

Nicotine biosynthesis is completed by cryptic activating glucosylation

Received: 11 November 2025

Accepted: 21 April 2026

Published online: 18 May 2026



Benjamin T. W. Schwabe^{1,2}, Isabelle M. Angstman³, Katharina Vollheyde³,
Zoe Ingold^{1,2}, Jiacheng Li², Ksenia S. Stankevich^{2,4},
Christopher D. Spicer^{2,4}, Martin A. Fascione^{2,4}, Gideon Grogan²,
Fernando Geu-Flores³ & Benjamin R. Lichman¹ ✉

Nicotine is a neuroactive alkaloid produced by tobacco (*Nicotiana tabacum*) as a defence against herbivory, and an addictive stimulant that has been used by humans for millennia. Despite its significance, the core steps of its biosynthesis have remained elusive. Here, we demonstrate the *in vitro* reconstruction of a four-enzyme stereoselective biocatalytic cascade that forms (S)-nicotine from nicotinic acid and N-methylpyrrolinium. This cascade includes two glucose-processing enzymes that participate in a cryptic activating glucosylation step. We also reconstruct this pathway *in planta* and present high resolution X-ray structures of the key carbon-carbon bond forming reductase-oxidase pair bound to their substrate and product, respectively. This work establishes the complete biosynthetic pathway to nicotine, providing new gene targets for controlling alkaloid production in *Nicotiana* and unlocking enzymatic routes to pyridine alkaloids.

Nicotine (**1**) is a plant-derived alkaloid of profound societal and scientific importance. It is the major neuroactive and addictive component of cultivated tobacco (*Nicotiana tabacum*) products, which have been used by humans for over 10,000 years¹ and are responsible for an ongoing global health epidemic². Its neurological properties are due to its agonistic interaction with eponymous nicotinic acetylcholine receptors³. The alkaloid protects plants from herbivory⁴ and its insecticidal activity led to its use as an agricultural pesticide, prior to replacement by neo-nicotinoids⁵.

Tobacco and its alkaloids have formed the basis of extensive investigations into plant biology including chemical ecology⁶, metabolite transport⁷, genetic regulation⁸ and metabolic evolution⁹. Recently, tobacco species, especially *Nicotiana benthamiana*, have gained traction as chassis for the recombinant production of therapeutic molecules and proteins, including vaccines¹⁰. However, the presence of nicotine can cause challenges for downstream processing¹¹. Despite its notability, nicotine's biosynthesis has not been elucidated: solving it will enable genetic manipulation of the pathway

in native and heterologous producers, and will provide biocatalytic tools for the formation of pyridine-containing compounds.

Alkaloid scaffolds are typically formed via Mannich-like reactions, wherein an iminium electrophile is attacked by a carbon nucleophile, forming a carbon-carbon bond and a chiral center¹². Nicotine (**1**) consists of pyridine and N-methylpyrrolidine rings, derived from nicotinic acid (**2**) and N-methylpyrrolinium (**3**) respectively (Fig. 1)^{13,14}. Labelled precursor feeding studies indicate that **2** undergoes C6-reduction and decarboxylation to yield the unusual nucleophile 1,2-dihydropyridine (**4a**), which reacts at C3 with the electrophilic iminium **3** to form 3,6-dihydronicotine (**5a**) followed by stereoselective removal of the C6-hydrogen to form **1** (Fig. 1, red hydrogen)^{15–22}. The formation of **1** appears to be stereoselective, as the (S)-**1** enantiomer makes up 96% of **1** formed in *N. tabacum* roots, prior to its enrichment in leaves through enantioselective demethylation of (R)-**1**²³. Related pyridine alkaloids anabasine (**6**) and anatabine (**7**) are formed when **3** is replaced by lysine-derived Δ^1 -piperidine (**8**) or nicotinic acid-derived 2,5-dihydropyridinium (**9**) respectively (Fig. 1)²⁴.

¹Centre for Novel Agricultural Products, Department of Biology, University of York, York, UK. ²Department of Chemistry, University of York, York, UK. ³Section for Plant Biochemistry and Copenhagen Plant Science Centre, Department of Plant and Environmental Sciences, University of Copenhagen, Frederiksberg, Denmark. ⁴York Biomedical Research Institute, University of York, York, UK. ✉e-mail: benjamin.lichman@york.ac.uk

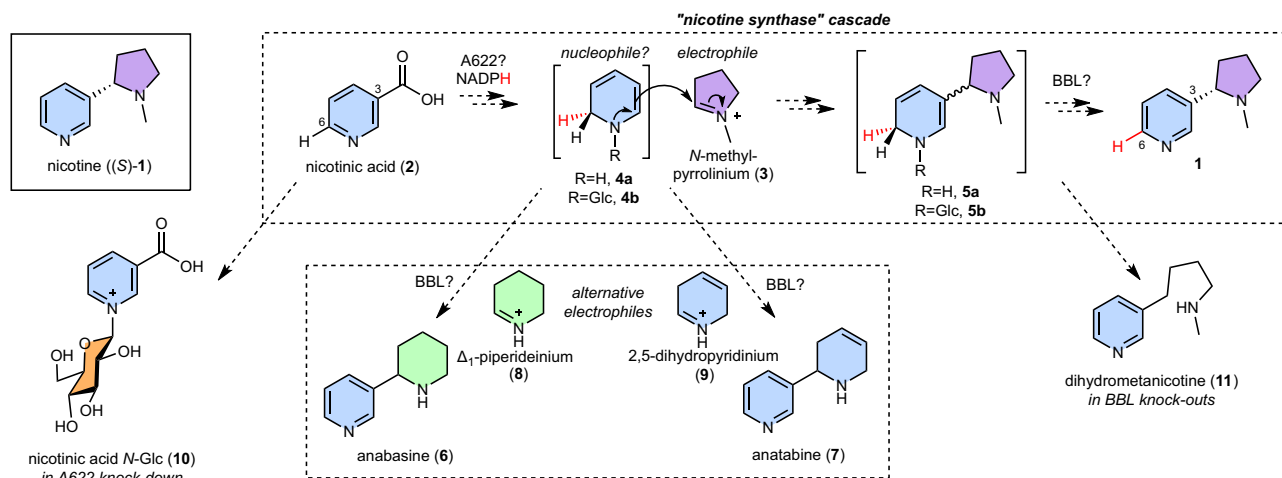


Fig. 1 | Proposed biosynthesis of nicotine. Hypothesized biosynthetic route to (S)-nicotine ((S)-1), showing origins of the pyridine (blue) and N-methylpyrrolidine (purple) rings. The Mannich-like scaffold formation is proposed to proceed via a

multi-step ‘nicotine synthase’ cascade¹⁶. Compounds that accumulate in *Nicotiana* silencing/knock-out experiments are depicted. The red hydrogens follow the fate of the hydride transferred via enzymatic reduction of **2**.

In 1990, the isolation of ‘nicotine synthase’ was reported, a crude enzyme preparation from *N. tabacum* capable of forming (S)-**1** from **2** and **3**¹⁶. Subsequently, A622 and BBL were described, two oxidoreductases involved in nicotine scaffold formation, first identified in low-nicotine tobacco mutants^{25–31}. However, their substrates and products have not been determined. Silencing of A622 and its close homolog A622L in hairy root cultures led to a reduction in pyridine alkaloids and an increase in nicotinic acid N-glucoside (**10**) (Fig. 1)²⁶, suggesting that A622, a cytosolically localised isoflavone reductase-like enzyme³², is responsible for the reduction of **2** or a derivative. However, recombinant enzyme assays failed to demonstrate activity of A622 with **2** or **10**²⁶. BBLs encode vacuolar-localised flavin-containing oxidases from the berberine bridge enzyme family³³. RNAi-based suppression or CRISPR-based inactivation of BBLs led to a decrease in alkaloid content and an accumulation of dihydrometanicotine (**11**), which may be derived from dihydronicotine (**5a**) (Fig. 1)^{30,31}. Furthermore, residual **1** in quintuple BBL knockouts was racemic and displayed a labelling pattern consistent with disruption of stereoselective H-loss at C6³¹. Single gene knockouts suggest that BBLa and BBLb are responsible for (S)-**1** formation whereas BBLc may be responsible for (R)-**1** formation³⁴. Whilst it appears that BBL paralogs have roles in both oxidation and stereochemical control, the accumulation of **11** in knockouts suggests they are not responsible for the ring coupling step.

It is nearing two hundred years since the first isolation of nicotine, in 1828, by Posselt and Reimann³⁵. Since then, the nicotine biosynthesis pathway has been under intense scrutiny due to its commercial and scientific significance. Nevertheless, progress in defining the crucial step in its formation—the coupling of the two rings—has been remarkably static, with current conceptions remaining essentially unchanged since those proposed in the 1960s¹⁵.

In this work, we resolve this enduring mystery, discovering two glucosyl-processing enzymes that complete nicotine biosynthesis via a cryptic activating glucosylation. We demonstrate this by reconstructing in vitro a four-enzyme stereoselective biocatalytic cascade that forms (S)-nicotine from nicotinic acid and N-methylpyrrolinium. Furthermore, X-ray crystal structures of the key oxidoreductases, alongside *in planta* pathway reconstruction and metabolite analysis, reveal the molecular and enzymatic basis for nicotine biosynthesis.

Results

Proposal of glucosylated intermediates in nicotine biosynthesis

Genes involved in plant specialised metabolism pathways can be arranged in biosynthetic gene clusters, where non-homologous but

functionally related genes are found in close genomic proximity³⁶. We investigated the genomic context of nicotine biosynthesis genes (Supplementary Data 1), and identified a cluster consisting of A622 (884g0010 from Edwards et al.³⁷), the nicotine-related tonoplasmic MATE1 transporter (884g0030)^{26,38,39} and a previously unreported gene predicted to encode a vacuolar localised β -glucosidase (β -GD1, 884g0020) (Fig. 2A and Supplementary Fig. 1). We identified the homeologous cluster on chromosome 12, featuring A622L (31g0370), MATE2 (31g0420) and a β -GD paralog (β -GD2, 31g0400) (Fig. 2A). All clustered genes, as well as an unclustered β -GD paralog (β -GD3, 795g0060) (Supplementary Data 1), had highest expression in roots and co-expressed tightly with known nicotine biosynthesis genes (Fig. 2B, C and Supplementary Data 2)⁴⁰.

The clustering and co-expression pattern of the β -GDs suggested they are involved in nicotine biosynthesis, implying on-pathway deglucosylation. This led us to reconsider nicotinic acid N-glucoside (**10**) as the substrate of A622, hypothesising that 1,2-dihydropyridine glucoside (**4b**, R=Glc) and 1,2-dihydropyridine glucoside (**5b**, R=Glc) may be pathway intermediates (Fig. 1). Therefore, the remaining step requiring a candidate gene was the glucosylation of nicotinic acid (**2**). We identified a UDP-dependent glucosyltransferase (UGT) candidate (UGT1, 6222g0020) whose expression mirrored A622 expression after topping, a process that elicits nicotine biosynthesis (Fig. 2D)⁴¹. UGT1 expression closely correlates with nicotine biosynthesis genes across multiple tissues (Fig. 2B, C and Supplementary Data 3). It was given the systematic UDP-glucosyltransferase name UGT709L18. Having identified β -GD and UGT candidates and proposed functions for A622 and BBL, we set out to test our hypothesis through pathway reconstruction.

In vitro reconstitution of nicotine biosynthesis

We first reconstructed the nicotine biosynthesis pathway in vitro, selecting this simple system as it lacks interfering enzymes, non-relevant metabolites and complications of subcellular and tissue localisation. For expediency, we elected to test a single representative paralog of each enzyme type. Four heterologously expressed and purified enzymes (UGT1, A622, BBLa and β -GD1, Supplementary Data 4, and Supplementary Fig. 2) were combined in one-pot reactions alongside co-substrates (UDP-glucose, NADPH) and precursors (Fig. 3). In the presence of all four enzymes, (S)-nicotine ((S)-**1**) was formed from nicotinic acid (**2**) and N-methylpyrrolinium (**3**), reconstituting the activity previously observed in the ‘nicotine synthase’ crude enzyme preparation (Figs. 3; 4A and Supplementary Fig. 3)¹⁶.

