ORIGINAL ARTICLE

Biosynthesis of *Veratrum californicum* specialty chemicals in *Camelina sativa* seed

Megan M. Augustin[1] · Ashutosh K. Shukla[1,2] · Courtney M. Starks[3] ·
Mark O'Neil-Johnson[3] · Linna Han[1,4] · Cynthia K. Holland[1,5] · Toni M. Kutchan[1]

Abstract Economically feasible systems for heterologous production of complex secondary metabolites originating from difficult to cultivate species are in demand since *Escherichia coli* and *Saccharomyces cerevisiae* are not always suitable for expression of plant and animal genes. An emerging oilseed crop, *Camelina sativa*, has recently been engineered to produce novel oil profiles, jet fuel precursors, and small molecules of industrial interest. To establish *C. sativa* as a system for the production of medicinally relevant compounds, we introduced four genes from *Veratrum californicum* involved in steroid alkaloid biosynthesis. Together, these four genes produce verazine, the hypothesized precursor to cyclopamine, a medicinally relevant steroid alkaloid whose analogs are currently being tested for cancer therapy in clinical trials. The future supply of this potential cancer treatment is uncertain as *V. californicum* is slow-growing and not amendable to cultivation. Moreover, the complex stereochemistry of cyclopamine results in low-yield syntheses. Herein, we successfully engineered *C. sativa* to synthesize verazine, as well as other *V. californicum* secondary metabolites, in seed. In addition, we have clarified the stereochemistry of verazine and related *V. californicum* metabolites.

Significance statement Biosynthesis of verazine in *Camelina sativa* seeds by heterologous expression of *Veratrum californicum* genes CYP90B27, CYP94N1, CYP90G1, GABAT1 and the *Arabidopsis thaliana* gene GAD2 was achieved. This is the first report of engineering *C. sativa* for the production of medicinally relevant compounds. In addition, the conformation of the synthesized verazine at C-20 was confirmed to be (*S*) and the stereochemistry of 22-keto-26-hydoxycholesterol at C-25 was determined to be most likely (*R*).

Abbreviations
GAD2   Glutamate decarboxylase 2 from *Arabidopsis thaliana*
GABA   γ-Aminobutyric acid

Electronic supplementary material The online version of this article (doi:10.1007/s11816-017-0427-x) contains supplementary material, which is available to authorized users.

✉ Toni M. Kutchan
 tmkutchan@danforthcenter.org

 Megan M. Augustin
 maugustin@danforthcenter.org

 Ashutosh K. Shukla
 ashupov@yahoo.com

 Courtney M. Starks
 cstarks@sequoiasciences.com

 Mark O'Neil-Johnson
 moj@sequoiasciences.com

 Linna Han
 helensdutcm@gmail.com

 Cynthia K. Holland
 cynthia.holland@wustl.edu

[1] Donald Danforth Plant Science Center, 975 North Warson Road, St. Louis, MO 63132, USA

[2] CSIR-Central Institute of Medicinal and Aromatic Plants, P.O. CIMAP, Lucknow, Uttar Pradesh 226015, India

[3] Sequoia Sciences, 1912 Innerbelt Business Center Dr., St. Louis, MO 63114, USA

[4] College of Pharmacy, Shandong University of Traditional Chinese Medicine, 4655 Daxue Road, Jinan 250355, People's Republic of China

GABAT1	γ-Aminobutyrate transaminase 1 from *Veratrum californicum*
CYP90B27	Cholesterol 22-hydroxylase from *Veratrum californicum*
CYP94N1	22-Hydroxycholesterol 26-hydroxylase/oxidase from *Veratrum californicum*
CYP90G1	22-Hydroxy-26-aminocholesterol 22-oxidase from *Veratrum californicum*
CO	Binary vector containing CYP90B27 and CYP94N1
CTOG	Binary vector containing CYP90B27, CYP94N1, GABAT1, and GAD2
CXTOG	Binary vector containing CYP90B27, CYP94N1, GABAT1, GAD2, and CYP90G1

Introduction

Industrial production of plant-derived natural products for medicinal use can be unachievable by either traditional chemical synthesis or agricultural techniques. Chemical reactions often result in toxic byproducts while the complex stereochemistry of many natural products reduces the yield of synthesis. In addition, direct extraction of medicinal compounds from source plant species is not always a viable alternative, as many produce low amounts and are not amendable to cultivation (Atanasov et al. 2015). The development of a low input heterologous eukaryotic system is attractive for the production of high value medicinal compounds. A potential production system is the emerging oilseed crop *Camelina sativa*.

C. sativa is an attractive alternative to current microbial heterologous production systems for plant natural products due to post-translational protein processing similar to that of most plant species as well as low input growth requirements. The close phylogenetic relatedness to the well-studied model organism *Arabidopsis thaliana*, in addition to ease of transformation and a sequenced genome, offers a foundation of genetic and molecular tools that can be utilized for metabolic engineering (Kagale et al. 2014; Lu and Kang 2008). Advantages to cultivation of *C. sativa* include low water- and nutrient input and its compatibility with current agricultural practices (Putnam et al. 1993). In addition, its short growing season (*ca*. 100 days) along with cold tolerance has shown its potential for growth and harvest prior to the normal growing season, therefore, maximizing land use without effecting normal crop production (Putnam et al. 1993).

C. sativa is under extensive investigation for its use in biofuels, industrial oil/lubricants, a replacement for fish oil in aquaculture, and high value metabolite production. Genetic manipulation by heterologous expression of fatty acid synthesis genes has altered fatty acid content in *C. sativa* to mimic hydrocarbons present in Jet A fuel and accumulate valuable liquid wax esters while RNAi suppression has been shown to alter the fatty acid profile, which increased attractive chemical properties including reduced freezing point and viscosity (Iven et al. 2015; Kim et al. 2015; Liu et al. 2015). In addition, accumulation of highly desired omega-3 long chain polyunsaturated fatty acids by heterologous gene expression has yielded ω3 LC-PUFA lines that have been shown to successfully replace fish oil in aquaculture (Betancor et al. 2015a, b; Petrie et al. 2014; Ruiz-Lopez et al. 2014). Moreover, production of the monoterpene limonene and the sesquiterpene cadinene in transgenic *C. sativa* could provide a robust and renewable source for these industrially useful terpenes in addition to their potential use in jet fuel (Augustin et al. 2015a). *C. sativa* was also engineered to produce the biodegradable polymer poly-3-hydroxybutyrate (PHB), a renewable component in bio-plastics with other industrial uses (Malik et al. 2015). Additionally, production of these metabolites in seeds lends to easy harvest and long term storage, desirable traits for large scale production. Based on the previous success of metabolic engineering of *C. sativa*, we have now engineered *C. sativa* to synthesize medicinally useful small molecules in seed.

The stereochemically complex steroidal alkaloid cyclopamine is a medicinal compound currently under clinical investigation whose supply is predicted to fall short of demands upon FDA approval (Heretsch et al. 2010). Cyclopamine is best known for its teratogenic effects in lambs born to pregnant ewes that ingested the source plant *Veratrum californicum* (Keeler and Binns 1968). Cyclopamine inhibits the hedgehog pathway, which is mainly active during embryonic development, by direct binding to the transmembrane receptor Smoothend (Chen et al. 2002). Cyclopamine and its semi-synthetic analog IPI-926 have shown promise in cancer therapy where mutations in the hedgehog pathway cause over activation leading to tumor growth, including pancreatic cancer, medulloblastoma, basal cell carcinoma, leukemia, colon cancer, and small cell lung cancer (Bahra et al. 2012; Batsaikhan et al. 2014; Berman et al. 2002; Jimeno et al. 2013; Lin et al. 2010; Olive et al. 2009; Taipale et al. 2000; Tremblay et al. 2009; Watkins et al. 2003). Cultivation of *V. californicum* for the production of cyclopamine has yet to be successful, and cell cultures produce only trace amounts of cyclopamine (Ma et al. 2006; Song et al. 2014). Four genes (CYP90B27, CYP94N1, CYP90G1, and GABAT1) have been discovered in the

