New Phytologist

The metabolite chemotype of *Nicotiana benthamiana* transiently expressing artemisinin biosynthetic pathway genes is a function of *CYP71AV1* type and relative gene dosage

Hieng-Ming Ting[1], Bo Wang[1], Anna-Margareta Rydén[1,2], Lotte Woittiez[1], Teun van Herpen[1], Francel W. A. Verstappen[1], Carolien Ruyter-Spira[1], Jules Beekwilder[1,2], Harro J. Bouwmeester[1] and Alexander van der Krol[1]

[1]Laboratory of Plant Physiology, Wageningen University, Droevendaalsesteeg 1, 6708 PB, Wageningen, the Netherlands; [2]Plant Research International, Droevendaalsesteeg 1, 6708 PB, Wageningen, the Netherlands

Author for correspondence:
Harro J. Bouwmeester
Tel: +31 317 489 859
Email: harro.bouwmeester@wur.nl

Received: *27 November 2012*
Accepted: *10 March 2013*

New Phytologist (2013) **199:** 352–366
doi: 10.1111/nph.12274

Key words: *Artemisia annua*, artemisinin, CYP71AV1, *Nicotiana benthamiana*, transient expression.

Summary

- *Artemisia annua*, which produces the anti-malaria compound artemisinin, occurs as high-artemisinin production (HAP) and low-artemisinin production (LAP) chemotypes. Understanding the basis of the difference between these chemotypes would assist breeding and optimising artemisinin biosynthesis.
- Here we present a systematic comparison of artemisinin biosynthesis genes that may be involved in determining the chemotype (*CYP71AV1*, *DBR2* and *ALDH1*). These genes were isolated from the two chemotypes and characterized using transient expression *in planta*. The enzyme activity of DBR2 and ALDH1 from the two chemotypes did not differ, but structural differences in CYP71AV1 from LAP and HAP chemotypes (AMOLAP and AMOHAP, respectively) resulted in altered enzyme activity.
- AMOLAP displays a seven amino acids N-terminal extension compared with AMOHAP. The GFP fusion of both proteins show equal localization to the ER but AMOHAP may have reduced stability.
- Upon transient expression in *Nicotiana benthamiana*, AMOLAP displayed a higher enzyme activity than AMOHAP. However, expression in combination with the other pathway genes also resulted in a qualitatively different product profile ('chemotype'); that is, in a shift in the ratio between the unsaturated and saturated (dihydro) branch of the pathway.

Introduction

Based on the content of artemisinin and its precursors, two chemotypes of *Artemisia annua*, can be distinguished: the low-artemisinin production (LAP) chemotype and a high-artemisinin production (HAP) chemotype (Wallaart *et al.*, 2000). Both chemotypes contain artemisinin and arteannuin B, but the HAP chemotype has a relatively high content of artemisinin and its presumed precursor dihydroartemisinic acid (DHAA), while the LAP chemotype has a high content of arteannuin B and its presumed precursor artemisinic acid (AA).

The artemisinin biosynthesis pathway has been largely elucidated and the genes required for production of dihydroartemisinic acid, the most likely precursor of artemisinin (*ADS*, *CYP71AV1*, *DBR2* and *ALDH1*) have all been described (Bouwmeester *et al.*, 1999; Teoh *et al.*, 2006, 2009; Zhang *et al.*, 2008; Rydén *et al.*, 2010). Artemisinin is a sesquiterpene lactone endoperoxide, which is synthesized in the cytosol from the general isoprenoid precursors IPP and DMAPP. These are converted to FPP and the first committed step in the artemisinin biosynthetic pathway is the cyclization of FPP to amorpha-4,11-diene (AD) by amorphadiene synthase (Fig. 1; Bouwmeester *et al.*, 1999; Mercke *et al.*, 2000). In the subsequent step, AD is oxidized by the cytochrome P450 enzyme, CYP71AV1/AMO, to artemisinic alcohol (AAOH), artemisinic aldehyde (AAA) and artemisinic acid (AA; Fig. 1; Ro *et al.*, 2006; Teoh *et al.*, 2006). However, the latter mainly occurs in the LAP chemotype. In the HAP chemotype only very little of the AAA is converted to AA, as most of the AAA is converted to dihydroartemisinic aldehyde (DHAAA) by DBR2, the enzyme that reduces the exocyclic double bond of AAA (Fig. 1; Bertea *et al.*, 2005; Zhang *et al.*, 2008). Supposedly, DHAAA is subsequently oxidized by alcohol dehydrogenase ALDH1 to the final intermediate dihydroartemisinic acid (DHAA; Bertea *et al.*, 2005; Zhang *et al.*, 2008). The conversion of DHAA to artemisinin, is believed to be a nonenzymatic and spontaneous photo-oxidation reaction (Wallaart *et al.*, 1999; Sy & Brown, 2002). Similarly, in the LAP chemotype, AA is likely spontaneously converted to arteannuin B.

Recently we reported on the (transient) reconstruction of the artemisinin biosynthetic pathway in *Nicotiana benthamiana*

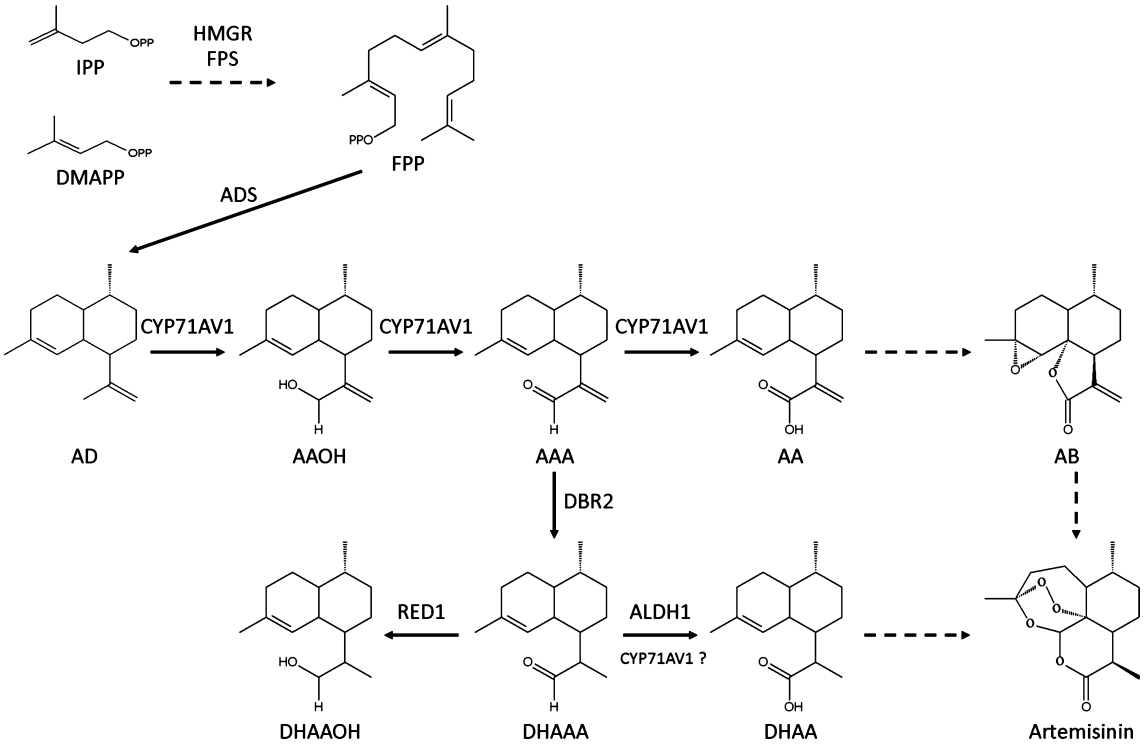

Fig. 1 Artemisinin biosynthetic pathway in *Artemisia annua*. IPP, isopentenyl diphosphate; DMAPP, dimethylallyl diphosphate; FPP, farnesyl diphosphate; AD, amorpha-4,11-diene; AAOH, artemisinic alcohol; AAA, artemisinic aldehyde; AA, artemisinic acid; AB, arteannuin B; DHAAOH, dihydroartemisinic alcohol; DHAAA, dihydroartemisinic aldehyde; DHAA, dihydroartemisinic acid; FPS, farnesyl diphosphate synthase; HMGR, 3-hydroxy-3-methylglutaryl-CoA reductase; ADS, amorphadiene synthase; CYP71AV1, amorphadiene oxidase; DBR2, artemisinic aldehyde double-bond reductase; RED1, dihydroartemisinic aldehyde reductase 1; ALDH1, aldehyde dehydrogenase 1. Broken arrows indicate the involvement of more than one step (Nguyen *et al.*, 2011).

leaves, resulting in up to 39.5 mg kg^{-1} FW of AA (van Herpen *et al.*, 2010). In the present work we analyse the role of DBR2, ALDH1 and CYP71AV1 in determining the 'chemotype' (as defined by the AA and DHAA ratio) of *N. benthamiana* leaves agro-infiltrated with artemisinin biosynthesis genes. Results show that the chemotype is a function of the *CYP71AV1* type and relative dosage of *DBR2* and *ALDH1*.

Materials and Methods

Cloning of the ADS + FPS + HMGR expression construct *AmFH*

Cloning of the *AmFH* expression construct which contains *ADS* with CoxIV mitochondrial targeting signal, mitochondrial targeted *FPS* and a cytosolic (truncated) *HMGR* under control of the CaMV35S promoter has been described previously (van Herpen *et al.*, 2010).

Identification and cloning of *CYP71AV1* from HAP and LAP chemotypes

The *AMOHAP* EST sequence was identified in the sequence database of an *Artemisia annua* L. (HAP chemotype) glandular trichome cDNA library (Bertea *et al.*, 2006). This sequence has been deposited in the GenBank database (JQ254992). The full length sequence of *AMOHAP* was obtained by RACE PCR (Clontech, Mountain View, CA, USA). We note that the race PCR did not yield any sequences that indicated the presence of a *AMOLAP* version in our cDNA library. Subsequently, the full length coding region was amplified from *A. annua* trichome cDNA by PCR using Phusion polymerase (Finnzymes, Espoo, Finland) using primers 1 + 2 (Supporting Information Table S1). The full length *AMOLAP* coding region was cloned by primers 3 + 4. After confirmation of the correct sequences both the *AMOHAP* and *AMOLAP* coding region were isolated from pGEMT/AMOHAP and pGEMT/AMOLAP using *Bam*HI and *Kpn*I for cloning into the yeast expression vector pYEDP60 (Pompon *et al.*, 1996), resulting in pYEDP60-AMOHAP and pYEDP60-AMOLAP.

For the cloning into a plant binary expression vector, the genes were first introduced into ImpactVectorC3.1 (http://www.wageningenur.nl/en/show/Productie-van-farmaceutische-en-industriele-eiwitten-door-planten.htm). For this a *Kpn*I site was introduced into C3.1, resulting in C3.1/*Kpn*I. Hereto, two oligos (GATCCATTTCGGTACCAATTAGC and GGCCG CTAATTGGTACCGAAATG) were hybridized, kinase-treated and ligated into C3.1 digested with *Bam*HI and *Not*I. The *Bam*HI/*Kpn*I fragment of pGEMT/AMOHAP and pGEMT/AMOLAP was isolated and ligated into the vector C3.1/*Kpn*I between the CaMV35S promoter and Rbcs1 terminator. The resulting plasmids C3.1/AMOHAP and C3.1/AMOLAP were

digested with *Asc*I and *Pac*I and the full gene sequence was cloned into the *Asc*I and *Pac*I site of the pBinPlus binary vector (van Engelen *et al.*, 1995).

