fluorescent signal exhibited a punctated, sickle-shaped aspect as previously observed for the THAS1/SGD interaction (Fig. 9e,f) and for SGD localization<sup>34</sup>. In contrast, no interactions were detected when the BiFC assay was conducted with a downstream enzyme from this biosynthetic pathway, 16-hydroxytabersonine 16-O-methyltransferase (16OMT), that is not expected to interact with SGD (Fig. 9g,h).

Double BiFC assays were performed to combine the study of THAS2 and HYS interactions, as well as their interactions with SGD. After transformation into plant cells (16h), we noted the formation of a dual fluorescent signal for THAS2, both in the cytosol and as punctates in the nucleus that may correspond to the superposition of the signal observed for THAS2 self-interactions and THAS2–SGD interaction (Fig. 9i,j) as confirmed by multicolour BiFC (mBiFC) assays (Supplementary Fig. 15). Increased time of expression (36 h) progressively resulted in the disappearance of the cytosolic signal, and it is intriguing to speculate that this implies a recruitment of THAS2 by SGD (Fig. 9q,r). A similar phenomenon was observed for HYS and THAS1 (Fig. 9k–n,s–v– Supplementary Fig. 15). While self-interactions of 16OMT were detected, no nuclear signal was

recovered, confirming the specificity of THAS1–, THAS2–, HYS–SGD interactions (Fig. 9o,p).

**In planta silencing of HYSs.** The expression levels of the HYS transcripts do not suggest whether a specific HYS is the most biologically relevant (Supplementary Fig. 12). To establish whether any of these enzymes synthesize the expected metabolic product *in planta* the genes encoding active HYSs were silenced. For many medicinal plants, including *C. roseus*, virus-induced gene silencing (VIGS) is the only established method to silence genes in the whole plant. In *C. roseus*, the effect of VIGS is temporally and spatially limited to the first two leaves that emerge immediately after infection<sup>35</sup>. Each of the genes encoding a biochemically active enzyme (*THAS1*, *THAS2*, *THAS3*, *THAS4* and *HYS*) was subjected to VIGS in *C. roseus* seedlings and the effect on alkaloid production was monitored by MS. Since *HYS* and *THAS4* were too similar to silence separately, one common gene fragment was used for silencing both genes simultaneously. Successful silencing of the genes was confirmed by quantitative reverse transcription–PCR (qRT–PCR) (Supplementary Fig. 16). Aside from a small degree of cross-silencing between *THAS2* and *THAS3* (12%), all of the target genes were silenced selectively, as measured by qRT–PCR (Supplementary Fig. 16).

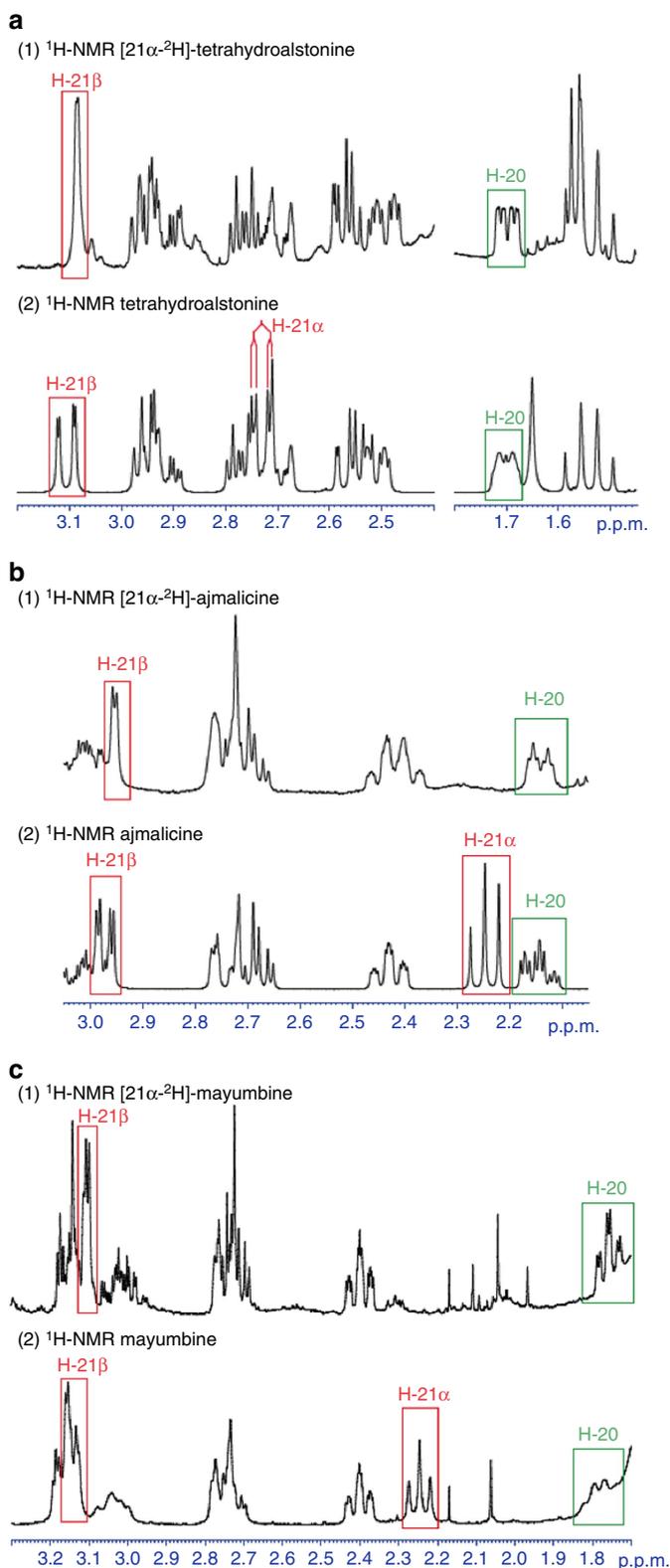


Due to the inherent reactivity of the HYS substrate (strictosidine aglycone), changes in the level of this compound *in planta* are difficult to accurately detect. Instead, the effect of silencing must be established by quantitatively measuring decreases in heteroyohimbine levels. Previously for *THAS1*, we measured the combined peak for heteroyohimbines, since the diastereomers were difficult to resolve on a reverse phase LC column under the reported conditions<sup>24</sup>. However, after substantial optimization (see Methods), an LC-MS method was developed to separate ajmalicine and tetrahydroalstonine in crude leaf extracts (19-epi-ajmalicine/mayumbine is not observed in *C. roseus* leaves, Supplementary Fig. 17). Ajmalicine and particularly tetrahydroalstonine are observed in low levels even in empty vector control samples compared with the major leaf monoterpene indole alkaloids vindoline and catharanthine, and in addition, heteroyohimbine composition varied substantially among individual leaves. Therefore, accurate measurement of decreases in ajmalicine and tetrahydroalstonine levels is challenging. There was no evidence for a decrease of ajmalicine or tetrahydroalstonine when *THAS2* and *THAS3* were silenced. However, for *HYS*, there was a statistically significant decrease (*t*-test 0.0275) in ajmalicine, and no change in tetrahydroalstonine levels. Surprisingly, a statistically significant decrease in ajmalicine, as well as tetrahydroalstonine (*t*-test 0.0277 and 0.0276, respectively) was also noted for *THAS1* (Supplementary Fig. 16). While the results are statistically significant, the leaf-to-leaf variability, the low level of endogenous production, and the catalytic redundancy of these

enzymes make it difficult to draw firm conclusions from these VIGS data. In addition, regulatory factors that impact the ratio of ajmalicine and tetrahydroalstonine, such as transport mechanisms and/or further derivatization to other products, cannot be ruled out. Additional silencing systems, in different tissues, will be required to more firmly establish the physiological function of these enzymes. Nevertheless, we can state that silencing of *HYS* and *THAS1* impacts alkaloid production in *C. roseus* leaves.

## Discussion

Here we report several medium chain dehydrogenases/reductases that produce the heteroyohimbine stereoisomers ajmalicine and/or tetrahydroalstonine, thereby providing a framework to understand the enzymatic control over stereoselectivity in this metabolic pathway. It is notable that we have identified four enzymes that generate tetrahydroalstonine, yet ajmalicine is the more abundant isomer *in planta* (Supplementary Fig. 17). Expression profile data of the genes identified in this study suggest that *HYS*, which produces ajmalicine, is not expressed at higher levels than the other synthases (Supplementary Fig. 12). There may be additional ajmalicine synthases that are not related to the MDR superfamily homologues identified in this study. Alternatively, tetrahydroalstonine could be shuttled into another pathway or degraded, thereby resulting in the observed lower levels that accumulate *in planta*.



**Figure 6 | Deuterium labelling of THAS1 and HYS products using pro-*R*-NADPD.** Comparison of selected regions of  $^1\text{H-NMR}$  spectra of labelled (a) tetrahydroalstonine, (b) ajmalicine, (c) mayumbine. The spectra indicate that C21 is labelled with deuterium in the pro-*R* position.

