TECHNISCHE UNIVERSITÄT MÜNCHEN

TUM School of Life Sciences

Comprehensive functional characterization of a glucosyltransferase from *Nicotiana benthamiana*

Jieren Liao

Vollständiger Abdruck der von der TUM School of Life Sciences der Technischen Universität München zur Erlangung einer

Doktorin der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz:	Prof. Dr. Michael Rychlik
Prüfende der Dissertation:	1. Prof. Dr. Wilfried Schwab
	2. Prof. Dr. Wolfgang Liebl

Die Dissertation wurde am 24.01.2024 bei der Technischen Universität München eingereicht und durch die TUM School of Life Sciences am 20.06.2024 angenommen.

Contents

Abstract

Protein conformational dynamics play a pivotal role in enzyme catalysis and hold significant implications in bioengineering. Notably, scopoletin exerts a substantial inhibitory effect on a UDP-sugar dependent glucosyltransferase from *Nicotiana benthamiana* (NbUGT72AY1), prompting the exploration of two intriguing hypotheses: allosteric conformational shift of the protein and substrate-induced conformational shift.

In the initial phase of this study, a homolog from *Solanum tuberosum* (StUGT72AY2), which shares an impressive 80% amino acid sequence similarity with NbUGT72AY1, was found to exhibit limited substrate inhibition when exposed to scopoletin. Through the utilization of advanced techniques such as HDX-MS (Hydrogen-Deuterium Exchange Mass Spectrometry), mutant enzyme generation, and docking, pivotal amino acids critical for enzymatic activity and scopoletin binding were identified. Mutation analysis identified key amino acids influencing the enzyme's activity. Remarkably, the amino acid exchange of the F87I led to an increase in glucosyltransferase activity while reducing substrate inhibition of NbUGT72AY1. Furthermore, the creation of chimera I155F-C207D virtually eradicated substrate inhibition. The data obtained by the UDP-Glo^{TM} glycosyltransferase assay and iquid chromatography-mass spectrometry (LC-MS) analysis were fitted using a two-binding site equation, providing a better understanding of the enzymatic behavior.

In the second part of this thesis, a series of substrates characterized by different chemical structures were employed to investigate the kinetics. This resulted in a wide spectrum of profiles, ranging from classical Michaelis-Menten behavior (MM) with hydroquinones to substrates exhibiting weak inhibition (sinapaldehyde, guaiacol, o-cresol) and those displaying strong substrate inhibition (scopoletin, eugenol, and vanillin). Notably, vanillin, which shares a common phenol 2-methoxy structure with scopoletin and exhibits a similar substrate inhibition pattern to the hydroxycoumarins, was studied as a substrate in detail. Coumarins and fatty acids were investigated as effectors of NbUGT72AY1 and its mutations. While coumarins had no discernible impact on vanillin enzymatic activity of NbUGT72AY1, fatty acids led to an increase in glucosyltransferase activity but did not

affect substrate inhibition. Remarkably, alpha- and beta-ionol not only enhanced scopoletin glucosyltransferase activity but also reduced substrate inhibition. Furthermore, the introduction of the F87I mutation and the chimeric I155F-N191S mutation demonstrated reduced substrate inhibition and glucosyltransferase activity in the presence of stearic acid. Overall, these findings suggest that the allosteric binding hypothesis may not be applicable to NbUGT72AY1. Furthermore, they confirm the multi-substrate functionality of NbUGT72AY1, shedding light on the enzyme's capacity to be finely tuned by external metabolites. This research offers valuable insights into the complex interplay between enzyme kinetics, substrate recognition, and the regulatory influence of various metabolites on NbUGT72AY1.

Zusammenfassung

Die konformationelle Dynamik von Proteinen spielt eine entscheidende Rolle bei der Enzymkatalyse und hat bedeutende Auswirkungen auf die Bioverfahrenstechnik. So weist Scopoletin eine starke Substratinhibierung bei der Umsetzung mit der UDP-Zucker abhängigen Glukosyltransferase NbUGT72AY1 aus *Nicotiana benthamiana* auf, was zur Untersuchung von zwei faszinierenden Hypothesen führte: der allosterischen Konformationsänderung Veränderung von Proteinen und der durch Substrate hervorgerufenen Strukturänderung.

In der ersten Phase dieser Studie wurde festgestellt, dass eine homologe Glukosyltransferase aus *Solanum tuberosum* StUGT72AY2, welche eine beeindruckende Aminosäure-Sequenzähnlichkeit von 80% mit NbUGT72AY1 aufweist, eine weitaus geringere Substrathemmung zeigt, wenn sie mit Scopoletin umgesetzt wird. Fortschrittliche Techniken wie HDX-MS (Wasserstoff-Deuterium-Austausch-Massenspektrometrie), Herstellung von Mutanten und *In-silico*-Docking wurden verwendet, um entscheidende Aminosäuren zu identifizieren, die für die Bindung von Scopoletin durch NbUGT72AY1 entscheidend sind. Die Analyse von Mutanten führte zur Identifizierung von Schlüsselaminosäuren, die die Aktivität des Enzyms beeinflussen. Bemerkenswerterweise führte die Einführung der Aminosäurenaustausch F87I zu einer Zunahme der Glucosyltransferase-Aktivität und einer Reduzierung der Substrathemmung. Darüber hinaus eliminierte die Schaffung der Chimäre I155F-C207D praktisch die Substrathemmung. Die durch UDP-Glo™ Glykosyltransferase-Assay und Flüssigchromatographie-Massenspektrometrie (LC-MS) Analyse erhaltenen Daten konnten mithilfe einer Zwei-Bindungsstellen-Gleichung angenähert werden, was zu einem besseren Verständnis des Verhaltens des Enzyms führte.

Im zweiten Teil dieser Arbeit wurde eine Vielzahl von Substraten mit unterschiedlichen chemischen Strukturen verwendet, um deren Kinetiken zu untersuchen. Dies führte zu einer breiten Palette von Profilen, die von klassischem Michaelis-Menten-Verhalten (MM) bei Hydrochinonen bis hin zu Substraten reichten, die eine schwache Hemmung zeigten,

(Sinapaldehyd, Guajakol, o-Cresol) und solchen, die eine starke Substrathemmung aufwiesen (Scopoletin, Eugenol und Vanillin). Bemerkenswerterweise zeigte auch Vanillin, das wie Scopoletin eine Phenol-2-Methoxygruppe aufweist eine ähnliche Substrathemmungskinetik wie Hydroxycumarine und wurde eingehender als Substrat untersucht. Neben Cumarinen wurden auch Fettsäuren als Effektoren der NbUGT72AY1 und ihren Mutanten analysiert. Während Cumarine keine erkennbare Auswirkung auf die enzymatische Umsetzung von Vanillin durch NbUGT72AY1 hatten, führten Fettsäuren zu einer Zunahme der Glucosyltransferase-Aktivität, ohne die Substrathemmung zu beeinflussen. Bemerkenswerterweise steigerten Alpha- und Beta-Ionol nicht nur die Glucosyltransferase-Aktivität von Scopoletin, sondern reduzierten auch die Substrathemmung. Darüber hinaus führte die Einführung der Mutation F87I und der chimären Mutation I155F-N191S in Anwesenheit von Stearinsäure zu einer reduzierten Substrathemmung und Glucosyltransferase-Aktivität.

Diese Ergebnisse deuten insgesamt darauf hin, dass die Hypothese der allosterischen Bindung möglicherweise nicht auf NbUGT72AY1 zutrifft. Sie bestätigen außerdem die multifunktionale Eigenschaft von NbUGT72AY1 und zeigen seine Fähigkeit zur Feinabstimmung durch externe Metaboliten. Diese Ergebnisse liefern wertvolle Einblicke in das komplexe Zusammenspiel zwischen Enzymkinetik, Substraterkennung und dem regulierenden Einfluss verschiedener Metaboliten auf NbUGT72AY1.

Abbreviations

ABA	Abscisic acid
CAST/ISM	Combinatorial active site saturation test/iterative saturation mutagenesis
CAZy	Carbohydrate-Active Enzyme database
CD	Circular Dichroism
CDART	Conserved Domain Architecture Retrieval Tool
Cryo-EM	Cryo-electron miscrocopy
DEDAL	Deep embedding and differentiable alignment
E	Enzyme
EC	Enzyme Commission
ECatDB	Enzyme Catalytic-mechanism Database
EI	Enzyme-intermediate complex
EP	Enzyme–product complex
ES	Enzyme–substrate complex
FRET	Fluorescence resonance energy transfer
GT	Glycosyltransferase
HAMAP	High-quality Automated and Manual Annotation of Proteins
HDX-MS	Hydrogen-Deuterium Exchange Mass Spectrometry
http	HyperText Transfer Protocol
IAA	Indole-3-acetic acid
KEGG	Kyoto Encyclopedia of Genes and Genomes
LC-MS	Liquid Chromatography Mass Spectrometry
M-CSA	Mechanism and Catalytic Site Atlas
MACiE	Mechanism, Annotation and Classification in Enzymes
MM	Michaelis–Menten
NCBI	National Center for Biotechnology Information
NMR	Nuclear Magnetic Resonance
PDB	Protein Data Bank
PSPG	Plant secondary product GT motif
SAG	Salicylic acid glucoside
SAR	Systemic acquired resistance
SCOP	Structure Classification of Proteins
SFLD	Structure Function Linkage Database
TS	Transition State

1. Introduction

1.1. Enzyme classification and functional annotation

1.1.1. Enzyme definition and classification

Enzymes are proteins that catalyze various biochemical reactions in all living organisms [1]. These macromolecular biological catalysts are selective, and each catalyzes the conversion of only a limited number of substrates. In addition to substrate specificity, enzymes also exhibit product selectivity as well as regio- and enantioselectivity [2]. While product specificity results from the fact that an enzyme catalyzes only one particular reaction, regioselectivity refers to the preferential conversion of one functional group in the reactant by the catalyst over a second identical group.

The use of enzymes, i.e. biocatalysis is critical to a green, sustainable, biobased economy, and this has led to major advances in biotechnology and biocatalysis over the past decades [3]. Biocatalysis offers numerous advantages, including higher selectivity and specificity, lower operating costs, and reduced toxicity, all of which lead to a lower environmental impact from industrial processes. In order to use enzymes efficiently in biotechnology, a comprehensive knowledge of their biochemical properties is essential.

Enzymes are assigned to groups according to their different properties [3]. The most prevalent method is a hierarchical categorization, which classifies proteins based on evolutionary links and clusters them based on class, architecture, and fold type. The phylogenetic technique generates an evolutionary tree representing different protein families based on their evolutionary history [4]. It can be used to define protein structure, as well as to group proteins based on their structural properties. Here, homologous protein families with the same 3D fold are classified together [5]. Domains can also be seen by sequence analysis. These domains are referred to as "units of evolution" [6]. A protein typically has one or more domains.

As genome projects progress, the number of protein sequences is rapidly increasing, and the majority of them are awaiting experimental characterization. Protein function, protein

structure, and sequence similarity are all described using the Enzyme Commission (EC) classification [1]. Proteins are classified in the EC system according to the reaction they catalyze. The EC system is a four-digit numerical code that represents different levels of chemical functional description. The first digit represents the overarching type of enzyme reaction, such as transferase or hydrolase reaction, whereas the second and succeeding digits represent the specific subclass and serial number allocated to each individual enzyme. The EC system is widely recognized as the gold standard for explaining enzyme function, which includes sequencing, structural, and metabolic pathways [7].

Several protein databases are now publicly available. Based on the type of information stored, protein databases can be classified into different categories. Some of the most common protein databases are as follows. M-CSA (Mechanism and Catalytic Site Atlas) is an enzyme reaction mechanism database that provides annotation to proteins, catalytic residues, cofactors, and the reaction mechanism of hundreds of enzymes [8]. Metal-MACiE (Mechanism, Annotation and Classification in Enzymes) is a new publicly available web-database that aims to consolidate available knowledge on metal characteristics and roles in the context of metalloenzyme catalytic processes [9]. The Structure Function Linkage Database (SFLD) organizes enzymes based on their shared chemical activities [10]. The EzCatDB (Enzyme Catalytic-mechanism Database) database comprises enzyme catalytic processes, and proteins are structured by evaluating and categorizing data from the Protein Data Bank (PDB) [11]. The database Structural Classification of Proteins (SCOP) gives a precise and complete description of structural and evolutionary links for proteins with known structure [12]. Query3D was created to analyze functional residues in protein structures [13] and the Carbohydrate-Active Enzyme database (CAZy) CAZy describes the families of structurally-related catalytic and carbohydrate-binding modules of enzymes that degrade, modify, or create glycosidic bonds [14].

1.1.2. Enzyme functional annotation

Enzymes are categorized into seven groups depending on the chemical reaction they catalyze:

Oxidoreductases, Transferases, Hydrolases, Lyases, Isomerases, Ligases, and Translocases [15]. Much work remains to be done to create credible functional annotations for new protein sequences, particularly on algorithm and machine learning methods. Annotation packages such as GeneQuiz, Rulebase, and HAMAP (High-quality Automated and Manual Annotation of Proteins) are used to correct sequence composition bias and update databases [16]. Functional annotations of proteins, among others, are essential for understanding actual cellular activities [17]. Metabolic pathways are predicted using databases such as Ecocyc [18] and KEGG (Kyoto Encyclopedia of Genes and Genomes) [19]. Furthermore, CDART (Conserved Domain Architecture Retrieval Tool), NCBI (National Center for Biotechnology Information) and InterPro-Scan are well-known databases that provide information about protein families, domains, and functional locations that can be used to annotate unknown protein sequences [20].

1.2. Enzyme analysis

1.2.1. Enzyme kinetics

The best known equation (Eq.1) for characterizing enzymology under the premise of rapid equilibrium or steady state conditions is the Michaelis-Menten equation, proposed by Leonor Michaelis and Maud Menten [21]. The equation describes substrate binding and subsequent catalysis at the active site. However, the equation has its limitations. For example, the time scale of chemical reactions ranges from microseconds to hundreds of milliseconds. However, in the Michaelis-Menten equation, the enzyme-substrate complex is assumed to form rapidly under conditions where a significant excess of substrate is present, and to reach a quasi-steady state in which d[ES]/dt = 0 [22].

$$v = \frac{v_{max} * \frac{[S]}{K_m}}{1 + \frac{[S]}{K_m}} \qquad \text{(Eq. 1)}$$

In this equation, [S] represents the variable substrate concentration, v_{max} is the maximum

reaction rate, and K_m the substrate concentration at which the reaction rate v is half of v_{max}.

Allosteric control is applied to an enzyme when the binding of a molecule at one binding site changes the enzyme's affinity for its substrate and so affects enzyme activity. In this scenario, the Hill equation (Eq. 2) is applicable [23].

$$v = \frac{v_{max} * \frac{[S]^n}{K_m^n}}{1 + \frac{[S]^n}{K_m^n}} \qquad \text{(Eq. 2)}$$

In this case, n is the Hill coefficient, which indicates the degree of cooperation. When an enzyme has more than one site to which a substrate can bind, positive cooperativity (n> 1) occurs, and the binding of one molecule boosts the rate of binding of other substrates. If n=1 or n< 1, there is no or negative cooperativity, respectively. A partial uncompetitive inhibition model (Eq. 3) is employed to analyze atypical Michaelis-Menten substrate inhibition data [23].

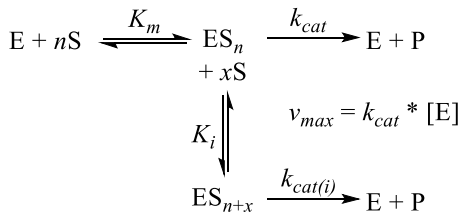
$$v = \frac{v_{max} * \frac{[S]}{K_m} + v_i * \frac{[S] * [S]}{K_m * K_i}}{1 + \frac{[S]}{K_m} + \frac{[S] * [S]}{K_m * K_i}} \qquad \text{(Eq. 3)}$$

To describe the substrate inhibition phenomena, a two-site model is assumed (Scheme 1; n=x=1). The metric v_i represents the reaction velocity in the presence of inhibition, while K_i is the inhibitor concentration required to reduce the maximal rate of the reaction to half of its uninhibited value. The equation assumes sequential substrate molecule binding, which means that the inhibitory site cannot be occupied until the reaction site is filled. When the cooperativity-describing Hill equation and the partial uncompetitive inhibition model are combined, Eq. 4 is obtained [24, 25].

$$v = \frac{v_{max} * \frac{[S]^n}{K_m^n} + v_i * \frac{[S]^n * [S]^x}{K_m^n * K_i^x}}{1 + \frac{[S]^n}{K_m^n} + \frac{[S]^n * [S]^x}{K_m^n * K_i^x}} \qquad \text{(Eq. 4)}$$

The superscript n is a Hill coefficient, and x is another Hill coefficient that allows for the

possibility of cooperative substrate binding in the inhibitory mode [25]. To achieve convergence for Eq. 4, the value of x is set to an integral number that is empirically established to have the best fit (lowest variance).



(Scheme 1)

The catalytic constants k_{cat} and $k_{cat(i)}$ correspond to the maximal reaction rates v_{max} and v_i, respectively.

1.2.2. Protein conformational change

In 1894, the lock and key model was the first model to describe the interaction between enzyme and substrate [26] (Figure 1A). However, when the substrate binds, a conformational shift occurs. An induced fit is required to establish a new conformation before the substrate transition state is altered [27] (Figure 1B). Furthermore, the active site in the enzyme's tertiary structure is designed like a "cleft" for substrate and product diffusion. Active site amino acids interact with substrates depending on a variety of variables such as pH, temperature, and ionic strength. The adjustment of these parameters is critical to the function of the enzyme [28].

In reality, the importance of protein conformational change for catalysis is difficult to demonstrate, and several theories have been proposed, such as the preexisting equilibrium hypothesis and the dynamic population shift model [29] (Figure 1C). The reversible formation of an enzyme-substrate complex marks the start of a catalytic cycle. When products are formed, the enzyme quickly returns to its original state for another catalytic cycle or is transformed to a second conformational state. Analysis of the diversity of protein conformations and studies of the relationships between different structural conformers and protein function, as well as technologies for homology modeling, are a current area of research. [30-32]. In addition, protein apo and holo structure identification and modeling studies are essential for virtual screening and detection of novel protein functions [33, 34].

Binding interactions between proteins and ligands, such as other proteins, carbohydrates, lipids, and tiny molecules normally cause protein conformational changes [35]. Proteins with allosteric properties, first described in 1963, have a more sophisticated regulatory mechanism. The first and most frequently described protein with allosteric interaction is hemoglobin [36]. The Monod-Wyman-Changeux (MWC) model, the Koshland-Nemethy-Filmer (KNF) model, and the dynamic shift model were among the first theories that attempted to explain protein-ligand binding with protein conformational changes [37, 38]. Mutations, machine learning, and energy landscape theory are used to study the physical effects of ligand binding on conformational dynamics and ligand-induced alterations in conformational ensembles [39-43]. In addition, the adaptability and specificity of protein functions and the trade-off between activity and stability in enzymes have been investigated for future applications such as drug discovery [44, 45]. Protein flexibility is currently being used to better understand protein interactions through studies on side chains, protein loops and helices, and dynamic networking [46-49].

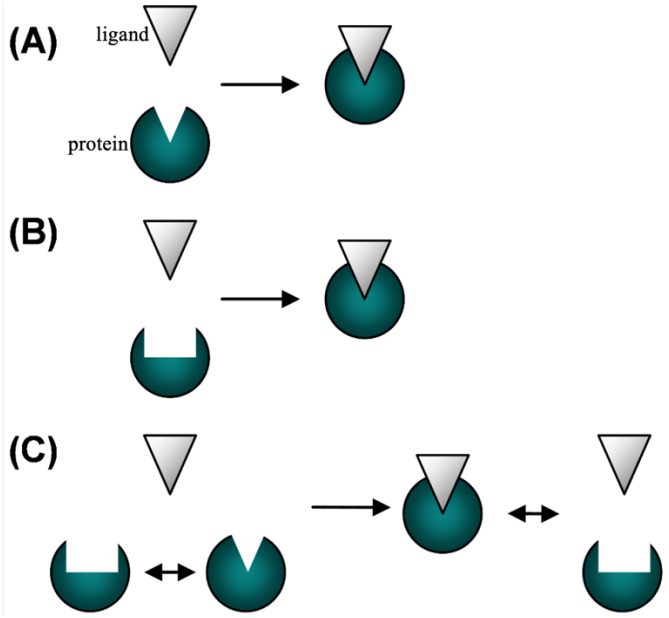


Figure 1. Models of protein binding mechanisms. (A) Lock and key model. (B) Induced-fit model. (C) Pre-existing equilibrium model. Adapted from[50]

1.2.3. Protein analysis by hydrogen/deuterium exchange-mass spectrometry

Hydrogen deuterium exchange mass spectrometry (HDX-MS) is widely used for investigating protein structure and dynamics. HDX-MS can be used to investigate protein dynamics by monitoring solvent exchange over time. HDX-MS is always combined with orthogonal structural, biophysical, and biochemical techniques, as well as molecular dynamic simulations [51].

HDX relies on the inherent ability of hydrogen atoms within protein amides and the backbone to readily exchange with deuterium when introduced into an aqueous deuterated solution. Labile protons, such as those found on primary amines, typically undergo exchange at a considerably faster rate compared to the backbone. In the case of backbone amides, regions of the protein that are exposed and characterized as "dynamic" will experience rapid exchange, while regions that are shielded and considered "rigid" will exhibit a slower exchange rate. In the MS-based HDX approach (Figure 2), the protein solutions are initially equilibrated under room temperature and physiological pH conditions in an H_2O buffer. The exchange process is then initiated by diluting the solution into a D_2O buffer, thereby promoting the kinetics of hydrogen-to-deuterium exchange. This exchange is allowed to progress for various durations before it is halted by acidifying the solution to pH 2.5 and lowering the temperature to 0°C. These quenching conditions ensure the preservation of the deuterium labeling pattern imprinted on the protein backbone for subsequent MS analysis. The quenched protein sample can be directly introduced into a mass spectrometer (utilizing liquid chromatography as the interface) to determine the mass of the intact protein. Alternatively, it can undergo proteolysis with trypsin to generate short peptide segments, which are then subjected to HPLC-MS analysis [52].

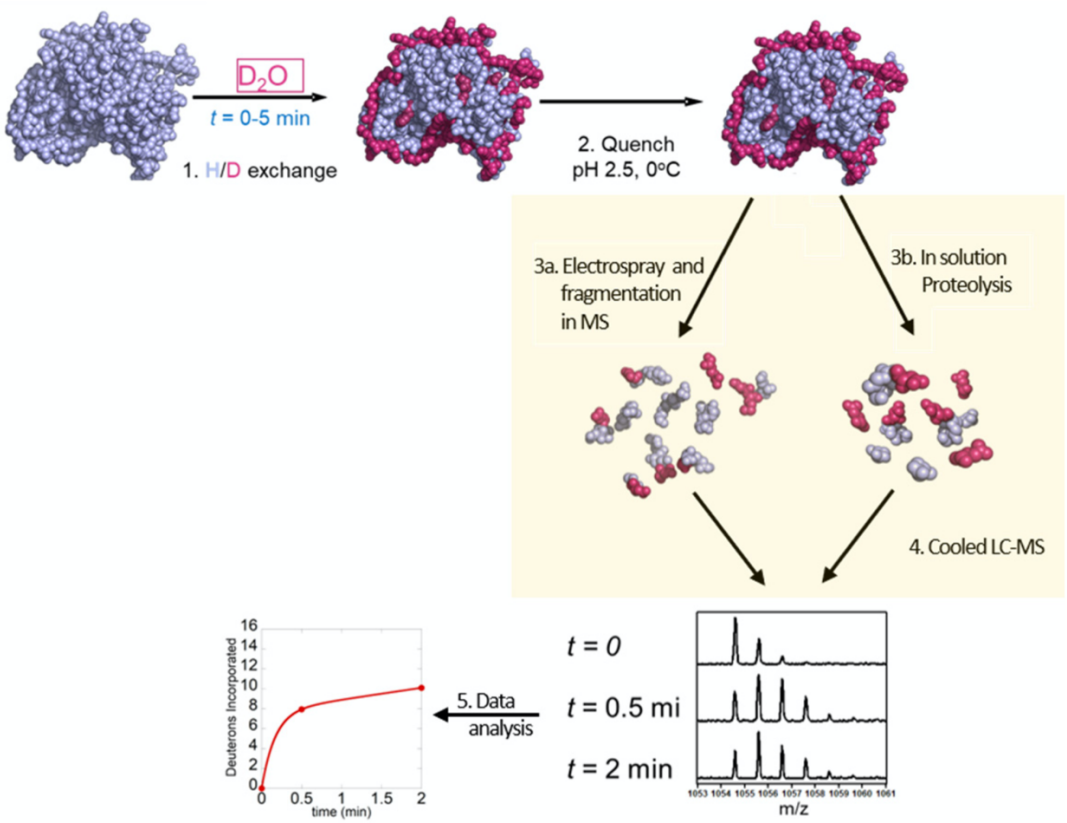


Figure 2. Workflow of a typical HDX-MS experiment. Adapted from[52]

Machine learning ensemble technology has been used to model protein conformation based on HDX-MS data [53]. HDX-MS was utilized to investigate the in solution-dynamics of the GT MshA which catalyzes the first step in the biosynthesis of mycothiol, a low-molecular-weight compound in Actinomycetales, including Mycobacterium tuberculosis. The results provide a better mechanistic understanding of CgMshA, highlighting a conformational change of the flexible β4/α5-loop in the C-terminal domain that interacts with the donor sugar. HDX-MS is useful for the analysis of allosteric enzymes [54, 55].

1.2.4. Homology modelling

Homology modeling is based on the fundamental principle that protein sequences that share close evolutionary relationships tend to adopt analogous three-dimensional structures. Consequently, it becomes feasible to generate a homology model for a protein by employing related protein structure(s) as templates for the modeling process. Sequence alignment and

loop building are the main difficulties in homology modelling since there is no human input to correct errors.

There are many sequence-structure alignment and homology modelling programs, like 3D-Coffee [56], SAlign [57], Nest [58], Modeller [59], and Swiss-Model [60]. Evaluating the different homology modelling programs, on average Modeller performs marginally better in overall modelling than the others tested. However, on average Nest produces the best loops with an 8% improvement of the root-mean-square deviation (RMSD) compared to the loops generated by Builder. The well-known machine learning approach-AlphaFold incorporates physical and biological knowledge about protein structure [61]. DEDAL (deep embedding and differentiable alignment) exhibits a notable enhancement, improving the accuracy of the alignment by a factor of two to three compared to existing methods, particularly on remote homologs [62]. Furthermore, it excels in distinguishing remote homologs from sequences that are evolutionarily unrelated. These advancements open the door to potential improvements in various downstream tasks that heavily depend on sequence alignment, particularly in the realms of structural and functional genomics [62].

1.2.5. Allostery

Allosteric regulation typically involves changes in binding affinity resulting from the binding of an effector molecule at a distinct site. Allosteric control allows specific regulation without directly targeting the active site. Allosteric sites are distal to the active site but coupled via a dynamic network of inter-atomic interactions between residues in the protein [63]. Through site-directed mutagenesis of crucial residues, it becomes possible to selectively influence the equilibrium between different states, as exemplified by the mutation of a critical thrombomodulin-binding residue in thrombin from Glu192 to Gln. This mutation demonstrates how targeted alterations can favour one state over another in an allosteric system [64].

X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy are used to observe protein structures, but they are time-consuming, expensive and not sufficient to determine the structural and dynamic changes occurring in complex allosteric system [65-67]. Computational methods are explored to predict allostery, allosteric communication pathways and quantitative assessment of protein-energy dynamics [68-73]. G-coupled receptors and thrombin are typical allosteric proteins [64, 74]. Thrombin was observed to be an allosteric protein in 1980, followed by discovering of a cofactor called thrombomodulin.

1.2.6. Protein engineering

Chemistry is essential for understanding the structures and properties of substances at the atomic and molecular levels. Merging with biology improves scientific innovation by exploring its endless possibilities. Biocatalysts are being used in an increasing number of fields, such as the health, energy and environmental sectors for the production of valuable substances due to advances in protein technology and environmental and economic requirements [75]. Nowadays, protein engineering focuses on the selection of new functional enzymes and the modification of enzyme thermostability, stability towards proteases, and enzyme activity [76]. The value k_{cat} represents the rate of reaction at saturating substrate concentration and k_{cat}/K_m is the second-order rate constant describing the reaction rate at negligible substrate concentration. The k_{cat}/K_m values determine the specificity of the enzyme for each substrate. The higher this value the more specific the enzyme is for that substrate. This is because a high value of k_{cat} and a low value of K_m are expected for the best substrates. If k_{cat}/K_m approaches the diffusion limit (~ 10^8–10^9 M^{-1} s^{-1}), the enzyme cannot catalyze the reaction any better and is said to have reached 'catalytic perfection'[77]. Proteins are modified to improve the affinity between the enzyme and the substrate (K_m) through interactions with amino acids and substrate and thus to increase enzyme activity k_{cat} [78-8078-8078-8078-8079-81]. The first enzyme was found in 1833 [81, 82], and the "lock and key model" of enzymes and substrates

was proposed in 1894 [26] (Figure 3). Protein and DNA sequencing technology was developed around 1977 [83], followed by site-directed mutagenesis in 1978. Prior to 1991, only a few modifications of proteins were routinely performed to expand enzyme libraries. Directed evolution of enzymes was developed by Frances Arnold in 1990, and with the advancement of protein structure identification technologies, rapid and effective enzyme engineering approaches have been described, such as the combinatorial active site saturation test/iterative saturation mutagenesis (CAST/ISM) strategy in 2006 [84, 85]. With the rapid progress of computer science in the twenty-first century, computational protein engineering was launched. Computational protein design [86], protein-structure prediction algorithm Alphafold [61], and the artificial intelligence ChatGPT (https://www.technologynetworks.com/informatics/news/chatgpt-for-proteins-predicts-structure-371718) are being used to an ever-increasing extent for protein structure modeling and function prediction. Two broad approaches have been used for enzyme engineering, namely, rational design and evolutional design [87] (Figure 3).

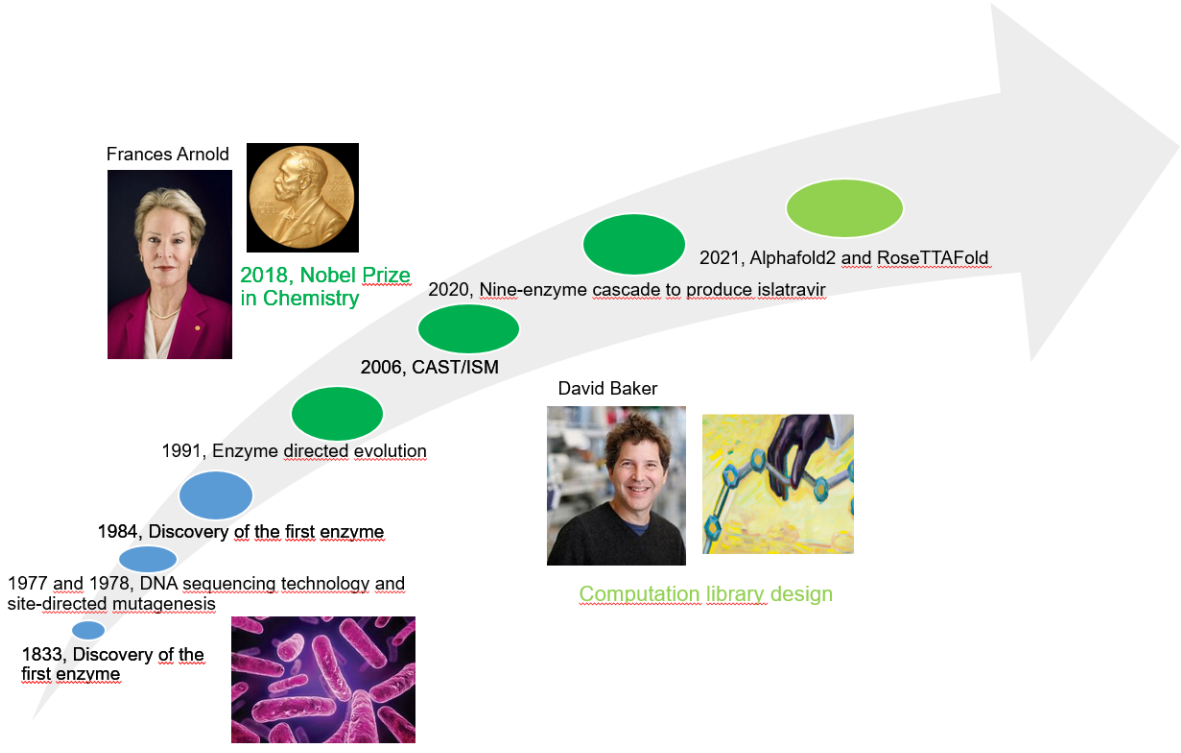


Figure 3. The history about enzyme engineering.

Protein structures can be determined by various technologies, which are listed in Supplementary Table 1 with explanations. Besides, websites about enzymes and enzyme engineering methods are summarized in the Supplementary Table 2 and 3.

1.3. Glycosyltransferase

1.3.1. Glycosylation

Glycosylation is the key to changing the bioactivity of small molecules found in nature [88]. Glycosylation changes the polarity, volatility, solubility and reactivity of the parent compounds [89]. It plays an important role in plant cells, for example in modifying secondary metabolites, thereby altering their storage and intercellular transport. Thus, some secondary metabolites accumulate in plant vacuoles as glycosides [90]. Due to the structural and functional diversity of natural products, glycosylation is intensively studied in drug discovery and development [91]. Glycosyltransferases (GTs) catalyse the relocation of a sugar moiety from an activated nucleotide diphosphate sugar donor, usually UDP-glucose to acceptor molecules. UDP-sugars (typically UDP-D-glucose, UDP-D-glucuronic acid, UDP-D-xylose, UDP-L-rhamnose, UDP-L-arabinose, and UDP-D-galactose) act as sugar donors in plants. Their sugar residues are transferred to acceptors by glycosyltransferases, resulting in the formation of *O*-, *S*-, *N*-, and *C*-glycosides and sugar esters [92]. Many of these glycosyltransferases are promiscuous, i.e. they transfer the sugar to different acceptors which contributes to the diversity of secondary metabolism [93].

At present, there are more than 45,000 putative Family 1 glycosyltransferases in the protein databank CAZy (http://www.cazy.org/GT1_structure.html), of which 159 3D structure are available in PDB (https://www.rcsb.org/) but the functions of most of these GTs have not been verified [94]. Ultimately, these enzymes could be used as tools to generate new glycoforms of natural products by catalysing "unnatural" coupling reactions through the exploitation of the broad substrate specificity and the alteration of GT specificity through genetic engineering. GT enzymes decorate natural products with a remarkable array of sugar moieties, including D/L-configured, amino-, deoxy- , and methoxy sugars [92].

1.3.2. Structures of glycosyltransferases

GTs have been grouped into 116 families (http://www.cazy.org/). Plant GTs that glycosylate natural products are grouped into family 1. They show a horseshoe-like structure with N- and C-terminal domains connected directly via an interdomain linker. Additionally, two α-helices at the C-terminus stabilize the conformation of the GTs. The plant CAZy Family 1 GT substrates include terpenoids, alkaloids, cyanohydrins and thiohydroximic acids as well as flavonoids, isoflavonoids and phenylpropanoids. They have a conserved so-called plant secondary product glycosyltransferase (PSPG) motif towards the C-terminus in common (Figure 4). A three-dimensional model has revealed a direct interaction between the uracil residue of the co-substrate UDP–glucose and the highly conserved first amino acid of the PSPG motif (W) and an interaction of the conserved sequence HCGWNS and the diphosphate group of UDP [95]. Correspondingly, the discrimination between the donor UDP–glucose and UDP–galactose is probably dictated by the last amino acid of the PSPG box [92].

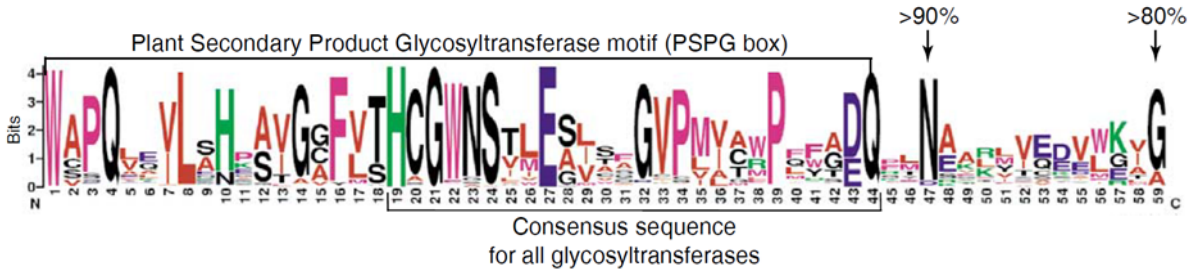


Figure 4. Plant secondary product glycosyltransferase (PSPG) motif. Adapted from [68].

In addition to small molecules, proteins can also be glycosylated. Glycans can be covalently attached to the amide nitrogen of Asn residues (*N*-glycosylation), to the hydroxyl oxygen of, typically, Ser or Thr residues (*O*-glycosylation), and, in rare cases, to the indole C2 carbon of Trp through a C–C linkage [95]. In eukaryotes and archaea, the Asn residues are located in an Asn-X-Ser/Thr motif (where X represents any amino acid except Pro), whereas the bacterial system requires an extended *N*-glycosylation consensus sequence that contains an Asp or Glu at the –2 position (Asp/Glu-X_1 -Asn-X_2-Ser/Thr, where X_1 and X_2 represent any amino acid except Pro) [96].

The Protein Data Bank (PDB; https://www.rcsb.org/) contains over 150 GT crystal structures

classified into the GT-A, GT-B, and GT-C fold subfamilies. GT-A enzymes comprise two abutting β/ α /β Rossman- like domains and are generally divalent metal ion dependent. The metal ion is coordinated by a highly conserved DXD motif within the GT active site and aids leaving group departure by stabilizing the charged phosphate groups in the nucleotide sugar donor. GT-B enzymes consist of two β/ α /β Rossman-like domains that face each other. The active site lies in the cleft between the two domains. GT-B enzymes are generally metal ion independent, with active site residues acting to aid leaving group departure. The active site of fold B UGTs contains a catalytically active His, contained in the N-terminal domain [89]. The His residue acts as a general base by removing a proton from the acceptor molecule to produce an oxyanion nucleophile that combines with the UDP-sugar, allowing for a direct displacement S_N2-like process [94]. In most cases, an Asp can help deprotonate acceptor molecules by generating an electron transfer chain. More recently a third (GT-D) and forth (GT-E) fold of GT enzymes, was identified [97]. Structural representatives of GT-C enzymes are hydrophobic integral membrane proteins, and perhaps not surprisingly all GT-C enzymes characterised to date use lipid phosphate-linked sugar donors [98].

1.3.3. Mechanisms of glycosyltransferases

Sugar conjugation results in increased stability, polarity and water solubility, and has a major impact on biological activity or toxicity. Understanding the mechanisms of GTs is beneficial for engineering the active site. Two main groups of GTs have been discovered, including retaining GTs, that transfer sugar residues with retention of the anomeric configuration at C1, and inverting GTs, that relocate sugar moieties with inversion of the anomeric configuration. The classification relies on the configuration of the anomeric functional group of the glycosyl donor and of the resulting glycoconjugate product (Figure 5). Glycosylation reactions are regio- and stereo-specific. Hybrid quantum mechanics/molecular mechanics (QM/MM) methods and associated experimental data have been combined to explain glycosylation mechanism. These methods generally model enzymatic reactions through calculation of the

electronic structure of the active site [99].

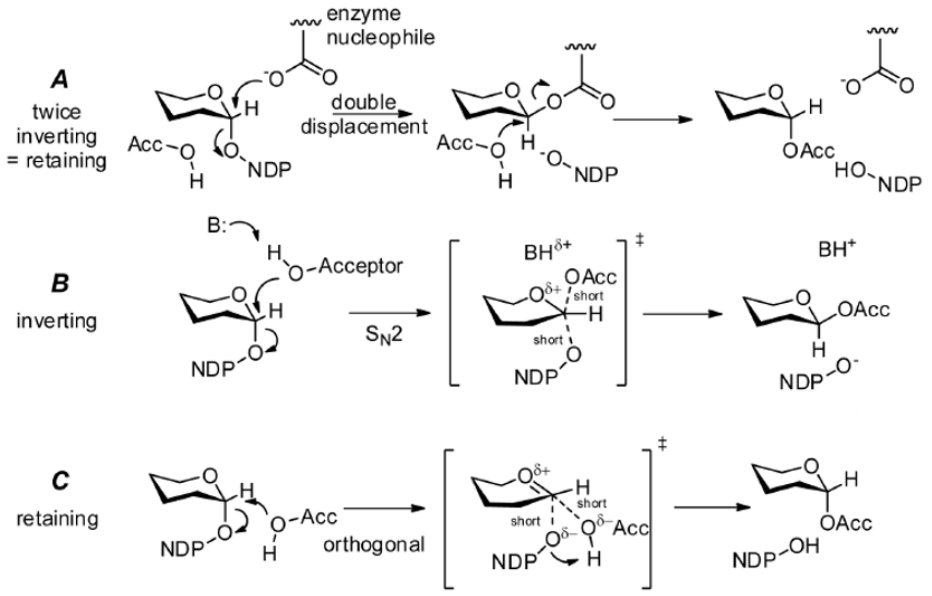


Figure 5. Mechanisms of glycosyltransferases NDP nucleotide diphosphate, Acc-OH acceptor substrate. Adapted from [100]

The mechanism of inverting GTs has been explained, among others by a direct displacement S_N2-like mechanism [101, 102] (Figure 5B). An oxocarbenium-ion transition state (TS) forms with the help of a catalytic base usually provided by an active-site side chain (such as Asp, Glu or His) of the GTs. The catalytic base abstracts a proton from an OH-group of the acceptor, facilitating nucleophilic attack at the sugar anomeric C1, forming a glycosidic bond between the sugar donor and the acceptor with the inversion of the configuration at C1. This suggests that the nucleophilic addition and glycosidic bond cleavage occur almost simultaneous, accompanied by proton transfer. During the formation of the TS, the anomeric C1 atom moves towards the nucleophilic oxygen accompanied by the rotation of the diphosphate group, which promotes the glycosylation reaction. Structural studies showed that the catalytic bases are usually located near the acceptor OH-group. Classically, the catalytic His residue in some GT-B members can form a hydrogen bond with an Asp residue, balancing the charge on the His after proton abstraction. In general, the negative charge on

the phosphate group can be stabilized by a divalent metal ion (generally Mn^{2+} or Mg^{2+}) for most GT-A proteins or positive amino acids/helix dipole for GT-B proteins. The S_N2-like mechanism[103103103103104], the candidate catalytic residue, and the catalytic role of beta-phosphate have been proposed several times [99, 103]. Besides, alpha-phosphate of UDP-GlcNAc serves as the proton acceptor with the help of an essential Lys residue during the glycosylation catalysed by human O-GlcNAc transferase (OGT) [104].

