THE C-GLYCOSYLATION OF FLAVONOIDS IN CEREALS*

Melissa Brazier-Hicks, Kathryn M. Evans, Markus C. Gershater, Horst Puschmann, Patrick G. Steel, Robert Edwards

Centre for Bioactive Chemistry, Durham University, Durham DH1 3LE, UK.

Running head: C-Glycosylation of Flavonoids in Cereals

Address correspondence to: Robert Edwards, Centre for Bioactive Chemistry, Durham University, Durham DH1 3LE, UK. E-mail: Robert.Edwards@durham.ac.uk

Flavonoids normally accumulate in plants as O-glycosylated derivatives, but several species, including major cereal crops, predominantly synthesize flavone-C-glycosides, which are stable to hydrolysis and are biologically active both in planta and as dietary components. An enzyme (OsCGT) catalyzing the UDP-glucose-dependent C-glucosylation of 2-hydroxyflavanone precursors of flavonoids has been identified and cloned from rice (Oryza sativa ssp. indica), with a similar protein characterized in wheat (Triticum aestivum L.). OsCGT is a 49 kDa family 1 glycosyltransferase related to known O-glucosyltransferases (OGTs). The recombinant enzyme C-glucosylated 2-hydroxyflavanones, but had negligible OGT activity with flavonoid acceptors. Enzyme chemistry studies suggested that OsCGT preferentially C-glucosylated the dibenzoylmethane tautomers formed in equilibrium with 2-hydroxyflavanones. The resulting 2-hydroxyflavonone-C-glycosides were unstable and spontaneously dehydrated in vitro to yield a mixture of 6C- and 8C-glycosyl derivatives of the respective flavones. In contrast, in planta only the respective 6C-glycosides accumulated. Consistent with this selectivity in glycosylation product, a dehydratase activity which preferentially converted 2-hydroxyflavonone-C-glycosides to the corresponding flavone-6C-glycosides was identified in both rice and wheat. Our results demonstrate that cereal crops synthesize C-glycosylated flavones through the concerted action of a CGT and dehydratase acting on activated 2-hydroxyflavanones, as an alternative means of generating flavonoid metabolites.

INTRODUCTION

The glycosylation of natural products with sugars through carbon-carbon bonds is a biochemically demanding reaction which gives rise to stable metabolites exhibiting the combined activity of both the secondary metabolite acceptor and sugar (1). C-Glycosides are formed in microbes, plants and insects, where they serve a diverse range of functions including acting as siderophores, antibiotics, antioxidants, attractants and feeding deterrents (1, 2). Despite their importance in conferring biological activity, the C-glycosyltransferases (CGTs) responsible for forming these glycosidic bonds have attracted relatively little attention. As a rare exception, a CGT which catalyzed the C-glycosylation of the siderophore enterobactin has been characterized in Escherichia coli (3). Similarly, an enzyme (UrdGT2) which C-conjugated a polyketide intermediate with D-olivose has also been identified as a component of the pathway leading to the biosynthesis of the antibiotic urdamycin A in Streptomyces fradiae (4, 5). Analysis of the amino acid sequences of these two CGTs place them in family 1 of the 91 glycosyltransferase families classified to date (6). Family 1 enzymes are invertase glycosyltransferases which utilise nucleotide-diphospho-sugars as activated donors to conjugate small molecule acceptors, most typically to form ether glycosidic bonds. The fact that microbial CGTs are related to enzymes which exhibit O-glycosyltransferase (OGT) activity suggests that relatively minor modifications to active site chemistry facilitate the more unusual C-conjugation. In the case of UrdGT2, CGT activity appears to be associated with the presence of a unique aspartate residue which activates the acceptor for C9 glycosylation (5). Intriguingly UrdGT2 also O-glycosylates artificial substrates (5), confirming that in microorganisms, there are
no fundamental differences in the evolutionary origins of CGTs and OGTs. However, while the enzyme chemistry of OGTs has been well described, the exact mechanism by which C-glycosylation is achieved is still poorly understood (7).

C-glycosylation in plants has received little attention, despite the common occurrence of such secondary metabolites in major cereal crops and medicinal species (2). The most commonly abundant C-glycosylated natural products in plants are the flavonoids, a large group of polyphenolic compounds with diverse protective and attractant functions (8). Flavonoids normally accumulate in the vacuoles of plant tissues as their respective O-linked glycosidic conjugates (Fig. 1A; compound 3). However, in at least 20 families of angiosperms, flavonoids also accumulate as the respective C-glycosides (8). As such, these derivatives are major secondary metabolites in maize, wheat and rice (2,8). In these cereals, C-glycosides of the simple flavones apigenin and/or luteolin predominate, with conjugation occurring singly or doubly at the C-8 and/or C-6 position (Fig. 1B). Activities ascribed to these plant secondary metabolites include them functioning as antioxidants (9, 10), insect feeding attractants (11), antimicrobial agents (12), promoters of mycorrhizal symbioses (13) and UV-protective pigments (14). From a dietary perspective, these compounds have also been ascribed both positive and negative biological activities. Thus, in vitro, flavone-C-glycosides can counteract tissue oxidation (15), inflammation and cancer development (16). However, millet diets containing high levels of C-glucosylflavones have been shown to suppress thyroid iodine uptake in rats and have the potential to cause goitrogenic effects (17).

Relatively little is known about flavone-C-glycoside biogenesis. The flavonones, which are core intermediates of the flavonoid pathway are the most likely precursors (Fig.1A; 1). Studies in buckwheat (Fagopyrum esculentum), demonstrated that 2-hydroxyflavanones (Fig. 1B; 4a/b) underwent enzyme catalyzed C-glycosylation and were the direct precursors of flavone-C-glycosides (18,19). However, the identity of the respective CGTs has not been determined and the biochemistry underlying this unusual conjugation remains unresolved. With an interest in the natural products chemistry and the biotechnological applications of this important branch of plant secondary metabolism, we now report on the purification, identification and characterization of CGTs responsible for flavone-C-glycoside synthesis in rice (Oryza sativa ssp. indica) and wheat (Triticum aestivum L.).

