Phytochemistry xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Phytochemistry

journal homepage: www.elsevier.com/locate/phytochem

Biosynthesis of natural and novel *C*-glycosylflavones utilising recombinant *Oryza sativa C*-glycosyltransferase (OsCGT) and *Desmodium incanum* root proteins

B. Hao [a,b], J.C. Caulfield [b], M.L. Hamilton [b], J.A. Pickett [b], C.A.O. Midega [c], Z.R. Khan [c], J. Wang [a,*], A.M. Hooper [b,*]

[a] College of Science, Northwest A&F University, Yangling, Shaanxi 712100, China
[b] Department of Biological Chemistry and Crop Protection, Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK
[c] International Centre of Insect Physiology and Ecology, P.O. Box 30772, Nairobi, Kenya

ARTICLE INFO

Article history:
Received 7 December 2015
Received in revised form 12 January 2016
Accepted 19 February 2016
Available online xxxx

Keywords:
Oryza sativa
Poaceae
C-Glycosyltransferase
OsCGT
2-Hydroxyflavanone
C-Glucosylflavone
Di-*C*-glycosylflavone
Biotechnological synthesis
Substrate specificity
Chemical space

ABSTRACT

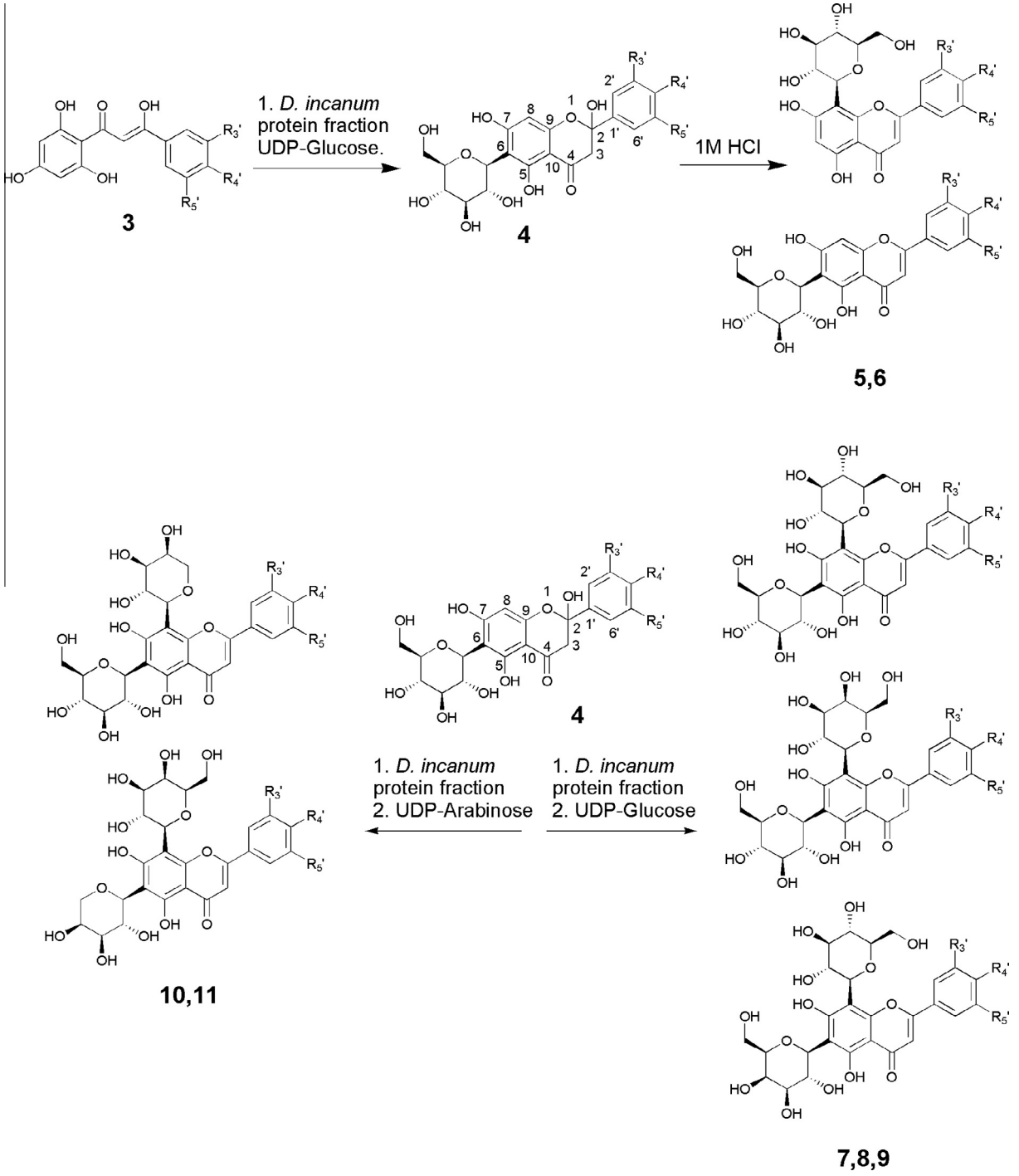
The rice *C*-glycosyltransferase (OsCGT) is one of only a small number of characterised plant *C*-glycosyltransferases (CGT) known. The enzyme *C*-glucosylates a 2-hydroxyflavanone substrate with UDP-glucose as the sugar donor to produce *C*-glucosyl-2-hydroxyflavanones. We tested substrate specificity of the enzyme, using synthetic 2-hydroxyflavanones, and showed it has the potential to generate known natural CGFs that have been isolated from rice and also other plants. In addition, we synthesised novel, unnatural 2-hydroxyflavanone substrates to test the B-ring chemical space of substrate accepted by the OsCGT and demonstrated the OsCGT capacity as a synthetic reagent to generate significant quantities of known and novel CGFs. Many B-ring analogues are tolerated within a confined steric limit. Finally the OsCGT was used to generate novel mono-*C*-glucosyl-2-hydroxyflavanones as putative biosynthetic intermediates to examine the potential of *Desmodium incanum* biosynthetic CGTs to produce novel di-*C*-glycosylflavones, compounds implicated in the allelopathic biological activity of *Desmodium* against parasitic weeds from the *Striga* genus.

1. Introduction

In our work to characterise the allelopathic root exudates of the *Desmodium* genus and their impact on *Striga* inhibition in subsistence farming technologies in East Africa (Khan et al., 2002, 2014), we have elucidated a number of *C*-glycosylflavones (CGFs) from the active root exudates (Hooper et al., 2015). In addition their biosynthesis *in planta* was elucidated and shown to be initiated by a *C*-glycosyltransferase (CGT) that acts upon 2-hydroxynaringenin (**3a**) using the sugar donor UDP-glucose (Hooper et al., 2009) (Scheme 1). Subsequent *C*-glycosylations act upon the product of this *C*-glucosylation, 6-*C*-glucosyl-2-hydroxynaringenin (**4a**), and produce di-*C*-glycosylflavones (**7a**–**11a**) through the addition of UDP-galactose, UDP-arabinose or UDP-glucose (Hamilton et al., 2012; Hao et al., 2015). The rice *C*-glycosyltransferase (OsCGT) is one of only a small number of characterised plant *C*-glycosyltransferases (CGT) known (Brazier-Hicks et al., 2009) and catalyses the *C*-glucosylation of 2-hydroxyflavanones with UDP-glucose as do plants from the *Desmodium* genus. Recently the field has expanded to include two characterised CGTs from *Fagopyrum esculentum* (UGT708C1, UGT708C2) (Nagatomo et al., 2014), a bifunctional *C*- and *O*-glucosyltransferase from *Zea mays* (UGT708A6) (Ferreyra et al., 2013), and one from *Glycine max* (UGT708D1) (Hirade et al., 2015), *Gentiana triflora* (GtUF6CGT1) (Sasaki et al., 2015) and *Mangifera indica* (MiUGT13) (Chen et al., 2015). The CGTs UGT708C1, UGT708C2, UGT708A6, UGT708D1 and GtUF6CGT1 are all capable of producing CGFs although GtUF6CGT1 does so *via* a direct flavone glycosylation reaction and not by *C*-glucosylation of a 2-hydroxyflavanone. The biological activities reported for the CGF class of molecules is various, hence the renewed interest in recent years in this class of chemistry and the enzymes responsible for their production. CGFs are relevant to

* Corresponding authors.
E-mail addresses: bing.hao@hotmail.com (B. Hao), john.caulfield@rothamsted.ac.uk (J.C. Caulfield), mary.hamilton@rothamsted.ac.uk (M.L. Hamilton), john.pickett@rothamsted.ac.uk (J.A. Pickett), cmidega@icipe.org (C.A.O. Midega), zkhan@icipe.org (Z.R. Khan), wangjunru@nwsuaf.edu.cn (J. Wang), tony.hooper@rothamsted.ac.uk (A.M. Hooper).

http://dx.doi.org/10.1016/j.phytochem.2016.02.013
0031-9422/



Scheme 1. Production of di-*C*-glycosides by the soluble root protein of *D. incanum* and suitable sugar donor cofactors from the subtrate generated by recombinant OsCGT.

plant ecology, as allelochemicals against *Striga* (Hooper et al., 2010, 2009), as phytoalexins (McNally et al., 2003), insect antifeedants (Byrne et al., 1996; Chul-sa et al., 2009), soil-plant interactions with arbuscular mycorrhizal fungi (Akiyama et al., 2002) and other pharmacological properties (Talhi and Silva, 2012) such as antioxidant activities (Kitta et al., 1992; Kim et al., 2005) and protection against obesity and cancer (Lee et al., 2010; Nichenametla et al., 2006). We now aim to exploit the potential of the *Oryza sativa*, rice OsCGT in synthetic applications and in biosynthetic studies on the production of CGFs *in planta*.

We examined the substrate specificity of the recombinant His-tagged OsCGT enzyme through a range of synthetic 2-hydroxyflavanones (**3a**–**x**), to investigate whether it had the capacity (i) to generate other natural *C*-glucosylflavones found *in planta*, either from rice or other plants; (ii) to generate novel *C*-glucosylflavones to probe the B-ring chemical space tolerated by this enzyme and to utilise it as a synthetic tool; (iii) to investigate whether sufficient flexibility exists to build a suitable scaffold biotechnologically for further functionalization; (iv) to use OsCGT to generate mono-*C*-glucosyl-2-hydroxyflavanone intermediates (**4**) to probe

Please cite this article in press as: Hao, B., et al. Biosynthesis of natural and novel *C*-glycosylflavones utilising recombinant *Oryza sativa* *C*-glycosyltransferase (OsCGT) and *Desmodium incanum* root proteins. Phytochemistry (2016), http://dx.doi.org/10.1016/j.phytochem.2016.02.013

the capacity of *D. incanum* to generate novel di-*C*-glycosylflavones, allelopathic compounds produced in *Desmodium* root exudates that inhibit *Striga* parasitism of subsistence cereal crops (Hooper et al., 2015).

2. Results and discussion

2.1. Synthesis of 2-hydroxyflavonone substrates

Novel and nature-identical 2-hydroxyflavanones were synthesised as previously reported according to the synthetic strategy shown (Scheme 2) (Hamilton et al., 2009). The divergent chemical synthesis involves preparation of 2,4-dibenzyloxy-6-hydroxyacetophenone which is esterified by a range of aromatic acids to generate **1a**–**x** which undergo a Baker–Venkataraman rearrangement to the subsequent propanediones **2a**–**x**. These were deprotected by one of two methods. The first was using 10% Pd on charcoal under a hydrogen atmosphere after which filtration and purification on silica gel furnished the substrate **3a**–**t**. However, some substrates were not stable to this and so were produced by using platinum dihydroxide under hydrogen, evaporating the solvent and washing the catalyst with ethanol under nitrogen to obtain the products **3u**–**x** that were separated from the catalyst using a centrifuge and therefore not purified before immediate use.

2.2. Sugar substrate specificity

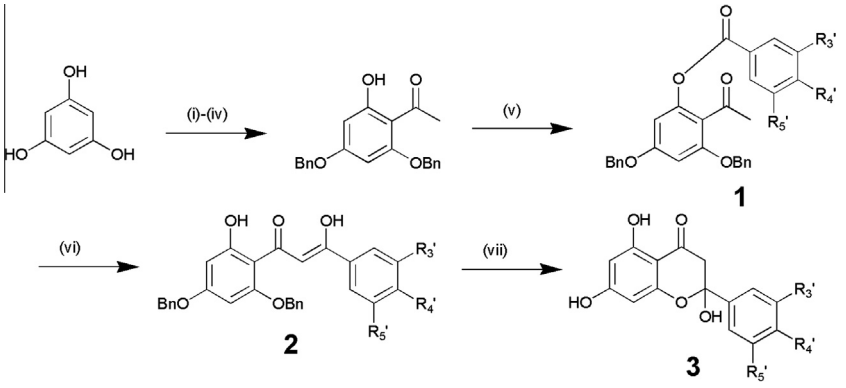
The substrate specificity of OsCGT was assessed by challenging the enzyme with synthetic 2-hydroxynaringenin (**3a**) and a range of commercially available sugar donors. No new products were found using UDP-glucuronic acid, UDP-galactose, UDP-arabinose or UDP-*N*-acetylglucosamine and so OsCGT showed specificity to UDP-glucose as the sugar donor in its glycosyltransferase activity.

2.3. OsCGT C-Glucosylation of 2-hydroxyflavanones derived from known natural product structures

Oryza sativa in rice produces CGFs from 2-hydroxynaringenin (**3a**) and from 2-hydroxyeriodictyol (**3b**), which are *C*-glucosylated to vitexin **5a**, isovitexin **6a**, orientin **5b** and isoorientin **6b**, products that may be further glycosylated to more complex metabolites (Brazier-Hicks et al., 2009). Rice also produces 3′,5′-dimethoxyapigenin (tricin) but no *C*-glucosylated derivatives of that B-ring structure, only two *O*-glucosylated derivatives. This may be because the intermediate **3e** is not present or not available for *C*-glucosylation. Any late-stage B-ring oxidation of a flavone generates a substrate that cannot be *C*-glucosylated as *C*-glucosylation must occur before flavone formation. *C*-glycosylflavonoid plant natural products derived from the substrates **3c**–**g** and not found in rice are affinetin (**5c**) (*Trichomanes venosum*), isoaffinetin (**6c**) (Polygonum affine), scoparin (**5d**) (*Sarothamnus scoparius*), isoscoparin (**6d**) (*Hordeum vulgare*), pyrenin (**5e**) (*Gentiana pyrenaica*), isopyrenin (**6e**) (*Gentiana pyrenaica*), cytisocise (**5f**) (*Cytisus laburnum*), isocytisocide (**6f**) (*Fortunella margarita*), 8-*C*-glucosylchrysin (**5g**) (*Scutellaria amoena*), 6-*C*-glucosylchrysin (**6g**) (*Scutellaria baicalensis*) and many more natural products derived from these parent molecules. We therefore used the appropriate 2-hydroxyflavanone precursors **3a**–**g** to test the OsCGT substrate specificity to 2-hydroxyflavanones with varying hydroxyl and methoxyl groups that may be present naturally in other plants. The substrates were incubated with OsCGT and UDP-glucose as before and all of them were efficiently *C*-glucosylated by the OsCGT enzyme with good yields (Table 1). The yields were obtained by HPLC integration of the products at 350 nm, close to the flavone maximum, and show the enzyme could generate *C*-glucosyl-2-hydroxyflavanones **4a**–**g**, if the analogous 2-hydroxyflavanone were present, which were dehydrated chemically to **5a**–**5g** and **6a**–**6g** (Scheme 3). In order to verify that the products were the expected 6-*C*- and 8-*C*-glucosylflavones, the enzyme reactions involving substrates **3a**, **3b**, **3d**, **3f**, **3g** were scaled up and the products isolated for NMR spectroscopic characterisation. The generation of milligram quantities shows that in some instances, where yields are high, the use of OsCGT as a synthetic reagent is feasible. The 2-hydroxyflavanones known in nature from rice are **3a** and **3b**, however, the capacity of a CGT to *C*-glucosylate other 2-hydroxyflavanones that would give rise to CGFs found as natural products means that this biosynthetic possibility, as well as late stage B-ring oxidation, cannot be precluded.