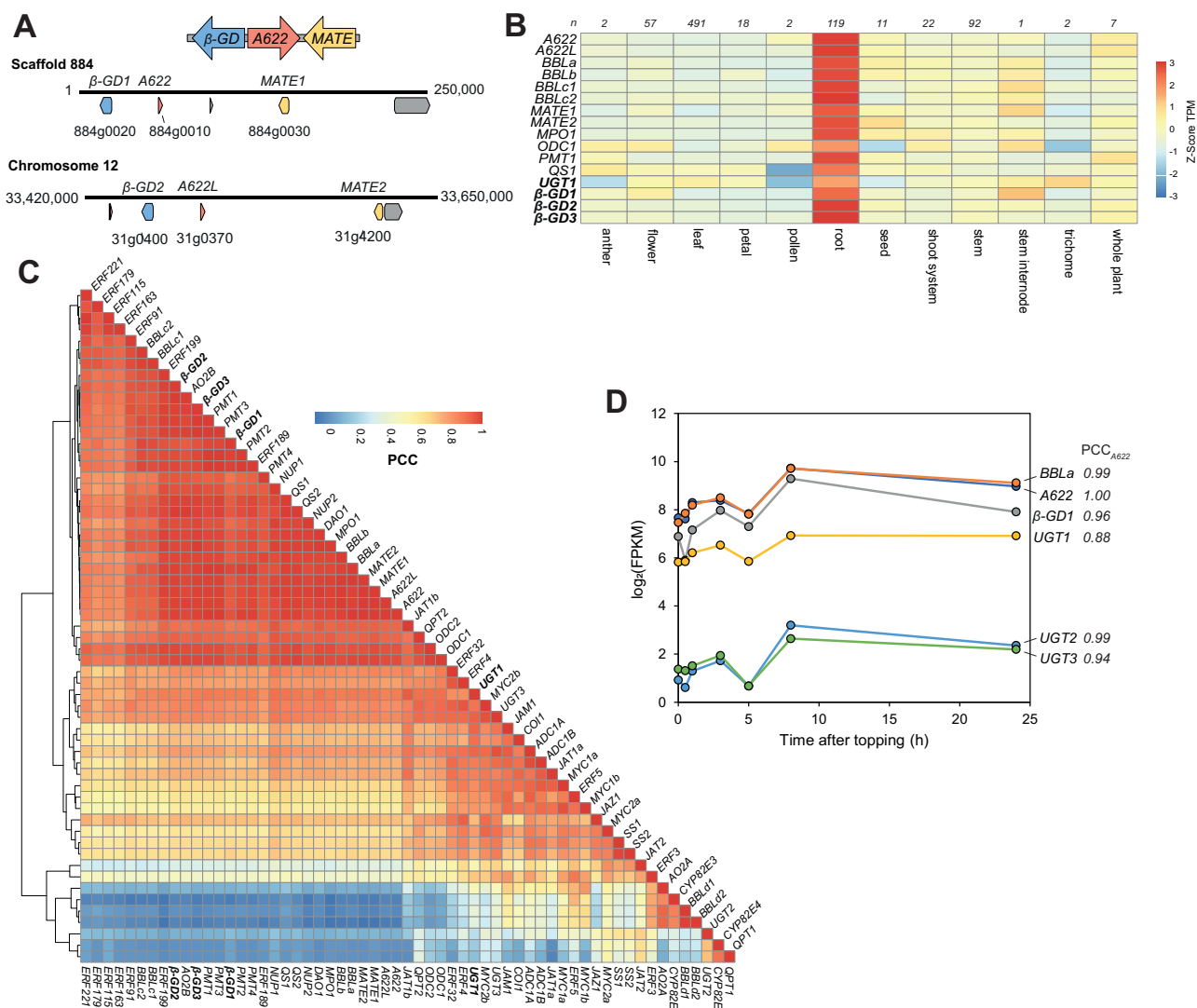


Fig. 2 | Identification of biosynthetic gene clusters and gene candidates.

A Homeologous biosynthetic gene clusters identified in the *N. tabacum* 2017 genome assembly³⁷ consisting of β -glucosidase (β -GD, blue), reductase (A622, red) and transporter (MATE, yellow) genes. Scaffold/chromosome number is described above the linear genome strand, with numbers at either end representing nucleotide position. Scaffold 884 is unplaced in this assembly but in the Sierró et al. 2024 assembly it is part of chromosome 16⁷⁵. The gene model between A622 and MATE1 in scaffold 884 (884g0040) encodes a protein that contains a domain with homology to a reverse transcriptase domain (PANTHER ID PTHR33116). **B** Heatmap

of expression of selected nicotine genes, including β -GD and UGT candidates across tissue types. Average expression calculated in each tissue type and values normalized per gene. Number of samples in each category above each column.

C Clustered heatmap showing all-by-all correlations (PCC) of nicotine biosynthesis and related genes across 824 samples. TPM values obtained from PEO⁴⁰. **D** Root expression levels of selected nicotine biosynthesis genes and UGT candidates following topping treatment, and their correlation (PCC) with A622 responses. Data obtained from Qin et al. 2020⁴¹. Details of genes available in Supplementary Data 1.

The UGT1 enzyme was determined to be a nicotinic acid *N*-glucosyltransferase (NaGT), producing nicotinic acid *N*-glucoside (**10**) from nicotinic acid (**2**) and UDP-glucose (Fig. 3 and Supplementary Fig. 4). NaGT could also catalyse the formation of pyridine glucoside (**12**) from pyridine and UDP-glucose, but could not form nicotinic glucoside (**13**) from nicotine and UDP-glucose (Supplementary Fig. 4).

A622 was determined to be a nicotinic acid *N*-glucoside reductase (NaGR), catalysing the NADPH-dependent reduction of **10** (Fig. 3 and Supplementary Fig. 5), with the inferred product 1,2-dihydropyridine glucoside (**4b**) undergoing non-enzymatic oxidation to yield pyridine glucoside (**12**). No peak corresponding to the expected mass of 1,2-dihydropyridine glucoside (**4b**) ($[M+H]^+ = m/z$ 244) could be observed in NaGR-catalysed reactions. The formation of **4b** from dihydronicotinic acid *N*-glucoside (**14**) could occur non-enzymatically through an enamine-imine tautomerisation followed by

decarboxylation (Supplementary Fig. 5C). The subsequent oxidation of **4b** to form **12** likely occurs through reaction with molecular oxygen.

When *N*-methylpyrrolinium (**3**) was included in reactions with NaGR and **10**, the levels of **12** decreased, and a diastereomeric mixture of (*R,S*)-nicotine glucoside ((*R,S*)-**13**) appeared (Figs. 3 and 4B). This is likely due to a non-stereoselective Mannich reaction between **3** and **4b**, forming (*R,S*)-dihydronicotinic glucoside (**5b**), which undergoes non-enzymatic oxidation to yield (*R,S*)-**13**. We were also able to identify a small peak matching the expected mass of dihydronicotinic glucoside (**5b**) ($[M+H]^+ = m/z$ 327) but this peak was assigned as its isomer dihydrometanicotinic *N*-glucoside (Glc-**11**) via comparison to a chemically synthesised standard (Supplementary Fig. 5). We also identified traces of the aglycone dihydrometanicotinic (**11**) (Supplementary Fig. 5). We propose that (*R,S*)-dihydronicotinic glucoside (**5b**) can either decompose into Glc-**11** or oxidise into (*R,S*)-**13**. The former can occur

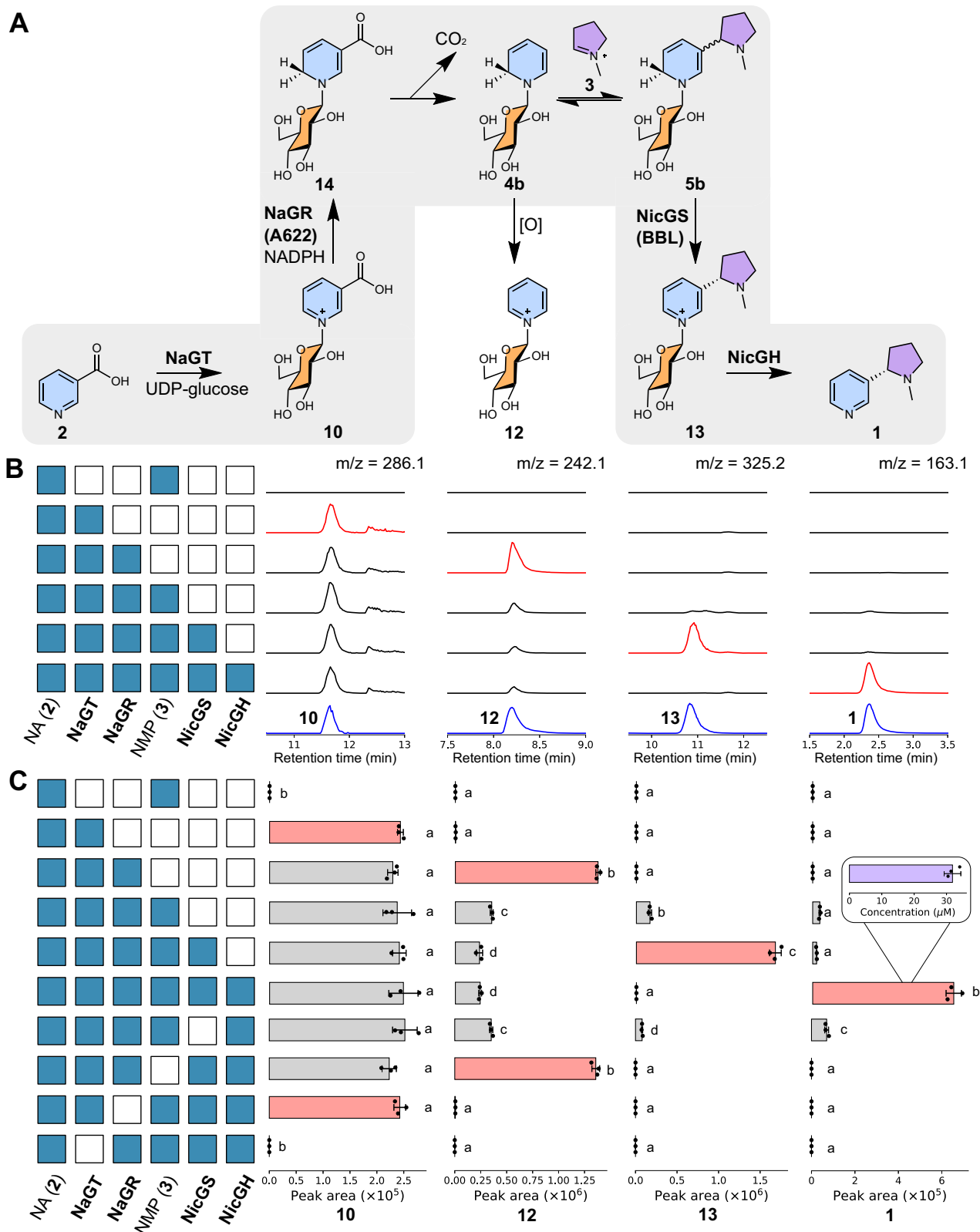


Fig. 3 | In vitro reconstitution of nicotine biosynthesis. A Proposed chemical transformations in nicotine biosynthesis reconstitution. **B** Products of in vitro cascades. Each row corresponds to an in vitro reaction, with the matrix showing presence/absence (blue/white) of reaction components, alongside UDP-Glc and NADPH. Each column shows extracted ion chromatograms (EICs, $m/z \pm 0.15$). Blue chromatograms are chemically verified standards, scaled for clarity. Red chromatograms highlight key products. MS² data is available in Supplementary Fig. 3. **C** Quantification of in vitro cascades. Each row of bars corresponds to an in vitro reaction, with the matrix showing presence/absence (blue/white) of reaction components. Each

column of bars shows the peak area of EICs ($m/z \pm 0.15$) corresponding to the masses of the chemical depicted above (left-to-right: **10**, **12**, **13**, **1**). Compound identity was validated by standard retention time, and MS² spectra (Supplementary Fig. 3). Red bars highlight key products. The letters above bars are significant groupings of peak area, determined per chemical ($p < 0.05$, Tukey post-hoc test, see Supplementary Data 5). The concentration of nicotine produced (from 1 mM of **2**) was determined via an external standard curve; error bars show standard error (inset). Bars show mean EIC peak area ($n = 3$); points are individual replicates; error bars show standard deviation. Source data are provided as a Source Data file.