Importantly, the pharmacological activity of heteroyohimbines is impacted by the stereochemistry. Ajmalicine has recently been used in combination with almitrine in post-stroke treatments,

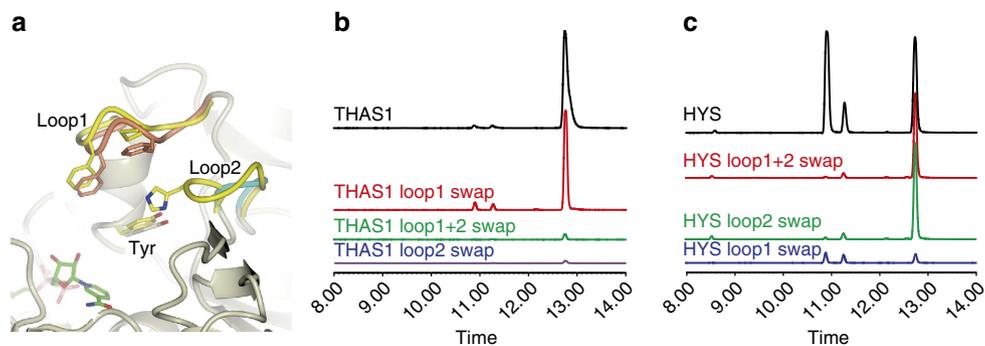
though the side effects caused by almitrine resulted in widespread withdrawal of the drug in 2013 (ref. 4). While tetrahydroalstonine has no reported pharmacological function, its oxidized product, alstonine (Fig. 1), has recently been shown to act by a unique mechanism for modulating dopamine uptake and shows potential as an anti-psychotic drug<sup>13</sup>. The heteroyohimbines have excellent promise as a scaffold for pharmacological activity. The discovery of the HYSs, along with recently developed heterologous production platforms for monoterpene indole alkaloids<sup>36</sup>, now allows the possibility of generating these alkaloids and unnatural derivatives through metabolic engineering/synthetic biology strategies.

The crystal structures of three HYSs reveal the potential of biosynthetic machinery to generate stereochemical variation. Flexible loop regions can be the key to unlocking chemical diversity: as we have demonstrated here, mutating the extended loop over the HYS active site (loop 2 in Fig. 3) impacts stereochemical outcome. Notably, the MDRs that we have identified demonstrate high variability at this region (Fig. 3). Phylogenetic analysis (Supplementary Fig. 18) suggests that these HYSs, which appear to have originated from a common ancestor, may have undergone neo-functionalization through mutation in this loop region. This loop could potentially be harnessed in protein engineering efforts to generate novel catalytic activity.

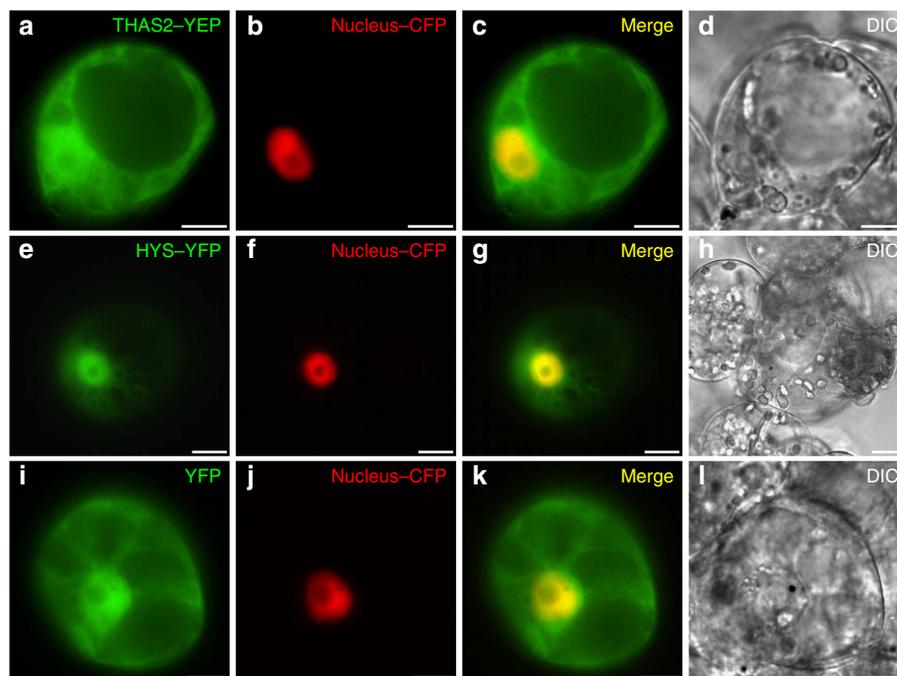
While the *in planta* function of heteroyohimbines is unknown, deglycosylated strictosidine is toxic and may act as a defense compound<sup>34</sup>, similar to the defense roles of the aglycones of the iridoids from which stictosidine is derived<sup>37,38</sup>. SGD is expressed in most tissues (Supplementary Fig. 12), suggesting that the plant must have evolved mechanisms to control the levels of the toxic strictosidine aglycone. In directed overflow metabolism, excess reactive intermediates are converted into non-reactive byproducts<sup>39</sup>. It is intriguing to speculate that monoterpene indole alkaloid biosynthesis may have initially arisen as a mechanism for handling overflow of strictosidine aglycone. The HYSs perform a single, chemically straightforward reduction reaction that immediately neutralizes the reactivity of strictosidine aglycone/cathenamine. The co-localization and interaction of THAS1, THAS2 and HYS with SGD reinforces the hypothesis of an evolutionary mechanism deployed by strictosidine-accumulating plants to manage the reactivity of the strictosidine aglycone. It also raises the question of a possible competition between HYSs for recruitment by SGD when distinct enzymes are co-expressed in the same tissue/cells. Whether the heteroyohimbines serve an active biological function in the plant, or whether they are simply the end product of directed overflow metabolism, or both, remains to be investigated. Regardless, it is clear that MDRs play an important role in the generation of a wide variety of chemical structures. Duplication of the evolutionary dehydrogenase ancestor may have given rise to multiple HYSs, along with MDRs with other biosynthetic activities, such as tabersonine-3-reductase that is involved in the biosynthesis of the anti-cancer alkaloid vinblastine (Supplementary Fig. 18)<sup>40</sup>.

## Methods

**Selection and cloning of candidate MDRs.** The nucleotide and the protein sequences of THAS1 were subjected to a BLAST search against the *C. roseus* Sunstorn Apricot V1.0 Transcript sequences (<http://medicinalplantgenomics.msu.edu>) and the MDR sequences with the highest identity to THAS1 at the active site and which showed non-negligible expression levels in young and mature leaves were selected as candidates for cloning and expression. The protein sequence of SAD was blasted against the same database and MDRs were also selected based on their active site similarity to that of SAD. The genes coding the candidate MDRs were amplified from *C. roseus* leaf cDNA and cloned into the *E. coli* expression vector pOPINF using the In-Fusion cloning kit (Clontech Takara)<sup>41</sup> by using primers designed based on the transcript sequences (Supplementary Table 1).



**Figure 7 | Product profiles of THAS1 and HYS loop swap mutants.** (a) Shown is the apo THAS1 structure (magnolia) with loops 1 and 2 highlighted in orange and cyan, respectively. For clarity, only the corresponding loops of the HYS apo structure are shown in yellow after superposition. Similarly, only the cofactor from the superposed holo THAS1 structure is shown for reference (green carbons). The side chains of important residues are also shown. (b) Representative LC-MS chromatograms of assays with THAS1 mutants in which loop 1, loop 2 or both have been swapped with the corresponding sequences from HYS. (c) Representative LC-MS chromatograms of assays with HYS mutants in which loop 1, loop 2 or both have been swapped with the corresponding sequences from THAS1.

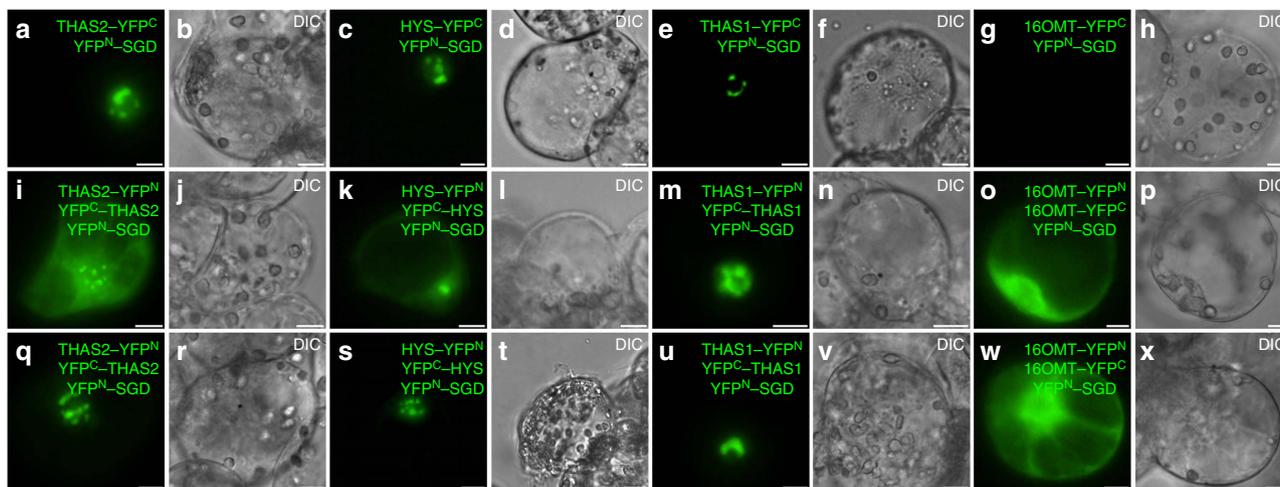


**Figure 8 | THAS2 displays nucleocytoplasmic localization while HYS is preferentially targeted to the nucleus.** *C. roseus* cells were transiently co-transformed with plasmids expressing either THAS2-YFP (a), HYS-YFP (e) or YFP (i) and the plasmid encoding the nuclear-CFP marker (b,f,j). Co-localization of the fluorescence signals appears in yellow when merging the two individual (green/red) false colour images (c,g,k). Cell morphology is observed with differential interference contrast (DIC) (d,h,l). Scale bars, 10  $\mu$ m.

