

UCL FACULTY OF LIFE SCIENCE

LONDON



UCL

Structure-based Enzyme Engineering of Glycosyltransferases

Yixi Zhang

Thesis submitted in accordance with the requirements of University
College London for the degree of Doctor of Philosophy

Supervisor

Dr Min Yang

Prof Simon Gibbons

Declaration

I, Yixi Zhang confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signature:

Date: 13/09/2018

Abstract

Background: Plant UDP-dependent glycosyltransferases (UGTs) play important roles in biology via the glycosylation of secondary metabolites. The future prospects of UGTs look promising, for example, it may serve as promising path for progress in expanding drug targets and synthesising glycan-based drug with enhanced bioactivity.

Nevertheless, the current poor understanding of UGTs at molecular level (e.g. kinetic and structure) has led to a limited understanding of their biological roles and has also hampered their potential applications.

Aims: This project aims to 1) build up a mass spectrometry (MS) based approach to study families of UGTs including their substrate specificities, kinetic parameters and mechanisms of action (Chapter 3); 2) identify catalytic key amino acids (ckAAs) in the various UGTs (Chapter 4); and finally, 3) apply the methods above in the study of selected Rhamnosyltransferases (RhaTs) 78D1 and 89C1 (Chapter 5).

Methodology: A triple quadrupole MS (QQQ-MS) was used as this instrument required limited modification of substrates and provided direct monitoring of the glycosylated product. 'Full scan mode' gave the initial screening of any potential glycosylated product, and the 'product ion' mode provided additional confirmation of the formation of the glycosylated product. The 'multiple reaction monitoring (MRM)' mode quantified the products formation as a function of reaction time and provided kinetic data of the UGTs (Chapter 3).

The study of catalytic key amino acids (ckAAs) was based on the multiple sequence alignment (MSA) method via the AA sequence comparison with template UGTs that have known crystal structures. Subsequent site-directed mutagenesis (SDM) was used to substantiate/disprove the functional role of potential ckAAs: mutants (with potential ckAAs mutated) were checked by MS to find out whether the original activities were maintained and/or new activities were gained. A further activity comparison (k_{cat}/K_M) between the active mutant and the wild type (WT) could indicate the influence from a particular AA (Chapter 4).

Results and conclusions: 29 recombinant UGTs from groups B, D, F, H and L were examined. New donor activities (e.g. UDP-GlcNAc towards 73B4 and 78D2) were reported. Full kinetic studies of these WT UGTs indicated that they followed the Bi-Bi sequential mechanism. Whilst most followed the Bi-Bi random sequential mechanism, exceptions could be found such as that for 73B4 facilitating the UDP-GlcNAc reaction (Chapter 3).

Based on the interactions between the donor and template UGT (e.g. VvGT1, PDB 2C1Z), mutations of the potential ckAAs oriented towards the sugar (positions C2, C3, C4 and C6), phosphate and uridine were designed with alanine and an AA with a similar structural and chemical character. An activity comparison (k_{cat}/K_M) between the WT and active mutants indicated that most of the potential ckAAs within the PSPG motif inferred from MSA, were conserved and possibly followed a similar interaction pattern. However, exceptions could be found (e.g. 78D2 D380). Taking the results of both MSA and SDM together, the ckAAs in the active sites in each target UGT towards a specific donor were identified (Chapters 4 and 5).

Additionally, the study of Rhamnosyltransferases (RhaT) 78D1 and 89C1 showed that new activities were acquired by point mutagenesis: 78D1 N375Q acquired UDP-Glc and UDP-GlcNAc activities; and 89C1 H357Q acquired UDP-Glc activity (Chapter 5).

Impact Statement

It is almost redundant to state the significance of glycans in nature. Thus, the enzyme that is responsible for the synthesis of glycans, i.e. glycosyltransferases (GTs), attracts much attention. GTs are widely distributed in all kingdoms of life and play numerous biological roles *in vivo* (e.g. xenobiotic detoxification in human and defence in plant). They may be disease-associated. Additionally, GTs may provide alternative tool for the synthesis of glycan-based drugs. Thus, the understanding of GTs at the molecular level, through kinetic, catalytic mechanism and active site, is significant for the understanding of their biological roles and may further shed light on potential application such as drug discovery.

This project contributed the understanding of GTs, using plant UDP-dependent GTs (UGTs).

The study of characterisation of wild type UGTs will be helpful to understand the biological functions of UGTs. In addition, this research may shed light on glycan-based drug discovery, as plant UGTs may serve as potential tools in chemo-enzymatic synthesis of glycans with pharmaceutical effects.

The identification of catalytic key amino acids (ckAAs) of UGTs lay the foundation of enzyme catalytic mechanism and specificity of sugar donors and acceptors. It will further shed light on many applications, such as GT inhibitor and the genetic engineering of GTs.

Although this project focused on plant UGTs, it will inform subsequent exploitation of other GTs (e.g. animal GTs) as they share similarities that possibly result from the process of evolution (e.g. arms race and co-evolution).

Table of Contents

Abstract	3
Impact Statement	5
Acknowledgements	12
List of the frequent abbreviations	13
Chapter 1 Introduction	14
1.1 General principles of glycans and importance of GTs.....	16
1.1.1 Prevalence and structural diversity of glycans.....	16
1.1.2 Biological roles of glycans.....	17
1.1.3 Glycosyltransferases	18
1.2 Plant UGTs.....	21
1.2.1 Plant secondary metabolism and glycosylation therein	21
1.2.2 Substrate specificity of plant UGTs	22
1.2.3 Kinetic study of UGTs	25
1.2.3.1 MS based GT characterisation assays	27
1.2.4 Crystal structures and active sites of plant UGTs	29
1.2.4.1 Structure of plant UGTs.....	29
PSPG motif of plant UGTs	31
1.2.4.2 Catalytic key amino acids in plant UGTs	32
1.2.4.3 Genetic engineering of plant UGTs	34
1.3 General scheme of the project.....	37
Chapter 2 Materials and methods	39
2.1 Materials	40
Chemicals.....	40
Apparatus/Instrument.....	40
2.2 Methods.....	41
2.2.1.1 Plasmid transformation	42
2.2.1.2 Protein expression	42
2.2.1.3 Protein purification and storage	43

2.2.1.4 Protein verification and concentration test	44
SDS-PAGE	44
Bradford assay	45
2.2.2 MSA and SDM	46
2.2.2.1 MSA	46
2.2.2.2 SDM	48
WT-encoding plasmid extraction	48
Primer design	49
Exponential amplification (EA)	57
EA product verification – agarose gel	57
Treatment and enrichment	57
Transformation and mutant plasmid verification	58
2.2.3 MS-based enzyme activity test	58
2.2.3.1 Substrate specificity (GAR) screening	58
Sample preparation	58
‘Full scan’ mode to detect the product formation	64
‘Product ion’ mode to confirm the product formation	64
2.2.3.2 Full kinetics	65
MRM parameter optimisation	65
Selection of enzyme concentrations ([E])	66
Selection of the substrate concentration ([S]) – Ion suppression	67
Full UGT kinetics	68
Product inhibition assays of UGTs	69
2.2.3.3 Activity comparison of WT and mutants	69
2.2.4 Homology modelling and protein overlay	70
2.2.4.1 Homology modelling	70
2.2.4.2 Protein overlay	71
Chapter 3 <i>In vitro</i> characterisation of UGTs	72
3.1 Method development of MS-based GT activity assay	74
3.1.1 Qualitative analysis for enzyme activity	75
3.1.1.1 Sample preparation	75
Selection of buffer and its concentration	75
Temperature and reaction time	76

3.1.1.2 MS conditions	77
3.1.2 Quantitative analysis for enzyme activity	78
3.1.2.1 MS conditions (MRM parameters set up).....	78
3.1.2.2 LC condition	79
3.1.2.3 Calibration curve.....	80
3.1.2.4 Kinetic study of UGTs	81
Kinetic model.....	81
Kinetic mechanisms and parameter calculations	83
Kinetics conditions.....	83
Selection of enzyme concentration ([E])	84
Selection of substrate concentration ([S]).....	85
Determination of enzyme full kinetics.....	88
3.2 Substrate specificity GAR screening	91
3.2.1 Donor specificity	92
3.2.2 Acceptor specificity	96
3.2.2.1 Acceptor specificities – group D UGTs.....	96
3.2.2.2 Acceptor specificity screening – groups B, F, H and L UGTs	97
3.3 Multi-substrate enzyme kinetics	101
3.3.1 Full kinetic study – Group D UGTs.....	101
UDP-Glc activity in Group D UGTs	104
UDP-Glc and UDP-GlcNAc activity of 73B4	106
UDP-Glc and UDP-Gal activity of 73C5.....	107
3.3.2 Full kinetic study – Group B, F and H UGTs	108
UDP-Glc activity in group B, F and H	110
UDP-Glc and UDP-Gal activity of 78D2	111
3.3.3 Error analysis	112
3.4 Conclusions.....	116
Chapter 4 Identification of catalytic key amino acids (ckAAs) in UGTs.....	118
4.1 Method development	120
4.1.1 Template UGTs and MSA	122
4.1.1.1 Template UGTs.....	122
4.1.1.2 MSA	123

4.1.2 Mutants	125
4.1.2.1 Mutant plasmids	125
Exponential amplification (EA)	125
Treatment and enrichment	127
Transformation	128
Mutant plasmid verification	128
4.1.2.2 Production of mutant protein	129
4.1.3 Mutant activity	131
4.1.4 Determination of ckAAs in UGTs	132
4.2 Identification of ckAAs – sugar moiety	135
4.2.1 ckAAs towards C2 and C3 positions of the sugar moiety (the AA 44 within the PSPG motif)	135
4.2.1.1 MSA	135
4.2.1.2 SDM	136
Alanine scanning	136
Effect on Q/N swap and Q/H swap	143
4.2.2 ckAAs towards the C3 and C4 positions of the sugar moiety (the AA 43 within the PSPG motif)	147
4.2.2.1 MSA	148
4.2.2.2 SDM	149
Alanine scanning	149
Effect on D/E swap	152
4.2.3 ckAAs towards C6 position of sugar moiety	155
4.2.3.1 MSA	155
4.2.3.2 SDM	156
4.2.4 Cysteines in the PSPG motif – disulphide bridge	159
4.2.5 Summary –ckAAs in UGTs (sugar moiety).....	161
4.3 Identification of ckAAs – phosphate moiety	166
4.3.1 MSA	166
4.3.2 SDM	167
4.3.3 Summary – ckAAs in UGTs (phosphate moiety + sugar moiety)	169
4.4 Identification of ckAAs – uridine moiety	172
4.4.1 MSA	172

4.4.2 SDM.....	173
4.4.3 Summary – ckAAs in UGTs (uridine moiety + phosphate moiety + sugar moiety)	174
4.5 Homology model validation	177
4.6 Conclusions.....	180
Chapter 5 Rhamnosyltransferases 78D1 and 89C1	185
5.1 <i>In vitro</i> characterisation of 78D1 and 89C1.....	189
5.1.1 Donor specificity of 78D1 and 89C1	189
5.1.2 Kinetic study	190
5.1.3 Experiment optimisation.....	193
5.2 Identification of catalytic key amino acids (ckAAs) in 78D1 and 89C1	194
5.2.1 Identification of ckAAs – sugar moiety.....	194
5.2.1.1 ckAAs towards C2-C4 positions of the sugar moiety.....	195
UGT 78D1: ckAAs towards C2-C4-positions of the sugar moiety	196
UGT 89C1: ckAAs towards C2-C4-positions of sugar moiety	199
5.2.1.2 ckAAs towards C6 position of sugar moiety	199
5.2.2 Identification of ckAAs: the phosphate moiety	201
UGT 78D1: ckAAs towards phosphate moiety	201
UGT 89C1: ckAAs towards phosphate moiety	203
5.2.3 Identification of ckAAs: the uridine moiety	204
UGT 78D1: ckAAs towards the uridine moiety	205
UGT 89C1: ckAAs towards the uridine moiety	206
5.2.4 Homology model validation	208
5.3 Conclusions.....	210
Final Conclusions	213
Future work.....	215
Appendix.....	217
A.1 Bi-Bi enzymatic reaction kinetic equations	217
A.2 Confirmation of positive substrate specificity	218
A.3 MS of RhaTs	221

References.....224

Acknowledgements

This thesis would not have been possible without the help of so many people in different ways. I would like to express my immeasurable appreciation and deepest gratitude for the help and support.

The first and largest thank you is, of course, to my principal supervisor Dr Min Yang for his considerable expertise, advice and eternal encouragement and enthusiasm. I have furthermore to thank my co-supervisor, Professor Simon Gibbons, to give me valuable advice on the project. I would like to thank the Research Department of Pharmaceutical and Biological Chemistry for giving me permission to commence this thesis and to do the necessary research work.

The next major thanks go to our collaborators from Shandong University and York University, who provided the valuable materials. I also wish to express my sincere thanks to my colleagues in the study space and labs, who gave me the necessary support to overcome all the obstacles in my work.

In addition, I would like to express the deepest thanks to my friends in London. I feel so fortunate to encounter Xuexia Huang and to share so many occasions of happiness, sorrow and struggling. Without her, I cannot have had such a wonderful experience. I would like to thank Lin Luo who shared so many days with me since I came to London. Your lovely daughter is the first child in our generation. We all wish her health, beauty and happiness. I would like to show my gratitude to Zilan when I corrected my thesis in despair. You helped me to update the understanding of the world and the attitude towards study and career. Thanks for sharing your room, your bed and bedbug with me. Especially, I would like to thank Peixi Cong; your valuable support and care are two of the most important things I got in London.

Lastly, I dedicate this thesis to my parents – who give me the most and best love from the world.

List of the frequent abbreviations

AA: amino acid;
bAA: binding amino acid;
cAA: catalytic amino acid;
ckAA: catalytic key amino acid;
dTDP: deoxythymidine diphosphate;
dTDP-Rha: dTDP-Rhamnose;
GDP: guanosine diphosphate;
GDP-Fuc: GDP-Fucose;
GlcT: glucosyltransferase;
GT: glycosyltransferase;
KMP: kaempferol;
MRM: multiple reaction monitoring;
MS: mass spectrometry;
MSA: multiple sequence alignment;
MW: molecular weight;
PSPG: plant secondary product glycosyltransferase;
RhaT: rhamnosyltransferase;
SDM: site-directed mutagenesis;
UDP: uridine Diphosphate;
UDP-Gal: UDP-Galactose;
UDP-Glc: UDP-Glucose;
UDP-GalNAc: UDP-*N*-acetylgalactosamine;
UDP-GlcNAc: UDP-*N*-acetylglucosamine;
UDP-Rha: UDP-Rhamnose;
UGT: UDP-dependent glycosyltransferase.
WT: wild type.

Chapter 1 Introduction

Glycans are fundamental macromolecules on which all life depends. They participate in a vast array of biological processes, are prevalent on the cell surface (e.g. glycolipids and glycoproteins)¹, modulate cell adhesion (e.g. attachment of glycans to polysialic acid to regulate cell adhesion in neural cells²), and are specifically recognised by glycan-binding proteins (e.g. during pathogen evasion)³. In 2012, a report⁴ from the US National Research Council suggested that the study of glycoscience is highly significant for the study of life science and may bring advancements in many fields, such as human health.

As a significant toolkit for the development of glycoscience, glycosyltransferases (GTs) are involved in the biosynthesis of glycans by transferring sugars from a donor to an acceptor. The understanding of GTs at the molecular level, through kinetic, catalytic mechanism and active site, is significant for the understanding of their biological roles and may further shed light on potential application such as the synthesis of biologically relevant glycans.

This chapter firstly describes the general principles of glycans, and then followed by a more detailed introduction of plant UGTs. Finally, the general scheme of this project is described.

1.1 General principles of glycans and importance of GTs

Over the past decades, studies have shown that glycans functions pervade biology at all levels⁵ (e.g. the structural and modulatory properties⁶ and specific recognition⁶). More importantly, it has been shown that the aberrant glycosylation contributes to aberrant glycans, which underlines the aetiology of various diseases such as infectious diseases⁷, cancers⁸, diabetes⁹, and immune diseases¹⁰. These discoveries emphasise the need for studies of glycosylation, particularly of the GTs involved in diseases and those useful for practical applications (e.g. biomarkers¹¹).

This section briefly introduces important principles of glycans, including their prevalence, structural diversity, and biological functions, and the importance of GTs.

1.1.1 Prevalence and structural diversity of glycans

Glycans are ubiquitous in biology. No example of a living cell is not coated with a layer of glycans on their surfaces in 3 billion years of evolution (Figure 1.1A)⁴: glycolipids¹² (Figure 1.1B) and glycoproteins¹³ (Figure 1.1B) are frequently present in the outer leaflet of the eukaryotic plasma membrane, with their carbohydrate portions exposed on the cell surface¹². In addition, several types of glycans are found in cytoplasm¹⁴ [e.g. nucleocytoplasmic *O*-linked β -*N*-acetylglucosamine (*O*-GlcNAc), see in Figure 1.1B] and extracellular matrix¹⁵ (e.g. hyaluronic acid, see in Figure 1.1B).

Glycans exhibit great structural diversity (Figure 1.1B). This feature mainly stems from 1) the linkage between monosaccharides, i.e. a monosaccharide may link to different groups of the other (e.g. the core of N-glycan); 2) anomeric state, i.e. a monosaccharide can attach to another residue through either α or β glycosidic linkage at anomeric centre (e.g. the core of N-glycan); 3) the monosaccharide composition [e.g. glucose (Glc) or galactose (Gal)]; 4) the aglycone part (e.g. protein or lipid).

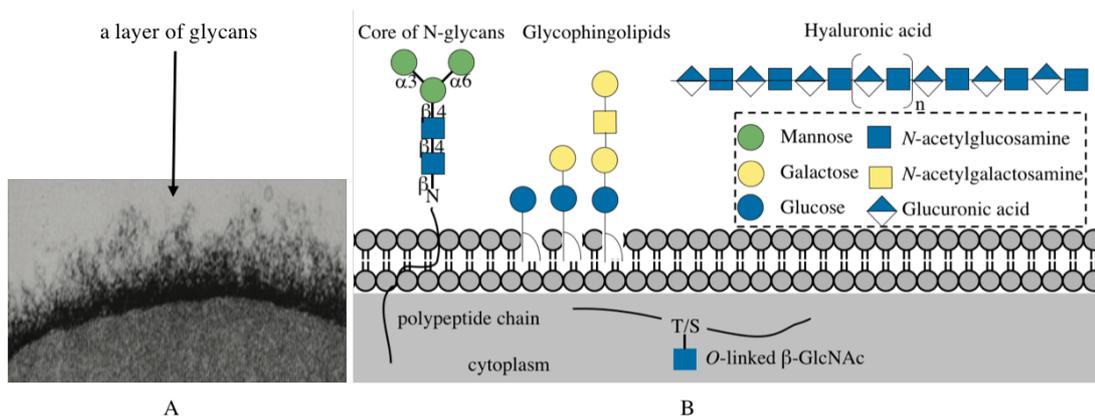


Figure 1.1: The prevalence and structure diversity of glycans. (A): Image of a red blood cell showing that a layer of glycans on the membrane surface¹⁶. (B): Some common classes of glycans in mammalian cells. Glycans can be found in various types of macromolecules with high level of structure diversity.

1.1.2 Biological roles of glycans

As fundamental molecules in nature, it is not surprising that glycans play a variety of biological roles. These biological roles range from being relatively subtle to crucial for the development, growth, functioning, or survival of organisms that synthesise them⁶. Given the enormous diversity of glycans, understanding all specific biological roles of glycans is therefore challenging and not yet evident. Three broad categories of glycan biological roles are recognised⁶: 1) the structural and modulatory properties of glycans; 2) the specific recognition of glycans; and 3) molecular mimicry of host glycans. There is a degree of overlap among these categories, such as the ‘structural property’ is required for the ‘specific recognition of glycans’ in some cases⁶.

Glycans have a variety of biological effects based on their primary structural properties, and/or the modulating functions of the proteins and lipids to which they are attached⁶. For example, neural cell surface glycans can physically mask sites on proteins that are susceptible to cleavage by extracellular matrix proteases². Neural cell surface glycans also act as modulators of cell adhesion². For example, they can attach to the long chains of polysialic acid, which prevents the homophilic interactions with neural cell adhesion molecules on opposing cells.

Another significant category of biological function is the specific recognition of glycans, which is accomplished by glycan-binding proteins (GBPs), including intrinsic GBPs (responsible for cell-cell interactions and glycan recognition by extracellular molecules and the same cell) and extrinsic GBPs (responsible for

recognition and communication with pathogenic microbial adhesins, agglutinins, or toxins)⁶ (Figure 1.2). As an extrinsic substance, a pathogen must specifically recognise and then evade the host cell. Before specific recognition, pathogens often decorate themselves with similar or identical glycan structures as those present on host cell surfaces³ as part of the molecular mimicry function of host glycans (the 3rd category of glycan biological function). Taken together, the pathogen can avoid the immune surveillance, recognise the host cells, and finally evade the immune system of the host. For example, Nontypeable *Human influenza* (NTHi) mimics the molecular signature of the host by incorporating sialic acid into the bacterial lipooligosaccharide¹⁷. Then, an NTHi protein (HMW1) interacts with sialylated lacto/neolactoglycolipids that located on the host cells to facilitate infection¹⁸.

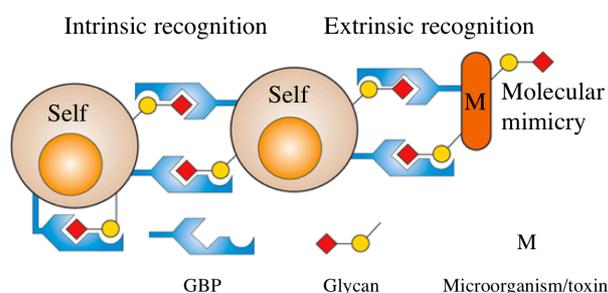


Figure 1.2: GBPs in recognising glycans⁶: intrinsic recognition between ‘self’ cells (left) and extrinsic recognition between ‘self’ cell and ‘M’ (right). Molecular mimicry of host glycans adds further complexity to potential roles.

1.1.3 Glycosyltransferases

Because of the importance of glycans, an increasing number of studies has been conducted to examine glycosyltransferases (GTs), as they determine the biosynthesis of glycans¹⁹ and play significant roles *in vivo*. For example, different blood group GTs can determine blood group by glycosylating H-antigen (the core precursor antigen that is present in blood group O)²⁰, such as GTB transfers galactose (Gal) to the terminal of H-antigen and determines blood group B²¹.

Recently, GTs have been recognized for their important roles in the aetiologies of many diseases¹⁵. Aberrant expression of GTs is associated with diseases such as cancer²². GnT-III is one of good exemplifications (Figure 1.3). It transfers GlcNAc to N-glycan via a β 1,4-linkage and finally establishes a ‘bisecting GlcNAc’ structure.

The ‘bisecting GlcNAc’ is not or only minimally elongated²³. GnT-III cannot be or is only minimally detected in the normal liver and hepatocytes, but is highly expressed in hepatoma cells²⁴. In hepatoma cells, GnT-III reduces cancer growth and aggressiveness by increasing cell adhesion through functional modification of some cell adhesion molecules (e.g. E-cadherin)²⁵, inhibiting some other GTs (e.g. GnT-V) that promote cell proliferation and growth²⁶, and inhibiting α 2,3-sialylation²⁷. Interestingly, GnT-III does not always act as a tumour-suppressing factor, but enhances cancer progression on specific cancer types (e.g. spleen cancer)²⁸ and specific environment around the tumour (e.g. diethylnitrosamine-induced liver cancer model)²⁹. Recently, a study of Alzheimer’s disease has revealed a broadened role for GnT-III in disease^{30,31}, of which GnT-III deficient mice showed less amyloid- β formation in the brain and improved short-term memory.

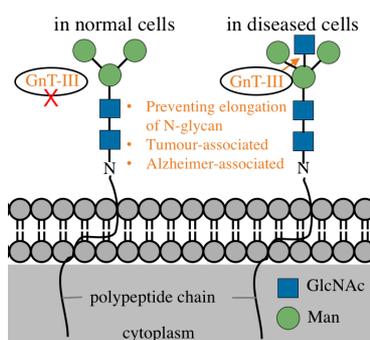


Figure 1.3: The role of GnT-III in normal cells (not expressed) and diseased cells (highly expressed). GnT-III produces ‘bisecting GlcNAc’ structure, which is associated with tumour and Alzheimer.

Therefore, the understanding of GTs in diseases will further contribute to explaining the mechanisms of disease, and more importantly, it enable the development practical applications such as early diagnosis³², glycan inhibitors and antagonists³³ and glycan-function modulators³⁴.

So far, we have briefly introduced some general principles of glycans and emphasised the significance of GTs in biology. The following section will focus on some knowledge of GTs, more specifically, on plant UDP-dependent GTs (UGTs). The study of plant UGTs is not only helpful to understand their biological functions in planta (e.g. regulating the bioactivity of secondary metabolites, hormones, and

toxins³⁵), but also benefits the understanding of other GTs (e.g. animal GTs) as they share similarities that possibly result from the process of evolution (e.g. arms race and co-evolution)³⁶. In addition, plant UGTs are soluble and readily to obtain, hence, they may serve as potential tools for further applications, such as the synthesis of glycans³⁷.

1.2 Plant UGTs

As at November 2017, there are more than 370,000 known and putative GT genes. They have been classified into 105 distinct families [based on amino acid (AA) sequence homology] by the Carbohydrate Active enZYmes database (CAZy)³⁸. Therein, family 1 contains the largest number of GT genes (more than 13,000).

As mentioned above, the study of plant UGTs may serve as promising path for GT study and for further applications. The present section will introduce the current understanding of plant UGTs, from the aspects of their function, catalytic behaviour, and structure. Furthermore, we will briefly discuss the techniques that used in our project.

1.2.1 Plant secondary metabolism and glycosylation therein

Plant secondary metabolites (SMs) are a diverse group of organic compounds such as flavonoids, terpenes, phenolic compounds, and alkaloids. They play significant roles in plant survival process such as by providing herbivore and pathogen protection^{39,40}, acting as endogenous regulators of plant growth⁴¹, and pollinator attractors⁴⁰. For example, flavonoids are responsible for the petal pigments in different *Papaver nudicaule* cultivars, which is helpful for pollinator attraction⁴².

In plants, SMs are frequently modified by glycosylation that catalysed by plant UGTs. Glycosylation can increase the solubility and mobility of these SMs. For example, glycosylated flavonoids are able to move from the endoplasmic reticulum to different cellular compartments and are secreted through the plasma membrane and the cell wall⁴³. Furthermore, plant UGTs may be involved in plant defence mechanisms. For example, tobacco salicylic acid- and pathogen-inducible UGTs (TOGTs) can mediate glucosylation that enables scopoletin accumulation in cells with *Tobacco mosaic virus* lesions, enhancing oxidative stress and weakening anti-viral resistance⁴⁴.

In addition to their many important biological roles, glycosylated SMs may have pharmaceutical effects for human health. For example, the cornel iridoid glycoside (CIG) treatment improves memory deficits in fimbria-fornix transection rats, by modulating the expression of related proteins (e.g. increasing NGF and decreasing Cyt

c)⁴⁵. However, obtaining glycosylated SMs can be challenging: their isolation from plants involves complex purification steps and is limited by the supply of plant materials³⁷. Traditional chemical synthesis requires complicated reaction controls in order to produce glycans with proper regio- and stereo- chemistries⁴⁶. Thus, studies on plant UGTs may suggest a potential approach to facilitate the synthesis of glycans⁴⁷, such as UGT94F2 is an iridoid-specific GT, which may be a potential biocatalyst in the synthesis of iridoid glycosides⁴⁸.

1.2.2 Substrate specificity of plant UGTs

The first question in UGT study is that a large number of UGTs' functions, *in vivo* or *in vitro*, are still unknown. For example, in the model organism *Arabidopsis thaliana* (*At*), 119 UGT genes have been identified, and the overall organisation and evolutionary relationships among individual members can be illustrated by phylogenetic tree (Figure 1.4). However, more than half of them have not been assigned with a function. In the 29 UGTs that will be investigated in our project (the UGTs in dotted rectangles, Figure 1.4), less than half of them are function-known in literature (the UGTs in black dotted rectangles, Figure 1.4).

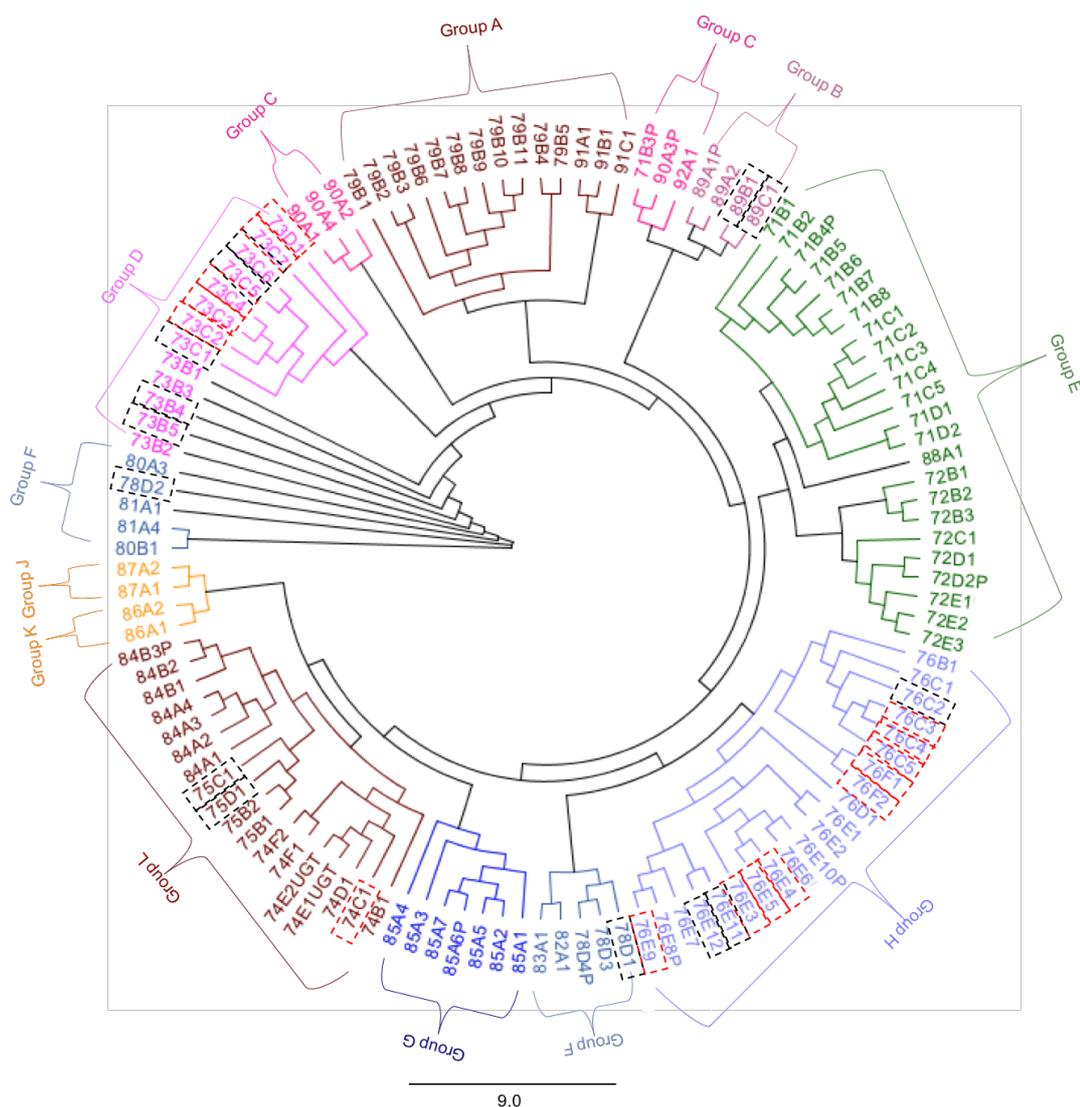


Figure 1.4: Unrooted phylogenetic tree (showing as rooted) of 119 *At* UGTs. Geneious software⁴⁹ with the neighbour-joining method was used. The bar 9.0 at the bottom means the scale of tree view. The UGTs that will be studied in this project are highlighted in dotted rectangles: the activities of the ones with black dotted rectangles have been identified in previous research, while the activities of the ones with red dotted rectangles have not been identified in previous research.

As literature shows, UGTs utilise UDP-dependent sugars. For example, UDP-glucose (UDP-Glc), UDP-galactose (UDP-Gal), UDP-rhamnose (UDP-Rha), and UDP-xylose (UDP-Xyl) can serve as donors for UGTs (Figure 1.5, left)⁵⁰. Generally, UGTs show high selectivity for a specific donor. Although some UGTs have broad donor ranges, they often display a preference for a specific donor. For example, *Vitis vinifera* VvGT1 exhibits activity towards a broad range of donors including UDP-Glc, dTDP-Glc, UDP-GlcNAc, UDP-Gal, UDP-mannose (UDP-Man), GDP-Glc, dTDP-Xyl, and UDP-Xyl. However, it shows a high donor preference towards UDP-Glc, with catalytic efficiency (k_{cat}/K_M) towards the other molecules of less than 6% of that

towards UDP-Glc (except for the catalytic efficiency with dTDP-Glc, which is 42% of that with UDP-Glc)³⁵. Another example is UGT88A1, which uses UDP-Glc, UDP-Gal, and UDP-glucuronic acid (acceptor: apigenin). The catalytic efficiency of UDP-Gal and UDP-GA are only 3% and 2% compared to that of UDP-Glc, respectively⁵¹.

Compared to donor, the available acceptor molecules are much more diverse (examples in Figure 1.5, right). Plant SMs are frequently used acceptors which include flavonoids⁵², terpenoids⁵³, and many other low-molecular-weight substrates⁵⁴⁻⁵⁶. UGTs can attach a glycosyl moiety to groups containing O (-OH and -COOH)⁵², N (-NH₂)⁵⁷, S (-SH)⁵⁸, and C (C-C in aromatic molecules)⁵⁹ atoms.

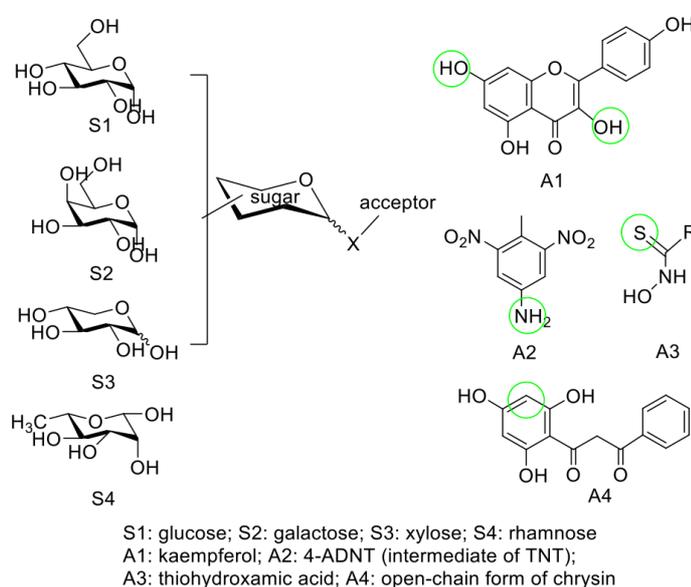


Figure 1.5: Examples of available sugars (left) and acceptors (right) of plant UGTs (the groups attaching sugar moiety are highlighted in green circle).

Since enzymes from the same family frequently exhibit different substrate specificities, precise functional predictions solely based only on the AA sequence are not reliable or accurate⁶⁰. The correlation between the degree of AA sequence identity and function is complex. The same substrate specificity may be observed in highly divergent UGTs, while different substrate specificity may be observed in highly homologous UGTs. For example, *Dorotheanthus bellidiformis* betanidin 6-O-glucosyltransferase (6-GT) has similar substrate specificity (betacyanin, anthocyanidins, and flavonols) as betanidin 5-O-glucosyltransferase (5-GT), although they only share 19% AA identity⁶¹ (Figure 1.6, top). On the contrary, *Streptomyces fradiae* UrdGT1b and UrdGT1c share

91% AA sequence identity, however, UrdGT1b uses dNDP-D-olivose, while UrdGT1c uses dNDP-L-rhodinose⁶² (Figure 1.6, bottom). Thus, in order to characterise the substrate specificities of UGTs, experimental *in vitro* or *in vivo* assays are strongly required.

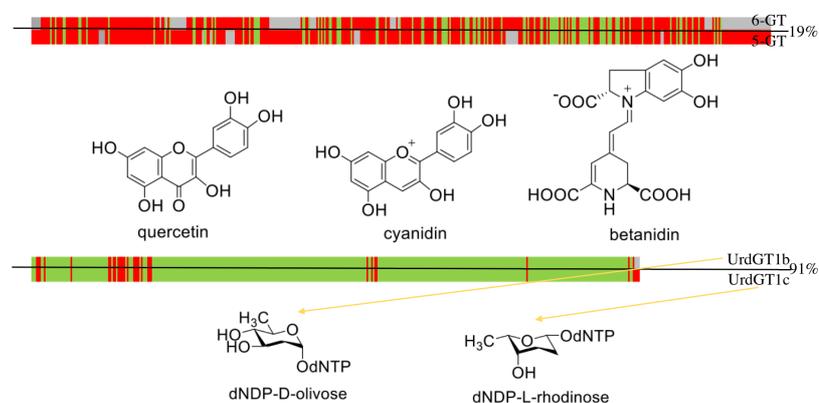


Figure 1.6: The correlation between the degree of AA sequence identity and function is complex: 6-GT and 5-GT share low AA identity but using the same acceptors (top). UrdGT1b and UrdGT1c share high AA identity but using different donors (bottom). The green region is where two UGTs have the same AA. The red region is where two UGTs have different AA. The grey region is where the gap located.

1.2.3 Kinetic study of UGTs

Enzyme kinetics focuses on the rate of enzyme-catalysed reactions. The understanding of *in vitro* UGT kinetics is significant in shedding light on their catalytic process *in vivo* and their potential applications, such as synthesis of glycans⁶³.

Cleland (1963) has classified enzyme-catalysed reactions and designated them using the terms Uni, Bi, Ter, and Quad, according to the number of substrates and products involved⁶⁴. The Uni-substrate enzyme kinetic model, a Michaelis-Menten model, provides a basis for complex enzyme kinetic models⁶⁵. The Michaelis-Menten model explains the correlation between the reaction velocity and substrate concentration ($[S]$), and describes enzyme behaviour using kinetic parameters, such as K_M and V_{max} . K_M has two meanings: 1) it is the concentration of substrate at which half of the active sites are occupied and reveals the $[S]$ required for significant catalysis to occur; 2) it is an indication of the affinity between the substrate (S) and the enzyme (E) under the condition that the rate constant for $ES \rightarrow E + S \gg$ rate constant for $ES \rightarrow E + P$. V_{max} reflects the turnover number of an enzyme, i.e. the amount of product formed in a unit

1.2.3.1 MS based GT characterisation assays

To evaluate GT-catalysed reactions, a broad range of assays have been developed to probe substrate specificity and kinetics. Thermodynamic or structural characteristics of the transient intermediate state of an enzyme during the reaction leads to a more direct and accurate kinetic understanding of the enzyme, however, the detection of intermediates is usually difficult because of their fleeting nature⁶⁶. Thus, monitoring of product formation over time is the primary method used to test GT substrate specificity and GT kinetics⁶⁷.

There are many GT assays to monitor the reaction product, generally divided into label (e.g. fluorescence-based method⁶⁸) and label-free methods. Label-free GT assays are attractive as the sample is relatively simple to prepare and can be used directly. In this research, we used the Agilent Triple Quadrupole LC/MS system to monitor the GT reaction, which uses LC to separate the products as traditional HPLC and uses MS to accurately quantify the product.

MS is an analytical technique that gives the mass-to-charge ratio (m/z) of charged particles. A sample is transformed into ions via ionisation [e.g. electrospray ionisation (ESI)]. Subsequently, ions travel through an MS analyser (e.g. quadrupole) and are separated based on their m/z ratios. The ion signals can be quantified in the MS detector, and are represented as a spectrum of ion abundance (y-axis) vs. m/z (x-axis)⁶⁹.

A quadrupole is an MS analyser that consists of four rods with direct current (DC) and radio frequency (RF) voltages applied⁷⁰. Ions with specific m/z values can pass through the quadrupole to reach the detector only when their corresponding fixed DC and RF are applied. Thus, a range of DC and RF voltage settings is used to scan the ions in a sample, while a fixed DC and RF voltage setting monitors specific ions in the MS. A triple quadrupole MS ($Q_1q_2Q_3$) is composed of three quadrupoles. The arrangement of $Q_1q_2Q_3$ allows for different analysis modes: full scan, product ion scan, precursor ion scan, neutral loss scan, and multiple reaction monitoring scan (MRM) modes (Table 1.1)⁷⁰.

Table 1.1: The arrangement of Q₁q₂Q₃ in different modes

	Q ₁	q ₂	Q ₃
Full scan	Scan	CID off	Scan
Product ion scan	Fixed	CID on	Scan
Precursor scan	Scan	CID on	Fixed
Neutral loss scan	Scan at $m/z=m_{\text{product}}$	CID on	Scan at $m/z=m_{\text{product}}-m_{\text{neutral molecule}}$
MRM scan	Scan at $m/z=m_{\text{precursor}}$	CID on	Scan at $m/z=m_{\text{product}}$

To examine the GT substrate specificity, the ‘full scan’ mode is used for the initial screening to determine whether a potential glycosylated product is formed in the reaction containing the tested substrates and GT. Once the potential glycosylated product ion (with an appropriate MS modification, e.g. $[M+H]^+$ or $[M-H]^-$ in positive-ion or negative-ion modes, respectively) is detected in ‘full scan’ mode, a subsequent ‘product ion’ mode is used to fragment the potential glycosylated product to confirm substrate utilisation by the enzyme by verifying aglycone ion formation after fragmentation. However, collision-induced dissociation (CID) in q₂ may result in the heterolytic or homolytic cleavage of the glycosidic bond, resulting in 1 and/or 2 Da mass shifts. For example, previous studies have shown that flavonoid glycosides yield aglycone fragments with 1 and/or 2 Da mass shifts from the parental deprotonated molecules in the negative-ion mode (Figure 1.8)⁷¹.

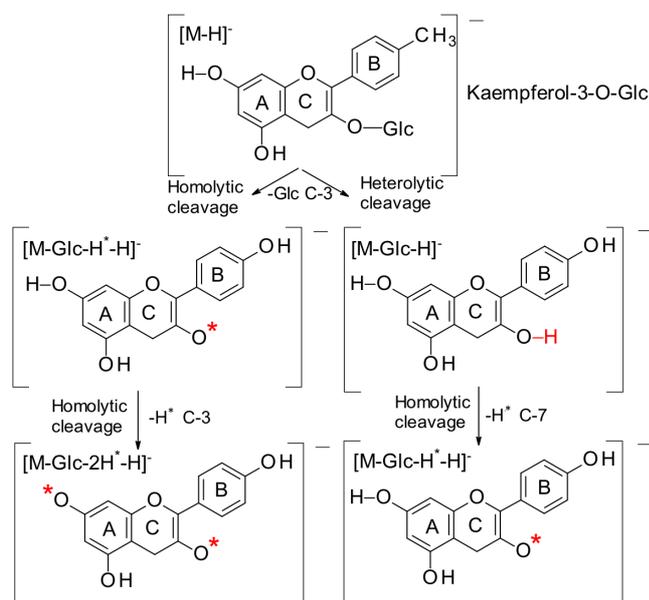


Figure 1.8: Proposed fragmentation pathway of KMP-3-O-Glc by homolytic and heterolytic cleavage.

In addition to qualitative analysis of product formation, Q₁q₂Q₃ can quantify the formation of glycosylated product thus to obtain GT kinetic data via the MRM mode.

The MRM mode functions as a double mass filter, with the collision energy applied in q2. For example, glycosylated product can be fragmented into aglycone in q2. To quantify glycosylated product, it would be first selected by Q1 and dissociated into aglycone in q2. Aglycone would then be selected for detection by Q3. Thus, the ‘glycosylated product → aglycone’ transition can be monitored for the quantification of glycosylated product, which dramatically reduces the noise and increases selectivity of the method, as has been widely acknowledged⁷². Typically, the chance of an isobaric interference at the same exact mass as the fragmentation ion is remote. Compared to the ‘product ion’ mode, MRM allows quantification of the glycosylated product instead of qualitative analysis. This method has been widely used in previous GT studies. For example, the kinetic behaviour of PgUGT74AE2 were obtained by monitoring its ginsenoside products, such as Rh₂ (transition at m/z 645.3 → 23.2) and F₂ (transition at m/z 807.5 → 627.5), using the MRM mode⁷³. The kinetics of UGT71C1 (transferring Glc to lariciresinol) were evaluated by monitoring the transition of the glycosylated product at m/z 521.1 (MW of lariciresinol glucoside) → 359.2 (MW of aglycone fragment) in the MRM mode⁷⁴.

1.2.4 Crystal structures and active sites of plant UGTs

In 1994, Vrieland research group unveiled the discovery of GT structure by reporting the first crystal structure of GT (bacteriophage T4-GlcT)⁷⁵. To date, 28 crystal structures have been solved for GT1 family proteins: 19 are from bacteria, one from human, two from yeasts, and seven from plants. The solved crystal structures provide important information regarding GT structures and interactions with substrate.

1.2.4.1 Structure of plant UGTs

The crystal structures of seven plant UGTs have been solved (Table 1.2): *Medicago truncatula* (*Mt*) UGT 71G1⁷⁶, 85H2⁷⁷, and 78G1⁷⁸; *Vitis vinifera* VvGT1³⁵; *Clitoria ternatea* UGT 78K6 (*Ct*3GT-A)^{79,80}; *At* UGT 72B1⁸¹; and *Oryza sativa* Os79⁸².

Table 1.2: Plant UGTs with known crystal structure

Plant UGT	Function	PDB	In complex with
UGT 71G1 ⁷⁶	Saponin biosynthesis	2ACV	UDP
		2ACW	UPG
UGT 78G1 ⁷⁸	Flavonoid 3GlcT	3HBJ	UDP
		3HBF	UDP, MYC
UGT 85H2 ⁷⁷	Flavonoid 3GlcT	2PQ6	-
VvGT1 ³⁵	Anthocyanin 3GlcT	2C1X	UDP
		2C1Z	U2F, KMP
		2C9Z	UDP, QUE
UGT 78K6 (<i>Ct</i> 3GT-A) ^{79,80}	Anthocyanin 3GlcT	3WC4	GOL, ACT
UGT 72B1 ⁸¹	<i>O</i> -GT/ <i>N</i> -GT	4REL	KMP, GOL, ACT
		4REM	DLM, GOL
		4REN	P5M, GOL
		2VCE	U2F, TC7, EDO
		2VCH	UDP, EDO
		2VG8	UDP, TRS, EDO
<i>Os</i> 79 ⁸²	Flavonoid GlcT	5TMB	UDP
		5TMD	U2F, 7ED
		5TME	UDP

All solved plant UGTs adopt a GT-B fold that the two Rossmann-like domains (C- and N-terminal) face towards each other and are loose-associated (Figure 1.9, left)¹⁹. The C-terminal domain is usually structurally conserved and responsible for the nucleotide sugar-donor binding, while the N-terminal domain is more pronouncedly variable as it has evolved to accommodate a variety of acceptors⁸³. A cleft between the two domains is responsible for substrate binding. A PSPG motif, mainly responsible for donor binding, is conserved found and located in the cleft close to the C-terminal.

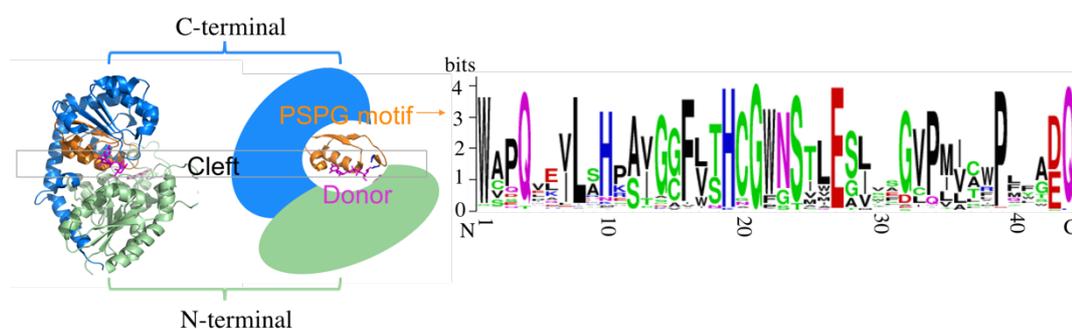


Figure 1.9: Representative structure of plant UGTs (*Vv*GT1, PDB: 2C1Z) (left) and AA constitution of 119 *At* UGT PSPG motifs (right). The overall structure of plant UGTs comprises N- and C-terminal domain. A cleft between the two domains is responsible for substrate binding. The conserved PSPG motif is located in the cleft close to C-terminal domain and binds the donor. The AA constitution of the PSPG motif shows its conservation: the size of letter depends on the degree of conservation of AAs. Colours of letters represent AAs with different chemical properties. Green: polar AAs; Blue: basic AAs; Red: acidic AAs; Black: hydrophobic AAs.

Plant UGT structures are highly similar, particularly in their C-terminal domains⁸⁴. For example, an overall structure comparison between *Vv*GT1 and 85H2³⁸ revealed 24% ‘structurally equivalent and identical residues’ and 44% ‘structurally equivalent and

similar residues'. Especially, 57% 'structurally equivalent and identical residues' and 58% 'structurally equivalent and similar residues' are in the C-terminal domain, indicating that the C-terminal domains of VvGT1 and 85H2 are more conserved than N-terminal domains.

Some flexible regions are present in different UGTs which typically vary in sequence, length, and conformation. The differences can be pronounced on loops N3, N5, and N5 α , which are allocated in the N-terminal domain. For example, 85H2 and 72B1 contain a short α -helix in the N5 loop, while two short stretches of β sheets are present in the corresponding loop of 71G1. Further, differences may also occur around the substrate binding cleft. For example, the linker region in 85H2 is longer than that in VvGT1 and 71G1, and contains several deletions and insertions. Specifically, insertions of additional residue are observed between C β 2 and C α 2, with a deletion between C β 6 and C α 6. The differences among these plant UGTs indicate that the active site for substrate binding is flexible and varies in each enzyme.

PSPG motif of plant UGTs

As mentioned above, a 44-AA conserved PSPG sequence is observed in plant UGTs⁸⁵. This motif is thought to have evolved from a common ancestor, and has been defined as a signature motif of plant UGTs that participate in the glycosylation of SMs⁸⁶. Figure 1.9 (right) shows the conserved PSPG motif in 119 *At* UGTs: 'W-2x-Q-3x-L-1x-H-1x-A/S-1x-G-G/C-F-L/V/W-T/S-H-C-G-W-1x-S/T-2x-E-4x-G-V/C/L-P/Q-4x-P-3x-D/E-Q'. Further, *in silico* motif diversity analysis using the MEME tool revealed that the PSPG motif is consistently present in all UGT sequences in the C-terminal domain⁸⁷. The presence of the PSPG motif has been widely accepted and applied in bioinformatics studies to identify new UGTs⁸⁸.

In addition to in AA sequence analysis, the PSPG motif is observed in crystal structures. This motif is located at the C-terminal of UGTs, where most donor-binding AAs are located. Figure 1.10 gives an example of the conserved manner by which the PSPG motif binds UDP-Glc in 71G1 and VvGT1. UDP-Glc is stacked in the PSPG motif and directly interacts with some AAs within the motif as follows: AAs 1, 11, and 34 directly bind the uridine portion of UDP-Glc; AAs 19, 23, and 24 interact with the phosphate portion of UDP-Glc; AAs 43 and 44 interact with the glucose moiety of

UDP-Glc. Although other AAs within the PSPG motif do not directly bind the donors, they play critical roles in stabilising the motif structure and appropriately stacking UDP-Glc. For example, 71G1 E356 (AA 18 within the PSPG motif) does not directly interact with the uridine moiety in the crystal structure; however, the E356P mutant exhibits no detectable UGT activity⁷⁶. This is likely because E356 is within the loop region of AAs 339–334 (WAPQVE), which acts as a hydrophobic platform for stacking of uridine and helps W339 (AA 1 within the PSPG motif) to form π - π interactions with the uracil ring. Thus, the E356P substitution disturbs the loop conformation and results in loss of activity, resulting in a loss of activity⁷⁶.

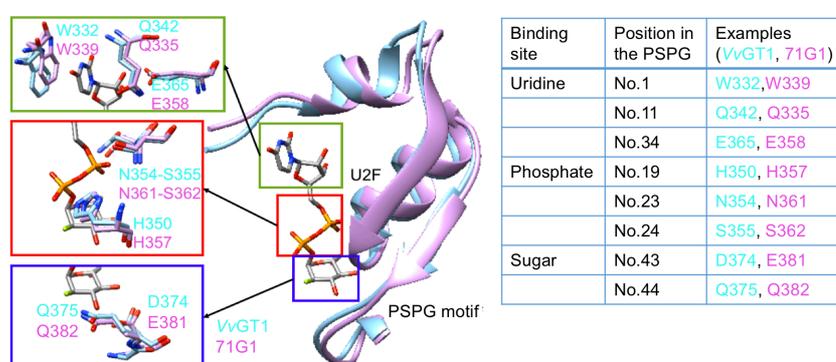


Figure 1.10: Structure of the PSPG motif in VvGT1 (light blue) and 71G1 (plum) and key binding AAs in donor binding.

1.2.4.2 Catalytic key amino acids in plant UGTs

In an enzymatic reaction, substrates are accommodated in the active site of the enzyme and undergo chemical reactions. The active site of an enzyme is composed of binding site and catalytic site. AAs in the binding site (binding AAs, bAAs) interact with the substrate to form an enzyme-substrate complex, such as by forming temporary bonds (e.g. hydrogen bonds⁸⁹) or by non-covalent interactions (e.g. van der Waals forces⁹⁰). AAs in the catalytic site (catalytic AAs, cAAs) catalyse a reaction such as by acting as a Brønsted base to activate the substrate for a nucleophilic attack⁹¹. Both bAAs and cAAs are key factors in enzymatic reactions and are often referred to as the catalytic key AAs (ckAAs)^{92,93}.

In enzymatic reactions catalysed by plant UGTs, donor binding is generally conserved. UGTs mainly interact with a nucleotide sugar donor via hydrogen-bonding and π -stacking⁹⁴, or, in some cases, via hydrophobic stacking interactions^{95,96}.

Figure 1.11 shows some key bAAs in the active site of UGTs, as exemplified by VvGT1 (tan)³⁵, 71G1 (pink)⁷⁶, and 72B1 (cyan)⁸¹. The binding patterns of plant UGT are generally conserved, but with some variations possible. The last two AAs (D/E-Q) within the PSPG motif are conserved for binding of Glc-O2, O3, and O4. However, the binding pattern at the C6 position of the sugar is not as conserved as that at the C2–C4 positions. In VvGT1 and 71G1, T141 and T143 bind Glc-O6, while 72B1 does not appear to have any bAA at the C6 position. The relatively lower conservation of the binding pattern at the C6 position of the sugar may be because bAAs binding C6 binding are at the N-terminal. As described above, the N-terminal domain of UGTs is mainly responsible for binding the acceptor and is more variable than the donor-binding C-terminal domain. As shown in the crystal structures of VvGT1 and 71G1, T141 and T143 are in the N-terminal domain but interact with Glc-O6 by crossing the entire N-terminal domain.

T280 (VvGT1 numbering) and AA 19 of the PSPG motif (e.g. VvGT1 H350) are conserved for phosphate binding. Another bAA (e.g. VvGT1 T19), located in the N-terminal domain, is not as highly conserved. In the three example UGTs, only VvGT1 T19 binds phosphate, while 71G1 and 72B1 do not. In the other crystals, 78G1 T25 and 78K6 S16 (the equivalent sites of VvGT1 T19) bind phosphate in the same manner as VvGT1 T19.

The binding pattern of uridine is most highly conserved, with AAs 1 (e.g. VvGT1 W332) and 34 (e.g. VvGT1 E358) interacting with the uracil ring and ribose by forming a π - π bond and hydrogen bond, respectively.

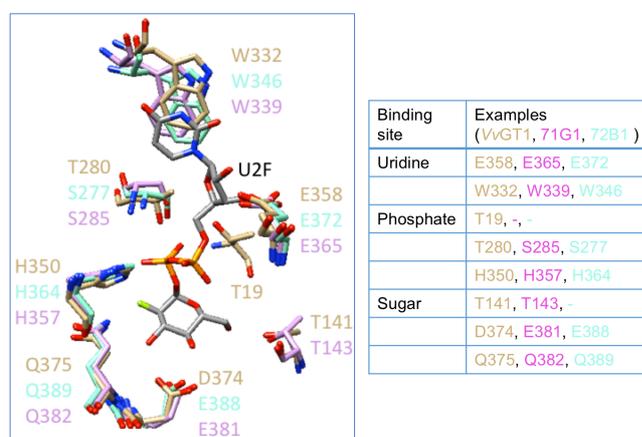


Figure 1.11: Some key bAAs in VvGT1 (tan), 71G1 (pink) and 72B1 (cyan).

Apart from some bAAs for the donor, interactions with donors can be determined by simple steric occlusion in some cases. This might be a possible explanation of how GTs are able to transfer sugars indiscriminately in some cases^{97,98}, and even transfer some non-carbohydrates (e.g. glycerol)⁹⁹.

1.2.4.3 Genetic engineering of plant UGTs

Although plant UGTs may have many commercial and economic applications, the stringent specificity or low performance of individual UGT is disadvantageous and limits glycoengineering⁴⁷. The function and performance of individual UGT have evolved to satisfy their natural roles, but not industrial biotechnology (human activity)¹⁰⁰. Thus, the manipulation of UGTs to create novel 'chimera' enzymes with desired or enhanced activity is an important future prospect for potential applications, such as to provide new tools for synthesising of glycans.

With the availability of protein structure, particularly the interaction between enzyme and substrate, a rational approach is used to propose substitutions resulting in enhanced or desired traits. This approach has the significant advantage of increasing the probability of success with less effort and time¹⁰⁰. For instance, F148 and Y202 in 71G1 are in close proximity and are located at one end of a channel shown to accommodate quercetin (Que) in a docking experiment based on the solved crystal structure of 71G1 (Figure 1.12, A)¹⁰¹. Subsequently, the following substitution to alter the regio-selectivity of Que glycosylation was proposed: reducing the size of either F148 or Y202 to expand the volume of the binding pocket where the C-ring of Que is located, so that the volume change of the binding pocket would allow the -OH at the C3 position to move closer to the C1 reaction centre on UDP-Glc. Thus, F148V and Y202A mutants were constructed and the regio-selectivity of Que glycosylation was successfully altered (from 3'-O-position of the B-ring to the C-O position of the C ring)¹⁰¹.

Another successful example is 85H2. Based on 85H2 crystal structure, the docking data shows that I305 is involved in the substrate binding pocket and resides in the proximity to the acceptor (Figure 1.12, B)⁷⁷. Structure analysis and docking experiment further indicated that changing I to T/S might result in a more favourable hydrogen bonding interaction and environment than in the original protein.

Subsequently, I305T was constructed and proved the prediction, as the UDP-Glc activity towards KMP or biochanin A increased 37- and 19- fold, respectively¹⁰².

However, manipulation of the active site cannot absolutely and solely rely on crystal structures, as the number of solved crystal structures is limited. Thus, a semi-rational approach is required¹⁰⁰. Semi-rational approaches combine some known knowledges, such as consensus sequence and function, and helps for predicting AA substitutions with the desired properties.

A well-known example is that of switching the donor specificity of AcGaT (UDP-Gal to UDP-Glc activity) by altering H374 to Q¹⁰³. The idea initially originated from multiple sequence alignment (MSA) analysis of GlcTs and GalTs, which revealed that Q and H were conserved as AA 44 within the PSPG motifs in GlcTs and GalTs, respectively (Figure 1.12, C). Thus, it was inferred that this site confers the sugar-specific recognition of Glc and Gal. Point mutagenesis experiments were subsequently designed, wherein AcGaT H374 was substituted with Q. Subsequently, the donor specificity of AcGaT was successfully switched from UDP-Gal to UDP-Glc.

More examples of active site manipulation using semi-rational approaches exist. Using the known crystal structure of VvGT1 as a template, the structure of VvGT5 was constructed and its substrates (UDP-glucuronic acid and KMP) were docked in the active site. The model predicted that an interaction existed between the guanidinium group of R140 and the carboxyl group of glucuronic acid. Further, at the equivalent site of VvGT1, W140 was present, which was oriented towards the bound U2F, while R140 was observed in VvGT5 instead (Figure 1.12, D). These observations suggested the possible important role of R140 in recognising UDP-glucuronic acid. To validate this prediction, point mutagenesis to introduce the R140W substitution was performed. The mutant not only exhibited a pronounced Glc transfer activity (the same activity as VvGT1), but also lost its original glucuronic acid transfer activity¹⁰⁴.

1.3 General scheme of the project

As introduced above, improved understanding of plant UGTs will help enhance the understanding of GTs and enable potential applications, such as probing of their *in vivo* biological activity, synthesis of valuable glycoconjugates, and manipulation of the active site for desired or enhanced activity. The present project aimed to broaden the understanding of plant UGTs, focusing on substrate specificity, kinetic behaviour, and identification of catalytic key amino acids (ckAAs).

In the current project, 29 recombinant *At* UGTs (Figure 1.4) from group B (89B1 and 89C1), D (73B4, 73B5, 73C1, 73C2, 73C3, 73C4, 73C5, 73C6, 73C7, and 73D1), F (78D1 and 78D2), H (76C2, 76C3, 76C4, 76C5, 76E3, 76E4, 76E5, 76E9, 76E11, 76E12, 76F1, and 76F2) and L (74C1, 75C1, and 75D1) were studied. Generally, groups D, H, and L are the three main groups and comprise large numbers of *At* UGTs (11%, 18%, and 15%, respectively, of 119 *At* UGTs). These three groups represent *At* UGTs with different functions (e.g. UDP-Glc and UDP-Rha activity). In addition, we aimed to identify the function for these function-unknown UGTs (the UGTs in red dotted rectangles, Figure 1.4). Although they have been predicted to use flavonoid from the existing genomic sequences in the UniProt database, no experimental data was found. Furthermore, we aimed to draw the complete picture of kinetic behaviours of tested UGTs if active.

In addition to the interest to probe the functions and kinetics of tested UGTs, the correlation between AA sequence and function interest us. This will provide insight useful for UGT genetic engineering. Thus, UGTs from the relatively small groups, groups B and F, were also chosen for analysis, as AA sequences of UGTs from these small groups are more closely related than those of large groups. Thus, the differences in their functions may have indicated the correlation between AA sequence and function (e.g. 78D1 and 78D2 share high AA identity (71%) but with different donor specificities).

This thesis will report the *in vitro* substrate specificity of the 29 UGTs listed above determined using a donor library (α -D-UDP-Glc, α -D-UDP-Gal, α -D-UDP-GlcNAc, α -D-UDP-GalNAc, β -L-UDP-Rha, β -L-dTDP-Rha, α -D-GDP-Glc, and β -L-GDP-Fuc)

and acceptor library (48 acceptors, including alcoholic, anthraquinone, coumarin, alkaloid, saponin, antibiotic, flavonoid, cyanogenic, plant hormone, and other phenolic compounds). Based on their substrate specificities, kinetics analyses with positive substrates were carried out (Chapter 3).

Apart from substrate specificity and kinetic studies, ckAAs in the active sites of UGTs have been identified. The UGTs used were the same as those in kinetic studies, whose substrate specificities have been identified. Studies of GlcTs 73B4, 73B5, 73C1, 73C5, 73C6, 76E11, 76E12, 78D2, and 89B1 are described in Chapter 4, while RhaTs 78D1 and 89C1 are described in Chapter 5. Rather than using a computational method, e.g. molecular modelling, the potential ckAAs were identified by multiple sequence alignment (MSA) with seven crystal structures of known plant UGTs as the templates, and examined experimentally for the initial rate comparisons (k_{cat}/K_M) of the WT and mutants (the potential ckAAs were changed to other AAs).

Chapter 2 Materials and methods

2.1 Materials

Chemicals

Glycerol, bromophenol blue, sodium dodecyl sulfate (SDS), acrylamide, ammonium persulfate (APS), tetramethylethylenediamine (TEMED), Luria-Bertani (LB) medium, agar, ampicillin, Tris, uridine 5'-diphosphoglucose disodium salt hydrate, uridine 5'-diphosphogalactose disodium salt hydrate, guanosine 5'-diphosphoglucose sodium salt, guanosine 5'-diphospho- β -L-fucose sodium salt, uridine 5'-diphospho-*N*-acetylglucosamine sodium salt, uridine 5'-diphospho-*N*-acetylgalactosamine sodium salt, kaempferol (KMP), KMP-3-glucoside, quercetin, MgCl₂, liquid chromatography/mass spectrometry (LC-MS) solvents (water and acetonitrile), Coomassie Brilliant Blue G, bovine serum albumin (BSA), quinine, atropine, anthranilic acid, and caffeic acid were from Sigma Aldrich (United Kingdom); deoxynivalenol, aloe-emodin, galantamine hydrobromide, and diacerein were purchased from Cambridge Bioscience; gossypol, theobromine, acetone cyanohydrin, and gibberellin A4 were purchased from Insight Biotechnology Ltd.; benzoic acid and hydroquinone were purchased from VWR International Ltd.; dTDP-rhamnose was purchased from Carbosynth; UDP-rhamnose was synthesised by Dr Min Yang's group. The GStrap fast flow column was from GE Healthcare; VivaSpin concentrators were from Generon; the QIAprep miniprep kit and XL-1 blue competent cells were obtained from Agilent; Instant Blue was ordered from BioLab. The Kinetex 5 μ m C18 100A New Column (50 \times 4.6 mm) was obtained from Phenomenex; UGT recombinant plasmids (*Arabidopsis* UGT-encoding sequences inserted in pGEX-2T vector) were kindly provided by the University of York and Shandong University.

Apparatus/Instrument

The 6400-series triple quadrupole LC-MS (QQQ-LC-MS) was from Agilent. Mini-PROTEAN® Tetra Cell systems for SDS electrophoresis and Mini-Sub® Cell GT cell for agarose electrophoresis were obtained from Bio-Rad. The cell shaker was from Thermo Fisher Scientific. The autoclave, cells incubator, and sonicator were supplied by Wolf Labs. The ultracentrifuge was obtained from Beckman.

2.2 Methods

The methods in this project are split into four sections, as follows. 1) protein production (Section 2.2.1): purified WT and protein mutants were obtained following the same procedures, including plasmid transformation, protein expression, and purification. 2) multiple sequence alignment (MSA) and site-directed mutagenesis (SDM) (Section 2.2.2): MSA was used to infer the potential catalytic key amino acids (ckAAs) interacting with the donor. The potential ckAAs were substituted using SDM. This method was used to identify the ckAAs in UGTs. 3) MS-based enzyme activity test (Section 2.2.3): by checking glycosylated product formation, both qualitatively and quantitatively, information on substrate specificity, kinetics, and activity comparison were obtained. Qualitative analysis was used to check substrate specificities of WT and mutant enzymes. Quantitative analysis was used for the kinetic study (Chapter 3) and activity comparisons of WT and protein mutants (Chapter 4). 4). Homology modelling and protein overlay (Section 2.2.4): construction of a model of target UGT and overlying it on template UGT allowed a convenient way of displaying and discussing the interactions between target UGTs and substrates.