2.4. OsCGT C-glucosylation of novel 2-hydroxyflavanones

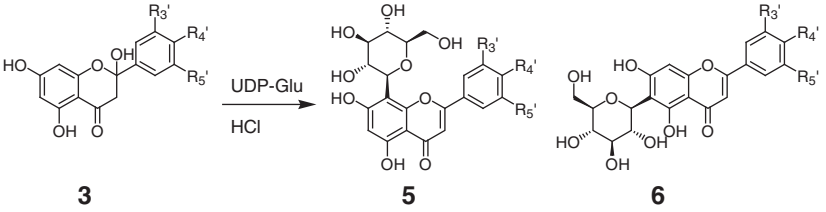
Compounds **3h**–**x** were prepared to examine further the tolerance of the OsCGT to B-ring analogues of its natural substrate and assays with OsCGT were conducted in the same way as described above. Substrates **3g** and **3o**–**q** show a sequence of increasing steric bulk of an alkyl group in the R4′ position and as the steric bulk increases, the yield of CGFs drops with a *t*-butyl group (**3q**) implying that this group surpasses the steric constraints that are acceptable to the enzyme. This is verified by substrates **3a**, **3f**, **3k**, **3l**, **3m**, **3n** which have an increasing alkoxy steric bulk at R4′. This time, as steric bulk increases, so the yield decreases through the series until the limit is reached between a C_3 (**3l**) and C_6 (**3m**,**n**) group. The electronic requirements of the B-ring were also tested using substrates **3r**–**3x** through the addition of halogen moieties. Interestingly, all the fluorinated substrates were



Scheme 2. (i) Ac_2O, pyridine; (ii) NaH, H_2O, BnBr, 99% (2 steps); (iii) TFA, TFAA, 87%; (iv) $TiCl_4$, 76%; (v) EDC, DMAP, $ArCO_2H$, 48–99%; (vi) NaOH, pyridine, 7–70%; (vii) 10% Pd/C or $Pt(OH)_2$, 19–84%.

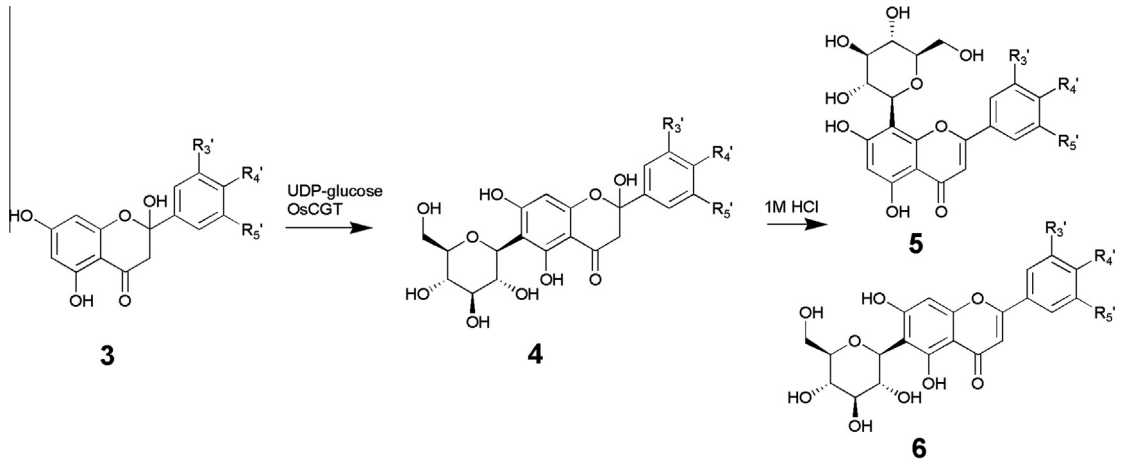
Please cite this article in press as: Hao, B., et al. Biosynthesis of natural and novel *C*-glycosylflavones utilising recombinant *Oryza sativa* *C*-glycosyltransferase (OsCGT) and *Desmodium incanum* root proteins. Phytochemistry (2016), http://dx.doi.org/10.1016/j.phytochem.2016.02.013

Table 1
Generation of natural product mono-*C*-glucosylflavones from 2-hydroxyflavanones and UDP-glucose by the *Oryza sativa* OsCGT.



	R3′	R4′	R5′	8-*C*-Glucoside %	6-*C*-Glucoside %
3a	H	OH	H	**5a**; 44*	**6a**; 53*
3b	OH	OH	H	**5b**; 43*	**6b**; 56*
3c	OH	OH	OH	**5c**; 40	**6c**; 27
3d	OMe	OH	H	**5d**; 41*	**6d**; 53*
3e	OMe	OH	OMe	**5e**; 69†	**6e**; 21
3f	H	OMe	H	**5f**; 44*	**6f**; 52*
3g	H	H	H	**5g**; 33*	**6g**; 47*

* Products characterised by NMR spectroscopy.
† Not pure by NMR spectroscopy.



Scheme 3. Synthesis of glucosylated 2-hydroxyflavanones using OsCGT and dehydration to produce **4** for biosynthetic experiments with *Desmodium incanum* proteins or **5** and **6** as standards for characterization.

efficiently *C*-glucosylated, despite having an electron withdrawing and deactivating effect on the ring which is contrary to the electron donating and activating natural substrates. Chlorinated and brominated analogues were not turned over by the OsCGT. As before, using substrates **3o**, **3p**, **3r** and **3t**, where yield were high, products were isolated from scaling up enzyme assays and the products characterised by NMR spectroscopy to verify structure. Substrates **3u**–**3x** were designed to produce halogenated CGFs that may be used to attach biologically active glucosylated flavones to other biologically active compounds through Pd-catalysed cross-coupling chemistry. However, these results mean using OsCGT as a synthetic tool in this respect appears not to be feasible.

2.5. Synthesis of 6- and 8-C-glucosylflavones from synthetic 2-hydroxyflavanones by D. incanum soluble root proteins

In our previous work demonstrating the biosynthesis of allelopathic *C*-glycosylflavonoids found in the root exudates of *Desmodium spp*., we demonstrated *C*-glucosylation in this genus requires a 2-hydroxyflavonone substrate (Hamilton et al., 2009). We therefore challenged the soluble root proteins of *D. incanum* with the synthetic 2-hydroxyflavanones to assess whether there are significant differences in the capacity of *D. incanum* to *C*-glucosylate 2-hydroxyflavanones compared to the rice recombinant OsCGT protein (Table 3). Where standards were not available from the experiments with OsCGT which allowed NMR characterisation, the two products were analysed by high cone voltage ESIMS experiments to verify the *C*-glucosyl link to the flavonoid core through the loss of 90 and 120 amus (Fig. 1) (Cuyckens and Claeys, 2004) (Supplementary data). An example of the mass spectrometric analysis is shown for incubation assays with **3h** (3′,4′-methylenedioxy) (Fig. 2) and in this case the products are fully characterised as **5h** and **6h**. The differences between the two systems are small but relevant. OsCGT cannot *C*-glucosylate the sterically bulky **3n** (*t*-butyl) and **3q** (phenyl) while *D. incanum* can, implying a relaxed steric constraint. However, this is minimal as **3p** (hexamethylenoxyl) still cannot be *C*-glycosylated by *D. incanum* as is the case with OsCGT. Minor differences include the acceptance by OsCGT of a

Please cite this article in press as: Hao, B., et al. Biosynthesis of natural and novel *C*-glycosylflavones utilising recombinant *Oryza sativa C*-glycosyltransferase (OsCGT) and *Desmodium incanum* root proteins. Phytochemistry (2016), http://dx.doi.org/10.1016/j.phytochem.2016.02.013

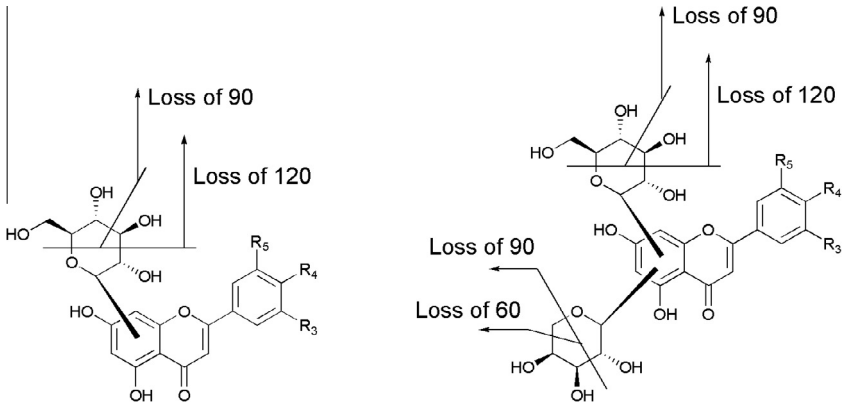


Fig. 1. Fragmentation of *C*-hexosides and *C*-pentosides under high cone voltage ESIMS conditions to generate fragments losses of 60, 90 and 120 amus.

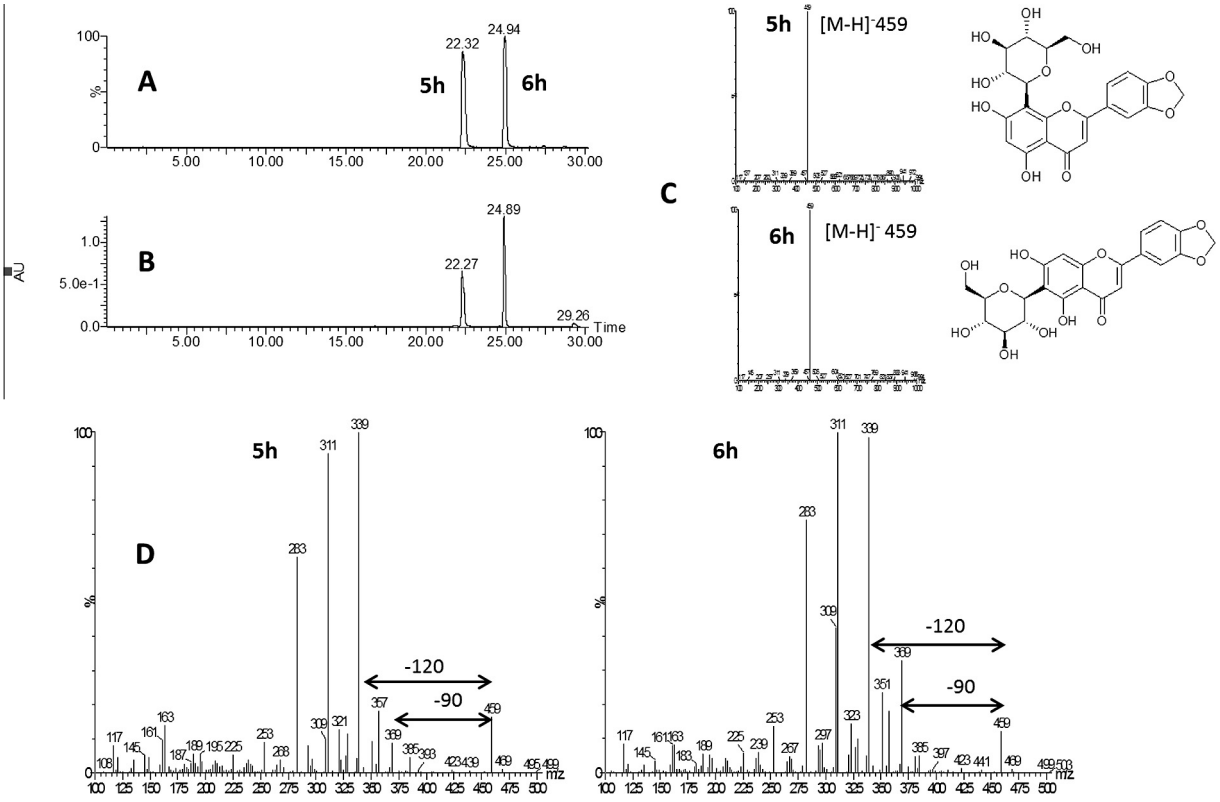


Fig. 2. (A) Ion scan for $[M-H]^-$ 459 showing two new metabolites found in (B) the LCMS ion trace. (C) Molecular ions of two new metabolites and (D) fragmentation under high cone voltage showing *C*-link fragments.

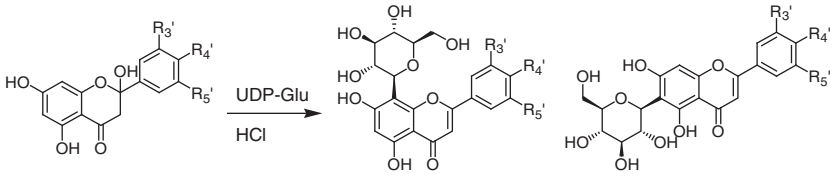
brominated substrate **3v** although at a very low level and **3i** (3′,5′-hydroxy) which was surprisingly not *C*-glycosylated by *D. incanum* proteins. A significant difference was the ability of *D. incanum* proteins to turn over chlorinated substrates which would allow the potential for generating chlorinated biologically active synthons for cross-coupling. These results demonstrate that crude *D. incanum* soluble protein has a different, though very similar capacity to accept novel substrates. These differences may be due to a single protein having a less sterically hindered substrate specificity for the substrate B-ring (**3n** and **3q**) along with some different electronic requirements (**3i**, **3v**, **3w**, and **3x**) or it may be due to different *C*-glycosyltransferases with a combined greater range of acceptable chemical space for 2-hydroxyflavanones. In any case, the plant has a greater capacity for producing novel *C*-glucoslyflavones than those possible using the single rice recombinant OsCGT.

2.6. Synthesis of 6,8-di-C-glycosylflavones using 6-C-glucosyl-2-hydroxyflavanone substrates created by OsCGT and CGT enzymes from D. incanum

The results of incubating synthetic 2-hydroxyflavanones with UDP-glucose and OsCGT (Tables 1 and 2) revealed substrate that can be used to generate *C*-glucosyl-2-hydroxyflavanones (**4**) that may be themselves substrates for a second glycosylations. In our work to understand the biosynthesis of di-*C*-glycosylflavones in the allelopathic root exudates of *D. incanum* we have demonstrated that the soluble plant enzyme extracts can *C*-glucosylate,

Please cite this article in press as: Hao, B., et al. Biosynthesis of natural and novel *C*-glycosylflavones utilising recombinant *Oryza sativa C*-glycosyltransferase (OsCGT) and *Desmodium incanum* root proteins. Phytochemistry (2016), http://dx.doi.org/10.1016/j.phytochem.2016.02.013

Table 2
Generation of novel mono-*C*-glucosylfavones from 2-hydroxyflavanones and UDP-glucose by the *Oryza sativa* OsCGT.

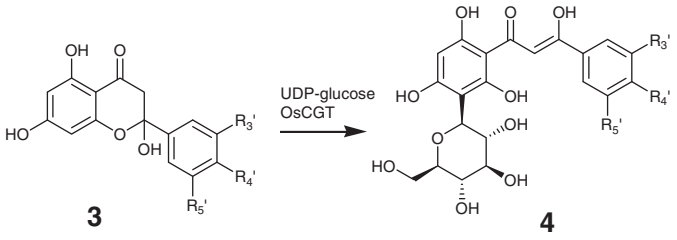


3 **5** **6**

	R3′	R4′	R5′	8-*C*-Glucoside %	6-*C*-Glucoside %
3h	O-CH_2–	–O	H	**5h**; 44*	**6h**; 52*
3i	OH	H	OH	**5i**; 47	**6i**; 43
3j	OH	H	H	**5j**; 14	**6j**; 25
3k	H	OEt	H	**5k**; 34	**6k**; 38
3l	H	O(CH_2)$_3$OH	H	**5l**; 17	**6l**; 27
3m	H	O(CH_2)$_6$OH	H	**5m**; 0	**6m**; 0
3n	H	OPh	H	**5n**; 0	**6n**; 0
3o	H	Me	H	**5o**; 25*	**6o**; 33*
3p	H	Et	H	**5p**; 33*	**6p**; 41*
3q	H	*t*-Bu	H	**5q**; 0	**6q**; 0
3r	F	OH	H	**5r**; 42*	**6r**; 56*
3s	H	CF_3	H	**5s**; 34	**6s**; 42
3t	H	F	H	**5t**; 45*	**6t**; 47*
3u	H	Br	H	**5u**; 0	**6u**; 0
3v	Br	OH	H	**5v**; 4	**6v**; 4
3w	H	Cl	H	**4w**; 0	**6w**; 0
3x	Cl	OH	H	**5x**; 0	**6x**; 0

* Products characterised by NMR spectroscopy.