**Site-directed mutagenesis of THAS1 and HYS.** THAS1 mutants were generated by overlap extension PCR. Briefly, the codon to be mutated was selected and two primers, one reverse and one forward (Supplementary Table 4), were designed to overlap and introduce the mutation. A first PCR was carried out using the reverse mutant primer and the 5' forward gene-specific primer (Supplementary Table 1), thus generating the 5' half of the gene carrying the mutation. In parallel, the 3' half of the mutated gene was generated by PCR using the forward mutant primer and the 3' reverse gene-specific primer (Supplementary Table 4). PCR products were gel purified and used for the second PCR overlap reaction for generation of the full-length mutated gene, where the 5' and 3' halves of the mutated genes were mixed in a PCR reaction in equimolar amounts (~100 ng per fragment) and five cycles of PCR were carried out without including primers. After the 5 overlap PCR cycles, the forward and reverse gene-specific primers were added to the mix and further 30 cycles were performed. Full-length PCR products were gel purified, ligated into pOPINF expression vector and transformed into competent *E. coli* Stellar strain cells (Clontech Takara). HYS point mutants were obtained as gene fragments (Integrated DNA Technologies, Belgium) with the H127 or F128 codons mutated

(H127A CAT->GCA; H127N CAT->AAC; F128A TTT->GCT; F128Y TTT->TAC) and the pOPINF overhangs included at the 3' and 5' extremities. The THAS1 and HYS double loop mutants were generated by first making their loop 1 mutant genes and then inserting the second loop 2 swap following the same procedure described above. Mutant constructs were sequenced to verify the mutant gene sequence and correct insertion.

**Enzyme activity assays.** All candidate enzymes and mutants were expressed in SoluBL21 (DE3) *E. coli* cells (Genlantis) grown in 2  $\times$  YT medium. Protein production was induced by addition of 0.2 mM IPTG and the cultures were shaken at 18  $^{\circ}$ C for 16 h. Cells were collected by centrifugation, lysed by sonication in Buffer A (50 mM Tris-HCl pH 8, 50 mM glycine, 500 mM NaCl, 5% v/v glycerol, 20 mM imidazole) supplemented with EDTA-free protease inhibitor (Roche Diagnostics) and 0.2 mg ml<sup>-1</sup> lysozyme. Soluble proteins were purified on Ni-NTA agarose (Qiagen) and eluted with Buffer B (50 mM Tris-HCl pH 8, 50 mM glycine, 500 mM NaCl, 5% v/v glycerol, 500 mM imidazole). Eluates were analysed



**Figure 9 | THAS2 and HYS interact with SGD in the nucleus.** THAS2/SGD (a,i,q) and HYS/SGD (c,k,s) interactions were analysed by BiFC in *C. roseus* cells transiently transformed by distinct combinations of plasmids encoding fusions with the two split YFP fragments, as indicated on each fluorescence picture. THAS1/SGD (e,m,u) and 16OMT/SGD (g,o,w) interactions were studied to evaluate the specificity of THAS2/SGD and HYS/SGD interactions. Single BiFC assays showing interactions with SGD (upper row) and double BiFC assays highlighting both interactions with SGD and THAS2, HYS, THAS1, 16OMT self-interactions were conducted and observed 16 h (middle row) and 36 h (lower row) post-transformation. Cell morphology is observed with differential interference contrast (DIC) (b,d,f,h,j,l,n,p,r,t,v,x). Scale bars, 10  $\mu\text{m}$ .

by SDS-PAGE to verify the purity and the molecular weight of the purified proteins. All proteins were dialysed in Buffer C (50 mM phosphate pH 7.6, 100 mM NaCl) and concentrated. Protein concentration was measured with Bradford reagent (Sigma-Aldrich) according to the manufacturer's instructions. Purified proteins were divided in 20  $\mu\text{l}$  aliquots, fast-frozen in liquid nitrogen and stored at  $-20^{\circ}\text{C}$ .

Candidate MDR enzymes and the selected mutants were screened for activity against deglycosylated strictosidine. The substrate was generated by deglycosylating strictosidine (300  $\mu\text{M}$ ) by the addition of purified SGD in the presence of 50 mM phosphate buffer (pH 6.5) at room temperature for 10 min. The reactions were started by the addition of MDR enzyme (1  $\mu\text{M}$ ) and NADPH (5 mM). Caffeine (50  $\mu\text{M}$ ) was used as internal standard. All reactions were performed in triplicate. Aliquots of the reaction mixtures (10  $\mu\text{l}$ ) were sampled 1 and 30 min after addition of MDR enzyme. The reactions were stopped by the addition of 10  $\mu\text{l}$  of 100% MeOH. Samples were diluted 1:5 in mobile phase ( $\text{H}_2\text{O} + 0.1\%$  formic acid) and centrifuged for 10 min at 4,000g before UPLC-MS injection (1  $\mu\text{l}$ ). The activity of MDR enzymes and mutants was measured by UPLC-MS. Enzymes exhibited a loss of activity after one freeze thaw cycle.

The initial velocity was determined for the wild-type enzymes that displayed HYS activity (THAS1-4 and HYS), as well as for the mutants THAS1 Y56S, E59A and HYS loop 2 swap. Reactions were monitored at 340 nm at room temperature using a (Cary 50 Bio, Varian) spectrophotometer. Strictosidine (50  $\mu\text{M}$ ) was incubated with purified SGD in 50 mM phosphate buffer (pH 7.0) for  $\sim 30$  min at  $30^{\circ}\text{C}$ . NADPH (100  $\mu\text{M}$ ) was added, mixed by pipetting and the absorbance at 340 nm was monitored until stabilized (minimum of 2 min). A predetermined amount of enzyme (10–400 nM) was then added, mixed and the reaction was monitored for a minimum of 5 min at 340 nm. The resulting slope was calculated during the linear reaction range, usually 0–60 s after addition of the enzyme. The reactions were replicated five times and the initial velocity was calculated for each replicate after accounting for the background. The concentration of substrate (50  $\mu\text{M}$  strictosidine) was determined to be saturating for all wild-type enzymes under these conditions.

**Protein crystallization.** Proteins for crystallization were purified from 21 cultures in  $2 \times \text{YT}$  medium. Protein expression was induced by addition of IPTG and the cultures were grown for 16 h at  $18^{\circ}\text{C}$ . Cells were collected by centrifugation and lysed by sonication in 50 ml Buffer A supplemented with EDTA-free protease inhibitor and 10 mg of Lysozyme. Lysates were clarified by centrifugation at 17,000g for 20 min. Two-dimensional automated purification was performed on an AKTApurifier (GE Healthcare). The IMAC step was performed on HisTrap HP 5 ml columns (GE Healthcare) equilibrated with Buffer A. Proteins were step-eluted with Buffer B and directly injected on a gel filtration column equilibrated with Buffer D (20 mM HEPES, 150 mM NaCl, pH 7.5). Fractions were collected and analysed by SDS-PAGE and those containing pure protein were pooled and concentrated in a 10 kDa membrane filter—Millipore filter (Merck Millipore).

Purification of HYS required the addition of 1 mM DTT to all purification buffers and dialysis in Buffer D containing 0.5 mM *tris*(2-carboxyethyl)phosphine (TCEP) before crystallization and storage.

Crystallization screens were conducted by sitting-drop vapour diffusion in MRC2 96-well crystallization plates (Swissci) with a mixture of 0.3  $\mu\text{l}$  well solution from the PEGs (Qiagen), PACT (Qiagen) and JCSG (Molecular Dimensions) suites and 0.3  $\mu\text{l}$  protein solution. Protein concentrations were adjusted to 7–10  $\text{mg ml}^{-1}$  while  $\text{NADP}^+$  (Sigma-Aldrich) was added to a final concentration of 1 mM for co-crystallization studies. Solutions were dispensed either by an OryxNano or an Oryx8 robot (Douglas Instruments).

THAS1 apo crystals were obtained from His<sub>6</sub>-tag cleaved THAS1 (3C protease) in a solution containing 0.1 M MES, pH 6.5, 15% w/v PEG 2000. THAS1-NADP<sup>+</sup> crystals were obtained from a solution containing 0.2 M potassium/sodium tartrate with 20% w/v PEG 3350. THAS2 crystals (with and without NADP<sup>+</sup>) were obtained from a condition containing 0.2 M lithium chloride and 20% w/v PEG 3350. HYS crystals were obtained after removal of the His<sub>6</sub>-tag (using 3C protease) in 0.1 M MMT buffer, pH 5 and 15% w/v PEG 3350. All crystals were cryoprotected by soaking in crystallization solution containing 25% v/v ethylene glycol before flash-cooling in liquid nitrogen.

