Appl Microbiol Biotechnol (2017) 101:2021–2032
DOI 10.1007/s00253-016-7997-3

Expression in *Pichia pastoris* and characterization of two novel dirigent proteins for atropselective formation of gossypol

Isabelle Effenberger[1] · **Michael Harport**[1] · **Jens Pfannstiel**[2] · **Iris Klaiber**[2] · **Andreas Schaller**[1]

Received: 4 August 2016 / Revised: 30 October 2016 / Accepted: 5 November 2016 / Published online: 17 November 2016

Abstract We established an efficient fed-batch fermentation process for two novel dirigent proteins from cotton plants, GbDIR2 from *Gossypium barbadense* and GhDIR3 from *G. hirsutum*, using the engineered *Pichia pastoris* GlycoSwitch® SuperMan$_5$ strain to prevent hyperglycosylation. The two (His)$_6$-tagged proteins were purified by metal-chelate affinity chromatography and obtained in quantities of 12 and 15 mg L^{-1} of culture volume, respectively. Glycosylation sites were identified for the native and for the enzymatically deglycosylated proteins by mass spectrometry, confirming five to six of the seven predicted glycosylation sites in the NxS/T sequence context. The predominant glycan structure was Man$_5$GlcNAc$_2$ with, however, a significant contribution of Man$_{4–10}$GlcNAc$_2$. Both dirigent proteins (DIRs) mediated the formation of (+)-gossypol by atropselective coupling of hemigossypol radicals. Similar to previously characterized DIRs, GbDIR2 and GhDIR3 lacked oxidizing activity and depended on an oxidizing system (laccase/O_2) for the generation of substrate radicals. In contrast to DIRs involved in the biosynthesis of lignans, glycosylation was not essential for function. Quantitative enzymatic deglycosylation yielded active GbDIR2 and GhDIR3 in excellent purity. The described fermentation process in combination with enzymatic deglycosylation will pave the way for mechanistic and structural studies and, eventually, the application of cotton DIRs in a biomimetic approach towards atropselective biaryl synthesis.

Keywords Atropselectivity · Biaryls · Dirigent protein · Fermentation · N-glycosylation · *Pichia pastoris*

Electronic supplementary material The online version of this article (doi:10.1007/s00253-016-7997-3) contains supplementary material, which is available to authorized users.

✉ Andreas Schaller
andreas.schaller@uni-hohenheim.de

[1] Institute of Plant Physiology and Biotechnology, University of Hohenheim, 70593 Stuttgart, Germany

[2] Core Facility Hohenheim, Mass Spectrometry Unit, University of Hohenheim, 70593 Stuttgart, Germany

Introduction

Oxidative phenol coupling for C-C bond formation contributes to the biosynthesis of many natural products with interesting biological and pharmacological activities. One of the most prominent examples is morphine, a benzylisoquinoline alkaloid and powerful analgesic in opium poppy (*Papaver somniferum*). During morphine biosynthesis, the internal C-C phenol linkage of salutaridine is formed by salutaridine synthase with (*R*)-reticuline as the substrate (Beaudoin and Facchini 2014; Gesell et al. 2009). It is this intramolecular phenol coupling reaction, in addition to the presence of five stereocenters, that renders total synthesis of morphine commercially unfeasible (Gesell et al. 2009). Bimolecular phenol couplings also are widespread in plant natural products (Li and Lei 2014; Vassao et al. 2010). Prime examples include podophyllotoxin in *Podophyllum* (mayapple) and gossypol in *Gossypium* (cotton) species that are derived from the phenylpropanoid and terpenoid pathways, respectively. Podophyllotoxin, a precursor of semisynthetic cancer chemotherapeutics (Petersen and Alfermann 2001; Vassao et al. 2010), belongs to the structurally diverse class of lignans that are generated via bimolecular phenoxy radical coupling of two hydroxycinnamoyl alcohols (monolignols). In this particular example, two coniferyl alcohol radicals are coupled in a stereochemically controlled reaction resulting in (+)-pinoresinol, the central precursor of numerous 8–8′-linked

lignans. Gossypol, on the other hand, is a sesquiterpene dimer that is generated by C-C coupling of two hemigossypol radicals (Wagner et al. 2012). In the resulting tetra-*ortho*-substituted biaryl system (Fig. 1a), rotation around the central C-C bond is hindered and, therefore, two atropisomers of gossypol exist, which differ in toxicity and biological activity. While the two isomers are equally effective as chemoprotective agents in plant defense (Hagenbucher et al. 2013; Puckhaber et al. 2002; Stipanovic et al. 2006; Stipanovic et al. 2008), only (−)-gossypol is pharmacologically relevant with strong antispermatogenic and antiviral activities and high potential for the treatment of various cancers (Keshmiri-Neghab and Goliaei 2014; Lan et al. 2015; Wang et al. 2009). (−)-Gossypol is also more toxic than the (+)-isomer to non-ruminant animals, and the toxicity of (−)-gossypol thus limits the use of cottonseed oil and protein for food and feed (Zhou et al. 2013).

The oxidative coupling of phenols also is the most direct approach towards biaryl formation in organic synthesis. As compared to traditional transition metal-catalyzed approaches, the environmentally more benign coupling of unfunctionalized arenes offers many advantages, but it also presents formidable challenges with respect to regio- and enantioselectivity (Ashenhurst 2010; Bringmann et al. 2011; Kozlowski et al. 2009). In biological systems, stereochemical control is provided by proteins, and these proteins thus have

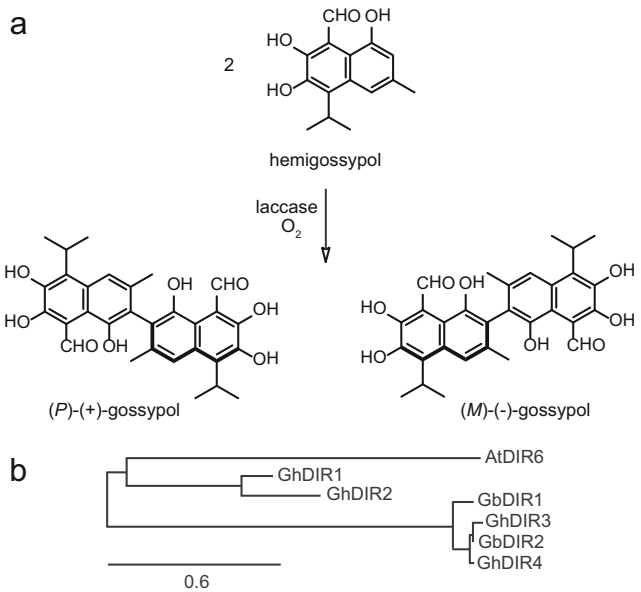


Fig. 1 Gossypol formation and phylogenetic relationship of DIRs involved in the process. **a** Oxidative coupling of two molecules of hemigossypol results in two atropisomers of gossypol, (*P*)-(+)-gossypol and (*M*)-(−)-gossypol. **b** The phylogenetic relationship is shown between GbDIR2 and GhDIR3 and other previously characterized *Gossypium* DIRs, as well as the pinoresinol-forming AtDIR6. The phylogenetic tree was generated at http://www.phylogeny.fr using the programs MUSCLE for multiple alignment, PhyML for tree building, and TreeDyn for tree rendering (Dereeper et al. 2008). Branch length is proportional to the number of substitutions per site

high potential for biotechnological approaches towards stereoselective biaryl synthesis (Aldemir et al. 2014; Wezeman et al. 2015). Nature's toolbox includes enzymes that provide both, oxidizing activity for phenoxy radical formation and stereochemical control during radical coupling. Specific cytochrome P450 monooxygenases, for example, catalyze the abovementioned intramolecular coupling reaction of (*R*)-reticuline, as well as intramolecular coupling of (*S*)-reticuline, during morphine and magnoflorine biosynthesis, respectively (Gerardy and Zenk 1992; Gesell et al. 2009; Ikezawa et al. 2008). Bacterial P450s are responsible for intramolecular C-C coupling in the biosynthesis of antibiotics (Bischoff et al. 2001; Howard-Jones and Walsh 2006; Li et al. 2011) and for intermolecular coupling in the biosynthesis of dimeric pre-anthraquinones (Präg et al. 2014). In fungi, the cytochrome P450s KtnC and DesC catalyze the regio- and stereo-selective bi-molecular coupling of dimethyl siderin to *P*-orlandin and *M*-desertorin, respectively (Gil Girol et al. 2012; Mazzaferro et al. 2015). In addition to cytochrome P450s, some laccases also were shown to provide both oxidizing capacity and stereochemical control, for example, during the formation of (−)-dalesconols in *Daldinia eschscholzii* (Fang et al. 2012) and of ellagitannins in *Tellima grandiflora* (Niemetz and Gross 2003a; Niemetz and Gross 2003b), and finally, unidentified peroxidases were shown to catalyze the oxidative dimerization of korupensamines A and B to michellamines A and C (Schlauer et al. 1998).

