

This electronic thesis or dissertation has been downloaded from the King's Research Portal at <https://kclpure.kcl.ac.uk/portal/>



Development of carbohydrate-based inhibitors and probes for the interrogation of the bacterial proteome

Metier, Camille

Awarding institution:
King's College London

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

END USER LICENCE AGREEMENT



Unless another licence is stated on the immediately following page this work is licensed

under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

licence. <https://creativecommons.org/licenses/by-nc-nd/4.0/>

You are free to copy, distribute and transmit the work

Under the following conditions:

- Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
- Non Commercial: You may not use this work for commercial purposes.
- No Derivative Works - You may not alter, transform, or build upon this work.

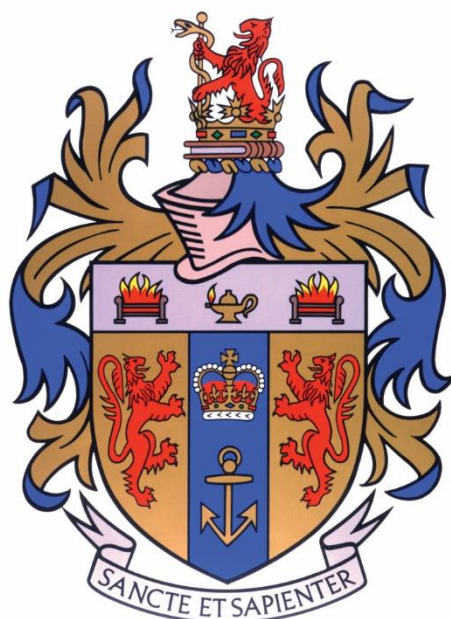
Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Development of carbohydrate-based inhibitors and probes for the interrogation of the bacterial proteome

Camille C. Métier



King's College London

Department of Chemistry

School of Natural and Mathematical Sciences

A thesis submitted in fulfilment of the requirements for the degree of 'Doctor in Philosophy' in Chemistry

September 2019

Table of Contents

Acknowledgements.....	7
Abstract.....	9
Abbreviations.....	11
1. General introduction.....	14
1.1. Antimicrobial resistance, current challenges	15
1.1.1. Examples of drug-resistant pathogens	16
1.1.1.1. <i>Neisseria meningitidis</i>	17
1.1.1.2. <i>Haemophilus influenzae</i>	17
1.2. Mechanisms of AMR.....	18
1.3. Anti-virulence strategies	19
1.4. Relationship between AMR and virulence	20
1.5. Glycoconjugates: LOS and LPS at the surface of Gram-negative bacteria	22
1.6. Glycosyltransferases (GTs).....	24
1.6.1. Mechanism of GTs.....	25
1.6.2. Classification according to amino acid sequence: the CAZy database	27
1.6.3. Leloir vs non-Leloir GTs.....	28
1.6.4. 3D-structures of GTs.....	29
1.7. LgtC from <i>Neisseria meningitidis</i>	33
1.7.1. LgtC and its role in virulence.....	33
1.7.2. Phase variation	35
1.7.3. Structure of LgtC from <i>Neisseria meningitidis</i> ⁵⁴	36
1.7.3.1. Donor binding site	38
1.7.3.2. Acceptor binding site	39
1.7.4. Catalytic mechanism of LgtC.....	41
1.7.5. Inhibitors of LgtC.....	42
1.7.5.1. Donor-like LgtC-inhibitors.....	42
1.7.5.2. Acceptor-like inhibitors of GalT and LgtC	45
1.7.5.3. A novel, non-substrate-related class of covalent inhibitors of LgtC....	46
1.8. Irreversible enzyme modification: covalent drugs	48
1.9. Covalent probes and applications.....	51
1.10. Aims and objectives.....	53
2. Disaccharide analogues as a scaffold for the covalent design of LgtC inhibitors	58
2.1. Introduction	59
2.1.1. Carbohydrates as a template for inhibitor design.....	59

2.1.2.	Targeted covalent inhibitors: design and carbohydrate-based examples	62
2.1.3.	Docking programmes for TCI design	64
2.1.4.	TCIs against drug resistance	65
2.2.	Objectives	67
2.3.	Structure-based design: Covalent docking	68
2.4.	Chemical synthesis of disaccharide chemical targets	71
2.4.1.	Mechanism of the Heyns Rearrangement	72
2.4.2.	Discussion of disaccharide chemical targets synthesis: cmpds 1 and 2	73
2.4.2.1.	Step (i): Addition of benzylamine to lactulose	74
2.4.2.2.	Step (ii): Heyns rearrangement	75
2.4.2.3.	Characterisation of cmpd 5 : 2-N-benzyl lactosamine	81
2.4.2.4.	Step (iii): debenylation by hydrogenation	83
2.4.2.5.	Step (iv): Chemical attachment of the electrophilic warhead	84
2.5.	Biochemical testing	84
2.5.1.	Phosphatase-coupled biochemical assay	85
2.5.2.	Evaluation of inhibition activity	86
2.5.3.	Evaluation of substrate activity	88
2.6.	Summary and conclusions	91
3.	From disaccharides to monosaccharides	93
3.1.	Introduction	94
3.1.1.	Kinetic model of covalent inhibition	94
3.1.2.	Electrophilic warheads (WH) for TCI design	95
3.1.2.1.	Epoxides	95
3.1.2.2.	Haloacetamides	96
3.1.2.3.	Michael acceptors	98
3.1.2.4.	Sulfonyl fluorides	98
3.1.3.	Reversible TCIs	99
3.1.4.	Monosaccharide substrates for LgtC	101
3.2.	Objectives	102
3.3.	Chemical synthesis	103
3.4.	Initial biochemical testing: inhibition and substrate activity	105
3.5.	Investigations into the mode of action of monosaccharide inhibitors	106
3.5.1.	Diafiltration	107
3.5.2.	Time dependant inhibition	108
3.5.3.	Mass spectrometry	109
3.6.	Investigation into the reactivity of the monosaccharides towards cysteine	110
3.7.	Reactivity model study	112

3.8.	Attempted synthesis of bromoacetamide monosaccharide analogue	115
3.9.	Substrate competition experiments	117
3.10.	Inhibition of non-electrophilic analogues	119
3.11.	Preliminary crystallography work	120
3.12.	Summary and conclusions	122
4.	Carbohydrate-based fluorescent probes for the labelling of recombinant LgtC	125
4.1.	Introduction	126
4.1.1.	Covalent probes for glycosyl hydrolases (GHs).....	127
4.1.2.	Covalent probes for glycosyltransferases.....	131
4.1.3.	Copper-Catalyzed Azide–Alkyne Cycloaddition (CuAAC).....	133
4.2.	Objectives	135
4.3.	Probe design.....	136
4.4.	Chemical synthesis of fluorescent probes	137
4.5.	Structural and spectroscopic characterisation of probes.....	142
4.5.1.	Characterisation of 13	142
4.5.2.	Identification of α - β anomers	143
4.5.3.	UV-Fluorescence spectroscopy	145
4.6.	Labelling of recombinant LgtC.....	147
4.6.1.	Effect of probe concentration on labelling	148
4.6.2.	Effect of enzyme concentration on labelling.....	149
4.6.3.	Effect of incubation time on labelling	150
4.6.4.	Effect of reducing environment and labelling specificity.....	151
4.6.5.	Effect of pH on the labelling	152
4.6.6.	Competitive labelling.....	153
4.6.7.	Two-step vs one step labelling	157
4.7.	Labelling of LgtC mutants and attempted identification of target residue	159
4.8.	Labelling of LgtC in cell lysates and through live <i>E. coli</i> cells.....	161
4.8.1.	Labelling of LgtC in <i>E. coli</i> cell lysates	161
4.8.2.	Labelling of LgtC in live <i>E. coli</i> cells	163
4.9.	Further indications of the binding differences between probe and inhibitor .	168
4.10.	Summary and conclusions.....	169
5.	Carbohydrate-based fluorescent probes for bacterial proteomic studies on <i>H. influenzae</i> R2866.....	173
5.1.	Introduction	174
5.1.1.	Chemical proteomics and application to bacterial studies.....	174
5.1.2.	Carbohydrate binding proteins and their role in bacterial virulence.....	177
5.1.2.1.	Proteins with catalytic activity.....	177

5.1.2.2.	Proteins with no catalytic activity.....	179
5.1.3.	Carbohydrate-based probes for bacterial proteomic studies.....	181
5.1.4.	H. influenzae R2866 strain.....	181
5.2.	Objectives	182
5.3.	Chemical synthesis of galactosamine-based probes	183
5.4.	Structural insight into the sugar scaffold for both probes.....	186
5.5.	Chemical proteomic study: bacterial protein profiling	189
5.5.1.	Bacterial growth for H. influenzae R2866.....	190
5.5.2.	Through-cells protein labelling	191
5.5.2.1.	Effect of probe concentration on labelling profile	194
5.5.2.2.	Effect of incubation time on labelling profile	195
5.5.3.	Bacterial cell lysate labelling.....	196
5.5.4.	Whole cell versus cell lysates labelling.....	198
5.5.1.	Two-step labelling	200
5.5.2.	Mass spectrometry analysis	202
5.6.	Summary and conclusions	206
6.	Conclusions and outlook.....	209
7.	Material and methods	215
7.1.	Covalent docking	216
7.2.	Chemical synthesis: protocols and compound characterisation	216
7.2.1.	Chapter 2 compounds: disaccharides.....	217
7.2.2.	Chapter 3 compounds: monosaccharides	219
7.2.3.	Chapter 4 compounds: glucosamine-based probes and fluorophores.....	222
7.2.4.	Chapter 5 compounds: galactosamine-based probes	229
7.3.	Expression and purification of LgtC	237
7.4.	Biochemical colorimetric assay: general protocols.....	239
7.4.1.	General assay procedure.....	240
7.4.2.	Standard procedure for the collection and analysis data.....	240
7.4.3.	Activity assay.....	241
7.4.4.	Inhibition assay.....	242
7.4.5.	CIP control experiment.....	243
7.4.6.	Substrate assay.....	243
7.4.7.	Dialysis.....	244
7.4.8.	Substrate kinetics.....	245
7.4.9.	WH reactivity monitoring by ¹ H NMR.....	246
7.5.	Crystallography	246
7.6.	Monitoring of attachment of linker on cmpd 9 by HPLC	247

7.7. Fluorescence scanning.....	248
7.8. LgtC-labelling experiments	248
7.8.1. General procedure for labelling of recombinant LgtC.....	249
7.8.2. LC-MS/MS analysis for identification of target residue	251
7.8.3. Labelling of non-recombinant LgtC: through cells or in cell lysates.....	252
7.9. Bacterial growth curve	254
7.10. Labelling of H. influenzae R2866 proteins	254
7.10.1. Protein labelling in intact cells	254
7.10.2. Protein labelling in cell lysates.....	255
7.10.3. Periplasmic extraction.....	255
7.10.4. Protein identification by MS	256
References	258
Appendices	272

Acknowledgements

First and foremost, I would like to thank King's College London, the Faculty of Natural and Mathematical Sciences and the department of Chemistry for funding my studentship. My thanks go to Dr. Gerd Wagner for giving me the opportunity to undertake this PhD. I also would like to express my appreciation to the people who contributed to this work. Jiaming Peng, Victoire Riesi and Hayley Wootton, thank you for your invaluable contributions towards the synthesis of intermediates and target probe molecules. I would also like to thank Dr. Jennifer Dow and the group of Prof. Brendan Wren at LSHTM for the extensive microbiology training that enable me to carry out the experiment with *Haemophilus influenzae* described in Chapter 5. Thank you for making me feel so welcome, for your support and patience. Deepest thanks go to Nicola Evans, for her help with the crystallography. Thanks for sacrificing so much of your precious time for me. Many thanks to Dr. Andrew Atkinson, the NMR facility manager at King's College London, for his help on several occasions, when high resolution NMR was required. Thank you to Steve Lynham, from the Centre of Excellence for Mass Spectrometry at King's College London for the MS proteomics analysis and for a detailed discussion on the results. Thank you to Dr. Mirella Vivoli & Dr. Nicholas J. Harmer from the University of Exeter for the generation of LgtC mutants. Thank you also goes to the Bioorganic and Medicinal Chemistry Sector of the Royal Society of Chemistry for a generous travel award which allowed me to present part of my research at the 2018 International Carbohydrate Symposium (Lisbon, Portugal). Thank you to the technical team at Britannia House for providing training on equipment and for their continuous assistance. Thank you to Dr. Sarah Barry, and Dr Manuel Muller for their support and useful suggestions during group meetings.

I wish to express my deepest appreciation to Prof. Roger Morris, Dr. Alice Collier, Dr. Graeme Hogarth and Dr. Helen Coulshed for their continuous support throughout my PhD.

You have been a great inspiration. Teaching with you and learning from you has been a true privilege.

Thank you to all the special people at Britannia House starting with Yong, I can never thank you enough for spending a lot of time (when you had none to spare) helping me in the lab and teaching me the biochemical assay. I have been lucky to share the lab with the most incredible master's students. Jiaming, Vic and Hayley thank you for all your hard work and for all the laughing and crying (often simultaneously). Special thank you to Georgia, the best thesis writing partner I could ever hope for. Georgia, Monica, Janina and Kristine we have always supported and protected each other in this adventure when mechanisms for others to do so did not exist and I am forever grateful for this. To all of you and all others I did not name, this journey would not have been the same without you.

To my best friends, Nelly, Julia, Romy and Cathy thank you for your constant support and for always believing in me. I can never thank you enough for giving me the strength to pursue this adventure until the end. Phil, my husband to be, thank you for feeding me when I had no time to cook. Thank you for proof reading these 250 pages at times when you could have done something else. And above all thank you for being who you are, I am a lucky bunny. The Conti-Ramsden-Moulin family, thank you for your continuous support and love throughout the years. Dad, I promise you I will soon stop being a student and get a "proper" job. Mum, Dad, Thomas and Clement, thank you for being the best family, je vous aime. Manet, Mel, you have been my greatest source of inspiration. Your unconditional love has carried me throughout my whole life and your determination has given me the strength to undertake this PhD. I am so proud to be your granddaughter and hope that from where you are, you too, are proud of me.

This journey has not been a walk in the park, but I feel truly lucky and privileged to have received such a great academic education and to have been surrounded by the most wonderful people.

Abstract

Over the last few decades, the increasing emergence of multidrug-resistant bacteria has become one of the major threats to human health worldwide. Developing new antibiotic agents and understanding the biological mechanisms by which bacteria develop resistance are of great interest from both an academic and clinical perspective. Chemistry and chemical biology have played an essential role in biomedical research, as small molecular probes are widely used to uncover new potential therapeutics and probe virulence mechanisms in pathogens. Small molecules are ideal tools for exploring the underlying pathogenicity of biological mechanisms due to their ability to modulate or monitor enzyme activity levels. At the interface between chemistry and biology, the work presented herein aims to provide an example regarding how chemistry, through the rational design of small molecules, can be used as a powerful tool to explore virulence factors in Gram-negative bacteria. In bacteria, glycosyltransferases (GTs) and associated carbohydrate-active enzymes are responsible for the biosynthesis and modulation of surface glycoconjugates. Thus, this class of enzymes often play a key role in cellular adhesion, infection and virulence. LgtC is a virulent galactosyltransferase expressed in Gram negative bacteria such as *Neisseria meningitidis* and *Haemophilus influenzae*. It is responsible for the biosynthesis of a cell surface digalactoside epitope associated with serum resistance and evasion of the host's immune response. Small molecular inhibitors of LgtC are therefore sought after as chemical tools for the study of glycosylation processes involved in virulence. LgtC possesses many structural similarities with other bacterial GTs and is therefore an excellent model to develop a general strategy for the study of bacterial GTs and related enzymes.

In this thesis we explore the development and evaluation of carbohydrate-based inhibitors and chemical probes for LgtC, and their application as tool compounds for

chemical biology and microbiology. Recently, a non-substrate-based inhibitor of LgtC was identified from screening. Strong evidence suggests that it reacts covalently with the non-catalytic Cys246 in the acceptor binding site. Inspired by this, we designed a novel type of carbohydrate-based covalent inhibitor for LgtC, which combines an acceptor substrate scaffold with an electrophilic warhead. We reasoned that such substrate-based covalent inhibitors may exhibit greater target selectivity and potency than the existing non-substrate-based chemotypes. In Chapter 2, disaccharide candidates were rationally designed and 2-lactosamine-based inhibitors, with an acryl- or chloro-acetamide warheads attached *via* an amide bond were synthesised. This series of direct substrate analogues exhibits residual substrate activity which makes them unattractive as inhibitor candidates. In Chapter 3, monosaccharide analogues, with no residual substrate activity were synthesised. Investigation into their mode of action strongly suggests a non-covalent mode of inhibition which was an unexpected finding. Preliminary substrate competition experiments indicated a potential allosteric binding mode for the monosaccharides which may rationalise the non-covalent inhibition mode observed. In Chapter 4, a fluorescent probe was synthesised based on the scaffold of the chloroacetamide-containing monosaccharide inhibitor. The probe was found to label recombinant LgtC covalently. Substrate competition studies suggest binding occurs in the active site of the enzyme. Further investigations enabled confident conclusions that the inhibitor and the derived fluorescent probe exhibited a different binding mode. This provides additional evidence for a binding of the monosaccharide outside of the active site of LgtC. In Chapter 4 and 5, the covalent probe was utilised for the chemical proteomics profiling of non-pathogenic and pathogenic bacterial strains.

Our results provide great insight into the design of carbohydrate-based covalent probes. The deep understanding of the challenges faced allowed us to propose methods for the design of optimised probes as ideal tools to enable the identification of new virulence factors for the development of novel antimicrobial strategies.

Abbreviations

Å: angstrom

ABPP: activity-based protein profiling

ABPs: activity-based probes

Ac: acetyl(ated)

Acc: acceptor

AEBSF: 4-(2-aminoethyl)benzenesulfonyl fluoride

AfBPs: affinity-based probes

AMR: antimicrobial resistance

Aq: aqueous

Bn: benzyl

CAZy: carbohydrate-active enzyme database

CEL: chicken egg-white lysosome

CEMS: Centre of excellence for mass spectrometry

CIP: calf intestine phosphatase

Cmpd: compound

COSY: correlation spectroscopy NMR

CDCl₃: deuterated chloroform

Cys: cysteine

d: doublet

DCM: dichloromethane

deAc: deacetylated

dd: doublet of doublet

DMSO: dimethyl sulfoxide

Don: donor

dt: doublet of triplet

DTT: dithiothreitol

D₂O: deuterium oxide

E. coli: *Escherichia coli*

EDG: electron donating group

EDTA: ethylenediaminetetraacetic acid

eq: equivalent

EWG: electron withdrawing group

FSBA: 5'-fuorosulfonylbenzoyl 5'-adenosine

Gal: galactose

GalT: galactosyltransferase

GHS: Glycosyl hydrolases

Glc: Glucose

GlcNAc: N-acetyl glucosamine

GTs: glycosyltransferases

HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HI: *Haemophilus influenzae*

HPLC: high pressure liquid chromatography

HRMS: high resolution mass spectroscopy

Hz: hertz

IPTG: Isopropyl β -D-1-thiogalactopyranoside

J: coupling constant

kDa: kilo Dalton (unit)

Lac: lactose

LOS: lipooligosaccharide

LC-MS: liquid chromatography coupled with mass spectroscopy

LG: leaving group

LgtC: Lipopolysaccharyl- α -1,4-galactosyltransferase C

LPS: lipopolysaccharide

m: multiplet

MA: Michael acceptor

MBPs: mechanism-based probes

MeOD: deuterated methanol

MoA: mode of action

MS: mass spectroscopy

MLG: malachite green

MRSA: methicillin-resistant *S. aureus*
NMR: nuclear magnetic resonance
MW: molecular weight
NTHi: non-typeable *Haemophilus influenzae*
OD: optical density
PEG: polyethylene glycol
PBPs: penicillin-binding proteins
PG: peptidoglycan
Pi: inorganic phosphate
Pyr: pyridine
ppm: part per millions
s: singlet
SAR: structure relation activity
SDS: sodium dodecyl sulphate
SEC: size exclusion column
SFs: sulfonide fluorides
SHG: second harmonic generation
TCl: targeted covalent inhibitors
TIC: total ion chromatogram
TCEP: Tris(2-carboxyethyl) phosphine
td: triplet of doublet
THF: tetrahydrofuran
THPTA: Tris(3-hydroxypropyltriazolylmethyl)amine
TLC: thin layer chromatography
Tris: tris(hydroxymethyl)aminomethane
UDP: uridine diphosphate
UDP-Gal: uridine diphosphate galactose
UDP-2FGal: UDP-2-deoxy-2-fluoro- α -D-galactopyranose
UV: ultraviolet
WH: warhead
WT: wild type
XS: excess

Chapter 1

General introduction

1.1. Antimicrobial resistance, current challenges

The discovery of antibiotics in the 20th century was one of the greatest ever medical advances¹. It revolutionised healthcare, as common yet frequently deadly illnesses such as pneumonia and tuberculosis became efficiently treatable. Conventional antibiotics target cell viability by either killing bacteria (bacteriocidal) or inhibiting their growth (bacteriostatic). In both cases, essential bacterial functions are prevented such as cell wall synthesis, DNA replication or protein synthesis^{2,3}. Although initially highly effective, the mode of action of traditional antimicrobial agents imposes a selective pressure that prompts the emergence of resistant strains. Alarming, the introduction of a novel antibiotic has typically been followed by clinically significant resistance within as little as two years⁴ (Figure 1.1).

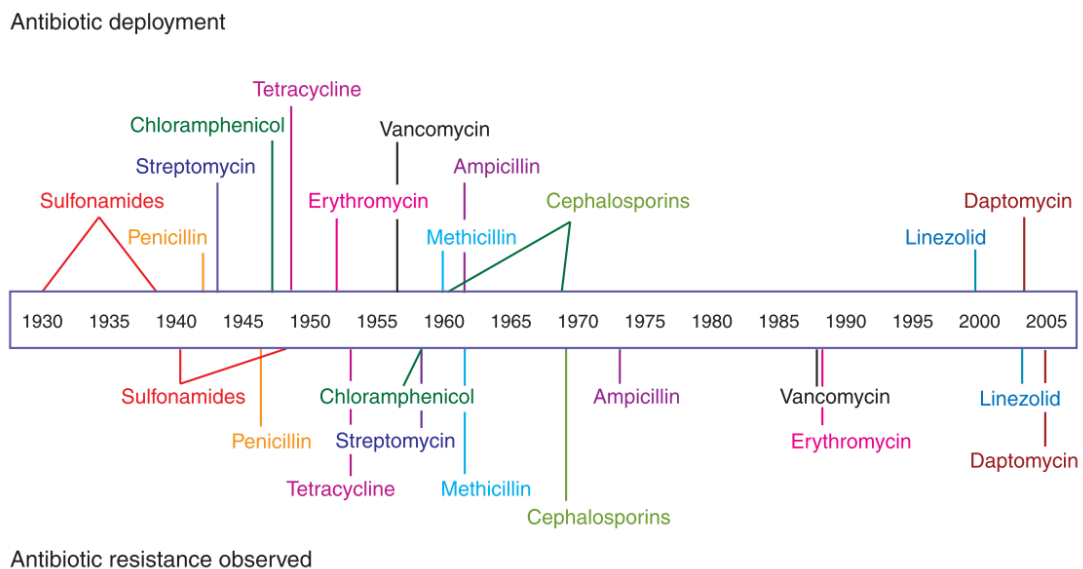


Figure 1.1. Timeline of antibiotic deployment and the evolution of antibiotic resistance²

Antimicrobial resistance (AMR) has been fuelled over the years by extensive and frequent misuse of antimicrobials by the community³. In the mid-20th century, diseases caused by pathogenic bacteria were controlled by the discovery of many new antibiotic scaffolds and derivatives. At the time, AMR was not the object of great concern as resistance was always countered by the next generation of antibiotics. However, in over 45 years now, except

for the recent development of daptomycin⁵ and linezolid⁶ effective against Gram-positive bacterial infections, no new classes of clinically relevant antibiotics have been discovered². Today, we are witnessing an alarming increase in the number multi-drug resistant strains with limited treatment options. AMR is now believed to claim at least 50,000 lives each year across Europe and the US alone, with many more in the third world⁴. Today the Infectious Disease Society of America estimates that around 70% of hospital-acquired infections in the US are resistant to at least one antibiotic². Other sources claim that in “fifteen European countries more than 10% of bloodstream *Staphylococcus aureus* infections are caused by methicillin-resistant *S. aureus* (MRSA), with several of these countries seeing resistance rates closer to 50%”⁷. There is therefore an urgent need for the discovery of novel and effective antimicrobial agents.

1.1.1. Examples of drug-resistant pathogens

Among the most dangerous pathogens exhibiting multidrug resistance and virulence are *Enterococcus faecium* and *Staphylococcus aureus* (Gram-positive bacteria), as well as *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter spp* (Gram-negative bacteria), collectively known as ESKAPE pathogens⁸. They are particularly challenging as they have developed a high level of drug resistance, limiting their treatment options and increasing morbidity and mortality⁹. ESKAPE pathogens are the highest cause of nosocomial infections throughout the world. In addition to these, other pathogens have recently gained further global attention for their increased observed resistance such as the Gram-negative bacteria *Neisseria meningitidis*¹⁰, responsible for invasive meningococcal diseases, and *Haemophilus influenzae*¹¹, causing community-acquired pneumonia (CAP) and meningitis.

In this thesis, most of the work is focused on the design of inhibitors and probes for the study of a virulence factor in *Neisseria meningitidis*. In the last chapter of the thesis, chemical probes are applied to whole cell studies on *Haemophilus influenzae* R2866.

1.1.1.1. *Neisseria meningitidis*

Neisseria meningitidis is a common Gram-negative bacterium that lives in the nasopharyngeal tract of humans. If spreads to other parts of the body, it can cause invasive and life-threatening infections. *N. meningitidis* is the leading cause of meningitis and rapidly fatal sepsis worldwide with over 500,000 meningococcal cases occurring each year¹². If untreated, such infections are fatal in up to 50% of cases¹⁰. In 1887, *N. meningitidis* was the cause of epidemic cerebrospinal fever ¹³, and still today, the prevention of human infections due to this pathogen remains a global challenge. The situation has recently been aggravated as just last year in Brazil, two ciprofloxacin-resistant *N. meningitidis* strains were discovered¹⁴. *N. meningitidis* serogroups A, B, C, W, X and Y cause most of epidemics and their distribution varies with geography.

1.1.1.2. *Haemophilus influenzae*

Haemophilus influenzae is a Gram-negative coccobacillus responsible for invasive diseases such as pneumonia and meningitis. This bacterium can be divided into two categories: encapsulated strains, containing a distinct capsular antigen, and unencapsulated strains, also known as non-typeable (NTHi) strains due to the absence of capsular serotypes. While only six types of encapsulated strains are known (from a to f), the genetic diversity of unencapsulated strains is greater¹⁵. Childhood pneumonia, meningitis, and sepsis associated with *H. influenzae* are mostly caused by type b strains (Hib) while community-acquired pneumonia (CAP) in adults and acute exacerbations of chronic bronchitis (AECB) are typically caused by non-typeable strains¹⁶. Even after the introduction of Hib vaccines, Hib is still responsible for approximately 370,000 deaths per

year and 60% of cases were meningitis¹⁶. Additionally, acquisition of new AMR to drugs such as fluoroquinolones¹⁷ and ampicillin¹⁸ has been described for *Haemophilus influenzae*.

Clear conclusions can be drawn: our current strategies for antimicrobial design impose a great evolutionary pressure on bacteria and the emergence of resistant strains is inevitable. Additionally, such survival mechanisms have been recently shown to predate the discovery of antibiotics as genes encoding resistance to β -lactam, tetracycline, and glycopeptide antibiotics have been identified in an ancient DNA sample from 30,000-year-old Beringian permafrost sediments¹⁹. This finding provides evidence that resistance is a natural phenomenon and cannot be completely avoided. It also reinforces the fact that new strategies, not compromising cell viability and therefore less prone to induce resistance, must now be employed for the discovery of novel antibiotics. Understanding the mechanisms of pathogenicity and resistance spreading is critical for better selection of the next generation of antibiotic drugs.

1.2. Mechanisms of AMR

The remarkable ability of bacteria to overcome the killing effects of antimicrobial agents and transmit resistance is caused by sophisticated survival mechanisms not yet fully understood. Bacterial physiology can confer intrinsic resistance as is the case for Gram-negative bacteria and the low permeability of their outer membrane which serves as the first line of defence against antibiotics²⁰. Resistance can also arise from spontaneous mutations in the protein or gene target of the antibiotic which prevents drug binding²¹. Such unpredictable mutations are the result of the strong evolutionary pressure exerted by antimicrobials and are a normal adaptive response and a clear manifestation of the principles of evolution. Although resistant genes can be inherited vertically from parent to progeny, horizontal gene transfer between species is responsible for most of the acquired antibiotic resistance often due to the polymicrobial nature of infections²².

Bacteria also have the ability to modify the mode of action of the antibiotic drug. Specific enzymes may be expressed to introduce chemical modification to the molecular scaffold of the drug making it inactive²². β -lactamase, for example, cleaves the amide bond in the β -lactam ring of penicillin derivatives rendering it ineffective. Bacteria have also developed sophisticated mechanisms to decrease cell uptake of antibiotics, either preventing the antibiotic from reaching its intracellular or periplasmic target, or excluding it from the cell *via* a complex efflux system^{22,23}. Additionally, when under stress, bacteria are organised in a surface adherent biofilm structure. Their formation occurs through a series of events coordinated through a cell-cell communication process known as quorum sensing. Biofilms are up to 1000 times more resistant to antibiotic than planktonic cells²⁴.

In addition to this, virulence factors, essential for infection and host colonization in pathogenic bacteria, have been widely studied in the past decade due to their role in resistance^{25,26}. The development of small molecular tools to probe biological targets involved in pathogenicity and virulence may provide a stepping stone towards the development of novel antibiotics^{25,26}.

1.3. Anti-virulence strategies

The disruption of host / pathogen interactions, mediated by virulence factors, is an attractive option for the development of new antimicrobial agents that is being increasingly explored. The concept of virulence is understood by the ability of bacteria to cause disease and infiltrate a host² (degree of pathogenicity). In practice, once at the site of infection, pathogenic bacteria initiate a specific mechanism for the activation of virulence traits making the host ill. Therefore, anti-virulence strategies, preventing the expression or activation of these virulence traits, interfere with the ability of the pathogen to recognise and colonise its host²⁵. In other words, anti-virulence therapies aim to “disarm” the pathogen preventing it from causing direct harm to its host. The comparison

of anti-virulence agents with vaccination is an interesting analogy as in both cases, the host is infected with an attenuated and non-illness inducing pathogen². The non-virulent infection can then be cleared by the host immune response with little to no impact on the normal human microbiome. Importantly, the inhibition of virulence, rather than growth, does not affect cell viability which imposes a milder evolutionary pressure to the pathogen making it less prone to develop resistance^{24,25}. Due to this incredible potential, biological pathways involved in virulence such as toxin function²⁷ and delivery²⁸, regulation of virulence expression²⁹ and bacterial adhesion³⁰ have been thoroughly investigated. Additionally, new anti-virulence candidates have recently been identified (Figure 1.2).

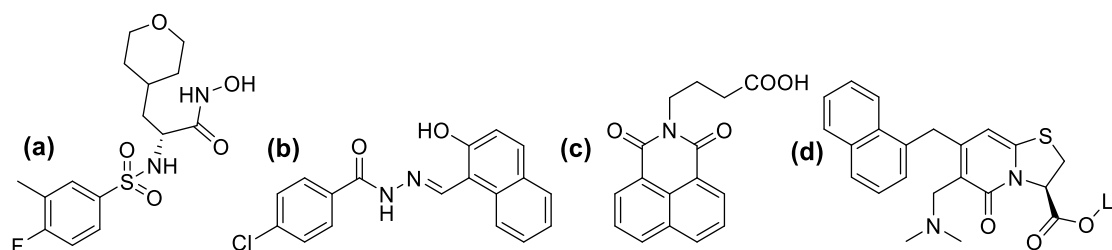


Figure 1.2. Example of anti-virulence candidates recently identified. (a) *B. anthracis* lethal factor inhibitor²⁷ (Toxin inhibitor) (b) the T3SS inhibitor INP0400 (prevent toxin delivery)²⁸ (c) the quorum sensing inhibitor furanone C-30 (regulator of virulence expression) (d) a pilicide (a bicyclic 2-pyridone compound) (inhibit adhesion of *E. coli* to bladder).

1.4. Relationship between AMR and virulence

Both AMR and virulence are associated with infection and host pathogenesis and both are necessary biological processes when bacterial survival is compromised^{24,31,32}: virulence mechanisms overcome the host's defence system while development of AMR allows pathogenic bacteria to evade and adapt to antibiotic pressure. Quorum sensing, a key biological process involved in AMR for the formation of pathogenic biofilm structures, is also used by bacteria as a mean to increase virulence, as observed in a recent study of seed-borne pathogen *Acidovorax citrulli*³³. The regulation of virulence factors and the expression of AMR genes can indeed influence each other in a complex manner²⁴. The

combined action of biofilm genes and quorum sensing molecules regulates the expression of virulence and AMR genes (Figure 1.3). Therefore, the relationship between AMR and virulence is somehow linked to pathogenic biofilm formation^{24,31}. However, other characteristics are also common to both AMR and virulence such as efflux pumps and cell wall alterations³¹. In practice, the direct association of resistance and virulence is unsurprising as antibiotics are typically administered when the host is infected by virulent pathogens. This co-occupation of the host most likely explains why AMR and virulence share many common biological processes. Also, in addition to evading the host immune response, the rapid expression of virulence traits by pathogens may render them too invasive to be treated by common antibacterial strategy therefore leading to resistance. Many studies have shown that increased resistance is observed alongside increased virulence^{34,35}.

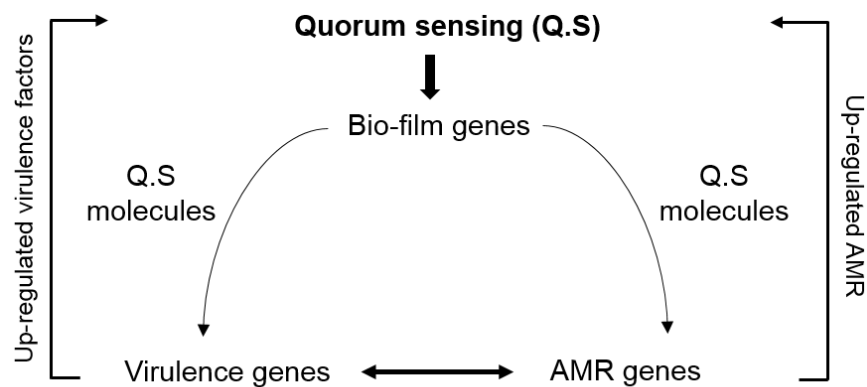


Figure 1.3. Regulatory mechanisms act as a connection between virulence and antibiotic resistance (adapted from Brooks *et al.*, Genes 2017²⁴)

Virulence and antibiotic resistance are therefore closely related and designing small molecules to probe, and study virulence factors will allow the development of more specific and directed antimicrobial treatments³².

1.5. Glycoconjugates: LOS and LPS at the surface of Gram-negative bacteria

Bacteria establish a close contact with their host's mucous membrane for the manipulation of their metabolism and immune system. Gleaning insight into the role of the molecules present at the surface of bacteria is essential for developing a better understanding of host / pathogen interactions. Glycoconjugates cover the vast majority of the bacterial surface and can be divided in several categories such as peptidoglycan (PG), exopolysaccharides (EPS), capsular polysaccharides (CPSs), lipopolysaccharides (LPS), and lipooligosaccharides (LOS) (Figure 1.4)³⁶. The highly conserved PG structure forms the basis of the cell wall for the anchorage of other molecules. CPS, EPS or glycoproteins are found in all types of species while glycosylated teichoic acids (Tas) are typical of Gram-positive bacteria. LOS and LPS structures are found exclusively at the surface of Gram-negative bacteria (Figure 1.4). The enormous variety of glycoconjugates combined with their key location on the bacterial cell wall indicate their role as unique barcode for the recognition of host cells and subsequent processes such as adhesion and immunomodulation³⁷.

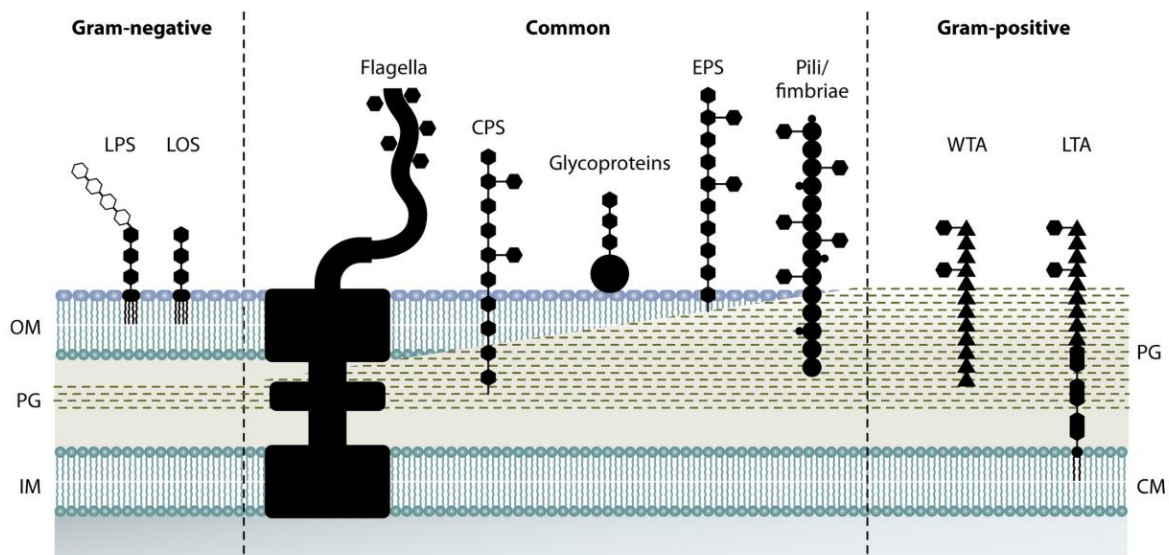


Figure 1.4³⁶ Bacterial glycoconjugates. The round dots represent proteins, the triangles are ribitol phosphate or glycerol phosphate moieties and the hexagones are carbohydrates (from Tytgat & Lebeer *Microbiol. Mol. Biol. Rev.* 78, 372–417, 2014)

Gram-negative bacteria, responsible for many nosocomial infections, are characterised by the presence of LPS and LOS glycan structures on their surface. These glycoconjugates cover about 75% of the cell surface and exert essential functions such as cell wall integrity and resistance against environmental stress. Additionally, because of their key role in pathogen / host interactions, the surface glycans have been identified as important virulence factors^{36,38}. LPS are built out of a lipid A anchor upon which long polysaccharides are covalently attached: an inner and outer core oligosaccharide comprising 2-keto-3-deoxyoctulosonic acid (KDO), followed by the O-antigen polysaccharide (Figure 1.5). The O-antigen is the most diverse section of the LPS structure and together with the CPS, determines the serotype specificity of strains. The O-antigen is also one of the major surface antigens inducing a prompt immune response by the host. For this reason, LOS, a short version of LPS lacking the O-antigen, (Figure 1.5) are found at the surface of virulent pathogens such as *C. jejuni*, *Haemophilus*, and *N. meningitidis*¹³. This ingenious variation allows them to evade the immune response in the host triggered by the presence of the O-antigen^{36,38}. Additionally, both LOS and LPS structures can recruit unusual sugars mimicking mammalian glycans. This molecular mimicry is another strategy by which virulent pathogens can prolong their persistence and evade host immunity^{36,39,40}.

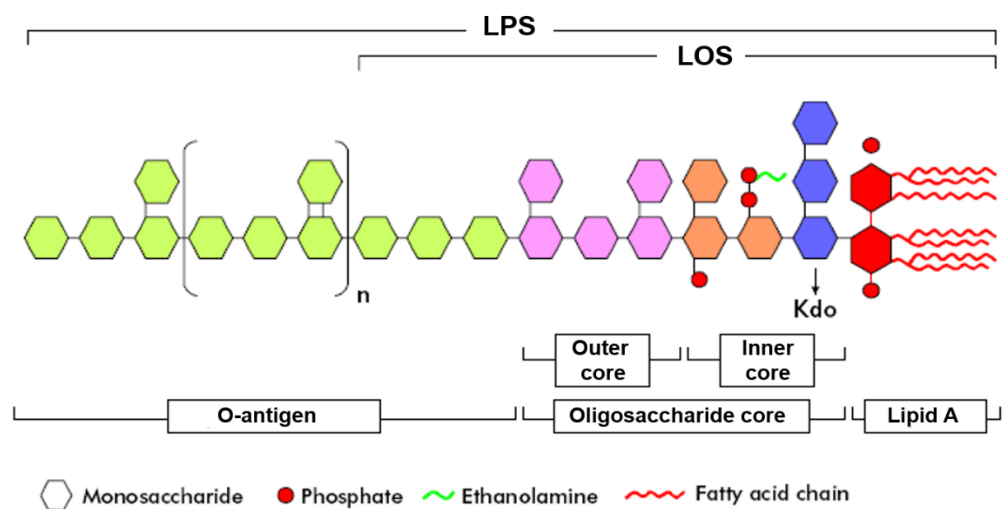
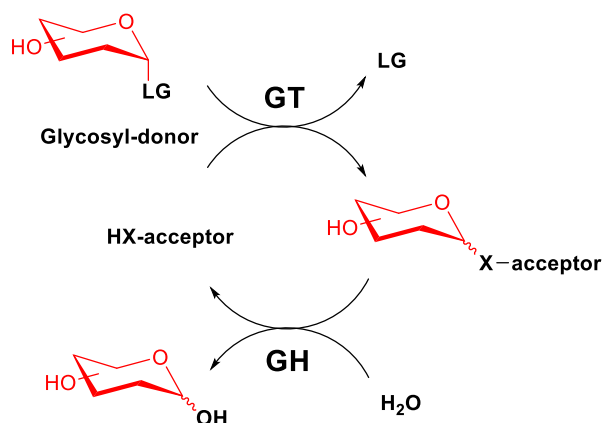


Figure 1.5⁴¹. General structure of LOS and LPS

Surface glycan structures secreted by Gram negative pathogens, such as LOS and LPS, are key mediators in host / pathogen interactions enabling pathogenesis and virulent bacterial colonization. The biosynthesis of such complex glycans in bacteria, and other domains of life, is the result of the naturally orchestrated action of many carbohydrate-active enzymes such as and glycosyl hydrolases (GHs), also known as glycosidases, and glycosyltransferases (GTs).

1.6. Glycosyltransferases (GTs)

GTs are Nature's glycosylation agents. Along with glycosidases (GHs) they are believed to constitute about 3% of the bacterial genome, reflecting the importance of the role of glycans in bacteria as previously discussed³⁶. The known complexity and variety of glycans present in the biosphere is the result of the combined action of many GTs and GHs. In all domains of life, while GTs catalyse glycosylation reactions, GHs hydrolyse glycosidic bonds (Scheme 1.1).



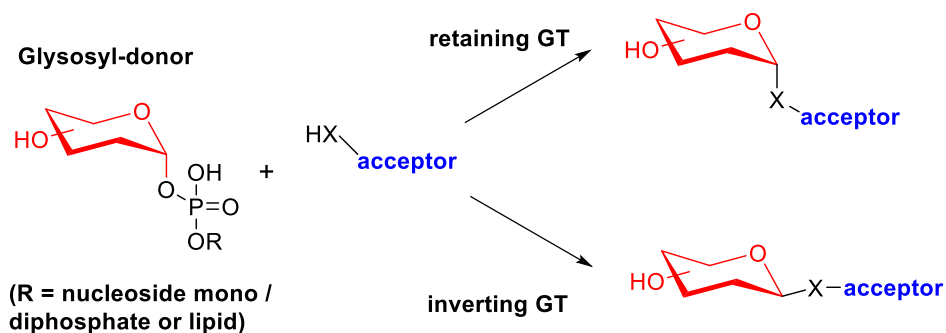
Scheme 1.1. Mechanism of glycosyltransferase (GTs) versus glycosydases (GHs)

GTs are responsible for the elongation of glycans by adding monosaccharides to the non-reducing ends of acceptor substrates. They transfer a sugar moiety from a glycosyl donor to a variety of suitable acceptor molecules (sugar, protein, lipid or DNA) for the biosynthesis of vital and complex oligosaccharides^{42,43}. This family of carbohydrate-active enzyme can be classified in four different categories depending on: the mechanism by

which the enzymatic reaction occurs (inverting or retaining GT, see 1.6.1), their amino acid sequences (CAZy classification, see 1.6.2), the type of sugar donor molecules they use (Leloir and non-Leloir GTs, see 1.6.3) and their 3D structures (GT-A, GT-B or GT-C architecture, see 1.6.4).

1.6.1. Mechanism of GTs

In contrast with the well characterized hydrolysis mechanism of GHs, the mechanism by which GTs glycosylate their substrate remains less understood⁴². The transfer of the sugar moiety by GTs is both regio- and stereospecific and two stereochemical outcomes are possible^{42,43,44}. The anomeric centre of the bio-product can either have a “retaining” or “inverting” stereochemistry with respect to the original donor sugar^{42,43,44}. Thus, depending on the reaction outcome, GTs are divided into either “retaining” or “inverting” enzymes (Scheme 1.2).



Scheme 1.2. Stereochemical outcomes for inverting and retaining GTs

Although still partially unclear, both catalytic mechanisms have been extensively studied^{42,43}. Like inverting GHs, the mechanism of inverting GTs is an S_N2 -like single displacement and occurs *via* an oxocarbenium ion transition state. The abstraction of a proton from the acceptor OH-group by a catalytic base (usually provided by an active-site side chain such as Asp, Glu or His) facilitates its attack at the anomeric position of the

donor with net inversion of the stereochemistry (Figure 1.6). Theoretical studies indicate that the nucleophilic addition and the LG departure are nearly simultaneous⁴⁵. Additionally, the negative charge on the leaving phosphate is stabilised by a divalent metal (generally Mn^{2+} or Mg^{2+}) for GT-A (see in 1.6.4) or positive amino acids/helix dipoles for GT-B proteins.

On the other hand, the mechanistic characterisation of retaining GTs has proven to be challenging and two hypothetical mechanisms are still debated today. By direct comparison to retaining GHs, a double displacement mechanism for retaining GTs has been proposed⁴². This involves two subsequent S_N2 reactions and a short-lived glycosyl-enzyme covalent intermediate following the attack of a specific catalytic nucleophile (Figure 1.6). However, despite numerous efforts using techniques which have proven successful for the characterisation of retaining GHs, no glycosyl-enzyme intermediate has yet been trapped, except for one instance, involving two blood group-synthesizing GT mutants⁴⁶, nor has a conserved catalytic nucleophilic residue been identified^{36,47}. This may provide evidence against this mechanism for retaining GTs but could also be explained by an inapplicability of these techniques to GT targets. As a result of this evidence, a second proposed mechanism emerged, the direct attack of the acceptor *via* an S_{Ni} -like reaction (Figure 1.6). The mechanism involves an intermolecular interaction between the departing phosphate and the incoming acceptor molecule leading to an enzyme-stabilised oxocarbenium. Beyond stabilising the intermediate, the enzyme also protects one face of the oxocarbenium, consequently forcing the nucleophilic attack on the opposite face resulting in a retention of configuration⁴⁷. The contradictions that have developed over the years of studies have suggested that, unlike GHs, there may not be a single uniform mechanism for retaining GTs.

GTs of bacterial origin exhibit both catalytic mechanisms with approximately twice as many inverting enzymes as retaining enzymes being reported to date (CAZy database).

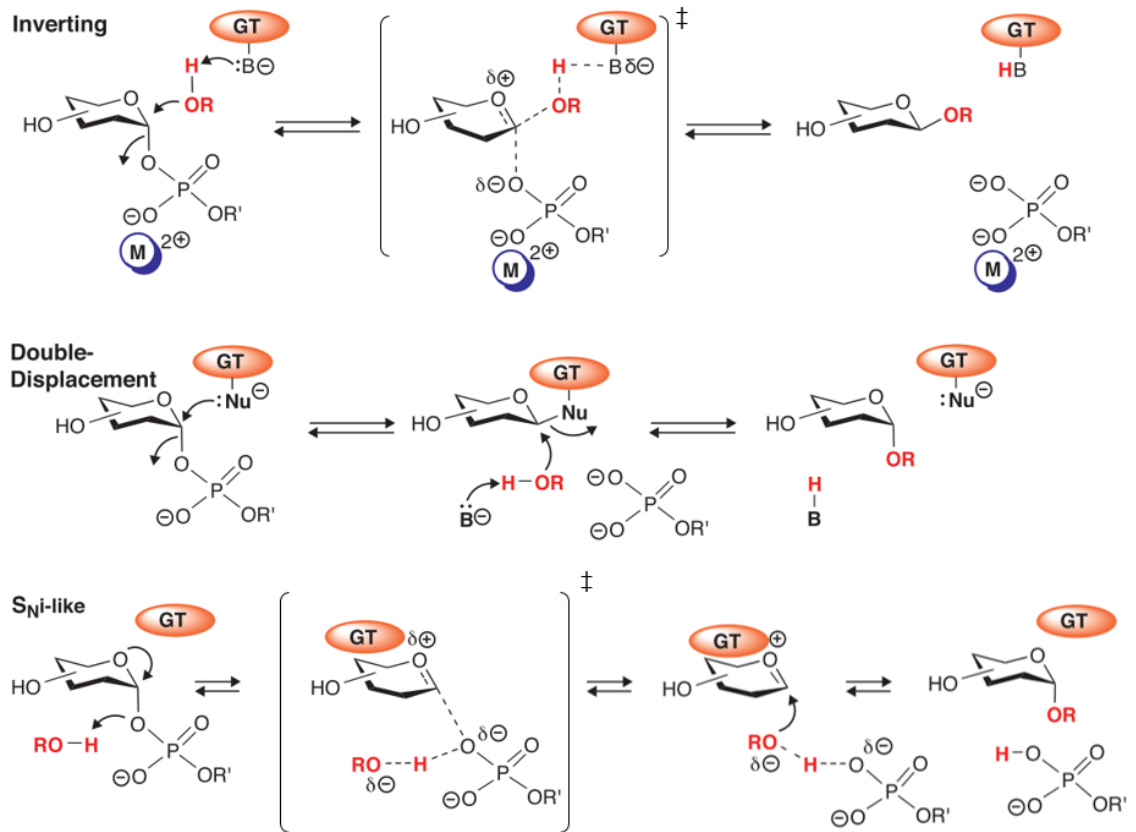


Figure 1.6⁴³ Suggested mechanisms for inverting and retaining GTs (sugar acceptor in red)

1.6.2. Classification according to amino acid sequence: the CAZy database

GTs can be classified into many distinct families depending on their amino acid sequence similarities, reflecting the variety of substrates that can be used⁴⁸. For all species, the CAZY (carbohydrate-active enzymes) database records, to date, more than 570,000 GT sequences grouped into 108 families. Additionally, 12,000 more protein sequences with putative GT activity but yet unknown biochemical function have been identified to date (August 2019). Only two years ago, 340,000 sequences had been identified with 7,000 putative GTs. These ever-growing numbers demonstrate the extensive arsenal available in nature for the biosynthesis of complex oligosaccharides.

86% of all recorded GT sequences in the CAZY database are from bacterial origin, the 14% left are shared between eukaryotes, viruses and archaea. For GHs, bacterial enzymes also

contribute to the majority of recorded proteins. This indicates the immense arsenal of existing bacterial GTs and GHs and reflects the great ability of bacteria to modulate the structure of their surface glycoconjugates and ensure virulent host infection. Interestingly, 58% of all bacterial GTs recorded on the database are gathered in only two large families: GT2 with 164,000 enzymes and GT4 with 128,000. Additionally, 82% of them are found in eight families (GT1, GT2, GT4, GT8, GT9, GT28, GT35, GT51) indicating that the 100 other families only account for 18% of all known GTs. The eight GT families, accounting for most of bacterial GTs recorded to date, almost exclusively contain bacterial GTs (except for GT1) (Table 1.1). This contrasts with GT5 family, which for example contains 0.8% of Archaeal, 56% of bacterial and 42% of eukaryotic GTs. All this suggests a high degree of sequence similarities among bacterial GTs not necessarily shared with human GTs. This can be exploited for the design of small molecular tools and the study of virulence factors.

Table 1.1. Percentage of bacterial GTs in GT1, 2, 4, 8, 9, 28, 35, 51 families

CAZy family	GT1	GT2	GT4	GT8	GT9	GT28	GT35	GT51
% of bacterial GTs	63	92	95	84	~100	99	94	~100

The prediction of GTs' functions using this classification may be challenging, as within the same family (similar sequence), GTs catalysing different reactions are found (alongside different substrate specificity)⁴⁸. Therefore, beyond the similarity of amino acid sequence, GTs can also be classified according to their donor substrate types.

1.6.3. Leloir vs non-Leloir GTs

GTs can be classified in two categories depending on the nature of their glycosyl-donor substrate. GTs using sugar nucleotides such as UDP-Gal or GDP-Man are called "Leloir type" enzymes in honour of Luis F. Leloir, who won a Nobel prize in 1970 for his

contributions to our understanding of glycoside biosynthesis and sugar metabolism^{42,43}. On the other hand, non-Leloir GTs use non-nucleotide donors such as polyprenol pyrophosphate and sugar-1-phosphate^{42,44}. In both cases, for the glycosyl transfer reaction to occur, the donor sugar is activated by the highly energetic glycosyl-phosphate bond⁴⁴.

Bacterial GTs use mainly sugar nucleotide diphosphates as glycosyl donors such as UDP-galactose for β -1,4 galactosyltransferase (GT8), GDP-mannose for dolichyl-phosphate β -D-mannosyltransferase (GT2), UDP-N-acetyl-D-glucosamine for lipopolysaccharide N-acetylglucosaminyltransferase (GT9). In some cases, sugar nucleotide monophosphate donors are used. For example, GTs in families 73 and 107 are exclusively bacterial KDO-transferases, involved in the synthesis of the LPS and LOS envelopes and use CMP- β -KDO, a sugar nucleotide monophosphate, as the glycosyl donor. Additionally, polysialyltransferases, in GT38 family, use CMP-beta-N-acetylneuraminate as the sugar donor. Although sugar nucleotides are the donor scaffold of choice for bacterial GTs, enzymes in GT51, a large family of over 42,000 murein (peptidoglycan) polymerases exclusively from bacterial origin, use undecaprenyl-pyrophosphate-MurAc-(GlcNAc)-pentapeptide, a lipid-phosphate sugar, as the donor substrate.

1.6.4. 3D-structures of GTs

There is a wide variety of sequences in the GT enzyme class, reflected by the increasing number of families. However, GTs are characterized by a more conserved three-dimensional architecture as only three protein folds have been identified so far.

It was not until the late 1990s, when the first X-ray crystal structures of GTs were published that the structural characterization of GTs was possible. Unlike GHs, GTs exhibit a narrow spectrum of folds. All Leloir enzymes solved to date can be split into only two families based on their common three-dimensional fold: GT-A and GT-B. The GT-A fold

was first identified when the 3D structure of the family 2 enzyme SpsA from *B. subtilis* was solved in 1999⁴⁹. This fold comprises two tightly associated $\beta/\alpha/\beta$ Rossmann-like domains, typical of nucleotide-binding proteins (Figure 1.7)^{50,43}. The two domains, whose sizes vary, abut closely, forming a central β -sheet responsible for the description of the GT-A fold as “a single domain fold”. However, distinct donor and acceptor binding sites are present. GT-A are generally divalent metal ion dependent (generally Mn^{2+} or Mg^{2+}). The ion is coordinated by the carboxylate of the Asp-X-Asp (D-X-D) signature motif within the GT active site and promotes leaving group departure by stabilizing the phosphate groups in the nucleotide sugar donor. As seen for GT-A enzymes, the structure of GT-B enzymes consists of two $\beta/\alpha/\beta$ Rossmann-like domains; however, in this case, the two domains are facing each other making them less tightly bound (Figure 1.7)^{50,43}. For GT-B, the active site is located within the cleft formed in between the two domains. GT-B enzymes are generally metal ion independent, with active site residues acting to promote leaving group departure.

More recently, GT-C, a third fold of hydrophobic integral membrane GT was identified (Figure 1.7)⁵⁰. GTs with that architecture are exclusively the less common non-Leloir enzymes which are defined by being non-nucleotide dependant. To date, the few solved GT-C structures are large hydrophobic integral membrane proteins and, not surprisingly, utilise lipid phosphate-linked sugars as donor substrates. They contain 8 to 13 trans-membrane helices with the active site located on a long-loop region^{50,42}. Most GT-C enzymes are involved in the synthesis of polysaccharide derivatives of dolichol phosphate, mostly found in eukaryotes and in archaea⁵¹. Therefore, GT-C appears to be a highly specialized and evolutionarily recent group of GTs.