**Data collection and structure determination.** X-ray data sets were recorded on one of three beamlines at the Diamond Light Source (Oxfordshire, UK) (2F13, I04; 2F15, I03; 5H81 I04-1; 5H82, I04-1; 5H83, I04-1) at wavelengths of 0.9000–0.976  $\text{\AA}$  (2F13, 0.900  $\text{\AA}$ ; 2F15, 0.976  $\text{\AA}$ ; 5H81, 0.920  $\text{\AA}$ ; 5H82, 0.920  $\text{\AA}$ ; 5H83, 0.920  $\text{\AA}$ ) using either a Pilatus 6M or 2M detector (Dectris) with the crystals maintained at 100 K by a Cryojet cryocooler (Oxford Instruments). Diffraction data were integrated using XDS<sup>42</sup> and scaled and merged using AIMLESS<sup>43</sup> via the XIA2 expert system<sup>44</sup>; data collection statistics are summarized in Supplementary Table 3. Initially the THAS1-NADP<sup>+</sup> data set was automatically processed at the beamline by fast\_dp<sup>45</sup> to 1.12  $\text{\AA}$  resolution and a structure solution was automatically obtained by single wavelength anomalous dispersion phasing using the SHELX suite<sup>46</sup> via the fast\_ep pipeline (Winter, manuscript in preparation). Despite being collected at a wavelength somewhat remote from the zinc K X-ray absorption edge (theoretical wavelength 1.284  $\text{\AA}$ ), the anomalous signal was sufficient for fast\_ep to locate four zinc sites and calculate a very clear experimentally phased electron density map (Fig. 4a). This was available to view at the beamline in the ISPyB database<sup>47</sup> via the SynchWeb interface<sup>48</sup> within a few minutes of completing the data collection. The map was of sufficient quality to enable 94% of the residues expected for a THAS1 homodimer to be automatically fitted using BUCCANEER<sup>49</sup>. The model was finalized by manual rebuilding in COOT<sup>50</sup> and restrained refinement using anisotropic thermal parameters in REFMAC5 (ref. 51) against the same data set reprocessed to a resolution of 1.05  $\text{\AA}$  as described above (Supplementary Table 3), and contained 97% of the expected residues, with one NADP<sup>+</sup> molecule and two zinc ions per subunit. All the remaining structures were solved by molecular replacement using PHASER.<sup>52</sup> In each case, the asymmetric unit corresponded to the biological dimer and the preliminary models were obtained by searching for two copies of a monomer template. For THAS1 apo, THAS2 NADP<sup>+</sup> and HYS apo, a THAS1-NADP<sup>+</sup> protein-only monomer model was used as the basis for the template, although in the latter two cases a homology model of the target structure was generated from the THAS1 template using the Phyre2 server<sup>53</sup> (<http://www.sbg.bio.ic.ac.uk/~phyre2>) before running PHASER.

For solving the THAS2 apo structure, a THAS2 NADP<sup>+</sup> monomer was used as the template. In contrast to THAS1-NADP<sup>+</sup>, these four structures were refined in REFMAC5 with isotropic thermal parameters and TLS group definitions obtained from the TLS-MD server<sup>54</sup>. Model geometries were validated with the MOLPROBITY<sup>55</sup> tool before submission to the PDB. The statistics of the final models are summarized in Supplementary Table 3. Additional statistics for  $R_{\text{pim}}$ : 5F13, 0.020 (0.517); 5F15, 0.038 (0.600); 5H81, 0.041 (0.349); 5H82, 0.033 (0.439); 5H83, 0.068 (0.664) and  $CC_{1/2}$ : 2F13, 0.999 (0.510); 2F15, 0.999 (0.523); 5H81, 0.998 (0.725); 5H82, 0.999 (0.639); 5H83, 0.996 (0.510) (where values in parentheses are for highest-resolution shell) were also noted. Ramachandran statistics (favoured/allowed/outlier (%)) are 5F13, 96.8/3.2/0.0; 5F15, 96.0/4.0/0.0; 5H81, 96.2/3.8/0.0; 5H82, 96.1/3.9/0.0; 5H83, 96.6/3.1/0.3. All structural figures were prepared using CCP4mg<sup>56</sup>.

Simulated annealing omit maps were calculated for the active site regions of all five structures presented in this study. For all structures, selected residues bordering the active site (and the cofactors in the case of the two holoenzyme structures) were deleted from the coordinates of the final models. The resultant PDB files were used as inputs to simulated annealing refinement with PHENIX (<https://www.phenix-online.org>) from a starting temperature of 5,000 K after applying small random shifts to the model ('shake' term set to 0.3). The resultant  $mF_{\text{obs}}-dF_{\text{calc}}$  difference electron density maps (contoured at  $\sim 3\sigma$ ) are displayed superposed on the final coordinates, where the corresponding omitted atoms are shown in stick representation. In each case, structurally equivalent residues were omitted, with the exception of HYS where His127 from loop 2 was also omitted.

**UPLC-MS analysis.** Enzyme assays and plant tissue samples from VIGS experiments on *C. roseus* plants were analysed by UPLC-MS. UPLC-MS analysis was carried out on a UPLC (Waters) equipped with an Acquity BEH C18 1.7  $\mu\text{m}$  2.1  $\times$  50 mm column connected to Xevo TQS (Waters). MS detection was performed in positive ESI. Capillary voltage was 3.0 kV; the source was kept at 150 °C; desolvation temperature was 500 °C; cone gas flow, 50 l h<sup>-1</sup>; and desolvation gas flow, 800 l h<sup>-1</sup>. Unit resolution was applied to each quadrupole. Multiple reaction monitoring signals were used for detection and quantification of caffeine ( $m/z$  195 > 110, 138) and heteroyohimbine alkaloids (353 > 117, 14).

For rapid dereplication of active enzymes and mutants, a linear gradient method (Method 1) was used at a flow rate of 0.6 ml min<sup>-1</sup> using a binary solvent system in which solvent A1 was 0.1% formic acid in water and solvent B1 was acetonitrile. The gradient profile was: 0 min, 5% B1; from 0 to 3.5 min, linear gradient to 35% B1; from 3.5 to 3.75 min, linear gradient to 100% B1; from 3.75 to 4 min, wash at 100% B1; back to the initial conditions of 5% B1 and equilibration for 1 min before the next injection. Column temperature was held at 30 °C. The injection volume for both the solutions of standard compounds and the samples was 1  $\mu\text{l}$ . Samples were kept at 10 °C during the analysis.

For separation of the different heteroyohimbines, a different chromatographic method was applied that was adapted from the work of Sun J. *et al.*<sup>57</sup> In this method (Method 2) solvent A2 was 0.1% NH<sub>4</sub>OH and solvent B2 was 0.1% NH<sub>4</sub>OH in acetonitrile. A linear gradient from 0 to 65% B2 in 17.5 min was applied for separation of the compounds followed by an increase to 100% B2 at 18 min, a 2-min wash step and a re-equilibration at 0% B2 for 3 min before the next injection. The column was kept at 60 °C throughout the analysis and the flow rate was 0.6 ml min<sup>-1</sup>.

**<sup>2</sup>H labelling experiments.** Deuterated Pro-R-NADPD was regenerated in solution by *Thermoanaerobacter brockii* alcohol dehydrogenase (50 U, Sigma) using 400  $\mu\text{M}$  NADP<sup>+</sup> and 1% v/v [<sup>2</sup>H<sub>6</sub>]-isopropanol (CIL). The NADPD regeneration was monitored by ultraviolet spectroscopy at 340 nm. Strictosidine (19.9 mg) was incubated with 1.27 nM SGD in 94 ml of 50 mM phosphate buffer (pH 6.5). THAS1 enzyme was added to the reaction (final concentration of 1.65  $\mu\text{M}$ ) and the mixture was incubated at 35 °C. The reaction was monitored for completeness by UPLC-MS and after 5 h no strictosidine or deglycosylated strictosidine was observed. The reaction was stopped by addition of 100 ml of methanol and reaction mixture was concentrated to dryness. The dried reaction mixture was resuspended in 15 ml H<sub>2</sub>O and extracted with 3  $\times$  15 ml of ethyl acetate and the EtOAc fraction was dried. [21 $\alpha$ -<sup>2</sup>H<sub>1</sub>]-tetrahydroalstonine was isolated by preparative TLC (nano-silica plate, Sigma-Aldrich), as previously described<sup>24</sup>. The band of [21 $\alpha$ -<sup>2</sup>H<sub>1</sub>]-tetrahydroalstonine was excised from the plate and THA was extracted with EtOAc multiple times (total volume 40 ml). The EtOAc fraction was filtered and dried under high-vacuum overnight. The [21 $\alpha$ -<sup>2</sup>H<sub>1</sub>]-tetrahydroalstonine was dissolved in 600  $\mu\text{l}$  of CDCl<sub>3</sub> and <sup>1</sup>H-NMR was measured.

