Organic & Biomolecular Chemistry

PAPER

Cite this: *Org. Biomol. Chem.*, 2016, **14**, 6252

UGT74B1 from *Arabidopsis thaliana* as a versatile biocatalyst for the synthesis of desulfoglycosinolates†

Sami Marroun,‡[a] Sabine Montaut,‡[b] Stéphanie Marquès,‡[c] Pierre Lafite,[c] Gaël Coadou,[a] Patrick Rollin,[c] Guillaume Jousset,[c] Marie Schuler,[c] Arnaud Tatibouët,*[c] Hassan Oulyadi*[a] and Richard Daniellou*[c]

Received 8th May 2016,
Accepted 24th May 2016

DOI: 10.1039/c6ob01003b

www.rsc.org/obc

Thioglycosides, even if rare in Nature, have gained increased interest for their biological properties. Chemical syntheses of this class of compounds have been largely studied but little has been reported on their biosynthesis. Herein, combining experiments from the different fields of enzymology, bioorganic chemistry and molecular modeling, we wish to demonstrate the versatility of the glucosyltransferase UGT74B1 and its synthetic potency for the preparation of a variety of natural and unnatural desulfoglycosinolates.

Introduction

Glycoconjugates are involved in a wide range of biological processes in Nature.[1,2] Among this large and diverse family of compounds, *S*-glycosides that bear a sulphur atom at the anomeric position are of particular interest.[3] These compounds are more stable towards enzymatic or chemical hydrolysis and exhibit the same conformation as their widespread *O*-glycoside homologues.[4] So far, only a few examples of natural *S*-glycosides have been identified in organisms, thus being considered as rare glycosides:[2] some bacteriocin glycopeptides such as sublancin,[5] glycocin F,[6] or thurandacin[7] possess a glycosylated cysteine thiol group, but the most diversified and studied family of *S*-glycosides is glucosinolates.[8]

These plant secondary metabolites found in the order Brassicales participate in the plant's natural defense. Their enzymatic hydrolysis leads to a range of compounds that are toxic to bacteria or insects.[9] Interest in these compounds, well known for decades, has grown recently thanks to their chemopreventive action towards cancer, and cardiovascular or neurological diseases.[10]

In plants, glucosinolates are biosynthesized from amino acids (Scheme 1).[8] One of the critical steps of this biosynthesis is the formation of the thioglycosidic bond, catalyzed by *S*-glucosyltransferase (*S*-GT) from UDP-α-D-glucopyranoside (UDP-Glc, sugar donor) and thiohydroxamic acid (acceptor). Apart from *S*-GT involved in glucosinolate biosynthesis, a few glycosyltransferases have been reported to be able to catalyse the formation of the *S*-glycosidic linkage, either using a wild-type enzyme[11,12] or its bioengineered variant.[13] Several genes coding for *S*-GT have been identified in *Arabidopsis thaliana*,[14,15] *Brassica napus*,[16] or *Brassica rapa*.[17] Among *A. thaliana* *S*-GT, UGT74B1 has been more extensively studied and characterized.[14,18] This enzyme belongs to the superfamily of UDP-dependent glycosyltransferases (UGT) that regroup in the *A. thaliana* genome more than 100 enzymes that are usually responsible for the last steps of natural product biosynthesis.[19]

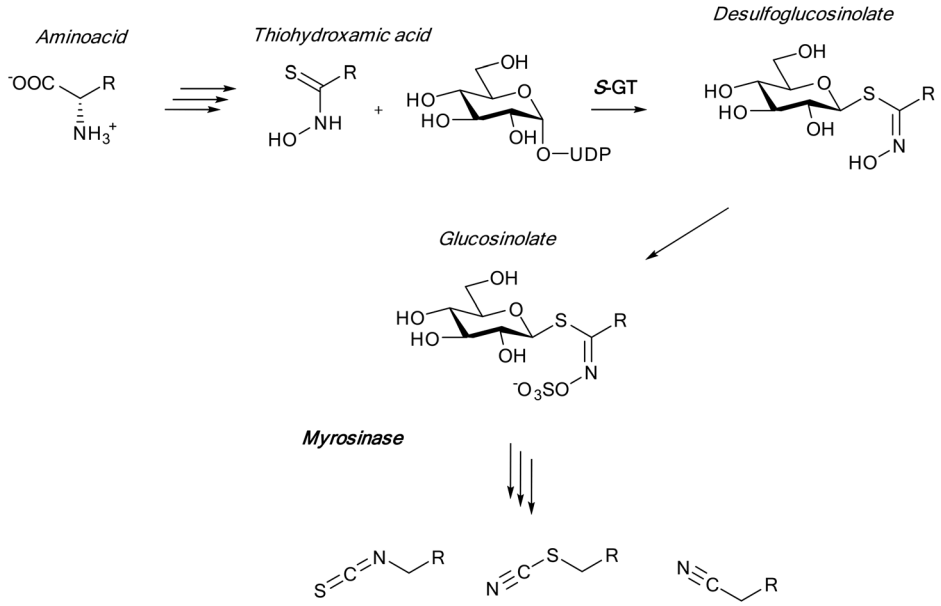
In the search for assessing the influence of glucosinolate structures on their biological roles, chemical strategies to synthesize natural glucosinolates and their structural analogues have been described earlier in the literature.[20–26] Some of these glucosinolates were prepared for characterization purposes, to give access to specific standards, for studying their biological properties or to explore myrosinase, the hydrolase involved in glucosinolate breakdown.[27–32] In all the reported studies, the desulfoglucosinolate precursor was synthesized as an intermediate, through multistep synthesis (from 2 to 5 steps) using protecting groups to prevent side reactions from the reactive sugar hydroxyls. It is now well admitted that biocatalytic systems (*e.g.* enzymes) can efficiently catalyse specific reactions without the need for protective groups. Working in aqueous buffers and at moderate temperatures, they can greatly improve the ecological footprint of chemical reactions.

From a synthetic point of view, previous studies have reported the use of 2 enzymes to generate thiohydroxymates.[33] Yet, the preparative *S*-glycosylation of these substrates remains undone using biocatalysts. In this study, we have focused on

[a]*Normandie Univ, COBRA, UMR 6014 et FR 3038; Univ Rouen; INSA Rouen; CNRS, IRCOF, 1 rue Tesnière, 76821 Mont Saint Aignan Cedex, France*
[b]*Department of Chemistry and Biochemistry, Biomolecular Sciences Programme, Laurentian University, 935 Ramsey Lake Road, Sudbury, ON P3E 2C6, Canada*
[c]*Université d'Orléans, CNRS, ICOA, UMR 7311, Orléans, France.*
E-mail: richard.daniellou@univ-orleans.fr

†Electronic supplementary information (ESI) available. See DOI: 10.1039/c6ob01003b

‡These authors contributed equally to this work.



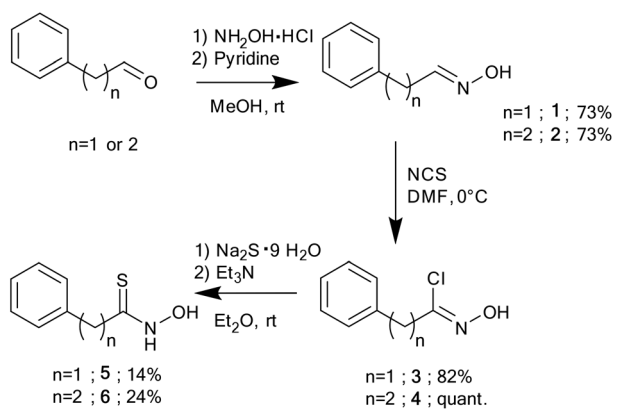
Scheme 1 Synthesis and degradation of glucosinolates in plants.

A. thaliana UGT74B1 in order to assess its potency in the chemoenzymatic synthesis of desulfoglucosinolate analogues, by varying the sugar nature of the *S*-glycoside.