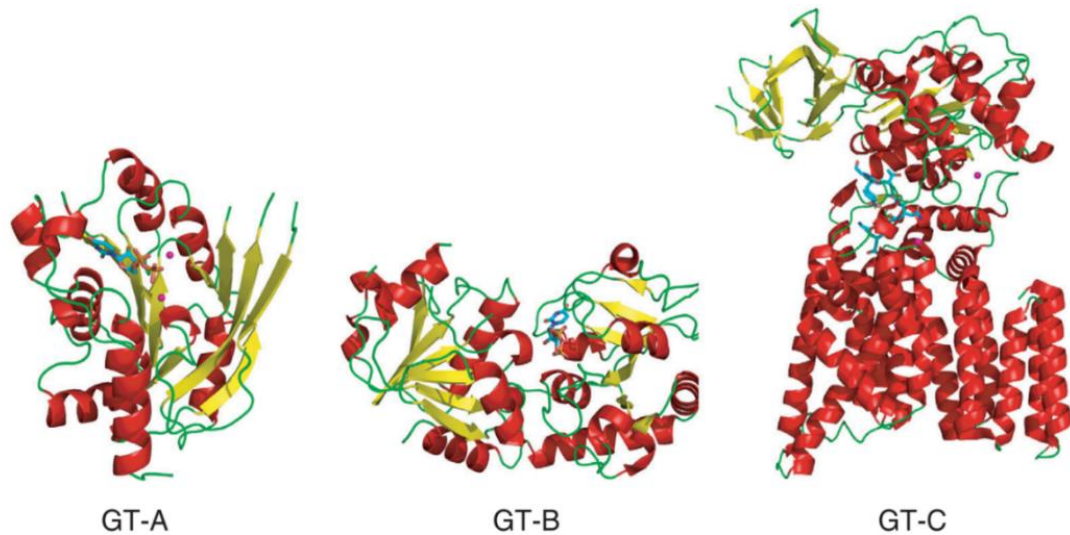


Figure 1.7⁵⁰. Representative folds of GTs

The number of GT folds identified compared with the recent increase in reported enzyme structures is strikingly low. Furthermore, protein threading analysis has revealed that many of the yet uncharacterized GTs most likely belong to one of the two main architectures (GT-A or GT-B)⁴². This suggests that most GTs may have evolved from only a few progenitor sequences. Some even propose that GT-A and GT-B enzymes may have evolved from a unique ancestral gene due to them sharing a Rossmannoid structure. Experts have expressed different opinions regarding the origin and evolution of GTs. One suggestion is that the GT-A fold was the earliest to develop as it represents the majority of the GTs identified to date. The GT-B structure may therefore have originated from gene duplication of an ancestral GT-A enzyme⁴³. On the other hand, GT-B enzymes are mostly involved in the synthesis of core glycan structures such as the PG layer, while GT-A enzymes are responsible for most of the glycoconjugate elongation and terminal decoration. Therefore, based on the role of enzymes found in both families the opposite argument is also defensible⁴³.

Only 1175 bacterial 3D GT structures have been solved so far (PDB). This corresponds to 42% of all available GT structures. 54% are eukaryotic enzymes while 1 and 3% are from viral and Archaeal origin. Interestingly, although eukaryotic GTs correspond to only 10%

of recorded enzymes (CAZy), more structures have been solved. By comparing the numbers between the PDB and CAZy database, it can be concluded that only 0.2% of all known bacterial GTs have been solved structurally. This indicates the significant lack of structural information available for GT from bacterial origin. Among the solved structures, *Escherichia coli* is the organism with the most solved GT structures (12% of all solved bacterial GTs). On the other hand, 3% of the available structures belong to the *Neisseria* genus. Bacterial GTs, using sugar nucleotides as their glycosyl donor of choice in most cases, are typically characterised by a GT-A or GT-B fold. While solved bacterial structures in families GT1, GT4, GT9, GT28 and GT35 all have a GT-B structure, available structures in families GT2, and GT8 possess a GT-A fold. Accounting for all families containing bacterial GTs, approximately 70% of bacterial enzymes have a GT-B fold. GT-C folds are rare in prokaryotes and are limited to parasitic mycobacteria⁵¹. Interestingly, in the non-Leloir family GT51, containing exclusively peptidoglycan GTs with only 8 solved structures, a distinct fold from GT-A, B or C has been identified. This is the case of Pbp2, a transpeptidase penicillin-binding protein from *S. aureus*⁵² whose structure has been characterised as lysosome-like, because of a significant level of similarities with lysosome structures.

Therefore, our understanding of bacterial GTs at the structural level remains poor. In comparison, almost twice as many bacterial GHs have been solved to date (PDB). This, in addition to a more complex double substrate mechanism for GTs, could explain why despite numerous efforts, the elucidation of GTs catalytic mechanism remains partially unclear. Additionally, as will be described later in the thesis, small molecular tools are significantly more developed for GHs than for GTs. An understanding of the enzyme at the structural level provides great grounds towards the design of chemical probes. Therefore, our currently poor knowledge of bacterial GT 3D structures may account for the lack of biochemical tools available for the study bacterial GTs.

1.7. *LgtC* from *Neisseria meningitidis*

As part of this thesis, the work was focussed on the study of the bacterial lipopolysaccharyl α -galactosyltransferase C (LgtC) from *Neisseria meningitidis*.

1.7.1. *LgtC* and its role in virulence

LgtC is a retaining galactosyltransferase (GalT), responsible for the transfer of α -galactose from uridine 5'-diphospho- α -galactose (UDP-Gal) to a terminal lactose acceptor of the bacterial cell-wall (Figure 1.8). With 311 residues, LgtC is a monomer belonging to GT family 8. Found exclusively in Gram-negative pathogenic bacteria and more specifically in *Haemophilus*, *Neisseria*, *Pasteurella* and *Brachyspira* genera, LgtC plays an important role in the biosynthesis of the LOS envelope on the bacterial surface. A study on *Haemophilus influenzae* NTHI strain R2866, isolated from a child with meningitis, showed that LgtC is a critical factor in the high-level serum resistance observed for this strain⁹. MS analysis showed that the terminal digalactoside epitope, resulting from the reaction catalysed by LgtC (Figure 1.8) was present in strain R2866 but absent in R3392, a strain derived from R2866 with increased sensitivity to human serum⁹. In addition to this, *in vivo* infection studies demonstrated that the expression of the digalactoside epitope resulted in increased virulence⁵³. It is believed that the digalactoside epitope mimics the human Pk blood group glycolipid⁵⁴ therefore enabling the pathogen to attach to host receptors and evade the immune response. The digalactoside epitope, and by association LgtC, have therefore been identified as virulence factors^{53,9}.

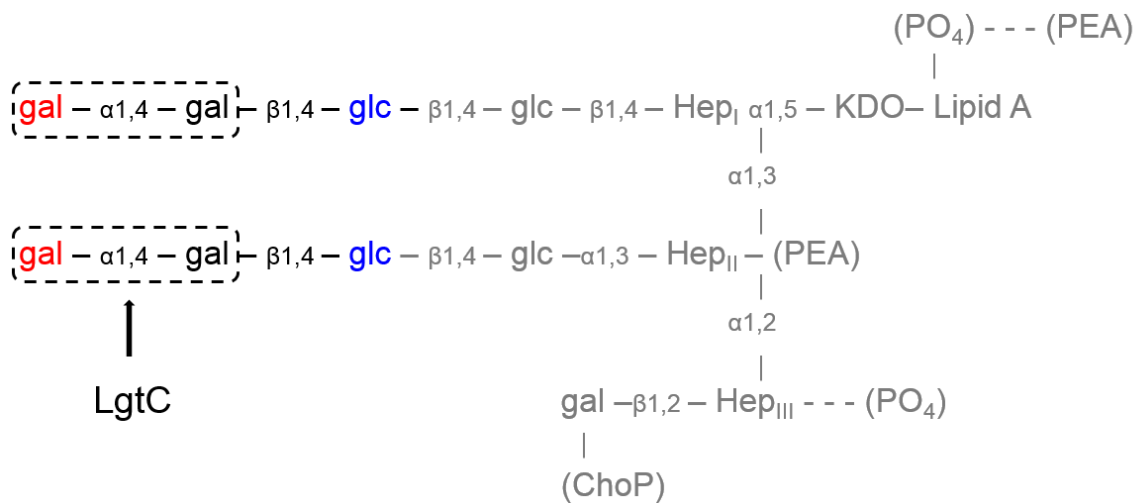
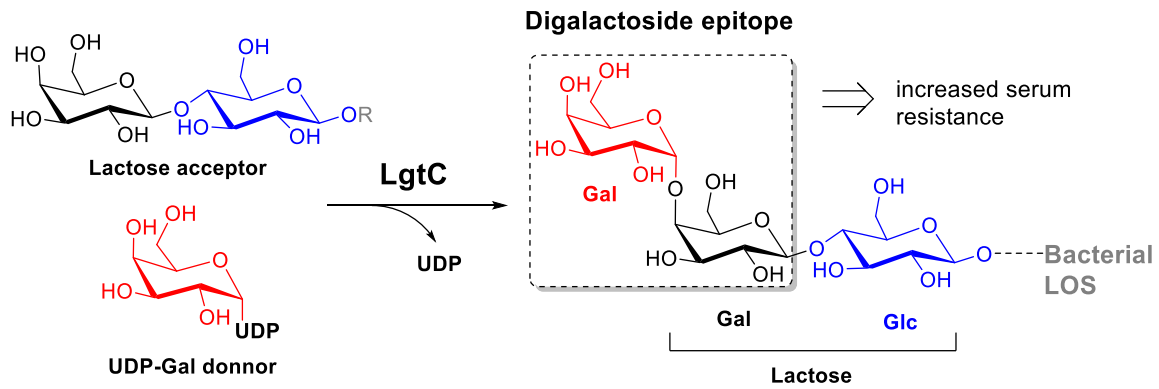


Figure 1.8. Reaction catalysed by LgtC and schematic representation of the structure of the fully extended LPS glycoforms of *H. influenzae* type b strains RM153 and RM7004. Hep, LD-heptose; Glc, glucose; Gal, galactose; P, phosphate; ChoP, phosphorylcholine; PEA, phosphoethanolamine. A dotted line indicates the substituents that are variably present. The places of action of the proximal-to-distal heptoses are numbered I to III accordingly (adapted from Moxan *et al.*, Infect. Immun, 2005⁵³)

Lipooligosaccharide (LOS), is one of the major virulence factors involved in meningococcal pathogenesis and both the morbidity and mortality of meningococcal sepsis are highly dependent on levels of circulating meningococcal LOS¹². For this reason, the discovery of small molecules with the ability to block the biosynthesis of LOS in pathogenic bacteria, such as LgtC inhibitors, may be sought after in drug discovery for the discovery of new anti-virulence agents.

On the other hand, the structures of LOS are extensively heterogeneous due to phase-variation: the ability of a pathogen to alter surface-exposed molecules. The variable expression of LgtC gene has been widely reported for *Haemophilus* and *Neisseria*^{9,55,56}.

1.7.2. Phase variation

In the previously described study⁹, the variable expression of LgtC in *Haemophilus influenzae* NTHI strain R2866 was described. Lic3A, the only *H. influenzae* sialyltransferase reported to be phase variable at the time, was used as a positive control to evaluate the phase variation associated with LgtC and the subsequent impact on serum resistance. In the virulent strain R2866, the increased serum resistance is found to be due to the expression of LgtC rather than lic3A, as while it remains high in lic3A deletion mutant, it is significantly decreased in LgtC-Off colonies (Figure 1.9). Also, in the non-virulent *H. influenzae* parent strain R3392, with identical multilocus sequence as R2866, LgtC was not expressed (leading to the absence of galactoside epitope observed for this strain by MS). This indicated that, in *H. influenzae*, the serum resistance controlled by LgtC is variable.

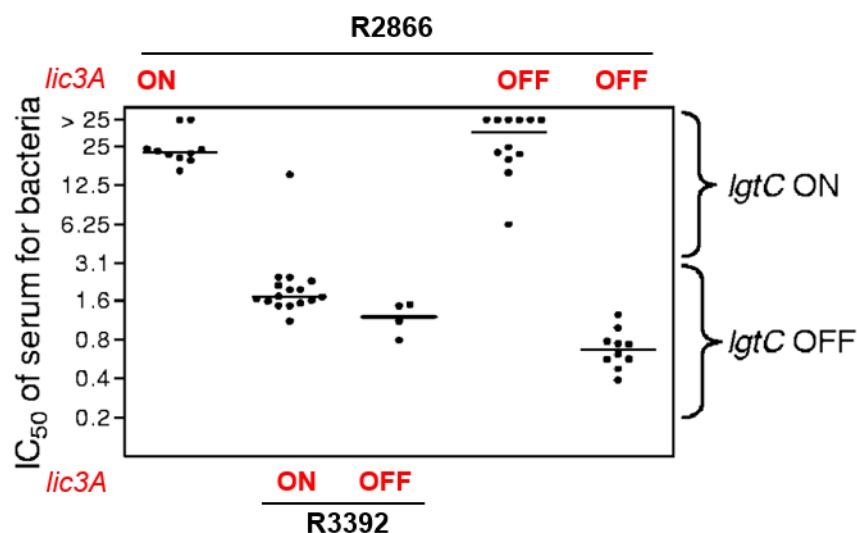


Figure 1.9⁹ IC₅₀ of normal human serum for cultures derived from colonies of R2866 and R3392. OFF: Deletion mutants. ON: Colonies expressing protein of interest. High level of serum resistance is observed in strains expressing LgtC (LgtC ON strains) which is independent on the expression levels of lic3A. LgtC is not expressed in non-virulent strain R3392 (adapted from Erwin *et al.*, Infect. Immun 2006)

A study examined nine lgt genes at three chromosomal loci (lgt-1, 2, 3) in 26 *Neisseria meningitidis*, 51 *Neisseria gonorrhoeae* and 18 commensal *Neisseria* strains⁵⁷. The authors found that in *N. meningitidis*, lgt-1 and lgt-3 loci were hypervariable genomic regions therefore leading to a high degree alteration in the LOS structure. Lgt-2 locus, on the other hand, was conserved among all *N. meningitidis* strains⁵⁷. Interestingly, in *N. gonorrhoeae*, no significant variability in the composition of all three lgt loci was observed, and all strains expressed LgtC. Unsurprisingly, lgt genes were detected only in very few commensal *Neisseria* species⁵⁷.

Despite its key role in virulence, LgtC has a poor prospective as therapeutic target for anti-microbial strategy due the high degree of phase variation associated with its expression in virulent strains. However, LgtC can serve as an indicator of the degree of virulence of a specific strain. Additionally, it possesses many structural and mechanistic commonalities with other bacterial GTs (CAZy). LgtC is also practically accessible as efficient protocols for its expression and purification have been developed⁵⁸. Importantly, it is one of the very few bacterial GTs with a fully solved crystal structure⁵⁴ making the rational design of inhibitors and probes for LgtC possible. For these reasons, LgtC is of high interest as a model system for the design of small molecular tools for exploring bacterial GTs involved in virulence.

LgtC from *Neisseria meningitidis* was the first retaining Leloir GT structure to be solved with a high resolution. More importantly, it was the first GT of any type with structural data for both substrates⁵⁴.

1.7.3. Structure of LgtC from *Neisseria meningitidis*⁵⁴

In 2001, the structure of LgtC from *Neisseria meningitidis* in complex with manganese and a donor analogue in the presence and absence of acceptor analogue was solved⁵⁴. The C-terminus, involved in the attachment of LgtC to the cell wall, is subjected to proteolysis.

For this reason, the 25 last residues of the protein were deleted, and the published structure is that of 286 residues. The deleted part is rich in basic and hydrophobic residues, suggesting that LgtC anchors to the membrane via hydrophobic and electrostatic interactions.

In order to prevent the catalysis from occurring, LgtC was crystallised with substrate mimics. Non-cleavable UDP-2FGal, with the hydroxyl at the position 2 of the galactose moiety substituted for a fluorine was used as donor analogue (Figure 1.10.C). The fluorine atom inductively destabilizes the oxocarbenium ion, reducing the reaction rate. UDP-2FGal is an inhibitor of LgtC with a similar binding to UDP-Gal but does not allow the gal transfer to occur. Similarly, 4'-deoxylactose, lacking the hydroxyl group receiving the Gal upon catalysis, was used as acceptor analogue (Figure 1.10.C). 4'-deoxylactose cannot function as a substrate for LgtC but does act as an inhibitor ($K_i = 16\text{mM}$). Its binding affinity is very similar to that of lactose ($K_m = 20\text{mM}$), indicating that interactions at the 4'-position are not crucial for binding.

The overall structure of LgtC was found to be that of a typical GT-A architecture. This monomer consists of fourteen α -helices and nine β -strands, and it has been organised into two domains: a large N-terminal domain (residues 1–247) and a smaller C-terminus (residues 248–282) (the last four residues are disordered) (Figure 1.10.A). The core of the N-terminal domain is a central seven-stranded sheet ($\beta_3, \beta_2, \beta_1, \beta_4, \beta_7, \beta_6, \beta_8$) surrounded by helices and contains the active site. The C-terminus is formed of only 33 residues, is mainly helical and mediates membrane attachment.

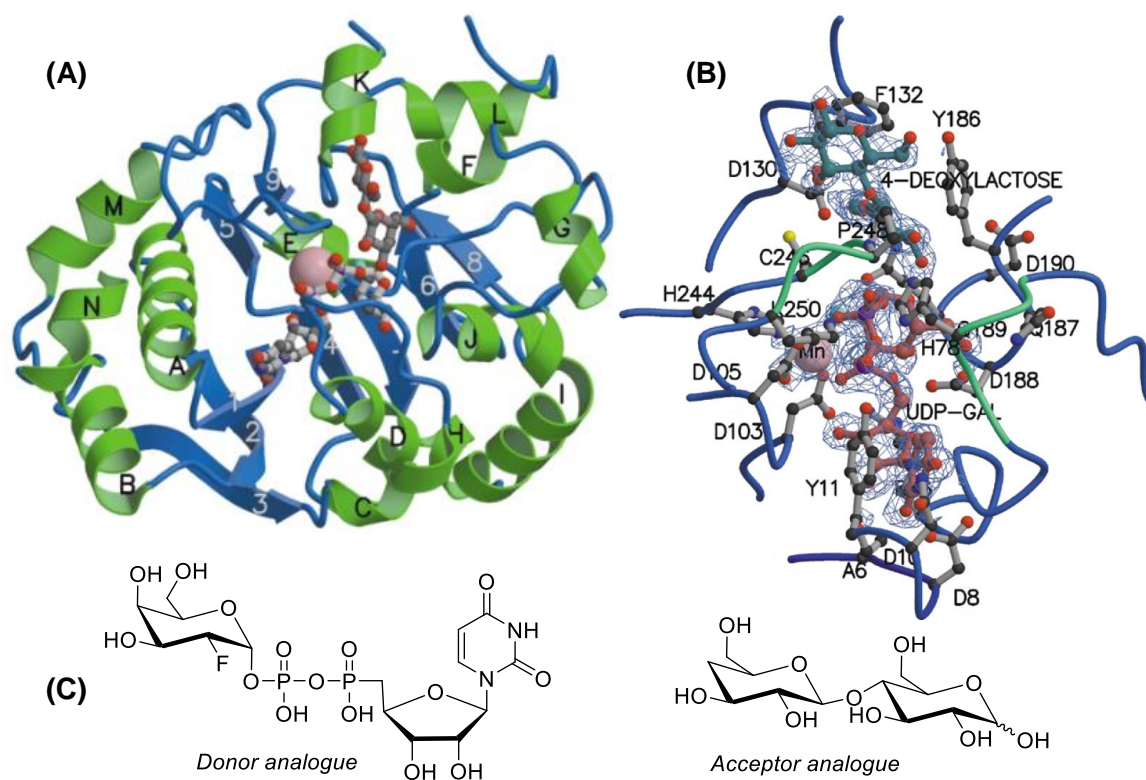


Figure 1.10⁵⁴. Solved crystal structure of LgtC from *Neisseria meningitidis*. (A) Stereo view of the LgtC structure with bound substrate analogs. The substrates are depicted in CPK representation: UDP-2FGal (light grey), 4'-deoxylactose (dark grey), Mn^{2+} (pink). Strands (blue), and helices (green) are labeled. (B) The donor sugar and the acceptor sugar lactose are shown, in stereo, as ball-and-stick models (with red and green sticks, respectively) in a refined $2F_o - F_c$ map contoured at 1.2σ . Amino acids interacting with the substrates are labeled. The loops that fold over the active site (residues 75–80 and 246–251) are colored green. (C) Structures of substrate analogues.

1.7.3.1. Donor binding site

First, the structure of LgtC with UDP-2FGal and Mn^{2+} was solved in the absence of acceptor analogue. Interestingly, the donor binding pocket was found in a deep, solvent shielded cleft of the enzyme with only 1.5% of UDP-2F-Gal exposed to water (Figure 1.11). This contrasts with the much shallower and more solvent exposed binding pockets identified at the time for inverting GTs. The donor mimic adopts an unusual folded conformation with a 160° angle between the galactose moiety and the phosphate groups (Figure 1.11). This was unlike the previously observed fully extended structures of donor substrates bound to inverting GTs. Most of the interactions between the enzyme and the nucleotide occur with residues located at the C-terminus of $\beta 1$ sheet and the N-terminus of helix A.

Additionally, two loops, from opposite sides of the structure, fold over the donor active site and form a tight envelope (Figure 1.10.B (shown in green) & Figure 1.11). In the absence of donor substrate, the authors claim that these loops would be disordered.

In the solved structure, a single Mn^{2+} is coordinated by the two phosphate oxygens of the donor analogue and by the side chain atoms of three residues: Asp103, His244, and Asp105. These three amino acids form the well-known D-X-D motif characteristic of GT-A enzymes. Manganese is required for the stability of the fold and activity of LgtC and other related glycosyltransferases. Unsurprisingly, the above D-X-D motif is conserved among members of GT family 8.

1.7.3.2. Acceptor binding site

The binding of the acceptor to the LgtC – UDP-2FGal complex results in no significant structural changes (root mean square deviation of 0.16 Å). However, the crystallisation of the LgtC-acceptor complex, in the absence of donor analogue, was not successful. This is consistent with an ordered bi-bi kinetic mechanism for LgtC in which the binding of the donor substrate initiates the movement of two loop regions, allowing for the formation of the full active site.

The acceptor is significantly more accessible to water than the donor with 28% of the entire molecular surface being exposed (Figure 1.11). It is bound in a large open pocket near the galactose moiety of the donor mimic. Both the non-reducing terminal galactose and the reducing glucose moieties of the lactose scaffold adopt a full chair conformation with the former interacting with Asp130 and Gln189 and the latter with Phe132 Pro211 and Pro248. Mutation of Asp130 to an Alanine residue drastically reduced protein expression, which indicates the importance of this residue for the structural integrity of the protein. Upon binding of the acceptor to the LgtC-donor complex, all hydrogen bonds between the donor and the enzyme are maintained, with additional bonds observed between the phosphate and both Tyr11 and the carbonyl of His78 *via* a water molecule.

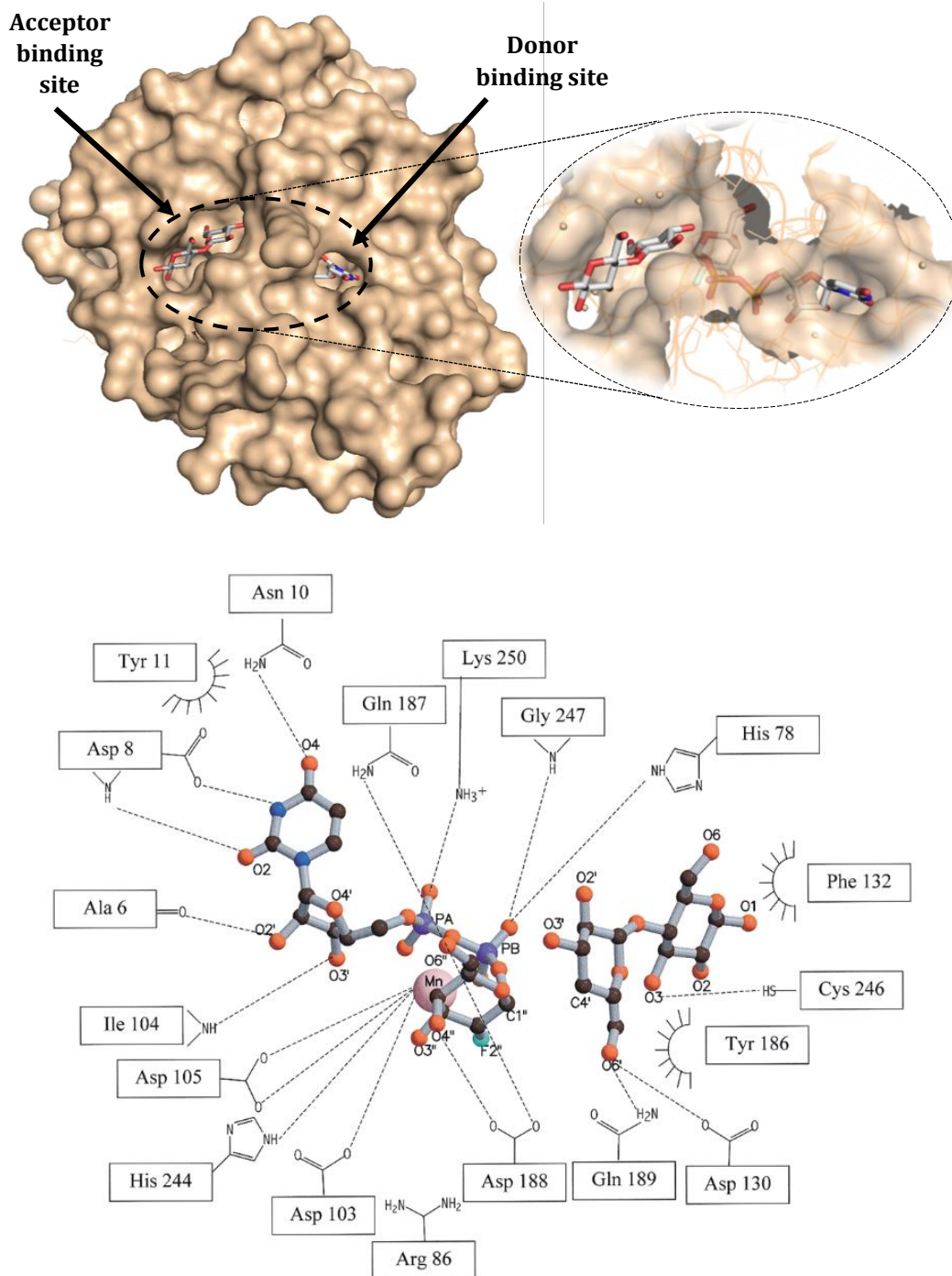


Figure 1.11. Top: Acceptor and donor binding pocket of LgtC (wheat). Bottom⁵⁴: Schematic representation of the interactions between the enzyme and the substrate analogs. Hydrogen bonds ($<3.1 \text{ \AA}$ for all bonds, except Cys $<3.5 \text{ \AA}$) are indicated by lines. Van der Waals contacts are shown as nested half circles. Water molecules have not been included.

In the donor-enzyme-acceptor complex, the side chain of Cys246 adopts a new conformation to form a hydrogen bond with the glucose moiety of lactose (at hydroxyl O3) (Figure 1.11). Additionally, in the donor-enzyme complex, an acetate ion (most likely

from NH₄OAc used as the crystallisation solvent) interacts with Asp130 and Gln189. Upon binding, the acceptor analogue displaces the acetate ion, and hydroxyl O6' from the galactose end of lactose now forms a hydrogen bond with Asp 130 and Gln189. As part of their attempts to decipher the catalytic mechanism of LgtC, the authors used 6'-deoxylactose in the LgtC reaction. No turnover, nor inhibition activity was observed for the molecule indicating that the above interactions are crucial for binding⁵⁴.

1.7.4. Catalytic mechanism of LgtC

Over the past decade, the retaining catalytic mechanism of LgtC from *Neisseria meningitidis* has been extensively studied^{59,60,61,62,63}. The elucidation of the crystal structure of LgtC in complex with both donor and acceptor analogues in 2001 provided key insight into the galactosyl transfer mechanism^{54,60}. Different hypothetical mechanisms were proposed over the past 15 years, but in 2012, after thorough kinetic and dynamic investigations, Withers *et al.* proposed that the retaining mechanism of LgtC involved a “front side attack, S_Ni-like mechanism with a short-lived oxocarbenium-phosphate ion pair intermediate”⁶¹. Their research also indicated that the protein would adopt multiple interconverting conformational states⁶². At the same time, Hansel Gomez and colleagues used density functional theory (DFT) combined with quantum mechanics / molecular mechanics (QM / MM) to study and compare different theories for the catalytic mechanism of LgtC and their conclusions were consistent with a S_Ni-like reaction as described by Withers⁶³.

The elucidation of LgtC's structure as well as our improved understanding of its catalytic cycle provide the foundation for the development of small molecule inhibitors of LgtC.

1.7.5. Inhibitors of LgtC

Only a limited number of inhibitors have been reported so far. The vast majority of GalT inhibitors found in the literature are non-covalent derivatives of the donor substrate given the strong recognition and binding of the nucleotide^{64,65,66}. Herein, a few examples of such inhibitors are described.

1.7.5.1. Donor-like LgtC-inhibitors

Fluorinated derivatives of donor substrates have been identified as good inhibitors of GTs⁶⁴. The replacement of hydroxyl group by a highly electronegative fluorine at the C-2 position of the sugar donor scaffold destabilises the oxocarbenium ion generated in the transition state and hinders the hydrolysis of glycosidic linkages (Figure 1.12). Non-cleavable UDP-2FGal, a good inhibitor of LgtC ($K_i = 2 \mu\text{M}$), was used for successfully solving the crystal structure of LgtC⁵⁴ (see 1.7.3).

Vidal *et al.* are highly interested in targeting the specific interaction of phosphate groups with divalent cations in GT-A enzymes. They design neutral pyrophosphate mimics for the development of GalTs inhibitors with increased cell permeability^{67,68,69}. They have developed a series of UDP-Gal- and UDP-Glc-based inhibitors using pyridine and triazole moieties as pyrophosphate surrogates (Figure 1.12). These inhibitor candidates were tested against five different GT-A enzymes including LgtC. For LgtC, the best candidates were both the α and the β anomers of the galactose-based analogue of **11** with IC_{50} values of $1597 \mu\text{M}$ and $546 \mu\text{M}$ respectively (Figure 1.12)⁶⁸. Compared to the IC_{50} value of UDP ($60 \mu\text{M}$), these inhibitors displayed only weak activity against LgtC, suggesting that the pyrophosphate bonds are critical for the binding of inhibitors to the enzyme. Although their inhibitor design strategy is interesting, further optimisation is needed in this domain for the design of more potent GalT inhibitors.

In 2010, Wagner *et al.* reported a base-modified UDP-Gal donor analogue **I3**, with a potent inhibitory activity against five GalTs, including LgtC ($K_i = 0.45 \mu\text{M}$) (Figure 1.12)⁷⁰. The characteristic structural feature of this novel class of UDP-Gal derivatives is the addition of a formylthienyl substituent in position 5 of the uracil base⁷⁰. The high-resolution crystal structure of **I3** bound to a representative mammalian GalT, AA(Gly)B, suggests that the additional substituent interferes with a π - π stacking interaction between Trp181 and Arg352 in the donor binding site⁷⁰. This prevents the folding of an internal loop and blocks the fully closed enzymatic conformation necessary for the catalytic activity (Figure 1.12)⁷⁰. The presence of **I3** locks the enzyme in an unreactive conformation and therefore, inhibits the galactosyl transfer.

The presence of a structural loop, essential for catalytic activity, was also identified in the crystal structure of LgtC (see 1.7.3). Additionally, sequence alignment of the five tested GTs showed that this flexible loop was a conserved motif⁷⁰, which suggests that this new mode of inhibition could be applied to a wider range of GTs⁷⁰. A structure activity relation (SAR) study was performed to explore the scope of this new mode of GalT inhibition⁷¹. Three structural analogues of **I3** were synthesised (Figure 1.12) and tested for inhibitory activity towards different GTs⁷¹ in order to gain a better understanding of the new mode of inhibition previously reported⁷⁰. **I4** and **I5** are modified at the R group while **I6** is modified at the pyrophosphate region for increased chemical stability and membrane permeability.

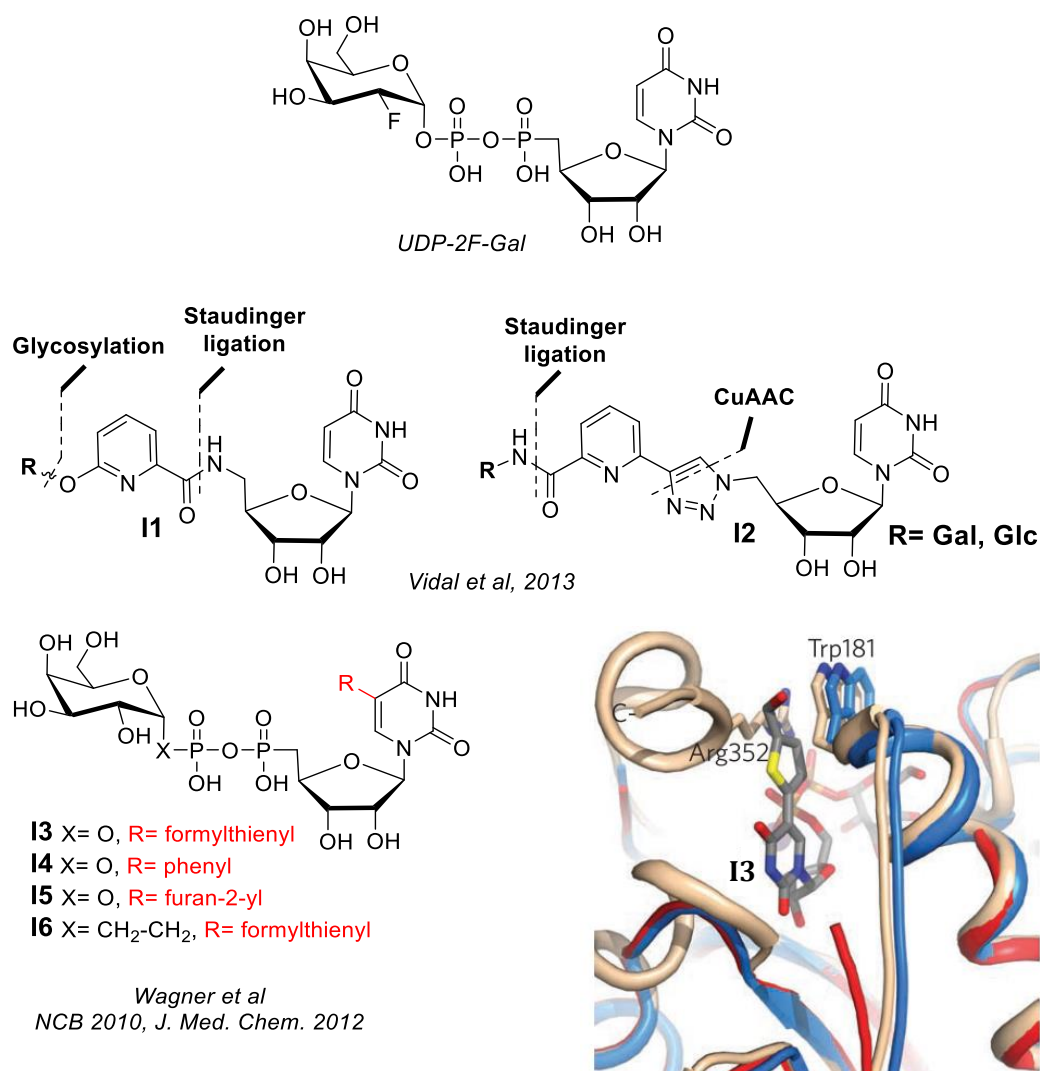


Figure 1.12. Examples of donor-like inhibitors of LgtC. **I1** and **I2** are pyrophosphate mimic inhibitors (Vidal *et al.*, 2013). **I3-6** are uracil modified analogues of UDP-Gal. Bottom right: X-ray structure of AA(Gly)B (GalT) in fully closed conformation (light brown), semi closed conformation (blue) and in complex with **I3** (red). Trp 181 in the internal loop, stacking to Arg352 in the C-terminus and ligand **I3** are shown in sticks⁷⁰.

Among the three different analogues, only **I6** retains a similar inhibitory activity against a panel of GTs, and more importantly against LgtC ($K_i = 1.7 \mu\text{M}$ for **I6** vs $K_i = 0.45 \mu\text{M}$ for **I3**). Computational simulations showed that the 5-substituent in **I3** and **I6** is well positioned to not only prevent the closing of the flexible loop as reported before⁷⁰, but also to form specific interactions with individual loop residue⁷¹ (intermolecular hydrogen bond between the formylthienyl group (H-bond acceptor) and Arg77 (H-bond donor) which is unique to **I3** and **I6**⁷¹).

1.7.5.2. Acceptor-like inhibitors of GalT and LgtC

On the other hand, far fewer acceptor-like inhibitor of GTs have been reported^{72,64}. Their mode of action is based on the chemical modification of the catalytic centre to prevent sugar transfer by means such as deoxygenation or fluorination. The methylation of the neighbouring carbon for induction of steric bulk is another employed strategy⁷³(Figure 1.13). 4-deoxylactose, which cannot act as a galactose acceptor, is an inhibitor of LgtC ($K_i=16\text{mM}$) and was used for the elucidation of the structure of the enzyme⁵⁴ (see 1.7.3). To the best of my knowledge, this is the only acceptor-like inhibitor of LgtC reported. However, such strategies do not always work. This is the case of β -(1-4)-GalT, for which the 4-deoxy substrate scaffold was not an inhibitor indicating that the hydroxyl group was essential for enzyme recognition. Instead, inhibition of β -(1-4)-GalT was achieved by decorating the acceptor substrate (Figure 1.13). It is to be noted that in most cases, acceptor-like inhibitors exhibit mM range potency.

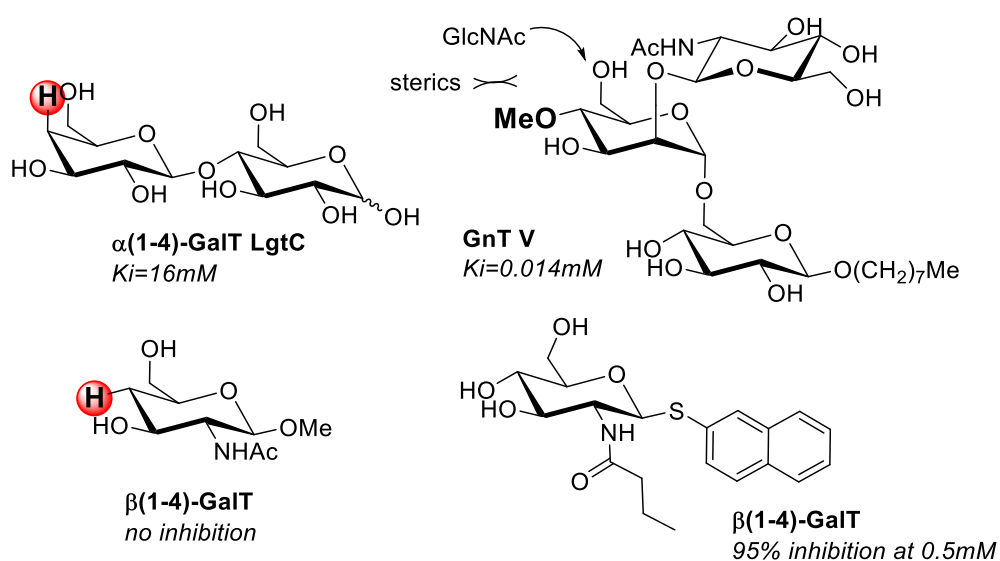


Figure 1.13. Examples of acceptor based inhibitor of GTs

In the case of LgtC, and for all GTs in general, far more donor-like inhibitors have been discovered than acceptor-based inhibitors. Non-substrate related inhibitors of GTs have also been reported for a relatively small number of GTs including LgtC⁷⁴.

1.7.5.3. A novel, non-substrate-related class of covalent inhibitors of LgtC

All LgtC inhibitors reported to date are non-covalent inhibitors. No covalent inhibitors had been described until recently, when a pyrazolone scaffold was identified from screening as a novel small molecular inhibitor chemotype of LgtC^{75,76} (Figure 1.14). The new inhibitors, which are structurally unrelated to both substrates of LgtC, have low micromolar inhibitory activity. Experimental evidence all converged towards a covalent mode of action towards LgtC which is supported by the presence of an electrophilic Michael acceptor in the inhibitor scaffold (Figure 1.14, red). Enzymological data and docking simulation strongly suggests a binding in the acceptor active site with irreversible enzyme modification at the non-catalytic Cys246⁷⁵ (Figure 1.14). An SAR study was performed with modification of the inhibitor scaffold on region A, B and C⁷⁶ (Figure 1.14). For region A, the small and electron withdrawing trifluoromethyl motif exhibited the best potency, most likely by increasing the reactivity of the electrophilic moiety without interfering with binding. Subsequently, alteration of ring B had little to no impact on the potency of the molecule while substitution of ring C significantly affected the binding of the inhibitor in the active site. These results are consistent with the docking simulations which predict ring B to be oriented away from the binding pocket towards solvent and ring A to be located towards the donor substrate. **17** was identified from the SAR study as the best molecular target ($IC_{50}=6.4\mu M$) for this class of inhibitors (Figure 1.14).

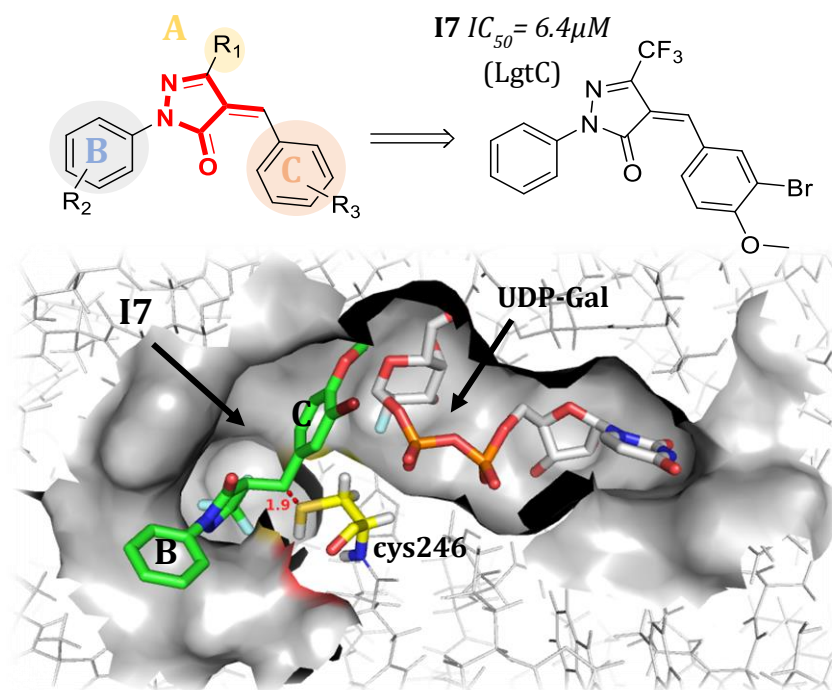


Figure 1.14. Structure of **I7**, covalent inhibitor of LgtC. *Bottom:* docking solution of **I7** (green) in binding pocket of LgtC (grey) for the targeting of Cys246 (yellow)

A covalent mode of inhibition is described by a two-step model: the non-covalent binding and the subsequent irreversible inactivation step⁷⁷ (see 3.1.1). The first step is quantitatively described by the inhibition constant K_i , while the inactivation rate k_{inact} characterise the second step⁷⁸. To understand the contribution of each part of the inhibitor scaffold for binding, the authors calculated the K_i and k_{inact} values for eight selected derivatives ($R_1 = \text{CF}_3$, modified at region C)⁷⁶. Interestingly, within the tested set of 5- CF_3 pyrazol-3-ones, k_{inact} values did not vary significantly while a considerable spread in K_i values was obtained (5–150 μM). This indicates that modification of region C does not impact on the rate of covalent bond formation. However, structural alterations at region C directly affect the reversible binding affinity of these inhibitors for LgtC. Therefore, these results suggest that the non-covalent binding drives the inhibition in this series of pyrazol-3-ones.

This is the first example of covalent inhibitor developed for LgtC. Analysis of available sequence and structural data reveals that non-catalytic cysteines are a common motif in

the active site of many bacterial glycosyltransferases⁷⁵. Therefore, targeting such cysteine may be a good strategy for the rational design of covalent inhibitors for a broad range of bacterial glycosyltransferases including potential therapeutic targets.

1.8. Irreversible enzyme modification: covalent drugs

Irreversible inhibition via covalent bond formation between the inhibitor and the target receptor is an effective strategy for the development of successful therapeutics^{78, 79, 80, 81, 82}. By targeting a specific residue in the protein of interest, covalent inhibition provide many therapeutic benefits such as greater potency and selectivity and longer biological half-life^{78, 80}. The high biochemical efficiency of irreversible inhibitors may require lower doses which reduces toxicity. They can also address undruggable targets with shallow binding sites⁸³. Despite these advantages, the potential for off-target reactivity which comes with a risk of toxicity has engendered anxiety among medicinal chemists^{78, 82}. There has been a clear reluctance from pharmaceutical companies to develop covalent therapeutics in drug discovery programmes⁷⁸. However, interestingly, aspirin, the oldest pain killer and the most widely used medication in the world, is a covalent drug^{78, 82}. Covalent drugs available on the market before 2010 were discovered by serendipity as structural hits from library screening⁷⁸ (Figure 1.15). In the past, this class of therapeutics was never designed to exhibit a covalent mode of inhibition and their molecular mechanism was elucidated long after their identification as a structure of interest⁷⁸.

Between historical success and recent scepticism, covalent drug design is currently undergoing a resurgence^{84, 85, 78}. Because of the pharmacodynamic and kinetic advantages described above, it is now recognised that drugs with a covalent inhibition mode have great potential and new approaches for the discovery of such therapeutics have recently emerged. A novel and rapidly developing strategy for the identification of selective covalent inhibitors, known as Targeted Covalent Inhibitors (TCIs), focuses on specific

non-catalytic nucleophiles⁷⁸, instead of targeting catalytic residues. TCIs function by exploiting non-covalent binding interactions to position a reactive warhead (WH) near a poorly conserved nucleophilic residue in the active site of their biological target⁸⁶. Such WHs, chemically attached to a carefully selected position of a non-covalent inhibitor, enable the subsequent irreversible modification of the protein, *via* the formation of a covalent linkage with the targeted residue. TCI candidates are now widely explored in drug discovery programmes and some TCI drugs have recently been approved by the FDA such as afatinib (Boehringer Ingelheim 2013) and neratinib (Pfizer 2017) (Figure 1.15). These examples are covalent inhibitors of tyrosine kinases, targeting a cysteine residue at the periphery of the active site. Such drugs are currently the frontline treatment in oncology⁸⁷ and have a major impact on human health as well as on pharmaco-economics⁷⁸. The expansion of TCIs has been facilitated by the recent development of advanced and specialized computational tools^{81,82} offering a new perspective to covalent drug design^{81,88,84,82}.

Interesting systematic studies have recently been published in which the authors have tuned the reactivity of electrophile warheads toward a non-catalytic residue of different enzyme targets^{81,89,90,91}. It has been proven that reactive electrophiles do not react indiscriminately with any biological nucleophile which suggests that selectivity of covalent drugs can be achieved. Such studies provide insights into the safety and efficacy profiles of covalent drugs and encourage the expansion of this therapeutic class.

More details regarding the design and potential applications of TCIs will be provided in chapter 3, alongside relevant examples.

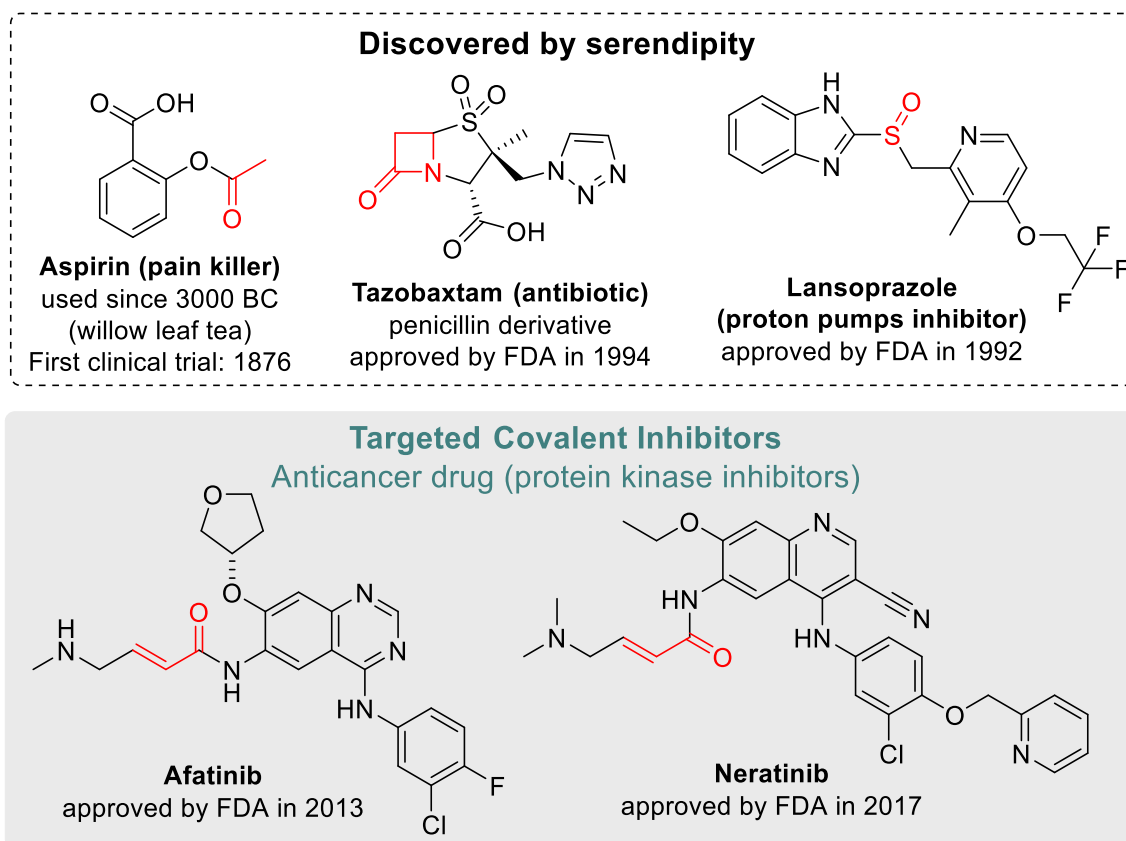
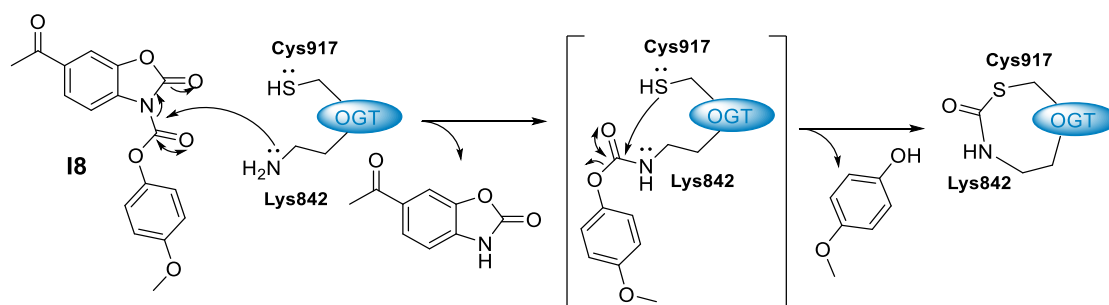


Figure 1.15. Example of covalent drugs. The site of covalent interaction is shown in red.

Despite the key role of GTs in virulence and infection and the great potential of covalent inhibitors, irreversible modifiers of bacterial GTs remain extremely rare to date, with only two known examples: **I7**, presented in the previous section, and **I8**, originally designed for a human GT (Scheme 1.3). **I8** was identified as a covalent modifier of eukaryotic O-linked N-acetylglucosamine transferase (OGT), a human enzyme responsible for the transfer of β -O-GlcNAc to different protein acceptors thereby modulating cell signalling pathways⁹². The five-heteroatom dicarbamate of **I8** acts as a neutral diphosphate mimic and is readily attacked by Lys842, an essential active site residue in the donor binding pocket. The resulting intermediate adduct is then displaced by the neighbouring Cys917 forming an irreversible thiocarbamate cross-link in the active site of OGT (Scheme 1.3).



Scheme 1.3. Mode of action of compd **18**, covalent modifier of human OGT⁹² (adapted from Walker *et al.*, NCB 2012)

Interestingly, **18** also had the ability to inactivate MurG, a key bacterial GlcNAc transferase with a similar structure to OGT⁹². Therefore, this study provides evidence that the rational design of covalent inhibitors on a model system can be successfully applied to a wider range of structurally related enzymes. Additionally, there is clearly a discrepancy between the great potential for covalent inhibitors for bacterial GTs and the current significant lack of such molecules.

1.9. Covalent probes and applications

The recent growing interest in irreversible modifiers for drug design purposes, and the associated increase in the development of covalent inhibitors, has provided a sophisticated molecular arsenal to enable the development of covalent chemical probes. In chemical biology, the design of small molecular tools, commonly called chemical probes, for direct analysis of enzyme activity in native biological systems has undergone explosive growth in recent years⁹³. This is an area where organic chemistry contributes immensely to chemical biology research. Usually derived from a covalent inhibitor scaffold, covalent probes have the ability to irreversibly attach to their target protein enabling a wide range of applications such as proteomics profiling^{94,95}, target identification^{96,97} and live cell labelling and imaging⁹⁸. Covalent probes can be classified into two categories: activity-based probes (ABPs) (also known as mechanism-based

probes) and affinity-based probes (A/BPs). The design of ABPs, structurally derived from substrate analogues, requires a full understanding of enzymatic mechanisms. To exert their labelling functions, ABPs require to be applied on catalytically active enzymes as they used the enzymatic machinery to form a covalent complex with their target. ABPs remain chemically inert and cannot function as covalent probes on inactive enzymes. In contrast, A/BPs do not require participation in the enzymatic mechanism. In this case, labelling is achieved by binding and subsequent non-specific covalent attachment at a specific site on a protein (not necessarily an enzyme active site). The distinction resides in the need for a catalytically-active enzyme: ABPs are dependent on a catalytically active enzyme while A/BPs are not⁹⁹.

Such chemical probes are carefully designed and contain a reactive warhead (WH) attached to a recognition motif for the binding and covalent modification of its biological target. They also possess a reporter tag, to facilitate target characterization, such as fluorophores, biotin (1-step labelling) or latent analytical handles such as alkynes or azides, which can be modified by click chemistry methods to visualize protein targets post-labelling (2-step labelling). A linker region is also present to provide enough spatial separation between the reporter tag and the recognition motif, while also modulating the reactivity and specificity of the reactive / binding group.

Activity- and affinity-based probes provide a “chemoproteomic” means for the detection and monitoring of specific proteins in a complex sample (cell extract, living cells or animal models). Activity-based protein profiling (ABPP) has recently emerged as a powerful technique allowing the global profiling and quantitative readout of specific enzyme families in complex proteomes. As this field moves into the post-genomic era, scientists are facing exciting challenges to decipher the emergent properties of the genome at the molecular level. Small molecular probes for the interrogation of the proteome are a mean towards achieving this. This will be discussed further in Chapter 5.

Additionally, ABPP has begun to find applications in the study of pathways involved in pathogenesis as beyond their use to perturb protein function, small molecular probes can also be applied to profile the activity levels of enzymes involved in different stages of virulence processes^{100,101}. More specifically, because of the relevance of carbohydrates in bacterial virulence, glycoprobes, for the labelling of sugar binding proteins, are powerful tools for exploring biological mechanisms underlying pathogenicity¹⁰². While a sizable number of chemical probes have been developed for GHs¹⁰³, only a limited number of molecular tools are available for GTs due to their complex multi-substrate mechanisms, their conformational plasticity¹⁰⁴ and the lack of structural information for this class of enzyme. A more detailed review on available chemical probes for GHs and GTs will be provided in Chapter 4.

