Chemical Probing of Glycogenesis



A thesis submitted to the Board of the Faculty of Physical Sciences in partial fulfilment of the requirements for the degree of Doctor of Philosophy at the University of Oxford

Matthew Bilyard

New College, University of Oxford

Trinity Term 2018

Author Declaration

The work presented in this thesis was carried out at the Chemistry Research Laboratory, University of Oxford, UK under the supervision of Professor Benjamin Davis. All the work contained herein is my own, unless clearly stated otherwise (see "collaborations", page xii), and has not been submitted for any other degree at this university or at any other.

Matthew Bilyard

Abstract

Chemical Probing of Glycogenesis

Matthew Bilyard, New College

Submitted towards the degree of D.Phil. (Organic Chemistry) in Trinity Term 2018

The polysaccharide molecule glycogen represents a crucial glucose (and hence energy) storage reserve present in a wide variety of organisms. In eukaryotes, its biosynthesis is initiated by the glycosyltransferase enzyme Glycogenin, which unusually catalyses its own autoglucosylation to form a short maltosaccharide extending from a tyrosine residue. Glycogenin's privileged position at the heart of carbohydrate metabolism has sparked wide interest in elucidating its catalytic mechanism. However, prior studies have been hindered by an inability to access homogeneous glycoforms of the enzyme, restricting precise mechanistic analyses.

Suzuki-Miyaura cross-coupling represents a widely utilised reaction in organic chemistry and has more recently emerged as a powerful tool for the site-selective modification of proteins. In this thesis, we present the application of this methodology to the mechanistic study of Glycogenin, utilising a "genetically tagged" variant of the latter bearing an unnatural *p*-iodophenylalanine residue in place of the catalytic tyrosine. Preliminary studies demonstrate the potential of this enzyme to undergo efficient cross-coupling under benign, "enzyme-compatible" conditions. This leads to the application of this reaction as part of a novel palladium-mediated enzyme activation process for Glycogenin, allowing unprecedented access to different, homogeneously glucosylated enzyme states. These closely mimic intermediates at different stages of the autoglucosylation pathway, and are therefore catalytically active.

Comparative kinetic analyses of such Glycogenin constructs reveal unusual triphasic kinetics for the initiation of glycogen biosynthesis. Subsequent studies on both a non-natural autogalactosylation pathway and on autoglucosylation pathways initiating from non-glucose carbohydrate glycosyl acceptors reveal a surprising degree of enzyme plasticity towards glycosyl donor and acceptor, and suggest that the three kinetic phases may be additionally defined in this regard.

We also apply our methodology to investigation of a Glycogenin mutation implicated in a pathological condition arising from impaired autoglucosylation. Comparison of previously inaccessible glycoforms of this mutant challenges prior beliefs that it is entirely incapable of autoglucosylation and highlights further differences between early catalytic phases of catalysis. We anticipate that our approach may in the future enable mechanistic investigation of the other metabolic disorders and neurodegenerative conditions associated with malfunctions in glycogen biosynthesis.

Acknowledgements

There are a number of people without whose help, guidance and support this thesis would almost certainly not exist. Firstly, I would like to thank my supervisor, Prof. Ben Davis, for giving me the opportunity to work in his group for the best part of 4 years on a fascinating research project. I enjoyed significant freedom to develop the project as I wished, whilst equally appreciated his guidance when needed. I would also like to thank my collaborator at the Structural Genomics Consortium, Prof. Wyatt Yue, for his regular input and advice, providing different and interesting perspectives on my work, and also for his generous donation of plasmids. In particular, I am extremely grateful to the large body of work carried out by his DPhil student, Henry Bailey; I will be forever grateful for the incredible amount of protein expression work you carried out for this project.

I would like to thank my examiners, Prof. Hagan Bayley and Prof. Steve Withers, for taking the (considerable) time to read this thesis and for a very interesting viva with lots of helpful suggestions. A special thanks to Prof. Withers for flying over from western Canada for this!

Much of the work described herein would have been substantially more challenging than it already was without the excellent work carried out on Suzuki-Miyaura cross-coupling by previous group members. In this regard, I extend my thanks to Dr. Justin Chalker, Dr. Anaelle Dumas and especially Dr. Chris Spicer, who provided much useful guidance in the early stages of this work.

The work in this thesis owes much to numerous departmental staff. In particular, I thank Dr. James Wickens and the mass spectrometry team for putting up with my incessant questions and demands in this area. Equally, I am indebted to the NMR team: Prof. Tim Claridge, Dr. Barbara Odell, and Tina Jackson. I regularly placed the toughest demands on their NMR service (I did really appreciate that ^{13}C NMR on less than 1 mg of hexasaccharide…), and they always delivered. I must apologise to Barbara for giving her a particularly terrible experiment on her final day at work in the department. If it is any consolation, it really was useful!

There are a lot of people in the research group, past and present, who I would like to thank for a number of reasons. Firstly, I am grateful to all of you who took the time to proofread sections of this thesis: Sadra Hamedzadeh, Dr. Mylène Richard, Dr. Charlie Fehl, Sarah Faulkner, and (especially, for voluntarily reading 2 chapters) Dr. Jitka Dadova. I owe you all drinks (and if this is the case when you read this, do remind me). Secondly, I thank any of you who have either helped me with LC-MS issues or listened to me moan

about them (special mention here to Dr. Seb Galan, you know why all too well…). Thirdly, thanks to Dr. Anuchit Phanumartwiwath (I do hope I spelled it correctly this time…) for essentially teaching me practical biology from scratch when I joined. Finally, I'd simply like to thank all the members of the group during my time there, for making the group such a fun place to be both in and out of work. Pretty much everyone here has helped me out at some point, too many to thank individually so I'm afraid you'll have to make do with a general thank you! In addition, I've been lucky to meet a number of great people via my involvement with New College and through music, and I thank you guys for reminding me of the life that exists outside of science!

Of course, I extend massive thanks to my family for their unwavering support during my time in Oxford, through good times and bad. A special mention to my grandfather, for representing in recent years an exceptional example of how to just pick yourself up and carry on in the face of tough situations, and to my late grandmother – you're not forgotten, ever, and I hope you would have been proud. And of course, I would like to say an enormous thank you to my better half, Tanja. You, more than anyone, have been there throughout and supported me always, whilst at the same time dealing with your own DPhil, and without this I may well have not reached this point!

Table of Contents

Author Declaration i

Abstract ii

Acknowledgements iii

Table of Contents v

List of Abbreviations viii

Collaborations xii

Publications and Presentations xii

1. Introduction 1

1.1 Glycogen: an Energy Storage Molecule 3

1.2 Glycogenin: an Unusual Autocatalytic Glycosyltransferase 12

1.3 Strategies to Access Homogeneous Glycoproteins 25

1.4 Site-selective Chemical Modification Approaches to Glycoprotein Synthesis 38

1.5 Suzuki-Miyaura Cross-coupling: from Synthesis to Biology 49

1.6 Aims of this Thesis 54

1.7 References 56

2. Suzuki-Miyaura Cross-coupling to Glycogenin-1 75

2.1 Introduction 78

2.2 Site-specific Incorporation of *p*IPhe into Glycogenin-1 82

2.3 Suzuki-Miyaura Cross-coupling: Preliminary Studies 84

2.4 GYG-3HMPB: a Putative "Minimal Glycomimetic" 87

2.5 Development of an Enzyme-compatible Cross-coupling 92

2.6 Investigating the Activity of Minimal GYG Glycomimetics 97

2.7 Conclusion 102

2.8 Experimental Section 103

2.9 References 162

3. Palladium-mediated Activation of Glycogenin-1 166

3.1 Introduction 168

3.2 Synthesis of a Glucose-derived Boronic Acid 174

3.3 Pd-mediated Enzyme Activation 176

3.4 Conclusion and Outlook 189

3.5 Experimental Section 190

3.6 References 228

4. Comparison of GYG Glycoform Mimetics Reveals a Triphasic Mechanism 232

4.1 Introduction 234

4.2 Generation of an Active GYG-disaccharide Conjugate 237

4.3 Investigating the Catalytic Activity of Early GYG Glycoforms 242

4.4 Conclusion and Outlook 259

4.5 Experimental Section 260

4.6 References 301

5. Investigating a GYG-oligosaccharide Conjugate 304

5.1 Introduction 306

5.2 Synthesis of a Hexasaccharide Coupling Partner 308

5.3 Investigating the Generation of GYG-Maltohexaose 312

5.4 Catalytic Activity of GYG-Maltohexaose 318

5.5 Conclusion and Outlook 327

5.6 Experimental Section 329

5.7 References 358

6. Exploring Unnatural GYG Glycoforms 361

6.1 Introduction 363

6.2 Investigating Unnatural GYG-Carbohydrate Acceptors 370

6.3 Investigation of Catalytically Impaired GYG-T83M Glycoforms 380

6.4 Conclusion and Outlook 387

6.5 Experimental Section 388

6.6 References 426

7. Conclusion and Perspectives 430

7.1 Summary of Findings 432

7.2 Discussion 435

7.3 Future Perspectives 438

7.4 References 442

Appendix 445

1. General Procedures for Small Molecule Chemistry 446

2. General Procedures for Biological Work and Protein Modification 448

3. Protein LC-MS 449

List of Abbreviations

2-HMPB	2-(hydroxymethyl)phenylboronic acid cyclic monoester
3-HMPB	3-(hydroxymethyl)phenylboronic acid
3-HPB	3-hydroxyphenylboronic acid
4-HMPB	4-(hydroxymethyl)phenylboronic acid
4-HPB	4-hydroxyphenylboronic acid
Ac	Acetyl
ADP	Adenosine diphosphate
AGL	Glycogen debranching enzyme
AMP	Adenosine monophosphate
ATP	Adenosine triphosphate
aaRS	Aminoacyl tRNA synthetase
aq.	Aqueous
Aha	Azidohomoalanine
C	Celsius
CHO	Chinese hamster ovary
CD	Circular dichroism
CuAAC	Copper-catalysed azide-alkyne cycloaddition
cAMP	Cyclic adenosine monophosphate
Da	Dalton(s)
DEA	Diethanolamine
Dha	Dehydroalanine
DNA	Deoxyribonucleic acid
DCM	Dichloromethane
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
E. coli	*Escherichia Coli*
EDTA	Ethylenediaminetetraacetic acid
Endo	Endo-β-*N*-acetylglucosaminidase

eq.	Equivalents
ESI	Electrospray ionisation
g (in context of masses)	Gram(s)
g (in context of centrifugation speed)	Gravitational constant
G1P	Glucose-1-phosphate
G6P	Glucose-6-phosphate
Gal	Galactose
GBE	Glycogen branching enzyme
GDP	Guanosine diphosphate
Glc	Glucose
GlcNAc	*N*-acetylglucosamine
GlgA	Bacterial homolog of GYS
GlgB	Bacterial homolog of GBE
GLUT	Glucose transporter
GNIP	Glycogenin interacting protein
GP	Glycogen phosphorylase
GT	Glycosyltransferase
GYG	Glycogenin
GYG-1	Glycogenin-1
GYG-2	Glycogenin-2
GYG-wt	Wild-type Glycogenin
GYS	Glycogen Synthase
GYS-1	Glycogen Synthase 1
GYS-2	Glycogen Synthase 2
h	Hour(s)
HEK	Human embryonic kidney
HIV	Human immunodeficiency virus
HRMS	High resolution mass spectrometry
Hz	Hertz

IEDDA	Inverse electron demand Diels-Alder
IgG	Immunoglobulin G
IPTG	Isopropyl β-D-thiogalactopyranoside
IR	Infrared
J	Coupling constant
kDa	Kilodalton(s)
l	Litre
LB	Lysogeny broth
LC-MS	Liquid chromatography – mass spectrometry
LRMS	Low resolution mass spectrometry
lit.	Literature
M	Molar
M. jannaschii	*Methanococcus jannaschii*
m.p.	Melting point
Man	Mannose
MBP	Maltose binding protein
mg	Milligram(s)
MHz	Megahertz
MIDA	*N*-methyliminodiacetic acid
ml	Millilitre(s)
mM	Millimolar
mmol	Millimole(s)
mol	Mole(s)
μM	Micromolar
μmol	Micromole(s)
NCL	Native chemical ligation
NDP	Nucleoside diphosphate
NHC	*N*-heterocyclic carbene
nm	Nanometre(s)
nM	Nanomolar

nmol	Nanomole(s)
OD_{600}	Optical Density at 600 nm
OmpC	Outer membrane protein C
PBS	Phosphate buffered saline
PEG	Polyethylene glycol
PGM	Phosphoglucomutase
P_i	Inorganic phosphate
Pin	Pinacol
*p*IPhe	*p*-iodophenylalanine
pp.	Pages
ppm	Parts per million
PTM	Post-translational modification
RNA	Ribonucleic acid
rpm	Revolutions per minute
r.t.	Room temperature
s	Seconds
SAC	*S*-allylcysteine
SB	Super broth
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SMC	Suzuki-Miyaura cross-coupling
SPAAC	Strain promoted azide-alkyne cycloaddition
SPPS	Solid phase peptide synthesis
t	Time
TBS	*tert*-Butyldimethylsilyl
tRNA	Transfer-ribonucleic acid
UAA	Unnatural amino acid
UDP	Uridine diphosphate
UDPG-P	Uridine diphosphate glucose pyrophosphorylase
UDPG-PPase	Uridine diphosphate glucose pyrophosphatase
wt	Wild type

Collaborations

Chapter 2: Expression of his-tagged GYG-Y195*p*IPhe was performed in collaboration with Mr. Henry Bailey, a DPhil student in the research group of Prof. Wyatt Yue (Structural Genomics Consortium, University of Oxford). Specifically, smaller-scale batches were carried out by myself at the Chemistry Research Laboratory, and one larger-scale batch was carried out by Henry.

Plasmids encoding GYG-wt and GYG-Y195F were generously donated by Prof. Wyatt Yue. Plasmid GYG-Y195X was previously made in our laboratory (using GYG-wt as template) by Mr. Gustaf Hemberg, a current DPhil and former MSci student in the Ben Davis group.

Chapter 3: TEV protease cleavage of his-tagged GYG-Y195*p*IPhe was performed in collaboration with Mr. Henry Bailey, a DPhil student in the research group of Prof. Wyatt Yue (Structural Genomics Consortium, University of Oxford).

Chapter 7: Any mentions of ongoing computational work are being carried out at present by collaborators Prof. Carme Rovira, Mr. Lluis Raich and Dr. Javier Iglesias-Fernandez, University of Barcelona, as part of the manuscript under revision listed below. This manuscript is referenced when mentioned.

Publications and Presentations

Publications

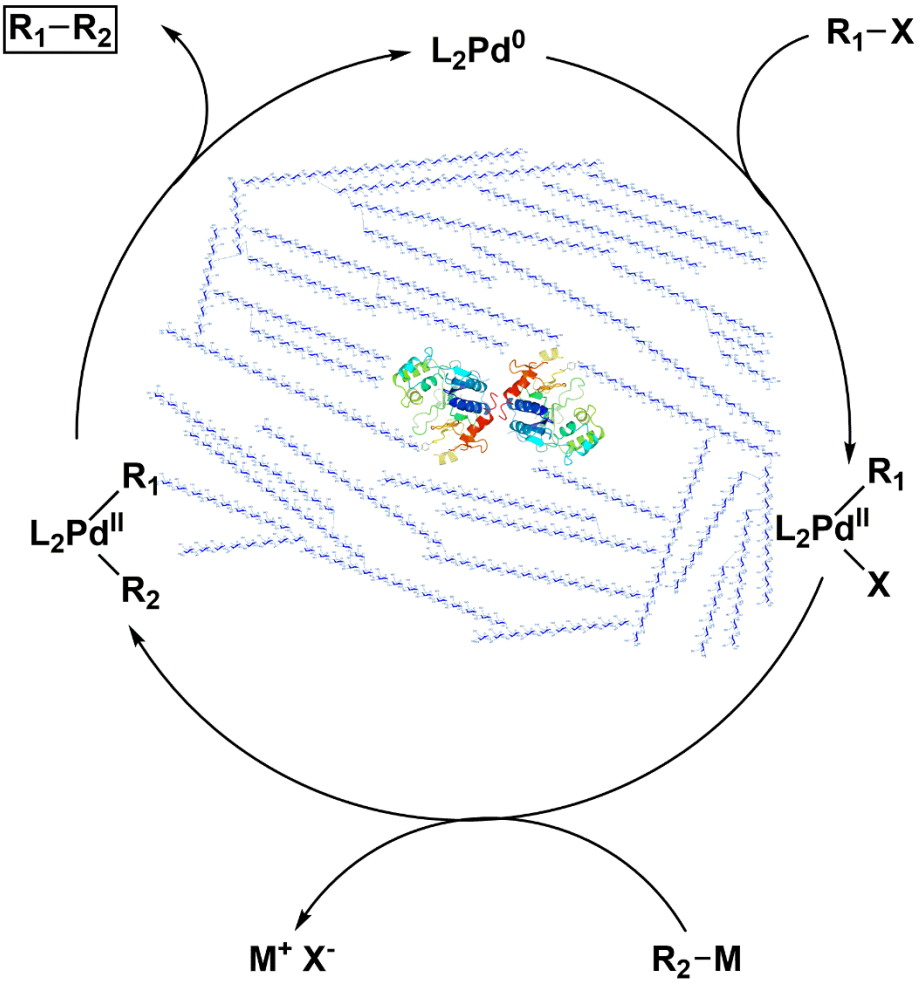
Bilyard, M.K., *et al*, "Palladium-mediated enzyme activation suggests multiphase initiation of glycogenesis." *Manuscript under revision (Nature)*, 2018.

Presentations

28th International Carbohydrate Symposium, New Orleans, *2016*. Oral presentation: "Bilyard, M.K., Yue, W.W. and Davis, B.G. Probing Glycogen Biosynthesis via a Palladium-mediated Enzyme Switch"

Eli Lilly CASE day, Erl Wood R&D site, Windlesham, UK, *2014*. Poster presentation: "Bilyard, M.K., Yue, W.W. and Davis, B.G. Investigating the Catalytic Mechanism of Glycogenin through Pallado-biology"

Chapter 1



Introduction

Contents

1.1 Glycogen: an Energy Storage Molecule 3

1.1.1 The Structure and Function of Glycogen 3

1.1.2 Glycogen Biosynthesis and Degradation: Glycogenesis and Glycogenolysis 5

1.1.3 GYG, GYS and GBE 9

1.2 Glycogenin: an Unusual Autocatalytic Glycosyltransferase 12

1.2.1 Common Mechanisms of Glycosyltransferases 12

1.2.2 Catalytic Mechanism of Glycogenin 15

1.2.3 Challenges in the Mechanistic Study of GYG: the Need for New Methods 21

1.3 Strategies to Access Homogeneous Glycoproteins 25

1.3.1 Chemical Synthesis of Glycoproteins 25

1.3.2 Drawing Inspiration from Nature: Enzymatic Extension and Glycoprotein Remodelling 29

1.3.3 Genetic Incorporation: Codon Suppression Technology 31

1.3.4 Engineering Glycosylation Pathways 33

1.3.5 Site-selective Chemical Modification of Proteins 34

1.4 Site-selective Chemical Modification Approaches to Glycoprotein Synthesis 38

1.4.1 Cysteine as a Tag for Homogeneous Glycoprotein Generation 39

1.4.2 Tags Involving Cysteine Derivatives 39

1.4.3 Unnatural Amino Acids as Tags for Glycoprotein Synthesis 42

1.4.4 The Choice of a Suitable Strategy for Glycogenin Modification 46

1.5 Suzuki-Miyaura Cross-coupling: from Synthesis to Biology 49

1.5.1 Suzuki-Miyaura Cross-coupling: An Overview 49

1.5.2 From Organic Synthesis to Chemical Biology 53

1.6 Aims of this Thesis 54

1.7 References 56

1.1 Glycogen: an Energy Storage Molecule

Across all living organisms, there is a fundamental requirement for energy in order to carry out and maintain biological processes needed for their survival and reproduction.[1] Much of this energy arises from food molecules (especially sugars), which act as "cellular fuel molecules" in that their controlled oxidation leads to a release of free energy that is transferred to convenient activated carrier molecules, notably ATP (a universal "currency of energy").[2] Energetically favourable hydrolysis of the latter is in turn used to drive a range of energetically unfavourable processes, for example protein biosynthesis, or active transport.[2] Autotrophic organisms are capable of synthesising such food molecules from simple environmentally-available substances, using energy from either sunlight (photoautotrophs) or inorganic chemical reactions (chemoautotrophs). Heterotrophs, in contrast, must ultimately obtain these molecules second-hand from autotrophic species.[3]

At various times, however, the availability of food molecules to serve as fuel does not match energy demand; for example, plants must survive overnight when they cannot use sunlight to photosynthesise such species, whilst animals have only periodic access to food.[2] This fact explains, and indeed necessitates, the existence of molecules that serve as longer-term stores of metabolic energy. In times when availability of fuel molecules exceeds current energy demands, organisms biosynthesise such storage molecules, which are rich in potential energy in the form of chemical bonds. They may then be disassembled to release energy when demand for the latter exceeds the intake of food molecules.[2]

Examples of molecules that serve this function include triacylglycerols, which act as long-term energy stores in both animals and plants, and carbohydrate polymers, notably starch (in plants) and glycogen, which represent more readily mobilised, shorter-term energy reserves.[2,4–7] It is the biosynthesis of this latter molecule that forms the backdrop to the work described in this thesis, and as such the remainder of this section explores glycogen, its degradation and (principally) its biosynthesis in greater depth.

1.1.1 The Structure and Function of Glycogen

The vast proportion of glycogen consists of a branched polysaccharide comprising exclusively glucose monomers; glycogen may thus ultimately be assembled from, and disassembled to, free glucose, and it therefore represents a key mechanism of glucose storage across a wide variety of organisms, from humans to yeast to bacteria.[8,9] Through careful physiological regulation, glycogen is selectively assembled during

times of glucose abundance and disassembled during times of glucose depletion; these pathways are referred to as glycogenesis and glycogenolysis respectively.[5] Glucose represents a particularly crucial cellular fuel molecule; it feeds directly into the glycolysis pathway and thence into the TCA cycle and oxidative phosphorylation pathway, enabling biosynthesis of significant quantities of ATP.[2,5,10] As such, it is apparent that glycogen represents a readily mobilised intracellular energy reserve.[7]

In mammals, glycogen deposits are found in a variety of tissue types, however it is at its most abundant in liver and skeletal muscle.[8,11] In the former, glycogen metabolism plays a key role in maintaining blood glucose levels in accordance with the energy requirements of the organism as a whole, ensuring that cells in different organisms receive sufficient quantities of glucose as fuel. This is particularly crucial for the brain given that, in contrast to other organs, glucose represents virtually its sole fuel (and it lacks significant glycogen deposits).[10,12] In muscle, by contrast, glycogen serves as the primary source of glucose (and thus ATP) to serve the needs of muscle itself during exercise.[10]

In a structural sense, glycogen differs markedly from biological macromolecules such as proteins and nucleic acids; whereas the latter are assembled in a strictly templated fashion to yield well-defined products, glycogen particles form a naturally heterogeneous population.[8] Whilst this renders detailed structural characterisation complicated – crystallography, for example, is not an option – a number of features are nonetheless well-established. Fundamentally, the polysaccharide of glycogen comprises chains of α(1→4) linked glucose monomers, with additional α(1→6) linkages distributed in a non-uniform fashion providing "branch points", as depicted in **figure 1.1**.[13] This polysaccharide is covalently linked to a central protein core, known as Glycogenin (GYG); the role of this species will be described in significantly greater detail in subsequent sections, but essentially it serves as the initial anchor point from which the polysaccharide extends.[14] On average, an individual chain has a length of ca. 13 α(1→4) linked glucose residues;[8,15,16] the branch points serve as cross-links between different chains.

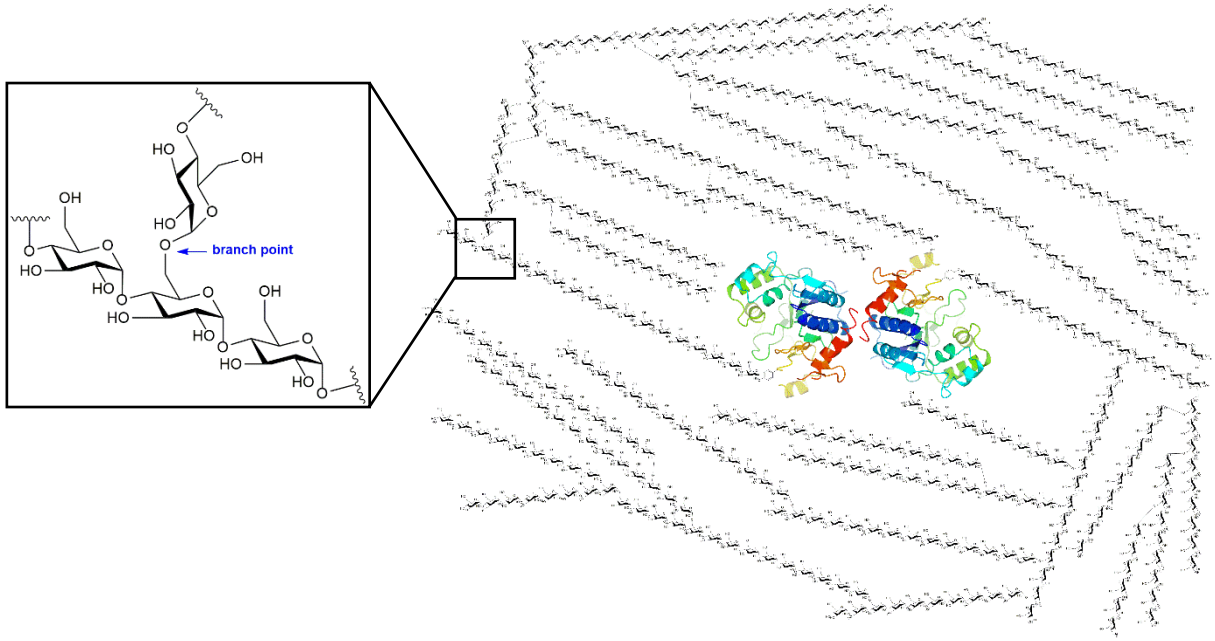


Figure 1.1 Schematic illustration of the structure of glycogen. Polysaccharide chains are linked to a protein core (Glycogenin), and comprise α(1→4) linked glucose residues, with additional α(1→6) linkages serving as branch points.

In a commonly accepted model, glycogen particles are often considered as consisting of two types of maltosaccharide chain: inner "B-chains" containing approximately two branch points per chain, and outer "A-chains" lacking any such branching (note that these are not shown on the schematic model in the figure above).[17–19] It should also be noted that, in addition to glucose monomers and the protein core, glycogen additionally consists of trace amounts of other species, in particular glucosamine and phosphate;[20,21] excessive levels of the latter have been implicated in the neurodegenerative condition Lafora disease.[22–25]

1.1.2 Glycogen Biosynthesis and Degradation: Glycogenesis and Glycogenolysis

In eukaryotes, the biosynthesis of glycogen – glycogenesis – can itself be further divided into two distinct metabolic pathways, referred to as the direct and indirect pathways dependent on the nature of the starting precursor.[26–28] In the former case, this is free glucose itself, transported into cells via glucose transporters (GLUTs) and subsequently converted by hexokinase to glucose-6-phosphate (G6P).[29] However, this intermediate may also be accessed via the gluconeogenesis pathway from precursors such as lactate, alanine and pyruvate; this constitutes the indirect pathway.[26,28] From this point onwards, the pathways merge. As such, through the actions of first phosphoglucomutase (PGM), then UDP-glucose pyrophosphorylase (UDPG-P), G6P is converted to the key glycosyl donor UDP-glucose (this may be recycled to glucose-1-

phosphate (G1P) and UMP through the action of UDP-glucose pyrophosphatase (UDPG-PPase)).[8] Polymerisation of this species into mature glycogen particles is catalysed first by GYG, and latterly by the downstream enzymes glycogen synthase (GYS) and glycogen branching enzyme (GBE) (figure 1.2).[8,29]

In a similar vein to glycogenesis, the catabolic pathway – glycogenolysis – may also be divided into two distinct pathways (**figure 1.2**, highlighted in red). In the more studied pathway, glycogen phosphorylase (GP) and glycogen debranching enzyme (AGL) catalyse the degradation of glycogen particles, releasing either glucose-1-phosphate (G1P) or free glucose directly from the degradation of α(1→4) or α(1→6) linkages respectively. G1P can either be converted to G6P for subsequent entry into other metabolic pathways (e.g. glycolysis, pentose phosphate), or may – in the liver only – be dephosphorylated to yield glucose.[5,29] A lesser studied glycogenolysis pathway is known to occur in lysosomes, and involve the direct hydrolysis of glycogen to glucose via a lysosomal α-glucosidase.[30]

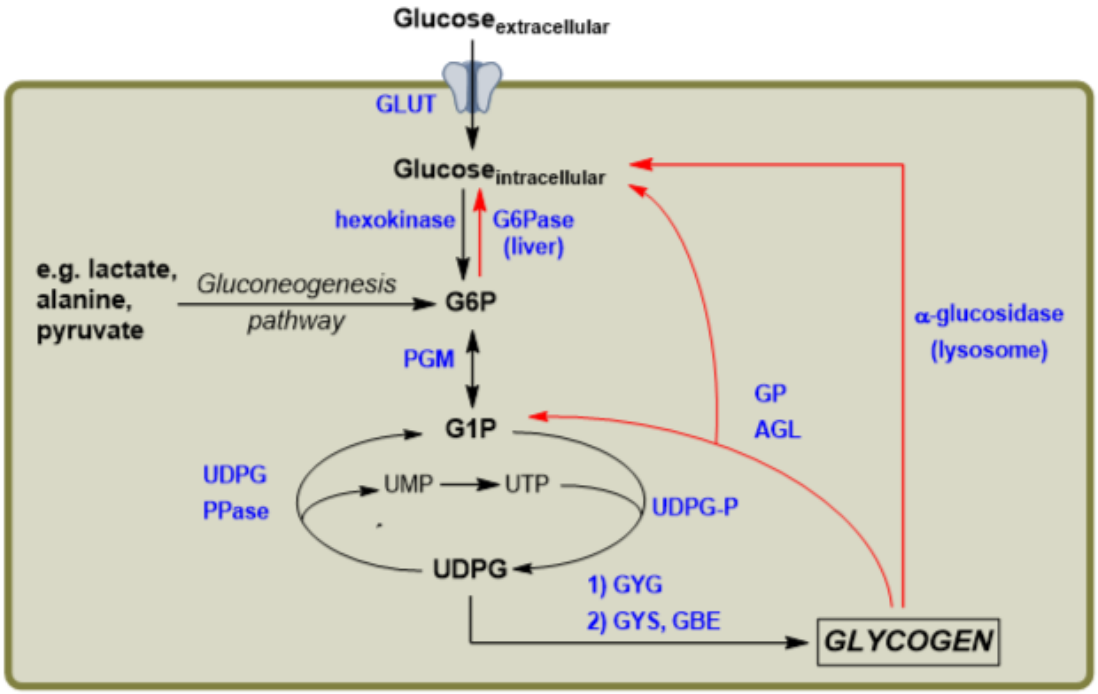


Figure 1.2 Biosynthesis and degradation of glycogen in eukaryotes. Arrows in red indicate the degradation (glycogenolysis) pathway. The full names for the compounds and enzymes abbreviated in the above figure can be found in the text above, and also in the list of abbreviations preceding this chapter.

It should be emphasised that the above description of glycogenesis and glycogenolysis refers primarily to the eukaryotic pathways; whilst the prokaryotic pathway is believed to be broadly similar, subtle differences are known.[9] Firstly, the nucleotide sugar donor is not UDP-glucose but ADP-glucose.[31] Secondly, no homolog of Glycogenin is known to exist in bacterial systems; the polymerisation of ADP-glucose is thus catalysed solely by Glycogen Synthase and Branching Enzyme (the bacterial versions are referred to as GlgA and GlgB respectively).[9,32] However, the work in this thesis pertains solely to (an enzyme involved in) eukaryotic glycogen metabolism, and as such the prokaryotic variant is not discussed further.

Glycogenesis and glycogenolysis are tightly and reciprocally regulated to ensure that glycogenesis is indeed favoured during times of glucose abundance, whilst glycogenolysis is favoured during times of glucose depletion. Such regulation is effected at different points in each pathway by both allosteric and hormone-mediated mechanisms. Allosteric control ensures the needs of the individual cell are met, whereas hormonal control caters for the demands of the organism as a whole.[11] A complete overview of the regulatory mechanisms involved in glycogenesis and glycogenolysis is beyond the scope of this thesis – this topic is comprehensively reviewed elsewhere[9–11,13,33] – however some illustrative examples (**tables 1.1, 1.2**, shown on the subsequent two pages) serve to highlight some basic principles.

As exemplified in **table 1.1** and **table 1.2**, hormonal regulation is exerted by insulin, glucagon (in liver) and epinephrine (in muscle), which primarily act by promoting dephosphorylation (insulin) or phosphorylation (glucagon, epinephrine, via a cAMP-dependent cascade) of relevant enzymes, switching them between inactive and active states.[10,34,35] Insulin is released in response to high blood glucose levels, thus favours glycogenesis; glucagon and epinephrine are both released to promote glycogen degradation, either in response to low blood glucose (glucagon) or as part of the "fight or flight" response (epinephrine).[10,13] Allosteric effectors principally act on glycogen synthase (GYS) and glycogen phosphorylase (GP), and include ATP and G6P, which act as signals of high glucose abundance and thus act to inhibit glycogenolysis and/or promote glycogenesis, and AMP, P_i and G1P, high levels of which signal a need for further glucose to fuel ATP production; these thus promote glycogenolysis.[36–38] Additionally, free glucose may act as an allosteric inhibitor of glycogenolysis in liver.[29]

A further source of regulation in muscle is provided by Ca^{2+} ions, which stimulate muscular contraction and are therefore indicative in high levels of a need for glucose in order to synthesise ATP as fuel for such contraction. As such, this ion inhibits GYS) through its phosphorylation (via activation of Calmodulin-dependent protein kinase) yet activates GP through binding to secondary effector proteins (e.g. Calmodulin).[5,39–41]

Requirement for more glucose: glycogenesis inhibited, glycogenolysis promoted.

Effector(s)	Enzyme affected	Effect	Physiological Outcome
AMP (muscle only), glucose-1-phosphate (G1P), P_i [35,37]	Glycogen Phosphorylase (GP)	Allosteric activators	**Promotes glycogenolysis**
Glucagon (liver only, released in response to low blood glucose) [34,35,37]	1) Glycogen Phosphorylase (GP)	Promotes phosphorylation → activation.	**Promotes glycogenolysis**
	2) Glycogen Synthase (GYS)	Promotes phosphorylation → inhibition.	**Inhibits glycogenesis**
Epinephrine (muscle, released as "fight or flight" response) [10,35,37]	1) Glycogen Phosphorylase (GP)	Promotes phosphorylation → activation.	**Promotes glycogenolysis**
	2) Glycogen Synthase (GYS)	Promotes phosphorylation → inhibition.	**Inhibits glycogenesis**
Calcium (muscle, utilised in muscular contraction)[5,39–41]	1) Glycogen Phosphorylase (GP)	Indirect allosteric effector (binds calcium-binding proteins that in turn bind GP).	**Promotes glycogenolysis**
	2) Glycogen Synthase (GYS)	Promotes phosphorylation → inhibition.	**Inhibits glycogenesis**

Table 1.1 Some factors in the regulation of glycogenesis and glycogenolysis during times of glucose need.

The focus of this thesis lies in the biosynthesis of glycogen, rather than its degradation, and as such the latter will not be discussed in detail any further. Furthermore, with reference to figure 1.2 above, the conversion of UDP-glucose to glycogen may be seen as the first "committed" step of glycogenesis, since prior to this step UDP-glucose may still have a number of alternative fates in the cell (e.g. conversion to UDP-galactose, UDP-glucuronic acid, use by other glycosyltransferase enzymes).[5] It was this step that attracted our attention, and the specific roles of the three enzymes involved are overviewed below.

Glucose abundance: glycogenesis promoted, glycogenolysis inhibited.

Effector	Enzyme affected	Effect	Physiological Outcome
ATP [35,36]	Glycogen Phosphorylase (GP)	Allosteric inhibitor	**Inhibits glycogenolysis**
Glucose-6-Phosphate (G6P) [35,36,38,42]	1) Glycogen Phosphorylase (GP)	Allosteric inhibitor of GP	**Inhibits glycogenolysis**
	2) Glycogen Synthase (GYS)	Allosteric activator of GYS	**Promotes glycogenesis**
Insulin[35,43,44]	1) Glycogen Phosphorylase (GP)	Dephosphorylation → inhibition	**Inhibits glycogenolysis**.
	2) Glycogen Synthase (GYS)	Dephosphorylation → activation	**Promotes glycogenesis**.
	3) GLUTs	Translocation to plasma membrane	**Promotes glycogenesis**

Table 1.2 Some factors in the regulation of glycogenesis and glycogenolysis during times of glucose surplus.

1.1.3 GYG, GYS and GBE

The triumvirate of enzymes responsible for the conversion of the ubiquitous glucose donor UDP-glucose to mature glycogen particles in eukaryotes – Glycogenin (GYG), Glycogen Synthase (GYS) and Glycogen Branching Enzyme (GBE) – catalyse reactions that are fundamentally similar, insofar as they involve the transfer of a carbohydrate donor to a particular nucleophile. GYG and GYS belong to the GT-8 and GT-3 families of glycosyltransferase enzymes respectively;[45,46] they catalyse the transfer of an activated donor sugar substrate to a suitable glycosyl acceptor, leading to the formation of a glycosidic bond (see section 1.2.1 for further discussion).[47] Whilst GBE is classed as belonging to the GH-13 family of glycosyl hydrolases on the basis of amino acid sequence conservation, its overall activity is best described as transferase-type; it transfers a portion of the α(1→4)-linked chain to a 6'-OH group elsewhere in glycogen.[46]

The conversion of UDP-glucose to glycogen begins with the glycosyltransferase enzyme Glycogenin.[48–51] Through an unusual stepwise autoglucosylation, an enzyme-linked maltosaccharide comprising ca. 10-20[8] α(1→4) linked glucose residues is built up.[8,52,53] The glycosyl acceptor in each step is the 4'-OH group of the terminal glucose residue, other than for the first step, in which a tyrosine residue of the enzyme itself serves this function.[54] This catalysis is believed to be dependent on the binding of a metal cofactor (likely Mn^{2+}), which serves to activate the UDP leaving group.[48,55] As such, it is GYG that acts as a "primer" for glycogen biosynthesis, providing the initial anchor point from which the dendron-like poly-glucose structures that comprise mature glycogen particles extend (**figure 1.3**).

In higher primates, GYG is known to exist as two distinct isoforms, namely GYG-1 and GYG-2, that differ primarily with respect to localisation – GYG-1 exists predominantly in muscle, whilst GYG-2 is found mostly in liver.[56–58] Other minor differences do exist, for example with regards to autoglucosylation site, which is Tyr195[i] in GYG-1,[59] but Tyr228 in GYG-2[57], size (GYG-2 is larger), and extent of autoglucosylation (reduced in GYG-2).[57] In a similar vein, minor organism-specific differences exist; for example, yeast GYG homologues may initiate glucosylation from two tyrosines (Tyr228 and Tyr230), in contrast to mammalian counterparts.[53,60,61] However, the fundamental mechanism and function remains as outlined above; metal-dependent α-(1,4)-glucan formation from initial tyrosine(s) (**figure 1.3**).[57,61,62] It should additionally be noted that *in vivo* studies in yeast identified a "glycogenin-interacting protein" (GNIP) that was found to stimulate GYG autoglucosylation, although the relevance of this protein in other organisms is unknown.[63]

Association of the downstream enzyme GYS to GYG is mediated by the C-terminal domain and oligosaccharide moieties of the latter.[64] Like GYG, this enzyme is also a retaining glycosyltransferase (see section 1.2.1), which catalyses further α(1→4) glucosylation of the GYG-linked maltosaccharide, and also exists as two isoforms (GYS-1 and GYS-2) that also differ primarily with regards to location in the same sense as the GYG isoforms.[65] However, unlike GYG, the catalysis effected by GYS is not dependent on a metal cofactor; the UDP leaving group is instead activated via hydrogen bonding to proximal amino acid residues.[66] Furthermore, in eukaryotes, GYS is not an autocatalytic enzyme in the sense that GYG is, thus the mature glycogen particles remain covalently bound to only the latter enzyme. This is not the case in

[i] This residue is variously referred to as either Tyr195 (e.g. refs 2, 52, 54) or Tyr194 (e.g. refs 49, 75, 78) in existing literature, depending on the author's amino acid numbering (whether the initial methionine is included). According to our plasmid sequence (see chapter 2), we refer to it as Tyr195 (including the initial Met). In a similar vein, other amino acid numberings utilised in this text may differ by one from some prior literature; these will be noted as and when encountered.

prokaryotes, where, owing to their lack of a homolog of GYG, their GYS homolog (GlgA) is believed to effectively carry out both autoglucosylation and subsequent extension.[9,32]

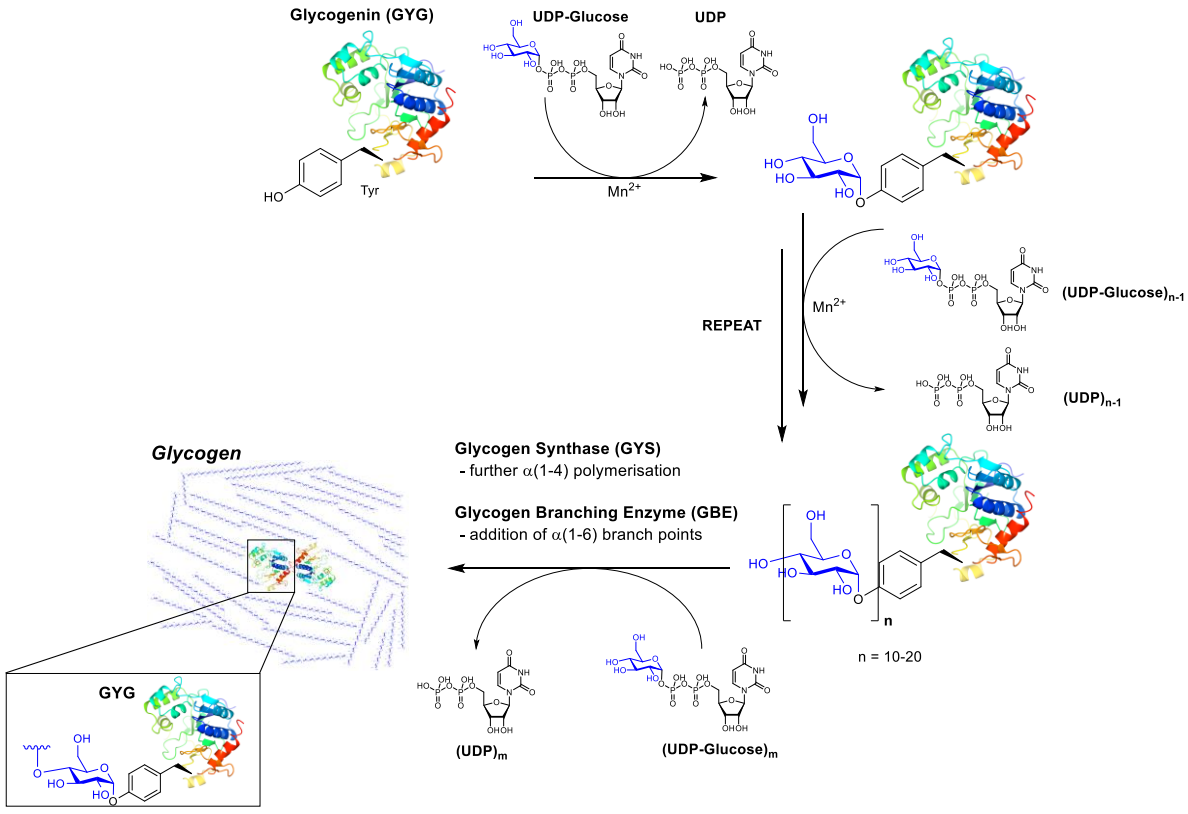


Figure 1.3 The roles of Glycogenin, Glycogen Synthase and Branching Enzyme in (eukaryotic) glycogenesis. Glycogenin (GYG) catalyses its own stepwise autoglucosylation from a tyrosine residue in order to synthesise an enzyme-linked maltosaccharide. This serves as a primer for GYS, which catalyses further α(1→4) glucosylation of this carbohydrate, whilst GBE introduces branch points in the form of α(1→6) glycosidic linkages.

Meanwhile, GBE acts in concert with GYS to introduce branch points in the form of α(1→6) glycosidic linkages between individual α(1→4) linked maltosaccharide chains. This occurs with retention of configuration and presumably involves a so-called double displacement mechanism, discussed further in section 1.2.1 below but briefly outlined here for this specific case.[67] The enzyme first cleaves the maltosaccharide (every 8-14 residues) via nucleophilic attack of a conserved Asp residue at C1 of a particular glycosidic linkage (**figure 1.4**).[67] This yields an enzyme-bound donor moiety, which may now be transferred to a 6-OH acceptor within the maltosaccharide chain – transglucosylation – to form the new branch point.[67] A conserved Glu residue acts as an acid/base catalyst (**figure 1.4**).[67] The second step can occur in an *intra* or *inter* sense (the former is illustrated below), depending on whether the donor is transferred to the same maltosaccharide chain from which it was cleaved, or instead to a different chain.

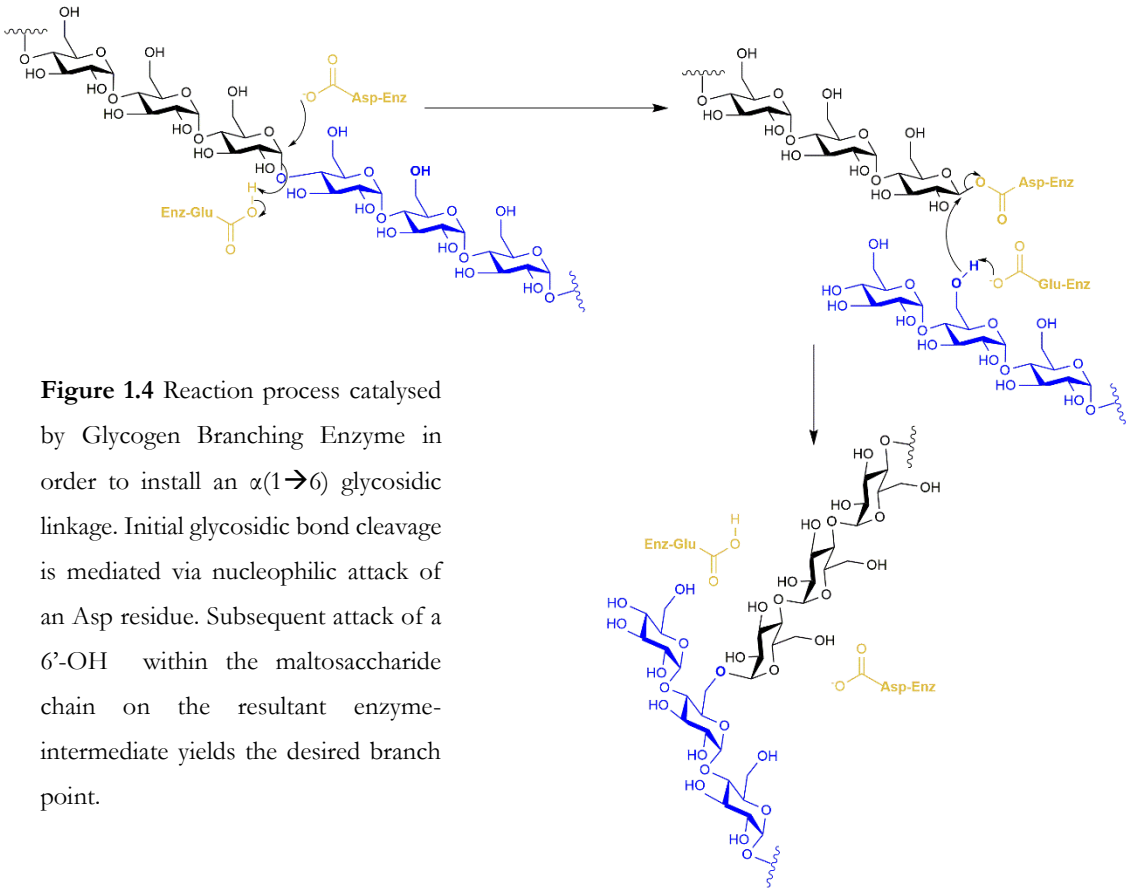


Figure 1.4 Reaction process catalysed by Glycogen Branching Enzyme in order to install an α(1→6) glycosidic linkage. Initial glycosidic bond cleavage is mediated via nucleophilic attack of an Asp residue. Subsequent attack of a 6'-OH within the maltosaccharide chain on the resultant enzyme-intermediate yields the desired branch point.

Given the evident importance of these three enzymes in such a fundamental biological mechanism of energy storage, it goes without saying that a complete mechanistic understanding of each is desirable, a point highlighted further by the large number of diseases related to deficiencies in glycogen biosynthesis that in many cases arise from malfunction of one or more of these enzymes.[5,68] We were particularly drawn to the specific case of GYG, as the enzyme responsible for the first, crucial steps of glycogenesis. In addition to the privileged biological position this affords to the enzyme, as a highly unusual autocatalytic glycosyltransferase its catalytic mechanism is of significant inherent interest in its own right. A more thorough examination of current knowledge regarding this mechanism forms the basis of the following section, placed in the context of general mechanisms of glycosyltransferase enzymes.

1.2 Glycogenin: an Unusual Autocatalytic Glycosyltransferase

1.2.1 Common Mechanisms of Glycosyltransferases

A glycosyltransferase may be defined as an enzyme catalysing the transfer of a suitable nucleophile with an activated donor sugar substrate;[47] GYG and GYS, as noted previously, represent examples thereof. This donor is frequently activated as a nucleoside diphosphate (e.g. UDP-glucose, GDP-mannose), although

alternative donors include nucleoside monophosphate variants, lipid phosphates and even unsubstituted phosphate itself.[69] Glycosyltransferase enzymes have been classified into families based on amino acid sequence similarities; to date, 103 such families exist.[46] However, only three general folds (GT-A, GT-B and GT-C) have been observed for glycosyltransferase enzymes in structures solved thus far.[70] The GT-A fold consists of two β/α/β Rossmann domains that are closely associated, with distinct acceptor-binding and nucleotide-binding domains observed.[71] Whilst the GT-B fold also comprises two β/α/β Rossmann domains, these are less tightly associated and face each other, forming a cleft in which the active site is located.[72] The GT-C fold, identified in more recent crystal structures, consists of a large number (8-13) of transmembrane helices, with an active site located on a long-loop region.[73]

The catalytic mechanism of glycosyltransferase enzymes can be divided into two classes, depending on the stereochemistry at the anomeric centre of the carbohydrate donor species (**figure 1.5**). Inverting glycosyltransferases catalyse glycosylation with stereochemical inversion at this centre (e.g. α-NDP-sugar → β-glycoside). However, the analogous reaction with stereochemical retention at this centre (e.g. α-NDP-sugar → α-glycoside) is also a possibility, and is catalysed by retaining glycosyltransferases (e.g. GYG, GYS).[69]

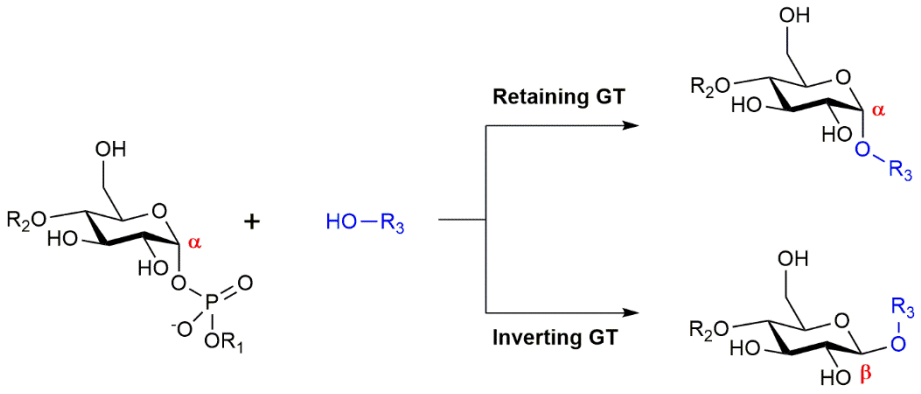


Figure 1.5 Reactions catalysed by retaining (top) and inverting (lower) glycosyltransferases.

The commonly accepted mechanism for inverting glycosyltransferases involves a single step, involving direct displacement of the leaving group by the incoming nucleophile in an S_N2 reaction, thus accounting for the stereochemical inversion observed. An appropriately positioned basic residue serves to deprotonate the glycosyl acceptor, facilitating its attack at the anomeric carbon with concomitant departure of the leaving group, thus forming the glycosidic linkage (**figure 1.6 a**).[69,74,75] The reaction is believed to proceed via an oxocarbenium-like transition state.

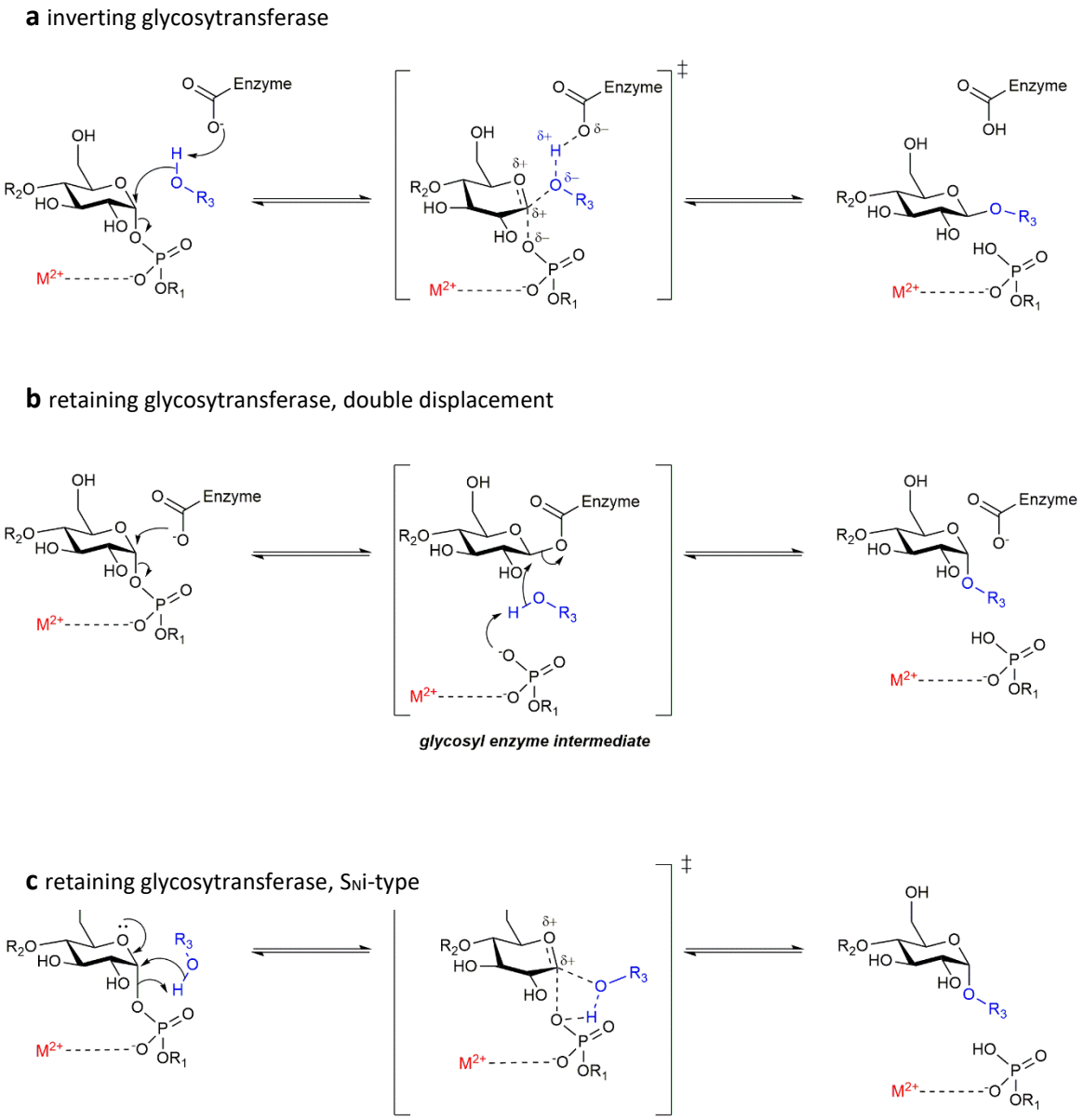


Figure 1.6 Common mechanisms of glycosyltransferase enzymes. In the examples shown, a divalent metal cation serves to stabilise the departing phosphate leaving group, however this function may in some instances be carried out instead by positively charged side chain residues or helix dipoles, as mentioned in the main text.

For retaining glycosyltransferases, the mechanism is less clearly defined. One proposal involves a double displacement process involving two sequential S_N2 processes: 1) attack of a suitably nucleophilic enzyme residue on the glycosyl donor to generate a covalently bound enzyme intermediate, and 2) attack of the glycosyl acceptor on this species, releasing enzyme and product (**figure 1.6 b**); the glycosyl acceptor is most likely deprotonated by the departing phosphate group.[69,76] However, the lack of a clear, suitably positioned

candidate for the nucleophilic residue in some cases led to an alternative proposition involving an S_Ni-type reaction (**figure 1.6 c**).[77–79] In this case, attack of the glycosyl acceptor occurs at the same face from which the leaving group is lost, with the alternative face being shielded by the enzyme itself.

Stabilisation of the departing (negatively charged) phosphate-based leaving group may be mediated by its coordination to a divalent metal cation (as illustrated in **figure 1.6**); this is common in GT-A and GT-C enzymes, which frequently possess a "D-X-D" amino acid motif for binding this metal cofactor.[70,80,81] However, GT-B enzymes typically do not require a metal cofactor to effect catalysis, instead utilising positively charged amino acid side chains or helix dipoles to coordinate and stabilise the leaving group; GYS represents one such example.[47,69]

1.2.2 Catalytic Mechanism of Glycogenin

1.2.2.1 Substrate Binding and Catalytic Mechanism

In broader respects, the catalytic mechanism of GYG is encompassed within the typical glycosyltransferase mechanisms outlined above. GYG is categorised as belonging to the GT-8 family of glycosyltransferases, members of which exhibit the canonical GT-A fold previously discussed.[46] Consistent with this, and as previously mentioned, GYG requires a metal cofactor – believed to be Mn^{2+} - in order to effect catalysis. As previously noted, this moiety serves to coordinate and stabilise the pyrophosphate leaving group of UDP-Glucose during the glycosyl transfer reaction.

The following discussion focusses on the better-studied GYG-1 isoform, since this is the enzyme that will be utilised throughout this thesis. In general, from this point, any mention of "GYG" should be assumed to refer to GYG-1 unless stated otherwise. However, the general principles are likely applicable to GYG-2 (though detailed mechanistic studies on this isoform are lacking).

Crystal structures of mammalian GYG-1, obtained by Gibbons *et al*, showed this metal cofactor to be coordinated to the enzyme via a D-X-D motif comprised of Asp102, Asp104 and His212 (**figure 1.7**);[ii] knockout of one of the aspartate residues abolishes activity, highlighting its essential role.[55] Given its additional coordination to the UDP-Glucose pyrophosphate moiety, the metal overall exhibits an octahedral coordination pattern.[55]

[ii] See footnote (i): may be referred to as Asp101, Asp103, and His211 in prior literature owing to differences in amino acid numbering.[55,82] This is true for all other amino acids listed also.

Similarly, crystal structures have provided information regarding the mechanism of UDP-Glucose binding by GYG-1.[55,59] Fundamentally, this nucleotide donor is bound within an N-terminal β-α-β Rossmann-like domain. The uracil moiety is stabilised by contacts with residues between 11 and 15 (**figure 1.7**), with Asn12 being particularly crucial in determining specificity for **U**DP-Glucose versus, for example, **C**DP-Glucose.[55] Meanwhile, the ribose portion is stabilised through contacts with Leu9, Ala103, and Asp104 (**figure 1.7**), the latter notable given its additional involvement in Mn^{2+} coordination as noted above.[55] Stabilisation of the glucose moiety appears effected most notably via coordination of Asp160 and Asp163, since mutation of these residues essentially abolishes activity, though other residues (e.g. Ser134, Gln133) are also involved.[59,82] The "active" conformation of UDP-Glucose within GYG has been shown crystallographically to involve a "folding back" of the glucose moiety back beneath the pyrophosphate group.[59]

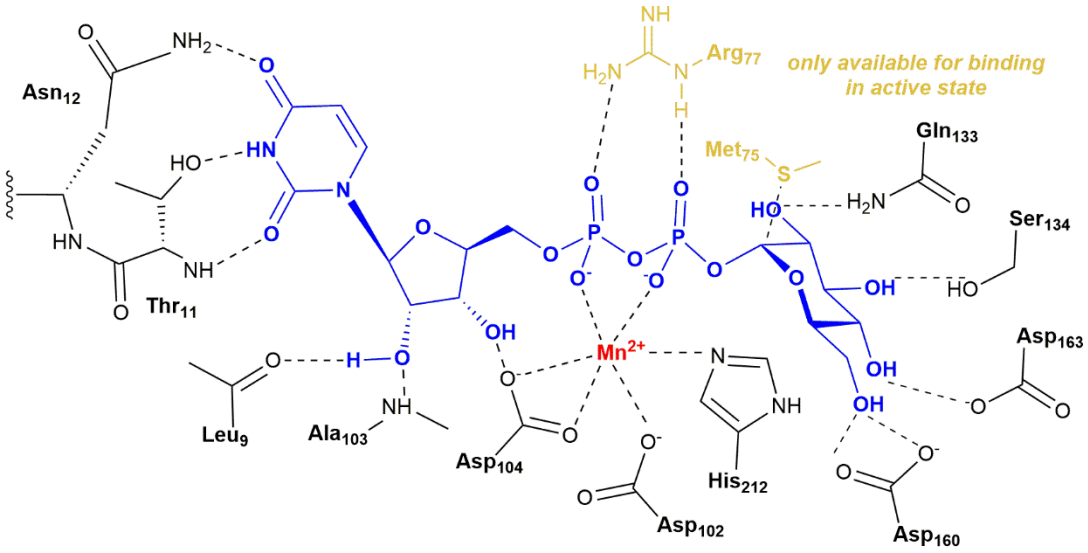


Figure 1.7 Some key residues involved in binding UDP-Glucose within GYG-1. UDP-G is shown in blue the manganese co-factor in red. Residues in yellow are known to be available for UDP-G binding only in the "active" state of the enzyme, not the "ground" state (discussed below). Other residues are shown in black.

Whilst the crystal structures of Gibbons *et al* provided great insight into specific residues involved in substrate binding within GYG-1,[55] for example, they could not provide information regarding the dynamics of the enzyme during catalysis. More recently, however, Chaikuad *et al* successfully captured crystallographic "snapshots" of the enzyme during different stages of its catalytic cycle.[59] These revealed that structural dynamics play a crucial role in GYG catalysis. Specifically, two distinct states of the enzyme exist: "ground", and "active". Transition from the inactive ground state to the competent active state involves structural rearrangement of three distinct regions: 1) a 32-residue "lid segment" from residues 60-91; 2) a helix-turn-

helix (HTH) motif from residues 189-207, which notably includes the catalytic Tyr195 residue; and 3) a C-terminal loop (residues 233-243) located close to the arm bearing the Tyr195 acceptor.[59]

For the lid, the ground-to-active reorganization results firstly in a change from an "open" to a "closed" conformation, sequestering the active site from surroundings, and secondly in the freeing up of two residues – Arg75 and Met77 – to play roles in substrate binding (**figure 1.7**). Meanwhile, both the HTH motif and C-terminal loop are disordered in the ground state but ordered in the active state. Notably, the HTH motif – containing the catalytic tyrosine – exhibits a variety of conformations, affording flexibility to the tyrosine (or glucosylated Tyr) acceptor enabling it to adopt a suitable orientation to attack the UDP-Glucose donor.[59]

GYG (like GYS) is a retaining glycosyltransferase, and originally the glycosyl transfer mechanism was proposed to involve a double displacement process (**figure 1.8 a**, see also **figure 1.6 b**). An active site aspartate residue (Asp163[ii]) was implicated as the putative nucleophilic catalyst based on loss of activity observed upon its mutation – not due to any structural change – and a high degree of conservation of this residue.[55,82] However, a substantial movement of this residue would be necessary for it to act as nucleophilic catalyst (i.e. to attack the C1 of UDP-Glucose), yet it remains far from this C1 in both ground and active state, despite the dynamic transition between these.[59] As such, an S_Ni-type mechanism (**figure 1.8 b,** see also **figure 1.6 c**) has more recently been suggested as more plausible.[59] In this case, the Arg77 and Met75 residues – freed up upon transition to the active enzyme state, as noted above – along with Asp163, are suggested to be well-placed to stabilise (partial) charges developing as glycosyl transfer proceeds. In either mechanism, the pyrophosphate leaving group is implicated as the base required to deprotonate the glycosyl acceptor.[55,59]

1.2.2.2 Oligomeric State of GYG

The catalytic mechanism of GYG is further complicated by the fact that, across organisms and isoforms, this enzyme can and does form dimers, as evidenced both by crystal structures and gel filtration studies in solution.[55,57,61,83–85] The K_D for the GYG-1 dimer is between 0.8 μM (non-glucosylated) and 0.6 μM (glucosylated);[83] at concentrations below this, gel filtration indicates largely monomeric enzyme as the more thermodynamically favoured species. This K_D is a relatively high value (i.e. not nM), indicating fairly weak affinity between monomers within a dimer. At the other end of the spectrum, higher order oligomers – notably, a decamer – have been previously observed in crystallographic studies.[55] However, gel filtration analysis of GYG-wt has yet to detect any significant quantity of such species other than dimer even at concentrations of up to 20 μM,[82,83] and it is thus believed that such a species is likely a crystallography artefact and not biologically relevant.[55] Thus, the dimer appears, with respect to monomer and oligomer,

the thermodynamically favoured species at least above 0.8 μM and below 20 μM (though transient formation of other species may occur). For reference, all assays within this thesis, as well as the majority of prior solution-phase studies,[iii] fall well within this concentration range in which dimers are thermodynamically favoured.

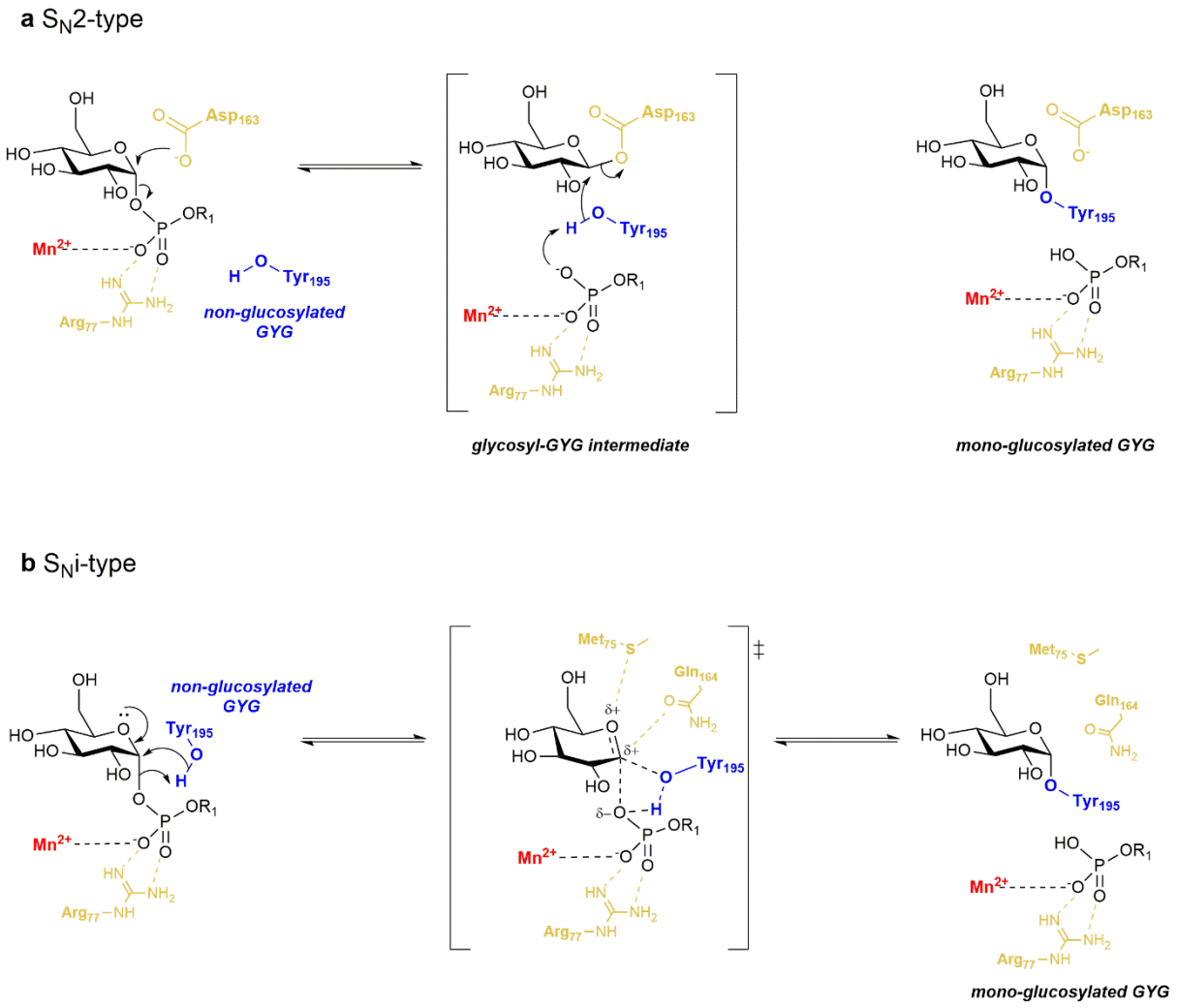


Figure 1.8 Possible glycosyl transfer mechanisms for GYG-1 catalysis, based on those shown in **figure 1.6**: (a) S_N2-like and (b) S_Ni-like, shown here for the initial Tyr195 glucosylation of GYG for simplicity. The Tyr195 acceptor is shown in blue, the UDP-Glucose in black (R_1 represents the remaining uridine-Pi moiety). Residues in yellow are those implicated as being directly involved in the reaction. The Mn^{2+} cofactor is shown in red.

The regions of GYG-1 involved in the dimer interface have been identified through prior crystallographic studies.[55] These include three loop regions (aa125-132, aa160-165 and aa176-193), and two α-helices (aa194-

[iii] See for example refs 82, 84-85, 88-89..

198 and aa200-205). Of note here is the inclusion in this interface firstly of the acceptor tyrosine (Tyr195 for GYG-1), and secondly of two other residues identified through mutagenesis as critical for catalytic activity (Asp160, Asp163) owing to their roles in substrate binding.[59,82]

For dimeric GYG, various distinct types of glucosylation mechanism can be envisaged, depending on whether a GYG monomer within a dimer glucosylates itself ("intra-monomer"), glucosylates the other monomer within the *same* dimer ("inter-monomer"), or even glucosylates a monomer within a *different* dimer ("inter-dimer") (**figure 1.9**).[85] In the latter case, this would presumably involve only transient dimer-dimer association in general, given that, as noted above, formation of higher order GYG oligomers appears unfavourable across a range of studied concentrations.[82–84]

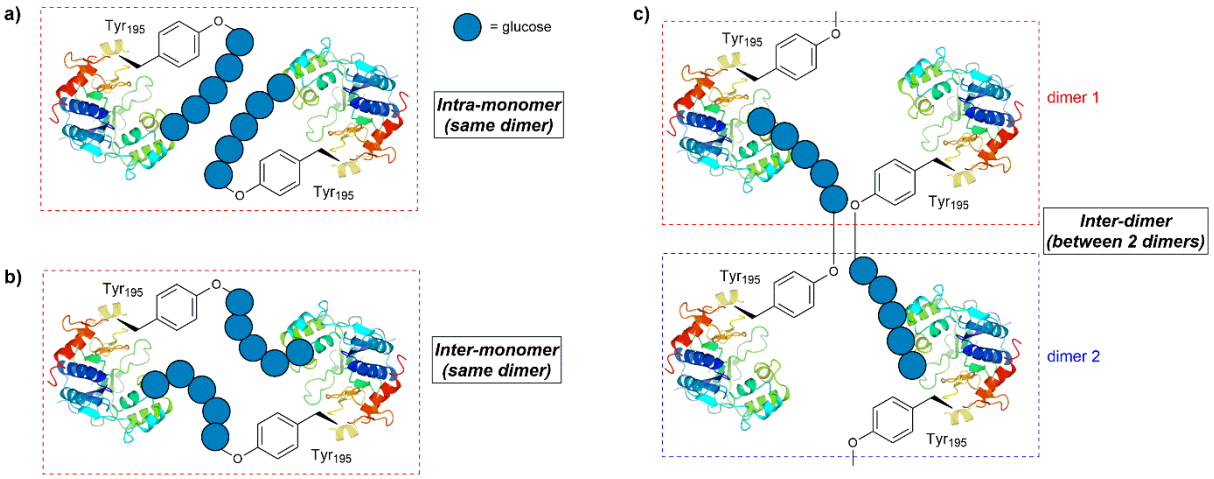


Figure 1.9 Schematic depiction of possible modes of glucosylation within a GYG dimer: (a) intra-monomer, (b) inter-monomer, (c) inter-dimer.[85] Dimers are shown within dotted boxes. In the inter-dimer case, only one glucosyl chain from each dimer is shown for clarity.

Elegant experiments by Issoglio *et al* have demonstrated the existence of both intra-monomeric and inter-monomeric modes of glucosylation within a GYG dimer.[84] They constructed two different mutant GYG heterodimers, one capable only of glucosylation via inter-monomer mechanism and one capable of glucosylation only via intra-monomer. The results indicated that, whilst both mechanisms were viable, intra-monomeric glucosylation could not significantly *initiate* catalysis (i.e. from Tyr195), while inter-monomeric glucosylation could initiate catalysis but yielded reduced final autoglucosylation extents.[84] This highlights the apparent need for both mechanisms within a GYG dimer for full catalysis; different mechanisms may operate at different oligosaccharide lengths. In a similar vein, the existence of inter-subunit glucosylation was also demonstrated by Lin *et al* through persistence of activity upon mixing one GYG mutant only capable of acting as a "donor" and one only capable of acting as an "acceptor".[85]

It is of note that the monomeric enzyme, despite only having an intra-monomeric mode of glucosylation available, in fact exhibits full glucosylation capacity, in contrast to the "intra-only" dimer.[84] This highlights two things. Firstly, the mechanisms for monomeric and dimeric GYG are not equivalent; in the latter, there are additional steric constraints on both the orientation of the oligosaccharide chain and on the enzyme scaffold imposed by the other monomer within a dimer. These are not present for an isolated monomer.[84] Secondly, in combination with the establishment that GYG dimers undergo intra-monomer glucosylation, it emphasises the "genuinely" autocatalytic nature of the enzyme; one GYG subunit can indeed glucosylate itself. If only inter-subunit mechanisms existed one might question such a label, but that is not the case.

Crystallographic data has also confirmed the existence of both intra-monomer and inter-monomer modes of glucosylation, and implicated a switch between these during catalysis.[59] Crystal structures of GYG after 4 and 6 glucosylation steps – i.e. with 4 or 6 glucoses attached at Tyr195 – were obtained through crystallisation of heterogeneously glucosylated GYG-wt.[59] For the GYG-tetrasaccharide species, the oligosaccharide is oriented towards the active site of the same monomer, indicating intra-monomeric catalysis. However, for the hexasaccharide variant, it is instead oriented towards the active site of the other monomer within the dimer, indicating an inter-monomeric mechanism. This again indicates multiple mechanisms of catalysis within a GYG dimer, at different oligosaccharide lengths. Such a mechanistic switch appears primarily mediated by a striking flexibility of the acceptor tyrosine (Tyr195), flipping from an "in" orientation (intra-monomer) to an "out" one (inter-monomer), and is likely necessitated to avoid a steric clash of the longer hexasaccharide chain from one monomer with the other monomer.[59]

With regards to inter-*dimer* glucosylation, kinetic evidence suggests that this may not occur; GYG glucosylation has been seen to be largely independent of GYG concentration,[49,85] except for a decrease at very low concentrations, reflecting instead dissociation to the monomer.[54,83] For inter-dimer catalysis, some level of rate dependence on GYG concentration, even when existing as dimers, would be more likely. However, it has nonetheless been hypothesised that inter-dimeric catalysis may operate during the earliest stages of GYG catalysis, e.g. the initial tyrosine glucosylation, given the very large distance of this tyrosine from the active site of either monomer.[55] As noted above, the lack of higher order oligomers as thermodynamically favoured species in solution does mean that any such mechanism must likely involve transient formation of these.

A final point to briefly note concerns the kinetics of GYG dimer association and dissociation. Whilst this appears little-studied in general, prior data implies that subunit exchange between dimers does occur on reasonably short (minutes) timescales; even well above K_D, a given dimer is not "fixed".[84,85] Issoglio *et al*

investigated glucosylation of a mutant GYG heterodimer, capable only of inter-monomeric glucosylation, with a large excess of either the "donor" monomer or of the "acceptor" monomer.[84] If no exchange occurred, the extent of glucosylation, relative to concentration of heterodimer, would be the same in each case (equal concentrations of heterodimer and and no glucosylation possible outside these). However in fact it proved 5-6 fold greater after one hour with excess acceptor than with excess donor. This is indicative of dimer subunit exchange, such that some of the excess acceptor could also be glucosylated by the donor monomer.[84] The 5-6 fold difference after 1 hour implies that each heterodimer undergoes exchange of the acceptor subunit at least every 10-12 minutes. This notion of reasonably facile dimer subunit exchange is also supported by prior failed attempts to purify well-defined GYG heterodimers, attributable to subunit exchange on the timescales of column chromatography (e.g. 1 hr).[85]

1.2.3 Challenges in the Mechanistic Study of GYG: the Need for New Methods

The preceding section serves, above all, to highlight the complexity surrounding the mechanism of GYG catalysis. This reflects in no small part its ability to form dimers and glucosylate through multiple mechanisms, and also the dynamic nature of the catalytic cycle. Yet more generally, the very fact that GYG is an autocatalytic enzyme immediately renders it somewhat unique; whilst other autocatalytic glycosyltransferases are known, they are not common.[86,87] Furthermore, GYG must necessarily catalyse two very distinct glucosylations; that of the initial tyrosine residue, and that of the 4'-OH of subsequently added glucose moieties.[54]

These facts highlight the concept that each glucosylation step catalysed by GYG is distinct from those preceding and following it; the enzyme (specifically, its glycosyl acceptor oligosaccharide moiety) is structurally non-identical for each step (indeed, strictly speaking it is not a "catalyst" in the true sense). It is precisely this fact that opens up the possibility for the existence of multiple phases and mechanisms during GYG catalysis, a concept that as discussed above has been demonstrated in practice.[59,84,85] Whilst the existence of dimers may open up *possibilities* for different mechanisms, for example, there would be no need to switch between these were the enzyme identical at each step; this is not the case.

Much progress has undoubtedly been made regarding understanding of the catalytic mechanism of GYG since its initial discovery, and as noted above the existence in practice of a yet a number of aspects remain poorly understood. The existence of multiple glucosylation mechanisms indeed implies a multiphasic nature of GYG catalysis. Yet precise understanding of the timing and indeed number of such phases is hindered

by inherent limitations of conventional biological approaches to isolate GYG for studies of its glucosylation, as detailed below (**figure 1.10**). As mentioned previously, we focus on the better-studied GYG-1 isoform that will be used in this work.

Conventional over-expression of GYG-1 in *E. coli* typically yields a heterogeneous mixture of glucosylated glycoforms at the glycosyl acceptor tyrosine (Tyr195), reflecting partial autoglucosylation within the expression host (**figure 1.10 a**).[49,59,88] Such heterogeneity renders precise mechanistic analyses (particularly kinetics-based approaches) challenging, especially of the earliest glucosylation steps, since they have thus already occurred prior to isolation of the enzyme. Alternatively, expression in bespoke *E. coli* strains deficient in UDP-glucose pyrophosphorylase may instead afford glycan-free enzyme (**figure 1.10 b**).[59,88,89] However, whilst this avoids issues associated with glycoform heterogeneity, a different problem arises; prior kinetic data for GYG-1 indicates that the initial tyrosine glucosylation is relatively slow,[89] acting as a "lag phase" that may obscure subsequent faster steps. It is readily apparent, therefore, that both methods suffer from drawbacks, particularly with regards to kinetics-based approaches and study of the earliest glucosylation steps.

More generally, it is evident that such limitations reflect the fact that these approaches allow controlled access to only one state of the enzyme as a homogeneous species – the glycan-free variant. (Crystallography of heterogeneous GYG-wt may yield structures of defined glycoforms,[59] but without any control over which glycoform, and with no means to access these outside the limited crystallography setting). Crucially, they cannot enable access to any homogeneous GYG glycoforms, and it is precisely this fact that leads to an inability to examine multiple individual glucosylation steps in detail. As discussed in chapter 4, a pre-requisite for such an ability is access to the relevant starting GYG enzyme for the glucosylation step of interest as a homogeneous species, which can enable monitoring of the particular glucosylation step through tracking consumption of this starting enzyme. At present, therefore, only the initial tyrosine glucosylation may be monitored in this sense. This is a particular drawback given the aforementioned inherent uniqueness of each glucosylation step, and the potential for distinct mechanisms and phases in GYG catalysis.

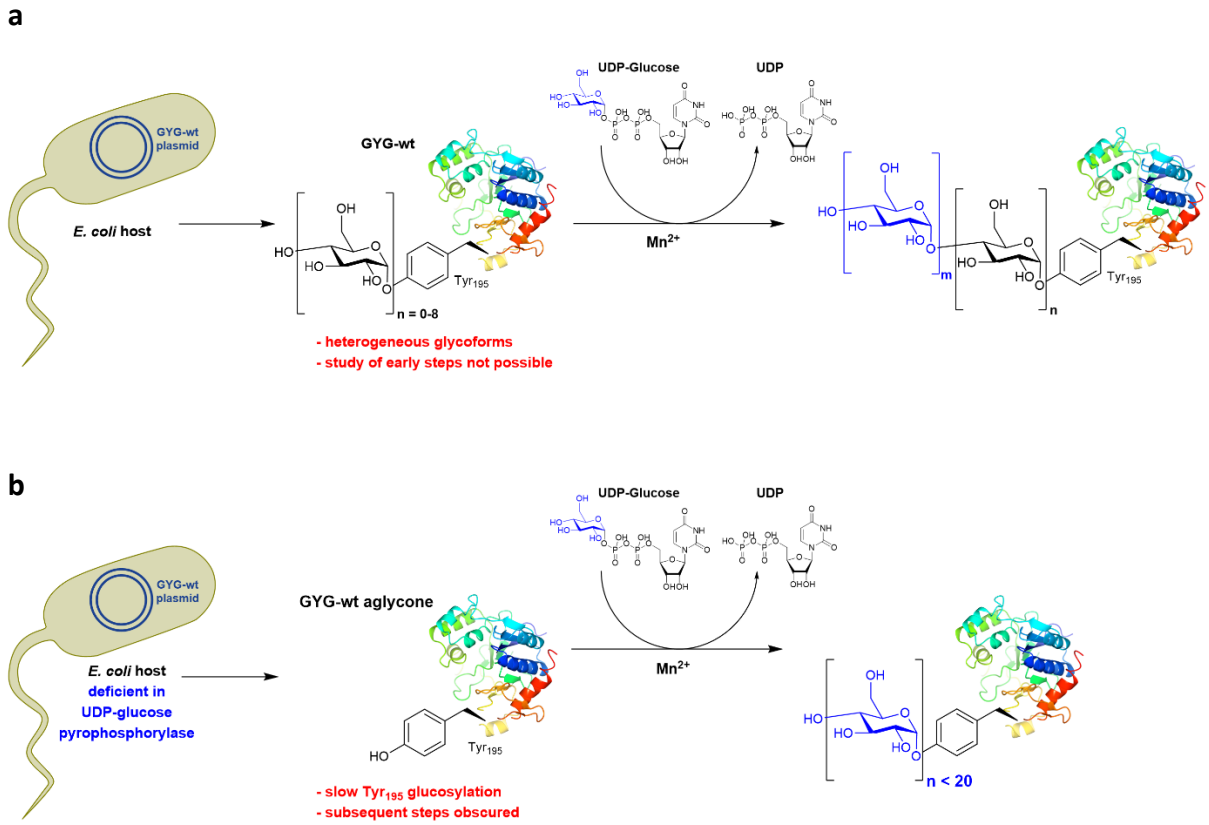


Figure 1.10 Conventional biological methods for the isolation and study of GYG-1. The enzyme can either be isolated as a mixture of heterogeneous glycoforms (a), or as the aglycone (b), both of which have associated drawbacks (listed in red) with regards to study of the enzymatic mechanism.

We considered that a new approach capable of enabling access to, and isolation of, multiple different GYG glycoforms as homogeneous variants could greatly enhance our understanding of the catalytic mechanism for this enzyme. This would correspond to unprecedented, direct access to intermediates along the GYG autoglucosylation pathway, and could enable characterisation of distinct glucosylation steps with a level of precision not previously possible (**figure 1.11**). Such a method would circumvent aforementioned problems of glycoform heterogeneity and the slow initial tyrosine glucoslyation step, whilst additionally enabling detailed study of the least well-understood early catalytic steps. As noted previously, the research described in this thesis focusses exclusively on the more abundant mammalian GYG-1 isoform; this represents the better-studied variant from which the majority of mechanistic conclusions to date have been drawn,[53] and therefore enables better alignment of any novel methodology with existing methods.

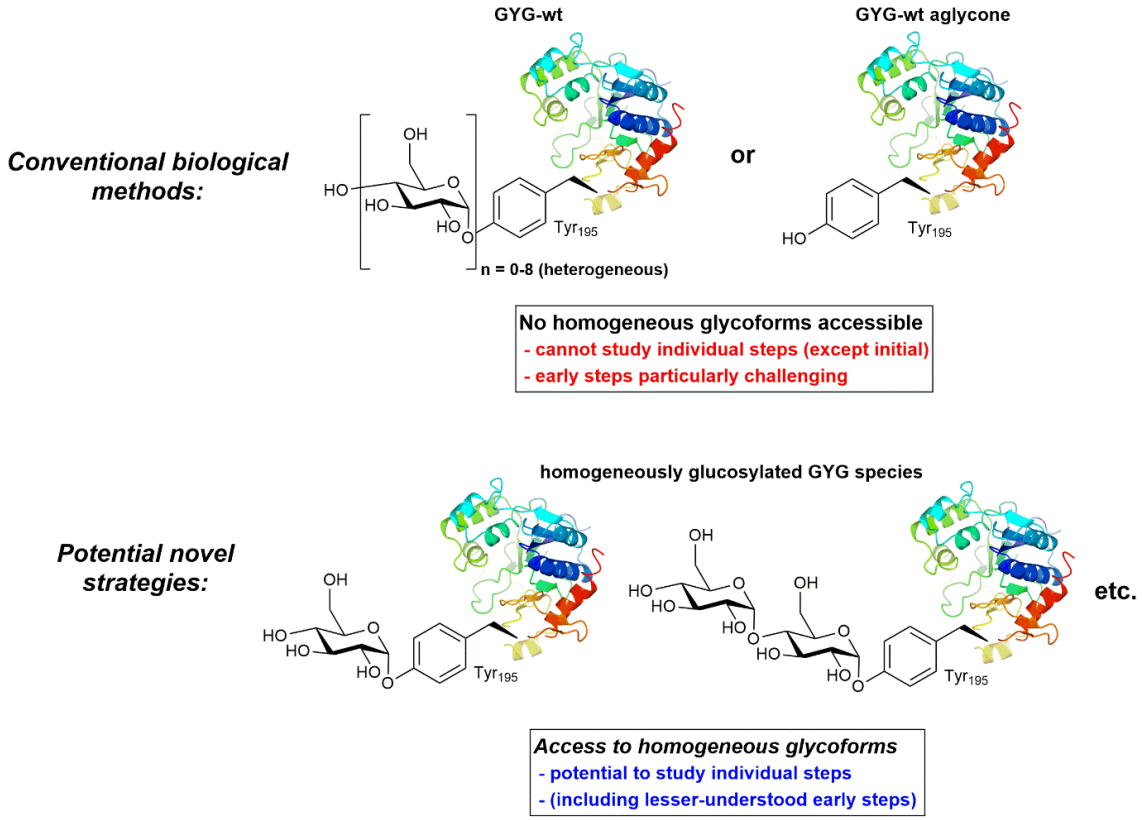


Figure 1.11 Novel strategies could enable access to homogeneous GYG glycoforms, facilitating study of individual steps and thus overcoming the limitations of conventional biological approaches.

Any successful approach in this regard must enable not only the installation of a carbohydrate moiety in a site-selective fashion at the catalytic tyrosine residue (Tyr195 for GYG-1), but additionally, the GYG glycoforms thus generated must be catalytically active in order to be of biological relevance. A number of potential approaches that might be utilised in this regard can be envisaged. The following section (1.3) outlines five such general strategies, discussing their applicability or otherwise to this work. Subsequently, the most promising of these is examined in greater depth (section 1.4), ultimately arriving at an overview of the precise method of choice (section 1.5).

1.3 Strategies to Access Homogeneous Glycoproteins

The challenge of accessing homogeneous protein glycoforms is not in itself a new one. Protein glycosylation represents one of the most abundant protein modifications in eukaryotes,[90] yet glycoproteins are typically produced as heterogeneous mixtures of glycoforms.[91] This reflects the fact that, whereas the biosynthesis of e.g. proteins occurs in a strictly templated fashion, that of the glycan portion of glycoproteins does not.[92] The biosynthesis of both *N*-linked and *O*-linked glycoproteins involves a complex myriad of glycosyltransferases and glycosidases (see ref. 93 for an overview).[93,94] A description of all of these is beyond the scope of this thesis, however the key point is that factors such as competition between these with regards to substrate/donor specificity, local substrate availability, and varying expression levels, activation and localisation of the enzymes may all lead to significant structural heterogeneity of the glycan portion for a given protein scaffold.[92,94,95] Given the integral role of protein glycosylation in a range of biological processes such as cell-cell adhesion,[96,97] signalling,[98] trafficking,[99] immune response,[100,101] as well as its impact on protein stability and folding,[99,102] it is unsurprising that significant efforts have been devoted to the generation of homogeneous glycoforms of proteins to better understand such phenomena.

As such, a variety of approaches to access homogeneous glycoproteins exist, yet, as ever, the optimal approach is very much dependent on the specific nature of the protein in question. As an enzyme, and an unusual autocatalytic one at that, GYG presents a somewhat unique case, in which the glycan moiety fulfils an integral catalytic role. Any successful strategy must, as noted above, fulfil at least two basic requirements: 1) the carbohydrate must be site-selectively installed at the catalytic tyrosine (Tyr195 of GYG-1), and 2) the resultant GYG glycoconjugates generated must be catalytically active to be of any biological relevance. This section outlines five classes of approach that have be used to obtain homogeneous glycoproteins in general, and evaluate their applicability – or otherwise – with respect to the synthesis of GYG glycoforms, in so doing rationalising our choice of approach.

1.3.1 Chemical Synthesis of Glycoproteins

Whilst undoubtedly representing a relatively time-consuming approach, total chemical synthesis has been used to prepare a number of glycoproteins as homogeneous glycoforms,[103–105] a particularly notable example being the total synthesis of the highly glycosylated signalling protein erythropoietin as reported by the Danishefsky group.[106] The basic concept of chemical protein synthesis is summarised graphically below (**figure 1.12**).[107,108] A number of peptide fragments are synthesised, typically by standard SPPS techniques,

although fragments may instead be expressed in *E. coli*.[109] For smaller proteins (typically < 50 amino acid residues[92,110]) this may itself be sufficient to generate the target molecule. For larger peptides and proteins, SPPS alone is too inefficient, however they may be accessed through chemoselective ligation of the SPPS-generated peptides. Native chemical ligation (NCL) represents a particularly powerful and widely-utilised method to accomplish this,[111–113] although other notable methods (e.g. serine/threonine ligation, KAHA ligation, formation of oxime or hydrazine linkages) also exist.[114–116] The former strategy classically involves trans-thioesterification between a peptide bearing a C-terminal thioester and a second peptide bearing an N-terminal cysteine, with a subsequent *S,N*-acyl shift yielding a native amide bond (for reviews see refs. 111-113).[111] Whilst this therefore traditionally required, in a retrosynthetic sense, disconnection adjacent to a cysteine residue, the development of methods to convert cysteine to other amino acids post-ligation render alternative disconnections possible.[113,117–119] Irrespective of method, a final refolding step is typically required to attain the protein with correct secondary and tertiary structure.

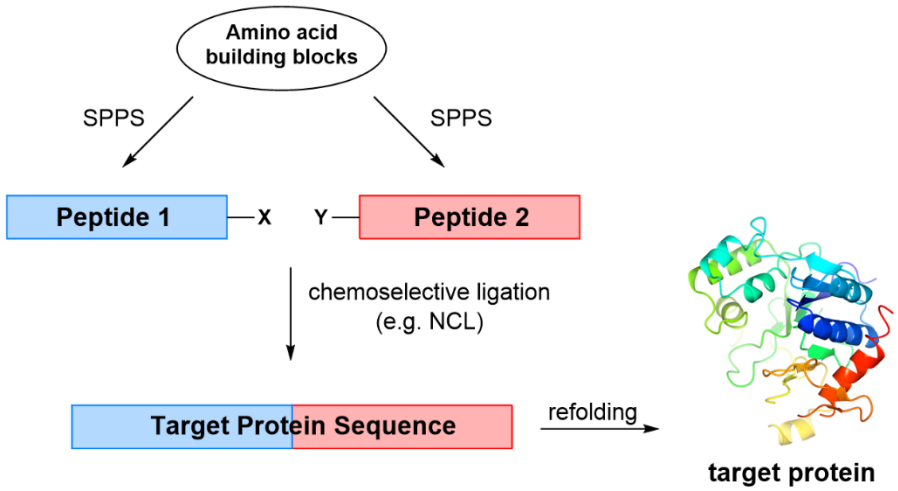


Figure 1.12 The use of chemical synthesis to access protein targets. Peptides are assembled through SPPS techniques. Subsequent ligation (for larger peptides/proteins) and post-synthesis refolding can yield the desired target protein.

Glycoprotein chemical synthesis follows this same concept, with the addition of glycosylated amino acid residues at desired positions on the initial peptides (peptide 1 and/or peptide 2 in **figure 1.12**). Naturally, a key part of this approach is thus the generation and installation of such residues onto these peptides. One may envisage two basic conceptual approaches that may be utilised in this regard (**figure 1.13**).[92,110] In a "linear assembly" method, the glycan is coupled to a (protected) amino acid, which is then used as an amino acid building block in SPPS-based construction of the peptide (**figure 1.13 a**). Such an approach was utilised to prepare the entirety of a 47-aa HIV-1 V3 domain.[120] Alternatively, in a "convergent assembly" approach,

the carbohydrate moiety is ligated to a pre-formed peptide fragment in a site-selective fashion (**figure 1.13 b**). For example, Lansbury and co-workers developed a protocol enabling direct coupling of glycosyl amines with an Asp residue on a peptide;[121] this approach was later employed by the group of Danishefsky to the synthesis of complex prostate-specific antigen glycopeptides.[122] Clearly, in both cases, the peptides thus formed may be further extended (via SPPS) and/or ligated to other peptides as above if desired. Whilst not illustrated, in certain instances a mixture of convergent and linear approaches may prove beneficial. For example, in a synthesis of mucin-like glycoprotein mimetics, the glycan was installed on a tripeptide fragment, which was then subjected to polymerisation to afford the target species.[123]

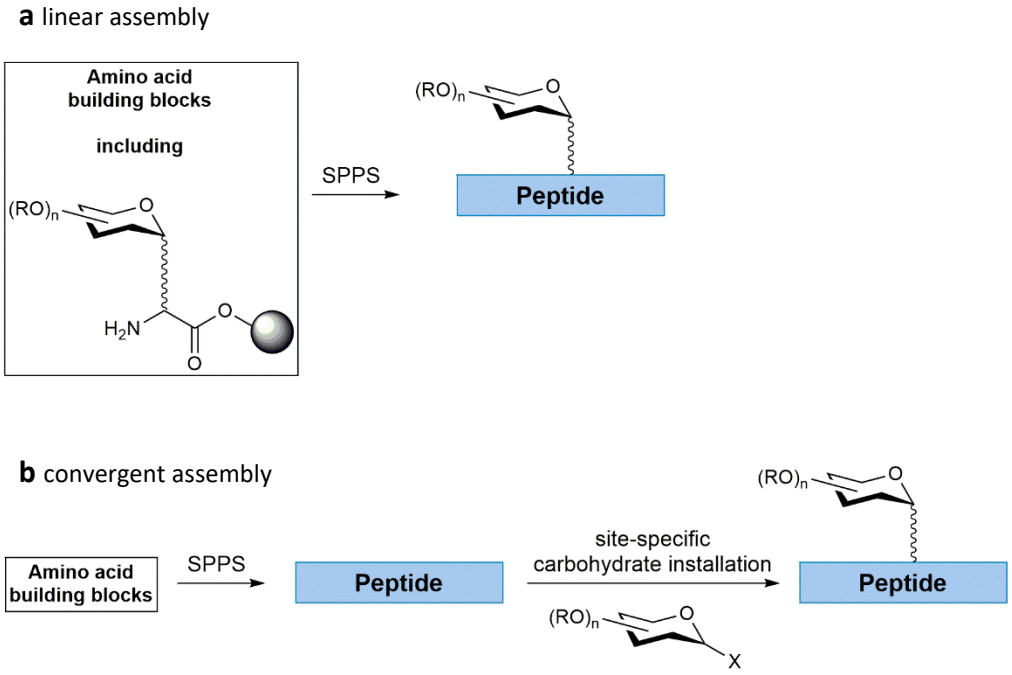


Figure 1.13 Conceptual strategies for the synthesis of glycopeptides: linear assembly (a) and convergent assembly (b).

The precise glycosylation method utilised to ligate glycan and amino acid depends on the nature of the glycosidic linkage. Naturally occurring glycoproteins fall into two classes: *N*-linked and *O*-linked.[93] In the former, the glycan is linked to the protein via an amide linkage at asparagine. Common methods for the synthesis of this native linkage include the coupling of glycosyl amines to Asp derivatives (as noted above),[121] or Staudinger ligation of glycosyl azides to either Asn or Asp derivatives,[124,125] in each case forming an *N*-glycosylated asparagine residue (**figure 1.14 a**); such methods have found application in convergent strategies.[122,124,125] Further alternatives include direct glycosylation of primary amides,[126] or amidation of aspartic anhydrides.[127] *O*-linked glycoproteins involve linkages at threonine, serine or (as in GYG) tyrosine.

These can be synthesised through glycosylation of protected threonine, serine or tyrosine derivatives with a range of suitable glycosyl donors (e.g. thioglycosides, trichloroacetimidates, halides, **figure 1.14 b**);[95,128–130] such approaches are usually employed in linear-based strategies.

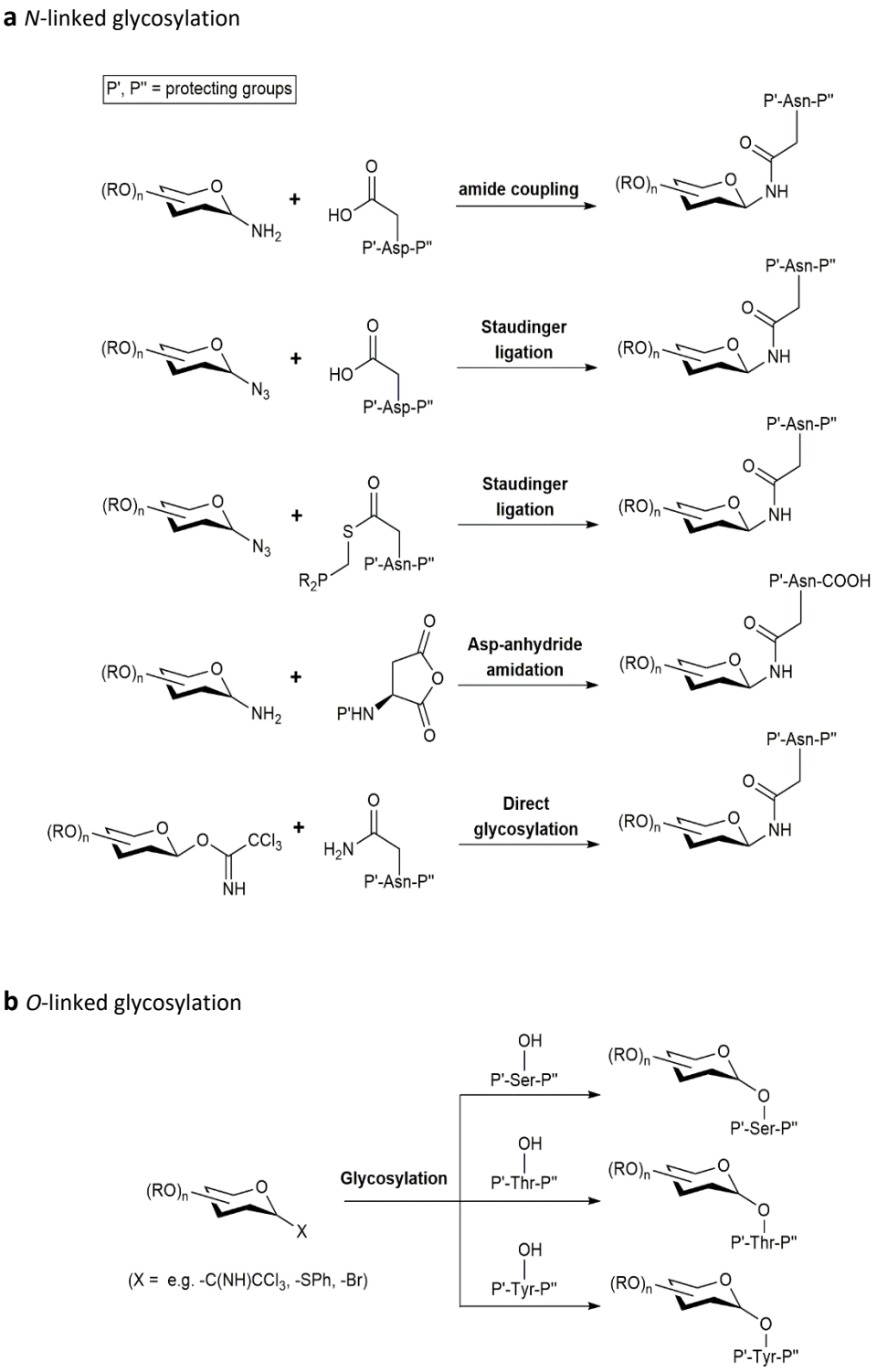


Figure 1.14 Strategies for the installation of *N*-glycosidic (a) and *O*-glycosidic (b) linkages to amino acids, in the context of chemical protein synthesis.

Given the fact that total chemical synthesis can evidently enable incorporation of a well-defined glycosylated amino acid at a precisely defined site in a protein, one might envisage that this methodology could readily yield homogeneous GYG-1 glycoforms through incorporation of glycosylated tyrosine in place of the catalytic Tyr195. Indeed, the chemical synthesis of GYG-1 peptide fragments containing Tyr195-glucose and Tyr195-maltose derivatives has been previously reported,[130] and there is no intrinsic reason why total synthesis could not afford homogeneous GYG glycoforms that exactly replicate corresponding *in vivo* species.

However, whilst recent years have seen the chemical synthesis of increasingly challenging, large protein targets,[131–133] there remain restrictions on the size of proteins accessible through this approach;[113,134,135] the largest chemically synthesised protein to date contains 352 amino acid residues.[131] Whilst the smallest possible functional, and thus useful, GYG protein (the catalytic domain of the GYG-1 isoform) comprises 262 amino acids,[59] this nonetheless remains towards the upper end of what is presently feasible, especially given additional complexity afforded by the need for glycan introduction. As such the synthesis of *multiple* GYG-1 glycoforms would surely represent a highly challenging and time-consuming approach. This represented the principal reason for not pursuing this approach in this research (particularly in light of significantly more facile, rapid alternatives as discussed subsequently), however other drawbacks include a requirement for protection/deprotection regimes (especially for installation of *O*-glycosylated amino acids),[92,130] and the frequent requirement for a refolding step post-synthesis.[134,136]

1.3.2 Drawing Inspiration from Nature: Chemoenzymatic Approaches

Nature uses glycosyltransferase and glycosidase enzymes in the biosynthesis of *N*-linked and *O*-linked glycoproteins, yet as noted previously highly heterogeneous mixtures of glycoforms typically result.[91,93] Homogeneous glycoproteins can however be accessed in some instances through expression of these types of enzymes and their application in a controlled, ordered fashion to proteins of interest in an *in vitro* setting.[92,95,137] Furthermore, glycosyltransferase enzymes may be utilised to elaborate simpler glycan structures installed on glycopeptides or glycoproteins through other methods, such as those outlined in the previous and subsequent sections.[138–140]

In certain cases, glycosyltransferase enzymes may be utilised to directly install a specific, homogeneous carbohydrate moiety onto a protein, rather than merely elaborate or alter a pre-existing glycan. This concept has been demonstrated for both the cases of glycopeptides and full-length glycoproteins;[141,142] the latter is

well illustrated by the use of the glycosyltransferase PglB to transfer defined oligosaccharides to a protein substrate.[143] However, in such cases, the site-specificity is typically defined by a specific consensus sequence, limiting their broader application.

A related and more widely applicable concept is that of glycoprotein remodelling,[144,145] in which the glycan moieties of a heterogeneous mixture of glycoforms are first trimmed through use of suitable enzymes, thus yielding a homogeneously glycosylated population. The carbohydrate may subsequently be elaborated through the use of glycosyltransferase enzymes, so as to effect transfer of a homogeneous synthesised glycan to the trimmed enzyme, yielding a homogeneous glycoprotein (**figure 1.15**). A key requirement of this method is that the trimming step does not remove the entire glycan, but leaves at least one carbohydrate residue – typically GlcNAc[137] – to act as a "tag", in order to facilitate subsequent glycan extension specifically from this site.

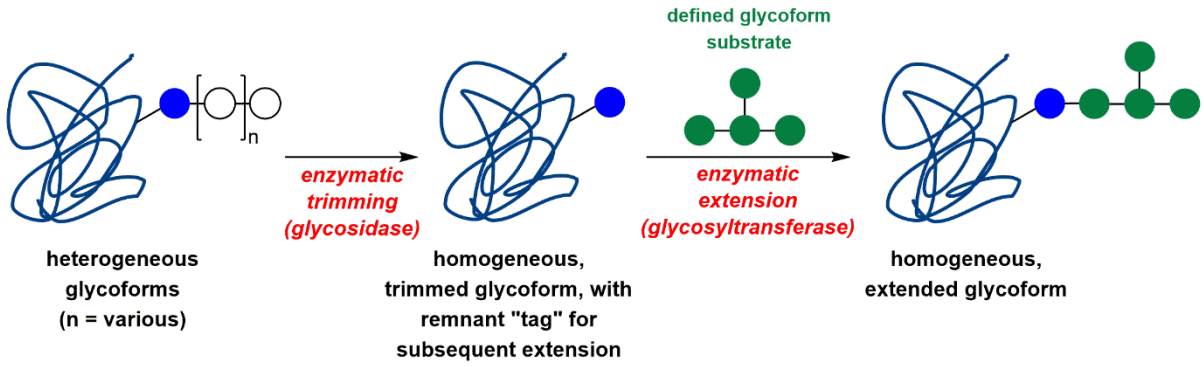


Figure 1.15 The concept of glycoprotein remodelling. Heterogeneous glycoforms are subjected to enzymatic "trimming" to yield a homogeneous, trimmed species bearing (at least) one remaining carbohydrate residue. This serves as a "tag" to direct subsequent enzymatic extension with a defined, homogeneous substrate can yield more elaborate homogeneous glycoforms.

This method has been widely utilised, in particular through the use of endo-β-*N*-acetylglucosaminidase (Endo) enzymes (e.g. Endo S, Endo H), which may cleave large glycans to yield only a single GlcNAc core.[146–148] Notably, Endo enzymes (e.g. Endo M, Endo A, Endo S) can also catalyse the subsequent elongation step from this residue, using different (oligosaccharide) oxazolines as substrates.[146] This methodology has been applied to the generation of highly pure glycoforms of the antibody Herceptin, enabling generation of homogeneous antibody-drug conjugates (ADCs).[149] Whilst Endo-catalysed product hydrolysis can lower yields, the use of mutant Endo enzymes for the elongation step can overcome this limitation.[149–151]

Such approaches, if applicable to the protein of interest, have great synthetic power. However, key requirements are either a requisite consensus sequence (direct enzymatic glycosylation), or the existence of a suitable trimming enzyme capable of trimming most – but not all – of the glycan (glycoprotein remodelling).

With regards to the former, GYG – in common with most *O*-linked glycoproteins[93] – lacks a particular consensus sequence from which glycosyl transfer by another enzyme could be effected (only GYG itself can effect glycosyl transfer to the catalytic tyrosine), and the introduction of such a sequence would require extensive protein engineering. For the latter case, whilst enzymes do exist that are known trim the carbohydrate moiety of heterogeneously glucosylated GYG-wt obtained from expression *in vitro* – isoamylase[152] and amylase[57] – the former cleaves the entire oligosaccharide chain to yield glycan-free GYG (thus affording no advantage over expression in *E. coli*, and failing to leave a "tag" to direct subsequent extension), whilst the latter yields heterogeneous mixtures.[54,57] Indeed, this strategy has found greatest application for *N*-linked glycoproteins, and that the trimming step almost always yields a core GlcNAc residue.[95,153] As such, whilst the broader utility of such methods cannot be disputed, they are somewhat inapplicable to the specific goals of this research.

1.3.3 Genetic Incorporation: Codon Suppression Technology

A potential alternative method to access homogeneous glycoproteins revolves around the direct genetic incorporation of a glycosylated amino acid into the protein of interest during protein biosynthesis, utilising the natural cellular biosynthetic machinery.[154,155] The fact that 3 of the 64 naturally occurring codons – TAG (amber), TGA (opal) and TAA (ochre) – do not encode an amino acid partner, instead acting as "stop codons" to cause termination of protein translation, opens up the possibility for the reassignment of one of these to other non-natural amino acids. The incorporation of such an amino acid in response to such a codon is referred to as codon suppression technology; amber codon suppression represents the most popular option owing to the rarity of this codon in the common expression host *E. coli*.[156] The mechanism of such codon suppression is discussed in greater detail in chapter 2 in a different context, and a number of comprehensive reviews on the topic exist.[157–159] The basic principles are however briefly overviewed here, set in the context of glycoprotein synthesis via glycosylated amino acid incorporation (although the concept is the same for incorporation of other amino acids).

Natural translation relies upon the recognition of a codon within the mRNA template via base pairing to the anticodon loop of a tRNA molecule charged with the requisite amino acid. This amino acid will then be incorporated into the nascent polypeptide chain during translation on the ribosome. For codon suppression, a "suppressor tRNA", i.e. a tRNA capable of recognising either of the amber, opal or ochre stop codons is required, and this tRNA must be charged with the non-natural glycosyl amino acid.[157] A DNA template containing one such codon at the desired position is transcribed to mRNA (containing the corresponding codon). During subsequent translation, the charged suppressor tRNA recognises the stop codon, leading to incorporation of the glycosyl amino acid into the protein during its biosynthesis on the ribosome (see also chapter 2).[157-159] Given that stop codons can be uniquely introduced into template DNA at precise locations via site-directed mutagenesis, this strategy can enable generation of homogeneous glycoforms. This concept is illustrated below, for the example of amber codon suppression (**figure 1.16**).

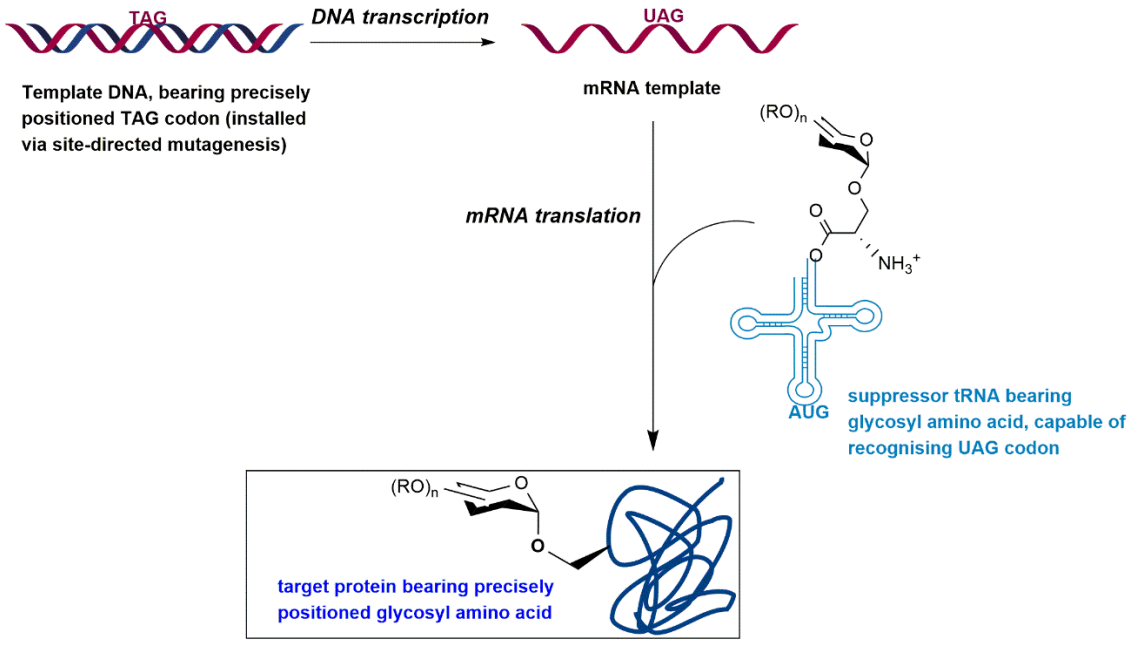


Figure 1.16 Codon suppression technology applied to the synthesis of homogeneous glycoproteins, illustrated for the specific example of amber codon suppression. A template DNA bearing the amber codon is transcribed and translated; addition of a tRNA charged with glycosyl amino acid and capable of recognising the amber codon leads to incorporation of this amino acid during protein translation.

This approach has been successfully applied for the site-specific introduction of glycosylated amino acids to yield homogeneous glycoforms of proteins of interest (e.g. firefly luciferase, DHFR).[154,155] However, in both cases, the tRNA was "pre-charged" with the unnatural amino acid by chemoenzymatic synthesis prior to addition to the protein expression system. This is in itself far from straightforward, and additionally limits

application of the method largely to specialised *in vitro* transcription/translation systems, not to more commonly utilised *in vivo* expression systems (e.g. *E coli*). Nature uses dedicated aminoacyl synthetase (aaRS) enzymes to charge specific tRNA molecules with their partner amino acid; the expansion of codon suppression technology to such *in vivo* systems thus requires the evolution of an aaRS that can specifically charge the suppressor tRNA with the non-natural amino acid (see chapter 2).[160] Whilst such pairs now exist for a large variety of amino acids, allowing their facile incorporation *in vivo* without recourse to chemical tRNA synthesis, no such pair has been successfully obtained for glycosylated amino acids; indeed, one notable study claiming to have achieved this was later retracted.[161]

Given the above limitations, this method has yet to be widely utilised for homogeneous glycoprotein synthesis, although ongoing development in codon suppression technology may change this in the future. It remains in the relatively early stages of development, and is somewhat limited in its generality. There is no inherent reason why this approach could not nonetheless be adopted for the synthesis of homogeneous GYG glycoforms that could exactly replicate native variants; one would require a tRNA charged with, for example, glucosyl-tyrosine moiety that could be incorporated in place of the native Tyr195 of GYG-1, in a system lacking UDP-glucose to avoid further "in-expression" glucosylation. Indeed, this avenue of research represents an ongoing interest within our research group, and preliminary experiments to this end have been carried out.[162] However, we considered that other, more established methods may prove more fruitful for the present work, allowing relatively more facile construction of GYG glycoforms and study thereof.

1.3.4 Engineering Glycosylation Pathways

Genetic incorporation of glycosylated amino acids relies on perturbation of the biosynthesis of the glycoprotein at the level of protein translation. An alternative approach successfully utilised for some glycoproteins is to perturb instead the biosynthetic pathway of exclusively the carbohydrate moiety to yield homogeneous glycoproteins. There are a wide variety of strategies by which this may be accomplished that have been comprehensively reviewed elsewhere;[95,144,145] some key ones are outlined below.

In mammalian cells, the biosynthesis of glycoproteins involves a complex array of glycan-processing enzymes.[93,94] However, the generation of mutant mammalian cell lines deficient in one or more such enzymes offers the potential to simplify this pathway, yielding defined glycoforms. For example, homogeneously glycosylated rhodopsin was obtained through expression in a mutant HEK cell line deficient in *N*-acetylglucosaminyltransferase activity,[163] and a wide range of mutant CHO lines with altered

glycosylation profiles have been generated.[164] Alternatively, the use of small-molecule inhibitors of glycosidase enzymes has been demonstrated to yield homogeneous IgG glycoforms following expression in mammalian cells.[165]

A different approach involves the use of non-mammalian expression systems. For example, knockout of plant-specific glycosylation enzymes and introduction of specific human glycosyltransferases can yield plant lines capable of producing "humanized" glycoforms in high homogeneity; this principle was applied to the production of homogeneously glycosylated erythropoietin.[166] A similar principle can also be applied to afford yeast strains capable of biosynthesising complex and highly uniform human glycoproteins,[167–169] an approach that has been applied to the synthesis of homogeneously glycosylated human antibodies.[170]

However, it is readily apparent that such strategies, whilst of undoubted value, are inapplicable to the issue of GYG glycoform heterogeneity under investigation in this work. Glycoengineering approaches are principally of value when glycoform heterogeneity is the result of pathways involving multiple glycan-processing enzymes, such that simplification of such pathways (as described above) can ameliorate such heterogeneity. For GYG, the entire glycan portion is constructed by one enzyme – GYG itself – and the source of heterogeneity thereof is the non-uniform repetition of its glycosyltransferase action during expression. Thus, the biosynthetic pathway for this glycan cannot be readily simplified further, other than by preventing GYG activity during expression to yield glycan-free enzyme[89] which, as discussed earlier (section 1.2.3), has its own associated issues. Furthermore, such glycoform heterogeneity for GYG arises even in simple bacterial hosts (*E. coli*),[49,59,88] since it is not dependent on other cellular glycan-processing enzymes, thus altering the expression host is not a productive option.

1.3.5 Site-selective Chemical Modification of Proteins

The strategies outlined thus far can be crudely seen as representing a spectrum of approaches from "chemical" (protein synthesis) to "biological" (glycoengineering, genetic incorporation). An approach that falls somewhere around the midpoint of such a spectrum is the precise, site-selective chemical modification of a protein scaffold. With regards to the convergent and linear assembly approaches discussed previously in the context of chemical synthesis of glycoproteins (section 1.3.1), this might be considered as the ultimate example of a convergent method; the carbohydrate moiety is ligated not just to a pre-formed peptide fragment, but instead to the entire protein.

Precise chemical modification of protein substrates is frequently discussed in the context of the so-called "tag-and-modify" concept.[171] This involves the initial installation of a uniquely reactive amino acid residue "tag" at a defined, specific site in a protein substrate. Subsequent site-selective chemical modification of this tag can generate a homogeneously modified construct (**figure 1.17 a**). Thus, if this modification step involves a carbohydrate derivative, a homogeneous glycoprotein can result (**figure 1.17 b**).

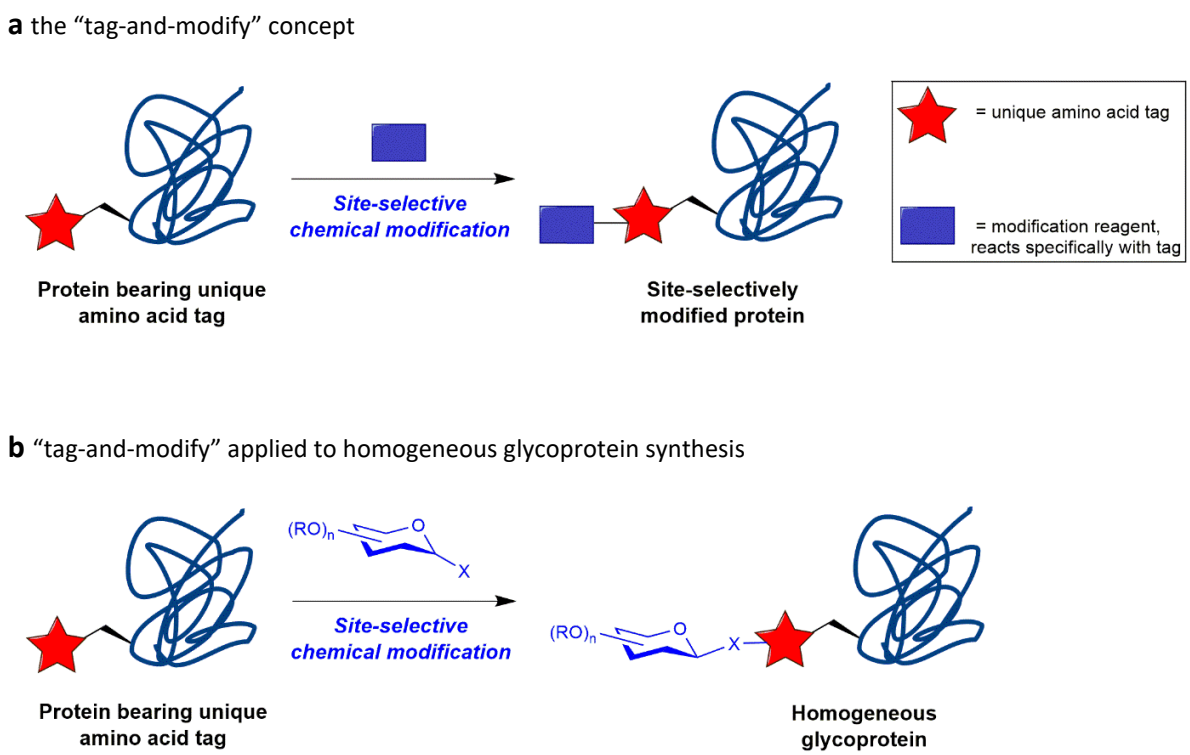


Figure 1.17 (a) the "tag-and-modify" concept for site-selective chemical modification of proteins. (b) the application of this concept to homogeneous glycoprotein synthesis.

The precise (chemical) natures of tags commonly utilised, and strategies employed for their selective modification, are discussed in greater depth in section 1.4. At this point, it is simply stated that commonly utilised tags may be natural amino acids (or derivatives thereof), or unnatural amino acids. Briefly, tags in the first category most commonly include cysteine and derivatives thereof; the former may if not already present, be installed at a desired position through site-directed mutagenesis.[171,172] With regards to the latter category, whilst the challenges of the use of codon suppression in the direct incorporation of glycosyl amino acids were discussed previously (section 1.3.3), this approach is far more applicable to the site-specific incorporation of a wide range of other unnatural amino acid tags,[158,159] with the existence of $tRNA_{suppressor}$/aaRS pairs for such species enabling their incorporation in *in vivo* expression systems.

Alternatively, some unnatural amino acids may be incorporated through the use of auxotrophic bacterial strains, exploiting the flexibility of existing tRNA/aaRS pairs for alternative amino acids.[173,174]

The previous methods outlined were either entirely infeasible for the generation of GYG glycoforms (glycoprotein remodelling, direct enzymatic extension, and glycoengineering) or had significant drawbacks that were discussed previously (chemical synthesis is challenging for larger proteins, genetic incorporation is at present a nascent technology for glycoprotein synthesis). In contrast, we felt that chemical protein modification could represent a comparatively facile method in this regard. Specifically, it was envisaged that the tyrosine glycosyl acceptor of GYG (i.e. Tyr195 of GYG-1) could be readily replaced with a suitably unique amino acid tag, incorporated via standard biological methods (mutagenesis, codon suppression, or auxotrophic incorporation). This tag may be subsequently modified to generate homogeneous glycoforms or close mimetics thereof that may then be utilised for mechanistic studies, circumventing the aforementioned disadvantages associated with conventional approaches for GYG isolation (section 1.2.3). This concept is illustrated for a generic, hypothetical example below (**figure 1.18 a**, generating a mono-glucosylated species), however it is evident that, by variation of the carbohydrate reagent, a range of distinct glycoforms could be accessed.

Furthermore, whilst the ultimate goal would indeed involve the ligation of carbohydrate derivatives to tagged GYG species as illustrated (**figure 1.18 a**), one might envisage that the carbohydrate moiety may be crudely mimicked by simpler, more readily available structures bearing a pendant hydroxyl group of suitable orientation and pK_A to mimic the carbohydrate glycosyl acceptor function; this may enable more facile generation of "cruder" mimetics of enzyme glycoforms that may yet also exhibit activity (**figure 1.18 b**). This is an avenue that will be discussed and explored in depth in chapter 2, although in following discussions (section 1.4) the focus will remain the ultimate goal of carbohydrate-protein ligation.

A further interesting aspect of such a tag-and-modify strategy is the potential for development of a novel mode of chemical enzyme activation. Replacement of the natural glycosyl acceptor tyrosine (Tyr195 in GYG-1) with a suitable "tag" should yield a GYG species incapable of being glucosylated (assuming the tag itself is not a glycosyl acceptor). The purpose of chemical modification of this tag is precisely to generate a catalytically species for subsequent biological study, thus it is immediately apparent that such a modification could represent an "off → on" enzyme switch (**figure 1.19**). This concept will be discussed and placed in context in chapter 3, however at present we note that most such "enzyme activation" strategies rely on uncaging of a suitably protected residue, as opposed to a tag-and-modify approach.[175–178] As such, this

rendered the application of a chemical modification strategy to GYG of additional interest from a methodological perspective.

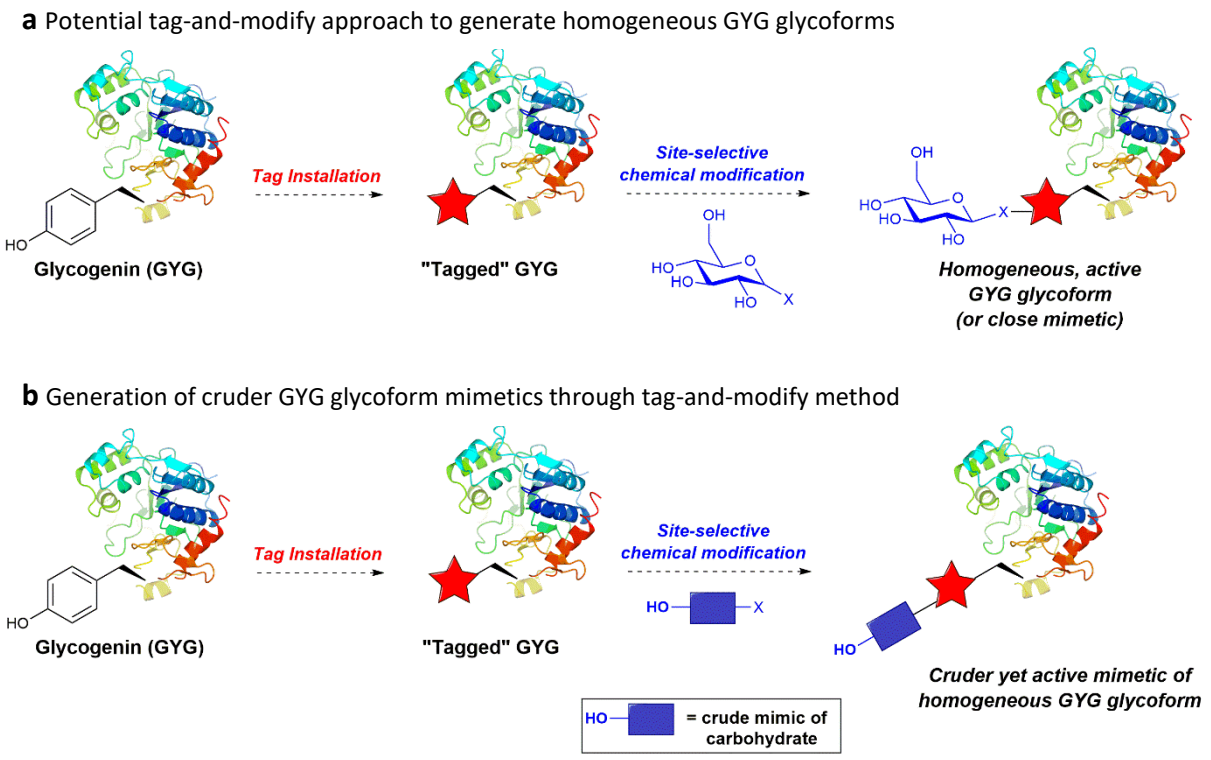


Figure 1.18 Tag-and-modify can allow access to homogeneous, potentially active GYG glycoforms (a). Alternatively, cruder mimetics of GYG glycoforms could be generated that simply bear a glycosyl acceptor at the active site, which may be sufficient to lead to catalytic activity (b).

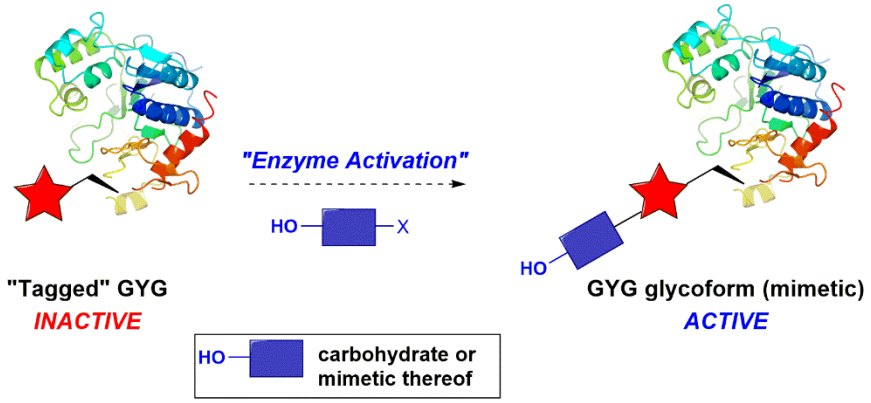


Figure 1.19 Chemical modification of GYG has the potential to serve as a novel "off → on" switch.

Site-selective chemical modification as applied to homogeneous glycoprotein synthesis is not without its drawbacks. Principal amongst these is the fact that the linker between the protein and carbohydrate is frequently non-identical to that found in the corresponding natural variant, reflecting inherent limitations in the chemistries available to us in this context. However, as discussed in depth in chapter 3, numerous examples highlight the fact that exact structural correspondence is not a pre-requisite for replication of native biological functionality of glycoproteins.[139,179–181]

We considered that such a factor was outweighed by the undoubted strengths of site-selective chemical modification as a tool for glycoprotein generation. In particular, a tag-and-modify approach enables the comparatively rapid generation (c.f. total synthesis) of multiple modified (glyco)protein variants from a common tagged precursor, itself often straightforwardly generated. Furthermore, a large number of different tag-and-modify chemistries are now well-established; many involve the use of amino acid tags for which robust methods for their installation exist (e.g. mutagenesis, or increasingly well-established amber codon suppression or auxotrophic incorporation systems),[171,182] and many have previously been applied to glycoprotein generation (see section 1.4).[92,183–185] Finally, as noted above, this strategy contained within it the potential to develop an exciting, novel "enzyme switch" methodology.

Such factors, and the drawbacks or inapplicability of alternative methods noted previously, highlight chemical modification as a suitable approach for this research. The success of such a strategy could depend critically on the precise nature of the approach used. To this end, the following section provides an overview of site-selective modification techniques utilised for homogeneous glycoprotein generation, in order to rationalise our ultimate choice of chemical approach for the generation of homogeneous GYG glycoforms.

1.4 Site-selective Chemical Modification Approaches to Glycoprotein Synthesis

As outlined in the previous section, site-selective chemical approaches for protein modification, with reference to the "tag-and-modify" concept, may be broadly divided into two distinct classes: those using natural amino acid tags – typically cysteine or derivatives – and those using tags based on unnatural amino acids. Here, the chemical nature of such tags and the chemistries utilised to site-selectively modify them, in the context of glycoprotein synthesis, are outlined in greater detail for each class. Subsequently, our final choice of tag-and-modify strategy for GYG is discussed and rationalised.

1.4.1 Cysteine as a Tag for Homogeneous Glycoprotein Generation

Of the 20 canonical amino acids, cysteine represents one of the most uncommon,[186] rendering it suitable as a tag for site-selective modification. A cysteine residue may be introduced at the desired modification site – if not already present – via site-directed mutagenesis. In order to ensure product homogeneity, it is necessary that, in addition to the modification being Cys-selective, the installed tag represents the only (reactive) cysteine residue. However, given the rarity of the amino acid, this may well be the case for a number of protein substrates.

A number of methods exist for the selective modification of cysteine tags to generate glycoproteins, most of which exploit the inherent nucleophilicity of its thiol moiety. These include reaction with glycosyl iodoacetamide reagents (**figure 1.20 a**),[187,188] with glycosyl maleimides (**figure 1.20 b**),[189,190] and formation of disulphide species. The latter may be achieved via a variety of methods, including reaction with carbohydrate dithiopyridyl reagents[191] (**figure 1.20 c**), with glycosyl selenosulfide species (**figure 1.20 d**),[192] or with glycosyl methyl- or phenylthiosulfonates (respectively referred to as glyco-MTS and glyco-PTS reagents, **figure 1.20 e**).[193,194] Notably, the use of a glyco-MTS reagent enabled site-selective modification of a cysteine residue in the cell adhesion protein PSGL-1 as part of an impressive dual modification strategy, enabling construction of accurate, functional mimics of this naturally glycosylated protein.[179] Finally, in addition to its more conventional reaction as a nucleophile, cysteine may also react as a thiyl radical in the context of a thiol-ene reaction (**figure 1.20 f**). Dondoni *et al* demonstrated the application of this reaction to glycoprotein synthesis, however owing to the existence of 3 naturally occurring cysteine residues in the protein substrate the reaction was not site-selective;[195] nonetheless, its potential for homogeneous glycoprotein synthesis using single-Cys substrates is evident.

1.4.2 Tags Involving Cysteine Derivatives

A number of techniques for glycoprotein synthesis via site-selective chemical modification involve the further derivatisation of the cysteine tag prior to the final modification step (with regards to terminology, "modification" in to context of glycoprotein synthesis is taken to refer to reaction with the carbohydrate reagent; any steps prior to this represent "tag derivatisation"). Prominent amongst these is the chemical formation of the uniquely reactive amino acid dehydroalanine (Dha) from cysteine, usually generated by treatment of the protein with a bis-alkylating agent that may undergo subsequent elimination (**figure 1.21**).[196]

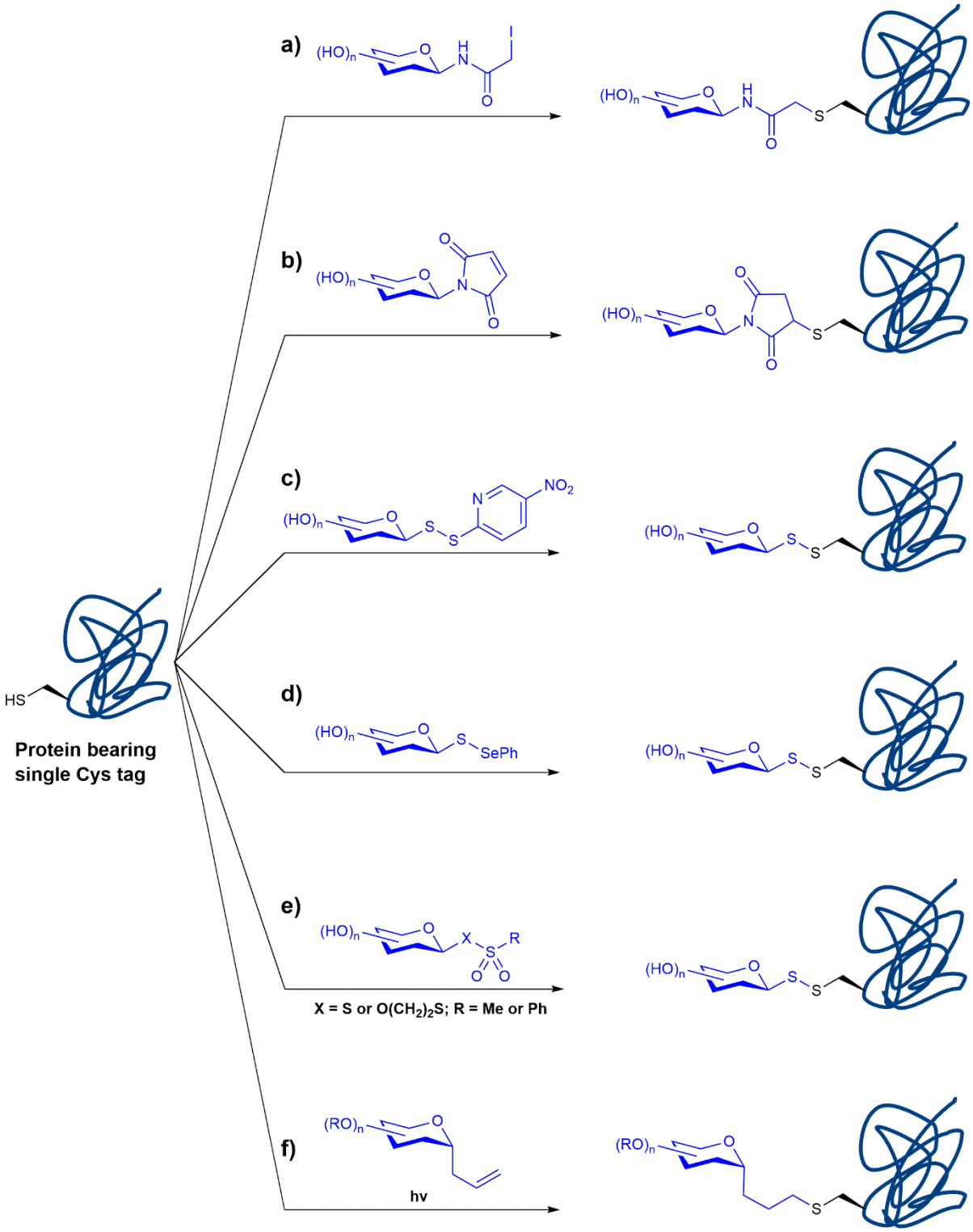


Figure 1.20 Methods for the synthesis of glycoproteins using a Cysteine tag: alkylation with iodoacetamides (a) or maleimides (b), disulphide bond formation using a variety of reagents (c-e), or thiol-ene chemistry (f).

Subsequent modification of the Dha tag thus installed can be affected by two classes of reaction. The first exploits the electrophilic, Michael-acceptor properties of Dha in its reaction with nucleophilic species, predominantly thiols. In the context of glycoproteins, this implies the use of readily attainable thioglycosides as reactive partners; a number of examples exist, particularly using GlcNAc thiol (**figure 1.21**).[138,180,196–198] Alternatively, Dha may act as a radical acceptor in reaction with carbon-based radicals derived typically from alkyl iodides, facilitating a novel C(sp^3) – C(sp^3) bond formation process.[134,199] The broad tolerance of this reaction enables site-selective installation of a huge range of amino acid side-chains, including natural, phosphorylated, and glycosylated.[134] It should be noted, however, that reactions at Dha on protein substrates are believed to yield epimeric products.

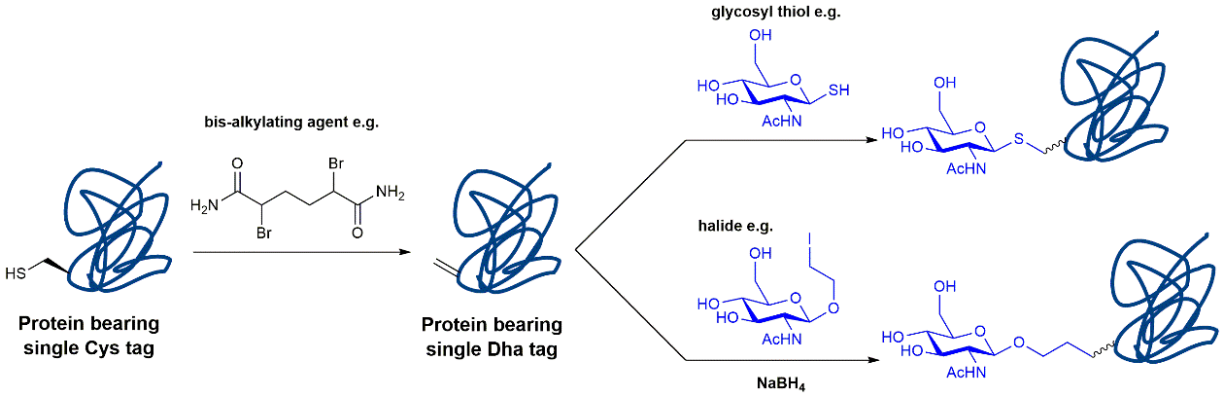


Figure 1.21 Accessing glycoproteins from dehydroalanine via either thia-Michael addition or radical chemistry.

Alternative tags derived from cysteine include the unnatural amino acid *S*-allylcysteine (Sac), which can be installed via simple *S*-alkylation or via generation of Dha and reaction with mercapto thiol (**figure 1.22 a**).[200–202] Subsequent homogeneous glycoprotein formation can be effected through olefin metathesis chemistry. Similarly, the cysteine tag may be converted to *p*-iodobenzylcysteine, which represents a substrate for Suzuki-Miyaura cross-coupling to a variety of boron derivatives, including carbohydrate species (**figure 1.22 b**).[203] This latter reaction will be re-visited in subsequent sections in different contexts (section 1.4.3, section 1.5).

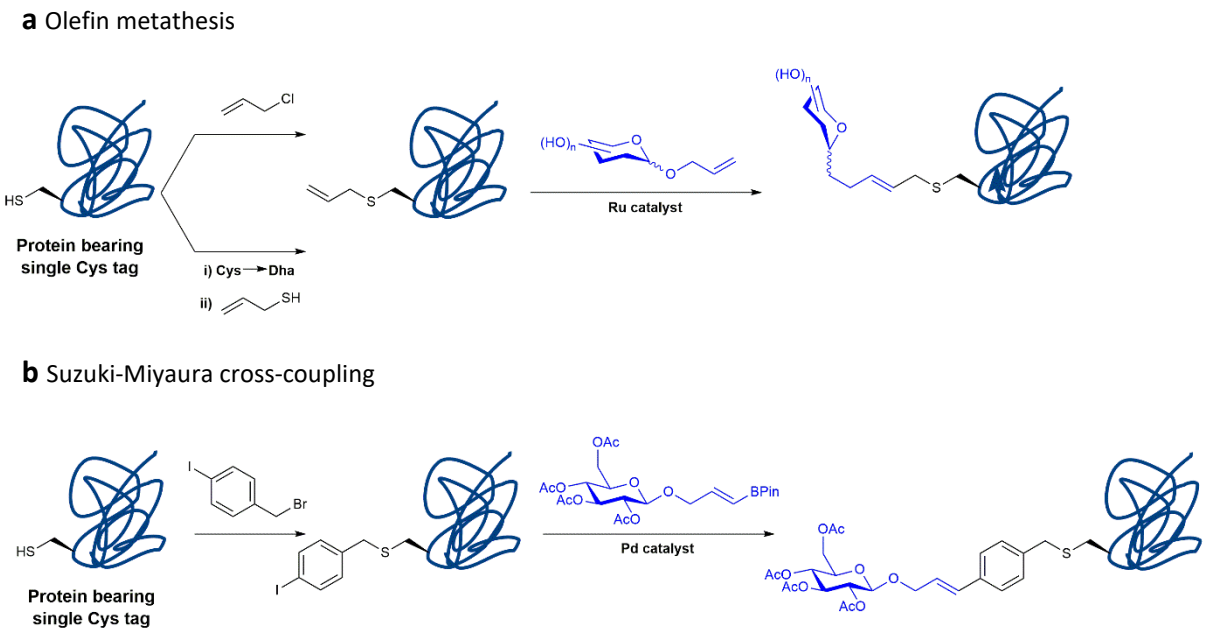


Figure 1.22 The application of olefin metathesis (a) and Suzuki-Miyaura cross-coupling (b) to glycoprotein generation using tags derived from cysteine.

1.4.3 Unnatural Amino Acids as Tags for Glycoprotein Synthesis

Methods dependent upon the use of cysteine or derivatives as tags for homogeneous glycoprotein generation do however suffer from inherent drawbacks. Firstly, whilst in numerous cases proteins bearing single – and thus unique – cysteines are readily attainable, other proteins may contain multiple reactive cysteine residues. This may either render reaction at cysteine or Cys-derived tags non-site-selective, or necessitate multiple site-directed mutagenesis steps to eliminate other Cys residues. The latter may increasingly perturb protein structure or function, especially true given the frequent occurrence of Cys residues in structurally important disulphide bridges. Secondly, regardless of whether cysteine or a derivatve is ultimately used as the tag, a necessary chemical step involves the reaction of cysteine as a nucleophile. This may potentially lead to chemoselectivity issues, since other naturally occurring amino acids are nucleophilic; for example, histidine and lysine may also react with iodoacetamides.[188,204] The use of unnatural amino acid tags opens up new reaction manifolds that are entirely bio-orthogonal to naturally occurring residues.[205,206]

As discussed previously, unnatural amino acids may be incorporated into protein substrates via either codon suppression or auxotrophic replacement techniques. The latter approach leads to the global replacement of a natural amino acid – usually methionine – and as such may still cause issues with site-selectivity of the reaction. In contrast, codon suppression can enable mono-incorporation of an unnatural amino acid at a precise site through introduction of a stop codon at a defined position in the DNA sequence via mutagenesis and recognition of this codon by an appropriate tRNA/aaRS pair bearing the unnatural amino acid (see section 1.3.3 and chapter 2 for further details). Nonetheless, both approaches have found utility in the context of glycoprotein synthesis via chemical modification. Some examples are discussed and illustrated below (**figure 1.23**).

1) Staudinger Ligation

The Staudinger ligation, a modification of the Staudinger reaction,[207] refers to the reaction of an azide with a triarylphosphine in order to generate an amide linkage.[208] Of importance is the presence of a suitable electrophilic trap on an aryl group of the phosphine moiety, which enables formation of this linkage instead of simple hydrolysis to the free amine. A "traceless" variant of the Staudinger ligation was subsequently reported, in which the presence of the phosphine oxide moiety in the product is avoided;[209] this was later applied to the synthesis of homogeneous *N*-linked glycoprotein constructs via ligation to azidohomoalanine (Aha) (**figure 1.23 a**),[210] which may be incorporated in place of methionine via auxotrophic methods.[211] Furthermore, *p*-azidophenylalanine (*p*AzPhe) may be incorporated via codon suppression, bringing associated advantages mentioned above with regards to site-selectivity; whilst Staudinger ligations to *p*AzPhe-containing proteins are known, this is not the case in the context of glycoprotein synthesis.[212] However, the Staudinger ligation has fallen out of favour somewhat, as a consequence of factors such as relatively slow kinetics, potential for side-reactions, incompability with certain amino acids,[210] and development of alternative methods for azide-based protein modification, not least cycloaddition approaches as discussed below.

2) Azide-Alkyne Cycloaddition

The copper-catalysed azide-alkyne cycloaddition (CuAAC) has been, and continues to be, widely used the context of protein modification for a variety of purposes, a consequence of its desirable characteristics (rapid kinetics under benign conditions, excellent selectivity, aqueous compatibility). The reaction involves, as the name would suggest, the formation of a triazole from alkyne and azide partners. Glycoproteins may thus be

generated through the reaction of alkyne-containing carbohydrates with Aha-containing proteins (**figure 1.23 b**).[179,183,213] However, given that the amino acid homopropargylglycine (Hpg) may also be incorporated into proteins via auxotrophic replacement of methionine, they may be additionally formed via CuAAC between Hpg-containing proteins and glycosyl azides (**figure 1.23 c**).[214]

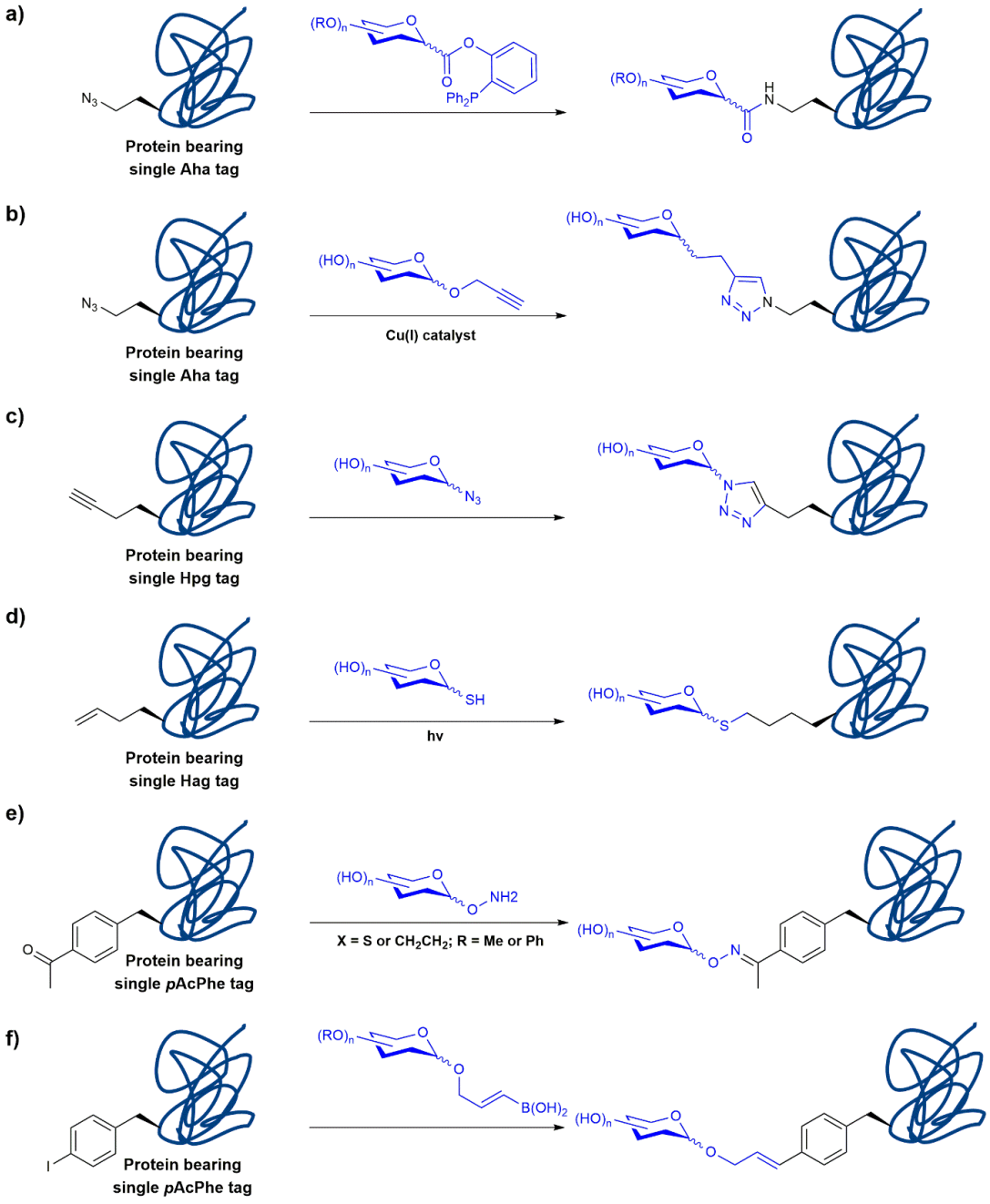


Figure 1.23 The use of unnatural amino acid tags in the generation of glycoproteins. Examples include the Staudinger ligation to Aha (a); CuAAC to Aha (b) or Hpg (c); thiol-ene chemistry to Hpg (d); oxime formation at *p*AcPhe (e) or Suzuki-Miyaura cross-coupling to *p*IPhe (f). Amino acid tags are incorporated via auxotrophic means (a-d) or amber codon suppression (e, f); whilst azide (*p*AzPhe) and alkyne (*O*-propargyl Tyr) tags may additionally be incorporated via the latter method, these have not been utilised to date in glycoprotein synthesis

Aryl-containing azide (*p*AzPhe) and alkyne handles (*O*-propargyl tyrosine) may additionally be incorporated through codon suppression techniques, and the utility of both in CuAAC reactions has been generated, although surprisingly not for the synthesis of glycoproteins.[215] It should also be noted that copper-free cycloaddition reactions have more recently been applied to bioconjugation, in particular strain-promoted cycoadditions (SPAAC), involving reaction of azide-containing proteins with highly strained cyclic alkynes, and inverse electron demand Diels-Alder reactions (IEDDA) involving reacting of genetically encoded cyclic alkenes with tetrazines.[216,217] However, presumably owing to the difficulty of synthesising requisite carbohydrate derivatives, these alternatives have yet to find significant application in the generation of protein glycoconjugates.

3) Thiol-ene Reaction

The use of the thiol-ene reaction as a means for glycoprotein generation utilising a cysteine tag was described previously. Here, we briefly note that the reaction can also be applied in a reverse sense, since an alkene-containing amino acid – homoallylglycine (Hag) – can be incorporated auxotrophically, and may be ligated to thioglycosides to thus generate glycoproteins (**figure 1.23 d**).[218]

4) Oxime Formation

The reaction of hydroxylamines with either aldehydes or ketones leads to the formation of oxime linkages. Bertozzi *et al* demonstrated the utility of this approach in coupling an aldehyde-containing disaccharide to a glycopeptide bearing a pendant hydroxylamine group to facilitate elaboration of the carbohydrate moiety.[219] The subsequent establishment of a method to incorporate the ketone-containing unnatural amino acid *p*-acetylphenylalanine (*p*AcPhe) via amber codon suppression opened up new possibilities for site-selective chemical modification.[220] Proteins do not naturally contain any reactive carbonyl groups (many carbonyl moieties are present but in the context of poorly electrophilic amides or carboxylates), rendering chemoselective modification of *p*-AcPhe distinctly plausible for homogeneous glycoprotein synthesis, as demonstrated by Schultz and coworkers (**figure 1.23 e**).[221]

5) Pd-Catalysed Cross-Coupling

The use of the Pd-catalysed Suzuki-Miyaura cross-coupling (SMC) between an aryl halide moiety and a boronic acid derivative in glycoprotein generation was briefly mentioned previously, but in the context of

modification of a tag installed via derivatisation of cysteine. However, the amino acid L-*p*-iodophenylalanine (*p*IPhe) can be incorporated through amber codon suppression (see chapter 2) in a site-specific manner.[222] This greatly enhanced the applicability of this methodology, affording as it did the potential for absolute site-selectivity of modification without requirement for a single-cysteine substrate.[223] Cross-coupling to carbohydrate-derived boronic acids can be effected in a site-selective fashion to yield glycoproteins (**figure 1.23 f**), as previously demonstrated by Spicer *et al* in an on-cell cross-coupling to a *p*IPhe-containing *E. coli* membrane protein.[224]

*p*IPhe-containing proteins have also found application as substrates for Sonogashira cross-coupling to terminal alkynes.[225] Indeed, this reaction has been demonstrated not only in an *in vitro* context but also inside *E. coli* cells.[226] However, to the best of our knowledge, its application to glycoprotein synthesis remains unexplored, and as such it is not discussed further.

1.4.4 The Choice of a Suitable Strategy for Glycogenin Modification

From the above overview, it is evident that an impressive and varied myriad of chemical modification techniques for the synthesis of well-defined glycoproteins exist. When considering which of these might be best employed to our overall goal – the generation of homogeneous GYG glycoforms at the catalytic tyrosine, through replacement of this tyrosine with a suitable tag and subsequent modification – a number of factors must be considered.

Firstly, in order that the glycoforms generated are indeed homogeneous, it is necessary that any tag-and-modify approach ensures site-selective modification at the position of the natural acceptor tyrosine (Tyr195) alone; thus, the tag must be specifically present only at this position. Yet the primary amino acid sequence of mammalian GYG-1 (the focus of this research) contains a total of 6 cysteine residues and 8 methionine residues; similarly, whilst not explicitly studied in this thesis, GYG-2 contains 10 cysteines and 7 methionines.[227] As such, approaches relying on the use of cysteine or derivatives thereof as tags, or on the use of unnatural amino acid tags incorporated via auxotrophic methods (which typically involve global replacement of methionine), are not suitable; the tag would not be unique without recourse to extensive, potentially structure- and function-altering, mutagenesis. Furthermore, such tags do not retain the aryl moiety of the natural tyrosine residue, and were deemed additionally disadvantageous for this reason.

In contrast, unnatural amino acids incorporated through codon suppression offer precise positioning of a unique tag at a defined site; as such, our choice was restricted to methods based on such tags (**figure 1.24**).

Of the reactions outlined above, these most clearly include Suzuki-Miyaura cross-coupling (SMC) to *p*IPhe and oxime formation at *p*AcPhe (**figure 1.24 a, b**). However, whilst in the context of glycoprotein synthesis the Staudinger ligation and CuAAC have been applied to tags incorporated using auxotrophic methods (see **figure 1.23** and previous discussion), such reactions have also been utilised to install alternative modifications from tags installed via codon suppression (*p*AzPhe, *O*-propargyl tyrosine).[212,215] It is evident that these tags could potentially be utilised in GYG glycoprotein generation (**figure 1.24 c-e**). The conceptual application of each of these reactions in the hypothetical ligation of a generic monosaccharide coupling partner to GYG to generate a mimetic of mono-glucosylated GYG is illustrated on the following page (**figure 1.24**); for comparative purposes, the corresponding natural glycoform is also shown.

A number of reaction characteristics were considered for these reactions. The reaction should naturally exhibit fast kinetics and be highly selective for the tag used, and the product thus formed should contain a stable carbohydrate-protein linkage. The requisite tagged protein and carbohydrate reagents should ideally be readily accessible. Finally, it is evident that, for all the reactions considered, there is a non-natural "linker" moiety between carbohydrate and GYG enzyme that is not present in natural glycoforms (**figure 1.25**, also see **figure 1.24**). As discussed previously, this is an inherent limitation of chemical modification in glycoprotein generation. Thus, any GYG glycoforms generated through this method are better termed "GYG glycomimetics". In order to create such glycomimetics that resemble native glycoforms as far as possible – desirable to generate catalytically active, biologically relevant enzyme constructs – this linker should clearly be as structurally insignificant as possible. This is discussed further in chapter 3.

We felt that SMC offered optimal characteristics in many of the above regards. The tRNA/aaRS system for the incorporation of the requisite *p*IPhe tag into GYG via amber codon suppression is robust and well-established[228] (and, more pragmatically, was readily available to us, in contrast to systems for *p*AzPhe and *p*AcPhe). Straightforward synthetic routes towards carbohydrate-derived boronic acids are known,[224] and the protein-carbohydrate linkage formed from their cross-coupling to GYG, resulting from C-C bond formation, represents a stable and irreversible one. The reaction has previously been demonstrated at physiological pH and temperature on protein substrates,[223] and is typically complete within 2-3 hours, minimising exposure to reagents and catalyst species. Furthermore, the sole (ideal) functional group pre-requisite for the non-natural linker moiety is a vinyl (or aryl) moiety, (since sp^3 couplings on proteins remain at present an unsolved challenge), the former of which we considered to represent a relatively unobtrusive group in a structural sense.

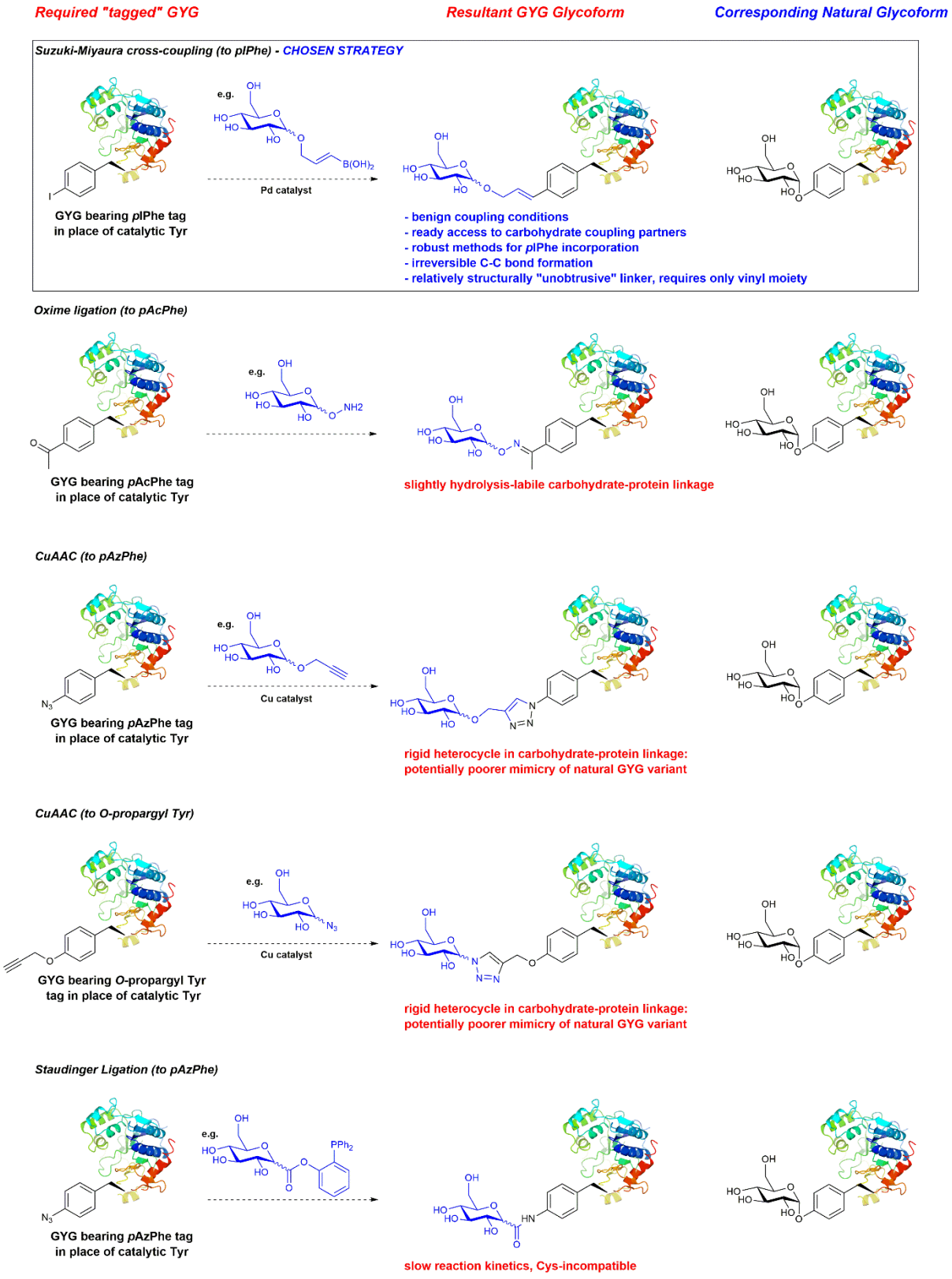


Figure 1.24 Methods considered for the generation of GYG glycoforms, based on tags incorporated through amber codon suppression to ensure site-specific incorporation, illustrated here for the hypothetical generation of a GYG-monosaccharide conjugate; the corresponding natural glycoform is shown for comparison. Advantages of our chosen strategy (Suzuki-Miyaura coupling) are highlighted, along with reasons for rejecting alternative approaches.

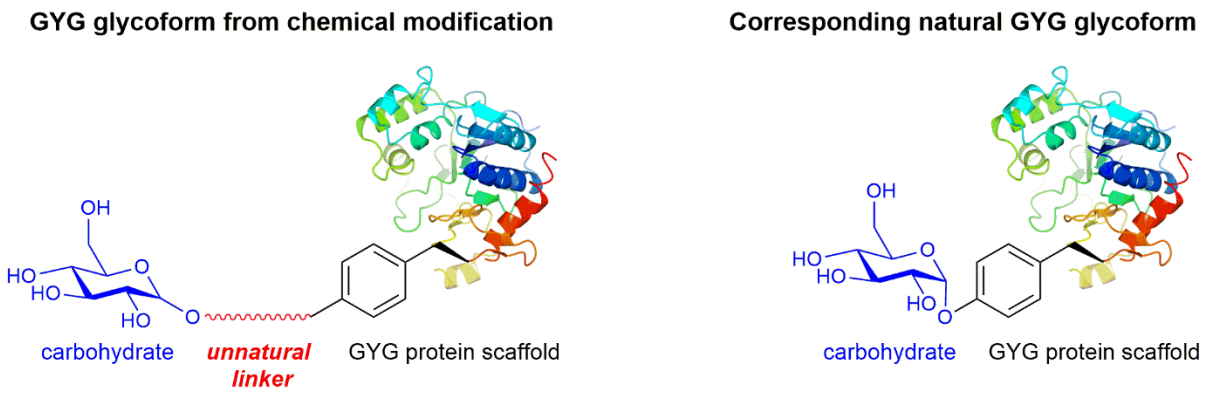


Figure 1.25 Comparison of GYG glycoforms generated through chemical modification to corresponding natural glycoforms, highlighting the presence of an unnatural "linker" in the former between protein and carbohydrate.

In contrast, other methods suffered from at least one notable drawback. The disadvantages of the Staudinger ligation with regards to incompatibility with other residues (notably Cys) and slow kinetics have been previously discussed.[210] Oxime formation from *p*-AcPhe leads to the formation of a protein-carbohydrate linkage that is not entirely irreversible, being to some extent labile towards hydrolysis.[229] Whilst CuAAC-based methods exhibit comparable advantages to SMC in many regards, we felt that the unavoidable formation of a heterocycle moiety (a triazole) in the former as part of the carbohydrate-protein linker (see **figure 1.24**) could lead to a lower degree of structural resemblance between the GYG glycoforms generated via this method and the natural variants thereof. These factors, in combination with the strengths of SMC, led us to select the latter as our method of choice for the chemical modification of GYG.

1.5 Suzuki-Miyaura Cross-coupling: from Synthesis to Biology

1.5.1 Suzuki-Miyaura Cross-coupling: A Powerful Pd-catalysed Reaction

Since their inception, palladium-catalysed cross-coupling reactions have vastly transformed the field of organic synthesis, opening as they do novel retrosynthetic disconnections that were not previously possible. Indeed, they represent a widely-utilised class of reactions within pharmaceutical and agrochemical industries.[230] The original discovery of the Pd-mediated coupling of an aryl or alkenyl halide to an alkene, which came to be known as the Mizorocki-Heck reaction,[231,232] would go on to inspire the subsequent development of a wide range of Pd-catalysed cross-coupling reactions that, broadly speaking, adhere to the fundamental mechanism illustrated below (**figure 1.26**). This involves: 1) oxidative addition of Pd(0) to a

carbon-halogen (or pseudohalogen) bond in the first substrate; 2) transmetalation of the second coupling partner, leading to bond formation between the latter and the Pd catalyst and 3) reductive elimination of the Pd-bonded coupling partners to yield cross-coupled product, in so doing regenerating the active Pd(0) catalytic species.

As illustrated, the initial Pd(0) species is itself frequently generated *in situ* via reduction of a more stable Pd(II) "pre-catalyst".[233] Such a process may be effected through a variety of mechanisms, including reduction by phosphine ligands,[234] by alcoholic solvents,[235] or through homo-coupling of the cross-coupling partner (R_2-M in **figure 1.26**).[236,237] Ligand dissociation steps may also be required to generate the 14-electron $L_2Pd(0)$ species.

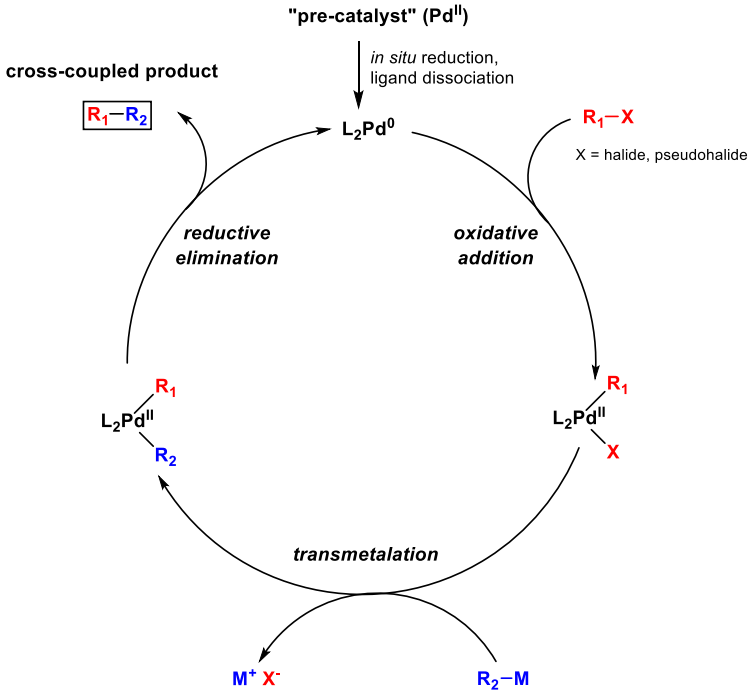


Figure 1.26 General cross-coupling cycle for a wide range of Pd-catalysed C-C bond-forming coupling reactions, including Hiyama (M = SiR_3), Suzuki-Miyaura (M = $B(OR)_2$), Kumada (M = MgX), Negishi (M = ZnX), Stille (M = SnR_3) and Sonogashira (M = Cu) couplings.

Representative examples of cross-coupling reactions include the coupling of organohalides or triflates to organozinc species (Negishi coupling),[238] to organosilane species (Hiyama coupling),[239] to organomagnesium species (Kumada coupling),[240] to terminal alkynes with copper co-catalysis (Sonogashira coupling),[241] to organotin species (Stille coupling),[242] and to organoboron species (Suzuki-Miyaura coupling).[243] The latter, developed through a series of publications commencing in 1979,[243–245] has remained

highly utilised in organic synthesis and medicinal chemistry, whilst also being widely applied in the field of chemical biology, as discussed later.

Whilst the mechanism of the Suzuki-Miyaura cross-coupling is considered to broadly follow the catalytic cycle depicted above,[246,247] some additional details are apparent. In particular, the subsequent transmetalation step requires that the organoboron species is first activated as its more reactive tetra-coordinated boronate form. Two potential pathways can be envisaged; in the "borate" pathway, activation occurs prior to coordination to the Pd(II) species and subsequent transmetalation, whereas in the second "oxo-palladium" pathway, such activation occurs as a direct consequence of coordination by a suitable ligand on the palladium catalyst, again followed by transmetalation (**figure 1.27**).[236,246,248]

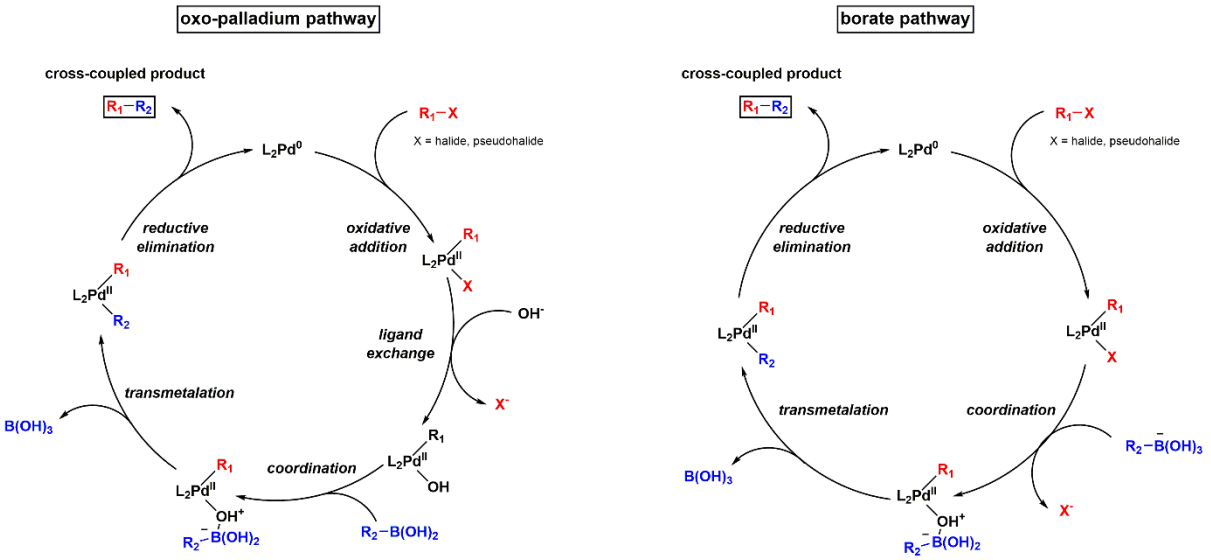


Figure 1.27 Oxo-palladium (left) and borate (right) mechanisms for Suzuki-Miyaura cross-coupling.

Naturally, this cross-coupling reaction is not perfect, and certain side-reactions are sufficiently general and prevalent to be of note. The boronic acid partner may be consumed via cleavage of the C-B bond via oxidation or protodeboronation,[249–251] and through Pd-mediated homocoupling,[236] whilst the halide partner is similarly susceptible to cleavage of the C-X bond via hydrodehalogenation (**figure 1.28 a**).[250,252] The latter reaction is of greater significance in the context of protein modification, since the protein typically contains the aryl halide moiety, and indeed its occurrence has been previously noted in this context.[182,253–255] It is proposed to result from an offshoot of the cross-coupling catalytic cycle in which the R_1-L_2Pd(II)-X complex first undergoes ligand exchange to coordinate a ligand containing a suitable β-hydrogen to enable

subsequent β-hydride elimination; subsequent reductive elimination yields the dehalogenated species R_1-H (**figure 1.28 b**).[252]

Nowadays, a wide variety of boron derivatives can be utilised in SMC.[236] Examples aside from conventional boronic acids[245] include pinacol boronates,[256,257] MIDA boronate esters,[258,259] potassium trifluoroborates,[260] and conveniently pre-activated triol borates[261] (it should be noted that trifluoroborates typically undergo *in situ* hydrolysis to the boronic acid, and the latter is the reactive species).[262] These are frequently bench-stable solids and easy to handle, as well as being relatively straightforward to synthesise; in addition, a number are now commercially available.

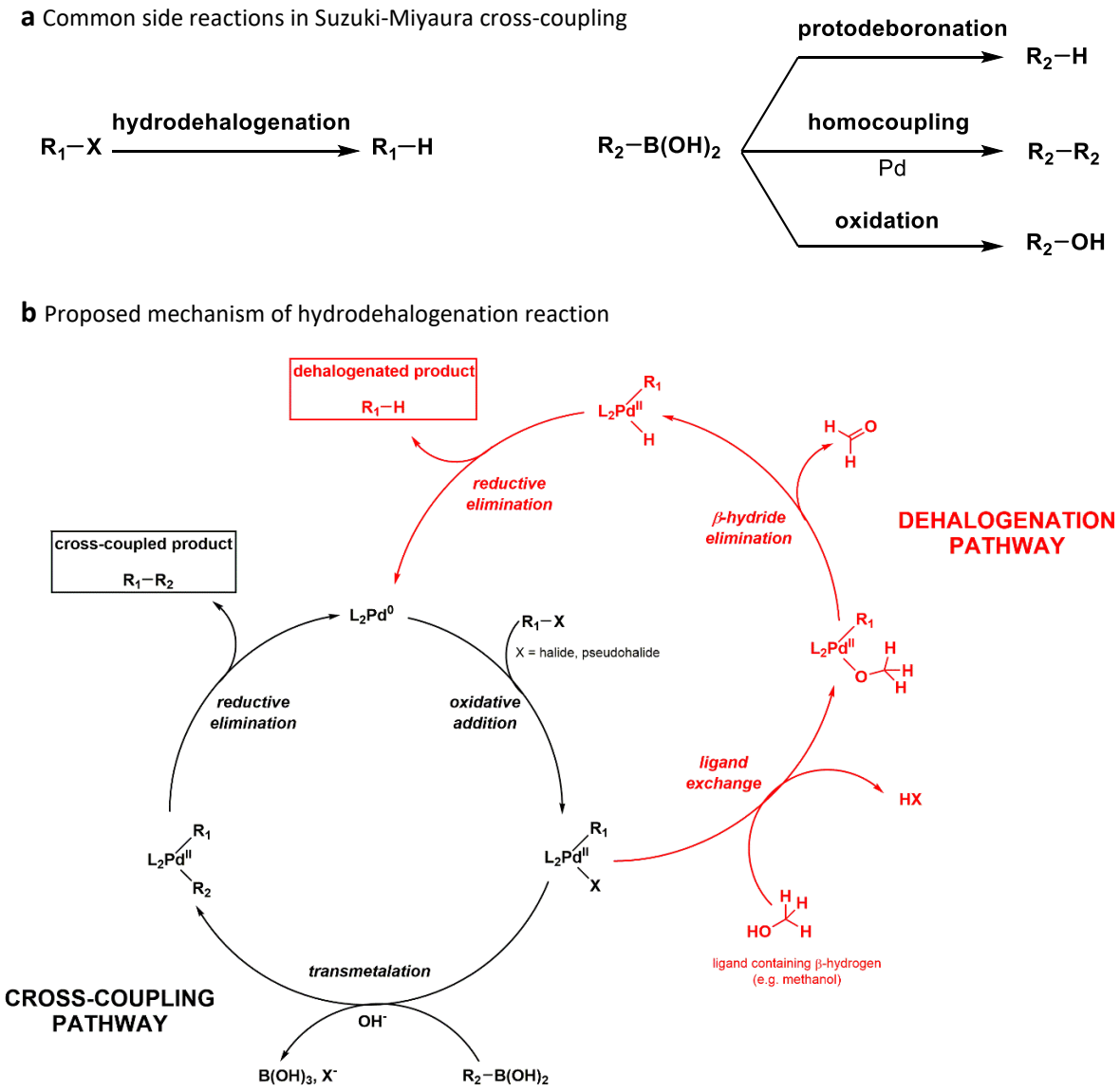


Figure 1.28 Common side-reactions in Suzuki-Miyaura cross-coupling (a), and proposed mechanism of the common hydrodehalogenation side-reaction as an offshoot of the cross-coupling pathway (b).

1.5.2 From Organic Synthesis to Chemical Biology

The transition of SMC from organic synthesis to chemical biology applications necessitated the development of protein-compatible conditions for this reaction. An optimal protein bioconjugation reaction should work at low (physiological) temperatures and close-to-neutral pH, should exhibit rapid reaction kinetics, and – perhaps most crucially – should ideally be feasible in purely aqueous buffers; these are all features desirable to minimise degradation and/or precipitation of the protein, and preserve its native tertiary structure. The latter point is of particular relevance, and research focussed on the development of aqueous-compatible SMC in the context of organic chemistry would pave the way for its subsequent use as a viable bioconjugation method. Indeed, this continues to remain an active area of research in the context of green chemistry.[263]

The development of SMC reactions in water benefitted from the fact that, in contrast to coupling partners for certain other Pd-catalysed couplings (e.g. Negishi, Kumada), boronic acids typically exhibit good aqueous stability and some degree of aqueous solubility.[249] A more significant hurdle instead was the development of catalyst ligands that were themselves water-soluble and water-stable; comprehensive reviews exist on this topic.[263,264] Water-soluble phosphine ligands have been developed, bearing hydrophilic structural moieties including carbohydrates,[265] sulfonates[266] and quaternary ammonium groups.[267] Yet P-donor ligands may be susceptible to decomposition in aqueous media,[263] and, in part as a consequence, N-donor ligands, whilst less electron-donating than phosphine ligands, have seen something of a resurgence in use, along with N-heterocyclic carbene (NHC) donors. Within the former category, a range of ligand types exist, such as pyridine derivatives,[268] pyrimidines,[269] amines,[270] and oxime ethers.[271] NHC ligands represent an intriguing alternative option, and a numer of water-soluble NHC ligands have been reported that are capable of effecting purely aqueous SMC.[272,273] It should also be noted that heterogeneous catalyst systems have also been utilised in the context of aqueous SMC,[274–277] thus circumventing issues of aqueous solubility of catalyst/ligand, although solid-supported catalysts do often suffer from lower activity and selectivity with respect to their homogeneous counterparts.[263]

In 2005, Ojida *et al* reported the first SMC on a protein surface under mild, aqueous conditions.[278] However, a lack of methods for installation of the requisite iodine tag into the protein other than by SPPS limited its application at this stage, and further development of this reaction did not occur for some time after. Subsequently, the group of Schultz demonstrated SMC to a genetically encoded boronate residue,[279] however this required high temperatures incompatible with most proteins. In 2009, Chalker *et al* reported the discovery of a water-soluble and phosphine-free palladium catalyst system, based around a pyrimidine

ligand (**L1, figure 1.29**), that was capable of effecting on-protein SMC to a chemically-installed *p*-iodobenzylcysteine residue to a broad range of boronic acid substrates under benign, biocompatible conditions.[203]

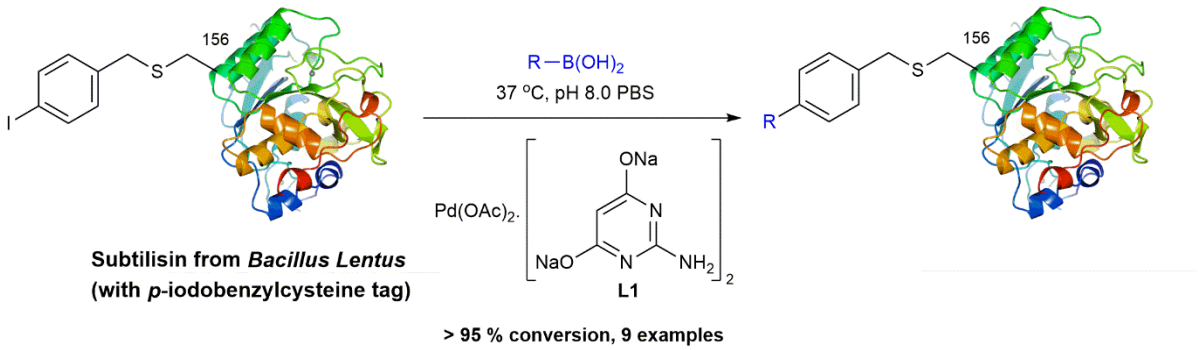


Figure 1.29 Water-soluble, phosphine-free catalyst system for on-protein Suzuki-Miyaura cross-coupling, as reported by Chalker *et al.*[203]

Subsequently, as previously mentioned (section 1.4.3), cross-couplings to *p*-iodophenylalanine (*p*IPhe), which may be genetically encoded at a unique and precise position in a protein substrate, were developed,[223] further increasing the potential of this reaction. Indeed, a range of biological applications of this reaction have since been demonstrated, including protein PEGylation,[254] cell surface fluorescence labelling,[280] and radiolabelling.[253] Whilst protein SMC has yet to be realised in an intracellular context, encouragement in this regard comes from the work of Bradley and co-workers, who have utilised Pd(0) microspheres to successfully effect intracellular SMC of small molecules,[281] as well as the successful application of the Pd-catalysed Sonogashira cross-coupling to intracellular protein labelling.[226] Yet, if its application to *in vivo* reactions remains in the early stages of development, it is evident that Suzuki-Miyaura cross-coupling represents an increasingly well-established method for the covalent, site-selective modification of protein substrates in an *in vitro* setting.

1.6 Aims of this Thesis

This introduction has outlined the crucial biological context of the unusual autocatalytic enzyme Glycogenin (GYG), whilst highlighting the need for new tools with which to better understand its catalytic mechanism. Prior mechanistic studies, carried out predominantly on the GYG-1 isoform, have been hindered, fundamentally, by an inability to access homogeneous glycoforms of the enzyme, restricting detailed analysis of individual glucosylation steps, in particular the earlier ones. Following a survey of existing methodologies

for glycoprotein generation, Suzuki-Miyaura cross-coupling to genetically encoded *p*-iodophenylalanine was alighted upon as an appealing option for the prospective generation of homogeneous GYG glycoconjugates.

Thus, the fundamental goal of this thesis is the application of SMC to the generation of distinct, homogeneous GYG glycoforms (or, as noted previously in section 1.3.5, mimetics thereof), via selective replacement of the native tyrosine glycosyl acceptor (Tyr195 in GYG-1) with *p*-iodophenylalanine and subsequent coupling to suitable boronic acids (**figure 1.30**), and mechanistic study of the enzymes thus formed, should they exhibit activity. The generation of catalytically active GYG species through cross-coupling to GYG-*p*IPhe – itself lacking the native glycosyl acceptor, and thus presumably inactive – would in itself represent a novel, Pd-mediated enzyme activation mode (see section 1.3.5 and chapter 3). We anticipated that an unprecedented ability to isolate multiple different GYG glycoforms would afford a corresponding ability to investigate multiple glucosylation steps, primarily via a kinetics-based approach, enabling a deeper understanding of the catalytic mechanism. In particular, we reasoned that comparative analysis of different individual steps, never previously possible, would facilitate a far clearer understanding of the existence or otherwise of multiple "phases" to GYG catalysis, which have been suggested previously but lack unambiguous, direct definition in terms of, for example, comparative rates or substrate tolerance.

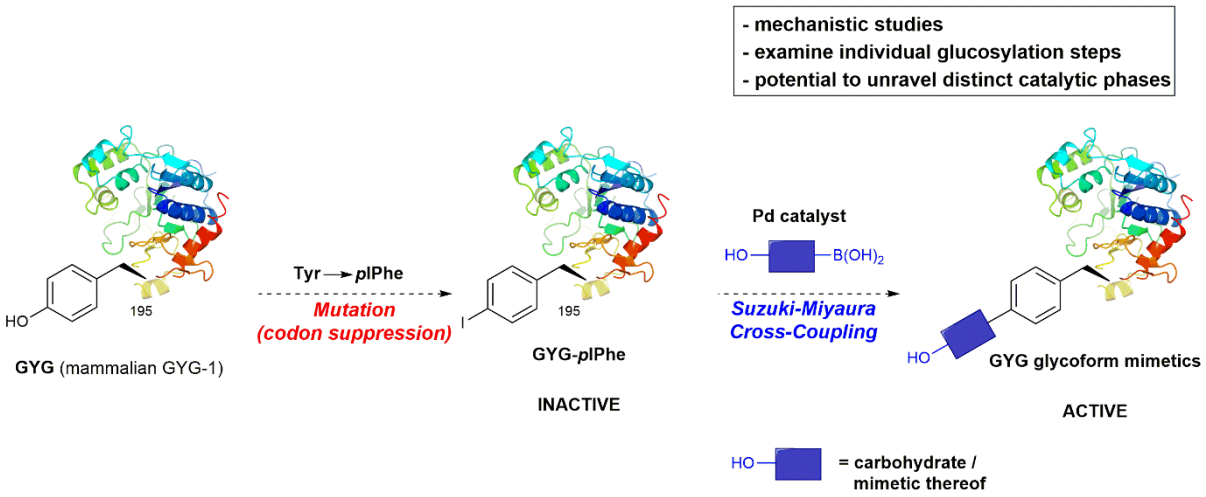


Figure 1.30 Graphical summary of the fundamental goals of this thesis. Installation of a *p*IPhe tag in place of the native tyrosine glycosyl acceptor of GYG (specifically, GYG-1) would yield an inactive, "tagged" GYG-*p*IPhe species. Subsquent cross-coupling to appropriate boronic acids (carbohydrates or mimetics thereof) would yield homogeneous GYG glycoforms (or mimetics thereof), facilitating subsequent mechanistic studies. The Pd-mediated conversion of inactive to active enzyme represents a novel enzyme activation mode.

As noted at various points within this chapter, the enzyme under study in this thesis is the mammalian GYG-1 isoform. However, given aforementioned similarities between GYG-1 and GYG-2, and across

organisms (see section 1.1.3), we anticipate that the proposed methodology could prove equally applicable to such species. Indeed, the investigation of GYG-2 in comparison to GYG-1, whilst unfortunately beyond the scope of this work, represents an exciting future avenue.