Strictosidine (39.3 mg) was incubated with 1 nM of SGD and 500  $\mu\text{M}$  NADP<sup>+</sup> with 50 U of *T. brockii* ADH and 1% v/v [<sup>2</sup>H<sub>6</sub>]-isopropanol in a total volume of 148 ml of 50 mM HEPES buffer (pH 7.5). HYS was added (final concentration 1.71  $\mu\text{M}$ ) and the reaction was incubated at 37 °C with shaking and monitored for completeness by UPLC-MS. After 6 h, the reaction was complete and was stopped by addition of 150 ml of methanol. [21 $\alpha$ -<sup>2</sup>H<sub>1</sub>]-tetrahydroalstonine, [21 $\alpha$ -<sup>2</sup>H<sub>1</sub>]-ajmalicine and [21 $\alpha$ -<sup>2</sup>H<sub>1</sub>]-mayumbine were isolated by preparative TLC and <sup>1</sup>H-NMR spectra measured as described above.

**<sup>15</sup>N labelling experiments.** *C. roseus* tryptophan decarboxylase (TDC) was cloned into pOPINF vector, expressed in *E. coli* and purified as described above for the MDRs. [alpha-<sup>15</sup>N]-tryptophan (CIL, 50 mg) was incubated with 500 nM of TDC,

400  $\mu\text{M}$  pyridoxal-5'-phosphate in 100 ml 50 mM phosphate buffer (pH 7.5) at 35 °C. The reaction was monitored by MS, continued through completion after 4 h and terminated by addition of 50 ml MeOH. [alpha-<sup>15</sup>N]-tryptamine (34 mg) was isolated by preparative HPLC. The isolated [alpha-<sup>15</sup>N]-tryptamine was incubated with 3 mM secologanin and 200 nM strictosidine synthase in 100 ml 50 mM phosphate buffer (pH 7.0) at 30 °C overnight. The reaction was terminated by addition of 50 ml MeOH. [4-<sup>15</sup>N]-strictosidine (62 mg) was isolated by preparative HPLC. [4-<sup>15</sup>N]-strictosidine was then assayed with SGD and the product was extracted the ethyl acetate, dissolved in MeOH-*d*<sub>4</sub> and characterized by <sup>1</sup>H,<sup>15</sup>N-HMBC as described above.

**Compound characterization.** High-resolution electrospray ionization MS spectra were measured with a Shimadzu IT-TOF mass spectrometer. NMR spectra were acquired using a Bruker Avance NMR instrument operating at 400 MHz for <sup>1</sup>H equipped with a BBFO plus 5-mm probe. The number of scans depended on the concentration of the sample. The <sup>1</sup>H,<sup>15</sup>N-HMBC experiment was acquired with a spectral width 6,009 Hz in the F2 (<sup>1</sup>H) dimension and 30,410 Hz in the F1 (<sup>15</sup>N), with an acquisition time of 0.09 s and 360 scans per increment. The long range delay was optimized after a series of experiments with [4-<sup>15</sup>N]-strictosidine using a range of different mixing times and finally was adjusted for a coupling of 5 Hz. The relaxation delay was 2.5 s, the data collection matrix was 1024  $\times$  64, the t1 dimension was zero filled to 1k real data points and a  $\pi/2$  square sine bell window was applied in both dimensions. The <sup>1</sup>H-NMR spectra were compared with those of standards and literature data.

[21 $\alpha$ -<sup>2</sup>H<sub>1</sub>]-tetrahydroalstonine. TLC (EtOAc: C<sub>6</sub>H<sub>14</sub>: Et<sub>3</sub>N, 24: 75: 1 v/v):  $R_f$  = 0.58; HR-MS (IT-TOF): found for [M + H]<sup>+</sup>: [C<sub>21</sub>H<sub>24</sub>DN<sub>2</sub>O<sub>3</sub>]<sup>+</sup> = 354.1921: calcd 354.19225; <sup>1</sup>H-NMR (400 MHz, CDCl<sub>3</sub>):  $\delta$  7.80 (br s, 1H), 7.56 (s, 1H), 7.45 (dd,  $J$  = 7.6 Hz,  $J$  = 1.8 Hz, 1H), 7.28 (dd,  $J$  = 7.6 Hz,  $J$  = 1.8 Hz, 1H), 7.12 (ddd,  $J$  = 7.6 Hz,  $J$  = 7.6 Hz,  $J$  = 1.8 Hz, 1H), 7.07 (ddd,  $J$  = 7.6 Hz,  $J$  = 7.6 Hz,  $J$  = 1.8 Hz, 1H), 4.50 (dq,  $J$  = 10.4 Hz,  $J$  = 6.3 Hz, 1H), 3.74 (s, 3H), 3.36 (dd,  $J$  = 12.0 Hz,  $J$  = 3.0 Hz, 1H), 3.08 (d,  $J$  = 2.0 Hz, 1H), 2.95 (ddd,  $J$  = 12.2 Hz,  $J$  = 6.1 Hz,  $J$  = 2.5 Hz, 1H), 2.89 (ddd,  $J$  = 14.3 Hz,  $J$  = 6.1 Hz,  $J$  = 2.5 Hz, 1H), 2.77 (ddd,  $J$  = 12.0 Hz,  $J$  = 4.5 Hz,  $J$  = 3.0 Hz, 1H), 2.69 (ddd,  $J$  = 14.3 Hz,  $J$  = 10.7 Hz,  $J$  = 3.6 Hz, 1H), 2.56 (ddd,  $J$  = 12.2 Hz,  $J$  = 10.4 Hz,  $J$  = 4.8 Hz, 1H), 2.49 (ddd,  $J$  = 12.0 Hz,  $J$  = 4.5 Hz,  $J$  = 3.0 Hz, 1H), 1.70 (ddd,  $J$  = 10.4 Hz,  $J$  = 4.7 Hz,  $J$  = 2.0 Hz, 1H), 1.53 (dd,  $J$  = 12.0 Hz,  $J$  = 12.0 Hz, 1H), 1.40 (d, 6.3, 3H).

[21 $\alpha$ -<sup>2</sup>H<sub>1</sub>]-ajmalicine. TLC (EtOAc: C<sub>6</sub>H<sub>14</sub>: Et<sub>3</sub>N, 24: 75: 1 v/v):  $R_f$  = 0.41; HR-MS (IT-TOF): found for [M + H]<sup>+</sup>: [C<sub>21</sub>H<sub>24</sub>DN<sub>2</sub>O<sub>3</sub>]<sup>+</sup> = 354.1921: calcd 354.19225; <sup>1</sup>H-NMR (400 MHz, CDCl<sub>3</sub>):  $\delta$  7.94 (br s, 1H), 7.53 (d,  $J$  = 1.8 Hz, 1H), 7.46 (dd,  $J$  = 7.7 Hz,  $J$  = 1.5 Hz, 1H), 7.30 (dd,  $J$  = 7.7 Hz,  $J$  = 1.5 Hz, 1H), 7.13 (ddd,  $J$  = 7.7 Hz,  $J$  = 7.7 Hz,  $J$  = 1.5 Hz, 1H), 7.08 (ddd,  $J$  = 7.7 Hz,  $J$  = 7.7 Hz,  $J$  = 1.5 Hz, 1H), 4.43 (dq,  $J$  = 13.3 Hz,  $J$  = 6.7 Hz, 1H), 3.74 (s, 3H), 3.41 (ddd,  $J$  = 11.3 Hz,  $J$  = 4.3 Hz,  $J$  = 1.8 Hz, 1H), 3.22 (m, 1H), 3.10 (m, 1H), 3.00 (m, 1H), 2.95 (d,  $J$  = 3.1 Hz, 1H), 2.75 (m, 1H), 2.68 (m, 1H), 2.42 (m, 1H), 2.14 (ddd,  $J$  = 13.3 Hz,  $J$  = 11.4 Hz,  $J$  = 3.1 Hz, 1H), 1.32 (dd,  $J$  = 11.4 Hz,  $J$  = 11.4 Hz, 1H), 1.19 (d, 6.7, 3H).