Protein production

29 WT recombinant UGTs (UGT73B4, 73B5, 73C1, 73C2, 73C3, 73C4, 73C5, 73C6, 73C7, 73D1, 74C1, 75C1, 75D1, 76C2, 76C3, 76C4, 76C5, 76E3, 76E4, 76E5, 76E9, 76E11, 76E12, 76F1, 76F2, 78D1, 78D2, 89B1, and 89C1) and 106 mutants (listed in Table 2.1) were obtained using the following standard operating procedures: 1) plasmid transformation; 2) protein expression; and 3) protein purification and storage.

Table 2.1: List of 106 Mutants

WT	73B4	73B5	73C1	73C5	73C6	76E11	76E12	78D2
Mutants	R7A	G154A	G148A	G150A	G151A	T134A	T135A	T22A
	G127A	E394A	D393A	D397A	D397A	D373A	D375A	T286A
	S270A	E397D	D393E	D397E	D397E	D373E	D375E	W338A
	W331A	E397E	D393K	D397K	D397K	D373K	D375K	H356A
	H349A	E397N	Q394A	Q398A	Q398A	D373N	D375N	E264A
	E357A	Q398A	Q394E	Q398E	Q398E	Q374A	Q376A	D380A
	E373A	Q398N	Q394H	Q398H	Q398H	Q374E	Q376E	D380E
	E373D	Q398H	Q394N	Q398N	Q398N	Q374H	Q376H	D380K
	E373Q	Q398E				Q374N	Q376N	D380N
	E373K							Q381A
	Q374A							Q381E
	Q374N							Q381H
	Q374H							Q381N
	Q374E							C357A- C374A
WT	78D1	89B1	89C1					
Mutants	A23T	P145A	G20A					
	A141T	D389A	F126A					
	T280A	D389E	S250A					
	W332A	D389K	W314A					
	H350A	D389N	H332A					
	E358A	Q390A	E340A					
	D374A	Q390E	D356A					
	N375A	Q390H	H357A					
	N375Q	Q390N	H357Q					

2.2.1.1 Plasmid transformation

Plasmid encoding WT recombinant or mutant UGTs (1 μL) was gently mixed with XL-1 blue competent cells (20 μL). After placing the plasmid/cell mixture on ice for 30 min, it was treated by heat shock at 42°C for 30 s. The plasmid/cell mixture was then incubated on ice for another 5 min, mixed with LB media (950 μL) without any antibiotics, and incubated in a shaker at 160 rpm (37°C) for 1 h. Subsequently, the plasmid/cell mixture (50 μL) was spread onto an agar plate with 50 $\mu\text{g}/\text{mL}$ of ampicillin, and incubated overnight at 37°C. A single colony was then picked from the plate and incubated overnight in a liquid LB medium with 50 $\mu\text{g}/\text{mL}$ of ampicillin with shaking at 160 rpm (37°C), and, finally, stored in 80% glycerol.

2.2.1.2 Protein expression

The glycerol stocks of UGTs were inoculated into LB medium (10 mL) with 50 $\mu\text{g}/\text{mL}$ of ampicillin and incubated overnight in a shaker at 160 rpm (37°C). Then, the seed

culture was transferred into LB medium (500 mL) and grown under the same conditions until the culture OD₆₀₀ reached 0.4–0.6, i.e. the cells entered the logarithmic phase of growth. The system was cooled to 20°C for 1 h, and protein production was then induced by the addition of isopropyl-β-D-1-thiogalactopyranoside (IPTG) (final concentration: 0.1 mM) overnight.

2.2.1.3 Protein purification and storage

Cells were harvested by centrifuging at 9,000 rpm (4°C) for 30 min. The cell pellet was collected and re-suspended in Tris buffer (20 mM, pH 7.8, 10 mL). The suspended cells were broken by sonication at an amplitude 80, on ice, and subsequently centrifuged at 14,000 × *g* (4°C) for 30 min. The supernatant was collected for purification.

Proteins in the above supernatant were purified using GSTrapTM FF column via the following steps:

1. Buffer preparation: binding buffer (PBS, pH 7.3) and elution buffer (50 mM Tris with 10 mM reduced glutathione; final pH 8.0) were prepared.
2. Sample preparation: the above supernatant was filtered through a 45-μm filter to prevent any precipitate clogging the column.
3. A syringe was filled with the binding buffer and was connected to the column. Five column volumes (one column volume: 1 mL) of the binding buffer, were used to equilibrate the column.
4. The sample from step 2 was applied onto the column, at a flow rate of 0.2–1 mL/min. In this step, any proteins with a GST tag would bind to the resin in the column.
5. To prevent binding of unspecific contaminants, the column was washed with 5 to 10 column volumes (5–10 mL) of the binding buffer at a flow rate of 1–2 mL/min.
6. Five column volumes (5 mL) of the elution buffer were applied to the column at 1–2 mL/min. In this step, the targeted proteins that bound to the column in step 4 were eluted. The first 3 mL of protein effluent were collected.
7. Cleaning and storage of the column: the column was washed with two column volumes (2 mL) of washing buffer (6 M guanidine hydrochloride), followed

by five column volumes (5 mL) of PBS. This step was required for the removal of hydrophobic bound substances.

8. The column was finally stored in 20% ethanol.

The purified protein in the elution buffer was transferred to a storage buffer (1 mM Tris, 1 mM MgCl₂, and 10% glycerol, pH 7.6) via Vivaspin (30,000 molecular weight cut-off), for improved protein storage environment and further application.

A Vivaspin tube was filled with the above-purified protein and centrifuged at 10,000 rpm for 5 min. The liquid in the filtrate tube was removed and protein was collected in the concentrate. Finally, the concentrated protein was transferred into the storage buffer prepared beforehand.

2.2.1.4 Protein verification and concentration test

SDS-PAGE

To test the purity of the protein, SDS polyacrylamide gel electrophoresis (SDS-PAGE) was performed.

1. Glass cassette and casting stand

The short plate was positioned in front of the spacer plate and placed in the casting frame (Bio-Rad Mini-PROTEAN®). The casting frame was inserted into the casting stand and locked.

2. Gel preparation

Separating gel (10%) and stacking gel (5%) solutions were prepared as described in Table 2.2. APS and TEMED were added last, as they initiate the gel polymerisation. Firstly, the separating gel solution was poured between the glass plates. After it cooled down, the stacking gel was poured over the separating gel, and a comb was immediately inserted on top. The comb was removed after the gel had set.

Table 2.2: SDS gels recipe

Stacking gel (5%)		Separating gel (10%)	
30% polyacrylamide (mL)	0.850	30% polyacrylamide (mL)	3.333
1 M Tris, pH 6.8 (mL)	0.625	1.5 M Tris, pH 8.8 (mL)	2.500
10% APS (mL)	0.050	10% APS (mL)	0.100
10% SDS (mL)	0.050	10% SDS (mL)	0.100
TEMED (mL)	0.005	TEMED (mL)	0.004
H ₂ O (mL)	3.420	H ₂ O (mL)	3.967
Total volume (mL)	5.000	Total volume (mL)	10.000

3. Electrophoresis

For each protein sample, protein (5 μ L), loading buffer (5 μ L), and DTT (1 M, 2 μ L) were mixed, and then heated at 95°C for 10 min to completely denature the proteins.

Each sample was loaded into a specific well of the prepared gel. The electrophoresis was performed at 40 V for the first 20 min, to let the samples reach the separating gel. The voltage was then increased to 150 V and maintained until the end of the electrophoresis.

4. Gel staining

The gel was stained with Instant Blue solution for 1 h, rinsed with ddH₂O for 1 h, and its image was then acquired using an optical camera.

Bradford assay

Coomassie reagent was prepared by dissolving Coomassie Blue G (30 mg) in absolute ethanol (100 mL). Then, phosphoric acid (55 mL) was added, followed by ddH₂O, for the final volume of 1 L.

BSA was used to generate the calibration curve. Six samples (in triplicate) were prepared, containing 1, 2, 3, 4, 5, or 6 μ g of BSA dissolved in 100 μ L of the protein storage buffer (1 mM Tris, 1 mM MgCl₂, and 10% glycerol, pH 7.6). Then, 100 μ L of the sample was mixed with 1 mL of the Bradford reagent. The average OD value from

triplicate samples at 595 nm (control: blank, protein storage buffer only) was plotted against its corresponding BSA concentration (Figure 2.1).

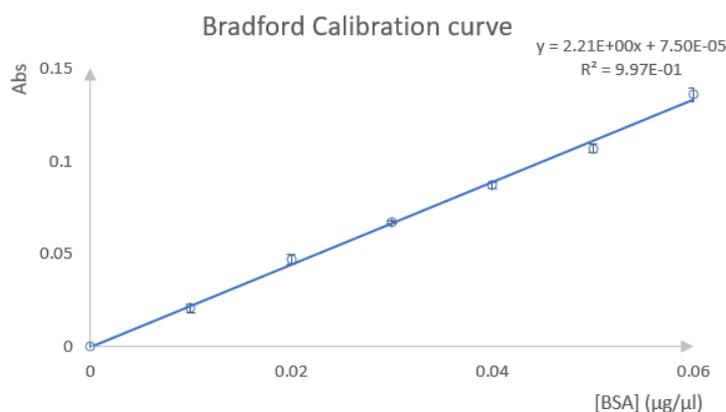


Figure 2.1: Bradford calibration curve. The experiments were run in triplicate and average reading value from triplicate samples under 595 nm was plotted.

To determine the UGT concentrations, 10 µL of UGT preparations were diluted 10 times by the addition of 90 µL of storage buffer, and mixed with 1 mL of the Bradford reagent. The OD value at 595 nm was obtained for each sample. Protein concentration in each sample was calculated using the equation obtained from the calibration curve.

2.2.2 MSA and SDM

2.2.2.1 MSA

MSA was performed using 11 target UGTs (73B4, 74B5, 73C1, 73C5, 73C6, 76E11, 76E12, 78D1, 78D2, 89B1, and 89C1) and seven template UGTs (VvGT1, 71G1, 72B1, 78G1, 78K6, 85H2, and OS79). Their AA sequences were first revised to the FASTA format (Figure 2.2).

```

>73B4
MAKLFARRGAKSTLLTTPINAKILEKPIEAFKVNQPDLEIGIKILNFCVLELPEGCENRDFINSYQKSDSFDLFLKFLFSTKYMKQLESFIETTKPSALVADMFPPWA
TESAEKIGVPRLVFHGTSSFFALCCSYNMRHHPKHKVASSSTPFVIPGLPGDIVITEDQANVTNEETPFPGKFWKEVRESESSFGVLVNSFYELESSYADFYRSFVAK
KAWHIGPLSLNRRGIAEKAGRGKKANIDEQECLKWLDSKTPGSSVYLSFGSGTGLNPQELLEIAFAGLEGGQNFIVVVSKNENQVGTGENEDWLKPGFEEERKNGK
GLIRGWAPQVLIDHKAIGGFVTHCGWNSTLEGIAAGLPMVTPWPMGAEQFYNEKLLTKVLRIGVNVGATELVKKGKLSRAQVEKAVREVIGGEKAEERRLRAKELG
EMAKAAVEEGSSYNDVNKFMEELENGRK

>73B5
MNREVERSERIHILFFPMAQGHMIPMDIARLLAQRGVITITVTPQNAGRFKNVLSRAIQGLPINLVQVKFPPQESGSPGQENLDLDSLGLSTFFKAFSLLEEP
YMKQQLSEFIETTKPSALVADMFPPWATESAEKLGVPRLVHGTSSFFALCCSYNMRHHPKHKVATSTPFVIPGLPGDIVITEDQANVAKEETPMGKFMKEVRESE
NSFGVLVNSFYELESAYADFYRSFVAKRAWHIGPLSLNRELGEKARRGKKANIDEQECLKWLDSKTPGSSVYLSFGSGTNTFDNQLLEIAFAGLEGGQSFIVVVR
KNENQGDNEEVLPEGFKERTTGKGLIIPGWAPQVLIDHKAIGGFVTHCGWNSTLEGIAAGLPMVTPWPMGAEQFYNEKLLTKVLRIGVNVGATELVKKGKLSRAQV
EKAVREVIGGEKAEERRLWAKKLGEMAKAAVEEGSSYNDVNKFMEELENGRK

>73C1
MASEFRPPLHFLVFPFMAQGHMIPMDIARLLAQRGVITITVTPQNAGRFKNVLSRAIQGLPINLVQVKFPPQESGSPGQENLDLDSLGLSTFFKAFSLLEEP
VEKLLKEIQPRPNICIIDMCLPYTNRIAKNLGIPKIIHGMCCFNLLCTHIMHQNHFEFTIESDKEYFPINFPDRVEFTKSQOLPMVLAAGDWDKDFDGMTEGDNSTY
GVIVNTFEELEPAYVRDYKVKAGKIWSIGPVSCLCNKLGEDQAERGNKADIDQDECIKWLDSKEEGSVLYVCLGSIICNPLSQLKELGLGLEESQRPFIWVIRGWEK
YNELLEWISSESGYKERIKERGLLITGWSQMLLTHPAVGGFLTHCGWNSTLEGITAGPLLTWPLFADQFCNEKLVQILKAGVRAGVEESMRWGEEEEKIGVLVDK
EGVKKAVEELMGDSNDAKERRRKRKELGELAHKAVEEGSSHSNITFLLQDIMLEQPKK

>73C5
MVSETTKSSPLHFLVFPFMAQGHMIPMDIARLLAQRGVITITVTPHNAARFKNVLSRAIQGLPINLVQVKFPPQESGSPGQENLDLDSLGLSTFFKAFSLLEEP
PVQKLIIEEMNPRPSCISDFCLPYTSKIAKFNIPKILFHGMGCFCLLCMHVLRKNREILDNLKSDKELFTVPDFPDRVEFTKSQOLPMVLAAGDWDKDFDGMTEGDNSTY
ETSYGIVNSFQLEPAYAKDYKVERSGKAWTIGPVSCLCNKVGADKAERGNKSDIDQDECIKWLDSKKHGSLVLYVCLGSIICNPLSQLKELGLGLEESQRPFIWVIRG
GWEKYKELVEWFSSESGFEDRIQDRGLLIGWSPQMLLTHPAVGGFLTHCGWNSTLEGITAGPLLTWPLFADQFCNEKLVQILKAGVRAGVEESMRWGEEEEKIGVLVDK
GVLVDKEGVKKAVEELMGESDDAKERRRRAKELGSAHKAVEEGSSHSNISLQDIMLAEPNN

>73C6
MAFEKNEPFPPLHFLVFPFMAQGHMIPMDIARLLAQRGVITITVTPHNAARFKNVLSRAIQGLPINLVQVKFPPQESGSPGQENLDLDSLGLSTFFKAFSLLEEP
EPVGNLIEEMNPRPSCISDFCLPYTSKIAKFNIPKILFHGMGCFCLLCMHVLRKNREILDNLKSDKELFTVPDFPDRVEFTKSQOLPMVLAAGDWDKDFDGMTEGDNSTY
TSYGIVNSFQLEPAYAKDYKVERSGKAWTIGPVSCLCNKVGADKAERGNKSDIDQDECIKWLDSKKHGSLVLYVCLGSIICNPLSQLKELGLGLEESQRPFIWVIRG
GWEKYKELVEWFSSESGFEDRIQDRGLLIGWSPQMLLTHPAVGGFLTHCGWNSTLEGITAGPLLTWPLFADQFCNEKLVQILKAGVRAGVEESMRWGEEEEKIGVLVDK
VDKEGVKKAVEELMGESDDAKERRRRAKELGSAHKAVEEGSSHSNISLQDIMLAEPNN

>76E11
MEEKPAGRRVVLVAVPAQGHISPMQLAKTLHLKGFSTIAQTKFNYFSPSDDFTDFQFVTIPESLPESDFELGPIEFLHKLNEKQVSKFDCLGQLLQQGNEIACV
VYDEFMYFAEAAAKEFKLPNIFSTTSATAFVCRSAFDKLYANSILTPLEKPKQQNELVPEFHLRCKDFPVSHWASLESMMELYRNTVDKRTASSVIINTASCLSS
SLSRLQQQLQIPVYPIGPHLHVASASTSLEENKSCIEWLNKQKNSVIFVSLGSLALMEINEVIEITALGLDSSKQQLWVIRPGSVRGSEWENLPKEFSKIISGRGYIV
KWAPQKEVLSHPAVGGFVSHCGWNSTLESIGEGVPMICKPFSDDQMVNARYLECVWIKIQVEGDLDRGAVERAVRRLMVEEAGEEMRKRAISLKEQLRASVIS
GSSHSNLEEFVHYMRTL

>76E12
MEEKPARRSVVLVAVPAQGHISPMQLAKTLHLKGFSTIAQTKFNYFSPSDDFTDFQFVTIPESLPESDFELGPIEFLHKLNEKQVSKFDCLGQLLQQGNEIACV
CVIYDEFMYFAEAAAKEFKLPNIFSTTSATAFACRSVFDKLYANNVQAPLKETKGQEEELVPEFYPLRYKDFPVSRAFASLESIMEVYRNTVDKRTASSVIINTASCLSS
SSLSFLQQQLQIPVYPIGPHLHVASAPTSLLEENKSCIEWLNKQKNSVIFVSLGSLALMEINEVIEITALGLDSSKQQLWVIRPGSVRGSEWENLPKEFSKIISGRGYIV
GYIVKWAPQKEVLSHPAVGGFVSHCGWNSTLESIGEGVPMICKPFSDDQMVNARYLECVWIKIQVEGDLDRGAVERAVRRLMVEEAGEEMRKRAISLKEQLRASVIS
SVKGGSSHSNLEEFVHFIRTL

>78D1
MTKFSSEPIRDSHVAVLAFPPGAHAGPLAVTRRLAASPSTIFSFFNTARSNASLFSDDHPENIKVHDVSDGVPEGTMLGNPLEMVELFLEAAPRIFRSEIAAAAEIV
GKVTMLTDAFFWFAADIAAELNATWVAFWAGGANSLSLAHLYTDLIREITIGLKVDSMEETLGFIPGMENYRVKDIPEEVVFEDLDSVFPKALYQMSLALPRASAVFI
SSFEELPTLNYNLSKLRFLNIAPLTLSTSEKEMRDPHGCFAWMGKRSAASVAYISFGTVMPEPPEELVAIAQGLESSKVPFVWSLKEKNMVHLKPGFLDRTR
EQGIVVPWAPQVELLKHAMGVNVTHCGWNSVLESVAGVPMIGRPILADNRLNGRAVEVWVKVGMMDNGVFTKEGFEKCLNDVVFVHDDGKTKMANAKKLE
KLQEDFSMKGSSLENFKILLDEIVK

>78D2
MTKFSSEPIRDSHVAVLAFPPGAHAGPLAVTRRLAASPSTIFSFFNTARSNASLFSDDHPENIKVHDVSDGVPEGTMLGNPLEMVELFLEAAPRIFRSEIAAAAEIV
AETEVEGTEVKCLMTDAFFWFAADIAAELNATWVAFWAGGANSLSLAHLYTDLIREITIGLKVDSMEETLGFIPGMENYRVKDIPEEVVFEDLDSVFPKALYQMSLALPRASAVFI
PRATAVINSFEDLPTLNTLNSRFRKRYLNIGPLGLLSSTLQQLVQDPHGLAWMEKRSRSGVAYISFGTVMTPPPGELAAIAEGLESSKVPFVWSLKEKSLVQLPK
GFLDRITREQGIVVPWAPQVELLKHAMGVNVTHCGWNSVLESVAGVPMIGRPILADNRLNGRAVEVWVKVGMMDNGVFTKEGFEKCLNDVVFVHDDGKTKMANAKKLE
KKLRELAYEAVSSKGRSSENFRLDVAVNI*

>89B1
MKVNEENKPTKTHVLIFPPAQGHMIPMDIARLLAQRGVITITVTPHNAARFKNVLSRAIQGLPINLVQVKFPPQESGSPGQENLDLDSLGLSTFFKAFSLLEEP
SWITSHSPPPVAIVSDFFLGWTKNLGIPRFDSPSAAITCCILNTLWIEPMTKINEDDDNEILHFKIPKPCPKYRFDQISSLYRSYVHGDPWAFEFIRDSFRDNVAVSWGLV
VNSFTAMEGVYLEHLKREMGHDIRVWAVGPIIPLSGDNRGGPTSVSDHVMWLDAREDNHVYVCFGSQVLTKEQLALASGLEKSGVHFIWAVKEPVEKDST
RGNILDRGDFDRVAGRGLVIRGWAPQVAVLRHRAVGAFLTHCGWNSVLEAVVAVGLMLTWPMDRQDYTASLVVDELKVGVRACEGPDVTPDPDELARVAFDSVT
RQNTKERIKAVELRKAALDAIQERSSVNDLDGFIQHVLSGLNK*

>89C1
MTTFTTKKPHVLVIFPPAQGHMIPMDIARLLAQRGVITITVTPHNAARFKNVLSRAIQGLPINLVQVKFPPQESGSPGQENLDLDSLGLSTFFKAFSLLEEP
RQPPSPLDPAIGSSFLSPWINKVADAFSISIFLPIAHSISVMWAQEDRSFFNDLETATTESYGLVINSFYDLEPEFVETVKTRFLNHHRIWTVGLLPFGAGVDR
GGQSSIPPAKVASWLDSCPEDNSVYVYGFSGQIRLTAETAALAAALEKSSVRFIWAVRDAAKVNSSDNSVEEDVIPAGFEERVKEKGLVIRGWAPOTMILEHRAV
GSYLTHLGGVSVLEGMVGGVMLLAWPMQADHFFNTLIVDKLRAAVRVGENRDSVPSDKLARILAESAREDLPERVTLMLKREKAMEAIEGGSSYKNLDELVA
EMCL

```

Figure 2.2: FASTA format of AA sequences of target UGTs: sequence of AAs follows > (the identifier of the sequence) + the name of the UGT

The AA sequences of target UGTs in the FASTA format were imported to the sequence alignment tool Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>).

Parameter settings are shown in Figure 2.3. The outcomes of sequence alignment are presented after automated calculation.

<u>OUTPUT FORMAT</u> Clustal w/o numbers ▼	
<u>DEALIGN INPUT SEQUENCES</u> no ▼	<u>MBED-LIKE CLUSTERING GUIDE-TREE</u> yes ▼
<u>MAX GUIDE TREE ITERATIONS</u> default ▼	<u>MAX HMM ITERATIONS</u> default ▼
<u>MBED-LIKE CLUSTERING ITERATION</u> yes ▼	
<u>NUMBER of COMBINED ITERATIONS</u> default(0) ▼	
<u>ORDER</u> aligned ▼	

Figure 2.3: Parameter settings for MSA.

In target UGTs, at sites equivalent to those where ckAAs are located in the template UGTs, i.e. T19, T141, T280, W332, H350, E358, D374, and Q375 (VvGT1 numbering), potential ckAAs were selected and subsequently substituted using SDM (as summarised in Table 2.1).

2.2.2.2 SDM

SDM was performed to obtain specific protein mutants, generally following the standard procedure: WT-encoding plasmid extraction (used as template plasmid in SDM), primer design, and SDM.

WT-encoding plasmid extraction

The WT-encoding plasmids for target UGTs (73B4, 73B5, 73C1, 73C5, 73C6, 76E11, 76E12, 78D1, 78D2, 89B1, and 89C1) were extracted from the bacterial glycerol stocks, and purified prior to SDM.

Using a QIAprep® Spin Miniprep kit, the plasmids were extracted at room temperature, as follows: 1) bacterial culture (5 mL) was incubated at 37°C overnight in LB medium (10 mL) with 50 µg/mL of ampicillin. Cells were harvested by centrifugation at 14,000 rpm for 3 min; 2) the cells were re-suspended in buffer P1

(250 μ L) of the kit; 3) buffer P2 (250 μ L, with LyseBlue reagent as an indicator) was added, and the sample mixed thoroughly until the solution became clear blue, this was usually controlled within 5 min; 4) buffer N3 (350 μ L) was added, and the sample mixed until the solution turned colourless; 5) the mixture was centrifuged for 10 min at 13,000 rpm, and supernatant (800 μ L) was transferred to the QIAprep spin column; 6) the column was washed with buffer PE (750 μ L) and centrifuged for 1 min at 12,000 rpm, discarding the flow-through; 7) another 1-min at 12,000 rpm centrifugation was employed to remove the residual washing buffer; 8) the column was placed in a clean 1.5-mL Eppendorf tube, ddH₂O (50 μ L) was added to the centre of the column, and left for 1 min; the plasmid was eluted after a 1-min centrifugation at 12,000 rpm.

Primer design

Primers to generate the mutants (listed in Table 2.1) were designed based on the back-to-back principle, using the primer design tool NEBaseChanger™ (<http://nebasechanger.neb.com/>). The cDNA sequences of 11 target UGTs were obtained from <http://www.p450.kvl.dk/UGT.shtml>. The DNA sequences that coded the sites to be mutated were selected, and replaced by DNA sequences that coded the desired AA.

Herein, the primer design for 73B4 R7A is exemplified to demonstrate the standard procedure for this step (Figure 2.4): 1) 73B4 cDNA was imported into NEBaseChanger™; 2) the cDNA region encoding R7 was selected (AA 19–21, AGA); 3) the 3-AA sequence GCC was set as the desired sequence (alanine: GCA or GCT, or GCC, or GCG); 4) the forward and reverse primers were automatically designed, with their specific properties also provided, e.g. melting temperature (T_m) and annealing temperature (T_A), that were required for the next step.

NEBaseChanger v1.2.5

Quick-Start Steps

- 1 Input starting sequence
- 2 Set mutagenesis type
- 3 Set mutagenesis region
- 4 Set desired seq [optional]
- 5 View primers

Input

Click and drag to set mutagenesis region

>73B4 1383 bp

ATGGCCAAAGCTTTGCTGCTAGAGAGAGGAGCCAAATCAACTCTCCTCACAAAC
 CCCAATAAATGCTAAGATCTGGAGAAACCCATTGAAAGCATTCAAAGTTC
 AAAATCCTGATCTCCGAAATCGGAATCAAGATCCTCAATTTCCCTTGTGTA
 GAGCTTGGATTGCCAGAAAGGATGCGAGAACCGGACTTCATTAACCTATA
 CCAAAAATCTGACTCATTGACTTGTCTTGAAGTTTCITTTCTACCA
 AGTATATGAAACAGCAGTTGGAGAGTTTCATTGAAACAAACCAACCGAGT
 GCTCTTGTAGCCGATATGTTCTCCCTGGGCAACAGAATCCCGGAGAA
 GATCGGTGTTCCAAAGACTTGTGTTCCAGGCCACATCATCCTTGGCTTGT
 GTTGTGTTAACAATGAGGATTCATAAGCCACACAGAAAGTCGTTG
 AGTCTACTGCTATTGTAATCCCTGGTCTCCCTGGAGACATAGTTATTAC
 AGAAGACCAAGCCAAATGTCACCAACGAAAGAACTCCATTGCGAAAGTTT
 GGAAGAAGTCAGGGAAATCAGAGACCAAGTAGCTTGTGTTGTTGGAAT
 AGCTTCTCAGAGCTGGAATCATCTTATGCTGATTTTACCGTAGTTTGT
 GCGAAGAAAGCCGTTGATATAGTCCACTTTCATATCCAAACAGAGGGA
 TTCCAGAGAAAGCCGAGAGGAGAAAGGCAAAACATTGATGCAAGAA
 TGCTCAAATGGCTTGAATCTAAGACACCTGGCTCAGTAGTTACTTGTG

73B4 1383 bp

Substitution Insertion Deletion

Find:

Start and end positions included in substitution.
 Start (5') 19 End (3') 21

Desired Sequence: GGC
 Genomic acid coding alanine

Common Peptide Tags

Result

R K * M A K L F A A R G A K
 * K V D G Q A F R C Q R S Q I
 V E S R W P S F S L P E E P N
 GTAGAAAGTAGATGGCCAAAGCTTTGCTGccAGAGAGCCAAATC
 CATCTTTCATCTACCGTTGAAAAGCGACGGTCTCCTCGGTTAG

Required Primers

Name (F/R)	Oligo (Uppercase = target-specific primer)	Len	% GC	Tm	Ta *
QSSDM_10/13/2016_F	GCTTTTCGCTGccAGAGGAGCCAAATC	27	56	63°C	60°C
QSSDM_10/13/2016_R	TTGGCCATCTACTTTCTAC	19	42	59°C	

* Ta (recommended annealing temperature)

Physical Properties of primers

Figure 2.4: Primer design for 73B4 R7A. The genetic code of R is highlighted and shown in amber character. They are replaced by the genetic code of A (highlighted in red rectangle). The physical properties of primers are also calculated (highlighted in blue rectangle)

Following the same procedure, primers for other mutants were designed and summarised in Tables 2.3–2.13. The designed primers were synthesised at Eurofins.

Table 2.3: Primer design of 73B4 mutants. Genetic codes of mutant site were underlined.

Mutants	Targeted sequence	Forward primer	Reverse primer	T _A (°C)
R7A	AGA (19 th -21 st)	GCTTTTCGCTGCCAGAGGAGCCAAATC	TTGGCCATCTACTTTCTAC	60
G127A	GGC (379 th -381 st)	TGTGTTCCACGCCACATCATCCT	AGTCTTGGAACACCGATC	66
S270A	AGT (808 th -810 th)	GTCCTTTGGTGCCGGAACCGGCT	AAGTAAACTACTGAGCCAGG	63
W331A	TGG (991 st -993 rd)	AATACGCGGAGCCGCCCCGCAAG	ATCAGCCCTTTTCCTTTATTC	61
H349A	CAT (1045 th -1047 th)	ATTTGTGACGGCCTGCGGATGGAAC	CCTCCGATTGCTTTGTGG	59
E357A	GAG (1069 th -1071 st)	CTCGACTTTGGCAGGCATTGCCG	TTCCATCCGCAATGCGTC	63
E373A	GAA (1117 th -1119 th)	GATGGGGGCAGC <u>AC</u> AGTTCTACAA	GGCCAAGTCACCATAGGC	67
E373K		GATGGGGGCAAA <u>AC</u> AGTTCTACAA		65
E373D		GATGGGGGCAG <u>ACC</u> AGTTCTACAA		67
E373Q		GATGGGGGCACA <u>AC</u> AGTTCTACAA		66
Q374A		GATGGGGGCAGAA <u>GC</u> ATTCTACAACGAG		ATCGGCCAAGTCACCATAG
Q374N	GATGGGGCAGAAA <u>ACT</u> TCTACAACGAG	61		
Q374H	GATGGGGCAGAA <u>CAC</u> TCTACAACGAG	63		
Q374E	GATGGGGCAGAA <u>AG</u> TTCTACAACGAG	65		

Table 2.4: Primer design of 73B5 mutants. Genetic codes of mutant site were underlined.

Mutants	Targeted sequence	Forward primer	Reverse primer	T _A (°C)
G154A	GGT (460 th -462 nd)	TGTGTTCCACGCCACATCTTTCTTTTC	AGTCTTGGTACACCGAGC	63
E397A	GAA (1189 th -1191 st)	AATGGGGGCAG <u>CCC</u> AGTTCTACA	GGCCATGTTACCATAGGC	65
E397D		AATGGGGGCAG <u>ACC</u> AGTTCTACA		65
E397K		AATGGGGGCAAA <u>AC</u> AGTTCTACA		65
E397N		AATGGGGGCAA <u>ACC</u> AGTTCTACA		64
Q398A		GATGGGGCAGAA <u>GC</u> ATTCTACAATGAG		ATTGGCCATGTTACCATAG
Q398N	GATGGGGCAGAAA <u>ACT</u> TCTACAATGAG	59		
Q398H	GATGGGGCAGAA <u>CAC</u> TCTACAATGAG	61		
Q398E	GATGGGGCAGAA <u>AG</u> TTCTACAATGAG	61		

Table 2.5: Primer design of 73C1 mutants. Genetic codes of mutant site were underlined.

Mutants	Targeted sequence	Forward primer	Reverse primer	T _A (°C)
G148A	GGC (442 nd -444 th)	ATCTTTCAT <u>GGC</u> ATGTGTGCTTC	ATTTTGGTATAACCAAGATTCTTG	60
D393A	GAC (1177 th -1179 th)	ACTGTTTGGAG <u>CCCA</u> ATTCTGCAATG	GGCCACGTGAGTAATGGA	65
D393E		ACTGTTTGGAGAG <u>CA</u> ATTCTGCAATG		63
D393K		ACTGTTTGGAA <u>AGCA</u> ATTCTGCAATG		64
Q394A		GTTTGGAGAC <u>GC</u> ATTCTGCAATGAGAAATTGGCG		AGTGGCCACGTGAGTAATG
Q394E	GTTTGGAGAC <u>GAG</u> TTCTGCAATGAGAAATTGGCG	63		
Q394H	GTTTGGAGAC <u>CACT</u> TTCTGCAATGAGAAATTGGCG	65		
Q394N	GTTTGGAGAC <u>AACT</u> TTCTGCAATGAGAAATTGGCG	61		

Table 2.6: Primer design of 73C5 mutants. Genetic codes of mutant site were underlined.

Mutants	Targeted sequence	Forward primer	Reverse primer	T _A (°C)
G150A	GGC (448 th -450 th)	CCTCTCCATGCCATGGGTTGCT	ATCTTTGGGATATTGAACTTCTTG	63
D397A	GAC (1189 th -1191 st)	GCTATTCGAG <u>CCCA</u> ATTCTGCAATG	GGCCATGTAAGTAGCGGT	66
D397E		GCTATTCGAGAG <u>CA</u> ATTCTGCAATG		67
D397K		GCTATTCGCA <u>AAGCA</u> ATTCTGCAATG		66
Q398A		ATTTCGAGAC <u>GC</u> ATTCTGCAATGAGAAATTGGTTCG		AGCGCCATGTAAGTAGC
Q398E	ATTTCGAGAC <u>GAG</u> TTCTGCAATGAGAAATTGGTTCG	62		
Q398H	ATTTCGAGAC <u>CACT</u> TTCTGCAATGAGAAATTGGTTCG	61		
Q398N	ATTTCGAGAC <u>AACT</u> TTCTGCAATGAGAAATTGGTTCG	62		

Table 2.7: Primer design of 73C6 mutants. Genetic codes of mutant site were underlined.

Mutants	Targeted sequence	Forward primer	Reverse primer	T _A (°C)
G151A	GGC (451 st -453 rd)	CCTCTTCCAT <u>GCC</u> ATGGGTTGCT	ATCTTTGGTATTTGAACTTCTTGG	66
D397A	GAC (1189 th -1191 st)	ACTATTTGCAG <u>CCCA</u> ATTCTGCAAC	GGCCATGTAAGCATTGGTAG	63
D397E		ACTATTTGCAG <u>GAGCA</u> ATTCTGCAAC		61
D397K		ACTATTTGCA <u>AAGCA</u> ATTCTGCAAC		62
Q398A		ATTTGCAGAC <u>GCA</u> TTCTGCAACGAGAAACTGG		AGTGGCCATGTAAGCATTG
Q398E	ATTTGCAGAC <u>GAGTT</u> CTGCAACGAGAAACTGG	62		
Q398H	ATTTGCAGAC <u>CACTT</u> CTGCAACGAGAAACTGG	61		
Q398N	ATTTGCAGAC <u>AACTT</u> CTGCAACGAGAAACTGG	61		

Table 2.8: Primer design of 76E11 mutants. Genetic codes of mutant site were underlined.

Mutants	Targeted sequence	Forward primer	Reverse primer	T _A (°C)
T134A	ACC (400 th -402 nd)	CATTTTCAGC <u>GCC</u> ACAAGTGCCA	ACGTTTGAAGCTTAAACTCTTG	64
D373A	GAT (1117 th -1119 th)	GTTTTCCAGT <u>GCCC</u> AAATGGTGAATG	GGCTTGCAAATCATTGGAAC	60
D373E		GTTTTCCAGT <u>GAGCA</u> AAATGGTGAATG		63
D373K		GTTTTCCAGT <u>AAGCA</u> AAATGGTGAATG		59
D373N		GTTTTCCAGT <u>AACCA</u> AAATGGTGAATG		60
Q374A	CAA (1120 th -1122 nd)	TTCCAGTGAT <u>GCA</u> ATGGTGAATGCG	AACGGCTTGCAAATCATTG	61
Q374E		TTCCAGTGAT <u>GAGAT</u> GGTGAATGCG		59
Q374H		TTCCAGTGAT <u>CACAT</u> GGTGAATGCG		60
Q374N		TTCCAGTGAT <u>AATAT</u> GGTGAATGCG		60

Table 2.9: Primer design of 76E12 mutants. Genetic codes of mutant site were underlined.

Mutants	Targeted sequence	Forward primer	Reverse primer	T _A (°C)
T135A	ACA (403 rd -405 th)	CATTTTCAGCGCAACAAGTGCCAC	ATGTTTGGGAAGCTTACACTC	66
D375A	GAT (1123 rd -1125 th)	ATTTTCGGGTGCCCAAAGGTGAAC	GGCCTGCAGATCATTGGA	67
D375E		ATTTTCGGGTGAGCAAAGGTGAAC		66
D375K		ATTTTCGGGT <u>AAG</u> CAAAGGTGAAC		60
D375N		ATTTTCGGGT <u>AAC</u> CAAAGGTGAAC		61
Q376A		CAA (1126 th -1128 th)		TTCGGGTGATGCAAAAGGTGAACG
Q376E	TTCGGGTGATGAGAAGGTGAACG		61	
Q376H	TTCGGGTGATCACAAGGTGAACG		63	
Q376N	TTCGGGTGAT <u>AATA</u> AGGTGAACG		59	

Table 2.10: Primer design of 78D1 mutants. Genetic codes of mutant site were underlined.

Mutants	Targeted sequence	Forward primer	Reverse primer	T _A (°C)
A23T	GCT (67 th -69 th)	CCCCGTTGGC <u>ACC</u> CATGCCGGTC	AAAAACGCGAGAAGTGCAC	67
A141T	GCC (421 st -423 rd)	TGCCTTCTGG <u>ACG</u> GGCGGAGCAA	ACCCAAGTCGCGTTCAGC	69
T280A	ACC (838 th -840 th)	TAGCTTCGGC <u>GCC</u> GTCATGGAAC	ATGTACGCTACAGAAGCAGCTG	69
W332A	TGG (994 th -996 th)	AGTGGTTCCTG <u>CCG</u> TCCACAAG	ATCCCTTGCTCTCTTGTC	58
H350A	CAT (1048 th -1050 th)	GAATGTGACAG <u>CCT</u> GTGGATGGAAGTCAAGTGTG	ACACCCATGCTCCTCGTGT	63
E358A	GAG (1072 nd -1074 th)	CTCAGTGTGGCCAGTGTGTCGG	TTCCATCCACAATGTGTC	59
D374A	GAT (1120 th -1122 nd)	GATTTTGGCGGCAATAGGCTCAACG	GGTCTGCCGATCATCGGT	65
N375A	AAT (1123 rd -1125 th)	TTTGGCGGATGCAAGGCTCAACG	ATCGGTCTGCCGATCATC	67
N375Q		TTTGGCGGAT <u>CAA</u> AGGCTCAACG		67

Table 2.11: Primer design of 78D2 mutants. Genetic codes of mutant site were underlined.

Mutants	Targeted sequence	Forward primer	Reverse primer	T _A (°C)
T22A	ACT (64 th -66 th)	TCCTTTTCGGCGCCCATGCAGCTC	AAAGCGAGAACTGCCACG	66
T286A	ACG (856 th -858 th)	TAGCTTTGGTGCCGTCATGACACCG	ATGTACGCCACAGAACCAG	63
W338A	TGG (1012 nd -1014 th)	AGTGGTTCCAGCCGCACCGCAAG	ATCCCTTGCTCTCTTGTC	60
H356A	CAT (1066 th -1068 th)	GTTTGTGACGGCCTGTGGATGGAACCTCGG	ACACCCGTTGCTTCGTGT	65
E364A	GAG (1090 th -1092 nd)	CTCGGTGTTGGCCAGTGTATCGG	TTCCATCCACAATGCGTC	63
D380A	GAT (1138 th -1140 th)	ATTTTTTGGGGCCAGAGATTGAACGG	GGCCTGCAAATCATCGGT	63
D380E		ATTTTTTGGGGAGCAGAGATTGAACGG		63
D380K		ATTTTTTGGGAAAGCAGAGATTGAACGG		63
D380N		ATTTTTTGGGAAACCAGAGATTGAACGG		64
Q381A	CAG (1141 st -1143 rd)	TTTTGGGGATGCCAGATTGAACGGAAG	AATGGCCTGCAAATCATC	60
Q381E		TTTTGGGGATGAGAGATTGAACGGAAG		61
Q381H		TTTTGGGGATCACAGATTGAACGGAAG		59
Q381N		TTTTGGGGATAACAGATTGAACGGAAG		60
C357A-C374A	TGT (1069 th -1071 st)- TGC (1120 th -1122 nd)	GTATCGGGTGGTGTACCGATGATTGCCAGGCCATT TTTTGGGGATC	ACTCTCCAACACCGAGTTCATCCA GCAATGCGTCACAAACACACC	63

Table 2.12: Primer design of 89B1 mutants. Genetic codes of mutant site were underlined.

Mutants	Targeted sequence	Forward primer	Reverse primer	T _A (°C)
P145A	CCC (433 rd -435 th)	CGATTCTCTGCCTCCGCTGCTA	AAACGAGGGATTCCGAGG	64
D389A	GAC (1165 th -1167 th)	GATGAGAGCTGCCAGTACACTGAC	GGCCACGTCAGCATCAAA	63
D389E		GATGAGAGCTGAGCAGTACACTGAC		65
D389K		GATGAGAGCTAAGCAGTACACTGAC		62
D389N		GATGAGAGCTAACCAGTACACTGAC		63
Q390A	CAG (1168 th -1170 th)	GAGAGCTGACGCCTACACTGACGCGTC	ATCGGCCACGTCAGCATC	67
Q390E		GAGAGCTGACGAGTACACTGACGCGTC		66
Q390H		GAGAGCTGACCACTACACTGACGCGTC		65
Q390N		GAGAGCTGACAACACTACACTGACGCGTC		65

Table 2.13: Primer design of 89C1 mutants. Genetic codes of mutant site were underlined.

Mutants	Targeted sequence	Forward primer	Reverse primer	T _A (°C)
G20A	GGT (58 th -60 th)	TCCACAATCC <u>GCC</u> CACATGGTTC	AACGGTATCACCAGAACG	60
P147A	CCC (439 th -441 st)	TTAGTTTCTTAG <u>CC</u> ATCAATGCTCATTC	AGGATGGCGTCGGGGAGA	64
W314A	TGG (940 th -942 nd)	GATAAGAGGAG <u>CCG</u> CCCCACAAACTATG	ACGAGTCCTTTCCTCCTC	63
H332A	CAT (994 th -996 th)	TTACCTAACTG <u>CCT</u> TGGGTTGGGGTTCG	GATCCAACGGCTCGATGC	64
E340A	GAA (1018 th -1020 th)	TTCGGTTCTG <u>GCC</u> GGAATGGTTCG	CCCCAACCCAAATGAGTTAG	61
D356A	GAC (1066 th -1068 th)	GATGCAAGCAG <u>CCC</u> ATTTCTTTAAC	GGCCACGCTAGCAACATA	62
H357A	CAT (1069 th -1071 st)	GCAAGCAGACG <u>CAT</u> TCTTTAACAC	ATCGGCCACGCTAGCAAC	61
H357Q		GCAAGCAGAC <u>CAG</u> TCTTTAACAC		62

Exponential amplification (EA)

The sample for EA was prepared as follows: Q5[®] Hot Start High-Fidelity 2x master mix (12.5 μL) was mixed with WT-encoding plasmid (1 μL , purified as above, 1–25 ng/ μL), and forward and reverse primers (10 μM stock, 1.25 μL each); ddH₂O was added for a final volume of 25 μL .

The prepared sample was performed to a standard EA process (Figure 2.5): 1) initial denaturation (98°C for 30 s); 2) denaturation (98°C for 10 s); 3) primer annealing (50–72°C for 20 s, according to the annealing temperature, T_A , of primers for specific mutants, listed in Tables 2.3–2.13); 4) extension (72°C for 161 s); 5) final extension (72°C for 2 min); 6) standing at 4°C. Steps 2 to 4, were repeated 25 times.

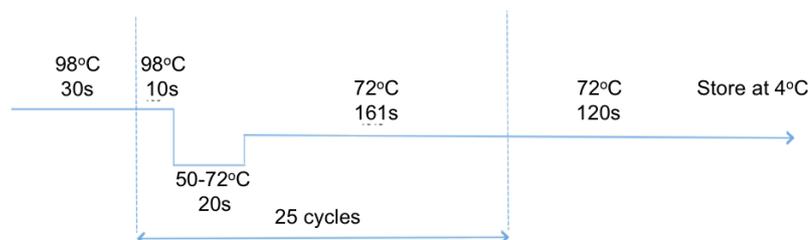


Figure 2.5: Schematic diagram of EA process.

EA product verification – agarose gel

The EA product was checked by agarose gel electrophoresis. Briefly, 0.8% agarose gel with 0.1% SYBR[®] Safe DNA gel stain was prepared in TBE buffer, cooled, and solidified in a gel caster. EA product (5 μL) and loading buffer (5 μL) were mixed, and loaded in a gel lane. Electrophoresis was performed at 35 V in TBE tank buffer for 3 h. The gel was evaluated using the Gel Imager, and a clear gel band at around 6 kb indicated successful EA process.

Treatment and enrichment

EA product (verified by agarose gel; 1 μL) was mixed with 2x KLD reaction buffer (5 μL) and 10x KLD enzyme mix (1 μL). Nuclease-free water (3 μL) was added for a final volume of 10 μL . The sample was then incubated for 5 min at room temperature and used in transformation.

Transformation and mutant plasmid verification

The mutated plasmids (products of the ‘treatment and enrichment’ step) were transformed into XL-1 blue competent cells, using the same procedure as in Section 2.2.1.1. A few representative mutated plasmids were selected and sent to Source Bioscience for DNA sequencing for verification. These mutated plasmids were 73C6 Q398A, 78D1 N375Q, 78D2 Q381E, and a plasmid encoding a double-point mutant 78D2 C357A-C374A (Table 2.14).

Table 2.14: Verified mutant plasmids by Sanger sequencing.

Mutant plasmid	Targeted sequences (in WT)	Outcome sequences (in mutant, verified by Sanger sequencing)
73C6 Q398A	CAA (1192 nd -1194 th)	GCA (1192 nd -1194 th)
78D1 N375Q	AAT (1123 rd -1125 th)	CAG (1123 rd -1125 th)
78D2 Q381E	CAG (1141 st -1143 rd)	GAG (1141 st -1143 rd)
78D2 C357A-C374A	TGT (1069 th -1071 st)- TGC (1120 th -1122 nd)	GCA (1069 th -1071 st) - GCC (1120 th -1122 nd)

2.2.3 MS-based enzyme activity test

2.2.3.1 Substrate specificity (GAR) screening

Substrate specificity screening was performed using LC-MS/MS by monitoring the formation of the product (glycosylated product) in the negative-ion mode. The ‘full scan’ mode was used for an initial screening of the UGT’s activity to determine whether a potential glycosylated product had formed in the reactions containing the tested UGTs and tested substrates. If such a glycosylated product was detected, the ‘product ion’ mode was subsequently applied to fragment the potential glycosylated product, to confirm the activity by determining whether aglycone was produced during this step. The result was displayed as a GAR panel, using colours red (no activity), amber (ambiguous activity), and green (positive activity).

Sample preparation

Donor GAR screening

Donor GAR screening was used to examine the substrate specificities of 29 WT UGTs and 106 UGT mutants (acceptor: KMP). The donor library contained eight sugar

donors (Figure 2.6 and Table 2.15): α -D-UDP-Glucose (UDP-Glc), α -D-UDP-Galactose (UDP-Gal), α -D-UDP-*N*-acetylglucosamine (UDP-GlcNAc), α -D-UDP-*N*-acetylgalactosamine (UDP-GalNAc), β -L-UDP-Rhamnose (UDP-Rha), β -L-dTDP-Rhamnose (dTDP-Rha), α -D-GDP-Glucose (GDP-Glc), and β -L-GDP-Fucose (GDP-Fuc).

KMP (10 μ L, 100 μ M), donor (10 μ L, 100 μ M), and UGT (10 μ L, 0.01–0.1 mg/mL) were mixed in a buffer (50 μ L, 1 mM Tris with 1 mM MgCl₂, pH 7.6). The samples were incubated at 37°C overnight.

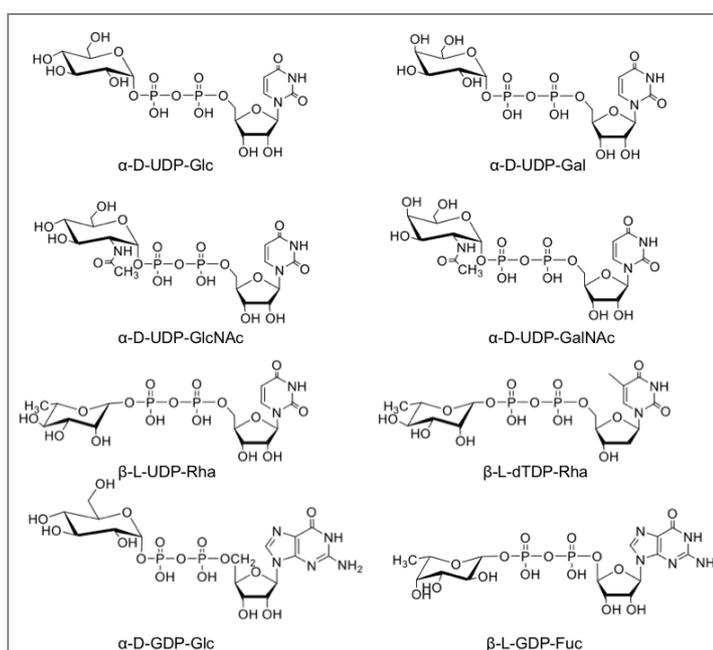


Figure 2.6: Donor library.

Table 2.15: Exact molecular weight of glycosylated products with different sugar donors (KMP as acceptor).

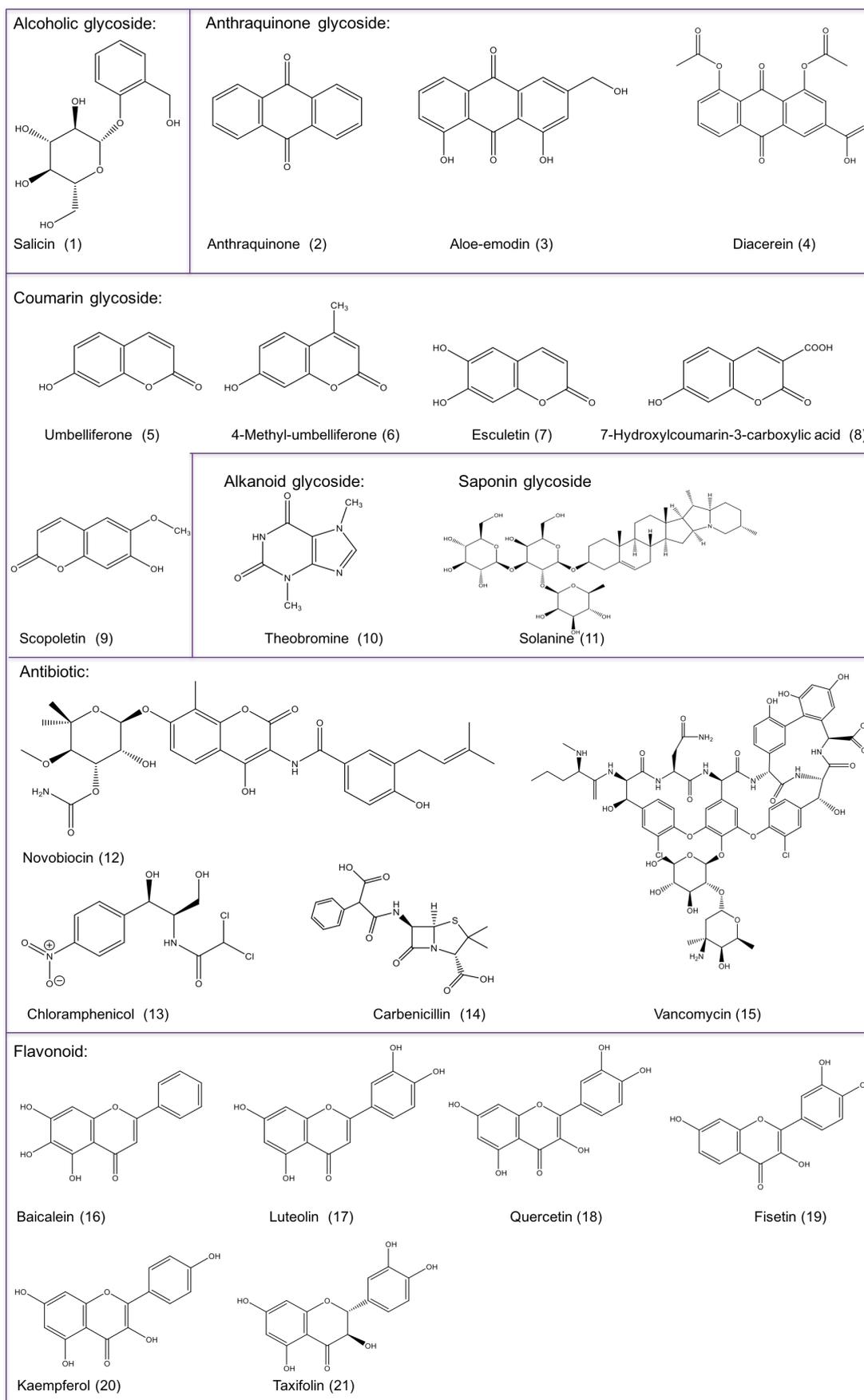
Donor	UDP-Glc	UDP-Gal	GDP-Glc	GDP-Fuc	UDP-GlcNAc	UDP-GalNAc	UDP-Rha	dTDP-Rha
Exact MW (Dalton)	448	448	448	432	489	489	432	432

Acceptor GAR screening

Acceptor GAR screening was used to examine the substrate specificities of 29 WT UGTs. UDP-Glc was used as the donor for 27 WT UGTs, and UDP-Rha was used as

the donor for 78D1 and 89C1. The acceptor library contained 48 acceptors, listed in Figure 2.7 and Table 2.16.

Similar to the donor GAR screening, acceptor (10 μ L, 100 μ M), donor (10 μ L, 100 μ M), and UGT (10 μ L, 0.01–0.1 mg/mL) were mixed in a buffer (50 μ L, 1 mM Tris with 1 mM MgCl_2 , pH 7.6). The samples were incubated at 37°C overnight.



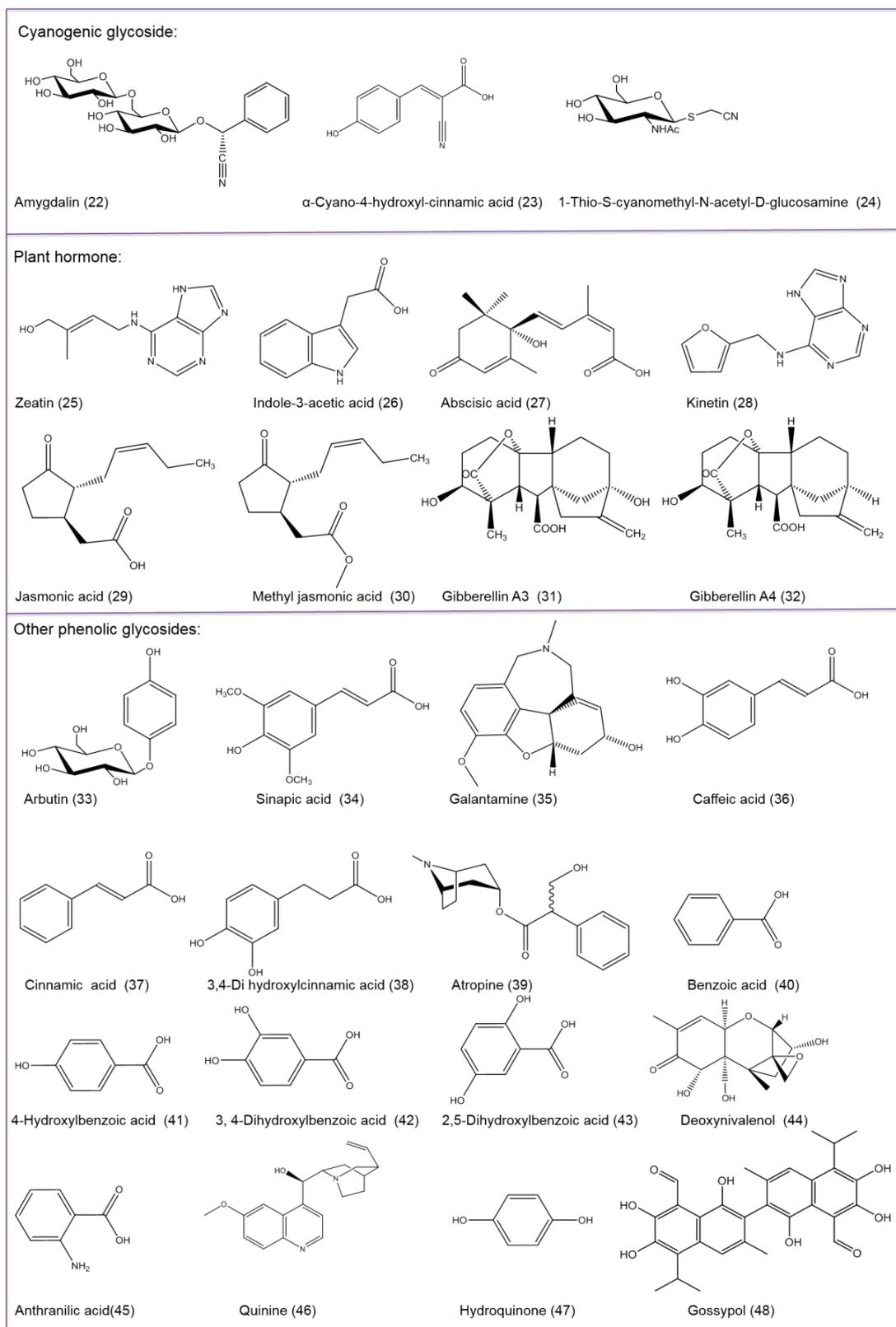


Figure 2.7: Sugar acceptor library– the numbers in brackets correspond to the number in the subsequent green-amber-red (GAR) screen assay.

Table 2.16: Exact molecular weight of glycosylated product with different acceptors.

No.	Chemical name	Formula	MW (Dalton)	MW+Glc (Dalton)	MW+Rha (Dalton)
1	Salicin	C ₁₃ H ₁₈ O ₇	286.28	448.28	432.28
2	Anthraquinone	C ₁₄ H ₈ O ₂	208.22	370.22	354.22
3	Aloe-emodin	C ₁₅ H ₁₀ O ₅	270.24	432.24	416.24
4	Diacerein	C ₁₉ H ₁₂ O ₈	368.30	530.30	514.30
5	Umbelliferone	C ₉ H ₆ O ₃	162.14	324.14	308.14
6	4-methyl-umbelliferone	C ₁₀ H ₈ O ₃	176.14	338.14	322.14
7	Esculetin	C ₉ H ₆ O ₄	178.14	340.14	324.14
8	7-Hydroxylcoumerin-3-carboxylic acid	C ₁₀ H ₆ O ₅	206.15	368.15	352.15
9	Scopoletin	C ₁₀ H ₈ O ₄	192.16	354.16	338.16
10	Theobromine	C ₇ H ₈ N ₄ O ₂	180.16	342.16	326.16
11	Solanine	C ₄₅ H ₇₃ NO ₁₅	868.06	1030.06	1014.06
12	Novobiocin	C ₃₁ H ₃₆ N ₂ O ₁₁	612.62	774.62	758.62
13	Chloramphenicol	C ₁₁ H ₁₂ Cl ₂ N ₂ O ₅	323.13	485.13	469.13
14	Carbenicillin	C ₁₇ H ₁₈ N ₂ O ₆ S	378.40	540.40	524.40
15	Vancomycin	C ₆₆ H ₇₅ Cl ₂ N ₉ O ₂₄	1449.30	1611.30	1595.30
16	Baicalein	C ₁₅ H ₁₀ O ₅	270.24	432.24	416.24
17	Luteolin	C ₁₆ H ₁₀ O ₆	286.24	448.24	432.24
18	Quercetin	C ₁₅ H ₁₀ O ₇	302.24	464.24	448.24
19	Fisetin	C ₁₆ H ₁₀ O ₆	286.24	448.24	432.24
20	Kaempferol	C ₁₆ H ₁₀ O ₆	286.24	448.24	432.24
21	Taxifolin	C ₁₅ H ₁₂ O ₇	304.25	466.25	450.25
22	Amygdalin	C ₂₀ H ₂₇ NO ₁₁	457.43	619.43	603.43
23	α -Cyano-4-hydroxyl-cinnamic acid	C ₁₀ H ₇ NO ₃	189.17	351.17	335.17
24	1-Thio-S-cyanomethyl-N-acetyl-D-glucosamine	C ₁₀ H ₁₆ N ₂ O ₅ S	276.08	438.08	422.08
25	Zeatin	C ₁₀ H ₁₃ N ₅ O	219.25	381.25	365.25
26	Indole-3-acetic acid	C ₁₀ H ₉ NO ₂	175.19	337.19	321.19
27	Abscisic acid	C ₁₅ H ₂₀ O ₄	264.32	426.32	410.32
28	Kinetin	C ₁₀ H ₉ N ₅ O	215.22	377.22	361.22
29	Jasmonic acid	C ₁₂ H ₁₈ O ₃	210.27	372.27	356.27
30	Methyl jasmonic acid	C ₁₃ H ₂₀ O ₃	224.14	386.14	370.14
31	Gibberellin A3	C ₁₉ H ₂₄ O ₆	348.16	510.16	494.16
32	Gibberellin A4	C ₁₉ H ₂₄ O ₅	332.16	494.16	478.16
33	Arbutin	C ₁₂ H ₁₆ O ₇	272.25	434.25	418.25
34	Sinapic acid	C ₁₁ H ₁₂ O ₅	224.21	386.21	370.21
35	Galantamine	C ₁₇ H ₂₁ NO ₃	287.35	449.35	433.35
36	Caffeic acid	C ₉ H ₈ O ₄	180.16	342.16	326.16
37	Cinnamic acid	C ₉ H ₈ O ₂	148.16	310.16	294.16
38	3,4-Di hydroxylcinnamic acid	C ₈ H ₈ O ₃	152.05	314.05	298.05
39	α -Cyano-4-hydroxyl-cinnamic acid	C ₁₀ H ₇ NO ₃	189.04	351.04	335.04
40	Benzoic acid	C ₇ H ₆ O ₂	122.12	284.12	268.12
41	4-Hydroxylbenzoic acid	C ₈ H ₈ O ₂	136.05	298.05	282.05
42	3,4-Dihydroxylbenzoic acid	C ₈ H ₈ O ₃	152.05	314.05	298.05
43	2,5-Dihydroxylbenzoic acid	C ₈ H ₈ O ₃	152.05	314.05	298.05
44	Deoxynivalenol	C ₁₅ H ₂₀ O ₆	296.32	458.32	442.32

45	Anthranilic acid	C ₇ H ₇ NO ₂	137.14	299.14	283.14
46	Quinine	C ₂₀ H ₂₄ N ₂ O ₂	324.42	486.42	470.42
47	Hydroquinone	C ₆ H ₆ O ₂	110.11	272.11	256.11
48	Gossypol	C ₃₀ H ₃₀ O ₈	518.56	680.56	664.56

‘Full scan’ mode to detect the product formation

After incubation overnight, samples were passed through a column (Kinetex 5 μ m C18 100A New Column, 50 \times 4.6 mm) in MS experiments.

LC conditions: the LC system was equipped with a binary solvent: water (with 0.1% formic acid) and acetonitrile (ACN). From 0.0–1.0 min, the percentage of ACN increased gradually from 30% to 45%; from 1.0–2.5 min, the percentage of ACN remained at 45%; from 2.5–5.0 min, the percentage of ACN decreased gradually to the initial percentage (30%). All analyses were performed in this gradient condition at a flow rate of 0.5 mL/min at room temperature. The sample volume injected was 10 μ L, with a run time of 5 min.

MS conditions: The ‘full scan’ mode of the quadrupole MS analyser was used in the negative-ion mode, with a range of 100–1000 MW for the donor GAR screening and 100–2000 MW for the acceptor GAR screening [fragmentor voltage (Frag) 135 V and cell accelerator voltage (CAV) 7 V]. The source parameters were set as follows: gas temperature of 300°C and flow of 5 L/min; nebuliser pressure of 45 psi; sheath gas temperature of 250°C at a flow of 11 L/min; the capillary voltage at 3500 V, negative; and the nozzle voltage at 500 V, negative.

Once the glycosylated product peak (the exact MW are shown in Tables 2.15 and 2.16) was detected in the ‘full scan’ mode, ‘product ion’ mode analysis was performed to double-check the existence of the glycosylated product, as follows.

‘Product ion’ mode to confirm the product formation

Samples that exhibited positive activity, as described above, were passed through the same column as above.

LC conditions: the same as above.

MS conditions: The ‘product ion’ mode of MS was used. Ions of the glycosylated product MW-1 were selected as the ‘precursor ions’ (e.g. KMP-Glc, 447; KMP-Gal, 447; KMP-Rha, 431; and KMP-GlcNAc, 488), allowing them to travel through the collision cell at collision energy (CE) of 10 eV, to produce ‘product ions’. The conditions of source, nebuliser pressure, sheath gas, capillary, and nozzle voltage were the same as above.

2.2.3.2 Full kinetics

MRM parameter optimisation

To determine the optimal MRM parameters for the glycosylated product and the internal standard, the software MassHunter Optimizer was used to optimise the Frag and CE for MRM transition and conditions.

A standard sample was prepared containing UDP-Glc (10 μ L, 100 μ M), KMP (10 μ L, 100 μ M), quercetin (10 μ L, 100 μ M), commercial KMP-3-Glc standard (10 μ L, 100 μ M), UDP (10 μ L, 100 μ M), and buffer (50 μ L, 1 mM Tris and 1 mM MgCl₂, pH 7.6). This sample was used to probe the MRM parameters of KMP-Glc and quercetin. Since KMP-GlcNAc, KMP-Gal, and KMP-Rha were not commercially available, their MRM parameters were probed using a glycosylated product generated by an enzymatic reaction. To probe the MRM parameters of KMP-GlcNAc, a sample containing UDP-GlcNAc, KMP, and 78D2 incubated together overnight at 37°C was used. To probe the MRM parameters of KMP-Rha, a sample containing UDP-Rha, KMP, and 78D1 incubated together overnight at 37°C was used. To probe the MRM parameters of KMP-Gal, a sample containing UDP-Gal, KMP, and 73C5 incubated together overnight at 37°C was used.

A range of MRM parameters were set, including a coarse range from 100 to 300, collision energy from 5 to 100, and a low mass cut-off of 100. Using different MRM parameters, the Glycosylated-KMP underwent different transitions (precursor ion \rightarrow product ion). The transition with the highest abundance under specific conditions was used in subsequent experiments.

Using the MRM parameter optimiser, the optimal MRM transitions and conditions for every analyte were obtained (Table 2.17):

Table 2.17:MRM parameters of analyte

Compound	MRM Transition	Frag (eV)	CE (eV)	CAV (eV)
KMP-Glc	447.1 → 284.1	235	33	4
KMP-GlcNAc	488.0 → 285.0	200	24	4
KMP-Gal	447.1 → 284.1	235	30	4
KMP-Rha	431.0 → 285.0	200	16	4
Quercetin	301.0 → 151.1	170	17	4

Other settings: the source parameters were set to gas temperature of 300°C and flow of 5 L/min; nebuliser pressure of 45 psi; sheath gas temperature of 250°C at a flow of 11 L/min; capillary voltage at 3500 V, negative; and nozzle voltage at 500 V, negative.

