Structural and Biochemical studies of
Arabidopsis thaliana Glycosyltransferases

Aishat Mojisola Akere

This thesis is submitted in accordance with the requirements of University College London
for the degree of Doctor of Philosophy

Supervisors
Prof. Shozeb Haider (Primary)
Prof. Michael Heinrich (Secondary)
Declaration

I, Aishat M. Akere confirm that the work presented in this thesis is my own. I also confirm that I have indicated in the thesis where information has been derived from other sources.

Signed: ..................

Date: 23/03/2020
Abstract

Background: Glycosylation of secondary metabolites involves plant UDP-dependent glycosyltransferases (UGTs). UGTs have shown promising potential as drug targets as well as catalysts in the synthesis of glycosides of medicinal importance. However, limited understanding at the molecular level due to insufficient biochemical and structural information has hindered potential applications of most of these UGTs. For example, only two crystal structures of Arabidopsis thaliana UGTs are currently solved of the 122 genes available. In addition, more than half of these UGTs are yet to be biochemically characterised.

Aims: This research aims to i) investigate qualitatively substrate specificities of Arabidopsis thaliana UGTs from selected families using mass spectrometry (MS) based methods and to study the kinetic parameters via bioluminescence (Chapter 3); ii) produce and study homology models of the UGTs (novel) to further understand their substrate preferences and key catalytic amino acid residues involved (Chapter 4); and lastly iii) manipulate rationally the active sites of UGTs to engineer mutant UGTs of improved donor substrate activity (Chapter 5).

Methodology: Direct monitoring of products of glycosylation was done using triple quadrupole mass spectrometry (QQQ-MS) as it involves limited substrate modification. Full scan and product ion screening modes identifies the potential glycosylated product and confirms the product formation respectively. The kinetic data of the UGTs was determined via UDP-Glo glycosyltransferase assay which measured the amount of UDP released as a function of time (Chapter 3). Homology modeling was employed in the absence of experimental crystal structures to identify structural differences in these UGTs which drive substrate preferences. Docking of ligand substrates into the model UGTs was done to understand interactions at the molecular level (Chapter 4). Site directed mutagenesis was used to produce mutant UGTs to substantiate the functional roles of potential key amino acids. These mutations were rationally (sequence-based and structure-based) designed (Chapter 5).

Results and conclusions: 22 recombinant UGTs from groups L, H and D were selected for substrate screening. 15 of these were successfully expressed while 8 UGTs show glycosylation activity. 76E1 displayed the highest acceptor substrate recognition while both 76E5 and 76E1 showed highest donor recognition. Very low Km at μM scale suggests enzymes good affinity for the donor substrates with 76E5 showing stronger preference for UDP-Gal and UDP-GlcNAc (Chapter 3). Homology models of five group H UGTs were constructed, validated and substrate ligands docked into them. With a focus on donor sugar interactions, key amino acid residues interacting at specific positions of each model UGT were shown. In addition, a major structural difference in N3/Nα3 region of 76E1 was found which may be responsible for its higher acceptor substrate recognition (Chapter 4).
The usefulness and predictive power of these models helped design mutant UGTs. Rationally designed mutant UGTs such as 76E2 N320S, 76E4 K275L, 76D1 P129T and 76E2 D374E displayed improved substrate recognition which also highlights the functional roles of those amino acid residues (Chapter 5).
Impact statement

This research has contributed to the understanding of mechanism of plant GTs. With the growing usefulness and potential of natural products in disease management and treatment, glycosylation of plant secondary metabolites forms a crucial mechanism in biogeneration of their glycoconjugates with medicinal properties. This study has identified wildtype UGT76E1 which show promising substrate recognition (both acceptor and donor). Such broad spectrum of activity with plant secondary metabolites has a great biotechnological relevance in the synthesis of glycoconjugates. In addition, broad spectrum GTs are acceptable candidates for enzyme manipulation to create novel and improved enzymes in glycoengineering.

Glycosyltransferase research is drifting towards focus on the synthesis of bioactive glycoconjugates via metabolic engineering and manipulation of enzyme's active site to create improved/desirable catalytic properties. Therefore, in this research such active sites manipulation has led to discovery of some mutant UGTs 76E4 K275L, 76E2 N320S, 76E4 Q375A and 76D1 P129T with improved properties. These can serve as basis for biocatalytic studies of their usage in bioactive glycosides production. The numerous advantages of glycosylation in plant secondary metabolomics indicates the rising importance of the GTs. In the near future, the enzyme superfamily may serve as promising path for widening drug targets for pharmacophore discovery and development.

Glycosylation of some bioactive molecules at specific position alter their pharmaceutical properties. Achieving this specific glycosylation via organic chemistry has presented challenges. However, enzymatic glycosylation is a better option than organic chemistry commercially in the synthesis of specific glycoconjugates. Therefore, the in vitro biochemical characterisation studies of GTs presented in this research may help in achieving this. Substrate preferences of GTs have been investigated and our findings will guide further usage in this regard.

Since structural information is key to understanding enzyme catalytic mechanism, the utilization of protein modeling and docking studies helped in unravelling structural complexity, enzyme functionality and designing rational manipulations. Hence, the challenge of absence of crystal structures for the studied UGTs was surmounted. This research has further confirmed the benefit of usage of homology modeling in GTs structural studies.
Acknowledgements

I give all thanks to Allaah, who without His favours and sustenance this would not have been possible.

I am grateful to the Federal Scholarship Board/Presidential Special Scholarship Scheme for Innovation and Development (PRESSID) Nigeria, for funding my postgraduate studies. An immense appreciation goes to my primary supervisor Prof. Shozeb Haider who came at the best and most needed period of this research. I have learnt a lot from your expertise in a short time and got a lot of help which has helped my research to completion. I also thank my secondary supervisor Prof. Michael Heinrich for his valuable advice and support through the years. I acknowledge my former primary supervisor Dr. Min Yang for the knowledge imparted and the Research Department of Pharmaceutical and Biological Chemistry for the facilities provided which ensured the execution of this research work. I also thank my colleagues in the study space 424 and lab 409 who has kept my company over the years.

How could I have done this without the inestimable support of my family? I express my deepest heartfelt gratitude to my husband Hussein Akere, who has supported me in every way to achieve this big dream. The huge sacrifices you have endured and still enduring to keep me going is praiseworthy. Please keep being you!

To my mum who has trained and nurtured me, who has travelled several times across the seas to babysit for me, thanks for your encouragement and prayers always.

To my sons, Jaabir and Zayd for the family time lost and every other sacrifices, I promise to make it up to you.

To my sisters (Zaynab, Hannat and Rasheedah) who have stayed ever supportive over the years, thanks for the beautiful love.

To my friends and in-laws, thanks for the well wishes.

To the Olosos who received me on my first day this postgraduate sojourn began, I won’t forget this favour.

To everyone who believe in me, who have encouraged me and helped me along this journey

Lastly, I dedicate this thesis to my late dad who I feel sad that he is never able to read this piece.
Current publication from work presented in this thesis

# Table of Contents

Declaration ......................................................................................................................... 2

Abstract .............................................................................................................................. 3

Impact statement ................................................................................................................ 5

Acknowledgement .............................................................................................................. 6

Current publications from work presented in this thesis .................................................. 7

Table of Content ............................................................................................................... 8

List of figures .................................................................................................................... 12

List of tables ..................................................................................................................... 15

List of abbreviations ........................................................................................................ 16

Chapter 1 Introduction .................................................................................................... 18

1.1 Glycosyltransferases (GTs) ....................................................................................... 19

1.2 Plant UGTs ................................................................................................................. 20

1.2.1 General features and importance ......................................................................... 20

1.2.2 Crystal structures .................................................................................................. 25

1.2.3 Active sites in plant UGTs .................................................................................... 26

1.2.3.1 The catalytic amino acids .............................................................................. 26

1.2.3.2 Sugar donor pocket; PSPG motif .................................................................... 27

1.2.3.3 Interdomain linker ......................................................................................... 29

1.2.3.4 Other conserved C-terminal and less conserved N-terminal loops .................. 30

1.2.3.5 Acceptor pocket environment ........................................................................ 32

1.2.3.6 Inter and intradomain interactions .................................................................. 33

1.2.4 Substrate specificity studies .................................................................................. 34

1.3 Structural studies of UGTs using homology modeling .............................................. 35

1.4 Structure-based enzyme engineering of plant UGTs ................................................ 37

1.4.1 Mutagenesis: a tool for UGTs enzyme engineering ............................................... 40
### Chapter 2 Mutagenesis process

- 2.2.4.1 Primer design
- 2.2.4.2 Site-directed mutagenesis (SDM)
- 2.2.4.3 Mutant plasmid DNAs extraction via miniprep
- 2.2.4.4 DNA sequencing

References

### Chapter 3 Substrate studies of UGTs; In vitro biochemical characterisation

- 3.1 Mass spectrometry (MS)-based GT activity
- 3.1.1 Qualitative analysis
- 3.1.1.1 Reaction conditions
- 3.1.1.2 Conditions of MS
- 3.1.2 Quantitative analysis
- 3.1.2.1 Calibration curve
- 3.1.2.2 Kinetic study: mechanisms and conditions
- 3.2 Substrate specificity studies
- 3.2.1 Acceptor specificity
- 3.2.1.1 Group H UGTs
- 3.2.1.2 Group L UGTs
- 3.2.2 Donor specificity
- 3.3 Kinetics
- 3.3.1 Analysis of kinetic data
- 3.4 Conclusions

References

### Chapter 4 Homology modeling and docking studies of UGTs

- 4.1 Homology modeling of UGTs
- 4.1.1 Model construction
- 4.1.2 Secondary structure; model-template comparison
4.1.3 Model validation..................................................................................................................114
4.2 Docking studies and molecular interactions; Donor sugar interactions..................117
4.2.1 Key amino acids identification – Sugar moiety.................................................................117
  4.2.1.1 C2 and C3 positions in solved crystal structures............................................................117
  4.2.1.2 C2 and C3 positions in studied models............................................................................118
  4.2.1.3 C3 and C4 positions in solved crystal structures..............................................................119
  4.2.1.4 C3 and C4 positions in studied models............................................................................120
  4.2.1.5 C6 positions in solved crystal structures..........................................................................122
  4.2.1.6 C6 positions in studied models........................................................................................123
  4.2.2 Key amino acids identification – phosphate moiety........................................................124
  4.2.3 Key amino acids identification – uridine moiety..............................................................128
  4.2.4 Acceptor compounds.......................................................................................................131

Conclusion..................................................................................................................................138

References...................................................................................................................................139

Chapter 5 Enzyme engineering: Improving donor substrate recognition via mutagenesis
studies........................................................................................................................................140

5.1 Enzyme engineering...........................................................................................................142
  5.1.1 MSA-based mutation experiments.....................................................................................142

5.2 Structure-guided mutation experiments.............................................................................146
  5.2.1 Based on other plant UGT crystal structures.................................................................146
  5.2.2 Based on model structures of studied UGTs....................................................................150

5.3 Conclusion.............................................................................................................................160

Final thesis conclusions...........................................................................................................161

Future work................................................................................................................................162

References...................................................................................................................................163

Appendix....................................................................................................................................166

Publication.................................................................................................................................205
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fig 1.1</td>
<td>Types of GT-folds</td>
<td>20</td>
</tr>
<tr>
<td>Fig 1.2</td>
<td>Carbohydrate numbering convention</td>
<td>21</td>
</tr>
<tr>
<td>Fig 1.3</td>
<td>Domains in UGTs</td>
<td>21</td>
</tr>
<tr>
<td>Fig 1.4</td>
<td>Nomenclature of UGT superfamily</td>
<td>22</td>
</tr>
<tr>
<td>Fig 1.5</td>
<td>PSPG motif in nine UGTs crystal structures</td>
<td>24</td>
</tr>
<tr>
<td>Fig 1.6</td>
<td>Cleft region in between N-and C-terminal domains</td>
<td>26</td>
</tr>
<tr>
<td>Fig 1.7</td>
<td>Catalytic triads in selected plant UGTs</td>
<td>26</td>
</tr>
<tr>
<td>Fig 1.8</td>
<td>Catalytic residues (histidine and aspartic acid) in nine UGTs structure</td>
<td>27</td>
</tr>
<tr>
<td>Fig 1.9</td>
<td>All PSPG motif residues in nine UGTs structure</td>
<td>28</td>
</tr>
<tr>
<td>Fig 1.10</td>
<td>Three PSPG motif residues interacting with varying part of donor</td>
<td>28</td>
</tr>
<tr>
<td>Fig 1.11</td>
<td>Linker proximity to UDP-donor sugar</td>
<td>30</td>
</tr>
<tr>
<td>Fig 1.12</td>
<td>Loop regions in N- and C-domains</td>
<td>30</td>
</tr>
<tr>
<td>Fig 1.13</td>
<td>C1 loop residue interaction</td>
<td>31</td>
</tr>
<tr>
<td>Fig 1.14</td>
<td>Conservation of C1 loop region</td>
<td>32</td>
</tr>
<tr>
<td>Fig 1.15</td>
<td>Hydrophilic environment around acceptor substrates</td>
<td>33</td>
</tr>
<tr>
<td>Fig 1.16</td>
<td>Intra- and inter domain interactions</td>
<td>34</td>
</tr>
<tr>
<td>Fig 1.17</td>
<td>A. thaliana UGT genes grouped</td>
<td>36</td>
</tr>
<tr>
<td>Fig 1.18</td>
<td>LC-MS system</td>
<td>42</td>
</tr>
<tr>
<td>Fig 1.19</td>
<td>General thesis outline</td>
<td>44</td>
</tr>
<tr>
<td>Fig 2.1</td>
<td>SDS Protein Gels</td>
<td>59</td>
</tr>
<tr>
<td>Fig 2.2</td>
<td>Donor compounds library</td>
<td>61</td>
</tr>
<tr>
<td>Fig 2.3</td>
<td>Calibration curve creation</td>
<td>68</td>
</tr>
<tr>
<td>Fig 2.4</td>
<td>Amino acid sequences of studied UGTs in FASTA format</td>
<td>73</td>
</tr>
<tr>
<td>Fig 2.5</td>
<td>Primer design sample</td>
<td>74</td>
</tr>
<tr>
<td>Fig 2.6</td>
<td>Agarose gel of DNA plasmids</td>
<td>76</td>
</tr>
<tr>
<td>Fig 3.1</td>
<td>UGT enzyme activity test workflow</td>
<td>82</td>
</tr>
<tr>
<td>Fig 3.2</td>
<td>Mass spectrum showing full scan analysis</td>
<td>83</td>
</tr>
<tr>
<td>Fig 3.3</td>
<td>Mass spectrum showing product ion scan</td>
<td>84</td>
</tr>
<tr>
<td>Fig 3.4</td>
<td>Modes of MS data acquisition</td>
<td>85</td>
</tr>
<tr>
<td>Fig 3.5</td>
<td>Principle of UDP-Glo assay</td>
<td>87</td>
</tr>
<tr>
<td>Fig 3.6</td>
<td>Creation of calibration curve</td>
<td>88</td>
</tr>
</tbody>
</table>
Fig 3.7  UDP calibration curve ................................................................. 89
Fig 3.8  Summary of acceptor screening results ......................................... 93
Fig 3.9  Acceptor compounds showing positive activity ................................. 92
Fig 3.10 Major classes of plant secondary metabolites .................................. 94
Fig 3.11 Acceptor substrate recognition pattern in group H ............................ 95
Fig 3.12 Acceptor substrate recognition pattern in group L ............................ 97
Fig 3.13 Donor sugar library .................................................................. 97
Fig 3.14 Summary of donor screening results ............................................. 98
Fig 3.15 Distribution of nucleoside base recognition .................................... 99
Fig 3.16 Base diversity of donor sugars ..................................................... 100
Fig 3.17 Differences in structures of UDP-Glc, UDP-Gal and UDP-GlcNAc .... 101
Fig 4.1 UGT76E5 showing secondary structures ......................................... 113
Fig 4.2 Predicted secondary structures in all model UGTs ............................. 114
Fig 4.3 Superimposition of template and model structures ............................ 115
Fig 4.4 Ramachandran and Z-scores plots for all model UGTs ....................... 116
Fig 4.5 Component parts of UDP donor sugar ........................................... 117
Fig 4.6 C2/C3 and glutamine interactions .................................................. 119
Fig 4.7 Conserved aspartic/glutamic acid in model UGTs ............................ 120
Fig 4.8 C3/C4 and aspartic/glutamic acid interactions .................................. 121
Fig 4.9 C6 position interaction in crystal structures ..................................... 122
Fig 4.10 Conservation at C6 position ...................................................... 122
Fig 4.11 Residues at C6 position in model UGTs ......................................... 123
Fig 4.12 C6 and threonine interactions ..................................................... 124
Fig 4.13 Interacting residues with the phosphate moiety in crystal structures .... 125
Fig 4.14 Interacting residues with the phosphate moiety in model UGTs ........ 126
Fig 4.15 Interactions around phosphate moiety in model UGTs .................... 127
Fig 4.16 Conservation of residues around uridine in crystal structures ........... 129
Fig 4.17 C3/Cα3 region in model UGTs for stacking ..................................... 129
Fig 4.18 Interactions around uridine moiety in model UGTs .......................... 130
Fig 4.19 Conserved catalytic residues in model UGTs ................................ 131
Fig 4.20 N-terminal regions around acceptor substrate ................................ 132
Fig 4.21 N3 loop/Nα3 regions in model UGTs ............................................ 133
Fig 4.22  F77KLF81 of Na3 region in 76E1 ...................................................... 133
Fig 4.23  F77KLF81 in acceptor binding pocket in 76E1 ................................. 134
Fig 4.24  F77KLF81 substitutes in 76E2, 76D1, 76E4 and 76E5 ....................... 135
Fig 4.25  N4 loop region in model UGTs ..................................................... 135
Fig 4.26  76E1 Y114 interaction with acceptor compounds .......................... 136
Fig 4.27  76E2 Y115 interaction with acceptor compounds .......................... 137
Fig 4.28  C2 loop/Cα2 region in model UGTs .............................................. 142
Fig 5.1  Mutant UGT activity results (GAR screen) ..................................... 143
Fig 5.2  N320 and C2 loop/Cα2 region distance to donor sugar .................... 144
Fig 5.3  S320 in mutant 76E2 N320S distance to three donor sugars .............. 144
Fig 5.4  S318N and S311N show no direct interaction with UDP-Glc ............. 145
Fig 5.5  Side chain comparison in serine, alanine and asparagine ................. 146
Fig 5.6  Mutant UGT activity results ii (GAR screen) ................................ 147
Fig 5.7  Interactions of 76E1 Q373 and 76E5 Q366 with UDP-Glc .................. 147
Fig 5.8  Interactions of 76E1 Q373 and 76E5 Q366 with UDP-GlcNAc .......... 148
Fig 5.9  Interactions of 76E1 Q373 and 76E5 Q366 with UDP-Glc ................. 149
Fig 5.10 Differences in donor binding regions in model UGTs ...................... 151
Fig 5.11 Potential mutation sites in 76E4, 76D1 and 76E2 ........................... 152
Fig 5.12 Mutant UGT activity results iii (GAR screen) ................................ 153
Fig 5.13 C1 loop residues in crystal structures .......................................... 153
Fig 5.14 L275 in mutant 76E4 K275L with UDP-Glc .................................. 154
Fig 5.15 L275 in mutant 76E4 K275L with acceptor kaempferol .................. 155
Fig 5.16 T129 in mutant 76D1 P129T with UDP-Glc and UDP-Gal ............... 156
Fig 5.17 T129 interactions ........................................................................ 157
Fig 5.18 Side chain comparison in proline and threonine ............................ 157
Fig 5.19 Side chain comparison in aspartic acid and glutamic acid ............... 158
Fig 5.20 76E2 D374 and 76E2 E374 interactions with UDP-Gal ................. 158
Fig 5.21 C364 and C347 in mutant 76D1 G347C ......................................... 159
List of tables

Table 1.1  Plant UGT crystal structures currently on Protein data bank (PDB) ..................5
Table 1.2  Lengths of linker region ............................................................................27
Table 1.3:  Arrangement of the triple quadrupoles Q1q2Q3 ........................................... 42
Table 2.1  Chemicals and instruments ........................................................................48
Table 2.2  SDS-PAGE recipe ..................................................................................51
Table 2.3  Molecular weights of donor sugars .......................................................... 53
Table 2.4  Acceptor compounds library ....................................................................59
Table 2.5  Potential template UGTs for homology modeling .......................................62
Table 2.6  Components of PCR reaction ...................................................................66
Table 2.7  PCR cycling steps and conditions ...............................................................66
Table 2.8  Composition of KLD reaction ...................................................................67
Table 2.9  Primer design of mutants ....................................................................... 68
Table 3.1  Standard curve luminescence readings .....................................................79
Table 3.2  Kinetic study of UGTs .............................................................................91
Table 3.3  Michaelis Menten and Lineweaver-Burk plots for 76E1 .............................92
Table 3.4  Michaelis Menten and Lineweaver-Burk plots for 76E2 .............................93
Table 3.5  Michaelis Menten and Lineweaver-Burk plots for 76E5 (with UDP-Glc) ..94
Table 3.6  Michaelis Menten and Lineweaver-Burk plots for 76E5 (with UDP-Gal) ..95
Table 3.7  Michaelis Menten and Lineweaver-Burk plots for 76E5 (with UDP-GlcNAc) .96
Table 3.8  Michaelis Menten and Lineweaver-Burk plots for 76D1 ............................97
Table 4.1  Potential templates and identities to studied UGTs ...................................104
Table 4.2  Individual UGTs and identities to 74F2 ...................................................104
Table 4.3  Secondary structure comparison between template and models ...............105
Table 4.4  Residues in disallowed regions of Ramachandran plots ............................108
Table 4.5  Conserved glutamine in all crystal structures ..........................................111
Table 4.6  Conserved glutamine in model UGTs ......................................................111
Table 4.7  Residues interacting with phosphate moiety in crystal structures ..........118
Table 4.8  Residues interacting with phosphate moiety in model UGTs .................119
Table 4.9  Residues interacting with uridine moiety in crystal structures ................121
List of abbreviations

ATP: Adenosine triphosphate
DNA: deoxyribonucleic acid
DTT: Dithiothreitol
GAR: Green/Amber/Red
GDP: guanosine diphosphate
GDP-Fuc: GDP-Fucose
GDP-Glc: GDP-Glucose
GDP-Man: GDP-Mannose
GST: Glutathione S-transferase
GT: glycosyltransferase
HPLC: High performance liquid chromatography
IPTG: isopropyl-β-D-1-thiogalactopyranoside
Kcat: turnover number
KMP: Kaempferol
Km: Michaelis constant
MS: mass spectrometry
MSA: multiple sequence alignment
MW: molecular weight
PBS: Phosphate-buffered saline
PCR: Polymerase chain reaction
PDB: Protein data bank
PSPG: plant secondary product glycosyltransferase
QUE: Quercetin
QQQ: Triple quadrupole
SDM: site-directed mutagenesis
SDS-PAGE: sodium dodecyl sulfate polyacrylamide gel electrophoresis
dTDP: deoxythymidine diphosphate
TEMED: Tetramethylthlenediamine
UDP: uridine diphosphate
UDP-Glc: UDP-Glucose
UDP-Gal: UDP-Galactose
UDP-GlcNAc: UDP-\textit{N}-acetylglucosamine
UDP-Man: UDP-Mannose
UDP-Xyl: UDP-Xylose
UGT: UDP-dependent glycosyltransferase
WT: wild type
Chapter 1 Introduction
1.1 Glycosyltransferases (GTs)

Glycosyltransferases (GTs) are increasingly being studied because they control the biosynthesis of glycans (Varki, 2017). Biologically, glycans play vital but varying roles in the development, growth, functioning and survival of organisms producing them (Varki, 2017). GTs are a large family of structurally conserved enzymes responsible for catalysing the transfer of a sugar moiety from an activated donor sugar to an acceptor molecule (Osmani et al., 2009). The sugar transfer reaction, known as glycosylation, is quantitatively the most important reaction on earth as they account for the biosynthesis of the most of biomass (Kleene and Berger, 1993).

Different forms of classification have been performed on GTs; based on similarity of amino acids, mechanism of reaction, folding of 3D structures and substrate specificity (Tiwari et al., 2016). Based on the degree of amino acid (primary sequence) conservation, GTs have been grouped into families. Members can be found documented at the (Carbohydrate-Active enZymes) CAZy database (http://www.cazy.org/GlycosylTransferases). There are 108 families of GTs namely GT1 – GT108 (accessed September 2019). Another classification involves the mechanism of the catalytic reaction; described as either retaining or inverting (Coutinho et al., 2003). The mechanism type defines the stereochemistry of the activated donor sugar’s carbon atom involved in the newly formed glycosidic bond. This can either be retained (stereochemistry kept) or inverted (stereochemistry changed in an inversion). Although generally conserved within a CAZy family, prediction due to sequence comparison is not always definite (Breton et al., 2006).

Furthermore, GTs’ primary structure is poorly conserved which contrasts with their secondary and tertiary structure. Among members of the same CAZy family, fold structure is usually conserved as pointed out by sequence analysis as well as crystal structure information (Coutinho et al., 2003; Bourne and Henrissat, 2001). They all fold distinctively and can either be of GT-A, GT-B or GT-C fold (Breton et al., 2006). The GT-A fold is described as a single domain fold with seven stranded β-sheet bordered by α-helices and with a metal co-ordinating DXD motif. On the other hand, the GT-B fold contains of two separate N-terminal and C-terminal domains which are Rossman-like, of six/seven parallel β-sheet connected to α-helices. Both domains are joined by a linker region and an interdomain cleft (Lairson et al., 2008; Qasba et al., 2008). The GT-C fold is similar to the GT-A fold but with no DXD motif, showing a different type of α/β/α arrangement (Chiu et al., 2004). Examples showing these types of folds are shown in Fig 1.1.

More than 520,000 known and putative GT genes are currently known. (http://www.cazy.org/GlycosylTransferases).
Over the past decades, the significance of GTs has been studied. This include biosynthesis of glycoconjugates of phytochemicals, metabolic engineering of crops and also identification of prospective key targets in drug design (Bhat et al., 2013; Lim, 2005a; Williams et al., 2007). Insufficient biochemical information on individual members has however hindered further research for their functional understanding (Tiwari et al., 2016). GT1 family, comprising over 20,000 genes however contain the plant UDP-dependent glycosyltransferases (UGTs) (http://www.cazy.org/GT1_all; Osmani et al., 2009). The importance, activities, behaviour and potential applications of plant UGTs will be discussed in the next sections.

1.2 Plant UGTs

1.2.1 General features and importance

Glycosylation of plant secondary metabolites is catalysed by uridine diphosphate (UDP) glycosyltransferases (UGTs) belonging to family 1 glycosyltransferases (GTs) (Ross et al 2011). These plant secondary metabolites include alkaloids, terpenoids, flavonoids, isoflavonoids and other small molecules (Lim and Bowles, 2004). The biggest family of secondary metabolites glycosylated by plant UGTs are flavonoids (Tiwari et al., 2016) UDP activated sugars act as donors in the glycosylation reaction with UDP-glucose serving as the primary sugar donor. Others such as UDP-galactose, UDP-rhamnose, UDP-xylose and UDP-glucuronic acid follow respectively (Lim and Bowles, 2004). Typical numbering convention for UDP donor sugars is shown in Fig 1.2.
All available plant UGT crystal structures adopt the GTB fold, which comprises of a two-domain structure (N- and C-terminal) assuming α/β/α fold (Fig 1.3). The N-terminal domain being less conserved suggests involvement in binding of diverse acceptor compounds (Kumar et al., 2012). They are also inverting GTs. For the addition of sugar moiety (single or multiple), functional groups such as -OH, -NH₂, -COOH, -SH and C-C are favoured (Lim and Bowles, 2004).

Fig 1.3: Conserved two-domain structure of a representative of nine currently solved plant UGT crystal structures; The N- and C-terminal domains are (shown in green and orange respectively) with acceptor and donor substrates bound.
The P450 gene nomenclature designates numbering 1-50 for animal UGTs, 51-70 for yeast UGTs, 71-100 for plant UGTs and 101-200 for bacteria. Therefore, plant UGTs are allocated to UGT families 71 – 100. These plant UGTs are grouped together based on >40% amino acid sequence identity (Mackenzie et al., 1997). An example of nomenclature classification for a plant UGT is represented in Fig 1.4.

![Diagram illustrating current nomenclature of UGT superfamily specifically plant UGTs](image)

Fig 1.4: Diagram illustrating current nomenclature of UGT superfamily specifically plant UGTs

Nine plant UGT structures have currently been solved (Table 1.1). Although all the nine proteins share only 9% sequence identity, their secondary and tertiary structures are greatly conserved. A conserved motif found in the C-terminal domain in UGTs called the PSPG motif, is involved in the binding of the nucleotide sugar UDP moiety (Mackenzie et al., 1997). The PSPG (Plant Secondary Product Glycosyltransferases) box is regarded as “consensus sequence” and considered a characteristic structural feature of plant UGTs. The PSPG box contains 44 conserved amino acids (Hughes and Hughes., 1994). In silico analysis of the motif shows its conservation in the C-terminal of all UGT sequences (Kumar et al., 2012). The consensus sequence has also been applied in bioinformatics studies to identify plant glycosyltransferases from various databases (Tiwari et al., 2016). Some highly conserved amino acids within the PSPG box - HCGWNS - are vital for enzymatic function (Jadhav et al., 2012). The degree of conservation of the PSPG motif among the nine plant UGT crystal structures available as well as their aligned sequences are shown in Fig 1.5.
Several hundred thousand bioactive secondary metabolites are synthesised by plants, with a lot of them glycosylated at particular positions (Jones and Vogt., 2001). The stability, reactivity, solubility and bioactivity of these metabolites are modified by the addition of sugar moiety; conferring positive implications on their properties (Tiwari et al., 2016). In plants, the effects of glycosylation such as plant hormones and cell homeostasis regulation and xenobiotics detoxification have been studied (Jones and Vogt., 2001; Lim and Bowles., 2004).

<table>
<thead>
<tr>
<th>Plant UGT</th>
<th>Protein Databank Code</th>
<th>In complex with</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>AtUGT74F2</td>
<td>5V2K 5U6M 5U6N 5U6S 5V2J</td>
<td>(T15A) with UDP and 2-bromobenzoic acid UDP and salicylic acid (T15S) with UDP and salicylic acid UDP and 2-bromobenzoic acid (T15S) with UDP and 2-bromobenzoic acid</td>
<td>George-Thompson et al. (2017)</td>
</tr>
<tr>
<td>AtUGT72B1</td>
<td>2VCE 2VCH 2VG8</td>
<td>UDP and Tris buffer UDP-2-fluoro-glucose and TCP</td>
<td>Brazier-Hicks et al. (2007)</td>
</tr>
<tr>
<td>MtUGT78G1</td>
<td>3HBF 3HBJ</td>
<td>UDP and myricetin UDP</td>
<td>Modolo et al. (2007)</td>
</tr>
<tr>
<td>MtUGT71G1</td>
<td>2ACV 2ACW</td>
<td>UDP glucose</td>
<td>Shao et al. (2005)</td>
</tr>
<tr>
<td>MtUGT85H2</td>
<td>2PQ6</td>
<td>None</td>
<td>Li et al. (2007)</td>
</tr>
<tr>
<td>CjUGT78K6</td>
<td>3WC4 4REL 4REM 4REN 4WHM</td>
<td>None Kaempferol Delphinidin Petunidin UDP</td>
<td>Hiromoto et al. (2013), (2015)</td>
</tr>
<tr>
<td>VvGT1</td>
<td>2C1X 2C1Z 2C9Z</td>
<td>UDP UDP-2-fluoro-glucose and kaempferol UDP and quercetin</td>
<td>Offen et al. (2006)</td>
</tr>
<tr>
<td>Os79</td>
<td>5TMB 5TD 5TME 6BK0 6BK1 6BK2</td>
<td>UDP U2F and trichothecene UDP Q202A with UDP T291V with UDP H122A/L123A with UDP</td>
<td>Wetterhorn et al. (2016), (2017)</td>
</tr>
<tr>
<td>PtigS</td>
<td>5NLM</td>
<td>Magnesium ion and 3-sulfooxy-1H-indole</td>
<td>Hsu et al. (2018)</td>
</tr>
<tr>
<td><strong>AtUGT89C1</strong></td>
<td>6IJ7 6IJ9 6IJA 6IJD</td>
<td>None UDP UDP-rhamnose Quercetin</td>
<td>Zong et al., (2019)</td>
</tr>
</tbody>
</table>

Table 1.1: Nine solved plant UGT crystal structures. [AtUGT89C1 (in red) is the most recently added, making the 10th]

Many glycosides (sugar-containing metabolites) possess pharmaceutically interesting properties in humans, which is important in drug development (Thorson et al., 2001). Glycosylation of these biologically active compounds at specific positions regulate their pharmaceutical activity as well as optimising their potential use as drug compounds (Kren., 2001; Mijatovic et al., 2007; Weymouth-Wilson., 1997; Bonina et al., 2003; Egleton et al., 2000).
A very popular example are cardiac glycosides which have been in clinical use for many years in the treatment of heart failure and atrial arrhythmia. They are natural glycosides with sugars such as glucose, galactose, rhamnose, and mannose commonly attached. The sugars themselves have been found to have no activity however their addition influences the potency of the aglycone. For instance, the addition of rhamnose increased the potency several times (6-35 times), highlighting the importance of glycosylation of natural products (Prassas and Diamandis, 2008).