1.10. Aims and objectives

With the great challenges posed by AMR, there is an urgency to develop alternative antimicrobial strategies. Chemistry, *via* the synthesis of small molecular tools for the study of virulence and pathogenesis, can have a powerful impact on understanding how bacteria cause disease and spread resistance, which is essential for the discovery of efficient antibiotics. In pathogenic Gram-negative bacteria, for which inherent resistance is observed due to their physiology, surface glycoconjugates play a key role in pathogen / host interactions. LOS have been identified as a virulence factor, responsible for host immune evasion, and small molecular tools able to modulate their biosynthesis are of great interest for the development of anti-virulence agents. In *Neisseria meningitidis*, LgtC, a retaining galactosyltransferase, is responsible for the formation of a virulent digalactoside epitope in the LOS envelope and plays a key role in bacterial camouflage and immune suppression. Small molecular inhibitors of LgtC are therefore of great interest as chemical tools to study glycosylation pathways involved in bacterial virulence and pathogenicity. LgtC itself is, however, not considered as a great therapeutic target due to

the significant phase variability associated with its expression in pathogenic Gram-negative bacteria. On the other hand, the practical accessibility of LgtC combined with its fully solved structure and its commonalities with other bacterial GTs, makes it an ideal model system for the study of bacterial GTs *via* the rational design of inhibitors and probes.

Currently, there is only a limited number of LgtC inhibitors available, most of which are derived from the enzyme donor substrate. Recently, a non-substrate related inhibitor of LgtC was reported and evidence strongly suggests that it covalently modifies Cys246, a non-catalytic residue in the acceptor binding site of the protein. Despite the great potential of covalent inhibitors for the development of drug with increased potencies and prolonged effects, this is only the second example of a covalent inhibition for bacterial GTs. Although these examples have created an opportunity for the development of a new class of potent inhibitors for bacterial GTs, both current inhibitors have a flat, drug-like and non-substrate related scaffold which does not allow for an efficient occupation of the 3D binding pocket, nor takes advantage of the great selectivity of GTs for carbohydrate scaffolds. Additionally, the fully solved crystal structure of LgtC enables the rational design of inhibitors which remains to date under-exploited.

The aim of the project is the development of rationally designed carbohydrate-based covalent inhibitors of LgtC for the target of the non-catalytic Cys246 in the acceptor binding pocket. To achieve maximum selectivity, the acceptor substrate of LgtC, lactose, will be equipped with an electrophilic warhead to convert the molecule to a covalent inhibitor.

Towards this goal, three objectives were pursued:

- Rational design by covalent docking¹⁰⁵ for the identification of suitable positions on the lactose scaffold for the attachment of an electrophilic warhead
- Synthesis of the best chemical targets¹⁰⁶ identified through docking
- Biochemical testing¹⁰⁷ of synthesized molecules against recombinant LgtC

Sequence alignment and structural data comparison reveals that non-catalytic cysteines are a common motif in the active site of many bacterial glycosyltransferases. Therefore, converting the substrate scaffold to a covalent inhibitor by attachment of a chemical warhead for the targeting of such cysteines may be a good strategy for the rational design of covalent inhibitors for a broad range of bacterial GTs. For reasons mentioned above, LgtC is an excellent model for exploring this approach. This project may therefore serve as a proof of concept study towards rational design of covalent inhibitors for bacterial GTs involved in virulence.

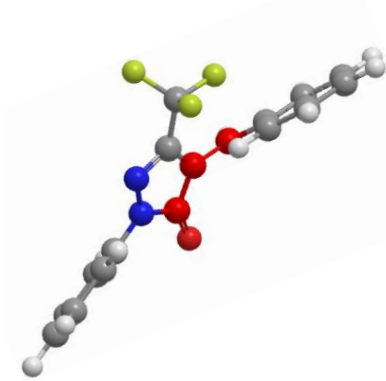
Covalent inhibitors could be highly useful as tool compounds for anti-virulence strategy in inhibition studies. However, another key aspect which can contribute to developing efficient therapeutics is the identification of resistance and virulence markers. Therefore, covalent inhibitors can be chemically tuned into affinity probes for the labelling, imaging and identification of bacterial GTs *in vivo* and in cell lysates. Towards this goal, carbohydrate-based probes are of high interest as they allow the selective targeting of carbohydrate-binding proteins, something hardly achievable with standard drug-like probes.

A secondary aim of the project is the synthesis of covalent carbohydrate-based probes for the labelling of LgtC and carbohydrate binding proteins recombinantly and *in vivo*.

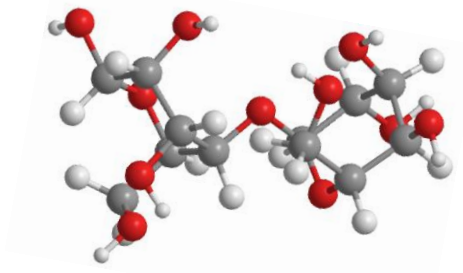
The description of this work will be divided into four experimental chapters (Figure 1.16):

- Chapters 2 and 3 describe the rational design of an acceptor-like covalent inhibitor of LgtC. The design relies on the chemical attachment of an electrophilic warhead at a specific position of the lactose acceptor scaffold for the covalent targeting of cys246 in the acceptor binding pocket of LgtC. The synthesis, characterisation and mode of action of a series of carbohydrate-based inhibitors of LgtC equipped with a chemical warhead will be discussed.
- In Chapter 4, a covalent probe derived from the inhibitor scaffold is synthesised and fully characterised. The probe is used for the tagging of purified LgtC and in cell lysate and *in vivo* labelling.
- Chapter 5 expands the applicability of the synthesised carbohydrate-based probe to more complex *in vivo* proteomics study on a pathogenic strain of *Haemophilus influenzae*.

Current covalent inhibitor of LgtC

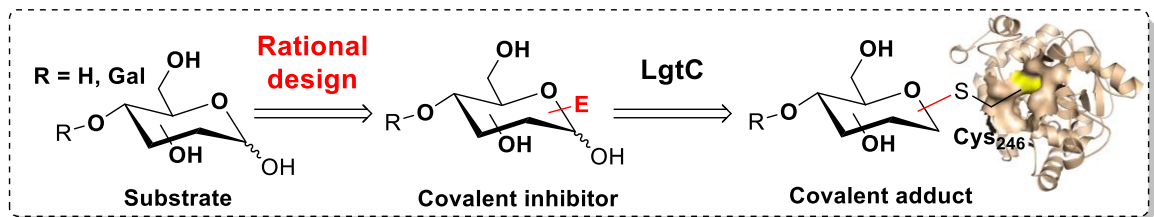


Carbohydrate-based inhibitors

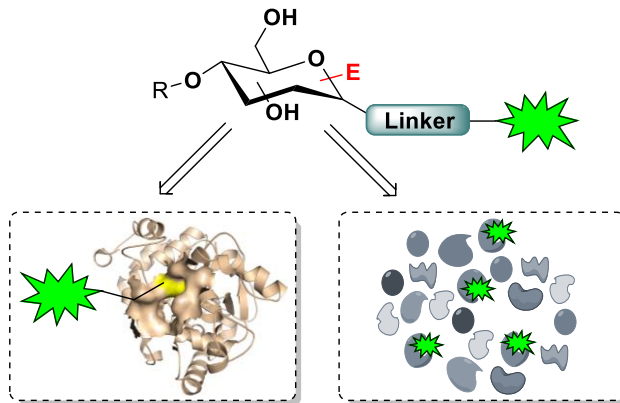


- ✓ Can fully exploit shape of the binding pocket
- ✓ Increased selectivity

Chapter 2 and 3 Design of carbohydrate-based covalent inhibitor of LgtC



Covalent probe design



Chapter 4

Design of carbohydrate-based covalent probe for the labelling of LgtC

Chapter 5

Carbohydrate-based covalent probe for the interrogation of the bacterial proteome

Figure 1.16. Illustration of the project objectives: structure of the thesis

Chapter 2

Disaccharide analogues as a scaffold for the covalent design of LgtC inhibitors

In this chapter, the rational design for covalent inhibition, chemical synthesis and initial biochemical testing of a disaccharide series LgtC-inhibitors is discussed. Before presenting the results, we will discuss carbohydrate scaffolds as a template for the design of potent inhibitors and illustrate this with a few relevant examples.

2.1. Introduction

2.1.1. Carbohydrates as a template for inhibitor design

Carbohydrates have been identified as a rich source of inspiration for the design of GT inhibitors⁶⁶. A great example is moenomycin, a carbohydrate-based natural product which, along with its derivatives, forms the only known group of antibiotics that inhibit bacterial peptidoglycan glycosyltransferases, involved in the penultimate step of bacterial cell wall biosynthesis¹⁰⁸. Its structure consists of a pentasaccharide and a chromophore group linked to a phosphoglycerate-lipid which mimicks the lipid IV chain of PG, the donor substrate of the enzyme (Figure 2.1.A). Effective against a variety of Gram-positive bacteria, moenomycin is one of the most active antibiotics (MIC of 0.05 µg/ml for *S. aureus*, 10–1000 times better than vancomycin)¹⁰⁹. However, it forms part of a relatively underexploited class of antibiotics due to its high lipophilicity and the unfavourable pharmacokinetic properties associated with this, which renders it unsuitable for systemic administration¹⁰⁸. On the other hand, the low rate of resistance observed for moenomycin derivatives has revived the general interest in this underused class of antibiotics following the rise of multidrug-resistant pathogens. Recently, a novel monosaccharide scaffold which mimics the essential structure features of moenomycin was reported¹¹⁰ (Figure 2.1.A). This was created based upon the structures of both moenomycin scaffold and previously reported GT sugar-based inhibitors¹¹¹. With a smaller, more drug-like, hydrophobic tail, the novel class of compound exhibits *in vitro* inhibition comparable to moenomycin, with low toxicity and good efficacy in several *in vivo* models of infection¹¹⁰.

Kdo analogues are another good example of how carbohydrate can be used as templates for inhibitor design. Kdo, an ulosonic acid, is a critical component of the LPS envelope in Gram-negative bacteria (Figure 2.1.B). The biological activity of synthetic Kdo analogues has been tested over the years, highlighting the potential of this pathway for the discovery of new antibiotics¹¹². The breakdown of Kdo biosynthesis is believed to result in accumulation of lipid A precursors, resulting in inhibited cell growth and reduction in pathogenicity, rendering the bacteria more susceptible to antibiotics. A study has shown the *in vitro* synergistic effect of Kanamycin and Fosfomycin with Kdo synthase inhibitors as antibiotic adjuvants, on the production and release of Vero toxins by *E. coli* O157¹¹³. With a short peptide sequence attached on their scaffold, this series of Kdo analogues was designed to permeate the cell through an oligopeptide permease system and be subsequently metabolised *in vivo* (Figure 2.1.B). While Kdo synthase inhibitors and antibiotics have no biological effects on their own, both antibiotics demonstrated remarkably strong inhibition of Vero toxin's release through synergistic collaboration with KDO synthase inhibitors¹¹³. This series showed good antibacterial activity against *Salmonella* and *E. coli* strains. More interestingly, virulent strains of *Salmonella typhimurium* became non-virulent by inhibition of Kdo synthase, due to lipid A precursor accumulation in the outer membrane, causing structural perturbations. However, this mechanism of uptake provided a route to resistance (e.g. transporter mutations), and these compounds have not yet found clinical applications. Very recently, a concise route for the synthesis of Kdo analogues, comprising a tetrazole moiety as bioisosteric replacement for the carboxylic acid group was reported¹¹⁴ (Figure 2.1.C). Tetrazole compounds are more lipophilic and readily pass the phospholipid bilayer of biological membranes. Therefore, this may circumvent the observed resistance of Kdo analogues due to transporter mutations and allow such candidates to potentially find clinical applications in the future.

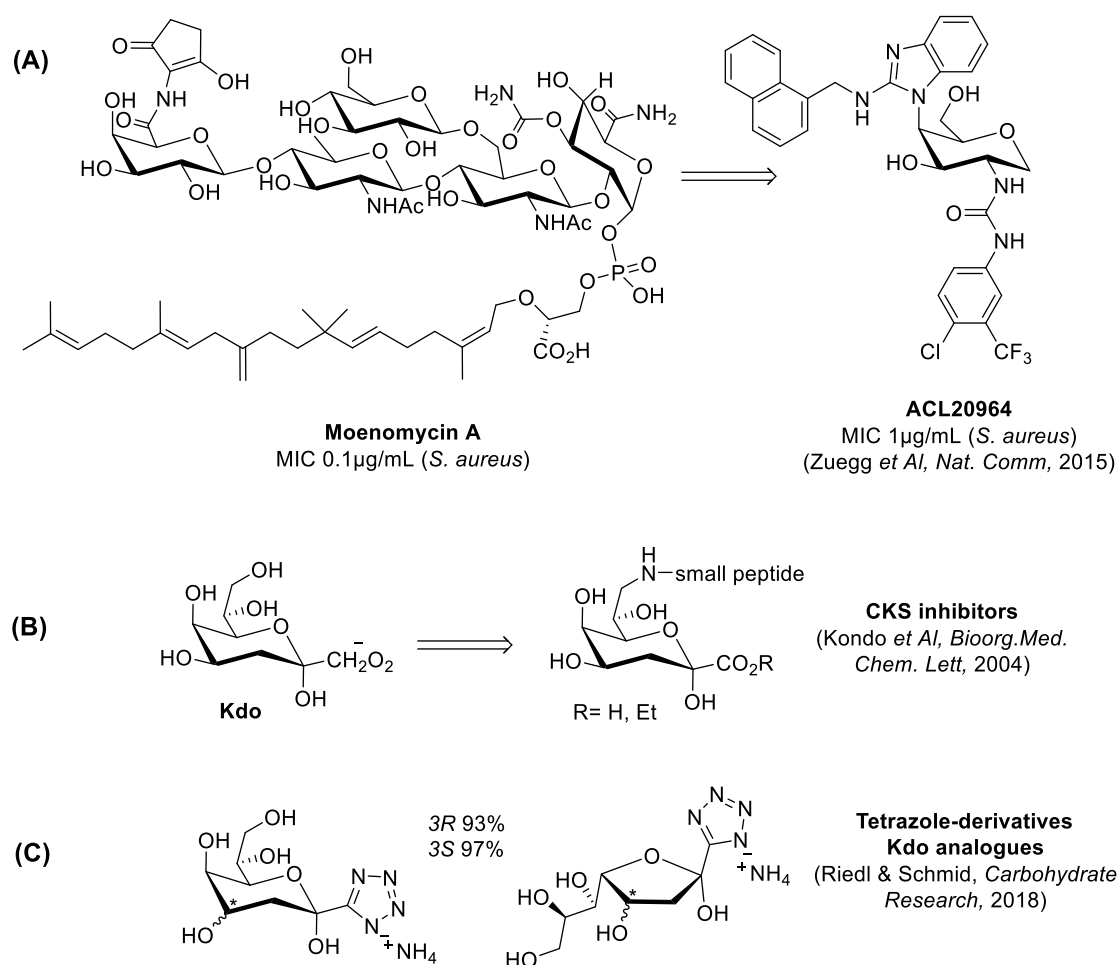


Figure 2.1. Examples of carbohydrate-based inhibitors of bacterial carbohydrate-active enzymes involved in resistance and virulence.

The above examples, alongside many others⁶⁶ indicate that sugar-based molecules are a great template for the design of potent carbohydrate-active enzyme inhibitors. Conventional “drug-like” non-carbohydrate screening libraries, with their planar and stereochemically indistinct structures, have failed to mimic the natural properties of the glycospace and therefore have had limited success against carbohydrate-active targets¹¹⁵. On the other hand, non-planar carbohydrate-based inhibitors are able to mimic the shape and binding of the substrate they replace which increases the level of interaction with their biological target. On top of this, to reach the desired biological effect, the inhibitor also needs to contain additional groups rendering them structurally different from the natural substrate and allowing target inhibition.

In the context of this project, for the rational design of carbohydrate-based covalent inhibitor of LgtC, an electrophilic warhead will act as the additional moiety required for inhibition and anchoring.

2.1.2. Targeted covalent inhibitors: design and carbohydrate-based examples

Targeted covalent inhibitors (TCI) are a recent and rapidly developing approach towards the design of covalent inhibitors. As described in Chapter 1, the development of TCIs relies on a structure-based design by exploiting non-covalent binding interactions to position a reactive warhead (WH), whose position on the inhibitor scaffold has been carefully selected, near a poorly conserved nucleophilic residue in the active site of their biological target⁸⁶. This strategy enables the subsequent irreversible modification of the enzyme, *via* the formation of a covalent linkage with the targeted residue (Figure 2.2).

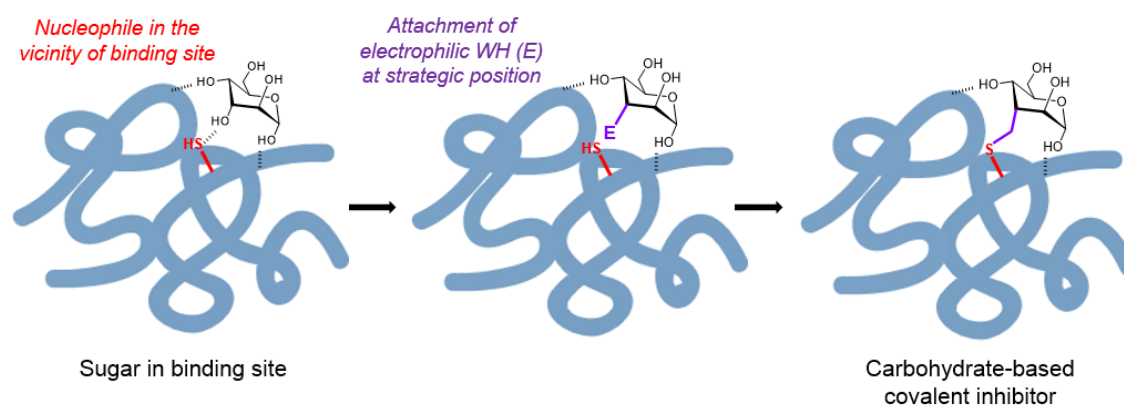


Figure 2.2. Structure-based design of TCI

Acrylamides, which are Michael acceptors, haloacetamide and epoxide motifs are among the most commonly used WHs. They typically modify cysteine residues, with examples of threonine and histidine modification reported for epoxide containing inhibitors (see 3.1.2). Recent *in vitro*⁹¹ and computational⁸¹ studies have shown that the initially feared promiscuous reactivity and off-target selectivity of electrophilic WHs may well be overestimated. A systematic study has tested a combinatorial library of 72 fragment-like

compounds from the reaction of six electrophilic moieties (including chloroacetamide and acrylamide WHs) with eleven unreactive moieties, against eleven enzymes from viral, bacterial and human origin⁹¹. The aim of the study was to assess the selectivity and feared promiscuous binding of electrophiles to proteins. Interestingly, the authors found that the degree of promiscuity was significantly lower than anticipated. Acryl- and chloroacetamides specifically are believed to be highly reactive motifs and exhibit a surprisingly low non-specific reactivity. This indicates the need of a specific scaffold for target recognition allowing a strong non-covalent binding from which the irreversible modification can occur. These results suggest that chloroacetamide and acrylamide are highly suitable WHs for the design of TCIs⁹¹.

Recently, Titz *et al.* have developed the first cysteine-targeting covalent inhibitor of carbohydrate binding sites by rational structure-based design. They report a galactose-based covalent inhibitor of LecA, a virulent bacterial lectin from *P. aeruginosa* involved in cell adhesion and biofilm formation¹¹⁶ (Figure 2.3.A). The epoxide-containing sugar targets Cys62 in the carbohydrate binding domain. This elegant work combines the idea of exploiting the glycospace for the design of inhibitors for carbohydrate-active proteins, with the attachment of an electrophilic WH at a key position for the targeting a cysteine residue. A derived fluorescent probe was used for the LecA-specific *in vitro* imaging of biofilms formed by *Pseudomonas aeruginosa*¹¹⁶ (see 5.1.2.2). Very recently, the same group has developed a mannose-based inhibitor for LecB from *P. aeruginosa*¹¹⁷, a less conserved protein among *P. aeruginosa* strains than LecA and directly involved in biofilm formation¹¹⁸ (Figure 2.3.B). The molecule was a potent inhibitor of biofilm formation without affecting bacterial viability, making it an attractive candidate towards anti-virulence therapeutics with reduced associated resistance. Despite its acrylamide-containing structure, no covalent inhibition properties were reported for this inhibitor, however residence times to target were significantly increased¹¹⁷.

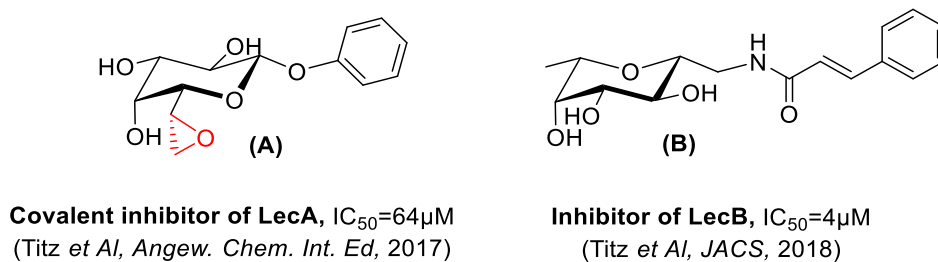


Figure 2.3. Inhibitors of bacterial lectins (reported by Titz *et al.*)

To the best of my knowledge, the covalent inhibitor of LecA designed by Titz *et al.* is the only example of TCI design applied to a carbohydrate scaffold for the targeting of carbohydrate-binding proteins involved in virulence¹¹⁶. However, this strategy has not yet been successfully applied to proteins with catalytic activity.

2.1.3. Docking programmes for TCI design

The growing interest in covalent drugs has resulted in a surge in the development of computational tools for the design of TCIs¹¹⁹. Although still significantly more developed for non-covalent drug design, docking programmes have evolved towards facilitating TCI design by enabling the rational selection of positions within the inhibitor scaffold for the attachment of an electrophilic trap. Docking programs predict binding and affinity of ligands and rely on two fundamental steps: the sampling of the conformational space of the ligand-protein complex, and the ranking of these conformations *via* a scoring function to predict the likeliness of a binding mode. Algorithms have recently been modified to accommodate bond formation between the target and the ligand, resulting in the identification of covalent poses which can be scored and ranked. CovalentDock^{82,105} is able to screen a library consisting of both covalent and non-covalent binders but only supports two covalent modifications: Michael addition (on cysteine residues) and β -lactam ring opening (on serine residues). CovalentDock has been used to discover novel Michael acceptor-containing inhibitors of RecA, involved in growth and survival of *Mycobacterium*

*tuberculosis*¹²⁰. The docking program has also been exploited for the study of 76 covalent complexes, yielding an average root mean square deviation of 1.68 Å¹¹⁹.

2.1.4. TCIs against drug resistance

Covalent drugs may find a powerful application for the development of efficient treatments for infectious diseases caused by drug-resistant pathogens^{78,83}. TCIs offer the possibility of dual inhibitory mechanism which confers such drugs with a great potential for clinical success. In addition to the high affinity reversible occupancy of the binding site, irreversible target modification provides a second mechanism for inhibition should the pathogen attempt to evade the effect of the drug by mutation. Irreversible modifiers maintain activity against mutation acquired after treatment with traditional inhibitors as sufficient exposure to the covalent drug will result in fully inhibited targets even for resistant mutants reacting at a slower rate⁷⁸. Additionally, repeated periods of temporary enzyme inhibition most likely promote the development of resistance mutations. As opposed to this, permanent target inactivation may prove extremely efficient towards preventing the emergence of resistance. Recent advances in Hepatitis C virus (HCV) research provide a great example of how irreversible inhibitors can be applied for the treatment of infectious diseases that have developed resistance to existing therapies. As a matter of fact, a novel irreversible Hepatitis C Virus protease inhibitor, active against clinical mutants resistant to standard treatment, has recently been reported⁸⁹.

Additionally, a covalent inhibitor of transcription factor LasR, a key component in the initiation of quorum sensing in *Pseudomonas aeruginosa*, has recently been reported¹²¹ (Figure 2.4). Several potent reversible inhibitors of LasR had been identified prior to this¹²², but the high affinity of quorum sensing messenger 3-oxo-C12-HSL, the natural ligand of the protein (also known as OdDHL), had always been an obstacle for the discovery of efficient inhibitors. This novel inhibitor is a mimic of 3-oxo-C12-HSL with an additional tag targeting the covalent modification of Cys79 near the binding site (Figure

2.4). The irreversible modification of LasR results in the inhibition of gene expression regulated by quorum sensing therefore preventing the production of virulence factors and biofilm formation^{121,123}. This was the first example of covalent inhibition of a quorum sensing receptor and provides a stepping stone towards novel strategies to combat antibiotic resistance as well as a tool to further unravel the complicated quorum sensing regulation in *P. aeruginosa* and other bacterial pathogens.

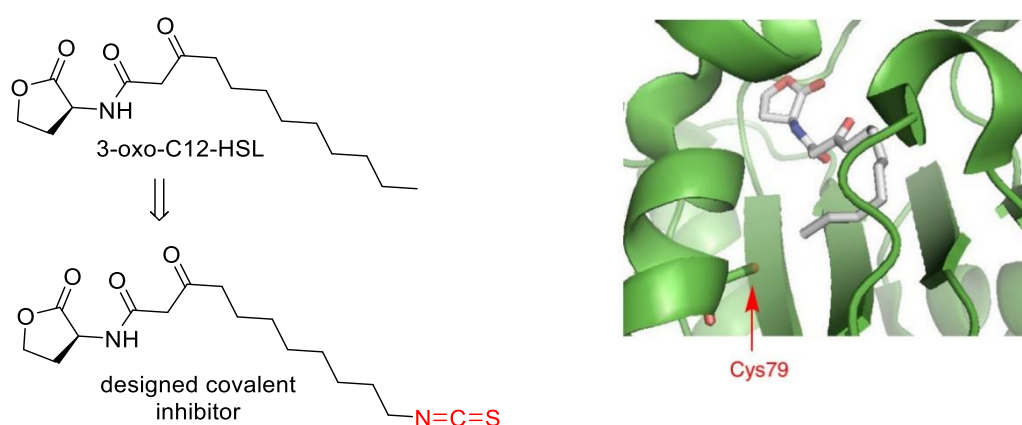
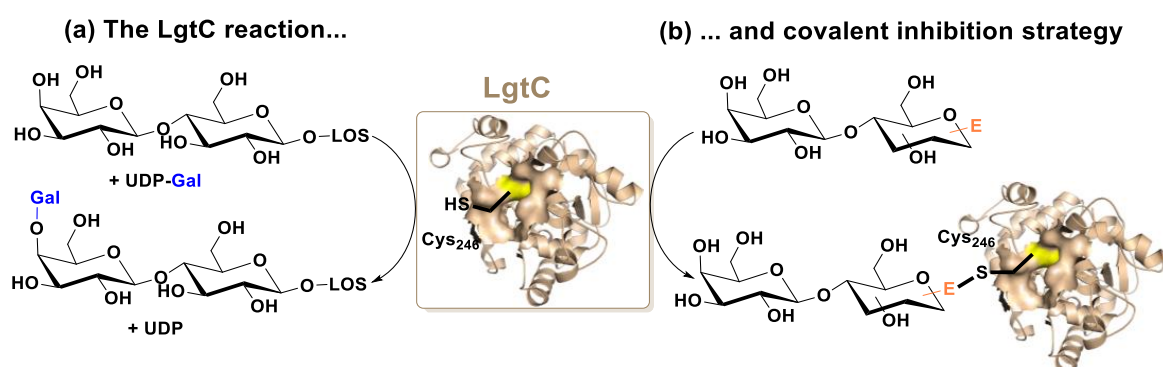


Figure 2.4. Left: TCI for LasR in *Pseudomonas aeruginosa* based on natural ligand scaffold (3-oxo-C12-HSL) Right: 3-oxo-C12-HSL (grey) in binding pocket of LasR (green)

TCI typically relies on the targeting of non-conserved amino acids as a strategy for achieving selectivity for covalent inhibitors. However, the rarity of the targeted residue may indicate that it is not essential for the function of the protein, therefore providing a route towards developing TCI resistance. On the other hand, in bacteria, many non-catalytic residues are well-conserved among an enzyme family which suggests an important yet obscure role in the fitness of the organism⁷⁸. Targeting such conserved non-catalytic residues may therefore prove useful for the design of potent and effective inhibitors for bacterial GTs.

2.2. Objectives

Recently, a non-substrate based covalent inhibitor of LgtC from *Neisseria meningitidis*, a galactosyltransferase involved in virulence, was identified from screening⁷⁵ (see 1.7.5.3). The inhibitor was shown to target cys246, a non-catalytic residue in the acceptor binding site of the enzyme. Inspired by these findings, the aim of the next two chapters is to design carbohydrate-based covalent inhibitors of LgtC for the targeting of Cys246 in the lactose acceptor binding pocket of the enzyme. In this first chapter, we will investigate if the lactose scaffold can be converted from LgtC-substrate to LgtC-inhibitor targeting Cys246, upon chemical addition of an electrophilic warhead (Scheme 2.1). Unlike the previous project, in which the inhibitor was discovered by serendipity, the work described here relies on a rational approach as the target molecules are investigated by covalent docking.



Scheme 2.1. (a) LgtC reaction. LgtC catalyses the transfer of galactose from UPD-Gal donor to the terminal lactose of the lipooligosaccharide envelope. (b) Covalent inhibition strategy. Electrophilic lactose-like molecules are designed to covalently target Cys246, a non-catalytic residue located in the acceptor binding site. If the strategy is successful, a covalent adduct is formed *via* the irreversible condensation of Cys246 with E.

Three objectives will be pursued:

- 1) Structure-based design of electrophilic lactose-like compounds and identification of suitable positions on the lactose scaffold for the attachment of an electrophilic warhead by covalent docking¹⁰⁵ (see 2.3)
- 2) Synthesis of best chemical targets¹⁰⁶ identified through docking (see 2.4)

3) Biochemical testing¹⁰⁷ of synthesized molecules against LgtC (see 2.5)**2.3. Structure-based design: Covalent docking**

For the development of lactose analogue covalent inhibitors of LgtC, suitable positions of the scaffold for the attachment of an electrophilic warhead must be identified. This was investigated using the CovalentDock Cloud interface¹⁰⁵ (see 2.1.3). The structure of LgtC from *Neisseria meningitidis* solved in 2001 (PDB 1GA8) gave insight into its catalytic mechanism and provided the opportunity for the rational design of covalent inhibitors⁵⁴. A specific interaction between Cys246 and the glucosidic hydroxyl O₃ of lactose was reported⁵⁴ (Figure 1.11, chapter 1). This indicates the proximity of the Cys246 residue and the glucose moiety of lactose within the acceptor binding pocket and allows for the rational selection of candidates to be tested in docking simulations. Supposing that the candidates were to adopt a similar binding as the natural acceptor, three positions of the glucose moiety of lactose appeared well places and were tested for their suitability for the attachment of the electrophilic WH (the anomeric position 1, and positions 2 and 3) (Figure 2.5).

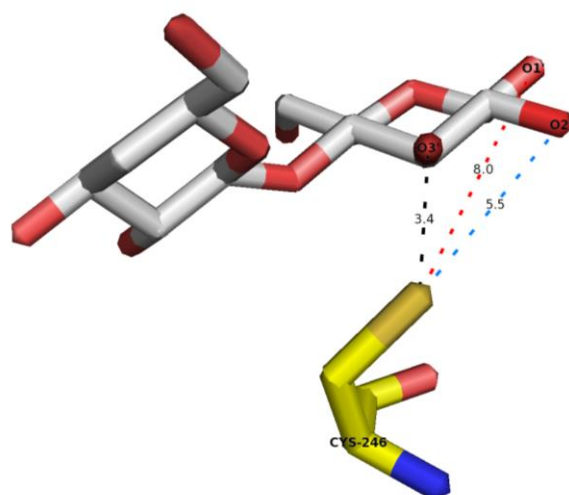


Figure 2.5. Three positions on the glucose moiety of lactose were tested for their suitability for the attachment of the electrophilic WH. Distance between hydroxyl group and thiol of cys246 (Å): 8.0 (OH-1), 5.5 (OH-2), 3.4 (OH-3)

For the covalent targeting of a cysteine residue in the receptor, the algorithm in place can only simulate a Michael addition. A small library of six lactose and lactosamine derivatives was therefore docked against LgtC, with a Michael acceptor motif attached at the three tested positions *via* an ester or an amide linkage (Figure 2.6). Both the PDB and the ligand files were prepared as described in the experimental chapter (see 7.1).

For each docking the program provided the following information:

- The number of docking solutions (also called clusters), indicating the poses adopted by the ligand into the binding site with respect to Cys246.
- The free energy which indicates the thermodynamic stability of each solution. The lower the free energy the more stable the ligand is in the binding site of the target protein.

For our six tested candidates the results of the docking simulations are listed in Table 2.1.

Table 2.1. Results from docking simulations.

Electrophilic WH (position-linkage)	Nb of clusters	ΔG range (kcal/mol)
1-NH	9	From -7,17 to -4,41
2-NH	3	From -10,45 to -5,48
3-NH	8	From -7,05 to -4,02
1-OH	8	From -7,40 to - 4,24
2-OH	6	From -7,07 to -3,99
3-OH	7	From -6,62 to -3,69

Compound **1**, with the electrophilic warhead attached in position 2 of lactosamine, (Figure 2.6, Table 2.1 highlighted in red) was identified as the best target for our strategy because of its minimal number of possible poses, lowest free energy and ideal position of electrophilic WH in relation to cys246. The binding calculated for **1** showed a high level of qualitative similarity to the binding of lactose, the natural acceptor substrate (Figure 2.6). This is an indication that our inhibitor candidates may show high affinity for the acceptor binding site of LgtC in which the cysteine of interest is located. Additionally, the docking solution of **1** indicates that the reactive end of the electrophilic warhead and the

nucleophilic thiol of Cys246 are separated by 1.8 Å which is ideal for a covalent bond formation (Figure 2.6).

From a synthesis perspective, the modification of a single glucosidic hydroxyl over others may be challenging because of selectivity issue. On the other hand, the greater nucleophilicity of amines over hydroxyls can be exploited for less challenging synthetic strategies. With the warhead attached *via* an amide bridge, the resulting inhibitor candidate would also show greater stability in solution as it would be much less subjected to hydrolysis. Because of the above, as well as the observations made from the docking simulations and the reported synthetic accessibility of 2-lactosamine¹⁰⁶, candidate inhibitor **1** was taken forward.

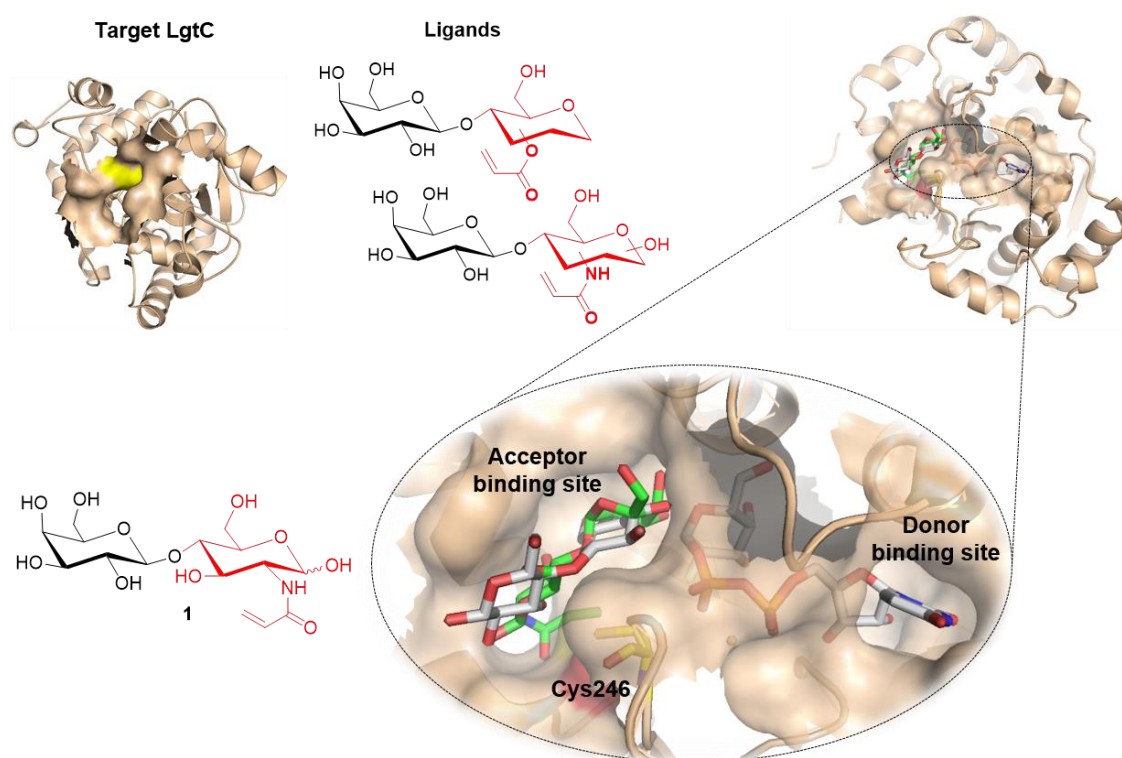
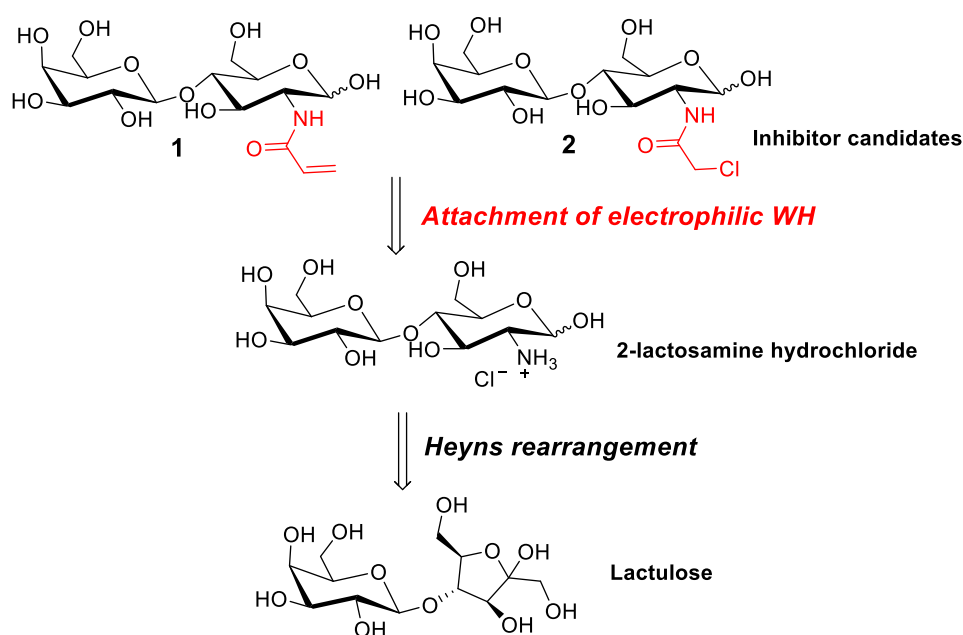


Figure 2.6. Structure-based design of disaccharide covalent inhibitors. Docking of a small library of lactose and lactosamine derivatives against LgtC (PDB 1GA8), with a Michael acceptor motif attached at different glucosidic positions of lactose. *Bottom left:* Structure of **1**, best target identified through docking. *Bottom right:* Docking simulation solution. Donor and acceptor substrates (grey) are shown in their respective binding pocket of LgtC (wheat). **1** (green) binds to acceptor binding site and reactive carbon of Michael Acceptor group is located 1.8 Å from cys246 (yellow)

2.4. Chemical synthesis of disaccharide chemical targets

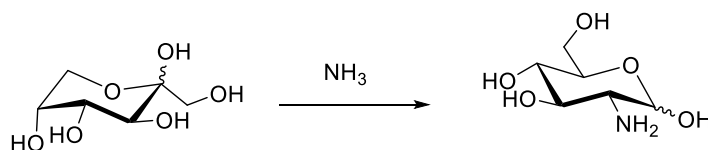
The structure-based design of disaccharide candidates described above identified 2-lactosamine derivatives as the best candidate to take forward. Two disaccharide derivatives have therefore been synthesised: **1** containing an acrylamide WH directly identified *via* docking and **2**, containing a chloroacetamide WH (Scheme 2.2). In the introduction to this chapter, chloroacetamide and acrylamide have both been described as highly suitable WHs for the design of TCIs⁹¹. **1** and **2** were synthesised from lactulose *via* the Heyns rearrangement, allowing the formation of key intermediate 2-lactosamine hydrochloride, from which the WHs can be attached¹⁰⁶. In this section, before discussing the chemical synthesis of **1** and **2** as well as the identification of a reaction by-product (not yet reported at the time). The mechanism by which the Heyns rearrangement occurs, will be briefly introduced.



Scheme 2.2. Retrosynthesis of disaccharide candidates **1** and **2**

2.4.1. Mechanism of the Heyns Rearrangement

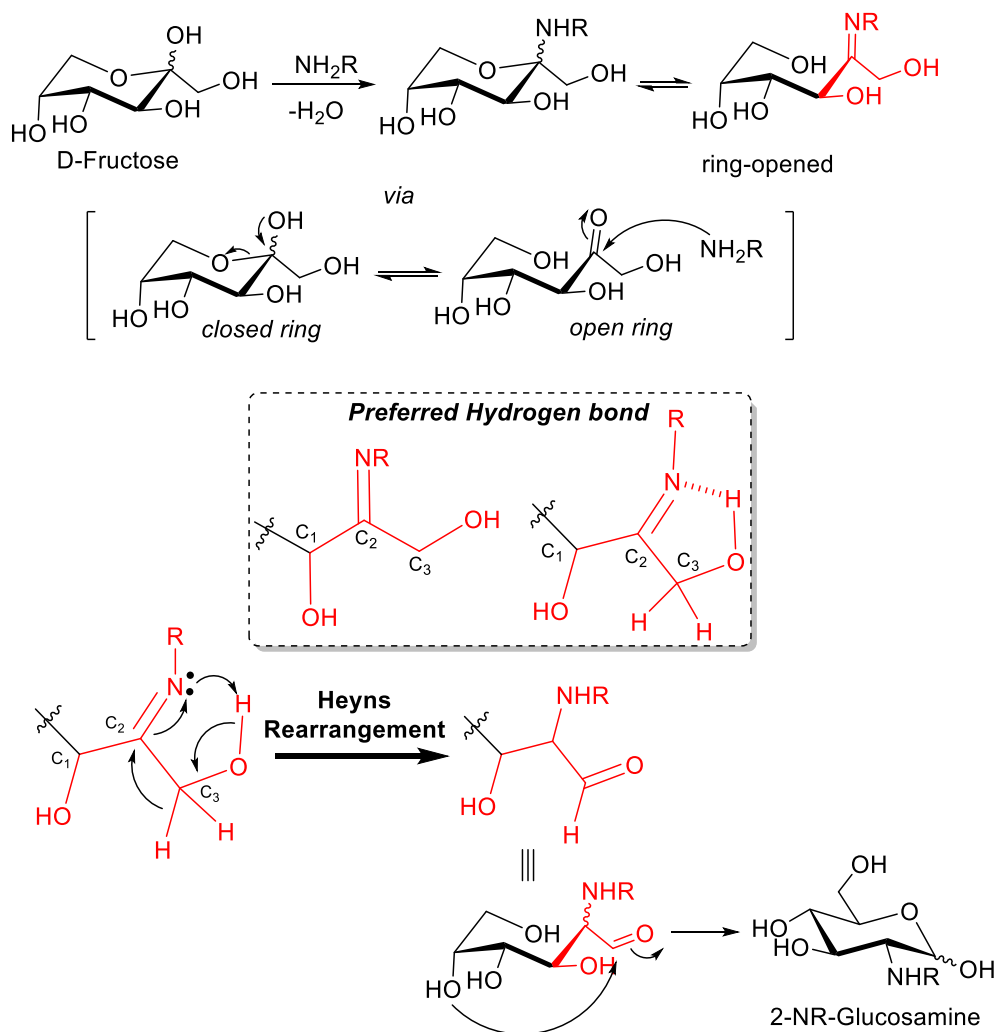
Because of their relevance in living organisms, the synthesis of 2-deoxy-2-amino sugars has been extensively studied. First discovered in the early 50's for the formation of 2-deoxy-2-aminoglucose by reaction of D-Fructose with ammonia^{124, 125, 126} (Scheme 2.3), the Heyns rearrangement has been described as the easiest and cheapest way to synthesise 2-deoxy-2-amino sugars¹⁰⁶.



Scheme 2.3. Heyns Rearrangement of D-Fructose

Initially, the reported reaction yields were low: around 10% with rare exceptions of up to 30% depending on the reaction conditions¹²⁵. In the late 90's Wrodnigg and Stütz developed modified reaction conditions and managed to substantially increase the reaction yields, making it viable on a preparative scale¹²⁷.

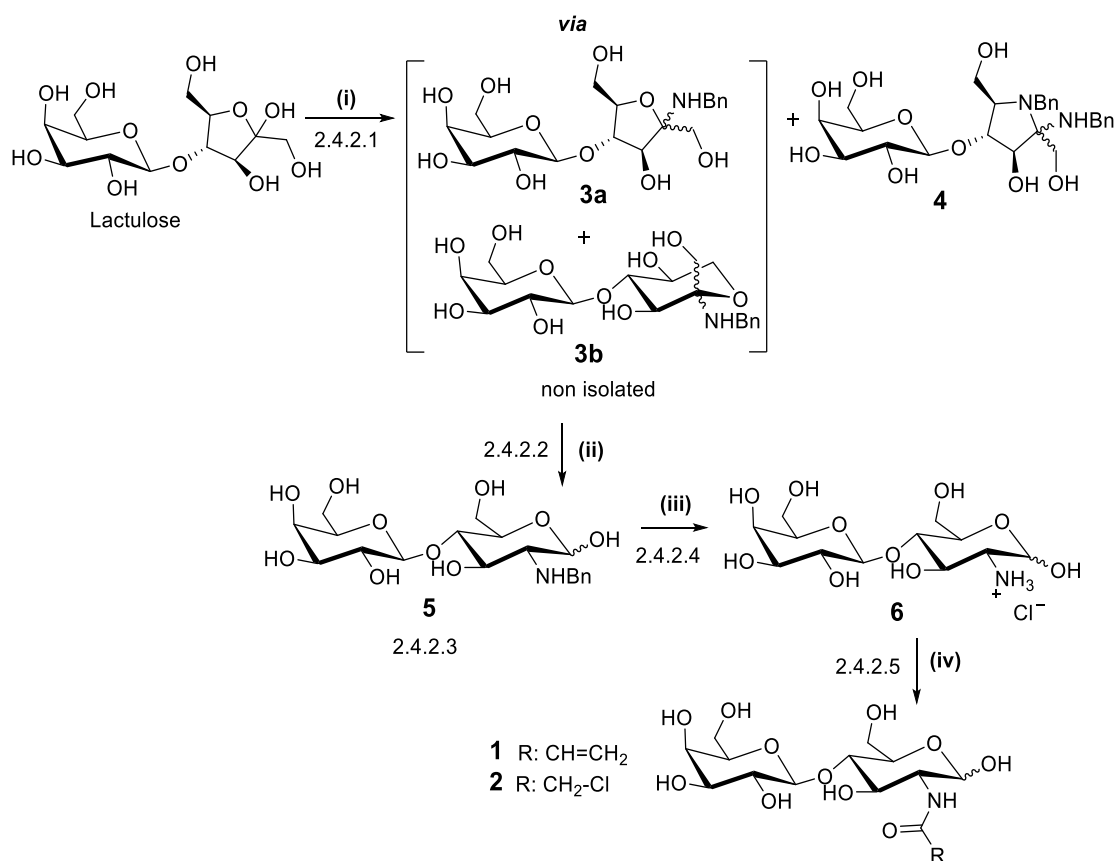
More recently, in 2015, Sanchez Viesca and Gomez proposed a novel reaction mechanism for the Heyns rearrangement and explained the observed regioselectivity for a C2-C3 reaction over a C1-C2 reaction¹²⁸ (Scheme 2.4). Due to its greater acidity, the primary alcohol at C3 is a better H-bond donor. The hydrogen bond with the primary alcohol at C3 is therefore preferred over a H-bond with the secondary alcohol at C1¹²⁸. The nitrogen atom acts as a Lewis base and promotes an internal oxido-reduction reaction: C3 is oxidized while the nitrogen atom is reduced. This 1,4 hydrogen transfer occurs via a cyclic, concerted five-membered reaction mechanism (internal catalysis)¹²⁸ (Scheme 2.4).



Scheme 2.4: Proposed mechanism for the Heyns Rearrangement

2.4.2. Discussion of disaccharide chemical targets synthesis: *cmpds 1 and 2*

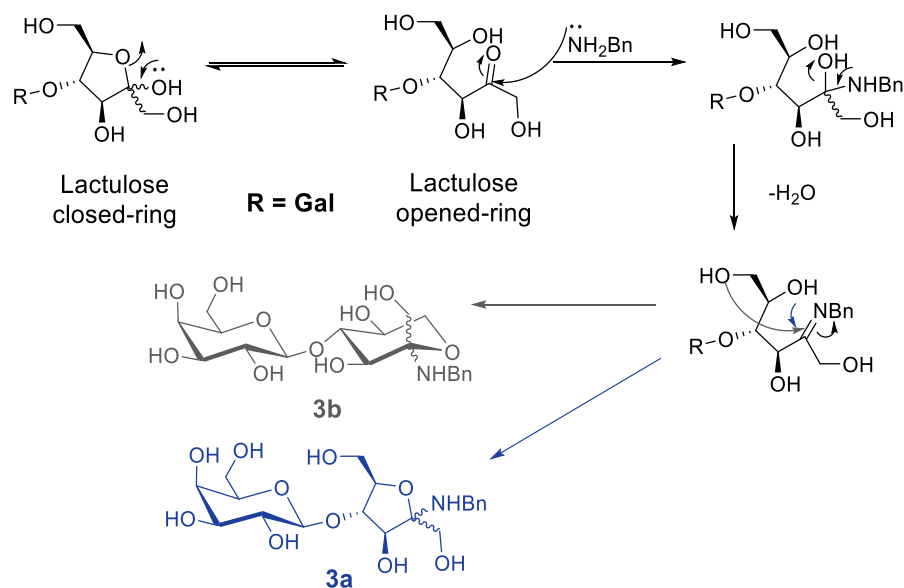
1 was prepared via the Heyns rearrangement from commercially available lactulose¹⁰⁶, using the route described in Scheme 2.5. The appealing factor of this route relies on the fact that compound **6** is a key intermediate from which different electrophilic warheads can be attached in order to generate a small library of electrophilic lactosamine-like compounds.



Scheme 2.5. Synthesis of target molecules **1** and **2**. *Reagents & Conditions:* (i) BnNH₂ (solvent), 45°C, 40 hours (ii) 10% glacial AcOH in MeOH, rt, 2 hours, 50%, (iii) H₂ (5 bars), Pd(OH)₂/C cat (0.3 eq), 2:1 water/dioxane, HCl pH=1-2, 24 hours, 90%, (iv) acid anhydride (3 eq), TEA (2eq) in DMF, 20-25%.

2.4.2.1. Step (i): Addition of benzylamine to lactulose

The first step involves the nucleophilic attack of benzylamine at the anomeric position of lactulose to yield intermediate **3**. The proposed mechanism for this step is shown in Scheme 2.6. Lactulose is in an equilibrium between two isomeric forms: the closed and the opened-ring states. In the opened-ring structure of lactulose, the carbonyl can be readily attacked by the excess of nucleophilic benzylamine present in solution. After release of water and ring closing, compound **3** is formed. The ring closure can be initiated by two distinct hydroxyl groups attacking the imine intermediate, therefore both isomers **3a** and **3b**, are generated.



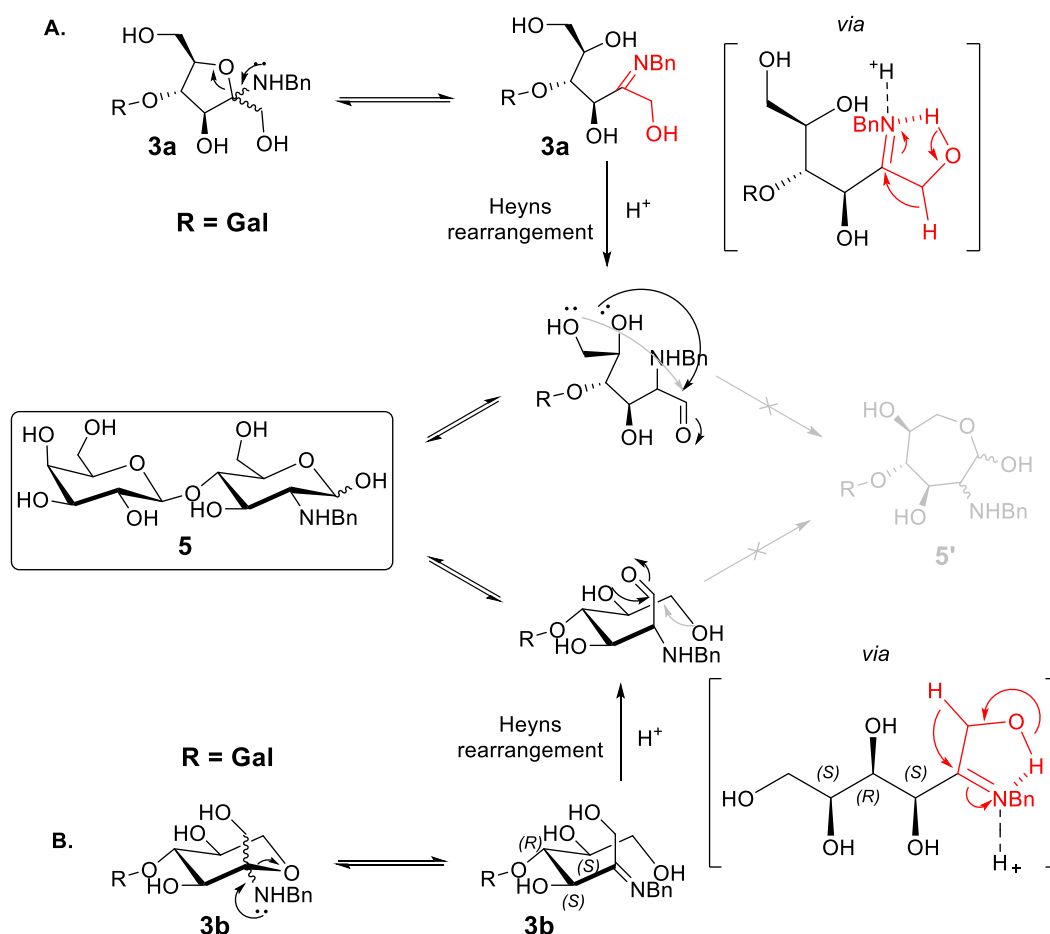
Scheme 2.6. Mechanism for step (i): formation of isomers **3a** and **3b**

Lactulose was heated in benzylamine for 40 hours at 45°C and the reaction was monitored by TLC (DCM: MeOH: NH₄OH, 1:1:0.1). After removal of the excess of benzylamine, TLC showed that the residue contained a mixture of two products (A, *r_f* = 0.4 and B, *r_f* = 0.8) (Figure 2.7). **3a** and **3b** are closely related structural isomers and are not expected to exhibit such a significant difference in retention on TLC. Therefore, this observation can only be explained by the presence of a secondary product in the reaction mixture which had to date, never been reported. This would also explain the relatively low yield obtained for this step (50%). Following the reaction procedure¹¹⁶, the residue obtained from step (i) was taken forward as is without further treatment. Therefore, the secondary product was isolated and characterised in the following step.

2.4.2.2. Step (ii): Heyns rearrangement

The second step is the Heyns rearrangement which generates **5** from **3**. The presence of acid in the reaction mixture promotes the internal catalysis initiating the formation of the Heyns product¹⁰⁶ (Scheme 2.7). The two isomers of compound **3** both lead to the formation of compound **5** under those conditions (Scheme 2.7). An alternative ring closing

process could lead to the formation of the isomer **5'**. However, being an unstable seven-membered ring, **5'** is unlikely to be formed.



Scheme 2.7. Mechanism for step (ii): formation of **5** from **3a** and **3b**

TLC (DCM: MeOH: NH_4OH , 1:1:0,1) show that product A formed in step (i) disappears to form a slightly less polar product: C ($r_f = 0.5$) (Figure 2.7). On the other hand, it is observed that product B formed in step (i) was not modified upon rearrangement conditions. This suggests that product A formed in step (i) was the expected product of the reaction: compound **3** (showing as a smear due to the presence of two isomers) while product B was most likely the secondary unreported product.