Results and discussion

Synthesis of thiohydroxamic acid acceptors

Phenylacetothiohydroxamic acid **5** and 3-phenylpropanothiohydroxamic acid **6** were respectively prepared from phenyl-acetaldehyde and 3-phenylpropanal following the reported but tedious procedures (Scheme 2).[34] Oximes **1** and **2** were obtained from the corresponding aldehydes after treatment with hydroxylamine hydrochloride (73% yield). The oximes were then chlorinated by *N*-chlorosuccinimide (NCS) in DMF to give the hydroximoyl chlorides **3** and **4** (82% and quantitative yields, respectively). Addition of sodium sulphide nonahydrate in the presence of triethylamine led to the formation of thiohydroxamic acids **5** and **6** in respectively 14% and 24% yields. It is noteworthy that these rather unstable compounds need to be handled carefully and stored at −20 °C avoiding light and air.



Scheme 2 Synthesis of UGT74B1 thiohydroxamic acid substrates.

Cloning and enzymatic characterization of UGT74B1

The *ugt74b1* gene coding for the UGT74B1 protein was amplified by PCR from *A. thaliana* cDNA, and subsequently cloned in the pET28a(+) bacterial expression vector. This vector, as well as the cloning strategy chosen, enabled us to fuse the *ugt74b1* gene with an N-terminal polyhistidine tag. After plasmid transformation in an expression strain, induction of the culture, and subsequent purification of the Ni-NTA affinity column, the recombinant UGT74B1 protein was obtained pure and in good yield (10 mg l^{-1} culture) (Fig. 1-A).

Consequently, the two thiohydroxamic acids **5** and **6** were tested as sugar acceptors in the UGT74B1 *S*-glycosylation reaction, and catalytic constants K_M and k_{cat} were then determined at 37 °C using 1 mM UDP-Glc (Table 1 and Fig. 1-B). *In vivo*, these two substrates are the respective precursors of natural glucosinolates, glucotropaeolin and gluconasturtiin.[9] The enzymatic activity was determined using an adaptation of a previously reported tri-enzymatic assay for glucosyltransferases.[35,36] This assay enables the enzymatic coupling of UDP production by UGT74B1 to the consumption of NADH (whose absorbance is monitored at 340 nm).

Thiohydroxamic acid **5** had been previously identified as a substrate for UGT74B1,[14,18] and kinetic constants calculated for this compound (K_M = 3.1 μM and k_{cat} = 280 min^{-1}) were close to those reported earlier (K_M of 6 μM and 14.5 μM and k_{cat} of 60 and 240 min^{-1}). Like **5**, thiohydroxamic acid **6** is a substrate of UGT74B1 exhibiting a lower Michaelis constant

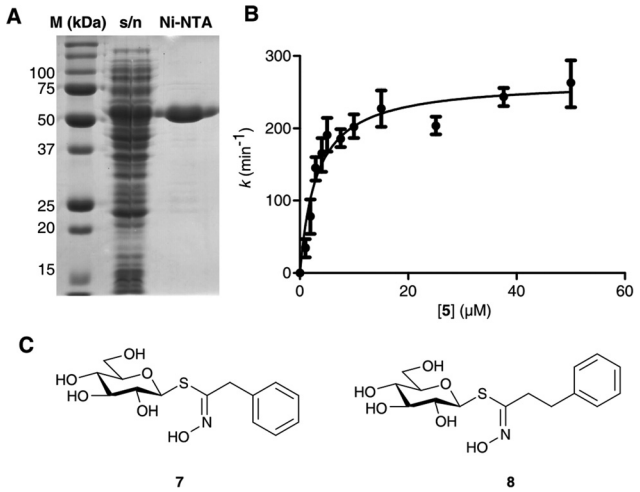


Fig. 1 (A) SDS-PAGE purity analysis of UGT74B1 in a culture supernatant (s/n), and after affinity chromatography (Ni-NTA); (B) Michaelis–Menten plot of the glycosylation of substrate **5** by UGT74B1. Error bars were calculated from 3 independent experiments. (C) Structures of desulfoglucosinolates **7** and **8** obtained from **5** and **6** as glucosyl acceptors.

Table 1 Kinetic constants for the glycosylation of thiohydroxamic acid acceptors **5** and **6** by UGT74B1

Substrate	K_M [a] (μM)	k_{cat} [a] (min^{-1})	k_{cat}/K_M (min^{-1} μM^{-1})	Ref.
5	3.1 ± 0.6	280 ± 58	104	This work
5[b]	5.8 ± 3.1	60	172	14
5[c]	14.5 ± 5.9	240 ± 90	16.6	18
6	2.2 ± 0.5	326 ± 95	148	This work

[a] Kinetic constants were determined as reported in Experimental procedures using a constant 1 mM concentration of UDP-α-D-glucose as a sugar donor. Means and standard deviations were calculated from 3 independent experiments. [b] Kinetics parameters were calculated in the presence of 0.5 mM UDP-α-D-glucose. [c] Kinetics parameters were calculated in the presence of 1 mM UDP-α-D-glucose.

(K_M = 2.2 μM) and a higher catalytic rate than **5** (k_{cat} = 280 min^{-1}). Overall, the specificity constant k_{cat}/K_M is higher for **6** compared to **5**. Indeed, it was previously shown by the mRNA expression profile that UGT74B1 is only able to *S*-glycosylate aromatic containing thiohydroxamates.[15] Therefore and as expected, modification of the length of the side chain of the thiohydroxamate does not change the affinity of the substrate for UGT74B1.

UGT74B1 is versatile for other UDP-sugar donors

To understand the structural requirements for sugar donor recognition by UGT74B1, several nucleotide sugars were screened as potential sugar donors: UDP-α-D-galactose (UDP-Gal) and UDP-α-D-*N*-acetylglucosamine (UDP-GlcNAc), UDP-α-D-glucuronic acid (UDP-GlcA) and GDP-α-D-mannose (GDP-Man).

When compared to UDP-Glc as the reference substrate, UDP-Gal exhibited 25% of the relative activity. UDP-GlcNAc exhibited a lower yet non-negligible activity (9%) whereas UDP-GlcA and GDP-mannose were not recognized as efficient substrates (<5%).

For all substrates, kinetic parameters of the enzymatic reaction were determined (Table 2). As expected from the preliminary results, even on increasing the donor concentration to 10 mM, no activity usable for such analysis could be determined for UDP-GlcA and GDP-Man.

Like other *S*-GT reported in the literature,[5,7] UGT74B1 is versatile towards several UDP-sugars, yet the natural substrate UDP-Glc shows the highest specificity constant (8.3 min^{-1} μM^{-1}). However, because of their lower yet non negligible specificity constants, the two alternative substrates UDP-Gal and UDP-GlcNAc are good candidates to be used for the

Table 2 Kinetics constants of NDP-sugar donors

NDP-sugar	K_M [a] (μM)	k_{cat} [a] (min^{-1})	k_{cat}/K_M (min^{-1} μM^{-1})
UDP-Glc	38 ± 2	233 ± 13	8.3
UDP-Gal	1427 ± 250	150 ± 17	0.11
UDP-GlcNAc	45 ± 18	11 ± 1	0.24
UDP-GlcA	n.d.[b]	n.d.	n.d.
GDP-Man	n.d.	n.d.	n.d.

[a] Kinetic constants were determined as reported in Experimental procedures using a constant 1 mM concentration of **5** as the acceptor. Means and standard deviations were calculated from 3 independent experiments. [b] n.d.: not determined.

chemoenzymatic synthesis of desulfoglucosinolate analogues. To better understand the influence of the sugar structure on substrate recognition, the molecular modeling of sugar analogues bound to the UGT74B1 active site was performed.

Modeling of the UGT74B1/donor complex reveals specific interactions with the sugar moiety

According to the CAZY database,[37] § UGT74B1 belongs to the GT1 family that contains all *A. thaliana* UGT proteins. Although more than 9000 genes have been identified to code for GT1 enzymes, only 58 structures (28 distinct proteins) have been deposited in the Protein Data Bank (PDB), which illustrates the low proportion of structural representatives of a given family for glycosyltransferases, when compared to other carbohydrate acting enzymes.