Other glycosides such as quercetin 3-glucoside, quercetin 3-galactoside and rhamnoside have shown antioxidant, anti-inflammatory and anti-viral activities respectively (Olthof et al., 2000; Kim et al., 2011; Choi et al., 2009). Also, stability is increased for vitamin C (Yamamoto et al., 1990) and anthocyanins (Zhao et al., 2014a) following glycosylation. In site-specific drug delivery, the opportunity to target galactosylated compounds to the liver makes an exciting prospect (Wong and Toth, 2001). Furthermore, numerous other examples of small molecule glycosylation especially their potentials and applications have been reported. These include glucosides of steviol, steviosides as sweeteners (Geuns, 2003), naringenin-3-O-xylloside possessing anti-tumour properties (Simkhada et al., 2009), kaempferol-3-O-glycoside is neuroprotective (Yu et al., 2013) esculin, an esculetin glucoside is anti-apoptotic (Zhao et al., 2007) while artemisinic acid (from precursor artemisinin with antimalarial activity) shows good anti-tumour activity (Zhu et al., 2014). The removal of sugar residues from saponins (terpene glycosides of antifungal properties) often results in the loss of their bioactivity (Osbourn, 2003).
Additionally, bioactive glycosides content of fruits and vegetables is reflected in their healthy beneficial effects on human health (Harbourne and William., 2000). The bitterness of grapefruit or tastelessness of mandarin is determined by the addition of rhamnose on the 2-OH or 6-OH of the glucose moiety of naringenin 7-O-glucoside (Frydman et al., 2004). Therefore, to design improved UGTs of desired pharmaceutical properties, a detailed knowledge of the activity as well as specificity of UGTs is pertinent (Bowles et al., 2005; Butelli et al., 2008; Thorson et al., 2004).

For most UGTs, in vivo and in vitro functions are still unknown (Zhang et al., 2018). Consistent with this intricacy, 122 UGT encoding genes have been currently identified in the model plant Arabidopsis thaliana genome (http://www.p450.kvl.dk) while more than 165 UGTs are available in the legume Medicago truncatula (Modolo et al., 2007). Of the 122 UGT genes identified in Arabidopsis thaliana, more than half have not been biochemically characterised yet (Zhang et al., 2018). Therefore, research on more plant UGTs gives a potential approach to aiding glycan synthesis (Tiwari et al., 2016).

1.2.2 Crystal structures

The crystal structures of nine plant UGTs have been solved and deposited in the Protein Data Bank (PDB). They are: Arabidopsis thaliana AtUGT74F2 (George-Thompson et al., 2017), AtUGT72B1 (Brazier-Hicks et al., 2007), Medicago truncatula MtUGT71G1 (Shao et al., 2005), MtUGT85H2 (Li et al., 2007), MtUGT78G1 (Modolo et al., 2009), Vitis vinifera VvGT1 (Offen et al., 2006), Oryza sativa Os79 (Wetterhorn et al., 2016), Clitoria ternatea UGT78K6 [Ct3GT-A] (Hiromoto et al., 2013) and Polygonum tinctorium PtigS (Welner et al., 2017). Details of these UGTs have been shown in Table 1.1. A new plant crystal structure whose detailed structural information is yet to be fully available on the PDB has recently been solved. AtUGT89C1 becomes the first plant rhamnosyltransferase crystal structure solved; others are mostly in complex with UDP or UDP-glucose (Zong et al., 2019).

All the crystal structures belong to the GT-B fold, which consists of two Rossmann-like domains; the C- and N-terminals (Lairson et al., 2008). A cleft located in between both domains is used for substrates binding (Fig 1.6). The N-terminal is more variable and can accommodate diverse acceptors while the more conserved C-terminal binds nucleotide sugar donors (Breton et al., 2006). Plant UGTs C-terminal domains are especially similar (Osmani et al., 2009). There is usually a linker region connecting both domains called the interdomain linker whose flexibility suggests domain movements important for substrate binding. Many regions are reported to diverge in the N-terminal of plant UGT structures (loops N3, N5, N5a and the interdomain linker - shown in Fig 1.10). These and other several regions within the primary sequence control substrate specificity in UGTs binding (Osmani et al., 2009).
1.2.3 Active sites in plant UGTs

1.2.3.1 The catalytic amino acids

Several plant UGT structures previously reported had unveiled two catalytic residues (histidine and aspartic acid) which both form a triad with the acceptor compound, as key to glycosylation activity (Shao et al. 2005; Offen et al., 2006; Brazier-Hicks et al., 2007; Li et al., 2007; Modolo et al., 2009; Hiromoto et al., 2013; Wetterhorn et al., 2016; George-Thompson et al., 2017 and Welner et al., 2017). Examples of catalytic triad formed in 72B1 and VvGT1 are indicated in Fig 1.7.

Fig 1.6: The cleft (in yellow) for substrate binding found in between the N- and C- terminal domains. The cleft represented is in the crystal structure of plant UGT74F2 (PDB: 5V2K).

Fig 1.7: Examples of catalytic triad in plant UGTs [histidine-aspartic acid-acceptor substrate] a) AaUGT72B1: H19-D117-Trichlorophenol acceptor triad b) VvGT1: H20-D119-Kaempferol acceptor triad
The acceptor functional group (−OH in these instances), which is open to glycosylation must be positioned near the histidine residue as well as the sugar donor. Histidine acts as the base deprotonating the acceptor compound. (Shao et al. 2005; Offen et al., 2006; Brazier-Hicks et al., 2007; Li et al., 2007). The precise positioning of the −OH group dictates whether a glycosylation reaction can occur (Osmani et al., 2009). The conservation of both amino acids in these crystal structures are shown in Fig 1.8. The catalytic triad is conserved in all plant UGT crystal structures except for the very recently published AtUGT89C1 crystal structure, which indicated a slightly different mechanism. The aspartic acid was replaced in the UGT with a serine; hence no interaction was established with histidine like in the others. Hence, a catalytic dyad (acceptor-Histidine) was utilised instead of a triad including aspartic acid (Zong et al., 2019).

Fig 1.8: Multiple sequence alignments showing positions of catalytic amino acids histidine-aspartic acid which are key to glycosylation reaction in all nine plant UGT structures. Both amino acids form catalytic triad with the acceptor substrate.

1.2.3.2 Sugar donor pocket; PSPG motif

As described in Fig 1.4, plant UGTs are defined by a 44-amino acid conserved sequence called the PSPG motif (Mackenzie et al., 1997). Being part of the C-terminal, PSPG residues offer bulk of the sugar donor interactions. Ten of these 44 amino acids directly interact with the donor while others are thought to stabilize and/or enable molecular interactions (Li et al., 2007). These are highlighted in red as shown in Fig 1.9.
Of these ten residues, three namely tryptophan, aspartic/glutamic acid and glutamine [22nd, 43rd and 44th] usually make hydrogen bonds with the varying part (sugar part) of the donor. The other seven residues then interact with the invariant part. Aspartic/glutamic acid, glutamine and some other unspecified residues have been determined to control sugar donor specificity in plant UGTs (Osmani et al., 2009). The positions of these three amino acids in the solved plant UGTs crystal structures are very similar as shown in Fig 1.10.

![Conserved Residues](image)

**Aspartic/Glutamic acid**

**Glutamine**

**Tryptophan**

**UDP-sugar**

**Fig 1.10:** Overlay of conserved residues (Tryptophan, Aspartic acid/Glutamic acid and Glutamine) from the nine plant crystal structures which interact with the sugar part of the UDP-donor sugar. The location of the three amino acids in these plant UGT crystals are sited almost exactly. Colour code: UDP-Glucose in CPK colours.
1.2.3.3 Interdomain linker

Despite the independent folding of the N- and C-terminals, interdomain interactions between both domains has been observed (Li et al., 2007). The linker region typically runs from the end of N-terminal to start of C-terminal. It is composed of amino acids from both the N- and C-terminal varying in length and sequence. This difference in length and sequence has been proposed to help substrate accommodation and positioning. In the nine crystallized plant UGTs, differing lengths (shown in Table 1.2) are observed with Os79 having the longest and MtUGT71G1 with the shortest. In addition, this region is often distorted in some plant UGT crystal structures suggesting flexibility (Osmani et al., 2009).

<table>
<thead>
<tr>
<th>UGT</th>
<th>Length of linker region</th>
</tr>
</thead>
<tbody>
<tr>
<td>AtUGT74F2</td>
<td>29AAs</td>
</tr>
<tr>
<td>AtUGT72B1</td>
<td>17AAs</td>
</tr>
<tr>
<td>VvGT1</td>
<td>17AAs</td>
</tr>
<tr>
<td>MtUGT71G1</td>
<td>15AAs</td>
</tr>
<tr>
<td>MtUGT78G1</td>
<td>17AAs</td>
</tr>
<tr>
<td>MtUGT85H2</td>
<td>28AAs</td>
</tr>
<tr>
<td>Os79</td>
<td>31AAs</td>
</tr>
<tr>
<td>CtUGT78K6</td>
<td>17AAs</td>
</tr>
<tr>
<td>PtigS</td>
<td>17AAs</td>
</tr>
</tbody>
</table>

Table 1.2: The lengths of linker region in the nine plant UGTs structures

The C-terminal part of the linker is however usually positioned around the uridine part of the UDP-donor sugar in many plant UGTs. Although not sufficiently near to make any direct interactions, the linker residues may be involved in some intramolecular interactions (Osmani et al., 2009). For example, in 71G1, tryptophan W339, which binds to UDP-donor sugar uridine is positioned near the linker loop. Fig 1.11 shows the position of tryptophan to the donor sugar, and the linker region surrounding it (Shao et al., 2005). Upon donor sugar binding, the linker loop goes through a conformational change. This is consistent with the suggestion of the region’s ability to adapt to conformational changes upon sugar binding (Li et al., 2007).
Fig 1.1: 3D structure of 71G1 showing the relative position of conserved W339, alongside interdomain linker region’s proximity to the UDP-sugar donor. Colour code: Linker loop in grey, UDP and W339 in CPK colours.

1.2.3.4 Other conserved C-terminal and less conserved N-terminal loops
Several loop regions in plant UGT structures (besides the PSPG motif residues) offer critical interactions with either or both substrates. Loops from both N- and C-terminal regions in 74F2 are shown in Fig 1.12.

Fig 1.12: Loop regions of the N- and C-terminal domains in 74F2. These loop regions generally offer vital interactions with both acceptor and donor substrates in plant UGTs.
An example of such loop region is the C-terminal C1 loop region whose residues interact with the donor sugar. C1 loop region (and residue T280) in VvGT1 interacts with the sugar donor (Fig 1.13).

Fig 1.13: C1 loop region (coloured in blue) residue T280 bound to UDP-Glucose in VvGT1 indicating interaction of the loop residues with the donor sugar

Across all nine plant UGT crystal structures, the length of this loop is highly conserved except UGT71G1 with an extra residue (Fig 1.14). The loop also contains a conserved glycine next to a serine/threonine, which directly binds to the β-phosphate part of sugar (Hiromoto et al., 2015; Offen et al., 2006; Shao et al., 2005; Brazier et al., 2007). The C1 loop region residues in plant UGTs are shown in Fig 1.14. Increased activity with more sugar donors observed in mutation of residues in this loop, indicates the general importance of this C-terminal loop (He et al., 2006).

The N-terminal domain is generally less conserved to accommodate several acceptor compounds. However, some loop regions like loops N1 and N5 in this domain offer vital interactions for catalysis and donor recognition respectively. In all nine plant UGT structures, the N1 loop (sometimes extending to Nα1) contains the very conserved histidine residue, the acceptor-deprotonating base crucial for glycosylation activity (Offen et al., 2006; Shao et al., 2005; Brazier et al., 2007; Li et al., 2007; Hiromoto et al., 2013; George-Thompson et al., 2017; Welner et al., 2017; Wetterhorn et al., 2016; Modolo et al., 2009).
Other N1 loop residues besides this catalytic histidine is not conserved (Osmani et al., 2009). Likewise, in N5 loop conserved serine/threonine bound to the donor sugar in VvGT1 (T141). Similar binding was predicted in 71G1 and 85H2 (Offen et al., 2006; Li et al., 2007; Shao et al., 2005).

![Multiple sequence alignment showing the conserved C1 loop residues of nine plant UGTs crystals. Serine/threonine (in green print and red box) which directly interacts with the β-phosphate part of donor sugar](image)

Fig 1.14: Multiple sequence alignment showing the conserved C1 loop residues of nine plant UGTs crystals. Serine/threonine (in green print and red box) which directly interacts with the β-phosphate part of donor sugar

### 1.2.3.5 Acceptor pocket environment

Hydrophobic environment was found to typically surround the acceptor pockets of plant UGTs structures. These hydrophobic interactions provide stability vital for enzymatic activity. In MtUGT78G1, aromatic (especially phenylalanine F) and other hydrophobic amino acids form the acceptor-binding pocket (Modolo et al., 2009).

In CtUGT78K6-delphinidin complex, the flavylum ring of the anthocyanin is surrounded by hydrophobic side chains from proline, leucine and phenylalanine (Hiromoto et al., 2013). This is illustrated in Fig 1.15. Similarly, in MtUGT71G1, triterpene acceptors are surrounded by residues such as proline, phenylalanine, leucine and isoleucine, which provide a hydrophobic environment (Shao et al., 2005). In the new rhamnosyltransferase 89C1 structure, the acceptor compound quercetin is immersed in similar hydrophobic cavity containing phenylalanines, alanine, leucine and isoleucine (Zong et al., 2019). The general fit of the acceptor into its binding pocket along with stability offered by hydrophobic interactions is crucial for activity (Osmani et al., 2009).
33

Fig 1.1: Hydrophobic environment (provided by proline, leucine, phenylalanine, isoleucine, tryptophan) around acceptor compound delphinidin in UGT78K6. The hydrophobic interactions provide stability vital for enzymatic activity.

1.2.3.6 Inter and intradomain interactions

Hydrogen bonds, disulphide bridges and salt bridges are examples of inter- and intradomain interactions, which support protein secondary and tertiary structure. These interactions are significant for specificity and activity (Osmani et al., 2009). Sometimes, intramolecular interactions comprising residues far from substrates and not in any direct contact affect acceptor specificity. This was studied with two C-terminal loop (C2) residues in AtUGT72B1, which were observed to control N-glucosylation, although too far to directly interact with the acceptor compound. Y315-P15 interacts to place catalytic H19 in a disadvantaged position for O-glucosylation, thereby favouring N-glucosylation. Y315 alongside N312 are vital for N-glucosylation without directly binding to the substrates (Brazier-Hicks et al., 2007).

Between the C- and N-terminal domains of 85H2, a salt bridge is formed between K192 and E413. This interdomain interaction was explained to provide improved stability (Li et al., 2007). An intradomain salt bridge is present in 78G1 between R89 and E192 of the acceptor pocket, played a role in the entry of acceptor myricetin (Modolo et al., 2009). Fig 1.16 shows the salt bridges in both UGTs described. Disulphide bridge such as that found between C379 and C396 of the PSPG motif in 85H2 increases the stability of the region (Li et al., 2007). All these inter- and intradomain interactions typically contribute to the activity and specificity of plant UGTs.
1.2.4 Substrate specificity studies

Substrate specificity and activity studies are complex, particularly the mechanism controlling them. In plant UGTs, substrate specificity studies involve recognition of the UDP-sugar donor and the acceptor compound (Osmani et al., 2009). UGTs are highly specific with respect to donor sugars. UDP-dependent sugars such as UDP-glucose, UDP-galactose, UDP-rhamnose, UDP-xylose have been previously used as sugar donors by UGTs (Lim et al., 2004). However, few plant UGTs such as AtUGT89C1 can accept UDP-, dTDP- and GDP sugars (Parajuli et al., 2016) although most plant UGTs will typically use UDP-glucose (Thorsoe et al., 2005). This was displayed in donor specificity screening of some Arabidopsis thaliana group H UGTs where all showed activity with UDP-glucose (Ake re et al., 2018). Plant UGTs also commonly show preference for a specific sugar donor even if a broad sugar donor activity is displayed. For example, Vitis vinifera VvGT1 demonstrates a great preference for UDP-glucose despite activity with myriad of other sugars like UDP-galactose, UDP-glcNAc, UDP-mannose, GDP-glucose, dTDP-glucose, dTDP-xylose and UDP-xylose (Offen et al., 2006).

On the contrary, plant UGTs can recognise very diverse acceptor compounds from plant secondary metabolites such as flavonoids, terpenoids, alkaloids, glucosinolates, cyanogenic glucosides, isoflavonoids and other phenylpropanoids (Lim and Bowles, 2004). Different varieties of glycosides are found in plants connoting their ability to glycosylate diverse compounds in vivo (Jones and Vogt., 2001).
In vitro, considerable differences are found in the range of acceptors of individual UGTs. Some are highly specific while others of broad specificity (Osmani et al., 2009). Attempts have been made to study the relationship between substrate specificity and high amino acid sequence identity.

Phylogenetic grouping of Arabidopsis thaliana UGTs has been done based on high similarity of sequence into fourteen groups designated A-N (Li et al., 2001; Ross et al., 2001). The groups comprise all 122 genes (and their families) of Arabidopsis thaliana as shown in Fig 1.17. Within these groups, substrate specificity has been studied with several acceptors (of varying classes of plant secondary metabolites). Highly identical A. thaliana UGTs have been found to identify structurally diverse acceptor compounds (Caputi et al., 2008; Lim et al., 2004; Lim et al., 2002). Similarly, AtUGTs 76E1 and 76E2 has a high degree of sequence similarity (88%) but UGT76E1 can utilize many more acceptor and donor substrates than UGT76E2 (Akere et al., 2018). The summary of findings has indicated that within the same group, A. thaliana UGTs recognised and glycosylated very different acceptor substrates while UGTs from different groups were observed to share same substrates. Substrate specificity studies in other plant UGTs have shown similar findings. For example, highly divergent Dorotheanthus bellidiformis UGTs, betanidin 5-O-glucosyltransferase and betanidin 6-O-glucosyltransferase glycosylated the same substrates. Both share only 19% sequence similarity (Vogt, 2002). Allium cepa UGTs 73G1 and 71J1 show broad and narrow acceptor specificity respectively while of the same family (Kramer et al., 2003).

All these studies have further confirmed the complexity surrounding the correlation between the degree of sequence identity and substrate specificity. The understanding of factors dictating substrate preferences within UGT groups is predicted to improve with more biochemical data on individual UGTs (Lim et al., 2003a; Vogt and Jones., 2000). Additionally, due to the wide range nature of potential substrates that needs to be screened, biochemical characterisation of the substrate specificity can also be demanding. A more targeted and faster means would be therefore through substrate interactions from 3D structures (Osmani et al., 2009). Since only a limited number of solved X-ray crystal structures for plant UGTs are available, homology modeling using solved crystal structures as templates suggests an alternative method in understanding substrate specificity (Breton et al., 2006; Imberty et al., 2006; Kopp and Schwede., 2004).

1.3. Structural studies of UGTs using homology modeling

Homology modeling is a computational method, which predicts protein 3D models using known experimental structures to provide insight into structural and functional analysis.
This offers a cost-effective option when there is a lack of experimental structures (Cavasotto and Phatak., 2009). As an attractive alternative, homology modeling deeply relies on secondary and tertiary structure conservation between query and template proteins (Tiwari et al., 2016).

Fig 1.17: Phylogenetic tree showing 122 Arabidopsis thaliana UGT genes showing grouped into A-N (circled) created with Geneious software. Eight UGTs from groups H and L screened later in this report are printed in blue and red respectively. Groups H and L UGTs constitute the largest of the 14 groups accounting for 16% and 14% respectively.
In structure-based studies, homology modeling has effectively helped in hit identification, guided ligand optimisation and assisted mutagenesis experiments notwithstanding challenges associated. Such problems include identifying right templates, refining loops and carrying out accurate sequence alignments (Cavasotto and Phatak., 2009). Generally, models of between 25-50% sequence identities are deemed suitable for designing mutagenesis experiments (Hillisch et al., 2004). Since mutagenesis is vital for recognising amino acids with important biological function, structural information obtained from homology modeling aids in the identification of residues for such studies (Cavasotto and Phatak., 2009).

Homology modeling has been used to study plant UGTs. UGT73A5 from *Dorotheanthus bellidiformis* and UGT85B1 from *Sorghum bicolor* both were the first homology-based model structures for plant GTs built prior to the solving of crystal structure of any plant UGT. Both models (made using bacterial GT templates) highlighted the predictive power of homology models, identifying the important residues (Hans et al., 2004; Thorsoe et al., 2005). Site-directed mutagenesis confirmed the importance of many of the predicted residues. Loop modelling was however difficult, especially as the sequence similarity was as low as 14-15% (Hans et al., 2004; Thorsoe et al., 2005).

Subsequently, the first homology model of a plant UGT made based on crystal structure coordinates of MtUGT71G1 and WGT1 was UGT94B1 from *Bellis perennis*. Residues crucial to substrate specificity and activity of enzyme were predicted and mutational studies confirmed their importance. R25 was found to be specific for UDP-glucuronic acid sugar donor (Osmani et al., 2008). In order to gain better insight into UDP-sugar specificity mechanism, there is a need for crystal structures of UGTs complexed with substrates other than UDP-glucose (Osmani et al., 2009). Jadhav et al., (2012) has reported 3D modeling and ligand docking studies for WsFGT, a flavonoid glycosyltransferase from *Withania somnifera*. MtUGT85H2 was used as a template with 31% identity. The binding pockets of various flavonoid acceptor compounds as well as the PSPG motifs’ interactions with UDP-glucose were highlighted.

1.4 Structure-based enzyme engineering of plant UGTs

Recently, extensive development has been made in structural studies of plant UGTs. This is to advance understanding of function, substrate specificity and inhibition/induction properties of UGTs (Fujiwara et al., 2016). This has also been progressing the mechanism of glycosylation and directing UGT engineering to aid the production of bioactive glycosides. In plant UGTs, their C-terminal domains are similar leading to their recognising same or similar donors.
Few other regions in the N-terminal have been shown to also participate in donor sugar recognition. Manipulation around these regions influences activity and varies donor sugar specificity in UGTs (Wang et al., 2009).

Enzyme engineering via structure-guided mutation in UGTs (particularly aiming at their active sites) is a method to manipulate enzyme activities and specificities generating newly enhanced mutant UGTs. Besides the benefit of synthesizing bioactives, structure-based enzyme engineering also helps plant metabolic engineering towards improving the production and quality of crop plants (Wang 2011). There are two basic methods to engineer enzymes, via rational design and directed evolution. Rational design involves the introduction of mutations at specific position(s) in a protein based on the knowledge of possible relationships such as sequence, structure, function as well as catalytic mechanism. Directed evolution approach includes recurrent series of random mutagenesis of and/or recombination with the gene variants to produce a gene library of slightly different sequences. Each method has their strengths and limitations (Damian-Almazo and Saab-Rincon, 2013). In selecting amino acids for mutation, any structural, biochemical or protein sequence information may be valuable. A typical approach is to concentrate on the catalytic region. Amino acid residues that modify substrate specificity are usually non-conserved residues often near catalytic residues or active site. Another way is to identify conserved sequence motifs (Morley & Kazlauskas, 2005; Paramesvaran et al., 2009; Park et al., 2005; Saravanan et al., 2008).

Utilizing the rational design approach, several studies have reported UGTs manipulation using either crystal structures and/or homology models to engineer mutants of improved sugar substrate recognition. Lamiales plant flavonoid 7-O-glucuronosyltransferases (F7GAT) are a group of UGTs which use donor sugar UDP-glucuronic acid (UDPGA). Homology modeling of a F7GAT, PfUGT88D7 was made while UDPGA and apigenin aglycone were docked. Interactions via docking pointed to R350 and S127 to be important in UDPGA binding. Single mutant UGTs PfUGT88D7 R350W and S127T and also a double mutant S127T R350W were made. R350W showed reduced glucuronosyl but gained glucosyl activity. S127T effect on activity was minor, but still a decreased glucuronosyl and increased glucosyl activity. Interestingly, the double mutant S127T R350W displayed a more remarkable decrease in glucuronosyl and more enhanced glucosyl activity. Similar mutations in F7GAT such as SiUGT88D6 and AmUGT88D4 resulted in a shift in donor sugar specificity from UDP-glucuronic acid to UDP-Glc (Noguchi et al., 2009).

Another UDP-glucuronosyltransferase BpUGT94B1 used molecular modeling to recognise an N-terminal residue R25 to be crucial and specific for UDP-glucuronic acid activity.
Mutants R25S/G/K all displayed very weak UDP-glucuronic acid activity and improved UDP-Glc activity. The rotation of donor sugar specificity was exhibited via enzyme engineering (Osmani et al., 2008). However, attempts were made to engineer mutants with UDP-glucuronosyl activity.

At position mutated in PfUGT88D7 (i.e. R and S), mutations were not successful as T139S and W367R in PfUGT88A7, T150S and W371R in GmUGT88E3 and P26R in SbUGT85B1 did not gain UDP-glucuronosyl activity. This explained that possibly several other amino acids are essential for donor sugar recognition albeit a single amino acid may perform a key role (Noguchi et al., 2009; Osmani et al., 2008). Similar manipulations have also been done around the acceptor binding pocket regions, affecting UGT enzyme activity and altering substrate specificity. While UGT71G1 glucosylated flavonoid genistein at the 7-OH position, its mutant Y202A was able to add sugar at an additional position 5-OH; giving two products 7-O-glucoside and 5-O-glucoside (He et al., 2006).

These mutations, all guided by structures resulted in the development of new UGTs with enhanced activity. These new UGTs may be exploited for enzymatic synthesis of glycosides (Lim et al., 2005). Other enzyme engineering strategies include domain swapping, which involves sequence alignments and molecular modelling, generating UGT chimeras of enhanced or altered glycosylation activity (Hansen et al., 2009). In addition, reversibility (glycosylation and deglycosylation) has been investigated via engineering works on some UGTs. The findings may be applied in designing effective mutants with reversibility to synthesize activated donor sugars (Modolo et al., 2009; Modolo et al., 2009a; Wang, 2009). Furthermore, the design and generation of specific enzyme mutants with novel activity and specificity and their use as biocatalysts in glycosylation may produce non-natural aglycone derivatives; which are attractive targets in drug synthesis and discovery (Wang et al., 2011).

Multiple sequence alignment (MSA)-based mutations (based on knowledge of possible sequence relationship) and/or structure-based mutations (based on structure relationship) constitute the rational approach. Evolution of proteins involves changes such as single residue mutations, insertions and deletions which can overtime make sequence similarities identification cumbersome (Pascarerella and Argos, 1992). Protein structure is, nevertheless, more conserved than protein sequence hence offers more precise information on potential mutation sites. With structural alignments, the identification of residues in the active sites and/or in direct contact with substrate can be easily observed, unlike with an MSA-based only mutation. Subsequently, site-directed mutagenesis offers a powerful means in the design study of novel enzymes including their functions (Damian-Almazo and Saab-Rincon, 2013).
With methods such as identifying amino acids involved in binding as well as conserved amino acids, modification of enzyme properties for biotechnological and industrial applications (such as in UGTs) have become possible.

1.4.1 Mutagenesis: a tool for UGTs enzyme engineering

Apart from the synthesis of beneficial glycosides for example cardiac glycosides, GTs contribute immensely to plant defence mechanism, regulation of hormone, modification of xenobiotics and detoxification of pollutants, secondary metabolite biosynthesis, stabilization of secondary metabolites and metabolic engineering of crops (Tiwari et al., 2016).

However, the pharmacological application is our focus as many phytochemicals such as flavonoids, alkaloids, hormones, antibiotics, and sweeteners occur as glycosides (Blanchard and Thorson, 2006). A key condition for the pharmacological usage of drugs and its activity is the existence of a carbohydrate moiety (Kren and Martinkova, 2001). The addition of sugar to bioactive molecules boosts solubility and bioavailability and such sugar addition at specific positions modify bioactives’ pharmaceutical properties (Thorson et al., 2001; Kren and Martinkova, 2001; Mijatovic et al., 2007).

The usage of plant UGTs for small molecule glycosylation is therefore a better option than organic chemistry in the biosynthesis of the needed glycoconjugates due to the challenges encountered with the latter (Tiwari et al., 2016). This usage offers good prospects in rational drug designing in the pharmaceutical industry although limited nucleotide activated sugars are a challenge (Luzhetskyy et al., 2007). Active sites mutation studies would confer novel function to the present UGT enzymes. UGTs of broad activity are acceptable candidates for manipulation while those of strict specificity hamper glycoengineering. Several novel “chimeras” have been created through such mutation studies (Tiwari et al., 2016).

The structure of VvGT5 was modelled using the crystal structure of VvGT1 with substrates (kaempferol and UDP-glucuronic acid) docked. At the same position, W140 in VvGT1 faces to donor sugar Uridine-5’-Diphosphate-2-Deoxy-2-Fluoro-Glucose. The prediction of R140 interaction with UDP-glucuronic acid from the model suggested its potential function. Site-directed mutagenesis of the mutation R140W validated the roles of each amino acid with the mutant losing its glucuronic activity and exhibiting a distinct VvGT1 glucose transfer (Ono et al., 2010).
1.4.2 How mutagenesis takes place

Protein engineering involves the design of new proteins with novel or desirable functions usually based on the use of recombinant DNA technology to change amino acid sequences. Several methods of protein engineering are available today; the most classical method being rational design approach which comprises site-directed mutagenesis of proteins (Arnold, 1993). Site-directed mutagenesis (SDM) permits introduction of specific amino acids into a target DNA gene. Two common methods for SDM exists: the whole plasmid single round PCR and the overlap extension method. The overlap extension method involves two primers pairs with one pair containing the mutant codon with a mismatched sequence. Two PCRs take place and two double-stranded DNA products are made. However, the whole plasmid single round PCR forms the basis of many commercial kits such as Q5 SDM kit used in this work. Two oligonucleotide primers with the desired mutation (also complementary to the opposite strands of a double-stranded DNA plasmid template) are essential. PCR takes place using DNA polymerase with both DNA strand replicated. A mutated plasmid is obtained and selective digestion and transformation into competent cells occur thereafter (Antikainen and Martin, 2005). Sequencing studies will confirm the incorporation of right mutation(s). Expression of mutant proteins from the mutant DNA plasmid usually follow which would be used in further studies such as UGT assay.

1.5 MS-based GT characterisation assays

Activities of both wild-type and mutant UGTs (whether enhanced or not) produced were characterised via GT assays. To effectively test GT substrate specificity, the principal technique used is monitoring product formation over time. A wide variety of assays have been made to examine substrate specificity and even kinetics in GT-catalysed reactions (Johnson, 2013). These assays are typically label or label free methods (Zhang, 2018). Label GT assays include fluorescence-based method, radiochemical assays, spectrophotometric assays and so on (Schmid et al., 2016). However, label-free assays such as the use of mass spectrometry (MS) have shown high effect on GT screening and characterisation (Yang et al., 2005). The relatively easy sample preparation as well as direct usage has made label-free GT assays attractive (Zhang, 2018).

In this research, we have combined the usage of both label (UDP-Glo assay based on bioluminescence) and label-free assay (MS) for monitoring product formation (substrate specificity) respectively. MS was used only for qualitative studies (which was to check whether glycosylation occurred or not). Specifically, Triple Quadrupole LC/MS system was used; a combination of the separation power of LC and the detection power of MS.
LC generally separates components in a mixture through the chromatographic column which are then moved to the MS (the ion source). Fig 1.18 analyses the LC-MS system as used in this work.

Fig 1.18: LC-MS system

The sample is transformed into ions via electrospray ionisation (ESI). After ionisation, transfer of ions into mass analyser take place where separation of ions based on their mass to charge (m/z) ratio are done (Nikalje et al., 2018). Quadrupole is the most commonly used mass analyser, positioned between ion source and a detector, consists of four rods with direct current (DC) and radio frequency (RF) voltages applied. Here, the triple quadrupole MS has three quadrupoles (Q1,q2Q3) with the poles’ arrangement accounting for the different analysis modes (such as full scan, product ion scan, multiple reaction monitoring scan, precursor ion scan). The arrangement of Q1q2Q3 in the two modes used in this work is shown in Table 1.3. These details were given by Edmond (1996).

<table>
<thead>
<tr>
<th></th>
<th>Q1</th>
<th>q2</th>
<th>Q3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full scan</td>
<td>Scan</td>
<td>CID off</td>
<td>Scan</td>
</tr>
<tr>
<td>Product ion scan</td>
<td>Fixed</td>
<td>CID on</td>
<td>Scan</td>
</tr>
</tbody>
</table>

Table 1.3: Arrangement of the triple quadrupoles Q1q2Q3 which dictates analysis types in MS system

Subsequently, ion signals quantified in the detector are displayed as a spectrum of ion abundance against m/z on y and x axis respectively (Hisatomi et al., 2012; Edmond, 1996). In our GT substrate specificity studies, the initial screening which examines whether a potential glycosylated product was formed, uses the ‘full scan’ mode. Upon detection of the potential product (with suitable MS modification) in either positive [M+H]+ or negative-ion [M+H]+ modes, its fragmentation by verifying aglycone ion formation via ‘product ion’ mode follows. This would confirm the usage of substrate by a particular enzyme. CID (collision-induced dissociation) takes place in q2 and may result in 1 or 2 Da mass shifts known as heterolytic or homolytic cleavage of the glycosidic bond (Yamagaki and Watanabe, 2012).
For kinetics, the UDP-Glo assay based on bioluminescence is a homogenous, single reagent addition method which detect UDP formation in GT reactions. UDP product is converted to light in a luciferase reaction which is measured. The light output is proportional to GT activity which is detected using a luminometer. The main advantages of the assay are being highly sensitive, easy to use, a rapid method, reliable and reproducible data obtained and stability for more than 3 hours. However, only GT reactions involving UDP sugars can be used in this assay as it is unable to used GDP or TDP sugars. Also, it may be used with only purified GTs and not whole cells or cell extract. Biochemical values such as Km of UDP sugars and acceptor substrates are provided which helps in the GT mode of action studies (www.promega.com; Zegzouti et al., 2013).

1.7 Aims and objectives

The aims are firstly, to investigate UGT biochemical information via studying their:

i) substrate preferences and

ii) important amino acids driving the choices

Secondly, is to advance potential application of UGTs in the production of bioactive glycosides through:

i) structural studies via homology modeling

ii) rational manipulations of the active sites.