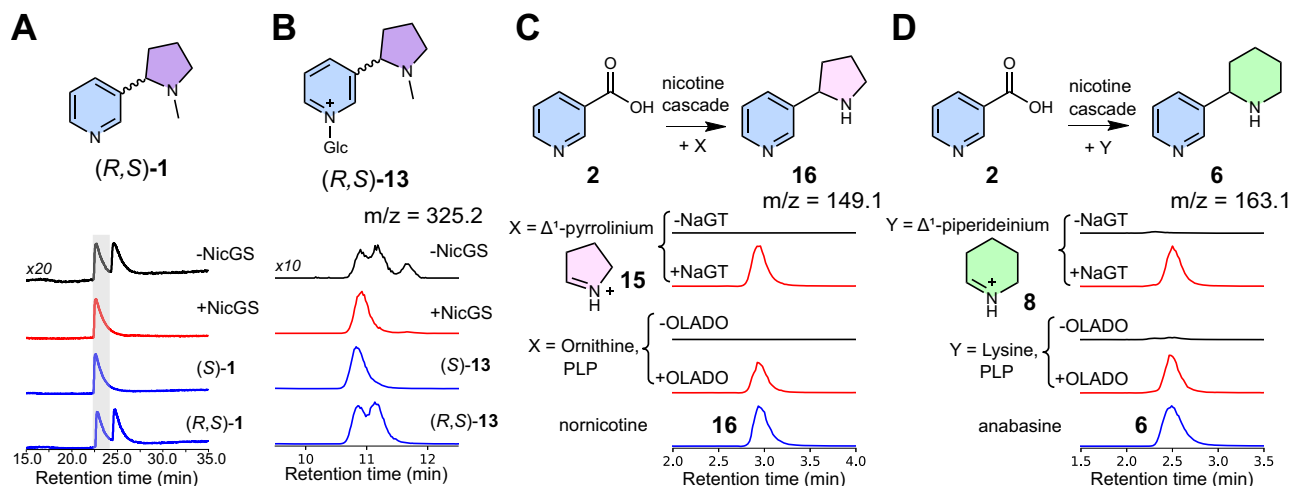


Fig. 4 | Stereoselectivity and substrate scope of nicotine biosynthetic cascade.

A Stereoselective formation of (S)-nicotine (**1**). Nicotine formation was achieved through one-pot enzymatic cascade (**2**, **3**, NaGT, NaGR, NicGH) with and without NicGS detected by chiral HPLC (260 nm). Signal intensity of peaks and standards (blue) was scaled for comparison. **B** Induction of stereoselectivity by NicGS. EIC (m/z 325.2 $\pm$ 0.15) showing (R, S)-nicotine glucoside (**13**) formation from **2** with

NaGT, NaGR and NMP, with or without NicGS. Chemically verified standards are blue. Signal intensity was scaled for comparison. **C** Formation of normicotine (**16**) and **D** anabasine (**6**) via the nicotine biosynthetic cascade (**2**, NaGT, NaGR, NicGH) with **15** or **8**, respectively, either synthesized or formed enzymatically by OLADO.

through a series of tautomerisation steps forming Glc-**11** and be followed by hydrolysis to yield **11** (Supplementary Fig. 5D). Notably, NaGR failed to catalyse the reduction of pyridine glucoside (**12**), confirming that this is a dead-end shunt metabolite (Supplementary Fig. 6A).

Inclusion of BBLA led to a significant decrease in **12**, increase in **13** and rendered the reaction diastereoselective, with the (S)-**13** isomer in excess (Figs. 3 and 4B), reflected in the exclusive production of the (S)-**1** enantiomer in the full pathway reconstruction (Fig. 4A). In the cascades, the peaks corresponding to **11** and Glc-**11** are absent in the presence of BBLA; yet Glc-**11** is not a substrate of the enzyme, as it cannot be converted into **13** by BBLA (Supplementary Fig. 6B). This suggests that BBLA's substrate is **5b**, which is upstream of Glc-**11**: conversion of **5b** to (S)-**13** prevents accumulation of Glc-**11** (Supplementary Fig. 5D). The presence of **11** in the absence of BBLA matches plant silencing experiments of the BBLs^{30,31}. These results redefine BBLA as (S)-nicotine glucoside synthase (NicGS).

The final step of the pathway is the deglucosylation of (S)-**13** to form (S)-**1**, which can be catalyzed by β -GDI, now redefined as nicotine glucoside hydrolase (NicGH) (Fig. 3). The enzyme NicGH can catalyze the deglucosylation of both diastereomers of **13** but cannot accept nicotinic acid *N*-glucoside **10**, dihydrometanicotine glucoside Glc-**11** or pyridine glucoside **12** as substrates (Supplementary Fig. 7).

In vitro biosynthesis of alternative tobacco alkaloids

We anticipated that we could exchange *N*-methylpyrrolinium (**3**) for different electrophiles to form related pyridine alkaloids. We demonstrated this with Δ^1 -pyrrolinium (**15**) and Δ^1 -piperidine (**8**), to form the natural products normicotine (**16**) and anabasine (**6**), respectively (Fig. 4C, D). We generated the electrophiles via two approaches, chemically and via in situ oxidative decarboxylation from ornithine and lysine respectively, catalysed by a recently discovered ornithine/lysine/arginine decarboxylase-oxidase (OLADO) from *N. tabacum*⁴².

In vitro assays with isotopically labelled nicotinic acid

Feeding experiments with isotopically labelled precursors have been crucial in establishing atomic-level information about nicotine biosynthesis, such as the loss of the C6 hydrogen from nicotinic acid (Fig. 1)^{17–19,31}. We reconstructed the four-enzyme nicotine biosynthesis cascade using isotopically labelled nicotinic acid-(ring- d_4) (**2-d₄**) as a substrate (Fig. 5 and Supplementary Fig. 8; Supplementary Data 6).

With the full cascade, we observed accumulation of the major product **1-d₃** (Fig. 5B, row i). A similar labelling pattern is observed for nicotine glucoside (**13**) in the absence of NicGH (Fig. 5B, row ii). In contrast, when NicGS (BBLA) was absent, then not only is the yield of **1** reduced, but the labelling pattern of **1** is changed, with **1-d₄** now the major isotopologue (Fig. 5B, row iii). This suggests that NicGS is responsible for the selective formation of **1-d₃** in the full cascade.

To pinpoint exactly which deuterium was being exchanged, we used the singly labelled nicotinic acid- d^6 (**2-d⁶**). With NicGS present, we saw formation of the major product **1-d₀**; without NicGS, **1-d₁** was the major product (Fig. 5C). This verifies that the isotope exchange determined by NicGS occurs at position C6 (Fig. 5C). This labelling pattern, where the C6 deuterium from labelled **2** is lost in the product **1**, is consistent with a cascade where NaGR (A622) catalyses reduction of **2**, transferring an unlabelled hydride (i.e. a protide) from NADPH onto C6, followed by the NicGS (BBLA) catalysing specific removal of the deuteride from the intermediate 1,2-dihydronicotine glucoside (**5b**) (Fig. 5A). The overall effect is a swapping of the deuteride for protide at C6, consistent with previous observations^{17–19,31}.

In the absence of NicGS, the ratio of isotopologues is not equal, but reversed, with the dominant isotopologue products being the fully labelled **1-d₁** or **1-d₄** from **2-d⁶** or **2-d₄** respectively (Fig. 5B, row iii, and Fig. 5C). This indicates that the non-enzymatic oxidation of **5b** favours the abstraction of the unlabelled protide (introduced by NaGR from NADPH) rather than the deuteride (from the labelled substrate **2**). This bias occurs due to the kinetic isotope effect, which renders carbon-deuterium bonds stronger than carbon-protium bonds, so the latter are broken more readily in the non-enzymatic reaction, where selectivity is determined by bond energy rather than spatial context.

When **2-d₄** is used a substrate, the NicGS (BBLA) also appears to have an effect on the labelling of pyridine glucoside **12**, which accumulates in the absence of *N*-methylpyrrolinium (**3**). Without NicGS (BBLA), the ratio of fully labelled **12-d₄** to partially labelled **12-d₃** is significantly greater than when NicGS (BBLA) is present (Fig. 5B, rows iv and v, Supplementary Data 6). This implies that NicGS can bind and catalyse the oxidation **4b** as well as **5b**. Cascade formation of normicotine- d_3 (**15-d₃**) and anabasine- d_3 (**6-d₃**) from **2-d₄** confirmed they are also formed via NicGS (BBLA)-catalysed oxidation (Fig. 5D, E).

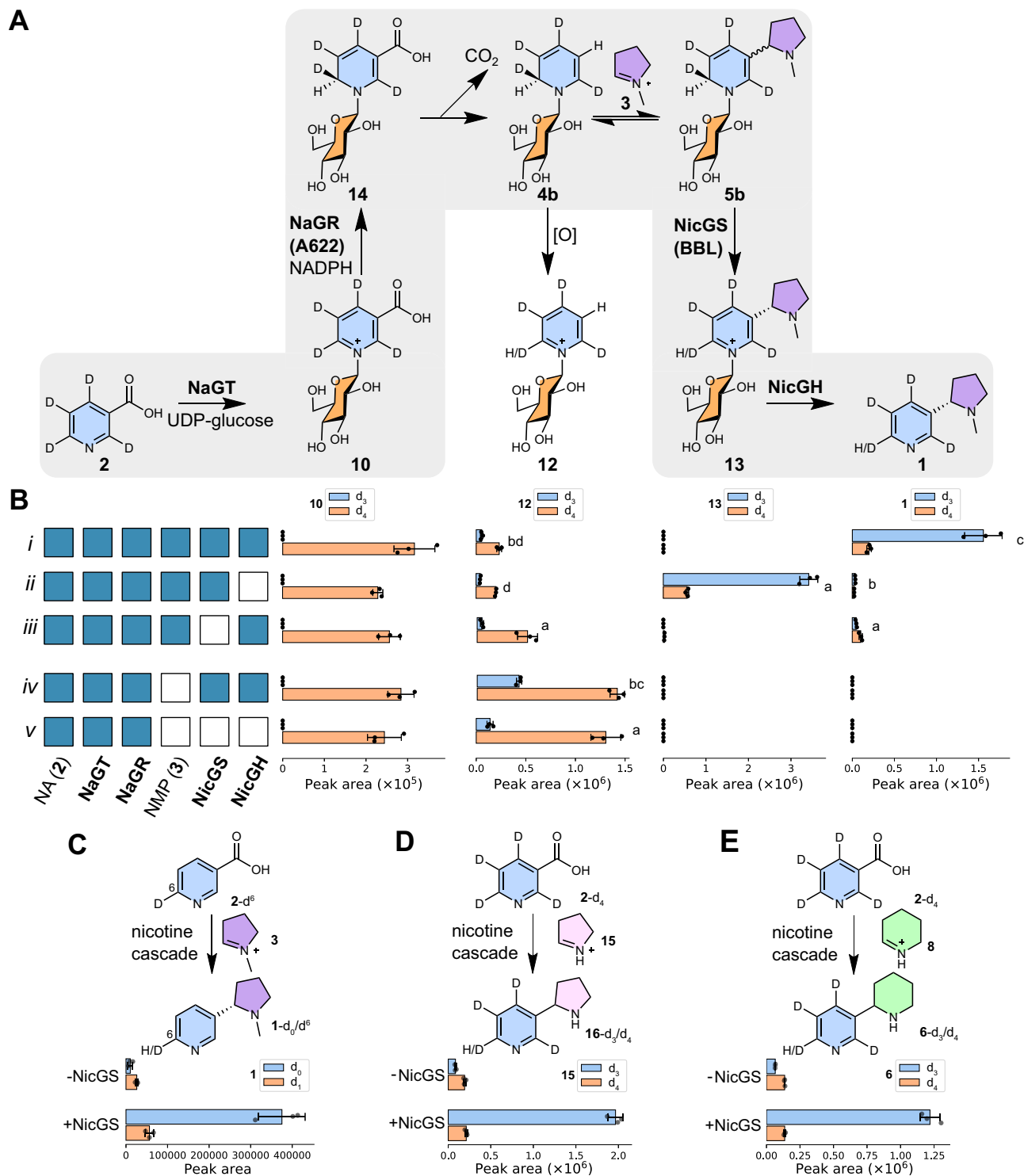


Fig. 5 | In vitro enzymatic cascade with labelled nicotinic acid. A Schematic of reaction cascade with nicotinic acid (*ring-d₄*) (**2-d₄**). **B** Selected in vitro cascades with **2-d₄** substrate. Bar charts show EIC peak areas of product (from L-to-R: **10**, **12**, **13**, **1**) isotopologues (all $m/z \pm 0.15$ except for d_3 -**12** which was $m/z \pm 0.5$ due to peak shape distortion). Reactions performed in triplicate, bars show mean peak area, points show individual measurements, and error bars show standard deviation. Each row corresponds to an in vitro reaction, with the matrix showing presence/absence (blue/white) of reaction components. Letters above bars are significant

groupings of d_3/d_4 isotopologue ratios, determined per chemical across all reaction combinations ($p < 0.05$, Tukey's HSD test, Supplementary Data 6, see Supplementary Fig. 8 for all reactions). **C–E** Conversion of nicotinic acid (**2**) isotopologues by the nicotine biosynthetic cascade (NaGT, NaGR, NicGH, NADPH and UDP-Glc) with and without NicGS. Bar charts show peak areas of products from EICs. Reactions were performed in triplicate (points); bars show mean peak area; error bars show standard deviation. Source data are provided as a Source Data file.