The lack of an experimental structure of UGT74B1 drove us to predict its three-dimensional structure using comparative homology modeling techniques. All templates identified for this protein belonged to the glycosyltransferase family GT1, but presented low to average sequence identity (21% to 30% from a BLAST search). Such low values meant that we entered the "twilight zone" of sequence alignment.[38] Nevertheless, studying the available templates revealed that they exhibited surprisingly high structural conservation (Tables S1 and S2 in the ESI†). This structural conservation within a multiple sequence alignment yielded a high accuracy and biologically significant sequence alignment (Fig. S1†). Kopycki *et al.* previously attempted to model UGT74B1 in 2013. However, their model was based on a single template using a purely automated homology modeling protocol not suitable for delicate cases in the twilight zone, such as this one.

Based on the homology model, UGT74B1 is a GT-B enzyme presenting two Rossmann fold domains, a 7-parallel strand C-terminal domain and a 6-parallel strand N-terminal domain separated by a median hinge region and wrapped by a C-terminal helix-loop-helix structure. The active site of the enzyme is situated in the cleft between these two domains. The active site can be visualized as a long channel in-between the two Rossmann folds with two regions, one for each ligand. The sugar donor occupies the larger part of the active site whereas the sugar acceptor occupies a relatively small pocket. This structure corresponds to the GT-B fold of glycosyltransferases in the GT1 family.[39] It is also noteworthy that based on our dual structure-sequence-based alignment, the sequence identity within the active site is relatively high (45.83% to 60.82%), further validating our model.

To understand the possible differences that could explain the drastic change in affinity (K_M) and reaction rates (k_{cat}) of the experimentally examined sugar donors, we docked UDP-Gal, UDP-GlcNac, UDP-GlcA and GDP-Man into the UGT74B1 homology model and compared their predicted binding modes to UDP-Glc using the Autodock Vina docking program.

§ http://www.cazy.org

Docking of the natural sugar donor UDP-Glc was compared to the crystallographic position of UDP-2-deoxy-2-fluoro-α-D-Glc from PDB structures 2C1Z and 2VCE. Binding was predicted correctly (RMSD = 0.62 Å over 35 atom pairs). In the active site, the uracil moiety of UDP-Glc is stabilized by an essential π-stacking interaction with the highly conserved tryptophan Trp336 as well as several hydrogen bonds with Cys337 involving uracil nitrogen and carbonyls. Both hydroxyls of the ribose moiety are stabilized by electrostatic interactions with the glutamic acid at position 362 whereas the diphosphate group is stabilized by several electrostatic interactions with several polar residues including Asn358, Ser359, His354 and Ser284. The glucosidic moiety establishes mainly electrostatic contacts with Ser377, Asn136, Asp378 and Trp357 (Fig. 2-A).

The binding modes for other sugar donors were verified by examining the correct docking of the nucleobase (when comparable), ribose and diphosphate moieties compared to the crystallographic ligands. When we compared the predicted binding modes of different sugar donors to UGT74B1, we confirmed that the slightest structural modification to the sugar donor had dire biological effects on both the affinity and activity.

First, concerning modifications to the nucleobase, it is well accepted that glycosyltransferases possess a binding pocket for the nucleotide moiety of the sugar donor that exhibits strong interactions.[40] This led us to postulate that using a purine (GDP) instead of a pyrimidine (UDP) must alter the binding efficiency of the sugar donor to the protein, leading to a loss of enzymatic activity. Inspecting the binding modes for GDP-glucose revealed that although the purine can bind within the nucleotide pocket conserving the essential π-stacking interaction with Trp336, its binding leads to a complete shift of the sugar donor within the active site and an occupation of the acceptor site by the sugar moiety (data not shown). This displacement renders any binding of the acceptor impossible and thus any catalytic activity unfeasible.

Concerning modifications to the sugar moiety, the addition of a carboxylic acid moiety at position C-6 in the case of UDP-GlcA, completely modifies the binding of the sugar donor. In fact, the carboxylic acid establishes an electrostatic bond with the basic and positively charged residue Arg182 (2.6 Å distance between the carboxylic acid oxygen and the guanidinium proton). This strong stabilizing interaction leads to a shift of the sugar within the active site and a displacement of the anomeric carbon. The C-6 position of the sugar moiety thus appears to be critical for activity. Indeed, the presence of electrostatic interactions leads to a complete decrease in the enzymatic activity (Table 2).

Concerning the modification of the sugar moiety at position C-4, compared to UDP-Glc, UDP-Gal presents a sugar moiety that was heavily displaced as compared to glucose. First, this displacement decreases the favorable contacts. In fact, the galactose moiety loses polar interactions with Glu362 and develops unfavorable contacts with Phe285, Trp376 and Trp357 because of the orientation of its sugar hydroxyls resulting in the biological observation of a heavily lowered affinity