Selection of enzyme concentrations ([E])

Since a suitable [E] is required to observe a clear progressive curve (the curve of product quantity Vs time) with a linear section, [E] selection was performed prior to a full kinetic study.

Below, [E] selection prior to the full kinetic study of 78D2 is described, to exemplify how this step was performed.

Three samples were prepared. They all contained UDP-Glc (10 μ M) and KMP (10 μ M) in a buffer (1 mM Tris and 1 mM MgCl₂, pH 7.6), and different concentrations of 78D2 (0.0002, 0.0010, and 0.0020 mg/mL). The final sample volume was 150 μ L.

The glycosylated product was sampled after 5, 15, 30, and 50 min. For example, after 5 min since the addition of 78D2, 25 μ L of mixture from the original (reaction) Eppendorf tube were withdrawn and transferred to a new Eppendorf tube. An equal volume of ACN was added to the new Eppendorf tube to precipitate the protein and prevent further reaction. The sample/ACN mixture was then centrifuged for 10 min at 14,000 rpm. The supernatant (40 μ L) was transferred to a well of a 96-well plate. Quercetin (the internal standard, 10 μ L from 100 μ M stock solution) and ddH₂O (50 μ L) were added to the well. The glycosylated products formed after 15, 30, and 50 min were sampled in the same manner.

Product formation was quantified using the MRM mode of MS, as described above (the LC conditions were the same as described in section 2.2.3.1).

The [E] yielding the optimal progressive curve was used in the subsequent kinetic study (discussed in Section 3.1.2). This initial evaluation was carried out to select the suitable [E] for the kinetic study for every enzyme. The enzyme concentrations used are summarised in Table 2.18.

Table 2.18: The final concentration of UGTs used in the full kinetics test.

Donor	UGTs	Concentration (mg/ml)
UDP-Glc	89B1	0.0030
	73B4	0.0030
	73B5	0.0015
	73C1	0.0015
	73C5	0.0015
	73C6	0.0015
	78D2	0.0010
	76E11	0.0055
	76E12	0.0015
UDP-Gal	73C5	0.0055
	78D2	0.0030
UDP-GlcNAc	73B4	0.0070
	78D2	0.0015

Selection of the substrate concentration ([S]) – Ion suppression

In order to explore whether and how ion suppression took place, and whether it subsequently impacted the linear response of ESI-MS in kinetic assays, correlation curves of concentrations of commercial KMP-3-*O*-Glc and KMP-3-*O*-Glc/quercetin MRM counts were prepared at different UDP-Glc concentrations.

Four sets of samples were prepared: 1) quercetin (10 μM) mixed with different concentrations of KMP-3-*O*-Glc (0.001, 0.002, 0.005, 0.01, 0.02, 0.05, and 0.1 μM) in seven individual Eppendorf tubes, no UDP-Glc; 2) quercetin (10 μM) mixed with different concentrations of KMP-3-*O*-Glc in seven individual Eppendorf tubes, with low UDP-Glc (5 μM); 3) quercetin (10 μM) mixed with different concentrations of KMP-3-*O*-Glc (0.001, 0.002, 0.005, 0.01, 0.02, 0.05, and 0.1 μM) in seven individual Eppendorf tubes, with intermediate UDP-Glc (50 μM); 4) quercetin (10 μM) was mixed with different concentrations of KMP-3-*O*-Glc (0.001, 0.002, 0.005, 0.01, 0.02, 0.05, and 0.1 μM) in seven individual Eppendorf tubes, with high UDP-Glc (100 μM).

The same LC-MS conditions were applied as above, to generate correlation curves of concentrations of commercial KMP-3-*O*-Glc and KMP-3-*O*-Glc/quercetin MRM counts.

Full UGT kinetics

To perform the full UGT kinetic study, the experiment was set up to use four groups of samples with KMP concentrations fixed at 2, 5, 10, and 20 μM . In each group, five samples with the following donor concentrations were used: 2, 5, 10, 20, and 50 μM (UDP-Glc for 73B4, 73B5, 73C1, 73C5, 73C6, 76E11, 76E12, 78D2, and 89B1; UDP-Gal for 73B4 and 78D2; and UDP-GlcNAc for 73B4 and 78D2). In total, 20 samples with different KMP and donor concentrations were prepared (Figure 2.8).

Each sample contained 10 μL of UGT (the final concentration varied, as listed in Table 2.18) and buffer (1 mM Tris and 1 mM MgCl_2 , pH 7.6) added for a final volume of 150 μL . The glycosylated products were collected at different time points, as described above (Figure 2.8). The experiment was conducted at room temperature.

KMP-Glc formation in all 80 samples (20 reaction samples \times four time points) was monitored under the LC-MS condition described above. Velocity at different concentration was calculated and, thus, the full enzyme kinetics were investigated.

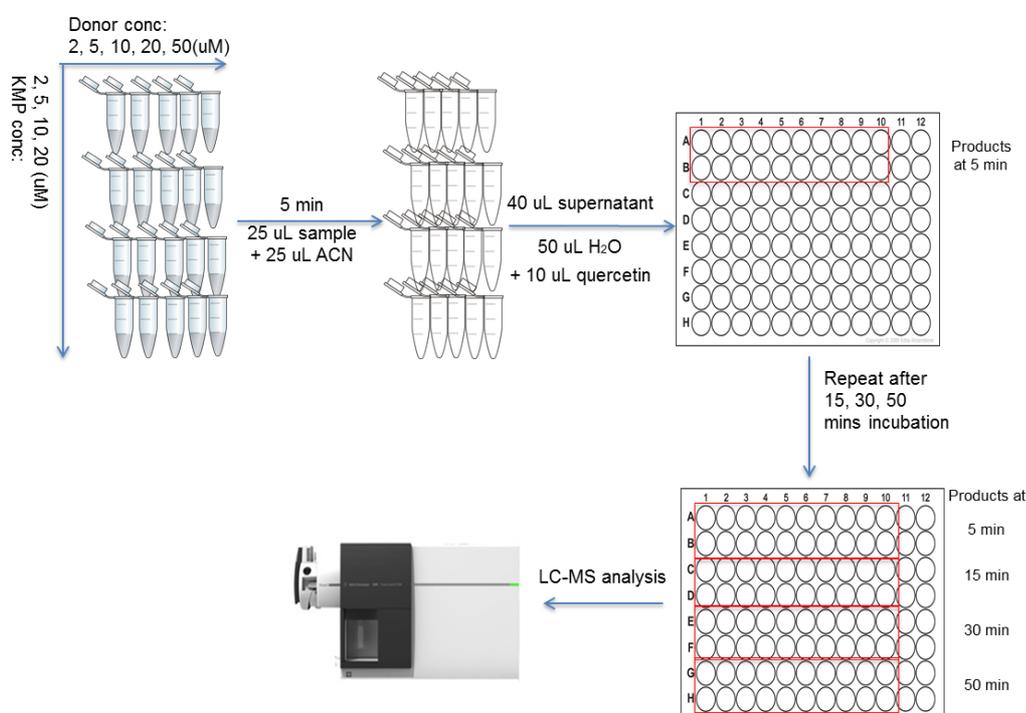


Figure 2.8: Schematic chart of sample preparation for the full kinetics test.

A calibration curve correlating the concentration of the glycosylated product and ions of the glycosylated product/internal standard was generated each time, in parallel to a full kinetic experiment. For that, six samples were prepared containing different concentrations of commercial KMP-3-*O*-Glc (0.001, 0.002, 0.005, 0.01, 0.02, or 0.05 μM), quercetin (10 μM), UDP-Glc (10 μM), and KMP (10 μM).

Product inhibition assays of UGTs

Product inhibition assays were performed for 73C5.

Four sets of kinetic experiments were carried out: 1) with UDP as the inhibitor (0, 50, and 200 μM), KMP (fixed at 100 μM), with varying concentrations of UDP-Glc (2, 5, 10, and 20 μM); 2) with UDP as the inhibitor (0, 50, and 500 μM), UDP-Glc (fixed at 100 μM), with varying concentrations of KMP (2, 5, 10, and 20 μM); 3) with KMP-3-Glc as the inhibitor (0, 1, and 5 μM), KMP (fixed at 100 μM), with varying concentrations of UDP-Glc (2, 5, 10, and 20 μM); and 4) with KMP-3-Glc as the inhibitor (0, 1, and 5 μM), UDP-Glc (fixed at 100 μM), with varying concentrations of KMP (2, 5, 10, and 20 μM).

The glycosylated products were collected at different time points, as described above, and the subsequent MS analysis procedure followed the same procedure as that described above.

2.2.3.3 Activity comparison of WT and mutants

Two sets of reaction samples were prepared: the control group (WT) and the experimental group (mutant). The control group contained the donor (15 μL , 1 mM), KMP (15 μL , 1 mM), buffer (105 μL , 1 mM Tris, and 1 mM MgCl_2 , pH 7.6), and WT enzyme (15 μL , the same concentration as used in the full kinetic study). The experimental group contained the same starting materials, but the enzyme used was the corresponding mutant (same concentration as WT).

For example, when comparing the activities of 73B4 WT and 73B4 R7A, the two groups of samples were: the control group, with UDP-Glc (10 μM), KMP (10 μM), and WT 73B4 (0.0030 mg/mL); and the experimental groups, with the same reactants but with the protein mutants 73B4 R7A (0.0030 mg/mL), accordingly.

Aliquots were withdrawn after 5, 15, 30, and 50 min, and were treated as described above. The products were quantified using the MRM mode of MS, as above. Progressive reaction curves with the two enzymes were then obtained. The initial rates of enzymatic velocity with different enzymes were compared in this closed system as a function of the reaction time.

Using this same approach, comparisons of the activities of WT 73B4, 74B5, 73C1, 73C5, 73C6, 76E11, 76E12, 78D1, 78D2, 89B1, and 89C1 with their corresponding mutants were performed, with the appropriate WT (control) and 1–5 corresponding mutants (the experimental group).

2.2.4 Homology modelling and protein overlay

2.2.4.1 Homology modelling

In the current project, Swiss-Model was used to build protein models of C-terminal of the target UGTs by homology modelling, using the C-terminal in crystal structures of template UGTs. Below, an example is provided on how to build a homology model of the C-terminal of UGT 73B4.

Since the homology modelling is more accurate if the template protein and target protein share a higher percentage of identical residues, the percentages of identical residues between C-terminal of 73B4 and template UGTs (71G1, 85H2, 78G1, VvGT1, 78K6, 72B1, and OS79) with known crystal structures were compared. The comparison revealed that the C-terminal of 73B4 and 71G1 shared the highest percentage of identical residues (46%). Hence, the C-terminal of 71G1 was used as the template for building the C-terminal model of 73B4.

The subsequent modelling was set up using the ‘User template’ program on Swiss-Model website (<https://swissmodel.expasy.org>) with default parameters. The C-terminal crystal structure of the template 71G1 (PDB: 2ACW) was used.

Following the same step-by-step procedure, C-terminal models of the other target UGTs were built, with their optimal template UGTs.

2.2.4.2 Protein overlay

To overlay the structure of two proteins, UCSF Chimera (Mac version) was used. In Section 2.2.4.1, 71G1 was used as the template to build a C-terminal model of 73B4 because of their high AA sequence identity. However, VvGT1 was used for the overlay in this step, as the VvGT1+U2F+KMP complex can reveal detailed information of how 73B4 might interact with its substrates (no 71G1+donor+acceptor complex was solved).

Thus, the protein overlay of VvGT1 and 73B4 was performed using ‘MatchMaker’ function of the chimera software with the default parameters (Figure 2.9).

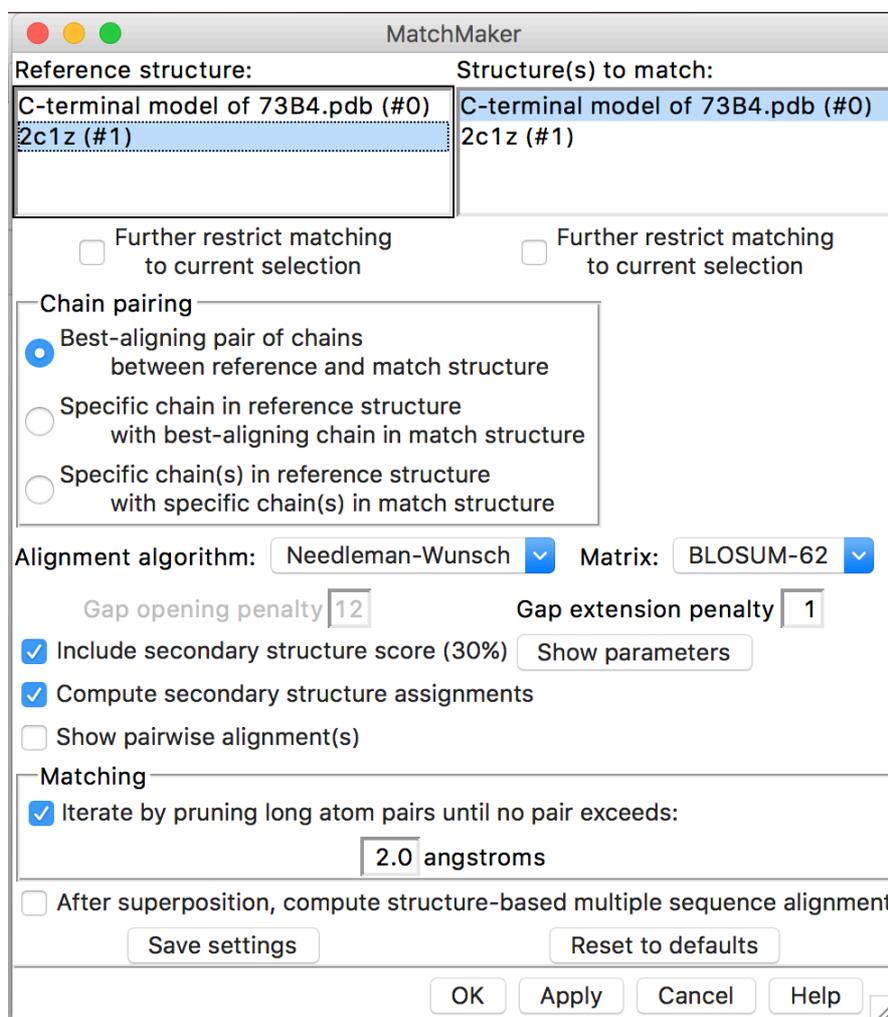


Figure 2.9: Parameter settings in ‘MatchMaker’

Using the same method, protein overlays of the model and template UGTs were generated to reveal possible interactions between the models and the substrates.

Chapter 3 *In vitro* characterisation of UGTs

As introduced in Chapter 1, plant UGTs catalyse the addition of a sugar moiety from sugar donors to sugar acceptors and are of significance in plant secondary metabolites. Owing to breakthroughs in whole genome sequencing, high-throughput screening and analysis of sequences, an increasing number of UGT genes have been identified^{105,106}. However, experimental data of their *in vitro* activities have not been fully obtained.

This chapter will describe the *in vitro* characterisation of recombinant UGTs from groups B, D, F, H and L of *Arabidopsis thaliana* (*At*), using mass spectrometry (MS)-based methods to investigate substrate specificity, kinetic parameters and catalytic mechanisms.

3.1 Method development of MS-based GT activity assay

The study of GT's glycosylation reactions was divided into qualitative and quantitative analyses, both achieved using MS. One advantage of MS is that it can conduct a label-free enzyme assay, allowing the rapid determination of substrate specificity and catalytic reaction rate for an enzyme with no influence on the reaction itself¹⁰⁷. This section describes the development of the MS-based method that was used to examine 1) enzyme substrate specificities and 2) full kinetics by monitoring the product formation qualitatively and quantitatively respectively. The general scheme of operation is shown in Figure 3.1, and details will be discussed in the following sections.

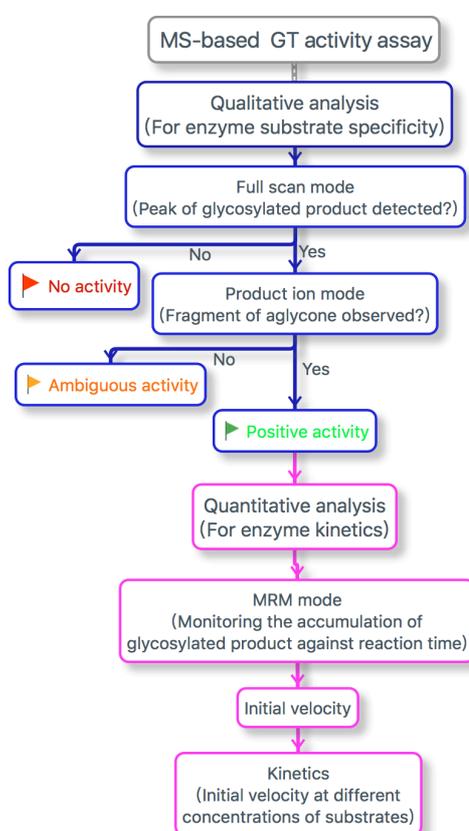


Figure 3.1: General scheme of the workflow of MS-based enzyme activity test.

3.1.1 Qualitative analysis for enzyme activity

Qualitative analysis was used to detect the formation of the desired glycosylated product against various substrates of interest. The positive activity of a UGT against a substrate was preliminarily identified by detecting the expected glycosylated product's mass ions (with appropriate modification by MS, e.g. $[M-H]^+$ in negative-ion mode) using the 'full scan' mode of MS. The formation of the potential glycosylated product was further confirmed by MS/MS analysis using the 'product ion' mode. The presence of the ions of the aglycone indicated a confirming rather than a false signal of the potential glycosylated product in the previous step.

As MS has unique requirements of experimental conditions, the conditions for the enzyme reaction and the MS analysis were not the same as other methods such as UV spectroscopy. It is therefore necessary to discuss the procedure of sample preparation and MS conditions in this research.

3.1.1.1 Sample preparation

Selection of buffer and its concentration

A buffer is important in maintaining the pH of a reaction. Common buffers used in previous literature for GT assays included Tris¹⁰⁸⁻¹¹², potassium phosphate¹¹³⁻¹¹⁶, PBS⁷³, TAPS¹¹⁷ and HEPES^{59,118}. In this project, Tris buffer was used and it has several advantages such as having good buffering capacity, low cost and most importantly is relatively MS friendly¹¹⁹. Although Tris buffer may be disadvantageous for some enzyme assays such as inhibiting certain enzymes (for example bimetallohydrolases¹²⁰ and *Bacillus licheniformis* alpha-amylase¹²¹) and incidentally destabilizing enzyme structure¹²², the literature indicated that Tris buffer is suitable for GT assays¹²³, and has been frequently used in GT assays.

Previous studies have shown that GTs can tolerate a wide range of pH. For example, one study showed that 74D1 was active from pH 6.0 to 9.0 (and even with a weak activity at pH 5.0), with its maximum activity was achieved in pH 6.0 (50 mM HEPES), followed by pH 7.0 (50 mM Tris or 50 mM MES)¹²⁴. Another study for the detoxification of 2,4,6-trinitrotoluene (TNT) in *At*, revealed that UGT73B4 and 73C1

(both of which were studied in this project) were active over the pH range of 6.5-8.5 either in Tris or in phosphate buffer⁵⁷.

Although buffer is indispensable for enzyme assays, it is not favourable for MS analysis. High concentrations of buffer and salts may exert adverse effects on MS performance such as making ion formation less reproducible and causing severe adduction or ion suppression¹²⁵. For example, non-volatile substances in buffers may not be fully evaporated by ESI and will add to the analyte that causes ion suppression¹²⁶. Additionally, the metal ions in salts may cause peak broadening with cluster and adduct formation¹²⁵. Hence, samples for MS are required to be desalted and the buffer must be changed in some cases¹²⁷. Alternatively, concentrations of buffer should be lowered to be MS friendly¹¹⁹.

From previous reports, a low concentration of buffer with low concentration of a metal cofactor (1 mM Tris, 1mM MgCl₂) was a compromise solution used for an enzyme reaction that was directly analysed by MS and used in many previous studies^{35,128-130}. Whilst the concentrations of Tris and metal were kept low, the conditions were sufficient to maintain a good environment for the enzyme reaction. In experiments in this section, a typical reaction contained 100 μM of substrate concentrations and this was 10% of the concentration of the buffer. If 100% product was produced, then 100 μM of H⁺ would be produced in the reaction and this would change the pH to 7.3 based on the Henderson-Hasselbalch equation ($\text{pH}=\text{pK}_a+\log\left(\frac{[\text{A}^-]}{[\text{HA}]}\right)$). Considering the turnover of reaction cannot be 100%, the buffer will be sufficient to support the reaction in a good environment during the reaction.

Temperature and reaction time

In general, enzyme activity changes with temperature. It was observed that the optimum temperature for enzyme activity is far from the physiological temperature of the plant source in some cases¹²². As a clear positive activity or inactivity was sufficient for qualitative analysis, 37°C was selected to increase the probability of detecting positive activity. This temperature is a common optimum temperature for plant GT enzymatic reactions^{131,132}, though, the ‘body temperature’ of plants rarely reaches 37°C except in desert and tropical environments.

The reaction was incubated overnight and this incubation time ensured that the enzyme had sufficient time to catalyse the reaction. A long-time of incubation has been used in many screening experiments previously^{73,115}. It is possible that the enzyme might become unstable or even degrade in such long-time incubation, leading to a decrease or loss of catalytic capability. However, the possible activity decrease of the enzyme was not the main concern in our qualitative analysis. The main concern of the qualitative analysis was whether the product could be formed rather than the amount or the rate of product formation. Hence, the possible activity decrease of the enzyme did not affect the judgement as to whether the product was formed.

Given the fact that the pH of Tris buffer is sensitive to temperature, the actual pH of the reaction might not be exactly 7.6, which was measured at 20°C, but may have approximately a 0.5 pH unit change; however, this change is still in the acceptable pH range for glycosylation.

Considering the above conditions, in this project, sample concentrations used for qualitative analysis were as follows: donor (100 μ M), acceptor (100 μ M), and enzyme in buffer (50 μ L, 1 mM Tris, 1 mM MgCl₂, pH 7.6). Samples were incubated at 37°C overnight and the reaction was followed by qualitative MS analysis.

3.1.1.2 MS conditions

After preparation of the conditions as detailed above, samples were examined by MS on whether there was a glycosylated product formed. Two modes of MS were performed: ‘full scan’ mode and ‘product ion’ mode.

‘Full scan’ mode was used to evaluate if glycosylated product was formed, by checking whether there was a peak observed with the corresponding exact molecular weight (MW) of glycosylated product and or its adduct minus one, as negative-ion mode was used. Negative-ion mode was used as the glycosylated product would readily give up a proton and therefore this mode was more sensitive than the positive-ion mode. ‘Full scan’ mode provided information about all ions in the mixture within the set detection range. The detection range should be large enough to include all possible ions. However, if the range was too large, the detection sensitivity would be lowered, since more ions need to be scanned per unit of time and the desired ions

would have less chance to accumulate in the detector. The scanning range for donor specificity screening was set at 100-1000 Daltons owing to the fact that with UDP-GlcNAc as the donor, doubly charged kaempferol-GlcNAc ($[2(\text{KMP-GlcNAc})-\text{H}^+]^+$), with a MW of 979 Da might be observed. The scanning range for acceptor specificity screening was set at 100-2000 Daltons, as the MW of product vancomycin-Glc is 1611.

'Product ion' mode, also known as MS/MS analysis, was then employed to fragment the potential glycosylated product. Since the newly-synthesised glycosidic bond was readily broken, the fragment with the MW of the aglycone minus one was expected. The collision energy should be suitable to break the newly synthesised glycosidic bond, thus confirming the glycosylated product. As this was a qualitative process, the collision energy was not optimised for each glycosylated product, but was set at 10 eV for all experiments in this mode. This level of collision energy ensured that the newly synthesised glycosidic bond would break, but was not so high that the glycosylated product would be fragmented into too small fragments that could not be attributed.

3.1.2 Quantitative analysis for enzyme activity

Quantitative analysis of enzyme activity delivers data of the initial rate of glycosylation at different substrate concentrations, from which the kinetics of an enzyme can be obtained. MRM mode was performed, which provided highly-targeted and accurate quantification analysis of the analyte. Two analytes were monitored: glycosylated product (Glycosylated-KMP) and internal standard quercetin (Que). The formation of glycosylated product was measured as a function of time. The details of the method development are discussed below.

3.1.2.1 MS conditions (MRM parameters set up)

To better monitor the Glycosylated-KMP and internal standard Que, their optimal MRM parameters were required, this included 1) optimal transition of analyte in MRM mode (precursor ions \rightarrow product ions) 2) optimal MS conditions applied (including fragmentor voltage (Frag), collision energy (CE) and acceleration voltage (CAV)).

A standard sample was prepared by mixing the glycosylated product and the internal standard. The software 'MassHunter Optimizer' automatically initiated the injection

of the sample, showing the abundance of analytes at different MS conditions. Each injection was performed in a certain MS condition, showing the abundance of analyte's transition in this MS condition. Finally, the optimal transition analyte was chosen (the one with highest abundance, highlighted characters in Table 3.1) and the corresponding MS condition was selected (highlighted characters in Table 3.1). These optimised MRM parameters were used to monitor analytes in the quantitative analysis.

Table 3.1: MRM optimiser results of Glycosylated-KMP (KMP-3-*O*-Glc, KMP-GlcNAc, KMP-Rha) and internal standard (Que). The optimal transition analyte and condition are highlighted in yellow (the one with highest abundance).

Compound	Exact MW (Dalton)	Precursor ion	Product ion	Frag (eV)	CE (eV)	CAV (eV)	Abundance (count)
KMP-3- <i>O</i> -Glc	448.1	447.1	284.1	235	33	4	142114
KMP-3- <i>O</i> -Glc	448.1	447.1	151.1	235	45	4	25369
KMP-3- <i>O</i> -Glc	448.1	447.1	107.1	235	57	4	11794
KMP-3- <i>O</i> -Glc	448.1	447.1	227.1	235	57	4	5576
KMP-GlcNAc	489.0	488.0	285.0	200	24	4	5520
KMP-GlcNAc	489.0	488.0	255.0	200	50	4	2063
KMP-GlcNAc	489.0	488.0	227.0	200	50	4	1871
KMP-GlcNAc	489.0	488.0	151.0	200	40	4	1238
KMP-Rha	432.0	431.0	285.0	200	16	4	13323
KMP-Rha	432.0	431.0	284.0	200	20	4	12757
KMP-Rha	432.0	431.0	255.0	200	36	4	12401
KMP-Rha	432.0	431.0	227.0	200	48	4	11221
Que	302.0	301.0	151.1	170	17	4	60052
Que	302.0	301.0	179.0	170	13	4	24715
Que	302.0	301.0	121.1	170	25	4	16953
Que	302.0	301.0	273.1	170	13	4	6243

3.1.2.2 LC condition

To measure the initial velocity of the enzymatic catalysed reaction, the quantity of Glycosylated-KMP was monitored as a function of time. Additionally, the quantity of the Que was also monitored, which was helpful to improve the accuracy and precision of the results in case that some unpredictable and uncontrollable errors of MS occurred¹³³. Que was used as the internal standard in this section, due to its structurally similar to KMP and its consistent and stable ionisation in MS.

A gradient condition was used as shown in Table 3.2. The sample volume injected was 10 μ L with a run time of 5.0 min. Glycosylated-KMP appeared at a retention time of \sim 1.4 min. The internal standard Que appeared at a retention time of \sim 2.6 min.

Table 3.2: Gradient elution of mobile phase

Time (min)	Water (+0.1% FA) / ACN %	Flow rate (mL/min)
0.0	70.0/30.0	0.05
1.0	55.0/45.0	0.05
2.5	55.0/45.0	0.05
5.0	70.0/30.0	0.05

3.1.2.3 Calibration curve

A calibration curve for Glycosylated-KMP was obtained to check the correlation of concentration and ion counts, using commercial KMP-3-O-glucoside as standard (Figure 3.2). Good linearity of KMP's concentration (in the range of 0.001 μM – 0.500 μM) and counts was achieved, with typical R^2 values of 9.98E-01 (with internal standard) and 9.85E-01 (without internal standard) for three runs. The data showed that the internal standard increased the precision and ability of the analytical method, by reducing the possible influence of the fluctuation of ionisation caused by the environment. With the internal standard, the calibration curve yielded a line with an average slope of 1.64 ± 0.03 and an intercept of 0.0004 ± 0.0050 , with a precision of 1.83%. This calibration curve provided evidence of the precision of measurement of the tested analyte.

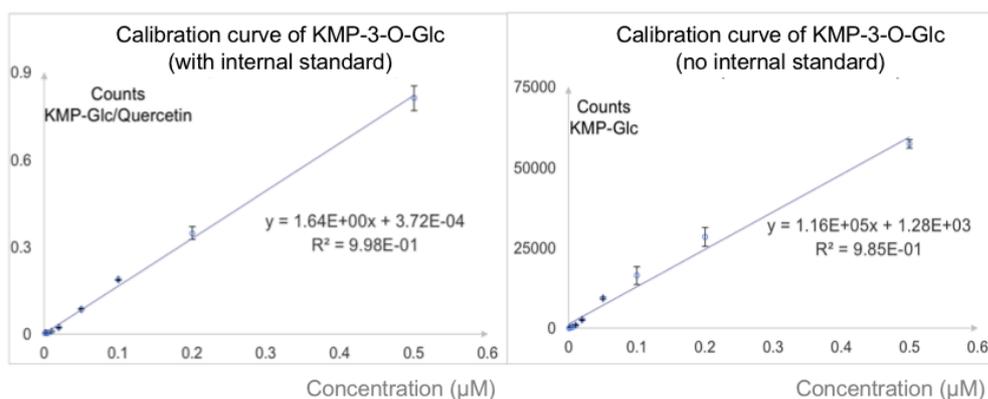


Figure 3.2: Calibration curve of commercial KMP-3-O-Glc (with or without internal standard). This calibration curve shows the evidence of precise measurement of KMP-3-O-Glc. The internal standard method increases the precision and ability of the analytical method.

However, MS is not a method without problems. For example, the system may drift from run to run¹³⁴. Ion suppression effects may happen in every run and may vary as ions are accumulated in the instrument gradually¹³⁵. Thus, the intra-day precision of MS may vary. To avoid this, a calibration curve was made along with each kinetic

measurement, thus ensuring the best correlation of the MS signal and the true concentration of products.

3.1.2.4 Kinetic study of UGTs

The LC-MS method developed above (Section 3.1.2.1 & 3.1.2.2) was used to quantitatively analyse enzyme kinetics by monitoring the formation of Glycosylated-KMP as a function of time with Que as an internal standard. The initial rate of each enzymatic reaction was obtained using the $d[\text{product}]/dt$ from the product formation curve at $t=0$ min. Mathematically, this method consists of approximating the slope (=first derivative) of the progressive curve at time $t=0$ by the average slope of the quasi-linear initial part¹³⁶. Below, a few factors such as the kinetic model, buffer, temperature, concentrations of the enzyme and substrates are discussed.

Kinetic model

GTs' Bi-Bi enzymatic reaction usually follows a Bi-Bi sequential mechanism, in which the EAB complex is formed in a catalytic process. This mechanism is either a Bi-Bi random sequential mechanism (if there is no compulsory order for substrates binding with enzymes) or a Bi-Bi ordered sequential mechanism (if there is a compulsory order for substrates binding with enzymes). A Bi-Bi Ping-pong mechanism is inferred when no substrate EAB complex is formed in the catalytic process, which is another possible but rather rare mechanism for GTs.

In Bi-Bi enzymatic reactions, no matter ordered or random sequential, the kinetic parameters can be calculated by the same equation:

$$v = \frac{V_{max}[A][B]}{K_{IA}K_B + K_B[A] + K_A[B] + [A][B]} \quad \text{Equation 3.1}$$

Where

K_{IA} : inhibitor parameter of enzyme-substrate A;

K_A : Michaelis parameter of substrate A; K_B : Michaelis parameter of substrate B;

V_{max} : maximum velocity reaction.

ordered mechanism. The relation of K_{IA} and K_A can be obtained not only from the kinetic parameter calculation but also from the pattern of reciprocal plots. If $K_A < K_{IA}$, the ratio of K_A/K_{IA} will be below 1, causing the lines of the reciprocal plots to intersect above x-axis, and vice versa (Figure 3.3).

Kinetic mechanisms and parameter calculations

Reciprocal plots provide a convenient way to distinguish the kinetic mechanism, as the family of plots intersects in the sequential mechanism and appears parallel in the Ping-pong mechanism.

However, the calculation of kinetic parameters from reciprocal form plots may not be precise as the reciprocal form model distorts the error structure of the data to some extent. For example, data points at high substrate concentrations are compressed into a small region, while data points at lower substrate concentrations are emphasised. Therefore, it is more precise to calculate the kinetic parameters by the original equation without reciprocal transformation (Equation 3.1).

Based on the Equation 3.1, there are two independent variables ([A] and [B]) and four parameters (K_A , K_B , K_{IA} , and V_{max}) that need to be calculated. Therefore, the parameter calculation concerns the interpolation and approximation of data in three dimensions. Consequently the surface fitting model was used to fit the kinetic model in this research¹³⁸.

Kinetics conditions

The buffer used for the kinetic study was the same as the qualitative analysis. However, the temperature used in this section was not 37°C, but room temperature instead (laboratory temperature controlled at 20°C by air conditioning). This is close to the plant's real environment, although plant UGTs probably have higher activity at 37°C^{131,132}. Room temperature was also frequently used in previous research for plant enzyme assays^{139,140}.

It should be noted that the use of room temperature means that the temperature is not vigorously controlled. Room temperature might be affected by many factors such as the condition of air conditioning, weather conditions and the room conditions. In order

to avoid the fluctuation, future optimisation includes a temperature control such as the use of a water bath to keep the temperature consistent.

Selection of enzyme concentration ([E])

A kinetic study of an enzyme requires a suitable and known [E] because 1) based on a kinetic model assumption, [E] should be kept as low as possible but remain within the detectable range¹²², and the kinetic parameter k_{cat} is calculated by dividing V_{max} by [E]; 2) high [E] leaves a limited time to observe a progressive curve (the curve of product quantity Vs time) with a clear linear part, which is key to the calculation of the initial rate. The reaction rate will be very fast with a high [E] and the reaction will enter the equilibrium phase within a very short time period or even at the onset of the experiment¹²². Therefore, [E] in a kinetic study should be carefully considered, to produce a significant progressive curve of the reaction with a clear linear part. Experiments were designed to select the most suitable [E] for every tested UGT. Herein, the selection of [E] of 78D2 with UDP-Glc and KMP is exemplified.

UDP-Glc (10 μ M), KMP (10 μ M) and 78D2 at a range of concentrations (0.0020, 0.0010 and 0.0002 mg/mL) were mixed and incubated in the buffer (1 mM Tris, 1 mM MgCl₂, pH 7.6). Aliquots were picked up at 5, 15, 30 and 50 min time intervals and treated as described in Chapter 2 to stop the reaction. Formation of the Glycosylated-KMP was quantified by the MRM mode. Progressive curves with different [E] values were then obtained (Figure 3.4).

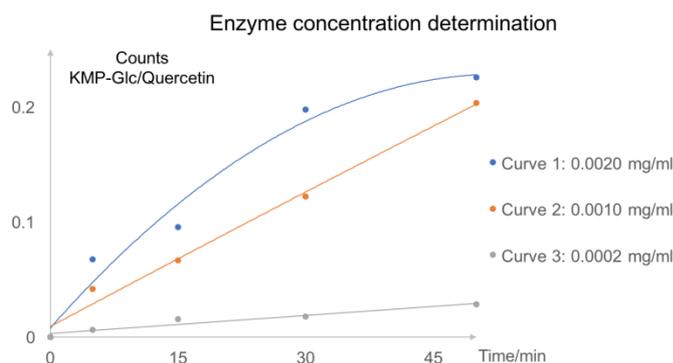


Figure 3.4: Selection of [E] in the kinetic study of 78D2 with UDP-Glc and KMP. Different concentrations of 78D2 were used: 0.0020 (curve 1), 0.0010 (curve 2) and 0.0002 (curve 3) mg/mL. Substrates were fixed at 10 μ M UDP-Glc and 10 μ M KMP. As this step was a quick test to show the [E] values required approximately, it was performed only once and therefore no error bars were displayed.

As Figure 3.4 shows, progressive curves with different shape profiles were observed at different concentrations of 78D2 at 0.0020, 0.0010 and 0.0002 mg/mL (curve 1, 2, and 3 respectively). Curve 1 shows the initial rate period (the linear part of progressive curve) lasted less than 30 mins, and was followed by a plateauing of the curve, which indicated that the reaction entered the equilibrium phase. Curve 2 and 3 did not reach a plateau within the monitoring period and good linear responses were recorded ($PMCC_{\text{curve 2}}=0.99$ and $PMCC_{\text{curve 3}}=0.96$).

For 10 μM UDP-Glc and 10 μM KMP, all three concentrations of 78D2 tested were suitable, as clear initial rate periods (the linear part of a progressive curve) were detected. However, a further full kinetic test was not limited to an $[\text{S}]$ of 10 μM , but ranged from low $[\text{S}]$ (e.g. 2 μM) to high $[\text{S}]$ (e.g. 50 μM). In this case, $[\text{E}]=0.0020$ mg/mL (curve 1) might be too high for experiments with a low $[\text{S}]$, as the reaction might happen so quickly and might enter the equilibrium phase so quickly even at the onset of the experiment. On the other hand, $[\text{E}]=0.0002$ mg/mL (curve 3) might be too low for experiments with a low $[\text{S}]$, as the reaction might happen so slowly that a low amount of product would be produced at the onset of the experiment, leading to a weak signal detection of the product. Besides, 0.0002 mg/mL (curve 3) might not be suitable for experiments with a high $[\text{S}]$, as such a low $[\text{E}]$ would probably become the reaction limiting factor, which would contradict the kinetic assumption that only $[\text{S}]$ correlated with the velocity rate. Hence, medium $[\text{E}]$ (0.0010 mg/mL, curve 2) was a good compromise to avoid possible problems and to be suitable in the full kinetic study of 78D2.

This step was carried out as a standard method to select the suitable $[\text{E}]$ for the kinetic study of every enzyme. The concentration used for every enzyme is summarised in Chapter 2.

Selection of substrate concentration ($[\text{S}]$)

Enzyme kinetics is described by the relationship between $[\text{S}]$ and its corresponding velocity. In a conventional enzyme assay, the selection of an $[\text{S}]$ range is an iterative process, with an optimal $[\text{S}]$ range covering 1/5-5 of the K_M value¹⁴¹. To find the first estimation, a wide range of $[\text{S}]$ values was used, and subsequent iterations were then performed, until an $[\text{S}]$ range of 1/5-5 K_M was covered. However, in a Bi-Bi kinetic

reaction, the first estimation is hard to achieved approximately through a simple experiment due to the complexity of the Bi-Bi kinetic model (there are 10 parameter rates in the Bi-Bi ordered sequential model (Appendix, Figure A.1) and 18 parameter rates in the Bi-Bi random sequential model (Appendix, Figure A.2)).

Consequently, an empirical estimation based on the previous literature was necessary. From previous literature, it was found that $[S]$ in the range of 1s-100s μM can cover K_M values of many UGTs^{59,73,108,110,111,113,116-119,142-144}. Hence, this $[S]$ range was used as a first estimation in this project. Below, the rationality of 1s-100s μM $[S]$ was examined from the aspects of the Bi-Bi kinetic model and MS practical considerations.

The $[S]$ cannot be too high

According to Equation 3.2 and Figure 3.3, the slope of reciprocal plots depends on the ratio of K_{IA} and $[A]$. If $[A]$ is much larger than K_{IA} , the variation in slopes of $1/v$ versus $1/[B]$ is non-responsive to a change in $[A]$ (Figure 3.5). Hence, for an enzyme with unknown kinetic parameters, $[S]$ should be kept low.

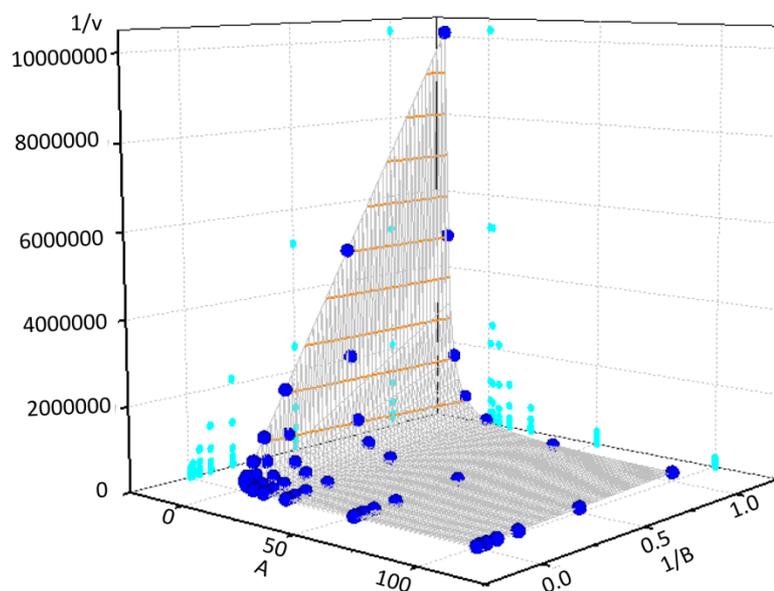


Figure 3.5: 3D plot of full kinetic simulation model ($K_A=10 \mu\text{M}$, $K_B=10 \mu\text{M}$, $K_{IA}=100 \mu\text{M}$, $V_{max}=1\text{E-}04 \mu\text{M}/\text{min}$). This figure shows a reciprocal form of kinetics with $[A]$ and $[B]$ vary. Contour amber lines show the major level of the z axis, indicating that $1/v$ Vs $1/B$ became non-responsive in a high $[A]$ range.

Apart from consideration of the kinetic model, the practical issues of ion suppression in the MS should be considered. Although an internal standard was added to minimise the impact of ion suppression, the impact of ion suppression might still have an adverse influence on the linear response of ESI-MS with increasing $[S]$ ¹⁴⁵.

In order to explore how ion suppression impacted the linear response of ESI-MS in this experiment, the counts of KMP-3-*O*-Glc/Que MRM were measured at different $[KMP-3-O-Glc]$ with different $[UDP-Glc]$ s added (Figure 3.6). It was observed that with an increasing $[UDP-Glc]$, both R^2 and PMCC in the curves that correlate MRM counts and concentration were decreased, indicating a decrease in a linear response of ESI-MS with increasing $[S]$ in our experiment.

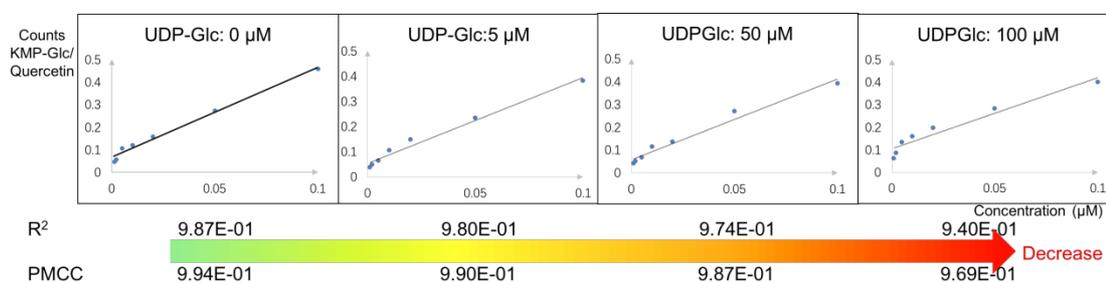


Figure 3.6: Correlation Curve of commercial KMP-3-*O*-Glc and counts KMP-Glc/Que. This series of curves show that linear response of ESI-MS gradually decreases with increasing $[UDP-Glc]$.

The $[S]$ cannot be too low

As discussed above, a high $[S]$ range was not recommended, to avoid the non-response of the reciprocal kinetic form and the linear response decrease of ESI-MS. However, $[S]$ cannot be too low, especially at the beginning of the reaction, as the product formation will fall below the detection limit of the MS. In the first few trials, we suggested keeping the $[S]$ above $1 \mu M$.

Taking all of these factors together, $[UDP-Glc]$ were set to 2, 5, 10, 20 and $50 \mu M$, while $[KMP]$ were set to 2, 5, 10 and $20 \mu M$. As discussed above, we considered that the empirical estimation of $[S]$ range was reasonable based on a previous study, the Bi-Bi kinetic model and practical considerations. Moreover, it had the advantage of simplifying the experimental design.

However, it cannot be ignored that an empirical estimation can bring some problems. Each enzyme has its own unique kinetic behaviour. And one enzyme may behave differently towards a different substrate. Consequently, if the same [S] range is used for all of the enzymes and all of the substrates, it may not be optimal for all, and the actual K_M value may drop out of the uniform [S] range. Specifically, the imperfect [S] range may bring two major problems: 1) the error in the kinetic calculation may be big, although a surface fitting model can still be applicable even if the K_M of the enzyme is out of the tested [S] range; 2) the reciprocal plots may not be clearly responsive, as K_{IA} of a different enzyme varies and results in $K_{IA}/[A]$ not being high enough to make a clear responsive correlation of $1/v$ Vs $1/[B]$.

With the disadvantage of an empirical estimation of [S] range in this step, the data quality of the enzyme varied. This will be further discussed in section 3.3 depending on the specific data of each enzyme.

Determination of enzyme full kinetics

This project provided the full kinetics of 73B4, 73B5, 73C1, 73C5, 73C6, 76E11, 76E12, 78D2 and 89B1 (acceptor: KMP; donor: UDP-Glc for all, UDP-Gal for 73C5 and 78D2, UDP-GlcNAc for 73B4). Herein, 78D2 glycosylation towards UDP-Glc is exemplified as a representative to illustrate the workflow in obtaining full kinetics results. The general procedure is summarised in Figure 3.7.

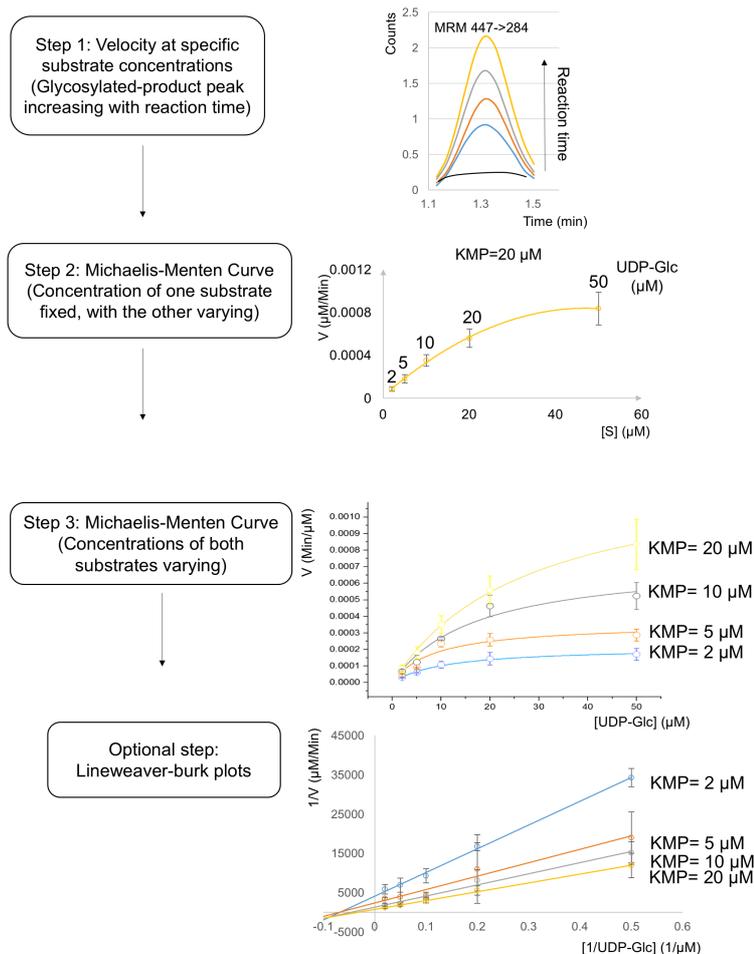


Figure 3.7: Workflow of 78D2 full kinetics (donor: UDP-Glc). 1) KMP-Glc was monitored in MRM mode (447→284) and quantified at a fixed concentration of acceptor and donor. As a function of reaction time, the initial rate of enzymatic velocity was calculated in these conditions; 2) with the data of initial rate, the Michaelis-Menten curve was obtained fixing KMP at 20 μM but varying UDP-Glc at 2, 5, 10, 20 and 50 μM ; 3) Michaelis-Menten curves of fixing KMP at 2, 5, 10 and 20 μM were obtained in the same way as above. If needed, the reciprocal form of the Michaelis-Menten curve (Lineweaver-burk plots) could be applied.

The full kinetics of 78D2 were measured with varying both concentrations of KMP and UDP-Glc. The experiment was set up by a fixed [KMP] at 2, 5, 10 and 20 μM , and there were five samples of each with a [UDP-Glc] of 2, 5, 10, 20 and 50 μM . In total, 20 reaction samples with a volume of 150 μL for each were assembled. Enzyme was added in the last step to facilitate the reaction. 25 μL of solution from each reaction sample was extracted at 5, 15, 30 and 50 minutes after the start of the reaction and was transferred to a new Eppendorf tube, then acetonitrile was added to precipitate the enzyme therefore stopping any further reaction.

KMP-Glc formation in all 80 samples (20 reaction samples x 4 time points) was monitored by the MRM mode. Figure 3.7 shows that the peak area of KMP-Glc

increased with a longer reaction time at specific [S]; thus, the velocity at this condition could be obtained with $(v=d[P]/dt)$ (Step 1). Subsequently, the Michaelis-Menten curve of the enzyme's velocity, with [KMP] fixed at $20\ \mu\text{M}$ against different [UDP-Glc] (varied at 2, 5, 10, 20 and $50\ \mu\text{M}$), was obtained (Step 2). Subsequently, the full set of kinetic results of the enzyme at different [KMP] was obtained following the same methodology (Step 3). Reciprocal plots provided a simple way to distinguish the enzymatic mechanism (Optional step): reciprocal plots intersected in the Bi-Bi sequential mechanism and were parallel in the Ping-pong mechanism.

3.2 Substrate specificity GAR screening

Characterisation of 29 purified recombinant UGTs from group B (89B1 and 89C1), D (73B4, 73B5, 73C1, 73C2, 73C3, 73C4, 73C5, 73C6, 73C7 and 73D1), F (78D1 and 78D2), H (76C2, 76C3, 76C4, 76C5, 76E3, 76E4, 76E5, 76E9, 76E11, 76E12, 76F1 and 76F2) and L (74C1, 75C1 and 75D1) was conducted, along with their substrate specificities in this section.

A sugar donor library was set up for donor specificity screening, containing eight nucleotide sugars (Figure 3.8): α -D-UDP-Glucose (UDP-Glc), α -D-UDP-Galactose (UDP-Gal), α -D-UDP-*N*-acetylglucosamine (UDP-GlcNAc), α -D-UDP-*N*-acetylgalactosamine (UDP-GalNAc), β -L-UDP-Rhamnose (UDP-Rha), β -L-dTDP-Rhamnose (dTDP-Rha), α -D-GDP-Glucose (GDP-Glc), and β -L-GDP-Fucose (GDP-Fuc). The diverse combination of glycosyl parts and nucleoside parts offered a broad scope for donor screening of UGTs.

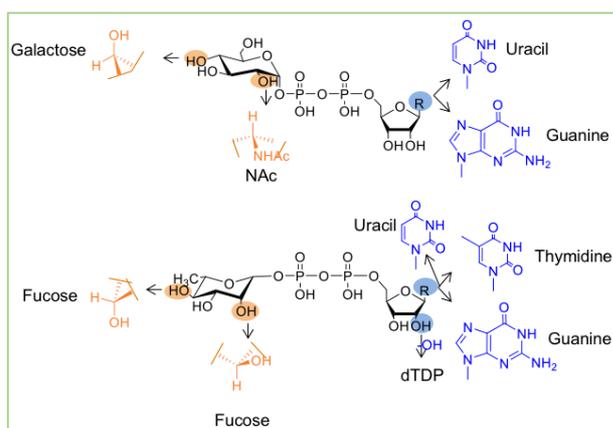


Figure 3.8: Summary of donor structures. (A) α -D form sugars: the core sugar structure is Glc. Gal differs in the $-OH$ orientation at C4 position of the Glc. GlcNAc replaces the $-OH$ at C2 position of the Glc with an $-NAc$ moiety. The nucleoside part can be replaced by UDP and GDP; (B) β -L form sugars: the core sugar structure is Rha. Fuc differs in the $-OH$ orientation at C2 and C4 positions of Rha. The nucleoside part can be replaced by UDP, dTDP and GDP.

The acceptor library mainly contained types of plant secondary metabolites, including anthraquinone, coumarin, alkaloid, saponin, flavonoid, cyanogenic and a plant hormone (structure shown in Chapter 2, Figure 2.7). These aglycones are frequently found as plant secondary metabolites and as hormone homeostasis metabolites. In this section, the experimental data are reported of UGTs' activities towards 48 acceptors

in vitro, using UDP-Glc as the donor (with the exceptions that 78D1 and 89C1 used UDP-Rha as donor).

3.2.1 Donor specificity

Of the 29 UGTs, twelve of these were shown to be GlcTs (73B4^{50,57,146}, 73B5^{50,147}, 73C1^{57,148}, 73C5^{50,148-150}, 73C6^{50,151}, 75C1^{152,153}, 75D1¹⁵⁴, 76C2¹⁴⁸, 76E11⁵⁰, 76E12⁵⁰, 78D2^{50,146,152}, and 89B1^{50,146}), whilst 78D1¹⁵¹ and 89C1¹⁵⁵ were RhaTs.

As discussed above, a sample was assembled with KMP (100 μ M), donor (100 μ M) and UGT (0.01~0.1 mg/mL) in the buffer (50 μ L, 1 mM Tris with 1 mM MgCl₂, pH 7.6) and incubated at 37°C overnight. Samples were analysed by ‘full scan’ mode and ‘product ion’ mode of MS, providing the results of positive/ambiguous/no activity towards the tested donor and UGT. Results are summarised and displayed as Green-Amber-Red, or ‘GAR’¹⁵⁶, denoted as the colour red (no activity), colour amber (ambiguous activity), and colour green (positive activity) (Figure 3.9).

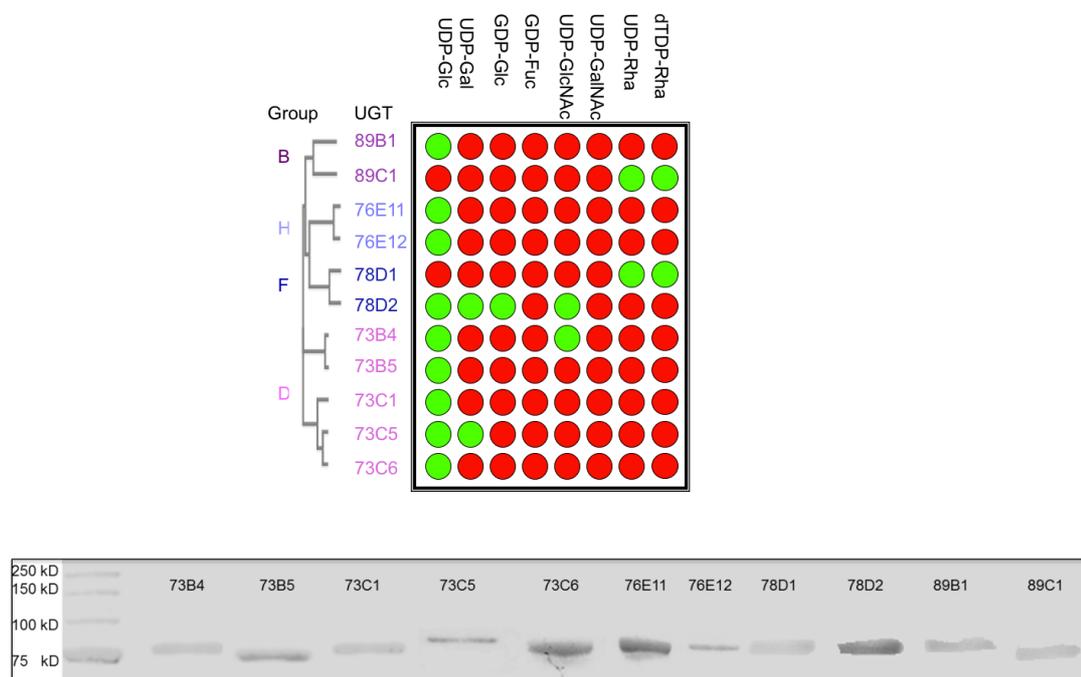


Figure 3.9: GAR screening result of donor specificity (KMP as the sugar acceptor) (top). The activity of UGTs is denoted as green (positive), amber (ambiguous) or red (no activity). The phylogenetic tree to the left reveals the origin and homology of the tested enzymes. SDS-PAGE of the 11 UGTs (purified UGTs display MW around 75 kD, with a 26 kD GST tag) is displayed to show the existence of protein (bottom).

Experimental data in this research revealed that UGT 73B4, 73B5, 73C1, 73C5, 73C6, 76E11, 76E12, 78D2 and 89B1 use UDP-Glc as their substrate, consistent with previous research. The confirmation of positive substrate specificity can be found in Appendix.

Other known GlcTs 75C1^{152,153}, 75D1¹⁵⁴, 76C2¹⁴⁸ did not show activity under current condition, probably because they could not use KMP as acceptor. Other UGTs (group D 73C2, 73C3, 74C4, 73C7 and 73D1; group H 76C2, 76C3, 76C4, 76C5, 76E3, 76E4, 76E5, 76E9, 76F1 and 76F2; group L 74C1, 75C1 and 75D1) did not show detectable glycosyl transfer activities in the current conditions.

In summary, the data shown in Figure 3.9 not only supports previous research, but also validates the methodology used. In addition, a broader donor range was revealed and is reported in this research for the first time: 73B4 and 78D2 with UDP-GlcNAc activity; 78D2 and 73C5 with UDP-Gal; 78D2 with GDP-Glc; 78D1 and 89C1 with dTDP-Rha.

From this result, it was observed that UDP-Glc is the most frequently used donor for the tested UGTs. This result is consistent with previous research which found that most plant UGTs favoured UDP-Glc as the donor, whereas mammalian UGTs often used UDP-Glucuronic acid¹⁵⁷. From the perspective of structure, the high selectivity and specificity of the enzymes may originate from their active sites¹⁵⁸. Thus, the substrate specificity can be a reflection of the flexibility of the active site of the enzyme. As suggested by evolutionary theory, the more promiscuous the active sites are, the more they facilitate change^{159,160}. In these results, 78D2 had the broadest range of donors among tested our UGTs. It can be suggested that the active site of 78D2 might be the most flexible in these enzymes, thus tolerating the most structural change and range of donors.

Apart from an overall reflection of the flexibility of the active site, the different structures of the donors revealed some important information about the binding sites of the enzyme. For example, in the crystal structures of β 4Gal-T1 (natural donor: UDP-Gal) complex with UDP-Glc and with UDP-Gal, a β 4Gal-T1+UDP-Gal+Mn²⁺ complex was observed pointing towards the C4 position of Gal and forming a with D318. However, the β 4Gal-T1+UDP-Glc+Mn²⁺ complex was observed pointing away

from the C4 position of Glc and no longer interacting with D318, due to the –OH orientation change at the C4 position¹⁶¹. This example revealed that the orientation of the –OH moiety at C4 position may influence the enzyme-substrate complex formation therefore influencing enzyme specificity. This is possibly why GlcTs cannot use UDP-Gal, even though there is only a subtle difference between Glc and Gal. Hence, according to the donor specificity results shown above, the active sites of 73B4, 73B5, 73C1, 73C6, 76E11, 76E12 and 89B1 are sensitive to the orientation of the –OH group at the C4 position, while the active sites of 73C5 and 78D2 can tolerate it. The tolerance of UDP-Gal for GlcTs was also observed in some studies, for example, ceramide GlcT showed dual activities of UDP-Glc and UDP-Gal, though the UDP-Gal activity was just 10% (k_{cat}/K_M) of that compared to UDP-Glc activity¹⁶².

Donor UDP-GlcNAc is a larger molecule than UDP-Glc, with an *N*-acetyl group rather than an –OH group at the C2 position. With the bulkier and more hydrophilic functional group, GlcNAc may make significant steric barrier in the place of Glc, thus most UGTs cannot accommodate UDP-GlcNAc and catalyse the reaction. In this project, only 73B4 and 78D2 showed UDP-GlcNAc activity, indicating that their active sites can tolerate the *N*-acetyl group at the C2 position of the sugar.

Apart from the sugar part, UGTs showed a high specificity towards UDP, as their names indicate. However, some UGTs also showed tolerance of other nucleosides. In this project, dTDP and GDP were tested. dTDP has subtle structural differences as compared to UDP: an additional CH₃ group attached at C5 of the uracil ring and the lack of an –OH at position C2' of the ribose. These slight structural changes may not make any significant steric barrier to dTDP in the place of UDP, and to support this, a crystal complex of the SpsA+dTDP complex showed an identical binding pattern to the SpsA+UDP complex¹⁶³. As experimental results showed, the compatibility of UDP and dTDP nucleosides was also suggested in the RhaTs 78D1 and 89C1. However, GDP was difficult for the UGTs to accommodate. GDP differs in many aspects compared to UDP, with an additional imidazole ring and an NH₂ replacement at C2 position of the nucleotide. With the big structural changes to the nucleotide, only 78D2 showed tolerance towards the GDP base, which indicated the possible flexibility of the nucleotide binding region in the active site of 78D2.

The correlation between the enzyme active site structure and function is not straightforward. However, the amino acid sequence and structure of the enzymes may provide some useful insights. This will be presented in the next Chapter, providing some possible explanations as to why different UGTs showed different donor specificity.

3.2.2 Acceptor specificity

3.2.2.1 Acceptor specificities – group D UGTs

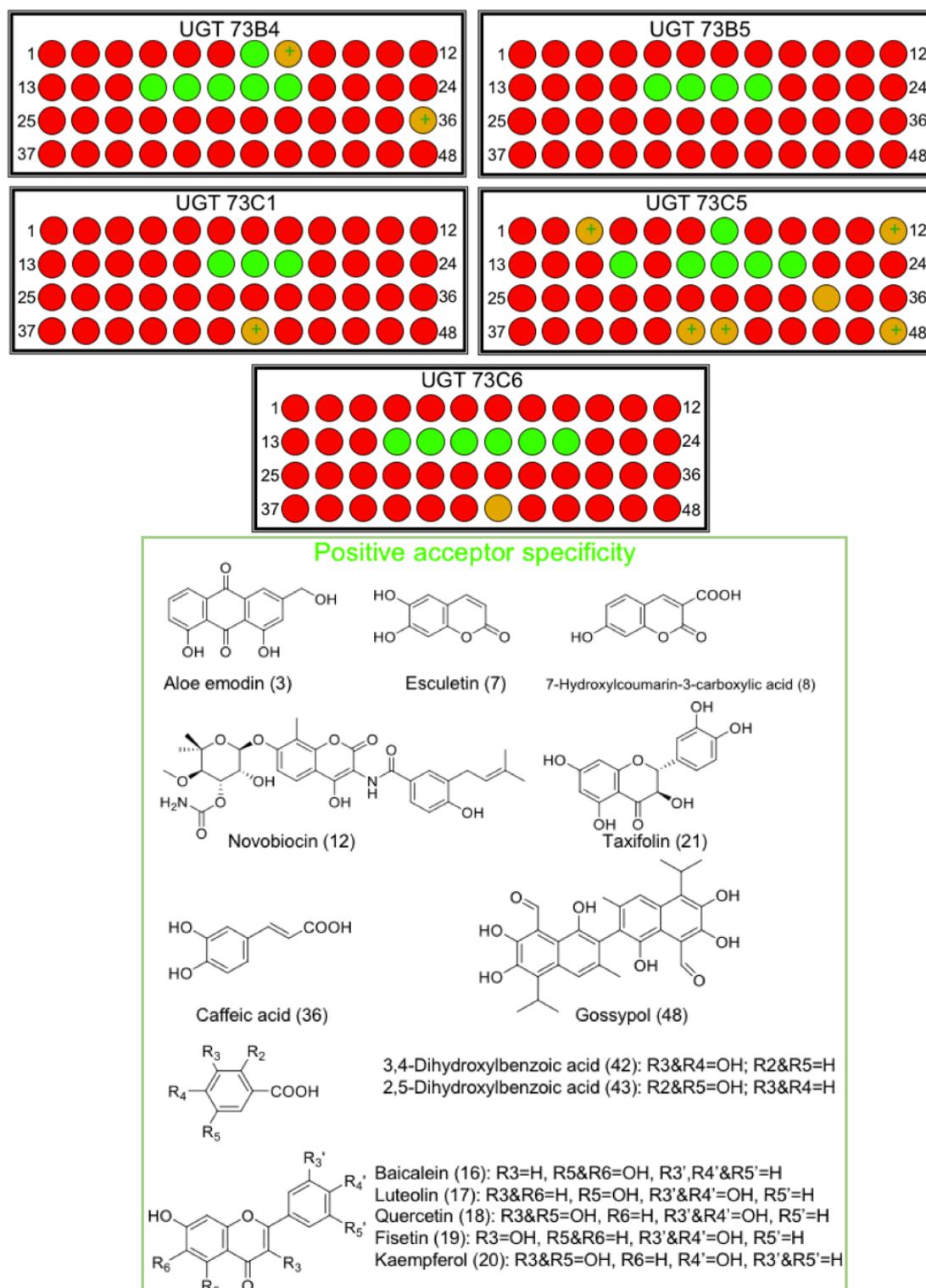


Figure 3.10: GAR screening results of acceptor specificities for the D group UGTs (UDP-Glc as donor) (top). The activity of UGTs is denoted as green (positive), amber (ambiguous) or red (no activity). A '+' in the amber circle means that the signal for the glycosylated product was very strong in the 'full scan' mode of MS, but was not detected in the 'product ion' mode of MS. The structures of the available acceptors are shown at the bottom.

From Figure 3.10, it was found that the group D UGTs can use many compounds and not only limited to flavonoids. Some new acceptor specificities of group D UGTs were discovered (with UDP-Glc as donor), including that 73B4 can use Esculetin, 7-hydroxycoumerin-3-carboxylic acid, and caffeic acid; 73C1 can use 2,5-dihydroxybenzoic acid with UDP-Glc as the donor; 73C5 had a broad acceptor specificity of aloe emodin, Esculetin, novobiocin (reported previously¹⁶⁴), gossypol, 3,4-dihydroxybenzoic acid and 2,5-dihydroxybenzoic acid.

Additionally, flavonoid activities were varied for different UGTs. There were six flavonoids tested in this study: baicalein (No. 16), luteolin (No. 17), Que (No. 18), fisetin (No. 19), KMP (No. 20), and taxifolin (No. 21). The tested group D enzymes showed selectivity towards the six flavonoids. For example, 73C1 only exhibited positive activity towards flavonoids with an –OH group at the C3 position (Que, fisetin and KMP), probably as a 3-O-flavonoid GlcT. In addition, the presence of chirality of the flavonoid was also important. Among the tested UGTs, only 73C5 and 73C6 could use taxifolin, indicating their tolerance towards the presence of a chiral centre in taxifolin.

3.2.2.2 Acceptor specificity screening – groups B, F, H and L UGTs

UGTs from four other groups were also examined using UDP-Glc as a donor (except 78D1 and 89C1, which used UDP-Rha) as the donor. Their activities are summarised below.