A completely different approach towards stereochemical coupling control is taken by dirigent proteins (DIRs). These proteins lack oxidizing activity and thus rely on, e.g., laccases or peroxidases for phenoxy radical formation. These radicals are then bound by DIRs and guided towards regio- and enantioselective product formation (Davin and Lewis 2005; Pickel and Schaller 2013). DIRs have first been described in *Forsythia intermedia* (FiDIR1 and 2), where they mediate the stereospecific formation of (+)-pinoresinol from two coniferyl alcohol radicals (Davin et al. 1997). The same specificity has been reported for DIRs from *Thuja plicata* (TpDIR1–8) (Kim et al. 2002), *Linum usitatissimum* (LuDIR1–4) (Dalisay et al. 2015), *Pisum sativum* (DRR206)(Seneviratne et al. 2015), and *Schizandra chinensis* (ScDIR1) (Kim et al. 2012). An enantiocomplementary activity, directing the bi-molecular coupling of coniferyl alcohol radicals towards the antipodal (−)-enantiomer of pinoresinol, has first been described in *Arabidopsis thaliana* (AtDIR6) (Pickel et al. 2010) and was then confirmed for additional DIRs from *Arabidopsis* (AtDIR5) (Kim et al. 2012) and *L. usitatissimum* (LuDIR5 and 6) (Dalisay et al. 2015). Very recently, DIRs were also shown to be involved in the biosynthesis of the sesquiterpenoid dimer gossypol in cotton (Effenberger et al. 2015; Liu et al. 2008). During oxidative coupling of hemigossypol, these DIRs confer stereospecificity to the formation of the central biaryl bond, yielding only one of two

possible gossypol atropisomers (Effenberger et al. 2015). These findings indicate that DIR-mediated stereoselective phenol coupling not only is relevant for phenylpropanoid metabolism in the biosynthesis of lignans but also extends to entirely different biosynthetic pathways and may thus play a more general role in plant secondary metabolism. Furthermore, DIRs have recently been implicated in certain developmentally controlled lignification processes, where they contribute to the formation of the Casparian strip in the developing *Arabidopsis* root (Hosmani et al. 2013) and to pod dehiscence in soybean (Funatsuki et al. 2014).

DIRs have great potential in organic synthesis for biomimetic approaches towards stereochemically controlled phenoxy radical coupling and biaryl synthesis (Pickel and Schaller 2013). However, there are two factors still precluding a more general application of DIRs as mediators of stereospecificity in organic synthesis. Firstly, we still do not know how they operate. Based on ligand binding studies and kinetic modeling of DIR1 from *Forsythia*, Lewis and coworkers suggested a model in which two phenoxy radicals are sequentially bound, one to each of the two subunits of the active DIR homodimer, and oriented in a way that favors a specific coupling mode (e.g., 8–8′ for pinoresinol formation in the lignan pathway) resulting in stereospecific product formation (Davin and Lewis 2005; Halls et al. 2004). However, this model was challenged by the recently solved crystal structure of DRR206, a (+)-pinoresinol-forming DIR from pea. The structure shows DRR206 as a homotrimer, in which the binding cavities of the three subunits are spatially far apart, not allowing the bound radicals to engage in the bi-molecular coupling reaction (Kim et al. 2015). Unfortunately, because several loops surrounding the potential binding cavity were not resolved in the structure, alternative modes of substrate binding and coupling control could not be deduced (Kim et al. 2015). Without such knowledge, it will be impossible to adapt DIRs for specific applications and to overcome their restricted substrate spectrum. In this respect, the engineering of cotton DIR specificity for 1- or 2-naphthols other than hemigossypol would be of particular interest, since enantiomerically pure binaphthols are widely used as ligands in asymmetric catalysis (Brunel 2005; Zhou 2011) and in the total synthesis of chiral biaryl natural products (Bringmann et al. 2011). More structural information is obviously needed to allow the engineering of DIRs by rational design.

Secondly, and this also relates to the accessibility of milligram quantities of homogenous protein for structure elucidation by X-ray crystallography or protein NMR, efficient expression systems are needed for the cost-effective, fast, and robust production of DIRs in quantities that are sufficient for biotechnological applications. While *Escherichia coli* may often be the first choice as a host for protein expression, it is not an option for DIRs, as previously characterized DIRs seem to rely on N-linked glycosylation for folding and/or solubility (Kazenwadel et al. 2013). The presence of a disulfide bond in, e.g., AtDIR6 may be an additional factor necessitating eukaryotic cells as expression system (Pickel et al. 2012). While functional DIRs have in fact been expressed in plant, insect, and yeast cell cultures, each of these systems suffered from serious limitations. As an obvious advantage of plant cells, native-like post-translational modifications can be expected. On the downside, the generation of transgenic plant cell cultures is very time-consuming, and reported expression levels did not exceed 0.2 mg of purified DIR protein per liter of culture volume (Effenberger et al. 2015; Kim et al. 2015; Pickel et al. 2010). While expression levels appeared to be somewhat higher (1 to 1.5 mg L^{-1}) in insect (*Spodoptera* Sf9 and *Drosophila* S2) cells, considerable heterogeneity was observed for DIRs expressed in these systems, due to differential glycosylation (Gang et al. 1999; Kim et al. 2002). In *Pichia pastoris*, very high expression was achieved for AtDIR6 (47 mg L^{-1}) by fed-batch fermentation. However, hyper-glycosylation that is frequently observed in this system resulted in an in-homogenous protein preparation (Kazenwadel et al. 2013).

In the present study, we tried to overcome these limitations by use of an engineered yeast strain, *Pichia* GlycoSwitch® SuperMan$_5$ (BioGrammatics Inc.). In this strain, the *OCH1* gene is disrupted, which codes for the α-1,6-mannosyltransferase that is responsible for hypermannosylation of glycans in yeast. In addition, by overexpression of an ER-resident α-1,2-mannosidase, glycans are trimmed to a more uniform $Man_5GlcNAc_2$ structure (Jacobs et al. 2009; Vervecken et al. 2004). Two novel DIRs were produced in this system by fed-batch fermentation, GhDIR3 from *Gossypium hirsutum* (upland cotton) and GbDIR2 from *Gossypium barbadense* (sea island cotton), and characterized with respect to their ability to mediate atropselective gossypol formation.

Methods

Materials

Analytical grade chemicals were purchased from Carl Roth GmbH (Karlsruhe, Germany) or Sigma-Aldrich Chemie GmbH (Munich, Germany). DNA modifying and restriction enzymes were from Fermentas (Thermo Fisher Scientific; Dreieich, Germany) and used according to the manufacturer's protocols. Oligonucleotides were obtained from Eurofins Genomics (Ebersberg, Germany). *E. coli* strain DH10B [F^- mcrA Δ(*mrr-hsd*RMS-*mcr*BC) φ80d*lac*ZΔM15 Δ*lac*X74 *end*A1 *rec*A1 *deo*R Δ(*ara*,*leu*)7697 *ara*D139 *gal*U *gal*K *nup*G *rps*L λ^-] and the *Pichia* GlycoSwitch® expression strain (*P. pastoris* SuperMan$_5$, *HIS*$^+$, *pep4*Δ, *mut*$^+$) were

obtained from Invitrogen (a Thermo Fisher Scientific company) and BioGrammatics Inc. (Carlsbad, CA, USA), respectively.

Expression constructs

The open reading frames (ORFs) of mature GbDIR2 (GenBank AAY44415) and GhDIR3 (AFP65784) were codon-optimized for expression in *P. pastoris* using the GeneOptimizer® algorithm (Raab et al. 2010). The ORFs were extended at the 3′-end with six codons for histidine and custom synthesized by GeneArt Gene Synthesis (Thermo Fisher Scientific), including *Xho*I and *Xba*I restriction sites at the 5- and 3′-ends, respectively. The ORFs were then cloned into pPICZαB (Invitrogen), under the control of the AOX1 promoter and terminator, and in frame with the α-factor signal peptide. The vector was amplified in *E. coli* DH10B, linearized with *Sac*I, and transformed into *Pichia* GlycoSwitch by electroporation according to the protocols of Invitrogen. Transformants were selected on yeast extract-peptone-dextrose-sorbitol (YPDS) agar containing 100 μg/mL Zeocin.