This thesis is essentially structured in two distinct parts, although it is not formally divided as such. Chapters 2 and 3 focus on the development of the SMC-based methodology for the generation of GYG-1 glycomimetics. Whilst the former chapter focusses predominantly on the incorporation of *p*IPhe into GYG and optimisation of SMC reactions to this "tagged" enzyme, the latter discusses the successful generation of a homogeneous, catalytically active GYG-monosaccharide conjugate. Subsequently, chapters 4 through 6 focus on mechanistic studies and applications of the methodology thus established. These include an investigation of the earlier catalytic steps through comparative analyses of GYG glycoconjugates bearing 1 or 2 glucose moieties (chapter 4), efforts to selectively investigate the later catalytic steps through the generation of a GYG-oligosaccharide species (chapter 5), and studies on GYG glycoforms bearing either non-natural carbohydrate acceptor structures or activity-perturbing, disease-relevant mutations (chapter 6). Finally, chapter 7 provides a conclusion and discussion of the work within this thesis.

1.7 References

1. Steven, A. C., Baumeister, W., Johnson, L. N. & Perham, R. N. *Molecular Biology of Assemblies and Machines*. (Garland Science, Taylor & Francis Group, LLC, 2016).

2. Alberts, B. *et al. Molecular Biology of the Cell, Sixth Edition.* (Taylor & Francis Group, 2014).

3. Alters, S. *Biology: Understanding Life.* (Jones and Barlette Publishers, 2000).

4. Flatt, J. P. Use and storage of carbohydrate and fat. *Am. J. Clin. Nutr.* **61,** 952S–959S (1995).

5. Voet, D., Voet, J. G. & Pratt, C. W. *Fundamentals of Biochemistry: Life at the Molecular Level, 5th Edition: Life at the Molecular Level.* (Wiley, 2016).

6. Jensen, J., Rustad, P. I., Kolnes, A. J. & Lai, Y-C. The role of skeletal muscle glycogen breakdown for regulation of insulin sensitivity by exercise. *Front. Physiol.* **2,** 112 (2011).

7. Preiss, J. & Walsh, D. A. in *Biology of Carbohydrates* (eds. Ginsburg, V. & Robbins, P.) 199–314 (Wiley and Sons, 1981).

8. Roach, P. J., Depaoli-Roach, A. A., Hurley, T. D. & Tagliabracci, V. S. Glycogen and its metabolism: some new developments and old themes. *Biochem. J.* **441,** 763–787 (2012).

9. Wilson, W. A. *et al.* Regulation of glycogen metabolism in yeast and bacteria. *FEMS Microbiol. Rev.* **34,** 952–985 (2010).

10. Berg, J. M., Tymoczko, J. L. & Stryer, L. *Biochemistry, Fifth Edition.* (W.H. Freeman, 2002).

11. Zois, C. E., Favaro, E. & Harris, A. L. Glycogen metabolism in cancer. *Biochem. Pharmacol.* **92,** 3–11 (2014).

12. Obel, L. F. *et al.* Brain glycogen - new perspectives on its metabolic function and regulation at the subcellular level. *Front. Neuroenergetics* **4,** 3 (2012).

13. Roach, P. J., Skurat, A. V & Harris, R. A. in *Handbook of Physiology* (eds. Jefferson, L. S. & Cherrington, A. D.) 609–647 (Oxford University Press, 2000).

14. Rodriguez, I. R. & Whelan, W. J. A novel glycosyl-amino acid linkage: Rabbit-muscle glycogen is covalently linked to a protein via tyrosine. *Biochem. Biophys. Res. Commun.* **132,** 829–836 (1985).

15. Meléndez-Hevia, E., Waddell, T. G. & Shelton, E. D. Optimization of molecular design in the evolution of metabolism: the glycogen molecule. *Biochem. J.* **295,** 477–483 (1993).

16. Romero, J. M., Issoglio, F. M., Carrizo, M. E. & Curtino, J. A. Evidence for glycogenin autoglucosylation cessation by inaccessibility of the acquired maltosaccharide. *Biochem. Biophys. Res. Commun.* **374,** 704–708 (2008).

17. Gunja-Smith, Z., Marshall, J. J., Mercier, C., Smith, E. E. & Whelan, W. J. A revision of the Meyer-Bernfeld model of glycogen and amylopectin. *FEBS Lett.* **12,** 101–104 (1970).

18. Meléndez, R., Meléndez-Hevia, E. & Cascante, M. How did glycogen structure evolve to satisfy the requirement for rapid mobilization of glucose? A problem of physical constraints in structure building. *J. Mol. Evol.* **45,** 446–455 (1997).

19. Goldsmith, E., Sprang, S. & Fletterick, R. Structure of maltoheptaose by difference Fourier methods and a model for glycogen. *J. Mol. Biol.* **156,** 411–427 (1982).

20. Kirkman, B. R. & Whelan, W. J. Glucosamine is a normal component of liver glycogen. *FEBS* **194,** 6–11 (1986).

21. Fontana, J. D. The presence of phosphate in glycogen. *FEBS Lett.* **109,** 85–92 (1980).

22. Turnbull, J. *et al.* Glycogen Hyperphosphorylation Underlies Lafora Body Formation. *Ann. Neurol.* **68,** 925–933 (2010).

23. Roach, P. J. Glycogen phosphorylation and Lafora disease. *Mol. Aspects Med.* **46,** 78–84 (2015).

24. Tagliabracci, V. S. *et al.* Abnormal Metabolism of Glycogen Phosphate as a Cause for Lafora Disease. *J. Biol. Chem.* **283,** 33816–33825 (2008).

25. Roach, P. J. Are there errors in glycogen biosynthesis and is laforin a repair enzyme? *FEBS Lett.* **585,** 3216–3218 (2011).

26. Tosh, D., Beresford, G. & Agius, L. Glycogen synthesis from glucose by direct and indirect pathways in hepatocyte cultures from different nutritional states. *Biochim. Biophys. Acta* **1224,** 205–212 (1994).

27. Huang, M. & Veech, R. L. Role of the Direct and Indirect Pathways for Glycogen Synthesis in Rat Liver in the Postprandial State. *J. Clin. Invest.* **81,** 872–878 (1988).

28. Kuwajimas, M., Newgardg, C. B., Foster, D. W. & Mcgarry, J. D. The Glucose-phosphorylating Capacity of Liver as Measured by Three Independent Assays. *J. Biol. Chem.* **261,** 8849–8853 (1986).

29. Rodwell, V., Bender, D., Botham, K. M., Kennelly, P. J. & Weil, P. A. *Harpers Illustrated Biochemistry 30th Edition*. (McGraw-Hill Education, 2015).

30. Singh, R. & Cuervo, A. M. Review Autophagy in the Cellular Energetic Balance. *Cell Metab.* **13,** 495–504 (2011).

31. Ballicora, M. A., Iglesias, A. A. & Preiss, J. ADP-Glucose Pyrophosphorylase , a Regulatory Enzyme for Bacterial Glycogen Synthesis. *Microbiol. Mol. Biol. Rev.* **67,** 213–225 (2003).

32. Ugalde, J. E., Parodi, A. J. & Ugalde, R. A. De novo synthesis of bacterial glycogen: agrobacterium tumefaciens glycogen synthase is involved in glucan initiation and elongation. *PNAS* **100,** 10659–10663 (2003).

33. Graham, T. E., Yuan, Z., Hill, A. K. & Wilson, R. J. The regulation of muscle glycogen: the granule and its proteins. *Acta Physiol.* **199,** 489–498 (2010).

34. Akatsuka, A., Singh, T. J., Nakabayashi, H., Lin, M. C. & Huang, K-P. Phosphorylation of Rat Liver Glycogen Synthase in Isolated Hepatocytes. *J. Biol. Chem.* **260,** 3239–3242 (1985).

35. Johnson, L. N. Glycogen phosphorylase: control by phosphorylation and allosteric. *FASEB J.* **6,** 2274–2282 (1992).

36. Buchbinder, J. L. & Fletterick, R. J. Role of the Active Site Gate of Glycogen Phosphorylase in Allosteric Inhibition and Substrate Binding. *J. Biol. Chem.* **271,** 22305–22309 (1996).

37. Vernier, J. M. & Sire, M. F. Etude 'in Vitro' de la Glycogène Phosphorylase Hépatique de la Truite Arc-en-ciel: son Côntrole par le Glucose, les Corticoïdes, l'Adrénaline, et le Glucagon. *Gen. Comp. Endocrinol.* **34,** 360–369 (1978).

38. Leloir, L. F., Olavarria, J. M., Goldemberg, S. H. & Carminatti, H. Biosynthesis of Glycogen from Uridine Diphosphate Glucose '. *Arch. Biochem. Biophys.* **81,** 508–520 (1959).

39. Picton, C., Klee, C. B. & Cohen, P. The regulation of muscle phosphorylase kinase by calcium ions, calmodulin and troponin-C. *Cell Calcium* **2,** 281–294 (1981).

40. Walsh, K. X., Millikin, D. M., Schlenders, K. K. & Reimann, E. M. Calcium-dependent Phosphorylation of Glycogen Synthase by Phosphorylase Kinase. *J. Biol. Chem.* **254,** 6611–6616 (1979).

41. Oron, Y., Emerson, J. F., Garrison, J. C. & Larner, J. Calcium control of glycogen synthase activities in mouse diaphragms, rat adipocytes and rat hepatocytes. *Cell Calcium* **5,** 143–158 (1984).

42. Villar-Palasi, C. On the mechanism of inactivation of muscle glycogen phosphorylase by insulin. *Biochim. Biophys. Acta* **1224,** 384–388 (1994).

43. Cohen, P. The Croonian Lecture 1998. Identification of a protein kinase cascade of major importance in insulin signal transduction. *Phil. Trans. R. Soc. Lond.* **354,** 485–495 (1999).

44. Huang, S. & Czech, M. P. The GLUT4 Glucose Transporter. *Cell Metab.* **5,** 237–252 (2007).

45. Gloster, T. M. Advances in understanding glycosyltransferases from a structural perspective. *Curr. Opin. Struct. Biol.* **28,** 131–141 (2014).

46. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42,** D490-D495 (2014).

47. Breton, C., Šnajdrová, L., Jeanneau, C., Koča, J. & Imberty, A. Structures and mechanisms of glycosyltransferases. *Glycobiology* **16,** 29–37 (2006).

48. Smythe, C. & Cohen, P. The discovery of glycogenin and the priming mechanism for glycogen biogenesis. *Eur. J. Biochem.* **200,** 625–631 (1991).

49. Cao, Y., Mahrenholz, A. M., DePaoli-Roach, A. A. & Roach, P. J. Characterization of rabbit skeletal muscle glycogenin: Tyrosine 194 is essential for function. *J. Biol. Chem.* **268,** 14687–14693 (1993).

50. Alonso, M. D., Lomako, J., Lomako, W. M. & Whelan, W. J. A new look at the biogenesis of glycogen. *FASEB J.* **9,** 1126–1137 (1995).

51. Lomako, J., Lomako, W. M. & Whelan, W. J. A self-glucosylating rabbit protein is the primer for rabbit muscle glycogen biosynthesis. *FASEB J.* **2,** 3097–3103 (1988).

52. Roach, P. J. & Skurat, A. V. Self-glucosylating initiator proteins and their role in glycogen biosynthesis. *Prog. Nucleic Acid Res. Mol. Biol.* **57,** 289–316 (1997).

53. Lomako, J., Lomako, W. M. & Whelan, W. J. Glycogenin: The primer for mammalian and yeast glycogen synthesis. *Biochim. Biophys. Acta - Gen. Subj.* **1673,** 45–55 (2004).

54. Alonso, M. D., Lomako, J., Lomako, W. M. & Whelan, W. J. Tyrosine-194 of glycogenin undergoes autocatalytic glucosylation but is not essential for catalytic function and activity. *FEBS Lett.* **342,** 38–42 (1994).

55. Gibbons, B. J., Roach, P. J. & Hurley, T. D. Crystal structure of the autocatalytic initiator of glycogen biosynthesis, glycogenin. *J. Mol. Biol.* **319,** 463–477 (2002).

56. Mu, J., Skurat, A. V. & Roach, P. J. Glycogenin-2, a novel self-glucosylating protein involved in liver glycogen biosynthesis. *J. Biol. Chem.* **272,** 27589–27597 (1997).

57. Nilsson, J. *et al.* LC-MS/MS characterization of combined glycogenin-1 and glycogenin-2 enzymatic activities reveals their self-glucosylation preferences. *Biochim. Biophys. Acta* **1844,** 398–405 (2014).

58. Mu, J., Roach, P. J. & Chem, J. B. Characterization of Human Glycogenin-2 , a Self-glucosylating Initiator of Liver Glycogen Metabolism. *Biochemistry* **273,** 34850–34856 (1998).

59. Chaikuad, A. *et al.* Conformational plasticity of glycogenin and its maltosaccharide substrate during glycogen biogenesis. *Proc. Natl. Acad. Sci. USA.* **108,** 21028–21033 (2011).

60. Cheng, C. Initiation of Glycogen Synthesis in Yeast. *J. Biol. Chem.* **271,** 26554–26560 (1996).

61. Albrecht, T. *et al.* Yeast glycogenin (Glg2p) produced in Escherichia coli is simultaneously glucosylated at two vicinal tyrosine residues but results in a reduced bacterial glycogen accumulation. *Eur. J. Biochem.* **271,** 3978–3989 (2004).

62. Mu, J., Skurat, A. V & Roach, P. J. Glycogenin-2, a Novel Self-glucosylating Protein Involved in Liver Glycogen Biosynthesis. *J. Biol. Chem.* **272,** 27589–27597 (1997).

63. Skurat, A. V, Dietrich, A. D., Zhai, L. & Roach, P. J. GNIP, a novel protein that binds and activates glycogenin, the self-glucosylating initiator of glycogen biosynthesis. *J. Biol. Chem.* **277,** 19331–19338 (2002).

64. Zeqiraj, E. *et al.* Structural basis for the recruitment of glycogen synthase by glycogenin. *Proc. Natl. Acad. Sci. USA.* **111,** E2831-E2840 (2014).

65. Gomis, R. R., Cid, E., García-Rocha, M., Ferrer, J. C. & Guinovart, J. J. Liver glycogen synthase but not the muscle isoform differentiates between glucose-6-phosphate produced by glucokinase or hexokinase. *J. Biol. Chem.* **277,** 23246–23252 (2002).

66. Baskaran, S., Roach, P. J., DePaoli-Roach, A. A. & Hurley, T. D. Structural basis for glucose-6-phosphate activation of glycogen synthase. *Proc. Natl. Acad. Sci. USA* **107,** 17563–17568 (2010).

67. Froese, D. S. *et al.* Structural basis of glycogen branching enzyme deficiency and pharmacological rescue by rational peptide design. *Hum. Mol. Genet.* **24,** 5667–5676 (2015).

68. Wolfsdorf, J. I. & Weinstein, D. A. Glycogen Storage Diseases. *Rev. Endocr. Metab. Disord.* **4,** 95–102 (2003).

69. Lairson, L. L., Henrissat, B., Davies, G. J. & Withers, S. G. Glycosyltransferases: Structures, Functions, and Mechanisms. *Annu. Rev. Biochem.* **77,** 521–555 (2008).

70. Liu, J. & Mushegian, A. Three monophyletic superfamilies account for the majority of the known glycosyltransferases. *Protein Sci.* **12,** 1418–1431 (2003).

71. Charnock, S. J. & Davies, G. J. Structure of the nucleotide-diphospho-sugar transferase, SpsA from Bacillus subtilis, in native and nucleotide-complexed forms. *Biochemistry* **38,** 6380–6385 (1999).

72. Vrielink, A., Rüger, W., Driessen, H. P. & Freemont, P. S. Crystal structure of the DNA modifying enzyme beta-glucosyltransferase in the presence and absence of the substrate uridine diphosphoglucose. *EMBO J.* **13,** 3413–3422 (1994).

73. Rosén, M. L., Edman, M., Sjöström, M. & Wieslander, A. Recognition of fold and sugar linkage for glycosyltransferases by multivariate sequence analysis. *J. Biol. Chem.* **279,** 38683–38692 (2004).

74. Liang, D-M. *et al.* Glycosyltransferases: mechanisms and applications in natural product development. *Chem. Soc. Rev.* **44,** 8350–8374 (2015).

75. Kozmon, S. & Tvaroška, I. Catalytic Mechanism of Glycosyltransferases: Hybrid Quantum Mechanical/Molecular Mechanical Study of the Inverting N-Acetylglucosaminyltransferase I. *J. Am. Chem. Soc.* **128,** 16921–16927 (2006).

76. Soya, N., Fang, Y., Palcic, M. M. & Klassen, J. S. Trapping and characterization of covalent intermediates of mutant retaining glycosyltransferases. *Glycobiology* **21,** 547–552 (2011).

77. Schuman, B., Evans, S. V & Fyles, T. M. Geometric attributes of retaining glycosyltransferase enzymes favor an orthogonal mechanism. *PLoS One* **8,** e71077 (2013).

78. Flint, J. *et al.* Structural dissection and high-throughput screening of mannosylglycerate synthase. *Nat. Struct. Mol. Biol.* **12,** 608–614 (2005).

79. Lee, S. S. *et al.* Mechanistic evidence for a front-side, SNi-type reaction in a retaining glycosyltransferase. *Nat. Chem. Biol.* **7,** 631–638 (2011).

80. Schmid, J., Heider, D., Wendel, N. J., Sperl, N. & Sieber, V. Bacterial glycosyltransferases: Challenges and Opportunities of a Highly Diverse Enzyme Class Toward Tailoring Natural Products. *Front. Microbiol.* **7,** 1–7 (2016).

81. Wiggins, C. A. R. & Munro, S. Activity of the yeast MNN1 α-1,3-mannosyltransferase requires a motif conserved in many other families of glycosyltransferases. *Proc. Natl. Acad. Sci. USA* **95,** 7945–7950 (1998).

82. Hurley, T. D., Stout, S., Miner, E., Zhou, J. & Roach, P. J. Requirements for catalysis in mammalian glycogenin. *J. Biol. Chem.* **280,** 23892–23899 (2005).

83. Bazán, S., Issoglio, F. M., Carrizo, M. E. & Curtino, J. A. The intramolecular autoglucosylation of monomeric glycogenin. *Biochem. Biophys. Res. Commun.* **371,** 328–332 (2008).

84. Issoglio, F. M., Carrizo, M. E., Romero, J. M. & Curtino, J. A. Mechanisms of monomeric and dimeric glycogenin autoglucosylation. *J. Biol. Chem.* **287,** 1955–1961 (2012).

85. Lin, A., Mu, J., Yang, J. & Roach, P. J. Self-glucosylation of glycogenin, the initiator of glycogen biosynthesis, involves an inter-subunit reaction. *Arch. Biochem. Biophys.* **363,** 163–170 (1999).

86. Wald, F. A., Kissen, R., Du Jardin, P. & Moreno, S. Characterization of UDP-glucose:protein transglucosylase genes from potato. *Plant Mol. Biol.* **52,** 705–714 (2003).

87. Selth, L. A., Dogra, S. C., Rasheed, M. S., Randles, J. W. & Rezaian, M. A. Identification and characterization of a host reversibly glycosylated peptide that interacts with the tomato leaf curl virus V1 protein. *Plant Mol. Biol.* **61,** 297–310 (2006).

88. Carrizo, M. E., Romero, J. M., Issoglio, F. M. & Curtino, J. A. Structural and biochemical insight into glycogenin inactivation by the glycogenosis-causing T82M mutation. *FEBS Lett.* **586,** 254–257 (2012).

89. Hurley, T. D., Walls, C., Bennett, J. R., Roach, P. J. & Wang, M. Direct detection of glycogenin reaction products during glycogen initiation. *Biochem Biophys Res Commun.* **348,** 374–378 (2006).

90. Okamoto, R., Izumi, M. & Kajihara, Y. Decoration of proteins with sugar chains: recent advances in glycoprotein synthesis. *Curr. Opin. Chem. Biol.* **22,** 92–99 (2014).

91. Rademacher, T. W., Parekh, R. B. & Dwek, R. A. Glycobiology. *Annu. Rev. Biochem.* **57,** 785–838 (1988).

92. Gamblin, D. P., Scanlan, E. M., Davis, B. G. Glycoprotein Synthesis : An Update. *Chem. Rev.* **109,** 131–163 (2009).

93. Taylor, M. E. & Drickamer, K. *Introduction to Glycobiology*. (OUP Oxford, 2011).

94. Lowe, J. B. & Marth, J. D. A Genetic Approach to Mammalian Glycan Function. *Annu. Rev. Biochem.* **72,** 643–691 (2003).

95. Schmaltz, R. M., Hanson, S. R. & Wong, C. H. Enzymes in the synthesis of glycoconjugates. *Chem. Rev.* **111,** 4259–4307 (2011).

96. Gu, J. *et al.* Potential roles of *N*-glycosylation in cell adhesion. *Glycoconj. J.* **29,** 599–607 (2012).

97. Bektas, M. & Rubenstein, D. S. The role of intracellular protein *O*-glycosylation in cell adhesion and disease. *J. Biomed. Res.* **25,** 227–236 (2011).

98. Boscher, C., Dennis, J. W. & Nabi, I. R. Glycosylation, galectins and cellular signaling. *Curr. Opin. Cell Biol.* **23,** 383–392 (2011).

99. Helenius, A. Intracellular Functions of *N*-Linked Glycans. *Science.* **291,** 2364–2369 (2001).

100. Rudd, P. M. Glycosylation and the Immune System. *Science.* **291,** 2370–2376 (2001).

101. Tan, F. Y. Y., Tang, C. M. & Exley, R. M. Sugar coating: Bacterial protein glycosylation and host-microbe interactions. *Trends Biochem. Sci.* **40,** 342–350 (2015).

102. Petrescu, A. J., Wormald, M. R. & Dwek, R. A. Structural aspects of glycomes with a focus on *N*-glycosylation and glycoprotein folding. *Curr. Opin. Struct. Biol.* **16,** 600–607 (2006).

103. Lewis, Y. E. *et al.* O-GlcNAcylation of α-synuclein at serine 87 reduces aggregation without affecting membrane binding. *ACS Chem. Biol.* **12,** 1020–1027 (2017).

104. Kajihara, Y. Studies on the Precise Chemical Synthesis of Human Glycoproteins. *Bull. Chem. Soc. Jpn.* **89,** 409–423 (2016).

105. Winans, K. A., King, D. S., Rao, V. R. & Bertozzi, C. R. A chemically synthesized version of the insect antibacterial glycopeptide, diptericin, disrupts bacterial membrane integrity. *Biochemistry* **38,** 11700–11710 (1999).

106. Wang, P. *et al.* Erythropoietin Derived by Chemical Synthesis. *Science.* **342,** (2013).

107. Borgia, J. A. & Fields, G. B. Chemical synthesis of proteins. *Annu Rev Biophys Biomol Struct.* **34,** 91–118 (2005).

108. Wilken, J. & Kent, S. B. H. Chemical protein synthesis. *Curr. Opin. Biotechnol.* **9,** 412–426 (1998).

109. Muir, T. W. Semisynthesis of Proteins by Expressed Protein Ligation. *Annu. Rev. Biochem.* **72,** 249–289 (2003).

110. Liu, L. *et al.* Advances in glycoprotein synthesis. *Chem. Commun.* **126,** 21–33 (2006).

111. Dawson, P., Muir, T., Clark-Lewis, I. & Kent, S. Synthesis of proteins by native chemical ligation. *Science.* **266,** 776–779 (1994).

112. Kent, S. Origin of the chemical ligation concept for the total synthesis of enzymes (proteins). *Biopolymers* **94,** iv–ix (2010).

113. Malins, L. R. & Payne, R. J. Recent extensions to native chemical ligation for the chemical synthesis of peptides and proteins. *Curr. Opin. Chem. Biol.* **22,** 70–78 (2014).

114. Zhang, Y., Xu, C., Lam, H. Y., Lee, C. L. & Li, X. Protein chemical synthesis by serine and threonine ligation. *PNAS* **110,** 6657–6662 (2013).

115. Shao, J. & Tam, J. P. Unprotected Peptides as Building Blocks for the Synthesis of Peptide Dendrimers with Oxime, Hydrazone, and Thiazolidine Linkages. *JACS* **117,** 3893–3899 (1995).

116. Bode, J. W., Fox, R. M. & Baucom, K. D. Chemoselective amide ligations by decarboxylative condensations of N-alkylhydroxylamines and ??-ketoacids. *Angew. Chemie. Int. Ed.* **45,** 1248–1252 (2006).

117. Chen, J., Wan, Q., Yuan, Y., Zhu, J. & Danishefsky, S. J. Native Chemical Ligation at Valine: A Contribution to Peptide and Glycopeptide Synthesis. *Angew. Chem. Int. Ed.* **47,** 8521–8524 (2008).

118. Rohde, H. & Seitz, O. Ligation-desulfurization: a powerful combination in the synthesis of peptides and glycopeptides. *Biopolymers* **94,** 551–559 (2010).

119. Thapa, P., Zhang, R. Y., Menon, V. & Bingham, J. P. Native chemical ligation: a boon to peptide chemistry. *Molecules* **19,** 14461–14483 (2014).

120. Li, H. *et al.* Chemoenzymatic synthesis of HIV-1 V3 glycopeptides carrying two *N*-glycans and effects of glycosylation on the peptide domain. *J. Org. Chem.* **70,** 9990–9996 (2005).

121. Cohen-Anisfeld, S. T. & Lansbury, P. T. A practical, convergent method for glycopeptide synthesis. *J. Am. Chem. Soc.* **115,** 10531–10537 (1993).

122. Dudkin, V. Y., Miller, J. S. & Danishefsky, S. J. Chemical Synthesis of Normal and Transformed PSA Glycopeptides. *J. Am. Chem. Soc.* **126,** 736–738 (2004).

123. Tachibana, Y. *et al.* Efficient and versatile synthesis of mucin-like glycoprotein mimics. *Tetrahedron* **58,** 10213–10224 (2002).

124. Doores, K. J. *et al.* Direct deprotected glycosyl–asparagine ligation. *Chem. Commun.* **167,** 1401 (2006).

125. Deras, I. L., Takegawa, K., Kondo, A., Kato, I. & Lee, Y. C. Synthesis of a high-mannose-type glycopeptide analog containing a glucose-asparagine linkage. *Bioorganic Med. Chem. Lett.* **8,** 1763–1766 (1998).

126. Tanaka, H., Iwata, Y., Takahashi, D., Adachi, M. & Takahashi, T. Efficient stereoselective synthesis of γ-N-glycosyl asparagines by N-glycosylation of primary amide groups. *J. Am. Chem. Soc.* **127,** 1630–1631 (2005).

127. Ibatullin, F. M. & Selivanov, S. I. Reaction of N-Fmoc aspartic anhydride with glycosylamines: a simple entry to N-glycosyl asparagines. *Tetrahedron Lett.* **50,** 6351–6354 (2009).

128. Mukhopadhyay, B. *et al.* From Solution Phase to 'On-Column' Chemistry : Trichloroacetimidate-based Glycosylation Promoted by Perchloric Acid - Silica. *J. Org. Chem* **70,** 9059–9062 (2005).

129. Cato, D., Buskas, T. & Boons, G. Highly Efficient Stereospecific Preparation of Tn and TF Building Blocks Using Thioglycosyl Donors and the Ph_2SO/Tf_2O Promotor System. *J. Carbohydr. Chem.* **24,** 503–516 (2005).

130. Jansson, A. M. *et al.* Solid-phase glycopeptide synthesis of tyrosine-glycosylated glycogenin fragments as substrates for glucosylation by glycogenin. *J. Chem. Soc. Perkin Trans. 1* (1996).

131. Xu, W. *et al.* Total chemical synthesis of a thermostable enzyme capable of polymerase chain reaction. *Cell Discov.* **3,** 17008 (2017).

132. Bondalapati, S., Jbara, M. & Brik, A. Expanding the chemical toolbox for the synthesis of large and uniquely modified proteins. *Nat. Chem.* **8,** 407–418 (2016).

133. Hemantha, H. P. *et al.* Nonenzymatic polyubiquitination of expressed proteins. *J. Am. Chem. Soc.* **136,** 2665–2673 (2014).

134. Wright, T. H. *et al.* Posttranslational mutagenesis: A chemical strategy for exploring protein side-chain diversity. *Science.* **354,** aag1465 (2016).

135. Liu, L. *Protein Ligation and Total Synthesis II*. (Springer International Publishing, 2015).

136. D'Andrea, L. D. & Romanelli, A. *Chemical Ligation: Tools for Biomolecule Synthesis and Modification*. (Wiley, 2017).

137. Wang, L. X. & Joseph, L. Emerging Technologies for Making Glycan-Defined Glycoproteins. *ACS Chem. Biol.* **7,** 110–122 (2012).

138. Fernández-González, M. *et al.* Site-selective chemoenzymatic construction of synthetic glycoproteins using endoglycosidases. *Chem. Sci.* **1,** 709-715 (2010).

139. Matsushita, T. *et al.* Construction of highly glycosylated mucin-type glycopeptides based on microwave-assisted solid-phase syntheses and enzymatic modifications. *J. Org. Chem.* **71,** 3051–3063 (2006).

140. Wei, Y. *et al.* Glycoengineering of human IgG1-Fc through combined yeast expression and in vitro chemoenzymatic glycosylation. *Biochemistry* **47,** 10294–10304 (2008).

141. Lomino, J. V. *et al.* A two-step enzymatic glycosylation of polypeptides with complex N-glycans. *Bioorganic Med. Chem.* **21,** 2262–2270 (2013).

142. Sørensen, A. L. *et al.* Chemoenzymatically synthesized multimeric Tn/STn MUC1 glycopeptides elicit cancer-specific anti-MUC1 antibody responses and override tolerance. *Glycobiology* **16,** 96–107 (2006).

143. Liu, F. *et al.* Rationally Designed Short Polyisoprenol-linked PglB Substrates for Engineered Polypeptide and Protein N-Glycosylation. *J. Am. Chem. Soc.* **136,** 566–569 (2014).

144. Rich, J. R. & Withers, S. G. Emerging methods for the production of homogeneous human glycoproteins. *Nat. Chem. Biol.* **5,** 206–215 (2009).

145. Wang, L. X. & Lomino, J. V. Emerging technologies for making glycan-defined glycoproteins. *ACS Chem. Biol.* **7,** 110–122 (2012).

146. Goodfellow, J. J. *et al.* An endoglycosidase with alternative glycan specificity allows broadened glycoprotein remodelling. *J. Am. Chem. Soc.* **134,** 8030–8033 (2012).

147. Li, B., Song, H., Hauser, S. & Wang, L. X. A highly efficient chemoenzymatic approach toward glycoprotein synthesis. *Org. Lett.* **8,** 3081–3084 (2006).

148. Huang, W., Giddens, J., Fan, S., Toonstra, C. & Wang, L. Chemoenzymatic Glycoengineering of Intact IgG Antibodies for Gain of Functions. *J. AM. CHEM. SOC.* **134,** 12308–12318 (2012).

149. Parsons, T. B. *et al.* Optimal synthetic glycosylation of a therapeutic antibody. *Angew. Chemie Int. Ed.* **55,** 2361–2367 (2016).

150. Umekawa, M. *et al.* Mutants of Mucor hiemalis endo-β-*N*-acetylglucosaminidase show enhanced transglycosylation and glycosynthase-like activities. *J. Biol. Chem.* **283,** 4469–4479 (2008).

151. Yamaguchi, T., Amin, M. N., Toonstra, C. & Wang, L.-X. Chemoenzymatic Synthesis and Receptor Binding of Mannose-6-Phosphate (M6P)-Containing Glycoprotein Ligands Reveal Unusual Structural Requirements for M6P Receptor Recognition. *J. Am. Chem. Soc.* **138,** 12472–12485 (2016).

152. Lomako, J., Lomako, W. M. & Whelan, W. J. The substrate specificity of isoamylase and the preparation of apo-glycogenin. *Carbohydr. Res.* **227,** 331–338 (1992).

153. Wang, L. X. & Amin, M. N. Chemical and chemoenzymatic synthesis of glycoproteins for deciphering functions. *Chem. Biol.* **21,** 51–66 (2014).

154. Fahmi, N. E., Dedkova, L., Wang, B., Golovine, S. & Hecht, S. M. Site-Specific Incorporation of Glycosylated Serine and Tyrosine Derivatives into Proteins. *JACS* **129,** 3586–3597 (2007).

155. Matsubara, T., Iijima, K., Watanabe, T., Hohsaka, T. & Sato, T. Incorporation of glycosylated amino acid into protein by an *in vitro* translation system. *Bioorganic & Medicinal Chemistry Letters* **23,** (2013).

156. Alff-Steinberger, C. & Epstein, R. Codon Preference in the Terminal Region of E. coli Genes and Evolution of Stop Codon Usage. *Journal of Theoretical Biology* **168,** 461–463 (1994).

157. Davis, L. & Chin, J. W. Designer proteins: applications of genetic code expansion in cell biology. *Nat. Rev. Mol. Cell Biol.* **13,** 168–182 (2012).

158. Wang, L. & Schultz, P. G. Expanding the genetic code. *Angew. Chem. Int. Ed. Engl.* **44,** 34–66 (2004).

159. Liu, C. C. & Schultz, P. G. Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* **79,** 413–444 (2010).

160. Wang, L., Brock, A., Herberich, B. & Schultz, P. G. Expanding the genetic code of Escherichia coli. *Science* **292,** 498–500 (2001).

161. Zhang, Z., Gildersleeve, J., Yang, Y. Y. & Xu, R. A new strategy for the synthesis of glycoproteins. *Science.* **303,** 371–373 (2004).

162. Hemberg, G. Investigating the Autocatalytic Priming Mechanism in Glycogenin. (University of Oxford, MSci Thesis, 2013).

163. Reeves, P. J., Callewaert, N., Contreras, R. & Khorana, H. G. Structure and function in rhodopsin: High-level expression of rhodopsin with restricted and homogeneous N-glycosylation by a tetracycline-inducible N-acetylglucosaminyltransferase I-negative HEK293S stable mammalian cell line. *Proc. Natl. Acad. Sci. USA* **99,** 13419–13424 (2002).

164. Patnaik, S. K. & Stanley, P. Lectin-Resistant CHO Glycosylation Mutants. *Methods Enzymol.* **416,** 159–182 (2006).

165. Bowden, T. A. *et al.* Chemical and structural analysis of an antibody folding intermediate trapped during glycan biosynthesis. *J. Am. Chem. Soc.* **134,** 17554–17563 (2012).

166. Parsons, J. *et al.* Moss-based production of asialo-erythropoietin devoid of Lewis A and other plant-typical carbohydrate determinants. *Plant Biotechnol. J.* **10,** 851–861 (2012).

167. Hamilton, S. R. *et al.* Production of complex human glycoproteins in yeast. *Science* **301,** 1244–1246 (2003).

168. Jacobs, P. P., Geysens, S., Vervecken, W., Contreras, R. & Callewaert, N. Engineering complex-type N-glycosylation in Pichia pastoris using GlycoSwitch technology. *Nat. Protoc.* **4,** 58–70 (2008).

169. Bobrowicz, P. *et al.* Engineering of an artificial glycosylation pathway blocked in core oligosaccharide assembly in the yeast Pichia pastoris: Production of complex humanized glycoproteins with terminal galactose. *Glycobiology* **14,** 757–766 (2004).

170. Li, H. *et al.* Optimization of humanized IgGs in glycoengineered Pichia pastoris. *Nat. Biotechnol.* **24,** 210–215 (2006).

171. Chalker, J. M., Bernardes, G. J. L. & Davis, B. G. A 'Tag-and-Modify' Approach to Site-Selective Protein Modification. *Acc. Chem. Res.* **44,** 730–741 (2011).

172. Chalker, J. M., Bernardes, G. J. L., Lin, Y. A. & Davis, B. G. Chemical modification of proteins at cysteine: opportunities in chemistry and biology. *Chem. Asian J.* **4,** 630–640 (2009).

173. Budisa, N. *et al.* Toward the experimental codon reassignment in vivo: protein building with an expanded amino acid repertoire. *FASEB J.* **13,** 41–51 (1999).

174. Van Hest, J. C. M., Kiick, K. L. & Tirrell, D. A. Efficient incorporation of unsaturated methionine analogues into proteins in vivo. *J. Am. Chem. Soc.* **122,** 1282–1288 (2000).

175. Deiters, A., Groff, D., Ryu, Y., Xie, J. & Schultz, P. G. A genetically encoded photocaged tyrosine. *Angew. Chemie. Int. Ed.* **45,** 2728–2731 (2006).

176. Lemke, E. A., Summerer, D., Geierstanger, B. H., Brittain, S. M. & Schultz, P. G. Control of protein phosphorylation with a genetically encoded photocaged amino acid. *Nat. Chem. Biol.* **3,** 769–772 (2007).

177. Arbely, E., Torres-Kolbus, J., Deiters, A. & Chin, J. W. Photocontrol of tyrosine phosphorylation in mammalian cells via genetic encoding of photocaged tyrosine. *J. Am. Chem. Soc.* **134,** 11912–11915 (2012).

178. Völker, T. & Meggers, E. Transition-metal-mediated uncaging in living human cells - an emerging alternative to photolabile protecting groups. *Curr. Opin. Chem. Biol.* **25,** 48–54 (2015).

179. van Kasteren, S. I. *et al.* Expanding the diversity of chemical protein modification allows post-translational mimicry. *Nature* **446,** 1105–1109 (2007).

180. Raj, R., Lercher, L., Mohammed, S. & Davis, B. G. Synthetic Nucleosomes Reveal that GlcNAcylation Modulates Direct Interaction with the FACT Complex. *Angew. Chemie. Int. Ed.* **55,** 8918–8922 (2016).

181. Parthasarathy, R., Rabuka, D., Bertozzi, C. R. & Groves, J. T. Molecular orientation of membrane-anchored mucin glycoprotein mimics. *J. Phys. Chem. B* **111,** 12133–12135 (2007).

182. Spicer, C. D. Suzuki Biology: Palladium Mediated Reactions under Ambient Conditions. (University of Oxford, DPhil Thesis, 2014).

183. van Kasteren, S. I., Kramer, H. B., Gamblin, D. P. & Davis, B. G. Site-selective glycosylation of proteins: creating synthetic glycoproteins. *Nat. Protoc.* **2,** 3185–3194 (2007).

184. Adamo, R. *et al.* Synthetically defined glycoprotein vaccines: current status and future directions. *Chem. Sci.* **4,** 2995 (2013).

185. van Kasteren, S. I., Garnier, P. & Davis, B. G. in *Protein Engineering* (eds. Köhrer, C. & RajBhandary, U. L.) 97–125 (Spinger-Verlag, 2009).

186. Miseta, A. & Csutora, P. Relationship between the occurrence of cysteine in proteins and the complexity of organisms. *Mol. Biol. Evol.* **17,** 1232–1239 (2000).

187. Wong, S. Y., Guile, G. R., Dwek, R. A. & Arsequell, G. Synthetic glycosylation of proteins using N-(beta-saccharide) iodoacetamides: applications in site-specific glycosylation and solid-phase enzymic oligosaccharide synthesis. *Biochem. J.* **300 (Pt 3)**, 843–850 (1994).

188. Macmillan, D., Bill, R. M., Sage, K. A., Fern, D. & Flitsch, S. L. Selective in vitro glycosylation of recombinant proteins: Semi-synthesis of novel homogeneous glycoforms of human erythropoietin. *Chem. Biol.* **8,** 133–145 (2001).

189. Shin, I., Jung, H. J. & Lee, M. R. Chemoselective ligation of maleimidosugars to peptides/protein for the preparation of neoglycopeptides/neoglycoprotein. *Tetrahedron Lett.* **42,** 1325–1328 (2001).

190. Ferna, V. *et al.* Polysaccharide Vaccine Against Haemophilus influenzae Type b. *Science.* **305,** 522–525 (2004).

191. Taniguchi, N., Nishikawa, A., Fujii, S. & Gu, J. Glycosyltransferase Assays Using Pyridylaminated Acceptors: N-Acetylglucosaminyltransferase III, IV, and V. *Methods Enzymol.* **179,** 397–408 (1989).

192. Boutureira, O., Bernardes, G. J. L., Fernández-González, M., Anthony, D. C. & Davis, B. G. Selenenylsulfide-Linked Homogeneous Glycopeptides and Glycoproteins: Synthesis of Human 'Hepatic Se Metabolite A'. *Angew. Chemie Int. Ed.* **51,** 1432–1436 (2012).

193. Gamblin, D. P. *et al.* Glycosyl phenylthiosulfonates (glyco-PTS): novel reagents for glycoprotein synthesis. *Org. Biomol. Chem.* **1,** 3642–3644 (2003).

194. Davis, B. G., Maughan, M. A. T., Green, M. P., Ullman, A. & Jones, J. B. Glycomethanethiosulfonates: Powerful reagents for protein glycosylation. *Tetrahedron Asymmetry* **11,** 245–262 (2000).

195. Dondoni, A., Massi, A., Nanni, P. & Roda, A. A new ligation strategy for peptide and protein glycosylation: Photoinduced thiol-ene coupling. *Chem. - A Eur. J.* **15,** 11444–11449 (2009).

196. Chalker, J. M. *et al.* Methods for converting cysteine to dehydroalanine on peptides and proteins. *Chem. Sci.* **2,** 1666 (2011).

197. Chalker, J. M., Lercher, L., Rose, N. R., Schofield, C. J. & Davis, B. G. Conversion of cysteine into dehydroalanine enables access to synthetic histones bearing diverse post-translational modifications. *Angew. Chemie Int. Ed.* **51,** 1835–1839 (2012).

198. Lercher, L. *et al.* Generation of a synthetic GlcNAcylated nucleosome reveals regulation of stability by H2A-Thr101 GlcNAcylation. *Nat. Commun.* **6,** 7978 (2015).

199. Yang, A. *et al.* A chemical biology route to site-specific authentic protein modifications. *Science.* **354,** 623–626 (2016).

200. Lin, Y. A., Chalker, J. M., Floyd, N., Bernardes, G. J. L. & Davis, B. G. Allyl Sulfides Are Privileged Substrates in Aqueous Cross-Metathesis: Application to Site-Selective Protein Modification. *J. Am. Chem. Soc.* **130,** 9642–9643 (2008).

201. Lin, Y. A., Chalker, J. M. & Davis, B. G. Olefin cross-metathesis on proteins: investigation of allylic chalcogen effects and guiding principles in metathesis partner selection. *J. Am. Chem. Soc.* **132,** 16805–16811 (2010).

202. Lin, Y. A., Chalker, J. M. & Davis, B. G. Olefin metathesis for site-selective protein modification. *Chembiochem* **10,** 959–969 (2009).

203. Chalker, J. M., Wood, C. S. C. & Davis, B. G. A convenient catalyst for aqueous and protein Suzuki-Miyaura cross-coupling. *J. Am. Chem. Soc.* **131,** 16346–16347 (2009).

204. Heinrikson, R. L. On the alkylation of amino acid residues at the active site of ribonuclease. *J. Biol. Chem.* **241,** 1393–1405 (1966).

205. de Graaf, A. J., Kooijman, M., Hennink, W. E. & Mastrobattista, E. Nonnatural amino acids for site-specific protein conjugation. *Bioconjug. Chem.* **20,** 1281–1295 (2009).

206. Spicer, C. D. *et al.* Selective chemical protein modification. *Nat. Commun.* **5,** 4740 (2014).

207. Staudinger, H. & Meyer, J. Uber neue organische Phosphorverbindungen III. Phosphinmethylenderivate und Phosphinimine. *Helv. Chim. Acta* **2,** 635–646 (1919).

208. Saxon, E. & Bertozzi, C. R. Cell Surface Engineering by a Modified Staudinger Reaction. *Science.* **287,** 2007–2010 (2000).

209. Saxon, E., Armstrong, J. I. & Bertozzi, C. R. A 'Traceless' Staudinger Ligation for the Chemoselective Synthesis of Amide Bonds. *Org. Lett.* **2,** 2141–2143 (2000).

210. Bernardes, G. J. L., Linderoth, L., Doores, K. J., Boutureira, O. & Davis, B. G. Site-Selective Traceless Staudinger Ligation for Glycoprotein Synthesis Reveals Scope and Limitations. *ChemBioChem* **12,** 1383–1386 (2011).

211. Kiick, K. L., Saxon, E., Tirrell, D. A. & Bertozzi, C. R. Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *Proc. Natl. Acad. Sci. USA.* **99,** 19–24 (2002).

212. Tsao, M.-L., Tian, F. & Schultz, P. G. Selective Staudinger Modification of Proteins Containing p-Azidophenylalanine. *ChemBioChem* **6,** 2147–2149 (2005).

213. Merkel, L., Beckmann, H. S. G., Wittman, V. & Budisa, N. Efficient N-terminal glycoconjugation of proteins by the N-end rule. *ChemBioChem* **9,** 1220–1224 (2008).

214. De Munari, S., Schiffner, T. & Davis, B. G. A triply divergent reagent for glycoprotein synthesis. *Isr. J. Chem.* **55,** 387–391 (2015).

215. Deiters, A. *et al.* Adding Amino Acids with Novel Reactivity to the Genetic Code of Saccharomyces Cerevisiae. **1,** 11782–11783 (2003).

216. Agard, N. J., Prescher, J. A. & Bertozzi, C. R. A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *J. Am. Chem. Soc.* **126,** 15046–15047 (2004).

217. Lang, K. *et al.* Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction. *Nat. Chem.* **4,** 298–304 (2012).

218. Floyd, N., Vijayakrishnan, B., Koeppe, J. R. & Davis, B. G. Thiyl glycosylate of olefinic proteins: S-linked glycoconjugate synthesis. *Angew. Chemie - Int. Ed.* **48,** 7798–7802 (2009).

219. Rodriguez, E. C., Winans, K. A., King, D. S. & Bertozzi, C. R. A strategy for the chemoselective synthesis of O-linked glycopeptides with native sugar-peptide linkages. *J. Am. Chem. Soc.* **119,** 9905–9906 (1997).

220. Wang, L., Zhang, Z., Brock, A. & Schultz, P. G. Addition of the keto functional group to the genetic code of Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 56–61 (2003).

221. Liu, H., Wang, L., Brock, A., Wong, C. & Schultz, P. G. A Method for the Generation of Glycoprotein Mimetics. *JACS* **125,** 1702–1703 (2003).

222. Xie, J. *et al.* The site-specific incorporation of p-iodo-L-phenylalanine into proteins for structure determination. *Nat. Biotechnol.* **22,** 1297–1301 (2004).

223. Spicer, C. D. & Davis, B. G. Palladium-mediated site-selective Suzuki-Miyaura protein modification at genetically encoded aryl halides. *Chem. Commun. (Camb).* **47,** 1698–700 (2011).

224. Spicer, C. D. & Davis, B. G. Rewriting the bacterial glycocalyx via Suzuki-Miyaura cross-coupling. *Chem. Commun. (Camb).* **49,** 2747–9 (2013).

225. Li, N., Lim, R. K. V, Edwardraja, S. & Lin, Q. Copper-free Sonogashira cross-coupling for functionalization of alkyne-encoded proteins in aqueous medium and in bacterial cells. *J. Am. Chem. Soc.* **133,** 15316–15319 (2011).

226. Li, J. *et al.* Ligand-free palladium-mediated site-specific protein labeling inside gram-negative bacterial pathogens. *J. Am. Chem. Soc.* **135,** 7330–8 (2013).

227. The Uniprot Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **45,** D158–D169 (2017).

228. Young, T. S., Ahmad, I., Yin, J. A. & Schultz, P. G. An Enhanced System for Unnatural Amino Acid Mutagenesis in E. coli. *J. Mol. Biol.* **395,** 361–374 (2010).

229. Kalia, J. & Raines, R. T. Hydrolytic stability of hydrazones and oximes. *Angew. Chemie Int. Ed.* **47,** 7523–7526 (2008).

230. Torborg, C. & Beller, M. Recent applications of palladium-catalyzed coupling reactions in the pharmaceutical, agrochemical, and fine chemical industries. *Adv. Synth. Catal.* **351,** 3027–3043 (2009).

231. Heck, R. F. & Nolley, J. P. Palladium-catalyzed vinylic hydrogen substitution reactions with aryl, benzyl, and styryl halides. *J. Org. Chem.* **37,** 2320–2322 (1972).

232. Heck, R. F. Arylation, Methylation, and Carboxyalkylation of Olefins by Group VIII Metal Derivatives. *J. Am. Chem. Soc.* **90,** 5526–5531 (1968).

233. Biscoe, M. R., Fors, B. P. & Buchwald, S. L. Communication A New Class of Easily Activated Palladium Precatalysts for Facile C-N Cross-Coupling Reactions and the Low Temperature Oxidative Addition of Aryl Chlorides. *J. Am. Chem. Soc.* **130,** 6686–6687 (2008).

234. Amatore, C., Jutand, A. & Thuilliez, A. Formation of palladium(0) complexes from Pd(OAc)2 and a bidentate phosphine ligand (dppp) and their reactivity in oxidative addition. *Organometallics* **20,** 3241–3249 (2001).

235. Melvin, P. R., Balcells, D., Hazari, N. & Nova, A. Understanding Precatalyst Activation in Cross-Coupling Reactions: Alcohol Facilitated Reduction from Pd(II) to Pd(0) in Precatalysts of the Type (η^3-allyl)Pd(L)(Cl) and (η^3-indenyl)Pd(L)(Cl). *ACS Catal.* **5,** 5596–5606 (2015).

236. Lennox, A. J. J. & Lloyd-Jones, G. C. Selection of boron reagents for Suzuki-Miyaura coupling. *Chem. Soc. Rev.* **43,** 412–43 (2014).

237. Adamo, C., Amatore, C., Ciofini, I., Jutand, A. & Lakmini, H. Mechanism of the Palladium-Catalyzed Homocoupling of Arylboronic Acids : Key Involvement of a Palladium Peroxo Complex. *JACS* **128,** 6829–6836 (2006).

238. King, A. O., Okukado, N. & Negishi, E. Highly general stereo-, regio-, and chemo-selective synthesis of terminal and internal conjugated enynes by the Pd-catalysed reaction of alkynylzinc reagents with alkenyl halides. *J. Chem. Soc. Chem. Commun.* 683–684 (1977).

239. Hatanaka, Y. & Hiyama, T. Cross-coupling of organosilanes with organic halides mediated by a palladium catalyst and tris (diethylamino) sulfonium difluorotrimethylsilicate. *J. Org. Chem.* **53,** 918–920 (1988).

240. Kiso, Y., Tamao, K. & Kumada, M. Effects of the nature of halides on the alkyl group isomerization in the nickel-catalyzed cross-coupling of secondary alkyl Grignard reagents with organic halides. *J. Organomet. Chem.* **94,** 4374–4376 (1972).

241. Sonogashira, K., Tohda, Y. & Hagihara, N. A convenient synthesis of acetylenes: catalytic substitutions of acetylenic hydrogen with bromoalkenes, iodoarenes and bromopyridines. *Tetrahedron Lett.* **16,** 4467–4470 (1975).

242. Milstein, D. & Stille, J. K. A general, selective, and facile method for ketone synthesis from acid chlorides and organotin compounds catalyzed by palladium. *J. Am. Chem. Soc.* **100,** 3636–3638 (1978).

243. Miyaura, N., Yamada, K. & Suzuki, A. A new stereospecific cross-coupling by the palladium-catalyzed reaction of 1-alkenylboranes with 1-alkenyl or 1-alkynyl halides. *Tetrahedron Lett.* 3437–3440 (1979).

244. Miyaura, N. and Suzuki, A. Stereoselective Synthesis of Arylated (E)-Alkenes by the Reaction of Alk-1-enylboranes with Aryl Halides in the Presence of Palladium Catalyst. *J. Chem. Soc., Chem. Commun.* 866–867 (1979).

245. Miyaura, N., Yanagi, T. & Suzuki, A. The Palladium-Catalyzed Cross-Coupling Reaction of Phenylboronic Acid with Haloarenes in the Presence of Bases. *Synth. Commun.* **11,** 513–519 (1981).

246. Miyaura, N., Yamada, K., Suginome, H. & Suzuki, A. Novel and convenient method for the stereo- and regiospecific synthesis of conjugated alkadienes and alkenynes via the palladium-catalyzed cross-coupling reaction of 1-alkenylboranes with bromoalkenes and bromoalkynes. *J. Am. Chem. Soc.* **107,** 972–980 (1985).

247. Suzuki, A. Organoborane coupling reactions (Suzuki coupling). *Proc. Japan Acad. Ser. B* **80,** 359–371 (2004).

248. Lennox, A. J. J. & Lloyd-Jones, G. C. Transmetalation in the Suzuki-Miyaura coupling: the fork in the trail. *Angew. Chem. Int. Ed. Engl.* **52,** 7362–7073 (2013).

249. Hall, D. G. in *Boronic Acids* 1–99 (Wiley-VCH Verlag GmbH & Co. KGaA, 2005).

250. Lennox, A. J. J. & Lloyd-Jones, G. C. The Slow-Release Strategy in Suzuki – Miyaura Coupling. *Isr. J. Chem.* **50,** 664–674 (2010).

251. Zhou, J., Guo, X., Tu, C., Li, X. & Sun, H. Aqueous Suzuki coupling reaction catalyzed by water-soluble diimine/Pd(II) systems. *J. Organomet. Chem.* **694,** 697–702 (2009).

252. Ahmadi, Z. & McIndoe, J. S. A mechanistic investigation of hydrodehalogenation using ESI-MS. *Chem. Commun.* **49,** 11488-11490 (2013).

253. Gao, Z., Gouverneur, V. & Davis, B. G. Enhanced aqueous Suzuki-Miyaura coupling allows site-

specific polypeptide 18F-labeling. *J. Am. Chem. Soc.* **135,** 13612–13615 (2013).

254. Dumas, A. *et al.* Self-liganded Suzuki-Miyaura coupling for site-selective protein PEGylation. *Angew. Chem. Int. Ed. Engl.* **52,** 3916–21 (2013).

255. Kodama, K. *et al.* Site-specific functionalization of proteins by organopalladium reactions. *ChemBioChem* **8,** 232–238 (2007).

256. Laulhé, S., Blackburn, J. M. & Roizen, J. L. Exhaustive Suzuki-Miyaura Reactions of Polyhalogenated Heteroarenes with Alkyl Boronic Pinacol Esters. *Chem. Commun.* 1–4 (2017).

257. Robbins, D. W. & Hartwig, J. F. A C-H borylation approach to suzuki-miyaura coupling of typically unstable 2-heteroaryl and polyfluorophenyl boronates. *Org. Lett.* **14,** 4266–4269 (2012).

258. Li, Y. *et al.* Palladacycle-catalyzed Suzuki–Miyaura reaction of aryl/heteroaryl halides with MIDA boronates in EtOH/H_2O or H_2O. *RSC Adv.* **4,** 36262–36266 (2014).

259. Isley, N. A., Gallou, F. & Lipshutz, B. H. Transforming Suzuki-Miyaura cross-couplings of MIDA boronates into a green technology: no organic solvents. *J. Am. Chem. Soc.* **135,** 17707–17710 (2013).

260. Molander, G. A. & Felix, L. A. Stereoselective Suzuki-Miyaura cross-coupling reactions of potassium alkenyltrifluoroborates with alkenyl bromides. *J. Org. Chem.* **70,** 3950–3956 (2005).

261. Yamamoto, Y., Takizawa, M., Yu, X.-Q. & Miyaura, N. Cyclic triolborates: air- and water-stable ate complexes of organoboronic acids. *Angew. Chem. Int. Ed.* **47,** 928–931 (2008).

262. Lennox, A. J. J. & Lloyd-Jones, G. C. Organotrifluoroborate hydrolysis: Boronic acid release mechanism and an acid-base paradox in cross-coupling. *J. Am. Chem. Soc.* **134,** 7431–7441 (2012).

263. Chatterjee, A. & Ward, T. R. Recent Advances in the Palladium Catalyzed Suzuki–Miyaura Cross-Coupling Reaction in Water. *Catal. Letters* **146,** 820–840 (2016).

264. Polshettiwar, V., Decottignies, A., Len, C. & Fihri, A. Suzuki-Miyaura Cross-Coupling Reactions in Aqueous Media: Green and Sustainable Syntheses of Biaryls. *ChemSusChem* **3,** 502–522 (2010).

265. Nishimura, M., Ueda, M. & Miyaura, N. Palladium-catalyzed biaryl-coupling reaction of arylboronic acids in water using hydrophilic phosphine ligands. *Tetrahedron* **58,** 5779–5787 (2002).

266. Fleckenstein, C. A. & Plenio, H. Aqueous cross-coupling : highly efficient Suzuki – Miyaura coupling of N -heteroaryl halides and N -heteroarylboronic acids {. *Green Chem.* **9,** 1287–1291 (2007).

267. Wu, W., Chen, S. & Tsai, F. Recyclable and highly active cationic 2,2-bipyridyl palladium(II) catalyst for Suzuki cross-coupling reaction in water. *Tetrahedron Lett.* **47,** 9267–9270 (2006).

268. Crisóstomo-Lucas, C., Toscano, R. A. & Morales-Morales, D. Synthesis and characterization of new potentially hydrosoluble pincer ligands and their application in Suzuki-Miyaura cross-coupling reactions in water. *Tetrahedron Lett.* **54,** 3116–3119 (2013).

269. Edwards, G. A. *et al.* Melamine and melamine-formaldehyde polymers as ligands for palladium and application to Suzuki-Miyaura cross-coupling reactions in sustainable solvents. *J. Org. Chem.* **79,** 2094–2104 (2014).

270. Li, X., Liu, C., Wang, L., Ye, Q. & Jin, Z. Palladium/Amine Complex Catalyzed Suzuki–Miyaura Reaction of Arylboron Compounds in Water. *Chem. Sel.* **2,** 4016–4020 (2017).

271. Mondal, M. & Bora, U. *O*-Aryloxime ether analogues as novel and efficient ligands for palladium-catalyzed Suzuki-Miyaura coupling in water. *Tetrahedron Lett.* **55,** 3038–3040 (2014).

272. Miyaura, S. *et al.* Facile-prepared sulfonated water-soluble PEPPSI-Pd-NHC catalysts for aerobic aqueous Suzuki–Miyaura cross-coupling reactions. *Green Chem.* **16,** 4955–4962 (2014).

273. Shi, J. *et al.* N-Heterocyclic Carbene Conjugated with Poly(ethylene glycol) for Palladium-Catalyzed Suzuki–Miyaura Coupling in Aqueous Solvents. *Catal. Letters* **144,** 158–164 (2013).

274. Tzschucke, C. C., Markert, C., Glatz, H. & Bannwarth, W. Katalyse in fluorigen Zweiphasensystemen ohne perfluorierte L^sungsmittel: Anwendung. *Angew. Chemie* **114,** 4678–4681 (2002).

275. Baleiza, C., Corma, A., Garcia, H. & Leyva, A. Oxime Carbapalladacycle Covalently Anchored to High Surface Area Inorganic Supports or Polymers as Heterogeneous Green Catalysts for the Suzuki Reaction in Water. *J. Org. Chem* **69,** 439–446 (2004).

276. Wang, L. & Cai, C. Journal of Molecular Catalysis A : Chemical Fluorous silica gel-supported perfluoro-tagged palladium nanoparticles catalyze Suzuki cross-coupling reaction in water. *J. Mol. Catal. A Chem.* **306,** 97–101 (2009).

277. Firouzabadi, H., Iranpoor, N., Gholinejad, M. & Kazemi, F. RSC Advances Agarose hydrogel as an effective bioorganic ligand and support for the stabilization of palladium nanoparticles . Application as a recyclable catalyst for Suzuki-Miyaura reaction in aqueous media. *RSC Adv.* **1,** 1013–1019 (2011).

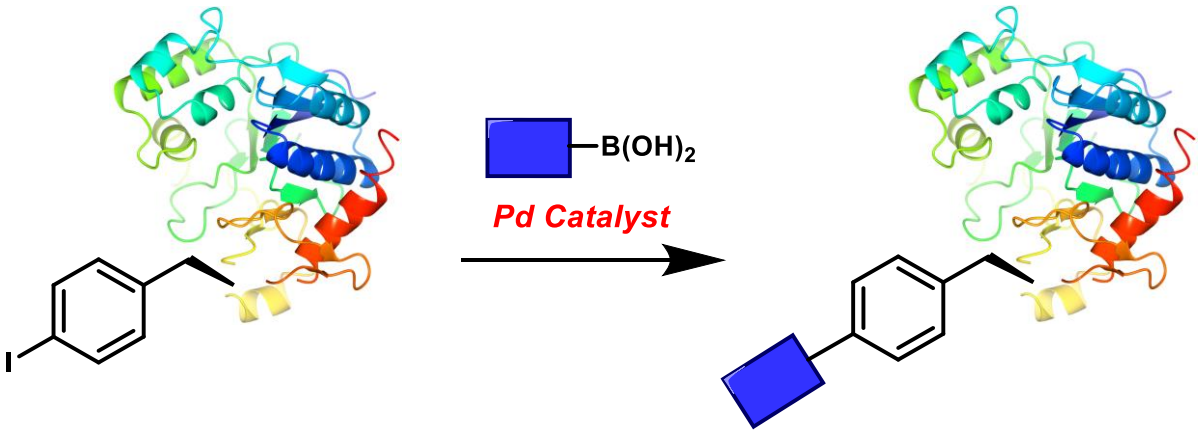
278. Ojida, A., Tsutsumi, H., Kasagi, N. & Hamachi, I. Suzuki coupling for protein modification. *Tetrahedron Letters* **46,** (2005).

279. Brustad, E. *et al.* A genetically encoded boronate-containing amino acid. *Angew. Chem. Int. Ed. Engl.* **47,** 8220–8223 (2008).

280. Spicer, C. D., Triemer, T. & Davis, B. G. Palladium-mediated cell-surface labeling. *J. Am. Chem. Soc.* **134,** 800–803 (2012).

281. Yusop, R. M., Unciti-Broceta, A., Johansson, E. M., Sanchez-Martin, R. M. & Bradley, M. Palladium-mediated intracellular chemistry. *Nat Chem* **3,** 239–243 (2011).

Chapter 2



Suzuki-Miyaura Cross-coupling to Glycogenin-1

Table of Contents

2.1 Introduction 78

2.1.1 Genetic Incorporation of L-*p*-Iodophenylalanine 78

2.1.2. *p*IPhe as a Reactive Handle for Cross-coupling: the Challenge of Enzymes 80

2.1.3 Accessing "Minimal GYG Mimics" 80

2.2 Site-Specific Incorporation of *p*IPhe into Glycogenin-1 82

2.3 Suzuki-Miyaura Cross-coupling: Preliminary Studies 84

2.4 GYG-3HMPB: a Putative "Minimal Glycomimetic" 87

2.5 Development of an Enzyme-Compatible Cross-coupling 92

2.6 Investigating the Activity of Minimal GYG Glycomimetics 97

2.7 Conclusion 102

2.8 Experimental Section 103

2.8.1 Chemical Synthesis 103

2.8.2 Plasmid DNA Details and Sequencing 106

2.8.3 Protein Expression and Characterisation 109

2.8.3.1 GYG-Y195pIPhe 109

2.8.3.2 GYG-wt 111

2.8.3.3 GYG-Y195F 113

2.8.3.4 Circular Dichroism Comparison of GYG-wt and GYG-Y195*p*IPhe 114

2.8.4 Cross-couplings to 3-furanylboronic acid 115

2.8.4.1 Initial Cross-coupling Test 115

2.8.4.2 Comparison of Catalyst Ligands 116

2.8.4.3 Effect of Boronic Acid Loading (1) (Table 2.1, entries 1-2) 118

2.8.4.4 Effect of Further Increase in Boronic Acid Loading (Table 2.1, entry 3) 119

2.8.4.5 Negative Control Reactions 120

2.8.5 Cross-couplings to 3-HMPB 123

2.8.5.1 Initial Cross-coupling Test 123

2.8.5.2 Optimisation of Boronic Acid Loading 124
2.8.5.3 Effect of pH and temperature variation 126
2.8.5.4 Effect of Protein Concentration 128
2.8.5.5 Example of Cross-coupling without stepwise boronic acid addition 129
2.8.5.6 Negative Control Reactions 130
2.8.6 An "Enzyme-compatible" SMC: Optimisations on GYG-wt 133
2.8.6.1 Method for Analysis of GYG-wt Autoglucosylation from LC-MS Spectra 133
2.8.6.2 Autoglucosylation Activity of GYG-wt 134
2.8.6.3 GYG-wt Actvity after SMC Treatment: Initial Attempt 135
2.8.6.4 Effect of Individual SMC Components on GYG-wt Activity 136
2.8.6.5 Effect of DTT Concentration on Catalytic Activity: With and Without Pd 138
2.8.6.6 Circular Dichroism: Effect of SMC Reaction Components 143
2.8.6.7 Concentration-dependent Effect of Palladium on GYG-wt Activity 144
2.8.6.8 Optimised SMC Protocol for Retention of GYG-wt Activity (with 3-HMPB) 146
2.8.6.9 Circular Dichroism of GYG-wt after SMC Conditions 147
2.8.7 Catalytic Activity of Minimal GYG Glycomimetics 148
2.8.7.1 General Procedure 148
2.8.7.2 GYG-3HMPB (boronic acid = 3-HMPB) 149
2.8.7.3 GYG-3HPB (boronic acid = 3-HPB) 150
2.8.7.4 GYG-2HMPB (boronic acid = 2-HMPB) 152
2.8.7.5 GYG-4HPB (boronic acid = 4-HPB) 154
2.8.7.6 GYG-4HMPB (boronic acid = 4-HMPB) 156
2.8.8 Further Evaluation of 4-HMPB and 4-HPB Systems 158
2.8.8.1 Glucosylation Activity of Non-treated GYG-wt, Utilised in 4-HMPB/4-HPB Experiments 158
2.8.8.2 Effect of Pd Loading on GYG-wt Activity after 4-HMPB/4-HPB SMC Conditions 160

2.9 References 162

2.1 Introduction

2.1.1 Genetic Incorporation of L-*p*-Iodophenylalanine

The advent and development of codon suppression technology has allowed the site-specific incorporation of a diverse range of unnatural amino acids into proteins. Such an approach relies on the fact that, of the 64 naturally occurring codons, only 61 encode for an amino acid; the remaining 3 lack an amino acid partner, instead causing termination of the translation process through the recruitment of release factors.[1] These codons are TAG (amber), TGA (opal) and TAA (ochre),[2] and their degeneracy opens up the possibility for their reassignment to other, unnatural, amino acids via use of a suitable suppressor tRNA (i.e. a tRNA that recognises a stop codon but does not lead to translation termination, instead inserting an amino acid at that site). Of the three, amber codon suppression represents the most popular option owing to the rarity of this codon in *E. coli*,[3] although approaches utilising opal and ochre codon suppression also exist.[4,5] This technology was discussed briefly in the previous chapter in the context of glycoprotein synthesis (section 1.3.3); here, a further overview is provided.

The first report of the application of amber codon suppression involved the use of a TAG-specific tRNA chemically acylated with the unnatural amino acid to be incorporated.[6] As noted in the preceding chapter,[7,8] such methodology is limited to either *in vitro* transcription-translation systems (or – given a need to micro-inject the tRNA – large cells) and its more general application to common *in vivo* protein expression systems additionally requires an aminoacyl-tRNA synthetase (aaRS) capable of charging the tRNA with the desired unnatural amino acid. [2,9] The tRNA/aaRS pair must be orthogonal to the host organism; that is, the aaRS must not recognise and charge any endogeneous tRNAs, and the suppressor tRNA must not be charged by any endogeneous aaRSs.[9]

If such conditions are met, then unnatural amino acid incorporation during *in vivo* protein expression can proceed as illustrated below (**figure 2.1**). The suppressor tRNA is charged by the corresponding aaRS with the unnatural amino acid (UAA). Subsequently, this aminoacyl-tRNA recognises the amber (UAG) codon of the template mRNA, in the ribosome. The peptidyltransferase activity of the latter[10] leads to incorporation of the UAA into the nascent polypeptide chain, ultimately leading, upon translation completion, to the generation of a protein bearing this unnatural amino acid as a "tag".[11,12]

A number of suitable such systems now exist for *E. coli* [13,14] and also for eukaryotic organisms;[15–17] one prominent example is the *M. jannaschii* tyrosyl tRNA/aaRS pair, orthogonal in *E. coli*.[13,18] This has been used to incorporate a variety of non-canonical amino acids, principally derivatives of phenylalanine.[9,19–22] Of

direct relevance for this work is the use of this system for incorporation of L- *p*-iodophenylalanine (*p*IPhe) via amber codon suppression, first reported by Wang *et al* in the context of heavy atom incorporation for crystallographic purposes.[23] Typically, unnatural acid incorporation results in lower protein yields as compared to wild-type protein, owing primarily to competition of stop codon-mediated incorporation with the natural function of that codon, i.e. translation termination resulting in truncated or degraded proteins. However, improvements in this regard for *p*IPhe and other amino acids resulted from the development of an optimised plasmid construct for unnatural amino acid mutagenesis, "pEVOL".[24] This plasmid encodes for two copies of the requisite tRNA/aaRS pair, one under the control of a constitutive promoter and the other following an inducible promoter. This would form the backbone of our efforts to incorporate *p*IPhe into GYG, to act as a tag for subsequent cross-coupling studies towards our goal of generating homogeneous glycoforms or mimics thereof.

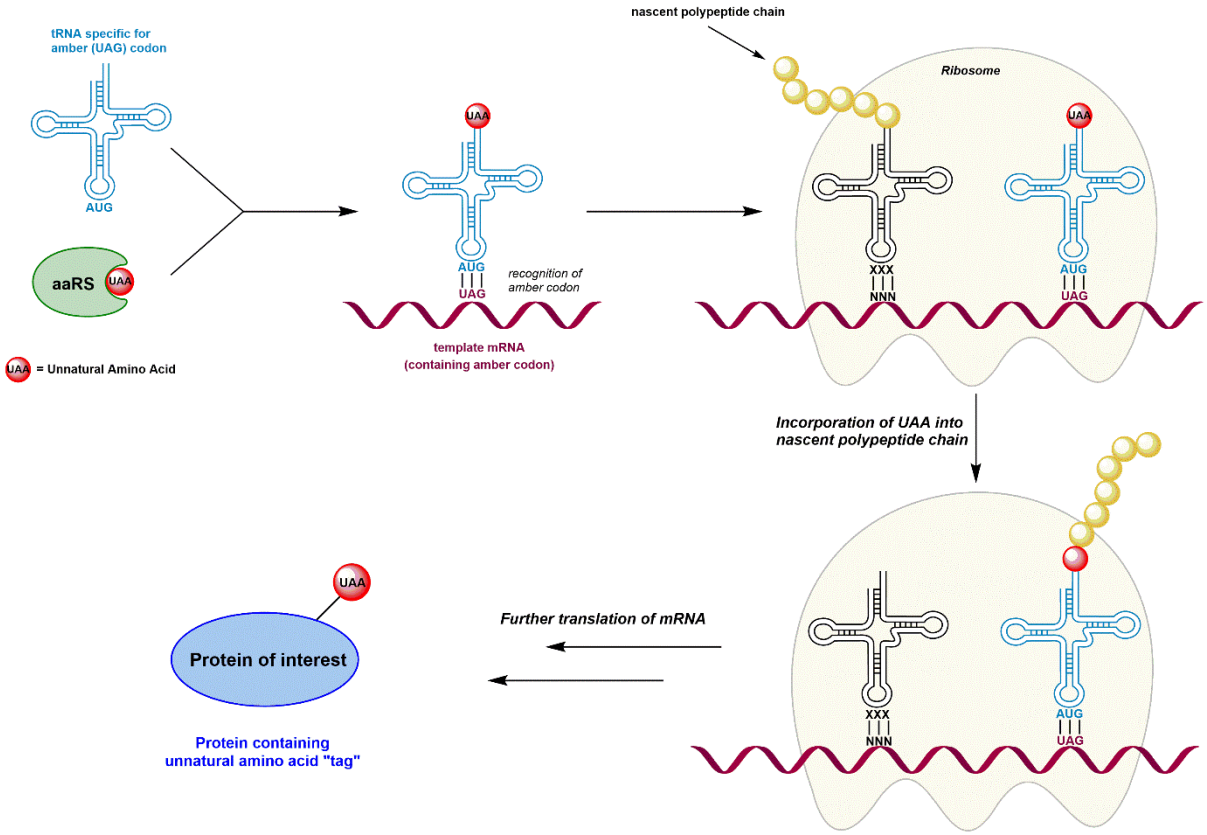


Figure 2.1 A generally applicable amber codon suppression method requires 2 added components: a tRNA specific for the amber codon (UAG), and an aaRS capable of charging this tRNA with the desired unnatural amino acid. This can lead to incorporation of the unnatural amino acid (UAA) into a protein during its biosynthesis in *in vivo* expression systems.

2.1.2. *p*IPhe as a Reactive Handle for Cross-coupling: the Challenge of Enzymes

Whilst the first reported genetic incorporation of L-*p*-iodophenylalanine into a protein substrate came in 2004,[23] it would be a further 7 years until this would be combined with its evident potential as a reactive handle for Suzuki-Miyaura cross-coupling (SMC). Spicer *et al* demonstrated site-specific incorporation of this amino acid into maltose binding protein (MBP), and its subsequent coupling to a model boronic acid,[25] taking advantage of an efficient, water-soluble palladium catalyst developed shortly beforehand (see chapter 1).[26] A key challenge encountered was the non-specific coordination of palladium to the protein surface,[27] precluding reaction analysis by LC-MS, however this was readily solved through the addition of a "palladium scavenger" (in this case 3-mercaptopropionic acid) post-reaction.

Subsequently, SMC has been demonstrated on other *p*IPhe-containing protein substrates, such as the pentapeptide repeat protein from *Nostoc Punctiforme,* Np275-276 or the outer membrane protein OmpC.[21-23] However, the translation of this reaction to an enzyme substrate such as GYG, particularly given our aim of generating catalytically active glycoconjugates, presents new challenges in addition to those associated with the transfer of SMC to any new protein (e.g. differing local environment and accessibility of *p*IPhe residue, differing surface charge that may affect Pd scavenging). In contrast to examples on model proteins such as MBP, it is in this case a pre-requisite that the reaction conditions themselves should yield the enzyme in a correctly folded, potentially active state, for fair evaluation of the GYG glycoconjugates generated.

Thus, the SMC reaction conditions must here be optimised in two senses. Firstly, in the conventional sense with respect to reaction conversion, but additionally with respect to "enzyme compatibility". Whilst a number of factors influencing the former in a protein modification context can be gauged from analogous studies on other protein substrates,[30–32] the latter represented relatively less well understood territory for us, and its exploration occupies a significant proportion of this chapter.

2.1.3 Accessing "Minimal GYG Mimics"

The fundamental initial goal of this thesis is the generation of homogeneous GYG glycoconjugates for mechanistic study, which may be achieved through the ligation, via SMC, of a suitable carbohydrate-derived boronic acid to *p*IPhe-tagged GYG (see **figure 1.28**, chapter 1). However, in the preceding chapter it was noted that the carbohydrate moiety might be more crudely approximated by more generic, simpler structures bearing a hydroxyl moiety of suitable orientation and pK_A to mimic the carbohydrate glycosyl acceptor function (see **figure 1.16 b**, chapter 1). By virtue of its autocatalytic nature, GYG must exhibit a

certain level of plasticity with respect to the glycosyl acceptor, since the latter's structure is necessarily altered with each glucosylation step. This requirement is further highlighted by the fact that GYG is capable of catalysing two chemically distinct glucosylation steps: that of the initial Tyr^{195} acceptor and that of the 4'-OH of a glucose residue.[33] As such, we considered that the precise structure of the enzyme-bound glycosyl acceptor might not be critical; rather, any correctly positioned –OH group may be sufficient for enzymatic activity. Investigation of this hypothesis, and thus of enzyme "plasticity", represented a point of interest.

Based on the above, it was envisaged that cross-coupling of simple hydroxyl-containing arylboronic acids to GYG might be sufficient to yield a catalytically active enzyme species (**figure 2.2**). Such cross-coupled enzymes can be considered as "minimal mimics" of the native mono-glucosylated GYG species, in which the presence of a glycosyl acceptor at the 195 position is retained, but the glucose structural unit is rather crudely replaced by an aryl moiety. Importantly, and in contrast to more complex carbohydrate-derived boronic acids, a number of hydroxyl-containing arylboronic acids are readily commercially available, rendering them better suited for initial optimisation studies on general SMC reaction parameters for a GYG-*p*IPhe protein substrate.

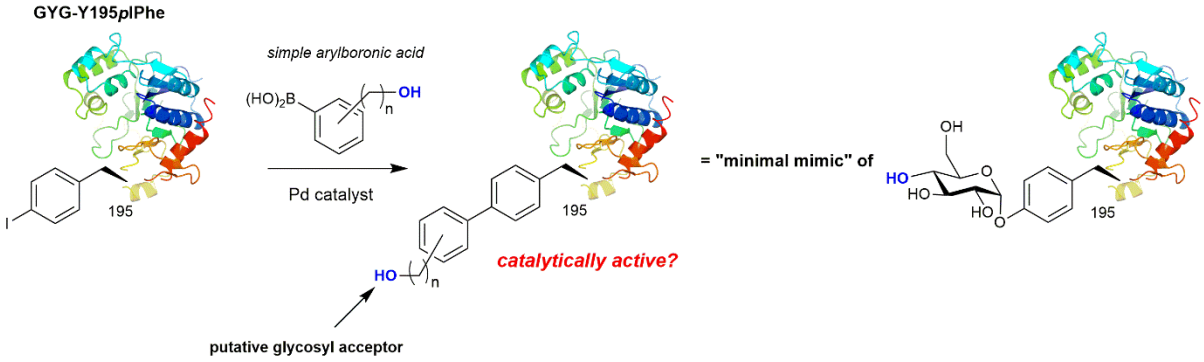


Figure 2.2 Cross-coupling of simple, commercially available arylboronic acids to GYG-Y195*p*IPhe could yield species bearing a putative glycosyl acceptor at the 195 position. Such enzymes may be considered as "minimal mimics" of mono-glucosylated GYG, retaining the glycosyl acceptor but lacking the carbohydrate structural motif.

As noted previously, this thesis deals specifically with the mammalian GYG-1 isoform, as opposed to other variants; in this species, the native glycosyl acceptor tyrosine is at the 195 position.[34] To this end, this chapter first discusses the generation of the "tagged" enzyme GYG-Y195*p*IPhe, in which this native glycosyl acceptor is replaced with the SMC-reactive amino acid L-*p*-iodophenylalanine through amber codon suppression technology (section 2.2). Subsequently, Suzuki-Miyaura cross-coupling to this species is

investigated and optimised, in terms of both conversion and enzyme compatibility; much of this work is carried out on one hydroxyl-containing arylboronic acid suitable for the generation of a minimal GYG glycomimetic as outlined above in **figure 2.2** (sections 2.3 – 2.5). Finally, through cross-coupling of GYG-Y195*p*IPhe to alternative hydroxyl-containing boronic acids, this chapter concludes with an investigation into the generation and activity of a number of different such minimal GYG mimics (section 2.6).

2.2 Site-Specific Incorporation of *p*IPhe into Glycogenin-1

In order to incorporate *p*IPhe (**2.1**) into GYG-1 through amber codon suppression, two plasmids are required: a pEVOL system as outlined above, specific for this amino acid, and a plasmid encoding GYG-1 bearing a "TAG" stop codon at the desired site of incorporation (position 195[34]). The first of these, pEVOL-*p*IPhe, was generously donated by Prof. Peter Schultz to our group some years ago; this encodes a tRNA/aaRS pair enabling incorporation of L-*p*-iodophenylalanine in response to a TAG amber codon installed in a given DNA sequence. The second, GYG-Y195X (where "X" = TAG codon), had been previously prepared within our group via standard site-directed mutagenesis methods from the corresponding wild-type GYG plasmid (see experimental section 2.8.2).[35] This encodes the catalytic domain of GYG-1 with an N-terminal 22-residue fragment containing a polyhistidine tag and TEV cleavage site; the catalytic domain has previously been shown to have near-identical properties to the full-length protein and to exhibit enhanced stability,[36,37] and has thus been widely used in GYG-1 mechanistic studies.[34,36–38] Co-expression of both plasmids in *E. coli* in the presence of the amino acid L-*p*IPhe (**2.1**) should thus generate GYG-Y195*p*IPhe with this amino acid being incorporated in response to the TAG codon at the 195 position (**figure 2.3**).

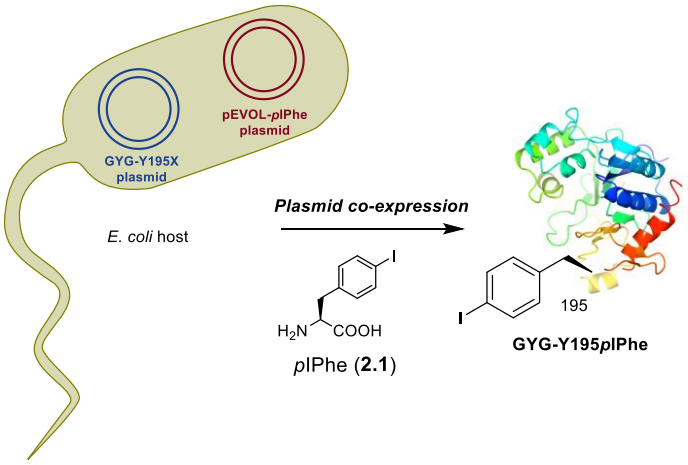


Figure 2.3 Incorporation of *p*IPhe into Glycogenin-1 (GYG-1) at the 195 position can be achieved through co-expression of two plasmids in *E. coli*: GYG-Y195X, encoding GYG bearing a "TAG" codon at this position, and pEVOL-*p*IPhe, encoding the requisite tRNA/aaRS pair for incorporation of *p*IPhe in response to this codon.

Expression of this protein was attempted using conditions based on those previously reported for *p*IPhe incorporation into other proteins, in collaboration with Prof. Wyatt Yue and Mr. Henry Bailey (Structural Genomics Consortium, University of Oxford, see "collaborations", page xii).[25] The corresponding wild-type enzyme, GYG-wt, was also expressed for comparison purposes, as well as the mutant GYG-Y195F, for later control studies. SDS-PAGE and LC-MS analysis after purification by Ni-NTA chromatography confirmed the successful isolation of all His-tagged species (**figure 2.4**). Note that amino acid numbering does not include 22-residue His-tag/TEV cleavage site). Whilst GYG-wt was isolated as a mixture of glycoforms owing to partial and variable extent of *in vivo* glucosylation in the expression host, the GYG-Y195*p*IPhe and GYG-Y195F mutants exhibited no such glucosylation, consistent with site-specific knockout of the native glycosyl acceptor Tyr195 residue. Optimised expression yields were typically 1 mg/L culture which, whilst not high, represented sufficient levels for future studies. For GYG-Y195*p*IPhe and GYG-Y195F, a low level of gluconoylation (+178 Da) was evident; this modification for His-tagged enzymes is a known feature of expression in *E. coli*.[39]

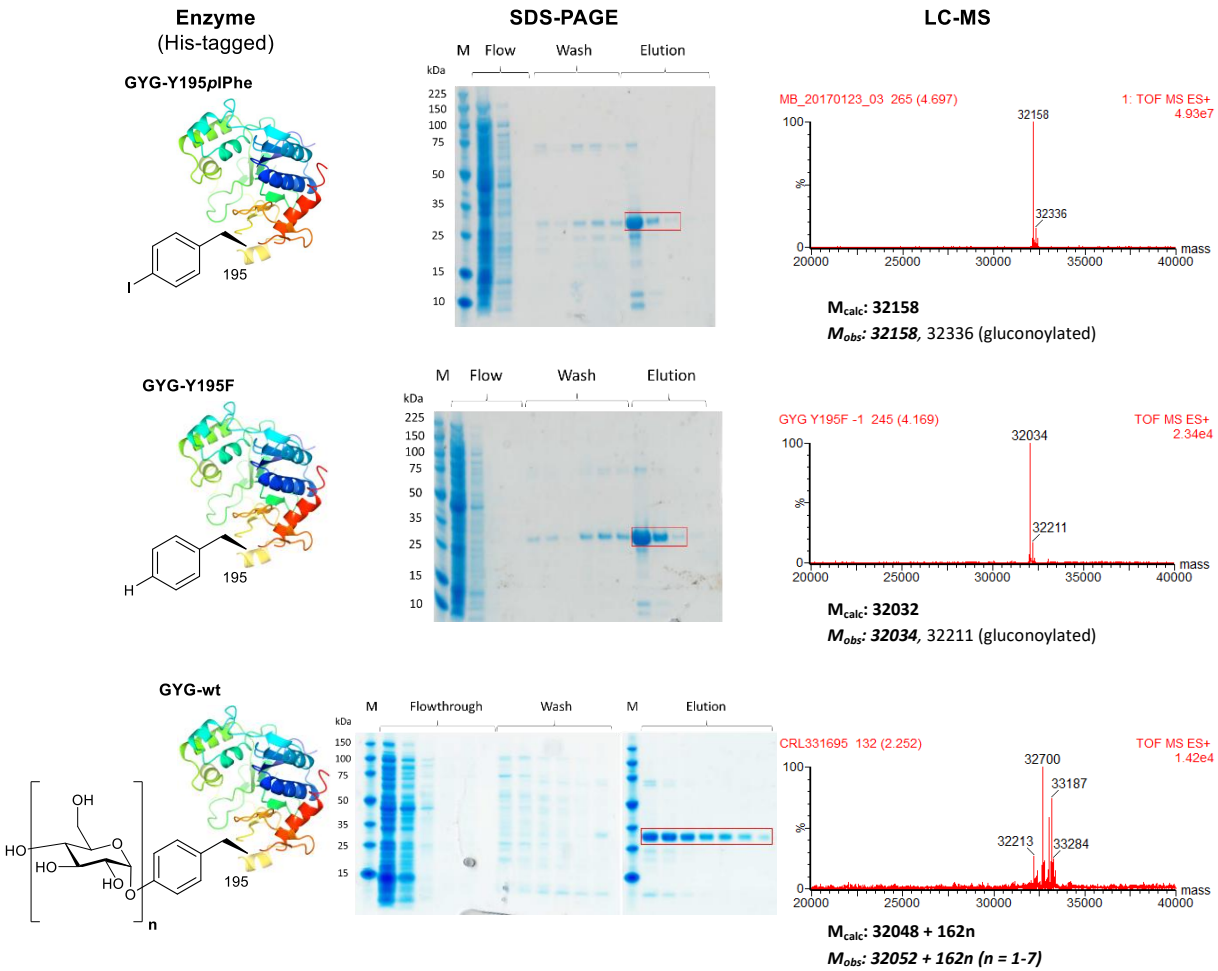


Figure 2.4 Expression and characterisation of GYG-Y195*p*IPhe via amber codon suppression (top), in addition to the corresponding wild-type enzyme GYG-wt (lower).

Comparison of the circular dichroism spectra of GYG-Y195*p*IPhe and the GYG-wt revealed negligible differences in secondary structure (**figure 2.5**). This highlights the fact that *p*IPhe incorporation did not lead to significant structural distortion of the enzyme.

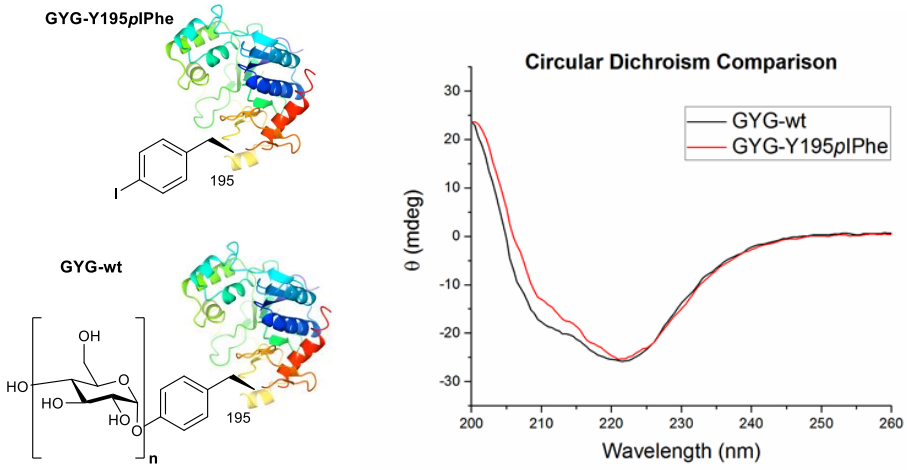


Figure 2.5 Circular dichroism illustrates negligible distortion of protein secondary structure upon incorporation of *p*IPhe, as revealed by comparison to the natural, catalytically active wild-type enzyme.

2.3 Suzuki-Miyaura Cross-Coupling: Preliminary Studies

With requisite protein in hand, its potential to undergo SMC at the *p*IPhe site was investigated. We initially selected a model coupling partner – 3-furanylboronic acid – known to be reactive in cross-couplings to other protein substrates.[25] Adapting previous conditions,[25,32] and utilising the catalyst initially developed by Chalker *et al*,[26] a preliminary cross-coupling test to GYG-Y195*p*IPhe was carried out. Pleasingly, after post-reaction Pd-scavenging with DTT (necessary for LC-MS analysis[25]), formation of cross-coupled product could be detected by LC-MS, albeit in moderate yields (47 % by LC-MS, **figure 2.6**). Without either boronic acid or palladium, no reaction was observed, confirming the genuine cross-coupling nature of the process (see experimental section 2.8.4.5). Consistent with previous work,[25] treatment of protein with palladium without palladium scavenger led to loss of the protein peak in the LC-MS chromatogram (section 2.8.4.5).

Principal side-products were attributable, on the basis of LC-MS analysis (**figure 2.6**), to non-reacted GYG-Y195*p*IPhe ($M_{calc} = M_{obs} = 32158$) and a competitive de-iodination pathway ($M_{calc} = 32032$, $M_{obs} = 32031$).[i]

[i] As noted, the starting material contains a +178 Da adduct attributable to N-terminal gluconoylation. In general, such adducts were insufficiently abundant to be detectable above the (increased) background noise in MS spectra after cross-coupling, although in rare instances they may be observed at low levels (see, for example, experimental section 2.8.7.4). Note that any reaction conversions consider only products and side-products derived from non-gluconoylated species (given that, in most cases, this is the only species detectable).

The latter is a well-established side-reaction in SMC reactions[40–43] (see chapter 1), and has previously been observed for various other protein substrates.[30–32] With reference to the accepted reaction mechanism for dehalogenation discussed in chapter 1, it is apparent that a species capable of coordinating to Pd and acting as a β-hydride donor is a pre-requisite for this process. In this case, the initial Pd ligands and boronic acid both fail to fulfil the latter requirement; most plausible candidates are possibly the palladium scavenger DTT or indeed the protein itself (e.g. the neighbouring Ser196 residue).

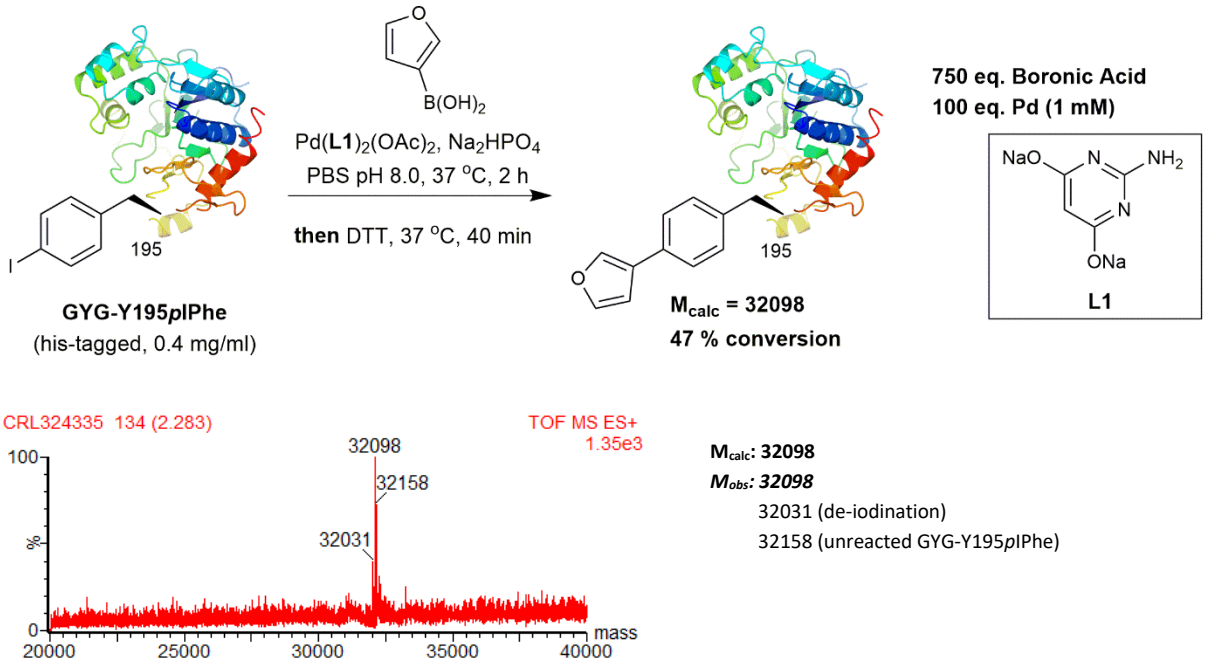


Figure 2.6: Initial cross-coupling between GYG-Y195*p*IPhe and 3-furanylboronic acid. LC-MS revealed masses corresponding to cross-coupled product (32098), unreacted GYG-Y195*p*IPhe (32158) and de-iodinated enzyme (32031).

Following on from the initial work of Chalker *et al*,[26] a number of alternative palladium catalysts have been developed for use in protein cross-coupling reactions that bear modified variants of the original ligand **L1**;[31,32,44] in particular its dimethylated analogue **L2** and "minimal" variants **L3** and **L4** (**figure 2.7**). For this system, the use of Pd(**L3**)$_2$(OAc)$_2$ and Pd(**L4**)$_2$(OAc)$_2$ catalysts led to precipitation of the protein, hence these were not pursued further, yet Pd(**L2**)$_2$(OAc)$_2$ afforded enhanced reaction rates in comparison to the analogous **L1**-based catalyst (**figure 2.8**).

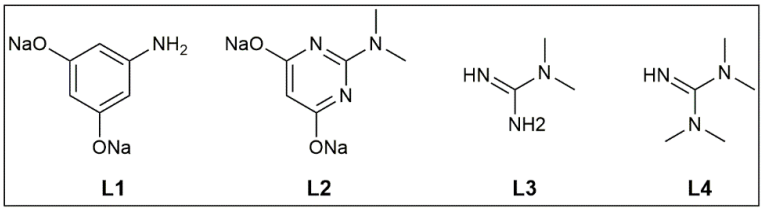


Figure 2.7 Alternative ligands utilised in Suzuki-Miyaura cross-coupling reactions to protein substrates.

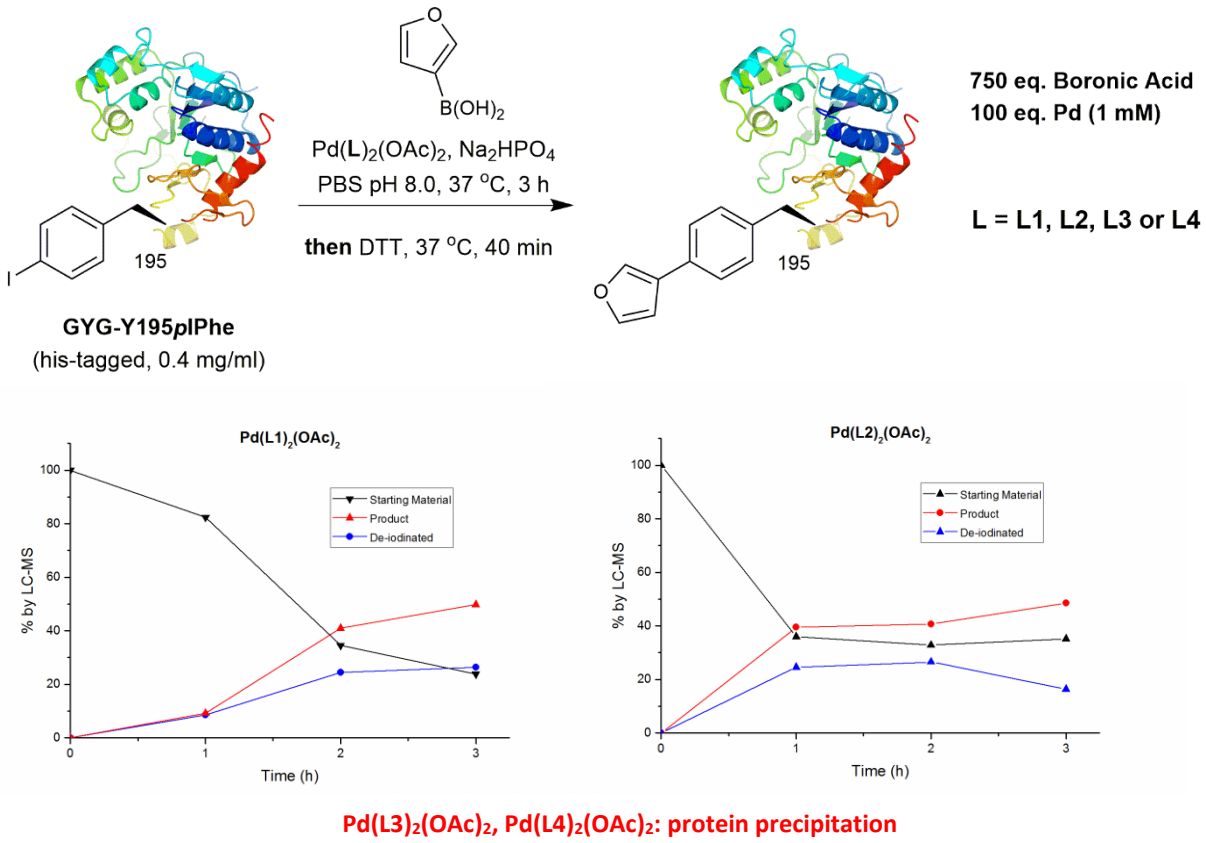
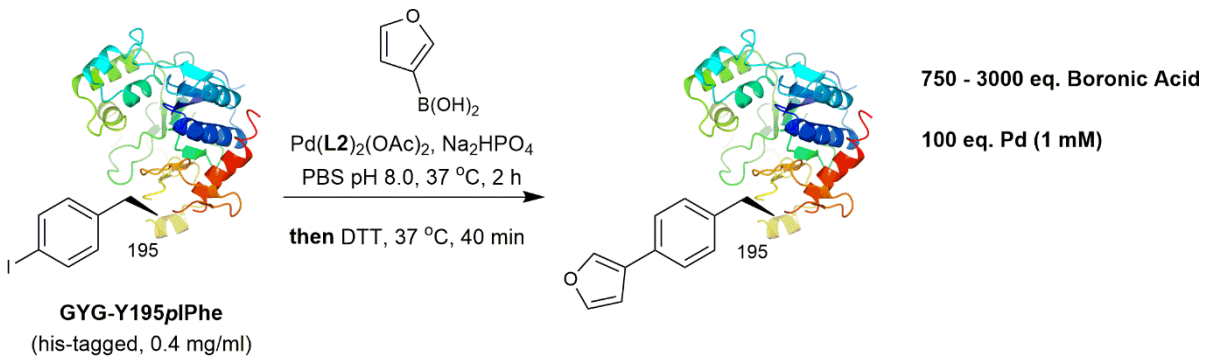


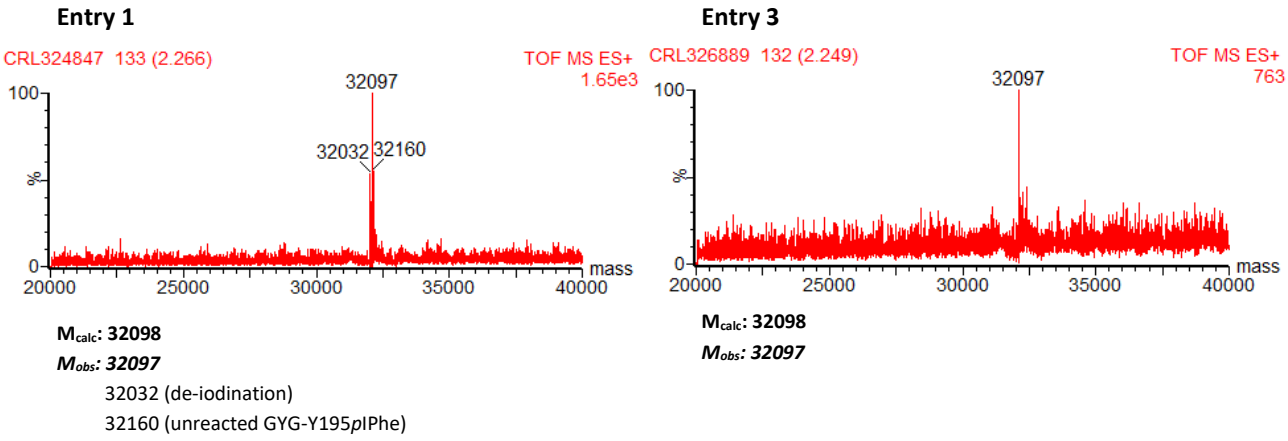
Figure 2.8 Comparison of different ligand systems for cross-coupling of GYG-Y195*p*IPhe to 3-furanylboronic acid.

An increase of boronic acid loading appeared to a slight increase in reaction conversion (**table 2.1** entries 1-2). This was a consequence of improved starting material consumption, consistent a role for the boronic acid in the reduction of the Pd(II) catalyst to the active Pd(0) species, as previously proposed.[30,43] As such, a further more substantial increase in boronic acid amount was utilised in an effort to further improve starting material consumption and reduce de-iodination (the latter is a result of lack of transmetalation to boronic acid), with stepwise addition in an effort to prevent any substrate exhaustion. This appeared to yield essentially quantitative conversion (**table 2.1** entry 3), although a concomitant drop in LC-MS quality reduced accuracy of estimated reaction conversions. Nonetheless, this informed subsequent studies.



Entry	Equivs. $B(OH)_2$	% Product	% Starting Material	% De-iodinated
1	750	**48**	26	26
2	1150	**52**	22	26
3	2250 (1200 + 3 x 350)	**100***	0*	0*

* *Reduced LC-MS quality at high boronic acid loadings may lower accuracy of analysis.*



M_{calc}: 32098
M_{obs}: 32097
32032 (de-iodination)
32160 (unreacted GYG-Y195*p*IPhe)

M_{calc}: 32098
M_{obs}: 32097

Table 2.1 Effect of boronic acid loading on cross-coupling to 3-furanylboronic acid. Representative LC-MS spectra are illustrated below.

2.4 GYG-3HMPB: a Putative "Minimal Glycomimetic"

The studies detailed above concerning cross-coupling to 3-furanylboronic acid were carried out first and foremost to investigate the inherent capacity of GYG-Y195*p*IPhe to undergo any cross-coupling at all. Having demonstrated that it indeed did represent a viable cross-coupling substrate, it was decided to shift focus towards a system of more direct relevance to the goal of "minimal GYG mimic" generation as

outlined previously in section 2.1.3. As previously discussed, it was considered that the cross-coupling of a hydroxyl-containing arylboronic acid can generate an enzyme containing a putative glycosyl acceptor at the correct site that could be considered a crude mimic of mono-glucosylated enzyme. We were interested to investigate whether such species indeed sufficiently mimic a native GYG glycoform to exhibit activity, and thus probe the "plasticity" of the enzyme.

One example of such a boronic acid is 3-(hydroxymethyl)phenylboronic acid (3-HMPB), whose cross-coupling to GYG-Y195*p*IPhe would yield an enzyme "GYG-3HMPB" bearing an aliphatic glycosyl acceptor, in common with that of glucose itself (**figure 2.9**). In light of other beneficial factors such as excellent water-solubility (> 15 mg/ml in our hands) and ready commercial availability, it was decided to utilise this substrate for further cross-coupling optimisation studies.

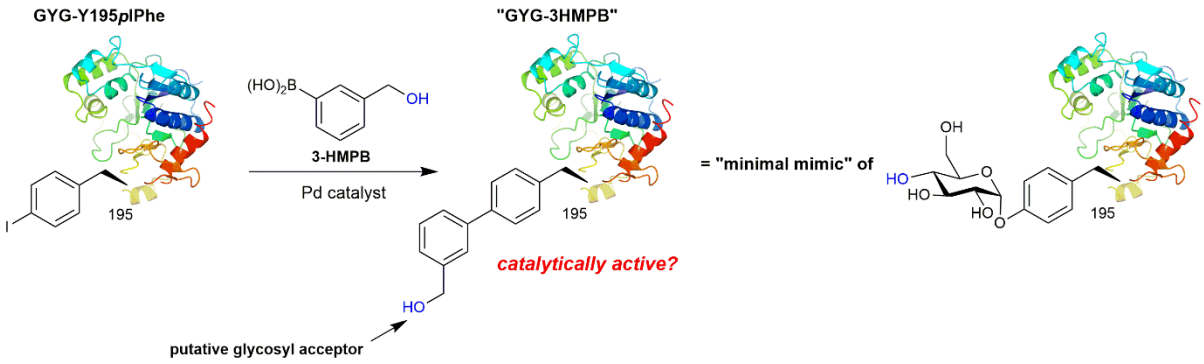


Figure 2.9 Cross-coupling of GYG-Y195*p*IPhe to commercially available boronic acid 3-HMPB could yield a "minimal mimic" of native mono-glucosylated GYG.

Initial studies on this cross-coupling were based on conditions utilised for the analogous 3-furanylboronic acid coupling involving high boronic acid loadings (see **table 2.1** entry 3). These had appeared to give high conversions, although the LC-MS quality was relatively poor. Pleasingly, when applied to cross-coupling between GYG-Y195*p*IPhe and 3-HMPB, LC-MS quality was significantly improved and confirmed high levels of reaction conversion to form the cross-coupled product (GYG-3HMPB), with no starting material detectable after 3 hours of reaction time (**figure 2.10**). As noted previously, a small amount of side-product attributed to a competitive de-iodination pathway was observed.

In a separate study, the effect of a further increase in boronic acid was investigated (**table 2.2**); an improvement in conversion was observed owing to reduction in de-iodination as would be expected (de-iodination results from lack of transmetalation with boronic acid), although this did plateau at high levels. In later experiments it became apparent that the further stepwise addition of boronic acid after the initial

addition was not in fact beneficial (see experimental section 2.8.5.5), and this was therefore ultimately abandoned. Evidently, any enhancements to conversion principally reflected differences in *initial* boronic acid loading, with at least 2000 eq. being of benefit.

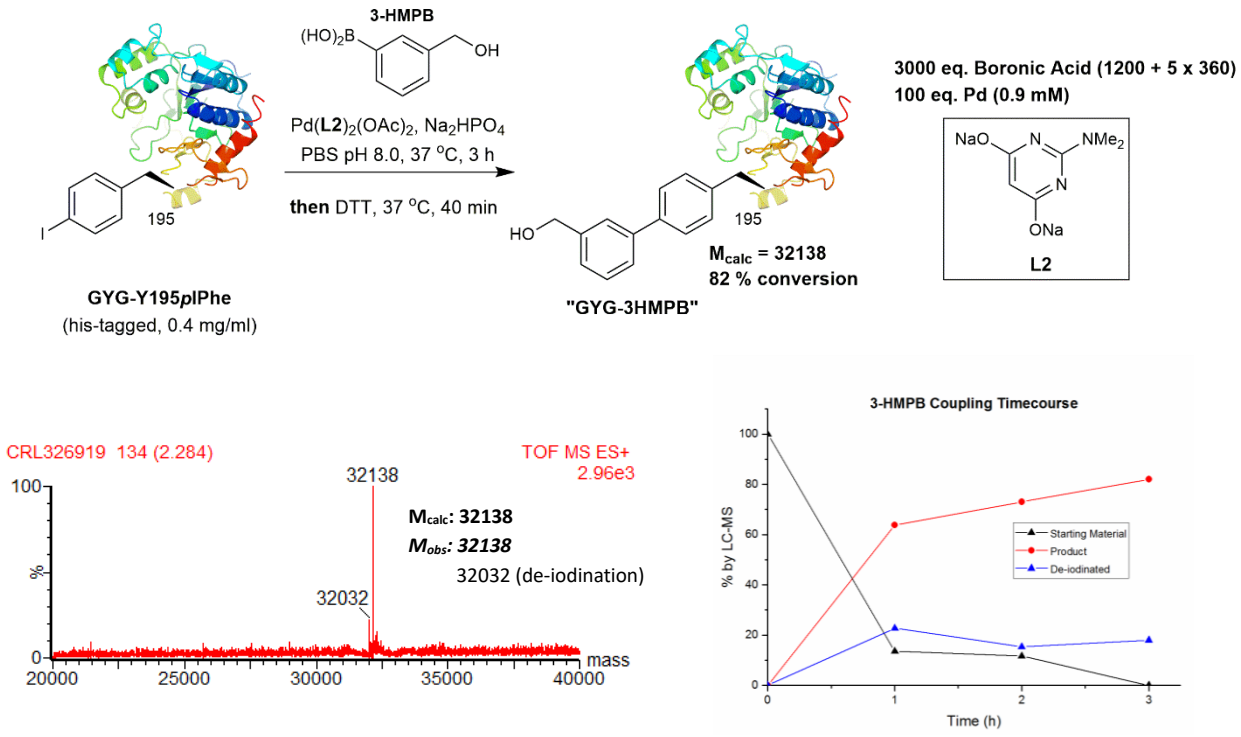
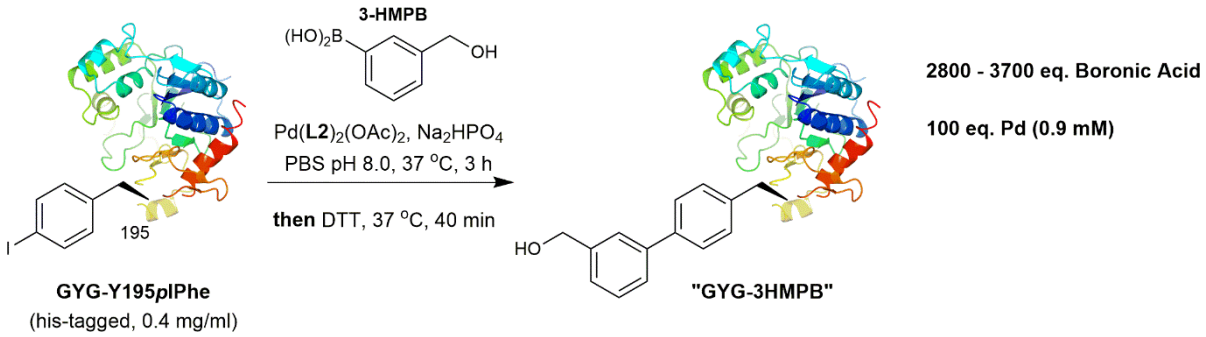


Figure 2.10 Initial cross-coupling between GYG-Y195*p*IPhe and boronic acid 3-HMPB.



Entry	Equivs. $B(OH)_2$	% Product	% Starting Material	% De-iodinated
1	2800 (1600 + 3 x 400)	**78**	0	22
2	3200 (2000 + 3 x 400)	**84**	0	16
3	3700 (2500 + 3 x 400)	**84**	0	16

Table 2.2 Optimisation of boronic acid loading for cross-coupling to 3-HMPB,

Consistent with the 3-furanylboronic acid system, cross-coupling to 3-HMPB did not proceed in the absence of Pd or boronic acid, further confirming a genuine cross-coupling process (see experimental section 2.8.5.6). Furthermore, treatment of the GYG-Y195F enzyme – which may be considered an "atomic level mutant" of GYG-Y195*p*IPhe in which the iodine is replaced with hydrogen – under the identical conditions utilised above used above did not lead to analogous reaction (experimental section 2.8.5.6), thus confirming formation of cross-coupling product at *p*IPhe.

Noticing that even at high boronic acid levels a low level of de-iodination persisted, we investigated other reaction parameters. However, neither an increase nor a decrease in reaction temperature were observed to be of benefit; surprisingly, increase appeared to reduce starting material consumption (**figure 2.11**). Similarly, alterations in pH caused little change to conversion, although based on a reduction in LC-MS quality it appeared that Pd-scavenging was more challenging at lower pH.[ii] An increase in protein concentration (maintaining other species at constant concentrations) also had no significant effect on de-iodination. However it was also not detrimental to reaction conversion despite the consequent reduced equivalents of all other reagents. Reasoning that this latter fact would most likely be of benefit with regards to maintaining the enzyme in an active state as desired, as well as being convenient for downstream sample handling, a higher initial protein concentration of 0.6 mg/ml was adopted from hereon in.

In spite of this persistent low-level de-iodination, the high conversions attained (80 – 85 %) were easily sufficient for the subsequent aim of evaluating whether the cross-coupled enzyme could indeed mimic the glycosyl acceptor capacity of glucosylated GYG, and thus undergo glucosylation. De-iodinated side-product corresponds to GYG-Y195F, which is known to be incapable of glucosylation owing to its lack of a correctly positioned glycosyl acceptor (its properties are discussed more in chapter 4).[34] As such, it should not interfere in any qualitative glucosylation assay; in a similar vein, neither should any remaining GYG-Y195*p*IPhe; this was subsequently seen to be the case (see section 2.6). Thus, whilst it is clearly desirable to achieve high SMC conversions, full conversion is not a pre-requisite in this system. Indeed, it is arguably of greater importance that the SMC conditions themselves do not inactivate the enzyme. Our studies in this latter area are discussed in the following section.

[ii] This presumably reflects greater proportion of thiol vs. thiolate at lower pH, and reduced ability of the former to coordinate Pd.

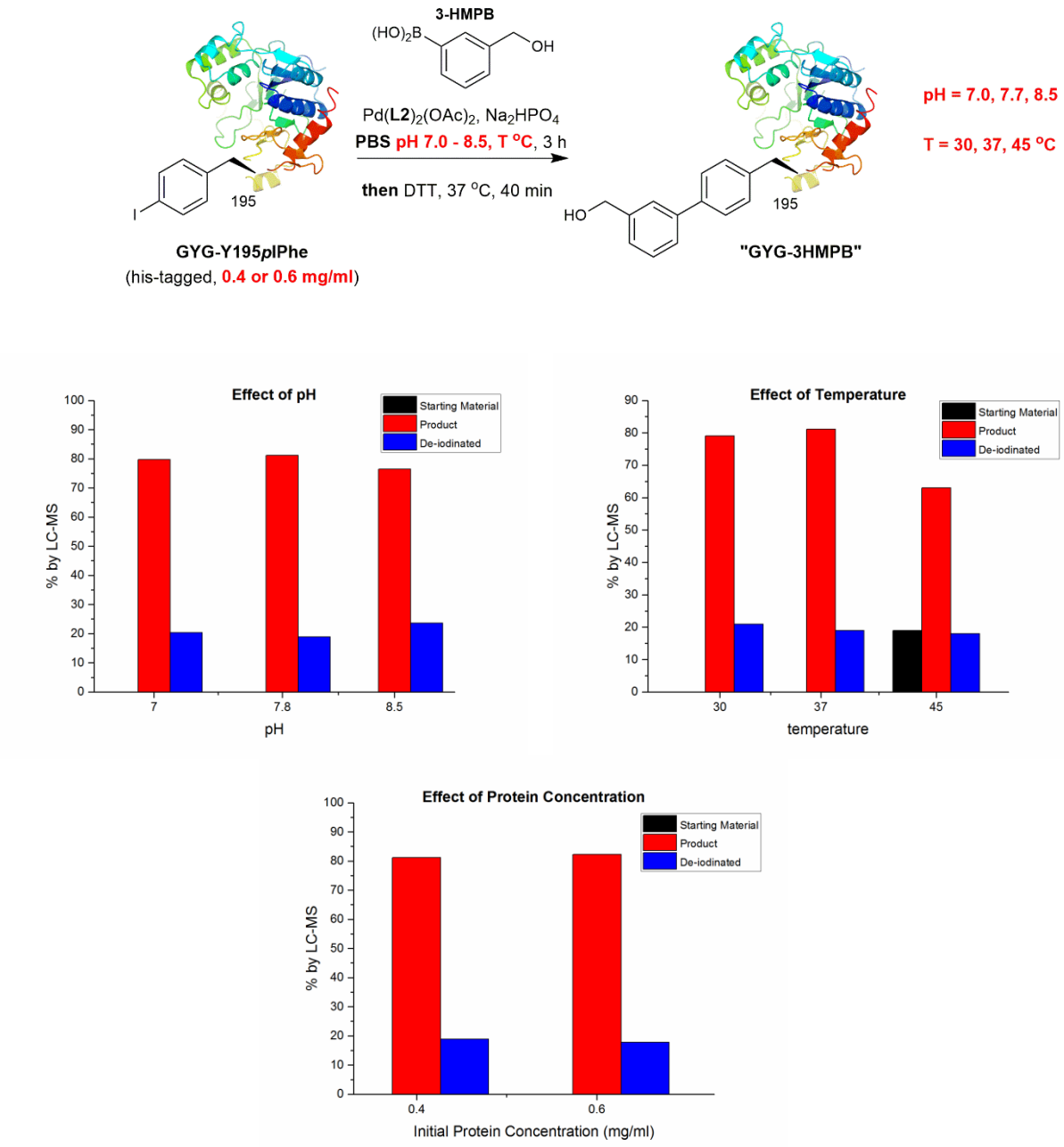


Figure 2.11 Effect of pH, temperature and protein concentration variation on cross-coupling between GYG-Y195*p*IPhe and 3-HMPB. For the latter case, it should be noted that all other reagents were maintained at constant concentrations as opposed to equivalents.

2.5 Development of an Enzyme-Compatible Cross-Coupling

Before any attempt could be made in order to assess the activity or otherwise of GYG-3HMPB and any similar constructs, it was first necessary to ensure that the cross-coupling conditions themselves did not inactivate the enzyme. In order to study this, we turned to the previously expressed wild-type enzyme (GYG-wt) for use as a positive control. LC-MS of this enzyme before and after autoglucosylation assay (involving incubation with 1 mM each of cofactor Mn^{2+} and substrate UDP-Glucose) confirmed its intrinsic activity (**figure 2.12**). The figure shows a graphical representation of the LC-MS spectra (see experimental section 2.8.6.2), illustrating the proportions of each GYG glycoform present prior to and following glucosylation assay. These are determined from their relative MS peak intensities in LC-MS spectra; average numbers of glucoses added per enzyme can also be determined from these (see experimental section 2.8.6.1 for full details). Note that all similar graphs in subsequent figures represent analogous analyses of LC-MS spectra (shown in the experimental section) before and after glucosylation assays according to the protocol detailed in section 2.8.6.1. A clear shift towards higher glycoforms is seen below, with all glycoforms present in the starting enzyme undergoing further glucosylation

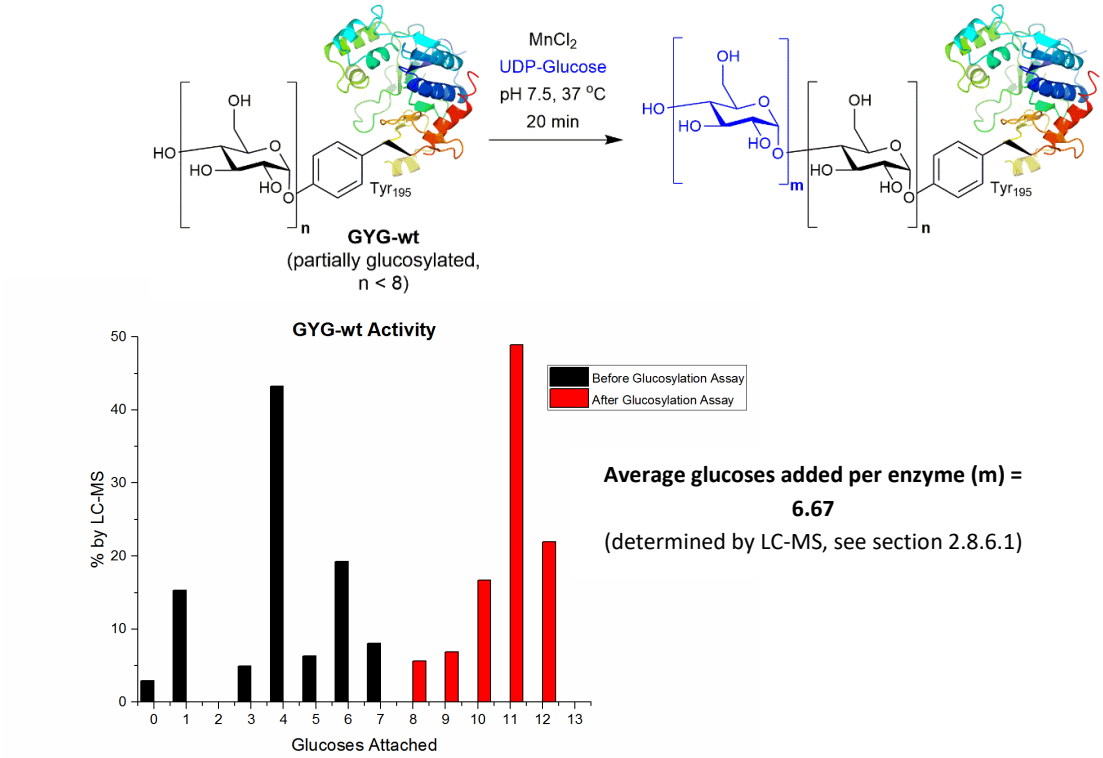


Figure 2.12 Catalytic activity of GYG-wt. The graph illustrates proportions of different GYG glycoforms present before (black) and after (red) glucosylation assay, as determined by LC-MS analysis. Average glucoses added per enzyme (m) was determined from LC-MS spectra (see section 2.8.6.2) as described in section 2.8.6.1.

However, to our disappointment, a sample of the same enzyme pre-treated with coupling conditions utilised for 3-HMPB exhibited very little glucosylation activity (**figure 2.13**), as again indicated by LC-MS (the same glycoforms were present before and after glucosylation assay).

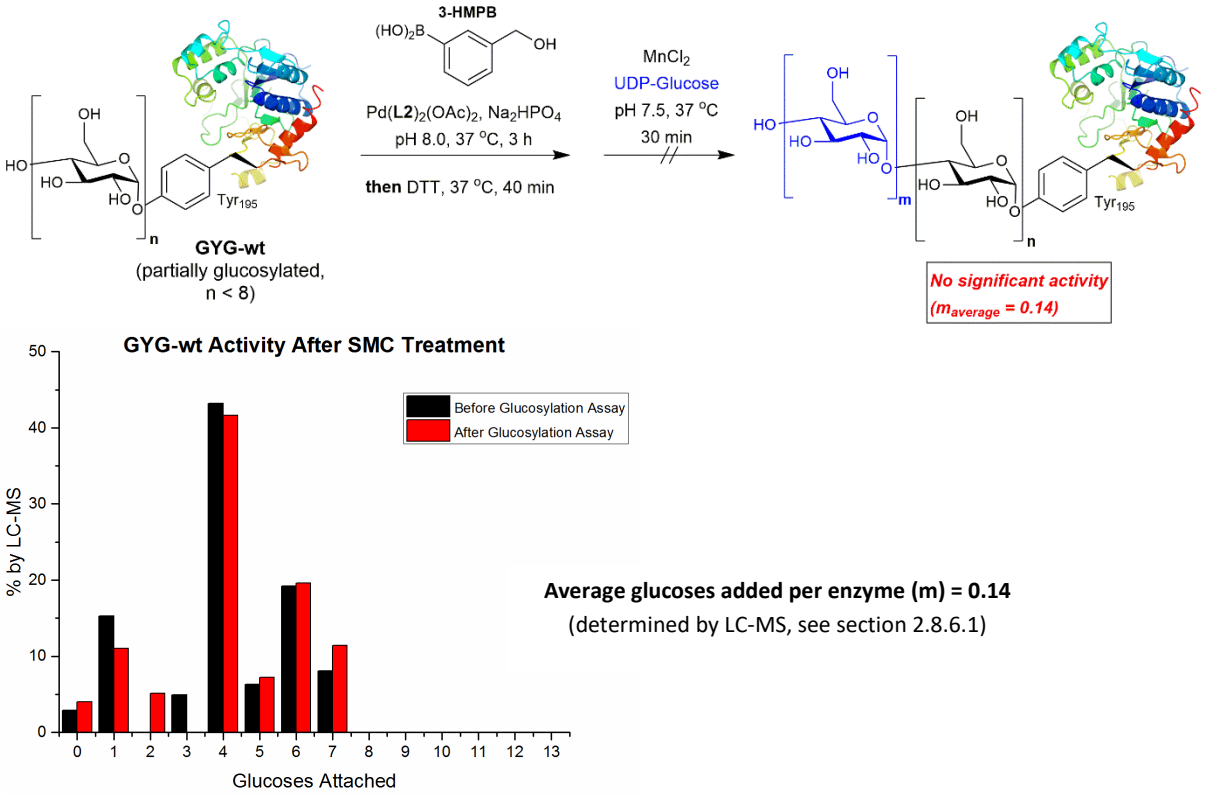
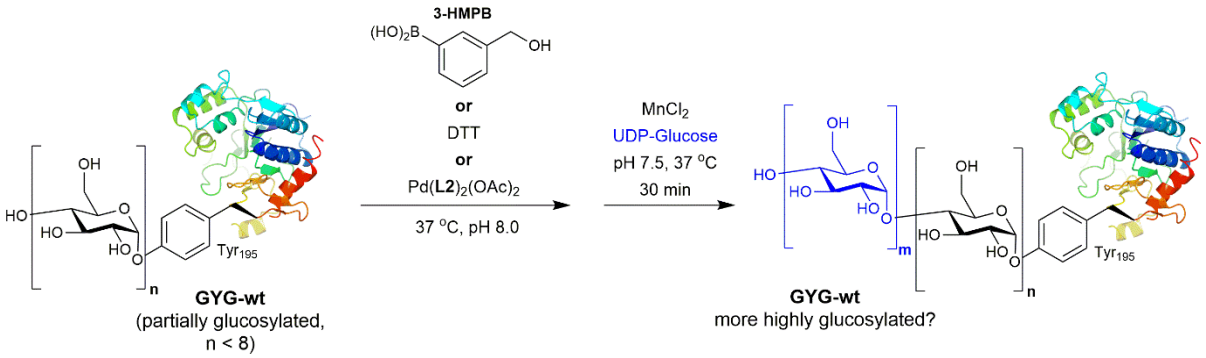


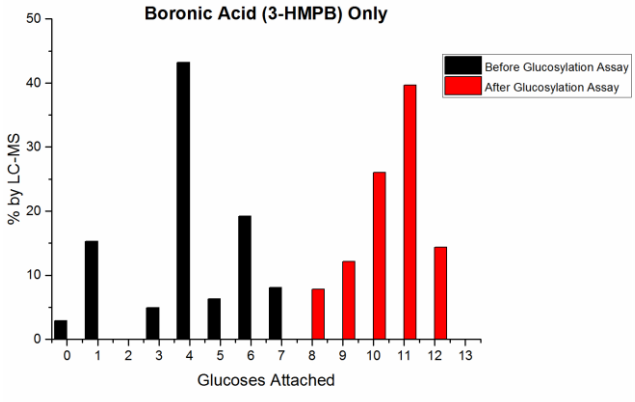
Figure 2.13 Catalytic activity of GYG-wt, after exposure to SMC conditions utilised for cross-coupling between GYG-Y195*p*IPhe and 3-HMPB. The graph illustrates proportions of different GYG glycoforms present before (black) and after (red) glucosylation assay, as determined by LC-MS analysis. Very limited glucosylation is apparent. Average glucoses attached per enzyme were determined from LC-MS spectra (see sections 2.8.6.1 and 2.8.6.3).

Individual cross-coupling reaction components were investigated to determine which of these caused this apparent enzymatic inactivation. To this end, samples of GYG-wt were pre-treated with either boronic acid 3-HMPB, palladium scavenger DTT, or palladium catalyst with DTT (in this latter case, DTT is required for LC-MS analysis), then assayed for autoglucosylation activity by LC-MS as previously outlined (**figure 2.14**, see experimental section 2.8.6.4). The samples were run in parallel to each other and to the sample outlined in **figure 2.13** in order to ensure, as far as possible, equivalence of the samples with respect to all factors other than those being intentionally varied.

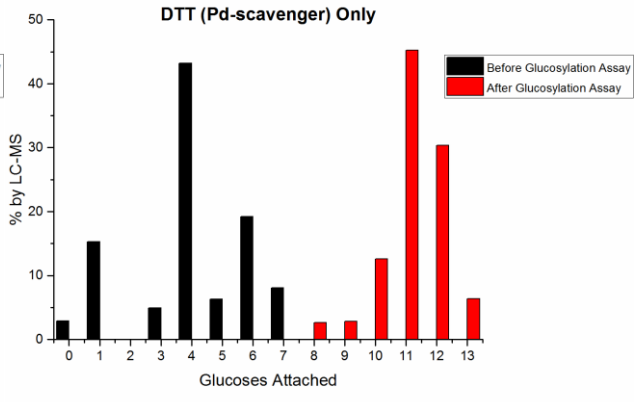
Pre-treatment of GYG-wt with either palladium scavenger DTT or boronic acid 3-HMPB in isolation did not cause enzymatic inactivation (**figure 2.14**, note that all glycoforms present in the starting material are entirely consumed). However, similar treatment with palladium catalyst and DTT seemingly caused

significant inhibition of activity; glycoforms present in the starting material were not entirely consumed upon glucosylation, and overall extent of glucosylation was markedly lower.

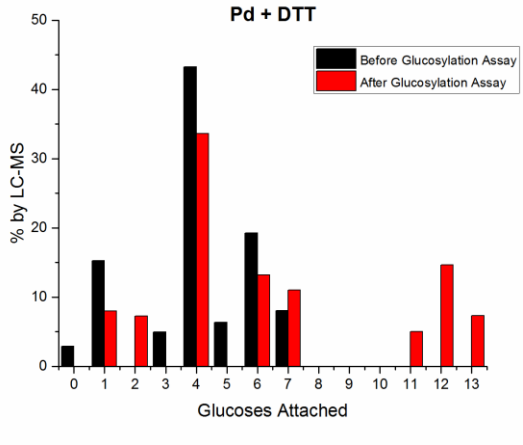




Average glucoses added (m) = 6.34



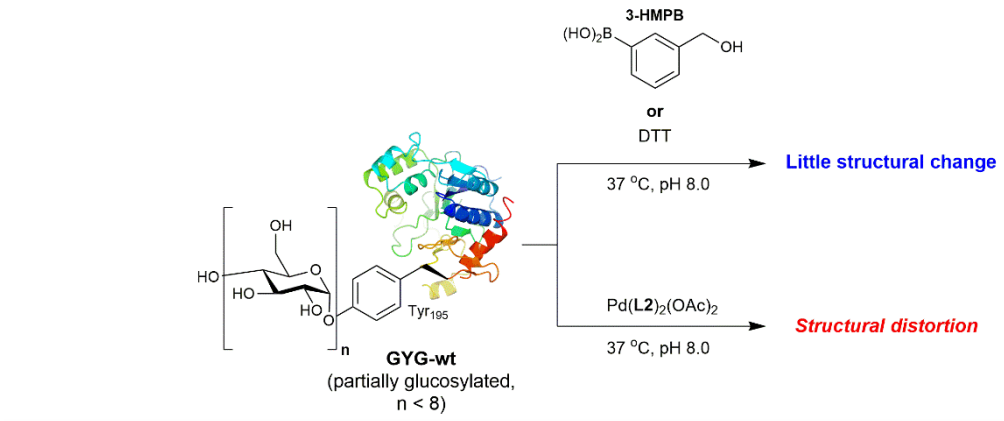
Average glucoses added (m) = 7.10



Average glucoses added (m) = 2.32

Figure 2.14 Effect of pre-treatment with SMC reaction components on enzymatic activity of GYG-wt. Graphs show proportion of GYG glycoforms before (black) and after (red) glucosylation assay, as determined by LC-MS analysis. Average glucoses added (m) determined from LC-MS spectra (see experimental sections 2.8.6.1 and 2.8.6.4).

This suggested to us that the palladium catalyst was predominantly responsible for the loss of activity noted in **figure 2.13**. Consistent with these observations, circular dichroism analysis revealed little structural change upon treatment of GYG-wt with DTT or boronic acid alone, but gross structural distortion after treatment with Pd catalyst (**figure 2.15**).



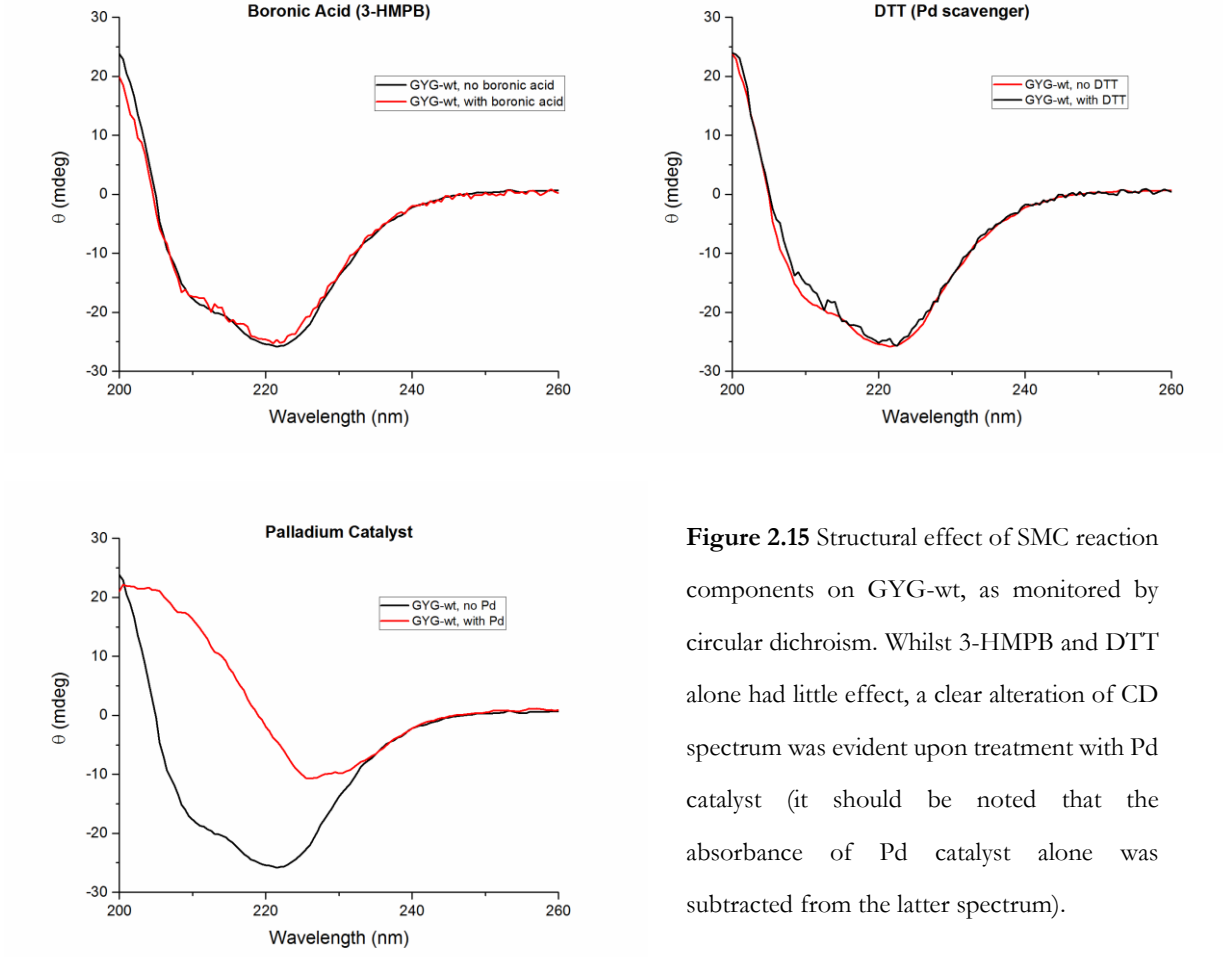


Figure 2.15 Structural effect of SMC reaction components on GYG-wt, as monitored by circular dichroism. Whilst 3-HMPB and DTT alone had little effect, a clear alteration of CD spectrum was evident upon treatment with Pd catalyst (it should be noted that the absorbance of Pd catalyst alone was subtracted from the latter spectrum).

We further probed the detrimental effect of the palladium catalyst through co-incubation of non-treated GYG-wt with UDP-Glucose, Mn^{2+} and varying concentrations of Pd. This revealed a palladium concentration-dependent reduction in autoglucosylation extent of the enzyme (**figure 2.16**), as determined through LC-MS analysis (section 2.8.6.7). Whilst at 0.1 mM Pd, little change was seen, moderate inhibition was noted at 0.2 mM Pd, and significant reduction in activity at 0.4 mM Pd. This may represent a competition between Pd^{2+} and the natural Mn^{2+} cofactor, or could instead simply arise from any Pd-

mediated structural distortion being more rapid at higher catalyst concentrations. In either case, it is evident that any such "residual" catalyst remaining after cross-coupling may inhibit subsequent glucosylation. In combination with the prior results concerning pre-treatment with Pd (see **figure 2.14**), this implied a need to not only minimise Pd loadings utilised during the SMC step, but also to ensure thorough post-SMC catalyst removal prior to glucosylation assay.

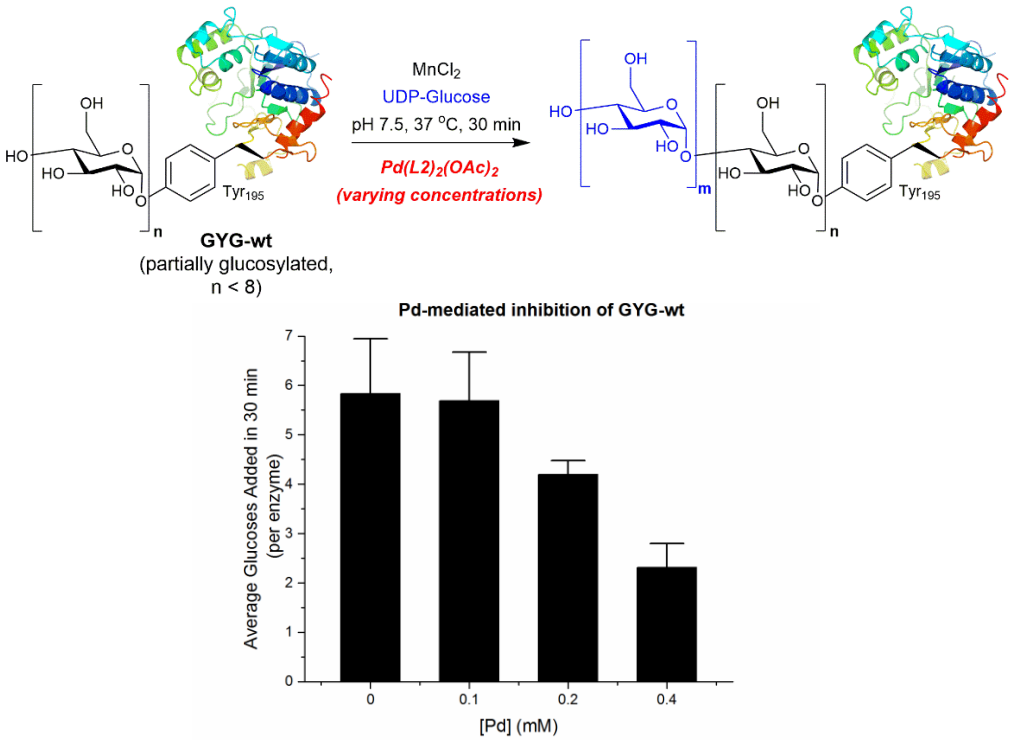


Figure 2.16 Inhibitory effect of Pd(**L2**)$_2$(OAc)$_2$ catalyst on GYG-wt glucosylation, as judged by extent of glucosylation (m) at different Pd concentrations; this was determined through LC-MS analysis according to the method outlined in section 2.8.6.1 (see section 2.8.6.7 for spectra). Error bars represent s.d. of independent triplicate samples.

Gratifyingly, treatment of a sample of GYG-wt with modified SMC conditions for boronic acid 3-HMPB, key alterations being reduced Pd concentration (0.8 mM) and the addition of more extended post-reaction Pd-removal/refolding steps, led to preservation of the enzyme in a largely active state as determined through LC-MS analysis (**figure 2.17**, see experimental section 2.8.6.8). Glycoforms present in the starting material prior to glucosylation were largely consumed, with concomitant formation of new, more highly glucosylated variants. Circular dichroism also confirmed a now largely unaltered secondary structure with respect to non-treated GYG-wt. It should be noted that concentrations of the Pd scavenger DTT were also ultimately lowered (from 12 mM to 8 mM); additional experiments involving treatment of GYG-wt with varying levels of DTT with or without other reagents suggested that this was of benefit with respect to enzymatic activity (experimental section 2.8.6.5), and 8 mM proved easily sufficient for Pd scavenging going forward (as judged by the fact that protein peaks remained readily detectable upon LC-MS analysis).

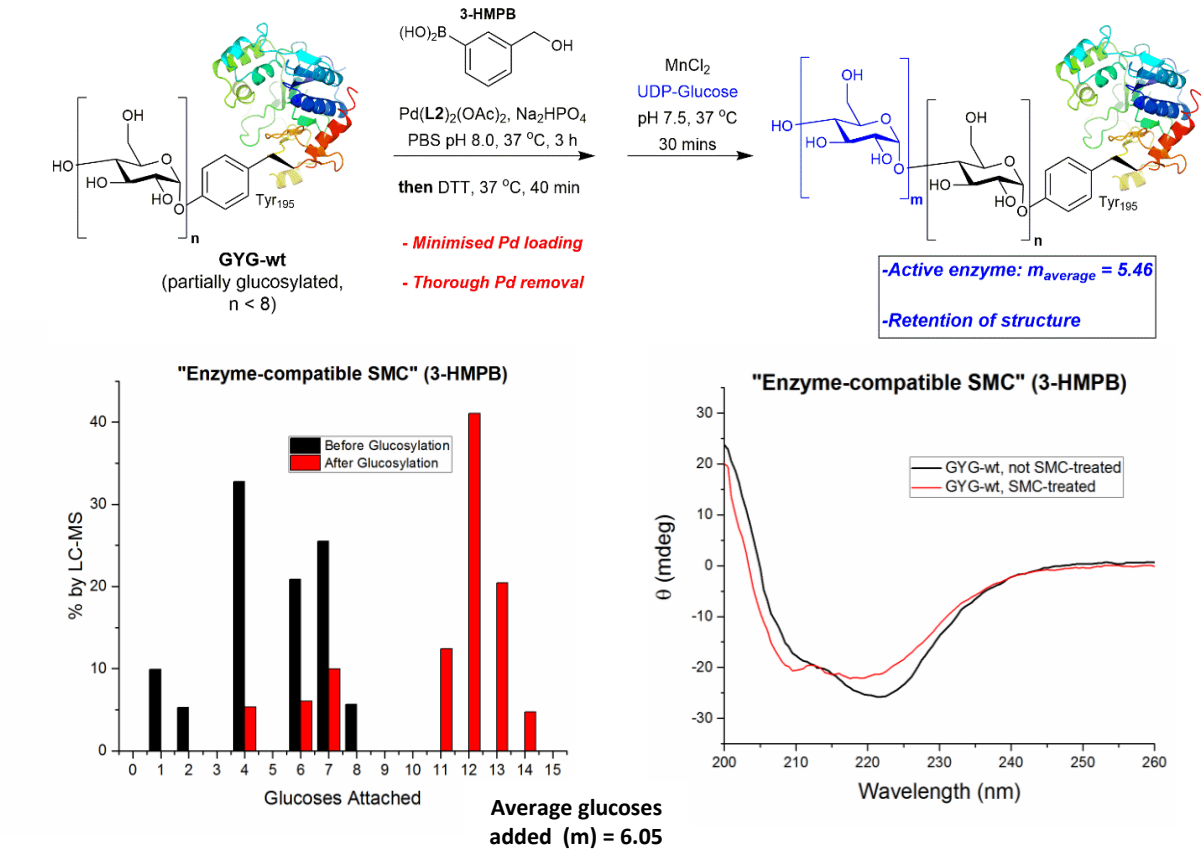


Figure 2.17 "Enzyme-compatible" SMC conditions for 3-HMPB. Treatment of GYG-wt under conditions involving lower Pd loadings and thorough Pd removal led to reproducible retention of catalytic activity (left) and largely unaltered secondary structure (right). Graphs show proportion of GYG glycoforms before (black) and after (red) glucosylation assay, as determined by LC-MS analysis (see section 2.8.6.8) using the method outlined in section 2.8.6.1. Average glucoses added (m) determined from LC-MS spectra.

2.6 Investigating the Activity of Minimal GYG Glycomimetics

Having developed an SMC protocol for the model system of 3-HMPB that not only afforded high reaction conversions but was also compatible with retention of GYG enzymatic structure and function, we were now in a position to investigate whether the cross-coupled enzyme, despite being a crude "minimal" mimetic of glucosylated GYG, nonetheless could undergo glucosylation. As such, the requisite cross-coupled enzyme "GYG-3HMPB" was generated from GYG-Y195*p*IPhe based on the "enzyme-compatible" conditions developed above and, following thorough Pd-removal, was subjected to glucosylation assay. Pleasingly, these SMC conditions remained compatible with high reaction conversions (77 %, **figure 2.18**). However, no glucosylation could be observed by LC-MS analysis. As a positive control, a sample of GYG-wt was treated under the same SMC conditions, then assayed for autoglucosylation

activity. This corresponds to a reproduction of the experimental conditions in **figure 2.17**, however it was important to repeat this here in parallel to the GYG-Y195*p*IPhe sample so as to ensure exact equivalence of conditions across both.[iii] This control enzyme retained good activity (**figure 2.18**), further confirming that inactivity of the cross-coupled enzyme did not reflect reaction conditions.

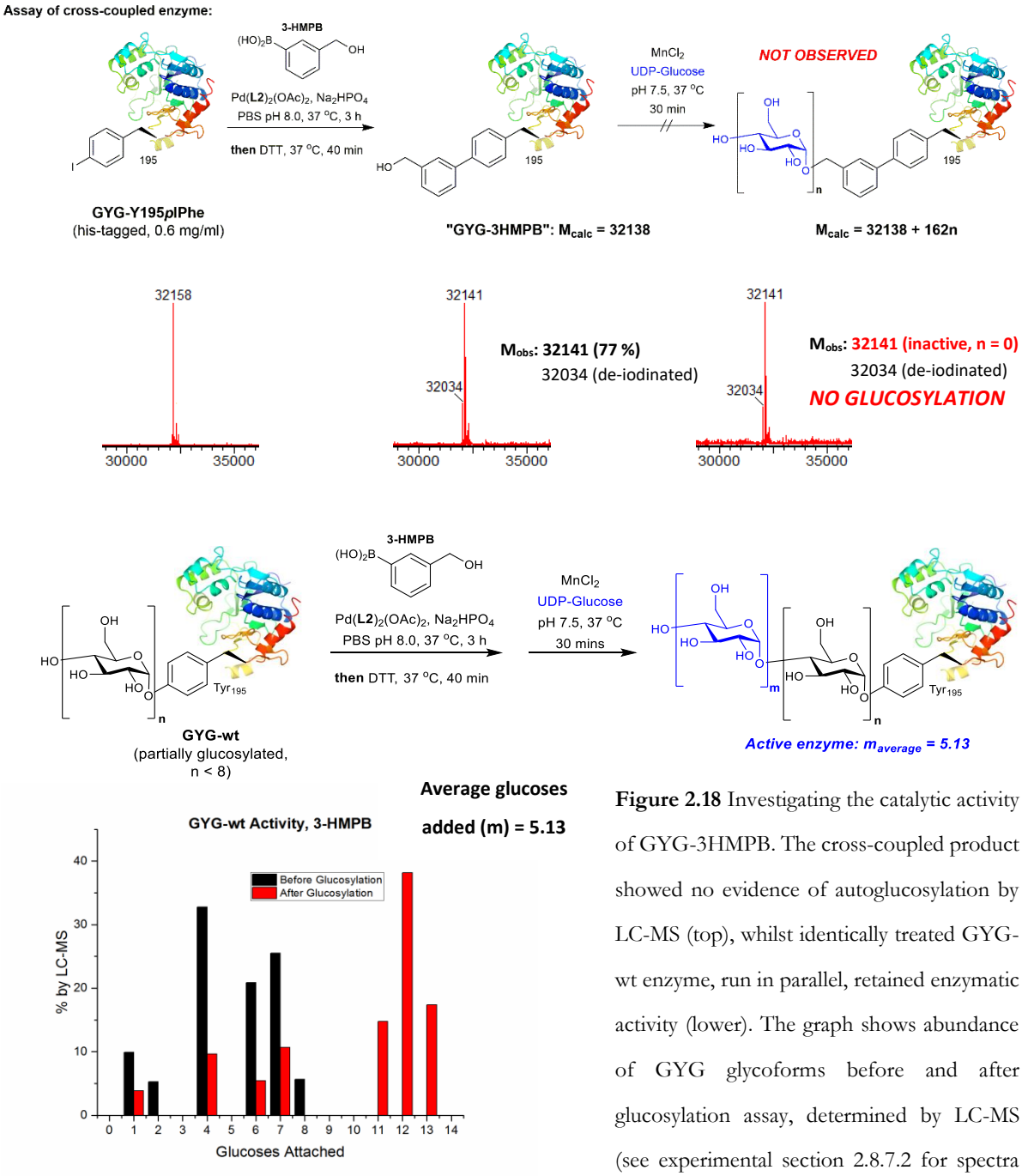


Figure 2.18 Investigating the catalytic activity of GYG-3HMPB. The cross-coupled product showed no evidence of autoglucosylation by LC-MS (top), whilst identically treated GYG-wt enzyme, run in parallel, retained enzymatic activity (lower). The graph shows abundance of GYG glycoforms before and after glucosylation assay, determined by LC-MS (see experimental section 2.8.7.2 for spectra and 2.8.6.1 for analytical method).

[iii] For example, **figures 2.17** and **2.18** do not utilise the same batch of Pd catalyst; it was important to confirm this had no effect on enzymatic activity. The graphs in **figures 2.17** and **2.18** thus represent independent replicates.

It was considered that the orientation of the hydroxyl group in GYG-3HMPB may be incorrect with respect to the bound UDP-Glucose donor. Alternatively, given that the enzymatic mechanism demands a proton transfer from glycosyl acceptor to the leaving group (UDP-Glucose),[34] an increase in acidity of the glycosyl acceptor may be of benefit. As such, the translation of the system developed for 3-HMPB to other structurally similar boronic acids – 2-HMPB, 4-HMPB, 3-HPB and 4-HPB (**figure 2.19**) was investigated in order to generate analogues differing with respect to orientation and/or acidity of the putative glycosyl acceptor.

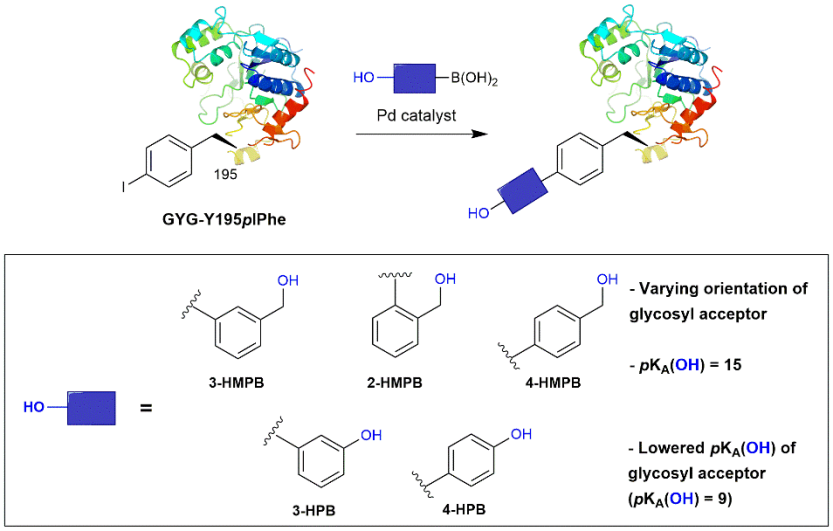
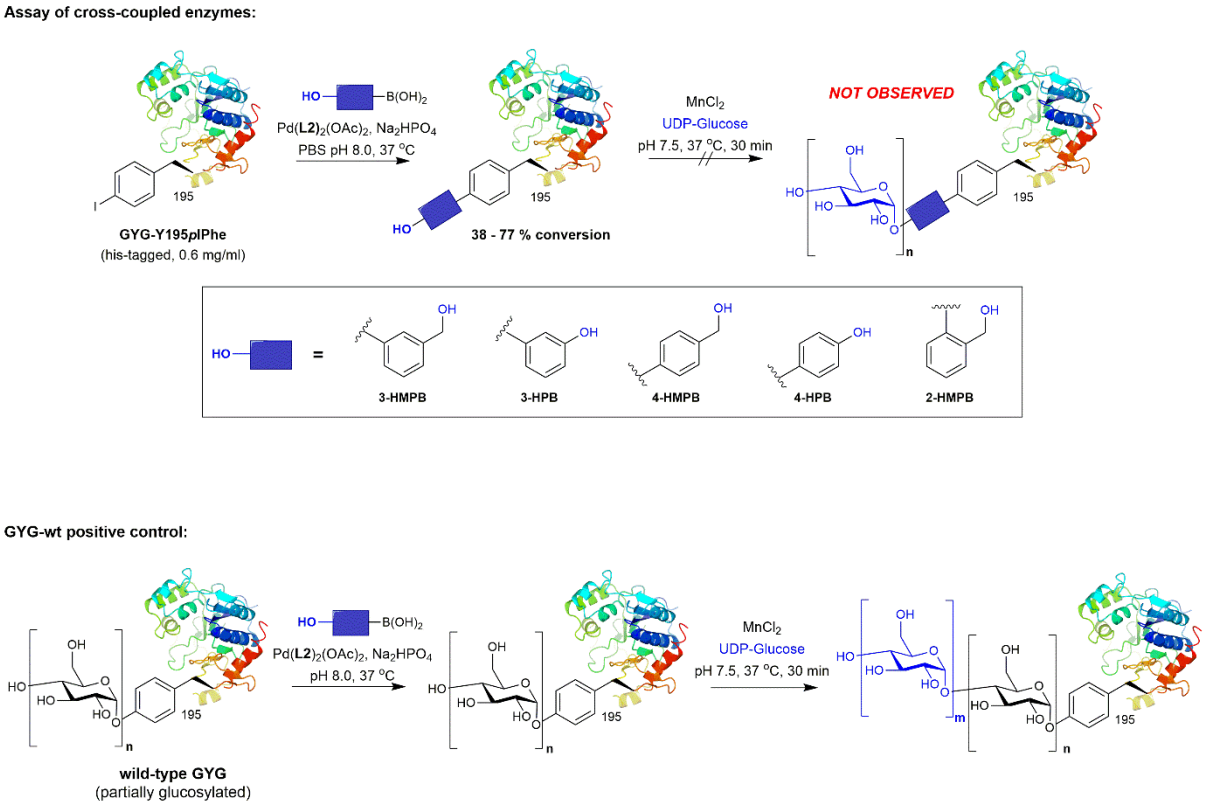


Figure 2.19 Variation of boronic acid can yield GYG glycosyl acceptors of differing orientations and/or acidity.

Results concerning the generation and activity of these analogues are tabulated below (**table 2.3**, see experimental section 2.8.7), alongside the prior results concerning GYG-3HMPB (**table 2.3** entry 1). Cross-coupling to 3-HPB proceeded with similar conversion to that of 3-HMPB under identical conditions (67 %, **table 2.3** entry 2). Analogous coupling to 2-HMPB proved less efficient but could be effected in moderate levels sufficient for subsequent study (45 %) under slightly higher Pd concentrations of 0.9 mM (**table 2.3** entry 3). Indeed, cross-couplings to such *ortho*-subsituted arylboronic acids are known to be challenging, likely due to increased steric hindrance.[45] However, in neither case could glucosylation activity be detected by LC-MS (experimental sections 2.8.7.3, 2.8.7.4).

For both these boronic acids, a positive control sample of identically treated GYG-wt, run in parallel to the test sample to ensure equivalence of conditions, retained at least moderate activity. Full graphical data for the activity of all such GYG-wt samples is additionally shown in **figure 2.20** (except that for 3-HMPB, which was shown previously in **figure 2.18**). For 3-HPB, clear glucosylation was observed with almost

complete consumption of glycoforms present in the starting material. In the case of 2-HMPB, this activity was reduced – consistent with the higher Pd loadings employed – yet still clearly apparent, since glycoforms not present in starting material were formed (**figure 2.20**). Thus, the lack of glucosylation of cross-coupled enzyme did not appear to reflect reaction conditions in either case.



Entry	Boronic Acid	Cross-coupling conversion[a]	Activity of cross-coupled enzyme	Activity of GYG-wt positive control, run in parallel (see also figure 2.20)
1	3-HMPB	77 %	None	Good (5.13 Glc added in 30 min)
2	3-HPB	67 %	None	Good (7.14 Glc added in 30 min)
3	2-HMPB[b]	45 %	None	Moderate (3.05 Glc added in 30 min)
4	4-HMPB[c]	35 %	None	Low (0.81 Glc added in 30 min)
5	4-HPB[c]	38 %	None	Low (1.19 Glc added in 30 min)

[a] *As determined by LC-MS.*

[b] *Carried out using a higher Pd loading of 0.9 mM in order to attain higher reaction conversions.*

[c] *Further improvement of coupling efficiency not attempted owing to incompatibility of GYG-wt to reaction conditions.*

Table 2.3 Generation and activity of various "minimal mimics" of glucosylated GYG. In all cases, no activity of cross-coupled enzyme was seen. Identically treated wild-type enzyme run in parallel retained at least some activity in all cases, although for 4-HMPB and 4-HPB this was very limited in extent and not readily discernible.

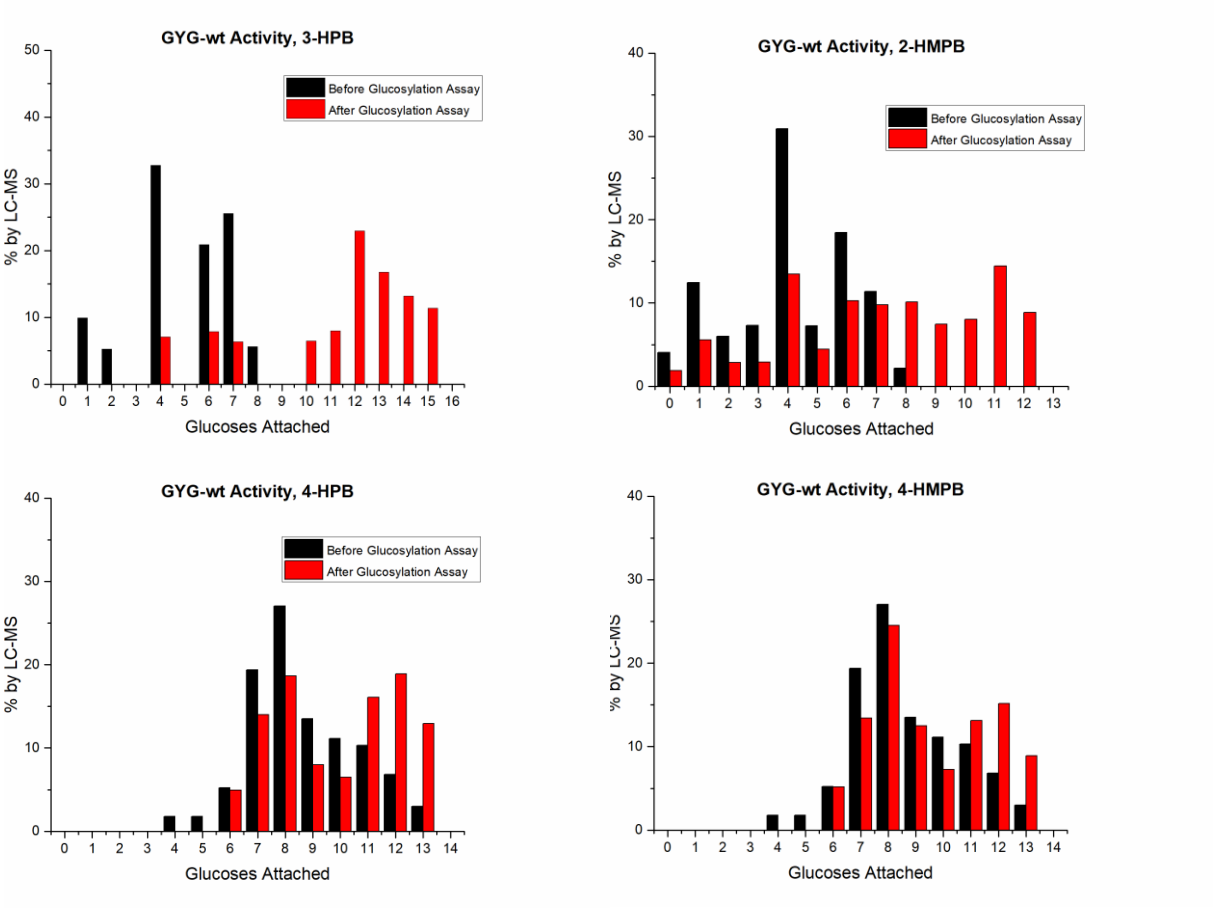


Figure 2.20 Graphical data for activities of GYG-wt positive control samples from **table 2.3** entries 2-5 (that from entry 1 was shown in **figure 2.18**). Graphs show abundances of GYG glycoforms before and after glucosylation assay, as determined by LC-MS (see section 2.8.7 for spectra, 2.8.6.1 for analytical method). Limited glucosylation is apparent for the 4-HMPB and 4-HPB cases, in contrast to 3-HPB and 2-HMPB.

In contrast, attempts to evaluate the 4-HMPB and 4-HPB systems (**table 2.3** entries 4-5) proved more challenging. Not only were cross-coupling efficiencies poor (35-38 %) under conditions identical to those optimised for 3-HMPB, but GYG-wt treated with these same SMC conditions, except for the identity of the boronic acid, exhibited poor levels of activity (see experimental sections 2.8.7.5, 2.8.7.6). The fact that glycoforms not already present in the starting material were *not* formed rendered this activity less evident than for prior systems (**figure 2.20**). As such, whilst the cross-coupled products GYG-4HMPB and GYG-4HPB did not undergo glucosylation as judged by LC-MS (experimental sections 2.8.7.5, 2.8.7.6), this may well reflect the reaction conditions used to generate them.

A separate experiment indicated that this surprising intolerance of GYG-wt to the reaction conditions for 4-HMPB and 4-HPB held true across a range of palladium concentrations equal to and lower than those utilised in prior systems (see experimental section 2.8.8.2). It may be noted that the sample of GYG-wt

utilised in these systems was more highly glucosylated following expression than prior batches – this presumably reflected unavoidable batch-to-batch variations of cell strain – yet this seemingly did not impede its ability to serve as a positive control, since unambiguous glucosylation was detectable in absence of cross-coupling conditions (experimental section 2.8.8.1).

2.7 Conclusion

As part of our overall aim of generating homogeneous GYG glycoform mimetics for subsequent mechanistic studies, in this chapter we set out to explore the viability of Suzuki-Miyaura cross-coupling (SMC) as a strategy in this regard. The feasibility of cross-coupling to GYG with respect to both product formation and yield, and compatibility with subsequent enzymatic function, was investigated. The unnatural amino acid *p*IPhe was successfully installed in a site-specific fashion into mammalian GYG-1 in place of the native glycosyl acceptor, Tyr195, generating an inactive enzyme, GYG-Y195*p*IPhe, bearing a reactive tag for subsequent SMC. Successful cross-coupling of this enzyme was demonstrated and optimised, enabling high conversions to be attained under "enzyme-compatible" conditions within a model system. Whilst the conditions developed using this system were not universally enzyme-compatible for all boronic acids, they proved to be so for the vast majority, as will become evident in the following chapters.

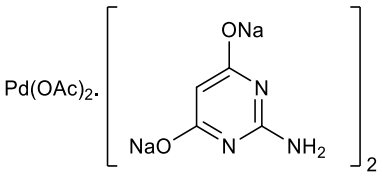
We investigated whether cross-coupled enzymes bearing simple biaryl-based glycosyl acceptors at the 195 site, with pendant hydroxyl groups, were capable of undergoing any glucosylation; it was considered that these may represent crude, "minimal mimetics" of mono-glucosylated GYG. However, no such activity was observed for any species tested, in spite of varying orientation and acidity of the glycosyl acceptor. In two cases, this was attributed to an intolerance of the enzyme towards the reaction conditions themselves across a range of Pd concentrations (reasons for this remain unclear, although it may be of relevance that, in contrast to other substrates, 4-HMPB was water-insoluble whilst 4-HPB was visually observed to degrade in aqueous solution, consistent with prior data[30]). However in other cases this was not the case, implying that the enzyme did not exhibit sufficient plasticity to effect glucosylation from these acceptors.

We hypothesised, therefore, that these cross-coupled species lacked sufficient structural resemblance to natural GYG glycoforms to exhibit glucosylation activity, and that increasing such resemblance could yet afford catalytically active enzymes. The application of Suzuki-Miyaura cross-coupling to the generation of a more accurate GYG glycoform mimetic forms the basis of the next chapter.

2.8 Experimental Section

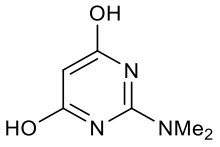
2.8.1 Chemical Synthesis

Palladium Catalysts



$Pd(OAc)_2(L1)_2$

In accordance with the protocol previously reported by our group,[26] 2-amino-4,6-dihydroxypyrimidine (**L1**, 12.7 mg, 0.1 mmol) was dissolved in 0.1M NaOH (2 ml, 0.2 mmol) and heated to 65 °C. $Pd(OAc)_2$ (11.2 mg, 0.05 mmol) was added and the mixture stirred for 30 minutes, after which time all solid dissolved. After cooling to room temperature, the mixture was diluted with 3 ml H_2O to give a total of 5 ml of 10 mM yellow catalyst solution, which was stored in a sealed vial at room temperature.



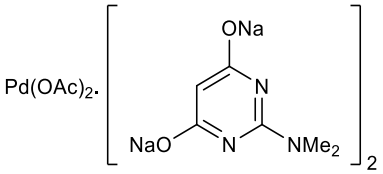
Ligand L2

N,*N*-dimethylguanidine sulphate (500 mg, 1.8 mmol) and diethyl malonate (0.56 ml, 3.7 mmol) were suspended in anhydrous MeOH (5 ml) and NaOMe (2.6 ml of 25 wt % in MeOH, 11.3 mmol) was added. The mixture was stirred for 16 h at r.t. under N_2, after which time it was evaporated to dryness *in vacuo*. The resultant white solid was taken up in H_2O (10 ml) and acidified to ca. pH 5 with 1M HCl, causing the product (ligand **L2**) to precipitate as a white solid; this was collected by filtration and dried in vacuum dessicator over Na_2SO_4. Yield: 280 mg, 1.80 mmol, quant.

^{1}H NMR (400 MHz, DMSO) δ 10.49 (s, 2H, OH), 4.67 (s, 1H, Ar-H), 3.02 (s, 6H, CH3)

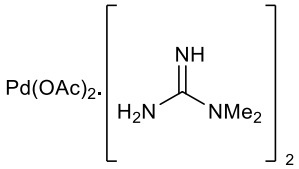
LRMS *m/z* (ESI+) found: 200.2 (M+2Na-H, calculated 200.2), 222.0 (M+3Na-2H, calculated 222.2)

m.p. >300 °C (lit.[46] >300 °C). Data consistent with that previously reported.[47]



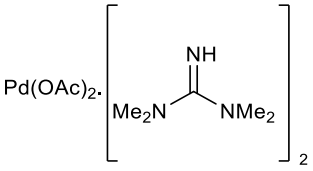
$Pd(OAc)_2(L2)_2$

Based on protocols previously reported by our group,[32,48] ligand **L2** (15.5 mg, 0.1 mmol) was dissolved in 0.05M NaOH (4 ml, 0.2 mmol) and heated to 65 °C. $Pd(OAc)_2$ (11.2 mg, 0.05 mmol) was added and the mixture stirred for 30 minutes, after which time all solid dissolved. After cooling to room temperature, the mixture was diluted with 1 ml H_2O to give a total of 5 ml of dark orange/brown 10 mM catalyst solution, which was stored in a sealed vial at room temperature.



$Pd(OAc)_2(L3)_2$

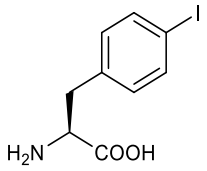
Based on protocols previously reported by our group,[31,48] 1,1-dimethylguanidine sulfate (**L3**, 13.6 mg, 0.1 mmol) was dissolved in 0.1M NaOH (2 ml, 0.2 mmol) and 1 ml H_2O, then heated to 65 °C. $Pd(OAc)_2$ (11.2 mg, 0.05 mmol) was added and the mixture stirred for 45 minutes, after which time all solid dissolved. After cooling to room temperature, the mixture was diluted with 2 ml H_2O to give a total of 5 ml of a yellow 10 mM catalyst solution, which was stored in a sealed vial at room temperature.



$Pd(OAc)_2(L4)_2$

Based on protocols previously reported by our group,[31,48] 1,1-dimethylguanidine sulfate (**L3**, 13.6 mg, 0.1 mmol) was dissolved in 0.1M NaOH (2 ml, 0.2 mmol) and 1 ml H_2O, then heated to 65 °C. $Pd(OAc)_2$ (11.2 mg, 0.05 mmol) was added and the mixture stirred for 45 minutes, after which time all solid dissolved. After cooling to room temperature, the mixture was diluted with 2 ml H_2O to give a total of 5 ml of a yellow 10 mM catalyst solution, which was stored in a sealed vial at room temperature.

L-*p*-iodophenylalanine (*p*IPhe, 2.1)



2.1

L-Phenylalanine (10 g, 60.5 mmol) was dissolved in acetic acid (60 ml) and conc. Sulphuric acid (7.5 ml). Sodium iodate (2.4 g, 12.1 mmol, 0.2 eq) and iodine (6.1 g, 24.2 mmol, 0.4 eq) were added, and the resultant purple solution stirred at 70 °C for 16 h. After this time, sodium periodate (0.53 g, 2.5 mmol, 0.04 eq) was added portionwise, and the solution stirred for a further 30 minutes, during which time it turned orange. The acetic acid was removed *in vacuo*, and the resultant mixture diluted with H_2O (100 ml) and washed with diethyl ether (2 x 50 ml) and DCM (2 x 750 ml). The aqueous layer was basified through the addition of 2M NaOH until an off-white precipitate formed, which was collected by filtration. Recrystallization from acetic acid gave the desired product **2.1** (***p*IPhe**) as an off-white solid (9.1 g, 31.3 mmol, 52 %).

^{1}H NMR (400 MHz, NaOD) δ 7.62 (d, *J* = 8.4 Hz, 2H, $\underline{H}_{Ar,meta}$), 6.95 (d, *J* = 8.4 Hz, 2H, $\underline{H}_{Ar,ortho}$), 3.38 (dd, *J* = 7.4, 5.6 Hz, 1H, $H_2NC\underline{H}$), 2.83 (dd, *J* = 13.5, 5.6 Hz, 1H, $C\underline{H}_2$), 2.70 (dd, *J* = 13.5, 7.4 Hz, 1H, $C\underline{H}_2$)

LRMS *m/z* (ESI+) found: 336.0 $(M+2Na-H)^+$, calculated 336.0 $(M+2Na-H)^+$

LRMS *m/z* (ESI-) found 290.0 (M-H), calculated 290.0 (M-H)

m.p. 260-263 °C decomp. (AcOH), lit.[25] 260-263 °C

$[\alpha]_D$ = +16 (c = 1.0, 0.5M HCl in EtOH), Lit.[49] +20.

Data consistent with that previously reported.[25,49]

2.8.2 Plasmid DNA Details and Sequencing

The plasmids utilised in the course of this work were generously prepared and generated by prior members of the Davis group, as follows:

pEVOL-*p*IPhe: donated by Dr. Chris Spicer, previously a generous gift from Prof. Peter Schultz.[24] This plasmid encodes two copies of the *p*IPhe-specific aaRS and the requisite *p*IPhe$_{TAG}$ tRNA, in a pEVOL vector.

GYG-wt, GYG-Y195F: donated by Dr. Seung Seo Lee, previously obtained from Structural Genomics Consortium (Oxford). These plasmids encode proteins GYG-wt and GYG-Y195F, contained in a pNIC28-bsa4 vector (human GYG-1 isoforms).

GYG-Y195X: donated by Gustaf Hemberg. This plasmid was previously prepared via standard site-directed mutagenesis of plasmid GYG-wt,[35] and encodes GYG-Y195X (where X = TAG codon) in a pNIC28-bsa4 vector (human GYG-1 isoform).

In order to confirm the correct mutations at the 195 site of the various GYG plasmids, they were all sent for Sanger sequencing to confirm the DNA sequence (T7 forward primer, and T7 reverse primer if required, multiple runs of the same sample may be necessary for full coverage).

The DNA and translated protein sequences for each plasmid thus obtained are illustrated below. The 195 site is in blue, and the N-terminal polyhistidine tag with TEV cleavage site (22 residues) is highlighted in yellow.

Note that amino acid numbering in the main text does not include the poly-histidine tag (22 residues). For consistency both within this text and with respect to prior literature, the numbering in sequences below similarly does not include this portion.

GYG-wt

```
    atgcaccatcatcatcattcttctggtgtagatctgggtaccgagaacctgtacttc
     M   H   H   H   H   H   H   S   S   G   V   D   L   G   T   E   N   L   Y   F
    caatccatgacagatcaggcctttgtgacactaaccacaaacgatgcctacgccaaaggt
     Q   S   M   T   D   Q   A   F   V   T   L   T   T   N   D   A   Y   A   K   G
19  gccctggtcctgggatcatctctgaaacagcacaggaccaccaggaggctggtcgtgctc
     A   L   V   L   G   S   S   L   K   Q   H   R   T   T   R   R   L   V   V   L
39  gccacccctcaggtctcagactccatgagaaaagttttagagacagtctttgatgaagtc
     A   T   P   Q   V   S   D   S   M   R   K   V   L   E   T   V   F   D   E   V
59  atcatggtagatgtcttggacagtggcgattctgctcatctaaccttaatgaagaggcca
     I   M   V   D   V   L   D   S   G   D   S   A   H   L   T   L   M   K   R   P
79  gagttgggtgtcacgctgacaaagctccactgctggtcgcttacacagtattcaaaatgt
     E   L   G   V   T   L   T   K   L   H   C   W   S   L   T   Q   Y   S   K   C
```

```
99  gtattcatggatgcagatactctggtcctagcaaatattgatgatcttttgacagagaa
     V  F  M  D  A  D  T  L  V  L  A  N  I  D  D  L  F  D  R  E
119 gaattgtcagcagcaccagacccagggtggcctgactgcttcaattccggagtcttcgtt
     E  L  S  A  A  P  D  P  G  W  P  D  C  F  N  S  G  V  F  V
139 tatcagccttcagttgaaacatacaatcagctgttgcatcttgcttctgagcaaggtagt
     Y  Q  P  S  V  E  T  Y  N  Q  L  L  H  L  A  S  E  Q  G  S
159 tttgatggtggggaccaaggcatactgaacacatttttagcagctgggcaacaacagat
     F  D  G  G  D  Q  G  I  L  N  T  F  F  S  S  W  A  T  T  D
179 atcagaaaacacctgccgtttatttataacctaagcagcatctctatatcctacctc
     I  R  K  H  L  P  F  I  Y  N  L  S  S  I  S  I  Y  S  Y  L
199 ccggcatttaaagtgtttggtgcaagtgccaaagttgtgcatttcctgggacgagtcaaa
     P  A  F  K  V  F  G  A  S  A  K  V  V  H  F  L  G  R  V  K
219 ccatggaattatacttatgatcccaaaacaaaaagtgtcaaaagtgaggcccatgatccc
     P  W  N  Y  T  Y  D  P  K  T  K  S  V  K  S  E  A  H  D  P
239 aacatgactcatccagagtttctcatcctgtggtggaacatctttaccaccaacgtttta
     N  M  T  H  P  E  F  L  I  L  W  W  N  I  F  T  T  N  V  L
259 cctctgcttcaatga
     P  L  L  Q  -
```

GYG-Y195F

```
    atgcaccatcatcatcattcttctggtgtagatctgggtaccgagaacctgtacttc
     M  H  H  H  H  H  H  S  S  G  V  D  L  G  T  E  N  L  Y  F
    caatccatgacagatcaggcctttgtgacactaaccacaaacgatgcctacgccaaaggt
     Q  S  M  T  D  Q  A  F  V  T  L  T  T  N  D  A  Y  A  K  G
19  gccctggtcctgggatcatctctgaaacagcacaggaccaccaggaggctggtcgtgctc
     A  L  V  L  G  S  S  L  K  Q  H  R  T  T  R  R  L  V  V  L
39  gccacccctcaggtctcagactccatgagaaaagttttagagacagtctttgatgaagtc
     A  T  P  Q  V  S  D  S  M  R  K  V  L  E  T  V  F  D  E  V
59  atcatggtagatgtcttggacagtggcgattctgctcatctaaccttaatgaagaggcca
     I  M  V  D  V  L  D  S  G  D  S  A  H  L  T  L  M  K  R  P
79  gagttgggtgtcacgctgacaaagctccactgctggtcgcttacacagtattcaaaatgt
     E  L  G  V  T  L  T  K  L  H  C  W  S  L  T  Q  Y  S  K  C
99  gtattcatggatgcagatactctggtcctagcaaatattgatgatcttttgacagagaa
     V  F  M  D  A  D  T  L  V  L  A  N  I  D  D  L  F  D  R  E
119 gaattgtcagcagcaccagacccagggtggcctgactgcttcaattccggagtcttcgtt
     E  L  S  A  A  P  D  P  G  W  P  D  C  F  N  S  G  V  F  V
139 tatcagccttcagttgaaacatacaatcagctgttgcatcttgcttctgagcaaggtagt
     Y  Q  P  S  V  E  T  Y  N  Q  L  L  H  L  A  S  E  Q  G  S
159 tttgatggtggggaccaaggcatactgaacacatttttagcagctgggcaacaacagat
     F  D  G  G  D  Q  G  I  L  N  T  F  F  S  S  W  A  T  T  D
179 atcagaaaacacctgccgtttatttataacctaagcagcatctctatattcctacctc
     I  R  K  H  L  P  F  I  Y  N  L  S  S  I  S  I  F  S  Y  L
199 ccggcatttaaagtgtttggtgcaagtgccaaagttgtgcatttcctgggacgagtcaaa
     P  A  F  K  V  F  G  A  S  A  K  V  V  H  F  L  G  R  V  K
219 ccatggaattatacttatgatcccaaaacaaaaagtgtcaaaagtgaggcccatgatccc
     P  W  N  Y  T  Y  D  P  K  T  K  S  V  K  S  E  A  H  D  P
239 aacatgactcatccagagtttctcatcctgtggtggaacatctttaccaccaacgtttta
     N  M  T  H  P  E  F  L  I  L  W  W  N  I  F  T  T  N  V  L
259 cctctgcttcaatga
     P  L  L  Q  -
```

GYG-Y195X (X = TAG codon, = *p*IPhe in expressed protein)

```
    atgcaccatcatcatcattcttctggtgtagatctgggtaccgagaacctgtacttc
     M  H  H  H  H  H  H  S  S  G  V  D  L  G  T  E  N  L  Y  F
    caatccatgacagatcaggcctttgtgacactaaccacaaacgatgcctacgccaaaggt
     Q  S  M  T  D  Q  A  F  V  T  L  T  T  N  D  A  Y  A  K  G
19  gccctggtcctgggatcatctctgaaacagcacaggaccaccaggaggctggtcgtgctc
     A  L  V  L  G  S  S  L  K  Q  H  R  T  T  R  R  L  V  V  L
39  gccacccctcaggtctcagactccatgagaaaagttttagagacagtctttgatgaagtc
     A  T  P  Q  V  S  D  S  M  R  K  V  L  E  T  V  F  D  E  V
59  atcatggtagatgtcttggacagtggcgattctgctcatctaaccttaatgaagaggcca
     I  M  V  D  V  L  D  S  G  D  S  A  H  L  T  L  M  K  R  P
79  gagttgggtgtcacgctgacaaagctccactgctggtcgcttacacagtattcaaaatgt
     E  L  G  V  T  L  T  K  L  H  C  W  S  L  T  Q  Y  S  K  C
99  gtattcatggatgcagatactctggtcctagcaaatattgatgatcttttgacagagaa
     V  F  M  D  A  D  T  L  V  L  A  N  I  D  D  L  F  D  R  E
119 gaattgtcagcagcaccagacccagggtggcctgactgcttcaattccggagtcttcgtt
     E  L  S  A  A  P  D  P  G  W  P  D  C  F  N  S  G  V  F  V
139 tatcagccttcagttgaaacatacaatcagctgttgcatcttgcttctgagcaaggtagt
     Y  Q  P  S  V  E  T  Y  N  Q  L  L  H  L  A  S  E  Q  G  S
159 tttgatggtggggaccaaggcatactgaacacatttttagcagctgggcaacaacagat
     F  D  G  G  D  Q  G  I  L  N  T  F  F  S  S  W  A  T  T  D
179 atcagaaaacacctgccgtttatttataacctaagcagcatctctatatagtcctacctc
     I  R  K  H  L  P  F  I  Y  N  L  S  S  I  S  I  -  S  Y  L
199 ccggcatttaaagtgtttggtgcaagtgccaaagttgtgcatttcctgggacgagtcaaa
     P  A  F  K  V  F  G  A  S  A  K  V  V  H  F  L  G  R  V  K
219 ccatggaattatacttatgatcccaaaacaaaaagtgtcaaaagtgaggcccatgatccc
     P  W  N  Y  T  Y  D  P  K  T  K  S  V  K  S  E  A  H  D  P
239 aacatgactcatccagagtttctcatcctgtggtggaacatctttaccaccaacgtttta
     N  M  T  H  P  E  F  L  I  L  W  W  N  I  F  T  T  N  V  L
259 cctctgcttcaatga
     P  L  L  Q  -
```

2.8.3 Protein Expression and Characterisation

2.8.3.1 GYG-Y195pIPhe

Expression of this protein was carried out in collaboration with Mr. Henry Bailey, a DPhil researcher in the group of Prof. Wyatt Yue, SGC, University of Oxford; certain expressions were carried out by myself, and others by him. The data below comes from a sample expressed by myself.

Transformation and Expression

BL21(DE3) cells were thawed on ice. A 20 μl aliquot was transferred to a pre-chilled 14 ml falcon tube. Plasmids pEVOL-pIPhe (1.5 μl, ca. 50 ng/μl) and GYG-Y195X (1.5 μl, ca. 50 ng/μl) were added, and the tube placed on ice for 30 minutes with occasional swirling. Cells were heat shocked (42 °C, 40 s) then immediately placed on ice for a further 2 minutes. 80 μl of SOC media was added, and cells incubated at 37 °C for 1 h with shaking (200 rpm). Cells were plated on LB-agar plates containing both chloramphenicol (25 μg/ml) and kanamycin (50 μg/ml). Plates were incubated at 37 °C overnight.

A single colony obtained from the above transformation was used to inoculate 5 ml LB media, supplemented with kanamycin (50 μg/ml) and chloramphenicol (25 μg/ml). Cells were grown at 37 °C overnight with shaking (200 rpm). Resultant cells were used to inoculate 1L of similarly supplemented SB media (see recipe below). Cells were grown (37 °C, 180 rpm) to a final OD_{600} of 0.6-0.7 (typically 3 h). For each L culture, L-Arabinose was added to a final concentration of 0.02 %, followed by L-*p*-iodophenylalanine ***p*IPhe** to a final concentration of 2 mM (580 mg was dissolved in 5 ml 1M NaOH). Cells were incubated at 30 °C, 180 rpm for a further 2 hours, before addition of IPTG to a final concentration of 0.5 mM. Cells were cooled to 18 °C and incubated with shaking at 180 rpm for 16 h. After this time, cells were harvested through centrifugation (7000 x g, 7 minutes, 4 °C). Cells may be flash frozen and stored at -80 °C prior to lysis and purification if desired.

SB media: (per L culture) 32 g tryptone, 20 g yeast extract, 5 g NaCl.

Cell Lysis and Purification

Cell pellets were resuspended in lysis buffer (approx. 15 ml per L culture) with addition of cOmplete™ EDTA-free protease inhibitor cocktail. Lysozyme (approx. 5 mg per L culture) and DNaseI (approx. 0.5 mg per L culture) were added and the suspension gently shaken at 4 °C for 30 minutes. Cells were then

sonicated (40 % amplitude, 5 s on, 10 s off, 12 cycles) and cell debris removed by centrifugation (50000 x g, 4 °C, 20 minutes).

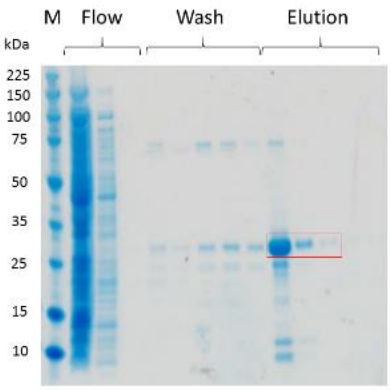
The clarified lysate was loaded onto a gravity flow Ni-NTA column. The flowthrough was collected, and the protein was purified using a step gradient of imidazole (20 mM – 40 mM – 60 mM – 250 mM, protein elutes at 250 mM). Fractions were analysed by SDS-PAGE. For isolation of His-tagged protein, fractions containing the desired protein were combined and dialysed against dialysis buffer (2 x 2L, 10000 MWCO, 4 °C, 2 h then overnight). Protein was concentrated by Vivaspin (10000 MWCO, Sartorius Stedem) to a concentration of >1 mg/ml, flash frozen and stored in 100 µl aliquots at – 80 °C. Typical protein yields were ca. 1 mg per L culture. LC-MS confirmed the isolation of the desired GYG-Y195*p*IPhe protein in good purity; a small amount of gluconoylation was visible (+178 Da, this is a known modification of N-terminal polyhistidine tags[39]).

Lysis Buffer: 20 mM imidazole, 50 mM Na-HEPES, 0.5 M NaCl, 0.5 mM TCEP, 5 % Glycerol, pH 7.5

Dialysis Buffer: 10 mM Na-HEPES, 0.5 M NaCl, 1 mM Na-EDTA, 5 % glycerol, 1 mM DTT, pH 7.5.

Characterisation Data:

SDS-PAGE:



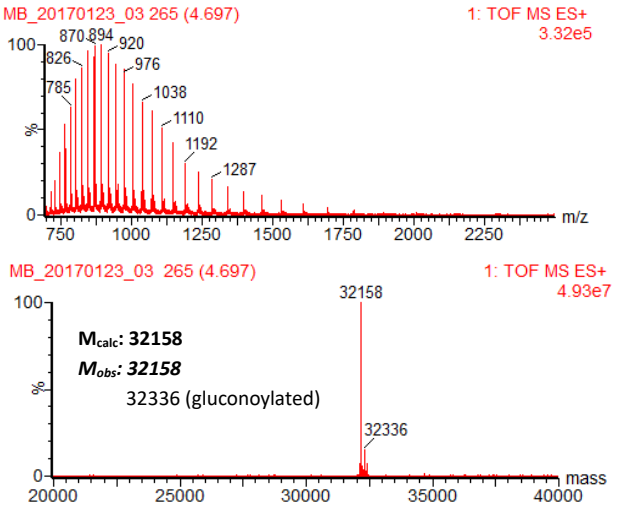
M = molecular weight markers

Flow = flowthrough (20 mM imidazole)

Wash = 40 mM and 60 mM imidazole

Elution = 250 mM imidazole

LC-MS:



2.8.3.2 GYG-wt

Transformation and Expression

BL21(DE3) cells were thawed on ice. A 50 μl aliquot was transferred to a pre-chilled 14 ml falcon tube. Plasmid GYG-wt (1.0 μl, ca. 100 ng/μl) was added, and the tube placed on ice for 10 minutes with occasional swirling. Cells were heat shocked (42 °C, 30 s) then immediately placed on ice for a further 10 minutes. 250 μl of SOC media was added, and cells incubated at 37 °C for 1 h with shaking (200 rpm). Cells were plated on LB-agar plates containing both kanamycin (50 μg/ml); 20 μl of the transformation mixture is typically ample to obtain many colonies. Plates were incubated at 37 °C overnight.