**EXPERIMENTAL PROCEDURES**

**CGT activity determination-** Acceptor substrates were purchased from Aldrich, Alfa Aesar and Apin chemicals. Benzyl 2,4,6-trihydroxybenzoate (20) and 2-hydroxyflavanones (18) were prepared using previously described methods, purified by reversed-phase HPLC and their identities confirmed by MS (21). Glucosyltransferase (CGT and OGT) activity was determined by incubating flavonoid acceptors (66 µM) with UDP-[14C]-glucose (50,000 dpm, 11.2 GBq mmol⁻¹) and assaying radioactive glycoside formation by scintillation counting after partitioning into organic solvent (22). To distinguish between O- and C-glucosylated products, the organic phase was dried down and treated with 6% HCl at 100 °C for 1 h, prior to selective recovery and assay of the acid stable C-glycosides. [14C]-glucosylated products were also analyzed by TLC and autoradiography (23).

**Purification and identification of a CGT-** Rice (Oryza sativa ssp. japonica cv. Nipponbare) and wheat (Triticum aestivum cv. Einstein) seedlings were grown at 25 °C for 10 days and the shoots extracted as described (21). Ammonium sulfate (0-80% saturation) protein precipitates were resuspended in buffer A = 20 mM Tris-HCl pH 8.0 containing 2 mM dithiothreitol (DTT), adjusted to 1 M (NH₄)₂SO₄ and applied (4 ml min⁻¹) onto a phenyl Sepharose column (40 ml). Retained protein was recovered by decreasing the (NH₄)₂SO₄ concentration (1.0 - 0 M) over 300 ml and fractions (8 ml) analyzed for CGT activity. Combined active fractions were dialyzed overnight in buffer A, applied onto a mono Q column (1 ml) and protein eluted with a linear gradient of 0 - 0.25 M NaCl (1 ml min⁻¹). Active fractions were then loaded on to a Superdex 200 column eluted in buffer A containing 0.15 M NaCl (0.5 ml min⁻¹). Fractions were monitored by SDS-PAGE and polypeptides whose relative abundance matched the elution of CGT activity.
excised and analyzed by MALDI-TOF-MS proteomics after tryptic digestion (22). Raw spectra were peak picked using MASCOT Wizard (www.matrixscience.com; 24) with a correlation threshold of 0.75 and data picked from m/z 800 - 3500. The peak list was used to search the NCBI database using MASCOT. Parameters used were a peptide tolerance of 50 ppm, a maximum of one missed cleavage site, fixed carbamidomethyl sites and an allowance made for the oxidation of methionine residues. The program MODELLER (25) was subsequently used to build a homology model based on the crystal structure of UGT72B1 (26). Model scores of 1 and the combined model quality score of 1.36 indicated reliability in the model as expected with a sequence identity of 35 % (27).

CGT cloning and expression- Primers 5'-cgcgcgecatagccgctctggcgacg-3' and 5'-cgcgcgectctgtgcaattgagctcctc-3' were used to amplify the coding sequence of ABC94602 from gDNA prepared from rice shoots (28). The 1.4 kb amplification product was cloned into a custom prepared pET-STRP3 vector for expression of the respective N-terminal Strep-tagged fusion protein, which was purified by affinity chromatography (29).

Characterization of C-glucosylation reaction products- Recombinant CGT (0.75 µg) was incubated with acceptor and UDP-glucose (both 1 mM). The products were analyzed before and after treatment with 0.3 M HCl (100 °C , 60 min), by LC-MS on an Acquity UPLC™ BEH C18 (1.7 µM, 2.1 x 100 mm) column eluted with a 25 min gradient of 5% - 95% acetonitrile in 0.5 % aqueous formic acid at 0.2 ml min⁻¹. The eluent was passed into a Micromass Q-TOF Premier spectrometer after electrospray ionisation (ESI; capillary 2.55 kV, sample cone 41 kV, extraction cone 5.0 kV, source 100 °C with desolvation at 180 °C). Samples were analyzed in negative ion mode, with collision energies ramped from 10 V to 30 V for fragmentation analysis. ¹H and ¹³C NMR spectra were measured on a Bruker Advance 500 or Varian Inova-500 instrument and assignments carried out using COSY, HSQC/HMQC, HMBC, and NOESY experiments. NMR samples were acquired using the deuterated solvent as the lock and the residual solvent as the internal reference (CD₃OD: δH = 3.34 ppm, δC = 49.9 ppm and CD₃CN: δH = 1.96 ppm, δC = 118.3 ppm). The purified flavone-C-glucosides were recovered from CD₃OD, with compound 7 (C₂₂H₂₇,4O₁₀,87₅) forming single crystals which were repeatedly re-crystallized from wet methanol until suitable for diffraction analysis. Crystals were mounted in inert oil and transferred to the cold gas stream of the diffractometer. Crystal data: C₂₂H₂₇,4O₁₀,87₅, Mᵣ = 476.35, Triclinic, a = 7.7417(3), b = 12.5623(4), c = 12.7758(4), α = 66.9400(10), β = 86.9340(10), γ = 82.442(2), U = 1133.28(7) Å³, T = 120(2), space group P1 (no.1), Z = 2, µ(Mo-Kα) = 0.112, 4932 reflections measured, 4932 unique (Rint = 0.0471) which were used in all calculations. The final wR(F²) was 0.1010 (all data) and the final R1(data > 2σ) was 0.0448. (CCDC entry No: 670429).