Table 3
2-Hydroxyflavanones (**3**) and *C*-glucosylflvanones (**4**), generated by OsCGT, introduced to *D. incanum* soluble root proteins with UDP-sugar donors to generate *C*-glucosylflavones (**5**,**6**) or di-*C*-glycosylflavones (**7**–**11**). *(Hao et al., 2015).



3 **4**

	R3′	R4′	R5′	6- and 8-*C*-glucosylflavones (**5**,**6**)		6,8-Di-*C*-hexosylflavones (**7**,**8**,**9**)	6,8-*C*-Glucosyl-*C*-arabinosylflavones (**10**,**11**)
3a	H	OH	H	✓*	**4a**	✓*	✓*
3b	OH	OH	H	✓	**4b**	✓	X
3d	OMe	OH	H	✓	**4d**	✓	X
3e	OMe	OH	OMe	✓	**4e**	X	X
3f	H	OMe	H	✓	**4f**	✓	✓
3g	H	H	H	✓	**4g**	✓	Trace
3h	O-CH_2–	–O	H	✓	**4h**	✓	✓
3i	OH	H	OH	X	**4i**	X	X
3j	OH	H	H	✓	**4j**	–	–
3k	H	OEt	H	✓	**4k**	✓	X
3l	H	O(CH_2)$_3$OH	H	✓	**4l**	X	X
3m	H	O(CH_2)$_6$OH	H	X	**4m**	–	–
3n	H	OPh	H	✓	**4n**	–	–
3o	H	Me	H	✓	**4o**	✓	X
3p	H	Et	H	✓	**4p**	✓	X
3q	H	*t*-Bu	H	Trace	**4q**	–	–
3r	F	OH	H	✓*	**4r**	–	–
3s	H	CF_3	H	✓	**4s**	✓	X
3t	H	F	H	✓*	**4t**	–	–
3v	Br	OH	H	X	**4v**	X	X
3w	H	Cl	H	✓	**4w**	–	–
3x	Cl	OH	H	✓	**4x**	–	–

Please cite this article in press as: Hao, B., et al. Biosynthesis of natural and novel *C*-glycosylflavones utilising recombinant *Oryza sativa* *C*-glycosyltransferase (OsCGT) and *Desmodium incanum* root proteins. Phytochemistry (2016), http://dx.doi.org/10.1016/j.phytochem.2016.02.013

C-galactosylate and *C*-arabinoslyate the mono-*C*-glucosylflavanones **4a** to produce di-*C*-glycosylflavonones that dehydrate to the respective flavones (Hamilton et al., 2012; Hao et al., 2015). We therefore used synthetic 2-hydroxyflavanones that are well accepted by the OsCGT to generate natural and novel *C*-glucosyl-2-hydroxyflavanones (Table 3). By using recombinant OsCGT, these unstable intermediates can be generated in good yields and purity without the need for separation from other flavonoids. The incubation of the substrates with UDP-glucose is expected to give three products (**7**–**9**), rather than one containing two glucose groups, as the *D. incanum* soluble enzyme extract is able to isomerise UDP-glucose to UDP-galactose which is subsequently incorporated and after dehydration can be fixed in either the C-6 or C-8 position to give two regiosiomers (Scheme 1) (Hao et al., 2015). The incorporation of UDP-arabinose likewise produces two products (**10**,**11**), regioisomers with the *C*-arabinose in either the 6 or 8 position. In all cases, where the correct number of novel products was detected with the correct molecular weight, the fragmentation pattern was subsequently carried out under high cone voltage ESIMS conditions to demonstrate that the products are indeed di-*C*-glycosylflavones. In the case of *C*-hexosylflavones, loss of 90 and 120 amus indicate the correct *C*-link and with *C*-arabinosylflavones, loss of 60 and 90 amus is indicative of the correct structure (Fig. 1). The results of these experiment showed that there was a greater relaxed specificity for *C*-glucosylation and *C*-galactosylation than for *C*-arabinosylation. Substrates **4a**,**b**,**d**,**f**,**g**,**h**,**k**,**o**,**p**,**s** all produced 3 new metabolites when incubated with UDP-glucose and *D. incanum* root protein, deduced to be structures **7**, **8** and **9** by ESIMS mass spectrometry (Supplementary data). In the case of **4b**, as the products are already present in *D. incanum* and are co-extracted with the proteins, the labelled sugar moiety UDP-α-D-[UL-$^{13}C_6$]glucose was used and incorporation judged by presence of the labelled $[M-H+6]^-$ 615 ion with six ^{13}C atoms. The ion trace, low cone voltage ESIMS spectrum and fragmented high cone voltage ESIMS spectra from compounds **7h**, **8h**, and **9h** produced from **4h** are shown (Fig. 3). The selective ion trace at 621 amus (Fig. 3A) shows three new metabolites with the correct ion ($[M-H]^-$ 621) which was shown to be the anticipated molecular ion (Fig. 3C). Fragmentation shows sequential losses of 90 ($[M-H-90]^-$ 531 and $[M-H-180]^-$ 441} and 120 amus ($[M-H-120]^-$ 501 and $[M-H-240]^-$ 381) and combinations of losses ($[M-H-90-120]^-$ 411 (Fig. 3D) indicating both sugars must be *C*-linked to the apigenin core (Fig. 1). This has previously been shown for **4a** where the three new metabolites produced were characterised through the NMR characterisation of natural standards as the di-*C*-glucosylated compound and two *C*-galactosylated regioisomers **7a**, **8a** and **9a** (Hooper et al., 2015). The substrates that were not incorporated were **4e** and **4i** which both possess two oxygen atoms in *meta* positions. The more sterically bulky substrate **4l** was also not *C*-hexosylated. In the case of *C*-arabinosylation, however, there

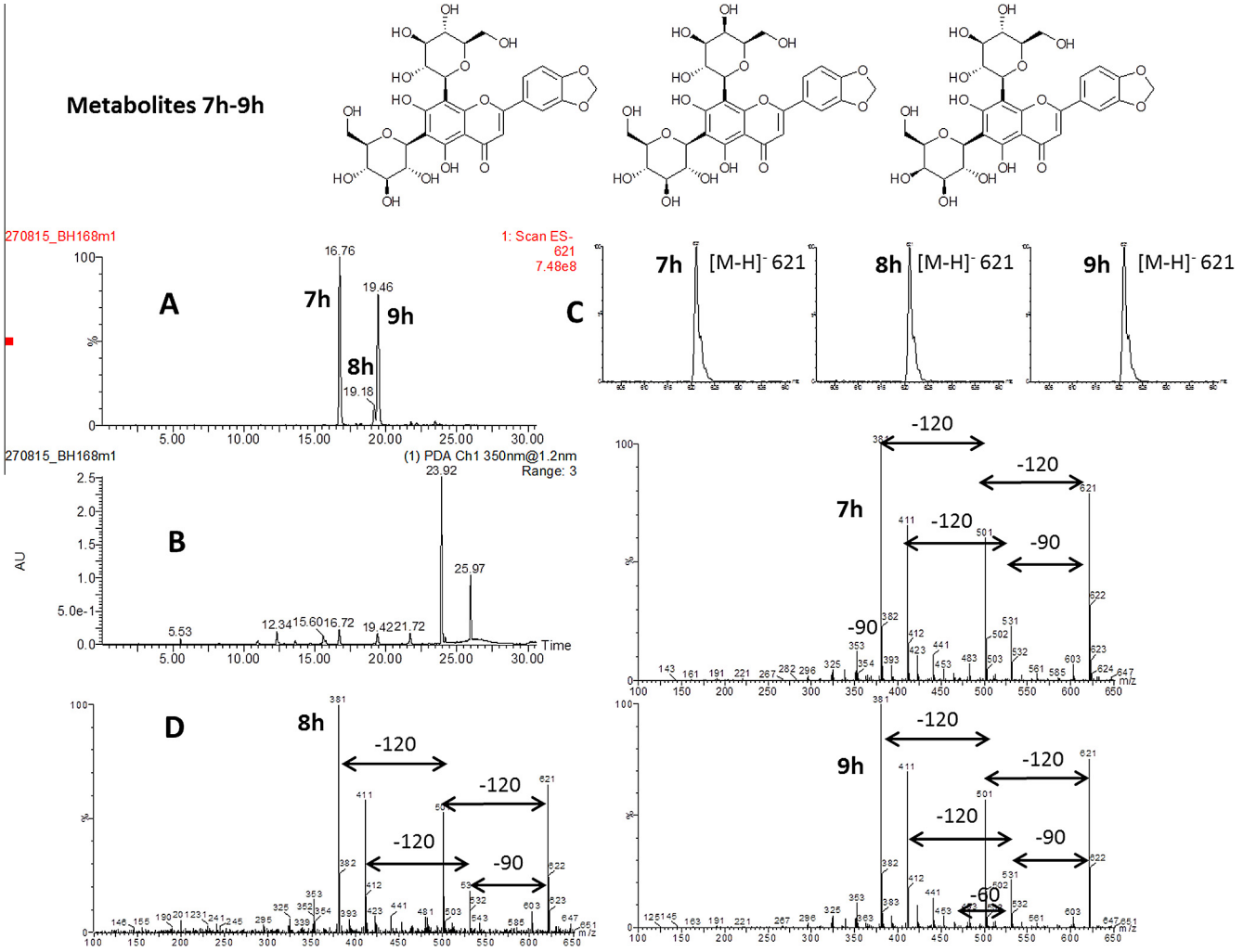


Fig. 3. (A) Ion scan for $[M-H]^-$ 621 showing three new metabolites found in (B) the LCMS ion trace. (C) Molecular ions of three new metabolites and (D) fragmentation under high cone voltage showing *C*-link fragments.

Please cite this article in press as: Hao, B., et al. Biosynthesis of natural and novel *C*-glycosylflavones utilising recombinant *Oryza sativa* *C*-glycosyltransferase (OsCGT) and *Desmodium incanum* root proteins. Phytochemistry (2016), http://dx.doi.org/10.1016/j.phytochem.2016.02.013

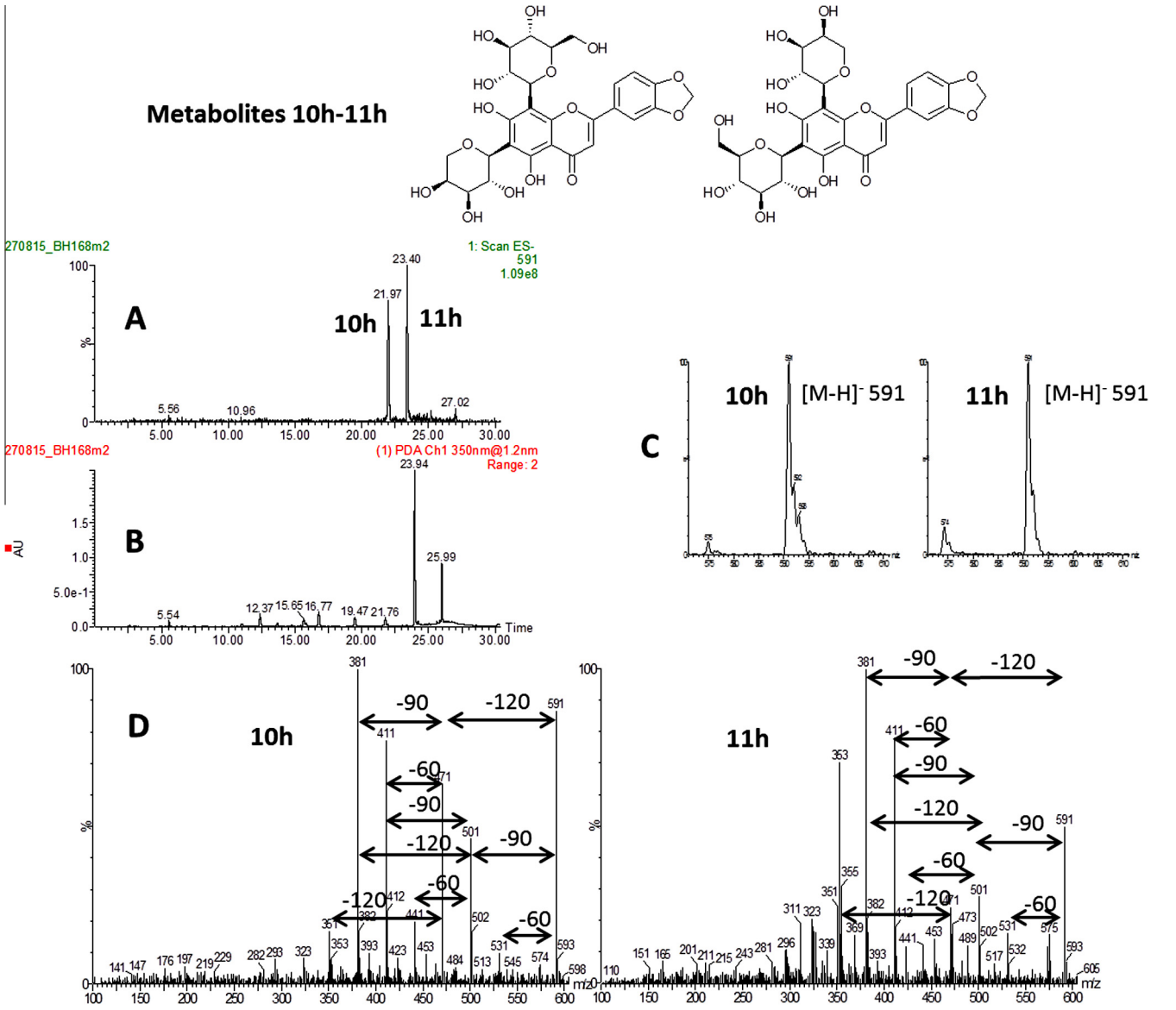


Fig. 4. (A) Ion scan for $[M-H]^-$ 591 showing two new metabolites found in (B) the LCMS ion trace. (C) Molecular ions of two new metabolites and (D) fragmentation under high cone voltage showing *C*-link fragments.

was a distinctly lower tolerance for different B-ring functionality with products only generated from substrates **4f**, **4h** and traces from **4g**, all very close in structure to the natural substrate **4a**. These three substrates were incorporated to yield **10f**,**g**,**h** and **11f**,**g**,**h** (Supplementary data). Representative data (Fig. 4) reveal two compounds in the ion trace (Fig. 4B) with ions of 591 (Fig. 4A) which are revealed as molecular ions $[M-H]^-$ 591 (Fig. 4C) and which fragment to give losses of 60 ($[M-H-60]^-$ 531, losses of 90 $[M-H-90]^-$ 501}, losses of 120 ($[M-H-120]^-$ 471 and combinations of losses ($[M-H-60-90]^-$ 441, $[M-H-60-120]^-$ 411 and $[M-H-90-120]^-$ 381 (Fig. 4D) indicating both sugars must be *C*-linked to the apigenin core (Fig. 1).

3. Conclusions

Work by others (Brazier-Hicks et al., 2009) revealed the substrates most suitable for *C*-glycosylation by the OsCGT enzyme possessed a 2,4,6-trihydroxylated benzene ring with a pendant chain likely to be the open form tautomer of a 2-hydroxyflavanone. They used naturally derived samples of **3a**,**b**,**g**, a dihydroxychalcone, a benzophenone or an ester. Critically, apart from 2-hydroxyflavanones which can exist either as open chain or cyclised forms, only un-cyclised substrates with a 2,4,6-trihydroxy moiety were *C*-glucosylated. Similarly, UGT708C1 and UGT708C2 from *F. esculentum* will *C*-glycosylate 2′,4′,6′-trihydroxyacetophenones and trihydroxydihydrochalcones but not dihydroxyacetophenone (Nagatomo et al., 2014), though the highly promiscuous MiUGT13 and the crude CGT activity demonstrated by *Pueria lobata*, which biosynthesises the *C*-glucosylisoflavone puerarin, both accept dihydroxybenzenoid substrates (Chen et al., 2015, 2010). Consistent with this model, we also synthesised 2,4′,7-trihydroxyflavanone which does not possess the trihydroxybenzenoid moiety and this substrate was not *C*-glucosylated by OsCGT, even though NMR data showed it existed in the open chain tautomeric form. The stability and solution availability of the cyclised tautomers of novel 2-hydroxyflavanones relative to their open chain form may therefore influence the rate at which the substrates are *C*-glucosylated. Our NMR characterisation of 2-hydroxyflavanone substrates (**3**) often had to be conducted at low temperature (−60 °C) to

Please cite this article in press as: Hao, B., et al. Biosynthesis of natural and novel *C*-glycosylflavones utilising recombinant *Oryza sativa* *C*-glycosyltransferase (OsCGT) and *Desmodium incanum* root proteins. Phytochemistry (2016), http://dx.doi.org/10.1016/j.phytochem.2016.02.013

stabilise the cyclised tautomer so preventing tautomeric interconversion and allowing structure confirmation. The flexibility of B-ring analogues being accepted by the rice OsCGT and by active CGTs in the soluble root proteins of *D. incanum* is again consistent with this model as the differences are distal from the open-chain 2,4,6-trihydroxybenzene active site, although analogues must remain within a steric restraint. As that limit is approached there is reduced activity of the CGT both in OsCGT and in the active *D. incanum* protein mixture. However, the added tolerance of different chemistries on the B-ring by *D. incanum* CGTS, in particular halogen introduction, may allow functionalization of a biotechnologically produced scaffold by cross-coupling if recombinant CGTs can be made from *D. incanum*.