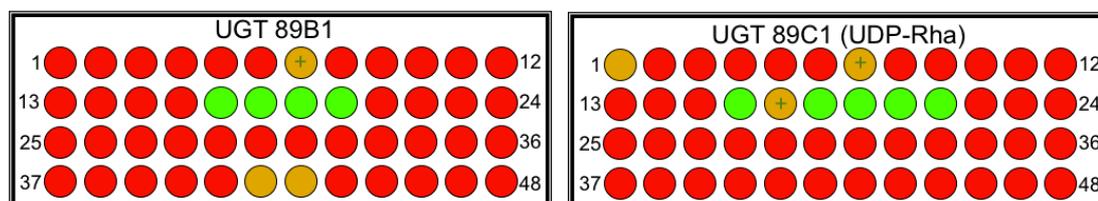


Figure 3.11: GAR screening results of the acceptor specificities for the B group UGTs (89B1 with UDP-Glc, 89C1 with UDP-Rha as donor). The activity of UGTs is denoted as green (positive), amber (ambiguous) or red (no activity). A '+' in the amber circle means that the signal for the glycosylated product was very strong in the 'full scan' mode of MS, but not detected in the 'product ion' mode of MS. Structures of the available acceptors are shown in Figure 3.17 bottom.

Group B 89B1 and 89C1 can also use Esculetin (Figure 3.11). Furthermore, enzymes showed selectivity towards flavonoids. 89B1 did not tolerate the presence of chirality in taxifolin, whilst 89C1 did.

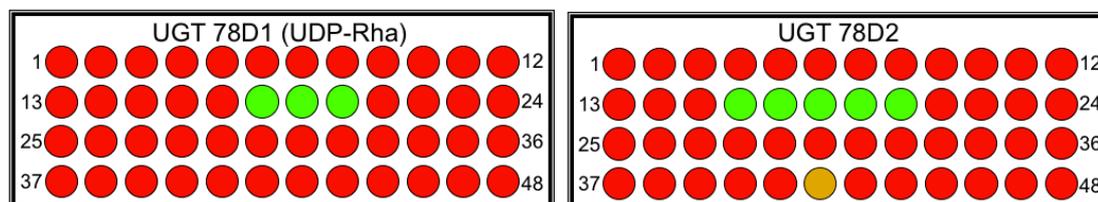


Figure 3.12: GAR screening results for the acceptor specificities for the F group UGTs (78D1 with UDP-Rha, and 78D2 with UDP-Glc as donor). The activity of the UGTs is denoted as green (positive), amber (ambiguous) or red (no activity). The structures of the acceptors are shown in Figure 3.17 bottom.

Group F 78D1 and 78D2 showed a narrow acceptor range (Figure 3.12). Both were sensitive towards the presence of chirality in taxifolin, as neither utilised taxifolin as a substrate in these conditions. 78D1 only exhibited positive activity towards flavonoids with an –OH group at the C3 position, possibly inferring that 78D1 is a 3-O-RhaT, which is consistent with previous research¹⁵¹.

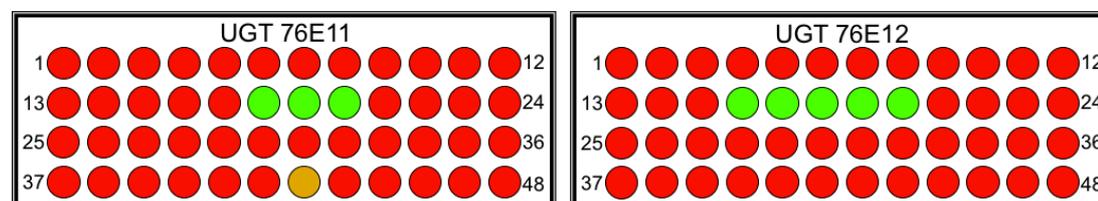


Figure 3.13: GAR screening results of acceptor specificities for H group UGTs (UDP-Glc as donor). The activity of UGTs is denoted as green (positive), amber (ambiguous) or red (no activity). Structures of the acceptors are shown in Figure 3.17 bottom.

Group H 76E11 and 76E12 also showed a narrow acceptor specificity (Figure 3.13). They could only use flavonoids as acceptors. Neither of them could tolerate the presence of chirality in taxifolin. 76E11 only displayed positive activity towards Que, fisetin and KMP, showing its possible characterisation as a 3-O-GlcT.

Apart from the enzymes discussed above, group D 73C2, 73C3, 74C4, 73C7 and 73D1, group H 76C2, 76C3, 76C4, 76C5, 76E3, 76E4, 76E5, 76E9, 76F1 and 76F2, and group L 74C1, 75C1 and 75D1 were examined but did not show detectable glycosyltransfer activities (donor: UDP-Glc) in the current conditions.

Generally, the same group of UGTs show a very high amino acid (AA) percentage sequence identity, implying that they may be evolved from the same ancestor and can be expected to have similar functions. However, prediction from AA sequence identity is not always reliable. Many UGTs have different acceptor specificity systems even though their AA sequences are highly similar.

For example, there are four UGTs belonging to group B (89A1P is the putative glycosyltransferase pseudogene) with an AA identity of approximately ~40%. Based on the phylogenetic tree, they are inferred to have evolved from the same ancestor and this would suggest a similar substrate specificity. This inference is true for 89B1 and 89C1, which both had flavonoid activity. However, it was not the case for 89A2 in a previous study, which only used benzoates and not flavonoids¹⁴⁶. These results suggested that the acceptor specificities for group B UGTs have commonalities to some degree, but vary depending on the specific enzyme.

As mentioned above, an enzyme's substrate specificity can originate from the active site of the enzyme. In UGTs, the acceptor binding region is located in the N-terminal domain and possesses relatively higher variations of AA sequence than that in the donor binding region (located in the C-terminal domain)¹⁹. Different UGTs have different AAs in their active sites to interact with different acceptors. Moreover, one UGT may have different AA in its active site to interact with similar acceptors. The comparison of crystal structures of the 72B1 complex with KMP (PDB: 4REL) and delphinidin (PDB: 4REM) is a good example to illustrate this phenomenon. Compared to KMP, delphinidin has additional –OH groups at C3' and C5' of ring B, but lacks a carbonyl group at C4 of ring C. These structure changes can shift their accommodation in 72B1, subsequently affecting some variations on specific sites: P78 bound to the –OH group at C4' of delphinidin, but not to the –OH group at C4' of KMP; D181 interacted with the –OH group at C7 of both delphinidin and KMP with different rotations that might result in a difference in the strength of the binding effect. This example clearly illustrated that different molecules have different ways of interacting with the same enzyme even though they have high similarities in structure. Consequently, any change of the acceptor structure may lead to the change of interaction and consequently affect the observed catalytic activity. This phenomenon was observed in the experimental data regarding the selectivity of flavonoids in this

section. Five flavonoids were tested and showed differential specificity towards different enzymes. This may be explained by differences in the active site of the specific enzyme from the perspective of a structure-based enzyme study.

3.3 Multi-substrate enzyme kinetics

In this part, full kinetics of UGTs were examined under varying concentrations of both substrates. Compared to a simplified pseudo-Uni kinetic model, a more detailed picture of the reaction kinetics could be obtained.

The full kinetics of nine active UGTs (73B4, 73B5, 73C1, 73C5, 73C6, 76E11, 76E12, 78D2, and 89B1) were examined, with the acceptor as KMP and corresponding donors.

3.3.1 Full kinetic study – Group D UGTs

The reciprocal forms of plots are displayed to distinguish between sequential and Ping-pong mechanisms (Figures 3.14-3.20). The kinetic parameters were calculated using Equation 3.1 by surface fitting model (Figure 3.21).

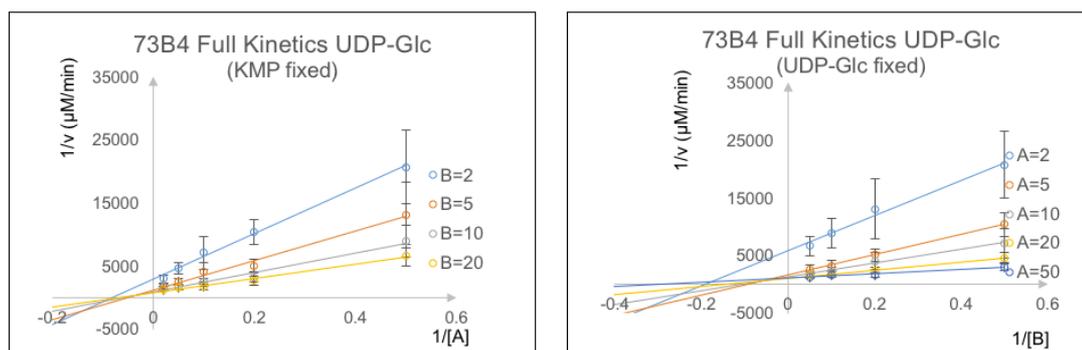


Figure 3.14: Reciprocal form plots of 73B4 full kinetics (donor: UDP-Glc). Left: concentration of KMP (B) is fixed at 2, 5, 10, and 20, with UDP-Glc (A) varying at 2, 5, 10, 20, and 50 μM . Right: concentration of UDP-Glc (A) is fixed, with KMP (B) varying. Their concentration range is the same as on the left.

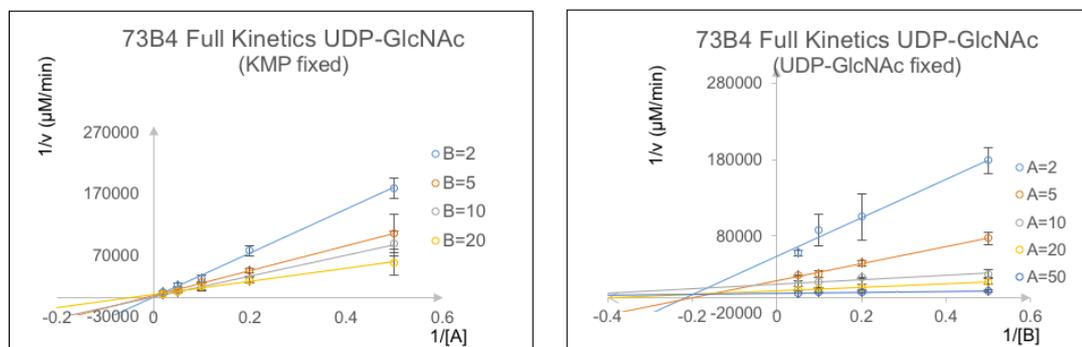


Figure 3.15: Reciprocal form plots of 73B4 full kinetics (donor: UDP-GlcNAc). Left: concentration of KMP (B) is fixed at 2, 5, 10, and 20, with UDP-GlcNAc (A) varying at 2, 5, 10, 20, and 50 μM . Right: concentration of UDP-GlcNAc (A) is fixed, with KMP (B) varying. Their concentration range is the same as on the left.

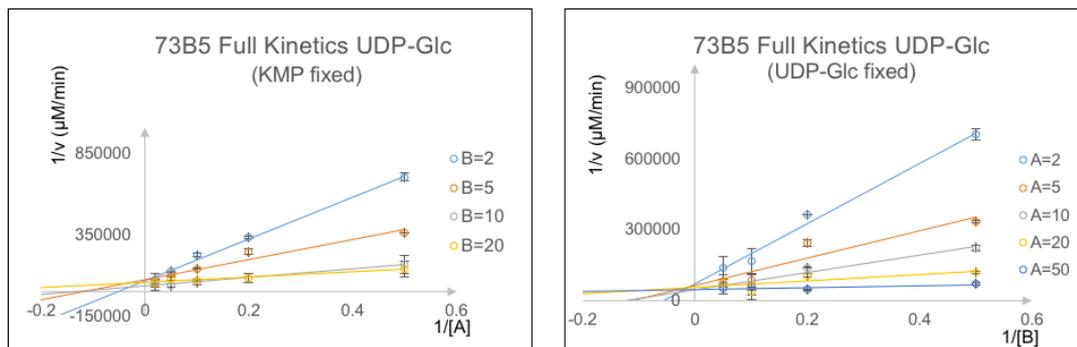


Figure 3.16: Reciprocal form plots of 73B5 full kinetics (donor: UDP-Glc). Left: concentration of KMP (B) is fixed at 2, 5, 10, and 20, with UDP-Glc (A) varying at 2, 5, 10, 20, and 50 μM . Right: concentration of UDP-Glc (A) is fixed, with KMP (B) varying. Their concentration range is the same as on the left.

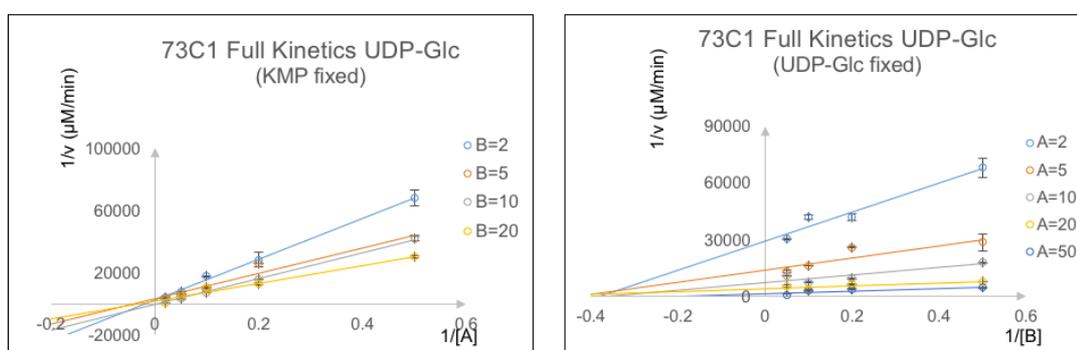


Figure 3.17: Reciprocal form plots of 73C1 full kinetics (donor: UDP-Glc). Left: concentration of KMP (B) is fixed at 2, 5, 10, and 20, with UDP-Glc (A) varying at 2, 5, 10, 20, and 50 μM . Right: concentration of UDP-Glc (A) is fixed, with KMP (B) varying. Their concentration range is the same as on the left.

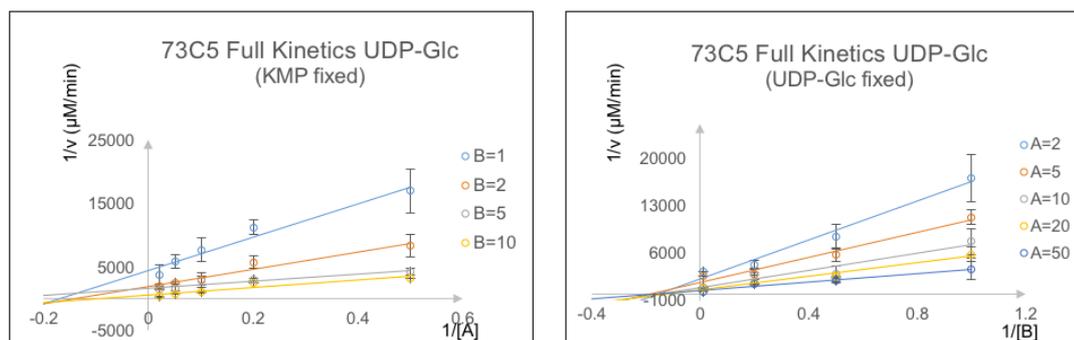


Figure 3.18: Reciprocal form plots of 73C5 full kinetics (donor: UDP-Glc). Left: concentration of KMP (B) is fixed at 1, 2, 5, and 10, with UDP-Glc (A) varying at 2, 5, 10, 20, and 50 μM . Right: concentration of UDP-Glc (A) is fixed, with KMP (B) varying. Their concentration range is the same as on the left.

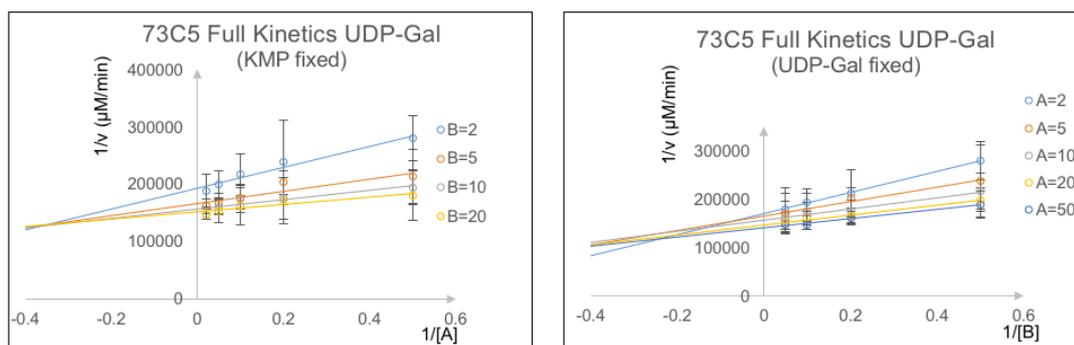


Figure 3.19: Reciprocal form plots of 73C5 full kinetics (donor: UDP-Gal). Left: concentration of KMP (B) is fixed at 2, 5, 10, and 20, with UDP-Gal (A) varying at 2, 5, 10, 20, and 50 μM . Right: concentration of UDP-Gal (A) is fixed, with KMP (B) varying. Their concentration range is the same as on the left.

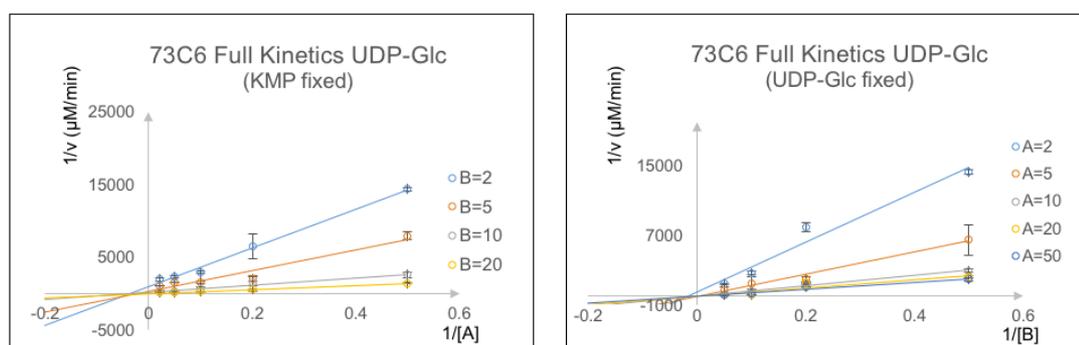


Figure 3.20: Reciprocal form plots of 73C6 full kinetics (donor: UDP-Glc). Left: concentration of KMP (B) is fixed at 2, 5, 10, and 20, with UDP-Glc (A) varying at 2, 5, 10, 20, and 50 μM . Right: concentration of UDP-Glc (A) is fixed, with KMP (B) varying. Their concentration range is the same as in the left.

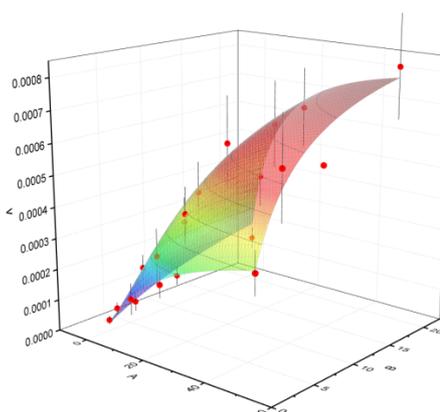


Figure 3.21: 3D plot of full kinetics of 73B4 towards UDP-Glc. This Figure shows that the two concentrations of donor (A) and acceptor (B) vary. Full kinetic parameters are calculated by a nonlinear regression analysis that allows simultaneous fitting of the data. This is representative of the surface fitting model.

Table 3.3: Kinetic parameters of Group D UGTs.

A: donor; B: acceptor (KMP)

* means the data not converged in surface modelling, and their parameters were obtained from pseudo single calculation.

Donor	UGT	K_A (M)	K_B (M)	K_{IA} (M)	V_{max} (M/S)	k_{cat} (s ⁻¹)	k_{cat}/K_A (s ⁻¹ M ⁻¹)
UDP-Glc	73B4	3.1E-06 ±1.9E-06	1.8E-06 ±9.5E-07	4.8E-05 ±3.1E-05	1.6E-11 ±1.4E-12	3.8E-04 ±3.2E-05	1.2E+02 ±7.6E+01
	73B5	4.0E-06 ±2.4E-06	1.5E-06 ±5.3E-07	2.4E-05 ±1.4E-06	4.5E-13 ±5.5E-14	1.8E-05 ±2.2E-06	4.5E+00 ±2.8E+00
	73C1*	2.0E-06 ±2.5E-06	4.7E-05 ±5.2E-05	-	7.3E-12 ±3.6E-12	3.0E-04 ±1.5E-04	1.5E+02 ±2.0E+02
	73C5	3.9E-06 ±5.0E-06	2.0E-06 ±1.2E-06	1.6E-05 1.3E-05	1.9E-11 ±4.3E-12	7.7E-04 ±1.8E-04	2.0E+02 ±2.6E+02
	73C6*	2.2E-05 ±1.1E-05	1.1E-04 ±1.2E-04	-	5.7E-10 ±4.0E-10	2.3E-02 ±1.6E-02	1.0E+03 ±9.4E+02
UDP-Gal	73C5	4.1E-07 ±1.3E-07	6.0E-07 ±1.2E-07	3.3E-06 ±1.5E-06	1.1E-13 ±2.5E-15	1.6E-06 ±3.6E-08	4.0E+00 ±1.3E+00
UDP-GlcNAc	73B4*	1.1E-06 ±1.5E-07	4.3E-05 ±6.8E-06	-	4.3E-12 ±2.1E-13	4.9E-05 ±2.4E-06	4.6E+01 ±6.7E+00

The families of reciprocal plots intersect in both conditions that donor concentration was fixed (acceptor concentration was varied) and acceptor concentration was fixed (donor concentration was varied). These results reveal that group D UGTs follow a Bi-Bi sequential mechanism, ruling out the possibility of a Ping-pong mechanism.

UDP-Glc activity in Group D UGTs

As introduced in Chapter 1, K_M values are the concentration of substrate at which half the active sites are filled and the requirement of significant catalysis to occur⁶⁵. Thus, K_A and K_B of UGTs imply the substrate concentration required for significant glycosylation to occur. The K_A of group D UGTs varies from 2 to 22 μ M and K_B varies from 1.8 to 110 μ M. There are indeed big differences between individual GTs although they belong to the same group. From the K_M data shown in Table 3.3, 73C1 requires the smallest concentration (2.0E-06 M) of UDP-Glc for the significant catalysis, while of 73C6 requires 11-fold higher concentration (2.2E-05 M) of UDP-Glc. For acceptor KMP, 73B4 requires the least (1.8E-06 M) while 73C6 needs 61-fold higher concentration (1.1E-04 M).

This fact more or less reflects the preference/sensitivity of an enzyme towards certain substrate, which can be a reference to some applications such as chemo-enzymatic synthesis. If the kinetic data are used to speculate enzyme *in vivo* activity, some

considerations are required. For example, although 73B4 seems to facilitate glycosylation easier than 73C6 from the aspect of K_M , it is not reasonable to infer that 73B4 plays a better or more important role *in vivo* than 73C6 as they are expressed in different tissue (73B4 is expressed in roots¹⁴⁷ while 73C6 was expressed in leaves and flowers¹⁵¹). To make the most use of kinetic data, specific considerations of further applications are required.

In addition, K_M values of UGTs can be used to suggest the Bi-Bi mechanism, i.e., whether it belongs to Bi-Bi ordered or Bi-Bi random mechanism¹¹⁹. A literature survey suggests that the K_M ratio [K_B/K_A (when $K_B > K_A$) or K_A/K_B ($K_A > K_B$)] can be used as an evidence to identify random or ordered sequential mechanism. Random sequential displayed a ratio in the range 1.1~7.5, while ordered sequential showed a ratio in the range 7.4~485¹¹⁹. This method can be employed for a quick distinction between random and ordered mechanisms. Alternatively, product inhibition can be used, which is effort and chemical consuming.

To validate the reliability of the method using K_M ratio to distinguish Bi-Bi ordered/random mechanism, traditional product inhibition assay of 73C5 was performed (Figure 3.22).

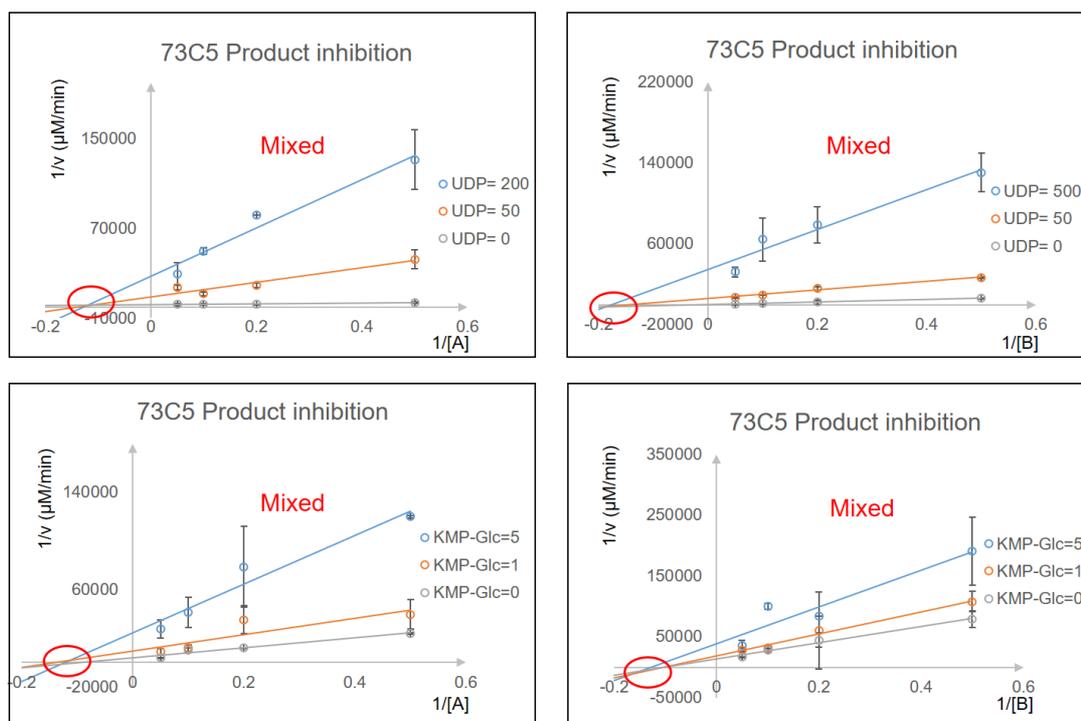


Figure 3.22: Product inhibition of 73C5. Mixed inhibition patterns are observed.

Product inhibition pattern of 73C5 showed all mixed patterns (Figure 3.22), indicating a Bi-Bi random sequential mechanism. This complies with the fact that the K_M ratio of 73C5 is 1.9 (± 2.7) that is in the range of K_M ratio range of Bi-Bi random mechanism. This experiment indicates that K_M ratio is a convenient, reliable tool to determine the Bi-Bi mechanism.

Following the same method, 73B4, 73B5, and 73C6 towards UDP-Glc are also suggested to follow a Bi-Bi random sequential mechanism as well, with K_M ratio of 1.7 (± 1.4), 2.7 (± 1.9) and 3.8 (± 6.3) respectively.

The turnover of an enzyme can also be reflected from the parameters such as V_{max} and k_{cat} ⁶⁵. It is the ability of an enzyme to convert the number of substrate molecules into product in a unit of time (typically second) in a saturated condition. However, V_{max} is not frequently used as it is [E] dependent. Hence, k_{cat} is introduced as it excludes the influence of enzyme concentration (by dividing V_{max} to [E]). The values of k_{cat} show that group D UGTs have different catalytic capability: 73B5 has the lowest catalytic capability ($1.8\text{E-}05\text{ s}^{-1}$) while 73C6 has the highest ($2.3\text{E-}02\text{ s}^{-1}$), which is 1300-fold higher.

k_{cat}/K_M is typically used to measure the catalytic efficiency⁶⁵ or catalytic performance towards certain substrate¹⁶⁵. Using this criterion, 73C6 has the highest catalytic efficiency/performance using UDP-Glc ($1000\text{ s}^{-1}\text{M}^{-1}$), while 73B5 has the lowest catalytic efficiency/performance ($4.5\text{ s}^{-1}\text{M}^{-1}$) that is 222-fold smaller than 73C6.

The K_{IA} of 73B4, 73B5 and 73C5 are larger than their corresponding K_A . The fact that $K_{IA} > K_A$ indicates that the complex of donor-enzyme may promote KMP subsequent binding (as discussed in Section 3.3.1), which can be called as ‘donor-driven promotion of acceptor binding’¹⁶⁶ that generally found in GT1 enzymes.

UDP-Glc and UDP-GlcNAc activity of 73B4

In addition to UDP-Glc, 73B4 also uses UDP-GlcNAc. K_A comparison between UDP-GlcNAc and UDP-Glc shows that a higher concentration of UDP-GlcNAc is required to facilitate significant catalysis (a 2.8-fold difference). The k_{cat} suggests that the catalytic capability of 73B4 towards UDP-Glc is 7.8-fold higher than that towards

UDP-GlcNAc. The K_{cat}/K_A highlights that 73B4 has a 2.6-fold greater catalytic efficiency/performance towards UDP-Glc than UDP-GlcNAc. Taking all parameters together, UDP-Glc is indicated as a better substrate for 73B4 compared to UDP-GlcNAc.

As for kinetic mechanism, it is suggested that 73B4 may facilitate UDP-GlcNAc transfer by the Bi-Bi ordered sequential mechanism, with K_M ratio of 47 (± 18). This is different from the kinetic mechanism by which 73B4 catalyses UDP-Glc, showing that 73B4 may follow a donor-dependent kinetic mechanism.

In previous studies, *O*-GlcNAc transferases (OGTs) were proposed to transfer UDP-GlcNAc in a Bi-Bi ordered sequential mechanism as well. They were initially predicted to follow a Bi-Bi ordered sequential mechanism via kinetic experiments, with K_M ratio of 18 ($K_{UDP-GlcNAc} = 6.0 \mu\text{M}$ and $K_{CKII} = 107 \mu\text{M}$)¹⁶⁷. This prediction from the perspective of kinetics was approved by the structure study with a substrate-enzyme complex in catalytic process by a ternary complex crystal structure of human OGT-UDP-CKII^{168,169}. UDP was bound firstly to human OGT, followed by its α -phosphate binding to the backbone amide of serine of CKII, thereby aligning CKII in the correct position. Once CKII was attached to the binding pocket, access to the active site was blocked. After the formation of the ternary substrate complex, the side chain of CKII attacked the anomeric carbon and formed a β -glycosidic linkage. The glycopeptide product was then released, followed by UDP.

In this research, 73B4 is also suggested to follow a Bi-Bi ordered sequential mechanism. The same Bi-Bi ordered mechanism might suggest the co-evolution relationship of UGTs³⁶ that with the same function (UDP-GlcNAc) but from different origins (plant and human).

UDP-Glc and UDP-Gal activity of 73C5

73C5 has dual donor activities towards UDP-Glc and UDP-Gal. K_A reveals that 73C5 requires an 11-fold higher concentration of UDP-Gal than UDP-Glc to facilitate significant catalysis. The values of k_{cat} and k_{cat}/K_A indicate that 73C5 uses UDP-Glc with a 480-fold higher catalytic capability and 50-fold higher catalytic

efficiency/performance than that of UDP-Gal. These experimental data indicate that UDP-Glc acts as a better donor candidate for 73C5 than UDP-Gal.

As for the kinetic mechanism, it is suggested that 73C5 uses the both donors in Bi-Bi random sequential mechanism with K_M ratio towards UDP-Glc (1.9 (± 2.7)) and UDP-Gal (1.5 (± 0.5)).

As discussed above, the relation between K_{IA} and K_A often suggests whether the already-bound substrate promotes or prevents the binding of the co-substrate. The values of K_{IA} of both UDP-Glc and UDP-Gal are larger than their corresponding values of K_A . Consequently, it is suggested that when the donor binds the enzyme, a lower affinity is required for the acceptor to bind the free enzyme.

3.3.2 Full kinetic study – Group B, F and H UGTs

UGTs from group B (89B1), H (76E11 and 76E12), and F (78D2) were also examined using the same method. The results are displayed in Figures 3.23-3.27.

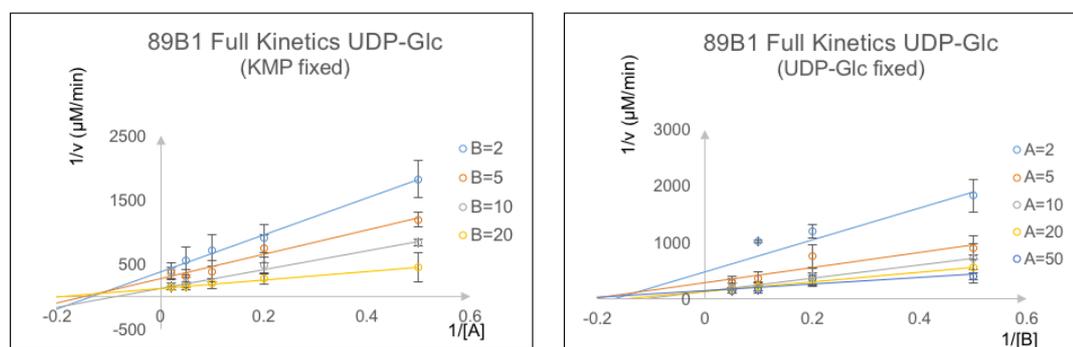


Figure 3.23: Reciprocal form plots of 89B1 full kinetics (donor: UDP-Glc). Left: Concentration of KMP (B) is fixed at 2, 5, 10, and 20, with UDP-Glc (A) varying at 2, 5, 10, 20, and 50 μM . Right: Concentration of UDP-Glc (A) is fixed, with KMP (B) varying. Their concentration range is the same as on the left.

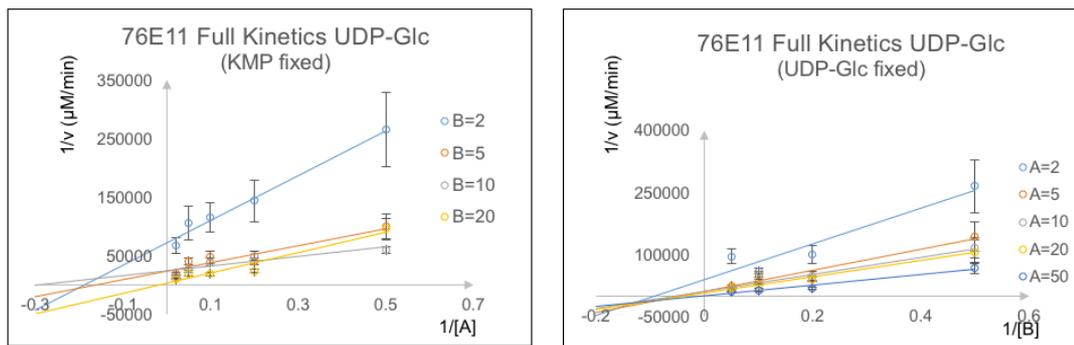


Figure 3.24: Reciprocal form plots of 76E11 full kinetics (donor: UDP-Glc). Left: Concentration of KMP (B) is fixed at 2, 5, 10, and 20, with UDP-Glc (A) varying at 2, 5, 10, 20, and 50 μM . Right: Concentration of UDP-Glc (A) is fixed, with KMP (B) varying. Their concentration range is the same as on the left.

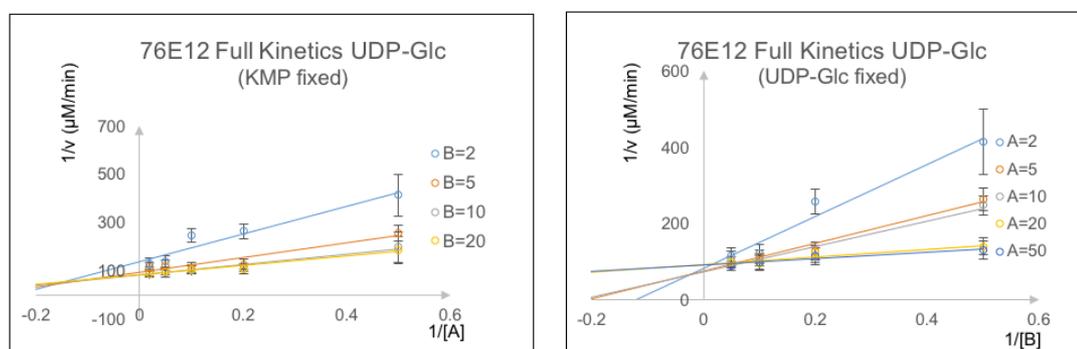


Figure 3.25: Reciprocal form plots of 76E12 full kinetics (donor: UDP-Glc). Left: Concentration of KMP (B) is fixed at 2, 5, 10, and 20, with UDP-Glc (A) varying at 2, 5, 10, 20, and 50 μM . Right: Concentration of UDP-Glc (A) is fixed, with KMP (B) varying. Their concentration range is the same as on the left.

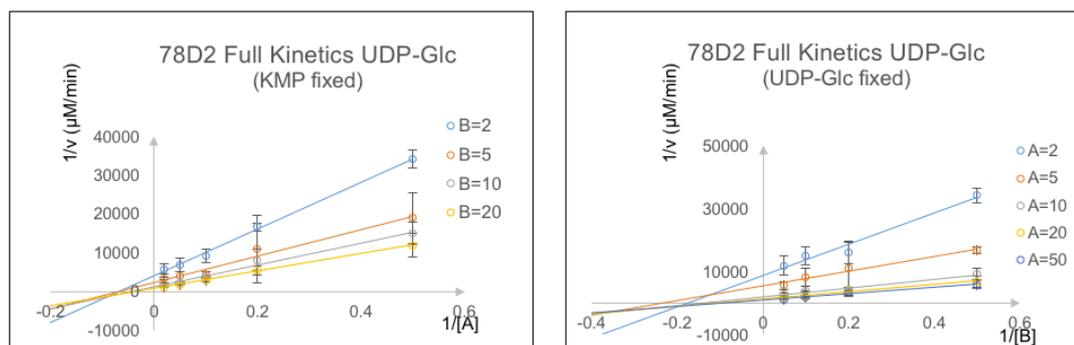


Figure 3.26: Reciprocal form plots of 78D2 full kinetics (donor: UDP-Glc). Left: Concentration of KMP (B) is fixed at 2, 5, 10, and 20, with UDP-Glc (A) varying at 2, 5, 10, 20, and 50 μM . Right: Concentration of UDP-Glc (A) is fixed, with KMP (B) varying. Their concentration range is the same as on the left.

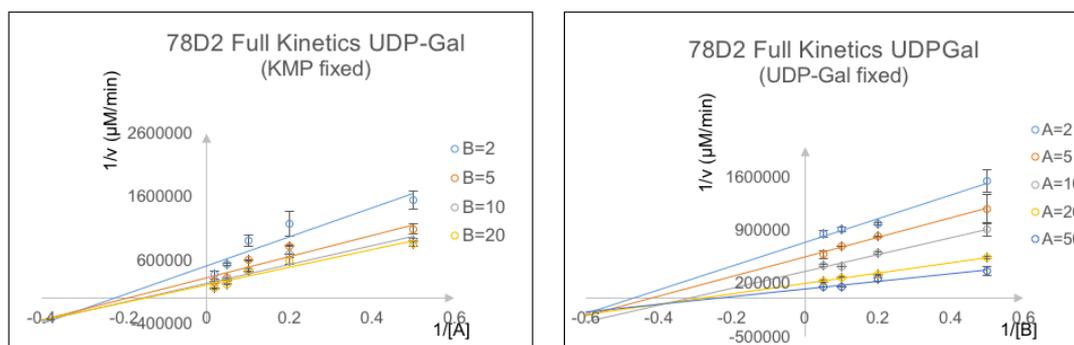


Figure 3.27: Reciprocal form plots of 78D2 full kinetics (donor: UDP-Gal). Left: Concentration of KMP (B) is fixed at 2, 5, 10, and 20, with UDP-Gal (A) varying at 2, 5, 10, 20, and 50 μM . Right: Concentration of UDP-Glc (A) is fixed, with KMP (B) varying. Their concentration range is the same as on the left.

Table 3.4: Kinetic parameters of group B, H and F

A: donor; B: acceptor (KMP)

* means the data is not converged in surface modelling, and their parameters were obtained from pseudo-Uni calculation

Donor	UGT	K_A (M)	K_B (M)	K_{IA} (M)	V_{max} (M/S)	k_{cat} (s^{-1})	k_{cat}/K_A ($s^{-1}M^{-1}$)
UDP-Glc	76E11	4.9E-06 $\pm 9.8E-06$	6.3E-06 $\pm 6.4E-06$	3.8E-05 $\pm 4.4E-05$	2.3E-12 $\pm 7.9E-13$	3.3E-05 $\pm 1.1E-05$	6.7E+00 $\pm 1.4E+01$
	76E12	6.9E-07 $\pm 2.1E-07$	1.4E-06 $\pm 4.0E-07$	5.9E-06 $\pm 2.7E-06$	2.0E-10 $\pm 9.5E-12$	7.9E-03 $\pm 3.7E-04$	1.1E+04 $\pm 3.5E+03$
	78D2	6.5E-06 $\pm 2.1E-06$	9.1E-06 $\pm 3.6E-06$	3.7E-05 $\pm 1.4E-05$	2.5E-11 $\pm 3.8E-12$	3.2E-03 $\pm 4.8E-04$	4.9E+02 $\pm 1.8E+02$
	89B1	3.1E-06 $\pm 3.3E-06$	7.7E-06 $\pm 3.7E-06$	1.0E-05 $\pm 7.4E-06$	1.9E-10 $\pm 3.5E-11$	4.7E-03 $\pm 8.8E-04$	1.5E+03 $\pm 1.6E+03$
UDP-Gal	78D2	3.7E-05 $\pm 1.4E-05$	6.3E-06 $\pm 3.2E-06$	1.1E-05 $\pm 1.1E-05$	2.5E-13 $\pm 5.5E-14$	6.3E-06 $\pm 1.4E-06$	1.7E-01 $\pm 7.5E-02$

UDP-Glc activity in group B, F and H

As Figures 3.23-3.27 show, families of reciprocal form plots are intersected, indicating group B, F, and H UGTs follow sequential mechanisms.

In four tested UGTs from these three groups, K_A of UDP-Glc lies in a range of 6.9E-07 M (76E12) \sim 6.5E-06 M (78D2), with a 9.4-fold difference. K_B of KMP is located in a narrower range of 1.4E-06 M (76E12) \sim 9.1E-06 M (78D2), with a 6.5-fold difference. According to the result, 76E12 needs the lowest concentration of substrates to facilitate significant catalysis, while 78D2 requires the highest.

The values of k_{cat} and k_{cat}/K_A indicate the catalytic capability and efficiency/performance of these tested UGTs. 76E12 has the highest catalytic capability and efficiency/performance, with k_{cat} $7.9\text{E-}03\text{ s}^{-1}$ and k_{cat}/K_A $1.1\text{E+}04\text{ s}^{-1}\text{M}^{-1}$, while 76E11 has the lowest with k_{cat} $3.3\text{E-}05\text{ s}^{-1}$ and k_{cat}/K_A $6.7\text{E+}00\text{ s}^{-1}\text{M}^{-1}$. The K_{IA} of these four enzymes are all larger than corresponding K_A , suggesting that donor-enzyme binding may promote acceptor-enzyme binding.

From K_M ratios, 76E11, 76E12, 78D2 and 89B1 are suggested to follow Bi-Bi random sequential mechanism, with K_M ratios as $1.3 (\pm 2.9)$, $2.0 (\pm 0.8)$, $1.4 (\pm 0.7)$ and $2.5 (\pm 2.8)$ respectively.

UDP-Glc and UDP-Gal activity of 78D2

For 78D2, UDP-Glc is believed to be a better substrate candidate than UDP-Gal in aspects of K_A , K_B , k_{cat} and k_{cat}/K_A : The K_A of UDP-Glc is lower than that of UDP-Gal (18-fold), showing that a lower concentration of UDP-Glc can facilitate the glycosylation. Besides, k_{cat} and k_{cat}/K_A also reveal that 78D2 can transfer Glc with 510-fold higher catalytic capability and 29,000-fold higher catalytic efficiency/performance than that of Gal.

Notably, K_{IA} of 78D2 towards UDP-Gal is smaller than its K_A , consistent with the intersecting point below the x-axis in Figure 3.27. This result indicates that the binding of the donor-enzyme may prevent the binding of the acceptor-enzyme, rather than the ‘donor-driven promotion of acceptor binding’¹⁶⁶ that generally found in GT1 enzymes.

From K_M ratio, 78D2 is suggested to facilitate Gal transfer in the Bi-Bi random sequential mechanism with a ratio value of $5.9 (\pm 3.7)$. The mechanism is the same as Glc transfer reaction of 78D2.

So far, the full kinetics of nine active UGTs (group D 73B4, 73B5, 73C1, 73C5 and 73C6; group H 76E11 and 76E12; group F 78D2, and group B 89B1) were examined. The *in vitro* kinetic study of UGTs can be used to not only describe the enzymatic process, but also shed light on many related research topics such as *in vivo* metabolism. Many plant GT researches showed the catalytic specificity of recombinant enzyme *in vitro* is the same/similar as that *in vivo*^{170,171}. In addition, many experimental evidence

suggested K_M provides an approximation of substrate *in vivo*⁶⁵. Apart from the reference to *in vivo* metabolism, some other applications such as chemo enzymatic synthesis can use the information derived from kinetics, such as how to select a suitable UGT with desired substrate and optimise a reaction condition for maximum efficiency. Moreover, the kinetic data obtained in this chapter will be further discussed along with the active site analysis of enzymes (Chapter 4).

3.3.3 Error analysis

It cannot be neglected that there were errors in the full kinetic study of UGTs as displayed above, which were generally not low. In this section, possible origins of these error will be discussed mainly including two areas: data fitting and experimental design.

Firstly, errors might originate from the process of data fitting mathematically. Compared to the single substrate kinetic model, the Bi-Bi mechanism kinetic model is multivariate hence very complicated (Section 3.1.2.4 and Appendix A.1). Regression analysis of such a complicated mathematic equation (Equation 3.1) requires a large amount of precisely measured data. However, the data amount was limited due to MS being used. Therefore, the regression could produce mathematical fittings with large error.

In addition, the MS method itself can contribute to error. The process of MS fragmentation can reduce some sensitivity of MS although MRM can help increase a confident detection, however, the error might be big especially at the initial rate stage where not many products were formed. Although measures that minimise the possible errors had been taken such as using internal standard, systematic errors can occur including 1) MS system may drift from run to run in many aspects such as sensitivity¹³⁴. This happens particularly for long time experiments. In a full kinetic experiment, 80 samples plus a calibration curve were measured, which lasted around 8 hours. Although the error from system drift could be reduced by using internal standard method, it could happen and be accumulated; 2) Ion suppression frequently occur and may vary during each experiment¹³⁴. This error was minimised by using internal standard and avoiding using a high concentration of substrates (discussed in Section 3.1.2). However, it is hard to be fully eliminated; 3) Measurement of a signal intensity

always carries a minimal intrinsic error¹³⁴. This error is inversely proportional to the square root of the number of ions detected for that signal. This error is relatively large in the quantification of a small amount of analyte and cannot be reduced by using an internal standard (Discussed in Section 3.1.2); 4) Contaminating substances may generate ion signals that interfere with the quantification of analyte ion¹³⁴. Although MRM mode can recognise an analyte specifically by its specific MRM transition, contamination might associate with analyte thus resulting in the change of MRM transition of analyte. The contamination is hard to be avoided in a shared laboratory, though a one-hour cleaning program was performed before each time; 5) Sample handling of MS by autonomous injection might cause unpredictable random errors¹⁷².

Although there are many other methods to test UGT kinetics, such as coupled assay and fluorescent-based UGT assay, they all have different limitations. For example, coupled assay requires additional coupled enzymes with 2 steps to turn into UV/VIS signal, which probably influence the kinetics of tested enzyme. A few trials were conducted using this method but it was decided not to go further due to various practical limitations. Finally, the MS was adopted in this research with several advantages especially on direct monitoring of the reaction without any modification and was proved as a robust method previously.

Another possible error might come from the experimental design – [S] range used in an experiment. With the complexity of a full kinetic model, it is hard to estimate a rough K_M experimentally then designing a [S] range of $1/5-5K_M$. Hence, an empirical estimation (with considerations previous studies, kinetic model and practical problems) of [S] range was performed (discussed in Section 3.1.2). Although kinetic parameter calculation is not affected by imperfect [S] range mathematically, the imperfect [S] range may affect the optimal performance of experiments and lead to large errors. According to the data generated above, some enzyme kinetics probably need to be double checked with the new $1/5-5K_M$ [S] range to obtain the best quality data (Table 3.5). This will be done as one of the future works.

Table 3.5: $1/5-5K_M$ [S] range based on the full kinetic experimental K_M value

Donor	UGT	Range [A] (μM)	Range [B] (μM)
UDP-Glc	73B4	0.6-16	0.4-9.0
	73B5	0.8-20	0.3-7.5
	73C1	0.4-10	9.4-240
	73C5	0.8-20	0.4-15
	73C6	4.4-110	22-550
	76E11	1.0-25	1.3-32
	76E12	0.1-3.5	0.3-7.0
	78D2	1.3-33	1.8-46
	89B1	0.6-16	1.5-39
UDP-Gal	73C5	0.1-2.1	0.1-3.0
	78D2	7.4-190	1.3-32
UDP-GlcNAc	73B4	0.2-5.5	8.6-220

Table 3.5 gives the optimal [S] range according to traditional kinetic theory. This guides the [S] range applied in future experiments, however, careful consideration should be taken: 1) As discussed in Section 3.1, too low [S] might not be enough to produce detectable signal of products especially at the onset of the experiment. For example, in the 73C5 catalysed UDP-Gal reaction, signal counts of KMP-Gal in MRM were 146 counts at $2 \mu\text{M}$ UDP-Gal and $2 \mu\text{M}$ KMP at time point 5 minutes. If [S] is re-designed according to Table 3.5, signal counts of KMP-Gal in MRM would be estimated (based on kinetic equation with 73C5 UDP-Gal kinetic parameters obtained above) only 1 count at $0.1 \mu\text{M}$ UDP-Gal and $0.1 \mu\text{M}$ KMP. Such low signal count is not acceptable due to MS detection limit (discussed in Section 3.1). Hence, the [UDP-Gal] concentration range cannot reach as low as $0.1 \mu\text{M}$. To avoid problem of detection limit, [S] concentration is recommended not below $1 \mu\text{M}$ to ensure that enough signal ions can be detected. Hence, UDP-Gal concentration range for 73C5 UDP-Gal kinetic test should be adjusted to $1.0-2.1 \mu\text{M}$ rather than $0.1-2.1 \mu\text{M}$; 2) Concentration of KMP cannot be too high due to its poor solubility in buffer. The precipitate was observed when the concentration near $200 \mu\text{M}$. Hence, the range of KMP concentration should always be below $200 \mu\text{M}$.

All in all, all kinetic results above were reported with final errors from experimental design and data fitting, which was not low indeed. For cases where the error was larger than the data itself, it is recommended to repeat the experiment for a better result.

It should be noted that enzyme activity can be influenced by manifold factors. Different enzymes show their highest activities under different optimum conditions. Any deviation from the optimum conditions leads to the reduction of enzyme activity

that may be relatively subtle or crucial. Hence, enzyme activity reported in this project is only applied to the conditions clarified.

3.4 Conclusions

In this study, 29 recombinant UGTs from group B (89B1 and 89C1), D (73B4, 73B5, 73C1, 73C2, 73C3, 73C4, 73C5, 73C6, 73C7 and 73D1), F (78D1 and 78D2), H (76C2, 76C3, 76C4, 76C5, 76E3, 76E4, 76E5, 76E9, 76E11, 76E12, 76F1 and 76F2), and L (74C1, 75C1, and 75D1) were examined in the aspects of substrate specificities, kinetic mechanism and parameters.

Donor library (UDP-Glc, UDP-Gal, UDP-GlcNAc, UDP-GalNAc, UDP-Rha, dTDP-Rha, GDP-Glc and GDP-Fuc) and acceptor library (48 acceptors included) were set for the substrate specificity screening of UGTs, and displayed as a GAR figure (Green: positive activity; Amber: ambiguous activity; Red: no activity). The product formation was detected by a combination of the ‘full scan’ mode and ‘product ion’ mode of MS.

Nine out of these 29 UGTs (except 78D1 and 89C1) used UDP-Glc, while 78D1 and 89C1 used UDP-Rha. Some new donor activities of these UGTs were discovered, including 73C5 and 78D2 with UDP-Gal, 73B4 and 78D2 with UDP-GlcNAc, 78D2 with GDP-Glc, 78D1 and 89C1 with dTDP-Rha. As for acceptor specificities, all of these 11 UGTs utilised flavonoids (UDP-Glc as the donor, with the exceptions that 78D1 and 89C1 used UDP-Rha). Additionally, coumarins were used in 73B4 (Esculetin, 7-hydroxycoumerin-3-carboxylic acid), 73C5 (Esculetin), 89B1 (Esculetin), and 89C1 (Esculetin). Caffeic acid could be used for 73B4. Aloe-emodin, gossypol and benzoic acid derivatives (3,4-dihydroxybenzoic acid and 2,5-dihydroxybenzoic acid) could be used for 73C5.

Full kinetics of nine UGTs (73B4, 73B5, 73C1, 73C5, 73C6, 76E11, 76E12, 78D2 and 89B1) were examined with KMP as the acceptor and their available donors. MRM mode of MS was performed to quantify the accumulated glycosylated product as a function of the reaction time. The study revealed the kinetic parameters of each UGT. Towards donor UDP-Glc, Michaelis parameter (K_A) lay in the range 10^{-7} ~ 10^{-5} M. Catalytic capability was in the range 10^{-5} ~ 10^{-2} s⁻¹. Catalytic efficiency (for donor UDP-Glc) varied in the range of 1 ~ 10^2 M⁻¹s⁻¹.

Intersecting Reciprocal form plots and the K_M ratio illustrated further mechanism of kinetics (validated by product inhibition assay of 73C5). UGTs generally mediated

glycosylation by Bi-Bi random sequential mechanism, however, exception was discovered such as 73B4 facilitated UDP-GlcNAc glycosylation via a Bi-Bi ordered mechanism.

In summary, we demonstrated an *in vitro* characterisation of UGTs in terms of substrate specificity, kinetic mechanism and parameters, shedding light on the understanding of UGT metabolism *in vivo* and potential applications such as an alternative approach for synthesis of valuable glycans.

**Chapter 4 Identification of catalytic key amino acids
(ckAAs) in UGTs**

In an enzymatic reaction, the substrates are accommodated into the active site of the enzyme and undergo chemical reactions. The active site of an enzyme can be divided into the binding site and catalytic site. AAs in the binding site (binding AAs, bAAs) can interact with the substrate to make an enzyme-substrate complex such as by forming temporary bonds (e.g. hydrogen bond⁸⁹) or non-covalent interactions (e.g. van der Waals⁹⁰). AAs in the catalytic site (catalytic AAs, cAAs) can catalyse a reaction such as acting as a Brønsted base to activate the substrate for nucleophilic attack⁹¹. Both bAAs and cAAs are key for an enzymatic reaction, and are frequently named as catalytic key amino acids (ckAAs)^{92,93}.

This chapter aims to identify ckAAs in target UGTs that showed positive activities in Chapter 3 (73B4, 74B5, 73C1, 73C5, 73C6, 76E11, 76E12, 78D2 and 89B1 all showed UDP-Glc activity. Additionally, 73B4 and 78D2 also showed UDP-GlcNAc activity. 73C5 and 78D2 showed UDP-Gal activity), using the multiple sequence alignment (MSA)-based method.

4.1 Method development

A crystal structure of an enzyme can provide the most direct and reliable evidence to identify ckAAs, by providing insight into the essence of a protein's structure and interactions with other molecules directly¹⁷³. However, obtaining a protein crystal can be a very difficult and expensive process¹⁷⁴. For examples, proteins must be produced in a relatively large amount with a high level of purity and the conditions for crystallization need to be optimized for each enzyme. In addition, crystals, even after they have been obtained, must diffract well to produce an electron density map with sufficiently high resolution¹⁷⁴. Thus, attempts to elucidate the molecular basis of proteins by crystallization frequently fall short even with much effort and time.

Without crystal structures, rational MSA-based methods are usually adopted and shown to be efficient tools to identify ckAAs in enzymes^{93,103,175-178}. Previously solved crystal structures of homologous proteins, in particular the complex of an enzyme-substrate, can be used as evidence to predict potential ckAAs in other homologous proteins. Further experimental data from site-directed mutagenesis (SDM) can provide experimental evidence to identify whether potential ckAAs (inferred from MSA) play a significant role in enzyme activity/performance, by checking whether mutation of potential ckAAs have an effect on enzyme activity/performance.

The MSA-based methodology has been widely used and acknowledged as a reliable experimental approach to identify ckAAs in UGTs. For example, the Thorsøe group identified ckAAs in UGT85B1 by using an MSA-based method with GtfA and GtfB as templates⁹³. At equivalent sites of ckAAs that were located in GtfA and GtfB, AAs in 85B1 were inferred as ckAAs. For instance, GtfA E318 and GtfB D332 were observed interacting with sugar in the crystal. At the equivalent site, E410 appeared in 85B1, thus to be inferred as a potential ckAA in 85B1 that might bind sugar as has been observed in GtfA and GtfB (also shown in homology modelling). Subsequently, alanine mutation in E410 was carried out by SDM. UDP-Glc activity of E410A showed a 225-fold decrease compared to the wild type (WT), thus substantiating the results from the MSA-based inference that E410 was a ckAA. Additionally, further research conducted by Osmani used the same method to identify ckAAs in 94B1 using 71G1 and VvGT1 as templates⁹². N123 and D152 were inferred as ckAAs in 94B1 by

binding to the acceptor according to the MSA results. Subsequently, experimental data from SDM agreed with the significant role of these potential ckAAs by detecting that mutants N123A and D152A had less than 15% of activity when compared to the WT.

In this Chapter, MSA was used to identify ckAAs in target UGTs. By comparing AA sequences of seven template plant UGTs whose crystal structures were already known, with target UGTs of 73B4, 74B5, 73C1, 73C5, 73C6, 76E11, 76E12, 78D2 and 89B1 that were to be tested in this project, potential ckAAs in target UGTs involved in binding could be predicted. Then, mutations of potential ckAAs were performed by SDM and the activities of the mutants were compared to their corresponding WT enzymes. The results of the comparison of activity between mutants and WT will show whether the predicted important roles of those potential ckAAs are correct (general Scheme in Figure 4.1). This methodology provides a reliable identification of ckAAs from known crystal structures and avoids the structure distorting prediction from computational method by using experimental data (experimental data of mutants) instead¹⁷⁹.

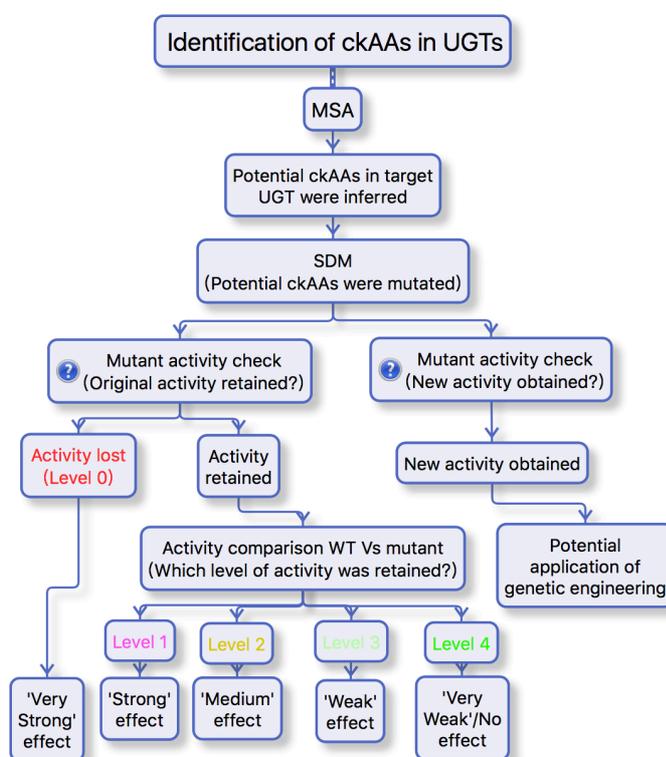


Figure 4.1: General scheme of Chapter 4.

4.1.1 Template UGTs and MSA

4.1.1.1 Template UGTs

MSA was performed on nine UGTs and seven crystal-structure-known template plant UGTs (*VvGT1*³⁵, 78K6^{79,80}, 71G1⁷⁶, 72B1⁸¹, 78G1⁷⁸, 85H2⁷⁷ and *OS79*⁸²). For example, one of the important template UGTs used was the red grape *Vitis vinifera* flavonoid 3-O-GT *VvGT1*³⁵, whose crystal structure was revealed with the substrates [kaempferol (KMP) and UDP-2-deoxy-2-fluoro-Glc (U2F)] presented in the active centre. This is one of the important templates that can provide valuable interaction information for enzymes and substrates, which makes it as the most frequently used template in many UGT researches¹⁸⁰⁻¹⁸⁵. In *VvGT1*, eight ckAAs were key to binding the donor (Figure 4.2): W332 and E358 bound to uridine via a pi-pi bond with the uracil ring and a hydrogen bond with ribose respectively; T19, T280 and H350 were responsible for binding phosphate via hydrogen bonds; D374, Q375 and T141 were involved in binding Glc by forming hydrogen bonds towards O3/O4, O2/O3 and O6 of Glc respectively.

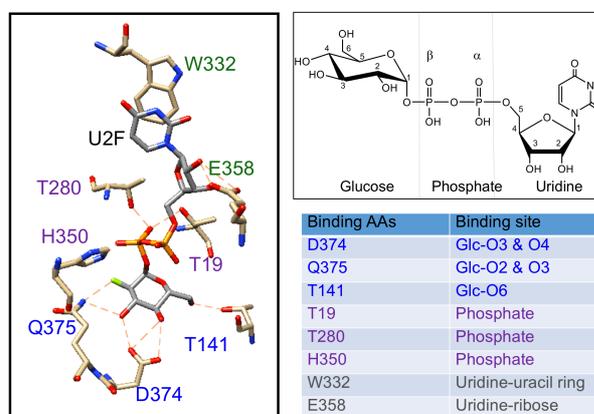


Figure 4.2: ckAAs in *VvGT1*: D374, Q375 and T141 bind to Glc (Glc-O3 and Glc-O4, Glc-O2 and Glc-O3, and Glc-O6, respectively); T19, T280 and H350 bind to phosphate; W332 and E358 bind to uridine (the uracil ring and ribose respectively).

In other template UGTs, some of their ckAAs were also observed at equivalent sites of where the eight ckAAs in *VvGT1* are listed above. For example, the D-Q (D374-Q375 *VvGT1* numbering) motif was highly conserved among template UGTs. In those crystal complex structures of the enzyme-substrate (72B1, 71G1 and *OS79*), the D-Q motif was also conserved to bind the -OH at the C2-C4 positions of the sugar in the same manner as what observed in *VvGT1*. In those crystal enzyme structures without

substrates (78G1 and 85H2 and 78K6), the D-Q motif was predicted to interact with donors (from crystal-based docking experiments) in the same pattern as for VvGT1. However, a few exceptions exist. These exceptions were ckAAs that were located at the N-terminal: VvGT1 T19 that bound to phosphate and T141 that bound to the Glc-O6. Conversely the equivalent sites of VvGT1 T141, 72B1 and 85H2 did not show corresponding ckAAs towards the Glc-O6. At the equivalent site of VvGT1 T19, no corresponding AAs that bound to phosphate were observed in 72B1, 71G1, 85H2 and OS79. These facts showed that ckAAs in UGTs towards a donor molecule followed a conserved fashion in general, but with some variations. Therefore, the ckAAs in individual UGT need to be investigated and identified specifically.

4.1.1.2 MSA

MSA was performed by Clustal Omega¹⁸⁶. This is the most recent algorithm in the Clustal family and enables aligning any number of AA sequences quickly and accurately¹⁸⁷. Once the FASTA format of UGTs (template and target) were inputted into the Clustal Omega website, the software would output a result of MSA by running the following procedures: 1) A MSA was produced by pairwise alignments using the k-tuple method; 2) The sequences were clustered using the mBed method followed by the k-means clustering method; 3) The guide tree was constructed using the UPGMA method; 4) The final MSA result was produced using the HHAlign package, which aligns two profile-hidden Markov models (HMM)¹⁸⁸.

After MSA, the result was generated in Clustal w/o numbers format while the sequences were compared according to their similarities (Figure 4.3). Gaps were introduced so that the optimal MSA score was achieved. From the resulting MSA, sequence homology can be inferred. The signature motif of UGTs, the PSPG motif was observed (highlighted in the amber rectangle in Figure 4.3) in both the template and target UGTs. Moreover, the potential ckAAs in the target UGTs (highlighted in the green rectangle in Figure 4.3) were found at the equivalent site of where ckAAs in the template UGTs were located.

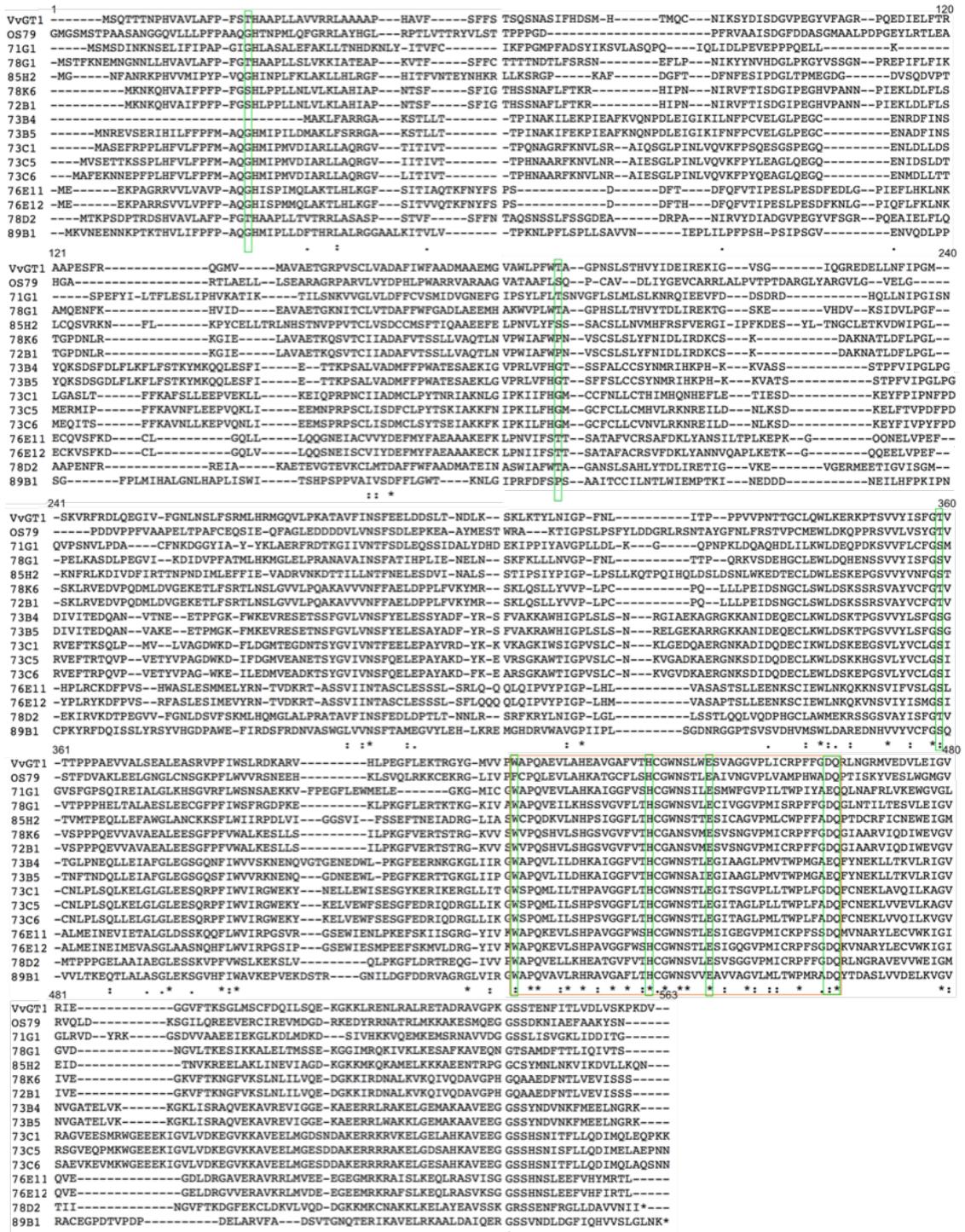


Figure 4.3: MSA among seven template UGTs and nine target UGTs. Potential ckaAs are highlighted in the green rectangle and the PSPP motif is highlighted in the amber rectangle.