5   Washington University, 1 Brookings Dr., St. Louis, MO 63130, USA

hypothesized biosynthetic pathway to cyclopamine (Augustin et al. 2015b). Together, the enzymatic products of these genes convert cholesterol to the predicted cyclopamine steroidal alkaloid precursor verazine (Fig. 1). By transforming

these four genes plus *A. thaliana* glutamate decarboxylase 2 (GAD2), an enzyme required to produce the co-substrate γ-aminobutyric acid (GABA) (Turano and Fang 1998), we have successfully engineered *C. sativa* to accumulate
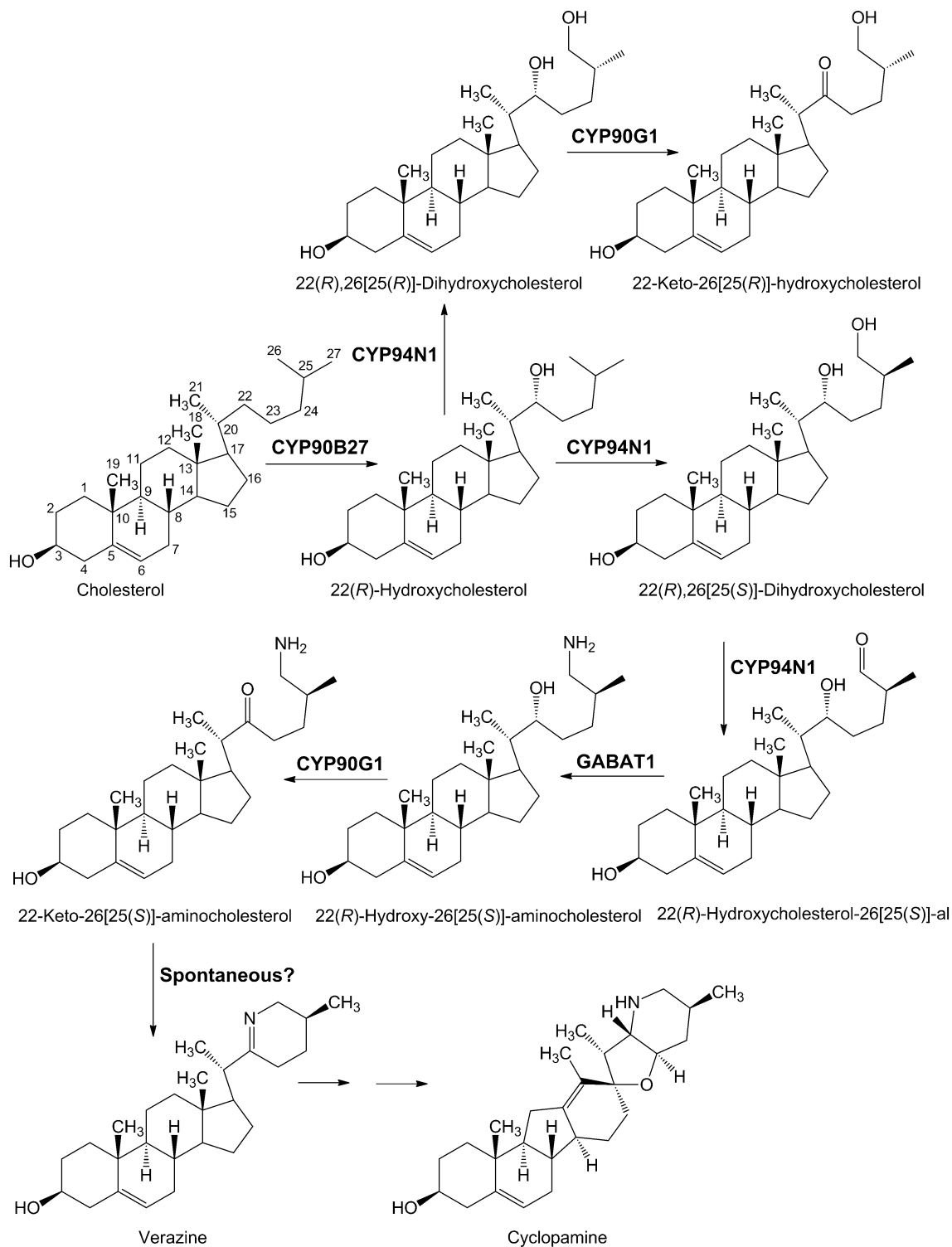

Fig. 1 Proposed biosynthetic pathway for cyclopamine

verazine in seed. In addition, the stereochemistry of the *V. californicum* metabolites produced herein are addressed.

Materials and methods

Cloning and plant transformation

CYP90B27, CYP94N1, CYP90G1, and γ-aminobutyrate transaminase 1 (GABAT1) were cloned from *Veratrum californicum* total root cDNA as previously described (Augustin et al. 2015b). GAD2 was cloned from *Arabidopsis thaliana* (L.) (Heyn) ecotype Columbia (Col-0) total leaf cDNA using the forward primer 5′-CACACATATGGTTTTGACAAAAACCGCAACGA-3′ and the reverse primer 5′-CACAGCGGCCGCTTAGCACACACCATTCATCTTCTT-3′, which incorporated the *Nde*I and *Not*I restriction sites, respectively. Arabidopsis tissue was obtained onsite (Donald Danforth Plant Science Center Greenhouse, St. Louis, MO, USA) and RNA was extracted using an RNeasy Plant Mini Kit (Qiagen). cDNA synthesis followed using MMLV-RT. Genes were first ligated into cassette vectors to provide each with a seed-specific promoter and terminator before transfer into the plant expression vector pRSe3, a binary vector with kanamycin and DsRed selection markers. pRSe3 is based on pRSe2 with an enhanced multiple cloning site (Augustin et al. 2015a). The following genes were ligated into each cassette vector providing the following seed specific promoters/terminators: CYP90B27 (Oleosin/Oleosin); CYP94N1 (Napin/Glycinin); GABAT1 (Napin/Glycinin); CYP90G1 (Napin/Glycinin); and GAD2 (Glycinin/Glycinin). The final constructs contained either two genes in *V. californicum* steroid alkaloid biosynthesis: CYP90B27 and CYP94N1 (designated CO), three genes in *V. californicum* steroid alkaloid biosynthesis: CYP90B27, CYP94N1, and GABAT1, plus GAD2 (designated CTOG), or four genes in *V. californicum* steroid alkaloid biosynthesis: CYP90B27, CYP94N1, GABAT1, and CYP90G1 plus GAD2 (CXTOG). Plants were transformed as previously described (Lu and Kang 2008). Transgenic seeds were screened as previously described (Augustin et al. 2015a).

DNA extraction and PCR for confirmation of genomic integration

DNA was extracted from five leaf samples of three selected CXTOG plant lines for the verification of construct integration using the DNeasy Plant Mini Kit (Qiagen). Lines were chosen based upon seeds exhibiting red fluorescence and alkaloid content. Polymerase Chain Reaction (PCR) was performed with primers listed in Table S1 using Taq DNA polymerase (NEB) and the following temperature program

parameters: 3 min 94 °C, 1 cycle; 30 s 94 °C, 30 s 55 °C, 1 min 72 °C, 35 cycles; final extension at 72 °C for 7 min.

Seed extraction for alkaloid analysis

Seeds of ten individual T2 (generation) lines of CO and CTOG (each) and 3 lines of T3 CXTOG (3 replicates for each line) were extracted based upon the protocol in the acyl-lipid metabolism chapter in The Arabidopsis Book (Li-Beisson et al. 2013). Hot isopropanol (1.5 ml, 75 °C) was added to 15–20 mg of seeds and incubated for 15 min. Seeds were then crushed with a glass rod followed by the addition of chloroform and H_2O (0.75 and 0.3 ml, respectively). Next, samples were sonicated in a sonication bath for 10 min followed by robust shaking for 1 h at room temperature (RT). Tubes were then vortexed for 10 s and centrifuged ($1500 \times g$, 2 min, RT). The liquid was moved to a fresh tube and the remaining tissue was re-extracted once with 2 ml of chloroform:methanol (2:1) and once with 1 ml of chloroform:methanol (2:1). Supernatants were combined prior to addition of 0.5 ml 1 M KCl. Samples were vortexed and centrifuged, and the aqueous upper phase was removed. This extraction was repeated once by the addition of 1 ml H_2O. The aqueous phase was removed and the extract was filtered with a 0.2 μm low protein binding hydrophilic LCR (PTFE) membrane (Millipore). 1/10 volume was removed and dried under N_2 for LC–MS/MS analysis while the remaining 9/10 were dried under N_2 and used for GC–MS.