Fed-batch fermentation

To generate the inoculum, a pre-culture was grown from a single colony in 100 mL buffered complex glycerol medium (BMGY) at 220 rpm and 28 °C. When OD_{600} was between 3 and 4, the pre-culture was added to 1 L basal salts medium (18.2 g K_2SO_4, 14.9 g $MgSO_4 \times 7H_2O$, 4.13 g KOH, 0.93 g $CaSO_4 \times 2H_2O$, 26.7 mL 85% H_3PO_4), supplemented with 4.35 mL PTM_1 trace salts (1 L PTM_1 contained 6 g CuSO4 × $5H_2O$, 0.08 g NaI, 3.0 g $MnSO_4 \times H_2O$, 0.5 g $CoCl_2$, 20.0 g $ZnCl_2$, 0.02 g H_3BO_3, 0.2 g $Na_2MoO_4 \times 2H_2O$, 65.0 g $FeSO_4 \times 7H_2O$, 0.2 g biotin and 5 mL H_2SO_4) and 4% (*w*/*v*) glycerol in a Infors HT Labfors-4 bioreactor (7.5 L vessel volume; Infors GmbH, Einsbach, Germany). Fermentation parameters were monitored and controlled with IRIS 2.0 (V5.3) software (Infors HT). Temperature was maintained at 28 °C, pH at 5.0 with aqueous NH_4OH (25%), and stirrer speed at 600 rpm. Initial airflow was set at 4 L/min. The culture was grown until the initial glycerol was consumed as indicated by a dissolved oxygen (DO) spike after 24–48 h. In the subsequent glycerol fed-batch phase, 50% (*w*/*v*) glycerol and 12 mL/L PTM_1 trace salts were fed at 18 mL/h for 5 h. When a DO spike indicated consumption of all glycerol, methanol (12 mL/L PTM1 salts in 100% methanol) was added to induce protein expression under control of the AOX1 promoter. The MeOH feeding rate needed some optimization. We started out with an initial rate of 1.8 mL/h that was increased after 3 to 5 h to 3.6 mL/h, as shown for the fermentation of GhDIR3 shown in Fig. 1b. However, GlycoSwitch $SuperMan_5$ turned out to be more sensitive to MeOH toxicity than conventional *Pichia* strains, and therefore, feeding rates were reduced to 0.8 mL/h, initially, and 1.8 mL/h at 3 to 5 h later. Continuous growth of the culture throughout the fed-batch fermentation phase was only observed with these reduced MEOH concentrations (compare Fig. 2a, b). Airflow was adjusted as needed to maintain DO above 20%. Stronger oscillation of the DO curve in Supplementary Fig. S1a as compared to Supplementary Fig. S1b is due to the increased metabolic activity at lower MeOH concentrations. Daily samples were taken to monitor the increase in cell mass and recombinant protein expression. After 5 to 6 days of methanol fed-batch fermentation, the culture was harvested and the culture supernatant containing the recombinant DIRs was separated from the cells by centrifugation at 3200×*g*, for 30 min at 4 °C.

Protein purification and quantification

Recombinant GbDIR2 and GhDIR3 were purified by metal chelate affinity chromatography on Ni^{2+}- nitrilotriacetic acid (NTA) agarose (Qiagen; Hilden, Germany). To this end, the

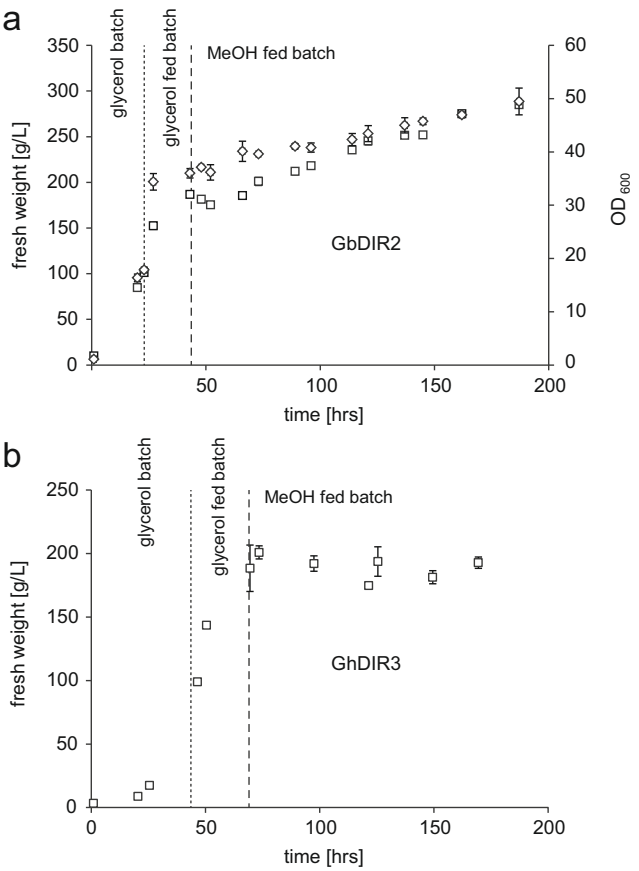


Fig. 2 Fed-batch fermentation of GbDIR2 and GhDIR3. The increase in cell density was monitored as OD_{600} (*diamonds*) and/or fresh weight (*squares*) for GbDIR2 (**a**) and GhDIR3 (**b**). Standard deviation is shown for the mean of three individual samples. *Dotted lines*, shift from glycerol batch to fed-batch process; *dashed lines*, shift to methanol fed-batch process. Dissolved oxygen, pH, temperature, stirrer speed, and feeding rate were monitored and controlled by IRIS 2.0 (Infors HT) and are shown in Supplementary Fig. S1

culture supernatant was adjusted to 10 mM imidazole and 100 mM potassium phosphate pH 6.0. After addition of Ni^{2+}-NTA agarose (6 mL/L), the slurry was incubated for 2 h at 4 °C under constant agitation. The matrix was recovered by centrifugation (100×*g*, 5 min, 4 °C) and washed three times in two volumes potassium phosphate buffer (100 mM, pH 6.0), 300 mM NaCl, and 20 mM imidazole. Bound proteins were eluted from the matrix by three additions of one volume 300 mM imidazole in the same buffer. The entire purification procedure was repeated as needed to recover all DIR protein from the culture supernatant. Fractions were analyzed for protein content using the Bradford procedure and bovine serum albumin as reference protein (Bradford 1976). Fractions that tested positive for DIR proteins on western blots were combined and stored at −80 °C.

SDS-PAGE and western blot analysis

SDS-PAGE analysis was performed on 14% polyacrylamide gels using the discontinuous buffer system described by Laemmli (1970). Gels were stained with Coomassie Brilliant Blue R250 and documented on a Gel Doc EZ system (BioRad; Munich, Germany). For western blot analysis, a trans-blot SD semidry transfer cell (BioRad) was used for protein transfer onto nitrocellulose membranes (Schleicher & Schuell Protran 85; Hartenstein GmbH, Würzburg, Germany). An affinity purified polyclonal antibody (Perbio Science; Bonn, Germany) directed against a peptide (DPLRTGPEPDSELIG) that is conserved among cotton DIRs was used at 1:500 as the primary antibody, with goat-anti-rabbit IgG horseradish peroxidase conjugate (1:10,000, Calbiochem/Merck; Schwalbach, Germany) as secondary antibody as described (Effenberger et al. 2015). Blots were developed by enhanced chemiluminescence (ECL) and exposure to X-ray film. Alternatively, the chemiluminescence signal was detected and quantified with a C-DiGit blot scanner (Li-COR; Bad Homburg, Germany).

Deglycosylation of DIRs

Affinity purified DIRs were dialyzed against 0.1 M MES (2-(*N*-morpholino)-ethanesulfonic acid) pH 5.5 and ultrafiltrated on Vivaspin 20 (10 kDa cutoff; Sartorius AG; Göttingen, Germany) to a final concentration of 2.5 mg/mL. For deglycosylation, 10 μg of each DIR was incubated with 1 U $EndoH_f$ (New England Biolabs; Ipswich, MA, USA) in 20 μL 1× G5 reaction buffer for 2 h at 30 °C according to the manufacturer's recommendations. Soluble and insoluble fractions were separated by centrifugation (14,000×*g*, 10 min). The sediment was washed with 500 μL reaction buffer. The supernatant was analyzed for dirigent activity using the standard assay system.

Assay of dirigent activity

The dirigent assay for atropselective formation of gossypol was based on Liu et al. (2008) and modified as described (Effenberger et al. 2015). The reaction mixture contained 80 μM hemigossypol, *Trametes versicolor* laccase (4 mU), and dirigent proteins at the indicated concentrations (0 to 5 μg) in a total volume of 300 μL 0.1 M MES pH 5.5. The reaction was started by addition of laccase. After 15 min at 30 °C under constant agitation, the reaction was terminated by three extractions with 300 μL *n*-hexane/ethyl acetate (1:1). The combined extracts were evaporated to dryness (Savant SpeedVac Concentrator; Thermo Scientific). The residue was resuspended in 100 μL derivatizing reagent (2 mL D-alaninol, 10 mL glacial acetic acid in 100 mL *N*,*N*-dimethylformamide) and incubated at 100 °C for 30 min. The resulting diastereomeric Schiff base adducts of substrate and products were analyzed by reversed-phase (RP) HPLC essentially as described (Hron et al. 1999). Briefly, 100 μL HPLC solvent was added (60% A (10 mM KH_2PO_4/H_3PO_4 pH 3.0), 40% B (acetonitrile; ACN)), and 100 μL of the solution was injected onto a Gemini-NX C18-column (250 × 4.6 mm, 5 μ) equipped with a corresponding guard column (4 × 3 mm; Phenomenex; Aschaffenburg, Germany) on a LaChrom HPLC system (D-7000 system manager, L-7000 interface, L-7100 pump, L-7250 autosampler, L-7455 diode array detector; Merck Hitachi, Mannheim, Germany) run at 1 mL/min with gradient elution (40–80% B in 15 min, 10 min isocratic hold at 80% B, 80–40% B in 10 min) and detection at 254 nm. The ratio of gossypol atropisomers was calculated as enantiomeric excess (ee) of one gossypol enantiomer over the other as ee (%) = $(a - b)/(a + b) \times 100$, with *a* and *b* corresponding to the peak areas for (+)- and (−)-gossypol, respectively.