Although MSA has been acknowledged as a method to infer the potential ckaAs in an unknown enzyme, the level of accuracy and reliability should be considered according to specific UGTs¹⁸⁹. This is due to there being an important assumption underlying MSA in that protein structure is related to AA sequences¹⁹⁰. Additionally,

MSA can only infer and exploit the relationships of UGTs based on existing sequences, but the ancestral sequences and the level of evolution might differ in different UGTs¹⁹⁰. Thus, further experiments (e.g. SDM) are required to validate the actual role that potential ckAA play in different UGTs.

4.1.2 Mutants

In this project, SDM was carried out to mutate each potential ckAAs inferred from the MSA. Mutant plasmids were transformed to XL-1 blue competent cells and mutant proteins were produced following the same procedures as for the WT.

4.1.2.1 Mutant plasmids

SDM was performed using the Q5® SDM kit (NEB) (Figure 4.4). Three steps were included: 1) exponential amplification; 2) treatment and enrichment; 3) transformation.

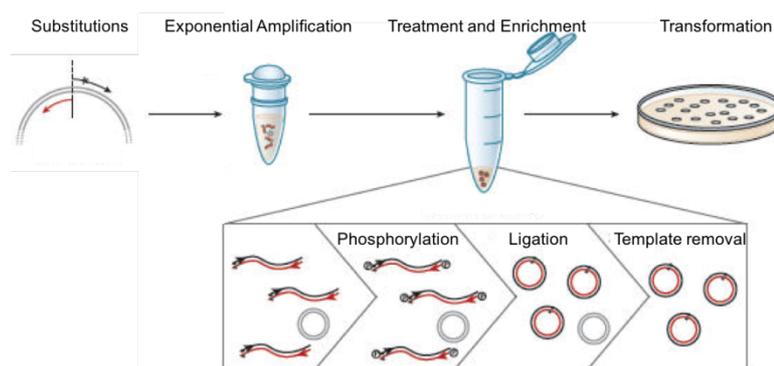


Figure 4.4: Schematic diagram of SDM. SDM undergoes exponential amplification, treatment and enrichment, and transformation (manual of Q5® SDM kit, NEB).

Exponential amplification (EA)

EA in SDM is based on the principle of a polymerase chain reaction (PCR), but with primers incorporating the desired mutation. After the process of EA, plasmids (both mutant and WT) were amplified and therefore the plasmid concentration increased. The results of EA can be analysed by agarose gel (Figure 4.5). Agarose gel provides a verification of successful EA at this stage, which is crucial to ensure the success of subsequent procedures and to avoid the propagation of errors.

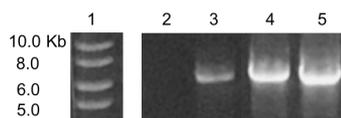


Figure 4.5: Agarose gel of WT 73C5 and its mutants. 1: DNA ladder; 2: Plasmid of 73C5 WT before EA; 3~5: EA product in SDM experiments 73C5 D397A, D397E and Q398A. This Figure is a representative of other experiments with successful amplification of the plasmid after the EA step. 73C5 WT before EA could not be observed on the agarose gel due to its low concentration, however, EA increased the concentration of the plasmid, thus making them visible in the gel (the band appears at approximately 6.0 kb compared to the DNA ladder).

In Figure 4.5, no band is shown in lane 2, as the concentration of plasmid was so low that it was out of the detection range of the SYBR® Safe DNA stain (500 pg/band) (manual of SYBR® Safe DNA stain, Thermo Fisher Scientific). After EA, a band around 6.0 Kb was observed in lanes 3-5 (the size of recombinant 73C5: 6.4 Kb), indicating that the concentration of the plasmid had increased after EA.

It is noteworthy that in some SDM kits, such as the Agilent Quickchange® SDM kit, agarose gel is not required as a band may or may not be visualized at this stage¹⁹¹. This is due to the amplification being linear but not exponential in these kits, thus the concentration increase of plasmids after amplification is not significant to be detected in agarose gel. The underlying reason for the linear amplification in these kits is the design of the overlapping primer: the entire plasmid with overlapping primers is amplified and generates a nicked circular DNA that cannot take part in the next cycle as a DNA template^{191,192} (Figure 4.6). However, back-to-back primer design was used in this experiment. Linear DNA was generated with a back-to-back primer, thus allowing exponential amplification of the plasmid (Figure 4.7)¹⁹³. Such amplification has many advantages: it makes the product verified at this stage, to ensure the success of further experimental steps and to avoid error propagation; also, it enables a high yield of products, increasing the proportion of mutant plasmids and decreasing the WT background.

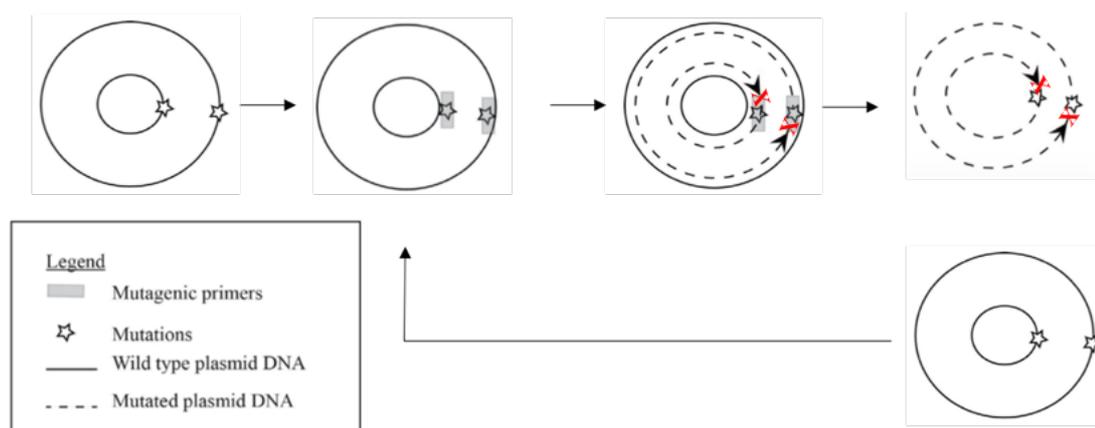


Figure 4.6: Schematic diagram of SDM using overlapping primers. Nicked (represented by red cross) circulated daughter plasmids are produced in the amplification and cannot be involved in the next cycle as template DNA. Thus, this amplification is a linear process.

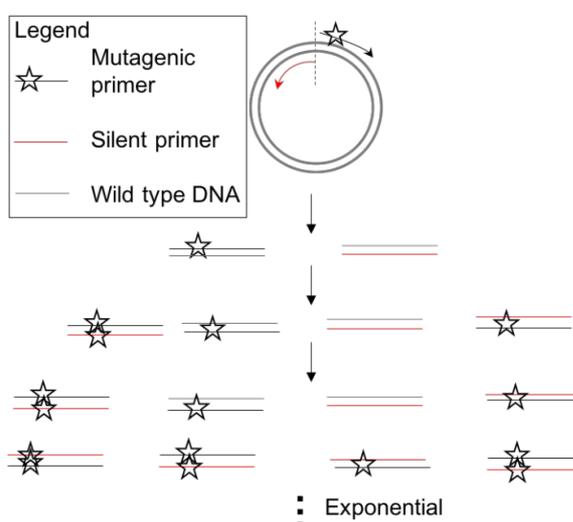


Figure 4.7: Schematic diagram of SDM using a back-to-back primer. Linear DNA is produced in the amplification and will be involved in the next cycle as template DNA. Thus, the amplification is an exponential process.

Treatment and enrichment

EA products existed as linear DNAs which could not be transformed into competent cells directly and contained WT. Therefore, EA products were incubated with the KLD mix (kinase, ligase and DpnI), with the aim of cyclising the DNA and removing the WT background.

After incubating with KLD mix, linear EA products were cyclised therefore could be transformed into competent cells. The cyclisation of EA products was accomplished by kinase and ligase in the KLD mix. Kinase is responsible for DNAs phosphorylation and then ligase can ligate and cyclise phosphorylated linear DNAs.

After cyclisation of linear DNAs, the next step was template removal. There are two types of WT plasmids that should be avoided: 1) ‘original WT plasmids’ that were extracted from cells and used as initial template plasmids at the outset of EA; 2) ‘obtained WT plasmids’ that were produced during EA when the silent primer was used as the template for DNA replication. Among which, the former (‘original WT plasmids’) could be digested by DpnI in the KLD mix, as they were methylated (plasmids amplified *in vivo* were methylated) but the EA products were not. However, the latter (‘obtained WT plasmids’) could not be recognized and digested by KLD mix as they were also amplified *in vitro*. Although this might be a problem of the SDM experiment, it could be omitted as the proportion of ‘obtained WT plasmids’ was negligible. As shown above (Figure 4.7), the amplification of mutant plasmids with respect to cycle (n) was shown by the equation ‘ $2^n - (n+1)$ ’, while the amplification of ‘obtained WT plasmids’ with respect to cycle (n) was shown by the equation ‘ $n-1$ ’. After 25 cycles in EA, ‘obtained WT plasmids’ were overwhelmed in terms of numbers by the millions of mutant plasmids ($24/3.36E+07=7.45E-07$).

Transformation

After treatment and enrichment, cyclised mutant plasmids were transformed into XL-1 blue competent cells, following the same procedures of WT plasmid transformation.

Mutant plasmid verification

Mutant plasmids were sent to the Source Biosource Company for verification using the Sanger sequencing method¹⁹⁴. A typical sequencing result with mutant 78D2 Q381E is shown in Figure 4.8.

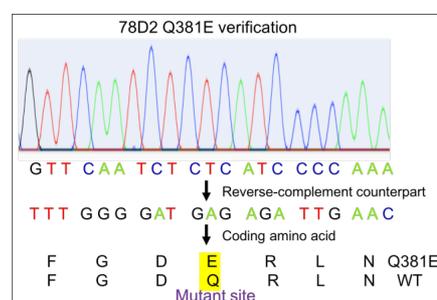


Figure 4.8: DNA sequencing results of 78D2 Q381E. At site 381 of 78D2, CAG (coding Q) was replaced by GAG (coding E) after SDM, indicating that 78D2 Q381 was successfully mutated by E.

In Figure 4.8, it is observed that GAG (E) replaced the original CAG (Q) at site 381 in 78D2 after SDM, showing that Q was successfully switched to E.

As a large number of mutants (106) was produced in this project, sequencing for all plasmids would be costly and time consuming. Hence, a few representative mutant plasmids were selected for verification. These mutant plasmids were 73C6 Q398A, 78D1 N374Q, 78D2 Q381E and the double point mutant 78D2 C357A-C374A. Sanger sequencing showed that all of these selected mutant plasmids were constructed correctly, indicating the reliability of the method. In addition, the same sequence checking was carried out by another PhD student's project in our group and the results showed that the mutant plasmids were constructed correctly, which demonstrated the robustness, reliability and reproducibility of our method.

In some previous studies¹⁹⁵⁻¹⁹⁷, the plasmid verification step was also omitted in the SDM experiment when using NEB Biolabs Q5 SDM kit as its highly efficient, reliable and robust (discussed above)¹⁹³.

Despite the low probability of wrong mutant plasmids produced by the NEB Biolabs Q5 SDM kit, it was possible that mutants might not be correctly constructed. An unsuccessful SDM may cause the same activity detected between the WT and mutant enzyme in the next step. This could lead to a confusion about the reason for unchanged activity which could be due to the failure of the mutation or due to the insignificant role the potential ckAA plays, which gives a wrong indication to the identification of ckAAs. Thus, the sequence should be double-checked, especially for those mutants with unchanged activity compared to the WT, in the future.

4.1.2.2 Production of mutant protein

The production of mutant proteins followed the same procedure of that for the WT (c.f. Chapter 2): mutant proteins were induced by IPTG, purified by a GST tag column and verified by SDS-PAGE.

Although previous research has shown a general trend that proteins manifested a robustness towards point mutations^{198,199}, It is possible that SDM may lead to instability of protein folding and the change of overall conformation, resulting in the malfunction of a protein²⁰⁰. This can lead to a confusion as to whether the change of

activity was due to the mutation of potential ckAAs or the possible improper folding and conformation of the protein.

Thus, there is an important assumption in this research that point mutagenesis did not change the stability and overall conformation of proteins in this project. This assumption has also been widely used in UGT research, such as in the study of VvGT1 (mutants: T141A, D374A, Q375H and Q385N), *Caenorhabditis elegans* (a xylosyltransferase)²⁰¹, LgtC from *Neisseria meningitides*²⁰² and human Gb3/CD77 synthase (α 1,4-GalTs)²⁰³.

In addition, a ‘Site Directed Mutator’ program (developed by the University of Cambridge) (<http://131.111.43.103>) was used to provide computational data to indicate whether point mutagenesis would affect protein stability. This program is based on a statistical potential energy function, to predict the effects of mutations on protein stability²⁰⁰. By giving the stability score DDG (PSEUDO Delta Delta G), the program indicates how mutants affect the stability of the protein with six classes: neutral ($-0.5 \leq \text{DDG} \leq 0.5$), slightly destabilizing ($-1.0 \leq \text{DDG} < -0.5$), slightly stabilizing ($0.5 < \text{DDG} \leq 1.0$), destabilizing ($-2.0 \leq \text{DDG} < -1.0$), stabilizing ($1.0 < \text{DDG} \leq 2.0$), highly destabilizing ($\text{DDG} < -2.0$) and highly stabilizing ($\text{DDG} > 2.0$)²⁰⁰. This program has been used in many studies²⁰⁴⁻²⁰⁶.

This program was firstly used in some template UGTs with mutagenesis data (Table 4.1). It showed that point mutagenesis may affect the stability of a protein to some extent, but no ‘highly destabilizing’ was observed.

Table 4.1: Mutations on protein stability of some template UGTs (predicted by SDM program)

WT	Mutants	DDG	Class
VvGT1	H20A	-0.91	Slightly destabilising
	T141A	-0.62	Slightly destabilising
	D374A	-0.21	Neutral
	Q375H	-0.15	Neutral
	Q375N	-1.49	Destabilising
71G1	Y202A	-1.19	Destabilising
78G1	H26A	-0.7	Slightly destabilising
	E192A	0.52	Slightly stabilising
	D376A	-0.21	Slightly destabilising

Subsequently, the program was run with 73B4 as an example to predict whether point mutagenesis affected the stability of 73B4 (WT 73B4 was modelled using 71G1 as the template) (Table 4.2).

Table 4.2: Mutations on protein stability of 73B4 (predicted by SDM program)

Mutants	DDG	Class
S270A	0.96	Slightly stabilising
W331A	8.95	Highly stabilising
H349A	-0.75	Slightly destabilising
E357A	0.96	Slightly stabilising
E373A	1.76	Stabilising
E373D	1.32	Stabilising
E373Q	0.28	neutral
E373K	-1.28	Destabilising
Q374A	2.81	Highly stabilising
Q374N	-0.65	Slightly destabilising
Q374H	0.15	neutral
Q374E	1.08	Stabilising

As Table 4.2 shows, point mutants of 73B4 maintained the stability of the protein at different levels but there was no ‘highly destabilizing’ level. Therefore, it can reasonably be assumed that point mutagenesis of 73B4 does not disrupt the stability of protein.

Although there is computational evidence to support the point mutagenesis does not significantly destabilise target UGTs, experimental result is required to improve the reliability of the research. Consequently, technology such as circular dichroism (CD) could be used as one of the optimisations in future work, to provide solid experimental evidence that point mutagenesis may be benign or harmful to protein function.

4.1.3 Mutant activity

Activities of mutants were checked both qualitatively and quantitatively (if active) with the MS method described in Chapter 3. Qualitative analysis checked the activities of mutants against sugar donors in the donor library (c.f. Chapter 3) with KMP as the acceptor. This qualitative analysis could lead to two discoveries 1) whether mutants lost/retained original activities and 2) whether mutants obtained new activities. Once mutants retained the original activities, the quantitative analysis then checked how much of the original activity was reduced/increased compare to WT.

The activity comparison was carried out between the mutant and WT (as the control) using a fixed substrate condition. Glycosylated-KMP counts in the MRM mode (c.f. Section 3.1.2) at different time intervals in the control (WT enzyme) and experimental groups (mutant enzyme), progressive curves (a function of MRM counts of glycosylated product/internal standard V s time) of WT and mutant enzymes, were obtained and initial rates were compared.

As the equation of initial rate $V_0 = (k_{cat}/K_M) * ([E][S])$ indicates, the direct initial rate comparison of WT and mutant enzymes in a closed system reflects their k_{cat}/K_M (unit: $s^{-1}M^{-1}$) comparison ($[E]$ and $[S]$ are fixed and the same)⁶⁵. As a higher k_{cat}/K_M indicated better enzyme activity/performance generally, k_{cat}/K_M can be used as an index to give an indication of the functional effects of enzyme mutations¹⁶⁵. Although separate k_{cat} and K_M provide more kinetic details of a mutant, the combination index k_{cat}/K_M would summarise the net advantage or disadvantage of a mutation to the enzyme in one constant. Hence, the interpretation of changes in enzymatic activity/performance by the k_{cat}/K_M index is straightforward and has a longstanding tradition in biochemistry practice, especially to guide directed mutagenesis research. Thus, this criterion was used in much research such as the enzyme activity/performance comparison between lipase and α -chymotrypsin²⁰⁷ and the comparison between p120GAP with the GTP-bound WT hERas and WT HRas²⁰⁸.

4.1.4 Determination of ckAAs in UGTs

Summarising steps above, the ckAAs in target UGTs were determined by 1) to infer potential ckAAs using MSA. The ckAAs in target UGTs were inferred at the equivalent site of the ckAAs in the template UGTs; 2) to obtain mutants (mutation to alanine or an AA with a similar character) at sites of those potential ckAAs using SDM; 3) to analyse the mutants' activities by comparing the initial kinetic rates of WTs.

If the mutant loses its activity entirely, it is possible that the mutated AA is significant, i.e. the mutated AA is a ckAA. If the original activity is retained, activity comparison between WT and mutant enzymes was carried out. In this project, five levels were set up to distinguish how significant a role the potential ckAA plays in enzyme activity/performance: Level 0 – mutants showing no activity ('Very Strong' effect in enzyme activity/performance); Level 1 – mutants retain <20% original activity

(‘Strong’ effect in enzyme activity/performance); Level 2 – mutants retain 20~50% original activity (‘Medium’ effect in enzyme activity/performance); Level 3 – mutants retain 50~80% of original activity (‘Weak’ effect in enzyme activity/performance) and Level 4 – mutants retain >80% original activity (‘Very Weak/No’ effect in enzyme activity/performance).

As introduced above, the specific roles of ckAAs included acting as either bAA (binding AA) or cAA (catalytic AA) or both. Initial rate comparison between WT and mutants in this Chapter might not provide enough evidence to distinguish the specific role of that potential ckAAs play. However, it is valuable to consider investigating ckAAs in this Chapter as bAAs for two main reasons 1) these ckAAs in target UGTs were found at the equivalent site of bAAs in template UGTs; 2) cAAs of UGTs are frequently act as a Brønsted bases that activate the acceptor for nucleophilic attack by deprotonation²⁰⁹. This means that cAAs are likely to be located in a region that stacks the acceptor molecule, such as VvGT1 H20-D121³⁵, CalG3 H11-E115²¹⁰. However, this Chapter focuses on the donor region, where cAAs are unlikely to be located.

Although ckAAs studied in this Chapter are more likely to act as bAAs rather than cAAs, the method of comparing k_{cat}/K_M between WT and mutant alone is not sufficient to provide solid evidence to confirm that they are bAAs. As the criterion of k_{cat}/K_M reflects the net effect of a mutation on the enzyme, the specific binding effect of a mutation on the enzyme is not accurately obtained. In order to confirm the specific role of potential ckAAs (cAAs or bAAs), further experiments should be done such as 1) measurement of detailed kinetic parameters (e.g. K_M). For example, UGT1A6 H371 was identified as a bAA with the evidence of kinetic parameters that K_M values of mutant H371A towards the acceptor (scopoletin) and donor (UDP-GA) increased by 4- and 11-fold respectively and K_d values for the enzyme+UDP-GA complex increased 9- fold²¹¹. These detailed kinetic values implied that the mutation severely impaired substrate-enzyme binding. Therefore, H371 was identified as a bAA rather than a cAA²¹¹. 2) Isothermal Titration Calorimetry (ITC). For example, the functional role of R135, D246 and R293 in α 1,4-*N*-acetylhexosaminyltransferase (EXTL2) were investigated by ITC: the alanine mutations at the sites decreased K_d values 2.25-, 5.43- and 3.87-fold respectively and increased K_M values 2.32-, 2.97- and 1.78-fold respectively compared to the WT (at pH 7.5), which confirmed the binding effect of

these three AAs²¹². 3) Saturation transfer difference NMR (STD NMR). As STD NMR can measure the affinity K_D , it can be used to detect the binding. For instance, the binding kinetics of disaccharides trehalose and trehalose-6-phosphate to repressor protein TreR were determined by STD NMR²¹³. 4) Molecular docking and dynamic simulation studies. For example, molecular docking and dynamic simulations were conducted to investigate the interaction of kinase inhibitors (KIs) with peroxisome proliferator-activated receptor gamma (PPAR γ). By analysing the energy contribution for 270 (207-476) AAs in the PPAR γ complex with KIs and specifically orsiglitazone, seven AAs (C285, R288, E295, I326, L330, I341 and S342) showed a binding free energy contribution of more than -2.0 kcal/mol. The seven AAs were supposed as to be energetically important for ligand binding in the active site via hydrophobic or hydrogen bond interactions²¹⁴.

The work in this Chapter will not only contribute valuable information to the understanding of ckAAs in the active site of UGTs, but also shed light on further applications of UGTs such as active site manipulation.

4.2 Identification of ckAAs – sugar moiety

In this section, the potential ckAAs towards the C2, C3, C4 and C6 positions of the sugar moiety in UGTs were investigated using the methodology discussed above.

4.2.1 ckAAs towards C2 and C3 positions of the sugar moiety (the AA 44 within the PSPG motif)

Previous studies showed that the AA 44 within the PSPG motif (a conserved motif for the donor binding in plant UGTs, c.f. Chapter 1) was responsible for binding the –OH at the C2 and C3 positions of the Glc moiety via hydrogen bonds. At this site, Q was highly conserved in template UGTs (Table 4.3).

Table 4.3: ckAAs that bind the –OH of the C2 and C3 positions of the sugar moiety in solved crystal structures

Plant UGT	ckAAs	Result origin
VvGT1 ³⁵	Q375	PDB (2C1Z)
UGT 72B1 ⁸¹	Q389	PDB (2VCE)
UGT 71G1 ⁷⁶	Q382	PDB (2ACW)
UGT 78G1 ⁷⁸	Q377	Docking
UGT 85H2 ⁷⁷	Q403	Docking
UGT 78K6 (<i>Ct3GT-A</i>) ^{79,80}	Q368	Overlay on VvGT1
<i>Os79</i> ⁸²	Q390	PDB (5TMD)

4.2.1.1 MSA

MSA was performed on the template UGTs (Figure 4.9, top) and the nine target UGTs (Figure 4.9, bottom) to indicate the potential ckAAs in target UGTs towards the C2 and C3 positions of the sugar moiety.

WvGT1	C	R	P	F	F	G	D	Q
71G1	T	V	P	I	Y	A	E	Q
72B1	A	W	P	L	Y	A	E	Q
78G1	S	R	F	F	F	G	D	Q
78K6	C	R	F	F	F	G	D	Q
85H2	C	W	P	F	F	A	D	Q
OS79	A	M	P	H	W	A	D	Q
73B4	T	W	P	M	G	A	E	Q
73B5	T	W	P	M	G	A	E	Q
73C1	T	W	P	L	F	G	D	Q
73C5	T	W	P	L	F	A	D	Q
73C6	T	W	P	L	F	A	D	Q
76E11	C	K	P	F	S	S	D	Q
76E12	C	R	P	F	S	G	D	Q
78D2	C	R	P	F	F	G	D	Q
89B1	T	W	P	M	R	A	D	Q
	:	.	.	.	:	.	.	*

Figure 4.9: MSA of potential ckAAs towards C2 and C3 positions of the sugar moiety. The colours of residues represent their different properties: small and hydrophobic AAs AVFPMILW are coloured red; acidic AAs DE are coloured blue; basic AAs RK are coloured magenta; hydroxyl/sulfhydryl/amine AAs STYHCNGQ are coloured green. Symbols in the last line denote the degree of conservation: conserved (*), conserved mutations (:), semi-conserved mutations (.), and non-conserved mutations (). This Figure demonstrates that the potential ckAAs towards the C2 and C3 positions of the sugar moiety are highly conserved as Q.

In similarity to template UGTs, Q is conserved in all target UGTs.

4.2.1.2 SDM

Alanine scanning

Alanine is widely used to replace an original AA in SDM with its characteristics as non-bulky, inert, and containing a methyl side chain²¹⁵. Alanine mutants mimic AA deletion by eliminating the side chain, but maintaining the primary secondary structure of the protein. Consequently, alanine screening can be used as an initial step to determine the contribution of a specific AA in an enzyme²¹⁵.

The potential ckAAs in the target UGTs were mutated to alanine firstly. The relative activities of alanine mutants (compared to the WT) were examined followed by and shown in Figure 4.10, along with the donor screening result.

	UDP-Glc	UDP-Gal	GDP-Glc	GDP-Fuc	UDP-GlcNAc	UDP-GalNAc	UDP-Rha	dUDP-Rha
73B4 Q374A	58 WT (±7)				WT			
73B5 Q398A	84 WT (±7)							
73C1 Q394A	WT							
73C5 Q398A	WT	WT						
73C6 Q398A	WT							
76E11 Q374A	WT							
76E12 Q376A	WT							
78D2 Q381A	58 WT (±3)	WT	WT		22 WT (±1)			
89B1 Q390A	110 WT (±5)							

Figure 4.10: Relative activities of mutations (WT activity set as 100%) of ckAAs towards C2 and C3 positions of the sugar moiety (alanine scanning). The green rectangle depicts that original activities were retained at level 4 (>80%); the light green rectangle means that original activities were retained at level 3 (51%-80); the yellow rectangle shows original activities were retained at level 2 (20%-50%); the red rectangle depicts original activities were retained at level 0 (no activity). The olive green 'WT' indicates positive activity was detected in the WT (data from Chapter 3).

Alanine mutants: UDP-Glc activity

Results showed that UGT 73C1 Q394A, 73C5 Q398A, 73C6 Q398A, and 76E11 Q374A lost all their UDP-Glc activities, indicating Q is highly important in these enzymes for enzymatic activity/performance. The side chain elimination of Q probably removed the binding to the Glc-O2 and Glc-O3 in the WT, thus leading to the absence of activities. To better visualise the interaction between Q and UDP-Glc, 73C1 (as a model) was used as an example to show the interaction of Q and UDP-Glc by overlaying the template VvGT1 (Figure 4.11). The side chain atom NE2 of 73C1 Q394 interacts with Glc-O2 and Glc-O3 in the same manner as that of VvGT1 Q375. Following the same inference, it can be inferred that Q is a ckAA in 73C1 (numbering 394), 73C5 (numbering 398), 73C6 (numbering 398) and 76E11 (numbering 374) by a possible role of binding Glc-O2 and Glc-O3 (based on MSA result). This result also complied with some previous reports that Q played a significant role in UGTs. For example, mutations of UGT2B7 at Q (Q399A and Q399L) abolished morphine glucosidation and glucuronidation activities completely²¹⁶.

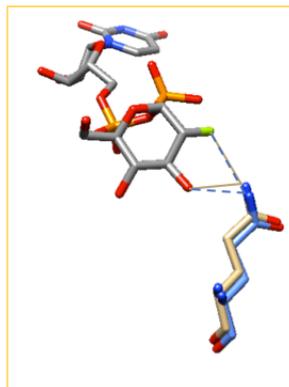


Figure 4.11: ckAA (Q) towards Glc-O2 & Glc-O3 in modelled 73C1 (cornflower blue) overlaying on the crystal VvGT1 complex U2F (UDP-2-deoxy-2-fluoro-Glc) (PDB: 2C1Z). The interactions of VvGT1 and 73C1 towards U2F are represented by a solid and dashed line, respectively. This Figure demonstrates that 73C1 Q394 binds Glc-O2 & Glc-O3 in the same manner as does VvGT1 Q375.

However, Q does not show a similar significance in some other target UGTs. For example, 73B4 Q374A (57%) and 78D2 Q381A (58%) still retained >50% original activities, which suggested that Q probably is not key in the enzyme and is not contributing to binding a Glc significantly. Especially in 73B5 and 89B1, the majority of original activities were retained (>80%, level 4), indicating that the elimination of the side chain Q was not sufficient to attenuate their activities. It is likely that Q in these enzymes is remote from the active site where the substrate is bound.

Alanine mutants: UDP-GlcNAc activity

As UDP-GlcNAc activity indicated, 73B4 and 78D2 had a region to accommodate UDP-GlcNAc and ckAAs to help binding and using UDP-GlcNAc. In crystal results of template GlcTs binding UDP-Glc^{35,76-82,217}, MurG binding UDP-GlcNAc²¹⁸ and many UGTs binding nucleotides (e.g. SpnP binding TDP²¹⁹, LanGT1 binding TDP²²⁰, SnogD binding dUDP²²¹ and OleI binding UDP²²²), the sugar donors were observed in a similar region of the UGTs. Thus, it was presumed that UDP-GlcNAc was located in the similar region where UDP-Glc is located in 73B4 and 78D2. The potential ckAAs towards UDP-Glc might be also important for 73B4 and 78D2 with UDP-GlcNAc activity.

As Glc and GlcNAc differ in the C2 position (with an –OH and *N*-Acetyl functional group respectively) structurally, ‘Q’ site attracted much attention as it was located in the vicinity of C2 and C3 positions of Glc and bound the –OH groups in template

GlcTs. As our study showed, alanine mutation of ‘Q’ site in 73B4 and 78D2 retained 0% and 22% original UDP-GlcNAc activity respectively (Figure 4.10), which indicated the significant role of ‘Q’ in UDP-GlcNAc activity. As suggested by template GlcTs, ‘Q’ site was probably located in the vicinity of C2 and C3 positions of GlcNAc. Then, 73B4 Q375 and 78D2 Q381 were possible to bind –OH groups at C3 position of GlcNAc, explaining why the alanine mutation of ‘Q’ site reduced UDP-GlcNAc activity. However, as GlcNAc has an *N*-acetyl group instead of –OH group at C2 position, ‘Q’ might not have the same interaction with *N*-acetyl group as that with –OH group of Glc. In the previous study of MurG+UDP-GlcNAc complex, the interaction equivalent to that of 73B4 Q375/78D2 Q381 was performed in MurG by a similarly located Q289²¹⁸. Q289 was observed interacting with –OH at C3 position of GlcNAc but not interacting with the *N*-acetyl group at C2 position, moreover, no AA was observed interacting with the *N*-acetyl group via hydrogen bond at C2 position of GlcNAc²¹⁸. Besides, another GlcNAc transfer enzyme UGT3A1 also showed Q394 (the equivalent site of 73B4 Q375 and 78D2 Q381) was responsible for GlcNAc-O3 binding (docking data)²²³. Based on these previous studies, 73B4 Q375 and 78D2 Q381 were likely to be in close proximity to the C2 and C3 positions of the sugar and bind GlcNAc-O3 in the same fashion to that of template GlcTs and GlcNAcTs (MurG and UGT3A1).

Although previous studies showed that there was no specific binding between C2 position of UDP-GlcNAc and enzyme, one would be interested in that what makes UDP-GlcNAc rarely used in most GlcTs. A previous research provided the clue: they observed that analogues with bulkier substitutions at C2 position of UDP-GlcNAc could not be used by human short OGT²²⁴, indicating that steric occlusion at C2 position might be the interaction pattern of GlcNAc. Thus, it was suggested that the interaction between target UGTs (73B4 and 78D2) and C2 position of GlcNAc was simple steric occlusion.

A further question is what kind of ‘steric occlusion’ can accommodate the *N*-acetyl group at C2 position of UDP-GlcNAc, and why only the active sites of 73B4 and 78D2 in all target UGTs can accommodate the *N*-acetyl group at C2 position. A close observation and comparison between GlcTs and GlcNAcTs was conducted subsequently.

In the previous study of MurG+UDP-GlcNAc crystal complex²¹⁸, an ‘HEQN’ structure was observed to directly cross from the hexose-binding site and provided the space to accommodate GlcNAc. This ‘HEQN’ AA sequence was conserved in target UGTs (Figure 4.12) and VvGT1³⁵ (a template GlcT with weak UDP-GlcNAc activity): H (AA 19 within the PSPG motif) – E (AA 27 within the PSPG motif) – Q (AA 44 within the PSPG motif) – N (AA 3 outside of the PSPG motif). The presence of ‘HEQN’ AA sequence provided the basic primary structure of active sites of UDP-GlcNAc in 73B4 and 78D2. However, as only 73B4 and 78D2 could use UDP-GlcNAc in tested UGTs, it was inferred that the secondary structure of ‘HEQN’ structure might vary in target UGTs and only ‘HEQN’ of 73B4 and 78D2 could form the conformation that provided the enough space to accommodate GlcNAc.

	GlcNAcTs	MurG	R S G A L T V S E I A A A G L P A L F V P F Q H K D R Q Q Y W N
		VvGT1	H C G W N S L W E S V A G G V P L I C R P F F - - - G D Q R L N
GlcTs (weaker GlcNAcT activity)		73B4	H C G W N S T L E G I A A G L P M V T W P M G - - - A E Q F Y N
		78D2	H C G W N S V L E S V S G G V P M I C R P F F - - - G D Q R L N
		78G1	H S G W N S V L E C I V G G V P M I S R F F F - - - G D Q G L N
		73B5	H C G W N S A I E G I A A G L P M V T W P M G - - - A E Q F Y N
		73C1	H C G W N S T L E G I T S G V P L L T W P L F - - - G D Q F C N
GlcTs		73C5	H C G W N S T L E G I T A G L P L L T W P L F - - - A D Q F C N
		73C6	H C G W N S T L E G I T A G L P M L T W P L F - - - A D Q F C N
		76E11	H C G W N S T L E S I G E G V P M I C K P F S - - - S D Q M V N
		76E12	H C G W N S T L E S I G Q G V P M I C R P F S - - - G D Q K V N
		89B1	H C G W N S V V E A V V A G V L M L T W P M R - - - A D Q Y T D

Figure 4.12: MSA of HEQN structure. HEQN sequence is shown conserved in both GlcTs and GlcNAcTs. The AA sequence (HKD) for an inserted extension in ‘C2 loop’ is observed in GlcNAcT MurG.

A further close look at the structure comparison between VvGT1 and MurG showed that the HEQN structure in MurG contained a longer loop (280-287: VPFQHKDR) surrounded the C2 position of hexose (named ‘C2 loop’ in this section) prior to the subsequent helix and provided more space to accommodate the *N*-acetyl group than that in VvGT1 (Figure 4.13). The longer ‘C2 loop’ in MurG resulted from an extended AA insertion HKD (284-286 Figures 4.12 and 4.13). As VvGT1 did not have such AA insertion, its ‘C2’ loop (369-374: RPFPGD) was shorter and provided less space. This was why its UDP-GlcNAc activity was much weaker (4%) than UDP-Glc in VvGT1, though the short ‘C2 loop’ could still provide the steric occlusion of *N*-acetyl group. This fact indicated one of reasons of why 73B4 and 78D2 could use UDP-GlcNAc: their HEQN structures contained ‘C2 loop’ that provided a space to accommodate *N*-acetyl group at C2 position of GlcNAc. It is tempting to speculate that an insertion of

'C2 loop' in 73B4 and 78D2 might improve the UDP-GlcNAc activity, as the insertion might provide more space to accommodate *N*-acetyl group.

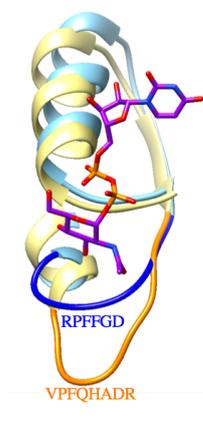


Figure 4.13: HEQN structure and 'C2 loop' comparison between GlcNAcT (MurG, light yellow) and GlcT with weaker UDP-GlcNAc activity (VvGT1, navy blue). The AA sequence for 'C2 loop' was written below respective loop (MurG: amber; VvGT1: navy blue). It is shown that MurG has an extended insertion in 'C2' loop, which might provide more space to *N*-acetyl group at C2 position.

Notably, it was thought that the AA constitution of the 'C2 loop' was not related to the secondary structure of 'C2 loop', i.e. the primary structure in this loop was not related to the secondary structure. 73B5 had almost the same AA sequence of 'HEQN' structure (27 out of 29 AAs are the same) as that of 73B4, especially they have the exactly the same AA sequence PMGAE (73B4 369-373 numbering; 73B5 395-397 numbering) in possible 'C2 loop'. However, UDP-GlcNAc activity was detected in 73B4 but not in 73B5 in the same condition, which suggested that AA sequence in 73B5 might not form a 'C2 loop' with proper secondary structure that is enough to accommodate *N*-acetyl group of GlcNAc while 73B4 could. The contrast example is 73B4 and 78D2: their AA sequences of the 'C2 loop' were very different (73B4 PMGAE 369-373 numbering; 78D2 PPFGD 364-380 numbering), however, it was inferred the secondary structure of their 'C2 loops' can provide the space to accommodate *N*-acetyl group of GlcNAc as they all possess UDP-GlcNAc activity

Apart from the 'C2 loop' discussed above, we are interested in whether there is an AA that discriminates Glc and GlcNAc specificity might be located at other sites. For example, a detailed analysis of the AA sequence of UGT3A1 (an GLCNACT) suggested that N391 might act in GlcNAc recognition of an AA²²³. The idea came from the observation that F was conserved in all GlcTs in the same family whilst N was observed in 3A1 (GlcNAcTs) at the equivalent site. Further mutation of N391F

successfully enhanced UDP-Glc activity but decreased its UDP-GlcNAc activity²²³. Interestingly, at the equivalent site of 3A1 N391, G371 and F378 were found in 73B4 and 78D2 respectively, but not the 'N' that potentially specifically recognised UDP-GlcNAc. This may explain why 73B4 and 78D2 do not recognize UDP-GlcNAc specifically and prefer UDP-Glc as their native donor rather than UDP-GlcNAc.

Alanine mutants: UDP-Gal activity

The 73C5 Q398A mutant showed the absence of UDP-Gal activity (Figure 4.10). This indicated that Q398 is probably a ckAA in 73C5 when using UDP-Gal. This result is consistent with the suggestion from MSA that Q398 may locate near the sugar C2 and C3 positions, as the mutant abrogated both UDP-Gal and UDP-Glc activities.

Contrary to 73C5, Q381 in 78D2 behaved differently towards UDP-Glc and UDP-Gal (Figure 4.10). 78D2 Q381A retained 58% of the original UDP-Glc activity but 0% original UDP-Gal activity. This may simply result from UDP-Glc being the main substrate for activity of 78D2 and being more resilient to mutation. Kinetic results showed that k_{cat} and k_{cat}/K_M of UDP-Gal were 510-fold and 29,000-fold lower than those of UDP-Glc (c.f. Chapter 3). Consequently, UDP-Glc is the main substrate with activity for 78D2. As previous studies suggested, the main activity of an enzyme is 'robust' towards mutation, whereas promiscuous activities are more 'plastic' from the aspect of evolution²²⁵. Therefore, UDP-Gal activity may show less tolerance to the introduction of a variance in the enzyme.

The only difference between Glc and Gal is in the orientation of the –OH group at the C4 position. However, it was reported that 'Q' (the AA binding for the C2 and C3 positions) may differentiate Glc and Gal⁹⁴. Much research has indicated that Q recognised Glc specifically whilst H recognised Gal specifically¹⁰³. Recently, an analysis of UGT8A1 (a GalT with 'H') gave the possible explanation of why 'Q' had a weaker binding with Gal but 'H' had a much stronger binding with Gal²²⁶. In their research, UDP-Gal was docked into the donor binding region of a UGT8A1 model. MD simulations of the full-length UGT8 model (with VvGT1 as the template) showed that 'H' formed a salt-bridge with its adjacent 'D', which acted by 'locking' the D-H dipeptide in the ligand free structure of UGT8A1. This might influence the spatial requirement for the preferential binding of Gal over Glc. The fact that UDP-Glc was

the main activity observed in 78D2 in this project might be aligned with the fact that ‘Q’ appears in this site of 78D2.

However, there are exceptions that ‘Q’ does not relate to or distinguish UDP-Glc and UDP-Gal specificities. For example, VvGT6 could use these two donors with Q373, and even displayed a slightly higher UDP-Gal activity against quercetin (1.4-fold and 3.7-fold higher compared to UDP-Glc in k_{cat} and k_{cat}/K_{donor} , respectively)¹⁰⁴. This raised a speculation that Q might not be the only AA to recognise Glc. This speculation was supported by the Pramod research group recently. They proposed that the AA 22 within the PSPG motif might be a potential AA to recognise the orientation of the –OH at the C4 position specifically. W was conserved as the AA 22 within the PSPG motif in UGT 2B7, 2A3 and 8A1 (all are GlcTs), and the backbone –NH of W was observed only to recognise the specific orientation of O4 sugars such as UDP-GlcUA, UDP-Glc, and UDP-Xyl and interacts with the sugar ring with a hydrophobic effect²²⁶. At the equivalent site of target UGTs, W was also observed, which possibly explained why target UGTs recognised UDP-Glc specifically rather than UDP-Gal (with the exception that 73C5 and 78D2 have UDP-Gal activity but with a much lower preference).

Effect on Q/N swap and Q/H swap

The detailed functional roles of Q were examined by replacing Q with AAs with similar properties – N (Figure 4.14).

Q→N

	UDP-Glc	UDP-Gal	GDP-Glc	GDP-Fuc	UDP-GlcNAc	UDP-GalNAc	UDP-Rha	dUDP-Rha
73B4 Q374N	110 ^{WT} (±8)				12 ^{WT} (±5)			
73B5 Q398N	110 ^{WT} (±7)							
73C1 Q394N	WT							
73C5 Q398N	WT	WT						
73C6 Q398N	<1 ^{WT} (±<1)							
76E11 Q374N	7 ^{WT} (±1)							
76E12 Q376N	38 ^{WT} (±1)							
78D2 Q381N	82 ^{WT} (±13)	WT	WT		120 ^{WT} (±13)			
89B1 Q390N	88 ^{WT} (±5)							

Figure 4.14: Relative activities of mutations (WT activity set as 100%) of ckAAs towards the C2 and C3 positions of sugar moiety (Q/N swap). The green rectangle depicts that original activities were retained at level 4 (>80%); the yellow rectangle indicates original activities were retained at level 2 (20%-50%); the pink rectangle indicates original activities were retained at level 1 (<20%); the red rectangle indicates original activities were retained at level 0 (no activity). The olive green 'WT' indicates positive activity was detected in the WT (data from Chapter 3).

N is very similar to Q in chemistry, but with one CH₂ of side chain being shorter. The mutant will best maintain the environment as well as the structure. This is why Q→N mutants relatively restored more original activity than Q→A mutants generally. However, a Q/N swap shortens the side chain of the AA, potentially increasing the distance between the enzyme and substrate therefore reducing the interaction. Hence, the retained activities for Q/N swap mutants might suggest the distance between the enzyme and substrate.

Q/N swap mutants: UDP-Glc activity

The Q/N swap resulted in a significant UDP-Glc activity loss in the following enzymes: 73C6 (1%) and 76E11 (7%) retained <10% of the original activity and 73C1, 73C5, and 76E12 lost all of UDP-Glc activity (Figure 4.14). The significant activity loss implied that 73C6 Q398, 76E11 Q394, 73C1 Q394, 73C5 Q398, and 76E12 Q376 may

be positioned in a place where a longer distance between the enzyme and the substrate cannot be tolerated. Similar results were found in previous research. For example, VvGT1 Q375 showed a distance of 3.0 Å and 2.9 Å between Q and Glc-O2 and Glc-O3 respectively in the WT. This distance is classified as a ‘moderate, mostly electrostatic’ hydrogen bond distance²²⁷. Q/N swap of VvGT1 resulted in only 1% of original activity (k_{cat}/K_M) being retained. Among all possibilities, it was very likely that the shorter side chain of N led to a greater distance between N and Glc in the mutant: with VvGT1 Q375N as the model (VvGT1 WT as the template), distances between N375 and Glc-O2 and Glc-O3 were changed to 3.5 Å (weak, mostly electrostatic hydrogen bonding) and 4.2 Å (out of the hydrogen bonding distance) respectively.

However, 78D2 Q381N (82%) and 89B1 Q390N (88%) retained >80% of the original activities, showing that the shortening of the side chain from Q to N can be tolerated. Unexpectedly, 73B4 Q374N and 73B5 Q398N showed that original activities were enhanced (110% and 110%, respectively). This is not consistent with the hypothesis that the shorter side chain of the AA will increase the distance from the AA to the substrate therefore reducing the original activity.

To explain this observation, it was considered that the mutation may be helpful to UDP-Glc activity. Figure 4.15 shows the possibility raised to explain how the mutant probably enhances UDP-Glc activity. The side chain atom of NE2 in the WT 73B4 Q374 and 73B5 Q398 interacted with the Glc-O2 and Glc-O3 originally. When Q was mutated to N, the distance between the side chain atom of NE2 in Q and Glc did indeed increase, however, the distance between the side chain atom of OD1 in N and Glc decreased. Thus, N might bring an additional atom (OD1) into the interaction, which compensated for the loss of activity caused by the longer distance. In 73B4 Q374N and 73B5 Q398N, the compensation of a binding effect became even stronger than previous binding, showing an enhanced activity. Additionally, protein may have enough flexibility that would allow site to change its conformation and distinguish changes due to the shorter side chain of N.

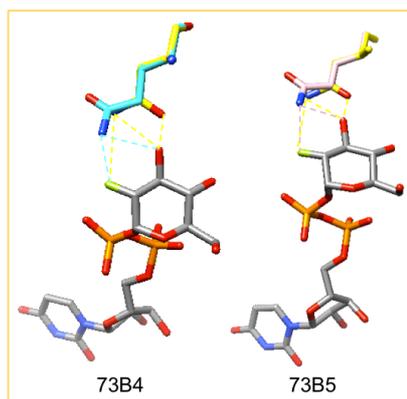


Figure 4.15: Overlay of modelled WT and Q→N mutants. Left: Overlay of 73B4 WT (cyan) and 73B4 Q374N (yellow). Right: Overlay of 73B5 WT (pink) and 73B5 Q398N (yellow). This Figure demonstrates that the OD1 of N in 73B4 Q374N and 73B5 Q398N exert additional binding to Glc-O3.

Q/N swap mutants: UDP-GlcNAc activity

Q/N swap in 73B4 decreased the original UDP-GlcNAc activity to 12%, implying that the longer distance between the AA and GlcNAc can be tolerated partially.

In 78D2, it was unexpected to find Q381N led to enhanced activity. This was possibly achieved if N could better bind UDP-GlcNAc with a strengthened hydrogen bonding (discussed above).

Q/N swap mutants: UDP-Gal activity

Q/N swap of 73C5 and 78D2 lost all UDP-Gal activities (Figure 4.14). This indicated that Q in 73C5 (numbering 398) and 78D2 (numbering 381) might interact with Gal with a limit of hydrogen bonding distance so that the longer distance between N and Gal cannot be tolerated.

Q/H swap

Apart from a potential role of binding the –OH at C2 and C3 positions of the sugar, Q was previously proposed to act as a UDP-Glc recognition AA (discussed above) and new activities might be conferred by point mutagenesis at this site. Some previous studies showed that a Q/H swap might switch the activity. For example, the point mutant H374Q of *Aralia cordata* UDP-Gal anthocyanin GalT successfully conferred new UDP-Glc activity¹⁰³. UGT8-Q383H decreased its original UDP-Glc activity by 50% to all bile acids but enhanced UDP-Gal activity²²⁸. However, there has always

been the debate whether point mutagenesis by a Q/H swap can surely switch Glc or Gal transfer activity. Unsuccessful cases have been observed as well. 71G1 Q382H^{76,229} and flavonoid GlcT UBGT Q382H¹⁰³ did not acquire UDP-Gal activity by point mutagenesis. This paradox might be due to the set of sugar recognition AAs being difficult to be substituted thoroughly during the evolution process, as many ckAAs are evolved to stay conserved in the active site²³⁰.

Herein, Q/H swap was subjected in some of the target UGTs, to investigate whether new donor specificity would be conferred in these UGTs (Figure 4.16).

	UDP-Glc	UDP-Gal	GDP-Glc	GDP-Fuc	UDP-GlcNAc	UDP-GalNAc	UDP-Rha	dTDP-Rha
73B4 Q374H	3 WT (±1)				WT			
73C1 Q394H	WT							
73C5 Q398H	WT	WT						
73C6 Q398H	WT							
78D2 Q381H	9 WT (±1)	WT	WT		1 WT (±1)			

Figure 4.16: Relative activities of mutations (WT activity set as 100%) of ckAAs towards the C2 and C3 positions of the sugar moiety (Q→H). The pink rectangle indicates original activities were retained at level 1 (<20%); the red rectangle means original activities were retained at level 0 (no activity). The olive green 'WT' indicates positive activity was detected in the WT (data from Chapter 3).

Although original UDP-Glc activities were greatly decreased, the Q/H swap did not confer any new activity. For 73C5 and 78D2 with UDP-Gal activity originally in the WT, a Q/H swap did not enhance their UDP-Gal activity but abrogated it conversely. The hypothesis that a Q/H swap leading to Glc/Gal specificity swap has not been demonstrated in 73B4, 73C1, 73C5, 73C6 and 78D2 in the current conditions.

4.2.2 ckAAs towards the C3 and C4 positions of the sugar moiety (the AA 43 within the PSPG motif)

According to previous studies in template UGTs, the AA 43 within the PSPG motif was responsible for binding –OH groups at C3 and C4 positions of Glc via hydrogen bonds (Table 4.4). A negatively charged AA (D/E) was conserved at this site.

Table 4.4: ckAAs that bind –OH of C3 and C4 positions of the sugar in solved crystal structures

Plant UGT	ckAAs	Result origin
VvGT1 ³⁵	D374	PDB (2C1Z)
UGT 72B1 ⁸¹	E388	PDB (2VCE)
UGT 71G1 ⁷⁶	E381	PDB (2ACW)
UGT 78G1 ⁷⁸	D376	Docking
UGT 85H2 ⁷⁷	D402	Docking
UGT 78K6 (<i>Ct3GT-A</i>) ^{79,80}	D367	Overlay on VvGT1
OS79 ⁸²	D389	PDB (5TMD)

4.2.2.1 MSA

With the seven template UGTs, MSA was generated on the nine target UGTs (Figure 4.17), to indicate the potential ckAAs towards C3 and C4 of the sugar moiety in target UGTs.

VvGT1	C	R	P	F	F	G	D	Q
71G1	T	V	P	I	Y	A	E	Q
72B1	A	W	P	L	Y	A	E	Q
78G1	S	R	F	F	F	G	D	Q
78K6	C	R	F	F	F	G	D	Q
85H2	C	W	P	F	F	A	D	Q
OS79	A	M	P	H	W	A	D	Q
73B4	T	W	P	M	G	A	E	Q
73B5	T	W	P	M	G	A	E	Q
73C1	T	W	P	L	F	G	D	Q
73C5	T	W	P	L	F	A	D	Q
73C6	T	W	P	L	F	A	D	Q
76E11	C	K	P	F	S	S	D	Q
76E12	C	R	P	F	S	G	D	Q
78D2	C	R	P	F	F	G	D	Q
89B1	T	W	P	M	R	A	D	Q
	:	.	:	:	:	:	*	

Figure 4.17: MSA of potential ckAAs towards C3 and C4 positions of the sugar moiety. The colours of residues represent their different properties: small and hydrophobic AAs AVFPMILW are coloured red; acidic AAs DE are coloured blue; basic AAs RK are coloured magenta; hydroxyl/sulphydryl/amine AAs STYHCNGQ are coloured green. The last line symbols denote the degree of conservation: conserved (*), conserved mutations (:), semi-conserved mutations (.), and non-conserved mutations (). This Figure demonstrates that the potential ckAAs towards C3 and C4 positions of the sugar moiety are highly conserved as D/E.

From the MSA result, it was observed that a negatively charged AA (D/E) is conserved as the AA 43 within the PSPG motif in target UGTs. With the exceptions of 73B4 and

73B5 showing E, the remaining target UGTs show D. Both D and E are negatively charged, only differing in structure that the side chain of E has one more CH₂ group.

4.2.2.2 SDM

Alanine scanning

The potential ckAAs associated with the C3 and C4 positions of the sugar moiety in target UGTs were replaced by alanine individually. Their relative activities compared to the WT were examined and summarised in Figure 4.18.

	UDP-Glc	UDP-Gal	GDP-Glc	GDP-Fuc	UDP-GlcNAc	UDP-GalNAc	UDP-Rha	dTDP-Rha
73B4 E373A	7 WT (±7)				WT			
73B5 E397A	16 WT (±13)							
73C1 D393A	WT							
73C5 D397A	WT	WT						
73C6 D397A	WT							
76E11 D373A	WT							
76E12 D375A	WT							
78D2 D380A	92 WT (±10)	<1 WT (±<1)	<1 WT (±<1)		10 WT (±1)			
89B1 D389A	<1WT (±<1)							

Figure 4.18: Relative activities of mutations (WT activity set as 100%) of ckAAs towards C3 and C4 positions of the sugar moiety (alanine scanning). The green rectangle indicates original activities were retained at level 4 (>80%); the pink rectangle shows original activities were retained at level 1 (<20%); the red rectangle means original activities were retained at level 0 (no activity). The olive green 'WT' indicates positive activity was detected in the WT (data from Chapter 3).

Alanine mutants: UDP-Glc activity

As shown above, the alanine mutation caused 73C1, 73C5, 73C6, 76E11, and 76E12 to lose their UDP-Glc activities completely. In addition, 89B1 (<1%), 73B4 (6.5%), and 73B5 (16%) retained their original activities <20% (level 1). Taking the MSA results together, the devoid or great decrease of activities by mutation can be explained by D/E→A preventing the hydrogen bonding between the D/E and Glc-O3 and Glc-

O4 (Figure 4.19). Thus, D/E are inferred as ckAAs in 73B4 (numbering 373), 73B5 (numbering 397), 73C1 (numbering 393), 73C5 (numbering 397), 73C6 (numbering 397), 76E11 (numbering 373), 76E12 (numbering 375) and 89B1 (numbering 389).

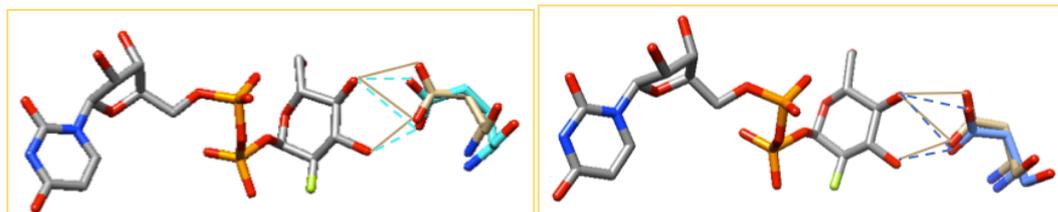


Figure 4.19: ckAA (D/E) towards Glc-O3 & Glc-O4 in modelled UGTs and crystal VvGT1 complex with U2F (PDB: 2C1Z). Left: Overlay of modelled 73B4 (cyan) E373 and crystal VvGT1 (tan) D374; Right: Overlay of modelled 73C1 (cornflower blue) D393 and crystal VvGT1 (tan) D374. The interaction of VvGT1 and model UGTs towards U2F is represented by a solid and dashed line, respectively. This Figure demonstrates that 73B4 E373 and 73C1 D393 bind Glc-O3 and Glc-O4 in the same manner as does VvGT1 D374.

However, one exception was observed – 78D2. 78D2 D380A showed no significant change in UDP-Glc activity (92%), which meant that the elimination of the side chain of D/E was not sufficient to reduce the UDP-Glc activity of 78D2 significantly. Thus, D380 is not inferred as a ckAA in 78D2, though the MSA result indicated that it may be involved in binding the substrate. It is suggested that D380 may be remote from the active site where the substrate is bound.

Alanine mutation at the D/E site revealed that D/E plays an important role as a ckAA in target UGTs with UDP-Glc activity (except for 78D2). Apart from the target UGTs above, previous studies also revealed the significant role of D/E. For example, UGT 73A5 E378¹⁷⁵ and 85B1 E410⁹³ showed they were responsible for Glc-O3 and Glc-O4 binding and their alanine mutants significantly reduced their UDP-Glc activities. Apart from plant UGTs, similar observations were also found in bacterial UGTs (e.g. GtfA, GtfB)^{231,232} and human UGTs²¹¹. For example, the alanine mutation of UGT1A6 (originating from humans but has a distant homologous relationship to the GT1 class) resulted in E379 retaining only 14% of its original activity²¹¹.

Alanine mutants: UDP-GlcNAc activity

Mutant 73B4 E373A had a dramatic effect on UDP-GlcNAc activity (0% activity retained), indicating E373 may be a ckAA in 73B4 with UDP-GlcNAc activity. Taking

the MSA result together with this finding, it indicated that E373 is probably located in proximity to C3 and C4 positions of the sugar and interacts with the sugar.

Reviewing data in the previous section, it is observed that the alanine mutation of 73B4 at site 373 had a different effect on UDP-Glc and UDP-GlcNAc: 6.5% and 0% retained respectively. This difference indicated that E373 can be more helpful in binding UDP-GlcNAc. A further hypothesis is that the UDP-GlcNAc activity is the promiscuous activity of 73B4 and therefore the activity is not as robust as the main activity (UDP-Glc), as discussed above.

78D2 D380A had a significantly different effect on UDP-Glc and UDP-GlcNAc activities: 92% and 10% respectively. This result suggests that the role of D380 in 78D2 is dependent on substrate. It is suggested that D380 may be remote from the active site where UDP-Glc is bound but be in close proximity to where UDP-GlcNAc is bound. Although the two substrates are very likely to be orientated in the same donor binding region as discussed above, the detailed orientation of UDP-Glc and UDP-GlcNAc in the region may have a shift, leading to the distance between D380 and Glc being out of the range that a hydrogen bond can form but that between D380 and GlcNAc is within the range that a hydrogen bond can form. The orientation shift of different substrates in the same region was observed in previous research. For example, in the UGT3A1 docking experiment, the distances between H35 to C1 position of UDP-Glc and UDP-GlcNAc were different (4.02 Å and 6.21 Å respectively), though the two donors were located in the same region²²³.

In comparison with other GlcNAcTs, AAs located at the equivalent site of 73B4 E373 and 78D2 D380 in GlcNAcTs were also observed to bind GlcNAc-O3 and -O4. For example, the crystal structure of the complex of MurG-UDP-GlcNAc (PDB: 1NLM) showed that polar residues in MurG anchored the -OH groups of C3 and C4 of GlcNAc, with Q288 specifically interacting with GlcNAc-O3 and -O4 by forming a hydrogen bond²¹⁸. Another example was UGT3A1, which showed an interaction between D393 and GlcNAc-O3 and -O4 (docking experiment)²²³. These studies indicated that 73B4 E373 and 78D2 D380 might also be in close proximity to C3 and C4 positions of GlcNAc and acted as ckAAs to help bind GlcNAc.

Alanine mutants: UDP-Gal activity

Mutant 73C5 D397A retained no UDP-Gal activity, implying that D397 is crucial to UDP-Gal activity. Based on the results from MSA, 73C5 D397 is proposed to be located in the vicinity of the C3 and C4 positions of the sugar. The interaction of 73C5 D397 with Gal might be analogous to the D interaction with sugar hydroxyls seen in GlcTs, which was observed in the docking experiment of UGT8A1 that D382 interacted with Gal-O3 and Gal-O4²²⁶. However, as the UDP-Gal activity is much weaker than UDP-Glc activity (The values of k_{cat} and k_{cat}/K_A of UDP-Gal is less than that of UDP-Glc with 480-fold and 50-fold respectively), the binding between D397 and Gal-O4 might be very weak as the orientation of –OH at C4 position varies in UDP-Gal compared to UDP-Glc.

Alanine mutation of 78D2 D380 showed that less than 1% original UDP-Gal activity was retained. It is possible that 78D2 D380 interacted with UDP-Gal in the same pattern as that for 73C5 D397, thus playing a catalytic key role in UDP-Gal activity. Or, the great reduction of UDP-Gal activity by alanine mutation was due to UDP-Gal being a promiscuous substrate (reflected in kinetics, in Chapter 3) and being ‘plastic’ towards mutation from the aspect of evolution (discussed above). This suggestion was inferred from the different effect of 78D2 D380A towards UDP-Glc and UDP-Gal (92% and <1% original activity retained).

In other GalTs, D has been found to be highly conserved as the AA 43 within the PSPG motif, such as *Petunia hybrida* GalTs (Accession No. AF165148), *Rattus norvegicus* ceramide GalTs (Accession No. L21698), and *Homo sapiens* ceramide GalTs (Accession No. U62899). The invariant ‘D’ indicated its conserved catalytic key role in Gal transfer activity.

Effect on D/E swap

A swap between D and E was tested (Figure 4.20). The two AAs have similar properties with negatively charged side chains. The only structural difference between these two AAs is the length of the side chain, as E has a longer side chain with one additional CH₂. It was expected that the D/E swap might restore some original activity and more than alanine mutation. The result of the D/E swap might indicate the distance between D/E and substrate by checking whether the shorter/longer distance between the AA and the substrate can be tolerated.

	UDP-Glc	UDP-Gal	GDP-Glc	GDP-Fuc	UDP-GlcNAc	UDP-GalNAc	UDP-Rha	dTDP-Rha
73B4 E373D	30 WT (±3)				WT			
73B5 E397D	35 WT (±6)							
73C1 D393E	WT							
73C5 D397E	7 WT (±2)	WT						
73C6 D397E	35 WT (±5)							
76E11 D373E	57 WT (±8)							
76E12 D375E	28 WT (±3)							
78D2 D380E	86 WT (±9)	<1 WT (±<1)	<1 WT (±<1)		52 WT (±10)			
89B1 D389E	WT							

Figure 4.20: Relative activities of mutations (WT activity set as 100%) of ckAAs towards the C3 and C4 positions of sugar moiety (D/E swap). The green rectangle means original activities were retained at level 4 (>80%); the light green rectangle indicates original activities were retained at level 3 (51%-80); the yellow rectangle shows original activities were retained at level 2 (20%-50%); the pink rectangle means original activities were retained at level 1 (<20%); the red rectangle indicates original activities were retained at level 0 (no activity). The olive green 'WT' indicates positive activity was detected in the WT (data from Chapter 3).

D/E swap mutants: UDP-Glc activity

73B4 E373D and 73B5 E397D lost UDP-Glc activity partially (30% and 35%, respectively), probably due to the longer distance between the AA and Glc caused by the shortening of the side chain. Combining the previous inference that E may be one of the ckAAs in 73B4 and 73B5 by binding Glc-O3 and Glc-O4, the D/E swap result indicated that E of 73B4 and 73B5 might interact with Glc closely and the interaction is strong so that the longer distance can be tolerated (Figure 4.21, left).

D/E swap of 89B1, 73C1, and 73C5 completely or significantly reduced their UDP-Glc activities. This result suggested that D in 89B1, 73C1, and 73C5 are located very close to Glc so that no additional place could tolerate one additional CH₂. Thus, when D was replaced by E, the extended side chain might be too long and too flexible in the active site, causing steric hindrance and inhibiting the interaction with UDP-Glc. By comparing the 89B1 WT (model) and 89B1 D389E (model) (Figure 4.21, right), it was

observed that the side chain bond CB-CG of E in the mutant did not position in the same way as the side chain bond CB-CG bond of D in the WT, but rotated to another direction with an actual increasing distance towards the donor (which may be due to a repulsive effect produced by the longer side chain). This explains why the extended side chain did not decrease the distance to enhance the activity as expected, but instead caused the reduction of activity.

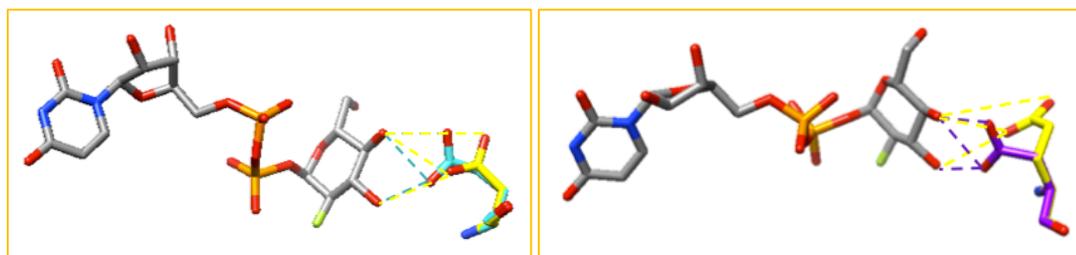


Figure 4.21: Overlay of modelled WT UGTs and D/E swap mutants. Left: Overlay of 73B4 WT (cyan) and 73B4 E373D (yellow). Right: Overlay of 89B1 WT (purple) and 89B1 D389E (yellow). This Figure demonstrates 1) In 73B4, the shorter chain of D causes a longer distance towards Glc-O3 and Glc-O4, thus reducing activity; 2) In 89B1, the longer chain of E has a rotation of CB-CG bond and removes the binding effect towards Glc-O3 and Glc-O4 as WT, thus destroying the activity.

However, the mutation of D was not fatal for all UGTs, as 73C6 D397E (35%), 76E11 D373E (57%), and 76E12 D375E (28%) still retained partial activities. This indicated that the distance between D in 73C6, 76E11 and 76E12 and Glc was not close as the UGTs above (89B1, 73C1, and 73C5), which had space to tolerate one additional CH₂.

D/E swap mutants: UDP-GlcNAc activity

73B4 E373D lost all UDP-GlcNAc activities. Taking the data above that mutation of E→A also lost UDP-GlcNAc activity, E373 is inferred as crucial to 73B4 towards UDP-GlcNAc activity. Even with this data, still 30% of UDP-Glc activity was retained for mutant E373D, it is therefore suggested that the distance between E373 and Glc/GlcNAc differ. It is possible that E373 is located closer to Glc compared to GlcNAc, thus the longer distance caused by mutation D was tolerated in UDP-Glc activity but not in UDP-GlcNAc activity.