GC–MS and LC–MS/MS analysis

Veratrum californicum metabolites in dried *C. sativa* seed extracts for GC–MS analysis were derivatized and analyzed as previously described (Augustin et al. 2015b). Dried *C. sativa* seed extracts for LC–MS/MS were first re-suspended in 50 μl of 80% methanol and diluted 1/10 prior to qualitative analysis by QTRAP 6500 (ABSciex) as previously described (Augustin et al. 2015b) except the TurboIonSpray ionization source temperature was set to 550 °C. For the quantitation of verazine in CXTOG lines, samples were prepared by diluting ½ vs. 1/10 to bring the verazine concentration into the linear range of the instrument. Quantitation was performed using pure verazine standard provided by Dr. David Kingston (Virgnia Tech, Blacksburg, VA, USA). The amount of verazine was too small to weigh, therefore, the concentration was determined using ELSD (Evaporative Light Scattering Detector) (Sedere, Sedex 75; France) with a hydrocortisone standard curve. For ELSD, samples were separated using a Hypersil Gold PFP (Thermo Electron Corporation), 250×4.6 mm 5 μ column and solvents identical to those used with the QTRAP 6500 using a gradient of 30–95% B over 50 min. LC–MS/MS

data were analyzed with Analyst 1.6.2. Selected CXTOG seed extracts were additionally analyzed by high-resolution mass spectrometry using a Q-Exactive (Thermo Fisher) coupled to an Agilent1200 microLC system. Q-Exactive samples were separated by a PLRP-S column (100×0.5 mm, 100 Å, 3 μ; Higgins Analytical) with a flow rate of 15 μl/min and the following solvent/gradient system: solvent A (0.05% formic acid/0.01% ammonium hydroxide v/v in H$_2$O); solvent B (0.05% formic acid/0.01% ammonium hydroxide v/v in 90% acetonitrile) where solvent B was held at 10% for 4 min, then 4–10 min 10–40% B, 10–18 min 40–45% B, 18–20 min 45–100% B, 20–21 min 100% B, 21–22 min 100−10% B and held at 10% B for an additional 10 min. Results were analyzed using Xcalibur 3.0.36 (Thermo Scientific) for data generated by Q-Exactive. Verazine, 22-keto-26[25(*R*)]-hydroxycholesterol, and 22(*R*)-hydroxy-26[25(*S*)]-aminocholesterol were heterologously produced in *Spodoptera frugiperda* (Sf9) insect cells and extracted as previously described (Augustin et al. 2015b). Frozen *V. californicum* root tissue was ground to a fine powder using a mortar and pestle under liquid nitrogen. The powder was weighed quickly and 70% ethanol was added in a 1:2 w/v ratio. The sample was vortexed at top speed for 5 min followed by centrifugation at 14,000×*g* for 10 min at RT. 200 μl of supernatant was then filtered through a 0.2 μm low protein binding hydrophilic LCR (PTFE) membrane prior to LC–MS/MS analysis.

Mosher esters for 22-keto-26[25(*R*)]-hydroxycholesterol stereochemical analysis

Purified 22-keto-26-hydroxycholesterol (50 μg) was dissolved in dichloromethane: pyridine (1:1, 200 μl) and divided in half. To each half was added 2 μl of (*R*)-MTPA-Cl or (*S*)-MTPA-Cl (90-fold excess, assuming two hydroxyls per molecule). The reactions were allowed to proceed overnight at room temperature, and were then quenched with water and extracted with ethyl acetate. Each organic layer was dried in a speed-vac, and then re-dissolved in methanol and purified by HPLC. HPLC was done with a Hypersil Gold PFP column (4.6×250 mm) eluted with a 30–95% gradient of acetonitrile in water. Both solvents contained 0.05% TFA. Fractions were collected from 0 to 64 min and dried down, and fractions corresponding to the largest peak visible by ELSD detection were pooled for NMR. Metabolite purification and NMR parameters were as previously described (Augustin et al. 2015b). Chemical shift assignments and integration of the ^{1}H spectrum indicated that for each compound, MTPA esters formed at both the 3 and 26 positions. Finamore et al. described using a small difference in the spread between the two H-26 signals in the *R* vs. *S* ester to assign the C-25 configuration of sterols oxygenated at C-26 (Finamore et al. 1991). In cases in
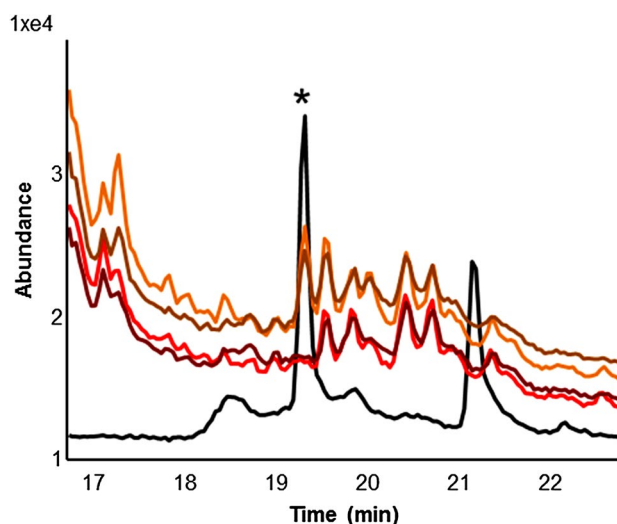

Fig. 2 GC–MS analysis of 22(*R*),26-dihydroxycholesterol found in transgenic *Camelina sativa* seeds expressing *Veratrum californicum* cytochrome P450 enzymes CYP90B27 and CYP94N1. Overlay of wild type *C. sativa* seed extract 1 (*light red*), wild type *C. sativa* seed extract 2 (*dark red*), transgenic *C. sativa* (CO) seed extract line 2 (*light orange*), transgenic *C. sativa* (CO) seed extract line 3 (*dark orange*), and enzyme assay using CYP94N1 and CPR expressed in *S. frugiperda* Sf9 insect cells utilizing 22(*R*)-hydroxycholesterol as substrate to produce 22(*R*),26-dihydroxycholesterol (*Black*). Samples were extracted and derivatized with Sylon HTP before GC–MS analysis detecting ion 99.1 with SIM scan. CPR refers to the cytochrome P450 reductase from *Eschscholzia californica*. The peak for 22(*R*),26-dihydroxycholesterol is indicated by an *asterisk*

which the H-26 signals are more separated in the (*S*) ester than in the (*R*) ester, they have assigned the 25 *S* configuration. In the MTPA esters of 22-keto-26-hydroxycholesterol, we observe greater separation between the H-26 signals in the (*R*) ester (4.15, 4.25) than in the (S) ester (4.15, 4.24), leading to a tentative assignment of 25*R* configuration. Because the difference in separation is so small, this assignment should be treated as tentative.

Results

Vector construction, plant transformation, and confirmation of construct integration