The rice OsCGT can accept synthetic 2-hydroxyflavanones that allow it to produce the basic *C*-glucosylflavonoid natural product metabolites found in other plants and not previously in rice. In plants, the generation of functionality on the B-ring may be through late stage modification, particularly oxidation and methylation, of CGFs produced from 2-hydroxynaringenin, but these results reveal the *C*-glucosylation of functionalised natural 2-hydroxyflavanones is also a possibility. In biosynthetic experiments to generate multiply glycosylated flavones using *D. incanum* soluble root proteins, the crude nature of the proteins makes the quantification of products difficult. However, there is clear tolerance on B-ring changes when introducing a second hexose that is not seen in experiments with UDP-arabinose. We suggest that this is due to a different enzyme controlling *C*-arabinosylation as compared with *C*-hexosylation which may be carried out by one CGT that can perform both *C*-glucosylation and *C*-galactosylation or by separate enzymes. Using this knowledge the biotechnological production of novel metabolites *in vitro* or in the field by transgenic plants may be pursued.

The high yields of products in our un-optimised reaction conditions show that as a synthetic reagent, the rice OsCGT has the potential to generate a class of compounds to which many biological activities have been attributed in addition to the activity against parasitic weeds from the *Striga* genus. A recent study has demonstrated this potential by optimising one such chemical reaction. The dihydrochalcone phloretin was 3′-*C*-β-D-glucosylated to nothofagin using OsCGT, *in situ* production of UDP-glucose and with β-cyclodextrin complexation to increase substrate availability (Bungaruang et al., 2015). Alternative methods to synthesise CGFs synthetically have relied on two approaches. One strategy is heavily reliant on protection group chemistry, with multiply protected activated sugar donors reacting with protected electron rich aromatic moieties. It requires many chemical steps and often suffers low yields due to the requirement for *O*-glycosylation followed by a 1,2-Fries shift to form the *C*-glycosidic link (Kumazawa et al., 1995, 2001; Mahliing et al., 1995; Sato et al., 2006a; Furuta et al., 2009). The other strategy shows greater potential and relies on Lewis acid catalysed glycosylations ($Sc(OTf)_3$, $Yb(OTf)_3$ or $Pr(OTf)_3$) of an electron rich aromatic moiety with the unprotected glycoside, but which must be subsequently converted to the desired flavonoid using protection chemistry (Sato et al., 2004, 2006b; Sato and Koide, 2010; Santos et al., 2013). Within this framework, the rice OsCGT or other recombinant *C*-glycosyltransferase enzymes also have the potential to contribute to synthetic strategies.

4. Experimental

General experimental procedures. Electrospray ionisation mass spectra (ESIMS) were recorded in negative ionisation mode using the Micromass Quattro Ultima. LCMS analysis was performed on the Micromass Quattro Ultima bench top triple quadrupole mass spectrometer attached to Waters Acquity UPLC system (Ultra Performance Liquid Chromatography).

4.1. Synthesis of 2-hydroxyflavanones **3a–x**

4.1.1. Synthesis of esters **1a–1x**

2,4-Dibenzyloxy-6-hydroxyacetophenone (1.0 g, 2.87 mmol) was dissolved in 30 mL dry CH_2Cl_2 along with the selected aromatic acid (1.1 eq., 3.16 mmol), 4-dimethylaminopyridine (0.5 eq., 176 mg) and the coupling agent *N*-(3-dimethylaminopropyl)-*N*′-ethylcarbodiimide hydrochloride (1.25 eq., 689 mg). The mixture was refluxed overnight, concentrated *in vacuo* and purified by column chromatography using 20–50% ethyl acetate in petroleum ether to yield a viscous oil or waxy solid, **1a** (87%), **1b** (53%), **1c** (49%), **1d** (92%), **1e** (76%), **1f** (67%), **1g** (48%), **1h** (69%), **1i** (63%), **1j** (81%), **1k** (67%), **1l** (51%), **1n** (71%), **1o** (80%), **1p** (98%), **1q** (40%), **1r** (68%), **1s** (42%), **1t** (89%), **1u** (99%), **1v** (65%), **1w** (89%), **1x** (76%).

4.1.2. Synthesis or propanediones **2a–2x**

The product of the above reaction was dissolved in pyridine (15 mL) with 1.0 g of dry, powdered KOH, prepared by powdering it under ether and drying overnight at 60 °C. The mixture was heated under nitrogen at 85 °C for 1 h, cooled in ice, quenched with acetic acid (6 mL), and the product extracted from water with dichloromethane. The dichloromethane was washed with HCl 1 M, brine, dried ($MgSO_4$), filtered and the solvent evaporated *in vacuo*. The crude residue was crystallised from hot dichloromethane by the addition of petroleum ether to yield yellow or orange crystals of the propanedione. **2a** (76%), **2b** (50%), **2c** (37%), **2d** (25%), **2e** (50%), **2f** (11%), **2g** (21%), **2h** (34%), **2i** (29%), **2j** (33%), **2k** (31%), **2l** (44%), **2m** (70%), **2n** (19%), **2o** (31%), **2p** (20%), **2q** (39%), **2r** (45%), **2s** (9%), **2t** (29%), **2u** (15%), **2v** (7%), **2w** (29%), **2x** (17%).

4.1.3. Synthesis of 2-hydroxyflavanones **3a–3t**

100–300 mg of the propanedione was dissolved in a mixture containing equal volumes of MeOH and THF (10–30 mL). 20% by weight of 10% Pd/C was added to the solution which was stirred under H_2 and monitored by TLC. When the reaction was complete, the solution was filtered, the catalyst washed with acetone and the organic solvent concentrated *in vacuo* before column chromatography in EtOAc/petroleum ether. The purified product was weighed and portioned into glass ampoules which were sealed and kept at −20 °C before use. In the case of **3u–3x**, the catalyst used was a very small spatula of $Pt(OH)_2$ and when the reaction was complete by TLC, the solution was placed under nitrogen, the solvent concentrated *in vacuo*, the residue re-suspended in the volume of EtOH required to give a theoretical 20 mM solution of substrate, and the suspension was centrifuged. Decanting of the supernatant under nitrogen provided the substrate solution that was immediately used in enzyme assays.

3a 87%; ^{1}H NMR (500 MHz, $(CD_3)_2CO$, 213 K) 12.20 (1H, s, 5-OH), 10.30 (1H, s, 7-OH), 9.27 (1H, s, 4′-0H), 7.56 (2H, d, *J* = 8.5 Hz, 2′,6′-H), 7.00 (1H, s, 2-OH), 6.91 (2H, d, *J* = 8.4 Hz, 3′,5′-H), 5.98 (2H, m, 6,8-H), 3.29 (1H, dd, *J* = 2.8, 16 9 Hz, H-3a), 2.80 (1H, d, *J* = 16.9 Hz, 3-Hb); ^{13}C NMR (125 MHz, $(CD_3)_2CO$, 213 K) 196.3, 166.7, 164.2, 161.1, 158.1, 134.1, 127.6 (2C), 115.1 (2C), 102.7, 102.3, 96.4, 96.1, 49.2.

3b 60%; ^{1}H NMR (500 MHz, $(CD_3)_2CO$, 221 K) 12.18 (1H, s, 5-OH), 10.41 (1H, s, 7-OH), 8.87 (1H, s, 4′-OH), 8.83 (1H, s, 3′OH), 7.17 (1H, s, 2-OH), 7.03 (1H, d, *J* = 8.1 Hz, 2′-H), 6.87–6.85 (2H, m, 5′,6′-H), 5.98 (1H, m, 6-H), 5.97 (1H, s, 8-H), 3.11 (1H, dd, **J = 11.0** Hz, H-3a), 2.78 (1H, d, *J* = 17.3 Hz, 3-Hb); ^{13}C NMR (125 MHz, $(CD_3)_2CO$, 221 K) 195.9, 166.4, 163.8, 160.7, 145.5, 144.7, 134.3, 116.7, 114.6, 112.8, 102.2, 101.9, 96.0, 95.7, 48.9.

Please cite this article in press as: Hao, B., et al. Biosynthesis of natural and novel *C*-glycosylflavones utilising recombinant *Oryza sativa C*-glycosyltransferase (OsCGT) and *Desmodium incanum* root proteins. Phytochemistry (2016), http://dx.doi.org/10.1016/j.phytochem.2016.02.013

3c 19%; ^{1}H NMR (500 MHz, $(CD_3)_2CO$, 221 K) 12.18 (1H, s, 5-OH), 10.24 (1H, s, 7-OH), 8.68 (2H, s, 3′,5′-0H), 8.20 (1H, s, 4′-OH), 6.77 (1H, s, 2-OH), 6.73 (2H, s, 2′,6′-H), 5.98 (1H, m, 6-H), 5.97 (1H, s, 8-H), 3.22 (1H, dd, J = 2.7, 17.0 Hz, H-3a), 2.75 (1H, d, J = 16.9 Hz, 3-Hb); ^{13}C NMR (125 MHz, $(CD_3)_2CO$, 221 K) 195.9, 166.3, 163.8, 160.7, 145.2 (2C), 133.6, 133.0, 104.5 (2C), 102.2, 101.9, 96.0, 95.7, 48.9.

3d 49%; ^{1}H NMR (500 MHz, $(CD_3)_2CO$, 220 K) 12.20 (1H, s, 5-OH), 10.28 (1H, s, 7-OH), 8.54 (2H, s, 4′-0H), 7.31 (1H, d, J = 2.0 Hz, 2′-H), 7.18 (1H, dd, J = 2.0, 8.0 Hz, 6′H), 6.94 (1H, s, 2-OH), 6.89 (1H, d, J = 8.0 Hz, 5′-H), 2-OH), 5.99–5.98 (2H, m, 6,8-H), 3.86 (3H, s, OMe), 3.31 (1H, dd, J = 2.8, 16.9 Hz, H-3a), 2.82 (1H, d, J = 16.9 Hz, 3-Hb); ^{13}C NMR (125 MHz, $(CD_3)_2CO$, 220 K) 195.9, 166.4, 163.8, 160.7, 146.9, 146.7, 134.0, 117.9, 114.3, 108.7, 102.3, 101.9, 96.0, 95.7, 55.0, 48.9.

3e 33%; ^{1}H NMR (500 MHz, $(CD_3)_2CO$, 220 K) 12.20 (1H, s, 5-OH), 10.41 (1H, s, 7-OH), 8.17 (1H, s, 4′-0H), 7.01 (2H, s, 2′,6′-H), 6.96 (1H, s, 2-OH), 6.00–5.98 (2H, m, 6,8-H), 3.84 (6H, s, 3′,5′-OMe), 3.33 (1H, d, J = 17.0 Hz, H-3a), 2.81 (1H, d, J = 17.0 Hz, 3-Hb); ^{13}C NMR (125 MHz, $(CD_3)_2CO$, 220 K) 195.8, 166.5, 163.8, 160.7, 147.1 (2C), 135.3, 132.9, 102.4, 102.1 (2C), 101.8, 96.1, 95.8, 55.3 (2C), 48.8.

3f 52%; ^{1}H NMR (500 MHz, $(CD_3)_2CO$, 222 K) 12.18 (1H, s, 5-OH), 10.21 (1H, s, 7-OH), 7.66 (2H, d, J = 8.7 Hz, 2′,6′-H), 7.02 (1H, d, J = 8.8 Hz, 3′,5′-H), 6.94 (1H, s, 2-OH), 6.00 (1H, d, J = 1.8 Hz, 6-H), 5.98 (1H, d, J = 1.8 Hz, 8-H), 3.83 (3H, s, 4′-OMe), 3.29 (1H, d, J = 16.9 Hz, H-3a), 2.83 (1H, d, J = 17.0 Hz, 3-Hb); ^{13}C NMR (125 MHz, $(CD_3)_2CO$, 220 K) 195.8, 166.4, 163.8, 160.6, 159.8, 134.7, 127.0 (2C), 113.3 (2C), 102.2, 101.9, 96.0, 95.7, 54.8, 48.7.

3g 65%; ^{1}H NMR (500 MHz, $(CD_3)_2CO$, 292 K) 12.08 (1H, s, 5-OH), 9.70 (1H, s, 7-OH), 7.75 (2H, d, J = 7.0 Hz 2′,6′-H), 7.47 (2H, t, J = 7.2 Hz, 3′,5′-H), 7.42 (1H, t, J = 7.0 Hz, 4′-H), 6.56 (1H, s, 2-OH), 6.05 (1H, br s, 6′-H), 6.01 (1H, d, J = 1.6 Hz, 8-H), 3.24 (1H, d, J = 16.8 Hz, H-3a), 2.92 (1H, d, J = 16.9 Hz, 3-Hb); ^{13}C NMR (125 MHz, $(CD_3)_2CO$, 293 K) 195.1, 166.4, 164.0, 160.6, 142.7, 128.7, 128.5 (2C), 125.4 (2C), 102.1, 101.9, 96.1, 96.0, 48.6.

3h 76%; ^{1}H NMR (500 MHz, $(CD_3)_2CO$, 220 K) 12.16 (1H, s, 5-OH), 10.41 (1H, s, 7-OH), 7.24 (1H, s, 2′-H), 7.23 (1H, d, J = 7.9 Hz, 5′-H), 6.96 (1H, dd, J = 1.5, 7.6 Hz, 6′-H), 6.11 (2H, s, OCH_2O), 6.00–5.98 (2H, m, 6,8-H), 5.76 (1H, d, J = 3.1 Hz, 2-OH), 3.27 (1H, dd, J = 3.3, 17.2 Hz, H-3a), 2.83 (1H, d, J = 16.8 Hz, 3-Hb); ^{13}C NMR (125 MHz, $(CD_3)_2CO$, 220 K) 195.7, 166.5, 163.8, 160.5, 147.9, 147.6, 136.8, 119.1, 107.7, 106.4, 102.1, 101.8, 101.7, 96.0, 95.8, 48.6.

3i 70%; ^{1}H NMR (500 MHz, $(CD_3)_2CO$, 221 K) 12.16 (1H, s, 5-OH), 10.30 (1H, s, 7-OH), 9.05 (2H, s, 3′,5′-0H), 6.88 (1H, s, 2-OH), 6.68 (2H, d, J = 2.5 Hz, 2′,6′-H), 6.37 (1H, t, J = 2.1 Hz, 4′-H), 6.00 (1H, d, J = 2.1 Hz, 6-H), 5.98 (1H, d, J = 2.1 Hz, 8-H), 3.24 (1H, dd, J = 2.6, 16.9 Hz, H-3a), 2.79 (1H, d, J = 16.9 Hz, 3-Hb); ^{13}C NMR (125 MHz, $(CD_3)_2CO$, 221 K) 195.6, 166.4, 163.8, 160.5, 158.5 (2C), 145.1, 104.0 (2C), 102.4, 102.0, 101.8, 96.0, 95.8, 48.5.

3j 35%; ^{1}H NMR (500 MHz, $(CD_3)_2CO$, 292 K) 12.06 (1H, s, 5-OH), 9.75 (1H, s, 7-OH), 8.62 (1H, s, 4′-0H), 7.28 (1H, t, J = 7.9 Hz, 5′-H), 7.21–7.17 (2H, m, 2′,6′-H), 6.88 (1H, dd, J = 2.4, 8.0 Hz, 4′-H), 6.51 (1H, d, J = 1.9 Hz, 2-OH), 6.03 (1H, d, J = 2.1 Hz, 6-H), 5.99 (1H, d, J = 2.2 Hz, 8-H), 3.21 (1H, dd, J = 2.1, 16 9 Hz, H-3a), 2.88 (1H, d, J = 16.9 Hz, 3-Hb); ^{13}C NMR (125 MHz, $(CD_3)_2CO$, 213 K) 195.2, 166.4, 163.9, 160.6, 157.4, 144.3, 129.4, 116.5, 115.5, 112.5, 102.1, 101.8, 96.0, 96.0, 48.6.