1.8 General scheme of thesis

In this project, twenty-two recombinant AtUGTs from group L (84A2, 84A3, 84A1, 84A4, 74E2, 75B1, 74C1, 75D1, 74F2 and 75C1), group H (76E2, 76E1, 76D1, 76E5, 76E4, 76E3 and 76F2), group D (73D1, 73B1, 73C2 and 73C3) and group B (89A2) were studied (Fig 1.17). Groups H, L and D Arabidopsis thaliana UGTs constitute the largest of the 14 groups (Fig 1.17) accounting for 16%, 14% and 11% respectively.

This thesis is outlined as shown in Fig 1.19. This introductory chapter (Chapter 1) has given background knowledge of what GTs are and more importantly, the significance, substrate specificity and structural studies of plant UGTs. Thereafter, the aims of this project as well as the outline of the thesis are elucidated. This introduction is followed by chapter 2, which explains the materials and methods used in this research.
Chapter 2 highlights materials used in this project including chemicals and instruments. In addition, the experimental methods used ranging from protein production (plasmid transformation, expression and purification), site-directed mutagenesis, MS-based enzyme activity, and homology modelling are detailed. Following this closely is the chapter on in vitro biochemical characterisation of UGTs.

Chapter 3 expatiates our findings on the biochemical characterisation of selected Arabidopsis thaliana UGTs; their substrate (donor and acceptor) specificities and kinetics using liquid chromatography-mass spectrometry. Following initial screening, fifteen of the twenty-two UGTs were expressed successfully. Nine of these showed positive glycosylation activity and so proceeded for substrate specificity screening.

A 42-member acceptor library comprising mainly of plant secondary metabolites coupled with a seven-member donor library was utilised. After screening, some of these UGTs were selected for further structural studies reported in the following chapter 4.

Homology modelling of five selected UGTs is reported in Chapter 4. Models were validated, substrates docked into them and substrates-ligand interactions explained. Structural differences among UGTs, which may affect their glycosylation behaviour pattern, are presented. In addition, key amino acids involved in substrate specificities of the UGTs are explained. Enzyme engineering trailed this chapter.

The final chapter 5 describes all mutations carried out in the research; both structure-guided and sequence-guided. Mutant UGTs were engineered and the findings on improved substrate specificities obtained were highlighted. Future works, which could advance findings, are discussed.
References


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*. *Scientific Reports* 7, 46629 (2017).
Hughes, J., & Hughes, M.A. Multiple secondary plant product UDP-glucose glucosyltransferase genes expressed in cassava (Manihot esculenta Crantz) cotyledons. DNA Sequence 5, 41-49 (1994).


Chapter 2 Materials and methods
## 2.1 Materials

### 2.1.1 List of Chemicals and Instruments

<table>
<thead>
<tr>
<th>Chemicals and Instruments</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agar, Luria-Bertani (LB) medium, acrylamide, glycerol, sodium dodecyl sulphonate (SDS), ampicillin, Tris, tetramethylethylenediamine (TEMED), liquid chromatography/mass spectrometry (LC-MS) solvents [acetonitrile and water], MgCl₂, apomorphine, caffeic acid, benzyl cinnamate, quinine, dioxime, naringenin, pelargonidin, catechin, arbutin, umbelliferone, eriodictyol, chrysins, hesperetin, formononetin, cyanidin, dihydromyricetin</td>
<td>Sigma Aldrich (United Kingdom)</td>
</tr>
<tr>
<td>6400-series triple quadrupole LC-MS, QIAGen miniprep kit, XL-1 blue competent cells,</td>
<td>Agilent</td>
</tr>
<tr>
<td>Resveratrol</td>
<td>StressMarq Biosciences</td>
</tr>
<tr>
<td>GSTrap fastflow column, Vivaspin concentrators (6 ml)</td>
<td>GE healthcare</td>
</tr>
<tr>
<td>Novobiocin, Ferulic acid</td>
<td>ChemCruz</td>
</tr>
<tr>
<td>Kinetex 5µm C18 column</td>
<td>Phenomenex</td>
</tr>
<tr>
<td>Vivaspin Concentrators (500µl)</td>
<td>Generon</td>
</tr>
<tr>
<td>Q5 SDM kit, 1kb DNA ladder, DNA gel loading dye</td>
<td>NEBiolabs</td>
</tr>
<tr>
<td>SDS electrophoresis Mini-Protein Tetra Cell systems, Agarose electrophoresis Mini-Sub Cell GT cell, Mini-Protein precast gels, protein standard, agarose, Quickstart Bradford 1X dye reagent</td>
<td>Bio-Rad</td>
</tr>
<tr>
<td>Cell Shaker, formic acid</td>
<td>Thermo fisher Scientific</td>
</tr>
<tr>
<td>Pherastar microplate reader</td>
<td>BMG Labtech</td>
</tr>
<tr>
<td>Autoclave, cell incubator, sonicator</td>
<td>Beckman</td>
</tr>
<tr>
<td>Clear 96-well plates, white 384-well plate</td>
<td>Greiner</td>
</tr>
<tr>
<td>UDP-Glc, UDP-Gal, UDP-GlcNAc, UDP-Glo assay kit</td>
<td>Promega</td>
</tr>
<tr>
<td>InstantBlue gel stain</td>
<td>Expedeon</td>
</tr>
<tr>
<td>2,5-dihydrobenzoic acid, 3,4-dihydrobenzoic acid, glutathione, scopoletin</td>
<td>Acros Organics</td>
</tr>
<tr>
<td>Ulttrapure distilled water, SYBR safe DNA gel stain</td>
<td>Invitrogen</td>
</tr>
<tr>
<td>TBE buffer</td>
<td>Applichem</td>
</tr>
<tr>
<td>Salicin, apigenin, galantamine</td>
<td>LKT Lab</td>
</tr>
<tr>
<td>Quercetin, fisetin,isorhamnetin, myricetin, delphinidin, aloemodin, galangin, daidzein, biochanin A</td>
<td>Cayman Chemicals</td>
</tr>
<tr>
<td>Sinapic acid, esculetin, morin, genistein</td>
<td>SantaCruz Biotech</td>
</tr>
</tbody>
</table>

Table 2.1: Table containing chemicals and instruments used in this research.
2.2 Methods

2.2.1 Protein production

Twenty-two recombinant UGTs (UGT 76E1, 76E2, 76E4, 76E5, 76D1, 76E3, 76F2, 84A1, 84A2, 84A3, 84A4, 75D1, 75C1, 75B1, 73C3, 73B1, 73D1, 73C2, 74C1, 74E2, 74F2 and 89A2) from Arabidopsis thaliana groups H, L, D and B were produced in this work. Also, thirteen (13) mutant UGTs (76E4 K275L, 76D1 G347C, 76D1 P129T, 76E2 D374E, 76E4 Q375A, 76E4 E374A, 76E5 Q366A, 76E1 Q373A, 76E1 N320S, 76E5 S311N, 76E1 S318N, 76E5 S311A and 76E1 S318A) were made using the below procedures.

2.2.1.1 Cloning

Total RNA was extracted from the 14d seedlings of the Arabidopsis using the TRIzol reagent (Takara, Japan). First-strand cDNA was synthesized using PrimerScript RT reagent Kit with gDNA Eraser (Takara, Japan), pooled cDNA was used as PCR template. The cDNA sequences of UGT76E1, UGT76E2, UGT76E5 and UGT76D1 were amplified by TransStart Fast Pfu DNA Polymerase (TRANS, China). A BamHI and XhoI restriction site were included in the sequence of forward primer and the reverse primer used for PCR of UGT76E1 and UGT76E2. The PCR primers for UGT76E1 cloning are: forward GGATCCATGGAAGAACTAGGAGTGAAG and reverse CTCGAGCTACATGAAATTCATCATT; the PCR primers for UGT76E2 cloning are: forward GGATCCATGGAGGAAAACAAGTGAAG, reverse CTCGAGTCACATGGAATTAACAAAGTC. The BamHI and EcoRI were included in the sequence of forward primer and the reverse primer used for PCR of UGT76E5. The PCR primers for UGT76E5 cloning are: forward GGATCCATGGAGAAAAATGCAGAGAAG, reverse GAATTCTCAAGTATTTCTATACTCTGC. The BamHI and SacI were included in the sequence of forward primer and the reverse primer used for PCR of UGT76D1. The primers for UGT76D1 cloning are: forward CGGGATCCATGGCAGAGATTCGCCAG, reverse CGAGCTCTCATTTGTCGATTTGATC. The amplified products were cloned into the corresponding sites of the Blunt simple cloning vector, previously digested with appropriate restriction enzymes. After sequencing and alignment with reference sequence of the four genes shown in the Arabidopsis information resource, the right clones were chosen for further cloning. Lastly, the cDNA fragments of the target genes were inserted and ligated into the pGEX-3H plasmid previously digested with appropriate restriction enzymes, respectively. The resulting recombinant plasmids were transformed into competent E. coli BL cells (Akere et al., 2018).

2.2.1.2 Plasmid transformation

1 µL of plasmid encoding UGTs (recombinant or mutant) was mixed with 20 µL XL-1 blue competent cells. The mix was placed on ice for 30min and heat shocked for 30s at 42ºC.
This was incubated for another 5 minutes on ice, mixed with 950 μL LB media and thereafter incubated at 37°C (180rpm) in a shaker for 1 hour. 50 μL plasmid/cell mixture was then spread on an agar plate (containing 50 μg/mL ampicillin) and incubated at 37°C overnight. The next morning, a single colony was picked from the agar plate and grown overnight in LB medium (containing 50 μg/mL ampicillin), at 37°C (180rpm). This was then stored in 80% glycerol at −80°C.

2.2.1.3 Protein expression

Seed cultures were prepared by inoculating UGTs glycerol stocks (500 μL) into LB medium (containing 50 μg/mL ampicillin) and incubating overnight in a shaker at 37°C [180 rpm]. This was transferred into 1L LB medium and grown until 0.4-0.6 OD600 under same conditions. The protein production process was induced using isopropyl-β-D-1-thiogalactopyranoside (IPTG) at 0.1 mM final concentration after the system was cooled (to enhance proper protein folding) for 1 hour to 20°C. The induction took place for 16 hours thereafter (Akere et al., 2018).

2.2.1.4 Protein purification

Harvesting of cells followed via centrifugation at 5,000 rpm (4°C) for 20 minutes. Cell pellets were collected and re-suspended in 10 mL buffer (Tris 20mM, Ph7.8). Sonication to break the cells followed (on ice) at 12,000 x g for 20 minutes. The resulting supernatant was then collected for purification. Using the GSTrap™ FF column, the protein purification process was as follows:

- Elution buffer (50mM Tris; 10mM reduced glutathione; pH 8.0) and binding buffer (PBS, pH 7.3) were prepared.
- Supernatant was passed through a filter (45 μm) to prevent clogging of the column with precipitates.
- Five column volumes of binding buffer were used to equilibrate column (one column volume = 1 mL) through a connected syringe to the column.
- The filtered supernatant was then passed through the column at flow rate 0.2-1 mL/minute. Our GST tagged UGTs would bind to the column resin at this stage.
- The column was washed with binding buffer (5-10 mL; flow rate 1-2 mL/min) to prevent contaminants binding and remaining on the column.
- Elution buffer (five column volume 5 mL) was applied to the column (flow rate 1-2 mL/min). The column-bound target proteins were eluted and collected (the first 3 mL).
Lastly the column was washed in 2mL washing buffer (6M guanidine hydrochloride) and 5mL PBS to remove other hydrophobic bound substances. And thereafter stored in 20% ethanol.

### 2.2.1.5 Protein Storage

The storage buffer (1mM Tris, 1mM MgCl₂ and 10% glycerol pH 7.6) was prepared. The purified protein was then transferred into the storage buffer via Vivaspin tube (30,000 molecular weight). The storage buffer provides improved environment for the protein. In the vivaspin tube, the protein was centrifuged at 10,000 rpm for 10 minutes. The concentrated protein (between 20 mg/mL to 50 mg/mL) was then stored at 4 degrees prior to use.

### 2.2.1.6 Protein verification

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was done to verify the purity of protein obtained. Gel preparation (separating and stacking) was carried out with contents as shown in Table 2.2.

<table>
<thead>
<tr>
<th>Separating gel (10%)</th>
<th>Stacking gel (6%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ddH₂O</td>
<td>ddH₂O</td>
</tr>
<tr>
<td>3.8 mL</td>
<td>2.9 mL</td>
</tr>
<tr>
<td>40% acrylamide</td>
<td>40% acrylamide</td>
</tr>
<tr>
<td>2.0 mL</td>
<td>0.75 mL</td>
</tr>
<tr>
<td>1.5M Tris pH 8.8</td>
<td>1.5M Tris pH 8.8</td>
</tr>
<tr>
<td>2.0 mL</td>
<td>1.25 mL</td>
</tr>
<tr>
<td>10% SDS</td>
<td>10% SDS</td>
</tr>
<tr>
<td>80 μL</td>
<td>50 μL</td>
</tr>
<tr>
<td>10% APS</td>
<td>10% APS</td>
</tr>
<tr>
<td>80 μL</td>
<td>50 μL</td>
</tr>
<tr>
<td>TEMED</td>
<td>TEMED</td>
</tr>
<tr>
<td>8 μL</td>
<td>5 μL</td>
</tr>
<tr>
<td>Total volume</td>
<td>Total volume</td>
</tr>
<tr>
<td>8.0 mL</td>
<td>5.0 mL</td>
</tr>
</tbody>
</table>

Table 2.2: SDS-PAGE gel recipe

After gel preparation, the gel was cast and protein samples prepared to be run. 5 μL protein, 5 μL protein loading buffer and DTT (1M, 2 μL) were mixed and heated for 10 minutes (at 95°C) to denature protein. In separate wells of the prepared gel, each protein sample was then loaded. The electrophoresis process was run at 40 volts (first 20 minutes of run) to allow samples reach separating gel. For the rest of the run, the voltage was increased to 150 volts. At the completion of the run, the gel was transferred to glass plates and stained in instant blue solution. After 1-hour staining, the gel was rinsed in ddH₂O for another 1 hour. The gel image (proteins) was obtained using a camera. Examples of two verified UGTs after purification are shown in Fig 2.0. Other purified UGT images are in appendix figures li and lii.
2.2.1.7 Protein concentration determination

Quick Start Bradford protein assay by BIO-RAD was used to quantify protein concentration. It is an easy and accurate procedure involving the binding of Coomassie Brilliant blue dye to proteins which is detected at 595 nm using a spectrophotometer. The 1x dye reagent is removed from 4°C storage to warm to ambient temperature. In two 1 ml disposable cuvettes, pipette each standard (Quick start Bradford dye reagent) and unknown sample solution (20 μl) into them. Then the dye reagent is added to each cuvette. This is mixed well and incubated at room temperature for about 5 minutes. The spectrophotometer is set to 595nm and zeroed with the blank dye. The absorbance of the unknown protein solution is then measured. This is done in triplicates.
2.2.2 Mass-spectrometry based enzyme activity screening

2.2.2.1 Substrate specificity studies

The UGT enzyme assay included the following components: 1 mM Tris, 1 mM MgCl$_2$ (pH 8.0), 10 mM UDP-sugars, 10 mM acceptor compounds and purified target proteins.

For the acceptor screening, the 42 compounds screened are shown in the acceptor library (Table 2.3). The seven donor sugar compounds used in the donor screening are also indicated in Fig 2.1. The reaction mixture was incubated at 37° C for 3 and 15 hours at different times. The reaction was terminated with acetonitrile and centrifuged to remove proteins. The supernatant was then analysed with LCMS. Glycosylated products were identified by their molecular weights, and these target compounds were subsequently fragmented using MS/MS for confirmation.

- **HPLC-MS/MS**

Samples were analysed with Agilent 6400 triple quadrupole mass spectrometer coupled with HPLC system using phenomenex-C18 column (50 x 4.6 mm, kinetex 5u, 100A). The solvents used were as follows: (A) HPLC grade water containing 0.1% formic acid and (B) acetonitrile containing 0.1% formic acid. For the MS scan, both positive and negative spectra were obtained and run at a flow rate of 0.4 ml min$^{-1}$ in isocratic mode (10% A and 90% B). The injection volume was 10 ul, detection wavelength at 260 nm and column temperature of 20°C. Other details include start and end mass of 100 and 1000 respectively, scan time of 500 and cell accelerator voltage of 7. The MS/MS (product ion scan) had a flow rate 0.5 ml min$^{-1}$ at gradient mode (70% A: 30% B for 1 min, 55% A: 45% B for 1.50 min and lastly 70% A: 30% B for 2.50 min). The injection volume was maintained at 10 ul and only negative spectra were obtained here. The precursor ions were fragmented for confirmation within the range of m/z 100–1000. Scan segment details are as follows: scan time 500 fragmentor 135, collision energy 15 and cell accelerator voltage 7. All analyses were done in duplicates.

The glycosylated product formation was monitored in the reactions in the negative ion mode. This was made possible using the ‘full scan mode’ in initial screenings to observe if a potential glycosylated product is formed. The detection of a glycosylated product led to the application of the ‘product ion mode’ which fragments the potential glycosylated product. This fragmentation will confirm the enzymes activity if an aglycone was produced. The summary of each enzyme screening results is shown as a GAR (Green-Amber-Red) panel, where green means positive activity, amber means ambiguous activity and red means no activity.
• **Donor screening**

Donor substrate screening of nine wildtype UGTs and 15 mutant UGTs were determined. The donor library comprises of seven sugar donors: UDP-Glucose (UDP-Glc), UDP-Galactose (UDP-Gal), UDP-N-acetylg glucosamine (UDP-GlcNAc), GDP-Fucose (GDP-Fuc), GDP-Mannose (GDP-Man), GDP-Glucose (GDP-Glc) and UDP-Mannose (UDP-Man) for WT UGTs. For mutant UGTs, the donor library has three donor sugars which are UDP-Glucose (UDP-Glc), UDP-Galactose (UDP-Gal) and UDP-N-acetylg glucosamine (UDP-GlcNAc) (shown in fig 2.1 and table 2.3).

![Donor substrates](image_url)

**Fig 2.1:** Donor screening. (A). UDPGlc. (B). UDPGal. (C). UDPGlcNAc (D). GDPFuc. (E). GDPMan. (F). GDPGlc. (G). UDPMan.

<table>
<thead>
<tr>
<th>Donor</th>
<th>UDP-Glc</th>
<th>UDP-Gal</th>
<th>UDP-GlcNAc</th>
<th>GDP-Fuc</th>
<th>GDP-Man</th>
<th>GDP-Glc</th>
<th>UDP-Man</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Molecular weight (Da)</strong></td>
<td>448</td>
<td>448</td>
<td>489</td>
<td>589</td>
<td>605</td>
<td>605</td>
<td>566</td>
</tr>
</tbody>
</table>

Table 2.3: Molecular weight of different sugar donors

Two flavonoid acceptor compounds (kaempferol and quercetin) were used in the donor screening. In a buffer (50 µL, 1 mM Tris, 1 mM MgCl₂, pH 7.6), the acceptor compounds [kaempferol or quercetin] (10 µL, 100 µM), the donor compound (10 µL, 100 µM) and each UGT (10 µL, 0.01-0.1 mg/mL) were added. The sample mix were incubated overnight at 37°C.

• **Acceptor screening**

The acceptor substrates specificities of 22 WT UGTs were examined using UDP-Glc, which is the UGTs general donor compound as donor. The acceptor library contained 42 acceptors listed in table 2.4. In a buffer (50 µL, 1 mM Tris, 1 mM MgCl₂, pH 7.6), the acceptor compound (10 µL, 100 µM), the donor compound (10 µL, 100 µM) and each UGT (10 µL, 0.01-0.1 mg/mL) were added. The sample mix were incubated overnight at 37°C.
<table>
<thead>
<tr>
<th>Chemical name</th>
<th>Structure</th>
<th>Formula</th>
<th>MW (Dalton)</th>
<th>MW + Glc (Dalton)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Umbelliferone</td>
<td><img src="https://example.com/umbelliferone.png" alt="Umbelliferone" /></td>
<td>C_{9}H_{6}O_{3}</td>
<td>162.14</td>
<td>324.14</td>
</tr>
<tr>
<td>2. Sinapic acid</td>
<td><img src="https://example.com/sinapicacid.png" alt="Sinapic acid" /></td>
<td>C_{11}H_{12}O_{5}</td>
<td>224.21</td>
<td>386.21</td>
</tr>
<tr>
<td>3. 3,4-DHBA (dihydroxybenzoic acid)</td>
<td><img src="https://example.com/34-DHBA.png" alt="3,4-DHBA" /></td>
<td>C_{7}H_{6}O_{4}</td>
<td>154.12</td>
<td>316.12</td>
</tr>
<tr>
<td>4. 2,5-DHBA (dihydroxybenzoic acid)</td>
<td><img src="https://example.com/25-DHBA.png" alt="2,5-DHBA" /></td>
<td>C_{7}H_{6}O_{4}</td>
<td>154.12</td>
<td>316.12</td>
</tr>
<tr>
<td>5. Abscisic acid</td>
<td><img src="https://example.com/abscisicacid.png" alt="Abscisic acid" /></td>
<td>C_{15}H_{20}O_{4}</td>
<td>264.32</td>
<td>426.32</td>
</tr>
<tr>
<td>6. Zeatin</td>
<td><img src="https://example.com/zeatin.png" alt="Zeatin" /></td>
<td>C_{10}H_{13}N_{5}O</td>
<td>219.25</td>
<td>381.25</td>
</tr>
<tr>
<td>7. Quercetin</td>
<td><img src="https://example.com/quercetin.png" alt="Quercetin" /></td>
<td>C_{15}H_{10}O_{7}</td>
<td>302.24</td>
<td>464.24</td>
</tr>
<tr>
<td>8. Fisetin</td>
<td><img src="https://example.com/fisetin.png" alt="Fisetin" /></td>
<td>C_{15}H_{10}O_{6}</td>
<td>286.24</td>
<td>448.24</td>
</tr>
<tr>
<td>9. Kaempferol</td>
<td><img src="https://example.com/kaempferol.png" alt="Kaempferol" /></td>
<td>C_{15}H_{10}O_{6}</td>
<td>286.24</td>
<td>448.23</td>
</tr>
<tr>
<td></td>
<td>Compound</td>
<td>Molecular Structure</td>
<td>Formula</td>
<td>MW</td>
</tr>
<tr>
<td>---</td>
<td>--------------</td>
<td>---------------------</td>
<td>--------</td>
<td>-----</td>
</tr>
<tr>
<td>10</td>
<td>Scopoletin</td>
<td><img src="image" alt="Scopoletin structure" /></td>
<td>$C_{10}H_{10}O_4$</td>
<td>192.17</td>
</tr>
<tr>
<td>11</td>
<td>Esculetin</td>
<td><img src="image" alt="Esculetin structure" /></td>
<td>$C_9H_8O_4$</td>
<td>178.14</td>
</tr>
<tr>
<td>12</td>
<td>Novobiocin</td>
<td><img src="image" alt="Novobiocin structure" /></td>
<td>$C_{31}H_{36}N_2O_11$</td>
<td>612.62</td>
</tr>
<tr>
<td>13</td>
<td>Catechin</td>
<td><img src="image" alt="Catechin structure" /></td>
<td>$C_{15}H_{14}O_6$</td>
<td>290.27</td>
</tr>
<tr>
<td>14</td>
<td>Genistein</td>
<td><img src="image" alt="Genistein structure" /></td>
<td>$C_{15}H_{10}O_5$</td>
<td>270.24</td>
</tr>
<tr>
<td>15</td>
<td>Cyanidin</td>
<td><img src="image" alt="Cyanidin structure" /></td>
<td>$C_{15}H_{11}O_6^+$</td>
<td>287.24</td>
</tr>
<tr>
<td>16</td>
<td>Arbutin</td>
<td><img src="image" alt="Arbutin structure" /></td>
<td>$C_{12}H_{16}O_7$</td>
<td>272.25</td>
</tr>
<tr>
<td>17</td>
<td>Salicin</td>
<td><img src="image" alt="Salicin structure" /></td>
<td>$C_{13}H_{18}O_7$</td>
<td>286.28</td>
</tr>
<tr>
<td></td>
<td>Name</td>
<td>Chemical Structure</td>
<td>Chemical Formula</td>
<td>MW</td>
</tr>
<tr>
<td>---</td>
<td>---------------</td>
<td>---------------------</td>
<td>------------------</td>
<td>------</td>
</tr>
<tr>
<td>18</td>
<td>Aloe-Emodin</td>
<td><img src="image" alt="Aloe-Emodin" /></td>
<td>C_{15}H_{10}O_{5}</td>
<td>270.24</td>
</tr>
<tr>
<td>19</td>
<td>Pelargonidin</td>
<td><img src="image" alt="Pelargonidin" /></td>
<td>C_{15}H_{11}O_{5}^+</td>
<td>271.24</td>
</tr>
<tr>
<td>20</td>
<td>Formononetin</td>
<td><img src="image" alt="Formononetin" /></td>
<td>C_{16}H_{12}O_{4}</td>
<td>268.26</td>
</tr>
<tr>
<td>21</td>
<td>Ferulic Acid</td>
<td><img src="image" alt="Ferulic Acid" /></td>
<td>C_{10}H_{10}O_{4}</td>
<td>194.18</td>
</tr>
<tr>
<td>22</td>
<td>Apomorphine</td>
<td><img src="image" alt="Apomorphine" /></td>
<td>C_{17}H_{17}NO_{2}</td>
<td>267.32</td>
</tr>
<tr>
<td>23</td>
<td>Daidzein</td>
<td><img src="image" alt="Daidzein" /></td>
<td>C_{15}H_{10}O_{4}</td>
<td>254.23</td>
</tr>
<tr>
<td>24</td>
<td>Resveratrol</td>
<td><img src="image" alt="Resveratrol" /></td>
<td>C_{14}H_{12}O_{3}</td>
<td>228.25</td>
</tr>
<tr>
<td></td>
<td>Name</td>
<td>Molecular Formula</td>
<td>Molecular Weight</td>
<td>Charge</td>
</tr>
<tr>
<td>---</td>
<td>--------------</td>
<td>-------------------</td>
<td>------------------</td>
<td>--------</td>
</tr>
<tr>
<td>25</td>
<td>Apigenin</td>
<td>C_{15}H_{10}O_{5}</td>
<td>270.24</td>
<td>432.24</td>
</tr>
<tr>
<td>26</td>
<td>Isorhamnetin</td>
<td>C_{16}H_{12}O_{7}</td>
<td>316.26</td>
<td>478.26</td>
</tr>
<tr>
<td>27</td>
<td>Hesperetin</td>
<td>C_{16}H_{14}O_{6}</td>
<td>302.28</td>
<td>464.28</td>
</tr>
<tr>
<td>28</td>
<td>Naringenin</td>
<td>C_{15}H_{12}O_{5}</td>
<td>272.26</td>
<td>434.26</td>
</tr>
<tr>
<td>29</td>
<td>Delphinidin</td>
<td>C_{15}H_{11}O_{7}</td>
<td>303.24</td>
<td>465.24</td>
</tr>
<tr>
<td>30</td>
<td>Galangin</td>
<td>C_{15}H_{10}O_{5}</td>
<td>270.24</td>
<td>432.24</td>
</tr>
<tr>
<td></td>
<td>31</td>
<td>Morin</td>
<td><img src="image1.png" alt="Image" /></td>
<td>C_{15}H_{10}O_{7}</td>
</tr>
<tr>
<td>---</td>
<td>--------</td>
<td>----------------</td>
<td>----------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>Biochanin A</td>
<td><img src="image2.png" alt="Image" /></td>
<td>C_{16}H_{12}O_{5}</td>
</tr>
<tr>
<td></td>
<td>33</td>
<td>Eriodictyol</td>
<td><img src="image3.png" alt="Image" /></td>
<td>C_{15}H_{12}O_{6}</td>
</tr>
<tr>
<td></td>
<td>34</td>
<td>Diosmetin</td>
<td><img src="image4.png" alt="Image" /></td>
<td>C_{16}H_{12}O_{6}</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>Chrysin</td>
<td><img src="image5.png" alt="Image" /></td>
<td>C_{15}H_{10}O_{4}</td>
</tr>
<tr>
<td></td>
<td>36</td>
<td>Dihydromyricetin</td>
<td><img src="image6.png" alt="Image" /></td>
<td>C_{15}H_{12}O_{8}</td>
</tr>
<tr>
<td></td>
<td>37</td>
<td>Myricetin</td>
<td><img src="image7.png" alt="Image" /></td>
<td>C_{15}H_{12}O_{8}</td>
</tr>
<tr>
<td>38</td>
<td>Caffeic acid</td>
<td>C_9H_8O_4</td>
<td>180.16</td>
<td>342.16</td>
</tr>
<tr>
<td>----</td>
<td>-------------</td>
<td>-----------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>39</td>
<td>Galantamine</td>
<td>C_17H_21NO_3</td>
<td>287.36</td>
<td>449.36</td>
</tr>
<tr>
<td>40</td>
<td>Rutin</td>
<td>C_27H_30O_16</td>
<td>610.52</td>
<td>773.52</td>
</tr>
<tr>
<td>41</td>
<td>Morphine</td>
<td>C_17H_19NO_3</td>
<td>285.34</td>
<td>447.34</td>
</tr>
<tr>
<td>42</td>
<td>Quinine</td>
<td>C_20H_24N_2O_2</td>
<td>324.42</td>
<td>486.42</td>
</tr>
</tbody>
</table>

Table 2.4: Library of acceptor compounds screened in this work. The structures, molecular weights and products weights are also shown.
2.2.2.2 Pseudo-kinetics (1st-Order kinetics)

Kinetic study was performed using the UDP-Glo™ glycosyltransferases assay kit - following the procedure provided by the supplier. The UGT enzyme assay mixture for kinetics contained Tris buffer (50mM Tris, 5 mM MgCl2, pH 7.5), kaempferol (for 76E1 and 76E5) and quercetin (for 76E2 and 76D1). Kaempferol and quercetin were fixed at 100 µM. The donor substrates concentration; UDP Glc (five concentrations) was varied from 10–500 µM while UDP Gal and GlcNAc were varied from 0.1–10 µM. The UGT reaction was carried out at 25ºC and terminated using the UDP-Glo assay detection buffer. This kit detects UDP generated after UDP-sugar transfer by converting UDP to light (measured in relative luminescence units) in a luciferase type reaction. A linear calibration curve using 0–25 µM UDP was performed (fig ii). Following the kit’s protocol, the UGT assay was combined in a ratio 1:1 (10 ul:10 ul) with the UDPGlo™ detection reagent in individual wells (white flat bottom 384-well plate [Corning]). Luminescence signal was measured after 1-hour incubation using a Pherastar microplate.

- Calibration (standard) curve

The amount of UDP produced in the reaction was estimated by creating a UDP calibration curve. UDP standard concentrations (supplied in kit) were prepared in a 96-well plate and made in triplicates. 200µl of 25µM UDP solution was prepared in 1X glycosyltransferase buffer using the 10mM UDP standard. All 200µl of 25µM UDP solution was added to well B1 of a 96-well plate. 100µl of 1X glycosyltransferase buffer was added to each well of B2 through to B12. A serial twofold dilution was then done by transferring 100µl from well B1 to well B2, and from B2 to B3, pipetting to mix. This was repeated for wells B4 through B11. The extra 100µl from well B11 was discarded. Well B12 has no UDP in it and contained only the buffer. This is described as shown in Fig 2.2.

![Fig 2.2: Creation of a UDP calibration (standard) curve.](image)
2.2.3 Homology modelling and docking studies

The homology modeling process consists of the following steps: identification of template protein(s), sequence alignment of target and template protein, model building (based on template structure and alignment) and validation/refining/evaluation of models (Marti-Renom et al., 2000). Models being abstractions, contain errors and hence their accuracy can be compared to that of template (which is an actual experimental structure). The main sources of error in homology modeling include template choice, alignment inaccuracy and refinement inefficiency (Larsson et al., 2008; Hillisch et al., 2004). The quality of models partly depends on degree of sequence identity which at 25-50% make it useful for designing mutagenesis experiments (Burley et al., 2008; Larsson et al., 2008).

Molecular docking is a computational method utilised to predict ligand binding to a receptor (Magalhaes et al., 2004). Putative binding site as well as interactions of docked ligand to receptor is highlighted by docking results which helps understanding of many biochemical process (Hughes at al., 2011). Two stages are involved in automated docking: prediction of ligand position and orientation within a defined binding site via sampling algorithms and estimation of binding affinities of ligand and receptor (Meng e al., 2011). The most significant pitfall of docking lies in the inaccuracy of scoring function which ranks various binding poses (Haider, 2018).

Homology models would give insight into the structure and function of proteins, in the absence of experimental crystal structures. Various applications to drug discovery include the design of mutagenesis experiments, assessment of target druggability and study of protein function and mechanism (Cavasotto and Phatak, 2008). In this research, only models of group H UGTs 76E1, 76E2, 76E4, 76E5 and 76D1 were made to advance potential application of UGTs in the production of bioactive glycosides. The following sections will outline the modeling process.

2.2.3.1 Template selection

Protein sequences of UGTs 76E1 (Uniprot ID: Q9LTH3), 76E2 (Uniprot ID: Q9LTH2), 76E4 (Uniprot ID: Q9STE3), 76E5 (Uniprot ID: Q9STE6) and 76D1 (Uniprot ID: O48715) were downloaded from Uniprot (https://www.uniprot.org/downloads). BLAST algorithm against the protein data bank (PDB) was employed to carry out the homology search (https://blast.ncbi.nlm.nih.gov/BlastAlign.cgi). 3D structures and sequences of homologous proteins were extracted from the PDB and three templates with highest sequence similarity were selected. The three potential templates information is shown in Table 2.5.
Lastly, one of the potential templates 74F2 was finally chosen over 85H2 despite slightly higher sequence similarity. This is due to the presence of a ligand complex in its structure as well as belonging to the same family as studied UGTs. Multiple sequence alignments (of targets and chosen template) were carried out using sequencing tools implemented in the Molsoft ICM Pro 3.8-7 suite (http://www.molsoft.com/icm_pro.html).