Identification and cloning of *DBR2* from HAP and LAP chemotypes

We cloned a *DBR2HAP* from *A. annua* HAP chemotype (Bertea *et al.*, 2006) and *DBR2LAP* was isolated from *A. annua* LAP chemotype from Iran (Table S2). Both *DBR2* cDNA sequences were isolated by RT-PCR from cDNA constructed from RNA extracted from *A. annua* flowers isolated from either chemotype. For amplification of the *DBR2* sequence, primers 5 + 6 were used (Table S1), of which sequences were based on the published *DBR2* sequence from a HAP chemotype (Zhang *et al.*, 2008). These primer sets were also able to amplify a *DBR2* sequence from the LAP chemotype. The primers used introduce *Bam*HI and *Not*I restriction sites which were used for cloning into ImpactVectorpIV1A_2.1 (www.pri.wur.nl/UK/products/ImpactVector/). The resulting pIV1A_2.1/DBR2 was digested with *Asc*I and *Pac*I and the full gene was cloned into the *Asc*I and *Pac*I sites of the pBinPlus binary vector (van Engelen *et al.*, 1995). *DBR2* sequences identified here have been deposited in the GenBank database (KC505370, JX898526 and JX898527).

Identification and cloning of *ALDH1* from HAP and LAP chemotypes

ALDH1 sequences were isolated by RT-PCR on cDNA from floral RNA isolated from the *A. annua* HAP and LAP chemotypes using primers 7 + 8 (Table S1). The sequences of these primers were based on the published *ALDH1* sequence from a HAP chemotype (Teoh *et al.*, 2009) and could amplify an *ALDH1* sequence from both HAP and LAP cDNA. Inspection of the coding sequence revealed that there were no specific amino acid residue differences in ALDH1 from HAP and LAP (both identical to *ALDH1* GenBank: FJ809784). The *ALDH1* cDNA was cloned into pBinPlus binary vector (van Engelen *et al.*, 1995).

Cloning of CYP71AV1 GFP reporter constructs

For the GFP fusion protein expression constructs (NtermAMOHAP:GFP, NtermAMOLAP:GFP, AMOHAP:GFP and AMOLAP:GFP): the N-terminal domains of *AMOHAP* (first 43 codons) and *AMOLAP* (first 50 codons) were amplified using primers 9 + 10 and 11 + 12, respectively. After digestion with *Bam*HI and *Kpn*I the fragments were cloned into C3.1/*Kpn*I to form C3.1/NtermAMOHAP and C3.1/NtermAMOLAP.

The GFP coding sequence was amplified by PCR using primers 13 + 14 (Table S1) using pBin-Egfp as template (www.pri.wur.nl/UK/products/ImpactVector/). After digesting with *Kpn*I and *Not*I, GFP was subcloned into plasmid C3.1/NtermAMOHAP, C3.1/NtermAMOLAP, C3.1/AMOHAP and C3.1/AMOLAP to form 35S-NtermAMOHAP:GFP, 35S-NtermAMOLAP:GFP, 35S-AMOHAP:GFP and 35S-AMOLAP:GFP.

In order to remove the stop codon of *AMOHAP* and *AMOLAP*, and fuse them in-frame to the ATG start-codon of the GFP coding sequence in the 35S-AMOHAP:GFP and 35S-AMOLAP:GFP constructs, the 3′ends of the *AMOLAP* and *AMOHAP* were PCR amplified using Phusion DNA polymerase (Finnzymes) by primers 15 + 16, and 17 + 18, respectively.

The PCR products and plasmids 35S-AMOHAP:GFP and 35S-AMOLAP:GFP were digested with *Kpn*I and *Eco*RI, followed by gel purification. The PCR products were ligated into the plasmids and transformed into *Escherichia coli*. Positive colonies were analysed and sequenced to confirm that inserts were correct.

Transient expression in leaves of *Nicotiana benthamiana*

Agro-infiltration for transient expression in leaves of *Nicotiana benthamiana* Domin was carried out as described (van Herpen *et al.*, 2010). Briefly, individual *Agrobacterium tumefaciens* strains with different expression constructs (or empty vector as control) were co-infiltrated into *N. benthamiana* leaves using a syringe without needle. After 7 d of transient expression, leaves were harvested for chemical analysis. In each set of experiments the total dosage of *A. tumefaciens* between treatments was equalized by diluting with *A. tumefaciens* with empty vector where necessary. We note that leaves infiltrated with *AmFH+AMOLAP* developed necrotic lesions, indicating that at this time certain compounds started to accumulate to toxic concentrations in the infiltrated leaves.

CYP71AV1 subcellular localization studies

For subcellular localization of the AMO:GFP fusion proteins, *Arabidopsis thaliana* (L.) Heynh protoplasts were isolated and transfected with expression constructs 35S-NtermAMOHAP:GFP, 35S-NtermAMOLAP:GFP, 35S-AMOHAP:GFP and 35S-AMOLAP:GFP, based on a published protocol (Yoo *et al.*, 2007). The ER-YFP construct was used as reference for ER subcellular localization (Aker *et al.*, 2006). After transfection, protoplasts were analysed using Carl-Zeiss Confocal Scanning Laser Microscopy, with exitation of GFP at 488 nm and YFP at 514 nm. The fluorescence was detected via a band pass filter (GFP: 505–530 nm, YFP: 535–590 nm). Chlorophyll was detected using a 650 nm long pass filter.

Analysis of nonvolatile metabolites by LC-QTOF-MS/MS

Seven days after agro-infiltration of *N. benthamiana*, the infiltrated leaves were harvested, immediately frozen in liquid nitrogen and then ground to a fine powder. From each infiltrated leaf 100 mg of powder was extracted in 300 µl methanol : formic acid (1000 : 1, v/v). Nonvolatile compounds from the infiltrated leaves were analysed by LC-QTOF-MS as described (van Herpen *et al.*, 2010).

Data were processed using the protocol for untargeted metabolomics of plant tissues as described (De Vos *et al.*, 2007; van Herpen *et al.*, 2010). Briefly, LC-QTOF-MS data were analysed

using Masslynx v4.0 (Waters) and processed using MetAlign v1.0 (www.metAlign.nl) for baseline correction, noise elimination and subsequent spectral data alignment (De Vos *et al.*, 2007). The processing parameters of MetAlign for LC-QTOF-MS data were set to analyse from scan numbers 60–2590 (corresponding to retention time 1.15–49.16 min) with a maximum amplitude of 25 000. After MetAlign processing, masses were clustered using the Multivariate Mass Spectra Reconstruction (MMSR) approach (Tikunov *et al.*, 2005) to elucidate which mass signals originate from the same metabolite. The mass signal intensity differences between treatments were compared using Student's *t*-test. Mass-directed LC-QTOF-MS/MS analysis for further elucidation of metabolite identities was done on differential compounds with signal intensities > 500 ion counts per scan.

Quantification of artemisinin precursors by UPLC-MRM-MS

Targeted analysis of artemisinin precursors in agro-infiltrated *Nicotiana benthamiana* leaves was performed with a Waters Xevo tandem quadrupole mass spectrometer equipped with an electrospray ionization source and coupled to an Acquity UPLC system (Waters) as described (Kohlen *et al.*, 2011) with some modifications. For details and instrument settings see Methods S1.

Analysis of volatile metabolites by GC-MS

Extracts were analysed by GC-MS using a gas chromatograph (7890A; Agilent, Amstelveen, the Netherlands) equipped with a 30-m × 0.25-mm i.d., 0.25-mm film thickness column with 5-m guard column (Zebron ZB5-MS; Phenomenex, Utrecht, the Netherlands) and a mass selective detector (model 5965c, Agilent). The GC was programmed at an initial temperature of 80°C for 1 min, with a ramp of 5°C min^{-1} to 235°C and then a ramp of 25°C min^{-1} to 280°C with a final time of 5 min. The injection port temperature was 250°C, and the He inlet pressure was controlled with electronic pressure control to achieve a constant column flow of 1.0 ml min^{-1}. One microlitre of the extracts was injected in split mode with a split flow set at 9 ml min^{-1}. Scanning was performed from 45 to 450 atomic mass units.

Glycosidase treatment

Viscozyme L (Sigma) was used as glycosidase treatment to hydrolyse hexose-conjugated compounds for subsequent quantification using GC-MS. Hereto, 200 mg infiltrated leaf material from each treatment was incubated in 1 ml citrate phosphate buffer, pH 5.4 containing 200 µl of Viscozyme L as previously described (van Herpen *et al.*, 2010).

Glutathione conjugation assay

In vitro conjugation of metabolites to glutathione by glutathione transferase activity (GST) was performed as described (Liu *et al.*, 2011). In brief, glutathione (GSH; 150 mM) in 7 µl potassium phosphate buffer (100 mM; pH 6.5), and 30 mM of artemisinin precursor (AAA, AAOH, AA, DHAAOH, DHAAA and DHAA) in 7 µl ethanol were added to 200 µl potassium phosphate buffer (100 mM; pH 6.5). The reaction was initiated by adding 7 µl of glutathione transferase (GST; 1 g l^{-1}, in 100 mM KH_2PO_4 potassium phosphate buffer; pH 6.5) into the mixture. The controls were complete assay mixtures without GST enzyme or either of the substrates. After 15 min incubation at room temperature, samples were cooled to −20°C until LC-QTOF-MS analysis.

Results

Comparison of artemisinin biosynthesis protein sequences from HAP and LAP chemotypes reveals only relevant differences for CYP71AV1

Because the difference between the HAP and LAP chemotypes must arise after the ADS step in the biosynthesis pathway, and because no expression differences were found for *ADS* genes in HAP and LAP chemotypes (Maes *et al.*, 2011), here we focused on analysis of putative differences in biosynthesis genes downstream of *ADS* (e.g. *CYP71AV1*, *DBR2*, *ALDH1*) in search of an explanation for the two different chemotypes of *A. annua*. For this purpose *CYP71AV1*, *DBR2* and *ALDH1* were isolated from *A. annua* HAP and LAP chemotypes and the encoded protein sequences compared.

CYP71AV1 Analysis of the different *CYP71AV1* sequences that have been deposited in GenBank shows the occurrence of two major types of *CYP71AV1*, encoding two proteins which differ by a seven amino acids extension at the N-terminus of the protein (Fig. 2). The long version of CYP71AV1 (which we refer to as AMOLAP) has been isolated from *A. annua* Tanzania (Sandeman seed), which is a LAP chemotype (Ro *et al.*, 2006). The other long version of CYP71AV1 (which we refer to as AMOLAP.1) was isolated from a different *A. annua* LAP chemotype (Kim *et al.*, 1992; S-U. Kim, pers. comm.). Two versions of the CYP71AV1 (here referred to as AMOHAP and AMOHAP.1) were cloned from two different HAP chemotypes (Bertea *et al.*, 2006; Teoh *et al.*, 2006). The alignment of the AMOHAP and AMOLAP variants shows that none of the other single amino acid substitutions between the different AMOLAP and AMOHAP sequences are specific to the long or the short version of CYP71AV1 and therefore likely do not play a role in determining the LAP or HAP chemotypes (Fig. 2). RACE-PCR was used to analyse multiple *CYP71AV1* 5′ sequences amplified from RNA isolated from a HAP chemotype and only *AMOHAP* 5′ sequences were found (Fig. S1). Variation in the 5′ untranslated region (nt 56–60) could be an indication that two different alleles of *AMOHAP* are present in this chemotype, both translating into the short AMOHAP. The RACE sequence data were consistent with the recently published sequence of the *CYP71AV1* promoter cloned from an *A. annua* HAP chemotype (Wang *et al.*, 2011).