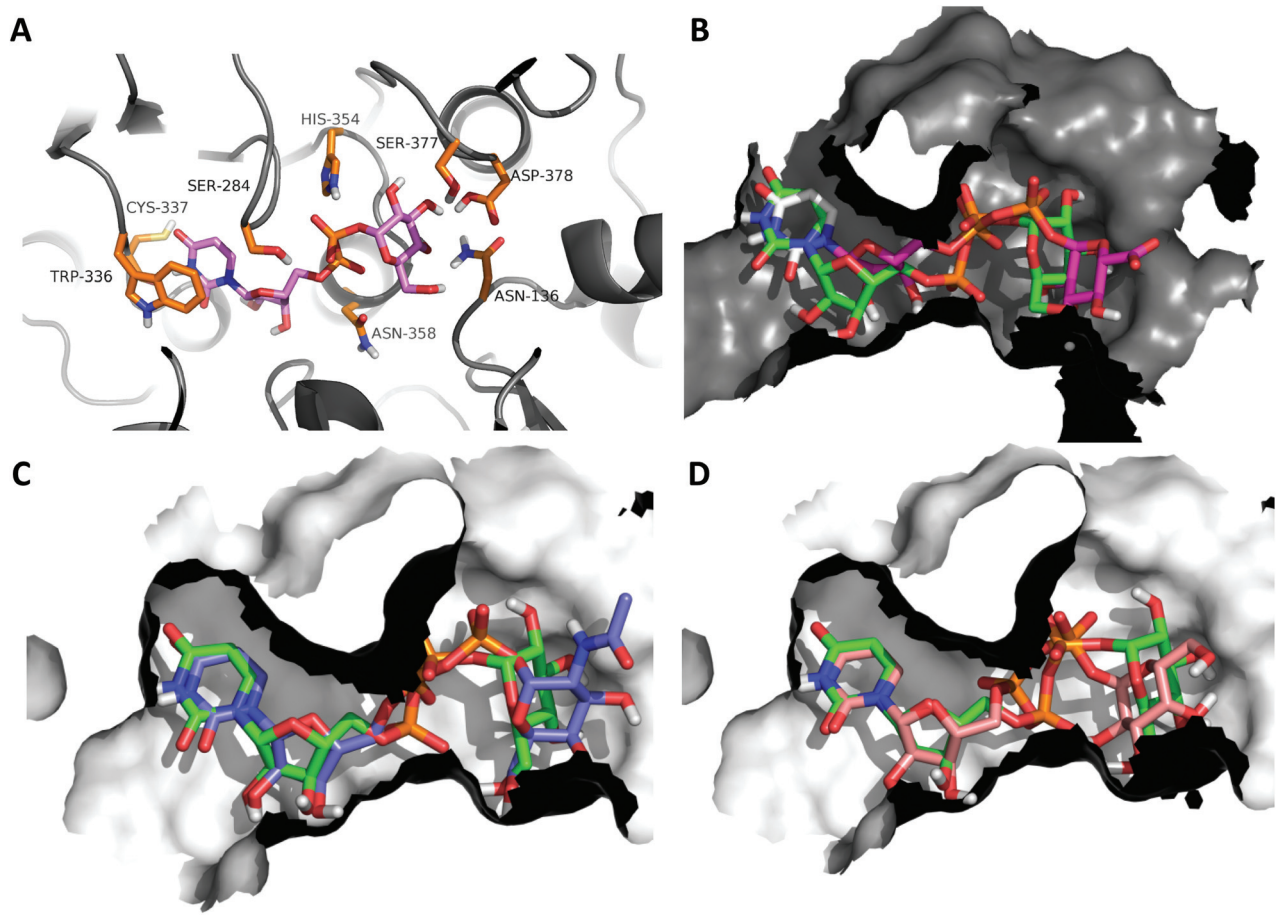


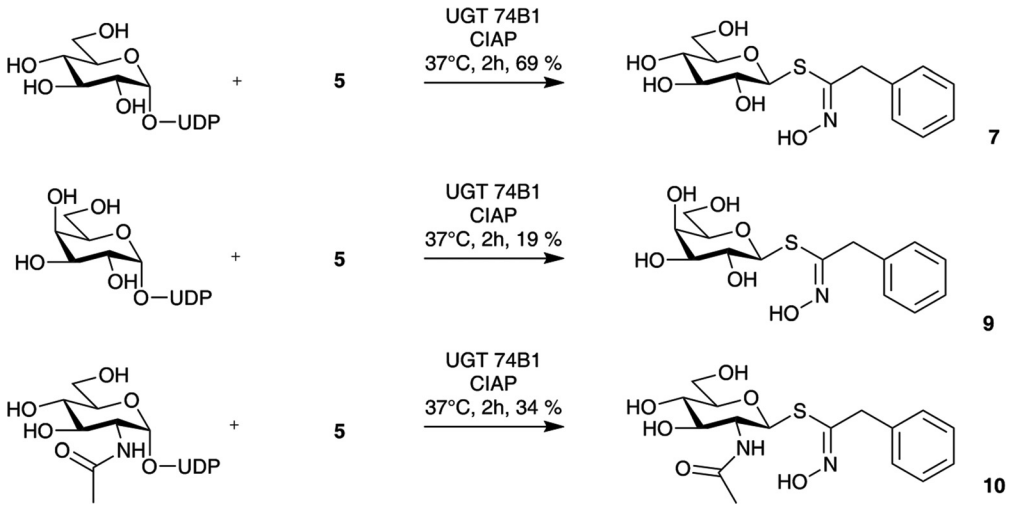
Fig. 2 (A) Major contacts of the sugar donor within the active site. (B) Relative position of UDP-GlcA (magenta) to the position of UDP-Glc (green) within the 74B1's active site. Binding modes of UDP-Glc superimposed to UDP-GlcNAc (C) and UDP-Gal (D) showing changes in steric hindrances as well as a small anomeric carbon displacement (1.4 Å) for UDP-GlcNAc and large displacement for UDP-Gal (2.5 Å) explaining the decrease of the binding strength and decrease in the catalytic activity respectively.

(the K_M value increases by a factor of 38 when compared to UDP-Glc, see Table 2). Second, this displacement results in a modified anomeric carbon position by 1 Å. This modification in the position of the anomeric carbon would have a moderate effect on the reaction rate, which correlates with a slightly lowered reaction rate (150 min^{-1} *vs.* 233 min^{-1} for UDP-Glc). This reveals the importance of the C-4 position of the sugar moiety which seems to be of importance for substrate binding in the UGT74B1 active site. Furthermore, our results suggest that this C-4 epimeric modification between Gal and Glc impacts only the binding of the donor into the active site and not the enzymatic mechanism.

Concerning the modification of the sugar moiety at position C-2 in the case of UDP-GlcNAc, the steric effect is less severe as is the displacement of the carbon giving this sugar donor a comparable affinity to UDP-Glc. When comparing Glc and GlcNAc, the C-2 position appears to be critical for the catalytic rate. Indeed, a 20-fold decrease in the catalytic rate k_{cat} is observed when using GlcNAc as the sugar. As an inverting GT glycosyltransferase, the UGT74B1 molecular mechanism is a SN_2-like reaction that involves the formation of an oxocarbenium ion-like transition state on the C-1 anomeric carbon.[39] Due to its proximity with the C-1 atom, modification of the functionality at the C-2 position modulates the physicochemical properties of the C-1 position. On the other hand, the K_M between Glc and GlcNAc remains similar. This indicates that modification of the C-2 hydroxyl group into an *N*-acetyl group does not change the molecular recognition of the substrate at the UGT74B1 active site. Finally, UDP-GlcA presents the most drastic difference compared to UDP-Glc. The sugar ring is completely flipped due to added steric hindrance. This steric hindrance leads to a complete loss of the enzymatic activity of UGT74B1 with this sugar donor.

Chemoenzymatic synthesis of desulfoglycosinolates

Still, several UDP-sugars are efficiently recognized by UGT74B1. To assess the potentiality of the enzyme to be a useful biocatalyst, 3 desulfoglycosinolates were synthesized, using **5** as the common thiohydroxamic acid acceptor. As expected from the kinetic constants' determination, UDP-Glc gives the highest yield (69%) after 2 h of reaction at 37 °C. The corresponding desulfoglucosinolate **7** was obtained after one



Scheme 3 Chemo-enzymatic synthesis of desulfo-benzyl glucosinolate **7** and non-natural desulfo-benzyl glycosinolates **9** and **10** by UGT74B1.

single reversed-phase chromatography on a C18 solid-phase extraction cartridge. This convenient purification system enables the fast recovery of the product without the use of additional protection/deprotection steps that usually decrease the overall yield of the desulfoglucosinolate.[26,28]

Two other UDP-sugars, that were identified as efficient substrates (UDP-Gal and UDP-GlcNAc) were also tested for the synthesis of desulfoglycosinolates from phenylacetothiohydroxamic acid **5**. Both UDP-sugars could lead to the corresponding desulfogalactosinolate **9** and desulfo-*N*-acetylglucosaminolate **10**. The yield for both reactions was lower than with UDP-Glc (19% and 34%, respectively), yet these one-step reactions followed by simple reverse phase purification enabled easy access to desulfoglycosinolate derivatives.

The structures of **7** and **9–10** were elucidated using NMR and mass spectrometry. The mass analysis of **7** gave a mass of 330 amu and the molecular formula $C_{14}H_{20}NO_6S$ $[M + H]^+$ was established by HRMS. The NMR spectra revealed the presence of five aromatic protons, a pair of protons from an isolated methylene group (δH 4.07 and δC 38.6), and signals attributable to a thio-glucose unit (see the Experimental section). The coupling constant of 9.6 Hz determined from the doublet of the anomeric proton, H-1, at 4.73 ppm, was indicative of a β-linked glucopyranosyl moiety. The HMBC spectrum showed couplings of CH_2Ph with aromatic carbons and the quaternary carbon of the thiohydroximate function (C-7), and of H-1 with C-7. Compound **7** was identified as desulfo-benzyl glucosinolate (Scheme 3), by comparison of spectroscopic data with published data for desulfoglucotropaeolin.[41] The mass spectrum of **9** also gave a mass of 330 amu and a molecular formula $C_{14}H_{20}NO_6S$ $[M + H]^+$ by HRMS. The NMR spectra revealed the presence of five aromatic protons, a pair of non-equivalent protons from an isolated methylene group (δH 4.11, 4.06 and δC 37.9), and signals attributable to a thio-galactose unit.[27] The coupling constant of 9.4 Hz determined from the doublet of the anomeric proton, H-1, at 4.68 ppm, was indicative of a β-linked galactopyranosyl moiety. Compound **9** was identified as desulfo-benzyl galactosinolate (Scheme 3). The mass spectrum of **10** gave a mass of 393 amu and the molecular formula $C_{16}H_{22}N_2O_6SNa$ $[M + Na]^+$ by HRMS. The NMR spectra revealed the presence of five aromatic protons, a pair of protons from an isolated methylene group (δH 4.04 and δC 37.7), and signals attributable to a thio-*N*-acetylglucosamine unit.[27] Compound **10** was identified as desulfo-benzyl *N*-acetyl-glucosaminolate (Scheme 3). We report for the first time the structures of **9** and **10**.