[21 $\alpha$ -<sup>2</sup>H<sub>1</sub>]-mayumbine. TLC (EtOAc: C<sub>6</sub>H<sub>14</sub>: Et<sub>3</sub>N, 24: 75: 1 v/v):  $R_f$  = 0.45; HR-MS (IT-TOF): found for [M + H]<sup>+</sup>: [C<sub>21</sub>H<sub>24</sub>DN<sub>2</sub>O<sub>3</sub>]<sup>+</sup> = 354.1921: calcd 354.19225; <sup>1</sup>H-NMR (400 MHz, CDCl<sub>3</sub>):  $\delta$  7.96 (br s, 1H), 7.56 (d,  $J$  = 1.8 Hz, 1H), 7.46 (dd,  $J$  = 7.6 Hz,  $J$  = 1.6 Hz, 1H), 7.30 (dd,  $J$  = 7.6 Hz,  $J$  = 1.6 Hz, 1H), 7.13 (ddd,  $J$  = 7.6 Hz,  $J$  = 7.6 Hz,  $J$  = 1.6 Hz, 1H), 7.09 (ddd,  $J$  = 7.6 Hz,  $J$  = 7.6 Hz,  $J$  = 1.6 Hz, 1H), 3.88 (dq,  $J$  = 12.0 Hz,  $J$  = 6.2 Hz, 1H), 3.73 (s, 3H), 3.40 (dd,  $J$  = 11.1 Hz,  $J$  = 4.4 Hz, 1H), 3.16 (ddd,  $J$  = 12.6 Hz,  $J$  = 3.0 Hz,  $J$  = 3.0 Hz, 1H), 3.14 (dd,  $J$  = 12.0 Hz,  $J$  = 6.0 Hz, 1H), 3.10 (d,  $J$  = 3.1 Hz, 1H), 3.03 (ddd,  $J$  = 14.8 Hz,  $J$  = 10.9 Hz,  $J$  = 2.5 Hz, 1H), 2.72 (dd,  $J$  = 16.3 Hz,  $J$  = 6.7 Hz, 1H), 2.40 (ddd,  $J$  = 13.8 Hz,  $J$  = 10.5 Hz,  $J$  = 3.4 Hz, 1H), 1.76 (ddd,  $J$  = 10.3 Hz,  $J$  = 10.3 Hz,  $J$  = 3.3 Hz, 1H), 1.37 (d, 6.3, 3H), 1.30 (ddd,  $J$  = 14.6 Hz,  $J$  = 11.0 Hz, 1H), 1.21 (dd,  $J$  = 16.2 Hz,  $J$  = 6.7 Hz, 1H).

[4-<sup>15</sup>N]-strictosidine. HR-MS (IT-TOF): found for [M + H]<sup>+</sup>: [C<sub>27</sub>H<sub>35</sub>N<sup>15</sup>N<sub>3</sub>O<sub>3</sub>]<sup>+</sup> = 532.2304: calcd 532.23074; <sup>1</sup>H-NMR (400 MHz, CD<sub>3</sub>OD):  $\delta$  10.5 (br s, 1H), 7.77 (s, 1H), 7.44 (dd,  $J$  = 7.8 Hz,  $J$  = 1.9 Hz, 1H), 7.30 (dd,  $J$  = 8.1 Hz,  $J$  = 1.6 Hz, 1H), 7.11 (ddd,  $J$  = 8.1 Hz,  $J$  = 7.0 Hz,  $J$  = 1.9 Hz, 1H), 7.02 (ddd,  $J$  = 7.8 Hz,  $J$  = 7.0 Hz,  $J$  = 1.6 Hz, 1H), 5.84 (ddd,  $J$  = 17.5 Hz,  $J$  = 10.4 Hz,  $J$  = 7.5 Hz, 1H), 5.83 (d,  $J$  = 9.0 Hz, 1H), 5.34 (dd,  $J$  = 17.4 Hz,  $J$  = 2.7 Hz, 1H), 5.26 (dd,  $J$  = 10.6 Hz,  $J$  = 2.6 Hz, 1H), 4.79 (d,  $J$  = 7.9 Hz, 1H), 3.96 (dd,  $J$  = 12.0 Hz,  $J$  = 2.2 Hz, 1H), 3.78 (s, 3H), 3.64 (dd,  $J$  = 11.8 Hz,  $J$  = 6.9 Hz, 1H), 3.42–3.30 (m, 3H), 3.23 (dd,  $J$  = 9.8 Hz,  $J$  = 8.9 Hz, 1H), 3.22 (dd,  $J$  = 9.3 Hz,  $J$  = 7.9 Hz, 1H), 3.10–3.00 (m, 2H), 2.97 (ddd,  $J$  = 15.9 Hz,  $J$  = 4.9 Hz,  $J$  = 1.3 Hz, 1H), 2.72 (ddd,  $J$  = 8.5 Hz,  $J$  = 7.9 Hz,  $J$  = 4.6 Hz, 1H), 2.27 (dd,  $J$  = 14.6 Hz,  $J$  = 11.5 Hz), 2.18 (dd,  $J$  = 11.4 Hz,  $J$  = 3.8, 1H); <sup>15</sup>N-NMR (40 MHz, CD<sub>3</sub>OD):  $\delta$  44.8.

**Subcellular localizations and analysis of protein-protein interactions by BiFC.**

Subcellular localization of THAS2 and HYS were studied by creating fluorescent fusion proteins using the pSCA-cassette YFPi plasmid<sup>58</sup>. The full-length open reading frame of THAS2 was amplified using the specific primers 5'-CTGAGAA CTAGTATGTCTTCAAAAATCAGCAAAAACAGTG-3' and 5'-CTGAGAACTA GTAGCAGATTTCAAATGTGTTTTCTATGCTCAAT-3', and HYS ORF with primers 5'-CTGAGAACTAGTATGGCTGCAAAAGTCACCTGAAAATGT ATAC-3' and 5'-CTGAGAACTAGTAAAGATGGGATTTGAGAGTG TTTCCTAC-3', which were designed to introduce the *SpeI* restriction site at both

cDNA extremities. PCR products were sequenced and cloned at the 5' end of the YFP-coding sequence to generate the THAS2–YFP, HYS–YFP fusion proteins or at the 3' end to express the YFP–THAS2 and YFP–HYS fusions.

The interaction of THAS2 and HYS with SGD were characterized by BiFC assays using the previously amplified THAS2 and HYS PCR products cloned via *SpeI* into the pSPYCE (M) vector<sup>34</sup>, which allows expression of THAS2 and YFP fused to the amino-terminal extremity of the split-YFP<sup>C</sup> fragment (THAS2–YFP<sup>C</sup>, HYS–YFP<sup>C</sup>, respectively), and into the pSPYNE(R)173–SGD plasmid<sup>34</sup> expressing SGD fused to the carboxy-terminal extremity of the split YFP<sup>N</sup> fragment (YFP<sup>N</sup>–SGD). Plasmids encoding THAS1–YFP<sup>N</sup>, THAS1–YFP<sup>C</sup>, YFP<sup>C</sup>–THAS1 and plasmids expressing 16OMT–YFP<sup>N</sup> and 16OMT–YFP<sup>C</sup> were used as controls and were constructed previously<sup>24,59</sup>.

THAS2 and HYS self-interactions were analysed via additional cloning of the THAS2 and HYS PCR products into the pSCA–SPYNE173, pSPYNE(R)173 and pSCA–SPYCE (MR) plasmids<sup>34,60</sup> to express THAS2–YFP<sup>N</sup>, HYS–YFP<sup>C</sup> and YFP<sup>C</sup>–THAS2, YFP<sup>C</sup>–HYS, respectively.

The capacities of THAS2 and HYS to interact with SGD were also characterized by double BiFC and mBiFC. The previously amplified THAS2 and HYS PCR products were fused to the coding sequences of the amino-terminal or carboxy-terminal of the split YFP fragments into the pSCA–SPYNE173, pSCA–SPYCE (M) and pSCA–SPYCE (MR) plasmids<sup>34,60</sup>, allowing expression of THAS2–YFP<sup>N</sup>, YFP<sup>C</sup>–THAS2, HYS–YFP<sup>N</sup> and HYS–YFP<sup>C</sup> respectively. SGD was subsequently fused to the carboxy-terminal extremity of the split YFP<sup>N</sup> fragment (YFP<sup>N</sup>–SGD) and the CFP<sup>N</sup> fragment (CFP<sup>N</sup>–SGD).

Transient transformation of *C. roseus* cells by particle bombardment and fluorescence imaging were performed following the procedures previously described<sup>58</sup>. Briefly, *C. roseus*-plated cells were bombarded with DNA-coated gold particles (1 µm) and 1,100 psi rupture disc at a stopping-screen-to-target distance of 6 cm, using the Bio-Rad PDS1000/He system. Cells were cultivated for 16–38 h before being harvested 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 CFP filters. The pattern of localization presented in this work is representative of circa 50 observed cells. The nuclear localizations of the different fusion proteins were confirmed by co-transformation experiments using a nuclear-CFP marker<sup>34</sup>. Such plasmid transformations were performed using 400 ng of each plasmid or 100 ng for BiFC assays.

**Agrobacterium VIGS and qPCR.** The THAS1, THAS2, THAS3 and THAS4–HYS silencing fragments were amplified with primers (Supplementary Table 6) and the resulting fragments were cloned into the pTRV2u vector as described<sup>61</sup>. Since THAS4 and HYS are ~91% identical it was not possible to design silencing fragments to avoid cross-silencing. Therefore, a common silencing fragment for both of the two genes was designed. The resulting pTRV2u constructs were used to silence the different THASs and HYS in *C. roseus* seedlings essentially as described before<sup>55</sup>. Leaves from the first two pairs to emerge following inoculation were harvested from eight plants transformed with the empty pTRV2u and pTRV2u carrying the silencing fragment. The collected leaves were frozen in liquid nitrogen, powdered using a pre-chilled mortar and pestle, and subjected to LC-MS and qRT–PCR analysis. The heteroyohimbine content of silenced leaves was determined by LC-MS. Leaf powder was weighed (10–20 mg), extracted with methanol (2 ml) and vortexed for 1 min. After a 10-min centrifugation step at 17,000g, an aliquot of the supernatant (20 µl) was diluted to 200 µl with methanol, filtered through 0.2-µm PTFE filters and analysed on Waters Xevo TQ-MS. The chromatographic separation and MS measurements were carried out as described above (method 2). To more comprehensively assess the global effect of silencing the HYS genes by VIGS on *C. roseus* metabolism, an untargeted metabolomics analysis by LC-MS was performed as previously reported<sup>24</sup>. However, aside from the changes in the heteroyohimbines reported in Supplementary Fig. 16, no substantial differences in metabolic profiles were noted using this approach.