DBR2 We cloned *DBR2* from the *A. annua* HAP chemotype (here referred to as *DBR2HAP.1*) and the sequence we obtained

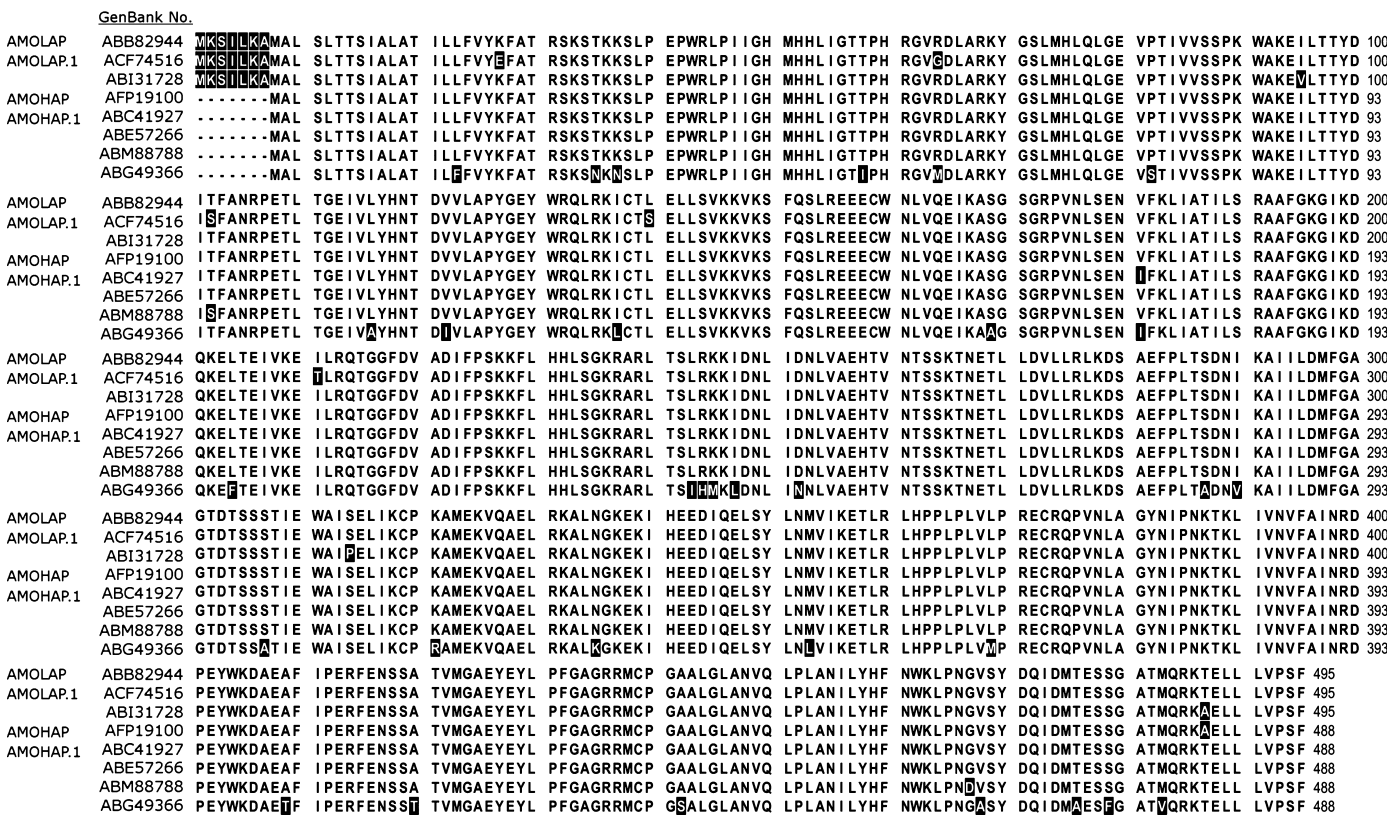

Fig. 2 Alignment of the AMOLAP and AMOHAP deduced amino acid sequences. All amino acid sequences used in the alignment were retrieved from GenBank (GenBank number given in front of each sequence). Two of the sequences were from confirmed LAP chemotypes: ABB82944, referred to as AMOLAP (Ro *et al.*, 2006) and ACF74516, referred to as AMOLAP.1 (Kim *et al.*, 1992). Two of the sequences were from confirmed HAP chemotypes: AFP19100, referred to as AMOHAP (Bertea *et al.*, 2006) and ABC41927, referred to as AMOHAP.1 (Teoh *et al.*, 2006). For the other sequences deposited in GenBank the chemotype of the plant from which they were isolated was not given.

Fig. 3 Alignment of the DBR2LAP and DBR2HAP deduced amino acid sequences. All amino acid sequences used in the alignment were retrieved from GenBank: DBR2LAP (JX898527), DBR2LAP.1 (KC505370), DBR2HAP (ACH61780) and DBR2HAP.1 (JX898526). DBR2LAP and DBR2LAP.1 were cloned from LAP, DBR2HAP was cloned from HAP (Zhang *et al.*, 2008) and DBR2HAP.1 was cloned from HAP (Bertea *et al.*, 2006). Alignment shows that one variant of DBR2LAP.1 is similar to DBR2HAP with only two AA differences. The second variant of the DBR2LAP has a two amino acids insertion at the C-terminus. No difference was detected in *in planta* activity of DBR2HAP and DBRLAP (Figs 5b,c and Supporting Information Fig. S3b).

was similar to the recently described *DBR2* (here referred to as *DBR2HAP*; Zhang *et al.*, 2008), with the exception of a one amino acid difference (Fig. 3). Using primers based on *DBRHAP* we isolated two variants of *DBR2* from an *A. annua* LAP chemotype (see Materials and Methods). Five of the 11 clones showed few amino acid differences with the published *DBR2HAP* (here referred to as *DBR2LAP.1*). However, in the remaining six clones (here referred to as *DBR2LAP*), the encoded protein sequence showed a number of amino acid residue differences with DBR2HAP, including two additional amino acids in position 295 (Fig. 3).

AlDH1 Cloning of *ALDH1* from *A. annua* has been described (Teoh *et al.*, 2009). We isolated *ALDH1* from both the HAP and

LAP chemotypes. Alignment of the AA-sequence showed no differences between the two proteins (data not shown).

Both AMOHAP and AMOLAP are localized to the endoplasmic reticulum

Because AMOLAP and AMOHAP only consistently differ in their N-terminal amino acid residues which supposedly encode the ER anchoring domain (Fig. 2), we investigated whether this difference causes altered subcellular targeting or difference in protein stability. Expression constructs encoding either full-length protein-GFP fusions or truncated N-terminal domain-GFP fusions of AMOLAP and AMOHAP were transiently expressed in *A. thaliana* protoplasts. Confocal microscopy of *Arabidopsis* protoplasts co-transfected with the full length AMOLAP protein fused to GFP (AMOLAP:GFP) showed co-localization of the GFP fluorescence signal with the fluorescence signal of the ER marker (ER:YFP; Fig. 4a). Similarly, full length AMOHAP protein fused to GFP (AMOHAP:GFP) also showed co-localization with the ER marker (ER:YFP; Fig. 4c). In addition, the truncated N-terminal portion of AMOLAP and AMOHAP were fused to GFP. When transfected into *Arabidopsis* protoplasts the NtermAMOLAP:GFP and NtermAMOHAP:GFP both showed co-localization with the ER:YFP ER marker (Fig. 4c,d). Localization experiments with 35S expression constructs may lead to artefacts such as cytosolic localization when ER import is saturated, however this was not observed in these experiments. Combined, the results demonstrate that AMOLAP and AMOHAP do not differ in subcellular targeting, as both proteins localize to the ER. Although the fluorescence signal varies between transfection assays, there were indications that AMOLAP may be more stable. For instance we do find a higher fluorescence signal for NtermAMOLAP:GFP than for the NtermAMOHAP:GFP.

Different product profiles *in planta* from AMOLAP and AMOHAP

Both the AMOLAP and the AMOHAP enzymes have been characterized in a yeast expression system and both were shown to be able to produce AAOH, AAA and AA from AD (Ro *et al.*, 2006; Teoh *et al.*, 2006). However, a direct comparison of variants AMOLAP and AMOHAP in the same expression system has not been performed until now. The different *in planta* expression studies using heterologous plant hosts (tobacco and *N. benthamiana*) have only been reported for the longer *AMOLAP* (van Herpen *et al.*, 2010; Zhang *et al.*, 2011). To test the effect of the seven AA extension of the AMOLAP protein we made two expression constructs, both based on the *AMOHAP* sequence but in one construct we introduced the seven AA extension to the protein sequence as found in AMOLAP, thus limiting

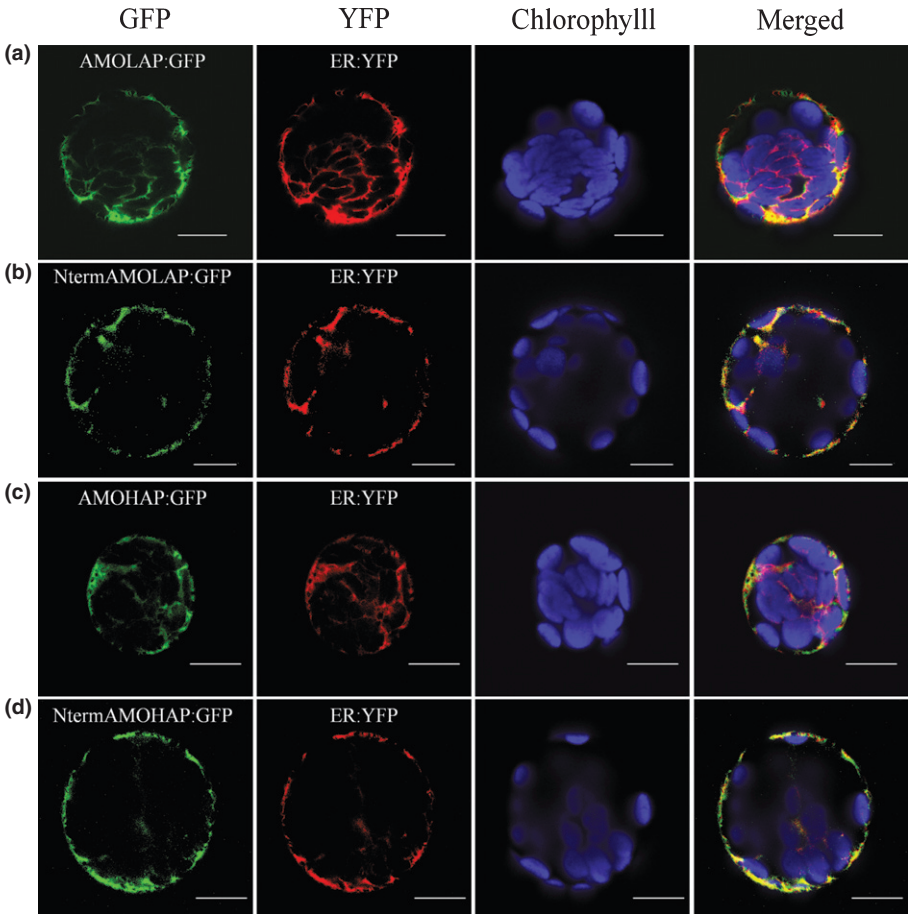

Fig. 4 Subcellular localization of AMOLAP and AMOHAP. Confocal microscopy analysis of *Arabidopsis* protoplasts co-transfected with (a) AMOLAP:GFP + ER:YFP; (b) Nterm AMOLAP:GFP + ER:YFP; (c) AMOHAP: GFP + ER:YFP; (d) NtermAMOHAP: GFP + ER:YFP. Artificial colours were given to GFP fluorescence (green), YFP fluorescence (red) and auto fluorescence of chloroplasts (purple). Merging of the pictures results in a yellow colour for GFP-YFP overlap. Bars, 10 μm.

the difference between the two forms to the N-terminal extension. The activity of these AMOLAP and AMOHAP genes was subsequently compared *in planta* by co-expression with *ADS*, using transient expression in *N. benthamiana* leaves. To achieve high concentrations of artemisinin precursor production, *ADS* was expressed with a mitochondrial targeting signal, and overexpression was combined with a mitochondrial targeted *FPS* and a truncated, cytosolic form of *HMGR*. *ADS*, *FPS* and *HMGR* were combined into a single 2A expression construct (*AmFH*) as described before (van Herpen *et al.*, 2010). Each expression construct (*AmFH* and *AMOLAP* or *AMOHAP*) was introduced into *A. tumefaciens* and *N. benthamiana* leaves were infiltrated with *AmFH+AMOLAP* or *AmFH+AMOHAP*.