A single colony obtained from the above transformation was used to inoculate 5 ml LB media, supplemented with kanamycin (50 μg/ml). Cells were grown at 37 °C overnight with shaking (200 rpm). Resultant cells were each used to inoculate 0.5 L of LB media similarly supplemented with kanamycin. Cells were grown (37 °C, 180 rpm) to a final OD_{600} of 0.6-0.7 (typically 2 h). For each culture, IPTG was added to a final concentration of 0.4 mM. Cells were cooled to 20 °C and incubated with shaking at 180 rpm for 16 h. Cells were then harvested through centrifugation (7000 x g, 7 minutes, 4 °C).

Cell Lysis and Purification

Cell pellets were resuspended in lysis buffer (approx. 15 ml per L culture) with addition of cOmplete™ EDTA-free protease inhibitor cocktail. Lysozyme (approx. 5 mg per L culture) and DNaseI (approx. 0.5 mg per L culture) were added and the suspension gently shaken at 4 °C for 60 minutes. Cells were then sonicated (40 % amplitude, 0.5 s on, 0.5 s off, 3 x 30 second cycles with 30 second pauses between) and cell debris removed by centrifugation (50000 x g, 4 °C, 20 minutes).

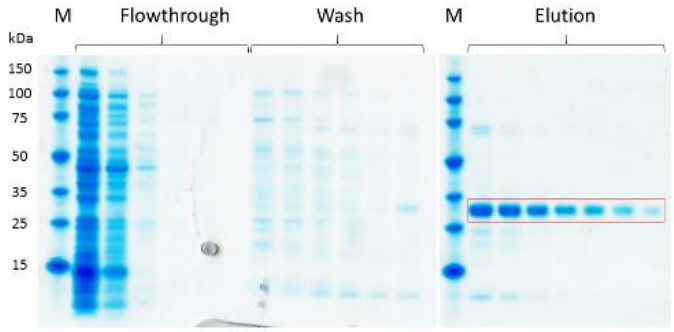
The clarified lysate was loaded onto a gravity flow Ni-NTA column. The flowthrough was collected, and the column washed with lysis buffer (ca. 40 ml per ml resin). Protein was purified using a step gradient of imidazole (10 mM – 35 mM – 60 mM – 160 mM – 250 mM – 500 mM imidazole).Fractions were analysed by SDS-PAGE. Fractions containing the desired protein were combined and dialysed against dialysis buffer (2 x 2L, 10000 MWCO, 4 °C, 2 h then overnight). Protein was concentrated by Vivaspin (10000 MWCO, Sartorius Stedem) to a concentration of >1 mg/ml, flash frozen and stored in aliquots at -80 °C. Typical protein yields were ca. 3 mg per L culture. LC-MS confirmed the isolation of the desired GYG-wt in good purity, as a mixture of glycoforms.

Lysis Buffer: 10 mM imidazole, 20 mM Tris-HCl, 0.5 M NaCl, pH 7.5

Dialysis Buffer: 10 mM Na-HEPES, 0.5 M NaCl, 1 mM Na-EDTA, 5 % glycerol, 1 mM DTT, pH 7.5.

Characterisation Data:

SDS-PAGE:



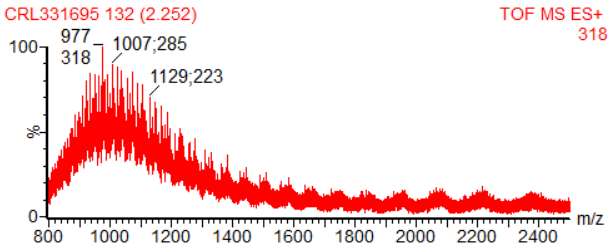
M = molecular weight markers

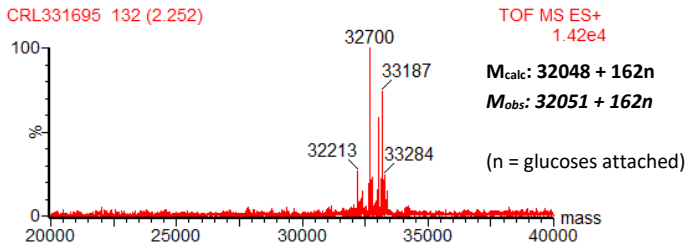
Flow = flowthrough (10 mM imidazole)

Wash = 35-60 mM imidazole

Elution = 160-500 mM imidazole

LC-MS:





M_{calc}: 32048 + 162n

M_{obs}: 32051 + 162n

(n = glucoses attached)

Note that precise glycoforms / extent of glucosylation may vary with different batches, even though the same protocol as above was always utilised.

2.8.3.3 GYG-Y195F

Transformation and Expression

BL21(DE3) cells were thawed on ice. A 50 μl aliquot was transferred to a pre-chilled 14 ml falcon tube. Plasmid GYG-Y195F (1.0 μl, ca. 80 ng/μl) was added, and the tube placed on ice for 30 minutes with occasional swirling. Cells were heat shocked (42 °C, 30 s) then immediately placed on ice for a further 2 minutes. 250 μl of SOC media was added, and cells incubated at 37 °C for 1 h with shaking (200 rpm). Cells were plated on LB-agar plates containing both kanamycin (50 μg/ml); 20 μl of the transformation mixture is typically ample to obtain many colonies. Plates were incubated at 37 °C overnight.

A single colony obtained from the above transformation was used to inoculate 10 ml LB media, supplemented with kanamycin (50 μg/ml). Cells were grown at 37 °C overnight with shaking (200 rpm). Resultant cells were used to inoculate 1 L of LB media. Cells were grown (37 °C, 180 rpm) to a final OD_{600} of 0.6-0.7. For each culture, IPTG was added to a final concentration of 0.5 mM. Cells were cooled to 30 °C and incubated with shaking at 180 rpm for 2 h, then cooled to 18 °C and incubated overnight. After this time, cells were harvested through centrifugation (7000 x g, 7 minutes, 4 °C).

Cell Lysis and Purification

Cell pellets were resuspended in lysis buffer (approx. 15 ml per L culture) with addition of cOmplete™ EDTA-free protease inhibitor cocktail. Lysozyme (approx. 5 mg per L culture) and DNaseI (approx. 0.5 mg per L culture) were added and the suspension gently shaken at 4 °C for 30 minutes. Cells were then sonicated (35 % amplitude, 5 s on, 10 s off, 12 cycles) and cell debris removed by centrifugation (50000 x g, 4 °C, 20 minutes).

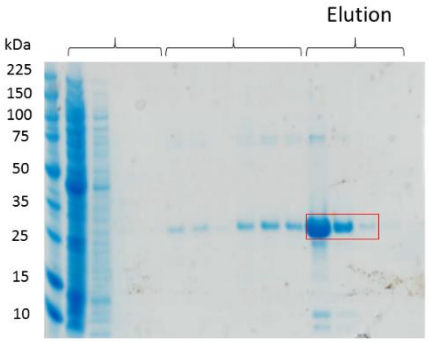
The clarified lysate was loaded onto a gravity flow Ni-NTA column. The flowthrough was collected, and the column washed with lysis buffer (ca. 40 ml per ml resin). Protein was purified using a step gradient of imidazole (20 mM – 40 mM – 60 mM – 250 mM imidazole).Fractions were analysed by SDS-PAGE. Fractions containing the desired protein were combined and dialysed against dialysis buffer (2 x 2L, 10000 MWCO, 4 °C, 2 h then overnight). Protein was concentrated by Vivaspin (10000 MWCO, Sartorius Stedem) to a concentration of >1 mg/ml, flash frozen and stored in aliquots at -80 °C. Typical protein yields were ca. 3 mg per L culture. LC-MS confirmed the isolation of the desired GYG-Y195F protein in good purity; a small level of N-terminal gluconoylation was visible (+178 Da, this is a known modification of N-terminal polyhistidine tags[39]).

Lysis Buffer: 50 mM Hepes pH 7.5, 500 mM NaCl, 20 mM Imidazole, 0.5 mM TCEP, 5% Glycerol

Dialysis Buffer: 10 mM Hepes pH 7.5, 500 mM NaCl, 1 mM DTT, 1 mM Na-EDTA, 5 % Glycerol

Characterisation Data:

SDS-PAGE:



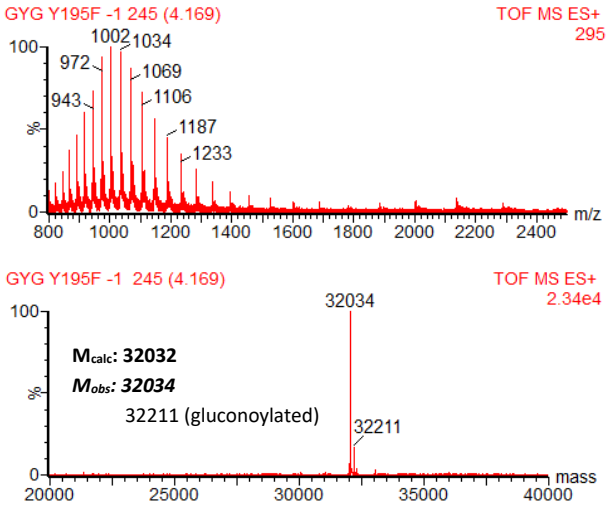
M = molecular weight markers

Flow = flowthrough (20 mM imidazole)

Wash = 40-60 mM imidazole

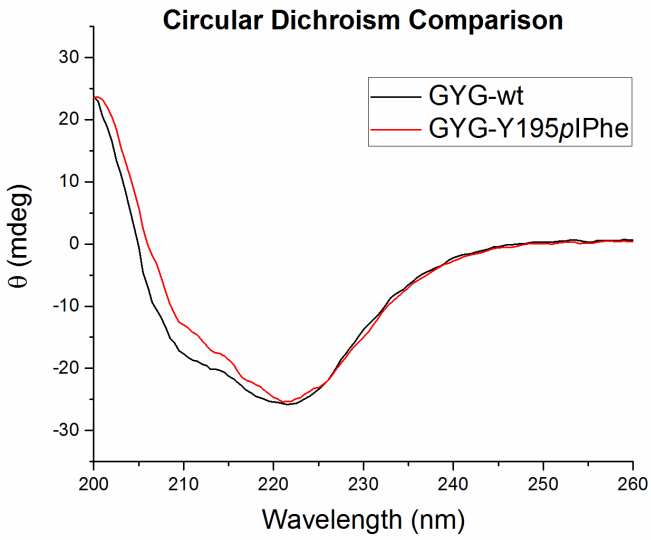
Elution = 250 mM imidazole

LC-MS:



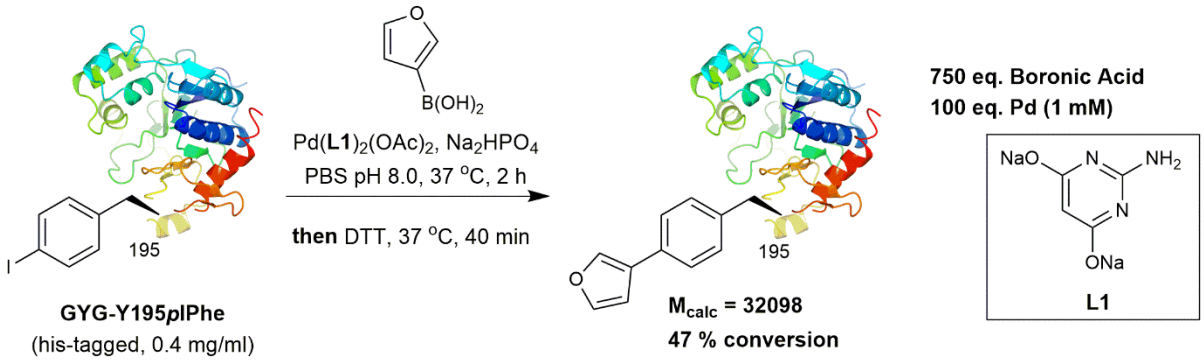
2.8.3.4 Circular Dichroism Comparison of GYG-wt and GYG-Y195*p*IPhe

GYG-wt or GYG-Y195*p*IPhe was buffer exchanged into "CD buffer" (50 mM NaH_2PO_4, 5 % v/v glycerol, pH 8.0) by PD SpinTrap G-25 (GE Healthcare). Circular dichroism spectrum of each protein was then recorded and analysed at a concentration of 0.2 mg/ml. Data represent mean averages of three runs (see also appendix).

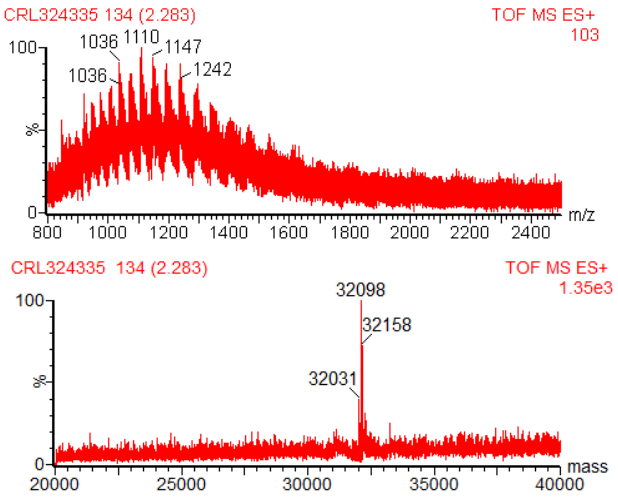


2.8.4 Cross-Couplings to 3-furanylboronic acid

2.8.4.1 Initial Cross-Coupling Test

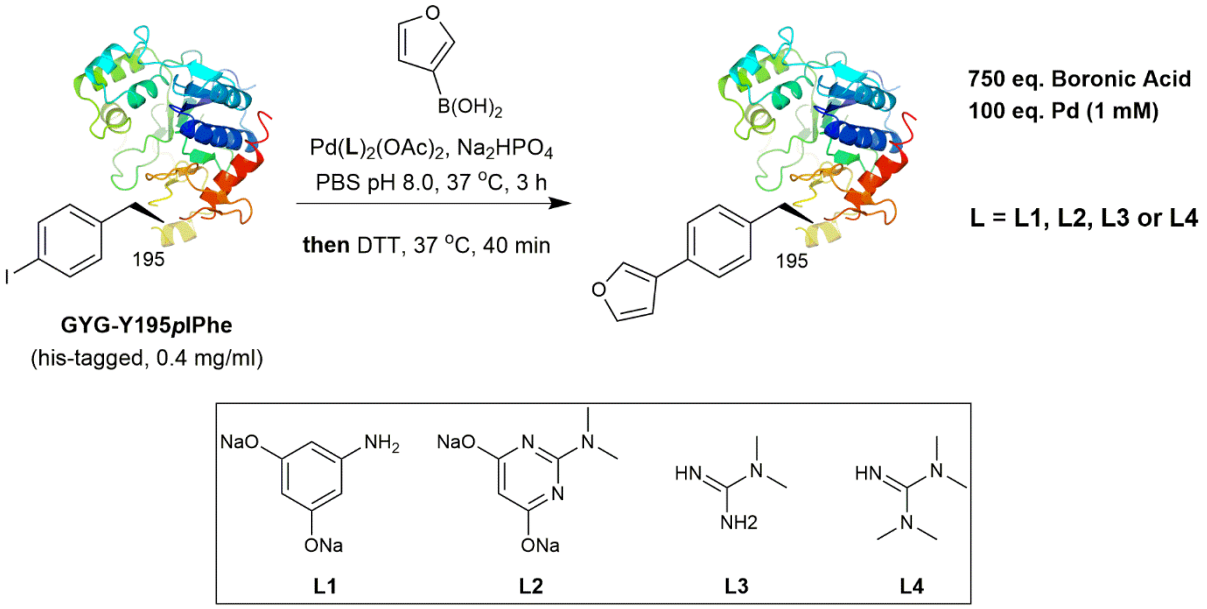


A stock solution of 3-furanylboronic acid was prepared by dissolving boronic acid (12 mg/ml) and Na_2HPO_4 (13 mg/ml) in H_2O. To His-tagged GYG-Y195*p*IPhe (35 µl of 0.4 mg/ml, 0.43 nmol) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was added 3.00 µl of this stock solution (320 nmol, 750 eq. boronic acid), followed by $Pd(L1)_2(OAc)_2$ catalyst (4.3 µl of 10 mM solution, 43 nmol, 100 eq.). The reaction mixture was briefly vortexed, then incubated at 37 °C with shaking (600 rpm). After 2 hours, the reaction was quenched by the addition of DTT (7.48 µl of an 8.85 mg/ml stock solution in H_2O, 430 nmol, 10 eq. w.r.t. Pd), shaken for a further 40 minutes at 37 °C, then analysed by LC-MS.

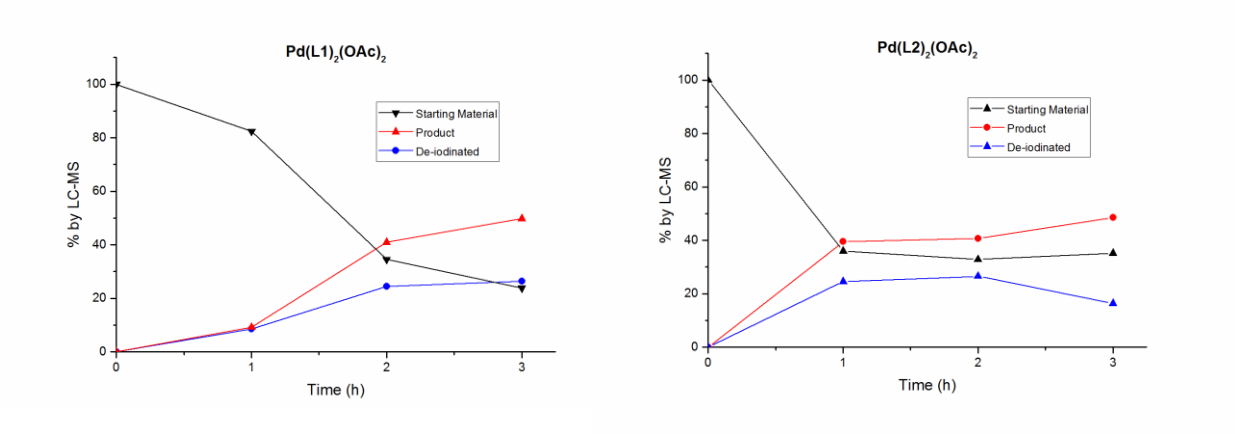


M_{calc}: 32098
M_{obs}: 32098 (47 %)
32031 (de-iodination)
32158 (unreacted GYG-Y195*p*IPhe)

2.8.4.2 Comparison of Catalyst Ligands

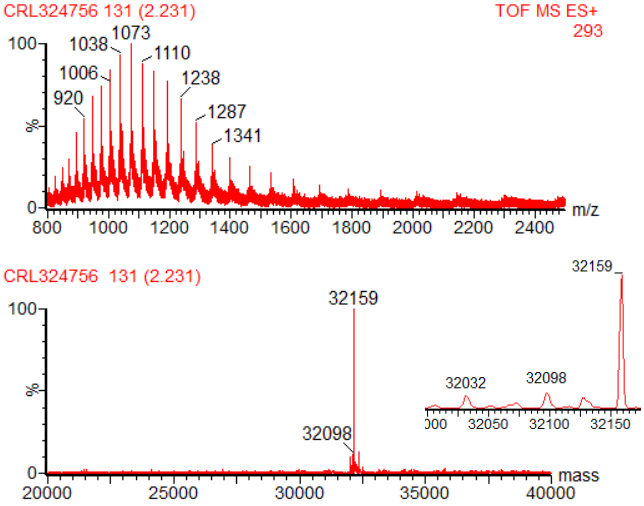


4 reactions were set up. In each case, a stock solution of 3-furanylboronic acid was prepared by dissolving boronic acid (12 mg/ml) and Na_2HPO_4 (13 mg/ml) in H_2O. To His-tagged GYG-Y195*p*IPhe (35 µl of 0.4 mg/ml, 0.43 nmol) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was added 3.00 µl of this stock solution (320 nmol, 750 eq. boronic acid), followed by either $Pd(L1)_2(OAc)_2$, $Pd(L2)_2(OAc)_2$, $Pd(L3)_2(OAc)_2$ or $Pd(L4)_2(OAc)_2$ catalyst (4.3 µl of 10 mM solution, 43 nmol, 100 eq.). The reaction mixtures were briefly vortexed, then incubated at 37 °C with shaking (600 rpm). After 1, 2 and 3 hours, 1/3 of the reaction volume was removed and quenched by addition of DTT (2.49 µl of 8.85 mg/ml stock solution in H_2O, 143 nmol, 10 eq. w.r.t. Pd), and shaken for a further 40 minutes at 37 °C, before being analysed by LC-MS. For the cases of $Pd(L3)_2(OAc)_2$ and $Pd(L4)_2(OAc)_2$, protein precipitation was observed inside 10 minutes and no LC-MS signal could be detected. For the other two catalysts, data are illustrated below.



Example LC-MS data:

Pd(**L1**)$_2$(OAc)$_2$, 1 hour:



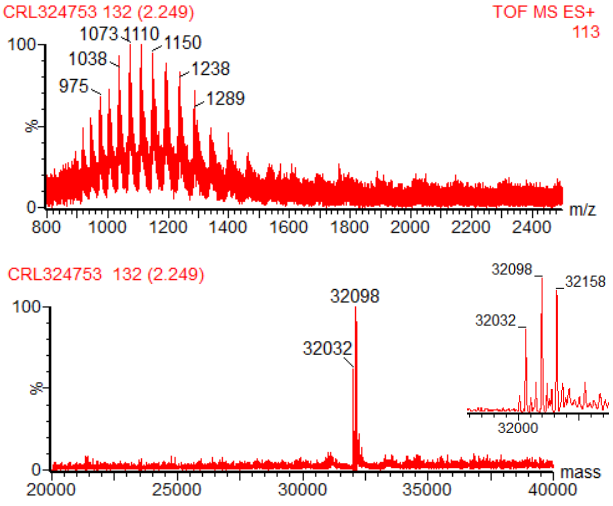
M$_{calc}$: 32098

M$_{obs}$: 32098

32032 (de-iodination)

32159 (unreacted GYG-Y195*p*IPhe)

Pd(**L2**)$_2$(OAc)$_2$, 1 hour:



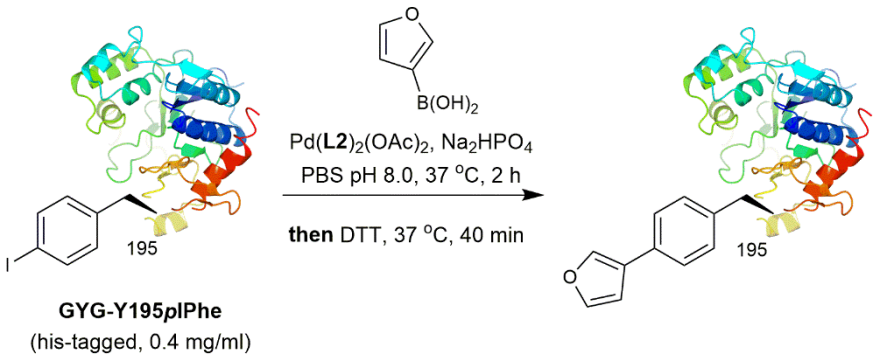
M$_{calc}$: 32098

M$_{obs}$: 32098

32032 (de-iodination)

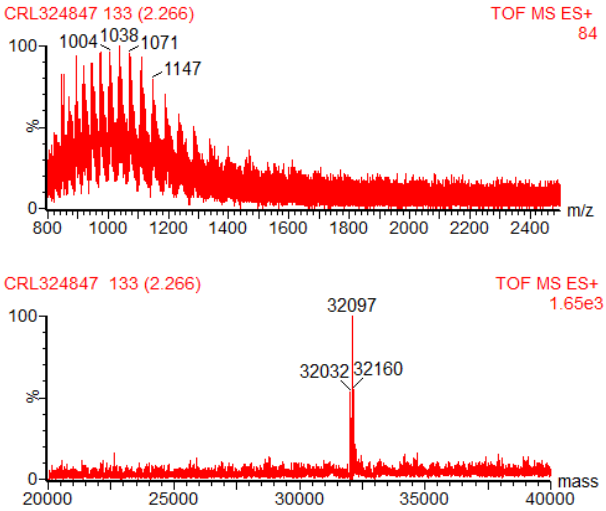
32158 (unreacted GYG-Y195*p*IPhe)

2.8.4.3 Effect of Boronic Acid Loading (1) (Table 2.1, entries 1-2)



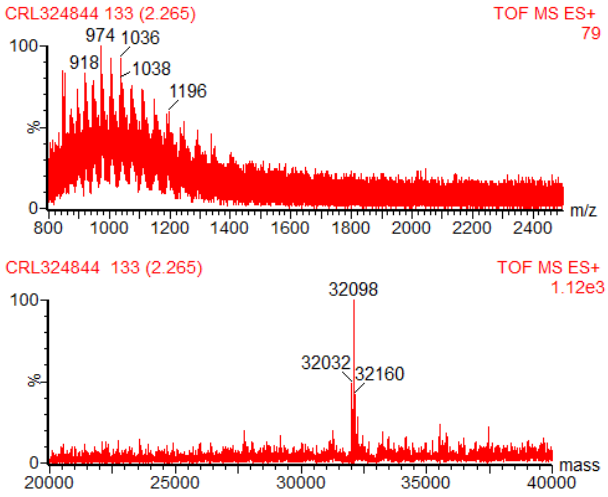
2 reactions were set up. In each case, a stock solution of 3-furanylboronic acid was prepared by dissolving boronic acid (11 mg/ml) and Na_2HPO_4 (15 mg/ml) in H_2O. To His-tagged GYG-Y195*p*IPhe (30 µl of 0.4 mg/ml, 0.37 nmol) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was added *either* 2.84 µl (275 nmol, 750 eq. boronic acid) *or* 4.06 µl (430 nmol, 1150 eq. boronic acid) of this stock solution, followed by $Pd(L2)_2(OAc)_2$ catalyst (3.7 µl of 10 mM solution, 37 nmol, 100 eq.). The reaction mixture was briefly vortexed, then incubated at 37 °C with shaking (600 rpm). After 2 hours, the reaction was quenched by the addition of DTT (7.48 µl of an 8.85 mg/ml stock solution in H_2O, 430 nmol, 10 eq. w.r.t. Pd), shaken for a further 40 minutes at 37 °C, then analysed by LC-MS.

Reaction 1 (750 eq.):



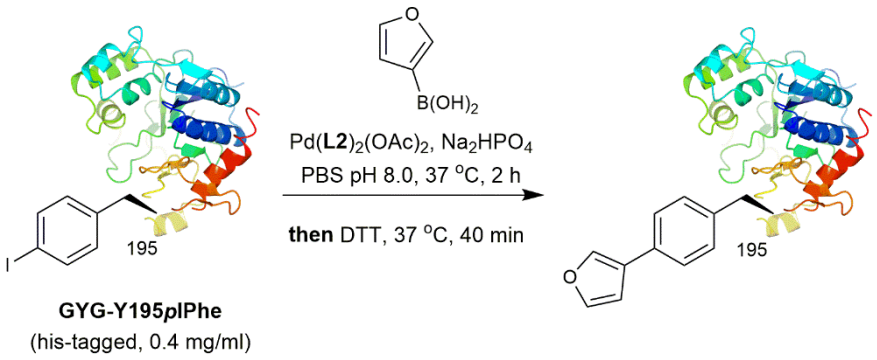
M_{calc}: 32098 (42 %)
M_{obs}: 32097
32032 (de-iodination)
32160 (unreacted GYG-Y195*p*IPhe)

Reaction 2 (1150 eq.):

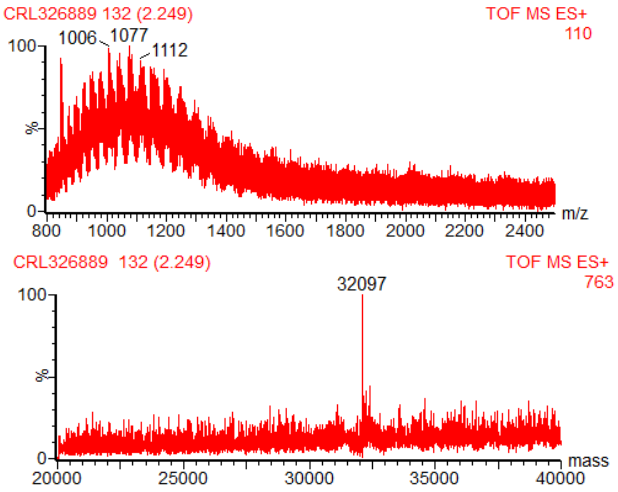


M_{calc}: 32098 (52 %)
M_{obs}: 32098
32032 (de-iodination)
32160 (unreacted GYG-Y195*p*IPhe)

2.8.4.4 Effect of Further Increase in Boronic Acid Loading (Table 2.1, entry 3)



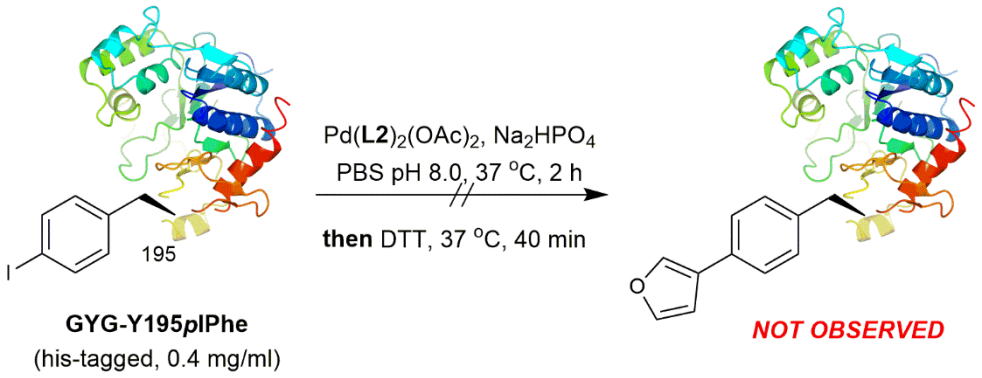
A stock solution of 3-furanylboronic acid was prepared by dissolving boronic acid (11 mg/ml) and Na_2HPO_4 (15 mg/ml) in H_2O. To GYG-Y195*p*IPhe (His-tagged) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was added 5.25 µl of this stock solution (515 nmol, 1200 eq. boronic acid), followed by $Pd(L2)_2(OAc)_2$ catalyst (4.3 µl of 10 mM solution, 43 nmol, 100 eq.). The reaction mixture was briefly vortexed, then incubated at 37 °C with shaking (600 rpm). After 30, 60 and 90 minutes, a further 360 eq. boronic acid solution was added. After 2 hours, an aliquot (1/3 of reaction volume) was removed and quenched by the addition of DTT (1.87 µl of a 115 mM stock solution in H_2O, 215 nmol, 15 eq. w.r.t. Pd), shaken for a further 40 minutes at 37 °C, then analysed by LC-MS.



M_{calc}: **32098**
M_{obs}: 32097

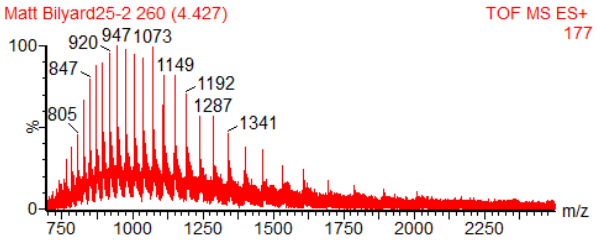
2.8.4.5 Negative Control Reactions

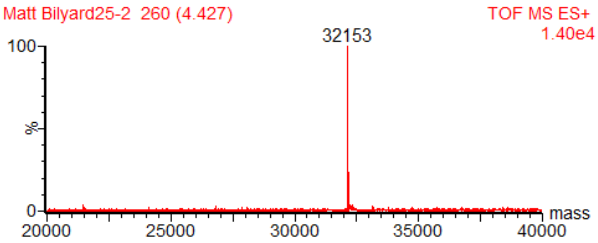
No boronic acid



Conditions are identical to those of table 2.1, entry 2, with no boronic acid. In presence of boronic acid, reaction was clearly observed (see section 2.8.4.3, reaction 2).

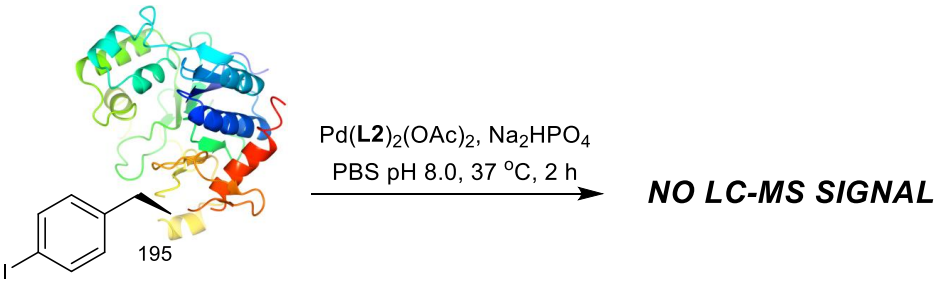
To His-tagged GYG-Y195*p*IPhe (30 µl of 0.4 mg/ml, 0.37 nmol) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was added 4.06 µl of Na_2HPO_4 solution (15 mg/ml) followed by $Pd(L2)_2(OAc)_2$ catalyst (3.7 µl of 10 mM solution, 37 nmol, 100 eq.). The reaction mixture was briefly vortexed, then incubated at 37 °C with shaking (600 rpm). After 2 hours, the reaction was quenched by the addition of DTT (7.48 µl of an 8.85 mg/ml stock solution in H_2O, 430 nmol, 10 eq. w.r.t. Pd), shaken for a further 40 minutes at 37 °C, then analysed by LC-MS. Only starting material was observed.





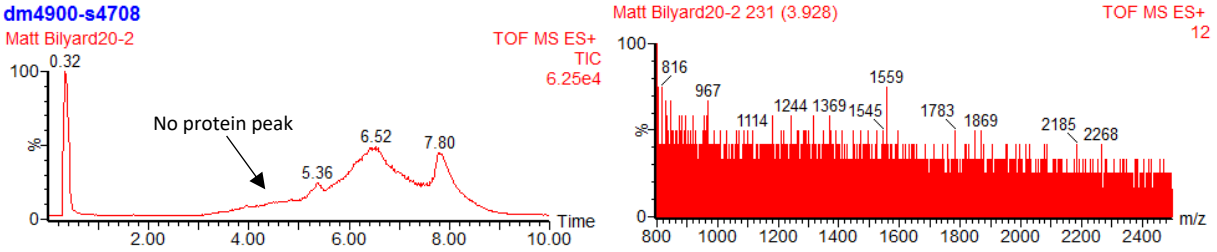
M_{calc} (GYG-Y195*p*IPhe) 32158
M_{obs}: 32153

Pd only, no Pd scavenger (or boronic acid)

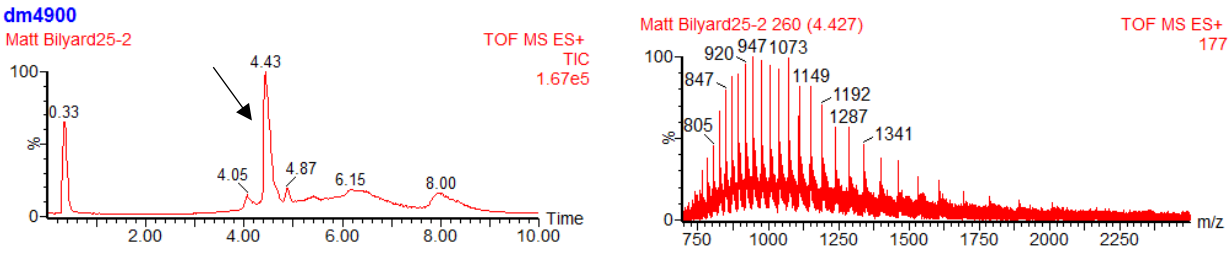


GYG-Y195*p*IPhe
(his-tagged, 0.4 mg/ml)

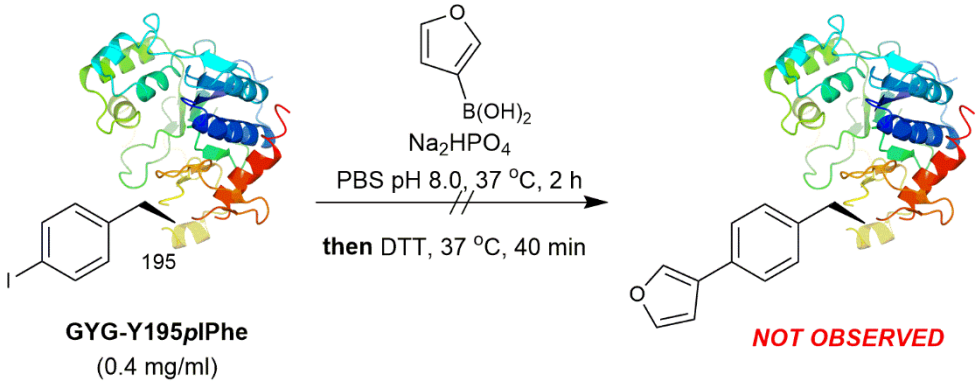
Reaction was carried out exactly as in "no boronic acid" above, but with no addition of DTT. No protein peak was observed upon LC-MS analysis, consistent with prior reports.[25]



For comparison, the chromatogram for the "no boronic acid" reaction above is shown (this is an identical reaction, but with addition of Pd scavenger). A clear protein peak is observed.

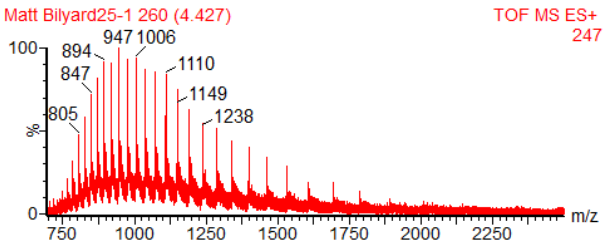


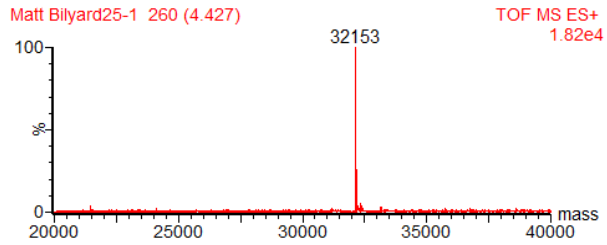
Boronic Acid only, no Pd



Conditions are identical to those of table 2.1, entry 2, with no Pd. In presence of palladium, reaction was clearly observed (see section 2.8.4.3, reaction 2).

A stock solution of 3-furanylboronic acid was prepared by dissolving boronic acid (12 mg/ml) and Na_2HPO_4 (13 mg/ml) in H_2O. To His-tagged GYG-Y195*p*IPhe (30 µl of 0.4 mg/ml, 0.37 nmol) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was added 4.06 µl (430 nmol, 1150 eq. boronic acid) of this stock solution, followed by H_2O (3.7 µl). The reaction mixture was briefly vortexed, then incubated at 37 °C with shaking (600 rpm). After 2 hours, DTT (7.48 µl of an 8.85 mg/ml stock solution in H_2O, 430 nmol, 10 eq. w.r.t. Pd) was added, the reaction shaken for a further 40 minutes at 37 °C, then analysed by LC-MS. Only starting material was observed.

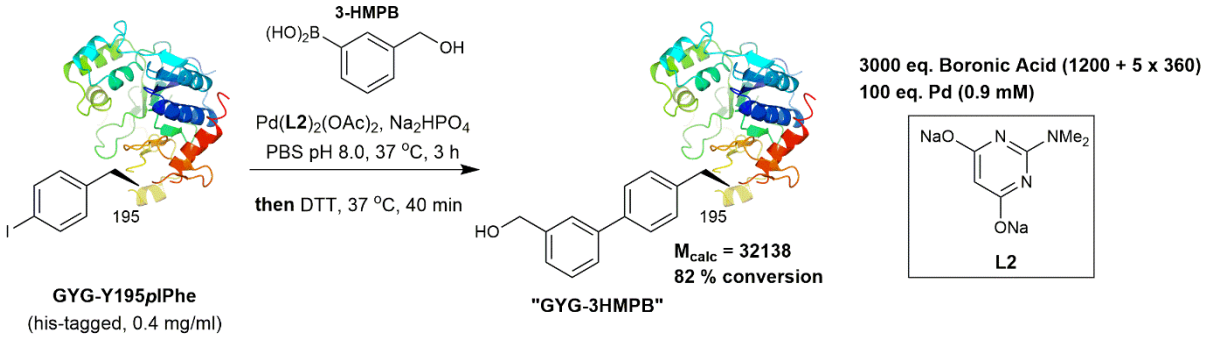




M_{calc} (GYG-Y195*p*IPhe) 32158
M_{obs}: 32153

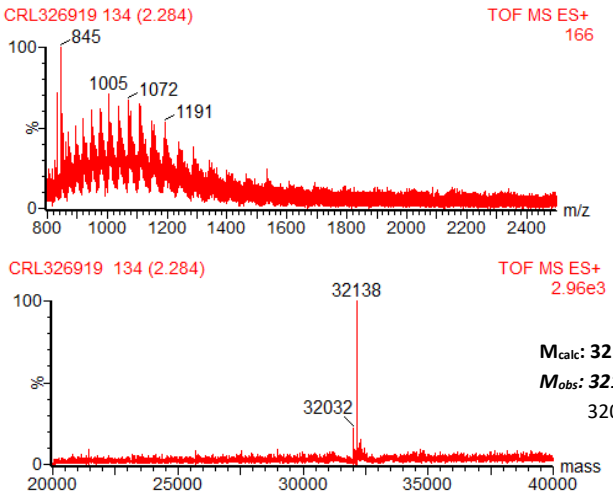
2.8.5 Cross-Couplings to 3-HMPB

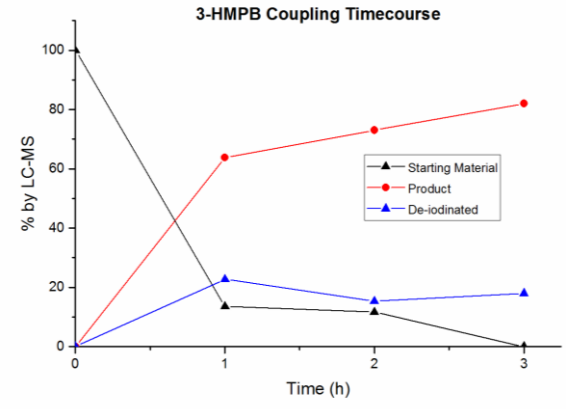
2.8.5.1 Initial Cross-Coupling Test



A stock solution of 3-(hydroxymethyl)phenylboronic acid (3-HMPB) was prepared by dissolving boronic acid (14.9 mg/ml) and Na_2HPO_4 (15 mg/ml) in H_2O. To His-tagged GYG-Y195*p*IPhe (35 µl of 0.4 mg/ml, 0.43 nmol) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was added 5.25 µl of this stock solution (515 nmol, 1200 eq. boronic acid), followed by $Pd(L2)_2(OAc)_2$ catalyst (4.3 µl of 10 mM solution, 43 nmol, 100 eq.). The reaction mixture was briefly vortexed, then incubated at 37 °C with shaking (600 rpm). After 30, 60, 90, 120 and 150 minutes, a further 360 eq. boronic acid solution was added. After 1, 2 and 3 hours, an aliquot (1/3 of reaction volume) was removed and quenched by the addition of DTT (1.87 µl of a 115 mM stock solution in H_2O, 215 nmol, 15 eq. w.r.t. Pd), shaken for a further 40 minutes at 37 °C, then analysed by LC-MS.

Example LC-MS Data (3 hours):



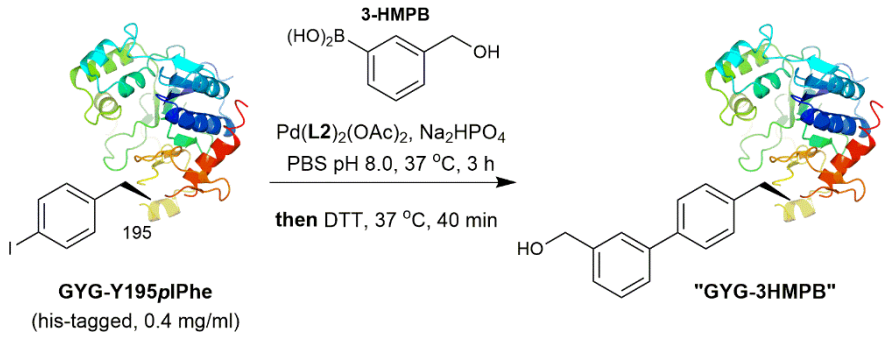


M_{calc}: 32138 (82 %)

M_{obs}: 32138

32032 (de-iodination)

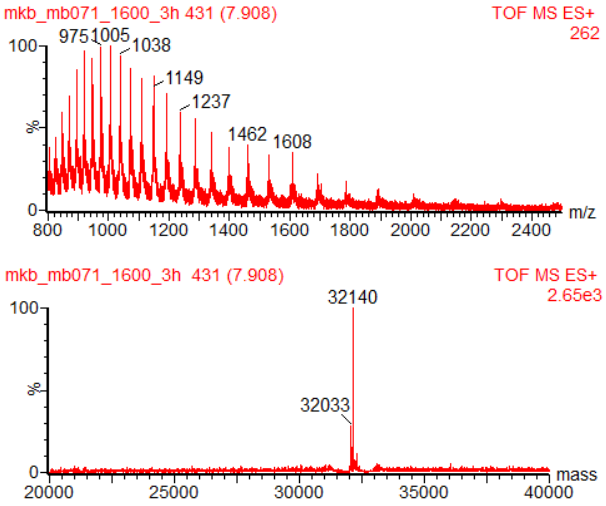
2.8.5.2 Optimisation of Boronic Acid Loading



3 reactions were set up. A stock solution of 3-(hydroxymethyl)phenylboronic acid (3-HMPB) was prepared by dissolving boronic acid (15.6 mg/ml[a]) and Na_2HPO_4 (15.7 mg/ml = 110 mM) in H_2O. To His-tagged GYG-Y195*p*IPhe (35 µl of 0.4 mg/ml, 0.43 nmol) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was added either 6.70, 8.38 or 10.48 µl of this stock solution (690/860/1075 nmol, 1600/2000/2500 eq. boronic acid), followed by $Pd(L2)_2(OAc)_2$ catalyst (4.3 µl of 10 mM solution, 43 nmol, 100 eq.). The reaction was vortexed and incubated at 37 °C with shaking (600 rpm). After 30, 60 and 90 minutes, a further 400 eq. boronic acid solution was added. After 3 hours, an aliquot (1/3 reaction volume) was removed and quenched by the addition of DTT (1.87 µl of a 115 mM stock solution in H_2O, 215 nmol, 15 eq. w.r.t. Pd), shaken for 40 minutes at 37 °C, then analysed by LC-MS.

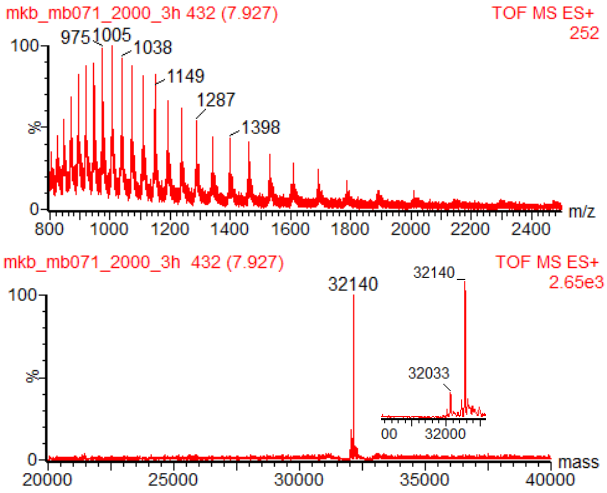
[a]*This is a saturated solution, so as to minimise significant dilution effects at high boronic acid loadings. This was utilised from hereon in.*

1600 eq. boronic acid initial loading (78 % conversion)



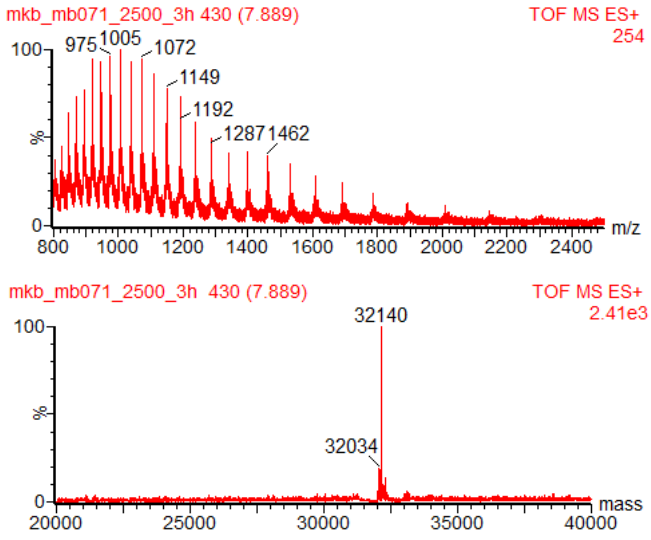
M_{calc}: 32138
M_{obs}: 32138
32033 (de-iodination)

2000 eq. boronic acid initial loading (84 % conversion):



M_{calc}: 32138
M_{obs}: 32140
32033 (de-iodination)

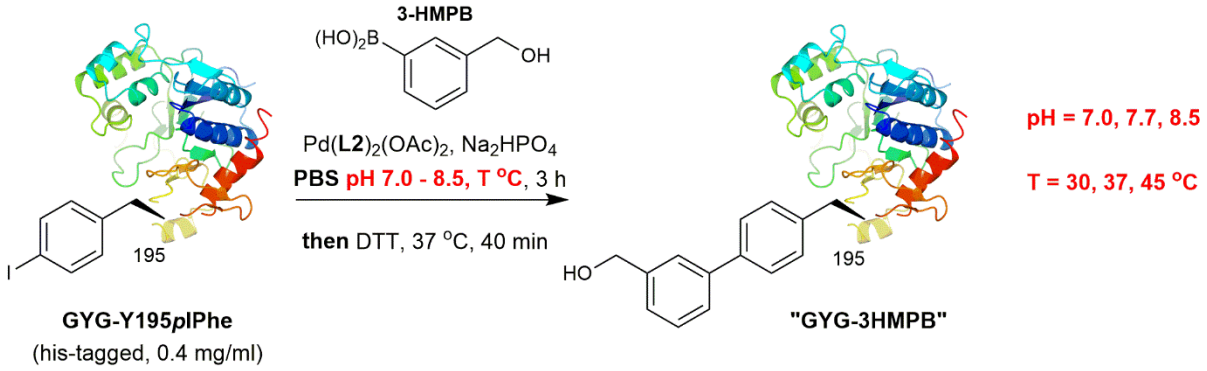
2500 eq. boronic acid initial loading (84 % conversion):



M_{calc}: 32138
M_{obs}: 32140
32034 (de-iodination)

2.8.5.3 Effect of pH and temperature variation

General Procedure

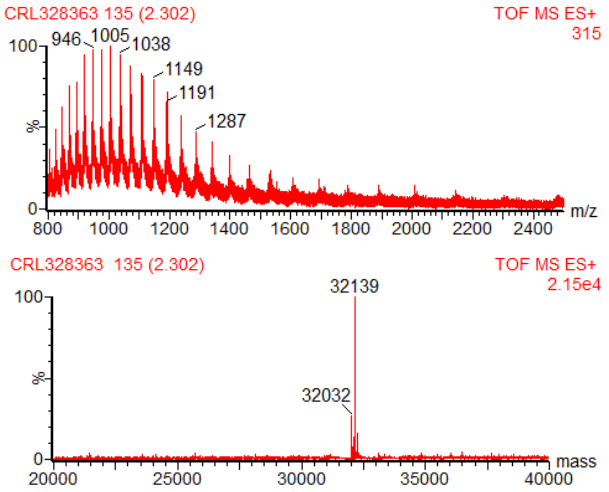


A stock solution of 3-(hydroxymethyl)phenylboronic acid (3-HMPB) was prepared by dissolving boronic acid (15.6 mg/ml) and Na_2HPO_4 (15.7 mg/ml = 110 mM) in H_2O. To His-tagged GYG-Y195*p*IPhe (35 µl of 0.4 mg/ml, 0.43 nmol) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was added 8.38 µl of this stock solution (860 nmol, 2000 eq. boronic acid), followed by $Pd(L2)_2(OAc)_2$ catalyst (4.3 µl of 10 mM solution, 43 nmol, 100 eq.). The reaction was vortexed and incubated at **T** °C with shaking (600 rpm). After 30, 60 and 90 minutes, a further 400 eq. boronic acid solution was added. After 3 hours, an aliquot (1/3 reaction volume) was removed and quenched by the addition of DTT (1.87 µl of a 115 mM stock solution in H_2O, 215 nmol, 15 eq. w.r.t. Pd), shaken for 40 minutes at 37 °C, then analysed by LC-MS.

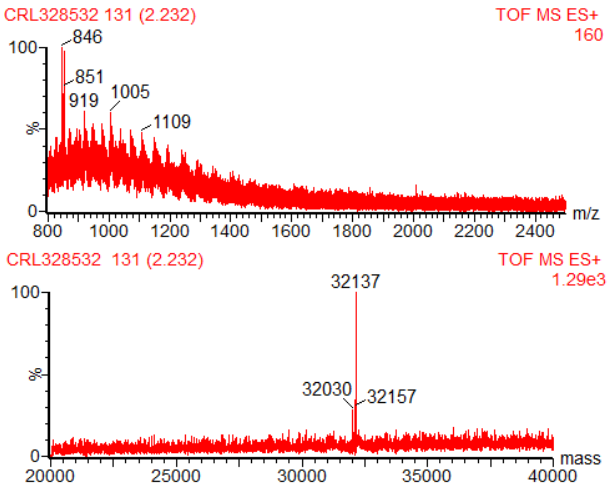
Variation of Temperature: 3 reactions were set up according to general procedure above, in which "**T** °C" was either 30, 37 or 45 °C. LC-MS spectra are shown below; for graphical data see **figure 2.11**.

Variation of pH: 3 reactions were set up according to general procedure above. In 2 cases, the pH was adjusted to 7.0 or 8.5 by addition of 3 µl 0.05 M HCl or 5 µl 0.05 M NaOH respectively after addition of Pd catalyst; in the third case, no base or acid was added. The volumes of acid/base required were determined by measuring the amount required to adjust the pH of a 10x volume mixture of buffer, boronic acid, and Pd catalyst in correct proportions (without protein) by pH meter, then dividing this by 10. LC-MS spectra are shown below; for graphical data please see **figure 2.11**.

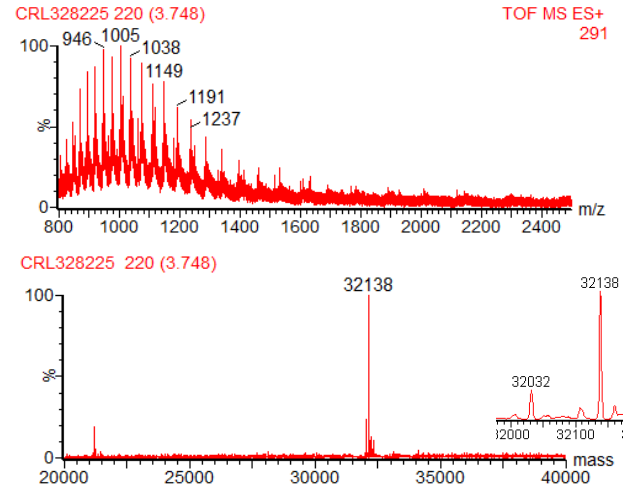
30 °C, pH 7.8 (79 % conversion)



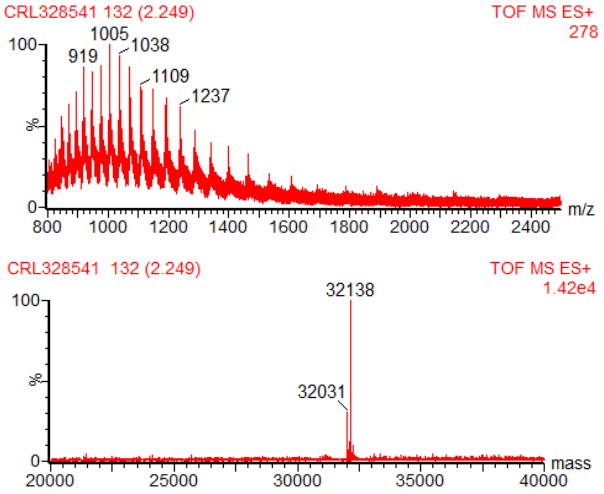
45 °C, pH 7.8 (63 % conversion)



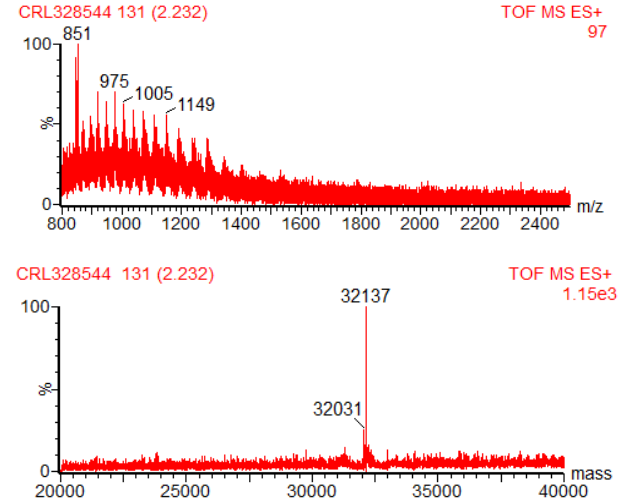
37 °C, pH 7.8 (81 % conversion)



37 °C, pH 8.5 (76 % conversion)



37 °C, pH 7.0 (80 % conversion)



M_{calc}: 32138

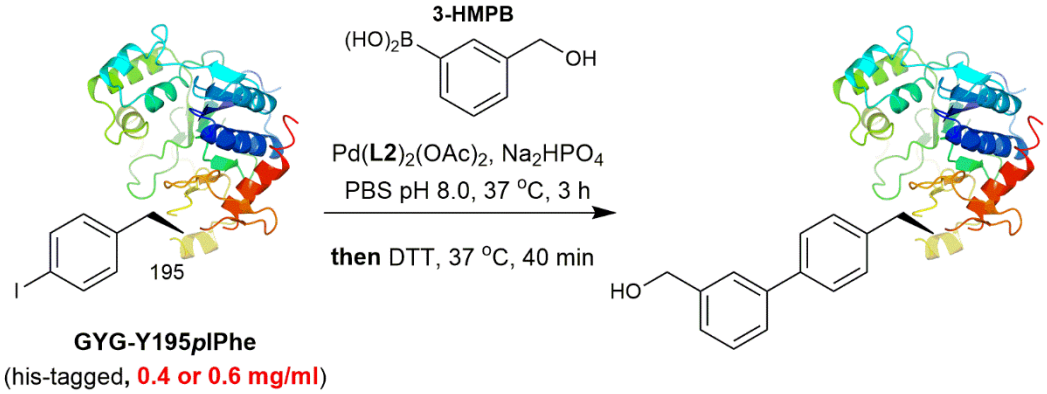
M_{obs}: 32137-32139

32030-32032 (de-iodination)

32157 (GYG-Y195*p*IPhe, 45 °C only)

(In all cases except 45 °C, only products are cross-coupling and de-iodination. For 45 °C, 19 % GYG-Y195*p*IPhe is present, and 18 % de-iodination)

2.8.5.4 Effect of Protein Concentration



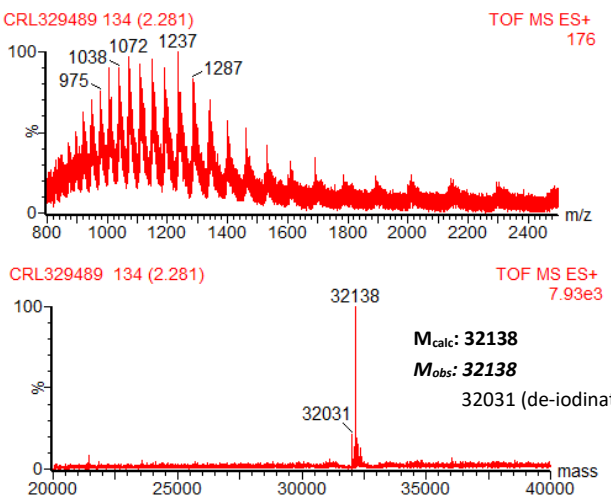
2 reactions were set up. A stock solution of 3-(hydroxymethyl)phenylboronic acid (3-HMPB) was prepared by dissolving boronic acid (15.6 mg/ml) and Na_2HPO_4 (15.7 mg/ml = 110 mM) in H_2O. To His-tagged GYG-Y195*p*IPhe (35 µl of 0.4 mg/ml, 0.43 nmol **or** 22.5 µl of 0.6 mg/ml, 0.42 nmol) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was added stock solution and Pd catalyst as follows:

Reaction 1 (0.40 mg/ml): 8.38 µl stock solution (860 nmol, 2000 eq., 18 mM), 4.3 µl 10 mM Pd catalyst solution (43 nmol, 100 eq., 0.9 mM).

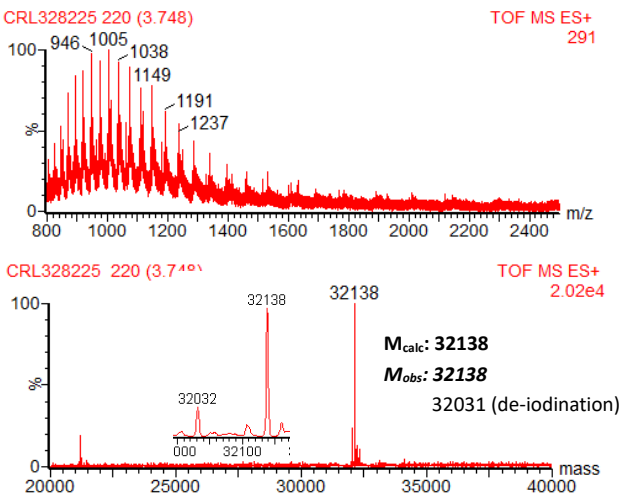
Reaction 2 (0.6 mg/ml): 5.36 µl stock solution (550 nmol, 1320 eq., 18 mM), 2.8 µl 10 mM Pd catalyst solution (28 nmol, 67 eq., 0.9 mM).

Reactions were incubated at 37 °C for 3 hours at 600 rpm. After 30, 60 and 90 minutes, a further 400 eq. (reaction 1) or 255 eq. (reaction 2) boronic acid was added. After 3 hours, the reactions were quenched through addition of DTT (5.60 or 3.58 µl of a 115 mM stock solution, 15 eq. w.r.t. Pd), shaken for a further 40 minutes and analysed by LC-MS.

0.6 mg/ml (82 % conversion):

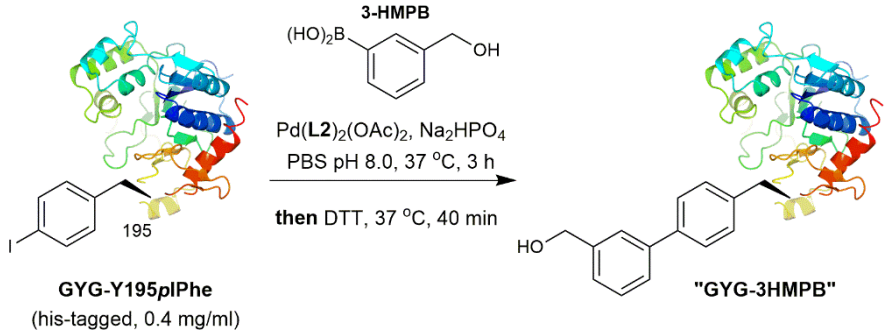


0.4 mg/ml (81 % conversion):



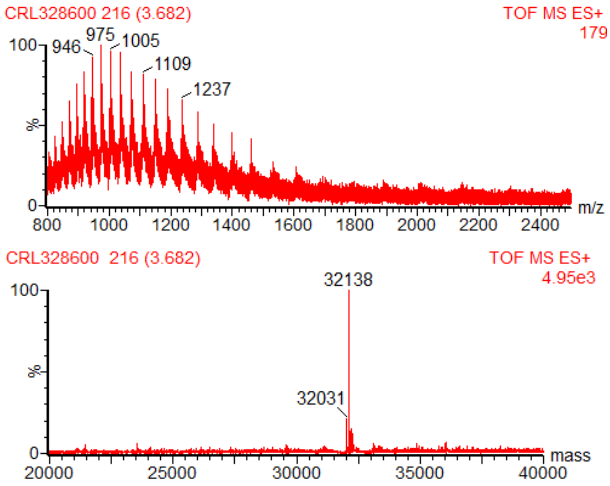
2.8.5.5 Example of Cross-coupling without stepwise boronic acid addition

In the text, it was stated that it was ultimately realised that stepwise addition of further boronic acid after the initial addition was not of significant benefit. An example without such further stepwise addition is illustrated below. This is identical to the protocols utilised for e.g. "37 °C, pH 7.8" (section 2.8.5.3), and "2000 eq. boronic acid" (section 2.8.5.2) above, but without the stepwise addition, and as such results should be compared to these.



A stock solution of 3-(hydroxymethyl)phenylboronic acid (3-HMPB) was prepared by dissolving boronic acid (15.6 mg/ml) and Na_2HPO_4 (15.7 mg/ml = 110 mM) in H_2O. To His-tagged GYG-Y195*p*IPhe (70 µl of 0.4 mg/ml, 0.86 nmol) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was added 16.76 µl of this stock solution (2000 eq. boronic acid), followed by $Pd(L2)_2(OAc)_2$ catalyst (8.6 µl of 10 mM solution, 86 nmol, 100 eq.). The reaction was vortexed and incubated at 37 °C with shaking (600 rpm). **No further boronic acid was added**. After 3 hours, the reaction was quenched by the addition of DTT (added from a 115 mM stock solution in H_2O, 15 eq. w.r.t. Pd), shaken for 40 minutes at 37 °C, then analysed by LC-MS.

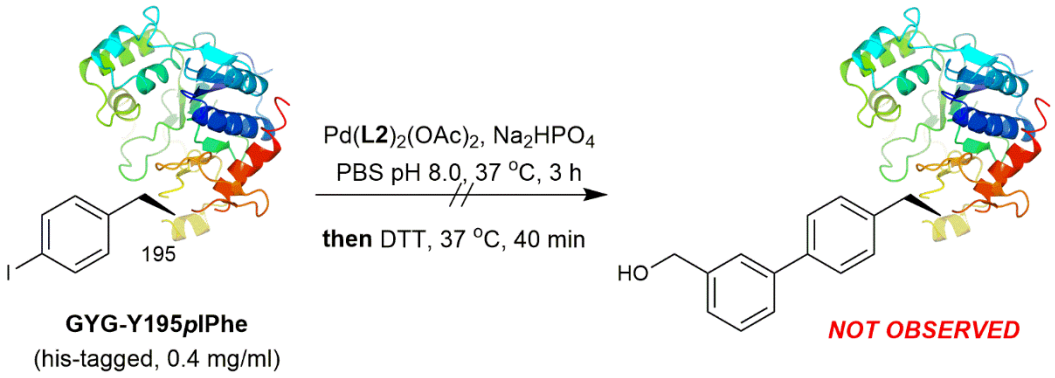
A conversion of **83 %** was attained. This is very similar to analogous conversions with further stepwise addition of boronic acid (see section 2.8.5.2, "2000 eq." and section 2.8.5.3 "pH 7.8, 37 °C).



M_{calc}: 32138
M_{obs}: 32138
32031 (de-iodination)

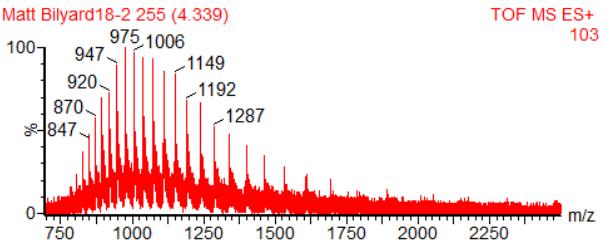
2.8.5.6 Negative Control Reactions

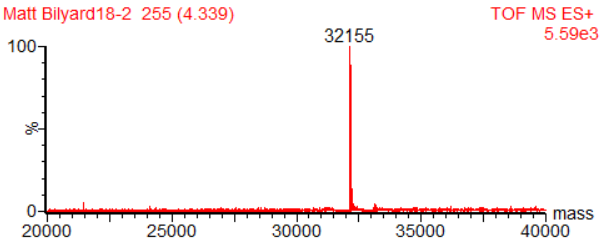
No Boronic Acid



Protocol identical to that utilised in section 2.8.5.2, "2000 eq.", with boronic acid absent. In presence of boronic acid, reaction was clearly observed (compare to section 2.8.5.2, "2000 eq. initial loading").

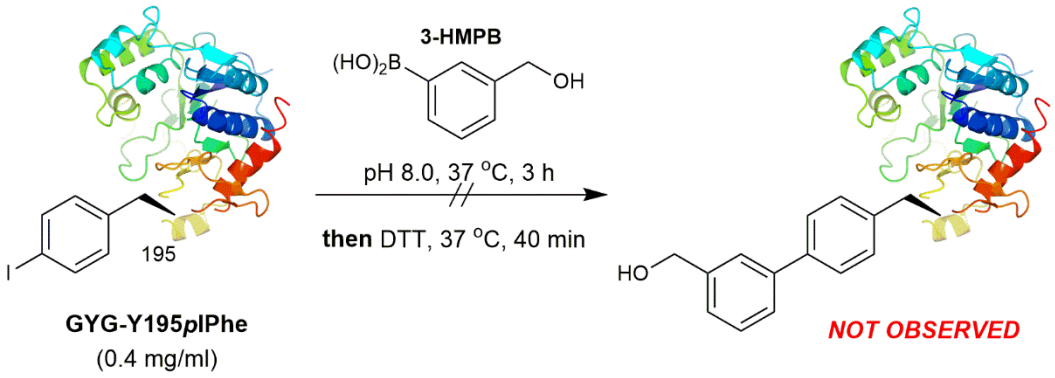
To His-tagged GYG-Y195*p*IPhe (35 µl of 0.4 mg/ml, 0.43 nmol) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was added 8.38 µl of aqueous Na_2HPO_4 (15.7 mg/ml) followed by $Pd(L2)_2(OAc)_2$ catalyst (4.3 µl of 10 mM solution, 43 nmol, 100 eq.). The reaction was vortexed and incubated at 37 °C with shaking (600 rpm). After 30, 60 and 90 minutes, a further 1.68 µl Na_2HPO_4 solution was added (this corresponds to 400 eq. boronic acid when using the 3-HMPB stock solution). After 3 hours, the reaction was quenched by the addition of DTT (5.6 µl of a 115 mM stock solution in H_2O, 15 eq. w.r.t. Pd), shaken for 40 minutes at 37 °C, then analysed by LC-MS. No reaction was observed.





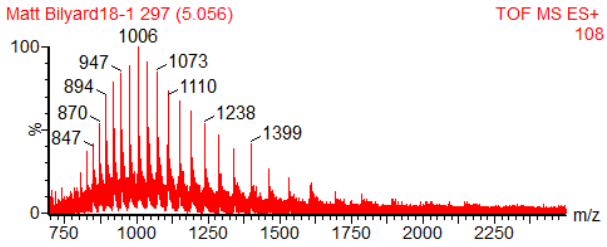
M_{calc} (GYG-Y195*p*IPhe) 32158
M_{obs}: 32155

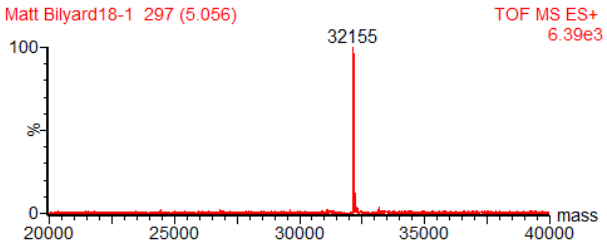
No Palladium



Protocol identical to that utilised in e.g. section 2.8.5.2, "2000 eq.", with palladium absent. In presence of palladium, reaction was clearly observed (compare to section 2.8.5.2, "2000 eq. initial loading").

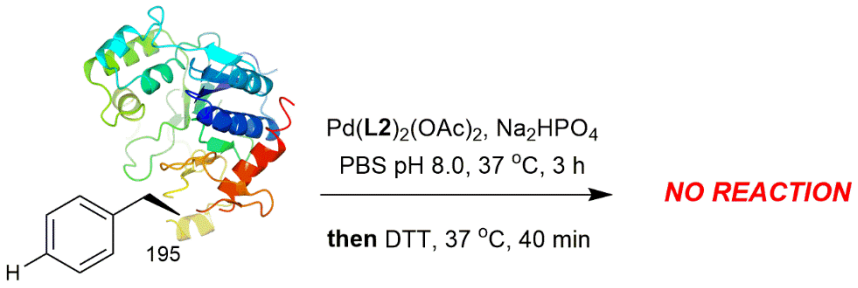
A stock solution of 3-(hydroxymethyl)phenylboronic acid (3-HMPB) was prepared by dissolving boronic acid (15.6 mg/ml) and Na_2HPO_4 (15.7 mg/ml) in H_2O. To His-tagged GYG-Y195*p*IPhe (35 µl of 0.4 mg/ml, 0.43 nmol) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was added 8.38 µl of this stock solution (860 nmol, 2000 eq. boronic acid), followed by H_2O (4.3 µl). The reaction was vortexed and incubated at 37 °C with shaking (600 rpm). After 30, 60 and 90 minutes, a further 400 eq. (1.68 µl) boronic acid solution was added After 3 hours, an aliquot was removed and DTT added (5.6 µl of a 115 mM stock solution in H_2O, 150 eq.), shaken for 40 minutes at 37 °C, then analysed by LC-MS. No reaction was observed.





M_{calc} (GYG-Y195*p*IPhe) 32158
M_{obs}: 32155

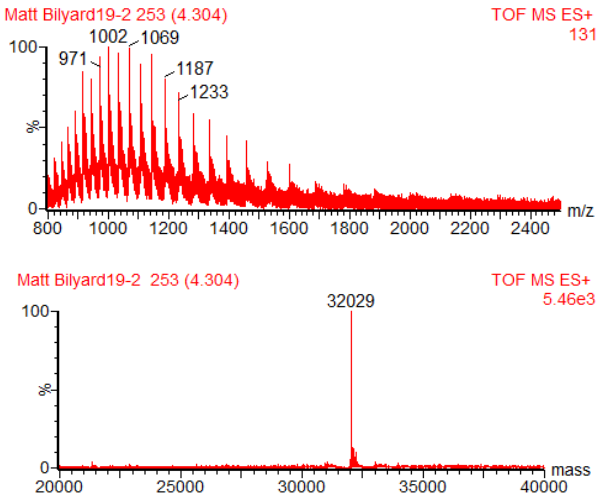
No Iodine (GYG-Y195F)



GYG-Y195F
(0.4 mg/ml, his-tagged)

Protocol identical to that utilised in e.g. section 2.8.5.2, "2000 eq.", with iodine replaced by hydrogen. In presence of iodine, cross-coupling (M_{calc} = 32138) and de-iodination (M_{calc} = 32032) peaks were clearly observed (see section 2.8.5.2, "2000 eq. initial loading"). If these peaks arose from reaction at a residue other than pIPhe*, one would expect similar reaction (and levels thereof) for GYG-Y195F. However, no reaction is seen.*

A stock solution of 3-(hydroxymethyl)phenylboronic acid (3-HMPB) was prepared by dissolving boronic acid (15.6 mg/ml) and Na_2HPO_4 (15.7 mg/ml) in H_2O. To His-tagged GYG-Y195F (35 µl of 0.4 mg/ml, 0.43 nmol) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was added 8.38 µl of this stock solution (860 nmol, 2000 eq. boronic acid), followed by $Pd(L2)_2(OAc)_2$ catalyst (4.3 µl of 10 mM solution, 43 nmol, 100 eq.). The reaction was vortexed and incubated at 37 °C with shaking (600 rpm). After 30, 60 and 90 minutes, a further 400 eq. boronic acid solution was added. After 3 hours, the reaction was quenched by the addition of DTT (5.6 µl of a 115 mM stock solution in H_2O, 15 eq. w.r.t. Pd), shaken for 40 minutes at 37 °C, then analysed by LC-MS. No reaction was observed.



M_{calc} (GYG-Y195F): 32032
M_{obs}: 32029

2.8.6 An "Enzyme-compatible" SMC: Optimisations on GYG-wt

2.8.6.1 Method for Analysis of GYG-wt Autoglucosylation from LC-MS Spectra

LC-MS data obtained from autoglucosylation assays of GYG-wt was used to determine the average number of glucoses attached to the cross-coupled enzyme according to the method previously outlined by Hurley *et al.*[50] This has been shown to agree well with alternative methods (e.g. radioassay),[50] and has been used in other studies to determine GYG-wt glucosylation extent.[37,38] See ref. 50 for a full description of the method, however it is reviewed briefly below.

The average number of glucoses attached to the enzyme either before or after glucosylation assay is obtained as follows. The proportion of each glycoform – as a fraction of the total glycoforms present – is obtained from the intensity of its corresponding peak in LC-MS spectrum, relative to the combined intensities of all such MS peaks:

$$proportion\ of\ glycoform = \frac{MS\ peak\ intensity\ of\ glycoform}{Total\ MS\ peak\ intensities\ of\ all\ glycoforms}$$

The above is used to construct the graphs shown repeatedly in this chapter, illustrating abundance of different glycoforms before and after glucosylation assay (proportions are there expressed as percentages).

For each glycoform, this proportion (as a value between 0 and 1, i.e. 1 = 100 % of the mixture) is multiplied by the number of glucoses associated with it. Summing the values thus obtained across all glycoforms yields the average number of glucose residues attached. This may be represented by the formula below, in which "i" refers to the number of glucose residues associated with a given glycoform:

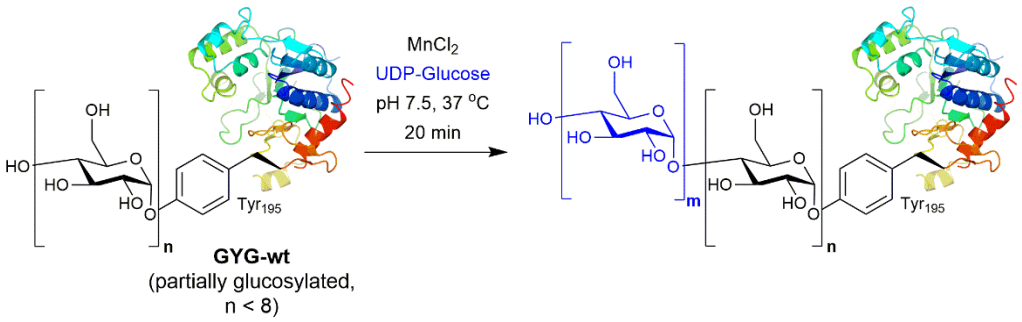
$$Average\ no.of\ glucoses\ attached = \sum_{i=1}^{n} i \times (proportion\ of\ glycoform)_i$$

This is effectively a weighted average, in which the "weights" are the number of glucose residues (i).

Therefore, the average number of glucoses *added* to GYG-wt during an assay, as quoted frequently in this chapter, is simply obtained from subtracting the number of glucoses attached before glucosylation assay from the number attached after assay, as obtained using the above method.

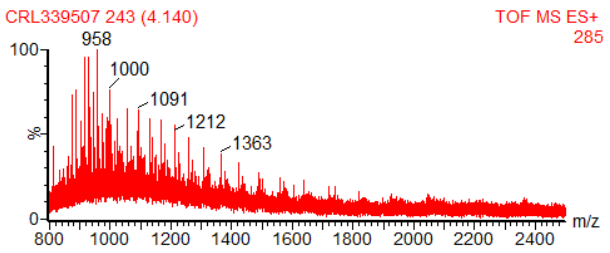
This protocol is used in all analyses of GYG-wt glucosylation detailed in the following sections. All "glycoform distribution" graphs originate from LC-MS spectra analysed according to this protocol.

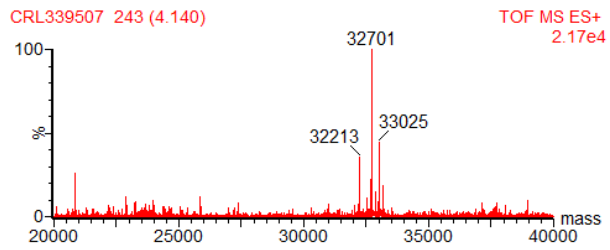
2.8.6.2 Autoglucosylation Activity of GYG-wt



GYG-wt in pH 7.5 25 mM Na-HEPES buffer was incubated with $MnCl_2$ and UDP-Glucose disodium salt at final concentrations of 1 mM each (added from 10 mM stock solution in H_2O). The reaction was incubated at 37 °C with shaking (500 rpm) for 20 minutes. After this time, Na-EDTA was added to a final concentration of 20 mM to quench glucosylation, and the reaction mixture analysed by LC-MS (see section 2.8.6.1).

Before assay:

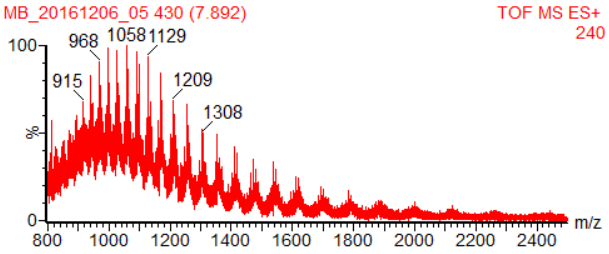


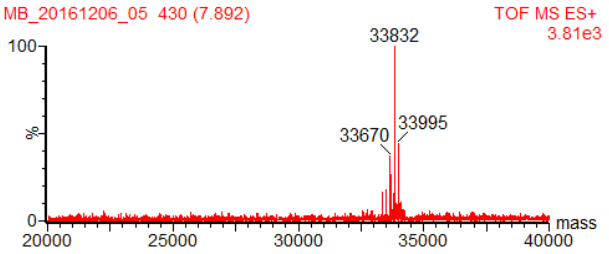


M_{calc}: 32048 + 162n
M_{obs}: 32051 + 162n

(n = glucoses attached = 1-7,
average = **4.07**)

After assay:

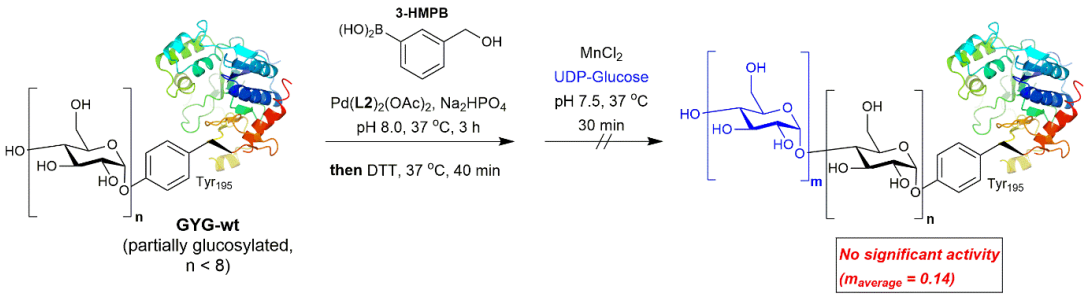




M_{calc}: 32048 + 162(n+m)
M_{obs}: 32050 + 162(n+m)

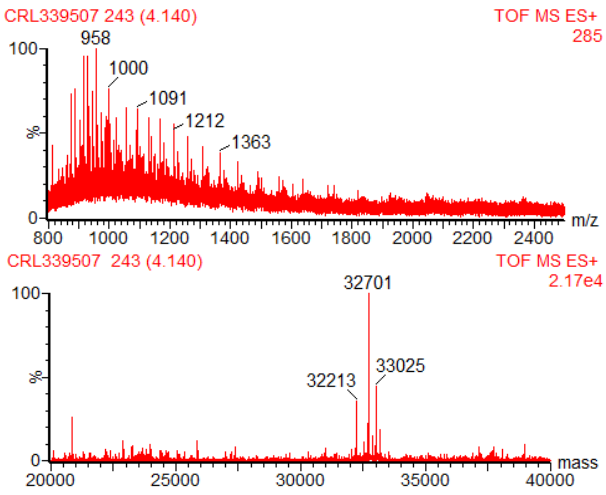
(n+m = glucoses attached = 8-12,
average = **10.74**)

2.8.6.3 GYG-wt Actvity after SMC Treatment: Initial Attempt



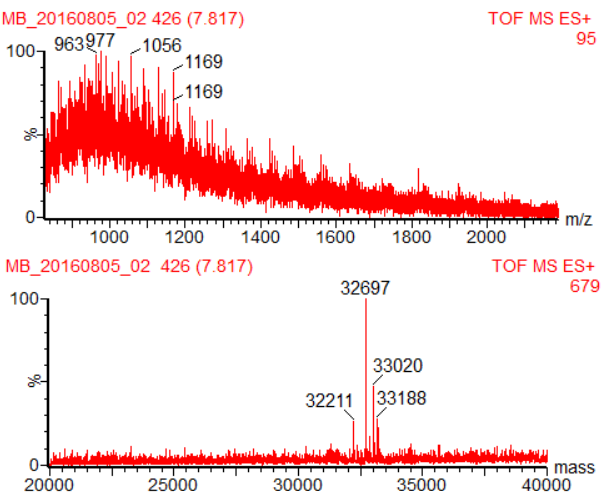
A stock solution of 3-(hydroxymethyl)phenylboronic acid (3-HMPB) was prepared by dissolving boronic acid (15.6 mg/ml) and Na_2HPO_4 (15.7 mg/ml) in H_2O. To His-tagged GYG-wt (70 µl of 0.6 mg/ml, 1.3 nmol) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was added boronic acid solution (16.8 µl, 1720 nmol) and $Pd(L2)_2(OAc)_2$ catalyst (8.6 µl of 10 mM stock solution, 86 nmol, 0.9 mM final concentration). The mixture was incubated at 37 °C for 3 hours (600 rpm). After this time, DTT was added as a Pd scavenger (11.0 µl of 115 mM stock solution in H_2O, 12 mM concentration). The reaction was shaken for a further 40 minutes, before being buffer exchanged into pH 7.5 25 mM Na-HEPES via PD SpinTrap G-25. 40 µl of resultant reaction mixture was incubated with $MnCl_2$ and UDP-Glucose disodium salt at final concentrations of 1 mM each (5 µl of 10 mM stock solution in H_2O was added). The reaction was incubated at 37 °C (500 rpm) for 30 minutes. Na-EDTA was then added to a final concentration of 20 mM to quench glucosylation, and the reaction mixture analysed by LC-MS (see section 2.8.6.1). Very little autoglucosylation was observed. LC-MS data are shown below; very little difference is apparent before and after glucosylation. See **figure 2.13** for graphical representation.

Before assay:



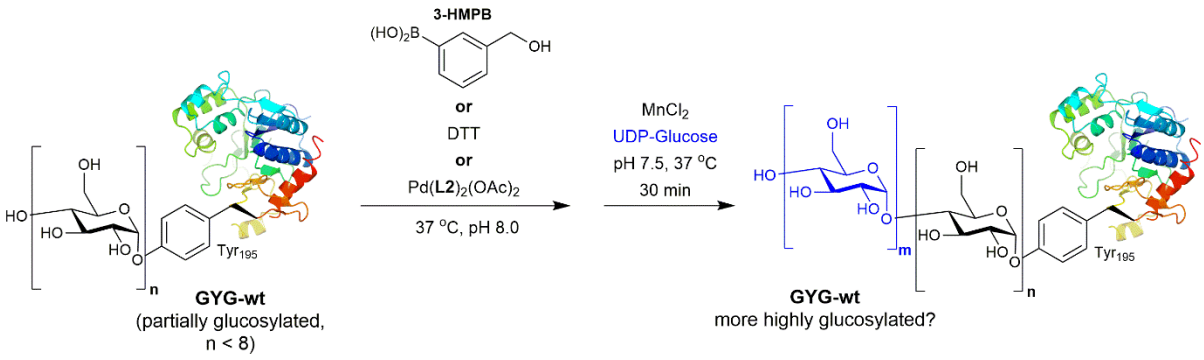
Mcalc: 32048 + 162n
Mobs: 32051 + 162n
(n = glucoses attached = 1-7,
average = **4.07**)

After assay:



Mcalc: 32048 + 162(n+m)
Mobs: 32049 + 162(n+m)
(n+m = glucoses attached = 1-7,
average = 4.21)

2.8.6.4 Effect of Individual SMC Components on GYG-wt Activity



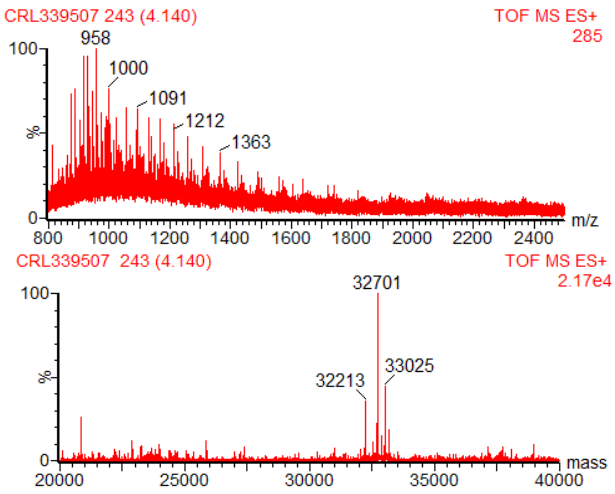
3 reactions were set up in parallel (and also in parallel to that in section 2.8.6.3). In each case, the reaction was carried out exactly as per the protocol described above in section 2.8.6.3, except for the following reaction-specific differences:

Reaction 1 (boronic acid): Palladium catalyst was replaced with an equal volume of H_2O. DTT was not added. A total of **6.34 glucose residues were added**.

Reaction 2 (Pd/DTT): Boronic acid was replaced with an equal volume of H_2O (DTT was still added). A total of **2.32 glucose residues were added.**

Reaction 3 (DTT only): Both boronic acid *and* palladium catalyst were replaced by equal volumes of H_2O. DTT was still added. A total of **7.10 glucose residues were added**.

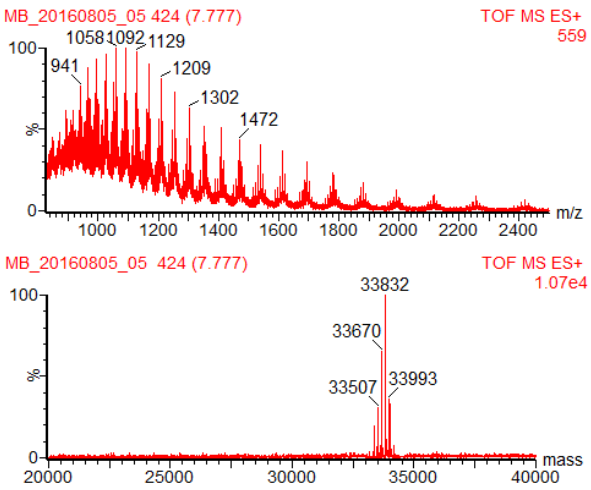
***Boronic Acid Only*: Before assay:**



M_{calc}: 32048 + 162n
M_{obs}: 32051 + 162n

(n = glucoses attached = 1-7,
average = **4.07**)

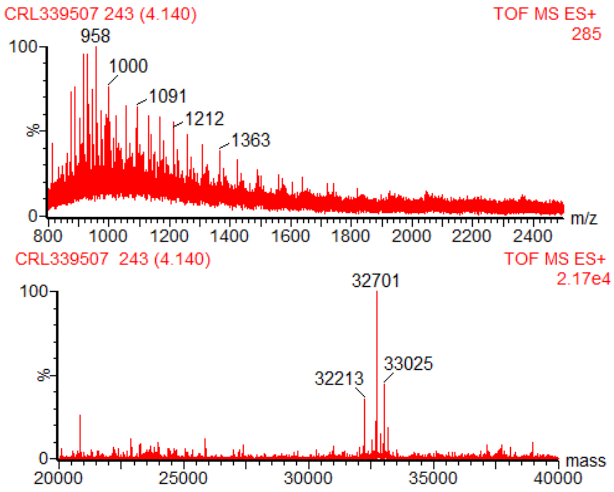
After assay:



M_{calc}: 32048 + 162(n+m)
M_{obs}: 32050 + 162(n+m)

(n+m = glucoses attached = 8-12,
average = **10.41**)

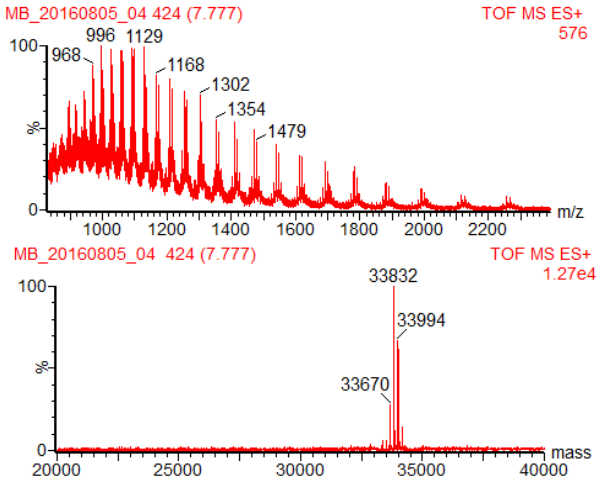
***DTT only*: Before assay:**



M_{calc}: 32048 + 162n
M_{obs}: 32051 + 162n#

(n = glucoses attached = 1-7,
average = **4.07**)

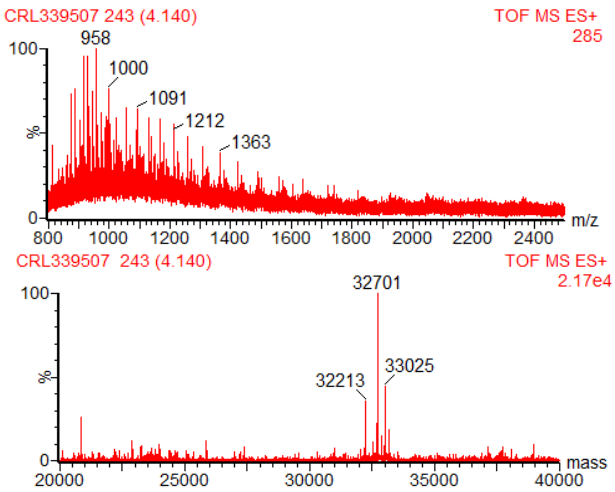
After assay:



M_{calc}: 32048 + 162(n+m)
M_{obs}: 32050 + 162(n+m)

(n+m = glucoses attached = 8-12,
average = **11.17**)

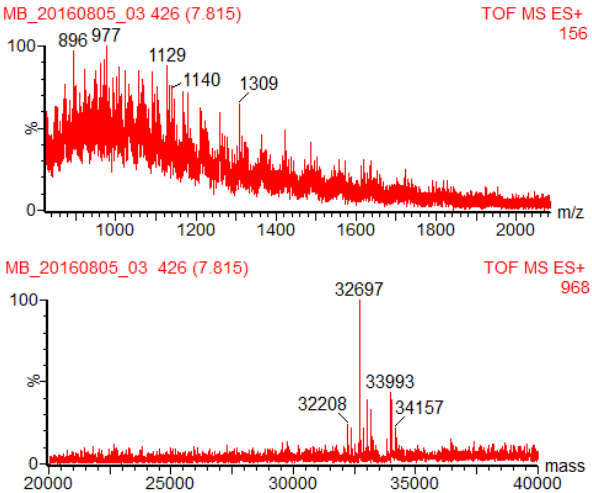
***Pd + DTT*: Before assay:**



M_{calc}: 32048 + 162n
M_{obs}: 32051 + 162n

(n = glucoses attached = 1-7,
average = **4.07**)

After assay:

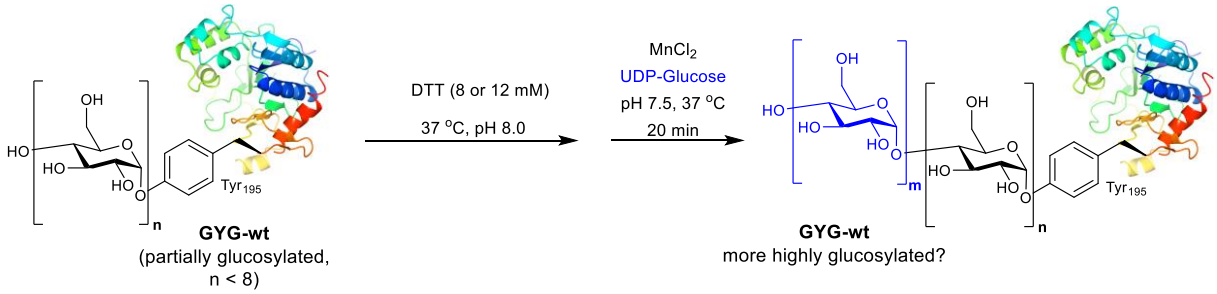


M_{calc}: 32048 + 162(n+m)
M_{obs}: 32049 + 162(n+m)

(n+m = glucoses attached = 1-13,
average = **6.39**)

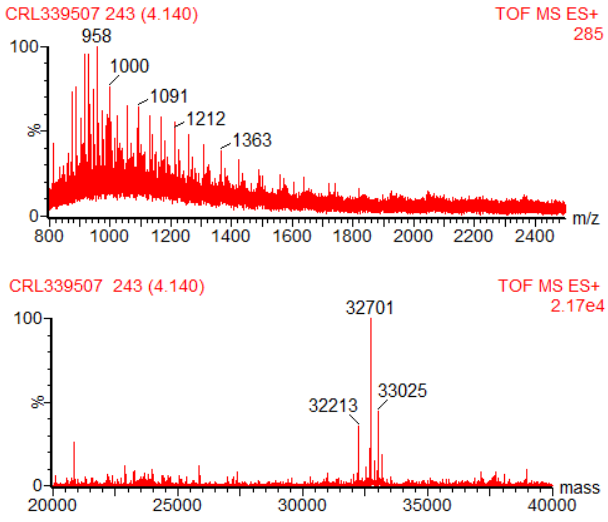
2.8.6.5 Effect of DTT Concentration on Catalytic Activity: With and Without Pd

a) Treatment with different concentrations of DTT, in isolation



His-tagged GYG-wt (0.6 mg/ml) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was incubated with DTT at final concentrations of either 8 or 11.5 mM, in a final volume of 75 µl (DTT was added from a 115 mM stock solution). Reactions were run in parallel to ensure conditions were otherwise identical. Samples were incubated at 37 °C for 1 hour with shaking at 600 rpm, before being buffer exchanged into pH 7.5 25 mM Na-HEPES via PD SpinTrap G-25. 40 µl of resultant reaction mixture was incubated with $MnCl_2$ and UDP-Glucose disodium salt at final concentrations of 1 mM each (5 µl of 10 mM stock solution in H_2O was added). The reaction was incubated at 37 °C with shaking (500 rpm) for 20 minutes. After this time, Na-EDTA was added to a final concentration of 20 mM to quench glucosylation, and the reaction mixture analysed by LC-MS (see section 2.8.6.1). Differences were slight, however a small decrease in glucosylation extent was apparent at higher DTT concentrations, and this was also consistent with the results obtained in section (b) below.

11.5 mM DTT, Before assay:

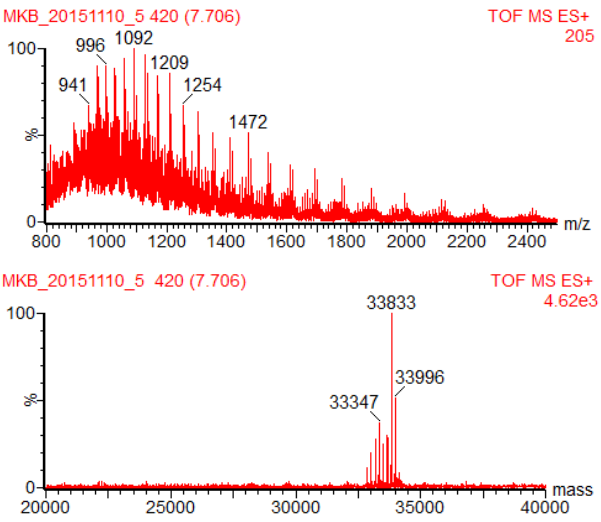


M_{calc}: 32048 + 162n

M_{obs}: 32051 + 162n

(n = glucoses attached = 1-7, average = **4.07**)

After assay:



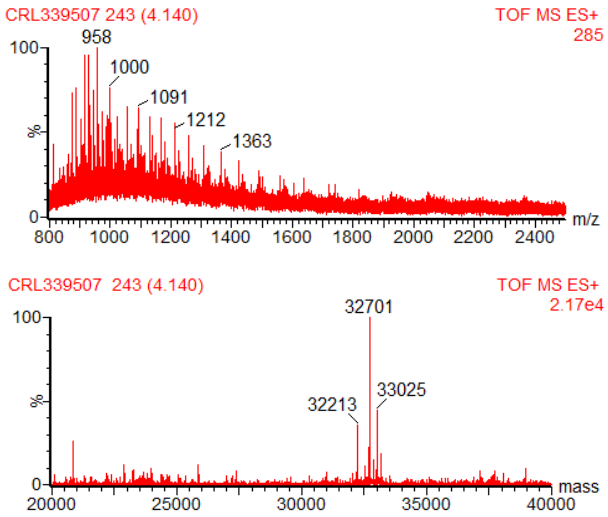
M_{calc}: 32048 + 162n

M_{obs}: 32051 + 162(n+m)

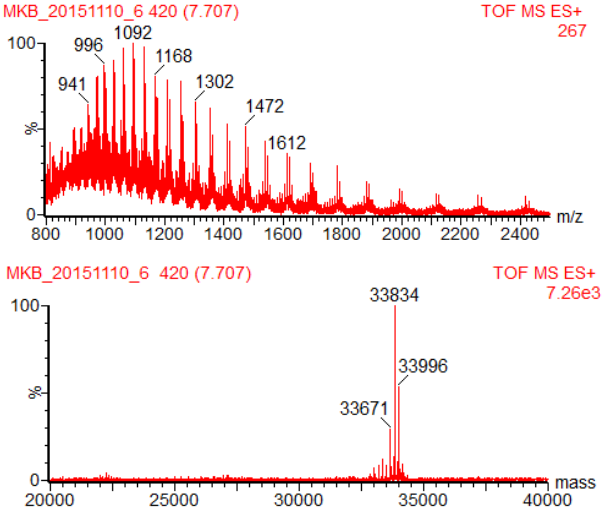
(n+m = glucoses attached = 5-12, average = **9.62**)

5.55 glucoses added.

8 mM DTT, before assay:



After assay:



M_{calc}: 32048 + 162n
M_{obs}: 32051 + 162n

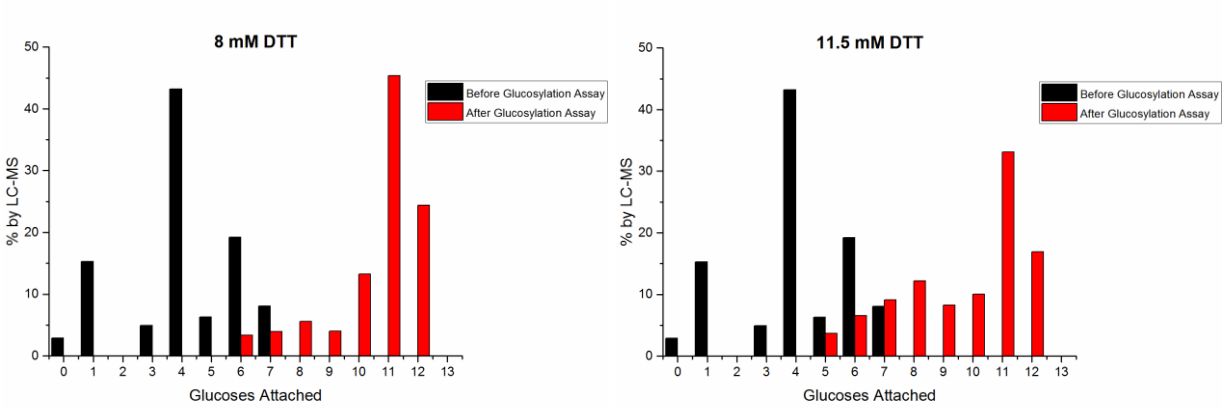
(n = glucoses attached = 1-7, average = **4.07**)

M_{calc}: 32048 + 162n
M_{obs}: 32051 + 162(n+m)

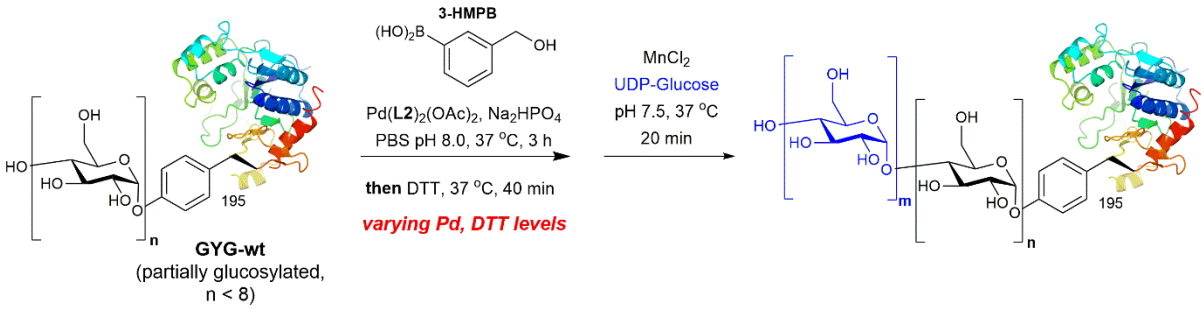
(n+m = glucoses attached = 6-12, average = **10.54**)

6.47 glucoses added.

Graphical summary:



b) Treatment with different concentrations of DTT, with other reagents



2 reactions were set up. A stock solution of boronic acid 3-HMPB was prepared by dissolving boronic acid (15.6 mg/ml) and Na_2HPO_4 (15.7 mg/ml = 110 mM, 1.1 eq. w.r.t. boronic acid) in H_2O. To his-tagged his-tagged GYG-wt (70 µl of 0.6 mg/ml, 1.30 nmol) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was added 19 µl of this stock solution (1950 nmol, 1500 eq. boronic acid). $Pd(L2)_2(OAc)_2$ catalyst was then added from a 10 mM stock solution (7.9 µl, 79 nmol, 0.8 mM final concentration), the reactions briefly vortexed, then incubated at 37 °C with shaking (600 rpm). After 3 hours, DTT was added to each reaction from a 115 mM solution in H_2O, as follows:

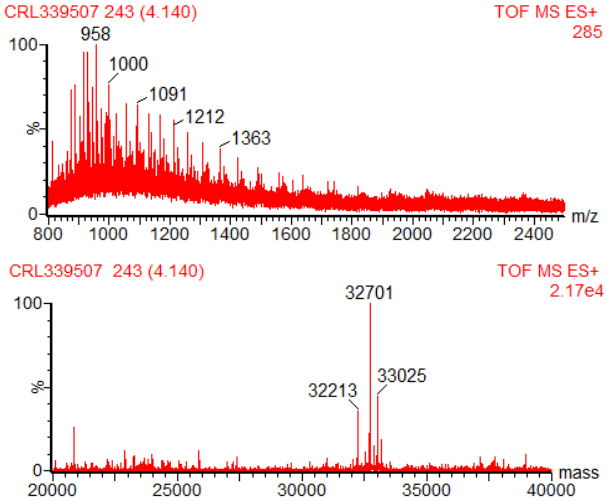
Reaction	Final concentration DTT / mM
1	**8** (= 7.3 µl 115 mM stock solution)
2	**11.5** (= 11 µl 115 mM stock solution)

The mixtures were shaken for a further 40 minutes, then buffer exchanged into pH 7.5 25 mM Na-HEPES by PD SpinTrap G-25. 100 µl of the sample was further desalted by dialysis against the same buffer (10000 MWCO, 4 °C, 2 h then 2 h then overnight, 3 x 300 ml buffer), using a Slide-A-Lyzer™ MINI dialysis device.

40 µl of resultant reaction mixture was incubated with $MnCl_2$ and UDP-Glucose disodium salt at final concentrations of 1 mM each (5 µl of 10 mM freshly prepared stock solution in H_2O was added). The reaction was incubated at 37 °C with shaking (500 rpm) for 30 minutes. After this time, Na-EDTA was added to a final concentration of 20 mM to quench glucosylation. Reactions were analysed by LC-MS to determine extent of glucosylation (see section 2.8.6.1). A slightly higher proportion of seemingly unreacted enzyme was present at the higher DTT concentrations.

0.8 mM Pd, 8 mM DTT

GYG-wt, before glucosylation assay

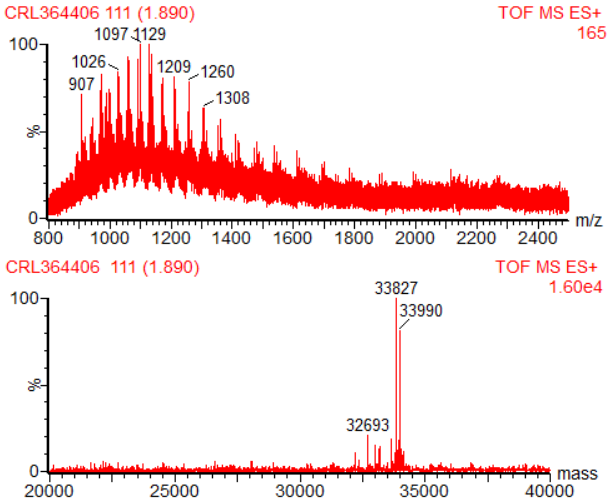


M_{calc}: 32048 + 162n

M_{obs}: 32051 + 162n

(n = glucoses attached = 0-7, average = **4.07**)

GYG-wt, after glucosylation assay



M_{calc}: 32048 + 162(n+m)

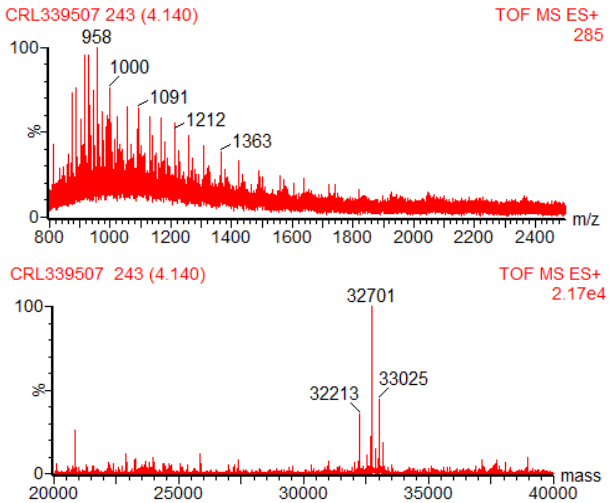
M_{obs}: 32045 + 162(n+m)

(n+m = glucoses attached = 1-13 average = **9.90**)

5.83 glucoses added.

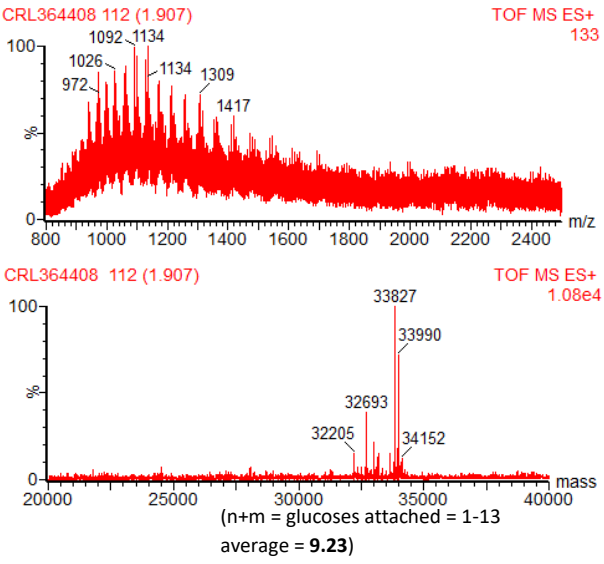
0.8 mM Pd, 11.5 mM DTT

GYG-wt, before glucosylation assay



(n = glucoses attached = 0-7, average = **4.07**)

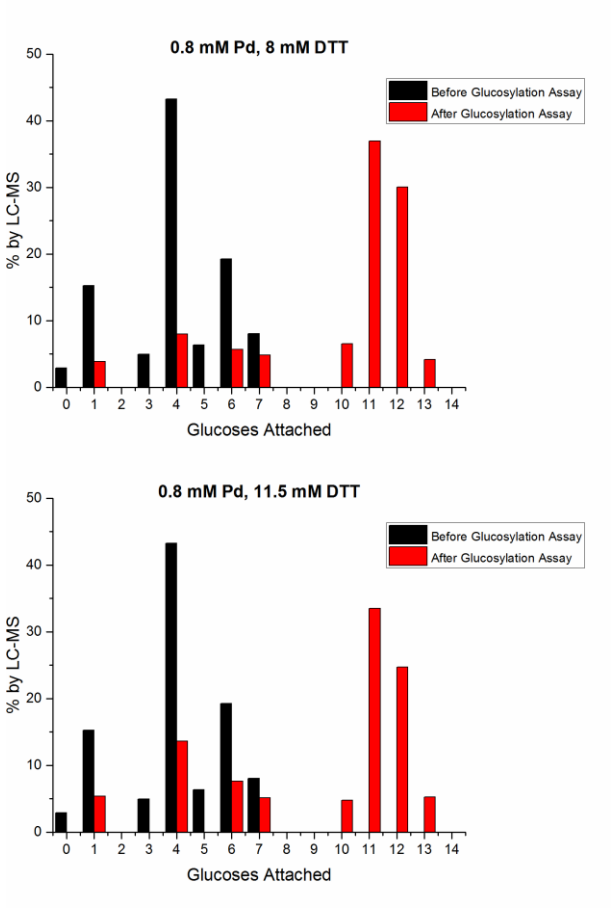
GYG-wt, after glucosylation assay



(n+m = glucoses attached = 1-13 average = **9.23**)

5.16 glucoses added.

Graphical summary:



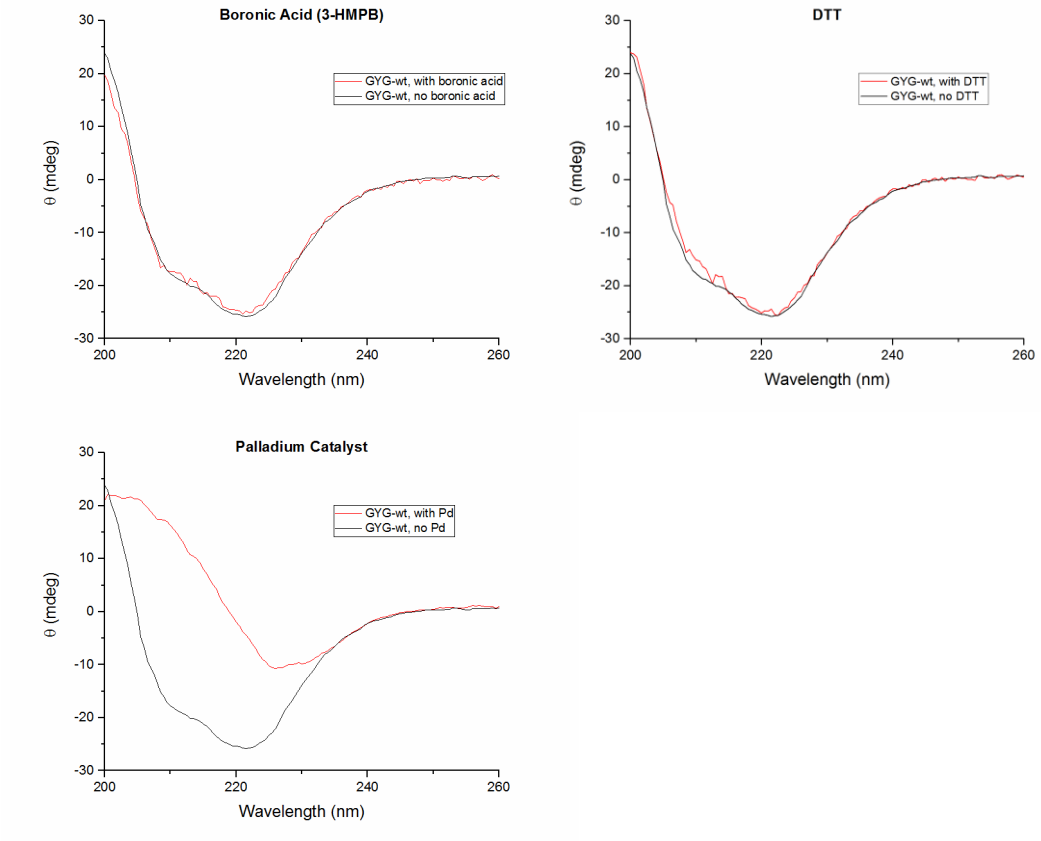
2.8.6.6 Circular Dichroism: Effect of SMC Reaction Components

Boronic Acid or DTT

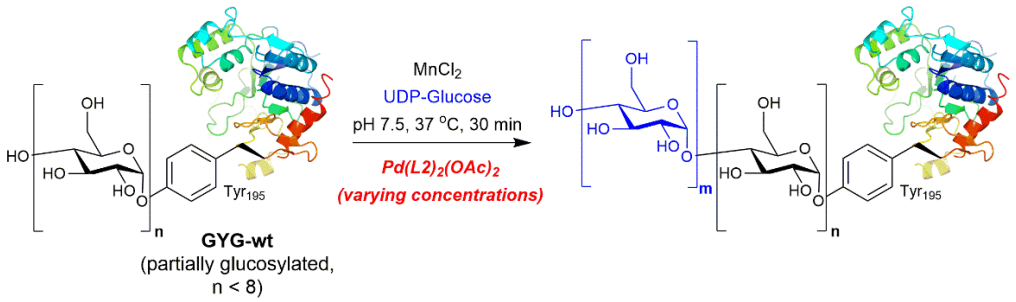
A stock solution of 3-(hydroxymethyl)phenylboronic acid (3-HMPB) was prepared by dissolving boronic acid (15.6 mg/ml) and Na_2HPO_4 (15.7 mg/ml = 110 mM) in H_2O. To His-tagged GYG-wt (100 µl of 0.6 mg/ml, 1.9 nmol) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was added *either* boronic acid solution (16.8 µl, 1720 nmol, 1320 eq.) *or* DTT (final concentration 12 mM, added from a 115 mM stock solution). The reactions were incubated for 1 hour (37 °C, 600 rpm) and were then buffer exchanged into CD buffer (50 mM NaH_2PO_4, 5 % v/v glycerol, pH 8.0) by PD SpinTrap G-25. Samples were then analysed by CD at concentrations of 0.2 mg/ml, using CD buffer as background, whichw as subtracted from the sample. Data represent mean averages of 3 runs.

Palladium Catalyst

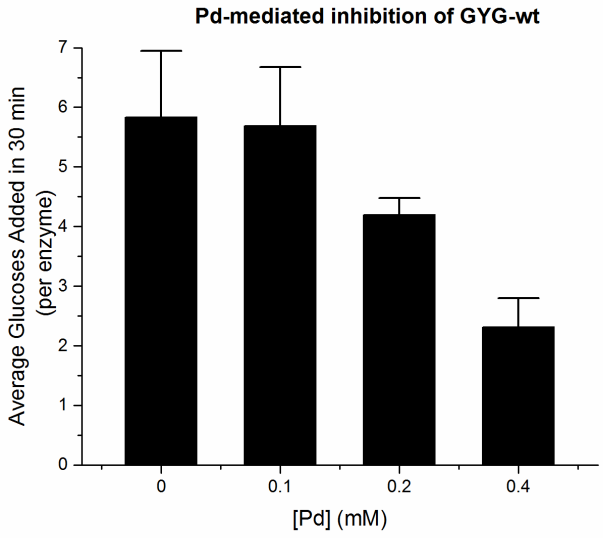
To His-tagged GYG-wt (100 µl of 0.6 mg/ml, 1.9 nmol) in CD buffer (50 mM NaH_2PO_4, 5 % v/v glycerol, pH 8.0) was added $Pd(L2)_2(OAc)_2$ catalyst (5.0 µl, 50 nmol, 0.5 mM). The protein was incubated for 1 hours (37 °C, 600 rpm), before being diluted to a concentration of 0.2 mg/ml with CD buffer and analysed by CD. The background was CD buffer containing an equal concentration of Pd catalyst, which was subtracted from the sample. Data represent mean averages of 3 runs.



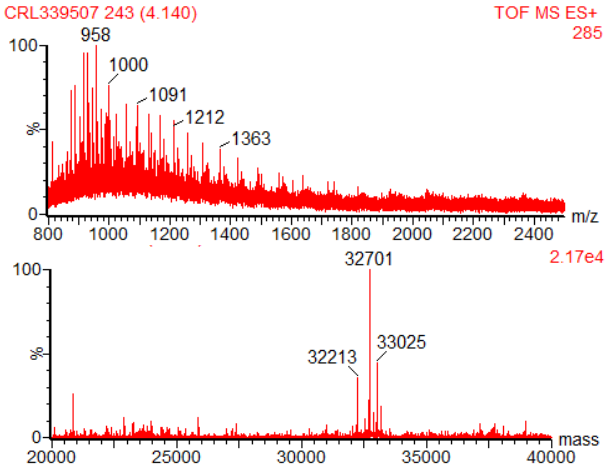
2.8.6.7 Concentration-dependent Effect of Palladium on GYG-wt Activity



4 reactions were set up. In each case, GYG-wt (0.4 mg/ml) in 25 mM Na-HEPES pH 7.5 was incubated with $MnCl_2$ (final concentration 1 mM), UDP-Glucose disodium salt (final concentration 1 mM). To 3 reactions, $Pd(L2)_2(OAc)_2$ was also added (final concentrations 0.4, 0.2, 0.1 mM). Reactions were incubated at 37 °C for 30 minutes with shaking (500 rpm). Glucosylation was then quenched by the addition of Na-EDTA to a final concentration of 20 mM. After a further 15 minutes shaking to ensure thorough quenching, DTT was added to scavenge Pd (final concentration ca. 20 mM), and the samples shaken for a further 30 minutes before being analysed by LC-MS to determine average numbers of glucoses attached (see section 2.8.6.1). Data points represent the average of three independent samples for each Pd concentration.



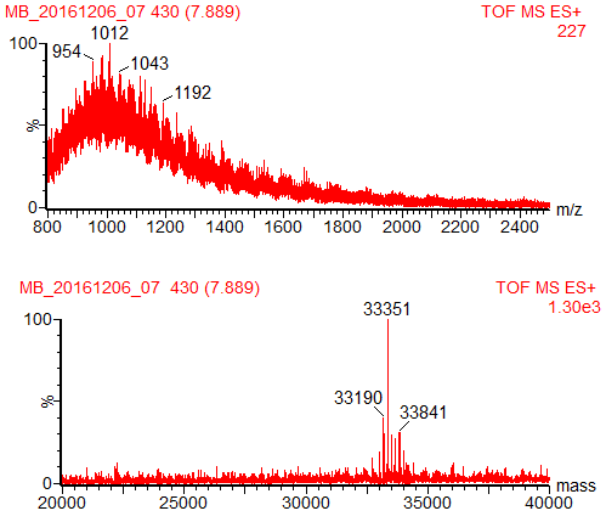
Example MS: before reaction



M_{calc}: 32048 + 162n
M_{obs}: 32051 + 162n

(n = glucoses attached = 1-7, average = **4.07**)

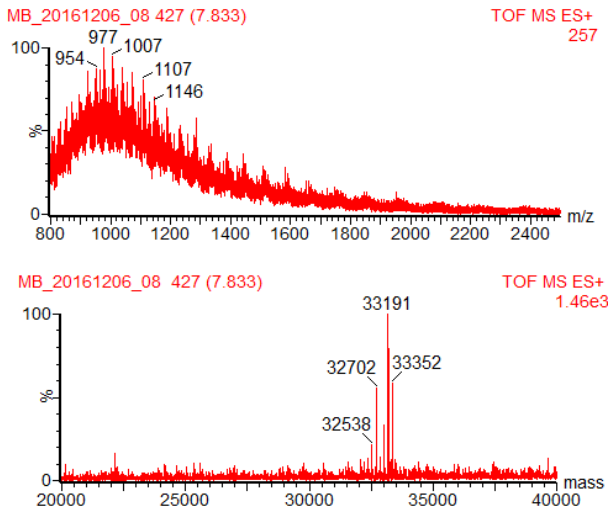
After reaction, 0.2 mM Pd:



M_{calc}: 32048 + 162(n+m)
M_{obs}: 32055 + 162(n+m)

(n+m = glucoses attached = 1-7, average = **8.41**)

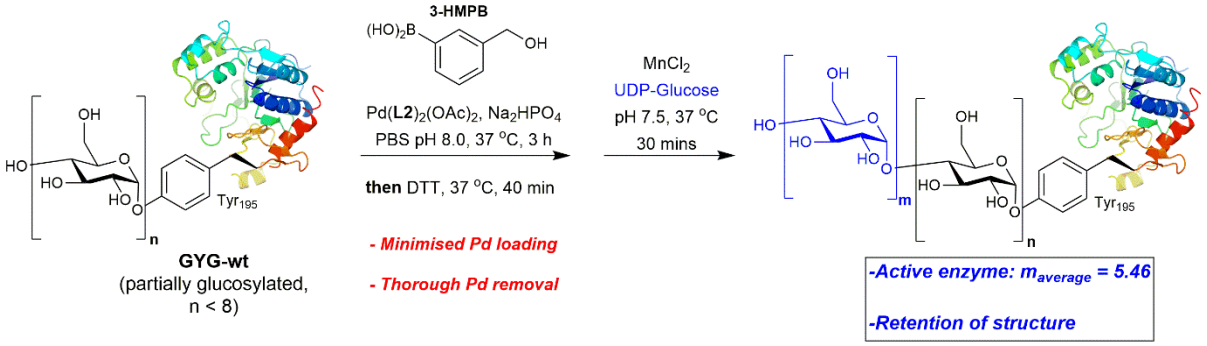
After reaction, 0.4 mM Pd:



M_{calc}: 32048 + 162(n+m)
M_{obs}: 32055 + 162(n+m)

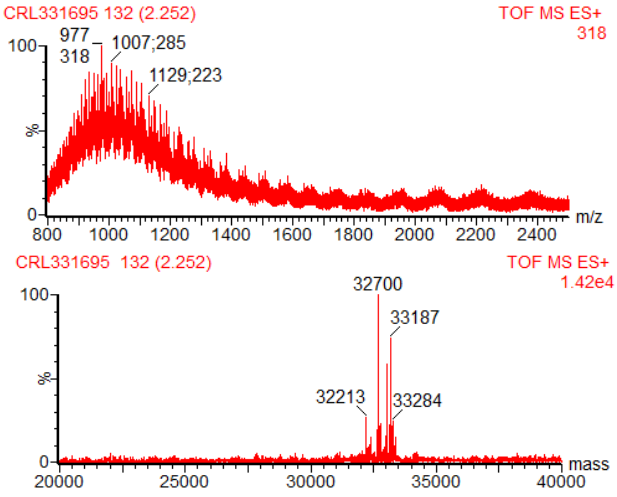
(n+m = glucoses attached = 1-7, average = **5.91**)

2.8.6.8 Optimised SMC Protocol for Retention of GYG-wt Activity (with 3-HMPB)



A stock solution of 3-(hydroxymethyl)phenylboronic acid (3-HMPB) was prepared by dissolving boronic acid (15.6 mg/ml) and Na_2HPO_4 (15.7 mg/ml = 110 mM) in H_2O. To His-tagged GYG-wt (100 µl of 0.6 mg/ml, 1.9 nmol) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was added boronic acid solution (27 µl, 2770 nmol) and $Pd(L2)_2(OAc)_2$ catalyst (11.2 µl of 10 mM stock solution, 0.8 mM final concentration). The mixture was incubated at 37 °C for 3 hours (600 rpm). After this time, DTT was added as a Pd scavenger (10.2 µl of a 115 mM stock solution in H_2O, 8 mM concentration). The reaction was shaken for 40 minutes, before being buffer exchanged into pH 7.5 25 mM Na-HEPES via PD SpinTrap G-25. 100 µl of the sample was further desalted by dialysis against the same buffer (10000 MWCO, 4 °C, 2 h then 2 h then overnight, 3 x 300 ml buffer), using a Slide-A-Lyzer™ MINI dialysis device. The resultant reaction mixture was incubated with $MnCl_2$ and UDP-Glucose disodium salt (final concentrations 1 mM each, added from 10 mM stock solutions). The reaction was incubated at 37 °C with shaking (500 rpm) for 30 minutes. Na-EDTA was then added to a final concentration of 20 mM to quench glucosylation, and the reaction mixture analysed by LC-MS (see section 2.8.6.1). See **figure 2.17** for graphical representation.

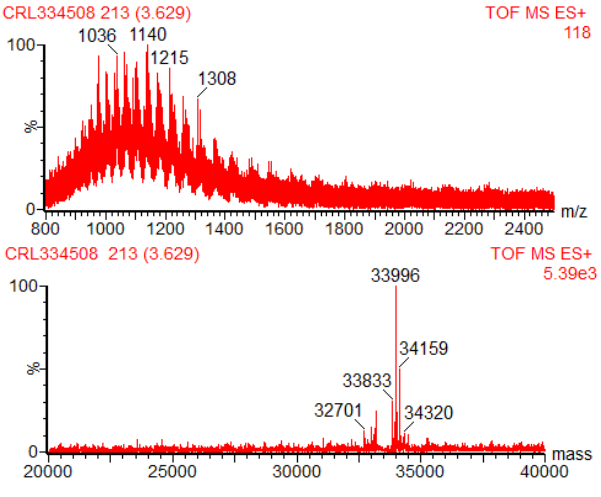
Before assay:



M_{calc}: 32048 + 162n
M_{obs}: 32050 + 162n

(n = glucoses attached = 1-7, average = **4.84**)

After assay:

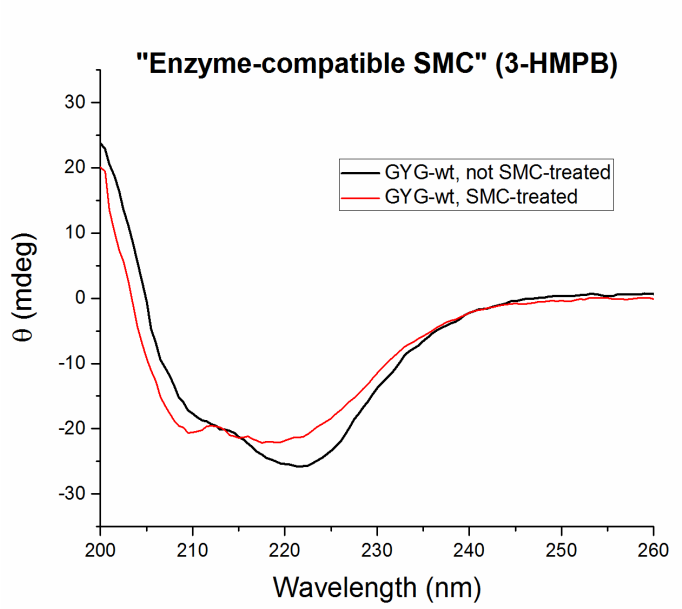


M_{calc}: 32048 + 162(n+m)
M_{obs}: 32050 + 162(n+m)

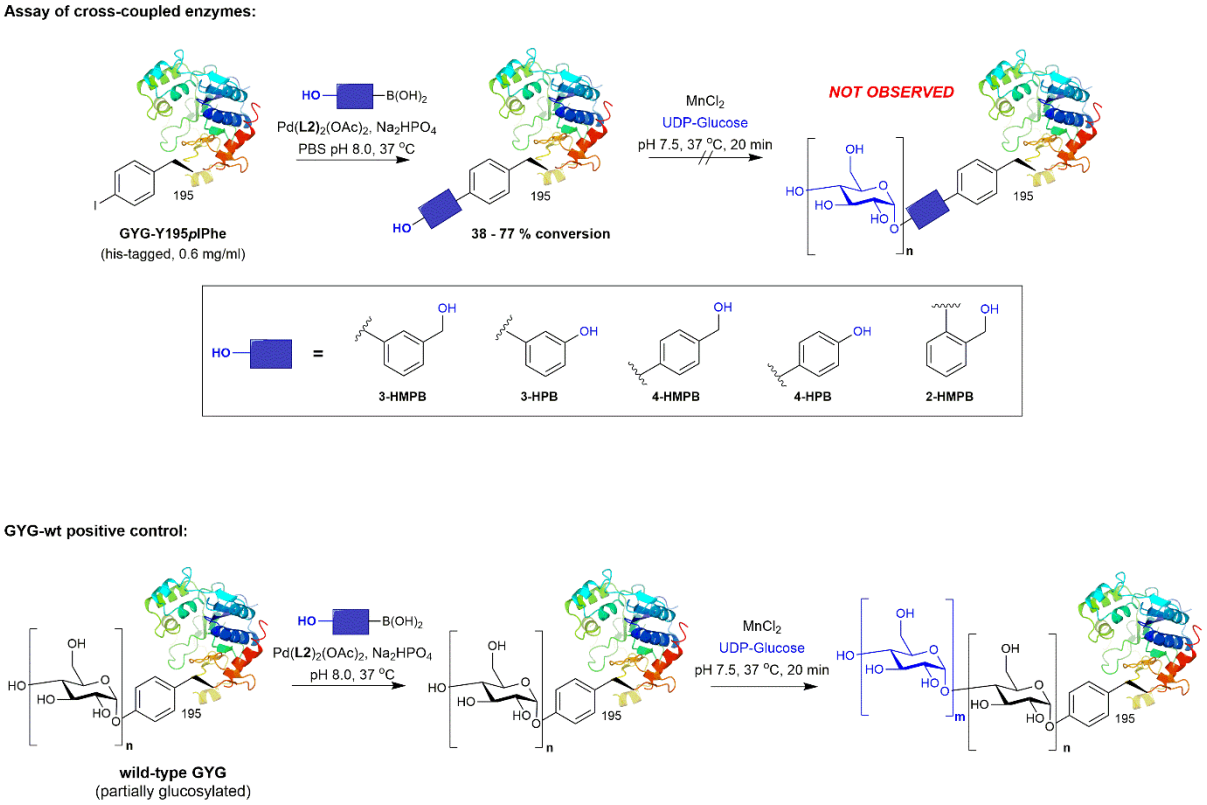
(n+m = glucoses attached = 1-13, average = **10.89**)

2.8.6.9 Circular Dichroism of GYG-wt after SMC Conditions

A stock solution of 3-(hydroxymethyl)phenylboronic acid (3-HMPB) was prepared by dissolving boronic acid (15.6 mg/ml) and Na_2HPO_4 (15.7 mg/ml = 110 mM) in H_2O. To His-tagged GYG-wt (100 µl of 0.6 mg/ml, 1.9 nmol) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was added boronic acid solution (27 µl, 2770 nmol) and $Pd(L2)_2(OAc)_2$ catalyst (11.2 µl of 10 mM stock solution, 0.8 mM final concentration). The mixture was incubated at 37 °C for 3 hours with shaking (600 rpm). After this time, DTT was added as a Pd scavenger (10.2 µl of a 115 mM stock solution in H_2O, 8 mM concentration). The reaction was shaken for a further 40 minutes, before being buffer exchanged into CD buffer (50 mM NaH_2PO_4, 5 % v/v glycerol, pH 8.0) via PD SpinTrap G-25. 100 µl of the sample was further desalted by dialysis against the same buffer (10000 MWCO, 4 °C, 2 h then 2 h then overnight, 3 x 300 ml buffer), using a Slide-A-Lyzer™ MINI dialysis device. The resultant reaction mixture was diluted to 0.2 mg/ml and the circular dichroism spectrum recorded (CD buffer was used for the background spectrum, which was subtracted from that of the sample). Data represents the mean average of 3 runs. The spectrum was compared to that previously recorded for GYG-wt in the absence of SMC conditions (section 2.8.3.4), revealing little significant difference.



2.8.7 Catalytic Activity of Minimal GYG Glycomimetics



2.8.7.1 General Procedure

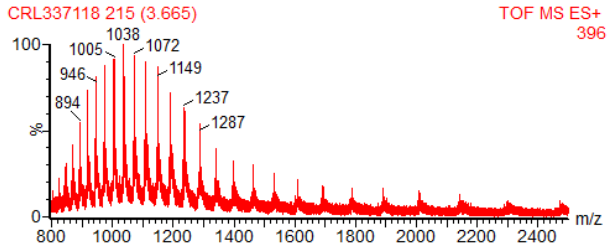
GYG-Y195*p*IPhe (His-tagged) *or* GYG-wt (His-tagged) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was combined with boronic acid (added from a stock solution in aq. Na_2HPO_4) and $Pd(L2)_2(OAc)_2$ catalyst (added from a 10 mM stock solution, final concentration 0.8 or 0.9 mM). The reaction mixture was briefly vortexed, then incubated at 37 °C with shaking (600 rpm). After 3 h, DTT was added to a final concentration of 8 mM (from a 115 mM stock solution) and the mixture shaken for a further 40 minutes. The mixture was buffer exchanged into pH 7.5 25 mM Na-HEPES by PD SpinTrap G-25; an aliquot was analysed by LC-MS. 100 µl of the sample was further desalted by dialysis against the same buffer (10000 MWCO, 4 °C, 2 h then 2 h then overnight, 3 x 300 ml buffer), using a Slide-A-Lyzer™ MINI dialysis device. 40 µl of resultant reaction mixture was incubated with $MnCl_2$ and UDP-Glucose disodium salt at final concentrations of 1 mM each (5 µl of 10 mM freshly prepared stock solution in H_2O was added). The reaction was incubated at 37 °C with shaking (500 rpm) for 30 minutes. After this time, Na-EDTA was added to a final concentration of 20 mM to quench glucosylation. Reactions were analysed by LC-MS (see section 2.8.6.1 for analysis method for GYG-wt glucosylation). ***Note: to ensure exact equivalence of conditions, both the GYG-Y195pIPhe and GYG-wt samples (referred to in sections below) were treated in parallel.***

2.8.7.2 GYG-3HMPB (boronic acid = 3-HMPB)

Carried out according to general procedure (section 2.8.7.1) using the following reagent amounts:

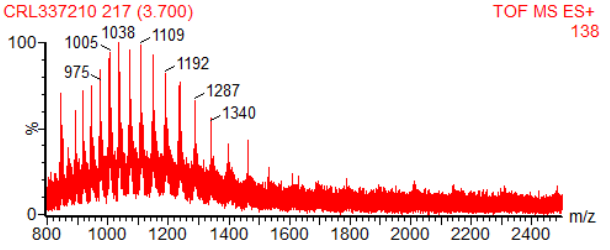
- GYG-Y195*p*IPhe or GYG-wt: 100 µl (0.6 mg/ml), 1.9 nmol
- Boronic acid **3-HMPB**: 27 µl of a 15.6 mg/ml stock solution, 2770 nmol, 1500 eq. Stock solution was prepared by dissolving boronic acid and Na_2HPO_4 (15.7 mg/ml = 110 mM) in H_2O.
- $Pd(L2)_2(OAc)_2$: 11 µl of 10 mM stock solution, 110 nmol, 59 eq., 0.8 mM final concentration.

GYG-3HMPB, after cross-coupling (77 % conversion):



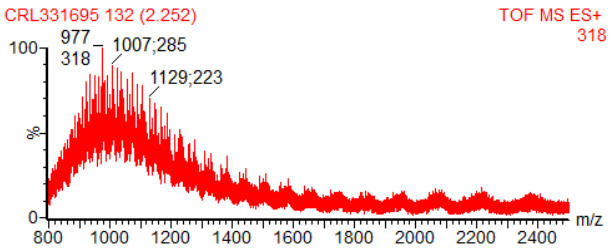
M_{calc}: 32138
M_{obs}: 32141 (77 %)
32034 (deiodination)

GYG-3HMPB, after glucosylation assay:



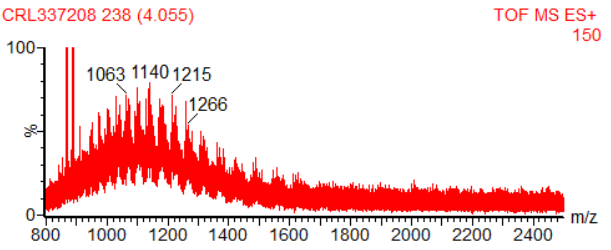
M_{calc} 32138 + 162n
M_{obs}: 32141 (non-glucosylated)
32034 (deiodinated)

GYG-wt, before glucosylation assay



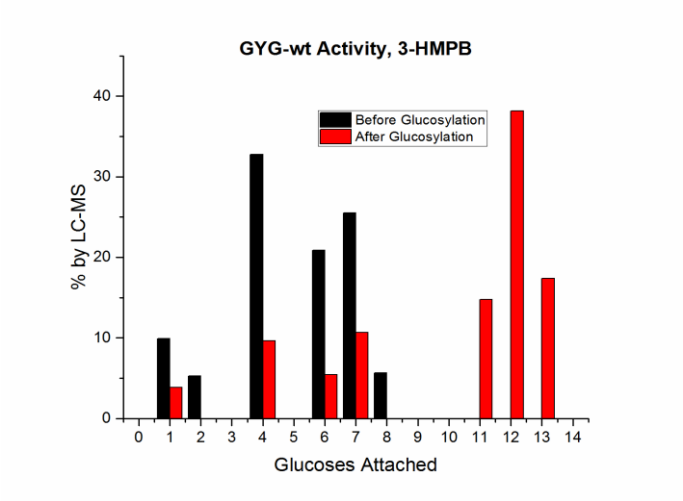
M_{calc}: 32048 + 162n
M_{obs}: 32050 + 162n
(n = glucoses attached = 1-7,
average = **4.84**)

GYG-wt, after glucosylation assay



M_{calc}: 32048 + 162(n+m)
M_{obs}: 32045 + 162(n+m)
(n+m = glucoses attached = 1-12
average = **9.97**)

Graphical Summary of GYG-wt activity from above data:

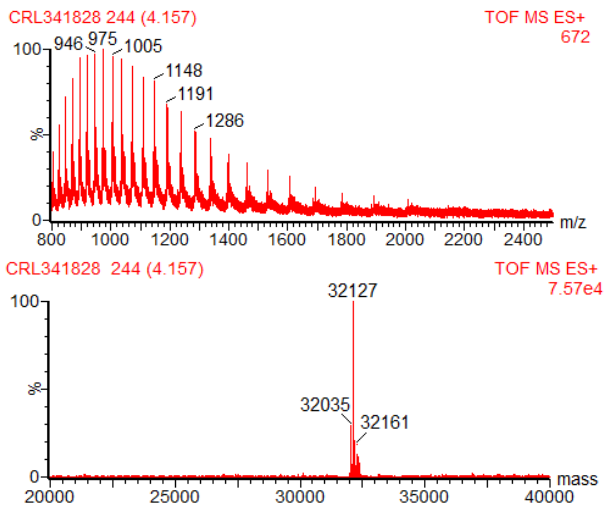


2.8.7.3 GYG-3HPB (boronic acid = 3-HPB)

Carried out according to general procedure (section 2.8.7.1) using the following reagent amounts:

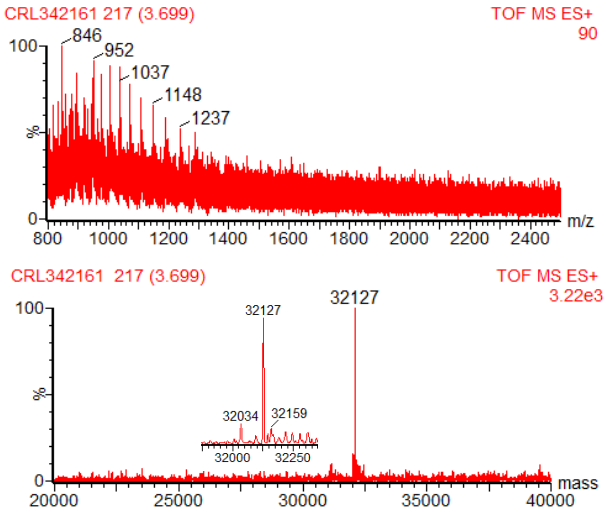
- GYG-Y195*p*IPhe or GYG-wt: 100 µl (0.6 mg/ml), 1.9 nmol
- Boronic acid **3-HPB**: 27 µl of a 14.2 mg/ml stock solution, 2770 nmol, 1500 eq. Stock solution was prepared by dissolving boronic acid and Na_2HPO_4 (15.7 mg/ml = 110 mM) in H_2O.
- $Pd(L2)_2(OAc)_2$: 11 µl of 10 mM stock solution, 110 nmol, 59 eq., 0.8 mM final concentration.

GYG-3HPB, after cross-coupling (67 % conversion):



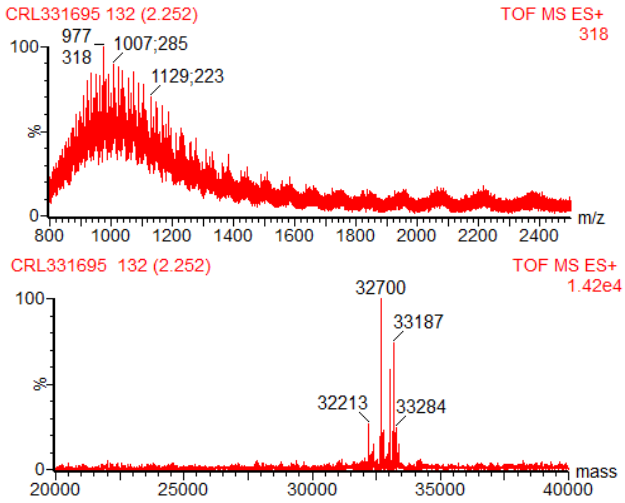
M_{calc}: 32124
M_{obs}: 32127 (67 %)
32035 (deiodination)
32161 (GYG-Y195*p*IPhe)

GYG-3HPB, after glucosylation assay:

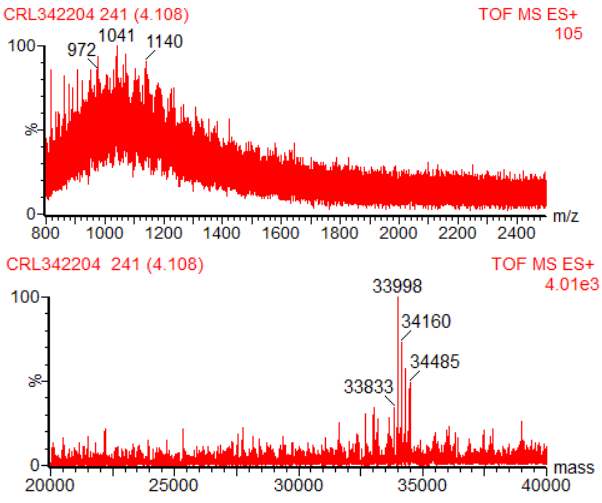


M_{calc} 32124 + 162n
M_{obs}: 32127 (non-glucosylated)
32035 (deiodinated)
32159 (GYG-Y195*p*IPhe)

GYG-wt, before glucosylation assay



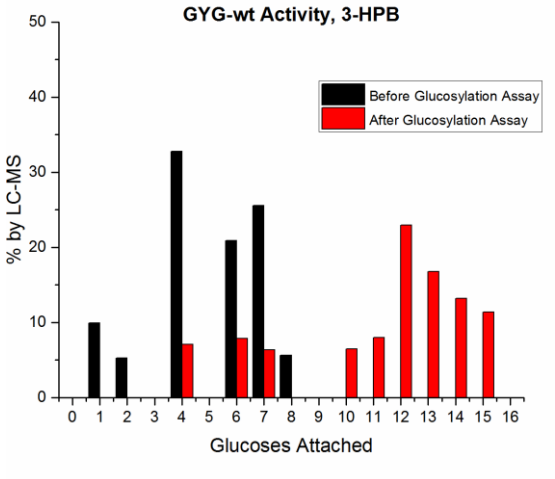
GYG-wt, after glucosylation assay



M_{calc}: 32048 + 162n
M_{obs}: 32050 + 162n
(n = glucoses attached = 1-7,
average = **4.84**)

M_{calc}: 32048 + 162(n+m)
M_{obs}: 32053 + 162(n+m)
(n +m = glucoses attached = 6-14
average = **11.98**)

Graphical Summary of GYG-wt activity from above data:

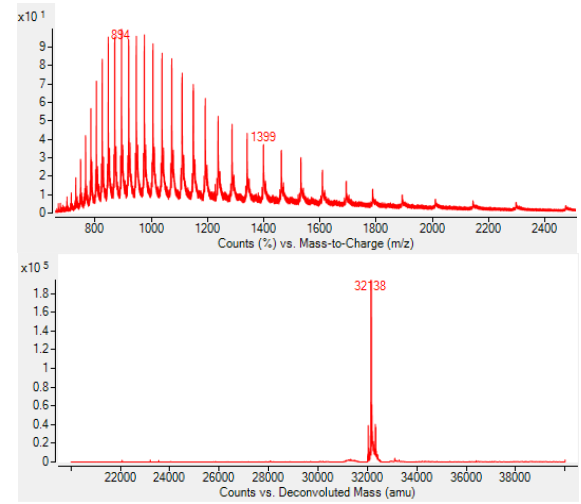


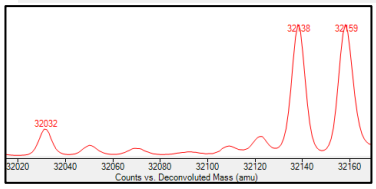
2.8.7.4 GYG-2HMPB (boronic acid = 2-HMPB)

Carried out according to general procedure (section 2.8.7.1) using the following reagent amounts:

- GYG-Y195*p*IPhe or GYG-wt: 40 µl (0.8 mg/ml), 1.0 nmol
- Boronic acid **2-HMPB**: 40 µl of a 6.7 mg/ml stock solution, 1940 nmol, 1950 eq. Stock solution was prepared by dissolving boronic acid and Na_2HPO_4 (7.4 mg/ml) in H_2O.
- $Pd(L2)_2(OAc)_2$: 8.0 µl of 10 mM stock solution, 80 nmol, 80 eq., 0.9 mM final concentration.

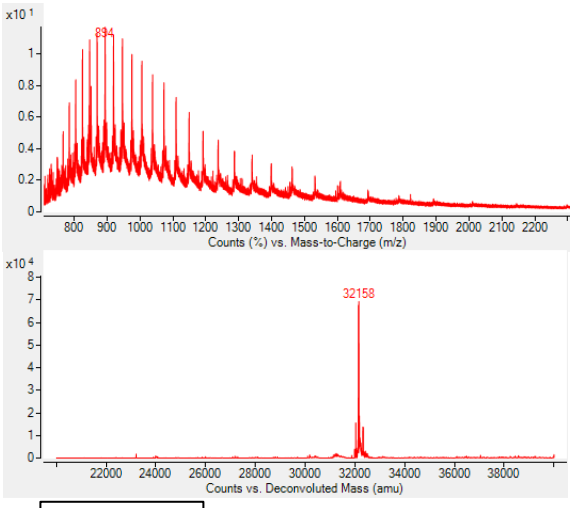
GYG-2HMPB, after cross-coupling (45 % conversion):

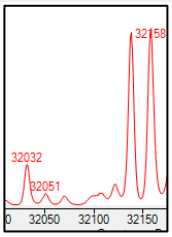




M_{calc}: 32138
M_{obs}: 32138 (45 %)
32032 (de-iodination)
32159 (GYG-Y195*p*IPhe)
Also see note below

GYG-2HMPB, after glucosylation assay:

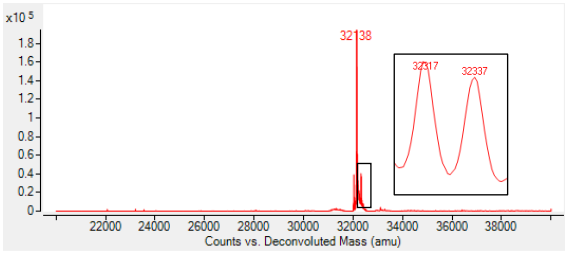




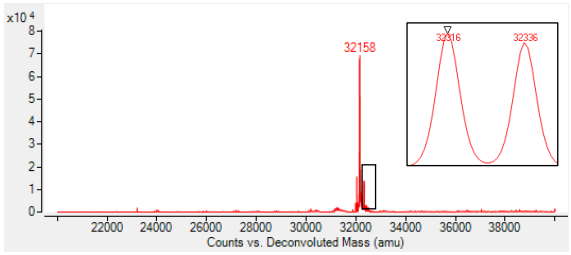
M_{calc} 32138 + 162n
M_{obs}: 32138 (non-glucosylated)
32032 (de-iodinated)
32158 (GYG-Y195*p*IPhe)
Also see note below

***Note**: it may be seen that small peaks at a higher Mw are present. As illustrated below and discussed in footnote (i), pp. 79, these reflect the fact that, as illustrated in **figure 2.4**, a small proportion of the GYG-Y195pIPhe starting enzyme may be gluconoylated (+178 Da) during expression. Typically, analogous adducts were not observable above background noise after cross-coupling, however in this case they are seen; the masses correspond to +178 Da adducts of cross-coupled enzyme (32316/7) and unreacted GYG-Y195pIPhe (32336/7). Note that reaction conversions refer only to reaction of majority, non-gluconoylated species (though they are not affected by inclusion of gluconoylated variant, which is firstly a minority species, and secondly appears to react identically). No conclusions concerning enzymatic activity are affected by the presence of this species; it is present in the same levels both prior to and following glucosylation, and the mass expected of glucose residue addition is +162 Da, not +178 Da.*

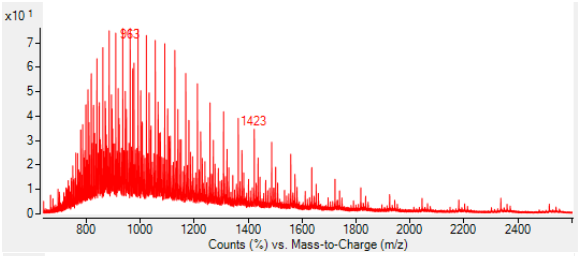
Cross-coupled (showing gluconoylation):



After glucosylation (showing gluconoylation):

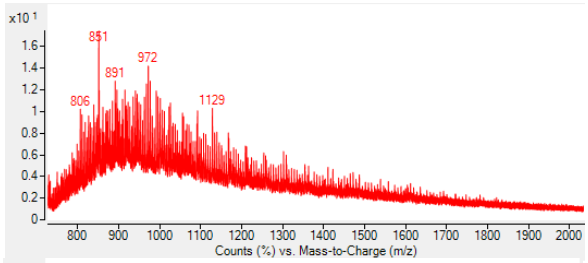


GYG-wt, before glucosylation assay



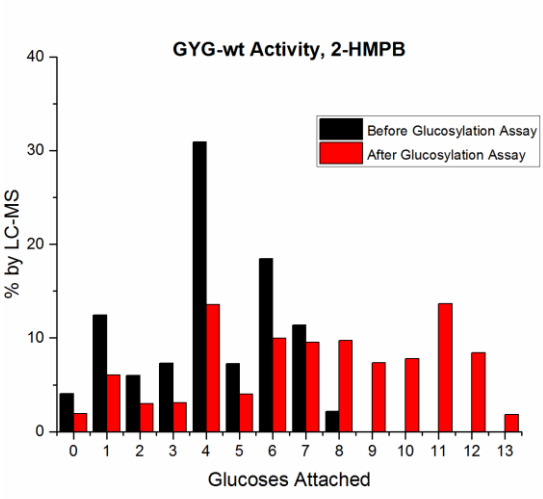
M_{calc}: 32048 + 162n
M_{obs}: 32050 + 162n
(n = glucoses attached = 0-8,
average = **4.14**)

GYG-wt, after glucosylation assay



M_{calc}: 32048 + 162(n+m)
M_{obs}: 32045 + 162(n+m)
(n+m = glucoses attached = 0-12
average = **7.19**)

Graphical Summary of GYG-wt activity from above data:

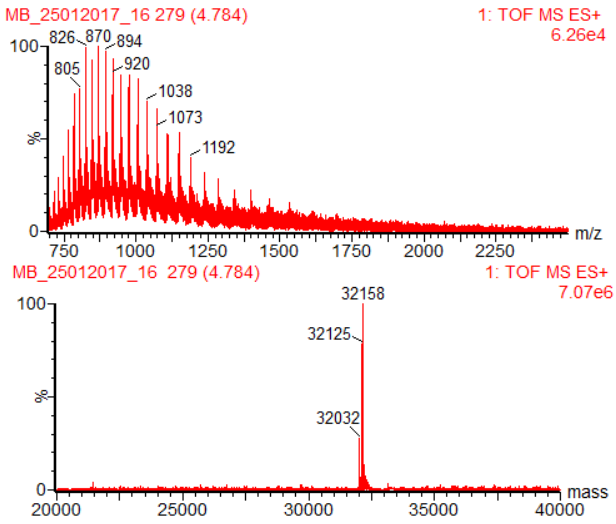


2.8.7.5 GYG-4HPB (boronic acid = 4-HPB)

Carried out according to general procedure (section 2.8.7.1) using the following reagent amounts:

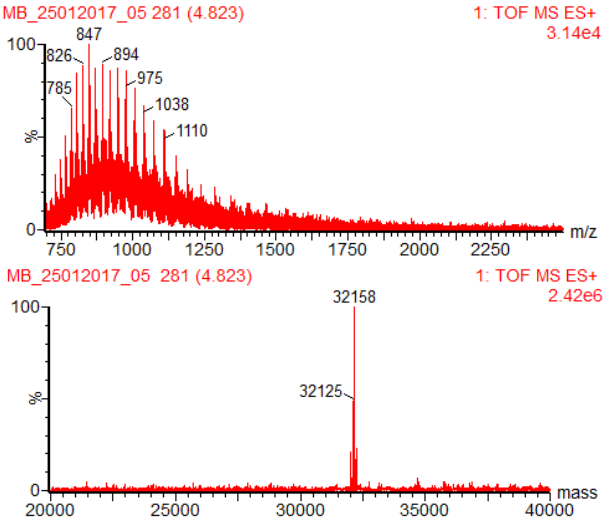
- GYG-Y195*p*IPhe or GYG-wt: 70 µl (0.6 mg/ml), 1.3 nmol
- Boronic acid **4-HPB**: 18.9 µl of a 14.2 mg/ml stock solution, 1940 nmol, 1500 eq. Stock solution was prepared by dissolving boronic acid and Na_2HPO_4 (7.4 mg/ml) in H_2O.
- $Pd(L2)_2(OAc)_2$: 7.7 µl of 10 mM stock solution, 77 nmol, 59 eq., 0.8 mM final concentration.

GYG-4HPB, after cross-coupling (38 % conversion):



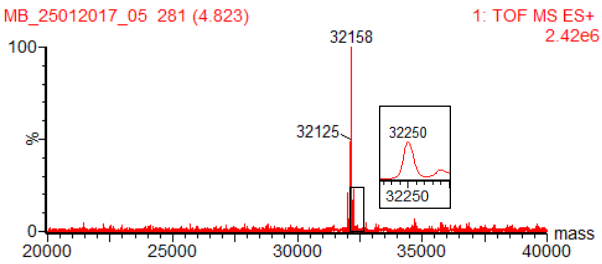
M_{calc}: 32124
M_{obs}: 32125 (38 %)
32032 (deiodination)
32158 (GYG-Y195*p*IPhe)

GYG-4HPB, after glucosylation assay:

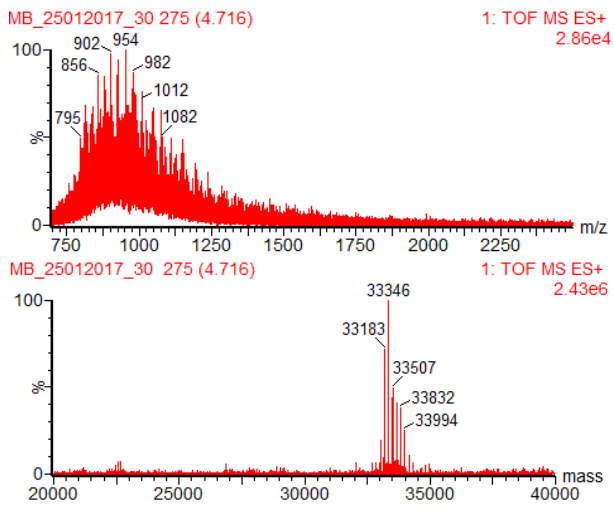


M_{calc} 32124 + 162n
M_{obs}: 32125 (non-glucosylated)
32032 (deiodinated)
32158 (GYG-Y195*p*IPhe)
Also see note below

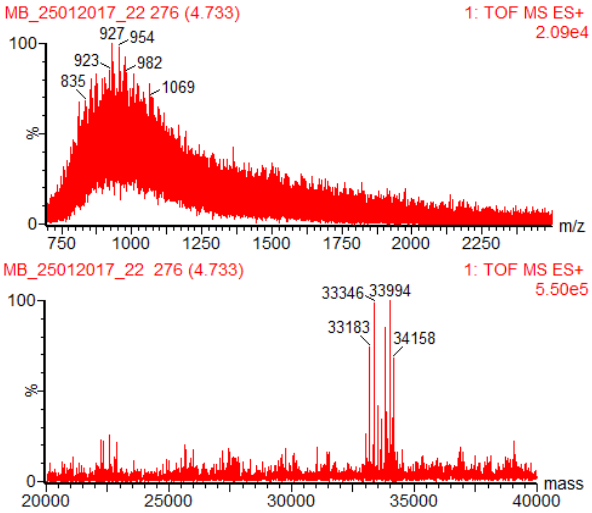
*Note: It may be observed that, after glucosylation assay, there is a small peak at higher Mw than cross-coupled product (and GYG-Y195*pIPhe*). This does not correspond to glucosylation (by mass, expected 32287, 32449 etc.), as shown below, nor does it arise from cross-coupling (it is not present on the left-hand spectrum above). It may well simply represent an artefact/adduct arising from MS analysis itself. It does not affect the conclusion that no glucosylation was observed.*



GYG-wt, before glucosylation assay



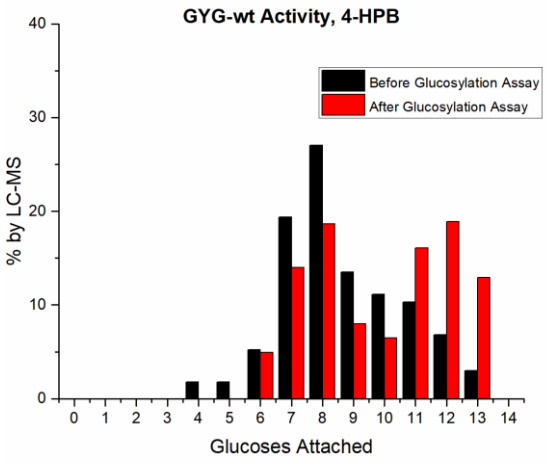
GYG-wt, after glucosylation assay



M_{calc}: 32048 + 162n
M_{obs}: 32050 + 162n
(n = glucoses attached = 4-12,
average = **8.67**)

M_{calc}: 32048 + 162(n+m)
M_{obs}: 32050 + 162(n+m)
(n +m = glucoses attached = 6-13
average = **9.86**)

Graphical Summary of GYG-wt activity from above data:

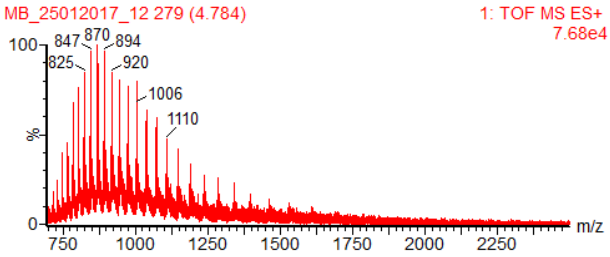


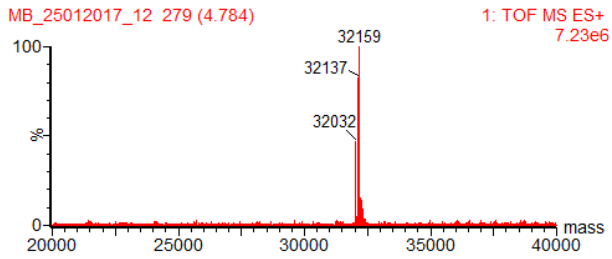
2.8.7.6 GYG-4HMPB (boronic acid = 4-HMPB)

Carried out according to general procedure (section 2.8.7.1) using the following reagent amounts:

- GYG-Y195*p*IPhe: 70 µl (0.6 mg/ml), 1.3 nmol
- Boronic acid **4-HMPB**: 0.3 mg, 1970 nmol, 1500 eq., added as solid owing to poor aqueous solubility. Na_2HPO_4 was thus added separately (18.9 µl of 15.7 mg/ml = 110 mM aq. solution).
- $Pd(L2)_2(OAc)_2$: 7.7 µl of 10 mM stock solution, 77 nmol, 59 eq., 0.8 mM final concentration.

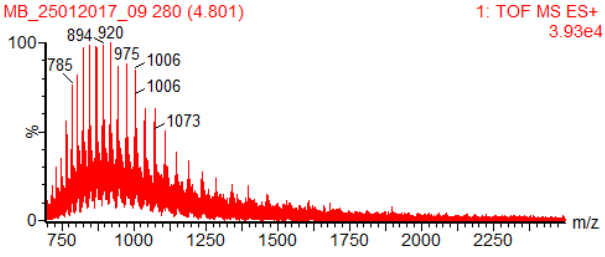
GYG-4HMPB, after cross-coupling (35 % conversion):

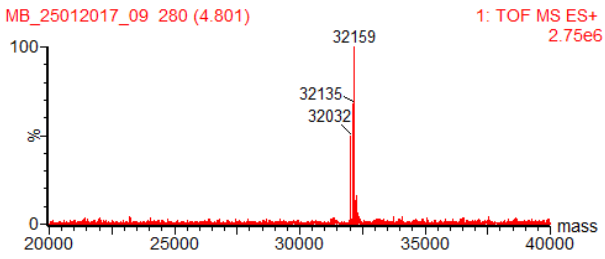




M_{calc}: 32138
M_{obs}: 32137 (35 %)
32032 (deiodination)
32159 (GYG-Y195*p*IPhe)

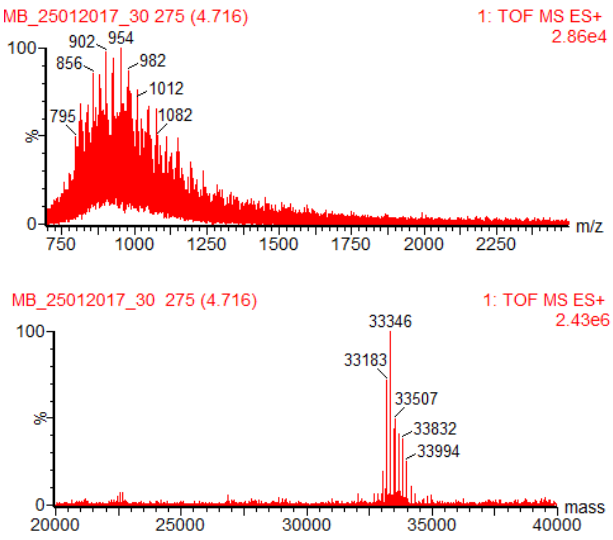
GYG-4HMPB, after glucosylation assay



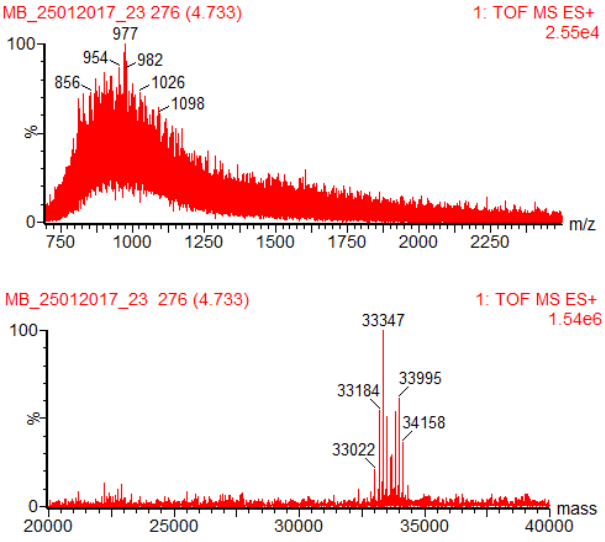


M_{calc} 32138 + 162n
M_{obs}: 32135 (non-glucosylated)
32032 (deiodinated)
32159 (GYG-Y195*p*IPhe)

GYG-wt, before glucosylation assay



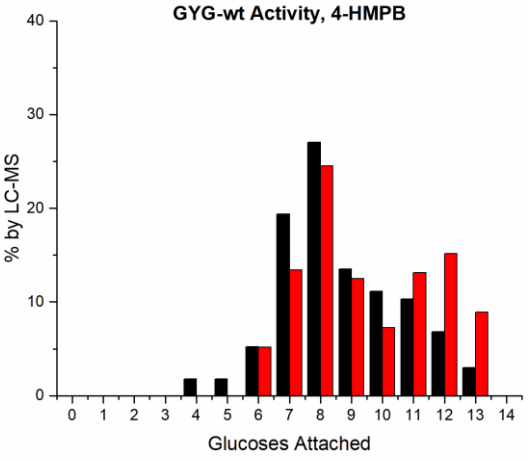
GYG-wt, after glucosylation assay



M_{calc}: 32048 + 162n
M_{obs}: 32050 + 162n
(n = glucoses attached = 4-12,
average = **8.67**)

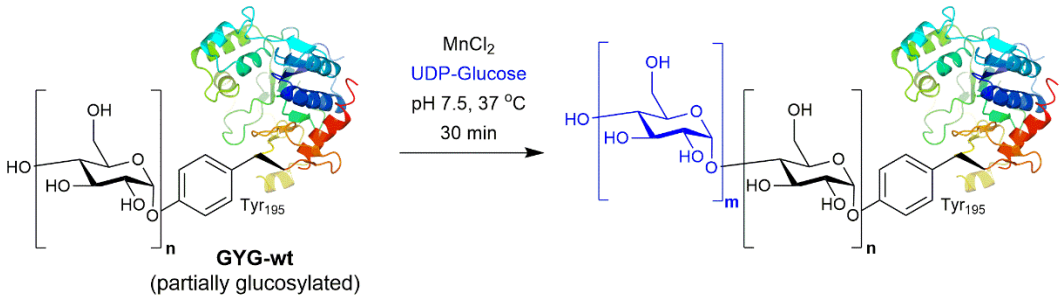
M_{calc}: 32048 + 162(n+m)
M_{obs}: 32049 + 162(n+m)
(n+m = glucoses attached = 6-13
average = **9.48**)

Graphical Summary of GYG-wt activity from above data:



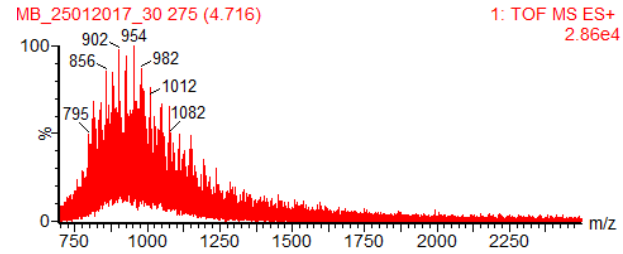
2.8.8 Further Evaluation of 4-HMPB and 4-HPB Systems

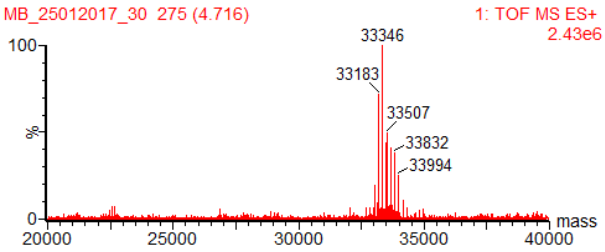
2.8.8.1 Glucosylation Activity of Non-treated GYG-wt, Utilised in 4-HMPB/4-HPB Experiments



GYG-wt (0.2 mg/ml) in pH 7.5 25 mM Na-HEPES was incubated with 1 mM $MnCl_2$ and 1 mM UDP-Glucose disodium salt (added from 10 mM stock solutions) at 37 °C for 30 minutes. After this time, Na-EDTA was added to a final concentration of 20 mM (from a 100 mM stock solution, pH 8.0), the mixture vortexed, and analysed by LC-MS (see section 2.8.6.1).

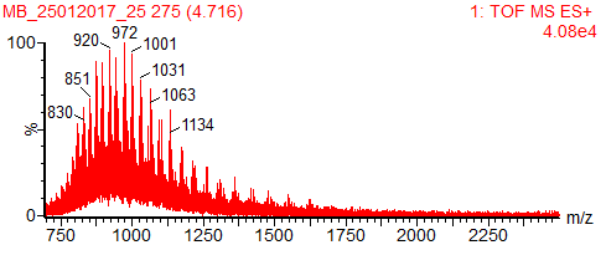
GYG-wt, before glucosylation assay

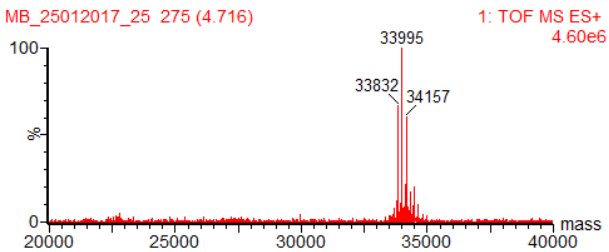




M_{calc}: 32048 + 162n
M_{obs}: 32050 + 162n
(n = glucoses attached = 4-12,
average = **8.67**)

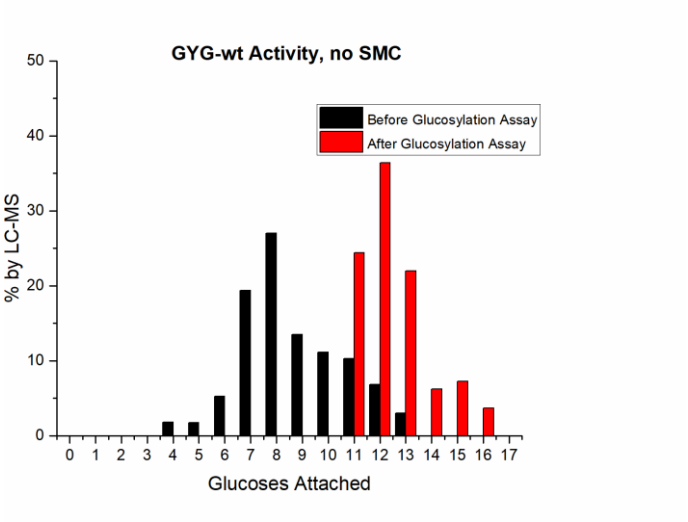
GYG-wt, after glucosylation assay



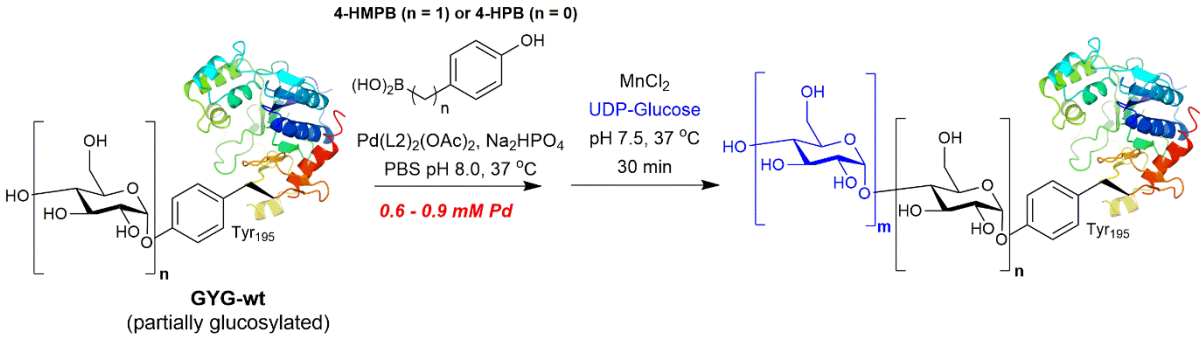


M_{calc}: 32048 + 162(n+m)
M_{obs}: 32049 + 162(n+m)
(n +m = glucoses attached = 11-16
average = **12.46**)

Graphical Summary of Above Data:

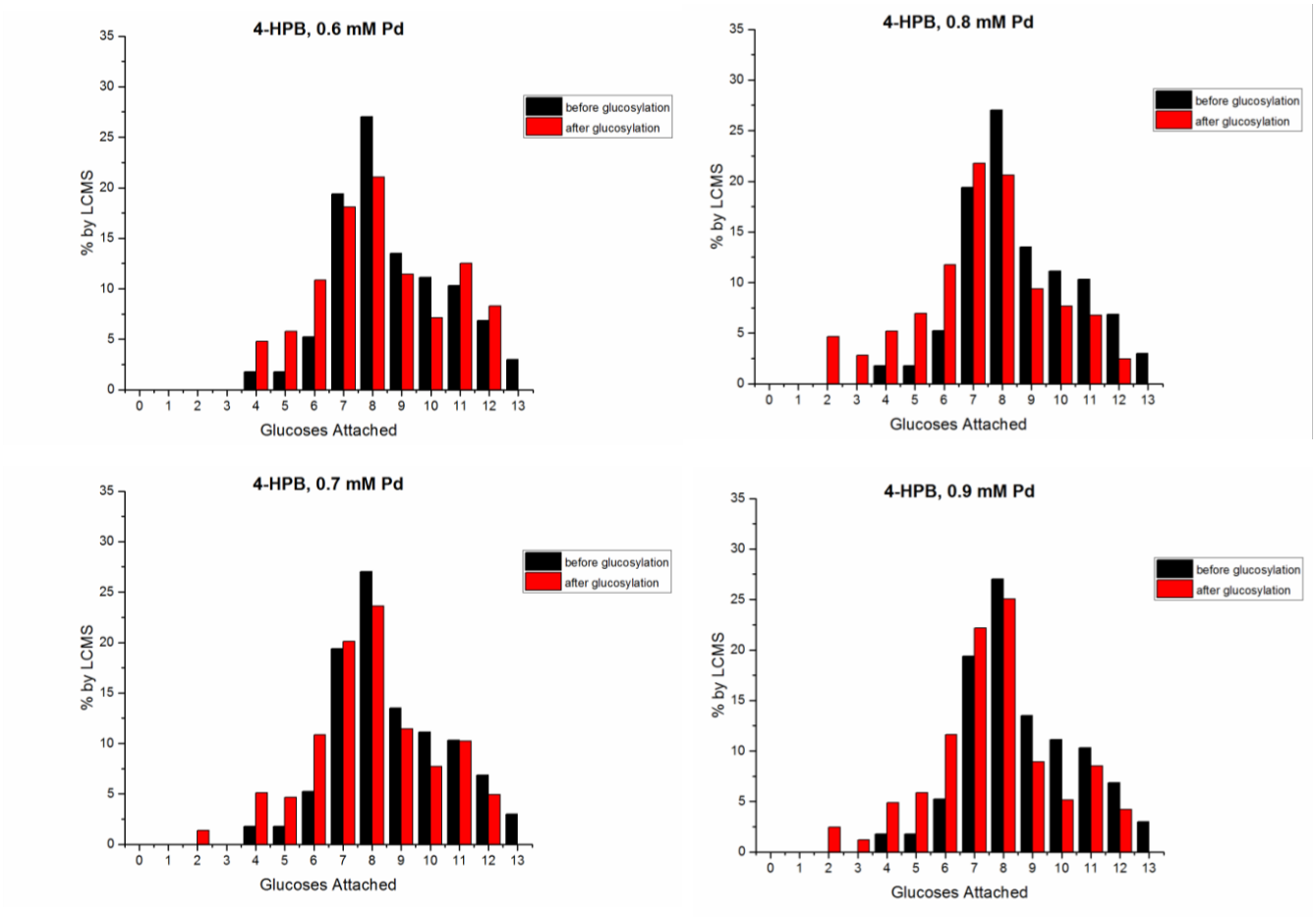


2.8.8.2 Effect of Pd Loading on GYG-wt Activity after 4-HMPB/4-HPB SMC Conditions

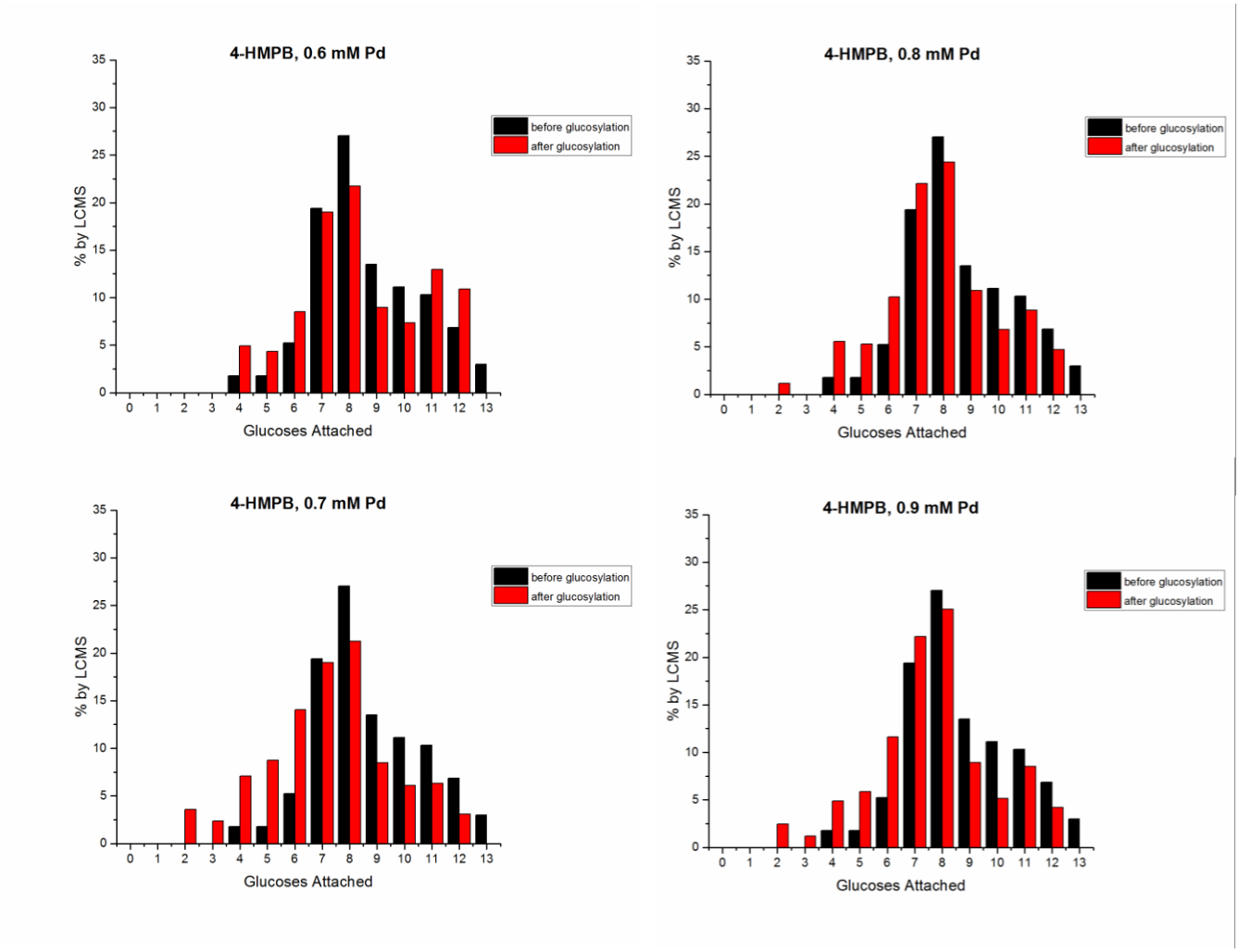


8 reactions were set up. 4 of these were carried out exactly as per section 2.8.7.5, but using GYG-wt only and varying amounts of Pd catalyst (either 5.7, 6.7, 7.7 or 8.8 µl of 10 mM stock solution, 0.6, 0.7, 0.8 or 0.9 mM final concentration). 4 reactions were carried out exactly as per section 2.8.7.6, but with these same variations. LC-MS data obtained were analysed according to section 2.8.6.1 to determine abundances of different glycoforms; results are illustrated graphically below. As can be seen, very little change in glycoform distribution before and after glucosylation assay is apparent, indicating limited if any activity across all Pd concentrations.

4-HPB



4-HMPB


4-HMPB, 0.6 mM Pd
4-HMPB, 0.8 mM Pd
4-HMPB, 0.7 mM Pd
4-HMPB, 0.9 mM Pd
before glucosylation
after glucosylation
% by LCMS
Glucoses Attached

2.9 References

1. Korostelev, A., Zhu, J., Asahara, H. & Noller, H. F. Recognition of the amber UAG stop codon by release factor RF1. *EMBO J.* **29,** 2577–2585 (2010).

2. Wals, K. & Ovaa, H. Unnatural amino acid incorporation in E. coli: current and future applications in the design of therapeutic proteins. *Front. Chem.* **2,** 1–12 (2014).

3. Alff-Steinberger, C. & Epstein, R. Codon Preference in the Terminal Region of E. coli Genes and Evolution of Stop Codon Usage. *Journal of Theoretical Biology* **168,** 461–463 (1994).

4. Wu, B., Wang, Z., Huang, Y. & Liu, W. R. Catalyst-Free and Site-Specific One-Pot Dual-Labeling of a Protein Directed by Two Genetically Incorporated Noncanonical Amino Acids. *ChemBioChem* **13,** 1405–1408 (2012).

5. O'Donoghue, P. *et al.* Near-cognate suppression of amber, opal and quadruplet codons competes with aminoacyl-tRNAPyl for genetic code expansion. *FEBS Lett.* **586,** 3931–7 (2012).

6. Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C., Schultz, P. G. & April, I. A General Method for Site-Specific Incorporation of Unnatural Amino Acids into Proteins. *Science.* **244,** 182–188 (1989).

7. Nour, E. F., Dedkova, L., Wang, B., Golovine, S. & Hecht, S. M. Site-Specific Incorporation of Glycosylated Serine and Tyrosine Derivatives into Proteins. *JACS* **129,** 3586-3597 (2007).

8. Matsubara, T., Iijima, K., Watanabe, T., Hohsaka, T. & Sato, T. Incorporation of glycosylated amino acid into protein by an in vitro translation system. *Bioorganic & Medicinal Chemistry Letters* **23,** 5634-5636 (2013).

9. Liu, C. C. & Schultz, P. G. Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* **79,** 413–444 (2010).

10. Alberts, B. *et al. Molecular Biology of the Cell, Sixth Edition:* (Taylor & Francis Group, 2014).

11. Davis, L. & Chin, J. W. Designer proteins: applications of genetic code expansion in cell biology. *Nat. Rev. Mol. Cell Biol.* **13,** 168–182 (2012).

12. Dumas, A., Lercher, L., Spicer, C. D. & Davis, B. G. Designing logical codon reassignment – Expanding the chemistry in biology. *Chem. Sci.* **6,** 50–69 (2015).

13. Wang, L., Brock, a, Herberich, B. & Schultz, P. G. Expanding the genetic code of Escherichia coli. *Science* **292,** 498–500 (2001).

14. Wang, Y.-S., Fang, X., Wallace, A. L., Wu, B. & Liu, W. R. A rationally designed pyrrolysyl-tRNA synthetase mutant with a broad substrate spectrum. *J. Am. Chem. Soc.* **134,** 2950–2953 (2012).

15. Wang, W. *et al.* Genetically encoding unnatural amino acids for cellular and neuronal studies. *Nat. Neurosci.* **10,** 1063–1072 (2007).

16. Liu, W., Brock, A., Chen, S., Chen, S. & Schultz, P. G. Genetic incorporation of unnatural amino acids into proteins in mammalian cells. *Nat. Methods* **4,** 239–244 (2007).

17. Ernst, R. J. *et al.* Genetic code expansion in the mouse brain. *Nat. Chem. Biol.* **12,** 1–5 (2016).

18. Wang, L., Magliery, T. J., Liu, D. R. & Schultz, P. G. A new functional suppressor tRNA/aminoacyl-tRNA synthetase pair for the in vivo incorporation of unnatural amino acids into proteins [16]. *J. Am. Chem. Soc.* **122,** 5010–5011 (2000).

19. Zeng, H., Xie, J. & Schultz, P. G. Genetic introduction of a diketone-containing amino acid into proteins. *Bioorganic Med. Chem. Lett.* **16,** 5356–5359 (2006).