<table>
<thead>
<tr>
<th>Plant UGTs (Template PDB)</th>
<th>% Similarity of templates to studied UGTs</th>
</tr>
</thead>
<tbody>
<tr>
<td>AtUGT74F2 (5V2K)</td>
<td>30 31 29 29 25</td>
</tr>
<tr>
<td>MtUGT85H2 (2PQ6)</td>
<td>33 33 32 32 28</td>
</tr>
<tr>
<td>MtUGT78G1 (3HBF)</td>
<td>28 27 27 27 25</td>
</tr>
</tbody>
</table>

Table 2.5: Top potential template UGT for homology modeling in order of % sequence similarity to studied UGTs

2.2.3.2 Secondary structure prediction

MSA was performed using five target UGTs (76E1, 76E2, 76E4, 76E5, 76D1) and template UGT 74F2. The amino acid sequences in FASTA format prior to sequencing is shown in Fig 2.3. These UGT sequences were moved to the sequence alignment tool of ICM. Secondary structures of models were predicted via the Molsoft ICM Pro 3.8-7 software. Protein secondary structure refers to the local conformation of its polypeptide backbone mainly α-helix, β-sheet and loop. These secondary structures provide information about protein relationships, activity and functions (Ma et al., 2018). Reliable and accurate structure prediction from protein sequences offers a significant first step towards tertiary structure prediction (Yang et al., 2016).

2.2.3.3 Modeling

The sequences of five UGTs 76E1, 76E2, 76E4, 76E5 and 76D1 (Uniprot ID: Q9LTH3, Q9LTH2, Q9STE3, Q9STE6 and O48715) were obtained from UniProt. The 3D models were generated using Modeler v 9.16, a modelling software (https://salilab.org/modeller/) based on the crystal structure of AtUGT74F2 (George-Thompson et al., 2017). Loop refinement of the models was performed using the loop refinement tools Molsoft ICM Pro 3.8-7 software.

2.2.3.4 Qualitative assessment of models

Models were validated using PROCHECK, which determines quality by validating stereochemical properties of models followed by comparing them with refined structures via Ramachandran plots. Ramachandran plots additionally evaluate quality of polypeptide backbone and side chains (Laskowski et al., 1993).
Furthermore, the model structures were also validated using PROSA II (Protein Structure Analysis) where Z-scores were calculated. Z-scores measure the compatibility between model sequence and structure and should be comparable to template’s Z-score (Wiederstein and Sippl., 2017). Pairwise structural superimposition of models with template 74F2 was done using MolSoft ICM Pro 3.8-7. This was to suggest functional and evolutionary relationships between both; showing regions of similarity between model and template (Koppensteiner et al., 2000).

2.2.3.5 Docking process

ICM docking today, is probably the most precise predictive tool of the binding geometry, consistently placing first in terms of precision compared to other leading docking software. The main pitfall being that the models does not reflect alternative confirmations of the receptor binding pocket. ICM algorithm tries to find the global minimum of the energy function that contains five grid potentials showing interaction of the flexible ligand with the receptor and the internal conformational energy of the ligand. A stack of alternative low energy confirmations are saved during the docking process (http://www.molsoft.com/icmpro/start-dock.html).

The PDB file you wish to dock to is read into ICM Molsoft. The PDB file is converted into an ICM object. Move the ligand out of the pocket; with two ICM objects a receptor and a ligand now in the workspace. Then select the ligand and click dock to start the process. Binding site residues can be done using the ICM pocket finder or the graphical selection tools. Receptor maps can be made by selecting the option and OK button clicked to start the process (Fig 2.3 shows the outlook after docking process is initiated). After docking, the energy units are in kcal/mol with the value reported in the score unitless. Docking is repeated 3 times and the lowest ICM score pose is selected. The best docking score to use for docking result analysis is the value under the heading SCORE. A score below -32 is regarded as a good docking score. The lowest score (below -32) for each UGT model was chosen.

Fig 2.3: ICM docking Interface at start of the docking process
2.2.3.5.1 Docking of donor compounds

Docking of sugar donor compounds UDP-Glc, UDP-Gal and UDP-GlcNAc into the models were done using the ICM Pro software (www.molsoft.com). Generally, depending on ligand size and the nature of the pocket, ligand docking takes approximately 10-30 seconds per ligand. Template-based docking protocol was used. The sugar donors were positioned (spatial orientation) in the models according to the location of UDP in 74F2 template. Default parameters for controlling docking process operations were used. Analysis of the hydrogen bonds contacts between UGT enzymes and ligand was done to identify optimal binding mode. Ligand binding sites were identified for all docked structures. Grid maps were created to define binding sites around the donor sugars. Then docking was run and conformation stored. A maximum of ten docked conformations was generated with a final conformation picked based on the strongest interaction energy. Docked poses were visualised using ICM Pro molecular modeling package. This protocol is identical for all models and donor sugars docked.

2.2.3.5.2 Docking of acceptor compounds

Due to the high acceptor substrate recognition by UGT76E1, this UGT is selected for further structural studies. This was carried out in order to recognise amino acid residues responsible for such better activity in 76E1. As reported in Chapter 3, 76E1 glycosylated more acceptor compounds with UDP-Glucose than other UGTs. Therefore, the acceptor substrates (kaempferol, quercetin, naringenin, diosmetin, apigenin, fisetin, hesperetin, esculetin and eriodictyol) were docked into 76E1 model alongside with donor sugar, UDP-glucose. Between acceptor substrates atoms and UGTs active sites, the hydrogen bonds, bond lengths and close contacts were analysed.
>76E1 (Uniprot ID: Q9LTH3)
MEELGKVRRIVLVPVAPQGHVTPIMQLGKALYSKGFSITVVLTQNYRVSSEEKDSDFHFLTIPGLSTES
DLKNGPFLFKLKNQICEASFQKCIQLQQDQQNDAQYDVEYMYSQAAKFQFLPSVLFSTTSA
TAFVCRSVELVSNVAESFLFDKMDPKVSDKEFPGLHPRLYKDLPTSAFQPSILKSVETVNTASA
ININSTCLESLLAWLQKLLQLQPVYPPGPIHIIAAASPSSLLEEDRSCLEWLKNQKIGSVYISGYSLGALME
TKDMLEMAWGLRNSNQFPLWIRPGSIPGSEWTESSLPEEFSLVSEERGIVKWAPQIEVLHPAVGG
FWSHCWGNSTLESIGEVPMPICRPFTGDQKVNARYLERVWRIQVGQLEGELDKGTHERAVELIMDE
EGAEMRKRVINLKEKASQVSRGGSSLDNFVNSLKMNNMF

>76E2 (Uniprot ID: Q9LTH2)
MEEKQKTRIVLVPVAPQGHVTPMQLGKALHSKSGFSITVVLTQNSRNVSSEEKDSDFHFLTIPGLST
ESDLQNLGPQKFLKNQCIESFKSCIGQLEHECNNDACVYDEYMYSHHAAKFQFLPSVVFST
TSATAFVCRSVELVSNVAESFLFDKMDPKVSDKEFSGLHPRLYKDLPTSAFQPIESTLVNTARTA
SAYINSASCSSSLARLQLQLQPVYPPGPIHITASAPSSLLEEDRSCLEWLKNQKIGSVYISGYSLGSL
ALMDTDMLEMAWGLRNSNQFPLWIRPGSIPGSEWTESSLPEEFLVSEERGIVKWAPQIEVLHPAVGG
FWSHCWGNSTLESIGEVEGMPICRPFTGDQKVNARYLERVWRIQVGQLEGELDKGTHERAVE
WLLVDIGSEMRARVIKLKTESVRSGRGSSLDNFVNSL

>76E4 (Uniprot ID: Q9STE3)
MEKRVEKKRIVLVPVAPQGHVTPMQLGKALQSKGGLFILTVAAQRQFNQGSLQHFPQFDFVTIPESLP
QSESKKLGPAYMLINNKLEASEFKECISQSMQGQNSIDYKMYFCAEAAKEFKISPISVIFSSSA
TQVCYCVLSESLAASKFLDPMQKDQKVLEHLPRLYKDLPTSGFQPLEPLLEMCREVNYKRTASA
VIINTASCSSSLALQLNLQFPLWIRPGSIPGSEWTESSLPEEFSLVSEERGIVKWAPQIEVLHPAVGG
FWSHCWGNSTLESIVEGMPICRPQEQKLNNYIESVWKGIQLEGELVERGVERAVKMLID
EGAAAMRER

>76E5 (Uniprot ID: Q9STE6)
MEKNAEKRRIVLVPFFLPQGHITPMMQLGQALNLKSGFSITVALGDSNRSSTQHFPGFQFVTIPETIPLS
QHEALGQEFVFVTNLKTSFSDKCIAHLLHGWIDNINDAICYDLEMFSETAKDLRISPFIVFTTGSSATN
HVCSCILSKNASEKFLDMPQEQNVLMVNEPLKIFCAEVNKRASAV
IINSSCLESLSWQLKQELIPYVPQPLHITASANFSLLEEDRSCLEWLKNQKLRFISVYVSGAHMETK
EVLMAWGLYNHSLQFPLWIRPGTESMPVETSQVSKIEGRTGIVKWAPQNEVLVHPAVGGFWSCGW
NSTLESIVEGMPICRPQFQSEKLNNYIESVWVRQVLLQGERGVERGVERAVKMLID
EGAAAMRER

>76D1 (Uniprot ID: O48715)
MAEIRQVRRLMVAPFPQGHLSMPLMSNLYAISQSGFISITVRFENFDKISHFNGPIKFTIKGRLSSESD
VKSLGLFLVLENSCSPEPLKKEFLTNHDVDFIIFYDEFYFPFRVAEDEMNLKPMVFSPPAATSISR
CVLMENQSNGLLQPDARQLEETVPEHFPFRRKDFLPTAYGSMERLMILYENVSNRAASSGGIHNSS
DCLENFSFTITAKQGKWPGYPLPQGLMNASCSPLSEERNKLEWLEKGQETSISVYISMSGLAMT
QDIEAVEMAGMFVQMNQPPLFLWIRPGSINGQESLDFLQEOQPQTVTDRGQFVKWAPQKEVLHR
AGVFVWNGWNCLESLISGSVMPICRPQSGQDQVRNTRLMSAWQYAEQUIGGERAVAVRRLID
VDQEGQEMRMATILKEEVEASVTTEGSSHNLNNLVHAIMMQIDEQ

Fig 2.4: Fasta format of amino acid sequences of UGTs studied structurally; amino acid sequence of each UGT follows identifier > and UGT name
2.2.4 Mutagenesis process

2.2.4.1 Primer design

Primer design was done using NEBaseChanger v1.2.7 (http://nebasechanger.neb.com/). Primers aid in generating the mutants and are designed on the back-to-back principle (Kunkel, 1985). Back to back principle prevents overlapping (the circularisation of product to form nicked plasmid which results in low efficiency transformation) thereby allowing exponential amplification to generate the desired product more (Kalnins et al., 1983; Dickinson et al., 2013).

From ‘The Arabidopsis’ database site (http://www.p450.kvl.dk/At_ugts/table.shtml), the cDNAs sequences of the WT UGTs to be mutated were extracted. DNA sequences coding for the desired amino acids were used to substitute the DNA sequences coding for the amino acids to be replaced. An example is highlighted for the design of 76E2 N320S primer (Fig 2.5), showing the standard technique for all primers designed in this report.

- UGT 76E2 cDNA was introduced into NEBaseChanger
- The N320 cDNA encoding region (amino acid 958 – 960, AAT) was selected
- Then the desired amino acid sequence AGT was set (serine: AGT or AGC)
- Automatically, the forward and reverse primers as well as properties (annealing temperature (Tₐ) and melting temperature (Tₘ) needed at the next stage were provided.

![NEBaseChanger](image_url)

**Fig 2.5:** Primer design for 76E2 N320S where N genetic code (highlighted in amber) is replaced by S genetic code (highlighted in blue box). Primer properties such as annealing temperature is calculated and highlighted in red box.
2.2.4.2 Site-directed mutagenesis (SDM)

Templates DNAs were extracted using QIAprep Spin miniprep kit protocol. Q5 Site-directed mutagenesis kit protocol was observed to make mutants from the wild type template DNAs. Mutagenesis was done in three stages of exponential amplification (polymerase chain reaction), digestion and transformation. The steps are shown in detail below:

**Step 1: Exponential Amplification** - Each mutant is assembled as shown in Table 2.6. The cycling conditions are highlighted in table vii as advised by the supplier:

<table>
<thead>
<tr>
<th>25 µL reaction</th>
<th>Final concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q5 Hot Start High-Fidelity 2X Master Mix</td>
<td>12.5 µL 1X</td>
</tr>
<tr>
<td>10 µM Forward Primer</td>
<td>12.5 µL 0.5 µM</td>
</tr>
<tr>
<td>10 µM Reverse Primer</td>
<td>1.25 µL 0.5 µM</td>
</tr>
<tr>
<td>Template DNA (1-25 ng/µL)</td>
<td>1.0 µL 1-25 ng</td>
</tr>
<tr>
<td>Nuclease-free water</td>
<td>9.0 µL</td>
</tr>
</tbody>
</table>

Table 2.6: Composition of the PCR reaction of SDM

<table>
<thead>
<tr>
<th>Step</th>
<th>Temperature</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial denaturation</td>
<td>98°C</td>
<td>30 seconds</td>
</tr>
<tr>
<td>25 cycles</td>
<td>98°C</td>
<td>10 seconds</td>
</tr>
<tr>
<td></td>
<td>50-72°C (according to the specific primer T&lt;sub&gt;A&lt;/sub&gt; listed in Table 2.9)</td>
<td>10-30 seconds</td>
</tr>
<tr>
<td></td>
<td>72°C</td>
<td>20-30 seconds/kb</td>
</tr>
<tr>
<td>Final extension</td>
<td>72°C</td>
<td>2 minutes</td>
</tr>
<tr>
<td>Hold</td>
<td>4-10°C</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.7: PCR cycling steps and conditions

*Exponential amplification products were verified using agarose gel electrophoresis. In TBE buffer, 0.8% agarose gel with 0.1% SYBR® safe DNA gel stain was prepared and then cooled and set in a gel cast. 5 µL amplification product mixed with 5 µL DNA loading buffer was loaded in a gel and run at 40 volts for 2 hours in TBE tank buffer in the electrophoresis process. A gel imager assessed the gel; a clear gel band at about 6kb showed a positive amplification process. After positive verification (example shown in Fig 2.6), the next step of mutagenesis (digestion) was carried out as shown in Table 2.6. This mix was incubated for 5 minutes at room temperature.
Step 2: KLD (Kinase, ligase and DpnI) reaction

After amplification via PCR, digestion followed which include rapid circularisation of the PCR product and removal of the template DNA. This composition of the reaction is shown in Table 2.8.

<table>
<thead>
<tr>
<th></th>
<th>Volume</th>
<th>Final Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR Product</td>
<td>1.0 µL</td>
<td></td>
</tr>
<tr>
<td>2X KLD Reaction Buffer</td>
<td>5.0 µL</td>
<td>1X</td>
</tr>
<tr>
<td>10X KLD Enzyme mix</td>
<td>1.0 µL</td>
<td>1X</td>
</tr>
<tr>
<td>Nuclease free water</td>
<td>3.0 µL</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.8: Composition of the KLD reaction of SDM

Step 3: Transformation

5 µL of the KLD mix was added to 50 µL of chemically competent cells. This was incubated on ice for 30 minutes. Heat shock followed at 42°C for 30 seconds. This was incubated again on ice for 5 minutes. 950 µL SOC media is added to the mix and gently shaken at 37°C for 1 hour. Thereafter, 50-100 µL of the mix is spread on an agar selection plate and incubated overnight at 37°C. A single colony is picked from the plate and cultured in LB media (containing 50 µg/mL ampicillin). 4ml of this culture is passed through the miniprep process for DNA extraction and that left is stored as glycerol stock as described earlier.

2.2.4.3 Mutant plasmid DNAs extraction via miniprep

Wild-type DNAs (used as template DNA in SDM) were extracted through the same method. 4ml bacterial culture was centrifuged at 13,000 rpm for 3 minutes at room temperature.
This was resuspended in 250 µL buffer P1 (resuspension buffer) and 250 µL buffer P2 added (lysing buffer) and mixed thoroughly. Buffer N3 (neutralisation buffer) was thereafter added which was followed by centrifugation for 10 minutes. A clear supernatant (containing DNA) forms on the upper part while a compact pellet remains in the bottom. 800 µL of the supernatant was transferred into spin column and centrifuged for 60 seconds. The flow-through was discarded and the spin column was washed with 750 µL wash buffer (buffer PE). Centrifugation for 60 seconds was repeated and flow-through discarded. Residual wash buffer was removed via centrifugation for 60 seconds again. The spin column was transferred to a clean 1.5 ml tube. 50 µL buffer EB (elution buffer 10 mM Tris-Cl, pH 8.5) was added to the spin column to elute the DNA. This was let to stand for 60 seconds and centrifuged for another 60 seconds. The extracted plasmid DNAs was then stored in clean tubes.

2.2.4.4 DNA sequencing

The mutant DNAs were sent for sanger sequencing (SourceBioScience) to confirm mutations. The primer sequences; forward primer and reverse primers (shown in Table 2.9) were confirmed from sequencing data to contain the desired mutation.

<table>
<thead>
<tr>
<th>Mutants</th>
<th>Forward primer</th>
<th>Reverse primer</th>
<th>TA (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>76D1 G347C</td>
<td>TTGGAACCATTGCGGATGGAAACTCGTG</td>
<td>AACCTCCCACGTCTCTA</td>
<td>64</td>
</tr>
<tr>
<td>76E2 D374E</td>
<td>TTTCAACGCGGAGCAGAAAGTCA</td>
<td>GGCCTACATATCATCGGAAC</td>
<td>64</td>
</tr>
<tr>
<td>76D1 P129T</td>
<td>GGTCTTTGAGACTTCCTTCCGGCG</td>
<td>ATCTTTGGCAAGATCATCTTCCG</td>
<td>66</td>
</tr>
<tr>
<td>76E4 K275L</td>
<td>CTTGGAACCCTTCCTGCACTTG</td>
<td>CTTATGTATATGACTGACCTTG</td>
<td>59</td>
</tr>
<tr>
<td>76E2 N320S</td>
<td>AGAGGAATTCAGTAGGTTGTTTC</td>
<td>GGTAAAGACTCTGTCCTTCC</td>
<td>58</td>
</tr>
<tr>
<td>76E5 S311N</td>
<td>AGTGGAAATGCAATAAGATTGTCCTC</td>
<td>GGCAATGACCTGACCTTG</td>
<td>57</td>
</tr>
<tr>
<td>76E1 S318N</td>
<td>GGAGGAATTCAATAGGGTTGTTT</td>
<td>GGTAAAGACTCTGTCCTTCC</td>
<td>56</td>
</tr>
<tr>
<td>76E5 S311A</td>
<td>AGTGGAAATGCGAACAGATGGCTTGCGGAAG</td>
<td>GGCAATGACCTGACCTTG</td>
<td>62</td>
</tr>
<tr>
<td>76E1 S318A</td>
<td>GGAGGAATTCCGCGAGTTGTTGGTGTCAG</td>
<td>GGTAAAGACTCTGTCCTTCC</td>
<td>61</td>
</tr>
<tr>
<td>76E4 Q375A</td>
<td>ACAAGGCCGAACGGGAAGTTAAATGCGATG</td>
<td>AAAGGGCTGCAAAATCATTG</td>
<td>59</td>
</tr>
<tr>
<td>76E4 E374A</td>
<td>TTTCAGGGCAGCAGAATTGAAATGCG</td>
<td>GGCGTCAATGCAATGGG</td>
<td>60</td>
</tr>
<tr>
<td>76E5 Q366A</td>
<td>TAACGTTGAGGCGGAAAGTAAACCGC</td>
<td>AAGGTCTGCAATCATTG</td>
<td>58</td>
</tr>
<tr>
<td>76E1 Q373A</td>
<td>TACCGGAGATCAGCGAAGATCAGTG</td>
<td>AAAGGGTCTGCAATCATTG</td>
<td>57</td>
</tr>
</tbody>
</table>

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

The sequencing primer was PGEX 3’ with sequence CCG GGA GCT GCA TGT GTC AGA GG. This step was followed by transformation and the process of protein production for mutant UGTs as explained in section 2.2.1. All wildtype and mutant sequences were compared using BLAST.
References


Chapter 3 Substrate studies of UGTs; \textit{In vitro} biochemical characterisation
As introduced in Chapter 1, the addition of sugar molecule from nucleotide sugar donors to plant secondary metabolites, which are the sugar acceptors are catalysed by plant UGTs. The importance of this catalysis has been highlighted in chapter 1 (1.2.1). A growing figure of UGT genes are being identified due to the advances made in genome sequencing, sequence analysis and high throughput screening (Tiwari et al., 2016). Not yet fully investigated is the in vitro activities of these UGTs, hence there is an inadequacy in the available experimental data on their biochemical behaviour.

This chapter reports on the in vitro characterisation of recombinant UGTs from Arabidopsis thaliana groups H and L. Mass spectrometry (MS) based methods were used to qualitatively examine substrate specificity and catalytic mechanism while bioluminescence methods were utilised to briefly investigate the kinetics of the UDP-donor substrate reactions.
3.1 Mass spectrometry (MS)-based GT activity

This section explains the MS-based method utilised in examining the substrate specificities via monitoring the formation of reaction products qualitatively. MS has the benefit of directing a label free assay, permitting a speedy determination of enzyme substrate specificity without altering the reaction (Wagner and Pesnot., 2010). Details of the general outline of procedure shown in Fig 3.1 will be discussed in the subsequent sections.

Fig 3.1: General scheme of workflow for UGT enzyme activity test
3.1.1 Qualitative analysis

The main goal of qualitative analysis is to identify the formation of intended glycosylated product against different substrates of interest catalysed by target UGTs. The presence of the glycosylated product indicates positive activity of a UGT against a substrate. Using the MS full scan analysis mode, a total ion current (TIC) plot is obtained which indicates all compounds present (in form of peaks molecular weight as well as signal intensity). An example of a full scan analysis TIC obtained for apigenin glycosylation is shown in Fig 3.2.

![Mass spectrum](image)

Fig 3.2: A mass spectrum (extracted from total ion count (TIC) plot) showing a full scan analysis of glycosylation of apigenin (MW 270). Apigenin glucoside (MW 432) was detected on a negative-ion mode.
Here, an expected glycosylated reaction product’s mass ion (with MS modifications such as the negative-ion mode [M-H\(^{+}\)]) was detected. Furthermore, an MS/MS analysis using the selected ion monitoring (via the product ion mode) was utilised to affirm the product’s detection. This is more precise as only compound(s) whose mass is/are selected is/are detected and plotted. Without MS/MS confirmation (fragmented form of aglycone ions), the possibility of a false signal exists in the first step (full scan analysis).

A SIM (product ion scan) of apigenin glucoside (MW 432) fragmentation is shown in Fig 3.3 where peak 431 (M-H\(^{-}\)) has been selected for fragmentation. The presence of aglycone apigenin (MW 269), a fragment from the glucoside confirmed the positive glycosylation activity (Fig 3.3).

![Product-ion scan showing the fragmentation of the selected apigenin glucoside (peak 431) resulting to aglycone apigenin (peak 269); confirming the glycosylation activity of a UGT](image)

Fig 3.3: A product-ion scan showing the fragmentation of the selected apigenin glucoside (peak 431) resulting to aglycone apigenin (peak 269); confirming the glycosylation activity of a UGT
The modes of data acquisition with MS utilised in this qualitative analysis is outlined in Fig 3.4 below.

![Diagram of MS data acquisition modes](image)

**Fig 3.4: Mode of acquiring MS and MS/MS data in this work**

MS has distinct conditions guiding its experiments therefore in relation to this work, these considerations will be discussed next.

### 3.1.1.1 Reaction conditions

For the UGT activity assay (for qualitative analysis), sample concentrations were as follows: acceptor compound (100 µM), donor compound (100 µM), and the UGT enzyme in buffer (43 µL, 1 mM Tris, 1 mM MgCl₂, pH 7.6). Thereafter, the samples were incubated at 37°C for 16 hours. Qualitative MS analysis followed for the reactions (full scan analysis and product ion scan). A long incubation period of 16 hours was used to guarantee that the enzymes have enough time for reaction catalysis. While it is possible that instability or degradation of enzyme resulting in decline in catalytic ability might occur during this long incubation, this would only affect rate of reaction and or/ amount of product released. Hence, the aim of the qualitative analysis which was to establish whether the product could be formed remains undisturbed. A long incubation time has been utilized in many previous screening experiments (Jung et al., 2014; Wei et al., 2015).

37°C was chosen for this qualitative analysis because it is a common optimum temperature for enzymatic reaction involving UGTs. For this qualitative analysis, this temperature was enough to show positive activity or inactivity (Yu et al., 2015; Crawford et al., 2012). Buffer is essential for enzyme assays to maintain the reaction pH.
Tris buffers have been reportedly used in many GT assays with the pros of low cost, relatively MS friendly and good buffering ability. Only low concentration of buffer works with MS (Yang et al., 2005). While few cons of Tris buffers namely inhibition and destabilization of structure have been indicated for certain enzymes, its suitability and frequent usage in GT assays in particular have been indicated in literature (Kim et al., 2014). A low concentration of metal cofactor alongside low concentration of buffer have been used in reactions catalysed by GTs. Low concentration of both such as 1 mM Tris, 1 mM MgCl₂ was adequate environment for buffering the enzyme reactions (Offen et al., 2006; Flint et al., 2005; Yang et al., 2005; Yang et al., 2007). A wide range of pH have been studied and described for GTs with different kinds of buffers. From 6.0 to 9.0 in HEPES to 7.0 in MES and 6.5 to 8.5 in Tris and phosphate buffers, glycosylation activity was found across board (Gandia-Herrero et al., 2008; Jin et al., 2013).

3.1.1.2 Conditions of MS
As explained in Fig 3.4, the usage of MS in this research involved the full scan analysis and product ion scan. Following the reaction conditions above, samples were observed for the presence of any glycosylated product formed. This was done in two parts; firstly, an observation of a peak matching the molecular weight (MW) of an expected glycosylation product using full scan mode. Many times, a negative-ion mode (of the full scan mode) gives better sensitivity than positive-ion mode due to proton loss. Therefore, a minus-one adduct of the expected glycosylation product appears in the negative-ion mode. The full scan mode typically shows all ions within detection range in the reaction mixture. This detection range should be wide enough to contain all possible ions.

In this work, the donor specificity screening was set at 100-1000 Daltons because with the largest sugar donor (UDP-GlcNAc) screened, quercetin-GlcNAc [(QUE-GlcNAc)-H⁺] of MW 504 might be observed. For the acceptor screening, the range was also set at 100 – 1000 Daltons as the MW of product novobiocin-Glc is 774. Secondly, the possible glycosylated product is then fragmented using the product ion mode known as the MS/MS analysis. The MW of the aglycone fragment is expected to be detected as the glycosidic bond produced will be broken here. The fragmentation of the glycosylated product to the original aglycone acceptor confirmed the glycosylation reaction. A level of collision energy enough to ensure fragmentation must be used. 10 eV was fixed and used in this work to safeguard against excessive fragmentation into undetectable smaller fragments.
3.1.2 Quantitative analysis

To study the kinetics of the donor screening reaction, UDP-Glo™ glycosyltransferase assay was employed. This assay uses a single-reagent addition procedure to detect UDP release in glycosyltransferase reactions. A UDP detection reagent converts UDP formed and converts to ATP, which in turn produce light via luciferase reaction. This assay can only be used with purified glycosyltransferases that utilise UDP-sugar as donor and works well with affinity tagged glycosyltransferases.

![UDP-Glo assay principle](image)

Fig 3.5: The principle of UDP-GloTM glycosyltransferase assay

As shown in Fig 3.5 above, the principle of UDP-Glo assay was outlined, where the ATP is converted into a luminescent signal which is proportional to the glycosyltransferase activity in the reaction. The assay involved a single step after the completed glycosyltransferase reaction where an equal volume of UDP detection reagent was added to the glycosyltransferase reaction. The UDP detection reagent consisted of nucleotide detection reagent (ATP detection substrate and lyophilized luciferase enzyme) and UDP-Glo working solution (UDP-Glo enzyme and enzyme dilution buffer). The detection reagent terminates the glycosyltransferase reaction. This was mixed with a plate shaker for 30 seconds and then incubated at room temperature for 60 minutes. The luminescence was then measured with a plate reading luminometer.

3.1.2.1 Calibration curve

The amount of UDP produced in the reaction was estimated by creating a UDP calibration curve. UDP standard concentrations (supplied in assay kit) were prepared in a 96-well plate and made in triplicates. The buffer suggested by the kit manufacturer was 1X glycosyltransferase reaction buffer (50mM Tris 5mM MgCl₂). 200µl of 25µM UDP solution was prepared in 1X glycosyltransferase buffer using the 10mM UDP standard. All 200µl of 25µM UDP solution was added to well B1 of a 96-well plate. 100µl of 1X glycosyltransferase buffer was added to each well of B2 through to B12. A serial twofold dilution was then done by transferring 100µl from well B1 to well B2, and from B2 to B3, pipetting to mix.
This was repeated for wells B4 through B11. The extra 100µl from well B11 was discarded. Well B12 has no UDP in it and contained only the buffer. The dilution scheme for is shown in Fig 3.6. The luminescence output of the assay is proportional to the concentration of UDP in the calibration curve, hence the volume of the UDP standards used was the same as the volume of the donor screening reactions tested. Therefore, a volume of 10µl of each UDP standard from the 96 well plate was transferred to a separate plate and then the assay protocol applied.

Fig 3.6: Creation of a UDP calibration (standard) curve

A calibration curve prepared over the indicated range of UDP concentrations was obtained to check the amount of UDP released via the luminescence output in relative light units (RLU) using UDP as standard. Good linearity of UDP’s concentration with typical R² values of 9.89E-01 showing the precision and ability of the analytical method (Fig 3.7).

<table>
<thead>
<tr>
<th>UDP Concentration (µM)</th>
<th>Luminescence (RLU)</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>1,231,237</td>
</tr>
<tr>
<td>12.5</td>
<td>706,608</td>
</tr>
<tr>
<td>6.25</td>
<td>209,590</td>
</tr>
<tr>
<td>3.13</td>
<td>124,714</td>
</tr>
<tr>
<td>1.56</td>
<td>51,560</td>
</tr>
<tr>
<td>0.78</td>
<td>29,578</td>
</tr>
<tr>
<td>0.39</td>
<td>14,090</td>
</tr>
<tr>
<td>0.20</td>
<td>7,494</td>
</tr>
<tr>
<td>0.10</td>
<td>2,757</td>
</tr>
<tr>
<td>0.05</td>
<td>1,473</td>
</tr>
<tr>
<td>0.02</td>
<td>581</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 3.1: Luminescence measured 1 hour after the addition of UDP Detection reagent for each UDP concentration of standard curve
Fig 3.7: UDP calibration curve was prepared over the range of UDP concentrations 0-25µM in 10µl of 1X glycosyltransferase reaction buffer

3.1.2.2 Kinetic study: mechanisms and conditions

Factors such as temperature influence the rate of luciferase reaction (UDP-Glo assay) and as such affects the stability and intensity of light output i.e. luminescent signal (Zegzouti et al., 2013). All reagents and assay plates were equilibrated to room temperature for consistency. The assay was also carried out at 20°C. Enzymatic rates and luminescence intensity would be affected by the chemical environment of the assay. Hence, the 1X glycosyltransferase buffer which provided chemical environment for the assay was fixed at pH 7.5 as a buffer pH of between 6 and 9 was recommended by the manufacturer. UDP-sugars being susceptible to hydrolysis could generate background in the assay by releasing UDP moiety. This is common when testing low-activity enzyme with high concentration UDP-sugar leading to diminished assay sensitivity. Usage of very pure and stabilised sugar donor (less than 0.005% UDP contamination) prevents low assay sensitivity, high background and UDP feedback inhibition of certain GTs. We have utilised ultra-pure UDP-sugar substrates (UDP-Glc, UDP-Gal and UDP-GlcNAc) from the manufacturer to avoid this potential challenge that could give flawed results. Solid white, multiwell plates were used in the assay as they offer maximum light reflection and gives higher output signal. Instrument optimisation was carried out through the UDP standard curve.
• Selection of enzyme concentration

To choose a suitable enzyme concentration for the kinetic analysis, each UGT enzyme is screened for a minimum enzyme concentration required for activity. The range of concentration between 0.001 – 0.1 mg/ml was used. Fixed substrate concentrations of 20µM KMP and 50µM donor sugar was utilised. Through a time of 0 – 30 minutes, the activity (rate of reaction) is determined through the estimation of amount of UDP released. The lowest enzyme concentration showing activity as shown through plot of UDP released against time is chosen as the optimum enzyme concentration for that UGT enzyme. This is because very little enzyme concentration is needed for the reaction catalysis.

• Selection of substrate concentration

This selection is in two parts; the acceptor substrate and the donor substrate. For each UGT enzyme, acceptor substrate (KMP or QUE) concentration was varied between 0.5µM and 500µM while concentrations of donor and enzyme fixed (enzyme concentration has earlier been determined). 20µM KMP/QUE was found to be high enough concentration for activity. This was fixed for the acceptor substrate throughout the kinetics reactions. Donor substrate concentration was varied between 1µM - 1000µM and a range over six-unit points (showing progressive reaction rates) is selected for each UGT enzyme.