Particularly with *AmFH+AMOLAP*, infiltrated leaves developed symptoms of necrosis around 7 d post infiltration (Fig. S2), suggesting the production of a toxic compound. Necrosis symptoms were stronger in leaves infiltrated with *AmFH+AMOLAP* than with *AmFH+AMOHAP*, suggesting that the products of both treatments may not be the same. Because necrosis started to appear after 7 d, in all our experiments the leaves were harvested at day 7 instead of day 10 after infiltration, as previously done (van Herpen *et al.*, 2010).

Analysis of free products In order to quantify the products from the ADS and AMOLAP/HAP transient enzyme activity in the infiltrated *N. benthamiana*, leaves were extracted with aqueous methanol for UPLC-MRM-MS analysis. Intriguingly, the distribution over the entire product range was different in leaves infiltrated with *AmFH+AMOLAP* and *AmFH+AMOHAP* (Table 1). Leaves infiltrated with *AmFH+AMOLAP* produced predominantly AA, while in leaves infiltrated with *AmFH+AMOHAP* AA concentrations were 50-fold lower. However, leaves infiltrated with *AmFH+AMOHAP* contained three-fold higher concentrations of AAOH and AAA than leaves expressing *AmFH+AMOLAP* (Table 1).

Also DHAAOH, DHAAA and DHAA were detected in the *AmFH+AMOLAP* leaf samples (Table 1), suggesting the presence of an endogenous *N. benthamiana* enzyme with carbon double bond reducing activity (catalysing the conversion of AAA to DHAAA just as DBR2 in *A. annua*), and enzymes similar to *A. annua* RED1 (catalysing the formation of DHAAOH from DHAAA) and *A. annua* ALDH1 (catalysing the conversion of DHAAA to DHAA; Fig. 1). Free DHAAOH concentrations were higher in *AmFH+AMOLAP* compared to *AmFH+AMOHAP* infiltrated leaves, suggesting that formation of DHAAOH is not directly related to the free DHAAA, AAOH or AAA concentrations in leaves, which were lower in *AmFH+AMOLAP*. DHAA was only detected in *AmFH+AMOLAP* infiltrated leaves.

Analysis of glycosylated products Previous results showed that most of the products of the ADS and AMOLAP activity in *N. benthamiana* agro-infiltration are present as glycosylated conjugates, mainly of AA (van Herpen *et al.*, 2010). Therefore, leaf material was also analysed by LC-QTOF-MS. *Nicotiana benthamiana* leaves co-expressing *AmFH+AMOLAP* indeed contain AA-12-β-diglucoside, as previously reported (van Herpen *et al.*, 2010). However, in addition several other AA-glycoside conjugates were detected, including conjugates with additional hexose units as well as malonylated hexoses. Also for the glycosylated products the distribution over the entire product range was different between *AmFH+AMOLAP* and *AmFH+AMOHAP* (Fig. 5a, Table 2). Leaves infiltrated with *AmFH+AMOLAP* produced more AA conjugates, while leaves infiltrated with *AmFH+AMOHAP* produced more AAOH conjugates (Table 2, Fig. S3a). For both treatments several DHAAOH and DHAA conjugates with hexose and malonyl groups were also detected, but no DHAAA conjugates (Table 2). Table 2 shows the mass fragmentation profiles of the detected products and their putative identification. MS/MS analysis was used to further confirm product identity and an example of the identification of one of the DHAA-hexose conjugates is shown in Fig. S5.

In order to quantify the concentrations of glycosylated products, samples were treated with a mix of glycosidases (Viscozyme L) and deglycosylated products were quantified using GC-MS. Note that the Viscozyme treatment only cleaves hexose conjugates but not malonylated hexose conjugates (Fig. S4). Results show that leaves infiltrated with *AmFH+AMOLAP* contained *c.* 40 mg kg^{-1} FW of AA (consistent with the previously reported 39.5 mg kg^{-1} FW of AA; van Herpen *et al.*, 2010), while the sensitivity of the GC-MS was not sufficient to detect any AA in leaves infiltrated with *AmFH+AMOHAP* (Table 3). GC-MS analysis after Viscozyme treatment confirmed that AAOH was

Table 1 Unconjugated artemisinin precursors produced in *Nicotiana benthamiana* as identified and quantified by UPLC-MRM-MS

(ng kg^{-1} FW)	AmFH + AMOLAP	AmFH + AMOHAP	AmFH + AMOLAP + DBR2	AmFH + AMOHAP + DBR2
AAOH	16 709 ± 3977	46 946 ± 5692	3596 ± 1247	16 554 ± 3233
AAA	5501 ± 1486	19 564 ± 6314	781 ± 246	3418 ± 890
AA	3969 ± 1391	74 ± 15	53 ± 16	ND
DHAAOH	1821 ± 576	597 ± 19	87 972 ± 15 014	57 289 ± 9455
DHAAA[a]	(5966 ± 1646)	(8067 ± 1341)	(220 347 ± 65 373)	(124 787 ± 26 145)
DHAA	17 ± 8	ND	838 ± 517	25 ± 9

AAOH, artemisinic alcohol; AAA, artemisinic aldehyde; AA, artemisinic acid; DHAAOH, dihydroartemisinic alcohol; DHAAA, dihydroartemisinic aldehyde; DHAA, dihydroartemisinic acid; ND, not detectable.

[a]The values for DHAAA are shown in brackets as they represent peak intensities and not concentrations. Results are means ± SD of three co-infiltrated leaves.

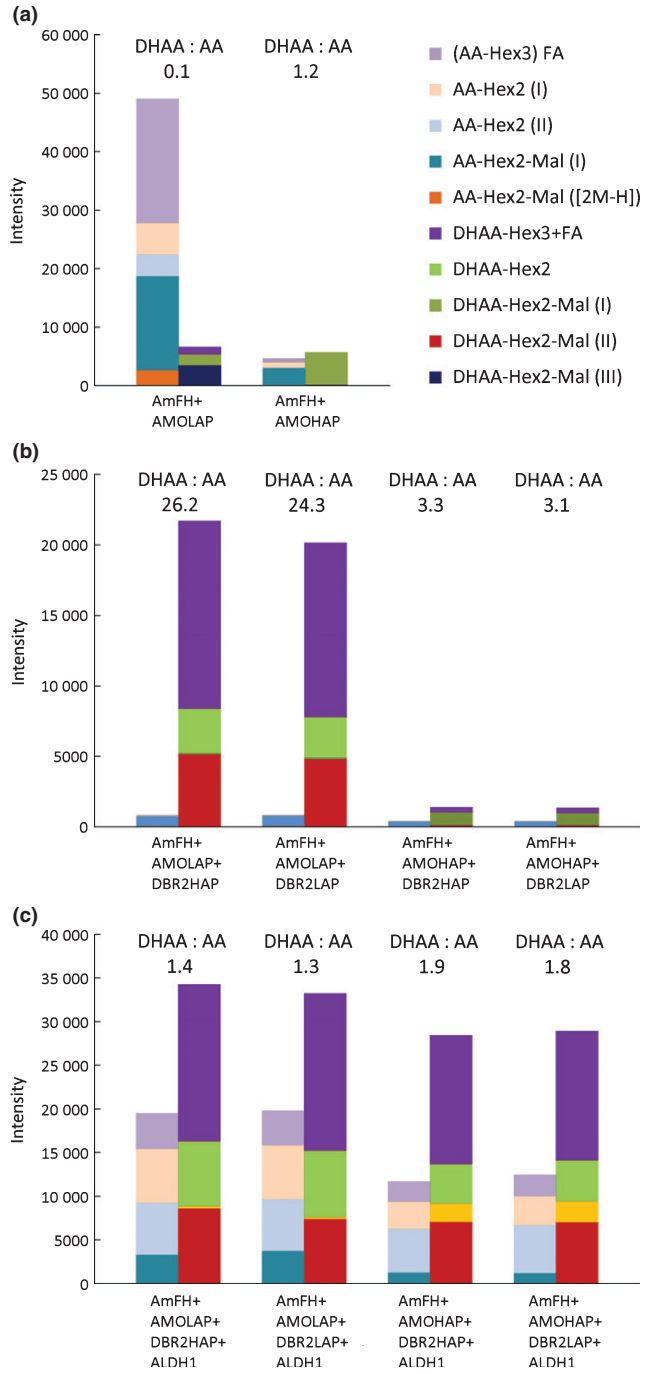

Fig. 5 Artemisinic acid (AA) and dihydroartemisinic acid (DHAA) related compounds in leaves of *Nicotiana benthamiana* agro-infiltrated with artemisinin biosynthesis genes as identified by LC-QTOF-MS. (a) Agro-infiltrated leaves with *AmFH* + *AMOLAP* and *AmFH* + *AMOHAP*. For all products identified by LC-QTOF-MS see Table 2. (b) Agro-infiltrated leaves with *AmFH* + *AMOLAP/HAP* + *DBR2LAP/HAP*. For all products identified by LC-QTOF-MS see Supporting Information Table S3. (c) Agro-infiltrated leaves with *AmFH* + *AMOLAP/HAP* + *DBR2LAP/HAP* + *ALDH1*. For all products identified by LC-QTOF-MS see Table S3. Data shown with peak intensity which analysed by LC-QTOF-MS. Peak intensities are the mean of three agro-infiltrated leaves.

the major glycosylated product in leaves infiltrated with *AmFH*+*AMOHAP* (as was suggested by Table 2) at 24 mg kg^{-1} FW.

No difference in DBR2 activity from HAP and LAP *A. annua* chemotypes

We compared the activity of the two variants of DBR2 by comparing product profiles of *N. benthamiana* leaves infiltrated with *AmFH*+*AMOLAP* in combination with either *DBR2HAP* or *DBR2LAP*. In addition we tested the two DBR2 variants in combination with *AmFH*+*AMOHAP*. Analysis of the conjugated products show that there is no difference in product profile between *DBR2HAP* and *DBR2LAP* (Figs 5b, S3b, Table S3), indicating that the two forms of DBR2 do not differ in enzymatic activity. The co-infiltration with *DBR2* relieved the necrosis symptoms caused by expression of *AmFH*+*AMOLAP* or *AmFH*+*AMOHAP* alone (Fig. S2), suggesting that additional DBR2 enzyme activity lowered the concentration(s) of the product(s) that cause necrosis. Product analysis in leaves agro-infiltrated with *AmFH*+*AMOLAP*+*DBR2* or *AmFH*+*AMOHAP*+*DBR2* showed that DBR2 activity resulted in a significant increase in DHAAOH, DHAAA and DHAA concentrations (Table 1) and this is also clear from LC-QTOF-MS analysis that shows a strong increase in DHAAOH and DHAA conjugates to hexose and malonyl groups (Table 2).