Three *C. sativa* expression vectors were assembled, each containing an increasing number of genes involved in *V. californicum* steroid alkaloid biosynthesis. Vector construction for plant transformation consisted of two parts. First, a gene was cloned into an initial vector providing it with a seed-specific promoter and terminator. Second, the expression cassette was amplified by PCR, digested by restriction enzymes, and then ligated into the multiple cloning
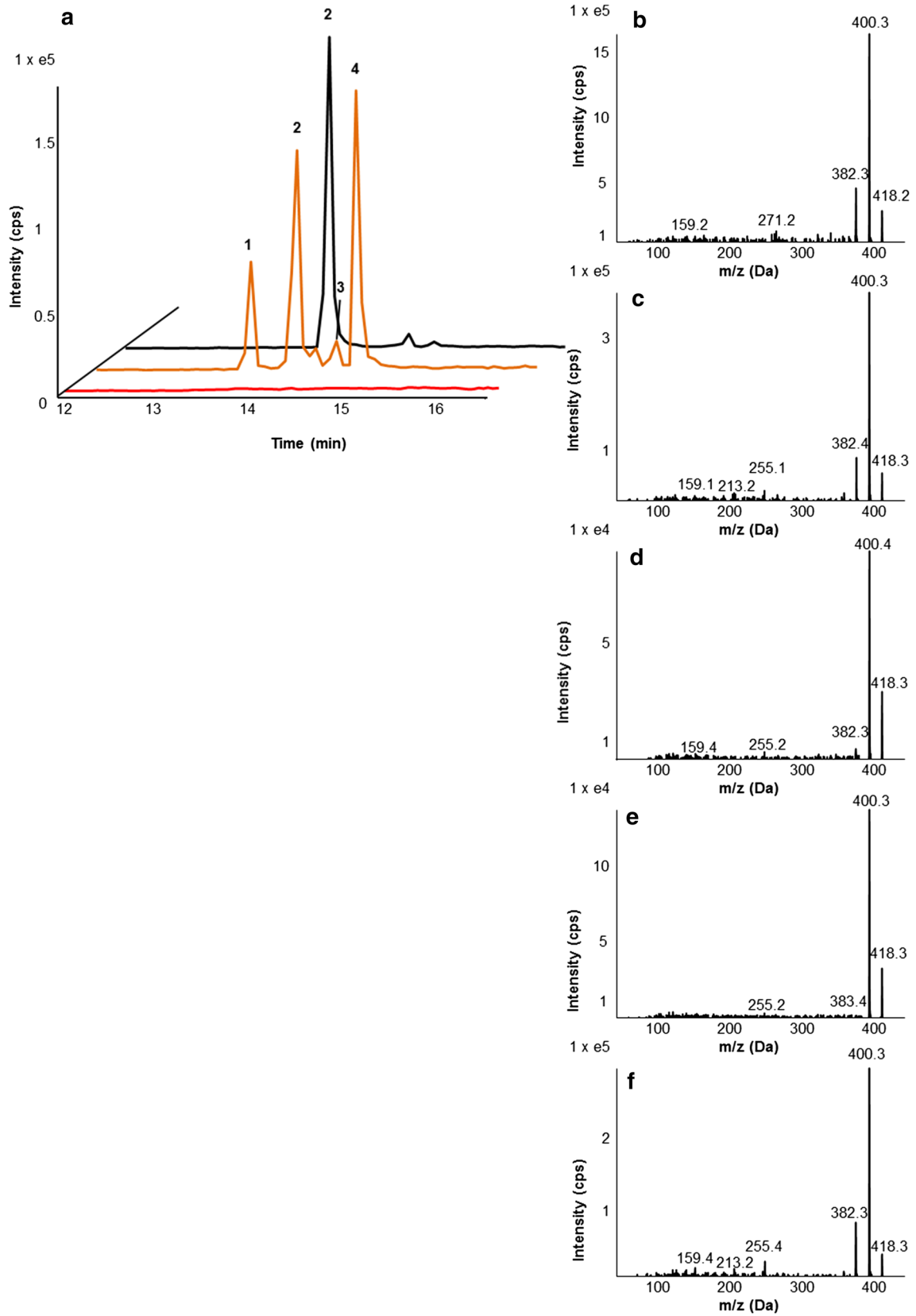
◀**Fig. 3** Chromatogram and mass spectra of 22(*R*)-hydroxy-26[25(*S*)]-aminocholesterol and similar compounds synthesized in *Camelina sativa* seed by expression of CYP90B27, CYP94N1, and GABAT1 from *Veratrum californicum* and GAD2 from *Arabidopsis thaliana*. Transgenic and wild type seeds were extracted and analyzed by LC–MS/MS with a QTRAP 6500 using targeted MRM scan for 418.3–400.0 *m/z* and EPI scan for 418.3 *m/z*. **a** Overlay of wild type *C. sativa* seed extracts (*red*), transgenic *C. sativa* expressing CYP90B27, CYP94N1, GABAT1, and GAD2 (CTOG) seed extracts (*orange*), and 22(*R*)-hydroxy-26[25(*S*)]-aminocholesterol produced in *S. frugiperda* Sf9 cells by expression of CYP90B27, CYP94N1, CYP90G1, GABAT1, and CPR (*Black*). Mass spectra of peak 1 (**b**), peak 2 in *C. sativa* seeds (**c**), peak 3 (**d**), peak 4 (**e**), and 22(*R*)-hydroxy-26[25(*S*)]-aminocholesterol (peak 2-produced in *S. frugiperda* Sf9 insect cells) (**f**). Selected fragment ions are shown. CPR refers to cytochrome P450 reductase from *E. californica* and GAD2 refers to glutamate decarboxylase 2

site (MCS) of the plant expression vector pRSe3 (Augustin et al. 2015a). Multiple expression cassettes were cloned into the MCS of a single plant transformation vector, allowing multiple genes to be transformed at once. Three vectors were constructed for transformation and analysis. One vector contained the first two genes in the biosynthetic pathway to verazine and was designated CO (**C** = **C**holesterol 22-hydroxylase; CYP90B27 and **O** = 22-Hydroxycholesterol 26-hydroxylase/**O**xidase; CYP94N1). The second vector contained the first three genes in the pathway to verazine and was designated CTOG (**T** = 22-Hydroxycholesterol-26-al **T**ransaminase; GABAT1 and **G** = **G**lutamate decarboxylase 2; GAD2). The third construct contained all four genes required for verazine biosynthesis and was designated CXTOG (**X** = 22-Hydroxy-26-aminocholesterol 22-o**X**idase; CYP90G1). GAD2 was included in both vector constructs containing GABAT1 to increase the concentration of GABA in seed, a co-substrate required by GABAT1. Preliminary work in which we transformed *C. sativa* with all four verazine biosynthetic genes but lacking GAD2 produced transgenic seeds where verazine was not detected. Production of the precursor cholesterol was not thought to be an issue as *C, sativa* is known to produce cholesterol (Mansour et al. 2014; Shukla et al. 2002). *C. sativa* gene expression data indicated that all homologs of GAD were either not expressed or had very low levels of expression in seed. We, therefore, chose to include GAD2 in our expression constructs. GAD2 was found to be highly expressed in mature plant tissues (expression data unpublished, used with permission of Noah Fahlgren, Donald Danforth Plant Science Center, St. Louis, MO, USA).

The pRSe3 vector is a modified, non-pathogenic, Ti (tumor inducing) plasmid from *Agrobacterium* that allows for genomic integration of the T-DNA without promoting bacterial infection. The vector also harbors a kanamycin resistance gene for selection in bacteria and the DsRed gene for selection in *C. sativa* seed. *C. sativa* plants were transformed as previously described (Lu and Kang 2008).

Initial confirmation of T-DNA integration into transformed *C. sativa* was achieved by visualization of DsRed by illuminating seeds with a green LED light and observing fluorescence through a red filter. Production of metabolites as described below confirmed integration and expression of the transgenes in T2 seeds. In addition, select CXTOG plants, based upon production of *V. californicum* metabolites, were propagated to T3 generation and subjected to PCR analysis to confirm integration of the construct (Figure S1).

GC–MS analysis of transgenic *C. sativa* seeds and detection of 22(*R*),26-dihydroxycholesterol in CO plants

Transgenic plants were first screened for biosynthesis of verazine precursors using GC–MS. GC–MS allows for detection of 22(*R*)-hydroxycholesterol, 22(*R*),26-dihydroxycholesterol, and 22-keto-26[25(*R*)]-hydroxycholesterol while 22(*R*)-hydroxy-26[25(*S*)]-aminocholesterol, 22(*R*)-hydroxycholesterol-26[25(*S*)]-al, and verazine are not detected by this method (Augustin et al. 2015b). Seeds from ten T2 lines of CO, ten T2 lines of CTOG, and three T3 lines of CXTOG (in triplicate) were extracted, derivatized, and analyzed by GC–MS for *V. californicum* metabolites. All ten lines of CO plants showed a peak consistent with authentic 22(*R*),26-dihydroxycholesterol, while the two wild type plants lacked a corresponding peak. Figure 2 shows two representative CO samples. 22(*R*)-hydroxycholesterol was not detected. Quantitation of 22(*R*),26-dihydroxycholesterol was not possible due to a lack of available standard. No *V. californicum* metabolites were detected from CTOG and CXTOG plants using GC–MS.

LC–MS/MS analysis for detection of *V. californicum* metabolites in transgenic *C. sativa* seeds