The double-displacement mechanism involving the formation of a covalent sugar–enzyme intermediate has been proposed for GT6 members, which have a carboxylic residue in the suitable position as a candidate nucleophile [105, 106] (Figure 5A). This mechanism proposes that the sugar moiety first binds to an appropriate part of the GTs with an inverting configuration at sugar C1. Then it is transferred to the acceptor with the C1 atom reverting to its original configuration. For most retaining GTs (Figure 5C), the catalytic base could not be identified in the catalytic site, thus an alternative internal return (S_Ni-like) mechanism was proposed. In this mechanism, the acceptor OH-group nucleophile attacks the sugar anomeric C1 atom on the same side that the sugar group leaves the donor. The reaction forms an oxocarbenium ion TS that is shielded on one face by the GTs, consequently protecting it against nucleophilic attack from the opposite face and resulting in retention of C1. The controversy over the retaining mechanism has persisted for years. Both the experimental and theoretical studies suggest that the leaving phosphate group could function as the catalytic base to deprotonate the acceptor OH-group [107, 108].

1.3.4. UDP-glucose hydrolase activity

In a conventional GT reaction, the glycoside product is released when the sugar has been transferred from the donor substrate to the acceptor substrate. However, it has been observed that GTs in the absence of the acceptor substrate exhibit background hydrolysis of the donor substrate, which can be viewed as an enzymatically catalyzed transfer of the sugar moiety to a water molecule [109, 110] (Figure 6).

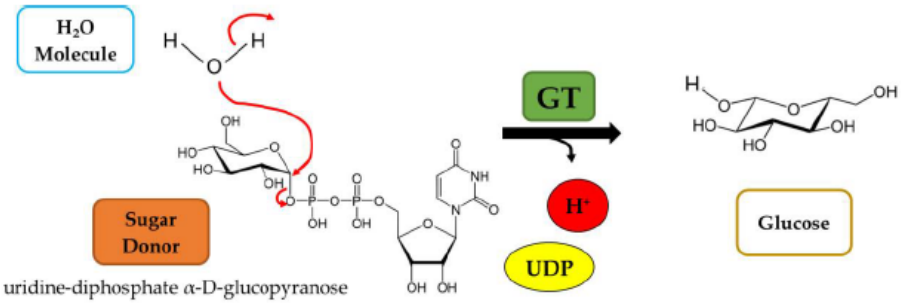


Figure 6. Reaction mechanism of GTs utilizing a water molecule as acceptor substrate. UGT family-1 GT-B inverting reaction mechanism employing direct displacement S_N2-like reaction with a water molecule resulting in UDP and H^+ by-products and no glycosidic product. Adopted from the dissertation of Kate McGraphery (https://www.mls.ls.tum.de/bina/forschung/dissertationen/).

1.3.5. Quantification of the glycosyltransferase reaction

Liquid chromatography coupled with mass spectrometry (LC-MS) and radiolabelled sugar donors combined with liquid scintillation counting were used to detect glycoside products [111]. The UDP by-product can be quantified with the UDP-Glo GT-assay, Phosphate GT Activity assay, pH-sensitive GT-assay, and UDP2-TR-FRET assay [91] (Figure 7). However, the pH-sensitive assay and the UDP2-TR-FRET assay are incomparable and unsuitable for high-throughput plant GT-1 family UGT screening.

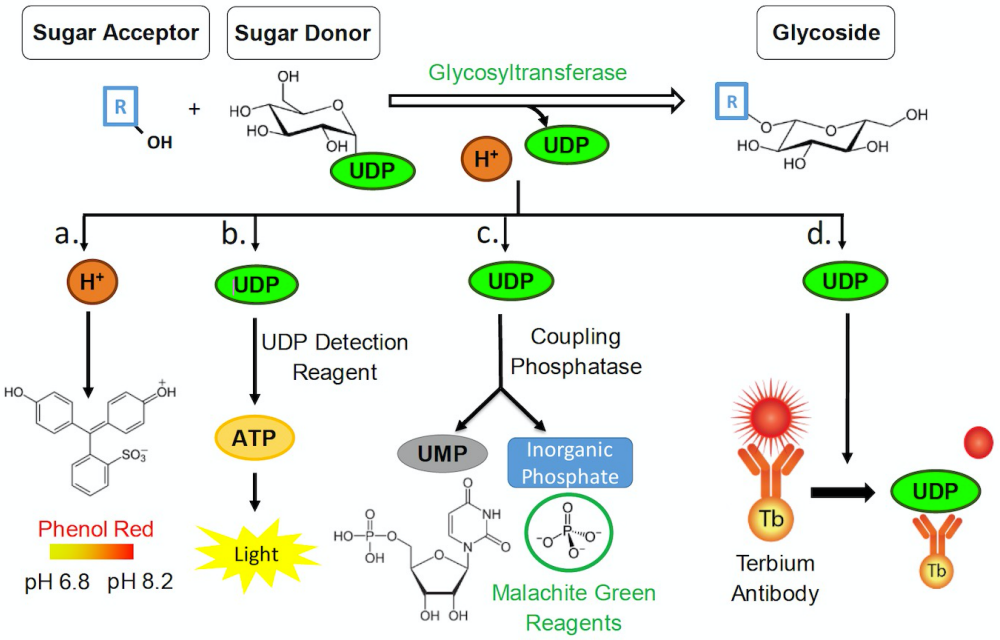


Figure 7. Glycosyltransferase reaction and four methods for the quantification of by-products. (a) Colorimetric pH-sensitive assay (b) UDP-Glo™ assay—UDP is converted to ATP eventually generating light in a luciferase

reaction. (c) Phosphate GT assay—Phosphate is released from UDP and detected by malachite green reagents. (d) UDP2 TR-FRET immunoassay—a competitive immunoassay for UDP detection. Adapted from [112].

1.3.6. Glycosyltransferase engineering

Enzymatic catalysis is increasingly used in industrial processes to manufacture chemicals, pharmaceuticals, and materials for human society [113]. Strategies, such as chemical modification, immobilization, site-directed mutagenesis, directed evolution of enzymes, artificial metalloenzymes, and computational design, are used to adapt enzymes for use as catalysts for industrial processes [114]. Editing enzymes to easily integrate biocatalysis with chemocatalysis is a potential way to apply enzymes in industry. Enzymes offer a unique strategy for catalyst optimization through protein engineering. In the last two decades, various studies have been conducted on the modification of natural products by glycosylation and numerous techniques have been applied, including enzymatic glycosylation using whole cells or single GTs, in vitro glycorandomization, and combinatorial biosynthesis in line with the idea of synthetic biology [89, 93]. In plants, GTs can glycosylate almost all major classes of secondary metabolites, such as phenylpropanoids, alkaloids, terpenoids, and polyketides [115]. Structure-guided alanine scanning and saturation mutations of the glycosyltransferase Yjic from *Bacillus subtilis* were performed and mutant M315F was found to efficiently synthesize ginsenoside Rh2 (~99%) and block the further glycosylation of C12–OH [116]. Structure-guided mutagenesis was also employed to modify the catalytic specificity of a *C*-glycosyltransferase from *Trollius chinensis* (TcCGT1), allowing for changes in *C*- and *O*-glycosylation. TcCGT1 possesses the capability to perform 8-*C*-glycosylation on 36 distinct flavonoids and *O*-glycosylation on a variety of phenolic compounds. This broad substrate range is attributed to the ample binding pocket, a characteristic elucidated through its crystal structure when bound to uridine diphosphate. The orientation of the substrate within this binding pocket determines whether *C*- or *O*-glycosylation activity is exhibited. By performing site-directed mutagenesis on two specific residues (I94E and G284K), it was possible to shift the enzyme's activity from *C*-glycosylation to *O*-glycosylation. This demonstrates how precise structural modifications can be used to fine-tune the enzymatic

specificity of TcCGT1, allowing for targeted alterations in its glycosylation capabilities [117]. Fortunately, genome sequencing and genome editing also contribute a lot.

The GT families are grouped hierarchically from their 3D structures (fold GT-A – GT-E) to their mechanism of reaction (inverting or retaining GTs) [118], as we mentioned above. As more research is done on the particular spatial arrangement of the active site, protein dynamics, conformational changes, and plasticity of GTs during substrate recognition and catalysis, the enzyme transition state complex will become less mysterious [119]. The first determination of the three-dimensional structure of a family 1 plant GT (VvGT1), including its "Michaelis" complex with a UDP-glucose-derived donor and the acceptor kaempferol and in complex with UDP and quercetin, and mutational analysis contributed decisively to uncovering the basis for the modification of plant natural products [120]. The structure-based rational design and directed evolution of GTs involved in the biosynthesis of glycosylated plant natural products have been recently summarized [91].

1.4. The glycosyltransferase UGT72AY1 from *Nicotiana benthamiana*

1.4.1. The substrate screening of NbUGT72AY1

Because tobacco plants produce high amounts of bioactive apocarotenoids in their leaves, most of which accumulate in glycosidic form, a publicly available transcriptome of *Nicotiana benthamina* was searched for glycosyltransferase genes that are particularly highly expressed in the leaves [96, 121]. Subsequently, total RNA was isolated from *N. benthamiana* leaves and the gene coding for *NbUGT72AY1* (Nbv6.1trP2283, **http://benthgenome.qut.edu.au/**) was amplified, among others, from cDNA. For recombinant His-tagged protein production, *Escherichia coli BL21(DE3)pLysS* cells were used [122, 123], and NbUGT72AY1 was identified as a promiscuous enzyme capable of glycosylating a wide range of small molecules, including phenolics (monolignols and hydroxycoumarins), isoprenoids (monoterpenoids and apocarotenoids), and aliphatic alcohols (hexanols and octanols), as shown by LC–MS analysis [122,121] (Figure 8).

Substrates	UGT 71AJ1	UGT 72AX1	UGT 72AY1	UGT 72B35	UGT 72B34	UGT 73A24	UGT 73A25	UGT 85A73	UGT 85A74	UGT 709Q1
Retinol	0	0	0	0	0	0	0	0	0	0
Farnesol	0	0	21	0	0	100	37	5	0	0
α-Ionol	2	0	51	2	2	60	100	16	0	0
β-Ionol	2	3	49	2	1	76	100	24	0	0
2-Phenylethanol	0	0	56	0	0	59	26	100	0	0
Furanmethanethiol	0	0	0	0	0	0	0	0	0	0
Carvacrol	31	33	50	48	58	100	42	2	0	0
Geraniol	1	2	23	1	11	81	47	100	0	0
Hydroquinone	13	22	55	81	100	0	0	0	0	0
Kaempferol	87	18	92	100	71	24	37	5	0	0
Scopoletin	8	32	72	61	79	100	82	4	0	0
Mandelonitrile	0	4	0	0	0	0	0	100	0	0
3-*cis*-Hexenol	0	17	41	0	35	23	14	100	0	0
Perillyl alcohol	0	0	100	3	16	82	74	90	0	0
1-Octen-3-ol	0	25	55	2	13	56	33	100	0	0
Benzyl alcohol	10	31	32	44	100	0	0	83	0	0
Mandelic acid	0	0	0	0	0	0	0	0	0	0
α-Bisabolol	0	0	0	0	0	0	0	0	0	0
Lavandulol	2	19	33	1	16	100	73	30	0	0
Borneol	0	0	0	0	0	79	75	100	0	0
Carveol	14	29	46	5	26	25	11	100	0	0
Fenchyl alcohol	0	0	0	0	0	0	0	0	0	0
4-Carvomenthenol	0	0	0	0	0	0	0	100	0	0
Tyrosol	0	0	100	10	54	33	16	16	0	0
Myrtenol	1	5	29	3	32	37	58	100	0	0
3-Oxo-α-ionol	0	0	100	0	0	46	27	0	0	0
Converted substrates	11	13	18	14	15	17	17	19	0	0

Figure 8. Screening of ten UGTs from *N. benthamiana* for potential substrates by LC-MS. The colour code shows non-reactive substrate enzyme combinations in red (0%; no product detected by LC-MS) and increasing reactivities from white to green colour (maximum activity in dark green corresponding to 100%). Adapted from [96].

Given the limitation of the LC–MS data, which hindered a direct quantitative comparison of NbUGT72AY1's enzymatic activity across different substrates due to varying ion yields of the products, we turned to the UDP-Glo™ assay as a solution. This assay assesses the formation of UDP during glucoside formation, providing a means for a direct comparison of catalytic activity (Figure 9a). In this assay, the release of UDP was quantified in mixtures containing the enzyme together with donor and acceptor substrates. (Figure 9b). To establish a baseline, the amount of UDP generated in a control sample containing only the enzyme and the donor substrate (referred to as "no-acceptor control") was subtracted from the values obtained in the presence of both substrates. The results from the UDP-Glo™ assay confirmed that scopoletin and carvacrol, among others, are indeed substrates of NbUGT72AY1. However, intriguingly, the assay revealed negative values for alpha-ionol, beta-ionol, farnesol, and retinol after subtracting the control values. This unexpected finding was

explained by the inherent UDP-glucose glucohydrolase activity of GTs (Figure 9b), which undergoes inhibition in the presence of the terpenoids. This insight provides valuable information about the enzyme's dual functionality and how specific compounds can modulate its activities in distinct ways[111].

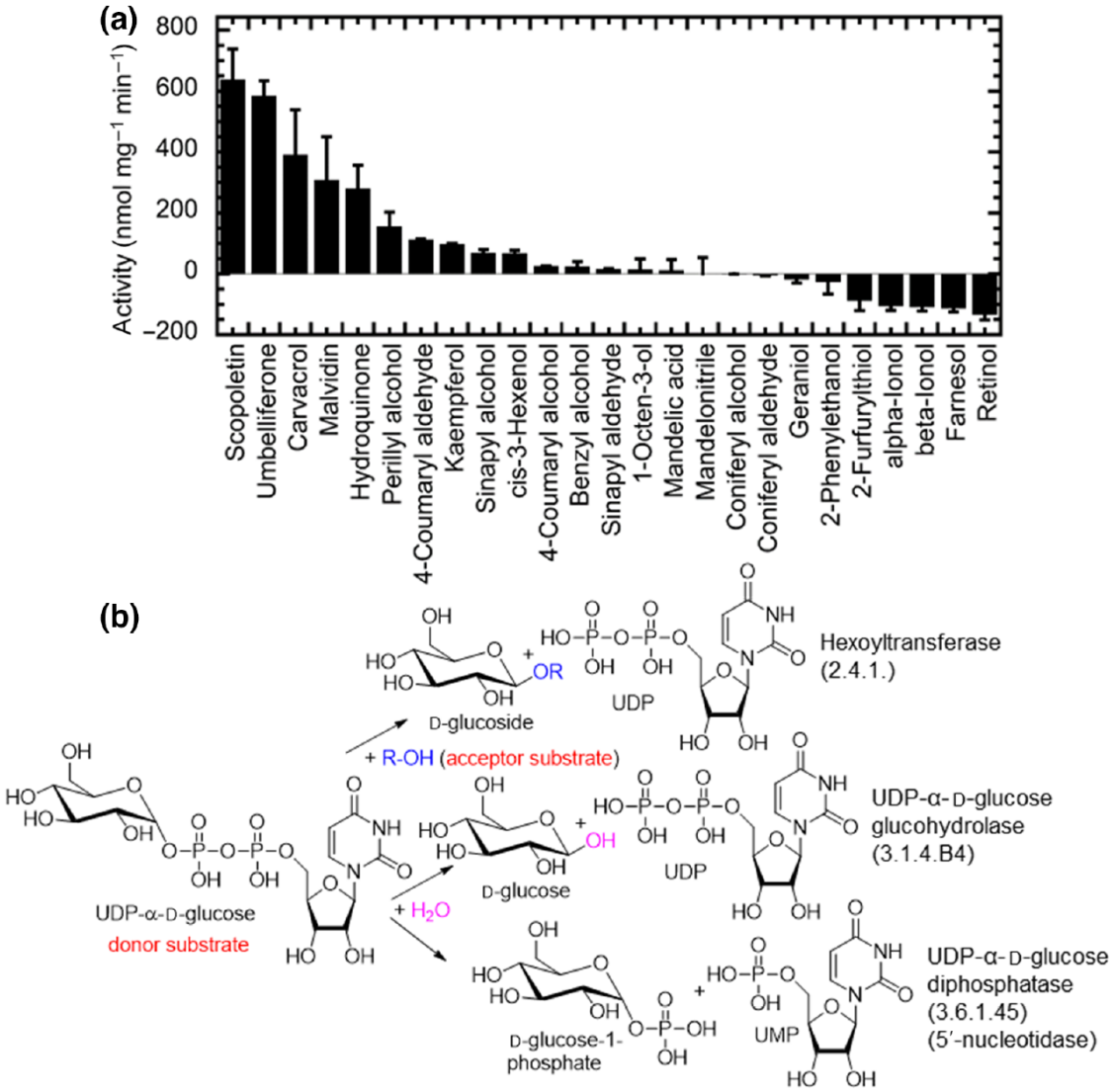


Figure 9. NbUGT72AY1 shows uridine 5′-diphosphate α-D-glucopyranose (UDP-glucose) glucohydrolase and glucosyltransferase activity. (a) Substrate screening of NbUGT72AY1 using the UDP-Glo™ assay. (b) Reaction catalyzed by Hexosyltransferase, UDP-α-D-glucose glucohydrolase, and UDP-α-D-glucose diphosphatase. Adapted from [111]

1.4.2. Substrate inhibition of glucosyltransferase NbUGT72AY1

UDP-Glo™ assay was used to measure NbUGT72AY1 kinetic parameters with

scopoletin. Substrate inhibition was observed with scopoletin, but also with structurally related acceptors such as umbelliferone, vanillin and carvacrol (Figure 10). The mechanisms of substrate inhibition appear to be very complex. For example, several (at least two) binding sites within the enzyme (allosteric binding), the formation of a ternary dead-end enzyme complex and/or the changes in enzyme conformation induced by the ligand could possibly lead to substrate inhibition [124].

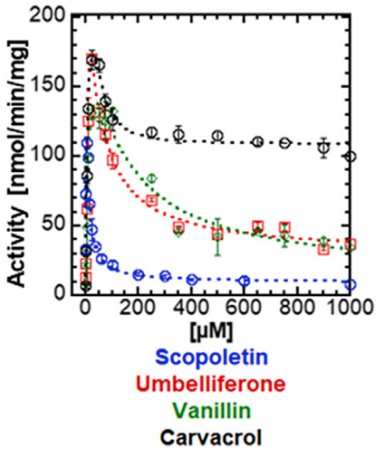


Figure 10. Kinetics for NbUGT72AY1. Adapted from[125]

1.4.3. Inhibition/activation of glucosyltransferase NbUGT72AY1

Since the production of alpha-ionol and beta-ionol β-D-glucosides by NbUGT72AY1 had already been confirmed by LC–MS analysis [122] (Figure 8), our results allowed us to conclude that the enzyme exhibits a high UDP-forming activity even in the absence of an acceptor molecule (Figure 9). This observation explained the divergent results obtained from both LC–MS and UDP-Glo™ analyses. Therefore, the impact of varying concentrations of apocarotenoid effectors (alpha- and beta-ionol, as well as alpha- and beta-ionone) on the UDP-glucose glucohydrolase activity (Figure 9b) and the glucosyltransferase activity of NbUGT72AY1 towards its preferred substrate, scopoletin were investigated (Figure 9b). The findings revealed that apocarotenoids enhance the catalytic activity of the UGT towards the hydroxycoumarin (scopoletin) (Figure 11). However, the glucohydrolase activity was inhibited in the absence of scopoletin. These results shed light on the intricate regulatory

mechanisms at play and their influence on the enzyme's activities in response to different substrate and effector concentrations (Figure 11).

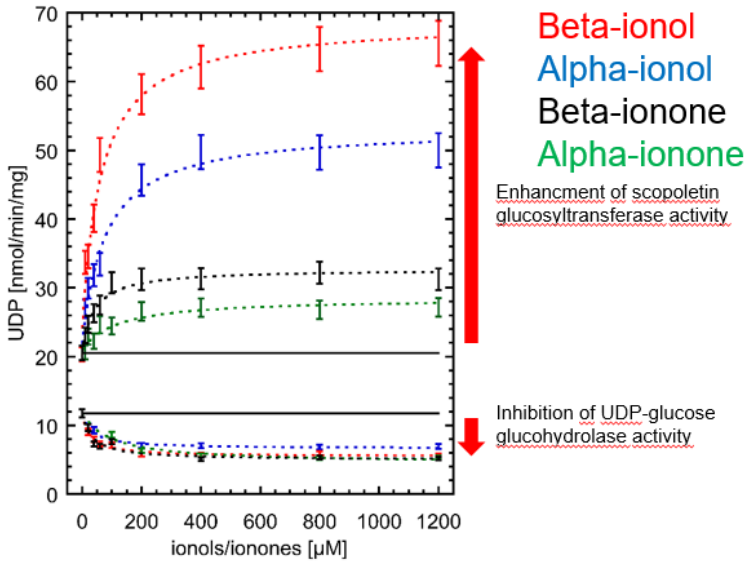


Figure 11. Effect of apocarotenoids on the glucosyltransferase and UDP-glucose glucohydrolase activity of NbUGT72AY1. Increase of the UDP-glucose: scopoletin glucosyltransferase activity of NbUGT72AY1 by ionols and ionones as a function of ionol/ionone concentration (upper part). Inhibition of the UDP-glucose glucohydrolase activity by ionols and ionones as a function of ionol/ionone concentration (lower part). Adapted from [111].

2. Aims of this study

Exploring the mysteries of protein-ligand interactions can help improve future protein engineering research. Preceding studies showed that NbUGT72AY1 is an unusual enzyme because it exhibits broad substrate tolerance [96,121] hydrolyzes UDP-glucose considerably in the absence of acceptor substrates [111] and is strongly substrate-inhibited by hydroxycoumarins [111]. However, this inhibition can be partially reversed by apocarotenoids [111]. Thus, NbUGT72AY1 appears to be an excellent model for studying critical amino acids and conformational changes during substrate inhibition. It may exhibit allosteric modulation with apocarotenoids. The following experiments were planned to test the hypothesis.

- perform homology modeling to obtain the most realistic 3D structure possible

- perform HDX-MS to locate the ligand binding site

- search for related sequences and compare enzymatic activity with that of NbUGT72AY1

- select candidate amino acids probably involved in ligand binding from HDX-MS results and amino acid alignment with related sequences.

- measure enzymes activities of mutants by UDP-Glo™ assay and LC-MS to confirm the proposed ligand binding site

- perform *in silico* docking on NbUGT72AY1 with scopoletin and additional ligands and analyze conformational changes.

- analyze the putative allosteric site using effectors and mutations.

- test other substrates for substrate inhibition

3. Subfunctionalization of a monolignol to a phytoalexin glucosyltransferase is accompanied by substrate inhibition

Abstract

The promiscuous monolignol/phytoalexin glucosyltransferase NbUGT72AY1 from *Nicotiana benthamiana* shows substrate inhibition with scopoletin at 4 µM, whereas the highly homologous StUGT72AY2 from *Solanum tuberosum* is only inhibited at a 50-fold higher concentration. In this research, we investigated the mechanism of substrate inhibition from the following perspectives:

Structural insights by HDX-MS: Hydrogen/Deuterium Exchange Mass Spectrometry (HDX-MS) is a powerful tool for deciphering the structural dynamics of enzymes. By identifying specific amino acids involved in ligand binding, the study unveiled critical regions that respond to scopoletin binding. This information aids in understanding how ligand-induced conformational changes impact the enzyme's function.

Homology modelling by in-silico docking: Computational simulations were used to elucidate conformational changes occurring during the interaction between NbUGT72AY1 and scopoletin. In particular, the movement of amino acids 305-325 are identified as a critical loop responsible for the open and close states of the protein. Furthermore, the rotation of W350 is shown to be a key factor that guides the donor substrate to the active site and facilitates the export of the reaction product UDP.

Reciprocal mutation analyses: The importance of amino acid F87I, positioned in the active site, and a distinct sequence segment, located outside the catalytic center, for substrate inhibition was validated by reciprocal mutation analyses. These results emphasize the complexity of the regulation of substrate inhibition, which affects both the catalytic and non-catalytic regions of the enzyme.

Subfunctionalization and extended promiscuity: The subfunctionalization within a monolignol glucosyltransferase in *Nicotiana* observed in this study underscores the evolutionary importance of enzyme diversification. This diversification can lead to extended

promiscuity and substrate inhibition patterns, reflecting the adaptability of these enzymes to evolving biochemical demands.

In summary, this comprehensive study sheds light on the intricate interplay between promiscuity, substrate inhibition, and structural dynamics within the NbUGT72AY1. These findings have broader implications for understanding enzyme evolution, substrate recognition, and the regulation of catalytic activities in various biochemical pathways.

This work has been published:

Title: Subfunctionalization of a monolignol to a phytoalexin glucosyltransferase is accompanied by substrate inhibition

Authors: **Jieren Liao**, Guangxin Sun, Elisabeth Kurze, Wieland Steinchen, Timothy D. Hoffmann, Chuankui Song, Zhiwei Zou, Thomas Hoffmann, Wilfried Schwab

First Published: 24 December 2022

Journal: Plant Communications

Publisher: Cell Press

DOI: 10.1016/j.xplc.2022.100506.

Link: https://www.sciencedirect.com/science/article/pii/S2590346222003534

The journal "Plant Communications" has information regarding 'author rights' to use their own work in support of their thesis.

J.L., G.S., E.K., T.H., and W.G.S. designed experiments and interpreted data; **J.L**., G.S., E.K., T.H., T.D.H., C.S., and W.G.S. wrote the manuscript; **J.L**., G.S., and E.K. performed biochemical analyses and analyzed the proteins; **J.L**. and E.K. generated mutant proteins; **J.L**., G.S., E.K., and T.H. performed LC-MS analyses. W.S. performed HDX-MS analyses; W.G.S. performed in silico analyses; C.S., T.D.H., and W.S. contributed to materials, instruments, data analysis, and the manuscript. All authors approved the final manuscript.

Plant Communications
Research article

CellPress
Partner Journal

Subfunctionalization of a monolignol to a phytoalexin glucosyltransferase is accompanied by substrate inhibition

Jieren Liao[1,5], Guangxin Sun[1,4,5], Elisabeth Kurze[1,5], Wieland Steinchen[2], Timothy D. Hoffmann[1], Chuankui Song[3], Zhiwei Zou[1], Thomas Hoffmann[1] and Wilfried G. Schwab[1,*]

[1]Biotechnology of Natural Products, Technische Universität München, Liesel-Beckmann-Str. 1, 85354 Freising, Germany

[2]Center for Synthetic Microbiology (SYNMIKRO) & Faculty of Chemistry, Philipps-University Marburg, Karl-von-Frisch-Straße 14, 35043 Marburg, Germany

[3]State Key Laboratory of Tea Plant Biology and Utilization, International Joint Laboratory on Tea Chemistry and Health Effects, Anhui Agricultural University, 230036 Hefei, Anhui, P. R. China

[4]Present address: Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, USA

[5]These authors contributed equally to this article.

***Correspondence: Wilfried G. Schwab (wilfried.schwab@tum.de)**

https://doi.org/10.1016/j.xplc.2022.100506

ABSTRACT

Uridine diphosphate-dependent glycosyltransferases (UGTs) mediate the glycosylation of plant metabolites, thereby altering their physicochemical properties and bioactivities. Plants possess numerous UGT genes, with the encoded enzymes often glycosylating multiple substrates and some exhibiting substrate inhibition kinetics, but the biological function and molecular basis of these phenomena are not fully understood. The promiscuous monolignol/phytoalexin glycosyltransferase NbUGT72AY1 exhibits substrate inhibition (K_i) at 4 μM scopoletin, whereas the highly homologous monolignol StUGT72AY2 is inhibited at 190 μM. We therefore used hydrogen/deuterium exchange mass spectrometry and structure-based mutational analyses of both proteins and introduced NbUGT72AY1 residues into StUGT72AY2 and vice versa to study promiscuity and substrate inhibition of UGTs. A single F87I and chimeric mutant of NbUGT72AY1 showed significantly reduced scopoletin substrate inhibition, whereas its monolignol glycosylation activity was almost unaffected. Reverse mutations in StUGT72AY2 resulted in increased scopoletin glycosylation, leading to enhanced promiscuity, which was accompanied by substrate inhibition. Studies of 3D structures identified open and closed UGT conformers, allowing visualization of the dynamics of conformational changes that occur during catalysis. Previously postulated substrate access tunnels likely serve as drainage channels. The results suggest a two-site model in which the second substrate molecule binds near the catalytic site and blocks product release. Mutational studies showed that minor changes in amino acid sequence can enhance the promiscuity of the enzyme and add new capabilities such as substrate inhibition without affecting existing functions. The proposed subfunctionalization mechanism of expanded promiscuity may play a role in enzyme evolution and highlights the importance of promiscuous enzymes in providing new functions.

Key words: glycosyltransferase, substrate inhibition, hydrogen/deuterium exchange mass spectrometry, protein morphing, protein conformer, scopoletin

Liao J., Sun G., Kurze E., Steinchen W., Hoffmann T.D., Song C., Zou Z., Hoffmann T., and Schwab W.G. (2023). Subfunctionalization of a monolignol to a phytoalexin glucosyltransferase is accompanied by substrate inhibition. Plant Comm. **4**, 100506.

INTRODUCTION

Glycosylation catalyzed by nucleoside diphosphate sugar-dependent glycosyltransferases (GTs) is an important physiological reaction that affects the solubility, stability, transport,

Published by the Plant Communications Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and CEMPS, CAS.

storage, reactivity, and bioactivity of sugar acceptors such as proteins, lipids, saccharides, and small molecules (Kurze et al., 2021; Putkaradze et al., 2021). In plants and animals, uridine diphosphate (UDP) sugar-dependent GTs (UGTs) catalyze the glycosylation, e.g., the glucosylation and glucuronidation, of natural products and xenobiotics (Meech et al., 2019; Wilson and Tian, 2019). Plant UGTs involved in carbohydrate transfer to small molecules are encoded by large multigene families and show a conserved signature motif in their amino acid sequences known as the plant secondary (specialized) product glycosyltransferase (PSPG) box, which groups them among the family 1 glycosyltransferases (GT1) according to the CAZy database (www.cazy.org/GlycosylTransferases.html) (Kurze et al., 2021). Glycosyltransferases can be classified into more than 100 families, of which GT1, containing only UGTs, is the most numerous in the plant kingdom (Yonekura-Sakakibara and Hanada, 2011). Plant and animal UGTs are often promiscuous enzymes and glycosylate more than one substrate, but multiple UGT enzymes can also convert the same substrate. This fact suggests that substrate availability is a critical factor for product formation in a cellular context and that redundancy reflects the plasticity of glycosylation (Bowles et al., 2006).

Protein promiscuity plays a central role in evolution, and this is reflected in proposed models of enzyme evolution, as three out of four models start from promiscuous enzymes (Glasner et al., 2020). Two subfunctionalization models have been proposed, differing in their mechanism of functional specialization, as well as an innovation-amplification-divergence model, and all three build on multifunctional predecessors. Although these models predict that the endpoints of each evolutionary pathway are functionally specialized, examples from the literature show that endpoints can still be promiscuous for the ancestral activity (Noda-Garcia and Tawfik, 2020).

The 3D crystal structures of glycosyltransferases published to date mainly adopt one of two folds, termed GT-A and GT-B. GT-C and GT-D folds have also been identified but have few structural representatives (Zhang et al., 2014). Plant glycosyltransferases categorized as GT1 in the CAZy database (Drula et al., 2022) show a GT-B fold and catalyze the enzymatic reaction using an inverting glycosylation mechanism (Liang et al., 2015). Regardless of the individual topology, the reaction usually proceeds according to a sequential bi-bi mechanism, during which the carbohydrate donor substrate and aglycone acceptor substrate are sequentially bound, followed by sugar transfer to the latter (Luukkanen et al., 2005). The glycoside product is set free, followed by release of the nucleotide. The N- and C-terminal domains of GT-B members bind the acceptor and carbohydrate donor, respectively, and the GT-B protein undergoes a series of conformational changes during the reaction (Albesa-Jové and Guerin, 2016). Binding of the carbohydrate donor, which allows the pyrophosphate to interact by hydrogen bonds with the N- and C-terminal domains and thus stabilize the catalytically active conformation, triggers the switch from open to closed conformation (Qasba et al., 2005). The conformational closure changes the size and shape of the active site and forms the actual acceptor binding sites, which are stabilized by entropic effects, in accordance with the induced-fit mechanism (Liang et al., 2015). A histidine residue at about position 20 in the N-terminal domain of GT1 enzymes is regarded as the catalytic base for deprotonation of the hydroxyl group at the acceptor molecule to enable nucleophilic attack at the anomeric center of the sugar donor (Offen et al., 2006). The reaction corresponds to a single displacement mechanism. Protonation of the histidine is subsequently stabilized through hydrogen-bonding interactions with an aspartate residue around position 120.

Plant UGTs are involved in various pathways of specialized metabolism and therefore play a central role in growth and development. They are implicated in homeostasis of plant hormones such as abscisic acid and auxins, defense against plant pathogens by glucosylation of phytoalexins, and biosynthesis of lignin, although their exact roles in lignin formation have not yet been clarified (Yonekura-Sakakibara and Hanada, 2011; Wilson and Tian, 2019). For lignin biosynthesis, the lignin monomers 4-coumaryl, coniferyl, and sinapyl alcohols/aldehydes (monolignols) must be translocated into the cell wall for polymerization to lignin (Le Roy et al., 2016). UGT72E2 and UGT72E3, two glycosyltransferases from *Arabidopsis thaliana*, glucosylate monolignols at the 4-*O* position, suggesting that these enzymes might play a role in lignin biosynthesis (Lanot et al., 2006, 2008). Downregulation of the corresponding genes led to reduced monolignol glucoside levels in transformed Arabidopsis plants, and *UGT72E1* and *UGT72E2* were co-expressed with the peroxidases *PRX49* and *PRX72*, which are major actors in lignin polymerization in the cell wall (Le Roy et al., 2016). Although the detailed mechanism underlying the relationship between lignin biosynthesis and monolignol glycosylation is not yet clear, it is generally accepted that UGTs are essential for lignification of the plant cell wall (Le Roy et al., 2016).

Inhibition of enzyme activity at high substrate and/or cofactor concentrations, also known as substrate inhibition, is a common phenomenon observed in over 20% of known enzymes, including dehydrogenases, P450 enzymes, and UGTs (Reed et al., 2010). A two-site binding model has been proposed, in which one substrate binding site is productive (catalytic site) whereas the other (inhibitory site) is suppressive (Wu, 2011; Dong and Wu, 2012). Formation of a ternary dead-end complex has also been discussed as another substrate inhibition mechanism whereby accumulation of the non-productive ternary complex slows catalysis to a significant extent (Luukkanen et al., 2005). However, these models do not account for conformational change of the proteins upon ligand binding.

In two previous studies, we investigated the promiscuous monolignol/phytoalexin glycosyltransferase NbUGT72AY1 from the tobacco plant *Nicotiana benthamiana* (Sun et al., 2019, 2020). In this work, we observed pronounced substrate inhibition kinetics of the enzyme with scopoletin as the acceptor substrate, whereas its most similar homolog StUGT72AY2 from the potato plant *Solanum tuberosum* was only slightly inhibited at high scopoletin concentrations and showed a different substrate spectrum. Examination of NbUGT72AY1 by hydrogen/deuterium exchange mass spectrometry (HDX-MS) revealed amino acids putatively involved in scopoletin and UDP binding. Mutational analyses of NbUGT72AY1 and StUGT72AY2 in combination with *in silico* modeling and morphing studies identified an allosteric site formed after the transition from open to closed protein

conformer and highlighted amino acids important for substrate preference and inhibition. Although members of the UGT72 family preferentially glucosylate monolignols (Speeckaert et al., 2022), our results show that single amino acid substitutions can contribute to the subfunctionalization of UGT72s (Glasner et al., 2020), resulting in additional conversion of hydroxycoumarins but with substrate inhibition.

RESULTS

NbUGT72AY1 and StUGT72AY2 share high sequence identity but exhibit distinct substrate preferences and enzyme kinetics

UGT72 members catalyze the glycosylation of monolignols such as 4-coumaryl alcohol, coniferyl alcohol, sinapyl alcohol, and their respective aldehydes and are therefore likely to be involved in lignin formation (Lim et al., 2005; Speeckaert et al., 2020, 2022). During comparative functional characterization studies on UGTs from *N. benthamiana*, NbUGT72AY1 attracted our particular attention because of its distinct, unusually strong scopoletin glycosylation activity (Sun et al., 2019). Recombinant NbUGT72AY1 produced in *E. coli* glycosylated six monolignols *in vitro* (Figures 1A and 1B); it is expressed primarily in stem tissue (Sun et al., 2019) (Supplemental Figure 1) and may thus play a role in lignification of this plant part.

Additional biochemical assays showed that the monoterpene alcohol carvacrol, the coumarin derivatives umbelliferone and scopoletin, and the phenol vanillin are preferred substrates of NbUGT72AY1, but they strongly inhibited the enzymatic activity of NbUGT72AY1 at higher concentrations (Figures 1A and 1C). NbUGT72AY1 exhibited atypical Michaelis–Menten kinetics, and the data did not conform to the typical substrate inhibition equation, implying that the uncompetitive inhibition was incomplete. In particular, the substrate scopoletin showed unusually strong substrate inhibition kinetics for NbUGT72AY1 (Figure 1C).

The equation (Equation 4) that best fit the data of all substrates combines the two-site kinetic model for sequential ordered binding (Equation 3) (Wu, 2011) and the Hill equation (Equation 2) (LiCata and Allewell, 1997). The equation contains two Hill coefficients, n and x, where x takes into account the possibility that binding of the substrate in the inhibitory mode can also be cooperative. The model (Scheme 1) and corresponding equation also explained well the kinetics of an anthocyanidin UGT from strawberry (*Fragaria vesca*) (Peng et al., 2016). The high activity and strong substrate inhibition of carvacrol, umbelliferone, scopoletin, and vanillin can also be inferred from their kinetics, as the preferred substrates exhibited higher maximal reaction rates (V_{max}) but lower inhibition constants (K_i) compared with the values for the monolignols (Supplemental Table 1).

The inhibition constant K_i is the concentration of inhibitor required to decrease the maximal rate of the reaction to half of the uninhibited value. The glucosylation of monolignols is inhibited at much higher substrate concentrations ($K_i > 500$ μM), and the inhibition appears to be complete ($V_i = 0$ nmol/min/mg).

To analyze the remarkable properties of NbUGT72AY1 in more detail by comparative analysis, we searched for similar enzymes using BLAST (blast.ncbi.nlm.nih.gov) and identified StUGT72AY2 from potato (*S. tuberosum*) (Figure 2), whose gene was also strongly expressed in stem tissue regardless of whether TPM or FPKM values were used (www.ebi.ac.uk/gxa/experiments/E-MTAB-552/; PGSC0003DMG401004500) (Supplemental Figure 1). NbUGT72AY1 and StUGT72AY2 consist of 477 and 474 amino acids, 383 of which are shared (80.3% identity), and have isoelectric points of 6.00 and 5.67, respectively. Both contain the catalytically active H18 as well as D118 that activates H18, the PSPG box (W352–Q395; consensus sequence), and the GSS motif, a feature of mono-glucosyltransferases (Kurze et al., 2021). However, *in vitro* enzyme activity assays revealed distinct substrate preferences and kinetics for StUGT72AY2 compared with NbUGT72AY1 (Figure 1A). StUGT72AY2 showed high catalytic activity toward 2-methoxy-phenols (vanillin, sinapyl aldehyde, sinapyl alcohol, coniferyl alcohol, coniferyl aldehyde, and scopoletin), as well as carvacrol (in which the methoxy group is replaced by a methyl group), which was also converted. By contrast, NbUGT72AY1 favored short-chain phenols (carvacrol and vanillin) and cyclic hydroxyl phenylpropanoids (umbelliferone and scopoletin), whereas open-chain hydroxyl phenylpropanoids were slowly transformed. StUGT72AY2 showed weak substrate inhibition and even Michaelis–Menten curves were obtained for vanillin, coniferyl alcohol, coniferyl aldehyde, and sinapyl alcohol (Figure 1C; Supplemental Table 1). Both enzymes produced exclusively 4-*O*-glucosides of monolignols, as confirmed by LC-MS analysis (Supplemental Figure 2).

Homology modeling of NbUGT72AY1 identified closed and open conformers

To understand the structural requirements and conformational changes that accompany substrate inhibition, we generated a homology model of NbUGT72AY1. The amino acid sequence of NbUGT72AY1 was submitted to the IntFOLD web server (https://www.reading.ac.uk/bioinf/IntFOLD/). Among the top five 3D models, model IntFOLD5_MUSTER_multi8_TS1.bfact (template 6JTD_A; coverage of target: 0.9937107; RMSD: 0.72; TM-score: 0.98727) was chosen on the basis of its high Global Model Quality Score (GMQS) of 0.59 and confidence p-value of 1.818×10^{-9}.

Scores greater than 0.4 are characteristic of more complete and confident models; thus, the predicted model should closely reflect the native protein 3D structure (McGuffin et al., 2019). With the exception of the first five amino acids at the N terminus and amino acids 310 to 315, no residue exceeded the disorder/order probability score of 0.5 (Supplemental Figure 3), and conserved regions of UGTs (catalytic H18, activating D118, and the PSPG box from W350–Q393) were located in highly ordered regions of the protein. Accordingly, these residues and regions also show low per-residue errors (Supplemental Figure 4). NbUGT72AY1 is a typical GT1 family member, as it adopts the GT-B fold, and it should follow an inverting mechanism (Figure 3A).