The analysis of deglycosylated extracts by GC-MS confirmed that co-expression of *DBR2* increased the concentrations of DHAAOH, DHAAA and DHAA at the expense of AAOH, AAA and AA concentrations (Table 3). The total yield of DHAA in leaves agro-infiltrated with *AmFH*+*AMOLAP*+*DBR2* as released by glycosidase treatment was *c.* 7.3 mg kg^{-1} FW, while DHAA in leaves agro-infiltrated with *AmFH*+*AMOHAP*+*DBR2* was below the level of detection by GC-MS. Combined, these results show that DBR2 further enhances the double bond reduction of the CYP71AV1 products that is also already catalysed by endogenous tobacco reductase activity. In addition, endogenous tobacco glycosyl and malonyl transferases modify these double-bond-reduced products which leads to DHAAOH and DHAA conjugates.

Increased DHAA and AA by combining *ADS*, *AMO* and *DBR2* with *ALDH1*

As already described, no differences were found between the ALDH1 protein sequence from LAP and HAP *A. annua* chemotypes. To test how the addition of *ALDH1* activity affects the product profile of the artemisinin HAP and LAP biosynthesis pathway, leaves were agro-infiltrated with *AmFH*+*AMOLAP*+*DBR2*+*ALDH1* or *AmFH*+*AMOHAP*+*DBR2*+*ALDH1*. After 7 d leaves were extracted and products were profiled by LC-QTOF-MS. The concentrations of conjugated DHAAOH products significantly decreased when *ALDH1* was added to *AmFH*+*AMOLAP*+*DBR2* (Table S3, Fig. S3b), coinciding with a substantial increase in glycosylated AA and DHAA product concentrations (Fig. 5c). This suggests that ALDH1 may be more efficient in the conversion of AAA to AA and DHAAA to DHAA than AMOLAP and AMOHAP as already suggested by the work of Teoh (Teoh *et al.*, 2009). Although the concentration of the presumed direct precursor of artemisinin (DHAA) was

New Phytologist (2013) 199: 352–366
www.newphytologist.com

Table 2 Conjugated artemisinin precursors produced in agro-infiltrated *Nicotiana benthamiana* leaf extracts fragmentation

Ret (min)	Detected mass (Da)[a]	MS/MS fragments	Mol form	ΔMass (ppm)	Putative ID	Intensity[b] AmFH + AMOLAP	AmFH + AMOHAP	AmFH + AMOLAP + DBR2	AmFH + AMOHAP + DBR2
27.85	543.2793	381[M-Hex-H]$^-$	$C_{27}H_{44}O_{11}$	1.3	AAOH-Hex2	319 ± 57	3715 ± 610	264 ± 51	1753 ± 823
28.87	629.2810	585[M-CO_2-H]$^-$,543[M-Mal-H]$^-$, 381[M-Mal-Hex-H]$^-$	$C_{30}H_{46}O_{14}$	0	AAOH-Hex2-Mal (I)[c]	1096 ± 288	2239 ± 124	861 ± 43	1378 ± 246
29.35	629.2810	585[M-CO_2-H]$^-$,543[M-Mal-H]$^-$, 381[M-Mal-Hex-H]$^-$	$C_{30}H_{46}O_{14}$	0	AAOH-Hex2-Mal (II)	606 ± 37	24 251 ± 76	627 ± 107	16 215 ± 5097
29.69	629.2810	585[M-CO_2-H]$^-$,543[M-Mal-H]$^-$, 381[M-Mal-Hex-H]$^-$	$C_{30}H_{46}O_{14}$	0	AAOH-Hex2-Mal (III)	439 ± 198	1109 ± 104	ND	515 ± 313
23.80	542.2536	272,254,210,179,143,128[d]	$C_{25}H_{41}N_3O_8S$	1.0	AAA-GSH-H_2O	2462 ± 323	7891 ± 65	130 ± 22	1970 ± 395
24.02	765.3181	719[M-H]$^-$,395[M-2Hex-H]$^-$, 233[M-3Hex-H]$^-$	$C_{34}H_{54}O_{19}$	0.1	(AA-Hex3) FA	21 190 ± 1730	612 ± 35	100 ± 80	ND
27.88	557.2598	395[M-Hex-H]$^-$,233[M-2Hex-H]$^-$	$C_{27}H_{42}O_{12}$	0	AA-Hex2 (I)[e]	5403 ± 1218	917 ± 71	121 ± 61	371 ± 186
28.48	557.2598	395[M-Hex-H]$^-$,233[M-2Hex-H]$^-$	$C_{27}H_{42}O_{12}$	0	AA-Hex2 (II)	3773 ± 882	110 ± 21	28 ± 15	ND
29.38	643.2602	599[M-CO_2-H]$^-$,395[M-Mal-Hex-H]$^-$, 233[M-Mal-2Hex-H]$^-$	$C_{30}H_{44}O_{15}$	0.5	AA-Hex2-Mal (I)	16 021 ± 320	2972 ± 162	1652 ± 327	1783 ± 139
29.78	1287.5282	643[M-H]$^-$,599[M-CO_2-H]$^-$, 395[M-Mal-Hex-H]$^-$, 233[M-Mal-2Hex-H]$^-$	$C_{60}H_{88}O_{30}$	4.5	AA-Hex2-Mal ([2M-H]$^-$)	2611 ± 1114	ND	129 ± 13	82 ± 9
24.86	753.3545	707[M-H]$^-$,545[M-Hex-H]$^-$, 383[M-2Hex-H]$^-$,221[M-3Hex-H]$^-$	$C_{34}H_{58}O_{18}$	0.9	(DHAAOH-Hex3) FA	145 ± 46	85 ± 8	14 248 ± 670	11 573 ± 206
28.80	545.2962	383[M-Hex-H]$^-$,221[M-2Hex-H]$^-$	$C_{27}H_{46}O_{11}$	1.8	DHAAOH-Hex2 (I)	343 ± 89	194 ± 46	14 321 ± 2617	14 968 ± 5018
30.27	631.2966	587[M-CO_2-H]$^-$,545[M-Mal-H]$^-$, 383[M-Hex-H]$^-$,221[M-2Hex-H]$^-$	$C_{30}H_{48}O_{14}$	1.8	DHAAOH-Hex2-Mal (I)	479 ± 116	266 ± 12	24 475 ± 120	24 394 ± 181
30.75	631.2966	587[M-CO_2-H]$^-$,545[M-Mal-H]$^-$, 383[M-Mal-Hex-H]$^-$, 221[M-Mal-2Hex-H]$^-$	$C_{30}H_{48}O_{14}$	0	DHAAOH-Hex2-Mal(II)	ND	18 ± 2	855 ± 156	919 ± 378
23.87	767.3338	721[M-H]$^-$,397[M-2Hex-H]$^-$, 235[M-3Hex-H]$^-$	$C_{34}H_{56}O_{19}$	1.6	(DHAA-Hex3) FA	1237 ± 110	ND	23 356 ± 1554	861 ± 185
27.28	559.2755	397[M-Hex-H]$^-$,235[M-2Hex-H]$^-$	$C_{27}H_{44}O_{12}$	0.3	DHAA-Hex2	ND	ND	5188 ± 2039	120 ± 54
26.58	645.2759	601[M-CO_2-H]$^-$,397[M-Mal-Hex-H]$^-$, 235[M-Mal-2Hex-H]$^-$	$C_{30}H_{46}O_{15}$	0	DHAA-Hex2-Mal (I)	1912 ± 244	5539 ± 174	236 ± 14	1645 ± 256
28.92	645.2759	601[M-CO_2-H]$^-$,397[M-Mal-Hex-H]$^-$, 235[M-Mal-2Hex-H]$^-$	$C_{30}H_{46}O_{15}$	0	DHAA-Hex2-Mal (II)	ND	20 ± 6	1303 ± 528	76 ± 35
29.83	645.2759	601[M-CO_2-H]$^-$,397[M-Mal-Hex-H]$^-$, 235[M-Mal-2Hex-H]$^-$	$C_{30}H_{46}O_{15}$	0	DHAA-Hex2-Mal (III)	3424 ± 549	64 ± 8	ND	ND

AAOH, artemisinic alcohol; AAA, artemisinic aldehyde; AA, artemisinic acid; DHAAOH, dihydroartemisinic alcohol; DHAAA, dihydroartemisinic aldehyde; DHAA, dihydroartemisinic acid; GSH, glutathione; FA, formic acid adduct; Ret (min), retention time, in minutes; Mol form, molecular formula of the metabolite; ΔMass (ppm), deviation between the detected mass and real accurate mass, in ppm; Putative ID, putative identification of metabolite; ND, not detectable.

[a]Detected mass (Da): The mass was detected in negative mode of LC-QTOF-MS.

[b]Peak intensities are the mean ± SD of three agro-infiltrated leaves.

[c]Hex, compound conjugated with hexose; Mal, compound conjugated with malonate; (I–III), different isobaric forms (i.e. identical accurate mass, but different retention times).

[d]The ions of a number of representative GSH adducts in the negative ion mode (Dieckhaus *et al.*, 2005).

[e]AA-Hex2, The structure of artemisinic acid-12-β-diglucoside was confirmed by NMR (van Herpen *et al.*, 2010).

Nonvolatile metabolites with mass intensity > 500 in LC-QTOF-MS, which were significantly increased in leaves agro-infiltrated with *AmFH + AMOLAP*, *AmFH + AMOHAP*, *AmFH + AMOLAP + DBR2*, or *AmFH + AMOHAP + DBR2* were targeted for analysis by LC-QTOF-MS/MS.

Table 3 Artemisinin precursors in *Nicotiana benthamiana* agro-infiltrated with artemisinin biosynthetic pathway genes

(mg kg^{-1} FW)	AmFH + AMOLAP	AmFH + AMOHAP	AmFH + AMOLAP + DBR2	AmFH + AMOHAP + DBR2
AAOH	8.1 ± 1.6	24.0 ± 3.5	5.1 ± 0.5	9.4 ± 2.1
AAA	1.6 ± 0.1	1.6 ± 0.1	ND	ND
AA	39.9 ± 9.8	ND	ND	ND
DHAAOH	1.6 ± 0.2	2.0 ± 0.2	42.9 ± 14.9	22.8 ± 8.6
DHAAA	ND	ND	4.0 ± 2.3	1.3 ± 0.6
DHAA	ND	ND	7.3 ± 2.2	ND

Agro-infiltrated leaves were treated with glycosidase (Viscozyme L.) and hydrolysed metabolites extracted and analysed by GC-MS.
AAOH, artemisinic alcohol; AAA, artemisinic aldehyde; AA, artemisinic acid; DHAAOH, dihydroartemisinic alcohol; DHAAA, dihydroartemisinic aldehyde; DHAA, dihydroartemisinic acid; ND, not detectable.
Results are means ± SD of three co-infiltrated leaves.

Table 4 Artemisinic acid (AA) and dihydroartemisinic acid (DHAA) produced in agro-infiltrated *Nicotiana benthamiana* as identified and quantified by UPLC-MRM-MS

(ng kg^{-1} FW)	AmFH + AMOLAP + DBR2 + ALDH1	AmFH + AMOHAP + DBR2 + ALDH1
AA	8836 ± 1730	2589 ± 563
DHAA	10 792 ± 341	2756 ± 547

Results are means ± SD of three co-infiltrated leaves.

substantially increased by ALDH1 (*c.* 13-fold by adding *ALDH1* to *AmFH+ AMOLAP+ DBR2* and *c.* 110-fold by adding *ALDH1* to *AmFH+ AMOHAP+ DBR2*), no artemisinin could be detected in *N. benthamiana* by UPLC-MRM-MS (Tables 1,4).