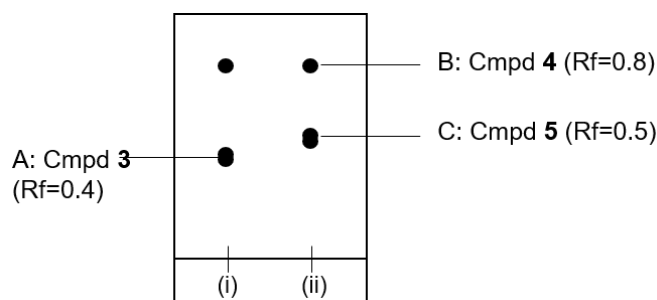
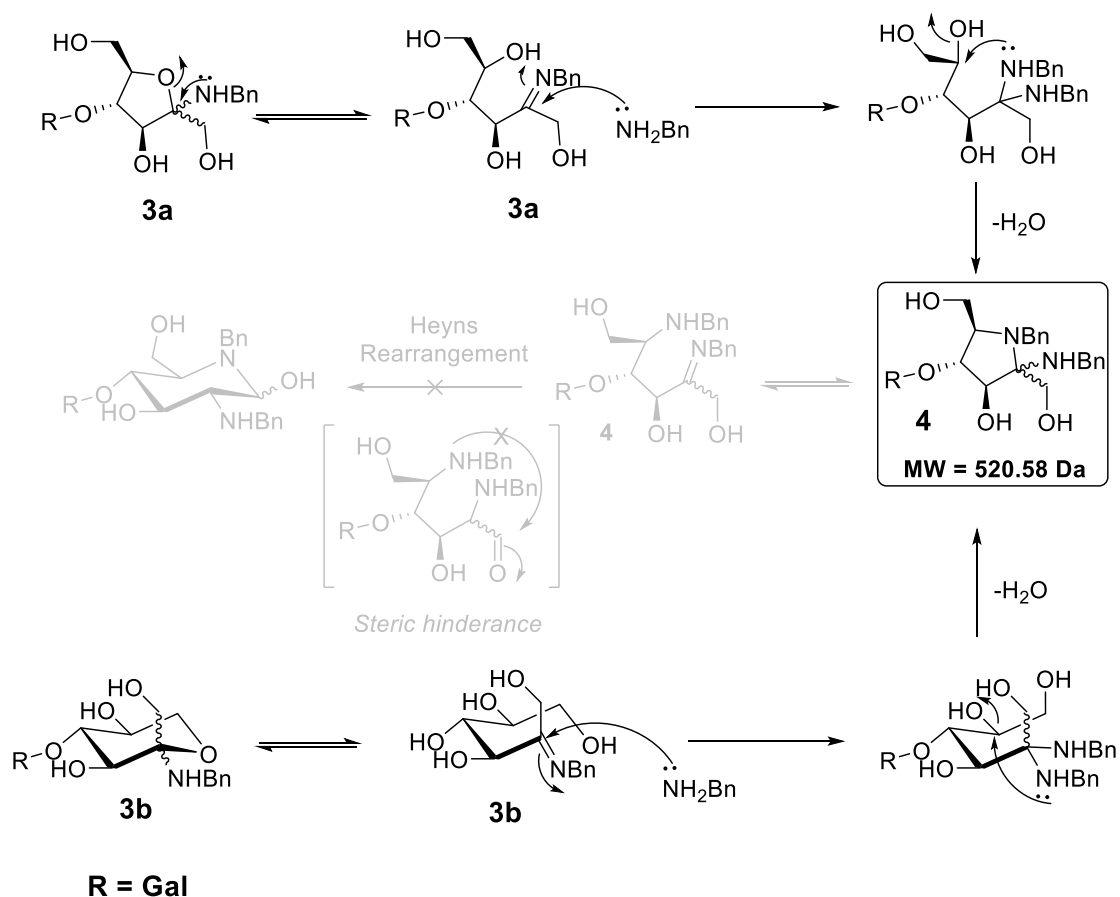


Figure 2.7. TLC (DCM: MeOH: NH₄OH, 1:1:0,1) after step (i) and step (ii) of disaccharide candidates synthesis. Cmpd **3** is converted to compound **5** under acidic conditions. By-product is unmodified.

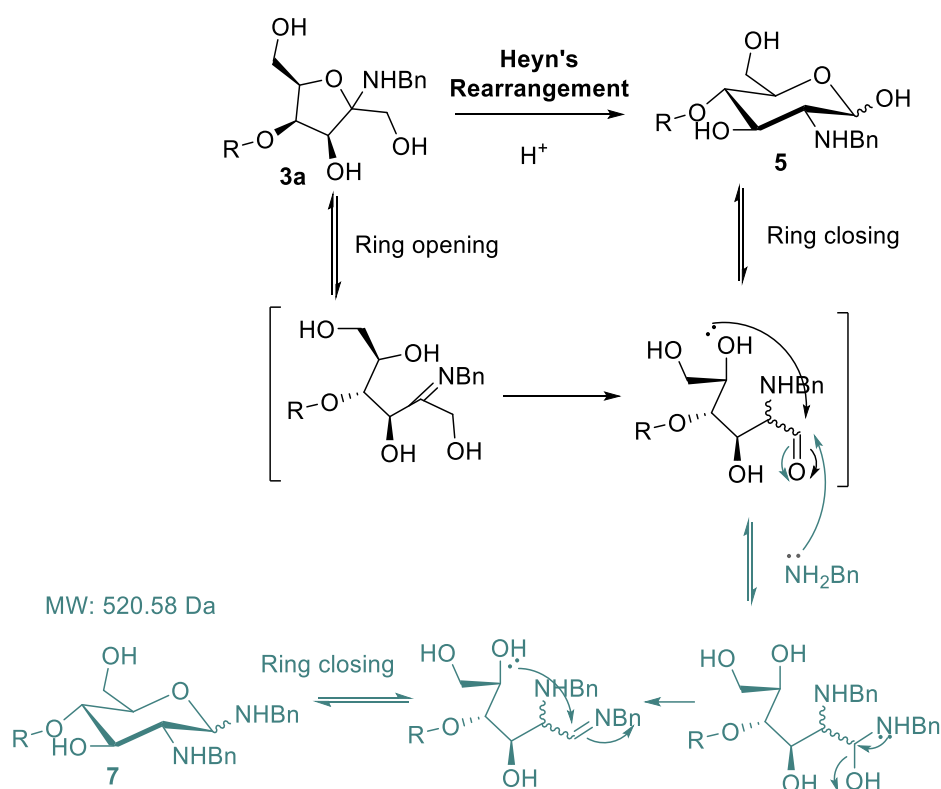
Successful purification by normal phase chromatography allowed for a good separation of both B and C. MS analysis of product C was consistent with the expected mass of compound **5** ($m/z = 432$, corresponding to the $[M+H]^+$ ion). Additionally, MS analysis revealed a mass of $m/z = 521$ for product B which gave insight into its structure. This mass corresponds to the addition of a second equivalent of benzylamine on **3** and the release of a water molecule ($431 + 107 - 18 = 520$). This is consistent with the ¹H NMR spectrum of the secondary product, clearly showing an additional aromatic region compared to compound **5** (Figure 2.8). Benzylamine is present on large excess in the first step of the reaction. Therefore, a second equivalent of benzylamine can attack the imine in the opened-ring form of **3**, which, followed by the loss of a water molecule, lead to the formation of **4** (Scheme 2.8).

The fact that by product **4** does not seem to be rearranged under acidic condition can be explained by steric effect (Scheme 2.8). The two benzylamine groups provide too much hindrance for **4** to undergo rearrangement. Alternatively, the rearrangement may have occurred, and unlike compounds **3** and **5**, **4** and its rearranged isomer may not be separable by normal phase under these conditions.



Scheme 2.8. Proposed mechanism for the formation of bis benzylated by product **4** from **3a** and **3b** in step (i) of the synthetic route.

During the course of these studies, the identification of by-product **4** had not yet been reported. These observations were therefore novel discoveries regarding the Heyns Rearrangement. Hederos *et al.* later reported similar observations, however the structure and mechanism by which the formation of their by-product, **cmpd 7**, is formed differs from our own observations¹²⁹. (Scheme 2.9).



Scheme 2.9. Reported mechanism for the formation of bis-benzylated by product **7** under acidic conditions

M. Hederos *et al.* describe the formation of compound **7** to occur from the rearranged compound, exclusively under acidic conditions¹²⁹. However, our observations were clear: the formation of the secondary product occurs during the first step of the synthesis during which the conditions are not acidic (BnNH_2 , 45°C). Additionally, all excess benzylamine was washed off by the time compound **3** is put through step (ii) conditions (10% glacial AcOH in MeOH). This confirms that the bis-benzylated by product can only be formed during step (i) and that its formation does not occur under acidic conditions. Additionally, as reported in the literature, my own observations show that the Heyns rearrangement only happens under acidic condition as no trace of compound **5** (TLC spot C) was observed when monitoring step (i) by TLC. Additionally, the ^1H NMR spectrum of the secondary product is consistent with the proposed structure for the secondary product as no anomeric proton signals are present in the molecule, indicating that the anomeric carbon is quaternary (Figure 2.8). My own observations all converge towards compound **4** being

the correct structure for the secondary product obtained rather than the reported compound **7**. A summary of these differences is laid out in Scheme 2.10.

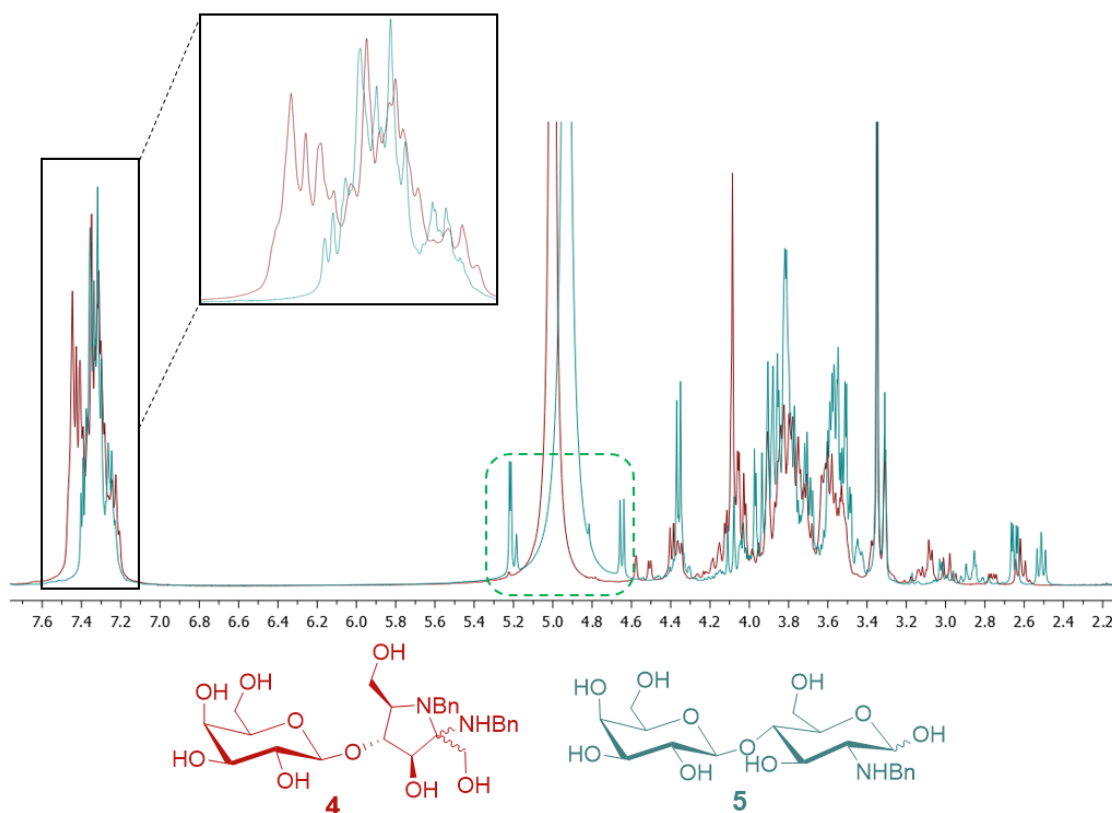
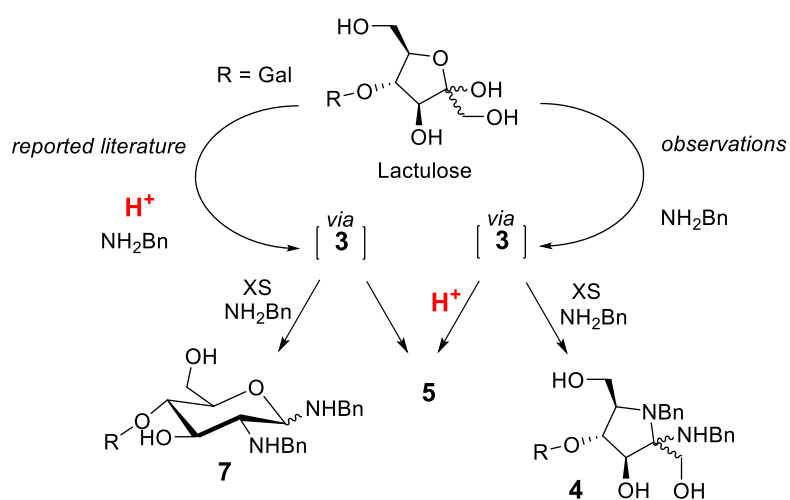


Figure 2.8. ^1H NMR of **4** (red) and **5** (green). An additional aromatic region is observed for **4** which is consistent with an extra benzyl group present on the molecule. No anomeric peak is observed for **4** which is consistent with the proposed structure in which the anomeric carbon is quaternary.



Scheme 2.10. Summary of reported literature and observations made regarding the nature of the bis-benzylated by product.

These observations, although interesting from a synthetic point of view are not crucial regarding the end goal of the project. Therefore, I analysed them to the best of my capacity and decided to move forward towards synthesising my target molecules from **5**.

2.4.2.3. Characterisation of cmpd **5**: 2-N-benzyl lactosamine

The free hydroxyl at the anomeric position of compound **5** confers a flexibility to the molecule as it can exist in both a closed and opened ring state. For this reason, the molecule epimerises making the stereochemistry at the anomeric position not fixed. As expected, compound **5** is obtained as a mixture of anomers, with an α : β ratio of 1.4:1 (Figure 2.9 A).

Although the complexity of the ^1H NMR spectra does not allow for the identification of each individual proton signal, the characteristic anomeric peaks are well defined (Figure 2.9 A). Due to a strong orbital overlap between H_1 and H_2 in the β -anomer, a large J_{1-2} coupling constant is observed ($J_{1-2} \beta = 8.5\text{Hz}$) compared to a smaller J_{1-2} in the case of the α -anomer ($J_{1-2} \alpha = 3.5\text{Hz}$) (Figure 2.9 A). Additionally, 2D correlation spectroscopy (COSY) provides a clear identification of H_2 proton for each anomer which allows for further confirmation of the α : β ratio (Figure 2.9 A and B).

The obtained α : β ratio is unsurprising. It can be explained by the anomeric effect, a stereoelectronic effect by which an heteroatom at the anomeric position of a sugar tends to adopt an axial orientation rather than the usual less hindered equatorial orientation due to the stabilising interaction between the lone pair of the endocyclic oxygen and the σ^* orbital of the exocyclic C-O bond (Figure 2.10). Additionally, the α anomer is stabilised by an intramolecular hydrogen bond between the amide group and the hydroxyl at the anomeric position, forming a favourable 5-membered ring. This stabilising H-bond cannot exist in the β anomer (Figure 2.10).

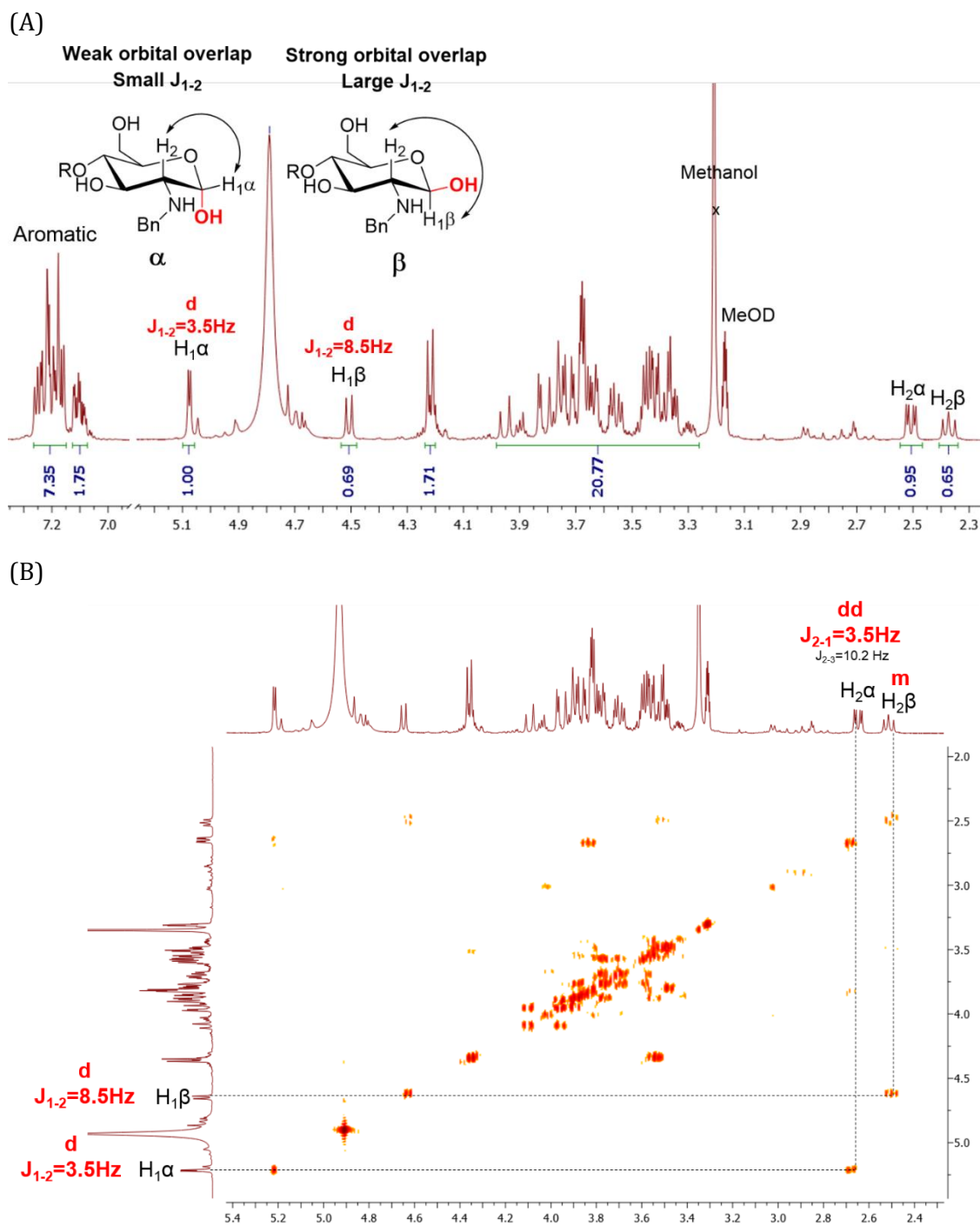


Figure 2.9. (A) ^1H NMR of **5** in D_2O . **5** is a mixture of α and β anomers in the ratio 1.4:1. (B) 2D correlation spectroscopy (COSY) of **5** which enables the identification of H_2 proton.

On the other hand, the analysis of the aromatic region in the ^1H NMR spectra of **5** provides confidence that all excess benzylamine has been removed (9.1 aromatic protons obtained, $5 + 5 \times 0.7 = 8.5$ aromatic protons expected) (Figure 2.9).

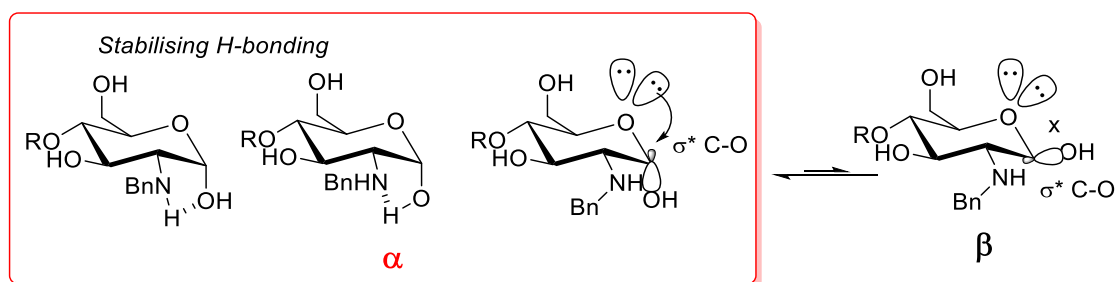


Figure 2.10. Structural and electronic explanations regarding the highest abundance of α -anomer in comparison to β -anomer.

2.4.2.4. Step (iii): debenzoylation by hydrogenation

The third step is the formation of the hydrochloride salt **6** by removal of the benzyl protecting group of **5**. This N-debenzoylation step is a heterogeneous catalytic hydrogenation using hydrogen gas as a reducing agent and a palladium catalyst. Pearlman's catalyst ($\text{Pd}(\text{OH})_2/\text{C}$) is used as it has been reported to be very efficient for the hydrogenolysis of benzyl-nitrogen bonds¹³⁰. It is important that the reaction proceeds under acidic conditions as it has been shown to drastically improve the reaction yields by increasing the polarity of the N-Bn bond and making it more electrophilic for surface hydride attack in hydrogenolysis¹³¹. The reaction was relatively slow (24 hrs) and required a high pressure of H_2 (5 bars) but gave quantitative yields. It is important to note the safety risks associated with this type of reactions: the combined use of hydrogen gas, pyrophoric catalyst and highly flammable solvent (dioxane) in a pressurised environment (6bars) may lead to fire and explosions, if caution is not applied¹³². The reaction was carried out in a contained fume hood, on a Paar apparatus (mechanical shaking) using a vessel suitable for high pressure. Oxygen was removed from the vessel prior to the addition of hydrogen gas, and hydrogen was evacuated from the vessel before oxygen was allowed back in. The catalyst was later filtered off on celite and placed in a wet metal waste container.

2.4.2.5. Step (iv): Chemical attachment of the electrophilic warhead

The last step is the attachment of the electrophilic warhead on the **6** *via* an amide formation. In this reaction the greater nucleophilicity of the amine over alcohol groups is exploited for the selective addition of the electrophilic warhead. The free amine is generated using triethylamine and both the acyl chloride and the acid anhydride were tried as a reaction partner. For the formation of both **1** and **2**, the acid anhydride gave better yields and purer compounds than the corresponding acyl chloride. However, the yields remained very low (25% highest yield obtained for this step), which is most likely explained by the use of methanol as the reaction solvent which may quench the anhydride *in situ*. The high polarity of the candidate did not allow for any other solvent to be used due to solubility issues. The reaction was attempted in DMF but due to its high boiling point, significant amount of solvent remained in the product. Additionally, **1** and **2** were purified by cation exchange column, and a large amount of resin was needed to remove all excess triethylammonium salt, which inevitably leads to product degradation.

An alternative synthetic strategy was attempted to overcome this polarity issue encountered: carrying out the synthesis on protected sugar (to allow for water extraction steps and normal phase chromatography purification) and deprotecting the molecules after attachment of the electrophilic warhead. **5** was successfully fully protected with Ac groups however the following N-debenzylation reaction of the protected sugar was unsuccessful even at 6 bars of H₂ pressure for 48hrs. Therefore **1** and **2** obtained from the unprotected sugar synthetic route, were biochemically tested for their inhibitory property against LgtC.

2.5. Biochemical testing

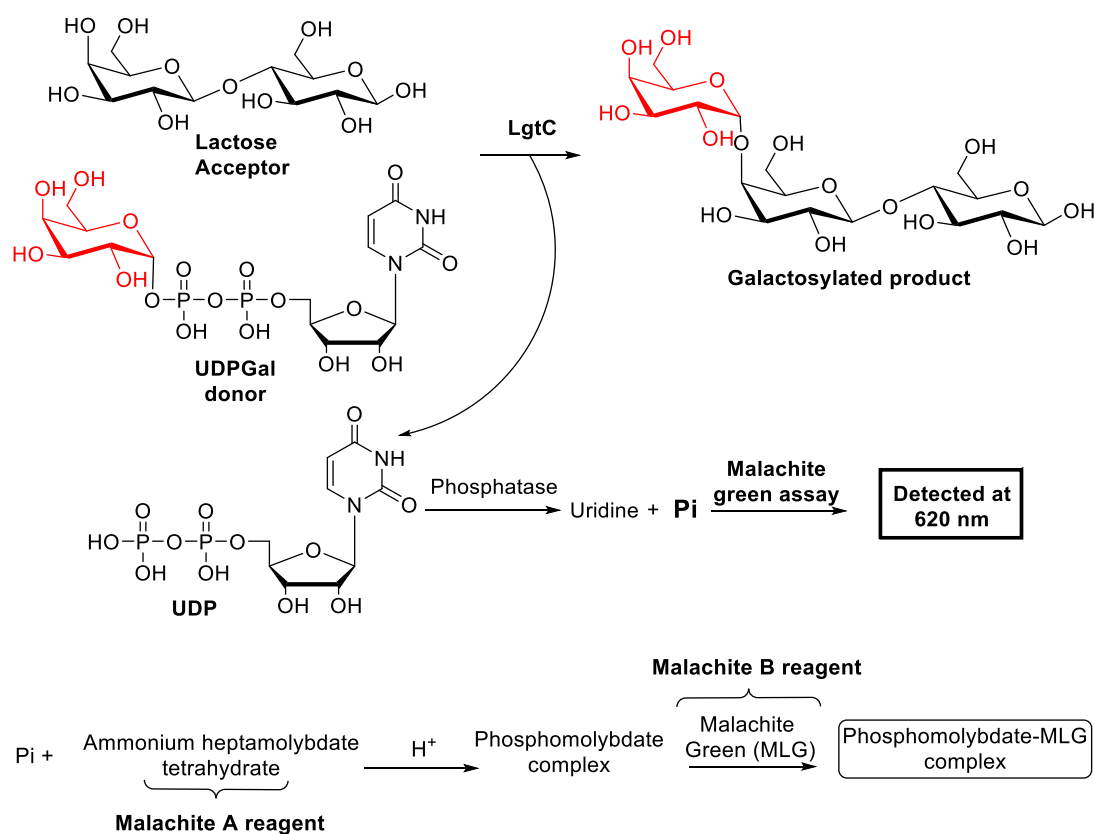
Before discussing the biochemical testing of cmpds **1** and **2**, the assay used for the evaluation of an inhibitor / substrate activity will be briefly introduced.

2.5.1. Phosphatase-coupled biochemical assay

Wu *et al.* recently reported a novel assay which quantifies the GT activity by detecting the release of inorganic phosphate, a secondary product of the enzymatic reaction¹³³. The principle relies on the addition of a UDP-selective phosphatase into the GT reaction, the free phosphate is then detected using the well-established colorimetric malachite green technique¹³⁴. The direct proportionality between the amount of phosphate released and the sugar nucleotide consumed makes this method very useful for the calculation of accurate kinetic parameters. Additionally, the assay can be performed on a 96-well plate and quantified by a plate reader making it High Throughput Screening (HTS) compatible. Additionally, as most GTs are Leloir enzymes and use sugar-nucleotides as their donor substrate, this assay is applicable to a wide variety of enzymes.

Our group has been working on the adaptation and optimisation of the assay developed by Wu *et al.*¹³³ for the identification and characterisation of small molecule glycosyltransferase inhibitors¹⁰⁷. The assay reproducibility has been significantly improved by addition of chicken egg-white lysozyme as a carrier protein to circumvent the protein losses through protein adhesion to plastic surfaces¹⁰⁷. Also, the running cost of the assay was substantially reduced by using the inexpensive calf intestinal phosphate (CIP)¹⁰⁷. This improved assay is now the standard technique used in the group for the biological testing of small molecules as potential GT inhibitors. The principle is shown in Scheme 2.11: LgtC transfers an α -galactose moiety from UDP-Gal to an acceptor, which results in the formation of UDP. The calf intestinal phosphatase (CIP) is used for the selective hydrolysis of UDP from which uridine and inorganic phosphate are generated. After a certain incubation time, Malachite A reagent, made of sulfuric acid and ammonium heptamolybdate tetrahydrate, is used to form a phosphomolybdate complex with the inorganic phosphate present in solution. Then Malachite B reagent (PVA and malachite green) is added and forms a phosphomolybdate-MLG complex which can be detected at 620 nm, a wavelength not typically absorbed by most small molecules. All inhibition

assays were carried out in the presence of surfactant to suppress non-specific aggregation and avoid assay artefacts¹⁰⁷. The assays were performed and analysed as detailed in the experimental section.



Scheme 2.11. Principle of the phosphatase-coupled GT assay and principle of malachite green phosphate detection

2.5.2. Evaluation of inhibition activity

Using the biochemical assay described¹⁰⁷ above, **1** and **2** were tested for their inhibitory activity against LgtC. Increasing concentrations of **1** and **2** were incubated for 20 mins at 30°C with DDT activated LgtC, lactose acceptor (2mM) UDP-Gal donor (28μM) and all other assay components¹⁰⁷. After addition of malachite reagent, the absorbance of each well was measured over 30mins. **1** and **2** were both identified as LgtC-inhibitors with an IC₅₀ of 1.2mM and 330μM respectively (enzyme turnover=25%) (Figure 2.11. A). The assay was also performed against CIP alone as a control experiment and no activity was

observed against the phosphatase (Figure 2.11. B). This indicates that the activity observed is genuine for LgtC.

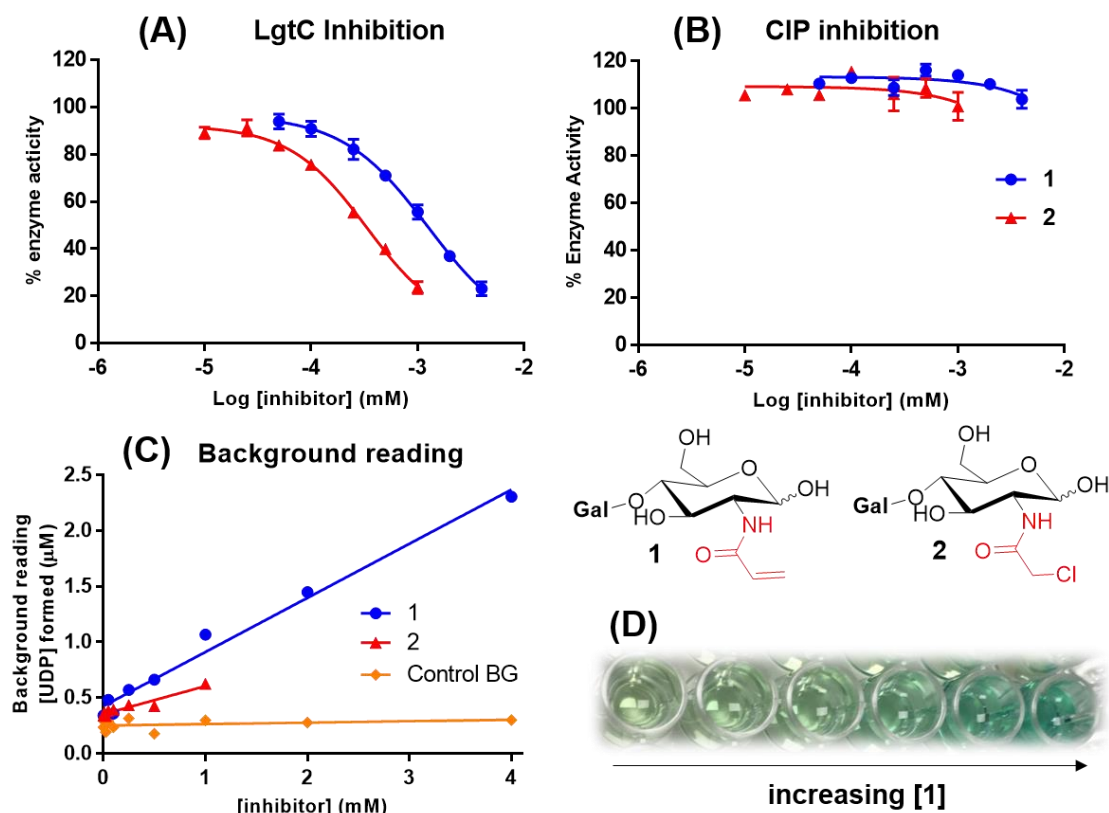


Figure 2.11. Biochemical testing of compounds **1** and **2**. (A) Inhibition assay against LgtC. *General conditions:* LgtC is activated with 5 mM DTT for 30 mins at 30°C and was incubated with lactose (2 mM), UDP-Gal (28 µM), MnCl₂ (5 mM), CIP (10 U/mL), CEL (1 mg/mL), Triton (0.01%) and inhibitor (0-1 or 4 mM) in 13 mM HEPES buffer (pH 7.0) for 20 min at 30 °C. The progress of the reaction was determined with malachite green as previously described. Each experiment was carried out in triplicate (technical). Error bars represent standard deviation. Data were plotted using GraphPad prism and fitted to “log(inhibitor) vs response variable slope (four parameters)” for the generation of the IC₅₀ values. *Results:* Both **1** and **2** show inhibitory activity against LgtC with an IC₅₀ of 1.2mM and 330µM respectively. (B) CIP control inhibition. The activity observed is due to a genuine inhibition of LgtC rather than an inhibition of the phosphatase used in the assay. (C) Unusual increasing in background reading upon increasing concentration of **1** and **2**. (D) Picture of background wells, showing an increase in colour intensity indicating an increase in release of UDP.

In the assay, each data point is corrected for a background reading which accounts for non-specific hydrolysis (and traces of UDP in UDP-Gal sample). Background wells contain all components of their corresponding assay well, except for lactose, to prevent the enzymatic reaction to occur. The background reading usually stays roughly constant (0.3 - 0.6µM) throughout the assay. In this case however, an unusual increase in the background reading was observed (Figure 2.11. C). This can be explained by mainly two

factors: the formation of precipitates in the well which affects the absorbance reading, or an unexpected amount of UDP being generated *in situ*. In this case, no precipitate was observed however, a clear colour change could be observed by the naked eye (Figure 2.11. D) indicating an increase in absorbance at 620nm. The only changing variable in all background well is the presence and increasing concentration of **1** and **2**. For lack of absorbance at 620nm for both **1** and **2** and as all other well components remain unchanged throughout, the above observation can only be explained by an increase in UDP generated *in situ* as a result of the increasing concentration of **1** and **2**. This in turn implies that the LgtC reaction has occurred in background wells which suggests a residual substrate activity for both disaccharide candidates as a consequence of their acceptor-based structure. Therefore, the LgtC-substrate activity of **1** and **2** needs to be assessed.

2.5.3. Evaluation of substrate activity

To better understand their behaviour towards LgtC, we tested compounds **1**, **2** and **6** in a substrate assay (Figure 2.12. A). DTT activated LgtC was incubated for 20 mins at 30°C with UDP-Gal donor (28µM) and increasing concentration of **1** and **2**, as the partner substrate of the reaction. We found that all 3 compounds show a residual substrate activity for LgtC. This is consistent with the observations previously made regarding the unexpected increase in background reading for the inhibition assay of **1** and **2**. Interestingly, **6**, the closest lactose analogue, was the most efficient altered substrate exhibiting a 14% UDP-Gal turnover at 2mM under the assay conditions (against 19% UDP-Gal turnover for natural lactose acceptor under the same conditions). At the same concentration, **1** and **2** were able to turnover 9- and 5% of donor respectively in the same assay conditions which suggests that LgtC also tolerates the presence of the warhead on its substrate scaffold. These results indicate the great versatility of LgtC to recognise not only its natural substrate but also a variety of structural analogues. Both the substrate promiscuity of LgtC and its potential application as a synthetic tool was described by

Withers *et al.* ^{135,136}. The wide range of analogues tested contained naturally occurring functional groups. However, we show herein that LgtC can also tolerate small synthetic functional groups. Although problematic in the context of this project, this suggests the potential widespread application of LgtC for synthetic chemistry purposes. More interestingly, these observations strongly suggest that the disaccharide inhibitors bind in the acceptor binding pocket of LgtC where Cys246 is present. However, using this biochemical assay, the residual substrate activity observed for **1** and **2** is not dissociable from their inhibitory properties, e.g. the data presented in Figure 2.11.A are not a good representation of the inhibitory activity of **1** and **2**. They were therefore re-analysed by accounting for the background in the absence of disaccharide candidates to allow for a more reliable quantification of their inhibitory activity (Figure 2.12. B).

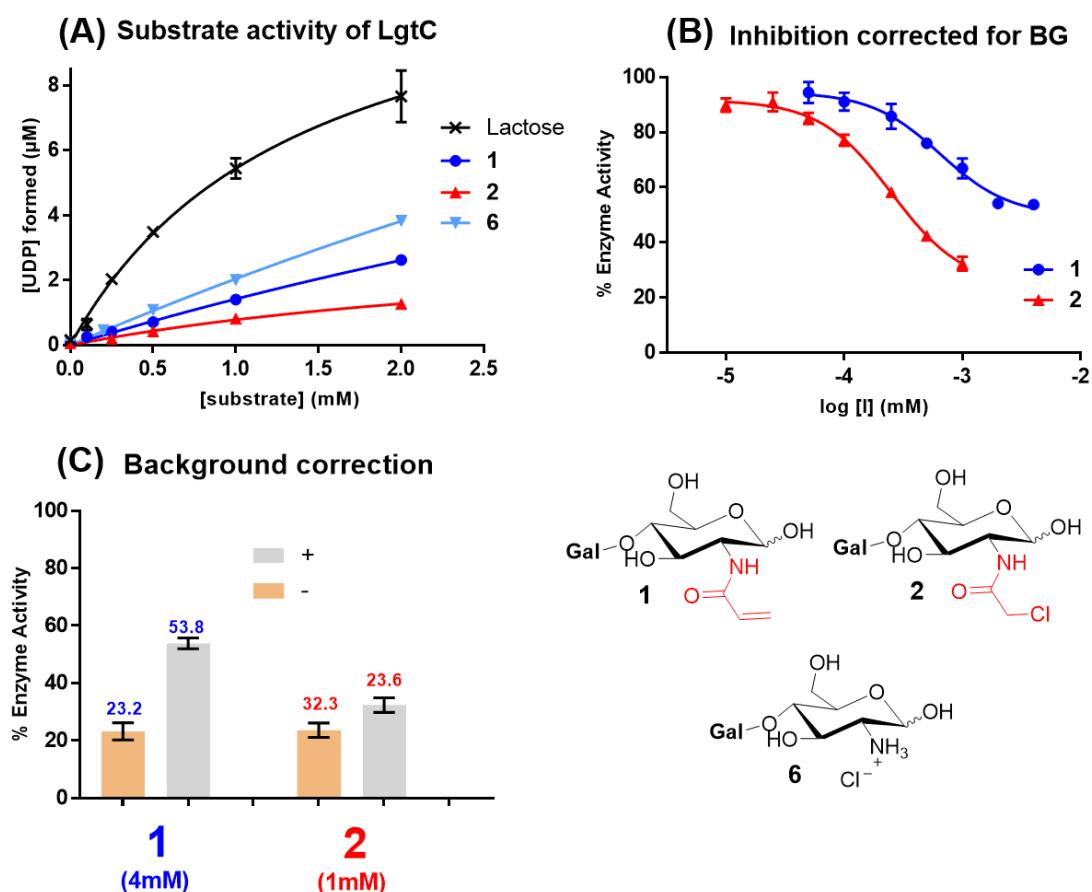


Figure 2.12. Biochemical testing of cmpds **1** and **2**. (A) Substrate assay of cmpd **1** and **2** for LgtC. *Substrate assay conditions:* LgtC is activated with 5 mM DTT for 30 mins at 30°C and was incubated with UDP-Gal (28 μM), MnCl₂ (5 mM), CIP (10 U/mL), CEL (1 mg/mL), Triton (0.01%) and substrate (0-2 mM) in 13 mM HEPES buffer (pH 7.0) for 20 min at 30 °C. The progress of the

reaction was determined with malachite green as previously described. Each experiment was carried out in triplicate (technical). Error bars represent standard deviation. Data were plotted using GraphPad prism and fitted to “log(inhibitor) vs response variable slope (four parameters)” for the generation of the IC_{50} values and to “Michaelis Menten cruve” for the generation of the K_m values. *Results:* Both **1** and **2** show residual substrate activity. (B) Inhibition of **1** and **2** corrected for background in the absence of disaccharide targets. (C) Difference in % of LgtC activity before and after background correction at 4mM of **1** and 1mM of **2**.

After re-analysis, the inhibition of both **1** and **2** was lower than initially observed. Unsurprisingly, correcting for the background in the absence of disaccharide analogue impacts more significantly on the activity of cmpd **1** as it was found to be a better LgtC-substrate than cmpd **2**. This was plotted in Figure 2.12. C. for the highest used concentration of **1** and **2** and the difference between non- and background corrected inhibition is indeed less significant for **2** than it is for **1**.

For direct comparison, **6** was also tested in an inhibition assay against LgtC (Figure 2.13.B). As expected, the increase in background reading is more significant for **6** than cmpd **1** or **2** as **6** is a better altered substrate (Figure 2.13.A). After standard data analysis, a sigmoidal curve is obtained suggesting an inhibition activity for **6**. However, when the data are corrected for the background in the absence of **6**, a shift in the curve towards greater enzyme activity is observed (from 1mM) indicating a stronger substrate than inhibitory activity for **6** at this concentration onwards (Figure 2.13.B).

Due to the dual nature of the disaccharide candidates, a robust experiment with our current biochemical assay cannot be performed as the concentration of disaccharide candidate present in solution at any time can never be known precisely. Additionally, the turnover of **1** and **2** leads to the presence of the trisaccharide enzymatic product in solution which may affect the outcome of the assay. Therefore, for all the above reasons, we believe that the dual nature of compounds **1** and **2** as substrates and inhibitors of LgtC does not allow for robust conclusions to be drawn as the quantification of one activity independently of the other is not achievable. Further analysis regarding the mode action of the disaccharide candidates were therefore not performed.

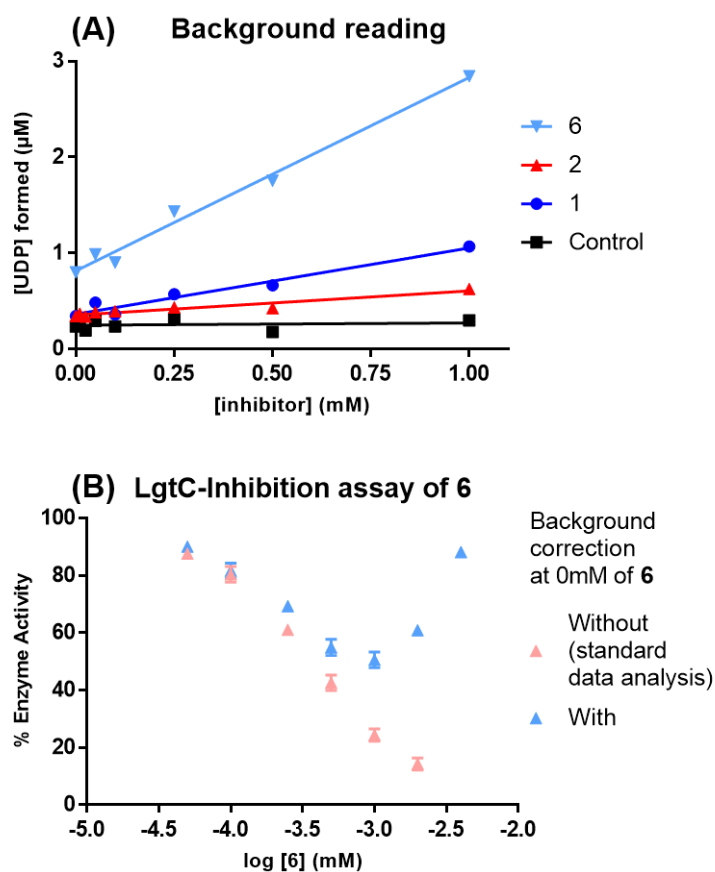


Figure 2.13. Biochemical testing of cmpds **6**. (A) Increase in background reading upon increasing concentration of **6**. Inhibition assay of **6** against LgtC (see general conditions in Fig 2.11) with and without background correction in the absence of **6**.

2.6. Summary and conclusions

In this chapter, the synthesis and initial biochemical testing of rationally designed covalent lactosamine-based inhibitors of LgtC was described. The design, derived from a standard TCI strategy, was based on the idea that if an electrophilic warhead was attached at a specific position of the substrate scaffold it may be possible to convert it to a covalent inhibitor for the targeting for cys246 in the acceptor binding pocket. We therefore investigated at which position of the lactose scaffold it was best to chemically attach the electrophilic warhead to achieve the covalent targeting of cys246 in the acceptor binding pocket. Compound **1**, with a Michael acceptor warhead attached *via* an amide linkage at position 2 of lactose, was expected by computational simulations to bind near the cys246 of interest in the acceptor binding site of LgtC. The chemical synthesis of cmpd **1** and its

chloroacetamide analogue (cmpd **2**) was achieved over 4 steps with a 9 and 11% yield respectively. Although partially reported in the literature, the synthesis and compound characterisation were challenging due to the polarity of such free sugar molecules and the fact that they exist as a mixture of diastereoisomers. A reaction by-product was identified and inconsistencies between the reported literature and my own observations have been described.

Both **1** and **2** are found to be inhibitors of LgtC but they also exhibit some non-negligible substrate activity for the enzyme. Complementing the reported literature¹³⁵, this indicates the great versatility of LgtC to tolerate non-naturally occurring functional group in the substrate scaffold which can be exploited for the chemical synthesis of complex oligosaccharides. In the case of the project, these molecules will not be used any further as, with our current assay detecting the release of inorganic phosphate, the quantification of their properties cannot be studied separately. It is to be noted that with a different biochemical assay, relying on substrate depletion or primary product detection for example, a better data analysis may be achievable.

Structural alignment and sequence analysis have shown that non-catalytic cysteines are a common motif in the active sites of many bacterial GTs⁷⁵. The commonality of non-catalytic residues among bacteria has been described before⁷⁸ and targeting these residues, as opposed to non-conserved ones (typical of TCI design), was identified as a potentially useful strategy for the design of potent and effective inhibitors for bacterial GTs⁷⁸. Although unsuccessful so far for LgtC, this strategy may therefore be applicable to other biological targets with higher substrate selectivity than LgtC. The structure-based method herein developed may serve as a stepping stone towards the systematic development of covalent and selective inhibitors for bacterial GTs.

Chapter 3

From disaccharides to monosaccharides

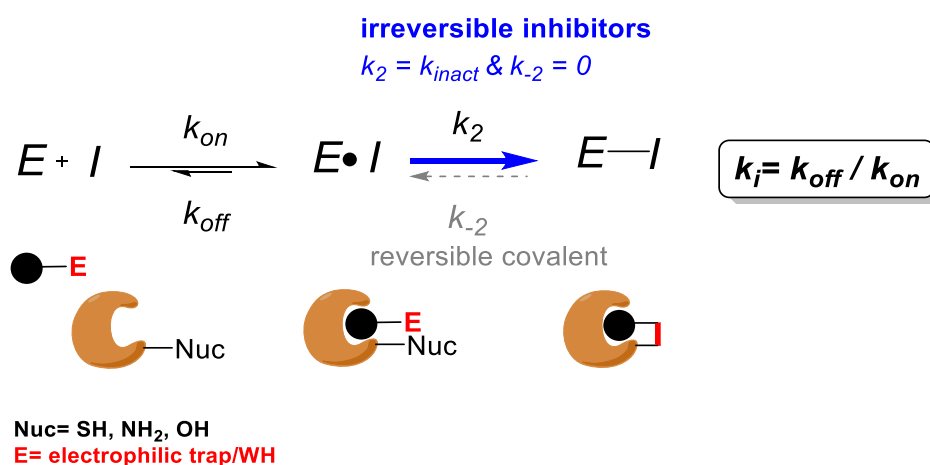
In the previous chapter, it was concluded that direct substrate analogues, although inhibitors of LgtC exhibited some residual substrate activity which unfortunately made them unfit for purpose. Their putative covalent mode of action (MoA) was therefore not analysed. In this chapter, the synthesis, biochemical testing and MoA analysis of a non-substrate, monosaccharide series of LgtC inhibitors will be described. Before discussing the results, a more detailed introduction on targeted covalent inhibitors (TCIs) will be provided.

3.1. Introduction

3.1.1. Kinetic model of covalent inhibition

TCIs exploit the two-step mode of action of covalent drugs, characterised by initial non-covalent binding and subsequent covalent bond formation (Scheme 3.1)⁷⁸. The first step involves non-covalent interactions between the inhibitor molecule and the protein and is governed by the constant k_i . These interactions stabilise a specific position of the complex from which the covalent bond can be formed. The subsequent formation of the covalent complex is characterised by the k_2 constant. For a fully irreversible inhibitor, the dissociation constant k_{-2} is essentially zero, while for a fully reversible inhibitor the covalent adduct E-I does not form therefore $k_2 = 0$. On the other hand, reversible covalent inhibitors have finite values for both k_2 and k_{-2} and encompass a range of behaviours between the two extremes for which the lifetime of the inhibited complex is typically governed by k_{-2} . A time-dependant inhibition, which shows a significantly increased inhibitory activity upon increasing enzyme-inhibitor pre-incubation is characteristic of a covalent mode of inhibition. The design of covalent drugs requires careful optimization of both the reversible binding affinity (k_i) and the reactivity of the electrophilic WH (k_2). The stronger the non-covalent interactions between inhibitor and enzyme the faster the

kinetics of the first step. The reactivity of the electrophilic warhead governs the kinetics of the second step.



Scheme 3.1. A two-step model for a covalent inhibition

3.1.2. Electrophilic warheads (WH) for TCI design

For the design of TCI drugs, a variety of reactive WHs that form irreversible links with a target are available. They differ in terms of their intrinsic reactivity and the type of amino acid they target. Among the most commonly used WH for the formation of irreversible bonds with biological targets are epoxides, Michael acceptors, haloacetamides and sulfonyl fluorides.

3.1.2.1. Epoxides

Due to their significant ring strain, epoxides are often regarded as too reactive for potential drug applications. However, some selective and relatively safe inhibitors bearing an epoxide group have been reported. A good example is fumagillin, an inhibitor of MetAP₂, an enzyme up-regulated in many cancers. Fumagillin targets His231 in the active site *via* ring opening¹³⁷ (Figure 3.1). Carfilzomib, a therapeutic agent for the treatment of multiple myeloma approved by the FDA in 2012, is another clinically relevant example¹³⁸ (Figure 3.1). Carfilzomib was rationally designed from the natural product Epoxomicin to target an N-terminal threonine residue of the proteasome resulting in the formation of an

irreversible ether linkage¹³⁸. Recently, a bacterial lectin covalent inhibitor, bearing an epoxide group has also been described¹¹⁶ (see 2.1.2, Figure 3.1).

3.1.2.2. Haloacetamides.

Alkylating agents, such as haloacetamide WHs, are another well-established class of irreversible modifiers typically targeting cysteine residues. Although generally associated with significant toxicities (chemical warfare in World War 1), some recent efforts have been made to improve the selectivity of alkylating agents. For example, haloacetamide WHs have found applications in the development of selective TCIs for protein kinases targeting cysteine residues^{139,140} (Figure 3.1). Haloacetamide containing molecules have also been used to target acquired cysteine mutations in human cancer as opposed to non-conserved cysteines in wild-type targets of interest. This has been demonstrated on an oncogenic G12C mutant of GTPase Ras, known to be exceptionally difficult to inhibit due to the high affinity of the protein for GTP/GDP (pM)¹⁴¹. A GDP analogue with an electrophilic chloroacetamide attached to the beta phosphate covalently modifies cys12 present in the K-Ras G12C mutant only (Figure 3.1)¹⁴¹. Another example reflecting the relevance of haloacetamide WHs in drug design is RRX-001, a bromoacetamide-containing small molecule identified as a prooxidative vascular anticancer agent which is currently in phase III of clinical trials¹⁴².

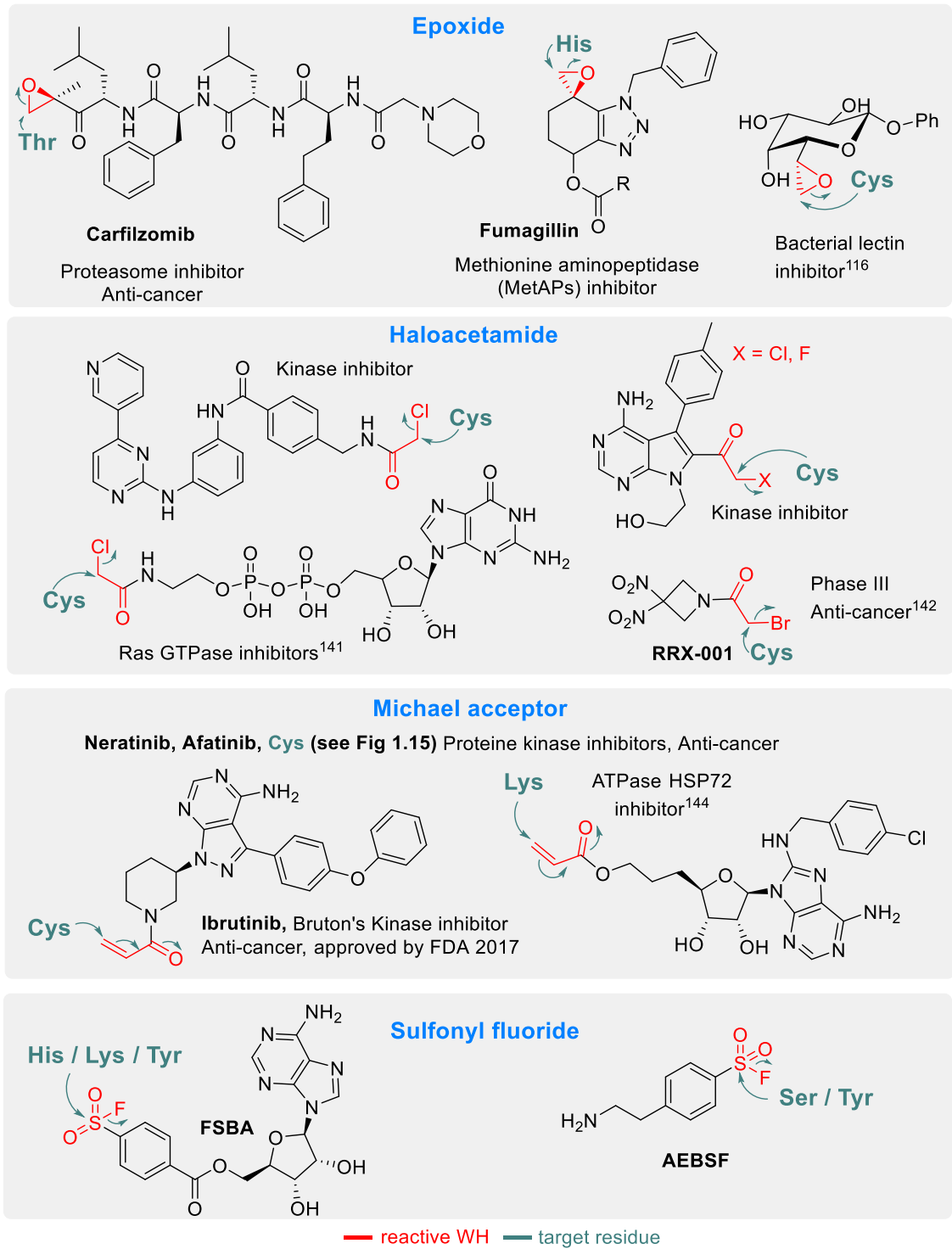


Figure 3.1. Examples of TCIs, their electrophilic WH and mode of action

3.1.2.3. Michael acceptors.

The third type of WH, which is amongst the most widely used nucleophilic trap for TCI design is the α - β unsaturated carbonyl. They are probably the most common WH and are Michael acceptors. The typical target of choice for these electrophiles is the thiol group from cysteine residues, the most nucleophilic among non-activated residues. Targeting catalytic or non-catalytic cysteines in an active site is becoming a popular strategy for TCI design. Examples of successful TCI research programmes are Afatinib, Imatinib and Ibrutinib, all FDA-approved anti-cancer drugs, which target cysteine residues in active site of protein kinases (Figure 3.1). In addition to this, many other Michael acceptor-containing TCIs are effective against kinases, ATP- and GTPases^{86,143}. Although Michael acceptors commonly react with cysteine residues, an unexpected instance of covalent modification of a lysine residue has been reported^{144,145}. While designed for the target of Cys17 in HSP72, a protein implicated in several cancers, the adenosine-based acrylate ester (Figure 3.1) reacted serendipitously with Lys56 instead¹⁴⁴.