The IntFOLD5-TS method works via iterative multi-template-based modeling (Buenavista et al., 2012) and uses target–template alignments. The templates (www.rcsb.org) 6JTD (He et al., 2019)

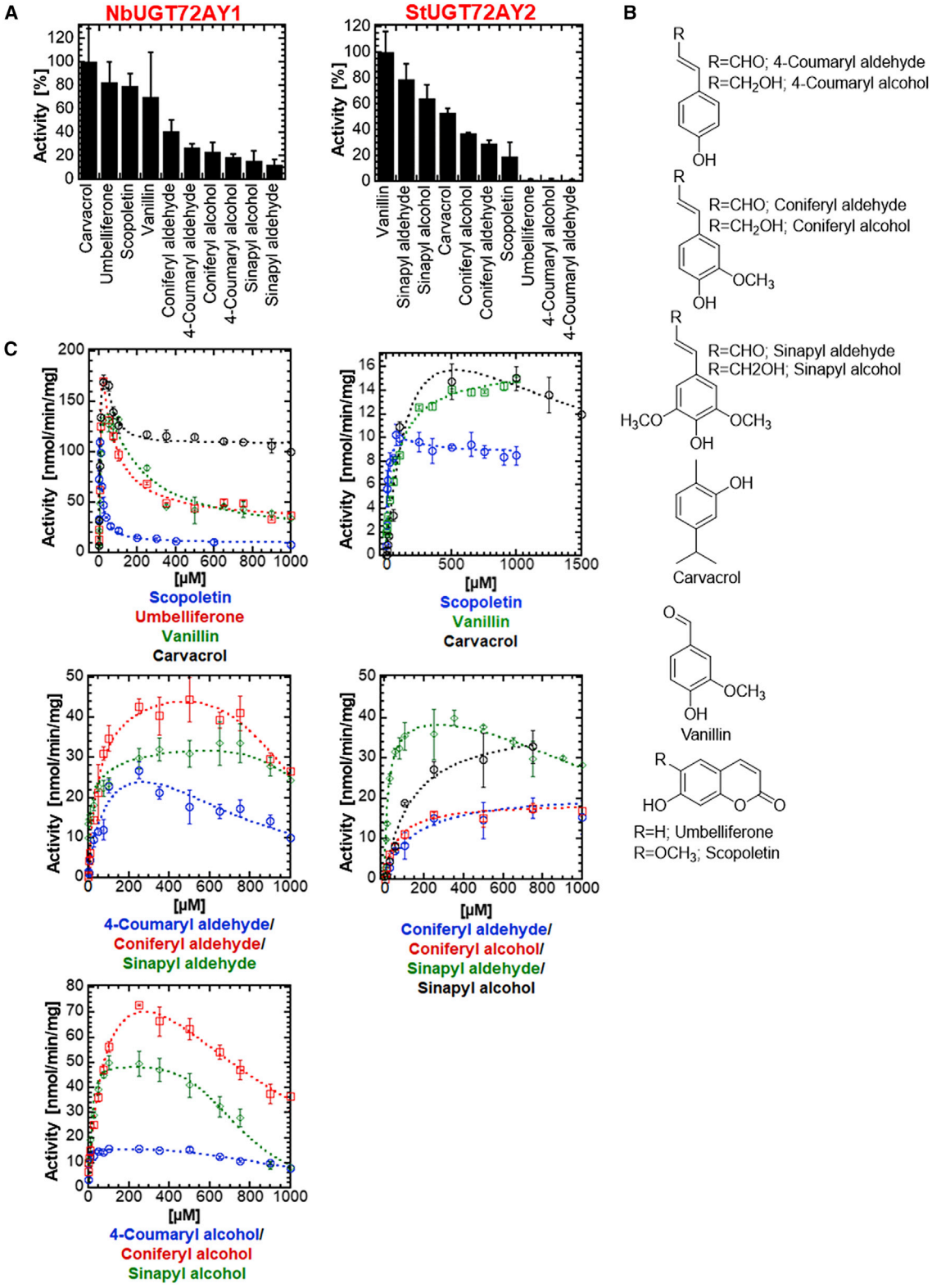


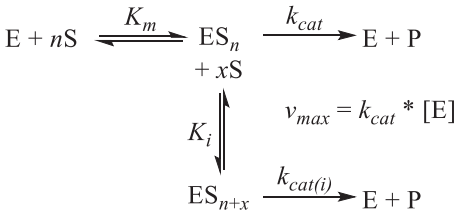
Figure 1. Substrate preferences and kinetics of NbUGT72AY1 and StUGT72AY2.

The first and second columns show the results for NbUGT72AY1 and StUGT72AY2, respectively.

(A) Substrate screening of NbUGT72AY1 and StUGT72AY2 using the UDP-Glo glycosyltransferase assay.

(B) Chemical structures of the acceptor substrates.

(C) Plots of acceptor substrate concentration versus reaction rate. Substrates are color coded. In **(A and C)**, data represent mean ± SD of n = 3 technical replicates.



Scheme 1. General reaction scheme for partial uncompetitive substrate inhibition

and 2VG8 (identical in sequence to 2VCE and 2VCH) (Brazier-Hicks et al., 2007) were preferred for generation of 3D structures because of their high amino acid sequence identities of 47.0% and 41.4%, respectively, with NbUGT72AY1. A search for suitable templates at the SWISS-MODEL server (https://swissmodel.expasy.org) and a detailed analysis revealed two additional protein sequences highly similar to NbUGT72AY1 whose 3D structures have been elucidated (Supplemental Figure 5).

The sequences of 6SU6 (identical to 6SU7 and 5NLM) (Hsu et al., 2018; Teze et al., 2021) and 6JEL (identical to 6JEN and 6JEM) (Maharjan et al., 2020) share 41.2% and 39.4% identical amino acids, respectively, with NbUGT72AY1. A closer inspection of the UGT protein structures suggested that 6JTD and 2VG8 show the enzyme in the closed form, whereas 6SU6 and 6JEL display the 3D model of the open conformer (Figure 3C). This assumption is supported by the greater distance between the N- and C-terminal domains of 2VG8/6JTD and 6SU6/6JEL (S62.A CB to E263.A CB in 6JTD: 19.1 Å, compared with S62.A CB to E262.A CB in 6SU6: 27.1 Å) and the opposite orientation of a loop in 2VG8/6JTD and 6SU6/6JEL, which is located between the domains. Furthermore, the side chain of the first amino acid of the PSPG box, Trp, orients differently to the donor substrate (Figure 3D). In the crystal structures of the putative open protein conformers (6SU6 and 6JEL), Trp is rotated about 180° compared with the 3D structures of the closed conformers (6JTD and 2VG8). This can explain why all soaking experiments with (closed) 6JTD were unsuccessful (He et al., 2019), whereas (open) 6SU6 (with UDP-Glc) and 6SU7 (with 3,4-dichloroaniline) complex structures were readily obtained after soaking of pre-formed *Polygonum tinctorium* UGT1 crystals with the ligands (Teze et al., 2021). Consequently, we also generated the 3D structure of the open conformer of NbUGT72AY1 based on 6SU6 using the SWISS-MODEL server (https://swissmodel.expasy.org), as the IntFOLD server does not accept predefined templates (Figure 3B). The GMQE and Qualitative Model Energy Analysis (QMEA) values were 0.71 and 0.71 ± 0.05 (Supplemental Figures 6 and 7), respectively. Among other differences, the open and closed NbUGT72AY1 conformers differ in the orientation of W350 and the flexible loop (amino acids 306 to 325, shown in red) (Figures 3A and 3B). The root-mean-square deviations (RMSDs) of atomic positions of closed and open NbUGT72AY1 compared with 6JTD and 6SU6 were 0.393 and 0.179 Å, respectively (Figure 3E).

HDX-MS analysis of NbUGT72AY1 revealed conformational changes upon substrate binding

It has been suggested that two substrate binding sites may contribute to substrate inhibition in the glycosylation reaction (Wu, 2011; Dong and Wu, 2012). To probe for substrate binding to NbUGT72AY1, we performed HDX-MS in solution. HDX-MS relies on exchange of the amide hydrogens of the protein backbone for hydrogens from the solvent, which can be quantified by MS using a solvent containing D_2O after proteolytic digestion of the protein owing to peptide mass shifts. The rate of H/D isotope substitution depends on the folded state and the dynamics of the protein (the stability of hydrogen bridge networks), and it may thus provide insights into protein structure or conformational changes upon ligand binding (Konermann et al., 2011; Masson et al., 2019).

NbUGT72AY1 was subjected to HDX-MS in the absence and presence of scopoletin or UDP ligands. After deuterium labeling and pepsin digestion, 150 peptides covering approximately 92% of the GST-tagged NbUGT72AY1 protein were obtained (Supplemental Figure 8A; Supplemental Table 2_HDX). The pattern of deuterium incorporation by NbUGT72AY1 agreed well with the predicted secondary structure, as rapid and high deuteration was apparent in protein parts without assigned secondary structures, e.g., the linker between the GST tag and NbUGT72AY1 and NbUGT72AY1 residues 244 to 255 and 305 to 328 (Supplemental Figure 8B).

Upon incubation with scopoletin, reduced HDX was observed in several areas of the N-terminal domain of NbUGT72AY1 (Supplemental Figure 9A), i.e., β-strand 1 and helix 1 (residues 8–29), the N-terminal portion of helix 4 (residues 83–91 and 94–98), the loop between β4 and α5 (residues 117–121), and parts of helices 6–10 and their interconnecting loops (residues 156–208). Most of these alterations should be caused by scopoletin binding to the enzyme active site, which is reflected in HDX changes surrounding the active site residues His18 and Asp118 (Figure 2). Correspondingly, no scopoletin-dependent perturbations in HDX were observed in peptides located in the GST tag or C-terminal domain of NbUGT72AY1. Changes in the HDX profile with respect to UDP were apparent for both N-terminal and C-terminal domains of NbUGT72AY1 (Supplemental Figure 9B) and similarly encompassed residues critical for coordination of this substrate (e.g., α-helix 17 with residues 371–376, and W350 located in the PSPG box, Figure 2). When both ligands (scopoletin/UDP) were applied at the same time, the effects were additive (Supplemental Figure 10). Collectively, these data suggest major conformational changes in NbUGT72AY1 caused by substrate binding, with scopoletin binding leading almost exclusively to transitions in the N-terminal part and UDP binding affecting both domains.

Scopoletin docking experiments on the predicted 3D model of NbUGT72AY1 to identify binding sites

Scopoletin docking was performed on the predicted closed 3D model of NbUGT72AY1 using the AutoDock Vina tool implemented in UCSF Chimera and the protein 3D structure 2VCE (www.rcsb.org) with the acceptor and donor ligands trichlorophenol and UDP-2-deoxy-2-fluoro-glucose, respectively. For the correct placement of UDP-Glc, 6SU6 (www.rcsb.org) was used as a template. The ligand scopoletin was installed in a plane with trichlorophenol in the active site (Supplemental Figure 11A). Comparison of the arrangement of the phenolic acceptor substrates in the active sites of the crystallized UGTs showed

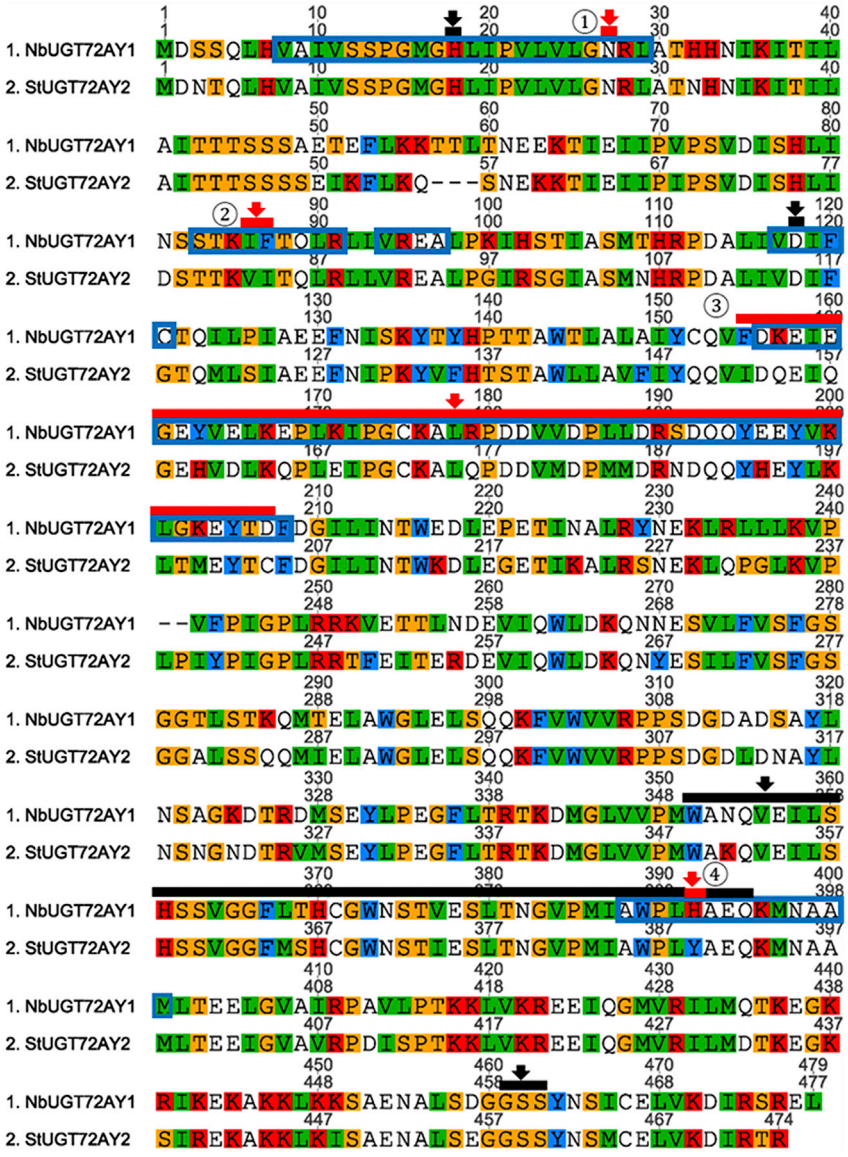


Figure 2. Amino acid sequence alignment of NbUGT72AY1 and StUGT72AY2 from *Nicotiana benthamiana* and *Solanum tuberosum*, respectively, and HDX results.

The catalytically active H18, activating D118, plant secondary product glycosyltransferase (PSPG) consensus sequence position 352–395, and G461/S462/S463 motif are highlighted by black bars and arrows; mutations ① N27D, ② IF/VI double mutant, ③ chimeric mutant, and ④ H/Y mutant are highlighted by red bars and arrows. Amino acids of NbUGT72AY1 that showed reduced hydrogen/deuterium exchange in HDX experiments in the presence of scopoletin are framed with blue boxes (Supplemental Figure S9A).

that they were all perfectly coplanar, with the hydroxyl group pointing toward the catalytically active His (Supplemental Figure 11B). The bond lengths (3.1 and 3.3 Å) and angle (102°) of the arrangement H18/scopoletin-OH/C1-UDP-Glc correspond well to the lengths (2.3 and 3.8 Å) and angle (111°) of the catalytic conformation in the crystal structure of 2VCE. A closer look at the active site shows that it is lined mainly with nonpolar amino acids (P14, G15, I86, F87, L90, I119, F120, P186, and A391), which presumably interact in a nonpolar (alkyl and π-π) and polar (van der Waals and carbon hydrogen) manner with the coumarin derivative (Supplemental Figure 11C). Scopoletin further forms a hydrogen bond with H18 that is important for catalysis, and D185, Q194, and H390 interact with the acceptor molecule via van der Waals forces and π-π stacking. The reaction between the acceptor and donor substrate takes place in a tunnel with openings at both ends (Supplemental Figure 11D).

The amino acids of NbUGT72AY1 that showed reduced HDX after the addition of scopoletin, UDP, and combinations of both were projected onto the predicted 3D structure of the open conformer of UGT72AY1 (Figure 4). The colors reflect the maximum amount of HDX regardless of the time course (Supplemental Figures 9 and 10). Most of the residues affected in their HDX by scopoletin (Figure 4A) are localized in the N-terminal domain and are part of α-helices (8–29, 83–91, 94–98, 117–121, and 386–399) that extend into the catalytic center. Interactions of amino acids with scopoletin in the active site stabilizes the hydrogen bonding networks, in particular the intrahelical hydrogen bonds, which results in significant decreases in deuterium incorporation (Skinner et al., 2012). However, residues 156–208 are rather remote from the active site and separated from it by α-helix 4 (Supplemental Figure 8A; Figure 4A). The effect of UDP on HDX in NbUGT72AY1 was more pronounced because the H/D exchange in amino acids of the N- and C-terminal domains was reduced (Figure 4B). In addition to amino acids whose HDX was also affected by scopoletin (N-terminal domain), numerous amino acids were affected at positions 243–377, a region that overlaps with the PSPG box, which is already known to interact with the donor substrate. Thus, UDP already stabilizes the hydrogen bonding networks in both domains. After co-addition of scopoletin and UDP, the effects were additive (Figure 4C).

Generation of NbUGT72AY1 and StUGT72AY2 mutants and their analysis highlight residues involved in substrate inhibition and promiscuity

To identify amino acids involved in scopoletin substrate inhibition and preference in NbUGT72AY1, the protein sequences of NbUGT72AY1 and StUGT72AY2 were compared in regions identified by HDX-MS as putatively important for ligand binding, and reverse mutations were generated when the sequences differed (Figure 2).

Deuterium incorporation of amino acids 87–97 in NbUGT72AY1 was strongly affected, but both sequences

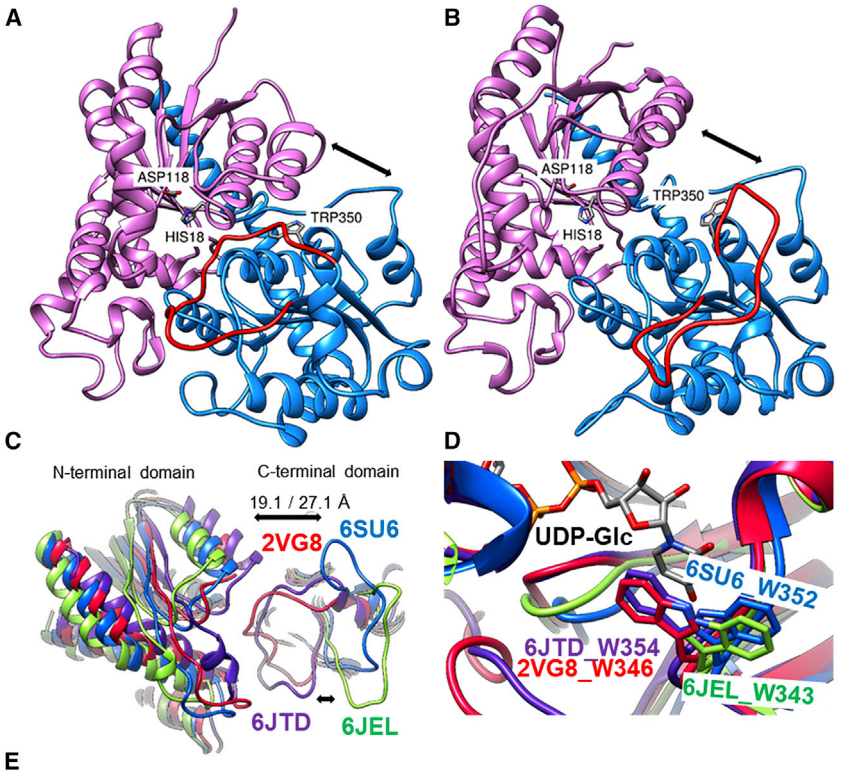


Figure 3. 3D UGT structures showing closed and open conformers.

(A) Prediction of the 3D structure of NbUGT72AY1 (closed conformer) was performed by the IntFOLD Integrated Protein Structure and Function Prediction Server (https://www.reading.ac.uk/bioinf/IntFOLD/) with default values based on 6JTD (www.rcsb.org). The result was visualized with UCSF Chimera (https://www.cgl.ucsf.edu/chimera). N- and C-terminal domains are shown in purple and blue, respectively. Important amino acids are marked, and the flexible loop covering the catalytic site is highlighted in red.

(B) Prediction of the 3D structure of NbUGT72AY1 (open conformer) was performed by the SWISS-MODEL Server (https://swissmodel.expasy.org) with default values based on 6SU6 (www.rcsb.org).

(C) Close-up of the superimposition of two putative open UGT conformers (6SU6 and 6JEL) and two closed UGT conformers (6JTD and 2VG8). Distances were measured between the N- and C-terminal domains of 6JTD and 6SU6 (S62.A CB to E263.A CB in 6JTD: 19.1 Å, in comparison to S62.A CB to E262.A CB in 6SU6: 27.1 Å).

(D) Tryptophan/uridine π-stacking interaction of the first amino acid of the PSPG box and UDP-Glc. In the crystal structures of the putative open protein conformers (6SU6 and 6JEL), W is rotated by 180° in comparison to the 3D structures of the closed conformers (6JTD and 2VG8).

(E) Calculation of mutual root-mean-square deviation (RMSD) values using UCSF Chimera.

E

root-mean-square deviation RMSD [Å]	NbUGT72AY1 closed	6JTD	NbUGT72AY1 open	6SU6
NbUGT72AY1 closed		0.393	1.105	1.131
6JTD	0.393		1.138	1.124
NbUGT72AY1 open	1.105	1.138		0.179
6SU6	1.131	1.124	0.179	

were identical in this region, except for position 87. Thus, the double mutants I86V_F87I of NbUGT72AY1 and V83I_I84F of StUGT72AY2 were generated, as these residues are within 5 Å of the catalytic center (② in Figure 2; Supplemental Figure 11C) and differ in both enzymes. In addition, a region from amino acids 156–208 of NbUGT72AY1 showed reduced HDX, and several residues were different in the corresponding region of StUGT72AY2. Therefore, this entire sequence part was exchanged, and chimeric mutant proteins were generated (③ in Figure 2). Scopoletin binding also affected deuterium incorporation of amino acids 386–399 located at the C-terminal end of the PSPG box. In this sequence segment, only the amino acid H390Y was different, and H390 was predicted to interact with scopoletin via π-stacking (Supplemental Figure 11C); therefore, a single mutant was created (④ in Figure 2). Because preliminary scopoletin docking experiments had revealed the N-terminal α-helix extending from amino acids 15–29 as a possible second interaction site, the polarity and charge of this site was altered by generating the single mutant N27D for both proteins (① in Figure 2). The kinetics of the four mutants of NbUGT72AY1 and StUGT72AY2 were analyzed with the substrates scopoletin and sinapyl aldehyde using the UDP-Glo assay (Figure 5).

Comparison of the V_{max}, V_i, K_m, and K_i values for scopoletin of the N27D mutant and the wild-type NbUGT72AY1 showed that the data were virtually identical. Although the I86 and F87 residues in NbUGT72AY1 are located near the substrate binding site, V_{max} was unchanged for I86V_F87I. However, V_i and K_i were significantly increased, resulting in a substantial decrease in substrate inhibition, similar to the chimeric mutant, in which K_m and V_{max} were also significantly increased and reduced, respectively, leading to an overall lower enzymatic activity. The elevated K_i value of the I86V_F87I mutant indicates that the two mutant amino acids are also located near the allosteric site. The catalytic activity of the H390Y mutant was reduced owing to a significantly lower V_{max} value, but substrate inhibition (K_i) was only slightly affected.

Interestingly, the cooperativity in the inhibitory mode (Hill coefficient x) was significantly increased for the chimeric mutant (Figure 5A). By contrast, the chimeric mutant of StUGT72AY2 showed no enzymatic activity, and the kinetic data of the N27D mutant of StUGT72AY2 closely resembled those of the wild-type enzyme.

For the double mutant V83I_I84F, increased V_{max} and V_i and a lower K_m value resulted in higher catalytic activity, whereas for

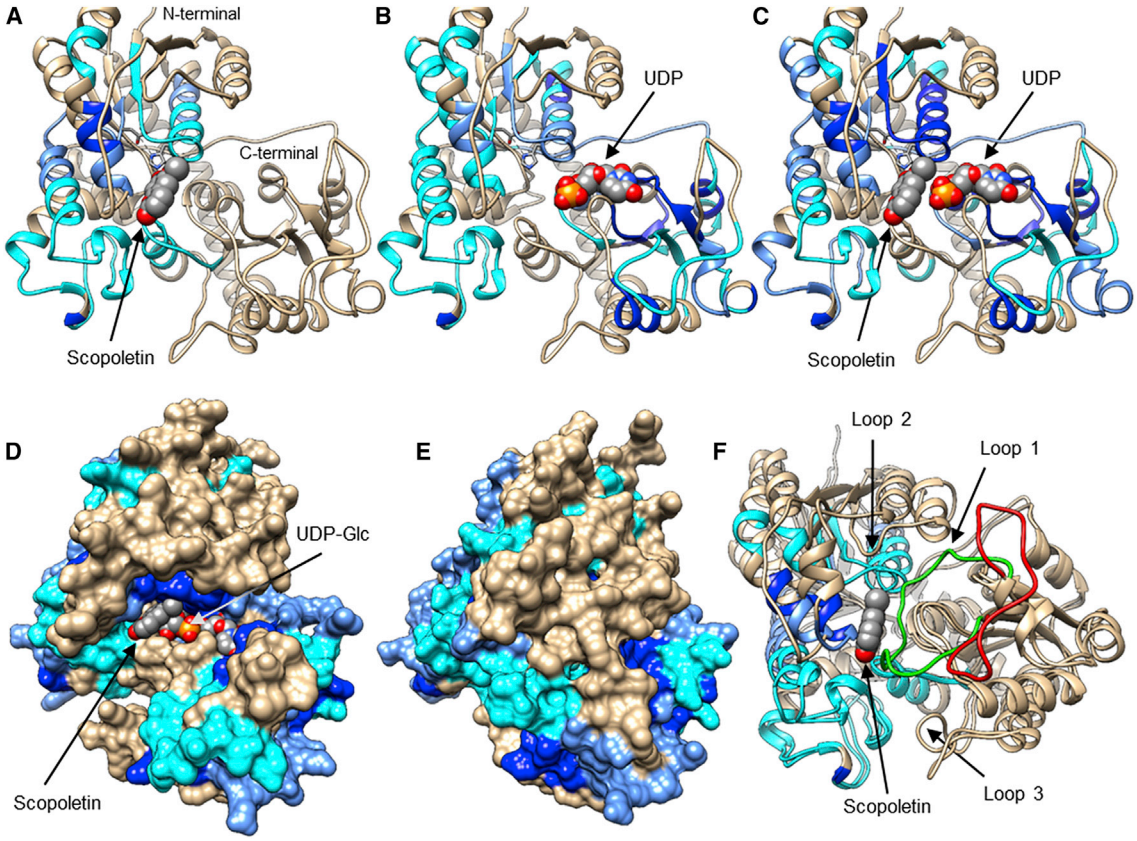


Figure 4. HDX results mapped on NbUGT72AY1.

Differential HDX results obtained after the addition of scopoletin, UDP, and a combination of both to the protein were color-mapped onto the predicted 3D structure of the open conformer of NbUGT72AY1.

(A) After addition of scopoletin (in sphere display), **(B)** after addition of UDP, and **(C)** after addition of scopoletin and UDP.

(D) Scopoletin and UDP-Glc (both in sphere display) are shown in the open conformer of the predicted structure of NbUGT72AY1 (HDX results of UDP/scopoletin are color-mapped).

(E) Scopoletin and UDP-Glc are buried in the closed conformer of NbUGT72AY1.

(F) Flexible loop (1) is highlighted in the NbUGT72AY1 model (green, closed conformation; red, open conformation; HDX results of scopoletin are color-mapped). Two additional closing loops are displayed.

the Y389H mutant, V_{max}, V_i, K_m, and K_i values increased along with the Hill coefficient x, leading to higher enzymatic activity.

LC-MS was used as a second independent measurement method because it directly quantifies the glucoside product of enzymatic catalysis. Although in the case of NbUGT72AY1 and its mutants, the kinetic data obtained by the UDP-Glo assay were confirmed, for StUGT72AY2 and its mutants, LC-MS revealed lower enzymatic activities for all enzymes in the concentration range of less than 200 μM scopoletin (Supplemental Figure 12). We therefore hypothesize that the UDP-glucose hydrolase side activity of StUGT72AY2 and its mutants is responsible for the higher overall activity in the UDP-Glo assay at low scopoletin concentrations. However, the general conclusion from the experiments was not compromised.

When sinapyl aldehyde was used as a substrate for the various mutant enzymes, the kinetic data differed less than with scopoletin. For I86V_F87I, substrate inhibition was completely abolished because V_i was greater than V_{max}, and K_i and the second Hill coefficient were significantly reduced. The enzymatic activity of the chimeric mutant was decreased owing to a higher K_m value. As with scopoletin, the chimeric mutant of StUGT72AY2 was inactive with sinapyl aldehyde, whereas the other mutants, except for Y389H, showed similar kinetics to their wild-type enzyme. Y389H exhibited an increased K_i value, resulting in higher enzymatic activity. Overall, these results highlighted the role of I86–F87, H390, and the sequence segment F155–D207 in substrate inhibition of scopoletin in NbUGT72AY1.

Morphing of open and closed NbUGT72AY1 conformers reveals conformational changes important for catalysis and substrate inhibition

Major changes in secondary, tertiary, and quaternary structure are essential for the functions of UGTs (Qasba et al., 2005, 2008). Because we obtained the 3D structures of the open and closed conformers of NbUGT72AY1 by homology modeling based on X-ray structures of proteins with sequence identities of >40%, we performed morphing of both extreme forms. Morphing involves calculation of a series of intermediate, interpolated structures between the original input structures. The series of structures can be rendered as a movie (Weiss and Levitt, 2009). The HDX results for scopoletin/UDP were color-mapped onto

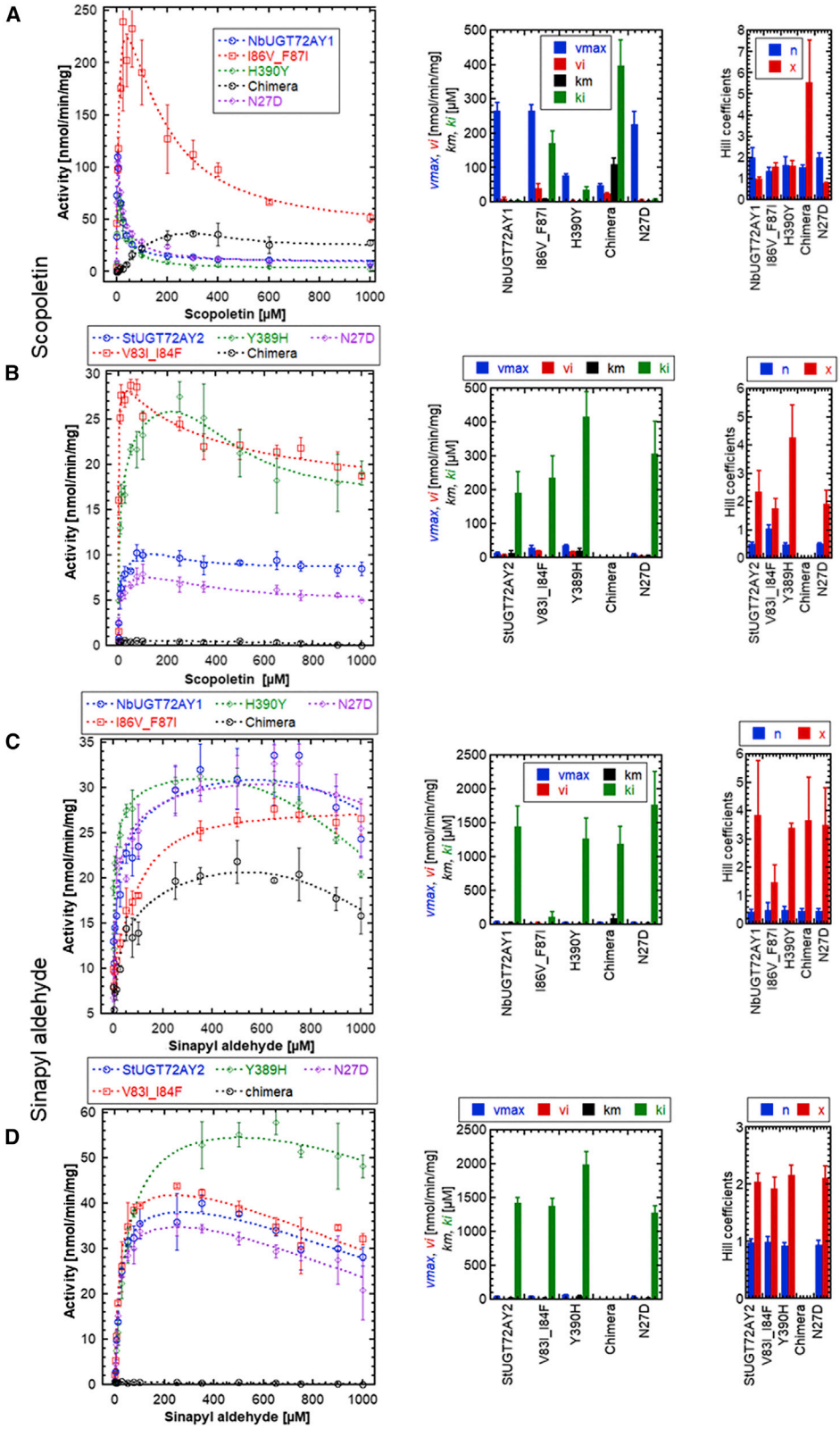


Figure 5. Kinetic parameters of NbUGT72AY1, StUGT72AY2, and mutant enzymes using scopoletin and sinapyl aldehyde as substrates.

(A) NbUGT72AY1 and its mutants were used to glucosylate scopoletin.

(B) StUGT72AY2 and its mutants were used to glucosylate scopoletin.

(legend continued on next page)

the 3D morphing structures to visualize structural and dynamic aspects of NbUGT72AY1 motions. The moving model also includes the ligands scopoletin and UDP-Glc in the active site to identify the second scopoletin binding site (Supplemental Video 1; Figures 4D and 4E). The animation shows that both the C- and N-terminal domains move toward each other, thus closing the cleft, which represents the active site between the two domains. In the animation, all amino acid residues are in motion, not only the amino acids, which are marked according to their HDX results. However, two regions (amino acids 87–91 and 346–353, labeled in dark blue, indicating strongly reduced deuterium incorporation, in the left and right corners of the active site, respectively) show strong conformational changes and are covered in the closed conformer. Three loops are primarily responsible for closure of the catalytic center (Supplemental Video 2; Figure 4F). A loop of amino acids 305–325 (right) approaches an opposite loop (amino acids 71–82) in the N-terminal domain (left), while the third loop (amino acids 410–420) covers the active site from below (Supplemental Figure 13). Overall, the N-terminal domain (left) undergoes a much stronger conformational change than the C-terminal part (right). A closer look at the catalytic center reveals a minimal distance between the proton acceptor Nε2-His and OH-scopoletin of 2.0 Å in the open enzyme conformer, whereas the distance to the accessory D118 (Nδ1-His to COOH-Asp) is 3.7 Å (Supplemental Video 3; Figures 6A and 6B). During closure, D118 approaches H18 to within 3.1 Å, and H18 simultaneously rotates and thus moves away from the OH of scopoletin (to within 3.1 Å), consistent with the presumed deprotonation of the hydroxyl group. D118 accepts a proton on its closest approach to H18 and then distances itself again, supporting the proton transfer to H18. After the acceptor substrate anion attacks the donor substrate UDP-Glc (not shown), the amino acids involved in catalysis return to their positions in the open UGT conformer.

From the animation, it can be concluded that catalysis already starts when the catalytic center closes. In addition, F87 and I86 mutated in the double mutant are involved in formation of the scopoletin binding pocket, which is completed only during active site closure. They increase the hydrophobicity of the catalytic center, facilitating acceptor substrate incorporation and catalysis. A feature that distinguishes closed UGT conformers from open ones is the orientation of the first amino acid Trp of the PSPG box to the uridine part of the donor substrate (Figure 3; Supplemental Video 4; Figure 6C and 6D). The animation reveals the approximate 180° rotation of W350 during active site closure, resulting in formation of a π-π stacking interaction of the phenyl and uracil rings (Harrus et al., 2018). In addition to the motions of H18, I86, and F87, the animation also shows the rotation of H390, an amino acid that was mutated because HDX results and comparative biochemical assays suggested that it might be involved in catalysis and substrate inhibition. H390 is located near the donor and acceptor substrate, I86, and the loop (amino acids 306–325) that covers the active site.

In search of the allosteric binding site, we took a closer look at the vicinity of F87 and I86 because the substrate inhibition of NbUGT72AY1 was significantly reduced in the double mutant (Supplemental Video 5; Figures 6E and 6F). R91, located at the same α-helix as F87, undergoes a dramatic conformational change during transition from the open to the closed conformer and eventually forms a cation-π interaction with F87 and F120 (Flocco and Mowbray, 1994; Gallivan and Dougherty, 1999; Steiner and Koellner, 2001). The dislocation opens a new binding site in the closed NbUGT72AY1 conformer, allowing attachment of a second scopoletin molecule (Supplemental Video 6; Figure 6G–6I). Formation of the allosteric binding site upon binding of scopoletin in the catalytic center is consistent with the uncompetitive substrate inhibition model (Scheme 1).

To confirm the detection of the allosteric site experimentally, we generated three R91 mutants, R91A, R91F, and R91M. Because A91 cannot block the allosteric site and F91, similar to R91, can form π-interactions with F87, we expected that only M91 would reduce substrate inhibition in comparison with the wild-type enzyme. The results showed that the mutants had similar V_{max} and K_m values, and therefore the catalytic efficiency of the mutants was not affected, whereas the V_i and in particular the K_i data differed (Supplemental Figures 14A and 14B). Overall, R91M exhibited reduced substrate inhibition, as V_i (13.1 ± 6.1 nmol/min/mg) and K_i (27.3 ± 4.2 μM) exceeded those of the wild-type enzyme (R91; 4.4 ± 2.3 nmol/min/mg and 10.7 ± 2.2 μM, respectively). R91A and R91F showed significantly stronger inhibition by scopoletin than the wild-type enzyme, reflected in lower K_i values (3.3 ± 0.7 μM and 0.9 ± 0.3 μM, respectively). After molecular docking of scopoletin in the putative allosteric site, the binding energies ΔG were used to calculate equilibrium constants K_D (Supplemental Figures 14C–14G). The K_D values determined *in silico* were considerably higher than the experimentally determined K_i values, but the ranking of the values was identical (R91F < R91A < R91 < R91M). Finally, we generated the single I86V and F87I mutants of NbUGT72AY1 and determined their kinetic parameters (Supplemental Figures 15A and 15B). Whereas I86V showed similar data to the wild-type enzyme, F87I exhibited lower substrate inhibition, similar to the double mutant I86V_F87I. Thus, only F87 is involved in formation of the second scopoletin binding site.

Subdivision of the chimeric mutants narrowed down the amino acids responsible for substrate inhibition

Because both chimeric mutants each contain 53 amino acids of the homologous UGT, we generated additional mutants to identify amino acid sequences that cause substrate inhibition in NbUGT72AY1. The newly produced mutants each carried only one half of the 53-amino-acid chimera sequence (Supplemental Figure 16C). Enzyme assays revealed that NbUGT72AY1-chimera A, carrying I152–N188 of StUGT72AY2, exhibited significantly reduced substrate inhibition kinetics, whereas the reverse mutant StUGT72AY2-chimera A showed atypical Michaelis–Menten kinetics in contrast to StUGT72AY2

(C) NbUGT72AY1 and its mutants were used to glucosylate sinapyl aldehyde.
(D) StUGT72AY2 and its mutants were used to glucosylate sinapyl aldehyde. Experimental data were obtained by UDP-Glo glycosyltransferase assay and fitted to the partial uncompetitive inhibition model (Equation 4). The amino acid sequence information of the mutants is shown in Figure 2. Data represent mean ± SD of n = 3 technical replicates.

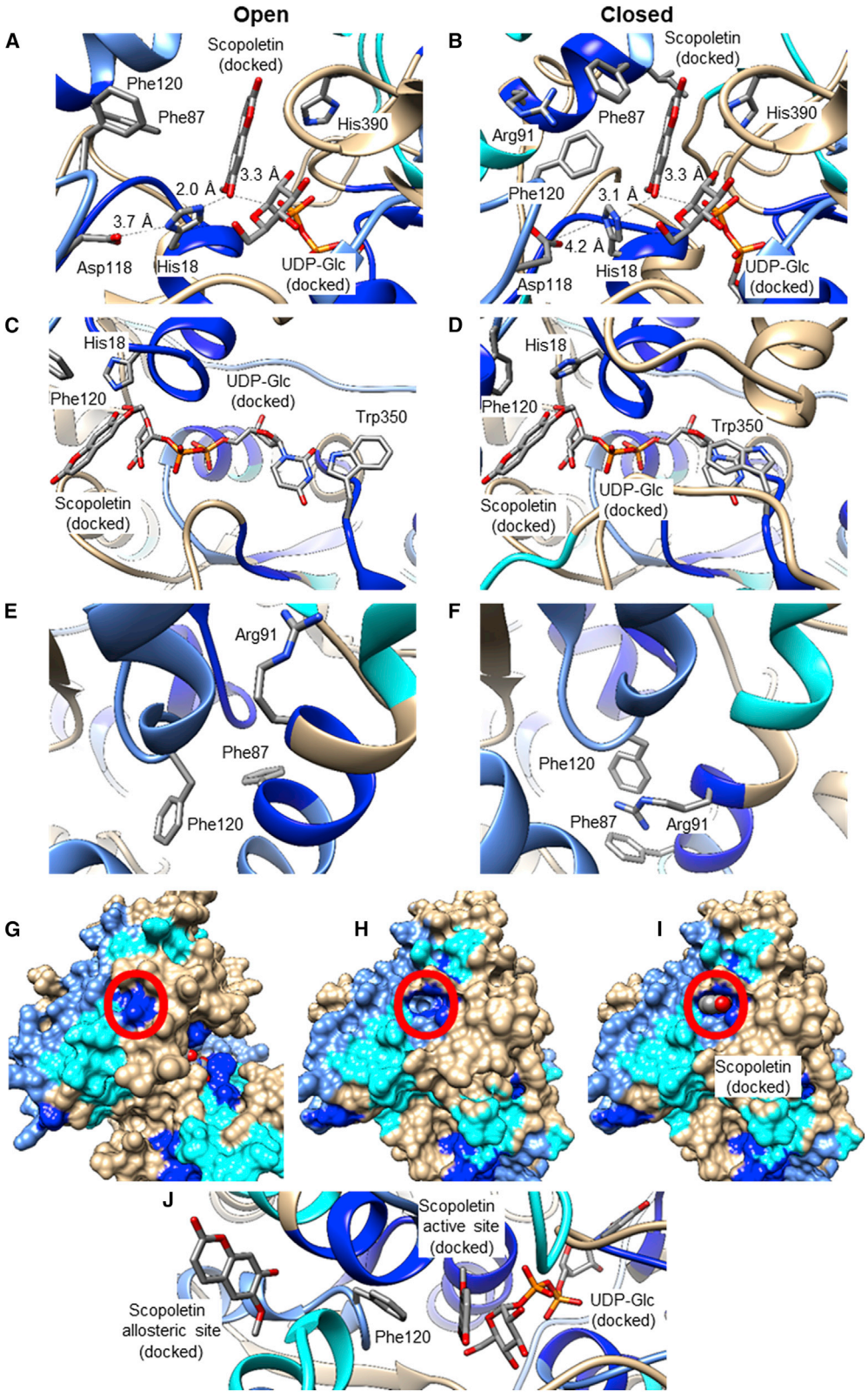


Figure 6. Visualization of conformational differences in the open and closed conformer of NbUGT72AY1.