Qualitative effects on product profile by *AMOLAP* dosage

The comparison of the total product concentrations produced by the reconstituted pathway with AMOLAP or AMOHAP suggests that AMOLAP has a higher enzyme activity than AMOHAP, in combination with a different product profile (Fig. 5a, Table 2). This difference could not be related to different subcellular localization (Fig. 4). To test if differences in relative enzyme activity within the pathway can affect the product profile in a qualitative way, we tested the effect of different dilutions of AMOLAP in combination with the rest of the biosynthesis pathway (*ADS+ AMOLAP+ DBR2HAP*). This was achieved by diluting the *Agrobacterium* strain carrying the *AMOLAP* expression construct with a suspension of an *Agrobacterium* strain carrying an empty expression vector to keep the total *Agrobacterium* dosage for infiltration the same. Results show that reduction of the AMOLAP agro-infiltration dosage to 1/2 had little effect, but that dilution up to 1/10 strongly decreased DHAA-glycoside production (Figs 6a, S3c; Table S4). The AA-glycoside concentrations were also reduced but to a much lower extent.

More AAA-related glutathione-conjugate from *AMOHAP* than from *AMOLAP*

Recently we described the reconstruction of the biosynthetic pathway of costunolide in *N. benthamiana* (Liu *et al.*, 2011) in which it was shown that the exocyclic carbon double bond of costunolide conjugates to glutathione (GSH). Because some of the products of ADS (AD) and AMOLAP/HAP enzyme activities also contain such an exocyclic double bond (AAA), but lack the hydroxyl or acid group used for glycosylation in AAOH and AA, we specifically looked for GSH conjugates of artemisinin intermediates in *N. benthamiana* leaves agro-infiltrated with the biosynthetic pathway genes. A putative GSH-conjugated compounds was detected by LC-QTOF-MS, both in negative mode ($m/z = 542.25$) and positive mode ($m/z = 544.24$; Fig. S6). The concentration of the GSH-conjugate ($m/z = 542.25$) was higher in *AmFH+ AMOHAP* than in *AmFH+ AMOHAP+ DBR2* agro-infiltrated leaves (Table 2). We tested the artemisinin biosynthetic pathway intermediates (AAA, AAOH, DHAAOH, AA, DHAA and DHAAA) in *in vitro* reactions for spontaneous or glutathione-S-transferase (GST) driven GSH conjugation. Only AAA formed an AAA–GSH conjugate similar to that extracted from agro-infiltrated leaves expressing the pathway genes ($m/z = 526.22$; Figs S6, S7). The mass of the major GSH conjugate formed *in planta* is 18 Da higher, suggesting an additional two protons and one oxygen atom, which could be explained by hydroxylation of the endocyclic double bond in AAA (Fig. S6). The concentration of the putative AAA glutathione conjugate was higher in leaves infiltrated with *AmFH+ AMOHAP* than in leaves infiltrated with *AmFH+ AMOLAP* (Table 2). Dilution of AMOLAP resulted in an increase in the concentration of the AAA glutathione conjugate, although not reaching the level of the full dosage of AMOHAP (Fig. 7).

Effects on product profile by *DBR2* or *ALDH1* dosage

We also tested the effect of *DBR2* and *ALDH1* gene dosage on the DHAA : AA related product ratio by testing dilutions of DBR2 and ALDH1 in combination with AMOLAP or AMOHAP. Product analysis by LC-QTOF-MS of glycosylated products showed that with full DBR2 and ALDH1 agro-infiltration dosage the ratio of DHAA : AA was more skewed towards the DHAA branch of the pathway. However, when the DBR2 infiltration dosage was diluted 10-fold, the ratio of DHAA : AA decreased, shifting the pathway activity more towards the AA branch, with relatively little effect on the total product concentration (Figs 1, 6b).

By contrast, 10-fold dilution of ALDH1 agro-infiltration dosage resulted in a strong decrease in glycosylated AA conjugates,

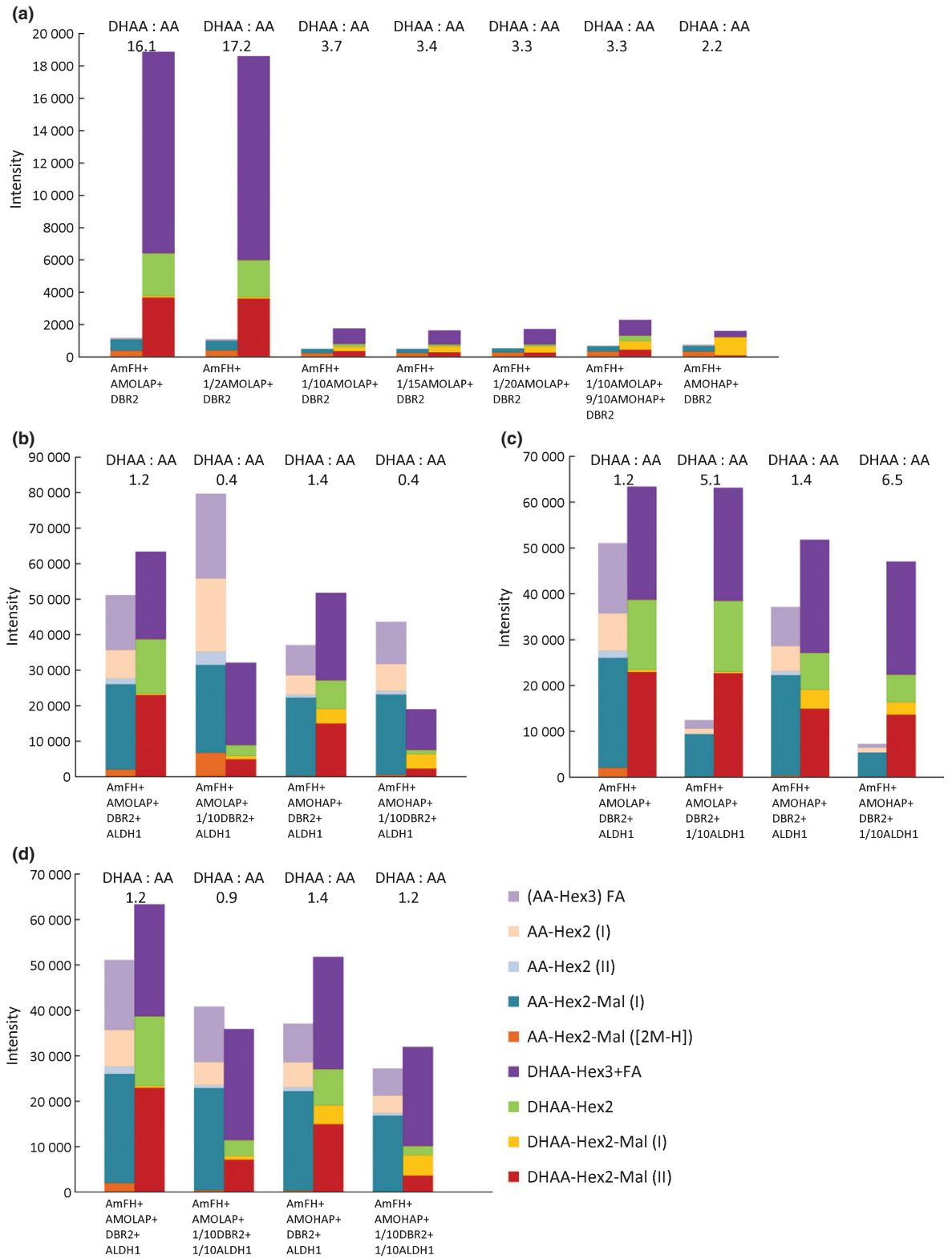

Fig. 6 Artemisinic acid (AA) and dihydroartemisinic acid (DHAA) conjugated compounds in leaves of *Nicotiana benthamiana* agro-infiltrated with AMOLAP/HAP and different dosage of DBR2 and ALDH1. (a) Comparison AA and DHAA conjugated compounds in agro-infiltrated leaves with dilution of AMOLAP/HAP. For all products identified by LC-QTOF-MS see Supporting Information Table S4. (b) Comparison AA and DHAA conjugated compounds in agro-infiltrated leaves with dilution of DBR2. For all products identified by LC-QTOF-MS, see Table S5. (c) Comparison AA and DHAA conjugated compounds in agro-infiltrated leaves with dilution of ALDH1. For all products identified by LC-QTOF-MS see Table S5. (d) Comparison AA and DHAA conjugated compounds in agro-infiltrated leaves with dilution of DBR2 and ALDH1. For all products identified by LC-QTOF-MS see Table S5. Data shown with peak intensity which analysed by LC-QTOF-MS. Data represent peak intensities for each of the compounds in LC-QTOF-MS analysis.

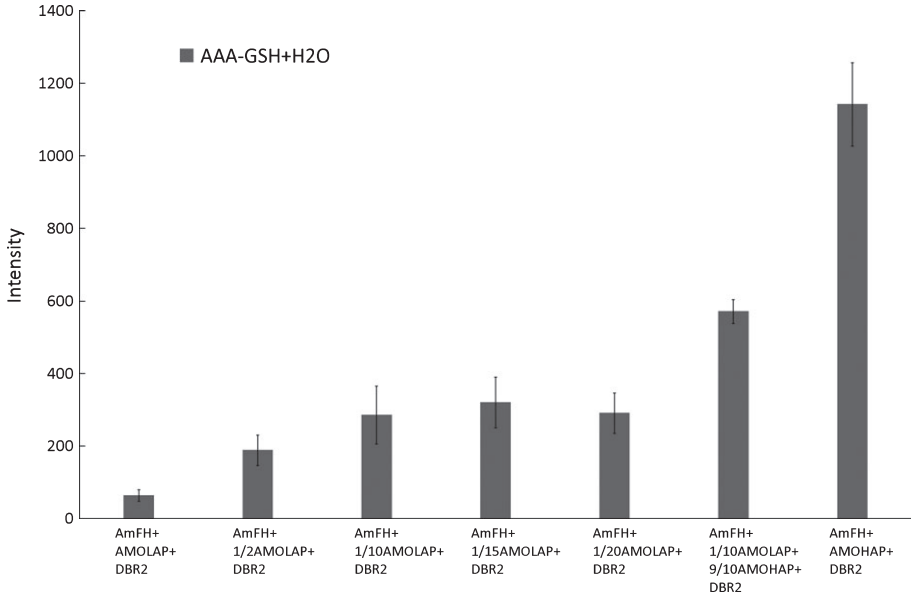

Fig. 7 Artemisinic aldehyde (AAA) conjugated compounds in leaves of *Nicotiana benthamiana* agro-infiltrated with dilution of AMOLAP/HAP. Data shown with peak intensity which analysed by LC-QTOF-MS. Data represent peak intensities for each of the compounds in LC-QTOF-MS analysis.

while the glycosylated DHAA conjugates were hardly affected (Fig. 6c), suggesting that ALDH1 has a preference for the DHAA substrate over the AAA substrate (Fig. 1).

When both DBR2 and ALDH1 agro-infiltration dosage were diluted 10-fold, the DHAA : AA related products ratio came close to one, with slight preference for the DHAA branch of the pathway with AMOHAP and slight preference for the AA branch of the pathway with AMOLAP (Fig. 6d). In combination, these results suggest that the agro-infiltrated leaf 'chemotype' is determined by a combination of the AMOLAP/AMOHAP catalytic effectivity in combination with especially *DBR2* gene dosage (Figs 6, S3d; Table S5).