Assay of the 2-hydroxyflavanone-C-glucoside dehydratase- Rice suspension cultures (21) were extracted with 3 v/v buffer B = 100 mM HEPES-OH, pH 7.5, containing 2 mM DTT, centrifuged (8500 g, 5 min, 4 °C) and the supernatant desalted on Sephadex G-25. The 2,5,7-hydroxyflavanone-C-glucoside substrate was biosynthesized by incubating 4 mM 2,5,7-hydroxyflavanone and 8 mM UDP-glucose with 20 µg recombinant OleCGT in buffer B (200 µl) for 4 h at 37 °C. On completion, an equal volume of methanol was added to give a final concentration of 2 mM 2,5,7-hydroxyflavanone-C-glucoside. Dehydratase activity was measured by mixing 0.2 mM 2,5,7-hydroxyflavanone-C-glucoside with 0.5 mg crude protein extract in buffer B (100 µl). After 20 min at 30 °C, the reactions were stopped with 1 v/v methanol and analyzed by HPLC-MS.

RESULTS

Flavone C-glycosylation in cereals. Rice seedlings were extracted and their component flavonoids separated by HPLC (Fig. 2A), prior to their identification by mass spectrometry (MS) with reference to published spectra (30, 31). In rice, a range of C-glycosylated-flavones derived from either apigenin or luteolin (Fig. 2B) were identified as the dominant UV-absorbing metabolites. In contrast, O-glycosylation was a minor route of conjugation, largely restricted to the poly-methylated flavone acceptor tricin (4',5,7-trihydroxy-3',5'-dimethoxyflavone).

Previous work on eight C-glycosylflavones from rice found that the 6C position was most
commonly substituted with one or more glucose residues, whereas arabinose was employed in the 8C position (30). Similarly in wheat, it has been demonstrated that the 6C position was consistently glucosylated, with xylose used in the 8C position (32). Based on the conservation in 6C-glucosylation observed in wheat and rice it was decided to focus on characterizing the associated glucose C-conjugating activity in these two plants.

Total glucosyltransferase (OGT plus CGT) activities were determined in wheat and rice using UDP-[14C]-glucose as the donor and the flavones chrysin, apigenin and luteolin as acceptors (Fig. 1). For chrysin and apigenin, the respective 2-hydroxyflavanones (Fig. 1; 4a/b), were also synthesized and tested as substrates. CGT could then be discriminated from OGT activity by quantifying the reaction products with and without a treatment with acid, which hydrolyzes O-glycosidic linkages but not C-linked conjugates (33). Based on these analyses and the resolution of the radioactive reaction products by TLC, it could be demonstrated that the extracts tested catalyzed the O-glucosylation of the flavone acceptors, while both C- and O-glucosides were formed with the 2-hydroxyflavanones (Supplemental Fig. 1). When the reaction products formed by the action of the rice preparations were analyzed by LC-MS, mass ions corresponding to the respective glucosides of both chrysin (m/z 415) and 2,5,7-trihydroxyflavone (m/z 433) were confirmed. In the case of chrysin, increasing the ionization energy caused the ether glycosidic bond to fragment, with the associated loss of 162 Da to give m/z 253. In contrast, the C-C linked glucoside formed with 2-hydroxyflavanone gave a very different fragmentation, with characteristic losses of 90 Da and 120 Da as opposed to the cleavage of the intact pyranoid ring (loss of 162 Da) (34). As determined by TLC, when 2-hydroxyflavanones were used as acceptors, a single acid-resistant C-glucoside was formed in rice extracts, while in wheat two glucosylated conjugates could be resolved, one of which was susceptible to chemical-, or enzyme- (β-glucosidase) mediated hydrolysis (Supplemental Fig. 1). Based on these studies, it was decided to purify the UDP-glucose-dependent CGT activity toward 2-hydroxyflavanone substrates from rice.

Isolation of a CGT from cereals. Using 2,5,7-trihydroxyflavonone as the substrate, CGT activity was purified 450-fold in 10% yield from extracts of rice shoots, using a combination of hydrophobic interaction, anion exchange and size exclusion chromatography (Fig. 3A to C). In each case, the majority of activity eluted as a single peak, with the analysis of the acid-stable [14C-glucose]-labelled reaction products at each stage confirming that the enzyme being purified was indeed a CGT. By collecting multiple fractions during the final purification step, it was possible to match changes in the relative abundance of four polypeptides, with molecular masses of between 45 and 55 kDa, with the elution of CGT activity (Fig. 3D). When a similar protein purification was applied to wheat shoots, the CGT was resolved into two peaks of activity, which were enriched 75-fold and 56-fold respectively, though in yields of less than 2% (Supplemental Fig. 2). Polypeptide content was also profiled against enzyme activity in the purified wheat preparations, with a 52 kDa polypeptide co-eluting with the CGT activity in the second peak (Supplemental Fig. 2). The rice and wheat proteins were then individually subjected to MALDI-TOF-MS or MS-MS based proteomics and assigned putative identities after interrogating available genome and EST databases (Supplemental Fig. 3A). Although the wheat proteins could not be identified, the four purified rice polypeptides were characterized (Fig. 3D), with bands 1 and 4 corresponding to family 1 UGTs (Genbank accessions EAZ03128 and ABC94602 respectively). The common occurrence of a 50 kDa polypeptide in the enriched CGT preparations from rice (band 4) and wheat focussed attention on the respective protein, with the coding sequence then amplified from rice foliage by PCR. The resulting 1.4 kb product, termed OsCGT, was identical to that predicted from the open reading frame of ABC94602 except for the single amino acid substitution G325D (Supplemental Fig. 3B; FM17912). This amino acid change appeared to be due to a polymorphism derived from the rice cultivar used, with the substitution reliably obtained in multiple independent amplifications.