3.1.2.4. Sulfonyl fluorides

In contrast to cysteine targeting WHs, sulfonyl fluorides (SFs) are known to target not only serine residues (their most common target) but also lysine, tyrosine and in very few cases histidine residues¹⁴⁶. Although not widely used for drug discovery purposes, SFs have found significant applications in the design of chemical probes in chemical biology and molecular pharmacology. Unlike sulfonyl chlorides, they are resistant to reduction and exhibit significantly improved thermodynamic stability¹⁴⁶. The most common pharmacophores used to design SF containing probes are (2-aminoethyl)benzenesulfonyl fluoride (AEBSF) and 5'-fuorosulfonylbenzoyl 5'-adenosine (FSBA) (Figure 3.1). FSBA is a commonly used covalent inhibitor and activity-based probe of ATP-binding proteins. The SF motif replaces the phosphoryl groups of ATP and peptide mapping has shown that the conserved lysine in the ATP-binding site of kinases is the targeted residue¹⁴⁷. In a recent

report, FSBA was used to label Hepatitis C Virus replicase by targeting residues Tyr382 and Lys491¹⁴⁸. Although the adduct of SF with a histidine residue is less stable than with Tyr and Lys, the FSBA analogue targets residue His130 in 5-phosphoribosyl- α -1-pyrophosphate synthetase of *Salmonella typhimurium*¹⁴⁹. Originally discovered by serendipity, the reactivity of SF has made them great tools for chemoproteomics. However, caution should be exercised when using fragment-like SF inhibitors due to their intrinsic promiscuity.

3.1.3. Reversible TCIs

Some other electrophilic WHs have the ability to form reversible covalent interactions with their biological targets *via* the formation of labile adducts¹⁵⁰. A good example is the reversible transformation of an aldehyde motif with a serine residue for the formation of a hemiacetal. However, as aldehydes are too reactive and show poor selectivity, they have been replaced by boronic acid moieties, which are milder electrophiles and exhibit greater selectivity. This is the case of Bortezomib¹⁵¹, an FDA-approved drug covalently inhibiting the 20S proteasome (Figure 3.2). Recently, the combined action of boronic acid and aldehyde WHs was implemented in the development of a lysine targeting inhibitor of MCL-1, a key factor in a wide range of human cancers responsible for protein-protein interactions¹⁵² (Figure 3.2). *Via* the formation of a stable iminoboronate, the boronic acid drastically slows down the rate of equilibrium of the imine formation reaction (reaction of lysine with aldehyde), prolonging on-target residence times for superior efficacy.

Nitrile groups, with their mild reactivity, require proper positioning in the active site in order to trigger covalent bond formation via a (reversible) Pinner reaction with a serine residue. A good example of this is Saxagliptin¹⁵³, a nitrile-containing covalent inhibitor of the serine protease dipeptidyl peptidase IV (Figure 3.2).

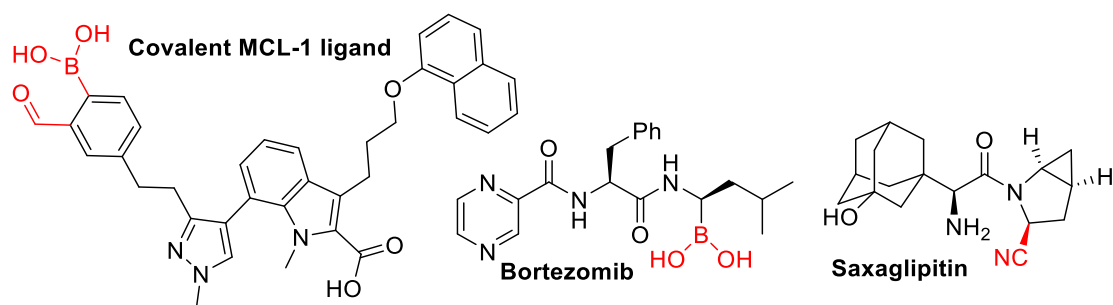


Figure 3.2. Examples of reversible covalent inhibitors

Taunton *et al.* showed that when combined with a Michael acceptor WH, nitrile groups increase the reactivity of the electrophile and tune it to react with cysteine residue in a reversible manner⁹⁰. The electron withdrawing effect of the nitrile moiety increases the acidity of the α -hydrogen promoting the cleavage of the covalent adduct (Figure 3.3). Additionally, by manipulating the size of the substituent near the WH, it was possible to modulate the residence time. A large group increases the steric bulk around the cyanoacrylamide WH, creating a “cap” and preventing the breakdown of the adduct by neighbouring residues or water molecules¹⁵⁴ (Figure 3.3).

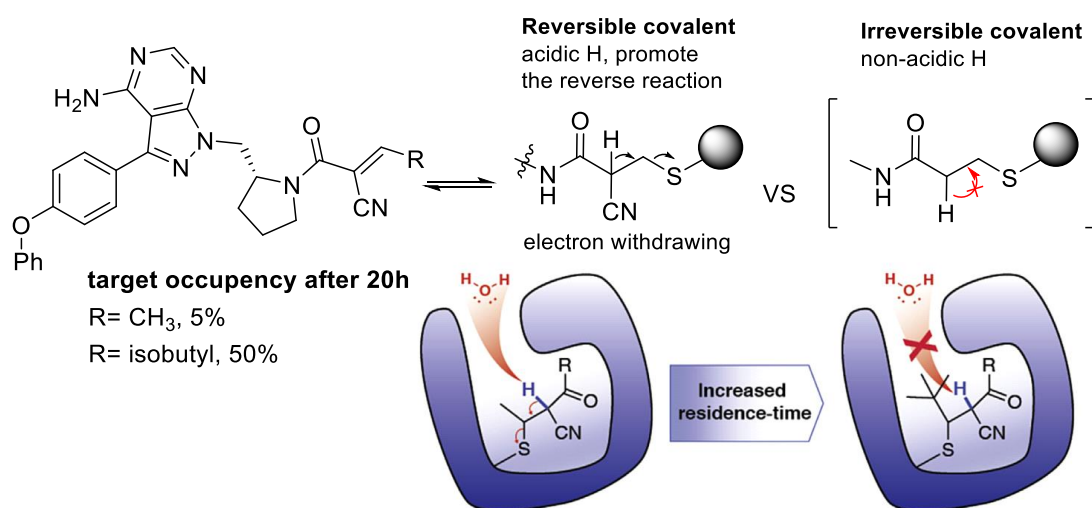


Figure 3.3. Nitrile-containing Michael acceptor for the WH reactivity tuning

3.1.4. Monosaccharide substrates for LgtC

In a relatively recent report, different carbohydrate scaffolds were investigated for their suitability as acceptor substrates for LgtC¹³⁵. For example, benzyl β -D-cellobioside, whose structure differs from the natural acceptor by an equatorial hydroxyl group at position 4 and an added benzyl group at the anomeric position, exhibits over 140-fold decrease in activity compare to lactose (Figure 3.4). As the anomeric position bearing the benzyl group points away from the binding pocket (see Figure 1.11), it is believed that the main reason for the reduction in activity is the inversion of stereochemistry at the galactose transfer centre. Monosaccharides were also tested in this study and all non-substituted sugars showed drastic reduction in activity (Figure 3.4). This was even more pronounced if the stereochemistry at position 4 was inverted compared to lactose (1200-fold decrease for D-glucose vs 800-fold for D-galactose). Interestingly, xylose showed the same level of activity as benzyl β -D-cellobioside and anomeric substitution resulted in largely increased substrate activity for monosaccharides¹³⁵.

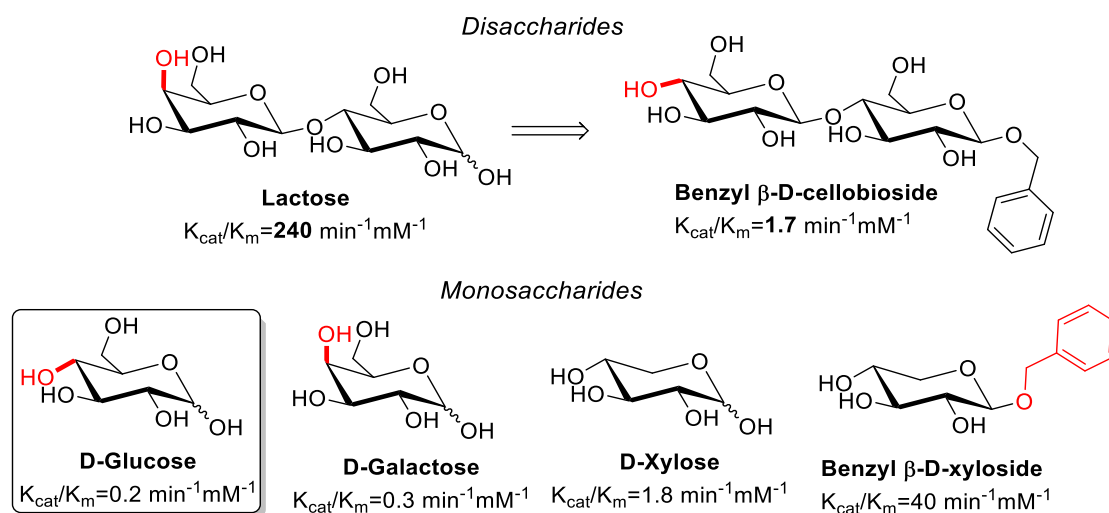


Figure 3.4. Kinetic parameters for various carbohydrate-based LgtC acceptor substrates (from Lairson *et al. Nat. Chem. Biol.*, 2006)

Following the disappointing results obtained in chapter 2, this study provides ground for the development of non-substrate carbohydrate-based inhibitors of LgtC.

3.2. Objectives

In the previous chapter, it was concluded that compounds **1** and **2**, direct substrate analogues of LgtC exhibit dual properties: they are inhibitors as well as substrates of LgtC. For these reasons, they are unattractive as molecular target for the purposes of the project. In this new chapter, the objective was to develop another class of carbohydrate-based inhibitors with no residual substrate activity and to study their inhibition activity and mode of action against LgtC. Unlike the lactosamine substrates described in the previous chapter, monosaccharides were identified as very poor acceptors of LgtC, with, in most cases, such slow kinetics that their substrate activity can be negligible¹³⁵ (Figure 3.4). We therefore wanted to exploit this for the design of carbohydrate-based LgtC-inhibitors with no residual substrate activity. The structure of LgtC shows the proximity of the Cys246 residue and the glucose motif of lactose within the acceptor binding pocket⁵⁴, which suggests that conserving a glucose moiety may be reasonable. Therefore, this, combined with the reported negligible substrate activity of glucose (Figure 3.4) and the synthetic accessibility of glucosamine derivatives, provided enough evidence that we should next focus on glucosamine-based candidates for the design of non-substrate inhibitors of LgtC.

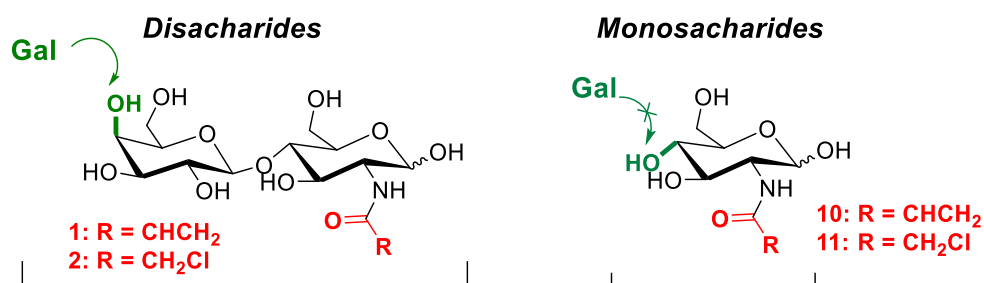
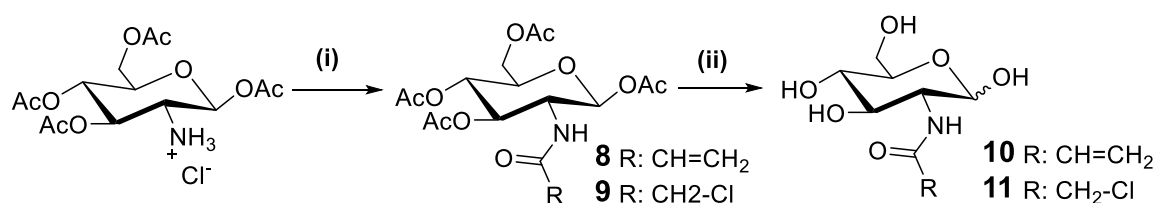


Figure 3.5. Structural differences between mono- and disaccharide LgtC inhibitor candidates

In line with the literature¹³⁵, we hypothesised that, unlike lactosamine derivatives, a glucosamine-based inhibitor may not be recognised by LgtC. This hypothesis relies on two main structural differences: the significant difference in the size of the molecules and the inversion of the stereochemistry at the Gal transfer hydroxyl (Figure 3.5). This chapter will discuss the chemical synthesis of compounds **10** and **11**, monosaccharide analogues of compounds **1** and **2** (chapter 2), then the analysis of their inhibition, mode of action and binding against LgtC will be studied in detail.

3.3. Chemical synthesis

10 and **11** are known compounds, already reported in the literature, therefore their synthesis will be described briefly. **10** is a known anti-arthritic¹⁵⁵ and has been used in two studies as a monomer for polymerisation strategies^{156,157} and **11** is a known fragment electrophile for proteomic studies⁹⁵. In this project, **10** and **11** were synthesised in two steps from the commercially available per acetylated glucosamine hydrochloride salt with a 60% overall yield¹⁵⁸. The first step is the attachment of the electrophilic warhead *via* amide bond formation on a protected sugar, followed by the removal of the sugar protecting groups by hydrolysis (Scheme 3.2).



Scheme 3.2. Synthetic route for the formation of **10** and **11**. *Reagents & Conditions:* (i) acyl chloride (3 eq), TEA (2 eq) in DCM, 1 hour, 75-90%, (ii) MeONa (4 eq), MeOH, 10 mins, 70%.

As observed for the disaccharide series, both **10** and **11** were obtained as a mixture of α and β anomers (small and large J_{1-2} constant coupling respectively). The α anomer was the major product due to the stability provided by the anomeric effect as described in Chapter

2 (ratio $\alpha:\beta = 2:1$) (Figure 3.6). Additionally, not unlike **1** and **2**, the identification of individual proton signals in the ^1H NMR spectrum of both **10** and **11** is challenging due to their unprotected carbohydrate structure and the presence of two anomers. Accounting for the presence of a mixture of anomers in the ratio 2 :1, the correct number of protons is identified for both **10** (10 protons for the α anomer + 0.5 x 10 for the β anomer = 15 protons expected in total and 15.5 counted on spectrum) and **11** (9 protons for the α anomer + 0.5 x 9 for the β anomer = 13.5 protons expected in total and 13.5 counted on spectrum) (Figure 3.6). This, combined with the expected ^{13}C NMR and MS spectra provided enough evidence that the right target molecules were synthesised. **10** and **11** were then tested for their activity against LgtC.

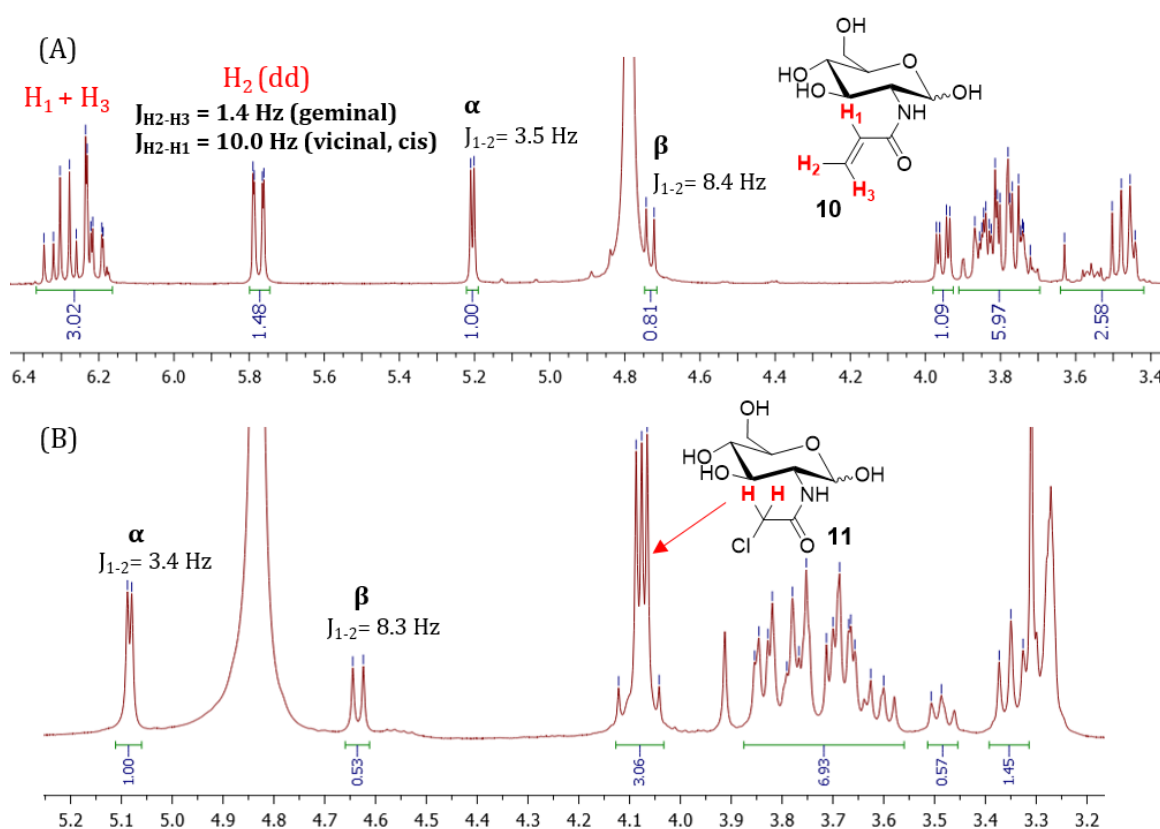


Figure 3.6. ^1H NMR (D_2O) spectra of compound **10** (A) and compound **11** (B). In both cases, a mixture of α and β anomers is obtained in the ratio 2 :1 For compd **10** the water peak does not allow for a good integration of the $\text{H}_{1\beta}$ signal. The signal corresponding to the electrophilic warhead attached is framed in red.

3.4. Initial biochemical testing: inhibition and substrate activity

Using the malachite green colorimetric biochemical assay described in Chapter 2¹⁰⁷ (Scheme 2.11), the LgtC-substrate activity of **10** and **11** was first assessed. DDT activated LgtC was incubated for 20 mins at 30°C with UDP-Gal donor (28µM) and increasing concentrations of **10** and **11**, as the partner substrate of the reaction. Increasing concentrations of the monosaccharide as the acceptor substrate partner does not result in any release of UDP, the secondary product of the enzymatic reaction (Figure 3.7). As hypothesised, this shows that neither of the monosaccharides **10** nor **11** have residual substrate activity for LgtC. In line with the literature, this indicates that, glucosamides **10** and **11** are structurally too distinct from the natural substrate scaffold to be recognised and turned over by LgtC.

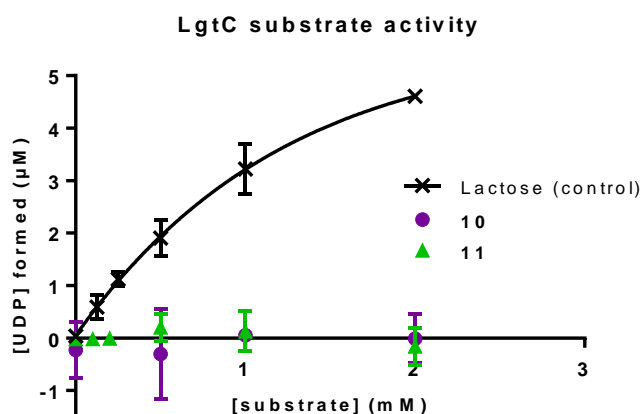


Figure 3.7. Substrate assay of compd **10** and **11** for LgtC (see Fig 2.12 for general conditions and data analysis). Neither of **10** nor **11** show residual substrate activity. Each experiment was carried out in triplicate (technical).

Having assessed that neither of **10** nor **11** had any substrate activity towards LgtC, their inhibitory properties were then tested. Increasing concentrations of **1** and **2** were incubated for 20 mins at 30°C with DDT activated LgtC, lactose acceptor (2mM) UDP-Gal donor (28µM) and all other assay components¹⁰⁷. Both candidates exhibited an inhibitory activity against LgtC with an IC₅₀ of 1.0 mM and 170 µM for **10** and **11** respectively (enzyme turnover=25%) (Figure 3.8. A). The assay was also performed against the CIP

alone as a control experiment and no inhibition of the phosphatase was observed (Figure 3.8.B). This indicates that the activity observed is genuine for LgtC.

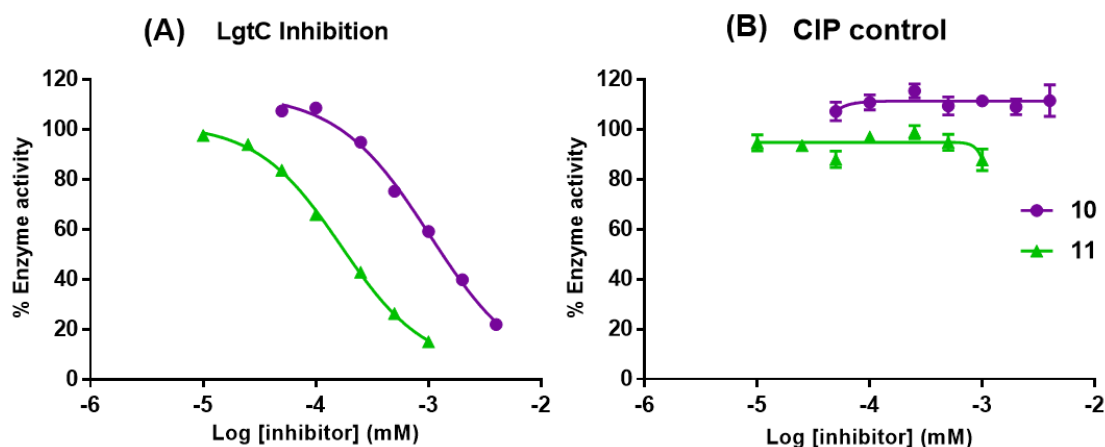


Figure 3.8. Biochemical assay of cmpds **10** and **11**. (A) Inhibition assay against LgtC (see Fig 2.11 for general conditions and data analysis) Both **10** and **11** show inhibitory activity against LgtC with an IC_{50} of 1.0mM and 170 μ M respectively. (B) CIP control inhibition. The activity observed is due to a genuine inhibition of LgtC rather than an inhibition of the phosphatase used in the assay.

Interestingly, **11** exhibited a 6-fold increase in activity compared to **10**, indicating that the structure of the warhead plays an important role in inhibition. This may be explained by a greater intrinsic reactivity of the chloroacetamide WH over the Michael acceptor (MA) group. The greater potency of chloroacetamide was also observed for the disaccharide series. However, the dual nature of **1** and **2** does not allow for a clear and quantifiable understanding of their activity towards LgtC. For this reason, the behaviour of mono- and disaccharides will not be further compared.

3.5. Investigations into the mode of action of monosaccharide inhibitors

The non-substrate nature of both monosaccharide inhibitors allowed their putative covalent mode of inhibition to be thoroughly investigated. Three experiments were conducted towards assessing the irreversible modification of LgtC by cmpds **10** and **11**:

- *Diafiltration*: The enzyme activity in the assay mixture with the most potent inhibitor **11** was compared before and after dialysis.

- *Time dependant inhibition*: The inhibition activity of **10** and **11** was investigated with increasing enzyme-inhibitor pre-incubation time
- *Mass spectroscopy*: The detection of the covalent adduct after irreversible modification of LgtC was attempted by MS.

3.5.1. Diafiltration

A dialysis experiment was first performed to gain insight into the mode of action of **11**, the more potent inhibitor. LgtC was incubated with **11** (0.5mM) under the standard assay conditions laid out previously and the enzyme activity was recorded. The mixture was then filtrated and rinsed in order to remove all free assay components. Following this, the resulting activity of the enzyme was recorded again. Although the filtration was performed at 4°C, the enzyme inevitably loses activity over the course of the experiment, and initially, the basal enzyme activity levels after diafiltration were too low to allow for a robust interpretation of the data. To address this, the reaction volume was minimised to lower the subsequent filtration time. This resulted in a sufficient post-dialysis enzyme activity to allow for a confident analysis of the data.

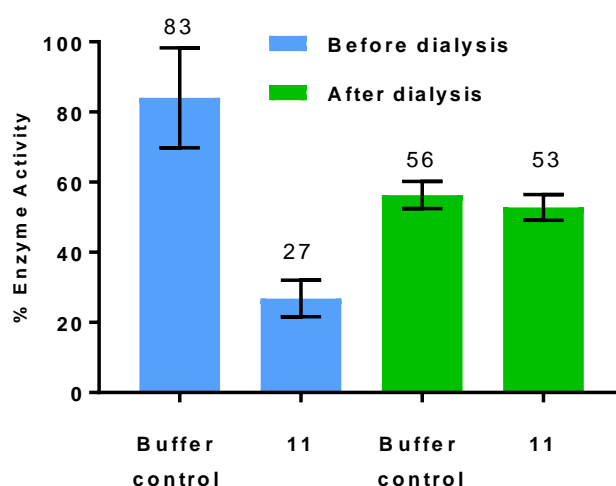


Figure 3.9. MoA studies of monosaccharide inhibitors against LgtC. LgtC activity in the presence of **11** before and after dialysis. *General conditions*: LgtC (activated with 5 mM DTT) was pre-incubated with **11** (1 mM) or HEPES buffer, in the presence of UDP-Gal (28 μ M), for 30 min at 30 °C. Lactose (2 mM) and others assay components. The reactions were incubated for 20 min at 30 °C. These samples were set up in duplicates for (i) the determination of enzyme activity before

diafiltration under standard assay conditions (ii) performing diafiltration: reaction mixtures were combined, washed with HEPES buffer and concentrated (x2) The enzyme activity after diafiltration (final residual volume) was determined under standard assay conditions. *Results:* LgtC activity is retrieved to basal level after dialysis.

The inhibition of LgtC observed prior to dialysis was lost after several washes and the enzyme activity was retrieved to basal level (Figure 3.9). For a covalent inhibitor, due to the non-reversible enzyme inactivation, the inhibition is typically maintained after diafiltration. Therefore, this unexpectedly suggests a non-covalent mode of inhibition for **11**.

3.5.2. Time dependant inhibition

As described in the introduction to this chapter, a covalent inhibition follows a two-step model (see 3.1.1). Increasing enzyme-inhibitor incubation prior to conducting the assay, allows for a greater proportion of the non-covalently bound inhibitor (1st step) to subsequently undergo the irreversible enzyme modification (2nd step). Therefore, over increasing enzyme-inhibitor pre-incubation time, a shift of the inhibition curve towards the left, corresponding to a reduction in the IC₅₀ value, is expected. In order to further assess the mode of action of our monosaccharide inhibitors, a time dependant inhibition assay was then performed with increasing enzyme-inhibitor pre-incubation periods for both **10** and **11**. Said pre-incubation was performed in the presence of UDP-Gal as it is known that the acceptor binding pocket of LgtC (assumed binding site of **10** and **11**) forms in the presence of the donor substrate¹⁰⁴ (see 1.7.3.1). Lactose is then added to the reaction mixture in order to start the enzymatic reaction.

We observed that increasing the enzyme-inhibitor pre-incubation time had no significant effect on the IC₅₀ values for compd **11** (Figure 3.10). For **10**, there may be a shift in the curve from 90 mins, but this needs to be confirmed by repeating the assay with longer pre-incubation times (Figure 3.10). However, after 100mins of incubation at 30°C, the enzyme degrades and loses most activity. For a covalent inhibitor, a time dependant inhibition is expected due to the non-reversibility of the inhibition. Therefore, our results, in line with

the previous dialysis experiment, strongly suggest a non-covalent inhibition mode for our monosaccharide inhibitors.

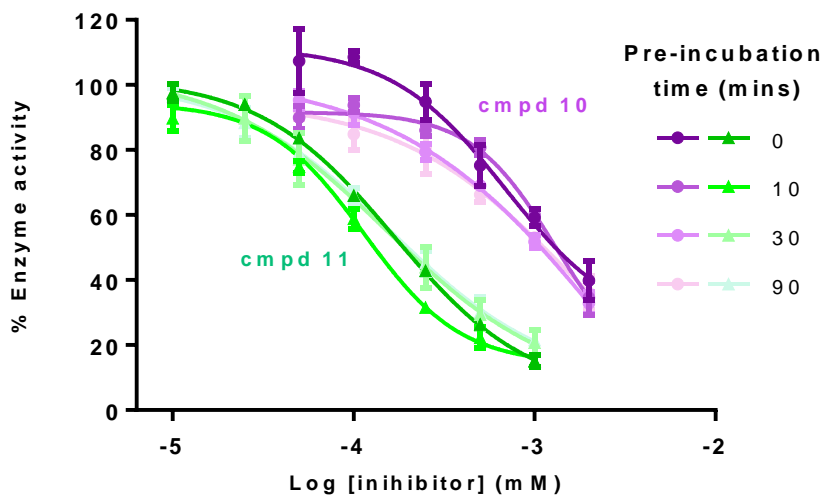


Figure 3.10. MoA studies of monosaccharide inhibitors against LgtC. Enzyme-inhibitor pre-incubation inhibition assay. *Assay conditions:* adapted from conditions given in Fig 2.11. After activation with DTT, LgtC was incubated with inhibitor (0-1 or 4mM) and all other assay components except lactose for a given time (0-90mins) at 30°C. Only then, lactose (2 mM), was added and the reactions were incubated for 20 min at 30°C. *Results:* Inhibition of **10** and **11** is independent of **10/11**-enzyme pre-incubation which is consistent with a non-covalent inhibition mode.

3.5.3. Mass spectrometry

Finally, both **10** and **11** were incubated in a 10:1 ratio with DDT-activated LgtC for 1 hour at 30°C and the resulting reaction mixture was analysed by electrospray MS. For both monosaccharides, no covalent adduct with LgtC was detected by mass spectroscopy (see Appendix 3 Fig A3.1 & A3.2). In agreement with previous observations, this suggested a non-covalent inhibition for **10** and **11**.

Combined, the results obtained so far all indicate a non- (or very slow) covalent inhibition mode for our monosaccharide candidates, which was unexpected. In addition to this, a fluorescent probe derived from the most potent inhibitor **11** was synthesised as a fourth strategy towards assessing of the mode of action of our inhibitors. For the purpose of clarity, the ability of the probe to label LgtC will be described in chapter 4.

3.6. Investigation into the reactivity of the monosaccharides towards cysteine

To gain a better understanding of the intrinsic reactivity of our monosaccharide inhibitors, both **10** and **11** were incubated with L-cysteine for 16 hours at 80mM in buffered D₂O (pH7) and the reaction was monitored over time by ¹H NMR.

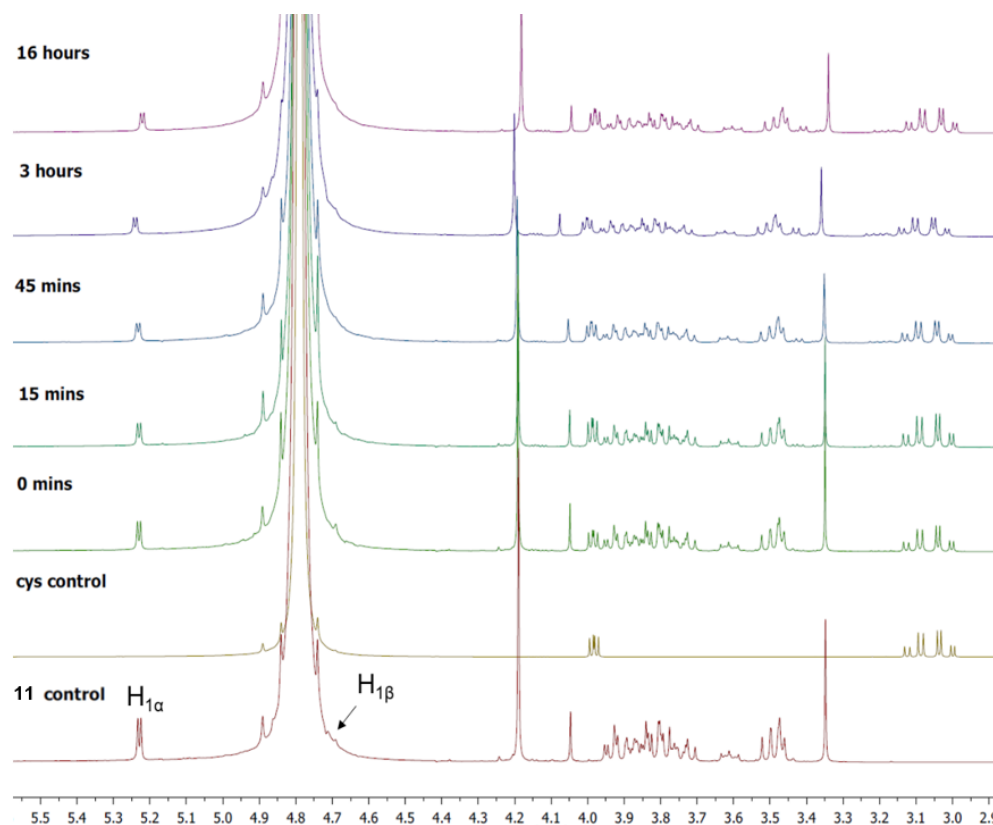


Figure 3.11. Monitoring of the reaction of **11** with cysteine (1:1, 80mM) by ¹H NMR in buffered D₂O. After 16 hours, no covalent adduct were observed.

In the case of **11**, no covalent adduct formation was observed by ¹H NMR over a period of 16 hours (Figure 3.11). Although this is consistent with previous observations, it was unexpected due to the wide use of chloroacetamide WHs for the targeting of cysteine residues for covalent inhibition^{159,86,91,160}. However, it is important to note that the reactivity of an isolated cysteine can be significantly weaker than the amino acid within an enzyme, which can be activated by neighbouring residues.

Regarding compound **10**, traces of covalent adduct formed over time, evidenced by depletion of the alkene signals as well as formation of new aliphatic peaks in the NMR spectrum (20% conversion over 16 hours) (Figure 3.12). This suggests that the barely significant effect observed at 90 mins pre-incubation time in the time dependant inhibition assay may be genuine (Figure 3.10).

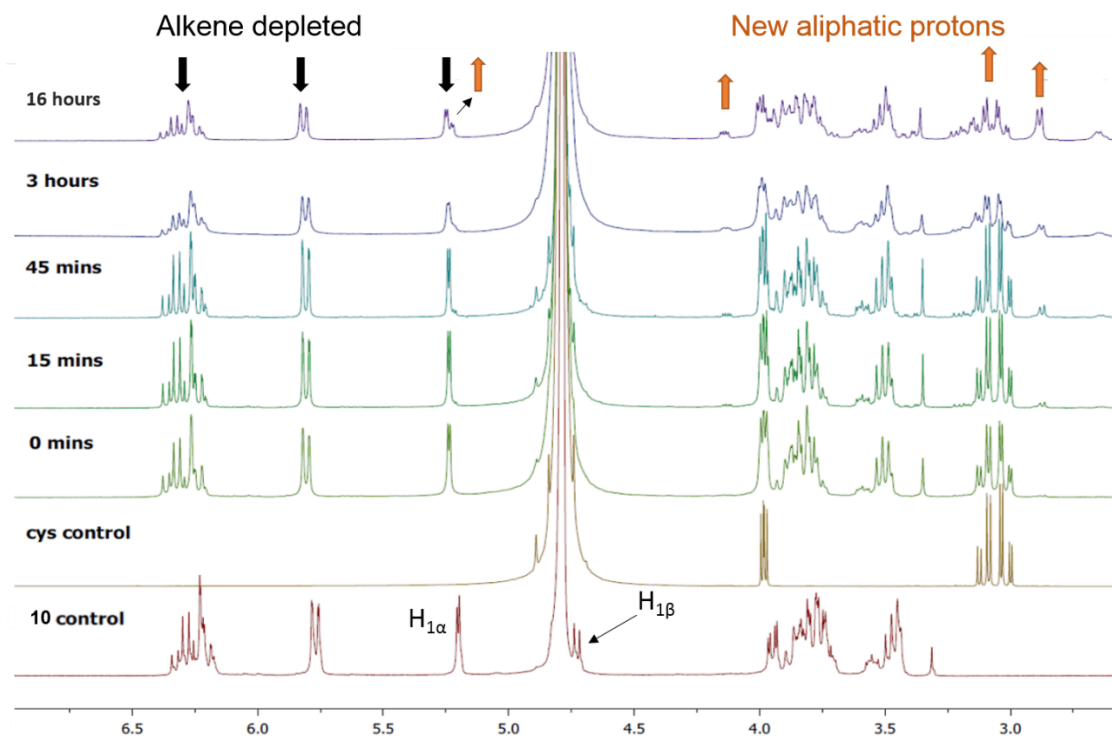


Figure 3.12. Monitoring of the reaction of **10** with cysteine (1:1, 80mM) by ^1H NMR in buffered D_2O . After 16 hours, traces of covalent adduct were observed.

The above experiments have allowed us to monitor the reactivity of our inhibitor towards cysteine over a long period of time (16hours). This was not possible with LgtC directly as the enzyme quickly degrades in solution. The results suggest a weak reactivity of **10**, and more particularly **11**, towards an isolated cysteine residue. Inspired by a recent report on the reactivity of covalent WHs towards different isolated amino acids¹⁶¹, we performed a model study with fragment WHs to investigate the reactivity of our electrophiles. This enabled us to investigate the WH as an individual entity, without considering the effect that the sugar group may have on reactivity.

3.7. Reactivity model study

An NMR-based model study was performed with commercially available warhead fragments (enamide and chloroacetamide) and their reactivity towards cysteine was investigated. Each fragment was incubated with L-cysteine at a concentration of 80mM in buffered D₂O (pH7) and the reaction was monitored overtime by ¹H NMR.

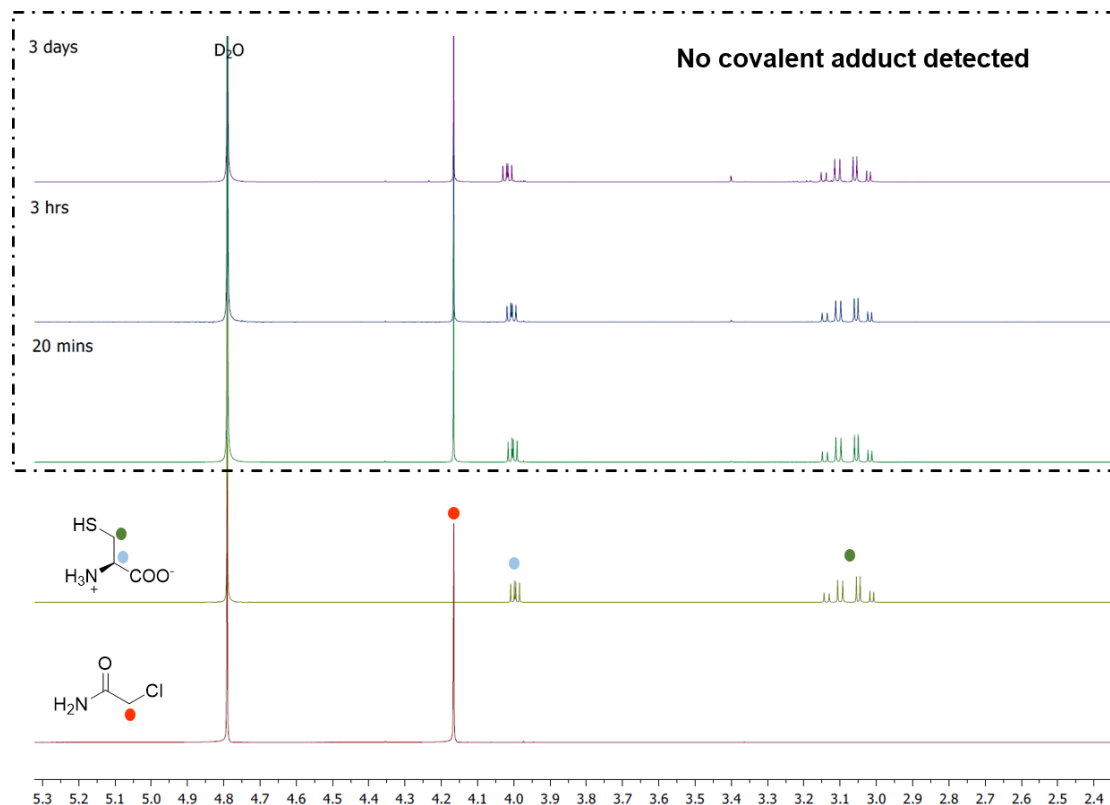


Figure 3.13. Monitoring of the reaction of chloroacetamide WH fragment with cysteine by ¹H NMR in D₂O. After 3 days, no traces of covalent adduct is observed.

While no covalent adduct was observed for the chloroacetamide warhead after 3 days (Figure 3.13), a slow reaction occurred for the enamide motif (50% conversion after 5 hours) (Figure 3.14, *top*). In line with previous observations this indicates the greater reactivity of the enamide motif towards cysteine compared to the chloroacetamide WH.

Interestingly, for the enamide group, the warhead alone reacted much faster than when attached to a sugar motif: 50% conversion after 5 hours vs 20% conversion over 16 hours respectively (Figure 3.14, *bottom*).

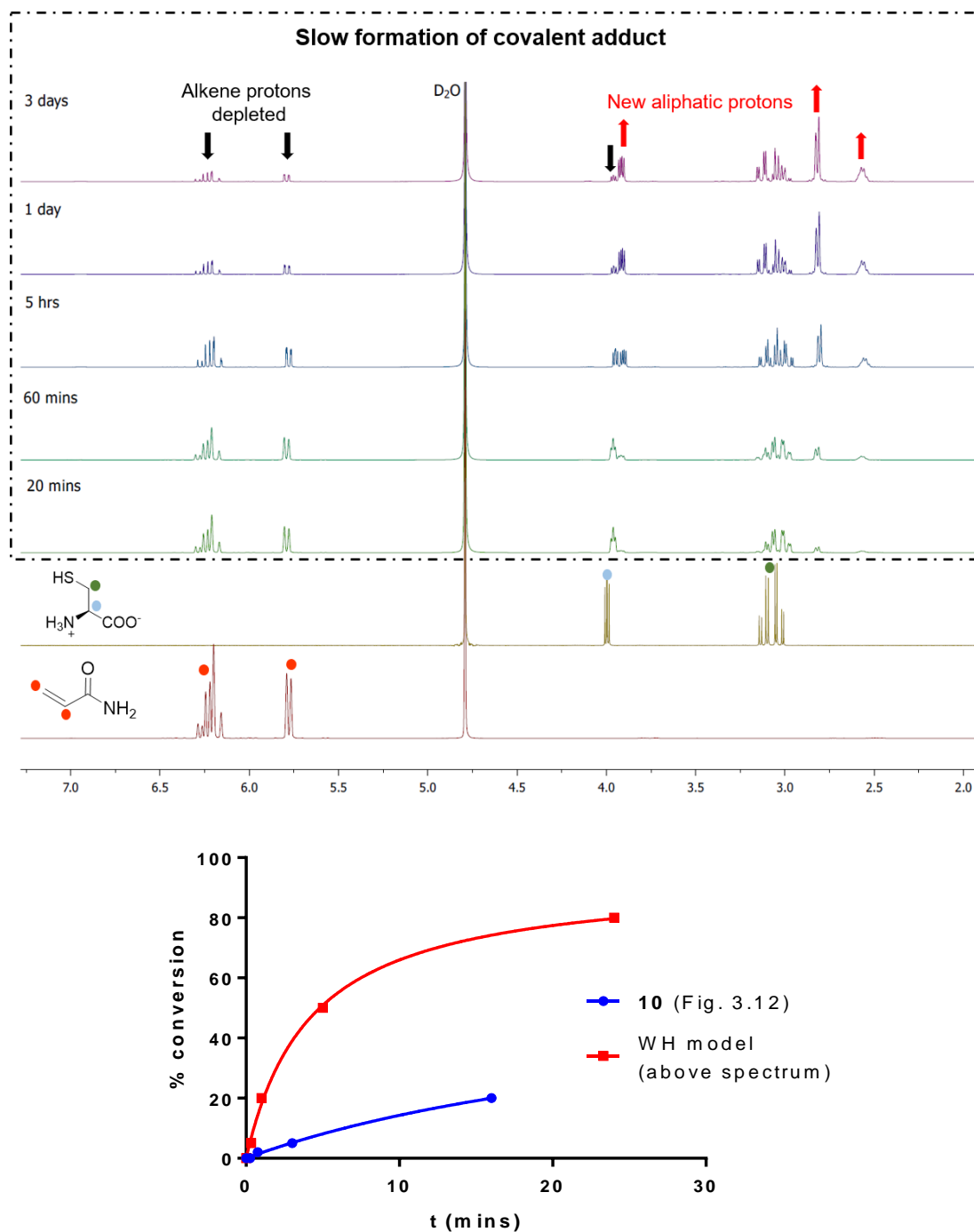


Figure 3.14. *Top:* Monitoring of the reaction of enamide WH fragment with cysteine (1:1) by ^1H NMR in D_2O . After 3 days, 50% conversion to covalent adduct is observed. *Bottom:* Plotting of % conversion (based on integrations) vs incubation time for **10** and corresponding WH model. The model warhead is more reactive towards cysteine than cmpd **10**.

Unsurprisingly, and in line with the literature¹⁶¹, a large increase in reactivity was observed for both ketone-based WHs fragments towards L-cysteine. Chloroacetone reacted relatively fast with cysteine (70% conversion after 80 mins) (see Appendix 3 Fig A3.3) while the enone fragment showed complete conversion to covalent adduct after

20 mins (see Appendix 3 Fig A3.4). The electron donating effect of the amide group strongly lowers the reactivity of both electrophilic groups by stabilising the partial charge on the reacting carbon. However, such ketone based WHs are not used for chemical biology purposes as their extremely high reactivity would result in poor selectivity.

Combined, the results of the NMR-based experiments strongly indicate the greater reactivity of the Michael acceptor WH in comparison to the chloroacetyl WH towards a cysteine residue. Additionally, the ability to target a particular amino acid depends on both the WH itself and the chemical environment around it. We have shown that the WH alone (fragment), reacted much faster with isolated cysteine than when attached to the free sugar motif. We hypothesised that the electron donating effect of a free sugar motif would stabilise the electropositive carbon of the WH, therefore making it less reactive towards a nucleophilic residue (Figure 3.15). For direct comparison, and confirmation of the hypothesis, the acetylated precursor of **10**, compd **8**, could be tested in a similar NMR-based study. In theory, the electron withdrawing effect of the protecting group should have the reverse electronic effect and increase the reactivity of the WH in comparison to the fragment (Figure 3.15). However, this has yet to be tested and optimisation of the conditions will be required as compd **8** is not soluble in aqueous solutions.

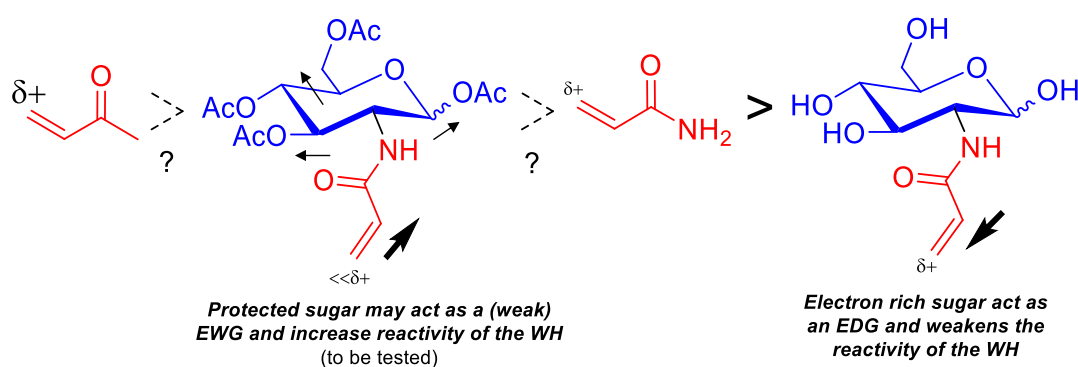


Figure 3.15. MA WH placed in order of increasing reactivity (proven and putative)

Although these data are a model study on an isolated cysteine residue and do not reflect what happens in a much more complex protein environment, they provide insight

regarding the reactivity of the candidate compounds corroborate the non-covalent inhibition mode observed for **11** in the time scale tested. The observations gathered raise important questions regarding the design of carbohydrate-based covalent inhibitors as the results suggest that their reactivity may be much lower than expected.

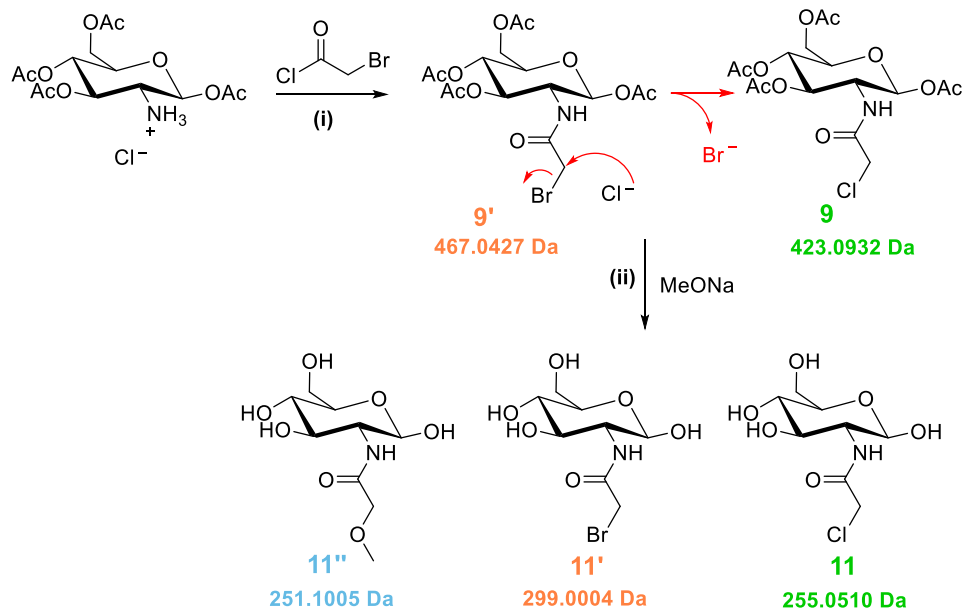
3.8. Attempted synthesis of bromoacetamide monosaccharide analogue

To address the issues encountered above, the synthesis of **11'**, an analogue of **11** containing a bromoacetamide WH, was attempted. Bromide ion is a better leaving group than chloride ion, therefore the reactivity of **11'** towards a nucleophilic residue is greater. Although less widely used than chloroacetamide, the covalent modification of biological targets using bromoacetamide-based probe has been reported^{162,163}. Additionally, a bromoacetamide-containing molecule is currently in phase III of clinical trials for anti-cancer therapy (see 3.1.2.2).

Peracetylated glucosamine hydrochloride salt was therefore reacted with bromoacetyl chloride for the formation of compd **9'** (Figure 3.16.A, step (i)) under the conditions previously described in Scheme 3.2. ¹H NMR after normal phase chromatography shows the presence of a mixture of two products (ratio 3:1) with the same proton signals suggesting a very high level of structural similarity (See Appendix 2 Fig A2.1). Both compounds have a β -configuration at the anomeric position which excludes the possibility that a mixture of **9'** anomers was formed (See Appendix 2 Fig A2.1). All proton signals are superimposed except for the WH peaks (NH and CH₂-X) which suggests that the structural difference is likely in this part of the molecule. The CH₂-X signal of the minor product (3.98 ppm) indicates that it is likely to be compd **9** previously isolated, while the major product (CH₂-X, 3.76 ppm) appears to be the expected compd **9'**. CH₂-X protons are expected to be more shielded when X=Br than X=Cl. Mechanistically, this observation can be explained by the presence of chloride ions in solution (from the salt) which could displace the

bromide in **9'** for the formation of **9** (Figure 3.16.A). The greater reactivity of **9'** compared to **9** make this explanation plausible.

(A)



(B)

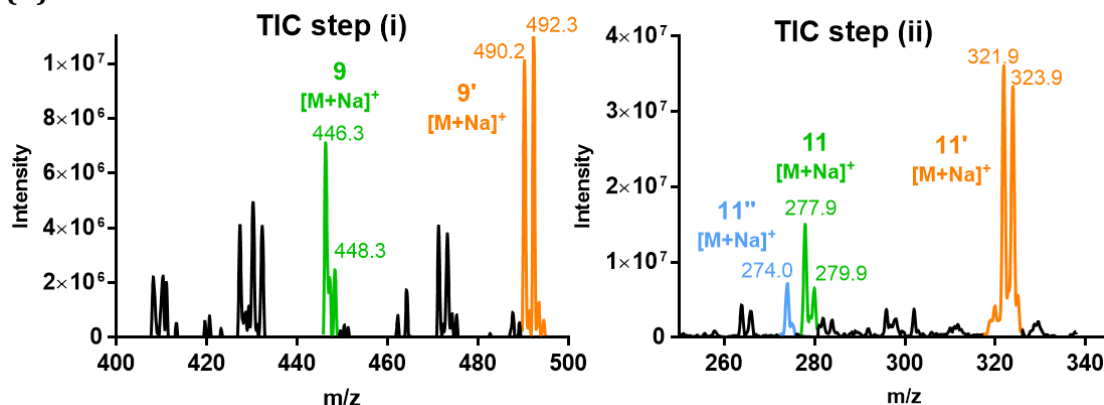


Figure 3.16. Attempted synthesis of bromoacetamide analogue. (A) Synthetic route for the formation of **11'**. Reagents & Conditions: (i) acyl chloride (3 eq), TEA (2 eq) in DCM, 1 hour, 75-90%, (ii) MeONa (4 eq), MeOH, 10 mins, 70%. (B) LC-MS analysis of pure sample obtained after normal phase purification. Conditions: Eclipse XDB-C8 column, gradient: MeOH (0.1% TFA) in H₂O (0.1% TFA) flow rate: of 1 mL/min, detecting at 210 nm.

MS analysis confirms the hypothesis that the mixture after step (i) is composed of compd **9** (Cl) and **9'** (Br) (Figure 3.16.B). These two compounds were not separable by normal phase chromatography under the conditions attempted (gradient 50-100% EtOAc in hexane and 10-30% Acetone in toluene). After deacetylation (step (ii)), a mixture of compd **11** (Cl), **11'** (Br) and **11''**(OMe) are observed by MS (Figure 3.16.B). Due to the high

reactivity of **9'**, methoxide ions can displace the bromide ion under these conditions resulting in the formation of **11''** (Figure 3.16.A). Therefore, after deacetylation, a complex mixture of **11**, **11'** and **11''**, was obtained. To date, the purification conditions attempted were unsuccessful (preparative HPLC, and normal phase chromatography), and therefore further investigation of **11'** was not possible.

A study has shown that acryl- and chloroacetyl amides exhibit a surprisingly low non-specific reactivity suggesting the need for a strong non-covalent binding to position the WH in the vicinity of the target residue for successful covalent modification⁹¹ (see 2.1.2). Therefore, the above results may surprisingly indicate that these conditions are not met in our case which could be explained by a different binding mode than expected. Unlike compound **1** and **2**, neither **10** nor **11** have residual substrate activity, therefore, so far, no evidence was gathered regarding the binding mode of the monosaccharides to the enzyme. To gain insight into this, substrate competition experiments were conducted.