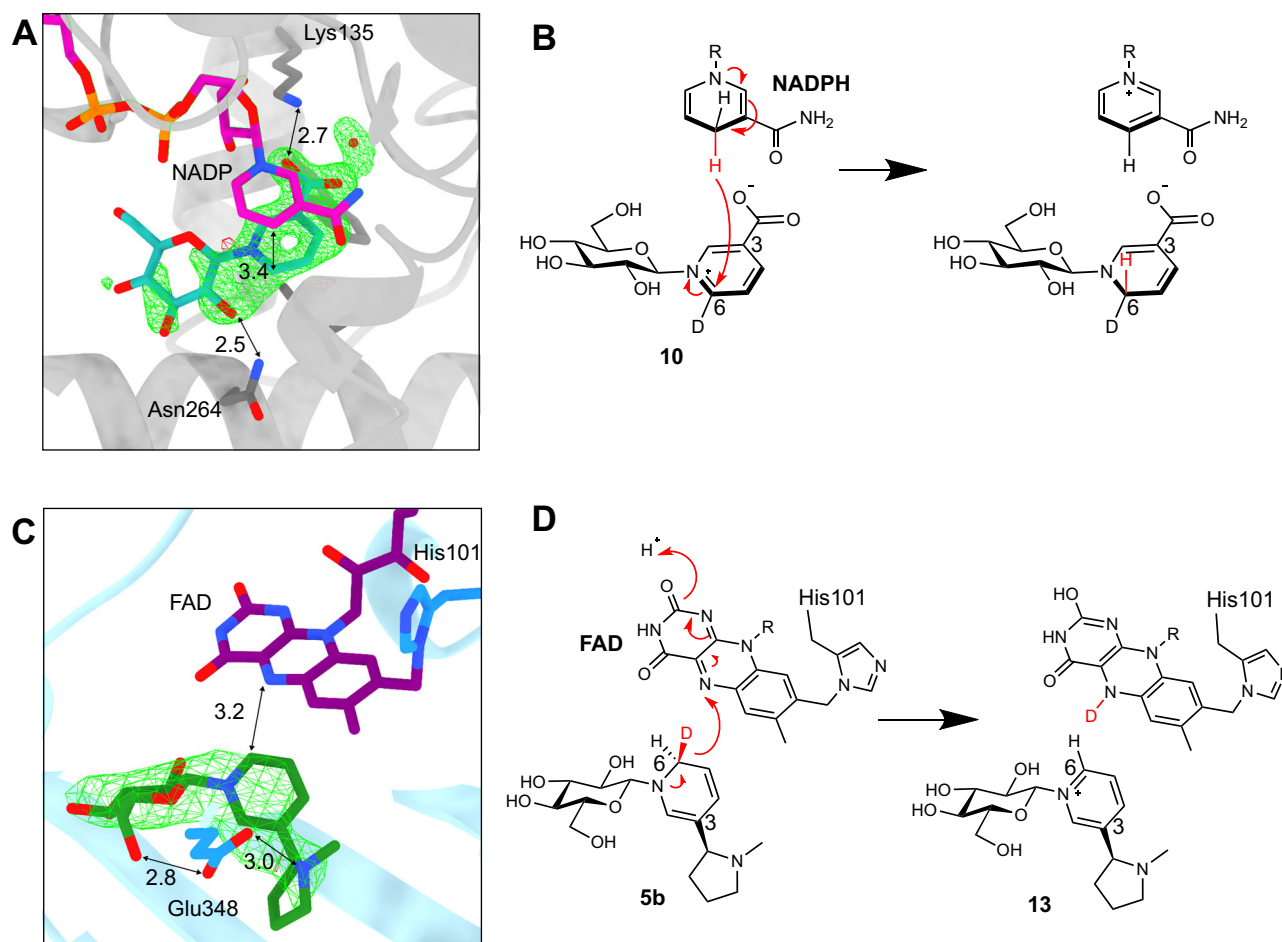


Fig. 6 | X-ray crystal structures of NaGR (A622) and NicGS (BBLA). **A** The crystal structure of NaGR bound to NADP⁺ and nicotinic acid *N*-glucoside (green net shows unrefined ligand density Fo-Fc at 3 σ). **B** Proposed mechanism of NaGR: NADPH reduces nicotinic acid *N*-glucoside-d⁶ on the *re*-face, adding a hydride in the pro-*R*

position. **C** The crystal structure of NicGS with covalently bound FAD and ligand (*S*)-nicotine glucoside (green net shows unrefined ligand density Fo-Fc at 3 σ). **D** Proposed mechanism of NicGS, abstracting the pro-*S* hydride from dihydronicotine glucoside-d⁶ to yield nicotine glucoside.

Structure of NaGR (A622)

To shed light on the mechanisms of NaGR (A622) and NicGS (BBLA), we determined their structures in complex with co-factors and relevant ligands using X-ray crystallography (Supplementary Data 7). In the case of NaGR (A622), we obtained a 1.3 Å resolution crystal structure in complex with NADP⁺ and nicotinic acid *N*-glucoside (**10**) (9RDD) (Supplementary Fig. 9). NaGR (A622) is most similar to dehydrogenases of the isoflavone reductase family (e.g. 2GAS, 57% sequence identity)⁴³, with the cofactor NADP⁺ bound between the Rossmann fold and the substrate binding domain. Omit density corresponding to the ligand nicotinic acid *N*-glucoside **10** was observed adjacent to the cofactor, with the C6 atom of the pyridine ring poised to receive hydride from C4 of the nicotinamide ring of the cofactor at a distance of 3.4 Å (Fig. 6A). This matches the regio- and stereoselectivity of the hydride transfer from NADPH onto the *re*-face of **10** at C-6, and into the pro-*R* position (Fig. 6B). The carboxylate is bound by the side chain of Lys135. The sugar density was not complete, but an interaction was observed between the side chain of Asn264 and the C2 hydroxyl of the sugar.

Structure of NicGS (BBLA)

We obtained 2.5 Å resolution structures of NicGS (BBLA) in complex with FAD (9RUW), and with both FAD and the putative reaction product (*S*)-nicotine glucoside (**13**) (9RDR) (Supplementary Data 7 and Fig. 6C; Supplementary Fig. 9). NicGS is structurally most similar to the

reticuline oxidase/berberine bridge enzyme from *Eschscholzia californica* (3D2D, 44.6% amino acid identity)⁴⁴, but the FAD is not bi-covalently bound as in the case of 3D2D, with C166 replaced by Gl65 in NicGS. Omit density was observed adjacent to the FAD density that was successfully modelled as (*S*)-**13** (Fig. 6C). The C-6 position of **13** is 3.2 Å from the reactive nitrogen of the FAD cofactor, with the *si*-face of the pyridine ring facing the FAD. This proximity indicates that the C6 pro-*S* hydride of the proposed dihydronicotine glucoside (**5b**) substrate reduces the FAD cofactor (Fig. 6D). The side chain of E348, not conserved in 3D2D, has a role in substrate recognition, interacting with the nitrogen of the pyrrolidine ring and also the exocyclic C6 hydroxyl of **13** at distances of 3.0 and 2.8 Å respectively. In the absence of ligand binding (9RUW) the loop containing this residue is unstructured (Supplementary Fig. 9). Overall, the structures of NaGR (A622) and NicGS (BBLA), with bound glucoside substrates or products lend strong support to our pathway proposal. Furthermore, the atomic details provide a structural/mechanistic validation of isotope labelling experiments, where NicGS (BBLA) abstracts the opposite hydride to that added by NaGR (A622).

In planta pathway reconstitution

To demonstrate that the pathway assembled *in vitro* could be operational *in planta*, we reconstituted nicotine biosynthesis in leaves of *N. benthamiana* via transient gene expression (Fig. 7 and Supplementary

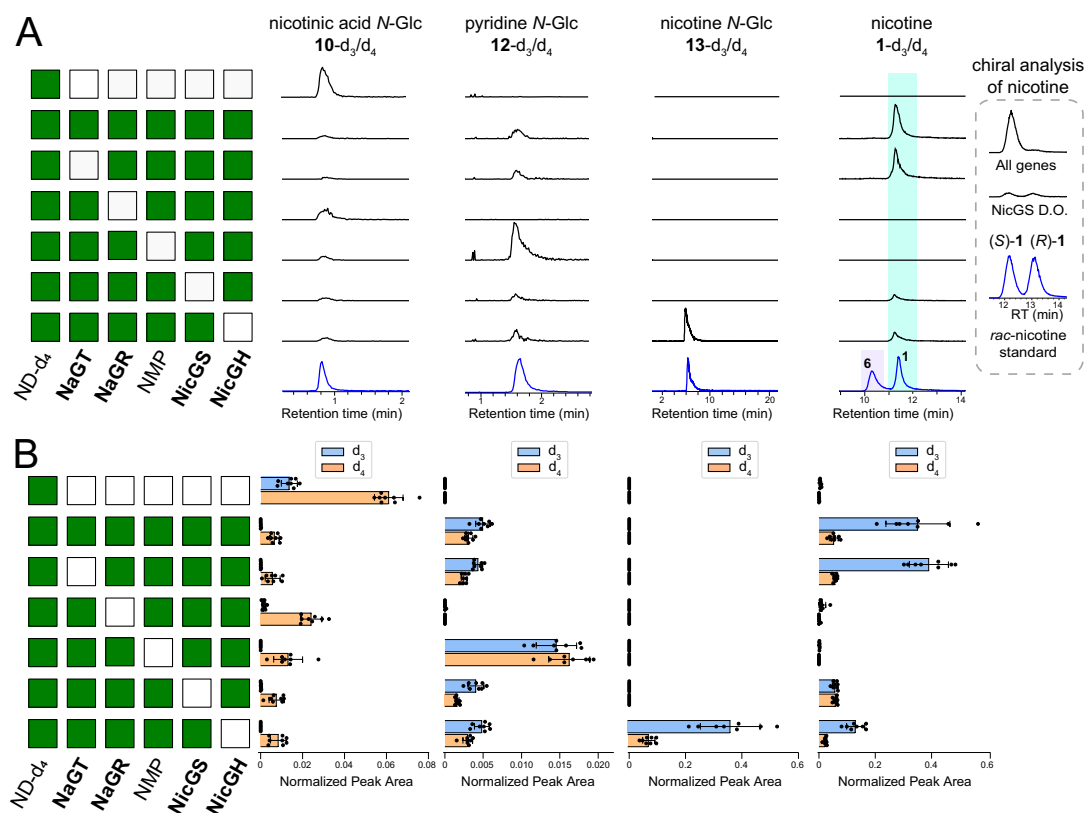


Fig. 7 | *In planta* reconstitution of nicotine biosynthesis. Leaves of *N. benthamiana* were agroinfiltrated with a combination of nicotine biosynthesis genes (day 0), fed with nicotinamide- d_4 (ND- d_4) at day 4, and harvested for metabolite analysis at day 6. The gene combination producing the *N*-methylpyrrolinium (ODC, PMT and MPO) is abbreviated as NMP. **A** Accumulation of the labelled compounds is indicated at the top (d_3 and d_4 versions together). Each row corresponds to the gene combination shown in the matrix on the left (green/white indicates presence/absence). Traces are representative extracted ion chromatograms (EICs, $m/z \pm 0.005$) corresponding to the labelled compounds (black) in comparison to chemically synthesized or commercial standards (blue) (see Supplementary Fig. 12

for MS² spectra). Signal intensities are normalized to the largest peak on each column; standards were rescaled to match the largest peak. The nicotine and anabasine standards were analysed as a mixed sample. The insert on the right shows the chiral analysis of nicotine in the presence of all genes and in the NicGS drop-out (D.O.). **B** Individual quantification of d_3 and d_4 isotopologues corresponding to the compounds and gene combinations shown in **A**. Bars show mean EIC peak area ($n=8$); points are individual replicates; error bars show standard deviation (see statistics in Supplementary Data 11). Source data are provided as a Source Data file.

Data 8)⁴⁵. Whilst *N. benthamiana* is capable of producing nicotine, we were able to independently study the reconstituted pathway as the key endogenous genes *A622* and *BBL* are not expressed in leaves (Supplementary Fig. 10 and Supplementary Data 9)⁴⁶. Furthermore, we could distinguish between the native nicotine transported into the leaves and the nicotine produced in situ by feeding the engineered leaves with a labelled precursor (nicotinamide- d_4) and analysing labelled nicotine.

To produce *N*-methylpyrrolinium (**3**), we co-expressed the known tobacco enzymes ODC, PMT and MPO (Supplementary Data 1 and 8)^{47–49}. When the entire pathway was expressed, we observed substantial production of (*S*)-nicotine- d_3 ((*S*)-**1**- d_3) (Fig. 7 and Supplementary Fig. 12), validating our in vitro observations. Additionally, stepwise reconstruction of the pathway mirrored results from the in vitro stepwise reconstruction, including production of anabasine- d_3 (**6**- d_3) enabled by co-expression of OLADO in place of the enzymes forming **3** (Supplementary Fig. 11 and Supplementary Data 10).