To our knowledge, this is the first report of the use of a natural *S*-glycosyltransferase as a biocatalytic tool to synthesize *S*-glycosides with a structural diversity on the sugar moiety. When compared with the traditional chemical synthesis of glucosinolate analogues,[21] this methodology proved to give similar yields in fewer steps, using an eco-compatible catalysis system.

Conclusions

Thioglycosides have become valuable tools as stable compounds towards chemical and enzymatic hydrolyses and have been used in a vast array of biochemical applications, from stable intermediate ligands in the X-ray crystallographic analysis of proteins[42,43] to competitive inhibitors of a wide range of glycoside hydrolases involved in diseases.[44] Chemical syntheses of this class of compounds have been studied since decades, however, the use of biocatalysts as enzymes has recently emerged as one of the most eco-friendly and efficient methodologies.[45] Enzymatic *S*-glycosylation was previously demonstrated using bioengineered glycoside hydrolases – through the thioglycoligase or thioglycosynthase methodology[46–55] – or glycosyltransferases.[12,13] In the latter case, GTs either exhibited a broad substrate versatility for acceptor function (alcohol or thiol) or were bioengineered from *O*-GT to increase

their aglycon promiscuity. The ability of UGT74B1 to recognize and utilize a diversity of nucleotide-sugar donors makes this enzyme a promising tool for the glycorandomization of *S*-glycosides.[13,40] Thus, these results provide an alternative method, which is complementary to classical synthetic approaches available to date. Although the chemical structures of the thiol acceptor and of the obtained *S*-glycosides are identical to the natural compounds, the versatility of UGT74B1 towards other thiol acceptors is currently under investigation.

Experimental procedures

Materials

UDP-α-D-glucopyranoside disodium, UDP-α-D-galacto-pyranoside disodium, and UDP-α-D-*N*-acetylglucosamine disodium salts were purchased from Carbosynth. Phenylacetaldehyde ≥95%, 3-phenylpropanal ≥95%, $Na_2S{\cdot}9H_2O$, and pyruvate kinase/lactate dehydrogenase suspension were purchased from Sigma-Aldrich. Calf Intestine Alkaline Phosphatase (CIAP) was purchased from ThermoScientific. Flash column chromatography was performed either on silica gel 60N (spherical, neutral, 40–63 μm, Merck) or C_{18} or with a Reveleris® flash chromatography system. TLC (precoated aluminium backed plates, Merck Kieselgel $60F_{254}$) was visualized under UV light (254 nm) and by charring after exposure to 10% H_2SO_4 ethanol solution or 1% potassium permanganate aqueous solution or 5% phosphomolybdic acid ethanol solution. Solvents were dried by standard methods: pyridine was dried using potassium hydroxide, *N*,*N*-dimethylformamide was dried using molecular sieves. Molecular sieves were activated prior to use by heating for 4 h at 500 °C. *N*-Chlorosuccinimide (NCS) was recrystallized prior to use from boiling acetic acid. All other commercially available solvents and reagents were used without further purification. ^{1}H NMR and ^{13}C NMR spectra were recorded on Bruker Avance II 400 or Bruker DPX 250 spectrometers. Assignments are based on the DEPT 135 sequence and on homo- and heteronuclear correlations (COSY, HSQC, HMBC). Chemical shifts are reported in parts per million (ppm) from tetramethylsilane as the internal standard in $CDCl_3$ or from residual solvents (DMSO or acetone) for ^{13}C NMR in D_2O. The reported coupling constants (*J*) are expressed in Hertz (Hz) and splitting patterns are designated as br (broad), s (singlet), d (doublet), dt (doublet of triplets), t (triplet) and m (multiplet). HRMS spectra were obtained from the "Fédération de Recherche" ICOA/CBM (FR2708) platform with a Maxis Bruker 4G instrument in the electrospray ionization (ESI) mode. IR spectra were recorded with a Thermo Scientific Nicolet iS10 spectrophotometer. The following chemicals are abbreviated: diethyl ether (Et_2O), dimethylsulfoxide (DMSO), *N*,*N*-dimethylformamide (DMF), dithiothreitol (DTT), ethyl acetate (EtOAc) and petroleum ether (PE).

Preparation of UGT74B1 acceptor substrates

Phenylacetaldoxime (**1**) and 3-phenylpropanaldoxime (**2**) were prepared according to the published literature.[29]

General procedure 1 for the preparation of hydroximoyl chlorides 3–4. The preparation of phenylethanohydroximoyl chloride (**3**) and 3-phenylpropanohydroximoyl chloride (**4**) was adapted from the published literature.[34] In a round-bottom flask covered with aluminum foil and under an argon atmosphere, 1/5th of the NCS was added initially to a solution of the aldoxime (1 equiv.) in DMF (3–5.5 mL). The mixture was cooled to 0 °C and stirred for about 30 min. The remaining 4/5th of the NCS was then added. A total of 1.05–1.86 equiv. of NCS was used. The reaction mixture was poured into ice-cold water (10 mL) and extracted with EtOAc (3 × 10 mL). The combined organic phases were washed with brine (10 mL), dried over anhydrous $MgSO_4$, filtered and concentrated under reduced pressure. The hydroximoyl chloride was purified by silica gel column chromatography when necessary.

Phenylethanohydroximoyl chloride (3). This was obtained after silica gel column chromatography (PE/EtOAc 19/1) as a whitish powder (412 mg, 82%); R_f 0.25 (PE/EtOAc 19/1).[29]

3-Phenylpropanohydroximoyl chloride (4). General procedure 1 was followed with 3-phenylpropanaldoxime **2** (600 mg, 4.02 mmol, 1 equiv.) and NCS (575 mg, 4.30 mmol, 1.07 equiv.) in anhydrous DMF (4 mL). After work up, crude **4**, obtained as a light yellow oil, was used right away in the next step.[34]

General procedure 2 for the preparation of thiohydroxamic acids 5–6. Thiohydroxamic acids **5** and **6** were prepared according to the published procedure.[34] In a round-bottom flask covered with aluminum foil and under an argon atmosphere, an aqueous solution of $Na_2S{\cdot}9H_2O$ (3–3.1 equiv.) followed by Et_3N (1 equiv.) was added to a stirred solution of the hydroximoyl chloride **3** or **4** (1 equiv.) in Et_2O (0.07–0.08 M). The reaction mixture was stirred at r.t. until complete consumption of the starting material, and then poured into a separatory funnel. After discarding the Et_2O layer, EtOAc (15 mL) was added to the aqueous phase, which was acidified to pH 2 by adding 2 M aqueous HCl (5 equiv.) to the stirred mixture. The EtOAc solution was separated and the aqueous phase was extracted with EtOAc (2 × 15 mL). The combined organic phases were dried over Na_2SO_4, filtered and concentrated under reduced pressure. The thiohydroxamic acids were purified by column chromatography on the C18 reverse phase (H_2O 100% to CH_3CN 100%).

Phenylacetothiohydroxamic acid (5). This was obtained, following general procedure 2 (for 14 h), after chromatographic purification on the C18 reverse phase (H_2O 100% to CH_3CN 100%) to afford **5** as a white powder (57 mg, 14%); R_f 0.15 (PE/EtOAc 1 : 1).