To determine if OsCGT was indeed a CGT, it was then cloned into a modified pET vector as an N-terminal Strep-tag II fusion protein (29). The recombinant tagged protein was abundantly
expressed in the soluble fraction (6 mg protein L⁻¹ culture) and was purified to homogeneity using affinity chromatography. When analyzed by MS, the parent polypeptide (predicted mass = 51295.63 Da) was found to be processed to a mixture of derivatives arising from the cleavage of the terminal methionine (51164.01 Da) and the partial acetylation of the exposed alanine (51203.71 Da). Similar post-translationally modified products have been observed with other Strep-tagged plant proteins expressed in E. coli (29).

Characterization of a recombinant rice CGT. The purified Strep-tagged rice protein OsCGT was incubated with a range of flavonoid acceptors in the presence of UDP-[⁴C-glucose] and the reaction products quantified by radio-assaying (Fig. 4). The enzyme was highly active toward all three 2,5,7-trihydroxy-substituted flavanones (Fig. 1; 4a/b). In contrast, the enzyme showed negligible conjugating activity with flavanones having lower levels of oxygenation in the A ring (Fig. 4; substrates 1a-c), or with flavones (ii). Activities toward other flavonoids were also tested (Fig. 4). While naringenin chalcone (iv) proved to be a poor substrate, the corresponding reduced 2',4',6'-trihydroxydihydrochalcone was C-glucosylated, with an even higher activity observed with its 4-hydroxylated analogue, phoretin (vb). To characterize the reaction products in greater detail, all substrates which could be glucosylated by the recombinant enzyme were incubated with higher concentrations of unlabeled UDP-glucose. LC-MS/MS analysis was carried out on the freshly-derived reaction products. In all cases a single reaction product was observed which underwent losses of [M-H⁻’-90] and [M-H⁻’-120] ions as a result of the fragmentation of the glucose moiety. This, together with the absence of the [M-H⁻’-162] ion corresponding to the cleavage of the ether-linked glucoside confirmed that C-glucosides rather than O-glucosides had been produced. Assays with a further 28 OGT acceptor substrates further confirmed that the enzyme was an exclusive CGT (see supplemental data). The limited commercial availability of UDP-sugars made it difficult to exhaustively test the specificity of the OsCGT for sugar donors. In an alternative approach, the UDP-sugars present in rice seedlings which were accumulating both flavone-C-glucosides and -C-arabinosides (Fig. 2), were isolated (35). When these preparations were incubated with the OsCGT and the 2,5,7-trihydroxyflavanones tested previously (Fig. 4B), the reaction products were found to be identical to the C-linked conjugates formed when the enzyme was incubated with UDP-glucose. No other reaction products were observed and this experiment therefore demonstrated that OsCGT showed a marked preference for UDP-glucose over the other sugar donors isolated from the host plant.

Further analysis was focused on characterizing the CGT reaction with 2,5,7-trihydroxyflavanone as substrate (Fig. 1; 4, R¹ = R² = H). When this substrate was incubated with the recombinant OsCGT and then analysed by LC-MS (Fig. 5A) a single reaction product (peak 5) which increased over time (Supplemental Fig. 4A) was observed of parent mass m/z [M-H⁻’] 433. MS-MS of this product produced the characteristic fragmentation of the C-glucoside ([M-H⁻’-90] and [M-H⁻’-120]) ions along with a species of m/z 287 (Fig. 5B). This ion was consistent with fragmentation of the conjugate in the open chain dibenzoylmethane form (Fig. 1; 5a), rather than as the 2-hydroxyflavanone (5b/c). Similar fragmentation patterns were observed for the products formed from the other 2-hydroxyflavanones.

In the course of purifying the reaction product, peak 5 was converted into two resolvable compounds (Fig. 5A; peaks 5 and 6), which had identical masses, m/z [M-H⁻’] 415 (Fig. 5C), and accumulated in a ratio of 1:0.6. As compared with peak 5, the loss of 18 Da in peaks 6 and 7 was consistent with dehydration of the 2-hydroxyflavanone glucoside. Based on previous studies (17), it was concluded that these derivatives were the respective flavone 8C- and 6C-glucosidic Wessely-Moser isomers (Fig. 1; 7 & 6 respectively). Following their purification, peaks 5 and 7 were analyzed by ¹H and ¹³C NMR, using 2D methods (COSY, HMOC, HMBC), permitting detailed characterization of each conjugate (Supplemental Table 1). Peak 6 (Fig. 5A) was identified as the 6-C-glucosylflavone (Fig. 1; 6), with ¹H/¹³C bond correlations determined between the anomeric proton, H-1” (δ 4.97 ppm) and C-6 (δ 100.4 ppm). Additional correlations of H-1” were observed by HMBC to the signals corresponding to C-5 and C-7. The characteristic chemical shift of the anomeric
carbon, C-1’’ (observed at δ 76.1 ppm and confirmed by correlation to H-1’’ in the HMQC spectra), was in full agreement with a glucoside linkage at C-6 (36). The connectivity of the sugar was then characterized by COSY and NOESY experiments, with the chemical shifts of the respective carbon resonances being in agreement with those expected for a β-glucoside (36). Peak 7 (Fig. 5A) had similar NMR spectra to that of the 6C-glucosylflavone, with correlations in the HMBC spectrum determined between H-1’’ with C-8, C-7 and C-8a consistent with 8C-glucosylation (Fig 1; 7). Similarly, the downfield shift associated with H-8 in peak 6 compared to that of H-6 in peak 7 is also in agreement with the 8C-substitution of the flavone (37). Further confirmation of the identity of the C-glucosylchrysin isomers was sought by carrying out single crystal x-ray diffraction with the putative 8-C-glucoside (CCDC No: 670429: final R1 agreement factor at 4.48%). This confirmed the identity of peak 7 as chrysin-8C-β-glucoside, with the C-8 to C1´´ bond length of 1.505 Å, consistent with that expected for an sp3–sp2 C-C bond (Supplemental Fig. 5). These studies therefore confirmed that OsCGT catalyzed the C-glucosylation of 2,5,7-trihydroxyflavanone to finally yield the reaction products chrysin-8C-glucoside (Fig 1; 7) and chrysin-6C-glucoside (Fig. 1; 6) respectively.