Mass spectrometric analysis and glycan structure

Deglycosylated GbDIR2 and GhDIR3 were in-gel-digested with chymotrypsin (Roche, Penzberg, Germany) according to Shevchenko et al. (1996) using a modified incubation buffer (100 mM Tris/HCl pH 7.6, 10 mM $CaCl_2$). After digestion, the gel pieces were extracted with 50% ACN/50% 0.1% formic acid (FA) (*v*/*v*) for 15 min. The supernatant was collected, and the gel pieces were covered with 5% FA for 15 min before the same volume of ACN was added. After incubation for 10 min, the supernatant was collected. The pooled supernatants were then dried in a vacuum centrifuge and stored at −20 °C. Dried samples were dissolved in 0.1% FA.

The in-solution digests of DIRs were performed in 100 mM Tris/HCl pH 7.6 and, 10 mM $CaCl_2$. After reduction with 10 mM DTT and alkylation with 30 mM chloroacetamide, samples were digested overnight at 37 °C by adding 1 μg chymotrypsin (Roche, Penzberg, Germany). The reaction

was stopped by addition of trifluoroacetic acid (TFA). The peptides were desalted over C18-STAGE tips (Rappsilber et al. 2003), dried in a vacuum concentrator, and dissolved in 5% ACN and 0.1% acetic acid. Mass spectrometry (MS) analysis was performed as described (Cedzich et al. 2009; Pickel et al. 2010), and data were analyzed using Mascot (Matrix Science, UK) and Byonic (Protein Metrics, USA) as search engines.

Accession numbers

The codon-optimized sequences for GhDIR3 and GbDIR2 were submitted the European Nucleotide Archive (http://www.ebi.ac.uk/ena) and were assigned accession numbers LT605201 and LT605202, respectively. The GenBank accessions for the original unmodified sequences are JX096805 and DQ018709, respectively.

Results

We recently cloned and characterized GhDIR4, a dirigent protein from *G. hirsutum* var. marie-galante that mediates atropselective coupling of hemigossypol radicals to (+)-gossypol in vitro (Effenberger et al. 2015). Consistent with the specificity of GhDIR4, this particular variety of upland cotton accumulates (+)-gossypol in up to 90% enantiomeric excess (Cass et al. 1991). An enantiocomplementary DIR may thus be expected in plants that contain an excess of the antipodal (−)-gossypol. Therefore, we looked at DIRs from *G. barbadense* which is exceptional among cotton species since it produces gossypol in a 3:2 ratio of the (−)- over the (+)-isomer (Cass et al. 2004; Jaroszewski et al. 1992). However, GbDIR1 that was cloned from *G. barbadense* and transiently expressed in *Nicotiana benthamiana* did not show any dirigent activity with respect to gossypol formation, suggesting that hemigossypol is not a substrate (Effenberger et al. 2015). In the present study, we addressed the function of GbDIR2, a second DIR from *G. barbadense*, and compared it to GhDIR3, the closest homolog in *G. hirsutum* (Fig. 1b). *P. pastoris* was chosen as host for heterologous expression of GbDIR2 and GhDIR3.

Our initial attempts to express His-tagged GhDIR4 in *P. pastoris* failed. The same system (pPICZαA vector for methanol-inducible expression in *P. pastoris* strain X-33) that was previously employed for high-level expression of AtDIR6 (Kazenwadel et al. 2013) did not result in satisfactory expression levels, and furthermore, strong hyper-glycosylation of the recombinant protein was observed (Supplementary Fig. S2). Therefore, to enhance expression levels, we optimized the sequences of GbDIR2 and GhDIR3 for codon usage in *P. pastoris* (GeneOptimizer®; Raab et al. 2010) and used an engineered *P. pastoris* strain to prevent hyper-glycosylation. In *Pichia* GlycoSwitch® (SuperMan$_5$ (pep4-); BioGrammatics Inc., Carlsbad, CA, USA), N-linked glycosylation entails more uniform $Man_5GlcNAc_2$ modifications at NxS/T consensus sites (Jacobs et al. 2010).

Fed-batch fermentation in *P. pastoris* SuperMan$_5$

For optimum yields, high-density fed-batch fermentation of *P. pastoris* SuperMan$_5$, bearing methanol-inducible expression constructs for GbDIR2 and GhDIR3, was performed in a Labfors4 bioreactor. Both fermentations proceeded comparably (Fig. 2; Supplementary Fig. S1). Cells were initially grown in batch culture on glycerol as carbon source. When the initial glycerol was depleted, after 23 h for GbDIR2 (Fig. 2a) and 47 h for GhDIR3 (Fig. 2b), more glycerol was fed over 5 h to allow for a further increase in cell density. Upon consumption of this glycerol as indicated by a spike in dissolved oxygen after 43 h for GbDIR2 and 70 h for GhDIR3 (Supplementary Fig. S1a, b), the methanol feed was initiated to induce protein expression from the AOX1 promoter. Optimized methanol concentrations supported continuous growth of the culture and a moderate increase in cell density during the methanol fed-batch phase for GbDIR2 (Fig. 2a), while no further increase was observed for GhDIR3 (Fig. 2b). Final cell densities reached 285 and 193 g L^{-1} for GbDIR2 and GhDIR3, respectively. After 5 to 6 days of methanol fed-batch processing, the culture was harvested and the cells were separated from the culture supernatant containing the recombinant DIRs by centrifugation. Recombinant DIRs were purified by metal chelate affinity chromatography on Ni-NTA agarose resulting in 12 and 15 mg L^{-1} of culture volume for GbDIR2 and GhDIR3, respectively.

Dirigent activity of GbDIR2 and GhDIR2 expressed in *P. pastoris*

The activity of GbDIR2 and GhDIR3 with respect to atropselective synthesis of gossypol from hemigossypol was tested in presence and absence of laccase from *T. versicolor* (Fig. 3a). In absence of laccase, hemigossypol was not consumed and gossypol formation was not observed. With laccase alone, the two gossypol atropisomers were formed in equal amounts. In presence of laccase with either GbDIR2 or GhDIR3, there was a clear increase in (+)-gossypol at the expense of the (−)-enantiomer (Fig. 3a). The data indicate that GbDIR2 and GhDIR3, like previously described DIRs, lack oxidizing activity and depend on laccase/O_2 for the formation of substrate radicals by one-electron oxidation of hemigossypol. Both DIRs direct the radical coupling reaction towards (+)-gossypol formation and thus resemble the previously characterized GhDIR4 and GaDIR1 from *G. arboreum* with respect to enantioselectivity (Effenberger et al. 2015). The enantiomeric excess of (+)-gossypol depended on the

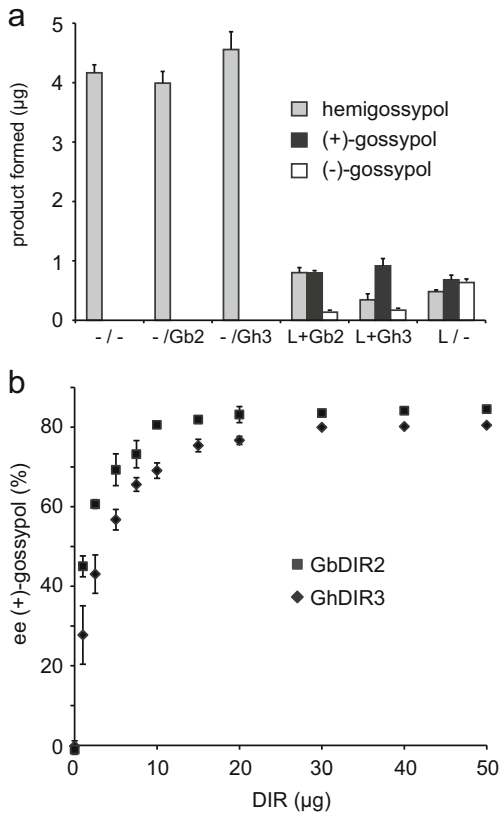
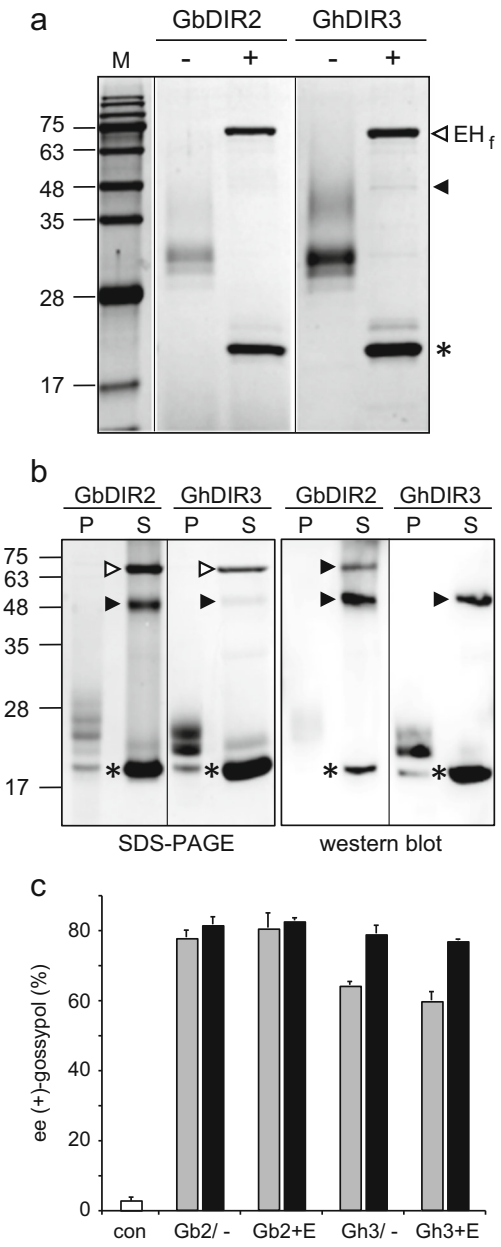