3-Phenylpropanothiohydroxamic acid (6). This was obtained following general procedure 2 (for 1 h), after chromatographic purification on the C18 reverse phase (H_2O 100% to CH_3CN 100%) to afford **6** as a white powder (176 mg, 24% over the last two steps); R_f 0.18 (PE/EtOAc 1/1). ^{1}H NMR (400 MHz, acetone-d6): δ 7.29–7.18 (m, 5H, H-Ar), 3.06 (t, 2H, J_{3-2} = 7.6, H-3), 2.86 (t, 2H, J_{2-3} = 7.6, H-2). ^{13}C NMR (100 MHz, acetone-d6): δ 184.8 (C-1), 141.3 (*ipso*-C-Ph), 129.3 (CH-Ph), 129.2 (CH-Ph), 127.1 (*para*-CH-Ph), 41.5 (C-2), 35.1(C-3). IR (neat) ν cm^{-1}: 3165 (OH), 3025 (NH), 1584 (C═N), 1201 (C═S), 1076 (N–O).[34]

Chemoenzymatic synthesis of desulfoglycosinolates 7, 9–10

General procedure 3 for the preparation of desulfoglycosinolates 7, 9–10. In a 0.5–2.0 mL conical tube covered with aluminium foil, reagents were added in the following order: UDP-monosaccharide disodium salt (1.1 equiv.), 0.1 M phosphate buffer (pH 7), phenylacetothiohydroxamic acid **5** (1 equiv.), DMSO, 0.1 mM DTT, 15 μM UGT74B1, and calf intestine phosphatase. After stirring at 37 °C for 2 h, the reaction mixture was purified on a HyperSep C18 cartridge (H_2O 100% to MeOH 100%) to afford the desulfoglycosinolate.

Desulfoglucotropaeolin (7). General procedure 3 was followed with UDP-glucose di Na (19.9 mg, 0.033 mmol, 1.1 equiv.), 3.4 mL 0.1 M phosphate buffer (pH 7), thiohydroxamic acid **5** (5 mg, 0.0299 mmol, 1 equiv.), DMSO (400 μL), 0.5 M DTT (80 μL), 1 mg ml^{-1} UGT74B1 (200 μL), and 1 unit per μl CIAP (10 μL). Purification on the HyperSep C18 cartridge (H_2O 100% to MeOH 100%) afforded **7** as an amorphous white powder (6.6 mg, 69% yield); R_f 0.37 (MeOH/EtOAc 1 : 19). ^{1}H NMR (400 MHz, D_2O): δ 7.48–7.38 (m, 5H, H-Ar), 4.73 (d, 1H, J_{1-2} = 9.6, H-1), 4.07 (s, 2H, CH_2Ph), 3.72 (dd, 1H, J_{6a-6b} = 12.4, J_{6a-5} = 2.5, H-6a), 3.66 (dd, 1H, J_{6b-6a} = 12.4, J_{6b-5} = 4.9, H-6b), 3.45–3.41 (m, 1H, H-4), 3.38–3.32 (m, 2H, H-2, H-3), 3.30–3.26 (m, 1H, H-5). ^{13}C NMR (100 MHz, D_2O): δ 154.9 (C-7), 136.8 (*ipso*-C-Ph), 129.7 (CH-Ph), 128.7 (CH-Ph), 128.0 (*para*-CH-Ph), 81.7 (C-1), 80.4 (C-5), 77.7 (C-3), 72.7 (C-2), 69.6 (C-4), 61.0 (C-6), 38.6 (*C*H_2Ph). HRMS-ESI$^+$ *m*/*z* calcd for $C_{14}H_{20}O_6SN$: 330.100585; found 330.100620 $[M + H]^+$.[41,56]

Desulfogalactotropaeolin (9). General procedure 3 was followed with UDP-galactose di Na (10 mg, 0.016 mmol, 1.1 equiv.), 1.7 mL 0.1 M phosphate buffer (pH 7), thiohydroxamic acid **5** (2.5 mg, 0.015 mmol, 1 equiv.), DMSO (200 μL), 0.5 M DTT (40 μL), 1 mg ml^{-1} UGT74B1 (100 μL), and 1 unit per μl CIAP (5 μL). Purification on the HyperSep C18 cartridge (H_2O 100% to MeOH 100%) afforded **8** as an amorphous white powder (0.7 mg, 19% yield); R_f 0.13 (EtOAc). ^{1}H NMR (400 MHz, D_2O): δ 7.49–7.39 (m, 5H, H-Ar), 4.68 (d, 1H, J_{1-2} = 9.4, H-1), 4.11 (d, 1H, J_{gem} = 16.4, Ha-CH_2Ph), 4.06 (d, 1H, J_{gem} = 16.4, Hb-CH_2Ph), 3.96 (d, 1H, J_{4-3} = 3.1, H-4), 3.72–3.50 (m, 5H, H-2, H-3, H-5, H-6a, H-6b). ^{13}C NMR (100 MHz, D_2O): δ 154.5 (C-7), 129.1 (CH-Ph), 128.0 (CH-Ph), 127.3 (*para*-C-Ph), 81.5 (C-1), 78.9 (C-5), 73.7 (C-3), 69.3 (C-2), 68.2 (C-4), 60.4 (C-6), 37.9 (*C*H_2Ph). HRMS-ESI$^+$ *m*/*z* calcd for $C_{14}H_{20}O_6SN$: 330.100585; found 330.100677 $[M + H]^+$; calcd for $C_{14}H_{19}O_6SNaN$: 352.082529; found 330.082400 $[M + Na]^+$.

Desulfo *N*-acetylglucosaminotropaeolin (10). General procedure 3 was followed with UDP-*N*-acetylglucosamine di Na (11 mg, 0.018 mmol, 1.1 equiv.), 1.9 mL 0.1 M phosphate buffer (pH 7), thiohydroxamic acid **5** (2.75 mg, 0.016 mmol, 1 equiv.), DMSO (220 μL), 0.5 M DTT (44 μL), 1 mg ml^{-1} UGT74B1 (300 μL), and 1 unit per μl CIAP (5.5 μL). Purification on the HyperSep C18 cartridge (H_2O 100% to MeOH 100%) afforded **9** as an amorphous white powder (2 mg, 34% yield); R_f 0.19 (MeOH/EtOAc 1 : 19). ^{1}H NMR (250 MHz, D_2O); δ 7.47–7.35 (m, 5H, H-Ar), 4.79 (under the solvent peak, 1H, H-1), 4.04 (s, 2H, CH_2Ph), 3.79–3.65 (m, 3H, H-2, H-6a, H-6b), 3.51–3.38 (m, 2H, H-3, H-4), 3.35–3.27 (m, 1H, H-5), 1.96 (s, 3H, *C*H_3CO). ^{13}C NMR (62.5 MHz, D_2O): δ 174.3 (CH$_3$*C*O), 153.9 (C-7), 138.1 (*ipso*-C-Ph), 129.1 (CH-Ph), 127.9 (CH-Ph), 127.3 (*para*-CH-Ph), 80.2 (C-1), 79.8 (C-5), 74.6 (C-3 or C-4), 69.2 (C-4 or C-3), 60.4 (C-6), 54.2 (C-2), 37.7 (*C*H_2Ph), 22.0 (*C*H_3CO). HRMS-ESI$^+$ *m*/*z* calcd for $C_8H_{14}O_5N$: 204.086649; found 204.086877 $[N$-Ac-Glc$]^+$, calcd for $C_{16}H_{22}O_6SNaN_2$: 393.109078; found 393.108913 $[M + Na]^+$.

Cloning and overexpression of UGT74B1

The *ugt74b1* gene cDNA from *A. thaliana* (locus AT1G24100) cloned in pENTR/SD-TOPO was obtained from the Arabidopsis Biological Resource Center (clone U15624). The gene was amplified by PCR using the following primers: 5′-cg**gaattc**gcggaaacaactcccaaagtgaa-3′ (forward) and 5′-cg**ctcgag**ttacttcccctaaactctctataaact-3′ (reverse). EcoRI and XhoI restriction sites (indicated in bold) were respectively added up- and downstream of the gene for subsequent cloning in the pET-28a(+) expression vector (Novagen), that adds a N-terminal His-Tag to the recombinant protein. Rosetta(DE3) *E. coli* cells transformed with the cloned plasmid were grown in LB medium at 37 °C until OD600 reached 0.5, and then were induced by 0.2 mM IPTG at 25 °C overnight. Cells were harvested and lyzed by several freeze–thaw cycles. The lysate was clarified by centrifugation (20 000*g*, 30 min). The supernatant was filtered (0.45 μm), loaded on a HisPur Ni-NTA column (Thermo), and then bound protein was eluted by an imidazole gradient (10 to 500 mM). The protein concentration and purity were respectively assessed by the Bradford assay[57] and SDS-PAGE analysis.