20. Tsao, M.-L., Tian, F. & Schultz, P. G. Selective Staudinger Modification of Proteins Containing p-Azidophenylalanine. *ChemBioChem* **6,** 2147–2149 (2005).

21. Brustad, E. *et al.* A genetically encoded boronate-containing amino acid. *Angew. Chemie. Int. Ed.* **47,** 8220–8223 (2008).

22. Santoro, S. W., Wang, L., Herberich, B., King, D. S. & Schultz, P. G. An efficient system for the evolution of aminoacyl-tRNA synthetase specificity. *Nat. Biotechnol.* **20,** 1044–1048 (2002).

23. Xie, J. *et al.* The site-specific incorporation of p-iodo-L-phenylalanine into proteins for structure determination. *Nat. Biotechnol.* **22,** 1297–1301 (2004).

24. Young, T. S., Ahmad, I., Yin, J. A. & Schultz, P. G. An Enhanced System for Unnatural Amino Acid Mutagenesis in E. coli. *J. Mol. Biol.* **395,** 361–374 (2010).

25. Spicer, C. D. & Davis, B. G. Palladium-mediated site-selective Suzuki-Miyaura protein modification at genetically encoded aryl halides. *Chem. Commun. (Camb).* **47,** 1698–1700 (2011).

26. Chalker, J. M., Wood, C. S. C. & Davis, B. G. A convenient catalyst for aqueous and protein Suzuki-Miyaura cross-coupling. *J. Am. Chem. Soc.* **131,** 16346–16347 (2009).

27. Antos, J. M. & Francis, M. B. Transition metal catalyzed methods for site-selective protein modification. *Curr. Opin. Chem. Biol.* **10,** 253–262 (2006).

28. Spicer, C. D., Triemer, T. & Davis, B. G. Palladium-mediated cell-surface labeling. *J. Am. Chem. Soc.* **134,** 800–803 (2012).

29. Spicer, C. D. & Davis, B. G. Rewriting the bacterial glycocalyx via Suzuki-Miyaura cross-coupling. *Chem. Commun. (Camb).* **49,** 2747–9 (2013).

30. Spicer, C. D. Suzuki Biology: Palladium Mediated Reactions under Ambient Conditions. (DPhil thesis, University of Oxford, 2014).

31. Gao, Z., Gouverneur, V. & Davis, B. G. Enhanced aqueous Suzuki-Miyaura coupling allows site-specific polypeptide 18F-labeling. *J. Am. Chem. Soc.* **135,** 13612–13615 (2013).

32. Dumas, A. *et al.* Self-liganded Suzuki-Miyaura coupling for site-selective protein PEGylation. *Angew. Chem. Int. Ed. Engl.* **52,** 3916–3921 (2013).

33. Alonso, M. D., Lomako, J., Lomako, W. M. & Whelan, W. J. Tyrosine-194 of glycogenin undergoes autocatalytic glucosylation but is not essential for catalytic function and activity. *FEBS Lett.* **342,** 38–42 (1994).

34. Chaikuad, A. *et al.* Conformational plasticity of glycogenin and its maltosaccharide substrate during glycogen biogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 21028–33 (2011).

35. Hemberg, G. Investigating the Autocatalytic Priming Mechanism in Glycogenin. (MSci thesis, University of Oxford, 2013).

36. Hurley, T. D., Stout, S., Miner, E., Zhou, J. & Roach, P. J. Requirements for catalysis in mammalian glycogenin. *J. Biol. Chem.* **280,** 23892–23899 (2005).

37. Issoglio, F. M., Carrizo, M. E., Romero, J. M. & Curtino, J. A. Mechanisms of monomeric and dimeric glycogenin autoglucosylation. *J. Biol. Chem.* **287,** 1955–1961 (2012).

38. Bazán, S., Issoglio, F. M., Carrizo, M. E. & Curtino, J. A. The intramolecular autoglucosylation of monomeric glycogenin. *Biochem. Biophys. Res. Commun.* **371,** 328–332 (2008).

39. Geoghegan, K. F. *et al.* Spontaneous alpha-N-6-phosphogluconoylation of a 'His tag' in Escherichia coli: the cause of extra mass of 258 or 178 Da in fusion proteins. *Anal. Biochem.* **267,** 169–184 (1999).

40. Ahmadi, Z. & McIndoe, J. S. A mechanistic investigation of hydrodehalogenation using ESI-MS. *Chem. Commun.* **49,** 11488-11490 (2013).

41. Lennox, A. J. J. & Lloyd-Jones, G. C. Transmetalation in the Suzuki-Miyaura coupling: the fork in the trail. *Angew. Chem. Int. Ed. Engl.* **52,** 7362–7370 (2013).

42. Schönfelder, D., Nuyken, O. & Weberskirch, R. Heck and Suzuki coupling reactions in water using poly(2-oxazoline)s functionalized with palladium carbene complexes as soluble, amphiphilic polymer supports. *J. Organomet. Chem.* **690,** 4648–4655 (2005).

43. Leadbeater, N. E. Fast, easy, clean chemistry by using water as a solvent and microwave heating: the Suzuki coupling as an illustration. *Chem. Commun.* 2881-2902 (2005).

44. Li, N., Lim, R. K. V, Edwardraja, S. & Lin, Q. Copper-free Sonogashira cross-coupling for functionalization of alkyne-encoded proteins in aqueous medium and in bacterial cells. *J. Am. Chem. Soc.* **133,** 15316–15319 (2011).

45. Moore, L. M. & Shaughnessy, K. H. Efficient Aqueous-Phase Heck and Suzuki Couplings of Aryl Bromides Using Tri(4,6-dimethyl-3- sulfonatophenyl)phosphine Trisodium Salt (TXPTS). *Org. Lett.* **6,** 225-228 (2004).

46. Tedeschi, D. H. Compositions and methods for producing tranquilizing activity. U.S. Patent 3303095 (1967).

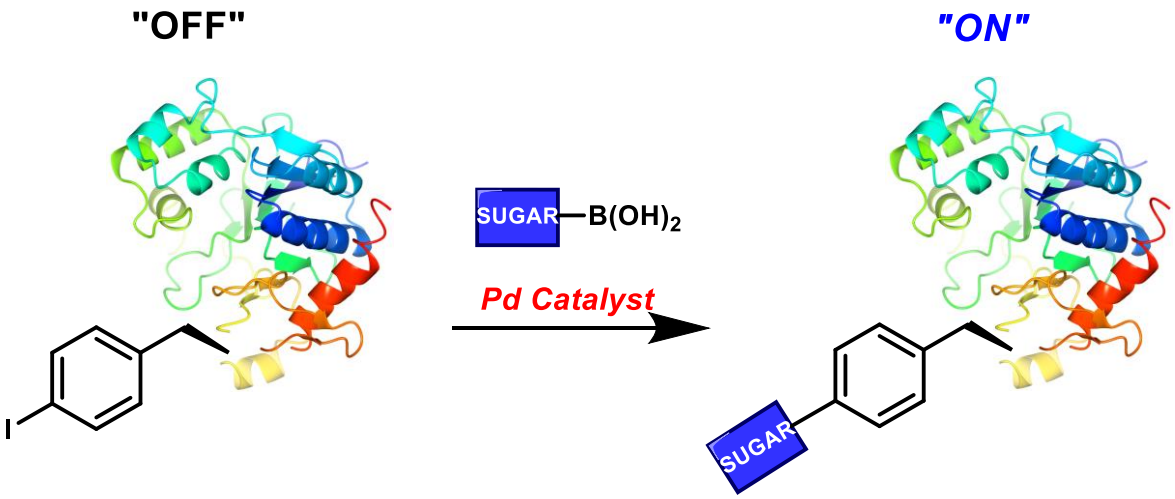
47. Mochita Pharmaceutical Company, Ltd. (Tokyo, Japan). Pyrimidine derivatives. U.S. Patent 5397781 (1995).

48. Willwacher, J., Raj, R., Mohammed, S. & Davis, B. G. Selective Metal-Site-Guided Arylation of Proteins. *J. Am. Chem. Soc.* **138,** 8678–8681 (2016).

49. Smith, H. E. & Neergaard, J. R. Optically active amines. 41. Application of the benzene chirality rule to chiral ring-substituted benzylcarbinamines and benzylcarbinamine salts. *J. Am. Chem. Soc.* **119,** 116–124 (1997).

50. Hurley, T. D., Walls, C., Bennett, J. R., Roach, P. J. & Wang, M. Direct detection of glycogenin reaction products during glycogen initiation. *Biochem Biophys Res Commun.* **348,** 374–378 (2006).

Chapter 3



Palladium-mediated

Activation of Glycogenin-1

Contents

3.1 Introduction 168

3.1.1 Design of More Accurate GYG Glycoform Mimetics 168

3.1.2 Selective Activation of Enzymes: Towards a Novel Methodology 172

3.2 Synthesis of a Glucose-Derived Boronic Acid 174

3.3 Pd-Mediated Enzyme Activation 176

3.3.1 Chemical Generation of a Competent GYG Glycoform 176

3.3.2 Optimisation of Pd-mediated Enzyme Activation 185

3.4 Conclusion and Outlook 189

3.5 Experimental Section 190

3.5.1 Chemical Synthesis 190

3.5.2 TEV Cleavage of his-tagged GYG-Y195*p*IPhe 195

3.5.3 Generation and Enzymatic Activity of his-tagged GYG-Glucose 196

3.5.3.1 Initial cross-couplings: optimisation of Boronic Acid Loading 196

3.5.3.2 Enzymatic Activity of his-tagged GYG-Glucose: Initial Attempt (0.8 mM Pd) 199

3.5.3.3 Optimisation Studies on GYG-wt: Pd/DTT Loading 202

3.5.3.4 Cross-Coupling of GYG-Y195*p*IPhe to Boronic Acid **3.1**: Minimising Pd Concentration 205

3.5.3.5 Enzymatic Activity of his-tagged GYG-Glucose: Optimised (0.6 mM Pd) 207

3.5.4 Generation and Enzymatic Activity of non-his-tagged GYG-Glucose 209

3.5.4.1 Initial Cross-Coupling of his-tag-cleaved GYG-Y195*p*IPhe to Glucose Boronic Acid **3.1**, with Enzymatic Activity Test 209

3.5.4.2 Optimisations of Cross-Coupling to non-his-tagged GYG-Y195*p*IPhe 211

3.5.4.3 Enzymatic Activity of GYG-Glucose: Fully Optimised Procedure 221

3.5.5 Further Investigations of Cross-coupling 223

3.5.5.1 Negative Control Studies: GYG-Y195F 223

3.5.5.2 Effect of LC-MS Mobile Phase on Cross-coupling side products 225

3.6 References 228

3.1 Introduction

3.1.1 Design of More Accurate GYG Glycoform Mimetics

The work described in the preceding chapter demonstrated the evident potential of Suzuki-Miyaura cross-coupling (SMC) as a method for the chemical modification of GYG under benign, enzyme-compatible conditions. However, it was equally apparent that that the enzyme did not exhibit sufficient plasticity to recognise and thus effect catalysis from the simple biaryl-type glycosyl acceptors generated thus far, presumably on account of their limited structural resemblance to the native, glucose-derived GYG acceptor species. As such, we turned our focus towards the generation of more structurally accurate mimetics of GYG glycoforms, which may be more likely to display catalytic activity and thus represent biologically relevant, functional species.

Neither SMC nor any other current protein chemical modification method is capable of affording an exact replica of naturally glucosylated GYG; this is an inherent limitation of the chemistries, and thus the retrosynthetic disconnections, currently available in the context of protein chemistry. This issue extends well beyond this particular example, and indeed beyond the field of glycoprotein synthesis; the synthetic installation of any post-translational modification through such a chemical modification approach can generally afford at best a close mimic of the natural variant, but not an exact one. (An exception in this regard is a recently reported radical-based method for installation of a wide variety of PTMs,[1] yet even this approach yields a mixture of the native PTM and its non-natural epimer).

However, a number of examples have demonstrated that exact structural mimicry is not a pre-requisite for a wide range of synthetic PTMs to be successfully recognised as natural (figure 3.1). Alkylation of a cysteine residue, installed by site-directed mutagenesis, on Histone H3 with methylthiocarbonyl-aziridine afforded an acetyl-Lys mimic that is recognised by acetyltransferases and site-specific antiacetyl-Lys antibodies (**figure 3.1 a).**[2] A similar alkylation approach can also yield mimics of mono-, di- or tri-methylated lysine that functionally resemble their natural counterparts (**figure 3.1 b**).[3] Recent work in the Davis group has demonstrated that the chemical installation of a thiophosphate moiety via thia-Michael addition to dehydroalanine (Dha) mimics the functional properties of threonine phosphorylation of the kinase p38α (**figure 3.1 c**),[4] whilst Brik and co-workers have utilised chemical modification in order to site-specifically ubiquitinate protein substrates (**figure 3.1 d**);[5] in spite of a non-natural linker between protein and ubiquitin, the protein is nonetheless targeted for degradation, consistent with the consequences of the corresponding natural modification.[6]

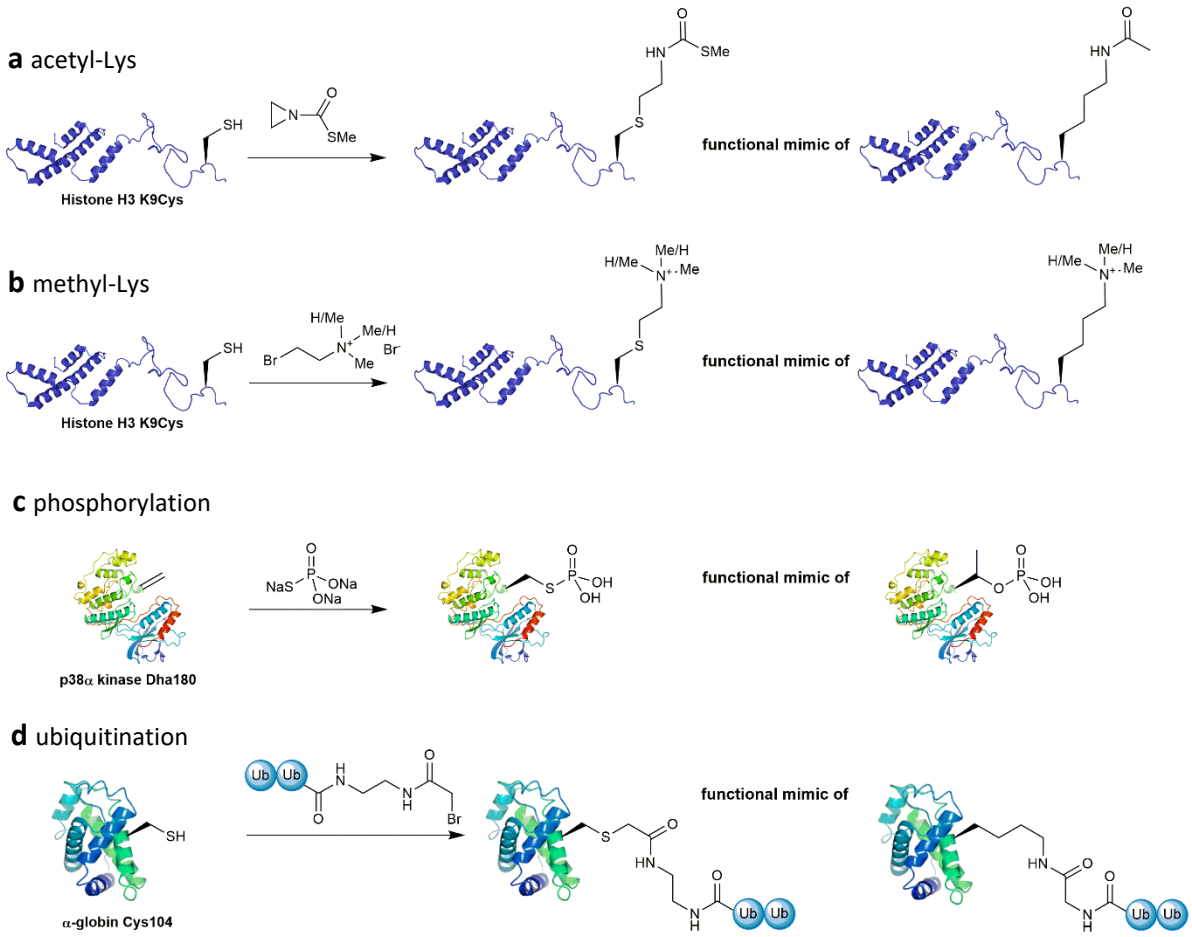


Figure 3.1 Chemical modification of proteins can afford close, but non-exact mimics of natural post-translational modifications; examples include acetyl-lysine (a), methyl-lysine (b), phosphorylation (c), and ubiquitination (d). In spite of non-exact structural correspondence, the chemically installed modifications retained the function of the native variants.

Returning to the more immediately relevant case of glycosylation, similar observations can be made; whilst some studies have resorted to total synthesis in order to generate exact structural mimics of glycoproteins,[7,8] other examples demonstrate the application of chemical modification to install glycosylation PTMs that, in spite of non-exact structural mimicry, nonetheless reproduce the native biological function and can be recognised by relevant protein partners (**figure 3.2**).[9–12] For example, synthetic installation of an *O*-GlcNAc PTM on the nucleosome enabled probing of its native biological function, in spite of a non-natural thioether linkage between carbohydrate and protein (**figure 3.2 a**).[10,11] Notably, whilst not representing an example of chemical protein modification, the group of Bertozzi have demonstrated that mucins, highly complex glycoproteins serving numerous functions on cell surfaces, can be successfully mimicked by glycopolymer mimetics;[13,14] such species do not even contain a single peptide bond, yet are nonetheless recognised as

native and have been successfully applied in probing possible reasons for the frequent elevation of mucins in tumour cells (**figure 3.2 b**).[15,16] Of particular relevance to this work is the report of Spicer *et al* demonstrating the on-cell modification of the bacterial cell-surface protein OmpC with carbohydrate moieties via SMC; the proteins thus generated were sufficient mimics of native variants to be recognised by native binding partners (**figure 3.2 c**).[12] It should be noted, however, that, owing to the *in vivo* nature of the study, no quantification of the cross-coupling was obtained.

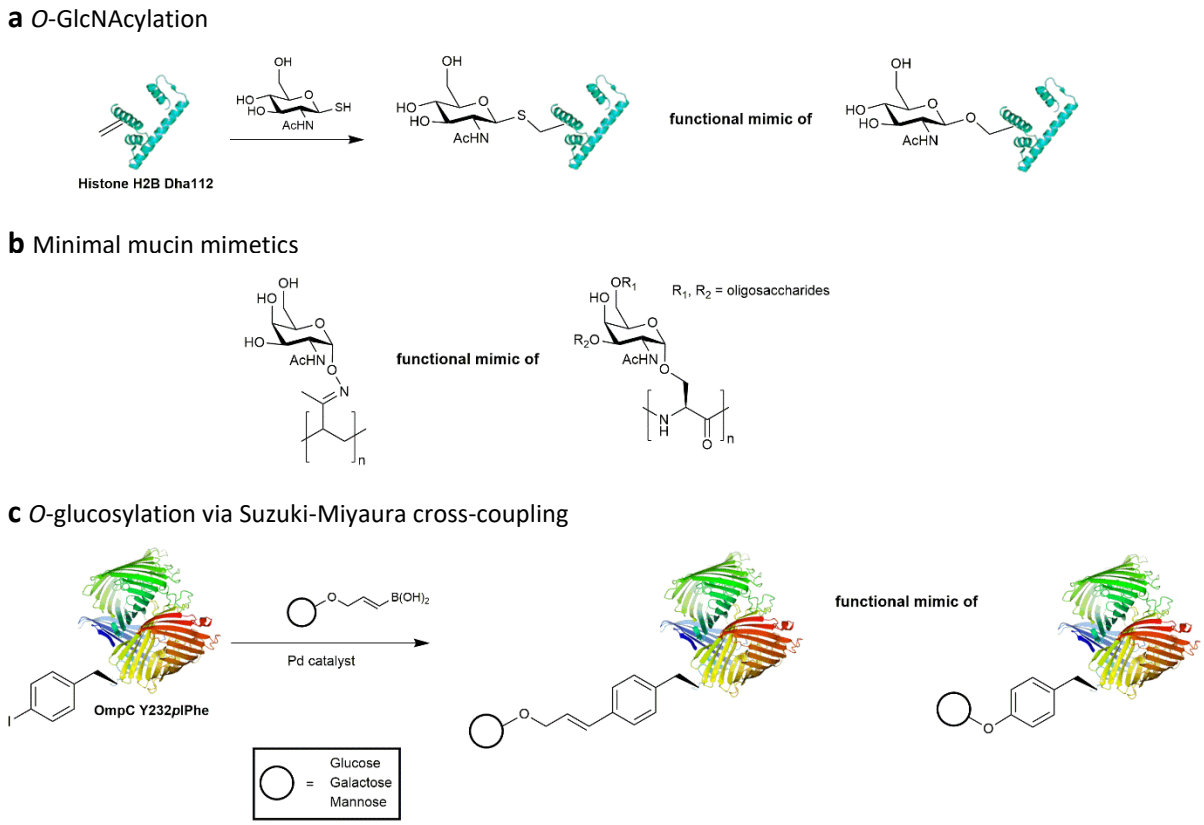


Figure 3.2 Chemical approaches can generate mimics of glycoprotein species that, in spite of non-exact structural correspondence to native variants, retain the biological function of the latter. Examples include: the installation of *O*-GlcNAc via chemical modification (a), the chemical synthesis of minimal mimetics of mucins, retaining native biological function in spite of the absence of a single peptide bond (b), and – of particular relevance to this work – the use of Suzuki-Miyaura cross-coupling (SMC) to access glycoprotein mimetics that could be recognised by native binding partners.

Indeed, it was this latter study that would inspire much of the work described herein. Furthermore, a cursory overview of the enzymatic mechanism of GYG (see chapter 1) reveals that it must recognise at least two distinct structural motifs as glycosyl acceptors; the phenolic acceptor of Tyr[195], and the 4'-OH of a glucose moiety.[17] Given that both steps are known to proceed without requirement for any other enzyme,[18] this implies the necessity of at least some inherent flexibility of the enzyme with respect to glycosyl acceptor, further implicating a potential tolerance of chemically installed variants thereof.

It was evident to us that a cross-coupling of the glucose-derived boronic acid **3.1**, also employed by Spicer *et al* (**figure 3.2 c**),[12] to GYG-Y195*p*IPhe could yield a species "GYG-Glucose" bearing a striking structural resemblance to native, mono-glucosylated GYG (**figure 3.3 a**); the latter represents the first catalytic intermediate of GYG autoglucosylation. Indeed, comparing such a putative cross-coupled product ("chemical") with its corresponding native variant ("biological") reveals that both cases fundamentally involve a linkage of the same protein scaffold to a glucose moiety at the 195 position; the sole difference is the presence of a minimal non-natural "linker" moiety in the former, which we envisaged may be sufficiently small so as to not significantly perturb enzyme function (**figure 3.3.b**).

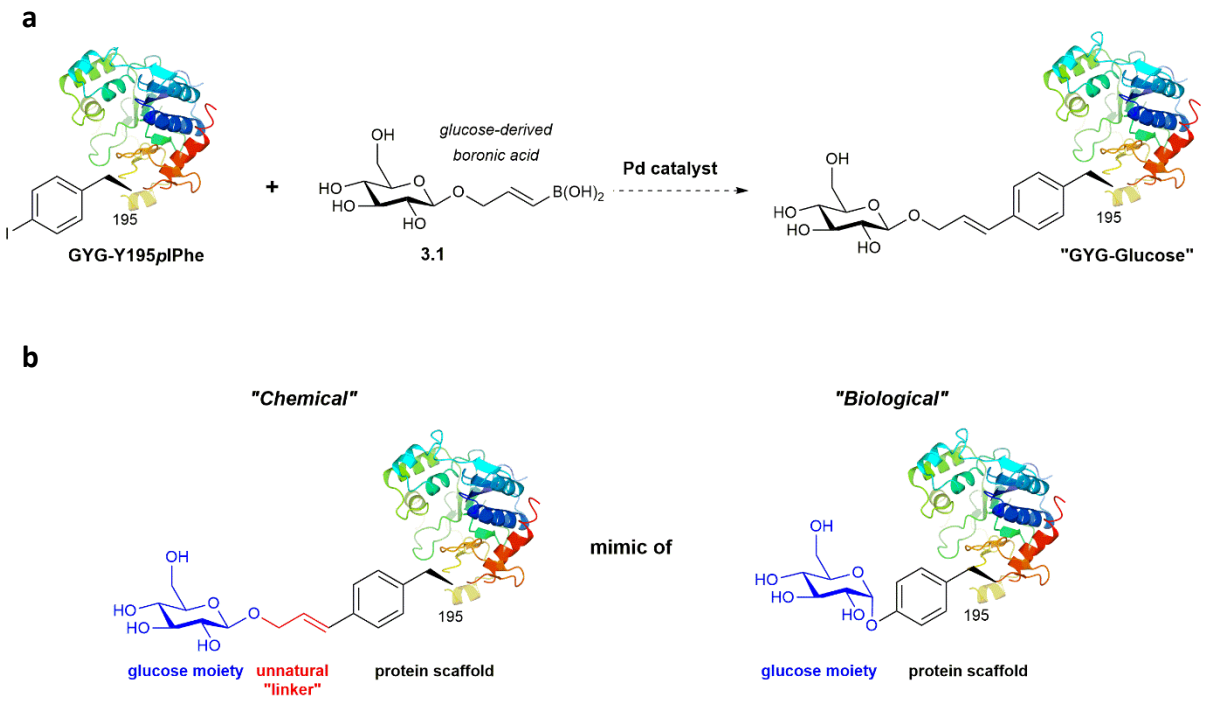


Figure 3.3 (a) cross-coupling between GYG-Y195*p*IPhe and glucose boronic acid **3.1** could yield a species "GYG-Glucose". Such a chemically generated enzyme exhibits striking structural mimicry to the native "biological" variant, with the sole difference being the presence of a short, unnatural "linker" in the former (b).

A brief discussion concerning the nature of this linker between carbohydrate and boronic acid moieties is warranted. Clearly, in order to best mimic the natural enzyme, the linker should be of minimal length; however, equally, the carbohydrate boronic acid must remain synthetically tractable. Additionally, considering that Suzuki-Miyaura cross-couplings involving sp^3 alkylboronic acids, whilst feasible in organic chemistry,[19–21] represent an as yet unsolved challenge in protein chemistry; the presence of an alkene functionality proximal to the boronic acid is beneficial. The only analogue of **3.1** having a shorter linker yet still conforming to this latter requirement is the vinyl glycoside **3.2**; cross-coupling of this species to GYG-Y195*p*IPhe may in principle yield a slightly improved mimic of the natural variant, yet such a species represents a somewhat more challenging synthetic target than **3.1** (**figure 3.4**). Conversely, as mentioned, variants of the form **3.3** bearing longer linkers yield poorer mimics of the natural enzyme upon cross-coupling. Thus, **3.1** represented a suitable compromise of the relevant factors.

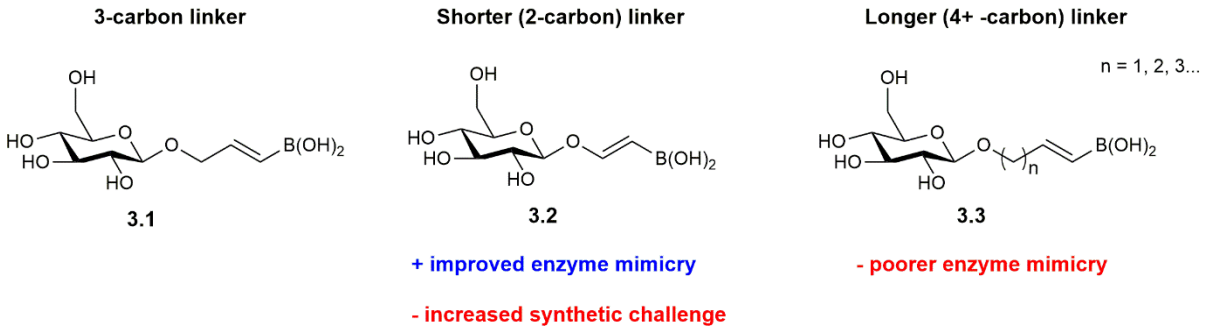


Figure 3.4 Potential candidate boronic acids for cross-coupling to GYG-Y195*p*IPhe in order to generate mimics of mono-glucosylated GYG, with varying linker lengths. The 3-carbon linker variant represents an optimal compromise of synthetic tractability and efficient mimicry of the native mono-glucosylated enzyme.

We also note that the linker does contain a β-glycosidic linkage, rather than the natural α-linkage found in GYG. This is based on two factors. Firstly, the β-glycosides represent substantially more tractable synthetic targets. Secondly, this linkage is a small part of an already non-natural – yet flexible – linker. These two factors combined led us to pursue this route.

3.1.2 Selective Activation of Enzymes: Towards a Novel Methodology

Whilst the proposed generation of homogeneous, catalytically active glycoforms from the cross-coupling of boronic acids such as **3.1** to GYG-Y195*p*IPhe primarily represented to us a means to attain superior mechanistic understanding of the enzyme, the significant inherent interest of the methodology itself did not escape our attention. The GYG-Y195*p*IPhe enzyme is catalytically inactive towards autoglucosylation, owing to the lack of any glycosyl acceptor at the natural modification site (195). Thus, any SMC to this site

that affords an active enzyme construct would represent an intriguing example of Pd-mediated chemical enzyme activation (**figure 3.5**).

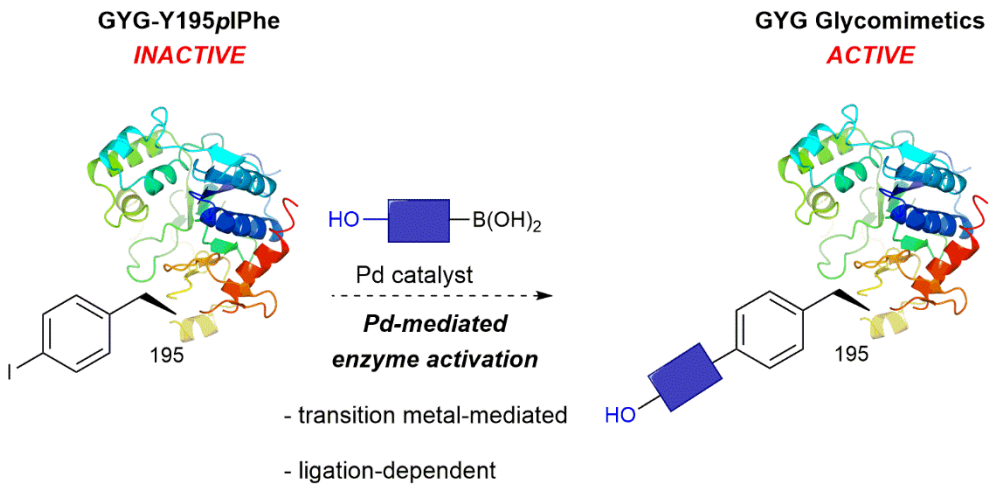


Figure 3.5 Our methodology represents a potential example of Pd-mediated chemical enzyme activation, given the necessity for activity of the GYG glycomimetic products and, in contrast, the inactivity towards autoglucosylation of the tagged precursor, GYG-Y195*p*IPhe. Its combination of transition metal-dependence and ligation-dependent reaction mode represents a highly novel approach.

Such controllable protein activation is not in itself an unknown concept. This may be mediated through the action of a particular small molecule,[22,23] or alternatively, genetic incorporation of photocaged amino acid analogues can enable the selective activation of enzymes through subsequent light-mediated "unmasking" of key catalytic residues.[24–28] The use of transition metals in this context represents a less well-established field, although their use in "activation" of small molecules is known (e.g. to induce fluorescence, including in a cellular context.[29,30] Indeed, Li *et al* recently demonstrated the use of palladium chemistry in the selective activation of a protein substrate,[31] whereby genetic replacement of a key lysine residue with a propargylated analogue thereof rendered the protein inactive, but selective Pd-catalysed deprotection of this species regenerates the lysine and activates the protein.

Yet, in common with the vast majority of current methods for enzyme activation, this too relies on deprotection chemistry of a suitably caged amino acid residue. It is difficult to find examples of the use of ligation-based chemistry (e.g. cross-coupling) to effect such a transformation from inactive to active enzyme.[32] Our proposed methodology, if successful, would represent an unprecedented mode of enzyme

activation that is both ligation-dependent and transition metal-mediated, and as such would represent the first application of SMC in this regard. The remainder of this chapter outlines the journey towards the successful realisation of this approach, leading to the first generation of a homogeneously glucosylated, catalytically active variant of GYG.

3.2 Synthesis of a Glucose-Derived Boronic Acid

Spicer *et al* have previously reported two general convergent synthetic routes towards monosaccharide boronic acids, including glucose boronic acid **3.1**, starting from the desired pentaacetate (**figure 3.6**).[12] These both utilise the same synthetic steps – glycosylation, hydroboration, boronic ester hydrolysis and global acetate deprotection – and merely differ with respect to the order of the first two steps. In the first route, glycosylation of the requisite pentaacetyl sugar with propargyl alcohol affords a substrate for subsequent hydroboration, yielding an intermediate pinacol ester. This may alternatively be accessed via hydroboration of protected propargyl alcohol **3.4**, followed by glycosylation of the deprotected product **3.5** to the peracetate. The routes converge hereafter; deprotection of first the boron pinacol ester, and subsequently of the acetate groups, affords carbohydrate-derived boronic acids.

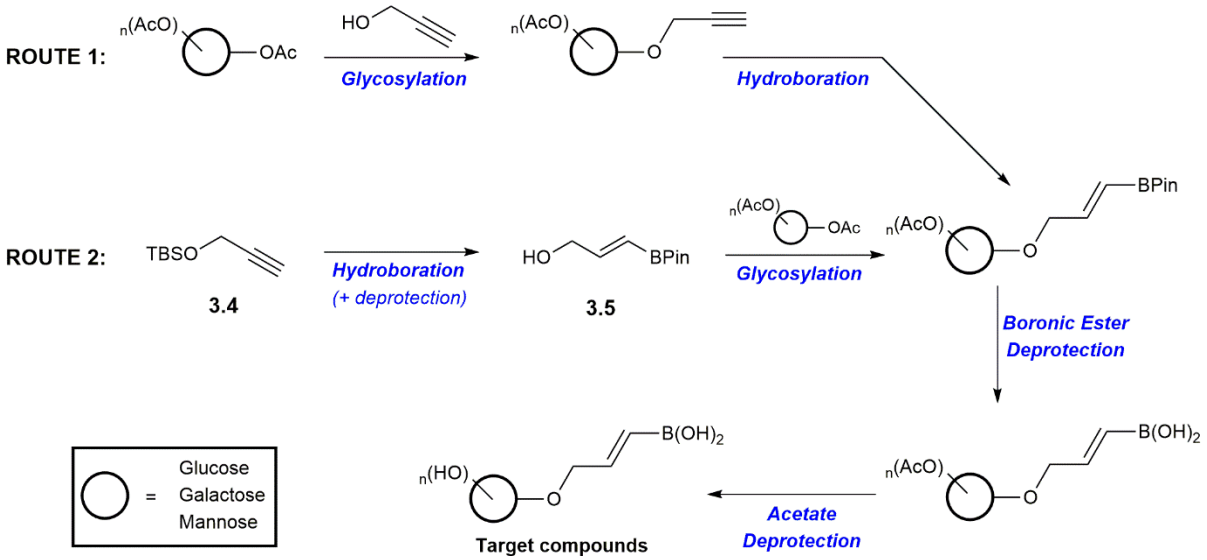


Figure 3.6 General methods for the synthesis of monosaccharide-derived boronic acids, as previously reported by Spicer *et al*.[12]

For the specific case of glucose boronic acid **3.1**, the former of these two routes was previously found to proceed in slightly higher overall yield;[12] this route was therefore employed in this work. Thus, glycosylation of commercially available β-D-glucose pentaacetate **3.6** with propargyl alcohol proceeded to afford *O*-glycoside **3.7**, which could be recrystallized to yield solely the β-anomer in good yield (**figure 3.7**). Zr-mediated hydroboration[33] proceeded in moderate yield, consistent with that previously reported and reflecting challenges in the purification of this compound.[12] However, in our hands, this proceeded with concomitant deacetylation, as confirmed by MS and NMR analysis; NMR indicated the 6'-OAc as the likely site of de-acetylation to give **3.8**. As such, prior to further characterisation, this species was subjected to a re-acetylation step, yielding fully protected pinacol ester **3.9**.

Deprotection of the boronic ester was effected through treatment with diethanolamine; this leads to selective precipitation of the intermediate diethanolamine ester, which is readily hydrolysed upon exposure to a biphasic solution of HCl and Et_2O, yielding peracetylated boronic acid **3.10**. Subsequent global deacetylation under Zemplén conditions afforded the desired glucose boronic acid **3.1**. It should be noted that, in comparison to the protocol utilised by Spicer *et al*,[12] we found an additional brief purification step of **3.1** to be of critical importance for the efficiency of the subsequent cross-coupling. Furthermore, whilst one might initially question the need for re-acetylation of the mono-deacetylated intermediate **3.8**, we found this to be necessary for the success of the subsequent diethanolamine-mediated hydrolysis step; attempts to effect the latter directly on **3.8** routinely failed to proceed. This might reflect coordination of the free 6'-OH group of **3.8** to the boronic acid moiety, thus precluding reaction of the latter with the diethanolamine reagent and formation of the diethanolamine ester.

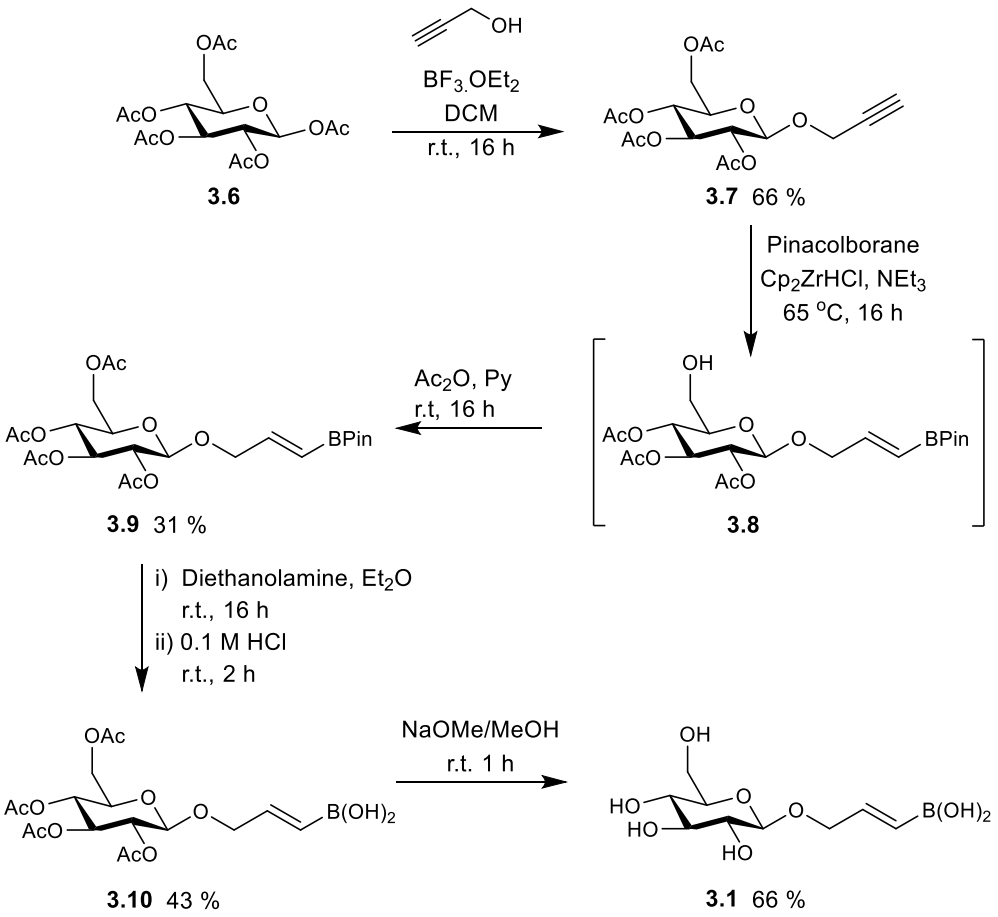


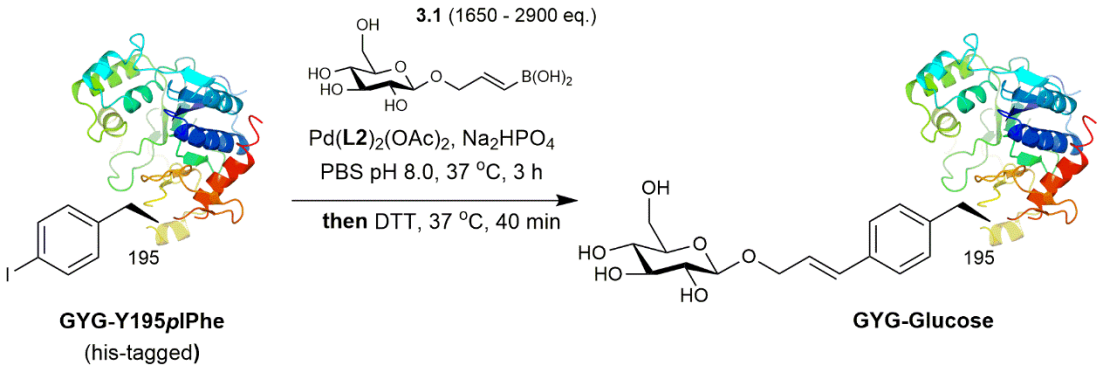
Figure 3.7 Synthesis of glucose-derived boronic acid **3.1**.

3.3 Pd-Mediated Enzyme Activation

3.3.1 Chemical Generation of a Competent GYG Glycoform

With requisite boronic acid partner now in hand, investigation of its cross-coupling to GYG-Y195*p*IPhe could be undertaken. Whilst, as mentioned previously, the on-protein cross-coupling of glucose boronic acid **3.1** has been previously reported,[12] the lack of any quantification of reaction conversions in this case rendered inferences concerning its intrinsic reactivity in Suzuki cross-couplings impossible. Nonetheless, we were cautiously optimistic that this boronic acid might exhibit at least comparable, and possibly greater reactivity towards SMC to GYG-Y195*p*IPhe than the hydroxylated arylboronic acid species explored in the previous chapter, as a consequence of both its excellent water solubility (> 30 mg/ml in our hands) and the relatively sterically unhindered, vinylic nature of the boronic acid functionality in this molecule.

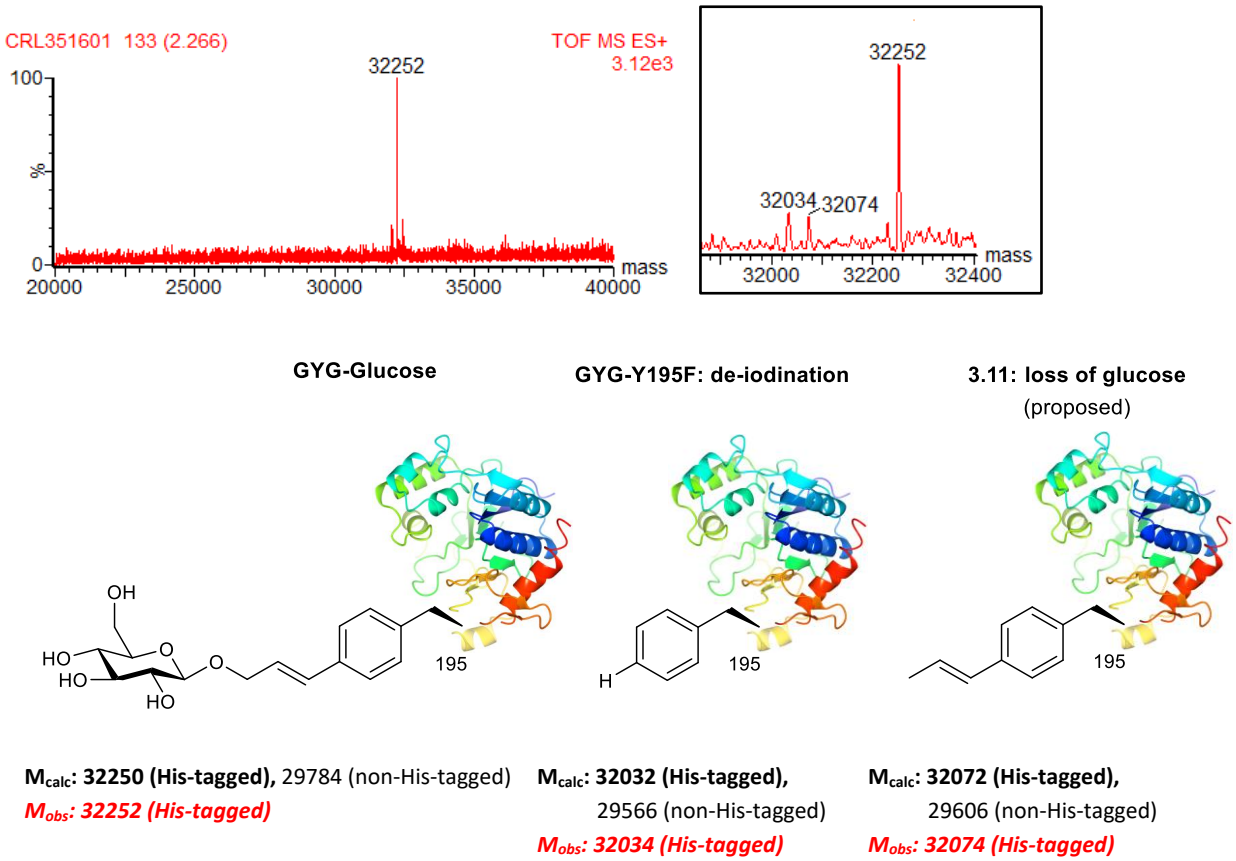
Conditions from the enzyme-compatible coupling of 3-HMPB to his-tagged GYG-Y195*p*IPhe (see chapter 2) were translated to the analogous cross-coupling with **3.1** (**table 3.1**, entry 1); in addition, couplings with increased boronic acid loadings were run (**table 3.1**, entries 2-3). Pleasingly, successful formation of cross-coupled enzyme GYG-Glucose was observed in all cases, with a boronic acid loading of 2500 eq. proving optimal (**table 3.1**, entry 2). The deconvoluted LC-MS spectrum for this entry, illustrating formation of GYG-Glucose, is shown below (**figure 3.8**).



Entry	Equivs. $B(OH)_2$	% Conversion[a]
1	1650	68
2	*2500*	*74*
3	2900	73

[a]*as measured by LC-MS intensity relative to side-products.*

Table 3.1 Cross-couplings of his-tagged GYG-Y195*p*IPhe to glucose boronic acid **3.1** with varying boronic acid loading.



M_{calc}: 32250 (His-tagged), 29784 (non-His-tagged)
M_{obs}: 32252 (His-tagged)

M_{calc}: 32032 (His-tagged), 29566 (non-His-tagged)
M_{obs}: 32034 (His-tagged)

M_{calc}: 32072 (His-tagged), 29606 (non-His-tagged)
M_{obs}: 32074 (His-tagged)

Figure 3.8 Deconvoluted LC-MS spectrum for cross-coupling to GYG-Glucose (from entry 2, table 3.1), with close-up to illustrate masses of reaction side-products. For reference in later sections, masses for both his-tagged and non-his-tagged enzyme are shown.

Close inspection of this spectrum as compared to that of starting material (chapter 2) reveals that the incomplete conversion arose not from incomplete starting material consumption, but from side-product formation (**figure 3.8**).[i] Lack of any observable analogous reaction for iodine-deficient mutant GYG-Y195F implicated both cross-coupled product and side-products formation as occurring at *p*IPhe195 (see experimental section 3.5.5.1). Consistent with prior cross-couplings (chapter 2), a side-product peak corresponding to de-iodination was observed (M_{obs} = 32034); as detailed previously, this is a well-documented side-reaction of Suzuki-Miyaura cross-coupling.[34,35] Unreacted GYG-Y195*p*IPhe (M_{calc} = 32158) was not observed here, but may be observed in certain cases (see later). In addition, a new, previously unobserved side-product (M_{obs} = 32074) was also noted. The precise structural identity of this latter could

[i] A small peak at M = 32430 (+178 Da adduct of coupled enzyme) was additionally noted (see experimental section 3.5.3.1). This arises from the fact that a small proportion (ca. 5-15 %) of the GYG-Y195*p*IPhe starting material is gluconoylated (see chapter 2, **figure 2.4**) – *N*-terminal gluconoylation of His-tagged species being a well-known protein expression artefact[53] – and that this reacts analogously to non-gluconoylated enzyme. As such, it is not a side-product arising from the reaction itself. Consistent with this fact, in later reactions on non-His-tagged enzyme, this adduct was never observed. In general, this adduct was only rarely sufficiently abundant to be clearly visible above MS background noise (only seen in reactions from **table 3.1**). Reaction conversions refer to reaction of the majority, non-gluconoylated enzyme only (though they are essentially unaffected if gluconoylated enzyme is also included).

not be unambiguously confirmed even with knowledge of both mass and reaction site (due to lack of direct literature precedent and more than 1 plausible species affording this mass and modification site), however we believe a likely explanation to be loss of glucose to yield an allyl-linked GYG species such as **3.11** (**figure 3.8**). This is consistent with both the observed mass change and aforementioned site-specificity, and, notably, the fact that this species was never noted for non-carbohydrate cross-couplings (see chapter 2).

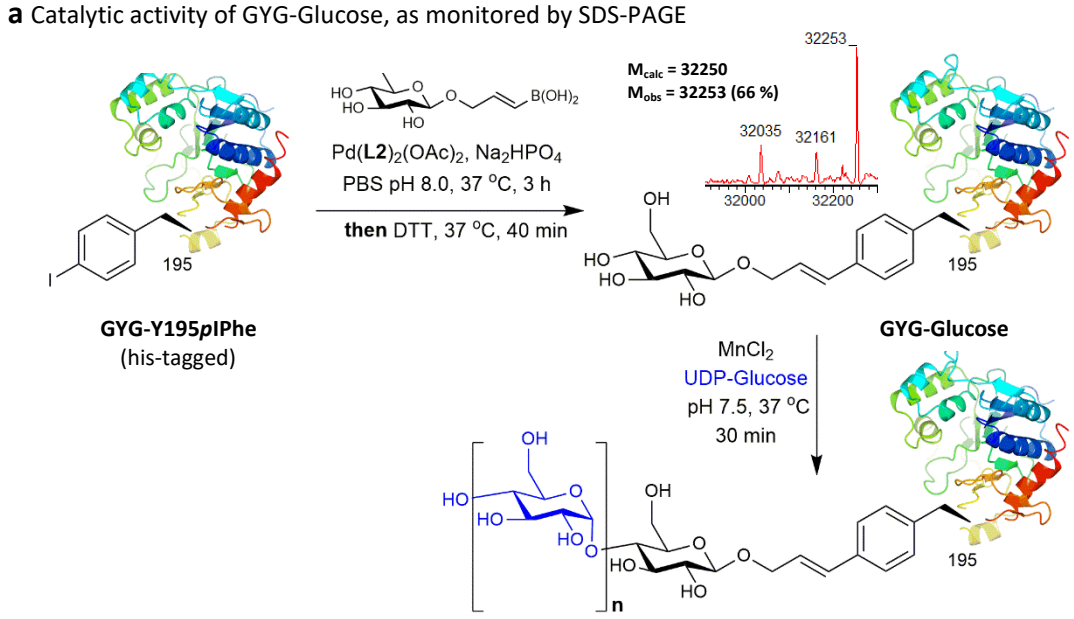
Formation of **3.11** might be rationalised from the well-documented ability of Pd(0)-Pd(II) systems to promote cleavage of allylic glycosides, such as in the product GYG-Glucose, across the C_{allyl}-O bond, via nucleophilic attack of a range of nucleophiles at C_{allyl}.[36–38] Whilst not explicitly demonstrated for glycosides,[ii] a wide range of studies on other allyl alcohol derivatives demonstrate that a number of hydride donors can serve as this nucleophile,[39–43] via a Pd-H species.[41,42] Such C_{allyl}-O hydrogenolytic cleavage in GYG-Glucose would afford **3.11** (the likely presence of Pd-H species may be inferred from the observation of de-iodination product and their known role in this reaction).[34] Whilst a notable alternative explanation for the mass observed could be coupling of a persistent GYG-aryl-Pd intermediate to MeCN upon LC-MS analysis (MeCN-protein adducts have been previously noted,[44] and Pd-catalysed coupling between aryl halides and MeCN has been reported)[45,46], this option is ruled out since this side-product was observed upon LC-MS analysis in the absence of MeCN (see experimental section 3.5.5.2). Thus, we believe **3.11** to represent a mechanistically plausible and highly likely option (especially given its formation only during carbohydrate cross-couplings, but, as seen in later chapters, for multiple different carbohydrate species).

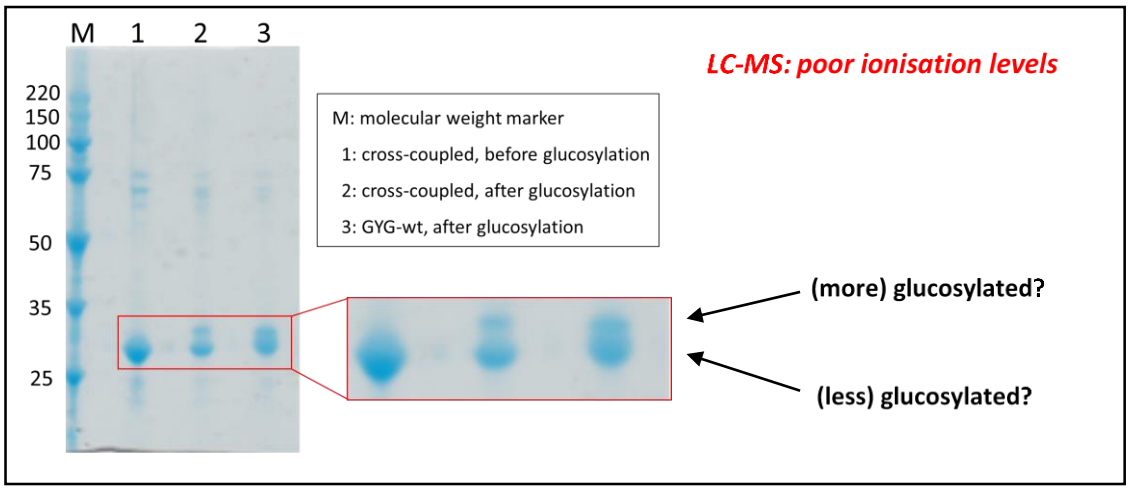
The primary interest for the present study, however, was whether any side-products produced were substrates for glucosylation, and could thus interfere from in subsequent (qualitative) autoglucosylation assays of cross-coupled product. As demonstrated below, we confirmed that this was not the case (consistent with their lack of a glycosyl acceptor at the 195 residue). Indeed, they therefore represented a convenient *in situ* negative control during autoglucosylation assays of cross-coupled product.

We next turned to investigation of the enzymatic activity of GYG-Glucose. As such, it was generated under "enzyme-compatible" conditions (see chapter 2), affording requisite enzyme in 66 % conversion (**figure 6.9**, experimental section 3.5.3.2). Then an autoglucosylation assay was carried out (**figure 3.9 a**), involving incubation of the cross-coupled enzyme with 1 mM each of UDP-Glucose substrate and Mn^{2+} cofactor. As a positive control, a sample of GYG-wt was treated under identical conditions (**figure 3.9 b**).

[ii] This most likely reflects the probable incompatibility of hydride donors (e.g. $NaBH_4$) with protecting groups frequently found on carbohydrates and consequent preference for alternative nucleophiles.

a Catalytic activity of GYG-Glucose, as monitored by SDS-PAGE





b Catalytic activity of identically treated GYG-wt, as detected by LC-MS

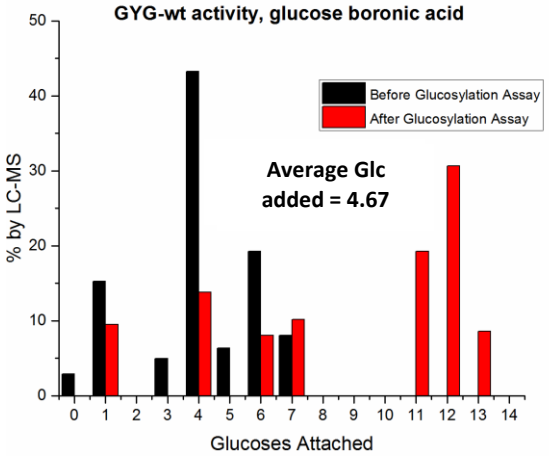


Figure 3.9 (a) SDS-PAGE analysis indicated possible autoglucosylation activity of the GYG-Glucose cross-coupled enzyme. (b) Identically treated wild-type enzyme exhibited clear catalytic activity, as determined by LC-MS.

Whilst GYG-wt was observed to largely retain catalytic activity, as judged by LC-MS (experimental section 3.5.3.2, see chapter 2, section 2.8.6.1 for analytical method), our conclusions regarding the cross-coupled enzyme were unfortunately hindered by poor MS ionisation. However, an SDS-PAGE analysis provided some encouragement. Cross-coupled enzyme before glucosylation showed only one band, yet an additional band of higher molecular weight appeared after glucosylation assay (**figure 3.9 a**). Notably, the same two bands were observed for GYG-wt after glucosylation. We reasoned that these bands may correspond to less glucosylated (or non-glucosylated) enzyme (lower) and highly glucosylated enzyme (upper), an observation that that would implicate catalytic activity of the cross-coupled enzyme. However, unambiguous LC-MS-based confirmation was evidently necessary.

Given the known capability of palladium to cause reduction in LC-MS quality (to the extent of complete signal elimination) through non-specific protein binding,[47,48] we envisaged that the poor LC-MS quality observed above might arise from the continuing presence of residual palladium. Levels of this might be reduced through an increase in the ratio of Pd-scavenger (DTT) to Pd catalyst, which can either be realised through an increase in DTT concentration employed, or a decrease in Pd loading. In the previous chapter, it was noted that higher DTT levels appeared to be slightly detrimental with respect to subsequent enzymatic activity (p. 91 and section 2.8.6.5). This is shown below (**figure 3.10**), in the context of autoglucosylation assays of GYG-wt following its treatment with SMC conditions for the model boronic acid 3-HMPB. Conversely, a reduction in Pd loading appeared to, if anything, be of benefit with regards to enzymatic activity, as judged through LC-MS analysis of SMC-treated GYG-wt before and after glucosylation assay (**figure 3.10**, note that all samples in this figure are reactions carried out in parallel). This therefore appeared to represent the more productive avenue to pursue. (Additionally, the total ion count upon LC-MS analysis was significantly increased upon reduction of Pd loading, but not upon increased DTT concentration, see experimental section 3.5.3.3).

As such, the cross-coupling of GYG-Y195*p*IPhe to **3.1** at lower Pd concentrations was investigated. Whilst a decline in cross-coupling efficiency with decreasing Pd catalyst concentration was observed, as would be expected, this was relatively shallow in nature (**table 3.2**). Indeed, at 0.6 mM Pd concentration, as utilised in the previous study on GYG-wt, a conversion of 65 % could still be attained (**table 3.2**, entry 2), which was deemed sufficient at this stage for qualitative evaluation of enzymatic activity, given that all side-products do not exhibit capacity for self-glucosylation; at lower Pd concentrations, a more significant drop-off in conversion was noted (**table 3.2**. entry 3).

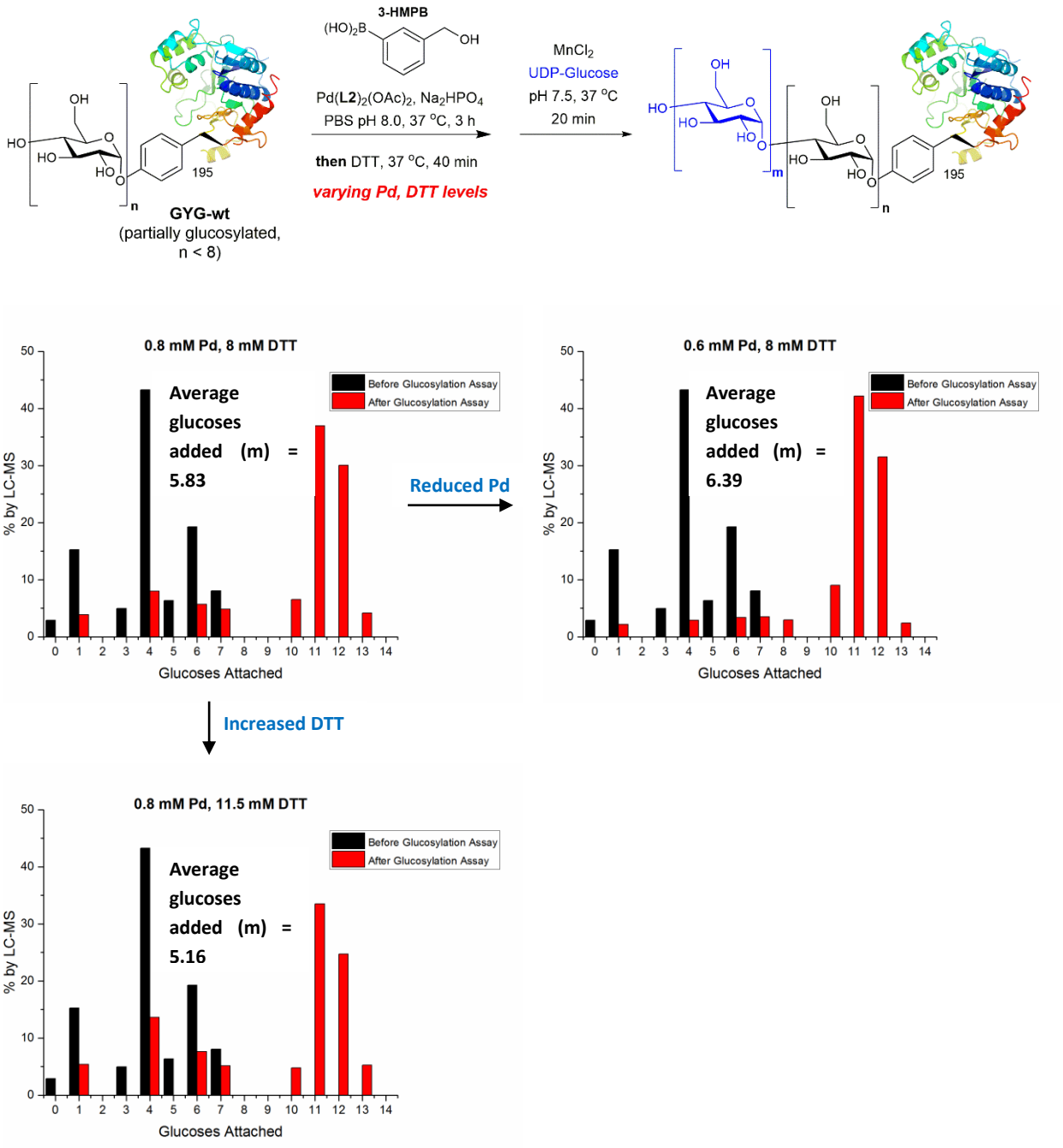
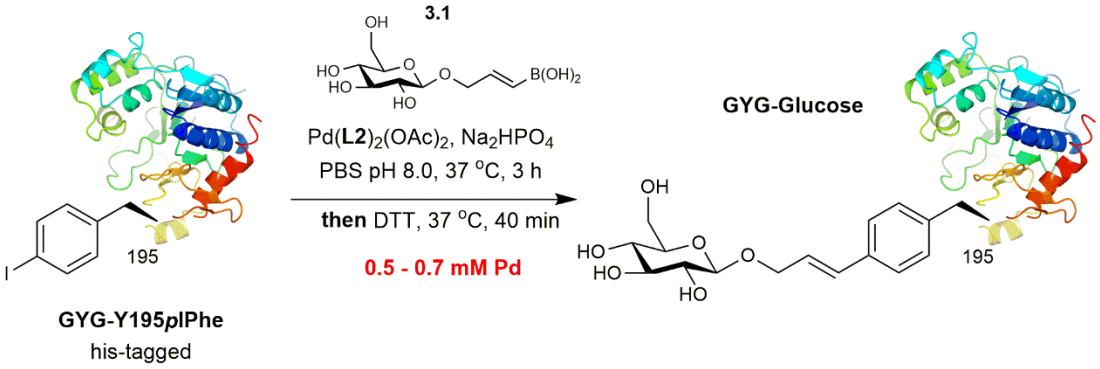


Figure 3.10 Effect of varying Pd and DTT levels on GYG activity, as judged through the use of wild-type GYG (GYG-wt) in the model system employed in the previous chapter. Samples were all run in parallel to each other. Graphs show GYG glycoform distributions before and after glucosylation assay, determined from LC-MS spectra (see section 3.5.3.3) according to the method detailed in the preceding chapter (section 2.8.6.1); this method was also used to determine average glucoses added.

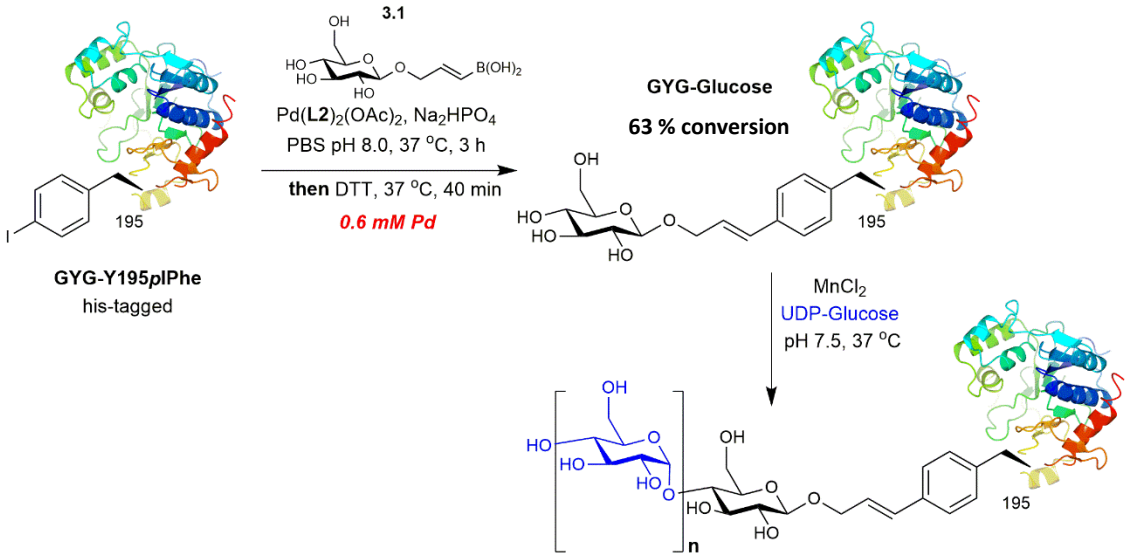


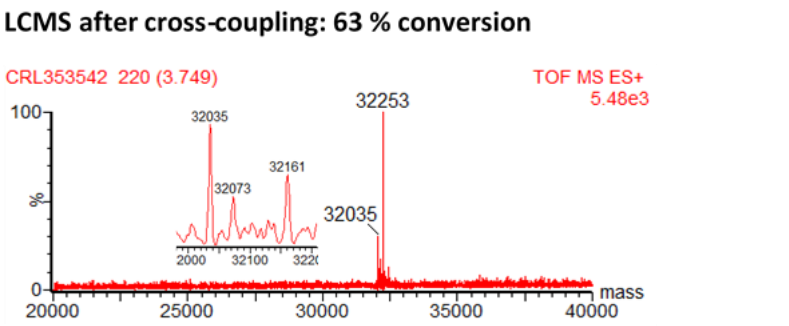
Entry	[Pd] / mM	% Product[a]
1	0.7	72
2	0.6	65
3	0.5	52

[a]as measured by LC-MS intensity relative to side-product peaks.

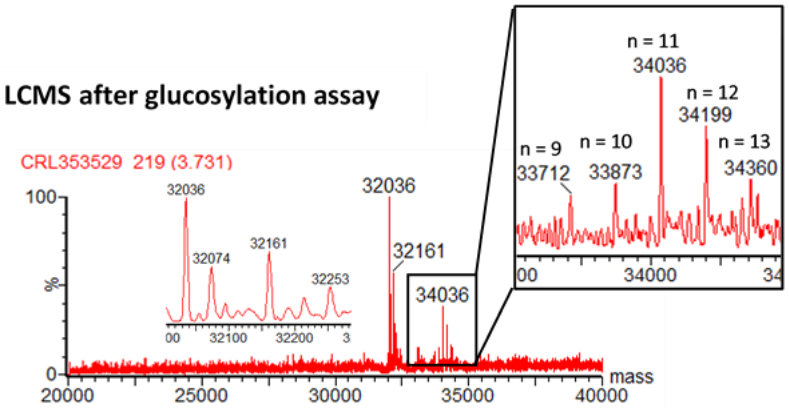
Table 3.2 Effect of Pd concentration on cross-coupling conversion for reaction between GYG-Y195*p*IPhe and boronic acid **3.1**.

The glucosylation assay outlined in **figure 3.9** was repeated on GYG-Glucose prepared utilising this lower Pd catalyst concentration. Whilst a low amount of unreacted GYG-Y195*p*IPhe now persisted after coupling, the LC-MS quality after glucosylation was indeed improved compared to prior attempts, and, most significantly, confirmed that this cross-coupled enzyme was indeed catalytically active (**figure 3.11**). Consumption of cross-coupled enzyme could be observed (relative to the catalytically inactive side-products), with concomitant appearance of peaks corresponding to addition of multiple glucose residues to this enzyme. In contrast, the side-products from the cross-coupling exhibited no catalytic activity, as no peaks corresponding to their glucosylation were observed, confirming specificity of glucosylation. Indeed, these inactive species afford the highest intensity peaks in the LC-MS spectrum, reflecting their lack of consumption; in contrast, the peak corresponding to GYG-Glucose is split into multiple peaks corresponding to different glycoforms. Though preliminary, this initial result represented the first confirmation that our methodology could indeed access catalytically active mimetics of GYG autoglucosylation intermediates, exemplified here through formation of a "homogeneously mono-glucosylated" species. This indeed represents an example of a ligation-dependent, Pd-mediated enzyme activation, from inactive GYG-Y195*p*IPhe to active GYG-Glucose.





M_{calc}: 32250
M_{obs}: 32253 (63 %)
+ non-glucosylatable side-products (32035, 32073, 32161), see pp. 172-3 and **figure. 3.8**



M_{calc}: 32250 + 162n
M_{obs}: 32254 + 162n (n = 0, 9-13)
+ non-glucosylated side-products (32036, 32074, 32161), see pp. 172-3 and **figure. 3.8**

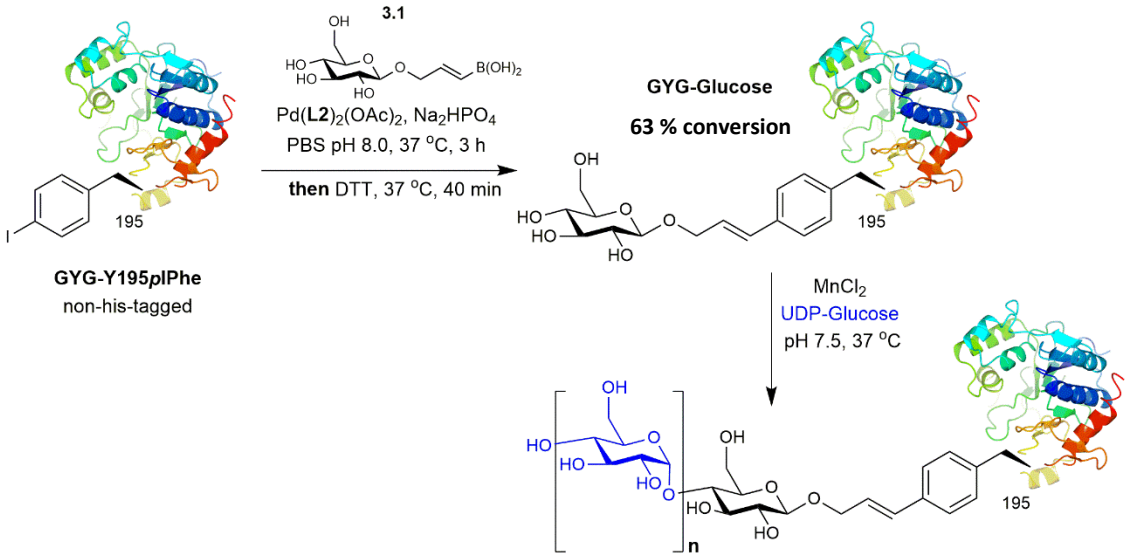
Figure 3.11 Catalytic activity of GYG-Glucose, as illustrated through LC-MS analysis. After glucosylation assay, the MS peak corresponding to GYG-Glucose splits into multiple peaks corresponding to addition of varying numbers of glucose residues. No glucosylation of cross-coupling side-products was seen, and, consistent with this, consumption of GYG-Glucose relative to such species was evident.

3.3.2 Optimisation of Pd-mediated Enzyme Activation

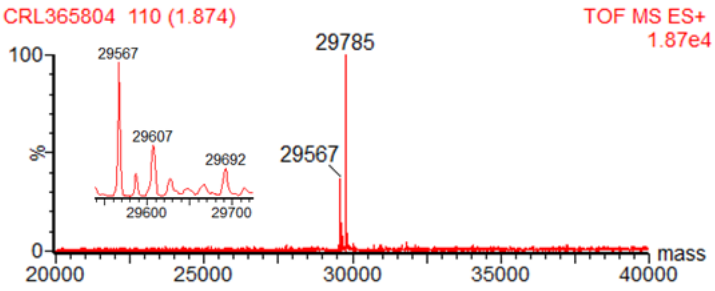
It was evident that improvements to the SMC-glucosylation procedure were both possible and desirable. In particular, upon further repeats, it quickly became apparent that further improvements in the quality (signal/noise ratio) for glucosylated enzyme peaks in LC-MS spectra would be necessary in order to carry out more detailed, quantitative analysis of glucosylation (see chapter 4) with a sufficient level of confidence. This could primarily be effected through improvements in the cross-coupling step, since the inactive side products thereof may dominate spectra obtained after glucosylation.

However, prior to this, given the known capacity for polyhistidine moieties to bind Pd[49,50] and the fact that non-specific coordination of Pd by proteins can cause suppression of MS ionisation,[47] we elected to cleave the his-tag from GYG-Y195*p*IPhe; this was effected via treatment with TEV protease (see experimental section 3.5.2; this was carried out as part of a collaboration by Mr. Henry Bailey, Structural Genomics Consortium, University of Oxford). The resultant non-his-tagged enzyme underwent cross-coupling with near-identical conversion under the same conditions, with the same side-product species observed (see **figure 3.8**), and exhibited catalytic activity as expected (**figure 3.12**). In general, a slight improvement in quality of MS spectra obtained after both cross-coupling and glucosylation was indeed noted.

Further cross-coupling optimisations were then carried out on this enzyme (**table 3.3**, see also experimental section 3.5.4.2), varying a number of different factors, without resorting to increasing the palladium loading, given the latter's previously observed detrimental effects on enzymatic activity and MS quality. It should be emphasised that the data in this table is derived from samples treated in parallel, in order to minimise any extraneous sources of conversion variation between samples and permit fair comparison. Increasing the protein concentration was found to be slightly detrimental, presumably due to concomitant reduction in the equivalents of other reagents which thus became conversion-limiting (**table 3.3**, entries 1, 2). Noting that the boronic acid is typically added from a stock solution in aqueous Na_2HPO_4 both in this work and in previous protocols,[12,47,51] the latter presumably serving as a base to promote the formation of the more reactive boronate species, we wondered if the addition of extra such base may prove beneficial to the reaction. This was not the case (**table 3.3**, entry 3), and conversely it was observed that total omission of this base led to an enhancement of reaction conversion (**table 3.3**, entry 4), through a reduction in the de-iodination pathway (see experimental section 3.5.4.2). Furthermore, a moderate increase in boronic acid loading was also observed to result in improved conversion (**table 3.3**, entry 5).

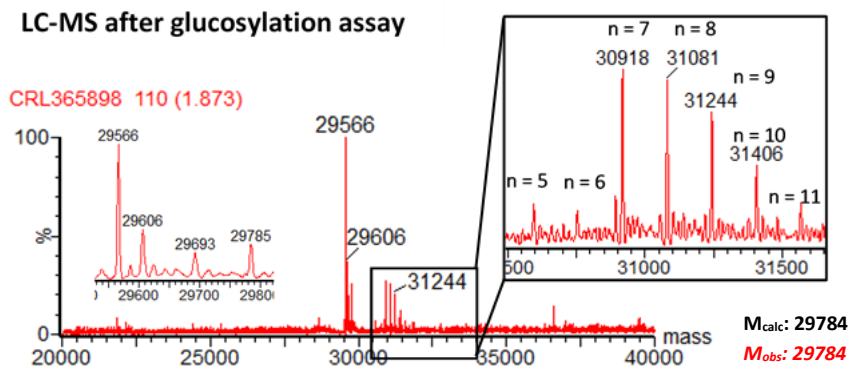


LC-MS after cross-coupling: 63 % conversion



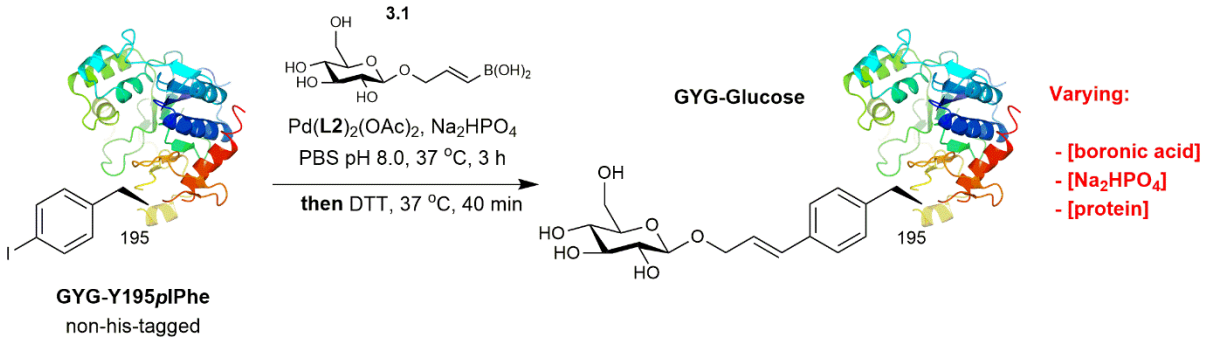
M_{calc}: 29784
M_{obs}: 29785 (63 %)
+ non-glucosylatable side-products (29567, 29607, 29692), see pp. 172-3 and **figure 3.8**

LC-MS after glucosylation assay



M_{calc}: 29784 + 162n
M_{obs}: 29784 + 162n (n = 0, 5-11)
+ non-glucosylated side-products (29566, 29606, 29693), see pp. 172-3 and **figure 3.8**

Figure 3.12 Cross-coupling to generate non-his-tagged GYG-Glucose and autoglucosylation activity thereof.

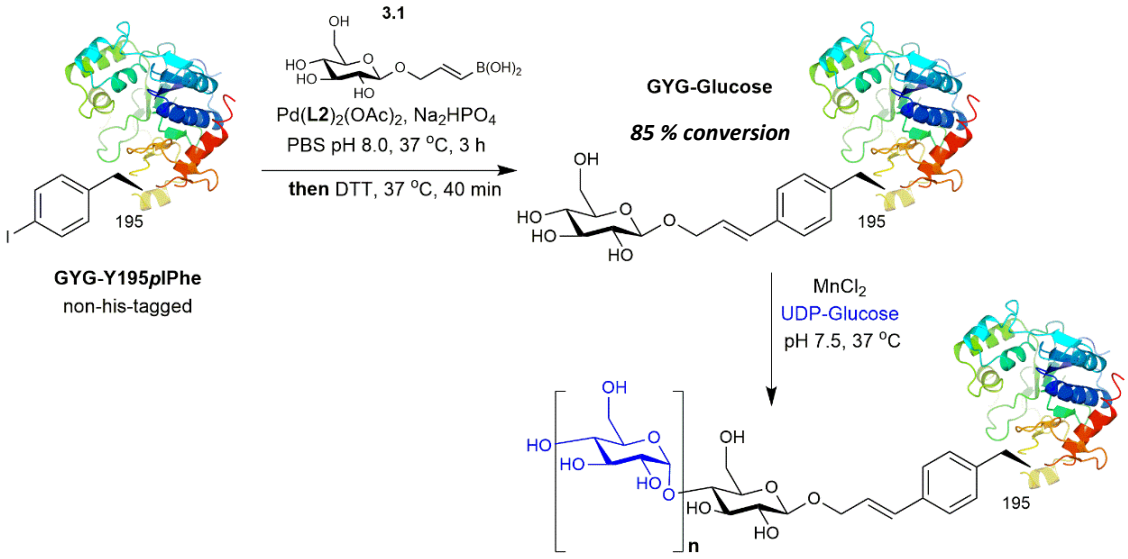


Entry	Equivs. $B(OH)_2$	[GYG-Y195*p*IPhe]	[Na_2HPO_4]	% Conversion[a]
1	2500	0.4 mg/ml	5.1 mg/ml	61
2	2500	**0.5 mg/ml**	5.1 mg/ml	57
3	2500	0.4 mg/ml	**7.7 mg/ml**	61
4	2500	0.4 mg/ml	**none**	69
5	**3400**	0.4 mg/ml	5.1 mg/ml	69

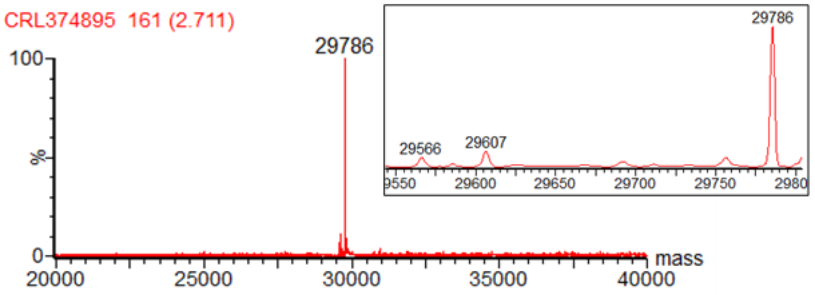
[a]as measured by LC-MS intensity with respect to side-products.

Table 3.3 Further optimisation of cross-coupling to boronic acid **3.1** using non-his-tagged GYG-Y195*p*IPhe. Samples were treated in parallel in order to minimise any extraneous sources of variation in conversion.

In separate experiments (i.e. not run in parallel to those in **table 3.3**), it was also confirmed that neither a further increase in boronic acid loading or reaction time had any beneficial effect on reaction conversion (see experimental section 3.5.4.3). As such, our optimised conditions combined the omission of Na_2HPO_4 (entry 4) with the boronic acid level utilised in entry 5. This ultimately enabled the generation of GYG-Glucose in excellent conversions of up to 85 %, in comparison to previous conversions of ca. 60 % (**figure 3.13**, see experimental section 3.5.3.4). Side-products, whilst reproducibly the same as previously observed, were minimal in abundance. Subsequent autoglucosylation highlighted the improved LC-MS quality resulting from a reduction in the quantities of catalytically inactive side-products (**figure 3.13**).

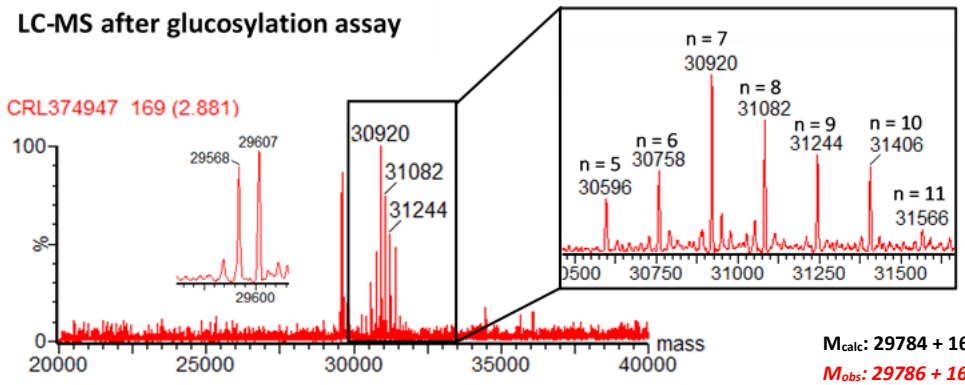


LC-MS after cross-coupling: *85 % conversion*



M_{calc}: 29784
M_{obs}: 29785 (85 %)
+ non-glucosylatable side-products (29566, 29607), see pp. 172-3 and **figure 3.8**

LC-MS after glucosylation assay



M_{calc}: 29784 + 162n
M_{obs}: 29786 + 162n (n = 5-11)
+ non-glucosylated side-products (29566, 29607), see pp. 172-3 and **figure 3.8**

Figure 3.13 Optimised cross-coupling and autoglucosylation assay, with comparison to prior unoptimised data.

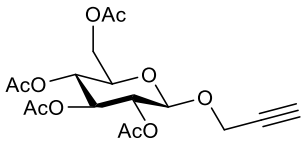
3.4 Conclusion and Outlook

Through the synthesis and cross-coupling of a glucose-derived boronic acid to GYG-Y195*p*IPhe, a mimic of a GYG glycoform resembling the natural mono-glucosylated species was generated. Preliminary studies revealed that this enzyme evidently bore sufficient resemblance to naturally glucosylated GYG to exhibit catalytic activity, and as such its formation represented a novel example of palladium-mediated enzyme activation. Further optimisation of the cross-coupling step led to improvement of reaction conversions from moderate (ca. 60%) to high levels (85 %), and greatly enhancing our ability to analyse the glucosylation through LC-MS methods. Cross-coupling side-products derived from de-iodination, lack of coupling, and a hypothesised cleavage of the glucose moiety to afford a "GYG-allyl" species **3.11**. However, these were not only present in very low levels under optimised conditions, but additionally did not represent substrates for glucosylation.

Transition metal-mediated enzyme activation remains at present a developing field rather than a well-established one, and examples of the use of ligation chemistry to activate enzymes are all but absent.[32] To the best of our knowledge, this work represents the first example of a combination of these two approaches in one strategy. Yet of arguably still greater significance is the fact that this methodology enables unprecedented access to a catalytically active, homogeneously glucosylated glycoform of GYG, representing (a mimic of a) catalytic intermediate of autoglucosylation of the latter. With the methodology optimised for the particular example of GYG-Glucose, we were in a position not only to attempt more detailed, quantitative analysis of the glucosylation of this enzyme, but also to expand it to the generation and similar analysis of other GYG glycoform mimetics. Evaluation and comparison of such species offers the potential to gain new insights into the complex catalytic mechanism of the enzyme, and forms the basis of the subsequent chapters of this thesis.

3.5 Experimental Section

3.5.1 Chemical Synthesis



3.7

β-*D*-Glucose pentaacetate (12 g, 30.8 mmol) was dissolved in anhydrous DCM in a flame-dried flask and cooled to 0 °C. Propargyl alcohol (3.00 ml, 51.1 mmol) then $BF_3.OEt_2$ (5.78 ml, 46.0 mmol) were added slowly. The reaction mixture was stirred for 16 h at room temperature under argon. K_2CO_3 (4 g) was added to quench the reaction, and the suspension stirred for 0.5 h. The mixture was filtered, and the organic layer washed with H_2O (2 x 100 ml). The aqueous layer was then back-extracted with DCM (100 ml). The combined organics were washed with brine (100 ml), dried ($MgSO_4$), filtered and concentrated *in vacuo* to afford a brown solid. Recrystallisation from EtOH gave the desired product **3.7** as an off-white solid (7.8 g, 66 %), as exclusively the β-anomer.

^{1}H NMR (400 MHz, Chloroform-*d*) δ 5.22 (t, *J* = 9.5 Hz, 1H, H3), 5.08 (t, *J* = 9.7 Hz, 1H, H4), 4.99 (dd, *J* = 9.6, 8.0 Hz, 1H, H2), 4.76 (d, *J* = 8.0 Hz, 1H, H1), 4.35 (d, *J* = 2.4 Hz, 2H, CH_2CCH), 4.25 (dd, *J* = 12.4, 4.6 Hz, 1H, H6a), 4.12 (dd, *J* = 12.3, 2.4 Hz, 1H, H6b), 3.71 (ddd, *J* = 10.0, 4.6, 2.4 Hz, 1H, H5), 2.46 (t, *J* = 2.4 Hz, 1H, CH_2CCH), 2.06 (s, 3H, OAc), 2.04 (s, 3H, OAc), 2.00 (s, 3H, OAc), 1.98 (s, 3H, OAc).

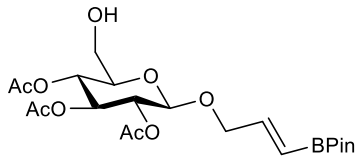
LRMS *m/z* (ESI+) found: 409.1 (M+Na, calculated 409.1)

IR (neat) $\tilde{\nu}$ = 3273 (C≡**C-H)**, 2970 (C-H), 2119 (C≡C), 1732 (C=O) cm^{-1}

m.p. 114-116 °C (EtOH), lit.[12] 114-115 °C

$[\alpha]_D$ = −42 (*c* = 0.5, $CHCl_3$), lit.[12] = -39.

Data consistent with that previously reported.[12]

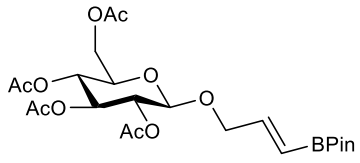


3.8

Pinacolborane (1.50 ml, 1.34 g, 10.4 mmol) was added to **3.7** (1.00 g, 2.6 mmol) in a flame-dried flask under N_2. Schwartz's reagent (168 mg, 0.66 mmol) and triethylamine (0.092 ml, 66 mg, 0.66 mmol) were then added, and the reaction stirred at 65 ^oC under N_2 for 16 h. The resultant brown slurry was diluted with DCM (10 cm^3), filtered, and concentrated *in vacuo*. The residue was purified by flash column chromatography (SiO_2, eluting with 7:3 -> 1:1 Pet. Ether: EtOAc) to yield an off-white foam. ^1H NMR and LRMS indicated that mono-deacetylation had occurred during the reaction. The 6'-OAc group was indicated as the site of this by significant downfield shift of the H6 protons upon re-acetylation (other protons remained at similar shifts), as well as 2D NMR experiments. The resultant species **3.8** was therefore subjected directly to a subsequent re-acetylation step prior to full characterisation (however, its presence was confirmed by NMR and MS).

^1H NMR (400 MHz, $CDCl_3$) δ 6.58 (dt, *J* = 18.2, 4.8 Hz, 1H, PinB-CHCH), 5.66 (dt, *J* = 18.2, 1.8 Hz, 1H, PinB-CH), 5.24 (t, *J* = 9.6 Hz, 1H, H3), 5.02 (t, *J* = 9.7 Hz, 1H, H4), 5.00 (dd, *J* = 9.7, 7.9 Hz, 1H, H2), 4.56 (d, *J* = 8.0 Hz, 1H, H1), 4.37 (ddd, *J*=14.7, 4.6, 1.8 Hz, 1H, CH_2CHCH-BPin), 4.22 (ddd, *J* = 14.6, 5.0, 1.7 Hz, 1H, CH_2CHCH-BPin), 3.72 (dd, *J* = 12.6, 2.4 Hz, 1H, H6a), 3.60 (dd, *J* = 12.6, 5.0 Hz, 1H, H6b), 3.50 (ddd, *J*=10.1, 5.0, 2.4 Hz, 1H, H5), 2.06 (s, 3H, OAc), 2.04 (3H, s, OAc), 2.01 (s, 3H, OAc), 1.22 (s, 12H, BPin)

LRMS *m/z* (ESI+) found: 495.2 (M+Na, calculated 495.2).



3.9

Intermediate **3.8** (490 mg, 1.04 mmol) was dissolved in anhydrous pyridine (2.7 ml) and acetic anhydride (0.3 ml) was added at 0 ^oC. The reaction was stirred at r.t. for 16 h. After this time, the mixture was diluted with EtOAc (10 ml) and H_2O (10 ml). The organic layer was separated and the aqueous layer extracted with EtOAc (3 x 40 ml). The combined organics were dried ($MgSO_4$), filtered and concentrated *in vacuo*. The

resultant oil was azeotroped with toluene (3 x 20 ml) to remove residual pyridine, leading to the isolation of pinacol ester **3.9** as an off-white solid (410 mg, 0.80 mmol, 31 % from **3.7**) as exclusively the β-anomer.

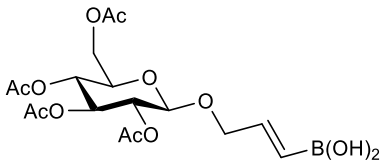
^{1}H NMR (400 MHz, Chloroform-*d*) δ 6.55 (ddd, *J* = 18.2, 5.1, 4.2 Hz, 1H, PinBCH=CH), 5.67 (dt, *J* = 18.1, 1.8 Hz, 1H, PinBCH=CH), 5.20 (t, *J* = 9.4 Hz, 1H, H3), 5.09 (t, *J* = 9.7 Hz, 1H, H4), 5.03 (dd, *J* = 9.6, 7.9 Hz, 1H, H2), 4.54 (d, *J* = 7.9 Hz, 1H, H1), 4.40 (ddd, *J* = 14.7, 4.2, 1.9 Hz, 1H, BCH=CHCH$_2$), 4.25 (dd, *J* = 12.3, 4.7 Hz, 1H, H6a), 4.18 (ddd, *J* = 14.8, 5.1, 1.7 Hz, 1H, BCH=CHCH$_2$), 4.13 (dd, *J* = 12.3, 2.4 Hz, 1H, H6b), 3.67 (ddd, *J* = 9.9, 4.7, 2.4 Hz, 1H, H5), 2.09 (s, 3H, OAc), 2.07 (s, 3H, OAc), 2.02 (s, 3H, OAc), 2.00 (s, 4H, OAc), 1.26 (s, 12H, BPin).

LRMS *m/z* (ESI+) found: 537.2 (M+Na, calculated 537.3)

m.p. 146-147 °C (toluene), lit.[52] 140-141 °C.

[α]$_D$ = -20 (c = 1.0, $CHCl_3$).

Data consistent with that previously reported.[12,52]



3.10

Diethanolamine (113 mg, 1.08 mmol) was added to a dried flask, followed by anhydrous Et_2O (18 ml). Pinacol ester **3.9** (360 mg, 0.70 mmol) was added as a solid. The suspension was stirred vigorously under N_2 for 18 h. The resultant white precipitate was collected by gravity filtration and washed with 2 x 30 ml Et_2O. ^{1}H NMR confirmed the presence of the intermediate diethanolamine ester (210 mg, 0.42 mmol, 60 %): **^{1}H NMR** (400 MHz, DMSO-d_6) δ 6.76 (br.s., 1H, NH), 5.73 (dt, *J* = 18.0, 5.1 Hz, 1H, R$_2$BCH=CH), 5.57 (d, *J* = 18.0 Hz, 1H, R$_2$BCH=CH), 5.27 (dt, *J* = 9.4, 4.8 Hz, 1H, H2), 4.90 (t, *J* = 9.4 Hz, 1H, H3), 4.77 (dd, *J* = 4.8, 1.2 Hz, 1H, H1), 4.14 - 4.22 (m, H, H4), 4.05 – 3.94 (m, 2H, BCH=CHCH$_2$), 3.73-3.71 (m, 2H, BOCH$_2$), 3.61-3.59 (m, 2H, BOCH$_2$), 3.46 (t, *J* = 5.8 Hz, 1H, H5), 3.03 – 2.94 (m, 2H, HNCH$_2$), 2.78 - 2.69 (m, 2H, HNCH$_2$), 2.60 (app t, *J* = 5.8 Hz, 2H, H6), 2.04 (s, 3H, OAc), 2.02 (s, 3H, OAc), 1.99 (s, 3H, OAc), 1.96 (s, 3H, OAc).

160 mg of this intermediate was resuspended in a biphasic mixture of Et_2O / 0.1M HCl (3:2 v/v, total volume 18.75 ml). The reaction was stirred at r.t. until complete dissolution of the solid was observed (2 h). At this point, the organic layer was separated and the aqueous extracted with Et_2O (3 x 40 ml). The combined organics were dried ($MgSO_4$), filtered and concentrated *in vacuo* to yield protected boronic acid **3.10** as a white solid (128 mg, 0.30 mmol, 71 % w.r.t. diethanolamine ester intermediate, 43 % w.r.t. **3.9**) as exclusively the β-anomer.

^{1}H NMR (500 MHz, Acetone-d_6) δ 6.48 (dt, J = 18.1, 4.8 Hz, 1H, BCH=C<u>H</u>), 5.63 (dt, J = 18.2, 1.8 Hz, 1H, $(HO)_2$B-C<u>H</u>), 5.23 (t, J = 9.6 Hz, 1H, <u>H3</u>), 5.00 (t, J = 9.8 Hz, 1H, <u>H4</u>), 4.90 (dd, J = 9.9, 8.0 Hz, 1H, <u>H2</u>), 4.75 (d, J = 8.0 Hz, 1H, <u>H1</u>), 4.33 (ddd, J = 14.3, 4.4, 1.9 Hz, 1H, BCH=CHC<u>H_2</u>), 4.24 (dd, J = 12.3, 4.8 Hz, 1H, <u>H6a</u>), 4.15 (ddd, J = 14.4, 5.2, 1.7 Hz, 1H, BCH=CHC<u>H_2</u>), 4.10 (dd, J = 12.3, 2.5 Hz, 1H, <u>H6b</u>), 3.93 (ddd, J = 9.9, 4.7, 2.4 Hz, 1H, <u>H5</u>), 2.01 (s, 3H), 2.00 (s, 3H), 1.97 (s, 3H), 1.93 (s, 3H).

^{13}C NMR (126 MHz, Acetone-d_6) δ 169.2, 168.6, 168.3, 168.2 (all O<u>Ac</u> C=O), 143.2 (BCH=<u>C</u>H), 123.5 (B<u>C</u>H=CH, weak), 98.2 (<u>C1</u>), 71.2 (<u>C3</u>), 69.9 (<u>C5</u>), 69.9 (<u>C2</u>), 69.2 (<u>C</u>H_2CHCHB$(OH)_2$), 67.0 (<u>C4</u>), 60.4 (<u>C6</u>), 18.4, 18.3, 18.3 (all O<u>Ac</u> CH_3).

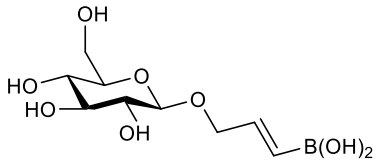
HRMS *m/z* (ESI+) found: 455.1332 (M+Na, calculated 455.1331)

IR (neat) $\tilde{\nu}$ = 3458 (O-H), 2965 (C-H), 1752, 1727 (C=O), 1633 (C=C) cm^{-1}

m.p. > 300 °C (Et_2O), phase transition to amorphous solid observed at 95-98 °C.

$[\alpha]_D$ = −21 (c = 0.6, $CHCl_3$)

It should be noted that, whilst this compound has been previously synthesised,[12] its characterisation was not previously reported, thus it is not considered as a "known" compound here.



3.1

Protected boronic acid **3.10** (100 mg, 0.23 mmol) was dissolved in anhydrous MeOH (3.3 ml) under argon. NaOMe (0.06 ml of a 25 wt% solution in MeOH, 14 mg, 0.26 mmol) was added dropwise and the yellow solution stirred for 2 h at r.t., after which time TLC and LC-MS indicated reaction completion. Pre-activated DOWEX®-H^+ (see note below) was added until pH was neutral, and the suspension stirred for a further 30 minutes. The resin was removed by filtration, and washed with MeOH. The filtrate was concentrated *in vacuo* to afford 60 mg of crude product. This was redissolved in 2 ml H_2O and passed through a 0.2 μm filter to remove small insoluble particulate impurities. The filtrate was concentrated *in vacuo* to yield boronic acid **3.1** as a white crystalline solid (40 mg, 0.15 mmol, 66 %).

Note: DOWEX®-H^+ was activated as follows. Unactivated resin (ca. 5 g) was stirred in ca. 200 ml of MeOH for 5 min, before being filtered. This was repeated with ca. 200 ml acetone. The resin thus obtained was stirred in 2M HCl for 30 minutes, before being filtered. Resin was washed with H_2O until the pH of the filtrate was neutral, and was then dried on the filter and stored in a sealed tube. Resin should ideally be prepared fresh before use to avoid any leaching of colour into isolated compounds. Any future use of DOWEX®-H^+ refers to resin activated by this method.

^{1}H NMR (500 MHz, Deuterium Oxide) δ 6.54 (dt, *J* = 18.3, 5.1 Hz, 1H, BCH=C<u>H</u>), 5.76 (dt, *J* = 18.3, 1.6 Hz, 1H, BC<u>H</u>=CH), 4.49 (d, *J* = 7.9 Hz, 1H, <u>H1</u>), 4.45 (dd, *J* = 4.8, 1.7 Hz, 1H, BCH=CHC<u>H_2</u>), 4.33 (ddd, *J* = 14.2, 5.4, 1.5 Hz, 1H, BCH=CHC<u>H_2</u>), 3.91 (dd, *J* = 12.3, 2.2 Hz, 1H, <u>H6a</u>), 3.72 (dd, *J* = 12.3, 5.9 Hz, 1H, <u>H6b</u>), 3.49 (t, *J* = 9.1 Hz, 1H, <u>H3</u>), 3.45 (ddd, *J* = 10.0, 6.0, 2.2 Hz, 1H, <u>H5</u>), 3.38 (dd, *J* = 9.8, 8.9 Hz, 1H, <u>H4</u>), 3.31 (dd, *J* = 9.3, 8.0 Hz, 1H, <u>H2</u>).

LRMS *m/z* (ESI+) found: 287.1 (M+Na, calculated 287.1)

m.p. >300 °C (H_2O), lit.[12] >300 °C

$[\alpha]_D$ = −15 (*c* = 0.6, H_2O), lit.[12] -18.

Data consistent with that previously reported.[12]

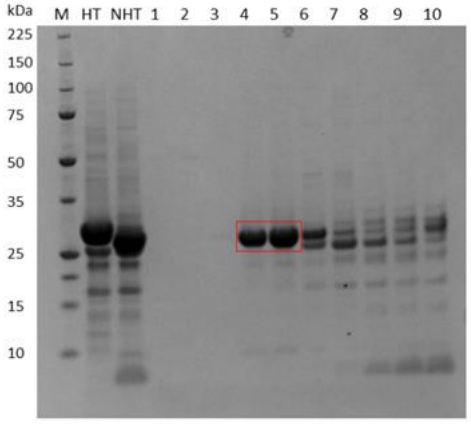
3.5.2 TEV Cleavage of his-tagged GYG-Y195*p*IPhe

This experiment, with the exception of the LC-MS analysis, was carried out by Mr. Henry Bailey, a DPhil student in the group of Prof. Wyatt Yue (Structural Genomics Consortium, Oxford), as part of a collaboration.

His-tagged GYG-Y195*p*IPhe was expressed and purified as reported in the previous chapter (see section 2.8.3.1). However, directly after purification by Ni-NTA chromatography, instead of being dialysed, fractions containing GYG-Y195*p*IPhe were concentrated to a volume of 5 ml and loaded onto an S200 16/600 gel filtration column in GF buffer. Eluant fractions were pooled (6.75 mg) and concentrated to 9 mg/ml. 3.6 mg of protein was incubated with 0.4 mg of TEV protease overnight on ice. The sample was then passed over 500 µl of Ni-NTA resin. The flow through was collected and the resin was washed with 2 x 1 ml GF buffer, 2 x 1 ml GF buffer with 20 mM imidazole, 1 x 1ml GF buffer with 40 mM imidazole, 1 x 1ml GF buffer with 60 mM imidazole, 1 x 1ml GF buffer with 80 mM Imidazole and 2 x 1 ml GF buffer with 250 mM imidazole. Fractions containing the desired protein were combined, divided into 0.5 ml aliquots, flash frozen in liquid N_2 and stored at -80 °C.

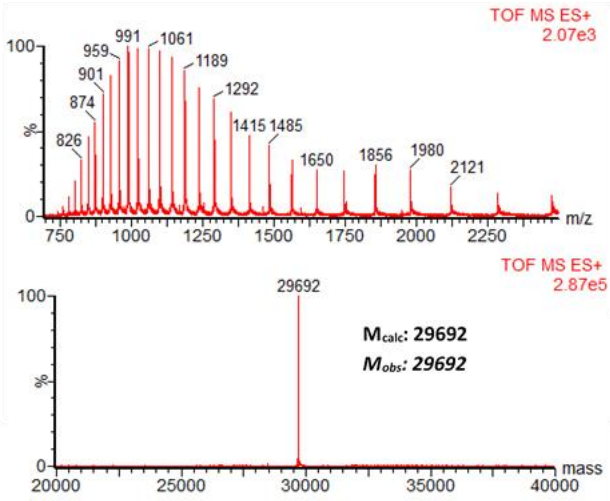
GF buffer: 50 mM Hepes pH 7.5, 500 mM NaCl, 0.5 mM TCEP, 5% Glycerol pH 7.5

SDS-PAGE:



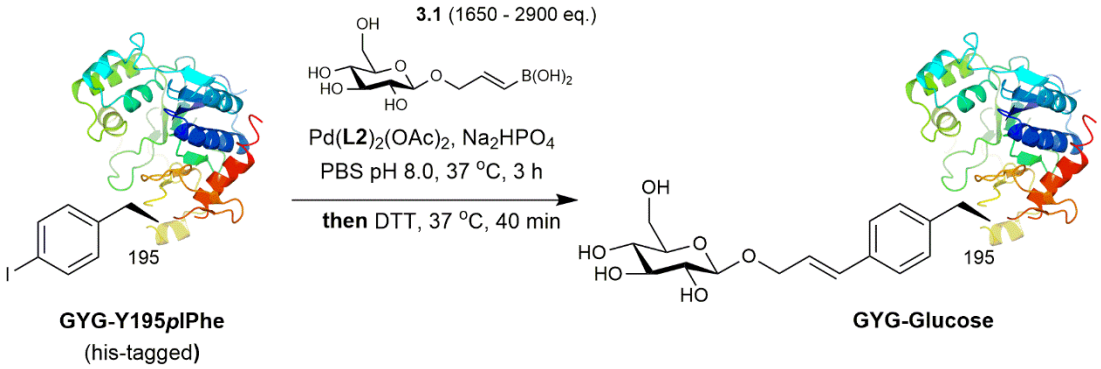
M = molecular weight markers
HT = his-tagged protein (reference)
NHT = non-his-tagged protein (reference)
1 = flowthrough
2, 3 = GF buffer wash
4, 5 = GF buffer + 20 mM imidazole
6 = GF buffer + 40 mM imidazole
7 = GF buffer + 60 mM imidazole
8 = GF buffer + 80 mM imidazole
9, 10 = GF buffer + 250 mM imidazole

LC-MS:



3.5.3 Generation and Enzymatic Activity of his-tagged GYG-Glucose

3.5.3.1 Initial cross-couplings: optimisation of Boronic Acid Loading

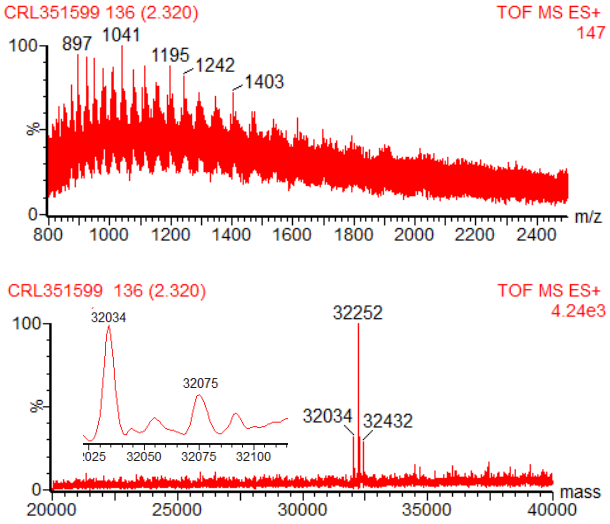


3 reactions were set up. A stock solution of glucose boronic acid **3.1** was prepared by dissolving boronic acid (27 mg/ml) and Na_2HPO_4 (15.7 mg/ml = 110 mM, 1.1 eq. w.r.t. boronic acid) in H_2O. Varying volumes of his-tagged GYG-Y195*p*IPhe (0.56 nmol) and boronic acid solution were combined as follows:

Reaction	Amount GYG-Y195*p*IPhe	Amount of Boronic Acid 3.1
1	30 µl of 0.6 mg/ml, 0.56 nmol	9.0 µl of 27 mg/ml stock solution[a], 920 nmol, 1650 eq.
2	25.5 µl of 0.7 mg/ml, 0.56 nmol	13.5 µl of 27 mg/ml stock solution[a], 1380 nmol, 2500 eq.
3	23 µl of 0.78 mg/ml, 0.56 nmol	16.0 µl of 27 mg/ml stock solution[a], 1640 nmol, 2900 eq.

$Pd(L2)_2(OAc)_2$ catalyst was added to each reaction (3.4 µl of 10 mM solution, 34 nmol, 60 eq., 0.8 mM concentration). The reaction mixture was briefly vortexed, then incubated at 37 °C with shaking (600 rpm). After 3 hours, DTT was added to a final concentration of 8 mM (from a 115 mM solution in H_2O), the mixture shaken for a further 40 minutes, then analysed directly by LC-MS.

Reaction 1



M_{calc}: 32250

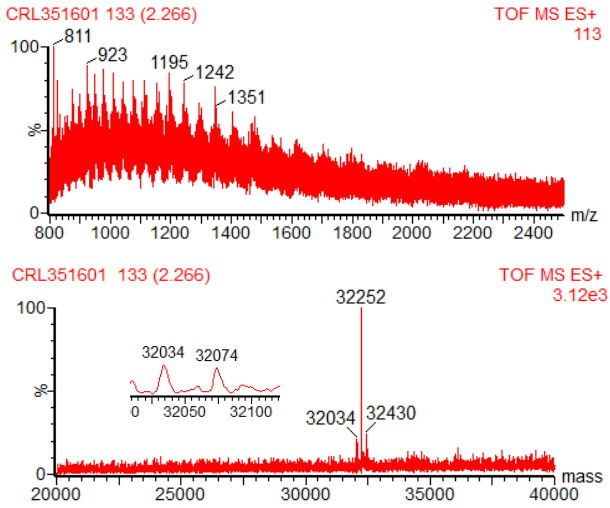
M_{obs}: 32252 (68 %)

32034 (de-iodination)

32075 (side-product, see pp. 172-3, **figure 3.8**)

32432 – gluconoylation, see note below

Reaction 2



M_{calc}: 32250

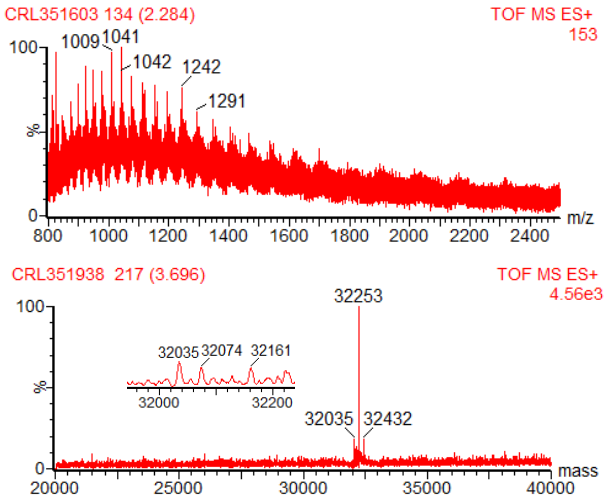
M_{obs}: 32252 (74 %)

32034 (de-iodination)

32074 (see see pp. 172-3, **figure 3.8**)

32430 – gluconoylation, see note below

Reaction 3



M_{calc}: 32250

M_{obs}: 32253 (73 %)

32034 (de-iodination)

32161 (GYG-Y195*p*IPhe)

32074 (see pp. 172-3, **figure 3.8**)

32432 – gluconoylation, see note below

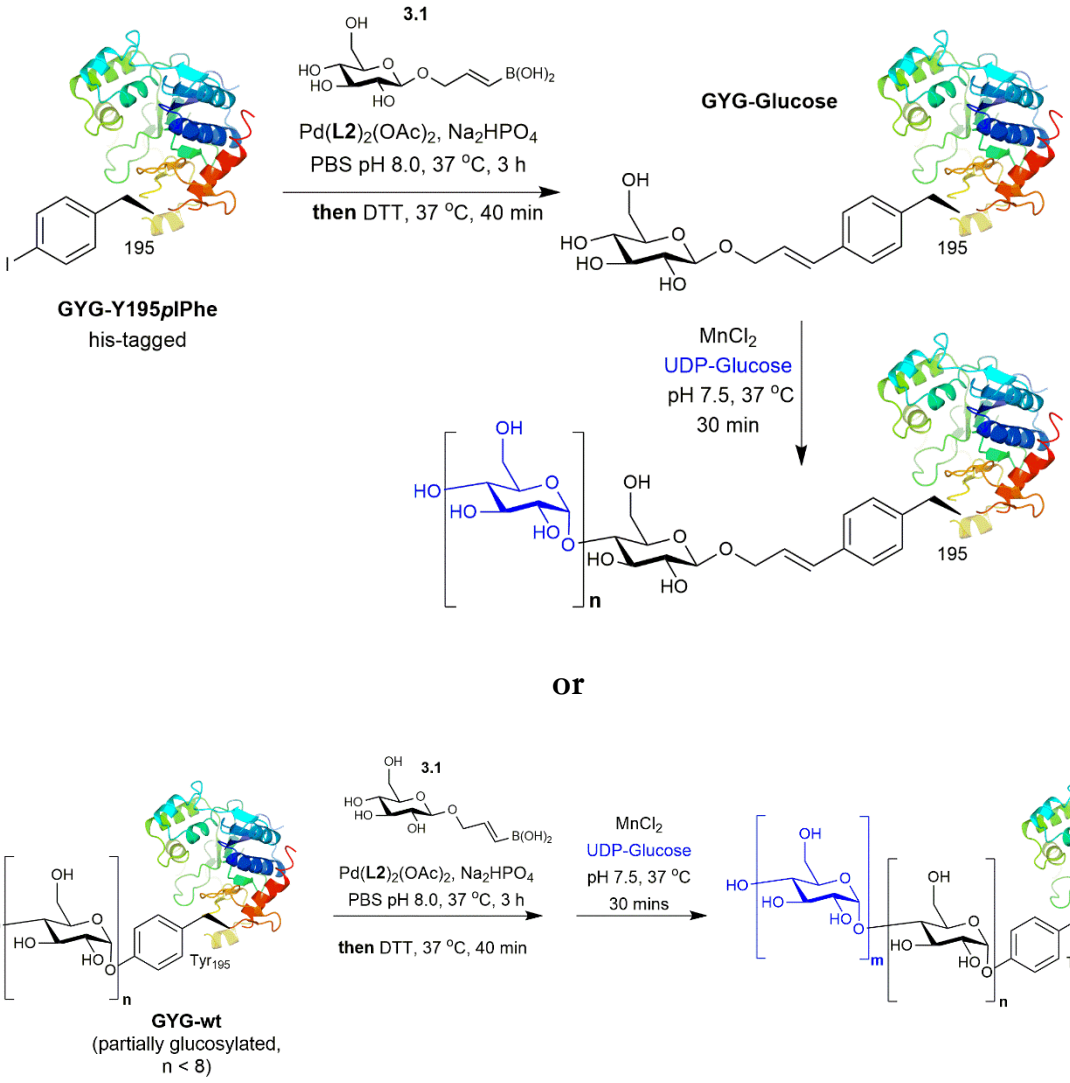
Note concerning the peaks at M_{obs} = 32430/2

As stated in footnote (i) in the main text (see pp. 172), the small peak at 32430/32432 in the reactions above arises from the fact that a small proportion of the GYG-Y195*p*IPhe starting material is N-terminally gluconoylated (+178 Da, see chapter 2, **figure 2.4**). This is a well-known artefact arising during bacterial expression of His-tagged proteins.[53] Thus, the peak at 32430/2 is, within error, a +178 Da adduct of the cross-coupled enzyme reflecting the fact that the small proportion of gluconoylated starting material reacts analogously to the majority, non-gluconoylated species.

This peak was only observed in certain samples (and was naturally never seen for non-His-tagged enzyme), reflecting variability in levels of gluconoylation present in the starting material (ca. 5-15 %), and presumably variation in MS signal/noise ratio (i.e. sensitivity). Indeed, with the exception of the samples above, it was not readily detectable above background noise.

It should be noted that reaction conversions refer to ratios of cross-coupled product with respect to side-products arising from the reaction; gluconoylation, being present in the starting material, is not a reaction side-product. **For consistency in comparisons, conversions are throughout calculated only from peaks arising from the majority, non-gluconoylated enzyme** (it should be emphasised, however, that its inclusion or omission does not affect conversions to any significant degree (+/- 1-2 %), and furthermore all samples under direct comparison use the same starting material so are affected identically). This decision was primarily taken since, whilst gluconoylated and non-gluconoylated enzymes should react identically, side-products for the former (e.g. de-iodination) are too low in abundance to detect, even if presumably present.

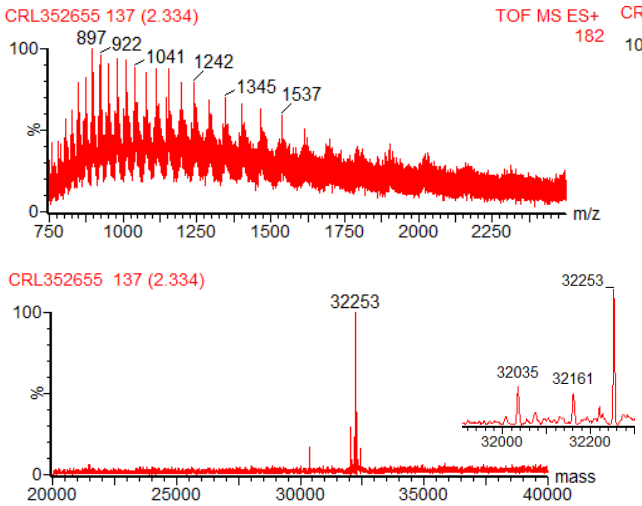
3.5.3.2 Enzymatic Activity of his-tagged GYG-Glucose: Initial Attempt (0.8 mM Pd)



A stock solution of glucose boronic acid **3.1** was prepared by dissolving boronic acid (27 mg/ml) and Na_2HPO_4 (15.7 mg/ml = 110 mM, 1.1 eq. w.r.t. boronic acid) in H_2O. To either his-tagged GYG-Y195*p*IPhe or his-tagged GYG-wt (64 µl of 0.7 mg/ml, 1.39 nmol) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was added 34 µl of this stock solution (3480 nmol, 2500 eq. boronic acid), followed by $Pd(L2)_2(OAc)_2$ catalyst (8.5 µl of 10 mM solution, 85 nmol, 60 eq., 0.8 mM concentration). The reaction mixture was briefly vortexed, then incubated at 37 °C with shaking (600 rpm). After 3 hours, DTT was added to a final concentration of 8 mM (from a 115 mM solution in H_2O), the mixture shaken for a further 40 minutes, then buffer exchanged into pH 7.5 25 mM Na-HEPES by PD SpinTrap G-25. 100 µl of the sample was further desalted by dialysis against the same buffer (10000 MWCO, 4 °C, 2 h then 2 h then overnight, 3 x 300 ml buffer), using a Slide-A-Lyzer™ MINI dialysis device. An aliquot was analysed by LC-MS.

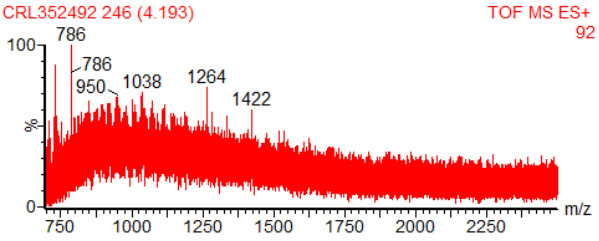
40 µl of resultant reaction mixture was incubated with $MnCl_2$ and UDP-Glucose disodium salt at final concentrations of 1 mM each (5 µl of 10 mM freshly prepared stock solution in H_2O was added). The reaction was incubated at 37 °C with shaking (500 rpm) for 30 minutes. After this time, Na-EDTA was added to a final concentration of 20 mM to quench glucosylation. Reactions were analysed by LC-MS to determine if any glucosylation had occurred (for GYG-wt, extent of glucosylation and proportions of glycoforms were determined according to the protocol outlined in chapter 2, section 2.8.6.1).

GYG-Glucose, Before Glucosylation Assay



M_{calc}: 32250
M_{obs}: 32253 (66 %)
32035 (de-iodination)
32161 (GYG-Y195*p*IPhe)

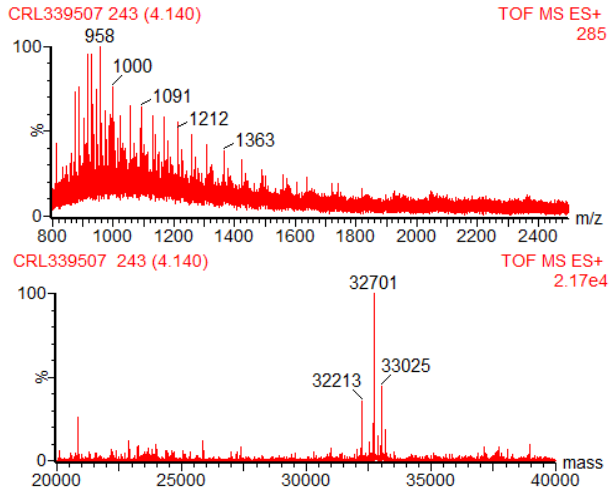
GYG-Glucose, After Glucosylation Assay



DID NOT DECONVOLUTE

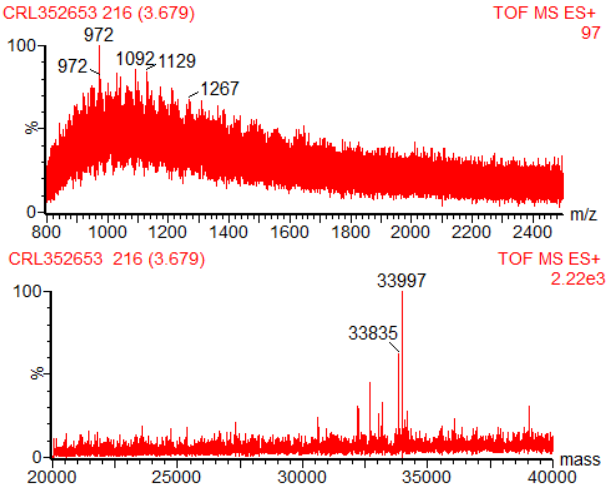
M_{calc} 32250 + 162n
M_{obs}: N/A

GYG-wt, before glucosylation assay



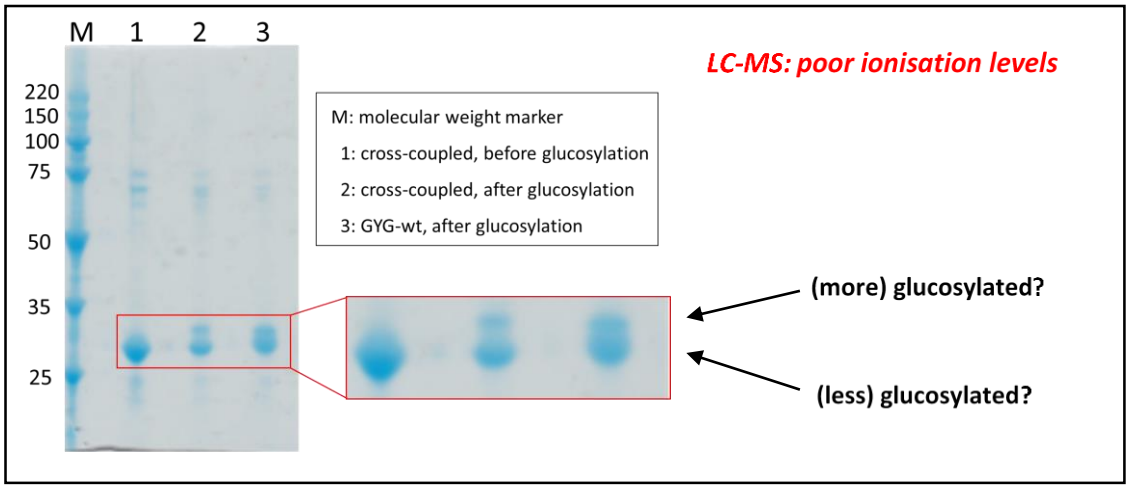
M_{calc}: 32048 + 162n
M_{obs}: 32051 + 162n
(n = glucoses attached = 1-7,
average = **4.07**)

GYG-wt, after glucosylation assay



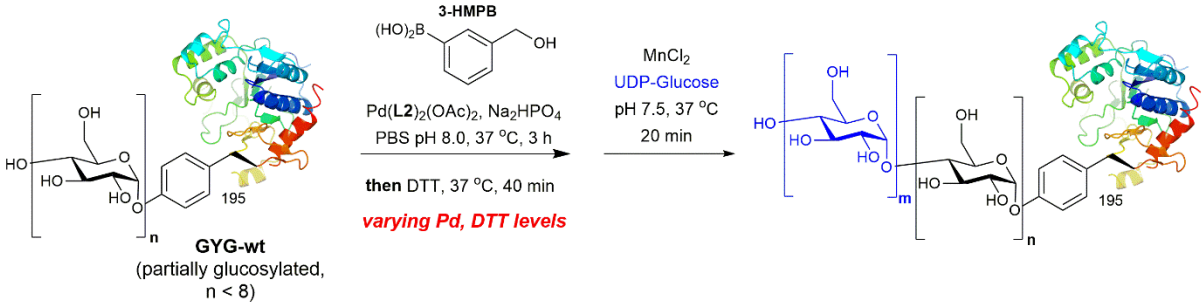
M_{calc}: 32048 + 162(n+m)
M_{obs}: 32053 + 162(n+m)
(n+m = glucoses attached = 1-12
average = **8.74**)

Given the lack of protein signal following LC-MS analysis of GYG-Glucose glucosylation, samples of GYG-Glucose (before glucosylation), GYG-Glucose (after glucosylation) and GYG-wt (after glucosylation) were additionally analysed by SDS-PAGE (for general procedure for SDS-PAGE analysis, please see appendix section). A 10 % bis-tris gel (Invitrogen) was used, with MOPS as the running buffer. The gel was stained with InstantBlue (Expedeon) for 1 h.



GYG-Glucose shows only 1 band (lane 1), however after glucosylation an additional higher Mw band is visible (lane 2), potentially corresponding to more highly glucosylated species; the same two bands were seen for glucosylated GYG-wt.

3.5.3.3 Optimisation Studies on GYG-wt: Pd/DTT Loading



Note that "reaction 1" and "reaction 2" below refer to protocols previously described in chapter 2 (see experimental section 2.8.6.5). However, these are reproduced here for clarity. All 3 reactons were carried out in parallel.

3 reactions were set up. A stock solution of boronic acid 3-HMPB was prepared by dissolving boronic acid (15.6 mg/ml) and Na_2HPO_4 (15.7 mg/ml = 110 mM, 1.1 eq. w.r.t. boronic acid) in H_2O. To his-tagged his-tagged GYG-wt (70 µl of 0.6 mg/ml, 1.30 nmol) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was added 19 µl of this stock solution (1950 nmol, 1500 eq. boronic acid). $Pd(L2)_2(OAc)_2$ catalyst was then added from a 10 mM stock solution in amounts indicated in the table below:

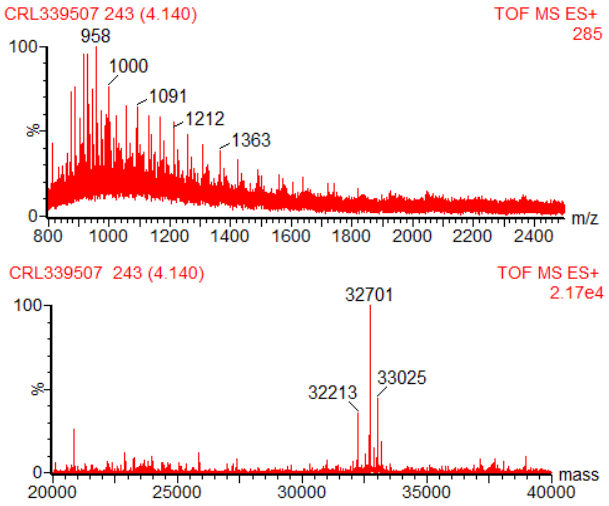
Reaction	Final concentration Pd / mM	Final concentration DTT / mM
1	**0.8** (= 7.9 µl 10 mM stock solution)	**8** (= 7.3 µl 115 mM stock solution)
2	**0.8** (= 7.9 µl 10 mM stock solution)	**11.5** (= 11 µl 115 mM stock solution)
3	**0.6** (= 5.7 µl 10 mM stock solution)	**8** (= 7.3 µl 115 mM stock solution)

The reaction mixtures were briefly vortexed, then incubated at 37 °C with shaking (600 rpm). After 3 hours, DTT was added to each reaction from a 115 mM solution in H_2O the mixture shaken for a further 40 minutes, then buffer exchanged into pH 7.5 25 mM Na-HEPES by PD SpinTrap G-25. 100 µl of the sample was further desalted by dialysis against the same buffer (10000 MWCO, 4 °C, 2 h then 2 h then overnight, 3 x 300 ml buffer), using a Slide-A-Lyzer™ MINI dialysis device.

40 µl of resultant reaction mixture was incubated with $MnCl_2$ and UDP-Glucose disodium salt at final concentrations of 1 mM each (5 µl of 10 mM freshly prepared stock solution in H_2O was added). The reaction was incubated at 37 °C with shaking (500 rpm) for 30 minutes. After this time, Na-EDTA was added to a final concentration of 20 mM to quench glucosylation. Reactions were analysed by LC-MS to determine extent of glucosylation and proportions of glycoforms present, according to the protocol described in chapter 2, section 2.8.6.1. Please consult **figure 3.10** for graphical representation.

0.8 mM Pd, 8 mM DTT

GYG-wt, before glucosylation assay

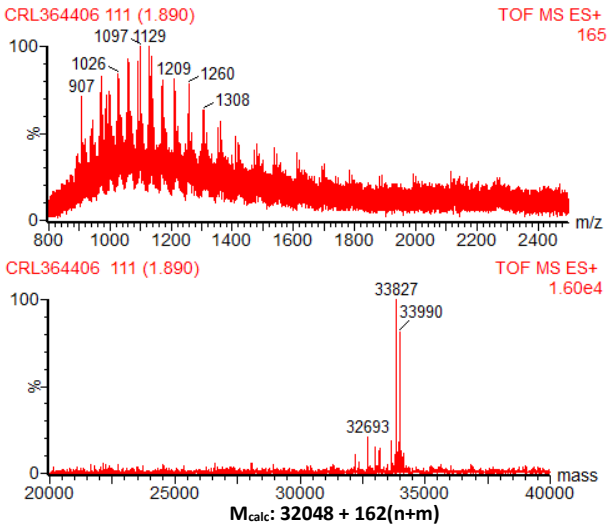


M_{calc}: 32048 + 162n

M_{obs}: 32051 + 162n

(n = glucoses attached = 0-7,
average = **4.07**)

GYG-wt, after glucosylation assay



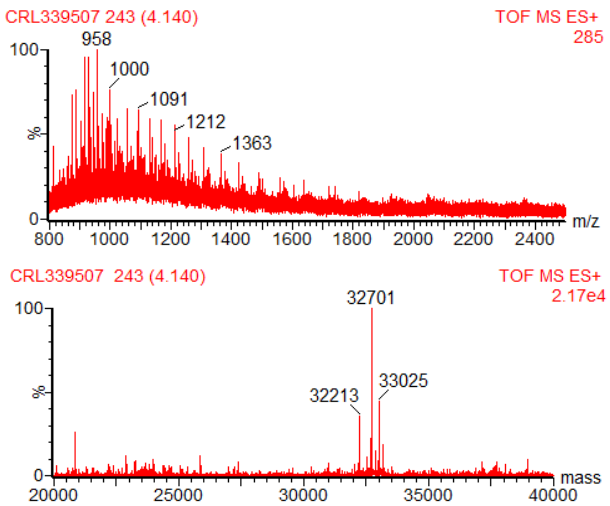
M_{calc}: 32048 + 162(n+m)

M_{obs}: 32045 + 162(n+m)

(n+m = glucoses attached = 1-13
average = **9.90**)

0.8 mM Pd, 11.5 mM DTT (raised DTT)

GYG-wt, before glucosylation assay

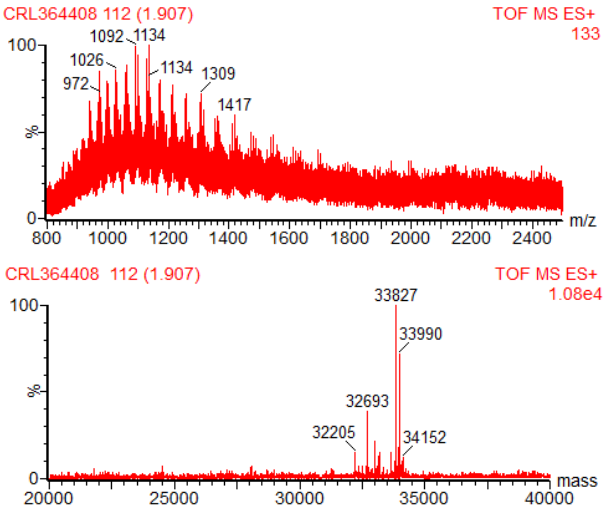


M_{calc}: 32048 + 162n

M_{obs}: 32051 + 162n

(n = glucoses attached = 0-7,
average = **4.07**)

GYG-wt, after glucosylation assay



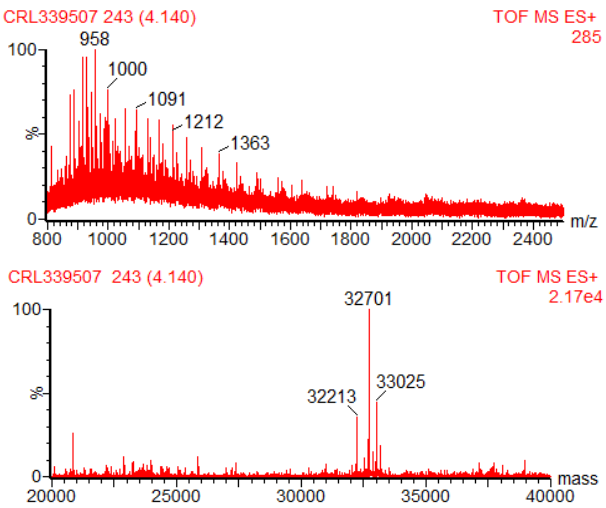
M_{calc}: 32048 + 162(n+m)

M_{obs}: 32045 + 162(n+m)

(n+m = glucoses attached = 1-13
average = **9.23**)

0.6 mM Pd, 8 mM DTT

GYG-wt, before glucosylation assay

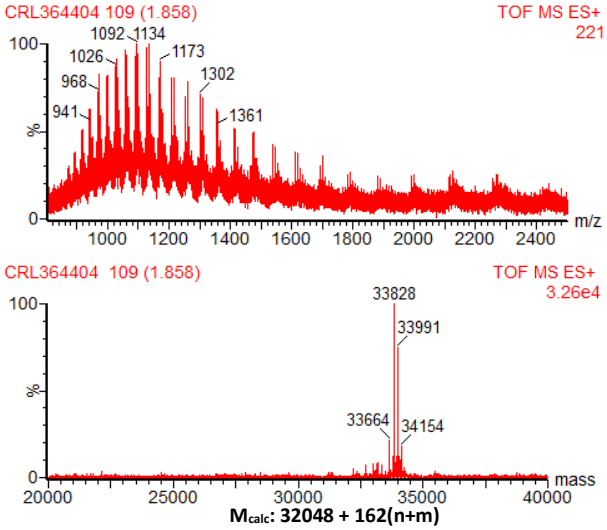


M_{calc}: 32048 + 162n

M_{obs}: 32051 + 162n

(n = glucoses attached = 0-7, average = **4.07**)

GYG-wt, after glucosylation assay



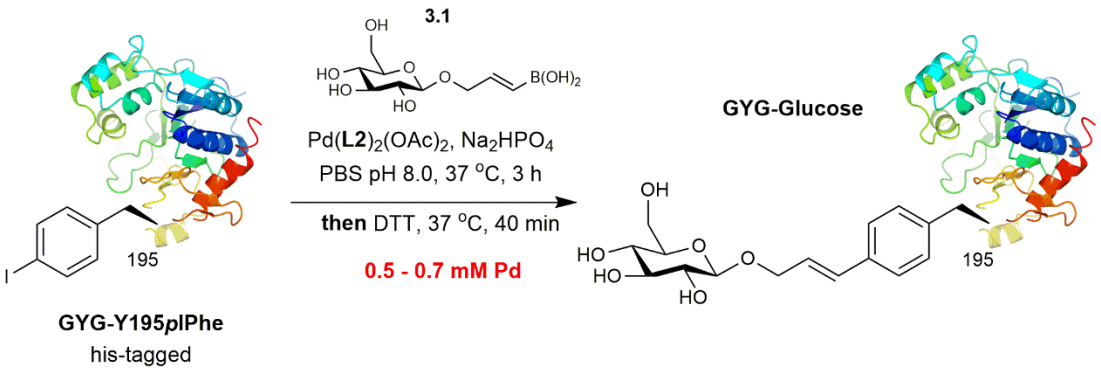
M_{calc}: 32048 + 162(n+m)

M_{obs}: 32046 + 162(n+m)

(n+m = glucoses attached = 1-13 average = **10.46**)

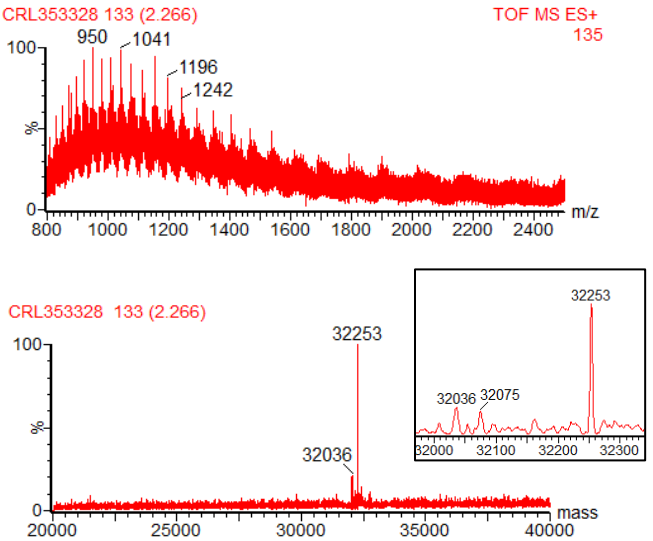
Final concentration Pd / mM	Final concentration DTT / mM	Average Glucoses Added (in 20 minutes)
0.8	8	5.83
0.8	11.5	5.16
0.6	8	6.39

3.5.3.4 Cross-Coupling of GYG-Y195*p*IPhe to Boronic Acid 3.1: Minimising Pd Concentration



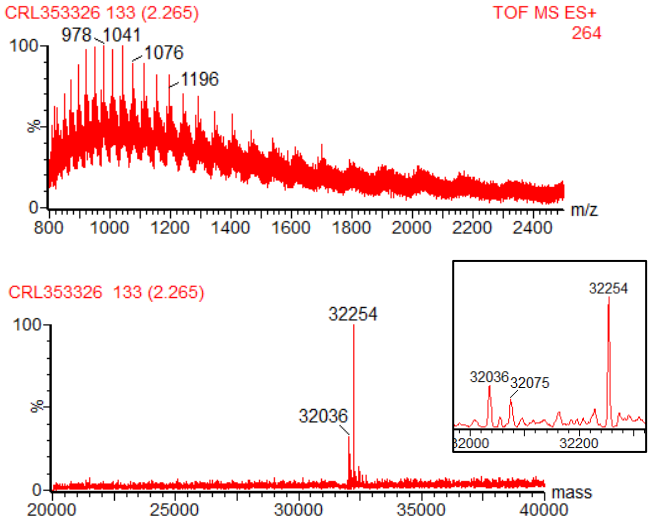
3 reactions were set up. A stock solution of glucose boronic acid **3.1** was prepared by dissolving boronic acid (27 mg/ml) and Na_2HPO_4 (15.7 mg/ml = 110 mM, 1.1 eq. w.r.t. boronic acid) in H_2O. To his-tagged GYG-Y195*p*IPhe (25.5 µl of 0.7 mg/ml, 0.56 nmol) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was added 13.5 µl of this stock solution (920 nmol, 1650 eq. boronic acid), followed by $Pd(L2)_2(OAc)_2$ catalyst (either 3.0, 2.6 or 2.2 µl of 10 mM solution, 0.7, 0.6 or 0.5 mM final concentration). The reaction mixtures were briefly vortexed, then incubated at 37 °C with shaking (600 rpm). After 3 hours, DTT was added to a final concentration of 8 mM (from a 115 mM solution in H_2O), the mixture shaken for a further 40 minutes, then analysed directly by LC-MS.

0.7 mM Pd



M_{calc}: 32250
M_{obs}: 32253 (72 %)
32036 (de-iodination)
32075 (see pp. 172-3, **figure 3.8**)

0.6 mM Pd



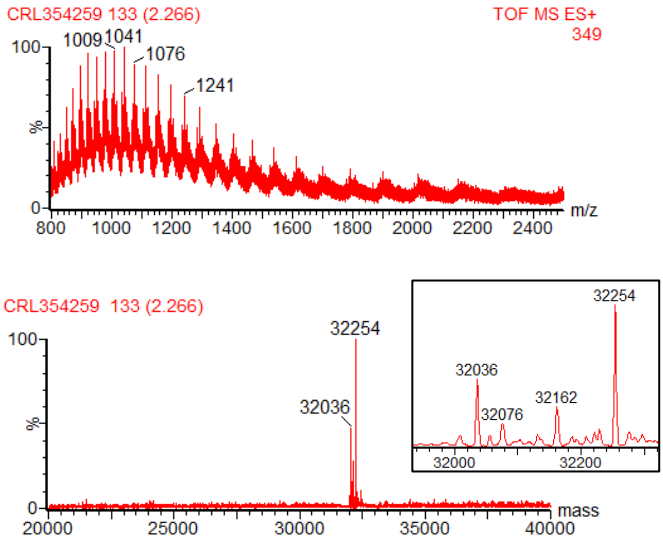
M_{calc}: 32250

M_{obs}: 32254 (65 %)

32036 (de-iodination)

32075 (see pp. 172-3, **figure 3.8**)

0.5 mM Pd



M_{calc}: 32250

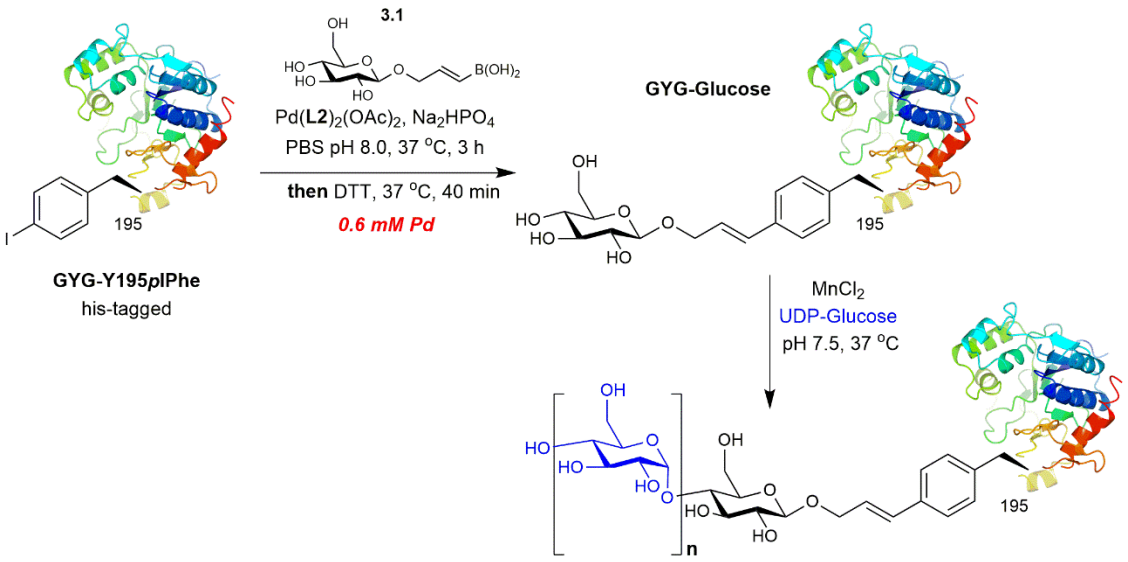
M_{obs}: 32254 (52 %)

32036 (de-iodination)

32162 (GYG-Y195*p*IPhe)

32076 (see pp. 172-3, **figure 3.8**)

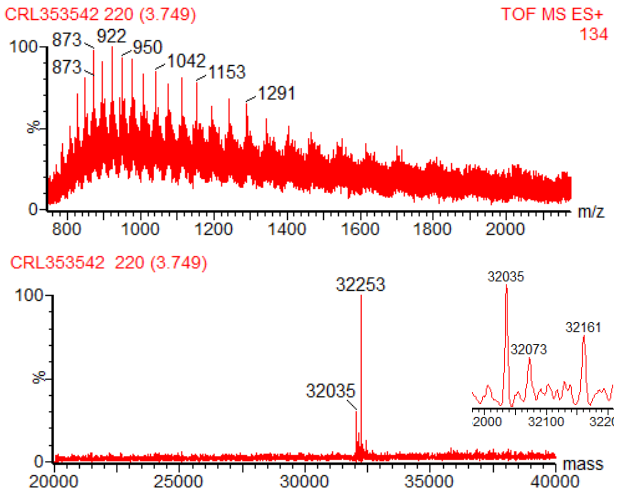
3.5.3.5 Enzymatic Activity of his-tagged GYG-Glucose: Optimised (0.6 mM Pd)



A stock solution of glucose boronic acid **3.1** was prepared by dissolving boronic acid (27 mg/ml) and Na_2HPO_4 (15.7 mg/ml = 110 mM, 1.1 eq. w.r.t. boronic acid) in H_2O. To his-tagged GYG-Y195*p*IPhe (64 µl of 0.7 mg/ml, 1.39 nmol) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was added 34 µl of this stock solution (3480 nmol, 2500 eq. boronic acid), followed by $Pd(L2)_2(OAc)_2$ catalyst (6.5 µl of 10 mM solution, 65 nmol, 0.6 mM concentration). The reaction mixture was briefly vortexed, then incubated at 37 °C with shaking (600 rpm). After 3 hours, DTT was added to a final concentration of 8 mM (from a 115 mM solution in H_2O), the mixture shaken for a further 40 minutes, then buffer exchanged into pH 7.5 25 mM Na-HEPES by PD SpinTrap G-25. 100 µl of the sample was further desalted by dialysis against the same buffer (10000 MWCO, 4 °C, 2 h then 2 h then overnight, 3 x 300 ml buffer), using a Slide-A-Lyzer™ MINI dialysis device. An aliquot was analysed by LC-MS.

40 µl of resultant reaction mixture was incubated with $MnCl_2$ and UDP-Glucose disodium salt at final concentrations of 1 mM each (5 µl of 10 mM freshly prepared stock solution in H_2O was added). The reaction was incubated at 37 °C with shaking (500 rpm) for 20 minutes. After this time, Na-EDTA was added to a final concentration of 20 mM to quench glucosylation. Reactions were analysed by LC-MS to determine if any glucosylation had occurred.

GYG-Glucose, Before Glucosylation Assay



M_{calc}: 32250

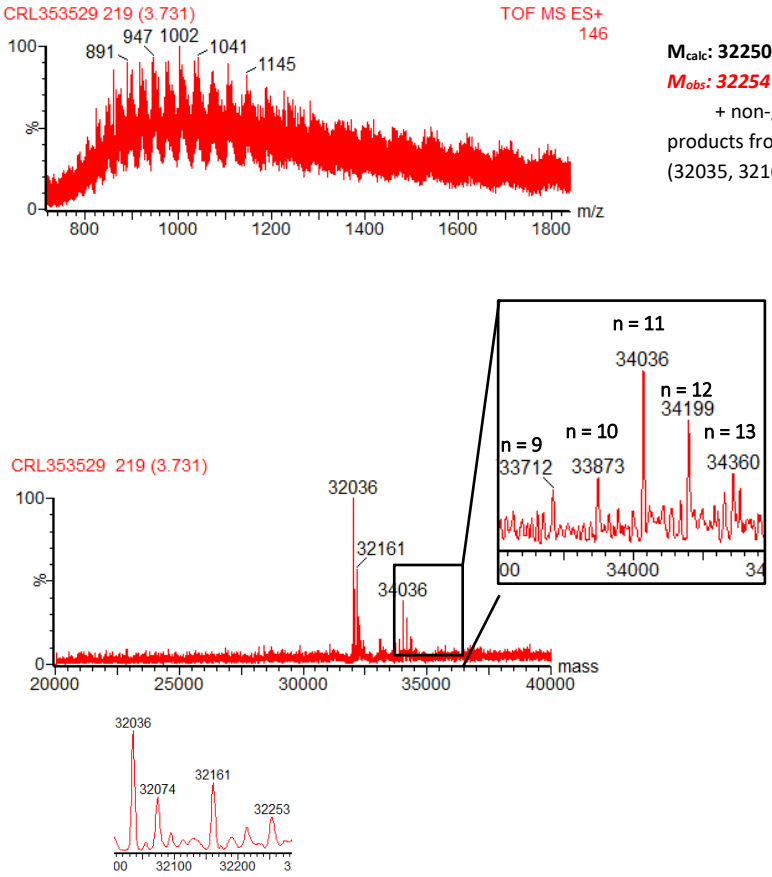
M_{obs}: 32253 (63 %)

32035 (de-iodination)

32161 (GYG-Y195*p*IPhe)

32073 (see pp. 172-3, **figure 3.8**)

GYG-Glucose, After Glucosylation Assay



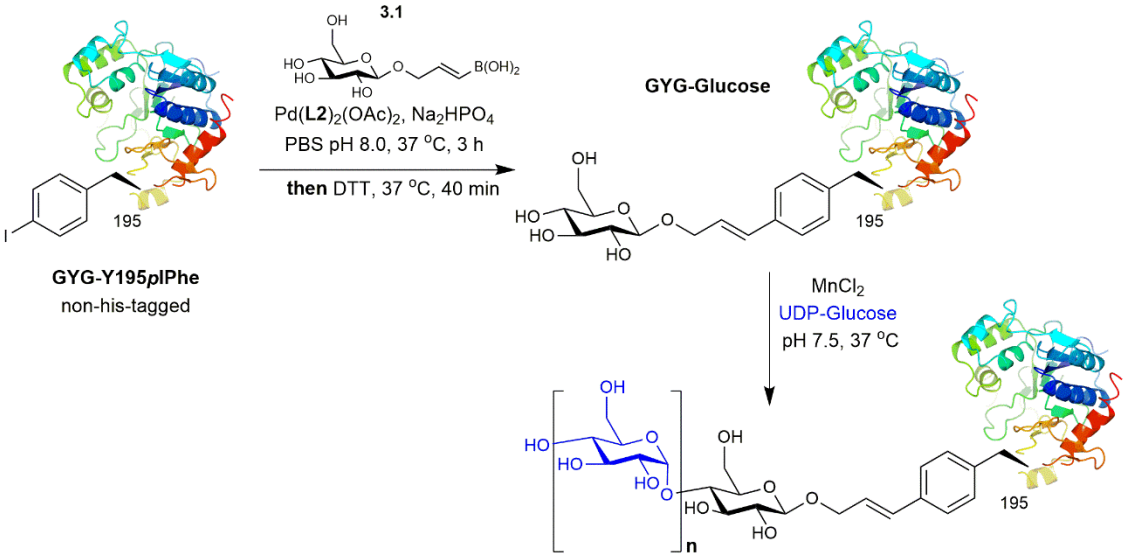
M_{calc}: 32250 + 162n

M_{obs}: 32254 + 162n (n = 0, 9-13)

+ non-glucosylatable side-products from cross-coupling (32035, 32161, 32073), see above

3.5.4 Generation and Enzymatic Activity of non-his-tagged GYG-Glucose

3.5.4.1 Initial Cross-Coupling of his-tag-cleaved GYG-Y195*p*IPhe to Glucose Boronic Acid 3.1, with Enzymatic Activity Test

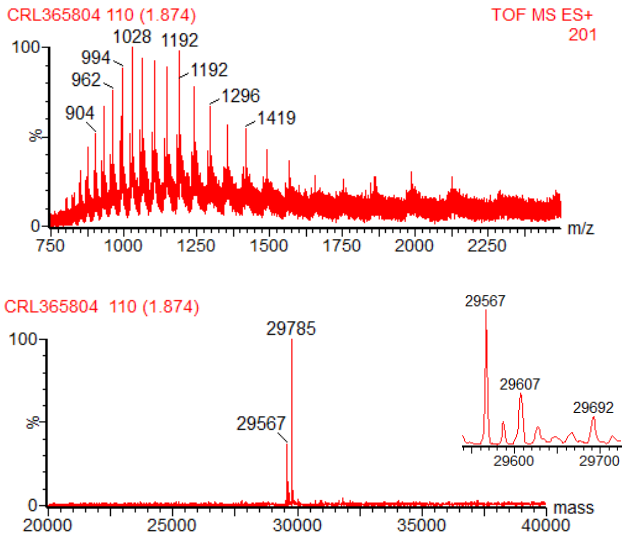


A stock solution of glucose boronic acid **3.1** was prepared by dissolving boronic acid (27 mg/ml) and Na_2HPO_4 (15.7 mg/ml = 110 mM, 1.1 eq. w.r.t. boronic acid) in H_2O. To **non-his-tagged** GYG-Y195*p*IPhe (63 µl of 0.65 mg/ml, 1.38 nmol) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was added 33 µl of this stock solution (3380 nmol, 2450 eq. boronic acid), followed by $Pd(L2)_2(OAc)_2$ catalyst (6.2 µl of 10 mM solution, 62 nmol, 45 eq., 0.6 mM concentration). The reaction mixture was briefly vortexed, then incubated at 37 °C with shaking (600 rpm). After 3 hours, DTT was added to a final concentration of 8 mM (from a 115 mM solution in H_2O), the mixture shaken for a further 40 minutes, then buffer exchanged into pH 7.5 25 mM Na-HEPES by PD SpinTrap G-25. 100 µl of the sample was further desalted by dialysis against the same buffer (10000 MWCO, 4 °C, 2 h then 2 h then overnight, 3 x 300 ml buffer), using a Slide-A-Lyzer™ MINI dialysis device. An aliquot was analysed by LC-MS.

40 µl of resultant reaction mixture was incubated with $MnCl_2$ and UDP-Glucose disodium salt at final concentrations of 1 mM each (5 µl of 10 mM freshly prepared stock solution in H_2O was added). The reaction was incubated at 37 °C with shaking (500 rpm) for 10 minutes (see note). After this time, an aliquot was removed and quenched by the addition of Na-EDTA to a final concentration of 20 mM, before being analysed by LC-MS.

Note: the glucosylation time was not extended further since an aliquot taken at this time already revealed unambiguous significant glucosylation activity.

GYG-Glucose, Before Glucosylation Assay



M_{calc}: 29784

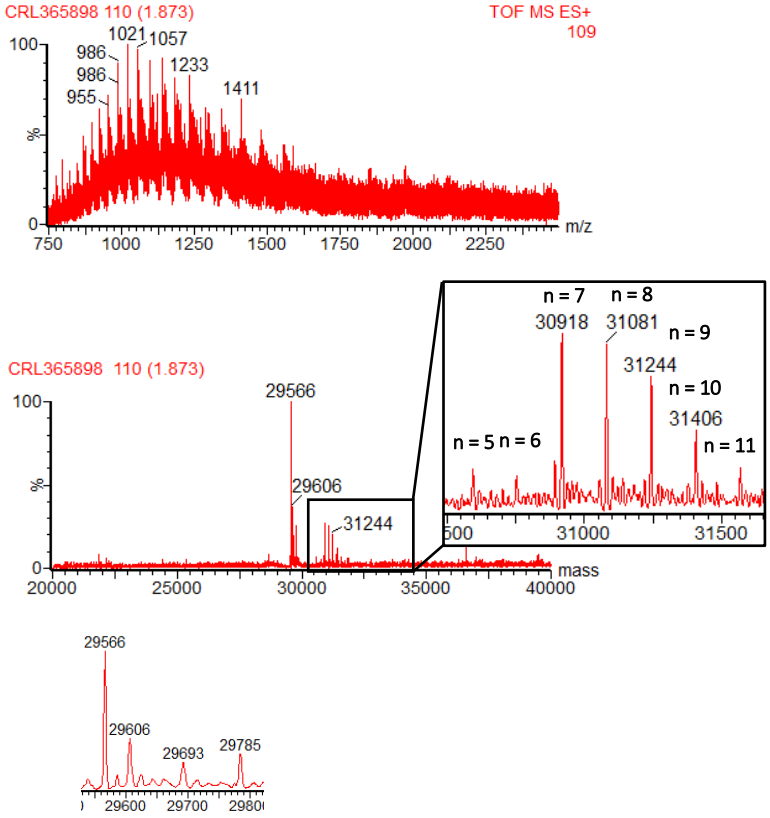
M_{obs}: 29785 (63 %)

29567 (de-iodination)

29692 (GYG-Y195*p*IPhe)

29607 (see pp. 172-3, **figure 3.8**)

GYG-Glucose, After Glucosylation Assay

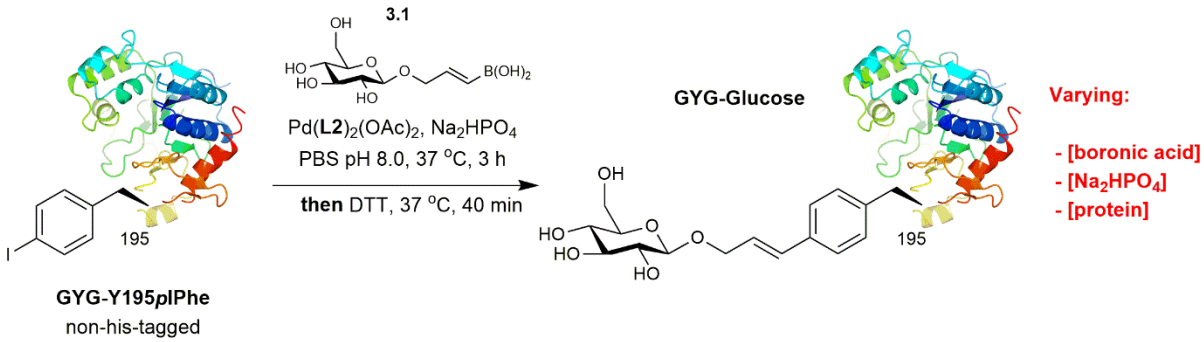


M_{calc}: 29784 + 162n

M_{obs}: 29784 + 162n (n = 0, 5-11)

+ inactive side-products from cross-coupling (29566, 29606, 29693), see above.

3.5.4.2 Optimisations of Cross-Coupling to non-his-tagged GYG-Y195*p*IPhe in table 3.3



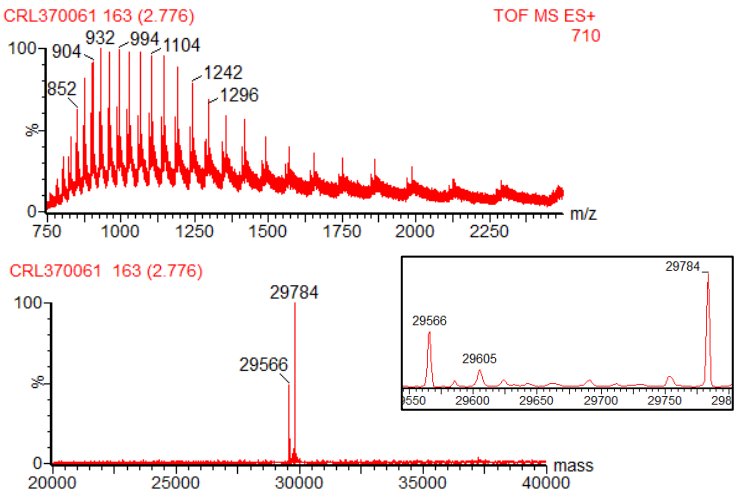
Experiments in this section represent samples treated in parallel, and thus under as identical conditions as possible except for the desired alterations, in order to minimise any extraneous sources of variation between samples and permit accurate evaluation of the desired parameters. Those in section 3.5.4.3 represent separate experiments carried out at different times, and thus should not be directly compared to those in this section.

a) Protein Concentration (Table 3.3, entries 1, 2)

2 reactions were set up. A stock solution of glucose boronic acid **3.1** was prepared by dissolving boronic acid (27 mg/ml) and Na_2HPO_4 (**15.7 mg/ml = 110 mM**, 1.1 eq. w.r.t. boronic acid) in H_2O. To non-his-tagged GYG-Y195*p*IPhe (*either* 30 µl of 0.65 mg/ml, 0.66 nmol *or* 30 µl of 0.80 mg/ml, 0.81 nmol) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was added 16 µl of this stock solution (1640 nmol boronic acid), followed by $Pd(L2)_2(OAc)_2$ catalyst (3.0 µl of 10 mM solution, 30 nmol, 0.6 mM concentration). Final protein concentrations were thus either 0.4 mg/ml or 0.5 mg/ml. The reaction mixtures were briefly vortexed, then incubated at 37 °C with shaking (600 rpm). After 3 hours, DTT was added to a final concentration of 7.5 mM[iii] (from a 115 mM solution in H_2O), the mixtures shaken for a further 40 minutes, then analysed by LC-MS.

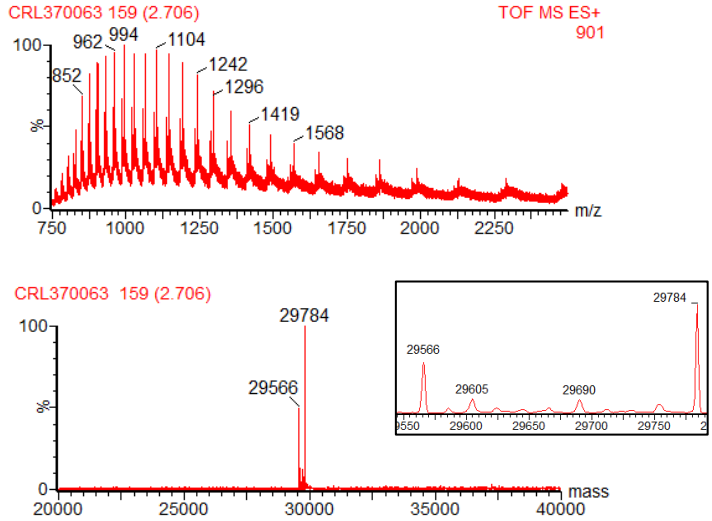
[iii] In earlier experiments, a concentration of 8 mM was utilised. However, it became gradually apparent that 7.5 mM DTT was sufficient for Pd scavenging and, based on results outlined in **figure 3.10**, this further lowering of DTT concentration would likely be of benefit for enzymatic activity.

0.65 mg/ml: 61 % conversion (table 3.3, entry 1)



M_{calc}: 29784
M_{obs}: 29784 (61 %)
29566 (de-iodination)
29605 (see pp. 172-3, **figure 3.8**)

0.8 mg/ml: 57 % conversion (table 3.3, entry 2)



M_{calc}: 29784
M_{obs}: 29784 (57 %)
29566 (de-iodination)
29690 (GYG-Y195*p*IPhe)
29605 (see pp. 172-3, **figure 3.8**)

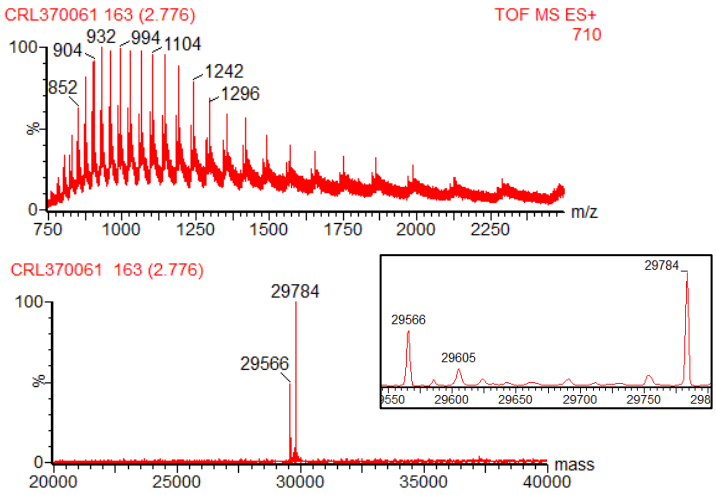
b) [Boronic Acid] (table 3.3, entries 1 and 5)

2 reactions were set up. A stock solution of glucose boronic acid **3.1** was prepared by dissolving boronic acid (27 mg/ml) and Na_2HPO_4 (15.7 mg/ml = 110 mM, 1.1 eq. w.r.t. boronic acid) in H_2O. To non-his-tagged GYG-Y195*p*IPhe (0.66 nmol) was added stock solution as indicated in the table below:

Reaction	GYG-Y195*p*IPhe	Boronic Acid Stock Solution
1	30 µl of 0.65 mg/ml, 0.66 nmol	16 µl, 1640 nmol, 2500 eq.
2	23 µl of 0.85 mg/ml, 0.66 nmol	22 µl, 2250 nmol, 3400 eq.

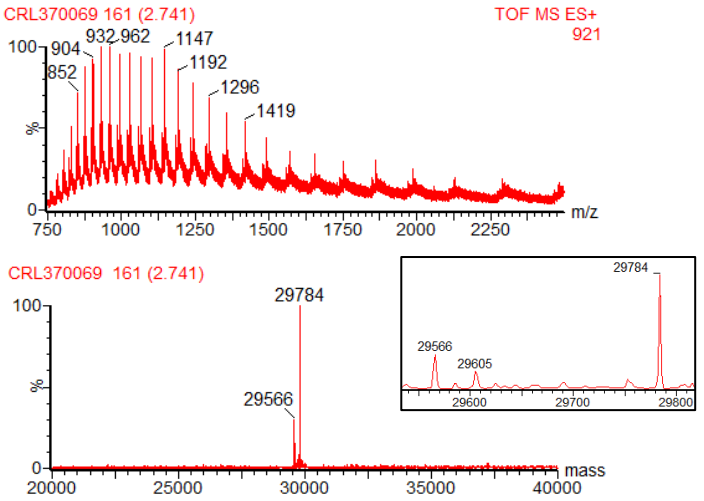
To each reaction was then added $Pd(L2)_2(OAc)_2$ catalyst (3.0 µl of 10 mM solution, 30 nmol, 0.6 mM concentration). The reaction mixtures were briefly vortexed, then incubated at 37 °C with shaking (600 rpm). After 3 hours, DTT was added to a final concentration of 7.5 mM (from a 115 mM solution in H_2O), the mixtures shaken for a further 40 minutes, then analysed by LC-MS (see overleaf).

2500 eq.: 61 % conversion (table 3.3, entry 1)



M_{calc}: 29784
M_{obs}: 29784 (61 %)
29566 (de-iodination)
29605 (see pp. 172-3, **figure 3.8**)

3400 eq.: 69 % conversion (table 3.3, entry 5)



M_{calc}: 29784
M_{obs}: 29784 (69 %)
29566 (de-iodination)
29605 (see pp. 172-3, **figure 3.8**)

c) [Na_2HPO_4] (table 3.3, entries 1, 3 and 4)

2 stock solutions of glucose boronic acid **3.1** were prepared:

Stock solution 1: boronic acid (27 mg/ml) and Na_2HPO_4 (15.7 mg/ml = 110 mM, 1.1 eq. w.r.t. boronic acid) in H_2O.

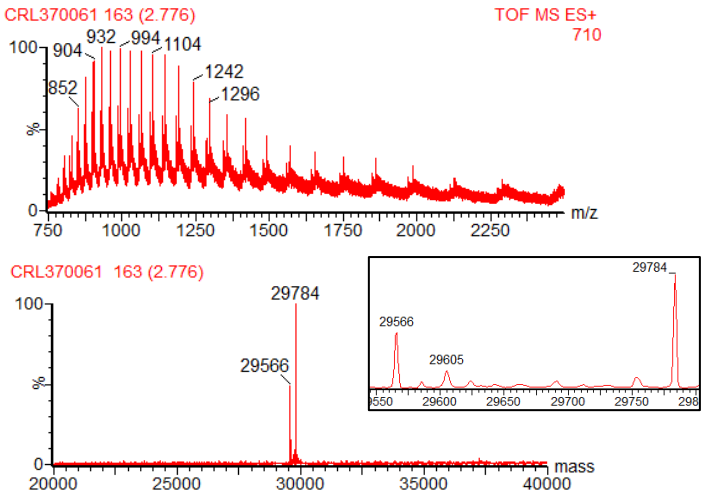
Stock solution 2: boronic acid (27 mg/ml) *only* in H_2O.

3 reactions were then set up by combining components in the amounts indicated in the table below:

Reaction	GYG-Y195*p*IPhe	Boronic Acid Stock Solution	Additional Na_2HPO_4
1	30 µl of 0.65 mg/ml, 0.56 nmol	16 µl of **stock solution 1**, 1640 nmol, 2500 eq.	None
2	30 µl of 0.65 mg/ml, 0.56 nmol	16 µl of **stock solution 2**, 1640 nmol, 2500 eq.	None
3	22 µl of 0.89 mg/ml, 0.56 nmol	16 µl of **stock solution 1**, 1640 nmol, 2500 eq.	8 µl of **15.7 mg/ml = 110 mM** solution in H_2O.

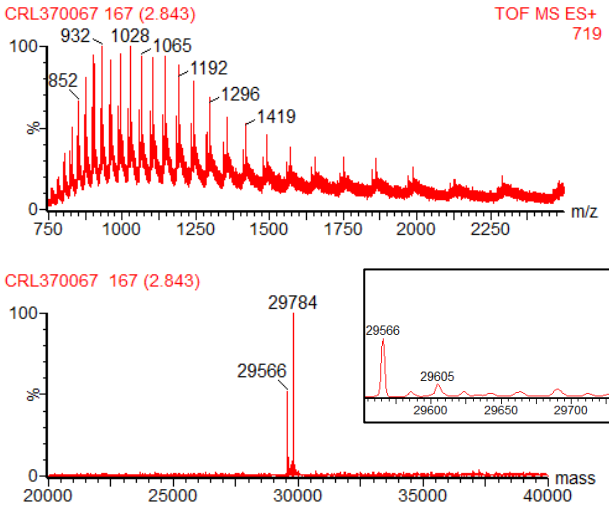
To each reaction was then added $Pd(L2)_2(OAc)_2$ catalyst (3.0 µl of 10 mM solution, 30 nmol, 0.6 mM concentration). The reaction mixtures were briefly vortexed, then incubated at 37 °C with shaking (600 rpm). After 3 hours, DTT was added to a final concentration of 7.5 mM (from a 115 mM solution in H_2O), the mixtures shaken for a further 40 minutes, then analysed by LC-MS.

Reference (reaction 1): 61 % conversion (table 3.3, entry 1)



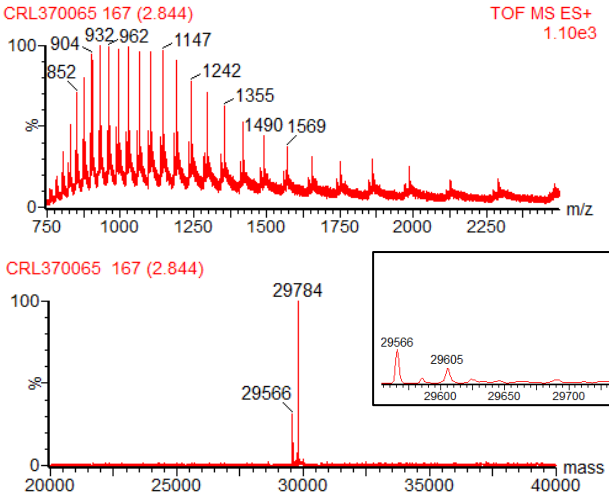
M_{calc}: 29784
M_{obs}: 29784 (61 %)
29566 (de-iodination)
29605 (see pp. 172-3, **figure 3.8**)

Extra Na_2HPO_4 (reaction 3): 61 % conversion (table 3.3, entry 3)



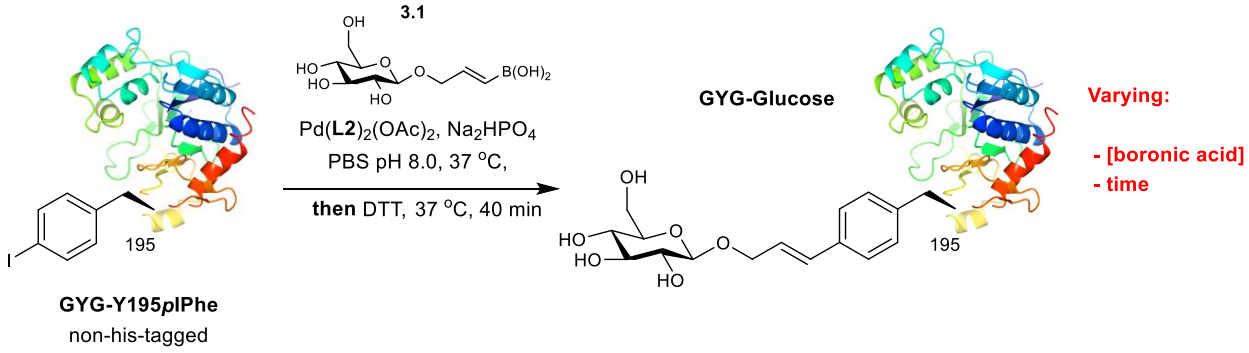
M_{calc}: 29784
M_{obs}: 29784 (61 %)
29566 (de-iodination)
29605 (see pp. 172-3, **figure 3.8**)

No Na_2HPO_4 (reaction 2) 69 % conversion (table 3.3, entry 4)



M_{calc}: 29784
M_{obs}: 29784 (69 %)
29566 (de-iodination)
29605 (see pp. 172-3, **figure 3.8**)

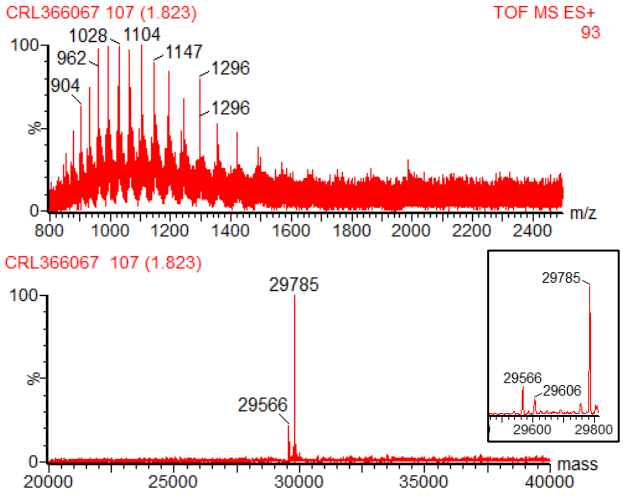
3.5.4.3 Further Optimisation Studies



a) Reaction time

A stock solution of glucose boronic acid **3.1** was prepared by dissolving boronic acid (27 mg/ml) and Na_2HPO_4 (15.7 mg/ml = 110 mM, 1.1 eq. w.r.t. boronic acid) in H_2O. To non-his-tagged GYG-Y195*p*IPhe (15 µl of 0.65 mg/ml, 0.33 nmol) was added 8 µl of this stock solution (820 nmol boronic acid, 2500 eq.), followed by $Pd(L2)_2(OAc)_2$ catalyst (1.5 µl of 10 mM solution, 15 nmol, 0.6 mM concentration). The reaction mixtures were briefly vortexed, then incubated at 37 °C with shaking (600 rpm). After either 3 or 4.5 hours, DTT was added to an aliquot of the reaction mixture to final concentration of 7.5 mM (from a 115 mM solution in H_2O). After shaking this aliquot for 40 minutes, it was analysed by LC-MS.

3 hours: 75 % conversion



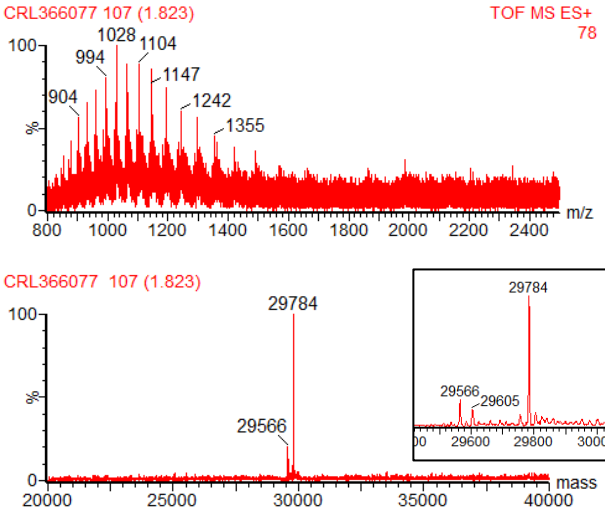
M_{calc}: **29784**

M_{obs}: 29785 (75 %)

29566 (de-iodination)

29606 (see pp. 172-3, **figure 3.8**)

4.5 hours: 76 % conversion



M_{calc}: 29784
M_{obs}: 29784 (75 %)
29566 (de-iodination)
29605 (see pp. 172-3, **figure 3.8**)

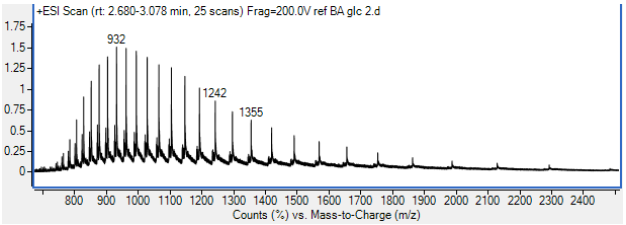
b) [Boronic Acid] – further increase

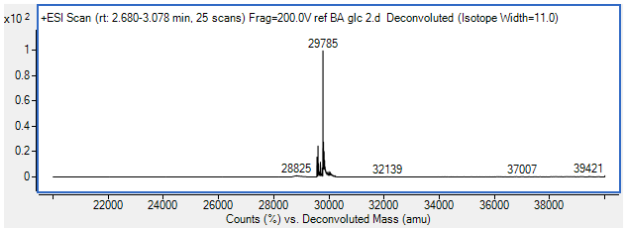
2 reactions were set up. 2 stock solutions of glucose boronic acid **3.1** was prepared by dissolving boronic acid (*either* 27 mg/ml *or* 36 mg/ml) in H_2O. To non-his-tagged GYG-Y195*p*IPhe (0.33 nmol) was added stock solution as indicated in the table below:

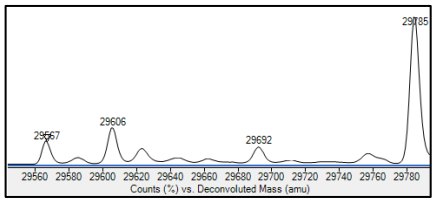
Reaction	GYG-Y195*p*IPhe	Boronic Acid Stock Solution
1	15 µl of 0.85 mg/ml, 0.43 nmol	14.2 µl of 27 mg/ml stock, 1450 nmol, 3400 eq.
2	15 µl of 0.85 mg/ml, 0.33 nmol	14.2 µl of 36 mg/ml stock, 1940 nmol, 4500 eq.

To each reaction was then added $Pd(L2)_2(OAc)_2$ catalyst (1.9 µl of 10 mM solution, 19 nmol, 0.6 mM concentration). The reaction mixtures were briefly vortexed, then incubated at 37 °C with shaking (600 rpm). After 3 hours, DTT was added to a final concentration of 8 mM (from a 115 mM solution in H_2O), the mixtures shaken for a further 40 minutes, then analysed by LC-MS.

3400 eq.: 66 % conversion







M_{calc}: 29784

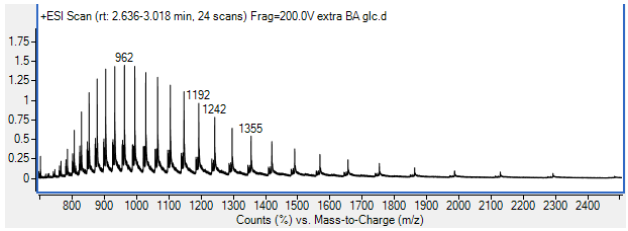
M_{obs}: 29785 (66 %)

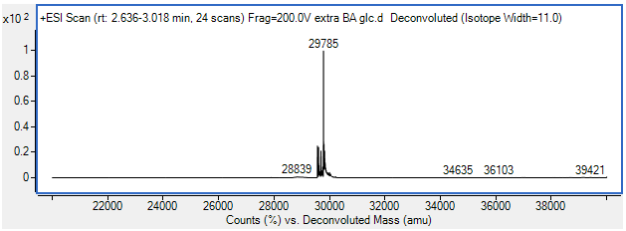
29567 (de-iodination)

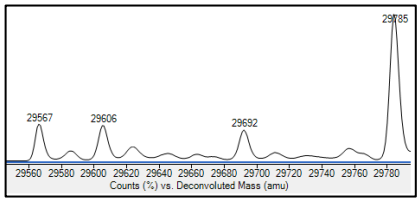
29692 (GYG-Y195*p*IPhe)

29606 (see pp. 172-3, **figure 3.8**)

4500 eq.: 59 % conversion







M_{calc}: 29784

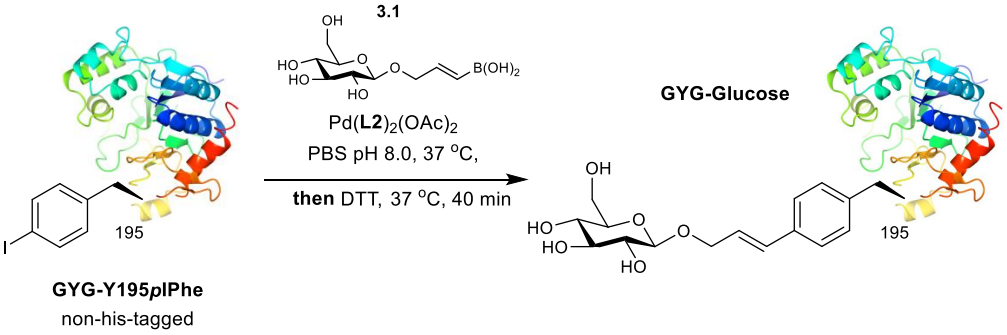
M_{obs}: 29785 (59 %)

29567 (de-iodination)

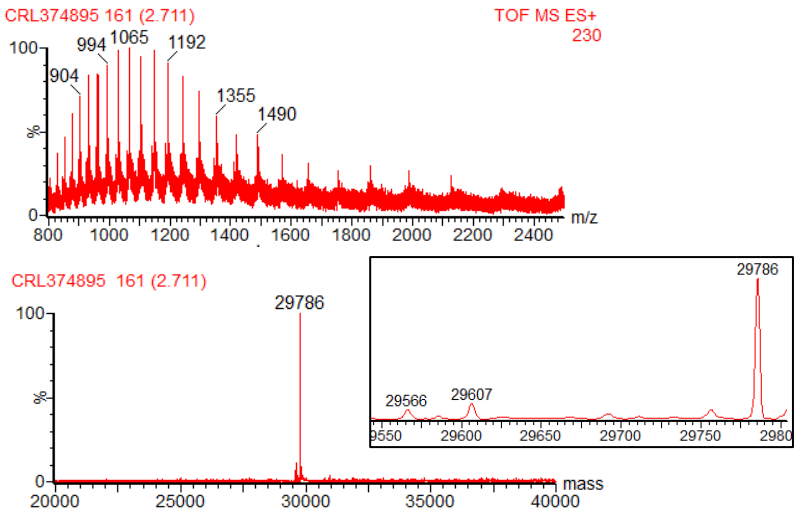
29692 (GYG-Y195*p*IPhe)

29606 (see pp. 172-3, **figure 3.8**)

3.5.4.4 Optimal Cross-coupling Procedure

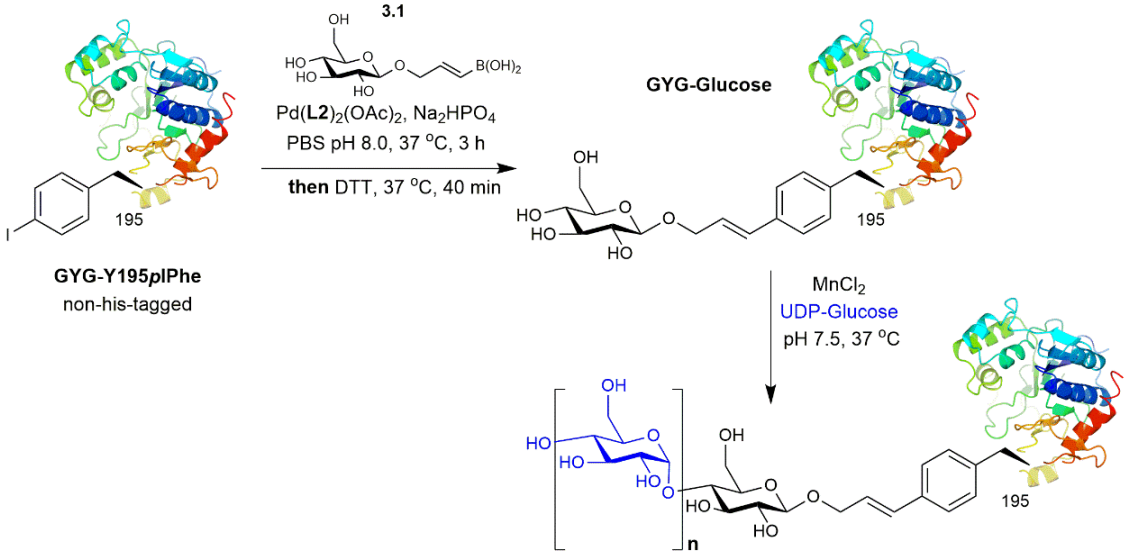


A stock solution of glucose boronic acid **3.1** was prepared by dissolving boronic acid (27 mg/ml) in H_2O. To non-his-tagged GYG-Y195*p*IPhe (50 µl of 0.85 mg/ml, 1.43 nmol) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was added 48 µl of this stock solution (4900 nmol, 3400 eq. boronic acid), followed by $Pd(L2)_2(OAc)_2$ catalyst (6.3 µl of 10 mM solution, 63 nmol, 44 eq., 0.6 mM concentration). The reaction mixture was briefly vortexed, then incubated at 37 °C with shaking (600 rpm). After 3 hours, DTT was added to a final concentration of 7.5 mM (from a 115 mM solution in H_2O), the mixture shaken for a further 40 minutes, then analysed by LC-MS. A conversion of **85 %** was achieved.



M_{calc}: 29784
M_{obs}: 29785 (85 %)
29566 (de-iodination)
29607 (see pp. 172-3, **figure 3.8**)

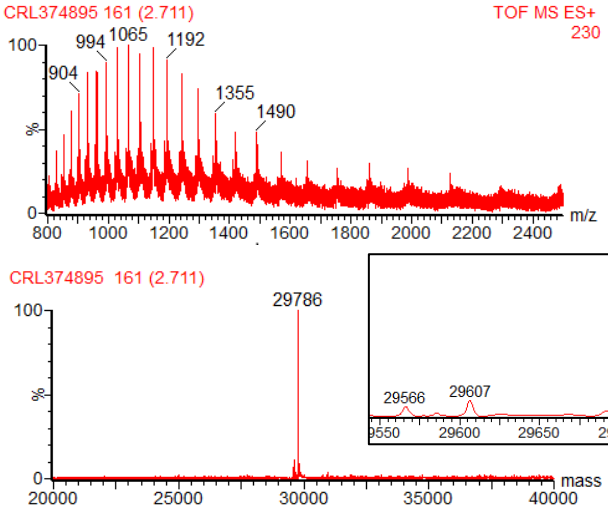
3.5.4.5 Enzymatic Activity of GYG-Glucose: Fully Optimised Procedure



A stock solution of glucose boronic acid **3.1** was prepared by dissolving boronic acid (27 mg/ml) in H_2O. To non-his-tagged GYG-Y195*p*IPhe (50 µl of 0.85 mg/ml, 1.43 nmol) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was added 48 µl of this stock solution (4900 nmol, 3400 eq. boronic acid), followed by $Pd(L2)_2(OAc)_2$ catalyst (6.3 µl of 10 mM solution, 63 nmol, 44 eq., 0.6 mM concentration). The reaction mixture was briefly vortexed, then incubated at 37 °C with shaking (600 rpm). After 3 hours, DTT was added to a final concentration of 7.5 mM (from a 115 mM solution in H_2O), the mixture shaken for a further 40 minutes, then buffer exchanged into pH 7.5 25 mM Na-HEPES by PD SpinTrap G-25. 100 µl of the sample was further desalted by dialysis against the same buffer (10000 MWCO, 4 °C, 2 h then 2 h then overnight, 3 x 300 ml buffer), using a Slide-A-Lyzer™ MINI dialysis device. An aliquot was analysed by LC-MS.