Gene silencing was confirmed by qRT–PCR. qRT–PCR was also used to check the expression of the other HYS genes to ensure that no cross-silencing occurred. RNA extraction was performed using the 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) using the SSO Advanced SYBR Green Supermix (Bio-Rad). Each reaction was performed in a total reaction volume of 20 µl containing an equal amount of cDNA, 0.25 mM forward and reverse primers and 1 × Sso Advanced SYBR Green Supermix (Bio-Rad). The reaction was initiated by a denaturation step at 95 °C for 10 min followed by 41 cycles at 95 °C for 15 s and 60 °C for 1 min. Melting curves were used to determine the specificity of the amplifications. Relative quantification of gene expression was calculated according to the delta–delta cycle threshold method using the 40S ribosomal protein S9 (RPS9). All primer pair (Supplementary Table 7) efficiencies were between 98 and 108%, 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.

**pH effect on product profile.** Strictosidine was deglycosylated using purified SGD for 25 min at room temperature using assay conditions as described above.

Strictosidine aglycone was then incubated at a final concentration of 300 µM at pH 5, 6, 7 and 8 in a buffer mix to avoid buffer ingredient effect on activity ((50 mM phosphate buffer, 50 mM citric acid, 50 mM HEPES). Caffeine (50 µM) was used as an internal standard.

At time 0 the enzyme, either THAS1 or HYS (1 µM final concentration), premixed with NADPH (500 µM) was added to the substrate solution. In parallel, a chemical reducing agent, NaBH<sub>4</sub> (3 mM final concentration), was added to deglycosylated strictosidine as a control reaction. All reactions were carried out in triplicate. An end-point sample (10 µl) was taken for each assay and prepared for UPLC-MS by addition of 10 µl of 100% MeOH to stop the reaction, and then diluted 1:5 with H<sub>2</sub>O, and centrifuged for 10 min at 4,000 r.p.m. UPLC-MS and data collection were performed as described above for heteroyohimbine separation and quantification.

**CD spectra and analysis.** Far ultraviolet CD spectra of the wild-type enzymes THAS1 and HYS, as well as the loop mutants of THAS1 and HYS were recorded on a Chirascan Plus spectropolarimeter (Applied Photophysics) at 20 °C in 10 mM potassium phosphate buffer pH 7.0. Samples were analysed from 180 to 260 nm using a 0.5-nm step at a speed of 1 s per step. Four replicate measurements were performed on each sample and baseline correction was applied to all data. Spectra are presented as the CD absorption coefficient calculated on a mean residue ellipticity basis.

Melting curves of HYS and the HYS loop 2 swap mutant were also acquired by CD. The samples were subjected to temperature ramping at the rate of 1 °C min<sup>-1</sup> from 20 to 90 °C. Data collection was done from 260 to 201 nm using a 1-nm step and 0.75 s time per point. Data were analysed using the Global 3 software. HYS melting point was measured as 61.0 ± 0.1 °C; enthalpy 351.5 ± 3.6 KJ mol<sup>-1</sup>. HYS loop 2 swap melting point was measured at 62.0 ± 0.1 °C; enthalpy 535.8 ± 4.5 KJ mol<sup>-1</sup>.

**Protein sequence alignments and phylogenetic tree.** Protein sequence alignment was generated using ClustalW algorithm with Geneious v.8 (<http://www.geneious.com>)<sup>62,63</sup>. The alignment was edited manually using Seaview V4 (ref. 64) and secondary structure depiction was added using ESPript V3 (<http://esprict.ibcp.fr>)<sup>65</sup>. Phylogenetic analysis was performed using the neighbour-joining<sup>66</sup> algorithm and Bootstrap analysis with 1,000 replicates.

**Docking of cathenamine in THAS1–NADP<sup>+</sup> structure.** Cathenamine was docked into the THAS1–NADP<sup>+</sup> crystal structure using Autodock 4.2 (ref. 67). The ligand (cathenamine) was prepared with two torsions at the C16, the rest of the molecule being rigid and the receptor consisted of the desolvated high-resolution crystal structure. The search space was defined by a 40 × 40 × 40 Å box with a 0.375-Å grid spacing, centred between the nicotinamide ring and the side chain of Tyr56, and encompassed the entire active site cavity. Searches were performed using the Lamarckian Genetic Algorithm, consisting of 100 runs with a population size of 150 and 2,500,000 energy evaluations. A total of 27,000 generations were analysed and clustered with an RMS tolerance of 2 Å per cluster. This resulted in just two distinct clusters, which constituted 98% and 2% of the resultant poses, respectively. The latter cluster placed the indole moiety such that the nitrogen atom was closest to the cofactor. Thus, this cluster was eliminated since it was inconsistent with the results of the deuterium labelling experiments (Fig. 6). The poses contained within the major cluster were all deemed to be 'productive' since they placed the indole moiety of cathenamine towards the entrance of the active site and the C20 and C21 3.3 Å above the nicotinamide C4 atom. The top ranked pose (with an estimated free energy of binding = -8.76 kcal mol<sup>-1</sup>), as selected by the software, is used in the structures illustrated here (Fig. 4b).

**Data availability.** The atomic coordinates and structure factors of the five X-ray structures described in this manuscript have been deposited in the Protein Data Bank (<http://www.pdb.org/>), with accession codes 5FI3, 5FI5, 5H81, 5H82 and 5H83. Accession numbers: THAS1 (AKF02528.1); THAS2 (KU865323); THAS3 (KU865322); THAS4 (KU865324); HYS (KU865325); Cro\_017994 (KU865326); Cro\_011702 (KU865327); Cro\_030442 (KU865328); Cro\_006840 (KU865329); Cro\_022770 (KU865330); Cro\_033537 (KU865331); Cro\_027234 (AHK60846); Cro\_033062 (KU865332); Tabersonine-3-reductase (AKM12281). Data supporting the findings of this study are available within the article and its Supplementary Information files and from the corresponding author upon reasonable request.

## References

- Shamma, M. & Richey, J. M. The stereochemistry of the heteroyohimbine alkaloids. *J. Am. Chem. Soc.* **85**, 2507–2512 (1963).
- Allain, H. & Bentue-Ferrer, D. Clinical efficacy of almitrine-raubasine. An overview. *Eur. Neurol.* **39**, 39–44 (1998).
- Benzi, G. Pharmacological features of an almitrine-raubasine combination. Activity at cerebral levels. *Eur. Neurol.* **39**, 31–38 (1998).
- Li, S. *et al.* Assessment of the therapeutic activity of a combination of almitrine and raubasine on functional rehabilitation following ischaemic stroke. *Curr. Med. Res. Opin.* **20**, 409–415 (2004).