(A) Distance of Nε2-His to OH-scopoletin, Nδ1-His to COOH-Asp, and OH-scopoletin to C1-UDP-Glc in the ternary complex (NbUGT72AY1●scopoletin●UDP-Glc); HDX results with scopoletin/UDP are color-mapped, and the open conformer is shown.

(B) Same as in **(A)**, but the closed conformer is shown.

(legend continued on next page)

and StUGT72AY2-chimera B. The reduction in substrate inhibition resulted in significantly increased V_i and K_i values in NbUGT72AY1-chimera A. The kinetics of StUGT72AY2 and its chimera B mutant can best be described by the Michaelis–Menten equation (Equation 1) but also by Equation 4 if $V_{max} = V_i$, $K_m = K_i$, and $n = x = 1$ are assumed. None of the amino acids that differ in the chimera A sequence in NbUGT72AY1 and StUGT72AY2 are components of the catalytic or postulated allosteric site (Supplemental Figure 16D). Because the different mutants examined in the present study could not completely reverse the substrate inhibition kinetics of the enzymes from tobacco and potato in each other, more than one amino acid exchange must be responsible.

DISCUSSION

UGTs represent a superfamily of enzymes found in all kingdoms of life. They comprise several groups and subgroups owing to the large number of acceptor substrates (Bowles et al., 2006). *N. benthamiana* transcribes about 290 UGTs (https://sefapps02.qut.edu.au/atlas/benthgenereturn6.php), some of which were recently characterized (Sun et al., 2019, 2020). Among them, one translated enzyme (NbUGT72AY1) stood out because it exhibited strong substrate inhibition with several substrates (Figure 1).

UGT72 enzymes are probably involved in lignin biosynthesis

NbUGT72AY1 is a homolog of AT5G66690.1 from Arabidopsis (AtUGT72E2), an enzyme shown to be involved in lignin metabolism (Baldacci-Cresp et al., 2020). AtUGT72E2 glucosylates sinapyl and coniferyl aldehydes as well as their corresponding alcohols. A knockdown mutant line (72E2KD) produced by RNAi silencing showed a two-fold reduction in coniferyl alcohol 4-*O*-glucoside and sinapyl alcohol 4-*O*-glucoside compared with the wild type (Lanot et al., 2006, 2008). Similarly, NbUGT72AY1 and its ortholog from potato, StUGT72AY2, transfer glucose to monolignols, preferring coniferyl and sinapyl derivatives (Figure 1A), and their products were identified as 4-*O*-glucosides (Supplemental Figure 2). Because of their high sequence identity, the two enzymes have been assigned consecutive numbers by the UGT Nomenclature Committee (https://prime.vetmed.wsu.edu/resources/udp-glucuronsyltransferase-homepage). They are also likely to be involved in lignin biosynthesis because their sequences are highly similar to those of monolignol UGTs (Speeckaert et al., 2020, 2022) and they efficiently glucosylate phenylpropanoids *in vitro* (Figure 1). They are constitutively expressed in the stem (Supplemental Figure 1), and the potato enzyme interacts with nine peroxidases that are thought to be required for lignin formation (https://string-db.org/network/4113.PGSC0003DMT400011466). However, UGT72 members are promiscuous and also glycosylate flavonoids and other phenolics (Speeckaert et al., 2022).

Identification of open and closed protein conformers enabled the observation of enzyme dynamics by morphing

Because of the high substrate similarity of NbUGT72AY1 and StUGT72AY2 but their drastically different enzyme kinetics toward scopoletin, HDX analyses in combination with mutation studies appeared to be a promising approach for elucidating the molecular mechanism of substrate inhibition. GT1 family members catalyze the reaction by a compulsory ordered bi-bi mechanism in which UDP-sugar is the first binding substrate (Luukkanen et al., 2005; Hashimoto et al., 2010; Liang et al., 2015). However, similar to (Albesa-Jové et al., 2017), our HDX results demonstrated that acceptor substrate binding is possible in the absence of UDP/UDP-Glc (Figure 4A; Supplemental Figure 9A). NbUGT72AY1 was found to have more than 40% sequence identity with four crystallized UGTs whose 3D structures have already been elucidated (Supplemental Figure 5). Interestingly, two structures (6SU6 and 6JEL) represented UGTs with an open catalytic site, whereas the other two structures (6JTD and 2VG8) showed closed conformers. These templates provided a unique opportunity to model both conformer states of NbUGT72AY1 and visualize the dynamics of enzyme catalysis by morphing (Supplemental Videos 1 and 2; Figure 4).

Conformational changes and enzyme catalysis

Conformational changes of UGTs during catalysis have been reported for GT-A enzymes, including the human beta-1,4-galactosyltransferase (B4GalT1) (Harrus et al., 2018) and the glucosyl-3-phosphoglycerate synthase (GpgS) from *Mycobacterium tuberculosis* (Albesa-Jové et al., 2017). These proteins contain short dynamic loops inserted between two Rossmann fold domains, which undergo an open-to-closed motion essential for acceptor and donor substrate recognition and assembly of the reaction center. A limited number of GT-B-folded UGT structures show that these enzymes often exhibit global domain motion upon substrate binding and differ in the type and extent of movement (Chang et al., 2011). It was hypothesized that this open-to-closed conformational transition brings acceptor and donor substrate into close proximity and is accompanied by multiple loop displacements (Bolam et al., 2007). We observed that the distance between the catalytically active His and C1 of the acceptor substrate scopoletin is smaller in the open enzyme conformer (2.0 Å) than in the closed form (3.1 Å; Supplemental Video 3; Figures 6A and 6B). This indicates that deprotonation

(C) UDP-Glc binding pocket, showing the rotation of Trp350; HDX results with scopoletin/UDP are color-mapped, and the open conformer is shown.
(D) Same as in **(C)**, but the closed conformer is shown.
(E) Putative allosteric binding site formed after rotation of Arg91; HDX results with scopoletin/HDX are color-mapped, and the open conformer is shown.
(F) Same as in **(E)**, but the closed conformer is shown.
(G) Surface presentation of the open conformation of NbUGT72AY1; HDX results with scopoletin/UDP are color-mapped. Red circle shows the putative allosteric site, which is closed in the open conformer.
(H) Surface presentation of the closed conformation of NbUGT72AY1. The new binding site is marked by a red circle.
(I) Same as in **(H)**, with scopoletin bound in the putative allosteric binding site.
(J) Phe120 is located between the catalytic and allosteric sites.

and therefore enzyme catalysis already starts with the motion of the loop, which is suggested by distance measurements between Nε2-His and C1 of UDP-Glc in the structures 6SU6 (open), 2ACW (closed), and 6LFZ (closed) of 4.6, 5.5, and 5.6 Å, respectively.

Particularly striking in the transition from open to closed form is the movement of a loop (amino acids 305–325; Supplemental Figure 13) toward the active center (Supplemental Videos 1 and 2, Figure 4F). Consequently, water is likely to be displaced from the active site, leading to the conclusion that the two small openings that can be identified in the closed enzyme conformer (Supplemental Figure 11D) are not substrate entry ports but probably drainage channels (Breton et al., 2006). Amino acids located at the two apertures could affect the efflux of water from the active site and thus control the inherent UDP-Glc hydrolase activity of UGTs (Sheikh et al., 2017). The closing loop, albeit of varying length, is present in various already crystallized GT-B-fold UGTs from plants (Supplemental Figure 5) and should behave similarly in them, as shown for flexible structures in GT-A-fold enzymes (Chang et al., 2011). In structures that do not contain this loop, alternative loops can be found at other positions, which presumably have a similar function.

When comparing the open and closed NbUGT72AY1 structural models, the position of the first amino acid of the PSPG box W350, which participates in fixation of the donor substrate by π-π interaction with the uracil group, is apparent (Supplemental Video 4; Figures 6C and 6D). This arrangement, together with the position of the loop (amino acids 305–325), can be used to distinguish open and closed GT-B folded UGTs based on their crystal structure. The Trp rotation could guide the donor substrate into the active site, but rotation could also be triggered only by the incorporation of the donor substrate into the active site, and/or the back rotation of W350 could facilitate the export of the reaction product UDP (Harrus et al., 2018).

An amino acid that affects mainly catalytic activity, but to a lesser extent substrate inhibition, of NbUGT72AY1 is H390. It moves during the transition from the open to the closed form (Supplemental Video 4) and interacts with the phosphate group adjacent to glucose in UDP-Glc, F87 in the active site, and the flexible loop (amino acids 303–325). Mutation to Tyr, as in the enzyme from potato, decreases catalytic activity by decreasing V_{max} (75.5 versus 264.3 nmol/min/mg) and increasing K_i (36.2 versus 4.5 μM), explainable by reduced stabilization of the donor substrate and altered enzyme dynamics due to interaction with amino acids of the flexible loop (S315 and A316).

Identification of the second binding site

It has been proposed that two substrate binding sites and the formation of a dead-end ternary complex might contribute to substrate inhibition in the glycosylation reaction (Wu, 2011; Dong and Wu, 2012). Our results indicate two dependent substrate binding sites: after binding of scopoletin in the active site, conformational changes in which R91 plays an important role result in formation of a second allosteric binding site (Figures 6E and 6F). The conformational change is promoted by the formation of two cation-π interactions, structures commonly found in 3D protein structures (Supplemental Videos 5 and 6) (Flocco and Mowbray, 1994; Gallivan and Dougherty, 1999; Steiner and Koellner, 2001). This mechanism is consistent with the partial uncompetitive inhibition model used for calculating the kinetic parameters (Scheme 1) (Peng et al., 2016). The importance of R91 as a key amino acid for substrate inhibition has been clearly demonstrated by mutational studies (Supplemental Figure 14). Therefore, our study provides new insights into the dynamics of the UGT catalytic mechanism and the conformational changes that lead to substrate inhibition. The arrangement of the three V-shaped helices enclosing the allosteric center in NbUGT72AY1 is evident in 3D structures of GT-B-folded UGTs (Supplemental Figure 17). Because substrate inhibition is observed in numerous plant (Peng et al., 2016) and human UGTs (Wu, 2011), it can be assumed that this conserved spatial structure is also used as an allosteric center in other UGTs. The allosteric binding site is formed by R91 spatially approaching and forming a π-stacking interaction with F87 during the transition from the open to the closed conformation. Biochemical analysis of the single mutants I86V and F87I confirmed this hypothesis (Supplemental Figure 15). In StUGT72AY1, F87 is replaced by I84, which is not able to make this kind of interaction with R91 (Figure 2; Supplemental Figure 18).

Our studies show that the allosteric site is presumably located in close proximity to the catalytic center so that, for example, Phe120 is within 5 Å of both binding sites (Figure 6J). This means that mutations to reduce substrate inhibition will inevitably have an effect on catalytic activity, as has been shown for UGTs and other enzymes (Barnett et al., 2004; Miller et al., 2008; Chen et al., 2014). In our study, based on structural information and structure-activity experiments, we identified a position that exclusively affects substrate inhibition.

Chimeric mutants and sinapyl aldehyde glucosylation

All four parameters, V_{max}, V_i, K_m, and K_i, were severely altered in the chimeric mutant of NbUGT72AY1, providing the explanation for its decreased substrate inhibition and reduced catalytic efficiency (Figure 5A). Because the amino acids exchanged in the chimeric mutant are not located directly in the catalytic and allosteric centers, we assume that the dynamics of protein conformational change during catalysis are altered in this mutant, e.g., by changing the quaternary structure. The high Hill coefficient for the inhibited reaction, which is only observed in the chimeric mutant of NbUGT72AY1, hints at changes in quaternary structure (Figure 5A). In this regard, there are multiple examples in which UGTs adopt a quaternary structure, e.g., by homodimerization (Liu et al., 2016; Wang et al., 2020), and in which substrate specificities, reaction rates, and types of reaction products are influenced by oligomerization (Fujiwara et al., 2016; Smith et al., 2020). The number and/or oligomerization positions of protomers may have been altered by the mutation. The chimeric mutant of StUGT72AY2, on the other hand, was enzymatically inactive, potentially because of similar alterations in oligomerization (Figure 5B).

Additional mutants that split the chimera mutant into two parts were generated and analyzed. NbUGT72AY1-chimera A showed significantly reduced substrate inhibition, whereas the reciprocal StUGT72AY2-chimera A was inhibited by scopoletin (Supplemental Figure 16). This result indicates that amino acids

F155–S191 contribute to substrate inhibition in NbUGT72AY1 in addition to F87.

Overall, scopoletin substrate inhibition was not completely abolished by the generated mutants. The strongest effects on substrate inhibition were observed with the double and especially the chimera/chimera A mutants of the tobacco enzyme, as evidenced by their increased V_i values. The postulated allosteric site of NbUGT72AY1 differs from the corresponding site in StUGT72AY2 by 87Phe and amino acids in the C-terminal sequence of the chimera (Glu196, Val199, Gly202, Lys203, and Asp207) (Figure 2). However, the side chains of the latter amino acids are located outside the putative allosteric site, and the reciprocal chimera potato mutant was inactive, whereas the double mutant exhibited apparently enhanced substrate inhibition due to increased V_{max} and lower K_m values. These results illustrate that no single amino acid alone is responsible for substrate inhibition, which instead arises from the interplay of different amino acid residues that influence the dynamics of protein movement and interaction with additional substrate molecules. The different mutations could contribute additively or synergistically to substrate inhibition.

The mutants of the two UGTs did not exhibit strong changes in kinetic data for the monolignol sinapyl aldehyde, which showed little substrate inhibition of the two wild-type enzymes (Figures 5C and 5D). However, results obtained with the I86V_F87I mutant of NbUGT72AY1 and the chimeric mutant of StUGT72AY2 using sinapyl aldehyde as a substrate confirmed the importance of the mutant amino acids for catalytic reaction and allosteric control.

Subfunctionalization of NbUGT72AY1

Evolutionary innovations arise from gene duplications whereby the copies undergo different fates during evolution. First, one copy can retain the original function so that the other copy is freed from purifying selection and can be altered by mutations that create a new function for the gene (neofunctionalization). Second, if the original gene had multiple activities, duplication can result in the different functions being shared, and sometimes optimized, among the different copies (subfunctionalization). Third, the two copies may retain the same function, resulting in redundancy and/or increased activity of the gene (gene dosage) (Voordeckers and Verstrepen, 2015).

NbUGT72AY1 likely arose from a promiscuous monolignol UGT and probably optimized some of its additional activities (subfunctionalization) (Glasner et al., 2020), as this enzyme can also effectively glucosylate coumarin derivatives, although it exhibits strong substrate inhibition for these substrates (Figure 1). Both the double mutant V83I_I84F and the single mutant Y389H of StUGT72AY2 (monolignol UGT) had significantly increased scopoletin glycosylation activity, but both modifications also led to substrate inhibition (Figure 5B). By contrast, monolignol UGT activity of StUGT72AY2 was almost unaffected (Figure 5D). The NbUGT72AY1/StUGT72AY2 example shows that minor changes in amino acid sequences can increase the catalytic activity of a substrate without compromising existing functions. Subfunctionalization in the case of NbUGT72AY1 results in a multifunctional enzyme with altered substrate preference, activity, and kinetics from a promiscuous UGT that can be shared by lignin and scopoletin metabolism. This result supports the hypothesis that the enzymes of natural product biosynthesis are trapped in the generalist state (Noda-Garcia and Tawfik, 2020).

One hundred NbUGT72AY1-like protein sequences were extracted from Solanaceae databases using BLAST. After manual editing based on sequence lengths, gaps, and insertions, 62 sequences were aligned, and a phylogenetic tree was generated with *Ipomoea* sequences as outgroups (Supplemental Figures 19 and 20; Figure 7). The sequences grouped into four classes: the *Ipomoea* outgroup and three other classes that sub-clustered into *Nicotiana*, *Solanum*, *Datura*, and *Capsicum* sub-groups. NbUGT72AY1 (class 2) and the most similar UGTs from other *Nicotiana* species, including a sequence from *Datura stramonium* (class 2), have Phe exclusively at position 87 (position 91 in the consensus sequence), the variant that has the highest scopoletin UGT activity. The related UGTs from *Solanum* (class 2) and *Ipomoea* species (class 1) have Phe, Val, Ile, and Leu at this position, whereas proteins from classes 3 and 4, including other *Nicotiana* species, do not show Phe at position 87 (91 in the consensus sequence). Phe87 thus appears to have been fixed in NbUGT72AY1 and its closest relatives by selection pressure during evolution and arose from a precursor that probably contained a different amino acid at this position, as indicated by the variability of the amino acid in related sequences. We assume that after duplication of a promiscuous precursor gene, one gene copy could mutate owing to lack of selection pressure, resulting in different enzyme variants with altered substrate preference, activity, and kinetics (substrate inhibition). Because of the need to effectively glucosylate scopoletin, the Phe variant could then have been positively selected in *Nicotiana* species. The large number of NbUGT72AY1-like sequences in *Nicotiana* species corroborates this hypothesis (Figure 7). Similarly, the His390 mutation may have been co-fixed (position 406 in the consensus sequence) with Phe87 (Figure 7). Moreover, the promiscuity of secondary metabolism UGTs argues against the mechanism of neofunctionalization. It is therefore very likely that the precursor exhibited substrate tolerance, whereas the diversity of similar but not functionally identical UGTs contradicts the gene dosage mechanism. Co-expression analysis of the *NbUGT72AY1* ortholog in an *N. attenuata* transcriptome database revealed a feruloyl ortho-hydroxylase-1-like gene as the most similarly expressed (Supplemental Figure 21). The encoded enzyme catalyzes the first reaction of the scopoletin biosynthetic pathway (Kai et al., 2008) and confirms the involvement of NbUGT72AY1 in the formation of scopoletin glucoside.

Tobacco plants produce the antifungal phytoalexin scopoletin in response to pathogen attack (Sun et al., 2014), and glycosylation appears to confer a selective advantage on *N. benthamiana*, as it probably protects the plant from the phytotoxin (Graña et al., 2017). Overexpression of a different scopoletin UGT in *N. tabacum* (TOGT; UGT73 homolog) resulted in early lesion formation during the hypersensitive response to tobacco mosaic virus but had no effect on virus resistance (Gachon et al., 2004). In a similar study, transgenic tobacco plants overexpressing TOGT showed similar formation of necrotic leaf lesions after inoculation with potato virus Y but significantly decreased levels of virus coat protein compared with control

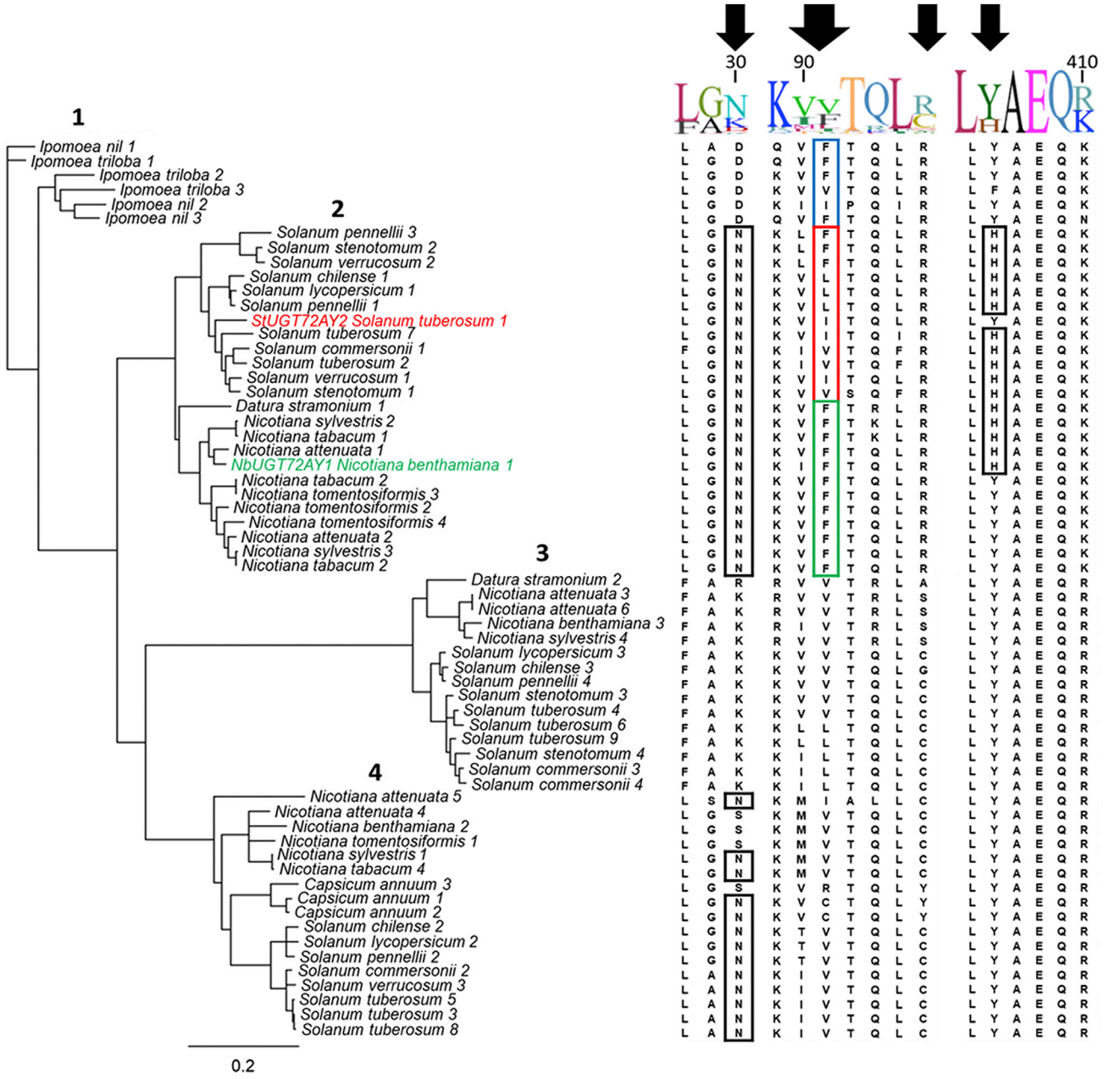


Figure 7. Phylogenetic analysis of the closest homologs of NbUGT72AY1.

The tree shows the plant species from which NbUGT72A1-like protein sequences were extracted by BLAST (blast.ncbi.nlm.nih.gov/Blast.cgi). Regions of the amino acid sequence alignment are shown, with the positions of the mutated amino acids indicated by black arrows. Accession numbers are included in the supplementary information (Supplemental Figure 20).

plants (Matros and Mock, 2004). These controversial results may have been obtained because tobacco plants express multiple UGTs that glucosylate hydroxycoumarins (Sun et al., 2019). Control mechanisms of the redundant UGT activities could have had different effects in the respective experiments.

Possible biological function of substrate inhibition

Enzymes of many classes are inhibited by their own substrates (Wu, 2011). Substrate inhibition has often been regarded as a biochemical curiosity and an experimental annoyance; this has changed, and substrate inhibition is now believed to have an important biological function (Reed et al., 2010). Scopoletin exhibits strong antifungal activity against *Alternaria alternata,* a necrotrophic fungus that causes severe yield losses in *Nicotiana* species *in vitro* and *in vivo* (Sun et al., 2014). The phenolic compound accumulates in large quantities during the hypersensitive response to plant pathogens, together with its glucoside (Gachon et al., 2004), but it is phytotoxic (Graña et al., 2017). Therefore, it is conceivable that at low scopoletin concentrations, neighboring plant cells protect themselves against the phytoalexin by glucosylation, whereas at higher

pathogen infestations and scopoletin concentrations, it is more advantageous for the plants that the cells die and thus form a barrier against the invader. The role of the tobacco glucosyltransferase TOGT that catalyzes the glucosylation of scopoletin has been studied by both loss- and gain-of-function approaches (Chong et al., 2002; Gachon et al., 2004; Matros and Mock, 2004). Although the results were not consistent, it appears that TOGT activity represents an important step in the containment of viral pathogens. Similarly, leaf necrosis was recently shown to result from downregulation of the poplar glycosyltransferase UGT72A2 (Behr et al., 2022).

Although scopoletin levels ranging from 0.07 ng/g fresh weight (FW) to 139 μg/g FW have been quantified in several plant species (Gnonlonfin et al., 2012), healthy, uninfected *Nicotiana* leaves contain only 7.3 ng/g FW of hydroxycoumarin (Großkinsky et al., 2013). However, in tobacco plants, tissues exhibiting localized acquired resistance (LAR) induced by tobacco mosaic virus or an elicitor accumulate 20–40 μg/g FW scopoletin (Costet et al., 2002; Ménard et al., 2004). From this, conservative estimation yields a concentration of 30 nM scopoletin in uninfected tobacco leaves and 100–200 μM in LAR leaves, suggesting that substrate inhibition of NbUGT72AY1 is relevant in the natural context.

Although NbUGT72AY1 and StUGT72AY2 glucosylate monolignols with similar effectiveness, they show very different activities and kinetics when converting scopoletin and other low molecular weight phenols (Figure 1). NbUGT72AY1 shows significantly higher turnover rates at very low hydroxycoumarin concentrations but behaves similarly to StUGT72AY2 at high scopoletin concentrations. The monolignol StUGT72AY2 exhibits little side activity for scopoletin, whereas NbUGT72AY1 has been additionally adapted to hydroxycoumarin by mutations. The effective detoxification of low levels of scopoletin appears to confer a selective advantage to tobacco plants in this regard because scopoletin not only possesses antifungal, antiviral, and cytotoxic properties but also behaves in an auxin-like manner (Graña et al., 2017). Finally, with the evolutionary acquisition of the allosteric property, the ability to regulate enzyme activity is also acquired (LiCata and Allewell, 1997; Cornish-Bowden, 2014).

MATERIALS AND METHODS

Cloning of *UGT72AY1* and *UGT72AY2* and production of the mutant proteins

Cloning of *NbUGT72AY1* from *Nicotiana benthamiana* (accession MT945401) and production of the NbUGT72AY1 protein was performed according to Sun et al. (2019). Genewiz, Leipzig, Germany (www.genewiz.com) synthesized *StUGT72AY2* (PGSC0003DMG401004500). The gene was ligated via *Bgl*II at the 5′ end and the *Xho*I site at the 3′ end into the pGEX-4T-1 vector. *NbUGT72AY1-I86V_F87I*, *-H390Y*, *-N27D*, *-R91A*, *-R91M*, *-R91F*, *-I86V*, and *-F87I* and *StUGT72AY2-V83I_I84F*, *-N27D*, and *-Y389H* were generated by site-directed mutagenesis following the QuickChange protocol (Agilent Technology, Santa Clara, CA). The temperature program was 3 min at 94°C, one cycle; 30 s at 94°C, 30 s at 65°C, 8 min at 72°C, 30 cycles; 10 min at 72°C, one cycle; and a final temperature of 4°C, using appropriate primers (Supplemental Table 3). After Dpn I digestion of templates, the PCR products were transformed into *E. coli* NEB 10 beta, followed by colony PCR and sequence confirmation. Cloning of *NbUGT72AY1-chimera* (Tobacco-Potato-Tobacco) and *StUGT72AY2-chimera* (Potato-Tobacco-Potato) involved six steps. First, the target amino acids were divided into three parts using the following PCR program and primers shown in Supplemental Table 1: 3 min at 94°C, one cycle; 30 s at 94°C, 40 s at 60°C, 1.5 min at 72°C, 40 cycles; 10 min at 72°C, one cycle; and a final temperature of 4°C. Then, the first part and the final part were separately joined to the middle part using the same program to obtain two longer fragments. Finally, the two longer fragments were used as templates to generate the target fragments. After gel extraction and Vector PGEX-4T1 DNA digestion, ligation was performed according to (Sun et al., 2020).

Protein production

Protein expression was performed using *E. coli* BL21(DE3) pLysS cells transformed with pGEX-4T-1 UGT72AY1, pGEX-4T-1 UGT72AY2, or their corresponding mutants. After pre-culturing overnight at 37°C and 150 rpm in Luria-Bertani medium containing 100 μg/mL ampicillin and 34 μg/mL chloramphenicol, 10 mL of the pre-culture was added to 1 L of the main culture containing the corresponding antibiotics and incubated at 37°C and 120 rpm until OD600 reached 1 in a chicane flask. For UGT72AY2 and its mutants, the cells were cultured in a 5-L fermenter at 600 rpm and pH 7.0. Gene expression was induced with 1 mM isopropyl-β-D-thiogalactopyranoside, and cultures were incubated overnight at 18°C and 150 rpm. Cells were harvested via centrifugation and stored at −80°C. Recombinant fusion proteins with an N-terminal GST tag were purified with Novagen GST Bind Resin following the manufacturer's instructions. After resuspension, the cells were disrupted by sonication. After centrifugation, the crude protein extract was incubated overnight at 4°C with the resin to bind the GST fusion protein and eluted with GST elution buffer containing reduced glutathione. The quality of the purified proteins was verified by SDS–PAGE (Supplemental Figure 22), and the protein concentration was determined with Roti-Nanoquant (Carl Roth, Karlsruhe, Germany) in 96-well microtiter plates according to the manufacturer's instructions. Absorption was measured at 450 nm and 590 nm using a CLARIOstar plate reader (BMG Labtech, Germany).

UDP-Glo glycosyltransferase assay

The enzymatic reaction was performed according to (Sun et al., 2019) with minor modifications and quantified by the UDP-Glo Glycosyltransferase Assay (Promega, Mannheim, Germany). Assays with NbUGT72AY1 and its mutants were performed at 40°C for 10 min in 50 mM Tris-HCl (pH 7.5) containing 100 μM UDP-glucose, substrate (dissolved in DMSO), and 0.5 μg purified protein, made up to 100 μL with water. The assays with StUGT72AY2 and its mutants were performed at 30°C for 20 min in 50 mM Tris-HCl (pH 7.0) containing 100 μM UDP-glucose, substrate (dissolved in DMSO), and 2 μg purified protein, made up to 100 μL with water. Controls did not contain the enzyme. The reaction was stopped by addition of 12.5 μL 0.6 M HCl and further neutralization with 1 M Trizma base. Five microliters of the UGT reaction was pipetted into a 384-well plate. The luminescence reaction was started by adding 5 μL UDP-Glo detection reagent and incubated for 30 min in the dark. The luminescence signal was detected with a CLARIOstar plate reader (McGraphery and Schwab, 2020). The calculation of kinetic data was performed with KaleidaGraph (https://www.synergy.com/; v4.5.4). The pH optima of NbUGT72AY1 and StUGT72AY2 were determined using different protein amounts from 0.5 μg to 4 μg and different buffers for various pH ranges: 50 mM sodium phosphate (pH 6–8) and 50 mM Tris-HCl (pH 6–8.5) in 0.5-unit intervals. The optimal temperature was evaluated from 15 to 45°C. The optimal time was measured from 5 min to 30 min.

Enzyme kinetics analysis

Normal kinetics data were fitted to the Michaelis–Menten equation (Equation 1).

$$v = \frac{vV_{max} * \frac{[S]}{K_m}}{1 + \frac{[S]}{K_m}} \quad \text{(Equation 1)}$$

In this equation, [*S*] is the concentration of the varied substrate, V_{max} represents the maximal reaction rate, and K_m is the substrate concentration

at which the reaction rate is half of V_{max}. An enzyme is under allosteric control if the binding of a molecule at one binding site alters the affinity of the enzyme for its substrate and thus regulates the enzyme activity. In this case, the Hill equation applies (Equation 2).

$$v = \frac{V_{max} * \frac{[S]^n}{K_m^n}}{1 + \frac{[S]^n}{K_m^n}} \quad \text{(Equation 2)}$$

Here, n is the Hill coefficient indicating the degree of cooperativity. Positive cooperativity ($n > 1$) occurs when an enzyme has more than one site to which a substrate can bind, and the binding of one molecule increases the rate of binding of other substrates. No or negative cooperativity is observed if $n = 1$ or $n < 1$, respectively. A partial uncompetitive inhibition model (Equation 3) was used to analyze atypical Michaelis–Menten substrate inhibition data (Wu, 2011).

$$v = \frac{V_{max} * \frac{[S]}{K_m} + V_i * \frac{[S]*[S]}{K_m * K_i}}{1 + \frac{[S]}{K_m} + \frac{[S]*[S]}{K_m * K_i}} \quad \text{(Equation 3)}$$

Here, a two-site model is assumed to explain the substrate inhibition phenomenon (Scheme 1; $n = x = 1$). The parameter V_i is the reaction velocity in the presence of inhibition, K_i is the inhibition constant, which is the inhibitor concentration required to decrease the maximal rate of the reaction to half of the uninhibited value. The equation presumes the sequential binding of substrate molecules, i.e., the inhibitory site cannot be occupied until the reaction site is filled. Combining the cooperativity-describing Hill equation and the partial uncompetitive inhibition model yields Equation 4, which best described the measured data (LiCata and Allewell, 1997; Kapelyukh et al., 2008; Peng et al., 2016).

$$v = \frac{V_{max} * \frac{[S]^n}{K_m^n} + V_i * \frac{[S]^n * [S]^x}{K_m^n * K_i^x}}{1 + \frac{[S]^n}{K_m^n} + \frac{[S]^n * [S]^x}{K_m^n * K_i^x}} \quad \text{(Equation 4)}$$

The superscript n is a Hill coefficient, and x is another Hill coefficient that allows for the possibility that binding of substrate in the inhibitory mode may also be cooperative (Peng et al., 2016). To obtain convergence for Equation 4, the value of x was fixed to an integral number that was determined empirically to give the best fit (lowest variance). The general reaction scheme for Equation 4 looks as follows (LiCata and Allewell, 1997).

The maximal reaction rates V_{max} and V_i correspond to the catalytic constants k_{cat} and $k_{cat(i)}$, respectively. The kinetic parameters were determined under optimum conditions and were calculated with KaleidaGraph v4.5.4. The data were derived from at least three repeats. Statistical analysis was performed using Student's t test ($p < 0.05$).

LC-MS analysis

For the initial substrate screening, the UGT reaction was performed in a final volume of 100 μL 100 mM Tris-HCl buffer (pH 7.5) containing 5 μg of purified recombinant protein, 1 mM UDP-Glc, and 600 μM substrate dissolved in DMSO. The reaction was incubated at 30°C with constant shaking at 400 rpm overnight. The enzymatic reaction was stopped by heat inactivation for 10 min at 95°C. After centrifugation, the supernatant was analyzed via LC-MS analysis performed according to (Huang et al., 2018). Products were identified using authentic reference materials (Supplemental Table 4 and Supplemental Figure 2).

HDX-MS

Ligands were used as 50-mM concentrated stock solutions dissolved in DMSO (scopoletin) or double-distilled water (UDP). Prior to HDX-MS, 194 μL of purified NbUGT72AY1 was mixed with 2 μL of ligand stock solution or solvent to reach final concentrations of 35 μM (NbUGT72AY1) and 500 μM (ligands), then incubated for 5 min at ambient temperature. Preparation of exchange reactions for HDX-MS was aided by a two-arm robotic autosampler (LEAP Technologies). The protein solution (7.5 μL) was mixed with 67.5 μL of D_2O-containing buffer (20 mM Tris-Cl [pH 7.5], 2% [v/v] DMSO), which also contained 500 μM of ligands to prevent their dilution during HDX, and incubated for 10/30/95/1000/10 000 s at 25°C. The completed HDX reactions were added to 55 μL quench solution (400 mM KH_2PO_4/H_3PO_4, 2 M guanidine-HCl, pH 2.2) kept at 1°C, and 95 μL of the resulting mixture was injected into an ACQUITY UPLC M-class system with HDX technology (Waters) (Wales et al., 2008). NbUGT72AY1 was digested online with a column (2 mm × 2 cm) packed with immobilized porcine pepsin and separated by reversed-phase HPLC followed by mass spectrometric analysis as described previously (Osorio-Valeriano et al., 2019; Skotnicka et al., 2020). Peptides were identified from the undeuterated samples with ProteinLynx Global Server (Waters) software as described previously (Osorio-Valeriano et al., 2019). For quantification of deuterium incorporation with DynamX 3.0 (Waters), peptides had to fulfill the following criteria: minimum intensity of 5000 counts; maximum length of 30 amino acids; minimum number of products of three; maximum mass error of 25 ppm; and retention time tolerance of 0.5 min. After automated data processing with DynamX 3.0, the mass spectra were manually inspected and, if necessary, peptides were omitted, e.g., in the case of low signal-to-noise ratio or presence of overlapping peptides. Raw HDX-MS data are supplied in the supplemental dataset (Masson et al., 2019).

In silico homology modeling and ligand docking

Three-dimensional structure homology models of NbUGT71AY1 (GenBank accession MT945401; Niben101Scf06112g01008.1; https://solgenomics.net/tools/blast) and StUGT72AY2 (GenBank accession XP_015164078.1, UniProt: M0ZZL3) were produced using the IntFOLD (McGuffin et al., 2015) and Swiss-Model (Waterhouse et al., 2018) servers. Four proteins with known crystal structures were found that show sequence identities between 39.4% and 47.0% with NbUGT72AY1, and StUGT72AY2 shares 80.3% identical amino acids with NbUGT72AY1. Two each were crystalized in their open (PDB: 6SU6/6SU7/5NLM and PDP: 6JEL/6JEM/6JEN) and closed (PDB: 6JTD and PDB: 2VG8/2VCE/2VCH) conformation. Therefore, PDB: 6JTD and PDB: 6SU6 were used as templates to guide the modeling of both NbUGT72AY1 and StUGT72AY2, each in the closed and open form. Models with the highest confidence scores were uploaded into UCSF Chimera 1.15 (https://www.cgl.ucsf.edu/chimera) for visualization and comparative analysis (Pettersen et al., 2004; Goddard et al., 2018). Ligand docking was performed with the AutoDock Vina tool implemented in UCSF Chimera 1.15 (Trott and Olson, 2010). Binding energies (ΔG) calculated by UCSF Chimera 1.15 were used to calculate equilibrium constants K_D by $K_d = e^{-\Delta G/R/T}$ and $K_D = K_d/c$ with R = 1.986 cal/mol/K, T = 298.15 K, and the standard reference concentration c = 1 mol/L. Videos (mp4) were recorded with UCSF Chimera 1.15.

DATA AVAILABILITY

All relevant data generated in this study are provided in the main text and the supplemental information.

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Plant Communications Online*.

FUNDING

DFG SCHW 634/34-1.

AUTHOR CONTRIBUTIONS

J.L., G.S., E.K., T.H., and W.G.S. designed experiments and interpreted data; J.L., G.S., E.K., T.H., T.D.H., C.S., and W.G.S. wrote the manuscript; J.L., G.S., and E.K. performed biochemical analyses and analyzed the proteins; J.L. and E.K. generated mutant proteins; J.L., G.S., E.K., and T.H. performed LC-MS analyses. W.S. performed HDX-MS analyses; W.G.S. performed *in silico* analyses; C.S., T.D.H., and W.S. contributed

to materials, instruments, data analysis, and the manuscript. All authors approved the final manuscript.

ACKNOWLEDGMENTS

We thank DFG for financial support. No conflict of interest declared.

Received: April 27, 2022
Revised: November 16, 2022
Accepted: December 16, 2022
Published: December 24, 2022

REFERENCES

Albesa-Jové, D., and Guerin, M.E. (2016). The conformational plasticity of glycosyltransferases. Curr. Opin. Struct. Biol. **40**:23–32.

Albesa-Jové, D., Romero-García, J., Sancho-Vaello, E., Contreras, F.-X., Rodrigo-Unzueta, A., Comino, N., Carreras-González, A., Arrasate, P., Urresti, S., Biarnés, X., et al. (2017). Structural snapshots and loop dynamics along the catalytic cycle of glycosyltransferase GpgS. Structure **25**:1034–1044.e3.

Baldacci-Cresp, F., Le Roy, J., Huss, B., Lion, C., Créach, A., Spriet, C., Duponchel, L., Biot, C., Baucher, M., Hawkins, S., et al. (2020). UDP-glycosyltransferase 72E3 plays a role in lignification of secondary cell walls in Arabidopsis. Int. J. Mol. Sci. **21**:6094.

Barnett, A.C., Tsvetanov, S., Gamage, N., Martin, J.L., Duggleby, R.G., and McManus, M.E. (2004). Active site mutations and substrate inhibition in human sulfotransferase 1A1 and 1A3. J. Biol. Chem. **279**:18799–18805.

Behr, M., Speeckaert, N., Kurze, E., Morel, O., Prévost, M., Mol, A., Mahamadou Adamou, N., Baragé, M., Renaut, J., Schwab, W., et al. (2022). Leaf necrosis resulting from downregulation of poplar glycosyltransferase UGT72A2. Tree Physiol. **42**:1084–1099.

Bolam, D.N., Roberts, S., Proctor, M.R., Turkenburg, J.P., Dodson, E.J., Martinez-Fleites, C., Yang, M., Davis, B.G., Davies, G.J., and Gilbert, H.J. (2007). The crystal structure of two macrolide glycosyltransferases provides a blueprint for host cell antibiotic immunity. Proc. Natl. Acad. Sci. USA **104**:5336–5341.

Bowles, D., Lim, E.-K., Poppenberger, B., and Vaistij, F.E. (2006). Glycosyltransferases of lipophilic small molecules. Annu. Rev. Plant Biol. **57**:567–597.

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). Characterization and engineering of the bifunctional N- and O-glucosyltransferase involved in xenobiotic metabolism in plants. Proc. Natl. Acad. Sci. USA **104**:20238–20243.

Breton, C., Snajdrová, L., Jeanneau, C., Koca, J., and Imberty, A. (2006). Structures and mechanisms of glycosyltransferases. Glycobiology **16**:29R–37R.

Buenavista, M.T., Roche, D.B., and McGuffin, L.J. (2012). Improvement of 3D protein models using multiple templates guided by single-template model quality assessment. Bioinformatics **28**:1851–1857.