Discussion

Here we have compared the different proteins from the artemisinin biosynthesis pathway encoded by genes isolated from high- and low-artemisinin producing *A. annua* chemotypes. The expression concentrations of artemisinin biosynthesis genes (*ADS*, *AMO*, *DBR2* and *ALDH1*) were recently analysed in *A. annua* HAP and LAP chemotypes showing that the chemotype identities of *A. annua* did not correlate with differences in expression level of these genes in the absence of stress (Maes *et al.*, 2011). Nevertheless, total product yield in 5-wk-old *A. annua* leaves is higher in LAP than HAP chemotypes (see Table S6, Maes *et al.*, 2011). Therefore, differences in protein activity rather than differences in gene expression level may account for differences in chemotype. In the present work, two *DBR2* variants were identified (Fig. 3) but characterization of the *in planta* activity showed that they have equal activity (Figs 5b, S3b). No difference was found for the ALDH1 amino acid sequence from *A. annua* HAP and LAP chemotypes. Therefore, the only consistent difference in artemisinin biosynthesis proteins in the branched pathway after ADS is in the AMOHAP and AMOLAP from the *A. annua* HAP and LAP chemotypes, respectively (Fig. 2).

In the LAP chemotype CYP71AV1 is consistently seven amino acids longer than CYP71AV1 from the HAP chemotype (Fig. 2). When expressed together with *ADS* (+*FPS*+ *HMGR*) the total product yield (based on the cumulative concentrations of AAOH, AAA, AA, DHAAOH and DHAA released from conjugated products) in leaves co-infiltrated with *AMOLAP* was approximately twice as high as in leaves co-infiltrated with *AMOHAP* (Table 3). This suggests a lower enzyme activity for AMOHAP than for AMOLAP (Fig. 5a). AMOHAP and AMOLAP seem to anchor equally well to the ER membrane (Fig. 4). The observed difference in efficiency may be caused by a different stability of the two proteins. Nevertheless, dilution of AMOLAP gene dosage in agro-infiltration experiments did not fully mimic the AMOHAP phenotype (Figs 6a, S3c). In theory, the lower efficiency of AMOHAP could also be due to a less efficient interaction with cytochrome P450 reductase (CPR; Lengler *et al.*, 2006). Alignment of the AMOLAP and AMOHAP protein sequences with other related sesquiterpene oxidases like germacrene A oxidase (GAO) from several different Asteraceae (Nguyen *et al.*, 2010) shows that the N-terminal extension in AMOLAP is the exception. However, the CPR interaction domain does not map to the N-terminus (Sevrioukova *et al.*, 1999). Finally, it may be that the substrate entry or release of AMOHAP is compromised.

Variable 'chemotype' of leaves expressing artemisinin biosynthesis genes

We defined the 'chemotype' of the *N. benthamiana* agro-infiltrated leaves based on the DHAA and AA glycoside conjugates (Fig. 5). We note that the peak intensity is a relative quantification as detection efficiency (e.g. ionization) may differ between compounds. However, comparison of the relative quantifications based on peak intensities from LC-QTOF-MS and the absolute quantification of DHAA and AA products by UPLC-MRM-MS or GC-MS (Tables 1–3) indicates a good correlation between the two analytical techniques. Expression of *ADS*+ *AMOHAP* results

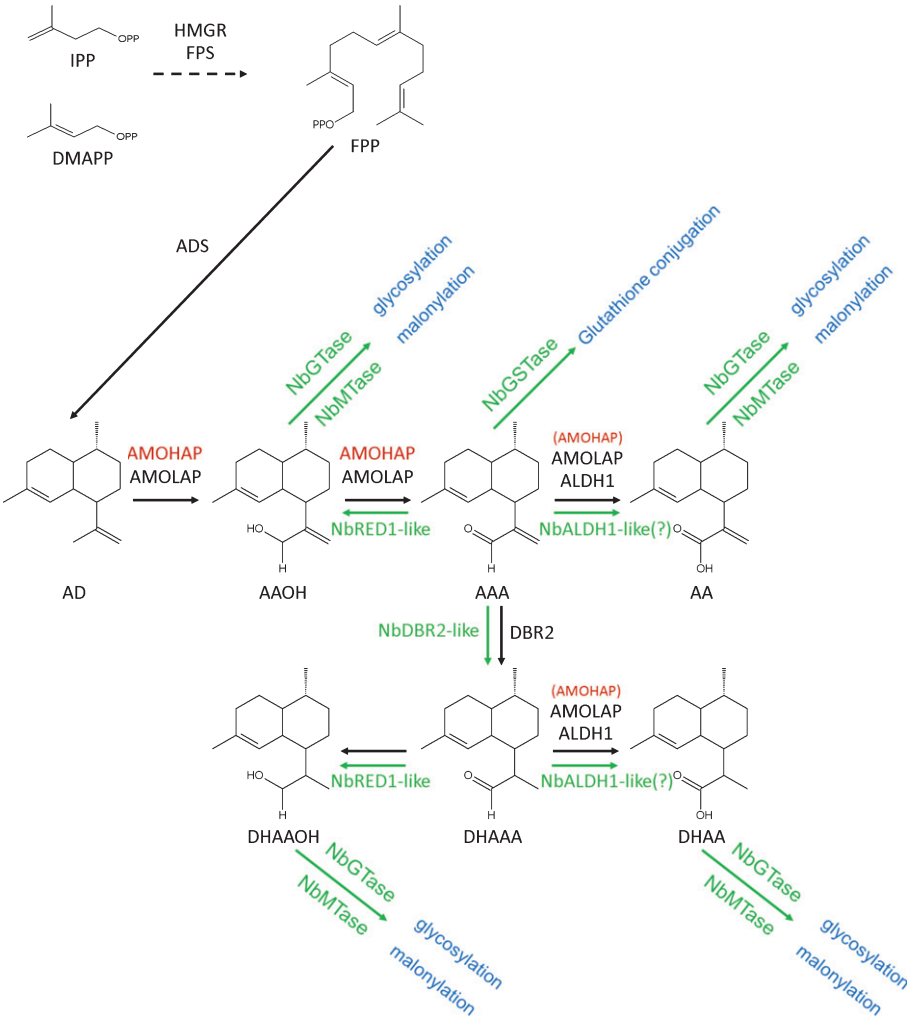

Fig. 8 Schematic representation of the artemisinin precursor biosynthetic pathway in *N. benthamiana*. NbDBR2-like, *N. benthamiana* artemisinic aldehyde double-bond reductase-like enzyme; NbRED 1-like, *N. benthamiana* dihydroartemisinic aldehyde reductase 1-like enzyme; NbALDH1-like, *N. benthamiana* aldehyde dehydrogenase 1-like enzyme; NbGTase, *N. benthamiana* glycosyl transferase; NbMTase, *N. benthamiana* malonyl transferase; NbGSTase, *N. benthamiana* glutathione-S-transferase; Enzyme names in brackets indicates low activity of the enzyme for the corresponding step.

in a HAP chemotype (more DHAA than AA) and expression of *ADS*+*AMOLAP* in a LAP chemotype (more AA than DHAA; Fig. 5a). However, when *DBR2* is included, the chemotype for both the combination *ADS*+*AMOHAP*+*DBR2* and *ADS*+*AMOLAP*+*DBR2* is changed to a HAP chemotype with relatively higher DHAA concentration (Fig. 5b). In all of these combinations, the overall yield of DHAA was always lower for *AMOHAP* (*c.* 15-fold). When *ALDH1* was included, the chemotype remained that of HAP, but the relative yield of the gene combinations that include AMOHAP was substantially increased and was now comparable to that of the gene combinations with AMOLAP (Fig. 5c). Because in the agro-infiltration assay in *N. benthamiana* addition of new genes to the pathway leads to a reduction in the relative dosage of the other genes infiltrated into the leaf, we tested whether the relative gene dosage affects the product profile. Lowering the dosage of AMOLAP with similar dosage of ADS and DBR2 resulted in a profile more closely related to that of AMOHAP, but also resulted in lower product yield of AA and DHAA conjugates (Fig. 6a). Lowering the relative dosage of DBR2 resulted in a reversion of the infiltrated leaf chemotype from HAP to LAP (Fig. 6b), while lowering the relative dosage of ALDH1 did not change the chemotype, but did decrease AA conjugates more than the DHAA conjugates (Fig. 6c). When both DBR2 and ALDH1 dosage were reduced, AMOHAP resulted in a more HAP related chemotype, while the combination with AMOLAP resulted in a more LAP related chemotype (Fig. 6d).

Nicotiana benthamiana enzyme activities both enhance and limit the DHAA 'chemotype'

AMOHAP seems to be less efficient in the conversion to AA, presumably resulting in an early release of AAA. Indeed free AAA and an AAA glutathione conjugate were present at higher concentrations when the pathway was expressed in combination with AMOHAP than with AMOLAP (Table 1, Fig. 7). The AAA released by AMOHAP, in particular, may subsequently be substrate for double-bond reductases (endogenous from *N. benthamiana* or the co-expressed *DBR2*) to produce DHAAA and DHAAA derived products (Tables 1–3). The endogenous DBR2-like activity is far from saturating, as introduction of *A. annua DBR2* greatly enhanced the conversion to DHAAOH, DHAAA and DHAA (Table 1). The early release of AAA by AMOHAP also reveals the activity of an endogenous *N. benthamiana* reductase, similar to the *A. annua* aldehyde reductase (RED1) that catalyses the conversion of DHAAA to DHAAOH (Bertea *et al.*, 2005; Rydén *et al.*, 2010). Indeed, the presence of a RED1-

like activity in tobacco was previously demonstrated through feeding experiments: tobacco leaves supplied with DHAAA and AAA form DHAAOH and AAOH, respectively (Zhang *et al.*, 2011). This suggests that in *N. benthamiana* the elevated pools of AAOH and DHAAOH (and glycosylated derivatives) may be the result of a reverse product flux from AAA back to AAOH and DHAAA to DHAAOH (Fig. 1). If this is the case, the affinity of AMOHAP for the AAOH substrate may be underestimated.

DHAA (conjugates) were detected in the leaves infiltrated with *AmFH+AMOLAP*. This suggests the presence of an endogenous aldehyde dehydrogenase, similar to ALDH1 from *A. annua*, which can produce DHAA from DHAAA (Teoh *et al.*, 2009). Alternatively, the low concentrations of DHAA could be the result of AMOLAP catalysing oxidation of DHAAA. Work in yeast showed that AMOLAP is far less effective in the conversion of DHAAA to DHAA than in the conversion of AAA to AA (Teoh *et al.*, 2009). If we assume that also AMOHAP catalyses this step less effectively, expression of *ALDH1* in a HAP background should further enhance the production of DHAA in the HAP chemotype. Indeed, our data show that product flow towards DHAA increased *c.* 110-fold when expression of *AmFH+AMOHAP+DBR2* was combined with *ALDH1* (Table 4).

Conjugating activities limit precursor pool for artemisinin production

Most of the products formed upon agro-infiltration of the artemisinin pathway genes are present in the form of hexose and/or hexose/malonyl conjugates. This explains why a previous heterologous expression study in tobacco detected only limited concentrations of free AD, AAOH, DHAAOH and no AA or DHAA (Zhang *et al.*, 2011). The deglycosylation experiments show that the conjugated forms accumulate to concentrations up to 4–10-fold higher than the corresponding free forms (Tables 1–3). The glycosylation/conjugation is most likely a response to the production of potentially toxic compounds, as leaves expressing *AmFH+AMOLAP/HAP* showed signs of necrosis (Fig. S2). Necrosis was stronger in leaves containing higher concentrations of free (and conjugated) AA and indeed presence of *DBR2* reduced both the necrotic phenotype and the free and conjugated AA concentrations (Fig. S2, Table 1). In addition to the AA/DHAA glycosides, AAA glutathione conjugates were detected. All these conjugating activities limit accumulation of DHAA, the direct precursor of artemisinin (Fig. 8). Interestingly, in extracts from *A. annua* HAP flowers, no glycosides of AAOH, AA, DHAAOH and DHAA or glutathione conjugates of AAA were detected (Fig. S8). This indicates that the cells of *A. annua* that produce artemisinin either do not have competing glycosyl/malonyl/GSH transferase activity or, perhaps more likely, product flux within them is protected from such competing activities.