• Determination of pseudo kinetics

In this research, the acceptor substrate concentration is fixed while that of donor sugar is varied since the rate of reaction depends on amount of UDP released. Pseudo-kinetic analysis including a series of fixed concentration of a substrate and varying the concentrations of the other is required to give a comprehensive kinetic picture. A 2nd order reaction can be challenging due to difficulty in measuring both reactants concurrently. The cost implications of measuring changing amounts for both reactants add to the challenge. Therefore, pseudo-first order reaction which involves treating a 2nd order reaction like a 1st order reaction was employed. This meant that one of the reactants would have a remarkably high concentration and the other a remarkably low concentration. Furthermore, an assumption that the reactant of high concentration remaining constant during the reaction was adopted since its consumption would be so minute prompting a negligible change in concentration. Therefore, the rate of reaction was measured solely based on one reactant (one of low concentration). This is referred to as pseudo-first order reaction.
The kinetic profiles of some of these UGTs (from group H) were obtained by fixing a high concentration for one substrate to create a pseudo-Uni substrate condition. Although the usage of the data is limited, they can nevertheless highlight the UGTs scale of preference for a sugar donor. This is important to type of glycosylated product formed from a GT reaction. This work reports the pseudo kinetics of 76E1, 76E2, 76E5 and 76D1 donor reactions (acceptor: KMP or QUE and UDP-Glc for all). 76E4 kinetics was not done as it showed no positive glycosylation activity with KMP/QUE and UDP-Glc. The general technique employed in carrying out this is summarised in Fig 3.5.

3.2 Substrate specificity studies

Twenty-two (22) recombinant UGTs from group L (84A2, 84A3, 84A1, 84A4, 74E2, 75B1, 74C1, 75D1, 74F2 and 75C1), group H (76E2, 76E1, 76D1, 76E5, 76E4, 76E3 and 76F2), group D (73D1, 73B1, 73C2 and 73C3) and group B (89A2) were selected for the project. Protein expression was attempted for all UGTs but only fifteen (15) were successfully expressed. Therefore, fifteen (15) UGTs 76E2, 76E1, 76D1, 76E5, 76E4, 84A2, 84A3, 84A1, 74C1, 73C2, 76E3, 76F2, 73D1, 75D1 and 89A2 proceeded to the initial screening stage. Initial glycosylation activity check using protein lysate followed for these UGTs. UDP-glucose was set up as the donor sugar alongside with kaempferol and scopoletin as acceptor compounds. Only 8 of fifteen initially screened UGTs exhibited glycosylation. They are UGTs 76E2, 76E1, 76D1, 76E5, 76E4, 84A2, 84A3 and 84A1 of groups L and H only. These UGTs therefore proceeded to the substrate (donor and acceptor) specificity screening and results presented as follows.

A sugar donor library containing seven nucleotide sugars (Fig 3.13) was set up for donor specificity screening. They are UDP-Glucose (UDP-Glc), UDP-Galactose (UDP-Gal), UDP-N-acetylglucosamine (UDP-GlcNAc), UDP-Mannose (UDP-Man), GDP-Glucose (GDP-Glc), GDP-Mannose (GDP-Man) and GDP-Fucose (GDP-Fuc). These donor sugars present a quite diverse scope for UGT donor screening due to varying glycosyl and nucleoside parts.

The acceptor library mainly comprises different kinds of plant secondary metabolites, including flavonoids, coumarins, alkaloids, cinnamic acids, plant hormones, cinnamic acids, benzoic acids and an antibiotic. Details of the structure, names and molecular weights of the 42 compounds of acceptor library is as shown in Chapter 2 Table 2.4. The next sections will describe the experimental data of UGTs’ activities towards 42 acceptor compounds in vitro using UDP-Glc as donor compound.
3.2.1 Acceptor specificity

Samples were analysed on both full scan and product ion MS modes. The results were subsequently abridged and presented on a GAR screen (Yang et al., 2005). Green/amber/red colours show positive/ambiguous/no activity of screened UGTs towards the acceptor compounds. 38 of the 42 acceptor compounds screened mainly belonged to various classes of plant secondary metabolites. Other four non-plant secondary metabolites were screened including plant hormones, antibiotic, vitamin and an anthraquinone. The summary of the UGTs activity with the acceptor compounds is described in Fig 3.8. A total of 19 acceptor compounds showed positive activity with all UGTs is shown in Fig 3.9 (as numbered in the GAR screen).

![Diagram of acceptor compounds]

**Fig 3.9:** Acceptor compounds displaying positive activity across all UGTs

- **[Phenolic acids (including coumarins)]**
  - 2,5 DHBA, Scopoletin, Esculetin, Sinapic acid, Caffeic acid, 3,4 DHBA, Ferulic acid

- **Flavonoids**
  - Kaempferol, Quercetin, Fisetin, Apigenin, Diosmetin, Hesperetin, Naringenin, Eriodictyol, Myricetin, Diadzein

- **Stilbenes**
  - Resveratrol

- **Alkaloids**
  - Galantamine

19 acceptors with positive activity
| Kaempferol | Quercetin | Fisetin | Apigenin | Diosmetin | Hesperetin | Narigenin | Eriodictyol | 2,5 DHBA | Scopoletin | Esculetin | Myricetin | Diadzein | Galanthamine | Sinapic acid | Caffeic acid | 3,4 DHBA | Ferulic acid | Resveratrol |
|------------|-----------|---------|----------|-----------|------------|-----------|-------------|----------|------------|----------|-----------|----------|----------|-------------|-------------|-------------|----------|------------|-----------|
| 76E1       |           |         |          |           |            |           |             |          |            |          |           |          |          |             |             |             |          |            |           |          |
| 76E5       |           |         |          |           |            |           |             |          |            |          |           |          |          |             |             |             |          |            |           |          |
| 76E2       |           |         |          |           |            |           |             |          |            |          |           |          |          |             |             |             |          |            |           |          |
| 76D1       |           |         |          |           |            |           |             |          |            |          |           |          |          |             |             |             |          |            |           |          |
| 84A1       |           |         |          |           |            |           |             |          |            |          |           |          |          |             |             |             |          |            |           |          |
| 84A2       |           |         |          |           |            |           |             |          |            |          |           |          |          |             |             |             |          |            |           |          |
| 84A3       |           |         |          |           |            |           |             |          |            |          |           |          |          |             |             |             |          |            |           |          |
| 76E4       |           |         |          |           |            |           |             |          |            |          |           |          |          |             |             |             |          |            |           |          |

Fig 3.8: Summary of GAR screening results of acceptor specificities of all UGTs (UDP-Glc as donor)

- **Green**: Positive activity
- **Yellow**: Ambiguous activity
- **Red**: No activity
3.2.1.1 Group H UGTs

The acceptor preferences for the five group H UGTs namely 76E1, 76E2, 76E5, 76E4 and 76D1 members are shown (Fig 3.8). Except for 76E4, these findings have been reported (Akere et al., 2018). Major classes of plant secondary metabolites are represented in Fig 3.10. The acceptors recognised in this group were mainly of a mix of flavonoids and phenolic acids groups, of the polyphenols. Other plant secondary metabolites tested such as alkaloids and plant hormones were not used by these UGTs.

Some new acceptor specificities (with UDP-Glc as donor) of group H UGTs were discovered in this research. This includes apigenin, diosmetin, hesperetin, naringenin and eriodictyol for 76E1, myricetin for 76E2, scopoletin and 2,5-DHBA for 76E5, fisetin for 76D1 and 2,5-DHBA for 76E4. Other acceptor activities already reported and also found in this work include kaempferol, quercetin, fisetin and esculetin for 76E1, kaempferol, quercetin and fisetin for 76E2, kaempferol and quercetin for 76E5, scopoletin for 76E4 and quercetin and fisetin for 76D1 (Yang et al., 2019; Lim et al., 2003; Lim et al., 2004).

Fig 3.10: Major classes of plant secondary metabolites
76E1 recognised the most compounds among all Group H UGTs. All nine (9) positive activities are of flavonoids and a coumarin. 76E4 on the other hand displayed the least positive activity with only two compounds (scopoletin and 2,5-DHBA), neither being flavonoids. 76E2 could only recognise flavonoids. Worthy of important note from this group is 76E1, which aside being able to recognise more acceptor compounds also could tolerate the presence of chiral centre in some flavonoids especially flavanones (hesperetin, naringenin and eriodictyol). These findings are summarised in Fig 3.11.

About more than half of drugs currently in use are chiral compounds. Most isomers of chiral drugs exhibit marked differences in biological activities (Nguyen et al., 2006). These isomers of chiral compounds, also referred to as enantiomers, can lead to different biological effects making chirality of potential importance (Ribeiro, 2017).

![Acceptor compound recognition among group H UGTs](chart.png)

**Fig 3.11**: Bar chart showing pattern of acceptor substrate recognition among group L UGTs

UGTs show regioselectivity in the event of an acceptor presenting multiple binding sites for a donor sugar (Lim et al., 2003). That means, they are selective for a particular OH group out of many possibilities. Due to numerous therapeutic and commercial benefits, flavonoids glycosylation has been fairly studied. For example, kaempferol-3-glucoside (astragalin), a bioactive flavonoid has immense pharmacological properties which include antioxidant, anti-inflammatory, anticancer, neuroprotective and cardioprotective properties. Various *in vitro* and *in vivo* researches have explained its medicinal characteristics and mechanism of actions (Riaz et al., 2018).
Also, reports have indicated that flavonoids in general have shown various biological effects such as antiallergic, antibacterial, antiviral, anti-inflammatory, antithrombic, hepatoprotective and antioxidant activities (Aruoma et al., 2006).

Flavonoids produced in Arabidopsis thaliana constitutes one of the major secondary metabolites, projected to exceed 5000. Flavonoids possess a variety of biological activities not only to plant but to animals who take flavonoids as part of their diets (Wang et al., 2009; Saito et al., 2010). As explained in chapter 1 (1.2.1), glycosylation improves pharmacological properties of these plant secondary metabolites (PSM). Hence, UGTs which can add sugar to improve the bioactivities of these medicinal PSM especially flavonoids are of immense importance.

3.2.1.2 Group L UGTs

The acceptor screening results of the three members of group L, UGTs 84A1, 84A2 and 84A3 are summarised in Fig 3.12. The choice of acceptor substrates is more diverse here with a mix of flavonoids, phenolic acids, alkaloid and stilbene activity observed. 84A2 recognised ten (10) acceptor compounds, mainly flavonoids and phenolic acids. 84A1 and 84A3 exhibited a little more diversity showing positive activity with an alkaloid and a stilbene respectively.

New acceptor specificities from this study include daidzein and galantamine for 84A1, naringenin, eriodictyol, myricetin, 3,4-DHBA for 84A2 and apigenin, scopoletin and resveratrol activity for 84A3. We also found cinnamic acid activity (ferulic acid, sinapic acid, caffeic acid) for the three UGTs as previously reported by Li et al., 2001. 84A2 just as in 76E1 (group L) accepts flavonoids such as naringenin and eriodictyol with chiral centres.

Apart from the UGT enzymes discussed above, all other UGTs 74C1, 73C2, 76E3, 76F2, 73D1, 75C1 and 89A2 were examined but did not show detectable glycosyltransferase activity (UDP-Glc) in the current conditions. In many acceptor compounds, more than one hydroxyl (OH) functional group, which is the site of glycosylation is present.

When this occurs, UGTs display regioselectivity by transferring sugar to a specific position on the acceptor. Regioselectivity has been studied in coumarins where UGT 71C1 was found to be selective for 3OH and not 4OH of caffeic acid (Lim et al., 2003). Quercetin, a common flavonoid acceptor compound with five OH groups is often glycosylated to increase its bioavailability, stability and solubility. The position of OH glycosylation has a great effect on its biological activity and potential medical benefits to humans (Day et al., 2003; Vogt et al., 1997).
3.2.2 Donor specificity

Some of the studied UGTs have previously been reported to be glucosyltransferases such as UGTs 76E1, 76E2 (Lim et al., 2003, Akere et al., 2018; Yang et al., 2019), 84A3, 84A1 (Lim et al., 2001; Lim et al., 2003; Yang et al., 2019), 84A2 (Lim et al., 2001; Yang et al., 2019), 76D1 (Lim et al., 2003; Akere et al., 2018; Huang et al., 2019), 76E4 (Yang et al., 2019) and 76E5 (Akere et al., 2018; Yang et al., 2019). Although UGTs are generally specific with donor sugar preference, activity with donors other than UDP-Glc is known (Ross et al., 2001). Therefore, all eight UGTs were screened through a library of donor sugars (Fig 3.13) to establish their preferences.
Following UGT activity assay, samples were analysed on both full scan and product ion MS modes to detect products of glycosylation reaction. The results were subsequently abridged and presented on a GAR screen (Yang et al., 2005). Green/amber/red colours indicate positive/ambiguous/no activity towards screened UGTs, and donor compounds as shown in Fig 3.14.

Apart from supporting previous works, our experimental data shown below validates the methodology employed in this research. UDP-Glc was the most commonly used donor for the screened UGTs in agreement with previous research which highlighted its frequent use among plant UGTs as the favoured donor compound (Li et al., 2001). UGTs 76E1, 76E2, 76D1, 84A1, 84A2, 84A3 and 76E5 were able to use UDP-Glc as their substrate, conforming with previous research. Unlike others, 76E4 specifically did not show positive activity with UDP-Glc under existing conditions. This could be probably due to non-usage of kaempferol and/or quercetin as acceptor.

In this project, only donor sugars of UDP and GDP nucleoside bases were utilised in the screening. 7 of 8 UGTs studied recognised donor sugars of UDP nucleoside base (Fig 3.15). Our findings further corroborate the strict specificity of UGTs for the UDP base. UGT 76E4 which exhibited a GDP-fucose activity (with kaempferol acceptor substrate) at the time of writing has not been previously reported.
This is the only UGT in this study capable of using a donor sugar of GDP nucleoside base, representing only 12.5% of screened UGTs (Fig 3.15). While the reason(s) for 76E4 GDP recognition is not yet clear, it is believed that molecular interactions from structural studies would shed more light on this.

As implied by the name, UGTs are generally highly specific for donor sugars with the nucleoside base UDP although some do tolerate GDP and dTDP donor sugars. Compared to UDP, GDP has an additional imidazole ring and an NH$_2$ replacement at C2 while dTDP also has an extra CH$_3$ group on uracil ring C5 and an absent -OH on ribose C2 as shown in Fig 3.16.

A recently reported thirteen (13) donor-sugar library screening containing all the nucleoside bases over a fifty-four (54) UGT enzyme panel revealed an interaction pattern therein. 100% UDP, 25% dTDP and only 7.4% GDP utilisation recorded, highlighted how stringently regulated the nucleoside component of the donor sugar can be (Yang et al., 2019). Broader donor activity was found for 76E1 and 76E5 and this novel behaviour has been reported (Akere et al., 2018). Structurally, the flexibility of the enzymes’ active site can reflect high substrate recognition (Hedstrom., 2010). Hence both UGTs might be the most capable of tolerating more donor compounds due to flexibility in the C-terminal domain coupled with other unexplained factors. In addition, the structural differences of the donor sugars may affect the binding site regions of the enzyme.

Fig 3.15: Pie chart indicating the distribution of nucleoside base recognition across screened UGTs
UGTs 76E1, 76E5 and 84A2 showed UDP-Gal activity. Structurally, UDP-Gal only slightly differs from UDP-Glc, mainly with the OH group at the C4 orientating at a different direction (Fig 3.17). Sensitivity to this change in C4 OH’s orientation by the binding site residues may influence an enzyme’s specificity for this sugar donor. 76E2, 76D1, 84A1, 84A3 and 76E4 non-activity with UDP-Gal might be associated with their non-tolerance of orientation change of the OH group at the C4 position. On the other hand, the active sites of 76E1, 76E5 and 84A2 could be able to accommodate the change.

Specificity for UDP-Gal has been unclear due to insufficient data to determine which plant UGT will accommodate the donor sugar. A fine interplay of stereochemistry and conformation of the donor sugars may be likely involved in UGT sugar specificity (Offen et al., 2006; Davies et al., 2005). With an N-acetyl group on the C2 instead of an -OH, UDP-GlcNAc is bigger in size than UDP-Glc. As the functional group is larger and more polar, it often makes substantial steric barrier compared to Glc (Zhang 2019). Therefore, majority of UGTs is unable to accommodate UDP-GlcNAc. Our findings indicate that only 76E1 and 76E5 exhibited UDP-GlcNAc activity, suggesting that their active sites can tolerate the donor sugar’s C2 positioned N-acetyl group.

All mass spectra of positive results for the acceptor and donor screening are shown in appendix figures i-xl.


3.3 Kinetics

Some functional and biochemical activities of group L members have been previously reported (Meißner et al., 2008; Brazier-Hicks et al., 2007). As group H is less studied than L, further studies in this section and the rest of the thesis chapters will focus on this group. This indicates the novelty of the kinetics findings as reported here.

3.3.1 Analysis of kinetic data

Km (Michaelis constant) is the substrate concentration at which half of the active sites are occupied. It reveals the substrate concentration required for significant catalysis to occur. Also, it indicates the affinity between substrate and enzyme. Kcat, referred to as the turnover number is the number of substrate molecule each enzyme site converts to product per unit time while kcat/Km is the catalytic efficiency, a useful index for measuring of enzyme performance. The higher the kcat/Km value, the more efficient an enzyme is. Table 3.2 shows details of 76E1 kinetics with UDP-Glc. Kinetic parameters were analysed by OriginPro™ (a graphing and analysis) software.
UGT 76E1 UDP-Glc

Michaelis menten plot

<table>
<thead>
<tr>
<th>S (µM)</th>
<th>V (µM/min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>0.0032</td>
</tr>
<tr>
<td>20</td>
<td>0.0042</td>
</tr>
<tr>
<td>50</td>
<td>0.0058</td>
</tr>
<tr>
<td>100</td>
<td>0.006</td>
</tr>
<tr>
<td>200</td>
<td>0.0064</td>
</tr>
<tr>
<td>500</td>
<td>0.0065</td>
</tr>
</tbody>
</table>

Lineweaver Burk plot

<table>
<thead>
<tr>
<th>1/S</th>
<th>1/V</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>312.5</td>
</tr>
<tr>
<td>0.05</td>
<td>238.0952</td>
</tr>
<tr>
<td>0.02</td>
<td>172.4138</td>
</tr>
<tr>
<td>0.01</td>
<td>166.6667</td>
</tr>
<tr>
<td>0.005</td>
<td>156.25</td>
</tr>
<tr>
<td>0.002</td>
<td>153.8462</td>
</tr>
</tbody>
</table>

Km 10.5±1.38µM

kcat/KM 0.79 s⁻¹mM⁻¹

kcat 0.00832 s⁻¹

Table 3.2: Michaelis Menten and Lineweaver-Burk plots for 76E1

The remaining sets of data (Michaelis-menten and Lineweaver Burk plots) for 76E2, 76E5 and 76D1 are shown in appendix (Table i-v). The kinetics parameters for these were also estimated using OriginPro™. On the pseudo-kinetic assumption, the kinetic behaviour of group L UGTs (except 76E4 which did not use UDP sugar) have been determined. The summary of all findings for all UGT kinetics is found in Table 3.3.
Table 3.3: Kinetic study of UGTs

As shown in the plots Michaelis–Menten and Lineweaver–Burk (Tables 3.3 – 3.8 Appendix) and summarised in Table 3.3 under the current condition, all the enzymes showed relatively low $K_M$ at μM scale. A low $K_M$ indicated that an enzyme requires a small amount of substrate to become saturated. Hence, the maximum velocity is reached at relatively low substrate concentrations kinetics data revealed relatively low $K_M$ values, suggesting enzymes have good affinity for the substrates UDP-Glc, UDP-Gal and UDP-GlcNAc. All 76E1, E2 and E5 enzymes share very similar $K_M$-UDPGlc values (~10 μM), which is a measure of enzymes’ substrate preference and catalytic efficiency. However, 76E5 showed strong preferences to both UDP-Gal and UDP-GlcNAc, with low sub μM $K_M$ and much higher $K_{cat}/K_M$ (>20 s$^{-1}$ mM$^{-1}$). This kinetic data strongly suggests that 76E5 is probably more of UDP-GlcNAc and UDP-Gal transferase than UDP-glucosyltransferase. 76D1 has much high $K_M$ values (>25 μM) and low $K_{cat}/K_M$ (0.185 s$^{-1}$ mM$^{-1}$).

76C1 and 76C2 are the only group H members previously characterised and are plant hormone GTs. The kinetic parameters here are with respect to the plant hormone substrates; hence unsuitable for comparison with our studied UGTs (Hou et al., 2004). There is no other kinetics data available for group H members for comparison. When compared with VvGT1 ($K_{cat}/K_M$ 0.124 s$^{-1}$ mM$^{-1}$) [Offen et al., 2006], our UGTs $K_{cat}/K_M$ values are in the same range for substrate UDPGlc. 76E5 has a far higher $K_{cat}/K_M$ with substrates UDP Gal and UDP-GlcNAc (>2000 fold) than VvGT1; affirming that UDP Gal and UDP-GlcNAc seems to be natural substrates to 76E5. Functional studies for 76E5 in vivo will reveal more in the near future (Akere et al., 2018). Additionally, the UDP-Gal and UDP-GlcNAc kinetic activity of 76E1 could not be obtained for this work.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Substrate varied</th>
<th>$K_M$ (μM)</th>
<th>$K_{cat}$ (s$^{-1}$)</th>
<th>$K_{cat}/K_M$ (s$^{-1}$ mM$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>76E1</td>
<td>UDP-Glc</td>
<td>10.5 ± 1.38</td>
<td>0.00832</td>
<td>0.79</td>
</tr>
<tr>
<td>76E2</td>
<td>UDP-Glc</td>
<td>8.75 ± 1.38</td>
<td>0.0107</td>
<td>1.23</td>
</tr>
<tr>
<td>76E5</td>
<td>UDP-Glc</td>
<td>8.79 ± 1.64</td>
<td>0.0194</td>
<td>2.21</td>
</tr>
<tr>
<td></td>
<td>UDP-Gal</td>
<td>0.57 ± 0.19</td>
<td>0.0130</td>
<td>22.8</td>
</tr>
<tr>
<td></td>
<td>UDP-GlcNAc</td>
<td>0.78 ± 0.28</td>
<td>0.0166</td>
<td>21.3</td>
</tr>
<tr>
<td>76D1</td>
<td>UDP-Glc</td>
<td>25.77 ± 8.82</td>
<td>0.00476</td>
<td>0.185</td>
</tr>
</tbody>
</table>
3.4 Conclusions

In this chapter, eight UGTs (76E1, 76E2, 76E4, 76E5, 76D1 – group H and 84A1, 84A2 and 84A3 – group L) have been studied for their choices of substrates. For the acceptor compounds screened, 76E1 recognised the most compounds within its group while 76E4 showed the least recognition. Within this group, flavonoids were mainly used suggesting their potential in flavonoid glycosides biosynthesis. Many flavonoid glycosides have been studied and have shown immense pharmacological benefits such as kaempferol-3-O glucoside and quercetin-3-O-glucoside with antioxidant and neuroprotective activities respectively (Yu et al., 2013; Olthof et al., 2000). 76E1 therefore, stands out in this group with broader acceptor substrate specificity. Group L UGTs displayed wider recognition with not only flavonoids but alkaloids and stilbenes. 84A2 recognised more acceptor compounds within group L. Generally, most of the UGTs studied in this acceptor screening have shown broad specificity to natural products, a potential which may be explored in the synthesis of bioactive glycosides.

Since the type of sugar added to these natural compounds often affects their pharmacological properties (Olthof et al., 2000), all eight UGTs were screened for their donor sugar preference. All UGTs but one show activity with UDP-Glc, confirming the donor sugar as the most commonly recognised among UGTs. Group H UGTs 76E1 AND 76E5 displayed activity with the most donor sugar screened (about 43%) showing broader donor substrate specificity. This is important as more varieties of glycosides can be made using the UGTs studied. 76E4 showed an unusual activity with GDP-Fucose, a donor sugar rarely recognised by plant UGTs. The substrate screening results (acceptor and donor) for these eight UGTs have laid bare glycosylation potential of UGTs towards specific small molecules and will serve as a foundation for further applications.

Group H UGTs 76E1, 76E2, 76E5 and 76D1 have low Km values indicating great affinities for the donor sugar. This means little amount of donor sugar substrate is required to reach maximum velocity. Since a major constraint to the use of in vitro glycosylation system in the synthesis of glycosides is the expensive cost of UDP-sugars (Ruffing and Chen, 2006), the effective use of donor sugars in small amount by these UGTs may strengthen their potential usage. 76E5 showed higher catalytic efficiencies for UDP-Gal and UDP-GlcNAc indicating its preference for these donor sugars over UDP-Glc. This is quite interesting as many plant UGTs reported prior show preference for UDP-Glc over other donor sugars. Understanding of in vitro kinetics of UGT is significant in shedding light on their catalytic process in vivo.

The mechanism dictating sugar donor as well as acceptor specificity and activity is multifaceted. Better understanding of the determinants of donor sugar specificity will become clearer as more crystal structures of plant UGTs of broader sugar recognition are solved. This would in turn advance and simplify structure prediction using homology modeling.
Inferences based on structural studies buttressed by biochemical studies of the substrate specificity of wild type (as well as) mutated proteins will greatly contribute to our general understanding of substrate specificities of plant UGTs (Osmani et al., 2009). UGTs with broad substrate specificity are potential biocatalysts in the synthesis of bioactive molecules. Useful insights to help understanding the relationship between function (findings from biochemical characterisation) and structure can be derived from UGTs structure (including amino acid sequence). Furthermore, a thorough look at plant UGTs capable of recognising diverse donor sugars of UDP, GDP and dTDP base extraction such as UGT89C1 (Parajuli et al., 2016) could beam a searchlight into understanding this difference. This UGT’s crystal structure in complex with UDP-Rha and acceptor quercetin has fortunately been recently solved (Zong et al., 2019). This will be laid bare as the next chapter offers possible explanations for the different UGTs behaviour with regards to substrate specificity.
References


Chapter 4 Homology modeling and docking studies of UGTs
This chapter describes structural modeling of the studied UGTs in the absence of experimental crystal structures to further understand their substrate specificities. Employing homology modeling and docking, interactions between substrate ligands and UGTs were studied. More importantly, key amino acids involved and driving the specificities were explored across the UGTs. These residues were compared with those previously reported in literature among plant UGTs crystal structures. Structural differences observed among the UGTs were utilised to explain behaviour displayed during substrate screening.
4.1 Homology modeling of UGTs

4.1.1 Model construction

Homology modeling depends on an evolutionary relationship between protein sequences of interest and other proteins in the same family with experimentally solved structures (Osmani et al., 2009). It is therefore important to search for such experimentally solved plant UGTs in the protein data bank (PDB). A Blast search through the protein data bank (PDB) for potential templates revealed *Medicago truncatula* UGT85H2, *Arabidopsis thaliana* UGT74F2 and *Vitis vinifera* VvGT1 for UGTs (76E1, 76E2, 76E4, 76E5 and 76D1). The sequence identities range from 28-33%, 25-31% and 28-30% for 85H2, 74F2 and VvGT1 respectively (Table 1). *AtUGT74F2* was chosen as a template for this work over *MtUGT85H2* (with slightly higher sequence identity) due to the presence of substrate ligand in its structure. The absence of conformational changes in unbound protein structure (which occur in ligand-bound protein structures) preclude it from being used as a basis for structure-based drug design. The availability of ligand-bound protein structure is a prerequisite currently for a successful docking (Seeliger and de Groot, 2010). In addition, *AtUGT74F2* belong to *Arabidopsis thaliana* and of phylogenetic group L whose class members recognise flavonoids as substrates (Fig 1.16 section 1.2.3).

<table>
<thead>
<tr>
<th>Plant UGTs</th>
<th>Template PDB</th>
<th>% Identity to studied UGTs</th>
</tr>
</thead>
<tbody>
<tr>
<td>AtUGT74F2</td>
<td>5V2K</td>
<td>25 – 31</td>
</tr>
<tr>
<td>MtUGT85H2</td>
<td>2PQ6</td>
<td>28 - 33</td>
</tr>
<tr>
<td>VvGT1</td>
<td>2C1X</td>
<td>28 - 30</td>
</tr>
</tbody>
</table>

Table 4.1: Potential template UGTs identified using BLAST their and sequence identities to studied UGTs

4.1.2 Secondary structure; model-template comparison

*AtUGT74F2* structure has 444 amino acids residues (original sequence has 449 amino acids). The first three and last two residues are missing in the crystal structure. The N-terminal domain runs from residues 4-245 while the C-terminal domain runs from 246-447. It exhibits the GT-B fold and the PSPG motif is from 324-367. The C-terminal domain has seven α-helices and six β-sheets while the N-terminal domain has eight α-helices and seven β-sheets. The interdomain linker loop runs from 232- 250 (George-Thompson et al., 2017).
Table 4.2: Sequence identity of studied UGTs to selected template UGT 74F2

<table>
<thead>
<tr>
<th>Model UGTs</th>
<th>% identity to template AtUGT74F2</th>
<th>RMSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>76E1</td>
<td>30</td>
<td>0.018</td>
</tr>
<tr>
<td>76E2</td>
<td>31</td>
<td>0.019</td>
</tr>
<tr>
<td>76E4</td>
<td>29</td>
<td>0.015</td>
</tr>
<tr>
<td>76E5</td>
<td>29</td>
<td>0.006</td>
</tr>
<tr>
<td>76D1</td>
<td>25</td>
<td>0.010</td>
</tr>
</tbody>
</table>

Table 4.2 shows how similar the five studied UGTs sequences are to template 74F2. The structure prediction run on a structural visualisation program, ICM Pro Browser showed that the UGTs consists of α-helix and β-strands as shown in Table 4.3.

Table 4.3: Secondary structure comparison between template and model UGTs structures

<table>
<thead>
<tr>
<th>Model UGT</th>
<th>N-terminal</th>
<th>C-terminal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>α-helix</td>
<td>β-sheet</td>
</tr>
<tr>
<td>74F2 (template)</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>76E1</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>76E2</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>76E4</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>76E5</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>76D1</td>
<td>8</td>
<td>7</td>
</tr>
</tbody>
</table>

Residues (6-448) 76E1, (7-443) 76E2, (6-443) 76E4, (6-442) 76E5 and (5-448) 76D1 were modeled in the structures. This is similar to as found in the template structure 74F2 where the first few residues are missing (residues 4-447 present). The N-terminal domain is predicted to fold into a seven-stranded parallel β-sheet (labelled N β1–Nβ7 on Fig 4.2) surrounded by seven α-helices (labelled Na1, Na2, Na3, Na4, Na5, Na5a and Na6) except in 76E5 where two α-helices are located between β3 and β4. The C-terminal domain folds into a six-stranded parallel β-sheet (labelled Cβ1 – Cβ6 on Fig 4.2) surrounded by seven α-helices (labelled Ca0, Ca1, Ca2, Ca3, Ca4, Ca5, Ca6 and Ca7) and an eighth (Ca8) folding back into N-terminal domain (except in 76E2 and 76E4). This general representation is shown in Fig 4.1. 76E5 and 76D1 however, have Ca2 absent (Fig 4.2). Generally, these secondary structures show similarity to solved structure template AtUGT74F2 (George-Thompson et al., 2017).
Less conserved loop regions link the conserved β-sheets and α-helices which display some diversity across our UGT models. Except for N-terminal N5a and N5b loops, the loops have been labelled according to the previous β-strand as shown in Fig 4.2. These loops differ in both length and amino acid composition. Differences predicted in the N-terminal loops include that in N3 (76E5 has a small α-helix in between), N5a (76E1 is a longer than others) and N5 (a conserved threonine reportedly involved in donor binding is a proline in 76D1). Also, loop N2 is predicted to be completely absent in all models. The absence of N2 loop was similarly reported in VvGT1 and MtUGT85H2 (Offen et al., 2006; Li et al., 2007).

Fig 4.1: Ribbon diagram of the model 76E5 secondary structures with bound UDP-glucose
Fig 4.2: Predicted secondary structures of our five model UGTs, α-helices in turquoise and β-strands in blue.

4.1.3 Model validation

An assessment of the modelled protein structures is vital to highlight the overall quality and identify regions that may require further careful investigation. PROCHECK can verify the details of the stereochemistry of structures, both experimentally solved and models. The Ramachandran plot is the first of the plots produced by PROCHECK, which highlights regions in the 'disallowed' regions (Laskowski, 1993).

The plots for each of our models are shown in Fig 4.4. The red regions in the graph represent the most allowed regions, while additional allowed, generous allowed and disallowed regions are indicated as yellow, light yellow and white fields respectively. Model 76E2 has the least number of residues in disallowed region (2) while model 76D1 have the most (9) as shown in Table 4.4. Residues in the most favoured regions are between 85-90% in all models which confirms the overall good quality of the homology models.

<table>
<thead>
<tr>
<th>Model</th>
<th>Residues in disallowed region (Ramachandran plots)</th>
</tr>
</thead>
<tbody>
<tr>
<td>76E1</td>
<td>S33, T181, N292, R289, L313, W309</td>
</tr>
<tr>
<td>76E2</td>
<td>L161, D71</td>
</tr>
<tr>
<td>76E4</td>
<td>K160, V231, K178, Q71, R398, V306,</td>
</tr>
<tr>
<td>76E5</td>
<td>Q70, R389, F244</td>
</tr>
<tr>
<td>76D1</td>
<td>K172, D69, S66, F176, Q6, R394, D370, Q287, N440</td>
</tr>
</tbody>
</table>

Table 4.4: Residues in disallowed regions as indicated by Ramachandran plots
The structural superimposition of template on each of the models is shown in Fig 4.3.

![Superimposition of 3D structures showing similarity](image)

**Fig 4.3:** Superimposition of 3D structures showing similarity a) All models and template 74F2 b) Each model with template 74F2 (Colour code: 74F2-purple, 76E1-blue, 76E2-yellow, 76E4-pink, 76E5-brown and 76D1-grey)

In addition to the Ramachandran plots, PROSA validated the quality of the homology model structures (Z-scores) by assessing overall model quality of Cα positions (Wiederstein and Sippl., 2007). Z-score of models can be then be compared with range of scores typical of native proteins of similar sizes. The Z-scores of models and that of nine available plant UGTs crystal structures when compared, the values indicate a quite close range. The plots also showed that the models’ Z-scores fall within the range of PDB-existing structures Z-scores (Appendix). This further highlights the good quality of the homology models. Fig 4.4 shows the Ramachandran plot as well as Z-score plot for model 76E2.
Fig 4.4: PROCHECK Ramachandran plots for model 76E2 and its Z-score is shown here. Only two residues lie in the disallowed region. The black dot in plot represents each model's Z-score in a plot containing the Z-scores of all experimentally determined protein structures currently in the PDB. Different sources of PDB structures (i.e X-ray and NMR) are distinguished by different shades of blue.
4.2 Docking studies and molecular interactions; Donor sugar interactions

In the absence of crystal structures of UGTs, the interactions between the UDP donor sugar and UGTs was investigated via automated ligand docking.