3.9. Substrate competition experiments

To better understand the binding mode of our monosaccharides to LgtC, kinetic competition experiments were performed for the most potent inhibitor **11**. First, the effect that the concentration of lactose substrate may have on inhibition was investigated under initial velocity conditions (enzyme turnover <10%) and fixed concentration of UDP-Gal (100 μ M). The velocity of the reaction was calculated over 3 time points (0, 10 and 20 mins of incubation) and corresponds to the gradient of the linear curve [UDP] formed = $f(t)$. Surprisingly, we observed that the inhibition cannot be overcome by increasing lactose concentration. The Michealis Menten curve in Figure 3.17 A. shows that V_{max} is decreased with increasing inhibitor concentration, but the substrate binding is mostly unaffected (K_m unchanged). This is characteristic of a non-competitive inhibition mode for lactose and raises the possibility that, unexpectedly, **11** may bind outside the acceptor

binding site. The same experiment was therefore performed to test for potential donor competition at low enzyme turnover (<10%) and fixed concentration of Lactose (5mM). Similarly, we found that the inhibition of **11** was non-competitive for UDP-Gal, the donor substrate of LgtC (Figure 3.17 B). The above results therefore indicate that **11** may bind outside the binding pocket of LgtC.

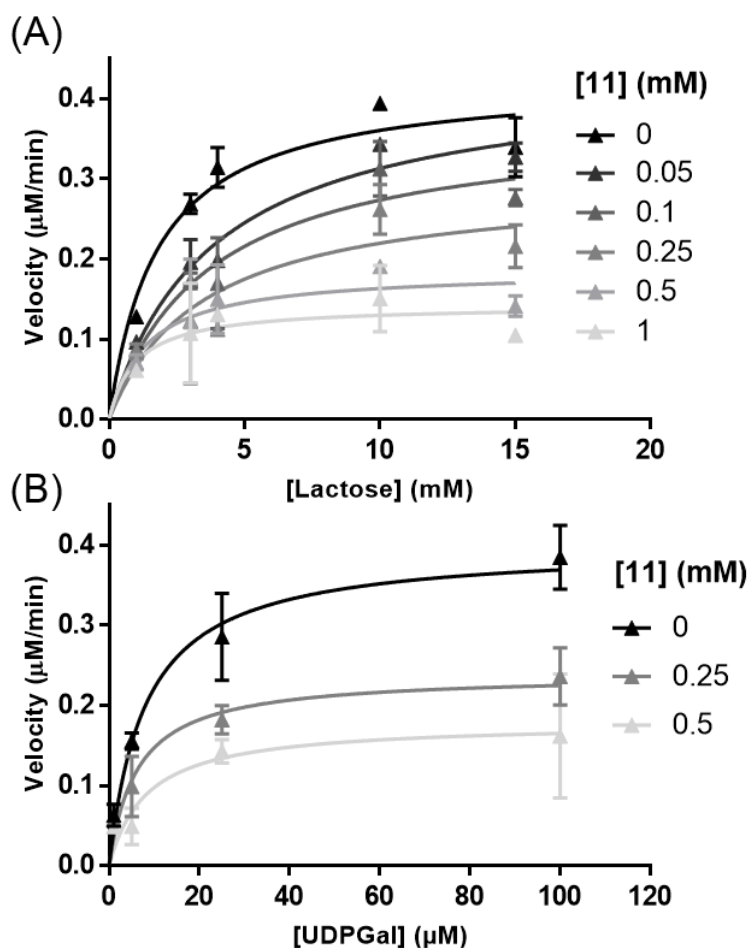


Figure 3.17. Substrate competition experiments for **11** (A) Lactose competition *Assay conditions:* Activated LgtC was incubated with **11** (0-1 mM), UDP-Gal (100 μM), MnCl_2 (5 mM), CIP (10 U/ml), CEL (1 mg/ml), Triton (0.01%) and lactose acceptor (1-15 mM) at 30 $^\circ\text{C}$ in 13mM HEPES buffer (pH 7.0). Velocity ($d[\text{UDP}]_{\text{formed}}/dt$) was calculated at three time points 0, 10 and 20 mins of incubation. UDP-Gal turnover was kept <10%. Data were fitted to a Michaelis Menten plot with GraphPad Prism. *Conclusions:* **11** is non-competitive for Lactose. (B) UDP-Gal competition (see (A) for assay conditions) *Alterations:* Fixed lactose concentration (5mM), increasing donor concentrations UDP-Gal donor (0-100 μM). *Results:* **11** is non-competitive for UDP-Gal.

A potential explanation for these unexpected observations is that **11** may have the ability to occupy both active sites. However, this is unlikely as when assessing the competition for one substrate we ensured that the partner substrate was saturating its binding site.

These preliminary results suggest that **11** may have a very different binding mode than anticipated and may behave as an allosteric inhibitor, which would also be consistent with the absence of residual substrate activity observed for **11**. Because of the relatively small size of our monosaccharide targets and the known function of LgtC as a carbohydrate binding protein, an allosteric binding mode for **11** may be possible. However, no other binding domain is known for LgtC, therefore this proposed binding mode remains highly hypothetical.

3.10. Inhibition of non-electrophilic analogues

Lastly, towards gaining a better understanding of the inhibition mode and the role of the WH for inhibition, two non-electrophilic monosaccharide analogues were tested: the saturated analogue of **10**, cmpd **10s**, synthesised in house using the route described in Scheme 3.2, and the commercially available N-acetyl-glucosamine (GlcNAc). These were tested against LgtC in an inhibition assay as previously described, including cmpd **10** as a positive control. Increasing concentrations of **10**, **10s** and GlcNAc were incubated for 20 mins at 30°C with DDT activated LgtC, lactose acceptor (2mM) UDP-Gal donor (28µM) and all other assay components¹⁰⁷.

Interestingly, while GlcNAc had no inhibitory activity against LgtC up to 10mM, cmpd **10s** inhibited LgtC with an activity comparable to its electrophilic analogue **10** (Figure 3.18). These results indicate the great importance of the presence of a substituent at the β -position of the carbonyl for inhibition; as for GlcNAc, which lacks this feature, all inhibitory activity was lost. Additionally, the observed activity of non-electrophilic cmpd **10s** may provide further evidence of the non-covalent inhibition mode for this series of inhibitors.

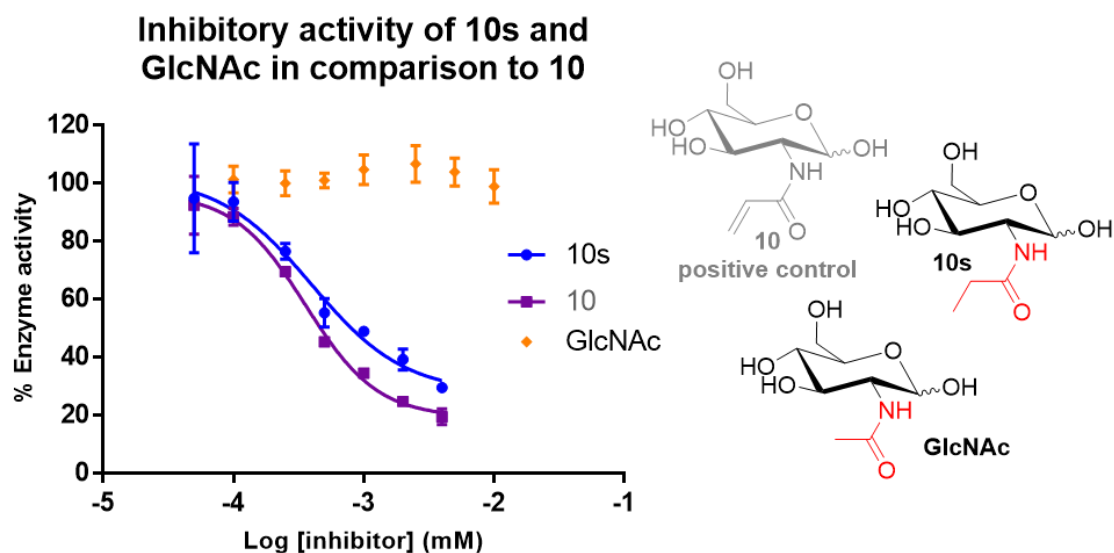


Figure 3.18. Inhibition assay of GlcNAc and cmpds **10** and **10s** against LgtC (see Fig 2.11 for general conditions). **10s** is an inhibitor of LgtC while GlcNAc has no inhibition activity.

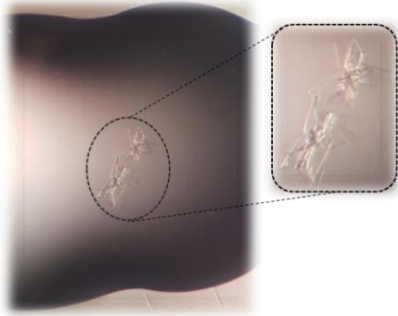
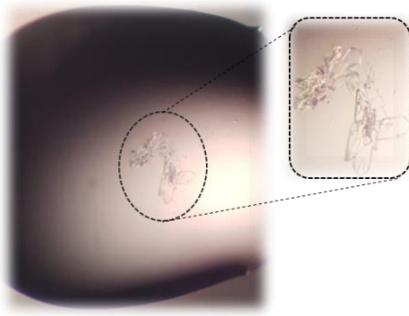
The dataset gathered so far for the monosaccharide series is highly intriguing and raises some important questions regarding both their mode of action and their binding. In order to get a definitive answer regarding these two aspects, some preliminary crystallography work with the aim of obtaining the structure of LgtC in complex with **11** was performed.

3.11. Preliminary crystallography work

LgtC (15mg/mL) was incubated with **11** (20mM) and the reported successful conditions for the crystallisation of LgtC in complex with analogue substrates were tried (50 mM NaOAc, pH 5.0, and 5–20% (w/v) PEG monomethylether 2000). Unfortunately, no crystals were obtained using this method. Coarse screen crystallography trays were then set up using the commercially available screening kits INDEX, PACT, JSCG and SALTRX. Out of the 384 conditions screened, crystals had grown in two wells within 1 hour (Table 3.1). Interestingly, the two wells containing crystals corresponded to very similar screened conditions (~1M ammonium phosphate dibasic, pH=8.5). A fine screen plate was therefore set up (0.5 - 2M ammonium phosphate dibasic, pH= 7 - 10 in both solvents) with and without proteins (control wells). Unfortunately, the control wells (absence of protein)

also contained similar plate-shaped crystal which indicates that the obtained crystals are from salt rather than protein.

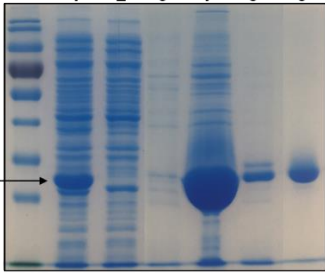
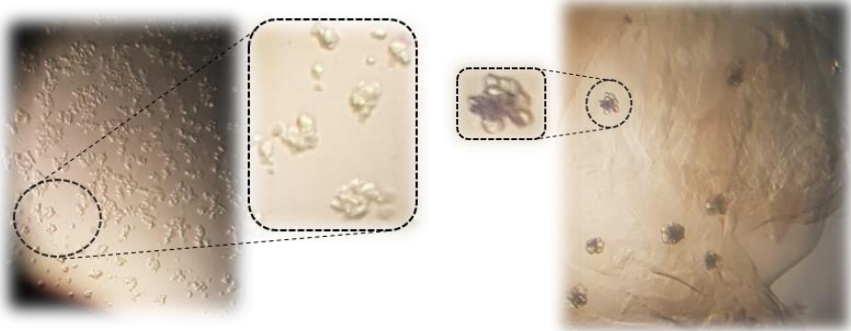
Table 3.1. Crystallisation of LgtC in complex with **11**: First set of coarse screens

Crystals obtained		
Conditions	<p>JCSG E.8 1 M Ammonium phosphate dibasic, 0.1 M sodium acetate pH 8.5</p>	<p>Salt RX E.3 1.5 M Ammonium phosphate dibasic 0.1 M Tris pH 8.5</p>

To increase our chances, fresh enzyme was expressed and purified with an additional size exclusion chromatography step (compared to standard protocol described in Chapter 7). This resulted in an enzyme batch of very high purity (Table 3.2, SDS-page lane 6). The newly expressed batch of LgtC was used for crystallography without freezing, concentrated to 15mg/mL and incubated with **11** (20mM). Coarse screen crystallography trays were set up as previously and stored at 16°C. This time, in addition to the salt crystals obtained in wells JCSG E.3 and Salt RX E.3, new crystals formed in wells PACT H.6 and INDEX H.9 within 3 days (Table 3.2). Unlike the previously obtained crystals which were hard, these were soft and malleable, which is characteristic of protein crystals. Additionally, formed over longer periods of time, these crystals have grown out of protein precipitation. They have formed under very different wells conditions and have a different shape than the previously obtained salt crystals. All the observations above indicate that the observed crystals may be of protein and similar morphology pseudo- nanocrystals have been reported before for CTB-MPR, a protein involved in HIV¹⁶⁴. Further optimisation of the conditions developed for CTB-MPR protein led to full size crystals

which have allowed for the structure of the protein to be fully elucidated¹²⁰. Therefore, the above results obtained for LgtC (Table 3.2) are promising and merit further investigation.

Table 3.2. Crystallisation of LgtC in complex with **11**: Second set of coarse screens

LgtC purification		
Crystals obtained		
Conditions	<p align="center">Pact H.6 0.2 M Sodium formate 0.1 M Bis Tris propane pH 8.5, 20% (w/v) PEG 3350</p>	<p align="center">Index H.9 0.05 M Zinc acetate dehydrate 20% PEG 3350</p>

3.12. Summary and conclusions

In this chapter, the synthesis, biochemical testing and mode of action study of two monosaccharide inhibitors of LgtC was described. **10** (Michael acceptor WH) and **11** (chloroacetamide WH) are the glucosidic analogues of compd **1** and **2** respectively (Chapter 2). They were chosen as suitable candidates following the very weak LgtC-substrate properties reported for glucose¹³⁵. We therefore hypothesised that an additional WH on a glucosamine scaffold will not display any substrate activity. **10** and **11** were synthesised in two steps from commercially available per acetylated glucosamine hydrochloride, in 60% yield. Unlike their disaccharide analogues, compounds **1** and **2** described in chapter 2, and in line with our suppositions, neither of the monosaccharides showed residual

substrate activity for LgtC. Additionally, both monosaccharides exhibited inhibition of LgtC with an IC_{50} of 1.0 mM and 170 μ M for **10** and **11** respectively. An NMR study into the mode of action of fragment WHs showed that on their own, acryl- and chloroacetamides are not highly reactive WHs, which is in agreement with the reported literature⁹¹. We have also identified that the sugar scaffold, with its electron donating properties, results in a further decrease in the reactivity of the WHs towards isolated cysteine residues. For acryl- and chloroacetamide-containing TCIs, this highlights the need for a strong ligand/enzyme non-covalent binding, to specifically position the WH in the vicinity of the activated target residue for successful covalent modification⁹¹. However, surprisingly, the analysis of the mode of action of **10** and **11** towards LgtC allows us to conclude with confidence that their inhibition mode was non-covalent, which may indicate a different binding mode than initially anticipated. In fact, the preliminary results of substrate competition assays suggest that the most potent inhibitor **11**, may exhibit an unprecedented allosteric inhibition of LgtC. An unexpected binding with no nucleophilic residue in the vicinity may also explain the non-covalent mode of action observed for **11**. However, at this stage, this remains highly hypothetical. Additionally, the biochemical testing of two non-electrophilic analogues (**10s** and GlcNAc) gave insight into the key role of a substituent in β -position to the carbonyl for inhibition, as the absence of such substituent resulted in total loss of inhibition. In order to gain a definite answer regarding the mode of action of the monosaccharide series, some preliminary crystallography work was performed, and promising conditions were identified for the formation of protein crystals which warrant further investigation.

As mentioned in p.108, in parallel to the work discussed in this chapter, a fluorescent probe, derived from the structure of the most potent inhibitor **11**, was synthesised as an additional method for assessing the inhibition mode of the monosaccharide. On the basis of the results obtained in this chapter, knowing that **11** does not bind covalently to LgtC, it may now seem out of place to design a covalent probe from a non-covalent inhibitor.

However, at the time no conclusion regarding the mode of action of **11** had been clearly drawn. In this chapter, we provided evidence that the chemical environment around the WH can have a significant impact on the its reactivity. Therefore, the addition of a chemical fluorophore on **11**'s scaffold may affect the WH reactivity as well has the binding of the probe. The next chapter will discuss the synthesis and biochemical labelling of LgtC using said probe.

Chapter 4

Carbohydrate-based fluorescent probes for the labelling of LgtC

In parallel to the work described in the previous chapter, a fluorescent chemical probe, derived from **11**'s scaffold, was designed with the primary aim of further assessing the mode of action of the monosaccharide towards LgtC. Additionally, should this probe be successful for the labelling of LgtC, it could be applicable to the labelling of a wider range of bacterial GTs. This was the secondary aim we were keen to investigate when designing the probe. This chapter will describe the synthesis of the probe, its structural and spectroscopic characterisation as well as its ability to label recombinant LgtC, in cell lysate and *in vivo*. Before discussing the results, existing covalent probes for GTs and GHs will be reviewed. The use of CuAAC for the design of potent carbohydrate-based probes and inhibitors will also be discussed.

4.1. Introduction

In chapter 1, the importance of phase variation as a key adaptation mechanism of pathogenic bacteria was described (see 1.7.2). We saw that the expression of LgtC was significantly subject to this biological phenomenon in a variety of *Haemophilus* and *Neisseria* pathogenic strains. This suggests that inhibitors of LgtC only have a limited potential as anti-virulence therapeutics. However, the development of covalent probes for the labelling of LgtC and other phase variable enzymes *in vivo* and in cell lysate can have far reaching applications, such as the identification and monitoring of resistance and virulence markers. Carbohydrates are a good template for probing virulence factors as a large class of virulence traits are involved in the regulation of surface glycoconjugates and are therefore carbohydrate binding proteins. Thus, although the impact of this project on the direct development of new therapeutics is limited, it can help towards a greater understanding of the molecular mechanisms underlying bacterial infections and host interactions, which is key for the development of novel and efficient antimicrobials.

In chapter 1, a general introduction on covalent probes was also provided (see 1.9). We discussed that they can be divided into two categories: activity-based probes (ABP), also known as mechanism-based probes (MBPs) and affinity labels (AfBP). ABPs are structural analogues and require activation by the enzyme while AfBP can be either substrate-based or non-substrate related. For covalent cross-linking, AfBP can be sub-divided into external trigger-dependant probes (i.e. UV radiation for photoaffinity labels) or intrinsically reactive probes. The latter react spontaneously with their target protein upon binding, due to the presence of an electrophilic WH for the targeting of noncatalytic residues such as cysteine or lysine.

Because of their important role in virulence, many efforts have been applied towards the development covalent probes for carbohydrate-active enzymes. Over the last couple of decades, a sizable number of chemical probes have been developed for GHs^{103,165}, however the same cannot be said for GTs. This is due to a number of different factors: the complex and not yet fully understood multi-substrate mechanisms of GTs, their conformational plasticity¹⁰⁴ and the significant lack of structural information for this class of enzyme. In addition to this, catalytic nucleophiles are more common for GHs than GTs¹⁶⁵, which facilitate the development of covalent probes for the former enzyme class.

4.1.1. Covalent probes for glycosyl hydrolases (GHs)

Numerous examples of both activity-based probes (ABP) and affinity labels (AfBP) have been reported for GHs^{103,166–168}. They have been widely applied to proteomics-based profiling and live cell labelling. Unlike for GTs, the catalytic mechanism of GHs is fully understood. While, inverting GHs follow a direct displacement mechanism, retaining GHs undergo a double displacement mechanism *via* the formation of a covalent intermediate initiated by a catalytic nucleophile (often a carboxylate motif) (Scheme 4.1).

Photoactivatable affinity-based probes (AfBPs) have been explored as tools for glycosidase profiling. Typically, the design of such probes relies on a sugar targeting group decorated with a suitable photoreactive group and a reporter group. Photoaffinity labels can be designed for both retaining and inverting glycosidases which allows them to be widely applicable. However, due to the unpredictable target of the cross-linking event, photoaffinity labels lack specificity. A recent example is the glucolipid designed by Sakurai *et al.* equipped with either a diazirine or a benzophenone as the photoreactive group and a BODIPY motif as a sensitive fluorescent reporter group (Figure 4.1)¹⁶⁹. Their probe was used for the tagging of almond β -glucosidase, catalyzing the hydrolysis of the β -glucosyl ceramide.

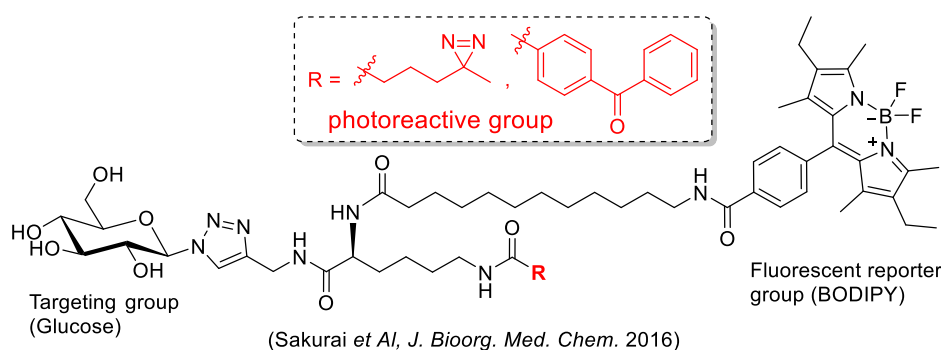
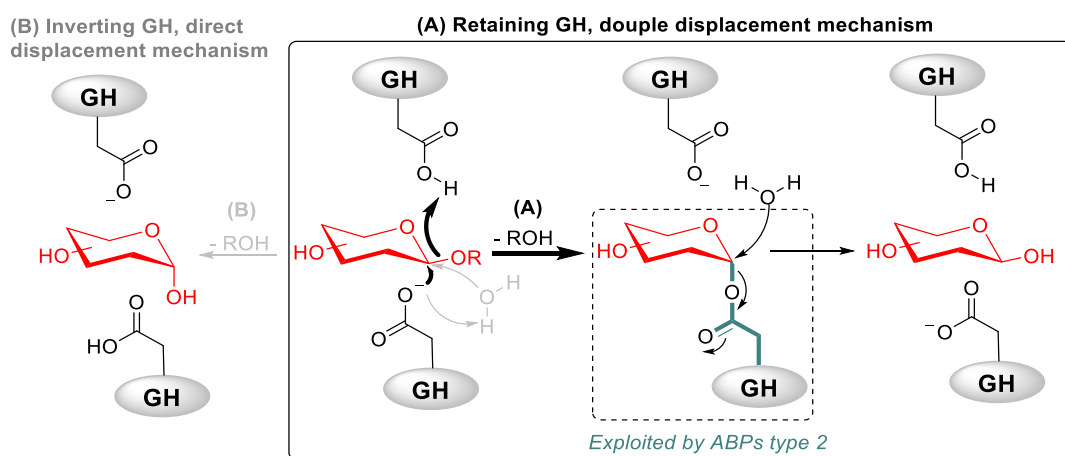


Figure 4.1. Example of AfBP for GHs

Unlike affinity labels, ABPs require catalytically active enzymes for the cross-linking event to occur. ABPs for GHs are of high interest as they can provide insights into the role of these enzymes in complex biological systems and enable the identification of previously unknown GHs. Two types of ABPs have been reported for GHs:

- *Type 1:* These probes undergo enzymatic hydrolysis and the released aglycone reacts covalently with a nucleophilic target residue in the vicinity of the active site.
- *Type 2:* These probes exploit the mechanism of retaining GHs by forming a stable covalent bond directly with the catalytic nucleophile of the enzyme (Scheme 4.1).

Therefore, ABPs of the first type can be designed for both retaining and inverting GHs, while the second type is limited to retaining GHs.



Scheme 4.1. Catalytic mechanism of GHs

A recent example of the first type of ABP, was developed for the labelling of an α -fucosidase of *Thermotoga maritima* in protein extracts¹⁷⁰. This probe was designed with α -L-fucose head group serving as recognition motif and a latent trapping moiety containing a *o*-fluoro- or *o*-difluoromethyl group, connected *via* a PEG linker to a BODIPY fluorophore (Figure 4.2). The labelling relies on the formation of a reactive quinone methide fragment upon enzymatic hydrolysis of the sugar motif from the trapping unit. The *o*-fluoromethylphenol derivative exhibited a significantly higher labelling efficiency and was successfully applied for the visualization and localisation of lysosomal α -L-fucosidase activity in human cells¹⁷⁰.

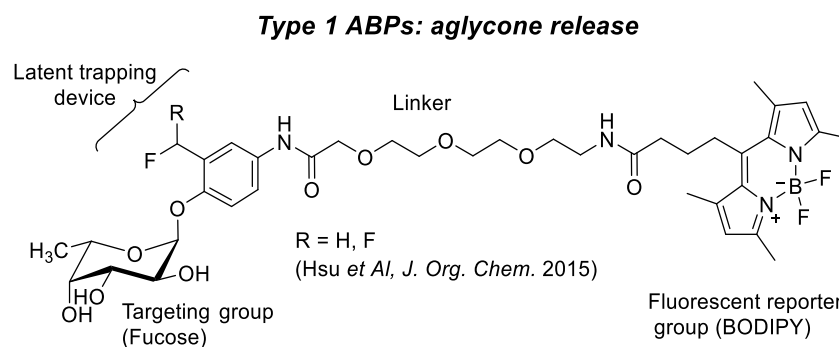


Figure 4.2. Example of a type 1 ABP for GHs

A wide range of scaffolds have been used for the development of the second type of GHs ABPs. Among them, activated 2-deoxy-2-fluorosugar are probably the most commonly used molecules. The design of such probes relies on the combination of an activated anomeric leaving group (LG) with an electron-withdrawing fluorine substituent in the C-2 position of the sugar unit. The LG promotes the formation of the covalent intermediate and the fluorine atom drastically reducing the rate for subsequent hydrolysis. This strategy enables the extension of the covalent 'glycosyl'-enzyme intermediate lifetime. A recent example where this strategy was employed for the labelling of retaining GHs is the ^{18}F -radiolabeled substrate analogue of acid β -glucocerebrosidase¹⁷¹, which catalyses the hydrolysis of β -glucosylceramide to ceramide and β -D-glucose, a deficient enzyme in Gaucher Disease (Figure 4.3.B).

However, in some cases, fluorosugars, are turned over too rapidly to be of use as probes¹⁶⁸. ABPs employing an electrophilic trap mechanism for target labelling can circumvent this. Cyclophellitol, a β -D-glucose with a highly electrophilic anomeric epoxide instead of an activating aglycone is a great example of this. Retaining β -glucosidase inactivation proceeds *via* a stereospecific ring opening of the epoxide by the enzyme's catalytic nucleophile, leading to irreversible alkylation of this residue. The first cyclophellitol derived ABP was reported in 2010 by Witte *et al.*, with a BODIPY reported group at the C6 position of the cyclophellitol glucose motif¹⁷² (Figure 4.3.A). This epoxide-based ABP was highly potent and selective for lysosomal glucocerebrosidase. However, the bulky C6 substituent had a strong impact on the probe binding against other retaining β -glucosidases, making this first generation of cyclophellitol-derived ABPs not applicable to a wider target range. Kah-Yee *et al.* addressed this by developing N-acylated cyclophellitol aziridines in which the reporter group occupies the position of the substrate aglycon, away from the binding pocket, leading to broad spectrum ABPs for the labelling of all retaining β -glucosidases¹⁷³ (Figure 4.3.A).

Type 2 ABPs: direct target of catalytic nucleophile (retaining GHs)

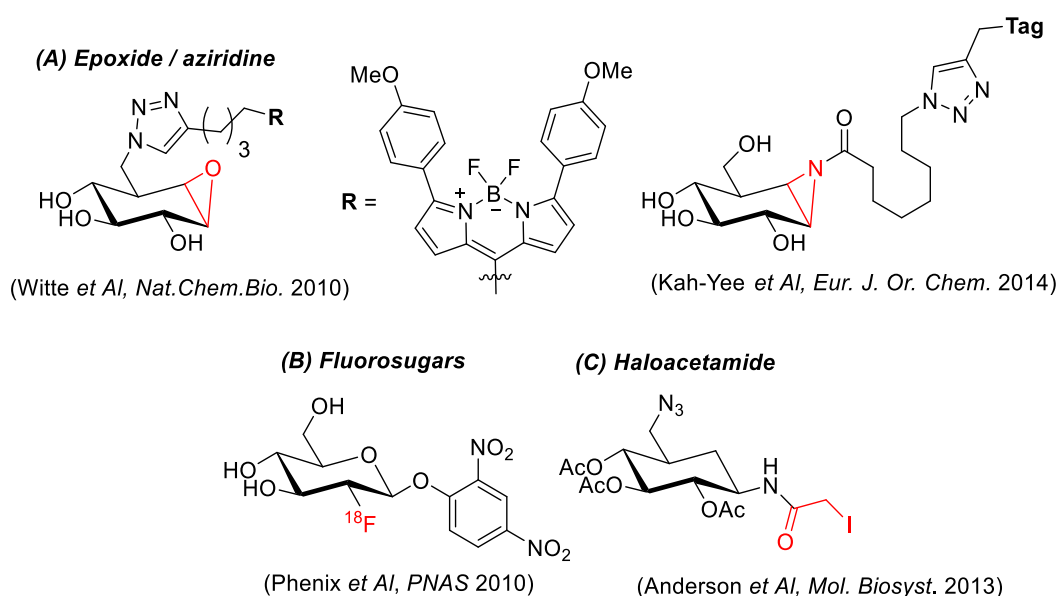


Figure 4.3. Examples of ABPs for GHs

Although there are relatively few reports of ABPs based upon haloacetamide-based inhibitors, Anderson *et al.* reported the use of an azido N-iodoacetyl-glycosylamine probe clicked *in vitro* with tetramethylrhodamine for the profiling of secreted cellulolytic enzymes of several *Trichoderma reesei* strains¹⁷⁴ (Figure 4.3.C).

ABPs for retaining glycosidases can be successfully applied for activity-based protein profiling studies (ABPP). 2-deoxy-2-fluoroglycosides and cyclophellitol derivatives, which form a selective and irreversible adduct with their target, have been applied to the global profiling of their target in cells and cell extracts. This was first described by Vocadlo and Bertozzi in 2004¹⁷⁵.

4.1.2. Covalent probes for glycosyltransferases

In comparison to GHs, covalent probes for GTs remain scarce. The design of ABPs for retaining GHs relies on the presence of a catalytic nucleophile, a feature which is absent in most retaining GTs. Alternative strategies are therefore required for the design of ABPs for GTs. The complex multi-substrate mechanism of GTs, their conformational

plasticity¹⁰⁴, and the significant lack of structural information may have also hindered the development of covalent inhibitors and probes for this class of enzymes.

Most covalent probes designed for GTs are UDP-sugar analogues¹⁷⁶. [β -³²P]5N₃UDP-Glc has been reported as an photoaffinity probe for the labelling of human liver microsomal UDP-glycosyltransferases and membrane-associated UDP-glucose:dolichylphosphate glucosyltransferase¹⁷⁷ (Figure 4.4.A). Another radiolabelled photoaffinity probe was designed by incorporating a ¹²⁵I-containing azido salicylic acid moiety in the UDP-GlcNAc scaffold for the labelling of GlcNAc Transferase I¹⁷⁸ (Figure 4.4.B). Upon exposure to UV light, a highly reactive nitrene intermediate is generated through activation of the azido group. Because of its extremely short half-life, the nitrene intermediate reacts with any amino acid side-chain or peptide backbone atoms, making it poorly selective. More recently, a UDP-Gal analogue was modified at the sugar moiety for the rational design of an affinity probe successfully labelling human recombinant galactosyltransferase (β GalT-1) at Trp310 residue¹⁷⁹ (Figure 4.4.C).

Very recently, a non-substrate-like covalent inhibitor of LgtC, targeting non catalytic Cys246 in the acceptor binding pocket was discovered,⁷⁵ providing ground for the development of a novel class of GT AfBPs. Structure–activity relationship studies in this series of inhibitors suggested that the phenyl group directly attached to the of pyrazol-3-one scaffold pointed away from the binding site (see 1.7.5.3), making it a suitable position for the installation of a non-binding interfering reporter group. A PEG-based linker containing a latent alkyne handle was therefore attached to the phenyl ring for the attachment of a fluorophore reporter group *via* CuAAC *in vitro*¹⁶⁵ (Figure 4.4.D). The resulting probe was used for the successful fluorescent labelling of LgtC both recombinantly and in cell lysates and could be applied for the direct imaging of phase variation in clinical isolates. Because of the relative ubiquity of a non-catalytic cysteine in the active site of bacterial GTs, this strategy may in theory be applicable to a wide range

of both retaining and inverting GTs. However, the drug-like and non-substrate related scaffold of the probe may be a drawback for such applications due to selectivity issues. A carbohydrate-based probe, however, may have a great potential towards the above biological applications.

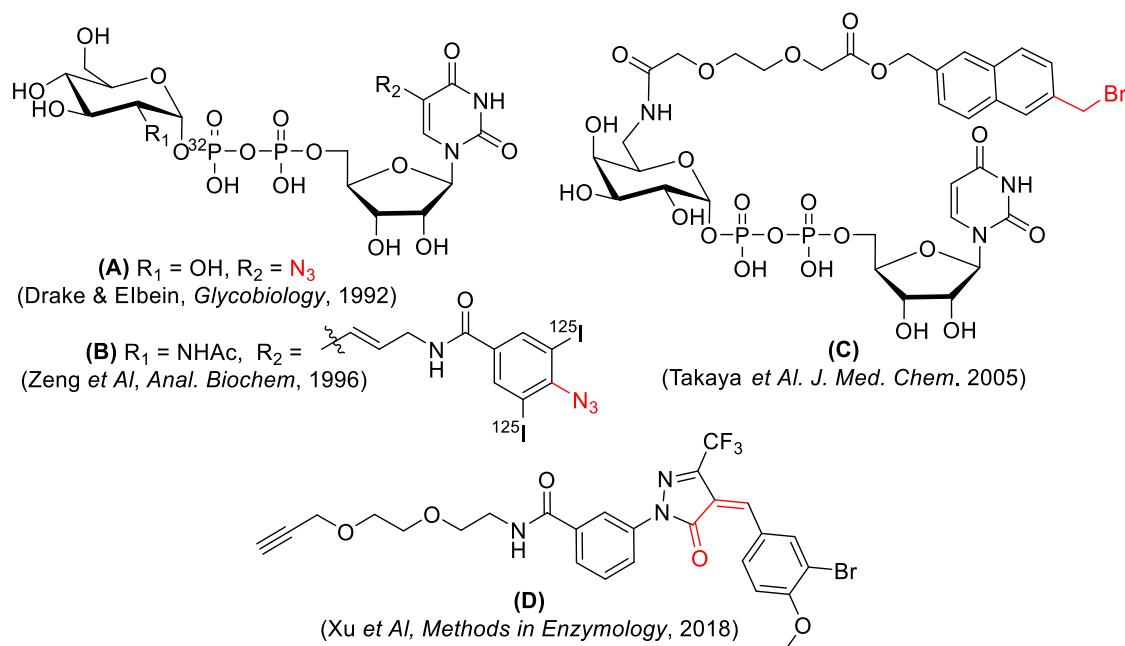
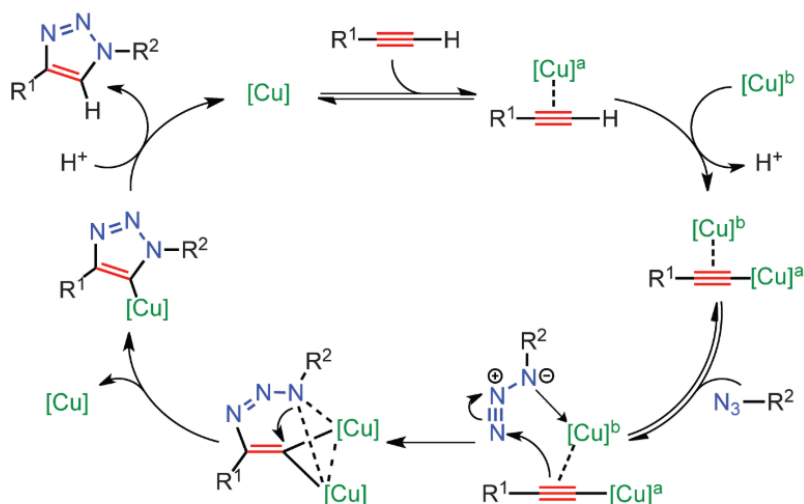


Figure 4.4. AFBPs for GTs

4.1.3. Copper-Catalyzed Azide–Alkyne Cycloaddition (CuAAC)

The CuAAC reaction, also known as click chemistry, is among the recently discovered transformations that have rapidly expanded to a wide range of discipline. Discovered by Sharpless in 2001¹⁸⁰, this reliable, and straightforward way for making covalent connections between building blocks of various functionalities, is now extensively used in organic, medicinal, surface and polymer chemistry as well as bioconjugation applications^{181–183}. The broad application spectrum is due to the great number of practical advantages of the reaction such as its high yields, the breadth of its scope, and the use of an aqueous solvent system as well as the simple purification process¹⁸². It was not until 2013, more than a decade after the discovery of the reaction, that Worrell *et al.*, finally elucidated

the catalytic mechanism of the CuAAC reaction¹⁸⁴. Unlike originally believed, the reaction occurs via a two-Cu-based mechanism: An σ -bound copper acetylide, bearing a π -bound copper, coordinates the azide for the formation of an unusual six-membered copper metallacycle. Ring contraction to a triazolyl-copper derivative is followed by protonolysis, generating the triazole product. The active Cu(I) catalyst is typically generated *in situ* from Cu(II) salts using sodium ascorbate as the reducing agent¹⁸⁴ (Scheme 4.2).



Scheme 4.2¹⁸⁴. Catalytic mechanism of the CuAAC reaction

Over the last decade, CuAAC has been increasingly employed for the development of functional carbohydrate derivatives¹⁸⁵⁻¹⁸⁷. The 1,4-disubstituted triazole, formed in the reaction acts as both a chemically stable linker between a carbohydrate and a functional molecule. In some cases, it can act as a pharmacophore itself, enhancing bioactivity. The triazole-containing carbohydrate molecules exhibit a diverse range of bioactivities such as anti-cancer¹⁸⁸, anti-viral¹⁸⁹ and anti-microbial, *via* the inhibition of disease-related GTs and GHs.

CuAAC has been used for the development of potent 1,6-oligomannoside-based inhibitors for mannosyltransferases involved in the biosynthesis of the cell envelope of *Mycobacterium tuberculosis*¹⁹⁰ (Figure 4.5.A). In a recent study, the activity of triazole-containing semi synthetic tobramycin, an aminoglycoside antibiotic, were found to

suppress the growth of both Gram-positive and Gram-negative bacterial strains, the strongest activity being observed for *P. aeruginosa* (strain I) (MIC = 64 $\mu\text{g}/\text{mL}$)¹⁹¹ (Figure 4.5.B). Interestingly, a triazolyl glycolipid derivative has been found to enhance the drug susceptibility of methicillin-resistant *Staphylococcus aureus* (MRSA)¹⁹² (Figure 4.5.C). The synergistic action of the synthesised triazolyl glycolipid lowered the MIC of a range of commercial β -lactam antibiotics, up to 256-fold on clinical MRSA isolates¹⁹².

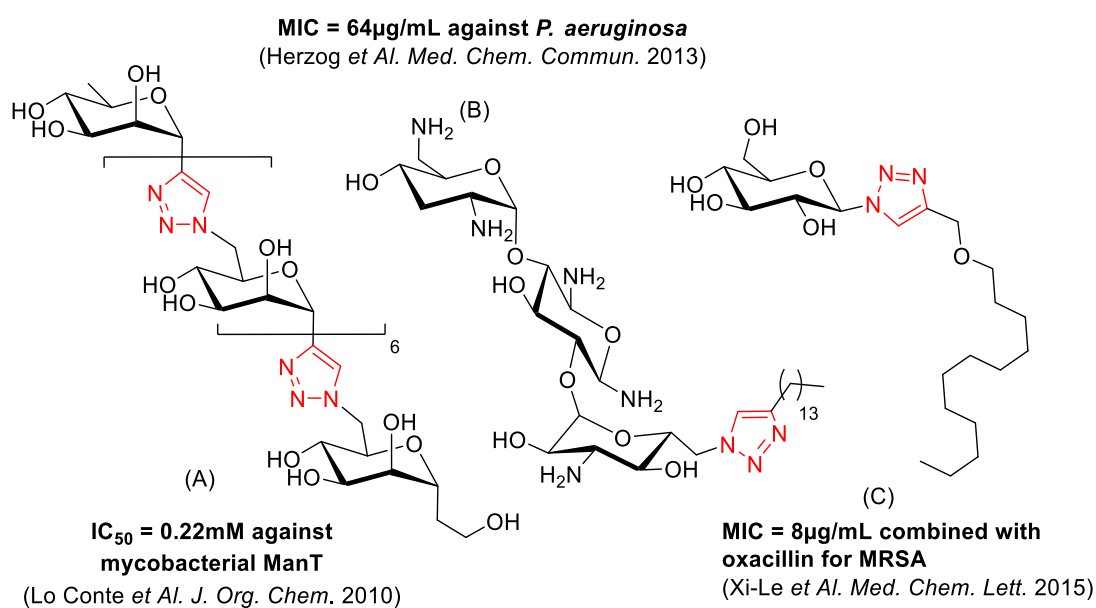


Figure 4.5. Examples of triazole-containing carbohydrate molecules with antimicrobial activity

4.2. Objectives

In the previous chapter, monosaccharides **10** and **11** were designed as covalent inhibitors of LgtC. As part of the study of their mode of action (described in Chapter 3) a fluorescent probe derived from the scaffold of the inhibitors was designed. The synthesis of these probes was completed by the time a surprising non-covalent inhibition mode was concluded for **10** and **11**. In chapter 3, the chemical environment around the WH was shown to have a great impact on its reactivity, Therefore the probe may still bind covalently to LgtC. Another unexpected conclusion drawn from the previous chapter is that the LgtC-binding of **11** remains unclear and kinetics data have suggested that it may

be an allosteric inhibitor. If this is true, then the non-covalent inhibition mode observed for **11** may be simply explained by the absence of nucleophilic residue in the vicinity. Facing all these yet unanswered questions, the probes may provide us with a great opportunity to finally learn more about the behaviour of the monosaccharide inhibitors.

Therefore, the fluorescent probe was synthesised towards two specific aims:

- providing further evidence regarding the behaviour of **10** and **11** towards LgtC.
- developing a carbohydrate-based AfBP for the labelling of purified and in cell lysate LgtC.

In this chapter, the design, synthesis and characterisation of the probes will be discussed in detail. Their binding mode will also be studied and compared to the binding of **11**. The ability of the probes to label recombinant LgtC under different conditions (concentration, pH, reducing environment) will then be tested. Finally, the labelling of LgtC in both a complex cell lysate and through cell membrane will be assessed and discussed.

4.3. Probe design

The designed probe is formed of three different components (Figure 4.6):

- (1) The monosaccharide scaffold containing: a targeting or recognition motif (glucose group) to recognise LgtC and **an anchoring group** (electrophilic WH) for the irreversible attachment to LgtC.
- (2) A chemical fluorophore motif behaving as a **reporter group** to allow for a visual detection of the labelled protein
- (3) A linker also called **chemical spacer** providing spatial separation between the recognition motif and the reporter group reducing its interference on binding.

The attachment of the fluorophore reporter group to the targeting motif will be performed via CuAAC ligation.

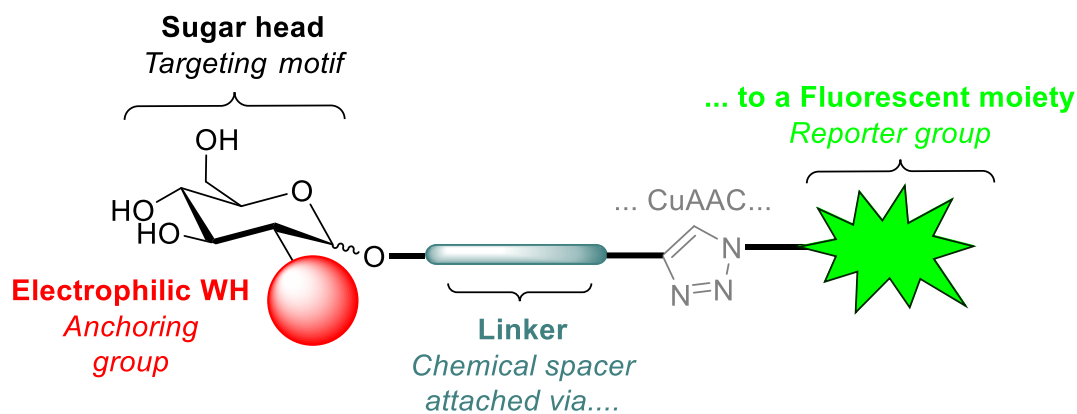
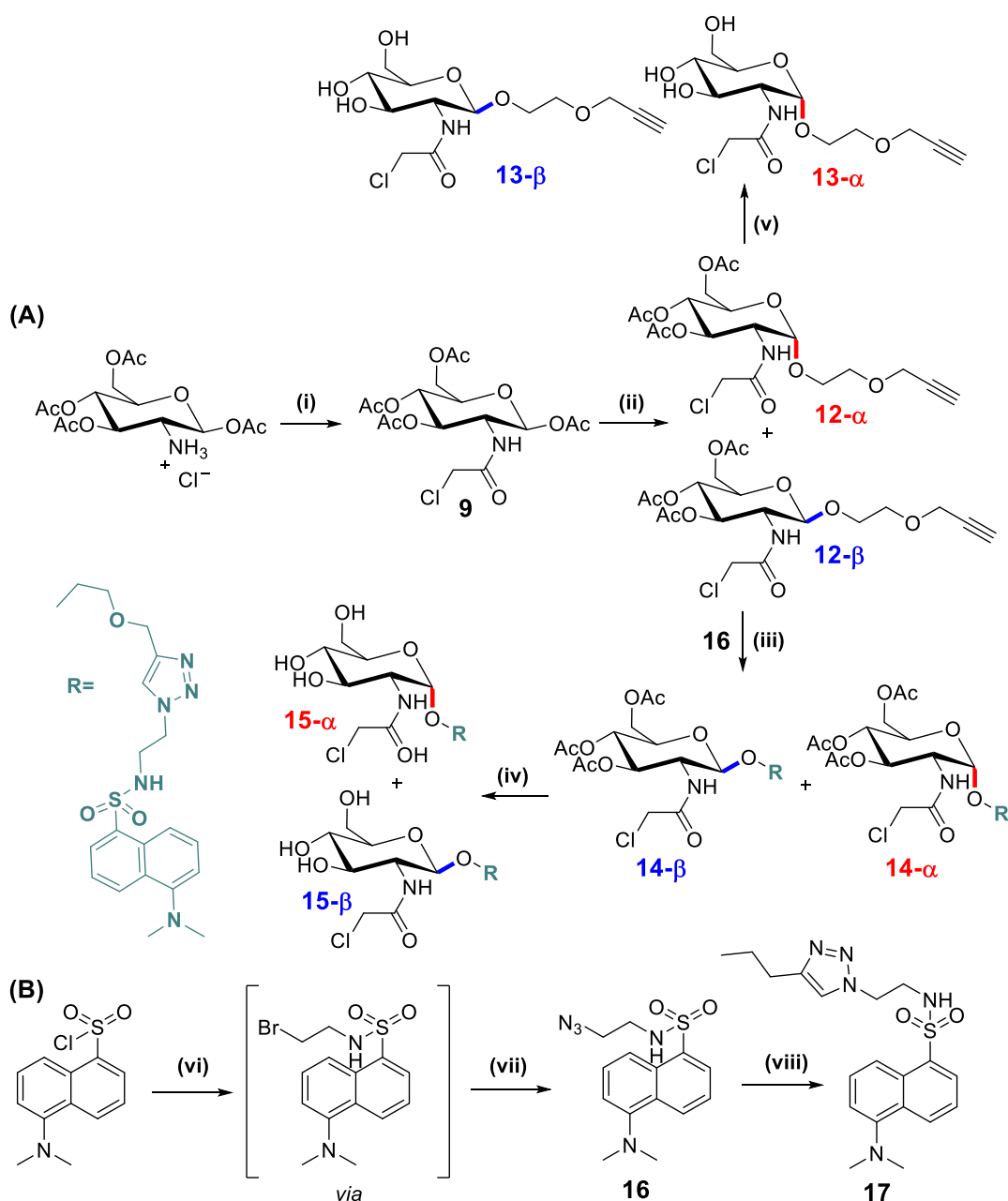


Figure 4.6. Probe design

4.4. Chemical synthesis of fluorescent probes

Derived from the structure of **11**, the most potent inhibitor, probe **15** (chloroacetamide WH) was successfully synthesised in four steps from the commercially available glucosamine hydrochloride salt (Scheme 4.3). The first step is the attachment of the electrophilic warhead for the formation of compound **9** as described in chapter 3. The second step is a glycosylation reaction for the attachment of an alkyne containing linker onto the sugar moiety. Propynol ethoxylate, a commercially available PEG-based linker, was chosen to provide spatial separation between the sugar motif and the fluorophore. However, its purity was not satisfactory (~70%), and the glycosylation reaction did not proceed smoothly when the linker was used as received. It was therefore re-purified in house by distillation to yield a high purity compound (>95%) which resulted in the successful glycosylation of **9** (see Appendix 4 Fig A4.1).



Scheme 4.3. (A) Synthesis of fluorescent probe **15 α/β** . *Reagents & Conditions:* (i) Chloroacetyl chloride (3 eq), TEA (2 eq), DCM, rt, 2 hours, 80% (ii) propynol ethoxylate (3 eq), $\text{BF}_3\text{Et}_2\text{O}$ (4 eq), DCM, 40°C, 8 hours, 72% (iii) **16** (1 eq), $\text{CuSO}_4\cdot 5\text{H}_2\text{O}$ (1 eq), sodium ascorbate (1.5 eq), DIPEA (3 eq), THF : water 10:1, rt, 3 hours, 72% (iv) MeONa (3 eq), MeOH, 10 mins, 0°C, 63%, (v) MeONa (3 eq), MeOH, 10 mins, 0°C, 55%, (B) Synthesis of dansyl azide fluorophore **16**. *Reagents & Conditions:* (vi) 2-bromoethylamine hydrobromide (1 eq), Et_3N (2 eq), DCM, rt, 4 hours, (vii) NaN_3 (2.5 eq), MeCN, reflux, overnight, 87%. (viii) pentyne (1.5 eq), $\text{CuSO}_4\cdot 5\text{H}_2\text{O}$ (1 eq), sodium ascorbate (1.5 eq), DIPEA (3 eq), THF : water 10:1, rt, 3 hours, 85%.

The glycosylation reaction produced the α and β anomers of **12** which are separable by normal phase chromatography. The alkyne signals are detected by ^1H NMR (CDCl_3) at 2.48 and 2.45 ppm for the α and β anomers respectively. The reaction was monitored by TLC and HPLC (Figure 4.7) and the identification of the reaction mixture components at

various time points was analysed against pure samples of **12- α** and **12- β** (Figure 4.7. B). It was found that **9** is very quickly converted to **12- β** (no remaining of **9** after 20 mins) and over time the formation of **12- α** was observed.

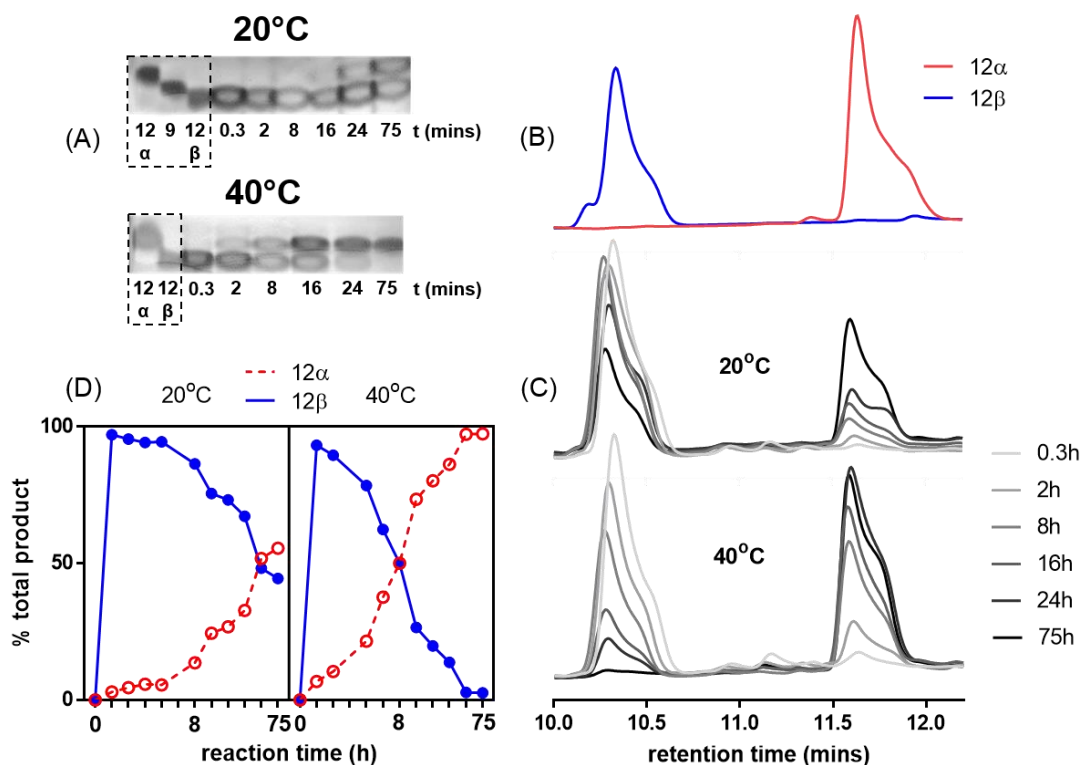
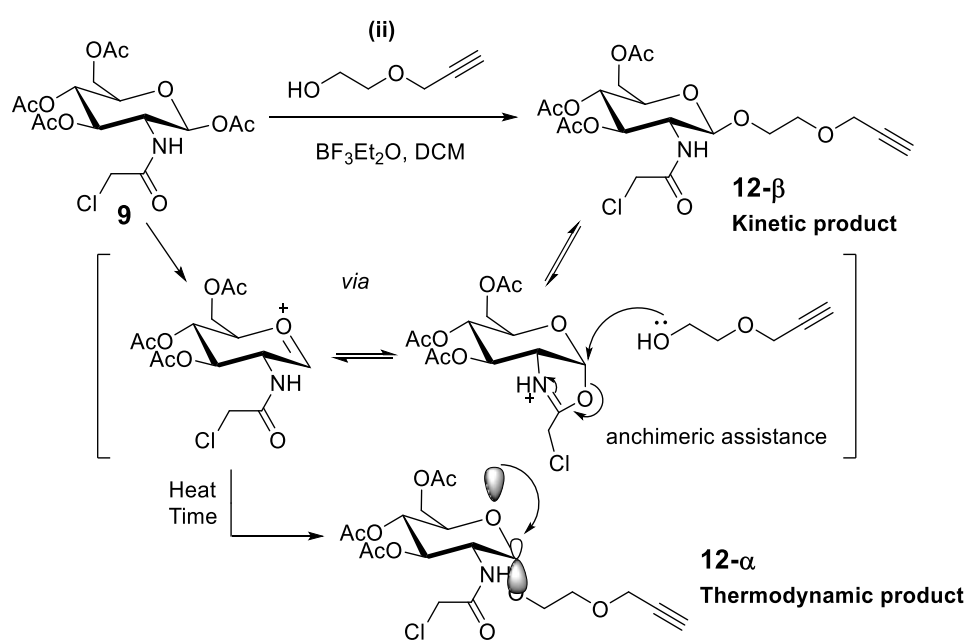


Figure 4.7. Monitoring of the attachment of the linker on **cmpd 9** (step (ii), Scheme 4.3) (A) Monitoring of glycosylation reaction of **9** by TLC at 20°C (top) and 40°C (bottom) *Conditions:* pre-coated aluminium plates (silica gel 60 F254, Merck), solvent system 20% acetone in toluene. (B) HPLC traces of purified **12- α** and **12- β** samples. (C) Monitoring of glycosylation reaction of **9** by HPLC at 20 °C (top) and 40 °C (bottom). (D) Percent of **12- α** and **12- β** of total product present over time at 20 °C (left) and 40°C (right) *Conditions:* SB300-C18 column, injection volume: 10 μ L, gradient: 10-70% MeOH (0.1% TFA) in H₂O (0.1% TFA) over 10 min, flow rate: of 1 mL/min, detecting at 210 nm.