We then conducted a drop-out experiment to validate the role of the individual components (Fig. 7 and Supplementary Data 11). Drop-out of NicGH (β -GD1) led to a decrease in labelled nicotine (**1**) and accumulation of labelled nicotine glucoside (**13**), which we did not detect in any other combination. Dropping out NicGS (*BBLa/b*), resulted in a similar reduction in labelled nicotine but the residual nicotine was racemic (Fig. 7A). Omission of the *N*-

methylpyrrolinium forming enzymes eliminated nicotine production and led to the accumulation of labelled pyridine glucoside (**12**). Dropping out NaGR (*A622*) abolished production of labelled pyridine glucoside (**12**), blocking the pathway at nicotinic acid *N*-glucoside (**10**). It was not possible to confirm the *in planta* activity of NaGT (*UGT1*), as control leaves fed with labelled precursor accumulated labelled **10**, likely due to high leaf expression of the native *UGT1* homolog (Supplementary Fig. 10). Similar to the in vitro experiments, the *in planta* experiments were able to ascribe the loss of deuterium to NicGS (*BBLa/b*) (Fig. 7 and Supplementary Fig. 11; Supplementary Data 10 and 11).

Metabolites in wild-type and mutant plants

We further probed our pathway hypothesis by searching for proposed intermediates or shunt metabolites in wild-type roots of *N. tabacum* and *N. benthamiana*. We were able to detect nicotine glucoside (**13**) in *N. tabacum* and pyridine glucoside (**12**) in both *N. tabacum* and *N. benthamiana* (Fig. 8 and Supplementary Figs. 13; S14). We also examined roots of existing *N. benthamiana* knockout lines in NaGR (*A622*/*A622L*)²⁹ and NicGS (*BBLa/b/c/d/d'*)³¹. The NaGR (*A622*) knockout accumulated nicotinic acid *N*-glucoside (**10**) but not pyridine glucoside (**12**), supporting the identity of **12** as a shunt product of NaGR (*A622*) (Fig. 8A). The NicGS (*BBL*) knockout accumulated pyridine glucoside (**12**) and nicotine glucoside (**10**), similar to wildtype *N. tabacum* roots (Fig. 8A).

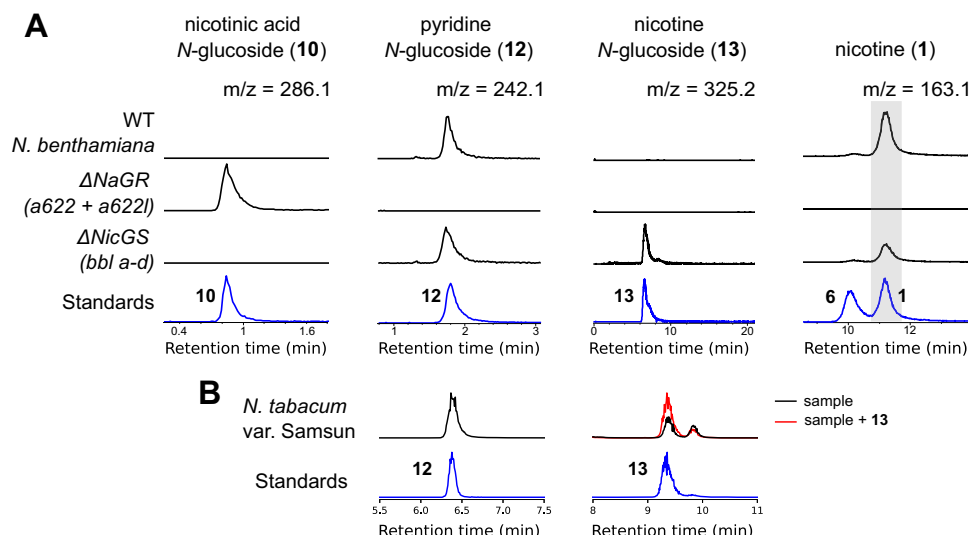


Fig. 8 | Identification of pathway intermediates and shunt products in *Nicotiana* roots. **A** Metabolite analysis of roots from *N. benthamiana* wild type, Δ NaGR²⁹ and Δ NicGS³¹ lines (black) in comparison to standards (blue) (see MS² verification in Supplementary Fig. 13). Traces are representative extracted ion chromatograms (EICs, $m/z \pm 0.005$) and were normalized to the largest peak in each column; standards were re-scaled to match the largest peak. **B** Metabolite analysis

of roots from *N. tabacum* var Samsun (black), compared to standards (blue) (see MS² verification in Supplementary Fig. 14). The red trace is root extract sample spiked with 10 μ M nicotine *N*-glucoside (13) to verify peak identity. Traces are representative EICs ($m/z \pm 0.005$); standards were re-scaled to match the corresponding peaks in the *N. tabacum* samples.

Discussion

It is remarkable that, despite the massive global use of tobacco, its economic importance, and the volume of investigation into nicotine biosynthesis, key steps in its formation have remained enigmatic. Here, we have reconstituted, *in vitro* and *in planta*, a minimal nicotine biosynthetic cascade, forming (*S*)-nicotine (**1**) from nicotinic acid (**2**) and *N*-methylpyrrolinium (**3**). It consists of four enzymes: NaGT, NaGR, NicGS, and NicGH.

Whilst we elected to test only a single representative paralog for each enzyme type in our *in vitro* assays, there is considerable evidence to suggest that the highly similar paralogs (or homeologs, i.e., paralogs derived from whole genome duplications) are functionally equivalent. NaGR (A622) has 95.5% protein sequence identity to its homeolog A622L and has similar expression patterns (Supplementary Data 2). Genome editing experiments show that both homeologs must be knocked out to guarantee ultra-low nicotine content in mutant tobacco lines^{28,50}. Similarly, BBLa (NicGS) and BBLb are homeologs with 92.8% protein sequence identity, similar expression patterns (Fig. 2C), and genome editing experiments implicate both genes in the formation of (*S*)-nicotine³⁴. The newly described NicGH (β -GDI) shares 91.8% with its homeolog β -GD2 and 83.6% protein sequence identity with its non-syntenic paralog β -GD3; all genes have similar expression patterns across tissues (Fig. 2C). During the revision of this manuscript, a report was published describing the systematic assessment of β -glucosidase genes in *N. tabacum*⁵¹. It revealed that the expression of NicGH (β -GDI) and its paralogs β -GD2 and β -GD3 (named NtBGLU51, 50 and 49, respectively in the β -GD manuscript) were all strongly induced by methyl jasmonate, supporting their role in defence response.

Cryptic biosynthetic steps cannot be predicted based on the structures of precursors or products, making them a bottleneck in pathway elucidation. In nicotine biosynthesis, NaGT and NicGH are responsible for a cryptic glucosylation, which acts like an activating group, increasing the electrophilicity of nicotinic acid (**2**) and making it susceptible to reduction by NADPH. Glucosylation steps are common in natural product biosynthesis, but the use of the glucoside in nicotine biosynthesis as an activating group is distinct from other known examples of glucosylation. Natural product glucose conjugates are usually seen in the context of detoxification, allowing the host plant to

store otherwise highly toxic molecules as inactivated glucosides, ready for future activation by deglucosylation⁵². For example, the benzoxazinoids DIBOA and DIMBOA are produced as defense compounds in grasses. These are sequestered in the plant vacuole as inactivated glucosides, which can be released by a specific glycosyl hydrolase for defense-related purposes⁵³. This kind of conditional detoxification/activation is also seen for cyanogenic glucosides e.g. in sorghum⁵⁴ and in the regulation of brassinosteroid and cytokinin hormone homeostasis⁵⁵.

In plant metabolism, glucose esters are commonly used as acyl donors. This involves a glucose moiety being added to a carboxylic acid, and the glucose ester product undergoes transesterification, a step catalysed by serine carboxypeptidase-like acyltransferases (SCPL-ATs)⁵⁶. This is analogous to coenzyme A thioesters, which are acyl donors that undergo transesterification catalysed by BAHD-acyltransferases⁵⁶. Glucose ester intermediates feature in the biosynthesis of sinapate (sinapoylcholine) and the UV-protectant sinapylmalate from sinapic acid via the activated sinapoylglucose ester intermediate in Brassicaceae⁵⁷, as well as in hormone homeostasis, forming IAA-myo-inositol conjugates via indoleacetic acid-glucoside⁵⁸. Interestingly, a glucoside ester is an intermediate in tropane alkaloid biosynthesis, a pathway from Solanaceae species that shares many features with nicotine biosynthesis. In particular, a UGT glucosylates phenyllactate to phenyllactylglucoside, enabling SCPL-AT catalysed transesterification to littorine with tropine⁵⁹. The glucosyl group in glucose esters acts to activate the carboxylic acid by forming an ester and becoming a favourable leaving group, with glucose being a product of the transesterification reaction. This is distinct from the proposed glucose activation in nicotine biosynthesis, where the glucose is not a direct product of the core reaction but instead modulates the activity of intermediates and requires a dedicated removal step: NicGH catalysed hydrolysis.

A cryptic glucosylation was recently described in Solanaceae steroidal glycoalkaloid biosynthesis, involving the cholesterol glucuronosyl transferase GAME15^{60,61}. A cryptic glucosylation was also discovered in the biosynthesis of the yeast antibiotic anisomycin⁶². In both of these examples, it appears that the glycosyl moiety acts mostly as a protecting group/recognition handle for the downstream

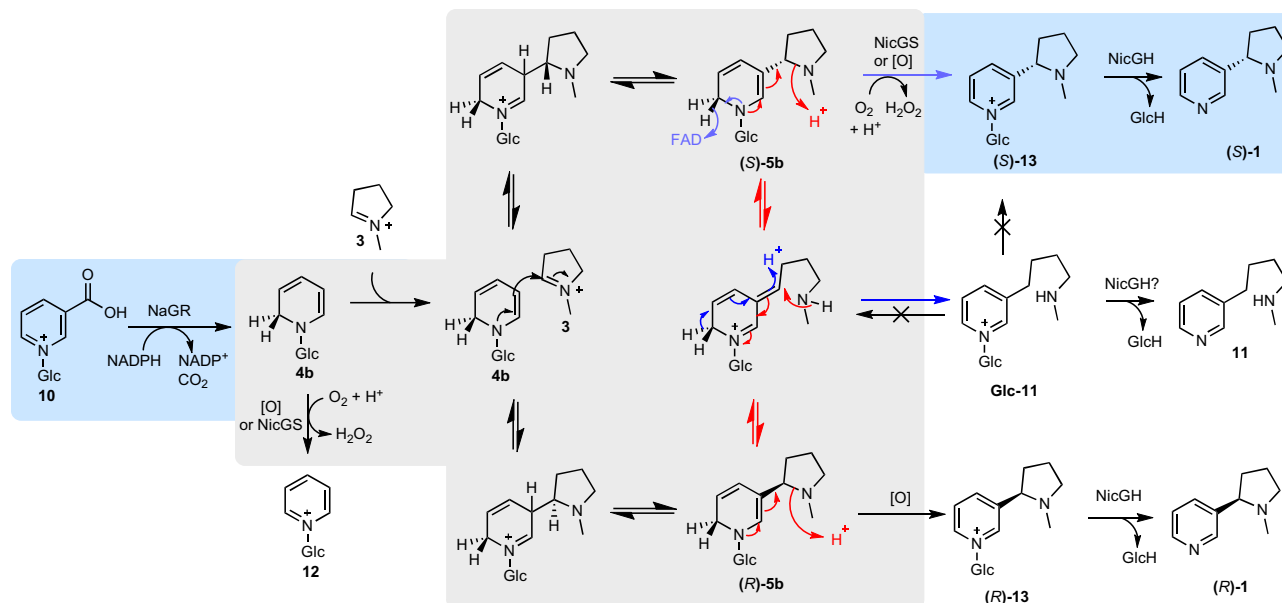


Fig. 9 | Putative activities of NicGS (BBL). Possible mechanistic routes towards (S)-1 from 10. The blue background shows key NaGR and NicGH reactions and major products. The grey background shows possible reactions that are either non-enzymatically or catalysed by NicGS. Only compounds with blue or white background have been directly detected. To induce the stereoselective outcome, NicGS may operate through: dynamic kinetic resolution (catalysing just purple arrows and

allowing other intermediates to non-enzymatically equilibrate), epimerisation of 5b via a ring opened intermediate (catalysing red and purple arrows), catalysis of the stereoselective Mannich reaction between 3 and 10 prior to oxidation, reversal of the non-desired Mannich reaction (opening (R)-5b into 3 and 10). Blue arrows show the predicted mechanism of Glc-11 formation.

biosynthetic enzymes, rather than having an activation effect as we propose for nicotine biosynthesis.