3k 82%; ^{1}H NMR (500 MHz, $(CD_3)_2CO$, 220 K) 12.18 (1H, s, 5-OH), 10.52 (1H, s, 7-OH), 7.63 (2H, d, J = 8.7 Hz, 2′,6′-H), 7.04 (2H, d, J = 8.8 Hz, 3′,5′-H), 6.99 (1H, s, 2-OH), 6.00 (1H, s, 6-H), 5.99 (1H, s, 8-H), 4.09 (2H, q, J = 7.0 Hz, CH_2O), 3.26 (1H, d, J = 16.9 Hz, H-3a), 2.85 (1H, d, J = 16.9 Hz, 3-Hb), 1.39 (3H, t, J = 7.0 Hz, CH_3); ^{13}C NMR (125 MHz, $(CD_3)_2CO$, 293 K) 195.4, 166.4, 163.9, 160.7, 159.3, 134.8, 126.8 (2C), 113..9 (2C), 102.1, 101.9, 96.0, 95.9, 63.2, 48.7, 14.2.

3l 75%; ^{1}H NMR (500 MHz, $(CD_3)_2CO$, 222 K) 12.18 (1H, s, 5-OH), 10.29 (1H, s, 7-OH), 7.64 (2H, d, J = 8.6 Hz, 2′,6′-H), 7.01 (2H, d, J = 8.6 Hz, 3′,5′-H), 7.00 (1H, s, 2-OH), 6.01 (1H, s, 6-H), 5.99 (1H, d, J = 1.7 Hz, 8-H), 4.08 (2H, t, J = 6.0 Hz, CH_2OAr), 3.74 (2H, q, J = 5.5 Hz, CH_2OH), 3.26 (1H, dd, J = 2.6, 16.9 Hz, H-3a), 2.83 (1H, d, J = 16.9 Hz, 3-Hb), 1.96 (2H, qu, J = 6.0 Hz, CH_2); ^{13}C NMR (125 MHz, $(CD_3)_2CO$, 220 K) 195.8, 166.4, 163.8, 160.6, 159.3, 134.6, 127.0 (2C), 113.7 (2C), 102.2, 101.9, 96.0, 95.8, 64.1, 57.7, 48.8, 32.1.

3m 44%; ^{1}H NMR (500 MHz, $(CD_3)_2CO$, 298 K) 12.07 (1H, s, 5-OH), 9.60 (1H, s, 7-OH), 7.63 (2H, d, J = 8.8 Hz, 2′,6′-H), 6.99 (2H, d, J = 8.6 Hz, 3′,5′-H), 6.44 (1H, s, 2-OH), 6.02–5.99 (2H, m, 6,8-H), 4.05 (2H, t, J = 6.4 Hz, CH_2OAr), 3.56 (2H, t, J = 6.5 Hz, CH_2OH), 3.21 (1H, d, J = 16.9 Hz, H-3a), 2.89 (1H, d, J = 16.9 Hz, 3-Hb), 1.80 (2H, m, CH_2), 1.58–1.55 (6H, m, $3CH_2$); ^{13}C NMR (125 MHz, $(CD_3)_2CO$, 298 K) 195.3, 166.4, 163.9, 160.7, 159.5, 134.7, 126.8 (2C), 114.0 (2C), 102.1, 102.0, 96.0, 95.9, 67.7, 61.5, 48.8, 32.8, 28.5, 25.8, 25.6.

3n 50%; ^{1}H NMR (500 MHz, $(CD_3)_2CO$, 294 K) 12.07 (1H, s, 5-OH), 9.60 (1H, s, 7-OH), 7.75 (2H, d, J = 8.7 Hz, 2′,6′-H), 7.43 (2H, t, J = 8.2 Hz, Ph), 7.19 (1H, t, J = 7.4 Hz, Ph), 7.07 (4H, d, J = 8.7 Hz, 3′,5′-H, Ph), 6.55 (1H, s, 2-OH), 6.04 (1H, d, J = 2.0 Hz, 6-H), 6.01 (1H, d, J = 2.0 Hz, 8-H), 3.26 (1H, d, J = 16.9 Hz, H-3a), 2.93 (1H, d, J = 16.8 Hz, 3-Hb); ^{13}C NMR (125 MHz, $(CD_3)_2CO$, 293 K) 195.1, 166.4, 164.0, 160.5, 157.8, 156.9, 137.6, 130.0 (2C), 127.3 (2C), 123.8, 119.1 (2C), 118.0 (2C), 102.1, 101.8, 96.0, 96.0, 48.7.

3o 90%; ^{1}H NMR (500 MHz, $(CD_3)_2CO$, 292 K) 12.07 (1H, s, 5-OH), 9.60 (1H, s, 7-OH), 7.61 (2H, d, J = 8.2 Hz, 2′,6′-H), 7.27 (2H, d, J = 8.0 Hz, 3′,5′-H), 6.48 (1H, s, 2-OH), 6.03 (1H, d, J = 1.9 Hz, 6-H), 5.99 (1H, d, J = 2.0 Hz, 8-H), 3.22 (1H, d, J = 16.9 Hz, H-3a), 2.90 (1H, d, J = 16.9 Hz, 3-Hb), 2.37 (3H, s, Me); ^{13}C NMR (125 MHz, $(CD_3)_2CO$, 292 K) 195.2, 166.3, 163.9, 160.7, 139.9, 138.3, 128.8 (2C), 125.4 (2C), 102.1, 102.0, 96.0, 95.9, 48.7, 20.2.

3p 73%; ^{1}H NMR (500 MHz, $(CD_3)_2CO$, 293 K) 12.08 (1H, s, 5-OH), 9.65 (1H, s, 7-OH), 7.64 (2H, d, J = 8.1 Hz, 2′,6′-H), 7.30 (2H, d, J = 8.1 Hz, 3′,5′-H), 6.48 (1H, s, 2-OH), 6.04 (1H, d, J = 1.9 Hz, 6-H), 6.01 (1H, d, J = 2.0 Hz, 8-H), 3.23 (1H, d, J = 16.9 Hz, H-3a), 2.91 (1H, d, J = 16.9 Hz, 3-Hb), 2.69 (2H, q, J = 7.6 Hz, $ArCH_2$), 1.24 (3H, t, J = 7.6 Hz, Me); ^{13}C NMR (125 MHz, $(CD_3)_2CO$, 293 K) 195.2, 166.4, 163.9, 160.7, 144.8, 140.1, 127.7 (2C), 125.5 (2C), 102.1, 102.0, 96.0, 96.0, 48.7, 28.2, 15.2.

3q 51%; ^{1}H NMR (500 MHz, $(CD_3)_2CO$, 220 K) 12.18 (1H, s, 5-OH), 10.31 (1H, s, 7-OH), 7.66 (2H, d, J = 8.3 Hz, 2′,6′-H), 7.53 (2H, d, J = 8.3 Hz, 3′,5′-H), 6.96 (1H, s, 2-OH), 6.00 (1H, s, 6-H), 5.99 (1H, d, J = 2.0 Hz, 8-H), 3.31 (1H, dd, J = 2.6, 16.9 Hz, H-3a), 2.85 (1H, d, J = 17.0 Hz, 3-Hb), 1.32 (9H, s, $(CH_3)_3$); ^{13}C NMR (125 MHz, $(CD_3)_2CO$, 221 K) 195.7, 166.4, 163.8, 160.6, 151.4, 139.8, 125.4 (2C), 125.3 (2C), 102.2, 101.9, 96.0, 95.8, 48.5, 34.4, 30.7 (3C).

3r 82%; ^{1}H NMR (500 MHz, $(CD_3)_2CO$, 221 K) 12.16 (1H, s, 5-OH), 10.38 (1H, s, 7-OH), 9.67 (1H, s, 4′OH), 7.49 (1H, dd, J = 2.0, 12.2 Hz, 2′-H), 7.39 (1H, dd, J = 1.6, 8.4 Hz, 6′-H), 7.08 (1H, t, J = 8.7 Hz, 5′-H), 7.07 (1H, d, J = 2.6 Hz, 2-OH), 6.01 (1H, d, J = 2.0 Hz, 6-H), 5.99 (1H, d, J = 2.1 Hz, 8-H), 3.28 (1H, dd, J = 16.9 Hz, H-3a), 2.85 (1H, d, J = 16.9 Hz, 3-Hb); ^{13}C NMR (125 MHz, $(CD_3)_2CO$, 236 K) 195.5, 166.4, 163.8, 160.4, 150.6 (d, J = 239 Hz), 139.8 (d, J = 5 Hz), 122.0 (d, J = 3 Hz), 117. 2, 113.8 (d, J = 20 Hz), 101.9, 101.5, 96.0, 95.9, 48.5.

3s 84%; ^{1}H NMR (500 MHz, $(CD_3)_2CO$, 221 K) 12.92 (1H, s, 5-OH), 11.18 (1H, s, 7-OH), 8.78 (2H, d, J = 8.1 Hz, 2′,6′-H), 7.69 (2H, d, J = 8.3 Hz, 3′,5′-H), 8.10 (1H, s, 2-OH), 6.83 (1H, d, J = 1.5 Hz, 6-H), 6.80 (1H, d, J = 1.7 Hz, 8-H), 4.13 (1H, d,

Please cite this article in press as: Hao, B., et al. Biosynthesis of natural and novel *C*-glycosylflavones utilising recombinant *Oryza sativa C*-glycosyltransferase (OsCGT) and *Desmodium incanum* root proteins. Phytochemistry (2016), http://dx.doi.org/10.1016/j.phytochem.2016.02.013

J = 16.9 Hz, H-3a), 2.85 (1H, d, J = 16.9 Hz, 3-Hb); ^{13}C NMR (125 MHz, $(CD_3)_2CO$, 294 K) 194.5, 166.5, 163.9, 160.2, 146.8, 130.2 (q, J = 32 Hz), 126.5 (2C), 125.3 (2C, q, J = 4 Hz), 124.3 (2C, q, J = 278 Hz), 102.0, 101.3, 96.1 (2C), 48.3.

3t 48%; ^{1}H NMR (500 MHz, $(CD_3)_2CO$, 293 K) 12.05 (1H, s, 5-OH), 9.73 (1H, s, 7-OH), 7.79 (2H, dd, J = 7.5, 8.8 Hz, 2′,6′-H), 7.22 (2H, t, J = 8.8 Hz, 3′,5′-H), 6.66 (1H, s, 2-OH), 6.04 (1H, d, J = 2.1 Hz, 6-H), 6.01 (1H, d, J = 2.1 Hz, 8-H), 3.24 (1H, d, J = 16.9 Hz, H-3a), 2.92 (1H, d, J = 17.0 Hz, 3-Hb); ^{13}C NMR (125 MHz, $(CD_3)_2CO$, 294 K) 194.9, 166.4, 163.9, 162.7 (d, J = 246 Hz), 160.4, 139.0 (d, J = 3 Hz), 127.8 (2C, d, J = 9 Hz), 114.9 (2C, d, J = 21 Hz), 102.0, 101.6, 96.1 (2C), 48.6.

*4.2. Generation of 6- and 8-C-glucosylflavones from **3** using recombinant OsCGT*

2-Hydroxyflavanone **3** (40 μL of a 20 mmol solution) in ethanol was added to 1 mg of UDP-glucose dissolved in 40 μL buffer with 50 μg recombinant OsCGT (Hamilton et al., 2012) dissolved in 200 μl 100 mM HEPES pH 7.5 buffer and incubated at 30 °C overnight. The reaction was terminated by addition of 1 M HCl (250 μL) and the products isolated on a semi-preparative HPLC column, ACEQ C-18 column (250 mm × 10 mm, 5 μm), using the analytical method at 4 mL/min (see 4.7), to yield 6- and 8-C-glucosylflavones that were characterised by NMR spectroscopy.

*4.2.1. 8-C-Glucosylapigenin (vitexin) **5a***

UV (λ_{max} MeOH/H_2O) 271, 349 nm. ^{1}H NMR (500 MHz, d_6-DMSO, 300 K) δ 7.96 (2H, J = 8.4 Hz, H2′, H6′), 6.91 (2H, d, J = 8.5 Hz, H3′, H5′), 6.63 (1H, s, H3), 6.15 (1H, s, H6), 8-C-β-Glu 4.70 (1H, d, J = 9.8 Hz, H1″), 3.89 (1H, t, J = 9.0 Hz, H2″), 3.76 (1H, d, J = 11.4, H6a″), 3.55 (1H, dd, J = 5.7, 11.7 Hz, H6b″), 3.39 (1H, t, 9.3 Hz, H3″), 3.31 (1H, t, J = 8.5 Hz, H4″), 3.28 (1H, m, H5″). ^{13}C NMR (125 MHz, d_4-MeOD, 350 K) 181.6 (C4), 164.2 (C2), 164.0 (C7), 162.2 (C4′), 161.1(C5), 156.9 (C9), 129.3 (C2′, C6′), 122.5 (C1′), 116.7 (C3′, C5′), 105.7 (C8), 103.6 (C10), 103.0 (C3), 100.1 (C6). 6-C-β-Glu 74.8 (C1″), 72.2 (C2″), 79.8 (C3″), 71.6 (C4″), 82.6 (C5″), 62.4 (C6″). ESIMS m/z (CV = 150, rel. int.): 431 $[M-H]^-$ (57), 341 $[M-H-90]^-$ (11), 311 $[M-H-120]^-$ (100).

*4.2.2. 6-C-Glucosylapigenin (isovitexin) **6a***

UV (λ_{max} MeOH/H_2O) 271, 349 nm. ^{1}H NMR (500 MHz, d_4-MeOD, 330 K) δ 7.79 (2H, J = 8.7 Hz, H2′, H6′), 6.93 (2H, d, J = 8.5 Hz, H3′, H5′), 6.57 (1H, s, H3), 6.49 (1H, s, H8), 8-C-β-Glu 4.94 (1H, d, J = 9.9 Hz, H1″), 4.13 (1H, t, J = 9.2 Hz, H2″), 3.91 (1H, d, J = 11.8, H6a″), 3.79 (1H, dd, J = 5.1, 12.1 Hz, H6b″), 3.55–3.47 (2H, m, H3″, H4″), 3.33 (1H, m, H5″). ^{13}C NMR (125 MHz, d_4-MeOD, 330 K) 182.5 (C4), 164.8 (C2), 163.1 (C7), 161.1 (C4′), 160.5(C5), 157.2 (C9), 127.8 (C2′, C6′), 121.9 (C1′), 115.6 (C3′, C5′), 107.8 (C8), 103.9 (C10), 102.6 (C3), 94.1 (C8). 81.0 (C5″), 78.7 (C3″), 74.0 (C1″), 71.4 (C2″), 70.4 (C4″), 61.5 (C6″). ESIMS m/z (CV = 150, rel. int.): 431 $[M-H]^-$ (57), 341 $[M-H-90]^-$ (11), 311 $[M-H-120]^-$ (100).

*4.2.3. 8-C-Glucosylluteolin (orientin) **5b***

UV (λ_{max} MeOH/H_2O) 298, 349 nm. ^{1}H NMR (500 MHz, d_4-MeOD, 350 K) δ 7.51 (1H, br s, H2′), 7.46 (1H, dd, J = 2.0, 8.4 Hz, H6′), 6.95 (1H, d, J = 8.4 Hz, H5′), 6.54 (1H, s, H3), 6.31 (1H, s, H6), 8-C-β-Glu 5.06 (1H, d, J = 10.1 Hz, H1″), 4.08 (1H, t, J = 9.1 Hz, H2″), 3.95 (1H, dd, J = 2.4, 12.1 Hz, H6a″), 3.85 (1H, dd, J = 5.2, 11.9 Hz, H6b″), 3.67 (1H, t, J = 9.0 Hz, H4″), 3.58 (1H, t, J = 8.9 Hz, H3″), 3.54 (1H, m, H5″). ^{13}C NMR (125 MHz, d_4-MeOD, 350 K) 182.5 (C4), 165.1 (C2), 163.1 (C7), 161.5 (C5), 156.4 (C9), 148.8 (C4′), 145.6 (C3′), 122.9 (C1′), 119.5 (C6′), 115.5 (C5′), 113.6 (C2′), 104.4 (C10), 103.2 (C8), 102.8 (C3), 99.0 (C6), 81.3 (C5″), 79.0 (C3″), 74.7 (C1″), 71.9 (C2″), 70.6 (C4″), 61.7 (C6″). ESIMS m/z (CV = 50, rel. int.): 493 $[M-H+46$ (formic acid)$]^-$ (100), 447 $[M-H]^-$ (17); ESIMS m/z (CV = 150, rel. int.): 493 $[M-H+46]^-$ (14), 447 $[M-H]^-$ (3), 403 $[M-H+46-90]^-$ (9), 373 $[M-H+46-120]^-$ (10), 357 $[M-H-90]^-$ (12), 327 $[M-H-120]^-$ (98).