4.2.1 Key amino acids identification – Sugar moiety

As implied in the name, UDP donor sugar is typically composed of nucleobase uracil, pentose sugar ribose and pyrophosphate - the three forming UDP (Fig 4.5). Additionally, a simple sugar is also present at the end. The simple sugar typically has six carbon atoms present.

![Sugar components](image)

Fig 4.5: Structural representation of components of a typical UDP donor sugar (UDP-Glucose structure is shown here).

The main amino acids in the UGTs interacting with the OHs at C2, C3, C4 and C6 positions of the sugar donor were studied via automated ligand docking. The sugar donor represented here is UDP-Glc since this is the most commonly used donor sugar among the UGTs according to the screening results (chapter 3). A comparison of the interacting amino acids with that of solved crystal structures will follow in the next sections.

4.2.1.1 C2 and C3 positions in solved crystal structures

PSPG motif residues have been reported to bind and form hydrogen bonds with the -OH at the C2 and C3 positions of the Glc moiety. The residue at position 44 in the PSPG motif is a conserved glutamine which binds at the C2 and/or C3 position in the available crystal structures of UGTs (Table 4.5). This is one of the residues that have been consistently reported to be involved in donor specificity and recognition (Shao et al., 2005). Additionally, it may also be suggestive of preference to use UDP-Glc specifically (Wetterhorn et al., 2016).
<table>
<thead>
<tr>
<th>Plant UGT</th>
<th>Result origin</th>
<th>Amino acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>UGT 74F2</td>
<td>Docking of Glc moiety</td>
<td>Q367</td>
</tr>
<tr>
<td>UGT 78G1</td>
<td>Docking of Glc moiety</td>
<td>Q377</td>
</tr>
<tr>
<td>VvGT1</td>
<td>PDB ID: 2C1Z</td>
<td>Q375</td>
</tr>
<tr>
<td>UGT 71G1</td>
<td>PDB ID: 2ACW</td>
<td>Q382</td>
</tr>
<tr>
<td>UGT 72B1</td>
<td>PDB ID: 2VCE</td>
<td>Q389</td>
</tr>
<tr>
<td>UGT 78K6</td>
<td>Overlay on VvGT1</td>
<td>Q368</td>
</tr>
<tr>
<td>UGT 85H2</td>
<td>Docking of Glc moiety</td>
<td>Q403</td>
</tr>
<tr>
<td>Os79</td>
<td>PDB ID: 5TMD</td>
<td>Q386</td>
</tr>
<tr>
<td>PtigS</td>
<td>Docking of Glc moiety</td>
<td>Q395</td>
</tr>
</tbody>
</table>

Table 4.5: Table showing positions of conserved glutamine Q in all plant UGTs crystal structures

4.2.1.2 C2 and C3 positions in studied models

As found in plant UGTs crystal structures, the residue at position 44 in the PSPG motif was also conserved glutamine in the five studied model UGTs. The conserved glutamine is present at the following positions as shown in Fig 4.6.

Table 4.6: Figure showing positions of conserved glutamine (in red box) in studied UGTs.

The formation of potential hydrogen bonds at C2 and C3 at these positions in each of the models is indicated in Fig 4.6. While the glutamine interacts with both C2 and C3 -OH group in 76E5, 76E4 and 76D1, it only interacts with C3 –OH group in 76E2 and 76E1. Few mutations involving glutamine in UGTs 76E5, 76E1 and 76E4 were carried out to ascertain the importance of this residue on donor substrate recognition. This will be reported in detail in Chapter 5.
Fig 4.6: Hydrogen bonding by conserved glutamine in all model UGTs (as labelled) to the C2 and/or C3 positions in UDP-Glc

4.2.1.3 C3 and C4 positions in solved crystal structures

The residue at position 43 in the PSPG motif binds to the -OH groups at the C3 and C4 positions of Glc in many previously solved plant UGTs structures. The amino acid is usually conserved as aspartic acid/glutamic acid as seen in Table 4.6. Both amino acids aspartic acid and glutamic acid are negatively charged and similar except that E has an extra CH2 group in its side chain. Through hydrogen bond formation, this negatively charged aspartic acid/glutamic acid interacts with the sugar donor. Alongside with Glutamine (section 4.2.1.2), it has been dubbed as the vital player in sugar recognition (Offen et al., 2006).
### Table 4.6: Table showing positions of conserved aspartic acid/glutamic acid (D/E) in all plant UGTs crystal structures

<table>
<thead>
<tr>
<th>Plant UGT</th>
<th>Result origin</th>
<th>Amino acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>UGT 74F2</td>
<td>Docking of Glc moiety</td>
<td>D366</td>
</tr>
<tr>
<td>UGT 78G1</td>
<td>Docking of Glc moiety</td>
<td>D376</td>
</tr>
<tr>
<td>VvGT1</td>
<td>PDB ID: 2C1Z</td>
<td>D374</td>
</tr>
<tr>
<td>UGT 71G1</td>
<td>PDB ID: 2ACW</td>
<td>E381</td>
</tr>
<tr>
<td>UGT 72B1</td>
<td>PDB ID: 2VCE</td>
<td>E388</td>
</tr>
<tr>
<td>UGT 78K6</td>
<td>Overlay on VvGT1</td>
<td>D367</td>
</tr>
<tr>
<td>UGT 85H2</td>
<td>Docking of Glc moiety</td>
<td>D402</td>
</tr>
<tr>
<td>Os79</td>
<td>PDB ID: 5TMD</td>
<td>D385</td>
</tr>
<tr>
<td>PtigS</td>
<td>Docking of Glc moiety</td>
<td>E394</td>
</tr>
</tbody>
</table>

#### 4.2.1.4 C3 and C4 positions in studied models

In the five (5) models, the conserved amino acid is present at the following positions [Fig 4.7] where amino acid present as glutamic acid in 76E4 and 76E5 and in others, as aspartic acid.

![Multiple sequence alignment showing positions of conserved aspartic acid/glutamic acid in studied UGTs](image)

**Fig 4.7: Multiple sequence alignment showing positions of conserved aspartic acid/glutamic acid in studied UGTs**

The formation of potential hydrogen bonds with hydroxyl groups at C3 and C4 in each of the models are illustrated in Fig 4.8. In 76E4 and 76E5, the interaction with C3 –OH group shows weaker hydrogen bonds with distances of 3.68 and 3.81Å respectively while 76E1 D372, 76E2 D374 and 76D1 D370 showed strong to moderate bonding with distances 2.65, 2.55 and 2.36Å respectively. At the C4 position, 76E4 E374 and 76E5 E365 show moderate hydrogen bonds like others. Jeffrey (1997) classified hbonds with donor-acceptor distances of 2.2-2.5 Å as "strong, mostly covalent", 2.5-3.2 Å as "moderate, mostly electrostatic", and 3.2-4.0 Å as "weak, electrostatic".
Fig 4.8: Hydrogen bonding by conserved aspartic/glutamic acid in studied UGTs to the hydroxyl groups at C3 and C4 positions in UDP-Glucose

Mutation of residue D374 to A (alanine) in VvGT1 completely abolished detectable catalytic activity, affirming the critical presence of this amino acid at this position in donor substrate studies (Offen et al., 2006). Since Glutamic acid has a longer side chain and an extra CH₂, the distance to the substrate might have an impact on the activity. The effect of Asp/Glu swap on donor sugar recognition was carried out on some UGTs via mutations (Zhang, 2018). Mutations involving this amino acid and interactions with C3 and/or C4 position in model UGTs 76E2 and 76E4 will be presented in Chapter 5 to shed light on the significance of the interactions with these positions on donor sugar recognition.
4.2.1.5 C6 positions in solved crystal structures

Amino acids interacting with the C6 –OH group of the sugar molecule varies in the available plant UGT crystal structures. This is because often, N-terminal residues (which are more variable and less conserved) are involved. In 71G1, 78G1 and VvGT1, T143, T141 and T141 were observed to interact with the C6 position respectively (Shao et al. 2015; Modolo et al., 2009; Offen et al., 2006). In 78K6, N137 is present while Os79 has S142 and Q143 (Hiromoto et al., 2015; Wetterhorn et al., 2016). No amino acids made interactions with the C6 position in some crystal structures such as 85H2 and 72B1. The presence and absence of interaction in VvGT1 and 85H2 is shown in Fig 4.9. Compared to amino acids interacting at C2/C3 and C3/C4 positions (which are more conserved), a more diverse residues interact at the C6 position (Fig 4.10).

![Fig 4.9: The presence and absence of interaction at C6 position of VvGT1 and 72B1 respectively depicting less conservation of interaction at that position](image)

The interactions with the C6 position has been described to be of less importance in catalysis. When compared to C2/C3 and C3/C4 positions, mutation T141A in VvGT1 reduces the activity only six-fold (Offen et al., 2006). In contrast, a total loss of UDP-Glc activity of 76E1 T134A. This on the contrary highlights a significant role of T141 in this UGT (Majeed et al., 2015). In addition, the presence of residues other than threonine suggest the changing nature of amino acid bound at C6 position (Fig 4.10).

![Fig 4.10: Conservation of amino acid interacting at C6 position across plant UGTs. The amino acid residues here are less conserved.](image)
4.2.1.6 C6 positions in studied models

In modeled UGTs 76E1, 76E2, 76E4 and 76E5, a conserved Threonine interacts with the C6 position. This is like as observed in 74F2, 71G1, VvGT1 and 78G1. However, only 76D1 differs, having a proline present at this position (Fig 4.11) and is similar to 72B1, 78K6 and PtigS (Fig 4.12). Structurally, proline side chain is non-reactive. Threonine’s side chain is polar and can form hydrogen bonds with substrates (Betts and Russell, 2003).

Fig 4.11: Multiple sequence alignment showing positions of residues interacting at C6 position in model UGTs; This residue is conserved as threonine in 76E1, 76E2, 76E4 and 76E5 while 76D1 has a proline present.

Docking studies to investigate residues interacting with the hydroxyl group at C6 position of sugar donor across the five model UGTs in shown in Fig 4.12.

Just as observed in 72B1 and 78K6, proline is present at the equivalent position in 76D1. Proline barely forms hydrogen bonds directly to a substrate (Holliday et al., 2007). As predicted, P129 in 76D1 does not make interactions with the –OH group at C6 position of UDP-Glc while in others, hydrogen bonds were established (Fig 4.14). Previous studies have pointed to the involvement of Threonine at this position in binding and recognition of UDP-Glc (Shao et al. 2015; Modolo et al., 2009; Offen et al., 2006). Hence, no UDP-Glc activity should be expected in 76D1 with a Proline at this position. However, 76D1 displayed UDP-Glc activity even with Proline at the position, in fact being the only donor recognised out of seven donor compounds screened (Chapter 3; Akere et al., 2018). This may suggest that in this UGT, other residues than threonine may have controlled UDP-Glc activity. The effect of substituting Proline in 76D1 will shed more light on its involvement on donor recognition. This will be reported in Chapter 5.

On the other hand, mutation of T134A completely abolished UDP-Glc activity in 76E1. This assumed that the interaction between the T134 and Glc-O6 was disrupted (Majeed et al., 2015).
4.2.2 Key amino acids identification – phosphate moiety

A typical component of nucleotide sugar donors is the phosphate moiety and as such, is of great importance. Amino acids interacting with the phosphate moiety are anticipated to be highly conserved across UGTs. Serine, asparagine and histidine are typically involved in interactions with the phosphate group. For example, in 74F2, the sides chains of S347, N346, S273 and H342 binds to the phosphate (George-Thompson et al., 2017). Table 4.6 shows the amino acids binding the phosphate moiety in plant UGT structures.
The residue interacting with phosphate moiety presents as either a serine/threonine. S273 in 74F2 is conserved in its interaction with the phosphate. At the equivalent position, threonine is present in the crystal structures of VvGT1, 78K6 and Os79 (Table 4.7). Threonine and serine both have polar side chains with the only difference being an added CH₃ in threonine.

<table>
<thead>
<tr>
<th>Plant UGT</th>
<th>Amino acids</th>
<th>Origin of result</th>
</tr>
</thead>
<tbody>
<tr>
<td>74F2</td>
<td>S273, H342, N346, S347</td>
<td>George-Thompson et al., 2017</td>
</tr>
<tr>
<td>78G1</td>
<td>H352, N356, S357</td>
<td>Modolo et al., 2009</td>
</tr>
<tr>
<td>71G1</td>
<td>S285, H357, N361, S362, Y379</td>
<td>Shao et al., 2005</td>
</tr>
<tr>
<td>VvGT1</td>
<td>T19, T280, H350</td>
<td>Offen et al., 2006</td>
</tr>
<tr>
<td>85H2</td>
<td>S304, N382, S383, H378,</td>
<td>Li et al., 2007</td>
</tr>
<tr>
<td>78K6</td>
<td>S16, T273, H343, S348</td>
<td>Hiromoto et al., 2015</td>
</tr>
<tr>
<td>Os79</td>
<td>T291, H361, N365, S366</td>
<td>Wetterhorn et al., 2016</td>
</tr>
<tr>
<td>72B1</td>
<td>S277, H364</td>
<td>Brazier-Hicks et al., 2007</td>
</tr>
<tr>
<td>PtigS</td>
<td>S282, H370</td>
<td>Hsu et al., 2018</td>
</tr>
</tbody>
</table>

Table 4.7: Table showing positions of residues interacting with the phosphate moiety in all plant UGTs crystal structures

Other interactions of residues such as Histidine, Asparagine and Serine (residues at positions 19, 23 and 24 in the PSPG motif) are also found in phosphate. These are highly conserved (George-Thompson et al., 2017, Modolo et al., 2009, Shao et al., 2005; Li et al., 2007; Wetterhorn et al., 2016). Some less conserved N-terminal residues interactions to the phosphate moiety such as (T19 and S16 of VvGT1 and 78K6) have also been reported (Offen et al., 2006; Hiromoto et al., 2015). Fig 4.13 compares conservation among all the amino acids interacting with donor sugar phosphate moiety across plant UGTs.

Fig 4.13: Comparison of conservation of amino acids binding to the phosphate moiety
Studies have shown the significance of the amino acid residues binding to the phosphate on the donor substrate recognition. In 78D2, T22A mutant completely lost UDP-Gal and GDP-Glc activities while only retaining 20% and 5% of UDP-Glc and UDP-GlcNAc activity respectively (Zhang et al., 2018).

This may suggest that T22 here was involved in direct interactions with phosphate and hence impacting substrate recognition. Furthermore, mutation of T286A in 78D2 resulted in 85% loss of UDP-Glc activity while S270A in 73B4 showed total loss of UDP-GlcNAc activity. Mutant 73B4 H349A lost UDP-Glc and UDP-GlcNAc activity whereas 78D2 H356A only lost its UDP-GlcNAc activity (Zhang et al., 2018). The significant activity loss following these mutations was attributed to impairment of the interactions between the enzyme and the phosphate. This highlights the importance of the residues binding to the sugar phosphate region and their roles in sugar donor recognition (Zhang et al., 2018).

Using data from docking studies, interactions predicted for the target UGTs and the sugar donor phosphate are quite similar to those observed in the crystal structures. The amino acids predicted to be involved in interactions to the phosphate moiety in our studied models are shown in Table 4.8.

<table>
<thead>
<tr>
<th>Plant UGT</th>
<th>Amino acids</th>
</tr>
</thead>
<tbody>
<tr>
<td>76E1</td>
<td>S272, H348, N352, S353</td>
</tr>
<tr>
<td>76E2</td>
<td>S274, H350, N354, S355</td>
</tr>
<tr>
<td>76E4</td>
<td>T274, H350, N354, S355</td>
</tr>
<tr>
<td>76E5</td>
<td>S272, H341, N345, S346</td>
</tr>
<tr>
<td>76D1</td>
<td>S269, H346, N350, S351</td>
</tr>
</tbody>
</table>

Table 4.8: Table showing amino acids binding to the phosphate moiety in studied UGTs

Residue conservation check around these interacting amino acids is shown in Fig 4.14. Among these five UGTs, UGT 76E4 has a distinctly different polar residue, lysine next to the interacting T274. Other model UGTs however, have hydrophobic leucine/isoleucine which is conserved.

![Fig 4.14: Amino acids interacting with the donor sugar phosphate moiety in model UGTs](image)
In addition to being next to the interacting T274, K275 is a C1 loop residue. C1 loop residues offer critical interactions with the donor sugar (Chapter 1, section 1.2.3.4; Fig 1.11). Therefore, the significance of this residue difference in UGT 76E4 is worth investigating as it is the only UGT with the least glycosylation activity to UDP-sugars as reported in chapter 3.

As highlighted in Table 4.7, a pattern of interaction similar to those of crystal structures have been observed in all five model UGTs. These are shown in Fig 4.15.

![Fig 4.15: Hydrogen bonding by conserved residues histidine, serines and asparagine in model UGTs to the phosphate region of UDP-Glucose](image-url)
4.2.3 Key amino acids identification – uridine moiety

Just like the phosphate moiety, the uridine moiety is a constant part of the nucleotide sugar. The uridine moiety in UDP sugar donors is expected to exhibit a conserved interaction pattern. Generally, in all solved plant UGT crystal structures, the uridine moiety consistently interacts with two PSPG residues (Table 4.7). These are tryptophan and glutamic acid (the first and 27th PSPG residue respectively).

Both residues interact via π-π (pi-pi) stacking interaction and hydrogen bonding respectively. Tryptophan forms a hydrophobic platform upon which the uracil base stacks and makes π-π stacking interactions. This is seen in VvGT1, 85H2 and 71G1 (Offen et al., 2006; Li et al., 2007; Shao et al., 2005). Additionally, alanine/cysteine and glutamine (the 2nd and 4th PSPG residue) may also form hydrogen bonds with the uracil ring as found in 78G1, 71G1, 85H2 and Os79 although this is less conserved (Table 4.9).

<table>
<thead>
<tr>
<th>Plant UGT</th>
<th>Amino acid</th>
<th>Origin of result</th>
</tr>
</thead>
<tbody>
<tr>
<td>74F2</td>
<td>W324 E350</td>
<td>George-Thompson et al., 2017</td>
</tr>
<tr>
<td>78G1</td>
<td>W334 E360 A335 Q337</td>
<td>Modolo et al., 2009</td>
</tr>
<tr>
<td>71G1</td>
<td>W339 E365 A340 Q342</td>
<td>Shao et al., 2005</td>
</tr>
<tr>
<td>VvGT1</td>
<td>W332 E358 Q335</td>
<td>Offen et al., 2006</td>
</tr>
<tr>
<td>85H2</td>
<td>W360 E386 C361 Q363</td>
<td>Li et al., 2007</td>
</tr>
<tr>
<td>78K6</td>
<td>W325 E351</td>
<td>Hiromoto et al., 2015</td>
</tr>
<tr>
<td>Os79</td>
<td>F343 E369 C344</td>
<td>Wetterhorn et al., 2016</td>
</tr>
<tr>
<td>72B1</td>
<td>W346 E372</td>
<td>Brazier-Hicks et al., 2007</td>
</tr>
<tr>
<td>PtigS</td>
<td>W352 E378</td>
<td>Hsu et al., 2018</td>
</tr>
</tbody>
</table>

Table 4.9: Table showing residues interacting with the uridine moiety in all plant UGTs crystal structures

Glutamine may be involved in aiding the maintenance of the hydrogen bonding interactions formed by glutamic acid. For example, in VvGT1, Q335 helps E358 in sustaining correct binding orientation to ribose hydroxyls without any direct interaction with the sugar (Offen et al., 2006). Like Q335 in VvGT1, Q328 in 78K6 may also be involved in helping E351 bind to the hydroxyl groups in ribose (Hiromoto et al., 2015). The conservation of amino acids tryptophan and glutamic acid interacting with the uridine moiety across plant UGTs crystal structures is presented in Fig 4.1f. In Os79 however, tryptophan is replaced by phenylalanine, another aromatic residue that is capable of making π-π stacking interactions.
Part of the C3 loop /Cα3 region, constituting the first six amino acids of the PSPG motif (residues 332-337 as found in VvGT1) forms the hydrophobic platform for uracil base stacking. Similarly, in 71G1, the loop region amino acids 339-344 aided W339 in stacking uridine, thereby forming a hydrophobic platform (Shao et al., 2005). This important region for stacking in all five UGTs is represented by the residues shown in Fig 4.17.

Docking studies have predicted interactions with the uridine moiety for all five model UGTs as illustrated in Fig 4.18.
Fig 4.18: Interactions of tryptophan, alanine, glutamic acid and glutamine with the uridine moiety of UDP-glucose in model UGTs. Tryptophan stacks the uracil, alanine forms hydrogen bond with the uracil ring while glutamine and glutamic acid form hydrogen bonds with the ribose sugar.

Analysing the interactions (Fig 4.18), tryptophan may have facilitated the stabilization of the uracil ring of UDP-donor sugar by making π stacking interactions with its indole ring in all five model UGTs. In addition, structural flexibility upon binding to donor sugar with this conserved tryptophan in plant UGTs has been observed. The occurrence or absence of the donor sugar determines the positioning of tryptophan. Such flexibility is a feature of regions connected with substrate binding (Li et al., 2007).
About 60% of aromatic amino acid side chains (such as tryptophan, histidine, phenylalanine) are projected to be involved in π-stacking interactions in proteins (Burley and Petsko, 1985). Such stacking interaction confers stability in nucleobases interactions (Li et al., 2009). Alanine residues (76E1 A331, 76E2 A333, 76E5 A324, 76D1 A329 and 76E4 A333) formed hydrogen bonds (via main-chain N and C atoms) with the uracil ring in all model UGTs. Glutamine also appeared to be stabilising glutamic acids’ hydrogen bonding to ribose in all UGTs except 76E1 as observed in Fig 4.20. It is unclear why this interaction is absent in 76E1. The absence of glutamine interaction to the ribose was also reported for 74F2 (George-Thompson et al., 2007).

Mutations W331A and E357A in 73B4 significantly reduced/abolished UDP-Glc activity (Zhang et al., 2018). Likewise, in 71G1 where W339 stacks uridine with the aid of the loop region WAPQVE, E365P abolished all UDP-Glc activity (Shao et al., 2005). However, 78D2 W338A retained UDP-Glc activity. This may be due to the extreme stability of the stacking loop where the destruction of the π-π interaction is not sufficient to cause total activity loss (Zhang et al., 2018).

4.2.4 Acceptor compounds

As discussed in chapter 3, UGT76E1 glycosylated more acceptor compounds than other UGTs using UDP-Glc as donor (Akere et al., 2018). This section highlights major structural differences observed in model UGT76E1 which may explain its broader acceptor substrate activity.

The acceptor compound functional group, which is a hydroxyl group, must be correctly oriented for glycosylation to occur. This accepting functional group requires to be near the donor sugar C1 and the deprotonating amino acid in enzyme (which act as a general base enabling acceptor compound deprotonation). Histidine is commonly this general base in plant UGTs (Brazier-Hicks et al., 2007; Shao et al., 2005; Offen et al., 2006; Li et al., 2007). The histidine interacts by forming a hydrogen bond with an acidic aspartic acid (histidine-aspartic acid dyad) in the acceptor compound deprotonation (Shao et al., 2005; Offen et al., 2006). Both residues are crucial for UGTs enzymatic activity (Li et al., 2007) and well conserved in all solved plant UGTs as well as our five modelled UGTs. The conservation of both residues in the model structures are shown in Fig 4.19.

![Fig 4.19: Multiple sequence alignment showing catalytic residues histidine and aspartic acid (in red block) in all five model UGTs](image-url)
The less conserved N-terminal residues predominantly forms the acceptor binding pocket. However, few C-terminal residues (especially from loops C1 and C5) which interacts with the sugar donor also contribute in part to the formation of the acceptor pocket. Residues from the loop regions such as loops N1, N2, N3, N4 and N5 as well as the Na3, Na5 and Na5a regions have been reported to be the main players of the acceptor pocket (Fig 4.1). These residues also determine acceptor specificity (Osmani et al., 2009). The acceptor pocket structural features are likely to control the acceptor specificities on enzymes (Li et al., 2007). These features include the backbone structure which determines the total shape of the acceptor pocket. This backbone structure is reportedly highly conserved despite the region’s poor amino acid conservation in solved plant UGTs structures. Variations in backbone structure and residues side chains (which either stabilise interactions or restrict space) controls the acceptor compound that can be accommodated in any particular UGT (Osmani et al., 2009). The N-terminal regions around the acceptor substrate in the modelled UGTs are shown in Fig 4.2. These are regions typical of acceptor substrate binding area in plant UGTs.

The N2 loop is absent in our UGTs where the Na2 helix trail the Nβ2 (Fig 4.2). Similar absence of N2 loop was reported in VvGT1 and MtUGT85H2 although in other plant UGTs, this region is in close proximity to the acceptor (Shao et al., 2005; Brazier-Hicks et al., 2007).

Fig 4.20: N-terminal regions which typically surrounds the acceptor substrate in plant UGTs are highlighted in the model UGTs (Colour code: 76E1-blue, 76E2-yellow, 76E4-pink, 76E5-brown, 76D1-grey and acceptor Quercetin in CPK colour)

The N3/Na3 forms a main part of the acceptor pocket in plant UGTs and also in the family 1 bacterial GTs. Amino acids cited here have been proposed to be key controllers of acceptor specificity in family 1 GTs (Osmani et al., 2009). In Streptococcus antibioticus UGTs OleD and OleI show good sequence similarity but differing acceptor specificity.
Analysis of both crystal structures indicated that the variance in acceptor substrate recognition was caused by the difference in the N3/Nα3 regions (Bolam et al., 2007). This region has varying length in MtUGT71G1, MtUGT85H2, VvGT1 and AtUGT72B1 although lining part of the acceptor pocket in all plant UGTs (Osmani et al., 2009).

In the modelled structures, the N3/Nα3 residues have been critically studied. 76E1 which can accommodate a lot more acceptors (Akere et al., 2018) has this region longer than others (Fig 4.21).

![Image: N3 loop/Nα3 region in all model UGTs showing longer length in 76E1 (circled). This difference in 76E1 is further studied in order to understand its high acceptor recognition (Colour code: 76E1-blue, 76E2-yellow, 76E4-pink, 76E5-brown, 76D1-grey and acceptor Quercetin in CPK colour)](image)

More importantly, a part of the Nα3 in 76E1 contain residues F\textsuperscript{77}KFLF\textsuperscript{81} which is present in the acceptor binding pocket of all the different acceptor compounds glycosylated. These residue combination FKFLF, which is unique to 76E1 (Fig 4.22) is particularly present around less than 4.5Å distance to all acceptor compounds.

![Image: MSA showing FKFLF residues in Nα3 of 76E1](image)
Docking studies have confirmed that these two phenylalanines F77 and F81 are in the binding pocket of all acceptor compounds recognised by 76E1 (Fig 4.25). These residues two are not found in the other four model UGTs. Both amino acid residues are constantly present in the binding pocket of all nine acceptor compounds recognised by 76E1. 

The nine acceptor compounds being of various kinds (Chapter 3) were recognised. The acceptor compounds recognised by 76E1 are kaempferol, quercetin, fisetin, apigenin, diosmetin, hesperetin, naringenin, eriodictyol and esculetin. Fig 4.23 shows the presence of phenylalanines F77 and F81 in the binding regions of all acceptor compounds to 76E1. How these residues F77 and F81 may have strengthened the stability of the acceptor binding pocket is however, unclear particularly the type of interactions involved.

On the contrary, in the other four UGTs (Fig 4.24), non-aromatic residues such as Q78 and L82 in 76E2, V77 and V81 in 76E5, L76 and L80 in 76D1 and A78 and M82 in 76E4 are found at those positions. Apart from the non-aromaticity of the side chains, their positions are also farther from the acceptor substrate aromatic rings making no interactions. Even in 76E2 where Q78 is present, no interactions occurred as well.
Fig 4.2: a) Q78 and L82 residues in 76E2 b) V77 and V81 residues in 76E5 c) LEFVL residues in 76D1 d) AEYLM residues in 76E4. Quercetin is the acceptor in a, b and c while scopoletin is in d. All acceptor ligands are coloured in CPK.

It will be interesting to find out how mutation of these unique phenylalanine residues will impact the high acceptor substrate recognition observed in 76E1. Residues from the N4 loop and Nα5 has been reported to be important in acceptor specificity (Osmani et al., 2009). N4 loop is next to Nβ4 which houses the catalytic aspartic acid involved in catalytic triad formation in the model UGTs (Fig 1.2 and Fig 4.2). A look into the N4 loop in all model UGTs is shown in Fig 4.25.

Fig 4.25: Multiple sequence alignment showing amino acids in N4 loop in all model UGTs. Residues in red box surround the acceptor binding pockets with tyrosine in 76E1 and 76E2 interacting with the acceptor ligand.
Tyrosine is present in 76E1 and 76E2 while leucine and phenylalanine are found in 76E5, 76E4 and 76D1 respectively. Y114 as shown in docking studies, interacts with all acceptors recognised by 76E1. Y114 may either be involved in a π-π interaction with the B ring of the flavonoid acceptor backbone (edge-to-face). In addition to stacking interaction, Y114 has an extra ability to form hydrogen bonds with both the B and C ring of the flavonoid acceptor backbone because of its OH group.

Examples of 76E1 Y114 interactions with both kaempferol and quercetin are illustrated in Fig 4.26. Similar interactions of Y114 are observed with diosmetin, hesperetin, naringenin and apigenin. This tyrosine interactions, in addition to the Nα3 F77KFLF81 support seemed to have helped 76E1 acceptor recognition.

76E2 has a tyrosine (Y115) at the same position like 76E1 Y114. Docking of an acceptor compound recognised in 76E2 screening indicated that Y115 formed hydrogen bond interaction just as 76E1 Y114 (Fig 4.27). However, the absence of Nα3 phenylalanines in 76E2 seem to make the difference. Hence, 76E2 showed low acceptor recognition just as others. The L114 was present at the same position in the acceptor binding pocket (kaempferol and quercetin) in 76E5, therefore neither hydrogen bond nor π-π interaction occurred here as expected. Similarly, L115 and F110 were found in 76E4 and 76D1 respectively.
Fig 4.27: Y115 interactions (in 76E2) with acceptor compound quercetin; hydrogen bonds formation and π-π interaction as displayed in 76E1 was also seen here.

In depth mutation studies focusing on the amino acid residues around the binding sites of acceptor substrates of these UGTs will explain further the influence of these residues as suggested by molecular interactions. This chapter has laid a foundation to such studies which will in the future, facilitate better understanding of the mechanism and substrate specificities of these UGTs.
Conclusion

In this chapter, models of group H UGTs (76E1, 76E2, 76E4, 76E5 and 76D1) were made via homology modelling in order to gain insight into how their structure drives substrate preference. Secondary structures were also predicted with the aid of ICM Molsoft as accurate prediction is a key element in tertiary structure prediction. The secondary structures of the UGTs showed a great similarity to that of AtUGT74F2. Using MODELLER, the most widely employed tool for homology modeling (due to its being freely available, powerful features and reliable output), the models were built based on the crystal structure of AtUGT74F2. The five UGTs show 25-31% sequence similarity to this template. These models were validated using PROCHECK Ramachandran plots and PROSA z-scores. 85-90% of the regions in the models had residues in the most favoured regions while the z-scores also are within the range typical of PDB native proteins. Both tools confirmed the quality of all homology models. Validation is important prior to further usage of the homology models.

Molecular interactions between the UGT models with donor sugar UDP-Glc were studied. This was done via automated ligand docking and the key residues involved in interaction with various parts of the donor sugar were explored. The three main typical parts of donor sugar are uridine, pyrophosphate base and sugar ribose. More particular attention was paid to the hydroxyl groups at C2, C3, C4 and C6 of the sugar ribose and the residues interacting with them. This was because residues at these positions drive donor substrate specificity. These interactions were then compared with that of experimental plant UGT crystal structures. This is vital to understanding if the models were similar in interactions to known plant UGTs or not. In summary, residues in the models interacting with the main parts of donor sugar were found to be similar to as found in the experimental crystal structures. Therefore, residues from other parts (especially C-terminal domain) of the UGT models may likely be responsible for the differences in donor substrate recognition.

Since UGT 76E1 displayed a much higher acceptor substrate recognition, its model was explored to observe structural features which may have enhanced this. All acceptor compounds recognised by 76E1 were docked into the model. Two N-terminal region (Nα3) residues phenylalanines F77 and F81 were noticed as distinct and present in the binding pockets of all acceptor compounds recognised by this UGT. How exactly these residues may have improved acceptor substrate recognition is unclear. However, future mutation studies may shed light on this.
References


Chapter 5  Enzyme engineering: Improving donor substrate recognition via mutagenesis studies
This chapter reports all mutations carried out in this research. As established earlier, no crystal structure for the studied UGTs is available. Therefore, homology modeling was employed to understand the UGTs choices of substrate through their molecular interactions. To further establish the usefulness and the predictive power of the model structures towards their substrate specificities, mutations were rationally designed based on the knowledge from sequence and structures. More importantly, the effect of both kinds of mutations (sequence-based and structure-based) on substrate recognition was studied. Mutant UGTs with improved donor recognition will advance our existing knowledge on UGT enzyme engineering.
5.1 Enzyme engineering

As introduced in Chapter 1 (section 1.4), enzyme engineering via rational design may be sequence-based or structure-based mutations (Pascarerella and Argos, 1992). This has been extensively used in UGT engineering (Noguchi et al., 2009; Osmani et al., 2008; He et al., 2006; Lim et al., 2005; Modolo et al., 2009; Modolo et al., 2009a; Wang, 2009).

With a focus on the donor sugar binding sites, the next sections will highlight some mutations around donor binding area and mostly in the C-terminal domain. This includes both structure-based mutations as well as MSA-based mutations. The outcome of these mutations on potentially improving donor sugar specificity and advancing enzyme engineering is explained.