Chang, A., Singh, S., Phillips, G.N., and Thorson, J.S. (2011). Glycosyltransferase structural biology and its role in the design of catalysts for glycosylation. Curr. Opin. Biotechnol. **22**:800–808.

Chen, C., Joo, J.C., Brown, G., Stolnikova, E., Halavaty, A.S., Savchenko, A., Anderson, W.F., and Yakunin, A.F. (2014). Structure-based mutational studies of substrate inhibition of betaine aldehyde dehydrogenase BetB from Staphylococcus aureus. Appl. Environ. Microbiol. **80**:3992–4002.

Chong, J., Baltz, R., Schmitt, C., Beffa, R., Fritig, B., and Saindrenan, P. (2002). Downregulation of a pathogen-responsive tobacco UDP-Glc:phenylpropanoid glucosyltransferase reduces scopoletin glucoside accumulation, enhances oxidative stress, and weakens virus resistance. Plant Cell **14**:1093–1107.

Cornish-Bowden, A. (2014). Understanding allosteric and cooperative interactions in enzymes. FEBS J. **281**:621–632.

Costet, L., Fritig, B., and Kauffmann, S. (2002). Scopoletin expression in elicitor-treated and tobacco mosaic virus-infected tobacco plants. Physiol. Plant. **115**:228–235.

Dong, D., and Wu, B. (2012). What are the real causes of substrate inhibition in the glucuronidation reaction? Single Cell Biol. **1**:e105.

Drula, E., Garron, M.-L., Dogan, S., Lombard, V., Henrissat, B., and Terrapon, N. (2022). The carbohydrate-active enzyme database: functions and literature. Nucleic Acids Res. **50**:D571–D577.

Flocco, M.M., and Mowbray, S.L. (1994). Planar stacking interactions of arginine and aromatic side-chains in proteins. J. Mol. Biol. **235**:709–717.

Fujiwara, R., Yokoi, T., and Nakajima, M. (2016). Structure and protein-protein interactions of human UDP-glucuronosyltransferases. Front. Pharmacol. **7**:388.

Gachon, C., Baltz, R., and Saindrenan, P. (2004). Over-expression of a scopoletin glucosyltransferase in Nicotiana tabacum leads to precocious lesion formation during the hypersensitive response to tobacco mosaic virus but does not affect virus resistance. Plant Mol. Biol. **54**:137–146.

Gallivan, J.P., and Dougherty, D.A. (1999). Cation-pi interactions in structural biology. Proc. Natl. Acad. Sci. USA **96**:9459–9464.

Glasner, M.E., Truong, D.P., and Morse, B.C. (2020). How enzyme promiscuity and horizontal gene transfer contribute to metabolic innovation. FEBS J. **287**:1323–1342.

Gnonlonfin, G.J.B., Sanni, A., and Brimer, L. (2012). Review scopoletin – a coumarin phytoalexin with medicinal properties. Crit. Rev. Plant Sci. **31**:47–56.

Goddard, T.D., Huang, C.C., Meng, E.C., Pettersen, E.F., Couch, G.S., Morris, J.H., and Ferrin, T.E. (2018). UCSF ChimeraX: meeting modern challenges in visualization and analysis. Protein Sci. **27**:14–25.

Graña, E., Costas-Gil, A., Longueira, S., Celeiro, M., Teijeira, M., Reigosa, M.J., and Sánchez-Moreiras, A.M. (2017). Auxin-like effects of the natural coumarin scopoletin on Arabidopsis cell structure and morphology. J. Plant Physiol. **218**:45–55.

Großkinsky, D.K., Edelsbrunner, K., Pfeifhofer, H., van der Graaff, E., and Roitsch, T. (2013). Cis- and trans-zeatin differentially modulate plant immunity. Plant Signal. Behav. **8**:e24798.

Harrus, D., Khoder-Agha, F., Peltoniemi, M., Hassinen, A., Ruddock, L., Kellokumpu, S., and Glumoff, T. (2018). The dimeric structure of wild-type human glycosyltransferase B4GalT1. PLoS One **13**:e0205571.

Hashimoto, K., Madej, T., Bryant, S.H., and Panchenko, A.R. (2010). Functional states of homooligomers: insights from the evolution of glycosyltransferases. J. Mol. Biol. **399**:196–206.

He, J.-B., Zhao, P., Hu, Z.-M., Liu, S., Kuang, Y., Zhang, M., Li, B., Yun, C.-H., Qiao, X., and Ye, M. (2019). Molecular and structural characterization of a promiscuous C-glycosyltransferase from Trollius chinensis. Angew. Chem. Int. Ed. Engl. **58**:11513–11520.

Hsu, T.M., Welner, D.H., Russ, Z.N., Cervantes, B., Prathuri, R.L., Adams, P.D., and Dueber, J.E. (2018). Employing a biochemical protecting group for a sustainable indigo dyeing strategy. Nat. Chem. Biol. **14**:256–261.

Huang, F.-C., Giri, A., Daniilidis, M., Sun, G., Härtl, K., Hoffmann, T., and Schwab, W. (2018). Structural and functional analysis of UGT92G6 suggests an evolutionary link between mono- and disaccharide glycoside-forming transferases. Plant Cell Physiol. **59**:857–870.

Kai, K., Mizutani, M., Kawamura, N., Yamamoto, R., Tamai, M., Yamaguchi, H., Sakata, K., and Shimizu, B.i. (2008). Scopoletin is

biosynthesized via ortho-hydroxylation of feruloyl CoA by a 2-oxoglutarate-dependent dioxygenase in Arabidopsis thaliana. Plant J. **55**:989–999.

Kapelyukh, Y., Paine, M.J.I., Maréchal, J.D., Sutcliffe, M.J., Wolf, C.R., and Roberts, G.C.K. (2008). Multiple substrate binding by cytochrome P450 3A4: estimation of the number of bound substrate molecules. Drug Metab. Dispos. **36**:2136–2144.

Konermann, L., Pan, J., and Liu, Y.-H. (2011). Hydrogen exchange mass spectrometry for studying protein structure and dynamics. Chem. Soc. Rev. **40**:1224–1234.

Kurze, E., Wüst, M., Liao, J., McGraphery, K., Hoffmann, T., Song, C., and Schwab, W. (2021). Structure-function relationship of terpenoid glycosyltransferases from plants. Nat. Prod. Rep. **39**:389–409.

Lanot, A., Hodge, D., Jackson, R.G., George, G.L., Elias, L., Lim, E.-K., Vaistij, F.E., and Bowles, D.J. (2006). The glucosyltransferase UGT72E2 is responsible for monolignol 4-O-glucoside production in Arabidopsis thaliana. Plant J. **48**:286–295.

Lanot, A., Hodge, D., Lim, E.-K., Vaistij, F.E., and Bowles, D.J. (2008). Redirection of flux through the phenylpropanoid pathway by increased glucosylation of soluble intermediates. Planta **228**:609–616.

Le Roy, J., Huss, B., Creach, A., Hawkins, S., and Neutelings, G. (2016). Glycosylation is a major regulator of phenylpropanoid availability and biological activity in plants. Front. Plant Sci. **7**:735.

Liang, D.-M., Liu, J.-H., Wu, H., Wang, B.-B., Zhu, H.-J., and Qiao, J.-J. (2015). Glycosyltransferases: mechanisms and applications in natural product development. Chem. Soc. Rev. **44**:8350–8374.

LiCata, V.J., and Allewell, N.M. (1997). Is substrate inhibition a consequence of allostery in aspartate transcarbamylase? Biophys. Chem. **64**:225–234.

Lim, E.-K., Jackson, R.G., and Bowles, D.J. (2005). Identification and characterisation of Arabidopsis glycosyltransferases capable of glucosylating coniferyl aldehyde and sinapyl aldehyde. FEBS Lett. **579**:2802–2806.

Liu, Y.-Q., Yuan, L.-M., Gao, Z.-Z., Xiao, Y.-S., Sun, H.-Y., Yu, L.-S., and Zeng, S. (2016). Dimerization of human uridine diphosphate glucuronosyltransferase allozymes 1A1 and 1A9 alters their quercetin glucuronidation activities. Sci. Rep. **6**:23763.

Luukkanen, L., Taskinen, J., Kurkela, M., Kostiainen, R., Hirvonen, J., and Finel, M. (2005). Kinetic characterization of the 1A subfamily of recombinant human UDP-glucuronosyltransferases. Drug Metab. Dispos. **33**:1017–1026.

Maharjan, R., Fukuda, Y., Shimomura, N., Nakayama, T., Okimoto, Y., Kawakami, K., Nakayama, T., Hamada, H., Inoue, T., and Ozaki, S.-I. (2020). An ambidextrous polyphenol glycosyltransferase PaGT2 from Phytolacca americana. Biochemistry **59**:2551–2561.

Masson, G.R., Burke, J.E., Ahn, N.G., Anand, G.S., Borchers, C., Brier, S., Bou-Assaf, G.M., Engen, J.R., Englander, S.W., Faber, J., et al. (2019). Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments. Nat. Methods **16**:595–602.

Matros, A., and Mock, H.-P. (2004). Ectopic expression of a UDP-glucose:phenylpropanoid glucosyltransferase leads to increased resistance of transgenic tobacco plants against infection with Potato Virus Y. Plant Cell Physiol. **45**:1185–1193.

McGraphery, K., and Schwab, W. (2020). Comparative analysis of high-throughput assays of family-1 plant glycosyltransferases. Int. J. Mol. Sci. **21**:2208.

McGuffin, L.J., Adiyaman, R., Maghrabi, A.H.A., Shuid, A.N., Brackenridge, D.A., Nealon, J.O., and Philomina, L.S. (2019). IntFOLD: an integrated web resource for high performance protein structure and function prediction. Nucleic Acids Res. **47**:W408–W413.

McGuffin, L.J., Atkins, J.D., Salehe, B.R., Shuid, A.N., and Roche, D.B. (2015). IntFOLD: an integrated server for modelling protein structures and functions from amino acid sequences. Nucleic Acids Res. **43**:W169–W173.

Meech, R., Hu, D.G., McKinnon, R.A., Mubarokah, S.N., Haines, A.Z., Nair, P.C., Rowland, A., and Mackenzie, P.I. (2019). The UDP-Glycosyltransferase (UGT) superfamily: new members, new functions, and novel paradigms. Physiol. Rev. **99**:1153–1222.

Ménard, R., Alban, S., de Ruffray, P., Jamois, F., Franz, G., Fritig, B., Yvin, J.-C., and Kauffmann, S. (2004). Beta-1, 3 glucan sulfate, but not beta-1, 3 glucan, induces the salicylic acid signaling pathway in tobacco and Arabidopsis. Plant Cell **16**:3020–3032.

Miller, G.P., Lichti, C.F., Zielinska, A.K., Mazur, A., Bratton, S.M., Gallus-Zawada, A., Finel, M., Moran, J.H., and Radominska-Pandya, A. (2008). Identification of hydroxywarfarin binding site in human UDP glucuronosyltransferase 1a10: phenylalanine90 is crucial for the glucuronidation of 6- and 7-hydroxywarfarin but not 8-hydroxywarfarin. Drug Metab. Dispos. **36**:2211–2218.

Noda-Garcia, L., and Tawfik, D.S. (2020). Enzyme evolution in natural products biosynthesis: target- or diversity-oriented? Curr. Opin. Chem. Biol. **59**:147–154.

Offen, W., Martinez-Fleites, C., Yang, M., Kiat-Lim, E., Davis, B.G., Tarling, C.A., Ford, C.M., Bowles, D.J., and Davies, G.J. (2006). Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification. EMBO J. **25**:1396–1405.

Osorio-Valeriano, M., Altegoer, F., Steinchen, W., Urban, S., Liu, Y., Bange, G., and Thanbichler, M. (2019). ParB-type DNA segregation proteins are CTP-dependent molecular switches. Cell **179**:1512–1524.e15.

Peng, H., Yang, T., Whitaker, B.D., Shangguan, L., and Fang, J. (2016). Calcium/calmodulin alleviates substrate inhibition in a strawberry UDP-glucosyltransferase involved in fruit anthocyanin biosynthesis. BMC Plant Biol. **16**:197.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera–a visualization system for exploratory research and analysis. J. Comput. Chem. **25**:1605–1612.

Putkaradze, N., Teze, D., Fredslund, F., and Welner, D.H. (2021). Natural product C-glycosyltransferases - a scarcely characterised enzymatic activity with biotechnological potential. Nat. Prod. Rep. **38**:432–443.

Qasba, P.K., Ramakrishnan, B., and Boeggeman, E. (2005). Substrate-induced conformational changes in glycosyltransferases. Trends Biochem. Sci. **30**:53–62.

Qasba, P.K., Ramakrishnan, B., and Boeggeman, E. (2008). Structure and function of β-1, 4-galactosyltransferase. Curr. Drug Targets **9**:292–309.

Reed, M.C., Lieb, A., and Nijhout, H.F. (2010). The biological significance of substrate inhibition: a mechanism with diverse functions. Bioessays **32**:422–429.

Sheikh, M.O., Halmo, S.M., Patel, S., Middleton, D., Takeuchi, H., Schafer, C.M., West, C.M., Haltiwanger, R.S., Avci, F.Y., Moremen, K.W., et al. (2017). Rapid screening of sugar-nucleotide donor specificities of putative glycosyltransferases. Glycobiology **27**:206–212.

Skinner, J.J., Lim, W.K., Bédard, S., Black, B.E., and Englander, S.W. (2012). Protein hydrogen exchange: testing current models. Protein Sci. **21**:987–995.

Skotnicka, D., Steinchen, W., Szadkowski, D., Cadby, I.T., Lovering, A.L., Bange, G., and Søgaard-Andersen, L. (2020). CdbA is a DNA-binding protein and c-di-GMP receptor important for nucleoid

organization and segregation in Myxococcus xanthus. Nat. Commun. **11**:1791.

Smith, A.D., Page, B.D.G., Collier, A.C., and Coughtrie, M.W.H. (2020). Homology modeling of human uridine-5'-diphosphate-glucuronosyltransferase 1A6 reveals insights into factors influencing substrate and cosubstrate binding. ACS Omega **5**:6872–6887.

Speeckaert, N., Adamou, N.M., Hassane, H.A., Baldacci-Cresp, F., Mol, A., Goeminne, G., Boerjan, W., Duez, P., Hawkins, S., Neutelings, G., et al. (2020). Characterization of the UDP-glycosyltransferase UGT72 family in poplar and identification of genes involved in the glycosylation of monolignols. Int. J. Mol. Sci. **21**:5018.

Speeckaert, N., El Jaziri, M., Baucher, M., and Behr, M. (2022). UGT72, a major glycosyltransferase family for flavonoid and monolignol homeostasis in plants. Biology **11**:441.

Steiner, T., and Koellner, G. (2001). Hydrogen bonds with pi-acceptors in proteins: frequencies and role in stabilizing local 3D structures. J. Mol. Biol. **305**:535–557.

Sun, G., Putkaradze, N., Bohnacker, S., Jonczyk, R., Fida, T., Hoffmann, T., Bernhardt, R., Härtl, K., and Schwab, W. (2020). Six uridine-diphosphate glycosyltransferases catalyze the glycosylation of bioactive C13-apocarotenols. Plant Physiol. **184**:1744–1761.

Sun, G., Strebl, M., Merz, M., Blamberg, R., Huang, F.-C., McGraphery, K., Hoffmann, T., and Schwab, W. (2019). Glucosylation of the phytoalexin N-feruloyl tyramine modulates the levels of pathogen-responsive metabolites in Nicotiana benthamiana. Plant J. **100**:20–37.

Sun, H., Wang, L., Zhang, B., Ma, J., Hettenhausen, C., Cao, G., Sun, G., Wu, J., and Wu, J. (2014). Scopoletin is a phytoalexin against Alternaria alternata in wild tobacco dependent on jasmonate signalling. J. Exp. Bot. **65**:4305–4315.

Teze, D., Coines, J., Fredslund, F., Dubey, K.D., Bidart, G.N., Adams, P.D., Dueber, J.E., Svensson, B., Rovira, C., and Welner, D.H. (2021). O -/N -/S -specificity in glycosyltransferase catalysis: from mechanistic understanding to engineering. ACS Catal. **11**:1810–1815.

Trott, O., and Olson, A.J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem. **31**:455–461.

Voordeckers, K., and Verstrepen, K.J. (2015). Experimental evolution of the model eukaryote Saccharomyces cerevisiae yields insight into the molecular mechanisms underlying adaptation. Curr. Opin. Microbiol. **28**:1–9.

Wales, T.E., Fadgen, K.E., Gerhardt, G.C., and Engen, J.R. (2008). High-speed and high-resolution UPLC separation at zero degrees Celsius. Anal. Chem. **80**:6815–6820.

Wang, Z.-L., Gao, H.-M., Wang, S., Zhang, M., Chen, K., Zhang, Y.-Q., Wang, H.-D., Han, B.-Y., Xu, L.-L., Song, T.-Q., et al. (2020). Dissection of the general two-step di-C-glycosylation pathway for the biosynthesis of (iso)schaftosides in higher plants. Proc. Natl. Acad. Sci. USA **117**:30816–30823.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., et al. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. **46**:W296–W303.

Weiss, D.R., and Levitt, M. (2009). Can morphing methods predict intermediate structures? J. Mol. Biol. **385**:665–674.

Wilson, A.E., and Tian, L. (2019). Phylogenomic analysis of UDP-dependent glycosyltransferases provides insights into the evolutionary landscape of glycosylation in plant metabolism. Plant J. **100**:1273–1288.

Wu, B. (2011). Substrate inhibition kinetics in drug metabolism reactions. Drug Metab. Rev. **43**:440–456.

Yonekura-Sakakibara, K., and Hanada, K. (2011). An evolutionary view of functional diversity in family 1 glycosyltransferases. Plant J. **66**:182–193.

Zhang, H., Zhu, F., Yang, T., Ding, L., Zhou, M., Li, J., Haslam, S.M., Dell, A., Erlandsen, H., and Wu, H. (2014). The highly conserved domain of unknown function 1792 has a distinct glycosyltransferase fold. Nat. Commun. **5**:4339.

4. Acceptors and effectors alter substrate inhibition kinetics of a plant glucosyltransferase NbUGT72AY1 and its mutants

Abstract

Building on previous research, the study of NbUGT72AY1 unravels further intricacies of its substrate inhibition properties and sheds light on potential allosteric binding sites for various ligands. The following results were obtained:

Diverse substrate inhibition profiles: The remarkable substrate inhibition behavior of NbUGT72AY1 extends to a range of compounds, including hydroxycoumarins, vanillin, carvacrol, and monolignols. The results of the UDP-Glo^{TM} glucosyltransferase assay revealed a wide range of different activity profiles, ranging from Michaelis-Menten (MM) kinetics (hydroquinones) to weak substrate inhibition (sinapaldehyde, guaiacol, and o-cresol) to strong substrate inhibition (scopoletin, eugenol, and vanillin). These results shed light on the peculiarities of the enzyme's catalytic behavior with different substrates.

Mutant-based insights: Through the utilization of mutants of NbUGT72AY1, the study identified potential commonalities in the binding sites of vanillin and scopoletin within the active site of the glucosyltransferase. This provides a molecular-level understanding of how these ligands interact with the enzyme and trigger substrate inhibition.

Effector selection: Different effectors were selected that could interact with NbUGT72AY1 during glucosylation of vanillin. While coumarins had a minimal effect on enzymatic activity, fatty acids emerged as significant modulators of substrate inhibition kinetics. This suggests that the second binding site model may not apply to NbUGT72AY1 and that fatty acids play a regulatory role in fine-tuning enzyme behaviours.

Conformational change model: In view of the results, a new model based on conformational change was proposed for NbUGT72AY1 catalysis in the presence of scopoletin. This model offers insights into how the enzyme undergoes structural alterations during the reaction with

scopoletin, contributing to a deeper understanding of the catalytic process.

In summary, the comprehensive investigation into NbUGT72AY1's substrate inhibition behavior, ligand interactions, and conformational changes opens up new avenues for studying the regulatory mechanisms of this enzyme.

This work has been published:

Title: Acceptors and Effectors Alter Substrate Inhibition Kinetics of a Plant Glucosyltransferase NbUGT72AY1 and Its Mutants

Authors: **Jieren Liao**, Veronika Lederer, Alba Bardhi, Zhiwei Zou, Timothy D. Hoffmann, Guangxin Sun, Chuankui Song, Thomas Hoffmann and Wilfried Schwab

First Published: 31 May 2023

Journal: International Journal of Molecular Sciences

Publisher: MDPI

Link: https://www.mdpi.com/1422-0067/24/11/9542/html

Conceptualization, **J.L**., V.L., T.D.H., G.S., T.H. and W.S.; methodology, T.H.; software, T.H. and W.S.; validation, T.D.H.; formal analysis, **J.L**., V.L., A.B., Z.Z., T.D.H. and G.S.; investigation, **J.L**., V.L., A.B., Z.Z. and G.S.; resources, W.S.; data curation, **J.L**., V.L., T.H. and W.S.; writing—original draft preparation, **J.L**. and T.D.H.; writing—review and editing, T.D.H., T.H., C.S. and W.S.; visualization, W.S.; supervision, **J.L**., T.D.H. and T.H.; project administration, W.S.; funding acquisition, W.S. All authors have read and agreed to the published version of the manuscript.

International Journal of
Molecular Sciences

MDPI

Article

Acceptors and Effectors Alter Substrate Inhibition Kinetics of a Plant Glucosyltransferase NbUGT72AY1 and Its Mutants

Jieren Liao [1], Veronika Lederer [1], Alba Bardhi [1], Zhiwei Zou [1], Timothy D. Hoffmann [1], Guangxin Sun [1,†], Chuankui Song [2], Thomas Hoffmann [1] and Wilfried Schwab [1,*]

[1] Biotechnology of Natural Products, Technische Universität München, Liesel-Beckmann-Str. 1, 85354 Freising, Germany; jieren.liao@tum.de (J.L.); veronika.lederer@tum.de (V.L.); aalbabardhi10@gmail.com (A.B.); ga96loj@mytum.de (Z.Z.); timothy.d.hoffmann@tum.de (T.D.H.); sgx_here@hotmail.com (G.S.); tom.hoffmann@tum.de (T.H.)
[2] State Key Laboratory of Tea Plant Biology and Utilization, International Joint Laboratory on Tea Chemistry and Health Effects, Anhui Agricultural University, Hefei 230036, China; sckfriend@163.com
* Correspondence: wilfried.schwab@tum.de
† Current address: Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06510, USA.

Abstract: One of the main obstacles in biocatalysis is the substrate inhibition (SI) of enzymes that play important roles in biosynthesis and metabolic regulation in organisms. The promiscuous glycosyltransferase UGT72AY1 from *Nicotiana benthamiana* is strongly substrate-inhibited by hydroxycoumarins (inhibitory constant $Ki < 20$ μM), but only weakly inhibited when monolignols are glucosylated ($Ki > 1000$ μM). Apocarotenoid effectors reduce the inherent UDP-glucose glucohydrolase activity of the enzyme and attenuate the SI by scopoletin derivatives, which could also be achieved by mutations. Here, we studied the kinetic profiles of different phenols and used the substrate analog vanillin, which has shown atypical Michaelis–Menten kinetics in previous studies, to examine the effects of different ligands and mutations on the SI of NbUGT72AY1. Coumarins had no effect on enzymatic activity, whereas apocarotenoids and fatty acids strongly affected SI kinetics by increasing the inhibition constant *Ki*. Only the F87I mutant and a chimeric version of the enzyme showed weak SI with the substrate vanillin, but all mutants exhibited mild SI when sinapaldehyde was used as an acceptor. In contrast, stearic acid reduced the transferase activity of the mutants to varying degrees. The results not only confirm the multi-substrate functionality of NbUGT72AY1, but also reveal that the enzymatic activity of this protein can be fine-tuned by external metabolites such as apocarotenoids and fatty acids that affect SI. Since these signals are generated during plant cell destruction, NbUGT72AY1 likely plays an important role in plant defense by participating in the production of lignin in the cell wall and providing direct protection through the formation of toxic phytoalexins.

Keywords: glycosyltransferase; vanillin; coumarin; sinapaldehyde; fatty acid; effector; substrate inhibition

Citation: Liao, J.; Lederer, V.; Bardhi, A.; Zou, Z.; Hoffmann, T.D.; Sun, G.; Song, C.; Hoffmann, T.; Schwab, W. Acceptors and Effectors Alter Substrate Inhibition Kinetics of a Plant Glucosyltransferase NbUGT72AY1 and Its Mutants. *Int. J. Mol. Sci.* **2023**, *24*, 9542. https://doi.org/10.3390/ijms24119542

Academic Editor: Maria Lourdes Gómez-Gómez

Received: 3 May 2023
Revised: 22 May 2023
Accepted: 25 May 2023
Published: 31 May 2023

1. Introduction

Glycosylation describes a biochemical reaction that strongly alters the physicochemical properties of small molecules, such as water solubility, stability, volatility, bioactivity, and bioavailability, and has proven to be a unique strategy in nature for broadening the chemical spectrum of natural products. This is an important prerequisite for the successful selection of adapted metabolic pathways [1,2]. Several enzyme families have been discovered that can form a variety of glycoside bonds [3,4]. Among others, UDP-dependent glycosyltransferases (UGTs), one of the largest protein families in plants, produce glycosides by transferring a sugar moiety from a donor to an acceptor molecule via an SN2-like mechanism, resulting in an inversion of the configuration of the anomeric carbon [5–8].

Donors include UDP-glucose, but also UDP-xylose, UDP-glucuronic acid, etc., and acceptors are, e.g., proteins, carbohydrates, lipids and low-molecular metabolites (small molecules) that carry an -OH, -COOH, -SH and -NH_2 group whereby *O*-, *S*-, *N*-, but also *C*-glycosides and sugar esters, can be formed [4,9]. Enzyme-based glycosylation generally exhibits high stereo- and regioselectivity with both promiscuous UGTs found glycosylating numerous acceptors (generalists) and selective specialists converting few substrates. In general, UGTs show a higher selectivity towards the donor substrate while being more flexible with respect to the acceptor. Although there are several families of glycosyltransferases, the GT1 family in the CAZy classification (www.cazy.org), which includes the UGTs, is of particular interest, because many of its members can glycosylate industrially relevant substances [1,10,11]. UGTs contain a conserved 44-amino-acid-long motif called the PSPG (plant secondary product glycosylation) box, carry a catalytically active His in the N terminus, are inverting Leloir-type glycosyltransferases, and adopt the GT-B fold [11]. Notably, UGTs are involved in the biosynthesis of a number of plant metabolites, such as flavonoids, alkaloids, terpenoids, and polyphenols. Thus, UGTs promote plant growth and development by modifying, detoxifying, transporting, and storing secondary metabolites and volatiles, and by protecting against biotic and abiotic stresses [12,13].

Since the advent of genome sequencing, and due to advances in sequencing techniques, the number of putative UGTs has increased exponentially, but only a few of them have been studied in depth. The characterization of UGTs could reveal a wealth of enzymes that could be harnessed for industrial purposes. Recently, interest in UGTs has focused mainly on their applications as catalysts in the biotechnological production of physiologically active metabolites such as steviosides, cardiotonic steroids, and *C*-glycosides [14–16]. Important criteria in these studies were substrate tolerance, regioselectivity, and enzyme reaction mechanisms. In similar studies, we discovered the promiscuous UGT72AY1 from *Nicotiana benthamiana*, which is thought to be involved in lignin biosynthesis, as shown in its homologs from *Arabidopsis thaliana* [17,18]. The glycosylation of monolignols roughly followed a Michaelis–Menten (MM)-like scheme, whereas strong substrate inhibition (SI) was observed when hydroxycoumarins such as scopoletin were used as acceptors (inhibitory constant K_i < 20 µM) [19]. A detailed biochemical analysis of this unusual enzyme by hydrogen/deuterium exchange mass spectrometry (HDX-MS) and mutational analyses revealed the strong UDP-glucose glucohydrolase activity of the enzyme in the absence of an acceptor substrate, which could be attenuated by apocarotenoid effector molecules [20]. However, the same effectors increased the enzymatic activity of the protein toward hydroxycoumarins by increasing the inhibitory constant K_I and thus reducing SI. A similar effect was observed in an F87I and a chimeric mutant (chimera A; Figure S1) containing a sequence segment of a homologous enzyme that showed only weak SI. Based on HDX-MS analyses and in silico modelling that identify amino acids interacting with the substrate, mutants N27D, R91F, R91M, and R91A were generated. The amino acids, N27 and R91, are thought to play a role in SI [19].

Since 20% of enzymes are inhibited by their own substrates, SI appears to have important biological functions and probably represents a biologically relevant regulatory mechanism [21]. However, since the molecular causes of this inhibition have been poorly investigated, we have performed further studies to find additional substrates of NbUGT72AY1 that show SI, and to clarify whether vanillin, which is structurally related to scopoletin, behaves similarly to hydroxycoumarin when glucosylated by this enzyme. The aim of the study was to investigate the effects of coumarins and various mutants on the enzymatic activity of UGT72AY1 toward vanillin and sinapaldehyde and to analyze how apocarotenoid and fatty acid effector molecules alter the kinetics of the reaction. The results show that NbUGT72AY1 is a flexible protein whose catalytic properties are modified by substrate and effector molecules, allowing it to sense the environment and adapt the enzyme activity accordingly.

2. Results

2.1. The Substituents of the Phenolic Substrates Dictate the Enzyme Kinetics of NbUGT72AY1

In a previous study, we showed that the promiscuous NbUGT72AY1 can glucosylate phenolics as well as short-chain alcohol, terpenoids, and apocarotenoids [17]. Since NbUGT72AY1 was strongly substrate-inhibited by hydroxycoumarins, vanillin, and carvacrol, but only weakly by monolignols [19], we performed further biochemical studies to reveal structure–function relationships. Enzyme activity studies carried out by UDP-Glo™ glucosyltransferase assay, with naturally occurring substrates structurally related to scopoletin and vanillin, showed a wide range of different activity profiles, ranging from Michaelis–Menten (MM) kinetics (hydroquinones) over weakly inhibited (sinapaldehyde, guaiacol, and o-cresol) to strongly substrate-inhibited profiles (scopoletin, eugenol, and vanillin) (Figure 1a). The equation that best explained the data for all substrates combines the two-binding site kinetic model for sequential ordered binding [22] and the Hill equation [23]. The equation contains two Hill coefficients, n and x, where x accounts for the possibility that substrate binding can also be cooperative in the inhibitory mode (Figure 1b). The equation becomes an MM equation when n and x are equal to 1, V_i is equal to 0, and K_i approaches infinity, as in the case of hydroquinone (Figure 1c). Complete substrate inhibition can be recognized by the fact that V_i is equal to 0 (sinapaldehyde, guaiacol, and o-cresol). All ortho-substituted phenols show SI kinetics, while para-substituted hydroquinone is not inhibited by the substrate. Kaempferol is a special case, as two products, kaempferol-3-*O*-glucoside and -7-*O*-glucoside, were formed (Figure S2). The replacement of the methoxy group in guaiacol with a methyl group, as in o-cresol, led to a strong decrease in V_{max} (393 ± 69 vs. 72 ± 4 nmol/min/mg) and K_m (179 ± 44 vs. 17 ± 9 µM), while other parameters were not significantly different, highlighting the importance of an ortho-substituent for substrate binding in the catalytic site. A third substituent appears to enhance the SI, as shown by the comparisons of guaiacol with ethylguaiacol/eugenol/vanillin. A fourth substituent, as in scopoletin and sinapaldehyde, leads to different outcomes.

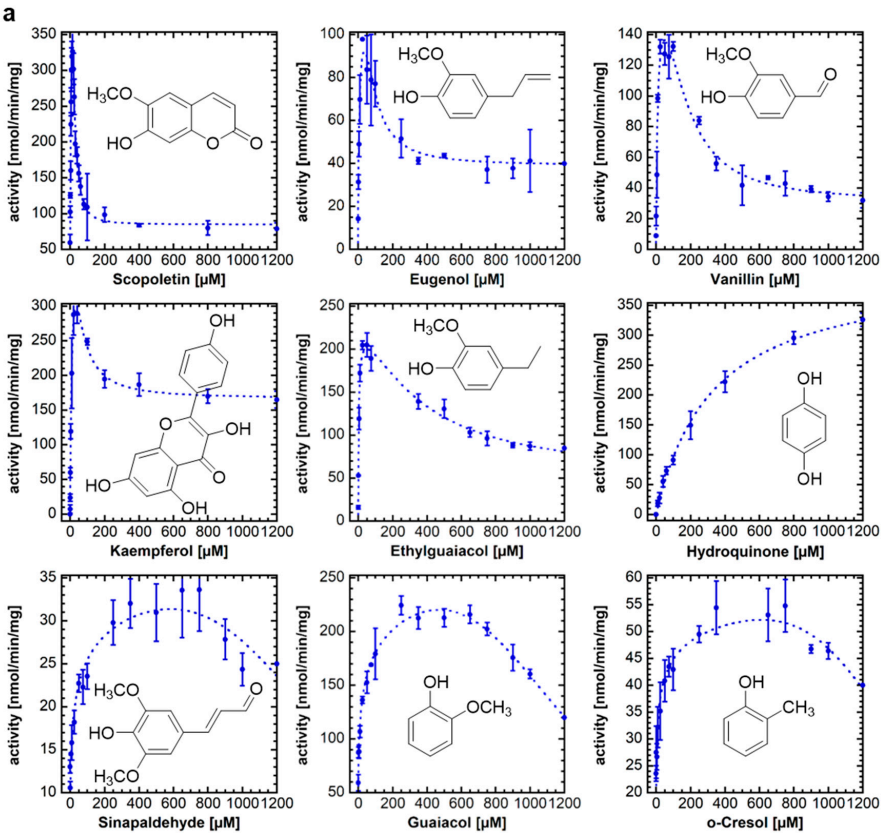


Figure 1. *Cont.*

b

$$v = \frac{V_{max} * \frac{[S]^n}{K_m^n} + V_i * \frac{[S]^n * [S]^x}{K_m^n * K_i^x}}{1 + \frac{[S]^n}{K_m^n} + \frac{[S]^n * [S]^x}{K_m^n * K_i^x}}$$

c

	Scopoletin	Eugenol	Vanillin	Kaempferol	Ethylguaiacol	Hydroquinone	Sinapaldehyde	Guaiacol	o-Cresol
Vmax	944 ± 62	171 ± 14	218 ± 14	515 ± 51	218 ± 8	414 ± 12	46 ± 2	393 ± 69	72 ± 4
Vi	86 ± 12	39 ± 4	31 ± 5	167 ± 5	72 ± 9	0	0	0	0
Km	16 ± 1	17 ± 2	9 ± 2	15 ± 3	3 ± 1	331 ± 25	56 ± 17	179 ± 44	17 ± 9
Ki	14 ± 1	62 ± 17	117 ± 24	46 ± 9	345 ± 58	∞	1359 ± 133	1022 ± 69	1368 ± 165
n	1 ± 0.1	0.6 ± 0.1	1 ± 0.1	1 ± 0.2	1 ± 0.1	1	0.4 ± 0.1	0.3 ± 0.1	0.3 ± 0.1
x	2 ± 0.2	1.7 ± 0.5	2 ± 0.2	1.5 ± 0.5	1.5 ± 0.5	1	4 ± 1.5	3.5 ± 1.0	4.5 ± 2
R	0.98	0.94	0.99	0.99	0.98	0.99	0.86	0.94	0.56

Figure 1. Enzymatic activity of NbUGT72AY1 toward various phenolic substrates. NbUGT72AY1 was incubated with increasing concentrations of substrates and the product UDP was determined by UDP-Glo^{TM} assay. (**a**) Plots of acceptor substrate concentration versus reaction rate. (**b**) Equation used for fitting the data (partial uncompetitive inhibition model with Hill coefficients) [23]. (**c**) Kinetic data obtained by fitting the data to the equation shown in (**b**).

2.2. Coumarins Have No Effect on UDP-Glucose Glucosyltransferase Activity and Substrate Inhibition of NbUGT72AY1

One possible hypothesis for the strong SI by scopoletin is based on a second (allosteric) substrate binding site on the NbUGT72AY1 protein [19]. To uncouple the catalytic and inhibitory modes of substrate binding, we used vanillin as the substrate and coumarins as possible inhibitors. However, enzyme activity assays performed by UDP-Glo^{TM} in the absence and presence of coumarin, 6-methoxycoumarin, and 7-methoxycoumarin showed that the kinetic parameters of UGT catalysis were not affected by the coumarins (Figure 2). Using a one-tailed *t*-test ($p < 0.01$), the V_{max}, V_i, K_m, and K_i values of the samples with coumarins were not significantly different from those of the samples without coumarins. Thus, we concluded that either there is no second binding site on NbUGT72AY1, or the coumarins are unable to bind to the allosteric site. Since the number and position of substituents on the phenol ring are essential for binding to the allosteric site of the enzyme, as shown in the previous section for hydroquinone, it is likely that coumarins, unlike hydroxycoumarins, do not interact with NbUGT72AY1.

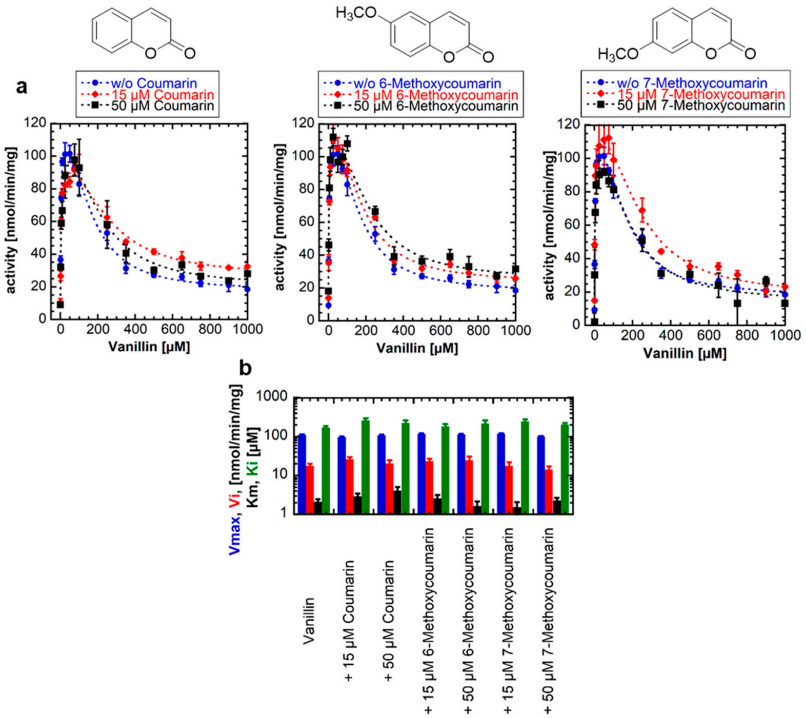


Figure 2. Enzyme activity of NbUGT72AY1 toward vanillin in the presence of different coumarins. (**a**) NbUGT72AY1 was used to glucosylate vanillin in the presence of different concentrations

of coumarin, 6-methoxycoumarin, and 7-methoxycoumarin. Enzyme activity was determined by UDP-Glo™ Glycosyltransferase assay and fitted to the equation shown in Figure 1. (**b**) Kinetic parameters of NbUGT72AY1 using vanillin as acceptor substrate in the presence of coumarins. Please note that the y-axis is displayed logarithmically to display the entire range of values. The colors of the bars are explained on the y-axis. Experimental values were fitted to the equation shown in Figure 1 (n = 1; x = 2). Parameters are not significantly different ($p < 0.01$).

2.3. Mutant NbUGT72AY1 Proteins Show Different Enzyme Kinetics with Vanillin but Similar Enzyme Reaction Curves with Sinapaldehyde

In a previous study, an F87I and a chimera A mutant of NbUGT72AY1 showed reduced SI with scopoletin compared with the wild-type (WT) enzyme [19]. In the F87I mutant, an essential amino acid in the active site was exchanged, whereas the chimera A mutant contained a sequence part of StUGT72AY2 from *Solanum tuberosum*, which is a homolog of NbUGT72AY1 but has only a weak SI. Therefore, in a new experiment, we investigated whether the mutants behave similarly when the natural substrate analog vanillin is used (Figure 3a).

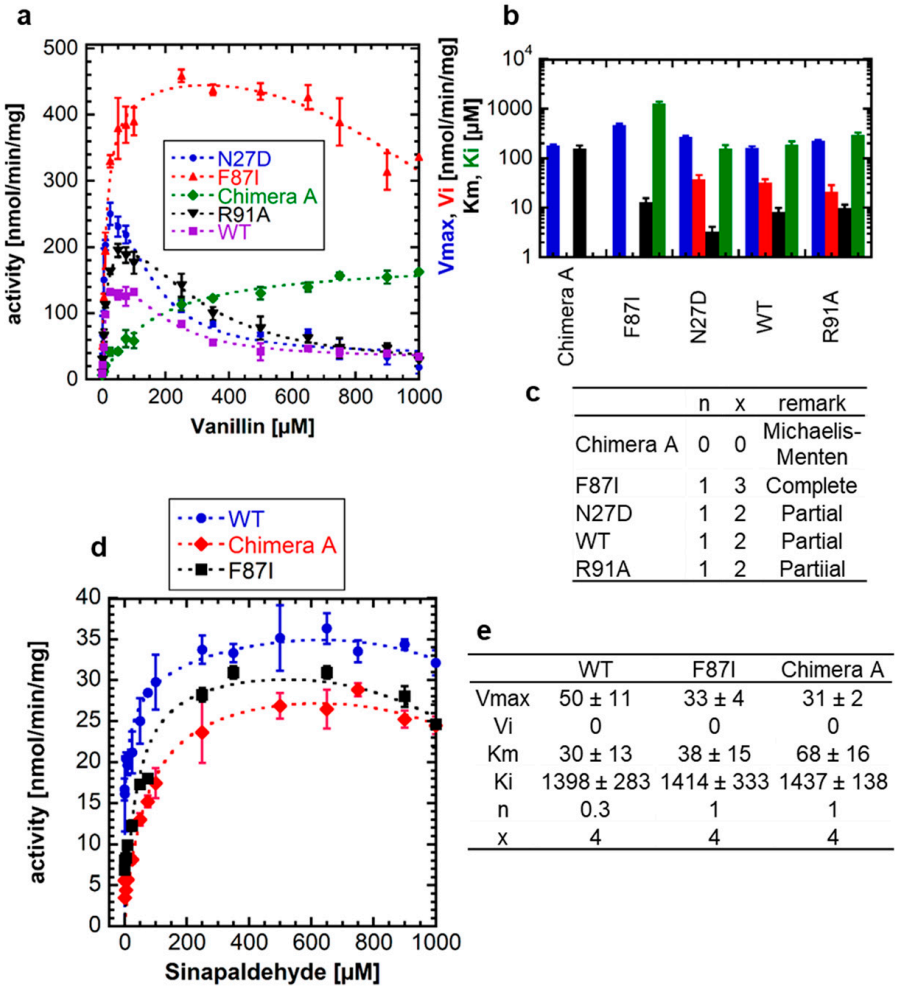
While mutants N27D and R91A showed a strong SI similar to the WT with comparable kinetic data, except for the V_{max} of N27D, which was significantly different from the value of the WT, F87I and chimera A showed a weak SI (K_i = 1261 ± 128 µM) and MM kinetics, respectively (Figure 3b,c). Thus, the catalytic activity of NbUGT72AY1 and its mutants toward vanillin resembles that of the enzymes toward scopoletin, implying that the phenolic aldehyde should interact with the same amino acids of the proteins as the phenolic lactone. In contrast, WT NbUGT72AY1, as well as the F87I and chimera A mutants showed mild SI with sinapaldehyde (Figure 3d).