78D2 D380E mutant retained 52% of its original UDP-GlcNAc activity. It is suggested that D380 interacts with GlcNAc with a tolerance for an additional CH₂ chain length.

D/E swap mutants: UDP-Gal activity

Neither 73C5 D397E nor 78D2 D380E showed positive UDP-Gal activity. It is suggested that D in 73C5 and 78D2 interact with UDP-Gal within a close distance, thus, the mutation of E caused steric hindrance and inhibited the interaction with UDP-Glc. Alternatively, the lack of UDP-Gal activity is simply due to promiscuous activity of enzymes being more "plastic" for mutants as discussed above.

4.2.3 ckAAs towards C6 position of sugar moiety

Compared to the high level of conservation of ckAAs at the C2-C4 positions of the sugar moiety, the ckAAs towards the C6 position of the sugar shows more variations in crystal structures (summarised in Table 4.5). In *VvGT1*, T141 interacted with Glc-O6 spatially by crossing the N-terminal domain. At the equivalent site of *VvGT1* T141, ckAAs towards Glc-O6 were also observed in 71G1 and *OS79*. However, unlike most ckAAs towards the donor concentrating in the C-terminal of UGT, the ckAAs towards the C6 position are located in the N-terminal of UGT where most ckAAs towards the acceptor are located. As a consequence of this, more variability happens. Some crystal structures did not show a clear indicator ckAA which is responsible for binding the –OH at the C6 position, such as 72B1 and 85H2.

Table 4.5: ckAAs that bind the –OH of C6 position of the sugar moiety in solved crystal structures

Plant UGT	ckAAs	Result origin
<i>VvGT1</i> ³⁵	T141	PDB (2C1Z)
UGT 72B1 ⁸¹	No AAs near Glc-O6	PDB (2VCE)
UGT 71G1 ⁷⁶	T143	PDB (2ACW)
UGT 78G1 ⁷⁸	T141	Docking
UGT 85H2 ⁷⁷	No data report	-
UGT 78K6 (<i>Ct3GT-A</i>) ^{79,80}	N137	Overlay on <i>VvGT1</i>
<i>OS79</i> ⁸²	S142, Q143	5TMD

4.2.3.1 MSA

As above, MSA was performed on the seven template UGTs and the nine target UGTs (Figure 4.22), to evaluate which of the potential ckAAs in the target UGTs may be involved in binding the –OH group at the C6 position of the sugar.

VvGT1	V	A	W	L	P	F	W	T
71G1	I	P	S	Y	L	F	L	T
72B1	V	P	P	Y	I	F	Y	P
78G1	A	K	W	V	P	L	W	T
78K6	V	P	W	I	A	F	W	P
85H2	L	P	N	V	L	I	F	S
OS79	V	P	P	Y	I	F	Y	S
73B4	V	P	R	L	V	F	H	G
73B5	V	P	R	L	V	F	H	G
73C1	I	P	K	I	L	F	H	G
73C5	I	P	K	I	L	F	H	G
73C6	I	P	K	I	L	F	H	G
76E11	L	P	N	V	I	F	S	T
76E12	L	P	N	I	I	F	S	T
78D2	A	S	W	I	A	F	W	T
89B1	I	P	R	F	D	F	S	P

Figure 4.22: MSA of potential ckAAs towards the C6 position of the sugar moiety. The colours of residues represent their different properties: small and hydrophobic AAs AVFPMILW are coloured red; acidic AAs DE are coloured blue; basic AAs RK are coloured magenta; hydroxyl/sulfhydryl/amine AAs STYHCNGQ are coloured green. This Figure demonstrates that the potential ckAAs towards the C6 position of the sugar moiety are not conserved.

In template UGTs, ckAAs that interacted with Glc-O6 are not highly conserved. Polar uncharged AA T or S (T has an additional CH₃ group) are present in most template UGTs. 72B1 and 78K6 display P at the equivalent site, however, P is hardly involved in directly binding of the substrate via a hydrogen bond²³³.

In target UGTs, 76E11, 76E12, and 78D2 show T at position 134, 135, and 145, respectively. Group D UGTs (73B4, 74B5, 73C1, 73C5, and 73C6) are conserved to show G and 89B1 shows P.

4.2.3.2 SDM

SDM was performed to mutate potential ckAAs and the activities of the mutants were examined (Figure 4.23).

	UDP-Glc	UDP-Gal	GDP-Glc	GDP-Fuc	UDP-GlcNAc	UDP-GalNAc	UDP-Rha	dTDP-Rha
76E11 T134A	WT							
76E12 T135A	WT							
78D2 T145A	104 WT (±15)	<1 WT (±<1)	<1 WT (±<1)		83 WT (±14)			
73B4 G127A	25 WT (±11)				WT			
73B5 G154A	66 WT (±5)							
73C1 G148A	110 WT (±9)							
73C5 G150A	37 WT (±8)	WT						
73C6 G141A	28 WT (±4)							
89B1 P147A	81 WT (±5)							

Figure 4.23: Relative activities of mutations (WT activity set as 100%) of ckAAs towards the C6 position of the sugar moiety (alanine scanning). The green rectangle indicates original activities were retained at level 4 (>80%); the light green rectangle means original activities were retained at level 3 (51%-80); the yellow rectangle demonstrates original activities were retained at level 2 (20%-50%); the pink rectangle means original activities were retained at level 1 (<20%); the red rectangle indicates original activities were retained at level 0 (no activity). The olive green 'WT' indicates positive activity was detected in the WT (data from Chapter 3).

Figure 4.23 shows that 76E11 T134A and 76E12 T125A retained no UDP-Glc activity, indicating the significant role of T. Based on the MSA result above, it is inferred that 76E11 T134 and 76E12 T125 are ckAAs that may interact with Glc-O6. Reviewing previous literature, similar results can also be found. 76E1, a UGT belonging to the same group of 76E11 and 76E12, showed a significant role of T134 with the experimental data that T134A abrogated all UDP-Glc activity and inferred the binding of Glc-O6 based on the MSA result²³⁴.

78D2 T145 did not show its significant role in UDP-Glc and UDP-GlcNAc activity as T145A retained most activity. As the functional group of C6 positions at the hexoses of UDP-Glc, UDP-GlcNAc, UDP-Gal and GDP-Glc are the same, the potential ckAA towards C6 positions of the four donors are expected to be the same. However, the experimental result did not come out as expected but 78D2 T145A lost the original GDP-Glc and UDP-Gal activities while retained most UDP-Glc and UDP-GlcNAc

activities. This confusing result might be explained by comparing the crystal complexes of β 4Gal-T1+UDP-Glc+Mn²⁺ and β 4Gal-T1+UDP-Gal+Mn²⁺¹⁶¹. The research observed that when UDP-Glc was bound, Glc-O4 was pointing away from the side chain of D318, resulting in a space between the C4 position of Glc and the side chain of D318, which was filled by the –OH group at C6 position of Glc. However, if –OH group was in axial orientation as in Gal, it would be pointing towards the side chain of D318, not leaving enough space between Gal-O4 and D381. Thus, the –OH group at C6 position must orient away from D318. As indicated, the allocation of C6 positions of Glc and Gal might be different in the active site of 78D2 thus mutation might affect the activity differently.

Additionally, another possible explanation is that the overall binding of UDP-Gal is very weak (kinetic results shown in Chapter 3), causing the activity to be too plastic to tolerate any change of protein. For GDP-Glc, as the GDP has many differences with UDP, GDP-linked nucleotides may have different accommodation in enzyme as discussed above. Thus, the mutagenesis can cause different activity change.

G is conserved to appear at the equivalent site that may be responsible for binding the C6 position in 73B4, 73B5, 73C1, 73C5, and 73C6. The alanine mutations of these UGTs at this site varied: 73C1 G148A, 73B5 G154A, 73B4 G127A, 73C5 G150A, and 73C6 G151A retained 110%, 66%, 25%, 37%, and 28%, of activities respectively. These variable results may due to the complexity of G function and its characteristics. The side chain of G just has a hydrogen atom, which is predicted not to be involved in substrate binding via a hydrogen bond. This might be the reason why 73C1 G148A did not show any significant activity change. However, G plays a significant role in providing conformational flexibility that may be required in an enzyme-catalysed reaction²³⁵, as an enzyme action generally follows an ‘induced fit’²³⁶. A closer look at the crystal structure of template UGTs found that the ckAAs of Glc-O6 in both VvGT1 (T141) and OS79 (S142) were located in a linker between β 5 and α 5³⁵. Moreover, VvGT1 H150 that is located at the α 5 helix interacted with the acceptor molecule (KMP) directly via a hydrogen bond. Consequently, the point mutation of G in a linker might not only affect the flexibility of the linker thereby affecting the twisting and rotating of the connected domains nearby, but also, the effect of some other ckAAs in connected domains might be affected, therefore changing the enzyme activity. For

example, 73B4 G127A only retained 25% of UDP-Glc activity. At the equivalent site of VvGT1 H150 (ckAA that interacts with KMP, located in the $\alpha 5$ helix that is connected by the linker with T141 on), 73B4 S136 was observed and its alanine mutation retained only 32% activity (data not shown). Consequently, the mutation of 73B4 G127 might affect the connected helix with 73B4 S136 on, thus reducing the activity of UDP-Glc.

However, MSA may not provide highly accurate and reliable potential ckAAs information concerning the C6 position, as potential ckAAs at the C6 position are located in the N-terminal that has a low level of conservation. For example, 78K6 shows P136 from the MSA at the equivalent site, however, its adjacent N137 is observed to bind Glc-O6 by overlaying its apo crystal structure on the crystal structure of VvGT1^{79,80}. In addition, the possible ckAA at C6 position of glucuronic acid in VvGT5 was approved as R140 (via a docking experiment)¹⁰⁴, which was located at the equivalent site of VvGT1 W140 rather than T141.

Consequently, more tests on the adjacent AAs in this region are required to provide more information of the ckAAs involved at the C6 position of the sugar moiety in an individual enzyme.

4.2.4 Cysteines in the PSPG motif – disulphide bridge

In enzyme catalysed reactions, AAs might act as spectator residues to stabilise the conserved motif of enzymes for the proper accommodation of a substrate²³⁷. The crystal structure of 85H2 showed two cysteines (Cs) within the PSPG motif forming a disulphide bridge to stabilise the PSPG motif. Herein, the study of the disulphide bridge is presented, which indicates the potential role of Cs within the PSPG motif in UGTs.

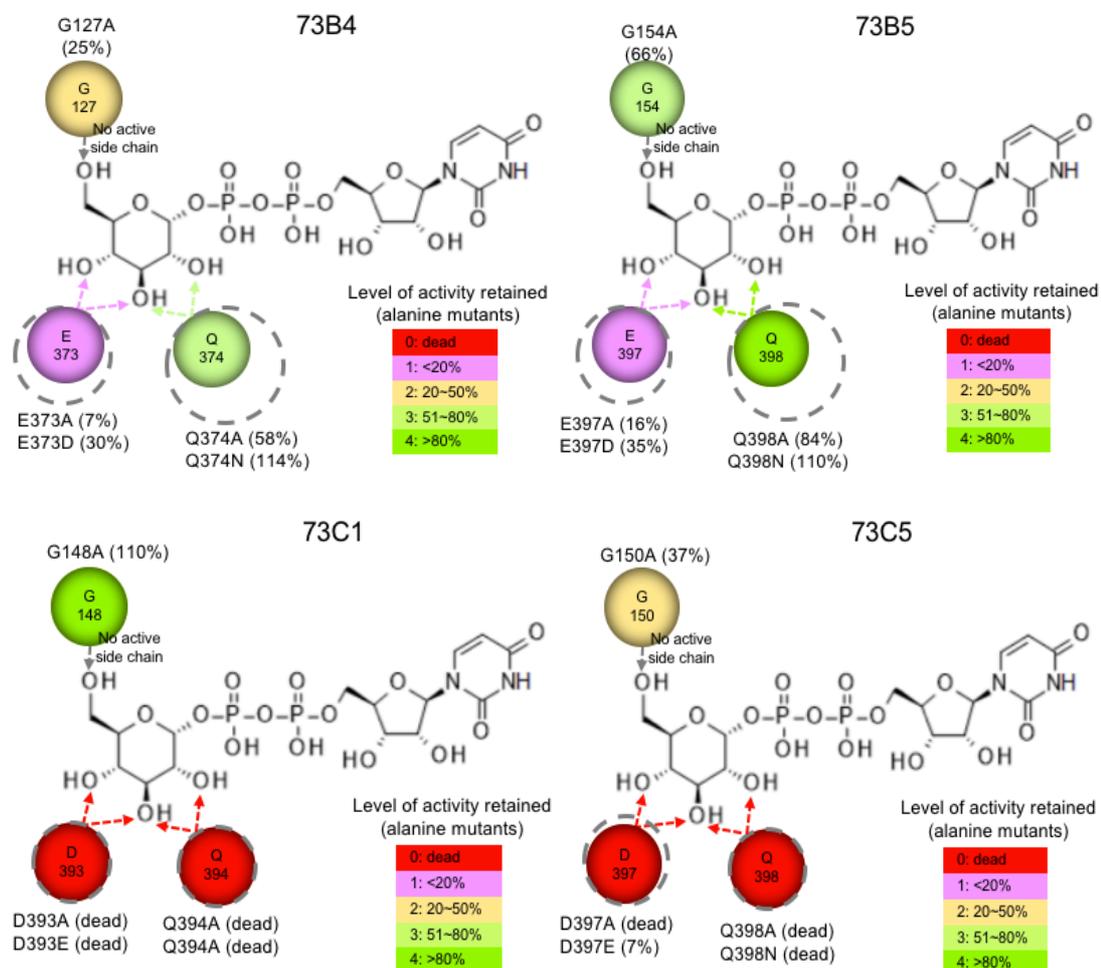
A disulphide bond is defined as a functional group with an R-S-S-R structure, stabilising the structures of enzymes in a great array of enzymatic processes²³⁸. Each disulphide bond may stabilise folded conformation by 2-5 kcal/mol²³⁹. Thus, it can not only reinforce local stability, but also create a suitable condition for local conformational changes relating enzymatic function²⁴⁰. Among 20 AAs, only two AAs incorporate sulphur elements - cysteine (C) and methionine (M). However, M never

Reviewing past literature, the significance of the disulphide bridge has been widely emphasised. Even an engineering called ‘disulphide engineering’ was applied to increase the stability of proteins²⁴¹. For example, Perry and Wetzel mutated I3 to C in T4 lysozyme. Mutant I3C then formed a disulphide with C97, leading to an increased stability of the T4 lysozyme in oxidizing conditions²⁴².

4.2.5 Summary –ckAAs in UGTs (sugar moiety)

UDP-Glc activity

The potential ckAAs in nine target UGTs towards the sugar moiety are summarised in Figures 4.25, 4.26 and 4.27. The potential ckAAs were deduced from MSA. The level of significance of the potential ckAAs in specific UGTs is indicated by the percentage of the retained original activities in the alanine mutants. The distance between ckAAs and substrates was suggested from the similar AA-swapped mutants (D/E and Q/N swap).



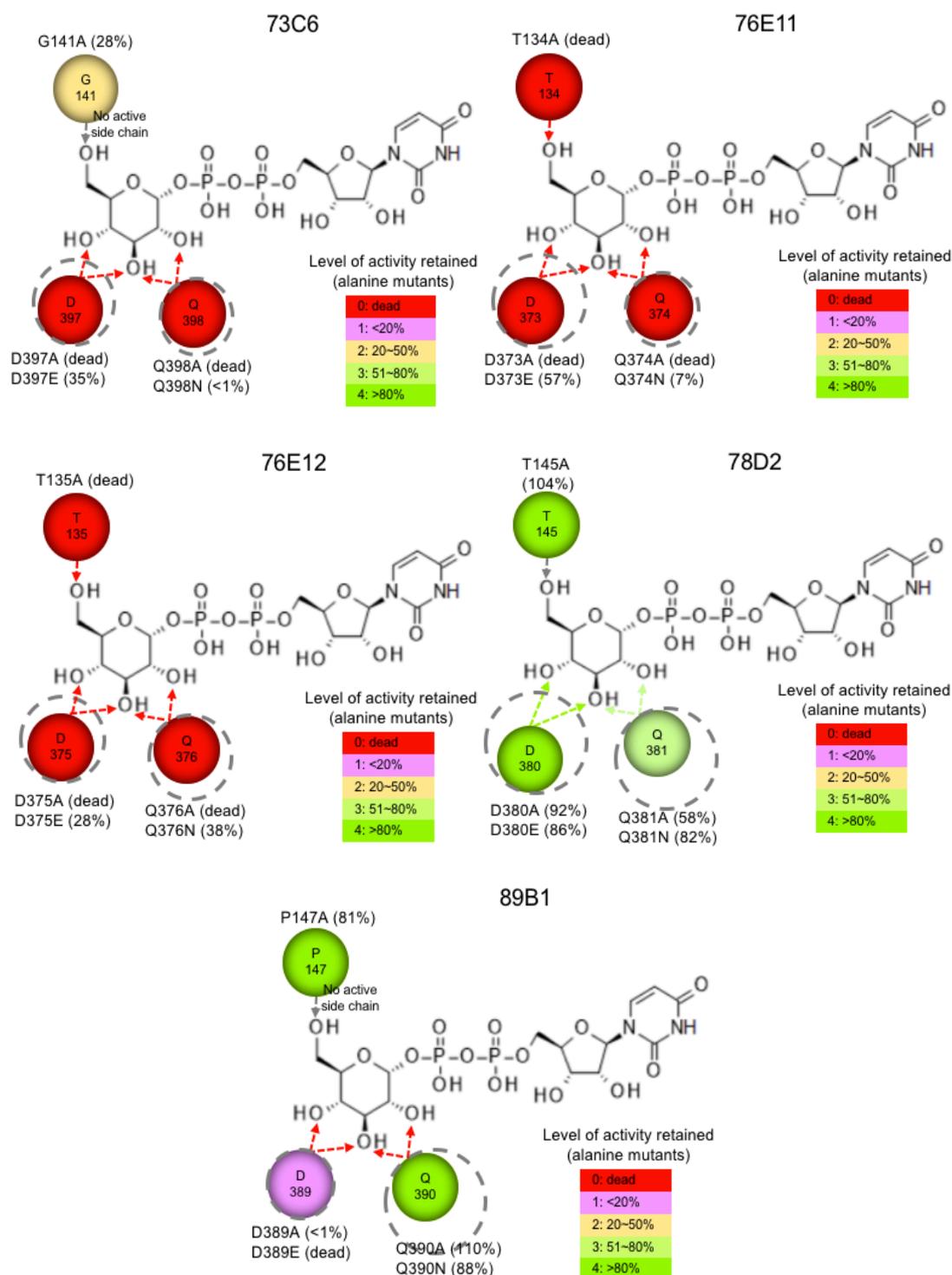


Figure 4.25: Potential cKAAs in UGTs towards UDP-Glc. The level of significance of potential cKAAs (from MSA) is based on the experimental results of alanine scanning and indicates in colour: Level 0, Very Strong (alanine mutants retained original activity 0%, in red); Level 1, Strong (alanine mutants retained original activity <20%, in pink); Level 2, Medium (alanine mutants retained original activity 20~50%, in yellow); Level 3, Weak (alanine mutants retained original activity 51~80%, in light green); Level 4, Very Weak/No (alanine mutants retained original activity >80%, in green). The diameter of the dotted circle suggests the distance (longer/shorter) between potential cKAAs and substrate based on the structure of D/E or Q/N swap.

As Figure 4.25 shows, potential ckAAs D/E-Q play different roles in different target UGTs, although they are conserved in MSA. For example, D/E was predicted to bind Glc-O3 and Glc-O4 in most UGTs, but the experimental results showed that not to be the case in 78D2 as 78D2 D380A retained UDP-Glc activity at level 4. Furthermore, although Q was highly conserved and proposed to bind Glc-O2 and Glc-O3, the elimination of its side chain did not lead to a significant loss of the original activity in 73B5 and 89B1.

Apart from alanine scanning, D/E and Q/N swap showed relatively modest activity loss due to their similarity in the structure. D/E and Q/N swap provided a possible estimation on the tolerance of the distance between AAs and a substrate. For example, 73C1 D393-Q394 showed no tolerance of the distance change from the D-Q motif to the substrate as any longer/shorter side chain of AA caused the absence of activity. However, 73B4, which is from the same group, showed tolerance of the distance change from the D-Q motif to the substrate as its D/E and Q/N swap restored partial original activity.

Unlike a relative clear picture of potential ckAAs towards the C2-C4 positions of the sugar, ckAAs towards the C6 position of the sugar were not highly conserved and remained obscure. At the equivalent site of ckAAs in template UGTs (e.g. *Vv*GT1 T141), only 76E11, 76E12, and 78D2 showed an identical AA (T) at 134, 135, and 145, respectively, and the SDM result substantiated only that T134 and T135 were potential ckAAs in 76E11 and 76E12 respectively. Besides, not all targeted UGTs showed AAs with active side chains (e.g. 73B4 G127 and 89B1 P147) that could be supposed to have a direct binding effect towards the donor via a hydrogen bond. This might suggest that the binding at the C6 position was not necessary or that the ckAA towards the C6 position was not at the equivalent position as in *Vv*GT1. It is recommended that future work should focus on probing the adjacent AAs in this region for individual enzymes.

UDP-GlcNAc activity

The ckAAs in 73B4 and 78D2 towards UDP-GlcNAc is summarised in Figure 4.26.

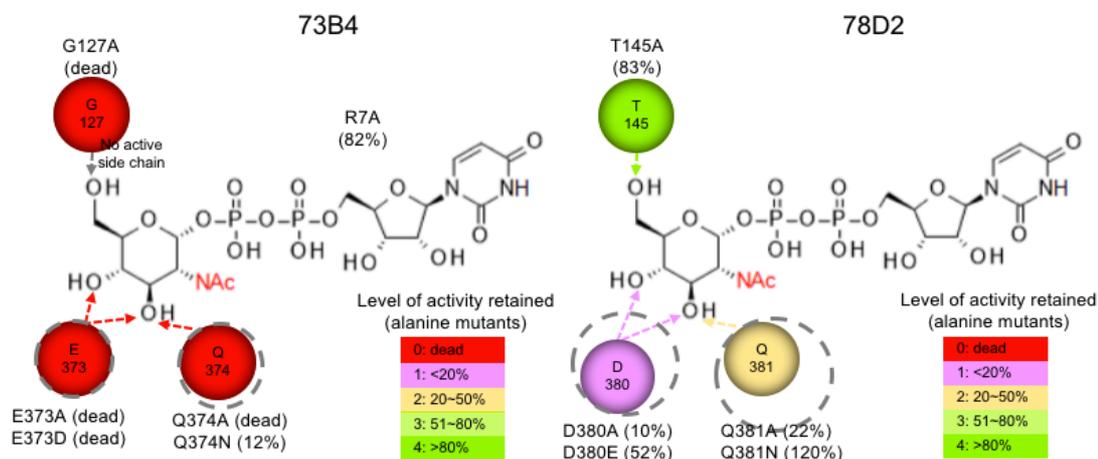


Figure 4.26: Potential ckAAs in UGTs towards UDP-GlcNAc. The level of significance of potential ckAAs (from MSA) is based on the experimental results of alanine scanning and is indicated by colour: Level 0, Very Strong (alanine mutants retained original activity 0%, in red); Level 1, Strong (alanine mutants retained original activity <20%, in pink); Level 2, Medium (alanine mutants retained original activity 20~50%, in yellow); Level 3, Weak (alanine mutants retained original activity 51~80%, in light green); Level 4, Very Weak/No (alanine mutants retained original activity >80%, in green). The diameter of the dotted circle suggests the distance (longer/shorter) between potential ckAAs and the substrate is based on the structure of D/E or Q/N swap.

73B4 and 78D2 had UDP-GlcNAc activity. The prediction from MSA was that D/E-Q bound to the C2-C4 positions of GlcNAc. SDM of 73B4 approved the significant role of D/E-Q as alanine mutations of E-Q abolished original activity. However, mutations of D-Q in 78D2 did not show the same level of activity loss, especially Q381A still retained the original activity at level 2. This data indicated the significance of potential ckAAs that varied in the different enzymes.

The study of ckAAs in 73B4 and 78D2 towards UDP-GlcNAc is important as GlcNAc transfer activity is important in plants. For example, GlcNAc transfer activity is involved in the gibberellin (GA) response pathway and takes part in regulating plant hormones such as the control of many plant processes like germination, growth, flowering, and seed development²⁴³. For example, some *Arabidopsis thaliana* GlcNAcTs mediate GlcNAcylation using UDP-GlcNAc as a donor to viral capsid proteins when infected by the Plum Pox virus²⁴⁴.

UDP-Gal activity

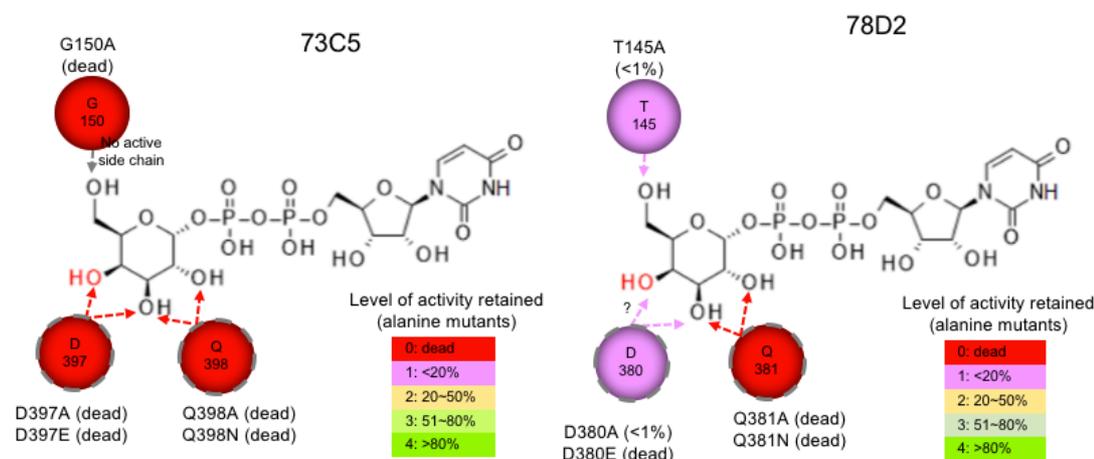


Figure 4.27: Potential cKAAs in UGTs towards phosphate, UDP-Glc (left, c.f. section 4.1), or UDP-GlcNAc (right, c.f. section 4.1). The level of significance of potential cKAAs (from MSA) is based on the experimental results of alanine scanning and is indicated by colour: Level 0, Very Strong (alanine mutants retained original activity 0%, in red); Level 1, Strong (alanine mutants retained original activity <20%, in pink); Level 2, Medium (alanine mutants retained original activity 20~50%, in yellow); Level 3, Weak (alanine mutants retained original activity 51~80%, in light green); Level 4, Very Weak/No (alanine mutants retained original activity >80%, in green). The diameter of the dotted circle suggests the distance (longer/shorter) between potential cKAAs and the substrate is based on the structure of D/E or Q/N swap.

As shown in Figure 4.27, the D-Q motif of 73C5 and 78D2 showed significance in UDP-Gal activity: only 78D2 Q380A retained the original activity at level 1 (<1%), while alanine mutations of D-Q in 73C5 and D in 78D2 completely abrogated original activity. As suggested by the MSA result, Q was located in the proximity to the C2 and C3 positions of the sugar, suggesting that it might interact with Gal-O2 and Gal-O3 as well. D was likely to be positioned in the vicinity of the C3 and C4 positions of the sugar, however, as the orientation of the -OH at the C4 position in Gal is different to that of Glc, the interaction effect of 73C5 D397 and 78D2 D380 binding Gal-O4 might vary.

Generally, UDP-Gal activity was shown to be easier to be reduced by mutation compared to UDP-Glc (e.g. 78D2 Q381A retained 58% original UDP-Glc activity but <1% original UDP-Gal activity), probably resulting from the observation that UDP-Gal was promiscuous activity with 73C5 and 78D2 (kinetic data in Chapter 3). As previous research suggested, promiscuous activity of an enzyme is more ‘plastic’ conferred by the mutation than the main activity of enzyme²²⁵.

4.3 Identification of ckAAs – phosphate moiety

The phosphate moiety is a common part of nucleoside sugar donors. Consequently, ckAAs for phosphate are expected to be highly conserved in UGTs. Previous research showed that T19 (VvGT1 numbering), T280 (VvGT1 numbering), and H350 (VvGT1 numbering) were three ckAAs that bind phosphate, however, T19 showed the least level of conservation (Table 4.6).

Table 4.6: ckAAs that bind the phosphate moiety in solved crystal structures

Plant UGT	ckAAs	Result origin
VvGT1 ³⁵	T19, T280, H350	PDB (2C1Z)
UGT 72B1 ⁸¹	S277, H364	PDB (2VCE)
UGT 71G1 ⁷⁶	S285, H357	PDB (2ACW)
UGT 78G1 ⁷⁸	T25, S282, H352	PDB (3HBF)
UGT 85H2 ⁷⁷	S304, H378	Docking
UGT 78K6 (<i>Ct3GT-A</i>) ^{79,80}	S16, T273, H343	PDB (4WHM)
<i>OS79</i> ⁸²	T291, H361	5TMB

4.3.1 MSA

Based on the AA sequence of template UGTs, ckAAs in the target UGTs towards phosphate were inferred from MSA (Figure 4.28).

VvGT1 T19		VvGT1 T280		VvGT1 H350	
VvGT1	F P F S T	VvGT1	I S F G T	VvGT1	A F V T H
78K6	F P F G S	78K6	V C F G T	78K6	V F V T H
71G1	A P G I G	71G1	L C F G S	71G1	G F V S H
72B1	S P G M G	72B1	V S F G S	72B1	G F L T H
78G1	F P F G T	78G1	I S F G S	78G1	V F L T H
85H2	Y P V Q G	85H2	V N F G S	85H2	G F L T H
OS79	P A A Q G	OS79	V S Y G T	OS79	C F L S H
73B4	K L F A R	73B4	L S F G S	73B4	G F V T H
73B5	F M A Q G	73B5	L S F G S	73B5	G F V T H
73C1	F M A Q G	73C1	V C L G S	73C1	G F L T H
73C5	F M A Q G	73C5	V C L G S	73C5	G F L T H
73C6	F M A Q G	73C6	V C L G S	73C6	G F L T H
76E11	V P A Q G	76E11	V S L G S	76E11	G F W S H
76E12	F P A Q G	76E12	I S M G S	76E12	G F W S H
78D2	F P F G T	78D2	I S F G T	78D2	V F V T H
89B1	F P A Q G	89B1	V C F G S	89B1	A F L T H
.	.	:	*	:	*

Figure 4.28: MSA of potential ckAAs towards the phosphate moiety. The colours of residues represent their different properties: small and hydrophobic AAs AVFPMILW are coloured red; acidic AAs DE are coloured blue; basic AAs RK are coloured magenta; hydroxyl/sulfhydryl/amine AAs STYHCNGQ are coloured green. This Figure demonstrates the degree of conservation of three main ckAAs towards phosphate binding follow the sequence as H250 (VvGT1 numbering), T280 (VvGT1 numbering), and T19 (VvGT1 numbering).

Figure 4.28 shows the degree of conservation of potential ckAAs relating to binding diphosphate. VvGT1 T19 is the least conserved. In the 16 UGTs displayed in Figure 4.28, only five UGTs (VvGT1, 78K6, 78G1, OS79, and 78D2) show T or S. The remaining ten UGTs (except 73B4 showing R) display G that is unlikely to be involved in direct binding via a hydrogen bond.

At the equivalent site of VvGT1 T280, a higher level of conservation was observed. At this site, only T and S were observed. These two AAs have similar properties with a polar uncharged side chain but with the subtle variation that T has an additional CH₃ group compared to S. As S is the fastest-evolving AA²⁴⁵, it is probably that the variability at this site is a consequence of evolution.

VvGT1 H380 is located in the 19th position within the PSPG motif and highly conserved: all UGTs show H in corresponding sites.

4.3.2 SDM

Potential ckAAs towards phosphate in 73B4 (R7, T270, and H349) and 78D2 (T22, T286, and H356) were mutated to alanine and their activities were examined (Figure 4.29).

	UDP-Glc	UDP-Gal	GDP-Glc	GDP-Fuc	UDP-GlcNAc	UDP-GalNAc	UDP-Rha	dTDP-Rha
73B4 R7A	88 WT (±8)				82 WT (±4)			
78D2 T22A	20 WT (±7)	WT	WT		5 WT (±1)			
73B4 S270A	26 WT (±4)				WT			
78D2 T286A	15 WT (±4)	WT	WT		19 WT (±3)			
73B4 H349A	3 WT (±1)				WT			
78D2 H356A	51 WT (±4)	WT	WT		<1 WT (±1)			

Figure 4.29: Relative activities of mutations (WT activity set as 100%) of ckAAs towards phosphate (alanine scanning). The green rectangle means original activities were retained at level 4 (>80%); the light green rectangle indicated original activities were retained at level 3 (51%-80); the yellow rectangle showed that original activities were retained at level 2 (20%-50%); the pink rectangle indicates that original activities were retained at level 1 (<20%); the red rectangle means original activities were retained at level 0 (no activity). The olive green 'WT' indicates positive activity was detected in the WT (data from Chapter 3).

At the equivalent site of *VvGT1* T19, 73B4 R7A did not display much activity change (88% activity retained towards UDP-Glc, and 82% activity retained towards UDP-GlcNAc). This result suggested that 73B4 R7 might not be a ckAA in 73B4 and did not participate directly in binding the phosphate as inferred from the MSA result. On the contrary, 78D2 T22A only retained original activities of UDP-Glc and UDP-GlcNAc with 20% and 5%, respectively, and activities of UDP-Gal and GDP-Glc were completely abrogated. Therefore, it could be suggested that T22 is more likely to be a ckAA in 78D2 and probably binds to the phosphate moiety inferred by MSA.

At the equivalent site of *VvGT1* T280, S and T appear in site 270 and 286 in 73B4 and 78D2, respectively. The SDM result substantiates that S or T plays a significant role in enzyme activity, as most original activities were reduced: 73B4 S270A retained 26% original UDP-Glc activity and 0% original UDP-GlcNAc activity. 78D2 T286A retained 15% original UDP-Glc activity and 19% original UDP-GlcNAc activity.

Based on the MSA result, it may be inferred that the reason for the significant activity loss resulted from 73B4 S270 and 78D2 T286 binding phosphate, however, alanine mutants prevent the hydrogen bond between the enzyme and phosphate.

The AA 19 within the PSPG motif, H350 (*VvGT1* numbering), shows a high degree of conservation in MSA. The subsequent SDM test substantiated the significance of this site: 73B4 H349A only retained 3% and 0% of activities towards UDP-Glc and UDP-GlcNAc respectively; 78D2 H356A retained 51% and <1% of activities towards UDP-Glc and UDP-GlcNAc respectively. Accordingly, H can be inferred as a ckAA for 73B4. However, H might not play a crucial role in 78D2 towards UDP-Glc, as >50% of the original activity was retained in the alanine mutant.

Phosphate binding is important in all UGTs, as it is the common component of a nucleoside sugar donor. Apart from plant UGTs, UGTs with other origins also showed a high level of conservation towards phosphate binding at the equivalent site of *VvGT1* T280 and H350, such as GtfA (PDB: 1PN3) S230 and H293²³¹, GtfD (PDB: 1RRV) S246 and H309²⁴⁶, and UDP-GlcNAc 2-epimerase H242 (PDB:3BEO)²⁴⁷. Furthermore, SDM showed that UDP-GlcNAc 2-epimerase H242A only retained 3% of the original UDP-GlcNAc activity. As the N-terminal domain of UGTs is variable, the role of T19 (*VvGT1* numbering) located in the N-terminal was less conserved and less reported in the literature, however, examples can still be found. At the equivalent site *VvGT1* T19, S was conserved in human UGTs and showed a similar binding pattern towards phosphate as T19 in *VvGT1* (e.g. the crystal structure of UGT2B7 showed S34 interacted with the phosphate part of UDP-glucuronic acid²⁴⁸).

4.3.3 Summary – ckAAs in UGTs (phosphate moiety + sugar moiety)

Until now, potential ckAAs in UGTs towards the phosphate moiety of 73B4 and 78D2 were identified and summarised in Figure 4.30, along with our previous study on ckAAs towards the sugar moiety.

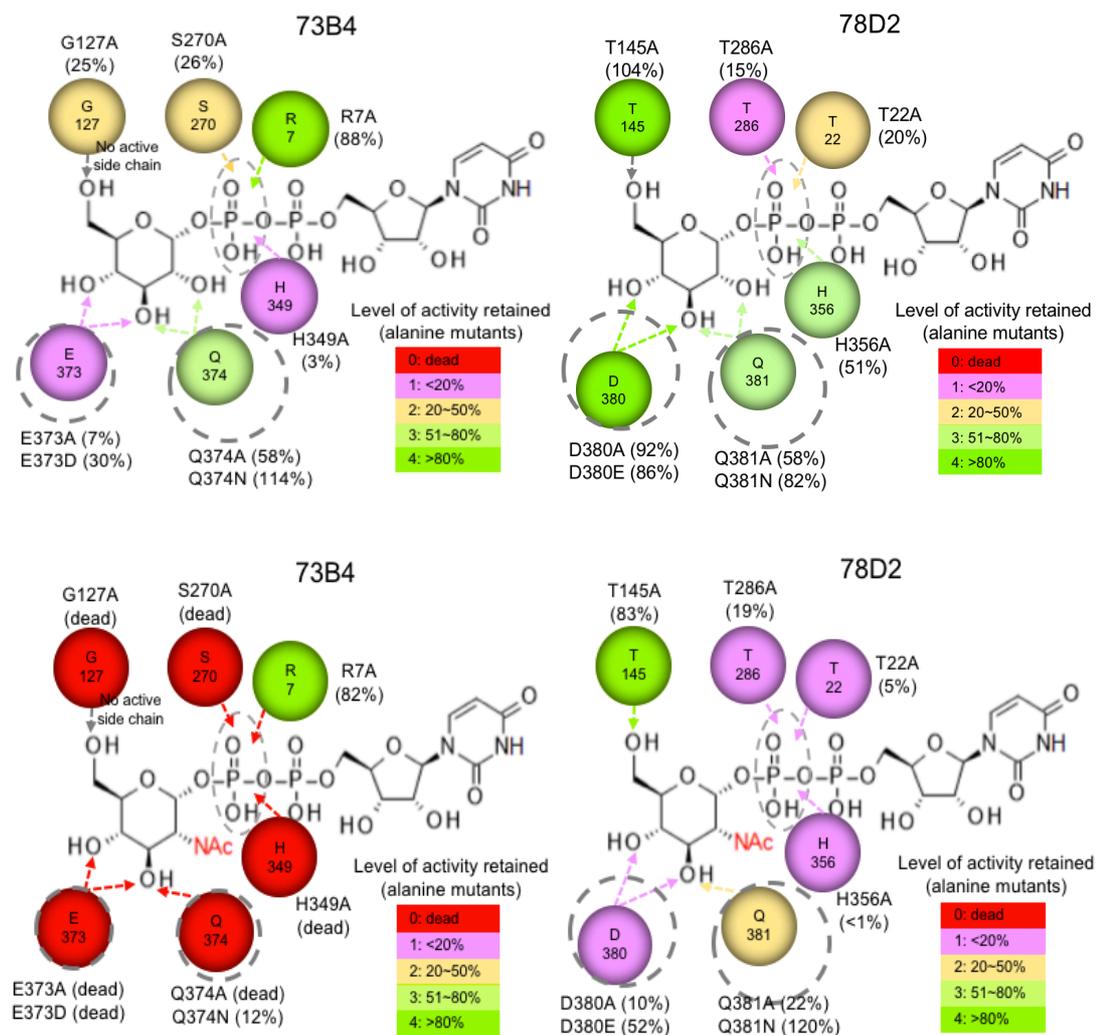


Figure 4.30: Potential ckaAs in UGTs towards phosphate, UDP-Glc (top, c.f. section 4.2), or UDP-GlcNAc (bottom, c.f. section 4.2). The level of significance of potential ckaAs (from MSA) is based on the experimental results of alanine scanning and is indicated by colour: Level 0, Very Strong (alanine mutants retained original activity 0%, in red); Level 1, Strong (alanine mutants retained original activity <20%, in pink); Level 2, Medium (alanine mutants retained original activity 20~50%, in yellow); Level 3, Weak (alanine mutants retained original activity 51~80%, in light green); Level 4, Very Weak/No (alanine mutants retained original activity >80%, in green). The diameter of the dotted circle suggests the distance (longer/shorter) between potential ckaAs and the substrate is based on the structure of D/E or Q/N swap.

Figure 4.30 demonstrates potential ckaAs towards the phosphate of 73B4 and 78D2 along with the ckaAs towards the sugar moiety as discussed in Section 4.2. They generally followed the similar interaction pattern as the template UGTs, but with some variations (e.g. 73B4 R7 was not a ckAA as the mutation of R did not show much effect on the enzyme activity, although MSA result indicated its potential role as a ckAA).

Phosphate is a common moiety both in UDP-Glc and UDP-GlcNAc, so it was expected that the mutation of potential ckAAs played an equal role in affecting enzyme activity/performance towards UDP-Glc and UDP-GlcNAc, however, the UDP-GlcNAc activity seemed to be more readily reduced by mutations in both 73B4 and 78D2. The possible explanation of this data is that UDP-GlcNAc activity is promiscuous activity (the kinetic data to support this is discussed in Chapter 3) but UDP-Glc activity is the main activity of 73B4 and 78D2. As discussed above, promiscuous activities of enzymes result from greater "plasticity" from the aspect of evolution²²⁵, therefore UDP-GlcNAc seemed to be affected by enzyme mutation more easily (this reason could be the explanation of why GDP-Glc and UDP-Gal activities of 78D2 were so easily affected).

4.4 Identification of ckAAs – uridine moiety

As the name indicates, the uridine moiety is the common part of a nucleoside sugar donor, and consequently one would suppose that the UDP binding pattern is highly conserved. This hypothesis has been supported by solved crystal structures where the AA 1 (generally W, except F in *OS79*) and AA 27 (E) within the PSPG motif were conserved and interacted with uridine by pi-pi interaction and hydrogen bond respectively (Table 4.7).

Table 4.7: ckAAs that bind uridine moiety in solved crystal structures

Plant UGT	ckAAs	Result origin
VvGT1 ³⁵	W332, E358	PDB (2C1Z)
UGT 72B1 ⁸¹	W346, E372	PDB (2VCE)
UGT 71G1 ⁷⁶	W339, E365	PDB (2ACW)
UGT 78G1 ⁷⁸	W334, E360	PDB (3HBF)
UGT 85H2 ⁷⁷	W360, E386	Docking
UGT 78K6 (<i>Ct3GT-A</i>) ^{79,80}	W325, E351	PDB (4WHM)
<i>OS79</i> ⁸²	F343, E369	5TMB

4.4.1 MSA

MSA was used to demonstrate the ckAAs in the target UGTs that might bind uridine (Figure 4.31).

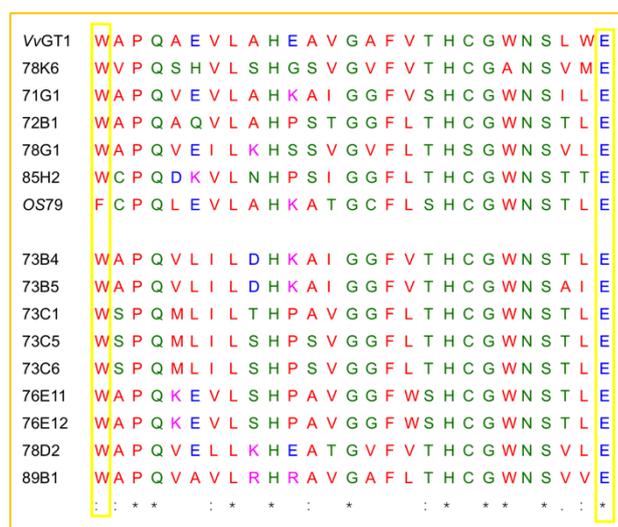


Figure 4.31: MSA of potential ckAAs towards the uridine moiety. The colours of residues represent their different properties: small and hydrophobic AAs AVFPMILW are coloured red; acidic AAs DE are coloured blue; basic AAs RK are coloured magenta; hydroxyl/sulphydryl/amine AAs STYHCNGQ are coloured green. This Figure demonstrates the two potential ckAAs towards uridine are located in the 1st and 27th position within the PSPG motif with a high degree of conservation.

MSA revealed that potential cKAAs towards uridine in targeted UGTs are highly conserved, with W and E as the 1st and the 27th AA within the PSPG motif.

4.4.2 SDM

SDM was performed on 73B4 W331 and E357, as well as 78D2 W338 and E364 (Figure 4.32).

	UDP-Glc	UDP-Gal	GDP-Glc	GDP-Fuc	UDP-GlcNAc	UDP-GalNAc	UDP-Rha	dUDP-Rha
73B4 W331A	WT				WT			
78D2 W338A	32 WT (±12)	WT	WT		<1 WT (±1)			
73B4 E357A	7 WT (±1)				WT			
78D2 E364A	9 WT (±1)	WT	WT		WT			

Figure 4.32: Relative activities of mutations (WT activity set as 100%) of cKAAs towards uridine (alanine scanning). The yellow rectangle shows original activities were retained at level 2 (20%-50%); the pink rectangle indicates original activities were retained at level 1 (<20%); the red rectangle means that original activities were retained at level 0 (no activity). The olive green 'WT' indicates positive activity was detected in the WT (data from Chapter 3).

As Figure 4.32 shows, alanine mutations of W and E significantly reduced original activity, with the only exception that 78D2 W338A retained above 20% of the original UDP-Glc activity (32%). The possible explanation why 78D2 W338A did not significantly reduce UDP-Glc activity is probably due to the region in 78D2 that stacks uridine being very stable. As shown in previous studies, the first six AAs within the PSPG motif were an important loop to stack the uracil ring in the hydrophobic platform. For example, in 71G1, the 339-344 loop region (WAPQVE) helped W339 form a hydrophobic platform to stack uridine. E356P abolished all UDP-Glc activity by disturbing the loop using P mutation²²⁹. Thus, it is possible that uridine binding is not only due to the pi-pi interaction formed by W, but may relate to the stability of the hydrophobic platform that the loop provided. Based on the fact that 78D2 W338A did not affect UDP-Glc activity significantly, it is suggested that the loop stacking of UDP

in 78D2 (WAPQVE) is extremely stable, therefore destruction of pi-pi interaction is not enough to result in a significant activity loss.

Reviewing previous studies, the critical role of W and E being responsible for interacting with uridine has been widely accepted. For example, SDM of UGT2B7 W356A and W356H resulted in attenuated activity towards UDP-glucuronic acid, although the modelling experiments showed that W356 did not interact with uracil²⁴⁸. The inconsistency of SDM and molecular modelling probably results from the conformational change of W in the apo form of UGT to the complex form of UGT+UDP-sugar (observed in VvGT1), which also indicated a disadvantage of molecular modelling and an advantage of SDM. Apart from W, the role of E was conserved throughout the UGT family, such as UGT2B7 E382²²⁶.

4.4.3 Summary – ckAAs in UGTs (uridine moiety + phosphate moiety + sugar moiety)

The potential ckAAs that might be involved in binding uridine in active sites of 73B4 and 78D2 were identified (Figure 4.33). Along with the potential ckAAs towards the phosphate and sugar moieties discussed in Sections 4.3 and 4.2, overall pictures of ckAAs at the active sites of 73B4 and 78D2 were speculated.

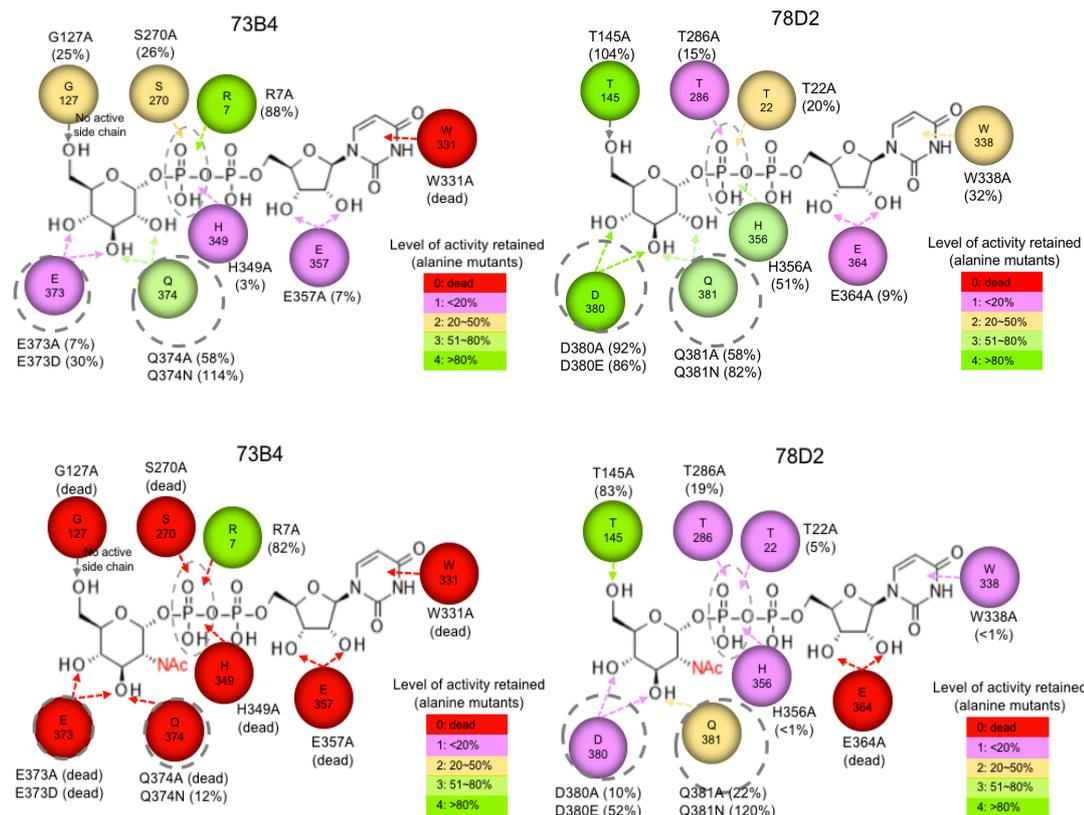


Figure 4.33: Potential cKAAs in UGTs towards uridine, phosphate, UDP-Glc (top, c.f. section 4.2), or UDP-GlcNAc (bottom, c.f. section 4.2). The level of significance of potential cKAAs (from MSA) is based on the experimental results of alanine scanning and is indicated by colour: Level 0, Very Strong (alanine mutants retained original activity 0%, in red); Level 1, Strong (alanine mutants retained original activity <20%, in pink); Level 2, Medium (alanine mutants retained original activity 20~50%, in yellow); Level 3, Weak (alanine mutants retained original activity 51~80%, in light green); Level 4, Very Weak/No (alanine mutants retained original activity >80%, in green). The diameter of the dotted circle suggests the distance (longer/shorter) between potential cKAAs and the substrate is based on the structure of D/E or Q/N swap.

Figure 4.33 shows potential cKAAs towards uridine of 73B4 and 78D2 along with the potential cKAAs towards the sugar and phosphate moieties as discussed above. Specifically, 73B4 W331 was inferred as a cKAA (by SDM data) probably interacting with the uracil ring via a pi-pi bond (by MSA data). 73B4 E357 and 78D2 E364 were identified as cKAAs (by SDM data) that might interact with ribose via a hydrogen bond (by MSA data). Again, mutation had a general stronger influence on UDP-GlcNAc activity than that of UDP-Glc activity, which probably could be explained from the aspect of evolution as discussed above.

Until now, the identification of potential cKAAs in nine target UGTs have been revealed by MSA and SDM. No crystal structures of target UGTs have been solved

successfully. Thus, MSA and SDM provide the best and fastest approach currently available to identify the potential ckAAs in UGTs.

4.5 Homology model validation

Although molecular modelling was not applied to guide the experiments presented here, it can give a useful way to visualise the interaction mentioned in the discussion. Consequently, homology modelling was inevitably used to construct models of UGTs without crystal structures.

Target UGTs in this section (73B4, 73B5, 73C1, 73C5, 73C6, 76E11, 76E12, 78D2, and 89B1) were constructed individually with their corresponding template UGTs (the one with the highest percentage of identical AAs) by the SWISS-MODEL program²⁴⁹. Notably, only the C-terminal of these UGTs was constructed as this research focused on donor binding and most ckAAs towards the donor were concentrated in the C-terminal of UGTs. The characteristic fold of UGT structure indicates the two terminal domains as independent halves²²⁶. C-terminal domain modelling can be accurate as no biasing would be introduced from the N-terminal domain that has a lower degree of homology.

According to the theory that Rost derived for a precise limit for homology modelling (Figure 4.34), two proteins that were practically guaranteed to adopt a similar structure the homology modelling could be applied, as long as the length of the two sequences and the percentage of the identical AAs fall in the ‘safe homology modelling zone’²⁵⁰. Taking the number of aligned AAs and the highest percentage of identical AAs with the template UGT together, all target UGTs were located in the ‘safe homology modelling zone’ and homology modelling was used to provide a reliable model (Figure 4.34)

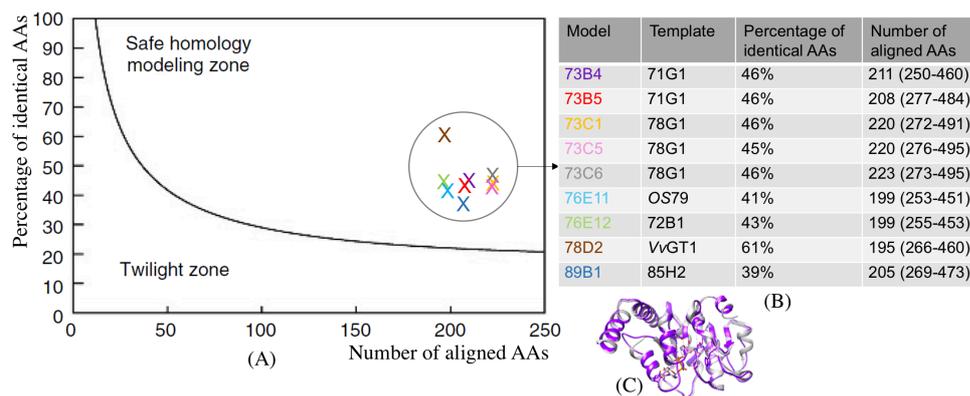


Figure 4.34: Homology modelling zone validation of target UGTs. (A): Limit for the homology modelling defined by Rost (1999). The location of 'X' means the point of the modelling zone where different target UGTs are located. They are represented by different colours corresponding to the name of the UGT shown in the right Table (B). All target UGTs are located in the safe homology modelling zone. (B): The percentage of identical residues between target and template UGT and the number of aligned residues of target UGT. (C): A representative homology model of the C-terminal of 73B4 (purple, model) overlaid on its template 71G1 (grey, crystal).

After cycles of energy minimisation, the final UGT models (C-terminal) were constructed and the model qualities were assessed (results are listed in Table 4.8). This comprehensive scoring indicated the level of confidence of the constructed models. The model quality can be estimated by QMEAN. QMEAN acts as a scoring function for global and local model quality estimation^{251,252}. It consists of four structural descriptors: local geometry, all-atom interaction, C-beta interactions, and burial status of residues. Furthermore, the model structures were examined using a Ramachandran map to reveal the rotations of the polypeptide backbone around the bonds between N-C α (Phi, ϕ) and C α -C (Psi, ψ)²⁵³. The map may find the torsion angles in disallowed regions and indicate any problems with the structure.

Table 4.8: Scoring of model quality validation

Model	Template	QMEAN	Residues in outlier region (Ramachandran plots)
73B4	71G1	-2.30	G241, K399
73B5	71G1	-3.33	K423
73C1	78G1	-3.10	V413, E415, V429
73C5	78G1	-3.79	E418, E427, V431
73C6	78G1	-2.71	S303, S339, W423, V431, L432, V433, D434, K435
76E11	OS79	-1.72	-
76E12	72B1	-1.31	-
78D2	VvGT1	-0.57	E315
89B1	85H2	-2.80	I296

Table 4.8 shows that all constructed models presented a QMEAN in the range of -3.33~-0.57. In the Swiss-Model program, a value between -4.00~+4.00 was

recommended, which indicated the reliability of the constructed models. In addition, more than 96% of all AAs were located in the allowed regions of the Ramachandran plots. Every potential ckAAs (C-terminal) in this project was in the allowed area of the Ramachandran plots. Therefore, the location of the ckAAs in the models displayed above are reliable. In summary, the scoring of model quality elucidates the reliability of the constructed model and the potential ckAAs in this Chapter. Consequently, the overlaying of the target UGTs on the template UGTs would be reasonable and could provide a useful interpretation of SDM data.

4.6 Conclusions

By combining MSA and SDM, potential ckAAs in the target UGTs were inferred and substantiated/disapproved. Potential ckAAs towards sugar, phosphate, and uridine moieties of an individual enzyme were summarised in Sections 4.2.5, 4.3.3 and 4.4.3, respectively. A general summary of the ckAAs towards different portions of the donor molecule is shown in Figure 4.35. A general summary of ckAAs towards different sugars is shown in Figure 4.36.

Chapter 4

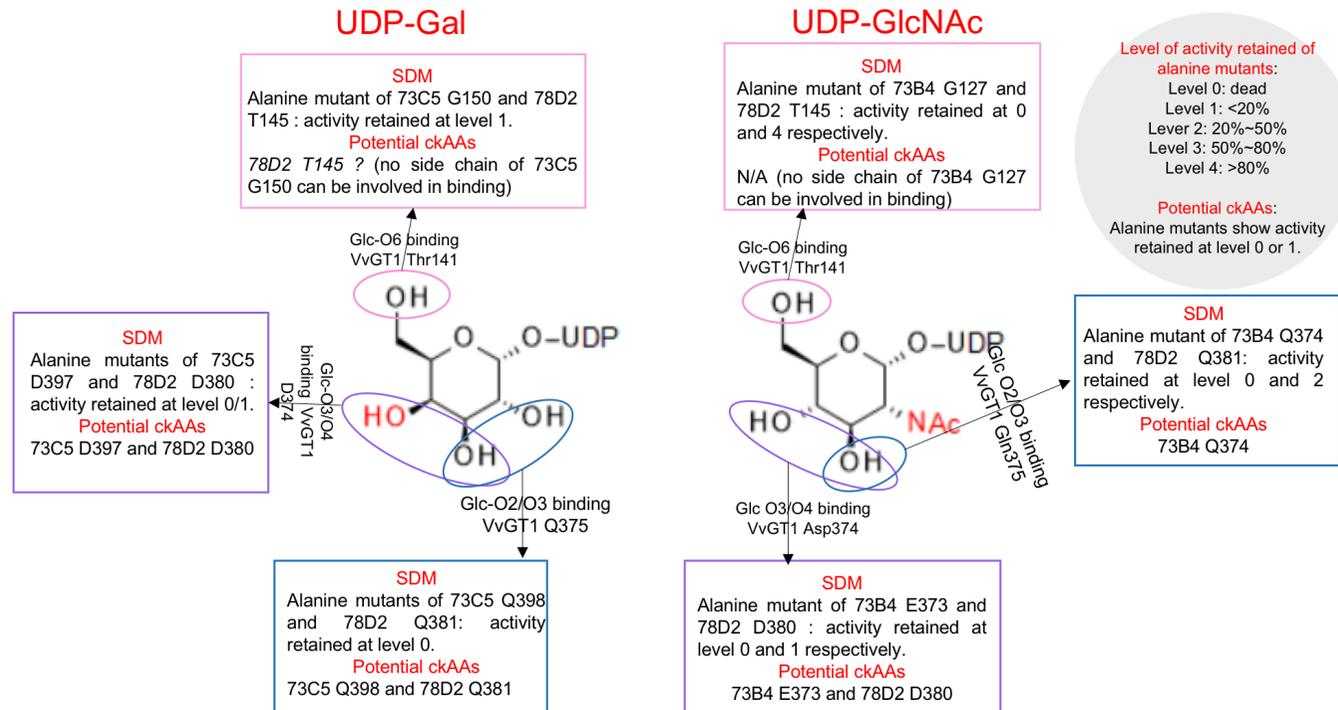


Figure 4.36: Summary of SDM and potential ckAAs towards Gal and GlcNAc moiety.

For potential ckAAs towards the C2 and C3 positions of the sugar, MSA showed that Q was highly conserved as the AA 44 within the PSPG motif. By SDM, 73C1 Q394, 73C5 Q398, 73C6 Q398, and 76E11 Q374 were inferred as potential ckAAs (alanine mutants retained original activity at 0/1), probably by orientation on the Glc-O2 and Glc-O3 positions (based on MSA result).

For potential ckAAs towards the C3 and C4 positions of the sugar, MSA showed that a negatively charged AA (D/E) was conserved as the AA 43 within the PSPG motif. Along with the result of alanine scanning, 73B4 E373, 73B5 E394, 73C1 D393, 73C5 D397 and 73C6 D397, 76E11 D373, 76E12 D375 and 89B1 D389 were inferred as potential ckAAs (alanine mutants retained the original activity at 0/1) probably by binding the Glc-O3 and Glc-O4 positions (based on MSA result).

Potential ckAAs towards the C6 position of the sugar moiety showed the least level of conservation: 76E11, 76E12, and 78D2 showed T; 73B4, 73B5, 73C1, 73C5, and 73C6 showed G; and 89B1 showed P. 76E11 T134 and 76E12 T135 were proposed as ckAAs (based on SDM result).

The phosphate binding site study was focused on the equivalent site of VvGT1 T19, T280, and H350. T19 (VvGT1 numbering) was the least conserved: 78D2 with T, 73B4 with R, and the remaining target UGTs with G. T280 (VvGT1 numbering) was semi-conserved as all target UGTs showed a polar uncharged AA (S/T). H350 (VvGT1 numbering) was the highest conserved as all target UGTs showed the same AA (H). SDM indicated that 78D2 T22, 78D2 T286 and 73B4 H349 were crucial for UDP-Glc activity and were proposed to be ckAAs.

Identification of ckAAs towards the uridine moiety included two AAs: the AAs 1 and 27 within the PSPG motif. They were all highly conserved as W (AA 1) and E (AA 27). SDM suggested that 73B4 W331, 73B4 E357, and 78D2 E364 were key for enzyme activity and could be inferred as potential ckAAs.

Apart from UDP-Glc, ckAAs towards other sugar donors were investigated. CkAAs towards other sugar donors generally followed the same fashion to those of UDP-Glc. However, there were some variations in details. For example, Q was inferred to be

involved in binding Glc-O2 and Glc-O3 moieties, which might not be applied directly to UDP-GlcNAc due to the *N*-acetyl group at the C2 position of GlcNAc might not be involved in the interaction via forming a bond as previous GlcNAcT studies indicated. Thus, Q was inferred to bind only the C3 position of UDP-GlcNAc rather than both the C2 and C3 positions. In addition, it was generally observed that the activities of other donors were more sensitive towards mutation compared to the activity of UDP-Glc. This probably could be explained from the aspect of evolution that promiscuous activity of an enzyme was not as 'robust' as the main activity, but appeared to be 'plastic' towards mutation.

Chapter 5 Rhamnosyltransferases 78D1 and 89C1

Many rhamnose (Rha)-containing natural compounds are present in both plants and bacteria. For example, L-Rha is a key component of rhamnogalacturanan I and II that are the primary constituents of pectins (mainly responsible for cell wall extension and plant growth) in the plant cell wall²⁵⁴. Further, L-Rha can be a part of *O*-linked glycoproteins (e.g. *O*-glycan chains of arabinogalactan proteins) that control many biological activities, such as growth, morphogenesis, and adaptation to stress²⁵⁵. In bacteria, several types of *O*-antigens containing Rha can be covalently linked to lipopolysaccharides in the outer leaflet of the outer cell membrane (e.g. the *Pseudomonas aeruginosa* core oligosaccharide of the lipopolysaccharide), which can facilitate bacterial evasion of the immune system defences²⁵⁶.

Rhamnosyltransferases (RhaTs) are enzymes involved in the synthesis of Rha-containing compounds. They catalyse the transfer of Rha from NDP-Rha to the acceptors. Thus far, sequences of over 3000 RhaTs have been deposited in the UniProt database; they originate from various organisms, such as *Arabidopsis thaliana* (*At*), *Escherichia coli* (strain K12), *Citrus maxima*, and *Mycobacterium tuberculosis*²⁵⁷.

Functions of some RhaTs have been characterised. For example, two RhaTs in citrus species (pomelo and grapefruit) determine the flavour of a plant: 1,2-RhaT, for the synthesis of the bitter flavanone-7-*O*-neohesperidoside, and 1,6-RhaT, for the synthesis of the tasteless flavanone-7-*O*-rutinoside²⁵⁸. *Glycine max* (soybean) UGT79A6 is a flavonol-3-*O*-glucoside (1→6) RhaT, which uses β -L-UDP-Rha and a flavonol 3-*O*- β -D-Glc (kaempferol or quercetin) to synthesise flavonol 3-*O*-(α -L-Rha-(1→6)- β -D-Glc)²⁵⁹. GmSGT3 (soybean) can convert soyasaponin III to soyasaponin I by using UDP- β -L-Rha as the sugar donor²⁶⁰.

Although the functions of some RhaTs have been identified, the number of available RhaTs crystal structures is limited. Currently, only two crystal structures of RhaTs from bacteria, β -1,2-RhaT WsaF (PDB: 2X0D, APO structure of WsaF; 2X0E, complex structure of WsaF with dTDP; 2X0F, complex structure of WsaF with dTDP- β -L-Rha)^{261,262} and *Saccharopolyspora spinosa* *SpnG* (PDB: 3TSA, APO structure of *SpnG*; 3UYK, complex structure of *SpnG* with spinosyn aglycone; 3UYL, complex structure of *SpnG* with dTDP)²⁶³ have been solved, and provide molecular-level insights into the understanding of RhaTs.

WsaF belongs to the GT4 family in the CAZy database. It is a β -retaining RhaT, responsible for the formation of β -1,2-linkage during the polyrhamnan biosynthesis for the S-layer protein SgsE in *Geobacillus stearothermophilus* NRS 2004/3a²⁶⁴. The growing Rha chain is used as a substrate by accepting the sugar from dTDP- β -L-Rha. WsaF adopts a GT-B fold with two Rossmann-fold domains ($\beta/\alpha/\beta$). A cavity in the cleft between the N-terminal and C-terminal domains is responsible for substrate accommodation. In the complex structure of WsaF with dTDP- β -L-Rha, the thymidine moiety forms polar interactions with the side-chain of K302 and the main-chain atoms of L303; O3 of the 2-deoxyribose interacts with E333 by hydrogen bonds; Rha sugar forms hydrophobic interactions with Y329, hydrogen bond with N227 (-OH groups at C2 and C3 positions), and hydrogen bond with K225 (-OH groups at C3 and C4 positions) (Figure 5.1, left).

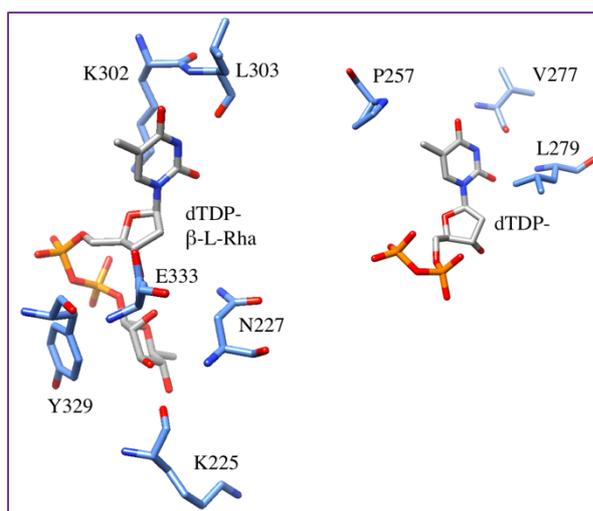


Figure 5.1: Some key AAs in the active sites in WsaF (left) and *SpnG* (right).