Metabolite screening using LC–MS/MS was performed on the same wild type and transgenic extracts described in "GC–MS analysis of transgenic *C. sativa* seeds and detection of 22(*R*),26-dihydroxycholesterol in CO plants". Extracts were analyzed for verazine, 22(*R*)-hydroxy-26[25(*S*)]-aminocholesterol, and 22-keto-26[25(*R*)]-hydroxycholesterol (Augustin et al. 2015b). No accumulation of these metabolites was detected in CO plants. In transgenic CTOG plants, 22(*R*)-hydroxy-26[25(*S*)]-aminocholesterol was detected along with three unknown compounds with the same mass and similar fragmentation patterns (Fig. 3). 22-Keto-26[25(*R*)]-hydroxycholesterol was not detected, but a small amount of verazine was detected (Fig. 4, Figure S2). In CXTOG plants, 22(*R*)-hydroxy-26[25(*S*)]-aminocholesterol was not detected, but 22-keto-26[25(*R*)]-hydroxycholesterol
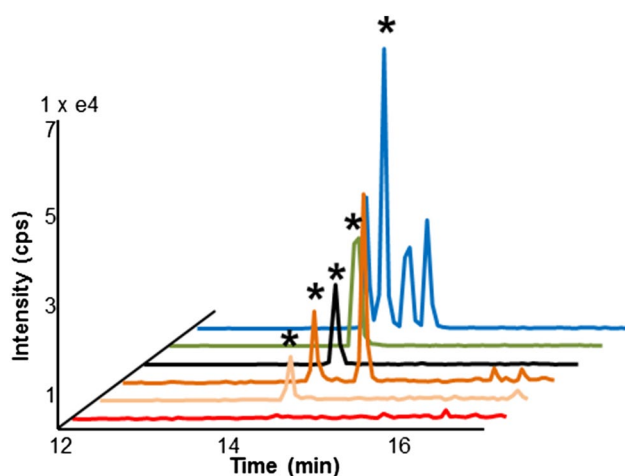
Fig. 4 Targeted LC–MS/MS analysis of verazine for structure validation and confirmation of presence in transgenic *Camelina sativa* seeds expressing select *Veratrum californicum* genes in conjunction with GAD2 from *Arabidopsis thaliana*. Extracted samples were analyzed by QTRAP 6500 using MRM scan for 398.3–159.2 *m/z*. Overlay of wild type *C. sativa* seeds (*red*), *C. sativa* seeds expressing CYP90B27, CYP94N1, GABAT1, and GAD2 (CTOG) (*light orange*), C. *sativa* seeds expressing CYP90B27, CYP94N1, CYP90G1, GABAT1, and GAD2 (CXTOG) (*dark orange*), *S. frugiperda* Sf9 insect cells expressing CYP90B27, CYP94N1, CYP90G1, CPR, and GABAT1 (*black*), structurally validated verazine (*green*), and *V. californicum* root extract diluted 1/10,000 (*blue*). CPR refers to cytochrome P450 reductase from *E. californica*, and GAD2 refers to glutamate decarboxylase 2. Verazine peaks are indicated by an *asterisk*, all other peaks are unknown

was found to accumulate (Fig. 5). An additional peak with the same mass and similar fragmentation pattern as 22-keto-26[25(*R*)]-hydroxycholesterol was also detected with a 1 min delay in retention time (Fig. 5). Similarly, verazine was detected alongside an additional peak having 1 min delay in retention time (Fig. 4, Figure S2). This unidentified compound also appeared to have the same mass and similar fragmentation pattern (Figs. 4, 6). The unstable metabolite 22-(*R*)-hydroxycholesterol-26[25(*S*)]-al was not detected in any sample. *V. californicum* root extracts were also analyzed for verazine and four distinct peaks were detected, one corresponding well with the heterologously produced verazine, but none with the unknown peak in CXTOG plants (Fig. 4).

Quantitation of verazine was performed in triplicate, on the same three CXTOG lines as above, using authentic verazine standard (Abdel-Kader et al. 1998). The first line analyzed revealed a verazine concentration of 41 ± 7 pg verazine/mg seed. The second line accumulated 67 ± 13 pg verazine/mg seed while the third produced 54 ± 6 pg verazine/mg seed.

Q-Exactive analysis for verazine confirmation

To verify the accumulation of verazine in CXTOG transgenic *C. sativa* seeds, extracts were analyzed by high-resolution mass spectrometry (Fig. 6). A peak with mass 398.3418 was detected in the CXTOG seed extracts, consistent with the calculated mass of 398.3417 ([M + H]$^+$) for verazine. Compellingly, a defining fragment peak of verazine, with mass 126.1279 (calculated mass 126.1277, [M + H]$^+$) was also detected in the extract. All mass measurement errors were well within the 5 ppm limit for confirming molecular formula.

Stereochemistry of 22-keto-26[25(*R*)]-hydroxycholesterol and verazine

It is possible that the compound produced by the recombinant enzymes is a stereoisomer of verazine, such as 20-*epi*-verazine; a naturally occurring compound possessing a 20(*R*) configuration. To help verify the stereochemical configuration of our product, we obtained a sample of purified verazine for which Dr. David Kingston's research laboratory (Abdel-Kader et al. 1998) had confirmed the configuration by NMR. The verazine from the Kingston laboratory, which has the *S*-configuration at C-20, had the same retention time (Fig. 4) and mass spectrum (Figure S2) as the verazine produced in Sf9 insect cells and transgenic *C. sativa*. Due to the instability of verazine, we could not isolate sufficient quantities for NMR, but these results, in conjunction with the known stereochemical configuration of the precursor molecule cholesterol, we conclude that the produced molecule is most likely verazine and not 20-*epi*-verazine.

The accumulation of 22-keto-26[25(*R*)]-hydroxycholesterol and its apparent inability to be converted into verazine (Augustin et al. 2015b) led us to question its configuration at the C-25 position. The configuration of verazine at C-25 is (*S*), and, therefore, the precursors 22(*R*), 26-dihydroxycholesterol and 22(*R*)-hydroxy-26[25(*S*)]-aminocholesterol are most likely also (*S*) at the corresponding position. To determine the configuration at C-25 of 22-keto-26[25(*R*)]-hydroxycholesterol, the (+)-(*R*)-MTPA and (-)-(*S*)-MTPA esters were formed and analyzed by NMR. The analysis suggested that the compound was most likely 25(*R*) as the (*R*) ester displayed a larger separation in H-26 signals (4.15, 4.25) than in the (*S*) ester (4.15, 4.24). Due to the small difference in H-26 signals, this assignment must be regarded as tentative; however, this stereochemical configuration is supported by its apparent lack of oxidation at C-26 by CYP94N1 and amination by GABAT1 (Augustin et al. 2015b).
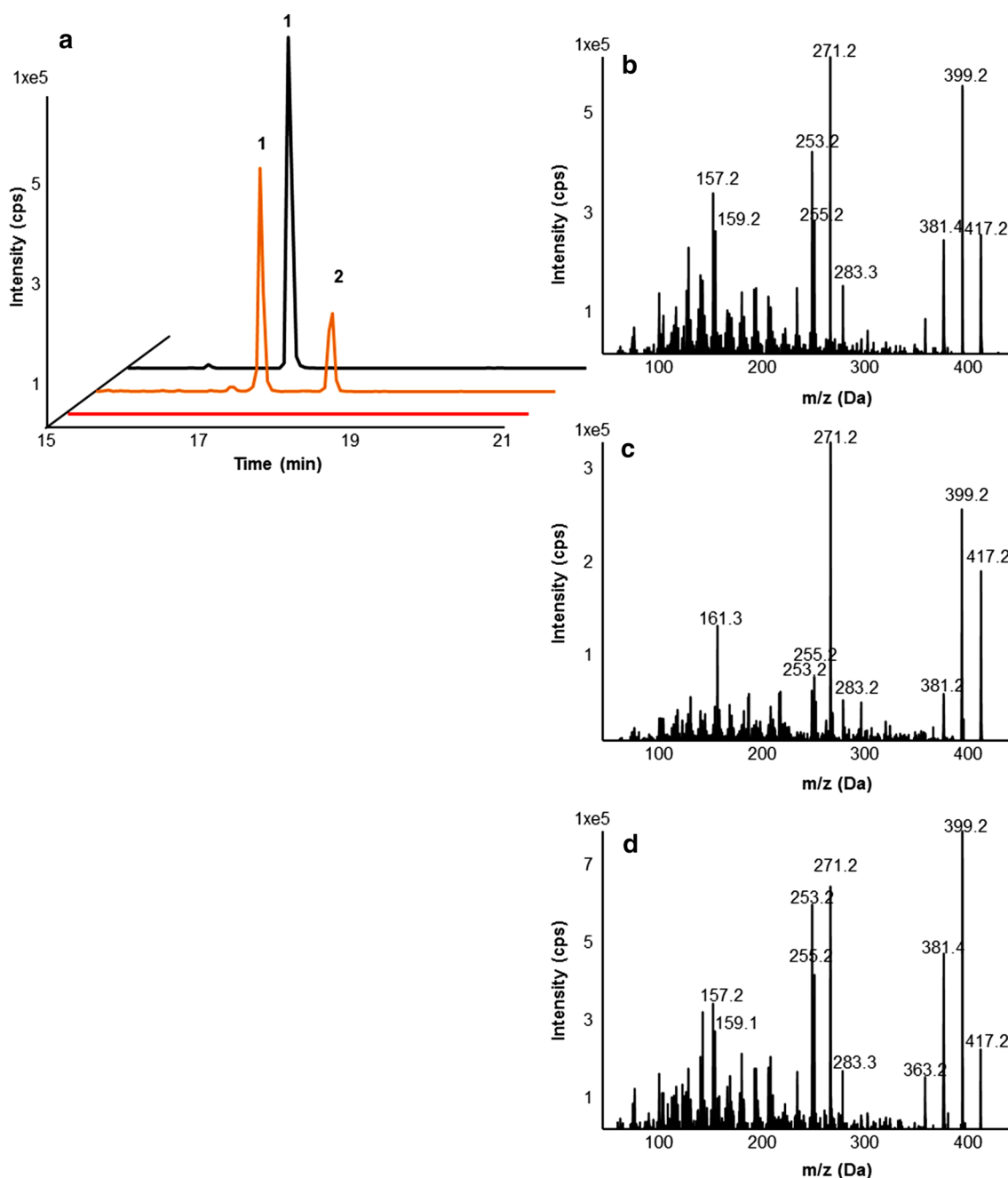
Fig. 5 Targeted LC–MS/MS analysis of 22-keto-26[25(*R*)]-hydroxycholesterol from transgenic *Camelina sativa* seeds expressing CYP90B27, CYP94N1, CYP90G1, and GABAT1 from *Veratrum californicum* and GAD2 from *Arabidopsis thaliana*. Samples were extracted and analyzed by QTRAP 6500 using MRM scan for 417.3–271.0 *m/z* and EPI scan for 417.3 *m/z*. **a** Overlay of wild type *C. sativa* seeds (*red*), *C. sativa* seeds expressing CYP90B27, CYP94N1, CYP90G1, GABAT1, and GAD2 (CXTOG) (*orange*), and 22-keto-26[25(*R*)]-hydroxycholesterol extracted from heterologous expression of CYP90B27, CYP94N1, CYP90G1, CPR, and GABAT1 in *S. frugiperda* Sf9 insect cells (*black*). Mass spectrum of (**b**) 22-keto-26[25(*R*)]-hydroxycholesterol from transgenic *C. sativa* (peak 1), **c** peak 2 in transgenic *C. sativa*, and **d** 22-keto-26[25(*R*)]-hydroxycholesterol from *S. frugiperda* Sf9 insect cells. Selected fragment ions are shown. CPR refers to cytochrome P450 reductase from *E. californica*, and GAD2 refers to glutamate decarboxylase 2