Enzymatic assay

The UGT74B1 activity was assayed at 37 °C in Tris buffer (20 mM, pH 8) containing 50 mM KCl and 20 mM $MgCl_2$, following and adapting the reported procedure for glucosyltransferases.[35,36,58] This procedure enables the enzymatic coupling of the UDP production by UGT74B1 to the oxidation of NADH, using pyruvate kinase (PK) and lactate dehydrogenase (LDH). The consumption of NADH can be monitored *in situ* using a spectrophotometer. Briefly, in a microplate well, the two substrates (UDP-sugar donor and acceptor) were mixed with 100 μM NADH, 0.5 mM DTT, 3.5 mM phosphoenolpyruvate, 5 U PK, and 10 U LDH. After preincubation at 37 °C, UGT74B1 (final concentration of 0.05 μM) was added to give a total volume of 200 μL. The activity was then recorded at 340 nm, using an extinction coefficient for NADH of 6220 M^{-1} cm^{-1}.

Data analysis. Kinetic parameters k_{cat} and K_M were obtained from 3 independent experiments, and determined using Prism 5 software (Graph Pad).

Molecular modeling

Suitable templates were first identified using the NCBI BLAST[59] software (Basic Local Alignment Search Tool) using a position-specific iterated protein search limited to proteins within the protein data bank with an expect threshold of 10^{-3} for the first BLAST run and a PSI-BLAST threshold of 10^{-5}. This search employed the BLOSUM62 (BLOcks SUbstitution

Matrix 62) substitution matrix for the first run and then iteratively constructed a substitution matrix based on the results of the first run.

Based on this first step, proteins with both the highest sequence identity and the highest query cover were selected namely: Uniprot P51094, a flavonoid 3-*O*-glucosyltransferase from *Vitis vinifera* (common grape vine)[60] – PDB codes 2C1X, 2C1Z and 2C9Z PDB codes 2C1X, 2C1Z and 2C9Z; Uniprot A6XNC5, a multifunctional (iso)flavonoid glycosyltransferase (UGT85H2) from *Medicago truncatula* (barrel clover)[61] – PDB code 2PQ6; Uniprot A4F1R4, a UDP-glucose anthocyanidin 3-*O*-glucosyltransferase from *Clitoria ternatea* (butterfly pea)[62,63] – PDB codes 3WC4, 4WHM, 4REL, 4REM and 4REN; Uniprot Q9M156, a bifunctional *O*- and *N*-UDP-glycosyltransferase (UGT72B1) from *Arabidopsis thaliana*[11] – PDB codes 2VCE, 2VCH and 2VG8, and Uniprot Q5IFH7 a multifunctional triterpene/flavonoid glycosyltransferase from *Medicago truncatula* (UGT71G1)[64] – PDB codes 2ACV and 2ACW.

To account for the low sequence identity but high structural conservation within the glycosyltransferase family of proteins, the TCOFFEE[65] sequence alignment software was used in the accurate mode. TCOFFEE in its accurate mode combines sequence, profile, and structural information to produce highly accurate and biologically significant alignments.

The structural correspondence of the final alignment was visualized using the Chimera 1.10.2 software's Multalign Viewer.[66] The structure prediction of UGT74B1 was carried out using comparative protein structure modeling by satisfaction of spatial restraints using the MODELLER[67] software. A total of 10 000 different UGT74B1 structures were produced based on the sequence alignment obtained by TCOFFEE and the multiple prepared templates. The best 100 models were selected based on a normalized DOPE score. All structures produced had a perfect GA341 score of 1.0. The chosen models were then clustered using the Calibur[68] program and the most representative structure of the largest cluster was chosen as the final homology model.

Ligands were drawn using the SybylX program and subjected to a minimization stage by using the MMFF force field using the BFGS algorithm in 1000 steps or until energy conversion (grad = 0.1).

Molecular docking simulations were run using the Autodock Vina[69] software using the default scoring function weights and an optimized exhaustiveness parameter of 16. For each ligand, 20 different poses were produced. The search space had the following dimensions $x \times y \times z = 29 \times 24 \times 28 = 19\,488$ Å^3 centered around the active site region.

Acknowledgements

This work has been partially supported by the Labex SynOrg (ANR-11-LABX-0029). We gratefully acknowledge the CRIANN computational centre for providing computational resources.

References

1 A. Varki, *Glycobiology*, 1993, **3**, 97–130.
2 P. Lafite and R. Daniellou, *Nat. Prod. Rep.*, 2012, **29**, 729–738.
3 G. Lian, X. Zhang and B. Yu, *Carbohydr. Res.*, 2015, **403**, 13–22.
4 L. Guillotin, P. Lafite and R. Daniellou, in *Carbohydrate Chemistry*, The Royal Society of Chemistry, 2014, vol. 40, pp. 178–194.
5 T. J. Oman, J. M. Boettcher, H. Wang, X. N. Okalibe and W. A. van der Donk, *Nat. Chem. Biol.*, 2011, **7**, 78–80.
6 J. Stepper, S. Shastri, T. S. Loo, J. C. Preston, P. Novak, P. Man, C. H. Moore, V. Havlicek, M. L. Patchett and G. E. Norris, *FEBS Lett.*, 2011, **585**, 645–650.
7 H. Wang, T. J. Oman, R. Zhang, C. V. Garcia De Gonzalo, Q. Zhang and W. A. van der Donk, *J. Am. Chem. Soc.*, 2014, **136**, 84–87.
8 B. A. Halkier and J. Gershenzon, *Annu. Rev. Plant Biol.*, 2006, **57**, 303–333.
9 F. S. Hanschen, E. Lamy, M. Schreiner and S. Rohn, *Angew. Chem., Int. Ed.*, 2014, **53**, 11430–11450.
10 A. T. Dinkova-Kostova and R. V. Kostov, *Trends Mol. Med.*, 2012, **18**, 337–347.
11 M. Brazier-Hicks, W. A. Offen, M. C. Gershater, T. J. Revett, E.-K. Lim, D. J. Bowles, G. J. Davies and R. Edwards, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 20238–20243.
12 K. Xie, R. Chen, J. Li, R. Wang, D. Chen, X. Dou and J. Dai, *Org. Lett.*, 2014, **16**, 4874–4877.
13 R. W. Gantt, R. D. Goff, G. J. Williams and J. S. Thorson, *Angew. Chem., Int. Ed.*, 2008, **47**, 8889–8892.
14 C. D. Grubb, B. J. Zipp, J. Ludwig-Müller, M. N. Masuno, T. F. Molinski and S. Abel, *Plant J.*, 2004, **40**, 893–908.
15 C. M. M. Gachon, M. Langlois-Meurinne, Y. Henry and P. Saindrenan, *Plant Mol. Biol.*, 2005, **58**, 229–245.
16 E.-F. Marillia, J. M. MacPherson, E. W. T. Tsang, K. Van Audenhove, W. A. Keller and J. W. D. GrootWassink, *Physiol. Plant.*, 2001, **113**, 176–184.
17 Y.-X. Zang, H. U. Kim, J. A. Kim, M.-H. Lim, M. Jin, S. C. Lee, S.-J. Kwon, S.-I. Lee, J. K. Hong, T.-H. Park, J.-H. Mun, Y.-J. Seol, S.-B. Hong and B.-S. Park, *FEBS J.*, 2009, **276**, 3559–3574.
18 J. Kopycki, E. Wieduwild, J. Kohlschmidt, W. Brandt, A. N. Stepanova, J. M. Alonso, M. S. Pedras, S. Abel and C. D. Grubb, *Biochem. J.*, 2013, **450**, 37–46.
19 S. Paquette, B. L. Møller and S. Bak, *Phytochemistry*, 2003, **62**, 399–413.
20 H. Streicher, L. Latxague, T. Wiemann, P. Rollin and J. Thiem, *Carbohydr. Res.*, 1995, **278**, 257–270.
21 P. Rollin and A. Tatibouët, *C. R. Chim.*, 2011, **14**, 194–210.
22 R. Elek, L. Kiss, J.-P. Praly and L. Somsák, *Carbohydr. Res.*, 2005, **340**, 1397–1402.
23 L. Brochard, B. Joseph, M.-C. Viaud and P. Rollin, *Synth. Commun.*, 1994, **24**, 1403–1414.