Although structurally related, monolignol is thus likely to bind more weakly to the putative allosteric site than scopoletin, since the K_i values are significantly higher (1398 ± 283 µM, 1414 ± 333 µM, and 1437 ± 138 µM for WT, F87I, and chimera A, respectively) (Figure 3e) than those for scopoletin to the WT variant (16 ± 1 µM) (Figure 1a,c).

2.4. Apocarotenoids and Fatty Acids Enhance the Glucosyltransferase Activity and Reduce Substrate Inhibition of NbUGT72AY1 with Vanillin

Apocarotenoids, including α- and β-ionol, have been shown to increase the glucosylation activity of NbUGT72AY1 toward scopoletin by decreasing SI due to increasing K_i levels [20] and UDP-glucuronosyltransferase 1A1 activity is inhibited by fatty acids [24]. Therefore, we analyzed the effect of the naturally occurring effectors α- and β-ionol (100 µM and 200 µM each) and fatty acids such as stearic acid, oleic acid, linoleic acid, and linolenic acid (20 µM and 100 µM each) on the enzymatic activity of NbUGT72AY1 toward vanillin (Figure 4a). The addition of α- and β-ionol to the reaction resulted in a concentration-dependent increase in K_i and reduced V_i to zero (complete uncompetitive SI) (Figure 4b). Similarly, the addition of C18 fatty acids altered the course of the enzymatic reaction curve, but in different ways.

Saturated stearic acid promoted the glycosylation activity of NbUGT72AY1 with vanillin as V_{max} increased in a concentration-dependent manner (Figure 4b). The unsaturated fatty acids also resulted in a statistically significant increase in V_{max} upon addition of 20 µM, but also in a concentration-dependent increase in K_i that was statistically significant at 100 µM. Moreover, the reaction curve changed to a complete uncompetitive SI upon addition of 100 µM of the unsaturated fatty acids. Thus, it appears that, in the case of oleic, linoleic, and linolenic acid, more than one cause contributes to the altered response curve. The increased catalytic activity of NbUGT72AY1, noted after addition of stearic and oleic acid, was independently confirmed by LC-MS analysis. The reduced product formation after excessive addition of oleic acid (100 µM) was also corroborated by LC-MS analysis (Figure S3).



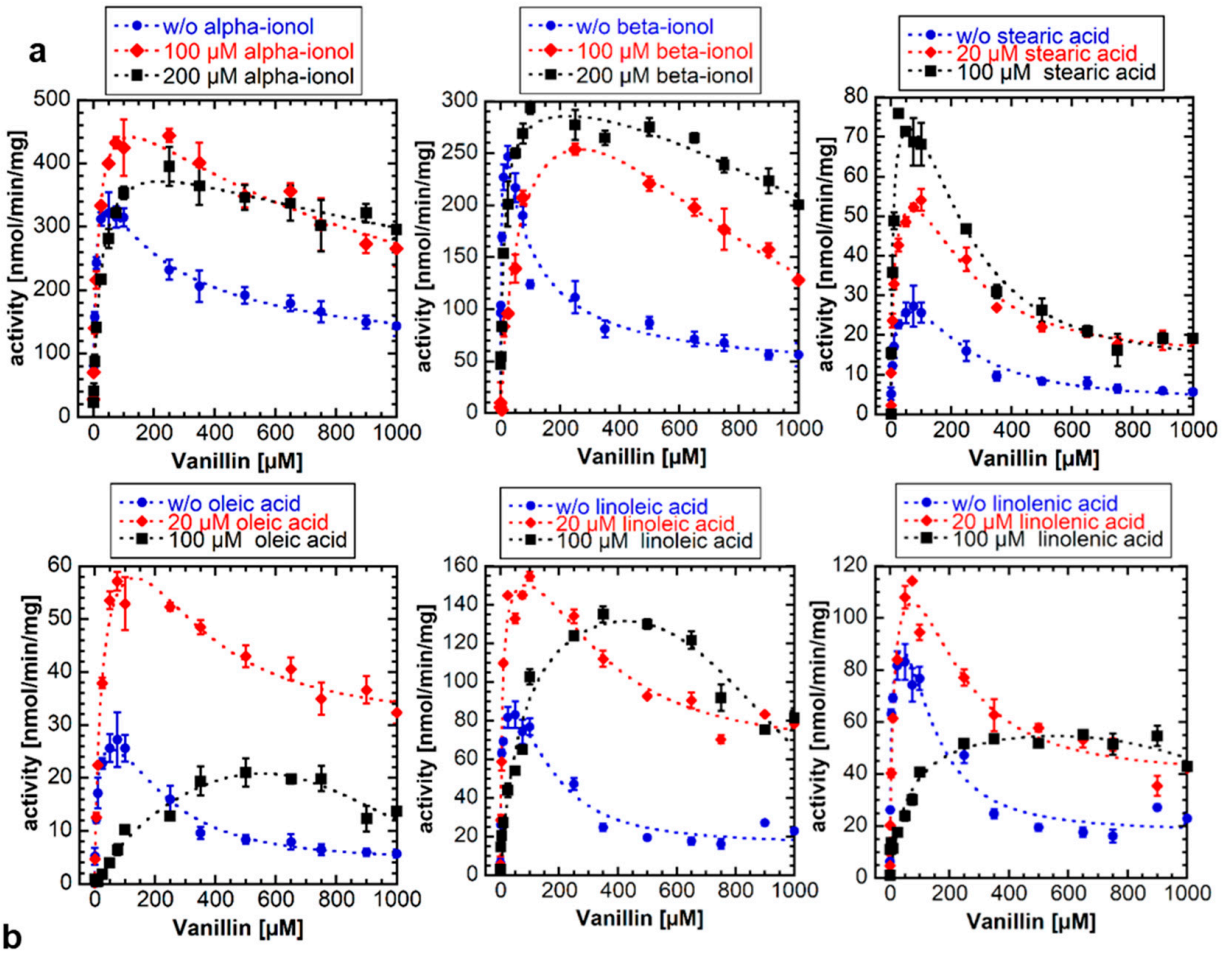
	n	x	remark
Chimera A	0	0	Michaelis-Menten
F87I	1	3	Complete
N27D	1	2	Partial
WT	1	2	Partial
R91A	1	2	Partiial

	WT	F87I	Chimera A
Vmax	50 ± 11	33 ± 4	31 ± 2
Vi	0	0	0
Km	30 ± 13	38 ± 15	68 ± 16
Ki	1398 ± 283	1414 ± 333	1437 ± 138
n	0.3	1	1
x	4	4	4

Figure 3. Enzyme activity of wild-type NbUGT72AY1 and selected mutants toward vanillin and sinapaldehyde. (**a**) Enzyme activity versus substrate (vanillin) concentration plot for wild-type (WT) NbUGT72AY1 and selected mutants. The N27D and R91A mutant show a similar substrate-inhibited enzyme kinetics as the WT protein. (**b**) Kinetic parameters of NbUGT72A1 and mutant enzymes using vanillin as acceptor substrate. Please note that the y-axis is displayed logarithmically to show the entire range of values. The colors of the bars are explained on the y-axis. (**c**) Values were fitted to the equation shown in Figure 1. Mutants show Michaelis–Menten, complete ($V_i = 0$) and partial uncompetitive substrate ($V_i > 0$) inhibition kinetics. (**d**) Enzyme activity versus substrate (sinapaldehyde) concentration plot for WT NbUGT72AY1 and selected mutants. Experimental data were determined by UDP-Glo^{TM} Glycosyltransferase assay. (**e**) Kinetic parameters of NbUGT72A1 and mutant enzymes using sinapaldehyde as acceptor substrate. Values were fitted to the equation shown in Figure 1. Parameters are not significantly different ($p < 0.01$).

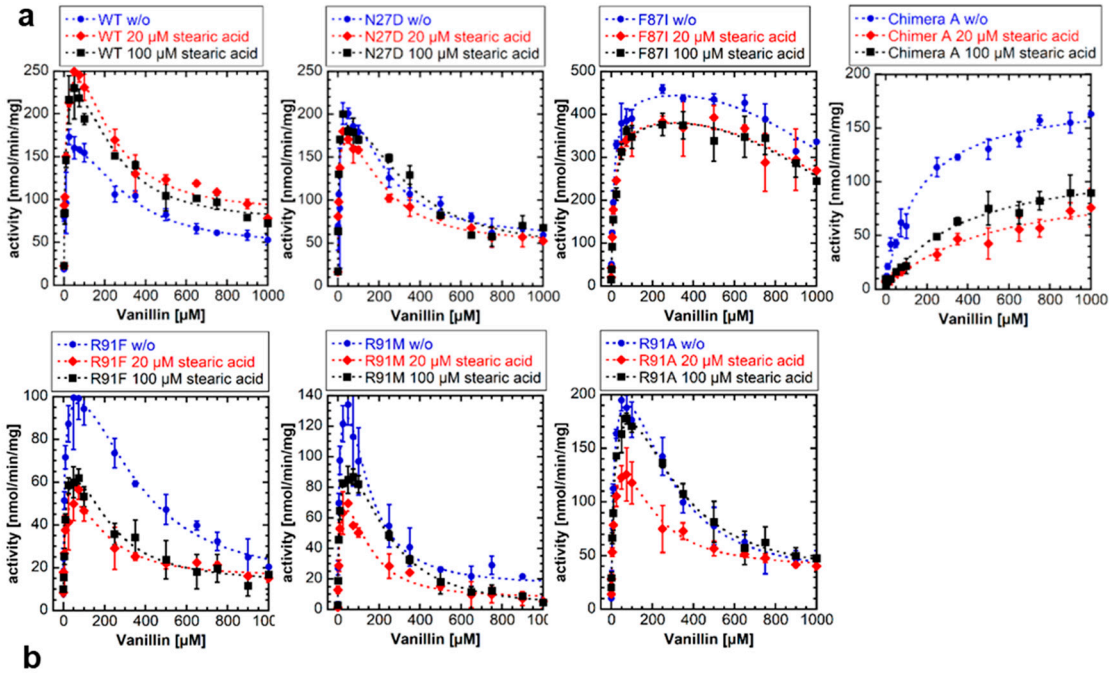
2.5. The Enzymatic Activity of NbUGT72AY1 Mutants Is Either Unaffected or Reduced by Fatty Acids

Since fatty acids such as stearic acid altered the enzymatic activity of WT NbUGT72AY1 by decreasing SI in the previous experiment, we next tested the effects on NbUGT72AY1 mutants (F87I and chimera A) that exhibit reduced SI (Figure 5a). The results of the enzyme activity assays confirmed the promoting effect of stearic acid for the NbUGT72AY1 WT, as explained by the significantly increased V_{max} values, and showed that the saturated acid did not affect the catalytic activity of the N27D mutant, but did alter the activities of the F87I and chimera A mutants (Figure 5b).



Enzyme	variable concentration	constant concentration	Km [µM]	Vmax [nmol/min/mg]	Ki [µM]	Vi [nmol/min/mg]	n	x	R
		release of UDP determined by UDP-Glo assay							
NbUGT 72AY1	vanillin	no	7 ± 1	29 ± 1	239 ± 29	4 ± 1	1	2	0.99
		20 µM stearic acid	8 ± 1	57 ± 2*	293 ± 46	13 ± 2	1	2	0.98
		100 µM stearic acid	6 ± 1	82 ± 2*	228 ± 27	13 ± 2	1	2	0.99
		no	7 ± 1	29 ± 1	239 ± 29	4 ± 1	1	2	0.99
		20 µM oleic acid	21 ± 3	71 ± 3*	328 ± 56	20 ± 2	1	2	0.99
		100 µM oleic acid	383 ± 126*	39 ± 7	858 ± 72*	0	1	4	0.99
		no	3 ± 1	91 ± 5	181 ± 31	15 ± 3	1	2	0.97
		20 µM linoleic acid	6 ± 1	167 ± 9*	339 ± 77	64 ± 9	1	2	0.98
		100 µM linoleic acid	76 ± 14*	161 ± 12	929 ± 59*	0	1	4	0.98
		no	3 ± 1	91 ± 5	181 ± 31	15 ± 3	1	2	0.97
		20 µM linolenic acid	11 ± 2	128 ± 9	222 ± 45	39 ± 4	1	2	0.98
		100 µM linolenic acid	62 ± 9*	62 ± 4	1367 ± 189*	0	1	4	0.97
		no	7 ± 2	428 ± 34	209 ± 87	89 ± 27	1	1	0.99
		100 µM α-ionol	15 ± 2	550 ± 24	1006 ± 135*	0	1	1	0.99
		200 µM α-ionol	25 ± 4	455 ± 23	2004 ± 422*	0	1	1	0.99
		no	3 ± 1	325 ± 87	75 ± 26	38 ± 20	1	1	0.89
		100 µM β-ionol	55 ± 16	330 ± 50	840 ± 146*	0	1	1	0.98
		200 µM β-ionol	11 ± 2	306 ± 12	1469 ± 188*	0	1	2	0.98

Figure 4. Enzyme activity of NbUGT72AY1 toward vanillin in the presence of apocarotenoids and fatty acids. (**a**) NbUGT72AY1 was used to glucosylate vanillin in the presence of α- and β-ionol, stearic acid, oleic acid, linoleic acid, and linolenic acid. Experimental data were determined by UDP-Glo^{TM} Glycosyltransferase assay and fitted to the partial uncompetitive inhibition model shown in Figure 1. (**b**) Kinetic parameters of NbUGT72AY1 using vanillin as substrate with various effectors. Asterisks (*) indicate that the values are statistically significantly different according to a *t*-test ($p < 0.01$) from the values for the samples without addition of the effectors.



b

Enzyme	variable concentration	constant concentration	Km [µM]	Vmax [nmol/min/mg]	Ki [µM]	Vi [nmol/min/mg]	n	x	R
release of UDP determined by UDP-Glo assay									
WT	vanillin	no	3 ± 1	167 ± 14	307 ± 105	45 ± 17	1	2	0.93
		20 µM stearic acid	7 ± 2	279 ± 23*	225 ± 64	85 ± 16	1	2	0.96
		100 µM stearic acid	6 ± 2	259 ± 10*	237 ± 70	73 ± 16	1	2	0.95
N27D		no	6 ± 2	221 ± 23	228 ± 75	57 ± 16	1	2	0.94
		20 µM stearic acid	2 ± 1	179 ± 11	235 ± 57	49 ± 10	1	2	0.97
		100 µM stearic acid	2 ± 1	198 ± 8	339 ± 62	41 ± 13	1	2	0.98
F87I		no	13 ± 2	458 ± 5	1296 ± 71	0	1	3	0.99
		20 µM stearic acid	14 ± 2	402 ± 4*	1249 ± 72	0	1	3	0.99
		100 µM stearic acid	17 ± 2	406 ± 3*	1207 ± 67	0	1	3	0.99
Chimera A		no	157 ± 23	181 ± 6	∞	0	1	1	0.99
		20 µM stearic acid	505 ± 172*	105 ± 10*	∞	0	1	1	0.97
		100 µM stearic acid	365 ± 65*	122 ± 4*	∞	0	1	1	0.99
R91F		no	5 ± 1	107 ± 3	364 ± 38	13 ± 4	1	2	0.99
		20 µM stearic acid	5 ± 2	56 ± 4*	207 ± 63	16 ± 3	1	2	0.95
		100 µM stearic acid	7 ± 2	65 ± 3*	227 ± 48	13 ± 3	1	2	0.97
R91M		no	5 ± 1	152 ± 10	147 ± 21	16 ± 4	1	2	0.99
		20 µM stearic acid	6 ± 2	79 ± 5*	150 ± 27	7 ± 2	1	2	0.98
		100 µM stearic acid	5 ± 1	100 ± 2	242 ± 12	1 ± 1	1	2	0.98
R91A		no	10 ± 2	228 ± 9	275 ± 27	30 ± 6	1	2	0.99
		20 µM stearic acid	8 ± 2	145 ± 8*	212 ± 36	39 ± 5	1	2	0.99
		100 µM stearic acid	11 ± 2	206 ± 10	315 ± 39	31 ± 8	1	2	0.99

Figure 5. Enzyme activity of NbUGT72AY1 mutants toward vanillin in the presence of putative effectors. (**a**) Wild-type (WT) and mutant proteins were used to glucosylate vanillin in the presence of stearic acid. Experimental data were determined by UDP-Glo™ Glycosyltransferase assay and fitted to the equation shown in Figure 1. (**b**) Kinetic parameters of WT and mutant proteins using vanillin as substrate with stearic acid. Asterisks (*) indicate that the values are statistically significantly different according to a *t*-test ($p < 0.01$) from the values for the samples without addition of the effectors.

The V_{max} values were significantly lowered after the addition of stearic acid in the F87I mutant, which has only a weak SI, whereas the V_{max} and K_m values were considerably reduced and increased, respectively, in the chimera A mutant. In the R91 mutants (R91F, R91M, and R91A), in which an amino acid presumably important for SI had been replaced, the addition of stearic acid had no effect or decreased activity, with the V_{max} values significantly reduced at 20 µM addition. Thus, it can be concluded that stearic acid has a positive, promoting effect on the catalytic activity of the WT NbUGT72AY1 enzyme, which exhibits strong SI but has no effect, or an inhibitory effect, on mutants in which amino acids putatively involved in SI have been mutated.

3. Discussion

Promiscuous enzymes are of particular interest for the biotechnological production of chemicals, agrochemicals, and pharmaceuticals, as they catalyze reactions in a stereoselective manner and can be used for the manufacture of a wide range of industrially relevant products due to their substrate tolerance. However, a significant number of biocatalysts are inhibited by their substrates at high concentrations, limiting their potential applications. Since the molecular mechanisms of SI are not fully understood, the targeted elimination of this limiting enzyme property by rational design is difficult. The best-known example of SI is the inhibition of phosphofructokinase by ATP, which leads to the suppression of glycolysis and, thus, to the cessation of ATP production [25]. The most commonly cited model of SI is the two binding site model with a catalytic and allosteric site [22]. The binding of the substrate to the allosteric binding site in the enzyme (E) or in the enzyme–substrate complex (ES) forms an inhibitory complex in which the catalyzed reaction is either very slow or completely suppressed (Figure 6). In alternative models, excess substrate molecules interact with enzyme forms other than the enzyme–substrate complex, such as the reaction intermediate (EI) or the enzyme–product complex (EP). All these models have in common a second substrate molecule that is bound to the enzyme. The inhibitory effects of substrates are attributed to the accumulation of a catalytically incompetent combination of enzyme, cofactor, and substrate. Such inappropriate termination complex formation has been reported for multi-substrate and multi-product enzymes with multiple binding sites [26]. However, a recent model based on conformational motions of proteins has shown that an allosteric site is not essential for SI. By using single-molecule FRET spectroscopy, it has been demonstrated that acceptor substrates can facilitate the domain closure of a kinase at lower concentrations of the donor substrate, which can affect the proper substrate-binding mechanics required for the reaction [27]. NbUGT72AY1 lends itself here as a model to better understand SI as it shows extreme SI (Figure 1).

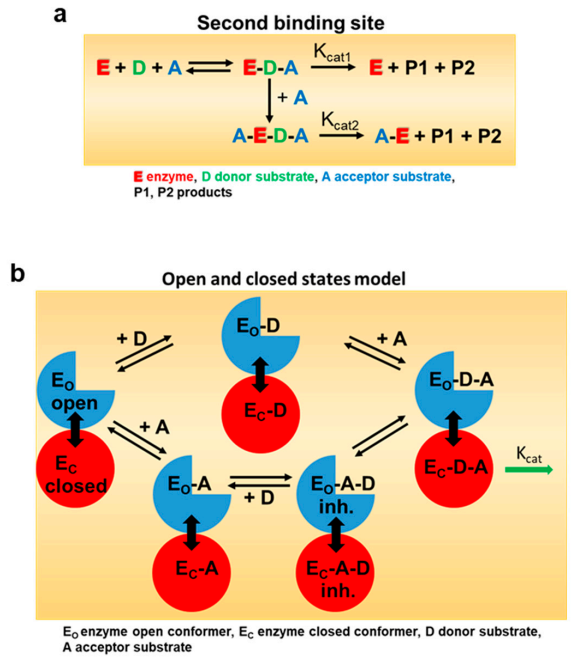


Figure 6. Model for enzymatic substrate inhibition. (**a**) In the classical model for substrate inhibition the enzyme has catalytic activity when one substrate is bound, but reduced ($K_{cat2} < K_{cat1}$) or even no activity $K_{cat2} = 0$) if two are bound (adapted from [22]). The color code is explained in the sub figure. (**b**) In the open and closed states model, inhibitory concentrations of acceptor substrate A lead to a faster and more cooperative domain closure by donor substrate D, leading, in turn, to an increased population of the closed inhibited state (E_C-A-D). Too rapid a premature closure of the domain could interfere with substrate-binding mechanisms (adapted from [27]).

3.1. NbUGT72AY1 Exhibits Michaelis–Menten and Substrate Inhibition Kinetics Depending on Substrate Structure

NbUGT72AY1 is a promiscuous enzyme that glucosylates phenols as well as short-chain alcohols and terpenes [17]. Studies on enzyme activity showed that MM and SI kinetics were obtained depending on the substrate structure (Figure 1). While only the para-substituted hydroquinone showed a hyperbolic curve, ortho- and tri-substituted phenols exhibited both weak and strong SI. Hydroquinone thus presumably binds exclusively in the catalytic center, or an excess of benzene-1,4-diol has no effect on the dynamics of domain closure. The measured K_m values for NbUGT72AY1 substrates are relatively low (<200 µM), with the exception of hydroquinone, suggesting that this enzyme has a high affinity for a number of ortho-substituted phenols. UGT72 enzymes are thought to be involved in the modification of flavonoids and the lignin formation by glucosylation of monolignols [28,29], as shown for homologous enzymes from *Arabidopsis thaliana* [30,31]. Similarly, NbUGT72AY1 glucosylates flavonoids and monolignols such as kaempferol and sinapaldehyde, albeit with different efficiencies and kinetics [19] (Figure 1). However, NbUGT72AY1 may also be implicated in the detoxification of the airborne phenols produced by forest fires, as guaiacol and ethylguaiacol are metabolized efficiently and the enzyme is constitutively expressed in the stem of the tobacco plant. The glucosylation of airborne volatiles after uptake by plants has been demonstrated in the leaves of grapevine, tomato and tea plants [32–34].

The strong SI of NbUGT72AY1 for scopoletin was related to its putative function in plant defense [13,20]. Thus, the neighboring cells of tissues damaged by an herbivore might protect themselves by glucosylation of the phytoalexin synthesized in response to the attack. However, if the damage is too severe, it might be more beneficial to build a physical barrier of dead cells, which could explain the significantly reduced activity at high scopoletin concentrations [35].

3.2. Blocking the Inhibitory Action of Substrates

Since both scopoletin and vanillin exhibit SI, we attempted to uncouple the inhibitory effect of vanillin from the catalytic activity by adding different coumarin derivatives to the reaction solution. However, the enzyme activity curves and kinetic parameters were not significantly different in the presence of the coumarins (Figure 2). The coumarins are either unable to replace vanillin at the allosteric site or the second binding site is not existent. In the case of the second hypothesis for SI, this would mean that coumarins have no influence on the dynamics of enzyme movement. However, mutants of NbUGT72AY1 generated to suppress SI in the WT enzyme toward scopoletin (F87I and chimera A) [19] also showed reduced SI or no inhibition at all with vanillin (Figure 3). However, the mutants did not exhibit altered kinetics of sinapaldehyde glucosylation compared with WT. In the N27D and R91A mutants, the amino acids putatively involved in SI had been exchanged [19], but only in N27D was V_{max} significantly increased compared with WT. Amino acid F87 is part of the active site and was identified by HDX-MS analysis [19], whereas chimera A contains a segment of a homologous enzyme that showed only weak SI [19]. Based on in silico analyses, it was hypothesized that NbUGT72AY1 has an allosteric site that shares F91 and amino acids of the sequence inserted into chimera A with the catalytic center [19]. The hypothesis is supported by the observation that the crystal structure of a human sulfotransferase (SULT1A1) contains two substrate molecules and the residue Phe-247 of SULT1A1, which interacts with both p-nitrophenol molecules, is important for substrate inhibition [26]. The results obtained for vanillin confirm the data determined for scopoletin and show that SI can be reduced or even abolished by the replacement of individual amino acids. This was also shown for tyrosine hydroxylase [36], betaine aldehyde dehydrogenase [37], salutaridine reductase [38], human sulfotransferase [26], D-3-phosphoglycerate dehydrogenase [39], lactate dehydrogenase [40], and haloalkane dehalogenase [41].

3.3. Effectors Increase the Enzymatic Activity of Substrate-Inhibited NbUGT72AY1 but Decrease the Catalytic Activity of Mutants That Exhibit Attenuated Substrate Inhibition

Recently, we showed that the SI of scopoletin in NbUGT72AY1 was decreased by apocarotenoids, which could be explained by an increase in the inhibitory constant K_i [20]. In terms of the two-substrate binding site model, this implies that binding to the allosteric site is restricted by apocarotenoids. In the case of the second model, this means that the effectors prevent the early closure of the catalytic center when the acceptor substrate is present in excess. In this study, we demonstrated that a decrease in SI and an increase in enzymatic activity is also possible through the addition of fatty acids (Figure 4). Thus, 20 µM of stearic, oleic, linoleic, and linolenic acids increased V_{max}. However, at 100 µM of unsaturated fatty acids, K_m and K_i increased significantly and V_i decreased to zero compared to the sample without effectors. An exception was stearic acid, where a concentration-dependent increase in enzyme activity was observed. A possible explanation for the different behavior is the micelle and vesicle formation of the long-chain fatty acids (Figure S4). Saturated and unsaturated fatty acids form self-assembling structures such as micelles, vesicles and oil droplets at different pH values of the medium above their critical micelle concentrations (CMC), and critical vesicle concentration (CVC), depending on the concentration of the acids, the temperature, and the ionic strength of the buffer [42,43] (Figure S4). Therefore, different CVCs and CMCs for fatty acids are found in the literature, but it seems that CMCs increase with the number of double bonds of the fatty acids and pH (Table S1). At concentrations below the CMC (20 µM), when the acids are dissolved as single molecules, they can readily interact with NbUGT72AY1 and promote the activity of the enzyme (Figure 4). At concentrations above the CMC (100 µM), when micelles and vesicles have formed, the diffusion from the micelles and vesicles, respectively, leads to an obvious increase in K_m and K_i values in the case of the unsaturated acids (Figures 4 and S4). Stearic acid enhances glucosylation activity even at a concentration of 100 µM. However, stearic acid does not promote activity in NbUGT72AY1 mutants in which the amino acids thought to play a role in SI have been mutated, but actually decreases V_{max} in the case of the F87I and chimera A mutant and increases K_m for chimera A (Figure 5). Thus, this is an un-competitive inhibition of the chimera A mutant (binding to the enzyme–substrate complex only), which exhibits MM kinetics and a non-competitive inhibition of F87I (binding to the enzyme and enzyme–substrate complex). The inhibition of enzyme activities by free fatty acids has long been known [44] and, recently, lipoxygenases were shown to be regulated by fatty acids through interaction with an N-terminal binding domain [45]. Furthermore, stearic acid suppressed the enzymatic activity of a thioesterase when a C-terminal lipid binding domain was absent, suggesting that this hydrophobic domain abolished the inhibitory effect of stearic acid [46]. In the future, HDX-MS will clarify whether the segment replaced in chimera A is a possible binding site for fatty acids.

4. Materials and Methods

4.1. Cloning and Protein Expression of UGT72AY1

Cloning of NbUGT72AY1 from *Nicotiana benthamina* (accession MT945401) with vector pGEX-4T-1, and the protein expression in *Escherichia coli BL21(DE3) pLysS*, were performed according to [17]. Mutants (N27D, F87I, chimera A, R91A, R91F, and R91M) were generated according to [17,19] (Figure S5).

4.2. Enzyme assays by UDP Glo™ Glycosyltransferase Assay

The kinetics of UGT72AY1 with vanillin and effectors were measured using the UDP-Glo™ Glycosyltransferase Assay (Promega, Mannheim, Germany). The optimal enzymatic reaction conditions were determined as described [17]. The 100 µL assay contained 50 mM Tris pH 7.5, 0.5 µg protein, a defined concentration of effector, and different concentrations of vanillin and 100 mM UDP-glucose, which was added to start the incubation at 40 °C and 500 rpm for 10 min. Finally, the reaction was stopped by 12.5 µL 0.6 M HCl and an addition of 12.5 µL 1 M Trizma base pH 10.7 was used to adjust the pH. UDP detection reagent (UDR)

was used to quantify the released UDP during the catalysis in a 384-well plate (384-Well Plates, Corning 4513, Sigma–Aldrich, Taufkirchen, Germany) and incubated for 30 min in the dark before the luminescence signal was measured by the CLARIOstar plate reader (BMG Labtech, Ortenberg, Germany). Kinetic data were calculated with KaleidaGraph (https://www.synergy.com/; accessed 22 May 2023; v4.5.4).

4.3. Liquid Chromatography-Mass Spectrometry Analysis

After centrifugation (20 min at 5000× *g*), the samples prepared for the UDP-Glo assay were used for LC-MS analysis according to [17]. Vanillin, sinapaldehyde, and kaempferol glucosides were identified according to [19] (Figure S2).

4.4. Enzyme Kinetics Analysis

A two-site model (Figure 1) was used to explain the substrate inhibition phenomenon of NbUGT72AY1 [19]. Here, [S] is the concentration of the varied substrate, V_{max} is the maximal reaction rate, and K_m represents the substrate concentration at which the reaction rate is $\frac{1}{2}$ V_{max}. The parameter V_i is the reaction velocity in the presence of inhibition, K_i is the inhibition constant which is the inhibitor concentration required to decrease the maximal rate of the reaction to $\frac{1}{2}$ of the uninhibited value. The equation presumes the sequential binding of substrate molecules, i.e., the inhibitory site cannot be occupied until the reaction site is filled. By adding cooperativity-describing Hill coefficients, an equation was obtained that best described the measured data. The superscript n is a Hill coefficient, and x is another Hill coefficient that allows for the possibility that the binding of substrate in the inhibitory mode may also be cooperative [47]. To obtain convergence for the equation in Figure 1, the value of x was fixed, which was determined empirically to give a best fit (lowest variance). The kinetic parameters were determined under optimum conditions and were calculated with KaleidoGraph version 4.5.4 from Synergy Software (Eden Prairie, MN, USA). The data were derived from at least three repeats. Statistical analysis was performed using Student's *t*-test ($p < 0.01$).

$$v = \frac{V_{max} * \frac{[S]^n}{K_m^n} + V_i * \frac{[S]^n * [S]^x}{K_m^n * K_i^x}}{1 + \frac{[S]^n}{K_m^n} + \frac{[S]^n * [S]^x}{K_m^n * K_i^x}}$$

5. Conclusions

Overall, the results show that vanillin behaves similarly to scopoletin as a substrate for NbUGT72AY1. These substrates show SI with the WT, which is reduced by apocarotenoids, is only weakly observed in the F87I mutant, and is not seen at all in the chimera A mutant. Coumarins do not affect enzymatic activity, but this does not completely rule out a second binding site. Fatty acids, on the other hand, promote catalytic activity depending on free fatty acid molecules, which is why aggregations of fatty acids above their CMC lead to altered reaction rates. Since the enzyme exhibits SI with a range of substrates, this opens up unique opportunities for the regulation of the enzyme by effectors that can attenuate inhibition. NbUGT72AY1 is a multi-substrate enzyme whose enzymatic activity can be fine-tuned by external, naturally occurring metabolites such as apocarotenoids and fatty acids that affect SI. These signals are generated upon plant cell destruction, which is why NbUGT72AY1 likely plays an important role in plant defense as it may be involved in the production of lignin in the cell wall and may provide direct protection through the formation of toxic phytoalexins.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms24119542/s1. References [48–52] are cited in the supplementary materials.

Author Contributions: Conceptualization, J.L., V.L., T.D.H., G.S., T.H. and W.S.; methodology, T.H.; software, T.H. and W.S.; validation, T.D.H.; formal analysis, J.L., V.L., A.B., Z.Z., T.D.H. and G.S.; investigation, J.L., V.L., A.B., Z.Z. and G.S.; resources, W.S.; data curation, J.L., V.L., T.H. and W.S.; writing—original draft preparation, J.L. and T.D.H.; writing—review and editing, T.D.H., T.H., C.S. and W.S.; visualization, W.S.; supervision, J.L., T.D.H. and T.H.; project administration, W.S.; funding acquisition, W.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by DEUTSCHE FORSCHUNGSGEMEINSCHAFT SCHW634/34-1.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article or Supplementary Material.

Acknowledgments: We thank Mechthild Mayershofer, Anja Forstner, and Ruth Habegger for their technical support.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. He, B.; Bai, X.; Tan, Y.; Xie, W.; Feng, Y.; Yang, G.-Y. Glycosyltransferases: Mining, engineering and applications in biosynthesis of glycosylated plant natural products. *Synth. Syst. Biotechnol.* **2022**, *7*, 602–620. [CrossRef] [PubMed]
2. Kurze, E.; Wüst, M.; Liao, J.; McGraphery, K.; Hoffmann, T.; Song, C.; Schwab, W. Structure-function relationship of terpenoid glycosyltransferases from plants. *Nat. Prod. Rep.* **2021**, *39*, 389–409. [CrossRef] [PubMed]
3. Danby, P.M.; Withers, S.G. Advances in Enzymatic Glycoside Synthesis. *ACS Chem. Biol.* **2016**, *11*, 1784–1794. [CrossRef] [PubMed]
4. Ati, J.; Lafite, P.; Daniellou, R. Enzymatic synthesis of glycosides: From natural O- and N-glycosides to rare C- and S-glycosides. *Beilstein J. Org. Chem.* **2017**, *13*, 1857–1865. [CrossRef] [PubMed]
5. Coines, J.; Cuxart, I.; Teze, D.; Rovira, C. Computer Simulation to Rationalize "Rational" Engineering of Glycoside Hydrolases and Glycosyltransferases. *J. Phys. Chem. B* **2022**, *126*, 802–812. [CrossRef]
6. McGraphery, K.; Schwab, W. Comparative analysis of high-throughput assays of family-1 plant glycosyltransferases. *Int. J. Mol. Sci.* **2020**, *21*, 2208. [CrossRef]
7. Lairson, L.L.; Henrissat, B.; Davies, G.J.; Withers, S.G. Glycosyltransferases: Structures, functions, and mechanisms. *Annu. Rev. Biochem.* **2008**, *77*, 521–555. [CrossRef]
8. Meech, R.; Hu, D.G.; McKinnon, R.A.; Mubarokah, S.N.; Haines, A.Z.; Nair, P.C.; Rowland, A.; Mackenzie, P.I. The UDP-Glycosyltransferase (UGT) superfamily: New members, new functions, and novel paradigms. *Physiol. Rev.* **2019**, *99*, 1153–1222. [CrossRef]
9. Teze, D.; Coines, J.; Fredslund, F.; Dubey, K.D.; Bidart, G.N.; Adams, P.D.; Dueber, J.E.; Svensson, B.; Rovira, C.; Welner, D.H. O -/ N -/ S -specificity in glycosyltransferase catalysis: From mechanistic understanding to engineering. *ACS Catal.* **2021**, *11*, 1810–1815. [CrossRef]
10. Liang, D.-M.; Liu, J.-H.; Wu, H.; Wang, B.-B.; Zhu, H.-J.; Qiao, J.-J. Glycosyltransferases: Mechanisms and applications in natural product development. *Chem. Soc. Rev.* **2015**, *44*, 8350–8374. [CrossRef]
11. Mestrom, L.; Przypis, M.; Kowalczykiewicz, D.; Pollender, A.; Kumpf, A.; Marsden, S.R.; Bento, I.; Jarzębski, A.B.; Szymańska, K.; Chruściel, A.; et al. Leloir Glycosyltransferases in Applied Biocatalysis: A Multidisciplinary Approach. *Int. J. Mol. Sci.* **2019**, *20*, 5263. [CrossRef] [PubMed]
12. Behr, M.; Neutelings, G.; El Jaziri, M.; Baucher, M. You Want it Sweeter: How Glycosylation Affects Plant Response to Oxidative Stress. *Front. Plant Sci.* **2020**, *11*, 571399. [CrossRef]
13. Behr, M.; Speeckaert, N.; Kurze, E.; Morel, O.; Prévost, M.; Mol, A.; Adamou, N.M.; Baragé, M.; Renaut, J.; Schwab, W.; et al. Leaf necrosis resulting from down-regulation of poplar glycosyltransferase UGT72A2. *Tree Physiol.* **2022**, *42*, 1084–1099. [CrossRef] [PubMed]
14. Putkaradze, N.; Teze, D.; Fredslund, F.; Welner, D.H. Natural product C-glycosyltransferases—A scarcely characterised enzymatic activity with biotechnological potential. *Nat. Prod. Rep.* **2021**, *38*, 432–443. [CrossRef] [PubMed]
15. Lee, S.G.; Salomon, E.; Yu, O.; Jez, J.M. Molecular basis for branched steviol glucoside biosynthesis. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 13131–13136. [CrossRef]
16. Wen, C.; Huang, W.; Zhu, X.-L.; Li, X.-S.; Zhang, F.; Jiang, R.-W. UGT74AN1, a Permissive Glycosyltransferase from Asclepias curassavica for the Regiospecific Steroid 3-O-Glycosylation. *Org. Lett.* **2018**, *20*, 534–537. [CrossRef]
17. Sun, G.; Strebl, M.; Merz, M.; Blamberg, R.; Huang, F.-C.; McGraphery, K.; Hoffmann, T.; Schwab, W. Glucosylation of the phytoalexin N-feruloyl tyramine modulates the levels of pathogen-responsive metabolites in Nicotiana benthamiana. *Plant J.* **2019**, *100*, 20–37. [CrossRef]