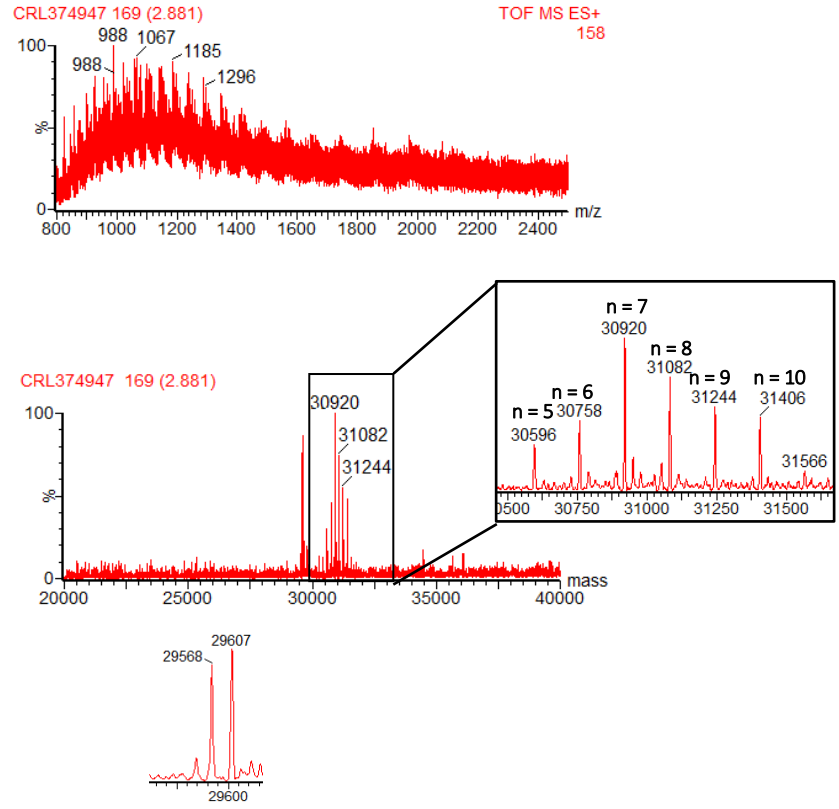
32 µl of resultant reaction mixture was incubated with $MnCl_2$ and UDP-Glucose disodium salt at final concentrations of 1 mM each (4 µl of 10 mM freshly prepared stock solution in H_2O was added). The reaction was incubated at 37 °C with shaking (500 rpm) for 10 minutes. After this time, an aliquot was removed and quenched by the addition of Na-EDTA to a final concentration of 20 mM, before being analysed by LC-MS.

GYG-Glucose, Before Glucosylation Assay (after cross-coupling):



M_{calc}: 29784
M_{obs}: 29785 (85 %)
29566 (de-iodination)
29607 (see pp. 172-3, **figure 3.8**)

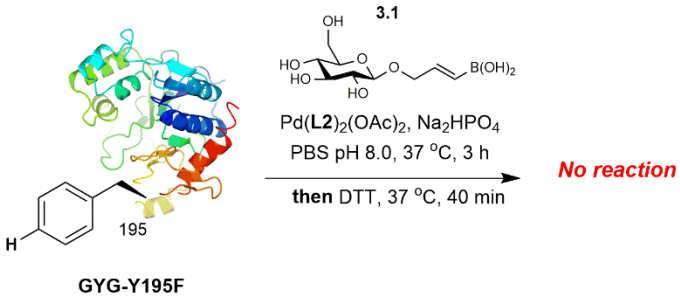
GYG-Glucose, After Glucosylation Assay



M_{calc}: 29784 + 162n
M_{obs}: 29786 + 162n (n = 5-10)
29568 (de-iodination)
29607 (see pp. 172-3, **figure 3.8**)

3.5.5 Further Investigations of Cross-coupling

3.5.5.1 Negative Control Studies: GYG-Y195F



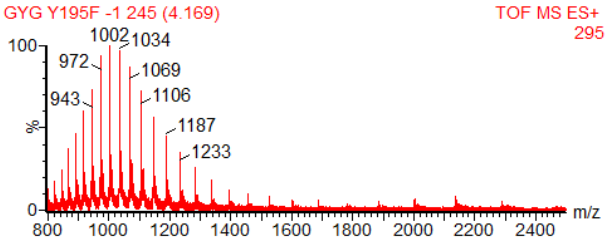
Explanation of Experiment:

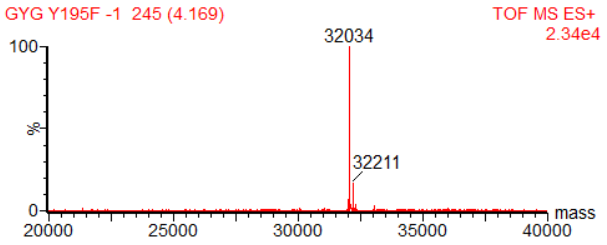
The lack of reaction of GYG-Y195F with **3.1** implies not only site-specificity of product formation but also site-specificity of side-product formation to the pIPhe residue. Whilst it is acknowledged that the optimised conditions employed give low (but reproducible) levels of side-products for analogous coupling to GYG-Y195pIPhe, they were always readily detectable; if formed from another amino acid site, we would expect to observe their formation also for GYG-Y195F.

Protocol:

A stock solution of glucose boronic acid **3.1** was prepared by dissolving boronic acid (27 mg/ml) in H_2O. To his-tagged GYG-Y195F (18 µl of 0.9 mg/ml, 0.51 nmol) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was added 17.3 µl of this stock solution (1770 nmol, 3450 eq. boronic acid), followed by $Pd(L2)_2(OAc)_2$ catalyst (2.3 µl of 10 mM solution, 23 nmol, 0.6 mM concentration). The reaction mixture was briefly vortexed, then incubated at 37 °C with shaking (600 rpm). After 3 hours, DTT was added to a final concentration of 8 mM (from a 115 mM solution in H_2O) and the mixture shaken for a further 40 minutes. The mixture was desalted (desalting buffer = phosphate buffer as used above) using a Micro Bio-Spin™ 6 column (BioRad), then analysed by LC-MS. No reaction was observed, indicating specificity of cross-coupling – and indeed side-reactions – to the *p*IPhe195 residue.

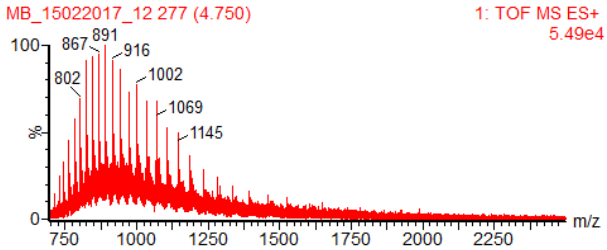
Before reaction:

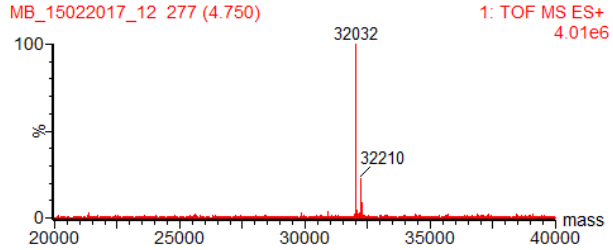




M_{calc}: 32032
M_{obs}: 32034
32211 (gluconoylated)

After reaction:





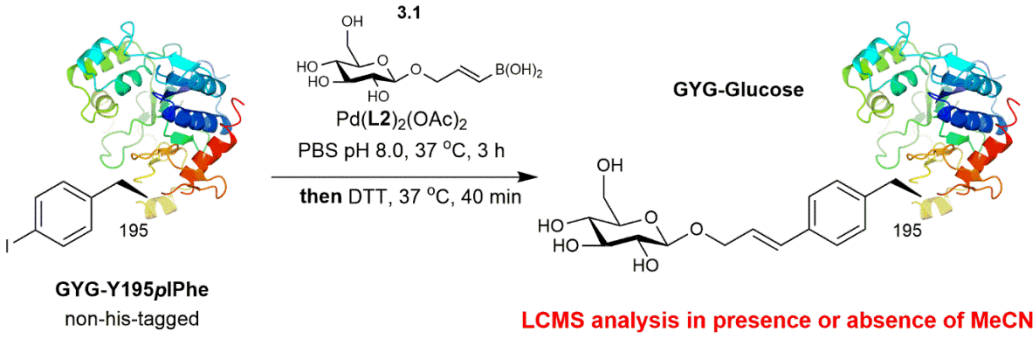
M_{calc} (GYG-Y195F): 32032
M_{obs}: 32032
32210 (gluconoylation)

The peak at 32210 corresponds to a mass difference of +178 Da relative to GYG-Y195F. This is attributable to N-terminal gluconoylation, and this peak is also clearly observed in the starting material as shown; this is observed in LC-MS spectra of sufficient signal/ noise ratio (typically at lower Pd loadings). As discussed previously, this is a known consequence of expression of his-tagged enzymes.[53]

Note: whilst levels of side-products in analogous cross-coupling to GYG-Y195*p*IPhe are low, they are readily observable, thus their lack of presence here implies that they result from the 195 site. Furthermore, in chapter 6, the same lack of reaction of GYG-Y195F will be shown in the context of a carbohydrate boronic acid whose reaction to GYG-Y195*p*IPhe naturally affords greater quantities of side-products.

3.5.5.2 Effect of LC-MS Mobile Phase on Cross-coupling side products

In the text it was noted that, whilst an alternative structure for the side-product with M_{calc} = 32074 (his-tagged) or 29606 (non-his-tagged) was simply an MeCN-adduct of de-iodinated enzyme formed upon LC-MS analysis, this possibility was ruled out by the fact that this side-product was still observed in the absence of MeCN in the LC mobile phase. This is detailed below.

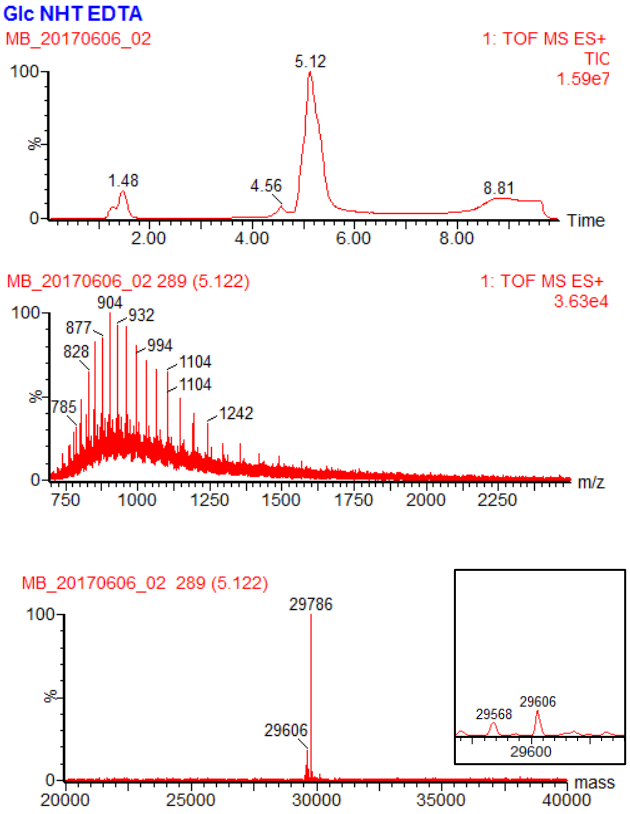


A stock solution of boronic acid **3.1** (27 mg/ml) was prepared in H_2O. Non-his-tagged GYG-Y195*p*IPhe (25.5 µl of 0.85 mg/ml, 0.73 nmol) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was combined with boronic acid stock solution (24.5 µl, 2500 nmol, 3400 eq.). $Pd(L2)_2(OAc)_2$ catalyst was added from 10 mM stock solution (3.15 µl, 32 nmol, 0.6 mM final concentration). The reaction mixtures were briefly vortexed, then incubated at 37 °C with shaking (600 rpm). After 3 hours, DTT was added to a final concentration of 7.5 mM (from a 115 mM solution in H_2O), the mixture shaken for a further 40 minutes. Samples were then dialysed overnight against pH 7.5 25 mM Na-HEPES containing 1 mM Na-EDTA (this was to confirm that no species was simply a metal adduct, which indeed was not the case).

Samples were analysed by LC-MS using one of two different gradient methods. All other parameters were maintained as identical (see appendix: for reference, the parameters used correspond to "method 4" in the section "Protein LC-MS").

Method 1: Solvent A = 0.1 % Formic Acid in H_2O, Solvent B = 0.1 % Formic Acid in MeCN.

Time (min)	% Solvent A	% Solvent B
0.0	95	5
1.0	95	5
7.0	5	95
8.0	5	95
8.1	95	5
10.0	95	5



M_{calc}: 29784

M_{obs}: 29786

29568 (de-iodination)

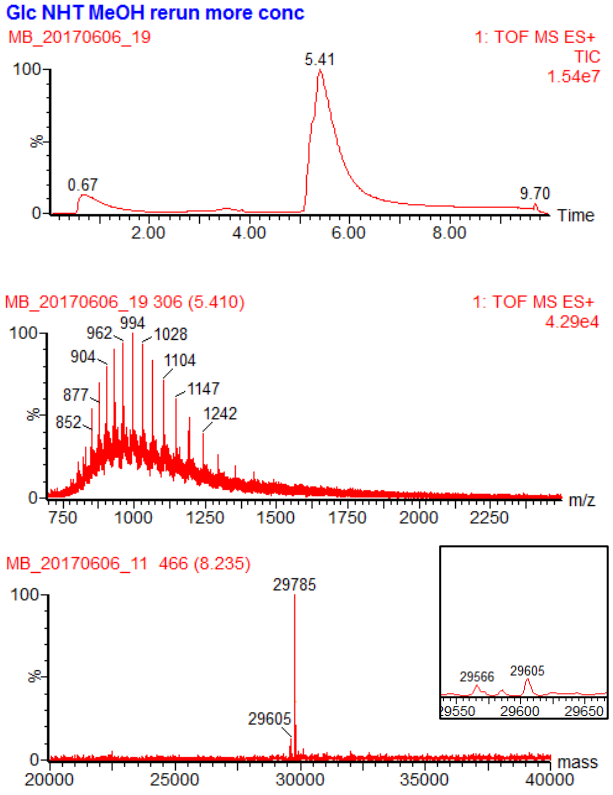
29606 (**3.11**or "MeCN adduct"?

see pp. **172-3, figure 3.8**)

Method 2: Solvent A = 0.1 % Formic Acid in H_2O, Solvent B = 0.1 % Formic Acid in MeOH.

Time (min)	% Solvent A	% Solvent B
0.0	95	5
1.0	95	5
4.0	5	95
8.0	5	95
8.1	95	5
10.0	95	5

Note: the sharper gradient was rendered necessary to ensure that the protein eluted fully from the column, given the reduced elution power of MeOH, see below.



M_{calc}: 29784
M_{obs}: 29785
29566 (de-iodination)
29605 (not MeCN adduct, likely 3.11 – see pp. **172-3, figure 3.8**)

3.6 References

1. Wright, T. H. *et al.* Posttranslational mutagenesis: A chemical strategy for exploring protein side-chain diversity. *Science.* **354,** aag1465 (2016).

2. Huang, R. *et al.* Site-specific introduction of an acetyl-lysine mimic into peptides and proteins by cysteine alkylation. *J. Am. Chem. Soc.* **132,** 9986–9987 (2010).

3. Simon, M. D. *et al.* The Site-Specific Installation of Methyl-Lysine Analogs into Recombinant Histones. *Cell* **128,** 1003–1012 (2007).

4. Chooi, K. P. *et al.* Synthetic Phosphorylation of p38a Recapitulates Protein Kinase Activity Synthetic Phosphorylation of p38α Recapitulates Protein Kinase Activity. *JACS* **136,** 1698–1701 (2014).

5. Hemantha, H. P. *et al.* Nonenzymatic polyubiquitination of expressed proteins. *J. Am. Chem. Soc.* **136,** 2665–2673 (2014).

6. Glickman, M. H. & Ciechanover, A. The Ubiquitin-Proteasome Proteolytic Pathway: Destruction for the Sake of Construction. *Physiol. Rev.* **82,** 373–428 (2002).

7. Lewis, Y. E. *et al.* O-GlcNAcylation of α-synuclein at serine 87 reduces aggregation without affecting membrane binding. *ACS Chem. Biol.* **12,** 1020–1027 (2017).

8. Wang, P. *et al.* Erythropoietin Derived by Chemical Synthesis. *Science.* **342,** 1357-1360 (2013).

9. van Kasteren, S. I. *et al.* Expanding the diversity of chemical protein modification allows post-translational mimicry. *Nature* **446,** 1105–1109 (2007).

10. Raj, R., Lercher, L., Mohammed, S. & Davis, B. G. Synthetic Nucleosomes Reveal that GlcNAcylation Modulates Direct Interaction with the FACT Complex. *Angew. Chemie. Int. Ed.* **55,** 8918–8922 (2016).

11. Lercher, L. *et al.* Generation of a synthetic GlcNAcylated nucleosome reveals regulation of stability by H2A-Thr101 GlcNAcylation. *Nat. Commun.* **6,** 7978 (2015).

12. Spicer, C. D. & Davis, B. G. Rewriting the bacterial glycocalyx via Suzuki-Miyaura cross-coupling. *Chem. Commun. (Camb).* **49,** 2747–2749 (2013).

13. Parthasarathy, R., Rabuka, D., Bertozzi, C. R. & Groves, J. T. Molecular orientation of membrane-anchored mucin glycoprotein mimics. *J. Phys. Chem. B* **111,** 12133–12135 (2007).

14. Rabuka, D. *et al.* Hierarchical assembly of model cell surfaces: Synthesis of mucin mimetic polymers and their display on supported bilayers. *J. Am. Chem. Soc.* **129,** 5462–5471 (2007).

15. Paszek, M. J. *et al.* The cancer glycocalyx mechanically primes integrin-mediated growth and survival. *Nature* **511,** 319–325 (2014).

16. Kufe, D. W. Mucins in cancer: function, prognosis and therapy. *Nat. Rev. Cancer* **9,** 874–885 (2009).

17. Roach, P. J., Skurat, A. V & Harris, R. A. in *Handbook of Physiology* (eds. Jefferson, L. S. & Cherrington, A. D.) 609–647 (Oxford University Press, 2000).

18. Alonso, M. D., Lomako, J., Lomako, W. M. & Whelan, W. J. Tyrosine-194 of glycogenin undergoes autocatalytic glucosylation but is not essential for catalytic function and activity. *FEBS Lett.* **342,** 38–42 (1994).

19. Kawada, H., Iwamoto, M., Utsugi, M., Miyano, M. & Nakada, M. Synthetic studies on the taxane skeleton: Construction of eight-membered carbocyclic rings by the intramolecular B-alkyl Suzuki-Miyaura cross-coupling reaction. *Org. Lett.* **6,** 4491–4494 (2004).

20. Molander, G. A. & Yun, C. S. Cross-coupling reactions of primary alkylboronic acids with aryl triflates and aryl halides. *Tetrahedron* **58,** 1465–1470 (2002).

21. Zou, G., Reddy, Y. K. & Falck, J. R. Ag(I)-promoted Suzuki-Miyaura cross-couplings of n-alkylboronic acids. *Tetrahedron Lett.* **42,** 7213–7215 (2001).

22. Qiao, Y., Molina, H., Pandey, A., Zhang, J. & Cole, P. Chemical Rescue of a Mutant Enzyme in Living Cells. *Science (80-.).* **311,** 1293–1298 (2006).

23. Pratt, M. R., Schwartz, E. C. & Muir, T. W. Small-molecule-mediated rescue of protein function by an inducible proteolytic shunt. *Proc. Natl. Acad. Sci. U.S.A.* **104,** 11209–11214 (2007).

24. Arbely, E., Torres-Kolbus, J., Deiters, A. & Chin, J. W. Photocontrol of tyrosine phosphorylation in mammalian cells via genetic encoding of photocaged tyrosine. *J. Am. Chem. Soc.* **134,** 11912–11915 (2012).

25. Gautier, A., Deiters, A. & Chin, J. W. Light-activated kinases enable temporal dissection of signaling networks in living cells. *J. Am. Chem. Soc.* **133,** 2124–2127 (2011).

26. Wu, N., Deiters, A., Cropp, T. A., King, D. & Schultz, P. G. A genetically encoded photocaged amino acid. *J. Am. Chem. Soc.* **126,** 14306–14307 (2004).

27. Lemke, E. A., Summerer, D., Geierstanger, B. H., Brittain, S. M. & Schultz, P. G. Control of protein phosphorylation with a genetically encoded photocaged amino acid. *Nat. Chem. Biol.* **3,** 769–772 (2007).

28. Deiters, A., Groff, D., Ryu, Y., Xie, J. & Schultz, P. G. A genetically encoded photocaged tyrosine. *Angew. Chemie - Int. Ed.* **45,** 2728–2731 (2006).

29. Weiss, J. T. *et al.* Extracellular palladium-catalysed dealkylation of 5-fluoro-1-propargyl-uracil as a bioorthogonally activated prodrug approach. *Nat. Commun.* **5,** 3277 (2014).

30. Yusop, R. M., Unciti-Broceta, A., Johansson, E. M., Sanchez-Martin, R. M. & Bradley, M. Palladium-mediated intracellular chemistry. *Nat Chem* **3,** 239–243 (2011).

31. Li, J. *et al.* Palladium-triggered deprotection chemistry for protein activation in living cells. *Nat. Chem.* **6,** 352–361 (2014).

32. Li, J., Jia, S. & Chen, P. R. Diels-Alder reaction–triggered bioorthogonal protein decaging in living cells. *Nat. Chem. Biol.* **10,** 1003–1005 (2014).

33. Wang, Y. D., Kimball, G., Prashad, A. S. & Wang, Y. Zr-Mediated hydroboration: Stereoselective synthesis of vinyl boronic esters. *Tetrahedron Lett.* **46,** 8777–8780 (2005).

34. Ahmadi, Z. & McIndoe, J. S. A mechanistic investigation of hydrodehalogenation using ESI-MS. *Chem. Commun.* **49,** 11488 (2013).

35. Lennox, A. J. J. & Lloyd-Jones, G. C. The Slow-Release Strategy in Suzuki – Miyaura Coupling. *Isr. J. Chem.* **50,** 664–674 (2010).

36. Williams, R. J., McGill, N. W., White, J. M. & Williams, S. J. Neighboring Group Participation in Glycosylation Reactions by 2,6-Disubstituted 2- O -Benzoyl groups: A Mechanistic Investigation. *J. Carbohydr. Chem.* **29,** 236–263 (2010).

37. Fukuda, K. *et al.* A recyclable heavy fluorous tag carrying an allyl alcohol pendant group: Design and evaluation toward applications in synthetic carbohydrate chemistry. *Carbohydr. Res.* **407,** 122–130 (2015).

38. Nakayama, K., Uoto, K., Higashi, K., Soga, T. & Kusama, T. A Useful Method for Deprotection of the Protective Allyl Group at the Anomeric Oxygen of Carbohydrate Moieties Using Tetrakis(triphenylphosphine)palladium. *Chem. Pharm. Bull. (Tokyo).* **40,** 1718–1720 (1992).

39. Mandai, T., Matsumoto, T., Kawada, M. & Tsuji, J. Stereocontrolled Formation of Cis and Trans Ring Junctions in Hydrindane and Decalin Systems by Palladium-Catalyzed Regioselective and Stereospecific Hydrogenolysis of Allylic Formates. *J. Org. Chemrg Chem* **57,** 1326–1327 (1992).

40. Keinan, E. & Roth, Z. Communications: Regioselectivity in Organo-Transition-Metal Chemistry. A New Indicator Substrate for Classification of Nucleophiles. *J. Org. Chem.* **48,** 1769–1772 (1983).

41. Guibé, F. Allylic protecting groups and their use in a complex environment Part II: Allylic protecting groups and their removal through catalytic palladium π-allyl methodology. *Tetrahedron* **54,** 2967–3042 (1998).

42. Tsuji, J. New general synthetic methods involving π-allylpalladium complexes as intermediates and neutral reaction conditions. *Tetrahedron* **42,** 4361–4401 (1986).

43. Tsuji, J. & Yamakawa, T. A convenient method for the preparation of 1-olefins by the palladium catalyzed hydrogenolysis of allylic acetates and allylic phenyl ethers with ammonium formate. *Tetrahedron Lett.* **20,** 613–616 (1979).

44. Li, B. *et al.* Histone H3 lysine 36 dimethylation (H3K36me2) is sufficient to recruit the Rpd3s Histone deacetylase complex and to repress spurious transcription. *J. Biol. Chem.* **284,** 7970–7976 (2009).

45. Culkin, D. A. & Hartwig, J. F. Synthesis, characterization, and reactivity of arylpalladium cyanoalkyl complexes: Selection of catalysts for the ??-arylation of nitriles. *J. Am. Chem. Soc.* **124,** 9330–9331 (2002).

46. Lei, C., Cao, J. & Zhou, J. Palladium-Catalyzed Arylation of Ketones and Acetonitrile with Ortho Alkylation of Aryl Rings: De Novo Synthesis of Tetralines and Benzocycloheptenes. *Org. Lett.* **18,** 6120–6123 (2016).

47. Spicer, C. D. & Davis, B. G. Palladium-mediated site-selective Suzuki-Miyaura protein modification at genetically encoded aryl halides. *Chem. Commun. (Camb).* **47,** 1698–1700 (2011).

48. Antos, J. M. & Francis, M. B. Transition metal catalyzed methods for site-selective protein modification. *Curr. Opin. Chem. Biol.* **10,** 253–262 (2006).

49. Kikot, P., Polat, A., Achilli, E., Fernandez Lahore, M. & Grasselli, M. Immobilized palladium(II) ion affinity chromatography for recovery of recombinant proteins with peptide tags containing histidine and cysteine. *J. Mol. Recognit.* **27,** 659–668 (2014).

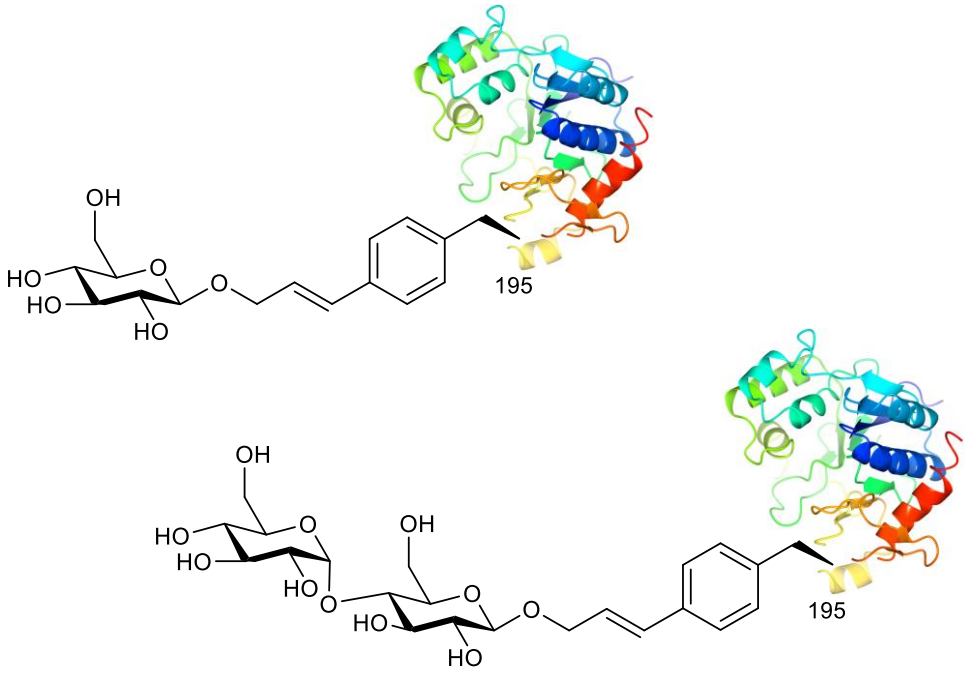
50. Chen, Y. *et al.* Coordination of two high-affinity hexamer peptides to copper(II) and palladium(II) models of the peptide - Metal chelation site on IMAC resins. *Inorg. Chem.* **39,** 1180–1186 (2000).

51. Dumas, A. *et al.* Self-liganded Suzuki-Miyaura coupling for site-selective protein PEGylation. *Angew. Chem. Int. Ed. Engl.* **52,** 3916–3921 (2013).

52. Chalker, J. M., Wood, C. S. C. & Davis, B. G. A convenient catalyst for aqueous and protein Suzuki-Miyaura cross-coupling. *J. Am. Chem. Soc.* **131,** 16346–7 (2009).

53. Geoghegan, K. F. *et al.* Spontaneous alpha-N-6-phosphogluconoylation of a 'His tag' in Escherichia coli: the cause of extra mass of 258 or 178 Da in fusion proteins. *Anal. Biochem.* **267,** 169–184 (1999).

Chapter 4



Comparison of GYG Glycoform Mimetics Reveals a Triphasic Mechanism

Contents

4.1 Introduction 234

4.2 Generation of an Active GYG-disaccharide Conjugate 237

4.2.1 Synthesis of Maltose Boronic Acid Coupling Partner 237

4.2.2 Generation and Catalytic Activity of GYG-Maltose 239

4.3 Investigating the Catalytic Activity of Early GYG Glycoforms 242

4.3.1 The Choice of an Assay Method 242

4.3.2 Comparing the Autoglucosylation Activity of GYG-Glucose and GYG-Maltose 245

4.3.3 Exploring an Unnatural Autogalactosylation Activity 249

4.3.4 Discussion 257

4.4 Conclusion and Outlook 259

4.5 Experimental Section 260

4.5.1 Chemical Synthesis 260

4.5.2 Suzuki-Miyaura Cross-coupling Optimisations 266

4.5.2.1 Initial Cross-couplings between GYG-Y195*p*IPhe and Maltose Boronic Acid 4.1 266

4.5.2.2 Cross-coupling Optimisation Studies (table 4.1) 267

4.5.2.3 Further Cross-coupling Studies: Effect of Reaction Time 271

4.5.3 General Protocol for Analysis of LC-MS Data for Glucosylation and Galactosylation Assays. 272

4.5.3.1 Lack of Effect of Number of Carbohydrate Residues Attached on MS Ionisation 273

4.5.4 Timecourse Autoglucosylation and Autogalactosylation Assays of GYG-Glucose and GYG-Maltose 275

4.5.4.1 General Protocol for Autoglucosylation and Autogalactosylation 275

4.5.4.2 Generation and Autoglucosylation Analysis of GYG-Glucose 277

4.5.4.3 Generation and Autoglucosylation Analysis of GYG-Maltose 282

4.5.4.4 Generation and Autogalactosylation Analysis of GYG-Glucose 288

4.5.4.5 Generation and Autogalactosylation Analysis of GYG-Maltose 292

4.5.4.6 Comparison of GYG-Maltose and GYG-Glucose Autoglucosylation, Equal Cross-Coupling 296

4.5.5 Autogalactosylation of wild-type GYG 298

4.6 References 301

4.1 Introduction

The privileged role of Glycogenin (GYG) in carbohydrate metabolism, as the initiator of a crucial glucose storage process, has sparked a desire to better understand the catalytic mechanism of this autoglucosylating enzyme. Studies to date, primarily on mammalian GYG-1 as used in this work, have suggested a complex catalytic process that is a direct consequence of its dimeric, autocatalytic nature.[1,2] As outlined in chapter 1, the former characteristic leads to the existence of distinct intra-subunit and inter-subunit modes of glucosylation within a GYG dimer, whilst the latter ensures that each glucosylation step is inherently unique.[3–5]

Crystallographic snapshots of the enzyme at different stages of its autoglucosylation process have highlighted the possibility of a switch between inter- and intra-subunit modes at an intermediate oligosaccharide chain length (between 4-6 glucose residues), as discussed in chapter 1; in a GYG-tetrasaccharide structure, the carbohydrate moiety extended towards the active site of the same monomer, whereas in a GYG-hexasaccharide it was instead oriented towards the other monomer of the GYG dimer.[5] Studies on mutant GYG heterodimers have also implied the necessity of both modes of glucosylation within a GYG dimer.[6] Such studies indicate the existence of at least two distinct mechanistic phases to GYG glucosylation, although questions necessarily arise concerning the extent to which inferences from single, static crystal structures can apply to a population of dynamic GYG molecules, and to which studies on mutant GYG species are relevant to the corresponding wild-type system.

One might envisage that any such distinct mechanistic phases – especially if connected with fundamental mechanistic changes – would likely exhibit different rates, and could thus be readily (further) distinguished and confirmed in this sense through kinetic analyses of the autoglucosylation process. However, studies adopting such an approach are notably and lacking,[7] a fact which – given the conceptual simplicity of such a study – is initially surprising. The reasons for this observation were briefly outlined in chapter 1 (see section 1.2.3), however are revisited here (**figure 4.1 a**). The heterogeneously glucosylated nature of GYG-wt isolated from *E. coli* expression[8] renders it clearly unsuitable for such studies.

An alternative approach adopted by Hurley *et al* that avoids this specific issue, in what remains to the best of our knowledge the only mechanistic study of GYG utilising a kinetics approach, is to isolate and utilise non-glucosylated enzyme (see chapter 1),[7] yet this too comes with inherent limitations. Whilst the authors were able to propose the existence of two distinct catalytic phases from timecourse analyses of the

glucosylation process, these could only be defined on the basis of accumulation of certain glycoforms (tetrasaccharide and hexasaccharide), and in terms of the nature of product distributions resulting from them ("non-distributive" or "uniform"); unambiguous phase definition in terms of global glucosylation rates was not possible.[7] It was evident that this reflected a notably slow initial glucosylation step of the acceptor tyrosine (Tyr195), which may act as a lag phase to obscure the details of immediately subsequent steps. Given a lack of unambiguous data concerning rates of other steps, however, this detail could not be incorporated into any kinetic definition of phases.[7]

Furthermore, as mentioned in chapter 1, the use of glycan-free GYG enables monitoring of only one glucosylation step on an individual level: the initial one (**figure 4.1 a**). This reflects the fact that the simplest method of following such a glucosylation step is to monitor the rate at which the starting enzyme is consumed, yet this requires access to that starting enzyme as a homogeneous species. Tracking individual glucosylation steps other than the first, therefore, requires access to homogeneous GYG glycoforms that themselves represent catalytic intermediates along the glucosylation pathway. Until now, this has not been possible.

In the previous chapter, an approach enabling access to a close mimetic of precisely such a species – GYG-Glucose – was outlined. Kinetic analysis of such an enzyme offers potential advantages over prior methods: 1) the slow Tyr195 glucosylation may be circumvented, enabling clearer visualisation of subsequent steps, and 2) an additional individual step – the addition of a glucose residue to GYG-Glucose, rather than to Tyr195 – may be studied, through monitoring the consumption of GYG-Glucose itself (**figure 4.1 b**). Furthermore, the earliest steps of GYG catalysis remain the most poorly understood, with even crystallographic studies failing to provide conclusive data;[5] the ability to isolate and analyse mimetics of early catalytic intermediates such as GYG-Glucose may remedy this.

However, it was equally apparent that our developed methodology could be rendered more powerful via its expansion to access additional GYG glycoforms corresponding to mimetics of other early catalytic intermediates, for example the disaccharide variant "GYG-Maltose" (**figure 4.1 b**). This would enable study of multiple individual glucosylation steps, whilst additionally facilitating direct, comparative analysis of such steps, potentially yielding significant insight into the catalytic mechanism. Furthermore, such a comparative approach would circumvent any issues relating to the effect of cross-coupling conditions on catalytic activity, since enzymes may be generated under essentially identical conditions.

As such, the following sections describe first the generation of such a "GYG-Maltose" enzyme. Whilst our approach could in theory enable access to a range of alternative GYG glycoforms, GYG-Maltose was

considered a particularly suitable target from both a biological perspective (the earliest catalytic steps are most poorly understood) and a chemical perspective (the requisite disaccharide boronic acid coupling partner represented a more synthetically tractable target than longer oligosaccharide variants). Subsequent comparative analysis of the catalytic activity of these homogeneous glycoforms, corresponding to mimetics of different early GYG catalytic intermediates, enables clear visualisation of the multiphasic nature of GYG catalysis from a kinetics perspective for the first time, whilst affording new insight into the little-understood earliest glucosylation steps.

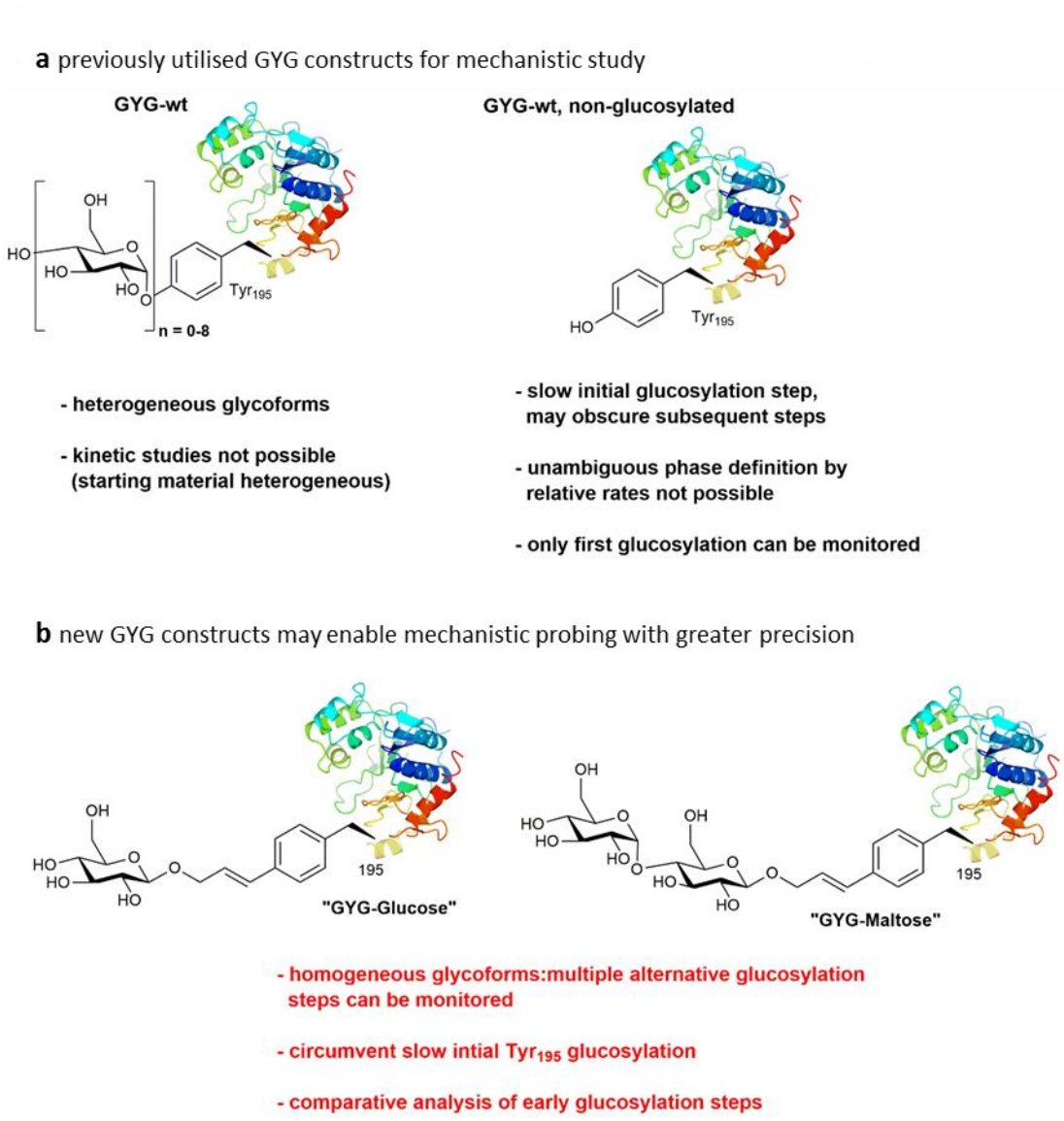


Figure 4.1 GYG constructs utilised in prior mechanistic studies have inherent limitations (a), especially with regards to kinetics-based studies. Study of homogeneous GYG glycoforms such as previously synthesised GYG-Glucose and disaccharide variant GYG-Maltose (synthesised in this chapter) may overcome these limitations, and enable mechanistic investigation with far greater precision (b).

4.2 Generation of an Active GYG-disaccharide Conjugate

4.2.1 Synthesis of Maltose Boronic Acid Coupling Partner

In an analogous fashion to the approach previously described for the generation of GYG-Glucose (chapter 3), the disaccharide glycoconjugate GYG-Maltose may be generated via cross-coupling of a maltose-derived boronic acid **4.1** to GYG-Y195*p*IPhe (**figure 4.2**). This enzyme represents the first catalytic intermediate of the autoglucosylation of GYG-Glucose, yet cannot be isolated in such a fashion, since enzymatic glucosylation yields distributions of different glycoforms.[7]

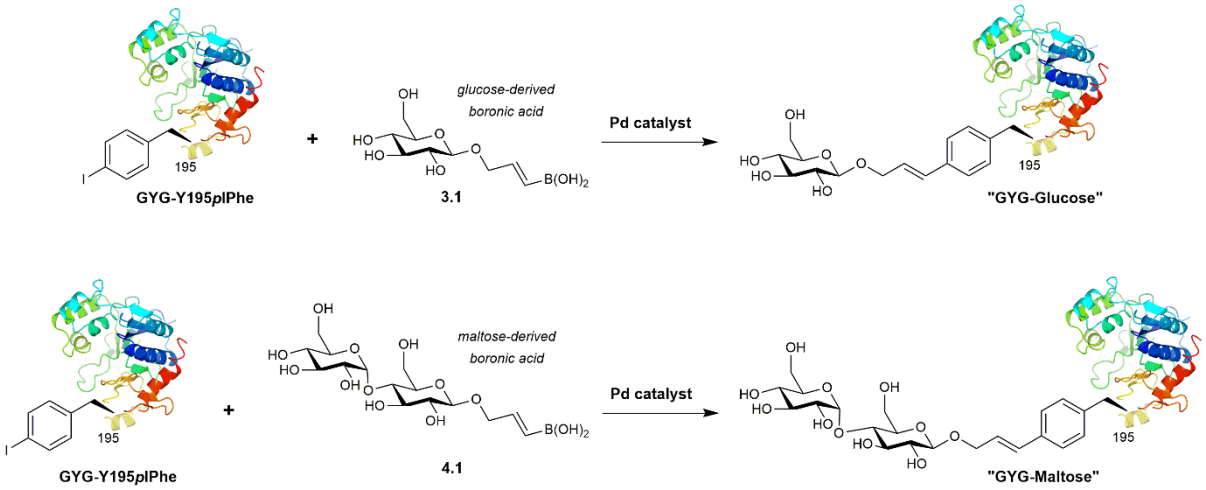


Figure 4.2 Generation of GYG-disaccharide conjugate GYG-Maltose (lower) may be carried out via an analogous cross-coupling to that used to generate GYG-Glucose (upper).

Our first focus was thus the generation of the requisite maltose boronic acid **4.1**. A synthesis for this species has not been previously reported, however we reasoned that synthetic routes towards monosaccharide analogues thereof might be readily adapted.[9] Thus, in an analogous fashion to that previously outlined (chapter 3, **figure 3.6**) two convergent syntheses might be envisaged (**figure 4.3**), again reliant on four key chemical transformations; glycosylation, hydroboration, boronic ester hydrolysis, and acetate deprotection. In the former route, which is the same as that utilised for glucose boronic acid **3.1**, glycosylation of β-D-maltose pentaacetate **4.2** with propargyl alcohol would yield glycoside **4.3**, which could be subjected to hydroboration to yield pinacol ester **4.4**; this is analogous to the route utilised towards glucose boronic acid **3.1** (chapter 3). However, this pinacol ester could instead be accessed via hydroboration of propargyl alcohol derivative **3.4** and subsequent glycosylation of peracetate **4.2** with its deprotected product **3.5**. Subsequently, boronic ester hydrolysis and acetate deprotection of the resultant intermediate **4.5** should afford the desired target compound.

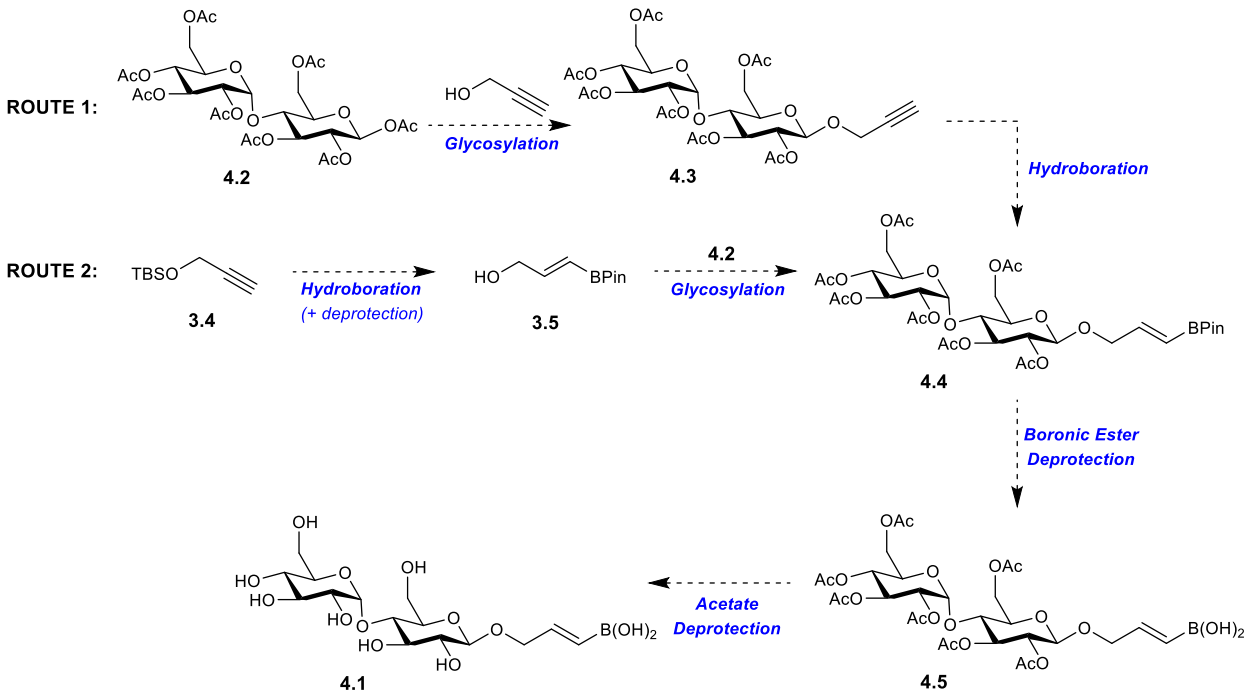
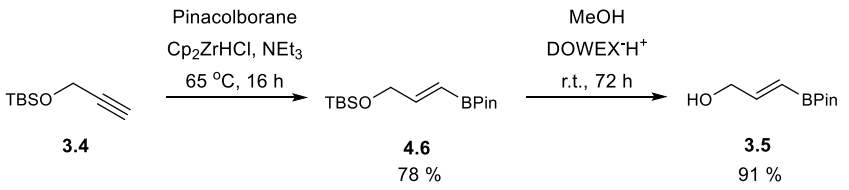


Figure 4.3 Possible routes towards maltose boronic acid **4.1**. Route 1 represents the same route as used towards glucose boronic acid **3.1**. Route 2 represents an alternative, ultimately employed in this case due to fewer manipulations of the more complex disaccharide. Note that these are the same types of routes as for monosaccharides (**figure 3.6**).

Reasoning that, for this more complex disaccharide target, minimising the number of synthetic steps involving manipulation of the disaccharide itself might be beneficial, it was decided to pursue the latter option (route 2 in **figure 4.3**), rather than that utilised for the glucose variant. The requisite glycosyl acceptor **3.5** was synthesised in two steps as previously reported from silyl-protected propargyl alcohol **3.4**, via protected pinacol ester **4.6** (**figure 4.4 a**).[9] $SnCl_4$, previously utilised as an effective promoter of glycosylations involving maltose peracetate donors,[10,11] successfully promoted glycosylation between **3.5** and commercially available β-D-maltose peracetate **4.2**; MS indicated low levels of mono-deacetylation, with ^{1}H NMR implicating this to occur at the 2'-OAc proximal to the site of glycosylation (see experimental section 4.5.1). Thus, a brief re-acetylation step afforded pinacol ester **4.4**. Given difficulties encountered during the purification of this species, the crude product was used directly in the subsequent step. Boronic ester deprotection via the diethanolamine ester as utilised previously (see chapter 3) failed to proceed, however treatment of crude pinacol ester **4.4** with $NaIO_4$ led to successful formation of protected boronic acid **4.5** in an overall yield of 16 % from peracetate **4.2**. Whilst this yield was relatively modest, sufficient quantities of pure target compound **4.1** could nonetheless be attained following global deprotection of **4.5** under Zemplen conditions (**figure 4.4 b**), as predominantly the β-anomer (> 90 % by ^{1}H NMR).

a



b

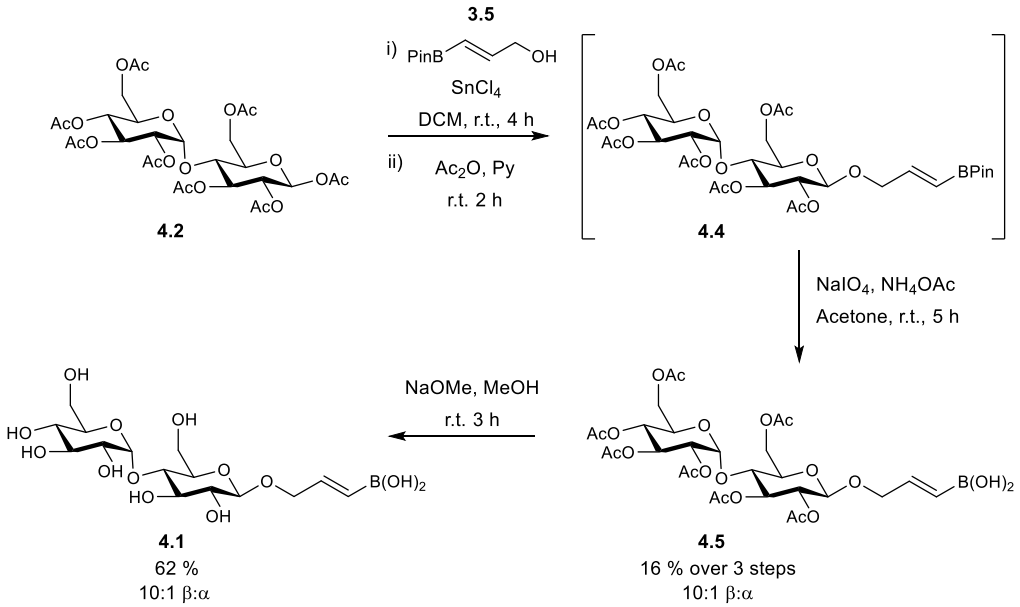


Figure 4.4 Synthesis of glycosyl acceptor **3.5** (a) and maltose boronic acid **4.1** (b).

4.2.2 Generation and Catalytic Activity of GYG-Maltose

With disaccharide boronic acid **4.1** to hand, cross-coupling to GYG-Y195*p*IPhe was attempted in order to generate the glycoform GYG-Maltose. Whilst it is likely that conditions optimal for generation of GYG-Glucose would also be optimal for GYG-Maltose formation, this is not necessarily the case. Therefore, as a starting point, conditions utilised for initial couplings in the generation of non-his-tagged GYG-Glucose were employed, although an increased boronic acid loading was employed since we anticipated that this disaccharide boronic acid may prove less reactive than its monosaccharide variant. Importantly, Pd loadings were restricted to 0.6 mM, identical to those utilised for GYG-Glucose; given the previously observed effects of Pd concentration on subsequent enzymatic activity (see chapter 2), this was considered important to permit fair comparisons of the two enzymes in this regard. It should be noted that, based on results obtained in the previous chapter, all work in this chapter and in subsequent chapters utilises non-his-tagged GYG-Y195*p*IPhe; this will not be explicitly stated on figures, but should be presumed from hereon in.

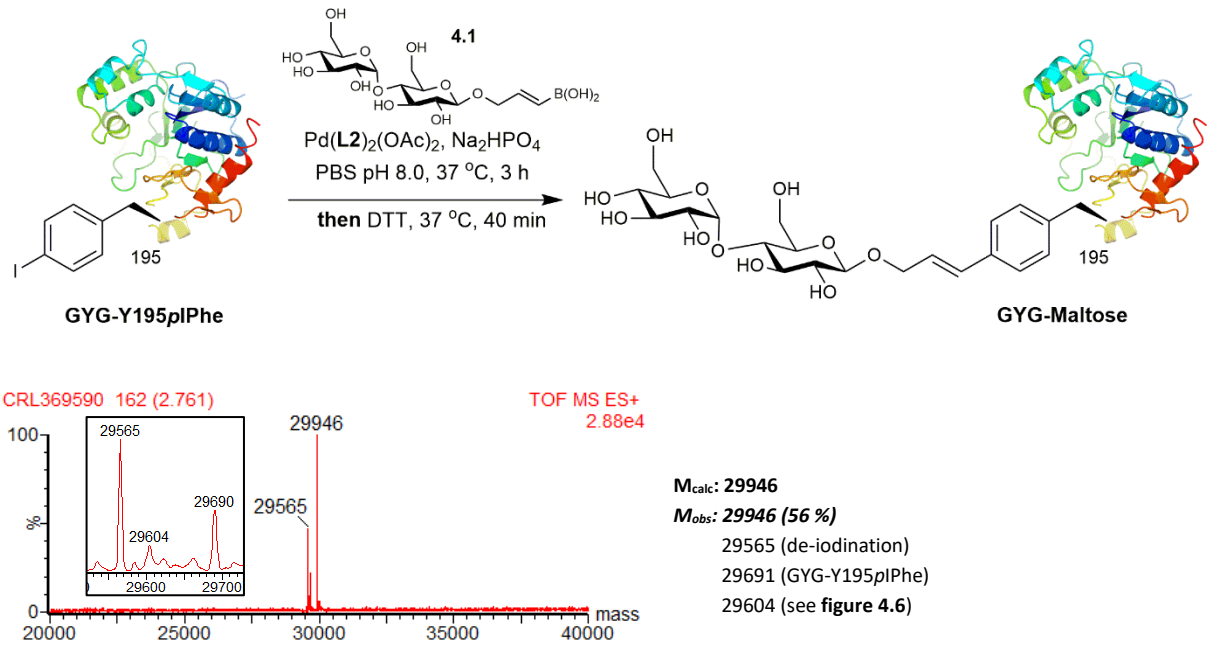


Figure 4.5 Initial cross-coupling attempts between GYG-Y195*p*IPhe and maltose boronic acid **4.1** to generate GYG-Maltose.

Pleasingly, an initial effort revealed the successful formation of the desired GYG-Maltose enzyme, as confirmed by LC-MS analysis, in moderate (58 %) conversion (**figure 4.5**). It is seen from the LC-MS spectrum above that side-products were the same as those observed for prior couplings to GYG-Glucose. These were discussed in greater detail in chapter 3, and result from de-iodination (M_{calc} = 29565), unreacted starting material (M_{calc} = 29692) and a hypothesised loss of the carbohydrate moiety via hydrogenolytic C_{allyl}-O cleavage of the cross-coupled product to afford allyl-linked GYG **3.11** (**figure 4.6**, also see pp. 172-3, chapter 3). Levels of the latter species are very low in this particular instance, however it is noted here since its presence is apparent in later samples, as will be seen later.

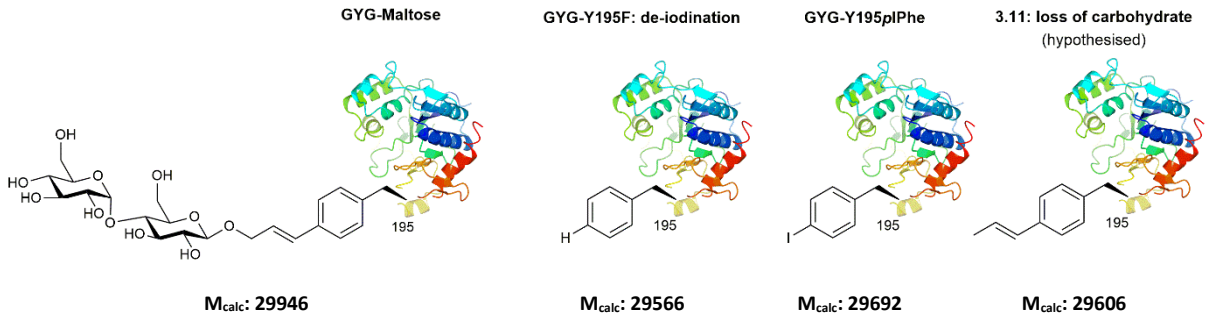
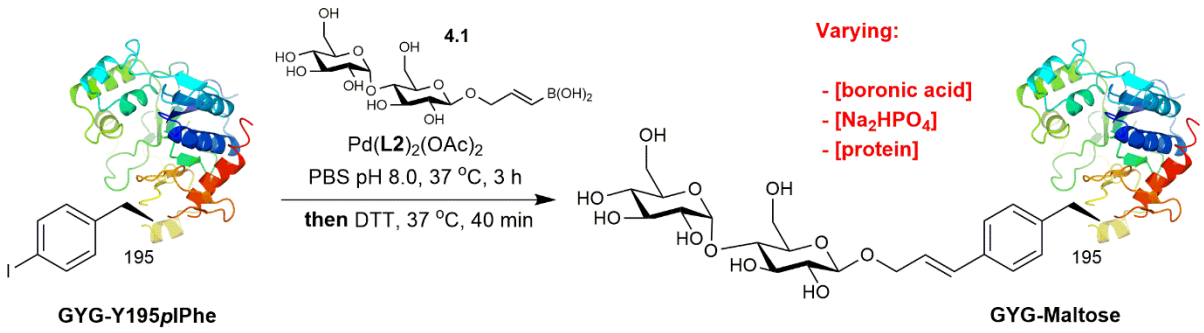


Figure 4.6 Species observed during cross-couplings to generate GYG-Maltose. See chapter 3 pp. **172-3** for more detailed discussion, especially concerning the structure **3.11**.

This reaction was then subjected to the same optimisation process as utilised for GYG-Glucose, in which factors such as protein concentration, level of Na_2HPO_4 and boronic acid loading were varied in small-scale (30 μl) optimisation studies (**table 4.1**). Reactions in this table were run in parallel so as to minimise any extraneous differences between them. As for the GYG-Glucose system, omission of Na_2HPO_4 from the reaction mixture proved beneficial (**table 4.1**, entry 4), whilst neither an increase in protein concentration nor in boronic acid loading had significant effect (**table 4.1**, entries 2, 5). Whilst full conversions were not attainable, optimal conditions for GYG-Maltose (**table 4.1**, entry 4) were essentially identical to those utilised for GYG-Glucose, ensuring that any differences observed between the catalytic activities of GYG-Glucose and GYG-Maltose were highly unlikely to reflect any differences in their preparation. In a separate experiment, it was also established that an increase in reaction time was not beneficial; whilst the ratio of product to known side-products present in all other samples improved slightly, other unidentified side-products began to appear (see experimental section 4.5.2.2)**.**



Entry	Equivs. $B(OH)_2$	Final [GYG-Y195*p*IPhe]	Final [Na_2HPO_4]	% Conversion[a]
1	3600	0.4 mg/ml	7.4 mg/ml	57
2	3600	**0.5 mg/ml**	7.4 mg/ml	56
3	3600	0.4 mg/ml	**10.4 mg/ml**	57
4	*3600*	*0.4 mg/ml*	***none***	***70***
5	**4400**	0.4 mg/ml	7.4 mg/ml	58

[a]*Determined by relative peak intensities in LC-MS spectra.*

Table 4.1 Optimisation of cross-coupling between GYG-Y195*p*IPhe and maltose boronic acid **4.1**. This table represents reactions run in parallel, so as to ensure equivalence of all conditions other than those intentionally varied (shown in red).

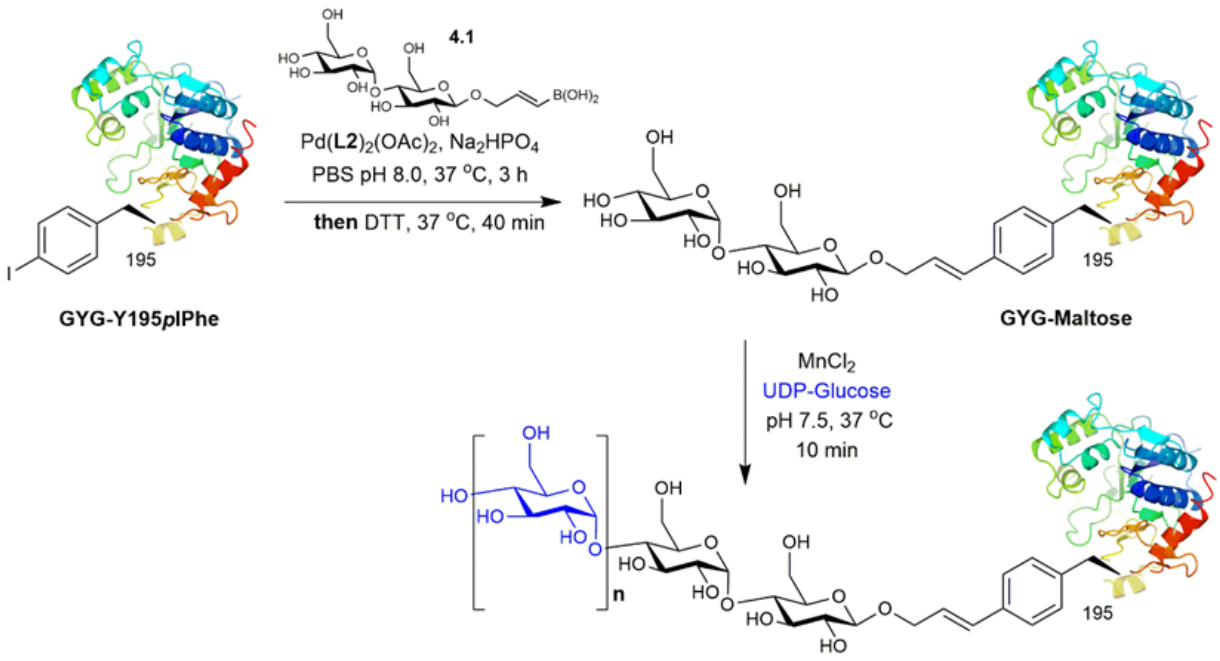
With optimised cross-coupling conditions in hand, we investigated whether the GYG-Maltose enzyme was also catalytically active. Employing scaled-up (100 μl) but otherwise identical cross-coupling conditions to those developed above (**table 4.1** entry 4), a maximum conversion of 76 % could be attained (**figure 4.7**, see also experimental section 4.5.4.3). This is slightly lower than that for GYG-Glucose (85 %), which is not overly surprising given that disaccharide boronic acid **4.1** represents a more sterically demanding cross-coupling partner than its monosaccharide analogue **3.1**.

Pleasingly, LC-MS analysis following glucosylation assay of the cross-coupled enzyme GYG-Maltose confirmed the activity of the latter (**figure 4.7**). Peaks corresponding to addition of multiple glucose residues were clearly observed. No such peaks were observed for the cross-coupling side-products, as expected given their lack of a correctly positioned glycosyl acceptor, and consumption of GYG-Maltose could be seen in relation to these species. The precise nature of these side-products is further discussed at the close of this chapter (section 4.3.4). At this point, we simply note that the generally similar cross-coupling conversions for GYG-Maltose and GYG-Glucose systems should largely negate any possible impact of these side-products during subsequent comparisons of these enzymes' activities (and that any such impact, as outlined in section 4.3.4 and investigated further in chapter 5, is seemingly minimal).

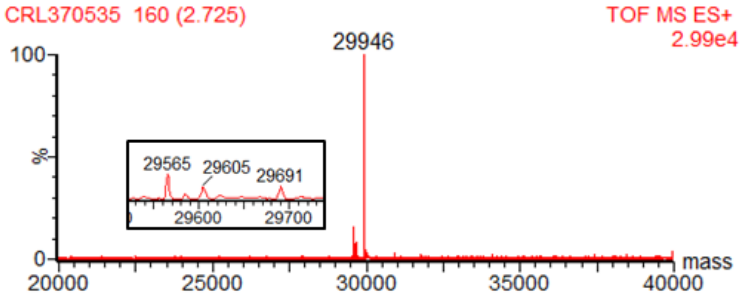
4.3 Investigating the Catalytic Activity of Early GYG Glycoforms

4.3.1 The Choice of an Assay Method

Having at this point successfully accessed two homogeneous, active GYG glycoforms, our focus turned to a more thorough investigation of their catalysis. As outlined previously, it was envisaged that timecourse-based analyses of GYG-Glucose and GYG-Maltose glucosylation may afford mechanistic insight with far greater clarity than previously possible.[7] This reflects both circumvention of the slow Tyr195 glucosylation step and an ability to directly compare different, individual (early) glucosylation steps of GYG catalysis (simply via monitoring the rate of consumption of either GYG-Maltose or GYG-Glucose). Specifically, this latter could provide information regarding the little-understood early steps of GYG catalysis: the "maltose → maltotriose" glucosylation step (GYG-Maltose) and the "glucose → maltose" glucosylation step (GYG-Glucose). Such analyses have not been previously possible.



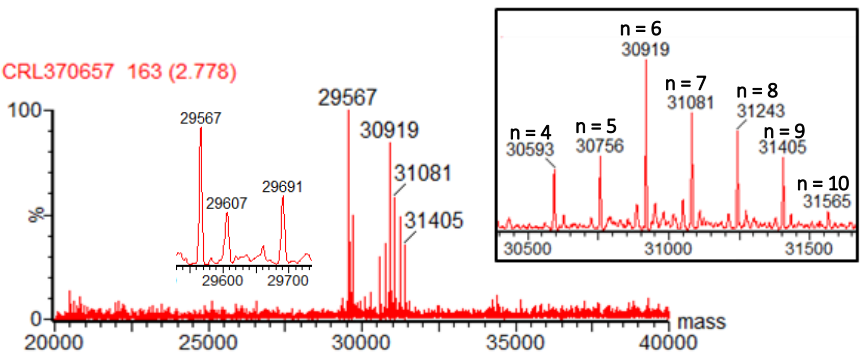
LC-MS after cross-coupling: 76 % conversion



M_{calc}: 29946
M_{obs}: 29946 (76 %)

\+ non-glucosylatable side-products (29565, 29605, 29691), see. **figure 4.6**

LC-MS after glucosylation:



M_{calc}: 29946 + 162n
M_{obs}: 29946 + 162n (n = 4-10)

\+ non-glucosylated side-products from cross-coupling (29567, 29607, 29692), **see figure 4.6**

Figure 4.7 Generation and autoglucosylation of GYG-Maltose.

Whilst access to such homogeneous GYG glycoforms undoubtedly opens up new possibilities, the utility of these is lost without a suitably powerful analytical method. Specifically, in order to follow individual steps, not only is access to homogeneous starting enzyme for each step required, but there is a need for an analytical approach that can directly track the levels of individual glycoforms of the enzyme. A wide range of methods for the assay of glycosyltransferase activities exist.[12] These include fluorescence-based assays (e.g. via the use of fluorescently labelled reagents),[13,14] direct monitoring of donor sugar nucleotide levels (or of its hydrolysed reaction product) through chromatographic methods,[15,16] and radiochemical methods based upon incorporation of a radiolabel from the donor sugar nucleotide into the reaction product.[17–20] The latter of these has been widely utilised for GYG, [2–4,21,22] however such an approach can afford data regarding only the average incorporation of glucose by the enzyme (since it monitors the UDP-glucose substrate, not the enzyme itself). It cannot follow the abundances of different, individual glycoforms as was desirable in this work.

LC-MS was previously utilised in this work as a qualitative measure of GYG autoglucosylation, however it has also been applied as a label-free method for quantification of glycosyltranserase activity for other enzymes[23,24] (in such non-autocatalytic cases, this implies monitoring of the small molecule substrates/products, not the enzyme itself). Notably, Hurley *et al* successfully developed an LC-MS-based approach to track GYG autoglucosylation, validating this method through a demonstration of its consistency with the alternative, widely-used radiochemical method.[7] Others have subsequently utilised this method, or an analogous approach based on LC-MS/MS of the relevant glucosylated GYG peptide, as a means to quantify glucosylation extent.[6,25]

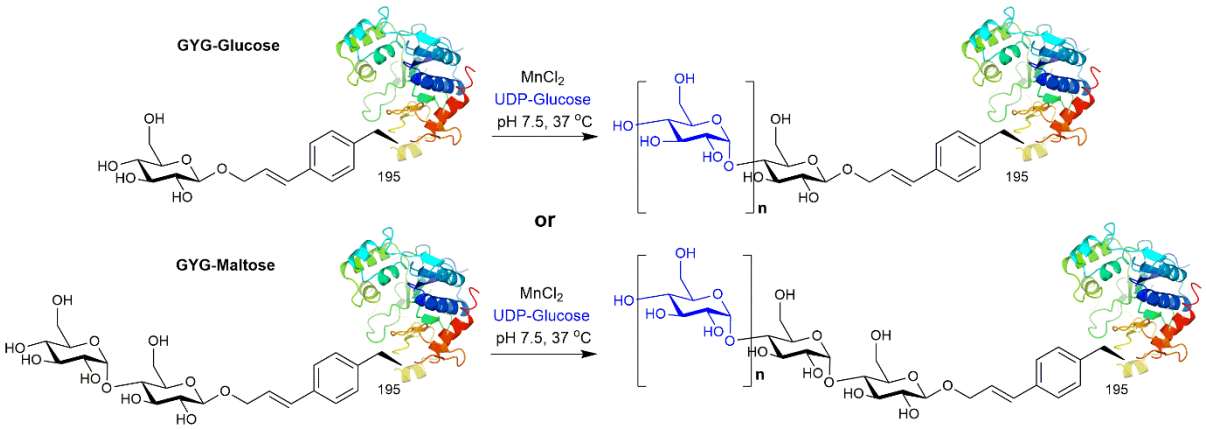
Such an LC-MS-based strategy offers a unique ability to track the levels of individual, distinct GYG glycoforms in the context of this work. This method has been described and validated in detail previously (and is also outlined in experimental section 4.5.3),[7] however in brief, the proportion of any one glycoform arising from cross-coupled enzyme (relative to all other glycoforms) is determined from the intensity of its corresponding MS peak relative to the combined intensities of all such glycoforms (i.e. total cross-coupled/glucosylated enzyme). This principle can be used to track rates of consumption of the homogeneous starting glycoforms GYG-Glucose and GYG-Maltose, which is a means of tracking and comparing the initial glucosylation step for each. It may also enable visualisation of the proportions of each glycoform present at different timepoints. Additionally, from the proportions of each glycoform present, it is possible to determine the overall number of monosaccharide residues attached to each GYG molecule (on average) at a given time point – the "overall glucosylation profile" – through multiplying the proportions

of each glycoform present by the number of monosaccharides associated with it (see ref. 7 and experimental section 4.5.3 for full details). A notable benefit of this method is its reliance only on calculating abundances of glycoforms relative to each other, as opposed to any absolute measures. This serves as an internal normalisation process, such that any run-to-run variations in MS ionisation, or in total level of starting enzyme (i.e. cross-coupling efficiency), are accounted for.

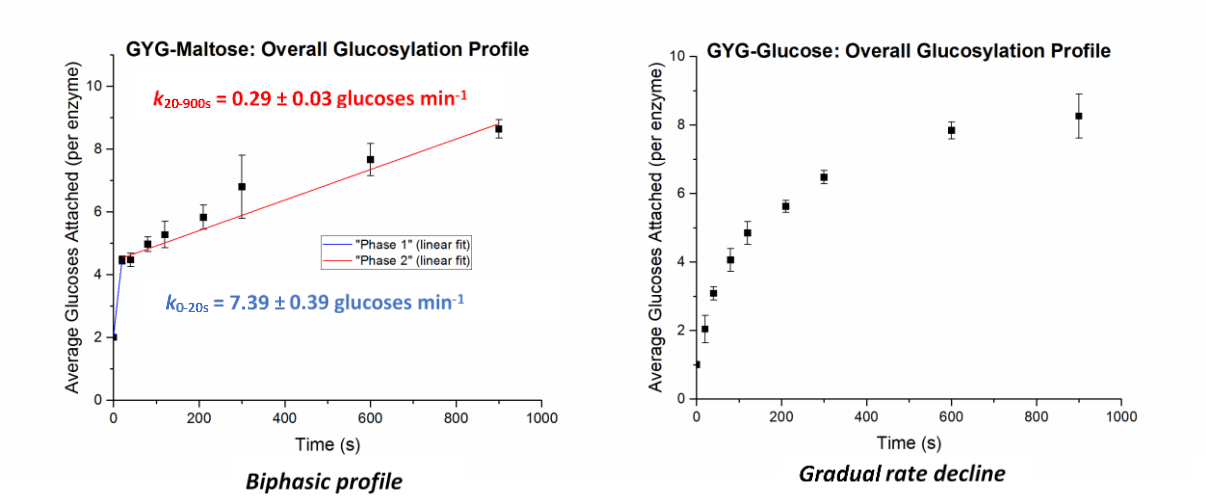
Based on the above, we elected to pursue LC-MS as a means of direct, real-time monitoring of GYG autocatalysis; the above approach was employed in all such studies in this regard throughout the remainder of this thesis. There is an inherent assumption in this methodology that each GYG glycoform ionises at comparable levels. However, that this is holds true may be inferred from known agreement of this method with radiochemical alternatives,[7] and indeed it has been explicitly demonstrated that the number of glucoses attached to a Tyr195-containing GYG peptide does not significantly impact MS ionisation.[25] Furthermore, we found this assumption to hold empirically; levels of glucosylated GYG, in relation to the non-glucosylated side-products (whose level should thus remain constant), remained similar at early and late time points (see experimental section 4.5.3.1).

4.3.2 Comparing the Autoglucosylation Activity of GYG-Glucose and GYG-Maltose

In order to compare the autoglucosylation activity of the GYG-Glucose and GYG-Maltose enzymes, both were generated according to their respective, near-identical optimised cross-coupling protocols, then – following thorough Pd-removal – subjected to autoglucosylation assay conditions. These involve incubation of the cross-coupled enzyme preparation with 1 mM of Mn^{2+} cofactor and 1 mM UDP-Glucose substrate, representing a ca. 200-fold excess of both with respect to enzyme, which is easily sufficient to ensure saturation of the donor substrate.[6] After various time-points, aliquots were quenched (via sequestration of the Mn^{2+} cofactor with excess EDTA[5,7]) and analysed by LC-MS. Identical concentrations of cross-coupling preparations were used throughout (including side-products, i.e. constant concentration of GYG species). Subjecting LC-MS spectra obtained at different time points from independent replicates to the analytical process outlined above (see experimental section 4.5.3) enabled real-time tracking of not only overall levels of glucosylation, but also of the consumption of GYG-Glucose and GYG-Maltose (**figure 4.8**). The latter, as discussed, represents a means of following the Glc1 → Glc2 and Glc2 → Glc3 glucosylation steps respectively.



a Total numbers of glucoses attached per enzyme



b Comparison of initial glucosylation step rates

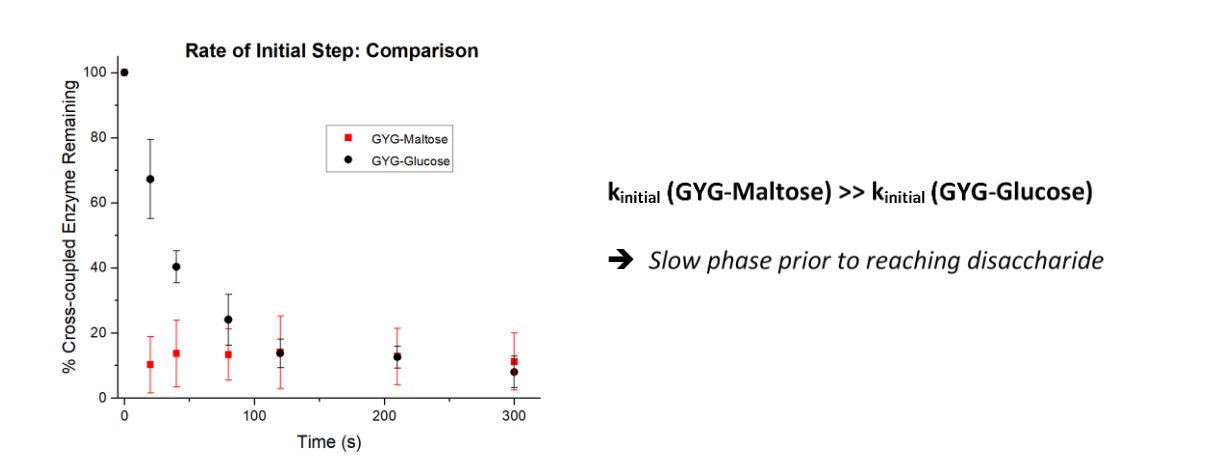


Figure 4.8 Comparison of autoglucosylation profiles of GYG-Maltose and GYG-Glucose reveals a distinct "fast → slow" biphasic profile for the former, but not the latter (a). This arises due to a markedly slower initial glucosylation step for GYG-Glucose (b), thus implicating an overall "slow → fast → slow" profile. Data points represent mean averages of at least 3 independent replicates, error bars are ± s.d. Lines on the upper left plot are linear regression fits (red), or simply joining of two adjacent data points (blue).

Results obtained proved revealing, particularly those pertaining to the GYG-Maltose enzyme. Monitoring the overall average number of glucoses attached per enzyme for this species, two very distinct glucosylation phases could be clearly distinguished (**figure 4.8 a**, **left**). There exists an initial rapid glucosylation phase, in which an average of 2.46 ± 0.13 glucose residues are added to the initial disaccharide in just 20 seconds (i.e. $k_{0\text{-}20}$ = 7.39 ± 0.39 glucoses min^{-1}). This is followed by an abrupt and significant rate decrease, and subsequent slower glucosylation. This latter phase may be, as illustrated, reasonably well approximated as linear ($k_{20\text{-}900}$ = 0.29 ± 0.03 glucoses min^{-1}). Species bearing up to 13 glucoses in total (i.e. 10 attached to the starting disaccharide) were detected after this final phase (see experimental section 4.5.4.3, also re-visited in **figure 4.13**).

Intriguingly, GYG-Glucose did not appear to exhibit such biphasic kinetics, despite representing the direct precursor of GYG-Maltose. Instead, a more gradual decline in enzyme turnover with increasing oligosaccharide length was noted (**figure 4.8 a, right**). This reflects a difference in the rates of initial glucosylation steps for each enzyme (**figure 4.8 b**). That for GYG-Maltose (Glc2 → Glc3) is particularly rapid, with almost all starting enzyme consumed within 20 seconds; indeed, it is this very fact that enables unambiguous visualisation of the subsequent rapid phase and biphasic profile. In contrast, that for GYG-Glucose (Glc1 → Glc2) is markedly slower, which serves to obscure visualisation of subsequent kinetic phases.

The transition from rapid, initial phase to slow latter phase for GYG-Maltose glucosylation is well illustrated through an examination of the glycoforms present and their abundances at early time points. LC-MS snapshots from t = 0 s to t = 80 s are shown below (**figure 4.9**); for purposes of clarity, only data for one replicate is shown, however good reproducibility in terms of glycoform distributions was observed between replicates at these timepoints (see experimental section 4.5.4.3). Thus, the initial 20 s sees significant glucosylation, with the conversion of the initial GYG-disaccharide into predominantly GYG-tetrasaccharide and GYG-pentasaccharide species (i.e. n = glucoses *added* = 2, 3). However, relatively little further change in glycoform distribution is seen upon extension to t = 40 s and t = 80 s. As such, the rapid initial phase would appear to correspond to the addition of mainly 2-3 glucose moieties to the initial GYG-Maltose disaccharide – generating GYG species bearing 4-5 glucoses in total.

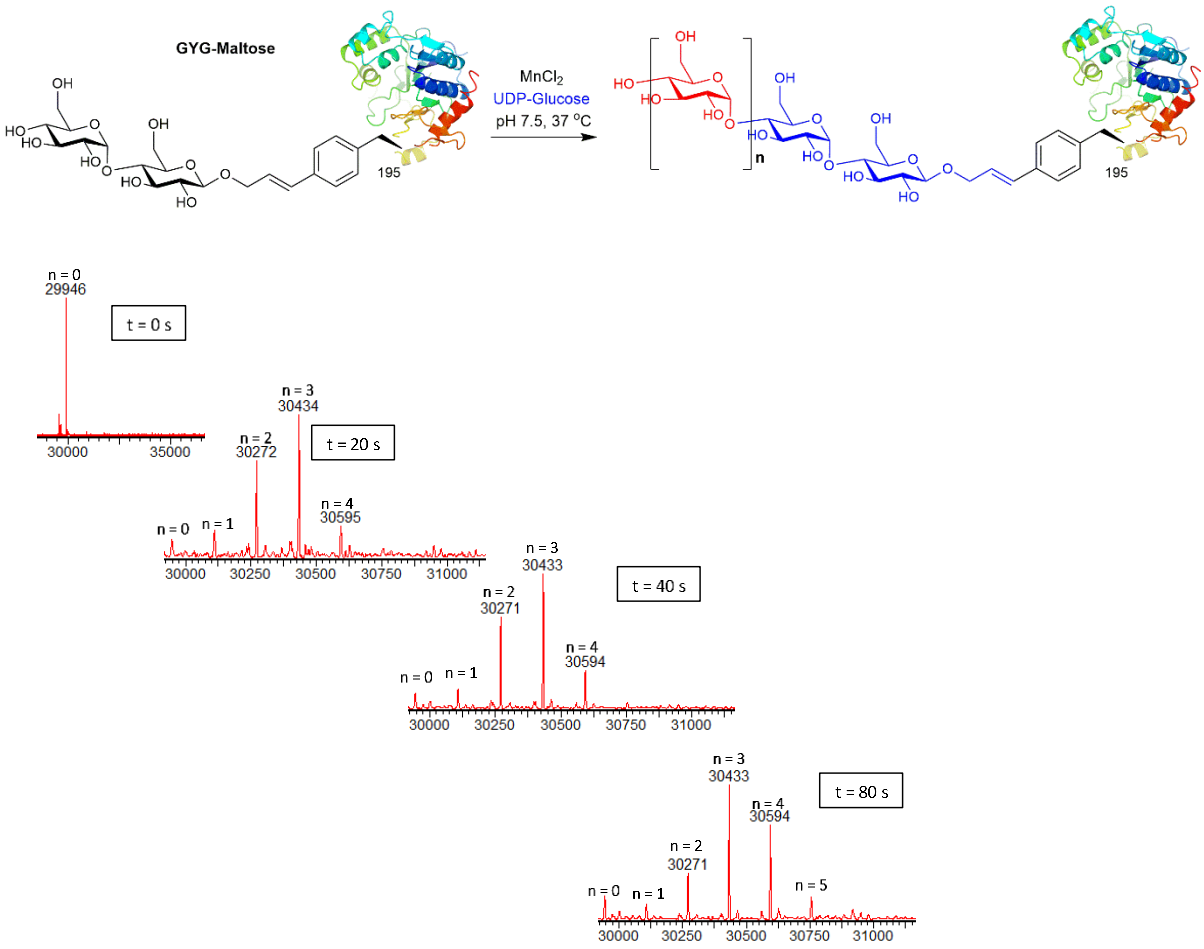


Figure 4.9 Typical GYG-glycoform distributions obtained during autoglucosylation of GYG-Maltose after various early time points, as illustrated by representative LC-MS spectra, serving to highlight a fast → slow transition after t = 20 s. Only portions of the MS corresponding to cross-coupled enzyme and glucosylation thereof are shown here, see experimental section 4.5.4.3 for corresponding full spectra

Collectively, the data for GYG-Maltose and GYG-Glucose suggests the existence of (at least) three distinct glucosylation phases exist for this catalytic system, distinguishable by their relative rates (**figure 4.10**). The data for the GYG-Maltose system, commencing from a disaccharide, reveal the latter two of these phases (rapid 2nd phase → slow final phase). However, these two phases must by definition also exist for GYG-Glucose even if not directly visible, given it is on the same glucosylation pathway. As such the slow initiation step for GYG-Glucose, in comparison for the rapid one for GYG-Maltose, must represent an additional slow phase prior to reaching the disaccharide. GYG-Maltose, therefore, does not exhibit this phase.

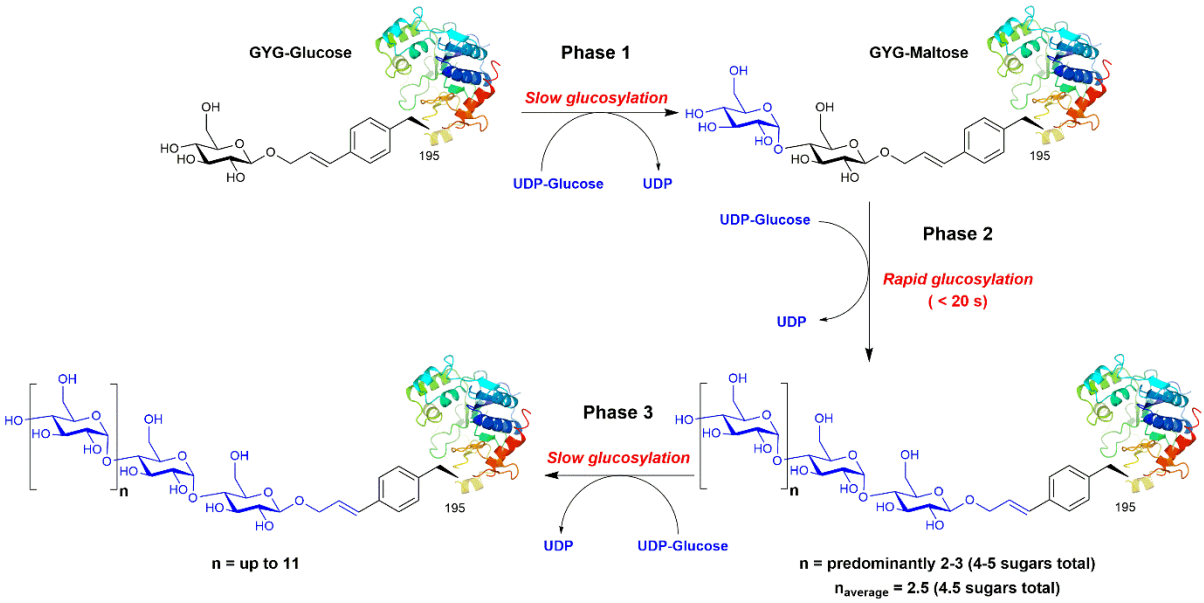


Figure 4.10 Proposed triphasic mechanism for GYG catalysis, as constructed from studies on GYG-Glucose (first phase) and GYG-Maltose (latter two phases), and illustrated by LC-MS snapshots (only catalytically active portion shown; for full LC-MS spectra, including inactive side products from cross coupling, see experimental section)

4.3.3 Exploring an Unnatural Autogalactosylation Activity

In order to further probe the nature of GYG catalysis, we turned to a similar yet distinct catalytic pathway. In addition to its natural autoglucosylation activity, GYG has also been reported as capable of catalysing an analogous autogalactosylation process, utilising UDP-Galactose as a donor substrate.[26] However, to our knowledge, this has not been investigated further to any significant degree; whilst it is known that autogalactosylation proceeds on average at a rate of about one seventeenth of that of autoglucosylation,[26] no data regarding the degree of oligosaccharide extension for this process – i.e. what GYG glycoforms are produced and when – is available. We felt that the MS-based approach employed above, in combination with the homogeneous glycoforms accessible through our methodology, could further elucidate this intriguing pathway, whilst also possibly further illuminating the catalytic mechanism as a whole. **Figure 4.11**, below, illustrates that both GYG-Glucose and GYG-Maltose were indeed capable of undergoing such galactosylation (assay conditions were identical to those employed for glucosylation, other than time points analysed and the use of UDP-Galactose as substrate).

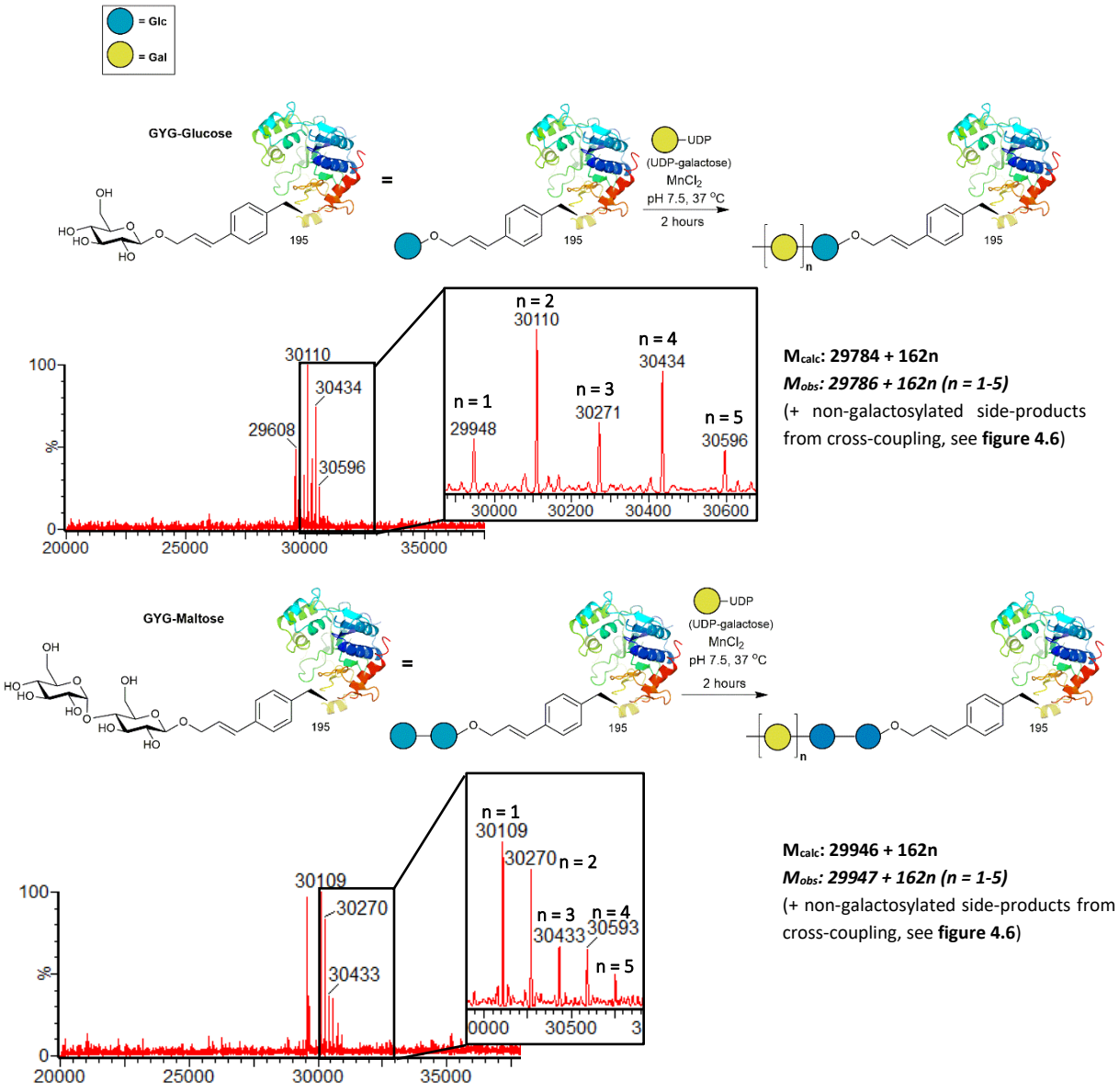


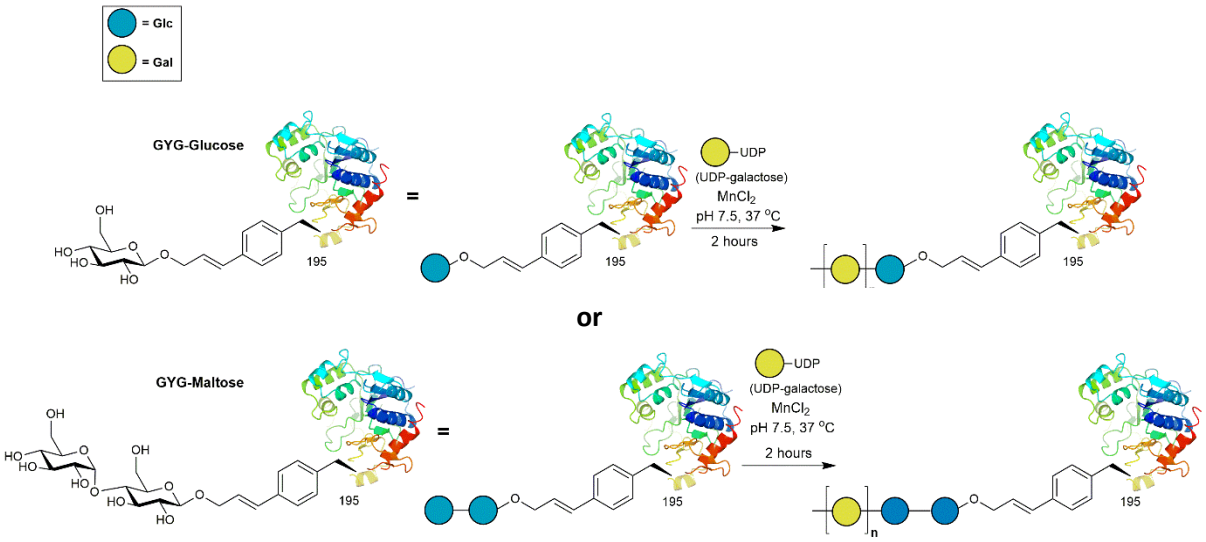
Figure 4.11 Autogalactosylation of GYG-Glucose (top) and GYG-Maltose (lower), as illustrated by LC-MS.

It will be noticed that in this figure and others relating to such autogalactosylation activity, the carbohydrate moieties attached to GYG are represented by symbols according to those conventionally utilised in glycan notation,[27] as opposed to chemical structures. This is at present necessitated by the fact that, in contrast to the well-characterised autoglucosylation pathway for which it is known that glucose monomers are connected via α-(1,4)-glycosidic linkages,[28,29] the connectivity of the Gal-to-Gal and initial Glc-to-Gal glycosidic linkages in GYG autogalactosylation remain at present unknown. Elucidation of this was beyond the scope of the present work, although further discussion comes at the end of this chapter.

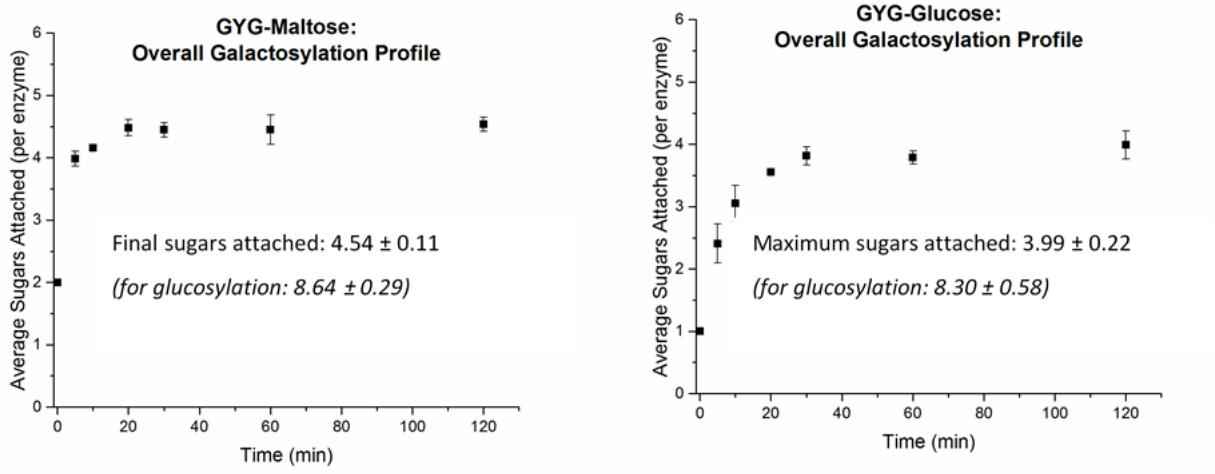
In an analogous fashion to that carried out for autoglucosylation (**figure 4.8**), the kinetics of this galactosylation process was investigated in more detail for both enzymes. Thus, both the overall galactosylation profiles and the rates of the initial galactosylation step were compared (**figure 4.12**); again, these were attained from LC-MS analysis at various timepoints of independent replicates as described above (see also experimental section 4.5.3). It should be noted that, for the former case, the graph refers to "average *sugars* attached", which reflects the fact that this includes both the Gal residues added enzymatically and the Glc residues ligated through cross-coupling.

The most noteworthy feature in both cases was a significantly reduced extent of galactosylation for both GYG-Maltose and GYG-Glucose, as compared to glucosylation; as seen, the maximum average number of carbohydrate residues attached to both enzymes was in the range of 4-4.6, in contrast to 8.3-8.6 for glucosylation (**figure 4.12 a**). It was notable that species bearing more than 8 carbohydrate residues (i.e. Gal6 for GYG-Maltose, Gal7 for GYG-Glucose) were never detected for autogalactosylation of either enzyme (i.e. irrespective of initial number of glucoses), and species bearing 7/8 residues were present only in low abundance from 20 minutes onwards (**figure 4.13**). This contrasts starkly to glucosylation, for which species bearing up to 12-13 glucose residues were detectable (**figure 4.13**). On the other hand, species bearing up to 5 sugars were routinely detected after the earliest timepoint (5 min), even whilst starting enzyme was still present in significant amounts (for GYG-Glucose). It should be noted that the concentrations of UDP-sugar employed were 1 mM for both UDP-Glucose and UDP-Galactose, which is a ca. 200-fold excess with respect to enzyme, thus substrate depletion is not responsible for the curtailment of galactosylation.

Other aspects were qualitatively similar to autoglucosylation; GYG-Maltose exhibited a fast → slow biphasic profile whilst GYG-Glucose did not, and GYG-Glucose exhibited a slower initial galactosylation step (**figure 4.12 b**). This highlights the fact that this latter enzyme appears to exhibit generally slower rates irrespective of donor sugar nucleotide. However, whilst the same features outlined above delineated a triphasic mechanism for the autoglucosylation of GYG-Glucose, this cannot be inferred for its autogalactosylation; GYG-Maltose is in this case not an on-pathway intermediate, thus its biphasic profile is not necessarily present for GYG-Glucose.



a Total numbers of carbohydrate (Glc/Gal) residues attached per enzyme.



b Comparison of initial galactosylation step rates.

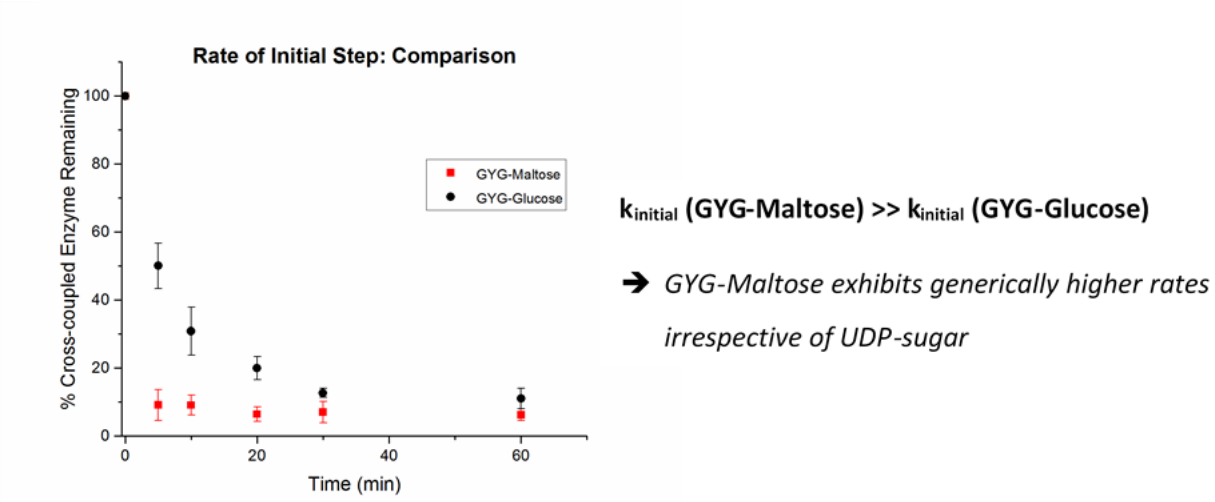
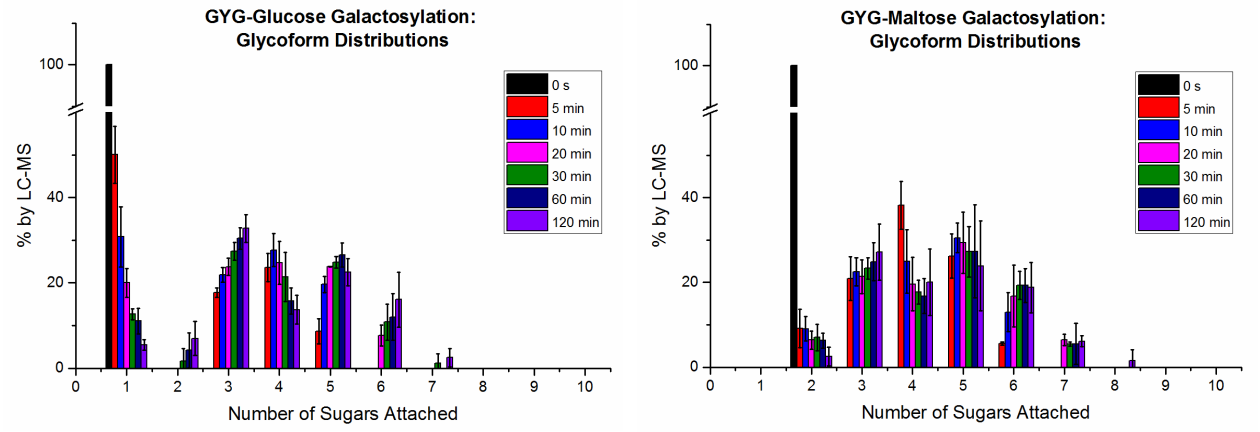


Figure 4.12 Autogalactosylation profiles for GYG-Maltose and GYG-Glucose (a), with comparison of initial step rates (b). Data represent mean averages of at least 3 independent experiments, error bars are ± s.d.

a Galactosylation: glycoform distributions



No more than 8 sugars (total, per enzyme) attached

b Glucosylation glycoform distributions, for comparison

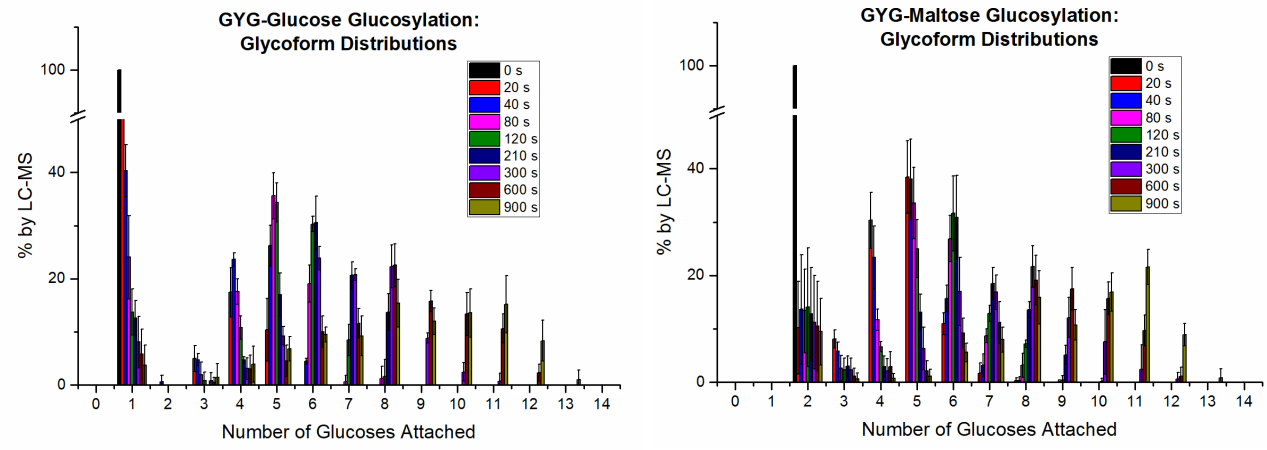


Figure 4.13 Analysis of the glycoforms present and their abundances during GYG-Glucose and GYG-Maltose galactosylation (a). No more than 8 sugars in total were ever attached to a given enzyme, in stark contrast to analogous glucosylation assays (b), in which up to 12-13 sugars were attached. Note that "sugars/glucoses attached" includes the initial glucose residues ligated through SMC. Error bars represent ± s.d. of at least 3 independent replicates; these were the same samples analysed to generate the data shown in **figure 4.8** and **figure 4.12**. Glycoform abundances were determined from LC-MS spectra as detailed in section 4.5.3.

The apparent reduced upper limit for autogalactosylation as opposed to autoglucosylation was intriguing, and further insight in this regard came from a prior, preliminary observation that wild-type Glycogenin – already bearing up to 7 glucose residues – underwent detectable but very limited autogalactosylation (**figure 4.14**). Notably, after 1 hour of autogalactosylation, no GYG species bearing any more than 8 monosaccharide residues was observed; thus, even GYG species bearing 7 *glucose* residues initially undergoes

very limited reaction. In contrast, autoglucosylation of this enzyme proceeds to a far greater extent (see chapter 2). This data implies that the lack of autogalactosylation at higher oligosaccharide lengths is independent of the nature (glucose or galactose) of this oligosaccharide; rather, the catalytic mechanism during the latter stages of GYG catalysis appears generically incompatible with autogalactosylation.

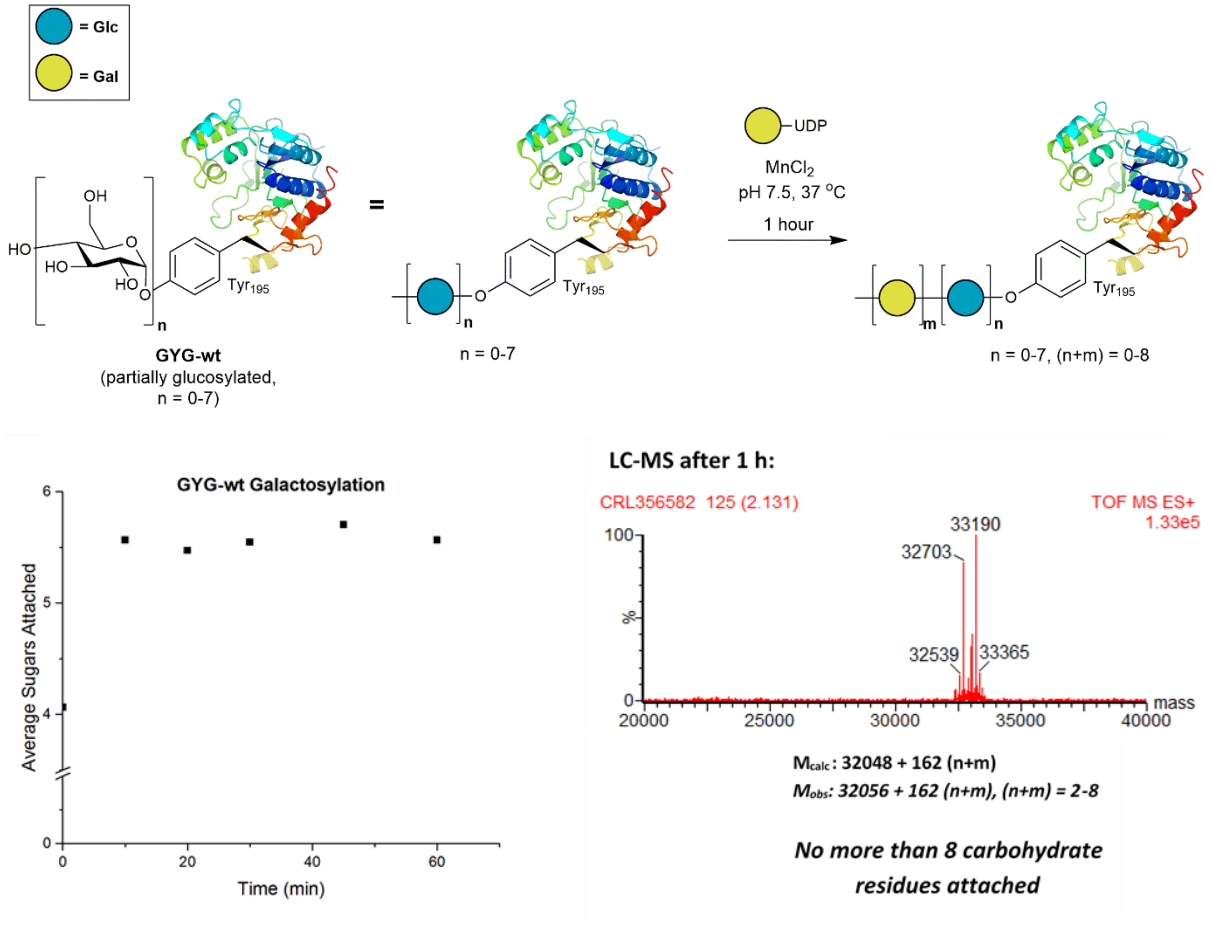


Figure 4.14 Autogalactosylation of GYG-wt. Despite already bearing up to 7 glucose residues, autogalactosylation was limited and no glycoforms bearing more than 8 carbohydrate moieties were observed. This implies a generic intolerance of the enzymatic mechanism towards autogalactosylation at higher oligosaccharide chain lengths.

The results obtained from the autogalactosylation studies may be combined with those obtained from autoglucosylation to afford a summary of the overall catalytic process as implied by the data presented in this chapter, commencing from GYG-Glucose (**figure 4.15**). Thus, as outlined previously, the natural glucosylation pathway seemingly involves a slow → fast → slow triphasic process. At various points in this pathway – including, but not necessarily limited to, GYG-Glucose and GYG-Maltose – offshoots to an unnatural galactosylation pathway may occur; this may also exhibit triphasic kinetics, though this cannot be judged from the data obtained, and alignment of the pathways in **figure 4.15** are principally for ease of comparison. This pathway, however, proceeds to a lesser extent than glucosylation, with GYG species

bearing more than 8 carbohydrate residues (initial, chemically ligated Glc and enzymatically added Gal) never observed across multiple replicates.

Indeed, this latter fact – and the previous observation that 7-sugar and 8-sugar GYG glycoforms were present only in very low abundances and after prolonged galactosylation assay times – indicates that, at oligosaccharide chain lengths corresponding to the 3rd kinetic phase of glucosylation (> ca. 5 sugar residues), galactosylation (or more specifically Gal-to-Gal transfer) is relatively limited in extent, and cannot occur throughout the whole phase. The only significant Gal-to-Gal transfer seen in this phase is from 5-sugar to 6-sugar (see **figure 4.13**); further transfers beyond this are apparently challenging (7-sugar, 8-sugar) or non-existent (beyond this).

In contrast, the preceding 2nd phase appears overall more tolerant of such Gal-to-Gal transfer. The latter may seemingly in principle readily occur across all oligosaccharide lengths encompassed by this phase (2 → 4/5 sugar residues); GYG species bearing up to 5 sugars were routinely observed in significant amounts from the earliest time points of galactosylation enzymes of both cross-coupled enzymes (see **figure 4.13**). Notably, this latter point holds true for both GYG-Maltose and GYG-Glucose; for the latter enzyme, galactosylation corresponding to the 2nd phase involves solely Gal-to-Gal transfer (**figure 4.15**). Therefore, the observations from galactosylation assays serve to suggest apparent further differences between the latter two phases of GYG catalysis within our system.

Regarding the first kinetic phase (< 2 sugars), no comment can be made concerning Gal-to-Gal transfer at such oligosaccharide chain lengths and its possible relationship in this regard to other stages of GYG catalysis, since at present autogalactosylation during this phase corresponds only to galactosylation of GYG-Glucose, i.e. a fundamentally distinct transfer of galactose to a *glucose* acceptor. This is, however, an area that will be explored further in chapter 6.

The plasticity of GYG with respect to non-glucose substrates, highlighted by its capacity to autogalactosylate to different degrees in different catalytic phases, could potentially have *in vivo* relevance; UDP-galactose is readily available in cells, being directly accessible from galactose itself via the Leloir pathway,[30] and as such GYG could, at least in principle, incorporate UDP-galactose *in vivo*. Whether or not this, and in particular the differing tolerances of catalytic phases to Gal-Gal transfer, in fact plays any *in vivo* role is not known, although some speculation is provided in chapter 7.

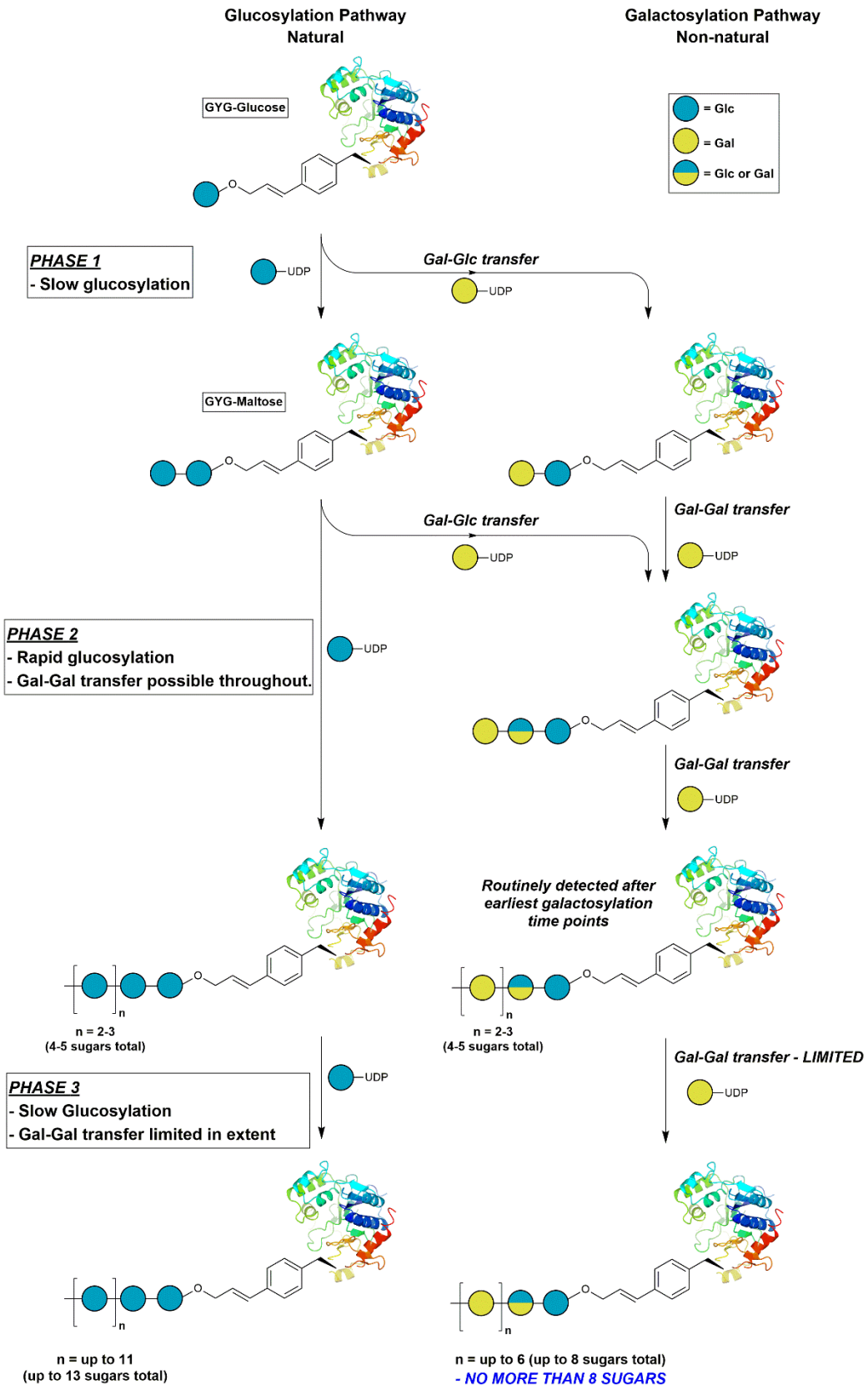


Figure 4.15 Summary of autoglucosylation and autogalactosylation pathways in GYG catalysis. Note that e.g. "Gal-Glc" refers to transfer of galactose *to* a glucose acceptor, and that galactosylation and glucsoylation pathways are aligned principally for ease of comparison.

4.3.4 Discussion

The results presented thus far implicate a triphasic kinetic mechanism for GYG catalysis, with the phases uniquely visible as a consequence of our ability to track individual glucosylation steps and circumvent the slow Tyr195 glucosylation step. Notably, the findings obtained in this work fit well with prior literature, building upon this rather than refuting it, as discussed below; this is also re-visited in chapter 7. This indicates that our system is a representative one, yielding relevant results in spite of currently unavoidable factors such as the non-natural "linker" moiety between carbohydrate and protein (see chapter 3), and the presence of small amounts of non-glucosylatable cross-coupling side-products (see **figure 4.6**), both discussed below.

With regards to prior studies, the "fast → slow" transition between 2nd and 3rd phases detailed in this work at ca. 5-6 glucose residues was never before unambiguously observed, yet is directly coincident with prior crystallographic data suggesting an intra-monomer to inter-monomer mechanistic transition at precisely this oligosaccharide chain length (as described in detail in chapter 1, section 1.2.2).[5] Furthermore, Hurley *et al* previously suggested a phase transition at the same point, although in their case this could only be defined through the natures of product distributions, not by relative rates of phases as achieved in the present study.[7] Meanwhile, the notion of a slow 1st phase for GYG catalysis is in itself not new – it is known that the initial Tyr195 glucosylation is slow[7]. Rather, the idea that it may encompass more than just this initial glucosylation has not previously been suggested, and the subsequent transition to a rapid phase upon reaching a disaccharide has never previously been observed.

As noted above, the GYG-Glucose and GYG-Maltose species do not represent exact mimics of the corresponding native glycoforms, given the presence of a short non-natural "linker" between carbohydrate and protein moieties, including an allyl moiety and a non-native β-glycosidic linkage (see chapter 3). Whilst the effect of this cannot be readily evaluated – it is impossible to isolate the corresponding natural variants for comparison – as discussed above, the results obtained from our system fit well with prior studies on the wild-type enzyme,[5–7] suggesting it is unlikely to be overly significant. Furthermore, the analytical methods employed, and conclusions drawn from these, are reliant on comparison of abundances of different glycoforms and, whether generated chemically (GYG-Glucose, GYG-Maltose) or enzymatically (from assays thereof), they all contain this same linker. It is possible, owing to the short additional length provided by the linker, that "GYG-Glucose", for example, in fact mimics GYG-wt with fractionally more than 1 glucose attached, but any such alteration would firstly be minimal and secondly would apply equally to all glycoforms, thus not affecting the conclusions drawn.

The nature of the cross-coupling side-products should also be discussed in slightly greater detail, in order to better evaluate any effects (or indeed lack thereof) arising from them. Whilst these indeed lack any capacity to be glucosylated, they are almost certainly capable of catalysing transfer of glucose from UDP-Glucose to the carbohydrate moiety of another GYG enzyme molecule (e.g. cross-coupled enzyme) in an inter-enzyme sense, i.e. acting as a "donor enzyme". Indeed, this capacity is specifically documented for GYG-Y195F[6] (de-iodination side-product). By extension, it presumably also exists for the other cross-coupling side-products, since these differ from GYG-Y195F (and GYG-wt) only at the 195 position, which plays no evident part in the key roles of the "donor enzyme" in inter-enzyme glucosylation (binding UDP-Glucose and carbohydrate acceptor from the other monomer, and catalysing transfer between these).[1,5]

However, numerous factors discussed below strongly indicate that these side-products, particularly their typically slightly differing levels in preparations of GYG-Maltose versus GYG-Glucose, do not affect our conclusions. Firstly, differences in cross-coupling efficiencies for these enzymes were minor – particularly in comparison to the very stark contrasts in activities of the two enzymes – and additionally the kinetic data was reproducible across replicates in spite of similar low-level variation in cross-coupling efficiency across replicates. Indeed, such variation meant that, for one of each of the GYG-Glucose and GYG-Maltose replicates in autoglucosylation, their cross-coupling efficiencies were essentially identical; comparison of these replicates in isolation yielded the same results as detailed previously, thus indicating that conclusions drawn are not influenced by cross-coupling side-product levels (see experimental section 4.5.4.6).

Secondly, from a theoretical perspective, the rate of glucosylation of a molecule of cross-coupled enzyme within a dimer would only be significantly affected by proportions of side-products if inter-enzyme glucosylation from the latter was notably different than that from cross-coupled enzyme itself (intra-monomer glucosylation rates would also be unaffected).[i] Given that cross-coupled products and side-products are structurally identical other than at the 195 position, and the lack of involvement of this residue from the perspective of the "donor enzyme", this is unlikely to be the case.

Thirdly, and most crucially, in the next chapter it will be empirically demonstrated, as an offshoot from a subsequent study on a third enzyme, that the deliberate "spiking" of either GYG-Glucose or GYG-Maltose with significant quantities of one representative such cross-coupling side-product – unreacted GYG-Y195*p*IPhe (the presence or absence of which represents a key difference between most GYG-Glucose and

[i] Note that, as discussed previously, any increase / decrease in level of cross-coupled enzyme within the mixture is accounted for by the "internal normalisation" of the LC-MS analytical method, see p. 239; all data is determined from proportions of glycoforms *relative* to total levels of glycoforms (i.e. total levels of coupled enzyme).

GYG-Maltose preparations) – has only minimal quantitative effects and does not affect any conclusions drawn in this work, indicating that such conclusions supersede any variations due to e.g. side-products. Finally, we again emphasise the fact that our results are well in line with existing literature as discussed above,[5,7] implying that our system is a representative one irrespective of any such side-products.

4.4 Conclusion and Outlook

In this chapter, our developed methodology, based on a palladium-mediated activation of GYG, was extended to enable access not only to a GYG-monosaccharide conjugate GYG-Glucose, but also to a new disaccharide variant, GYG-Maltose. Prior studies have suggested the existence of at least two phases for GYG catalysis, possibly resulting from distinct intra- and inter-monomer mechanisms, however unambiguous kinetic data to support this has been lacking.[5,7] Our approach allowed such data to be obtained, with comparative analyses of GYG-Glucose and GYG-Maltose implicating the existence of three distinct phases distinguished by relative rates in a "slow → fast → slow" profile. Discussion of possible reasons for this profile is provided in chapter 7, following further exploration in chapters 5 and 6. Further exploration revealed an interplay between autoglucosylation and an alternative, unnatural autogalactosylation pathway, the latter proceeding to a reduced extent and suggesting additional distinctions between the latter two kinetic phases in terms of their amenability towards Gal-to-Gal transfer.

Future directions stemming from this chapter include elucidation of the nature of the glycosidic linkages formed during autogalactosylation, which cannot be presumed to be α(1→4) as for glucosylation. Galactans comprising galactose monomers joined by precisely such a linkage do exist,[31] yet alternatives are also known (e.g. α(1→6), α(1→3)).[32,33] This could potentially be determined via protein NMR studies following galactosylation assays using ^{13}C-labelled UDP-galactose. It would also be of interest to investigate the other mammalian GYG isoform, GYG-2, and to evaluate its similarities and differences to the GYG-1 variant employed in the present study.[34]

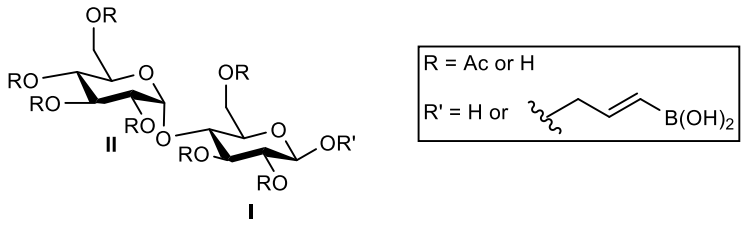
With our methodology having proven capable of accessing two distinct, catalytically active GYG glycoforms, the extent to which it can be further extended to access yet more GYG species was of interest. This final point forms the basis of the following two chapters of this thesis. In chapter 5, our efforts to generate and investigate a GYG-oligosaccharide conjugate are outlined, whilst chapter 6 focusses on the activity of "non-natural" GYG-carbohydrate species; that is, those deliberately intended *not* to mimic native glycoforms.

4.5 Experimental Section

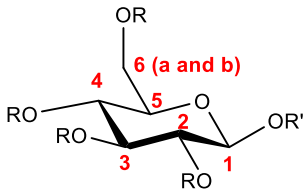
4.5.1 Chemical Synthesis

Note on NMR assignment nomenclature:

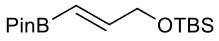
For the disaccharide compounds, superscript roman numerals are used to designate the ring from which a resonance belongs, according to the image below:



Numbers 1-6 are then used to designate the ring protons/carbons within a given hexose ring, as below:



For example, "H1^{II}" refers to the H1 proton within ring II above. "C6^{I}" refers to the C6 carbon in ring I above.



4.6

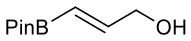
tert-Butyldimethyl(2-propynyloxy)silane (4 ml, 19.8 mmol) was charged under argon and pinacolborane (3.13 ml, 21.6 mmol) was added, followed by Schwartz's reagent (0.52 g, 1.98 mmol) and trimethylamine (0.276 ml, 1.98 mmol). The reaction mixture was stirred at 65 °C for 16 hours. After cooling to room temperature, pet. ether (15 ml) was added. The mixture was filtered through a short plug (ca. 20 g) of silica, which was rinsed with 60 ml of pet. ether. Any further remaining product was eluted through a further wash with 20% EtOAc in pet. ether. Concentration *in vacuo* of afforded silane **4.6** as a pale yellow oil (4.6 g, 15.4 mmol, 78%).

^{1}H NMR (400 MHz, Chloroform-*d*) δ 6.66 (dt, *J* = 18.0, 3.5 Hz, 1H, PinBCHCH), 5.74 (dt, *J* = 18.0, 2.1 Hz, 1H, PinBCH), 4.23 (dd, *J* = 3.5, 2.1 Hz, 2H, CH_2), 1.25 (s, 12H, BPin), 0.89 (s, 19H, *t*-Butyl), 0.04 (s, 6H, CH_3).

LRMS *m/z* (ESI+) 299.3 (M+H, calculated 299.2)

IR (neat) $\tilde{\nu}$ = 2979, 2957, 2931, 2887, 2858 (C-H), 1645 (C=C) cm^{-1}

Data consistent with that previously reported.[9]



3.5

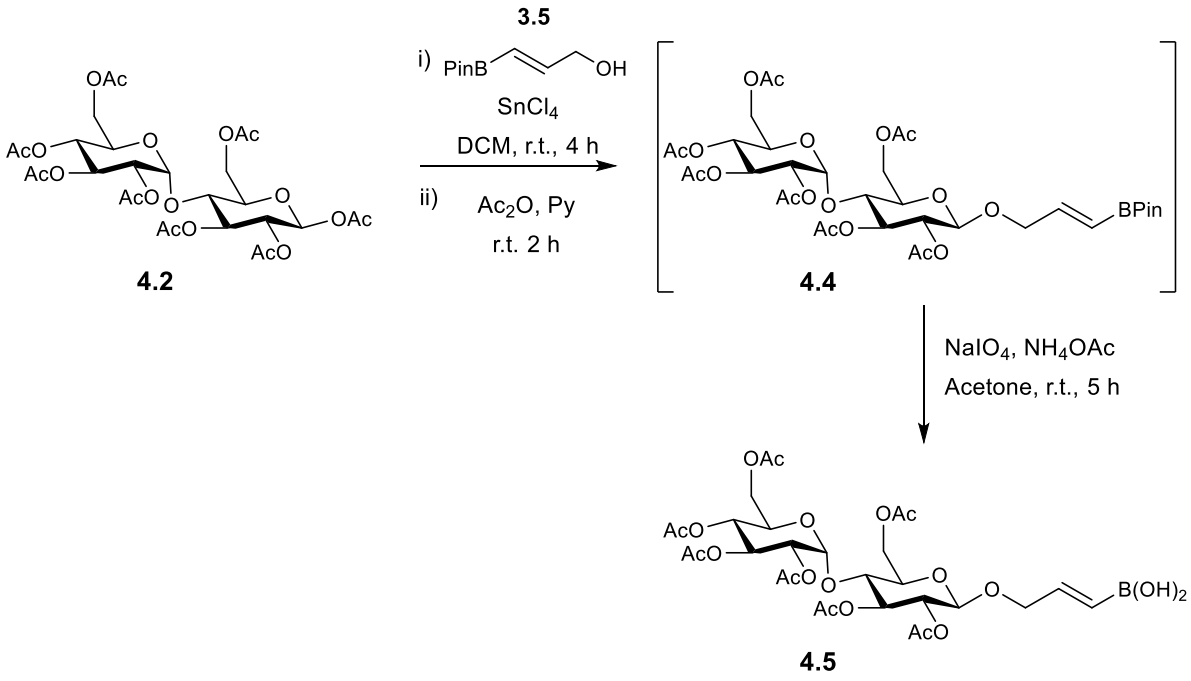
Silane **4.6** (3.9 g, 13.1 mmol) was dissolved in anhydrous MeOH (15 ml) and activated DOWEX-50WX8 added (ca. 2 g, for activation protocol please see chapter 3, section 3.5.1). The suspension was stirred at room temperature for 72 hours, after which time TLC showed the starting material to be almost fully consumed. The solution was filtered to remove the resin and the filtrate concentrated *in vacuo* to give an orange oil, which was purified by flash column chromatography (SiO_2, eluting with 7:3 Pet. Ether: EtOAc) to give alcohol **3.5** as a yellow oil (2.20 g, 12.0 mmol, 91 %).

^{1}H NMR (400 MHz, Chloroform-*d*) δ 6.72 (dt, *J* = 18.2, 4.2 Hz, 1H, PinBCHCH), 5.69 (dt, *J* = 18.2, 2.0 Hz, 1H, PinBCH), 4.22 (dd, *J* = 4.2, 2.0 Hz, 2H, CH_2), 1.25 (s, 12H, BPin)

HRMS *m/z* (ESI+) found: 185.1351 (M+H, calculated 185.1349).

IR (neat) $\tilde{\nu}$ = 3412 (O-H), 2979, 2932 (C-H), 1644 (C=C) cm^{-1}

Data consistent with that previously reported.[9]



Formation of intermediate **4.4**

Alcohol **3.5** was dried prior to use by azeotroping (2 x) with anhydrous toluene and then high vacuum (2 h). β-D-maltose octaacetate **4.2** (Carbosynth) was dried under high vacuum for at least 5 h. β-D-maltose octaacetate **4.2** (1.50 g, 2.20 mmol) was dissolved in anhydrous DCM (6.0 ml) in flame-dried flask under Ar. $SnCl_4$ (0.25 ml, 2.20 mmol) was added and the mixture stirred for 10 minutes at room temperature. Alcohol **3.5** (405 mg, 2.20 mmol) was added slowly as a solution in 4.0 ml DCM. The reaction was stirred at room temperature for 4 hours and was then diluted with EtOAc (40 ml) and quenched with sat. aq. $NaHCO_3$ (40 ml) before being filtered. The aqueous layer was extracted with EtOAc (3 x 50 ml) and the combined organics dried ($MgSO_4$) and concentrated *in vacuo*. NMR and MS indicated a low degree of mono-deacetylation (ca. 10 %); substantial downfield shift in NMR of the $H2^I$ proton before and after re-acetylation implicated the 2'-OAc group proximal to this as the likely site of this:

δ $H2^I$ (deacetylated): 3.51 (dd, *J* = 10.0, 3.9 Hz, α-anomer) and 3.44 (dd, *J* = 10.0, 8.0 Hz, β-anomer) (Chloroform-*d*, 400 MHz).

δ $H2^I$ (acetylated): 4.88 (dd, *J* = 9.4, 7.8 Hz, β-anomer) and 4.78 (dd, *J* = 10.1, 3.7 Hz, α-anomer) (Chloroform-*d*, 500 MHz).

The crude mixture thus obtained was redissolved in anhydrous pyridine (4.5 ml) and Ac_2O (0.5 ml) added dropwise at 0 °C. After stirring for 4 h at r.t., the reaction was diluted with EtOAc (15 ml) and quenched

with H_2O (ca. 15 ml). The organic layer was separated, and the aqueous extracted with EtOAc (3 x 40 ml). The combined organics were dried ($MgSO_4$) and concentrated *in vacuo*, then azeotroped with toluene (x 2) to remove pyridine and filtered through silica. ^{1}H NMR confirmed the presence of intermediate pinacol ester **4.4** (ca. 60% purity by ^{1}H NMR). This crude product was carried forward directly to the subsequent hydrolysis step without further purification. However, a separate small portion of crude product was subjected to purification by flash chromatography (SiO_2, 50→100% EtOAc in Pet. Ether) for characterisation purposes, data from which is displayed below (anomeric ratio of purified = 9:1 β:α). Anomeric ratio determined from integration of $H1^1_\alpha$ (δ 4.95, d, J = 3.7 Hz) and $H1^1_\beta$ (δ 4.56, d, J = 7.9 Hz).

^{1}H NMR for 4.4 (500 MHz, Chloroform-*d*, data for major anomer only shown) δ 6.54 (dt, J = 18.2, 4.6 Hz, 1H, BCH=C<u>H</u>), 5.66 (dt, J = 18.2, 1.8 Hz, 1H, BC<u>H</u>=CH), 5.40 (d, J = 3.9 Hz, 1H, <u>H1^{II}</u>), 5.35 (dd, J = 10.6, 9.5 Hz, 1H, <u>H3^{II}</u>), 5.24 (t, J = 9.1 Hz, 1H, <u>H3^{I}</u>), 5.04 (t, J = 9.9 Hz, 1H, <u>H4^{II}</u>), 4.88 (dd, J = 9.4, 7.8 Hz, 1H, <u>H2^{I}</u>), 4.84 (dd, J = 10.6, 4.2 Hz, 1H, <u>H2^{II}</u>), 4.56 (d, J = 7.9 Hz, 1H, <u>H1^{I}</u>), 4.46 (dd, J = 12.1, 2.8 Hz, 1H, <u>H6a^{I}</u>), 4.38 (ddd, J = 14.7, 4.3, 1.9 Hz, 1H, BCH=CHC<u>H$_2$</u>), 4.25 (dd, J = 12.4, 3.9 Hz, 1H, <u>H6a^{II}</u>), 4.21 (dd, J = 11.9, 4.3 Hz, 1H, <u>H6b^{I}</u>), 4.16 (ddd, J = 14.7, 5.1, 1.7 Hz, 1H, BCH=CHC<u>H$_2$</u>), 4.03 (dd, J = 12.4, 2.3 Hz, 1H, <u>H6b^{II}</u>), 4.00 (t, J = 9.2 Hz, 1H, <u>H4^{I}</u>), 3.95 (ddd, J = 10.3, 4.1, 2.3 Hz, 1H, <u>H5^{II}</u>), 3.66 (ddd, J = 9.6, 4.4, 2.7 Hz, 1H, <u>H5^{I}</u>), 2.14 (s, 3H, O<u>Ac</u>), 2.10 (s, 3H, O<u>Ac</u>), 2.04 (s, 6H, 2 x O<u>Ac</u>), 2.02 (s, 3H, O<u>Ac</u>), 2.00 (s, 3H, O<u>Ac</u>), 1.99 (s, 3H, O<u>Ac</u>), 1.26 (s, 12H, B<u>Pin</u>).

^{13}C NMR for 4.4 (126 MHz, Chloroform-*d*, major anomer only) δ 170.6, 170.5, 170.3, 170.0, 169.7, 169.5 (all O<u>Ac</u> C=O), 147.2 (BCH=<u>C</u>H), 120.2 (weak, B<u>C</u>H=CH), 99.2 (<u>C1^{I}</u>), 95.5 (<u>C1^{II}</u>), 83.4 (B<u>Pin</u>), 75.5 (<u>C3^{I}</u>), 72.7 (<u>C4^{I}</u>), 72.2 (<u>C2^{I}</u>), 72.1 (<u>C5^{I}</u>), 70.4 (BCH=CH<u>C</u>H$_2$), 70.0 (<u>C2^{II}</u>), 69.4 (<u>C3^{II}</u>), 68.5 (<u>C5^{II}</u>), 68.0 (<u>C4^{II}</u>), 62.8 (<u>C6^{I}</u>), 61.5 (<u>C6^{II}</u>), 24.8 (B<u>Pin</u>), 20.9, 20.9, 20.7, 20.7, 20.6, 20.6, 20.6 (all O<u>Ac</u> CH_3).

HRMS for 4.4 m/z (ESI+) found: 825.2955 (M+Na, calculated 825.2959)

IR for 4.4 (film) $\tilde{\nu}$ = 2978 (C-H), 1748 (C=O), 1647 (C=C) cm^{-1}

$[\alpha]_D$ for 4.4 = +47 (c = 1.0, $CHCl_3$)

Formation of boronic acid **4.5**

Crude pinacol ester **4.4** (1.05 g total mass, of which ca. 600 mg **4.4** by NMR, 0.75 mmol) was dissolved in acetone (15 ml), and 1M NH_4OAc (11 ml) was added, followed by $NaIO_4$ (800 mg, 3.7 mmol). The suspension was stirred vigorously at room temperature for 5 h. The acetone was removed *in vacuo* and the aqueous layer carefully acidified to pH 3 with 1M HCl before being extracted with EtOAc (3 x 40 ml). The combined organics were washed with brine, dried ($MgSO_4$) and concentrated *in vacuo*. The crude product was purified by column chromatography (SiO_2, 2.5-4-10 % MeOH in DCM). This afforded fractions containing pure product, and fractions containing product contaminated with a derivative of alcohol **3.5**. The latter were subjected to further purification by column chromatography (SiO_2, 22:3 toluene:acetone with 2% methanol). All fractions containing pure product were concentrated *in vacuo* to afford protected boronic acid **4.5** as a white crystalline solid (250 mg, 0.35 mmol, 16 % from β-D-maltose octaacetate, 10:1 β:α). Anomeric ratio determined from integration of $H1^{I}_{\alpha}$ (δ 5.00, d, J = 3.6 Hz) and $H4^{II}$ (δ 5.04, t, J = 9.8 Hz, both anomers), since the $H1^{I}_{\beta}$ proton partly overlapped with $H2^{I}$.

^{1}H NMR for 4.5 (500 MHz, Acetone-d_6, major anomer only) δ 6.51 (dt, J = 18.1, 4.8 Hz, 1H, BCH=C<u>H</u>), 5.65 (dt, J = 18.1, 1.8 Hz, 1H, BC<u>H</u>=CH), 5.37 (d, J = 3.8 Hz, 1H, <u>H1^{II}</u>), 5.35 (dd, J = 10.5, 9.8 Hz, 1H, <u>H3^{II}</u>), 5.30 (t, J = 9.1 Hz, 1H, <u>H3^{I}</u>), 5.04 (t, J = 9.8 Hz, 1H, <u>H4^{II}</u>), 4.86 (dd, J = 10.6, 3.9 Hz, 1H, <u>H2^{II}</u>), 4.82 (dd, J = 9.3, 8.0 Hz, 1H, <u>H2^{I}</u>), 4.76 (d, J = 7.9 Hz, 1H, <u>H1^{I}</u>), 4.53 (dd, J = 12.1, 2.8 Hz, 1H, <u>H6a^{I}</u>), 4.34 (ddd, J = 14.4, 4.5, 2.1 Hz, 1H, BCH=CHC<u>H$_2$</u>), 4.30 (dd, J = 12.1, 4.7 Hz, 1H, <u>H6b^{I}</u>), 4.22 (dd, J = 12.4, 4.5 Hz, 1H, <u>H6a^{II}</u>), 4.16 (ddd, J = 14.3, 5.2, 1.7 Hz, 1H, BCH=CHC<u>H$_2$</u>), 4.11 – 4.06 (m, 2H, <u>H6b^{II}</u> + <u>H5^{II}</u>), 4.03 (t, J = 9.2 Hz, 1H, <u>H4^{I}</u>), 3.92 (ddd, J = 9.6, 4.7, 2.8 Hz, 1H, <u>H5^{I}</u>), 2.10 (s, 3H, O<u>Ac</u>), 2.03 (s, 3H, O<u>Ac</u>), 2.01 (s, 3H, O<u>Ac</u>), 1.99 (s, 3H, O<u>Ac</u>), 1.98 (s, 3H, O<u>Ac</u>), 1.98 (s, 3H, O<u>Ac</u>), 1.95 (s, 3H, O<u>Ac</u>).

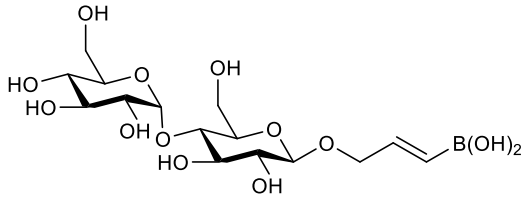
^{13}C NMR for 4.5 (126 MHz, Acetone-d_6, major anomer only) δ 170.2, 170.2, 170.0, 169.8, 169.5, 169.4, 169.2 (all O<u>Ac</u> C=O), 144.7 (BCH=<u>C</u>H), 125.0 (weak, B<u>C</u>H=CH), 99.3 (<u>C1^{I}</u>), 95.9 (<u>C1^{II}</u>), 75.1 (<u>C3^{I}</u>), 73.9 (<u>C4^{I}</u>), 72.0 (<u>C2^{I}</u> + <u>C5^{I}</u>), 70.5 (BCH=CH<u>C</u>H$_2$), 70.1 (<u>C2^{II}</u>), 69.2 (<u>C3^{II}</u>), 68.5 (<u>C5^{II}</u>), 68.2 (<u>C4^{II}</u>), 63.0 (<u>C6^{I}</u>), 61.7 (<u>C6^{II}</u>), 20.2, 19.9, 19.9, 19.8, 19.8, 19.7, 19.7 (all O<u>Ac</u> CH_3).

HRMS for 4.5 m/z (ESI+) found: 743.2177 (M+Na, calculated 743.2177)

IR for 4.5 (film) $\tilde{\nu}$ = 2956, 2917 (C-H), 1754 (C=O), 1640 (C=C) cm^{-1}

m.p. for 4.5 >300 °C ($CHCl_3$).

$[\alpha]_D$ for 4.5 = +83 (c = 1.0, $CHCl_3$)



4.1

Boronic acid **4.5** (100 mg, 0.14 mmol) was dissolved in anhydrous MeOH (2.0 ml) and NaOMe (8 mg, 0.032 ml of a 25 wt % solution in MeOH, 0.14 mmol) was added. The reaction mixture was stirred for 3 hours at room temperature, after which point TLC and MS indicated complete consumption of starting material. Pre-activated DOWEX®-H^+ was added until the pH was neutral, and the mixture stirred for 30 minutes. The DOWEX resin was removed by filtration and washed thoroughly with MeOH, and the eluate concentrated *in vacuo*. The crude mixture was redissolved in H_2O (2 ml) and filtered (0.2 μm syringe filter) to remove small insoluble impurities, before being concentrated *in vacuo* to give boronic acid **4.1** as a white solid (37 mg, 0.087 mmol, 62 %, 10:1 β:α). Anomeric ratio determined from integration of $H1^{I}_{\alpha}$ (δ 4.98, d, J = 3.9 Hz) and $H1^{I}_{\beta}$ (δ 4.52, d, J = 8.0 Hz)

^{1}H NMR (500 MHz, D_2O, major anomer only) δ 6.55 (dt, J = 18.3, 5.1 Hz, 1H, BCH=C<u>H</u>), 5.77 (dt, J = 18.2, 1.7 Hz, 1H, BC<u>H</u>=CH), 5.41 (d, J = 3.6 Hz, 1H, <u>H1^{II}</u>), 4.52 (d, J = 8.0 Hz, 1H, <u>H1^{I}</u>), 4.48 (ddd, J = 14.2, 4.8, 1.7 Hz, 1H, BCH=CHC<u>H$_2$</u>), 4.34 (ddd, J = 14.2, 5.4, 1.7 Hz, 1H, BCH=CHC<u>H$_2$</u>), 3.93 (dd, J = 12.2, 2.1 Hz, 1H, <u>H6a^{I}</u>), 3.85 (dd, J = 12.3, 2.2 Hz, 1H, <u>H6a^{II}</u>), 3.78 (dd, J = 9.6, 8.6 Hz, 1H, <u>H3^{I}</u>), 3.78 (dd, J = 12.3, 5.4 Hz, 1H, <u>H6b^{I}</u>), 3.77 (dd, J = 12.2, 5.2 Hz, 1H, <u>H6b^{II}</u>), 3.72 (ddd, J = 9.9, 5.3, 2.1 Hz, 1H, <u>H5^{II}</u>), 3.69 (dd, J = 9.9, 9.1 Hz, 1H, <u>H3^{II}</u>), 3.65 (dd, J = 9.8, 8.6 Hz, 1H, <u>H4^{I}</u>), 3.60 – 3.56 (m, <u>H5^{I}</u>), 3.59 (dd, J = 9.9, 4.0 Hz, 1H, <u>H2^{II}</u>), 3.42 (dd, J = 9.9, 9.1 Hz, 1H, <u>H4^{II}</u>), 3.35 (dd, J = 9.5, 8.0 Hz, 1H, <u>H2^{I}</u>).

^{13}C NMR (126 MHz, D_2O) δ 145.3 (BCH=<u>C</u>H), 124.1 (weak, B<u>C</u>H=CH), 101.2 (<u>C1^{I}</u>), 99.5 (<u>C1^{II}</u>), 76.7 (<u>C4^{I}</u>), 76.2 (<u>C3^{I}</u>), 74.5 (<u>C5^{I}</u>), 73.0 (<u>C2^{I}</u>), 72.8 (<u>C3^{II}</u>), 72.7 (<u>C5^{II}</u>), 71.6 (<u>C2^{II}</u>), 71.1 (BCH=CH<u>C</u>H$_2$), 69.3 (<u>C4^{II}</u>), 60.7 (<u>C6</u>), 60.4 (<u>C6^{II}</u>)

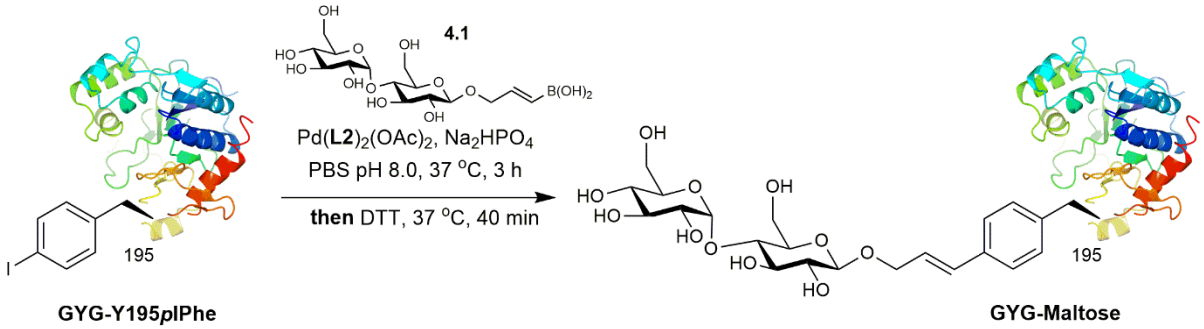
HRMS *m/z* (ESI+) found: 449.1347 (M+Na, calculated 449.1347)

IR (solid) $\tilde{\nu}$ = 3244 (O-H), 2926 (C-H), 1643 (C=C)

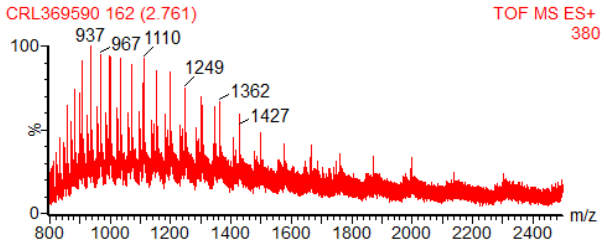
m.p. >300 °C (H_2O). **$[\alpha]_D$** = +77 (c = 0.3, H_2O)

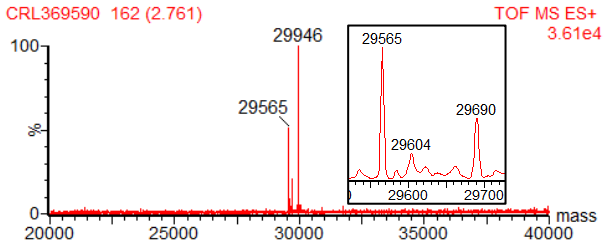
4.5.2 Suzuki-Miyaura Cross-coupling Optimisations

4.5.2.1 Initial Cross-couplings between GYG-Y195*p*IPhe and Maltose Boronic Acid 4.1



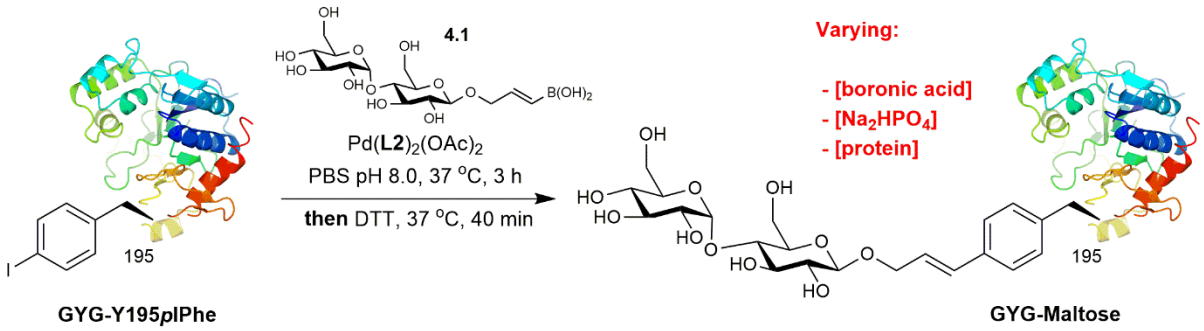
A stock solution of maltose boronic acid **4.1** was prepared by dissolving boronic acid (45 mg/ml) and Na_2HPO_4 (15.7 mg/ml) in H_2O. To non-his-tagged GYG-Y195*p*IPhe (15 µl of 0.86 mg/ml, 0.44 nmol) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was added 15 µl of this stock solution (1580 nmol, 3600 eq. boronic acid). $Pd(L2)_2(OAc)_2$ catalyst (2.0 µl of 10 mM solution, 0.6 mM concentration) was added. The reaction mixture was briefly vortexed, then incubated at 37 °C with shaking (600 rpm). After 3 hours, DTT was added to a final concentration of 11 mM (from a 115 mM solution in H_2O), the mixture shaken for a further 40 minutes, then the reaction was analysed directly by LC-MS.





M_{calc}: 29946
M_{obs}: 29946 (56 %)
29565 (de-iodination)
29691 (GYG-Y195*p*IPhe)
29604 (see **figure 4.6** and chapter 3 pp. 172-3)

4.5.2.2 Cross-coupling Optimisation Studies (table 4.1)

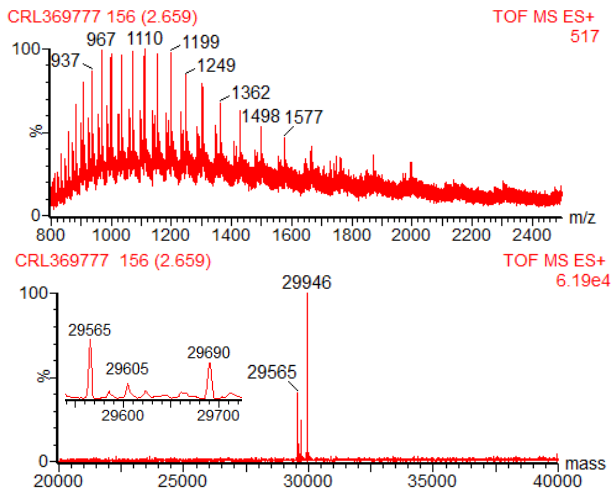


All reactions in this section were carried out in parallel, to minimise any differences between samples other than with regards to the factor(s) under investigation.

a) Protein Concentration

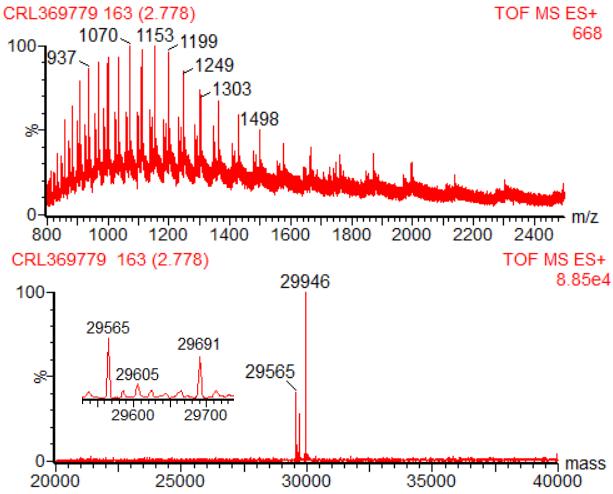
2 reactions were set up. A stock solution of maltose boronic acid **4.1** was prepared by dissolving boronic acid (45 mg/ml) and Na_2HPO_4 (15.7 mg/ml) in H_2O. To non-his-tagged GYG-Y195*p*IPhe (*either* 15 µl of 0.86 mg/ml, 0.44 nmol *or* 15 µl of 1.05 mg/ml, 0.53 nmol) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was added 15 µl of this stock solution (1580 nmol boronic acid), followed by $Pd(L2)_2(OAc)_2$ catalyst (2.0 µl of 10 mM solution, 0.6 mM concentration). Final protein concentrations were either 0.4 mg/ml or 0.5 mg/ml. The reaction mixtures were briefly vortexed, then incubated at 37 °C with shaking (600 rpm). After 3 hours, DTT was added to a final concentration of 11 mM (from a 115 mM solution in H_2O), the mixtures shaken for a further 40 minutes, then analysed by LC-MS.

0.4 mg/ml: 57 % conversion



M_{calc}: 29946
M_{obs}: 29946 (57 %)
29565 (de-iodination)
29605 (see **figure 4.6** and chapter 3 pp. 172-3)
29690 (GYG-Y195*p*IPhe)

0.5 mg/ml: 56 % conversion



M_{calc}: 29946
M_{obs}: 29946 (56 %)
29565 (de-iodination)
29605 (see **figure 4.6** and chapter 3 pp. 172-3)
29691 (GYG-Y195*p*IPhe)

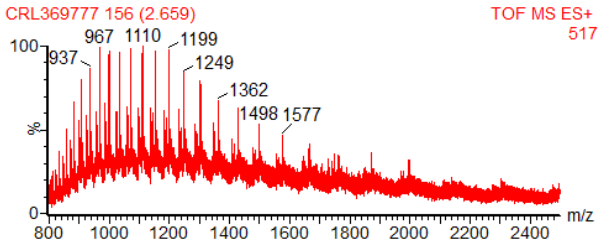
b) [Boronic Acid]

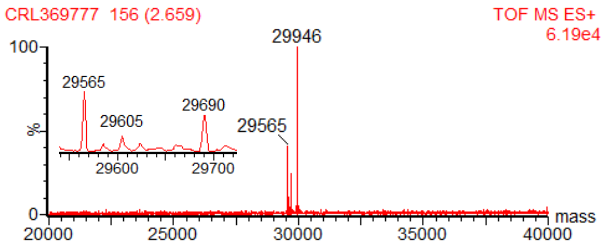
2 reactions were set up. A stock solution of maltose boronic acid **4.1** was prepared by dissolving boronic acid (45 mg/ml) and Na_2HPO_4 (15.7 mg/ml) in H_2O. To non-his-tagged GYG-Y195*p*IPhe (0.44 nmol) was added stock solution as indicated in the table below:

Reaction	GYG-Y195*p*IPhe	Boronic Acid Stock Solution
1	15 µl of 0.86 mg/ml, 0.44 nmol	15 µl, 1580 nmol, 3600 eq.
2	12 µl of 1.08 mg/ml, 0.44 nmol	18 µl, 1900 nmol, 4400 eq.

To each reaction was then added $Pd(L2)_2(OAc)_2$ catalyst (2.0 µl of 10 mM solution, 0.6 mM concentration). The reaction mixtures were briefly vortexed, then incubated at 37 °C with shaking (600 rpm). After 3 hours, DTT was added to a final concentration of 11 mM (from a 115 mM solution in H_2O), the mixtures shaken for a further 40 minutes, then analysed by LC-MS.

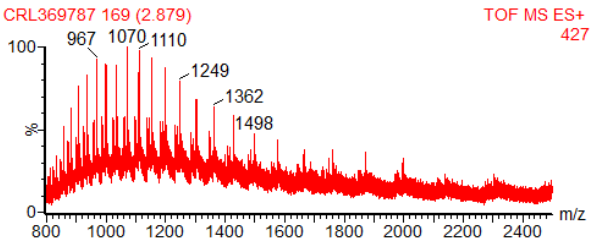
3600 eq.: 57 % conversion

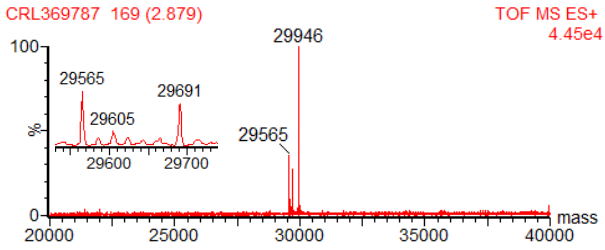




M_{calc}: 29946
M_{obs}: 29946 (57 %)
29565 (de-iodination)
29605 (see **figure 4.6** and chapter 3 pp. 172-3)
29690 (GYG-Y195*p*IPhe)

4400 eq.: 58 % conversion





M_{calc}: 29946
M_{obs}: 29946 (58 %)
29565 (de-iodination)
29605 (see **figure 4.6** and chapter 3 pp. 172-3)
29691 (GYG-Y195*p*IPhe)

c) [Na_2HPO_4]

2 stock solutions of maltose boronic acid **4.1** were prepared:

Stock solution 1: boronic acid (45 mg/ml) and Na_2HPO_4 (15.7 mg/ml) in H_2O.

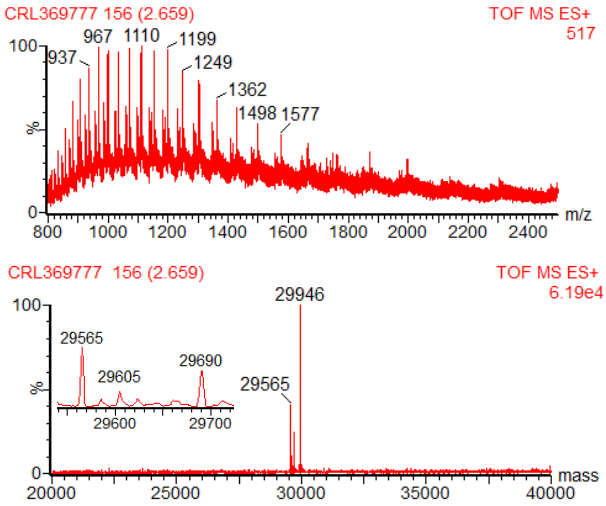
Stock solution 2: boronic acid (45 mg/ml) *only* in H_2O.

3 reactions were then set up by combining components in the amounts indicated in the table below:

Reaction	GYG-Y195*p*IPhe	Boronic Acid Stock Solution	Additional Na_2HPO_4
1	15 µl of 0.86 mg/ml, 0.44 nmol	15 µl of **stock solution 1**, 1580 nmol, 3600 eq.	None
2	15 µl of 0.86 mg/ml, 0.44 nmol	15 µl of **stock solution 2**, 1580 nmol, 3600 eq.	None
3	8.9 µl of 1.45 mg/ml, 0.44 nmol	15 µl of **stock solution 1**, 1580 nmol, 3600 eq.	6.1 µl of 15.7 mg/ml solution in H_2O.

To each reaction was then added $Pd(L2)_2(OAc)_2$ catalyst (2.0 µl of 10 mM solution, 0.6 mM concentration). The reaction mixtures were briefly vortexed, then incubated at 37 °C with shaking (600 rpm). After 3 hours, DTT was added to a final concentration of 11 mM (from a 115 mM solution in H_2O), the mixtures shaken for a further 40 minutes, then analysed by LC-MS.

Reference (reaction 1): 57 % conversion



M_{calc}: 29946

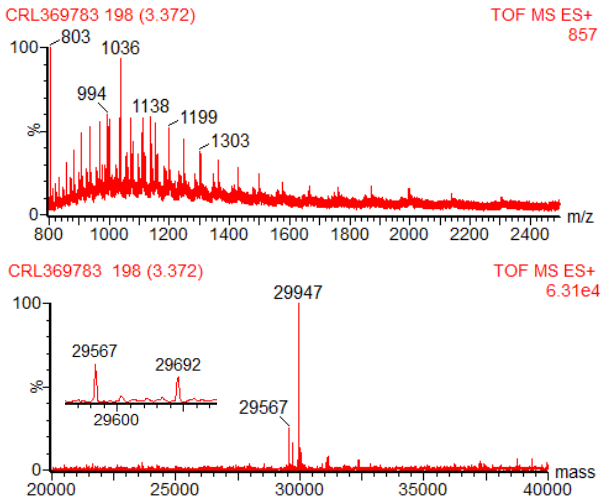
M_{obs}: 29946 (57 %)

29565 (de-iodination)

29605 (see **figure 4.6** and chapter 3 pp. 172-3)

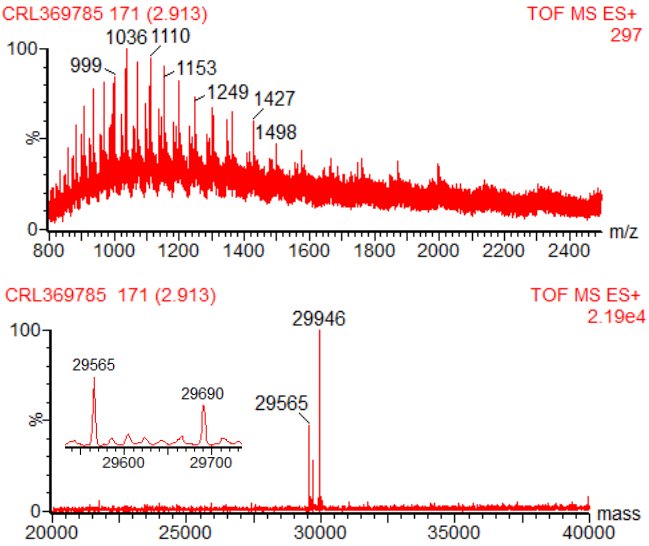
29690 (GYG-Y195*p*IPhe)

No Na_2HPO_4 (reaction 2) 70 % conversion



M_{calc}: 29946
M_{obs}: 29947 (70 %)
29567 (de-iodination)
29692 (GYG-Y195*p*IPhe)

Extra Na_2HPO_4 (reaction 3): 57 % conversion

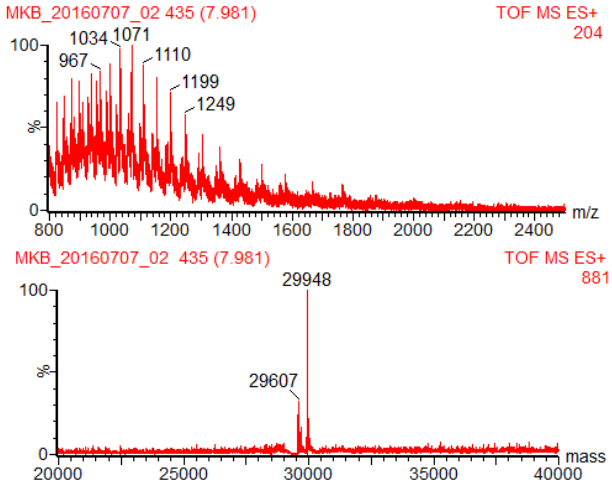


M_{calc}: 29946
M_{obs}: 29946 (57 %)
29565 (de-iodination)
29690 (GYG-Y195*p*IPhe)

4.5.2.3 Further Cross-coupling Studies: Effect of Reaction Time

A stock solution of maltose boronic acid **4.1** was prepared by dissolving boronic acid (45 mg/ml) in H_2O. To non-his-tagged GYG-Y195*p*IPhe (15 µl of 0.86 mg/ml, 0.44 nmol) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was added 15 µl of this stock solution (1580 nmol boronic acid, 3600 eq.), followed by $Pd(L2)_2(OAc)_2$ catalyst (2.0 µl of 10 mM solution, 0.6 mM concentration). Final protein concentrations were either 0.4 mg/ml or 0.5 mg/ml. The reaction mixtures were briefly vortexed, then incubated at 37 °C with shaking (600 rpm). After either 3 hours or 5 hours, an aliquot was removed and DTT was added to a final concentration of 7.5 mM (from a 115 mM solution in H_2O), the mixture shaken for a further 40 minutes, then analysed by LC-MS.

t = 3h



M_{calc}: 29946

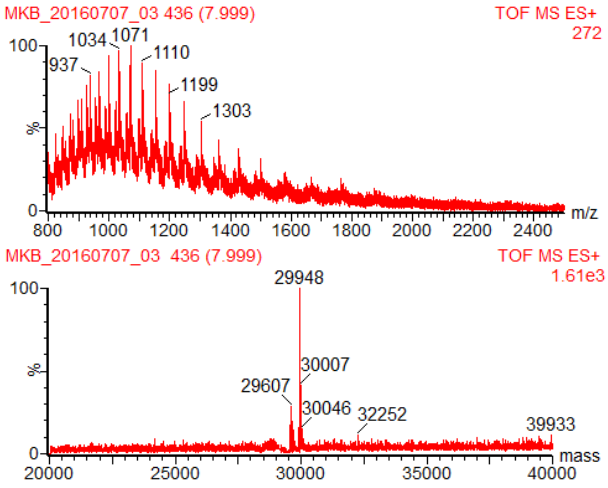
M_{obs}: 29948

29567 (de-iodination)

29607 (see **figure 4.6** and chapter 3 pp. 172-3)

29693 (GYG-Y195*p*IPhe)

t = 5h



M_{calc}: 29946

M_{obs}: 29948

29567 (de-iodination)

29607 (see **figure 4.6** and chapter 3 pp. 172-3)

29693 (GYG-Y195*p*IPhe)

30007 (unidentified side-product)

Whilst the proportion of product with respect to commonly occurring side-products (derived from de-iodination pathway or lack of reaction) increased slightly upon increased reaction time (58 % after 3 h, 62 % after 5 h), additional unidentified and not previously observed side-products appeared after 5 h.

4.5.3 General Protocol for Analysis of LC-MS Data for Glucosylation and Galactosylation Assays

This section explains how LC-MS data obtained from autoglucosylation/autogalactosylation assays of cross-coupled enzymes (see section 4.5.4 below) is utilised to generate the three distinct types of graphs throughout the text. **This is applicable to the following work in this chapter, and also in chapters 5 and 6.**

Deconvoluted LC-MS spectra from timepoints of autoglucosylation/autogalactosylation assays are analysed to determine: 1) the proportion of starting cross-coupled enzyme remaining (which serves as a means of comparing rates of initial catalytic steps); 2) the proportion of each glycoform present; 3) the average number of glucoses/galactoses attached to the cross-coupled enzyme. The general approach has been described previously (see ref. 7) – please refer to this for full details – however is briefly reviewed here. The analysis focusses solely on the glucosylation experienced by the cross-coupled enzyme, and therefore considers MS peaks arising from this and its glucosylated/galactosylated variants; cross-coupling side-products are necessarily not included in analysis, since they lack any capacity to undergo glucosylation/galactosylation.

1) "Glycoform Distributions" Graphs (figure 4.13)

For a given time point, the proportion of a given glycoform – as a fraction of the total glycoforms present – is obtained from the intensity of its corresponding peak in the deconvoluted LC-MS spectrum, relative to the combined intensities of all MS peaks arising from GYG glycoforms:

$$proportion\ of\ glycoform = \frac{MS\ peak\ intensity\ of\ glycoform}{Total\ MS\ peak\ intensities\ of\ all\ glycoforms}$$

2) "Rate of Initial Step" Graphs / % cross-coupled enzyme remaining (figures 4.8, 4.12)

The proportion of the cross-coupled enzyme remaining is calculated analogously to above (where the glycoform of interest is the cross-coupled enzyme itself). This is expressed as a percentage (i.e. multiplied by 100). I.e.:

$$\%\ cross-coupled\ enzyme\ glycoform\ remaining = \frac{MS\ peak\ intensity\ of\ cross-coupled\ glycoform}{Total\ MS\ peak\ intensities\ of\ all\ glycoforms}\ x\ 100$$

(Whilst it is possible to also calculate this from comparison of levels of the cross-coupled enzyme relative to the non-glucosylatable cross-coupling side-products, we preferred to adopt the approach above to ensure that the same method was utilised across all glycoforms, be they the cross-coupled enzyme or glycoforms generated from this enzymatically).

3) "Overall Glucosylation / Galactosylation Profile" Graphs (figures 4.8, 4.12, 4.14)

The average number of glucoses / galactoses attached (per enzyme molecule) is calculated as follows (see ref. 7 for full description). The proportion of each glycoform (calculated via "1" above, as a value between 0 and 1, i.e. 1 = 100 % of the mixture, e.g. GYG-Glucose at t = 0 s) is multiplied by the number of carbohydrate residues associated with it. E.g. GYG-Glucose = 1, GYG-Maltose = 2. The values obtained for all glycoforms are then summed. This effectively corresponds to a weighted average (in which "number of glc/gal attached" is the weight), and yields the average number of monosaccharide (Glc / Gal) residues attached. This may be represented by the formula below, in which "i" refers to the number of monosaccharide (Glc + Gal) residues associated with a given glycoform:

$$Average\ no.of\ glc/gal\ attached = \sum_{i=1}^{n} i \times (MS\ intensity\ of\ glycoform)_i$$

(Proportion of glycoform obtained from the equation shown in "1" above)

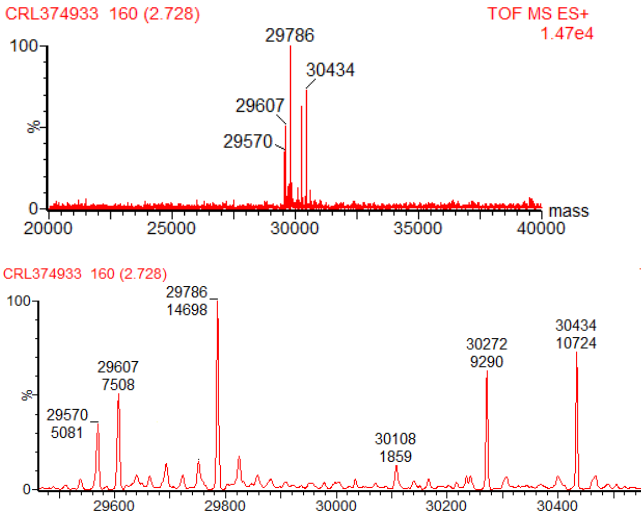
Note that the "average number of glc/gal attached" includes the carbohydrate residues ligated through cross-coupling. Hence, before any assay, this number is 1 for GYG-Glucose, and 2 for GYG-Maltose.

4.5.3.1 Lack of Effect of Number of Carbohydrate Residues Attached on MS Ionisation

As noted in the text (p. 239), an assumption in this LC-MS based protocol for glucosylation/galactosylation analysis is that each glycoform ionises on MS with comparable intensity; this has previously been validated directly and through comparison to other methods.[7,25] Experimentally, we did not notice significant differences in ionisation intensity of glycoforms between earlier time points (i.e shorter oligosaccharides) and later time points (i.e. longer oligosaccharides), as judged by comparing total MS intensities of all glycosylated enzyme to MS intensities of catalytically inactive cross-coupling side-products. The latter, not being substrates for glucosylation or galactosylation, are not consumed during autogalactosylation or

autoglucosylation assays, thus serve as an "internal standard" in this sense. Below is a representative example supporting this statement.

GYG-Glucose, t = 40 s (predominant glycoforms = 4-5 glucoses attached)



M_{calc}: 29784 + 162n

M_{obs}: 29786 + 162n (n = 0, 2, 3, 4, 5)

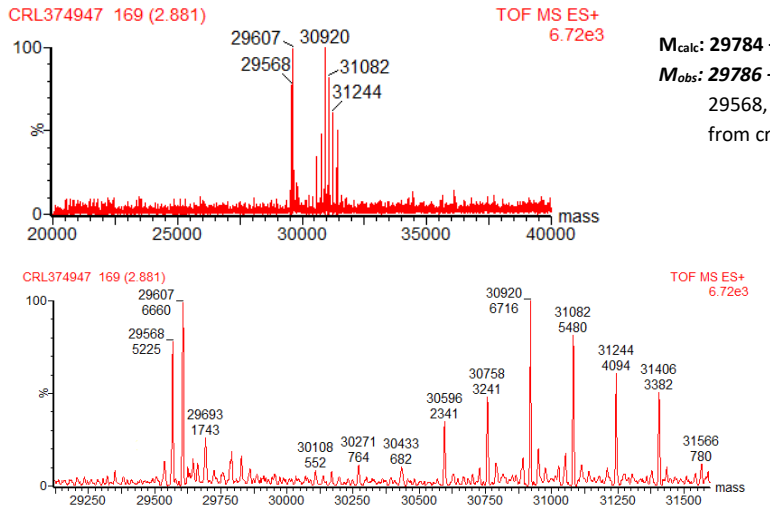
29568, 29607, 29693 = inactive side-products from cross-coupling (see **figure 4.6** and chapter 3 pp. 172-3)

Lower numbers = intensity

Upper numbers = M_w

Ratio of active glycoforms: inactive side products = 72:28

GYG-Glucose, t = 600 s (predominant glycoforms = 6-11 glucoses attached



M_{calc}: 29784 + 162n

M_{obs}: 29786 + 162n (n =2-11)

29568, 29607, 29693 = inactive side-products from cross-coupling (see **figure 4.6**)

Lower numbers = intensity

Upper numbers = M_w

Ratio of active glycoforms: inactive side products = 70:30

Thus, as judged relative to the non-glucosylatable side products, which are not consumed, the abundance of glycoforms is essentially constant irrespective of whether there are few (top) or many (lower) monosaccharide residues attached.

4.5.4 Timecourse Autoglucosylation and Autogalactosylation Assays of GYG-Glucose and GYG-Maltose

4.5.4.1 General Protocol for Autoglucosylation and Autogalactosylation

Generation of Cross-coupled Enzyme

A stock solution of boronic acid (either glucose boronic acid **3.1** or maltose boronic acid **4.1**) was prepared in H_2O (for synthesis of **3.1**, please see chapter 3). Non-his-tagged GYG-Y195*p*IPhe in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was combined with boronic acid added from stock solution, and $Pd(L2)_2(OAc)_2$ catalyst (0.6 mM final concentration, added from 10 mM stock solution). The reaction mixture was briefly vortexed, then incubated at 37 °C with shaking (600 rpm). After 3 hours, DTT was added to a final concentration of 7.5 mM (from a 115 mM solution in H_2O), the mixture shaken for a further 40 minutes, then buffer exchanged into pH 7.5 25 mM Na-HEPES by PD SpinTrap G-25. 100 µl of the sample was further desalted by dialysis against the same buffer (10000 MWCO, 4 °C, 2 h then 2 h then overnight, 3 x 300 ml buffer), using a Slide-A-Lyzer™ MINI dialysis device.

Enzymatic Activity Assay

Protein thus obtained was adjusted to a concentration of 0.2 mg/ml (Vivaspin 500, 10000 MWCO (Sartorius Stedem) were used for concentration if required). Protein was incubated with $MnCl_2$ and then either UDP-Glucose disodium salt (autoglucosylation) or UDP-Galactose disodium salt (autogalactosylation), all at final concentrations of 1 mM (added from freshly prepared 10 mM stock solutions in H_2O). The protein was incubated at 37 °C with shaking (500 rpm). After the desired time (indicated in the following sections), aliquots were quenched through the addition of Na-EDTA to a final concentration of 20 mM, and immediately placed on ice.

Quenched aliquots were analysed by LC-MS as described in more detail in section 4.5.3, in order to determine average numbers of glucoses/galactoses attached (including the initial chemically ligated glucose/maltose moiety) and the proportion of starting cross-coupled enzyme remaining at each time point (the latter serves as a means of monitoring the rate of the initial glucosylation or galactosylation step).

For sections 4.5.4.2 through 4.5.4.5 inclusive, this protocol was carried out at least in independent triplicate. As stated in the text, graphs show mean averages ± s.d. Section 4.5.4.6 explicitly compares only one such replicate each from sections 4.5.4.2 and 4.5.4.3. All LC-MS data was analysed as described in section 4.5.3.

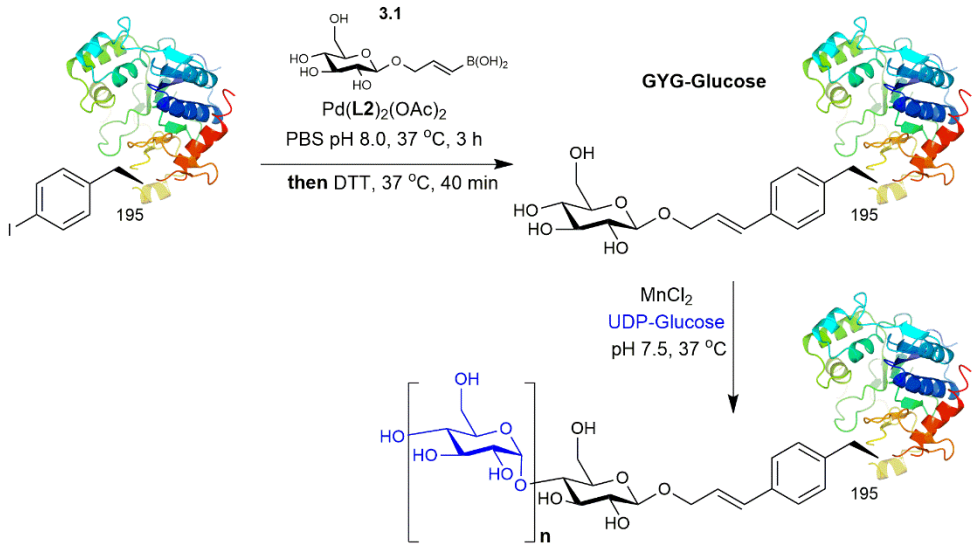
Additional Note: Explanation of formatting in the following sections (4.5.4.2 through 4.5.4.5):

In the following sections (4.5.4.2 through 4.5.4.5), autoglucosylation and autogalactosylation analyses of GYG-Glucose and GYG-Maltose are presented. These are formatted as follows:

1) The experimental protocol is outlined (including representative MS of cross-coupling sample).
2) Representative LC-MS data from kinetic assays are shown.
3) Graphical data (that shown in the main text), determined from LC-MS data across all replicates and timepoints via the methods outlined in section 4.5.3, is shown.
4) A detailed, "close-up" view of a representative LC-MS spectrum from glucosylation or galactosylation assay is provided, to illustrate that the non-glucosylated side-products do indeed correspond to those generated in cross-coupling.

Note that section 4.5.4.6 is simply a comparison of graphical data for one replicate each from sections 4.5.4.2 and 4.5.4.3, hence the formatting is not as above.

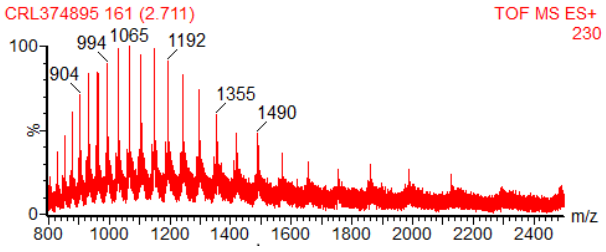
4.5.4.2 Generation and Autoglucosylation Analysis of GYG-Glucose

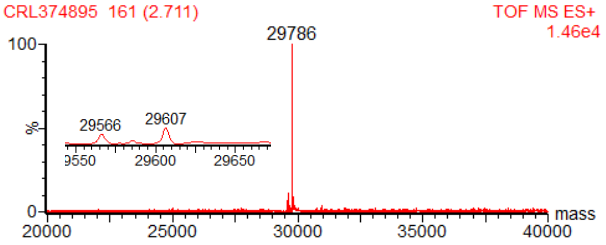


Carried out according to general protocol (section 4.5.4.1) using the following reagent amounts:

- GYG-Y195*p*IPhe: 50 µl (0.85 mg/ml), 1.4 nmol
- Boronic acid **3.1**: 48 µl of a 27 mg/ml stock solution prepared in H_2O, 4900 nmol, 3400 eq.
- $Pd(L2)_2(OAc)_2$: 6.3 µl of 10 mM stock solution, 63 nmol, 44 eq., 0.6 mM final concentration.

Example LC-MS after cross-coupling (85 % conversion):

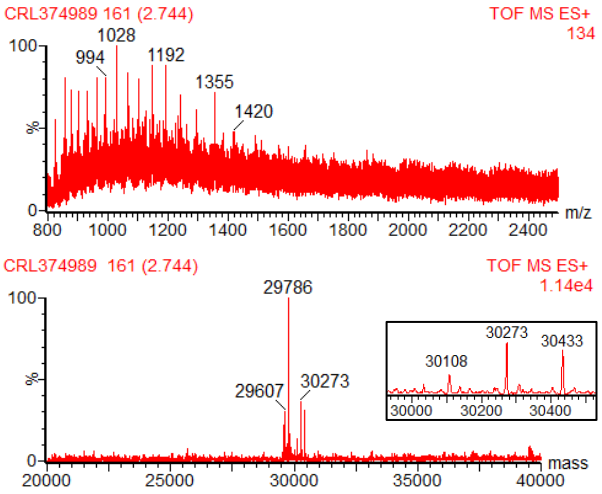




M_{calc}: 29784
M_{obs}: 29786 (85 %)
29566 (de-iodination)
29607 (see **figure 4.6** and chapter 3 pp. 172-3)

UDP-glucose was utilised as the donor sugar nucleotide during enzymatic activity assay. Aliquots were quenched and analysed after the following times of autoglucosylation assay: 20 s, 40 s, 80 s, 120 s, 210 s, 300 s, 600 s, 900 s. Representative LC-MS spectra and overall graphical data are shown below.

Representative LC-MS Spectra from glucosylation assay for kinetic analysis: t = 20 s

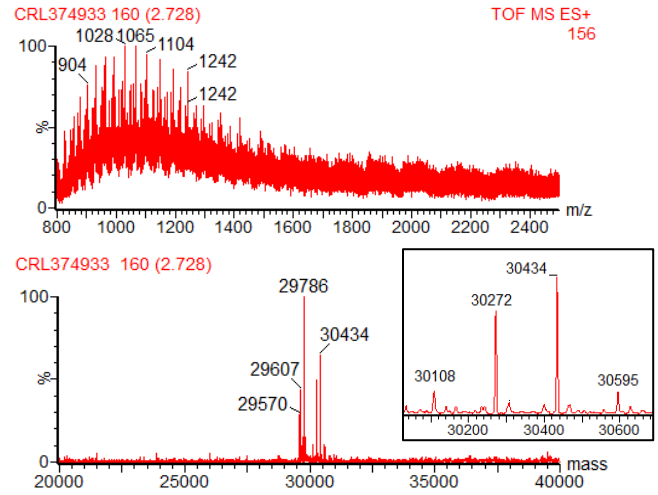


M_{calc}: 29784 + 162n

M_{obs}: 29786 + 162n (n = 0, 2, 3, 4)

+ inactive side-products from cross-coupling (29569, 29607: see "LC-MS after cross-coupling" above)

t = 40 s

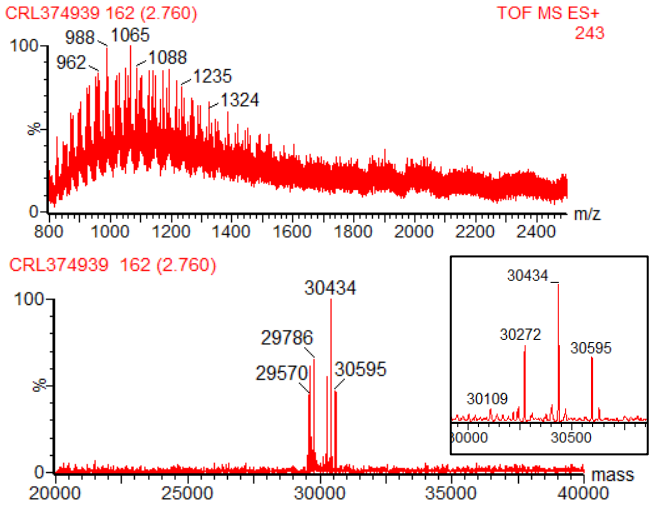


M_{calc}: 29784 + 162n

M_{obs}: 29786 + 162n (n = 0, 2-5)

+ inactive side-products from cross-coupling (29570, 29607: see "LC-MS after cross-coupling" above)

t = 80 s

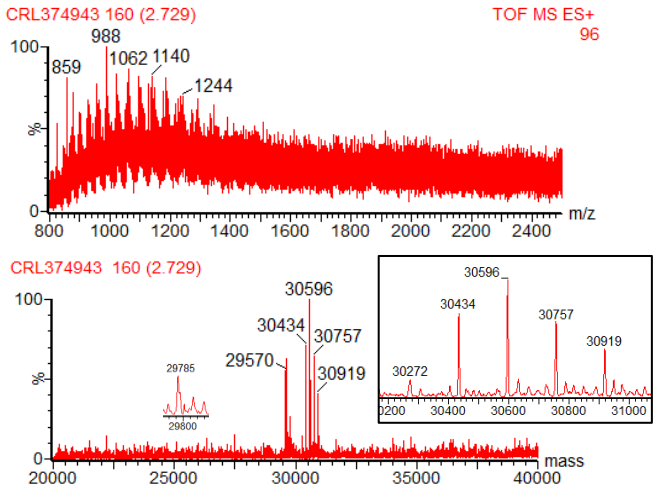


M_{calc}: 29784 + 162n

M_{obs}: 29786 + 162n (n = 0, 2-5)

+ inactive side-products from cross-coupling (29570, 29608: see "LC-MS after cross-coupling" above)

t = 210 s

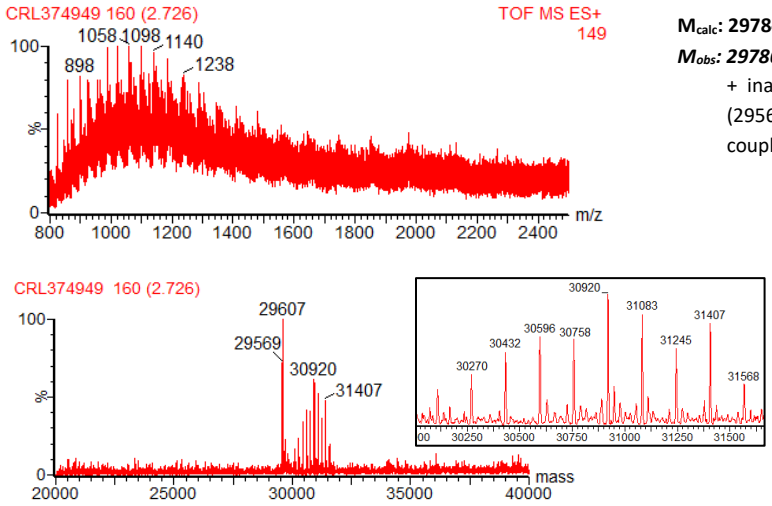


M_{calc}: 29784 + 162n

M_{obs}: 29785 + 162n (n = 0, 3-7)

+ inactive side-products from cross-coupling (29570, 29608: see "LC-MS after cross-coupling" above)

t = 900 s



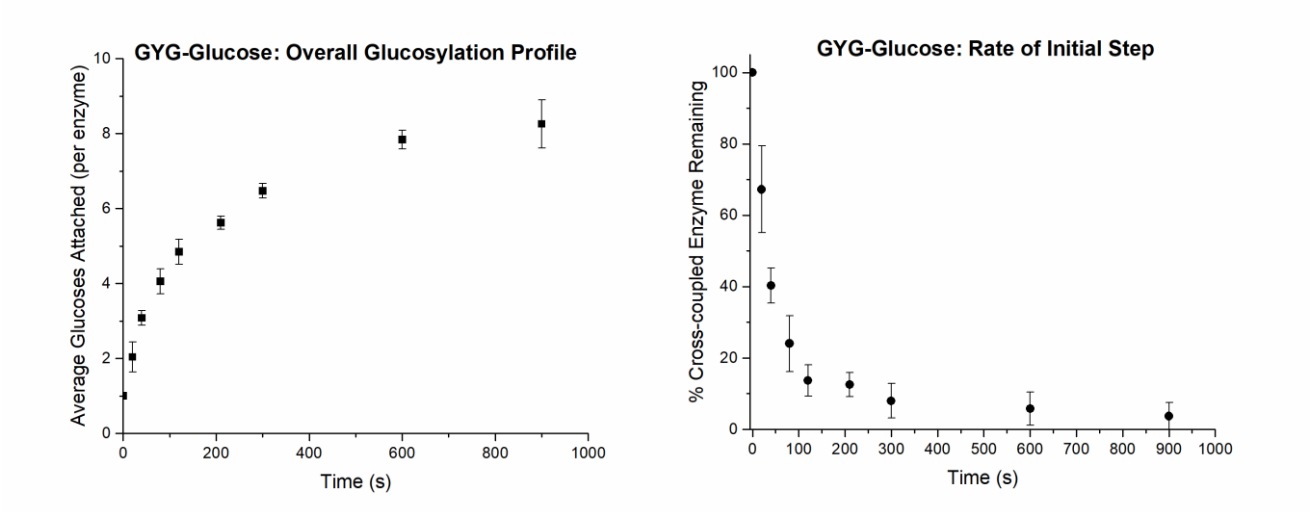
M_{calc}: 29784 + 162n

M_{obs}: 29786 + 162n (n = 3-11)

+ inactive side-products from cross-coupling (29569, 29607: see "LC-MS after cross-coupling" above)

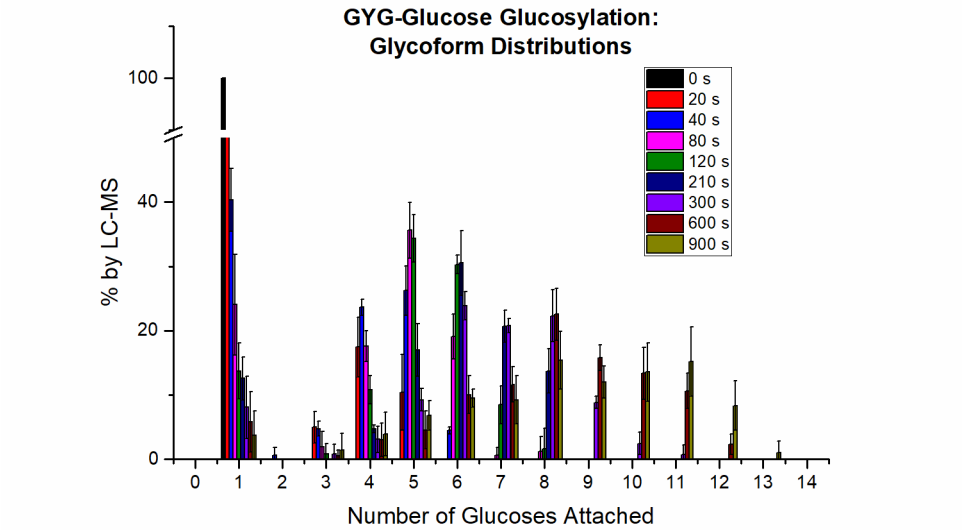
Graphical data from LC-MS from all timepoints, across all replicates (figure 4.8):

The methods for generation of this data from LC-MS spectra such as those shown above is described in section 4.5.3.