Discussion

Camelina sativa engineered to express CYP90B27, CYP94N1, GABAT1, and GAD2 (with and without CYP90G1), were shown to accumulate verazine in seed, the hypothesized precursor to the antineoplastic cyclopamine. Verazine accumulation in plants lacking CYP90G1 is most likely due to the oxidizing capabilities of CYP90B27
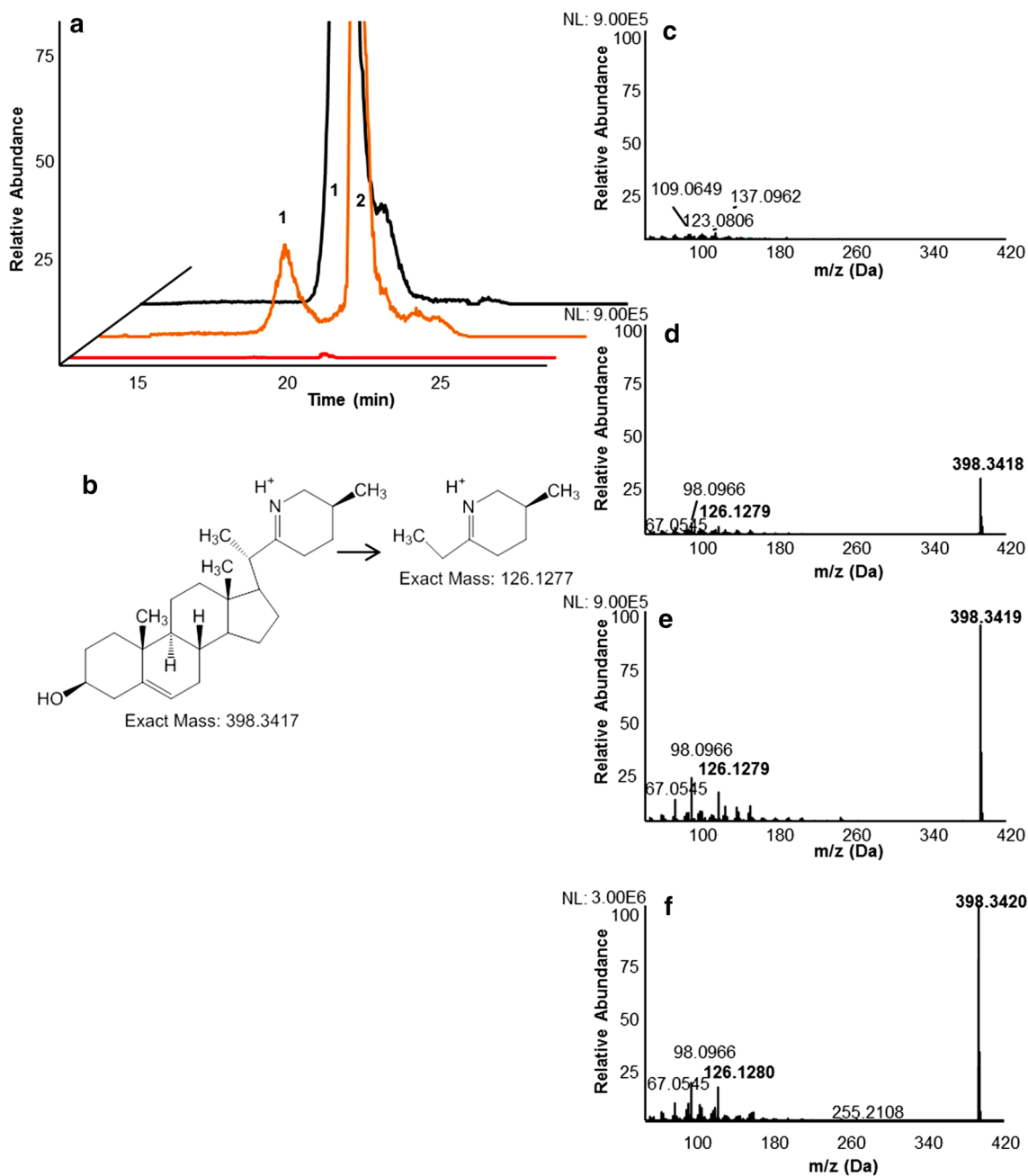
Fig. 6 Verazine synthesized in transgenic *Camelina sativa* seed. Wild type *C. sativa* seeds and *C. sativa* seeds transformed with CYP90B27, CYP94N1, CYP90G1, and GABAT1 from *Veratrum californicum* and GAD2 from *Arabidopsis thaliana* (CXTOG) were extracted and analyzed by high resolution mass spectrometry with the Q-Exactive. **a** Overlay of MS2 scan for 398.34 *m/z* of wild type *C. sativa* seeds (*red*), transgenic (CXTOG) *C. sativa* seeds (*orange*), and *S. frugiperda* Sf9 insect cells expressing CYP90B27, CYP94N1, CYP90G1, GABAT1, and CPR for production of verazine (*black*).

Peak 1 is verazine, peak 2 is unknown. **b** Calculated exact mass of verazine and key fragment ion. MS2 fragmentation pattern for peak 1 are shown on the right including, **c** wild type *C. sativa*, **d** transgenic *C. sativa*, and **e** Sf9 extracts. **f** Mass spectrum for peak 2 in transgenic *C. sativa*. Spectra were filtered for exact mass of 398.3417 with a mass tolerance of 20 ppm. Exact mass of verazine and key fragment are indicated in *bold*. CPR refers to cytochrome P450 reductase from *E. californica* and GAD2 refers to glutamate decarboxylase 2

at position C-22 (Augustin et al. 2015b). Not surprisingly, plants transformed with the first two genes in the pathway, CYP90B27 and CYP94N1 (CO), were found to contain 22(*R*),26-dihydroxycholesterol. As expected, plants

containing the first three genes CYP90B27, CYP94N1, GABAT1, and GAD2 (CTOG) accumulated 22(*R*)-hydroxy-26[25(*S*)]-aminocholesterol in addition to verazine. 22-Keto-26[25(*R*)]-hydroxycholesterol was also

detected in plants containing all genes (CXTOG). The lack of 22-keto-26[25(*R*)]-hydroxycholesterol detection in these plants by GC–MS is likely due to low sensitivity of the instrument. Interestingly, peaks with unknown identity were discovered in the transgenic *C. sativa* plants CTOG and CXTOG. A minimum of 3 additional peaks were detected alongside 22(*R*)-hydroxy-26[25(*S*)]-aminocholesterol, each appearing to have the same mass and similar fragmentation pattern but possessing different retention times. CXTOG plants contained one additional compound similar to 22-keto-26[25(*R*)]-hydroxycholesterol; the same was seen for verazine. These additional peaks suggest the possibility of alternative stereoisomers produced in *C. sativa*. Tautomerization could account for alternative stereochemical configurations, however, these unknown metabolites were not detected in Sf9 insect cells expressing the *V. californicum* genes. Therefore, they are most likely produced by plant specific enzymes present in *C. sativa*. Interestingly, the probable stereoisomer of verazine discovered in CXTOG lines was not detected in the CTOG plants while verazine was detected in both.