Identification of a selective dehydratase responsible for flavone-6C-glucoside formation. Under in vitro conditions, the 2-hydroxyflavanone conjugates underwent spontaneous dehydration to yield a mixture of flavone-6C- and -8C-glucosides. However, in planta, flavone-6C-glucosides were preferentially formed (Fig. 2), suggesting that this dehydration is a controlled process. To determine if this dehydrating activity was enzyme mediated, crude protein extracts from dark grown rice cell suspension cultures were incubated with the 2,5,7-trihydroxyflavanone-C-glucoside (Fig. 1; 5; R1= R2= =H). In boiled protein controls, two reaction products corresponding to 8C-glucosylchrysin (7) and 6C-glucosylchrysin (6) were formed in a ratio of 1:0.5 respectively (Fig. 5D). In contrast, when incubated with crude protein extracts there was a selective and time-dependent increase in the amount of 6C-glucosylchrysin (6) observed (Supplemental Fig. 4B), while the amount of 8C-glucosylchrysin (7) produced was unaffected (Fig. 5D). It could subsequently be calculated that the selective dehydration to the 6C-glucoside corresponded to a catalyzed activity of 0.27 ± 0.02 pkat mg⁻¹ crude protein. Interestingly, the ratio of 6C- to 8C-glucosides produced by the enzyme preparation closely matched the relative abundance of the two isomers formed from heating the 2,5,7-trihydroxyflavanone-C-glucoside in acid. An identical selective protein-dependent dehydratase activity was also determined in extracts from etiolated wheat shoots corresponding to a specific activity of 0.4 ± 0.01 pkat mg⁻¹ crude protein.

DISCUSSION

Evolutionary origins and mechanism of a plant CGT. The rice enzyme OsCGT catalyzes the formation of the precursors of flavone-C-glucosides, with a protein with similar activities and physical characteristics identified in wheat. On construction of a phylogenetic tree, OsCGT is a family 1 glycosyltransferase related to a cluster of UGTs in rice of unknown function (Supplemental Fig. 6). The most similar proteins in Arabidopsis thaliana (27-29% identity), are group E UGTs which catalyze the O-glucosylation of monolignols (38, 39). In addition, the E-group member UGT72B1 is able to both O- and N-glucosylate xenobiotic pollutants (26). None of the group E UGTs are known to C-glucosylate acceptors, with UGT72B1 showing no activity toward the 2-hydroxyflavanone substrates in our hands (data not shown). Phylogenetic analysis clearly shows that OsCGT is not closely related to the bacterial CGTs, urdGT2 and IroB, bearing only 14% and 10% sequence identity respectively (Supplemental Fig. 7). No evidence of CGT-specific sequence motifs could be demonstrated on aligning OsCGT with the bacterial CGTs after allowing for residues which were also found in the plant OGTs (Supplemental Fig. 7). The protein structure of OsCGT was then compared by homology modelling with its most closely related (35% identity) ortholog UGT72B (pdb code 2vch) (26). Modelling revealed the overall folds of the two β/α/β Rossman domains to be clearly conserved in the two proteins, with only a few smaller insertions and deletions in the loop regions. Similarly, there were no obvious differences in the conformation of the active site.
residues which could account for the unusual C-conjugating activity of OsCGT, or the potential of the enzyme to use sugar donors other than UDP-glucose in the PSPG binding domain (40).

Unlike the UrdGT2 from Streptomyces fradiae, which catalyzed both the O- and C-glycosylation of dihydroxanthraquinones (41), OsCGT was an obligate C-glucosyltransferase. As sequence and structural analysis gave no immediate clues as to the enzymic origins of CGT activity, attention was focussed on the chemical nature of the glucose acceptor. The large number of C-glycosylated flavonoids observed in plants commonly contain a 5,7-dioxygenated substitution pattern, with only a few having a 5-deoxy-7-oxy-8-glycosyl skeleton. Such a structural pattern is consistent with an electrophilic aromatic substitution pathway. The exclusive formation of the β-C-glucoside is similarly consistent with such a mechanism, reflecting a single inversion during conjugation. On this basis we suggest that C-glycosylation proceeds as outlined in Figure 1. Initial generation of the 2-hydroxyflavanone substrate \(4\) from core flavanone intermediates occurs through the action of a cytochrome P450 enzyme flavanone 2-hydroxylase (42). While this P450 activity has not yet been identified in rice, a gene apparently encoding such an enzyme can be readily identified based on homology with the isoflavone synthase from soybean which catalyzes the 2-hydroxylation of flavonones (43). On formation, the meta-stable 2-hydroxyflavanones (Fig. 1; 4) exist in equilibrium with their open chain dibenzoylmethane species (Fig. 1; 4b). The open chain form of the acceptor substrate would certainly provide the additional strongly activating substituent for the observed electrophilic aromatic substitution by the donor UDP-glucose. Consistent with the need for additional activation, 2-hydroxyflavanones lacking either, or both, 5- and 7-hydroxy groups (Fig 4; i a-e), were ineffective substrates. Further support for the dibenzoylmethane form being the actual acceptor species was obtained from the MS/MS fragmentation of the respective glucoside formed by the action of OsCGT (Fig. 5B). The formation of other C-glycosylated products, (Fig. 4; vi/vi/vii d), does suggest that the dicarbonyl function is not essential for ring activation and accounts for the formation of anthrone and xanthone C-glucosides (2). Whilst the equilibrium between 2-hydroxyflavanone and dibenzoylmethane increases the nucleophility of the A-ring, it is also conceivable that the open chain form allows the substrate to adopt a favorable binding coordination with the sugar donor. Consistent with this, neither naringenin nor naringenin chalcone in which the conformation of the B-ring has limited mobility were effective substrates. However, reduction of the double bond to afford 2′,4′,6′-trihydroxydihydrochalcone led to a 100-fold increase in activity. Moreover, phoretin, with an additional 4-hydroxy group, is yet more effective, suggesting that the active site of OsCGT contains a binding pocket for a suitably located B-ring bearing a phenolic hydroxyl group. In line with such a suggestion, the xanthone precursors 2,4,6-trihydroxybenzophenone and maclurin were both viable substrates (Fig. 4). The lower conjugating activity determined with maclurin was presumably due to the substrate binding OsCGT with the A-ring competitively occupying this second binding site. Significantly, all the partially deoxygentated glycosylflavonoids reported are 5-deoxy-8-glycosyl species and this conserved substitution has a parallel with the universal observation of ortho C-glycosylation in microbial systems (1, 6).