Another RhaT whose crystal structure had been solved is *SpnG*. It belongs to the GT1 family and catalyses 9-OH rhamnosylation to spinosyn by using dTDP-Rha as the sugar donor. Similar to WsaF, the architecture of *SpnG* also adopts a GT-B fold and contains two domains with two Rossmann-fold domains. The dTDP-binding pocket is composed of AAs that are mostly located in the C-terminal domain, i.e. A255–P257, S276–L279, and S295–T297. For example, thymine interacts with L279 (O2 of thymine) and V277 (N3 and O4 of thymine) via hydrogen bonding, and P257 (methyl of thymine) and L279 (the ring of thymine) by hydrophobic interactions (Figure 5.1, right). Because of a lack of the complex structure of *SpnG* with dTDP- β -L-Rha, dTDP-

Rha was docked along with the crystal research. The docking of dTDP-Rha into the crystal structure of *SpnG* revealed that D316–Q317 interact with C2–C4 positions of Rha. The D316–Q317 motif is also conserved in plant UGTs, as the last two AAs within the PSPG motif, and interacts with the C2–C4 positions of the sugar moiety as discussed in Chapter 4.

This chapter focuses on RhaTs 78D1 and 89C1, detailing kinetic study and the identification of catalytic key AAs (ckAAs). The methodology was the same as described in the previous chapters: the kinetic study was performed using multiple reaction monitoring (MRM) mode of mass spectrometry (MS) by monitoring the formation of kaempferol-rhamnose (KMP-Rha) as a function of time (c.f. Chapter 3). The ckAAs of enzymes were identified by multiple-sequence alignment (MSA)-based methodology (c.f. Chapter 4). This study expands the understanding of 78D1 and 89C1, shedding light on related research, such as the biological role of RhaTs, synthesis of Rha-containing glycoconjugates, and active site manipulation.

5.1 *In vitro* characterisation of 78D1 and 89C1

5.1.1 Donor specificity of 78D1 and 89C1

The donor specificities of UGT78D1 and 89C1 were examined using the donor library mentioned in Chapter 3 (with KMP as the acceptor), namely, α -D-UDP-Glc, α -D-UDP-Gal, α -D-UDP-GlcNAc, α -D-UDP-GalNAc, β -L-UDP-Rha, β -L-dTDP-Rha, α -D-GDP-GDP-Glc, and β -L-GDP-Fuc. Experimental data revealed that 78D1 and 89C1 are able to transfer the Rha moiety from both UDP-Rha and dTDP-Rha to KMP (see GAR results in Chapter 3). The glycosylated product KMP-Rha was expected to show a peak at 431 in the ‘full scan’ mode of MS in the negative-ion mode ($[MW-H^+]$). The subsequent ‘product ion’ mode of MS fragmented the KMP-Rha to KMP with a peak at 285 ($[MW-H^+]$), with the collision energy of 10 eV (See MS figure in Appendix, Figure A.7 and A.8). The experimental conditions were clarified and discussed in Chapter 3, including the buffer, temperature, MS, and LC conditions.

In our research, both 78D1 and 89C1 exhibited dual activities of UDP-Rha and dTDP-Rha. In previous studies, only UDP-Rha activity was reported^{151,155}. The dTDP-Rha activity had been predicted previously but never experimentally confirmed. In a previous *in silico* screening, 78D1 showed a dTDP-6-deoxytalose activity²⁶⁵, which indicated the tolerance of the dTDP nucleoside base in the active site of 78D1 and was consistent with the results of the current study.

With respect to the chemical structure, dTDP has an additional CH₃ group at C5 of uracil and lacks an –OH group at 2' of ribose (Figure 5.2). The structural difference of UDP and dTDP is tolerated by some enzymes, such as VvGT1³⁵, while some enzymes show strict selectivity towards specific donors, such as trehalose 6-phosphate synthase (Sco), which transfers Glc only from UDP-Glc but not dTDP-Glc²⁶⁶. The selectivity for UDP-sugar and dTDP-sugar donors might be associated with the active site of enzymes, which will be discussed below.

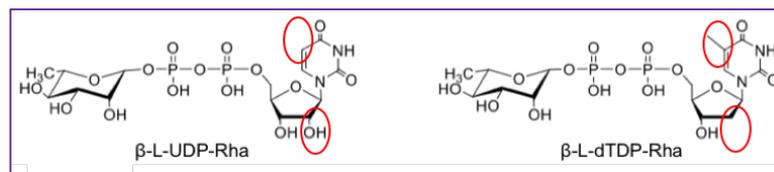


Figure 5.2: Structures of β -L-UDP-Rha (left) and β -L-dTDP-Rha (right).

Apart from the difference in structure, UDP-Rha and dTDP-Rha are different with respect to their distribution in nature. UDP-Rha is found in plants²⁶⁷, while dTDP-Rha is found in bacteria²⁶⁸. Although 78D1 and 89C1 show dual activities, they are only found in plants: 78D1 is produced in leaves, flowers, siliques, and stems¹⁵¹; 89C1 is highly produced in floral buds, but also produced in roots, stems, and leaves¹⁵⁵. Thus, dTDP-Rha cannot be the natural substrate of 78D1 and 89C1 *in vivo*. However, the newly reported dTDP-Rha activity may broaden the available materials in potential applications, such as synthesis of Rha-containing glycoconjugates.

5.1.2 Kinetic study

Rhamnosylation reactions catalysed by 78D1 and 89C1 are Bi-Bi enzymatic reactions, with two reactants and two products. A pseudo-Uni system was generated with a high fixed concentration of KMP (100 μ M) with varied UDP-Rha/dTDP-Rha concentrations. The K_M of KMP was fixed as this part of investigation mainly focused on the kinetic parameters of the donor molecule. According to previous studies (c.f. Chapter 3), most K_M values of KMP are below 10 μ M (except that for 73C1, which was 47 μ M), thus 100 μ M was estimated as the saturation level of KMP. The Michaelis-Menten curves (Figures 5.3 and 5.4) were used for the calculation of kinetic constants, based on the equation $v = \frac{V_{max} * S}{K_M + S}$. Other experimental conditions have been already discussed in Chapter 3, including, the buffer, temperature, MS, and LC conditions, and selection of enzyme concentration.

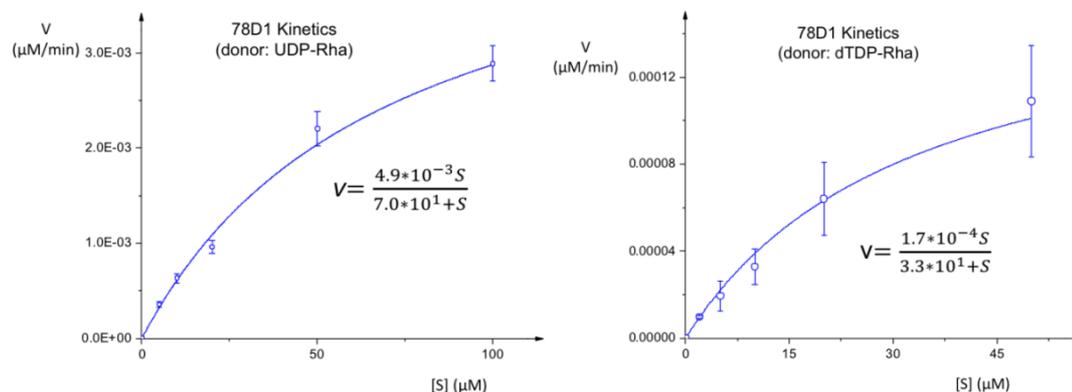


Figure 5.3: Michaelis-Menten curve of 78D1 kinetics towards UDP-Rha and dTDP-Rha.

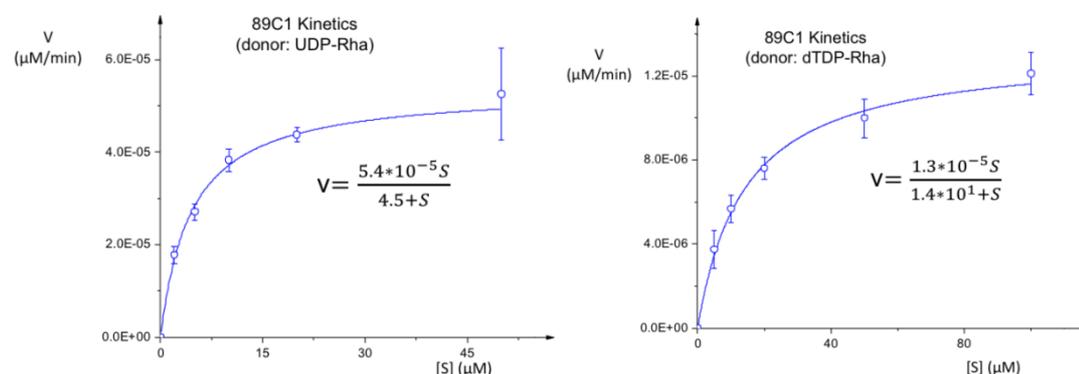


Figure 5.4: Michaelis-Menten curve of 89C1 kinetics towards UDP-Rha and dTDP-Rha.

Table 5.1: Kinetic parameters of 78D1 and 89C1

	Donor	K_M (M)	V_{max} (M/s)	k_{cat} (s^{-1})	k_{cat}/K_M
78D1	UDP-Rha	$7.0E-05$ ($\pm 1.6E-05$)	$8.1E-11$ ($\pm 1.3E-11$)	$2.0E-03$ ($\pm 3.3E-04$)	$2.9E+01$ ($\pm 8.1E+00$)
	dTDP-Rha	$3.3E-05$ ($\pm 7.2E-06$)	$2.8E-12$ ($\pm 5.5E-13$)	$7.0E-05$ ($1.4E-05$)	$2.1E+00$ ($\pm 6.2E-01$)
89C1	UDP-Rha	$4.5E-06$ ($\pm 5.1E-07$)	$9.0E-13$ ($3.3E-14$)	$2.2E-05$ ($8.3E-07$)	$5.0E+00$ ($\pm 6.1E-01$)
	dTDP-Rha	$1.4E-05$ ($\pm 1.7E-06$)	$2.2E-13$ ($\pm 9.5E-15$)	$5.5E-06$ ($\pm 2.4E-07$)	$3.9E-01$ ($4.9E-02$)

For 78D1, the K_M value indicated that a higher concentration of UDP-Rha is required to fill half of the active sites and for significant catalysis to occur than that of dTDP-Rha (2.1-fold). The values of k_{cat} and k_{cat}/K_M indicated that 78D1 transfers Rha from UDP-Rha with a higher catalysis capability (29-fold) and higher catalysis efficiency (14-fold) compared to dTDP-Rha. Thus, UDP-Rha is a better substrate candidate for 78D1 than dTDP-Rha.

For 89C1, the K_M value for UDP-Rha was 3.1-fold smaller than that for dTDP-Rha. This suggested that a lower concentration of UDP-Rha is required to fill half of the active sites of 89C1 and for significant catalysis to occur. In addition, the values of k_{cat} and k_{cat}/K_M indicated that UDP-Rha is a better substrate candidate of 89C1, with a 4-fold higher catalysis capability and a 13-fold higher catalysis efficiency compared to that of dTDP-Rha.

The comparison of 78D1 and 89C1 revealed that a lower concentration of UDP-Rha is required for a significant reaction to occur with 89C1 (16-fold lower than that of 78D1). However, the catalysis capability (k_{cat}) and efficiency (k_{cat}/K_M) of 89C1 were weaker than those of 78D1, by 91-fold and 6-fold, respectively. Reactions with the donor dTDP-Rha followed the same trend. Smaller concentrations of dTDP-Rha were required for significant reactions to occur with 89C1 (2.4-fold lower than that of 78D1). However, 78D1 had a 13-fold higher catalysis capability (k_{cat}) and a 5.4-fold higher efficiency (k_{cat}/K_M) than that of 89C1. These comparative results indicated that 78D1 is a more efficient enzyme facilitating the Rha transfer reaction than 89C1 *in vitro*.

Thus far, limited kinetic data have been obtained for RhaTs, and no kinetic data have been reported for 78D1 and 89C1. For some other RhaTs, the K_M values of donor are in the 1–100 μM range, as also observed for 78D1 and 89C1. For example, WbbL (catalysing Rha transfer from dTDP-Rha to decaprenyl-pyrophosphoryl-GlcNAc) exhibits K_M of 35 μM for dTDP-Rha²⁶⁹. *C. maxima* C12RT1 (catalysing Rha transfer from UDP-Rha to prunin and hesperetin-7-*O*-glucoside) showed K_M values of 1.3 and 1.1 μM for UDP-Rha with prunin and hesperidin-7-*O*-glucoside as acceptors, respectively²⁷⁰.

The detailed *in vitro* kinetic study of 78D1 and 89C1 presented in this Chapter may be useful for several fields, e.g. 1) the kinetic data of UDP-Rha can be indicative of the biological function of 78D1 and 89C1 *in vivo*, as K_M value may provide an approximation of substrate concentration *in vivo*⁶⁵. 78D1 and 89C1 are thought to play significant roles in hormone regulation during plant development, by flavonoid glycosylation. For example, 78D1 was relatively abundant at the shoot apex in *Arabidopsis*. It affects the growth of the plant by taking part in the synthesis of KMP-3-*O*-Rha-7-*O*-Rha that modulates auxin transport and affects the shoot phenotype²⁷¹.

Similarly, 89C1 indirectly induces auxin concentration change. It facilitates flavonol glycosylation, which would increase the concentration of auxin precursors and auxin metabolites²⁷². 2) as mentioned above, dTDP-Rha may broaden the array of materials available for such potential applications as the synthesis of Rha-containing glycoconjugates, although dTDP-Rha activity does not inform any biological functions of these enzymes as the compound is not present in plants. The significant role of flavonol rhamnosides has been emphasised in many studies. For example, quercetin-3-*O*-Rha might be an antimicrobial agent blocking the replication of the influenza virus^{273,274}. However, obtaining flavonol-Rha compounds is not easy: they were originally isolated from plants, which was followed by several complex purification steps and was limited by the supply of plant material³⁷. Thus, the *in vitro* activity and kinetic study of 78D1 and 89C1 may suggest an alternative method (e.g. a chemo-enzymatic method) for generation of such compounds.

5.1.3 Experiment optimisation

As discussed in Chapter 3, an accurate range of substrate concentration ([S]) can result in an improved experimental performance and smaller error; however, the [S] range used in this part of the investigation was not optimal. Thus, the kinetic study should probably be double-checked using the new $1/5-5K_M$ [S] range to obtain data of the best quality (Table 5.2). This will be performed as part of the future studies.

Table 5.2: $1/5-5K_M$ [S] range based on the full kinetic experimental K_M value

UGT	Donor	Range [S] (μM)
78D1	UDP-Rha	14-350
	dTDP-Rha	7-165
89C1	UDP-Rha	1-23
	dTDP-Rha	3-70

Table 5.2 provides the theoretical optimal [S] range. It guides the [S] range for future experiments; however, these concentrations cannot be implemented directly without practical consideration. For example, if the [UDP-Rha] or [dTDP-Rha] > 100 μM , MS ion suppression may occur, leading to the loss/decrease of a linear response of ESI-MS (discussed in Chapter 3). Thus, [UDP-Rha] or [dTDP-Rha] above 100 μM is not recommended for the MS approach.

5.2 Identification of catalytic key amino acids (ckAAs) in 78D1 and 89C1

Using the same MSA-based methodology as the one in Chapter 4, ckAAs in the active sites of 78D1 and 89C1 were identified. The rationality, accuracy, reliability, assumptions, and limitations of the methodology were discussed in Chapter 4.1.

First, selection of suitable enzymes as templates was required to predict the potential ckAAs in the active sites of 78D1 and 89C1. The plant template UGTs (*VvGT1*³⁵, *72B1*⁸¹, *71G1*⁷⁶, *78G1*⁷⁸, *85H2*⁷⁷, *78K6*^{79,80}, and *OS79*⁸²) used in Chapter 4 could have provided useful information for the prediction of the potential ckAAs in the active sites of 78D1 and 89C1 on account of their relatively high sequence identities, plant origins, and similar function of the UDP-linked sugars. However, the functions of 78D1 and 89C1 are not exactly the same as those of template UGTs, as the template UGTs transfer Glc, while the target UGTs transfer Rha. Thus, the binding between UGTs and sugar moiety may differ between template and target UGTs. Herein, *SpnG*, a RhaT, was used as another template. *SpnG* is a bacterial 9-OH RhaT that transfers Rha from dTDP-Rha to spinosyn. Although it shares low sequence identity with 78D1 (13%) and 89C1 (10%), the use of *SpnG* as a template was justified as it belongs to family 1 (as do 78D1 and 89C1) and may share a common ancestor with the target UGTs from the evolutionary perspective, i.e. the active site (sugar portion) of *SpnG* may have some commonalities with the those of the target UGTs.

As mentioned in the preface, another RhaT exists, *WsaF*, whose crystal has been solved. This enzyme was not used as a template mainly because it belongs to GT4 family, which is remote from the family to which the target UGTs belong (GT1). Consequently, low ancestral commonality was expected between *WsaF* and the target UGTs.

5.2.1 Identification of ckAAs – sugar moiety

The identification of 78D1 and 89C1 ckAAs towards sugar moiety focused on the ckAAs interacting with the –OH groups of Rha that are located at C2–C4 positions.

In addition to the absence of –OH group at C6 position of the sugar moiety, Glc and Rha have different forms in solution: Glc is present in a D-sugar form, while Rha is present in the L-sugar form. In Figure 5.5, UDP- β -L-Rha is shown in the 1C_4 chair conformation linking an equatorial UDP in solution. However, a possible alternative ring conformation might be existed (axial leaving-group orientation 4C_1)³⁵ as a nucleophilic attack during catalysis²⁷⁵.

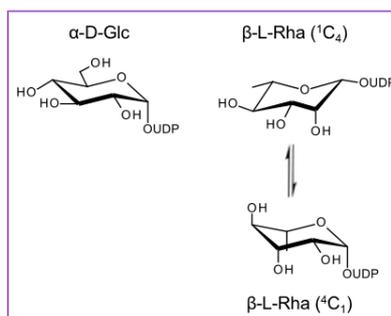


Figure 5.5: Schematic representations of UDP- α -D-Glc in 4C_1 conformation and UDP- β -L-Rha in both favoured 1C_4 conformation and 4C_1 conformation. Steric factors dictate that UDP- β -L Rha will exist in a conformation skewed between those drawn; one which allows an axial leaving-group orientation while fulfilling the stereo electronic considerations of an incipient oxocarbenium ion-like transition state³⁵.

5.2.1.1 ckAAs towards C2-C4 positions of the sugar moiety

In the *SpnG*-dTDP complex²⁶³, dTDP binds the C-terminal domain of *SpnG* with its diphosphate pointing towards the active area of the sugar moiety, which is consistent with template GlcTs, such as *VvGT1* (Figure 5.6). The docking analysis in the literature revealed that *SpnG* D316–Q317 facilitate Rha binding, as was observed in GlcTs: i.e. D316 binds Rha-O3 and O4, while Q317 binds Rha-O2 and O3.

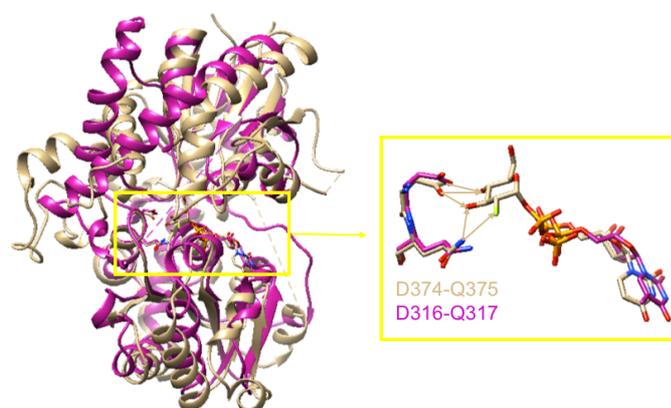


Figure 5.6: Overlay of overall crystal structure of *VvGT1* (tan colour), chain A of *SpnG* (violet red colour), and the DQ motif of *VvGT1* (D374-Q375) and *SpnG* (D316-Q317) in active site.

MSA was performed on template enzymes (seven GlcTs and *SpnG*) and target enzymes (78D1 and 89C1) (Figure 5.7). At sites equivalent to the D/E-Q motif in different templates, D374–N375 and D356–H357 dyads were observed in 78D1 and 89C1 respectively.

VwGT1	C R P F F G D Q		
71G1	T V P I Y A E Q		
72B1	A W P L Y A E Q		
78G1	S R F F F G D Q		
78K6	C R F F F G D Q		
85H2	C W P F F A D Q		
OS79	A M P H W A D Q	<i>SpnG</i>	V L P Q Y F D Q
78D1	G R P I L A D N	78D1	G R P I L A D N
89C1	A W P M Q A D H	89C1	A W P M Q A D H
	: .		: .

Figure 5.7: MSA of potential ckAAs towards C2-C4 positions of sugar moiety using templates of seven GlcTs (left) and *SpnG* (right). The colours of residues represent their different properties: small and hydrophobic amino acids AVFPMILW are coloured red, acidic amino acids DE are coloured blue, basic amino acids RK are coloured magenta, and hydroxyl/sulphydryl/amine amino acids STYHCNGQ are coloured green. The last line of symbols denotes the degree of conservation: conserved (*), conserved mutations (:), semi-conserved mutations (.), and non-conserved mutations (). This Figure demonstrates that the potential ckAAs towards C3 and C4 positions of the sugar moiety are highly conserved as D/E. The potential ckAAs towards C2 and C3 positions of sugar moiety are conserved in GlcTs as Q, but varied in RhaTs as Q (in *SpnG*), N (in 78D1) and H (in 89C1).

UGT 78D1: ckAAs towards C2-C4-positions of the sugar moiety

SDM was used to construct mutants 78D1 D374A, N375A, and N375Q to explore the role of the D-N dyad (Figure 5.8).

	78D1							
	UDP-Glc	UDP-Gal	GDP-Glc	GDP-Fuc	UDP-GlcNAc	UDP-GalNAc	UDP-Rha	dTDP-Rha
D374A	Red	Red	Red	Red	Red	Red	<1 WT (±1)	<1 WT (±1)
N375A	Red	Red	Red	Red	Red	Red	WT	WT
N375Q	Green	Red	Red	Red	Green	Red	4.3 WT (±1)	1.2 WT (±1)

Figure 5.8: Relative activities of mutations (WT activity set as 100%) of ckAAs towards C2-C4 positions of sugar moiety (alanine scanning). The pink rectangle means original activities were retained at level 1 (<20%); the red rectangle means original activities were retained at level 0 (no activity). The olive green 'WT' indicates positive activity was detected in the WT (data from Chapter 3).

Taken together, the MSA and SDM results indicated that 78D1 D374 is a ckAA that plays a significant role in Rha transfer (the original UDP-Rha and dTDP-Rha activities were significantly reduced to <1% by a D374 to A substitution), possibly interacting with Rha-O3 and Rha-O4.

Further, N375 in 78D1 may act as a ckAA, as suggested by the observation that A substitution of N375 completely abolished the original UDP-Rha and dTDP-Rha activities. Based on the MSA results, the loss of UDP-Rha and dTDP-Rha activities can be explained by the possible role of N375 in 78D1 in Rha-O2 and Rha-O3 binding, with the alanine substitution eliminating the interaction between N375 and Rha.

Notably, new UDP-Glc and UDP-GlcNAc activities were acquired by 78D1 (see MS in Appendix, Figure A.9) as a result of N375Q substitution, and the original Rha transfer activities were greatly simultaneously reduced (4.3% original UDP-Rha and 1.2% original dTDP-Rha activity retained). This result implies the possible 'recognition role' of the N residue, i.e. specifically recognising Rha.

Thus far, no clear conclusion has been made on how UGT recognises and even discriminates between the different sugar donors. Although some studies present successful cases of donor specificity switching by point substitutions (e.g. H374Q of *Aralia cordata* UDP-Gal anthocyanin GalT successfully acquiring a UDP-Glc activity¹⁰³), few studies successfully changed donor specificities of RhaTs and non-RhaTs.

A comparison of 78D1 and 78D2 revealed that they belong to the same UGT group (Group F) and share a high AA identity (71%), suggesting that one of them may have arisen from the other via gene duplication during evolution. However, they function quite differently: 78D1 transfers Rha, while 78D2 transfers Glc and GlcNAc. MSA revealed different AA at 44th position within the PSPG motif in these enzymes: 78D1 harbours N375 while 78D2 harbours Q381.

Interestingly, according to previous studies, AA 44 within the PSPG motif may be associated with donor specificity recognition and discrimination. For example, many studies claim that Q is responsible for recognising Glc, while H is responsible for recognising Gal (discussed in Chapter 4). Thus, a question arises whether the fact that 78D1 does not recognise and transfer Glc is because of the absence of Q as AA 44 within the PSPG motif. If so, 78D1 N375Q would acquire UDP-Glc activity. This was subsequently proved by the observation that 78D1 N375Q exhibited UDP-Glc and UDP-GlcNAc activities, and the original UDP-Rha and dTDP-Rha activities were significantly attenuated. However, the discrimination between transferring Rha and Glc/GlcNAc cannot be fully accomplished by point substitution, as 78D2 Q381N did not acquire UDP-Rha/dTDP-Rha activity (c.f. Chapter 4). This paradox may be linked to the difficulty of substituting the entire set of sugar recognition AAs during evolution, with many key sugar recognition AAs located at other sites²³⁰.

Apart from the discrimination of Rha and Glc by 78D1 and 78D2, previous studies also provided evidence that AA 44 within the PSPG motif might be associated with donor specificity. For example, 78D3 shares 75% AA identity with 78D2 but uses UDP-Arabinose (Ara). Ara is present in solution in the L- β form and compared with Glc lacks –OH at C6 position. In 78D3, AA 44 within the PSPG motif is H (site 386). A previous study demonstrated that 78D3 acquired new UDP-Glc activity when H was replaced with Q²⁷⁶, which was consistent with the result presented in the current chapter that UDP-Glc activity could be conferred on 78D1 by replacing of AA 44 within the PSPG motif with Q.

UGT 89C1: ckAAs towards C2-C4-positions of sugar moiety

Following the same methodology, D356 and H357 mutants of 89C1 were subjected to SDM, and their activities were examined (Figure 5.9).

	89C1							
	UDP-Glc	UDP-Gal	GDP-Glc	GDP-Fuc	UDP-GlcNAc	UDP-GalNAc	dTDP-Rha	
D356A							2.0 WT (±<1)	12 WT (±1)
H357A							WT	WT
H357Q							12 WT (±<1)	4.3 WT (±<1)

Figure 5.9: Relative activities of mutations (WT activity set as 100%) of ckAAs towards C2-C4 positions of sugar moiety (alanine scanning). The pink rectangle means original activities were retained at level 1 (<20%); the red rectangle means original activities were retained at level 0 (no activity). The olive green 'WT' indicates positive activity was detected in the WT (data from Chapter 3).

Based on the activity comparison between WT and mutants, and the MSA results, 89C1 D356 may be the ckAA as an alanine substitution resulted in a pronounced reduction of the original activities (2.0% original UDP-Rha and 12% original dTDP-Rha activities retained), with the possible role of binding Rha-O3 and Rha-O4.

The analysis indicated that 89C1 H357 might also be a ckAA (no Rha transfer activity of H357A was detected), possibly binding Rha-O2 and Rha-O3 (based on MSA data). Moreover, 89C1 H357 revealed the Rha recognition and discrimination role of this residue: the enzyme gained the UDP-Glc activity upon H357Q substitution (see MS in Appendix, Figure A.10). However, the Rha transfer activity of 89C1 was not fully abolished by substituting H with Q, as H357Q retained 12% of the original UDP-Rha activity and 4.3% of the original dTDP-Rha activity. It is possible that some piecemeal assembly of smaller functional domains in 89C1 from the perspective of enzyme evolution that may aid in Rha recognition, and thus retention of a partial Rha transfer activity²⁷⁷.

5.2.1.2 ckAAs towards C6 position of sugar moiety

The binding pattern at the C6 position of the sugar moiety is a major difference between GlcTs and RhaTs. Since there is no –OH group at C6 position of Rha, no direct hydrogen bonds can be formed between RhaTs and C6 position of Rha.

VvGT1	V	A	W	L	P	F	W	T										
71G1	I	P	S	Y	L	F	L	T										
72B1	V	P	P	Y	I	F	Y	P										
78G1	A	K	W	V	P	L	W	T										
78K6	V	P	W	I	A	F	W	P										
85H2	L	P	N	V	L	I	F	S										
OS79	V	P	P	Y	I	F	Y	S	<i>SpnG</i>	L	P	V	V	L	H	R	W	
78D1	A	T	W	V	A	F	W	A	78D1	M	S	L	A	L	P	R	-	
89C1	I	K	S	I	S	F	L	P	89C1	L	E	T	A	T	T	E	-	

Figure 5.10: MSA of potential ckAAs towards C6 position of sugar moiety using templates of seven GlcTs (left) and *SpnG* (right). The colours of residues represent their different properties: small and hydrophobic amino acids AVFPMILW are coloured red, acidic amino acids DE are coloured blue, basic amino acids RK are coloured magenta, and hydroxyl/sulphydryl/amine amino acids STYHCNGQ are coloured green. The last line of symbols denotes the degree of conservation: conserved (*), conserved mutations (:), semi-conserved mutations (.), and non-conserved mutations (o).

MSA analysis revealed that potential ckAAs towards C6 position of Rha are not conserved, with different results with different templates (Figure 5.10). When GlcTs were used as templates, 78D1 A141 and 89C1 P147 appeared at a site equivalent to VvGT1 T141 that is responsible for Glc-O6 binding. Since no hydrogen bonds can be formed between RhaTs and Rha, hydrophobic interactions may be mainly responsible for stacking the methyl group of C6²⁶³. This might explain why a handful of hydrophobic AAs appear in the vicinity of an equivalent site with potential ckAAs towards C6 position of Rha in 78D1 (136-WVAFWA-141) and 89C1 (143-ISFLPI-148). Subsequent SDM supported the important role of 78D1 A141 and 89C1 P147, as 78D1 A141T and 89C1 P147A abolished all Rha transfer activity (GAR panel, not shown: Rha activity was completely abolished in both mutants and no new donor activity was acquired). The activity loss may be explained by the destruction of the hydrophobic environment. In addition, 78D1 A141 and 89C1 P147 may be located in a linker that may affect the conformation of nearby domains (discussed in Chapter 4), thus affecting the location of ckAAs that bind the acceptor and abolishing the original activity.

A previous docking experiment of *SpnG* revealed the presence of the hydrophobic AA W142 near the methyl group at C6 position of Rha²⁶³. However, MSA analysis of *SpnG* and the target UGTs revealed no corresponding 78D1 and 89C1 AAs at sites equivalent to *SpnG* W142, which may be associated with a low conservation of ckAAs at C6 position and low sequence similarities of the N-terminal regions of the bacterial RhaT *SpnG* and plant RhaTs 78D1 and 89C1.

degree of conservation at T280 and H350, it is not surprising that T280A and H350A substitutions greatly reduced the original activity. Based on MSA prediction, T280 and H350 bound phosphate following the same pattern as that of VvGT1 (Figure 5.13).

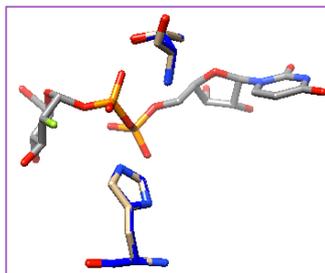


Figure 5.13: Phosphate binding site overlay of modelled 78D1 (blue) and crystal VvGT1 (tan) complex with U2F (PDB: 2C1Z). This Figure demonstrates 78D1 T280 and H350 are conserved to bind phosphate in the same manner as do VvGT1 T280 and H350.

A23 is observed at a site equivalent to VvGT1 T19. A hydrogen bond cannot be formed between A and phosphate, in contrast to that between T and phosphate, as the side chain of A does not have a propensity to form hydrogen bonds. However, mutant A23T only retained 2.4% of the original UDP-Rha activity and <1% of the original dTDP-Rha activity, indicating the significant role of A23 in 78D1. One possible explanation is that 78D1 A23 may be positioned near the phosphate group, providing a proper hydrophobic environment for phosphate accommodation. Thus, the substitution altered the hydrophobic environment and affected the original enzyme activities. Further, the additional side chain of T may be too long for the active site, causing steric hindrance, and disrupting the interaction between 78D1 and the substrate.

UGT 89C1: ckAAs towards phosphate moiety

	89C1							
	UDP-Glc	UDP-Gal	GDP-Glc	GDP-Fuc	UDP-GlcNAc	UDP-GalNAc	UDP-Rha	dTDP-Rha
G20A	Red	Red	Red	Red	Red	Red	120WT (±10)	120WT (±23)
S250A	Red	Red	Red	Red	Red	Red	29 WT (±1)	22 WT (±5)
H332A	Red	Red	Red	Red	Red	Red	5.5 WT (±1)	3.3 WT (±<1)

Figure 5.14: Relative activities of mutations (WT activity set as 100%) of ckAAs towards phosphate (alanine scanning). The green rectangle means original activities were retained at level 4 (>80%); the yellow rectangle means original activities were retained at level 2 (20%-50%); the pink rectangle means original activities were retained at level 1 (<20%); the red rectangle means original activities were retained at level 0 (no activity). The olive green 'WT' indicates positive activity was detected in the WT (data from Chapter 3).

At the equivalent ckAA site that is responsible for binding phosphate in template UGTs, 89C1 harbours G20, S250, and H332 (Figure 5.11). As indicated by MSA analysis, H is highly conserved and H332A substitution greatly reduced the original enzyme activities (5.5% of the original UDP-Rha activity and 3.3% of the original dTDP-Rha activity were retained). Hence, 89C1 H332 is regarded as a ckAA that may bind phosphate in the same manner as *Vv*GT1 H350 (Figure 5.14).

At the site equivalent to *Vv*GT1 T280, 89C1 harbours S, which possesses similar characteristics to T. Although S and T have different structures (T has an additional CH₃ group), the side-chain atoms N and OG of S can take the same positions as the side-chain atoms N and OG1 of T (Figure 5.15). This would indicate that S can interact with phosphate via a hydrogen bond. Subsequently, this hypothesis was confirmed by SDM data, since 89B1 S250A retained 29% of the original UDP-Rha activity and 22% of the original dTDP-Rha.

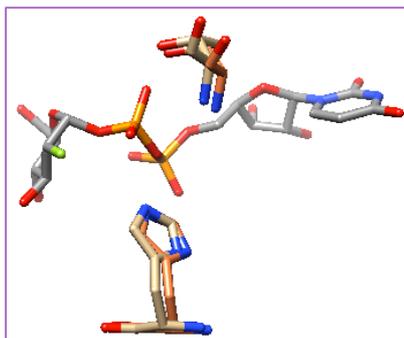


Figure 5.15: Phosphate binding site overlay of modelled 89C1 (coral) and crystal VvGT1 (tan) complex with U2F (PDB: 2C1Z). This Figure demonstrates 89C1S250 and H332 binds phosphate in the same manner as do VvGT1 T280 and H350.

The least conserved AA in 89C1 is G20. As an AA without any side chain, G probably does not interact with the substrate directly. Thus, most likely, G20 is not a ckAA interacting with phosphate; this was consistent with SDM revealing no activity loss by G20A. In fact, the enzyme activity was slightly enhanced in G20A. This may be because the nucleoside is always stacked in a hydrophobic environment and alanine substitution increased the degree of hydrophobicity of the environment.

5.2.3 Identification of ckAAs: the uridine moiety

As already mentioned, two AAs are responsible for binding uridine moiety in template UGTs: AAs 1 (conserved W) and 27 (conserved E) within the PSPG motif. Based on the MSA analysis (Figure 5.16), the potential ckAAs that interact with uridine at equivalent sites are also conserved in 78D1 and 89C1: both of them are W and E as the AAs 1 and 27, respectively, within the PSPG motif.

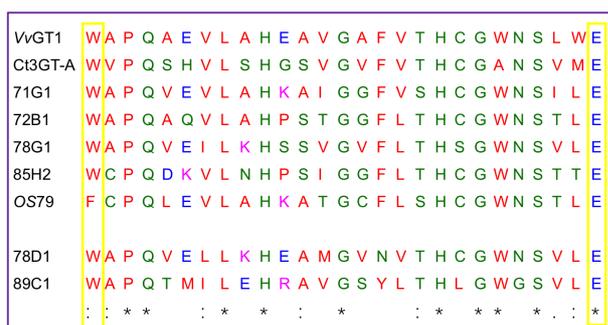


Figure 5.16: MSA of potential ckAAs towards uridine. The colours of residues represent their different properties: small and hydrophobic AAs AVFPMILW are coloured red; acidic AAs DE are coloured blue; basic AAs RK are coloured magenta; hydroxyl/sulphydryl/amine AAs STYHCNGQ are coloured green. The last line symbols denote the degree of conservation: conserved (*), conserved mutations (:), semi-conserved mutations (.), and non-conserved mutations (). This Figure demonstrates the two potential ckAAs towards uridine are located in the 1st and 27th position within the PSPG motif with a high degree of conservation.

UGT 78D1: ckAAs towards the uridine moiety

	78D1							
	UDP-glc	UDP-Gal	GDP-Glc	GDP-Fuc	UDP-GlcNAc	UDP-GalNAc	UDP-Rha	dTDP-Rha
W332A							<1 WT (±1)	<1 WT (±1)
E358A							14 WT (±2)	6.4 WT (±1)

Figure 5.17: Relative activities of mutations (WT activity set as 100%) of ckAAs towards uridine (alanine scanning). The pink rectangle means original activities were retained at level 1 (<20%); the red rectangle means original activities were retained at level 0 (no activity). The olive green 'WT' indicates positive activity was detected in the WT (data from Chapter 3).

Figure 5.17 shows the results of SDM analysis of potential ckAAs interacting with uridine in 78D1. Substitutions of both W332 and E358 greatly reduced the original enzyme activity (W332A retained <1% of the original UDP-Rha and dTDP-Rha activities; E358A retained 14% of the original UDP-Rha activity and 6.4% of the original dTDP-Rha activity). This indicated that W332 and E358 are important for enzyme activity, probably by binding uridine in the same manner as template UGTs, as suggested by MSA analysis. Further, this suggested that W332 might interact with the uracil ring by pi-pi interaction, with E358 interacting with ribose via hydrogen bonds (Figure 5.18). Moreover, substitution of W had a greater effect on enzyme activity than that of E, indicating that the impact of binding between W and the uracil ring may be greater than that of hydrogen bond formation between E and ribose.

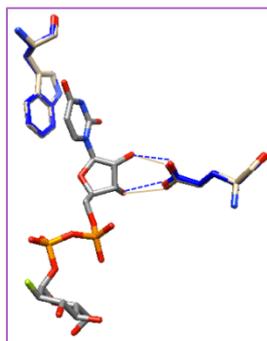


Figure 5.18: Uridine binding site overlay of modelled 78D1 (blue) and crystal VvGT1 (tan) complex with U2F (PDB: 2C1Z). The hydrogen bonds of VvGT1 and model UGTs towards U2F is represented by a solid and dash line respectively due to the interaction between VvGT1 and U2F being measured in crystal and the interaction between model UGTs and U2F being putative. This Figure demonstrates 78D1 W332 and E358 binds uridine in the same manner as VvGT1 W332 and E358.

UGT 89C1: ckAAs towards the uridine moiety

	89C1							
	UDP-Glc	UDP-Gal	GDP-Glc	GDP-Fuc	UDP-GlcNAc	UDP-GalNAc	UDP-Rha	dTDP-Rha
W314A							3.6 WT (±0.1)	2.9 WT (±0.4)
E340A							15 WT (±1)	17 WT (±2)

Figure 5.19: Relative activities of mutations (WT activity set as 100%) of ckAAs towards uridine (alanine scanning). The pink rectangle means original activities were retained at level 1 (<20%); the red rectangle means original activities were retained at level 0 (no activity). The olive green 'WT' indicates positive activity was detected in the WT (data from Chapter 3).

Alanine substitutions of 89C1 W314 and E340 (Figure 5.19) significantly reduced its original activity: 89C1 W314A retained 3.6% and 2.9% of the original UDP-Rha and dTDP-Rha activities, respectively; 89C1 E340A retained 15% and 17% of the original UDP-Rha and dTDP-Rha activities, respectively. Taken together with the MSA data, the reduction of activities by these substitutions may be associated with W314 and E340 binding to uracil and ribose, respectively, as the corresponding residues in VvGT1 (Figure 5.20). Further, A substitution of W resulted in a greater reduction of enzyme activity than that of E, which is consistent with the analysis of 78D1. This suggests that the pi-pi binding between W and the uracil ring plays a more significant role in the enzyme activity than that of E.

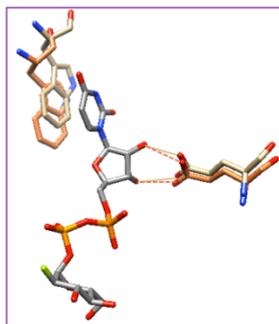


Figure 5.20: Uridine binding site overlay of modelled 89C1 (coral) and crystal VvGT1 (tan) complex with U2F (PDB: 2C1Z). The hydrogen bonds of VvGT1 and model UGTs towards U2F is represented by a solid and dash line respectively because the interaction between VvGT1 and U2F is measured in crystal and the interaction between model UGTs and U2F is putative. This Figure demonstrates 89C1W314 and E340 binds uridine in the same manner as VvGT1 W332 and E358.

Apart from UDP, 78D1 and 89C1 can also transfer Rha from dTDP. The structures of dTDP and UDP differ: dTDP has an additional CH₃ group at the C5 position of uracil and lacks an –OH group at the 2' position of ribose (Figure 5.21). Based on the results described above, both W and E play significant roles in dTDP-Rha activity, indicating that the potential ckAAs that interact with both UDP and dTDP may be conserved in 78D1 and 89C1. This suggests that the additional CH₃ group at the C5 position of uracil may not affect the formation of pi-pi interaction between W and the thymine ring, and the absence of –OH group at the 2' position of ribose may not disrupt the hydrogen bonding effect between E and 3' of ribose (Figure 5.21).

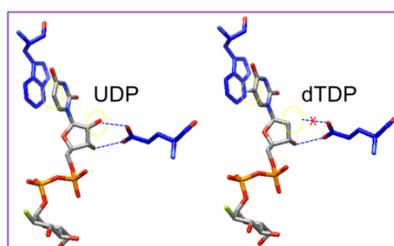


Figure 5.21: Putative UDP or dTDP binding site in 78D1 model. The structural difference of UDP has been highlighted in the yellow circle. This figure demonstrates how dTDP is tolerated in the active site of 78D1: the additional CH₃ group at C5 of the uracil ring in dTDP does not affect its pi-pi interaction with W. The missing –OH group at 2' of ribose disconnects its hydrogen bond towards E, but the hydrogen bond between 3' of ribose and E remains.

The model proposed above is supported by some examples. For instance, SpsA shows compatibility with UDP and dTDP sugars¹⁶³. In the crystal structure of the SpsA+UDP complex, the uracil ring is stabilised by an aromatic AA (F233) via pi-pi interaction and ribose is bound to a negatively charged AA (D98) via hydrogen bond¹⁶³, showing

the same binding pattern as template GlcTs. When dTDP was docked in the active site of SpsA, SpsA appeared to pose no steric barrier to dTDP occupying the place of UDP¹⁶³: the additional methyl group of dTDP is easily accommodated within the nucleotide-binding pocket and flanked on one side by the aromatic ring of F233; the 3' of ribose bound D98 in the same manner as that observed for UDP.

5.2.4 Homology model validation

As introduced in Chapter 4, a model is a good way to visualise the interaction between the enzyme and substrate, although it was not used to guide the design of the experiments. As discussed in Chapter 4, C-terminal models of 78D1 and 89C1 were constructed using the SWISS-MODEL²⁴⁹ program and the template UGTs that shared the highest percentage of identical AAs (Figure 5.22).

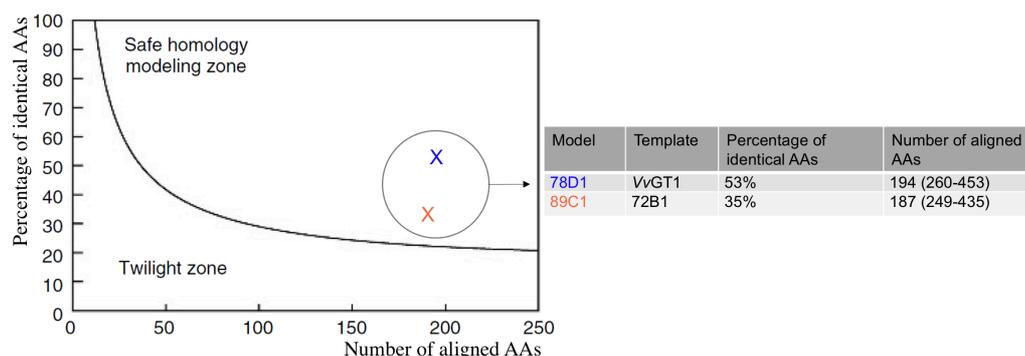


Figure 5.22: Homology modelling zone validation of 78D1 and 89C1. Left: Limit for the homology modelling defined by Rost (1999). The location of 'X' means the point of the modelling zone different target UGTs locate. They are represented by different colours corresponding to the name of the UGT shown in the table (right). All target UGTs are located in the safe homology modelling zone. Right: The percentage of identical residues between target and template UGT and the number of aligned residues of target UGT.

As shown above, the combinations of the percentage of identical AAs and the number of aligned AAs in both 78D1 and 89C1 fell in the 'safe homology modelling zone', by using VvGT1 and 72B1 as templates, respectively. This meant that homology modelling could indeed be applied for the template and target proteins²⁵⁰.

Table 5.3: Scoring of model quality validation

Model	Template	QMEAN	Residues in outlier region (Ramachandran plots)
78D1	VvGT1	-0.44	-
89C1	72B1	-2.76	1296

The qualities of 78D1 and 89C1 models (C-terminal) were determined. They both exhibited QMEAN scores that were in the recommended region in the Swiss-Model program (-4.00 to +4.00). More than 99% of all AAs were located in the allowed regions of Ramachandran plots. Every AA investigated in this chapter was in the allowed area based on Ramachandran plots. This scoring demonstrates the reliability of the constructed 78D1 and 89C1 models. Thus, the overlaying of target UGTs on template UGTs was reasonable and enabled useful interpretation of the SDM data.

5.3 Conclusions

As demonstrated in the current Chapter, 78D1 and 89C1 showed dual activities of UDP-Rha and dTDP-Rha (with KMP as the acceptor). The results of the MSA and SDM analyses were used to investigate the ckAAs of 78D1 and 89C1 (Figures 5.23 and 5.24).

The proposed active site of 78D1 is summarised in Figure 5.23. The potential ckAAs were deduced from MSA analysis. The levels of significance (shown by different colours) of the potential ckAAs are indicated by the percentage of the retained original activities in alanine mutants. Based on MSA, D374 and N375 might bind Rha; W332 and E358 probably bind uridine; and A23, W332, and H350 might interact with phosphate. Subsequently, SDM suggested that these potential ckAAs greatly affected enzyme activity (<20% of the original enzyme activities were retained by the alanine mutants) and could be regarded as ckAAs.

Furthermore, UDP-Rha and dTDP-Rha activities of 78D1 were successfully changed to UDP-Glc and UDP-GlcNAc activities, respectively, by substituting of N375 with Q.

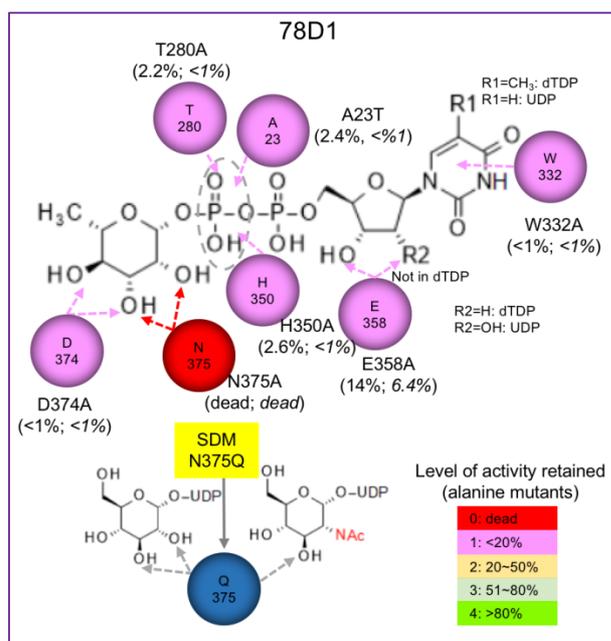


Figure 5.23: Potential ckAAs in 78D1 towards UDP-Rha and dTDP-Rha. The level of significance of potential ckAAs (from MSA) is based on the experimental results of alanine scanning and is indicated by colour: Level 0, Very Strong (alanine mutants retained original activity 0%, in red); Level 1, Strong (alanine mutants retained original activity <20%, in pink); Level 2, Medium (alanine mutants retained original activity 20~50%, in yellow); Level 3, Weak (alanine mutants retained original activity 51~80%, in light green); Level 4, Very Weak/No (alanine mutants retained original activity >80%, in green).

Similarly, the active site of 89C1 is proposed in Figure 5.24. The MSA analysis suggested the potential ckAAs. The percentage of the retained original activities in alanine mutants reflected the level of significance (shown by different colours). In summary, MSA analysis suggested that D356 and H357 are involved in Rha binding; W314 and E340 interact with uridine; and G20, S250, and H332 bind phosphate. Subsequent SDM analysis substantiated the notion that the AAs might be ckAAs (except for G20 and S250) as <20% of the original enzyme activities were retained by alanine mutants.

Furthermore, UDP-Glc activity was successfully conferred by the H357 point substitution.

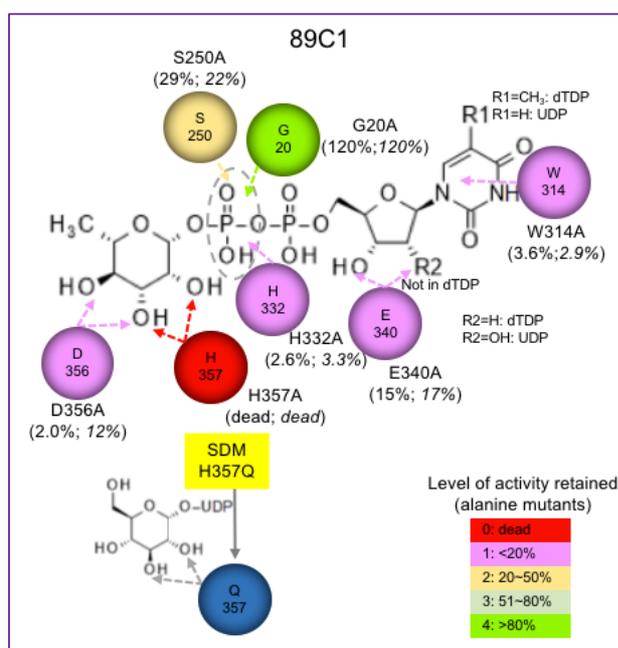


Figure 5.24: Potential ckAAs in 89C1 towards UDP-Rha and dTDP-Rha. The level of significance of potential ckAAs (from MSA) is based on the experimental results of alanine scanning and is indicated by colour: Level 0, Very Strong (alanine mutants retained original activity 0%, in red); Level 1, Strong (alanine mutants retained original activity <20%, in pink); Level 2, Medium (alanine mutants retained original activity 20~50%, in yellow); Level 3, Weak (alanine mutants retained original activity 51~80%, in light green); Level 4, Very Weak/No (alanine mutants retained original activity >80%, in green).

Final Conclusions

In this project, an MS-based method was developed to explore UGTs' substrate specificity and kinetics. Additionally, MSA using template UGTs with crystal structures was employed to investigate the ckAAs in target UGTs.

The MS-based method was developed in the study of WT recombinant UGTs from group B, D, F, H and L as the models (Chapter 3): 'full scan' mode of the MS was used to screen the potential glycosylated product after reaction; 'product ion' mode of the MS was used to confirm the formation of the glycosylated product; 'multiple reaction monitoring' (MRM) mode was used to quantify the formation of the glycosylated product as a function of reaction time. New activities were discovered (e.g. UDP-GlcNAc activity for 73B4 and 78D2 and dTDP-Rha activity for 78D1 and 89D1). Kinetic parameters not only revealed the catalytic capability and efficiency of specific UGTs, but also indicated the kinetic mechanisms (determined by the K_M ratio; e.g. GlcTs generally adopted a Bi-Bi random mechanism. 73B4 showed a donor dependent kinetic mechanism). The details of the full kinetic characterisation of tested UGTs were firstly described, providing clues to their metabolism *in vivo* and their use as biocatalysts in potential further applications such as glycoconjugates synthesis.

Due to the lack of structural information, MSA with existing crystal availability is probably one of the fastest approaches to understand ckAAs in target UGTs without crystallography data. By comparing target UGTs and seven template UGTs with known crystal structures, the potential ckAAs towards the donor (the C2, C3, C4 and C6 positions of sugar, phosphate and nucleotide base), were indicated. Subsequent alanine screening or similar AA swapping in the site of potential ckAAs (by SDM) substantiated/disproved the role of potential ckAAs. These findings showed that most potential ckAAs were conserved (especially those ckAAs within the PSPG motif) and might follow a similar or the same interaction pattern with template UGTs, in which alanine mutants showed a significant reduction in their original activity. However, a few (e.g. 78D2 D380, 73B5 Q398 and 89B1 Q390) did not follow the same trend, as their alanine mutants retained most activity, probably suggesting the catalytic mutation occurred during evolution. With a combination of MSA and SDM, this

project not only presented a systematic active site analysis of UGTs, but also showed the significance of specific ckAAs in active sites for individual UGTs (e.g. Figure 4.33), by means of experimentally-determined data, and shed light on the understanding of UGTs' active sites and gave detailed ckAAs information of specific UGTs.

Furthermore, we were also particularly interested in individual UGTs with specific functions. 78D1 and 89C1 transfer Rha rather than the Glc as do most UGTs. This project revealed that ckAAs in the RhaTs 78D1 and 89C1 followed a very similar interaction pattern to the GlcTs. Additionally, one of the ultimate targets is to manipulate GT activity freely based on an understanding of its ckAAs, as this will provide broader catalytic tools with desired activity. In this project, point mutagenesis on 78D1 N375Q acquired new UDP-GlcNAc and UDP-Glc activity, and on 89C1 H357Q acquired new UDP-Glc activity. Although the ckAAs for conferring sugar selectivity are still under investigation, the new activity shed fresh light on the active site manipulation with the UGTs.

In summary, the project not only provided specific knowledge of the individual GTs that have been tested, but also shed light on systematic understanding of UGTs, laying the foundation for a study of glycan structures and functions (especially in plants) as well as developing technologies with potential commercial value.

Future work

Although the aims setup in the beginning have been achieved, there are still many interesting questions and problems which arose during the study. We hereby propose the following areas to continue in this study.

1) To explore the *in vitro* activity of UGTs that showed no activity in the current setup. This will probably involve the optimisation of the experimental conditions (such as buffer, pH, temperature), checking protein folding (using CD) and using broader libraries (e.g. UDP-Xylose, UDP-Arabinose). Some of the current kinetic results will need to be optimised with a substrate range of $1/5-5K_M$ as discussed in Chapter 3. The *in vivo* function in plants will also be investigated (e.g. using gene knockouts).

2) To obtain crystal structures of representative UGTs, especially the complex with substrates in the active site. For example, 78D2 is an interesting enzyme, as it showed a broad range of donor specificity and relatively was not highly conserved in terms of its catalytic effects at some sites. Other interesting UGTs included 73B4 (dual activity with UDP-Glc and UDP-GlcNAc), 78D1 and 89C1 (dual activity with UDP-Rha and dTDP-Rha) and 73C5 (dual activity with UDP-Glc and UDP-Gal) due to their broad range of activities.

3) To continue exploring the ckAAs towards specific sites. A deeper understanding of the active site lays the foundation for a rational active site manipulation of UGTs with desired activities, which is crucial for the development of UGT genetic engineering. Some areas are recommended for further exploration: A) ckAAs towards the C6 position of the sugar moiety. As shown in Chapter 4, ckAAs towards the C6 position of the sugar were not highly conserved and MSA might not provide an accurate result due to the relatively low conservation in the N-terminal domain. Thus, manual MSA and the adjacent 5-10 AAs in the vicinity of the tested site in this project should be investigated; B) ckAAs towards the C2-C4 positions of the sugar moiety. As shown in Chapter 4, ckAAs towards these positions of sugar are significant as they determined the enzyme activity by interacting with the sugar moiety predominantly. In addition, this might relate to donor recognition and discrimination as donors differ mostly in the C2-C4 positions (e.g. Glc and GlcNAc). The identification of ckAAs

that 'recognise' a donor might expand to a region (by technology such as 'domain swapping') rather than a single ckAA.

Apart from these interesting areas, related experiments need to be optimised. For example, some technologies such as circular dichroism should be used to make sure that the point mutagenesis does not affect the stability and overall conformation of the protein. Detailed kinetic parameters (e.g. K_M and K_d) of the mutants should be measured to confirm the specific role of ckAAs (binding AA or catalytic AA), which could be achieved by associating with other technologies (e.g. Isothermal Titration Calorimetry, molecular docking and dynamic simulation studies)

Appendix

A.1 Bi-Bi enzymatic reaction kinetic equations

For Bi-Bi enzymatic reactions under steady-state assumptions, an ordered sequential reaction in both directions is:

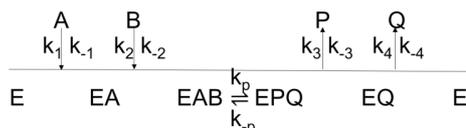


Figure A.1: Bi-Bi ordered sequential enzymatic mechanism. A, B: substrates; E: enzyme; P, Q: products. EAB: enzyme – substrate A – substrate B central complex; EPQ: enzyme – product P – product Q complex.

The velocity of the reaction is given:

$$v = \frac{k_1 k_2 k_3 k_4 [E_0][A][B] - k_{-1} k_{-2} k_{-3} k_{-4} [E_0][P][Q]}{k_{-1} k_4 (k_2 + k_3) + k_1 k_4 (k_{-2} + k_3)[A] + k_2 k_3 k_4 [B] + k_1 k_2 k_{-3} [A][P] + k_{-3} k_{-4} (k_{-1} + k_{-2}) [P][Q] + k_2 k_{-3} k_{-4} [B][P][Q] + k_1 k_2 k_{-3} [A][B][P] + k_1 k_2 (k_3 + k_4) [A][B] + k_2 k_3 k_{-4} [B][Q] + k_{-1} k_{-2} k_{-3} [P] + (k_{-2} + k_3) [Q]} \quad \text{Equation A.1}$$

For the initial velocity study, the product formation is very limited, i.e., $[P]=[Q]=0$.

The equation can be simplified:

$$v = \frac{V_{max}[A][B]}{K_{IA}K_B + K_B[A] + K_A[B] + [A][B]} \quad \text{Equation A.2}$$

Where

$$K_{IA} = \frac{k_{-1}}{k_1}: \text{inhibitor parameter of enzyme-substrate A;}$$

$$K_A = \frac{k_3 k_4 k_p}{k_1 (k_3 k_4 + k_3 k_p + k_4 k_{-p})}: \text{Michaelis parameter of substrate A;}$$

$$K_B = \frac{k_4 (k_{-2} k_3 + k_{-2} k_{-p} + k_3 k_p)}{k_2 (k_3 k_4 + k_3 k_p + k_4 k_{-p})}: \text{Michaelis parameter of substrate B;}$$

$$V_{max} = \frac{k_3 k_4 k_p}{k_3 k_4 + k_3 k_p + k_4 k_{-p}} * [E]: \text{maximum velocity reaction.}$$

In reciprocal form ($[B]$ is varied at fixed $[A]$),

$$\frac{1}{v} = \frac{K_B}{v_{max}} \left(1 + \frac{K_{IA}}{[A]} \right) \frac{1}{[B]} + \frac{1}{v_{max}} \left(1 + \frac{K_A}{[A]} \right) \quad \text{Equation A.3}$$

A random sequential mechanism, where the sequence of substrates binding with enzymes does not follow a compulsory order, has the following reaction schema:

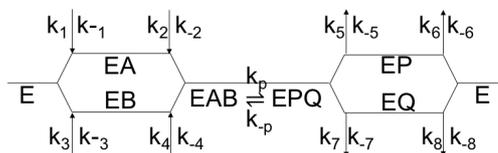


Figure A.2: Bi-Bi random sequential enzymatic mechanism. A, B: substrates; E: enzyme; P, Q: products. EAB: enzyme – substrate A – substrate B central complex; EPQ: enzyme – product P – product Q complex.

In the random sequential mechanism, the first binding step of one substrate towards an enzyme does not create an active enzyme complex. The transitory complex (EA complex or EB complex) will transfer to be active only when the co-substrate binds (Figure A.2). Hence, the first binding step is characterised by a pure dissociation parameter for the specific substrate. Due to the alternate reaction pathways, the random mechanism is defined by complex rate equations. Under the steady state assumption, a random Bi-Bi system will be defined by a very complicated rate equation. If there are no unusual rate constants, the random Bi-Bi system will appear as a rapid equilibrium random system²⁷⁸. The equation of Bi-Bi random sequential system can be simplified and made identical to that for the steady state ordered Bi-Bi system.

A.2 Confirmation of positive substrate specificity

Figures A.3, A.5 and A.6 show how the positive substrate specificities were confirmed using 78D2 as an example.

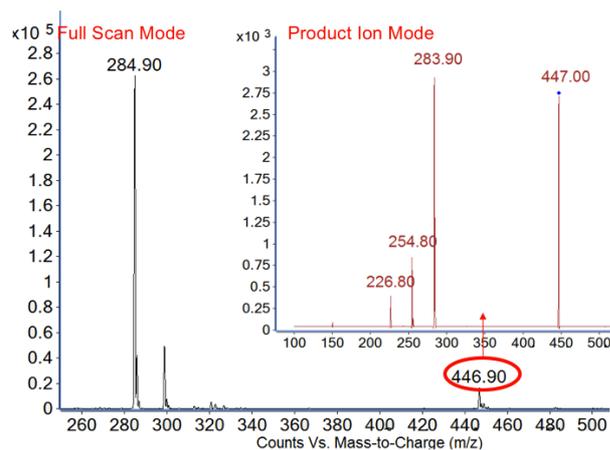


Figure A.3: MS of 78D2 positive activity with UDP-Glc. Peak at 447 in ‘full scan’ mode represents the product KMP-Glc, which fragments back to peak 284 (the aglycone) in ‘product ion’ mode. This chromatography is representative of other experiments with positive activity with UDP-Glc.

Figure A.3 demonstrates the positive activity of 78D2 towards UDP-Glc. In the negative-ion mode, the deprotonated molecular ions $[M-H^+]$ are usually the most abundant species. Thus, a peak of 447 $[KMP+Glc-H^+]$ revealed the presence of KMP-Glc. In some cases, the predominant ions were always accompanied by some salt adducts and, possibly, traces of dimeric or doubly charged materials. However, in this figure, these ions are too small to be observed.

Subsequent ‘product ion’ mode captured the ions of KMP-Glc (447) and fragmented them at a collision energy of 10 eV. Ions of 284/285 were observed, indicating the breaking of glycosidic bonds and the production of the aglycone (KMP). Consequently, the positive activity of 78D2 towards UDP-Glc was confirmed.

In ‘product ion’ mode, the predominant fragment was the m/z of the aglycone, i.e. KMP, which should give a peak of 285. However, a peak of 284 was also produced. This phenomenon indicated the glycosidic bond of KMP-Glc, when catalysed by 78D2, might be broken in distinct cleavage ways, i.e. homolytically and heterolytically. As introduced in Chapter 1, collision energy-induced homolytic or heterolytic cleavage might happen in negative electrospray ionisation tandem QQQ-MS for glycosylated flavonoids²⁷⁹. Thus, the fragments at 285 $[(KMP-Glc - Glc-H^+)]$ and 284 $[(KMP-Glc - Glc-H^+ - H^*)]$ were produced via heterolytic and homolytic cleavage of glycosidic bond respectively.

Apart from the predominant ions of KMP (284/285), ions of 227 and 255 were also observed. Previous studies showed that flavones generally exhibited neutral losses of CO (except genkwanin) that might possibly be attributed to the C ring²⁷⁹, which explained the peak at 255 representing the ion [KMP-H-CO-H*]⁻ and the peak at 227 representing the ion [KMP-H-2CO-H*]⁻ (Figure A.4).

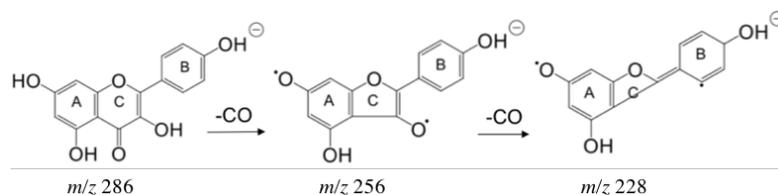


Figure A.4: Proposed fragmentation from KMP

Apart from UDP-Glc, 78D2 also showed positive activity towards UDP-Gal (Figure A.5 right), UDP-GlcNAc (Figure A.5 left) and GDP-Glc (Figure A.6), which were also confirmed in ‘full scan’ mode and ‘product ion’ mode of MS.

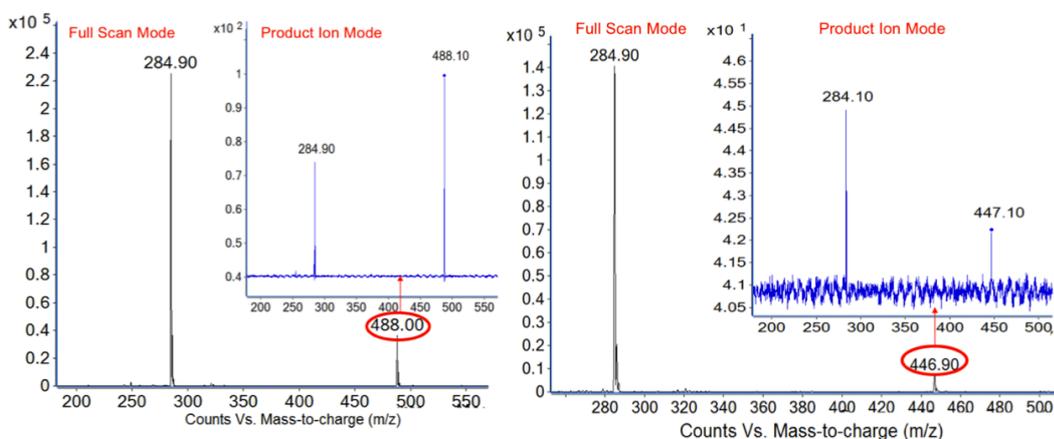


Figure A.5: MS of 78D2 positive activity with UDP-GlcNAc (left) and UDP-Gal (right). Peak of 488/447 in ‘full scan’ mode represents the product KMP-GlcNAc/KMP-Gal, which fragments back to peak 285 (aglycone) in ‘product ion’ mode. This chromatography is representative of other experiments with positive activity for UDP-GlcNAc and UDP-Gal.