An accumulation of 22-keto-26[25(*R*)]-hydroxycholesterol in both *C. sativa* and Sf9 insect cells expressing CYP90B27, CYP94N1, CYP90G1, GABAT1, and CPR (Augustin et al. 2015b) suggested that it was not a substrate for oxidation by CYP94N1 and transamination by GABAT1. It is possible that this compound accumulates because it has the wrong C-25 configuration for conversion to verazine. If CYP94N1 hydroxylates in a non-stereospecific manner, it would produce both 22(*R*),26[25(*R*)]-dihydroxycholesterol and 22(*R*),26[25(*S*)]-dihydroxycholesterol, isomers that we were unable to separate and distinguish. Subsequent oxidation by CYP94N1, however, may be stereospecific and only accept 22(*R*),26[25(*S*)]-dihydroxycholesterol. In *V. californicum*, additional unidentified enzymes may convert 22-keto-26[25(*R*)]-hydroxycholesterol to other secondary metabolites. Another possibility is that the equilibrium of these compounds may be altered by the presence of additional, subsequent enzymatic reactions, resulting in a higher accumulation of the 25(*S*) metabolites. The accumulation of 22-keto-26[25(*R*)]-hydroxycholesterol in transgenic *C. sativa* in conjunction with minor accumulation in *V. californicum* (Augustin et al. 2015b) also supports this conclusion.

The low level of verazine accumulated (average of 50 pg verazine per mg of seed) were suboptimal for production. Continued engineering and genetic manipulation of *C. sativa* is required for an increase in overall verazine yield. Addition of subsequent pathway enzymes may increase yield by potentially driving the equilibrium to production of metabolites with the correct stereochemistry, therefore, resulting in a reduction/ elimination of unwanted side products. A representative example was described in Kristensen et al., 2005 for *A. thaliana* plants engineered with metabolic genes for dhurrin synthesis, a cyanogenic glucoside from sorghum. Many unintended side products were found to accumulate in transgenic plants containing the first two genes in the biosynthetic pathway. When the third gene was introduced, and the pathway to dhurrin, therefore, complete, the unintended side products were no longer detected and end product accumulation was greatly enhanced (Kristensen et al. 2005). By introducing sequential cyclopamine biosynthetic enzymes (after they are discovered), similar results may be achieved.

Verazine production could also be enhanced in *C. sativa* by overexpression of genes that increase accumulation of cholesterol, a precursor molecule for steroid alkaloids. *C. sativa* has been previously found to contain nearly 200 μg of cholesterol per g of oil, a relatively high amount when compared to other oil crops such as coconut oil and palm oil, for example (Shukla et al. 2002), however, cholesterol was still identified as only a minor steroid in *C. sativa*, a possible limiting factor for triterpenoid metabolite production (Mansour et al. 2014; Shukla et al. 2002). Overexpression of oxidosqualene cyclases such as cycloartenol synthase or lanosterol synthase may increase the available cholesterol (Ohyama et al. 2009). Moreover, enhanced expression of the recently discovered Sterol Side Chain Reductase 2, a critical enzyme in plant cholesterol biosynthesis, could be another target for metabolic engineering (Sawai et al. 2014). In addition, upregulation of known bottleneck enzymes in the mevalonic acid pathway may enhance yield. As previously demonstrated by Augustin et al. 2015a, overexpression of the rate limiting enzyme 1-deoxy-D-xylulose-5-phosphate synthase (DXS) in *C. sativa* engineered to produce limonene and cadinene significantly increased terpene production. A deeper understanding of cholesterol biosynthesis in plants will enhance rational engineering in this regard.

Cultivation of *C. sativa* as an agricultural crop is industrially and environmentally attractive due to its low input requirements and biochemical properties. Efficient use of cultivatable land by growth and harvest of *C. sativa* prior to a typical growing season provides an opportunity for specialty chemical production without jeopardizing food crops or compromising undeveloped wilderness (Bansal and Durrett 2016). In addition, after chemical extraction, the remaining material can be used for hydrothermal processing to biofuels or other industrial applications, resulting in little waste (Asomaning et al. 2014). Herein, we demonstrated the biosynthesis of *V. californicum* secondary metabolites in *C. sativa*, highlighting the potential for future engineering and industrial production. In addition, we enhanced our knowledge of the stereochemical configuration of the produced metabolites.

Author contributions MMA, AKS, LH, and CKH constructed vectors for transformation. MMA and LH performed plant transformations and seed collection. MMA and LH extracted seeds for metabolite analysis. MMA ran all samples on the GC–MS and LC–MS/MS and analyzed the results. CMS and MOJ performed the Mosher ester analysis, NMR, and ELSD. MMA analyzed data and generated figures. MMA and TMK wrote the manuscript. TMK oversaw all work.

References

Abdel-Kader MS, Bahler BD, Malone S, Werkhoven MC, van Troon F, David, Wisse JH, Bursuker I, Neddermann KM, Mamber SW, Kingston DG (1998) DNA-damaging steroidal alkaloids from Eclipta alba from the suriname rainforest. J Nat Prod 61:1202–1208

Asomaning J, Mussone P, Bressler DC (2014) Two-stage thermal conversion of inedible lipid feedstocks to renewable chemicals and fuels. Bioresour Technol 158:55–62

Atanasov AG, Waltenberger B, Pferschy-Wenzig EM, Linder T, Wawrosch C, Uhrin P, Temml V, Wang L, Schwaiger S, Heiss EH, Rollinger JM, Schuster D, Breuss JM, Bochkov V, Mihovilovic MD, Kopp B, Bauer R, Dirsch VM, Stuppner H (2015) Discovery and resupply of pharmacologically active plant-derived natural products: A review. Biotechnol Adv 33:1582–1614

Augustin JM, Higashi Y, Feng X, Kutchan TM (2015a) Production of mono- and sesquiterpenes in *Camelina sativa* oilseed. Planta 242:693–708

Augustin MM, Ruzicka DR, Shukla AK, Augustin JM, Starks CM, O'Neil-Johnson M, McKain MR, Evans BS, Barrett MD, Smithson A, Wong GK, Deyholos MK, Edger PP, Pires JC, Leebens-Mack JH, Mann DA, Kutchan TM (2015b) Elucidating steroid alkaloid biosynthesis in *Veratrum californicum*: production of verazine in Sf9 cells. Plant J 82:991–1003

Bahra M, Kamphues C, Boas-Knoop S, Lippert S, Esendik U, Schuller U, Hartmann W, Waha A, Neuhaus P, Heppner F, Pietsch T, Koch A (2012) Combination of hedgehog signaling blockage and chemotherapy leads to tumor reduction in pancreatic adenocarcinomas. Pancreas 41:222–229

Bansal S, Durrett TP (2016) *Camelina sativa*: an ideal platform for the metabolic engineering and field production of industrial lipids. Biochimie 120:9–16

Batsaikhan BE, Yoshikawa K, Kurita N, Iwata T, Takasu C, Kashihara H, Shimada M (2014) Cyclopamine decreased the expression of Sonic Hedgehog and its downstream genes in colon cancer stem cells. Anticancer Res 34:6339–6344

Berman DM, Karhadkar SS, Hallahan AR, Pritchard JI, Eberhart CG, Watkins DN, Chen JK, Cooper MK, Taipale J, Olson JM, Beachy PA (2002) Medulloblastoma growth inhibition by hedgehog pathway blockade. Science 297:1559–1561

Betancor MB, Sprague M, Sayanova O, Usher S, Campbell PJ, Napier JA, Caballero MJ, Tocher DR (2015a) Evaluation of a high-EPA oil from transgenic in feeds for *Atlantic salmon* (L.): effects on tissue fatty acid composition, histology and gene expression. Aquaculture 444:1–12