*4.2.4. 6-C-Glucosylluteolin (isoorientin) **6b***

UV (λ_{max} MeOH/H_2O) 298, 352 nm. ^{1}H NMR (500 MHz, d_4-MeOD, 350 K) δ 7.40 (1H, d, J = 2.1 Hz, H2′), 7.38 (1H, dd, J = 2.2, 8.3 Hz, H6′), 6.94 (1H, d, J = 8.3 Hz, H5′), 6.55 (1H, s, H3), 6.52 (1H, s, H8), 6-C-β-Glu 4.96 (1H, d, J = 9.9 Hz, H1″), 4.08 (1H, dd, J = 8.8, 9.6 Hz, H2″), 3.89 (1H, dd, J = 2.7, 12.0, H6a″), 3.79 (1H, dd, J = 5.0, 12.0 Hz, H6b″), 3.55 (1H, t, J = 8.4 Hz, H4″), 3.52 (1H, t, J = 8.7 Hz, H3″), 3.46 (1H, ddd, J = 2.6, 4.9, 9.3 Hz, H5″). ^{13}C NMR (125 MHz, d_4-MeOD, 350 K) 182.4 (C4), 164.8 (C2), 163.0 (C7), 160.5 (C5), 157.4 (C9), 149.0 (C4′), 145.6 (C3′), 122.6 (C1′), 119.0 (C6′), 115.5 (C5′), 113.1 (C2′), 107.8 (C6), 104.0 (C10), 102.9 (C3), 94.1 (C8), 80.9 (C5″), 78.7 (C3″), 74.1 (C1″), 71.7 (C2″), 70.4 (C4″), 61.5 (C6″). ESIMS m/z (CV = 50, rel. int.): 447 $[M-H]^-$ (100); ESIMS m/z (CV = 150, rel. int.): 447 $[M-H]^-$ (16), 357 $[M-H-90]^-$ (33), 327 $[M-H-120]^-$ (72).

*4.2.5. 3′-O-Methyl-8-C-glucosylluteolin **5d***

UV (λ_{max} MeOH/H_2O) 269, 348 nm. ^{1}H NMR (500 MHz, d_4-MeOD, 350 K) δ 7.49 (1H, br d, J = 8.3 Hz, H6′), 7.56 (1H, br s, H2′), 6.98 (1H, br d, J = 7.3 Hz, H5′), 6.61 (1H, s, H3), 6.30 (1H, s, H6), 4.01 (3H, s, OMe), 8-C-β-Glu 5.09 (1H, d, J = 9.1 Hz, H1″), 4.09 (1H, br s, H2″), 3.93 (1H, br d, J = 11.9 Hz, H6a″), 3.80 (1H, br m, H6b″), 3.62 (1H, br t, J = 11.0 Hz, H4″), 3.55 (1H, br t, J = 7.2 Hz, H3″), 3.51 (1H, m, H5″). ^{13}C NMR (125 MHz, d_4-MeOD, 350 K, partial characterisation) 120.9 (C6′), 115.5 (C5′), 110.5 (C2′), 103.0 (C3), 99.3 (C6), 81.5 (C5″), 79.0 (C3″), 74.8 (C1″), 71.8 (C2″), 70.6 (C4″), 61.5 (C6″), 56.0 (OMe).). ESIMS m/z (CV = 50, rel. int.): 507 $[M-H+46$ (formic acid)$]^-$ (20), 461 $[M-H]^-$ (100); ESIMS m/z (CV = 150, rel. int.): 461 $[M-H]^-$ (5), 371 $[M-H-90]^-$ (2), 341 $[M-H-120]^-$ (20).

*4.2.6. 3′-O-Methyl-6-C-glucosylluteolin **6d***

UV (λ_{max} MeOH/H_2O) 272, 352 nm. ^{1}H NMR (500 MHz, d_4-MeOD, 330 K) δ 7.49 (1H, br d, J = 8.6 Hz, H6′), 7.46 (1H, br s, H2′), 6.95 (1H, d, J = 8.2 Hz, H5′), 6.60 (1H, s, H3), 6.51 (1H, s, H8), 3.97 (3H, c, OMe), 6-C-β-Glu 4.95 (1H, d, J = 9.8 Hz, H1″), 4.12 (1H, t, J = 8.8 Hz, H2″), 3.90 (1H, br d, J = 11.7, H6a″), 3.79 (1H, dd, J = 4.7, 11.9 Hz, H6b″), 3.55–3.50 (2H, m, H3″, H4″), 3.46 (1H, br m, H5″). ^{13}C NMR (125 MHz, d_4-MeOD, 330 K) 182.2 (C4), 164.7 (C2), 162.7 (C7), 160.4 (C5), 156.9 (C9), 150.2 (C4′), 148.1 (C3′), 122.1 (C1′), 121.5 (C6′), 116.6 (C5′), 110.8 (C2′), 107.7 (C6), 104.1 (C3), 103.6 (C10), 95.2 (C8), 82.2 (C5″), 79.9 (C3″), 75.2 (C1″), 72.7 (C2″), 71.6 (C4″), 62.6 (C6″). ESIMS m/z (CV = 50, rel. int.): 461 $[M-H]^-$ (100); ESIMS m/z (CV = 150, rel. int.): 461 $[M-H]^-$ (8), 371 $[M-H-90]^-$ (9), 341 $[M-H-120]^-$ (20).

*4.2.7. 4′-O-Methylvitexin **5f***

UV (λ_{max} MeOH/H_2O) 273, 333 nm. ^{1}H NMR (500 MHz, d_4-MeOD, 345 K) δ 7.97 (2H, d, J = 8.3 Hz, H2′, H6′), 7.06 (2H, d, J = 8.7 Hz, H3′, H5′), 6.58 (1H, s, H3), 6.27 (1H, s, H6), 3.87 (3H, s, Me), 8-C-β-Glu 5.02 (1H, d, J = 9.7 Hz, H1″), 4.05 (1H, br t, J = 8.8 Hz, H2″), 3.92 (1H, br d, J = 11.6 Hz, H6a″), 3.80 (1H, dd, J = 5.1, 12.0 Hz, H6b″), 3.63 (1H, br t, J = 8.8 Hz, H4″), 3.55 (1H, t, J = 8.8 Hz, H3″), 3.50 (1H, m, H5″). ^{13}C NMR (125 MHz, d_4-MeOD, 345 K) 181.3 (C4), 163.8 (C2), 162.2 (C7), 161.9 (C4′), 160.3 (C5), 155.4 (C9), 127.4 (C2′, C6′), 122.6 (C1′), 113.4 (C3′, C5′), 104.0 (C10), 102.7 (C8), 102.2 (C3), 98.1 (C6), 80.6 (C5″), 78.1 (C3″), 73.7 (C1″), 70.9 (C2″), 69.9 (C4″), 60.7 (C6″), 53.8 (OMe). ESIMS m/z (CV = 50, rel. int.): 445 $[M-H]^-$ (100); ESIMS m/z (CV = 150, rel. int.): 445 $[M-H]^-$ (7), 355 $[M-H-90]^-$ (3), 325 $[M-H-120]^-$ (55).

Please cite this article in press as: Hao, B., et al. Biosynthesis of natural and novel *C*-glycosylflavones utilising recombinant *Oryza sativa C*-glycosyltransferase (OsCGT) and *Desmodium incanum* root proteins. Phytochemistry (2016), http://dx.doi.org/10.1016/j.phytochem.2016.02.013

4.2.8. 4′-O-Methyiisovitexin ***6f***

UV (λ_{max} MeOH/H_2O) 271, 334 nm. ^{1}H NMR (500 MHz, d_4-MeOD, 345 K) δ 7.88 (2H, d, J = 8.6 Hz, H2′, H6′), 7.06 (2H, d, J = 8.6 Hz, H3′, H5′), 6.58 (1H, s, H3), 6.50 (1H, s, H8), 3.87 (3H, s, Me), 6-*C*-β-Glu 4.93 (1H, d, J = 10.0 Hz, H1″), 4.07 (1H, t, J = 8.9 Hz, H2″), 3.87 (1H, obsc., H6a″), 3.76 (1H, dd, J = 4.7, 12.0 Hz, H6b″), 3.52–3.49 (2H, m, H3″, H4″), 3.44 (1H, m, H5″). ^{13}C NMR (125 MHz, d_4-MeOD, 345 K) 181.3 (C4), 163.3 (C2), 162.6 (C4′), 162.3 (C7), 160.2 (C5), 156.8 (C9), 127.8 (C2′, C6′), 121.4 (C1′), 114.3 (C3′, C5′), 107.9 (C6), 103.2 (C3), 102.5 (C10), 94.1 (C8), 81.0 (C5″), 78.7 (C3″), 74.1 (C1″), 71.6 (C2″), 70.4 (C4″), 61.4 (C6″), 54.6 (OMe). ESIMS m/z (CV = 50, rel. int.): 445 $[M-H]^-$ (100); ESIMS m/z (CV = 150, rel. int.): 445 $[M-H]^-$ (7), 355 $[M-H-90]^-$ (21), 325 $[M-H-120]^-$ (52).

4.2.9. 8-C-Glucosylchrysin ***5g***

UV (λ_{max} MeOH/H_2O) 269, 315 nm. ^{1}H NMR (500 MHz, d_4-MeOD, 345 K) δ 8.03 (2H, br d, J = 6.4 Hz, H2′, H6′), 7.55–7.53 (3H, m, H3′, H4′, H5′), 6.70 (1H, s, H3), 6.30 (1H, s, H6), 8-*C*-β-Glu 5.04 (1H, d, J = 8.4 Hz, H1″), 4.06 (1H, t, J = 9.3 Hz, H2″), 3.91 (1H, d, J = 12.4 Hz, H6a″), 3.79 (1H, dd, J = 4.8, 12.5 Hz, H6b″), 3.61 (1H, t, J = 9.5 Hz, H4″), 3.54 (1H, t, J = 8.9 Hz, H3″), 3.50 (1H, m, H5″). ^{13}C NMR (125 MHz, d_4-MeOD, 345 K) 182.0 (C4), 164.3 (C2), 163.0 (C7), 161.3 (C5), 156.4 (C9), 131.4 (C4′), 131.1 (C1′), 128.5 (C3′, C5′), 126.4 (C2′, C6′), 104.7 (C3), 104.3 (C10), 103.1 (C8), 99.1 (C6), 81.3 (C5″), 78.9 (C3″), 74.6 (C1″), 71.8 (C2″), 70.7 (C4″), 61.6 (C6″). ESIMS m/z (CV 50, rel. int.): 415 $[M-H]^-$ (100); ESIMS m/z (CV = 150, rel. int.): 415 $[M-H]^-$ (2), 325 $[M-H-90]^-$ (2), 295 $[M-H-120]^-$ (25).

4.2.10. 6-C-Glucosylchrysin ***6g***

UV (λ_{max} MeOH/H_2O) 271, 317 nm. ^{1}H NMR (500 MHz, d_4-MeOD, 330 K) δ 7.93 (2H, br d, J = 9.0 Hz, H2′, H6′), 7.56–7.53 (3H, m, H3′, H4′, H5′), 6.70 (1H, s, H3), 6.52 (1H, s, H8), 8-*C*-β-Glu 4.92 (1H, d, J = 9.8 Hz, H1″), 4.11 (1H, t, J = 8.8 Hz, H2″), 3.87 (1H, d, J = 11.9 Hz, H6a″), 3.74 (1H, dd, J = 4.4, 11.9 Hz, H6b″), 3.52–3.48 (2H, m, H3″, H4″), 3.44 (1H, m, H5″). ^{13}C NMR (125 MHz, d_4-MeOD, 330 K) 182.5 (C4), 164.4 (C2), 163.6 (C7), 160.1 (C5), 156.6 (C9), 131.5 (C4′), 130.1 (C1′), 128.8 (C3′, C5′), 125.9 (C2′, C6′), 108.0 (C8), 104.8 (C3), 103.6 (C10), 94.2 (C6), 81.1 (C5″), 78.7 (C3″), 74.0 (C1″), 71.4 (C2″), 70.4 (C4″), 61.5 (C6″). ESIMS m/z (CV = 50, rel. int.): 415 $[M-H]^-$ (100); ESIMS m/z (CV = 150, rel. int.): 415 $[M-H]^-$ (1), 325 $[M-H-90]^-$ (5), 295 $[M-H-120]^-$ (20).

4.2.11. 3′-4′-Methylenedioxy-8-C-glucosylchrysin ***5h***

UV (λ_{max} MeOH/H_2O) 271, 348 nm. ^{1}H NMR (500 MHz, d_4-MeOD, 330 K) δ 7.66 (1H, d, J = 8.0 Hz, H6′), 7.55 (1H, s, H2′), 6.99 (1H, d, J = 8.2 Hz, H5′), 6.59 (1H, s, H3), 6.28 (1H, s, H6), 6.08 (2H, s, OCH_2O), 8-*C*-β-Glu 5.04 (1H, d, J = 9.5 Hz, H1″), 4.08 (1H, t, J = 9.3 Hz, H2″), 3.96 (1H, br d, J = 11.1 Hz, H6a″), 3.82 (1H, dd, J = 5.5, 11.9 Hz, H6b″), 3.64 (1H, t, J = 9.1 Hz, H4″), 3.56 (1H, t, J = 8.8 Hz, H3″), 3.54 (1H, m, H5″). ^{13}C NMR (125 MHz, d_4-MeOD, 330 K) 181.6 (C4), 164.2 (C7), 163.4 (C2), 160.1 (C5), 156.2 (C9), 150.6 (C4′), 148.4 (C3′), 124.9 (C1′), 121.9 (C6′), 108.2 (C5′), 106.3 (C2′), 103.8 (C10), 103.2 (C8), 103.3 (C3), 102.0 (OCH_2O), 99.2 (C6), 81.4 (C5″), 79.0 (C3″), 74.4 (C1″), 71.7 (C2″), 70.7 (C4″), 61.7 (C6″). ESIMS m/z (CV = 50, rel. int.): 459 $[M-H]^-$ (100); ESIMS m/z (CV = 150, rel. int.): 459 $[M-H]^-$ (18), 369 $[M-H-90]^-$ (7), 339 $[M-H-120]^-$ (100).

4.2.12. 3′-4′-Methylenedioxy-6-C-glucosylchrysin ***6h***

UV (λ_{max} MeOH/H_2O) 270, 348 nm. ^{1}H NMR (500 MHz, d_4-MeOD, 330 K) δ 7.52 (1H, d, J = 8.1 Hz, H6′), 7.38 (1H, s, H2′), 6.95 (1H, d, J = 8.1 Hz, H5′), 6.58 (1H, s, H3), 6.50 (1H, s, H8), 6.05 (2H, s, OCH_2O), 6-*C*-β-Glu 4.91 (1H, d, J = 9.8 Hz, H1″), 4.10 (1H, t, J = 8.8 Hz, H2″), 3.86 (1H, br d, J = 11.8 Hz, H6a″), 3.75 (1H, dd, J = 4.8, 11.8 Hz, H6b″), 3.51–3.48 (2H, m, H3″, H4″), 3.43 (1H, m, H5″). ^{13}C NMR (125 MHz, d_4-MeOD, 330 K) 182.3 (C4), 164.1 (C2), 163.3 (C7), 160.4 (C5), 156.9 (C9), 150.7 (C4′), 148.8 (C3′), 124.9 (C1′), 121.4 (C6′), 108.1 (C5′), 108.0 (C6), 105.7 (C2′), 103.8 (C10), 103.5 (C3), 93.9 (C8), 81.0 (C5″), 78.6 (C3″), 73.9 (C1″), 71.3 (C2″), 70.3 (C4″), 61.3 (C6″). ESIMS m/z (CV = 50, rel. int.): 459 $[M-H]^-$ (100); ESIMS m/z (CV = 150, rel. int.): 459 $[M-H]^-$ (12), 369 $[M-H-90]^-$ (35), 339 $[M-H-120]^-$ (100).