Probing the reaction mechanism of the CGT identified a further partner activity required to complete flavone-C-glycoside synthesis, namely the dehydratase acting on 2-hydroxyflavanone-C-glycosides (Fig. 1; 5a-c), to selectively form flavone-6C-glycosides (Fig. 1; 6). Such a selective dehydrating activity was identified in the current studies in crude protein preparations from both rice and wheat. Such enzyme-catalyzed dehydrations are not without precedence, with an analogous reaction identified in the conversion of 2-hydroxyisoflavanones to the respective isoflavones in licorice (44).

The generation of C-glycosides bearing different sugars at the 6C- and 8C-positions adds a further level of sophistication to this pathway. From a mechanistic standpoint both sugars must be conjugated to the reactive hydroxyflavanone acceptor intermediate. Such a dual conjugation could involve either two distinct CGTs, or single enzymes which can alternately use UDP-glucose or UDP-arabinose (rice), or UDP-glucose and UDP-xylose (wheat). Based on the protein modelling studies it was not possible to rule out the possibility that OsCGT could accept both
UDP-glucose and UDP-arabinose, though the studies with the recombinant enzyme incubated with UDP-sugars extracted from rice seedlings were only able to confirm glucose conjugation. Whether generated by one, or two C-glycosyltransferases, the resulting doubly conjugated product must then be selectively acted on by a dehydratase to generate the respective 6C-glucoside and 8C-arabinoside (rice)/ 8C-xylloside (wheat).

The identification of OsCGT and associated reaction products sheds new light on flavonoid synthesis in plants. By forming 2-hydroxylflavanone intermediates which are acted on by specific CGTs, cereals have effectively derived an alternative pathway to produce flavones (Fig. 1B). In most higher plants, flavones are generated from unconjugated flavanone intermediates by the action of flavone synthases (FNSs). Two independent classes of FNS are known, one of which act as cytochrome P450 mixed function oxidases and the other as dioxygenases (45). Flavanone-2-hydroxylation coupled to C-glucosylation and subsequent dehydration effectively provides a further route to generating flavones. The fact that rice and wheat produce both O- and C-glycosylated flavones (30, 31, 32), suggests that cereals use multiple pathways to generate these metabolites. Interestingly, a dioxygenase FNS which converted the flavanone naringenin directly to the flavone apigenin has recently been identified in rice (46).

It will now be of interest to apply this knowledge of flavone C-glucosylation in metabolic engineering experiments to generate flavone C-glucosides in recombinant plants and microbes. In addition to providing a tool to study the control and partitioning of flavonoid metabolism the generation of medicinally useful C-glycosylated phytoceuticals both in planta and in vitro may be of biotechnological interest.