5.1.1 MSA-based mutation experiments

As reported and discussed in Chapter 3 (substrate screening), UGTs 76E1 and 76E5 displayed better glycosylation activity than 76E2 and 76D1. Similar donor activity (UDP-Glc, UDP-Gal and UDP-GlcNAc) was observed in 76E1 and 76E5 while only UDP-Glc activity was seen in 76E2 and 76D1. A multiple sequence alignment (MSA) of the C-terminal domain indicated a spot in the Ca2 region (Fig 5.1), which highlights an amino acid residue similar to the donor sugar recognition pattern described above. In this region, serine was found in both 76E1 and 76E5 while asparagine was seen in 76D1 and 76E2 (Fig 5.1).

![Fig 5.1: Multiple sequence alignment showing serine and asparagine (in red box) in UGTs similar to donor recognition pattern observed in screening. This was flagged as a potential mutation site to improve donor recognition.](image)

As pointed out earlier in Fig 4.2 (chapter 4), C2 loop /Ca2 region is the most disordered and most varied part of the C-terminal domain. Similar divergence of the C2 loop /Ca2 region was observed in crystal structures of VvGT1, 71G1, 72B1 and 85H2 (Osmani et al., 2009). In 72B1, amino acid residues within this region influenced acceptor substrate specificity directly mainly via intramolecular interactions despite being too far for any direct interaction with the acceptor substrate, (Brazier-Hicks et al., 2007). Such indirect intramolecular interactions stabilised both secondary and tertiary structure and are important for activity and specificity (Osmani et al., 2009). The varying amino acids (serine and asparagine) observed in the UGTs as shown in Fig 5.1 interestingly belong to this loop C2/Ca2 region.
Therefore, asparagine (N320) in 76E2 was mutated to serine while serine (S318 and S311) in 76E1 and 76E5 respectively were mutated to asparagine. The effect of N320S, S318N and S311N mutation on the donor substrate recognition of these UGTs was studied. All mutant UGTs were designed and produced as described in sections 2.2.1 and 2.2.4 (chapter 2). UGT activity assay method was carried out for donor substrate screening as reported in section 2.2.2 (chapter 2).

Furthermore, alanine scanning, a technique used to determine the role of a specific residue’s side chain in bioactivity was further utilised to study serine significance at positions S318 and S311 in 76E1 and 76E5 respectively. Alanine mutation imitate deletion of an amino acid by eliminating the side chain but preserving proteins secondary structure. Hence, the mutation helps to determine an amino acid residue’s contribution in an enzyme function and structure (Morrison and Weiss, 2001).

The results of the mutant donor screening (compared to the wildtype UGTs) are summarised in the (Green/Amer/Red) GAR format as shown in Fig 5.2. Green, amber and red indicates positive, ambiguous and no activity as explained in Chapter 3. When compared to that of wildtype UGT 76E2, mutant UGT 76E2 N320S displayed improved donor substrate recognition, showing positive activity with UDP-Gal and UDP-GlcNAc. This suggests that serine had impacted the substrate recognition pattern.

Fig 5.2: GAR screen showing donor recognition activities of mutant UGTs as well as wildtype UGTs where green, amber and red indicates positive, ambiguous and no activity.
Probing further into how the mutation N320S has improved substrate recognition, docking studies indicated that in the wildtype 76E2, N320 (and indeed the entire C2 loop /Cα2 region) is quite far from the donor substrate. Hence, no direct interaction was observed between asparagine and the donor sugar (Fig 5.3).

Fig 5.3: Far position of N320 and C2 loop/Cα2 region in 76E2 to the donor substrate, UDP-Glucose.

In the mutant UGT 76E2 N320S, it was also observed that serine has no direct interaction with all donor sugars recognised (UDP-Glc, UDP-Gal and UDP-GlcNAc). This is illustrated in Fig 5.4.

Fig 5.4: Position of S320 C2 loop/Cα2 region in mutant UGT 76E2 N320S showing far distance to the donor sugars recognised. Figures are labelled as a) S320 and UDP-Glucose b) S320 and UDP-Galactose c) S320 and UDP-GlcNAc. No direct interactions were observed in between S320 and all donor substrates.
It is thought that indirect intramolecular interactions may be responsible for the massive improvement in the substrate recognition (both donor and acceptor). This is like as observed in the crystal structure of 72B1 where C2 residues Y315 and N312 improved substrate recognition without a direct interaction with substrates (Brazier-Hicks et al., 2007).

On the other hand, mutation of serine to asparagine in both 76E1 and 76E5 resulted in total loss of all activities. As shown in Fig 5.2, both 76E1 S318N and 76E5 S311N completely lost UDP-Glc, UDP-Gal and UDP-GlcNAc activities displayed by their respective wildtypes. Docking studies indicated that serines in both mutant UGTs were further away from all the donor sugars. Both mutant UGTs showing no direct interaction of N318 and N311 with donor sugar UDP-Glc is shown in Fig 5.5 even though total activity loss occurred. This further confirms the importance of serine in this region to donor substrate recognition, although no apparent direct interaction observed.

Both serine and asparagine are polar amino acids and are expected to interact in similar ways, thus having same effect on function. However, this does not seem to be so as shown by results in Fig 5.2. Structurally, serine side chain is of a smaller size (than asparagine) and is usually found within tight turns on the surface of protein (Betts and Russell, 2013). The OH group side chain forms a hydrogen bond with the protein backbone.

Mutation to alanine in both 76E1 S318A and 76E5 S311A showed activity to all three donor sugars being retained (Fig 5.2). Both mutants retaining all donor activities as found in wildtypes, indicated that alanine did not disrupt the structure (and by extension, the proposed intramolecular interaction) as asparagine did.

---

**Fig 5.5:** No interaction (direct) observed in mutant UGTs with UDP-Glucose despite total activity loss upon mutation a) in 76E1 S318N b) in 76E5 S311N
This suggests that mutation to alanine may not have disrupted both protein structures. The question of significance of serine comes to mind. How did mutation to asparagine result to loss of glycosylation activity but that to alanine didn’t?

Serine may be substituted and tolerated by amino acids with small side chains such as alanine, glycine, proline or cysteine. The effects of substitution of a small side chain for a larger one (such as in asparagine) can be devastating. Although alanine side chain is non-polar however, it is similar to serine in size as both classified as small amino acids (Betts and Russell, 2013). These structural differences inside chains size (Fig 5.6) may explain the loss of activity observed in 76E1 S318N and 76E5 S311N and the retaining of activity seen in S318A and 76E5 S311A.

![Serine, Alanine, Asparagine 2D structure](image)

Fig 5.6: Serine, alanine and asparagine 2D structure showing side chain comparison

### 5.2 Structure-guided mutation experiments

#### 5.2.1 Mutation based on other plant UGT crystal structures

As described in chapter 4, some amino acids have previously been reported in plant UGTs crystal structures to control donor sugar activity such as aspartic acid/glutamic acid and glutamine (the residues at position 43 and 44 in the PSPG motif). Therefore, few mutations were made at these sites for some of the model UGTs to further study the role of aspartic acid/glutamic acid and glutamine in donor sugar recognition. Results of the four mutant UGTs donor activities is shown in Fig 5.7. Mutant UGTs 76E1 Q373A and 76E5 Q366A both completely lost their glycosylation activities with all donor sugar screened. These finding also specifically confirms the role of glutamine Q in the UDP-Glc recognition of 76E1 and 76E5. Elimination of Q side chain may have disrupted interaction with O2 and O3 of glucose moiety, resulting to loss of activities. Similarly, total loss of UDP-Glc activity in particular, was reported for mutant UGTs 73C1 394A, 73C5 Q398A, 73C6 Q398A and 76E11 Q374A. This indicates the significance of Q in these enzymes (Zhang et al., 2018).
In chapter 4, it was established that Q373 in 76E1 binds to C3-OH only (and not to C2-OH as seen in others). To visualise this interaction, docking studies has shown that in mutant 76E1 Q373A, the hydrogen bond to the OH of glucose C3 was indeed disrupted upon mutation, resulting in the loss of activity. For the 76E5 mutant, the hydrogen bond between Q and OH (C3 and C4-OH here) was also abolished, accounting for the activity loss (Fig 5.8).

Fig 5.7: GAR screen results showing summary of mutant UGTs donor activities where green, amber and red indicates positive, ambiguous and no activity

Fig 5.8: Hydrogen bond interactions of Q373 and Q366 in wildtypes 76E1 and 76E5 respectively with UDP-Glucose shown in a and c. Alanine mutation abolishing hydrogen bond interactions with UDP-Glucose in 76E1 and 76E5 in b and d respectively.
Both mutant UGTs 76E1 Q373A and 76E5 Q366A lost their UDP-Gal and UDP-GlcNAc activities. UDP-Glc and UDP-GlcNAc differ at the C2 position. Glutamine usually reside around the C2 and C3 of Glc, binding either or both OH groups in template UGTs. UDP-GlcNAc has an N-acetyl group instead of an OH at C2 position. The loss of GlcNAc activity suggests an important role of glutamine in the activity. Some previous studies on GlcNAc transfer has observed glutamine interacting with OH at C3 position of the GlcNAc moiety (Hu et al., 2003; Meech et al., 2012).

As shown in Fig 5.9, 76E5 Q366 hydrogen bond to C3-OH in GlcNAc was impaired after mutation. Similar behaviour is predicted for 76E1 Q373. This may explain why mutation of glutamine led to loss of GlcNAc activity in both after mutation to alanine.

UDP-Gal also differs from UDP-Glc in the orientation of the -OH group at C4 position. The absence of UDP-Gal activity upon mutation indicate Q is significant for the activity in 76E1 and 76E5. 76E1 Q373 binds to C2 of Gal via hydrogen bonding, hence disruption of this bond is believed to have abolished Gal activity. Likewise, 76E5 Q366 interacts with Gal via hydrogen bonding to the C2 and C3-OH, which is absent in mutant 76E5 Q366A (Fig 5.10).
Although other residues might be involved alongside glutamine to define these activities, nevertheless glutamine plays enough vital role to uphold the activities. Information from crystal structure in the near future will unravel this.

Fig 5.10: Hydrogen bond interactions of Q373 and Q366 in wildtypes 76E1 and 76E5 respectively with UDP-Galactose shown in a and c. Alanine mutation abolishing hydrogen bond interactions with UDP-Galactose in 76E1 and 76E5 in b and d respectively.

Mutant UGTSs in 76E4, E374A and Q375A behaved very differently. The wildtype 76E4 showed no activity with UDP-Glc, UDP-Gal and UDP-GlcNAc (acceptor substrates kaempferol and quercetin). 76E4 E374A behaved exactly like the wildtype, displaying no glycosylation activity with these substrates. However, 76E4 Q375A showed positive activity with UDP-Glc and UDP-Gal but not UDP-GlcNAc. This means mutation at this point to alanine facilitated binding to UDP-Glc and UDP-Gal. This would also suggest that glutamine is not the key amino acid for UDP-Glc and UDP-Gal binding in 76E4. This is because both UDP-Glc and UDP-Gal activity is gained following the mutation Q375A. Docking studies was not able give any structural insight into how UDP-Gal recognition was facilitated. It is thought that it would be through some indirect intermolecular interactions as no direct bonds were formed between alanine (of mutant 76E4 Q375A) and the donor sugar.
The ability of wildtype UGT 76E4 to recognise both GDP- to UDP-donor sugars is nevertheless, remarkable. Crystal structure information in the future will shed light on the mechanism utilised by 76E4 to achieve this.

5.2.2 Mutation based on structural differences of model UGTs

Some potential key amino acids were inferred through the combination of MSA and model structure comparison among the studied UGTs. These amino acids were thought to affect the choice of donor substrate recognition. This may help to obtain mutants whose donor substrate recognition activity will be studied and then compared to the wild-type UGTs’ activities. The loss or retention of original activity as well as acquisition of new activity may explain the role of a particular amino acid residue. Activity comparison was limited to mainly qualitative analysis.

The modelled UGTs were studied for visible structural differences in the C-terminal domain to view potential residues involved in the pattern of donor sugar recognition. Some residues in the C1 loop, C4 loop and C5 loops were identified. Also, an N5 loop residue was marked as a potential key site. Fig 5.12 highlights the amino acids to be mutated which are believed to impact donor sugar recognition activity. Identification of residues and the prediction that they may be involved in substrate recognition guides experimental analysis and helps understanding function. A structural alignment along with MSA can expose conserved points which participate in binding and substrate recognition (Capra and Singh., 2007).

As observed in the MSA of model UGTs (Fig 5.11), amino acids K275 in 76E4, G347 in 76D1, P129 in 76D1, and D374 in 76E2 were identified as potential sites, which may affect substrate (donor sugar) activity. The positions of these residues in the UGTs and their proximity to the donor ligand further fuels this choice for mutation. As previously reported, our donor substrate screening results indicated 76D1 and 76E1 recognised only UDP glucose compared to 76E1 and 76E5 with wider donor activity (Akere et al., 2018). This was described in detail in Chapter 3. 76E4 however, could not utilise any of the donor substrates that were tested. The homology models of these UGTs have suggested various amino acids which may affect donor substrate recognition.
Fig 5.11: Multiple sequence alignment (MSA) of UGTs. Amino acid residues in blue boxes shows the regions of differences in the donor binding regions while the amino acid underlined in red indicates the mutated ones.
Fig 5.12: Potential mutation sites which may affect donor substrate activity. All residues in wildtypes a. 76E4 K275, b. 76D1 P129, c. 76D1 G347 and d. 76E2 D374 show proximity to the UDP donor sugar

Mutant UGTs 76D1 P129T, 76E4 K275L, 76E2 D374E and 76D1 G347C were designed and produced. Subsequently, a donor activity screening was carried out using three sugar donors; UDP-Glc, UDP-Gal and UDP-GlcNAc with acceptor substrates quercetin and kaempferol. Liquid chromatography-mass spectrometry/mass spectrometry (LC-MS/MS) was used to observe reaction products, hence, to confirm sugar addition. The findings are summarised in Fig 5.13.

While wildtype 76E4 have displayed no glycosylation activity in our donor screening with UDP-Glc, UDP-Gal and UDP-GlcNAc (acceptor compounds kaempferol and quercetin). 76E4 K275L recognised all donor sugars screened namely UDP-Glc, UDP-Gal and UDP-GlcNAc (with the same acceptor substrates). Furthermore, flavonoid acceptor substrates that were not recognised by wildtype 76E4 were utilised by the mutant 76E4 K275L.
Fig 5.1: GAR donor screening results of mutant UGTs 76E4 K275L, 76D1 P129T, 76D1 G347C and 76E2 D374E (wild type UGTs in box for comparison).

The importance of C1 loop residues in substrate recognition has been reported (He et al., 2006). The structural difference observed in C1 loop (lysine in place of leucine/isoleucine) predicted this mutation firstly (Fig 5.12). Secondly, is the role of C1 loop in substrate recognition has been previously reported (Osmani et al., 2009). As found in our UGTs as well as the nine solved plant UGTs structures, the conserved serine/threonine in C1 loop typically have an hydrophobic residue next to it (except in 85H2). The conserved serine/threonine in this loop generally forms hydrogen bond with the UDP-phosphate group in plant UGTs (George-Thompson et al., 2017). This is shown in Fig 5.14.

Fig 5.14: MSA showing C1 loop residues potentially involved in substrate recognition in nine solved plant UGTs
As polar lysine was found in 76E4, this C1 loop mutation (76E4 K275L) has shown tremendous improvement in substrate recognition. Docking interaction was unable to explain how exactly mutation K275L was able to influence mutant’s recognition of all donor sugar tested. However, L275 being next to T274 interacting with the phosphate group, its closeness of to the phosphate group and the C1 reaction centre of the donor sugars may all have contributed to the improved donor specificity. An example with UDP-Glc is illustrated in Fig 5.15.

Fig 5.15: L275 in mutant 76E4 K275L showing its closeness to the C1 reaction centre of UDP-Glucose which may have helped improved donor specificity

Such improvement was reported in UGT71G1, the mutation M286L (of C1 loop) increased activity with UDP-Gal and UDP-Glucuronic acid (He et al., 2006).

Although L275 is of the C-terminal domain, is has been found to be part of the acceptor binding pocket in mutant 76E4 K275L (Fig 5.16). Flavonoids such as kaempferol and quercetin were recognised by the mutant 76E4 K275L. This may be due to the interactions that L275 provides, suggesting that hydrophobicity in binding pockets may favour flavonoids recognition. With respect to the acceptor substrate recognition, wildtype 76E4 recognised coumarins but not flavonoid compounds (Chapter 3 acceptor screening results).
Fig 5.16: L275 (in black circle) and other residues in the acceptor’s (kaempferol) binding pocket in mutant 76E4 K275L, suggesting how hydrophobic L275 may have aided flavonoid kaempferol’s recognition. Kaempferol was not recognised by wildtype 76E4 which has K275.

Although bulk of the sugar donor-interacting residues are found in the C-terminal, some regions within the N-terminal are also involved in these important interactions. N1 and N5 loops are in the N-terminal domain regions, which can offer interactions with donor sugar due to their close proximity. VvGT1 crystal structure confirmed that N5 loop residue T141 interacted with 6’OH of the donor sugar (Offen et al., 2006). Frequently, this residue is conserved as threonine/serine and is found in AtUGT74F2, MtUGT71G1 and MtUGT85H2 (George-Thompson et al., 2017, Hans et al., 2004, Offen et al., 2006, Shao et al., 2005, Brazier-Hicks et al., 2007 and Li et al., 2007). Similarly, models 76E1, 76E2, 76E4 and 76E5 has threonine at this position as well. However, this N5 loop residue presents as a proline in 76D1. Therefore, mutant 76D1 P129T was produced and screened for its donor activity. The results indicated that an additional donor sugar was recognised (UDP-galactose) as shown in Fig 5.13. This results suggest that the N5 loop residue threonine may be involved in UDP-Gal recognition confirming the involvement of N5 loop in donor sugar recognition.

As explained in Chapter 4, threonine at this position (N5 loop) typically binds to the C6-OH of glucose moiety. Substitution of 76D1 P129 to T was expected to facilitate this binding. However, docking studies (Fig 5.17) shows that T129 in the mutant UGT seem not to directly interact with C6-OH of both glucose and galactose despite the mutant’s new activity (UDP-Gal activity).
However, in mutant 76D1 P129T, a hydrogen bond interaction formed between T129 and H19 (the catalytic amino acid acting as general base for deprotonation of acceptor compounds in UGTs – discussed in Chapter 1). For glycosylation activity to have occurred, a catalytic triad was established between H19, D107 and 3-0H quercetin, which must also be near the C1 atom of donor sugar. The hydrogen bond interaction between T129 and H19 may have aided to confer UDP-Gal specificity to the mutant UGT (Fig 5.18).
Structurally, proline has its side chain connected to backbone twice to form a five-membered nitrogen containing ring (Fig 5.19). Hence, it is unable to assume conformations like other amino acids and therefore resides in very tight turns in protein structures. In addition, proline is hydrophobic. On the other hand, threonine is polar, fairly reactive and able to form hydrogen bonds (Betts and Russell, 2013). Threonine in mutant 76D1 P129T may have increased flexibility to the conformation conferred on the protein by proline structure. This allows the protein to adopt conformations that helped in UDP-Gal recognition. T129 may also have formed hydrogen bonds with other than C6-OH of donor sugar moiety which may have helped improved its activity such as shown in Fig 18.

![Fig 5.18: Role of T129 in UDP-Gal recognition in mutant UGT 76D1 P129T. T129 interact with H19 involved in the formation of catalytic triad which is key to glycosylation activity.](image1)

![Fig 5.19: Structures of proline and threonine highlighting differences in side chains.](image2)
76E2 D374E mutant displayed a UDP-Gal activity that was absent in the wildtype 76E2. In addition, recognition of substrates kaempferol was observed in the mutant UGT. This residue belongs to the C5 loop, the end part of the PSPG motif. 76E2 wildtype however, only recognised UDP-Glc. Aspartic acid is very similar to glutamic acid as both are negatively charged, polar amino acids. However, glutamic acid has a longer side chain than aspartic acid by one carbon atom (Fig 5.20). Aspartic acid’s shorter side chain confers a slighter rigidity within protein structures (Betts and Russell, 2003). As shown in Fig 5.20, the longer side chain of glutamic acid may have assisted mutant 76E2 D374E to achieve a stronger hydrogen bond to UDP-Gal by moving closer to the C4-OH.

![Aspartic acid and Glutamic acid](image)

Fig 5.20: Structural difference in aspartic acid and glutamic acid highlighting difference in side chain.

Molecular interactions identified via docking explains that D374 in 76E2 wildtype did not interact with the C4-OH of UDP-Gal. However, with E374 in mutant UGT 76E2 D374E formed a hydrogen bond with C4-OH of UDP-Gal (Fig 5.21). The C4-OH interaction is believed to be central to UDP-Gal activity which may be why the wildtype didn’t recognise UDP-Gal.

![Molecular interactions](image)

Fig 5.21: a) D374 in wildtype 76E2 showing no interaction with UDP-Gal which may explain why the wildtype UGT did not show any activity with the donor sugar. b) E374 in mutant UGT 76E2 D3754E showing hydrogen bond interaction with C4-OH of UDP-Galactose which explains why the mutant recognised the donor sugar.
76D1 G347C displayed an “ambiguous” UDP-galactose activity in addition to UDP-glucose activity. This ambiguity is shown in the MS scan indicating positive glycosylation reaction but product ion scan was unable to confirm it. In addition, UDP-Glc activity with acceptor kaempferol was observed which was absent in wildtype 76D1. This may confirm our hypothesis on the importance of C4 loop cysteine on donor sugar recognition and in general substrate involvement.

Following mutation of G347 to C (cysteine), an unclear UDP-Gal was recognised as well as that of UDP-Glc and kaempferol (a new acceptor substrate). It was predicted that the formation of a disulphide bond with C364 may have improved the stability of the PSPG region. And ultimately, this may have fairly enhanced substrate recognition. However, docking studies as shown in Fig 5.22 indicates that no disulphide bond was formed following mutation. The distance between both cysteines was 3.97Å which is longer than that of a typical disulphide bond. A disulphide bond is about 2.05Å in length (Sevier and Kaiser, 2002). This may explain why not so significant donor recognition was achieved. However, the presence of cysteine has aided UDP-Glc recognition of another acceptor substrate. The mechanism by which this occurred is still unclear.

![Fig 5.22: Mutant 76D1 G347C showing potential interaction between two cysteines C364 and C347 which may have helped improved donor sugar recognition. This is bond length is however longer than a typical disulphide bond hence no disulphide bond was formed here.](image)

Intra-domain interactions such as disulphide bridges can confer stabilisation of secondary as well as tertiary structure of a protein, and they may be important for activity and specificity (Osmani et al., 2009).
The location of a disulphide bridge within a structure may influence its role in the stabilizing or folding of the protein. Disulphide bonds are formed between two cysteine residues. They help to stabilize the native conformation and maintain protein integrity, making them less susceptible to denaturation and degradation (Hogg., 2003). In MtUGT85H2, the stability of the PSPG motif region is said to be likely increased by the presence of a disulphide bridge between C349 and C366 (Li et al., 2007). Amino acids in enzyme catalysed reactions could act to stabilise conserved motifs for suitable substrate accommodation (Holliday et al., 2011). With future advances in techniques and more in-depth studies, the roles of disulphide bonds will be understood even more clearly (Khoo and Norton, 2012). The effect of the stability that disulphide bridges may confer on both UGTs is likely to have an effect on their donor substrate recognition. Although the presence of a pair of cysteine residues in the PSPG motif does not necessarily lead to the formation of disulphide bond. This was seen in VvGT1 where PSPG motif residues C351 and C368 were far apart (~3.4A) and do not form disulphide bridges (Offen et al., 2006). In mutation 76D1 G347C, C347 and C364 has also behaved similarly.

5.3 Conclusion

Mutations were designed in two ways; based on the UGT model structures and based on MSA and crystal structures (without using the models). Based on MSA design and crystal structures, mutant UGTs 76E2 N320S and 76E4 Q375A were observed to show improved substrate recognition. Based on the UGT models, mutant UGTs 76E4 K275L, 76D1 P129T, 76E2 D374E and 76D1 G347C displayed even better substrate recognition. Generally, the results indicate that mutation based on model structures enhanced substrate recognition more than the other. Docking studies explained how many of the mutant interactions may have confer such improved substrate recognition. This may include intramolecular interactions, hydrophobicity, shortening distance to aid establishment of bonding interactions and so on.

The improved substrate recognition observed in mutant UGTs explains how enzyme substrate specificity is enhanced via mutagenesis. Structure-guided enzyme engineering led to identifying novel mutant UGTs with improved donor substrate recognition. In addition, key amino acid residues driving substrate specificities in the studied UGTs have been revealed. The significance of homology modeling in the absence of crystal structures in understanding UGT substrate specificity has been described. It is hoped that the findings of this research will help improve the biotechnological and pharmaceutical potentials of the studied UGTs.
Final thesis conclusions

In this research, as specified in the aims and objectives (Chapter 1), the substrate preferences of selected UGTs (76E1, 76E2, 76E4, 76E5, 76D1, 84A1, 84A2 and 84A3) have been studied and reported (Chapter 3). Variety of acceptor substrates such as flavonoids, phenolic acids, alkaloids and stilbenes were used by the UGTs. Broad acceptor substrate specificity displayed by some of the UGTs (particularly 76E1 and 84A2) is important to their potential usage in the synthesis of diverse bioactive glycosides. In addition, wider donor substrate recognition (as seen in 76E5 and 76E1) could generally confer varying pharmacological activities. The kinetics data has indicated that lesser quantity (in uM volume) of the donor sugar substrates are required for reactions to take place. The implication of this is that the challenge on the high cost of UDP-donor sugars on in vitro glycosylation may be lessened with the use of these UGTs. These findings have contributed to the existing knowledge about biochemical information on these plant UGTs.

In addition, the structural studies have been carried out via homology modeling in the absence of experimental crystal structures (Chapter 4). Obtaining crystal structure can be time-consuming and challenging. Structural studies are vital to understanding functions of molecules. Therefore, as an alternative, homology modeling has been generally used to gain insight into structural information. Based on UGT74F2, models of only five UGTs (76E1, 76E2, 76E4, 76E5 and 76D1) were made. Validation and assessment of the models indicated good quality and via docking studies, structural interactions between substrates and model UGTs have been observed. 76E1 may have shown better acceptor substrate recognition due to some N-terminal (Nα3) phenylalanine residues which were noticed in the binding pockets of all recognised acceptors. Also, these residues are absent in other UGTs studied. Future mutation studies should explain the importance of these residues. This observation, which is based on information from the model structure, further show how helpful homology models can be in understanding UGTs substrate specificity.

Furthermore, some key amino acids have been discovered via rational manipulation which has improved substrate recognition thereby advancing UGT enzyme engineering (Chapter 5). Mutant UGTs such as 76E2 N320S, 76E4 K275L, 76D1 P129T, 76E2 D374E and 76E5 Q375A showed better substrate recognition than their wildtype UGTs. The design of these UGTs of broader spectrum would play important role in their biotechnological and industrial applications. These findings are believed, would serve as basis for further studies and usage of these UGTs.

Suggestions of future works which could follow up these submissions will be explained in the next section.
Future works

- Molecular dynamics (MD) simulations: Conformational changes upon substrate binding occur in UGTs like other proteins and this affects enzyme-substrate interactions. MD simulations will help in understanding how these conformational changes in structure affect substrate specificity in UGTs. Several studies have reported the use of MD simulations in GT studies (Sharma et al., 2014; Fujiwara et al., 2009; Tripathi et al., 2016). MD simulations have been used to explore aspects of conformational dynamics of the C-terminal domain (Haider et al., 2005). Therefore, this method will be useful to further understanding the mechanism of substrate specificity as aimed in this research.

Molecular dynamics (MD) simulation is a valuable tool in biology to understand structures of molecules and the microscopic interactions between them, serving as a complement to conventional experiments. This computer simulation method studies the physical movements of atoms and molecules; acting as a bridge between experiment and theory (Allen., 2004). Due to proteins’ flexibility and ability to undergo significant conformational changes, MD simulations play a part in elucidating their functionalities (Hospital et al., 2015).

- Effect of mutation of F77KLLF81 region on the acceptor substrate specificity of UGT76E1. Manipulations surrounding this N-terminal region in this UGT as postulated (Chapter 4) should have drastic effect on substrate recognition. Likewise, the impact of substitution of non-aromatic residues such as found in UGTs 76E2, 76E5, 76D1 and 76E4 with phenylalanine will be interesting to observe. This will advance UGT engineering.

- Structural studies of other UGTs from groups L via homology modeling. Since the substrate screening has been carried out (Chapter 3), homology modeling and docking studies will explain the choice of substrates, key amino acid residues as well as sites for potential mutations. Such mutation may help to improve substrate specificity studies.
References


APPENDIX
MUTANT UGTs – Mass Spectra [UGT activity assay – POSITIVE ACTIVITIES]

1) 76E4 K275L

Fig i: MS spectrum showing fragmentation of kaempferol glucoside (Product ion scan)

Fig ii: MS spectrum showing fragmentation of quercetin glucoside (Product ion scan)

Fig iii: MS spectrum showing fragmentation of kaempferol galactoside (Product ion scan)
Fig iv: MS spectrum showing fragmentation of quercetin galactoside (Product ion scan)

Fig v: MS spectrum showing fragmentation of quercetin GlcNAc glycoside (Product ion scan)

2) 76D1 P129T

Fig vi: MS spectrum showing fragmentation of quercetin glucoside (Product ion scan)
Fig vii: MS spectrum showing fragmentation of kaempferol glucoside (Product ion scan)

Fig viii: MS spectrum showing fragmentation of quercetin galactoside (Product ion scan)

Fig ix: MS spectrum showing fragmentation of kaempferol galactoside (Product ion scan)
3) 76E2 D374E

Fig x: MS spectrum showing fragmentation of quercetin galactoside (Product ion scan)

4) 76D1 G347C

Fig xi: MS spectrum showing fragmentation of kaempferol glucoside (Product ion scan)

Fig xii: MS spectrum showing fragmentation of quercetin glucoside (Product ion scan)
5a) 76E2 N320S (UDP-Glc K and Q)

Fig xiii: MS spectrum showing fragmentation of kaempferol glucoside (Product ion scan)
Fig xiv: MS spectrum showing fragmentation of quercetin glucoside (Product ion scan)
5b) 76E2 N320S (UDP-Gal K and Q)

Fig xv: MS spectrum showing fragmentation of kaempferol galactoside (Product ion scan)
Fig xvi: MS spectrum showing fragmentation of quercetin galactoside (Product ion scan)
5c) 76E2 N320S (UDP-GlcNAc K and Q)

Fig xvii: MS spectrum showing fragmentation of kaempferol GlcNAc glycoside (Product ion scan)
Fig xviii: MS spectrum showing fragmentation of quercetin GlcNAc glycoside (Product ion scan)
Acceptor screening result

A) 76E1

MS spectra; UDP Glucose (MW 162) was added to the test aglycones. MW of target peaks (glycosylated products) are shown in bold in the red boxes below.

i) Apigenin

![MS spectrum showing fragmentation of apigenin glucoside (Product ion scan).](image)

Fig xix: MS spectrum showing fragmentation of apigenin glucoside (Product ion scan).