Concluding remarks

Our results show that the chemical profile of agro-infiltrated *N. benthamiana* leaves is a function of both type and relative dosage of the expression constructs. Results still do not fully explain the difference in HAP and LAP chemotypes found in *A. annua*. The expression level of the different biosynthesis genes do not differ between *A. annua* chemotypes (Maes *et al.*, 2011) and therefore, at present, the difference in CYP71AV1 catalytic efficiency between HAP and LAP is the only identified factor that contributes to this difference in chemotype. However, other, as yet unidentified, factors in *A. annua* may further contribute to the chemical difference in the HAP and LAP varieties.

Acknowledgements

H-M.T. was funded by the graduate school of Experimental Plant Sciences (EPS). T.v.H. and J.B. were supported through the NWO-CW/ACTS IBOS programme (053.63.305) which was co-sponsored by Dafra Pharma, Turnhout, Belgium. We thank Ric de Vos for helpful discussions on LC-QTOF-MS/MS data, Bert Schipper for assistance in LC-QTOF-MS analysis, Desalegn Woldes Etalo, Ting Yang and Lemeng Dong for help in MS data analysis, and Qing Liu for support in the glutathione conjugation assay. We also would like to acknowledge Sajad Rashidi Manfard and Peter E. Brodelius (Linneaeus University, Sweden) for providing seeds of an *A. annua* LAP chemotype.

References

Aker J, Borst JW, Karlova R, de Vries S. 2006. The *Arabidopsis thaliana* AAA protein CDC48A interacts *in vivo* with the somatic embryogenesis receptor-like kinase 1 receptor at the plasma membrane. *Journal of Structural Biology* **156**: 62–71.

Bertea CM, Freije JR, van der Woude H, Verstappen FW, Perk L, Marquez V, De Kraker JW, Posthumus MA, Jansen BJ, de Groot A *et al.* 2005. Identification of intermediates and enzymes involved in the early steps of artemisinin biosynthesis in *Artemisia annua*. *Planta Medica* **71**: 40–47.

Bertea CM, Voster A, Verstappen FW, Maffei M, Beekwilder J, Bouwmeester HJ. 2006. Isoprenoid biosynthesis in *Artemisia annua*: cloning and heterologous expression of a germacrene A synthase from a glandular trichome cDNA library. *Archives of Biochemistry and Biophysics* **448**: 3–12.

Bouwmeester HJ, Wallaart TE, Janssen MH, van Loo B, Jansen BJ, Posthumus MA, Schmidt CO, De Kraker JW, Konig WA, Franssen MC. 1999. Amorpha-4,11-diene synthase catalyses the first probable step in artemisinin biosynthesis. *Phytochemistry* **52**: 843–854.

De Vos R, Moco S, Lommen A, Keurentjes J, Bino RJ, Hall RD. 2007. Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nature Protocols* **2**: 778–791.

Dieckhaus CM, Fernández-Metzler CL, King R, Krolikowski PH, Baillie TA. 2005. Negative ion tandem mass spectrometry for the detection of glutathione conjugates. *Chemical Research in Toxicology* **18**: 630–638.

van Engelen FA, Molthoff JW, Conner AJ, Nap JP, Pereira A, Stiekema WJ. 1995. pBINPLUS: an improved plant transformation vector based on pBIN19. *Transgenic Research* **4**: 288–290.

van Herpen TW, Cankar K, Nogueira M, Bosch D, Bouwmeester HJ, Beekwilder J. 2010. *Nicotiana benthamiana* as a production platform for artemisinin precursors. *PLoS ONE* **5**: e14222.

Kim NC, Kim JG, Kim S, Lim H, Hahm T. 1992. Production of secondary metabolites by tissue culture of *Artemisia annua* L. *Journal of the Korean Agricultural Chemical Society* **35**: 99–105.

Kohlen W, Charnikhova T, Liu Q, Bours R, Domagalska MA, Beguerie S, Verstappen F, Leyser O, Bouwmeester H, Ruyter-Spira C. 2011. Strigolactones are transported through the xylem and play a key role in shoot architectural response to phosphate deficiency in nonarbuscular mycorrhizal host Arabidopsis. *Plant Physiology* **155**: 974–987.

Lengler J, Omann M, Duvier D, Holzmuller H, Gregor W, Salmons B, Gunzburg WH, Renner M. 2006. Cytochrome P450 reductase dependent

inhibition of cytochrome P450 2B1 activity: implications for gene directed enzyme prodrug therapy. *Biochemical Pharmacology* **72**: 893–901.

Liu Q, Majdi M, Cankar K, Goedbloed M, Charnikhova T, Verstappen FWA, de Vos RCH, Beekwilder J, van der Krol S, Bouwmeester HJ. 2011. Reconstitution of the costunolide biosynthetic pathway in yeast and *Nicotiana benthamiana*. *PLoS ONE* **6**: e23255.

Maes L, Van Nieuwerburgh F, Zhang Y, Reed D, Pollier J, Vande Casteele S, Inze D, Covello P, Deforce DLDL, Goossens A. 2011. Dissection of the phytohormonal regulation of trichome formation and biosynthesis of the antimalarial compound artemisinin in *Artemisia annua* plants. *New Phytologist* **189**: 176–189.

Mercke P, Bengtsson M, Bouwmeester HJ, Posthumus MA, Brodelius PE. 2000. Molecular cloning, expression, and characterization of amorpha-4,11-diene synthase, a key enzyme of artemisinin biosynthesis in *Artemisia annua* L. *Archives of Biochemistry and Biophysics* **381**: 173–180.

Nguyen KT, Arsenault PR, Weathers PJ. 2011. Trichomes + roots + ROS = artemisinin: regulating artemisinin biosynthesis in *Artemisia annua* L. *In Vitro Cellular & Developmental Biology Plant* **47**: 329–338.

Nguyen DT, Göpfert JC, Ikezawa N, MacNevin G, Kathiresan M, Conrad J, Spring O, Ro DK. 2010. Biochemical conservation and evolution of germacrene A oxidase in Asteraceae. *Journal of Biological Chemistry* **285**: 16 588–16 598.

Pompon D, Louerat B, Bronine A, Urban P. 1996. Yeast expression of animal and plant P450s in optimized redox environments. *Methods in Enzymology* **272**: 51–64.

Ro DK, Paradise EM, Ouellet M, Fisher KJ, Newman KL, Ndungu JM, Ho KA, Eachus RA, Ham TS, Kirby J *et al.* 2006. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**: 940–943.

Rydén AM, Ruyter-Spira C, Quax WJ, Osada H, Muranaka T, Kayser O, Bouwmeester H. 2010. The molecular cloning of dihydroartemisinic aldehyde reductase and its implication in artemisinin biosynthesis in *Artemisia annua*. *Planta Medica* **76**: 1778–1783.

Sevrioukova IF, Li H, Zhang H, Peterson JA, Poulos TL. 1999. Structure of a cytochrome P450–redox partner electron-transfer complex. *Proceedings of the National Academy of Sciences, USA* **96**: 1863–1868.

Sy L-K, Brown GD. 2002. The mechanism of the spontaneous autoxidation of dihydroartemisinic acid. *Tetrahedron* **58**: 897–908.

Teoh KH, Polichuk DR, Reed DW, Covello PS. 2009. Molecular cloning of an aldehyde dehydrogenase implicated in artemisinin biosynthesis in *Artemisia annua*. *Botany* **87**: 635–642.

Teoh KH, Polichuk DR, Reed DW, Nowak G, Covello PS. 2006. *Artemisia annua* L. (Asteraceae) trichome-specific cDNAs reveal CYP71AV1, a cytochrome P450 with a key role in the biosynthesis of the antimalarial sesquiterpene lactone artemisinin. *FEBS Letters* **580**: 1411–1416.

Tikunov Y, Lommen A, De Vos C, Verhoeven HA, Bino RJ, Hall RD, Bovy AG. 2005. A novel approach for nontargeted data analysis for metabolomics. Large-scale profiling of tomato fruit volatiles. *Plant Physiology* **139**: 1125–1137.

Wallaart TE, Pras N, Beekman AC, Quax WJ. 2000. Seasonal variation of artemisinin and its biosynthetic precursors in plants of *Artemisia annua* of different geographical origin: proof for the existence of chemotypes. *Planta Medica* **66**: 57–62.

Wallaart TE, van Uden W, Lubberink HG, Woerdenbag HJ, Pras N, Quax WJ. 1999. Isolation and identification of dihydroartemisinic acid from *Artemisia annua* and its possible role in the biosynthesis of artemisinin. *Journal of Natural Products* **62**: 430–433.

Wang Y, Yang K, Jing F, Li M, Deng T, Huang R, Wang B, Wang G, Sun X, Tang KX. 2011. Cloning and characterization of trichome-specific promoter of *cpr71av1* gene involved in artemisinin biosynthesis in *Artemisia annua* L. *Molecular Biology* **45**: 751–758.

Yoo SD, Cho YH, Sheen J. 2007. Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nature Protocols* **2**: 1565–1572.

Zhang Y, Nowak G, Reed DW, Covello PS. 2011. The production of artemisinin precursors in tobacco. *Plant Biotechnology Journal* **9**: 445–454.

Zhang Y, Teoh KH, Reed DW, Maes L, Goossens A, Olson DJ, Ross AR, Covello PS. 2008. The molecular cloning of artemisinic aldehyde Δ11(13) reductase and its role in glandular trichome-dependent biosynthesis of artemisinin in *Artemisia annua*. *Journal of Biological Chemistry* **283**: 21 501–21 508.

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 Alignment of the 5′RACE sequences of *AMOHAP* in Clustal W.

Fig. S2 Necrotic symptom in *Nicotiana benthamiana*.

Fig. S3 All conjugated compounds grouped according to free precursor detected in agro-infiltrated *Nicotiana benthamiana*.

Fig. S4 Malonylated glycosylated compounds are not cleaved by Viscozyme L.

Fig. S5 MS/MS spectrum and MS/MS fragmentation of DHAA-Hex3.

Fig. S6 Comparison of *in vitro* and *in planta* formed AAA-GSH conjugates.

Fig. S7 *In vitro* conjugation of AAA with glutathione.

Fig. S8 Absence of artemisinin precursor conjugates in *Artemisia annua* HAP chemotype.

Table S1 Primers used in this study

Table S2 Unconjugated artemisinin precursors as identified and quantified by UPLC-MRM-MS in *Artemisia anuua*

Table S3 Peak intensity of conjugated artemisinin precursors produced in agro-infiltrated *Nicotiana benthamiana* leaves with *AmFH + AMOLAP/HAP + DBR2LAP/HAP* and *AmFH + AMOLAP/HAP + DBR2LAP/HAP + ALDH1*

Table S4 Peak intensity of conjugated artemisinin precursors produced in agro-infiltrated *Nicotiana benthamiana* leaves with *AmFH + AMOLAP/HAP + DBR2* compared with different dilution of *AMOLAP/HAP*

Table S5 Peak intensity of conjugated artemisinin precursors produced in agro-infiltrated *Nicotiana benthamiana* leaves with *AmFH + AMOLAP/HAP + DBR2 + ALDH1* with different dosage of *DBR2* and *ALDH1*

Table S6 Metabolite profiling of *Artemisia annua* plants

Methods S1 Details of fractionation and UPLC-MRM-MS analysis.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.