As well as modulating reactivity, the detoxification effect of glucoside-conjugates in plants can be attributed to driving vacuolar localisation of metabolites. This may also be true for nicotine biosynthesis, given the known subcellular localisation of NaGR (A622) to the cytosol³², and NicGS (BBL) to the vacuole³⁰, and also given that nicotine accumulates in vacuoles of tobacco root cells during biosynthesis³⁸. The predicted *N*-terminal signal peptide and *C*-terminal vacuolar sorting signal of NicGH strongly suggest it will also localise to the vacuole (Supplementary Fig. 1). The recent manuscript reporting tobacco β -glucosidases included an experiment where GFP was fused to the C-terminus of β -GD2 (NtBGLU50) and the fusion protein was transiently expressed in *N. benthamiana*. Surprisingly, the protein appeared to localise to the endoplasmic reticulum, despite its predicted vacuolar localisation sequence. Further investigation of the subcellular localisation of β -GDs is required for a conclusive outcome.

The division of the nicotine biosynthetic cascade across subcellular compartments—the cytosol and the vacuole—does not support the hypothesis that the cascade can operate as a structurally defined multi-enzyme complex. Furthermore, the reported ‘nicotine synthase’ enzyme isolated from tobacco roots by Friesen and Leete in 1990 was obtained with minimal purification: just filtration, desalting and crude ammonium sulfate protein fractionation, with activity observed in the 25–75% fraction¹⁶. Therefore, whilst the nicotine biosynthetic cascade enzymes (NaGT, NaGR, NicGS, NicGH) and their cofactors must have been present and soluble in this broad fraction, without further purification or alternative evidence, there is no reason to presume that the cascade acts as a ‘nicotine synthase’ enzyme complex.

The substrate for NaGR (A622) is nicotinic acid *N*-glucoside (10), which is consistent with its accumulation when A622 is silenced or knocked out²⁶. Previous attempts to characterise this reaction through spectrophotometry, typical for NADPH-dependent reactions, may have been hampered by overlapping absorbance spectra in the NADPH

substrate and the 1,2-dihydropyridine glucoside (4b) product. Furthermore, 4b appears unstable *in vitro*: here, its presence is inferred through the identification of pyridine glucoside (12) or through the reaction outcome in the presence of an electrophile. The cascade operates stereoselectively, with formation of (S)-enantiomer of 1 dominating both *in vitro* and *in planta* (Figs. 4A and 7A). NicGS (BBLa) is responsible for determining the stereoselectivity of the pathway, as we previously showed *in planta*³¹. The crystal structure of NicGS bound to (S)-nicotine glucoside ((S)-13) indicates it catalyses the diastereoselective oxidation of dihydronicotine glucoside (5b), but it is unclear whether it also catalyses a stereoselective Mannich reaction or a dynamic kinetic resolution process. The accumulation of the shunt product pyridine glucoside 12 is inhibited in the presence of 3, even in the absence of NicGS, indicating facile non-enzymatic coupling between 4b and 3. However, the decreased concentration of 12, higher conversion into 13 and induction of stereoselectivity in the presence of NicGS suggest enzymatic control of this step, as does the observation that NicGS can catalyse oxidation of 4b, which would match stepwise binding of 4b and 3. Further work is required to resolve the intricacies of the NicGS mechanism, given the complexity of the enzyme cascade (Fig. 9). The modularity of the cascade with respect to the electrophile has been demonstrated by the formation of anabasine and normicotine, and suggests the system could form the basis of a biocatalytic route towards diverse functionalised pyridines. New genetic understanding of nicotine biosynthesis also has the potential to inform engineering of *Nicotiana* sp., including the elimination or rerouting of alkaloid biosynthesis for new synthetic biology chassis. This work resolves a centuries-old puzzle in chemistry, plant biology and biochemistry, and will lead to the development of novel biotechnological and biocatalytic tools to produce valuable chiral chemicals.

In the period between completing final revisions to this manuscript and it being accepted, an article from the group of Dapeng Li was published describing their independent discovery of enzymes involved in nicotine biosynthesis⁶³.

Methods

Expression analysis

Genes of interest were defined by assessment of the literature related to nicotine biosynthesis (Supplementary Data 1). Protein sequences were obtained via GenBank or UniProt accession numbers and used as tblastn queries against the coding sequences from the 2017 *N. tabacum* genome³⁷ to obtain gene names and sequences. Biosynthetic gene clusters were identified using functional annotations and RNA-seq profiles of genes neighbouring genes of interest, using JBrowse hosted on the SolGenomics⁶⁴, querying the 2017 *N. tabacum* genome³⁷. Gene names were used as queries for the Plant Gene Expression Omnibus (PEO) (<https://peo.ku.dk/>)⁴⁰. The PEO *N. tabacum* dataset consists of TPMs for 824 samples over twelve tissue types. Precomputed Pearson Correlation Coefficient (PCC) values were used to identify top-correlating genes.

For gene co-expression analysis, TPMs of genes of interest were extracted from PEO and analysed in R-Studio (tidyverse and pheatmap packages). For plotting gene expression from *N. tabacum* across tissues, $\log_2(\text{TPM} + 1)$ values were grouped by Plant Ontology (i.e. tissue type), averaged, and plotted as a heatmap, with Z-score normalisation across tissue types. For expression correlation, an all-by-all PCC matrix was computed (824 samples, $\log_2(\text{TPM} + 1)$), depicted as a heatmap, with genes grouped with complete-linkage hierarchical clustering on Euclidean distance. *Nicotiana benthamiana* nicotine biosynthesis genes homologs were identified by top BLASTN hit using *N. tabacum* genes as queries of the *N. benthamiana* v1.0.1 genome⁶⁴. Gene names were used as queries for PEO and TPM values were extracted (476 samples), grouped by Plant Ontology and plotted as boxplots ($\log_2(\text{TPM} + 1)$)⁴⁰.

For assessment of gene expressions after topping, we extracted a list of differentially expressed genes (DEGs) in tobacco roots after topping, from Qin et al 2020⁴¹. Using the mean FPKM value at $t = 0$ we converted $\log_2$ fold change values to FPKM values across the time course and then calculated PCCs with A622 (Nitab4.5_0000884g0010) expression. Genes were filtered by annotation 'UDP-glucosyltransferase' and $\text{PCC}_{\text{A622}} > 0.85$. Top hits were assessed for their overall expression level and used as a query in PEO.

Protein purification

Codon-optimised NaGR (A622, UniProt Accession: IFRH_TOBAC), NaGT (UGT1, 2017 *N. tabacum* genome: Nitab4.5_0006222g0020³⁷) were expressed using the pET28a(+) vector as C-terminal (NaGR) or N-terminal (NaGT) His-tagged proteins, using *E. coli* B834 or SoluBL21, respectively. NaGR was expressed as a $\Delta 5$ N-terminal truncated form. Proteins were purified using nickel affinity chromatography followed by size-exclusion chromatography. NicGS (BBLA, 2017 *N. tabacum* genome: Nitab4.5_0006307g0010) and NicGH (β -GD1, 2017 *N. tabacum* genome: Nitab4.5_0000884g0020) were expressed as N-terminal His-tagged proteins using a *Komagataella phaffii* secretion system based on the pPICZ α vector. Proteins were purified from the media using tangential flow filtration, followed by nickel affinity chromatography and size-exclusion chromatography. OLADO (UniProt Accession: A0AIS3ZKS9_TOBAC) was expressed using the pET28a(+) vector as an N-terminal His-tagged protein using *E. coli* SoluBL21 and purified using nickel affinity chromatography followed by buffer exchange using a PD10 column⁴². Protein concentration was estimated using a Nanodrop spectrophotometer with predicted extinction coefficients derived from ExPASy ProtParam, and then stored at -70°C . Details of sequences are available in Supplementary Data 1 and 4), further methodological details are available in the Supplementary Information.

In vitro assays

In vitro enzyme assays and multi-enzyme cascades were carried out in 50 μL reactions, in 50 mM phosphate buffer, pH 7.4, 100 mM NaCl.

Reactions contained 1 mM nicotinic acid, 1 mM *N*-methylpyrrolinium, 0.5 mg mL⁻¹ NaGT (UGT), 0.5 mg mL⁻¹ NaGR (A622), 50 μg mL⁻¹ NicGS (BBLA) and 5 μg mL⁻¹ NicGH (β -GD). Other combinations were tested by removing component(s) of this mix to make different combinations of enzymes/substrates. The reactions were incubated for 16 hours at 37°C , then quenched (200 μL , 90:10 acetonitrile: water, 20 mM ammonium formate, pH 3), centrifuged ($21,000 \times g$, 10 min) and the supernatant was transferred to vials for LC-MS analysis. The *N*-methylpyrrolinium was prepared by incubation of 1 μL γ -methylaminobutyraldehyde diethyl acetal with 49 μL 1 M HCl at 70°C for 1 h before neutralisation with 40 μL 1 M NaOH, incubated on ice until adding to the assay on the same day. 1-Pyrroline for nicotine assays was prepared in the same way from aminobutyraldehyde diethyl acetal. Enzyme assays were analysed by LC-MS, with Hydrophilic Interaction Liquid Chromatography (HILIC) separation using a Waters XBridge BEH Amide column (5 μm , 2.1×100 mm) on a Thermo Scientific Vanquish UHPLC. Detection was performed on a Thermo Scientific LTQ XL Linear Ion Trap Mass Spectrometer. Details on LC-MS analytical procedures can be found in the Supplementary Information.

Chiral nicotine analysis of an in vitro reaction

In vitro enzyme assays for chiral analysis were carried out in 3×50 μL reactions for each condition, in 50 mM phosphate buffer, pH 7.4, 100 mM NaCl. Reactions contained 5 mM nicotinic acid, 5 mM *N*-methylpyrrolinium, 10 mM UDP-Glc, 2 U Glucose-6-phosphate dehydrogenase (Merck), 5 mM NADP and 10 mM glucose-6-phosphate, with 0.5 mg mL⁻¹ NaGT (UGT), 0.5 mg mL⁻¹ NaGR (A622), 50 μg mL⁻¹ NicGS (BBLA) and 5 μg mL⁻¹ NicGH (β -GD). Samples for chiral HPLC analysis were basified with 1 μL 1 M NaOH, then protein precipitated with 500 μL cold EtOH. After centrifugation ($21,000 \times g$, 10 min), the protein pellet was washed with a further 500 μL EtOH, the EtOH was then combined, evaporated to dryness and resuspended in 50 μL propan-2-ol and then loaded into vials for analysis. Assays with NicGS were diluted 10-fold further. Nicotine chiral analysis was carried out on an Agilent 1200 HPLC by normal phase isocratic elution using a CHIRALPAK[®] AD-H 250 $\times$ 4.6 mm column using 99:1 hexane: IPA at 1 mL min⁻¹. Nicotine peaks were detected by UV absorbance at 260 nm.

In vitro assay statistics

Statistical analysis of in vitro reaction product formation was performed using a one-way ANOVA (R-studio, function: aov) to determine EIC peak area differences across unique reactions, followed by a Tukey's HSD post-hoc test (function: TukeyHSD), with peak area grouping assigned letters (function: multcompLetters) based on significant differences ($p < 0.05$) in pairwise comparisons. Reactions were performed in triplicate ($n = 3$). For analysis of isotopologue ratios, d_3/d_4 isotopologue ratios were calculated per chemical and per sample through integration and division of EIC peak areas. Samples with zero peak area for either isotopologue were removed from the analysis. Then isotopologue ratios ($n = 3$) for each chemical were compared across unique in vitro reactions, analysed by analysis of variance and Tukey's HSD as described above. To calculate the nicotine concentration of the complete cascade, a standard curve of known concentrations was run in the same batch as the samples and fitted using log-log linear regression. Model parameters were used to convert sample peak areas into concentrations. To accurately estimate the mean and standard error, we first estimated the uncertainty of individual concentrations through the regression parameter errors and residual variability, before combining these using inverse-variance weighting of the triplicates to yield a pooled mean and standard error.

Protein crystallization

Initial screening of crystallization conditions was performed using commercially available INDEX (Hampton Research), PACT premier and CSSI/II (Molecular Dimensions) screens in 96-well sitting drop trays,

which were stored at 4 °C. Drops (300 nL) were set up with a 1:1 mixture of 12 mg/mL NaGR (A622) with precipitant mother liquor. Ligand complexes of A622 were obtained by co-crystallisation with 5 mM nicotinic acid *N*-glucoside **1**, derived from a 100 mM stock solution in H₂O, which was added to the NaGR (A622) solution and incubated on ice for 30 min before the trays were set up. Crystals were obtained in 0.1 M HEPES, pH 7.5 and 25 % (w/v) PEG 3350. NicGS (BBLa) protein was enzymatically deglycosylated before setting up crystallization screens. 53 µL 20 mg mL⁻¹ NicGS (BBLa) was incubated overnight at 37 °C with 7 µL EndoH and 7 µL GlycoBuffer 3 (New England Biolabs). After incubation, the deglycosylated BBL was used for crystallisation at 15 mg/mL. Initial screening of crystallization conditions was performed as for NaGR but at 16 °C. Initial *apo*-crystals of NicGS (BBLa) were obtained in drops containing 0.14 M LiSO₄, 0.1 M Bis-Tris pH 5.5, 21% PEG 3350, by co-crystallisation with 10 mM (*S*)-nicotine glucoside **13**, derived from a 100 mM stock solution in H₂O. Ligand-bound crystals were obtained by co-crystallisation as described, but with 50 mM (*S*)-nicotine glucoside, derived from a 1 M aqueous solution of (*S*)-nicotine glucoside. Crystals were harvested directly into liquid nitrogen with nylon CryoLoopsTM (Hampton Research), using the mother liquor without any further cryoprotectant (NaGR and NicGS *apo*) or by addition of 10% ethylene glycol (NicGS ligand-bound) to the drop prior to fishing.