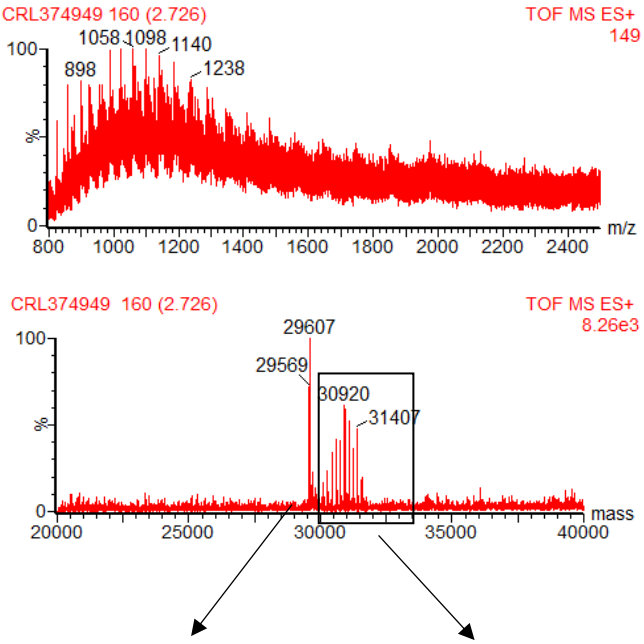
Extended Analysis of Glycoform Distributions (figure 4.13):

The method for generation of this data from LC-MS spectra such as those shown above is described in section 4.5.3.

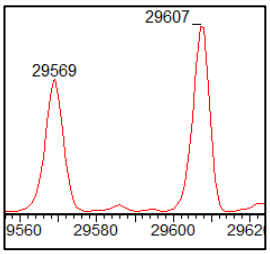


Detailed view of a representative LC-MS spectrum after glucosylation:

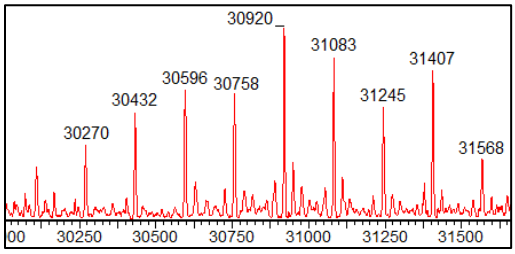
The inability of cross-coupling side-products to be glucosylated was demonstrated in chapter 3 for GYG-Glucose, which is the enzyme assayed here hence the side-products in this chapter are identical. Below is a detailed view of one of the spectra above (t = 900 s), illustrating that the catalytically inactive species visible in the above LC-MS spectra do indeed represent these cross-coupling side-products (they are the same for all time points).



Inactive Cross-Coupling Side-Products



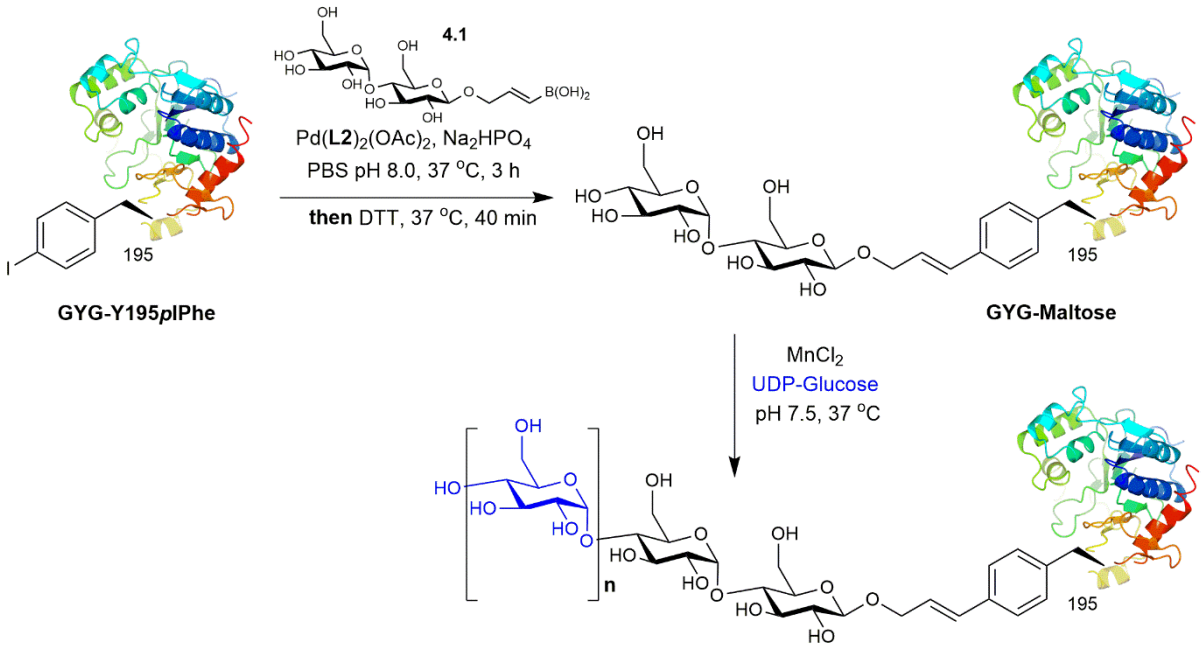
Catalytically Active Species



29569: de-iodinated
29607: see **figure 4.6** and chapter 3 pp. 172-3

30270-31568 = glucosylated GYG-Glucose (= 29785 + 162n, n = 3-11)

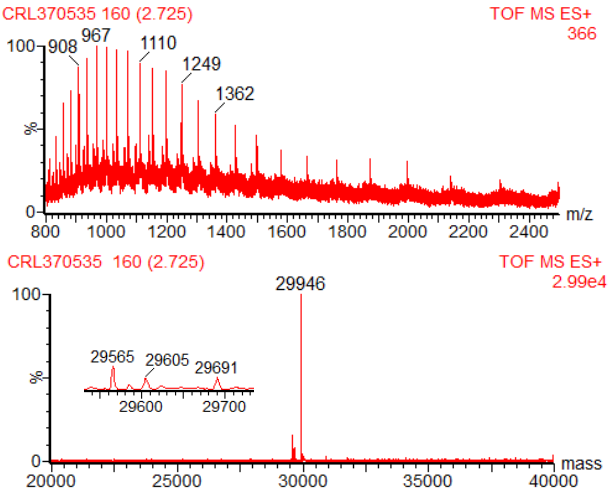
4.5.4.3 Generation and Autoglucosylation Analysis of GYG-Maltose



Carried out according to general protocol (section 4.5.4.1) using the following reagent amounts:

- GYG-Y195*p*IPhe: 48 µl (0.86 mg/ml), 1.4 nmol
- Boronic acid **4.1**: 48 µl of a 45 mg/ml stock solution prepared in H_2O, 5070 nmol, 3600 eq.
- $Pd(L2)_2(OAc)_2$: 6.2 µl of 10 mM stock solution, 62 nmol, 44 eq., 0.6 mM final concentration.

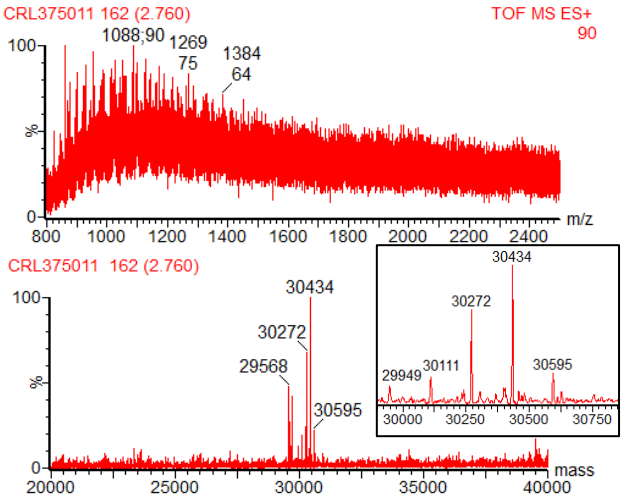
Example LC-MS after cross-coupling (76 % conversion):



M_{calc}: 29946
M_{obs}: 29946 (76 %)
29565 (de-iodination)
29605 (see **figure 4.6** and chapter 3 pp. 172-3)
29691 (GYG-Y195*p*IPhe)

UDP-glucose was utilised as the donor sugar nucleotide during enzymatic activity assay. Aliquots were quenched and analysed after the following times of autoglucosylation assay: 20 s, 40 s, 80 s, 120 s, 210 s, 300 s, 600 s, 900 s. Representative LC-MS spectra and overall graphical data are shown below.

Representative LC-MS spectra from glucosylation assay for kinetic analysis: t = 20 s

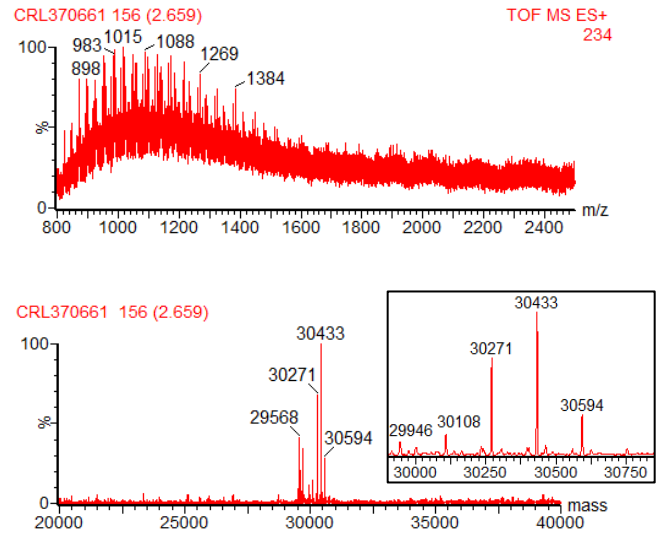


M_{calc}: 29946 + 162n

M_{obs}: 29949 + 162n (n = 0 – 4, mostly 2 and 3)

+ inactive side-products from cross-coupling (29568, 29607, 29693: see "LC-MS after cross-coupling" above)

t = 40 s

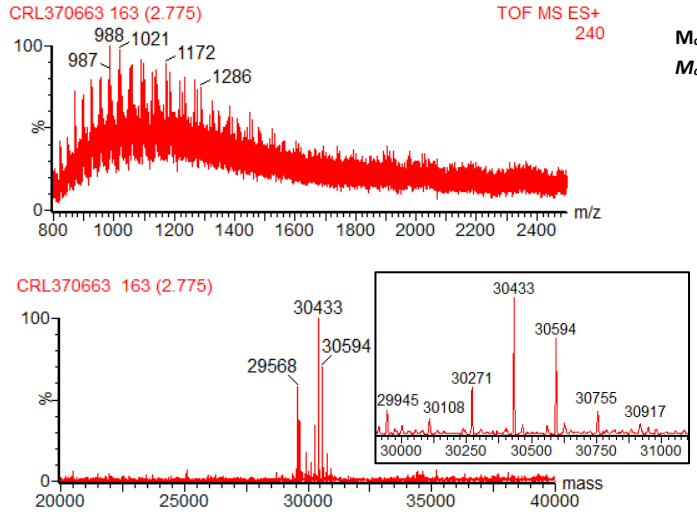


M_{calc}: 29946 + 162n

M_{obs}: 29946 + 162n (n = 0 – 4, mostly 2 and 3)

+ inactive side-products from cross-coupling (29568, 29606, 29691: see "LC-MS after cross-coupling" above)

t = 80 s

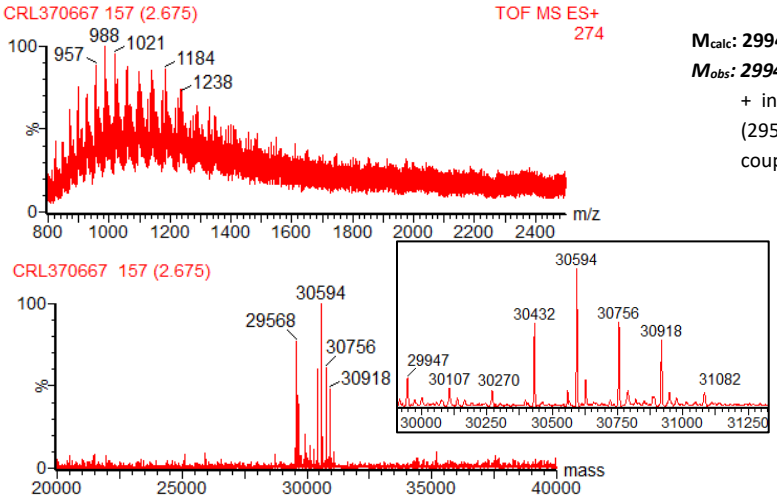


M_{calc}: 29946 + 162n

M_{obs}: 29945 + 162n (n = 0 – 6)

+ inactive side-products from cross-coupling (29568, 29606, 29691: see "LC-MS after cross-coupling" above)

t = 210 s

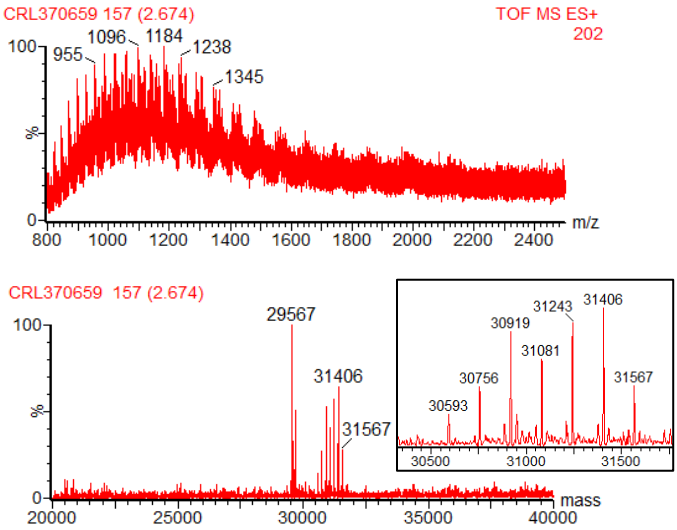


M_{calc}: 29946 + 162n

M_{obs}: 29947 + 162n (n = 0-7, mostly 3-6)

+ inactive side-products from cross-coupling (29568, 29606, 29691: see "LC-MS after cross-coupling" above)

t = 900 s



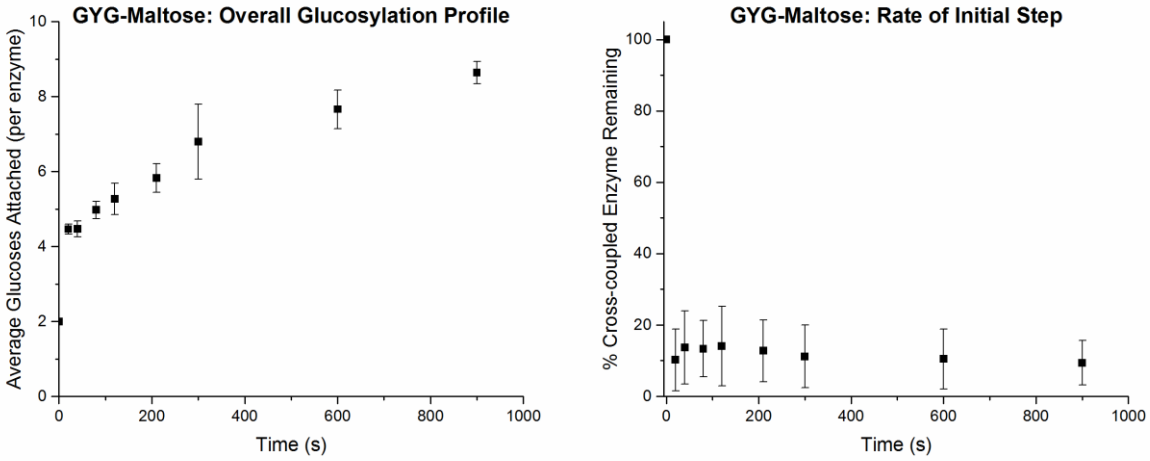
M_{calc}: 29946 + 162n

M_{obs}: 29948 + 162n (n = 4-10)

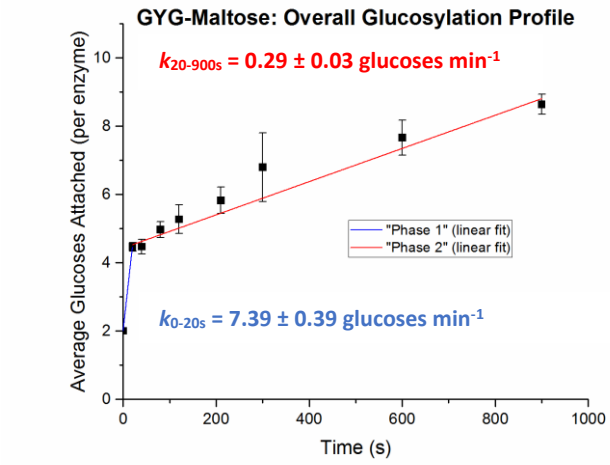
+ inactive side-products from cross-coupling (29567, 29606, 29691: see "LC-MS after cross-coupling" above)

Graphical data from LC-MS from all timepoints, across all replicates (figure 4.8):

The methods for generation of this data from LC-MS spectra such as those shown above is described in section 4.5.3.

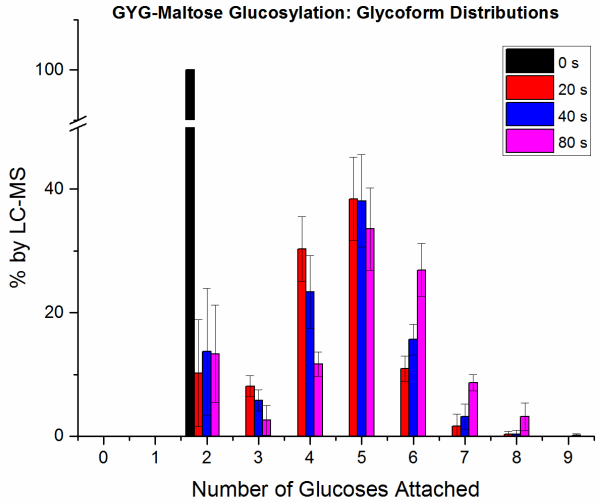


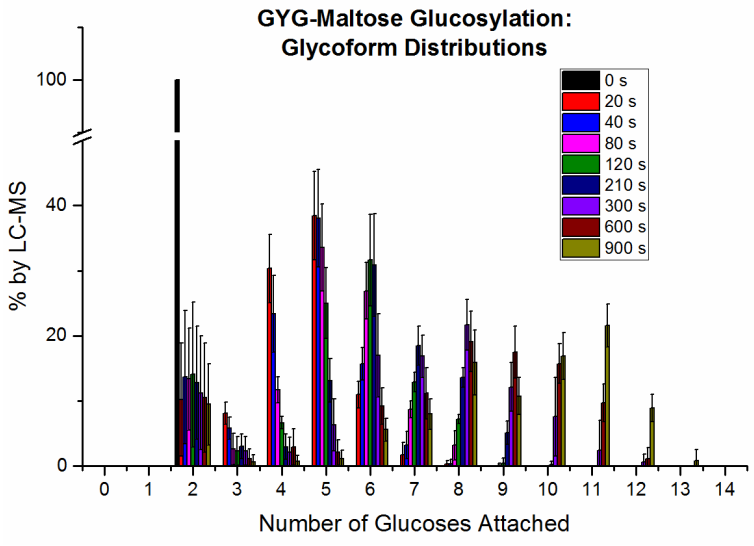
The overall glucosylation profile is biphasic (rapid → fast); these phases are illustrated below. The first phase involves only two timepoints (0 s, 20 s) and thus the rate simply reflects the number of glucoses added during this time. The second phase is approximately linear, and the rate shown is obtained from linear regression analysis (OriginPro 2015 software). It should be emphasised that the key conclusion here is not the precise numbers of the rates, but the significant difference between them (25-fold).



Extended analysis of glycoform distributions (see section 4.5.3 for methods of generation from LC-MS spectra) (figure 4.13):

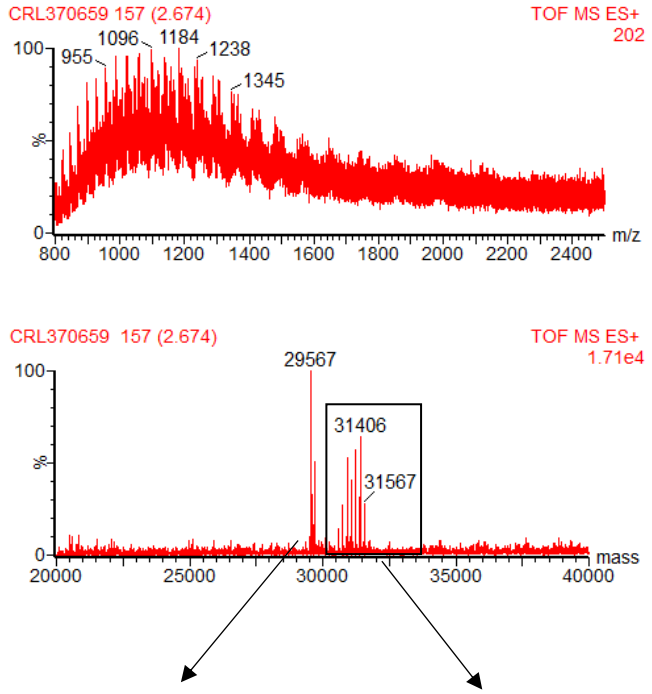
In section 4.3.2 (**figure 4.9**) it was noted that the transition from rapid to slow phases for GYG-Maltose glucosylation may be further illustrated via analysis of glycoform distributions at early time points. LC-MS snapshots from one run were illustrated to show that, whilst significant increase in glucosylation occurs during the first 20 seconds, the glycoform distribution change immediately after (40 s, 80 s) is limited. It was stated that this was reproducible; this is illustrated graphically below, showing average abundances of glycoforms across all replicates, as measured by LC-MS intensity (see section 4.5.3.2, the replicates are the same as utilised to generate the graphs above). As stated in section 4.5.3.1, error bars are s.d. of mean averages. For completeness, the analogous graph across all timepoints (shown in **figure 4.13**) is also shown.



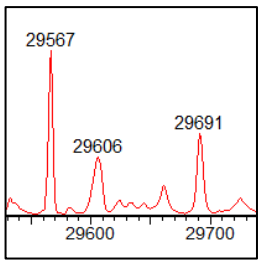


Detailed view of a representative LC-MS spectrum after glucosylation:

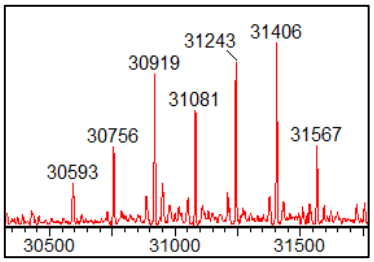
The inability of cross-coupling side-products to be glucosylated was demonstrated in chapter 3 for GYG-Glucose, and the side-products observed for GYG-Maltose here were identical (see section 4.2.2, **figure 4.6**). Below is a detailed view of one of the spectra above (t = 15 min), illustrating that the catalytically inactive species visible in the above LC-MS spectra do indeed represent these cross-coupling side-products (they are identical for all time points).



Inactive Cross-Coupling Side-Products



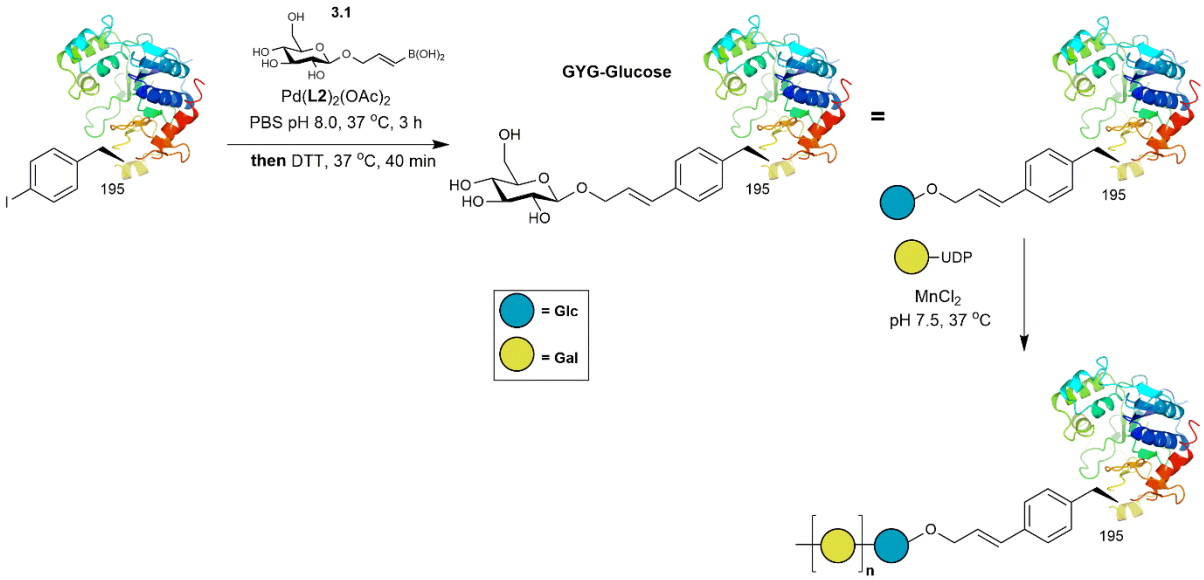
Catalytically Active Species



29567: de-iodinated
29606: see **figure 4.6** and chapter 3 pp. 172-3
29691: GYG-Y195*p*IPhe

30593-31567 = glucosylated GYG-Maltose (= 29948 + 162n, n = 4-10)

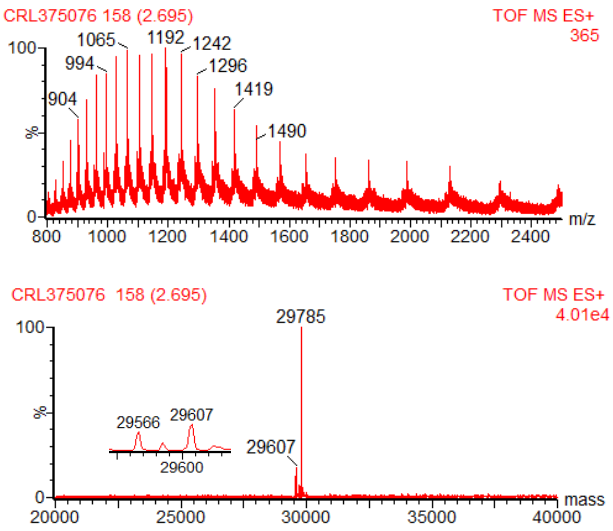
4.5.4.4 Generation and Autogalactosylation Analysis of GYG-Glucose



Carried out according to general protocol (section 4.5.4.1) using the following reagent amounts:

- GYG-Y195*p*IPhe: 50 µl (0.85 mg/ml), 1.4 nmol
- Boronic acid **3.1**: 48 µl of a 27 mg/ml stock solution prepared in H_2O, 4900 nmol, 3400 eq.
- $Pd(L2)_2(OAc)_2$: 6.3 µl of 10 mM stock solution, 63 nmol, 44 eq., 0.6 mM final concentration.

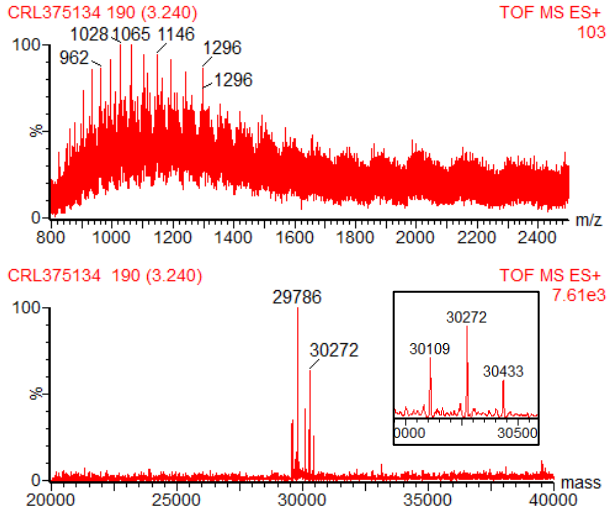
Example LC-MS after cross-coupling (79 % conversion):



M_{calc}: 29784
M_{obs}: 29785 (79 %)
29566 (de-iodination)
29607 (see **figure 4.6** and chapter 3 pp. 172-3)

UDP-galactose was utilised as the donor sugar nucleotide during enzymatic activity assay. Aliquots were quenched and analysed after the following times of autogalactosylation assay: 5 min, 10 min, 20 min, 30 min, 60 min, 120 min. Representative LC-MS spectra and overall graphical data are shown below.

Representative LC-MS spectra from galactosylation assay: t = 5 min

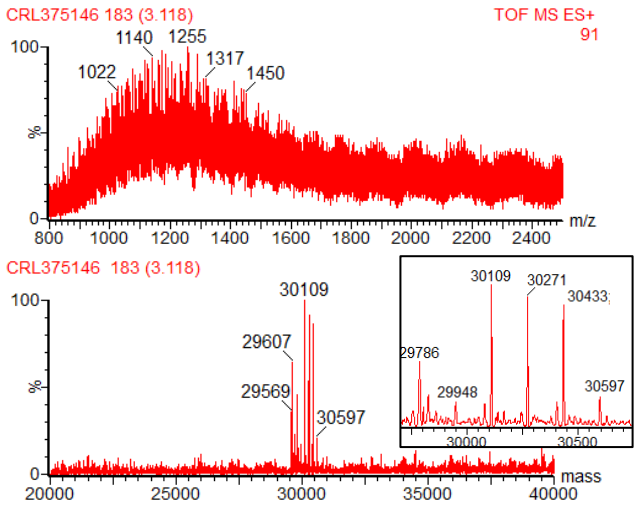


M_{calc}: 29784 + 162n

M_{obs}: 29786 + 162n (n = 0, 2, 3, 4)

+ inactive side-products from cross-coupling (29570, 29608: see "LC-MS after cross-coupling" above)

t = 30 min

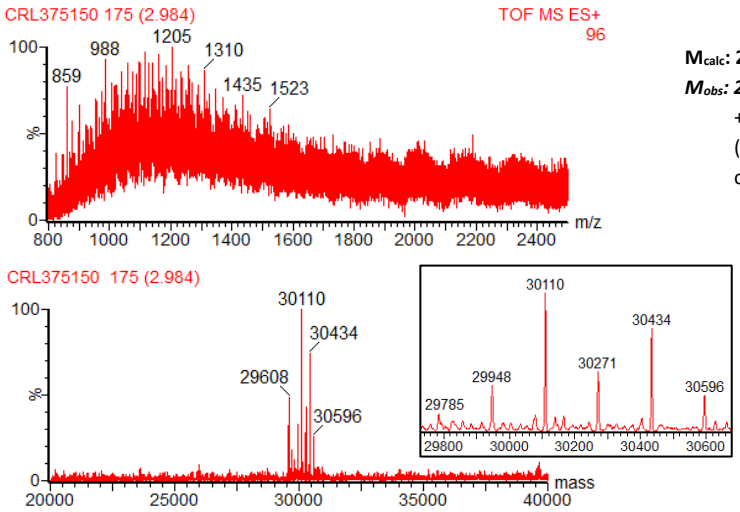


M_{calc}: 29784 + 162n

M_{obs}: 29786 + 162n (n = 0-5)

+ inactive side-products from cross-coupling (29569, 29607: see "LC-MS after cross-coupling" above)

t = 120 min



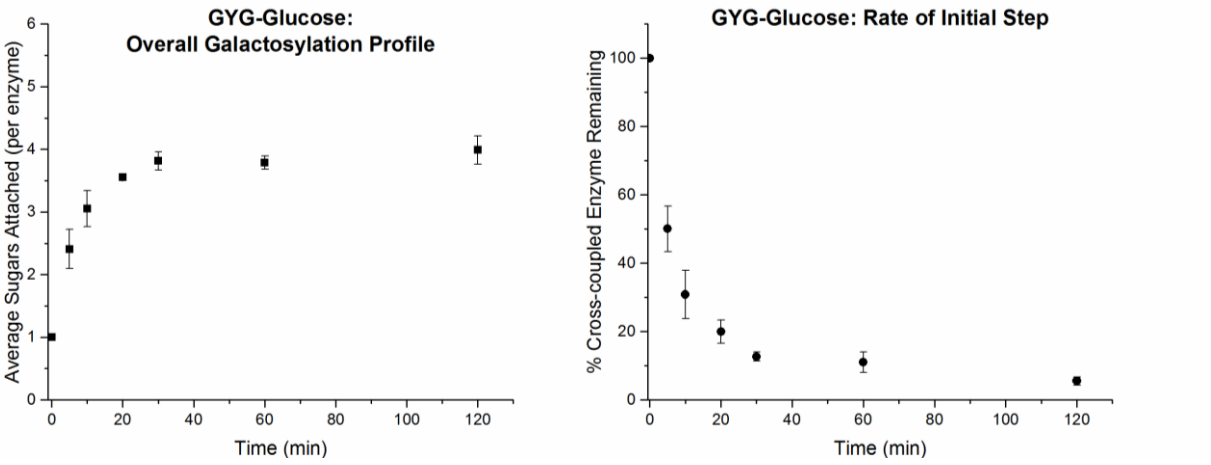
M_{calc}: 29784 + 162n

M_{obs}: 29785 + 162n (n = 0-5)

+ inactive side-products from cross-coupling (29570, 29608: see "LC-MS after cross-coupling" above)

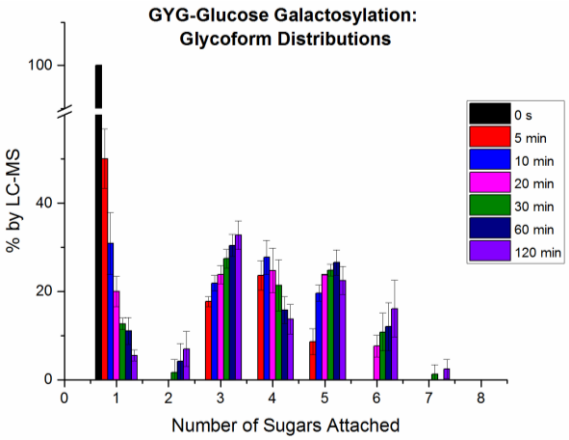
Graphical data from LC-MS from all timepoints, across all replicates (figure 4.12):

The methods for generation of this data from LC-MS spectra such as those shown above is described in section 4.5.3.



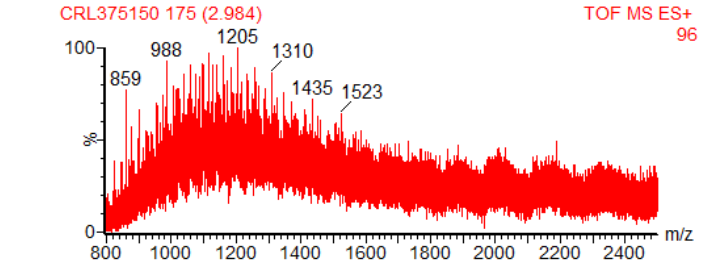
Extended analysis of glycoform distributions (figure 4.13):

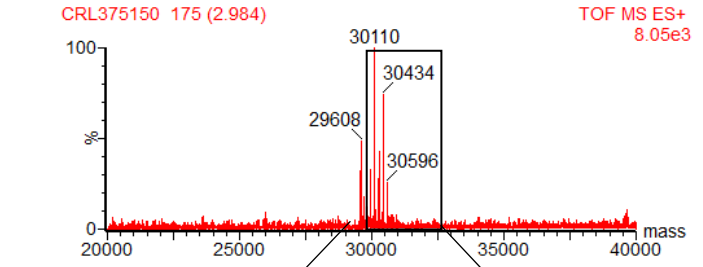
The methods for generation of this data from LC-MS spectra such as those shown above is described in section 4.5.3.



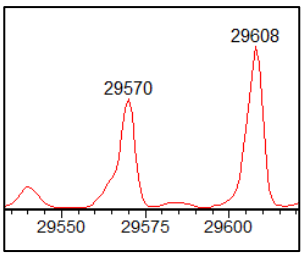
Detailed view of a representative LC-MS spectrum after galactosylation:

Below is a detailed view of one of the spectra above (t = 120 min), providing a zoomed-in view of not only the catalytically active, galactosylated species but also the inactive cross-coupling side-products (in order to confirm that these peaks do indeed correspond to these side-products; they are the same throughout all time points).

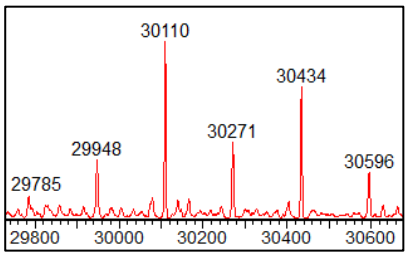




Inactive Cross-Coupling Side-Products



Catalytically Active Species

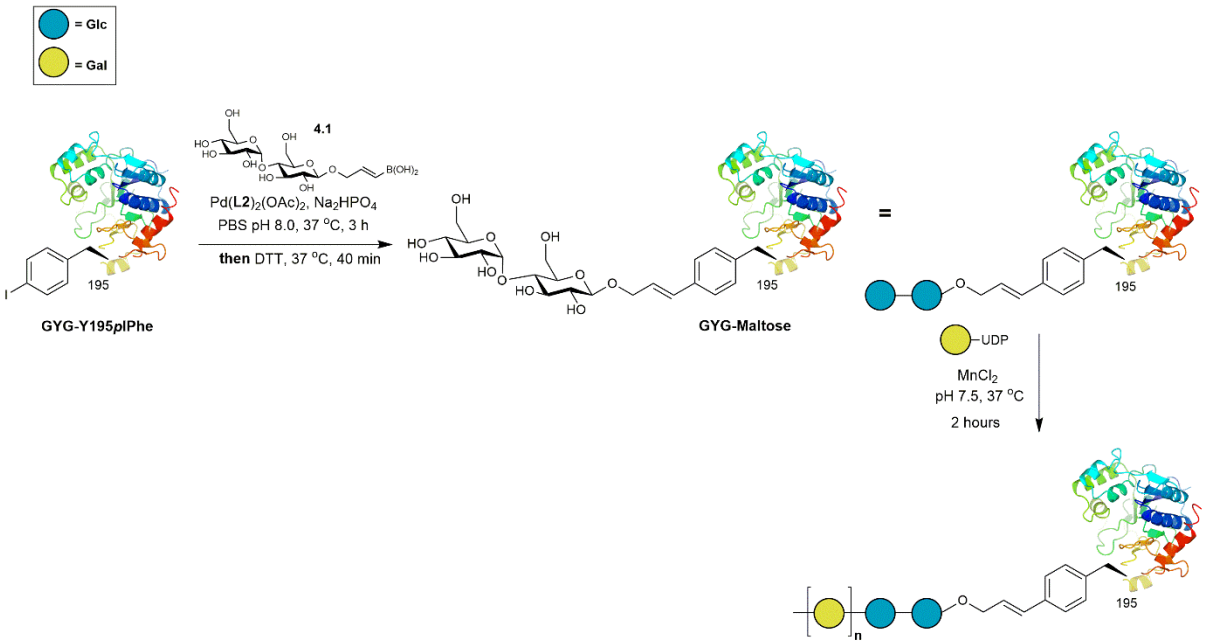


29570: de-iodinated
29608: see **figure 4.6** and chapter 3 pp. 172-3

29948 - 30596 = galactosylated GYG-Glucose (= 29785 + 162n, n = 1-5)

29785 = GYG-Glucose (n = 0)

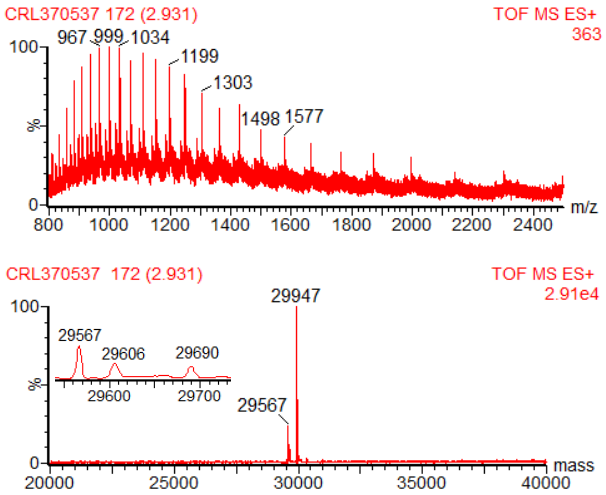
4.5.4.5 Generation and Autogalactosylation Analysis of GYG-Maltose



Carried out according to general protocol (section 4.5.4.1) using the following reagent amounts:

- GYG-Y195*p*IPhe: 48 µl (0.86 mg/ml), 1.4 nmol
- Boronic acid **4.1**: 48 µl of a 45 mg/ml stock solution prepared in H_2O, 5100 nmol, 3600 eq.
- $Pd(L2)_2(OAc)_2$: 6.2 µl of 10 mM stock solution, 62 nmol, 44 eq., 0.6 mM final concentration.

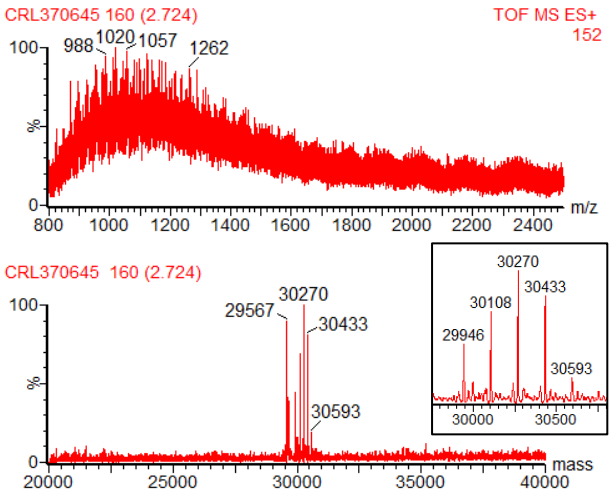
Example LC-MS after cross-coupling (73 % conversion):



M_{calc}: 29946
M_{obs}: 29947 (73 %)
29567 (de-iodination)
29606 (see **figure 4.6** and chapter 3 pp. 172-3)
29690 (GYG-Y195*p*IPhe)

UDP-galactose was utilised as the donor sugar nucleotide during enzymatic activity assay. Aliquots were quenched and analysed after the following times of autogalactosylation assay: 5 min, 10 min, 20 min, 30 min, 60 min, 120 min. Representative LC-MS spectra and overall graphical data are shown below.

Representative LC-MS spectra from galactosylation assay: t = 5 min

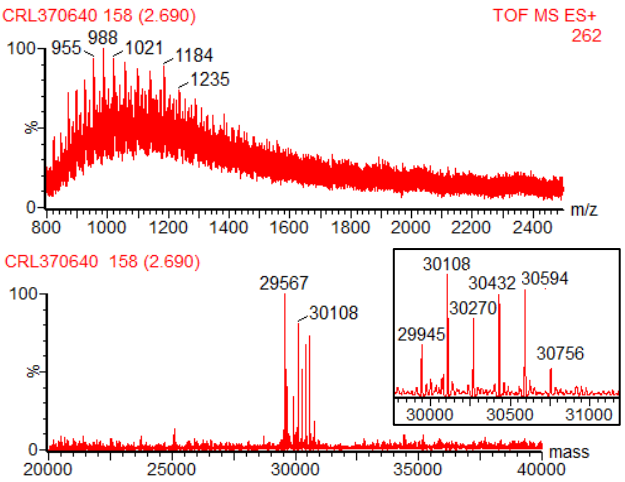


M_{calc}: 29946 + 162n

M_{obs}: 29946 + 162n (n = 0-4)

+ inactive side-products from cross-coupling (29567, 29606, 29690: see "LC-MS after cross-coupling" above)

t = 30 min

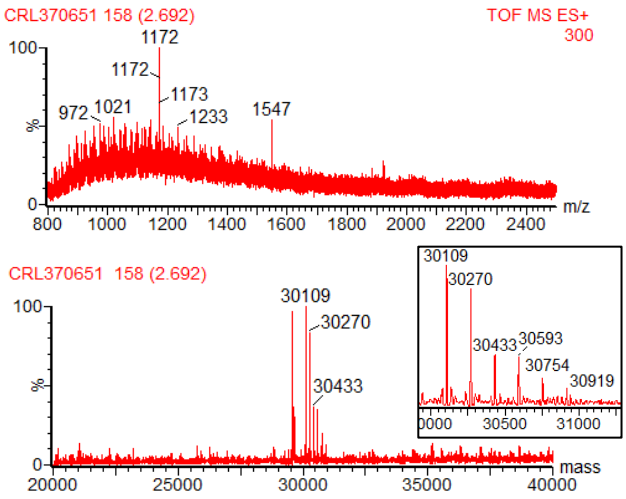


M_{calc}: 29946 + 162n

M_{obs}: 29945 + 162n (n = 0-5)

+ inactive side-products from cross-coupling (29567, 29606, 29690: see "LC-MS after cross-coupling" above)

t = 120 min



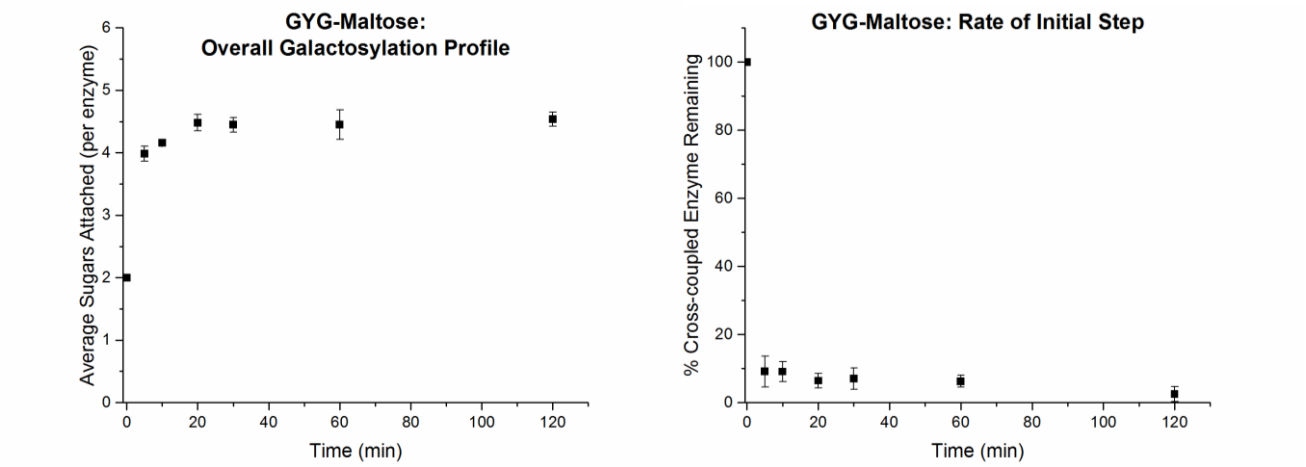
M_{calc}: 29946 + 162n

M_{obs}: 29946 + 162n (n = 1-6)

\+ inactive side-products from cross-coupling (29567, 29606, 29689: see "LC-MS after cross-coupling" above)

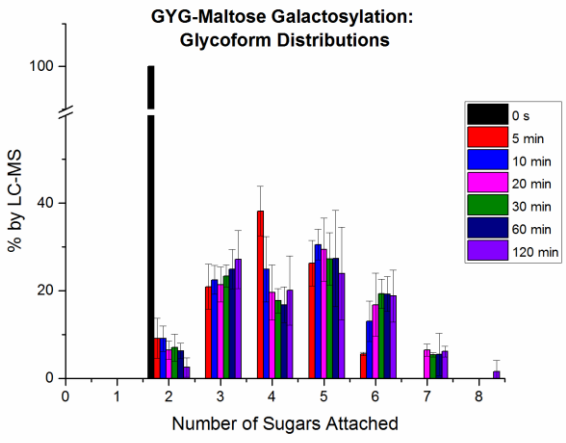
Graphical data from LC-MS from all timepoints, across all replicates (figure 4.12):

The methods for generation of this data from LC-MS spectra such as those shown above is described in section 4.5.3.



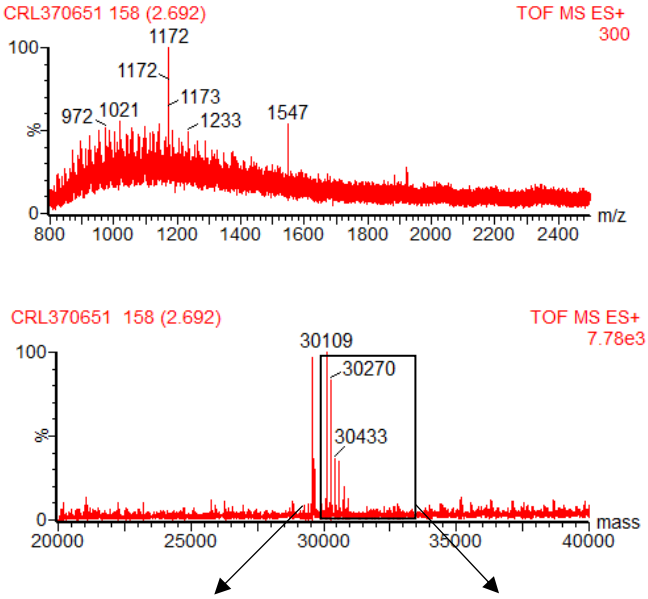
Extended analysis of glycoform distributions (figure 4.13):

The methods for generation of this data from LC-MS spectra such as those shown above is described in section 4.5.3.

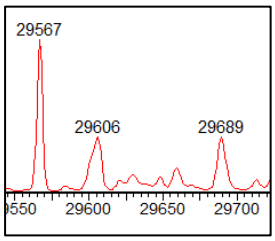


Detailed view of a representative LC-MS spectrum after galactosylation:

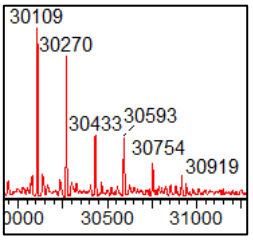
Below is a detailed view of one of the spectra above (t = 120 min), providing a zoomed-in view of not only the catalytically active, galactosylated species but also the inactive cross-coupling side-products (in order to confirm that these peaks do indeed correspond to these side-products; they are the same throughout all time points).



Inactive Cross-Coupling Side-Products



Catalytically Active Species



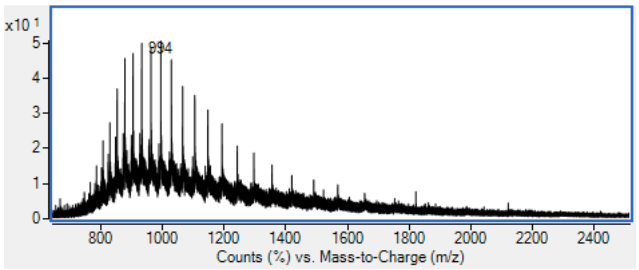
29567: de-iodinated
29606: see **figure 4.6** and chapter 3 pp. 172-3
29689: GYG-Y195*p*IPhe

30109 - 30919 = galactosylated GYG-Maltose (= 29946 + 162n, n = 1-6)

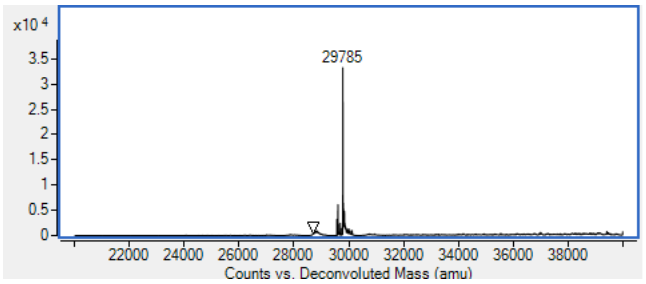
4.5.4.6 Comparison of GYG-Maltose and GYG-Glucose Autoglucosylation, Equal Cross-Coupling

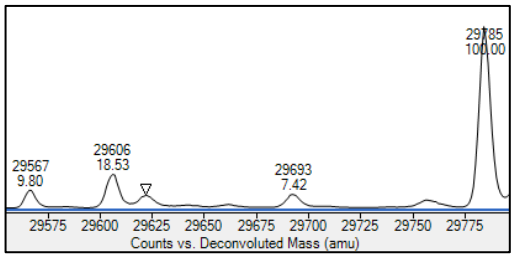
In the text (section 4.3.4), it was stated that one replicate from GYG-Glucose and GYG-Maltose autoglucosylation had almost identical cross-coupling efficiencies. Their autoglucosylation profiles showed the same differences as analogous profiles across all replicates, supporting the claim that levels of side-products are not responsible for differences between GYG-Maltose and GYG-Glucose in this regard. LC-MS data after cross-coupling for each of these isolated replicates, and corresponding autoglucosylation graphs, are shown below, to illustrate these points. Note that these represent one replicate of the protocol shown in section 4.5.4.2 (GYG-Glucose) and 4.5.4.3 (GYG-Maltose), i.e. the experimental procedure is identical.

LC-MS after cross-coupling, GYG-Glucose

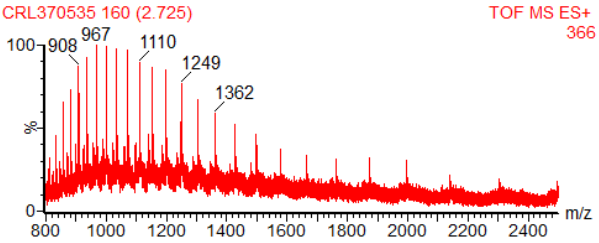


M_{calc}: 29784
M_{obs}: 29785 (74 %)
29567 (de-iodination)
29606 (see **figure 4.6** and chapter 3 pp. 172-3)
29693 (GYG-Y195*p*IPhe)

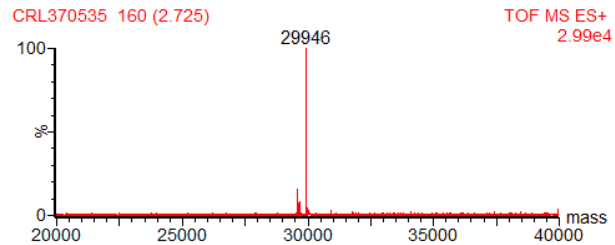


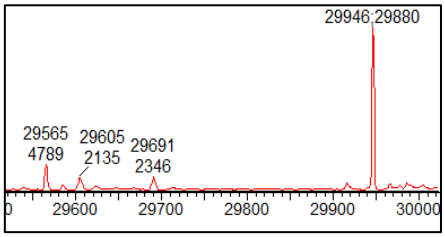


LC-MS after cross-coupling, GYG-Maltose

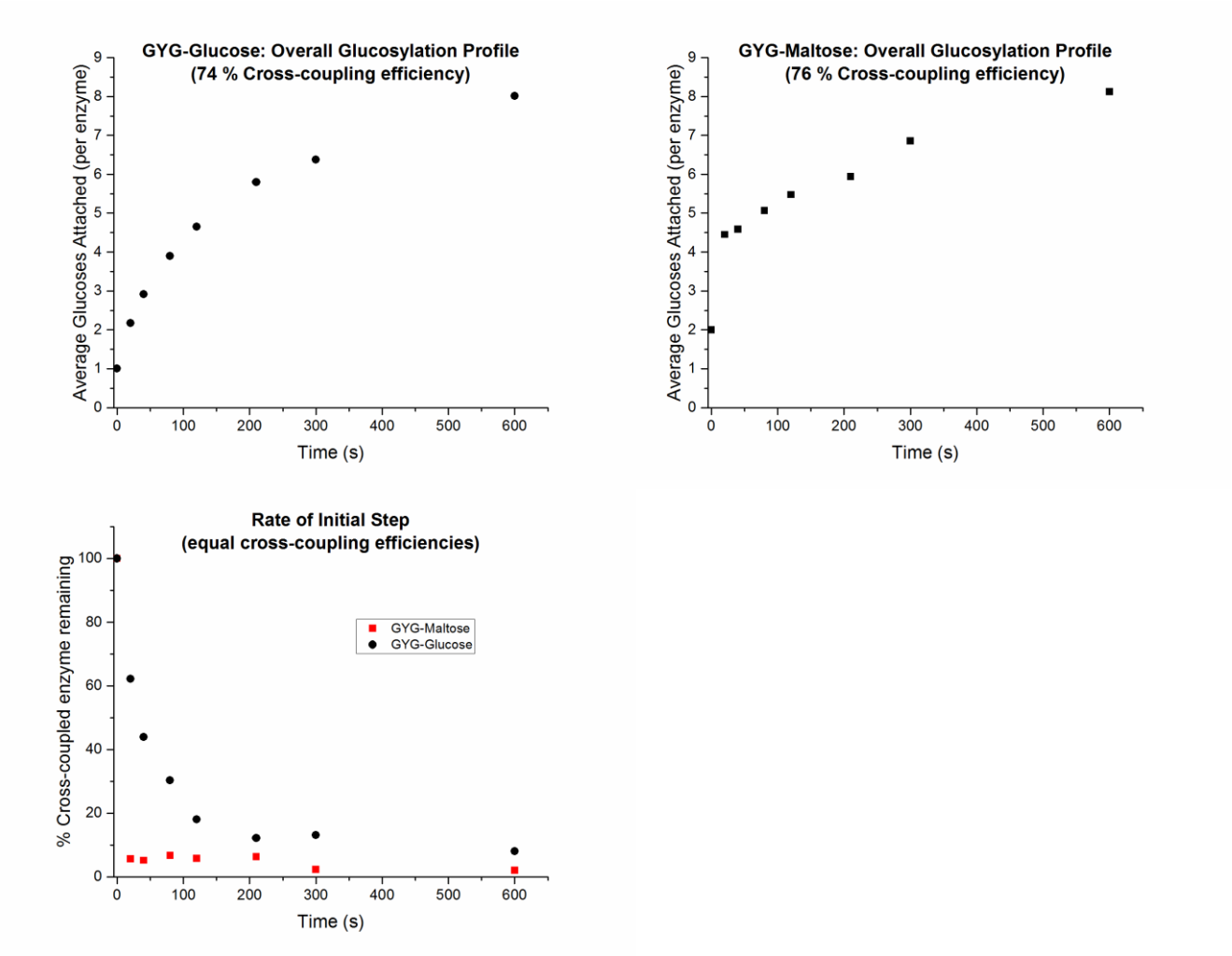


M_{calc}: 29946
M_{obs}: 29946 (76 %)
29565 (de-iodination)
29605 (see **figure 4.6** and chapter 3 pp. 172-3)
29691 (GYG-Y195*p*IPhe)





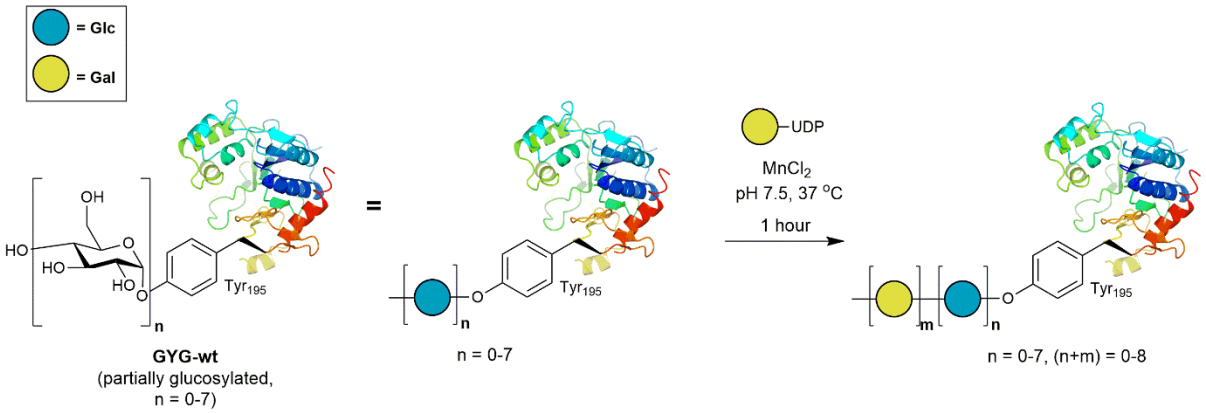
Autoglucosylation data for these cross-coupling samples:



The two enzymes exhibit precisely the same differences as observed across all replicates (**figure 4.8**), illustrating that these differences exist at near-identical levels of cross-coupling side-products, and thus highlighting that they do not arise from differences in levels of such species.

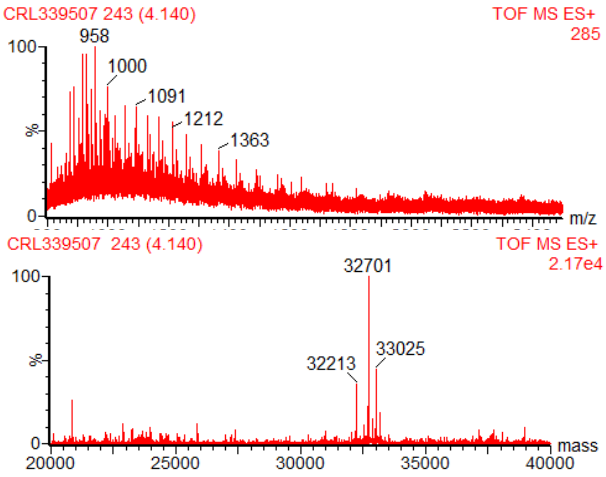
Specifically, GYG-Maltose exhibits a biphasic overall glucosylation profile with rapid initial step. GYG-Glucose exhibits a slower initial step and a glucosylation profile showing gradual decline in rate with increasing oligosaccharide length.

4.5.5 Autogalactosylation of wild-type GYG



To GYG-wt (0.2 mg/ml) in pH 7.5 25 mM Na-HEPES was added $MnCl_2$ and then UDP-Galactose disodium salt, both to final concentrations of 1 mM (added from freshly prepared 10 mM stock solutions in H_2O). The protein was incubated at 37 °C with shaking (500 rpm). After 10, 20, 30, 45 and 60 minutes, aliquots were quenched through the addition of Na-EDTA to a final concentration of 20 mM, and analysed directly by LC-MS. Note that this experiment, unlike previous, was carried out in 1x replicate only.

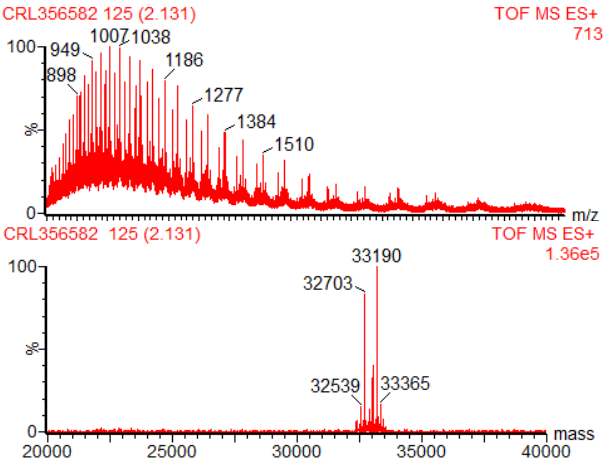
LC-MS data: t = 0 (before assay)



M_{calc}: 32048 + 162n
M_{obs}: 32051 + 162n

(n = glucoses attached = 1-7,
average = **4.07**)

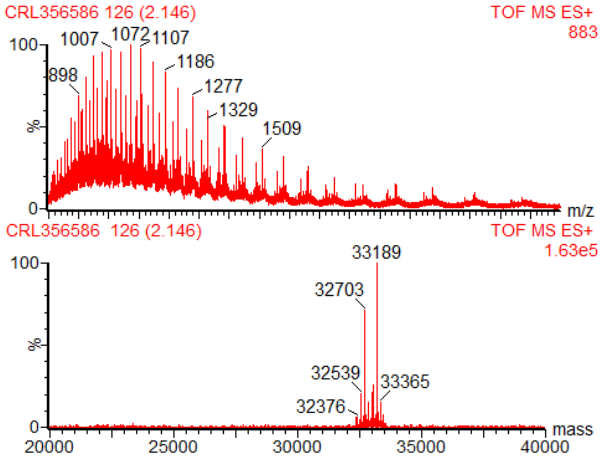
t = 10 min



M_{calc}: 32048 + 162(n+m)
M_{obs}: 32053 + 162(n+m)

n+m = glc + gal attached = 2-8
average = **5.57**

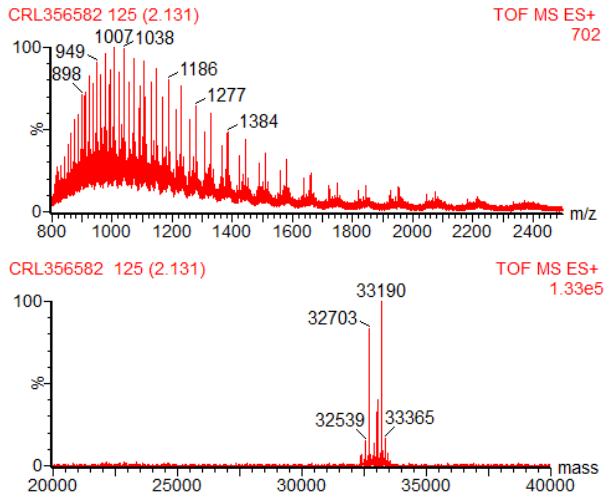
t = 30 min



M_{calc}: 32048 + 162(n+m)
M_{obs}: 32054 + 162(n+m)

n+m = glc + gal attached = 2-8
average = **5.55**

t = 60 min

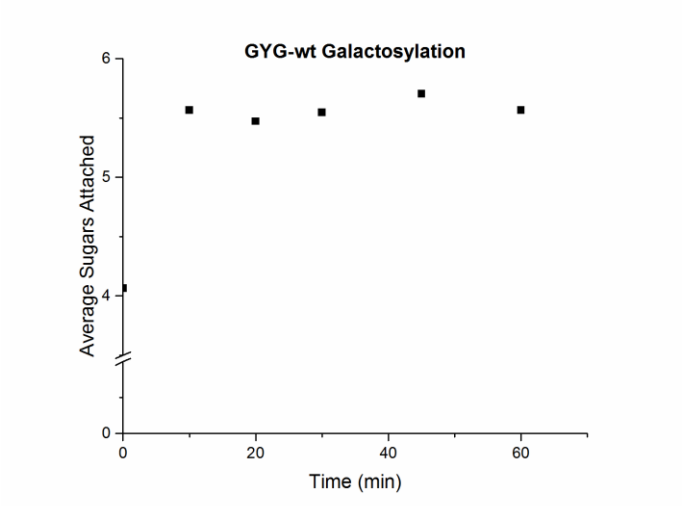


M_{calc}: 32048 + 162(n+m)
M_{obs}: 32055 + 162(n+m)

n+m = glc + gal attached = 2-8
average = **5.57**

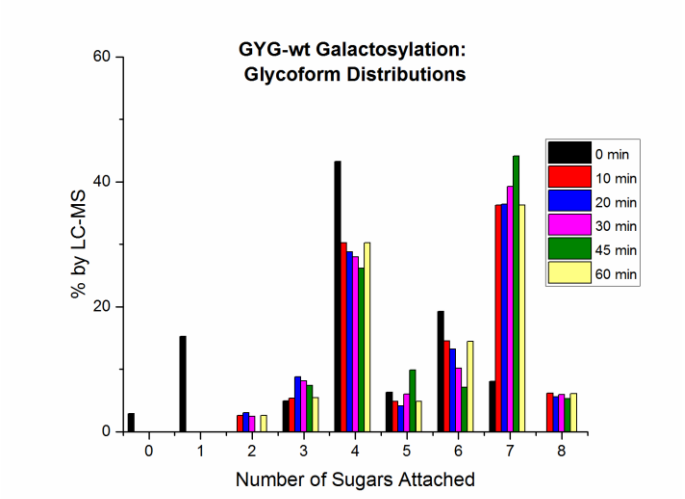
Graphical data (figure 4.14)

The method for generation of this data from LC-MS spectra such as those shown above is described in section 4.5.3.



Extended Analysis of Glycoform Distributions (figure 4.14).

The methods for generation of this data from LC-MS spectra such as those shown above is described in section 4.5.3.



4.6 References

1. Gibbons, B. J., Roach, P. J. & Hurley, T. D. Crystal structure of the autocatalytic initiator of glycogen biosynthesis, glycogenin. *J. Mol. Biol.* **319,** 463–477 (2002).

2. Alonso, M. D., Lomako, J., Lomako, W. M. & Whelan, W. J. Tyrosine-194 of glycogenin undergoes autocatalytic glucosylation but is not essential for catalytic function and activity. *FEBS Lett.* **342,** 38–42 (1994).

3. Lin, A., Mu, J., Yang, J. & Roach, P. J. Self-glucosylation of glycogenin, the initiator of glycogen biosynthesis, involves an inter-subunit reaction. *Arch. Biochem. Biophys.* **363,** 163–170 (1999).

4. Bazán, S., Issoglio, F. M., Carrizo, M. E. & Curtino, J. A. The intramolecular autoglucosylation of monomeric glycogenin. *Biochem. Biophys. Res. Commun.* **371,** 328–332 (2008).

5. Chaikuad, A. *et al.* Conformational plasticity of glycogenin and its maltosaccharide substrate during glycogen biogenesis. *Proc. Natl. Acad. Sci. U.S.A.* **108,** 21028–21033 (2011).

6. Issoglio, F. M., Carrizo, M. E., Romero, J. M. & Curtino, J. A. Mechanisms of monomeric and dimeric glycogenin autoglucosylation. *J. Biol. Chem.* **287,** 1955–1961 (2012).

7. Hurley, T. D., Walls, C., Bennett, J. R., Roach, P. J. & Wang, M. Direct detection of glycogenin reaction products during glycogen initiation. *Biochem Biophys Res Commun.* **348,** 374–378 (2006).

8. Cao, Y., Mahrenholz, A. M., DePaoli-Roach, A. A. & Roach, P. J. Characterization of rabbit skeletal muscle glycogenin: Tyrosine 194 is essential for function. *J. Biol. Chem.* **268,** 14687–14693 (1993).

9. Spicer, C. D. & Davis, B. G. Rewriting the bacterial glycocalyx via Suzuki-Miyaura cross-coupling. *Chem. Commun. (Camb).* **49,** 2747–2749 (2013).

10. Ma, D.-L., Shum, T. Y.-T., Zhang, F., Che, C.-M. & Yang, M. Water soluble luminescent platinum terpyridine complexes with glycosylated acetylide and arylacetylide ligands: photoluminescent properties and cytotoxicities. *Chem. Commun. (Camb).* 4675–4677 (2005).

11. Murakami, T., Hirono, R., Sato, Y. & Furusawa, K. Efficient synthesis of ω-mercaptoalkyl 1,2-trans-glycosides from sugar peracetates. *Carbohydr. Res.* **342,** 1009–1020 (2007).

12. Wagner, G. K. & Pesnot, T. Glycosyltransferases and their assays. *ChemBioChem* **11,** 1939–1949 (2010).

13. Williams, G. J. & Thorson, J. S. A high-throughput fluorescence-based glycosyltransferase screen and its application in directed evolution. *Nat. Protoc.* **3,** 357–362 (2008).

14. Ryu, J., Eom, M. S., Ko, W., Han, M. S. & Lee, H. S. A fluorescence-based glycosyltransferase assay for high-throughput screening. *Bioorganic Med. Chem.* **22,** 2571–2575 (2014).

15. Taniguchi, N., Nishikawa, A., Fujii, S. & Gu, J. Glycosyltransferase Assays Using Pyridylaminated Acceptors: N-Acetylglucosaminyltransferase III, IV, and V. *Methods Enzymol.* **179,** 397–408 (1989).

16. Pace, M., Agnellini, D., Gardana, C., Mauri, P. L. & Pietta, P. G. High-performance liquid chromatographic assay of glycosyltransferases using flavonoids as substrate. *J. Chromatogr. A* **691,** 331–336 (1995).

17. Walton, K. M. & Schnaar, R. L. Ganglioside glycosyltransferase assay using ion-exchange chromatography. *Anal. Biochem.* **152,** 154–159 (1986).

18. Miyashiro, M., Furuya, S. & Sugita, T. A high-throughput screening system for α1-3 fucosyltransferase-VII inhibitor utilizing scintillation proximity assay. *Anal. Biochem.* **338,** 168–170 (2005).

19. Hood, C. M., Kelly, V. A., Bird, M. I. & Britten, C. J. Measurement of alpha(1-3) fucosyltransferase activity using scintillation proximity. *Anal. Biochem.* **255,** 8–12 (1998).

20. Jain, J. C., Reed, D. W., GrootWassink, J. W. D. & Underhill, E. W. A radioassay of enzymes catalyzing the glucosylation and sulfation steps of glucosinolate biosynthesis in Brassica species. *Anal. Biochem.* **178,** 137–140 (1989).

21. Skurat, A. V, Dietrich, A. D., Zhai, L. & Roach, P. J. GNIP, a novel protein that binds and activates glycogenin, the self-glucosylating initiator of glycogen biosynthesis. *J. Biol. Chem.* **277,** 19331–8 (2002).

22. Hurley, T. D., Stout, S., Miner, E., Zhou, J. & Roach, P. J. Requirements for catalysis in mammalian glycogenin. *J. Biol. Chem.* **280,** 23892–9 (2005).

23. Yang, M., Brazier, M., Edwards, R. & Davis, B. G. High-throughput mass-spectrometry monitoring for multisubstrate enzymes: Determining the kinetic parameters and catalytic activities of glycosyltransferases. *ChemBioChem* **6,** 346–357 (2005).

24. Nagahori, N. & Nishimura, S. I. Direct and efficient monitoring of glycosyltransferase reactions on gold colloidal nanoparticles by using mass spectrometry. *Chem. - A Eur. J.* **12,** 6478–6485 (2006).

25. Nilsson, J. *et al.* LC-MS/MS characterization of combined glycogenin-1 and glycogenin-2 enzymatic activities reveals their self-glucosylation preferences. *Biochim. Biophys. Acta* **1844,** 398–405 (2014).

26. Carrizo, M. E., Miozzo, M. C., Goldraij, A. & Curtino, J. A. Purification of rabbit skeletal muscle proteoglycogen: Studies on the glucosyltransferase activity of polysaccharide-free and -bound glycogenin. *Glycobiology* **7,** 571–578 (1997).

27. Varki, A. *et al.* Symbol nomenclature for graphical representations of glycans. *Glycobiology* **25,** 1323–1324 (2015).

28. Roach, P. J., Depaoli-Roach, A. A., Hurley, T. D. & Tagliabracci, V. S. Glycogen and its metabolism: some new developments and old themes. *Biochem. J.* **441,** 763–787 (2012).

29. Roach, P. J., Skurat, A. V & Harris, R. A. in *Handbook of Physiology* (eds. Jefferson, L. S. & Cherrington, A. D.) 609–647 (Oxford University Press, 2000).

30. Frey, P. A. The Leloir pathway: a mechanistic imperative for three enzymes to change the stereochemical configuration of a single carbon in galactose. *FASEB J.* **10,** 461–470 (1996).

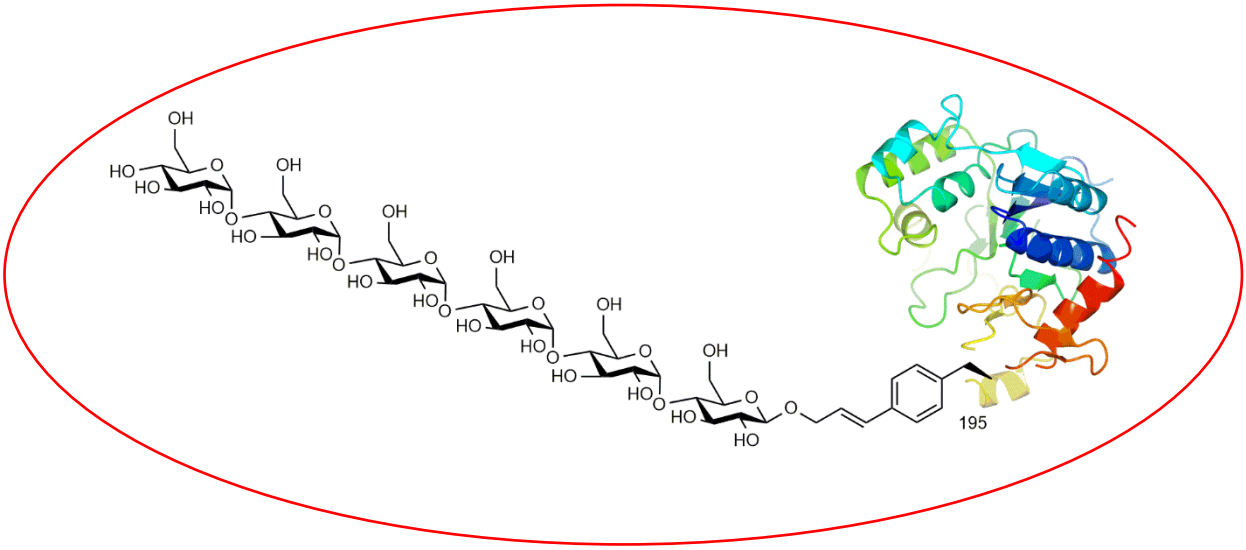
31. Rosado, F. R. *et al.* A partially 3-O-methylated (1-4)-linked α-D-galactan and α-D-mannan from *Pleurotus ostreatoroseus* Sing. *FEMS Microbiol. Lett.* **212,** 261–265 (2002).

32. Pomin, V. H. & Mourão, P. A. S. Structure, biology, evolution, and medical importance of sulfated fucans and galactans. *Glycobiology* **18,** 1016–1027 (2008).

33. Carbonero, E. R. *et al.* Unusual partially 3-O-methylated α-galactan from mushrooms of the genus Pleurotus. *Phytochemistry* **69,** 252–257 (2008).

34. Mu, J., Skurat, A. V & Roach, P. J. Glycogenin-2, a Novel Self-glucosylating Protein Involved in Liver Glycogen Biosynthesis. *J. Biol. Chem.* **272,** 27589–27597 (1997).

Chapter 5



Investigating a GYG-oligosaccharide Conjugate

Contents

5.1 Introduction 306

5.2 Synthesis of a Hexasaccharide Coupling Partner 308

5.3 Investigating the Generation of GYG-Maltohexaose 312

5.4 Catalytic Activity of GYG-Maltohexaose 318

5.4.1 Effect of Inactive Cross-coupling Side-product on Catalytic Activity of GYG-Maltose and GYG-Glucose 326

5.5 Conclusion and Outlook 327

5.6 Experimental Section 329

5.6.1 Chemical Synthesis 329

5.6.2 Cross-coupling Studies 337

5.6.2.1 Initial Cross-coupling of GYG-Y195pIPhe to maltohexaose boronic acid 5.1 337

5.6.2.2 Cross-coupling optimisations from table 5.1: Boronic Acid, Pd, Time 338

5.6.2.3 Further Cross-coupling Optimisation Studies 341

5.6.3 Enzymatic Activity Studies 345

5.6.3.1 Autoglucosylaton of GYG-Maltohexaose 346

5.6.3.2 Generation and autoglucosylaton of GYG-Glucose / GYG-Maltose under equivalent conditions 349

5.6.3.3 Effect of GYG-Y195*p*IPhe spiking on activity of GYG-Glucose and GYG-Maltose 354

5.7 References 358

5.1 Introduction

The successful delineation of the triphasic mechanism outlined in the previous chapter was greatly assisted by an ability to compare distinct, individual glucosylation steps along the same glycogenesis pathway. This was a direct consequence of an ability to chemically access distinct catalytic intermediates on the same glycogenesis pathway: GYG-Glucose, and GYG-Maltose. The wealth of information gained from these two enzymes alone reflected in part the fortunate finding that these two enzymes, despite differing by just one glucose residue, represented the onset of two different catalytic phases within a triphasic system (the first and second phases respectively).

Whilst the use of GYG-Maltose proved sufficient to additionally visualise the final, slower third phase (reflecting the absence of the slow first phase for this enzyme), the "missing enzyme construct" at this stage was one that fell within, or at the onset of, the third phase alone. Noting that, as discussed previously (chapter 4, **figures 4.8, 4.12**) the second, rapid glucosylation phase appeared to encompass GYG species bearing up to 4-5 glucose residues, we envisaged that chemical access to a GYG construct bearing at least this number of carbohydrate moieties (for example, "GYG-Maltohexaose") might enable selective "capture" and investigation of the third, final kinetic phase (**figure 5.1**).

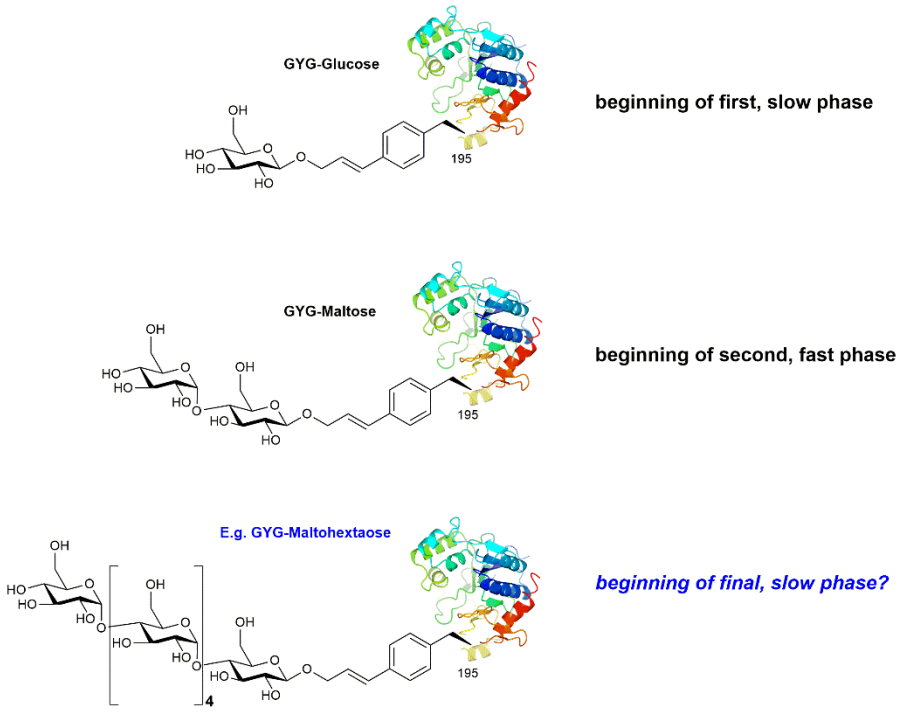


Figure 5.1 GYG-Glucose and GYG-Maltose represent enzymes at the onset of distinct phases identified in the previous chapter. Chemical access to a GYG species bearing a longer oligosaccharide, representing the onset of the final phase, could enable direct study of this phase in isolation.

Investigating the synthesis of such a GYG-oligosaccharide conjugate through Suzuki-Miyaura cross-coupling (SMC) would additionally be of inherent interest in a broader context than glycogenesis. It is well-established that the conjugation of oligosaccharides to protein carriers may lead to enhanced immunogenicity,[1–3] and to this end numerous chemical modification approaches to generate such glycoconjugates have been reported particularly in the context of vaccine development.[4–12] A number of these involve simple nucleophilic addition of naturally occurring amino acid residues (e.g. lysine,[6,10,11] cysteine[5,8,12]) to oligosaccharides bearing suitable electrophilic groups, for example maleimides,[5,8,12] squarate esters,[10,11] and iodoacetamides (**figure 5.2**).[13] However, these methods may yield heterogeneous mixtures of multiply glycosylated proteins since the amino acid residue may not be unique within the protein, potentially disadvantageous given the unknown impact of conjugation site on immunogenicity;[9] site-specific conjugation is desirable to generate synthetically defined vaccines that may exhibit greater efficacy and easier physicochemical characterisation.[4]

As discussed in chapter 1 in the broader context of glycoprotein synthesis, the use of unnatural amino acids as reactive "tags" for oligosaccharide conjugation may yield homogeneous species, e.g. via thiol addition to dehydroalanine,[9] or CuAAC to azidohomoalanine[14] (though Dha is of course typically derived from cysteine itself).[15] Yet SMC remains, to our knowledge, untested in this regard, and as such an evaluation of its potential towards more challenging oligosaccharide cross-couplings could not only yield glycoconjugates pertinent to GYG mechanistic investigation, but could afford useful information regarding its potential applicability to oligosaccharide-protein couplings for other purposes.

This chapter explores the synthesis of an oligosaccharide boronic acid for cross-coupling to GYG-Y195 *p*IPhe, in so doing evaluating the potential of this reaction as a tool for protein-oligosaccharide ligations in the context of the generation of a GYG-oligosaccharide glycoenzyme. We subsequently investigate the autoglucosylation activity of the coupled enzyme, and compare this to analogous activities of GYG-Maltose and GYG-Glucose in order to further probe the multiphasic nature of GYG catalysis.

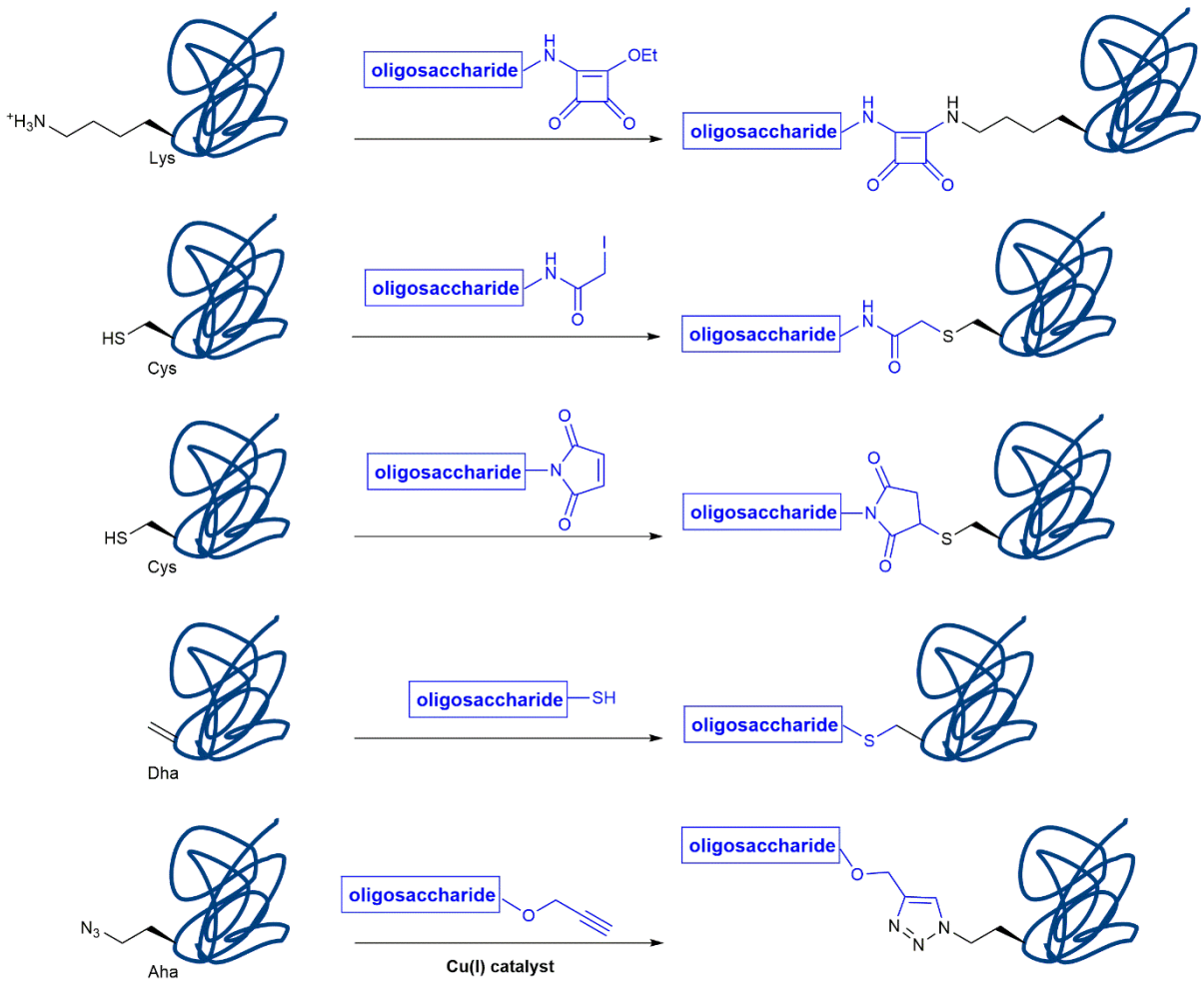


Figure 5.2 Some methods commonly utilised for the synthesis of oligosaccharide-protein conjugates.

5.2 Synthesis of a Hexasaccharide Coupling Partner

As mentioned, prior data obtained suggested that a transition between second and final catalytic phases occurred upon attachment of ca. 4-5 glucose moieties in total to GYG. In light of this, we reasoned that a GYG-maltohexaose species, accessible via cross-coupling of a maltohexaose-derived boronic acid **5.1** to GYG-Y195*p*IPhe, would represent a suitable target (**figure 5.3**); this enzyme should fall exclusively within the final phase, yet still near its onset. With GYG-Y195*p*IPhe already to hand, our intial efforts focussed on the synthesis of **5.1**.

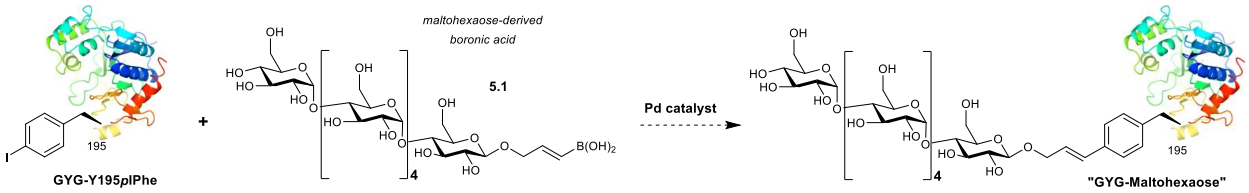


Figure 5.3 Proposed synthesis of a GYG-Maltohexaose glycoconjugate from GYG-Y195*p*IPhe and maltohexaose-derived boronic acid **5.1**.

Given that the same fundamental synthetic steps – hydroboration, glycosylation, boronic ester oxidation and deprotection – had enabled access to both monosaccharide and disaccharide boronic acids, it was reasoned that, with some modifications, these may also prove suitable for the synthesis of the more challenging hexasaccharide analogue. As such, a general synthetic route as outlined below (**figure 5.4**) could be envisaged. Glycosylation of a suitable peracetylated donor species **5.2** with pinacol ester **3.5** (the latter synthesised as described in the previous chapter) should yield intermediate pinacol ester **5.3**. Subsequent functional group conversion of the pinacol ester to the boronic acid (**5.4**), followed by global deprotection, would afford the desired target compound **5.1**.

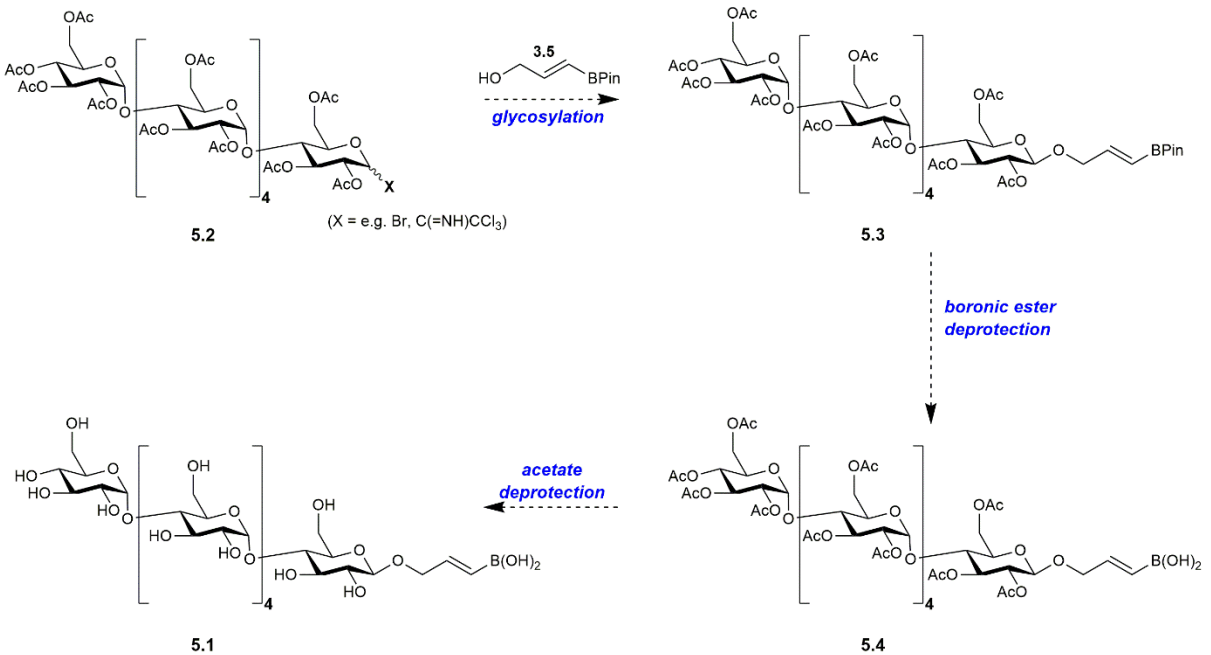


Figure 5.4 Proposed synthesis of maltohexaose boronic acid **5.1**.

In contrast to the disaccharide system, commercially available maltohexaose peracetate (**5.2**, X = OAc) itself proved insufficiently reactive to effect glycosylation to **3.5**. As such, the trichloroacetimidate donor **5.5** was synthesised in two steps from the peracetate (**figure 5.5**).[16,17] Selective deprotection of the anomeric acetate could be effected via treatment with hydrazine acetate, yielding hemiacetal **5.6**. Subsequent addition of the free hydroxyl to trichloroacetonitrile, catalysed by DBU, yielded the requisite glycosyl donor.

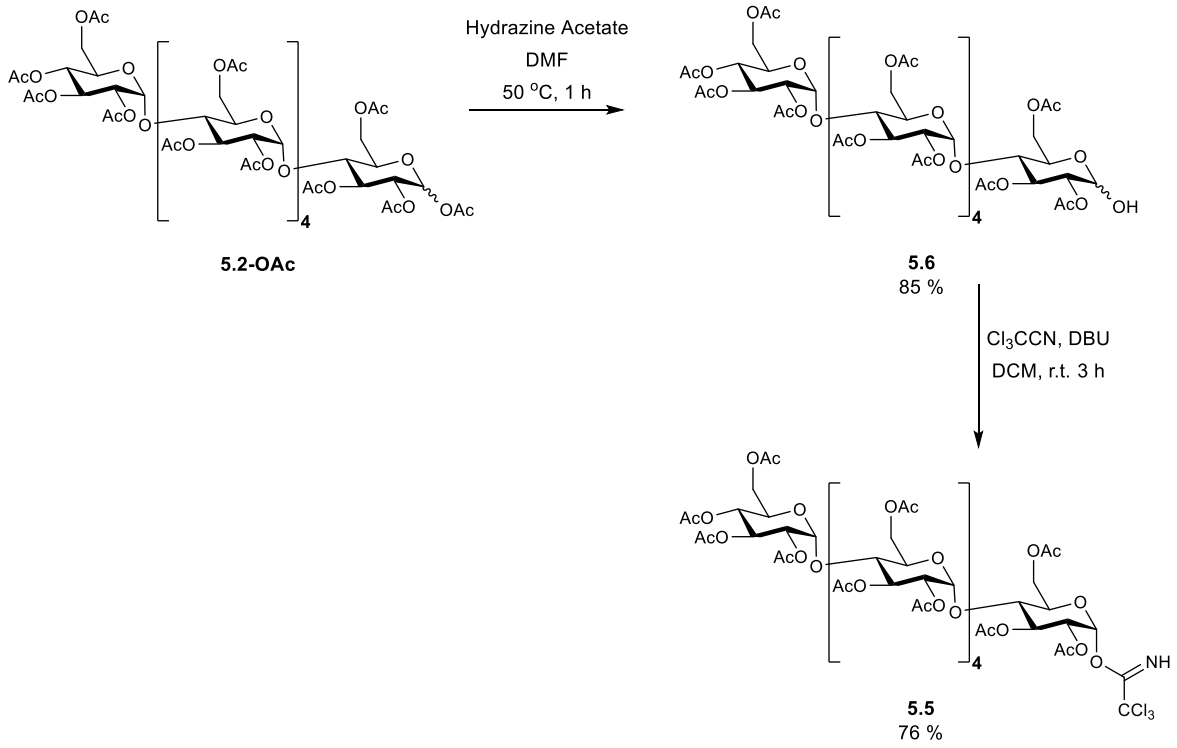


Figure 5.5 Synthesis of hexasaccharide imidate donor **5.5**.

Subsequent glycosylation of **5.5** with glycosyl acceptor **3.5** promoted by TMSOTf proceeded with a low level of concomitant mono-deacetylation as indicated by LC-MS (approx. 10-15 %). ^{1}H NMR indicated this to be occurring at the 2'-OH adjacent to the site of glycosylation (see experimental section 5.6.1), consistent with existing literature.[18] A subsequent re-acetylation step thus formed fully protected pinacol ester **5.3** (**figure 5.6**). This species was not purified at this stage, instead being subjected directly to subsequent boronic ester oxidation with $NaIO_4$, allowing more facile separation of boronic acid-containing compounds from other impurities (e.g. hydrolysed glycosyl donor). Full separation of the protected boronic acid product **5.4** from an impurity, indicated by ^{1}H, ^{13}C and 2D NMR analysis to correspond to **5.7** (derived from glycosyl acceptor **3.5**), could not be achieved at this stage. As such, the crude mixture was subjected directly to global deprotection under Zemplen conditions, affording the target compound **5.1** and impurity

indicated by ^{1}H NMR to correspond to **5.8** (as expected, see experimental section 5.6.1). Purification of **5.1** could be conveniently achieved through simple size-exclusion chromatography (PD Miditrap G-10, GE Healthcare), exploiting the large difference in molecular weights of the desired compound and impurity.[i] ^{1}H NMR indicated the β-anomer to be the major species (89:11 β:α; similar selectivity noted in a 2nd batch).

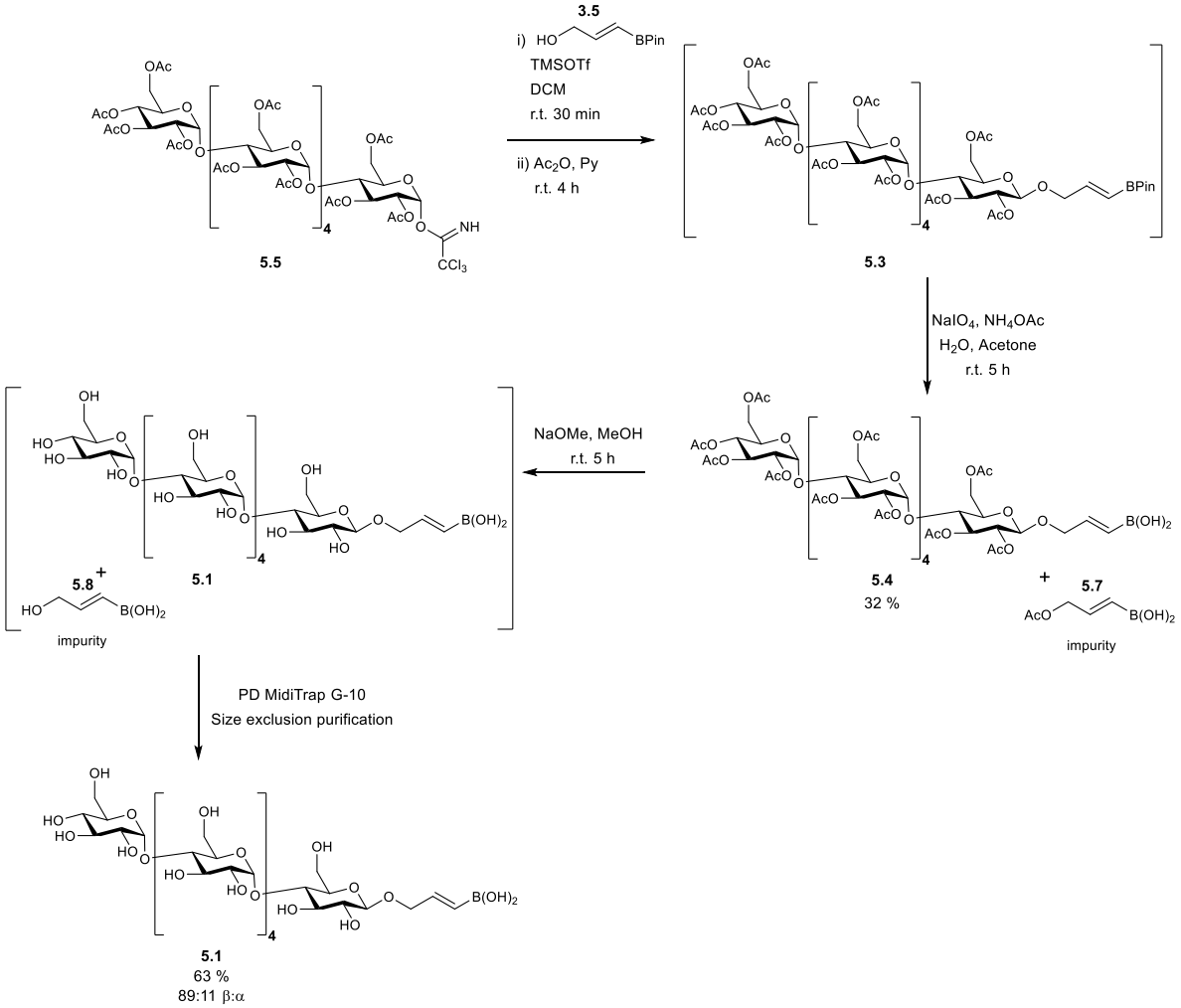


Figure 5.6 Synthesis of hexasaccharide boronic acid **5.1**.

[i] Note that this could not be so readily achieved prior to deprotection, since the Sephadex™ G-10 resin used in PD Miditrap G-10 columns is not generally compatible with organic solvents, and the protected species are not water-soluble.

5.3 Investigating the Generation of GYG-Maltohexaose

Anticipating that the cross-coupling of this hexasaccharide boronic acid to GYG-Y195*p*IPhe would likely be more challenging than previous couplings with smaller monosaccharide and disaccharide substrates, an initial test was carried out using the highest palladium loading previously shown to have compatibility with at least some subsequent enzymatic activity; 0.9 mM, as used for the cross-coupling of 2-HMPB (see chapter 2). With this exception, basic conditions were similar to those that proved optimal for the generation of both GYG-Glucose and GYG-Maltose, which were themselves the result of extensive optimisation work carried out on model systems in chapter 2. Pleasingly, a small amount of the desired GYG-Maltohexaose cross-coupled product could be detected upon LC-MS analysis (**figure 5.7**). However, reaction conversions as judged by LC-MS peak intensities were unfortunately low (17 %), with the principal side product being unreacted GYG-Y195*p*IPhe; a low level of de-iodination product was also visible.

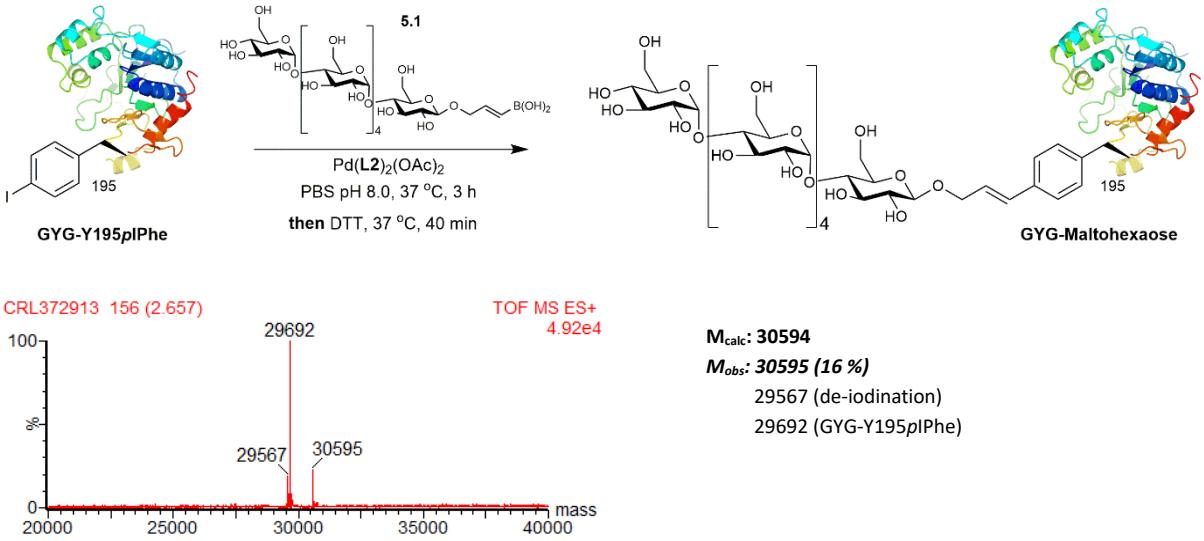
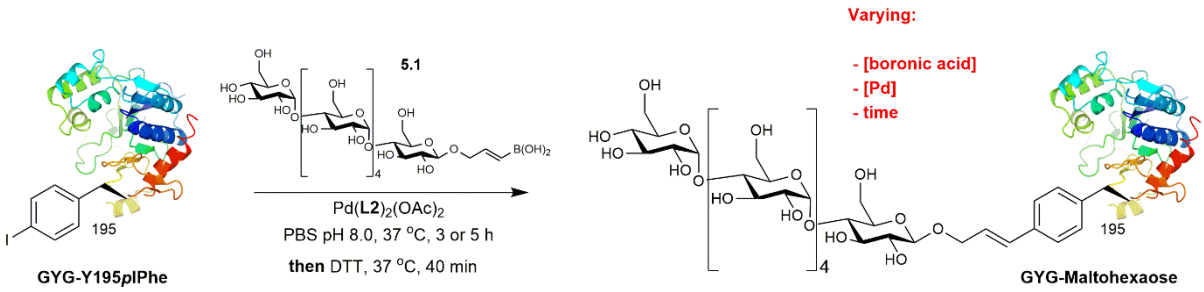


Figure 5.7 Initial cross-coupling between GYG-Y195*p*IPhe and maltohexaose boronic acid **5.1**, with deconvoluted LC-MS illustrated.

A number of factors affecting this conversion were thus investigated (**table 5.1**). A higher palladium loading led to little improvement (**table 5.1** entry 2, note that **table 5.1** entry 1 corresponds to the reaction shown in **figure 5.7**). In any case, this represented an undesirable avenue to pursue given the previously established detrimental effects of palladium on enzymatic activity (see chapter 2). Reduction in Pd loading was not investigated given prior indications that this typically lowers conversions, as would be expected (see chapter 3, **table 3.2**) An increase in boronic acid loading surprisingly led to a decrease in starting material consumption (**table 5.1**, entry 3), whilst a concomitant increase in reaction time had no significant effect (**table 5.1**, entry 4). We did not investigate variation of catalyst ligand to others known for protein cross-

couplings (see chapter 2, **figure 2.7**), since prior studies on both GYG (see chapter 2) and other proteins had indicated the current ligand, "L2", to be the most active.[19,20]



Entry	Equivs. $B(OH)_2$	$[Pd(L2)_2(OAc)_2]$ / mM	Reaction time / h	% Conversion[a]
1	3500	0.9	3	16
2	3500	**1.1**	3	18
3	**5000**	0.9	3	13
4	**5000**	0.9	**5**	14

[a]*as determined from peak intensities of deconvoluted LC-MS spectra.*

Table 5.1 Optimisation studies on cross-coupling between GYG-Y195*p*IPhe and boronic acid **5.1**. In entries 2-4, differences as compared to entry 1 are highlighted in red.

The detrimental effect of increased boronic acid loading was intriguing, and led us to wonder if a reduction might improve reaction conversions (**figure 5.8**). However, whilst unfortunately the presence of unidentified LC-MS adducts (masses between 29950-30020, see also experimental section 5.6.2.3)[ii] prevented accurate quantification of reaction conversion, it was immediately apparent that levels of GYG-Maltohexaose remained very low in relation to unreacted GYG-Y195*p*IPhe, which remained the predominant species present (and similar to levels of de-iodination product). For this reason, further investigation in this regard was therefore not pursued.

[ii] These were believed to arise from the machine itself, since they were observed across multiple different protein samples run during this time period by research group members. Their identity remains uncertain. Unfortunately, alternative machines were, at this time, not available or functioning sufficiently well to be of use.

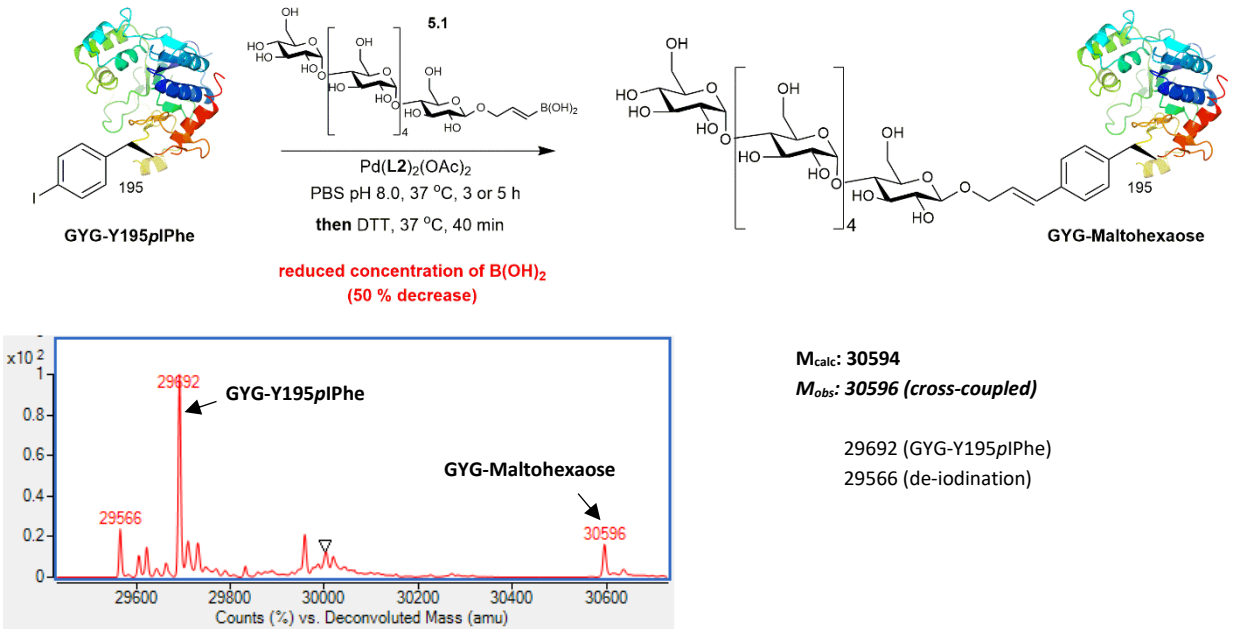


Figure 5.8 Reduction of boronic acid loading did not lead to significant improvement in levels of GYG-Maltohexaose formation, with the major species remaining GYG-Y195*p*IPhe. As discussed in the text (and footnote (i)), some LC-MS adducts were present (peaks between 29900-30050), precluding unambiguous quantitative analysis of conversion. However, it was clear that conversion of GYG-Y195*p*IPhe into GYG-Maltohexaose remained minimal.

The fact that the principal side-product was not de-iodination but unreacted GYG-Y195*p*IPhe suggests that the conversion-limiting step of the catalytic cycle (see chapter 1) is likely not transmetallation but the initial oxidative addition step preceding this. The $Pd(L2)_2(OAc)_2$ catalyst utilised is technically a "pre-catalyst", and, as discussed in chapter 1, such species must undergo an initial reduction from Pd(II) to Pd(0) in order to facilitate this step. This process is likely to be mediated by boronic acid homocoupling in this particular system,[19,21,22] which may itself be challenging for a large, sterically bulky boronic acid such as **5.1**.

As such, we wondered whether the addition of an alternative boronic acid as a "sacrificial reductant" may improve conversions to **5.1**. We first tested boronic acid 3-HMPB in this regard (**figure 5.9**), since this was previously shown to exhibit excellent reactivity (chapter 2). A low level of this species (350 eq., 1:9 ratio of 3-HMPB: **5.1**) was added so as to avoid competition in the cross-coupling step itself. We additionally trialled the use of crude **5.1** (prior to size exclusion chromatography) as a coupling partner, reasoning that the impurity **5.8** (see **figure 5.6**) might similarly serve as a reductant. In this latter case, levels of the "sacrificial" boronic acid were necessarily higher, since they reflected those naturally present in crude **5.1** (ca. 1:1 molar ratio by ^{1}H NMR), which could not be readily controlled.

Whilst the presence of LC-MS adducts (similarly at masses between 29900 – 30020, see also experimental section 5.6.2.3)[iii] again prevented precise quantification of reaction conversion, to our disappointment it was nonetheless immediately apparent that this strategy also failed to yield significant improvement. As illustrated, levels of the cross-coupled product GYG-Maltohexaose remained extremely low, in comparison to the significant quantities of unreacted GYG-Y195*p*IPhe persisting (**figure 5.9**). This approach was therefore not investigated further. When crude **5.1** was utilised as a cross-coupling partner, a mass consistent with coupling to **5.8** was observed; such coupling was not seen for 3-HMPB, however as noted the levels of the latter were far lower.

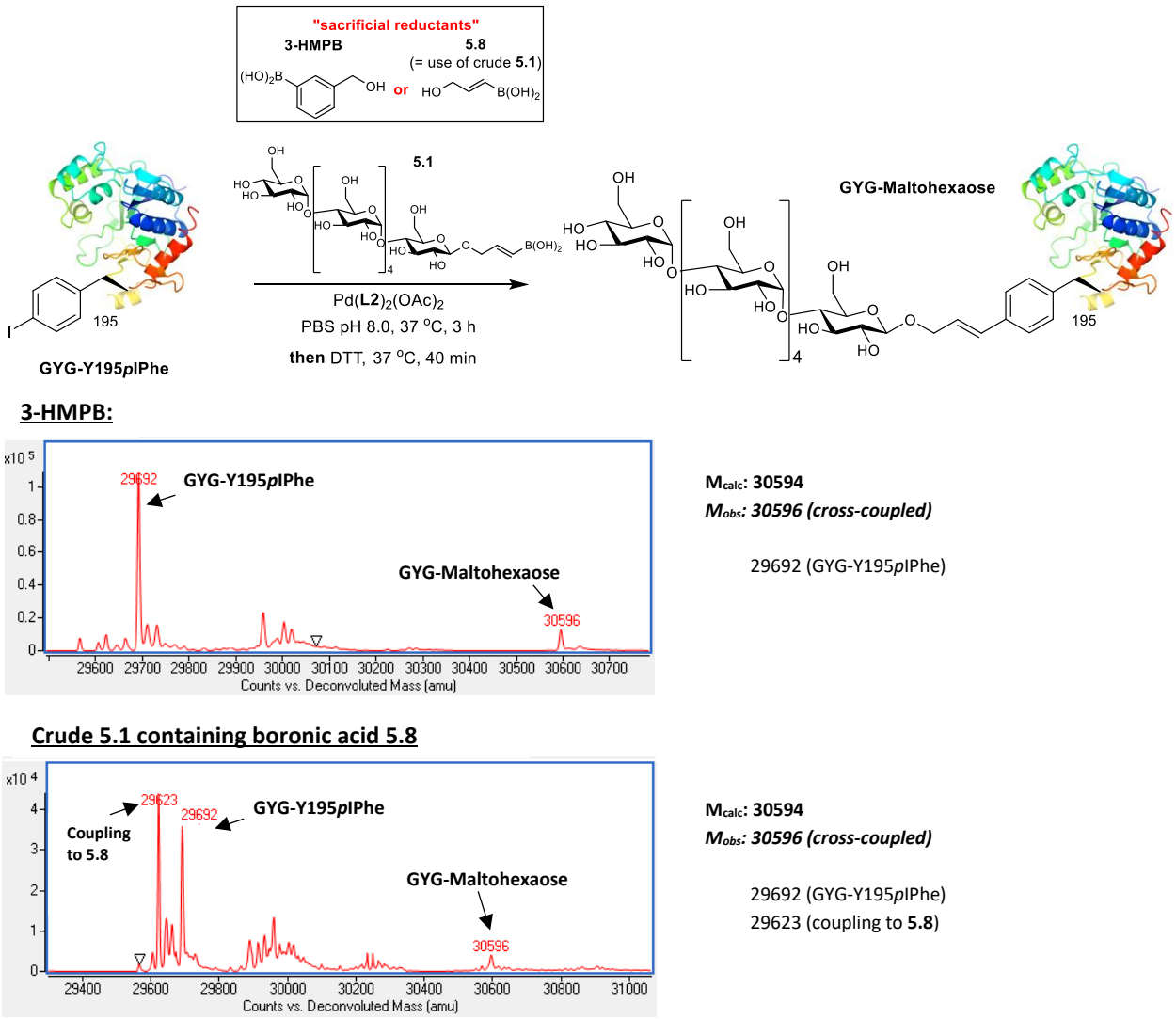


Figure 5.9 Addition of a "sacrificial" boronic acid to potentially aid in Pd(II) → Pd(0) pre-catalyst reduction. Whilst, as discussed in the text and in footnote (i), various LC-MS adducts were present (masses between 29900-30050), precluding quantitative analysis of conversion, these did not prevent the observation that levels of GYG-Maltohexaose remained low in both cases.

[iii] See footnote (i); we believe these to originate from the machine itself (although use of crude **5.1** may contribute)

Intriguing results arose from a competition experiment carried out in which equal amounts of **5.1** and boronic acid 3-HMPB were added to a cross-coupling reaction, and compared to the cross-coupling to 3-HMPB alone (**figure 5.10**, concentrations of 3-HMPB were identical in each case). Whilst a conversion of 66 % was attained in the latter case (**figure 5.10 a**), in the presence of **5.1** the level of cross-coupling to 3-HMPB dropped to 33 %; crucially, this decrease was not accounted for by cross-coupling to **5.1** itself, since the overall level of coupling (to both 3-HMPB and **5.1**) dropped to 43 % (**figure 5.10 b**).

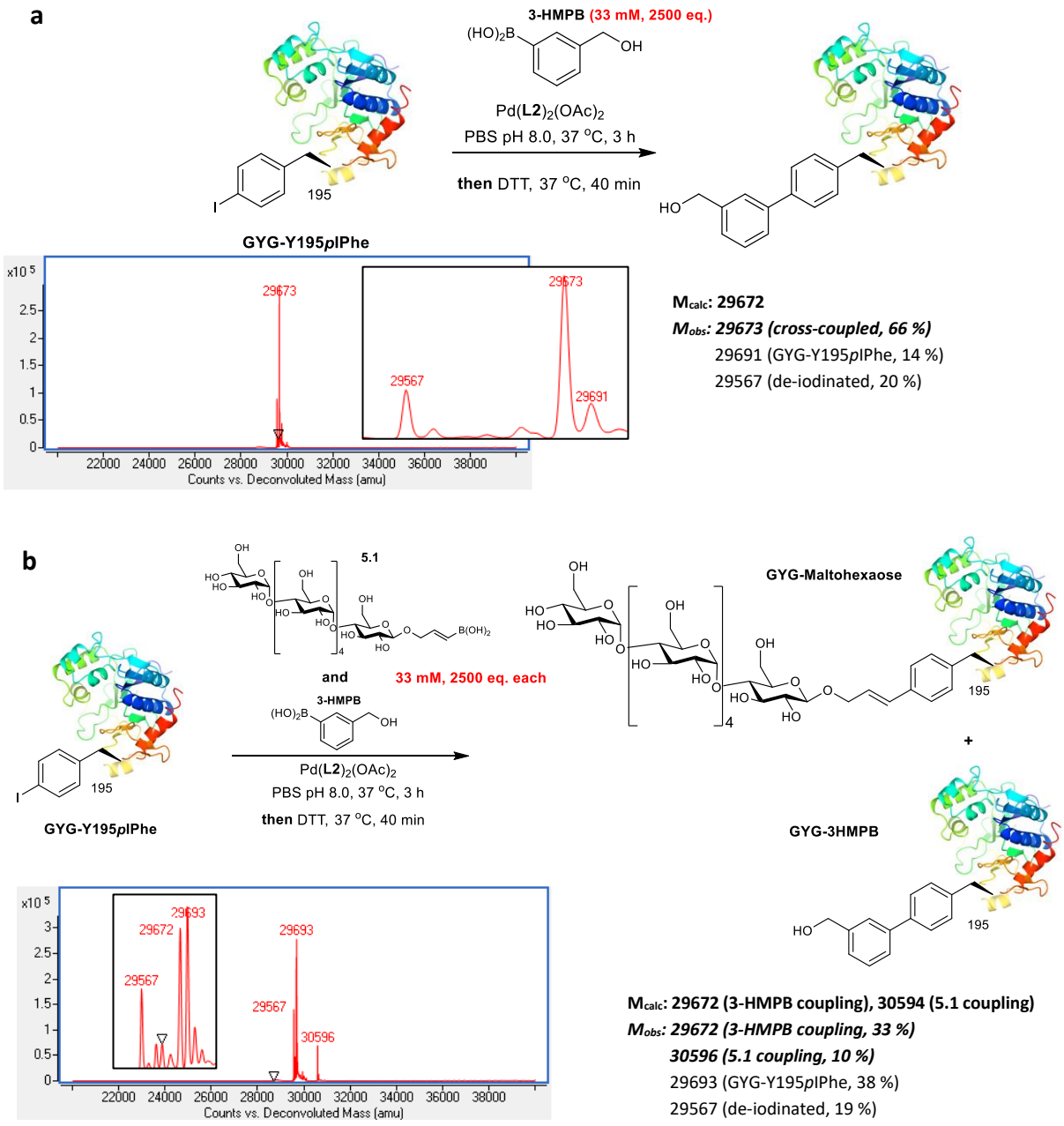


Figure 5.10 Competition experiment between cross-coupling of **5.1** and 3-HMPB. Compared to cross-coupling to 3-HMPB alone (top), both the cross-coupling of GYG-Y195*p*IPhe to 3-HMPB and the overall level of cross-coupling were reduced in the presence of **5.1** (lower).

This experiment implies that boronic acid **5.1** not only represents a challenging substrate for cross-coupling in its own right, but somewhat intriguingly acts more generally as a cross-coupling inhibitor (it does not do this by simple direct competition for coupling, since this would not explain the decrease in overall coupling conversion observed upon addition of **5.1**). Whilst this could quite feasibly reflect, for example, Pd-sequestration via the many carbohydrate hydroxyl groups present in this boronic acid – indeed, glucose-based Pd complexes are known[23] – in this scenario, one would expect increased Pd levels to notably improve conversions in couplings to this boronic acid, which was not the case (see **table 5.1**). It is additionally apparent that the continued poor cross-coupling to **5.1** in the presence of equimolar levels of 3-HMPB further confirms the result illustrated in **figure 5.9**, namely that 3-HMPB cannot improve coupling by assisting in Pd(II) → Pd(0) pre-catalyst reduction (even at the far higher levels used here).

A possible alternative might be that carbohydrate vicinal diol groups can react reversibly with any boronic acid moiety (including that on a molecule of **5.1** itself, or of e.g. 3-HMPB) to form boronic esters;[24,25] these are less reactive in cross-coupling than free boronic acids,[26] and by extension would also be less capable of reducing the Pd(II) precatalyst to the active Pd(0) species, should this occur via boronic acid homocoupling.[21,22] The proportion of less reactive boronic ester to more reactive boronic acid at a given time would presumably depend on the ratio of carbohydrate hydroxyl functionalities relative to the boronic acid functionality, which would in turn be consistent with the reduced reactivity of maltohexaose derivative **5.1** as compared to the monosaccharide and disaccharide variants used previously (see chapters 3, 4). However, this remains at present purely speculation; further mechanistic study would be necessary for a full understanding of the challenges involved in cross-couplings to oligosaccharide boronic acids such as **5.1**.

It is entirely possible that cross-coupling conversions to maltohexaose boronic acid **5.1** may be improved in the future, and prospects in this regard are discussed in section 5.5. It should be emphasised however that, whilst factors such as pH, temperature, buffer composition were not explicitly investigated for this particular coupling, such factors have been extensively optimised in the context of simpler model systems both in this work (chapter 2) and, moreover, in prior studies.[27] At this point, in the interests of time, it was decided to proceed directly towards an investigation of the catalytic activity of what cross-coupling product could be attained. Pleasingly, despite the low cross-coupling conversions, we were still able to gain useful information in this regard, as discussed in the following section.

5.4 Catalytic Activity of GYG-Maltohexaose

In spite of the generally poor cross-coupling between GYG-Y195*p*IPhe and maltohexaose boronic acid **5.1**, we were able to demonstrate that the cross-coupled enzyme exhibited autoglucosylation activity (**figure 5.11**). Upon subjection of cross-coupling preparation of GYG-Maltohexaose to autoglucosylation conditions (see chapter 4 and experimental section 5.6.3; 1 mM UDP-Glucose and 1 mM Mn^{2+} cofactor were used as before), and subsequent analysis by LC-MS, peaks corresponding to selective addition of glucose residues to GYG-Maltohexaose could be observed. As expected, the cross-coupling side-products did not undergo glucosylation, although their presence as a significant majority did render LC-MS analysis of the glucosylation of GYG-Maltohexaose challenging (owing to its consequent low abundance); nonetheless, peaks unambiguously corresponding to the addition of glucose moieties to GYG-Maltohexaose could be detected. Notably, the activity of the enzyme persisted in spite of the higher palladium loadings utilised in its formation.

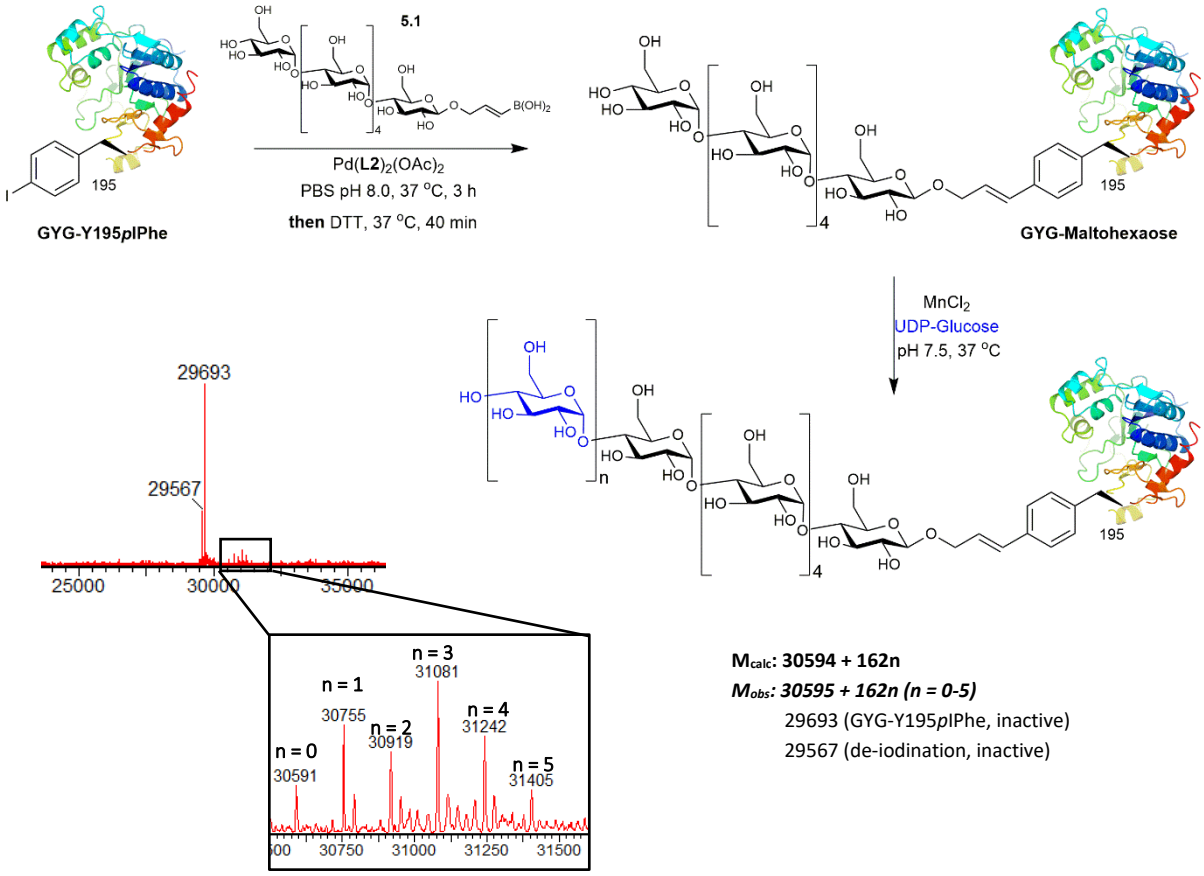


Figure 5.11 Catalytic activity of GYG-Maltohexaose, as illustrated by LC-MS analysis after glucosylation assay.

Subsequently, this glucosylation was examined in greater detail. In an analogous fashion to previous analyses for GYG-Glucose and GYG-Maltose (see chapter 4), both the overall glucosylation level of the active enzyme and the rate of consumption of the starting enzyme were analysed through LC-MS; the latter, in this case, provides a measure of the rate of the 6 → 7 glucosylation step. Both were determined from relative MS peak intensities of different glycoforms in LC-MS spectra from various assay time points, according to the method of Hurley *et al* (see chapter 4, experimental section 4.5.3 for details).[28] Whilst such MS-based quantification of glucosylation extent for this enzyme proved more challenging owing to the poor cross-coupling efficiency (and thus, the small proportion of the reaction mixture that corresponded to the active, coupled enzyme), sufficient data for duplicate analysis could still be obtained (**figure 5.12**).

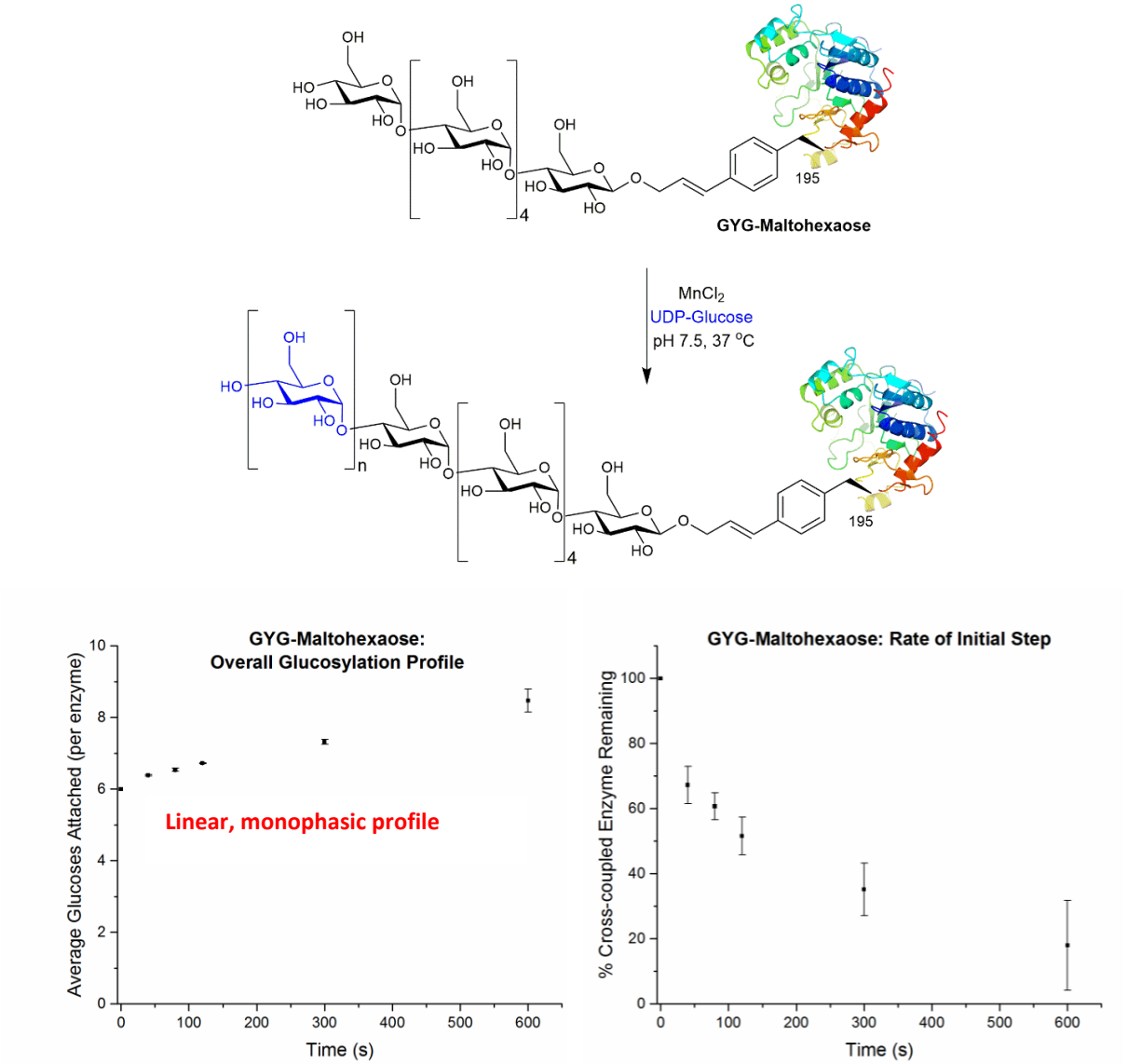


Figure 5.12 Analysis of the autoglucosylation activity of GYG-Maltohexaose. Average numbers of glucoses attached and proportion of cross-coupled enzyme remaining were determined from LC-MS analysis (see chapter 4 and experimental section). Data points represent mean averages of two independent replicates, error bars are ± s.d.

The overall glucosylation profile for this enzyme involves an essentially uniform autoglucosylation rate, irrespective of carbohydrate chain length; pleasingly, this is entirely consistent with our proposed triphasic kinetic profile outlined in the previous chapter, which is additionally illustrated below for purposes of clarity (**figure 5.13**). According to this profile, GYG-Maltohexaose should bypass the initial two phases, and thus should exhibit a monophasic glucosylation profile, precisely as observed above. This, importantly, also implies that the proposed final kinetic phase is indeed just one phase, and also indicates that it commences prior to attachment of 6 glucose residues.

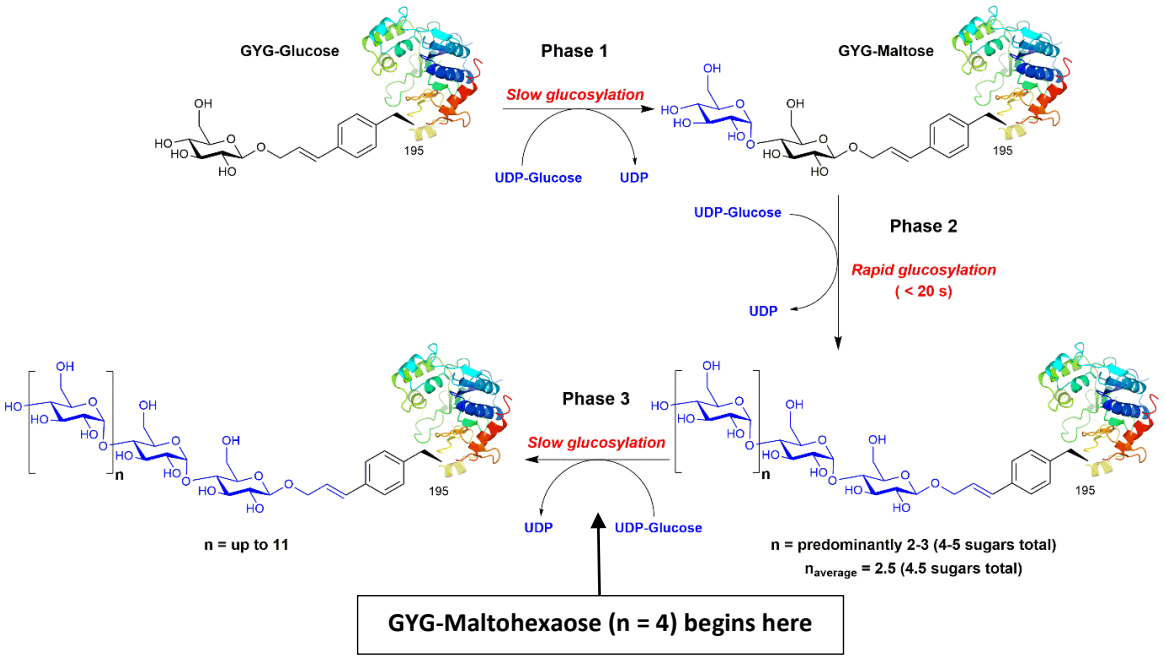


Figure 5.13 Reproduction of the triphasic kinetic profile delineated in the previous chapter. As indicated, according to this profile, GYG-Maltohexaose should bypass the initial two phases (since these involved extension only up to GYG-tetrasaccharide / GYG-pentasaccharide species).

Comparison with the data generated previously for GYG-Glucose and GYG-Maltose (chapter 4) serves to highlight the distinctions between the glucosylation profiles for these enzymes as compared to GYG-Maltohexaose (**figure 5.14**). Whilst the latter exhibits a monophasic profile corresponding to the final kinetic phase in isolation, GYG-Maltose exhibits a biphasic profile (including the phase before this), and GYG-Glucose exhibits a gradual decline in rate that, as discussed previously, reflects an amalgamation of all three phases. The initial glucosylation step for GYG-Maltohexaose, in common with that of GYG-Glucose, also appears significantly slower than that of GYG-Maltose, which would also be consistent with the catalytic profile depicted above (**figure 5.13**). Given that these three enzymes represent intermediates along the same pathway (in the order glucose → maltose → maltohexaose), this data further supports the

proposed "slow → fast → slow" kinetic profile. This is now visible for the first time through direct comparison of the initial steps of enzymes falling in each phase.

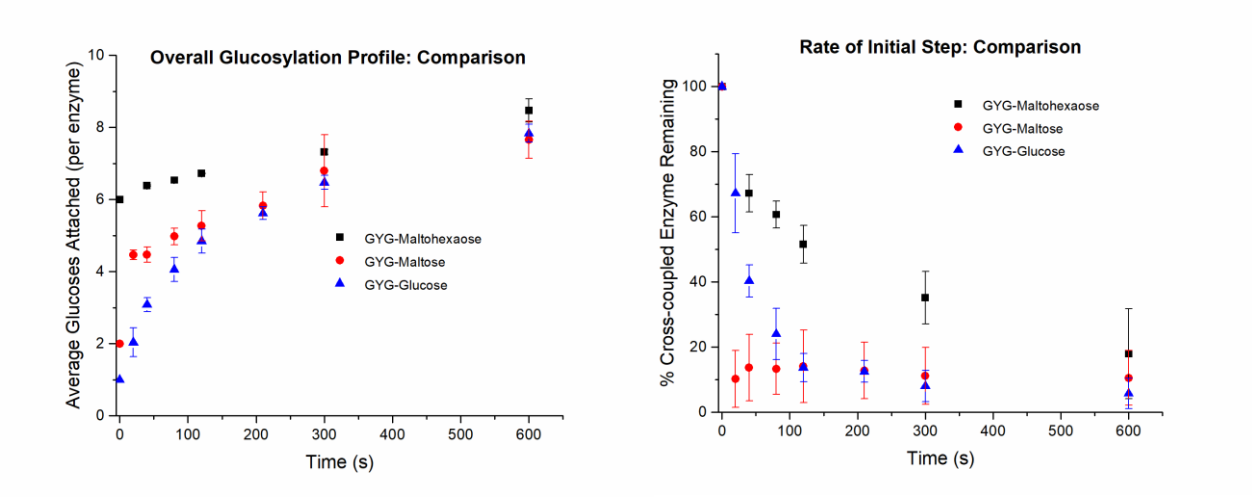


Figure 5.14 Compilation of glucosylation data generated for GYG-Maltohexaose, GYG-Maltose and GYG-Glucose. Data displayed is comprised from previous figures (**figure 4.8** and **figure 5.12**) and, other than being combined onto single graphs, is identical in content and format. Direct comparison should be carried out with caution owing to differing preparation and assay conditions between enzymes. Average numbers of glucoses attached and proportion of cross-coupled enzyme remaining were determined from LC-MS analysis (see chapter 4 and experimental section).

Two caveats to conclusions drawn from the above data must be acknowledged (which we attempt to address below). Firstly, the use of a higher palladium loading for GYG-Maltohexaose generation than for the other enzymes (0.9 mM vs 0.6 mM) may impact subsequent autoglucosylation, given the previously discussed effects of Pd on glucosylation activity of GYG-wt (see chapter 2). Secondly, whilst the cross-coupling side-products are not capable of their own autoglucosylation, it cannot be ruled out that the existence of such side-products (specifically, unreacted GYG-Y195*p*IPhe) as a significant majority in preparations of GYG-Maltohexaose, as opposed to distinct minority species for preparations of the other enzymes studied, could also lead to slight differences in the observed autoglucosylation profiles.

To investigate the extent to which such factors might influence – or not – for the conclusions drawn above, we generated the GYG-Glucose and GYG-Maltose enzymes under equivalent, higher Pd loadings as utilised for GYG-Maltohexaose (0.9 mM). Furthermore, the poor cross-coupling efficiencies observed for GYG-Maltohexaose were mimicked for GYG-Glucose and GYG-Maltose through "spiking" of cross-coupled samples thereof with additional GYG-Y195*p*IPhe prior to glucosylation; this is illustrated overleaf (**figure 5.15**).

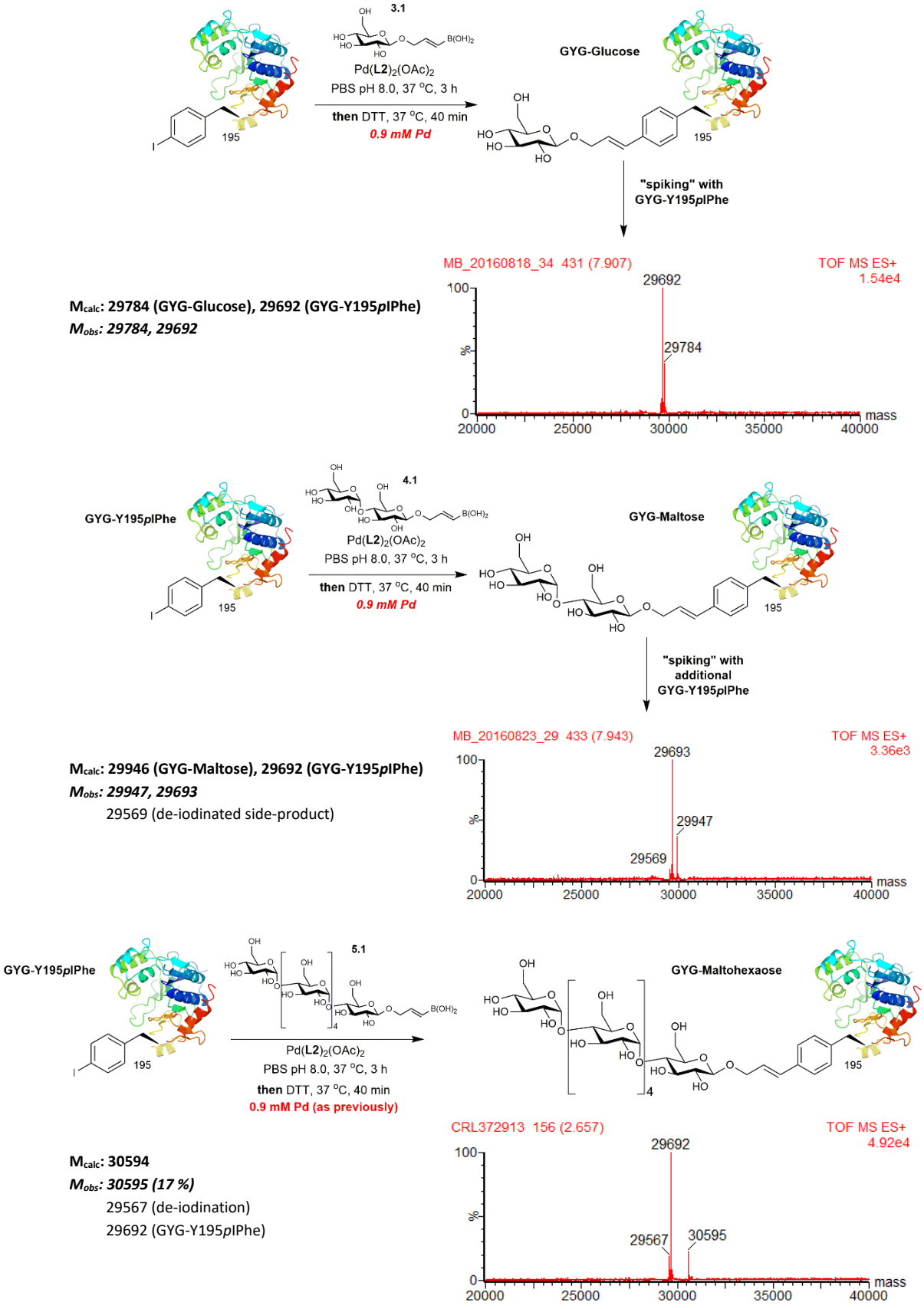


Figure 5.15 Preparation of GYG-Glucose, GYG-Maltose and GYG-Maltohexaose under equivalent Pd loadings. The former two samples were also "spiked" with GYG-Y195*p*IPhe in order to mimic poor cross-coupling of the latter.

We aimed to ensure that the levels of GYG-Y195*p*IPhe in GYG-Glucose and GYG-Maltose preparations were, if not necessarily identical, at least more similar to those present naturally in the GYG-Maltohexaose system, such that GYG-Y195*p*IPhe was substantially the major species in all cases. Proportions of cross-coupled enzyme and GYG-Y195*p*IPhe were judged solely by their relative peak intensities in deconvoluted MS spectra, given that, for the GYG-Maltohexaose system, this is necessarily the only means of evaluating such proportions. The "spiked" GYG-Y195*p*IPhe was "Pd-treated" (i.e. treated exactly as cross-coupled enzyme but without boronic acid) to try to minimise differences in MS ionisation with respect to enzymes formed in the cross-coupling reaction itself, given the well-established detrimental effects of Pd in this regard.[29] The autoglucosylation activity of the samples thus obtained was then assessed as outlined previously and in chapter 4 (utilising LC-MS to establish average glucoses attached and proportions of cross-coupled enzyme remaining) and compared to the results generated previously for GYG-Maltohexaose (**figure 5.16**).

We were firstly pleased to find that key points made previously concerning GYG-Glucose and GYG-Maltose (chapter 4) held true under these altered conditions. Specifically, in the previous chapter, it was stated that GYG-Glucose exhibited a slower initial glucosylation step than GYG-Maltose, and that the latter exhibited a biphasic glucosylation profile (which must necessarily be encompassed within the profile for GYG-Glucose); these factors formed the basis for our proposal of a triphasic catalytic process. Both of these features were again observed (**figure 5.16**). Crucially, this in itself suggests that the different conditions utilised to generate GYG-Maltohexaose did not have significant effects upon enzymatic activity, and were unlikely to account for previously noted – and substantial – differences in the activity of this enzyme with respect to GYG-Glucose / GYG-Maltose (**see figure 5.14**). Whilst an increased amount of apparently inactive enzyme is observed (**figure 5.16**, right graph (GYG-Maltose), compare to **figure 5.14**, right graph (GYG-Maltose)), this minor difference does not alter any of the above facts.

Furthermore, the points previously outlined concerning comparison of GYG-Maltohexaose activity to that of GYG-Glucose and GYG-Maltose (**figure 5.14**) remained valid under these altered, and more similar, assay conditions. Thus, the linear glucosylation profile of the former continues to represent a contrast to those of GYG-Glucose and GYG-Maltose (**figure 5.16**, left graph). Similarly, the initial glucosylation step for GYG-Maltohexaose continues to, like that of GYG-Glucose, be markedly slower than that of GYG-Maltose (**figure 5.16**, right graph). This remains consistent with GYG-Maltohexaose falling exclusively within the final phase of a "slow → fast → slow" triphasic glucosylation profile depicted in **figure 5.13**.

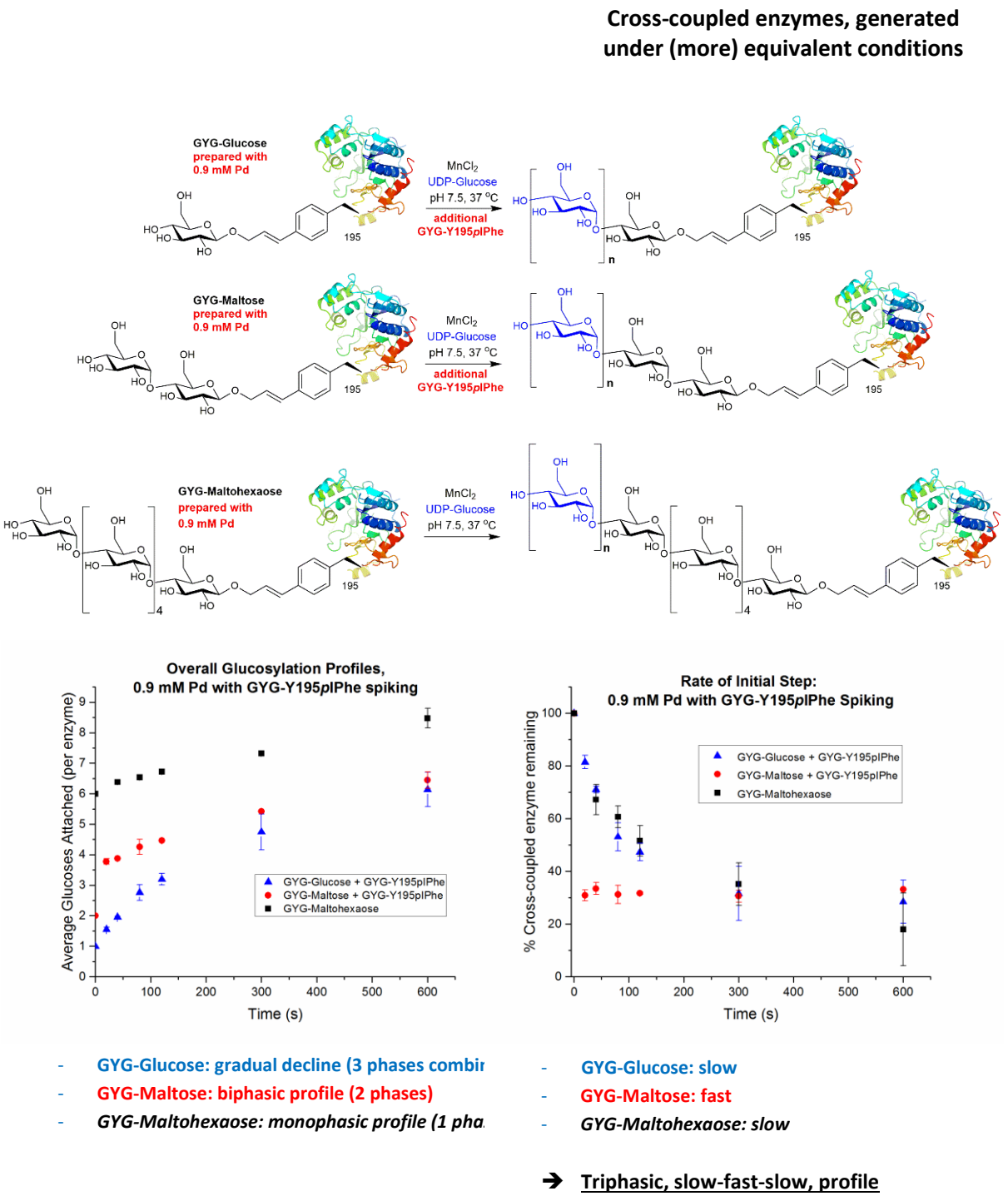


Figure 5.16 Comparison of autoglucosylation activity of GYG-Glucose, GYG-Maltose and GYG-Maltohexaose, under conditions designed to be as equivalent as possible. Average numbers of glucoses attached and proportion of cross-coupled enzyme remaining were determined from LC-MS analysis (see chapter 4 and experimental section). Data points represent mean averages of two independent replicates, error bars are ± s.d.

It must be acknowledged that the conditions under which the three enzymes are now compared may still, unavoidably, be slightly different. Spiking cross-coupled enzyme with "Pd-treated" GYG-Y195*p*IPhe is a reasonable means to mimic poor cross-coupling but is not identical, and proportions of GYG-Y195*p*IPhe may not be the same across samples (it is entirely possible that "Pd-treated" GYG-Y195*p*IPhe spiked into GYG-Glucose and GYG-Maltose samples still ionises with differing efficiency on MS to GYG-Y195*p*IPhe naturally present after GYG-maltohexaose cross-coupling).[iv] However, they are undeniably far more similar than previously (i.e. than comparisons in **figure 5.14**), and enable us to be more confident that differences in the glucosylation of GYG-Maltohexaose with respect to other enzymes do not reflect differing assay conditions. Indeed, the most important observation is arguably the very fact that, as noted, comparison of the glucosylation profiles of GYG-Glucose and GYG-Maltose to each other and to GYG-Maltohexaose, yielded the same essential results and conclusions with and without increased Pd loading and GYG-Y195*p*IPhe levels in the former two enzymes (compare **figure 5.14** and **figure 5.16**). This strongly indicates that the stark differences between the autoglucosylation properties of these three enzymes are highly unlikely to arise from such differences in reaction conditions, instead representing intrinsic differences between the enzymes that seemingly largely supersede such factors.

In summary, the autoglucosylation characteristics of GYG-Maltohexaose thus fit precisely into the triphasic mechanism originally defined from comparisons of GYG-Glucose and GYG-Maltose in the previous chapter, and, we believe, serve to further strengthen this proposal. As expected, access to the former enzyme results in a selective "trapping" of the final kinetic phase, as revealed by its monophasic glucosylation profile. This in turn implies that this is indeed one phase only, and that the preceding phases are complete prior to attachment of 6 glucose residues. The slow initial glucosylation step of GYG-Maltohexaose serves to further confirm that this is indeed a slow phase, consistent with the "slow → fast → slow" nature of the triphasic profile, which may now be directly visualised through comparison of catalytic activities of enzymes falling within each phase.

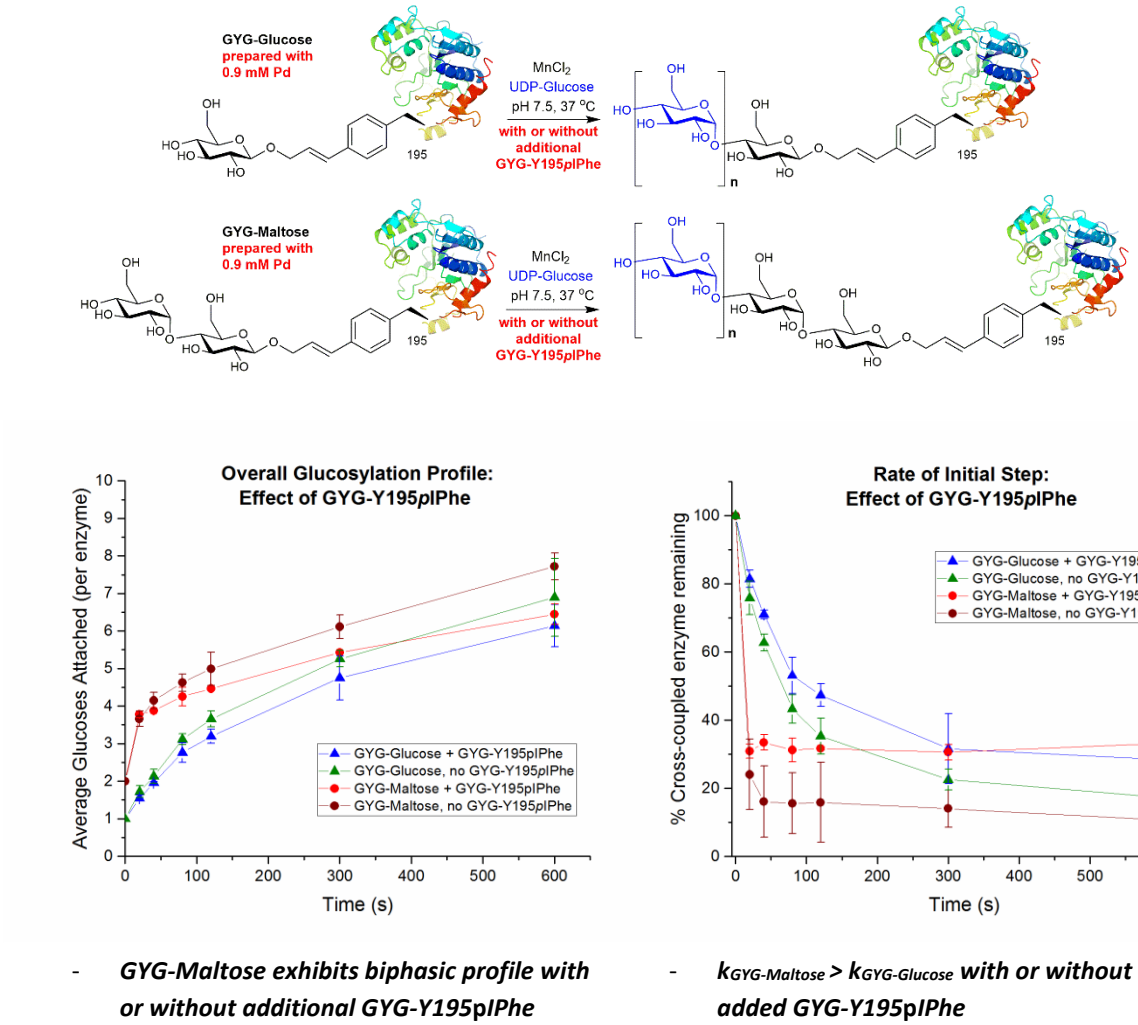
[iv] "Pd-treated" GYG-Y195*p*IPhe spiked into samples is, unlike GYG-Y195*p*IPhe remaining in GYG-Maltohexaose couplings, never exposed to boronic acid. By extension, the palladium species it is exposed to differ, especially given the believed role of boronic acid in reducing Pd(II) pre-catalyst to Pd(0).[22] The effectiveness of the Pd-scavenger may or may not be influenced by precise nature and oxidation state of the Pd complex,[32] and in turn, MS ionisation depends on the efficiency of Pd-scavenging from the protein.[29] Thus, "Pd-treated" GYG-Y195*p*IPhe may ionise differently to enzymes generated in genuine cross-coupling reactions.

5.4.1 Effect of Inactive Cross-coupling Side-product on Catalytic Activity of GYG-Maltose and GYG-Glucose

The preparations and assays of GYG-Glucose and GYG-Maltose described above involved both higher palladium loadings than prior assays (chapter 4) and increased levels of GYG-Y195*p*IPhe. Whilst these two combined changes did not alter any conclusions drawn regarding these two enzymes, we were interested to evaluate the effect of GYG-Y195*p*IPhe levels on subsequent activity in isolation, and appreciated that the data obtained above could be utilised to this effect. In particular, in the preceding chapter (section 4.3.4), the fact that conclusions drawn regarding activities of GYG-Glucose and GYG-Maltose were unlikely to be affected by the slightly higher proportions of cross-coupling side-products such as GYG-Y195*p*IPhe in preparations of the latter was discussed. Herein, further empirical data to support this claim is provided, utilising GYG-Y195*p*IPhe as a representative example of such a side-product.

The LC-MS-based timecourse data generated above for GYG-Glucose and GYG-Maltose, with GYG-Y195*p*IPhe spiking, was compared to new, analogous data generated from these enzymes prepared identically (i.e. same raised Pd loadings) but without any such spiking (**figure 5.17**). This experiment shows that the addition of extra GYG-Y195*p*IPhe appears to lead to a slight increase in amount of apparently non-reactive enzyme, as most clearly shown for GYG-Maltose (right graph). It is most likely this factor that accounts for the apparently slightly reduced glucosylation rates that were also observed, in terms of overall glucosylation (left graph) and rate of initial step (right graph, evident for GYG-Glucose), since this inactive enzyme is included in analysis.

However, differences are relatively minor in nature, and of most importance is the fact that the prior conclusions concerning these enzymes' autoglucosylation that formed the backbone of subsequent mechanistic proposals remain valid irrespective of whether additional GYG-Y195*p*IPhe has been added or not. Specifically, GYG-Glucose always exhibits a slower initial step than GYG-Maltose, whilst GYG-Maltose always exhibits a biphasic profile, with a significantly higher glucosylation rate in the first 20 seconds than thereafter. Whilst GYG-Y195*p*IPhe has here been utilised as an example of a cross-coupling side-product – and a highly relevant one, given that a principal difference between GYG-Glucose and GYG-Maltose preparations was low-level persistence of this species in the latter but not in the former – conclusions are likely applicable to the other cross-coupling side-products, given structural and functional similarities between them (see chapter 4: all differ solely at the 195 position, all cannot be glucosylated at this position yet all are potentially capable of catalysing glucosylation of another GYG substrate).



- ***GYG-Maltose exhibits biphasic profile with or without additional GYG-Y195pIPhe***
- ***$k_{GYG\text{-}Maltose} > k_{GYG\text{-}Glucose}$ with or without added GYG-Y195pIPhe***

Figure 5.17 Effect of additional GYG-Y195*p*IPhe on GYG-Glucose and GYG-Maltose glucosylation. Average numbers of glucoses attached and proportion of cross-coupled enzyme remaining were determined from LC-MS analysis (see chapter 4 and experimental section). Data points represent mean averages of at least two independent replicates, error bars are ± s.d. Lines are added purely to guide the eye.

5.5 Conclusion and Outlook

The work in this chapter investigated the potential of SMC as a tool for the conjugation of oligosaccharides to proteins in a site-specific fashion, in the context of generating a GYG-oligosaccharide conjugate for selective capturing of the proposed final catalytic phase. The synthesis of a challenging maltohexaose-derived boronic acid target was accomplished, applying and adapting methodology developed for analogous mono- and disaccharide systems. However, subsequent cross-coupling attempts to GYG-Y195*p*IPhe

yielded the desired product in low conversions under conditions compatible with subsequent enzymatic activity, owing to poor levels of starting material consumption. A variety of factors were investigated in efforts to improve this cross-coupling, however none proved successful in this regard.

Nonetheless, we were able to demonstrate the catalytic activity of the cross-coupled product GYG-Maltohexaose, with subsequent kinetic data suggesting that access to this enzyme did indeed represent a selective capturing of the final phase of the GYG catalytic mechanism. Comparison of its autoglucosylation activity to that of GYG-Glucose and GYG-Maltose, generated under as equivalent conditions as realistically possible, served to further emphasise the triphasic "slow-fast-slow" autoglucosylation mechanism for our system first proposed in the preceding chapter. Furthermore, similar comparisons yielded additional insight into the possible effects of increased levels of autoglucosylation-inactive cross-coupling side-products such as GYG-Y195*p*IPhe on subsequent catalytic activity.

Future work in this area should primarily focus on further optimisation of the cross-coupling. Whilst various factors were investigated in this regard with little success, and the conditions originally adopted themselves represented optimal conditions for both monosaccharide and disaccharide systems (and these, in turn, were based on the optimisation work on model systems carried out in chapter 2), other areas could be explored (although it should be noted that more fundamental parameters such as buffer composition, pH etc. have been extensively optimised previously within the Davis group).[22,27,29] In particular, the development of novel Pd catalysts for on-protein SMC could be pursued, with the aim of generating a catalyst exhibiting enhanced activity towards challenging substrates such as **5.1**. It is also possible that the coupling may proceed to a greater extent at higher temperatures, although this in turn may lead to problems with subsequent enzymatic activity, explaining why this was not pursued here. An alternative approach could instead be to attempt separation of cross-coupled enzyme from non-coupled enzyme, for example through the use of lectin affinity columns.

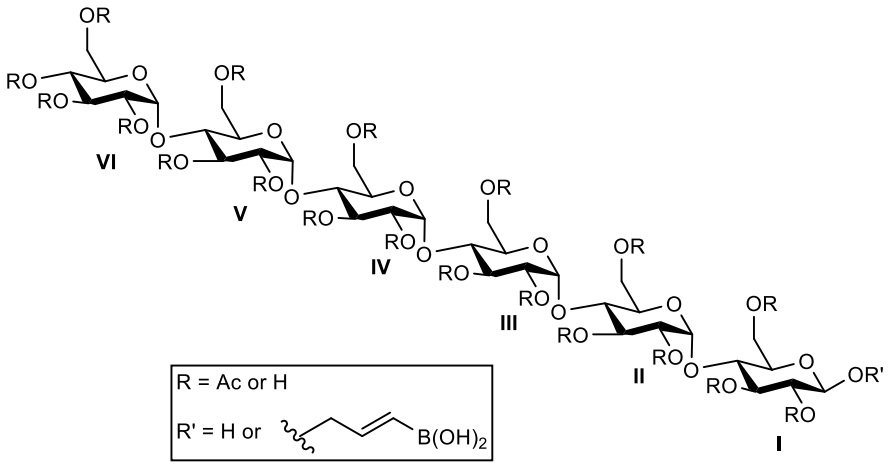
In general, it is apparent that cross-coupling efficiencies for this enzyme decline in the order GYG-Glucose > GYG-Maltose >> GYG-Maltohexaose. Whilst this may simply reflect increased steric hindrance and size of the boronic acid, we also found that GYG-Maltohexaose inhibited cross-couplings other than its own (that of 3-HMPB), suggesting that factors other than this may additionally be at play. Nonetheless, this highlights the fact that, at present, our methodology is most effective for the conjugation of mono- or disaccharides to GYG. In the following chapter, this area will be revisited, in the context of both different monosaccharide and different enzyme templates.

5.6 Experimental Section

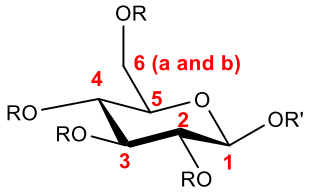
5.6.1 Chemical Synthesis

Note on NMR assignment nomenclature:

For the hexasaccharide substrates, superscript roman numerals are used to designate the ring from which a resonance belongs, according to the image below:

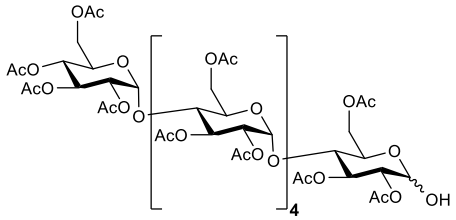


Numbers 1-6 are then used to designate the proton within a particular ring as below, according to standard convention:



For example, "$H1^{II}$" refers to the H1 proton within ring II above. "$C6^{V}$" refers to the C6 carbon in ring V above. "$H1^{II-V}$" would refer to the H1 protons from rings II-V *inclusive* (i.e. II, III, IV, V).

Any assignment refers to all listed resonances preceding it that are after the preceding assignment. E.g. "71.8, 71.7, 71.7 ($\underline{C3^{II-V}}$)" means that these three resonances are **all** accounted for by the C3 carbons from rings II-V inclusive.



5.6

Maltose eicosaacetate **5.2-OAc** (Carbosynth, 2.00 g, 1.09 mmol) was dissolved in anhydrous DMF (20 ml) and hydrazine monoacetate (150 mg, 1.63 mmol, 1.5 eq) was added. The mixture was stirred at 50 °C for 1 h, before being cooled to room temperature and diluted with EtOAc (ca. 70 ml). The organic layer was washed with brine (2 x 50 ml) then water (2 x 50 ml), before being dried ($MgSO_4$), filtered, and concentrated *in vacuo*. Purification by flash column chromatography (SiO_2, 1-8% MeOH in $CHCl_3$, linear gradient over 10 CV) yielded hemiacetal **5.6** as a white foam (1.15 g, 0.64 mmol, 85 %).

^{1}H NMR (400 MHz, Chloroform-*d*) δ 5.63 (dd, *J* = 10.0, 8.9 Hz, 1H, H3^{I}), 5.54 – 5.28 (m, 11H, H1^{I-VI} + H3^{II-VI}), 5.10 (t, *J* = 9.9 Hz, 1H, H4^{VI}), 4.89 (dd, *J* = 10.5, 4.0 Hz, 1H, H2^{VI}), 4.84 – 4.70 (m, 5H, H2^{I-V}), 4.62 – 4.49 (m, 6H, H6a^{I-VI}), 4.46 – 4.15 (m, 6H, H6b^{I-VI}), 4.14 – 3.89 (m, 11H, H4^{I-V} + H5^{I-VI}), 2.24-2.01 (14 x s, 57H, OAc).

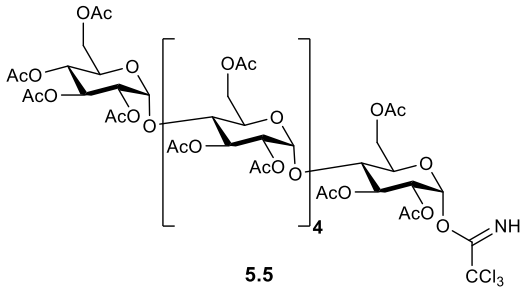
^{13}C NMR (101 MHz, $CDCl_3$) δ 170.9, 170.8, 170.8, 170.7, 170.6, 170.5, 170.4, 170.0, 169.7, 169.7, 169.6, 169.6 (OAc C=O), 95.8, 95.7 (C1^{II-VI}, C1^{I}β), 90.1 (C1^{I}α), 73.6, 73.5, 73.4, 73.3 (C4^{I-V}), 72.5, 72.4 (C3^{I}α, C3^{I}β), 71.8, 71.7, 71.7 (C3^{II-V}), 70.6, 70.5, 70.5 (C2^{I-V}), 70.1 (C2^{VI}), 69.5 (C3^{VI}), 69.0, 68.5 (C5^{I-VI}), 68.0 (C4^{VI}), 63.1, 63.0, 62.5, 62.4, 62.2, 61.4 (C6a^{I-VI}, C6b^{I-VI}), 21.0, 21.0, 21.0, 20.9, 20.8, 20.7, 20.7 (OAc CH_3).

LRMS *m/z* (ESI+) found: 1811 (M+Na, calculated 1811), 1789 (M+H, calculated 1789)

m.p. 133-136 °C ($CHCl_3$).

[α]$_D$ = +130 (c = 1.0, $CHCl_3$), lit.[17] = +140.

Data consistent with that previously reported.[17]



5.5

To a solution of hemiacetal **5.6** (600 mg, 0.34 mmol) in dry DCM (6.0 ml) was added trichloroacetonitrile (0.44 ml, 4.4 mmol, 13 eq.). The mixture was cooled to 0 °C and DBU (30 µl, 0.20 mmol, 0.6 eq.) was added. The mixture was stirred at room temperature for 3 hours and was then concentrated *in vacuo*. Purification by flash column chromatography (SiO_2, 50-100% EtOAc in Pet. Ether, linear gradient over 10 CV) yielded α-imidate **5.5** as an off-white solid (500 mg, 0.26 mmol, 76 %). Note that this purification should be carried out using a short column in order to minimise decomposition on silica. This compound should be stored at -20 °C and used within the next 1-2 days to prevent decomposition.

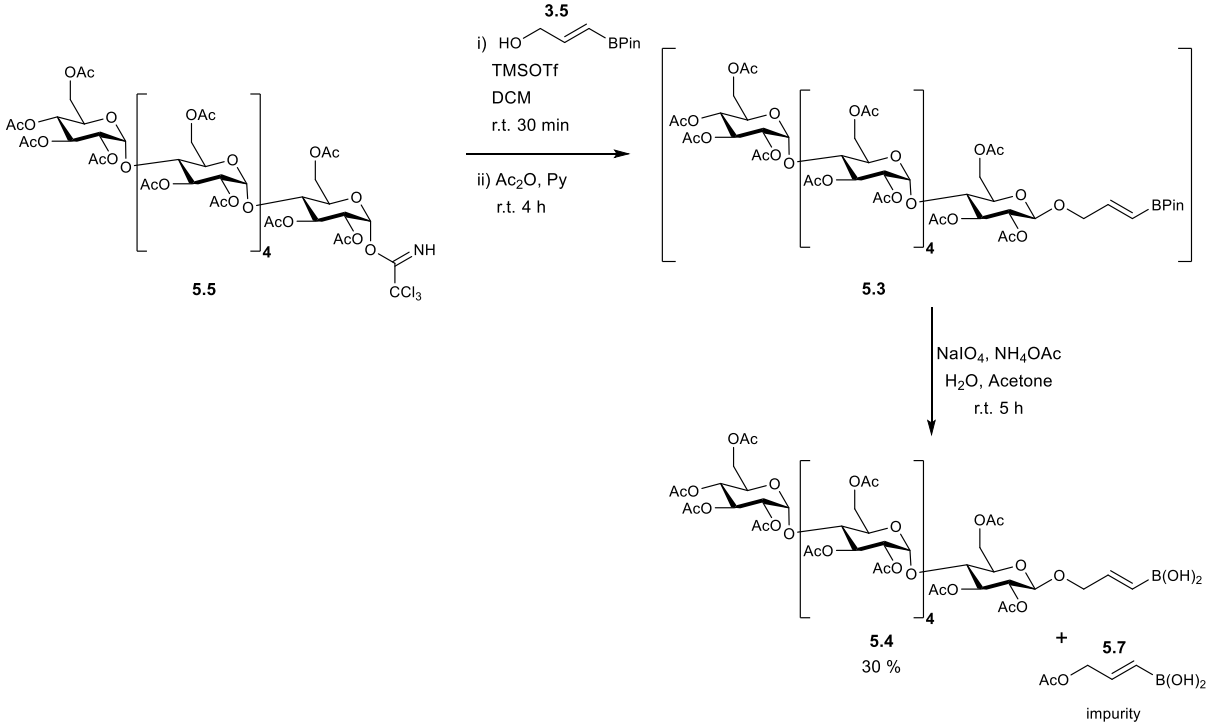
^{1}H NMR (400 MHz, Chloroform-*d*) δ 8.65 (s, 1H, NH), 6.46 (d, *J* = 3.8 Hz, 1H, $H1^{I}$), 5.57 (t, *J* = 9.4 Hz, 1H, $H3^{I}$), 5.47 – 5.22 (m, 10H, $H1^{II-VI}$ + $H3^{II-VI}$), 5.05 (t, *J* = 10.0 Hz, 1H, $H4^{VI}$), 4.99 (dd, *J* = 9.8, 3.8 Hz, 1H, $H2^{I}$), 4.83 (dd, *J* = 10.5, 4.0 Hz, 1H, $H2^{VI}$), 4.80 – 4.67 (m, 4H, $H2^{II-V}$), 4.514 – 4.43 (m, 5H, $H6a^{I-V}$), 4.39 – 3.85 (m, 16H, $H4^{II-V}$ + $H5^{I-VI}$ + $H6a^{VI}$ + $H6b^{I-VI}$), 4.00 (t, *J* = 9.5 Hz, 1H, $H4^{I}$) 2.18 – 2.06 (10 x s, 57H, OAc).

^{13}C NMR (101 MHz, $CDCl_3$) δ 170.8, 170.8, 170.6, 170.6, 170.5, 170.5, 170.4, 170.1, 169.9, 169.8, 169.7, 169.6, 169.6 (OAc C=O), 161.1 (C=NH), 96.0, 95.8, 95.7 ($C1^{II-VI}$), 92.9 ($C1^{I}$), 90.8 (CCl_3) 73.5, 73.3, 72.4 ($C4^{I-V}$), 72.0 ($C3^{I}$), 71.8, 71.7 ($C3^{II-V}$), 70.7 ($C5^{I}$), 70.6, 70.5, 70.5, 70.5 ($C2^{II-V}$), 70.2 ($C2^{I}$), 70.1 ($C2^{VI}$), 69.4 ($C3^{VI}$), 69.2, 69.0, 68.5 ($C5^{II-VI}$), 68.0 ($C4^{VI}$), 62.5, 62.4, 62.4, 62.2 ($C6^{I-V}$), 61.4 ($C6^{VI}$), 21.0, 20.9, 20.9, 20.8, 20.7, 20.6, 20.6, 20.4 (OAc CH_3).

LRMS (ESI+) found: 1956 (M+Na, calculated 1956)

$[\alpha]_D$ = 130 (c = 0.9, $CHCl_3$).

Data consistent with that previously reported.[36]



Formation of intermediate **5.3**

5.5 (860 mg, 0.45 mmol) and **3.5** (75 mg, 0.4 mmol, 0.9 eq.) were pre-dried through azeotropic removal of water with toluene (3 x) then high vacuum (> 1 h), and were then transferred to a flame-dried flask under Ar atmosphere as solutions in anhydrous CH_2Cl_2 (8.0 ml each, 16.0 ml total). Activated molecular sieves (3Å, ca. 2 g) were added and the solution stirred for 30 minutes. TMSOTf (20 µl, 0.11 mmol, 0.25 eq.) was added dropwise. The reaction was stirred at r.t. under Ar for 1 h, after which time it was quenched through the addition of pyridine (0.5 ml) and diluted with EtOAc (ca. 60 ml), then filtered through a pad of celite. The filtrate was washed successively with 1M HCl (50 ml), sat. $NaHCO_3$ (50 ml), and brine (30 ml), and was then dried ($MgSO_4$), filtered, and concentrated *in vacuo*. LC-MS analysis typically indicated low levels (ca. 10 – 15 %) of mono-deacetylation; a stark difference in ^{1}H NMR shifts for the H2^I proton before and after re-acetylation implicated the 2'-OH as the likely site of this:

δ H2^I (deacetylated): 3.47 (dd, *J* = 9.8, 3.6 Hz, α-anomer) and 3.40 (dd, *J* = 9.3, 7.6 Hz, β-anomer) (Chloroform-*d*, 400 MHz).

δ H2^I (acetylated): 4.83 (dd, *J* = 9.3, 7.6 Hz, β-anomer) and 4.77 – 4.73 (m, α-anomer, obscured by H2^{II-V} peaks, see below) (Chloroform-*d*, 400 MHz).

This mixture was therefore re-acetylated. The crude mixture (900 mg) was redissolved in anhydrous pyridine (2.7 ml) and Ac_2O (0.3 ml) was added dropwise at 0 °C. After stirring for 3 h at r.t., the reaction was diluted

with EtOAc (15 ml) and quenched with H_2O (ca. 15 ml). The organic layer was separated, and the aqueous extracted with EtOAc (3 x 30 ml). The combined organics were dried ($MgSO_4$) and concentrated *in vacuo*, then azeotroped with toluene (x 2) to remove pyridine and filtered through silica. Crude intermediate pinacol ester **5.3** and associated side-products were carried forward directly to subsequent hydrolysis step. The presence of **5.3** was confirmed by LRMS and ^{1}H NMR analysis of the crude mixture (note that some NMR peaks of **5.3** overlap with those of impurities, e.g. hydrolysed donor; peaks considered "diagnostic" for **5.3** are highlighted in bold): **^{1}H NMR** (400 MHz, Chloroform-*d*, only peaks corresponding to β-**5.3** are listed) δ **6.50 (dt, *J* = 18.3, 4.6 Hz, 1H, BCH=C<u>H</u>)**, **5.64 (dt, *J* = 18.3, 1.9 Hz, 1H, BC<u>H</u>=CH)**, 5.41 – 5.24 (m, 11H, <u>H3^{I-VI}</u> + <u>H1^{II-VI}</u>), 5.05 (t, *J* = 9.9 Hz, 1H, <u>H4^{VI}</u>), 4.83 (dd, *J* = 10.6, 4.1 Hz, 1H, <u>H2^{VI}</u>), **4.83 (dd, *J* = 7.8, 1.6 Hz, 1H, <u>H2^{I}</u>)**, 4.75 – 4.68 (m, 4H, <u>H2^{II-V}</u>), **4.54 (d, *J* = 7.8 Hz, 1H, <u>H1^{I}</u>)**, 4.50 – 4.44 (m, 5H, <u>H6a^{I-V}</u>), **4.37 (ddd, *J* = 14.9, 4.4, 1.8 Hz, 1H, BCH=CHC<u>H$_2$</u>)**, 4.33 – 4.09 (m, 8H, <u>H6b^{I-V}</u> + <u>H6a^{VI}</u> + BCH=CHC<u>H$_2$</u>), 4.07 – 3.84 (m, H, <u>H4^{I-V}</u> + <u>H5^{II-VI}</u> + <u>H6b^{VI}</u>), **3.66 (dt, *J* = 9.8, 3.7 Hz, 1H, <u>H5^{I}</u>)**, 2.21 – 1.88 (m, 57H, O<u>Ac</u>); **LRMS** *m/z* (ESI+) 1955.3 (M+H, calculated 1955.6), 1977.3 (M+Na, calculated 1977.6).

Note that anomeric ratio could not be quantitatively determined at this stage owing to extensive NMR peak overlap (with impurities), although from 2D NMR analysis it was apparent that the β-anomer was the major species. It was instead determined following the subsequent deprotection steps, however it can be inferred that it is similar after this step, since anomeric ratio is solely influenced by the glycosylation step itself.

<u>Formation of **5.4** (with impurity attributed to **5.7**)</u>

The crude product **5.3** (850 mg total) was dissolved in acetone (10 ml) and 1M NH_4OAc (7 ml) added. $NaIO_4$ (260 mg, 1.2 mmol) was added and the suspension stirred vigorously at room temperature for 5 hours. The acetone was removed *in vacuo* and the aqueous layer carefully acidified to pH 3 with 1M HCl before being extracted with EtOAc (3 x 40 ml). The combined organic layers were dried ($MgSO_4$), filtered and concentrated *in vacuo*. The residue was purified by flash column chromatography (SiO_2, 25-30-40% acetone in $CHCl_3$, then 4-6-10% MeOH in $CHCl_3$ (1 column)). Fractions containing boronic acid **5.4** were combined and concentrated *in vacuo*. This yielded **5.4** as a white solid as predominantly the β-anomer (91 % β), contaminated with a species that NMR analysis indicated to be **5.7** (presence of acetyl group was confirmed by HMBC analysis). This mixture was carried forward directly to the subsequent synthetic step, in which the impurity could be trivially removed. Yield of **5.4**: 260 mg total of which NMR indicated 242 mg **5.4** and 18 mg **5.7**, 0.13 mmol **5.4**, 32 % (based on limiting reagent **3.5**). ^{1}H NMR analysis indicated

that the product **5.4** was formed as 90 % β-anomer, 10 % α-anomer (determined through integration of the $H1^{I}{}_{\alpha}$ peak (δ 5.00, *J* = 3.6 Hz) and the $H5^{I}$ peak (both anomers: δ 3.92, dt, *J* = 9.3, 3.7 Hz)).