Fig. 3 Activity of GbDIR2 and GhDIR3. **a** GbDIR2 and GhDIR3 lack oxidizing activity. The conversion of hemigossypol to gossypol was monitored by derivatization of reaction products with D-alaninol. Diastereomeric Schiff base adducts of substrate (hemigossypol; *gray bars*) and products ((+)- and (−)-gossypol; *black and white bars*, respectively) were separated and quantified by RP-HPLC. The presence or absence of laccase (*L*) and addition of GbDIR2 (*Gb2*) or GhDIR3 (*Gh3*) is indicated. Values represent the mean ± standard deviation of three independent assays. **b** Atropselective formation of (+)-gossypol by GbDIR2 and GhDIR3. The enantiomeric excess (ee in percent) of (+)-gossypol formation is shown for increasing amounts (µg) of GbDIR2 (*squares*) or GhDIR3 (*diamonds*) in the assay. Values represent the mean ± standard deviation of three independent assays



DIR concentration reaching saturation at 80 and 85% ee with about 10 and 20 µg protein in the assay for GbDIR2 and GhDIR3, respectively (Fig. 3b). The specific activity of these protein preparations is thus about 10-fold lower than that of GhDIR4 that was purified to homogeneity from a plant cell culture expression system (Effenberger et al. 2015).

Enzymatic deglycosylation

SDS-PAGE analysis of the purified proteins showed a predominant but diffuse band with an apparent molecular mass of 30 kDa for both GbDIR2 and GhDIR3 (Fig. 4a). This is considerably more than the mass calculated for GbDIR2 (M_r = 16,208) and GhDIR3 (M_r = 16,359) suggesting that the proteins were glycosylated. Indeed, the apparent molecular weight was reduced upon enzymatic deglycosylation with EndoH_f resulting in a well-defined band at about 19 kDa for both proteins. Since glycosylation was reported to be required for solubility and function of the pinoresinol-forming AtDIR6 (Kazenwadel et al. 2013), we tested the effect of deglycosylation on solubility and activity. After treatment with EndoH_f, GbDIR2 and GhDIR3 were separated into soluble and insoluble fractions by centrifugation. Interestingly, western blot analysis detected both proteins in the supernatant, while only a minor, apparently partially deglycosylated fraction was found in the sediment (Fig. 4b). The deglycosylated proteins showed an increased tendency to form soluble, stable oligomers, presumably the trimer in case of GhDIR3 and trimer and tetramer for GbDIR2 (Fig. 4b).

◀ **Fig. 4** N-linked glycosylation of GbDIR2 and GhDIR3 is not required for activity. **a** SDS-PAGE analysis of GhDIR2 and GbDIR3. The DIRs were expressed in batch culture, purified by metal chelate affinity chromatography and separated by SDS-PAGE with (+) and without (−) deglycosylation by $EndoH_f$. A Coomassie-stained 14% gel is shown. **b** Deglycosylated DIRs are soluble. GbDIR2 and GhDIR3 (10 μg; affinity purified after fed-batch fermentation) were deglycosylated with $EndoH_f$ (1 U, 2 h, 30 °C). Soluble (*S*) and insoluble fractions (*P*) were separated by centrifugation (16,000×*g*, 10 min); aliquots (0.5 μg) were separated by SDS-PAGE (*left panel*) and analyzed by western blot (*right panel*) using an affinity-purified DIR antiserum and enhanced chemiluminescence (ECL) detection. **a**, **b** The position of the deglycosylated DIRs is marked by an *asterisk* and that of potential trimers and tetramers by *black triangles*. *Open triangles* mark the position of $EndoH_f$. The mass of marker proteins (*M*) is indicated in kilodalton. **c** Deglycosylated DIRs are active. The activity of 1 μg (*gray bars*) and 5 μg (*black bars*) GbDIR2 (*Gb2*) and GhDIR3 (*Gh3*) was analyzed before (*/ -*) and after deglycosylation (*+E*), and compared to the control without DIRs (*con*; *white bar*). Activity is shown in percent ee of (+)-gossypol formation. Values represent the mean of three assays ± standard deviation

Importantly, deglycosylation did not affect the activity of GbDIR2 and GhDIR3. The ee of (+)-gossypol formation was the same before and after $EndoH_f$ treatment (Fig. 4c).

Glycosylation sites and glycan structure

For the identification of glycosylation sites, we used chymotrypsin for an in-gel digest of the deglycosylated DIRs (cf. Fig. 4a). $EndoH_f$ cleaves between the two *N*-acetylglucosamine (GlcNAc) units in the diacetylchitobiose core of high-mannose glycans. This leaves one GlcNAc linked to the Asn resulting in a relative mass increase of 203.079 facilitating the identification of glycosylation sites by mass spectrometry (MS). In addition, the fully glycosylated forms of the two proteins were digested with chymotrypsin in solution. Chymotryptic peptides identified by LC-ESI-MS/MS covered 93 and 89% of the GbDIR2 and GhDIR3 sequences, respectively, allowing the identification of the same 5 to 6 glycosylation sites at N60, N64/N65, N104, N121, and N168 in GbDIR2 and GhDIR3, respectively (Fig. 5; Supplementary Figs. S3–S5). At N64/N65, there are two potential glycosylation sites flanking each other. Since differentiation between these two glycosylation sites by the search engines was based on a single low abundant fragment ion (Supplementary Fig. S5), MS/MS analysis of the glycosylated peptides could not discriminate unambiguously which of the two sites was modified. Therefore, we cannot say whether one of the sites is glycosylated preferentially or not. However, glycosylation of one of the two sites seems to preclude glycosylation of the other, since we did not identify any peptides carrying GlcNAc modifications at both Ns. In addition to these six sites, there is one more potential NxS/T glycosylation site in GbDIR2 and GhDIR3 at N159 (Fig. 5). In our MS analysis, chymotryptic peptides that included N159 could not be identified, neither in the glycosylated nor in the unmodified form. Therefore, it remains unclear whether or not this site is glycosylated as well.

The in-solution digest of the glycosylated DIRs also provided insight into the structure of the glycan chains linked to N121 and N168. MS/MS spectra indicated predominantly $Hex_5HexNAc_2$ modifications at N121 which is consistent with the structure expected for the engineered *Pichia* strain, i.e., five mannoses attached to the chitobiose core of two GlcNAc. However, in both GbDIR2 and GhDIR3, there was also a minor contribution of $Hex_7HexNAc_2$ at this site (data not shown). Glycan chains at N168 were more diverse, with 5 to 7 mannose moieties ($Hex_{5–7}HexNAc_2$) for GhDIR3 and 4 to 10 ($Hex_{4–10}HexNAc_2$) in GbDIR2 (Supplementary Fig. S6). The relative abundance of the different glycan chains at N168 of GbDIR2 was quantified on basis of the normalized signal intensities for the corresponding peptides. Consistent with the expected $Man_5GlcNAc_2$ structure, the glycoform with five hexoses amounted to more than 90% of all glycopeptides, while other glycoforms with 4, 6, 7, 8, 9, or 10 hexoses were detected at a much lower relative abundance of 0.4, 5, 0.6, 2, 0.6, and 0.2%, respectively (Supplementary Fig. S7).

```
                .         .         .         .         .         .
GbDIR1  MRGTLMLSWVLIICLSLVAVQSQYYSETLPYRPRPVKVTNLHFFMLEFTGITAVQVAQVN  60
GbDIR2  MRGTSVLSWILIICVCQVAVRSQYYSDTLPYQPRPVLVTNLHFYMHEFTGTTAVVLTQAN  60
GhDIR3  MRGTSVLSWILIICLSQVAVRSQYYSDTVPYHPRPVVVTNLHFYMHEFTGTTAVVLTQAN  60
        **** :***:****:. ***:*****:*:**:**** ******:* **** *** ::*.*
                .         .         .         .         .         .
GbDIR1  ITSSDNNSSVPFASLVAVNDPLRTGPEPDSELIGNVQGIALLAGTNASSTQYIDFGFNTG 120
GbDIR2  ITS--NNSSVPFATLVAVNDPLRTGPEPDSELIGNVQGISLLAGSNASSTQYIDFGFNTG 120
GhDIR3  ITS--NNSSVPFATLVAINDPLRTGPEPDSELIGNVQGMSLLAGSNASSTQYIDFEFNTG 120
        ***  ********:***:********************::****:********** ****
                .         .         .         .         .
GbDIR1  KLNGSSLSVFSRGEPGLAVVGGRGRFMMATGVALFNPILINATNVIIEFNVTVIHY 176
GbDIR2  KFNGSSLSVFSRGEAGLAVVGGRGQFAMATGTALFNPLLINATNVIIEFNVTVIHY 174
GhDIR3  KFNGSSLSVFSRGEAGLAVVGGRGQFAMATGTALFNPLLINATNVIMEFNFTVIHY 174
        *:************.*********:* ****.*****:********:***.*****
```

Fig. 5 Glycosylation sites of GbDIR2 and GhDIR3. A sequence alignment of GbDIR1, GbDIR2, and GhDIR3 was generated with Clustal Ω (Sievers et al. 2011). Identical, highly conserved, and less-conserved sites are marked with *asterisks*, *colons*, and *periods*, respectively. Signal peptide sequences are shown in *blue italics*. Potential N-linked glycosylation sites in the NxS/T sequence context are highlighted in *magenta*. Experimentally confirmed sites are shown in *bold* (color figure online)

Discussion

The goals of this study were 2-fold. Firstly, we aimed at improving heterologous expression of DIRs by using an engineered *P. pastoris* host strain for more uniform glycosylation, and secondly, we intended to functionally characterize two novel DIRs, GbDIR2 from *G. barbadense* and GhDIR3 from *G. hirsutum*.