MS/MS was used to fragment apigenin glucoside (MW 431). The observation of apigenin (MW 269) confirms the glucoside formation, and hence the glycosylation process.
ii) Esculetin

Fig xx: MS spectrum showing fragmentation of esculetin glucoside (Product ion scan)
iii) Quercetin

Fig xxi: MS spectrum showing fragmentation of quercetin glucoside (Product ion scan)
iv) Diosmetin

Fig xxii: MS spectrum showing fragmentation of diosmetin glucoside (Product ion scan)
Fig xxiii: MS spectrum showing fragmentation of hesperetin glucoside (Product ion scan)
Fig xxiv: MS spectrum showing fragmentation of hesperetin glucoside (Product ion scan)
vii) Eriodictyol

Fig xxv: MS spectrum showing fragmentation of eriodictyol glucoside (Product ion scan)
kaempferol

Fig xxvi: MS spectrum showing fragmentation of kaempferol glucoside (Product ion scan)
ix) Fisetin

Fig xxvii: MS spectrum showing fragmentation of fisetin glucoside (Product ion scan)
B) 76E2 Acceptor screening
i) Quercetin

Fig xxviii: MS spectrum showing fragmentation of quercetin glucoside (Product ion scan)
ii) Myricetin

Fig xxix: MS spectrum showing fragmentation of myricetin glucoside (Product ion scan)
Fig xxx: MS spectrum showing fragmentation of fisetin glucoside (Product ion scan)
C) 76E5 Acceptor screening

i) Kaempferol

Fig xxxi: MS spectrum showing fragmentation of kaempferol glucoside (Product ion scan)

ii) Quercetin

Fig xxxii: MS spectrum showing fragmentation of quercetin glucoside (Product ion scan)
iii) 2,5 DHBA

Fig xxxiii: MS spectrum of 76E5 glycosylation of 2,5 dihydrobenzoic acid (Full scan)
iv) Scopoletin

Fig xxxiv: MS spectrum of 76E5 glycosylation of scopolletin (Full scan)
B) 76D1 Acceptor screening

i) 2,5 dihydrobenzoic acid

Fig xxxv: MS spectrum showing fragmentation of 2,5 dihydrobenzoic acid glucoside (Product ion scan)
ii) **Quercetin**

Fig xxxvi: MS spectrum showing fragmentation of quercetin glucoside (Product ion scan)
iii) **Fisetin**

Fig xxxvii: MS spectrum showing fragmentation of fisetin glucoside (Product ion scan)
DONOR SCREENING RESULT AND GLYCOSYLATION POSITION

A) 76E1 Donor screening results

1) UDP glucose and Kaempferol

![MS spectrum showing fragmentation of kaempferol glucoside](image)

Fig xxxviii: MS spectrum showing fragmentation of kaempferol glucoside (Product ion scan)

2) UDP galactose and Quercetin

![MS spectrum showing fragmentation of quercetin galactoside](image)

Fig xxxix: MS spectrum showing fragmentation of quercetin galactoside (Product ion scan)
3) UDP glucose and Quercetin

Fig xi: MS spectrum showing fragmentation of quercetin glucoside (Product ion scan)

4) UDP GlcNAc and Quercetin

Fig xii: MS spectrum showing fragmentation of quercetin GlcNAc glycoside (Product ion scan)
B) 76E2 Donor screening
UDP glucose and Quercetin

![MS spectrum showing fragmentation of quercetin glucoside (Product ion scan)](image)

C) 76E5 Donor screening
1) a. UDP glucose and Kaempferol

![MS spectrum showing fragmentation of kaempferol glucoside (Product ion scan)](image)
b. UDP glucose and Quercetin

Fig xliv: MS spectrum showing fragmentation of quercetin glucoside (Product ion scan)

2) UDP galactose (Kaempferol and Quercetin)

Fig xlv: MS spectrum showing fragmentation of kaempferol galactoside (Product ion scan)

Fig xlvii: MS spectrum showing fragmentation of quercetin galactoside (Product ion scan)
3) UDP glcNAc (Kaempferol and Quercetin)

Fig xlvii: MS spectra showing fragmentation of kaempferol GlcNAc glycoside (Product ion scan)

Fig xlviii: MS spectra showing fragmentation of quercetin GlcNAc glycoside (Product ion scan)

D) 76D1 Donor screening

Fig xlix: MS spectrum showing fragmentation of quercetin glucoside (Product ion scan)
UGT 76E2 UDP-Glc

Michaelis menten plot

<table>
<thead>
<tr>
<th>S (uM)</th>
<th>V (uM/min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>0.0048</td>
</tr>
<tr>
<td>20</td>
<td>0.0064</td>
</tr>
<tr>
<td>50</td>
<td>0.0073</td>
</tr>
<tr>
<td>100</td>
<td>0.0076</td>
</tr>
<tr>
<td>200</td>
<td>0.0084</td>
</tr>
<tr>
<td>500</td>
<td>0.0084</td>
</tr>
</tbody>
</table>

Lineweaver Burk

\[
\frac{1}{S} \quad \frac{1}{V}
\]

<table>
<thead>
<tr>
<th>(\frac{1}{S})</th>
<th>(\frac{1}{V})</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>208.3333</td>
</tr>
<tr>
<td>0.05</td>
<td>156.25</td>
</tr>
<tr>
<td>0.02</td>
<td>136.9863</td>
</tr>
<tr>
<td>0.01</td>
<td>131.5789</td>
</tr>
<tr>
<td>0.005</td>
<td>119.0476</td>
</tr>
<tr>
<td>0.002</td>
<td>119.0476</td>
</tr>
</tbody>
</table>

\[K_m = 8.75 \pm 1.38 \mu M\]

\[k_{cat}/K_m = 1.23 \text{ s}^{-1}\text{mM}^{-1}\]

\[k_{cat} = 0.0107 \text{ s}^{-1}\]

Table i: Michaelis menten and Lineweaver-Burk plots for 76E2
UGT 76E5 UDP-Glc

Michaelis Menten plot

<table>
<thead>
<tr>
<th>S (uM)</th>
<th>V (uM/min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>0.0087</td>
</tr>
<tr>
<td>20</td>
<td>0.0116</td>
</tr>
<tr>
<td>50</td>
<td>0.014</td>
</tr>
<tr>
<td>100</td>
<td>0.0146</td>
</tr>
<tr>
<td>200</td>
<td>0.0151</td>
</tr>
<tr>
<td>500</td>
<td>0.0153</td>
</tr>
</tbody>
</table>

Lineweaver Burk

<table>
<thead>
<tr>
<th>1/S</th>
<th>1/V</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>114.9425</td>
</tr>
<tr>
<td>0.05</td>
<td>86.2069</td>
</tr>
<tr>
<td>0.02</td>
<td>71.42857</td>
</tr>
<tr>
<td>0.01</td>
<td>68.49315</td>
</tr>
<tr>
<td>0.005</td>
<td>66.22517</td>
</tr>
<tr>
<td>0.002</td>
<td>65.35948</td>
</tr>
</tbody>
</table>

Km \(8.79 \pm 1.64 \ \mu\text{M}\)

\(k_{\text{cat}}/K_M\) \(2.21 \ \text{s}^{-1}\text{mM}^{-1}\)

\(k_{\text{cat}}\) \(0.0194 \ \text{s}^{-1}\)

Table ii: Michaelis Menten and Lineweaver-Burk plots for 76E5
UGT 76E5 UDP-Gal

Michaelis Menten plot

<table>
<thead>
<tr>
<th>S (μM)</th>
<th>V (μM/min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.1</td>
<td>0.0019</td>
</tr>
<tr>
<td>0.2</td>
<td>0.0034</td>
</tr>
<tr>
<td>0.5</td>
<td>0.0046</td>
</tr>
<tr>
<td>1</td>
<td>0.0084</td>
</tr>
<tr>
<td>5</td>
<td>0.0102</td>
</tr>
<tr>
<td>10</td>
<td>0.0103</td>
</tr>
</tbody>
</table>

Lineweaver Burk

<table>
<thead>
<tr>
<th>1/S</th>
<th>1/V</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>526.3158</td>
</tr>
<tr>
<td>5</td>
<td>294.1176</td>
</tr>
<tr>
<td>2</td>
<td>217.3913</td>
</tr>
<tr>
<td>1</td>
<td>119.0476</td>
</tr>
<tr>
<td>0.2</td>
<td>98.03922</td>
</tr>
<tr>
<td>0.1</td>
<td>97.08738</td>
</tr>
</tbody>
</table>

Km = 0.57 ± 0.19 μM

k_{cat}/K_M = 22.8 s^{-1}mM^{-1}

k_{cat} = 0.0130 s^{-1}

Table iii: Michaelis Menten and Lineweaver-Burk plots for 76E5
UGT 76E5 UDP N-acetyl glucosamine

**Michaelis menten plot**

<table>
<thead>
<tr>
<th>S (uM)</th>
<th>V (uM/min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.5</td>
<td>0.006</td>
</tr>
<tr>
<td>1</td>
<td>0.007</td>
</tr>
<tr>
<td>2</td>
<td>0.0105</td>
</tr>
<tr>
<td>5</td>
<td>0.013</td>
</tr>
<tr>
<td>10</td>
<td>0.0131</td>
</tr>
</tbody>
</table>

**Lineweaver Burk**

<table>
<thead>
<tr>
<th>1/S</th>
<th>1/V</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>166.6667</td>
</tr>
<tr>
<td>1</td>
<td>142.8571</td>
</tr>
<tr>
<td>0.5</td>
<td>95.2381</td>
</tr>
<tr>
<td>0.2</td>
<td>76.92308</td>
</tr>
<tr>
<td>0.1</td>
<td>76.33588</td>
</tr>
</tbody>
</table>

**K_m**

$0.78 \pm 0.28 \mu M$

$k_{cat}/K_M$  

$21.3 \text{ s}^{-1}\text{mM}^{-1}$

$k_{cat}$

$0.0166 \text{ s}^{-1}$

Table iv: Michaelis menten and Lineweaver-Burk plots for 76E5
76D1 UDP-Glc

Michaelis menten plot

<table>
<thead>
<tr>
<th>S (uM)</th>
<th>V (uM/min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>0.0009</td>
</tr>
<tr>
<td>20</td>
<td>0.0016</td>
</tr>
<tr>
<td>50</td>
<td>0.0029</td>
</tr>
<tr>
<td>100</td>
<td>0.0034</td>
</tr>
<tr>
<td>200</td>
<td>0.0037</td>
</tr>
<tr>
<td>500</td>
<td>0.0037</td>
</tr>
</tbody>
</table>

Lineweaver Burk

<table>
<thead>
<tr>
<th>1/S</th>
<th>1/V</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>714.2857</td>
</tr>
<tr>
<td>0.05</td>
<td>625.0</td>
</tr>
<tr>
<td>0.02</td>
<td>344.8276</td>
</tr>
<tr>
<td>0.01</td>
<td>294.1176</td>
</tr>
<tr>
<td>0.005</td>
<td>270.2703</td>
</tr>
<tr>
<td>0.002</td>
<td>270.2703</td>
</tr>
</tbody>
</table>

K_m     25.77±8.82µM
K_cat/K_m 0.185 s⁻¹mM⁻¹
K_cat     0.00476

Table v: Michaelis menten and Lineweaver-Burk plots for 76D1
Fig L: PROCHECK Ramachandran plots for models 76E1, 76E2, 76D1, 76E4 and 76E5. Next to each models' Ramachandran plot is its Z-score. The black dot in each plot (circled in red) represents each model's Z-score in a plot containing the Z-scores of all experimentally determined protein structures currently in the PDB. Different sources of PDB structures (i.e X-ray and NMR) are distinguished by different shades of blue.
Fig li: Lane 1: Protein ladder; Lane 3: Purified 76E5 MW (80kda)

Fig lii: Lane 1: Protein ladder Lane 3: Purified 76E2 (MW 65kda)
High throughput mass spectrometry-based characterisation of *Arabidopsis thaliana* group H glycosyltransferases†

Aishat Akere, Qian Liu, Shibo Wu, Bingkai Hou and Min Yang

In this report, we cloned and characterised four members of group H glycosyltransferases (GTs) by studying their substrate specificities and kinetics. The formation of products and possible glycosylation position was confirmed using MS/MS. The results revealed that 76E1 and 76E5 have broader donor specificity, including UDP-glucose (UDPGlc), UDP-galactose (UDPGal) and UDP- N-acetylglucosamine (UDPGlcNAc) with various flavonoids as acceptor substrates. Pseudo-single substrate kinetics data showed a relatively low $K_{m}$, indicating a high affinity for substrate UDPGlc and also supported that 76E5 is more of a galactosyl and N-acetylglucosamine transferase. Sequence alignment and site-directed mutagenesis studies indeed suggested that serine is a crucial residue in the UDPGlcNAc and UDPGal activity.

Introduction

Glycosylation is a widespread modification of plant secondary metabolites involved in various functions such as the regulation of hormone homeostasis, the detoxification of xenobiotics and the biosynthesis and storage of secondary compounds. In plants, these reactions are controlled by a specific subclass of the glycosyltransferase family. Family 1 plant UDP dependent glycosyltransferases (UGTs) are responsible for the transfer of sugar to many bioactive natural products such as alkaloids, terpenoids, flavonoids and phenylpropanoids. UDP activated sugars are mainly used as donor compounds by UGTs. Generally, transfer of sugar molecules can be either to macromolecules (such as proteins, lipids) or small molecules (such as plant secondary metabolites – natural products, oligosaccharides). However, the products of glycosylation catalysed by UGTs are glycosides of small molecules.

Numerous pharmacological activities have been reported for natural products containing sugar residues. This is because glycosylation improves their stability, solubility in water and bioavailability; hence it has become an important process in drug research and development. In addition, through this process, glycosides have become attractive compounds used as nutraceuticals and food additives. Leaves and seeds of *Digitalis purpurea* and *Digitalis lanata* are the most significant sources of cardiac steroids important in the treatment of various heart conditions. The aglycone steroid however, binds and dissociates too quickly from the receptor. Attachment of the sugar moiety to the aglycone steroid improves its binding stability to receptor, though the sugar itself has no cardiac activity. Hence, sugar addition enhanced the pharmacokinetic behaviour of the aglycone steroid, leaving us with low dissociation potent glycosides. New and effective methods to produce glycosylated natural products have been studied over the years. Since glycosyltransferases are important biocatalysts in sugar addition, it is imperative to study UGTs more in depth.

Presently, 122 UGT encoding genes are found in the model plant *Arabidopsis thaliana*. Phylogenetic analysis of conserved amino acids in *Arabidopsis* UGTs led to the classification into 14 groups (ESI Fig. 1†). Although these enzymes have been studied for many years, to date only a few have been characterized in planta. Of the 19 genes found in group H, only 2 are fully characterised. Varieties of glycosides found in plants *in vivo* signify that plant UGTs can glycosylate diverse compounds. Although UGTs were believed to be highly specific in substrate recognition *in vivo*, reports have disclosed broader range promiscuity against both substrates. This substrate recognition includes recognition of both UDP sugar donor as well as acceptor. Generally, UGTs are highly specific for sugar donor although UDP-glucose is most commonly used. The screening studies which looked into substrate specificity within *Arabidopsis thaliana* phylogenetic groups found that UGTs of the same group glycosylate acceptors belonging to very different compound classes. These early studies challenged efforts to solely base substrate specificity on phylogeny. Although biochemical characterisation of the substrate specificity is quite challenging as it entails testing diverse substrates to fully understand specificity of individual UGTs, it will advance...
prediction of substrate specificity based on phylogeny. Comprehensive facts on the in vivo and in vitro activity and specificity of UGTs are important in the design of enhanced UGTs with desired properties.

In recent times, the importance of glycosyltransferases (GTs) has been in the spotlight. However, insufficient biochemical data on individual member enzyme has hindered further research into their functional understanding. Identifying and biochemically characterising some GT genes has helped in studying mechanism of glycosylation in planta. Biochemical data on substrate preferences has been relied on for prediction of substrate specificity based on phylogeny. Till date, most of Arabidopsis thaliana UGTs substrates remain unknown.

Here, we report the substrate specificity and kinetic analysis of four Arabidopsis thaliana group H UGTs (76E1, 76E2, 76E5 and 76D1). Substrate screening showed that the UGTs added sugar to acceptors from different classes of secondary metabolites ranging from flavonoids, cinnamic acids, coumarins to alkaloids. They also recognised varying donor compounds, with all accepting UDPGlc. In addition, 76E1 and 76E5 recognise UDPGaal and UDPGlcNac. Using tandem mass spectrometry, likely glycosylation positions of the acceptors were proposed. Further kinetic study measured each enzyme binding affinity with donor substrates, indicating their donor preferences and catalytic efficiencies. Knowledge of the activity and specificity of UGTs would provide a basis for their potential use in the design of improved UGTs to produce bioactive molecules for pharmaceutical purpose.

**Experimental**

(a) **Gene cloning procedure**

Total RNA was extracted from the 14d seedlings of the Arabidopsis using the TRIzol reagent (Takara, Japan). First-strand cDNA was synthesized using PrimerScript RT reagent Kit with gDNA Eraser (Takara, Japan), pooled cDNA was used as PCR template. The cDNA sequences of UGT76E1, UGT76E2, UGT76E5 and UGT76D1 were amplified by TransStart Fast Pfu DNA Polymerase (TRANS, China). A BamHI and XhoI restriction site were included in the sequence of forward primer and the reverse primer used for PCR of UGT76E1 and UGT76E2. The PCR primers for UGT76E1 cloning are: forward GGATCCATGGAGAAAAATGCA-CTCGAGTCACATGGAATTAACAAAGTC and reverse CTCGAGCTAACAGTGAATTCATCATT; the PCR primers for UGT76E5 were: forward CGGGATCCATGGAAATATGCA-CTCGAGTCACATGGAATTAACAAAGTC and reverse CTCGAGCTAACAGTGAATTCATCATT.

(b) **Expression and purification of AtUGT76E1, AtUGT76E2 and AtUGT76E5**

The recombinant plasmids were transformed into BL1 (DE3) cells for protein expression. Thereafter, they were verified by DNA sequencing via sanger sequencing (Source Bioscience Ltd). The bacterial cells were grown at 37 °C while being shaken at 200 rpm to an OD<sub>600</sub> of 0.6–0.8. Isopropyl 1-thiogalactopyranoside (IPTG) was added to the cell culture to a final concentration of 0.1 mM to induce protein expression. The culture was grown overnight at 20 °C and collected by centrifugation at 4 °C. The GST-tagged recombinant protein was purified by affinity chromatography (columns) and quantified using the Bradford assay according to the standard procedure provided by the suppliers. Cell lysate before purification and purified protein were analysed on sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) gels to verify protein production (ESI Fig. 2a–d†).

(c) **In vitro UGT reaction assay**

The UGT enzyme assay included the following components: 1 mM Tris 1 mM, MgCl<sub>2</sub> (pH 8.0), 10 mM UDP-sugars, 10 mM acceptor compounds and purified target proteins. For the acceptor screening, the 42 compounds screened are shown in the acceptor library (ESI Table 1†). The 7 sugar donor compounds used in the donor screening are also indicated in Fig. 2. The reaction mixture was incubated at 37 °C for 3 and 15 hours at different times. The reaction was terminated with acetonitrile and centrifuged to remove proteins. The supernatant was then analysed with LCMS. Glycosylated products were identified by their molecular weights, and these target compounds were subsequently fragmentated using MS/MS for confirmation.

(d) **UDP-Glo™ glycosyltransferase assay – for kinetics**

Kinetic study was performed using the UDP-Glo™ glycosyltransferases Assay kit-following the procedure provided by the supplier. The UGT enzyme assay mixture for kinetics contained Tris buffer (50mM Tris, 5 mM MgCl<sub>2</sub>, pH 7.5), kaempferol (for 76E1 and 76E5) and quercetin (for 76E2 and 76D1). Kaempferol and quercetin were fixed at 20 µM. The donor substrates concentration; UDP Glc (five concentrations) was varied from 10–500 µM while UDP Gal and GlcNAc were varied from 0.1–10 µM. The UGT reaction was carried out at 25 °C and terminated using the UDP-Glo assay detection buffer. This kit detects UDP generated after UDP-sugar transfer by converting UDP to light (measured in relative luminescence units) in a luciferase type reaction. A linear standard curve using 0–25 µM UDP was performed (ESI Fig. 17†). Following the kit’s protocol, the UGT
The amino acid sequences of the wild type UGTs were aligned using clustal omega by European Bioinformatics Institute (https://www.ebi.ac.uk/Tools/msa/clustalo/). The mutant and wild type DNA sequences were compared using NCBI nucleotide BLAST (National Centre for Biotechnology Institute).

**Results and discussion**

(a) **MS based HTS method**

The genes encoding 76E family and 76D1 were expressed as recombinant proteins in *Escherichia coli* and purified to electrophoretic homogeneity. *In vivo*, the functions of these UGTs are not known except for 76D1 whose functions have been recently published. The Green-Amber-Red ‘GAR’ screen was developed to probe potential substrates, *in vitro*, using a large acceptor and nucleotide-sugar donor library (Fig. 2, ESI Table 1†). A virtual colour, reflecting DNA micro-array practice, of Green-Amber-Red (GAR, green indicates positive, amber means not clear and red shows negative) is assigned reflecting the presence of selected ions of mass. This procedure using liquid chromatography triple quadrupole Mass Spectrometry (LC-MS/MS) was composed of two parts; a general MS scan to identify desired product peaks such as [M – H]⁻ or [M + Cl⁻]⁻ in negative mode and [M + H]⁺ or [M + Na]⁺ in positive mode first, and then a MS/MS scan to confirm the presence of acceptor precursor which indicates the loss of the sugar moiety to confirm the reaction product (Fig. 1a). This method, similar to the previous NLCT, could be used to determine the glycosylation site. No difference was noticed in the reaction products from 3 h and 15 h assay incubation. In addition, pseudo single substrate enzyme kinetics was measured by varying one substrate concentrations while saturating the other substrate using a commercial UDP-Glo® assay. An appropriate enzyme concentration was obtained in the function of time and initial rates were obtained. Kinetics parameters such as *V*ₘₐₓ/*K*ₐₙₜ and *K*ₐₚ were generated using either Lineweaver–Burk or Michaelis–Menten equation (ESI Fig. 13–16†).

(b) **GAR substrate specificity**

(1) **Acceptor screening result.** The enzyme activity towards acceptor compounds with UDPGlc as the donor using above MS/MS method was summarised in this GAR mode (Fig. 1a). The MS and MS/MS spectra were indicated in ESI Fig. 3–6.†

In this study, we have reported the acceptor substrate profile of UGTs 76E1, 76E2, 76E5 and 76D1 (Fig. 1b). Of the forty-two acceptor substrates screened, 76E1 recognised nine flavonoids and one coumarin. 76E2 only accepted avonoids and one coumarin. 76E5 and 76D1 recognised avonoids and one benzoic acid. Although there seem to exist an avonoid trend, recognition of benzoic acid by 76E5 and 76D1 was done in three stages of exponential amplification (polymerase chain reaction), digestion and transformation. Mutant DNAs were sent for sanger sequencing (SourceBioScience) to confirm the reaction product (Fig. 1a). This method, similar to the previous NLCT, could be used to determine the glycosylation site. No difference was noticed in the reaction products from 3 h and 15 h assay incubation. In addition, pseudo single substrate enzyme kinetics was measured by varying one substrate concentrations while saturating the other substrate using a commercial UDP-Glo® assay. An appropriate enzyme concentration was obtained prior to the kinetic study with fixed substrate concentrations (UDPGlc at 100 μM and KMP at 20 μM) while enzyme concentration varied. The formation of UDP was monitored in the function of time and initial rates were obtained. Kinetics parameters such as *V*ₘₐₓ/*K*ₐₚ and *K*ₐₚ were generated using either Lineweaver–Burk or Michaelis–Menten equation (ESI Fig. 13–16†).

**GAR substrate specificity**

(1) **Acceptor screening result.** The enzyme activity towards acceptor compounds with UDPGlc as the donor using above MS/MS method was summarised in this GAR mode (Fig. 1a). The MS and MS/MS spectra were indicated in ESI Fig. 3–6.†

In this study, we have reported the acceptor substrate profile of UGTs 76E1, 76E2, 76E5 and 76D1 (Fig. 1b). Of the forty-two acceptor substrates screened, 76E1 recognised nine flavonoids and one coumarin. 76E2 only accepted flavonoids as substrates while 76E5 on the other hand, recognised one benzoic acid, one coumarin and two flavonoids. Similarly, 76D1 accepted two flavonoids and one benzoic acid. Although there seem to exist a flavonoid trend, recognition of benzoic acid by 76E5 and 76D1 and non-recognition of other acceptor substrate class reinforces the established finding that substrate recognition cannot be predicted solely based on phylogenetic grouping.†

Our results showed that although acceptor compounds from varying classes were recognised, flavonoids appeared to be a common factor. According to Parajuli et al., expanded substrate study of a glycosyltransferase ATUGT99C1 revealed that it recognised five different classes of flavonoids and varying
NDP-sugars. Another Arabidopsis thaliana glycosyltransferases, AtUGT78D1 glycosylated several compounds of structural types; flavonoids, flavonoid glycosides and others.\textsuperscript{26} Furthermore, Jones et al.\textsuperscript{10} reported flavonoid glucoside/rhamnoside production from UGT73C6. All these studies corroborated that AtUGTs seemed to generally favour flavonoids O-glycosylation. Flavonoide glycosides possess promising beneficial health activities\textsuperscript{27} and regioselective glycosyltransferases offer to overcome limitations in their pharmacological potency.\textsuperscript{28} This implies that UGTs tested in group H may be useful in producing useful flavonoids glycosides.

(2) Donor screening result and glycosylation position. The donor screening was performed with two acceptors, kaempferol and quercetin which were confirmed to be the substrate from the acceptor result (Fig. 1). The MS and MS/MS spectra were indicated in the ESI 7–10.\textsuperscript{†} The spectral data is summarised in the GAR screen below (Fig. 2). ESI Table 2\textsuperscript{†} highlighted the precursor and product ions used for MS/MS screen. It was noticed that UDPGlc was the most useful donor but some enzymes did utilise more than one donor compound. 76E1 and 76E5 used UDP Gal and UDPGlcNAc in addition to UDP Glc. 76D1 was reported to use UDPGlc (and UDPXyl) in plant defence in vivo by glucosylation of DHBAs.\textsuperscript{18} This explains why 76D1 only recognised UDPGlc of all donor sugar compounds we screened.

Flavonoids have multiple glycosylation position which can change the product physical and biological activity. Mass spectra have been found to give information on structural characterisation, especially glycosylation position. Most sugar molecules in flavonoid O-glycosides bind at 3-, 7- or 4′- positions (Fig. 1c), leaving others such as 5-OH which show very low activity to glycosylation.\textsuperscript{29} Differentiating between 3-, 7- and 4′ – positions can be quite problematic; however, MS has shown advantage to provide possible solutions. In the MS2 fragment of a glycosylated flavonoid, radical aglycone ions ([M–H–H–Gly]\textsuperscript{−}, 284/300) are found to be more abundant for flavonol 3-O glycosides than regular aglycone ions ([M–H–Gly]\textsuperscript{−}, 285/301) which leaves 284 (kaempferol) or 300 (quercetin) instead of 285 or 301 (Fig. 3a). The formation of a radical aglycone ion is due to the homolytic cleavage of the glycosidic bond between the O-linked sugar and the aglycone moiety.\textsuperscript{29}

The presence of peak 179 (designated as the diagnostic ion, Fig. 3b) in retrocyclisation pathway involving loss of B-ring indicates specificity of 3′ or 4′ derivatives in negative ion ESI-MS/MS.\textsuperscript{30} Fig. 3 showed a few other examples of using the above
The presence of peak 179, a diagnostic ion for 4-O(284.9) and the presence of fragment of 179.0. (c) An example to show the 7-O glycosylation showing the glycosylation was not at the 3-O position. Since peak 301 is more prominent than radical aglycone ion 300, it will either be glycosylation at 7-O or 4-O position. It will either be glycosylation at 7-O or 4'-O position. The presence of peak 179, a diagnostic ion for 4'-O glycosylation suggests that it might be kaempferol 4'-O glucoside (Fig. 3b). Another example showing a glycosylation position was seen in a quercetin glycoside. Here, the regular aglycone ion peak 301 is more prominent than radical aglycone ion 300, showing the glycosylation was not at the 3-O position. Since peak 179 is absent here, we infer that the product was a quercetin 7-O glucoside (Fig. 3c).

From the acceptor screening (Fig. 2) it indicated that 76E1 can be more flexible in accommodating acceptors (either flat or curved). The result indicates the glycosylation position can be on 7-OH but no further information can be assured without NMR data. Interestingly, 76E1 is also very flexible in donor binding pocket with three (3) UDP-sugars used as substrate, but all on the 7-O position. 76E2 performed similarly to 76E1, glucosylating quercetin at 7-OH position. Similar flexibility was demonstrated for 76E5 in its donor binding pocket as 76E1 although 76E1 can utilise little more acceptors than 76E5. When UDPGlc was used as donor, it glycosylated kaempferol at 3-OH but with UDPGlcNAc as donor, the glycosylation position is on 4' or 7-OH. The acceptor screening prediction is challenging as it involves non-flavonoids (Table 1).

(3) **Protein sequence analysis and mutation studies.** Identification of residues responsible for protein function via protein sequence analysis is an important and widely studied concern. This is because it guides experimental analysis and gives insight to function prediction. Multiple sequence alignment (MSA) can expose sites that are conserved which can indicate residues of functional importance involved in binding. Comparing the MSA of the four studied UGTs showed few differences which seems to be in line with the donor recognised (Fig. 4). Serine (at points 311 and 318 in 76E1 and 76E5 respectively) is present in both UGTs recognising more donor compounds; UDP Gal and UDPGlcNAc. This residue is absent in 76D1 and 76E2 with no UDP Gal and UDPGlcNAc activity. However, 76D1 and 76E2 both has asparagine (N) at this point (circled in Fig. 4a). In our previous research (unpublished), we observed UDP Gal and UDPGlcNAc activity for 76E11 and 76E12, both of which contain serine at the same point as shown in Fig. 4b. The presence of serine at this point in two of studied UGTS with UDPGal and UDPGlcNAc activity from the same family, and its absence in two UGTs with no UDPGal and UDPGlcNAc activity could point at something important. Mutation of serine (S) to asparagine (N) at this site in 76E1 and 76E5 was done to establish the importance of serine in the enzyme substrate recognition, particularly if UDPGal and UDPGlcNAc activity will be lost, reduced or retained. In addition, mutation of asparagine (N) to serine (S) was done to observe if UDPGal and UDPGlcNAc activity will be acquired. Hence, mutants 76E1 S318N, 76E5 S311N and 76E2 N320S were made.

Mass spectra from UGT assay catalysed by mutant enzymes 76E1 S318N, 76E5 S311N and 76E2 N320S are shown in ESI Fig. 18–24.† Total loss of glycosylation activity (UDPGlc, UDPGal and UDPGlcNAc) was observed in 76E1 S318N and 76E5 S311N as summarised in Fig. 4c. This indeed demonstrates the strong link between these activities and serine at this point. More interestingly, UDPGal and UDPGlcNAc activity was acquired by 76E2 N320S; the WT enzyme 76E2 only has UDPGlc activity (Fig. 4c). Although the exact mechanism by which serine acts is still unclear, this finding confirms our hypothesis that serine at this point is important in the UDPGlcNAc and UDPgal activity.

(4) **Kinetics data.** As shown in the plots Michaelis–Menten and Lineweaver–Burk (Table 2, ESI Fig. 26–29f), under the current condition, all the enzymes showed relatively low $K_M$ at $\mu$M scale. A low $K_M$ indicates that an enzyme requires a small amount of substrate to become saturated. Hence, the maximum velocity is reached at relatively low substrate concentrations. Kinetics data revealed relatively low $K_M$ values, suggesting

<table>
<thead>
<tr>
<th>Enzymes</th>
<th>Donor</th>
<th>Acceptor</th>
<th>Peaks used in identification</th>
<th>Glycosylation position</th>
</tr>
</thead>
<tbody>
<tr>
<td>76E1</td>
<td>UDPGlc</td>
<td>Quercetin</td>
<td>285</td>
<td>7-OH</td>
</tr>
<tr>
<td>76E1</td>
<td>UDPGlc</td>
<td>Kaempferol</td>
<td>301</td>
<td>7-OH</td>
</tr>
<tr>
<td>76E2</td>
<td>UDPGlc</td>
<td>Quercetin</td>
<td>301</td>
<td>7-OH</td>
</tr>
<tr>
<td>76E5</td>
<td>UDPGlc</td>
<td>Kaempferol</td>
<td>284</td>
<td>3-OH</td>
</tr>
<tr>
<td>76E5</td>
<td>UDPGlc</td>
<td>Quercetin</td>
<td>300</td>
<td>3-OH</td>
</tr>
<tr>
<td>76E5</td>
<td>UDPGlc</td>
<td>Kaempferol</td>
<td>284</td>
<td>3-OH</td>
</tr>
<tr>
<td>76E5</td>
<td>UDPGlc</td>
<td>Quercetin</td>
<td>300</td>
<td>3-OH</td>
</tr>
<tr>
<td>76D1</td>
<td>UDPGlc</td>
<td>Kaempferol</td>
<td>285, 179</td>
<td>4'-OH</td>
</tr>
<tr>
<td>76D1</td>
<td>UDPGlc</td>
<td>Quercetin</td>
<td>301</td>
<td>7-OH</td>
</tr>
<tr>
<td>76D1</td>
<td>UDPGlc</td>
<td>Quercetin</td>
<td>301</td>
<td>7-OH</td>
</tr>
</tbody>
</table>

Fig. 3 The MS/MS method to determine the formation of product and potentially glycosylation position. (a) An example to show the 3-O glycosylation – the formation of radical aglycone (283.8). (b) An example to show the glycosylation on 4'-O position – the formation of aglycone (284.9) and the presence of fragment of 179.0. (c) An example to show the 7-O glycosylation – the formation of aglycone (300.8) only.
enzymes have good affinity for the substrates UDPGlc, UDPGal and UDPGlcNAc. All 76E1, E2 and E5 enzymes share very similar $K_M$/UDPgluc values (~10 μM), which is a measure of enzymes’ substrate preference and catalytic efficiency. However, 76E5 showed strong preferences to both UDPGal and UDPGlcNAc, with low sub μM $K_M$ and much higher $K_{cat}/K_M$ (> 20 s⁻¹ mM⁻¹). This kinetic data strongly suggests that 76E5 is probably more of UDPGlcNAc and UDPGal transferase than UDPglucosyltransferase. This is probably also the case in vivo. 76D1 has much high $K_M$ values (> 25 μM) and low $K_{cat}/K_M$ (0.185 s⁻¹ mM⁻¹). 76E1 and 76E2 are the only group H members previous characterised, which are plant hormone GTs. The kinetic parameters are with respect to the plant hormone substrates; hence unsuitable for comparison with our studied UGTs.

There is no other kinetics data available for group H members for comparison. When compared with VvGT1 ($K_{cat}/K_M$ 0.124 s⁻¹ mM⁻¹), our UGTs $K_{cat}/K_M$ values are in the same range for substrate UDPGlc. 76E5 has a far higher $K_{cat}/K_M$ with substrates UDP Gal and UDPGlcNAc (>2000 fold) than VvGT1; affirming that UDP Gal and UDPGlcNAc seems to be natural substrates to 76E5. Functional studies for 76E5 in vivo will reveal more in the near future.

**Conclusion**

Here, we have reported the substrate specificities and kinetics of four UGTs in group H of Arabidopsis thaliana (76E1, 76E2, 76E5 and 76D1). These UGTs added sugars to acceptors from different classes of secondary metabolites ranging from flavonoids, cinnamic acids, coumarins to alkaloids. They also recognised varying donors tested, with all recognising UDP glucose and 76E1 and 76E5 binding to UDP-galactose and UDPGlcNAc in addition. Using tandem mass spectrometry, the likely glycosylation positions on the acceptors were proposed. Further kinetic study measured each enzyme binding affinity with donor substrates, indicating their donor preferences and catalytic efficiencies. The distinctive amino acid sequences of both UGTs recognising UDP Gal and UDP GlcNAc was observed. Mutagenesis highlighted residue serine, which seem to be important for UDPGlcNAc and UDPGal activity. Knowledge of the activity and specificity of UGTs would provide a basis for their potential use to design improved UGTs to produce bioactive molecules for pharmaceutical purposes.

**Conflicts of interest**

There are no conflicts to declare.

**Acknowledgements**

This work was funded by the Nigerian Government, under the Presidential Special Scholarship Scheme for Innovation and Development (PRESSID) supervised by the Federal Scholarship Board (FSB).

**References**