4.2.13. 4′-Methyl-8-C-glucosylchrysin ***5o***

UV (λ_{max} MeOH/H_2O) 270, 317 nm. ^{1}H NMR (500 MHz, d_4-MeOD, 325 K) δ 7.97 (2H, br s, H2′,H6′), 7.39 (2H, br d, J = 8.0 Hz, H3′, H5′), 6.69 (1H, s, H3), 6.31 (1H, s, H6), 8-*C*-β-Glu 5.05 (1H, d, J = 9.1 Hz, H1″), 4.11 (1H, t, J = 8.7 Hz, H2″), 3.96 (1H, br d, J = 11.4, H6a″), 3.81 (1H, dd, J = 5.5, 12.0 Hz, H6b″), 3.65 (1H, m, H4″), 3.56 (1H, t, J = 8.8 Hz, H3″), 3.52 (1H, m, H5″), 2.44 (3H, s, Me). ^{13}C NMR (125 MHz, d_4-MeOD, 325 K) 182.6 (C4), 164.6 (C2), 162.5 (C7), 161.5 (C5), 156.7 (C9), 142.6 (C4′), 129.4 (C3′, C5′), 128.2 (C1′), 126.4 (C2′, C6′), 104.3 (C10), 103.8 (C3), 102.7 (C8), 98.7 (C6), 81.4 (C5″), 78.9 (C3″), 74.1 (C1″), 71.6 (C2″), 70.7 (C4″), 61.7 (C6″). ESIMS m/z (CV = 50, rel. int.): 429 $[M-H]^-$ (100); ESIMS m/z (CV = 150, rel. int.): 429 $[M-H]^-$ (1), 339 $[M-H-90]^-$ (2), 309 $[M-H-120]^-$ (22).

4.2.14. 4′-Methyl-6-C-glucosylchrysin ***6o***

UV (λ_{max} MeOH/H_2O) 272, 321 nm. ^{1}H NMR (500 MHz, d_4-MeOD, 325 K) δ 7.87 (2H, d, J = 8.0 Hz, H2′, H6′), 7.38 (2H, d, J = 7.8 Hz, H3′, H5′), 6.70 (1H, s, H3), 6.55 (1H, s, H8), 6-*C*-β-Glu 4.95 (1H, d, J = 9.7 Hz, H1″), 4.14 (1H, t, J = 8.7 Hz, H2″), 3.89 (1H, br d, J = 11.6, H6a″), 3.78 (1H, dd, J = 4.7, 11.8 Hz, H6b″), 3.55–3.49 (2H, m, H3″, H4″), 3.46 (1H, m, H5″), 2.45 (3H, s, Me). ^{13}C NMR (125 MHz, d_4-MeOD, 325 K) 182.2 (C4), 164.4 (C2), 163.4 (C7), 160.4 (C5), 157.3 (C9), 142.8 (C4′), 129.4 (C3′, C5′), 127.9 (C1′), 125.9 (C2′, C6′), 107.8 (C6), 103.9 (C10), 104.1 (C3), 94.0 (C8), 81.1 (C5″), 78.7 (C3″), 74.0 (C1″), 71.4 (C2″), 70.3 (C4″), 61.5 (C6″). ESIMS m/z (CV = 50, rel. int.): 429 $[M-H]^-$ (100); ESIMS m/z (CV = 150, rel. int.): 429 $[M-H]^-$ (1), 339 $[M-H-90]^-$ (6), 309 $[M-H-120]^-$ (23).

4.2.15. 4′-EthyI-8-C-glucosylchrysin ***5p***

UV (λ_{max} MeOH/H_2O) 272, 317 nm. ^{1}H NMR (500 MHz, d_4-MeOD, 350 K) δ 7.94 (2H, d, J = 7.8 Hz, H2′, H6′), 7.37 (2H, d, J = 7.9 Hz, H3′, H5′), 6.65 (1H, s, H3), 6.29 (1H, s, H6), 2.72 (2H. q, J = 7.6 Hz, Et), 1.26 (3H, t, J = 7.5 Hz, Et), 6-*C*-β-Glu 5.04 (1H, d, J = 9.6 Hz, H1″), 4.07 (1H, br t, J = 8.6 Hz, H2″), 3.92 (1H, br d, J = 11.9 Hz, H6a″), 3.79 (1H, dd, J = 5.0, 12.1 Hz, H6b″), 3.63 (1H, t, J = 9.0 Hz, H4″), 3.54 (1H, t, J = 8.8 Hz, H3″), 3.50 (1H, m, H5″). ^{13}C NMR (125 MHz, d_4-MeOD, 350 K) 182.3 (C4), 164.7 (C2), 163.3 (C7), 161.3 (C5), 156.4 (C9), 148 7 (C4′), 128.2 (C3′, C5′), 126.5 (C2′, C6′), 122.6 (C1′), 104.4 (C10), 104.0 (C3), 103.4 (C8), 99.0 (C6), 81.3 (C5″), 79.0 (C3″), 74.4 (C1″), 71.8 (C2″), 70.8 (C4″), 61.6 (C6″), 28.2 (CH_2), 13.9 (CH3). ESIMS m/z (CV = 50, rel. int.): 443 $[M-H]^-$ (100); ESIMS m/z (CV = 150, rel. int.): 443 $[M-H]^-$ (1), 353 $[M-H-90]^-$ (2), 323 $[M-H-120]^-$ (34).

4.2.16. 4′-EthyI-6-C-glucosylchrysin ***6p***

UV (λ_{max} MeOH/H_2O) 272, 317 nm. ^{1}H NMR (500 MHz, d_4-MeOD, 340 K) δ 7.83 (2H, d, J = 7.6 Hz, H2′, H6′), 7.36 (2H, d, J = 7.5 Hz, H3′, H5′), 6.64 (1H, s, H3), 6.50 (1H, s, H8), 2.72 (2H. q, J = 7.5 Hz, Et), 1.26 (3H, t, J = 7.5 Hz, Et), 6-*C*-β-Glu 4.93 (1H, d, J = 9.8 Hz, H1″), 4.08 (1H, t, J = 8.8 Hz, H2″), 3.87 (1H, br d, J = 11.8 Hz, H6a″), 3.76 (1H, dd, J = 4.0, 11.9 Hz, H6b″), 3.54–3.48 (2H, m, H3″, H4″), 3.44 (1H, m, H5″). ^{13}C NMR (125 MHz, d_4-MeOD, 340 K) 182.2 (C4), 164.3 (C2), 163.3 (C7), 160.2 (C5), 157.3 (C9), 148 8 (C4′), 128.2 (C3′, C5′), 126.0 (C2′, C6′), 121.2 (C1′), 108.0 (C6), 104.2 (C3), 104.0 (C10), 94.2 (C8), 81.0 (C5″), 78.7 (C3″), 74.1 (C1″), 71.6 (C2″), 70.4 (C4″), 61.4 (C6″), 28.2 (CH_2), 13.9

Please cite this article in press as: Hao, B., et al. Biosynthesis of natural and novel *C*-glycosylflavones utilising recombinant *Oryza sativa C*-glycosyltransferase (OsCGT) and *Desmodium incanum* root proteins. Phytochemistry (2016), http://dx.doi.org/10.1016/j.phytochem.2016.02.013

(CH3). ESIMS *m/z* (CV = 50, rel. int.): 443 $[M-H]^-$ (100); ESIMS *m/z* (CV = 150, rel. int.): 433 $[M-H]^-$ (1), 353 $[M-H-90]^-$ (2), 313 $[M-H-120]^-$ (36).

4.2.17. 3′-Fluorovitexin **5r**

UV (λ_{max} MeOH/H_2O) 269, 327 nm. ^{1}H NMR (500 MHz, d_4-MeOD, 325 K) δ 7.77 (1H, d, *J* = 13.0 Hz, H2′), 7.70 (1H, d, *J* = 8.7 Hz, H6′), 7.01 (1H, t, *J* = 8.7 Hz, H5′), 6.56 (1H, s, H3), 6.26 (1H, s, H6), 6-*C*-β-Glu 5.05 (1H, d, *J* = 9.9 Hz, H1″), 4.10 (1H, t, *J* = 9.4 Hz, H2″), 3.96 (1H, br d, *J* = 10.8, H6a″), 3.82 (1H, dd, *J* = 5.8, 12.1 Hz, H6b″), 3.65 (1H, t, *J* = 9.3 Hz, H4″), 3.56 (1H, t, *J* = 8.8 Hz, H3″), 3.54 (1H, m, H5″). ^{13}C NMR (125 MHz, d_4-MeOD, 325 K) 181.7 (C4), 163.8 (C2), 163.6 (C7), 161.2 (C5), 156.4 (C9), 151.4 (d, *J* = 151.4 Hz, C3′), 151.3 (C4″), 123.5 (C6′), 120.4 (C1′), 118.7 (C5′), 114.0 (C2′), 103.9 (C8), 103.5 (C10), 102.4 (C3), 99.3 (C6), 81.4 (C5″), 79.0 (C3″), 74.5 (C1″), 71.7 (C2″), 71.0 (C4″), 61.9 (C6″). ESIMS *m/z* (CV = 150, rel. int.): 449 $[M-H]^-$ (12), 359 $[M-H-90]^-$ (10), 329 $[M-H-120]^-$ (100).

4.2.18. 3′-Fluoroisovitexin **6r**

UV (λ_{max} MeOH/H_2O) 271, 332 nm. ^{1}H NMR (500 MHz, d_4-MeOD, 310 K) δ 7.69 (1H, dd, *J* = 1.7, 12.2 Hz, H2′), 7.65 (1H, br d, *J* = 8.7 Hz, H6′), 7.04 (1H, t, *J* = 8.6 Hz, H5′), 6.61 (1H, s, H3), 6.52 (1H, s, H8), 6-*C*-β-Glu 4.93 (1H, d, *J* = 9.9 Hz, H1″), 4.18 (1H, t, *J* = 9.1 Hz, H2″), 3.90 (1H, dd, *J* = 2.0, 12.0, H6a″), 3.78 (1H, dd, *J* = 5.1, 12.1 Hz, H6b″), 3.54–3.49 (2H, m, H3″, H4″), 3.45 (1H, m, H5″). ^{13}C NMR (125 MHz, d_4-MeOD, 310 K) 182.0 (C4), 164.1 (C7), 163.2 (C2), 160.4 (C5), 156.8 (C9), 152.3 (d, *J* = 230.0 Hz, C3′), 151.3 (C4″), 123.1 (C6′), 121.4 (C1′), 118.2 (C5′), 113.6 (C2′), 107.8 (C6), 103.6 (C10), 102.9 (C3), 94.0 (C6), 81.2 (C5″), 78.8 (C3″), 74.0 (C1″), 71.3 (C2″), 70.4 (C4″), 61.5 (C6″). ESIMS *m/z* (CV = 150, rel. int.): 449 $[M-H]^-$ (17), 359 $[M-H-90]^-$ (57), 329 $[M-H-120]^-$ (100).

4.2.19. 4′-Fluoro-8-*C*-glucosylchrysin **5t**

UV (λ_{max} MeOH/H_2O) 269, 316 nm. ^{1}H NMR (500 MHz, d_4-MeOD, 325 K) δ 8.17 (2H, br t, *J* = 6.6 Hz, H2′, H6′), 7.32 (2H, t, *J* = 8.7 Hz, H3′, H5′), 6.72 (1H, s, H3), 6.32 (1H, s, H6), 6-*C*-β-Glu 5.06 (1H, d, *J* = 10.0 Hz, H1″), 4.10 (1H, t, *J* = 9.3 Hz, H2″), 3.96 (1H, dd, *J* = 2.0, 12.1, H6a″), 3.85 (1H, dd, *J* = 5.1, 12.0 Hz, H6b″), 3.67 (1H, br t, 9.4 Hz, H4″), 3.57 (1H, t, *J* = 8.8 Hz, H3″), 3.52 (1H, m, H5″). ^{13}C NMR (125 MHz, d_4-MeOD, 325 K) 182.6 (C4), 165.0 (d, *J* = 250 Hz, C4′), 164.0 (C5), 163.5 (C7), 163.5 (C2), 156.7 (C9), 129.2 (d, *J* = 8.8 Hz, C2′, C6′), 128.0 (C1′), 115.7 (d, *J* = 22.5 Hz, C3′, C5′), 104.4 (C10), 104.4 (C3), 104.2 (C8), 98.9 (C6), 81.4 (C5″), 78.9 (C3″), 74.5 (C1″), 71.6 (C2″), 70.2 (C4″), 61.4 (C6″). ^{19}F NMR (470 MHz, d_4-MeOD, 325 K)at −109.9. ESIMS *m/z* (CV = 150, rel. int.): 433 $[M-H]^-$ (15), 343 $[M-H-90]^-$ (7), 313 $[M-H-120]^-$ (100).

4.2.20. 4′-Fluoro-6-*C*-glucosylchrysin **6t**

UV (λ_{max} MeOH/H_2O) 271, 317 nm. ^{1}H NMR (500 MHz, d_4-MeOD, 325 K) δ 8.01 (2H, dd, *J* = 5.2, 8.9 Hz, H2′, H6′), 7.27 (2H, t, *J* = 8.7 Hz, H3′, H5′), 6.69 (1H, s, H3), 6.52 (1H, s, H6), 6-*C*-β-Glu 4.92 (1H, d, *J* = 9.9 Hz, H1″), 4.11 (1H, t, *J* = 8.8 Hz, H2″), 3.86 (1H, dd, *J* = 2.2, 12.0, H6a″), 3.75 (1H, dd, *J* = 5.1, 12.2 Hz, H6b″), 3.52–3.46 (2H, m, H3″, H4″), 3.44 (1H, m, H5″). ^{13}C NMR (125 MHz, d_4-MeOD, 325 K) 182.8 (C4), 165.3 (d, *J* = 247.5 Hz, C4′), 163.5 (C7), 163.5 (C2), 160.9 (C5), 157.8 (C9), 128.5 (d, *J* = 10.0 Hz, C2′, C6′), 127.8 (C1′), 115.8 (d, *J* = 22.5 Hz, C3′, C5′), 108.3 (C6), 104.8 (C10), 104.6 (C3), 94.1 (C8), 81.1 (C5″), 78.8 (C3″), 74.0 (C1″), 71.4 (C2″), 70.4 (C4″), 61.4 (C6″). ^{19}F NMR (470 MHz, d_4-MeOD, 325 K)at −110.5 (dt, *J* = 8.3 5.7 Hz). ESIMS *m/z* (CV = 150, rel. int.): 433 $[M-H]^-$ (15), 343 $[M-H-90]^-$ (20), 313 $[M-H-120]^-$ (100).

4.3. Generation of 6-*C*-2-hydroxyglucosylflavanones from **3** using recombinant OsCGT

A His-tagged OsCGT gene from rice (a gift from R Edwards, Newcastle University) was transformed into *E. coli* strain BL21 (NEB). Cultures were grown in LB broth containing 50 μg/ml kanamycin at 30 °C and shaken at 200 rpm. At OD 600 nm 0.6–1, protein expression was induced by addition of 1 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG) and carried out for four hours. The OsCGT protein was purified according to precedent (Liu et al., 2002), with the following modifications; Tris buffer was substituted with 50 mM NaH_2PO_4 at pH 8. Cell lysis by sonication was performed in 50 mM NaH_2PO at pH 8, 300 mM NaCl, 10 mM imidazole, 1 mM benzamidine, 0.1 mM phenylmethanesulfonyl fluoride (PMSF) and 2 mM dithiothreitol (DTT). Purified protein was concentrated by centrifugation using 30 KDa cut-off columns (Vivaspin) and re-suspended in 50 mM HEPES buffer at pH 8. Protein was concentrated to approximately 1 mg/ml before being snap frozen in 0.2 ml aliquots and stored at −80 °C. 2-Hydroxyflavanone **3** (40 μL of a 20 mmol solution) in ethanol was added to 1 mg of UDP-glucose dissolved in 40 μL buffer with 50 μg recombinant OsCGT (Hamilton et al., 2012) dissolved in 200 μl 100 mM HEPES pH 7.5 buffer and incubated at 30 °C overnight. The assay was quenched with ice cold MeOH (250 μL) to precipitate protein which was removed by centrifugation at 4 °C. 1.5 assay equivalents of the crude *C*-glucosyl-2-hydroxyflavanone was then used directly in a single subsequent *D. incanum* enzyme assay.