Presence of **5.4** was confirmed through ^{1}H and ^{13}C NMR, IR and HRMS analysis of the mixture. This characterisation of the mixture of **5.4** and **5.7** is shown below. NMR assignments below refer to the product **5.4** except for those explicitly listed as arising from **5.7**

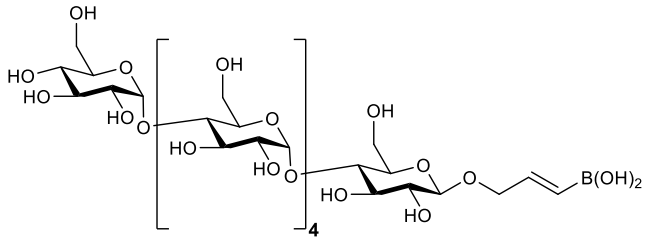
^{1}H NMR (700 MHz, Acetone-d_6, peaks for major β-anomer of **5.4** only and impurity **5.7** reported) δ 6.87 (s, 2H, $B(OH)_2$), δ 6.82 (s, 2H, $B(OH)_2$ of **5.7**), 6.52 (dt, *J* = 17.8, 4.8 Hz, 1H, BCH=CH of **5.7**), 6.51 (dt, *J* = 18.1, 4.8 Hz, 1H, BCH=CH), 5.66 (dt, *J* = 17.8, 1.7 Hz, 1H, BCH=CH), 5.65 (dt, *J* = 17.8, 1.8 Hz, 1H, BCH=CH of **5.7**), 5.46 – 5.39 (m, 4H, $H3^{II-V}$), 5.37 (d, *J* = 3.9 Hz, 1H, $H1^{VI}$), 5.33 (t, *J* = 10.0 Hz, 1H, $H3^{VI}$), 5.30 (t, *J* = 9.1 Hz, 1H, $H3^{I}$), 5.30 – 5.28 (m, 5H, $H1^{I-V}$), 5.04 (t, *J* = 9.9 Hz, 1H, $H4^{VI}$), 4.85 (dd, *J* = 10.6, 3.9 Hz, 1H, $H2^{VI}$), 4.81 (dd, *J* = 9.2, 8.0 Hz, 1H, $H2^{I}$), 4.79 (d, *J* = 7.9 Hz, 1H, $H1^{I}$), 4.78 – 4.74 (m, 4H, $H2^{II-V}$), 4.58 (dd, *J* = 4.9, 1.8 Hz, 2H, CH_2 from **5.7**), 4.51 (dd, *J* = 12.4, 2.3 Hz, 1H, $H6a^{I}$), 4.50 – 4.44 (m, 4H, $H6a^{II-V}$), 4.38 (dd, *J* = 12.7, 4.2 Hz, 1H, $H6b^{I}$), 4.35 (ddd, *J* = 14.4, 4.5, 1.8 Hz, 1H, BCH=CHCH_2), 4.38 – 4.32 (m, 2H, two of $H6b^{II-V}$), 4.31 (dd, *J* = 12.1, 3.4 Hz, 1H, one of $H6b^{II-V}$), 4.28 (dd, *J* = 12.3, 3.9 Hz, 1H, 1 of $H6b^{II-V}$), 4.22 (dd, *J* = 12.6, 4.4 Hz, 1H, $H6a^{VI}$), 4.17 (ddd, *J* = 14.4, 5.1, 1.7 Hz, 1H, BCH=CHCH_2), 4.14 – 4.00 (9H, $H5^{II-VI}$ + $H4^{II-V}$), 4.07 (dd, *J* = 12.2, 2.0 Hz, 1H, $H6b^{VI}$), 4.04 (t, *J* = 9.3 Hz, 1H, $H4^{I}$), 3.92 (dt, *J* = 9.6, 3.8 Hz, 1H, $H5^{I}$), 2.15 – 1.96 (19 x s, 57H, OAc), 2.03 (s, 3H, OAc of **5.7**).

^{13}C NMR (176 MHz, Acetone-d_6, peaks for major β-anomer of **5.4** only and impurity **5.7** reported) δ 171.0, 171.0, 171.0, 171.0, 170.9, 170.9, 170.9, 170.8, 170.7, 170.5, 170.5, 170.3, 170.3, 170.1, 170.0, 169.9 (OAc C=O from **5.4** and **5.7**), 145.7 (BCH=CH), 144.1 (BCH=CH from **5.7**), 125.8 (br weak, BCH=CH from **5.4**), 125.8 (br weak, BCH=CH from **5.7**), 100.2 ($C1^{I}$), 96.9, 96.9, 96.8, 96.8 ($C1^{II-VI}$), 75.9 ($C3^{I}$), 75.4, 75.2, 75.1 ($C4^{II-V}$), 74.4 ($C4^{I}$), 72.9 ($C5^{I}$), 72.9 ($C2^{I}$), 72.3, 72.2, 72.2, 72.2 ($C3^{II-V}$), 71.4, 71.4, 71.3 ($C2^{II-V}$ + BCH=CHCH_2), 70.9 ($C2^{VI}$), 70.1, 70.0, 70.0, 69.4, 69.0 ($C4^{VI}$), 66.2 (CH_2 of **5.7**), 64.0, 63.9, 63.8, 63.7, 63.5 ($C6^{I-V}$) 62.4 ($C6^{VI}$), 21.1, 21.0, 21.0, 20.9, 20.9, 20.8, 20.8, 20.8, 20.7, 20.7, 20.6, 20.6, 20.5, 18.9 (OAc CH_3 from **5.4** and **5.7**).

IR (chloroform) $\tilde{\nu}$ = 2961 (C-H), 1747 (C=O), 1647 (C=C) cm^{-1}

HRMS *m/z* (ESI+) found: 1895.5507 (**5.4**, M+Na, calculated 1895.5557)

*Owing to the presence of impurity **5.7**, $[a]_D$ and melting point were not recorded.*



5.1

Crude boronic acid **5.4** (50 mg, of which 4 mg impurity **5.7**, 24 μmol) was dissolved in anhydrous MeOH (2 ml). NaOMe (7.5 mg, 32 μl of 25 wt % solution in MeOH, 140 μmol) was added to give a final concentration of 0.07M and the reaction was stirred at room temperature. After 5 h LC-MS indicated completion of the reaction. The mixture was diluted with MeOH to ca. 5 ml and pre-activated DOWEX-H^+ resin was added until pH neutral. After stirring for 30 minutes, the resin was removed by filtration and washed thoroughly with MeOH, and the eluate concentrated *in vacuo* to give 28 mg crude material. This was taken up in H_2O (1.5 ml) and syringe filtered (0.2 μm filter), then concentrated to give **5.1** along with impurity **5.8**, the de-acetylated variant of **5.7** (see figure 5.6).This small molecule impurity was removed by PD-Miditrap G10 (GE Healthcare) according to manufacturer instructions; the desired product typically elutes after addition of ca. 1.6-3.0 ml H_2O, whereas the small molecule impurity is retained. Fractions containing **5.1** were combined and concentrated *in vacuo* to yield **5.1** as an off-white solid (16 mg, 15 μmol, 63 %) as predominantly the β anomer (88 % β). Anomeric ratio was determined through integration of the $H1^I$ protons for each anomer: $H1^I_\beta$: δ 4.46 (d, J =8.0 Hz); $H1^I_\alpha$: δ 4.91 (d, J = 3.8 Hz) (D_2O).

(Note that ^{1}H NMR was recorded and assigned at both 400 MHz and 700 MHz; both are reported below. Whilst the latter reduces peak overlap and thus improves assignment quality, it is understood that the former is typically more readily accessible; hence, if desired, identity of **5.1** might be confirmed more rapidly through comparison of 400 MHz spectra to that reported here).

^{1}H NMR (700 MHz, D_2O, only resonances for the major β-anomer are reported) δ 6.48 (dt, J = 18.3, 5.1 Hz, 1H, BCH=C<u>H</u>), 5.70 (dt, J = 18.3, 1.7 Hz, 1H, BC<u>H</u>=CH), 5.35 (app d, J = 3.6 Hz, 5H, <u>$H1^{II-VI}$</u>), 4.46 (d, J = 8.0 Hz, 1H, <u>$H1^I$</u>), 4.41 (ddd, J = 14.2, 5.1, 1.7 Hz, 1H, BCH=CHC<u>H_2</u>), 4.27 (ddd, J = 14.3, 5.4, 1.7 Hz, 1H, BCH=CHC<u>H_2</u>), 3.90 (2 x app t, J = 9.5 Hz, 4H, <u>$H3^{II-V}$</u>), 3.88 (dd, J = 12.4, 2.1 Hz, 1H, <u>$H6a^I$</u>), 3.80 – 3.69 (m, 15H, <u>$H6a^{II-VI}$</u> + <u>$H6b^{I-VI}$</u> + <u>$H5^{II-V}$</u>), 3.72 (t, J = 9.3 Hz, 1H, <u>$H3^I$</u>), 3.66 (ddd, J = 10.1, 5.2, 2.2 Hz, 1H, <u>$H5^{VI}$</u>), 3.63 (t, J = 9.7 Hz, 1H, <u>$H3^{VI}$</u>), 3.63-3.59 (m, 5H, <u>$H4^{I-V}$</u>), 3.58-3.55 (m, 4H, <u>$H2^{II-V}$</u>), 3.53 (app dd, J = 10.1, 3.9 Hz, <u>$H2^{VI}$</u> + <u>$H5^I$</u>), 3.36 (t, J = 9.5 Hz, 1H, <u>$H4^{VI}$</u>), 3.29 (dd, J = 9.5, 8.0 Hz, 1H, <u>$H2^I$</u>).

^{1}H NMR (400 MHz, D_2O, only resonances for the major β-anomer are reported) δ 6.48 (br d, *J* = 18.1 Hz, 1H, BCH=C<u>H</u>), 5.71 (dt, *J* = 18.5, 1.4 Hz, 1H, BC<u>H</u>=CH), 5.35 (app d, *J* = 3.6 Hz, 5H. <u>H1^{II-VI}</u>), 4.46 (d, *J* = 8.0 Hz, 1H, <u>H1^{I}</u>), 4.41 (ddd, *J* = 14.9, 4.8, 1.7 Hz, 1H, BCH=CHC<u>H$_2$</u>), 4.27 (ddd, *J* = 14.8, 5.4, 1.6 Hz, 1H, BCH=CHC<u>H$_2$</u>), 3.90 (app t, *J* = 9.5 Hz, 4H, <u>H3^{II-V}</u>), 3.88 (dd, *J* = 12.0, 1.6 Hz, 1H, <u>H6a^{I}</u>), 3.84 – 3.68 (m, 15H, <u>H6a^{II-VI}</u> + <u>H6b^{I-VI}</u> + <u>H5^{II-V}</u>), 3.75 – 3.71 (m, 1H, <u>H3^{I}</u>), 3.68 – 3.48 (m, 12H, <u>H5^{I}</u> + <u>H5^{VI}</u> + <u>H4^{I-V}</u> + <u>H2^{II-VI}</u>), 3.63 (t, *J* = 9.6 Hz, 1H, <u>H3^{VI}</u>), 3.36 (t, *J* = 9.4 Hz, 1H, <u>H4^{VI}</u>), 3.29 (dd, *J* = 9.3, 8.2 Hz, 1H, <u>H2^{I}</u>).

^{13}C NMR (176 MHz, D_2O, only resonances for the major β-anomer are reported) δ 145.2 (BCH=<u>C</u>H), 101.2 (<u>C1^{I}</u>), 99.6, 99.5, 99.4, 99.3 (<u>C1^{II-VI}</u>), 76.6, 76.6, 76.5, 76.5, 76.4 (<u>C4^{I-V}</u>), 76.1 (<u>C3^{I}</u>), 74.4 (<u>C5^{I}</u>), 73.2, 73.2, 73.2 (<u>C3^{II-V}</u>), 72.9 (<u>C2^{I}</u>), 72.7 (<u>C3^{VI}</u>), 72.6 (<u>C5^{VI}</u>), 71.6 (<u>C2^{VI}</u>), 71.4, 71.3 (<u>C2^{II-V}</u>), 71.1, 71.0 (<u>C5^{II-V}</u> + BCH=CH<u>C</u>H$_2$), 69.2 (<u>C4^{VI}</u>), 60.5 (<u>C6^{I}</u>), 60.3, 60.3 (<u>C6^{II-VI}</u>),

IR (neat) $\tilde{\nu}$ = 3291 (br, O-H), 2924 (C-H), 1643 (C=C) cm^{-1}

HRMS *m/z* (ESI+) found 1097.3538 (M+Na, calculated 1097.3549).

m.p. > 300 °C (H_2O).

$[\alpha]_D$ = +143 (*c* = 0.2, H_2O)

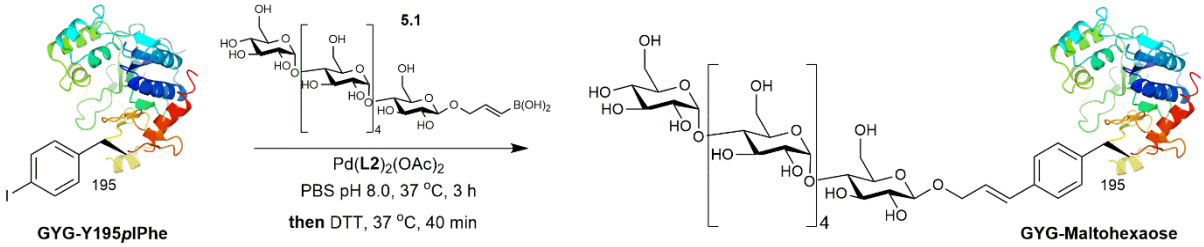
The identity of impurity **5.8**, the deprotected variant of **5.7**, (**figure 5.6**) was indicated by ^{1}H NMR analysis of the crude mixture prior to purification by PD Miditrap G-10. Its resonances are listed below:

^{1}H NMR (400 MHz, D_2O, resonances for impurity **5.8**) δ 6.54 (dt, *J* = 18.4, 4.2 Hz, 0H), 5.60 (dt, *J* = 18.3, 1.9 Hz, 0H), 4.14 (dd, *J* = 4.3, 1.9 Hz, 1H).

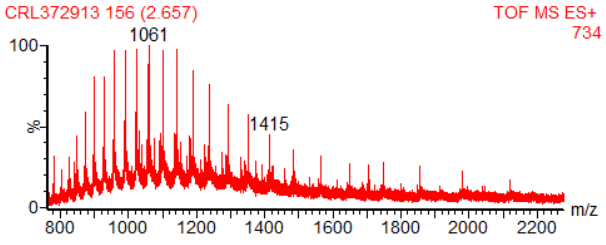
These are consistent with previously reported data.[30]

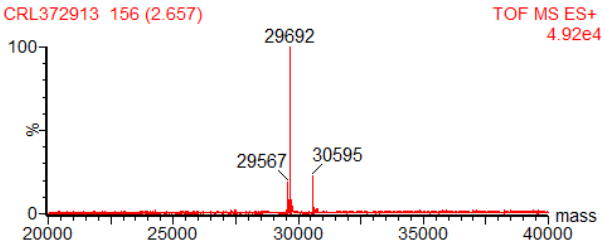
5.6.2 Cross-coupling Studies

5.6.2.1 Initial Cross-coupling of GYG-Y195pIPhe to maltohexaose boronic acid 5.1



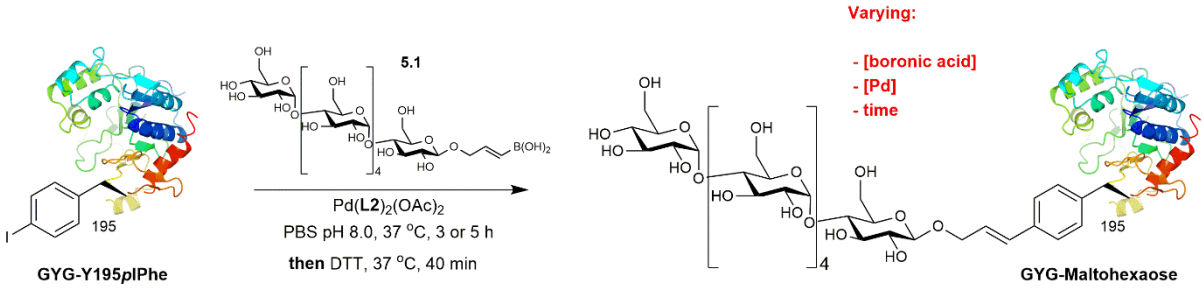
A stock solution of maltohexaose boronic acid **5.1** was prepared by dissolving boronic acid (110 mg/ml) in H_2O. To non-his-tagged GYG-Y195*p*IPhe (10 µl of 0.85 mg/ml, 0.29 nmol) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was added 10 µl of this stock solution (1020 nmol, 3500 eq. boronic acid), followed by $Pd(L2)_2(OAc)_2$ catalyst (2.0 µl of 10 mM solution, 20 nmol, 0.9 mM concentration). The reaction mixture was briefly vortexed, then incubated at 37 °C with shaking (550 rpm). After 3 hours, DTT was added to a final concentration of 12 mM (from a 115 mM solution in H_2O), the mixture shaken for a further 40 minutes, then analysed by LC-MS.





M_{calc}: 30594
M_{obs}: 30595 (16 %)
29567 (de-iodination)
29692 (GYG-Y195*p*IPhe)

5.6.2.2 Cross-coupling optimisations from table 5.1: Boronic Acid, Pd, Time

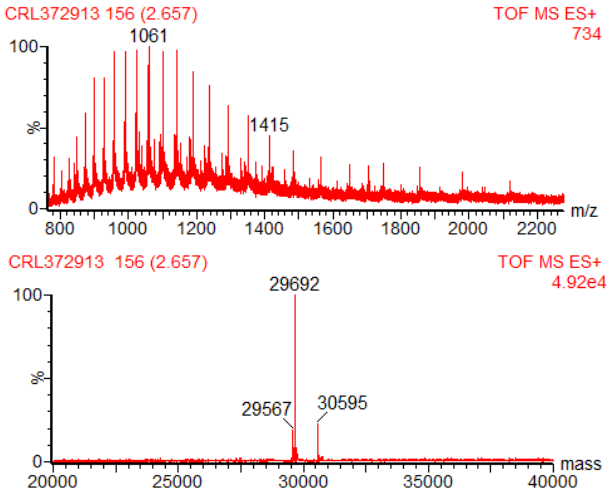


A stock solution of maltohexaose boronic acid **5.1** was prepared by dissolving boronic acid (110 mg/ml) in H_2O. 3 reactions were set up by combining non-his-tagged GYG-Y195*p*IPhe in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl), boronic acid stock solution and $Pd(L2)_2(OAc)_2$ catalyst in amounts indicated below.

Entry	Amount $B(OH)_2$	Amount $Pd(L2)_2(OAc)_2$	Amount GYG-Y195*p*IPhe	Time
1	10 µl stock solution, 1020 nmol, 3500 eq.	2.0 µl of 10 mM solution, 20 nmol, 0.9 mM	10.0 µl of 0.85 mg/ml, 0.29 nmol	3 h
2	10 µl stock solution, 1020 nmol, 3500 eq.	2.5 µl of 10 mM solution, 25 nmol, 1.1 mM	10.0 µl of 0.85 mg/ml, 0.29 nmol	3 h
3	14 µl stock solution, 1430 nmol, 5000 eq.	2.0 µl of 10 mM solution, 20 nmol, 0.9 mM	6.0 µl of 1.42 mg/ml, 0.29 nmol	3 h
4	14 µl stock solution, 1430 nmol, 5000 eq.	2.0 µl of 10 mM solution, 20 nmol, 0.9 mM	6.0 µl of 1.42 mg/ml, 0.29 nmol	5 h

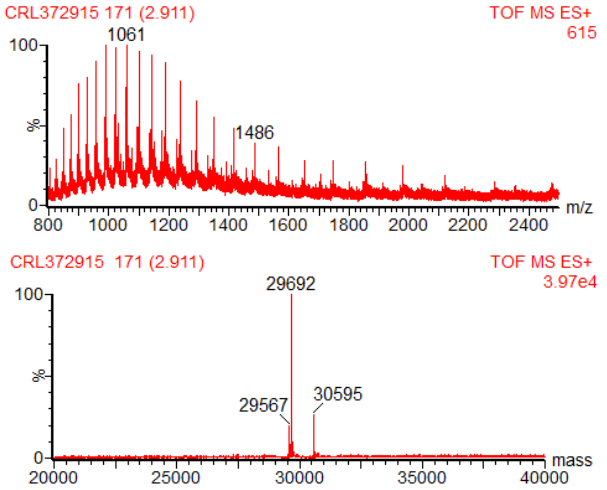
Reaction were briefly vortexed, then incubated at 37 °C with shaking (550 rpm). After the indicated time, DTT was added to an aliquot of each at a final concentration of 12 mM (from a 115 mM solution in H_2O), the aliquots shaken for a further 40 minutes, then analysed by LC-MS.

Entry 1 (Table 5.1 entry 1)



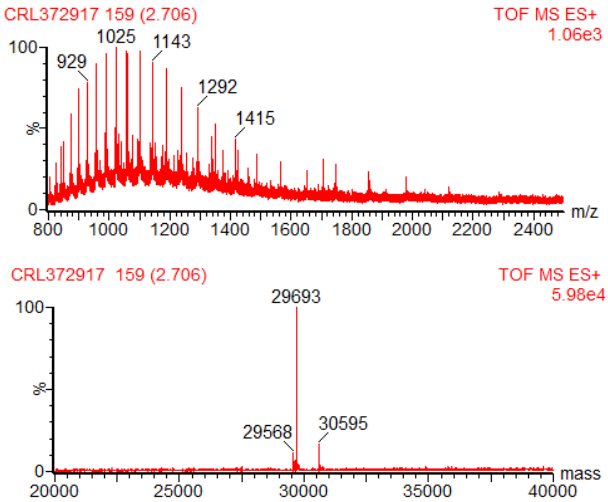
M_{calc}: 30594
M_{obs}: 30595 (16 %)
29567 (de-iodination)
29692 (GYG-Y195*p*IPhe)

Entry 2 (Table 5.1 entry 2, more Pd)



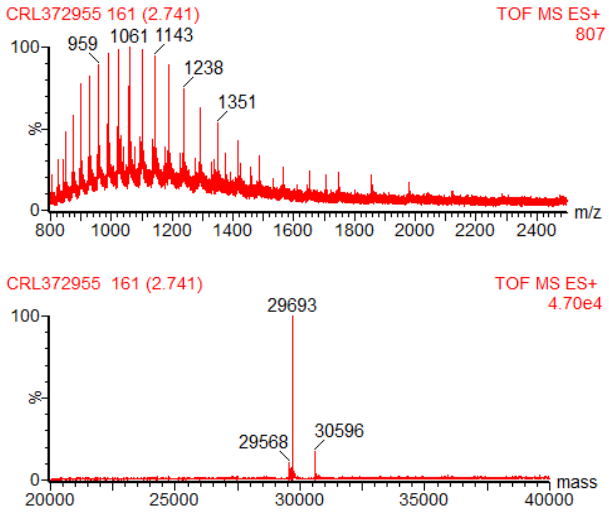
M_{calc}: 30594
M_{obs}: 30595 (18 %)
29567 (de-iodination)
29692 (GYG-Y195*p*IPhe)

Entry 3 (Table 5.1 entry 3, more boronic acid)



M_{calc}: 30594
M_{obs}: 30595 (13 %)
29568 (de-iodination)
29693 (GYG-Y195*p*IPhe)

Entry 4 (Table 5.1 entry 4, increased time)



M_{calc}: 30594

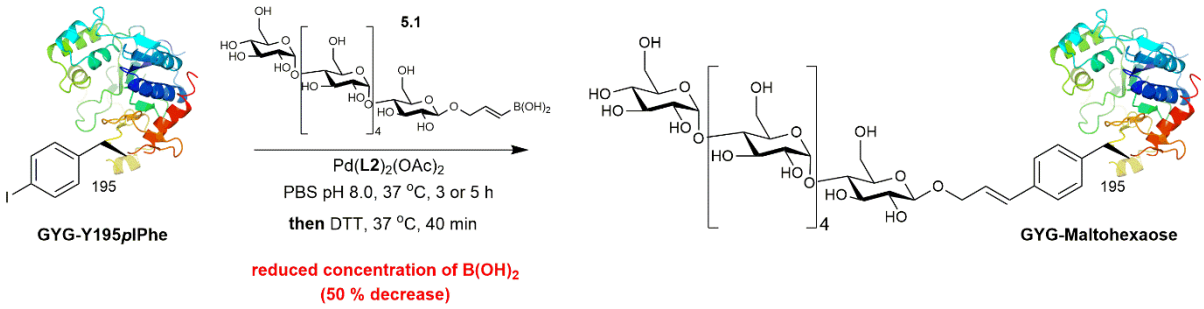
M_{obs}: 30596 (14 %)

29568 (de-iodination)

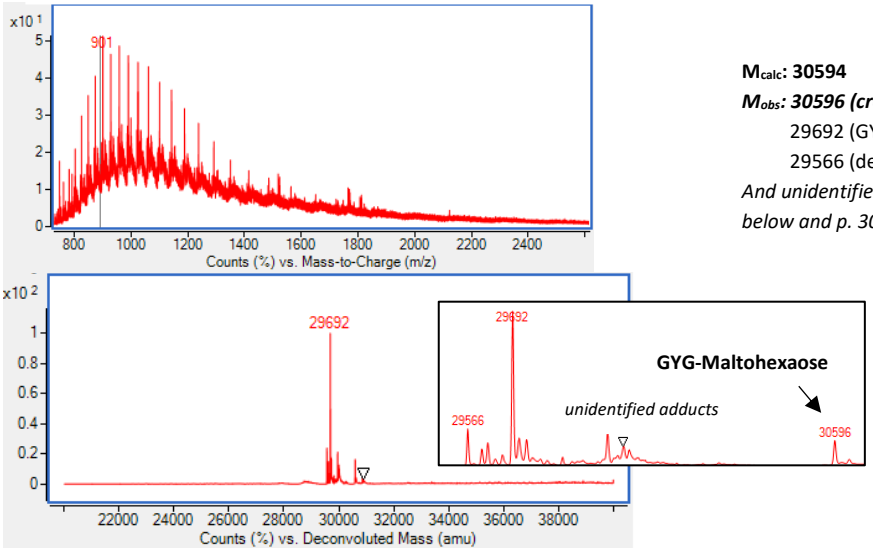
29693 (GYG-Y195*p*IPhe)

5.6.2.3 Further Cross-coupling Optimisation Studies

a) Lowering Boronic Acid



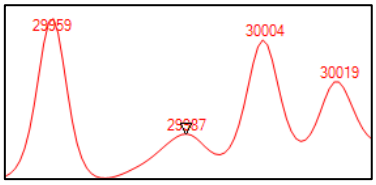
Reaction was carried out exactly as described in section 5.6.2.1, but using 5 µl boronic acid stock solution (510 nmol, 1750 eq.) rather than 10 µl; a further 5 µl of pH 8.0 phosphate buffer was also added to maintain the same reaction volume as used in section 5.6.2.1.



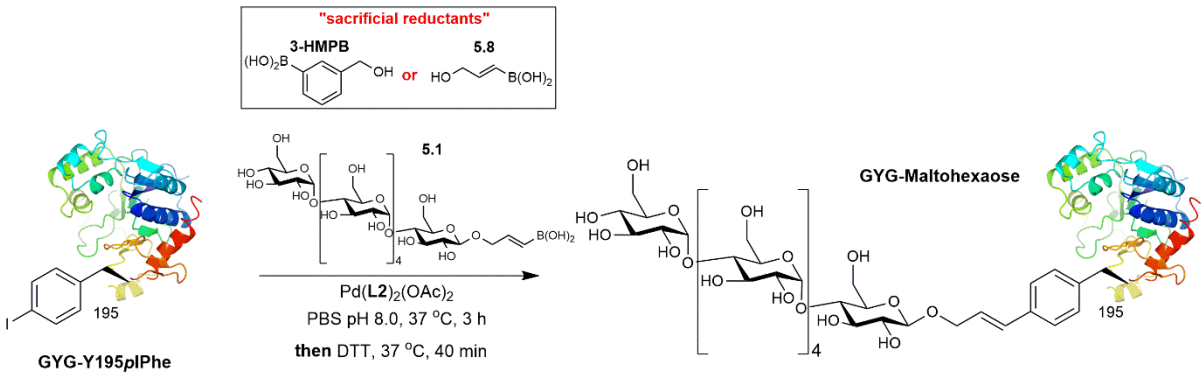
M_{calc}: 30594
M_{obs}: 30596 (cross-coupled)
29692 (GYG-Y195*p*IPhe)
29566 (de-iodination)
And unidentified adducts (29959 - 30019 , see note below and p. 307.

Note: LC-MS analysis of a range of protein samples during this time period was affected by the appearance of various adducts on the machine available, which we believe to result from the machine itself; such adducts were observed by other group members on different protein samples through this time period. Most notably, M + 266/7 Da adducts were repeatedly seen (the 29959 peak above = 29692 + 267, similar small peaks were seen for other species). Whilst this may preclude reliable quantitative analysis of reaction conversion, it does not affect the key observation that GYG-Maltohexaose levels remained very low. For this reason, this avenue was not pursued further. Natures of such adducts are illustrated below:

Adducts from above spectrum:



b) Effect of "Sacrificial" Boronic Acid as Putative Reductant



2 reactions were set up. Reactions were prepared according to the general protocol below, but with the following single differences:

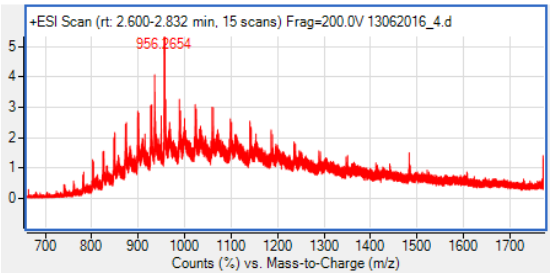
Reaction 1: the boronic acid stock solution was prepared using boronic acid **5.1** that had *not* been purified via PD-Miditrap G10, and therefore contained an approximately equimolar proportion of impurity **5.8** (see **figure 5.6**).

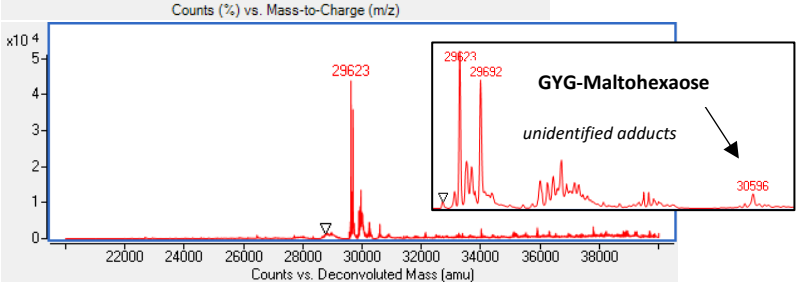
Reaction 2: an additional 350 eq. of boronic acid **3-HMPB** (1 µl of 15.6 mg/ml boronic acid stock solution in H_2O) was added 30 minutes following addition of Pd catalyst.

General protocol:

Boronic acid **5.1** stock solution was prepared by dissolving boronic acid (110 mg/ml) in H_2O. To non-his-tagged GYG-Y195*p*IPhe (10 µl of 0.85 mg/ml, 0.29 nmol) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was added 10 µl of this stock solution (1020 nmol, 3500 eq. boronic acid), followed by $Pd(L2)_2(OAc)_2$ catalyst (2.0 µl of 10 mM solution, 20 nmol, 0.9 mM concentration). The reaction mixtures were briefly vortexed, then incubated at 37 °C with shaking (550 rpm). After 3 hours, DTT was added to a final concentration of 12 mM (from a 115 mM solution in H_2O), the mixtures shaken for a further 40 minutes, then analysed by LC-MS.

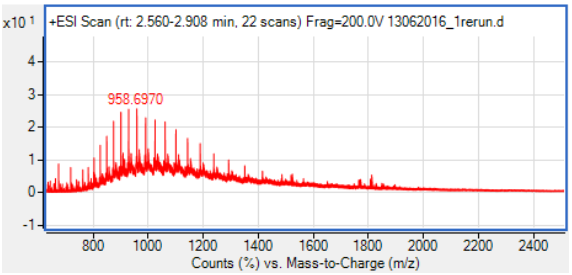
Reaction 1:

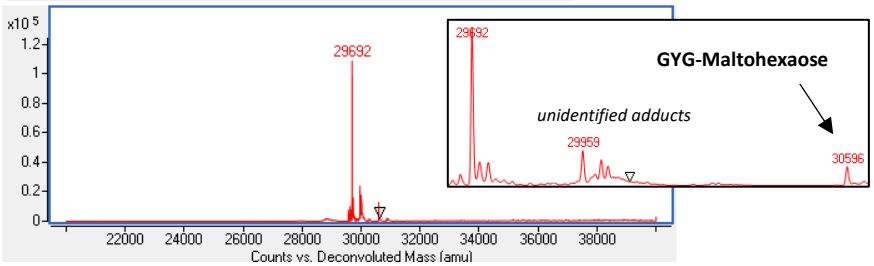




M_{calc}: 30594
M_{obs}: 30596 (cross-coupled)
29692 (GYG-Y195*p*IPhe)
29623 (coupling to impurity **5.8**)
+ various unidentified MS adducts (29890 – 30004), see below and p. 307

Reaction 2:

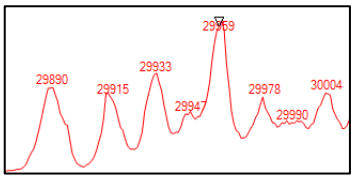




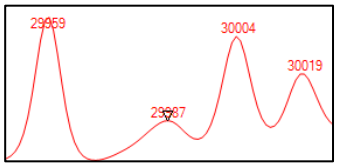
M_{calc}: 30594
M_{obs}: 30596 (cross-coupled)
29692 (GYG-Y195*p*IPhe)
+ unidentified MS adducts (29959 – 30019), see below and p. 307

As noted previously and in the main text (p. 307), LC-MS analysis of a range of protein samples during this time period was affected by the appearance of various adducts on the machine available, seemingly arising from the machine itself (although in reaction 1 it may arise from use of crude boronic acid). Notably, M + 266/7 Da adducts were repeatedly seen (the 29959 peak above = 29692 + 267, similar small peaks were seen for other species). Whilst this may preclude reliable quantitative analysis of reaction conversion, it does not affect the key observation that GYG-Maltohexaose levels remained very low. For this reason, this avenue was not pursued further. These adducts are illustrated below:

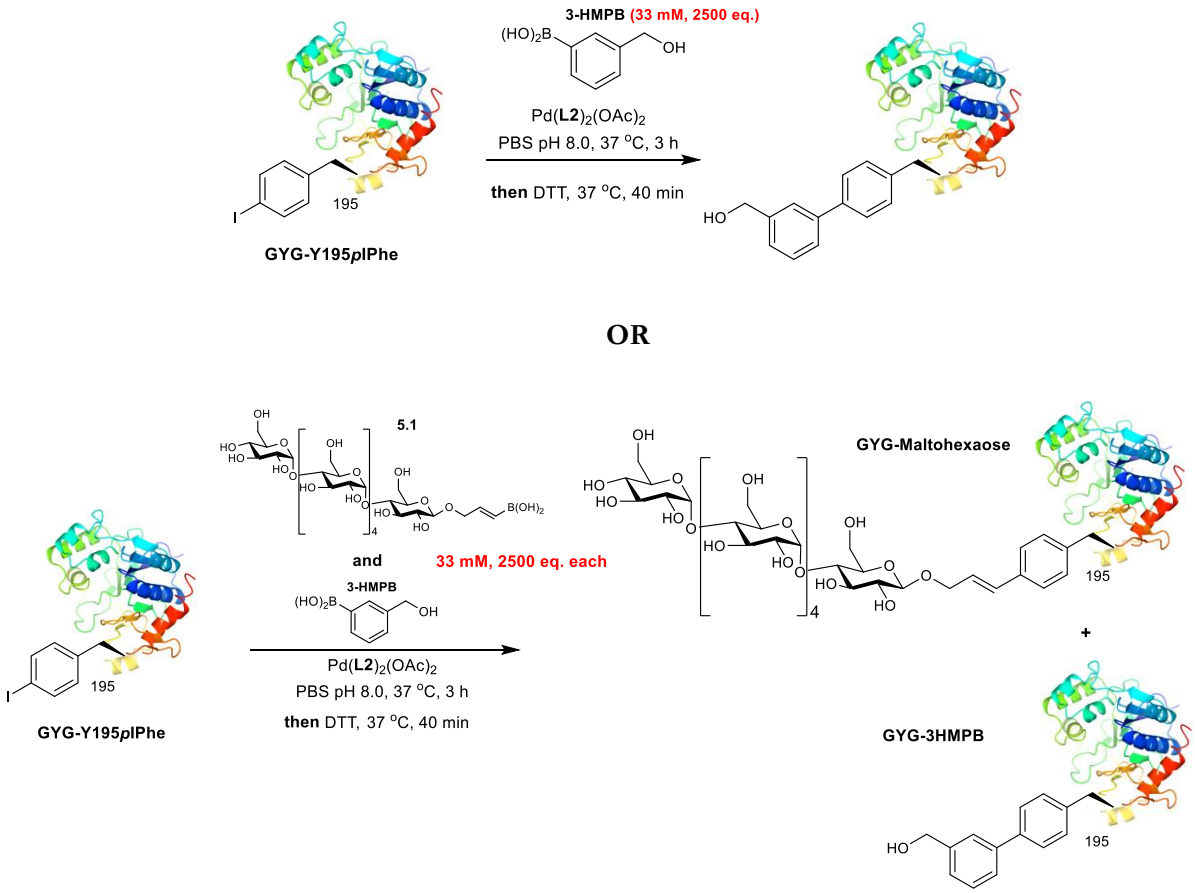
Reaction 1 adducts:



Reaction 2 aducts:

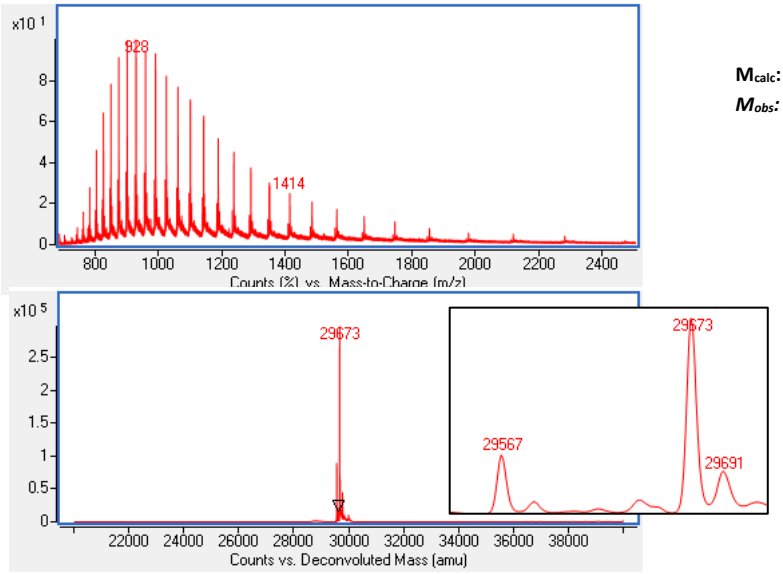


c) Competition Experiment with 3-HMPB



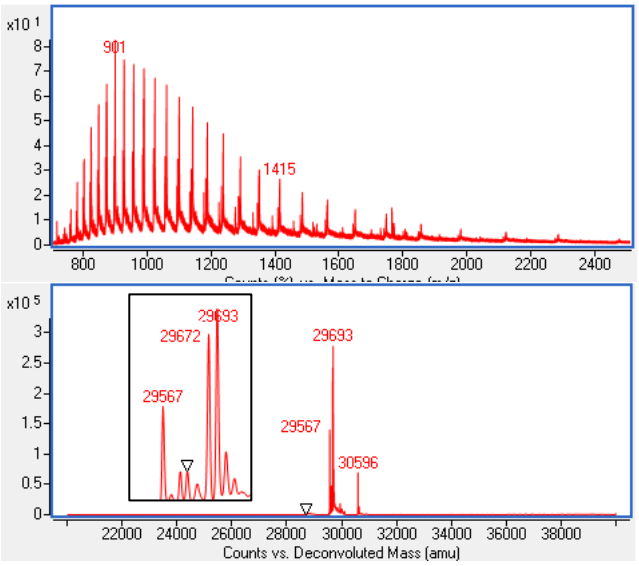
2 reactions were set up. A stock solution of maltohexaose boronic acid **5.1** was prepared by dissolving boronic acid (110 mg/ml) in H_2O, and a stock solution of **3-HMPB** was prepared by dissolving boronic acid (15.6 mg/ml) in aq. Na_2HPO_4 (110 mM). To non-his-tagged GYG-Y195*p*IPhe (6 µl of 1.42 mg/ml, 0.29 nmol) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was added 7 µl of stock solution of **3-HMPB** (720 nmol, 2500 eq. boronic acid). For one reaction, 7 µl of stock solution of **5.1** (720 nmol, 2500 eq. boronic acid) was then added; for the other reaction, this was not added and 7 µl of H_2O was added instead. $Pd(L2)_2(OAc)_2$ catalyst (2.0 µl of 10 mM solution, 20 nmol, 0.9 mM concentration) was added to both reactions, the reaction mixtures briefly vortexed, then incubated at 37 °C with shaking (550 rpm). After 3 hours, DTT was added to a final concentration of 12 mM (from a 115 mM solution in H_2O), the mixtures shaken for a further 40 minutes, then analysed by LC-MS.

3-HMPB only: conversion = 66 %



M_{calc}: 29672
M_{obs}: 29673 (cross-coupled, 66 %)
29691 (GYG-Y195*p*IPhe)
29567 (de-iodinated)

3-HMPB and 5.1: conversion (3-HMPB) = 33 %, conversion (5.1) = 10 %, overall coupling = 43 %.

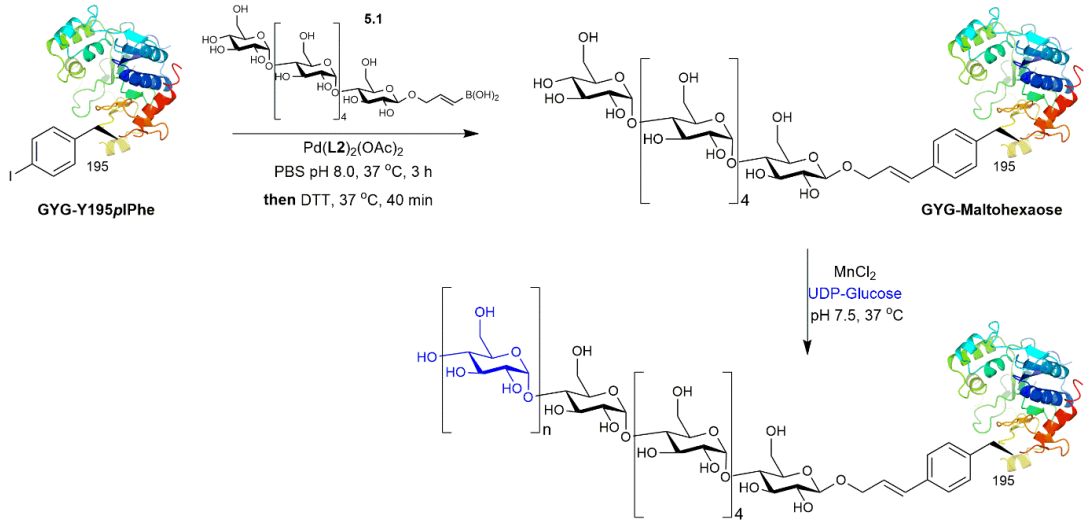


M_{calc}: 29672 (3-HMPB coupled), 30594 (5.1 coupled)
M_{obs}: 29672 (3-HMPB coupling, 33 %)
30596 (5.1 coupling, 10 %)
29693 (GYG-Y195*p*IPhe)
29567 (de-iodinated)

5.6.3 Enzymatic Activity Studies

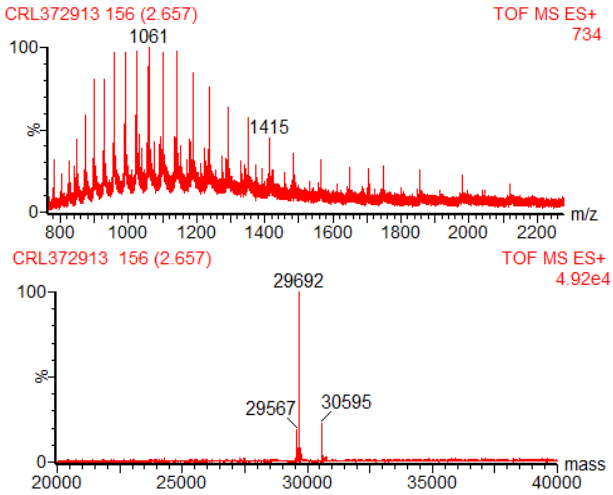
Analysis of LC-MS data generated through the below protocols was carried out as previously described in chapter 4, section 4.5.3. Please refer to this section for explanations of how graphs for overall glucosylation profiles and rates of initial glucosylation step (as measured by consumption of cross-coupled enzyme) are determined from such data. In brief, proportions of each glycoform – including cross-coupled enzyme – are determined from relative peak intensities, and these may be used to calculate average numbers of glucoses attached per enzyme).

5.6.3.1 Autoglucosylaton of GYG-Maltohexaose



A stock solution of maltohexaose boronic acid **5.1** was prepared by dissolving boronic acid (110 mg/ml) in H_2O. To non-his-tagged GYG-Y195*p*IPhe (10 µl of 0.85 mg/ml, 0.29 nmol) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was added 10 µl of this stock solution (1020 nmol, 3500 eq. boronic acid), followed by $Pd(L2)_2(OAc)_2$ catalyst (2.0 µl of 10 mM solution, 20 nmol, 0.9 mM concentration). The reaction mixture was briefly vortexed, then incubated at 37 °C with shaking (550 rpm). After 3 hours, DTT was added to a final concentration of 8 mM (from a 115 mM solution in H_2O), the mixture shaken for a further 40 minutes, then dialysed against pH 7.5 25 mM Na-HEPES buffer (10000 MWCO, 4 °C, 2 h then 2 h then overnight, 3 x 300 ml buffer), using a Slide-A-Lyzer™ MINI dialysis device.

LC-MS after cross-coupling (16 % conversion):

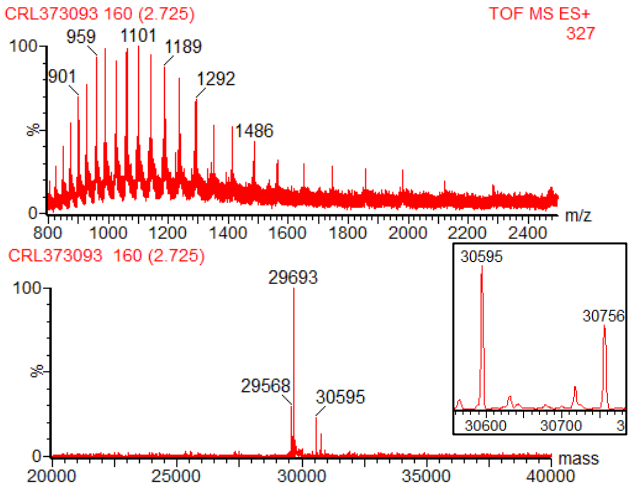


M_{calc}: 30594
M_{obs}: 30595 (16 %)
29567 (de-iodination)
29692 (GYG-Y195*p*IPhe)

Protein was adjusted to a concentration of 0.2 mg/ml (Vivaspin 500, 10000 MWCO (Sartorius Stedem) were used for concentration if required), and was incubated with $MnCl_2$ and UDP-Glucose disodium salt at final concentrations of 1 mM (added from freshly prepared 10 mM stock solutions in H_2O). The protein was incubated at 37 °C with shaking (500 rpm). After 40, 80, 120, 300 and 600 seconds, aliquots were taken and quenched through the addition of Na-EDTA to a final concentration of 20 mM, and placed on ice prior to LC-MS analysis, in order to determine average numbers of glucoses attached and the proportion of starting enzyme remaining at each time point. Please refer to chapter 4, section 4.5.3., for details regarding methods for analysis of LC-MS data; the method used is also described in detail in ref. 28.

Two independent replicates of the above protocol were carried out. Graphs are generated from LC-MS data from all time points across replicates according to chapter 4, section 4.5.3, and show means ±s.d. as stated in main text, where the same data is shown (as indicated by the figures referred to). Below is shown representative LC-MS data, followed by such graphical data.

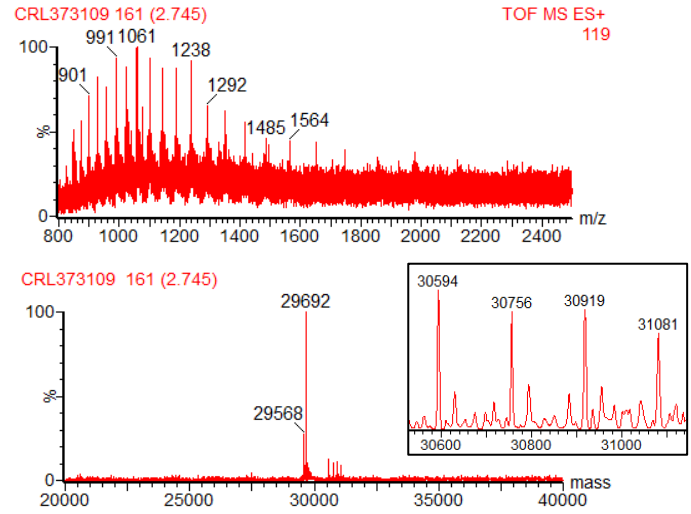
Representative LC-MS data from glucosylation assay: t = 80 s



M_{calc}: 30594 + 162n
M_{obs}: 30756 (n = 1), 30595 (n = 0)

+ non-glucosylated side-products from cross-coupling, see above.

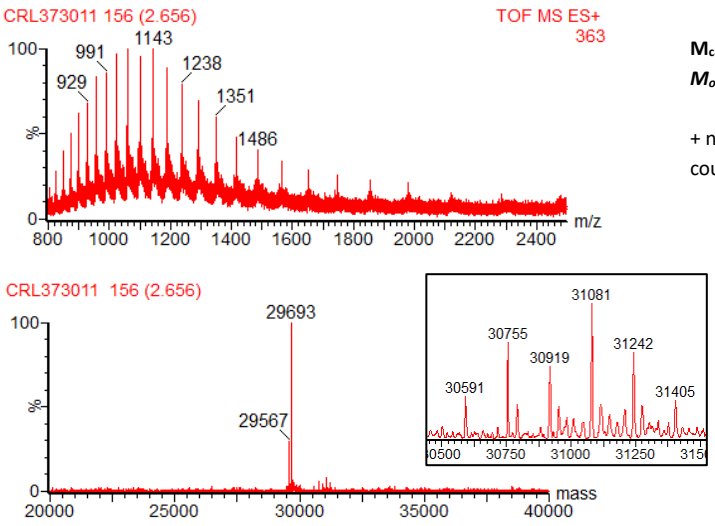
t = 5 min



M_{calc}: 30594 + 162n
M_{obs}: 30594 + 162n (n = 0-3)

+ non-glucosylated side-products from cross-coupling, see above.

t = 10 min



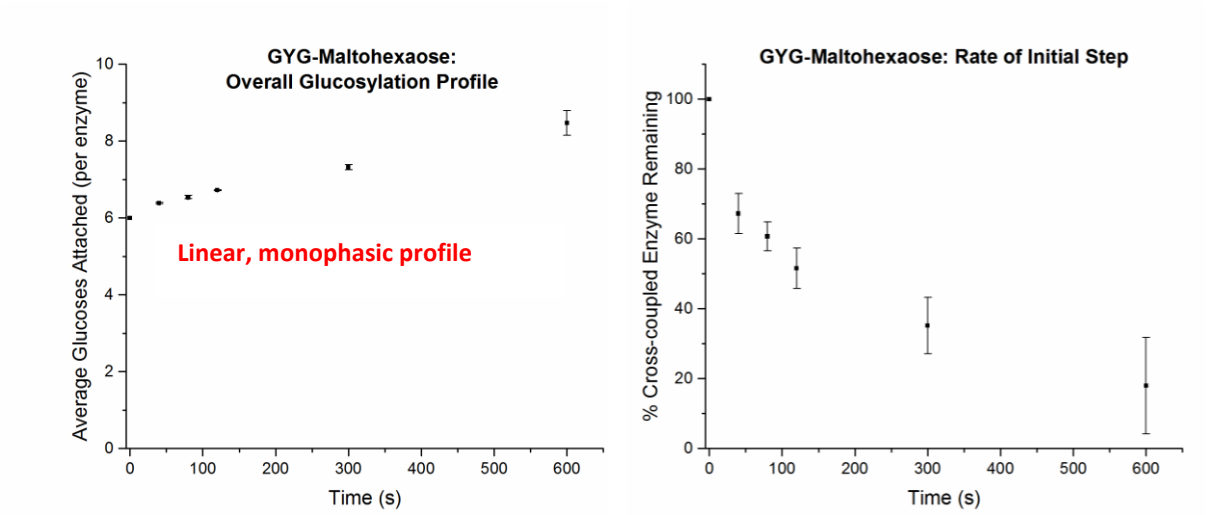
M_{calc}: 30594 + 162n

M_{obs}: 30591 + 162n (n = 0-5)

+ non-glucosylated side-products from cross-coupling, see above.

Graphical data from LC-MS from all time points and replicates (see also figure 5.12)

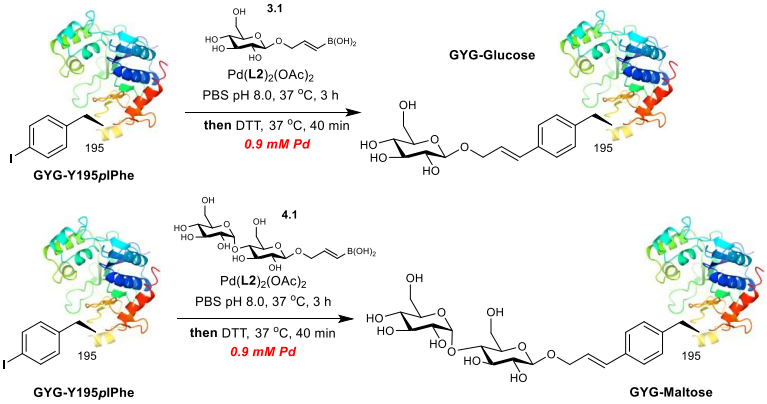
Methods for producing these types of graphs from LC-MS data such as that shown above were described in chapter 4, section 4.5.3. Please refer to this.



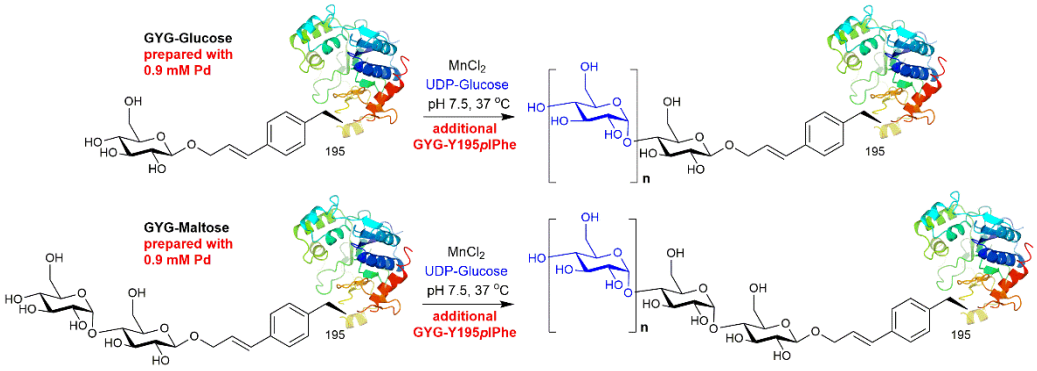
Note that the data shown in **figure 5.11** represents the t = 10 min timepoint from the protocol above.

5.6.3.2 Generation and autoglucosylation of GYG-Glucose / GYG-Maltose under equivalent conditions

1)



2)



4 reactions were set up by combining the following amounts of non-his-tagged GYG-Y195*p*IPhe in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl), boronic acid **3.1** or **4.1**, and $Pd(L2)_2(OAc)_2$ catalyst:

Reaction 1 (GYG-Glucose preparation):

- 16.6 µl of GYG-Y195*p*IPhe (0.85 mg/ml, 0.48 nmol)
- 16 µl of a 27 mg/ml stock solution of boronic acid **3.1** in H_2O (1640 nmol, 3400 eq.)
- 3.2 µl of 10 mM $Pd(L2)_2(OAc)_2$ catalyst solution (32 nmol, 0.9 mM final conc.)

Reaction 2 (GYG-Maltose preparation):

- 15 µl of GYG-Y195*p*IPhe (0.87 mg/ml, 0.44 nmol)
- 15 µl of a 45 mg/ml stock solution of boronic acid **4.1** in H_2O (1580 nmol, 3600 eq.)
- 2.95 µl of 10 mM $Pd(L2)_2(OAc)_2$ catalyst solution (29.5 nmol, 0.9 mM final conc.)

Reaction 3 (Pd-treatment of GYG-Y195*p*IPhe for use in GYG-Glucose autoglucosylation):

- 16.6 µl of GYG-Y195*p*IPhe (0.85 mg/ml, 0.48 nmol)
- 16 µl of H_2O only, **no boronic acid.**
- 3.2 µl of 10 mM $Pd(L2)_2(OAc)_2$ catalyst solution (32 nmol, 0.9 mM final conc.)

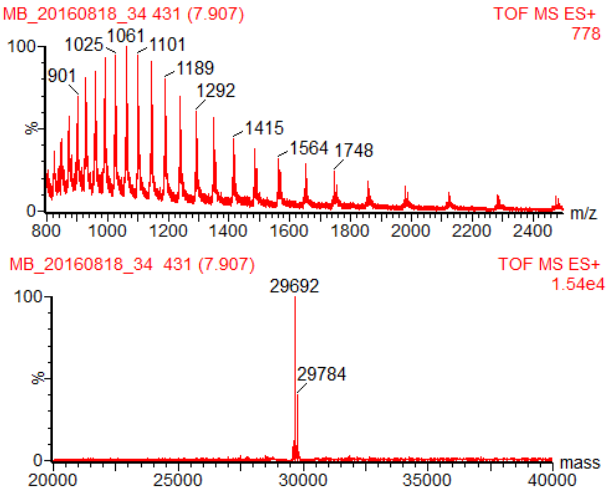
Reaction 4 (Pd-treatment of GYG-Y195*p*IPhe for use in GYG-Maltose autoglucosylation):

- 15 µl of GYG-Y195*p*IPhe (0.87 mg/ml, 0.44 nmol)
- 15 µl of H_2O only, **no boronic acid**.
- 2.95 µl of 10 mM $Pd(L2)_2(OAc)_2$ catalyst solution (29.5 nmol, 0.9 mM final conc.)

The reaction mixtures were briefly vortexed, then incubated at 37 °C with shaking (550 rpm). After 3 hours, DTT was added to a final concentration of 12 mM (from a 115 mM solution in H_2O), the mixtures shaken for a further 40 minutes, then dialysed against pH 7.5 25 mM Na-HEPES buffer (10000 MWCO, 4 °C, 2 h then 2 h then overnight, 3 x 300 ml buffer), using Slide-A-Lyzer™ MINI dialysis devices.

Cross-coupled enzyme from reaction 1 was combined with Pd-treated GYG-Y195*p*IPhe from reaction 3 – similarly, cross-coupled enzyme from reaction 2 was combined with GYG-Y195*p*IPhe from reaction 4 – such that the ratio of cross-coupled enzyme: GYG-Y195*p*IPhe was between 1:2.5 – 1:3, as judged by LC-MS intensities of their respective peaks (see additional note below). Lower proportions of cross-coupled enzyme led to insufficient LC-MS sensitivity.

Example LC-MS of GYG-Glucose with GYG-Y195*p*IPhe (reactions 1 + 3)

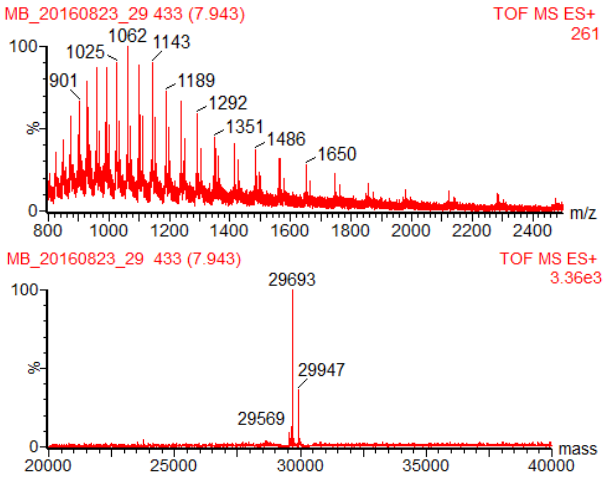


M_{calc}: 29784 (GYG-Glucose), 29692 (GYG-Y195*p*IPhe)
M_{obs}: 29784, 29692

Ratio GYG-Glucose: GYG-Y195pIPhe = 1:2.5

*Note: whilst e.g. de-iodination side-products may be present, they are too low in proportion after GYG-Y195*pIPhe *addition to be observable by LC-MS.*

Example LC-MS of GYG-Maltose with GYG-Y195*p*IPhe (reactions 2 + 4):



M_{calc}: 29946 (GYG-Maltose), 29692 (GYG-Y195*p*IPhe)
M_{obs}: 29947, 29693
29569 (de-iodinated)

Ratio GYG-Maltose: GYG-Y195pIPhe = 1:2.8

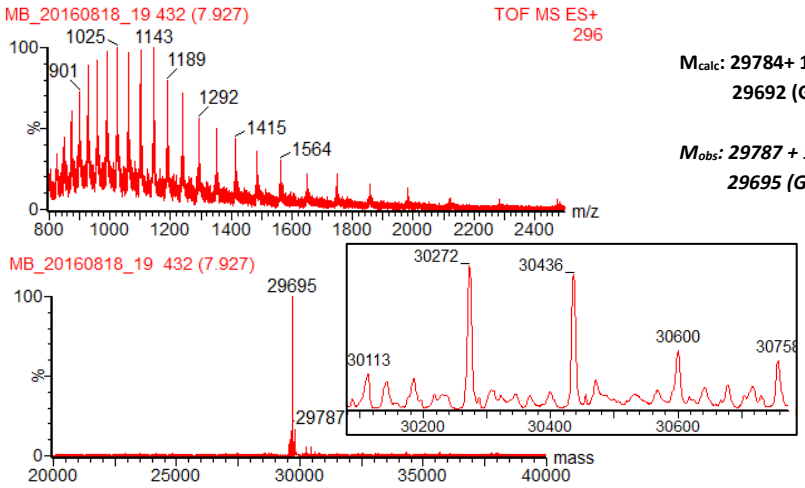
Combined Protein was adjusted to a concentration of 0.2 mg/ml (Vivaspin 500, 10000 MWCO (Sartorius Stedem) were used for concentration if required). Protein was incubated with $MnCl_2$ and UDP-Glucose disodium salt at final concentrations of 1 mM (added from freshly prepared 10 mM stock solutions in H_2O). The protein was incubated at 37 °C with shaking (500 rpm). After 20, 40, 80, 120, 300 and 600 seconds, aliquots were taken and quenched through the addition of Na-EDTA to a final concentration of 20 mM, and placed on ice prior to LC-MS analysis. Please refer to chapter 4, section 4.5.3., for details regarding methods for analysis of LC-MS data; the method used is also described in detail in ref. 28.

Two independent replicates of the entirety of the above protocol were carried out. Graphs are generated from LC-MS data from all time points across replicates according to methods in chapter 4, section 4.5.3, and show means ±s.d. as stated in main text, where the same data is also shown (as indicated by the figures referred to). Below is shown representative LC-MS data, followed by such graphical data.

*Additional Note: The ratios of GYG-Y195*pIPhe*: cross-coupled enzyme as judged by LC-MS were typically slightly higher (i.e. more GYG-Y195*pIPhe*) than when judged by Bradford assay; a molar ratio of ca. 1:1 cross-coupled enzyme: GYG-Y195*pIPhe, as judged by the latter was generally sufficient. As noted, we used MS proportions purely for consistency of analytical method with GYG-Maltohexaose samples. As alluded to in the text conditions (see p.319 and footnote iv), this difference may reflect the unavoidable fact that "Pd-treated" GYG-Y195*pIPhe spiked into GYG-Glucose and GYG-Maltose samples may perhaps ionise better than enzymes formed in genuine cross-coupling reactions, such as GYG-Y195*pIPhe remaining in GYG-Maltohexaose coupling, given slightly different reaction conditions. We note that it could in theory instead reflect intrinsic ionisation differences between GYG-Y195*pIPhe and cross-coupled enzymes, however this is unlikely given the known fact that the length of the glycan attached to GYG – including a length of zero, i.e. no glycan at all – has little impact on MS ionisation.*[31] *Given the single-atom difference between GYG-Y195*pIPhe and GYG with no glycan, and structural similarity of cross-coupled enzymes to GYG with longer glycans, it is expected that these should also ionise similarly.*

Representative LC-MS data from glucosylation assays:

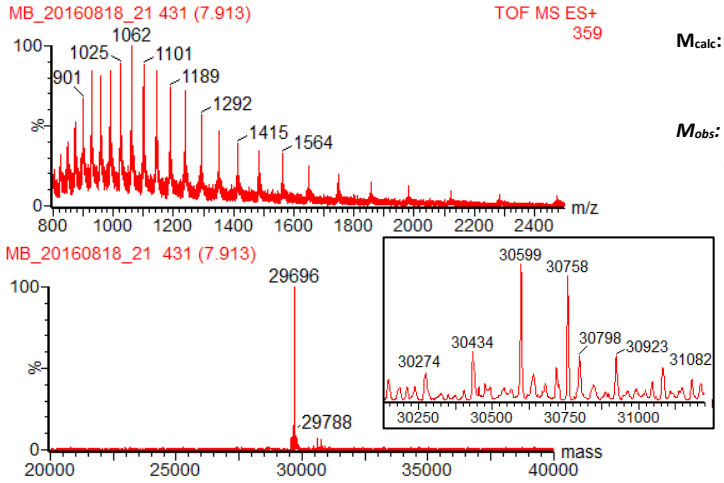
GYG-Glucose + GYG-Y195*p*IPhe, t = 80 s



M_{calc}: 29784+ 162n (glucosylated GYG-Glucose), 29692 (GYG-Y195*p*IPhe)

M_{obs}: 29787 + 162n (n = 0, 2-6)
29695 (GYG-Y195pIPhe, inactive)

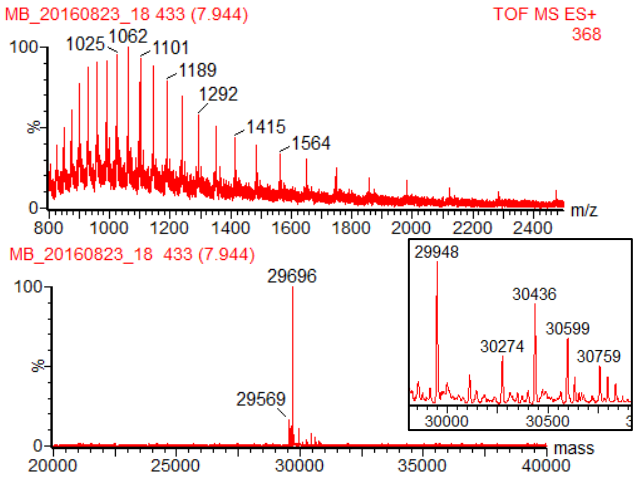
GYG-Glucose + GYG-Y195*p*IPhe, t = 300 s



M_{calc}: 29784+ 162n (glucosylated GYG-Glucose), 29692 (GYG-Y195*p*IPhe)

M_{obs}: 29788 + 162n (n = 0, 3-8)
29696 (GYG-Y195pIPhe, inactive)

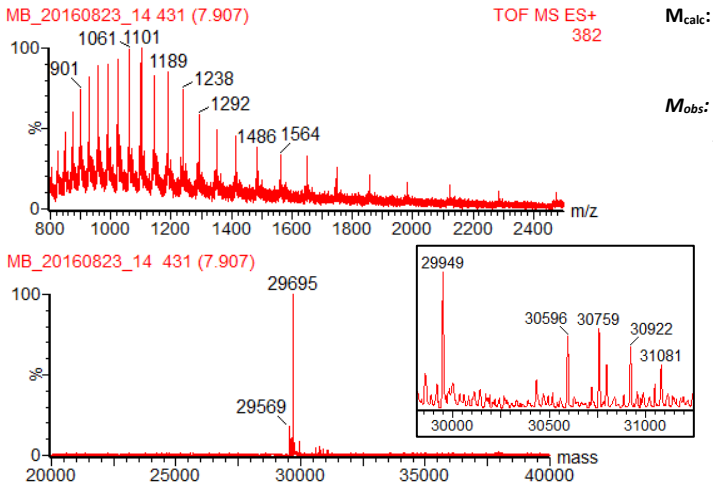
GYG-Maltose + GYG-Y195*p*IPhe, t = 80 s



M_{calc}: 29946+ 162n (glucosylated GYG-Glucose), 29692 (GYG-Y195*p*IPhe)

M_{obs}: 29948 + 162n (n = 0-5)
29696 (GYG-Y195pIPhe, inactive)
29569 (de-iodinated, inactive)

GYG-Maltose + GYG-Y195*p*IPhe, t = 300 s



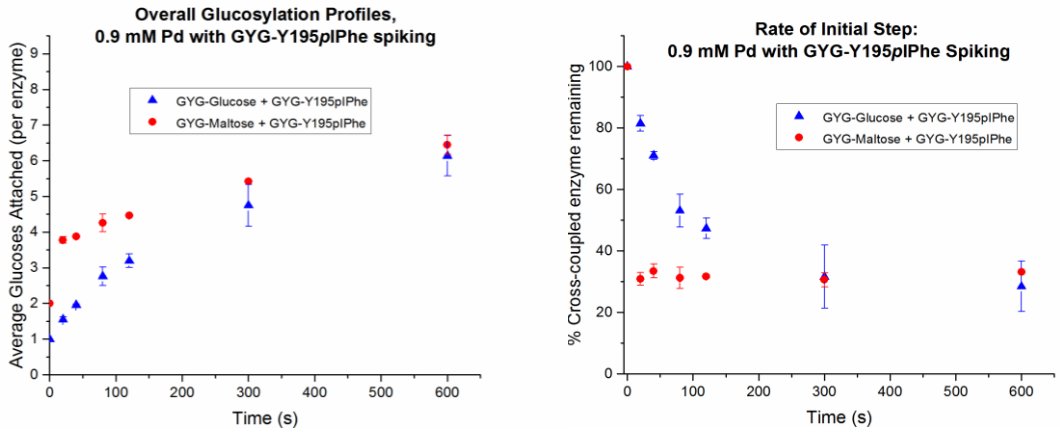
M_{calc}: 29946+ 162n (glucosylated GYG-Glucose), 29692 (GYG-Y195*p*IPhe)

M_{obs}: 29949 + 162n (n = 0, 3-7)
29696 (GYG-Y195pIPhe, inactive)
29569 (de-iodinated, inactive)

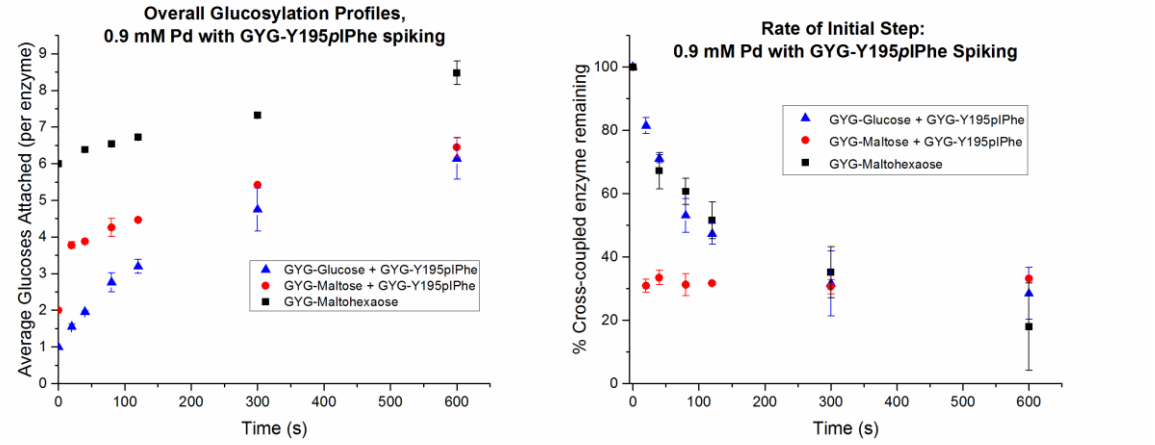
Graphical data from LC-MS from all time points and replicates:

Generated utilising the protocol previously described in chapter 4, section 4.5.3.

1) GYG-Glucose and GYG-Maltose in isolation (section 5.6.3.2 only)



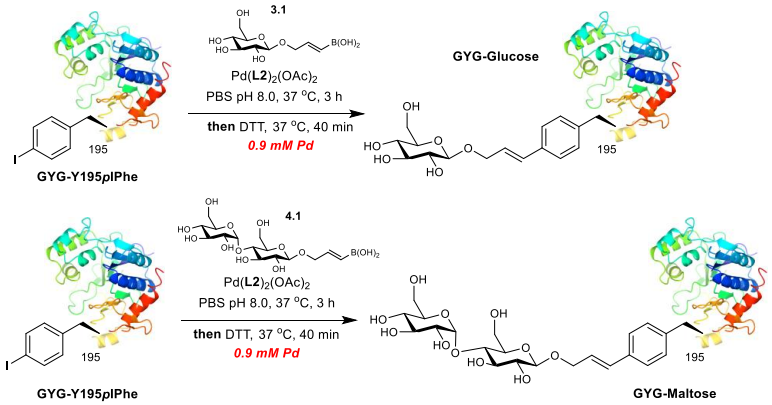
2) Comparison of above data to that of GYG-Maltohexaose from section 5.6.3.1 (figure 5.16)



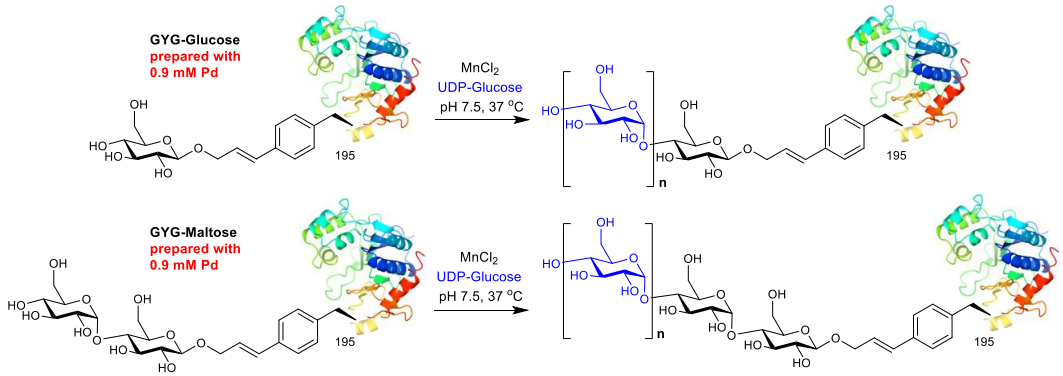
5.6.3.3 Effect of GYG-Y195*p*IPhe spiking on activity of GYG-Glucose and GYG-Maltose

In order to evaluate the effect of GYG-Y195pIPhe spiking on GYG-Glucose and GYG-Maltose autoglucosylation activity, these enzymes were prepared as described in section 5.6.3.2 but with no addition of GYG-Y195pIPhe, as detailed in full below. Data from their autoglucosylation was then compared to that generated in section 5.6.3.2 to assess the effect of the latter.

1)



2)



2 reactions were set up by combining the following amounts of non-his-tagged GYG-Y195*p*IPhe in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl), boronic acid **3.1** or **4.1**, and $Pd(L2)_2(OAc)_2$ catalyst:

Reaction 1 (GYG-Glucose preparation):

- 16.6 µl of GYG-Y195*p*IPhe (0.85 mg/ml, 0.48 nmol)
- 16 µl of a 27 mg/ml stock solution of boronic acid **3.1** in H_2O (1640 nmol, 3400 eq.)
- 3.2 µl of 10 mM $Pd(L2)_2(OAc)_2$ catalyst solution (32 nmol, 0.9 mM final conc.)

Reaction 2 (GYG-Maltose preparation):

- 15 µl of GYG-Y195*p*IPhe (0.87 mg/ml, 0.44 nmol)
- 15 µl of a 45 mg/ml stock solution of boronic acid **4.1** in H_2O (1580 nmol, 3600 eq.)
- 2.95 µl of 10 mM $Pd(L2)_2(OAc)_2$ catalyst solution (29.5 nmol, 0.9 mM final conc.)

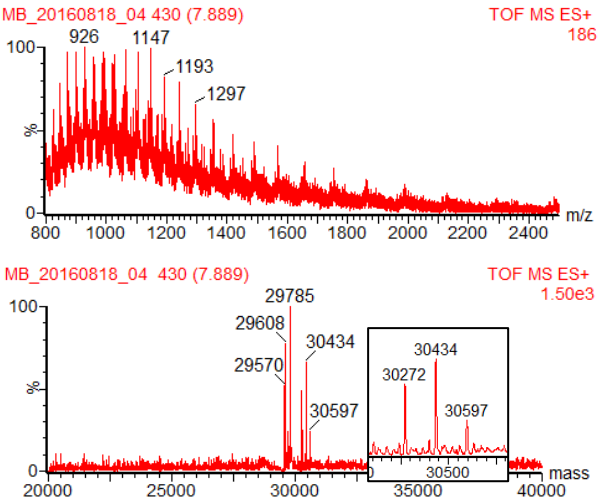
The reaction mixtures were briefly vortexed, then incubated at 37 °C with shaking (550 rpm). After 3 hours, DTT was added to a final concentration of 12 mM (from a 115 mM solution in H_2O), the mixtures shaken for a further 40 minutes, then dialysed against pH 7.5 25 mM Na-HEPES buffer (10000 MWCO, 4 °C, 2 h then 2 h then overnight, 3 x 300 ml buffer), using Slide-A-Lyzer™ MINI dialysis devices.

Protein was adjusted to a concentration of 0.2 mg/ml (Vivaspin 500, 10000 MWCO (Sartorius Stedem) were used for concentration if required), and was incubated with $MnCl_2$ and UDP-Glucose disodium salt at final concentrations of 1 mM (added from freshly prepared 10 mM stock solutions in H_2O). The protein was incubated at 37 °C with shaking (500 rpm). After 20, 40, 80, 120, 300 and 600 seconds, aliquots were taken and quenched through the addition of Na-EDTA to a final concentration of 20 mM, and placed on ice prior to LC-MS analysis. Please refer to chapter 4, section 4.5.3., for details regarding methods for analysis of LC-MS data; the method used is also described in detail in ref. 28.

Two independent replicates of the above protocol were carried out. Graphs are generated from LC-MS data from all time points across replicates according to chapter 4, section 4.5.3, and show means ±s.d. as stated in main text, where the same data is shown (as indicated by the figures referred to). Below is shown representative LC-MS data, followed by such graphical data.

Representative LC-MS data from glucosylation assays:

GYG-Glucose, t = 80 s

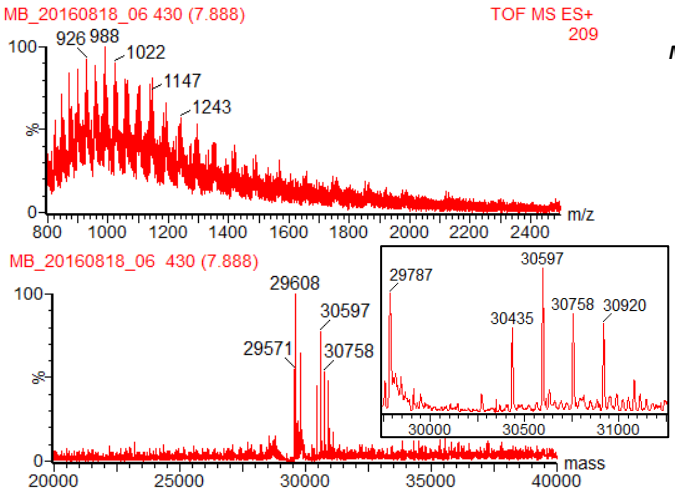


M_{calc}: 29784+ 162n (glucosylated GYG-Glucose),

M_{obs}: 29785 + 162n (n = 0, 3-5)

+ non-glucosylated side-products from coupling: 29570 (de-iodination), 29608 (**3.11,** see chapter 3 pp. 172-3)

GYG-Glucose, t = 300 s

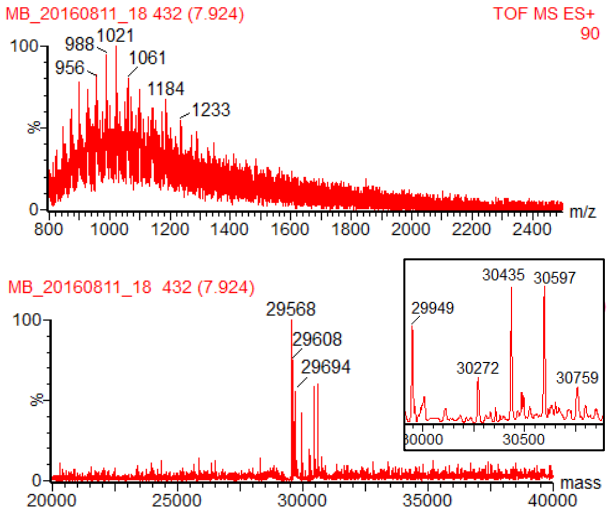


M_{calc}: 29784+ 162n (glucosylated GYG-Glucose),

M_{obs}: 29787 + 162n (n = 0, 3-7)

+ non-glucosylated side-products from coupling: 29571 (de-iodination), 29608 (**3.11,** see chapter 3 pp. 172-3, **figure 3.8**)

GYG-Maltose, t = 80 s

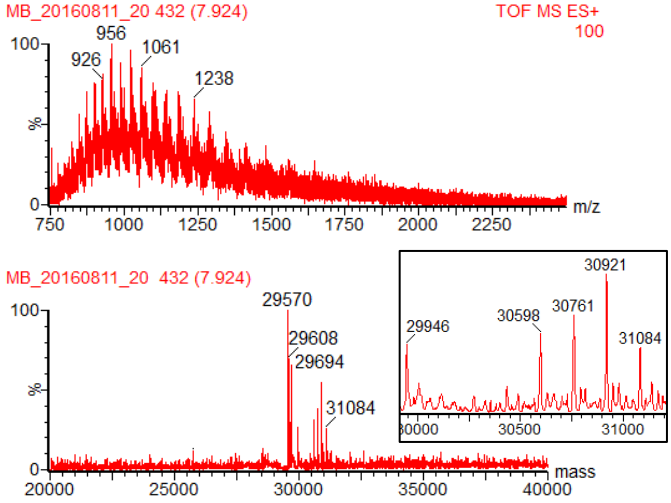


M_{calc}: 29946+ 162n (glucosylated GYG-Maltose),

M_{obs}: 29949 + 162n (n = 0, 2-5)

+ non-glucosylated side-products from coupling: 29568 (de-iodination), 29694 (GYG-Y195*p*IPhe), 29608 (**3.11,** see chapter 3 pp. 172-3, **figure 3.8**)

GYG-Maltose, t = 300 s



M_{calc}: 29946+ 162n (glucosylated GYG-Maltose),

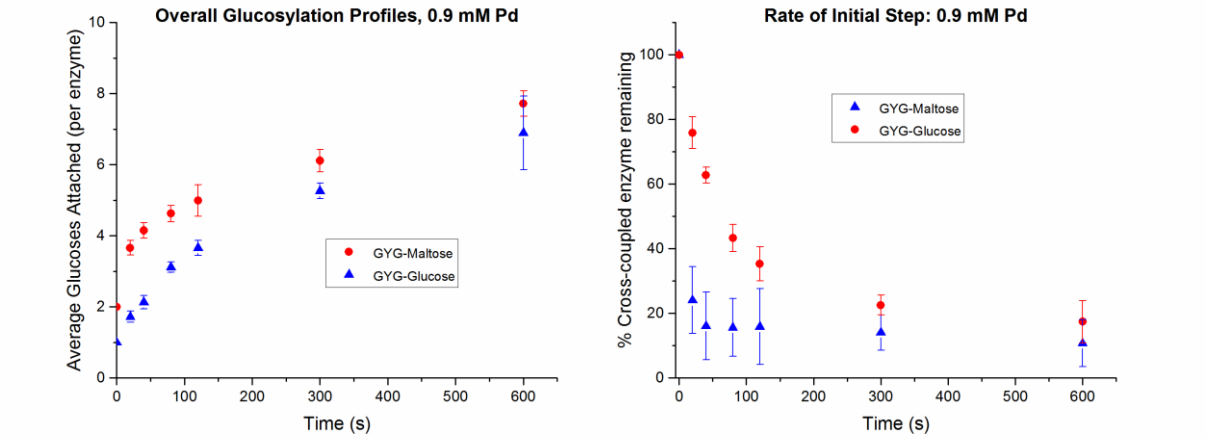
M_{obs}: 29946 + 162n (n = 0, 4-7)

+ non-glucosylated side-products from coupling: 29568 (de-iodination), 29694 (GYG-Y195*p*IPhe), 29608 (**3.11,** see chapter 3 pp. 172-3, **figure 3.8**)

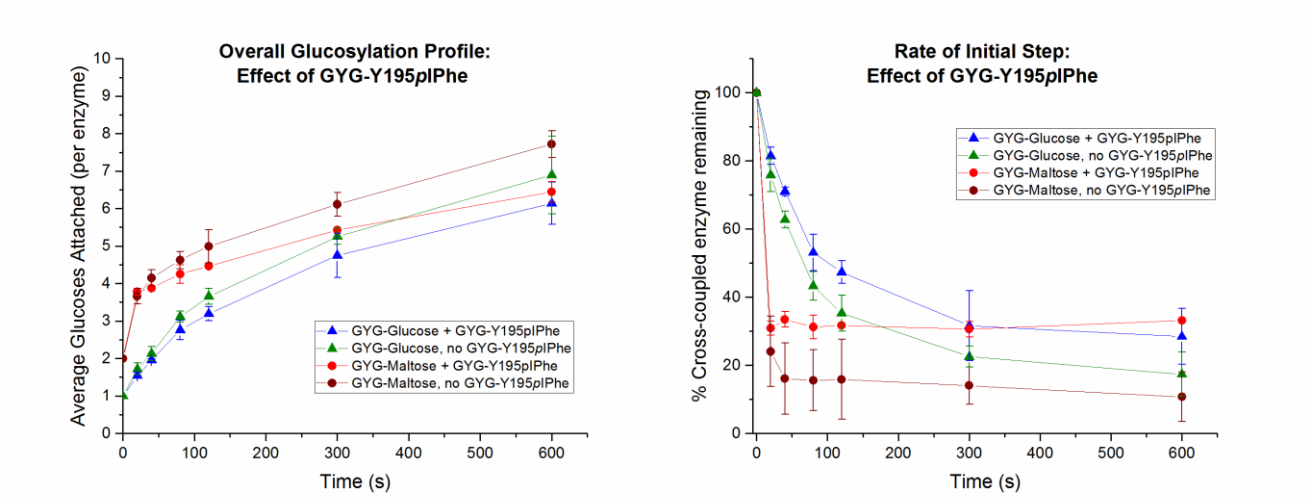
Graphical data from LC-MS from all time points and replicates:

Generated utilising the protocol previously described in chapter 4, section 4.5.3.

1) **Data from section 5.6.3.3 only**



2) **Data from sections 5.6.3.2 and 5.6.3.3 compared (figure 5.17):**



(lines are added purely to guide the eye).

5.7 References

1. Avery, O. T. & Goebel, W. F. Chemo-Immunological Studies on Conjugated Carbohydrate-Proteins : II. Immunological Specificity of Synthetic Sugar-Protein Antigens. *J. Exp. Med.* **50,** 533–550 (1929).

2. Pichichero, M. E. Protein carriers of conjugate vaccines: Characteristics, development and clinical trials. *Hum. Vaccin. Immunother.* **9,** 2505–2523 (2013).

3. Berti, F. & Adamo, R. Recent mechanistic insights on glycoconjugate vaccines and future perspectives. *ACS Chem. Biol.* **8,** 1653–1663 (2013).

4. Adamo, R. *et al.* Synthetically defined glycoprotein vaccines: current status and future directions. *Chem. Sci.* **4,** 2995 (2013).

5. Fernandez-Santana, V. *et al.* Antigenicity and Immunogenicity of a Synthetic Oligosaccharide-Protein Conjugate Vaccine against Haemophilus influenzae Type b Antigenicity and Immunogenicity of a Synthetic Oligosaccharide-Protein Conjugate Vaccine against Haemophilus influenzae Type b. *Infect. Immun.* **72,** 7115–7123 (2004).

6. Kong, L. *et al.* An antibacterial vaccination strategy based on a glycoconjugate containing the core lipopolysaccharide tetrasaccharide Hep2Kdo2. *Nat. Chem.* **8,** 242–249 (2016).

7. Schofield, L., Hewitt, M. C., Evans, K., Siomos, A. & Seeberger, P. H. Synthetic GPI as a candidate anti- toxic vaccine in a model of malaria. *Nature* **418,** 785–789 (2002).

8. Ferna, V. *et al.* Polysaccharide Vaccine Against Haemophilus influenzae Type b. *Science.* **305,** 522–525 (2004).

9. Grayson, E. J. *et al.* A coordinated synthesis and conjugation strategy for the preparation of homogeneous glycoconjugate vaccine candidates. *Angew. Chemie. Int. Ed.* **50,** 4127–4132 (2011).

10. Oberli, M. A. *et al.* A possible oligosaccharide-conjugate vaccine candidate for clostridium difficile is antigenic and immunogenic. *Chem. Biol.* **18,** 580–588 (2011).

11. Bundle, D. R. *et al.* Thiooligosaccharide conjugate vaccines evoke antibodies specific for native antigens. *Angew. Chemie. Int. Ed.* **44,** 7725–7729 (2005).

12. Ni, J., Singh, S. & Wang, L. X. Synthesis of maleimide-activated carbohydrates as chemoselective tags for site-specific glycosylation of peptides and proteins. *Bioconjug. Chem.* **14,** 232–238 (2003).

13. Wong, S. Y., Guile, G. R., Dwek, R. a & Arsequell, G. Synthetic glycosylation of proteins using N-(beta-saccharide) iodoacetamides: applications in site-specific glycosylation and solid-phase enzymic oligosaccharide synthesis. *Biochem. J.* **300 Pt 3,** 843–850 (1994).

14. van Kasteren, S. I. *et al.* Expanding the diversity of chemical protein modification allows post-translational mimicry. *Nature* **446,** 1105–1109 (2007).

15. Chalker, J. M. *et al.* Methods for converting cysteine to dehydroalanine on peptides and proteins. *Chem. Sci.* **2,** 1666 (2011).

16. Minamikawa, H., Murakami, T. & Hato, M. Synthesis of 1,3-di-O-alkyl-2-O-(β-glycosyl)glycerols bearing oligosaccharides as hydrophilic groups. *Chem. Phys. Lipids* **72,** 111–118 (1994).

17. Narumi, A., Satoh, T., Kaga, H. & Kakuchi, T. Glycoconjugated polymer. 3. Synthesis and amphiphilic property of core-glycoconjugated star-shaped polystyrene. *Macromolecules* **35,** 699–705 (2002).

18. Bérces, A. *et al.* Is acyl migration to the aglycon avoidable in 2- acyl assisted glycosylation reactions ? 1. *Can. J. Chem.* **82,** 1157–1171 (2004).

19. Dumas, A. *et al.* Self-liganded Suzuki-Miyaura coupling for site-selective protein PEGylation. *Angew. Chem. Int. Ed. Engl.* **52,** 3916–3921 (2013).

20. Gao, Z., Gouverneur, V. & Davis, B. G. Enhanced aqueous Suzuki-Miyaura coupling allows site-specific polypeptide 18F-labeling. *J. Am. Chem. Soc.* **135,** 13612–13615 (2013).

21. Miller, W. D., Fray, A. H., Quatroche, J. T. & Sturgill, C. D. Suppression of a palladium-mediated homocoupling in a Suzuki cross-coupling reaction. Development of an impurity control strategy supporting synthesis of LY451395. *Org. Process Res. Dev.* **11,** 359–364 (2007).

22. Chalker, J. M., Wood, C. S. C. & Davis, B. G. A convenient catalyst for aqueous and protein Suzuki-Miyaura cross-coupling. *J. Am. Chem. Soc.* **131,** 16346–16347 (2009).

23. Klüfers, P. & Kunte, T. A transition metal complex of D-glucose. *Angew. Chemie. Int. Ed.* **40,** 4210–4212 (2001).

24. Hall, D. G. in *Boronic Acids: Preparation and Applications in Organic Synthesis and Medicine* 1–134 (2005).

25. Sun, X., Zhai, W., Fossey, J. S. & James, T. D. Boronic acids for fluorescence imaging of carbohydrates. *Chem. Commun.* **52,** 3456–3469 (2016).

26. Lennox, A. J. J. & Lloyd-Jones, G. C. Selection of boron reagents for Suzuki-Miyaura coupling. *Chem. Soc. Rev.* **43,** 412–443 (2014).

27. Spicer, C. D. Suzuki Biology: Palladium Mediated Reactions under Ambient Conditions. (University of Oxford, DPhil Thesis, 2014).

28. Hurley, T. D., Walls, C., Bennett, J. R., Roach, P. J. & Wang, M. Direct detection of glycogenin reaction products during glycogen initiation. *Biochem Biophys Res Commun.* **348,** 374–378 (2006).

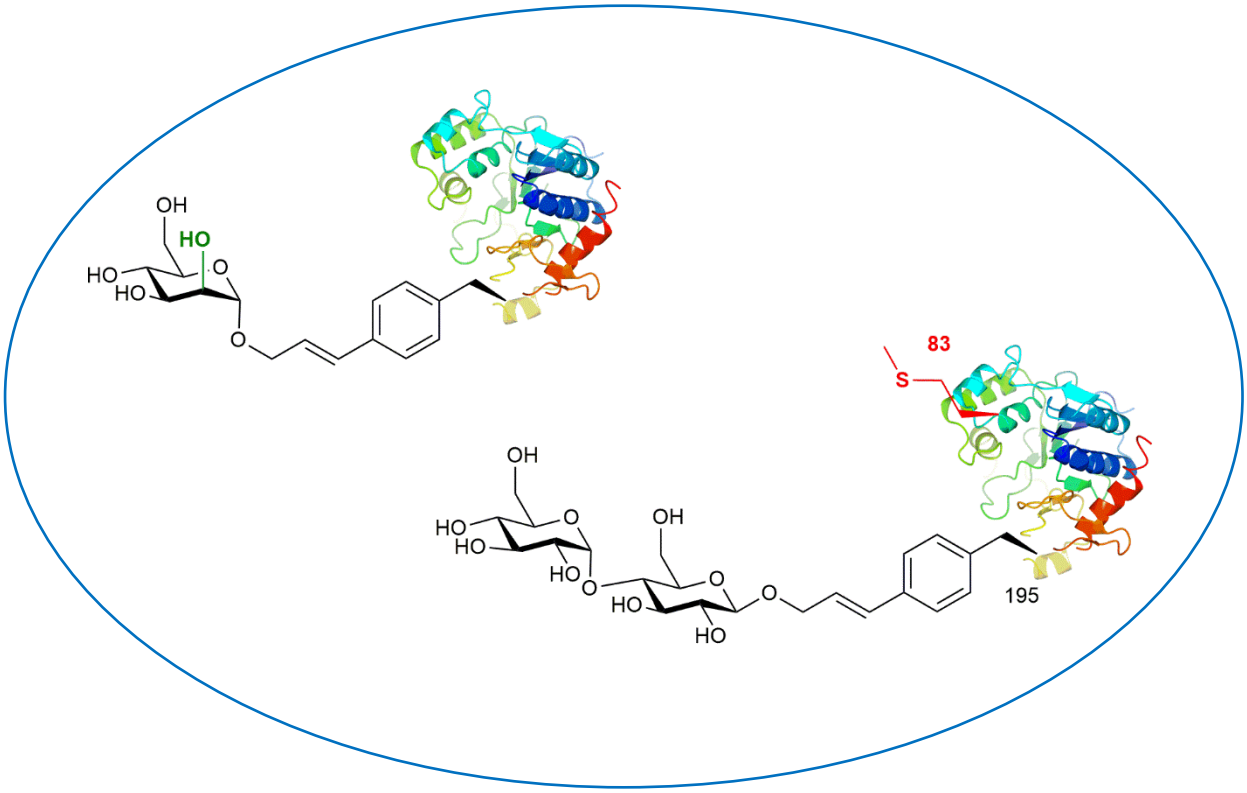
29. Spicer, C. D. & Davis, B. G. Palladium-mediated site-selective Suzuki-Miyaura protein modification at genetically encoded aryl halides. *Chem. Commun. (Camb).* **47,** 1698–1700 (2011).

30. Clarke, P. A., Rolla, G. A., Cridland, A. P. & Gill, A. A. An improved synthesis of (2E,4Z)-6-(benzyloxy)-4-bromohexa-2,4-dien-1-ol. *Tetrahedron* **63,** 9124–9128 (2007).

31. Nilsson, J. *et al.* LC-MS/MS characterization of combined glycogenin-1 and glycogenin-2 enzymatic activities reveals their self-glucosylation preferences. *Biochim. Biophys. Acta* **1844,** 398–405 (2014).

32. Pagliaro, M. *Silica-Based Materials for Advanced Chemical Applications*. (Royal Society of Chemistry, 2009).

Chapter 6



Exploring Unnatural GYG Glycoforms

Contents

6.1 Introduction 363
6.1.1 Tag and Modify: a Means for Rapid Protein Diversification 363
6.1.2 Variation of Acceptor Stereochemistry 366
6.1.3 Modifying the Enzyme Scaffold: a Catalytically Impaired GYG Mutant 367

6.2 Investigating Unnatural GYG-Carbohydrate Acceptors 370

6.3 Investigation of Catalytically Impaired GYG-T83M Glycoforms 380
6.3.1 Generation of GYG-T83M-Y195*p*IPhe 380
6.3.2 Catalytic Activity of GYG-T83M Mutant Glycoforms 382

6.4 Conclusion and Outlook 387

6.5 Experimental Section 388
6.5.1 Investigating Unnatural GYG-Carbohydrate Acceptors 388
6.5.1.1 Chemical Synthesis 388
6.5.1.2 Catalytic Activity of GYG-Galactose and GYG-Mannose 394
6.5.1.3 Other Studies on Cross-Couplings to GYG-Galactose and GYG-Mannose 407
6.5.2 Investigation of Catalytically Impaired GYG-T83M Glycoforms 416
6.5.2.1 Site-directed Mutagenesis to Generate GYG-T83M-Y195X (X = TAG) Plasmid 416
6.5.2.2 Expression of GYG-T83M-Y195*p*IPhe Protein (with his-tag) 418
6.5.2.3 Generation and Autoglucosylation Assay of GYG-T83M-Glucose 420
6.5.2.4 Inter-enzyme glucosylation of GYG-T83M-Glucose using GYG-Y195F 422
6.5.2.5 Generation and Autoglucosylation Assay of GYG-T83M-Maltose 424

6.6 References 426

6.1 Introduction

6.1.1 Tag and Modify: a Means for Rapid Protein Diversification

The "tag-and-modify" concept for the chemical modification of proteins, as discussed in chapter 1, involves the generation of a protein containing a uniquely reactive chemical "tag", and its subsequent specific (or, at least, selective) reaction with a particular class of modifying reagent.[1] A key advantage of such an approach is its ability to rapidly access a diverse library of protein constructs from a single precursor, through simple alteration of the modifying reagent used with the same tag. For example, a given protein bearing a dehydroalanine tag may be rapidly converted to many different variants thereof through reaction with thiol-containing compounds or (radicals derived from) alkyl iodides;[2,3] in a similar vein, CuAAC may enable access to multiple modified species from a common tagged precursor,[4] whilst Suzuki-Miyaura cross-coupling (SMC) to a given aryl halide-containing protein can generate diverse constructs through simple variation of the (small molecule) boronic acid partner.[5] This concept has underpinned the work in this thesis; Suzuki-Miyaura cross-coupling to a common tagged precursor (GYG-Y195*p*IPhe) has enabled access to a total of 8 distinct enzyme constructs (**figure 6.1**). Only the three constructs bearing a carbohydrate moiety represented functional enzyme species (chapters 3-5); the other five did not exhibit autoglucosylation activity (chapter 2) and were therefore of limited interest.

Two of the catalytically active species accessed thus far – GYG-Maltose and GYG-Maltohexaose – represent on-pathway intermediates in the glucosylation of a third (GYG-Glucose). These three species, all generated from GYG-Y195*p*IPhe, are thus inherently linked and form a class of GYG constructs differing from each other specifically with regards to the length of the chemically attached maltosaccharide (figure 6.2 a). Furthermore, they are all mimetics of natural glycoforms of the enzyme. In this chapter, we turn our attention instead to the deliberate generation and study of two different classes of GYG glycoforms that represent non-natural variants (**figure 6.2 b**).

Typically, accessing new protein variants via tag-and-modify involves variation of the "modifying" reagent, and indeed further alteration of the boronic acid coupling partners used could afford a class of GYG constructs differing not with regards to maltosaccharide chain length, but instead with regards to the stereochemistry of the initial carbohydrate glycosyl acceptor (e.g. GYG-Galactose, GYG-Mannose, **figure 6.2 b**). The investigation of the activity or otherwise of such constructs may yield further insight into the enzymatic mechanism, in a similar vein to that previously afforded via probing an unnatural autogalactosylation pathway (see chapter 4).

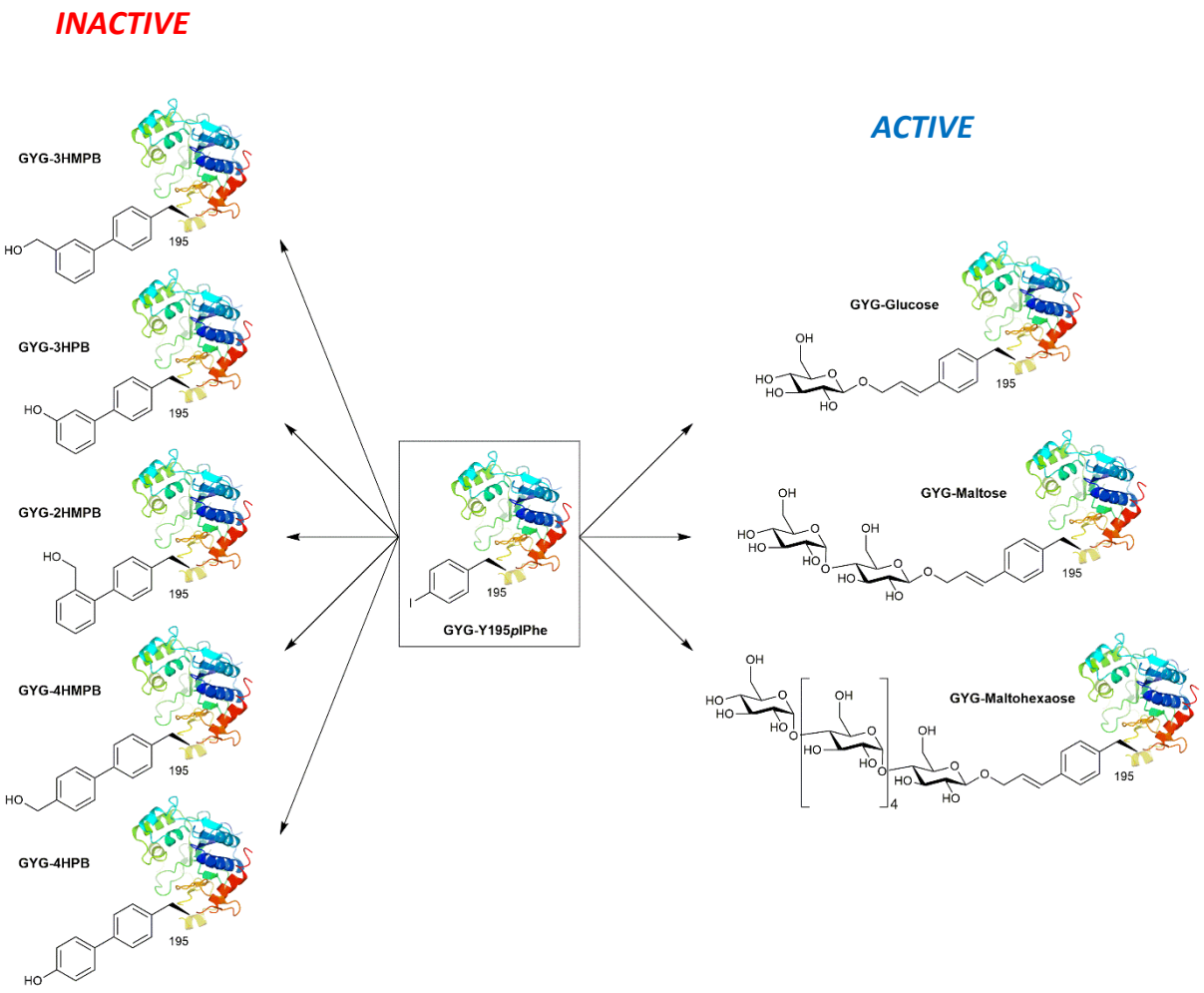
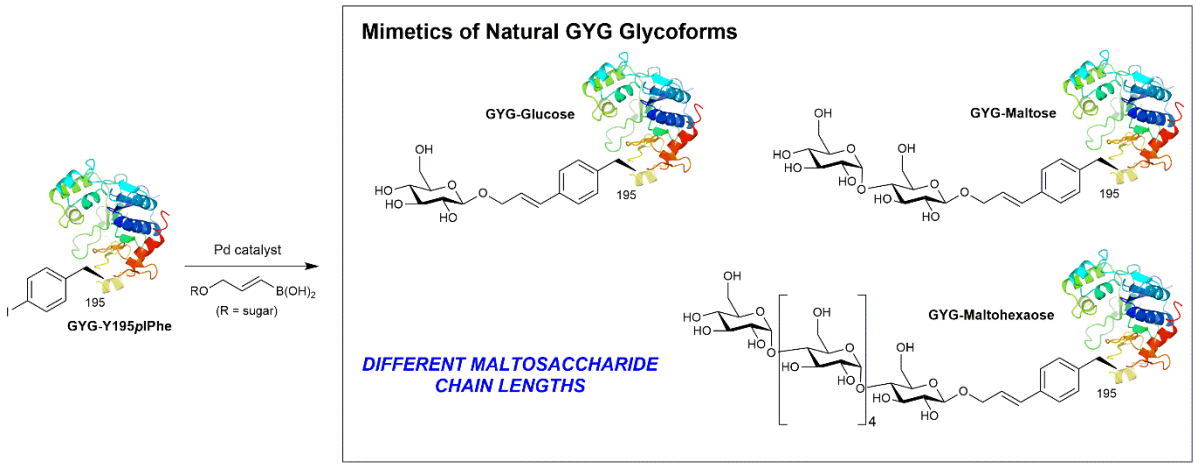


Figure 6.1 Summary of the GYG constructs accessed thus far from the common, tagged precursor GYG-Y195*p*IPhe.

However, an alternative approach would be to instead alter the tag itself (i.e. GYG-Y195*p*IPhe). For example, modification at the DNA level – mutagenesis – could yield a third class of GYG constructs that also differ with regards to maltosaccharide chain length, but are based on an altered enzyme scaffold (**figure 6.2 b**). Studies on catalytically impaired GYG mutants have previously lent significant insight into the catalytic mechanism, particularly with regards to existence of intra- and inter-enzyme glucosylation modes;[6,7] analogous studies on defined glycoform variants thereof accessible through our methodology could prove still more powerful in this sense.

In the following sections, the rationales behind such these two approaches are discussed in more detail (sections 6.1.2, 6.1.3). Subsequently, the results obtained and conclusions drawn from each approach – variation of glycosyl acceptor stereochemistry (section 6.2) and investigation of mutant GYG variants (section 6.3) – are outlined.

a Prior work: generation of natural GYG glycoform mimetics



b Present work: new classes of GYG glycoconjugates

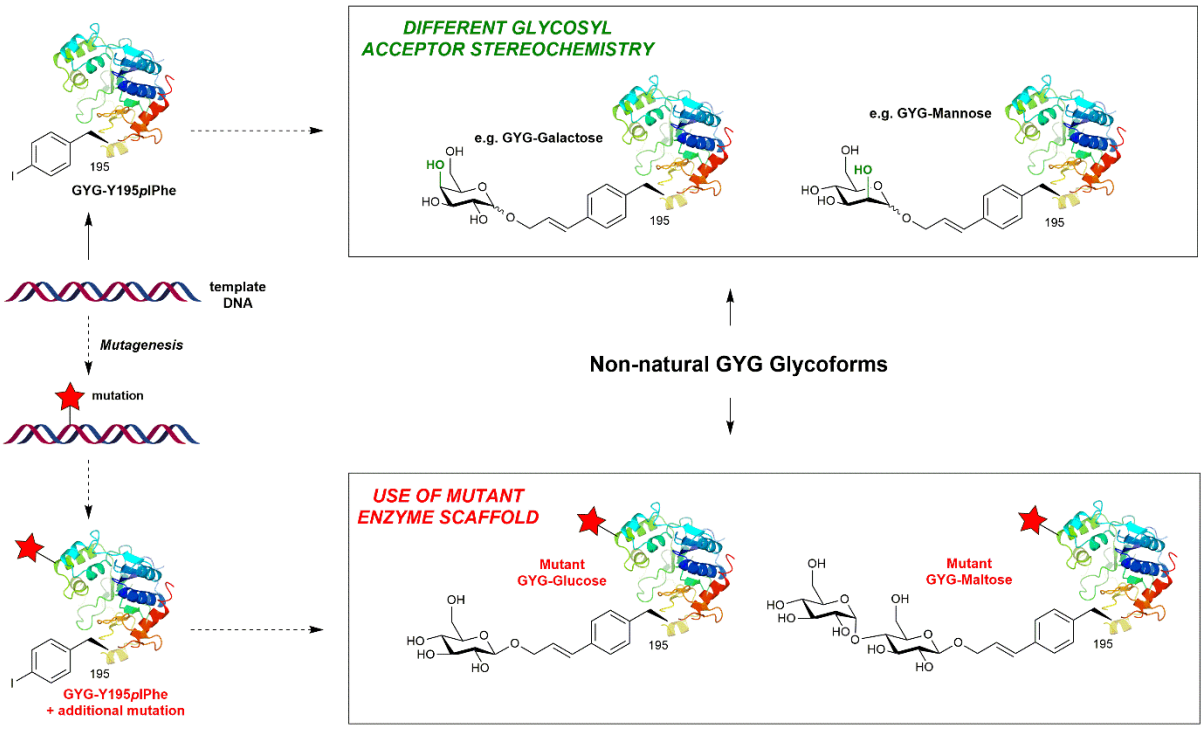


Figure 6.2 (a) Prior work has involved the generation of mimetics of natural GYG glycoforms, differing from each other solely in maltosaccharide chain length. (b) Work in this chapter focusses on the generation of two new classes of non-natural GYG glycoforms: those with altered glycosyl acceptor stereochemistry and those based on a mutant enzyme scaffold.

6.1.2 Variation of GYG Glycosyl Acceptor Stereochemistry

Glycosyltransferases were once considered to be entirely specific for a particular glycosyl donor and glycosyl acceptor pair, a concept that became known as the "one enzyme, one linkage" theory.[8] However, numerous examples have since demonstrated that a number of naturally occurring glycosyltransferases exhibit a certain degree of promiscuity with regards to both partners.[9–13] Equally, directed evolution approaches have been applied to generate glycosyltransferases displaying a broadened substrate tolerance.[14–16]

Such substrate plasticity is well-documented for GYG with respect to glycosyl donor; indeed, our own work on autogalactosylation (chapter 4), as well as prior work in this area,[17] highlighted this very fact. It is additionally known that the enzyme can tolerate other pyrimidine nucleotide donors (TDP-glucose and CDP-glucose),[18] and may catalyse autoxylosylation from UDP-xylose.[19,20] The ability of GYG to catalyse the glucosylation of multiple exogeneous, non-enzyme bound acceptors – particularly maltose and derivatives – is well documented.[17,21–23] However, little is currently known regarding the enzyme's ability or otherwise to *initiate* catalysis from non-natural, enzyme-linked acceptors, reflecting the lack of a method for their generation and isolation. Whilst it may be inferred from the ability of GYG to add multiple successive galactose residues to itself that galactose must be a suitable glycosyl acceptor, it is not known whether the enzyme would be capable of catalysing such galactosylation if the first carbohydrate moiety was also galactose; the limited prior work in this area – including our own – has investigated autogalactosylation of already partially glucosylated enzyme.[17]

In chapter 2, we utilised our methodology to probe non-carbohydrate-based glycosyl acceptors; these were insufficient for activity, implicating a need for at least a carbohydrate moiety or Tyr195 itself. However, we were interested to investigate whether the enzyme might be capable of initiating catalysis from carbohydrate structures exhibiting altered stereochemistry with respect to the natural glucosyl variant, and considered that our methodology might enable unique access to such GYG constructs. In the same way that cross-coupling of a glucose-derived boronic acid **3.1** affords a mimetic of a natural GYG glycoform (**figure 6.3 a**), we envisaged that analogous cross-coupling of non-glucose monosaccharide derivatives (e.g. galactose-derived, **6.1**, or mannose-derived, **6.2**) to GYG-Y195*p*IPhe could enable direct study instead of non-natural glycosyl acceptor species (**figure 6.3 b**).

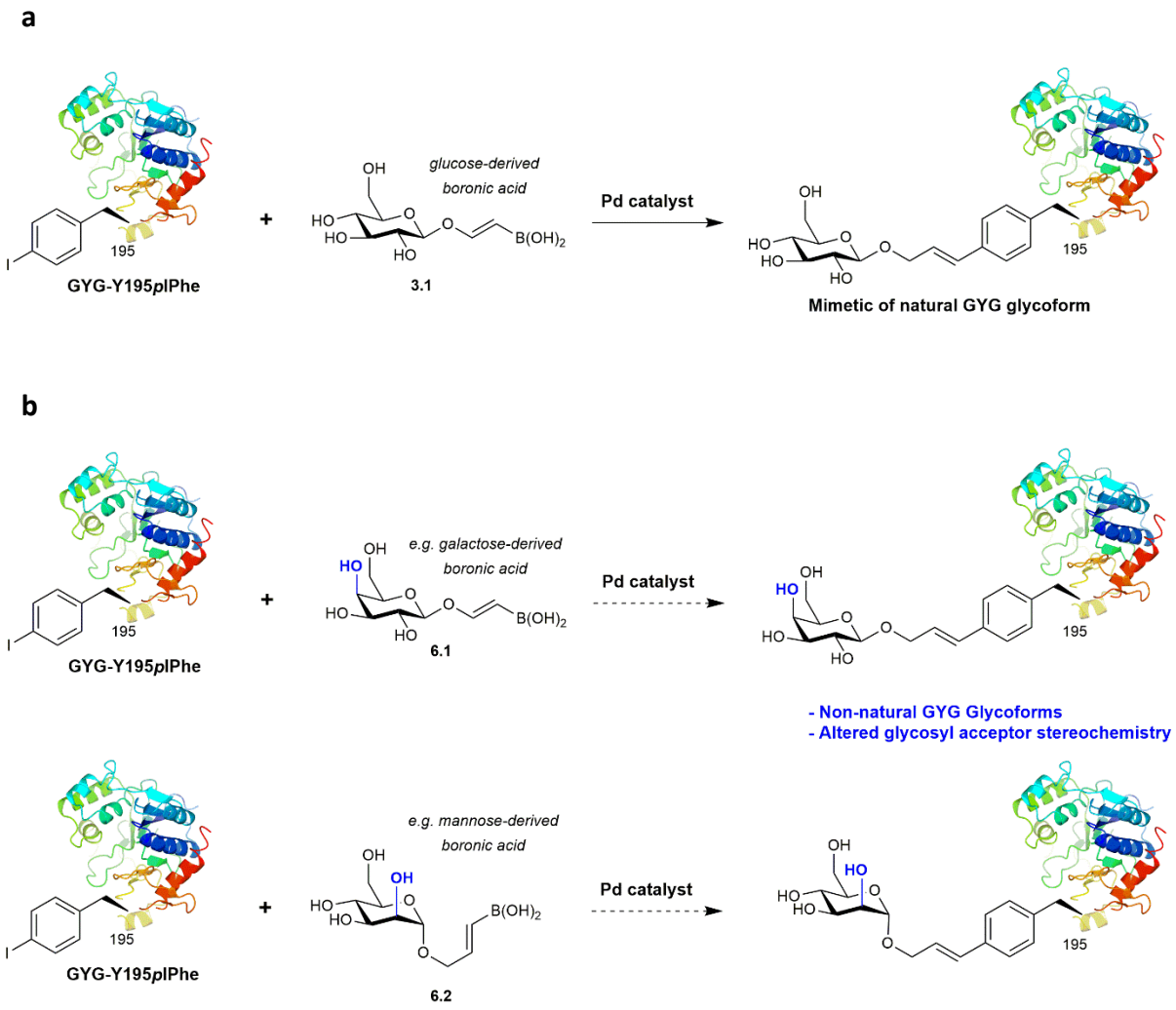


Figure 6.3 In the same way that, for example, cross-coupling of a glucose-derived boronic acid **3.1** to GYG-Y195*p*IPhe can yield a mimetic of a natural GYG glycoforms (a), non-natural GYG glycoforms may be accessed through cross-coupling of, for example, galactose- or mannose-derived boronic acids (**6.1** and **6.2**) to this tagged enzyme (b).

6.1.3 Modifying the Enzyme Scaffold: a Catalytically Impaired GYG Mutant

The GYG species generated in previous chapters and proposed in the section above represent a collectively diverse range, yet they are all ultimately generated from the same enzyme scaffold (GYG-Y195*p*IPhe). Mechanistic understanding of this enzyme, however, owes much to numerous studies not only on the wild-type system, but also on various catalytically impaired mutants.[24–28] For example, studies on Tyr195-

deficient GYG mutants were crucial in establishing this precise residue[i] as critical for function,[25,26] whilst the work of Hurley *et al* on GYG species lacking specific active site residues (Asp160 and Asp163) helped to define the precise catalytic roles of these residues.[24]

More recently, increased focus has been given to a Thr → Met mutation, T83M,[ii] known to prevent GYG from carrying out the initial autoglucosylation step, and thus all subsequent ones (**figure 6.4 a**).[7,29] This mutant GYG-T83M enzyme, in addition to lacking autoglucosylation activity, cannot glucosylate other GYG monomers in an inter-enzyme sense.[21] However, it can act as a substrate for glucosylation *by* other GYG monomers, a feature that Issoglio *et al* have successfully utilised in order to selectively investigate inter-enzyme glucosylation of GYG.[6]

Mechanistically, the lack of autoglucosylation activity has been shown to result from the loss of a key hydrogen bond between Thr83 and Asp162,[21] the latter known to be involved in glycosyl donor activation,[24] and from an impeded transition between "ground" and "active" states of the enzyme.[7] This fact serves to highlight how studies on such a mutant system may provide mechanistic information of broader relevance to GYG in general; the crucial role of this hydrogen bond in the catalytic mechanism of the wild-type enzyme, for example, is revealed precisely by the effects of its absence in GYG-T83M Furthermore, the T83M mutation gives rise to a variety of pathological symptoms (e.g. muscle weakness, cardiac arrhythmia) collectively referred to as Glycogen Storage Disease XV.[29,30] As such, not only are studies on the GYG-T83M system capable of providing insight in relation to GYG catalysis in general, but a complete mechanistic understanding of this mutant is desirable from a medical perspective.

Whilst it is known that the GYG-T83M enzyme cannot carry out initial Tyr195 autoglucosylation, no data exists concerning its ability or otherwise to carry out subsequent autoglucosylation steps (since the absence of the first step impedes study of those following); any presumption that the enzyme is generically inactive for all steps is not a valid one. We felt that our methodology could address this gap in current knowledge. Cross-coupling of carbohydrate-derived boronic acids – e.g. **3.1**, **4.1** – to a GYG-T83M-Y195*p*IPhe double mutant could afford unprecedented access to homogeneous, isolable GYG-T83M glycoforms, providing a means by which this question might be addressed (**figure 6.4 b**). Furthermore, should the mutant glycoforms prove inactive to autoglucosylation, one may then ask whether a specific glycoform is capable of acting as a substrate for inter-enzyme glucosylation, and thus potentially probe the existence of such a

[i] As noted in chapter 1, this is variously referred to as Tyr194 or Tyr195 in literature, depending on authors' amino acid numbering. According to our sequence (see chapter 2), Tyr195 is the correct annotation.
[ii] In a similar vein to note (i) above, "T83M" may in some instances be referred to as T82M in literature.[21] This refers to the same mutation; according to our sequence (see chapter 2), T83M is the correct annotation.

mechanism for *distinct* glucosylation steps. Given that the previously observed rate differences between GYG-Glucose and GYG-Maltose had served to define the first two phases of a triphasic kinetic profile for GYG (see chapter 4), we envisaged that generation and comparison of their mutant variants could further support and illuminate the differences between these two enzymes.

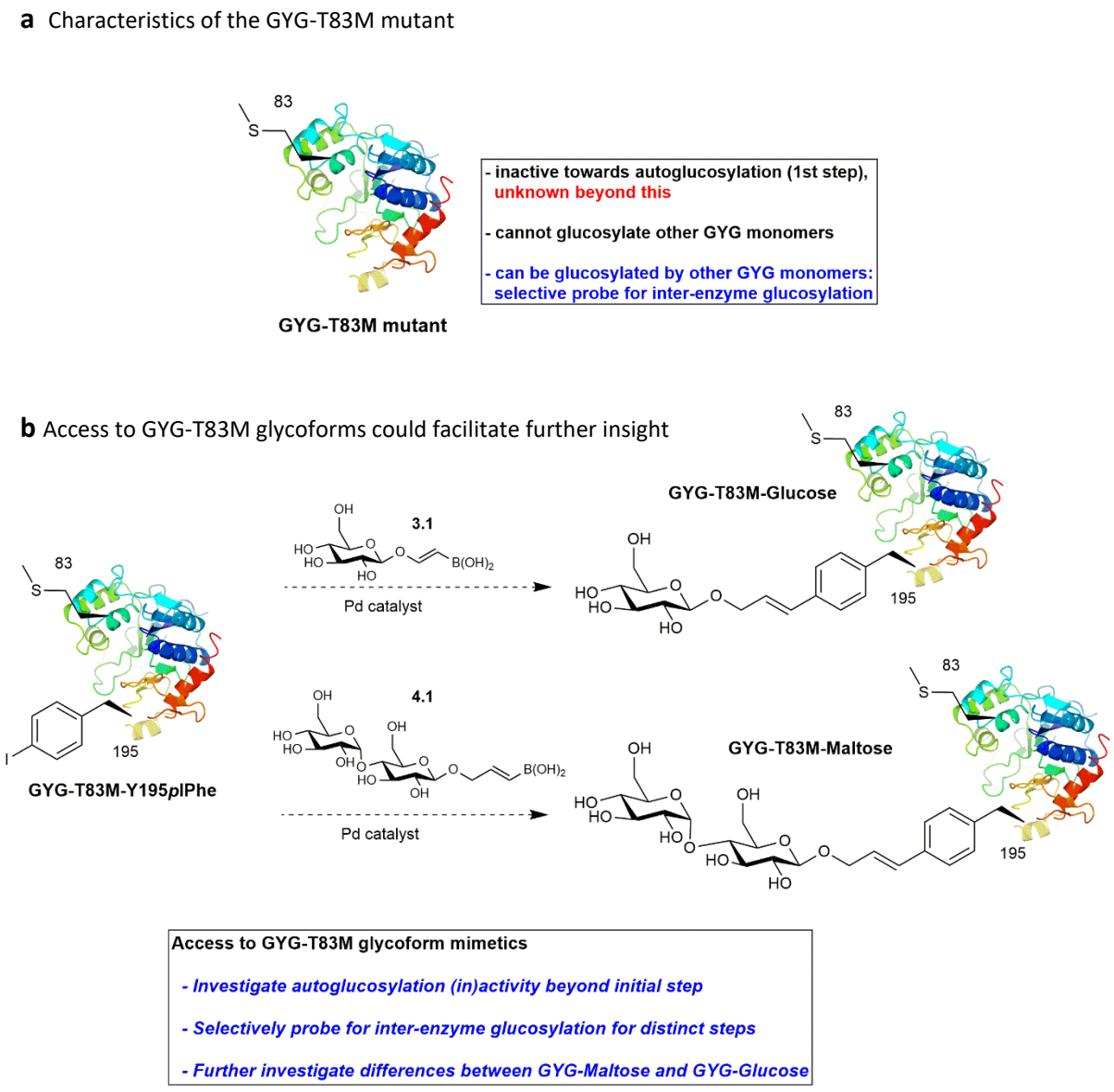


Figure 6.4 (a) The pathologically relevant GYG-T83M mutant cannot carry out the initial Tyr_{195} glucosylation (and thus those subsequent) and cannot glucosylate other GYG monomers, however its ability to be glucosylated by other GYG monomers renders it a selective "probe" for the existence of an inter-subunit glucosylation mode.[6] (b) Cross-coupling to carbohydrate boronic acids such as **3.1** and **4.1** could generate GYG-T83M glycoforms, enabling further study of the autoglucosylation (in)activity beyond the initial step, and of inter-subunit glucosylation for distinct glucosylation steps. Comparison of GYG-T83M-Maltose and GYG-T83M-Glucose may shed further light on the previously observed kinetic differences between these enzymes.

The remainder of this chapter describes the generation and investigation of non-natural GYG constructs according to the two strategies outlined above, beginning with stereochemical alteration of the initial glycosyl acceptor as discussed in section 6.1.2, and subsequently discussing modification of the enzyme scaffold itself. Through these approaches, we are able to further enrich our existing picture of GYG catalysis via the addition of still further non-natural pathways, in addition to the previously delineated autogalactosylation (chapter 4). Furthermore, evidence acquired for the GYG-T83M system suggests that this enzyme may not be generically inactive towards autoglucosylation, and enables further distinction of the GYG-Glucose and GYG-Maltose enzymes by means other than autoglucosylation rates.

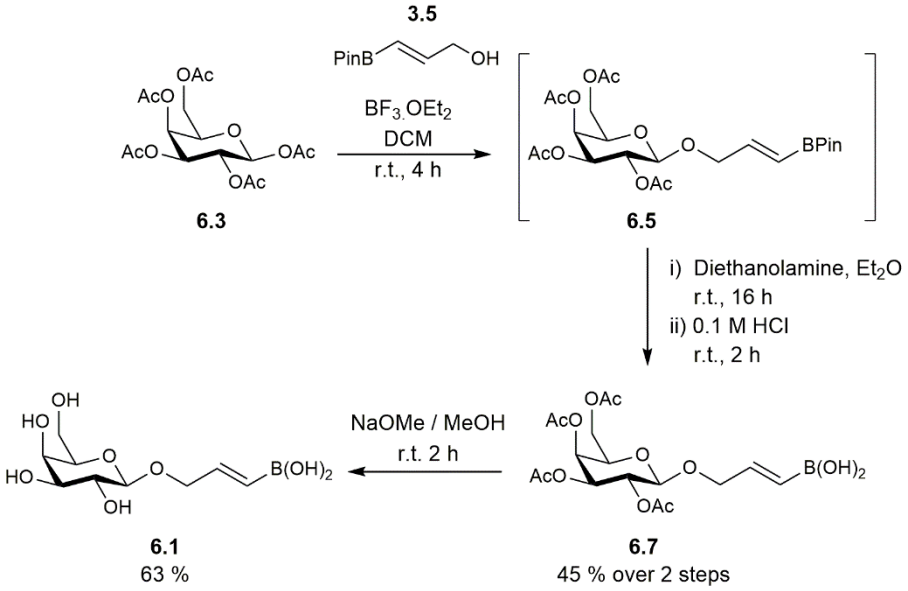
6.2 Investigating Unnatural GYG-Carbohydrate Acceptors

In section 6.1.2, the potential generation of GYG glycoforms bearing carbohydrate glycosyl acceptors of non-natural (i.e. non-glucose) stereochemistry via SMC was outlined (see **figure 6.3 b**). This concept was illustrated through the hypothetical use of galactose-derived and mannose-derived boronic acids **6.1** and **6.2** as examples, to generate GYG-Galactose and GYG-Mannose. Whilst this approach could be equally well applied to other stereoisomers of glucose, it was decided to pursue the investigation of these two GYG glycoforms. The study of GYG-Galactose held particular intrigue given the informative results previously obtained concerning autogalactosylation of GYG-Glucose and GYG-Maltose; investigation of the "reverse process" – utilising an initial galactose-based glycosyl acceptor, rather than glycosyl donor – was of interest.

For the synthesis of **6.1** and **6.2**, routes previously reported by Spicer *et al* were adopted;[31] pleasingly, these proved effective in our hands with few modifications required. Thus, either β-D-galactose pentaacetate **6.3** or α-D-mannose pentaacetate **6.4** was subjected to glycosylation with previously synthesised pinacol ester **3.5** (see chapter 4), affording the intermediate pinacol ester species **6.5** and **6.6** (**figure 6.5**). Subsequent transesterification with diethanolamine (DEA),[32] followed by acidic hydrolysis of the resultant DEA ester, afforded the peracetylated boronic acids **6.7** or **6.8**. Global deprotection then yielded the requisite target compounds as exclusively the β-anomer (**6.1**) or α-anomer (**6.2**). A notable improvement in the synthesis of **6.1** as compared to the prior iteration[31] is a lack of requirement for purification of the pinacol ester **6.5**; this is rendered possible by the fact that conversion of the pinacol ester into the boronic acid proceeds via a diethanolamine ester intermediate that selectively precipitates from solution; other boronic acid-lacking species remain in solution. This was not possible for the mannose system owing to incomplete consumption

of the glycosyl acceptor **3.5**, which also contains a pinacol ester functionality that could also form a diethanolamine ester and co-precipitate.

a Synthesis of galactose-derived boronic acid **6.1**



b Synthesis of mannose-derived boronic acid **6.2**

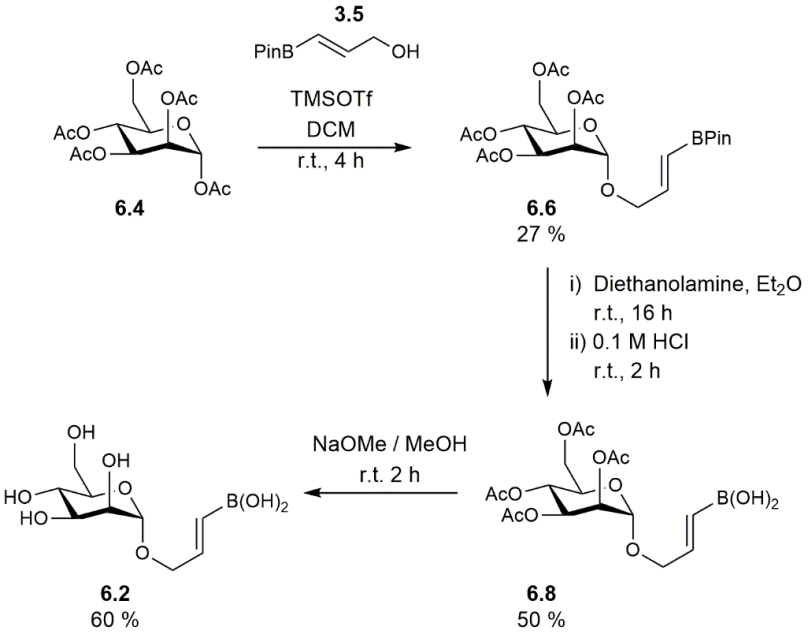


Figure 6.5 Synthesis of galactose-derived (a) and mannose-derived (b) boronic acids **6.1** and **6.2**.

Cross-coupling of **6.1** and **6.2** to GYG-Y195*p*IPhe was then attempted. Pleasingly, the conditions optimised for the GYG-Glucose and GYG-Maltose variants proved readily translatable to the generation of GYG-Mannose, affording the desired products in good to excellent conversions (82 % for GYG-Mannose, 74 % for GYG-Galactose, **figure 6.6**). In contrast to previous systems (and to GYG-Mannose, see experimental section 6.5.1.3), initial experiments indicated that the omission of Na_2HPO_4 from the reaction mixture was, if anything, slightly detrimental.

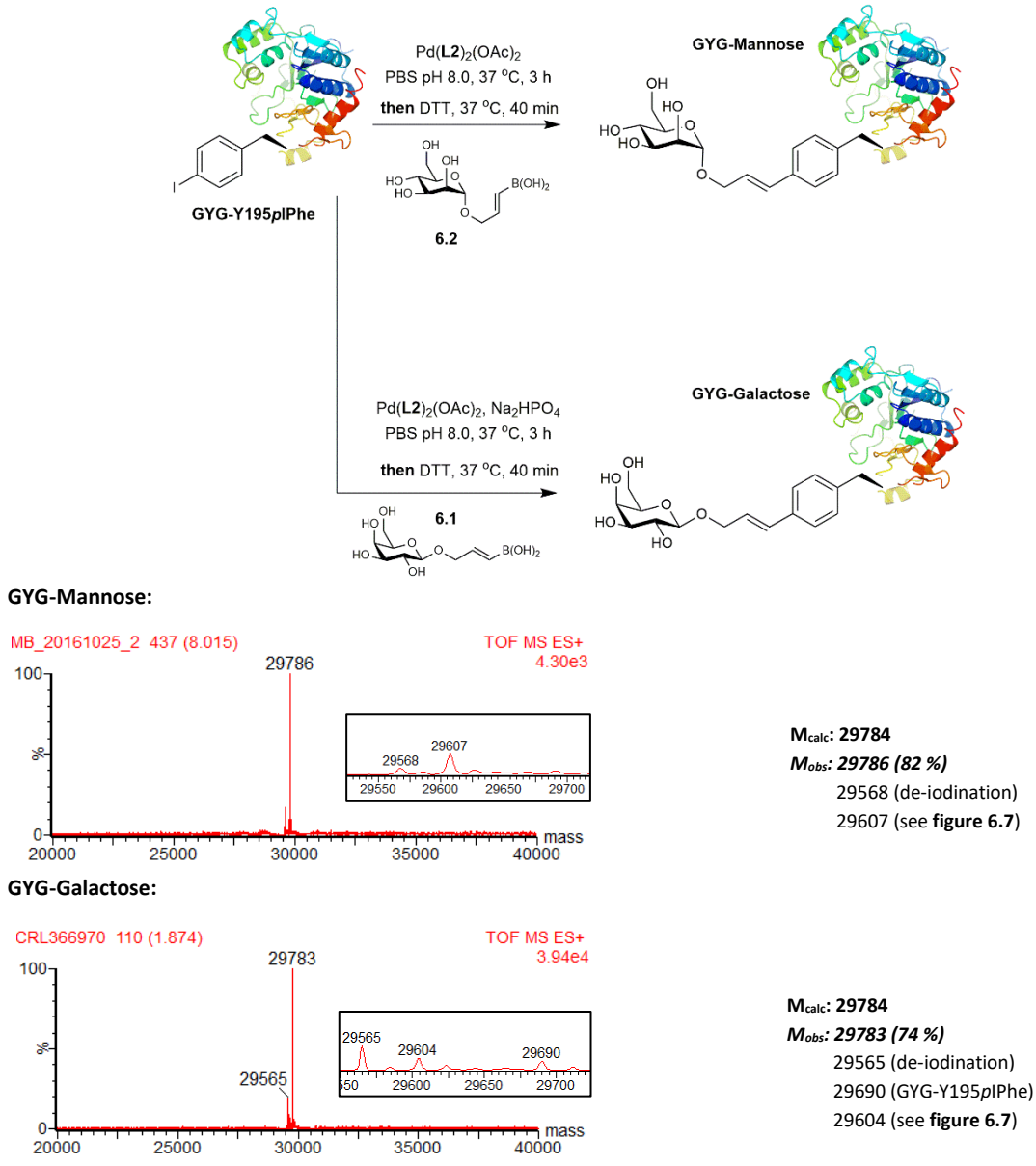


Figure 6.6 Cross-coupling of GYG-Y195*p*IPhe to generate GYG-Mannose and GYG-Galactose.

Closer inspection of LC-MS spectra confirms that the side-products formed are the same non-glucosylatable species as previously observed for all prior carbohydrate-based cross-couplings, namely de-iodination (M_{calc} = 29565), unreacted GYG-Y195*p*IPhe (GYG-Galactose only, M_{calc} = 29692), and another

species that, as discussed previously (see chapter 3 pp. 172-3), we believe to correspond to (hydrogenolytic) loss of the carbohydrate moiety to afford a structure such as **3.11** (**figure 6.7**).

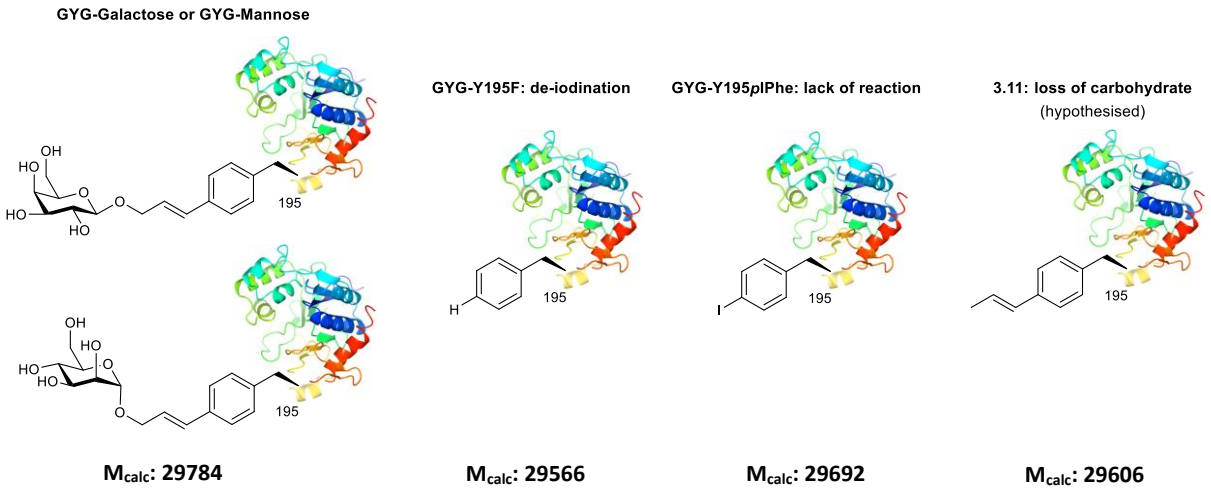


Figure 6.7 Illustration of species observed during cross-couplings of GYG-Y195*p*IPhe to **6.1** and **6.2**. These are identical to species observed in all prior carbohydrate couplings (see chapter 3, pp. 172-3 for further discussion).

In chapter 3, we further investigated the natures of the side-products in the context of the GYG-Glucose system (see pp. 172-3); key conclusions being that they all arose from the *p*IPhe195 site (GYG-Y195F failed to react), and that any "quenching" of an enzyme-aryl-Pd intermediate on MeCN upon LC-MS analysis was not a source of the M_{calc} = 29606 species (presence or absence of MeCN had no effect). In order to confirm these conclusions, we repeated experiments previously carried out on the GYG-Glucose system (see chapter 3 pp. 172-3) in the context of the GYG-Galactose system described above. Pleasingly, results obtained remained identical. Specifically, just as GYG-Y195F underwent no reaction under cross-coupling conditions with glucose boronic acid **3.1**, it also failed to react with galactose boronic acid **6.1**, further confirming *p*IPhe195 as the site of product and side-product formation; in addition, side-products formed during GYG-Galactose generation were not affected by presence or absence of MeCN during LC-MS analysis (see experimental section 6.5.1.4 for full details and spectra).

Subsequently, the autoglucosylation activity of these enzymes was investigated. To our initial surprise, LC-MS analysis after glucosylation assay (1 mM UDP-Glucose, 1 mM $MnCl_2$, as previously described in chapter 4) revealed that both of these enzyme species were capable of autoglucosylation (**figure 6.8**). As previously seen for other enzymes, the MS peak corresponding to cross-coupled enzyme split into multiple peaks corresponding to addition of glucose residues. As noted above and in previous chapters, cross-coupling side-products exhibited no such autoglucosylation, consistent with their lack of a glycosyl acceptor.

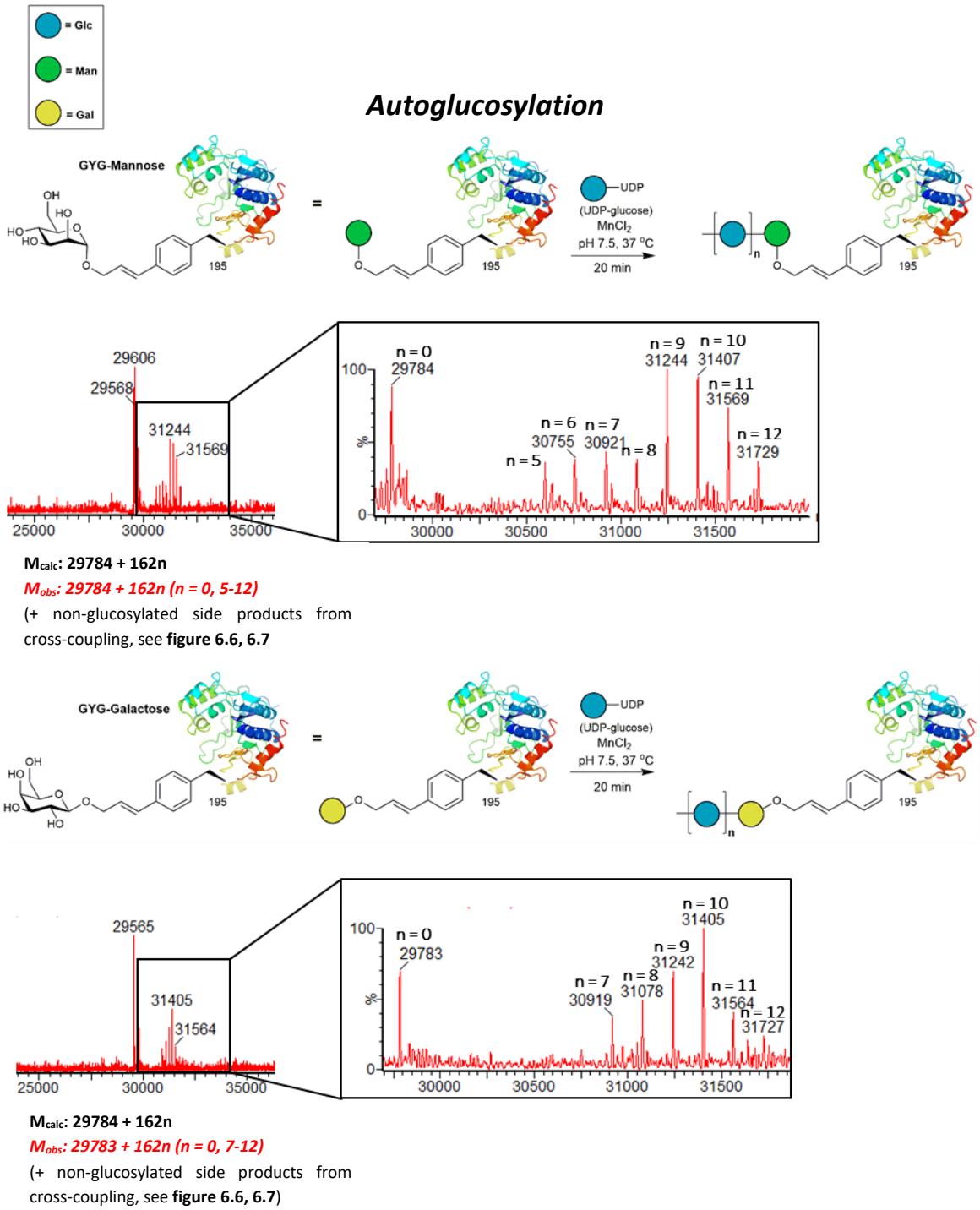


Figure 6.8 Autoglucosylation of GYG-Mannose (top) and GYG-Galactose (lower).

Kinetic evaluation of the autoglucosylation of these enzymes was undertaken, according to the LC-MS-based method outlined in chapter 4 (see pp. 238-9, and experimental section 4.5.3). Thus, in brief, LC-MS analysis of quenched aliquots of GYG-Galactose or GYG-Mannose after different times of glucosylation assay were analysed to determine proportions of the cross-coupled enzyme remaining and the average numbers of monosaccharide residues (Gal/Man and Glc) attached per enzyme (see chapter 4 for details).

Comparison to analogous prior data generated for GYG-Glucose (chapter 4) revealed the presence of a slow initial glucosylation step, significantly moreso than that observed for GYG-Glucose (**figure 6.9**). This is consistent with the non-natural nature of the glycosyl acceptor for this first step. As a consequence, the overall glucosylation profile was additionally altered with respect to GYG-Glucose, with a slower rate during the early stages of catalysis, reflecting the influence of this slow initial step.

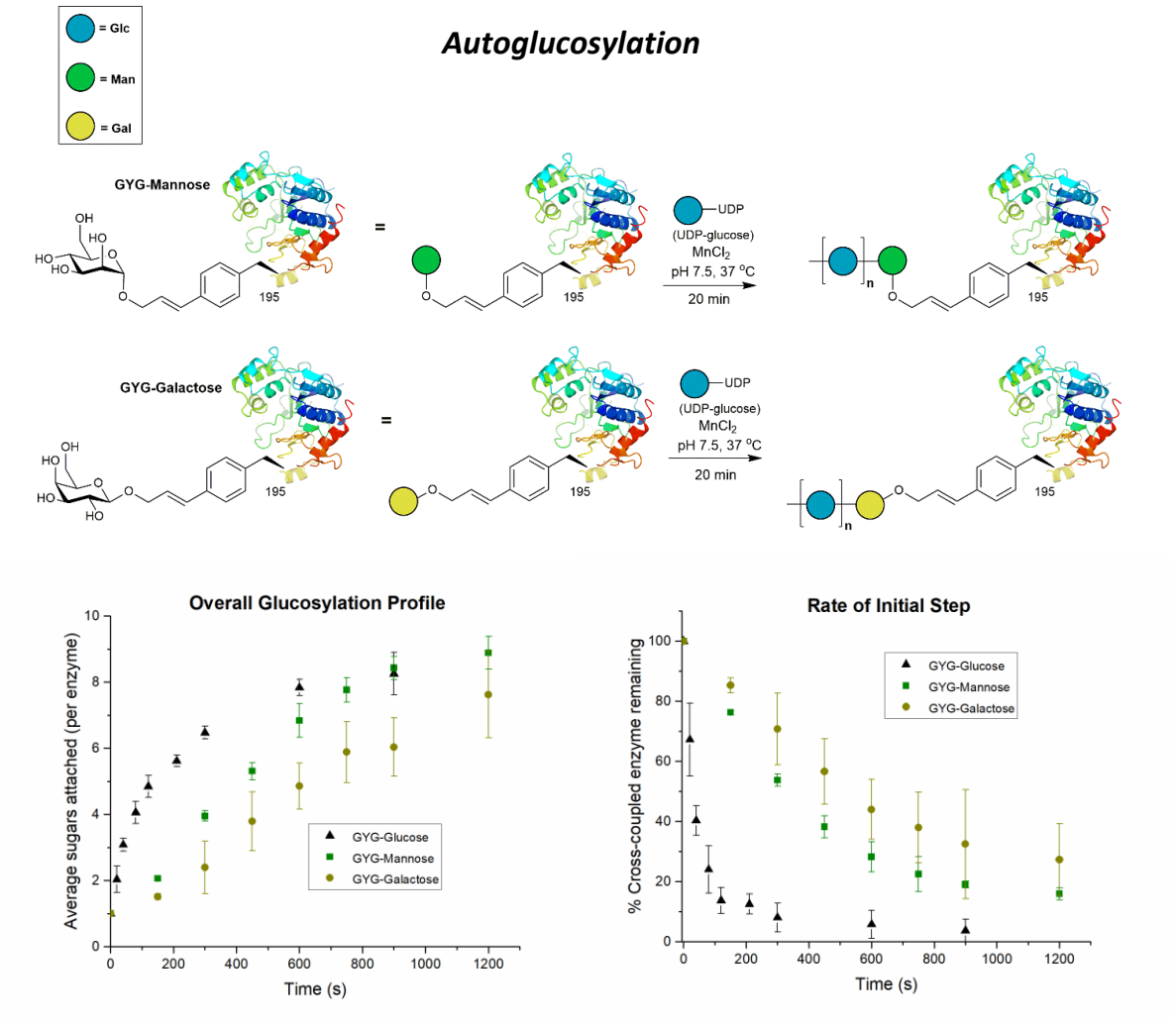
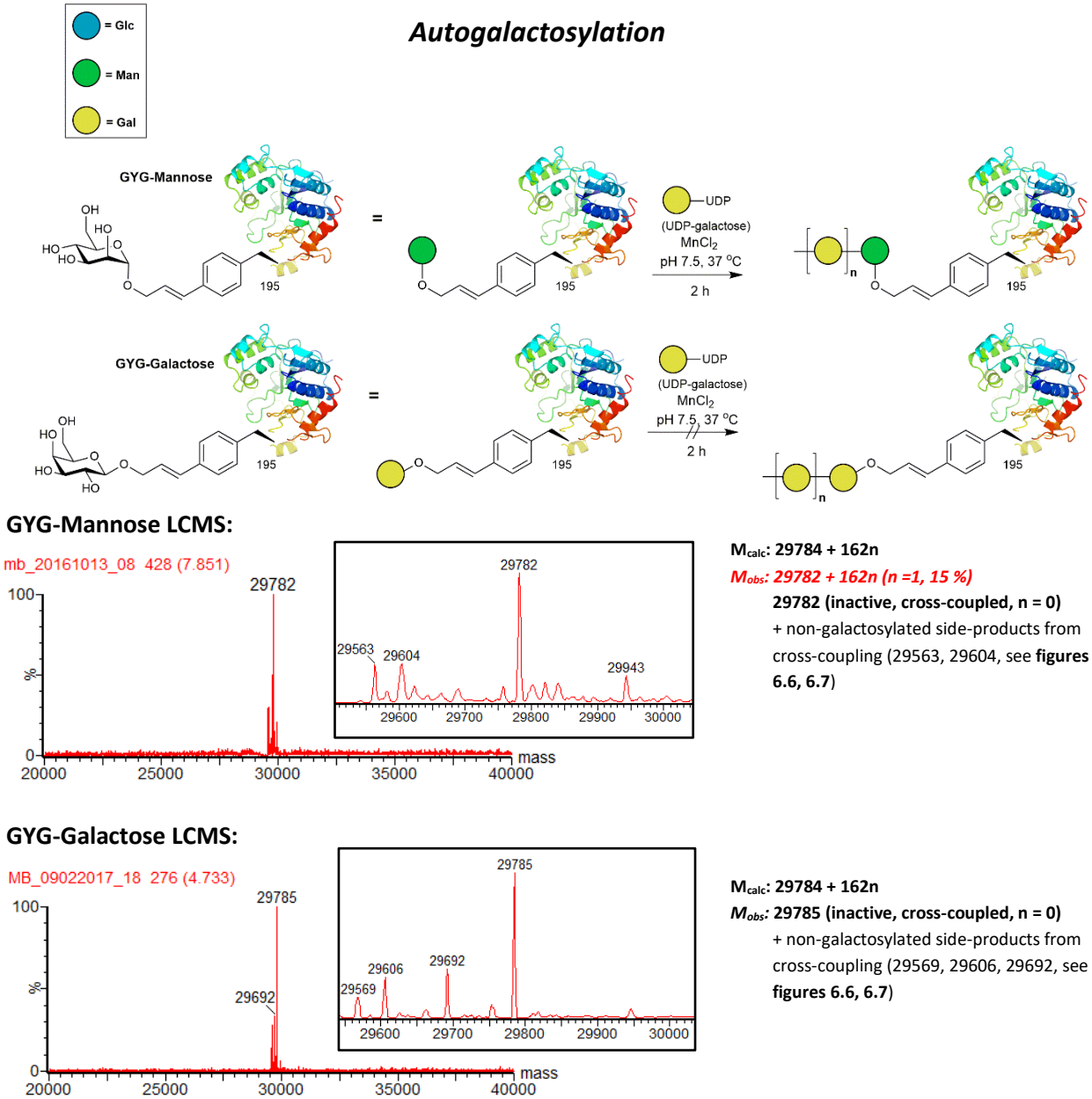


Figure 6.9 Autoglucosylation kinetics of GYG-Galactose and GYG-Mannose, compared to GYG-Glucose. Data represent mean averages of at least 3 independent replicates, error bars are ± s.d. (Data for GYG-Glucose is that from chapter 4).

It will be noted that, as per the autogalactosylation studies detailed in chapter 4, the monosaccharide residues are represented using glycan symbol notation;[33] this, again, reflects the fact that, at present, the precise nature of the glycosidic linkage between the initial mannose/galactose and the subsequent glucose residue remains unknown. Protein NMR studies using ^{13}C-labelled carbohydrates to potentially elucidate this represents a future avenue of interest; this was, however, beyond the scope of the present work.

In contrast to the above glucosylation, GYG-Galactose displayed no detectable capacity for autogalactosylation from UDP-galactose after 2 hours (**figure 6.10**); the implications of this result with regards to the overall catalytic mechanism are discussed on the following page. For GYG-Mannose, only limited levels of cross-coupled enzyme (15 %) underwent galactosylation, and this was restricted to mono-galactosylation only, i.e. no Gal-to-Gal transfer occurred. These results contrast with the ready autogalactosylation, including multiple such Gal-to-Gal transfers, of the corresponding "natural" enzyme GYG-Glucose (see chapter 4).



M_{calc}: 29784 + 162n

M_{obs}: 29782 + 162n (n =1, 15 %)

29782 (inactive, cross-coupled, n = 0)
+ non-galactosylated side-products from cross-coupling (29563, 29604, see **figures 6.6, 6.7**)

M_{calc}: 29784 + 162n

***M_{obs}:* 29785 (inactive, cross-coupled, n = 0)**
+ non-galactosylated side-products from cross-coupling (29569, 29606, 29692, see **figures 6.6, 6.7**)

Figure 6.10 Attempted autogalactosylation of GYG-Mannose and GYG-Galactose. After 2 hours, at best minimal galactosylation was observed, with no more than 1 Gal addition.

Prior work in this thesis delineated a triphasic "slow-fast-slow" kinetic profile for the natural autoglucosylation pathway of GYG. In addition, various "offshoots" to an unnatural autogalactosylation pathway were identified, contributing to an overall mechanistic picture outlined previously (**figure 4.12**, chapter 4). For reference in subsequent discussions, this figure is reproduced overleaf (**figure 6.11**). Furthermore, in chapter 4 it was discussed that such galactosylation– i.e. Gal-to-Gal transfer – appeared chalenging at oligosaccharide lengths corresponding to the 3rd kinetic phase of the natural glucosylation pathway, and could not proceed for much of this phase. In contrast, during the preceding 2nd phase, Gal-to-Gal transfer is possible throughout (species bearing up to 5 sugar residues – the close of this phase – were routinely detected at earliest time points of autogalactosylation). This was suggestive of additional differentiation of these two phases with regards to tolerance of this reaction.

Based on the data obtained in this chapter, this mechanistic picture for our system may be further embellished (**figure 6.12**); new aspects and conclusions of this figure with respect to **figure 6.11** are highlighted in red, whilst aspects previously delineated in chapter 4 are in black, hence the former should be focussed upon when examining this figure. Firstly, a second unnatural pathway may be added, involving autoglucosylation from non-natural enzyme-linked monosaccharide acceptors, namely galactose and mannose (this may well also be triphasic in nature, but at present this cannot be definitively stated). Secondly, the data above suggest that, although unnatural autogalactosylation pathways from GYG-Glucose and GYG-Maltose exist, these may not initiate from GYG-Galactose itself, and only to a very limited extent from GYG-Mannose. Indeed, whilst the natural autoglucosylation pathway may initiate from enzyme-linked monosaccharides of varying stereochemistry, highlighting plasticity of the enzyme, galactosylation activity would seem to largely require that the initial enzyme-linked carbohydrate residue is glucose.

We suggest that the three kinetic phases delineated for the natural autoglucosylation pathway exhibit differences not just in their relative autoglucosylation rates, but also in their tolerance of Gal-to-Gal transfer reactions. The apparent differentiation of the 2nd and 3rd such phases in this regard was discussed above and in chapter 4, on the basis of autogalactosylation studies on GYG-Glucose and GYG-Maltose. However, the same approach could not provide any information regarding Gal-to-Gal transfer at carbohydrate chain lengths corresponding to the 1st kinetic phase (< 2 sugars). Crucially, access to the construct GYG-Galactose, uniquely accessible through our methodology, enables the new observation that, for this first phase (< 2 sugars), such Gal-to-Gal transfer is not merely limited in extent but is not at all

possible (this key result held true across two replicates, see experimental section 6.5.1.2), further differentiating it from the phases following it.

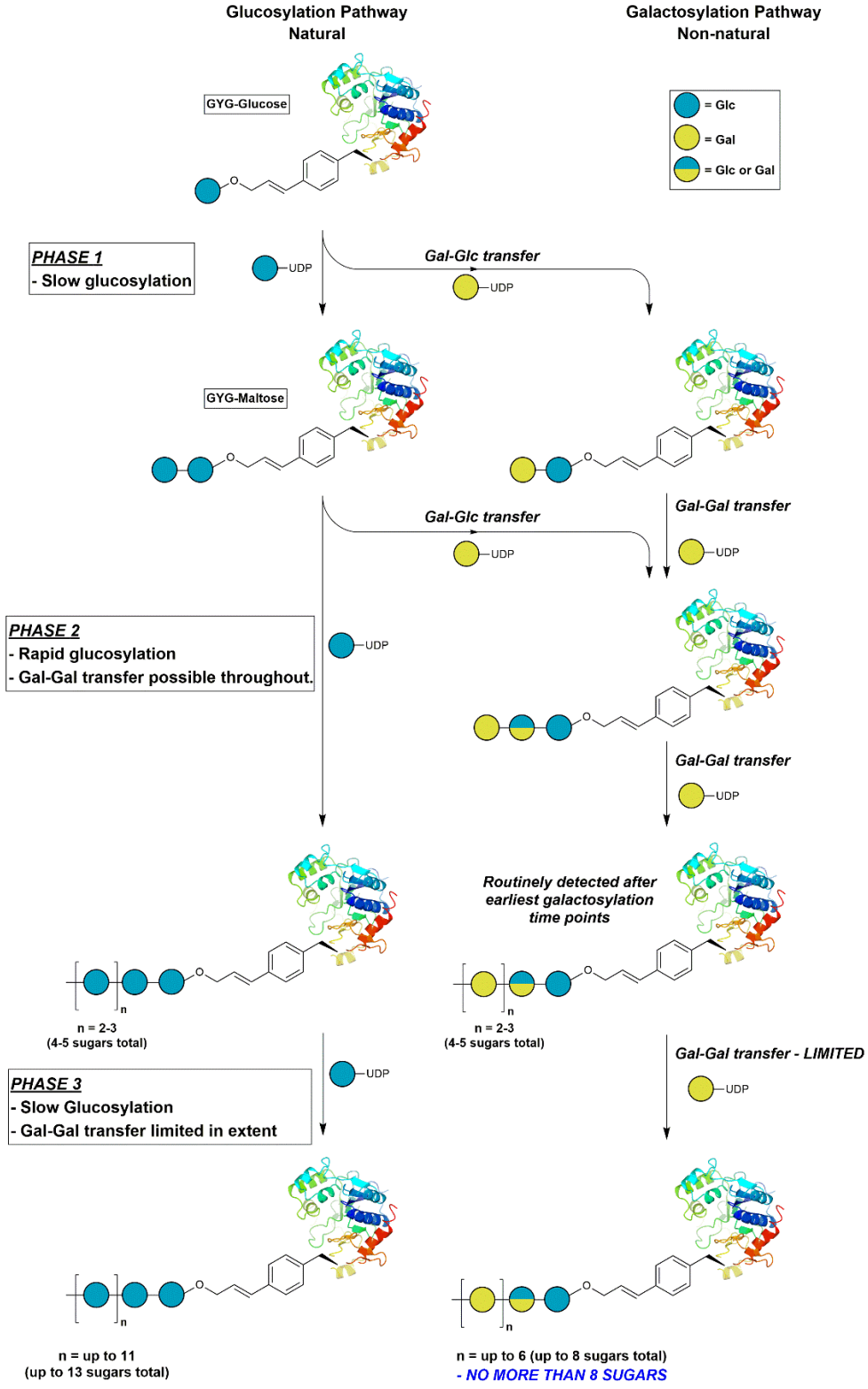


Figure 6.11 Reproduction of figure 4.12 (chapter 4), illustrating previous mechanistic picture for GYG catalysis prior to this chapter. Key points are highlighted in boxes (left), see chapter 4 for full discussion of these.

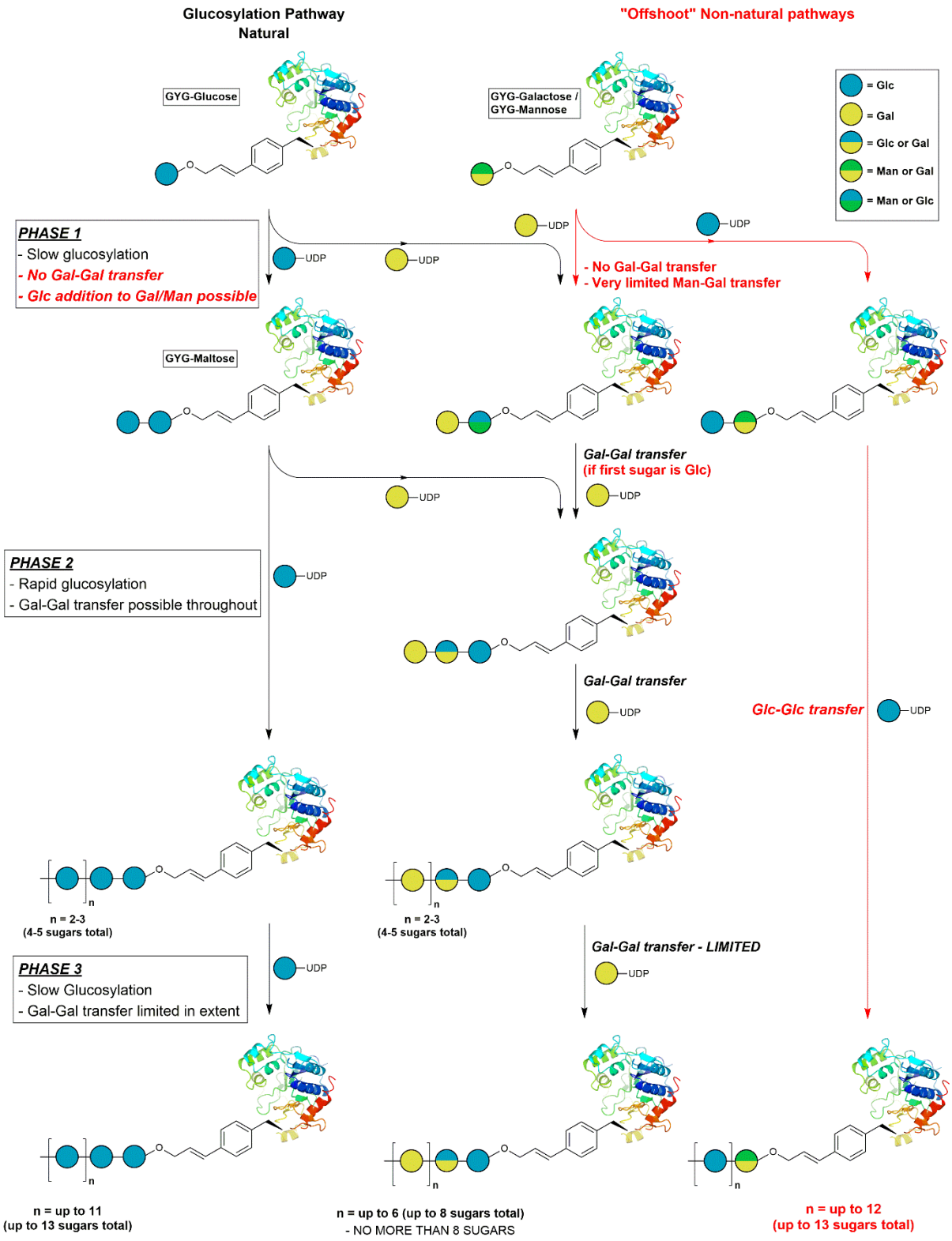


Figure 6.12 Embellished mechanistic picture for *in vitro* GYG catalysis within our system. New additions and conclusions with respect to figure 6.11 are highlighted in red; prior knowledge is in black. Key new points are thus the existence of new glucosylation pathways initiation from Gal / Man, and the lack of Gal-to-Gal transfer in the first phase; multiple Gal-to-Gal transfer can only occur if the initial glycosyl acceptor is glucose-derived.

It should be noted that at least the GYG-Mannose pathways almost certainly represent an "*in vitro* only" process; GYG will not incorporate mannose from its most abundant nucleotide GDP-mannose in an *in vivo* setting, owing to its well-established specificity for pyrimidine nucleotides.[18,34] In contrast, and as discussed previously in chapter 4, incorporation of galactose from UDP-galactose is, in principle, plausible, and possible *in vivo* relevance of this and the differing Gal-Gal transfer tolerances is speculated upon as part of the overall conclusion in chapter 7. It should be noted that at present we do not know whether a species corresponding to GYG-Galactose could in fact be generated by non-chemical means; this would require galactosylation of non-glucosylated GYG, which has not been studied in this work or previously. Further exploration of this latter enzyme in general, in light of the work in this thesis, represents a future avenue of interest.

6.3 Investigation of Catalytically Impaired GYG-T83M Glycoforms

We next turned to the study of a second class of non-natural GYG glycoforms, and the only enzyme constructs in this thesis not generated directly from GYG-Y195*p*IPhe; those derived instead from a mutant variant of this enzyme, GYG-T83M-Y195*p*IPhe. As discussed previously (section 6.1.3), we were particularly interested in the generation and study of T83M mutants of GYG-Glucose and GYG-Maltose, accessible via cross-coupling of a GYG-T83M-Y195*p*IPhe variant to previously synthesised glucose- and maltose-derived boronic acids **3.1** and **4.1** (see **figure 6.4**). Data obtained further highlights the distinctions between the identified first and second catalytic phases within our system, and, crucially, suggests that the GYG-T83M enzyme may not in fact be entirely autoglucosylation-inactive.

6.3.1 Generation of GYG-T83M-Y195*p*IPhe

In order to study the mutant enzymes "GYG-T83M-Glucose" and "GYG-T83M-Maltose", it was first necessary to generate their common precursor, GYG-T83M-Y195*p*IPhe. In an analogous fashion to the expression of the single mutant GYG-Y195*p*IPhe, this double mutant enzyme should be accessible through amber codon suppression technology.[35,36] Specifically, co-expression of the pEVOL-*p*IPhe plasmid, encoding a tRNA/aaRS pair enabling incorporation of *p*IPhe in response to a TAG codon installed in a DNA template (see chapter 2),[37] and a plasmid encoding GYG-T83M-Y195X (X = TAG codon), should thus yield GYG-T83M-Y195*p*IPhe.

The installation of the Thr → Met mutation at position 83 was achieved via standard site-directed mutagenesis techniques, using GYG-Y195X as the template DNA (see experimental section 6.5.2.1). Following transformation of the PCR product into *E. coli*, numerous colonies were observed; plasmid DNA from 4 of these was isolated and sequenced. In each case, the successful installation of the T83M mutation was confirmed, affording the plasmid GYG-T83M-Y195X (**figure 6.13**). It should be noted that, for consistency with previous chapters and prior literature,[7,24,28] amino acid numbering does not include the N-terminal 22-residue moiety containing the His-tag and TEV cleavage site (highlighted in yellow below).

DNA and Protein Sequence:

GYG-T83M-Y195X (X = TAG codon = *p*IPhe in expressed protein)

```
    atgcaccatcatcatcatcattcttctggtgtagatctgggtaccgagaacctgtacttc
     M  H  H  H  H  H  H  S  S  G  V  D  L  G  T  E  N  L  Y  F
    caatccatgacagatcaggcctttgtgacactaaccacaaacgatgcctacgccaaaggt
     Q  S  M  T  D  Q  A  F  V  T  L  T  T  N  D  A  Y  A  K  G
19  gccctggtcctgggatcatctctgaaacagcacaggaccaccaggaggctggtcgtgctc
     A  L  V  L  G  S  S  L  K  Q  H  R  T  T  R  R  L  V  V  L
39  gccacccctcaggtctcagactccatgagaaaagttttagagacagtctttgatgaagtc
     A  T  P  Q  V  S  D  S  M  R  K  V  L  E  T  V  F  D  E  V
59  atcatggtagatgtcttggacagtggcgattctgctcatctaaccttaatgaagaggcca
     I  M  V  D  V  L  D  S  G  D  S  A  H  L  T  L  M  K  R  P
79  gagttgggtgtcatgctgacaaagctccactgctggtcgcttacacagtattcaaaatgt
     E  L  G  V  M  L  T  K  L  H  C  W  S  L  T  Q  Y  S  K  C
99  gtattcatggatgcagatactctggtcctagcaaatattgatgatctttttgacagagaa
     V  F  M  D  A  D  T  L  V  L  A  N  I  D  D  L  F  D  R  E
119 gaattgtcagcagcaccagacccagggtggcctgactgcttcaattccggagtcttcgtt
     E  L  S  A  A  P  D  P  G  W  P  D  C  F  N  S  G  V  F  V
139 tatcagccttcagttgaaacatacaatcagctgttgcatcttgcttctgagcaaggtagt
     Y  Q  P  S  V  E  T  Y  N  Q  L  L  H  L  A  S  E  Q  G  S
159 tttgatggtggggaccaaggcatactgaacacatttttagcagctgggcaacaacagat
     F  D  G  G  D  Q  G  I  L  N  T  F  F  S  S  W  A  T  T  D
179 atcagaaaacacctgccgtttatttataacctaagcagcatctctatatagtcctacctc
     I  R  K  H  L  P  F  I  Y  N  L  S  S  I  S  I  -  S  Y  L
199 ccggcatttaaagtgtttggtgcaagtgccaaagttgtgcatttcctgggacgagtcaaa
     P  A  F  K  V  F  G  A  S  A  K  V  V  H  F  L  G  R  V  K
219 ccatggaattatacttatgatcccaaaacaaaaagtgtcaaaagtgaggcccatgatccc
     P  W  N  Y  T  Y  D  P  K  T  K  S  V  K  S  E  A  H  D  P
239 aacatgactcatccagagtttctcatcctgtggtggaacatctttaccaccaacgtttta
     N  M  T  H  P  E  F  L  I  L  W  W  N  I  F  T  T  N  V  L
259 cctctgcttcaatga
     P  L  L  Q  -
```

(**"tag"** = *p*IPhe in expressed protein. Yellow region = 22-residue section containing His-tag and TEV cleavage site. Amino acid numbering does *not* include this moiety. There are 20 residues per full line.)

Calculated mass of protein (X = *p*IPhe): 32188 (his-tagged), 29722 (his-tag-cleaved).

Figure 6.13 DNA and protein sequence details for the GYG-T83M-Y195X plasmid (X = TAG).

With requisite plasmids to hand, expression of the GYG-T83M-Y195*p*IPhe enzyme was attempted. Conditions previously optimised for the analogous expression of GYG-Y195*p*IPhe were adopted; thus, following growth in SB media (see experimental section) to an OD_{600} of 0.7, the expression of the pEVOL-*p*IPhe plasmid was induced (30 °C, 2 h) via addition of L-arabinose, with concomitant addition of L-*p*-iodophenylalanine, before induction of the GYG-T83M-Y195X plasmid (18 °C, overnight). Cells were lysed and the protein purified by Ni-NTA affinity chromatography. SDS-PAGE and LC-MS analysis revealed successful isolation of the desired double mutant enzyme, GYG-T83M-Y195*p*IPhe, and confirmed good levels of purity (**figure 6.14**). Yields were similar to those obtained for the single mutant GYG-Y195*p*IPhe (ca. 1 mg/L culture).

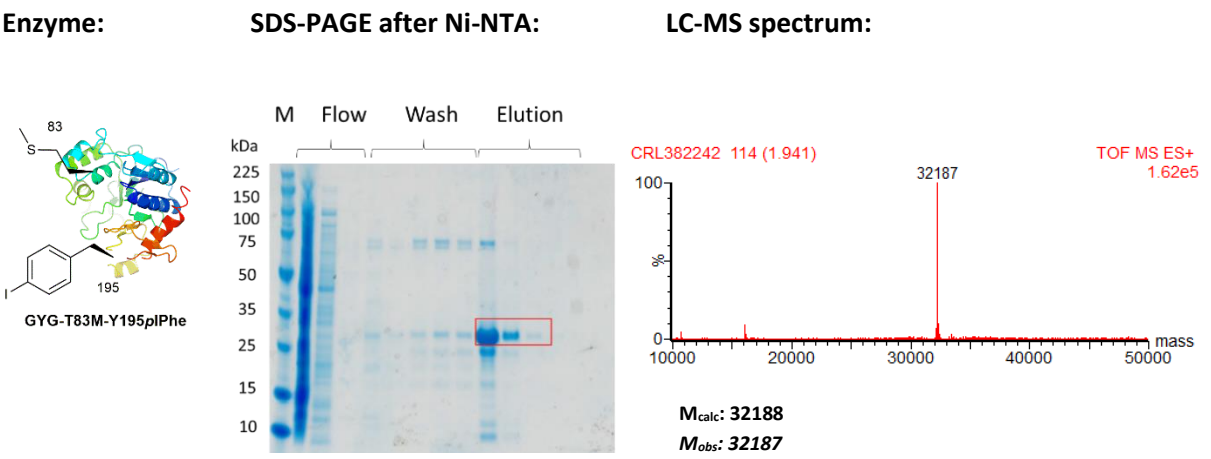


Figure 6.14 Expression of GYG-T83M-Y195*p*IPhe. SDS-PAGE after Ni-NTA affinity purification and deconvoluted LC-MS spectrum are shown.

6.3.2 Catalytic Activity of GYG-T83M Mutant Glycoforms

Consistent with prior studies on GYG-Y195*p*IPhe, it was decided to first proceed with investigations into cross-coupling and catalytic activity utilising the isolated his-tagged GYG-T83M-Y195*p*IPhe, and to only resort to (optimisation of) his-tag cleavage and subsequent further purification if rendered necessary (by e.g. low quality LC-MS spectra, as previously found for GYG-Glucose, see chapter 3). It should be noted that the assays outlined below are qualitative "active or inactive" in nature, much like those outlined in chapter 2 (in which his-tagged enzyme was used), and that our prior results for GYG-Glucose (see chapter 3) demonstrate that the presence or absence of a his-tag does not affect enzymatic activity in this sense. Furthermore, all conclusions from this section are derived only from internal comparisons of catalytic activity between similarly his-tagged enzymes, not comparisons to any prior data.

The cross-coupling of GYG-T83M-Y195*p*IPhe to glucose- and maltose-derived boronic acids **3.1** and **4.1** was therefore investigated, utilising reaction parameters based on those previously employed for analogous couplings of GYG-Y195*p*IPhe. Pleasingly, successful formation of cross-coupled product was observed in each case (**figure 6.15**). Formation of GYG-T83M-Glucose proceeded with goo conversions (70 %); in this case, the only observed side-product arose from de-iodination. Cross-coupling to generate GYG-T83M-Maltose proceeded with lower efficiency (50 %), principally reflecting higher levels of unreacted GYG-T83M-Y195*p*IPhe alongside de-iodination; nonetheless, sufficient quantities of the cross-coupled enzyme were generated for the purposes of subsequent investigation.

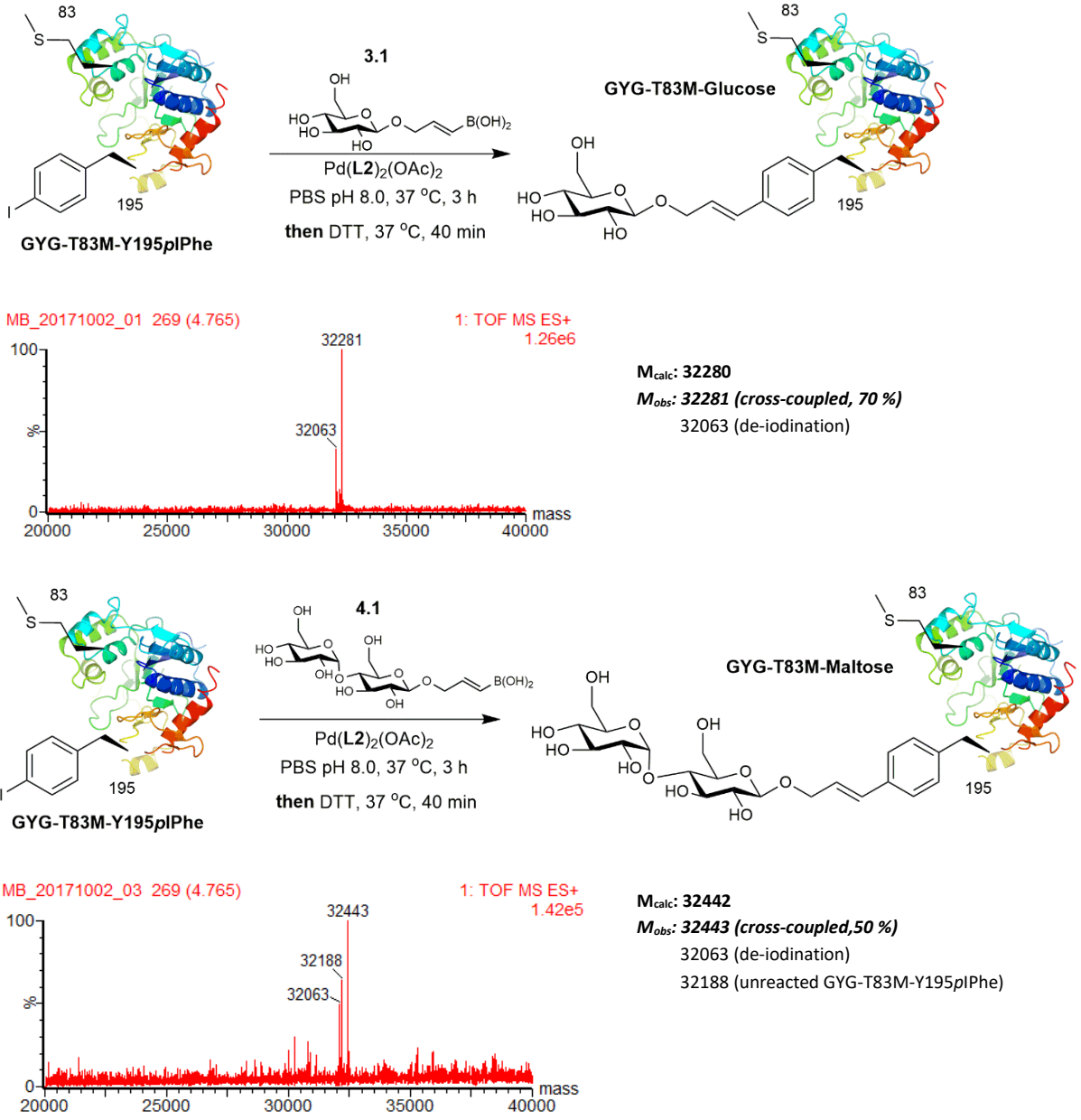


Figure 6.15 Cross-couplings between his-tagged GYG-T83M-Y195*p*IPhe and boronic acids **3.1** and **4.1**, to generate GYG-T83M-Glucose and GYG-T83M-Maltose respectively.

As discussed in prior chapters, owing to the lack of a suitable glycosyl acceptor present in such cross-coupling side products, they do not act as glucosylation substrates. Furthermore, given the very close structural correspondence between such species and wild-type GYG-T83M (de-iodination and unreacted enzyme represent minimal OH → H or OH → I "atomic mutations" of GYG-T83M), the inability of the wild-type GYG-T83M enzyme to catalyse glucosylation of an alternative GYG monomer should logically extend to these species.[6] As such, they should exhibit no activity in any sense, and thus their abundance relative to cross-coupled enzyme should not influence results pertaining to the latter.

The catalytic activity of the obtained cross-coupled enzymes was next investigated. As outlined above, whilst it might be expected that these enzymes would be entirely incapable of self-glucosylation given the known incapacity of GYG-T83M itself to autoglucosylate,[6,7,21] glycoforms of GYG-T83M have never previously been isolated and studied in this regard. We first investigated the monosaccharide glycoconjugate, GYG-T83M-Glucose. After 20 minutes incubation with UDP-Glucose and Mn^{2+}, no autoglucosylation was detectable by LC-MS (**figure 6.16**); by comparison, the analogous non-mutant GYG-Glucose, and all other GYG-carbohydrate enzymes, showed clear glucosylation within 20-40 seconds (chapters 3-5), as such this enzyme would appear to be essentially inactive.

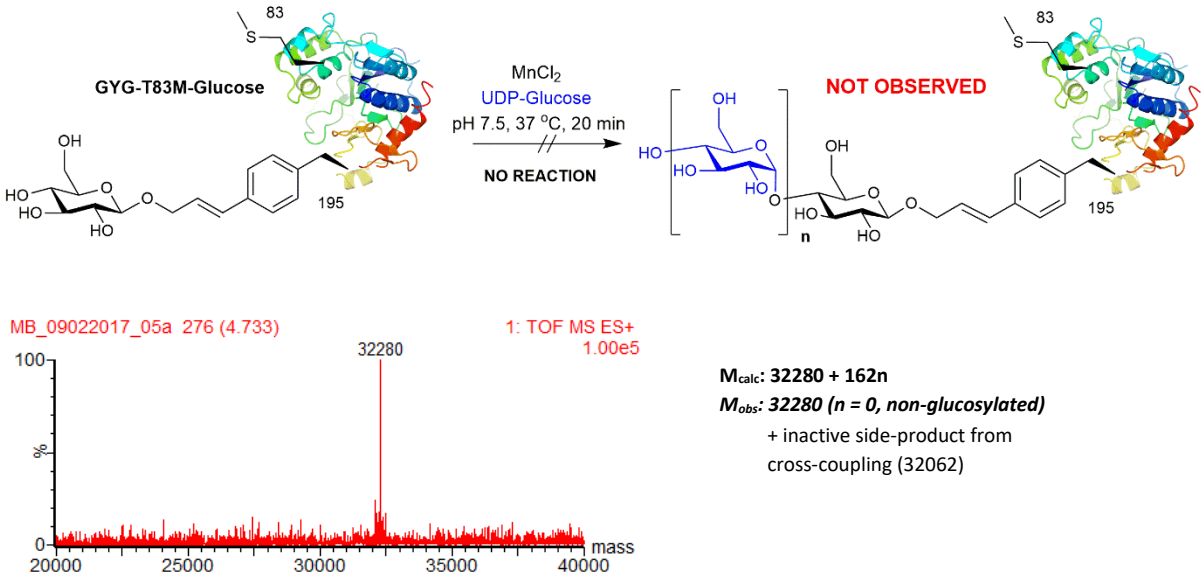


Figure 6.16 The GYG-T83M-Glucose construct exhibited no detectable autoglucosylation activity.

As such, its ability to be glucosylated in an inter-enzyme sense by a "donor" GYG species was determined. To this end, GYG-T83M-Glucose was combined with the previously expressed (chapter 2), his-tagged

GYG-Y195F mutant prior to glucosylation; the latter is capable of glucosylating other GYG monomers, but not itself.[6,27] As such, these two enzymes form a orthogonal pair, in which one can only *be* glucosylated but cannot glucosylate, and the other can only glucosylate but cannot *be* glucosylated. In this case, LC-MS analysis showed predominantly the non-glucosylated "donor" enzyme, GYG-Y195F, but additionally indicated partial glucosylation of the cross-coupled enzyme, with a mass corresponding to addition of 2 glucose residues being observed (**figure 6.17**, see experimental section 6.5.2.4). Thus, in a similar vein to GYG-T83M itself, GYG-T83M-Glucose is incapable of autoglucosylation, but can be glucosylated in an inter-enzyme sense. This suggests that the first kinetic phase of GYG autoglucosylation, from GYG-Glucose to GYG-Maltose (see **figures 6.11, 6.12**), may involve an inter-enzyme mechanism, although does not rule out other concomitant mechanisms.

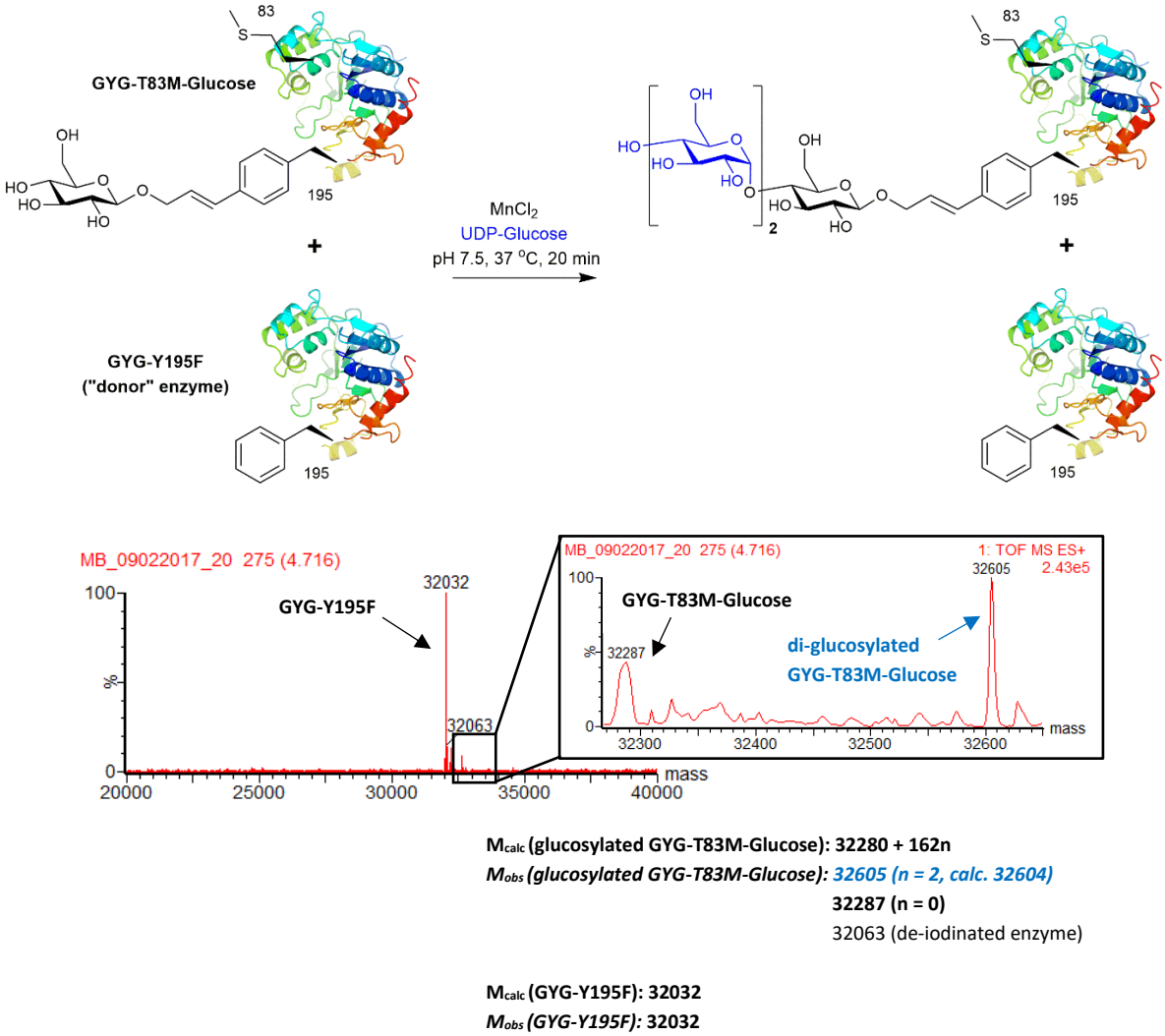


Figure 6.17 Inter-enzyme glucosylation of GYG-T83M-Glucose by the "donor" enzyme GYG-Y195F.

Intriguingly and somewhat unexpectedly, the disaccharide variant GYG-T83M-Maltose could catalyse its autoglucosylation, albeit to a limited extent; addition of just one glucose residue was observed in 20 minute (**figure 6.18**). As discussed above, whilst cross-coupling side-products are more abundant in this system, they should not play any role in observed activity, although further experiments to fully confirm this does represent a future direction. Whilst this highly surprising result unfortunately rendered any evaluation of inter-enzyme glucosylation for GYG-T83M-Maltose impossible – such glucosylation could not in any way be distinguished from this unexpected self-glucosylation – it is of clear significance in its own right, suggesting prior notions of a generalised inactivity of the GYG-T83M mutant is inaccurate.