REFERENCES

1. Hultin, P.G. (2005). Curr. Top. Med. Chem. 5, 1299-1331.
2. Franz, G., and Grun, M. (1983) Planta Med. 47, 131-140.
3. Fischbach, M.A., Lin, H., Liu, D.R., and Walsh, C.T. (2005) Proc. Natl. Acad. Sci. USA 102, 571-576.
4. Faust, B., Hoffmeister, D., Weitnauer, G., Westrich, L., Haag, S., Schneider, P., Decker, H., Künzel, E., Rohr, J., and Bechthold, A. (2000) Microbiol-S.G.M. 146, 147-154.
5. Mittler, M., Bechthold, A., and Schulz, G.E. (2007). J. Mol. Biol. 372,67-76.
6. Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) Nuc. Acids Res. 37, D233-D238.
7. Bililign, T., Griffith, B.R., and Thorsan, J.S. (2005) Nat. Prod. Rep. 22, 742-760.
8. Harborne, J.B. (1993) The Flavonoids: Advances in Research Since 1986, Chapman & Hall, London.
9. Kitta, K., Hagiwara, Y., and Shibamoto, T. (1992) J. Agric. Food. Chem. 40, 1843-1845.
10. Ramarathnam, N., Osawa, T., Namiki, M., and Kawakishi, S. (1989) J. Agric. Food. Chem. 37, 316-319.
11. Kim, M., Koh, H.S., and Fukmai, H., (1985) J. Chem. Ecol. 11, 441-452.
12. Dinda, B., Bhattacharya, A., De, U. C., Arima, S., Takayanagi, H., and Harigya, Y. (2006) Chem. Pharm. Bull. 54, 679-681.
13. Akiyama, K., Matsuoka, H., and Hayashi, H. (2002) Mol. Plant-Microbe Interact. 15, 334-340.
14. Les, D., and Sheridan, D.J. (1990) Am. J. Bot. 77, 453-465.
15. Kim, Y.-C., Jun, M., Jeong, W.-S., and Chung, S.-K. (2005) J. Food Sci. 70, S575-S580.
16. Manthey, J.A., Guthrie, N., and Grohmann, K. (2001). Curr. Med. Chem. 8,135-153.
17. Gaitan, E., Cooksey, R.C., Legan, J., and Lindsay, R.H. (1995) J. Clin. Endocrinol. Metab. 80, 1144-1147.
18. Kerscher, F., and Franz, G. (1987). Z Naturforsch. (C) 42, 519-524.
19. Kerscher, F., and Franz, G. (1988) J. Plant Physiol. 132,110-115.
20. Elix, J.A., Jiang, H., and Wardlaw, J.H. (1990) *Aust. J. Chem.* **43**, 1745-1758.
21. Brazier-Hicks, M., Evans, K.M., Cunningham, O. D., Hodgson, D.R.W., Steel, P.G., and Edwards, R. (2008) *J. Biol. Chem.* **283**, 21102-21112.
22. Brazier, M., Cole, D.J., and Edwards, R. (2002) *Phytochemistry* **59**, 149-156.
23. Loutre, C., Dixon, D. P., Brazier, M., Slater, M., Cole, D. J., and Edwards, R. (2003) *Plant J.* **34**, 485-493.
24. Perkins, D.N., Pappin, D.J.C., Creasy, D.M., and Cottrell, J.S. (1999) *Electrophoresis* **20**, 3551-3567.
25. Sali, A., and Blundell, T.L. (1993) *J. Mol. Biol.* **234**, 779-815.
26. Brazier-Hicks, M., Offen, W.A., Gershater, M.:C., Revett, T.J., Lim, E.-K., Bowles, D.J., Davies, G.J., and Edwards, R. (2007) *Proc. Natl. Acad. Sci. USA* **104**, 20238-20243.
27. Melo, F., Sánchez, R., and Sali, A. (2002) *Protein Sci.* **11**, 430-448.
28. Cone, K. (1989) *Maize Gen. Cooper. Newslett.* **63**, 68-69.
29. Dixon, D.P., Hawkins, T., Hussey, P.J., and Edwards, R. (2009) *J. Exp. Bot.* **60**, 1207 - 1218.
30. Besson, E., Dellamonica, G., Chopin, J., R. Markham, K.R., Kim, M., Koh, H.-S., and Fukami, H. (1985) *Phytochemistry* **24**, 1061-1064.
31. Markham, K. R., Tanner, G. J., Caasi-Lit, M., Whitecross, M. L., Nayudu, M., and Mitchell, K. A. (1998) *Phytochemistry* **49**, 1913-1919.
32. Cummins, I., Brazier-Hicks, M., Stobiecki, M., Franski, R., and Edwards, R. (2006) *Phytochemistry* **67**, 1722-1730.
33. Mabry, T.J., Markham, K.R., and Thomas, E.W. (1970) *The Systematic Identification of Flavonoids*, Springer-Verlag, Berlin.
34. Kazuno, S., Yanagida, M., Shindo, N., and Murayama, K. (2005) *Anal. Biochem.* **347**, 182-192.
35. Amino, S.i., Takeuchi, Y., and Komamine, A. (1985) *Physiol. Plant.* **64**, 197-201.
36. Rauter, A.P., Martins, A., Borges, C., Ferreira, J., Jorge, J., Bronze, M.R., Coelho, A.V., Choi, Y.H., and Verpoorte, R. (2005) *J. Chrom. A* **1089**, 59-64.
37. Shen, C.-C., Chang, Y.-S., and Ho, L.-K. (1993) *Phytochemistry*, **34**, 3, 843-845.
38. Ross, J., Li, Y., Lim, E., and Bowles, D.J. (2001) *Genome Biol.* **2**, Reviews 3004.
39. Lim, E.K., Jackson, R.G., and Bowles, D.J. (2005) *FEBS Lett.* **579**, 2802-2806.
40. Osmanli, S.A., Bak, S., and Moller, B.L. (2009) *Phytochemistry doi:10.1016*.
41. Dürr, C., Hoffmeister, D., Wohlert, S.-E., Ichinose, K., Weber, M., von Mulert, U., Thorson, J. S., and Bechtold, A. (2004) *Angew. Chem. Int. Ed.* **43**, 2962-2965.
42. Akashi, T., Aoki, T., and Ayabe, S. (1998) *FEBS Lett.* **431**, 287-290.
43. Ralston, L., Subramanian, S., Matsuno, M., and Yu, O. (2005) *Plant Physiol.* **137**, 1375-1388.
44. Akashi, T., Aoki, T., and Ayabe, S. (2005) *Plant Physiol.* **137**, 802-891.
45. Martens, S., and Mithöfer, A. (2005) *66*, 2399-2407.
46. Kim, J.H., Cheon, Y.M., Kim, B.-G., and Ahn, J.-H. (2008) *J. Plant Biol.* **51**, 97-101.

**FOOTNOTES**

*This work was supported by the Biotechnology and Biological Sciences Research Council (BBSRC; grant BB/F01094X/1). RE is a BBSRC research development fellow. The authors thank Dr Ehmke Pohl for homology modelling the structure of OsCGT.*

**Abbreviations:** CGT; C-Glucosyltransferase, OGT; O-Glucosyltransferase, UDP; uridine-diphospho, UGT; type 1 glycosyltransferase. Data deposition = CCDC No: 670429 entry for crystal structure of flavone-8C-glucoside; CGT from rice EMBL accession FM17912.
FIGURE LEGENDS

Fig. 1. Glycosylflavone biosynthesis in plants. (A) Flavone synthase (FNS) converts flavanones (1) to flavones (2), which are then conjugated by an O-glucosyltransferases (OGT). (B) In cereals the flavanone also undergoes conversion to the 2-hydroxyflavanone (4a), which exists in equilibrium with its open-chain form (4b), the latter apparently being acted on by the C-glucosyltransferase (CGT) to produce 2-hydroxyflavanone C-glucosides (5a-c). These are then dehydrated to yield the flavone-6C-glucoside (6) and flavone-8C-glucoside (7). (C) The hydroxylation and (D) C-glycosylation of the flavonoids referred to in the text.

Fig. 2. Flavonoids in 7 day old rice foliage. (A) HPLC profile of UV absorbing metabolites extracted with methanol from rice shoots. (B) Identification of the flavone structure and C-glycosylation pattern of key rice metabolites. ara; arabinose, glc; glucose.