18. Sun, G.; Putkaradze, N.; Bohnacker, S.; Jonczyk, R.; Fida, T.; Hoffmann, T.; Bernhardt, R.; Härtl, K.; Schwab, W. Six uridine-diphosphate glycosyltransferases catalyze the glycosylation of bioactive C13-apocarotenols. *Plant Physiol.* **2020**, *184*, 1744–1761. [CrossRef]
19. Liao, J.; Sun, G.; Kurze, E.; Steinchen, W.; Hoffmann, T.D.; Song, C.; Zou, Z.; Hoffmann, T.; Schwab, W.G. Subfunctionalization of a monolignol to a phytoalexin glucosyltransferase is accompanied by substrate inhibition. *Plant Commun.* **2022**, *4*, 100506. [CrossRef]
20. Sun, G.; Liao, J.; Kurze, E.; Hoffmann, T.D.; Steinchen, W.; McGraphery, K.; Habegger, R.; Marek, L.; Catici, D.A.M.; Ludwig, C.; et al. Apocarotenoids are allosteric effectors of a dimeric plant glycosyltransferase involved in defense and lignin formation. *New Phytol.* **2023**, *238*, 2080–2098. [CrossRef]
21. Reed, M.C.; Lieb, A.; Nijhout, H.F. The biological significance of substrate inhibition: A mechanism with diverse functions. *Bioessays* **2010**, *32*, 422–429. [CrossRef] [PubMed]
22. Wu, B. Substrate inhibition kinetics in drug metabolism reactions. *Drug Metab. Rev.* **2011**, *43*, 440–456. [CrossRef] [PubMed]
23. LiCata, V.J.; Allewell, N.M. Is substrate inhibition a consequence of allostery in aspartate transcarbamylase? *Biophys. Chem.* **1997**, *64*, 225–234. [CrossRef] [PubMed]
24. Shibuya, A.; Itoh, T.; Tukey, R.H.; Fujiwara, R. Impact of fatty acids on human UDP-glucuronosyltransferase 1A1 activity and its expression in neonatal hyperbilirubinemia. *Sci. Rep.* **2013**, *3*, 2903. [CrossRef]
25. Fenton, A.W.; Reinhart, G.D. Mechanism of substrate inhibition in Escherichia coli phosphofructokinase. *Biochemistry* **2003**, *42*, 12676–12681. [CrossRef]
26. Barnett, A.C.; Tsvetanov, S.; Gamage, N.; Martin, J.L.; Duggleby, R.G.; McManus, M.E. Active site mutations and substrate inhibition in human sulfotransferase 1A1 and 1A3. *J. Biol. Chem.* **2004**, *279*, 18799–18805. [CrossRef]
27. Scheerer, D.; Adkar, B.V.; Bhattacharyya, S.; Levy, D.; Iljina, M.; Riven, I.; Dym, O.; Haran, G.; Shakhnovich, E.I. Allosteric communication between ligand binding domains modulates substrate inhibition in adenylate kinase. *Proc. Natl. Acad. Sci. USA* **2023**, *120*, e2219855120. [CrossRef]
28. Speeckaert, N.; Adamou, N.M.; Hassane, H.A.; Baldacci-Cresp, F.; Mol, A.; Goeminne, G.; Boerjan, W.; Duez, P.; Hawkins, S.; Neutelings, G.; et al. Characterization of the UDP-glycosyltransferase UGT72 Family in Poplar and Identification of Genes Involved in the Glycosylation of Monolignols. *Int. J. Mol. Sci.* **2020**, *21*, 5018. [CrossRef]
29. Speeckaert, N.; El Jaziri, M.; Baucher, M.; Behr, M. UGT72, a Major Glycosyltransferase Family for Flavonoid and Monolignol Homeostasis in Plants. *Biology* **2022**, *11*, 441. [CrossRef]
30. Lanot, A.; Hodge, D.; Jackson, R.G.; George, G.L.; Elias, L.; Lim, E.-K.; Vaistij, F.E.; Bowles, D.J. The glucosyltransferase UGT72E2 is responsible for monolignol 4-O-glucoside production in Arabidopsis thaliana. *Plant J.* **2006**, *48*, 286–295. [CrossRef]
31. Lanot, A.; Hodge, D.; Lim, E.-K.; Vaistij, F.E.; Bowles, D.J. Redirection of flux through the phenylpropanoid pathway by increased glucosylation of soluble intermediates. *Planta* **2008**, *228*, 609–616. [CrossRef]
32. Härtl, K.; Huang, F.-C.; Giri, A.P.; Franz-Oberdorf, K.; Frotscher, J.; Shao, Y.; Hoffmann, T.; Schwab, W. Glucosylation of Smoke-Derived Volatiles in Grapevine (Vitis vinifera) is Catalyzed by a Promiscuous Resveratrol/Guaiacol Glucosyltransferase. *J. Agric. Food Chem.* **2017**, *65*, 5681–5689. [CrossRef]
33. Sugimoto, K.; Iijima, Y.; Takabayashi, J.; Matsui, K. Processing of Airborne Green Leaf Volatiles for Their Glycosylation in the Exposed Plants. *Front. Plant Sci.* **2021**, *12*, 721572. [CrossRef] [PubMed]
34. Sugimoto, K.; Matsui, K.; Iijima, Y.; Akakabe, Y.; Muramoto, S.; Ozawa, R.; Uefune, M.; Sasaki, R.; Alamgir, K.M.; Akitake, S.; et al. Intake and transformation to a glycoside of (Z)-3-hexenol from infested neighbors reveals a mode of plant odor reception and defense. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 7144–7149. [CrossRef] [PubMed]
35. Sun, H.; Wang, L.; Zhang, B.; Ma, J.; Hettenhausen, C.; Cao, G.; Sun, G.; Wu, J.; Wu, J. Scopoletin is a phytoalexin against Alternaria alternata in wild tobacco dependent on jasmonate signalling. *J. Exp. Bot.* **2014**, *65*, 4305–4315. [CrossRef]
36. Quinsey, N.S.; Luong, A.Q.; Dickson, P.W. Mutational analysis of substrate inhibition in tyrosine hydroxylase. *J. Neurochem.* **1998**, *71*, 2132–2138. [CrossRef] [PubMed]
37. Chen, C.; Joo, J.C.; Brown, G.; Stolnikova, E.; Halavaty, A.S.; Savchenko, A.; Anderson, W.F.; Yakunin, A.F. Structure-based mutational studies of substrate inhibition of betaine aldehyde dehydrogenase BetB from Staphylococcus aureus. *Appl. Environ. Microbiol.* **2014**, *80*, 3992–4002. [CrossRef] [PubMed]
38. Ziegler, J.; Brandt, W.; Geissler, R.; Facchini, P.J. Removal of substrate inhibition and increase in maximal velocity in the short chain dehydrogenase/reductase salutaridine reductase involved in morphine biosynthesis. *J. Biol. Chem.* **2009**, *284*, 26758–26767. [CrossRef]
39. Burton, R.L.; Chen, S.; Xu, X.L.; Grant, G.A. A novel mechanism for substrate inhibition in Mycobacterium tuberculosis D-3-phosphoglycerate dehydrogenase. *J. Biol. Chem.* **2007**, *282*, 31517–31524. [CrossRef]
40. Hewitt, C.O.; Eszes, C.M.; Sessions, R.B.; Moreton, K.M.; Dafforn, T.R.; Takei, J.; Dempsey, C.E.; Clarke, A.R.; Holbrook, J.J. A general method for relieving substrate inhibition in lactate dehydrogenases. *Protein Eng.* **1999**, *12*, 491–496. [CrossRef]
41. Kokkonen, P.; Beier, A.; Mazurenko, S.; Damborsky, J.; Bednar, D.; Prokop, Z. Substrate inhibition by the blockage of product release and its control by tunnel engineering. *RSC Chem. Biol.* **2021**, *2*, 645–655. [CrossRef] [PubMed]
42. Kundu, N.; Mondal, D.; Sarkar, N. Dynamics of the vesicles composed of fatty acids and other amphiphile mixtures: Unveiling the role of fatty acids as a model protocell membrane. *Biophys. Rev.* **2020**, *12*, 1117–1131. [CrossRef] [PubMed]
43. Mansy, S.S. Model protocells from single-chain lipids. *Int. J. Mol. Sci.* **2009**, *10*, 835–843. [CrossRef]

44. Pande, S.V.; Mead, J.F. Inhibition of Enzyme Activities by Free Fatty Acids. *J. Biol. Chem.* **1968**, *243*, 6180–6185. [CrossRef] [PubMed]
45. Offenbacher, A.R.; Holman, T.R. Fatty Acid Allosteric Regulation of C-H Activation in Plant and Animal Lipoxygenases. *Molecules* **2020**, *25*, 3374. [CrossRef] [PubMed]
46. Tillman, M.C.; Imai, N.; Li, Y.; Khadka, M.; Okafor, C.D.; Juneja, P.; Adhiyaman, A.; Hagen, S.J.; Cohen, D.E.; Ortlund, E.A. Allosteric regulation of thioesterase superfamily member 1 by lipid sensor domain binding fatty acids and lysophosphatidylcholine. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 22080–22089. [CrossRef]
47. Peng, H.; Yang, T.; Whitaker, B.D.; Shangguan, L.; Fang, J. Calcium/calmodulin alleviates substrate inhibition in a strawberry UDP-glucosyltransferase involved in fruit anthocyanin biosynthesis. *BMC Plant Biol.* **2016**, *16*, 197. [CrossRef]
48. Teo, Y.Y.; Misran, M.; Low, K.H.; Zain, S.M. Effect of Unsaturation on the Stability of C 18 Polyunsaturated Fatty Acids Vesicles Suspension in Aqueous Solution. *Bull. Korean Chem. Soc.* **2011**, *32*, 59–64. [CrossRef]
49. Theander, K.; Pugh, R.J. The Influence of pH and Temperature on the Equilibrium and Dynamic Surface Tension of Aqueous Solutions of Sodium Oleate. *J. Colloid Interface Sci.* **2001**, *239*, 209–216. [CrossRef]
50. Verhagen, J.; Vliegenthart, J.F.; Boldingh, J. Micelle and acid-soap formation of linoleic acid and 13-L-hydroperoxylinoleic acid being substrates of lipoxygenase-1. *Chem. Phys. Lipids* **1978**, *22*, 255–259. [CrossRef]
51. Fameau, A.-L.; Ventureira, J.; Novales, B.; Douliez, J.-P. Foaming and emulsifying properties of fatty acids neutralized by tetrabutylammonium hydroxide. *Colloids Surf. A Physicochem. Eng. Asp.* **2012**, *403*, 87–95. [CrossRef]
52. Glick, J.; Santoyo, G.; Casey, P.J. Arachidonate and related unsaturated fatty acids selectively inactivate the guanine nucleotide-binding regulatory protein. *Gz. J. Biol. Chem.* **1996**, *271*, 2949–2954. [CrossRef] [PubMed]

5. Discussion

5.1. NbUGT72AY1 is involved in lignin biosynthesis, and probably a component of plant defense

Glycosylation serves various functions in plants. Physiologically, glycosylation facilitates inter- and intracellular secondary metabolites transport, storage and accumulation in plant cells. Glycosylation is important for plant growth, stress regulation and detoxification (Figure 12) [126]. Glycosylation of flavonoids enhances their water solubility, allowing for efficient transport within plant cells and through the vascular system. The increased solubility facilitates their movement from leaves to other plant parts [127]. Alkaloids like nicotine in tobacco plants are glycosylated and stored in the vacuoles. Glycosylation reduces the toxicity of hazardous substances within the plant but allows their release when needed for defense against herbivores [128]. Transfer of sugars to plant constituents occurs in various organs such as flowers, fruits, or leaves, and mediates transport and storages processes into the vacuole [126, 129, 130]. UGTs are involved in plant growth and help plants to adapt to environmental stresses, e.g. by glycosylating phytohormones [131]. UGTs glycosylate salicylic acid to form salicylic acid glucoside (SAG). The glycosylation reduces the concentration of salicylic acid, which acts as signalling compound in plant defence against pathogens. Salicylic acid is a crucial component of systemic acquired resistance (SAR) in plants [132]. UGTs also glycosylate auxins like indole-3-acetic acid (IAA). Glycosylated IAA is less active than its free form, which helps regulate auxin-mediated processes like cell elongation and differentiation [133]. In addition, UGTs glycosylate abscisic acid (ABA) to form ABA-glucose ester. This modification impacts ABA's role in plant responses to drought stress and seed dormancy [134]. NbUGT72AY1 is homologous to AtUGT72E2, which is involved in lignin metabolism [135]. AtUGT72E2 catalyzes the formation of coniferyl alcohol 4-O glucoside and sinapyl alcohol 4-O-glucoside. A knockdown mutant line (72E2KD) created through RNAi silencing exhibited a notable twofold reduction in the levels of coniferyl alcohol 4-O-glucoside and sinapyl alcohol 4-O-glucoside when compared to the wild-type counterpart. This reduction

underscores the significance of AtUGT72E2, the target of the RNAi silencing, in the glycosylation of these monolignols [136]. Coniferyl alcohol 4-O-glucoside and sinapyl alcohol 4-O-glucoside are glycosylated forms of monolignols that play roles in lignin metabolism [137]. The glycosylation can influence the solubility, stability, and regulation of monolignols, which are critical components in the formation of lignin, a key structural component in plant cell walls. We found that, NbUGT72AY1 and StUGT72AY2 glycosylate monolignols, especially coniferyl and sinapyl derivatives, and are highly expressed in the stem. NbUGT72AY1 shows high activity and strong substrate inhibition with carvacrol, umbelliferone, scopoletin and vanillin, but low inhibition with monolignols [125]. Since NbUGT72AY1 glycosylates the phytoalexin scopoletin, this indicates that NbUGT72AY1 is involved in plant defense [122] [138]. Glycosylation could reduce the cytotoxic effect of the phytolaexin and allows transport into the vacuole and translocation from the aerial parts to the roots [139]. The glucosylated form of α-ionol has a more pronounced adverse impact on germination rates and embryo size than the unbound α-ionol. This is in contrast to 3-hydroxy-β-ionone and 3-oxo-α-ionol, which exhibit much weaker allelopathic activity than the aglycone compounds. The result suggests that glycosylation does not consistently imply detoxification and the subsequent decrease in phytotoxicity of the aglycones; instead, it can potentially augment phytotoxic effects [121].

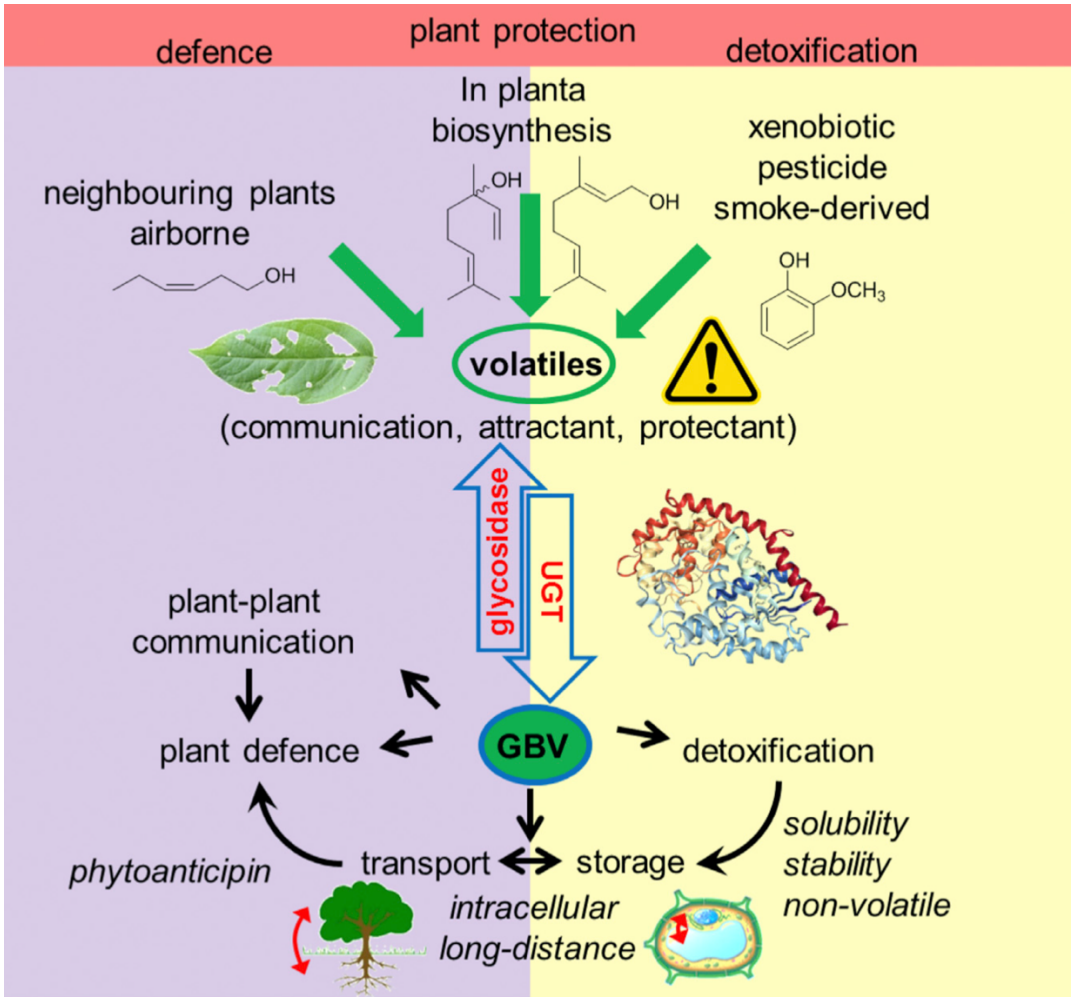


Figure 12. Proposed functions of glycosidically bound volatiles (GBV) and UGTs in plants. Adapted from[126]

5.2. NbUGT72AY1 exhibits substrate inhibition kinetics

Approximately 20% of enzymes exhibit reduced catalytic activity at excessive concentrations of substrates or cofactors, known as substrate inhibition, which probably has a biological function [124]. Substrate inhibition of DNA methyltransferase ensures accurate replication of DNA methylation patterns during cell division and at the same time prevents de novo methylation in promoter regions where methylation is not yet present [124]. In some cases, substrate inhibition may serve as a defense mechanism against the accumulation of toxic intermediates. If the substrate concentration becomes too high, enzymes with substrate inhibition properties can reduce the formation of harmful products. For example, ionol glucoside showed negative impact on germination rates[121]. NbUGT72AY1 is a promiscuous enzyme accepting a wide range of substrates, such as phenolics as well as short-chain

alcohols, terpenoids, and apocarotenoids [122], and is strongly substrate-inhibited by hydroxycoumarins, vanillin, and carvacrol, but only weakly by monolignols [125]. Similarly, FvUGT1 activity is inhibited by anthocyanidin.[25] Besides, NbUGT72AY1 showed a wide range of different activity profiles, ranging from Michaelis–Menten (MM) kinetics (hydroquinone) over weakly inhibited (sinapaldehyde, guaiacol, and o-cresol) to strongly substrate-inhibited profiles (scopoletin, eugenol, and vanillin) [140]. Thus, the tobacco UGT shows no or little substrate inhibition for its presumably original substrates, the monolignols, but substrate inhibition for its acquired glycosylation activity towards the phytoalexins scopoletin and umbelliferone. In the first phase of plant defense, scopoletin is produced, which also has a cytotoxic effect on neighboring plant cells that have not been attacked. Through glycosylation, the neighboring plant cells could protect themselves against scopoletin and even send a signal to other parts of the plant. However, if the infestation of the affected cells is too strong, i.e. an excess of scopoletin is produced and substrate inhibition occurs, it could be more advantageous for the plants if the outer cell layer dies and forms a barrier.

5.3. Advantage und disadvantage of homology modelling

Homology modeling is a valuable tool for predicting protein structures and gaining insights into protein function [141]. Numerous homology modelling tools are available, and comparative studies on the accuracy of these tools have been published [60]. Homology modelling is considered a versatile tool for drug discovery, as it can identify distant homologs by matching sequences to template structures, and contributes to consistent prediction of protein structure by modelling loops and side chains and detecting errors in models [142]. Artificial intelligence in particular contributes a great deal to solving difficult tasks such as the annotation of proteins or open reading frames [62]. Circular Dichroism (CD), X-ray Crystallography, NMR Spectroscopy and Cryo-EM are well-known protein measurement technologies that provide information about protein structure and conformation, but they are time-consuming and expensive [66, 143, 144]. Homology modelling is the better choice for reasons of time and cost.

However, the accuracy of homology modelling is directly influenced by sequence similarity and limitation of known protein structures [56]. Since protein conformational changes are complex, modelling flexible loop regions can be challenging [141]. Another challenge is the conformational changes induced by ligands. By homology modelling we were able to predict the open and closed form of NbUGT72AY1 and postulate a possible allosteric binding site that could explain the substrate inhibition of the enzyme by scopoletin. However, the results of the second study, in which the effect of coumarin effectors on catalytic activity was investigated, could not confirm this binding site. In our opinion, only crystal structures can provide definitive proof of the molecular basis of the substrate inhibition of NbUGT72AY1 by scopoletin and its attenuation by apocarotenoids.

5.4. Molecular basis of substrate inhibition

The open and closed UGT conformers were visualized using 3D structures to capture the dynamics of conformational changes during catalysis and substrate inhibition. In addition, HDX-MS was performed to analyze the conformational changes of NbUGT72AY1 upon binding of scopoletin. Homology modeling shows that NbUGT72AY1 is a typical GT1 family member with GT-B fold, that follows an inverting mechanism. Some GT-B folded UGT enzymes exhibit global domain motion on substrate binding and differ in types and extent of movement [145]. Multiple loop displacements during the transition from the open to the closed conformation of the protein bring the donor and acceptor into spatial proximity [146]. HDX-MS and homology modeling showed that amino acids 87-97, 156-208 and 386-399 are located in flexible regions (Figure 13). Amino acid W350 is the first amino acid in the PSPG box and rotates 180° during the conversion from open to closed conformation. It is involved in the binding of the donor substrate and forms a pi-pi complex with the uridine ring of UDP-glucose.

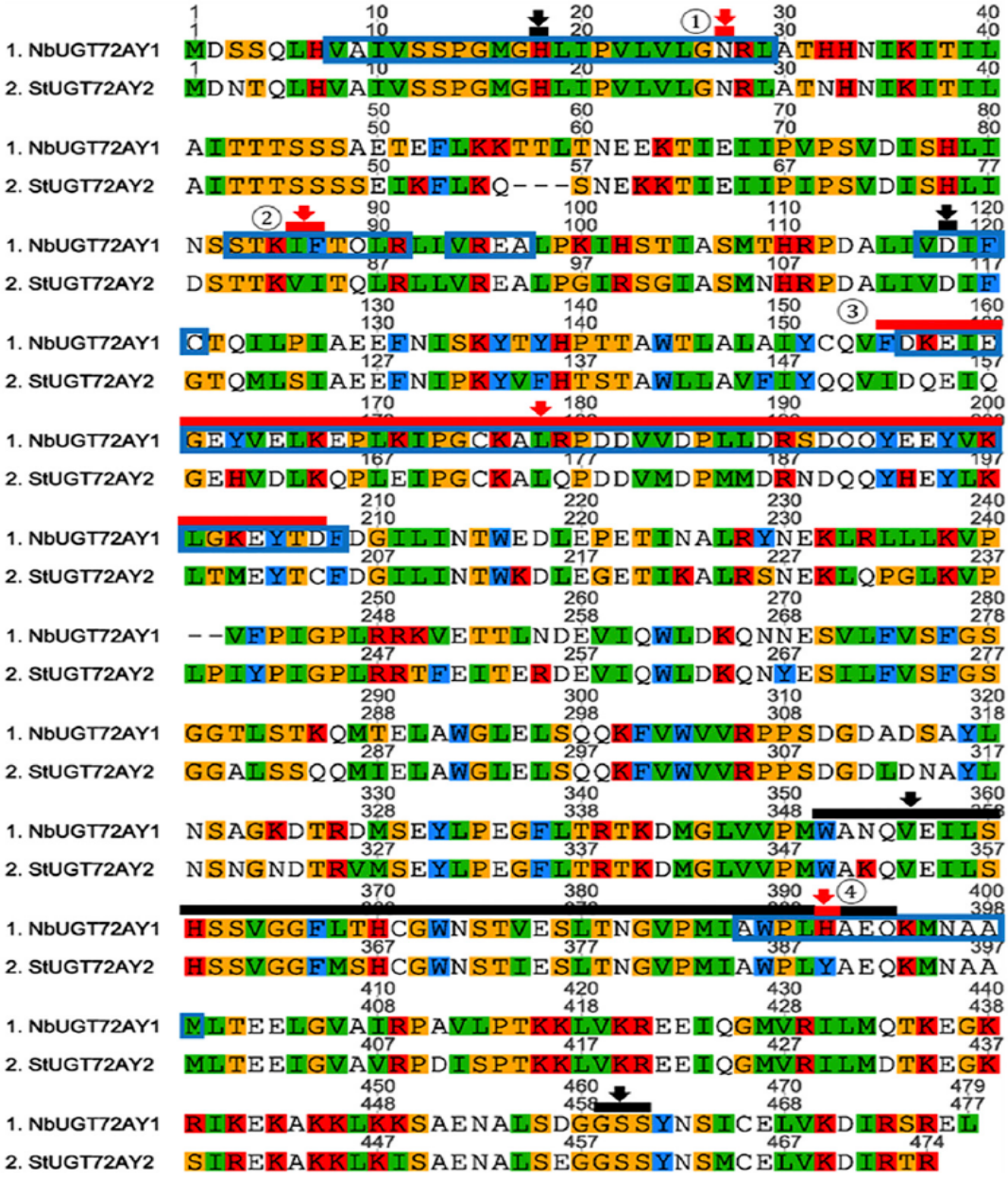


Figure 13. Amino acid sequence alignment of NbUGT72AY1 and StUGT72AY2 from *Nicotiana benthamiana* and *Solanum tuberosum*, respectively and HDX results. Catalytically active H18, activating D118, plant secondary product glycosyltransferase (PSPG) consensus sequence position 352-395, and G461/S462/S463 motif are highlighted by black bars and arrows; mutations 1 N27D, 2 IF/VI double mutant, 3 chimeric mutant, and 4 H/Y mutant are highlighted by red bars and arrows. Amino acids of NbUGT72AY1 that showed reduced hydrogen/deuterium exchange in HDX experiments in the presence of scopoletin are framed with blue boxes. Adpated from[125]

The simulated 3D model for NbUGT72AY1 with the ligand scopoletin showed that the active site is lined mainly with nonpolar amino acids (P14, G15, I86, F87, L90, I119, F120, P186, and A391) (Figure 14). The interactions of the amino acids with scopoletin in the active site stabilize the hydrogen bonding networks, resulting in significant decrease in deuterium

incorporation [147]. The amino acids I86 and F87 were found to be located in the active site, and H390 rotated in the animation, showing the transition from the open to the closed conformer. Since R91 shows a dramatic conformational change during the transition from the open to the closed conformer in the animation, creating a putative new binding site, this suggests an allosteric binding site that should allow the binding of a second scopoletin molecule. Therefore, three R91 mutants, R91A, R91F and R91M were generated. R91A and R91F were significantly stronger inhibited by scopoletin than the wild-type enzyme enzyme, which was reflected in lower Ki values. The strongest effects on substrate inhibition were observed for the chimera and chimera A mutant of the tobacco enzyme, as evidenced by the increased Ki values. In these mutants, sequence segments of the substrate-inhibited NbUGT72AY1 were exchanged with homologous segments of the non-inhibited enzyme StUGT72AY2. These results may indicate that NbUGT72AY1 has an allosteric site encompassing F91 and amino acids of the sequence inserted into chimera A. Proteins investigated with second binding site have been reported a lot. For example, pRib-AMP/ADP binding induces conformational changes in EDS1-PAD4 to allosterically promote its association with ADR1-L1 [148].

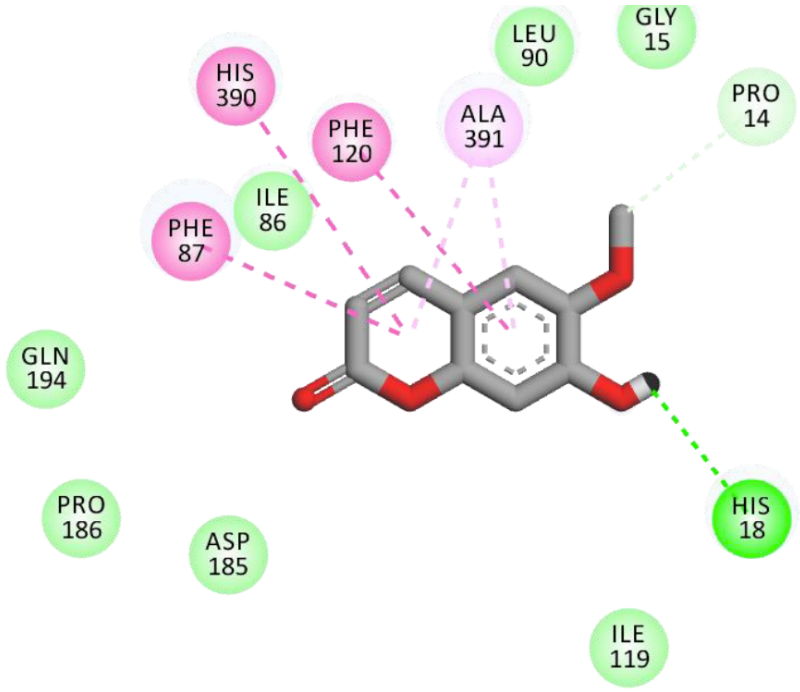


Figure 14. Ligand–protein interactions (scopoletin in the catalytic center) predicted by Discovery Studio Visualizer v19.1.0.18287 (https://discover.3ds.com/discovery-studiovisualizer-Download). Adapted from[125]

5.5. Structural requirements of ligands for substrate inhibition

Vanillin shares a common 2-methoxyphenol structure with scopoletin and also shows substrate inhibition with NbUGT72AY1 (Figure 15). One possible hypothesis for the strong substrate inhibition of scopoletin is a possible allosteric binding site at NbUGT72AY1, which could also bind vanillin. [125]. To confirm the catalytic and secondary binding site, we used vanillin as a substrate and coumarins as optional inhibitors. Coumarins had no effect on the glucosyltransferase activity and substrate inhibition of NbUGT72AY1. This means that coumarins do not affect enzymatic activity, but the result does not completely rule out a second binding site. In a previous study, apocarotenoid effectors were shown to inhibit the inherent UDP-glucose glucohydrolase activity of NbUGT72AY1 and attenuate substrate inhibition by scopoletin, which could also be achieved by mutations.

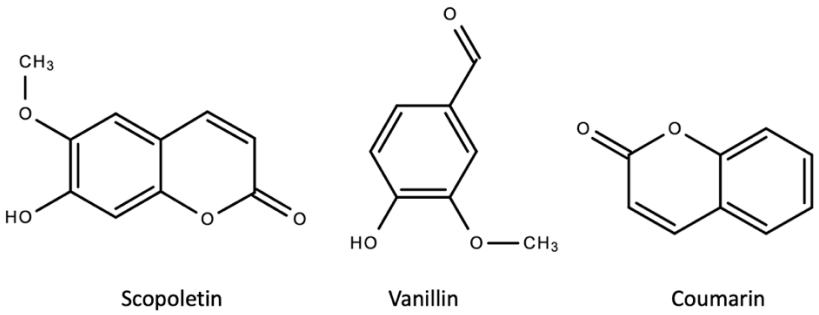


Figure 15. Chemical structures of scopoletin, vanillin, and coumarin.

In the follow-up study, fatty acids were selected and tested as effectors due to their structural similarity to carotenoids and their non-polar character. Fatty acids strongly influenced the substrate inhibition kinetics as they increased the inhibition constant Ki when vanillin was glucosylated by NbUGT72AY1. The molecular mechanism of substrate inhibition is still unclear. A well-known example of substrate inhibition is the inhibition of phosphofructokinase by ATP [124]. The model of two binding sites with one catalytic and one allosteric site is the best known model for inhibition by substrates [23]. Binding of the substrate to the allosteric binding site on the enzyme–substrate complex (E-D-A) results in the formation of an inhibitory complex (A-E-D-A). This complex slows down or completely suppresses the catalyzed reaction (Figure 16a). In alternative models, excess substrate

molecules interact with enzyme forms other than the enzyme–substrate complex, such as the reaction intermediate (EI) or the enzyme–product complex (EP). All of these models share a common feature: the presence of a second substrate molecule bound to the enzyme. The inhibitory effects of substrates are attributed to the accumulation of a catalytically incompetent combination of enzyme, cofactor, and substrate. A more recent model based on conformational motions could apply in our case (Figure 16b) [149]. Using single-molecule FRET spectroscopy, it was revealed that acceptor substrates alone cannot promote domain closure of a kinase, but inhibitory concentrations of acceptor lead to a faster and more cooperative domain closure by the donor, leading in turn to an increased population of the closed state [63]. In this context, NbUGT72AY1 could serve as an ideal model for deeper research into substrate inhibition dynamics in the future, as it exhibits pronounced substrate Inhibition kinetics.

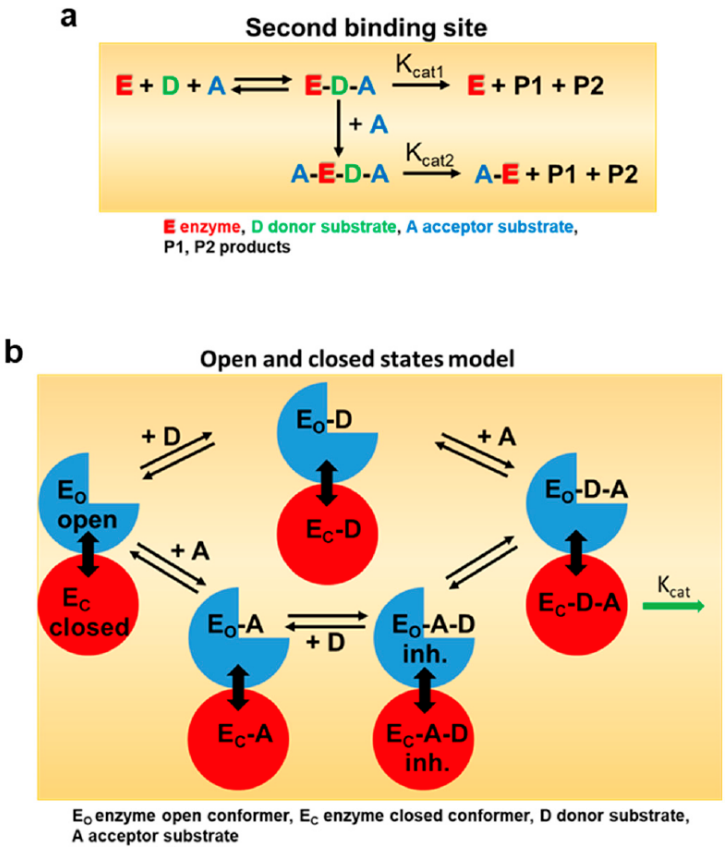


Figure 16. Models of enzymatic substrate inhibition. (a) Classical model in which a second substrate A is bound to an allosteric site on the enzyme-substrate complex (E-D-A) (adapted from[23]). (b) Inhibitory concentrations of the acceptor substrate A lead to faster and more cooperative domain closure by the donor substrate D, which in turn leads to an increased population of the closed inhibited state (E_c-A-D) (adapted from[63]). Adapted from[140]

5.6. Glucohydrolase activity of NbUGT72AY1

In a subsequent investigation, NbUGT72AY1 was identified to possess UDPG glucohydrolase activity, indicating the enzyme's ability to transfer glucose to water in the absence of an acceptor molecule. This phenomenon, observed in other transferases as well (Table 1), was effectively counteracted by apocarotenoids such as α- and β-ionol and α- and β-ionone [150]. Overall, in the absence of an acceptor substrate, GTs often exhibit background hydrolysis of the donor substrate, which can be regarded as a side-reaction with water as sugar-acceptor. Amino acids located at the two apertures could affect the efflux of water from the active site and thus control the inherent UDP-Glc hydrolase activity of UGTs [151].

Table 1. Glycosyltransferase with hydrolase activity

Glycosyltransferases	Hydrolase activity
AtFUT1	An atypical water-mediated fucosylation mechanism facilitated by an H-bonded network with S_N1-like reaction mechanism [101]
LtpM	Amino acids located at the two apertures could affect the efflux of water from the active site and thus control the inherent UDP-Glc hydrolase activity of UGTs [152].
Human blood group B galactosyltransferase	Hydrolysis may be understood as a glycosyl transfer reaction where water serves as universal acceptor, which indicates that this may point to a general mechanism used by retaining glycosyltransferases to discriminate acceptor substrates under physiological conditions [153].

6. Conclusion and Outlook

The precise 3D structure of NbUGT72AY1 is not yet known to explain the exact mechanism of substrate inhibition, UDP-glucose glucohydrolase activity and the effect of ligands at the molecular level. Currently, only HDX-MS and *in silico* simulations of the structure are available. In the future, the advantages of further techniques such as NMR or X-ray crystallography should be utilized to investigate the conformational changes during catalysis of NbUGT72AY1 with acceptor substrates, UDP-glucose, and effectors. Since the observed enzyme kinetics can be well explained by Hill equations without understanding the mechanistic basis, the elucidation of the mechanism could also contribute to the improvement of mathematical models and the generation of improved glycosyltransferases.

7. Publications

Below is a complete list of publications during my PhD:

1. **Jieren Liao**, Veronika Lederer, Alba Bardhi, Zhiwei Zou, Timothy D. Hoffmann, Guangxin Sun, Chuankui Song, Thomas Hoffmann, Wilfried Schwab. Acceptors and Effectors Alter Substrate Inhibition Kinetics of a Plant Glucosyltransferase NbUGT72AY1 and Its Mutants. Int. J. Mol. Sci. 2023, 24, 9542. https://doi.org/10.3390/ijms24119542

2. Guangxin Sun†, **Jieren Liao**†, Elisabeth Kurze, Timothy D. Hoffmann, Wieland Steinchen, Kate McGraphery, Ruth Habegger, Ludwig Marek, Dragana A. M. Catici, Christina Ludwig, Tingting Jing, Thomas Hoffmann, Chuankui Song, Schwab, Wilfried G .Apocarotenoids are allosteric effectors of a dimeric plant glycosyltransferase involved in defense and lignin formation New Phytologist 2023, https://doi.org/10.1111/nph.18875.

3. **Liao, Jieren**†; Sun, Guangxin†; Kurze, Elisabeth†; Steinchen, Wieland; Hoffmann, Timothy D.; Song, Chuankui; Zou, Zhiwei; Hoffmann, Thomas; Schwab, Wilfried G. Subfunctionalization of a monolignol to a phytoalexin glucosyltransferase is accompanied by substrate inhibition. Plant Communications 2022, doi: 10.1016/j.xplc.2022.100506.

4. Kurze, Elisabeth; Ruß, Victoria; Syam, Nadia; Effenberger, Isabelle; Jonczyk, Rafal; **Liao, Jieren**; Song, Chuankui; Hoffmann, Thomas; Schwab, Wilfried. Glucosylation of (±)-Menthol by Uridine-Diphosphate-Sugar Dependent Glucosyltransferases from Plants. Molecules 2021, doi: 10.3390/molecules26185511.

5. Kurze, Elisabeth; Wüst, Matthias; **Liao, Jieren**; McGraphery, Kate; Hoffmann, Thomas; Song, Chuankui; Schwab, Wilfried. Structure–function relationship of terpenoid glycosyltransferases from plants. Natural Product Reports 2021, doi: 10.1039/d1np00038a.

Oral Presentation
Research Seminar Food Chemistry and Food Quality 2023 on 17.07.2023 in Freising, germany

Poster Presentation
Agricultural seminar 2023 on 25.04.2023 in Freising

Acknowledgments

Four years of studying and living in Germany. I have experienced a lot. My gratitude for the people in my life over the past four years is hard to put into words.

I cannot thank Prof. Wilfried Schwab enough. I still recall the first time I met him at the Nanjing airport, and he gave me the opportunity to pursue a PhD. Additionally, starting university with poor English and tense interactions with peers is a challenge. Without his encouragement and guidance, I do not know how my life would be as a doctoral student. Moreover, I want to express my gratitude to Dr. Thomas Hoffmann, my mentor. He supported me in all of my experiments, including LC-MS and fermenter, and was always kind and understanding.

In addition, I would like to thank Dr. Elisabeth Kurze, from whom I learned a lot about designing experiments and how to conduct them. I consider myself very lucky to share an office with her and to have the opportunity to observe her rigid, highly successful, and responsible working style.

I would also thank Dr. Soraya Chebib, a very kind woman who always encouraged me and offered me help when I got into trouble and helped me solve my problems with the apartment rent. She gave me a lot of helpful advice since it was my first time living and studying abroad. I appreciate that I could work with Dr. Timothy Dennis Hoffmann. We shared the same office and worked on the same project. He is gentle and took care of the laboratory rules. He joined in our lab two years later than I did, we worked together and progressed together. I enjoyed discussing hobbies and social issues with him, which broaden my view.

Finally, I would like to thank Drs. Emilia Romer, Annika Haugeneder, Johanna Trinkl, Kate McGraphery, Julian Rüdiger, Martina Kolarek, Rafal Jonczyk, Guangxing Sun, Sadiq Mareai, and Nicolas Figueroa, and the doctoral candidates Martha Wulanjati, Shuai Zhao, and Xiran Wang. They made my transition from a new Ph.D. student to an independent Ph.D. student very easy.

Studying abroad thanks to the support of my family, friends and students.

Finally, I will never forget Heike Adamski, Mechthild Mayershofer, Ruth Habegger and Anja Forstner.

Our work cannot progress without them.

Time flies by. Thanks for the pleasant meetings with everyone at the most convenient time. Every memory with everyone is one to be treasured.

References

1. Martinez Cuesta, S.; Rahman, S. A.; Furnham, N.; Thornton, J. M., The Classification and Evolution of Enzyme Function. *Biophys J* **2015,** *109* (6), 1082–6.
2. Mu, R.; Wang, Z.; Wamsley, M. C.; Duke, C. N.; Lii, P. H.; Epley, S. E.; Todd, L. C.; Roberts, P. J., Application of Enzymes in Regioselective and Stereoselective Organic Reactions. *Catalysts* **2020,** *10* (8).
3. Mesbah, N. M., Industrial Biotechnology Based on Enzymes From Extreme Environments. *Front Bioeng Biotechnol* **2022,** *10*, 870083.
4. Caputi, L.; Malnoy, M.; Goremykin, V.; Nikiforova, S.; Martens, S., A genome-wide phylogenetic reconstruction of family 1 UDP-glycosyltransferases revealed the expansion of the family during the adaptation of plants to life on land. *Plant J* **2012,** *69* (6), 1030–42.
5. Lundin, D.; Poole, A. M.; Sjoberg, B. M.; Hogbom, M., Use of structural phylogenetic networks for classification of the ferritin-like superfamily. *J Biol Chem* **2012,** *287* (24), 20565–75.
6. Aziz, M. F.; Caetano-Anolles, G., Evolution of networks of protein domain organization. *Sci Rep* **2021,** *11* (1), 12075.
7. Donertas, H. M.; Martinez Cuesta, S.; Rahman, S. A.; Thornton, J. M., Characterising Complex Enzyme Reaction Data. *PLoS One* **2016,** *11* (2), e0147952.
8. Ribeiro, A. J. M.; Holliday, G. L.; Furnham, N.; Tyzack, J. D.; Ferris, K.; Thornton, J. M., Mechanism and Catalytic Site Atlas (M-CSA): a database of enzyme reaction mechanisms and active sites. *Nucleic Acids Res* **2018,** *46* (D1), D618–D623.
9. Andreini, C.; Bertini, I.; Cavallaro, G.; Holliday, G. L.; Thornton, J. M., Metal-MACiE: a database of metals involved in biological catalysis. *Bioinformatics* **2009,** *25* (16), 2088–9.
10. Akiva, E.; Brown, S.; Almonacid, D. E.; Barber, A. E., 2nd; Custer, A. F.; Hicks, M. A.; Huang, C. C.; Lauck, F.; Mashiyama, S. T.; Meng, E. C.; Mischel, D.; Morris, J. H.; Ojha, S.; Schnoes, A. M.; Stryke, D.; Yunes, J. M.; Ferrin, T. E.; Holliday, G. L.; Babbitt, P. C., The Structure-Function Linkage Database. *Nucleic Acids Res* **2014,** *42* (Database issue), D521–30.
11. Nagano, N., EzCatDB: the Enzyme Catalytic-mechanism Database. *Nucleic Acids Res* **2005,** *33* (Database issue), D407–12.
12. Holliday, G. L.; Brown, S. D.; Akiva, E.; Mischel, D.; Hicks, M. A.; Morris, J. H.; Huang, C. C.; Meng, E. C.; Pegg, S. C.; Ferrin, T. E.; Babbitt, P. C., Biocuration in the structure-function linkage database: the anatomy of a superfamily. *Database (Oxford)* **2017,** *2017* (1).
13. Ausiello, G.; Via, A.; Helmer-Citterich, M., Query3d: a new method for high-throughput analysis of functional residues in protein structures. *BMC Bioinformatics* **2005,** *6 Suppl 4* (Suppl 4), S5.
14. Drula, E.; Garron, M. L.; Dogan, S.; Lombard, V.; Henrissat, B.; Terrapon, N., The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res* **2022,** *50* (D1), D571–D577.
15. Gagler, D. C.; Karas, B.; Kempes, C.; Goldman, A. D.; Kim, H.; Walker, S. I., **2021**.
16. Gattiker, A.; Michoud, K.; Rivoire, C.; Auchincloss, A. H.; Coudert, E.; Lima, T.; Kersey, P.; Pagni, M.; Sigrist, C. J.; Lachaize, C.; Veuthey, A. L.; Gasteiger, E.; Bairoch, A., Automated annotation of microbial proteomes in SWISS-PROT. *Comput Biol Chem* **2003,** *27* (1), 49–58.
17. Hong, J.; Luo, Y.; Zhang, Y.; Ying, J.; Xue, W.; Xie, T.; Tao, L.; Zhu, F., Protein functional annotation of simultaneously improved stability, accuracy and false discovery rate achieved by a sequence-based deep learning. *Brief Bioinform* **2020,** *21* (4), 1437–1447.
18. Karp, P. D.; Ong, W. K.; Paley, S.; Billington, R.; Caspi, R.; Fulcher, C.; Kothari, A.; Krummenacker,

M.; Latendresse, M.; Midford, P. E.; Subhraveti, P.; Gama-Castro, S.; Muniz-Rascado, L.; Bonavides-Martinez, C.; Santos-Zavaleta, A.; Mackie, A.; Collado-Vides, J.; Keseler, I. M.; Paulsen, I., The EcoCyc Database. *EcoSal Plus* **2018,** *8* (1).