5. Roquebert, J. & Demichel, P. Inhibition of the  $\alpha_1$ - and  $\alpha_2$ -adrenoceptor-mediated pressor response in pithed rats by raubasine, tetrahydroalstonine and akuammigine. *Eur. J. Pharmacol.* **106**, 203–205 (1984).
6. Ai, J., Dekermendjian, K., Nielsen, M. & Witt, M. R. The heteroyohimbine mayubine binds with high affinity to rat brain benzodiazepine receptors in vitro. *Nat. Prod. Lett.* **11**, 73–76 (1997).
7. Dassonneville, L. *et al.* Stimulation of topoisomerase II-mediated DNA cleavage by three DNA-intercalating plant alkaloids: cryptolepine, matadine, and serpentine. *Biochemistry* **38**, 7719–7726 (1999).
8. Costa-Campos, L., Iwu, M. & Elisabetsky, E. Lack of pro-convulsant activity of the antipsychotic alkaloid alstonine. *J. Ethnopharmacol.* **93**, 3017–3310 (2004).
9. Elisabetsky, E. & Costa-Campos, L. The alkaloid alstonine: A review of its pharmacological properties. *eCAM* **3**, 39–48 (2006).
10. Herrmann, A. P. *et al.* Effects of the putative antipsychotic alstonine on glutamate uptake in acute hippocampal slices. *Neurochem. Int.* **61**, 1144–1150 (2012).
11. Linck, V. M. *et al.* Alstonine as an antipsychotic: effects on brain amines and metabolic changes. *eCAM*, 418597 (2011).
12. Linck, V. M. *et al.* 5-HT<sub>2A/C</sub> receptors mediate the antipsychotic-like effects of alstonine. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **36**, 29–33 (2012).
13. Linck, V. M. *et al.* Original mechanisms of antipsychotic action by the indole alkaloid alstonine (*Picralima nitida*). *Phytomedicine* **22**, 52–55 (2015).
14. Saxton, J. E. *The Chemistry of Heterocyclic Compounds, Indoles: the Monoterpenoid Indole Alkaloids* (Wiley, 2009).
15. Amer, M. A. & Court, W. E. P. Alkaloids of *Rauwolfia nitida* root bark. *Phytochemistry* **20**, 2569–2573 (1981).
16. Hochstein, F. A. Alkaloids of *Rauwolfia sellowii*. *J. Am. Chem. Soc.* **77**, 5744–5745 (1955).
17. Melchio, J., Bouquet, A., Pais, M. & Goutarel, R. Alcaloides indoliques CVI (1) Identité de la mayubine et de l'épi-19 ajmalicine. L'iso-3 rauniticine, un nouvel alcaloïde extrait du *Corynanthe mayumbensis* (R. Good) N. Hallé. *Tetrahedron Lett.* **18**, 315–316 (1977).
18. Phillipson, J. D. & Supavita, N. Alkaloids of *uncaria elliptica*. *Phytochemistry* **22**, 1809–1813 (1983).
19. Ponglux, D., Supavita, T., Verpoorte, R. & Phillipson, D. P. Alkaloids of *Uncaria attenuata* from Thailand. *Phytochemistry* **19**, 2013–2016 (1980).
20. Robinson, R. & Thomas, A. F. The alkaloids of *picralima nitida*, Stapf, Th. and H. Durand. Part I. The structure of akuammigine. *J. Chem. Soc.* 3479–3482 (1954).
21. Stoekigt, J., Husson, H. P., Kan-Fan, C. & Zenk, M. H. Cathenamine, a central intermediate in the cell free biosynthesis of ajmalicine and related indole alkaloids. *J. Chem. Soc. Chem. Commun.* 164–166 (1977).
22. O'Connor, S. E. & Maresh, J. J. Chemistry and biology of monoterpene indole alkaloid biosynthesis. *Nat. Prod. Rep.* **23**, 532–547 (2006).
23. Gerasimenko, I., Sheludko, Y., Ma, X. & Stöckigt, J. Heterologous expression of a *Rauwolfia* cDNA encoding strictosidine glucosidase, a biosynthetic key to over 2000 monoterpene indole alkaloids. *Eur. J. Biochem.* **269**, 2204–2213 (2002).
24. Stavrinides, A. *et al.* Unlocking the diversity of alkaloids in *Catharanthus roseus*: nuclear localization suggests metabolic channeling in secondary metabolism. *Chem. Biol.* **22**, 336–341 (2015).
25. Gongora-Castillo, E. *et al.* Development of transcriptomic resources for interrogating the biosynthesis of monoterpene indole alkaloids in medicinal plant species. *PLoS ONE* **7**, e25206 (2012).
26. Kellner, F. *et al.* Genome-guided investigation of plant natural product biosynthesis. *Plant J.* **82**, 680–692 (2015).
27. Bomati, E. K. & Noel, J. P. Structural and kinetic basis for substrate selectivity in *Populus tremuloides* sinapyl alcohol dehydrogenase. *Plant Cell* **17**, 1698–1611 (2005).
28. Pan, H. *et al.* Structural studies of cinnamoyl-CoA reductase and cinnamyl-alcohol dehydrogenase, key enzymes of monolignol biosynthesis. *Plant Cell* **26**, 3709–3727 (2014).
29. Youn, B. *et al.* Crystal structures and catalytic mechanism of the Arabidopsis cinnamyl alcohol dehydrogenases AtCAD5 and AtCAD4. *Org. Biomol. Chem.* **4**, 687–1697 (2006).
30. Auld, D. S. & Bergman, T. Medium- and short-chain dehydrogenase/reductase gene and protein families: The role of zinc for alcohol dehydrogenase structure and function. *Cell. Mol. Life Sci.* **65**, 3961–3970 (2008).
31. Eklund, H. & Ramaswamy, S. Medium- and short-chain dehydrogenase/reductase gene and protein families: Three-dimensional structures of MDR alcohol dehydrogenases. *Cell. Mol. Life Sci.* **65**, 3907–3917 (2008).
32. Stoekigt, J., Hemscheidt, T., Hoefle, G., Heinstejn, P. & Formacek, V. Steric course of hydrogen transfer during enzymatic formation of 3 $\alpha$ -heteroyohimbine alkaloids. *Biochemistry* **22**, 3448–3452 (1983).
33. Knowles, J. R. & Jencks, W. P. The intrinsic pka-values of functional groups in enzymes: Improper deductions from the pH-dependence of steady-state parameter. *Crit. Rev. Biochem.* **4**, 165–173 (1976).
34. Guirimand, G. *et al.* Strictosidine activation in Apocynaceae: towards a "nuclear time bomb"? *BMC Plant Biol.* **10**, 182 (2010).
35. 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).
36. Brown, S., Clastre, M., Courdavault, V. & O'Connor, S. E. De novo production of the plant-derived alkaloid strictosidine in yeast. *Proc. Natl Acad. Sci. USA* **112**, 3205–3210 (2015).
37. Konno, K., Hirayama, C., Yasui, H. & Nakamura, M. Enzymatic activation of oleuropein: A protein crosslinker used as a chemical defense in the privet tree. *Proc. Natl Acad. Sci. USA* **96**, 9159–9164 (1999).
38. Pankoke, H., Buschmann, T. & Muller, C. Role of plant beta-glucosidases in the dual defense system of iridoid glycosides and their hydrolyzing enzymes in *Plantago lanceolata* and *Plantago major*. *Phytochemistry* **94**, 99–107 (2013).
39. Frelin, O. *et al.* A directed-overflow and damage-control N-glycosidase in riboflavin biosynthesis. *Biochem. J.* **466**, 137–145 (2015).
40. Qu, Y. *et al.* Completion of the seven-step pathway from tabersonine to the anticancer drug precursor vindoline and its assembly in yeast. *Proc. Natl Acad. Sci. USA* **112**, 6224–6229 (2015).
41. Berrow, N. S. *et al.* A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res.* **35**, e45 (2007).
42. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
43. Evans, P. R. & Murshudov, G. N. Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1204–1214 (2013).
44. Winter, G. Xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43**, 186–190 (2010).
45. Winter, G. & McAuley, K. E. Automated data collection for macromolecular crystallography. *Methods* **55**, 81–93 (2011).
46. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. Sect. A* **64**, 112–122 (2008).
47. Delageniere, S. *et al.* ISPyB: an information management system for synchrotron macromolecular crystallography. *Bioinformatics* **27**, 3186–3192 (2011).
48. Fisher, S. J., Levik, K. E., Williams, M. A., Ashton, A. W. & McAuley, K. E. SynchWeb: a modern interface for ISPyB. *J. Appl. Crystallogr.* **48**, 927–932 (2015).
49. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. Sect. D* **62**, 1002–1011 (2006).
50. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
51. Winn, M. D., Murshudov, G. N. & Papiz, M. Z. Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol.* **374**, 300–321 (2003).
52. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
53. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
54. Painter, J. & Merritt, E. A. TLSMD web server for the generation of multi-group TLS models. *J. Appl. Crystallogr.* **39**, 109–111 (2006).
55. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).
56. McNicholas, S., Potterton, E., Wilson, K. S. & Noble, M. E. Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 386–394 (2011).
57. Sun, J., Baker, A. & Chen, P. Profiling the indole alkaloids in yohimbe bark with ultra-performance liquid chromatography coupled with ion mobility quadrupole time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **25**, 2591–2602 (2011).
58. Guirimand, 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).
59. Guirimand, G. *et al.* Spatial organization of the vindoline biosynthetic pathway in *Catharanthus roseus*. *J. Plant Physiol.* **168**, 549–557 (2011).
60. Waadt, R. *et al.* Multicolor bimolecular fluorescence complementation reveals simultaneous formation of alternative CBL/CIPK complexes in planta. *Plant J.* **56**, 505–516 (2008).
61. Geu-Flores, F. *et al.* An alternative route to cyclic terpenes by reductive cyclization. *Nature* **492**, 138–142 (2012).
62. Kears, M. *et al.* Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).
63. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
64. Gouy, M., Guindon, S. & Gascuel, O. Seaview version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **27**, 221–224 (2010).

65. Gouet, P., Robert, X. & Courcelle, E. ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res.* **31**, 3320–3323 (2003).
66. Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987).
67. Morris, G. M. *et al.* Autodock4 and autodocktools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* **30**, 2785–2791 (2009).

### Acknowledgements

We gratefully acknowledge support from the ERC (311363) and a BBSRC Institute Strategic Programme grant (MET; BB/J004561/1) to S.E.O.C. and from the Region Centre (France, ABISAL grant) to V.C. A.S. is supported by a BBSRC DTP studentship. The Diamond Light Source provided access to beamlines I03, I04 and I04-1 (proposal MX9475).

### Author contributions

A.S. and E.C.T. and S.E.O.C. designed the project; A.S., E.C.T. and L.C. performed molecular cloning/enzyme assays; L.C., A.S., C.E.M.S. and D.M.L. assisted with crystallization, X-ray data acquisition and structure refinement; E.C.T. and L.C. performed VIGS; E.C.T. performed NMR structural characterization; E.F. and V.C. performed all localization experiments; S.E.O.C. supervised the work; A.S. and E.C.T. and S.E.O.C. wrote the manuscript with input from all authors.

### Additional information

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing financial interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npj.nature.com/reprintsandpermissions>.

**How to cite this article:** Stavrinides, A. *et al.* Structural investigation of heteroyohimbine alkaloid synthesis reveals active site elements that control stereoselectivity. *Nat. Commun.* **7**:12116 doi: 10.1038/ncomms12116 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016