X-ray structure elucidation

The datasets described in this report were collected at the Diamond Light Source, Didcot, Oxfordshire, U.K., on beamline I03. Data were processed and integrated using XDS⁶⁵ and scaled using SCALA⁶⁶ included in the Xia2⁶⁷ processing system. Data collection statistics are provided in Supplementary Data 7. The crystals of NaGR (A622) were obtained in space group *P*1, with one molecule in the asymmetric unit; the crystals of NicGS (BBLa) were obtained in space group *P*2₁2₁2₁, with two molecules in the asymmetric unit. The structures were solved by molecular replacement using MOLREP⁶⁸ with the AlphaFold structures AF-P52579-F1-v4 and AF-F1T160-F1-v4 used as the models for NaGR and NicGS, respectively (<https://alphafold.ebi.ac.uk/>)⁶⁹. The structures were built and refined using iterative cycles in Coot⁷⁰ and REFMAC⁷¹ employing local NCS restraints in the refinement cycles where relevant. The final structure of NaGR (A622) ligand complex exhibited % $R_{\text{cryst}}/R_{\text{free}}$ values of 17.1/21.0; the final structures of *apo*-NicGS (BBLa) and the NicGS (BBLa) ligand complex exhibited % $R_{\text{cryst}}/R_{\text{free}}$ values of 29.3/32.4 and 21.5/26.9, respectively. Refinement statistics for the structures are presented in Supplementary Data 7. The structures of NaGR (A622)-NADP+nicotinic acid *N*-glucoside, NicGS (BBLa)-FAD nicotine *N*-glucoside and NicGS (BBLa)-FAD have been deposited in the Protein Databank (PDB) with accession codes 9RDD, 9RDR and 9RUW, respectively.

In planta pathway reconstruction

The genes for transient overexpression were synthesized and cloned into a pHREAC vector (Addgene plasmid #134908)⁷² by Twist Bioscience (South San Francisco, USA). Coding sequences of the genes of interest were obtained from the *N. tabacum* v1.0 Edwards 2017 genome (scaffold)³⁷ and confirmed/refined by examining mapped RNAseq data in the genome browser on the website of the Sol genomics Network⁶⁴. A list of gene identifiers and coding sequences can be found in Supplementary Data 8. The cloned sequences of all pHREAC gene inserts were checked by Sanger sequencing and the plasmids were separately transformed into the electrocompetent *Agrobacterium tumefaciens* strain AGL1. Transient expression was performed in leaves of 4-week-old greenhouse-grown wild-type *N. benthamiana* plants according to Chuang and Franke 2022⁷³. pHREAC-GFP (kindly provided by Hadrien Peyret and George Lomonosoff) was used as a control and as a placeholder in the drop-out experiment. Four days post-infiltration, leaves were reinfiltrated with 400 µM nicotineamide-d₄

(D4 98%; Cambridge Isotope Laboratories, Tewksbury, MA, USA) dissolved in infiltration buffer without acetosyringone. Leaf disks were harvested after an additional 2 days (6 disks/plant: 3 disks/leaf x 2 leaves per plant). The harvested samples were flash-frozen in liquid nitrogen and were stored at -70 °C until tissue homogenization using chrome balls and a TissueLyzer (Qiagen). The homogenized samples were stored at -70 °C until metabolite extraction.

Metabolite analysis of in planta pathway reconstruction

Metabolites were extracted from homogenized flash-frozen plant tissue ($n = 8$) as described by Vollheyde and Dudley et al. 2023³¹. For the analysis, 75 ppm (386.2 µM) caffeine was used as the internal standard in the extraction solution and the proportion of tissue to extractant (µL) was 40 mg/100 µL. After the incubation step, samples were spun down for 5 min at 16200 × *g* (max speed table top centrifuge) and supernatants were diluted in ultrapure water at a proportion of 1:15. Samples were stored at -70 °C after extraction until filtering and after filtering until further analysis. The methanolic plant extracts were analysed via reversed-phase LC-MS according to Vollheyde and Dudley et al. 2023³¹ with an injection volume of 10 µL. Compounds were identified by comparison to commercial and synthesized standards. The chiral LC-MS analysis of methanolic plant extracts was performed according to Vollheyde and Dudley et al. 2023³¹ with an injection volume of 10 µL.

Statistical analysis was performed in R Studio. For each isotope (D0, D3, D4), a one-way ANOVA was conducted to test for differences between gene combinations (groups = 7, $n = 8$). For the dropout, the construct containing all genes was set as the reference. For the step-wise reconstruction, the construct containing GFP only was set as the reference. Post-hoc comparisons were carried out using two-tailed Dunnett's test (multcomp package) to compare each construct directly against the reference group.

Nicotiana benthamiana metabolite analysis

N. benthamiana WT and quintuple *bbla-d'* knock-out seeds (Δ NicGS) were kindly provided by Nicola Patron³¹. Seeds of *N. benthamiana* a622 knock-out line 3-3-1 (Δ NaGR) were kindly provided by Boaz Negin and Georg Jander²⁹. Prior to germination, seeds of each line were surface sterilized according to Florentine et al. 2016⁷⁴. Several seeds per line were incubated in 500 µL 1% sodium hypochlorite solution for 2 min and afterward the seeds were washed three times with 500 µL ultrapure water. The seeds were germinated in petri dishes on three layers of filter paper wetted with ultra-pure water. The plates were sealed with Leucopore and kept in a growth cabinet under long-day conditions (16 h light, 8 h dark) at day: night temperatures of 22 °C: 21 °C, 60% humidity and 200 µmol light intensity. The germinated seedlings were transferred to soil and the plants were grown in a growth cabinet under long day conditions (16 h light, 8 h dark) at day: night temperatures of 22 °C: 21 °C, 60% humidity and 200 µmol light intensity. Leaf and root samples were harvested from about 2-month-old plants. The root material was mostly collected from the soil-pot interface. To clean the roots, the soil was washed off with water, and excess water was removed by taping the roots on tissue paper. The harvested samples were flash-frozen in liquid nitrogen and were stored at -70 °C until tissue homogenization using chrome balls and a TissueLyzer (Qiagen). The homogenized samples were stored at -70 °C until further use. The samples were extracted and analysed by LC-MS as described above.

Nicotiana tabacum metabolite analysis

Nicotiana tabacum var. *Samsun* was grown in a temperature-controlled glasshouse with 21/18 (SD ± 3) °C Day/Night and 16 h photoperiod. Further details of growing conditions are in the Supplementary Information. Plants' apical and axillary growing tips were removed as soon as flower buds appeared (after about 1 month) and

these were regularly removed (monthly) until roots were harvested from 3-month-old plants and freeze-dried. 10 mg of freeze-dried root tissue was extracted into 1 mL MeOH, centrifuged (21,000 × g, 10 min) and loaded into LCMS vials for analysis. LC-MS analysis was carried out with Hydrophilic Interaction Liquid Chromatography (HILIC) separation on a Waters ACQUITY UPLC I-Class with detection performed on a Thermo Scientific Orbitrap Fusion™ Tribrid™ mass spectrometer. Details on LC-MS analytical procedures can be found in the Supplementary Information.

Synthetic methods

Generally, glucoside intermediates were synthesised by glucosylation of the pyridine nitrogen using acetobromo- α -D-glucose, followed by deacetylation with aqueous HBr. Detailed synthetic methods and compound characterisation are available in the Supplementary Information.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Crystal structure files have been deposited on the RCSB Protein Data Bank under the accessions: NaGR *holo* (nicotinic acid-*N* glucoside) [9RDD](#) NicGS *holo* (nicotine *N*-glucoside) [9RDR](#) NicGS *apo* [9RUW](#). All other data are available in the main text, the supplementary information and from the corresponding author(s) upon request. Source data are provided with this paper.

References

- Duke, D. et al. Earliest evidence for human use of tobacco in the Pleistocene Americas. *Nat. Hum. Behav.* **6**, 183–192 (2022).
- Wipfli, H. & Samet, J. M. One hundred years in the making: The global tobacco epidemic. *Annu. Rev. Public Health* **37**, 149–166 (2016).
- Langley, J. N. On the reaction of cells and of nerve-endings to certain poisons, chiefly as regards the reaction of striated muscle to nicotine and to curari. *J. Physiol.* **33**, 374–413 (1905).
- Baldwin, I. T. An ecologically motivated analysis of plant-herbivore interactions in native tobacco. *Plant Physiol.* **127**, 1449–1458 (2001).
- Millar, N. S. & Denholm, I. Nicotinic acetylcholine receptors: targets for commercially important insecticides. *Invert. Neurosci.* **7**, 53–66 (2007).
- Kessler, D. et al. Unpredictability of nectar nicotine promotes out-crossing by hummingbirds in *Nicotiana attenuata*. *Plant J.* **71**, 529–538 (2012).
- Shitan, N., Hayashida, M. & Yazaki, K. Translocation and accumulation of nicotine via distinct spatio-temporal regulation of nicotine transporters in *Nicotiana tabacum*. *Plant Signal Behav.* **10**, e1035852 (2015).
- Shoji, T., Kajikawa, M. & Hashimoto, T. Clustered transcription factor genes regulate nicotine biosynthesis in tobacco. *Plant Cell* **22**, 3390–3409 (2010).
- Elser, D. et al. Evolutionary metabolomics of specialized metabolism diversification in the genus *Nicotiana* highlights *N*-acetylornicotine innovations. *Sci. Adv.* **9**, eade8984 (2023).
- Benvenuto, E. et al. Plant molecular farming in the wake of the closure of Medicago Inc. *Nat. Biotechnol.* **41**, 893–894 (2023).
- Golubova, D., Tansley, C., Su, H. & Patron, N. J. Engineering *Nicotiana benthamiana* as a platform for natural product biosynthesis. *Curr. Opin. Plant Biol.* **81**, 102611 (2024).
- Lichman, B. R. The scaffold-forming steps of plant alkaloid biosynthesis. *Nat. Prod. Rep.* **38**, 103–129 (2021).
- Dawson, R. F. et al. Biosynthesis of the pyridine ring of nicotine. *J. Am. Chem. Soc.* **78**, 2645–2646 (1956).
- Leete, E. Biosynthesis of the *Nicotiana* alkaloids. XI. Investigation of tautomerism in *N*-methyl- δ -pyrrolinium chloride and its incorporation into nicotine. *J. Am. Chem. Soc.* **89**, 7081–7084 (1967).
- Dawson, R. F. Biosynthesis of the *Nicotiana* alkaloids. *Am. Sci.* **48**, 321–340 (1960).
- Brent Friesen, J. & Leete, E. Nicotine synthase—an enzyme from *Nicotiana* species which catalyzes the formation of (S)-nicotine from nicotinic acid and 1-methyl- δ -pyrrolinium chloride. *Tetrahedron Lett.* **31**, 6295–6298 (1990).
- Yang, K. S., Gholson, R. K. & Waller, G. R. Studies on nicotine biosynthesis. *J. Am. Chem. Soc.* **87**, 4184–4188 (1965).
- Scott, T. A. & Glynn, J. P. The incorporation of [2,3,7- 14 C] nicotinic acid into nicotine by *Nicotiana tabacum*. *Phytochemistry* **6**, 505–510 (1967).
- Leete, E. The incorporation of [5,6- 13 C₂]Nicotinic acid into the tobacco alkaloids examined by the use of 13 C nuclear magnetic resonance. *Bioorg. Chem.* **6**, 273–286 (1977).
- Dawson, R. F., Christman, D. R., D’Adamo, A., Solt, M. L. & Wolf, A. P. The biosynthesis of nicotine from isotopically labeled nicotinic acids. *J. Am. Chem. Soc.* **82**, 2628–2633 (1960).
- Leete, E. & Liu, Y.-Y. Metabolism of [2- 3 H]- and [6- 3 H]-nicotinic acid in intact *Nicotiana tabacum* plants. *Phytochemistry* **12**, 593–596 (1973).
- Leete, E. Stereochemistry of the reduction of nicotinic acid when it serves as a precursor of anatabine. *J. Chem. Soc. Chem. Commun.* 610 (1978) <https://doi.org/10.1039/c39780000610>.
- Cai, B., Jack, A. M., Lewis, R. S., Dewey, R. E. & Bush, L. P. (R)-nicotine biosynthesis, metabolism and translocation in tobacco as determined by nicotine demethylase mutants. *Phytochemistry* **95**, 188–196 (2013).
- Leete, E. Biosynthesis of anatabine and anabasine in *Nicotiana glutinosa*. *J. Chem. Soc. Chem. Commun.* 9–10. <https://doi.org/10.1039/C3975000009> (1975).
- Hibi, N., Higashiguchi, S., Hashimoto, T. & Yamada, Y. Gene expression in tobacco low-nicotine mutants. *Plant Cell* **6**, 723–735 (1994).
- Kajikawa, M., Hirai, N. & Hashimoto, T. A PIP-family protein is required for biosynthesis of tobacco alkaloids. *Plant Mol. Biol.* **69**, 287–298 (2009).
- Deboer, K. D., Lye, J. C., Aitken, C. D., Su, A. K.-K. & Hamill, J. D. The A622 gene in *Nicotiana glauca* (tree tobacco): evidence for a functional role in pyridine alkaloid synthesis. *Plant Mol. Biol.* **69**, 299–312 (2009).
- Burner, N., Kernodle, S. P., Steede, T. & Lewis, R. S. Editing of A622 genes results in ultra-low nicotine whole tobacco plants at the expense of dramatically reduced growth and development. *Mol. Breed.* **42**, 20 (2022).
- Negin, B., Wang, F., Fischer, H. D. & Jander, G. Acylsugars, nicotine and a protease inhibitor provide variable protection for *Nicotiana benthamiana* in a natural setting. *Plant Cell Environ.* **48**, 1073–1087 (2025).
- Kajikawa, M., Shoji, T., Kato, A. & Hashimoto, T. Vacuole-localized berberine bridge enzyme-like proteins are required for a late step of nicotine biosynthesis in tobacco. *Plant Physiol.* **155**, 2010–2022 (2011).
- Vollheyde, K. et al. An improved *Nicotiana benthamiana* bioproduction chassis provides novel insights into nicotine biosynthesis. *N. Phytol.* **240**, 302–317 (2023).
- Shoji, T. et al. Expression patterns of two tobacco isoflavone reductase-like genes and their possible roles in secondary metabolism in tobacco. *Plant Mol. Biol.* **50**, 427–440 (2002).
- Daniel, B. et al. The family of berberine bridge enzyme-like enzymes: A treasure-trove of oxidative reactions. *Arch. Biochem. Biophys.* **632**, 88–103 (2017).