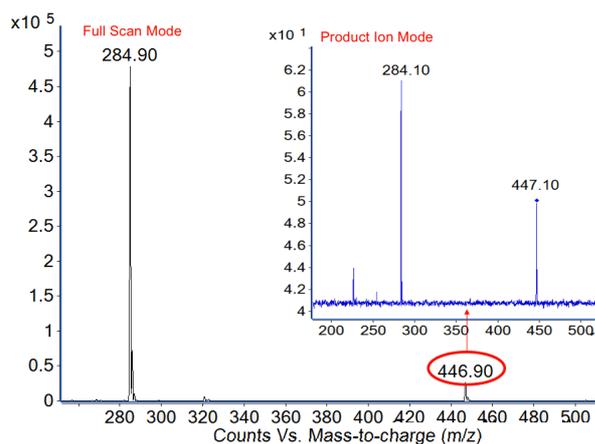


Figure A.6: MS of 78D2 positive activity with GDP-Glc. Peak at 447 in 'full scan' mode represents product KMP-Glc, which fragments back to peak 284 (aglycone) in 'product ion' mode.

Figures above demonstrate the confirmation of activity of 78D2 towards UDP-GlcNAc, UDP-Gal and GDP-Glc, with the ions of the Glycosylated-KMP (KMP-GlcNAc, KMP-Gal and KMP-Glc respectively) detected at 448, 447 and 447 respectively in 'full scan' mode of MS. Subsequently, the 'product ion' mode of MS showed that the aglycone KMP was observed, indicating that the glycosidic bonds of the Glycosylated-KMP were broken.

A.3 MS of RhaTs

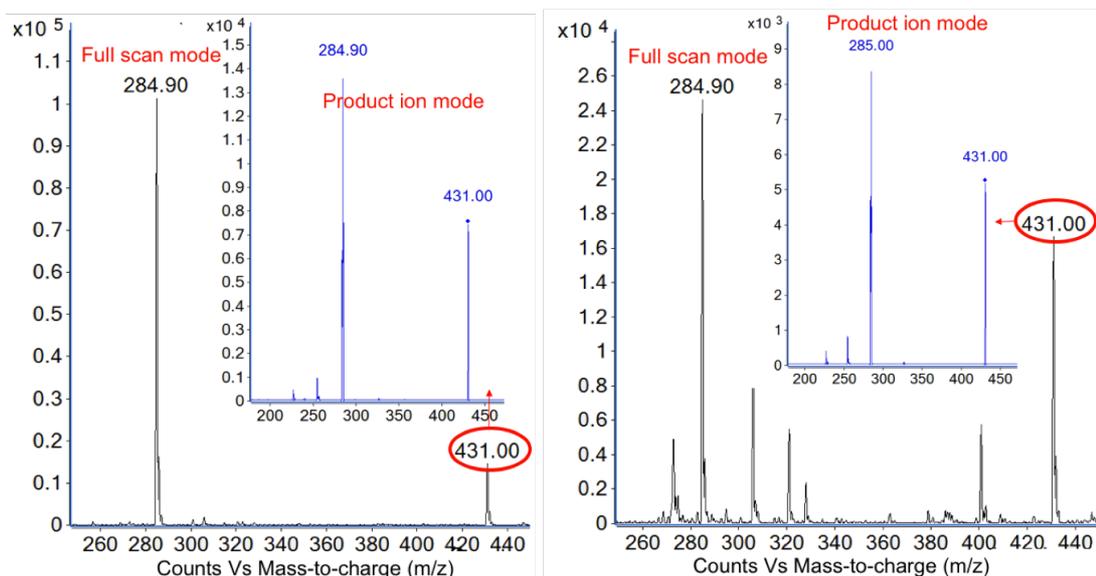


Figure A.7: MS of 78D1 positive activity with UDP-Rha (left) or dTDP-Rha (right) as donors. The product KMP-Rha is displayed as peak 431.00 in the negative-ion mode of MS. The 'product ion' shows that the product peak 431.00 can be fragmented into the aglycone fragment (284.90).

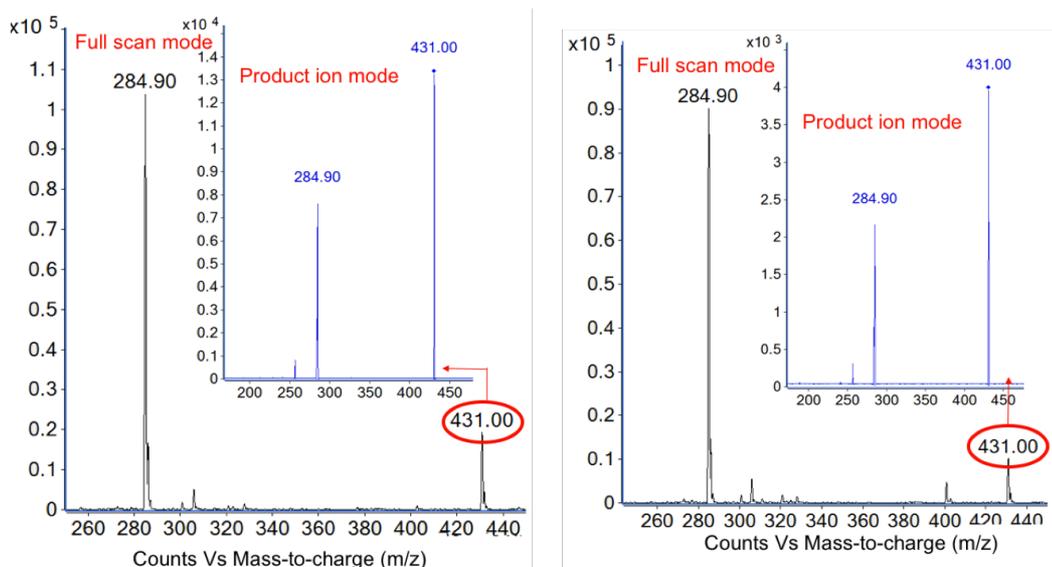


Figure A.8: MS of 89C1 positive activity with UDP-Rha (left) or dTDP-Rha (right) as donors. The product KMP-Rha is displayed as peak 431.00 in the negative-ion mode of MS. The ‘product ion’ mode shows the product peak 431.00 can be fragmented into the aglycone fragment (284.90).

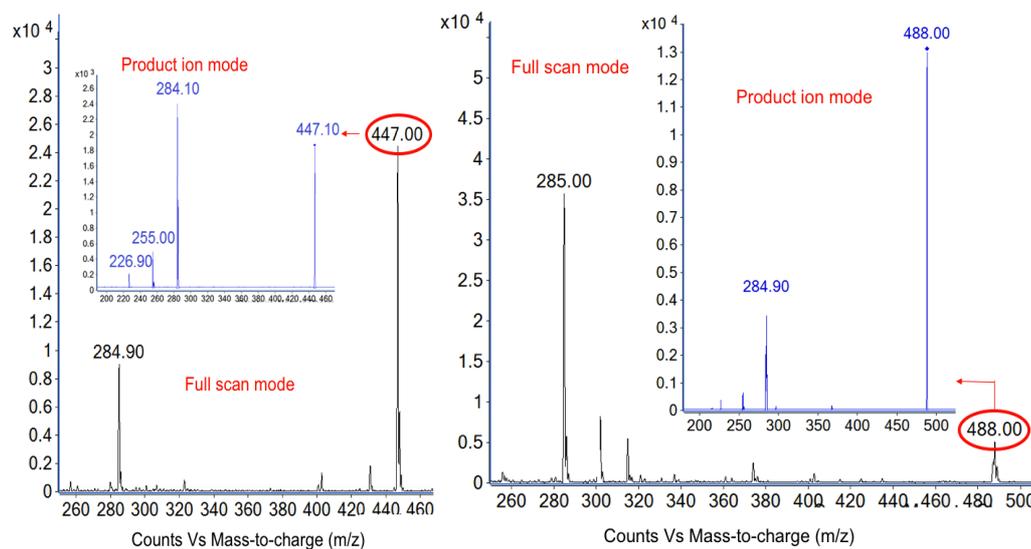


Figure A.9: MS of 78D1 N374Q positive activity with UDP-Glc (left) or UDP-GlcNAc (right) as donors. The products KMP-Glc and KMP-GlcNAc are displayed as peak 447.00 and 488.00 respectively in the negative-ion mode of MS. The ‘product ion’ mode shows the product peak can be fragmented into the aglycone fragment (284.10/284.90).

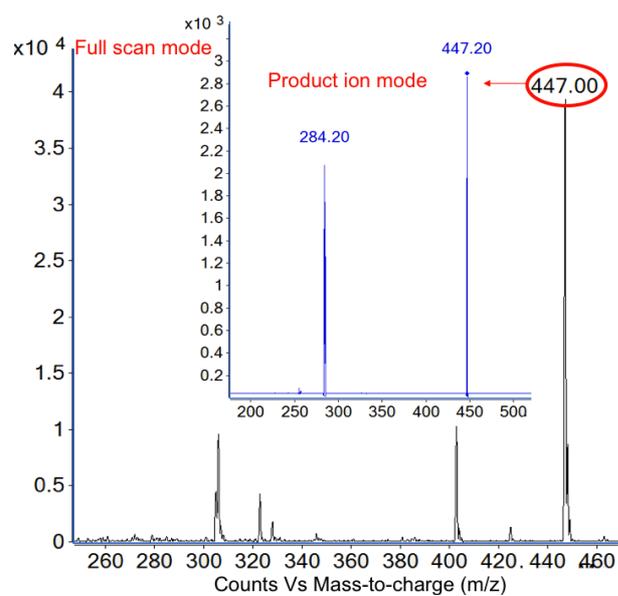


Figure A.10: MS of 89C1H357Q positive activity with UDP-Glc. The product KMP-Glc is displayed as peak 447.00. The 'product ion' mode of MS shows the peak of 447.00 can be fragmented as aglycone 284.20.

References

- 1 Varki, A. Evolutionary forces shaping the Golgi glycosylation machinery: why cell surface glycans are universal to living cells. *Cold Spring Harbor perspectives in biology* **3** (2011).
- 2 Huang, M. L. & Godula, K. Priming the cellular glycocalyx for neural development. *ACS Chem Neurosci* **5**, 873-875 (2014).
- 3 Cohen, M. Notable Aspects of Glycan-Protein Interactions. *Biomolecules* **5**, 2056-2072 (2015).
- 4 National Research Council (US) Committee. *Transforming Glycoscience: A Roadmap for the Future* (2012).
- 5 Hart, G. W. Thematic minireview series on glycobiology and extracellular matrices: glycan functions pervade biology at all levels. *The Journal of biological chemistry* **288**, 6903 (2013).
- 6 Varki, A. Biological roles of glycans. *Glycobiology* **27**, 3-49 (2017).
- 7 Kreisman, L. S. & Cobb, B. A. Infection, inflammation and host carbohydrates: a Glyco-Evasion Hypothesis. *Glycobiology* **22**, 1019-1030 (2012).
- 8 Varki, A., Kannagi, R., Toole, B. & Stanley, P. in *Essentials of Glycobiology* (eds rd et al.) 597-609 (2015).
- 9 Gul, A., Rahman, M. A. & Jaleel, A. Changes in glycosylated proteins in type-2 diabetic patients with and without complications. *J Ayub Med Coll Abbottabad* **17**, 33-37 (2005).
- 10 Marth, J. D. & Grewal, P. K. Mammalian glycosylation in immunity. *Nat Rev Immunol* **8**, 874-887 (2008).
- 11 Ruhaak, L. R., Miyamoto, S. & Lebrilla, C. B. Developments in the identification of glycan biomarkers for the detection of cancer. *Mol Cell Proteomics* **12**, 846-855 (2013).
- 12 Cooper, G. *The Cell: A Molecular Approach*. 832 (Sinauer Associates Inc; 7 Unbnd edition 2015).
- 13 Christiansen, M. N. et al. Cell surface protein glycosylation in cancer. *Proteomics* **14**, 525-546 (2014).
- 14 Ma, J. & Hart, G. W. O-GlcNAc profiling: from proteins to proteomes. *Clin Proteomics* **11**, 8 (2014).
- 15 Pinho, S. S. & Reis, C. A. Glycosylation in cancer: mechanisms and clinical implications. *Nat Rev Cancer* **15**, 540-555 (2015).
- 16 Berg, J. M., Tymoczko, J.L., Stryer, L. *Biochemistry*. 5 edn, (2002).
- 17 Varki, A. & Gagneux, P. Multifarious roles of sialic acids in immunity. *Annals of the New York Academy of Sciences* **1253**, 16-36 (2012).
- 18 Berenson, C. S. et al. Nontypeable Haemophilus influenzae-binding gangliosides of human respiratory (HEp-2) cells have a requisite lacto/neolacto core structure. *FEMS Immunol Med Microbiol* **45**, 171-182 (2005).
- 19 Lairson, L. L., Henrissat, B., Davies, G. J. & Withers, S. G. Glycosyltransferases: structures, functions, and mechanisms. *Annual review of biochemistry* **77**, 521-555 (2008).
- 20 Lloyd, K. O. The chemistry and immunochemistry of blood group A, B, H, and Lewis antigens: past, present and future. *Glycoconj J* **17**, 531-541 (2000).
- 21 Patenaude, S. I. et al. The structural basis for specificity in human ABO(H) blood group biosynthesis. *Nature structural biology* **9**, 685-690 (2002).
- 22 Hatano, K., Miyamoto, Y., Nonomura, N. & Kaneda, Y. Expression of gangliosides, GD1a, and sialyl paragloboside is regulated by NF-kappaB-dependent transcriptional control of alpha2,3-sialyltransferase I, II, and VI in human castration-resistant prostate cancer cells. *Int J Cancer* **129**, 1838-1847 (2011).
- 23 Stanley, P., Schachter, H. & Taniguchi, N. in *Essentials of Glycobiology* (eds nd et al.) (2009).
- 24 Miyoshi, E. et al. N-acetylglucosaminyltransferase III and V messenger RNA levels in LEC rats during hepatocarcinogenesis. *Cancer Res* **53**, 3899-3902 (1993).
- 25 Pinho, S. S. et al. The role of N-acetylglucosaminyltransferase III and V in the post-transcriptional modifications of E-cadherin. *Hum Mol Genet* **18**, 2599-2608 (2009).
- 26 Link-Lenczowski, P. et al. The glycomic effect of N-acetylglucosaminyltransferase III overexpression in metastatic melanoma cells. GnT-III modifies highly branched N-glycans. *Glycoconj J* **35**, 217-231 (2018).
- 27 Lu, J. et al. Expression of N-Acetylglucosaminyltransferase III Suppresses alpha2,3-Sialylation, and Its Distinctive Functions in Cell Migration Are Attributed to alpha2,6-Sialylation Levels. *The Journal of biological chemistry* **291**, 5708-5720 (2016).

- 28 Yoshimura, M. *et al.* Bisecting N-acetylglucosamine on K562 cells suppresses natural killer
cytotoxicity and promotes spleen colonization. *Cancer Res* **56**, 412-418 (1996).
- 29 Yang, X., Tang, J., Rogler, C. E. & Stanley, P. Reduced hepatocyte proliferation is the basis
of retarded liver tumor progression and liver regeneration in mice lacking N-
acetylglucosaminyltransferase III. *Cancer Res* **63**, 7753-7759 (2003).
- 30 Kizuka, Y. *et al.* An aberrant sugar modification of BACE1 blocks its lysosomal targeting in
Alzheimer's disease. *EMBO Mol Med* **7**, 175-189 (2015).
- 31 Kizuka, Y. *et al.* Bisecting GlcNAc modification stabilizes BACE1 protein under oxidative
stress conditions. *Biochem J* **473**, 21-30 (2016).
- 32 Taniguchi, N. & Kizuka, Y. Glycans and cancer: role of N-glycans in cancer biomarker,
progression and metastasis, and therapeutics. *Adv Cancer Res* **126**, 11-51 (2015).
- 33 Esko, J. D., Bertozzi, C. & Schnaar, R. L. in *Essentials of Glycobiology* (eds rd *et al.*)
701-712 (2015).
- 34 Jiang, H. *et al.* Modulating Cell-Surface Receptor Signaling and Ion Channel Functions by In
Situ Glycan Editing. *Angewandte Chemie* **57**, 967-971 (2018).
- 35 Offen, W. *et al.* Structure of a flavonoid glucosyltransferase reveals the basis for plant natural
product modification. *The EMBO journal* **25**, 1396-1405 (2006).
- 36 Bock, K. W. The UDP-glycosyltransferase (UGT) superfamily expressed in humans, insects
and plants: Animal-plant arms-race and co-evolution. *Biochemical pharmacology* **99**, 11-17
(2016).
- 37 Kim, B. G., Kim, H. J. & Ahn, J. H. Production of bioactive flavonol rhamnosides by
expression of plant genes in Escherichia coli. *J Agric Food Chem* **60**, 11143-11148 (2012).
- 38 <<http://www.cazy.org/GlycosylTransferases.html>>
- 39 Schlaeppli, K. & Mauch, F. Indolic secondary metabolites protect Arabidopsis from the
oomycete pathogen Phytophthora brassicae. *Plant Signal Behav* **5**, 1099-1101 (2010).
- 40 Wink, M. Evolution of secondary metabolites from an ecological and molecular phylogenetic
perspective. *Phytochemistry* **64**, 3-19 (2003).
- 41 Chu, H. Y., Wegel, E. & Osbourn, A. From hormones to secondary metabolism: the emergence
of metabolic gene clusters in plants. *The Plant journal : for cell and molecular biology* **66**, 66-
79 (2011).
- 42 Dudek, B., Warskulat, A. C. & Schneider, B. The Occurrence of Flavonoids and Related
Compounds in Flower Sections of Papaver nudicaule. *Plants (Basel)* **5** (2016).
- 43 Zhao, J. & Dixon, R. A. The 'ins' and 'outs' of flavonoid transport. *Trends in plant science* **15**,
72-80 (2010).
- 44 Chong, J. *et al.* Downregulation of a pathogen-responsive tobacco UDP-Glc:phenylpropanoid
glucosyltransferase reduces scopoletin glucoside accumulation, enhances oxidative stress, and
weakens virus resistance. *The Plant cell* **14**, 1093-1107 (2002).
- 45 Zhao, L. H., Ding, Y. X., Zhang, L. & Li, L. Cornel iridoid glycoside improves memory ability
and promotes neuronal survival in fimbria-fornix transected rats. *Eur J Pharmacol* **647**, 68-74
(2010).
- 46 Seeberger, P. H., Finney, N., Rabuka, D. & Bertozzi, C. R. in *Essentials of Glycobiology*
(eds nd *et al.*) (2009).
- 47 Tiwari, P., Sangwan, R. S. & Sangwan, N. S. Plant secondary metabolism linked
glycosyltransferases: An update on expanding knowledge and scopes. *Biotechnology
advances* **34**, 714-739 (2016).
- 48 Bhat, W. W. *et al.* Molecular characterization of UGT94F2 and UGT86C4, two
glycosyltransferases from Picrorhiza kurrooa: comparative structural insight and evaluation of
substrate recognition. *PloS one* **8**, e73804 (2013).
- 49 Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for
the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-1649 (2012).
- 50 Lim, E. K., Ashford, D. A., Hou, B., Jackson, R. G. & Bowles, D. J. Arabidopsis
glycosyltransferases as biocatalysts in fermentation for regioselective synthesis of diverse
quercetin glucosides. *Biotechnology and bioengineering* **87**, 623-631 (2004).
- 51 Noguchi, A. *et al.* Local differentiation of sugar donor specificity of flavonoid
glycosyltransferase in Lamiales. *The Plant cell* **21**, 1556-1572 (2009).
- 52 Cecilia, A., McIntosh, D. & Owens, K. Advances in flavonoid glycosyltransferase research:
integrating recent findings with long-term citrus studies. *Phytochemistry Reviews* **15**, 17
(2016).
- 53 Fan, J., Chen, C., Yu, Q., Li, Z. G. & Gmitter, F. G., Jr. Characterization of three terpenoid
glycosyltransferase genes in 'Valencia' sweet orange (Citrus sinensis L. Osbeck). *Genome /*

- National Research Council Canada = Genome / Conseil national de recherches Canada* **53**, 816-823 (2010).
- 54 Bowles, D., Lim, E. K., Poppenberger, B. & Vaistij, F. E. Glycosyltransferases of lipophilic small molecules. *Annual review of plant biology* **57**, 567-597 (2006).
- 55 Bowles, D., Isayenkova, J., Lim, E. K. & Poppenberger, B. Glycosyltransferases: managers of small molecules. *Curr Opin Plant Biol* **8**, 254-263 (2005).
- 56 Vogt, T. & Jones, P. Glycosyltransferases in plant natural product synthesis: characterization of a supergene family. *Trends in plant science* **5**, 380-386 (2000).
- 57 Gandia-Herrero, F. *et al.* Detoxification of the explosive 2,4,6-trinitrotoluene in Arabidopsis: discovery of bifunctional O- and C-glucosyltransferases. *The Plant journal : for cell and molecular biology* **56**, 963-974 (2008).
- 58 Marroun, S. *et al.* UGT74B1 from Arabidopsis thaliana as a versatile biocatalyst for the synthesis of desulfoglycosinolates. *Organic & biomolecular chemistry* **14**, 6252-6261 (2016).
- 59 Brazier-Hicks, M. *et al.* The C-glycosylation of flavonoids in cereals. *The Journal of biological chemistry* **284**, 17926-17934 (2009).
- 60 Glycosyltransferase family classification, <<http://www.cazy.org/GlycosylTransferases.html>>
- 61 Vogt, T. Substrate specificity and sequence analysis define a polyphyletic origin of betanidin 5- and 6-O-glucosyltransferase from Dorotheanthus bellidiformis. *Planta* **214**, 492-495 (2002).
- 62 Hoffmeister, D., Ichinose, K. & Bechthold, A. Two sequence elements of glycosyltransferases involved in urdamycin biosynthesis are responsible for substrate specificity and enzymatic activity. *Chem Biol* **8**, 557-567 (2001).
- 63 Gloster, T. M. Advances in understanding glycosyltransferases from a structural perspective. *Current opinion in structural biology* **28**, 131-141 (2014).
- 64 Cleland, W. W. The kinetics of enzyme-catalyzed reactions with two or more substrates or products. I. Nomenclature and rate equations. *BBA* **67**, 104-137 (1963).
- 65 Berg, J. M., Tymoczko, J. L. & Stryer, L. *The Michaelis-Menten Model Accounts for the Kinetic Properties of Many Enzymes* Ch. 8, (W. H. Freeman and Co. ; [Basingstoke : Palgrave] [distributor], 2001, 2002).
- 66 Truhlar, D. G. Transition state theory for enzyme kinetics. *Archives of biochemistry and biophysics* **582**, 10-17 (2015).
- 67 Johnson, K. A. A century of enzyme kinetic analysis, 1913 to 2013. *FEBS letters* **587**, 2753-2766 (2013).
- 68 Tom Zielinski, M. R., P. Scott Donover, and Robert G. Lowery. (2015).
- 69 Hisatomi, H. *et al.* Correlations between Chemical Compositions and Retention Times of Methacrylate Random Copolymers Using LC-ESI-MS. *Mass Spectrom (Tokyo)* **1**, A0012 (2012).
- 70 Edmond, D. H. Tandem mass spectrometry: A primer. *Journal of Mass Spectrometry* **31**, 129-137 (1996).
- 71 Yamagaki, T. & Watanabe, T. Hydrogen radical removal causes complex overlapping isotope patterns of aromatic carboxylic acids in negative-ion matrix-assisted laser desorption/ionization mass spectrometry. *Mass Spectrom (Tokyo)* **1**, A0005 (2012).
- 72 Blasco, C. & Picó, Y. in *Food Toxicants Analysis* 509-559 (Elsevier, 2007).
- 73 Jung, S. C. *et al.* Two ginseng UDP-glycosyltransferases synthesize ginsenoside Rg3 and Rd. *Plant & cell physiology* **55**, 2177-2188 (2014).
- 74 Okazawa, A. *et al.* Glucosyltransferase activity of Arabidopsis UGT71C1 towards pinoresinol and lariciresinol. *Plant Biotechnol* **31**, 561-566 (2014).
- 75 Vrieling, A., Ruger, W., Driessen, H. P. & Freemont, P. S. Crystal structure of the DNA modifying enzyme beta-glucosyltransferase in the presence and absence of the substrate uridine diphosphoglucose. *The EMBO journal* **13**, 3413-3422 (1994).
- 76 Shao, H. *et al.* Crystal structures of a multifunctional triterpene/flavonoid glycosyltransferase from Medicago truncatula. *The Plant cell* **17**, 3141-3154 (2005).
- 77 Li, L. *et al.* Crystal structure of Medicago truncatula UGT85H2--insights into the structural basis of a multifunctional (iso)flavonoid glycosyltransferase. *J Mol Biol* **370**, 951-963 (2007).
- 78 Modolo, L. V. *et al.* Crystal structures of glycosyltransferase UGT78G1 reveal the molecular basis for glycosylation and deglycosylation of (iso)flavonoids. *J Mol Biol* **392**, 1292-1302 (2009).
- 79 Hiromoto, T. *et al.* Structural basis for acceptor-substrate recognition of UDP-glucose: anthocyanidin 3-O-glucosyltransferase from Clitoria ternatea. *Protein science : a publication of the Protein Society* **24**, 395-407 (2015).

- 80 Hiromoto, T. *et al.* Crystal structure of UDP-glucose:anthocyanidin 3-O-glucosyltransferase
from *Clitoria ternatea*. *Journal of synchrotron radiation* **20**, 894-898 (2013).
- 81 Brazier-Hicks, M. *et al.* Characterization and engineering of the bifunctional N- and O-
glucosyltransferase involved in xenobiotic metabolism in plants. *Proceedings of the National
Academy of Sciences of the United States of America* **104**, 20238-20243 (2007).
- 82 Wetterhorn, K. M. *et al.* Crystal Structure of Os79 (Os04g0206600) from *Oryza sativa*: A
UDP-glucosyltransferase Involved in the Detoxification of Deoxynivalenol. *Biochemistry* **55**,
6175-6186 (2016).
- 83 Breton, C., Snajdrova, L., Jeanneau, C., Koca, J. & Imberty, A. Structures and mechanisms of
glycosyltransferases. *Glycobiology* **16**, 29R-37R (2006).
- 84 Osmani, S. A., Bak, S. & Moller, B. L. Substrate specificity of plant UDP-dependent
glycosyltransferases predicted from crystal structures and homology modeling.
Phytochemistry **70**, 325-347 (2009).
- 85 Paquette, S., Moller, B. L. & Bak, S. On the origin of family 1 plant glycosyltransferases.
Phytochemistry **62**, 399-413 (2003).
- 86 Hughes, J. & Hughes, M. A. Multiple secondary plant product UDP-glucose
glucosyltransferase genes expressed in cassava (*Manihot esculenta* Crantz) cotyledons. *DNA
Seq* **5**, 41-49 (1994).
- 87 Kumar, R., Sangwan, R. S., Mishra, S., Sabir, F. & Sangwan, S. 'In silico' motif diversity
analysis of the glycon preferentiality of plant secondary metabolic glycosyltransferases. *Plant
Omics* **5** (2012).
- 88 Kikuchi, N. & Narimatsu, H. Bioinformatics for comprehensive finding and analysis of
glycosyltransferases. *Biochim Biophys Acta* **1760**, 578-583 (2006).
- 89 Anand, P., Nagarajan, D., Mukherjee, S. & Chandra, N. ABS-Scan: In silico alanine scanning
mutagenesis for binding site residues in protein-ligand complex. *F1000Res* **3**, 214 (2014).
- 90 Fantini, J. & Barrantes, F. J. How cholesterol interacts with membrane proteins: an exploration
of cholesterol-binding sites including CRAC, CARC, and tilted domains. *Front Physiol* **4**, 31
(2013).
- 91 Pravda, L. *et al.* Anatomy of enzyme channels. *BMC Bioinformatics* **15**, 379 (2014).
- 92 Osmani, S. A., Bak, S., Imberty, A., Olsen, C. E. & Moller, B. L. Catalytic key amino acids
and UDP-sugar donor specificity of a plant glucuronosyltransferase, UGT94B1: molecular
modeling substantiated by site-specific mutagenesis and biochemical analyses. *Plant
physiology* **148**, 1295-1308 (2008).
- 93 Thorsoe, K. S. *et al.* Determination of catalytic key amino acids and UDP sugar donor
specificity of the cyanohydrin glycosyltransferase UGT85B1 from *Sorghum bicolor*.
Molecular modeling substantiated by site-specific mutagenesis and biochemical analyses.
Plant physiology **139**, 664-673 (2005).
- 94 Chang, A., Singh, S., Phillips, G. N., Jr. & Thorson, J. S. Glycosyltransferase structural
biology and its role in the design of catalysts for glycosylation. *Current opinion in
biotechnology* **22**, 800-808 (2011).
- 95 Guerin, M. E. *et al.* Molecular recognition and interfacial catalysis by the essential
phosphatidylinositol mannosyltransferase PimA from mycobacteria. *The Journal of biological
chemistry* **282**, 20705-20714 (2007).
- 96 Osawa, T. *et al.* Crystal structure of chondroitin polymerase from *Escherichia coli* K4.
Biochem Biophys Res Commun **378**, 10-14 (2009).
- 97 Gantt, R. W., Goff, R. D., Williams, G. J. & Thorson, J. S. Probing the aglycon promiscuity
of an engineered glycosyltransferase. *Angewandte Chemie* **47**, 8889-8892 (2008).
- 98 Williams, G. J., Zhang, C. & Thorson, J. S. Expanding the promiscuity of a natural-product
glycosyltransferase by directed evolution. *Nature chemical biology* **3**, 657-662 (2007).
- 99 Lovering, A. L. *et al.* Structure of the bacterial teichoic acid polymerase TagF provides
insights into membrane association and catalysis. *Nature structural & molecular biology* **17**,
582-589 (2010).
- 100 Steiner, K. & Schwab, H. Recent advances in rational approaches for enzyme engineering.
Comput Struct Biotechnol J **2**, e201209010 (2012).
- 101 He, X. Z., Wang, X. Q. & Dixon, R. A. Mutational analysis of the Medicago
glycosyltransferase UGT71G1 reveals residues that control regioselectivity for (Iso) flavonoid
glycosylation. *Journal of Biological Chemistry* **281**, 34441-34447 (2006).
- 102 Modolo, L. V., Escamilla-Trevino, L. L., Dixon, R. A. & Wang, X. Single amino acid
mutations of Medicago glycosyltransferase UGT85H2 enhance activity and impart
reversibility. *FEBS letters* **583**, 2131-2135 (2009).

- 103 Kubo, A., Arai, Y., Nagashima, S. & Yoshikawa, T. Alteration of sugar donor specificities of
plant glycosyltransferases by a single point mutation. *Archives of biochemistry and biophysics*
429, 198-203 (2004).
- 104 Ono, E. *et al.* Functional differentiation of the glycosyltransferases that contribute to the
chemical diversity of bioactive flavonol glycosides in grapevines (*Vitis vinifera*). *The Plant*
cell **22**, 2856-2871 (2010).
- 105 Barvkar, V. T., Pardeshi, V. C., Kale, S. M., Kadoo, N. Y. & Gupta, V. S. Phylogenomic
analysis of UDP glycosyltransferase 1 multigene family in *Linum usitatissimum* identified
genes with varied expression patterns. *BMC Genomics* **13**, 175 (2012).
- 106 Ricaldi, J. N. *et al.* Whole genome analysis of *Leptospira licerasiae* provides insight into
leptospirosis evolution and pathogenicity. *PLoS Negl Trop Dis* **6**, e1853 (2012).
- 107 Wagner, G. K. & Pesnot, T. Glycosyltransferases and their assays. *Chembiochem* **11**, 1939-
1949 (2010).
- 108 Dai, L. *et al.* Functional Characterization of Cucurbitadienol Synthase and Triterpene
Glycosyltransferase Involved in Biosynthesis of Mogrosides from *Siraitia grosvenorii*. *Plant*
& cell physiology **56**, 1172-1182 (2015).
- 109 Asada, K. *et al.* A 7-deoxyloganetic acid glycosyltransferase contributes a key step in
secologanin biosynthesis in Madagascar periwinkle. *The Plant cell* **25**, 4123-4134 (2013).
- 110 Nagatoshi, M. *et al.* UGT75L6 and UGT94E5 mediate sequential glucosylation of crocetin to
crocin in *Gardenia jasminoides*. *FEBS letters* **586**, 1055-1061 (2012).
- 111 Yu, H. S. *et al.* Characterization of glycosyltransferases responsible for salidroside
biosynthesis in *Rhodiola sachalinensis*. *Phytochemistry* **72**, 862-870 (2011).
- 112 Masada, S. *et al.* Functional and structural characterization of a flavonoid glucoside 1,6-
glucosyltransferase from *Catharanthus roseus*. *Plant & cell physiology* **50**, 1401-1415 (2009).
- 113 Ono, E., Ruike, M., Iwashita, T., Nomoto, K. & Fukui, Y. Co-pigmentation and flavonoid
glycosyltransferases in blue *Veronica persica* flowers. *Phytochemistry* **71**, 726-735 (2010).
- 114 Nakatsuka, T. & Nishihara, M. UDP-glucose:3-deoxyanthocyanidin 5-O-glucosyltransferase
from *Sinningia cardinalis*. *Planta* **232**, 383-392 (2010).
- 115 Wei, W. *et al.* Characterization of *Panax ginseng* UDP-Glycosyltransferases Catalyzing
Protopanaxatriol and Biosyntheses of Bioactive Ginsenosides F1 and Rh1 in Metabolically
Engineered Yeasts. *Molecular plant* **8**, 1412-1424 (2015).
- 116 Ohgami, S. *et al.* Volatile Glycosylation in Tea Plants: Sequential Glycosylations for the
Biosynthesis of Aroma beta-Primeverosides Are Catalyzed by Two *Camellia sinensis*
Glycosyltransferases. *Plant physiology* **168**, 464-477 (2015).
- 117 Augustin, J. M. *et al.* UDP-glycosyltransferases from the UGT73C subfamily in *Barbarea*
vulgaris catalyze sapogenin 3-O-glucosylation in saponin-mediated insect resistance. *Plant*
physiology **160**, 1881-1895 (2012).
- 118 Kovinich, N., Saleem, A., Arnason, J. T. & Miki, B. Functional characterization of a UDP-
glucose:flavonoid 3-O-glucosyltransferase from the seed coat of black soybean (*Glycine max*
(L.) Merr.). *Phytochemistry* **71**, 1253-1263 (2010).
- 119 Yang, M., Brazier, M., Edwards, R. & Davis, B. G. High-throughput mass-spectroscopy
monitoring for multisubstrate enzymes: Determining the kinetic parameters and catalytic
activities of glycosyltransferases. *Chembiochem* **6**, 346-357 (2005).
- 120 Desmarais, W. T. *et al.* The 1.20 Å resolution crystal structure of the aminopeptidase from
Aeromonas proteolytica complexed with tris: a tale of buffer inhibition. *Structure* **10**, 1063-
1072 (2002).
- 121 Ghalanbor, Z. *et al.* Binding of Tris to *Bacillus licheniformis* alpha-amylase can affect its
starch hydrolysis activity. *Protein Pept Lett* **15**, 212-214 (2008).
- 122 Bisswanger, H. Enzyme assays. *Perspectives in Science* **1**, 15 (2014).
- 123 Kim, N. A. *et al.* Effects of pH and buffer concentration on the thermal stability of etanercept
using DSC and DLS. *Biol Pharm Bull* **37**, 808-816 (2014).
- 124 Jin, S. H. *et al.* UGT74D1 is a novel auxin glycosyltransferase from *Arabidopsis thaliana*. *PLoS*
one **8**, e61705 (2013).
- 125 Sterling, H. J., Batchelor, J. D., Wemmer, D. E. & Williams, E. R. Effects of buffer loading
for electrospray ionization mass spectrometry of a noncovalent protein complex that requires
high concentrations of essential salts. *J Am Soc Mass Spectrom* **21**, 1045-1049 (2010).
- 126 Van Duijn, E. Current limitations in native mass spectrometry based structural biology. *J Am*
Soc Mass Spectrom **21**, 971-978 (2010).
- 127 UMassAmherst. <<https://www.umass.edu/massspec/sample-submission>> (2017).

- 128 Flint, J. *et al.* Structural dissection and high-throughput screening of mannosylglycerate
synthase. *Nature structural & molecular biology* **12**, 608-614 (2005).
- 129 Yang, M. *et al.* Probing the breadth of macrolide glycosyltransferases: in vitro remodeling of
a polyketide antibiotic creates active bacterial uptake and enhances potency. *J Am Chem Soc*
127, 9336-9337 (2005).
- 130 Yang, M., Davies, G. J. & Davis, B. G. A glycosynthase catalyst for the synthesis of flavonoid
glycosides. *Angewandte Chemie* **46**, 3885-3888 (2007).
- 131 Yu, M., Gao, G., Ding., Zhao, Y. & Sai, K. A review on body temperature of plants. *cje* **34**, 9
(2015).
- 132 Crawford, A. J., McLachlan, D. H., Hetherington, A. M. & Franklin, K. A. High temperature
exposure increases plant cooling capacity. *Curr Biol* **22**, R396-397 (2012).
- 133 Kirwan, A., Utratna, M., O'Dwyer, M. E., Joshi, L. & Kilcoyne, M. Glycosylation-Based
Serum Biomarkers for Cancer Diagnostics and Prognostics. *BioMed research international*
2015, 490531 (2015).
- 134 Zhang, H. *et al.* Methods for peptide and protein quantitation by liquid chromatography-
multiple reaction monitoring mass spectrometry. *Mol Cell Proteomics* **10**, M110 006593
(2011).
- 135 Pavarini, D. P., Pavarini, S. P., Niehues, M. & Lopes, N. P. Exogenous influences on plant
secondary metabolite levels. *Anim Feed Sci Tech* **176**, 5-16 (2012).
- 136 Nooralabettu, K. P. *Enzyme Technology: Pacemaker of Biotechnology.* (Prentice-Hall of
India Pvt.Ltd, 2011).
- 137 Bisswanger, H. *Enzyme Kinetics: Principles and Methods.* (2017).
- 138 Robert, J. & Renka. Multivariate interpolation of large sets of scattered data. *ACM*
Transactions on Mathematical Software (TOMS) **14**, 10 (1988).
- 139 Wang, Z. *et al.* Osmotic stress induces phosphorylation of histone H3 at threonine 3 in
pericentromeric regions of Arabidopsis thaliana. *Proceedings of the National Academy of*
Sciences of the United States of America **112**, 8487-8492 (2015).
- 140 Lee, S. Y. *et al.* Arabidopsis AtERF71/HRE2 functions as transcriptional activator via cis-
acting GCC box or DRE/CRT element and is involved in root development through regulation
of root cell expansion. *Plant Cell Rep* **34**, 223-231 (2015).
- 141 Brooks, H. B. *et al.* in *Assay Guidance Manual* (eds G. S. Sittampalam *et al.*) (2004).
- 142 Moraga, A. R., Mozos, A. T., Ahrazem, O. & Gomez-Gomez, L. Cloning and characterization
of a glucosyltransferase from Crocus sativus stigmas involved in flavonoid glucosylation.
BMC Plant Biol **9**, 109 (2009).
- 143 Loutre, C. *et al.* Isolation of a glucosyltransferase from Arabidopsis thaliana active in the
metabolism of the persistent pollutant 3,4-dichloroaniline. *The Plant journal : for cell and*
molecular biology **34**, 485-493 (2003).
- 144 Kanie, Y., Kirsch, A., Kanie, O. & Wong, C. H. Enzymatic assay of galactosyltransferase by
capillary electrophoresis. *Anal Biochem* **263**, 240-245 (1998).
- 145 Annesley, T. M. Ion suppression in mass spectrometry. *Clin Chem* **49**, 1041-1044 (2003).
- 146 Lim, E. K. *et al.* The activity of Arabidopsis glycosyltransferases toward salicylic acid, 4-
hydroxybenzoic acid, and other benzoates. *The Journal of biological chemistry* **277**, 586-592
(2002).
- 147 Langlois-Meurinne, M., Gachon, C. M. & Saindrenan, P. Pathogen-responsive expression of
glycosyltransferase genes UGT73B3 and UGT73B5 is necessary for resistance to
Pseudomonas syringae pv tomato in Arabidopsis. *Plant physiology* **139**, 1890-1901 (2005).
- 148 Hou, B., Lim, E. K., Higgins, G. S. & Bowles, D. J. N-glucosylation of cytokinins by
glycosyltransferases of Arabidopsis thaliana. *The Journal of biological chemistry* **279**, 47822-
47832 (2004).
- 149 Poppenberger, B. *et al.* Detoxification of the Fusarium mycotoxin deoxynivalenol by a UDP-
glucosyltransferase from Arabidopsis thaliana. *The Journal of biological chemistry* **278**,
47905-47914 (2003).
- 150 Poppenberger, B. *et al.* The UGT73C5 of Arabidopsis thaliana glucosylates brassinosteroids.
Proceedings of the National Academy of Sciences of the United States of America **102**, 15253-
15258 (2005).
- 151 Jones, P., Messner, B., Nakajima, J., Schaffner, A. R. & Saito, K. UGT73C6 and UGT78D1,
glycosyltransferases involved in flavonol glycoside biosynthesis in Arabidopsis thaliana. *The*
Journal of biological chemistry **278**, 43910-43918 (2003).

- 152 Tohge, T. *et al.* Functional genomics by integrated analysis of metabolome and transcriptome of Arabidopsis plants over-expressing an MYB transcription factor. *The Plant journal : for cell and molecular biology* **42**, 218-235 (2005).
- 153 Pourcel, L. *et al.* The formation of Anthocyanic Vacuolar Inclusions in Arabidopsis thaliana and implications for the sequestration of anthocyanin pigments. *Molecular plant* **3**, 78-90 (2010).
- 154 Messner, B., Thulke, O. & Schaffner, A. R. Arabidopsis glucosyltransferases with activities toward both endogenous and xenobiotic substrates. *Planta* **217**, 138-146 (2003).
- 155 Yonekura-Sakakibara, K., Tohge, T., Niida, R. & Saito, K. Identification of a flavonol 7-O-rhamnosyltransferase gene determining flavonoid pattern in Arabidopsis by transcriptome coexpression analysis and reverse genetics. *The Journal of biological chemistry* **282**, 14932-14941 (2007).
- 156 Yang, M., Brazier, M., Edwards, R. & Davis, B. G. High-throughput mass-spectrometry monitoring for multisubstrate enzymes: determining the kinetic parameters and catalytic activities of glycosyltransferases. *Chembiochem : a European journal of chemical biology* **6**, 346-357 (2005).
- 157 Li, Y., Baldauf, S., Lim, E. K. & Bowles, D. J. Phylogenetic analysis of the UDP-glycosyltransferase multigene family of Arabidopsis thaliana. *The Journal of biological chemistry* **276**, 4338-4343 (2001).
- 158 Hedstrom, L. Enzyme Specificity and Selectivity. *eLS* (2010).
- 159 Tokuriki, N. & Tawfik, D. S. Protein dynamism and evolvability. *Science* **324**, 203-207 (2009).
- 160 Tokuriki, N. & Tawfik, D. S. Stability effects of mutations and protein evolvability. *Current opinion in structural biology* **19**, 596-604 (2009).
- 161 Ramakrishnan, B., Shah, P. S. & Qasba, P. K. alpha-Lactalbumin (LA) stimulates milk beta-1,4-galactosyltransferase I (beta 4Gal-T1) to transfer glucose from UDP-glucose to N-acetylglucosamine. Crystal structure of beta 4Gal-T1 x LA complex with UDP-Glc. *The Journal of biological chemistry* **276**, 37665-37671 (2001).
- 162 Wu, K. *et al.* Histidine-193 of rat glucosylceramide synthase resides in a UDP-glucose- and inhibitor (D-threo-1-phenyl-2-decanoylamino-3-morpholinopropan-1-ol)-binding region: a biochemical and mutational study. *Biochem J* **341** (Pt 2), 395-400 (1999).
- 163 Tarbouriech, N., Charnock, S. J. & Davies, G. J. Three-dimensional structures of the Mn and Mg dTDP complexes of the family GT-2 glycosyltransferase SpsA: a comparison with related NDP-sugar glycosyltransferases. *J Mol Biol* **314**, 655-661 (2001).
- 164 Patel, S. M. *et al.* High throughput discovery of heteroaromatic-modifying enzymes allows enhancement of novobiocin selectivity. *Chem Commun (Camb)* **47**, 10569-10571 (2011).
- 165 Koshland, D. E. The application and usefulness of the ratio $k(\text{cat})/K(\text{M})$. *Bioorg Chem* **30**, 211-213 (2002).
- 166 Kopycki, J. *et al.* Kinetic analysis of Arabidopsis glucosyltransferase UGT74B1 illustrates a general mechanism by which enzymes can escape product inhibition. *Biochem J* **450**, 37-46 (2013).
- 167 Kreppel, L. K. & Hart, G. W. Regulation of a cytosolic and nuclear O-GlcNAc transferase. Role of the tetratricopeptide repeats. *The Journal of biological chemistry* **274**, 32015-32022 (1999).
- 168 Lazarus, M. B., Nam, Y., Jiang, J., Sliz, P. & Walker, S. Structure of human O-GlcNAc transferase and its complex with a peptide substrate. *Nature* **469**, 564-567 (2011).
- 169 Lazarus, M. B. *et al.* Structural snapshots of the reaction coordinate for O-GlcNAc transferase. *Nature chemical biology* **8**, 966-968 (2012).
- 170 Jackson, R. G. *et al.* Over-expression of an Arabidopsis gene encoding a glucosyltransferase of indole-3-acetic acid: phenotypic characterisation of transgenic lines. *The Plant journal : for cell and molecular biology* **32**, 573-583 (2002).
- 171 Tognetti, V. B. *et al.* Perturbation of indole-3-butyric acid homeostasis by the UDP-glucosyltransferase UGT74E2 modulates Arabidopsis architecture and water stress tolerance. *The Plant cell* **22**, 2660-2679 (2010).
- 172 Hoffmann, E. D. & Stroobant, V. t. *Mass spectrometry : principles and applications*. 3rd edn. (Wiley ; Chichester : John Wiley [distributor], 2007).
- 173 Rupp, B. *Biomolecular crystallography : principles, practice, and application to structural biology*. (Garland Science, 2010).
- 174 Slabinski, L. *et al.* The challenge of protein structure determination--lessons from structural genomics. *Protein science : a publication of the Protein Society* **16**, 2472-2482 (2007).

- 175 Hans, J., Brandt, W. & Vogt, T. Site-directed mutagenesis and protein 3D-homology modelling suggest a catalytic mechanism for UDP-glucose-dependent betanidin 5-O-glucosyltransferase from *Dorotheanthus bellidiformis*. *The Plant journal : for cell and molecular biology* **39**, 319-333 (2004).
- 176 Kerdpin, O., Mackenzie, P. I., Bowalgaha, K., Finel, M. & Miners, J. O. Influence of N-terminal domain histidine and proline residues on the substrate selectivities of human UDP-glucuronosyltransferase 1A1, 1A6, 1A9, 2B7, and 2B10. *Drug metabolism and disposition: the biological fate of chemicals* **37**, 1948-1955 (2009).
- 177 Korprasertthaworn, P. *et al.* Effects of amino acid substitutions at positions 33 and 37 on UDP-glucuronosyltransferase 1A9 (UGT1A9) activity and substrate selectivity. *Biochemical pharmacology* **84**, 1511-1521 (2012).
- 178 Lewis, B. C. *et al.* Amino terminal domains of human UDP-glucuronosyltransferases (UGT) 2B7 and 2B15 associated with substrate selectivity and autoactivation. *Biochemical pharmacology* **73**, 1463-1473 (2007).
- 179 Kufareva, I. & Abagyan, R. Methods of protein structure comparison. *Methods Mol Biol* **857**, 231-257 (2012).
- 180 Dai, X. *et al.* Identification of a Flavonoid Glucosyltransferase Involved in 7-OH Site Glycosylation in Tea plants (*Camellia sinensis*). *Sci Rep* **7**, 5926 (2017).
- 181 Mamoon Rehman, H. *et al.* Genome-wide analysis of Family-1 UDP-glycosyltransferases in soybean confirms their abundance and varied expression during seed development. *J Plant Physiol* **206**, 87-97 (2016).
- 182 Wang, L. *et al.* Comparing the acceptor promiscuity of a *Rosa hybrida* glucosyltransferase RhGT1 and an engineered microbial glucosyltransferase OleD(PSA) toward a small flavonoid library. *Carbohydrate research* **368**, 73-77 (2013).
- 183 Fan, B., Chen, T., Zhang, S., Wu, B. & He, B. Mining of efficient microbial UDP-glycosyltransferases by motif evolution cross plant kingdom for application in biosynthesis of salidroside. *Sci Rep* **7**, 463 (2017).
- 184 Magdalou, J., Fournel-Gigleux, S. & Ouzzine, M. Insights on membrane topology and structure/function of UDP-glucuronosyltransferases. *Drug metabolism reviews* **42**, 159-166 (2010).
- 185 Grandits, M., Michlmayr, H., Sygmund, C. & Oostenbrink, C. Calculation of substrate binding affinities for a bacterial GH78 rhamnosidase through molecular dynamics simulations. *J Mol Catal B Enzym* **92**, 34-43 (2013).
- 186 McWilliam, H. *et al.* Analysis Tool Web Services from the EMBL-EBI. *Nucleic acids research* **41**, W597-600 (2013).
- 187 Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539 (2011).
- 188 Jurate, D., Aisling, O. & Roy, D. An Overview of Multiple Sequence Alignments and Cloud Computing in Bioinformatics. *ISRN Biomath* **2013** (2013).
- 189 Kemena, C. & Notredame, C. Upcoming challenges for multiple sequence alignment methods in the high-throughput era. *Bioinformatics* **25**, 2455-2465 (2009).
- 190 Keith, J. M. *Bioinformatics*. (Humana ; [London : Springer, distributor], 2008).
- 191 Bauer, J. C., Wright, D. A., Braman, J. C. & Geha, R. S. (Google Patents, 1998).
- 192 Liu, H. & Naismith, J. H. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol* **8**, 91 (2008).
- 193 Tee, K. L. & Wong, T. S. Polishing the craft of genetic diversity creation in directed evolution. *Biotechnology advances* **31**, 1707-1721 (2013).
- 194 *Medical and health genomics*. Vol. 1102 (2012 American Chemical Society, 2012).
- 195 Al Dallal, S., Wolton, K. & Hentges, K. E. Zfp521 promotes B-cell viability and cyclin D1 gene expression in a B cell culture system. *Leuk Res* **46**, 10-17 (2016).
- 196 Hubbard, B. P. *et al.* Continuous directed evolution of DNA-binding proteins to improve TALEN specificity. *Nat Methods* **12**, 939-942 (2015).
- 197 Zhang, Y. *et al.* Site-directed mutagenesis of porcine pepsin: Possible role of Asp32, Thr33, Asp215 and Gly217 in maintaining the nuclease activity of pepsin. *Enzyme Microb Technol* **89**, 69-75 (2016).
- 198 Sinha, N. & Nussinov, R. Point mutations and sequence variability in proteins: redistributions of preexisting populations. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 3139-3144 (2001).

- 199 Hilser, V. J., Dowdy, D., Oas, T. G. & Freire, E. The structural distribution of cooperative interactions in proteins: analysis of the native state ensemble. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 9903-9908 (1998).
- 200 Worth, C. L., Preissner, R. & Blundell, T. L. SDM--a server for predicting effects of mutations on protein stability and malfunction. *Nucleic acids research* **39**, W215-222 (2011).
- 201 Wang, H., Julenius, K., Hryhorenko, J. & Hagen, F. K. Systematic Analysis of proteoglycan modification sites in *Caenorhabditis elegans* by scanning mutagenesis. *The Journal of biological chemistry* **282**, 14586-14597 (2007).
- 202 Persson, K. *et al.* Crystal structure of the retaining galactosyltransferase LgtC from *Neisseria meningitidis* in complex with donor and acceptor sugar analogs. *Nature structural biology* **8**, 166-175 (2001).
- 203 Kaczmarek, R. *et al.* Evaluation of an amino acid residue critical for the specificity and activity of human Gb3/CD77 synthase. *Glycoconj J* **33**, 963-973 (2016).
- 204 Verhamme, I. M. A novel antithrombin domain dictates the journey's end of a proteinase. *The Journal of biological chemistry* **292**, 16521-16522 (2017).
- 205 Pires, A. S., Porto, W. F., Franco, O. L. & Alencar, S. A. In silico analyses of deleterious missense SNPs of human apolipoprotein E3. *Sci Rep* **7**, 2509 (2017).
- 206 Pandurangan, A. P., Ascher, D. B., Thomas, S. E. & Blundell, T. L. Genomes, structural biology and drug discovery: combating the impacts of mutations in genetic disease and antibiotic resistance. *Biochemical Society transactions* **45**, 303-311 (2017).
- 207 Liljebblad, A., Kallio, P., Vainio, M., Niemi, J. & Kanerva, L. T. Formation and hydrolysis of amide bonds by lipase A from *Candida antarctica*; exceptional features. *Organic & biomolecular chemistry* **8**, 886-895 (2010).
- 208 Wey, M. *et al.* Kinetic Mechanism of Formation of Hyperactive Embryonic Ras in Cells. *Biochemistry* **55**, 543-559 (2016).
- 209 Malik, V. & Black, G. W. Structural, functional, and mutagenesis studies of UDP-glycosyltransferases. *Adv Protein Chem Struct Biol* **87**, 87-115 (2012).
- 210 Chang, A. *et al.* Complete set of glycosyltransferase structures in the calicheamicin biosynthetic pathway reveals the origin of regioselectivity. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 17649-17654 (2011).
- 211 Patana, A. S., Kurkela, M., Goldman, A. & Finel, M. The human UDP-glucuronosyltransferase: identification of key residues within the nucleotide-sugar binding site. *Molecular pharmacology* **72**, 604-611 (2007).
- 212 Sobhany, M., Dong, J. & Negishi, M. Two-step mechanism that determines the donor binding specificity of human UDP-N-acetylhexosaminyltransferase. *The Journal of biological chemistry* **280**, 23441-23445 (2005).
- 213 Perez-Victoria, I. *et al.* Saturation transfer difference NMR reveals functionally essential kinetic differences for a sugar-binding repressor protein. *Chem Commun (Camb)*, 5862-5864 (2009).
- 214 Mazumder, M. *et al.* Investigations on Binding Pattern of Kinase Inhibitors with PPARgamma: Molecular Docking, Molecular Dynamic Simulations, and Free Energy Calculation Studies. *PPAR Res* **2017**, 6397836 (2017).
- 215 Morrison, K. L. & Weiss, G. A. Combinatorial alanine-scanning. *Current opinion in chemical biology* **5**, 302-307 (2001).
- 216 Chau, N. *et al.* Morphine glucuronidation and glucosidation represent complementary metabolic pathways that are both catalyzed by UDP-glucuronosyltransferase 2B7: kinetic, inhibition, and molecular modeling studies. *J Pharmacol Exp Ther* **349**, 126-137 (2014).
- 217 George Thompson, A. M., Iancu, C. V., Neet, K. E., Dean, J. V. & Choe, J. Y. Differences in salicylic acid glucose conjugations by UGT74F1 and UGT74F2 from *Arabidopsis thaliana*. *Sci Rep* **7**, 46629 (2017).
- 218 Hu, Y. *et al.* Crystal structure of the MurG:UDP-GlcNAc complex reveals common structural principles of a superfamily of glycosyltransferases. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 845-849 (2003).
- 219 Isiorho, E. A., Jeon, B. S., Kim, N. H., Liu, H. W. & Keatinge-Clay, A. T. Structural studies of the spinosyn forosaminyltransferase, SpnP. *Biochemistry* **53**, 4292-4301 (2014).
- 220 Tam, H. K. *et al.* Structural characterization of O- and C-glycosylating variants of the landomycin glycosyltransferase LanGT2. *Angewandte Chemie* **54**, 2811-2815 (2015).
- 221 Claesson, M., Siitonen, V., Dobritzsch, D., Metsa-Ketela, M. & Schneider, G. Crystal structure of the glycosyltransferase SnogD from the biosynthetic pathway of nogalamycin in *Streptomyces nogalater*. *FEBS J* **279**, 3251-3263 (2012).

- 222 Bolam, D. N. *et al.* The crystal structure of two macrolide glycosyltransferases provides a blueprint for host cell antibiotic immunity. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 5336-5341 (2007).
- 223 Meech, R. *et al.* Identification of residues that confer sugar selectivity to UDP-glycosyltransferase 3A (UGT3A) enzymes. *The Journal of biological chemistry* **287**, 24122-24130 (2012).
- 224 Ma, X. *et al.* Substrate specificity provides insights into the sugar donor recognition mechanism of O-GlcNAc transferase (OGT). *PLoS one* **8**, e63452 (2013).
- 225 Aharoni, A. *et al.* The 'evolvability' of promiscuous protein functions. *Nat Genet* **37**, 73-76 (2005).
- 226 Nair, P. C., Meech, R., Mackenzie, P. I., McKinnon, R. A. & Miners, J. O. Insights into the UDP-sugar selectivities of human UDP-glycosyltransferases (UGT): a molecular modeling perspective. *Drug metabolism reviews*, 1-11 (2015).
- 227 George, A. J. *An introduction to hydrogen bonding.* (Oxford University Press, 1997).
- 228 Meech, R. *et al.* A novel function for UDP glycosyltransferase 8: galactosidation of bile acids. *Molecular pharmacology* **87**, 442-450 (2015).
- 229 He, X. Z., Wang, X. & Dixon, R. A. Mutational analysis of the Medicago glycosyltransferase UGT71G1 reveals residues that control regioselectivity for (iso)flavonoid glycosylation. *The Journal of biological chemistry* **281**, 34441-34447 (2006).
- 230 Hartl, D. L. Evolving theories of enzyme evolution. *Genetics* **122**, 1-6 (1989).
- 231 Mulichak, A. M. *et al.* Structure of the TDP-epi-vancosaminyltransferase GtfA from the chloroeremomycin biosynthetic pathway. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 9238-9243 (2003).
- 232 Mulichak, A. M., Losey, H. C., Walsh, C. T. & Garavito, R. M. Structure of the UDP-glucosyltransferase GtfB that modifies the heptapeptide aglycone in the biosynthesis of vancomycin group antibiotics. *Structure* **9**, 547-557 (2001).
- 233 Holliday, G. L., Almonacid, D. E., Mitchell, J. B. & Thornton, J. M. The chemistry of protein catalysis. *J Mol Biol* **372**, 1261-1277 (2007).
- 234 Majeed, H., Zia, M., Yang, M., Sheikh, M. & Bhatti, I. Cloning and Site Directed Mutagenesis of UGT76E1 Leads to Changed Substrate Activity in Arabidopsis thaliana. *International Journal of Agriculture & Biology* **17**, 1125 (2015).
- 235 Yan, B. X. & Sun, Y. Q. Glycine residues provide flexibility for enzyme active sites. *The Journal of biological chemistry* **272**, 3190-3194 (1997).
- 236 Koshland, D. E. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proceedings of the National Academy of Sciences of the United States of America* **44**, 98-104 (1958).
- 237 Holliday, G. L., Fischer, J. D., Mitchell, J. B. & Thornton, J. M. Characterizing the complexity of enzymes on the basis of their mechanisms and structures with a bio-computational analysis. *FEBS J* **278**, 3835-3845 (2011).
- 238 Witt, D. Recent Developments in Disulfide Bond Formation. *Synthesis* **16** (2008).
- 239 Creighton, T. E. Disulphide bonds and protein stability. *Bioessays* **8**, 57-63 (1988).
- 240 Thangudu, R. R. *et al.* Analysis on conservation of disulphide bonds and their structural features in homologous protein domain families. *BMC Struct Biol* **8**, 55 (2008).
- 241 Dombkowski, A. A., Sultana, K. Z. & Craig, D. B. Protein disulfide engineering. *FEBS letters* **588**, 206-212 (2014).
- 242 Perry, L. J. & Wetzel, R. Disulfide bond engineered into T4 lysozyme: stabilization of the protein toward thermal inactivation. *Science* **226**, 555-557 (1984).
- 243 Daviere, J. M. & Achard, P. Gibberellin signaling in plants. *Development* **140**, 1147-1151 (2013).
- 244 Chen, D. *et al.* Identification of secret agent as the O-GlcNAc transferase that participates in Plum pox virus infection. *J Virol* **79**, 9381-9387 (2005).
- 245 Creixell, P., Schoof, E. M., Tan, C. S. & Linding, R. Mutational properties of amino acid residues: implications for evolvability of phosphorylatable residues. *Philos Trans R Soc Lond B Biol Sci* **367**, 2584-2593 (2012).
- 246 Mulichak, A. M., Lu, W., Losey, H. C., Walsh, C. T. & Garavito, R. M. Crystal structure of vancosaminyltransferase GtfD from the vancomycin biosynthetic pathway: interactions with acceptor and nucleotide ligands. *Biochemistry* **43**, 5170-5180 (2004).
- 247 Velloso, L. M., Bhaskaran, S. S., Schuch, R., Fischetti, V. A. & Stebbins, C. E. A structural basis for the allosteric regulation of non-hydrolysing UDP-GlcNAc 2-epimerases. *EMBO Rep* **9**, 199-205 (2008).

- 248 Miley, M. J. *et al.* Crystal structure of the cofactor-binding domain of the human phase II drug-
249 metabolism enzyme UDP-glucuronosyltransferase 2B7. *J Mol Biol* **369**, 498-511 (2007).
- 249 Schwede, T., Kopp, J., Guex, N. & Peitsch, M. C. SWISS-MODEL: An automated protein
homology-modeling server. *Nucleic acids research* **31**, 3381-3385 (2003).
- 250 Essen, L. O. Structural Bioinformatics. Edited by Philip E. Bourne and Helge Weissig.
Angewandte Chemie **42**, 4993-4993 (2003).
- 251 Benkert, P., Biasini, M. & Schwede, T. Toward the estimation of the absolute quality of
individual protein structure models. *Bioinformatics* **27**, 343-350 (2011).
- 252 Benkert, P., Kunzli, M. & Schwede, T. QMEAN server for protein model quality estimation.
Nucleic acids research **37**, W510-514 (2009).
- 253 Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. The SWISS-MODEL workspace: a web-
based environment for protein structure homology modelling. *Bioinformatics* **22**, 195-201
(2006).
- 254 Caffall, K. H. & Mohnen, D. The structure, function, and biosynthesis of plant cell wall pectic
polysaccharides. *Carbohydrate research* **344**, 1879-1900 (2009).
- 255 Nguema-Ona, E. *et al.* Cell wall O-glycoproteins and N-glycoproteins: aspects of biosynthesis
and function. *Front Plant Sci* **5**, 499 (2014).
- 256 Kocincova, D. & Lam, J. S. Structural diversity of the core oligosaccharide domain of
Pseudomonas aeruginosa lipopolysaccharide. *Biochemistry (Mosc)* **76**, 755-760 (2011).
- 257
- 258 Frydman, A. *et al.* The molecular and enzymatic basis of bitter/non-bitter flavor of citrus fruit:
evolution of branch-forming rhamnosyltransferases under domestication. *The Plant journal :
for cell and molecular biology* **73**, 166-178 (2013).
- 259 Rojas Rodas, F. *et al.* Linkage mapping, molecular cloning and functional analysis of soybean
gene Fg2 encoding flavonol 3-O-glucoside (1 --> 6) rhamnosyltransferase. *Plant Mol Biol* **84**,
287-300 (2014).
- 260 Shibuya, M., Nishimura, K., Yasuyama, N. & Ebizuka, Y. Identification and characterization
of glycosyltransferases involved in the biosynthesis of soyasaponin I in *Glycine max*. *FEBS
letters* **584**, 2258-2264 (2010).
- 261 Steiner, K., Wojciechowska, A., Schaffer, C. & Naismith, J. H. Purification, crystallization
and preliminary crystallographic analysis of WsaF, an essential rhamnosyltransferase from
Geobacillus stearothermophilus. *Acta crystallographica. Section F, Structural biology and
crystallization communications* **64**, 1163-1165 (2008).
- 262 Steiner, K., Hagelueken, G., Messner, P., Schaffer, C. & Naismith, J. H. Structural basis of
substrate binding in WsaF, a rhamnosyltransferase from *Geobacillus stearothermophilus*. *J
Mol Biol* **397**, 436-447 (2010).
- 263 Isiorho, E. A., Liu, H. W. & Keatinge-Clay, A. T. Structural studies of the spinosyn
rhamnosyltransferase, SpnG. *Biochemistry* **51**, 1213-1222 (2012).
- 264 Steiner, K. *et al.* Molecular basis of S-layer glycoprotein glycan biosynthesis in *Geobacillus
stearothermophilus*. *The Journal of biological chemistry* **283**, 21120-21133 (2008).
- 265 Yoon, J. A. *et al.* Production of a novel quercetin glycoside through metabolic engineering of
Escherichia coli. *Appl Environ Microbiol* **78**, 4256-4262 (2012).
- 266 Abuelizz, H. A. & Mahmud, T. Distinct Substrate Specificity and Catalytic Activity of the
Pseudoglycosyltransferase VldE. *Chem Biol* **22**, 724-733 (2015).
- 267 Bar-Peled, M. & O'Neill, M. A. Plant nucleotide sugar formation, interconversion, and salvage
by sugar recycling. *Annual review of plant biology* **62**, 127-155 (2011).
- 268 Ma, Y., Pan, F. & McNeil, M. Formation of dTDP-rhamnose is essential for growth of
mycobacteria. *Journal of bacteriology* **184**, 3392-3395 (2002).
- 269 Grzegorzewicz, A. E. *et al.* Development of a microtitre plate-based assay for lipid-linked
glycosyltransferase products using the mycobacterial cell wall rhamnosyltransferase WbbL.
Microbiology **154**, 3724-3730 (2008).
- 270 Bar-Peled, M., Lewinsohn, E., Fluhr, R. & Gressel, J. UDP-rhamnose:flavanone-7-O-
glucoside-2"-O-rhamnosyltransferase. Purification and characterization of an enzyme
catalyzing the production of bitter compounds in citrus. *The Journal of biological chemistry*
266, 20953-20959 (1991).
- 271 Yin, R. *et al.* Kaempferol 3-O-rhamnoside-7-O-rhamnoside is an endogenous flavonol
inhibitor of polar auxin transport in *Arabidopsis* shoots. *The New phytologist* **201**, 466-475
(2014).

- 272 Kuhn, B. M. *et al.* 7-Rhamnosylated Flavonols Modulate Homeostasis of the Plant Hormone Auxin and Affect Plant Development. *The Journal of biological chemistry* **291**, 5385-5395 (2016).
- 273 Choi, H. J., Song, J. H., Park, K. S. & Kwon, D. H. Inhibitory effects of quercetin 3-rhamnoside on influenza A virus replication. *Eur J Pharm Sci* **37**, 329-333 (2009).
- 274 Choi, H. J., Song, J. H. & Kwon, D. H. Quercetin 3-rhamnoside exerts antiinfluenza A virus activity in mice. *Phytother Res* **26**, 462-464 (2012).
- 275 Davies, G. J., Ducros, V. M., Varrot, A. & Zechel, D. L. Mapping the conformational itinerary of beta-glycosidases by X-ray crystallography. *Biochemical Society transactions* **31**, 523-527 (2003).
- 276 Kim, H. S. *et al.* Engineering flavonoid glycosyltransferases for enhanced catalytic efficiency and extended sugar-donor selectivity. *Planta* **238**, 683-693 (2013).
- 277 Stafford, H. A. Flavonoid evolution: an enzymic approach. *Plant physiology* **96**, 680-685 (1991).
- 278 Segel, I. H. *Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems.* (1993).
- 279 Furstenberg-Hagg, J., Zagobelny, M. & Bak, S. Plant defense against insect herbivores. *International journal of molecular sciences* **14**, 10242-10297 (2013).