Fig. 3. Sequential purification of CGT activity from rice. CGT activity (■) was monitored along with elution of UV absorbing proteins (—) on (A) phenyl Sepharose, (B) mono Q and (C) Superdex 200 columns. (D) Analysis of polypeptides corresponding to the active fraction from the gel filtration chromatography (C) as determined by SDS-PAGE. Proteins arrowed were subjected to MALDI-TOF-MS sequencing (Fig. S2).

Fig. 4 CGT activities of pure recombinant OsCGT toward flavonoids and related compounds. (A) Respective specific activities were determined (B) and where possible kinetic data determined. (n.a.= not available). Bn; benzyl.

Fig. 5 Analysis of reaction products derived from the C-glucosylation of 2,5,7-trihydroxyflavanone. Determination by HPLC (A), with the initial product (5) spontaneously dehydrating to give peaks 6 and 7. The fragmentation of 5 (B) and mass ions associated with compounds 6 and 7 (C) are shown. Peak 5 (D) was then incubated with either an enzyme preparation from rice cultures (−) or boiled protein (−).
Figure 1

A. Chalcone

1. FNS

2. OGT

3. Glc

B. Flavone (2 R' R"

4a. CGT

4b. Flavone (2 R' R"

5a. Chrysin H H

5b. Apigenin OH H

5c. Luteolin OH OH

6. Chalcone

7. Flavone (2 R' R"

C. Flavone (2 R' R"

D. 8C
**Figure 2**

| Sugar linkage | Flavone structure | C-glycosylation pattern | Retention time (min) |
|---------------|-------------------|-------------------------|----------------------|
| C-gly        | Luteolin          | gluc                    | 13.3                 |
|              | Luteolin          | ara                     | 13.4                 |
|              | Luteolin          | (2"-O-glc)-glc         | 13.8                 |
|              | Apigenin          | gluc                    | 14.4                 |
|              | Apigenin          | ara                     | 14.5                 |
|              | Apigenin          | gluc                    | 15.1                 |
|              | Apigenin          | ara                     | 15.6                 |
|              | Apigenin          | (2"-O-(6-O-feruloyl)-glc)-glc | 17.0    |
|              | Apigenin          | (2"-O-(6-O-coumaryl)-glc)-glc | 18.6    |
| O-gly        | Tricin-O-glc     | H                       | 16.8                 |
|              | Tricin-O-glc     | H                       | 17.9                 |
|              | Tricin-O-glc     | H                       | 19.3                 |
Figure 3
Figure 4

A

![Chemical structures](attachment:chemical_structures.png)

B

|   | R¹ | R² | R³ | R⁴ | Specific activity (pkat mg⁻¹) | Kinetic data |
|---|----|----|----|----|-------------------------------|--------------|
| i | a  | H  | H  | H  | 2-Hydroxyflavanone           | 0 ± 0        |
|   | b  | H  | OH | H  | 2.5-Dihydroxyflavanone       | 0 ± 0        |
|   | c  | OH | H  | H  | 2.7-Dihydroxyflavanone       | 3 ± 1        |
| d | OH | OH | H  | H  | 2,5,7-Trihydroxyflavanone    | 382 ± 27     |
| e | OH | OH | OH | H  | 2-Hydroxyflavanone           | 2,5-Hydroxyflavanone     |
| f | OH | OH | OH | OH | 2-Hydroxyeriodictyol         | 466 ± 3      |
| ii| a  | H  | H  | -  | Chrysirin                   | 0 ± 0        |
|   | b  | OH | H  | -  | Apigenin                    | 0 ± 0        |
|   | c  | OH | OH | -  | Luteolin                    | 0 ± 0        |
| iii| -  | -  | -  | -  | Naringenin                  | 2 ± 0        |
| iv | -  | -  | -  | -  | Naringenin chalcone         | 3 ± 0        |
| v | a  | H  | -  | -  | 2',4',6'-Trihydroxyflavanone| 350 ± 45     |
|   | b  | OH | -  | -  | Phoretin                    | 587 ± 19     |
| vi| a  | H  | H  | -  | 2,4,6-Trithydroxybenzophenone| 48 ± 2       |
|   | b  | OH | OH | -  | Maclurin                    | 1 ± 0        |
| v  | a  | H  | -  | -  | 1,3,5-Trihydroxybenzoic acid| 0 ± 0        |
| b  | CO₂H| -  | -  | -  | 2,4,6-Trithydroxybenzoic acid| 0 ± 0        |
| c  | CO₂H₂| -  | -  | -  | 2,4,6-Trithydroxybenzoic acid| 0 ± 0        |
| d  | CO₂Bn| -  | -  | -  | 2,4,6-Trithydroxybenzoic acid| 0 ± 0        |
Figure 5

(A) Retention time vs. mAU (287 nm) for compound 5, 6, and 7. The peaks are well separated, indicating the resolution of the separation.

(B) Mass spectrum (m/z) showing the molecular ion at m/z 343 and fragment ions at m/z 167, 313, and 433. The [M-H]^- ion is also present at m/z 287.

(C) Percentage of compounds 5, 6, and 7 at different retention times. Compound 7 shows a higher percentage than compounds 5 and 6.

(D) Retention time vs. mAU (287 nm) for compound 5, 6, and 7. The peaks are consistent with those in (A) and suggest good reproducibility of the separation.

Chemical structures of compounds A, B, C, and D are shown, with functional groups highlighted in red and blue. M/z values are indicated for each peak.
The C-glycosylation of flavonoids in cereals
Melissa Brazier-Hicks, Kathryn M. Evans, Markus C. Gershater, Horst Puschmann, Patrick G. Steel and Robert Edwards

J. Biol. Chem. published online May 1, 2009

Access the most updated version of this article at doi: 10.1074/jbc.M109.009258

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2009/05/01/M109.009258.DC1