Both proteins were found to mediate atropselective radical coupling resulting in the formation of (+)-gossypol from hemigpossypol in presence of laccase/O_2 as the oxidant. Enantioselectivity is thus the same as for the previously characterized GhDIR4 and GaDIR1. This result was expected for GhDIR3 from *G. hirsutum* var. marie-galante, since this variety contains (+)-gossypol in large enantiomeric excess. The enantiomeric composition of the gossypol pool is thus consistent with the atropselectivity observed for GhDIR3 in vitro, and this protein may thus participate in gossypol synthesis in vivo. *G. barbadense*, on the other hand, contains a moderate 3:2 excess of (−)-gossypol in seeds and hairy root cultures (Cass et al. 2004; Frankfater et al. 2009; Jaroszewski et al. 1992), suggesting the involvement of a (−)-gossypol-forming DIR in gossypol biosynthesis. Hence, considering the scarcity of (+)-gossypol, the identification of GbDIR2 with (+)-selectivity was unexpected. However, the constitutive expression of GbDIR2 appears to be low and its contribution to gossypol biosynthesis in healthy plants may thus be limited (Zhu et al. 2007). GbDIR2 expression is strongly induced after *Verticillium dahliae* infection (Zhu et al. 2007), as part of the defense response that is mediated by the ethylene and jasmonate signaling pathways in cotton plants (Gao et al. 2013; Guo et al. 2016). Consistent with the induced expression of GbDIR2, the enantiomeric composition of the gossypol pool in *G. barbadense* hairy roots changes in response to methyl jasmonate treatment from scarcity (2:3) to an excess (3:2) of (+)- as compared to (−)-gossypol (Frankfater et al. 2009). The activity of GbDIR2 therefore seems to be relevant only after *Verticillium* infection, and by mediating (+)-gossypol formation, it may contribute to the resistance of *G. barbadense* to Verticillium wilt. The DIR that is responsible for the enantiomeric excess of (−)-gossypol in uninfected *G. barbadense* remains to be identified. The identification of such a DIR would be desirable from a biotechnological point of view, as it may pave the way for the biomimetic synthesis of (−)-gossypol, which is the pharmacologically relevant enantiomer (Keshmiri-Neghab and Goliaei 2014; Lan et al. 2015).

Any biotechnological application will rely on an efficient expression system for functional DIRs. Such a system was established here using the engineered *P. pastoris* SuperMan$_5$. Fed-batch fermentation yielded 12 mg L^{-1} GbDIR2 and 15 mg L^{-1} GhDIR3 after metal-chelate affinity purification of the $(His)_6$-tagged proteins. This translates into a 60- and 75-fold increase in protein production, respectively, as compared to the plant cell culture system that was used for the expression of the related GhDIR4 protein (Effenberger et al. 2015). Using *P. pastoris* SuperMan$_5$ as expression host, we achieved a reduction in hyper-glycosylation as compared to the commonly used strain X-33. SuperMan$_5$ lacks a functional *OCH1* gene coding for the mannosyltransferase that is responsible for hyper-glycosylation and, in addition, expresses an α-1,2-mannosidase that trims Man_8 glycans down to Man_5 resulting in >90% $Man_5GlcNAc_2$ (Jacobs et al. 2009; Vervecken et al. 2004). Greater than 90% $Man_5GlcNAc_2$ glycosylation was also reported for secreted granulocyte-macrophage colony-stimulating factor that was produced by the engineered *Pichia* strain in a fed-batch fermentation process similar to the one established here (Jacobs et al. 2010). Recombinant human mast cell chymase that was expressed in this system also showed a reduction in hyper-glycosylation, even though the composition of N-linked glycans was not analyzed in this study (Smith et al. 2014). In the present study, $Man_5GlcNAc_2$ was confirmed as the predominant modification of Gossypium DIRs produced in SuperMan$_5$, with >90% relative abundance at N168 of GbDIR2 (Supplemental Fig. S7), and only a minor contribution of other glycans carrying 4 to 10 mannose moieties. Therefore, much more uniform glycosylation was achieved in SuperMan$_5$ as compared to the hyper-glycosylation of GhDIR4 expressed in *Pichia* strain X-33 (Supplemental Fig. S2) and compared to the 9 to 13 mannose moieties that were observed for Arabidopsis DIR6 expressed in the latter system (Kazenwadel et al. 2013).

The choice of *P. pastoris* as the expression host offers an additional advantage over the plant cell culture system previously used for the expression of GhDIR4 (Effenberger et al. 2015). The complex glycans of higher eukaryotes are resistant to cleavage by EndoH. Plant glycans also are resistant to cleavage by the endoglycosidase PNGaseF, because of the plant-specific addition of α-1,3-linked fucose to the proximal GlcNAc of the conserved $Man_3GlcNAc_2$ core (Rayon et al. 1998; Wilson et al. 2001; Tretter et al. 1991). The high-mannose glycans in *Pichia*, on the other hand, are readily cleaved by PNGaseF and by $EndoH_f$ that was used in this study. Deglycosylation of GbDIR2 and GhDIR3, followed by chymotrypsin digestion and MS analysis of the resulting peptides identified five or six identical glycosylation sites in both proteins (the ambiguity being caused by the vicinal N64 and N65 sites). While these glycosylation sites appear to be conserved in gossypol-forming DIRs and may be relevant during protein folding and/or passage through the secretory pathway, the glycans are not required for the function of the mature proteins, as the deglycosylated DIRs retained full activity. This is in sharp contrast to pinoresinol-forming DIRs which possess only two conserved glycosylation sites. These sites are distinct from those in gossypol-forming DIRs, and importantly, they are essential for function. Deglycosylation of AtDIR6 with PNGaseF resulted in a progressive loss of

activity and solubility of the protein (Kazenwadel et al. 2013), while GbDIR2 and GhDIR3 retained activity after deglycosylation by EndoH$_f$. Therefore, N-linked glycan chains are dispensable for the function of GbDIR2 and GhDIR3; the single *N*-acetylglucosamine that remains attached to the asparagine after EndoH$_f$ treatment appears to be sufficient. The excellent expression levels that can be achieved by *Pichia* SuperMan$_5$ fed-batch fermentation concomitant with the superior homogeneity of the deglycosylated protein preparation will greatly facilitate future efforts to crystallize these gossypol-forming DIRs in order to elucidate their structure and mechanism of action.

Acknowledgements We thank Jutta Babo for cloning of the GhDIR4 expression construct and Benedikt Fabry for its transformation into *Pichia* X33.

Compliance with ethical standards

Funding This study was funded by Deutsche Forschungsgemeinschaft (DFG; grant number SCHA 591/10-1).