24 R. Iori, P. Rollin, H. Streicher, J. Thiem and S. Palmieri, *FEBS Lett.*, 1996, **385**, 87–90.
25 S. Cottaz, B. Henrissat and H. Driguez, *Biochemistry*, 1996, **35**, 15256–15259.
26 D. Cerniauskaite, J. Rousseau, A. Sackus, P. Rollin and A. Tatibouët, *Eur. J. Org. Chem.*, 2011, 2293–2300.
27 C. Gardrat, A. Quinsac, B. Joseph and P. Rollin, *Heterocycles*, 1993, **35**, 1015–1027.
28 S. Cottaz, P. Rollin and H. Driguez, *Carbohydr. Res.*, 1997, **298**, 127–130.
29 Q. V. Vo, C. Trenerry, S. Rochfort, J. Wadeson, C. Leyton and A. B. Hughes, *Bioorg. Med. Chem.*, 2013, **21**, 5945–5954.
30 Q. V. Vo, C. Trenerry, S. Rochfort, J. Wadeson, C. Leyton and A. B. Hughes, *Bioorg. Med. Chem.*, 2014, **22**, 856–864.
31 M. S. C. Pedras, Q. H. To and G. Schatte, *Chem. Commun.*, 2016, **52**, 2505–2508.
32 R. Nehmé, H. Nehmé, G. Roux, D. Cerniauskaite, P. Morin, P. Rollin and A. Tatibouët, *Anal. Chim. Acta*, 2014, **807**, 153–158.
33 J. Kopycki, J. Schmidt, S. Abel and C. Douglas Grubb, *Biotechnol. Lett.*, 2011, **33**, 1039–1046.
34 B. C. Lemercier and J. G. Pierce, *J. Org. Chem.*, 2014, **79**, 2321–2330.
35 N. L. Rose, R. B. Zheng, J. Pearcey, R. Zhou, G. C. Completo and T. L. Lowary, *Carbohydr. Res.*, 2008, **343**, 2130–2139.
36 S. Gosselin, M. Alhussaini, M. B. Streiff, K. Takabayashi and M. M. Palcic, *Anal. Biochem.*, 1994, **220**, 92–97.
37 V. Lombard, H. Golaconda Ramulu, E. Drula, P. M. Coutinho and B. Henrissat, *Nucleic Acids Res.*, 2014, **42**, D490–D495.
38 B. Rost, *Protein Eng.*, 1999, **12**, 85–94.
39 L. L. Lairson, B. Henrissat, G. J. Davies and S. G. Withers, *Annu. Rev. Biochem.*, 2008, **77**, 521–555.
40 A. Chang, S. Singh, G. N. Phillips Jr. and J. S. Thorson, *Curr. Opin. Biotechnol.*, 2011, **22**, 800–808.
41 N. Agerbirk, S. I. Warwick, P. R. Hansen and C. E. Olsen, *Phytochemistry*, 2008, **69**, 2937–2949.
42 H. Driguez, *ChemBioChem*, 2001, **2**, 311–318.
43 L. Guillotin, P. Lafite and R. Daniellou, *Biochemistry*, 2014, **53**, 1447–1455.
44 D. J. Wardrop and S. L. Waidyarachchi, *Nat. Prod. Rep.*, 2010, **27**, 1431–1468.
45 K. Pachamuthu and R. R. Schmidt, *Chem. Rev.*, 2006, **106**, 160–187.
46 M. Jahn, J. Marles, R. A. J. Warren and S. G. Withers, *Angew. Chem., Int. Ed.*, 2003, **42**, 352–354.
47 M. Jahn, H. Chen, J. Müllegger, J. Marles, R. A. J. Warren and S. G. Withers, *Chem. Commun.*, 2004, 274–275.
48 J. Müllegger, M. Jahn, H.-M. Chen, R. A. J. Warren and S. G. Withers, *Protein Eng., Des. Sel.*, 2005, **18**, 33–40.
49 Y.-W. Kim, H. Chen, J. H. Kim and S. G. Withers, *FEBS Lett.*, 2006, **580**, 4377–4381.
50 Y.-W. Kim, A. L. Lovering, H. Chen, T. Kantner, L. P. McIntosh, N. C. J. Strynadka and S. G. Withers, *J. Am. Chem. Soc.*, 2006, **128**, 2202–2203.
51 J. Müllegger, H. M. Chen, R. A. J. Warren and S. G. Withers, *Angew. Chem., Int. Ed.*, 2006, **45**, 2585–2588.
52 J. Müllegger, H.-M. Chen, W. Y. Chan, S. P. Reid, M. Jahn, R. A. J. Warren, H. M. Salleh and S. G. Withers, *ChemBioChem*, 2006, **7**, 1028–1030.
53 Y.-W. Kim, H.-M. Chen, J. H. Kim, J. Müllegger, D. Mahuran and S. G. Withers, *ChemBioChem*, 2007, **8**, 1495–1499.
54 Z. Armstrong, S. Reitinger, T. Kantner and S. G. Withers, *ChemBioChem*, 2010, **11**, 533–538.
55 M. Almendros, D. Danalev, M. François-Heude, P. Loyer, L. Legentil, C. Nugier-Chauvin, R. Daniellou and V. Ferrières, *Org. Biomol. Chem.*, 2011, **9**, 8371–8378.
56 G. Kiddle, R. N. Bennett, N. P. Botting, N. E. Davidson, A. A. B. Robertson and R. M. Wallsgrove, *Phytochem. Anal.*, 2001, **12**, 226–242.
57 M. M. Bradford, *Anal. Biochem.*, 1976, **72**, 248–254.
58 J. C. Krupa, D. Shaya, L. L. Chi, R. J. Linhardt, M. Cygler, S. G. Withers and J. S. Mort, *Anal. Biochem.*, 2007, **361**, 218–225.
59 S. F. Altschul, W. Gish, W. Miller, E. W. Myers and D. J. Lipman, *J. Mol. Biol.*, 1990, **215**, 403–410.
60 W. Offen, C. Martinez-Fleites, M. Yang, E. Kiat-Lim, B. G. Davis, C. A. Tarling, C. M. Ford, D. J. Bowles and G. J. Davies, *EMBO J.*, 2006, **25**, 1396–1405.
61 L. Li, L. V. Modolo, L. L. Escamilla-Trevino, L. Achnine, R. A. Dixon and X. Wang, *J. Mol. Biol.*, 2007, **370**, 951–963.
62 T. Hiromoto, E. Honjo, T. Tamada, N. Noda, K. Kazuma, M. Suzuki and R. Kuroki, *J. Synchrotron Radiat.*, 2013, **20**, 894–898.
63 T. Hiromoto, E. Honjo, N. Noda, T. Tamada, K. Kazuma, M. Suzuki, M. Blaber and R. Kuroki, *Protein Sci.*, 2015, **24**, 395–407.
64 H. Shao, X. He, L. Achnine, J. W. Blount, R. A. Dixon and X. Wang, *Plant Cell*, 2005, **17**, 3141–3154.
65 I. M. Wallace, O. O'Sullivan, D. G. Higgins and C. Notredame, *Nucleic Acids Res.*, 2006, **34**, 1692–1699.
66 E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J. Comput. Chem.*, 2004, **25**, 1605–1612.
67 P. Benkert, T. Schwede and S. C. Tosatto, *BMC Struct. Biol.*, 2009, **9**, 1–17.
68 S. C. Li and Y. K. Ng, *BMC Bioinf.*, 2010, **11**, 25.
69 O. Trott and A. J. Olson, *J. Comput. Chem.*, 2010, **31**, 455–461.