Betancor MB, Sprague M, Usher S, Sayanova O, Campbell PJ, Napier JA, Tocher DR (2015b) A nutritionally-enhanced oil from transgenic *Camelina sativa* effectively replaces fish oil as a source of eicosapentaenoic acid for fish. Sci Rep 5:8104

Chen JK, Taipale J, Cooper MK, Beachy PA (2002) Inhibition of Hedgehog signaling by direct binding of cyclopamine to Smoothened. Genes Dev 16:2743–2748

Finamore E, Minale L, Riccio R, Rinaldo G, Zollo F (1991) Novel Marine Polyhydroxylated Steroids from the Starfish Myxoderma-Platyacanthum. J Org Chem 56:1146–1153

Heretsch P, Tzagkaroulaki L, Giannis A (2010) Cyclopamine and hedgehog signaling: chemistry, biology, medical perspectives. Angew Chem Int Ed Engl 49:3418–3427

Iven T, Hornung E, Heilmann M, Feussner I (2015) Synthesis of oleyl oleate wax esters in *Arabidopsis thaliana* and *Camelina sativa* seed oil. Plant Biotechnol J 14:252–259

Jimeno A, Weiss GJ, Miller WH Jr, Gettinger S, Eigl BJ, Chang AL, Dunbar J, Devens S, Faia K, Skliris G, Kutok J, Lewis KD, Tibes R, Sharfman WH, Ross RW, Rudin CM (2013) Phase I study of the Hedgehog pathway inhibitor IPI-926 in adult patients with solid tumors. Clin Cancer Res 19:2766–2774

Kagale S, Koh C, Nixon J, Bollina V, Clarke WE, Tuteja R, Spillane C, Robinson SJ, Links MG, Clarke C, Higgins EE, Huebert T, Sharpe AG, Parkin IA (2014) The emerging biofuel crop *Camelina sativa* retains a highly undifferentiated hexaploid genome structure. Nat Commun 5:3706

Keeler RF, Binns W (1968) Teratogenic compounds in *Veratrum californicum* (Durand) V. Comparison of cyclopian effects of steroidal alkaloids from the plant and structurally related compounds from other sources. Teratology 1:5–10

Kim HJ, Silva JE, Vu HS, Mockaitis K, Nam JW, Cahoon EB (2015) Toward production of jet fuel functionality in oilseeds: identification of FatB acyl-acyl carrier protein thioesterases and evaluation of combinatorial expression strategies in Camelina seeds. J Exp Bot 66:4251–4265

Kristensen C, Morant M, Olsen CE, Ekstrom CT, Galbraith DW, Moller BL, Bak S (2005) Metabolic engineering of dhurrin in transgenic Arabidopsis plants with marginal inadvertent effects on the metabolome and transcriptome. Proc Natl Acad Sci USA 102:1779–1784

Li-Beisson Y, Shorrosh B, Beisson F, Andersson MX, Arondel V, Bates PD, Baud S, Bird D, Debono A, Durrett TP, Franke RB, Graham IA, Katayama K, Kelly AA, Larson T, Markham JE, Miquel M, Molina I, Nishida I, Rowland O, Samuels L, Schmid KM, Wada H, Welti R, Xu C, Zallot R and Ohlrogge J (2013) Acyl-lipid metabolism. Arabidopsis Book, vol 11, p e0161

Lin TL, Wang QH, Brown P, Peacock C, Merchant AA, Brennan S, Jones E, McGovern K, Watkins DN, Sakamoto KM, Matsui W (2010) Self-renewal of acute lymphocytic leukemia cells is limited by the Hedgehog pathway inhibitors cyclopamine and IPI-926. PLoS One 5:e15262

Liu J, Rice A, McGlew K, Shaw V, Park H, Clemente T, Pollard M, Ohlrogge J, Durrett TP (2015) Metabolic engineering of oilseed crops to produce high levels of novel acetyl glyceride oils with reduced viscosity, freezing point and calorific value. Plant Biotechnol J 13:858–865

Lu C, Kang J (2008) Generation of transgenic plants of a potential oilseed crop *Camelina sativa* by Agrobacterium-mediated transformation. Plant Cell Rep 27:273–278

Ma R, Ritala A, Oksman-Caldentey KM, Rischer H (2006) Development of in vitro techniques for the important medicinal plant *Veratrum californicum*. Planta Med 72:1142–1148

Malik MR, Yang W, Patterson N, Tang J, Wellinghoff RL, Preuss ML, Burkitt C, Sharma N, Ji Y, Jez JM, Peoples OP, Jaworski JG, Cahoon EB, Snell KD (2015) Production of high levels of poly-3-hydroxybutyrate in plastids of *Camelina sativa* seeds. Plant Biotechnol J 13:675–688

Mansour MP, Shrestha P, Belide S, Petrie JR, Nichols PD, Singh SP (2014) Characterization of oilseed lipids from "DHA-producing *Camelina sativa*": a new transformed land plant containing long-chain omega-3 oils. Nutrients 6:776–789

Ohyama K, Suzuki M, Kikuchi J, Saito K, Muranaka T (2009) Dual biosynthetic pathways to phytosterol via cycloartenol and lanosterol in *Arabidopsis*. Proc Natl Acad Sci USA 106:725–730

Olive KP, Jacobetz MA, Davidson CJ, Gopinathan A, McIntyre D, Honess D, Madhu B, Goldgraben MA, Caldwell ME, Allard D, Frese KK, Denicola G, Feig C, Combs C, Winter SP, Ireland-Zecchini H, Reichelt S, Howat WJ, Chang A, Dhara M, Wang L, Ruckert F, Grutzmann R, Pilarsky C, Izeradjene K, Hingorani SR, Huang P, Davies SE, Plunkett W, Egorin M, Hruban RH, Whitebread N, McGovern K, Adams J, Iacobuzio-Donahue C, Griffiths J, Tuveson DA (2009) Inhibition of Hedgehog signaling enhances delivery of chemotherapy in a mouse model of pancreatic cancer. Science 324:1457–1461

Petrie JR, Shrestha P, Belide S, Kennedy Y, Lester G, Liu Q, Divi UK, Mulder RJ, Mansour MP, Nichols PD, Singh SP (2014) Metabolic engineering *Camelina sativa* with fish oil-like levels of DHA. PLoS One 9:e85061

Putnam DH, Budin JT, Field LA and M. BW (1993) Camelina: a promising low-input oilseed. Wiley New York

Ruiz-Lopez N, Haslam RP, Napier JA, Sayanova O (2014) Successful high-level accumulation of fish oil omega-3 long-chain polyunsaturated fatty acids in a transgenic oilseed crop. Plant J 77:198–208

Sawai S, Ohyama K, Yasumoto S, Seki H, Sakuma T, Yamamoto T, Takebayashi Y, Kojima M, Sakakibara H, Aoki T, Muranaka T, Saito K, Umemoto N (2014) Sterol side chain reductase 2 is a key enzyme in the biosynthesis of cholesterol, the common precursor of toxic steroidal glycoalkaloids in potato. Plant Cell 26:3763–3774

Shukla VKS, Dutta PC, Artz WE (2002) Camelina oil and its unusual cholesterol content. J Am Oil Chem Soc 79:965–969

Song J, Naylor-Adelberg J, White S, Mann D, Adelberg J (2014) Establishing clones of *Veratrum californicum*, a native medicinal species, for micropropagation. In Vitro Cell Dev Biol Plant 50:337–344

Taipale J, Chen JK, Cooper MK, Wang B, Mann RK, Milenkovic L, Scott MP, Beachy PA (2000) Effects of oncogenic mutations in Smoothened and Patched can be reversed by cyclopamine. Nature 406:1005–1009

Tremblay MR, Lescarbeau A, Grogan MJ, Tan E, Lin G, Austad BC, Yu LC, Behnke ML, Nair SJ, Hagel M, White K, Conley J, Manna JD, Alvarez-Diez TM, Hoyt J, Woodward CN, Sydor JR, Pink M, MacDougall J, Campbell MJ, Cushing J, Ferguson J, Curtis MS, McGovern K, Read MA, Palombella VJ, Adams J, Castro AC (2009) Discovery of a potent and orally active hedgehog pathway antagonist (IPI-926). J Med Chem 52:4400–4418

Turano FJ, Fang TK (1998) Characterization of two glutamate decarboxylase cDNA clones from *Arabidopsis*. Plant Physiol 117:1411–1421

Watkins DN, Berman DM, Burkholder SG, Wang B, Beachy PA, Baylin SB (2003) Hedgehog signalling within airway epithelial progenitors and in small-cell lung cancer. Nature 422:313–317