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Aldemir H, Richarz R, Gulder TAM (2014) The biocatalytic repertoire of natural biaryl formation. Angew Chem Int Ed Engl 53:8286–8293. doi:10.1002/anie.201401075

Ashenhurst JA (2010) Intermolecular oxidative cross-coupling of arenes. Chem Soc Rev 39:540–548. doi:10.1039/b907809f

Beaudoin GAW, Facchini PJ (2014) Benzylisoquinoline alkaloid biosynthesis in opium poppy. Planta 240:19–32. doi:10.1007/s00425-014-2056-8

Bischoff D, Pelzer S, Bister B, Nicholson GJ, Stockert S, Schirle M, Wohlleben W, Jung G, Sussmuth RD (2001) The biosynthesis of vancomycin-type glycopeptide antibiotics—the order of the cyclization steps. Angew Chem Int Ed Engl 40:4688–4691. doi:10.1002/1521-3773(20011217)40:24<4688::AID-ANIE4688>3.0.CO;2-M

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72:248–254. doi:10.1016/0003-2697(76)90527-3

Bringmann G, Gulder T, Gulder TAM, Breuning M (2011) Atroposelective total synthesis of axially chiral biaryl natural products. Chem Rev 111:563–639. doi:10.1021/cr100155e

Brunel JM (2005) BINOL: A versatile chiral reagent. Chem Rev 105:857–897. doi:10.1021/cr040079g

Cass QB, Oliveira RV, De Pietro AC (2004) Determination of gossypol enantiomer ratio in cotton plants by chiral higher-performance liquid chromatography. J Agr Food Chem 52:5822–5827. doi:10.1021/jf049626p

Cass QB, Tiritan E, Matlin SA, Freire EC (1991) Gossypol enantiomer ratios in cotton seeds. Phytochemistry 30:2655–2657. doi:10.1016/0031-9422(91)85117-I

Cedzich A, Huttenlocher F, Kuhn BM, Pfannstiel J, Gabler L, Stintzi A, Schaller A (2009) The protease-associated (PA) domain and C-terminal extension are required for zymogen processing, sorting within the secretory pathway, and activity of tomato subtilase 3 (SlSBT3). J Biol Chem 284:14068–14078. doi:10.1074/jbc.M900370200

Dalisay DS, Kim KW, Lee C, Yang H, Rübel O, Bowen BP, Davin LB, Lewis NG (2015) Dirigent protein-mediated lignan and cyanogenic glucoside formation in flax seed: integrated omics and MALDI mass spectrometry imaging. J Nat Prod 78:1231–1242. doi:10.1021/acs.jnatprod.5b00023

Davin LB, Lewis NG (2005) Dirigent phenoxy radical coupling: advances and challenges. Curr Opin Biotech 16:398–406. doi:10.1016/j.copbio.2005.06.010

Davin LB, Wang HB, Crowell AL, Bedgar DL, Martin DM, Sarkanen S, Lewis NG (1997) Stereoselective bimolecular phenoxy radical coupling by an auxiliary (dirigent) protein without an active center. Science 275:362–366. doi:10.1126/science.275.5298.362

Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard J-F, Guindon S, Lefort V, Lescot M, Claverie J-M, Gascuel O (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucl Acids Res 36:W465–W469. doi:10.1093/nar/gkn180

Effenberger I, Zhang B, Li L, Wang Q, Liu Y, Klaiber I, Pfannstiel J, Wang Q, Schaller A (2015) Dirigent proteins from cotton (*Gossypium* sp.) for the atropselective synthesis of gossypol. Angew Chem Int Ed Engl 54:14660–14663. doi:10.1002/anie.201507543

Fang W, Ji S, Jiang N, Wang W, Zhao GY, Zhang S, Ge HM, Xu Q, Zhang AH, Zhang YL, Song YC, Zhang J, Tan RX (2012) Naphthol radical couplings determine structural features and enantiomeric excess of dalesconols in *Daldinia eschscholzii*. Nature Comm 3:1039. doi:10.1038/ncomms2031

Frankfater CR, Dowd MK, Triplett BA (2009) Effect of elicitors on the production of gossypol and methylated gossypol in cotton hairy roots. Plant Cell Tiss Organ Cult 98:341–349. doi:10.1007/s11240-009-9568-0

Funatsuki H, Suzuki M, Hirose A, Inaba H, Yamada T, Hajika M, Komatsu K, Katayama T, Sayama T, Ishimoto M, Fujino K (2014) Molecular basis of a shattering resistance boosting global dissemination of soybean. Proc Natl Acad Sci U S A 111:17797–17802. doi:10.1073/pnas.1417282111

Gang DR, Costa MA, Fujita M, Dinkova-Kostova AT, Wang H-B, Burlat V, Martin W, Sarkanen S, Davin LB, Lewis NG (1999) Regiochemical control of monolignol radical coupling: a new paradigm for lignin and lignan biosynthesis. Chem Biol 6:143–151. doi:10.1016/S1074-5521(99)89006-1

Gao W, Long L, Zhu L-F, Xu L, Gao W-H, Sun L-Q, Liu L-L, Zhang X-L (2013) Proteomic and virus-induced gene silencing (VIGS) analyses reveal that gossypol, brassinosteroids, and jasmonic acid contribute to the resistance of cotton to *Verticillium dahliae*. Mol Cell Proteomics 12:3690–3703. doi:10.1074/mcp.M113.031013

Gerardy R, Zenk MH (1992) Formation of salutaridine from (*R*)-reticuline by a membrane-bound cytochrome P-450 enzyme from *Papaver somniferum*. Phytochemistry 32:79–86. doi:10.1016/0031-9422(92)80111-Q

Gesell A, Rolf M, Ziegler J, Díaz Chávez ML, Huang F-C, Kutchan TM (2009) CYP719B1 is salutaridine synthase, the C-C phenol-coupling enzyme of morphine biosynthesis in opium poppy. J Biol Chem 284:24432–24442. doi:10.1074/jbc.M109.033373

Gil Girol C, Fisch KM, Heinekamp T, Günther S, Hüttel W, Piel J, Brakhage AA, Müller M (2012) Regio- and stereoselective

oxidative phenol coupling in *Aspergillus niger*. Angew Chem Int Ed 51:9788–9791. doi:10.1002/anie.201203603

Guo W, Jin L, Miao Y, He X, Hu Q, Guo K, Zhu L, Zhang X (2016) An ethylene response-related factor, GbERF1-like, from *Gossypium barbadense* improves resistance to *Verticillium dahliae* via activating lignin synthesis. Plant Mol Biol 91:305–318. doi:10.1007/s11103-016-0467-6

Hagenbucher S, Olson DM, Ruberson JR, Wäckers FL, Romeis J (2013) Resistance mechanisms against arthropod herbivores in cotton and their interactions with natural enemies. Crit Rev Plant Sci 32:458–482. doi:10.1080/07352689.2013.809293

Halls SC, Davin LB, Kramer DM, Lewis NG (2004) Kinetic study of coniferyl alcohol radical binding to the (+)-pinoresinol forming dirigent protein. Biochemistry 43:2587–2595. doi:10.1021/bi035959o

Hosmani PS, Kamiya T, Danku J, Naseer S, Geldner N, Guerinot ML, Salt DE (2013) Dirigent domain-containing protein is part of the machinery required for formation of the lignin-based Casparian strip in the root. Proc Natl Acad Sci U S A 110:14498–14503. doi:10.1073/pnas.1308412110

Howard-Jones AR, Walsh CT (2006) Staurosporine and Rebeccamycin aglycones are assembled by the oxidative action of StaP, StaC, and RebC on chromopyrrolic acid. J Am Chem Soc 128:12289–12298. doi:10.1021/ja063898m

Hron R, Kim H, Calhoun M, Fisher G (1999) Determination of (+)-, (−)-, and total gossypol in cottonseed by high-performance liquid chromatography. J Am Oil Chem Soc 76:1351–1355. doi:10.1007/s11746-999-0149-5

Ikezawa N, Iwasa K, Sato F (2008) Molecular cloning and characterization of CYP80G2, a cytochrome P450 that catalyzes an intramolecular C–C phenol coupling of (S)-reticuline in magnoflorine biosynthesis, from cultured *Coptis japonica* cells. J Biol Chem 283:8810–8821. doi:10.1074/jbc.M705082200

Jacobs P, Inan M, Festjens N, Haustraete J, Van Hecke A, Contreras R, Meagher M, Callewaert N (2010) Fed-batch fermentation of GM-CSF-producing glycoengineered *Pichia pastoris* under controlled specific growth rate. Microb Cell Factories 9:93. doi:10.1186/1475-2859-9-93

Jacobs PP, Geysens S, Vervecken W, Contreras R, Callewaert N (2009) Engineering complex-type N-glycosylation in *Pichia pastoris* using GlycoSwitch technology. Nat Protoc 4:58–70

Jaroszewski JW, Strøm-Hansen T, Hansen SH, Thastrup O, Kofod H (1992) On the botanical distribution of chiral forms of gossypol. Planta Med 58:454–458. doi:10.1055/s-2006-961512

Kazenwadel C, Klebensberger J, Richter S, Pfannstiel J, Gerken U, Pickel B, Schaller A, Hauer B (2013) Optimized expression of the dirigent protein AtDIR6 in *Pichia pastoris* and impact of glycosylation on protein structure and function. Appl Microbiol Biotechnol 97:7215–7227. doi:10.1007/s00253-012-4579-x

Keshmiri-Neghab H, Goliaei B (2014) Therapeutic potential of gossypol: an overview. Pharmaceut Biol 52:124–128. doi:10.3109/13880209.2013.832776

Kim K-W, Moinuddin SGA, Atwell KM, Costa MA, Davin LB, Lewis NG (2012) Opposite stereoselectivities of dirigent proteins in *Arabidopsis* and *Schizandra* species. J Biol Chem 287:33957–33972. doi:10.1074/jbc.M112.387423

Kim KW, Smith CA, Daily MD, Cort JR, Davin LB, Lewis NG (2015) Trimeric structure of (+)-pinoresinol-forming dirigent protein at 1.95 A resolution with three isolated active sites. J Biol Chem 290:1308–1318. doi:10.1074/jbc.M114.611780

Kim MK, Jeon J-H, Fujita M, Davin LB, Lewis NG (2002) The western red cedar (*Thuja plicata*) 8-8' DIRIGENT family displays diverse expression patterns and conserved monolignol coupling specificity. Plant Mol Biol 49:199–214. doi:10.1023/A:1014940930703