19. Kanehisa, M.; Sato, Y.; Kawashima, M., KEGG mapping tools for uncovering hidden features in biological data. *Protein Sci* **2022,** *31* (1), 47-53.

20. Chiu, S. H.; Chen, C. C.; Yuan, G. F.; Lin, T. H., Association algorithm to mine the rules that govern enzyme definition and to classify protein sequences. *BMC Bioinformatics* **2006,** *7*, 304.

21. Srinivasan, B., A guide to the Michaelis-Menten equation: steady state and beyond. *FEBS J* **2022,** *289* (20), 6086-6098.

22. Gunawardena, J., Time-scale separation--Michaelis and Mentens old idea, still bearing fruit. *FEBS J* **2014,** *281* (2), 473-88.

23. Wu, B., Substrate inhibition kinetics in drug metabolism reactions. *Drug Metab Rev* **2011,** *43* (4), 440-56.

24. Kapelyukh, Y.; Paine, M. J.; Marechal, J. D.; Sutcliffe, M. J.; Wolf, C. R.; Roberts, G. C., Multiple substrate binding by cytochrome P450 3A4: estimation of the number of bound substrate molecules. *Drug Metab Dispos* **2008,** *36* (10), 2136-44.

25. Peng, H.; Yang, T.; Whitaker, B. D.; Shangguan, L.; Fang, J., Calcium/calmodulin alleviates substrate inhibition in a strawberry UDP-glucosyltransferase involved in fruit anthocyanin biosynthesis. *BMC Plant Biol* **2016,** *16* (1), 197.

26. Kokkonen, P.; Bednar, D.; Pinto, G.; Prokop, Z.; Damborsky, J., Engineering enzyme access tunnels. *Biotechnol Adv* **2019,** *37* (6), 107386.

27. Laddach, A.; Chung, S. S.; Fraternali, F., Prediction of Protein-Protein Interactions: Looking Through the Kaleidoscope. In *Encyclopedia of Bioinformatics and Computational Biology*, 2019; pp 834-848.

28. Maiti, A.; Drohat, A. C., Dependence of substrate binding and catalysis on pH, ionic strength, and temperature for thymine DNA glycosylase: Insights into recognition and processing of G.T mispairs. *DNA Repair (Amst)* **2011,** *10* (5), 545-53.

29. Goh, C. S.; Milburn, D.; Gerstein, M., Conformational changes associated with protein-protein interactions. *Curr Opin Struct Biol* **2004,** *14* (1), 104-9.

30. Monzon, A. M.; Fornasari, M. S.; Zea, D. J.; Parisi, G., Exploring Protein Conformational Diversity. *Methods Mol Biol* **2019,** *1851*, 353-365.

31. Parisi, G.; Zea, D. J.; Monzon, A. M.; Marino-Buslje, C., Conformational diversity and the emergence of sequence signatures during evolution. *Curr Opin Struct Biol* **2015,** *32*, 58-65.

32. Saldano, T. E.; Tosatto, S. C. E.; Parisi, G.; Fernandez-Alberti, S., Network analysis of dynamically important residues in protein structures mediating ligand-binding conformational changes. *Eur Biophys J* **2019,** *48* (6), 559-568.

33. Guterres, H.; Park, S. J.; Jiang, W.; Im, W., Ligand-Binding-Site Refinement to Generate Reliable Holo Protein Structure Conformations from Apo Structures. *J Chem Inf Model* **2021,** *61* (1), 535-546.

34. Maria-Solano, M. A.; Serrano-Hervas, E.; Romero-Rivera, A.; Iglesias-Fernandez, J.; Osuna, S., Role of conformational dynamics in the evolution of novel enzyme function. *Chem Commun (Camb)* **2018,** *54* (50), 6622-6634.

35. Mobley, D. L.; Dill, K. A., Binding of small-molecule ligands to proteins: "what you see" is not always "what you get". *Structure* **2009,** *17* (4), 489-98.

36. Cornish-Bowden, A., Introduction: Enzyme catalysis and allostery: a century of advances in

molecular understanding. *FEBS J* **2014,** *281* (2), 433–4.
37. Marzen, S.; Garcia, H. G.; Phillips, R., Statistical mechanics of Monod–Wyman–Changeux (MWC) models. *J Mol Biol* **2013,** *425* (9), 1433–60.
38. Hilser, V. J.; Wrabl, J. O.; Motlagh, H. N., Structural and energetic basis of allostery. *Annu Rev Biophys* **2012,** *41*, 585–609.
39. Sikosek, T.; Chan, H. S., Biophysics of protein evolution and evolutionary protein biophysics. *J R Soc Interface* **2014,** *11* (100), 20140419.
40. Chu, W. T.; Yan, Z.; Chu, X.; Zheng, X.; Liu, Z.; Xu, L.; Zhang, K.; Wang, J., Physics of biomolecular recognition and conformational dynamics. *Rep Prog Phys* **2021,** *84* (12).
41. Khan, S. H.; Braet, S. M.; Koehler, S. J.; Elacqua, E.; Anand, G. S.; Okafor, C. D., Ligand–induced shifts in conformational ensembles that describe transcriptional activation. *Elife* **2022,** *11*.
42. Gupta, P.; Mohanty, D., Allosteric regulation of the inactive to active state conformational transition in CDPK1 protein of Plasmodium falciparum. *Int J Biol Macromol* **2022,** *215*, 489–500.
43. Wernimont, A. K.; Artz, J. D.; Finerty, P., Jr.; Lin, Y. H.; Amani, M.; Allali–Hassani, A.; Senisterra, G.; Vedadi, M.; Tempel, W.; Mackenzie, F.; Chau, I.; Lourido, S.; Sibley, L. D.; Hui, R., Structures of apicomplexan calcium–dependent protein kinases reveal mechanism of activation by calcium. *Nat Struct Mol Biol* **2010,** *17* (5), 596–601.
44. Wingert, B.; Krieger, J.; Li, H.; Bahar, I., Adaptability and specificity: how do proteins balance opposing needs to achieve function? *Curr Opin Struct Biol* **2021,** *67*, 25–32.
45. Yu, H.; Dalby, P. A., Exploiting correlated molecular–dynamics networks to counteract enzyme activity–stability trade–off. *Proc Natl Acad Sci U S A* **2018,** *115* (52), E12192–E12200.
46. ODonnell, T.; Cazals, F., Enhanced conformational exploration of protein loops using a global parameterization of the backbone geometry. *J Comput Chem* **2023,** *44* (11), 1094–1104.
47. Groban, E. S.; Narayanan, A.; Jacobson, M. P., Conformational changes in protein loops and helices induced by post–translational phosphorylation. *PLoS Comput Biol* **2006,** *2* (4), e32.
48. Ruvinsky, A. M.; Kirys, T.; Tuzikov, A. V.; Vakser, I. A., Side–chain conformational changes upon Protein–Protein Association. *J Mol Biol* **2011,** *408* (2), 356–65.
49. Bunzel, H. A.; Anderson, J. L. R.; Hilvert, D.; Arcus, V. L.; van der Kamp, M. W.; Mulholland, A. J., Evolution of dynamical networks enhances catalysis in a designer enzyme. *Nat Chem* **2021,** *13* (10), 1017–1022.
50. Savir, Y.; Tlusty, T., Conformational proofreading: the impact of conformational changes on the specificity of molecular recognition. *PLoS One* **2007,** *2* (5), e468.
51. Vinciauskaite, V.; Masson, G. R., Fundamentals of HDX–MS. *Essays Biochem* **2023,** *67* (2), 301–314.
52. Narang, D.; Lento, C.; D, J. W., HDX–MS: An Analytical Tool to Capture Protein Motion in Action. *Biomedicines* **2020,** *8* (7).
53. Salmas, R. E.; Harris, M. J.; Borysik, A. J., Mapping HDX–MS Data to Protein Conformations through Training Ensemble–Based Models. *J Am Soc Mass Spectrom* **2023,** *34* (9), 1989–1997.
54. Woods, V. A.; Abzalimov, R. R.; Keedy, D. A., Native dynamics and allosteric responses in PTP1B probed by high–resolution HDX–MS. *bioRxiv* **2023**.
55. Uggowitzer, K. A.; Shao, A. R. Q.; Habibi, Y.; Zhang, Q. E.; Thibodeaux, C. J., Exploring the Heterogeneous Structural Dynamics of Class II Lanthipeptide Synthetases with Hydrogen–Deuterium Exchange Mass Spectrometry (HDX–MS). *Biochemistry* **2022,** *61* (19), 2118–2130.
56. Dalton, J. A.; Jackson, R. M., An evaluation of automated homology modelling methods at low target

template sequence similarity. *Bioinformatics* **2007,** *23* (15), 1901–8.
57. Madhusudhan, M. S.; Webb, B. M.; Marti–Renom, M. A.; Eswar, N.; Sali, A., Alignment of multiple protein structures based on sequence and structure features. *Protein Eng Des Sel* **2009,** *22* (9), 569–74.
58. Larsson, P.; Wallner, B.; Lindahl, E.; Elofsson, A., Using multiple templates to improve quality of homology models in automated homology modeling. *Protein Sci* **2008,** *17* (6), 990–1002.
59. Webb, B.; Sali, A., Comparative Protein Structure Modeling Using MODELLER. *Curr Protoc Bioinformatics* **2016,** *54*, 5 6 1–5 6 37.
60. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T., SWISS–MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* **2018,** *46* (W1), W296–W303.
61. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Zidek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera–Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D., Highly accurate protein structure prediction with AlphaFold. *Nature* **2021,** *596* (7873), 583–589.
62. Llinares–Lopez, F.; Berthet, Q.; Blondel, M.; Teboul, O.; Vert, J. P., Deep embedding and alignment of protein sequences. *Nat Methods* **2023,** *20* (1), 104–111.
63. Godbole, S.; Dokholyan, N. V., Allosteric regulation of kinase activity in living cells. *bioRxiv* **2023**.
64. Komives, E. A., Dynamic allostery in thrombin–a review. *Front Mol Biosci* **2023,** *10*, 1200465.
65. Keedy, D. A., Journey to the center of the protein: allostery from multitemperature multiconformer X–ray crystallography. *Acta Crystallogr D Struct Biol* **2019,** *75* (Pt 2), 123–137.
66. Grutsch, S.; Bruschweiler, S.; Tollinger, M., NMR Methods to Study Dynamic Allostery. *PLoS Comput Biol* **2016,** *12* (3), e1004620.
67. Boulton, S.; Melacini, G., Advances in NMR Methods To Map Allosteric Sites: From Models to Translation. *Chem Rev* **2016,** *116* (11), 6267–304.
68. Reddy, B.; Fields, R., Multiple Sequence Alignment Algorithms in Bioinformatics. In *Smart Trends in Computing and Communications*, 2022; pp 89–98.
69. Nuin, P. A.; Wang, Z.; Tillier, E. R., The accuracy of several multiple sequence alignment programs for proteins. *BMC Bioinformatics* **2006,** *7*, 471.
70. Proctor, E. A.; Kota, P.; Aleksandrov, A. A.; He, L.; Riordan, J. R.; Dokholyan, N. V., Rational Coupled Dynamics Network Manipulation Rescues Disease–Relevant Mutant Cystic Fibrosis Transmembrane Conductance Regulator. *Chem Sci* **2015,** *6* (2), 1237–1246.
71. Shirvanyants, D.; Ding, F.; Tsao, D.; Ramachandran, S.; Dokholyan, N. V., Discrete molecular dynamics: an efficient and versatile simulation method for fine protein characterization. *J Phys Chem B* **2012,** *116* (29), 8375–82.
72. Wang, J.; Jain, A.; McDonald, L. R.; Gambogi, C.; Lee, A. L.; Dokholyan, N. V., Mapping allosteric communications within individual proteins. *Nat Commun* **2020,** *11* (1), 3862.
73. Kihn, K.; Marchiori, E.; Spagnolli, G.; Boldrini, A.; Terruzzi, L.; Lawrence, D. A.; Gershenson, A.; Faccioli, P.; Wintrode, P. L., Long–range allostery mediates the regulation of plasminogen activator inhibitor–1 by cell adhesion factor vitronectin. *J Biol Chem* **2022,** *298* (12), 102652.
74. Khan, S.; Gati, C., Decoding the mechanisms of allostery. *Elife* **2023,** *12*.
75. Choi, J. M.; Han, S. S.; Kim, H. S., Industrial applications of enzyme biocatalysis: Current status and

future aspects. *Biotechnol Adv* **2015,** *33* (7), 1443–54.
76. Dyer, R. P.; Weiss, G. A., Making the cut with protease engineering. *Cell Chem Biol* **2022,** *29* (2), 177–190.
77. Lorsch, J. R., Practical steady–state enzyme kinetics. *Methods Enzymol* **2014,** *536*, 3–15.
78. Aleksandra Maršavelski, a. S. L., b Marko Močibob,b Ivana Weygand– Đurašević,†b and Sanja Tomić*a, A single amino acid substitution affects the substrate specificity of the seryl–tRNA synthetase homologue. **2012, 10(12), 3207-16**.
79. Chen, Z.; Fu, Y.; Liu, S.; Huang, X.; Kong, X.; Mao, Z.; Hu, N.; Zhang, F.; Han, C., Characterization of aspartokinase double mutants using a combination of experiments and simulations. *Heliyon* **2023,** *9* (2), e13133.
80. Chen, Y.; Guo, X.; Gao, T.; Zhang, N.; Wan, X.; Schwab, W.; Song, C., UGT74AF3 enzymes specifically catalyze the glucosylation of 4–hydroxy–2,5–dimethylfuran–3(2H)–one, an important volatile compound in Camellia sinensis. *Horticulture Research* **2020,** *7* (1).
81. Zhou, Y.; Deng, R.; Xu, X.; Yang, Z., Enzyme Catalytic Efficiencies and Relative Gene Expression Levels of (R)–Linalool Synthase and (S)–Linalool Synthase Determine the Proportion of Linalool Enantiomers in Camellia sinensis var. sinensis. *J Agric Food Chem* **2020,** *68* (37), 10109–10117.
82. Heckmann, C. M.; Paradisi, F., Looking Back: A Short History of the Discovery of Enzymes and How They Became Powerful Chemical Tools. *ChemCatChem* **2020,** *12* (24), 6082–6102.
83. Heather, J. M.; Chain, B., The sequence of sequencers: The history of sequencing DNA. *Genomics* **2016,** *107* (1), 1–8.
84. Reetz, M. T.; Bocola, M.; Carballeira, J. D.; Zha, D.; Vogel, A., Expanding the Range of Substrate Acceptance of Enzymes: Combinatorial Active–Site Saturation Test. *Angewandte Chemie* **2005,** *117* (27), 4264–4268.
85. Acevedo–Rocha, C. G.; Sun, Z.; Reetz, M. T., Clarifying the Difference between Iterative Saturation Mutagenesis as a Rational Guide in Directed Evolution and OmniChange as a Gene Mutagenesis Technique. *Chembiochem* **2018,** *19* (24), 2542–2544.
86. Wang, J.; Cao, H.; Zhang, J. Z. H.; Qi, Y., Computational Protein Design with Deep Learning Neural Networks. *Sci Rep* **2018,** *8* (1), 6349.
87. Jemli, S.; Ayadi–Zouari, D.; Hlima, H. B.; Bejar, S., Biocatalysts: application and engineering for industrial purposes. *Crit Rev Biotechnol* **2016,** *36* (2), 246–58.
88. Sola, R. J.; Griebenow, K., Glycosylation of therapeutic proteins: an effective strategy to optimize efficacy. *BioDrugs* **2010,** *24* (1), 9–21.
89. Kurze, E.; Russ, V.; Syam, N.; Effenberger, I.; Jonczyk, R.; Liao, J.; Song, C.; Hoffmann, T.; Schwab, W., Glucosylation of (+/–)–Menthol by Uridine–Diphosphate–Sugar Dependent Glucosyltransferases from Plants. *Molecules* **2021,** *26* (18).
90. de Brito Francisco, R.; Martinoia, E., The Vacuolar Transportome of Plant Specialized Metabolites. *Plant Cell Physiol* **2018,** *59* (7), 1326–1336.
91. He, B.; Bai, X.; Tan, Y.; Xie, W.; Feng, Y.; Yang, G. Y., Glycosyltransferases: Mining, engineering and applications in biosynthesis of glycosylated plant natural products. *Synth Syst Biotechnol* **2022,** *7* (1), 602–620.
92. Schwab, W.; Fischer, T.; Wüst, M., Terpene glucoside production: Improved biocatalytic processes using glycosyltransferases. *Engineering in Life Sciences* **2015,** *15* (4), 376–386.
93. Thibodeaux, C. J.; Melancon, C. E., 3rd; Liu, H. W., Natural–product sugar biosynthesis and

enzymatic glycodiversification. *Angew Chem Int Ed Engl* **2008,** *47* (51), 9814–59.
94. Kurze, E.; Wust, M.; Liao, J.; McGraphery, K.; Hoffmann, T.; Song, C.; Schwab, W., Structure–function relationship of terpenoid glycosyltransferases from plants. *Nat Prod Rep* **2022,** *39* (2), 389–409.
95. Nothaft, H.; Szymanski, C. M., Protein glycosylation in bacteria: sweeter than ever. *Nat Rev Microbiol* **2010,** *8* (11), 765–78.
96. Kowarik, M.; Young, N. M.; Numao, S.; Schulz, B. L.; Hug, I.; Callewaert, N.; Mills, D. C.; Watson, D. C.; Hernandez, M.; Kelly, J. F.; Wacker, M.; Aebi, M., Definition of the bacterial N–glycosylation site consensus sequence. *EMBO J* **2006,** *25* (9), 1957–66.
97. Mestrom, L.; Przypis, M.; Kowalczykiewicz, D.; Pollender, A.; Kumpf, A.; Marsden, S. R.; Bento, I.; Jarzebski, A. B.; Szymanska, K.; Chrusciel, A.; Tischler, D.; Schoevaart, R.; Hanefeld, U.; Hagedoorn, P. L., Leloir Glycosyltransferases in Applied Biocatalysis: A Multidisciplinary Approach. *Int J Mol Sci* **2019,** *20* (21).
98. Uttaro, A. D., Biosynthesis of polyunsaturated fatty acids in lower eukaryotes. *IUBMB Life* **2006,** *58* (10), 563–71.
99. Liang, D. M.; Liu, J. H.; Wu, H.; Wang, B. B.; Zhu, H. J.; Qiao, J. J., Glycosyltransferases: mechanisms and applications in natural product development. *Chem Soc Rev* **2015,** *44* (22), 8350–74.
100. Schuman, B.; Evans, S. V.; Fyles, T. M., Geometric attributes of retaining glycosyltransferase enzymes favor an orthogonal mechanism. *PLoS One* **2013,** *8* (8), e71077.
101. Urbanowicz, B. R.; Bharadwaj, V. S.; Alahuhta, M.; Pena, M. J.; Lunin, V. V.; Bomble, Y. J.; Wang, S.; Yang, J. Y.; Tuomivaara, S. T.; Himmel, M. E.; Moremen, K. W.; York, W. S.; Crowley, M. F., Structural, mutagenic and in silico studies of xyloglucan fucosylation in Arabidopsis thaliana suggest a water–mediated mechanism. *Plant J* **2017,** *91* (6), 931–949.
102. Zou, Y.; Garcia–Borras, M.; Tang, M. C.; Hirayama, Y.; Li, D. H.; Li, L.; Watanabe, K.; Houk, K. N.; Tang, Y., Enzyme–catalyzed cationic epoxide rearrangements in quinolone alkaloid biosynthesis. *Nat Chem Biol* **2017,** *13* (3), 325–332.
103. UlugI˙ M. Unligil , S. Z.; Sivashankary Yuwaraj , M. S., HarrySchachter2,3andJamesM. Rini, X–ray crystal structure of rabbit N–acetylglucosaminyltransferase I: catalytic mechanismandanewproteinsuperfamily. **2000**.
104. Schimpl, M.; Zheng, X.; Borodkin, V. S.; Blair, D. E.; Ferenbach, A. T.; Schuttelkopf, A. W.; Navratilova, I.; Aristotelous, T.; Albarbarawi, O.; Robinson, D. A.; Macnaughtan, M. A.; van Aalten, D. M., O–GlcNAc transferase invokes nucleotide sugar pyrophosphate participation in catalysis. *Nat Chem Biol* **2012,** *8* (12), 969–74.
105. Soya, N.; Fang, Y.; Palcic, M. M.; Klassen, J. S., Trapping and characterization of covalent intermediates of mutant retaining glycosyltransferases. *Glycobiology* **2011,** *21* (5), 547–52.
106. Cote, J. M.; Taylor, E. A., The Glycosyltransferases of LPS Core: A Review of Four Heptosyltransferase Enzymes in Context. *Int J Mol Sci* **2017,** *18* (11).
107. Errey, J. C.; Lee, S. S.; Gibson, R. P.; Martinez Fleites, C.; Barry, C. S.; Jung, P. M.; OSullivan, A. C.; Davis, B. G.; Davies, G. J., Mechanistic insight into enzymatic glycosyl transfer with retention of configuration through analysis of glycomimetic inhibitors. *Angew Chem Int Ed Engl* **2010,** *49* (7), 1234–7.
108. Gomez, H.; Polyak, I.; Thiel, W.; Lluch, J. M.; Masgrau, L., Retaining glycosyltransferase mechanism studied by QM/MM methods: lipopolysaccharyl–alpha–1,4–galactosyltransferase C transfers alpha–galactose via an oxocarbenium ion–like transition state. *J Am Chem Soc* **2012,** *134* (10), 4743–52.
109. Brockhausen, I., Crossroads between Bacterial and Mammalian Glycosyltransferases. *Front Immunol*

2014, *5*, 492.
110. Leemhuis, H.; Dijkhuizen, L., Engineering of Hydrolysis Reaction Specificity in the Transglycosylase Cyclodextrin Glycosyltransferase. *Biocatalysis and Biotransformation* **2010,** *21* (4–5), 261–270.
111. Lee, H. S.; Thorson, J. S., Development of a universal glycosyltransferase assay amenable to high–throughput formats. *Anal Biochem* **2011,** *418* (1), 85–8.
112. McGraphery, K.; Schwab, W., Comparative Analysis of High–Throughput Assays of Family–1 Plant Glycosyltransferases. *Int J Mol Sci* **2020,** *21* (6).
113. Alcantara, A. R.; Dominguez de Maria, P.; Littlechild, J. A.; Schurmann, M.; Sheldon, R. A.; Wohlgemuth, R., Biocatalysis as Key to Sustainable Industrial Chemistry. *ChemSusChem* **2022,** *15* (9), e202102709.
114. Cao, Y.; Li, X.; Ge, J., Enzyme Catalyst Engineering toward the Integration of Biocatalysis and Chemocatalysis. *Trends Biotechnol* **2021,** *39* (11), 1173–1183.
115. Hoffmann, T. D.; Kurze, E.; Liao, J.; Hoffmann, T.; Song, C.; Schwab, W., Genome–wide identification of UDP–glycosyltransferases in the tea plant (Camellia sinensis) and their biochemical and physiological functions. *Front Plant Sci* **2023,** *14*, 1191625.
116. Ma, W.; Zhao, L.; Ma, Y.; Li, Y.; Qin, S.; He, B., Oriented efficient biosynthesis of rare ginsenoside Rh2 from PPD by compiling UGT–Yjic mutant with sucrose synthase. *Int J Biol Macromol* **2020,** *146*, 853–859.
117. He, J. B.; Zhao, P.; Hu, Z. M.; Liu, S.; Kuang, Y.; Zhang, M.; Li, B.; Yun, C. H.; Qiao, X.; Ye, M., Molecular and Structural Characterization of a Promiscuous C–Glycosyltransferase from Trollius chinensis. *Angew Chem Int Ed Engl* **2019,** *58* (33), 11513–11520.
118. Zhang, H.; Zhu, F.; Yang, T.; Ding, L.; Zhou, M.; Li, J.; Haslam, S. M.; Dell, A.; Erlandsen, H.; Wu, H., The highly conserved domain of unknown function 1792 has a distinct glycosyltransferase fold. *Nat Commun* **2014,** *5*, 4339.
119. Albesa–Jove, D.; Guerin, M. E., The conformational plasticity of glycosyltransferases. *Curr Opin Struct Biol* **2016,** *40*, 23–32.
120. Offen, W.; Martinez–Fleites, C.; Yang, M.; Kiat–Lim, E.; Davis, B. G.; Tarling, C. A.; Ford, C. M.; Bowles, D. J.; Davies, G. J., Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification. *EMBO J* **2006,** *25* (6), 1396–405.
121. Sun, G.; Putkaradze, N.; Bohnacker, S.; Jonczyk, R.; Fida, T.; Hoffmann, T.; Bernhardt, R.; Hartl, K.; Schwab, W., Six Uridine–Diphosphate Glycosyltransferases Catalyze the Glycosylation of Bioactive C(13)–Apocarotenols. *Plant Physiol* **2020,** *184* (4), 1744–1761.
122. Sun, G.; Strebl, M.; Merz, M.; Blamberg, R.; Huang, F. C.; McGraphery, K.; Hoffmann, T.; Schwab, W., Glucosylation of the phytoalexin N–feruloyl tyramine modulates the levels of pathogen–responsive metabolites in Nicotiana benthamiana. *Plant J* **2019,** *100* (1), 20–37.
123. Bourdenx, B.; Bernard, A.; Domergue, F.; Pascal, S.; Leger, A.; Roby, D.; Pervent, M.; Vile, D.; Haslam, R. P.; Napier, J. A.; Lessire, R.; Joubes, J., Overexpression of Arabidopsis ECERIFERUM1 promotes wax very–long–chain alkane biosynthesis and influences plant response to biotic and abiotic stresses. *Plant Physiol* **2011,** *156* (1), 29–45.
124. Reed, M. C.; Lieb, A.; Nijhout, H. F., The biological significance of substrate inhibition: a mechanism with diverse functions. *Bioessays* **2010,** *32* (5), 422–9.
125. Liao, J.; Sun, G.; Kurze, E.; Steinchen, W.; Hoffmann, T. D.; Song, C.; Zou, Z.; Hoffmann, T.; Schwab, W. G., Subfunctionalization of a monolignol to a phytoalexin glucosyltransferase is accompanied by

substrate inhibition. *Plant Commun* **2022**, 100506.
126. Song, C.; Hartl, K.; McGraphery, K.; Hoffmann, T.; Schwab, W., Attractive but Toxic: Emerging Roles of Glycosidically Bound Volatiles and Glycosyltransferases Involved in Their Formation. *Mol Plant* **2018,** *11* (10), 1225–1236.
127. Stafford, H. A., Flavonid Evolution: An Enzymic Approach. *Plant physiol.* **1991**.
128. Jensen, K.; Moller, B. L., Plant NADPH–cytochrome P450 oxidoreductases. *Phytochemistry* **2010,** *71* (2–3), 132–41.
129. Zagrobelny, M.; Bak, S.; Rasmussen, A. V.; Jorgensen, B.; Naumann, C. M.; Lindberg Moller, B., Cyanogenic glucosides and plant–insect interactions. *Phytochemistry* **2004,** *65* (3), 293–306.
130. Ohgami, S.; Ono, E.; Toyonaga, H.; Watanabe, N.; Ohnishi, T., Identification and characterization of Camellia sinensis glucosyltransferase, UGT73A17: A possible role in flavonol glucosylation. *Plant Biotechnology* **2014,** *31* (5), 573–578.
131. Hou, B.; Lim, E. K.; Higgins, G. S.; Bowles, D. J., N–glucosylation of cytokinins by glycosyltransferases of Arabidopsis thaliana. *J Biol Chem* **2004,** *279* (46), 47822–32.
132. Bowles, D.; Lim, E. K.; Poppenberger, B.; Vaistij, F. E., Glycosyltransferases of lipophilic small molecules. *Annu Rev Plant Biol* **2006,** *57*, 567–97.
133. Jackson, R. G.; Lim, E. K.; Li, Y.; Kowalczyk, M.; Sandberg, G.; Hoggett, J.; Ashford, D. A.; Bowles, D. J., Identification and biochemical characterization of an Arabidopsis indole–3–acetic acid glucosyltransferase. *J Biol Chem* **2001,** *276* (6), 4350–6.
134. Ali, S.; Hayat, K.; Iqbal, A.; Xie, L., Implications of Abscisic Acid in the Drought Stress Tolerance of Plants. *Agronomy* **2020,** *10* (9).
135. Baldacci–Cresp, F.; Le Roy, J.; Huss, B.; Lion, C.; Creach, A.; Spriet, C.; Duponchel, L.; Biot, C.; Baucher, M.; Hawkins, S.; Neutelings, G., UDP–GLYCOSYLTRANSFERASE 72E3 Plays a Role in Lignification of Secondary Cell Walls in Arabidopsis. *Int J Mol Sci* **2020,** *21* (17).
136. Lanot, A.; Hodge, D.; Jackson, R. G.; George, G. L.; Elias, L.; Lim, E. K.; Vaistij, F. E.; Bowles, D. J., The glucosyltransferase UGT72E2 is responsible for monolignol 4–O–glucoside production in Arabidopsis thaliana. *Plant J* **2006,** *48* (2), 286–95.
137. Guan, M.; Li, C.; Shan, X.; Chen, F.; Wang, S.; Dixon, R. A.; Zhao, Q., Dual Mechanisms of Coniferyl Alcohol in Phenylpropanoid Pathway Regulation. *Front Plant Sci* **2022,** *13*, 896540.
138. Beyer, S. F.; Beesley, A.; Rohmann, P. F. W.; Schultheiss, H.; Conrath, U.; Langenbach, C. J. G., The Arabidopsis non–host defence–associated coumarin scopoletin protects soybean from Asian soybean rust. *Plant J* **2019,** *99* (3), 397–413.
139. Grana, E.; Costas–Gil, A.; Longueira, S.; Celeiro, M.; Teijeira, M.; Reigosa, M. J.; Sanchez–Moreiras, A. M., Auxin–like effects of the natural coumarin scopoletin on Arabidopsis cell structure and morphology. *J Plant Physiol* **2017,** *218*, 45–55.
140. Liao, J.; Lederer, V.; Bardhi, A.; Zou, Z.; Hoffmann, T. D.; Sun, G.; Song, C.; Hoffmann, T.; Schwab, W., Acceptors and Effectors Alter Substrate Inhibition Kinetics of a Plant Glucosyltransferase NbUGT72AY1 and Its Mutants. *Int J Mol Sci* **2023,** *24* (11).
141. Fiser, A.; Sali, A., ModLoop: automated modeling of loops in protein structures. *Bioinformatics* **2003,** *19* (18), 2500–1.
142. V. K. VYAS, R. D. U., M. GHATE AND C. CHINTHA, Homology Modeling a Fast Tool for Drug Discovery: Current Perspectives. **2012**.
143. Danev, R.; Yanagisawa, H.; Kikkawa, M., Cryo–Electron Microscopy Methodology: Current Aspects

and Future Directions. *Trends Biochem Sci* **2019,** *44* (10), 837–848.
144. Shi, Y., A glimpse of structural biology through X–ray crystallography. *Cell* **2014,** *159* (5), 995–1014.
145. Chang, A.; Singh, S.; Phillips, G. N., Jr.; Thorson, J. S., Glycosyltransferase structural biology and its role in the design of catalysts for glycosylation. *Curr Opin Biotechnol* **2011,** *22* (6), 800–8.
146. Bolam, D. N.; Roberts, S.; Proctor, M. R.; Turkenburg, J. P.; Dodson, E. J.; Martinez–Fleites, C.; Yang, M.; Davis, B. G.; Davies, G. J.; Gilbert, H. J., The crystal structure of two macrolide glycosyltransferases provides a blueprint for host cell antibiotic immunity. *Proc Natl Acad Sci U S A* **2007,** *104* (13), 5336–41.
147. Skinner, J. J.; Lim, W. K.; Bedard, S.; Black, B. E.; Englander, S. W., Protein hydrogen exchange: testing current models. *Protein Sci* **2012,** *21* (7), 987–95.
148. Huang, S.; Jia, A.; Song, W.; Hessler, G.; Meng, Y.; Sun, Y.; Xu, L.; Laessle, H.; Jirschitzka, J.; Ma, S.; Xiao, Y.; Yu, D.; Hou, J.; Liu, R.; Sun, H.; Liu, X.; Han, Z.; Chang, J.; Parker, J. E.; Chai, J., Identification and receptor mechanism of TIR–catalyzed small molecules in plant immunity. *Science* **2022,** *377* (6605), eabq3297.
149. Scheerer, D.; Adkar, B. V.; Bhattacharyya, S.; Levy, D.; Iljina, M.; Riven, I.; Dym, O.; Haran, G.; Shakhnovich, E. I., Allosteric communication between ligand binding domains modulates substrate inhibition in adenylate kinase. *Proc Natl Acad Sci U S A* **2023,** *120* (18), e2219855120.
150. Sun, G.; Liao, J.; Kurze, E.; Hoffmann, T. D.; Steinchen, W.; McGraphery, K.; Habegger, R.; Marek, L.; Catici, D. A. M.; Ludwig, C.; Jing, T.; Hoffmann, T.; Song, C.; Schwab, W., Apocarotenoids are allosteric effectors of a dimeric plant glycosyltransferase involved in defense and lignin formation. *New Phytol* **2023,** *238* (5), 2080–2098.
151. Sheikh, M. O.; Halmo, S. M.; Patel, S.; Middleton, D.; Takeuchi, H.; Schafer, C. M.; West, C. M.; Haltiwanger, R. S.; Avci, F. Y.; Moremen, K. W.; Wells, L., Rapid screening of sugar–nucleotide donor specificities of putative glycosyltransferases. *Glycobiology* **2017,** *27* (3), 206–212.
152. Levanova, N.; Mattheis, C.; Carson, D.; To, K. N.; Jank, T.; Frankel, G.; Aktories, K.; Schroeder, G. N., The Legionella effector LtpM is a new type of phosphoinositide–activated glucosyltransferase. *J Biol Chem* **2019,** *294* (8), 2862–2879.
153. Sindhuwinata, N.; Munoz, E.; Munoz, F. J.; Palcic, M. M.; Peters, H.; Peters, T., Binding of an acceptor substrate analog enhances the enzymatic activity of human blood group B galactosyltransferase. *Glycobiology* **2010,** *20* (6), 718–23.
154. Dessau, M. A.; Modis, Y., Protein crystallization for X–ray crystallography. *J Vis Exp* **2011,** (47).
155. Wang, G.; Zhang, Z. T.; Jiang, B.; Zhang, X.; Li, C.; Liu, M., Recent advances in protein NMR spectroscopy and their implications in protein therapeutics research. *Anal Bioanal Chem* **2014,** *406* (9–10), 2279–88.
156. Damont, A.; Legrand, A.; Cao, C.; Fenaille, F.; Tabet, J. C., Hydrogen/deuterium exchange mass spectrometry in the world of small molecules. *Mass Spectrom Rev* **2023,** *42* (4), 1300–1331.
157. Schopper, S.; Kahraman, A.; Leuenberger, P.; Feng, Y.; Piazza, I.; Muller, O.; Boersema, P. J.; Picotti, P., Measuring protein structural changes on a proteome–wide scale using limited proteolysis–coupled mass spectrometry. *Nat Protoc* **2017,** *12* (11), 2391–2410.
158. Galazzo, L.; Bordignon, E., Electron paramagnetic resonance spectroscopy in structural–dynamic studies of large protein complexes. *Prog Nucl Magn Reson Spectrosc* **2023,** *134–135*, 1–19.
159. Georgieva, E. R., Protein Conformational Dynamics upon Association with the Surfaces of Lipid Membranes and Engineered Nanoparticles: Insights from Electron Paramagnetic Resonance

Spectroscopy. *Molecules* **2020,** *25* (22).
160. Greenfield, N. J., Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc* **2006,** *1* (6), 2876–90.
161. Gotz, M.; Wortmann, P.; Schmid, S.; Hugel, T., A Multicolor Single–Molecule FRET Approach to Study Protein Dynamics and Interactions Simultaneously. *Methods Enzymol* **2016,** *581*, 487–516.
162. Lazim, R.; Suh, D.; Choi, S., Advances in Molecular Dynamics Simulations and Enhanced Sampling Methods for the Study of Protein Systems. *Int J Mol Sci* **2020,** *21* (17).
163. Yang, K. K.; Wu, Z.; Arnold, F. H., Machine–learning–guided directed evolution for protein engineering. *Nat Methods* **2019,** *16* (8), 687–694.
164. Keefe, D. S. W. Z. A. D., Random Mutagenesis by PCR. *Current Protocols in Molecular Biology* **2000**.
165. Ouyang, F.; Zhao, M., Enhanced catalytic efficiency of CotA–laccase by DNA shuffling. *Bioengineered* **2019,** *10* (1), 182–189.
166. Sadanand, S., EvolvR–ing to targeted mutagenesis. *Nat Biotechnol* **2018,** *36* (9), 819.
167. Kwon, S.; Lee, B.; Yoon, S., CASPER: context–aware scheme for paired–end reads from high–throughput amplicon sequencing. *BMC Bioinformatics* **2014,** *15 Suppl 9* (Suppl 9), S10.
168. Zheng, L.; Baumann, U.; Reymond, J. L., An efficient one–step site–directed and site–saturation mutagenesis protocol. *Nucleic Acids Res* **2004,** *32* (14), e115.
169. Sandstrom, A. G.; Engstrom, K.; Nyhlen, J.; Kasrayan, A.; Backvall, J. E., Directed evolution of Candida antarctica lipase A using an episomaly replicating yeast plasmid. *Protein Eng Des Sel* **2009,** *22* (7), 413–20.
170. Kille, S.; Acevedo–Rocha, C. G.; Parra, L. P.; Zhang, Z. G.; Opperman, D. J.; Reetz, M. T.; Acevedo, J. P., Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis. *ACS Synth Biol* **2013,** *2* (2), 83–92.

Supplementary Information

Supplementary Tables

Supplementary Table 1. Technology widely applied in protein structure investigation

Technology	Usages/Advantages	Reference
X-ray crystallography	Any size molecules Atomic resolution but crystallization may take years and damage protein structure	[154]
NMR spectroscopy	Small molecules Dissolved in water	[155]
Cryo-electron microscopoy	Larger molecules Frozen in native state and near-atomic resolution	[143]
Hydrogen/deuterium exchange mass spectrometry	Protein conformation and dynamics Protein-protein interactions, protein-small molecule interactions, and protein-RNA interactions	[156, 157]
Limited proteolysis-coupled mass spectrometry (LiP-MS)	The identification of protein structural changes directly in their complex biological context on a proteome-wide scale	[158, 159]
Electron paramagnetic resonance	Providing structural and dynamic information about proteins in conditions similar to those of their physiological environment	
Circular dichroism	Rapidly evaluating the secondary structure, folding and binding properties of proteins	[160]
Single-molecule FRET experiment (microsecond-millisecond time scales)	Study the conformational heterogeneity and the state-to-state transition dynamics of proteins on the sub-millisecond to second timescales	[161]
Molecular dynamic simulations	A computer simulation method for analysing the physical movements of atoms and molecules	[162]
Machine learning	Molecular dynamic protein structure prediction	[163]

Supplementary Table 2. Well-known website about enzymes

Website	Application
Aggrescan3D (http://biocomp.chem.uw.edu.pI/A3D2)	Prediction of aggregation propensity in protein structures and rational design of protein solubility
DeepSoluE (http://lab.malab.cn/~wangchao/softs/DeepSoluE/)	Protein solubility prediction
NetSoIP (https://services.healthtech.dtu.dk/service.php?NetSolP)	Prediction of solubility and usability of proteins expressed in *E. coil*
Protein-Sol (https://protein-sol.manchester.ac.uk/)	Prediction of protein solubility
Scaneer (https://sbi.postech.ac.kr/w/SCANEER/)	Sequence co-evolutionary analysis to control the efficiency of enzyme reactions
FuncLib (http//: FuncLib.weizmann.ac.il)	Design and rank epistatic multipoint mutants at enzyme active sites using phylogenetic analysis and Rosetta atomistic design calculations
mCSM-lig (http://biosig.unimelb.edu.au/mcsm_lig/prediction)	Prediction of protein-ligand affinity change upon mutation
FireProt (https://loschmidt.chemi.muni.cz/fireprot/)	Computational design of stable proteins
iStable 2.0 (http://ncblab.nchu.edu.tw/iStable2/)	Prediction of protein stability by single mutation.
PremPS (https://lilab.jysw.suda.edu.cn/research/PremPS/)	Prediction of the effects of mutations on protein stability
DeepDDG (http://protein.org.cn/ddg.html)	Prediction of the stability change of protein point mutations using neural network

Supplementary Table 3. Enzyme engineering methods

Mutant methods	Special features	REFERENCE
Ep PCR (Erro-prone pcr)	Introducing random mutations into a defined long segment of DNA	[164]
DNA shuffling	Random recombination method to generate mutant genes, relying on the similarity of the DNA sequences	[165]
OrthoRep	An orthogonal DNA polymerase-plasmid pair in yeast that stably mutates around 1000,000-fold faster than the host genome *in vivo*, exceeding the error threshold of genomic replication that causes single-generation extinction.	[6]
EvolvR	Semi-random mutations in a small region downstream of any genomic or plasmid site that can be targeted by CRISPR-Cas9	[166]
CasPER	A genetic engineering technique in molecular biology by which the genomes of living organisms may be modified	[167]
Site-saturation mutagenesis	A random mutagenesis technique with a single codon or set of codons is substituted with all possible amino acids at the position	[168]
Combinatorial active-site saturation test (CAST)	A method based on semi-rational selection of two or more amino acids surrounding the catalytic site and simultaneous randomisation of these residue	[169]
Iterative saturation mutagenesis (ISM)	Performing iterative cycles of saturation mutagenesis at rationally chosen sites in an enzyme with one to three composed amino acid positions according to a Cartesian view of the protein structure	[85]
22c-trick	A method for saturation mutagenesis with a special mixture of three primers for creating a degeneracy of 22 unique codons coding for the 20 canonical amino acids	[170]