The formation of **12- α** is highly dependent on the temperature of the reaction mixture. While the levels of **12- α** are not detectable until 8 hours at 20°C, the presence of the α anomer is observed after less than 2 hours at 40°C (Figure 4.7 A and C). At this stage, both anomers are of interest as both probes will need to be tested for the labelling of LgtC. Therefore, the reaction was stopped when the reaction mixture contained a 1:1 ratio of α to β anomer. This is the case after 8 hours of stirring at 40°C while it took up to 70 hours

at room temperature (Figure 4.7.D). No product degradation was observed upon refluxing the reaction mixture in DCM, therefore the glycosylation of **9** was performed at 40°C.

A proposed reaction mechanism for the glycosylation of **9** under the described conditions is shown in Scheme 4.4. The anomeric position is activated using boron trifluoride etherate. This leads to the generation of the oxocarbenium intermediate which can be stabilised by anchimeric assistance (also known as neighbouring group participation). The fixed stereochemistry at the neighbouring carbon forces the linker to attack the anomeric position in an equatorial orientation leading to the fast and preferred formation of **12-β**: the kinetic product of the reaction (Scheme 4.4). The reaction being reversible in acidic conditions, the thermodynamic and most stable compound **12-α** will be formed over time and its formation is accelerated with heat.



Scheme 4.4. Proposed mechanism for the glycosylation reaction of **9** (step (ii))

12-α and **12-β** were individually either deacetylated to form the partial probe **13-α** and **13-β** or reacted in a 1,3-dipolar cycloaddition with the synthesised azide fluorophore **16** to form the acetylated full probes **14-α** and **14-β** (triazole proton signal detected at 7.60 and 7.66 ppm on ¹H NMR spectrum of **14-α** and **14-β** respectively, see Appendix 1 Fig

A1.26 & 28). Dansyl azide was chosen for its high quantum yield, and well-known strong fluorescence and long emission wavelength¹⁹³. After deacetylation of **14- α** and **14- β** , **15- α** and **15- β** were purified by preparative HPLC and obtained with high purity (>99% by HPLC and NMR) (Figure 4.8).

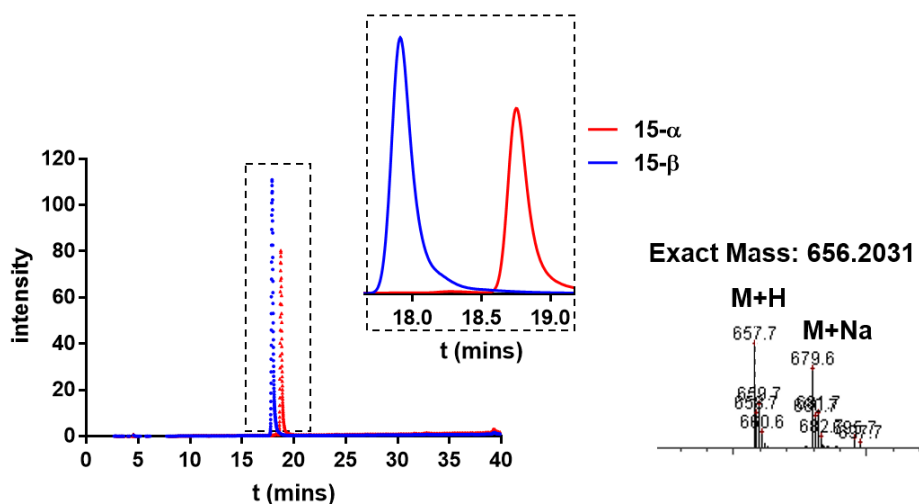


Figure 4.8. Separate overlaid traces of purified **15- α** and **15- β** probes (analysed by LCMS after purification *via* preparative HPLC). *Conditions:* XDB-C8 column, injection volume: 10 μ L, gradient: 10-30% MeOH (0.1% TFA) in H₂O (0.1% TFA) over 40 min, flow rate: of 1mL/min, detecting at 355 nm.

17 was also synthesised as a control molecule to later assess the potential nonspecific binding of the fluorophore alone to LgtC (Scheme 4.3).

Similarly, the synthesis of the corresponding probe with a Michael acceptor group in place was attempted but has unfortunately been unsuccessful to date. The glycosylation of **8** (acetylated precursor of inhibitor **10**, see Chapter 3) proceeded on a small scale (50mg) for the formation of the α - and β glycosylated anomers which were separated by normal phase chromatography and obtained separately with a relatively low yield (21% and 39% for α - and β respectively) and purity (see Appendix 2 Fig A2.2). When attempted on a large scale with the aim of increasing the product purity, the glycosylation reaction did not proceed. No product was obtained, and degradation products were visible on the baseline when monitoring the reaction by TLC. To date, no optimisations were attempted, and this will need to be further investigated.

4.5. Structural and spectroscopic characterisation of probes

4.5.1. Characterisation of **13**

Surprisingly, when characterising compound **13**, no alkyne signal was initially observed by ^1H NMR (Figure 4.9. A) and a very weak signal with the expected chemical shift was just detectable on the standard ^1H decoupled ^{13}C NMR spectra (Figure 4.9.B). The alkyne motif is essential if **13** was to be used for a two-step LgtC-labelling with *in situ* click ligation. The alkyne motif was present in the acetylated precursor **12**, and under the deacetylation conditions used, it is highly unlikely to have reacted. Due to its free sugar motif, NMR analysis of **13** was performed in D_2O . By zooming in the region 75-78 ppm of the ^1H decoupled ^{13}C NMR a triplet splitting pattern can be observed for C_{13} and C_{12} (Figure 4.9.B, red spectrum). This suggests that $^1\text{H}_{13}$ has been exchanged for a ^2H (terminal alkyne protons are known to be slightly acidic)¹⁹⁴ and therefore became ^1H NMR silent while also splitting C_{13} (and C_{12}) in ^{13}C NMR. Unsurprisingly, the splitting is smaller for C_{12} as it is not directly attached to $^2\text{H}_{13}$ (Figure 4.9. B red spectrum). A new ^{13}C NMR with ^2H decoupling was generated, and the disappearance of the triplet (towards a singlet) for C_{13} under these conditions suggests that $^1\text{H}_{13}$ has indeed been exchanged in D_2O (Figure 4.9. B, green spectrum). To provide further confirmation of this, NMR spectra of **13** in d_6 -DMSO were also generated and all the expected signals were observed (see Appendix 1 Fig A1.23). This provides enough evidence to confidently confirm the presence of the alkyne group in molecule **13**. This has been shown here for the α -anomer of compound **13** and the same observations were made for the β -anomer.

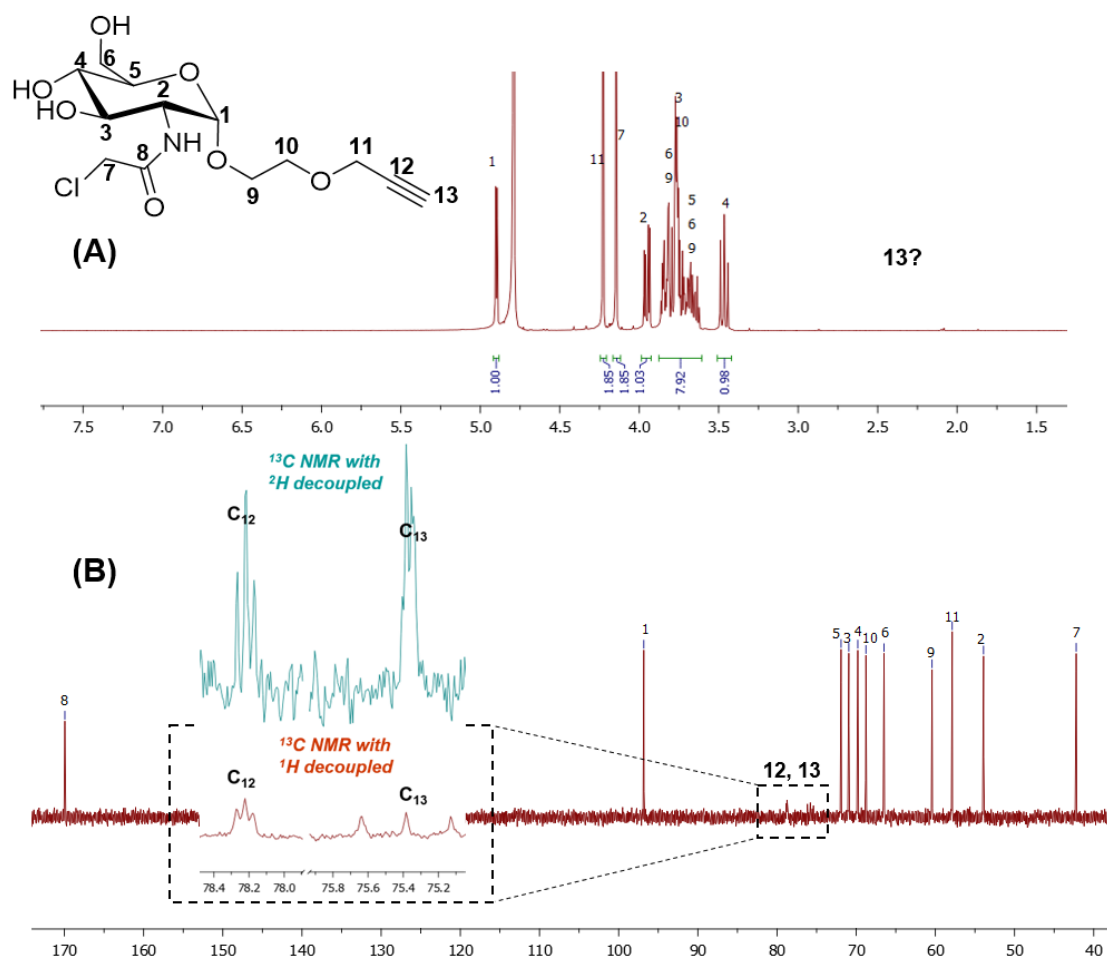


Figure 4.9. Characterisation of compd **13-α** (A) ¹H NMR of **13-α** in D₂O. Absence of alkyne signal. (B) *red spectrum*: ¹³C NMR with ¹H decoupling of **13-α** in D₂O. *Green spectrum*: Zoom in 75-78 ppm region of ¹³C NMR with ²H decoupling of **13-α** in D₂O.

4.5.2. Identification of α - β anomers

Compound **14** is used as a model to describe how anomers are identified (Figure 4.10). As described in chapter 2, for glucose derivatives, a large coupling constant between H₁ and H₂ is characteristic of the β -anomer ($J_{1-2} = 8.4\text{Hz}$) while a small coupling constant is associated with the α -anomer ($J_{1-2} = 3.6\text{Hz}$). Another identification method is possible by analysing the ROESY spectrum of each anomer, a strategy which was not achievable with previous compounds due to the complexity of their ¹H NMR spectrum (which made the identification of individual proton signals impractical). In glucosamine, a 1,3 diaxial interaction exists between H₃ and H₅ due to the structure of the molecule (Figure 4.10). In

the β -anomer, H_1 is in axial position, therefore in addition to the H_3 - H_5 control interaction an H_1 - H_3 as well as an H_1 - H_5 correlation is observed through space on the ROESY spectrum (Figure 4.10 A).

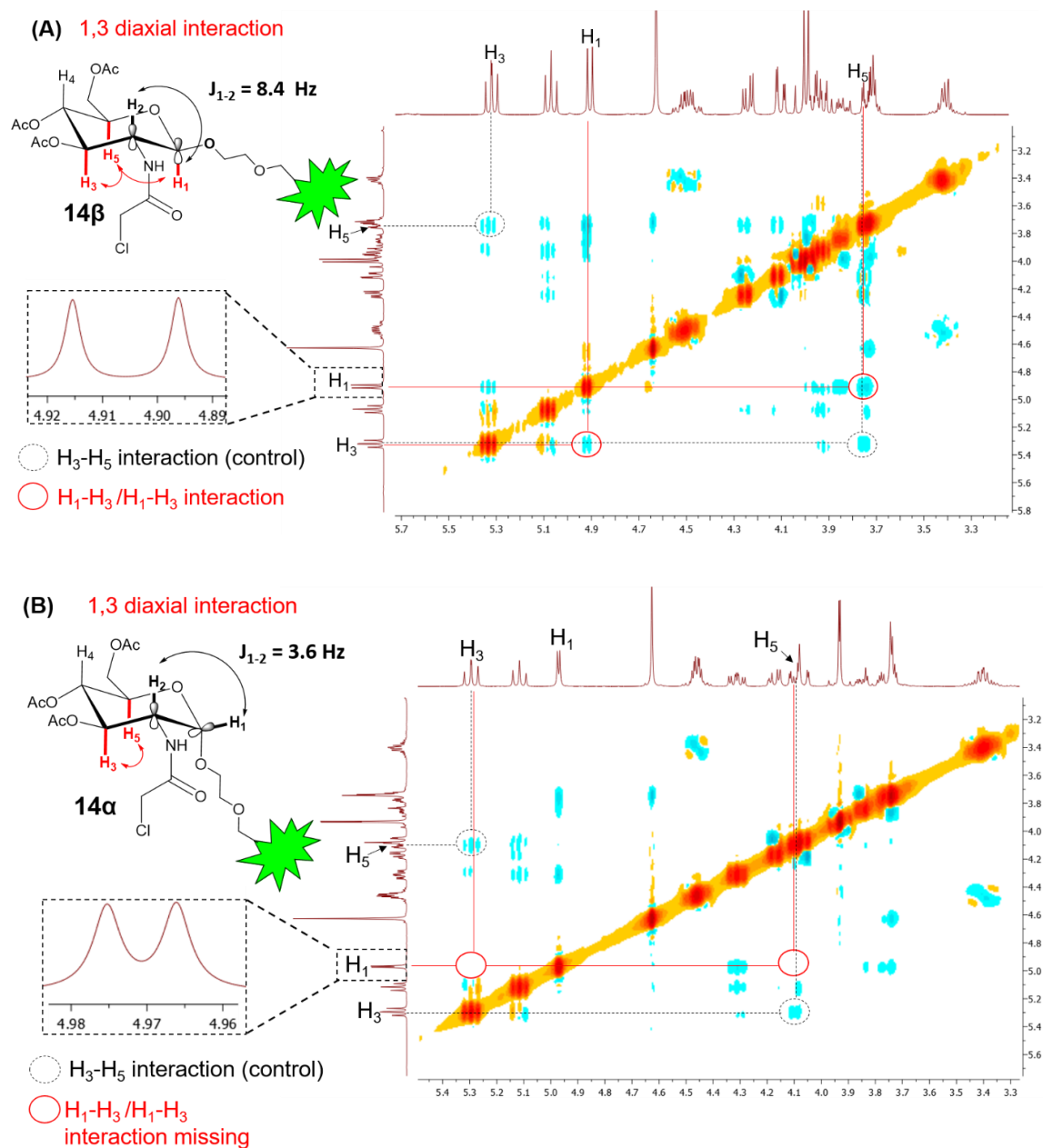


Figure 4.10. Identification of compound **14** anomers by ^1H NMR and 2D ROESY NMR, in CDCl_3 . (A) **14- β** (B) **14- α** .

However, for the α -anomer, H_1 is in equatorial position and therefore no additional 1,3 diaxial interactions is present (Figure 4.10 B). The same analysis was done for the identification of α - and β -anomer of compound **12**. For the deacetylated compounds **13**

and **15**, this identification technique is more challenging due to the proton signals of free sugar motifs being superimposable.

4.5.3. UV-Fluorescence spectroscopy

Having fully characterised the probes structurally, their fluorescence properties were investigated. This was tested in the same buffer as the subsequent LgtC labelling assays (13 mM HEPES buffer pH7). The effect of the sugar on the fluorescence of the Dansyl motif also needed to be assessed. However, due to the insolubility of the dansyl fluorophore in aqueous solvent systems, the fluorescence properties of the probe were also tested in DMSO to allow for a direct comparison with the fluorophore alone. Firstly, UV-vis absorbance spectra were obtained in order to identify the λ_{max} values later used as excitation wavelengths for the emission spectra of the probes. It was found that the λ_{max} value varied depending on the solvent used ($\lambda_{\text{max}} = 326\text{nm}$ and 341nm in HEPES buffer and in DMSO respectively) (Figure 4.11.A, B). These values were therefore used as excitation wavelengths to generate the emission spectra of the probes in both solvents (400 - 675nm emission range). In DMSO, **15- α** , **15- β** and **16**, all exhibit the same fluorescence properties with a maximum emission wavelength of 524nm (Figure 4.11.C). This indicates that the presence of the sugar group does not have any effect on the fluorescence emitted by the dansyl fluorophore. However, in HEPES buffer, the fluorescence of both probes is drastically reduced, indicating a quenching of the fluorescence in aqueous environment. A 10x more concentrated aqueous probe solution (100 μM) gave a maximum emission of half the intensity observed for DMSO (10 μM) (intensity 800 vs 400) (Figure 4.11C, D). This could either be a solvent effect or may be due to the self-assembly of the probe in water which would lead of self-quenching. **15- β** seems to be slightly more affected than **15- α** although the difference is minimal. Additionally, a shift in the maximum emission wavelength is observed compared to the DMSO sample ($\lambda_{\text{em}} = 557\text{nm}$ in HEPES buffer compared to 524nm in DMSO) (Figure 4.11.C,

D). The additional peak present at $\lambda=652\text{nm}$ in Figure 4.11.D is not characteristic of the sample but corresponds to the second harmonic generation signal (SHG) (photon with twice the initial energy, $\lambda_{\text{ex}}=326\text{nm}$). In DMSO this is not observed, as the SHG signal would show at 682nm , which is outside the measured wavelength range ($400\text{-}675\text{nm}$).

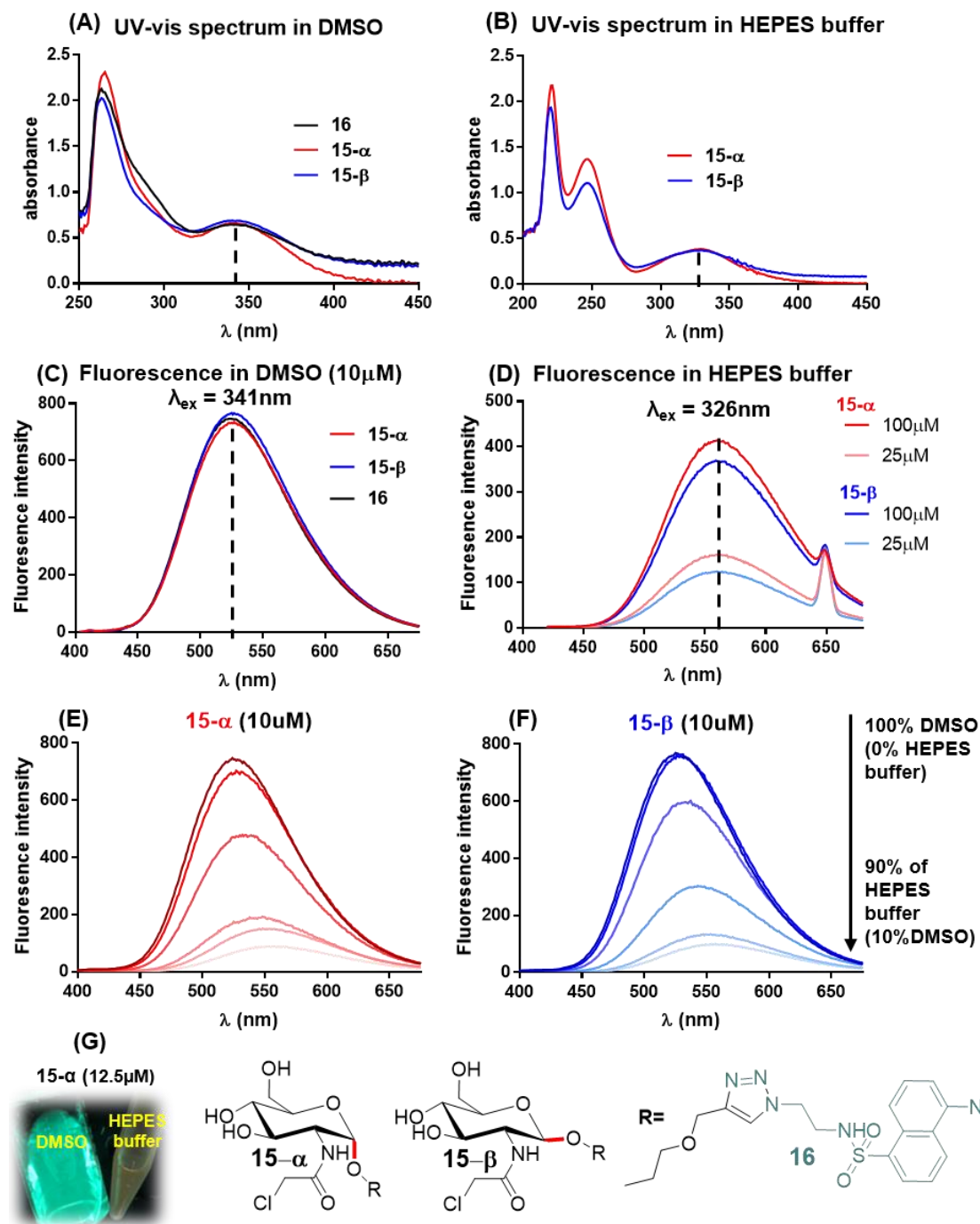


Figure 4.11. Fluorescence / UV-vis spectra of **15**. (A) UV-vis spectrum in DMSO ($10\mu\text{M}$). (B) UV-vis spectrum in HEPES buffer (13mM , $\text{pH}=7$). (C) Emission spectrum in DMSO ($\lambda_{\text{ex}}=341\text{nm}$). (D) Emission spectrum in HEPES buffer ($\lambda_{\text{ex}}=326\text{nm}$). (E) HEPES buffer titration in $10\mu\text{M}$ of $15\text{-}\alpha$ in

DMSO. (F) HEPES buffer titration in 10 μ M of **15- β** in DMSO. (G) **15- α** (10 μ M) in both DMSO and HEPES buffer observed under UV lamp (365nm).

In order to confirm the above observations, HEPES buffer was titrated into a solution of **15** in DMSO. While the concentration of **15** remains constant throughout, a progressive decrease in the emission intensity alongside a shift of the curve is observed for both anomers as the percentage of HEPES buffer increases (Figure 4.11.E, F). This is in line with our previous observations and confirms that the fluorescence of **15** is partially quenched in a protic solvent. Although such behaviour is not uncommon in fluorescent molecules¹⁹⁵, it is important to be aware of this as the probe will be used for biological assays in an aqueous environment. This apparent difference in the probes' fluorescent properties in both protic and aprotic solvents can also be observed by the naked eye when visualising both samples under the UV lamp (365nm) (Figure 4.11.G).

4.6. Labelling of recombinant LgtC

With compounds **15- α** and **15- β** now being fully characterised, they were then tested for their ability to label recombinant LgtC in a one-step labelling experiment. LgtC was incubated with full probe **15**, the protein was then denatured with non-dyed loading buffer (to avoid interference with fluorescence reading), and the mixture was separated on a 12% SDS page gel visualised by fluorescence read-out and Coomassie staining (Figure 4.12). Initially, both **15- α** and **15- β** (0.5mM) were incubated with LgtC (15 μ M) for 30 mins at 30°C (13mM HEPES buffer pH=7). Interestingly, the band corresponding to LgtC (33kDa) was detected during fluorescence read-out when incubated with the probe. The water negative control accounts for intrinsic fluorescence of the enzyme and in this case, none was observed (Figure 4.12). This suggests a genuine labelling of LgtC by the probe. Due to the denaturing conditions under which the experiment was conducted, the observed labelling can only be the consequence of a covalent modification of the enzyme, as any unbound probe is observed at the bottom of each lane (fluorescent blob below 17

kDa protein marker). Additionally, the intensity of the band fluorescence is significantly stronger for **15- α** than for **15- β** (Figure 4.12). This is not believed to be due to the slightly lower fluorescence properties of **15- β** in HEPES buffer showed in Figure 4.11. D, as the difference in band intensity observed on the gel is much more significant than the difference observed in the emission spectra. Also, more unbound probe is observed in the case of **15- β** than **15- α** (Figure 4.12, below 17 kDa protein marker). The above suggests a greater potency for the α -anomer potentially due to a greater binding affinity for the enzyme. This will be discussed in more detail later in this Chapter.

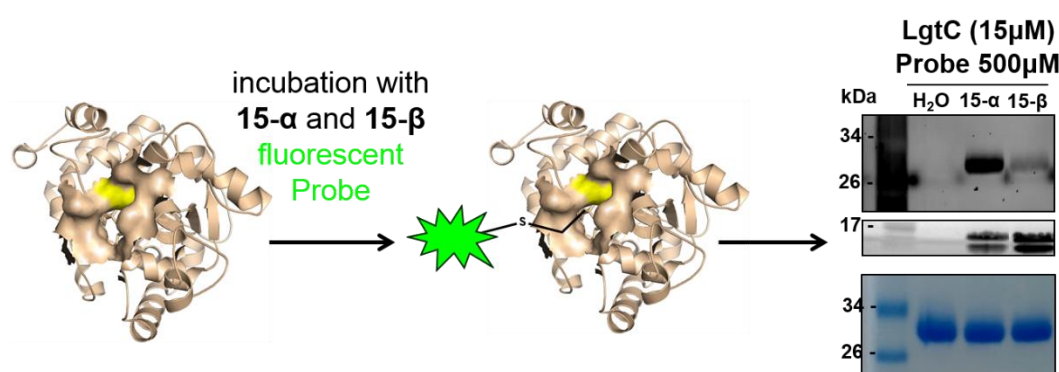


Figure 4.12. Principle for the labelling of LgtC with probe **15**. *General assay conditions:* **15- α/β** (1 μ L of 10X stock) is incubated with LgtC (1 μ L of 10X stock) in HEPES buffer (8 μ L, 13 mM, pH=7, unless stated otherwise) for 30mins (unless stated otherwise) at 30°C. Loading buffer (2.5 μ L) is added, the reaction mixture is incubated at 50°C for 10 mins and the resulting solution is loaded on 12% SDS-page gel and run at 160V for protein separation. The gel is visualised by fluorescence scanning (*top*) and coomassie staining (*bottom*).

Having established that the probe can label LgtC, we then assessed how altering conditions such as concentration, time, chemical environment or pH may affect the observed labelling.

4.6.1. Effect of probe concentration on labelling

The effect of the probe concentration on the labelling was first investigated. Decreasing concentrations of both probes were incubated with LgtC for 30 mins at 30°C. The data indicates that the labelling of **15- α** is concentration dependant. **17**, the fluorophore alone was also tested as a control and exhibited some non-specific labelling (which was non-

concentration dependant) (Figure 4.13 A). By using this non-specific labelling as the background reading and subtracting it to the band volume it can be concluded that the labelling ability of **15- β** was negligible at concentrations of less than 0.5mM (Figure 4.13 B). At 0.5 mM, the labelling of **15- α** is 8-fold greater than for **15- β** . Also, the labelling of LgtC by **15- α** was linear for the tested concentrations (Figure 4.13 C). 0.5mM was chosen as the standard probe concentration for the following experiments.

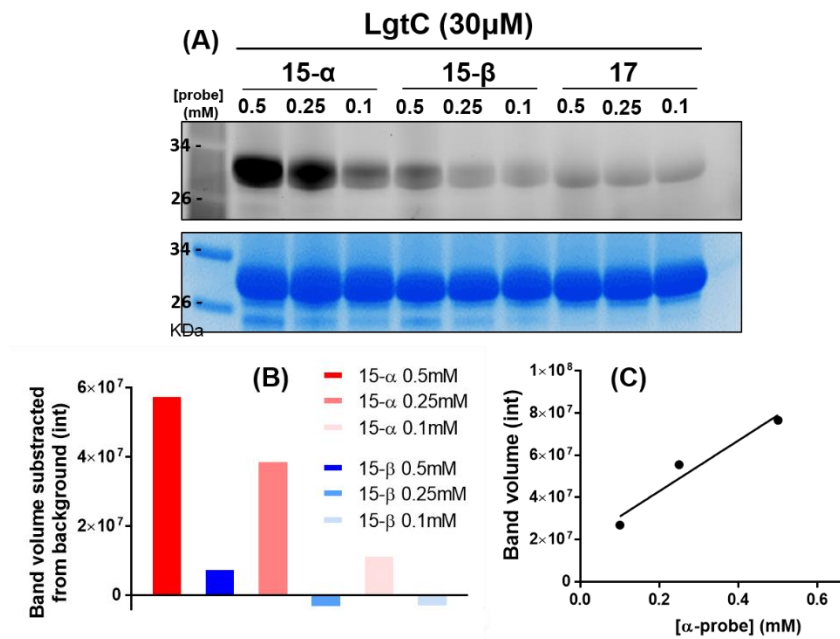


Figure 4.13. Effect of probe concentration on LgtC-labelling performed by probe **15**. (A) See general assay conditions in Fig 4.12. Decreasing probe concentration (1 μ L, 5 – 1mM stocks), LgtC stock concentration = 300 μ M. (B) The gel was visualised in ImageLab, band volume was subtracted from the background reading (**17**) and plotted using GraphPad prism. (C) The corrected band volume for **15- α** was plotted against the probe concentration using GraphPad prism.

4.6.2. Effect of enzyme concentration on labelling

Following this, the effect of the enzyme concentration on the labelling was investigated. Decreasing enzyme concentrations were incubated with a 500 μ M probe solution for 30 mins at 30 $^{\circ}$ C. An enzyme-concentration dependant labelling was observed for both **15- α** and **15- β** . The labelling is linear for the tested LgtC concentrations (Figure 4.14). 7.5 μ M LgtC concentration was the labelling detection limit under these conditions and 15-30 μ M LgtC was then used for the following experiments.

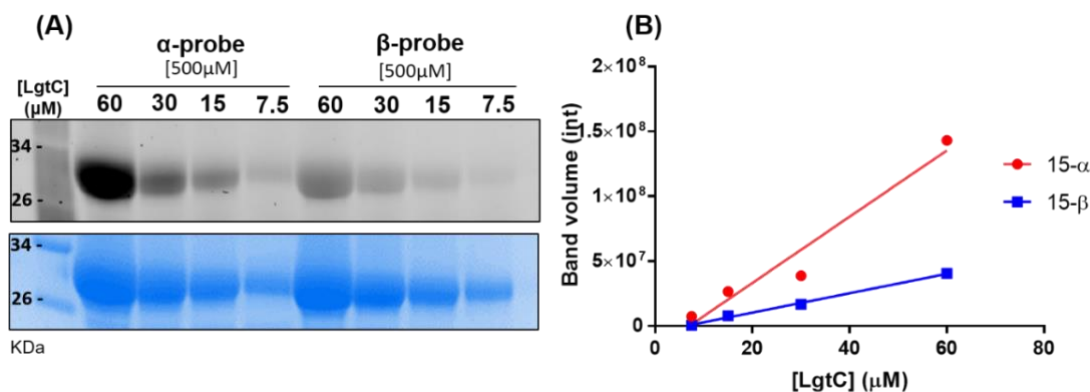


Figure 4.14. Effect of enzyme concentration on LgtC-labelling performed by probe **15** (determination of detection limit). (A) See general assay conditions in Fig 4.12. Probe stock concentration = 5mM. Decreasing LgtC concentration (600μM - 75μM stocks). (B) The band volume retrieved from ImageLab were plotted against the concentration of enzyme using GraphPad prism.

4.6.3. Effect of incubation time on labelling

The incubation time was initially set at 30 mins based on conditions previously used by the group. However, it was important to investigate the time dependency of the labelling. Both shorter and longer incubation times were tested, and the results are shown in Figure 4.15. While a time dependant labelling was observed for **15-α** throughout the whole tested time range (5 mins to 4 hrs) no time dependency was observed for **15-β** until 60 mins and as previously observed the labelling of **15-β** is within the background reading observed for **17**. Time dependent labelling is characteristic of covalent labelling. Due to the irreversibility of the labelling, as more time is allowed for the covalent bond to form, greater labelling is observed.

On the other end, increasing incubation times come alongside an increase in protein degradation, as observed after Coomassie staining the gel (Figure 4.15). This is in agreement with previous observations made in chapter 3: when performing biochemical assays at different pre-incubation times, the enzyme loses most of its activity after 100 mins. As no significant difference was observed between 30 mins 60 mins, 30 mins was kept as the standard incubation time for the rest of the work.

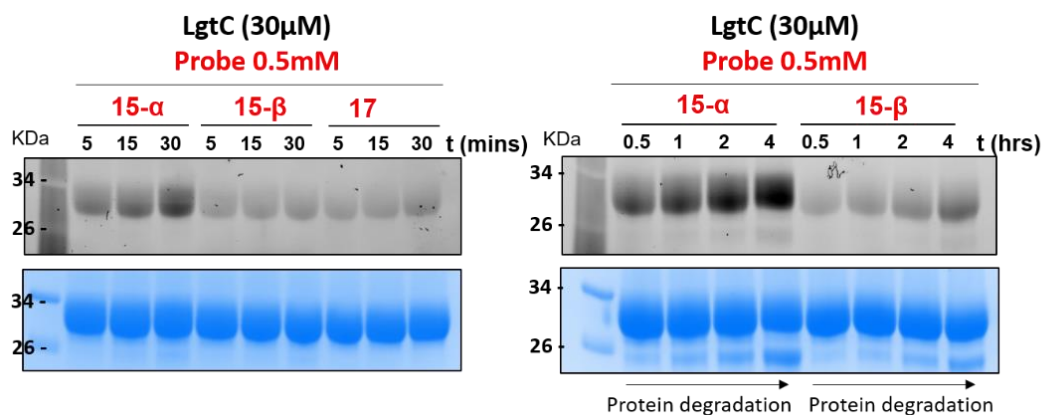


Figure 4.15. Effect of incubation time on LgtC-labelling performed by probe **15**. See general assay conditions in Fig 4.12. Probe stock concentration = 5mM, LgtC stock concentration = 300 μ M. Increasing incubation times (5 - 240 mins).

4.6.4. Effect of reducing environment and labelling specificity

Next, the effect of reducing conditions on the labelling of LgtC was investigated. No significant effect was observed when 0.1mM of either DTT or TCEP was added to the reaction mixture (Figure 4.16.A), suggesting that the target residue may not be subject to oxidation. This was repeated with higher concentrations of reducing agent. A decrease in the labelling of LgtC was observed upon increasing concentration of both DTT and TCEP, most likely due to reaction competition with the probe (Figure 4.16.B). The increase in band intensity observed at 0.8mM of TCEP is believed to be an artefact considering the large excess already present at 0.2mM at which no effect was observed.

A control experiment was also performed during which **15- α** was incubated with denatured enzyme (Figure 4.16.C). The fluorescence was drastically reduced to background level, indicating that **15- α** only has affinity for the native/folded protein. This suggests LgtC specific binding of the probe rather than non-specific covalent attachment to a surface residue.

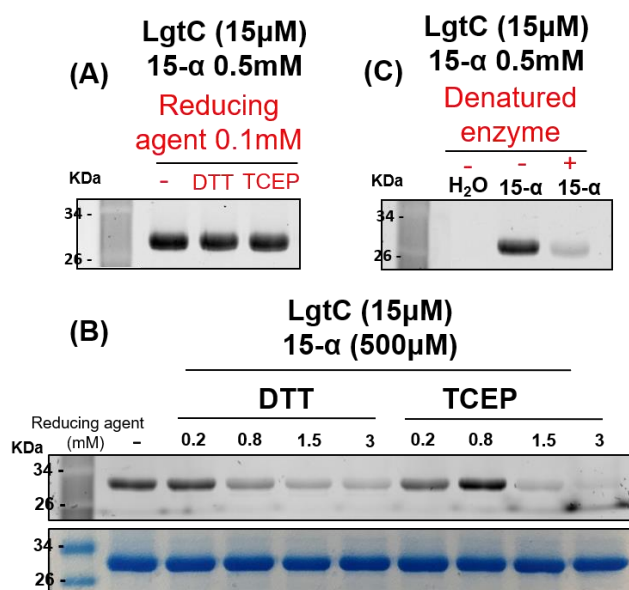


Figure 4.16. (A & B) Effect of reducing environment on LgtC-labelling performed by **15-α**. See general assay conditions in Fig 4.12. **15-α** stock concentration = 5mM, LgtC stock concentration = 150µM, DTT/TCEP (1µL, 1mM stock for (A) and 2 - 30mM stock for (B)), HEPES buffer (7µL, 13 mM, pH=7). (C) Labelling of **15-α** on folded and denatured LgtC. See general assay conditions in Fig 4.12. Probe stock concentration = 5mM, LgtC stock concentration = 150µM, LgtC is treated with loading buffer prior to probe incubation for labelling on denatured enzyme.

4.6.5. Effect of pH on the labelling

Following this, the effect of the pH on the observed labelling was investigated. So far, all experiments were performed in HEPES buffer at pH7. First, a range of pHs from acidic to basic was tested (3, 7, 11). For **15-α**, a significant increase in labelling was observed upon increase in buffer alkalinity (Figure 4.17.A). For **15-β**, while no difference was observed between pH3 and 7, a large increase in labelling was observed at pH11 (Figure 4.17.A). Subsequently, the labelling of LgtC at basic pH was assessed. For both probes, the labelling intensity remained similar between pH 9 and 11 while at pH 13, it became nonspecific as intense labelling was also observed for control (**17**) (Figure 4.17.B). Unsurprisingly, the above results suggest an important increase in labelling intensity at basic pH, explained most likely by deprotonation and activation of the nucleophilic target residue. This observation is therefore consistent with a covalent enzyme modification. The significant increase in labelling intensity between pH7 and 9 indicate that the pKa of the target residue may lie in that pH range. This may be consistent with a cysteine residue, however,

in the protein the pKa of acidic and basic residues can change significantly depending on the environment around it.

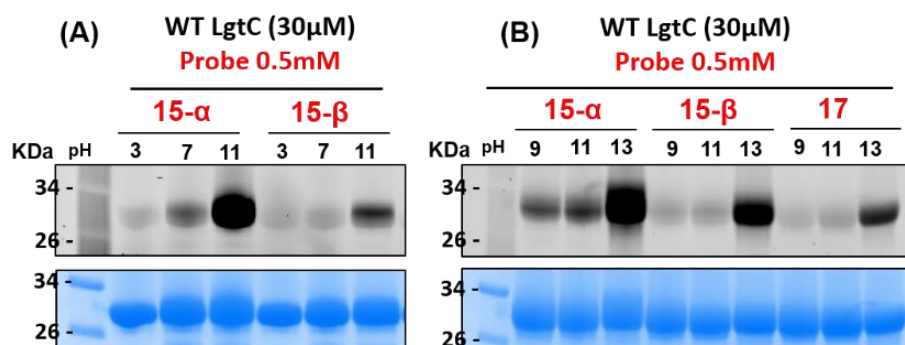


Figure 4.17. Effect of pH on LgtC-labelling performed by probe **15**. See general assay conditions in Fig 4.12. probe stock concentration = 5mM, LgtC stock concentration = 300 μM, HEPES Buffer (8 μL, 13mM, various pH).

Combined, all results obtained so far strongly suggest that the probe covalently binds to LgtC. The observed labelling is concentration and time dependant. While it is strongly affected by pH, the probe does not require the presence of a reducing agent for the successful labelling of LgtC. Importantly, we have shown that the labelling was specific for the native and folded enzyme as the probe could not label denatured LgtC.

The fact that probe **15-α** can covalently label LgtC is intriguing as the sugar alone (compound **11**, Chapter 3) was a non-covalent inhibitor of LgtC. This could be explained by two factors: (1) the probe is more reactive towards nucleophilic residue than inhibitor **11**, or (2) probe and inhibitor do not have the same LgtC-binding mode. To hopefully gain a better understanding of the binding mode of the candidates, competitive labelling experiments were performed.

4.6.6. Competitive labelling

To gain insight into the binding mode of the candidates, the competitiveness of the observed LgtC-labelling in the presence of **11** was investigated. **15-α** (0.5mM) was incubated with LgtC (15 μM) in the presence of increasing concentrations of **11** (0 to 5 mM). This was

conducted with and without **11**-enzyme pre-incubation (45mins) (Figure 4.18.A). As concluded in chapter 3, **11** does not bind covalently to LgtC therefore the pre incubation was not expected to impact on the outcome of this experiment. Indeed, no difference in the labelling pattern was observed between the pre-treated enzyme and the simultaneous component addition (Figure 4.18.A). More interestingly, upon increasing concentration of **11**, no reduction in the labelling of LgtC by **15- α** was observed (Figure 4.18.A). This indicates that the binding of **11** does not affect the binding of **15- α** . This non-competitiveness would be expected if **15** and **11** did not occupy the same binding site in the enzyme. As the probe was designed to reflect the behaviour of the inhibitor this was unexpected.

Therefore, we tested the competition of the observed labelling for LgtC substrates. When **15- α** (0.5mM) was incubated with LgtC (15 μ M) in the presence of an increasing concentration of UDP-Gal donor (0 to 0.5mM), an apparent decrease in LgtC-labelling was observed (Figure 4.18.B). This indicates that the binding of **15- α** is competitive for the donor substrate and that therefore **15- α** binds in the active site of LgtC. Additionally, the competition of the observed labelling was also tested for lactose, the partner substrate of LgtC. Initially, increasing lactose concentration had no significant effect on the observed labelling of LgtC (Figure 4.18.C). However, the superimposition of UDP-Gal and **15- α** (Figure 4.19.A) shows that **15- α** is too large to occupy the buried donor binding site alone therefore it must also occupy the surface exposed lactose binding pocket. As previously mentioned, it is known that the donor substrate induces a conformation change in the enzyme which allows the binding pocket of LgtC to form¹⁰⁴ (see 1.7.3.1). Therefore, the result shown in Figure 4.18.C may be unreliable as lactose may not bind to the enzyme under these conditions. For this investigation to be accurate and the acceptor binding pocket to form, the donor substrate must be present. However, if both substrates are incubated with LgtC then the enzymatic reaction takes place, and this will not allow for accurate and reliable observations. For this reason, UDP as a donor mimic was used, hoping it would trigger the conformation change in LgtC which allows lactose to bind. The

experiment was then repeated with a fixed concentration of UDP present (100 μ M) and increasing concentration of Lactose. Under these conditions, a significant reduction in the LgtC-labelling of **15- α** was observed (Figure 4.18.D). This suggests that **15- α** binds to the acceptor binding pocket of LgtC.

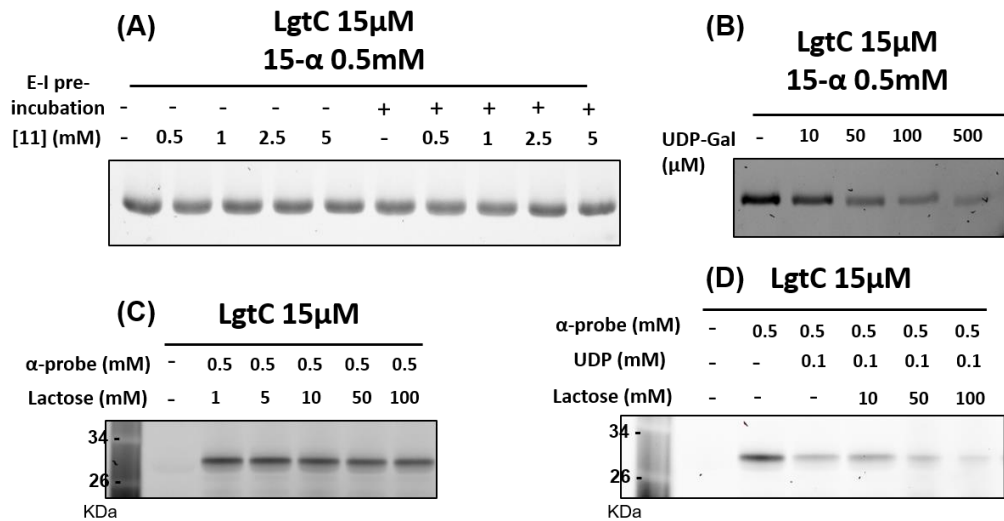


Figure 4.18. Assessment of inhibitor / substrate competition on LgtC-labelling performed by **15- α** . See general assay conditions in Fig 4.12. **15- α** stock concentration = 5mM, LgtC stock concentration = 150 μ M, HEPES Buffer (8 μ L, 13mM, pH=7) (A) Monosaccharide inhibitor **11** competition: **11** (1 μ L, 5 – 50mM stocks), performed with- or without 45 mins **11**-LgtC pre-incubation (B) UDP-Gal donor competition: UDP-Gal (1 μ L, 0.1 – 5mM stocks), (C) Lactose acceptor competition: Lactose (10 – 1000mM stocks) (D) Lactose acceptor competition in presence of donor mimic: UDP (1 μ L, 1mM stock), Lactose (100 – 1000mM stocks).

The results obtained in this section strongly suggest that **15- α** binds in the active site of LgtC. This fits with the docking simulations which predict a binding occupying both active sites and extending further on the surface of the protein with (Figure 4.19.B):

- the fluorophore in the donor binding site
- the lactose binding pocket accommodating the linker and a portion of the sugar head
- the rest of the sugar head extending outside the binding site (on the protein surface)

However, this indicates that the fluorophore plays an important role in protein recognition and that the binding is not, as designed, targeted by the sugar itself. Additionally, this proposed binding mode for the probe is in line with previous

observations regarding the absence of labelling on denatured enzyme (Figure 4.16) which indicate binding in the active site of the enzyme. The suggested binding mode can also rationalise why **15- α** is more potent than **15- β** . The stereochemistry at the anomeric position directs the orientation of the rest of the molecule. Both UDP-Gal and **15- α** have axial anomeric orientation, which suggest that **15- α** is a better mimic of UDP-Gal than its analogue. The linker, triazole and fluorophore groups attached to the sugar motif adopt the same orientation as the diphosphate and uridine motifs of the donor substrate (Figure 4.19.C) making **15- α** bind with more affinity into the binding site of UDP-Gal.

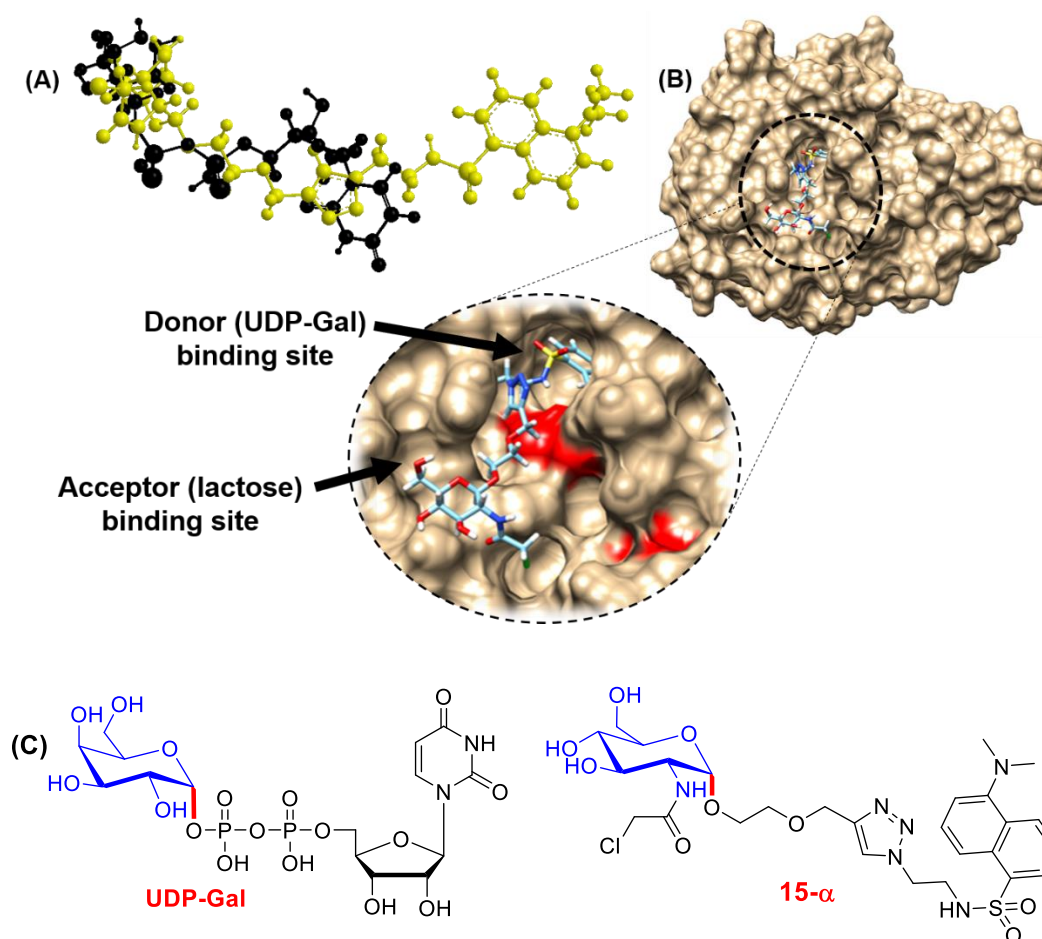


Figure 4.19. (A) Superimposition of UDP-Gal (black) and **15- α** (yellow) in chemdraw 3D (B) Docking simulation of probe **15- α** in LgtC, cys246 is shown in red. (C) Structure of **15- α** and UDP-Gal side by side. **15- α** is a good mimic of UDP-gal, both have a α -orientation at the anomeric position.

In addition to this, the data suggest that the binding mode of probe **15- α** and monosaccharide **11** are different, as in Chapter 3 the binding of **11** was not competitive

for either substrates. This indicates that probe **15- α** does not reflect the behaviour of **11** towards LgtC, as it was designed to. This fits with the absence of binding competitiveness observed between **15- α** and **11** and provides further evidence towards a hypothetical allosteric inhibition mode for monosaccharide **11**.

4.6.7. Two-step vs one step labelling

In order to gain further understanding of the role of the sugar motif in labelling by **15- α** , a two-step labelling protocol was performed. This involves the incubation of the partial probe **13- α** with LgtC (1 hour) followed by a biorthogonal click ligation *in situ* for the attachment of the fluorophore reporter group (1 hour) (Figure 4.20). This removes any potential influence of the fluorophore on the observed labelling which allows for the investigation into the role of the sugar targeting group independently.

The presence of DTT and related thiols may drastically reduce the rate of the Click reaction, therefore, for this experiment, LgtC was not activated with DTT prior to incubation with the probe (the activity of LgtC in the absence/presence of DTT has been compared and no significant differences were observed). This protocol is more challenging and time consuming than the standard 1-step protocol with the full probe. The non-bound partial probe is washed off before the *in situ* click reaction. Similarly, any excess of **16** and Cu is removed after the ligation and a significant amount of protein precipitates following this protocol. This leads to a significant amount of protein degradation as seen in Figure 4.20. To clearly compare the *two-step* and the *one-step* protocol, the samples were run on the same gel. However, for the same concentration of **13- α /15- α** (0.5mM) the labelling observed for the full probe in the one step protocol was much greater than the partial probe through the two-step labelling protocol. This resulted in a very poor-quality imaging and contrast modification still did not allow for a robust interpretation of the data. To address this, the concentration of full probe **15- α** in the one-

step labelling was reduced (0.1mM) while **13- α** was kept at 0.5mM. Under these conditions, and with a tedious image processing, a weak labelling is observed after the incubation of the partial probe **13- α** which indicates that the click ligation was successful. However, a non-negligible labelling is also present for the fluorophore controls. As the fluorophore alone has no mean to covalently modify LgtC, this corresponds to nonspecific background reading, and the labelling of **13- α** is even weaker than initially thought.

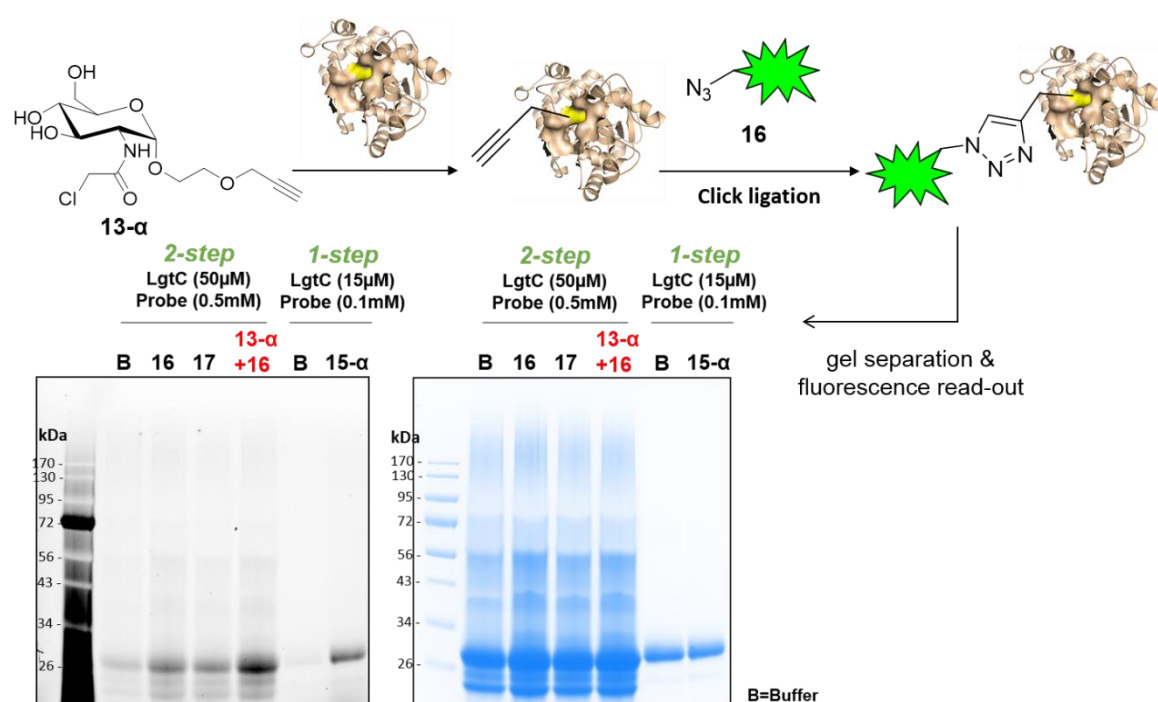


Figure 4.20. Two-step LgtC-labelling performed by **13- α** : workflow and results. *General assay conditions:* **13- α** / **17** / HEPES buffer (25 μ L, 10mM stock) is incubated with LgtC (100 μ L, 300 μ M stock) in HEPES buffer (375 μ L, 13 mM, pH=7) for 1 hour at 30°C. The reaction mixture is washed and concentrated, made up to 325 μ L with HEPES buffer and incubated with DMSO / **16** (25 μ L, 10mM stock), CuSO₄ (50 μ L, 2.5mM stock), THPTA (50 μ L, 12.5mM stock) and sodium ascorbate (50 μ L, 50mM stock) for 1 hour 30°C. The reaction mixture is washed and concentrated again. See Fig 4.12 for following sample treatment. Fluorescence scanning (*left*) and coomassie staining (*right*).

The above dataset suggests that the sugar alone has a drastically weaker labelling ability than the full probe. This agrees with previous observations regarding the fluorophore's key role on the binding of the probe. The docking simulations shown in Figure 4.19.B suggest that the fluorophore binds in donor binding pocket while sugar motif binds partially in the acceptor binding site. The work done in Chapter 3 provided evidence that

the sugar alone does not bind in the lactose binding pocket. Therefore, the barely significant labelling observed *via* the two-step protocol is consistent with the rest of the work, and the sugar alone does not behave as the probe's targeting motif. Because of the electrophilic WH attached, the sugar motif does act as the anchoring group, but the fluorophore alone seems to be mainly responsible for the enzyme recognition.

4.7. Labelling of LgtC mutants and attempted identification of target residue

Previous observations indicated that the covalent labelling was not affected by a reducing environment suggesting that the target residue of the probe was not subjected to oxidation. Additionally, the pH had a great impact on the labelling intensity, indicating that the target residue may be activated by deprotonation. We have also assessed that the probe binds in the binding site of LgtC, with the sugar head believed to be occupying the acceptor binding pocket. Thus, we next wanted to identify the target residue responsible for the covalent modification of the enzyme by the probe. To first gain insight into this, two LgtC mutants, C246S and C246A, were tested in a labelling protocol. Both mutants were incubated with **15- α** (0.5mM) in HEPES buffer for 30 mins at 30°C, at four different pH's. Interestingly, both mutants were labelled by the probe and, in line with the observations made for WT LgtC, the labelling of the mutants is greater at pH above 7 (Figure 4.21.A). This indicates that cys246 in the lactose binding pocket is not the target residue of the probe. The docking simulations (Figure 4.19.B) are line with this, as the chloroacetamide WH of probe **15- α** , seems to lie too far from cys246 (shown in red) to modify it. There are 4 other cysteine residues in addition to cys246 in the solved structure of LgtC. However, none of them seems close enough to the binding site to be the target of the probe (Figure 4.21.B), which indicates that the target residue may not be a cysteine. This is in line with the previous observations suggesting that the residue was not subject to oxidation.