Kozlowski MC, Morgan BJ, Linton EC (2009) Total synthesis of chiral biaryl natural products by asymmetric biaryl coupling. Chem Soc Rev 38:3193–3207. doi:10.1039/b821092f

Laemmli UK (1970) Cleavage of structural proteins during assembly of the head of bacteriophage T4. Nature 227:680–685. doi:10.1038/227680a0

Lan L, Appelman C, Smith AR, Yu J, Larsen S, Marquez RT, Liu H, Wu X, Gao P, Roy A, Anbanandam A, Gowthaman R, Karanicolas J, De Guzman RN, Rogers S, Aubé J, Ji M, Cohen RS, Neufeld KL, Xu L (2015) Natural product (−)-gossypol inhibits colon cancer cell growth by targeting RNA-binding protein Musashi-1. Mol Oncol 9:1406–1420. doi:10.1016/j.molonc.2015.03.014

Li C, Lei X (2014) Strategies toward the biomimetic syntheses of oligomeric sesquiterpenoids. J Org Chem 79:3289–3295. doi:10.1021/jo5002092

Li Z, Rupasinghe SG, Schuler MA, Nair SK (2011) Crystal structure of a phenol-coupling P450 monooxygenase involved in teicoplanin biosynthesis. Proteins 79:1728–1738. doi:10.1002/prot.22996

Liu J, Stipanovic RD, Bell AA, Puckhaber LS, Magill CW (2008) Stereoselective coupling of hemigossypol to form (+)-gossypol in moco cotton is mediated by a dirigent protein. Phytochemistry 69:3038–3042. doi:10.1016/j.phytochem.2008.06.007

Mazzaferro LS, Hüttel W, Fries A, Müller M (2015) Cytochrome P450-catalyzed regio- and stereoselective phenol coupling of fungal natural products. J Am Chem Soc 137:12289–12295. doi:10.1021/jacs.5b06776

Niemetz R, Gross GG (2003a) Ellagitannin biosynthesis: laccase-catalyzed dimerization of tellimagrandin II to cornusiin E in *Tellima grandiflora*. Phytochemistry 64:1197–1201. doi:10.1016/j.phytochem.2003.08.013

Niemetz R, Gross GG (2003b) Oxidation of pentagalloylglucose to the ellagitannin, tellimagrandin II, by a phenol oxidase from *Tellima grandiflora* leaves. Phytochemistry 62:301–306. doi:10.1016/S0031-9422(02)00557-5

Petersen M, Alfermann AW (2001) The production of cytotoxic lignans by plant cell cultures. Appl Microbiol Biotechnol 55:135–142. doi:10.1007/s002530000510

Pickel B, Constantin M-A, Pfannstiel J, Conrad J, Beifuss U, Schaller A (2010) An enantiocomplementary dirigent protein for the enantioselective laccase-catalyzed oxidative coupling of phenols. Angew Chem Int Ed Engl 49:202–204. doi:10.1002/anie.200904622

Pickel B, Pfannstiel J, Steudle A, Lehmann A, Gerken U, Pleiss J, Schaller A (2012) A model of dirigent proteins derived from structural and functional similarities with allene oxide cyclase and lipocalins. FEBS J 279:1980–1993. doi:10.1111/j.1742-4658.2012.08580.x

Pickel B, Schaller A (2013) Dirigent proteins: molecular characteristics and potential biotechnological applications. Appl Microbiol Biotechnol 97:8427–8438. doi:10.1007/s00253-013-5167-4

Präg A, Grüning BA, Häckh M, Lüdeke S, Wilde M, Luzhetskyy A, Richter M, Luzhetska M, Günther S, Müller M (2014) Regio- and stereoselective intermolecular oxidative phenol coupling in *Streptomyces*. J Am Chem Soc 136:6195–6198. doi:10.1021/ja501630w

Puckhaber LS, Dowd MK, Stipanovic RD, Howell CR (2002) Toxicity of (+)- and (−)-gossypol to the plant pathogen, *Rhizoctonia solani*. J Agr Food Chem 50:7017–7021. doi:10.1021/jf0207225

Raab D, Graf M, Notka F, Schödl T, Wagner R (2010) The GeneOptimizer algorithm: using a sliding window approach to cope with the vast sequence space in multiparameter DNA sequence optimization. Syst Synth Biol 4:215–225. doi:10.1007/s11693-010-9062-3

Rappsilber J, Ishihama Y, Mann M (2003) Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. Anal Chem 75:663–670

Rayon C, Lerouge P, Faye L (1998) The protein N-glycosylation in plants. J Exp Bot 49:1463–1472. doi:10.1093/jxb/49.326.1463

Schlauer J, Rückert M, Wiesen B, Herderich M, Assi LA, Haller RD, Bär S, Fröhlich K-U, Bringmann G (1998) Characterization of enzymes from *Ancistrocladus* (Ancistrocladaceae) and *Triphyophyllum* (Dioncophyllaceae) catalyzing oxidative coupling of naphthylisoquinoline alkaloids to michellamines. Arch Biochem Biophys 350:87–94. doi:10.1006/abbi.1997.0494

Seneviratne HK, Dalisay DS, Kim KW, Moinuddin SG, Yang H, Hartshorn CM, Davin LB, Lewis NG (2015) Non-host disease resistance response in pea (*Pisum sativum*) pods: biochemical function of DRR206 and phytoalexin pathway localization. Phytochemistry 113:140–148. doi:10.1016/j.phytochem.2014.10.013

Shevchenko A, Wilm M, Vorm O, Mann M (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. Anal Chem 68:850–858. doi:10.1021/ac950914h

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, Higgins DG (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539–10.1038/msb.2011.75

Smith ET, Perry ET, Sears MB, Johnson DA (2014) Expression of recombinant human mast cell chymase with Asn-linked glycans in glycoengineered *Pichia pastoris*. Protein Expr Purif 102:69–75. doi:10.1016/j.pep.2014.08.005

Stipanovic RD, López JD Jr, Dowd MK, Puckhaber LS, Duke SE (2006) Effect of racemic and (+)- and (−)-gossypol on the survival and development of *Helicoverpa zea* larvae. J Chem Ecol 32:959–968. doi:10.1007/s10886-006-9052-9

Stipanovic RD, López JD Jr, Dowd MK, Puckhaber LS, Duke SE (2008) Effect of racemic, (+)- and (−)-gossypol on survival and development of *Heliothis virescens* larvae. Environ Entomol 37:1081–1085. doi:10.1603/0046-225X(2008)37[1081:EORAGO]2.0.CO;2

Tretter V, Altmann F, MÄRz L (1991) Peptide-N4-(N-acetyl-β-glucosaminyl)asparagine amidase F cannot release glycans with fucose attached α1 → 3 to the asparagine-linked N-acetylglucosamine residue. Eur J Biochem 199:647–652. doi:10.1111/j.1432-1033.1991.tb16166.x

Vassao DG, Kim K-W, Davin LB, Lewis NG (2010) Lignans (neolignans) and allyl/propenyl phenols: biogenesis, structural biology, and biological/human health considerations. In: Townsend CA (ed) Comprehensive natural products II chemistry and biology, vol 1. Elsevier, Amsterdam, pp. 815–928

Vervecken W, Kaigorodov V, Callewaert N, Geysens S, De Vusser K, Contreras R (2004) *In vivo* synthesis of mammalian-like, hybrid-type N-glycans in *Pichia pastoris*. Appl Environ Microb 70:2639–2646. doi:10.1128/AEM.70.5.2639-2646.2004

Wagner TA, Liu J, Stipanovic RD, Puckhaber LS, Bell AA (2012) Hemigossypol, a constituent in developing glanded cottonseed (*Gossypium hirsutum*). J Agr Food Chem 60:2594–2598. doi:10.1021/jf2051366

Wang X, Howell CP, Chen F, Yin J, Jiang Y (2009) Gossypol—a polyphenolic compound from cotton plant. In: Taylor SL (ed) Advances in food and nutrition research, vol 58. Academic Press, New York, pp. 215–263

Wezeman T, Brase S, Masters K-S (2015) Xanthone dimers: a compound family which is both common and privileged. Nat Prod Rep. doi:10.1039/c4np00050a

Wilson IB, Zeleny R, Kolarich D, Staudacher E, Stroop CJ, Kamerling JP, Altmann F (2001) Analysis of Asn-linked glycans from vegetable foodstuffs: widespread occurrence of Lewis a, core alpha1,3-linked fucose and xylose substitutions. Glycobiology 11:261–274. doi:10.1093/glycob/11.4.261

Zhou M, Zhang C, Wu Y, Tang Y (2013) Metabolic engineering of gossypol in cotton. Appl Microbiol Biotechnol 97:6159–6165. doi:10.1007/s00253-013-5032-5

Zhou QL (2011) Privileged chiral ligands and catalysts. Wiley, New York

Zhu L, Zhang X, Tu L, Zeng F, Nie Y, Guo X (2007) Isolation and characterization of two novel dirigent-like genes highly induced in cotton (*Gossypium barbadense* and *G. hirsutum*) after infection by *Verticillium dahliae*. J Plant Pathol 89:41–45. doi:10.4454/jpp.v89i1.722