34. Allen, Z. et al. BBL enzymes exhibit enantiospecific preferences in the biosynthesis of pyridine alkaloids in *Nicotiana tabacum* L. *Phytochemistry* **232**, 114363 (2025).
35. Posselt, W. & Reimann, L. Chemische Untersuchung des Tabaks und Darstellung eines eigenthümlich wirksamen Prinzips dieser Pflanze. *Mag. f.ür. die neuesten Erfahrungen, Entdeckungen und Berichtigungen im. Geb. der Pharm.* **6**, 138–161 (1828).
36. Smit, S. J. & Lichman, B. R. Plant biosynthetic gene clusters in the context of metabolic evolution. *Nat. Prod. Rep.* **39**, 1465–1482 (2022).
37. Edwards, K. D. et al. A reference genome for *Nicotiana tabacum* enables map-based cloning of homeologous loci implicated in nitrogen utilization efficiency. *BMC Genomics* **18**, 448 (2017).
38. Shoji, T. et al. Multidrug and toxic compound extrusion-type transporters implicated in vacuolar sequestration of nicotine in tobacco roots. *Plant Physiol.* **149**, 708–718 (2009).
39. Kajikawa, M. et al. Genomic insights into the evolution of the nicotine biosynthesis pathway in tobacco. *Plant Physiol.* **174**, 999–1011 (2017).
40. Koh, E., Goh, W., Julca, I., Villanueva, E. & Mutwil, M. PEO: Plant Expression Omnibus - a comparative transcriptomic database for 103 Archaeplastida. *Plant J.* **117**, 1592–1603 (2024).
41. Qin, Y. et al. Transcriptome analysis reveals key genes involved in the regulation of nicotine biosynthesis at early time points after topping in tobacco (*Nicotiana tabacum* L.). *BMC Plant Biol.* **20**, 30 (2020).
42. Wood, C. X. et al. Parallel evolution of plant alkaloid biosynthesis from bacterial-like decarboxylases. *N. Phytol.* **249**, 2954–2973 (2026).
43. Wang, X. et al. Crystal structure of isoflavone reductase from alfalfa (*Medicago sativa* L.). *J. Mol. Biol.* **358**, 1341–1352 (2006).
44. Winkler, A. et al. A concerted mechanism for berberine bridge enzyme. *Nat. Chem. Biol.* **4**, 739–741 (2008).
45. Carlson, E. D., Rajniak, J. & Sattely, E. S. Multiplicity of the *Agrobacterium* infection of *Nicotiana benthamiana* for Transient DNA Delivery. *ACS Synth. Biol.* **12**, 2329–2338 (2023).
46. Dawson, R. F. The localization of the nicotine synthetic mechanism in the tobacco plant. *Science* **94**, 396–397 (1941).
47. Xu, B., Sheehan, M. J. & Timko, M. P. Differential induction of ornithine decarboxylase (ODC) gene family members in transgenic tobacco (*Nicotiana tabacum* L. cv. Bright Yellow 2) cell suspensions by methyl-jasmonate treatment. *Plant Growth Regul.* **44**, 101–116 (2004).
48. Junker, A., Fischer, J., Sichart, Y., Brandt, W. & Dräger, B. Evolution of the key alkaloid enzyme putrescine N-methyltransferase from spermidine synthase. *Front Plant Sci.* **4**, 260 (2013).
49. Naconsie, M., Kato, K., Shoji, T. & Hashimoto, T. Molecular evolution of N-methylputrescine oxidase in tobacco. *Plant Cell Physiol.* **55**, 436–444 (2014).
50. Jeong, J.-H. et al. Impact of CRISPR/Cas9-induced mutations in nicotine biosynthesis core genes A622 and BBL on tobacco: reduction in nicotine content and developmental abnormalities. *Curr. Plant Biol.* **38**, 100343 (2024).
51. Liu, J. et al. Systematic genomic characterization of β -glucosidase genes in tobacco reveals NtBGLU50 as a multifunctional regulator of glycoside metabolism and defense-associated pathways. *Ind. Crops Prod.* **240**, 122643 (2026).
52. Osbourn, A. E. Preformed antimicrobial compounds and plant defense against fungal attack. *Plant Cell* **8**, 1821 (1996).
53. von Rad, U., Hüttel, R., Lottspeich, F., Gierl, A. & Frey, M. Two glucosyltransferases are involved in detoxification of benzoxazinoids in maize. *Plant J.* **28**, 633–642 (2001).
54. Cicek, M. & Esen, A. Structure and expression of a dhurrinase (beta-glucosidase) from sorghum. *Plant Physiol.* **116**, 1469–1478 (1998).
55. Hou, B., Ingkai, Lim Gillian, E., ng-K. iat, Higgins Dianna, S. & Bowles, J. N-Glucosylation of cytokinins by glycosyltransferases of *Arabidopsis thaliana*. *J. Biol. Chem.* **279**, 27822–27832 (2004).
56. Bontpart, T., Cheynier, V., Ageorges, A. & Terrier, N. BAH or SCPL acyltransferase? What a dilemma for acylation in the world of plant phenolic compounds. *N. Phytol.* **208**, 695–707 (2015).
57. Regenbrecht, J. & Strack, D. Distribution of 1-sinapoylgucose: choline sinapoyltransferase activity in the Brassicaceae. *Phytochemistry* **24**, 407–410 (1985).
58. Michalczyk, L. & Bandurski, R. S. UDP-glucose: indoleacetic acid glucosyl transferase and indoleacetyl-glucose: myo-inositol indoleacetyl transferase. *Biochem. Biophys. Res. Commun.* **93**, 588–592 (1980).
59. Qiu, F. et al. Functional genomics analysis reveals two novel genes required for littorine biosynthesis. *N. Phytol.* **225**, 1906–1914 (2020).
60. Jozwiak, A. et al. A cellulose synthase-like protein governs the biosynthesis of Solanum alkaloids. *Science* **386**, eadq5721 (2024).
61. Boccia, M. et al. A scaffold protein manages the biosynthesis of steroidal defense metabolites in plants. *Science* **386**, 1366–1372 (2024).
62. Zheng, X. et al. Biosynthesis of the pyrrolidine protein synthesis inhibitor anisomycin involves a novel gene ensemble and cryptic biosynthetic steps. *Proc. Natl. Acad. Sci. USA* **114**, 4135–4140 (2017).
63. Chang, L. et al. Complete biosynthesis of nicotine. *Cell*. <https://doi.org/10.1016/j.cell.2026.03.034> (2026).
64. Fernandez-Pozo, N. et al. The Sol Genomics Network (SGN)-from genotype to phenotype to breeding. *Nucleic Acids Res.* **43**, D1036–41 (2015).
65. Kabsch, W. XDS. *Acta Crystallogr D. Biol. Crystallogr.* **66**, 125–132 (2010).
66. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr D. Biol. Crystallogr.* **62**, 72–82 (2006).
67. Winter, G. Xia2: An expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43**, 186–190 (2010).
68. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr D. Biol. Crystallogr.* **66**, 22–25 (2010).
69. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
70. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D. Biol. Crystallogr.* **60**, 2126–2132 (2004).
71. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D. Biol. Crystallogr.* **53**, 240–255 (1997).
72. Peyret, H., Brown, J. K. M. & Lomonossoff, G. P. Improving plant transient expression through the rational design of synthetic 5' and 3' untranslated regions. *Plant Methods* **15**, 108 (2019).
73. Chuang, L. & Franke, J. Rapid combinatorial coexpression of biosynthetic genes by transient expression in the plant host *Nicotiana benthamiana*. *Methods Mol. Biol.* **2489**, 395–420 (2022).
74. Florentine, S. K. et al. Influence of selected environmental factors on seed germination and seedling survival of the arid zone invasive species tobacco bush (*Nicotiana glauca* R. Graham). *Rangel. J.* **38**, 417 (2016).
75. Sierro, N., Auberson, M., Dulize, R. & Ivanov, N. V. Chromosome-level genome assemblies of *Nicotiana tabacum*, *Nicotiana sylvestris*, and *Nicotiana tomentosiformis*. *Sci. Data* **11**, 135 (2024).

Acknowledgements

We thank: Sam Hart and Dr. Johan P. Turkenburg for assistance with X-ray data collection; Diamond Light Source Didcot UK for access to beamline I03 under grant number mx24948; Jared Cartwright, Tony Larson and the Biosciences Technology Facility; Ed Bergstrom, Karl Heaton, Mariela González, Jack Olsen and the Centre of Excellence in

Mass Spectrometry; Matthew Davy and the NMR facility; Caragh Whitehead, Tessa Keenan, Zhouqian Jiang, Inesh Amarnath and Lachlan Waddell for preliminary work and materials; Peter Bølge Sørensen, Jason Daff and the University of York Horticulture team. We thank Hadrien Peyret and George Lomonosoff for the pHREAC-GFP plasmid, Georg Jander and Boaz Negin for the gift of the *N. benthamiana* a622 line, and Nicola Patron for the *N. benthamiana* wildtype and quintuple *bbla-d'* knock-out line. B.R.L. acknowledges funding from UKRI [MR/S01862X/1 and MR/X010260/1]. B.T.W.S. acknowledges funding from BBSRC [BB/T007222/1] and F.G-F. acknowledges funding from Independent Research Fund Denmark [DFF 2035-00038B, 0136-00410B and 5281-00267B].

Author contributions

B.T.W.S. performed chemical synthesis; protein expression, purification and crystallography; and in vitro enzyme assays, pathway reconstruction and analysis. I.M.A. performed and analysed *Nicotiana benthamiana* transient expression experiments and ran the analysis of mutant plants. K.V. designed plant expression constructs and co-supervised I.M.A. Z.I. contributed to construct design, protein purification and method development. J. L. developed methods for chiral analysis. K.S.S. and C.D.S. synthesised deuterated substrates. M.A.F. supervised the design and execution of chemical synthesis. G.G. supervised protein preparation, crystallography and contributed to the X-ray structural analysis. F.G-F. supervised the *in planta* transient expression and metabolic analysis. B.R.L. identified gene candidates, supervised the biocatalytic reconstruction and coordinated the project. All authors contributed to data interpretation and writing the manuscript.

Competing interests

C.D.S. and K.S.S. are inventors on a pending patent related to the production of deuterated compounds. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-026-72705-0>.

Correspondence and requests for materials should be addressed to Benjamin R. Lichman.

Peer review information *Nature Communications* thanks Yang Qu and the other anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2026