4.4. D. incanum soluble root protein

Roots of *D. incanum* were snap frozen in liquid nitrogen and blended in ice cold 5 v/w 100 mM HEPES pH 7.2, 2 mM DTT, 2.5% PVP, 0.5 mM EDTA, 1 mM benzamidine, 0.1 mM PMSF. Protein was precipitated between 40% and 80% $(NH_4)_2SO_4$ and the pellet recovered by centrifugation. The protein pellets were desalted in 20 mM HEPES-NaOH pH 8 containing 2 mM DTT using Sephadex G-25 gel filtration chromatography (PD-10 columns, GE Healthcare) and the desalted solution concentrated by centrifugation using 30 KDa cut-off columns (Vivaspin) for assays.

4.5. Mono-*C*-glycosylflavone biosynthesis assays with D. incanum root proteins

D. incanum root protein (1 g equivalent as obtained in 4.4) was added to 2-hydroxyflavanone **3** (40 μL of a 20 mmol solution in ethanol) with 1 mg of UDP-glucose dissolved in 40 μL buffer (100 mM HEPES, 2 mM DTT, pH 7.5) and the solution made up to 400 μl with more buffer. The assay was incubated at 30 °C overnight and quenched with 1 M HCl. The solution was cleared by centrifugation and the supernatant decanted, dried under a stream of nitrogen and dissolved in MeOH (250 μL) for LCMS analysis.

4.6. Di-*C*-glycosylflavone biosynthesis assays with D. incanum root proteins

D. incanum root protein (1 g equivalent as obtained in 4.4) was added to the 6-glucosyl-2-hydroxyflavanone substrate **4** from 4.3 (1.5 assay equivalents dissolved in buffer) with 1 mg of UDP-glucose dissolved in 40 μL buffer (100 mM HEPES, 2 mM DTT, pH 7.5) and the solution made up to 400 μl with more buffer. The assay was incubated at 30 °C overnight and quenched with 1 M HCl. In the case of assays with UDP-arabinose they were conducted with 0.1 mg UDP-arabinose due to the cost and availability of the sugar donor. In the case of substrate **4b**, the entire assay was carried out on 1/10th scale using UDP-α-D-[UL-$^{13}C_6$]glucose. All assays were quenched 1 M HCl (250 μl), centrifuged and decanted.

Please cite this article in press as: Hao, B., et al. Biosynthesis of natural and novel *C*-glycosylflavones utilising recombinant *Oryza sativa C*-glycosyltransferase (OsCGT) and *Desmodium incanum* root proteins. Phytochemistry (2016), http://dx.doi.org/10.1016/j.phytochem.2016.02.013

The resulting solution was evaporated under a stream of nitrogen and the residue re-dissolved in MeOH (250 µl) for LCMS analysis.

4.7. HPLC analysis of OsCGT enzyme assays

The residue from a single enzyme assay was dissolved in 500 µL 3:1 H_2O/MeOH for HPLC analysis on a Shimadzu VP series HPLC system using an ACEQ C-18 column (250 mm × 4.6 mm, 5 µm). The mobile phase used a gradient program at 1 mL/min, initially 95:5 (A:B), to 85:15 at 3 min, 75:25 at 13 min, 70:30 at 25 min, 45:55 at 35 min, 45:55 at 45 min, 5:95 at 46 min, 5:95 at 58 min, 95:5 at 60 min (Ferreres et al., 2003).

4.8. LCMS analysis of exudates

The enzyme assay products were subject to LCMS either as they were or diluted tenfold in MeOH. The mass spectrometer was operated in negative ion mode, with a capillary voltage of 2.7 kV, cone voltage 50–180 eV, mass range 50–1000 *m/z*. Source temperature 130 °C, desolvation temperature 350 °C, desolvation gas flow 1000 L/h (nitrogen) and cone gas flow 60 L/h (nitrogen). Where required, selected ions were admitted to the collision cell for MSMS analysis with argon admitted at a pressure of 2.1 e^{-3} mbar, causing CID. Samples were injected *via* the Acquity sample manager, injecting 2 µl onto an Acquity UPLC BEC HSS C18 1.8 µm 2.1 × 150 mm column. Run time was 42 min at a flow rate of 0.21 mL/min. Solvents used are defined A (water, 0.05% formic acid) and B (methanol). The mobile phase used a gradient program, initially 95:5 (A:B), to 85:15 at 1.8 min, 75:25 at 7.7 min, 70:30 at 17.9 min, 45:55 at 23.8 min, 45:55 at 26.8 min, 5:95 at 30.4 min, 5:95 at 37.5 min, 95:5 at 38.7 min, 95:5 at 40 min. Standards for comparison were obtained previously from *Desmodium* plant tissue and fully characterised by NMR spectroscopy (Hooper et al., 2015).

Acknowledgements

Rothamsted Research receives Grant-aided support from the Biotechnology and Biological Sciences Research Council (BBSRC). This project was funded through the Bill and Melinda Gates Foundation funded project 'Integrated Striga Management for Africa', and the European Union funded project 'Adapting Push-Pull Technology for Climate Change'.

BH was supported through a China Scholarship Council Award.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.phytochem.2016.02.013.

References

Akiyama, K., Matsuoka, H., Hayashi, H., 2002. Isolation and identification of a phosphate deficiency induced *C*-glycosylflavonoid that stimulates arbuscular mycorrhiza formation in melon roots. Mol. Plant Microbe Int. 15, 334–340.

Brazier-Hicks, M., Evans, K.M., Gershater, M.C., Puschmann, H., Steel, P.G., Edwards, R., 2009. The *C*-glycosylation of flavonoids in cereals. J. Biol. Chem. 284, 17926–17934.

Bungaruang, L., Gutmann, A., Nidetzky, B., 2015. β-Cyclodextrin improves the solubility and enzymatic *C*-glucosylation of the flavonoid phloretin. Adv. Synth. Catal. http://dx.doi.org/10.1002/adsc.201500838.

Byrne, P.F., McMullen, M.D., Snook, M.E., Musket, T.A., Theuri, J.M., Widstrom, N.W., Wiseman, B.R., Coe, E.H., 1996. Quantitative trait loci and metabolic pathways: genetic control of the concentration of maysin, a corn earworm resistance factor, in maize silks. Proc. Natl. Acad. Sci. 93, 8820–8825.

Chen, D., Chen, R., Wang, R., Lim, J., Kebo, X., Bian, C., Sun, L., Zhang, X., Liu, J., Ye, F., Yu, X., Dai, J., 2015. Probing the catalytic promiscuity of a region- and stereospecific *C*-glycosyltransferase from *Mangifera indica*. Angew. Chem. Int. Ed. 54, 12678–12683.

Chen, G., Wu, X., Zhou, W.-L., Li, L., 2010. Preparation and assay of *C*-glucosyltransferase from roots of *Pueria lobata*. J. Exper. Bot. 31, 655–660.

Chul-Sa, K., Kabir, M.A., Sachi, M., Shin-ichi, T., Hen-Sik, K., 2009. Antifeedants of Indian barnyard millet, *Echinochloa frumentacea* Link. aganst brown planthopper *Nilaparvata lugens* (Stal). Z. Naturforsch 63c, 755–760.

Cuyckens, F., Claeys, M., 2004. Mass spectrometry in the structural analysis of flavonoids. J. Mass Spectrom. 39, 1–15.

Ferreres, F., Silva, B.M., Andrade, P.B., Seabra, R.M., Ferreira, M.A., 2003. Approach to the study of *C*-glycosyl flavones by ion trap HPLC-PAD-ESI/MS/MS: application to seeds of quince (*Cydonia oblonga*). Phytochem. Anal. 14, 352–359.

Ferreyra, M.L.F., Rodriguez, E., Casas, M.I., Labadie, G., Grotewold, E., Casati, P., 2013. A bifunctional maize *C*- and *O*-glucosyltransferase. J. Biol. Chem. 288, 31678–31688.

Furuta, T., Nakayama, M., Suzuki, H., Tajimi, H., Inai, M., Nukaya, H., Wakimoto, T., Kan, T., 2009. Concise synthesis of chafurosides A and B. Org. Lett. 11 (11), 2233–2236.

Hamilton, M.L., Kuate, S.P., Brazier-Hicks, M., Caulfield, J.C., Rose, R., Edwards, R., Torto, B., Pickett, J.A., Hooper, A.M., 2012. Elucidation of the biosynthesis of the di-*C*-glycosylflavone isoschaftoside, an allelopathic component from *Desmodium* spp. that inhibits *Striga* spp. development. Phytochemistry 84, 169–176.

Hamilton, M.L., Caulfield, J.C., Pickett, J.A., Hooper, A.M., 2009. *C*-Glucosylflavonoid biosynthesis from 2-hydroxynaringenin by *Desmodium uncinatum* (Jacq.) (Fabaceae). Tetrahedron Lett. 50, 5656–5659.

Hirade, Y., Kotoku, N., Terasaka, K., Saijo-Hamano, Y., Fukumoto, A., Mizukami, H., 2015. Identification and functional analysis of 2-hydroxyflavanone *C*-glucosyltransferase in soybean (*Glycine max*). FEBS Lett. 589, 1778–1786.

Hooper, A.M., Hassanali, A., Chamberlain, K., Khan, Z.R., Pickett, J.A., 2009. New genetic opportunities from legume intercrops for controlling *Striga* spp. parasitic weeds. Pest Manag. Sci. 65, 546–552.

Hooper, A.M., Tsanuo, M.K., Chamberlain, K., Tittcomb, K., Scholes, J., Hassanali, A., Khan, Z.R., Pickett, J.A., 2010. Isoschaftoside, a *C*-glycosylflavonoid from *Desmodium uncinatum* root exudate, is an allelochemical against the development of *Striga*. Phytochemistry 71, 904–908.

Hooper, A.M., Caulfield, J.C., Hao, B., Pickett, J.A., Midega, C.A.O., Khan, Z.R., 2015. Isolation and identification of *Desmodium* root exudates from drought tolerant species used as intercrops against *Striga hermonthica*. Phytochemistry 117, 380–389.

Hao, B., Caulfield, J.C., Hamilton, M.L., Pickett, J.A., Midega, C.A.O., Khan, Z.R., Wang, J.R., Hooper, A.M., 2015. The biosynthesis of allelopathic di-*C*-glycosylflavones from the roots of *Desmodium incanum* (G. Mey.) DC. Org. Biomol. Chem. http://dx.doi.org/10.1039/C5OB01926E. Advance Article.

Khan, Z.R., Hassanali, A., Overholt, W., Khamis, T.M., Hooper, A.M., Pickett, J.A., Wadhams, L.J., Woodcock, C.M., 2002. Control of the witchweed *Striga hermonthica* by intercropping with *Desmodium* spp., and the mechanism defined as allelopathic. J. Chem. Ecol. 28 (9), 1871–1885.

Khan, Z.R., Midega, C.A.O., Pittchar, J.O., Murage, A.W., Birkett, M.A., Bruce, T.J.A., Pickett, J.A., 2014. Achieving food security for one million sub-Saharan African poor through push–pull innovation by 2020. Phil. Trans. Royal Soc.-B 1639, 369.

Kim, Y.-C., Jun, M., Jeong, W.-S., Chung, S.-K., 2005. Antioxidant properties of flavone *C*-glycosides from *Atractylodes japonica* leaves in human low-density lipoprotein oxidation. J. Food Sci. 70, 575–580.

Kitta, R., Hagiwara, Y., Shibamoto, T., 1992. Antioxidative activity of an isoflavonoid, 2″-*O*-glycosylisovitexin isolated from green barley leaves. J. Agric. Food Chem. 40, 1843–1845.

Kumazawa, T., Ohki, K., Ishida, M., Sato, S., Onodera, J., Matsuba, S., 1995. Practical synthesis of a *C*-glycosyl flavonoid via *O*-*C* glycoside rearrangement. Bull. Chem. Soc. Jpn. 68, 1379–1384.

Kumazawa, T., Kimura, T., Matsuba, S., Sato, S., Onodera, J., 2001. Synthesis of 8-*C*-glucosylflavones. Carbohydr. Res. 334, 183–193.

Lee, E.M., Lee, S.S., Chung, B.Y., Cho, J.Y., Lee, I.C., Ahn, S.R., Jang, S.J., Kim, T.H., 2010. Pancreatic lipase inhibition by *C*-glycosidic flavones isolated from *Eremochloa ophiuroides*. Molecules 15, 8251–8259.

Liu, Z., Zhang, J., Chen, X., Wang, P.G., 2002. Combined biosynthetic pathway for *de novo* production of UDP-galactose: catalysis with multiple enzymes immobilized on agarose beads. ChemBioChem 3 (4), 348–355.

Mahliing, J.-A., Jung, K.-H., Schmidt, R.R., 1995. Synthesis of flavone *C*-glycosides vitexin, isovitexin and isoembigenin. Liebigs Ann., 461–466

McNally, D.J., Wurms, K.V., Labbe, C., Belanger, R.R., 2003. Synthesis of *C*-glycosylflavonoid phytoalexins as a site-specific response to fungal penetration in cucumber. Physiol. Mol. Plant Pathol. 63, 293–303.

Nagatomo, Y., Usui, S., Ito, T., Kato, A., Shimosaka, M., Taguchi, G., 2014. Purification, molecular cloning and functional characterization of flavonoid *C*-glucosyltransferases from *Fagopyrum esculentum* M. (buckwheat) cotyledon. Plant J. 80, 437–448.

Nichenametla, S.N., Taruscio, T.G., Barney, D.L., Exon, K.H., 2006. A review of the effects and mechanisms of polyphenolics in cancer. Crit. Rev. Food Sci. Nutr. 46, 161–183.

Talhi, O., Silva, A.M.S., 2012. Advances in *C*-glycosylflavonoid research. Curr. Org. Chem. 16 (7), 859–896.

Sasaki, N., Nishizaki, Y., Yamada, E., Tatsuzawa, F., Nakatsuka, T., Takahashi, H., Nishihara, M., 2015. Identification of the glucosyltransferase that mediates direct flavone *C*-glucosylation in *Gentiana triflora*. FEBS Lett. 589, 182–187.

Sato, S., Akiya, T., Suzuki, T., Onodera, J., 2004. Environmentally friendly *C*-glycosylation of phloroacetophenone with unprotected D-glucose using scandium (III) trifluoromethanesulfonate in aqueous media: key compounds for the synthesis of mono- and di-*C*-glucosylflavonoids. Carbohydr. Res. 339, 2611–2614.

Sato, S., Koide, T., 2010. Synthesis of vicenin-1 and 3,6,8- and 8,6-di-*C*-β-D-(glucopyranosyl xylopyranosyl)-4′,5,7-trihydroxyflavones using two direct *C*-glycosylations of naringenin and phloroacetophenone with unprotected D-glucose and D-xylose in aqueous solution as the key reactions. Carbohydr. Res. 345, 1825–1830.

Sato, S., Hiroe, K., Kumazawa, T., Onodera, J., 2006a. Total synthesis of two isoflavone *C*-glycosides: genistein and orobol 8-*C*-β-D-glucopyranosides. Carbohydr. Res. 341, 1091–1095.

Sato, S., Akiya, A., Suzuki, N., Suzuki, T., 2006b. Total synthesis of three naturally occurring 6,8-Di-*C*-glycosylflavonoids: phloretin, naringenin and apigenin bis-*C*-β-D-glucosides. Carbohydr. Res. 341, 964–970.

Santos, R.G., Xavier, N.M., Bordado, J.C., Rauter, A.P., 2013. Efficient and first region- and stereoselective direct *C*-glycosylation of a flavanone catalysed by $Pr(OTf)_3$ under conventional heating or ultrasound irradiation. Eur. J. Org. Chem., 1441–1447.

Please cite this article in press as: Hao, B., et al. Biosynthesis of natural and novel *C*-glycosylflavones utilising recombinant *Oryza sativa C*-glycosyltransferase (OsCGT) and *Desmodium incanum* root proteins. Phytochemistry (2016), http://dx.doi.org/10.1016/j.phytochem.2016.02.013