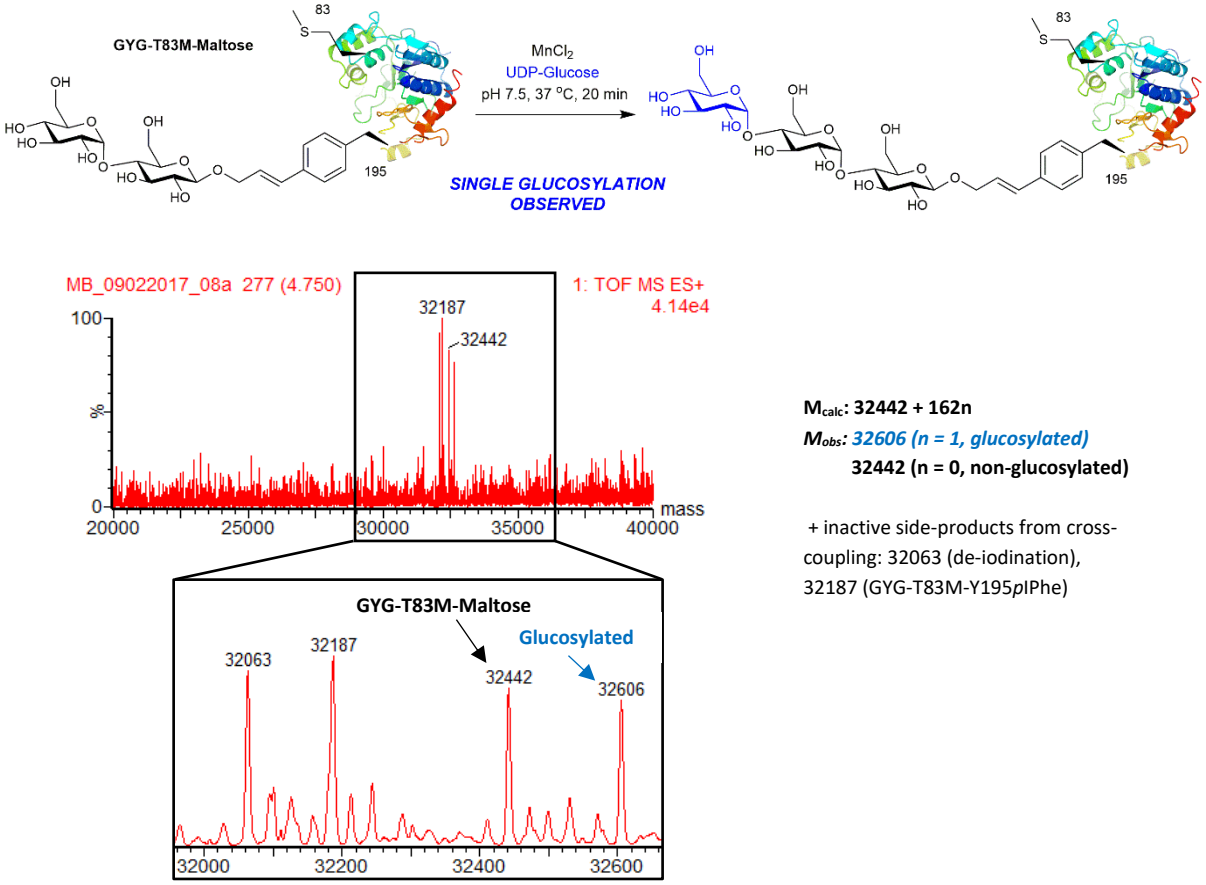


Figure 6.18 Autoglucosylation of GYG-T83M-Maltose.

An interesting observation for both systems is that, for both autoglucosylation (GYG-T83M-Maltose) and inter-enzyme glucosylation (GYG-T83M-Glucose), no species containing more than 3 glucose residues were observed. The rationale for this is at present unclear, and requires further investigation; it is possible that extension beyond a trisaccharide involves a distinct mechanism that is unavailable to the enzyme under the conditions employed. Nonetheless, this highlights the possibility that a triphasic profile for GYG

autoglucosylation (see **figure 6.12**) may still represent something of an oversimplification; the second phase, currently involving extension from 2 → 4/5 glucose residues, may in reality require further sub-division.

6.4 Conclusion and Outlook

The exploration of non-natural catalytic activities of enzymes has long been recognised as a valuable tool for dissecting their catalytic mechanisms. Numerous enzymatic studies exist involving the use of either non-natural substrates,[38–40] or mutant variants of the enzyme,[41–45] to enhance mechanistic understanding. Indeed, studies on mutant GYG species have proved invaluable in defining key residues involved in catalysis,[7,21,24] whilst previous work in this thesis concerning an autogalactosylation pathway highlighted the potential benefits of studying such non-natural catalytic activities.

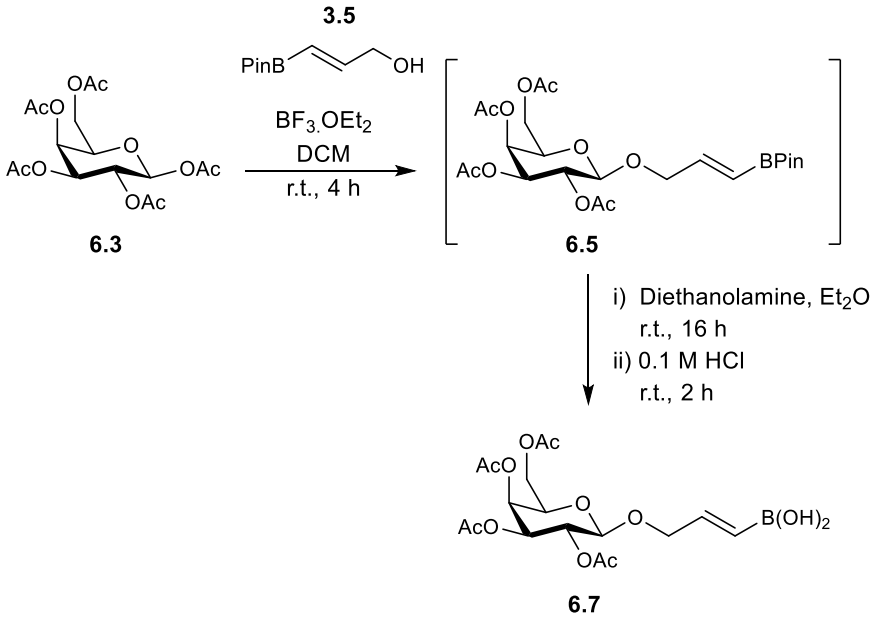
Previous chapters of this thesis have been focussed explicitly on the generation of mimics of natural glycoforms of GYG; the work outlined in this chapter instead demonstrates the application of our methodology to access non-natural variants inaccessible via other methods. Investigation of two enzymes in which the stereochemistry of the initial glucose acceptor was altered highlighted the plasticity of GYG in this regard, and enabled further distinction of catalytic phases with regards to tolerance of Gal-to-Gal transfer reactions. Subsequently, Suzuki-Miyaura cross-coupling enabled unprecedented access to mono- and di-glucosylated variants of a catalytically impaired mutant, GYG-T83M, studies on which suggested that this enzyme may not in fact be entirely inactive towards autoglucosylation, whilst further highlighting the evident distinctions between GYG-Glucose and GYG-Maltose variants. It is possible that such differences may arise from distinct mechanisms – one interpretation of the data for GYG-T83M-Glucose and GYG-T83M-Maltose autoglucosylation would be that the T83M mutation "shuts off" glucosylation mechanism(s) for the former (and also for the trisaccharide variant), but not for the latter – however our attempts to probe and compare inter-enzyme glucosylation for both species were hindered precisely by this unexpected autoglucosylation activity of GYG-T83M-Maltose.

As noted, the regioselectivity of the initial glucosylation of GYG-Galactose and GYG-Mannose is at present unknown (as is that of autogalactosylation); as discussed in chapter 4, NMR-based studies utilising ^{13}C-labelled glycosyl donors and/or acceptors could prove informative in this regard, although significant scale-up of procedures would be required. Furthermore, we envisage that crystallographic studies on GYG-T83M-Glucose and GYG-T83M-Maltose, and glucosylation products thereof, may prove informative with regards to dissecting mechanisms at play and distinctions between these.

6.5 Experimental Section

6.5.1 Investigating Unnatural GYG-Carbohydrate Acceptors

6.5.1.1 Chemical Synthesis



β-D-Galactose pentaacetate **6.3** (750 mg, 1.92 mmol) was dried under high vacuum for 2 h then dissolved in anhydrous DCM (4.50 ml). Alcohol **3.5** (387 mg, 2.10 mmol, see chapter 3 for synthesis) was dried via azeotropic removal of water with toluene (2 x) and was then added as a solution in 5.25 ml DCM. The reaction mixture was cooled to 0 °C. After 5 minutes, $BF_3.OEt_2$ (0.36 ml, 2.49 mmol) was added dropwise. The reaction was stirred for 4 hours at r.t after which time TLC (1:1 EtOAc: Pet. ether) indicated full consumption of starting material. The reaction was quenched through the addition of H_2O at 0 °C. The aqueous layer was extracted with DCM (3 x 30 ml), the combined organic layers dried ($MgSO_4$) and concentrated *in vacuo*. The resultant brown oil was dissolved in 60% EtOAc:Pet. Ether and filtered through a short plug of silica (ca. 15 g), which was then washed with ca. 3 x 30 ml of the solvent. Concentration *in vacuo* afforded crude product (930 mg) as a white foam (930 mg). ^{1}H NMR and LRMS confirmed the presence of pinacol ester **6.5**: **^{1}H NMR** (400 MHz, Chloroform-*d*) δ 6.59 (ddd, *J* = 18.1, 5.1, 4.2 Hz, 1H, PinBCH=C<u>H</u>), 5.70 (dt, *J* = 18.2, 1.8 Hz, 1H, PinBC<u>H</u>=CH), 5.41 (dd, *J* = 3.4, 1.2 Hz, 1H, <u>H4</u>), 5.27 (dd, *J* = 10.5, 7.9 Hz, 1H, <u>H2</u>), 5.04 (dd, *J* = 10.5, 3.4 Hz, 1H, <u>H3</u>), 4.54 (d, *J* = 8.0 Hz, 1H, <u>H1</u>), 4.44 (ddd, *J* = 14.7, 4.3, 1.9 Hz, 1H, PinBCH=CHC<u>H$_2$</u>), 4.24 (ddd, *J* = 14.7, 5.1, 1.7 Hz, 1H, PinBCH=CHC<u>H$_2$</u>), 4.21 – 4.14 (m, 2H, <u>H6a</u> + <u>H6b</u>), 3.91 (td, *J* = 6.7, 1.2 Hz, 1H, <u>H5</u>), 2.18 (s, 3H, O<u>Ac</u>), 2.11 (s, 3H, O<u>Ac</u>), 2.08 (s,

3H, OAc), 2.01 (s, 3H, OAc), 1.29 (s, 12H, BPin); **LRMS** *m/z* (ESI+) found: 537.2 (M+H, calculated 537.3).

This crude product (930 mg) was dissolved in anhydrous Et_2O (23 ml). Diethanolamine (0.26 ml, 2.70 mmol) was added as a liquid (m.p. 28 °C) and the reaction mixture stirred at r.t. After 10 minutes, formation of a white precipitate was noted. The suspension was stirred for a further 16 h. After this time, the precipitate was isolated from the mixture by filtration, and washed thoroughly with Et_2O to remove pinacol and other impurities carried forward from the glycosylation step. The presence and purity of the intermediate diethanolamine ester was confirmed by ^{1}H NMR. **^{1}H NMR** (400 MHz, DMSO-d_6) δ 6.73 (br s, 1H, NH), 5.72 (dt, *J* = 17.6, 5.3 Hz, 1H, BCH=CH), 5.56 (dt, *J* = 17.6, 1.3 Hz, 1H, BCH=CH), 5.24 (dd, *J* = 3.6, 1.1 Hz, 1H, H4), 5.15 (dd, *J* = 10.4, 3.6 Hz, 1H, H3), 4.94 (dd, *J* = 10.4, 8.0 Hz, 1H, H2), 4.67 (d, *J* = 8.0 Hz, 1H, H1), 4.19-4.13 (m, H5), 4.15 (ddd, *J* = 12.7, 5.3, 1.1 Hz, 1H, BCH=CHCH$_2$), 4.07 – 4.02 (m, 2H, H6a + H6b), 3.97 (ddd, *J* = 12.7, 5.7, 1.4 Hz, 1H, BCH=CHCH$_2$), 3.72 (td, *J* = 9.0, 5.4 Hz, 2H, BOCH$_2$), 3.60 (ddd, *J* = 9.7, 6.5, 3.8 Hz, 2H, BOCH$_2$), 2.98 (app dq, *J* = 12.8, 7.5 Hz, 2H, HNCH$_2$), 2.72 (app ddt, *J* = 11.7, 5.6, 3.0 Hz, 2H, HNCH$_2$), 2.11 (s, 3H, OAc), 2.02 (s, 3H, OAc), 2.01 (s, 3H, OAc), 1.91 (s, 3H, OAc). This species was resuspended in a biphasic mixture of Et_2O / 0.1M HCl (3:2 v/v, total volume 37.5 ml). The reaction was stirred at r.t. until complete dissolution of the solid was observed (2 h). The organic layer was separated and the aqueous extracted with Et_2O (3 x 40 ml). The combined organics were dried ($MgSO_4$), filtered and concentrated *in vacuo* to yield protected boronic acid **6.7** as a white solid (370 mg, 0.86 mmol, 45 % w.r.t. β-D-Galactose pentaacetate), as exclusively the β-anomer.

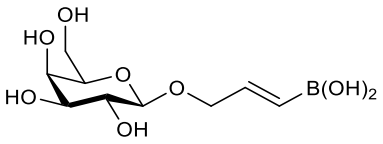
^{1}H NMR (400 MHz, DMSO-d_6) δ 6.61 (s, 1H, OH), 6.46 (dt, *J* = 18.1, 4.7 Hz, 1H, OCH$_2$CH=CH), 5.58 (dt, *J* = 18.1, 1.8 Hz, 1H, OCH$_2$CH=CH), 5.27 (dd, *J* = 3.6, 0.8 Hz, 1H, H4), 5.24 (dd, *J* = 10.4, 3.6 Hz, 1H, H3), 5.03 (dd, *J* = 10.4, 8.0 Hz, 1H, H2), 4.78 (d, *J* = 8.0 Hz, 1H, H1), 4.33 (ddd, *J* = 14.5, 4.7, 1.8 Hz, 1H, BCH=CHCH$_2$), 4.26 (dd, *J* = 6.6, 6.3, Hz, 1H, H5), 4.16 (ddd, *J* = 14.5, 4.7, 1.8 Hz, 1H, BCH=CHCH$_2$), 4.13 (dd, *J* = 11.0, 6.6 Hz, 1H, H6a), 4.09 (dd, *J* = 11.2, 6.3 Hz, 1H, H6b), 2.12-2.03 (4 x s, 12H, OAc)

LRMS *m/z* (ESI+) found: 455.1 (M+Na, calculated 455.1).

m.p. >300 °C (Et_2O), lit.[31] >300 °C. Phase transition to an amorphous solid was observed at 104-107 °C.

$[\alpha]_D$ = −20 (*c* = 1.0, H_2O), lit.[31] -18.

Data consistent with that previously reported.[31]



6.1

Protected boronic acid **6.7** (100 mg, 0.23 mmol) was dissolved in anhydrous MeOH (3.3 ml) and NaOMe (0.07 ml of a 25 wt % solution in MeOH, 0.34 mmol) was added dropwise. The reaction was stirred at room temperature for 3 h. TLC (10 % MeOH in EtOAc) and LC-MS indicated reaction completion. Pre-activated DOWEX®-H^+ was added until the pH was neutral, and the mixture stirred for a further 30 minutes. The resin was removed by filtration and washed with MeOH, and the eluate concentrated *in vacuo* to afford crude product. Further crude product was obtained from further washing of the filtered resin with H_2O and concentration *in vacuo* of the resultant filtrate. Combined residues were re-dissolved in H_2O to a concentration of ca. 30 mg/ml and passed through a 0.2 μm filter to remove small insoluble particles. The filtrate was concentrated *in vacuo* to yield boronic acid **6.1** as a white crystalline solid (38 mg, 0.14 mmol, 63 %), as exclusively the β-anomer.

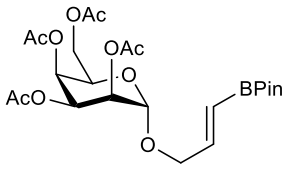
^{1}H NMR (400 MHz, Deuterium Oxide) δ 6.53 (dt, *J* = 18.3, 5.1 Hz, 1H, BCH=C<u>H</u>), 5.76 (dt, *J* = 18.3, 1.8 Hz, 1H, BC<u>H</u>=CH), 4.46 (ddd, *J* = 14.4, 4.9, 1.8 Hz, 1H, BCH=CHC<u>H$_2$</u>), 4.42 (d, *J* = 7.9 Hz, 1H, <u>H1</u>), 4.32 (ddd, *J* = 14.2, 5.5, 1.7 Hz, 1H, BCH=CHC<u>H$_2$</u>), 3.91 (dd, *J* = 3.4, 1.0 Hz, 1H, <u>H4</u>), 3.78 (dd, J = 11.7, 7.8 Hz, 1H, <u>H6a</u>), 3.73 (dd, *J* = 11.7, 4.4 Hz, 1H, <u>H6b</u>) 3.67 (ddd, *J* = 7.8, 4.4, 1.1 Hz, 1H, <u>H5</u>), 3.64 (dd, *J* = 10.0, 3.3 Hz, 1H, <u>H3</u>), 3.54 (dd, *J* = 10.0, 7.9 Hz, 1H, <u>H2</u>).

LRMS *m/z* (ESI+) found: 287.1 (M+Na, calculated 287.1)

m.p. >300 °C (H_2O), lit.[31] >300 °C

$[\alpha]_D$ = −13 (*c* = 0.5, H_2O), lit.[31] -12.

Data consistent with that previously reported.[31]



6.6

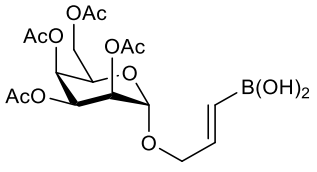
α-D-Mannose peracetate **6.4** (900 mg, 2.31 mmol) was dissolved in anhydrous DCM (14 ml) in a flame-dried flask under argon. Alcohol **3.5** (552 mg, 3.00 mmol) was added and the stirred mixture cooled to 0 °C. TMSOTf (667 mg, 0.54 ml, 3.00 mmol) was added dropwise, the reaction allowed to warm to r.t. and stirred for 4 h. After this time, H_2O was added (10 ml), the reaction diluted to 30 ml with DCM, and the organics washed with H_2O (20 ml). The aqueous layer was further extracted with DCM (2 x 30 ml). Combined organics were dried ($MgSO_4$), filtered and concentrated *in vacuo*. The residue was purified by flash column chromatography (SiO_2; 30-50-70% EtOAc in Pet. Ether) to yield pinacol ester **6.6** as a pale yellow oil (320 mg, 0.62 mmol, 27 %), as exclusively the α-anomer.

^{1}H NMR (400 MHz, Chloroform-*d*) δ 6.59 (dt, *J* = 18.2, 4.9 Hz, 1H, BCH=C<u>H</u>), 5.72 (dt, *J* = 18.1, 1.7 Hz, 1H, BC<u>H</u>=CH), 5.35 (dd, *J* = 9.9, 3.5 Hz, 1H, <u>H3</u>), 5.30 (t, *J* = 9.9 Hz, 1H, <u>H4</u>), 5.27 (dd, *J* = 3.5, 1.8 Hz, 1H, <u>H2</u>), 4.84 (d, *J* = 1.8 Hz, 1H, <u>H1</u>), 4.27 (ddd, *J* = 14.0, 4.7, 1.7 Hz, 1H, BCH=CHC<u>H$_2$</u>), 4.27 (dd, *J* = 12.2, 5.4 Hz, 1H, <u>H6a</u>), 4.09 (ddd, *J* = 13.8, 5.3, 1.6 Hz, 1H, BCH=CHC<u>H$_2$</u>), 4.08 (dd, *J* = 12.3, 2.7 Hz, 1H, <u>H6b</u>), 3.98 (ddd, *J* = 9.9, 5.4, 2.4 Hz, 1H, <u>H5</u>), 2.14 (s, 3H, -O<u>Ac</u>), 2.09 (s, 3H, -O<u>Ac</u>), 2.03 (s, 3H, -O<u>Ac</u>), 1.98 (s, 3H, -O<u>Ac</u>), 1.27 (s, 12H, B<u>Pin</u>)

LRMS *m/z* (ESI+) found: 537.2 (M+Na, calculated 537.2)

[α]$_D$ = +35 (c = 1.3); lit.[31] +26.

Data consistent with that previously reported.[31]



6.8

Pinacol ester **6.8** (310 mg, 0.60 mmol) was suspended in anhydrous Et_2O (10 ml). Diethanolamine (94 mg, 0.90 mmol) was added as a liquid (m.p. 28 °C), and the mixture stirred under argon for 16 h at room temperature. The resultant white precipitate was collected by gravity filtration and washed thoroughly with

Et_2O. ^1H NMR confirmed the presence of the intermediate diethanolamine ester: **^{1}H NMR** (400 MHz, DMSO-d_6) δ 6.79 (br s, 1H, NH), 5.80 (dt, *J* = 17.6, 5.4 Hz, 1H, BCH=CH), 5.64 (d, *J* = 17.5 Hz, 1H, BCH=CH), 5.13 – 5.10 (m, 2H, H3 + H4), 5.08 (dd, *J* = 3.4, 1.8 Hz, 1H, H2), 4.88 (d, *J* = 1.8 Hz, 1H, H1), 4.16 (dd, *J* = 12.3, 5.4 Hz, 1H, H6a), 4.11 (ddd, *J* = 12.3, 5.4, 1.3 Hz, 1H, BCH=CHCH_2), 4.05 (dd, *J* = 12.2, 2.6 Hz, 1H, H6b), 3.97 – 3.87 (m, 2H, BCH=CHCH_2 + H5), 3.73 (td, *J* = 9.1, 5.4 Hz, 2H, -BOCH_2), 3.61 (ddd, *J* = 9.7, 6.5, 3.7 Hz, 2H, -BOCH_2), 3.00 (ddt, *J* = 11.5, 8.7, 6.8 Hz, 2H, -NOCH_2), 2.73 (ddt, *J* = 11.7, 5.6, 3.0 Hz, 2H, -NOCH_2), 2.12 (s, 3H, -OAc), 2.04 (s, 3H, -OAc), 2.03 (s, 3H, -OAc), 1.95 (s, 3H, -OAc).This white solid (158 mg, 0.31 mmol) was suspended a biphasic mixture of Et_2O: 0.1 M HCl (3:2 v/v, total volume 18.75 ml). The mixture was stirred for 2 h at room temperature, during which time complete dissolution was observed. The organic layer was separated and the aqueous extracted with Et_2O (2 x 40 ml). Combined organics were dried ($MgSO_4$), filtered and concentrated *in vacuo* to yield boronic acid **6.8** as a white crystalline solid (129 mg, 0.30 mmol, 50 %), as exclusively the α-anomer.

^{1}H NMR (500 MHz, Acetone-d_6) δ 6.56 (dt, *J* = 18.0, 5.1 Hz, 1H, BCH=CH), 5.71 (dt, *J* = 18.1, 1.7 Hz, 1H, BCH=CH), 5.30 – 5.23 (m, 2H, H3 + H4), 5.22 (dd, *J* = 3.0, 1.8 Hz, 1H, H2), 4.89 (d, *J* = 1.7 Hz, 1H, H1), 4.28 (ddd, *J* = 13.9, 4.7, 1.7 Hz, 1H, BCH=CHCH_2), 4.21 (dd, *J* = 12.2, 5.3 Hz, 1H, H6a), 4.12 (ddd, *J* = 13.9, 5.4, 1.7 Hz, 1H, BCH=CHCH_2), 4.11 (dd, *J* = 12.2, 2.4 Hz, 1H, H6b), 4.01 (ddd, *J* = 8.2, 5.3, 2.5 Hz, 1H, H5), 2.10 (s, 3H, OAc), 2.03 (s, 3H, OAc), 2.02 (s, 3H, OAc), 1.93 (s, 3H, OAc).

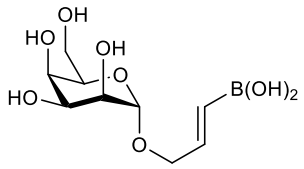
^{13}C NMR (126 MHz, Acetone-d_6) δ 170.4, 169.9, 169.8, 169.7 (all OAc C=O), 144.1 (BCH=CH), 126.1 (weak, BCH=CH), 96.6 (C1), 69.3 (C3), 69.2 (C2), 69.1 (BCH=CHCH_2), 68.6 (C5), 65.9 (C4), 62.3 (C6), 19.9, 19.8, 19.8 (all OAc CH_3).

HRMS *m/z* (ESI+) found: 455.1333 (M+Na, calculated 455.1331)

IR (film) $\tilde{\nu}$ = 3479 (O-H), 1748 (C=O), 1644 (C=C) cm^{-1}

m.p. >300 °C (Et_2O)

$[\alpha]_D$ = +49 (c = 1.0, $CHCl_3$)



6.2

Boronic acid **6.8** (60 mg, 0.14 mmol) was dissolved in anhydrous MeOH (1.85 ml). NaOMe (7.5 mg, 0.032 ml of a 25 wt % solution in MeOH, 0.14 mmol) was added dropwise and the mixture stirred under argon for 2 h, after which time TLC (10% MeOH in EtOAc) indicated complete consumption of starting material. Pre-activated DOWEX-H^+ was added portionwise until the pH was ca. 5, and the suspension stirred for a further 20 minutes, before removal of the resin by filtration. The filtrate was concentrated *in vacuo* and further azeotroped with DCM (4 x 5 ml) to give an off-white solid (34 mg). This was dissolved in H_2O (1.1 ml, 30 mg/ml) and insoluble material was removed via syringe filtration (0.2 μm filter). Concentration of the filtrate *in vacuo* afforded **6.1** as a white crystalline solid (22 mg, 0.083 mmol, 60 %), as exclusively the α-anomer.

^{1}H NMR (500 MHz, Deuterium Oxide) δ 6.46 (dt, *J* = 18.3, 4.9 Hz, 1H, BCH=CH), 5.65 (dt, *J* = 18.3, 1.6 Hz, 1H, BCH=CH), 4.82 (d, *J* = 1.7 Hz, 1H, H1), 4.23 (ddd, *J* = 14.3, 4.6, 1.7 Hz, 1H, BCH=CHCH$_2$), 4.09 (ddd, *J* = 14.4, 5.2, 1.6 Hz, 1H, BCH=CHCH$_2$), 3.90 (dd, *J* = 3.4, 1.7 Hz, 1H, H2), 3.80 (dd, *J* = 12.2, 1.4 Hz, 1H, H6), 3.77 – 3.72 (m, 1H, H3), 3.67 (dd, *J* = 11.9, 4.7 Hz, 1H, H6), 3.59 – 3.54 (m, 2H, H4 + H5).

LRMS *m/z* (ESI+) found: 287.1 (M+Na, calculated 287.1)

m.p. >300 °C (H_2O); lit.[31] >300 °C

$[\alpha]_D$ = +64 (c = 0.7, H_2O); lit.[31] +20 (see note below).

The $[\alpha]_D$ *value obtained here is notably higher than that reported previously. As such, this measurement was repeated on two independent samples, with 2 runs of 8 values for each sample, and the reported value is a mean average of the means for each sample. The values for each sample were highly consistent (+66, +62). Furthermore, we note that reported* $[\alpha]_D$ *values for other alkyl a-D-mannopyranosides are typically similar or higher, e.g. methyl a-D-mannopyranoside = +79 (H_2O),*[46] *octyl a-D-mannopyranoside = +56 (H_2O),*[47] *and the structurally highly related allyl a-D-mannopyranoside = +85 (H_2O).*[48] *Other spectroscopic data was identical to that previously reported.*[31]

6.5.1.2 Generation and Catalytic Activity of GYG-Galactose and GYG-Mannose

a) General Protocol for Autoglucosylation and Autogalactosylation

Cross-coupling step:

A stock solution of boronic acid (either galactose boronic acid **6.1** or mannose boronic acid **6.2**) was prepared either in H_2O (**6.2**) or in 15.7 mg/ml aqueous Na_2HPO_4 (**6.1**). Non-his-tagged GYG-Y195*p*IPhe in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was combined with boronic acid added from stock solution, and $Pd(L2)_2(OAc)_2$ catalyst (0.6 mM final concentration, added from 10 mM stock solution). The reaction mixture was briefly vortexed, then incubated at 37 °C with shaking (600 rpm). After 3 hours, DTT was added to a final concentration of 7.5 mM (from a 115 mM solution in H_2O), the mixture shaken for a further 40 minutes, then buffer exchanged into pH 7.5 25 mM Na-HEPES by PD SpinTrap G-25. 100 µl of the sample was further desalted by dialysis against the same buffer (10000 MWCO, 4 °C, 2 h then 2 h then overnight, 3 x 300 ml buffer), using a Slide-A-Lyzer™ MINI dialysis device.

Enzymatic activity assay:

Protein thus obtained was adjusted to a concentration of 0.2 mg/ml (Vivaspin 500, 10000 MWCO (Sartorius Stedem) were used for concentration if required). Protein was incubated with $MnCl_2$ and then either UDP-Glucose disodium salt (autoglucosylation) or UDP-Galactose disodium salt (autogalactosylation), all at final concentrations of 1 mM (added from freshly prepared 10 mM stock solutions in H_2O). The protein was incubated at 37 °C with shaking (500 rpm). After the desired time (indicated in the following sections), aliquots were quenched through the addition of Na-EDTA to a final concentration of 20 mM, and immediately placed on ice.

Quenched aliquots were analysed by LC-MS in order to determine average numbers of glucoses attached and the proportion of starting cross-coupled enzyme remaining at each time point (the latter serves as a means of monitoring the rate of the initial glucosylation step). For further details of procedures for such LC-MS analysis, please see chapter 4 (section 4.5.3).

For parts (b) and (c) (glucosylation) data points on graphs shown represent mean averages ± s.d of three independent replicates of the above protocol. All replicate LC-MS data for *glucosylation* assays was analysed according to the protocol detailed in chapter 4, section 4.5.3, in order to generate the "overall glucosylation profile" and "rate of initial step" graphs shown in the main text and reproduced here.

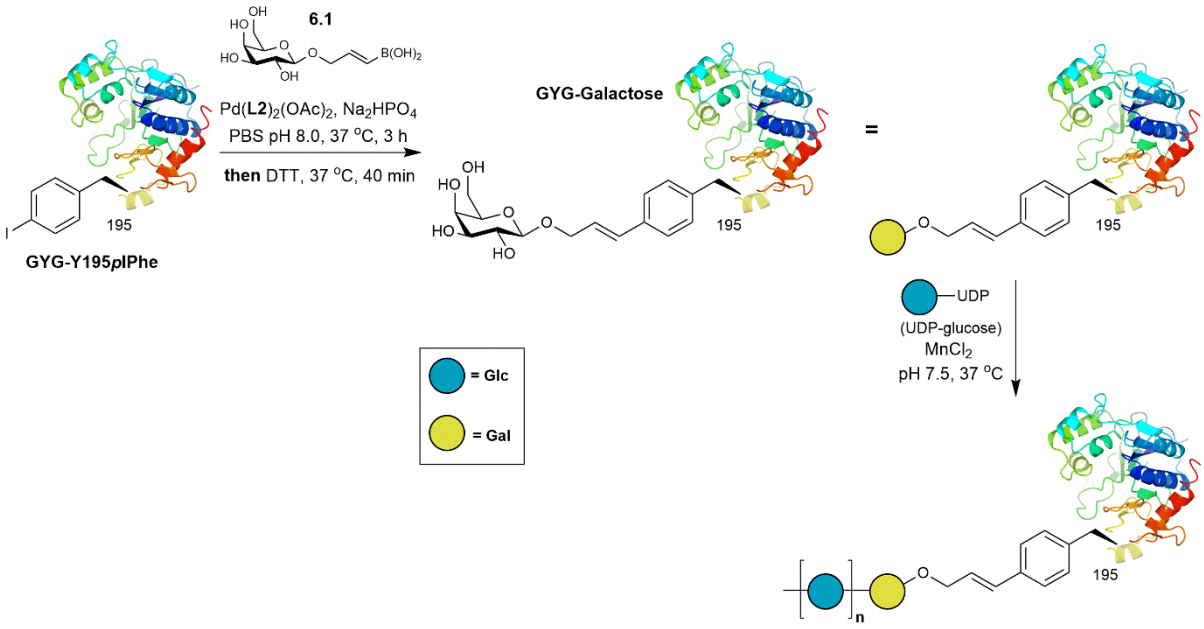
Note regarding formatting of data in the following sections

For sections (b) and (c), which represent autoglucosylation assays according to the above protocol, the data is presented as follows:

1) The experimental protocol is outlined (including representative MS of cross-coupling sample).
2) Representative LC-MS data from kinetic assays are shown.
3) Graphical data (that shown in the main text), determined from LC-MS data across all replicates and timepoints via the methods outlined in chapter 4, section 4.5.3, is shown (note that this also shown in the main text, as indicated by references to appropriate figures here).
4) A detailed, "close-up" view of a representative LC-MS spectrum from the glucosylation assay is provided, to illustrate that the non-glucosylated side-products do indeed correspond to those generated in cross-coupling.

For sections (d) and (e), which represent autogalactosylation assays according to the above protocol, the data is simply presented as the experimental protocol followed by LC-MS spectrum after galactosylation assay. This is because only one timepoint (2 h) was taken, since autogalactosylation did not occur to significant enough extents for any timecourse analyses to be undertaken.

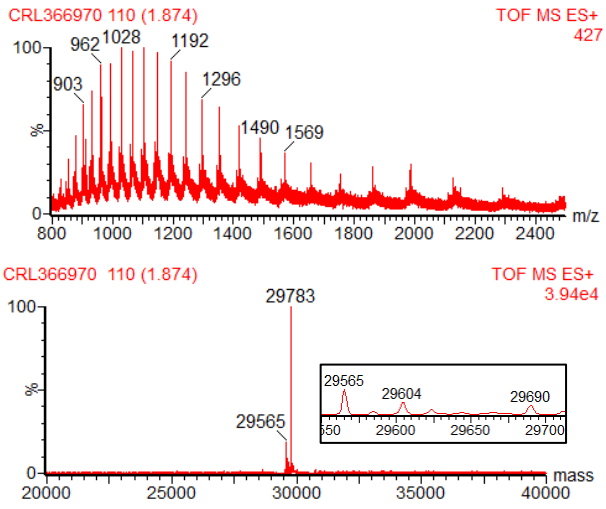
b) Generation and Autoglucosylation Analysis of GYG-Galactose



Carried out according to general protocol (section 6.5.1.2, (a)), using the following reagent amounts:

- GYG-Y195*p*IPhe: 50 µl (0.85 mg/ml), 1.4 nmol
- Boronic acid **6.1**: 48 µl of a 27 mg/ml stock solution prepared in 15.7 mg/ml Na_2HPO_4, 4900 nmol, 3400 eq.
- $Pd(L2)_2(OAc)_2$: 6.3 µl of 10 mM stock solution, 63 nmol, 44 eq., 0.6 mM final concentration.

LC-MS after cross-coupling showed 74 % conversion:



M_{calc}: 29784

M_{obs}: 29783 (74 %)

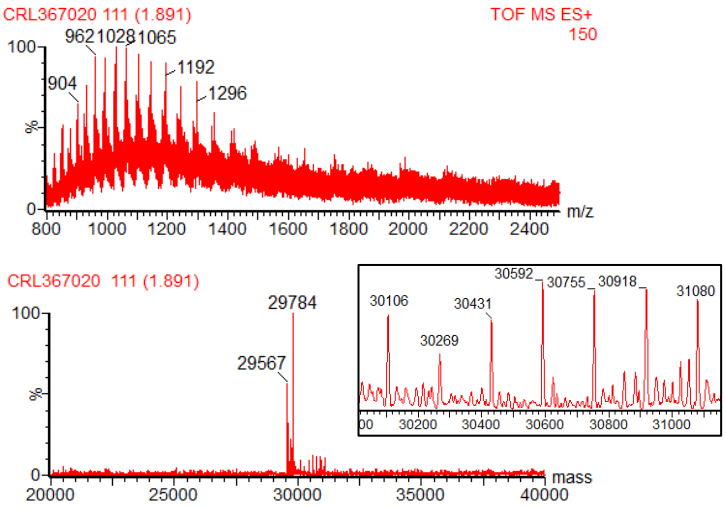
29566 (de-iodination)

29690 (GYG-Y195*p*IPhe)

29607 (see **figure 6.7** and chapter 3, pp. 172-3)

UDP-glucose was utilised as the donor sugar nucleotide during enzymatic activity assay. Aliquots were quenched and analysed after the following times of autoglucosylation assay: 2.5 min, 5 min, 7.5 min, 10 min, 12.5 min, 15 min, 20 min. Representative LC-MS spectra and overall graphical data are shown below. Cross-coupling conversions of up to 74 % were attained (see t = 0 spectrum below).

Representative LC-MS spectra from glucosylation assays: t = 5 min

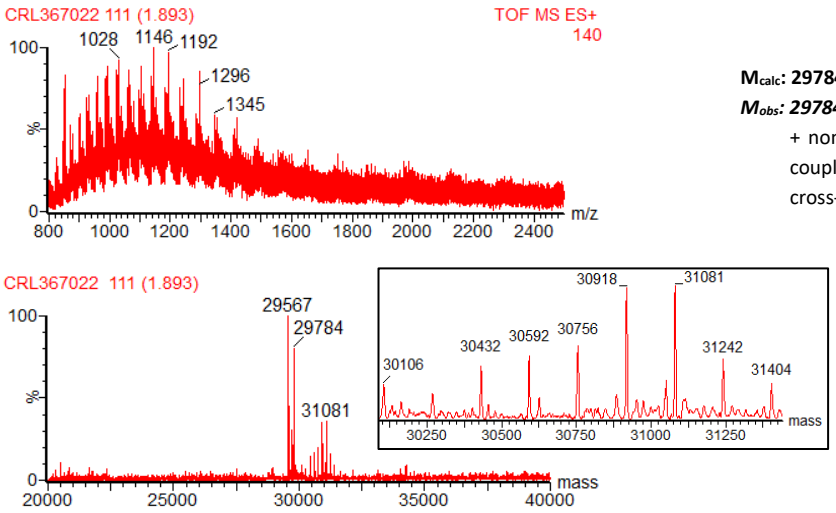


M_{calc}: 29784 + 162n

M_{obs}: 29784 + 162n (n = 0, 2-8)

+ non-glucosylated side-products from cross-coupling (29567, 29604), see "LC-MS after cross-coupling" above

t = 10 min

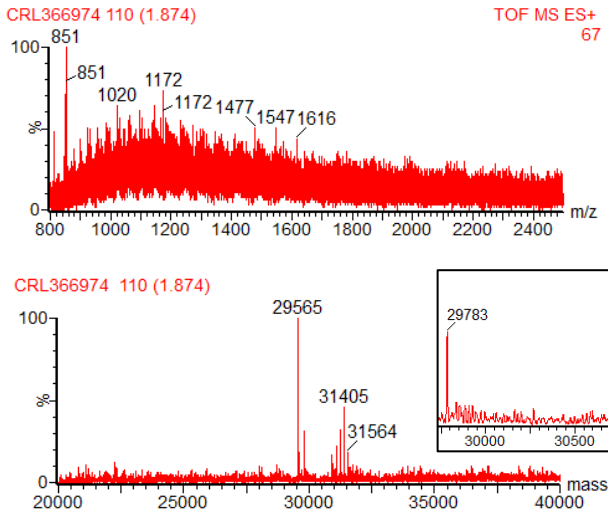


M_{calc}: 29784 + 162n

M_{obs}: 29784 + 162n (n = 0, 2-10)

+ non-glucosylated side-products from cross-coupling (29567, 29604), see "LC-MS after cross-coupling" above

t = 20 min



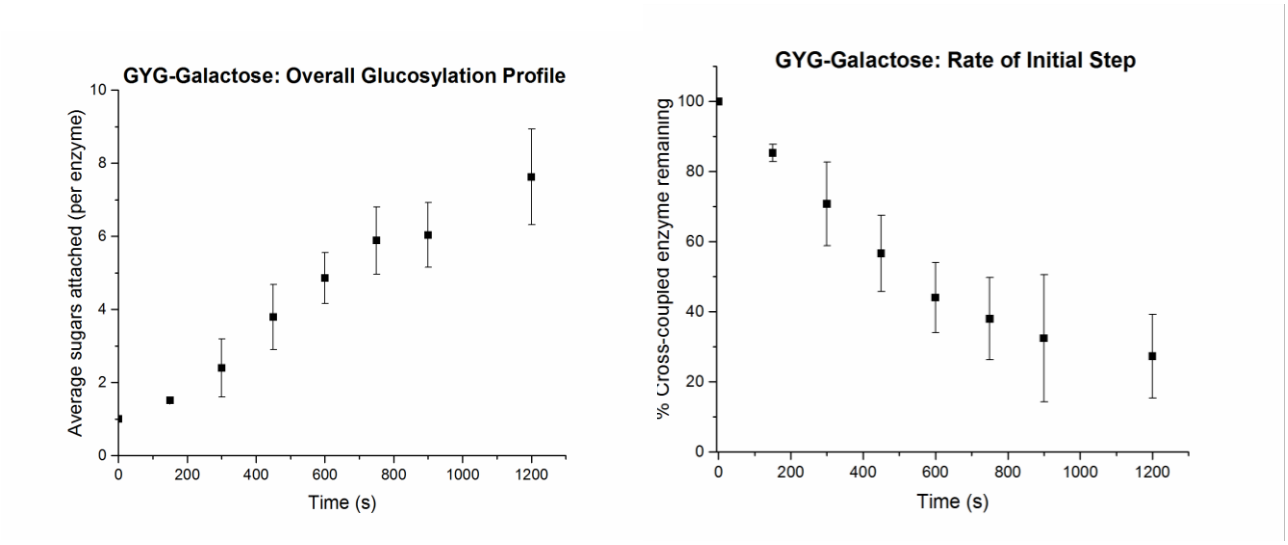
M_{calc}: 29784 + 162n

M_{obs}: 29783 + 162n (n = 0, 7-12)

+ non-glucosylated side-products from cross-coupling (29565, 29604), see "LC-MS after cross-coupling" above

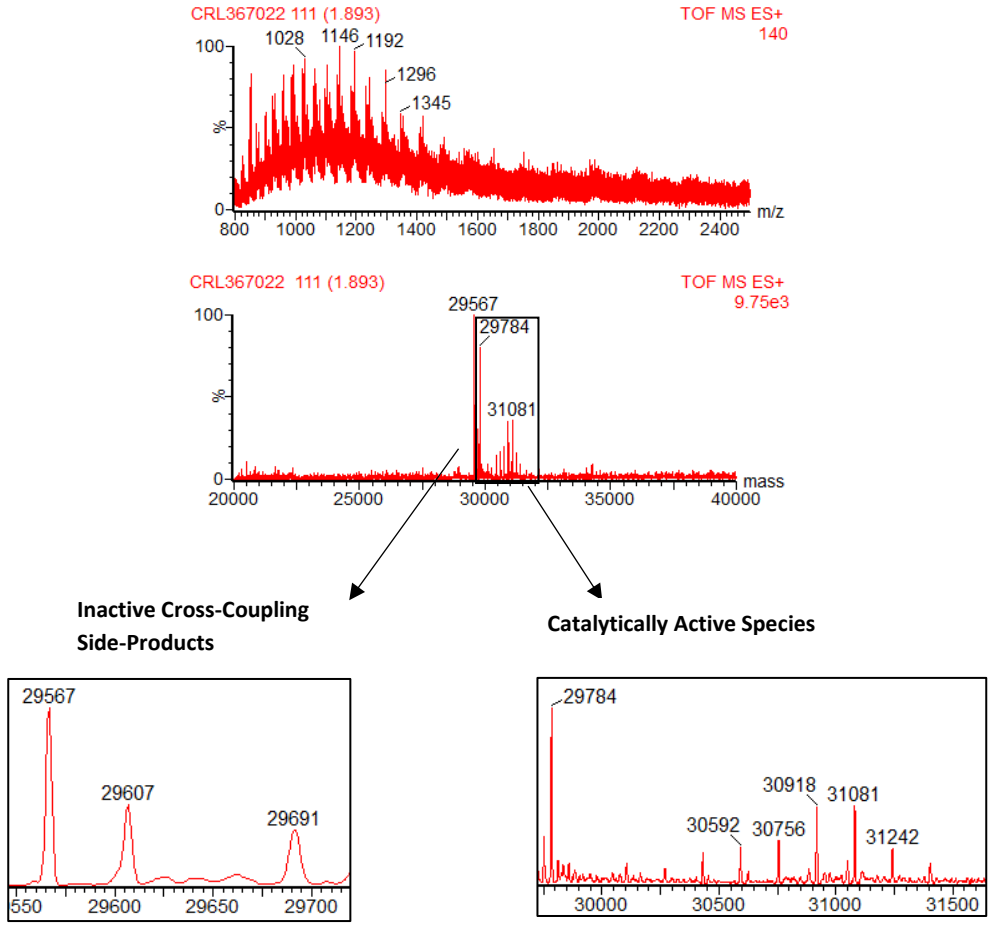
Graphical data from LC-MS replicates across all time points (figure 6.9):

This was generated from LC-MS spectra from all time points and replicates according to the analytical protocol detailed in chapter 4, section 4.5.3.



Detailed view of a representative LC-MS spectrum after glucosylation:

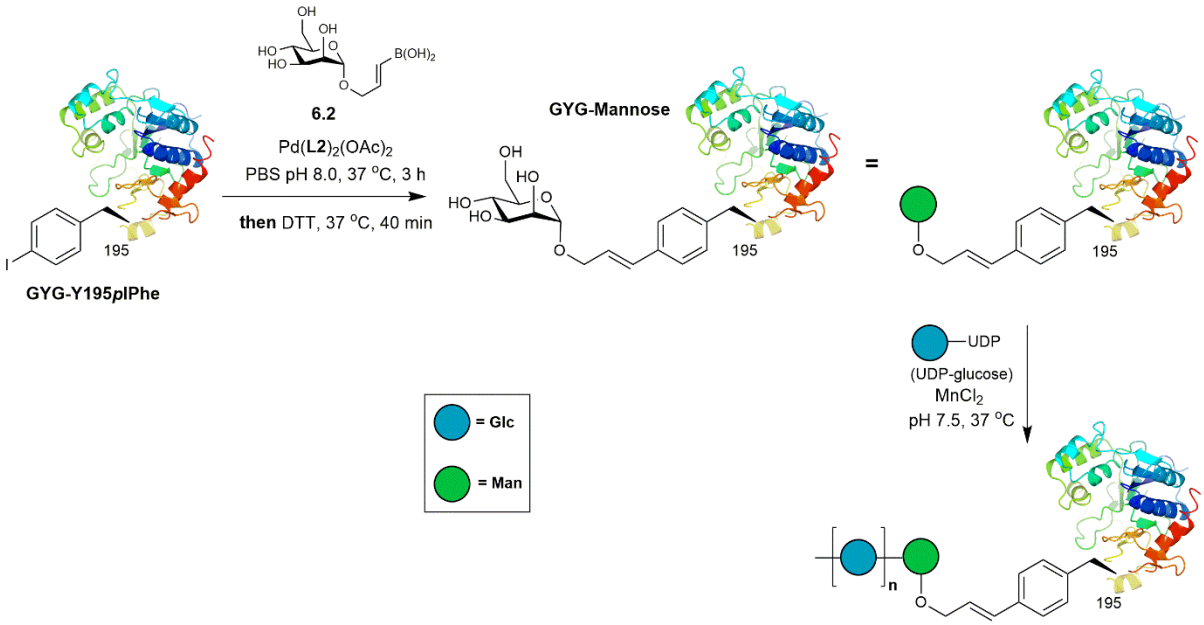
The inability of cross-coupling side-products to be glucosylated was demonstrated in chapter 3, and the side-products in this chapter are seemingly identical; the same inactivity was observed. Below is a detailed view of one of the spectra above (t = 10 min), illustrating that the catalytically inactive species do indeed represent these cross-coupling side-products.



29566: de-iodinated
29691: GYG-Y195*p*IPhe
29606: see fig. 6.7 and chapter 3, pp 172-3

29784 = GYG-Galactose
30430-31242 = glucosylated GYG-Galactose
(= 29783 + 162n, n = 4-9)

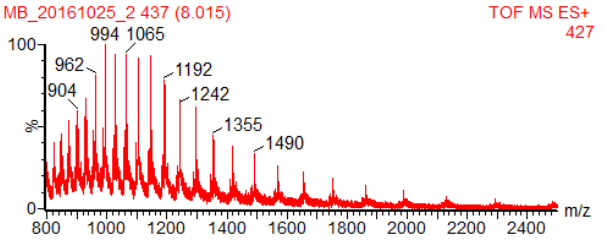
c) Generation and Autoglucosylation of GYG-Mannose

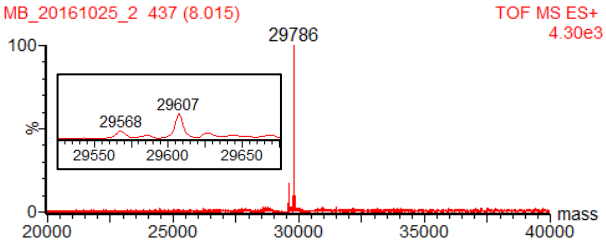


Carried out according to general protocol (section 6.5.1.2, (a)) using the following reagent amounts:

- GYG-Y195*p*IPhe: 50 µl (0.85 mg/ml), 1.4 nmol
- Boronic acid **6.2**: 48 µl of a 27 mg/ml stock solution prepared in H_2O, 4900 nmol, 3400 eq.
- $Pd(L2)_2(OAc)_2$: 6.3 µl of 10 mM stock solution, 63 nmol, 44 eq., 0.6 mM final concentration.

LC-MS after cross-coupling showed 82 % conversion:

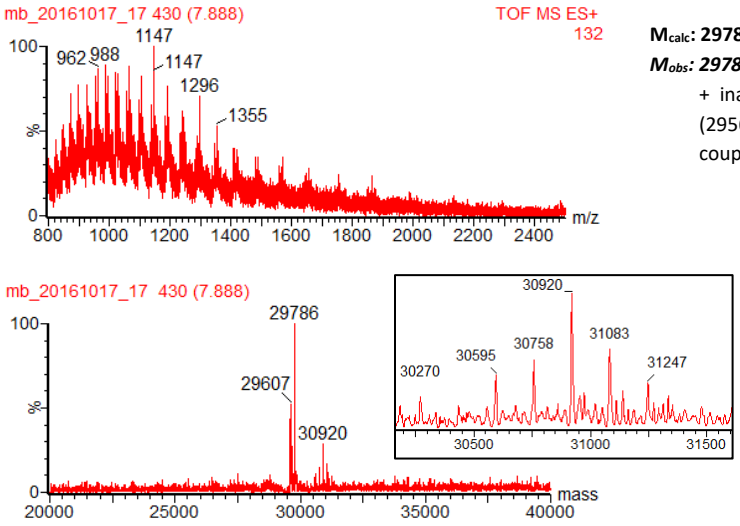




M_{calc}: 29784
M_{obs}: 29786 (82 %)
29568 (de-iodination)
29607 (see **figure 6.7** and chapter 3, pp. **172-3**)

UDP-glucose was utilised as the donor sugar nucleotide during enzymatic activity assay. Aliquots were quenched and analysed after the following times of autoglucosylation assay: 2.5 min, 5 min, 7.5 min, 10 min, 12.5 min, 15 min, 20 min. Representative LC-MS spectra and overall graphical data are shown below. Cross-coupling conversions of up to 82 % were attained (see t = 0 spectrum below).

Representative LC-MS spectra from glucosylation assays: t = 5 min

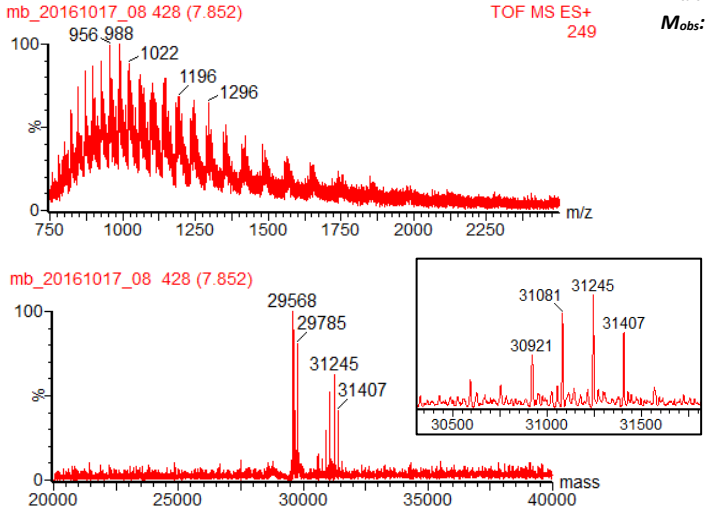


M_{calc}: 29784 + 162n

M_{obs}: 29786 + 162n (n = 0, 3-9)

+ inactive side-products from cross-coupling (29568, 29607, 29693), see "LC-MS after cross-coupling" above

t = 10 min

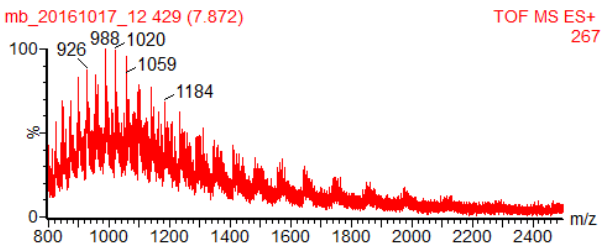


M_{calc}: 29784 + 162n

M_{obs}: 29785 + 162n (n = 0, 5-10)

+ inactive side-products from cross-coupling (29568, 29607, 29693), see "LC-MS after cross-coupling" above

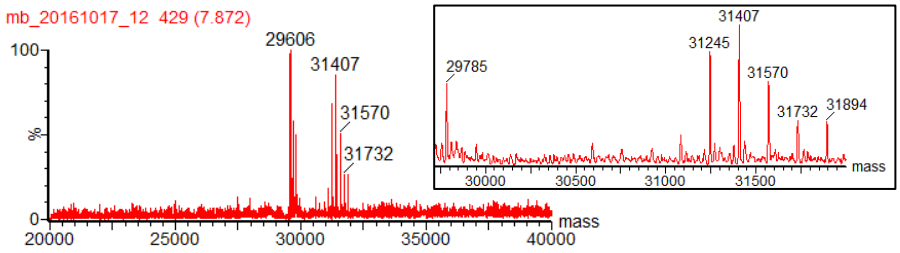
t = 20 min



M_{calc}: 29784 + 162n

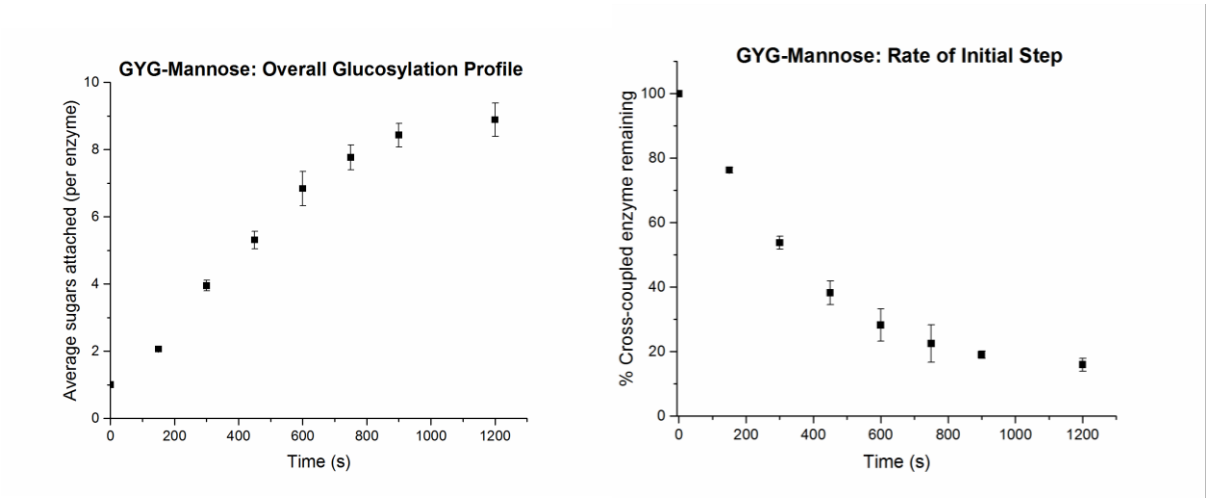
M_{obs}: 29785 + 162n (n = 0, 5-13)

+ inactive side-products from cross-coupling (29569, 29606, 29693), see "LC-MS after cross-coupling" above



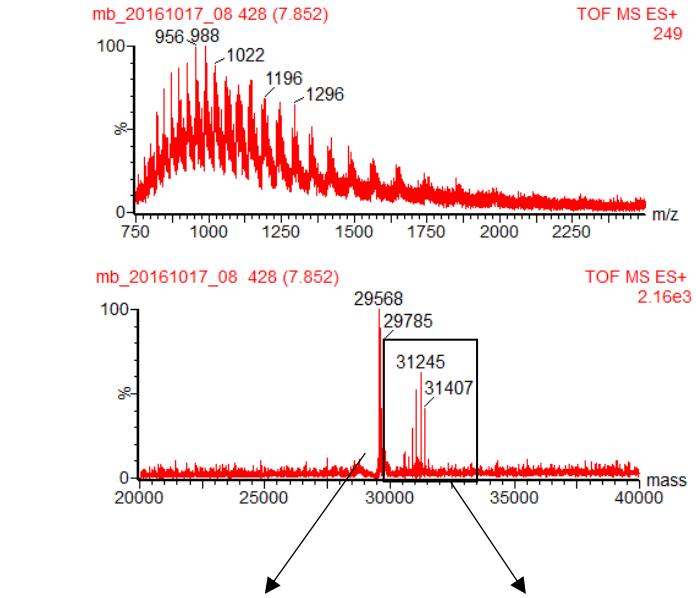
Graphical data from LC-MS replicates across all time points (figure 6.9):

This was generated from LC-MS spectra from all time points and replicates according to the analytical protocol detailed in chapter 4, section 4.5.3.

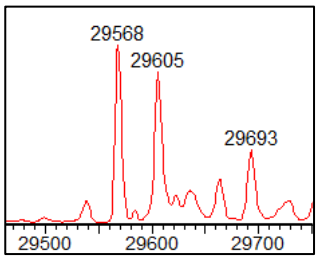


Detailed view of a representative LC-MS spectrum after glucosylation:

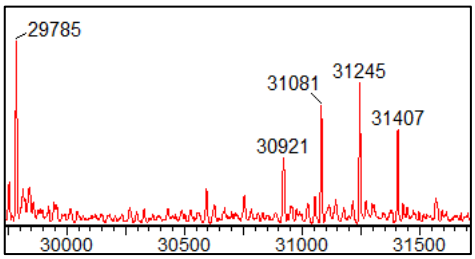
The inability of cross-coupling side-products to be glucosylated was demonstrated in chapter 3, and the side-products in this chapter are identical; the same inactivity was observed. Below is a detailed view of one of the spectra above (t = 10 min), illustrating that the catalytically inactive species do indeed represent these cross-coupling side-products.



Inactive Cross-Coupling Side-Products



Catalytically Active Species

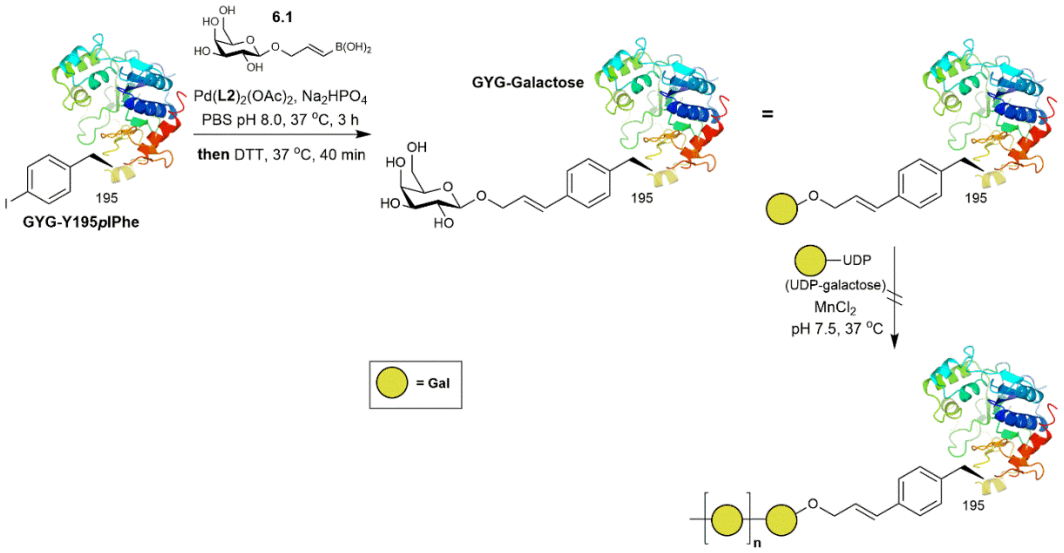


29568: de-iodinated
29692: GYG-Y195*p*IPhe*
29605: see fig. 6.7 and chapter 3 pp. 172-3

29785 = GYG-Mannose
30921-31407 = glucosylated GYG-Mannose
(= 29785 + 162n, n = 7-10)

**Note: the very low abundance of the GYG-Y195*pIPhe *relative to cross-coupled enzyme means that it is regularly not seen after cross-coupling, being indistinguishable from background noise, but may become apparent after glucosylation (since the cross-coupled enzyme peak splits into multiple glycoform peaks), as evident above.*

d) Generation and Autogalactosylation of GYG-Galactose

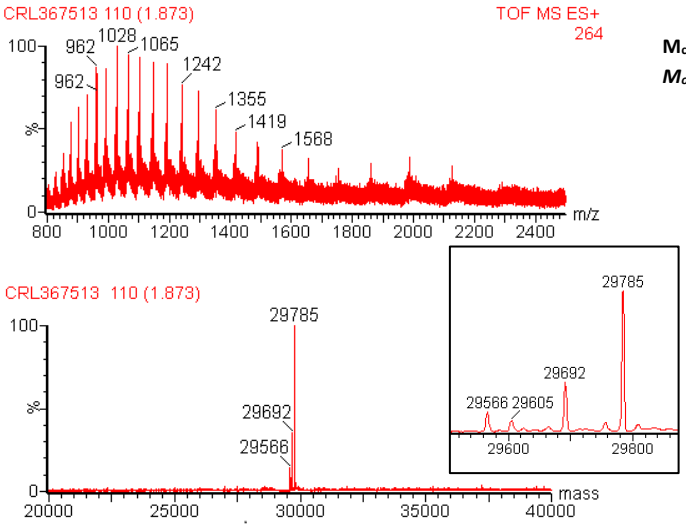


Carried out according to general protocol (section 6.5.1.2, (a)) using the following reagent amounts:

- GYG-Y195*p*IPhe: 50 µl (0.85 mg/ml), 1.4 nmol
- Boronic acid **6.1**: 48 µl of a 27 mg/ml stock solution prepared in 15.7 mg/ml Na_2HPO_4, 4900 nmol, 3400 eq.
- $Pd(L2)_2(OAc)_2$: 6.3 µl of 10 mM stock solution, 63 nmol, 44 eq., 0.6 mM final concentration.
-

UDP-galactose was utilised as the donor sugar nucleotide during enzymatic activity assay. This was an endpoint assay, thus only one aliquot was quenched after 2 hours reaction time. No significant autogalactosylation was observed.

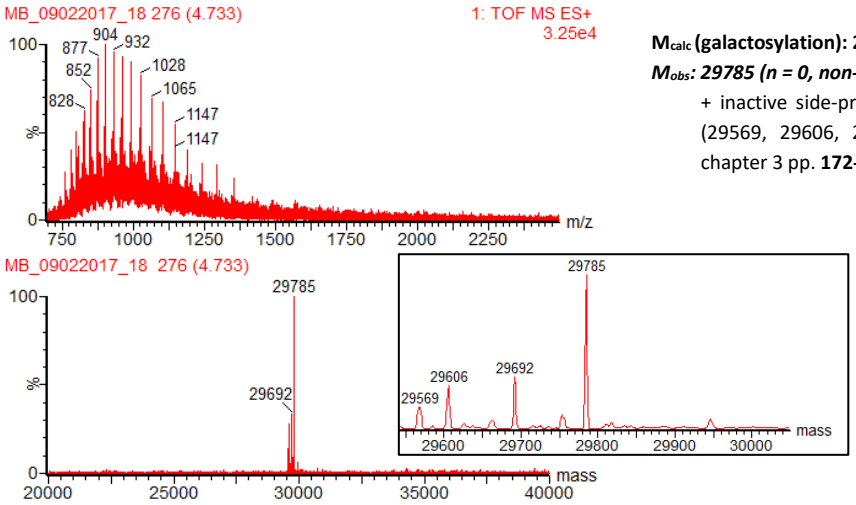
LC-MS spectrum after autogalactosylation



M_{calc} (galactosylation): 29784 + 162n
M_{obs}: 29785 (n = 0, non-galactosylated)
+ inactive side-products from cross-coupling, (29566, 29605, 29692, see **figure 6.7** and chapter 3 pp. **172-3**)

Given the crucial nature of this result in defining the three catalytic phases via their respective tolerances to Gal-to-Gal transfer (see p. 371), this experiment was independently repeated. The same result was obtained:

LC-MS spectrum after autogalactosylation (run 2):

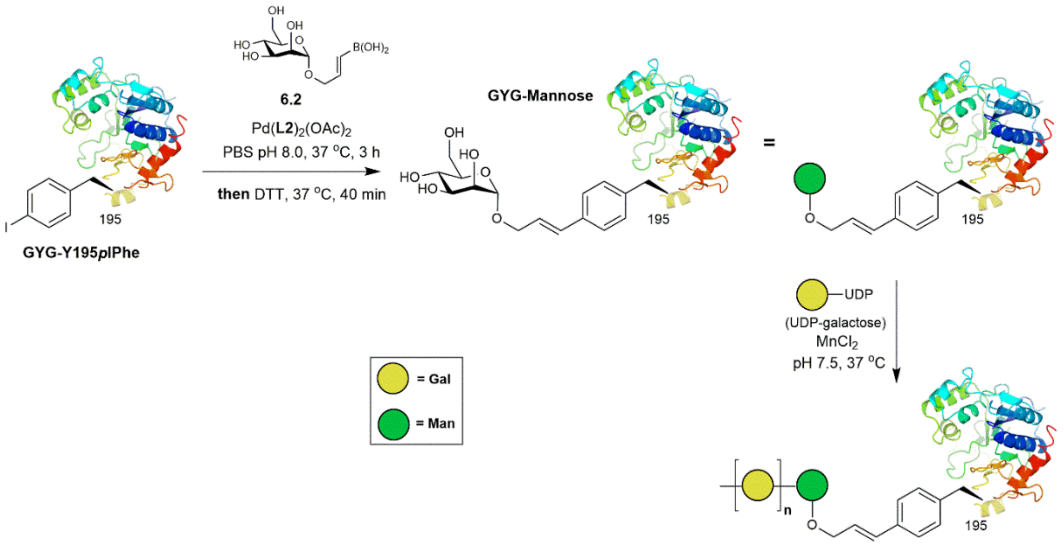


M_{calc} (galactosylation): 29784 + 162n

M_{obs}: 29785 (n = 0, non-galactosylated)

+ inactive side-products from cross-coupling, (29569, 29606, 29692, see **figure 6.7** and chapter 3 pp. **172-3**)

e) Generation and Autogalactosylation of GYG-Mannose

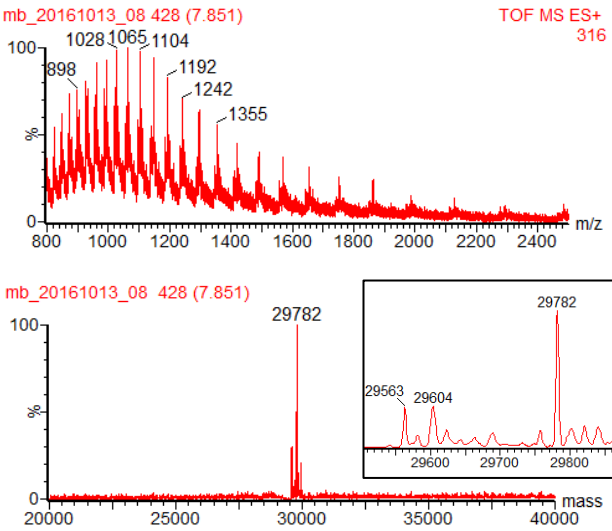


Carried out according to general protocol (section 6.5.1.2, (a)) using the following reagent amounts:

- GYG-Y195*p*IPhe: 50 µl (0.85 mg/ml), 1.4 nmol
- Boronic acid **6.2**: 48 µl of a 27 mg/ml stock solution prepared in H_2O, 4900 nmol, 3400 eq.
- $Pd(L2)_2(OAc)_2$: 6.3 µl of 10 mM stock solution, 63 nmol, 44 eq., 0.6 mM final concentration.

UDP-galactose was utilised as the donor sugar nucleotide during enzymatic activity assay. This was an endpoint assay, thus only one aliquot was quenched after 2 hours reaction time. Very limited autogalactosylation was seen (15 % of cross-coupled enzyme reacted). This was carried out in parallel to an aforementioned GYG-Mannose glucosylation assay, in order to rule out any generic erroneous enzyme inactivation as responsible for the limited galactosylation.

LC-MS spectrum after autogalactosylation:

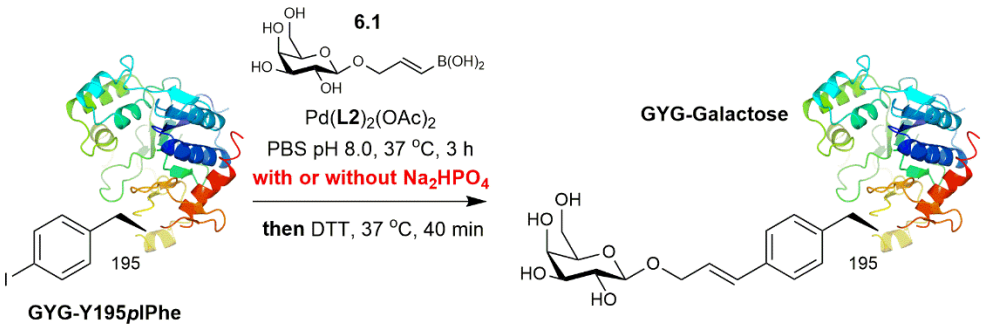


M_{calc} (galactosylation): 29784 + 162n
M_{obs}: 29943 (n = 1)
29782 (n = 0, non-galactosylated)
+ inactive side-products from cross-coupling:
(29563 29604, see **figure 6.7** and chapter 3 pp. **172-3**)

6.5.1.3 Effect of Na_2HPO_4 on Cross-Couplings to GYG-Galactose and GYG-Mannose

In the main text, it was mentioned that Na_2HPO_4 appeared beneficial to GYG-Galactose cross-coupling, whereas this was not the case for either GYG-Mannose formation. This is illustrated below for each enzyme in turn, comparing example cross-couplings with or without Na_2HPO_4 addition.

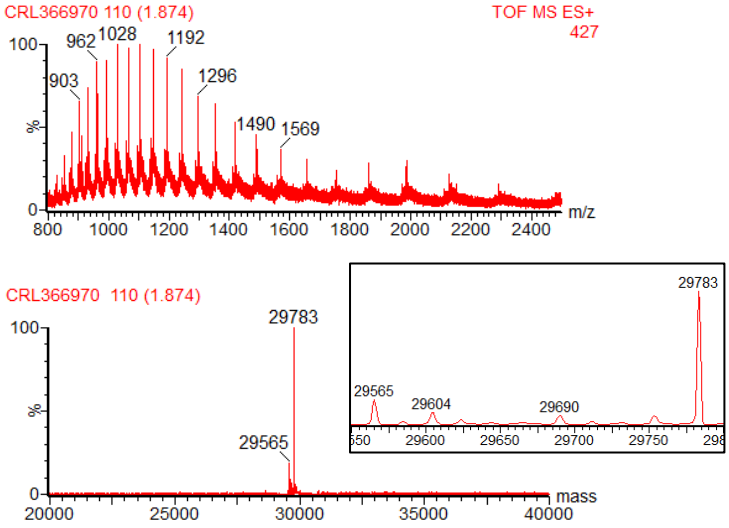
a) GYG-Galactose:



A stock solution of boronic acid **6.1** (27 mg/ml) was prepared either in H_2O or in 15.7 mg/ml aqueous Na_2HPO_4. Non-his-tagged GYG-Y195*p*IPhe (50 µl of 0.85 mg/ml, 1.43 nmol) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was combined with boronic acid added from one or the other stock solution (48 µl, 4900 nmol, 3400 eq.). $Pd(L2)_2(OAc)_2$ catalyst was added from 10 mM stock solution (6.3 µl, 63 nmol, 0.6 mM final concentration). The reaction mixtures were briefly vortexed, then incubated at 37 °C with shaking (600 rpm). After 3 hours, DTT was added to a final concentration of 7.5 mM (from a 115 mM solution in H_2O), the mixture shaken for a further 40 minutes, then analysed by LC-MS.

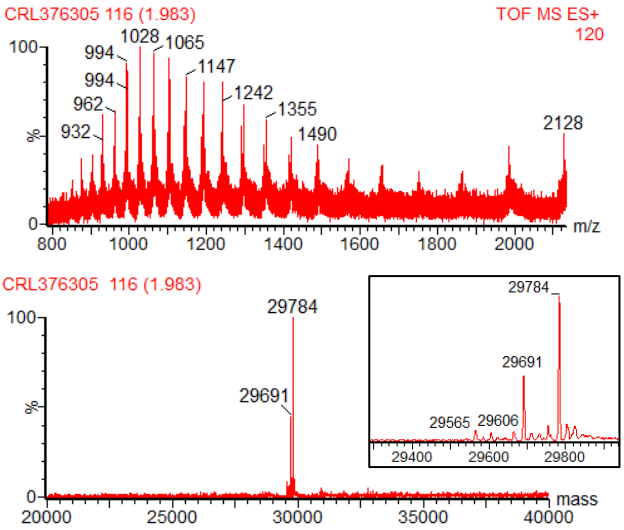
The absence of Na_2HPO_4 appeared to result in slightly lower conversion as a result of reduced starting material consumption (although levels of alternative reaction pathways were suppressed). Furthermore, as a result of more equal distributions of side-products with Na_2HPO_4, the cross-coupled enzyme was the main product by a greater majority with respect to any one side-product in this case.

With Na_2HPO_4: 74 % cross-coupled enzyme w.r.t. side-products



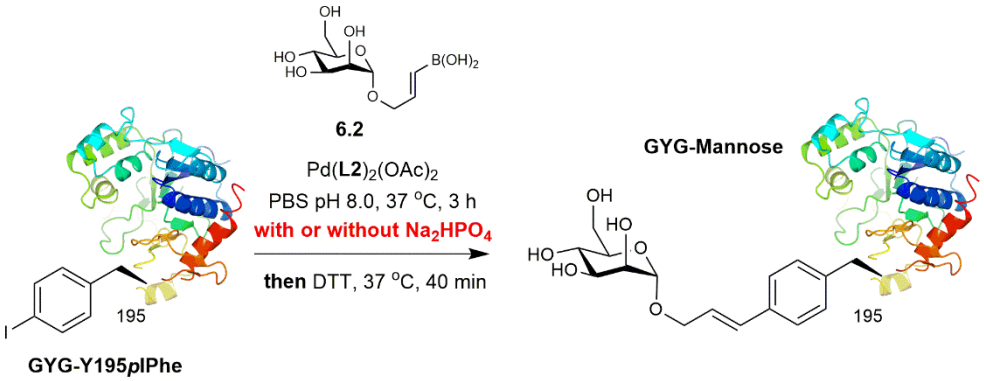
M_{calc}: 29784
M_{obs}: 29783 (74 %)
29565 (de-iodination)
29604 (see **figure 6.7**)
29690 (GYG-Y195*p*IPhe)

Without Na_2HPO_4: 63 % cross-coupled enzyme w.r.t. side-products



M_{calc}: 29784
M_{obs}: 29784 (63 %)
29565 (de-iodination)
29606 (see **figure 6.7**)
29691 (GYG-Y195*p*IPhe)

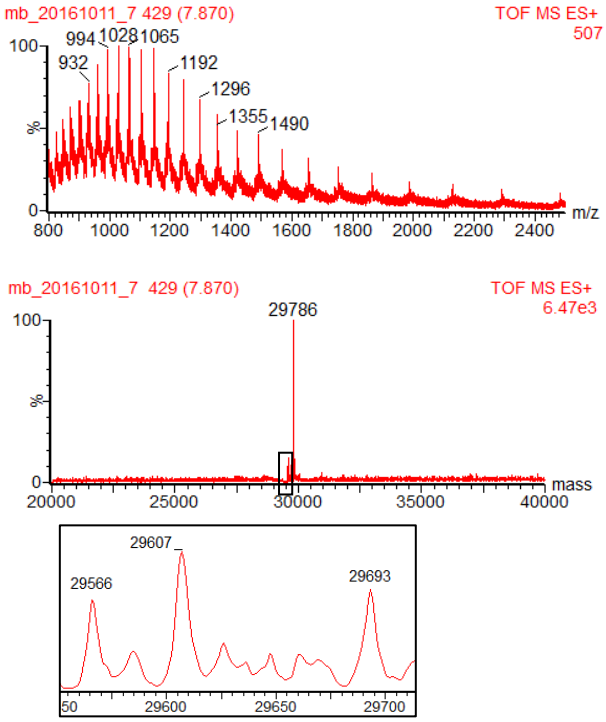
b) GYG-Mannose:



A stock solution of boronic acid **6.2** (27 mg/ml) was prepared either in H_2O or in 15.7 mg/ml aqueous Na_2HPO_4. Non-his-tagged GYG-Y195*p*IPhe (15 µl of 0.85 mg/ml, 0.43 nmol) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was combined with boronic acid added from one or the other stock solution (14.4 µl, 4900 nmol, 3400 eq.). $Pd(L2)_2(OAc)_2$ catalyst was added from 10 mM stock solution (1,9 µl, 19 nmol, 0.6 mM final concentration). The reaction mixtures were briefly vortexed, then incubated at 37 °C with shaking (600 rpm). After 3 hours, DTT was added to a final concentration of 7.5 mM (from a 115 mM solution in H_2O), the mixture shaken for a further 40 minutes, then analysed by LC-MS.

The absence of Na_2HPO_4 appeared to result in slightly higher conversion in this case, although differences are minimal. This is consistent with prior results for GYG-Glucose (chapter 3) and GYG-Maltose (chapter 4).

With Na_2HPO_4: 74 % cross-coupled enzyme w.r.t. side-products



M_{calc}: 29784

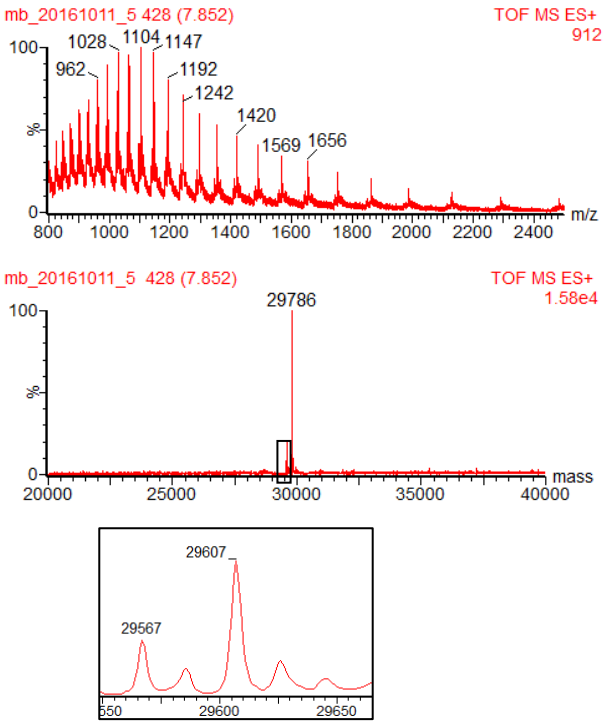
M_{obs}: 29786 (74 %)

29566 (de-iodination)

29607 (see **figure 6.7**)

29693 (GYG-Y195*p*IPhe)

Without Na_2HPO_4: 78 % cross-coupled enzyme w.r.t. side-products



M_{calc}: 29784

M_{obs}: 29786 (78 %)

29567 (de-iodination)

29607 (see **figure 6.7**)

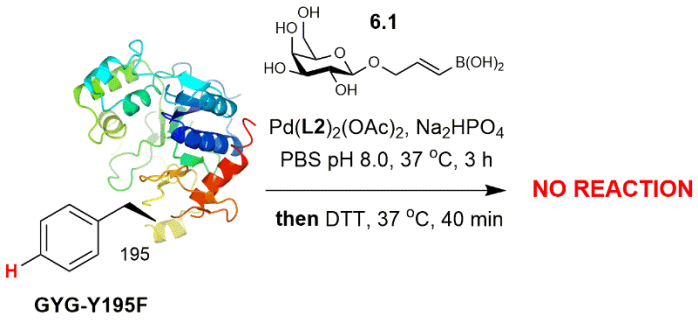
6.5.1.4 Further Investigation of Cross-Coupling Side-products

In section 6.2, it was noted that the iodine-lacking enzyme GYG-Y195F failed to react with galactose boronic acid **6.1**. Given the increased levels of side-products for this coupling w.r.t. couplings to generate GYG-Glucose, GYG-Mannose, this provides still stronger evidence for the site-specificity of both product and side-product formation.

It was also noted that the removal of MeCN from the solvent system during LC-MS analysis did not affect the side-products present, implying that the side product with M_{calc} = 29606 could not arise from "quenching" of an Aryl-Pd species on MeCN during LC-MS (i.e. essentially coupling between Ar-I and MeCN, this would give the correct mass). This further supports the notion that this mass may arise from loss of the carbohydrate via Pd-mediated C_{allyl}-O bond cleavage, with hydride displacing the carbohydrate (see chapter 3, pp. 172-3 for full discussion and chapter 6, fig. 6.8 for illustration of proposed structure).

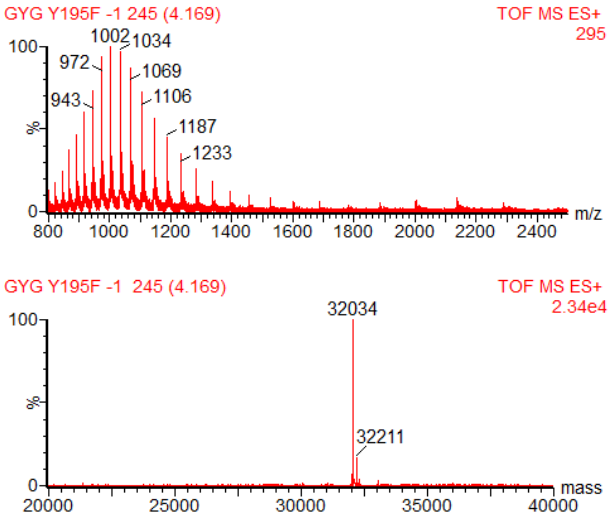
The sections below detail these experiments. It should be noted that these are directly analogous for those carried out on GYG-Glucose in chapter 3.

a) Negative Control Reaction with GYG-Y195F



A stock solution of boronic acid **6.1** (27 mg/ml) was prepared in 15.7 mg/ml aqueous Na_2HPO_4. Non-his-tagged GYG-Y195*p*IPhe (30 µl of 0.92 mg/ml, 0.86 nmol) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was combined with boronic acid stock solution (28.8 µl, 2950 nmol, 3400 eq.). $Pd(L2)_2(OAc)_2$ catalyst was added from 10 mM stock solution (3.8 µl, 38 nmol, 0.6 mM final concentration). The reaction mixture was briefly vortexed, then incubated at 37 °C with shaking (600 rpm). After 3 hours, DTT was added to a final concentration of 7.5 mM (from a 115 mM solution in H_2O), the mixture shaken for a further 40 minutes. The mixture was desalted using a Micro Bio-Spin® 6 column (BioRad) according to manufacturer instructions, using pH 8.0 phosphate buffer as eluant, before being analysed by LC-MS. No reaction was observed.

Before reaction:

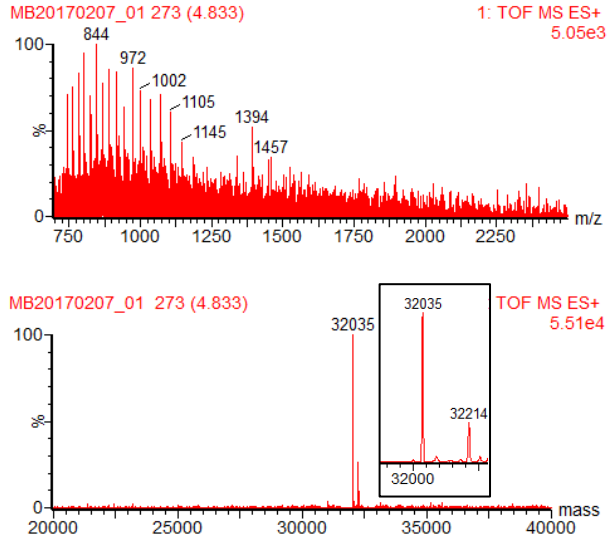


M_{calc} (GYG-Y195F): 32032

M_{obs}: 32034

32211 (gluconoylated GYG-Y195F)

After reaction:



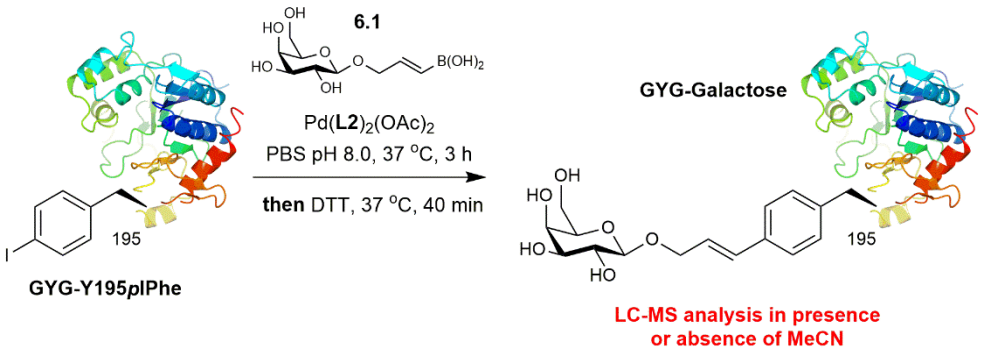
M_{calc} (GYG-Y195F): 32032

M_{obs}: 32035 (GYG-Y195F, unreacted)

32214 (gluconoylated GYG-Y195F, present in starting material, see above)

The peak at 32214 corresponds to gluconoylation of the protein following expression (M + 178/9 Da), as can be seen, it is present in the starting material; it does not result from reaction. As noted previously (chapter 2), gluconoylation of N-terminal polyhistidine tags during expression is a well-known phenomenon,[49] and was observed for his-tagged GYG-Y195pIPhe as well as GYG-Y195F here (see **figure 2.4**, chapter 2).

b) Cross-coupling to Galactose Boronic Acid 6.1: Varying the MS Mobile Phase

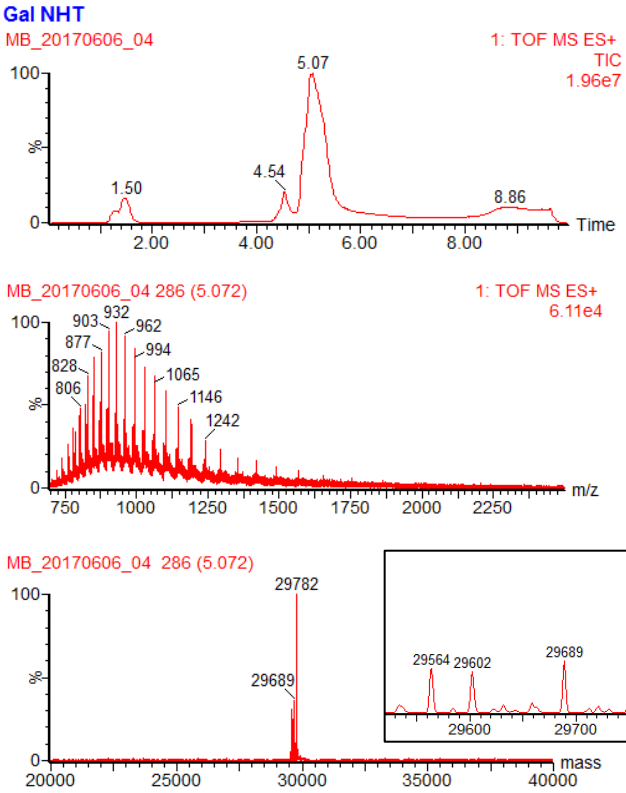


A stock solution of boronic acid **6.1** (27 mg/ml) was prepared in 15.7 mg/ml aqueous Na_2HPO_4. Non-his-tagged GYG-Y195*p*IPhe (25.5 µl of 0.85 mg/ml, 0.73 nmol) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was combined with boronic acid stock solution (24.5 µl, 2500 nmol, 3400 eq.). $Pd(L2)_2(OAc)_2$ catalyst was added from 10 mM stock solution (3.15 µl, 32 nmol, 0.6 mM final concentration). The reaction mixtures were briefly vortexed, then incubated at 37 °C with shaking (600 rpm). After 3 hours, DTT was added to a final concentration of 7.5 mM (from a 115 mM solution in H_2O), the mixture shaken for a further 40 minutes.

Samples were analysed by LC-MS using one of two different gradient methods. All other parameters were maintained as identical (see appendix: for reference, the parameters used correspond to "method 4" in the section "Protein LC-MS").

Method 1: Solvent A = 0.1 % Formic Acid in H_2O, Solvent B = 0.1 % Formic Acid in MeCN.

Time (min)	% Solvent A	% Solvent B
0.0	95	5
1.0	95	5
7.0	5	95
8.0	5	95
8.1	95	5
10.0	95	5



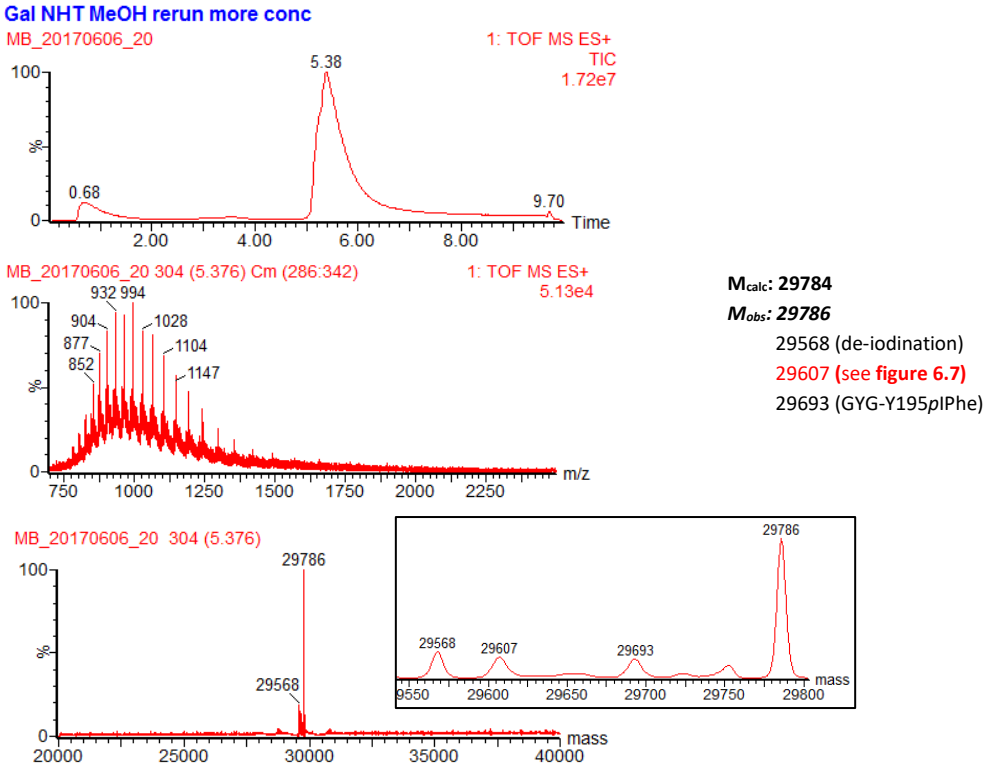
M_{calc}: 29784
M_{obs}: 29782
29564 (de-iodination)
29602 (see **figure 6.7**)
29689 (GYG-Y195*p*IPhe)

The cross-coupled product is seen, along with expected de-iodination side product, GYG-Y195*p*IPhe and the side-product under investigation (M_{obs} = 29602).

Method 2: Solvent A = 0.1 % Formic Acid in H_2O, Solvent B = 0.1 % Formic Acid in MeOH.

Time (min)	% Solvent A	% Solvent B
0.0	95	5
1.0	95	5
4.0	5	95
8.0	5	95
8.1	95	5
10.0	95	5

Note: the sharper gradient was rendered necessary to ensure that the protein eluted fully from the column, given the reduced elution power of MeOH.



Notably, the side-product under investigation was still observed in the absence of MeCN (M_{obs} = 29607). This means that it cannot result from an on-MS "quench" of an enzyme-Pd intermediate by MeCN; instead we believe it may arise from loss of the carbohydrate residue via C_{allyl}-O cleavage, leaving an "allyl-GYG" species **3.11** (see chapter 3 pp. 172-3 and **figure 6.7** for full details).

6.5.2 Investigation of Catalytically Impaired GYG-T83M Glycoforms

6.5.2.1 Site-directed Mutagenesis to Generate GYG-T83M-Y195X (X = TAG) Plasmid

Site-directed mutagenesis to generate plasmid GYG-T83M-Y195X was carried out using plasmid GYG-Y195X (pnic8-bsa4 vector) as a template (see chapter 2; this latter plasmid was generated previously in our group[50]). Mutagenesis was carried out using an Agilent QuikChange II XL kit, according to manufacturer instructions. 2 reactions were set up identically, except for the amount of template DNA, which was either 10 or 50 ng. The following primers, purchased from Invitrogen, were used. Numbers refer to the position of the mutation in the non-his-tagged protein (i.e. they do not include the polyhistidine tag, which is 22 residues long).

T83M:

Forward: 5'-ctgtgtcag**cat**gacacccaactctggcc-3'

Reverse: 5'-ggccagagttgggtgtc**atg**ctgacacag-3'

Nucleotides: 29

T_m = 78.8 °C

Transformation of the resultant PCR products into XL10 gold *e. coli* cells was carried out according to manufacturer instructions, using either 0.4 ng, 2.0 ng or 5.0 ng of DNA and selecting with kanamycin (50 μg/ml). Colonies were seen in all cases. From one such transformation, four colonies were used to inoculate 4 x 5 ml of LB media (1 colony for each culture) and grown at 37 °C overnight. Plasmid DNA from the cultures was isolated using Qiagen miniprep kit according to manufacturer instructions. Sequencing of the resultant DNA confirmed successful incorporation of the desired mutation in each case; the DNA sequence and translated protein sequence are shown below. The new mutation is shown in red, the original (Y195X) mutation is shown in blue, and the 22-residue polyhistidine tag is highlighted in yellow.

Amino acid numbers *do not* include this polyhistidine tag (i.e. "T83M" refers to the 83rd residue after this tag, which is thus the 105th residue in total); this is to avoid confusion with respect to prior literature.[7,21] There are 20 amino acids per line.

GYG-T83M-Y195X (X = TAG = *p*IPhe in expressed protein)

```
    atgcaccatcatcatcattcttctggtgtagatctgggtaccgagaacctgtacttc
    M  H  H  H  H  H  H  S  S  G  V  D  L  G  T  E  N  L  Y  F
    caatccatgacagatcaggcctttgtgacactaaccacaaacgatgcctacgccaaaggt
    Q  S  M  T  D  Q  A  F  V  T  L  T  T  N  D  A  Y  A  K  G
19  gccctggtcctgggatcatctctgaaacagcacaggaccaccaggaggctggtcgtgctc
    A  L  V  L  G  S  S  L  K  Q  H  R  T  T  R  R  L  V  V  L
39  gccacccctcaggtctcagactccatgagaaaagttttagagacagtctttgatgaagtc
    A  T  P  Q  V  S  D  S  M  R  K  V  L  E  T  V  F  D  E  V
59  atcatggtagatgtcttggacagtggcgattctgctcatctaaccttaatgaagaggcca
    I  M  V  D  V  L  D  S  G  D  S  A  H  L  T  L  M  K  R  P
79  gagttgggtgtcatgctgacaaagctccactgctggtcgcttacacagtattcaaaatgt
    E  L  G  V  M  L  T  K  L  H  C  W  S  L  T  Q  Y  S  K  C
99  gtattcatggatgcagatactctggtcctagcaaatattgatgatcttttgacagagaa
    V  F  M  D  A  D  T  L  V  L  A  N  I  D  D  L  F  D  R  E
119 gaattgtcagcagcaccagacccagggtggcctgactgcttcaattccggagtcttcgtt
    E  L  S  A  A  P  D  P  G  W  P  D  C  F  N  S  G  V  F  V
139 tatcagccttcagttgaaacatacaatcagctgttgcatcttgcttctgagcaaggtagt
    Y  Q  P  S  V  E  T  Y  N  Q  L  L  H  L  A  S  E  Q  G  S
159 tttgatggtggggaccaaggcatactgaacacatttttagcagctgggcaacaacagat
    F  D  G  G  D  Q  G  I  L  N  T  F  F  S  S  W  A  T  T  D
179 atcagaaaacacctgccgtttatttataacctaagcagcatctctatatagtcctacctc
    I  R  K  H  L  P  F  I  Y  N  L  S  S  I  S  I  -  S  Y  L
199 ccggcatttaaagtgtttggtgcaagtgccaaagttgtgcatttcctgggacgagtcaaa
    P  A  F  K  V  F  G  A  S  A  K  V  V  H  F  L  G  R  V  K
219 ccatggaattatacttatgatcccaaaacaaaaagtgtcaaaagtgaggcccatgatccc
    P  W  N  Y  T  Y  D  P  K  T  K  S  V  K  S  E  A  H  D  P
239 aacatgactcatccagagtttctcatcctgtggtggaacatctttaccaccaacgtttta
    N  M  T  H  P  E  F  L  I  L  W  W  N  I  F  T  T  N  V  L
259 cctctgcttcaatga
    P  L  L  Q  -
```

6.5.2.2 Expression of GYG-T83M-Y195*p*IPhe Protein (with his-tag)

Transformation and Expression

BL21(DE3) cells were thawed on ice. A 20 μl aliquot was transferred to a pre-chilled 14 ml falcon tube. Plasmids pEVOL-pIPhe (1.0 μl, ca. 50 ng/μl) and GYG-T83M-Y195X (0.4 μl, ca. 150 ng/μl) were added, and the tube placed on ice for 30 minutes with occasional swirling. Cells were heat shocked (42 °C, 40 s) then immediately placed on ice for a further 10 minutes. 80 μl of SOC media was added, and cells incubated at 37 °C for 1 h with shaking (200 rpm). Cells were plated on LB-agar plates containing both chloramphenicol (25 μg/ml) and kanamycin (50 μg/ml). Plates were incubated at 37 °C overnight.

Single colonies obtained from the above transformation were used to inoculate 3 x 10 ml LB media, supplemented with kanamycin (50 μg/ml) and chloramphenicol (25 μg/ml). Cells were grown at 37 °C overnight with shaking (200 rpm). Resultant cells were used to inoculate 3 x 1L of similarly supplemented SB media (see recipe below). Cells were grown (37 °C, 180 rpm) to a final OD_{600} of 0.6-0.7 (typically 3 h). For each L culture, L-Arabinose was added to a final concentration of 0.02 %, followed by L-*p*-iodophenylalanine (*p*IPhe, **2.1**) to a final concentration of 2 mM (580 mg was dissolved in 5 ml 1M NaOH). Cells were incubated at 30 °C, 180 rpm for a further 2 hours, before addition of IPTG to a final concentration of 0.5 mM. Cells were cooled to 18 °C and incubated with shaking at 180 rpm for 16 h. After this time, cells were harvested through centrifugation (7000 x g, 7 minutes, 4 °C). Cells may be flash frozen and stored at -80 °C prior to lysis and purification if desired.

SB media: (per L culture) 32 g tryptone, 20 g yeast extract, 5 g NaCl.

Cell Lysis and Purification

Cell pellets were resuspended in lysis buffer (50 ml, see recipe below) with addition of cOmplete™ EDTA-free protease inhibitor cocktail. Lysozyme (10 mg) and DNaseI (approx. 1 mg) were added and the suspension gently shaken at 4 °C for 30 minutes. Cells were then sonicated (40 % amplitude, 5 s on, 10 s off, 12 cycles) and cell debris removed by centrifugation (50000 x g, 4 °C, 20 minutes).

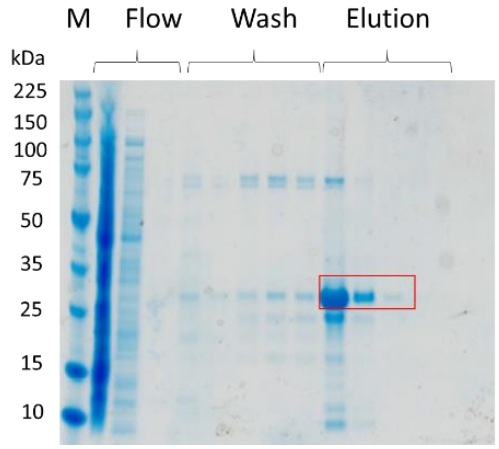
The clarified lysate was loaded onto a gravity flow Ni-NTA column. The flowthrough was collected, and the protein was purified using a step gradient of imidazole (20 mM – 40 mM – 60 mM – 250 mM, protein elutes at 250 mM). Fractions were analysed by SDS-PAGE. For isolation of his-tagged protein, fractions containing the desired protein were combined and dialysed against dialysis buffer (2 x 2L, 10000 MWCO,

4 °C, 2 h then overnight). Protein was concentrated by Vivaspin (10000 MWCO, Sartorius Stedem) to a concentration of >1 mg/ml, flash frozen and stored in 100 µl aliquots at – 80 °C. A protein yield of 1.1 mg/L culture was obtained. LC-MS confirmed the isolation of the desired his-tagged GYG-T83M-Y195*p*IPhe protein in good purity.

Lysis Buffer: 20 mM imidazole, 50 mM Na-HEPES, 0.5 M NaCl, 0.5 mM TCEP, 5 % Glycerol, pH 7.5

Dialysis Buffer: 10 mM Na-HEPES, 0.5 M NaCl, 1 mM Na-EDTA, 5 % glycerol, 1 mM DTT, pH 7.5.

SDS-PAGE after Ni-NTA purification:



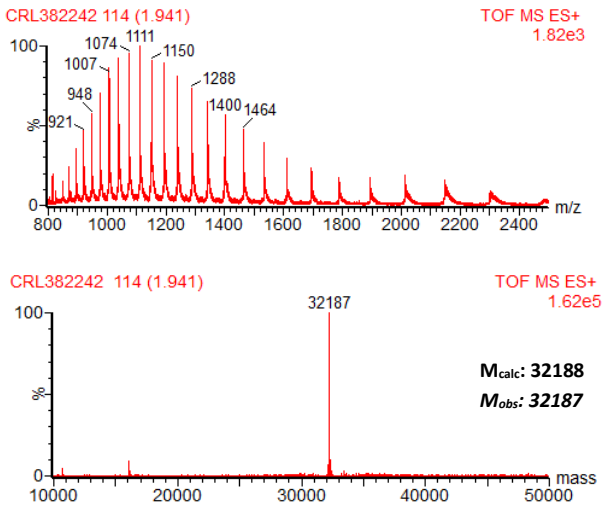
M = molecular weight markers

Flow = flowthrough (20 mM imidazole)

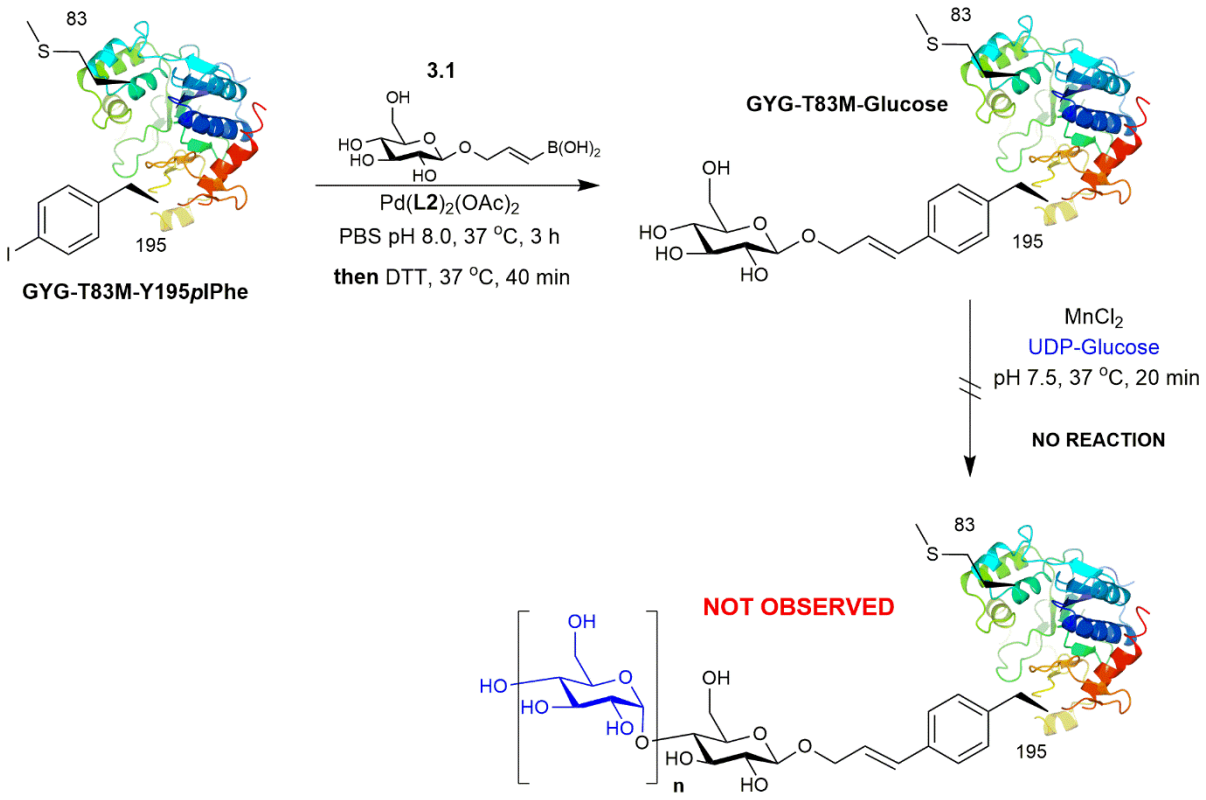
Wash = 40 mM and 60 mM imidazole

Elution = 250 mM imidazole

LC-MS spectrum of purified protein:



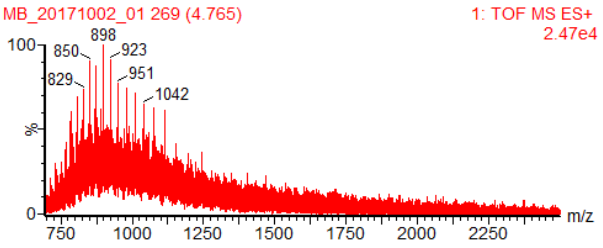
6.5.2.3 Generation and Autoglucosylation Assay of GYG-T83M-Glucose

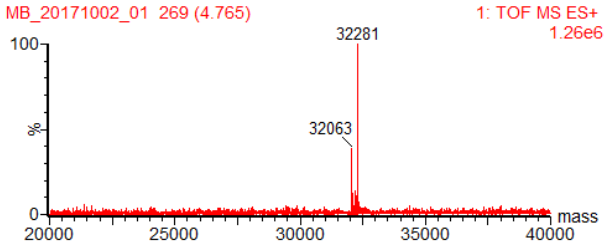


Cross-Coupling:

A stock solution of glucose boronic acid **3.1** (27 mg/ml) was prepared in H_2O. His-tagged GYG-T83M-Y195*p*IPhe (50 µl of 0.92 mg/ml, 1.43 nmol) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was combined with boronic acid stock solution (48 µl, 4900 nmol, 3400 eq.). $Pd(L2)_2(OAc)_2$ catalyst was added from 10 mM stock solution (6.3 µl, 63 nmol, 0.6 mM final concentration). The reaction mixtures were briefly vortexed, then incubated at 37 °C with shaking (600 rpm). After 3 hours, DTT was added to a final concentration of 7.5 mM (from a 115 mM solution in H_2O), and the mixture shaken for a further 40 minutes. The mixture was buffer exchanged into pH 7.5 25 mM Na-HEPES by PD SpinTrap G-25; an aliquot was analysed by LC-MS. 100 µl of the sample was further desalted by dialysis against the same buffer (10000 MWCO, 4 °C, 2 h then 2 h then overnight, 3 x 300 ml buffer), using a Slide-A-Lyzer™ MINI dialysis device. An aliquot was analysed by LC-MS, confirming formation of cross-coupled product:

LC-MS after cross-coupling



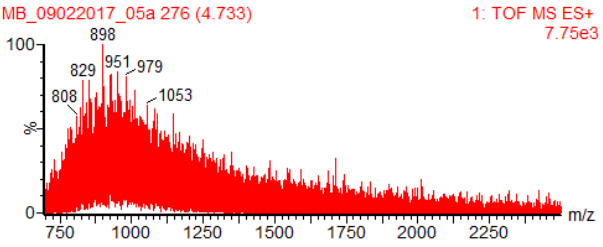


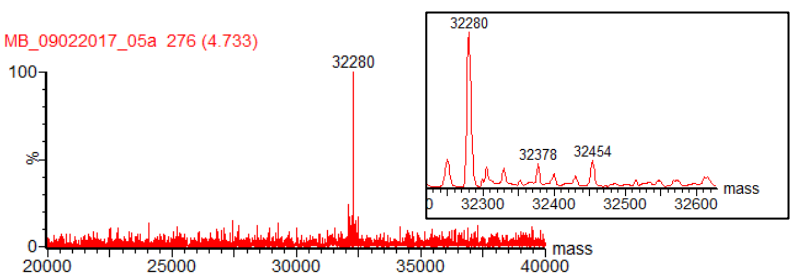
M_{calc}: 32280
M_{obs}: 32281 (70 %)
32063 (de-iodination)

Autoglucosylation:

40 µl of the resultant mixture was adjusted to a concentration of 0.2 mg/ml, before being incubated with $MnCl_2$ and UDP-Glucose disodium salt at final concentrations of 1 mM each (5 µl of 10 mM freshly prepared stock solution in H_2O was added). The reaction was incubated at 37 °C with shaking (500 rpm) for 20 minutes. After this time, Na-EDTA was added to a final concentration of 20 mM to quench glucosylation, and the solution placed on ice prior to analysis by LC-MS. No peaks corresponding to glucosylated products were observed:

LC-MS after autoglucosylation:

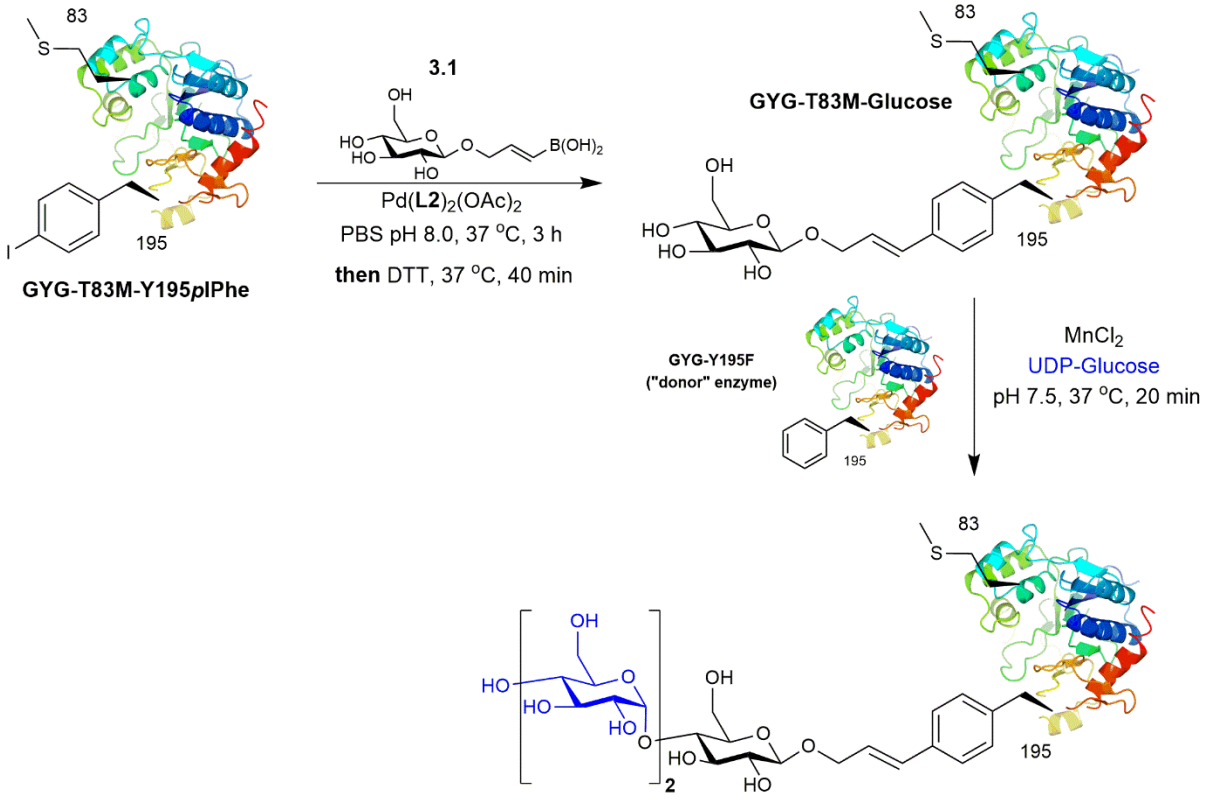




M_{calc}: 32280 + 162n (= 32442, 32604…)
M_{obs}: 32280 (n = 0, inactive)
Glucosylated products not seen.

+ inactive side-products from cross-coupling
(32063 = de-iodination)

6.5.2.4 Inter-enzyme glucosylation of GYG-T83M-Glucose using GYG-Y195F



Cross-coupling:

Cross-coupling of GYG-T83M-Y195*p*IPhe to boronic acid **3.1** was carried out exactly as described in section 6.5.2.3 "cross-coupling", to generate GYG-T83M-Glucose (the same sample of cross-coupled enzyme was used for both the autoglucosylation and inter-enzyme glucosylation experiments; LC-MS spectra are thus as shown above).

Glucosylation:

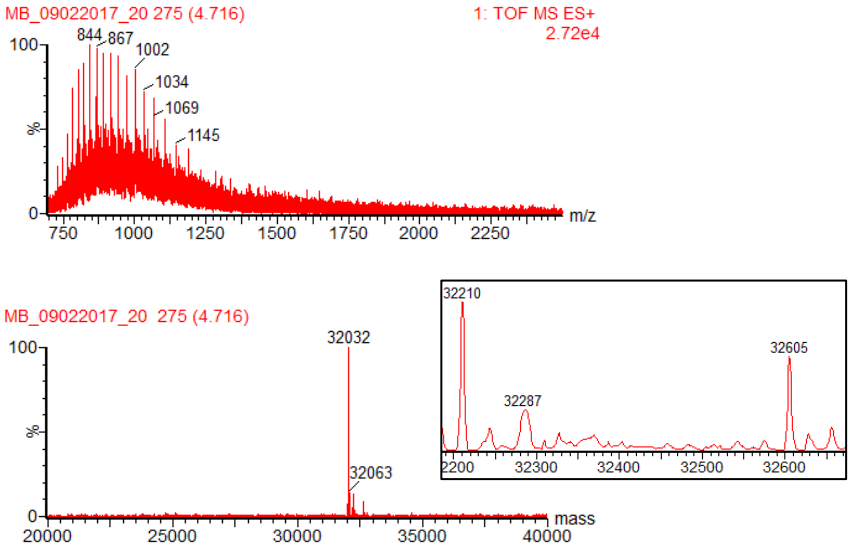
24 µl of resultant GYG-T83M-Glucose cross-coupling reaction mixture was adjusted to a concentration of 0.2 mg/ml. To this was added 8 µl of GYG-Y195F enzyme (0.2 mg/ml) in pH 7.5 25 mM Na-HEPES buffer (see note below). The enzymes were then incubated with $MnCl_2$ and UDP-Glucose disodium salt at final concentrations of 1 mM each (4 µl of 10 mM freshly prepared stock solution in H_2O was added). The reaction was incubated at 37 °C with shaking (500 rpm) for 20 minutes. After this time, Na-EDTA was added to a final concentration of 20 mM to quench glucosylation. The reaction was placed on ice prior to LC-MS analysis.

Note: owing to the superior MS ionisation of GYG-Y195F enzyme relative to cross-coupled enzyme, it was necessary to use a lesser amount of the former relative to the latter. This likely reflects the lack of Pd-treatment for GYG-Y195F (this was avoided in order to be sure that this enzyme retained 100% of its natural inter-enzyme glucosylation activity), given the known detrimental effects of Pd on protein MS ionisation.[51,52]

LC-MS shows the following:

Peaks belonging to GYG-Y195F: **32032** (GYG-Y195F), **32210** (gluconoylated GYG-Y195F, present in starting material as discussed in section 6.5.1.3 and chapter 2; N-terminal his-tag gluconoylation is a known feature of protein expression).[49]

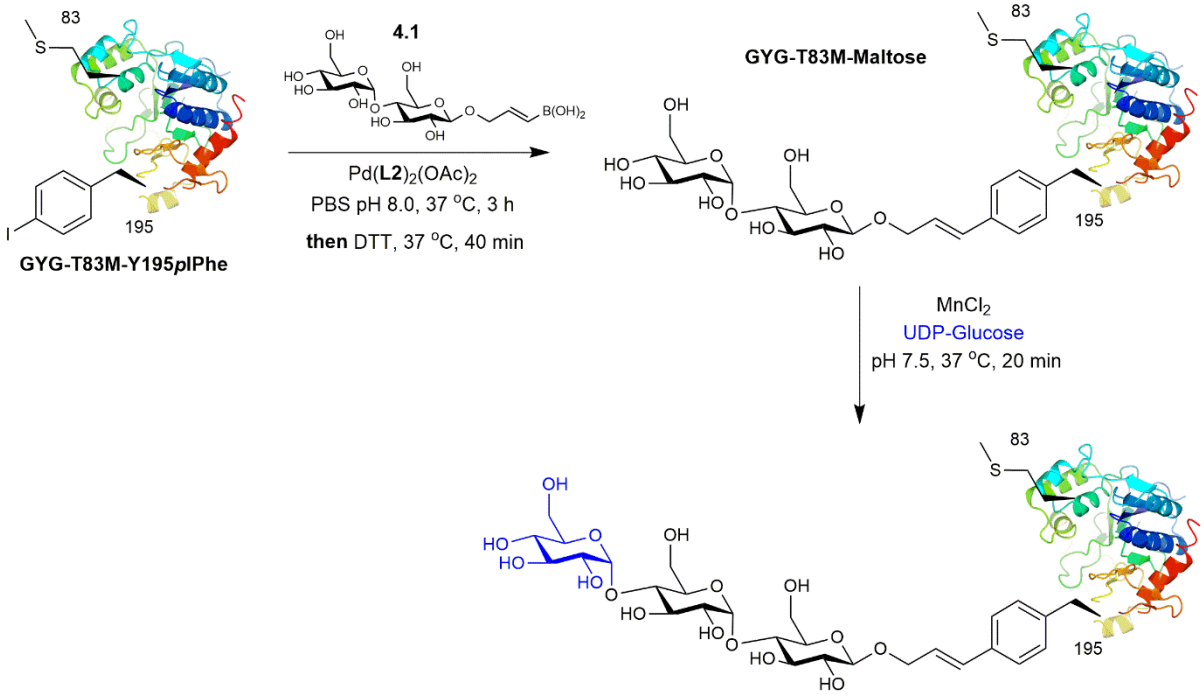
Peaks belonging to GYG-T83M-Glucose and derivatives: **32287** (GYG-T83M-Glucose), **32605** **(di-glucosylated GYG-T83M-Glucose)**, **32063** (de-iodinated enzyme).



M_{calc} (glucosylated GYG-T83M-Glucose): 32280 + 162n
M_{obs} (glucosylated GYG-T83M-Glucose): ***32605 (n = 2, calc. 32604)***
32287 (n = 0)
32063 (de-iodinated enzyme)

M_{calc} (GYG-Y195F): 32032
M_{obs} (GYG-Y195F): **32032**
32210 (gluconoylated)

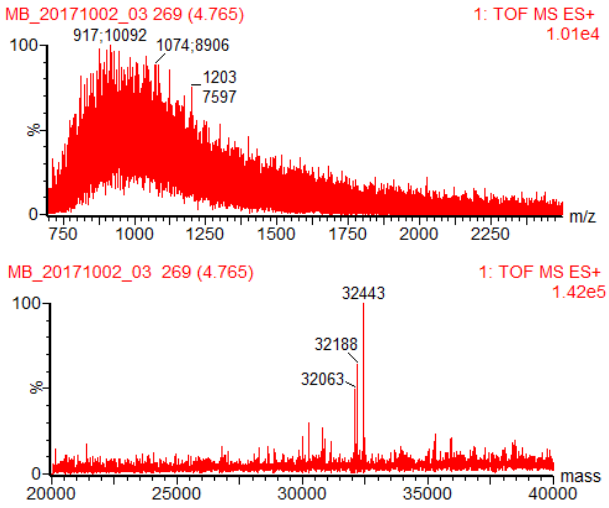
6.5.2.5 Generation and Autoglucosylation Assay of GYG-T83M-Maltose



Cross-Coupling:

A stock solution of maltose boronic acid **4.1** (45 mg/ml) was prepared in H_2O. His-tagged GYG-T83M-Y195*p*IPhe (40 µl of 0.94 mg/ml, 1.17 nmol) in pH 8.0 phosphate buffer (20 mM NaH_2PO_4, 50 mM NaCl) was combined with boronic acid stock solution (40 µl, 4230 nmol, 3600 eq.). $Pd(L2)_2(OAc)_2$ catalyst was added from 10 mM stock solution (5.14 µl, 51 nmol, 0.6 mM final concentration). The reaction mixtures were briefly vortexed, then incubated at 37 °C with shaking (600 rpm). After 3 hours, DTT was added to a final concentration of 7.5 mM (from a 115 mM solution in H_2O), and the mixture shaken for a further 40 minutes. The mixture was buffer exchanged into pH 7.5 25 mM Na-HEPES by PD SpinTrap G-25; an aliquot was analysed by LC-MS. 100 µl of the sample was further desalted by dialysis against the same buffer (10000 MWCO, 4 °C, 2 h then 2 h then overnight, 3 x 300 ml buffer), using a Slide-A-Lyzer™ MINI dialysis device. An aliquot was analysed by LC-MS, confirming formation of cross-coupled product, along with de-iodination side-product and unreacted GYG-T83M-Y195*p*IPhe.

LC-MS after cross-coupling

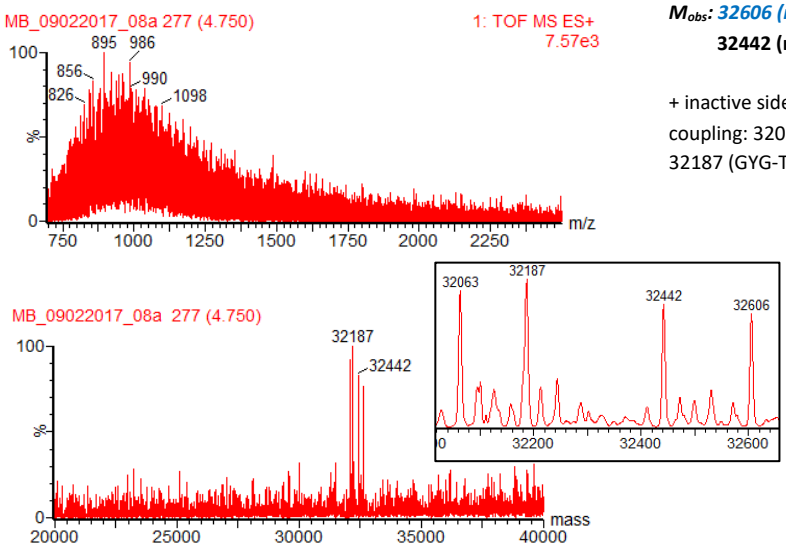


M_{calc}: 32442
M_{obs}: 32443 (50 %)
32063 (de-iodination)
32188 (GYG_Y195*p*IPhe)

Autoglucosylation:

40 µl of the resultant mixture was adjusted to a concentration of 0.2 mg/ml, before being incubated with $MnCl_2$ and UDP-Glucose disodium salt at final concentrations of 1 mM each (5 µl of 10 mM freshly prepared stock solution in H_2O was added). The reaction was incubated at 37 °C with shaking (500 rpm) for 20 minutes. After this time, Na-EDTA was added to a final concentration of 20 mM to quench glucosylation, and the solution placed on ice prior to analysis by LC-MS. A peak corresponding to the addition of a single glucose residue was observed, with concomitant consumption of GYG-T83M-Maltose (as judged by intensity relative to the inactive cross-coupling side-products):

LC-MS after autoglucosylation:



M_{calc}: 32442 + 162n
M_{obs}: 32606 (n = 1, glucosylated)
32442 (n = 0, non-glucosylated)

+ inactive side-products from cross-coupling: 32063 (de-iodination), 32187 (GYG-T83M-Y195*p*IPhe)

6.6 References

1. Chalker, J. M., Bernardes, G. J. L. & Davis, B. G. A 'Tag-and-Modify' Approach to Site-Selective Protein Modification. *Acc. Chem. Res.* **44,** 730–741 (2011).

2. Chalker, J. M., Lercher, L., Rose, N. R., Schofield, C. J. & Davis, B. G. Conversion of cysteine into dehydroalanine enables access to synthetic histones bearing diverse post-translational modifications. *Angew. Chemie - Int. Ed.* **51,** 1835–1839 (2012).

3. Wright, T. H. *et al.* Posttranslational mutagenesis: A chemical strategy for exploring protein side-chain diversity. *Science.* **354,** aag1465 (2016).

4. Gupta, S. Sen *et al.* Accelerated bioorthogonal conjugation: A practical method for the ligation of diverse functional molecules to a polyvalent virus scaffold. *Bioconjug. Chem.* **16,** 1572–1579 (2005).

5. Chalker, J. M., Wood, C. S. C. & Davis, B. G. A convenient catalyst for aqueous and protein Suzuki-Miyaura cross-coupling. *J. Am. Chem. Soc.* **131,** 16346–16347 (2009).

6. Issoglio, F. M., Carrizo, M. E., Romero, J. M. & Curtino, J. A. Mechanisms of monomeric and dimeric glycogenin autoglucosylation. *J. Biol. Chem.* **287,** 1955–1961 (2012).

7. Chaikuad, A. *et al.* Conformational plasticity of glycogenin and its maltosaccharide substrate during glycogen biogenesis. *Proc. Natl. Acad. Sci. U.S.A.* **108,** 21028–21033 (2011).

8. Roseman, S. The synthesis of complex carbohydrates by multiglycosyltransferase systems and their potential function in intercellular adhesion. *Chem. Phys. Lipids* **5,** 270–297 (1970).

9. Yang, M., Brazier, M., Edwards, R. & Davis, B. G. High-throughput mass-spectrometry monitoring for multisubstrate enzymes: Determining the kinetic parameters and catalytic activities of glycosyltransferases. *ChemBioChem* **6,** 346–357 (2005).

10. Fernández-González, M. *et al.* Site-selective chemoenzymatic construction of synthetic glycoproteins using endoglycosidases. *Chem. Sci.* **1,** 709 (2010).

11. Dewitte, G. *et al.* Screening of recombinant glycosyltransferases reveals the broad acceptor specificity of stevia UGT-76G1. *J. Biotechnol.* **233,** 49–55 (2016).

12. Song, C. *et al.* Functional Characterization and Substrate Promiscuity of UGT71 Glycosyltransferases from Strawberry (fragaria x ananassa). *Plant Cell Physiol.* **56,** 2478–2493 (2015).

13. Salas, A. P. *et al.* Deciphering the late steps in the biosynthesis of the anti-tumour indolocarbazole staurosporine: Sugar donor substrate flexibility of the StaG glycosyltransferase. *Mol. Microbiol.* **58,** 17–27 (2005).

14. Williams, G. J., Goff, R. D., Zhang, C. & Thorson, J. S. Optimizing Glycosyltransferase Specificity via 'Hot Spot' Saturation Mutagenesis Presents a Catalyst for Novobiocin Glycorandomization. *Chem. Biol.* **15,** 393–401 (2008).

15. Aharoni, A. *et al.* High-throughput screening methodology for the directed evolution of glycosyltransferases. *Nat. Methods* **3,** 609–614 (2006).

16. Yang, G. *et al.* Fluorescence Activated Cell Sorting as a General Ultra High-throughput Screening Method for Directed Evolution of Glycosyltransferases. *JACS* **132,** 10570–10577 (2010).

17. Carrizo, M. E., Miozzo, M. C., Goldraij, A. & Curtino, J. A. Purification of rabbit skeletal muscle proteoglycogen: Studies on the glucosyltransferase activity of polysaccharide-free and -bound glycogenin. *Glycobiology* **7,** 571–578 (1997).

18. Alonso, M. D., Lagzdins, E. J., Lomako, J., Lomako, W. M. & Whelan, W. J. New and specific nucleoside diphosphate glucose substrates for glycogenin. *FEBS Lett.* **359,** 110–112 (1995).

19. Meezan, E. *et al.* Xylosyl transfer to an endogenous renal acceptor. Characteristics of the reaction and properties of the product. *J. Biol. Chem.* **269,** 11503–11508 (1994).

20. Manzella, S. *et al.* Inhibition of glycogenin-catalyzed glucosyl and xylosyl transfer by cytidine 5'-diphosphate and related compounds. *Arch. Biochem. Biophys.* **320,** 361–368 (1995).

21. Carrizo, M. E., Romero, J. M., Issoglio, F. M. & Curtino, J. A. Structural and biochemical insight into glycogenin inactivation by the glycogenosis-causing T82M mutation. *FEBS Lett.* **586,** 254–7 (2012).

22. Lomako, J., Lomako, W. M. & Whelan, W. J. Substrate specificity of the autocatalytic protein that primes glycogen synthesis. *FEBS Lett.* **264,** 13–16 (1990).

23. Cao, Y., Steinrauf, L. K. & Roach, P. J. Mechanism of Glycogenin Self-Glucosylation. *Arch. Biochem. Biophys.* **319,** 293–298 (1995).

24. Hurley, T. D., Stout, S., Miner, E., Zhou, J. & Roach, P. J. Requirements for catalysis in mammalian glycogenin. *J. Biol. Chem.* **280,** 23892–23899 (2005).

25. Alonso, M. D., Lomako, J., Lomako, W. M. & Whelan, W. J. Tyrosine-194 of glycogenin undergoes autocatalytic glucosylation but is not essential for catalytic function and activity. *FEBS Lett.* **342,** 38–42 (1994).

26. Cao, Y., Mahrenholz, A. M., DePaoli-Roach, A. A. & Roach, P. J. Characterization of rabbit skeletal muscle glycogenin: Tyrosine 194 is essential for function. *J. Biol. Chem.* **268,** 14687–14693 (1993).

27. Lin, A., Mu, J., Yang, J. & Roach, P. J. Self-glucosylation of glycogenin, the initiator of glycogen biosynthesis, involves an inter-subunit reaction. *Arch. Biochem. Biophys.* **363,** 163–170 (1999).

28. Gibbons, B. J., Roach, P. J. & Hurley, T. D. Crystal structure of the autocatalytic initiator of glycogen biosynthesis, glycogenin. *J. Mol. Biol.* **319,** 463–477 (2002).

29. Moslemi, A-R. *et al.* Glycogenin-1 deficiency and inactivated priming of glycogen synthesis. *N. Engl. J. Med.* **362,** 1203–1210 (2010).

30. Wolfsdorf, J. I. & Weinstein, D. A. Glycogen Storage Diseases. *Rev. Endocr. Metab. Disord.* **4,** 95–102 (2003).

31. Spicer, C. D. & Davis, B. G. Rewriting the bacterial glycocalyx via Suzuki-Miyaura cross-coupling. *Chem. Commun. (Camb).* **49,** 2747–2749 (2013).

32. Sun, J., Perfetti, M. T. & Santos, W. L. A method for the deprotection of alkylpinacolyl boronate esters. *J. Org. Chem.* **76,** 3571–3575 (2011).

33. Varki, A. *et al.* Symbol nomenclature for graphical representations of glycans. *Glycobiology* **25,** 1323–1324 (2015).

34. Alonso, M. D., Lomako, J., Lomako, W. M. & Whelan, W. J. A new look at the biogenesis of glycogen. *FASEB J.* **9,** 1126–37 (1995).

35. Xie, J. *et al.* The site-specific incorporation of *p*-iodo-L-phenylalanine into proteins for structure determination. *Nat. Biotechnol.* **22,** 1297–301 (2004).

36. Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C., Schultz, P. G. & April, I. A General Method for Site-Specific Incorporation of Unnatural Amino Acids into Proteins. *Science.* **244,** 182–188 (1989).

37. Young, T. S., Ahmad, I., Yin, J. A. & Schultz, P. G. An Enhanced System for Unnatural Amino Acid Mutagenesis in E. coli. *J. Mol. Biol.* **395,** 361–374 (2010).

38. Ramaiah, D., Kan, Y., Koch, T., Orum, H. & Schuster, G. B. Enzymatic reaction with unnatural substrates: DNA photolyase (Escherichia coli) recognizes and reverses thymine [2+2] dimers in the DNA strand of a DNA/PNA hybrid duplex. *Proc. Natl. Acad. Sci. U.S.A.* **95,** 12902–12905 (1998).

39. Hartman, M. C. T., Jiang, S., Rush, J. S., Waechter, C. J. & Coward, J. K. Glycosyltransferase Mechanisms: Impact of a 5-Fluoro Substituent in Acceptor and Donor Substrates on Catalysis. *Biochemistry* **46,** 11630–11638 (2007).

40. Nielsen, M. M. *et al.* Substrate and metal ion promiscuity in mannosylglycerate synthase. *J. Biol. Chem.* **286,** 15155–15164 (2011).

41. Mayer, R. J., Chen, J.-T., Taira, K., Fierke, C. A. & Benkovic, S. J. Importance of a hydrophobic residue in binding and catalysis by dihydrofolate reductase. *Biochemistry* **83,** 7718–7720 (1986).

42. Waldman, A. D. B. *et al.* The use of a genetically engineered tryptophan to identify the movement of a domain of B. stearothermophilus lactate dehydrogenase with the process which limits the steady-state turnover of the enzyme. *Biochem. Biophys. Res. Commun.* **150,** 752–759 (1988).

43. Ferrer, S., Tuñón, I., Moliner, V. & Williams, I. H. Theoretical site-directed mutagenesis: Asp168Ala mutant of lactate dehydrogenase. *J. R. Soc. Interface* **5 Suppl 3,** S217-24 (2008).

44. Howell, E. E., Villafranca, J. E., Warren, M. S., Oatley, S. J. & Kraut, J. Functional Role of Aspartic Acid-27 in Dihydrofolate Reductase Revealed by Mutagenesis. *Science.* **231,** 1123–1128 (1986).

45. Lee, Y., Mehra-Chaudhary, R., Furdui, C. & Beamer, L. J. Identification of an essential active-site residue in the α-D-phosphohexomutase enzyme superfamily. *FEBS J.* **280,** 2622–2632 (2013).

46. Isobe, M., Ichikawa, Y., Kitamura, M. & Goto, T. Synthetic Studies Toward Maytansinoids: Preparation of the Optically Active Intermediates from D-Mannose. *Chem. Lett.* 457–460 (1981).

47. Oscarson, S. Syntheses of the octyl and tetradecyl glycosides A new way for 2,4-di-O-protection of mannopyranosides. *Carbohydr. Res.* **241,** 323–328 (1993).

48. Marwood, R. D., Riley, A. M., Jenkins, D. J. & Potter, B. V. L. Synthesis of adenophostin A and congeners modified at glucose. *J. Chem. Soc. Perkin Trans. 1* **3,** 1935–1947 (2000).

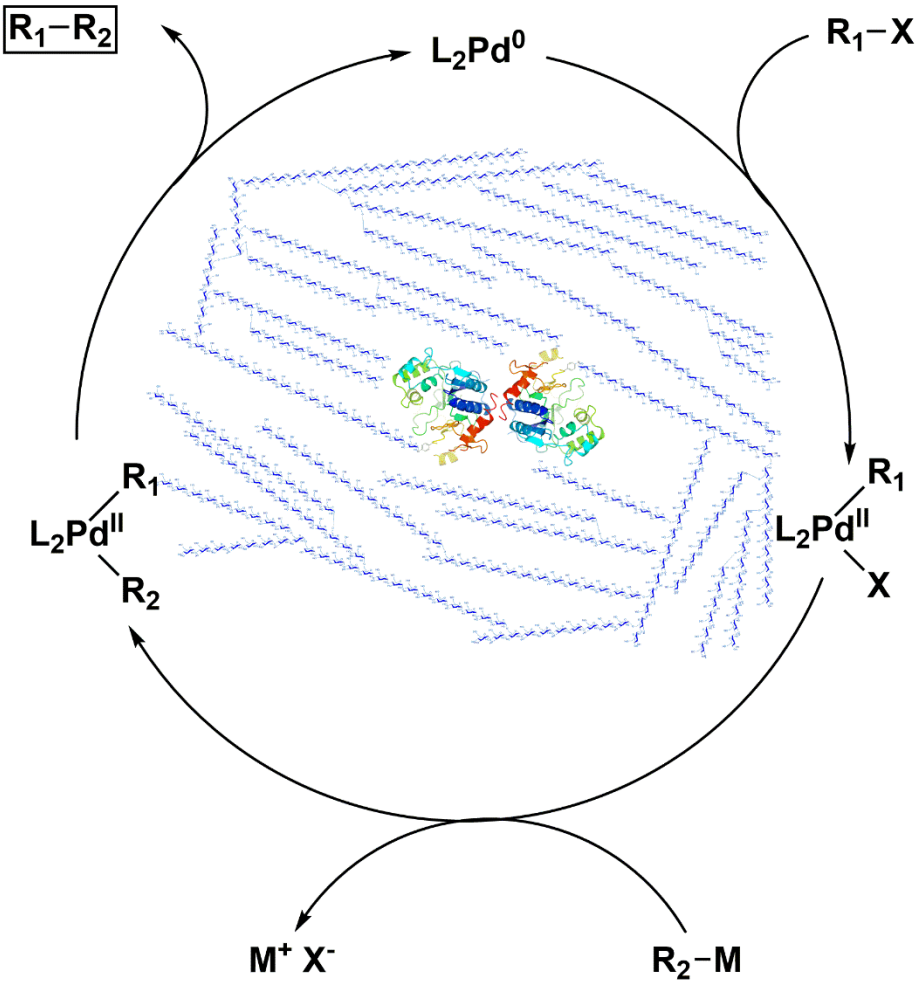
49. Geoghegan, K. F. *et al.* Spontaneous alpha-N-6-phosphogluconoylation of a 'His tag' in Escherichia coli: the cause of extra mass of 258 or 178 Da in fusion proteins. *Anal. Biochem.* **267,** 169–184 (1999).

50. Hemberg, G. *Investigating the Autocatalytic Priming Mechanism in Glycogenin*. (Univesity of Oxford, 2013).

51. Spicer, C. D. & Davis, B. G. Palladium-mediated site-selective Suzuki-Miyaura protein modification at genetically encoded aryl halides. *Chem. Commun. (Camb).* **47,** 1698–700 (2011).

52. Antos, J. M. & Francis, M. B. Transition metal catalyzed methods for site-selective protein modification. *Curr. Opin. Chem. Biol.* **10,** 253–262 (2006).

Chapter 7



Conclusions and Perspectives

Contents

7.1 Summary of Findings..... 432

7.2 Discussion..... 435

7.3 Future Perspectives..... 438

7.4 References..... 442

7.1 Summary of Findings

Glycogenin plays a fundamental role in glucose, and hence energy, storage, initiating the biosynthesis of glycogen via stepwise autoglucosylation from a tyrosine acceptor (Tyr195 in the GYG-1 isoform studied here).[1] Prior mechanistic studies have been hindered by an inability to isolate homogeneous GYG glycoforms, precluding analysis of individual catalytic steps and rendering distinct kinetic phases challenging to observe.[2–4] The fundamental goal of this thesis was to develop a novel, chemical modification approach to access such glycoforms and to utilise these in kinetic-based studies to enhance our understanding of the catalytic mechanism. This was accomplished through Suzuki-Miyaura cross-coupling (SMC) to the mutant GYG-Y195*p*IPhe (**figure 7.1**), enabling direct access to different homogeneously glycosylated GYG species and dissection of the catalytic mechanism with greater clarity.

The starting point for this approach lay in proof-of-concept studies demonstrating the capacity of GYG-Y195*p*IPhe to undergo cross-coupling to simple hydroxylated arylboronic acids (chapter 2). The resultant enzymes generated, representing "simple" mimics of glucosylated GYG (**figure 7.1, left**), were not active. However, the optimised conditions developed, enabling high conversions whilst remaining intrinsically compatible with retention of enzymatic activity and structure, formed the basis for subsequent work.

Subsequently, we were able to successfully generate a range of homogeneous GYG glycoconjugates via SMC of GYG-Y195*p*IPhe to carbohydrate-derived boronic acids (**figure 7.1, right**). Some of these glycoforms mimicked catalytic intermediates at different along the natural autoglucosylation pathway (chapters 3-5); others represented non-natural variants bearing non-glucose sugars (chapter 6). Crucially, all of these species exhibited catalytic activity towards autoglucosylation. Such defined glycoforms, isolated here through site-specific chemical ligation, are not accessible through more conventional biological methods.[2,3]

Comparative kinetic analyses of GYG glycoforms mimicking dautoglucosylation intermediates at different stages along the catalytic pathway implicated an unusual triphasic "slow-fast-slow" mechanism for GYG catalysis, the phases being: < 2 Glc; 2→4/5 Glc; > 4/5 Glc (**figure 7.2**). Observation of all three phases was a direct consequence of our ability to isolate and compare different GYG glycoforms falling within each phase, never previously possible. Notably, detection of the rapid, intermediate phase was mediated through circumvention of the slow Tyr195 glucosylation via chemical pre-installation of the initial sugars, not possible in prior kinetic studies.[2]

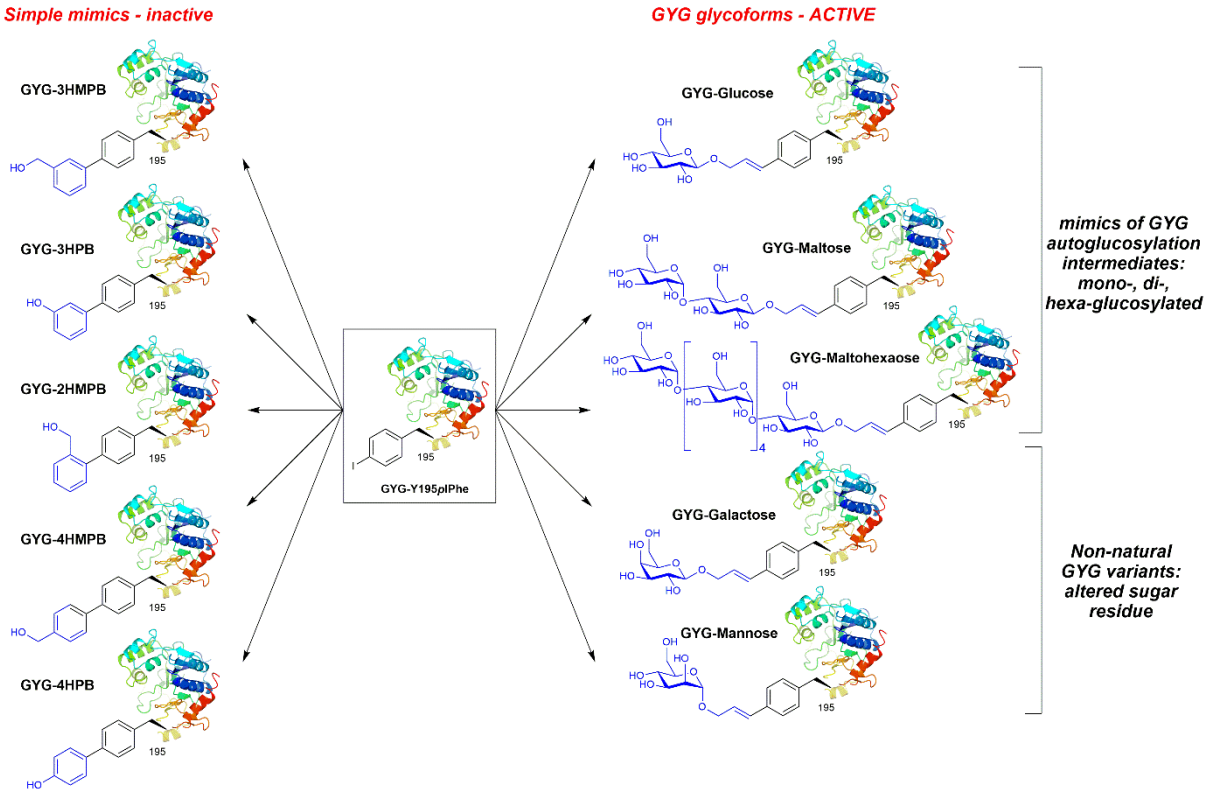


Figure 7.1 Summary of glycosylated GYG mimics accessed during this work. Crude mimics of glucosylated GYG, generated via cross-coupling of hydroxylated arylboronic acids to GYG-Y195*p*IPhe, were catalytically inactive (left). However, cross-coupling to various carbohydrate boronic acids enabled access to active GYG species (right). Both mimetics of intermediates along the autoglucosylation pathway, as well as non-natural variants bearing pendant non-glucose acceptors, could be accessed.

We showed that GYG exhibits striking plasticity, catalysing autogalactosylation, and glucosylation of non-glucose acceptors in addition to its natural glucosylation activity, although autogalactosylation was notably curtailed in extent. Study of such non-natural pathways led to the realisation that the three phases identified could be differentiated not only in terms of their relative rates but also in terms of such substrate plasticity, especially with respect to "Gal-Gal transfer" (**figure 7.2**). The intermediate, rapid phase exhibited markedly greater tolerance in this latter regard than the slower outer phases flanking it.

Finally, we applied our methodology to the investigation of a catalytically impaired, pathologically relevant mutant, GYG-T83M, implicated in glycogen storage disease XV.[5] Comparison of monosaccharide and disaccharide glycoforms of this mutant, uniquely accessible through our methodology, further highlighted differences between the first two phases illustrated in **figure 7.2**. Consistent with the known inactivity of the non-glucosylated mutant,[4,5] the mono-glucosylated variant (phase 1) was inactive towards autoglucosylation, yet the di-glucosylated species (phase 2) exhibited limited activity. This serves to

suggest that the inactivity of this mutant, rather than being global as previously assumed,[4] is in fact glycan-length-dependent, highlighting how our methodology might be used to understand disease.

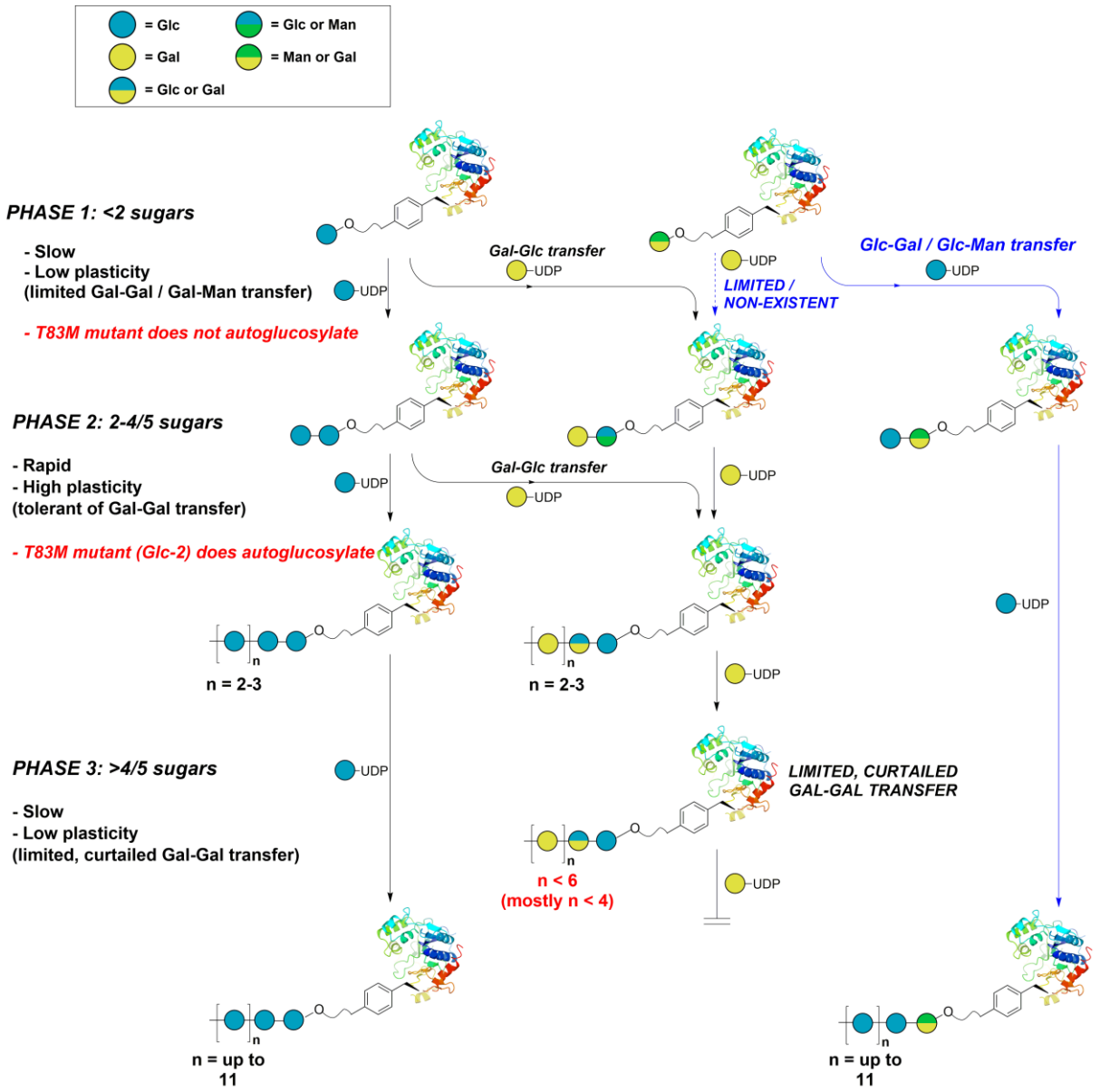


Figure 7.2 Summary of mechanism for GYG catalysis (sugars and linker shown schematically). This is similar to figure 6.12. In addition to natural autoglucosylation pathway (left), GYG may autogalactosylate or glucosylate from non-glucose acceptors. Three distinct phases – <2 Glc, 2→4/5 Glc, >4/5 Glc – are differentiated through rates (slow-fast-slow) and plasticity (particularly with regards to their tolerance for Gal-Gal transfer). Activity of T83M mutants (discussed below) also provide additional distinction of the early stages.

7.2 Discussion

The key contributions of this work can broadly be divided into two areas. Firstly, the methodology by which the GYG glycoforms were generated is of significant inherent interest in its own right. Yet of greater importance is the application of this methodology as a powerful tool for probing the mechanism of GYG catalysis and thus glycogen biogenesis. Implications within both areas, along with further interpretation of our mechanistic findings (**figure 7.2**) are discussed below.

As outlined in chapter 1, a number of site-specific chemical modification-based approaches to generate (homogeneous) glycoproteins exist.[6] Whilst SMC has previously been employed in this regard,[7] its application to this area has, to the best of our knowledge, been somewhat limited. This work more comprehensively highlights the evident potential of this method as a tool for chemical glycoprotein synthesis. This may pave the way for SMC to be utilised in the generation and detailed functional study of other biologically relevant glycoproteins. Ultimately, this may even be applicable to *in vivo* studies, given increasing demonstrations of Pd-chemistry *in cellulo* and *in vivo*.[8–10]

Furthermore, the conversion of catalytically *inactive* GYG-Y195*p*IPhe to *active* GYG-glycoform products represents a novel example of "chemical rescue". As discussed in chapter 3, whilst prior examples of such chemical modification-based enzyme activation exist – even involving Pd-chemistry[11] – these almost entirely involve deprotection of a "masked" residue,[11–17] most frequently of a photocaged residue upon light exposure.[12–14,16] Our approach, in contrast, involves a ligation-based activation of GYG-Y195*p*IPhe (i.e. cross-coupling); we believe such an approach to be essentially unprecedented. Such ligation-dependent activation may prove applicable to mechanistic study of other enzyme systems beyond GYG. In principle, any enzyme having a key activating residue or modification that could be installed chemically, by SMC or perhaps other means,[18] could represent a suitable system.

The key aim of this work, however, was the elucidation of the mechanism of GYG catalysis using the methodology developed. The following paragraphs provide further discussion of our findings in this area, summarised above (**figure 7.2**), to build on that within the preceding three chapters.

In apparent contrast to the triphasic mechanism proposed from this work, prior literature previously gave indications of a simpler biphasic process.[1,2,5] As outlined in chapter 4, previous kinetic data for aglycone GYG-wt, whilst hindered by a slow Tyr195 initial glucosylation obscuring subsequent steps, indicated one phase transition at 4-6 glucoses based on accumulation of glycoforms around this point.[2] This biphasic profile was then supported through crystallographic data; as discussed in detail in chapter 1, a mechanistic

"switch" from intra- to inter-monomeric mechanism was reported at 4-6 glucoses,[5] based on differing orientations of the GYG-bound maltosaccharide chain in crystal structures of GYG-wt dimers bearing 4 and 6 glucoses respectively.

Notably, our proposed triphasic profile, whilst novel, serves instead to build upon prior biphasic suggestions rather than refute them. Indeed, this serves to indicate that our "mimetic" system well-represents the natural version; for a full discussion/justification of this statement and of other supporting factors, please see chapter 4 section 4.3.4, and the recap in section 7.3. Thus, in the present work, the transition between 2nd and 3rd phases occurs ca. 4-5 glucoses (**figure 7.2**), which is is entirely consistent with the literature transition at 4-6 Glc,[2,5] though never previously defined through unambiguous kinetics However, our ability to access different early GYG glycoforms enabled us to detect the apparent existence of an additional, earlier phase transition (1st→2nd phases, 2 glucoses).[2,5] Specifically, chemical access to GYG-Glucose and GYG-Maltose enabled both observation of a slow 1st phase (GYG-Glucose) and, crucially, its circumvention allowing observation of the subsequent rapid 2nd phase (GYG-Maltose). The slow 1st phase is consistent with the known slow nature of the Tyr195 glucosylation,[2] yet the rapid 2nd phase – and hence any transition – was previously kinetically obscured precisely by this slow preceding phase (and not detectable crystallographically owing to inaccessibility of requisite crystal structures).

This highlights the power of chemical modification as a tool to enhance our understanding of these crucial early steps of glycogen biosynthesis, enabling visualisation of new aspects to the mechanism. Extension of the developed methodology could enable access to still further well-defined glycogen structures, facilitating still further mechanistic insight. Furthermore, our preliminary work on the pathologically relevant GYG-T83M system hints at the fact that this methodology may shine further light on the increasing body of glycogen-associated diseases.[19,20]

Our chemical modification approach also proved valuable in revealing non-natural autogalactosylation / "glucosylation of non-glucose acceptor" pathways for GYG (**figure 7.2**); defined non-glucose GYG acceptors are not accessible through other means. Such striking plasticity, to different degrees in different phases (**figure 7.2**), raises the question of whether this has any *in vivo* implications; in particular, as discussed in chapters 4 and 6, autogalactosylation would be feasible.[21] It is possible, though purely speculative, that the limited galactosylation of the final phase might be of importance given that later glucose residues, being further from Tyr195, may be more sterically accessible to e.g. Glycogen Synthase or Glycogen Phosphorylase, rendering it crucial that they are indeed glucose. It is admittedly less clear what role the same limited tolerance in the first phase might play. On the other hand, the ability to utilise

UDP-Galactose at all, e.g. in the more tolerant second phase where residues may be less accessible to downstream enzymes, may be advantageous if UDP-Glucose is urgently required for other cellular pathways.

Why though, mechanistically, might GYG catalysis be triphasic? We note first that the concentration of GYG in our assays is 5.4 μM, well in excess of the dimer K_D of 0.6-0.8 μM,[22] thus any interpretation of the above mechanism should invoke GYG dimers.[i] Dimeric GYG could undergo either intra-monomeric, inter-monomeric, and possibly inter-dimeric glucosylation, as discussed fully in chapter 1 (section 1.2.2). The practical existence of the first two of these is well-established.[3,5,22,23] Prior kinetic data and the lack of higher oligomers in solution at far higher concentrations than used here, suggests that inter-dimeric glucosylation is less likely play a significant role.[23,24] This said, its existence in early phases of catalysis has been occasionally hypothesised;[25] presumably requisite tetramers would have to form transiently.

Based on alignment with prior literature (see also chapter 1, section 1.2.2), the similar 1st and 3rd phases (slow, low plasticity), and the strikingly different 2nd phase (rapid, high plasticity), we speculate that the triphasic profile might reflect transitions between inter-monomer and intra-monomer phases, in an "inter→intra→inter" profile (**figure 7.3**). Prior crystallographic data for GYG dimers, as noted previously, highlighted an intra→inter "switch" occurring at oligosaccharide lengths of the 2nd→3rd phase transition in our system,[5] supporting the latter segment of this profile. The striking similarities of the 3rd and 1st phases suggest common mechanisms, thus we hypothesise that the 1st phase may also be inter-monomeric. Ongoing computational studies also suggest this may be the case.[26] This would be consistent with prior experiments on mutant GYG heterodimers capable of only inter-monomeric or intra-monomeric glucosylation (chapter 1, section 1.2.2), which imply that only inter-monomeric glucosylation is capable of **initiating** GYG catalysis within a dimer.[3] It would be similarly consistent with our observation that GYG-T83M-Glucose (1st phase), whilst incapable of autoglucosylation, may be glucosylated in an inter-enzyme sense. Nonetheless, we emphasise that this profile is speculative; suggestions for further exploration are provided later.

Such mechanistic transitions can be mediated through a known "flipping" of Tyr195.[5] Whilst intra-monomer is possible – and likely entropically favoured - during the 2nd phase, an increased steric hindrance upon coiling longer oligosaccharides into the active site of the same monomer can rationalise the intra→inter switch at 2nd→3rd phases.[5] It is less obvious why an inter-monomeric mechanism would also be favoured at very short oligosaccharide lengths; in crystal structures,

[i] Note that this is true for the majority of prior literature studies on solution-phase GYG. E.g. refs 2-4, 23.

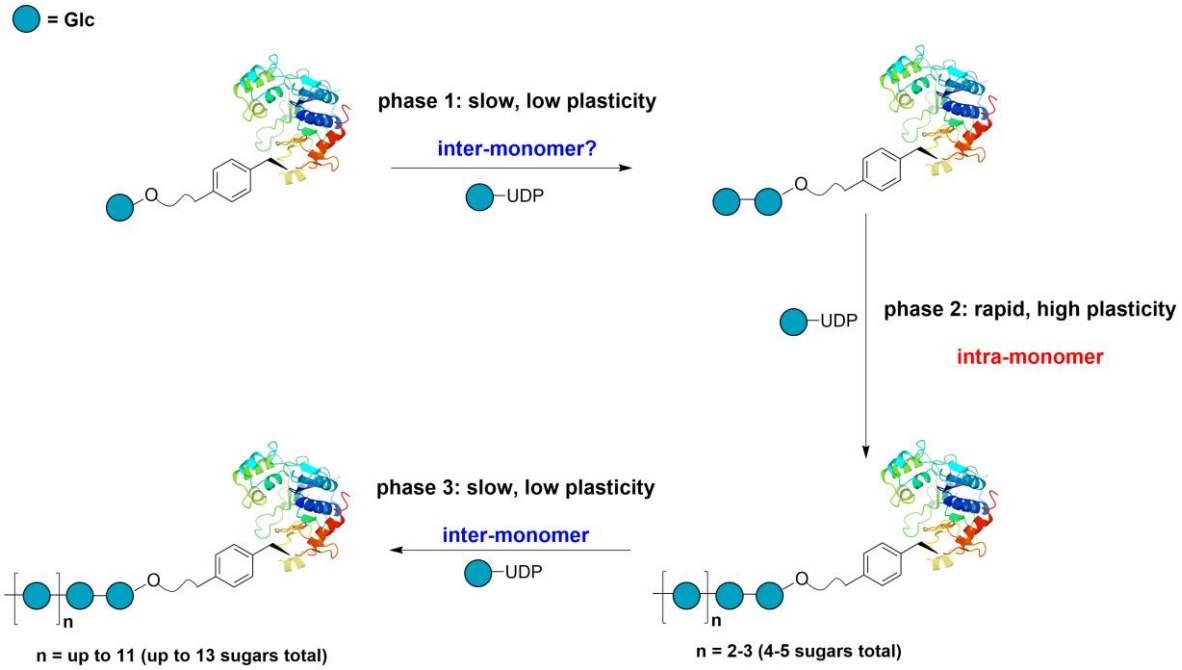


Figure 7.3 Similarities of the outer phases within our triphasic profile, their striking differences to the intermediate 2nd phase, and alignment with prior literature[2,3,5] lead us to hypothesise that such a triphasic profile may reflect transitions between inter- and intra-monomeric glucosylation mechanisms.

the initial Tyr195 is far from the active site of either monomer within a dimer.[5,25] This implies significant dynamic motion of the enzyme is required in these early phases, so perhaps the requisite large motion for intra-monomeric catalysis is simply not possible on steric grounds. We anticipate ongoing computational work will afford insight.[26] We cannot rule out alternative possibilities, however; for example, this phase may instead be inter-*dimer* (although as noted above prior data suggests this is unlikely[24,25]), or perhaps the aforementioned large distances from acceptor to active site render neither intra- nor inter-modes particularly favoured, accounting for the slow rate.

7.3 Future Perspectives

This work has developed a new tool for the study of GYG catalysis, and applied it to reveal previously unobserved catalytic phases, leading to the proposal of a novel triphasic mechanism that may reflect inter- and intra-monomeric mechanisms. However, there undoubtedly remain further avenues for exploration to address unanswered questions and limitations of the current work. Examples are provided below.

1) Exactly how good is the mimicry of the natural system?

One question arising from this research concerns exactly how well the GYG system utilised in this work mimics the corresponding natural one. There are admittedly two points of divergence: 1) there exist small

quantities (or larger, for GYG-Maltohexaose – see later) of cross-coupling side-products in our assays; 2) there is a non-natural linker between protein and carbohydrate in our system that does not exist in natural GYG glycoforms. We strongly believe these factors to have minimal effect on our results or conclusions, and thus to translate to the natural system, for a number of reasons; these were discussed in detail previously in chapter 4, section 4.3.4, hence they are only briefly recapped here. Nonetheless, there exist ways in which these aspects might be further investigated or indeed eliminated, as discussed below.

a) ***The presence of cross-coupling side-products***

The cross-coupling side-products, present at low levels with the exception of GYG-Maltohexaose, do not represent substrates for glucosylation. As a brief recap of chapter 4 (see section 4.3.4 for full detail), our reasons for believing their effect is negligible were: consistency of data cross replicates despite differing cross-coupling levels; lack of effect of spiking in extra GYG-Y195pIPhe side-product; lack of any theoretical reason for any rate difference of cross-coupled enzyme glucosylation when dimerised with itself or with side-product; general alignment with prior literature.[2,5] The last point, as discussed above (p. 435) and in chapter 4, refers to the fact that, our triphasic profile fits with and builds upon prior literature for GYG-wt, rather than contradicting it, implying our system is representative. For example, the aforementioned coincidence between the 2nd-3rd phase transition in our work and similar transitions implicated in prior crystallographic and kinetic data. An additional important reason not discussed previously, only becoming evident in light of all results, is the strong "internal consistency" of our results; e.g. in spite of differing side-product levels, GYG-Glucose, GYG-Maltose and GYG-Maltohexaose gave results consistent with each other. For example, the slow glucosylation of GYG-Maltohexaose aligns well with the slow final phase seen for GYG-Maltose, whilst the fast glucosylation of GYG-Maltose is reflected in its lack of accumulation during GYG-Glucose assays.

Nonetheless, elimination of such side-products would undoubtedly improve our system and also provide a means to further confirm our belief that they exert little effect. Further optimisation of the cross-coupling procedure to try to attain full conversions, at least for mono- and di-saccharide systems for which current conversions are already >75%, would be a logical direction. Such optimisation might include the development of further Pd catalysts for on-protein SMC, or perhaps the screening of alternative boron reagents to boronic acids, such as triol borates or trifluoroborates.[27,28]

b) *The presence of a non-natural "linker"*

The existence of a small, non-natural linker is necessitated by the chemical modification approach employed here. Again, we believe the effect of this to be minimal, as also discussed in detail in chapter 4, section 4.3.4. To briefly recap, the key reason is the fact that it is a common feature of all GYG glycoforms generated in this work, and that our methodology and conclusions rely on comparisons of these. Additionally, as with the side-products issue, alignment of aspects of our system with prior literature for GYG-wt[2,5] suggest our system is representative in spite of such a linker (chapter 4). However, any further empirical confirmation of the above would represent a means to enhance this work.

This is a challenging area to address; any chemical modification method will necessitate some sort of linker. One option might be to genetically incorporate a mono-glucosylated tyrosine residue into GYG through amber codon suppression (see chapter 1).[29] Whilst methods to accomplish such incorporation of glycosylated amino acids in *in vivo* expression systems remain unestablished, (chapter 1, section 1.3.3), it has proven possible, albeit challenging, to incorporate them using *in vitro* transcription-translation systems using chemically synthesised glycosyl-tRNA species.[30,31] Furthermore, GYG has previously been expressed using such cell-free systems.[32,33] A mono-glucosylated tyrosine residue incorporated in this fashion need not have any such linker and could be compared to our GYG-Glucose enzyme generated via SMC to evaluate any linker effect. Aside from this, exploration of genetic incorporation of glycosylated amino acids into GYG would be of inherent interest in its own right.

2) Optimisation of oligosaccharide cross-coupling

A limitation of this work is that the current cross-coupling conditions are far less effective for couplings between GYG and longer oligosaccharide boronic acids (chapter 5). Whilst kinetic studies may, and were, still be carried out on what GYG-oligosaccharide conjugate can be obtained, these are rendered more challenging owing to significant quantities of side-products (principally, unreacted GYG-Y195*p*IPhe). Whilst our data indicated that these didn't significant affect results (also see above), they may "swamp" LC-MS spectra. We anticipate that exploration of novel catalysts utilising new ligands might prove a fruitful avenue to explore to optimise these more challenging cross-couplings.

3) What is the regioselectivity of the linkages for galactosylation?

Whilst GYG glucosylation is known to involve formation of α(1→4) linkages,[1] the regioselectivity of the galactosylation reaction is not known. Elucidation of this was beyond the scope of this work, however it would be of interest to investigate this. One potential approach might involve NMR analysis of the

enzyme after autogalactosylation utilising ^{13}C-labelled UDP-Galactose. Regioselectivity of glycosidic linkages might be determined by, for example, HMBC analysis, possibly following e.g. tryptic digest to yield more manageable peptide fragments for NMR as opposed to the entire protein. As an alternative, crystallography of the enzyme (see below) could yield such information, although NMR might prove more facile given the lower enzyme concentrations likely required.

4) Further Evaluation of Inter- and Intra-Monomer Mechanisms

We suggested above that the triphasic profile of GYG catalysis may reflect transitions between intra- and inter-monomeric glucosylation mechanisms, but this is largely speculative (although supported to an extent by prior literature as discussed[3,5]). Our previous evaluation of GYG-T83M mutants was motivated partly by pathological relevance of this mutation, but an additional goal was to use such mutants – which prior literature suggested could only be glucosylated in an inter-monomeric fashion by a separate "donor" monome[3] – to investigate occurrence of inter-subunit glucosylation in different catalytic phases. The unexpected autocatalytic activity of GYG-T83M-Maltose restricted this (since this, being not known, could be inter- or intra-subunit).

A couple of approaches may shed further light. Firstly, alternative mutations of e.g. GYG-Glucose, GYG-Maltose – such as of Asp160 or Asp163, known to be critical to GYG catalysis[34] – might prove more fruitful.[3] However, a more powerful approach could well be to use crystallography. Whilst structures of GYG-maltotetraose and GYG-maltohexaose exist[5], no structures for earlier or later glycoforms have been accessible, and structures of e.g. GYG-Glucose, GYG-Maltose, may provide significant mechanistic insight. A related tactic, currently ongoing,[26] would be to computationally model and evaluate, based on prior crystal structures[5], some of the GYG species utilised in this work, to provide insight regarding the energy barriers for different mechanisms at different stages.

5) Do GYG glycoforms from non-natural pathways still interact with GYS?

An area not explored in this thesis concerns the interaction of GYG with the downstream enzyme Glycogen Synthase (GYS). It would be particularly interesting to ask whether GYG variants incorporating non-glucose moieties (e.g. galactosylated) can serve as substrates for further oligosaccharide extension by GYS. It should be noted that the work in this thesis utilised the catalytic domain of GYG-1, consistent with prior literature,[3,5,22,34] however this lacks the C-terminal domain that interacts with GYS.[35] Thus, it would be necessary to first translate the developed methodology to the full-length variant of GYG-1, generating glycoforms thereof, before investigating such downstream interactions. The ability of GYS to

extend GYG-bound oligosaccharides could be assayed through incubation with the requisite GYG partner and UDP-Glucose, followed by gel electrophoresis and PAS staining to detect the polysaccharide species present.[35] GYS activity would result in consumption of the band for the initial GYG-bound oligosaccharide, and its replacement with a far higher Mw band corresponding to glycogen.[35]

6) How does GYG-2 compare to GYG-1

As noted in chapter 1, GYG exists as two isoforms in higher primates: GYG-1, found mostly in muscle, and GYG-2, found more in liver.[36] This work focussed on GYG-1, but did not address GYG-2, which, whilst broadly similar, has some intriguing differences; notably, it exhibits reduced glucosylation with respect to GYG-1.[33] It would be interesting to translate the developed methodology to the generally lesser-studied GYG-2 isoform to gain a better understanding of the differences and similarities between these isoforms. This would therefore involve generation of "GYG2-Y228*p*IPhe" (Y228 being the catalytic tyrosine[33]), cross-coupling to sugars and evaluation of their activity, compared to analogous GYG-1 variants.

The present work has provided useful insight into the catalytic mechanism of this complex autocatalytic enzyme at the heart of carbohydrate mechanism. It is hoped that exploration of some of the above suggested areas may contribute towards a still fuller understanding of this area.

7.4 References

1. Roach, P. J., Depaoli-Roach, A. A., Hurley, T. D. & Tagliabracci, V. S. Glycogen and its metabolism: some new developments and old themes. *Biochem. J.* **441,** 763–787 (2012).

2. Hurley, T. D., Walls, C., Bennett, J. R., Roach, P. J. & Wang, M. Direct detection of glycogenin reaction products during glycogen initiation. *Biochem Biophys Res Commun.* **348,** 374–378 (2006).

3. Issoglio, F. M., Carrizo, M. E., Romero, J. M. & Curtino, J. A. Mechanisms of monomeric and dimeric glycogenin autoglucosylation. *J. Biol. Chem.* **287,** 1955–1961 (2012).

4. Carrizo, M. E., Romero, J. M., Issoglio, F. M. & Curtino, J. A. Structural and biochemical insight into glycogenin inactivation by the glycogenosis-causing T82M mutation. *FEBS Lett.* **586,** 254–257 (2012).

5. Chaikuad, A. *et al.* Conformational plasticity of glycogenin and its maltosaccharide substrate during glycogen biogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 21028–21033 (2011).

6. Gamblin, D. P., Scanlan, E. M. & Davis, B. G. Glycoprotein Synthesis : An Update. *Chem. Rev.* **109,** 131–163 (2009).

7. Spicer, C. D. & Davis, B. G. Rewriting the bacterial glycocalyx via Suzuki-Miyaura cross-coupling. *Chem. Commun. (Camb).* **49,** 2747–2749 (2013).

8. Yusop, R. M., Unciti-Broceta, A., Johansson, E. M., Sanchez-Martin, R. M. & Bradley, M. Palladium-mediated intracellular chemistry. *Nat Chem* **3,** 239–243 (2011).

9. Miller, M. A. *et al.* Nano-palladium is a cellular catalyst for in vivo chemistry. *Nat. Commun.* **8,** 15906 (2017).

10. Li, N., Ramil, C. P., Lim, R. K. V. & Lin, Q. A Genetically Encoded Alkyne Directs Palladium-Mediated Protein Labeling on Live Mammalian Cell Surface. *ACS Chem. Biol.* **10,** 379–384 (2015).

11. Li, J. *et al.* Palladium-triggered deprotection chemistry for protein activation in living cells. *Nat. Chem.* **6,** 352–361 (2014).

12. Arbely, E., Torres-Kolbus, J., Deiters, A. & Chin, J. W. Photocontrol of tyrosine phosphorylation in mammalian cells via genetic encoding of photocaged tyrosine. *J. Am. Chem. Soc.* **134,** 11912–11915 (2012).

13. Deiters, A., Groff, D., Ryu, Y., Xie, J. & Schultz, P. G. A genetically encoded photocaged tyrosine. *Angew. Chemie - Int. Ed.* **45,** 2728–2731 (2006).

14. Gautier, A., Deiters, A. & Chin, J. W. Light-activated kinases enable temporal dissection of signaling networks in living cells. *J. Am. Chem. Soc.* **133,** 2124–2127 (2011).

15. Delacour, Q. *et al.* Light-Activated Proteolysis for the Spatiotemporal Control of Proteins. *ACS Chem. Biol.* **10**, 1643-1647 (2015).

16. Lemke, E. A., Summerer, D., Geierstanger, B. H., Brittain, S. M. & Schultz, P. G. Control of protein phosphorylation with a genetically encoded photocaged amino acid. *Nat. Chem. Biol.* **3,** 769–772 (2007).

17. Li, J., Jia, S. & Chen, P. R. Diels-Alder reaction–triggered bioorthogonal protein decaging in living cells. *Nat. Chem. Biol.* **10,** 1003–1005 (2014).

18. Wright, T. H. *et al.* Posttranslational mutagenesis: A chemical strategy for exploring protein side-chain diversity. *Science.* **354,** aag1465 (2016).

19. Adeva-Andany, M. M., González-Lucán, M., Donapetry-García, C., Fernández-Fernández, C. & Ameneiros-Rodríguez, E. Glycogen metabolism in humans. *BBA Clin.* **5,** 85–100 (2016).

20. Zois, C. E., Favaro, E. & Harris, A. L. Glycogen metabolism in cancer. *Biochem. Pharmacol.* **92,** 3–11 (2014).

21. Frey, P. A. The Leloir pathway: a mechanistic imperative for three enzymes to change the stereochemical configuration of a single carbon in galactose. *FASEB J.* **10,** 461–470 (1996).

22. Bazán, S., Issoglio, F. M., Carrizo, M. E. & Curtino, J. A. The intramolecular autoglucosylation of monomeric glycogenin. *Biochem. Biophys. Res. Commun.* **371,** 328–332 (2008).

23. Lin, A., Mu, J., Yang, J. & Roach, P. J. Self-glucosylation of glycogenin, the initiator of glycogen biosynthesis, involves an inter-subunit reaction. *Arch. Biochem. Biophys.* **363,** 163–170 (1999).

24. Cao, Y., Mahrenholz, A. M., DePaoli-Roach, A. A. & Roach, P. J. Characterization of rabbit skeletal muscle glycogenin: Tyrosine 194 is essential for function. *J. Biol. Chem.* **268,** 14687–14693 (1993).

25. Gibbons, B. J., Roach, P. J. & Hurley, T. D. Crystal structure of the autocatalytic initiator of glycogen biosynthesis, glycogenin. *J. Mol. Biol.* **319,** 463–477 (2002).

26. Bilyard, M. K. *et al.* Palladium-mediated enzyme activation suggests multiphase initiation of glycogenesis. *Manuscript under revision (Nature).*

27. Yamamoto, Y., Takizawa, M., Yu, X.-Q. & Miyaura, N. Cyclic triolborates: air- and water-stable ate complexes of organoboronic acids. *Angew. Chem. Int. Ed. Engl.* **47,** 928–931 (2008).

28. Molander, G. A. & Felix, L. A. Stereoselective Suzuki-Miyaura cross-coupling reactions of potassium alkenyltrifluoroborates with alkenyl bromides. *J. Org. Chem.* **70,** 3950–3956 (2005).

29. Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C., Schultz, P. G. & April, I. A. General Method for Site-Specific Incorporation of Unnatural Amino Acids into Proteins. *Science.* **244,** 182–188 (1989).

30. Fahmi, N. E., Dedkova, L., Wang, B., Golovine, S. & Hecht, S. M. Site-Specific Incorporation of Glycosylated Serine and Tyrosine Derivatives into Proteins. *JACS* **129,** 3586–3597 (2007).

31. Matsubara, T., Iijima, K., Watanabe, T., Hohsaka, T. & Sato, T. Incorporation of glycosylated amino acid into protein by an in vitro translation system. *Bioorganic & Medicinal Chemistry Letters* **23,** 5634-5636 (2013).

32. Nilsson, J. *et al.* Molecular pathogenesis of a new glycogenosis caused by a glycogenin-1 mutation. *Biochim. Biophys. Acta* **1822,** 493–499 (2012).

33. Nilsson, J. *et al.* LC-MS/MS characterization of combined glycogenin-1 and glycogenin-2 enzymatic activities reveals their self-glucosylation preferences. *Biochim. Biophys. Acta* **1844,** 398–405 (2014).

34. Hurley, T. D., Stout, S., Miner, E., Zhou, J. & Roach, P. J. Requirements for catalysis in mammalian glycogenin. *J. Biol. Chem.* **280,** 23892–23899 (2005).

35. Zeqiraj, E. *et al.* Structural basis for the recruitment of glycogen synthase by glycogenin. *Proc. Natl. Acad. Sci. U. S. A.* **111,** E2831-2840 (2014).

36. Mu, J., Skurat, A. V & Roach, P. J. Glycogenin-2, a Novel Self-glucosylating Protein Involved in Liver Glycogen Biosynthesis. *J. Biol. Chem.* **272,** 27589–27597 (1997).

Appendix

Contents

1. General Procedures for Small Molecule Chemistry 446

2. General Procedures for Biological Work and Protein Modification 448

3. Protein LC-MS 449

3.1 Data Acquisition 449

3.2 Data Analysis 451

1. General Procedures for Small Molecule Chemistry

Chemicals and solvents were purchased from Sigma-Aldrich UK, Carbosynth, Acros UK, Alfa Aesar UK or Fisher UK and were used as received unless otherwise stated. All non-aqueous reactions were carried out in oven-dried or flame-dried glassware under an atmosphere of dry nitrogen or argon, as indicated. Room temperature (r.t) refers to ambient temperature. Temperatures of 0 °C were maintained using an ice-water batch. Pet Ether refers to the fraction of petroleum ether boiling in the range 40-60 °C. Deionized water was used in all chemical reactions.

Chromatography

Thin-layer chromatography was carried out using Merck aluminium-backed sheets coated with 60F254 silica gel. Compounds were visualised with either of the following stains: potassium permanganate; sulfuric acid; ceric ammonium molybdate (CAM). Flash column chromatography was performed using Geduran® silica 60 40-63 μm.

Melting Points and Optical Rotation

Melting points (m.p.) were recorded on a Leica Galen III hot stage microscope equipped with a Testo 720 temperature probe and are uncorrected. Optical rotation values were recorded on a Perkin-Elmer 241 polarimeter with a path length of 1.0 dm and are reported in units of. Samples were dissolved in either spectrophometric grade $CHCl_3$, deionized H_2O, or 0.5 M HCl in EtOH as indicated. The polarimeter was blanked with the corresponding solvent before use. Values represent mean averages from at least 8 successive readings.

IR Spectroscopy

Infrared (IR) spectra were recorded using a Bruker Tensor 27 Fourier Transform spectrophotometer. Absorption maxima are reported in wavenumbers (cm^{-1}); functional groups responsible for specific maxima are indicated. Solids were analysed via ATR in the solid state or as thin films obtained by evaporation of a solution. Oils were analysed as thin films.

Mass Spectrometry (Small Molecule)

Low-resolution mass spectra (LRMS) were recorded using an Agilent 6120 Quadrupole spectrometry using electrospray ionisation (ESI) in either positive and/or negative mode as indicated. High-resolution mass spectra (HRMS) were recorded on a Bruker MicroTOF ESI mass spectrometer using a lock-spray source.

^{1}H NMR Spectroscopy

Spectra were recorded on a Bruker AVIII HD 400 nanobay (400 MHz) or on a Bruker AVII 500 (500 MHz). Deuterated solvents were used as the lock. Spectra were analysed using MestReNova software. Chemical shifts (δ_H) are quoted in parts per million (ppm) to the nearest 0.01 ppm and are referenced to the centre of the residual non-deuterated solvent peak. Coupling constants (*J*) are reported in Hertz (Hz) to the nearest 0.1 Hz. Data are reported as follows: chemical shift, multiplicity (br = broad; s = singlet, d = doublet; t = triplet; q = quartet; qn = quintet; m = multiplet; or as a combination of these), coupling constant(s), integration and assignment. Spectra were assigned on the basis of a combination of the following: unambiguous chemical shift; 2D experiments (COSY, HSQC, HMBC, 2D TOCSY); 1D TOCSY; analogy to fully interpreted spectra of related compounds.

^{13}C Spectroscopy

Spectra were recorded on a Bruker AVIII 400 nanobay (101 MHz) or on a Bruker AVII 500 (126 MHz). Deuterated solvents were used as the lock. Spectra were analysed using MestReNova software. Chemical shifts (δ_C) are quoted in parts per million (ppm) to the nearest 0.1 ppm and are referenced to the centre of the residual non-deuterated solvent peak. Spectral assignment was based on chemical shift, DEPT editing, and, where appropriate, HMQC and HMBC experiments or by analogy to fully interpreted spectra of related compounds.

2. General Procedures for Biological Work and Protein Modification

All biological equipment and prepared biological solutions used in the handling of bacterial cultures were sterilized prior to use by autoclaving for 20 minutes at 121 °C in either a Falcon LTE or a Touchclave®-R tabletop autoclave. Biological work was carried out in a Kendro HERAsafe KS-12 Class II Laminar Flow Cabinet. Water was deionized and passed through a 0.22 μm Millipore filter before use. For growth of bacterial cultures, the following incubators were used: New Brunswick Scientific model G25, or New Brunswick Scientific Excella E25. Centrifugation of bacterial cultures and of cell lysates was carried out in a Beckmann Coulter Avanti-J25 centrifuge. JA10 and JA16.250 rotors were used for volumes of 200-500 ml, and a JA25.50 rotor was used for volumes of 25-50 ml.

SDS-PAGE

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was carried out using pre-cast Invitrogen Novex Bis-Tris gels (4-12 %). Samples were prepared by mixing with 5x SDS loading dye in a 4:1 sample:dye ratio, and were then heated at 90 °C for 5 minutes prior to gel loading. Perfect Protein™ markers were used as a reference. Gels were run at a voltage of 200V for 30-40 minutes, using 1x MES buffer. Proteins were visualised through staining with InstantBlue (Expedeon Ltd.) for ca. 60 minutes. Staining solution was then decanted and the gel washed with water (1 h then overnight) prior to scanning.

Protein Modification Reactions

All protein modification reactions were carried out in 0.5 ml Eppendorf® tubes. Water refers to MilliQ water. Protein concentrations were measured using the Bradford assay according to manufacturer protocol, with Bovine Serum Albumin (BSA) as a standard.

Circular Dichroism

Measurements were made at room temperature using a Chirascan CD-spectrophotometer (Applied Photophysics) using a 0.1 cm quartz cuvette. A background spectrum was recorded using 0.2 ml of appropriate buffer. Protein samples (0.2 ml) were run at concentrations of 0.2 mg/ml. Scans were run using

0.5 s per time point and 0.5 nm increments, scanning from 200nm to 260 nm. 3 repeats were carried out. For each protein sample measurement, the background spectrum was subtracted, then the runs averaged (mean) and the resultant data smoothed using the literature standard S.G algorithm (3 times).

3. Protein LC-MS

3.1 Data Acquisition

Liquid-chromatography mass spectrometry (LC-MS) was performed using the methods detailed below. Solvents A and B are water and acetonitrile respectively, both containing 0.1 % formic acid.

1) Waters LCT Classic coupled to a Shimadzu 20 Series HPLC using a Thermo Proswift (250 x 4.6 mm x 5 μm) column. Electrospray source parameters were as follows: capillary voltage 3000 V; cone voltage 25 V. Spectra were calibrated to at least 17 matched peaks of the multiply charged ion series of equine myoglobin run under equivalent conditions. The gradient program is shown below; a flow rate of 0.4 ml/min was used throughout.

Time (min)	% Solvent A	% Solvent B
0.0	95	5
1.0	95	5
7.0	5	95
11.0	5	95
12.0	95	5
17.0	95	5

2) Waters LCT Premier XE coupled to a Waters 1525 Micro HPLC using a Merck Chromolith C18 guard cartridge. Electrospray source parameters were as follows: capillary voltage 3000 V; cone voltage 100 V. Spectra were calibrated automatically. One of two gradient programs, shown below, were used, with a variable flow rate.

Gradient 1:

Time (min)	% Solvent A	% Solvent B	Flow rate (ml/min)
0.0	95	5	0.4
0.5	95	5	0.4
1.5	40	60	0.4
3.0	40	60	0.4
3.5	40	60	1.0
4.0	40	60	1.0
4.5	98	2	1.0

Gradient 2:

Time (min)	% Solvent A	% Solvent B	Flow rate (ml/min)
0.0	95	5	0.4
1.0	95	5	0.4
5.0	0	100	0.4
8.0	0	100	1.0
8.1	95	5	1.0
9.5	95	5	1.0

3) Agilent 6230 TOF MS coupled to a 1260 Infinity HPLC using a poroshell 5 μm 3000 B C3 column (21 x 75 mm), or using a Merck Chromolith C18 guard cartridge. Electrospray source parameters were as follows: capillary voltage 4000 V; nozzle voltage 1000 V. Spectra were calibrated through use of an internal lock-spray. The gradient program used is shown below.

Time (min)	% Solvent A	% Solvent B	Flow rate (ml/min)
0.0	95	5	0.4
0.5	95	5	0.4
1.5	40	60	0.4
3.0	40	60	0.4
3.5	40	60	0.35
4.0	40	60	0.35
4.5	2	98	0.35
6.5	95	5	0.4

4) Waters Xevo G2-QS QToF MS coupled to a Water Acquity UPLC using a **column**. Electrospray source parameters were as follows: capillary voltage 3000 C; cone voltage 20 V. Spectra were calibrated through use of an internal lock-spray. The gradient program is shown below; a flow rate of 0.3 ml/min was used throughout.

Time (min)	% Solvent A	% Solvent B
0.0	95	5
1.0	95	5
7.0	5	95
8.0	5	195
8.1	95	5
10.0	95	5

3.2 Data Analysis

Raw MS data containing multiply charged ion series were deconvoluted using the MaxEnt1 algorithm with either MassLynx (Waters) or MassHunter (Agilent) software. Spectra were deconvoluted between 20000 and 40000 Da according to manufacturer instructions.

Reaction conversions for cross-couplings were calculated from relative peak intensities of cross-coupled product, unreacted starting material and any species arising from side-reactions, in deconvoluted spectra. In glucosylation and galactosylation assays, proportions of each glycoform present were determined similarly from their relative peak intensities in deconvoluted MS spectra; these were in turn utilised to calculate average numbers of carbohydrate residues attached (per enzyme). Full details of the analytical procedures utilised for this are provided in the experimental sections of chapters 2 and 4.