In order to get a definite answer regarding the identity of the target residue, the labelled LgtC band was extracted from the gel and submitted to mass spectrometry analysis at the King's CEMS Proteomics Facility. Trypsin digestion followed by LC-MS/MS analysis of the resulting peptides were used to identify the residue containing the chemical modification.

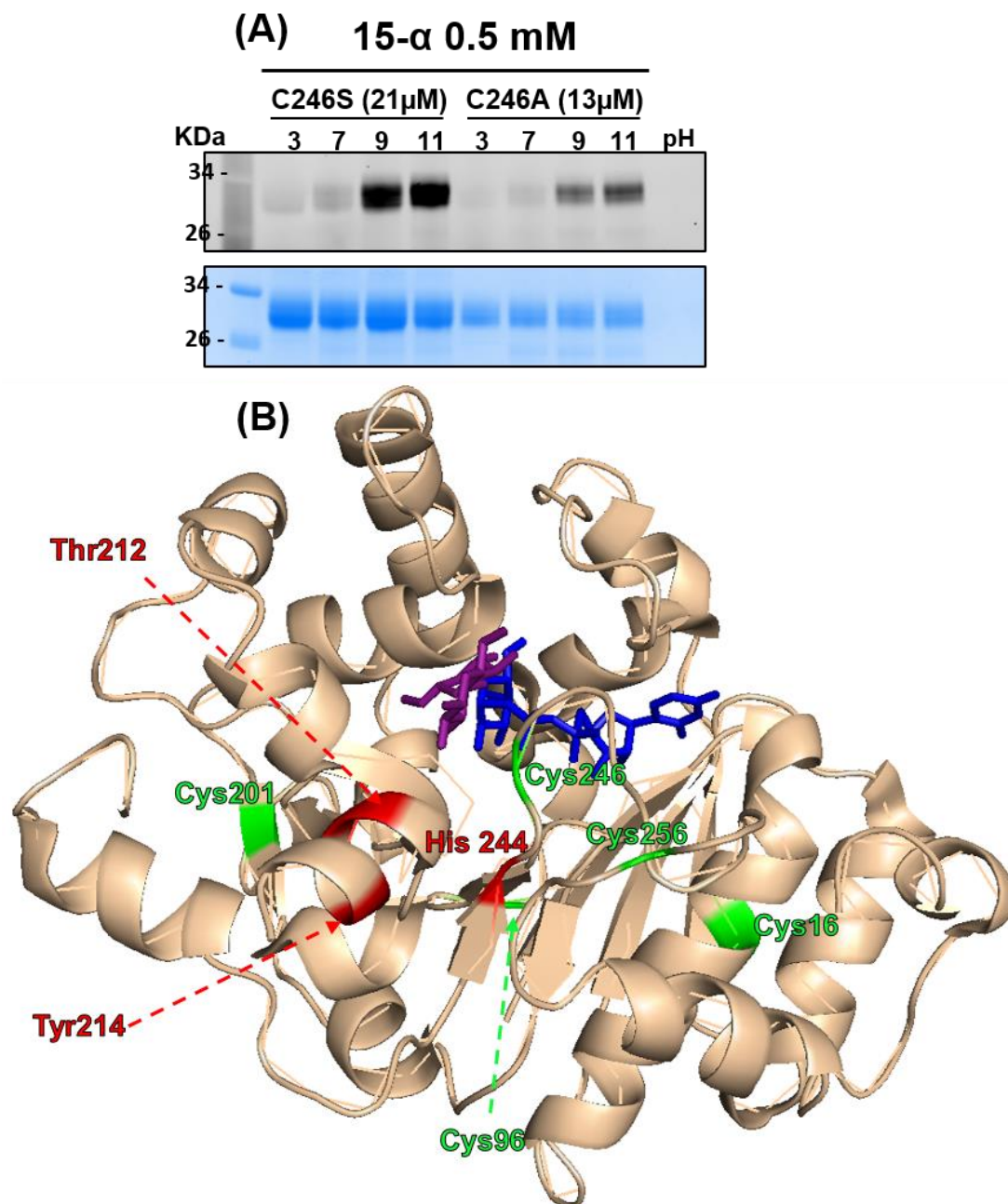


Figure 4.21. (A) Labelling of C246A and C246S LgtC mutants, performed by **15- α** at different pH. See general assay conditions in Fig 4.12. **15- α** stock concentration = 5mM, LgtC stock concentration = 210 μ M C246S mutant and 130 μ M for C246A mutant. (B) Visualisation of all cysteine residues (green) of LgtC (wheat) on Pymol. Possible probe target residues are shown in red, Acceptor and donor substrates are shown in purple and blue respectively

In direct comparison with the protein sequence of LgtC from *Neisseria meningitidis* retrieved from Uniprot (Q93EK7_NEIME), an 83% sequence coverage was obtained after enzymatic digestion. The last 25-C-terminal residues which are absent in the recombinant LgtC compared to the native enzyme⁵⁴ account for most of the 17% of unassigned sequence. MS analysis indicated that the modification did not occur on a cysteine residue as all 5 were found either trioxidised (cys96, cys201, +47.998Da) or unmodified (cys16, cys246, cys256). However, the absence of a signal in MS analysis does not necessarily indicate with confidence the absence of such modification. On the other hand, this may be the result of a very low amount of modified protein, its poor ability to fly in the instrument or the loss of the modification under the experiment conditions. Further work is therefore warranted in order to identify the exact target residue.

Should the above MS analysis be correct, three nucleophilic residues have been found in the vicinity of the supposed binding region of the probe's sugar-head: Thr212, Tyr214 and His244 (Figure 4.21.B). We hypothesise that one of these may be the target residue of the probe, but the labelling of appropriate LgtC mutants must be performed before any further conclusions can be drawn.

4.8. Labelling of LgtC in cell lysates and through live E. coli cells

So far, the labelling of LgtC with probe **15** has been exclusively studied on purified protein. In order to assess if our designed probe may be applicable for proteomic studies¹⁹⁶ we first investigated the cell lysates and live-cell labelling profile of **15** on standard lab strain *E. coli* BL21*.

4.8.1. Labelling of LgtC in E. coli cell lysates

Although the probe does not behave as it was designed to for LgtC, it does have the capacity to covalently label a recombinant protein. We then wanted to assess if the probe could label non-purified LgtC, and if it could to pull out other proteins in bacterial cell

lysates. For this, a mini culture of modified BL21* *E. coli* (containing the LgtC gene) was grown, induced for the expression of LgtC, harvested and lysed. Cell lysates were then incubated for 30 mins at 30°C with probe **15** (0.5mM) as well as protected probe **14** (acetylated precursor of **15**) (0.5mM) which will later be used for the intact cell labelling. The samples were analysed on a 12% SDS Page gel under denaturing conditions (Figure 4.22).

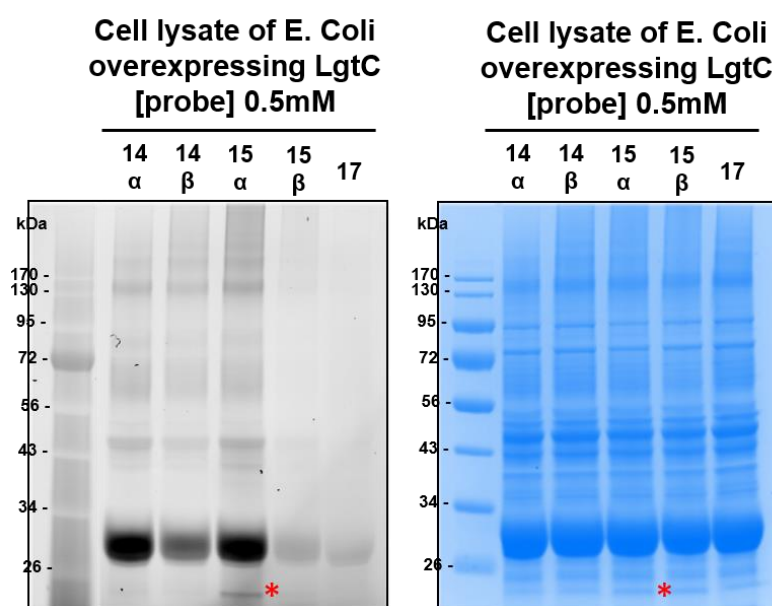


Figure 4.22. Labelling of non-purified LgtC in *E. coli* cell lysates performed by probe **14** and **15**. *General conditions:* *E. coli* cells overexpressing LgtC were lysed (8 hours after IPTC induction) using bugbuster (made from 10X stock in HEPES buffer) (100µL per 100mg of cell pellet) by shaking for 30mins at rt. Clear supernatant was collected and was incubated (9µL) with **15-α**/control (1µL, 5mM stock solution) for 1 hour at 30°C. See general assay conditions in Fig 4.12 for sample treatment. Fluorescence scanning (*left*), Coomassie stained (*right*). One low abundance protein (highlighted in red) is labelled exclusively by **15-α**.

The results show the ability of both **15** and **14** to label non purified LgtC in a complex cell lysates environment (Figure 4.22). As observed for the recombinant enzyme, the α -anomer of both **14** and **15** has a greater LgtC-labelling ability than the β -anomer. However, this difference in labelling is much less pronounced for probe **14** than probe **15**. This may be explained by the greater reactivity of protected probes due to the electron withdrawing effect of the Ac group, making the WH more electrophilic. Interestingly, these results provide further evidence that the observed labelling does not rely on sugar

recognition as both free sugar and sugar protected probes label LgtC. If the labelling was sugar-directed then a significantly weaker labelling would be expected for probe **14**, which in our case, is not observed.

On the other hand, in addition to LgtC, the probes also tagged some cytoplasmic *E. coli* proteins over a wide range of MW. Although, the labelling profile was highly similar for the four probes, one low abundance small protein (~20 kDa) was labelled exclusively by **15- α** (shown with a red star on Figure 4.22). This is of high interest as it indicates that for this protein at least, the recognition motif is the sugar scaffold and not the fluorophore as was the case for LgtC. These observations suggest the potential applicability of our designed probe for the labelling of both non- and carbohydrate-active enzymes in bacterial cell lysates.

4.8.2. Labelling of LgtC in live E. coli cells

Having assessed that the probe can label non-purified LgtC, as well as other bacterial cytoplasmic proteins, the probe was then applied to the labelling of protein in intact cells. A probe which can penetrate the permeable cell membrane and reach its target protein with minimal disturbance is of great interest for diagnostic applications as it can allow a quick identification of virulent bacterial strains. The ability of acetylated probe **14** to label LgtC in a whole cell labelling experiment was therefore investigated. Probe **14** was chosen instead of the deacetylated probe **15** for this experiment due to its greater lipophilicity which we hypothesised may allow it to penetrate the bilayer of the cell membrane. Once in the cell environment, we supposed the Ac protecting groups may be hydrolysed by bacterial esterases. To perform this experiment, cultures of modified BL21* *E. coli* (containing the lgt gene) were grown and harvested 0, 3 and 8 hours after IPTG induction. The cell pellets were then incubated with probe **14**, control **17** (1mM) and DMSO for 2 hours after which the excess probe was washed off. The cells were subsequently lysed

with Bug buster and the resulting cell lysates were separated on a 12% SDS-page gel and visualised by fluorescence read-out and Coomassie staining. This workflow is outlined in Figure 4.23.

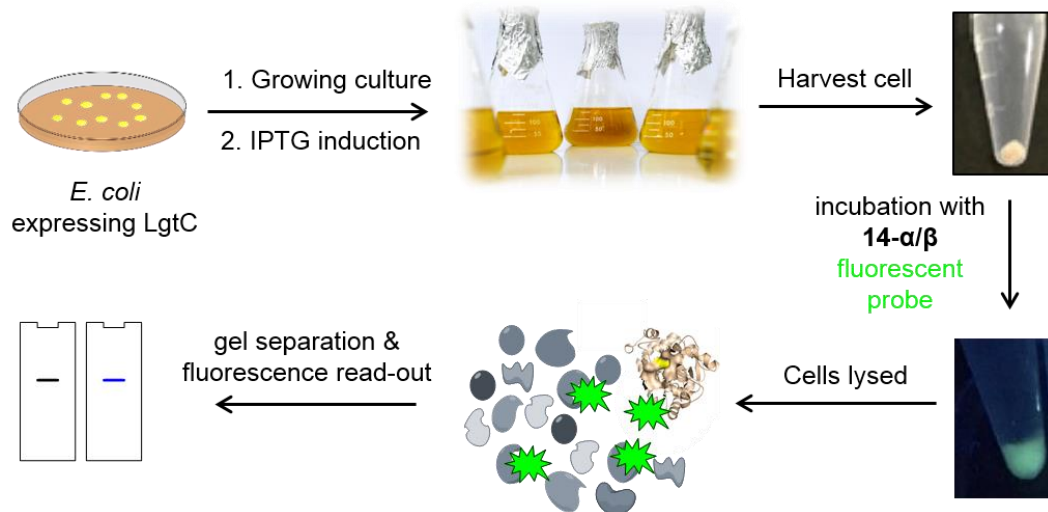


Figure 4.23. Workflow for the intact cells labelling of LgtC.

Interestingly, after successful removal of all unbound probe (achieved after 4 washes, Figure 4.24. A), the cell pellets incubated with both **14 α / β** and **17**, remained fluorescent while the control cells (incubated with DMSO) had no visible fluorescence. (Figure 4.24.B, observed under UV, 365nm). This indicates that probe has penetrated the cells and/or bound to the surface of the cells to some degree. Following subsequent cell lysis, it was observed that unlike the DMSO cell lysates, the lysates treated with **14 α / β** and **17** were fluorescent. This suggests that the probe has penetrated the periplasm and potentially the cytoplasm (Figure 4.24.C). The cell debris had some residual fluorescence, indicating that the probe may also tag membrane bound proteins. This is not unexpected as the probe is designed to label LgtC, a GT located in the LOS envelop of the cell membrane in Gram negative bacteria. In this context LgtC is expressed in the cytoplasm of *E. coli* therefore, only the soluble fraction was analysed. In the following chapter however, this will be discussed in more detail.

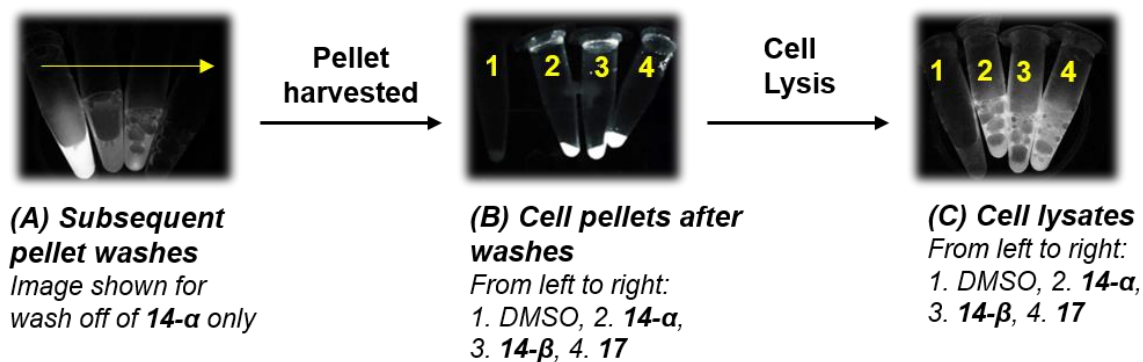


Figure 4.24. Observation of fluorescence of reaction mixture at different stages of the protocol (UV lamp, long wavelength 365nm). Images taken of the cell pellets harvested 8 hours after IPTG induction.

The cell lysates were then separated on a 12% SDS Page gel under denaturing conditions and initial results are shown in Figure 4.25. Interestingly, while no fluorescent band was present for either control (DMSO and **17**), a fluorescent read out was detected for the protein band corresponding to LgtC. This indicates that both anomers of **14** have penetrated not only the periplasm but have also reached and tagged LgtC in the cytoplasm. Additionally, as seen for the labelling in cell lysate, some other *E. coli* proteins were tagged by probe **14** both before and after inducing over-expression of LgtC (highlighted in purple in Figure 4.25). A similar and more resolved labelling profile of the probe is observed compared to the labelling in cell lysate (Figure 4.22).

Under these conditions, the stronger labelling ability of the α - over the β -anomer is not particularly marked. This had been observed for the labelling of LgtC in cell lysate by Ac probe **14** but not by deAc probe **15** (Figure 4.22). This may therefore suggest that the protecting groups on the sugar head have not been hydrolysed, and that probe **15** is not released *in vivo*. However, the presence of the low MW band labelled only by the α -probe (highlighted in red in Figure 4.25) suggests the contrary, as this was observed exclusively for probe **15- α** in cell lysate labelling (Figure 4.22).

These preliminary results are of high interests as they indicate that probe **14** can both penetrate the cell membrane of Gram-negative bacteria and covalently label non purified

LgtC as well as some, but not all, bacterial cytoplasmic proteins across a wide range of molecular weights.

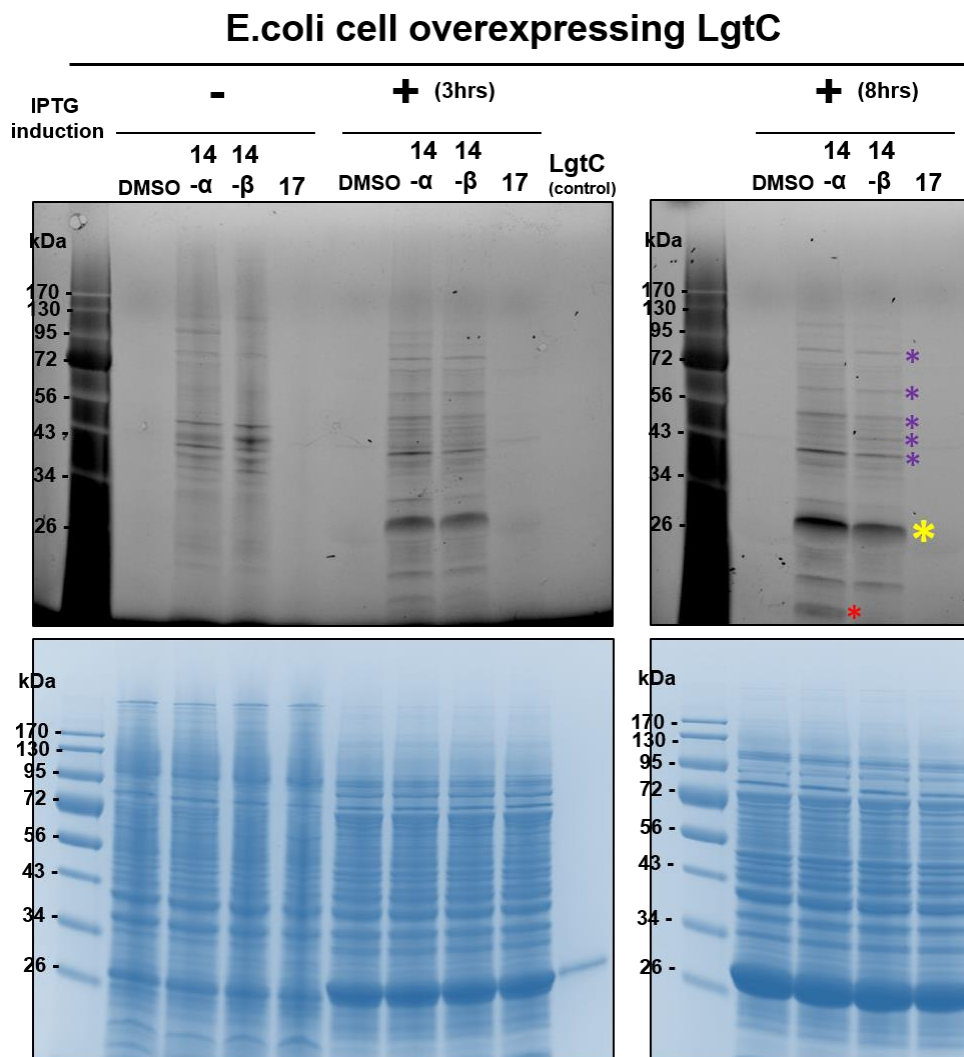


Figure 4.25. Intact cell labelling of LgtC. *General conditions:* *E. coli* cells overexpressing LgtC (100mg of pellets 0, 3 and 8 hours after IPTG induction suspended in 900 μ L of HEPES buffer) were incubated with **14- α** /**17**/ DMSO (100 μ L, 10mM stock solution) for 2 hours at 30 $^{\circ}$ C. The cells are subsequently lysed using bug buster (100 μ L, made from 10X stock in HEPES buffer) by shaking for 30mins at rt. Clear supernatant was collected. See general assay conditions in Fig 4.12 for sample treatment. *Top:* Fluorescence readout. LgtC (yellow) is labeled through cell by both **14- α** and **14- β** . Other *E. coli* cytoplasmic proteins (purple) are also labeled by the probes. One low MW protein is labelled exclusively by **14- α** (red)

Following this, the effect of the cell - probe incubation time on the observed labelling was investigated. Induced *E. coli* cells (harvested 8 hours after induction) were incubated with **14- α** and **14- β** (1mM) over a range of different time periods (30 – 240 mins) (Figure 4.26). As a control, the cells were also incubated and with compd **17** (1mM) for 240 mins. For compd **17** a very faint band was observed for LgtC after 240 mins of incubation but, as

expected, most of the probe was found unbound at the gel front (Figure 4.26). As seen for the labelling of recombinant LgtC, a time dependant in-cell labelling was observed which is consistent with a covalent enzyme modification. Within 30 mins, the probe had already penetrated the cells which suggests fast probe delivery to the cytoplasm. Additionally, the labelling for **14- β** is much weaker than for **14- α** at short incubation times. However, after 120 mins, the difference in the labelling profile through cells for each probe became non-significant (Figure 4.26). Although not observed for the recombinant protein, this is consistent with the results of the above preliminary experiment, performed at 120mins of incubation, for which both anomers exhibited a similar labelling intensity (Figure 4.25).

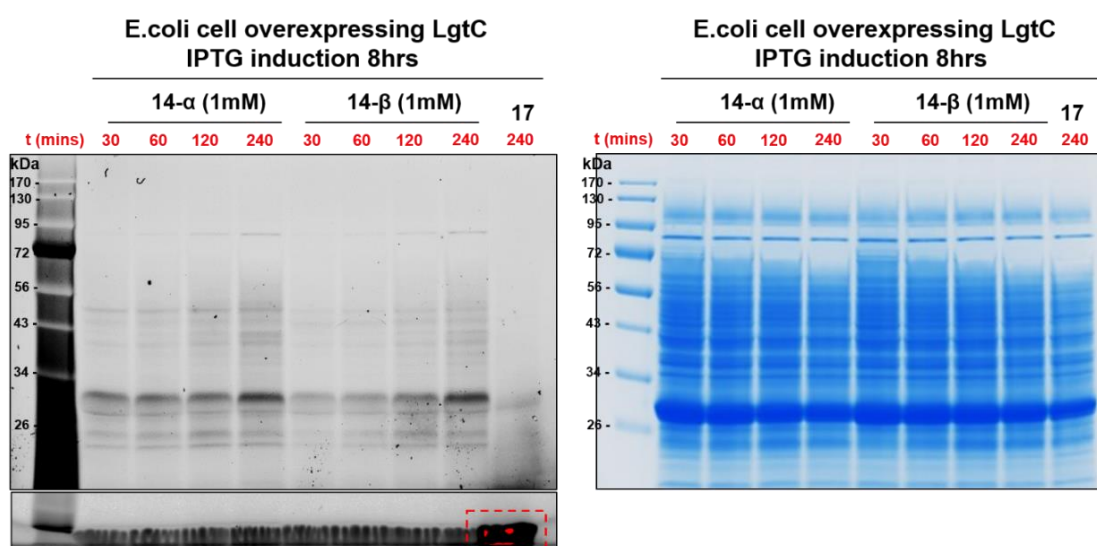


Figure 4.26.: Effect of incubation time on intact cell labelling of LgtC. See Fig 4.25 for general conditions. Alterations: Cells were incubated with probe / control for 0.5 - 4 hours at 30°C *Left:* Fluorescence readout. A time dependant in-cell labelling is observed for both **14- α** and **14- β** *Right:* Coomassie stained gel.

For both recombinant- and in-cell LgtC, at short incubation times the α -anomer is significantly more potent than the β -anomer. For recombinant LgtC, an increase in protein degradation was observed upon increasing incubation times (Figure 4.15), which affects the enzyme activity and folding and by association also impacts the probe's binding capacity. This inevitably reduces the accuracy of the observations made after long incubation times. In a cellular environment, LgtC remains stable and no degradation occurs even after long cells - probe incubation times (Figure 4.26). This may explain the

difference in the probe's observed labelling profile for recombinant and in-cell LgtC after long incubation times: The β -probe may be less active towards LgtC but if left in contact long enough, it can eventually reach the same amount of labelled LgtC as the α -anomer, except if the protein has degraded over time (which is the case for recombinant LgtC).

If **14** is hydrolysed to **15** *in vivo*, the above hypothesis may explain the fact that after long incubation times (>2hours), no significant difference between the labelling intensity of the two anomers is observed. Equally, this observation may also be due to the fact that **14** is not hydrolysed *in vivo* (Figure 4.27) as we have shown that the difference in labelling between the two acetylated probes (**14- α** and **14- β**) is not pronounced. Two analyses can be conducted in order to get a definitive answer: (1) Utilise probe **15** in a whole cell labelling experiment and compare with the current dataset. However, sugar-free probes may not penetrate the cell due to their low lipophilicity. (2) Extract the unbound probe at the bottom of the gel and analyse the sample by MS.

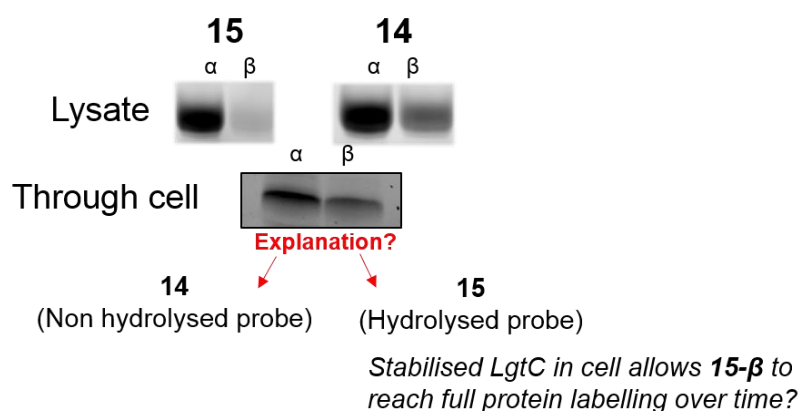


Figure 4.27. Hypothesis regarding the structure of the probe in cell

4.9. Further indications of the binding differences between probe and inhibitor

The findings regarding the binding of the probe discussed in the chapter inspired us to conduct two further analysis on the monosaccharide inhibitor **11** (See Appendix 4 Fig A4.2). This was performed with the aim of providing additional evidence that the probe and the inhibitor exhibit a different binding mode.

(1) Inhibitor **11** was tested in an inhibition assay against LgtC mutant (C246A) under standard procedure (described in chapter 2 and 7). Interestingly, all activity observed on WT LgtC was lost on the enzyme mutant (see Appendix 4 Fig A4.2). This was initially surprising as the current hypothesis is that the inhibitor does not bind in the active site. However, as the probe has the ability to label the LgtC mutant, this provides further evidence that inhibitor and probe have a different binding mode. Knowing that the probe binds in the active site, this is an additional indication that the inhibitor may bind elsewhere. We hypothesise that the mutation at position 246 results in a change in the overall structure of the enzyme which in turn prevents the binding of the monosaccharide inhibitor.

(2) Cmpd **9**, the acetylated precursor of inhibitor **11** was tested in an inhibition assay against WT LgtC under the standard procedure (described in chapter 2 and 7). Interestingly, a drastically reduced enzyme inhibition was observed for the Ac sugar (cmpd **9**) in comparison to free-sugar motif (cmpd **11**): at 1mM 15% of enzyme activity was observed for **11** against 89% of enzyme activity for **9** (see Appendix 4 Fig A4.2). However, this was not observed in the case of the probe, as Ac probe **14- α** labelled LgtC with equal intensity as the deAc probe **15- α** . This indicates that for the binding of the monosaccharide, the free sugar motif is essential while for the binding of the probe it has no effect. This provides further evidence for a different binding-mode of fluorescent probe **15** and inhibitor **11**.

4.10. Summary and conclusions

Fluorescent probe **15** was synthesised as part of the investigation into the mode of action of inhibitor **11** (Chapter 3). Both anomers of probe **15** were synthesised in three steps from cmpd **9** (Ac precursor of **11**) with 30% overall yield: 14% for **15- α** and 16% for **15- β** . The anomers were separated after the first glycosylation step and subsequent reactions

were performed on individual anomers. ^1H NMR combined with 2D ROESY NMR analysis has allowed for the unambiguous identification of each anomer. The strong fluorescence observed in DMSO was found to be quenched in an aqueous environment although at the tested concentration the probe was still highly fluorescent in HEPES buffer. Interestingly, probe **15** was found to covalently modify LgtC and a greater potency (8-fold) was observed for **15- α** compared to **15- β** . The labelling of recombinant LgtC has been shown to be concentration, time and pH dependant, with greater labelling observed at pH above 7. These results support a covalent enzyme modification given the increase in target modification over time and an activation of the target residue in basic pH. Substrate competition experiments provided confidence in the conclusion that the probe binds in the active site of the enzyme and that the fluorophore, instead of acting as a reporter group, plays a key role in binding (Figure 4.28). This proposed binding mode corroborates the greater potency and **15- α** (better UDP-Gal mimic) as well as the observed absence of labelling of denatured enzyme (active-site specific binding). The nucleophilic target residue of the probe has not yet been identified, but LC-MS/MS analysis indicated that it may not be a cysteine modification, and three residues have been identified as potential site of covalent enzyme modification. Initially designed to target cys246 in the acceptor binding site, enough evidence has now been gathered to conclude that this residue is not the probe target. Other studies have reported unexpected outcomes in covalent probe and inhibitor design¹⁴⁴.

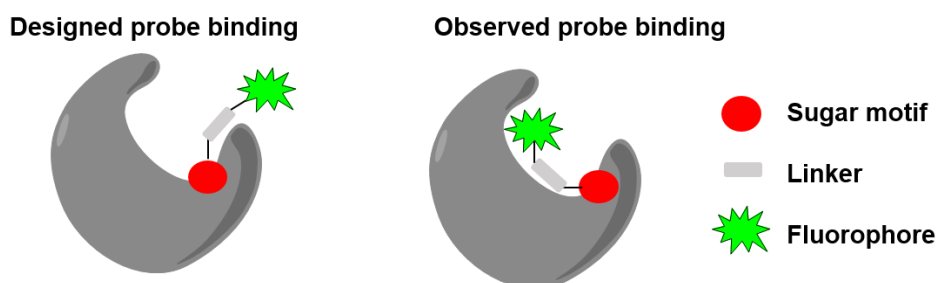


Figure 4.28. Difference between designed and observed probe binding

In addition to providing corroborating evidence for the binding mode of the probe, the work described in this chapter also gave further insight into the binding of the monosaccharide inhibitor **11**.

- (1) The binding of the probe was found to be non-competitive for inhibitor **11**.
- (2) While probe **15** labels C246A LgtC mutant, inhibitor **11** exhibits no inhibition of the mutant (Figure 4.29).
- (3) While cmpd **14**, the acetylated precursor of probe **15**, labels WT LgtC, a drastic loss in inhibition of LgtC was observed for cmpd **9**, the acetylated precursor of inhibitor **11** (Figure 4.29).

The above results strongly suggest a different binding mode for probe **15** and inhibitor **11**. Considering our proposed binding mode for the probe, in the binding pocket of LgtC, the dataset provides additional evidence towards a potential allosteric inhibition mode for the monosaccharide inhibitor **11**.

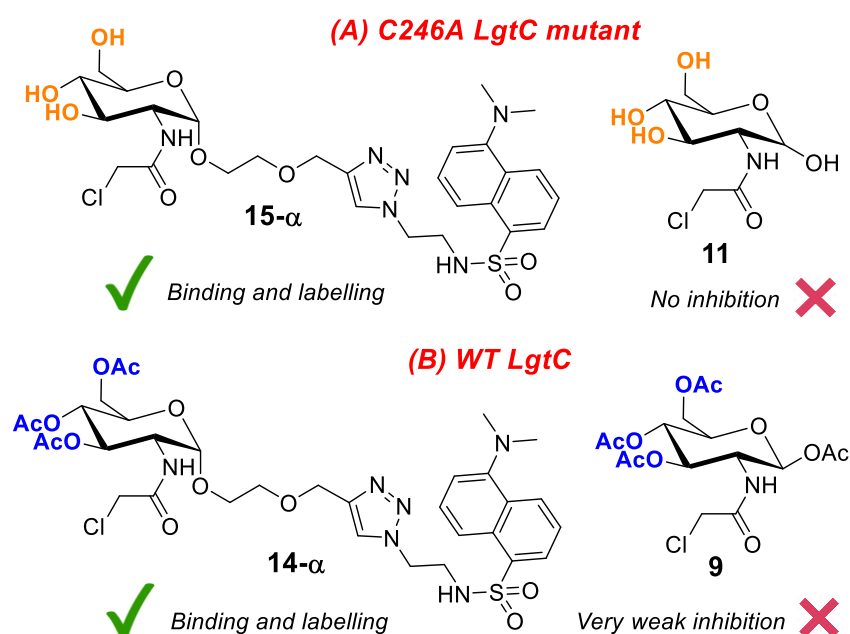


Figure 4.29. Dataset suggesting a difference in binding between probe **15** and inhibitor **11**. (A) **15** binds to C246A LgtC mutant while **11** does not. (B) **14** binds to WT LgtC while **9** does very poorly.

Finally, we have shown that the probe has the ability to label over-expressed LgtC in a complex cell lysate environment. Additionally, the acetylated precursor of the probe can penetrate the outer and inner membrane of *E. coli* and covalently modify LgtC in intact cells. It remains unclear if the free-sugar probe is released once in the cell (upon Ac hydrolysis) and additional experiments have been suggested in order to get a definitive answer. Whereas, for LgtC, the fluorophore takes part in binding, the probe may bind other proteins *via* the sugar head recognition motif, as designed. Importantly, this dataset has shown that the probe can penetrate the membrane of Gram-negative bacteria for the targeting of bacterial proteins. Therefore, this study on a non-pathogenic lab strain is a proof of concept towards the applicability of our probe for cell lysate and *in vivo* protein labelling. In the next chapter, the probe will be used in a chemical proteomic study on *H. influenzae* R2866, a pathogenic Gram-negative bacterium known to express LgtC with a high degree of phase variability.

Chapter 5

**Carbohydrate-based fluorescent probes for
bacterial proteomic studies on *H. influenzae*
R2866**

5.1. Introduction

5.1.1. Chemical proteomics and application to bacterial studies

The development of modern genomic technologies has enabled the sequencing of bacterial genetic code. In 2013, over 1,800 bacterial genomes had been entirely sequenced including important pathogens such as *Staphylococcus aureus*, *Mycobacterium tuberculosis* and *Clostridium difficile*¹⁹⁷. Although this acquired knowledge has enabled the differentiation of bacterial strains by genotypes, the bacterial genome only provides static genetic information and genetically similar strains can have remarkable differences in terms of pathogenicity and virulence¹⁹⁸. Therefore, the genetic information, which provides the primary sequence of a protein, is a very limited tool for the full understanding of the biological function of a pathogen. Proteins are, in most cases, regulated by post-translational modifications making it impossible to monitor their activity and predict their structures and functions *via* transcript of protein or expression levels alone^{198,199}. As opposed to the genome, the proteome is a dynamic system that is continuously regulated, and is directly related to cell functions in pathogens and all other domains of life^{198,199}. In addition to this, some clinically relevant bacteria are genetically intractable such as the Gram-positive bacterium *Clostridium difficile*, making genetic approaches obsolete for such pathogens¹⁰⁰.

Over the past decade, proteomic analysis has become an essential strategy for the investigation of many bacterial processes including bacterial virulence²⁰⁰, antibiotic-resistance²⁰¹ and antimicrobial drug discovery^{96,202}. This has led to a better understanding of the biology of pathogens essential for the development of novel classes of antibiotics²⁰³. Chemical proteomics for protein profiling in live cells or cell lysates, uses chemistry, *via* covalent probes often based on natural product²⁰⁴, and mass spectrometry (MS) analysis for target identification. This is a widely used strategy for the interrogation of the bacterial proteome^{200,205,206}. Chemical proteomics techniques can generally be classified into two

categories: gel-based proteomics and gel-free proteomics (also known as shotgun proteomics)²⁰². The former involves protein separation step by gel electrophoresis (one- or two-dimensional) prior to mass spectrometry target identification. Beyond proteome analysis from different species, this approach also enables the comparison of the proteome from the same species in different cell cycles, at different stages of cell growth, or under different conditions (i.e. pH, temperature, chemical environment). However, one of the main limitations of the gel-based method is the sensitivity of gel electrophoresis analysis, which may fail to detect low abundance proteins. Alternatively, in gel-free proteomic studies, the complex protein mixture is directly digested by proteases and analysed by MS.

Chemical proteomics requires the use of covalent probes, designed for the irreversible modification of a specific target protein (or class of targets). These probes are divided into two classes: activity-based probes (ABPs) and affinity-based probes (A/BPs) (discussed in Chapter 4). ABPs have been applied for protein profiling in bacterial samples (activity-based protein profiling (ABPP)) of both of Gram-negative and Gram-positive pathogens²⁰⁵.

A study from Staub and Sieber reported a chemical proteomic strategy for the labelling and identification of penicillin-binding proteins (PBPs) in *P. putida*, (Gram-negative) *L. welshimeri*, and *B. licheniformis* (Gram-positive)²⁰⁷. They synthesised a selection of natural and synthetic small β -lactam-based probes and incubated them with intact cells (Figure 5.1.A). Interestingly, subsequent MS analysis showed that while the natural probes labelled a variety of PBPs, synthetic probes exclusively tagged PBP unrelated enzymes such as the virulence-associated enzyme ClpP²⁰⁷. In a follow-up study, they applied their probes to the identification and functional characterisation of resistance associated enzymes in methicillin resistant *S. aureus* strains (MRSA)²⁰⁸. Interestingly, artificial probes labelled PBP2', which plays a key role in the β -lactam resistance of MRSA²⁰⁸. While these

probes could successfully monitor the activity and function of known resistance associated enzymes, they could also be applicable to the monitoring of unexplored and yet unidentified proteins associated with AMR in pathogenic bacteria. This would be of high interest for the discovery of novel therapeutic options.

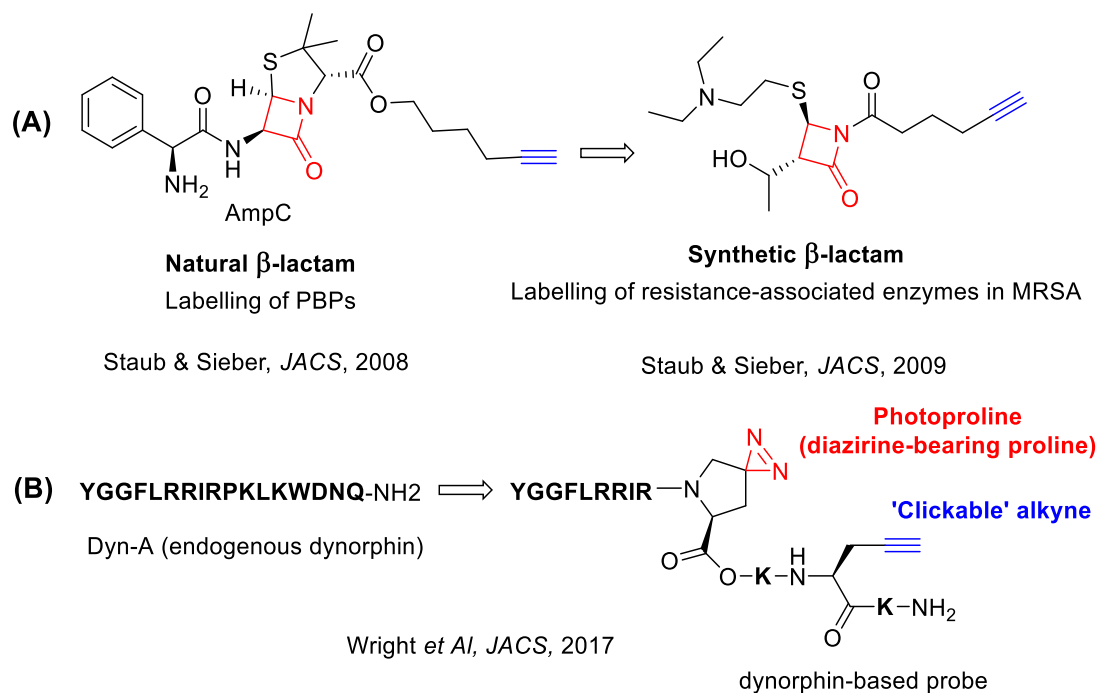


Figure 5.1. Examples of chemical probes applied for protein profiling in bacterial samples

More recently, Wright *et al.* reported a chemical proteomics study with the aim of understanding the mechanism by which human peptide dynorphin induces virulence in *P. aeruginosa*²⁰⁹. They used peptide-based photoaffinity probes (Figure 5.1.B) in whole cell studies to identify dynorphin-recognising proteins; and identified ParS, a largely uncharacterized membrane sensor kinase, as the binding partner of dynorphin. Subsequent full proteome studies revealed that dynorphin induces a specific defence response in *P. aeruginosa*, a mechanism not observed in ParS mutant strains which were more susceptible to dynorphin-induced toxicity²⁰⁹. Therefore, ParS has been identified as a sophisticated virulence mediator in *P. aeruginosa*, inducing antimicrobial-peptide resistance. *Via* the identification of ParS as a promising drug target for the prevention of

virulence induction and AMR, this study has provided further evidence that these two biological processes are directly associated.

5.1.2. Carbohydrate binding proteins and their role in bacterial virulence

There is increasing evidence showing that the regulation of bacterial virulence is directly associated with proteins involved in complex carbohydrate regulation. In the following section we will discuss carbohydrate binding enzymes and proteins and describe their role in bacterial virulence.

5.1.2.1. Proteins with catalytic activity

Carbohydrate-active enzymes catalyze the breakdown, biosynthesis or modification of carbohydrates and glycoconjugates. They can be divided into four main classes: glycosyl hydrolases (GHs), glycosyltransferases (GTs), polysaccharide lyases (PLs) carbohydrate esterases (CEs). While GTs catalyse the formation of glycosidic bonds, both PLs and GHs are responsible for their cleavage (Figure 5.2). GHs hydrolyse glycosidic bonds while PLs proceed *via* lytic β -elimination (non-hydrolytic cleavage)²¹⁰. As for, CEs, they catalyze the de-*O* or de-*N*-acylation by removing the ester decorations from carbohydrates²¹⁰ (Figure 5.2).

Because of their key role in biosynthesis and alteration of bacterial surface glycoconjugates, GTs²¹¹ and GHs²¹², but also CEs²¹³ and PLs²¹⁴ have often been associated with virulence. For example, a study has shown that Gram-positive *Streptococcus dysgalactiae* degrades host tissue polysaccharides by secreting poly- and oligosaccharide lyases²¹⁴. As seen previously, GTs play a key role in adhesion to host cells and evasion of the host immune response by sophisticated mechanisms of host mimicry⁵³. Bacterial biofilm formation mechanisms mediated by GTs have also been reported²¹⁵.

Additionally, protein glycosylation in pathogenic bacteria has been widely studied over the past decade, providing accumulating evidence of a direct association of this type of

post-translational modification with bacterial virulence^{216,217}. For example, virulent glycoproteins have been identified in pathogenic bacteria including *Neisseria meningitidis*²¹⁸ and *Haemophilus influenzae*²¹⁹. These glycosylation events are responsible for bacterial adhesion to host cells and hijack of the host's cellular processes²¹⁶ (by mimicry).

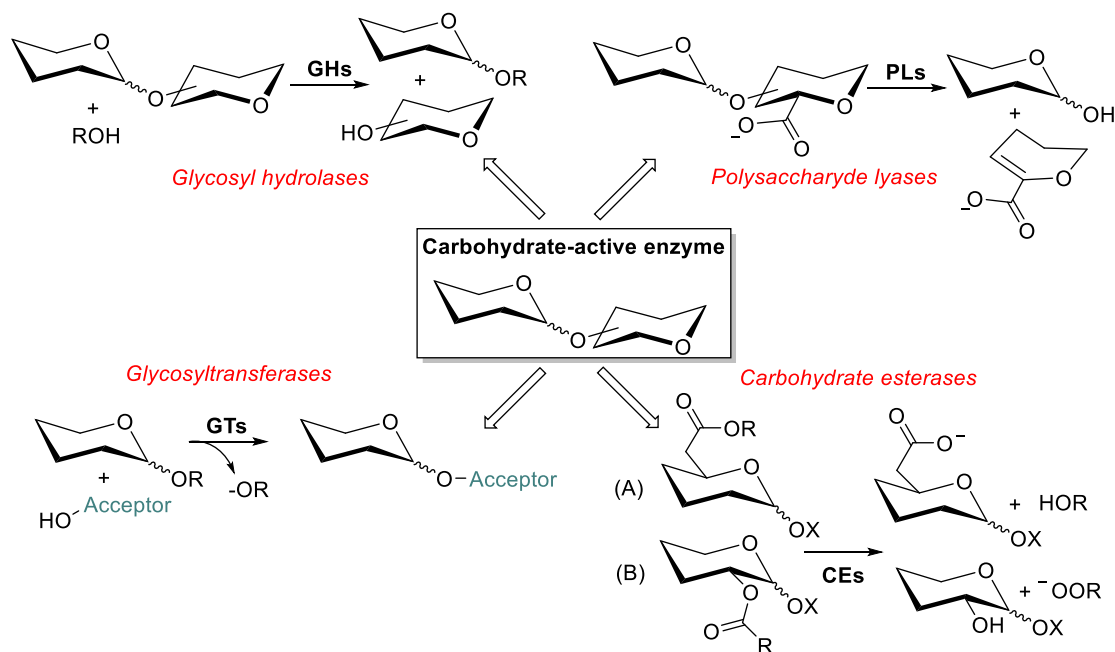


Figure 5.2. Classes of carbohydrate active enzymes. For CEs: (A) De-O-acetylation-sugar as acid (B) De-O-acetylation-sugar as alcohol

In addition to their participation in the initial steps of infection, relatively recent studies have shown that bacterial GTs also modulate host responses by the direct modification of key host signalling proteins²¹⁶. These, known as GT toxins, are major virulent factors used by pathogenic bacteria to manipulate the cellular functions of their eukaryotic host. In eukaryotic cells, protein glycosylation is operated by the highly conserved *O*-GlcNAc transferase (OGT), an inverting GT which forms β -anomeric sugar-protein bonds from UDP-GlcNAc. This can be rapidly reversed by glycosidase *O*-GlcNAcase (OGA), which exclusively cleaves β -anomeric sugar-protein bonds. In contrast, GT toxins are retaining enzymes and form α -anomeric sugar-protein linkages, which are not cleaved by OGA.

Thereby, this sophisticated mechanism allows bacteria to form a stable sugar–protein bond for a long-lasting toxin effect²²⁰. So far, only UDP-glucose and UDP-GlcNAc have been identified as sugar donor substrates of GT toxins, while the host protein acceptor appears to be more diverse. For example, *C. difficile* infections depend on the presence of the GT toxins TcdA and TcdB, active in the glucosylation of Rho proteins leading to inflammation and damage of the gut mucosa²²¹.

GHs in pathogenic bacteria have the ability to modify their host's glycan structures in order to favour bacterial survival and persistence²¹². A good example which demonstrates how GHs are involved in bacterial pathogenesis and immunomodulation is Endoglycosidase S from *Streptococcus pyogenes*, that cleaves glycans from human IgG and contributes to immune evasion²²². Additionally, the expression levels of a sialidase from *Pseudomonas aeruginosa* are highly correlated to the variations in adherence seen among strains²²³.

5.1.2.2. Proteins with no catalytic activity

Found in most organisms, lectins are the most common carbohydrate-binding protein devoid of catalytic activity. They bind mono- and oligosaccharides reversibly with high specificity but unlike antibodies, are not associated with immune response²²⁴. In bacteria, lectins mediate the attachment and binding of pathogens to their intended targets. Bacterial lectins occur commonly in the form of elongated proteins, known as fimbriae or pili which interact with glycoprotein and glycolipid receptors on host cells²²⁵. Bacterial surface lectins, occasionally referred to as sugar-specific adhesins, play a central role in the initiation of infection by regulating bacterial adherence to the host²²⁵. *E. coli* strain K99 illustrates the role of bacterial lectins in host-recognition as well as their high specificity. The pathogen binds to N-glycolyl-neuraminic-containing glycolipids only found on intestinal piglets' cells. Human cells contain N-acetylneuraminic acid analogues which

only differ from the former sugar scaffold by a single hydroxyl group. This explains at the molecular level why *E. coli* K99 can cause lethal diarrhoea in piglets but not in humans²²⁴.

Competition experiments have been performed to provide non-refutable evidence regarding the role of lectins in bacterial infections. The inactivation of bacterial lectins by administration of suitable sugars to the host has proven efficient at providing protection against infections²²⁴. In an interesting example, the oral administration of NeuAc(α 2-3)lactose to patients infected with *Helicobacter pylori*, a Gram-negative bacteria responsible for the development of gastroduodenal ulcers, has significantly decreased the gastric bacterial load²²⁶.

In drug-resistant *Pseudomonas aeruginosa*, galactose-binding LecA and LecB, cytotoxic lectins expressed in biofilm-grown cells, have been identified as virulence factors²²⁷. Towards developing “patho-blockers” and anti-virulence agents, Titz *et al.* have reported a covalent inhibitor of LecA, based on the addition of a reactive epoxide in position 6 of a galactose scaffold¹¹⁶ (see 2.1.2). A fluorescent glycoprobe (glyco-dye) derived from the inhibitor structure has been used for the labelling of LecA and the imaging of biofilms formed by *P. aeruginosa* *in vitro*¹¹⁶ (Figure 5.3). This work may be a stepping stone towards the development of pathogen-specific imaging agents to localize bacterial biofilm-associated infections inside an infected host.

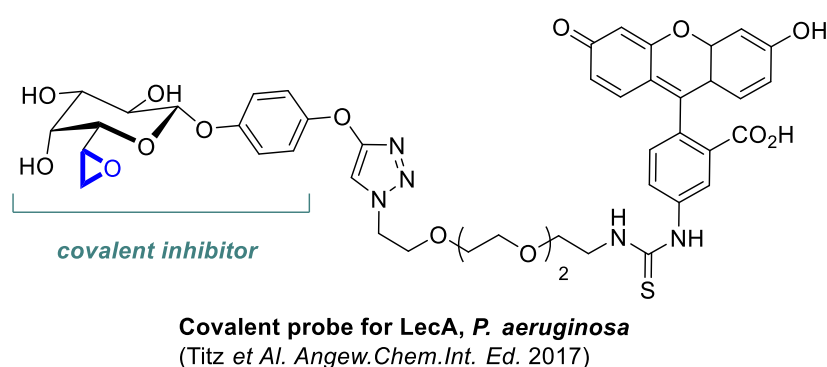


Figure 5.3. Fluorescent probe for the labelling of LecA from *Pseudomonas aeruginosa*

5.1.3. Carbohydrate-based probes for bacterial proteomic studies

Therefore, carbohydrate-binding enzymes and proteins play an essential role in bacterial virulence and AMR. In chapter 4, different strategies for the design of GH and GT probes were presented. The variety of anchoring moieties and their mode of action was discussed alongside relevant examples. Beyond their use for the labelling of recombinant proteins, well-designed glyco-based probes may be applied to whole cell / cell lysate chemical proteomic studies for the pull-out and identification of bacterial carbohydrate-binding proteins involved in virulence. For example, Stubbs *et al.* have reported a glucosamine-based probe the ability to pull out GH NagZ from in non-purified bacterial cell lysates of *P. aeruginosa*. NagZ is β -N-acetylhexosaminidase playing a critical role in peptidoglycan recycling and the induction of resistance to β -lactams²²⁸ (Figure 5.4).

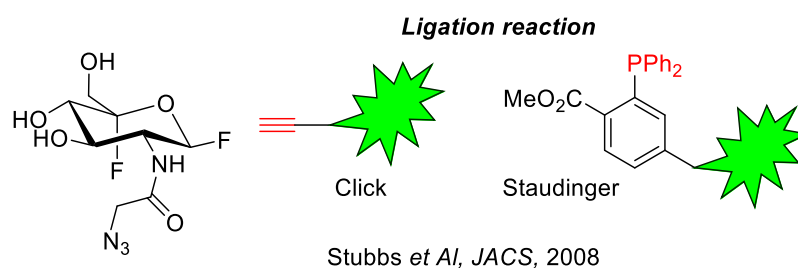


Figure 5.4. Example of a glycoprobe used in chemical proteomic studies on bacterial cell lysates

Although chemical proteomics studies have become a widely used strategy for the identification of virulence associated proteins in pathogenic bacteria, the use of carbohydrate-based probes towards that goal remains underexploited²²⁹. Glycoprobes have more commonly been used in glycoproteomic studies for the characterisation and identification of glycoproteins as a result of posttranslational modification²³⁰.

5.1.4. *H. influenzae* R2866 strain

H. influenzae R2866 is a virulent non-typeable *H. influenzae* (NTHi) strain first isolated from the blood of children with meningitis infection^{9,231}. Although this strain lacks a

capsular polysaccharide structure, it was reported to be serum-resistant, with a virulence level approaching that of encapsulated type b *H. influenzae* (Hib). The serum-resistance and its increased virulence are closely related to the terminal galactoside epitope of the outer-membrane LOS structure. As seen for *N. meningitidis* the galactosyltransferase LgtC is responsible for the biosynthesis of this virulent epitope⁹.

5.2. Objectives

In this last chapter, the synthesised glucosamine-based fluorescent probes were applied to a chemical proteomic study for the labelling of specific proteins of *H. influenzae* R2866. The aim of the study was to assess the ability of the probes to pull-out specific targets, and more interestingly, specific carbohydrate-binding proteins expressed in a virulent pathogen. To add more depth to the study and exploit the wide diversity of sugar scaffolds utilised by bacteria, an analogue series of galactosamine-based probes has been synthesised. This will allow the direct assessment of the role of the sugar recognition motif for binding and labelling. Additionally, with the modulation of their recognition motif, we hypothesised that, should the probes be successful for the labelling of specific *HI* proteins, their resulting protein profiling (the map of labelled protein) may be different. In addition to the testing of both probe-series on a single strain for the comparison of their protein profiling (Figure 5.5.A), the comparison of the tagged proteins performed by a single probe on different strains could be equally interesting (Figure 5.5.B). However, this was not performed as part of this preliminary study.

Haemophilus was chosen for this study as it is more easily cultured than *Neisseria* in a standard laboratory. Additionally, *Haemophilus* expresses LgtC, which is highly homologous to LgtC from *Neisseria*, successfully labelled by our α -Glc-based probe. Therefore, the direct labelling of endogenous LgtC as well as other phase-variable proteins, involved in serum resistance, is of high interest and may be attainable *via* this

strategy. This work was done in collaboration with Prof. Brendan Wren at the London School of Hygiene and Tropical Medicine (LSHTM).

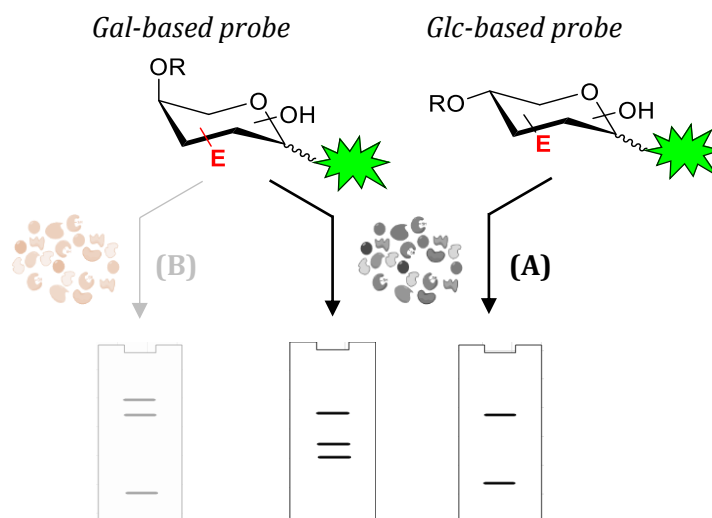
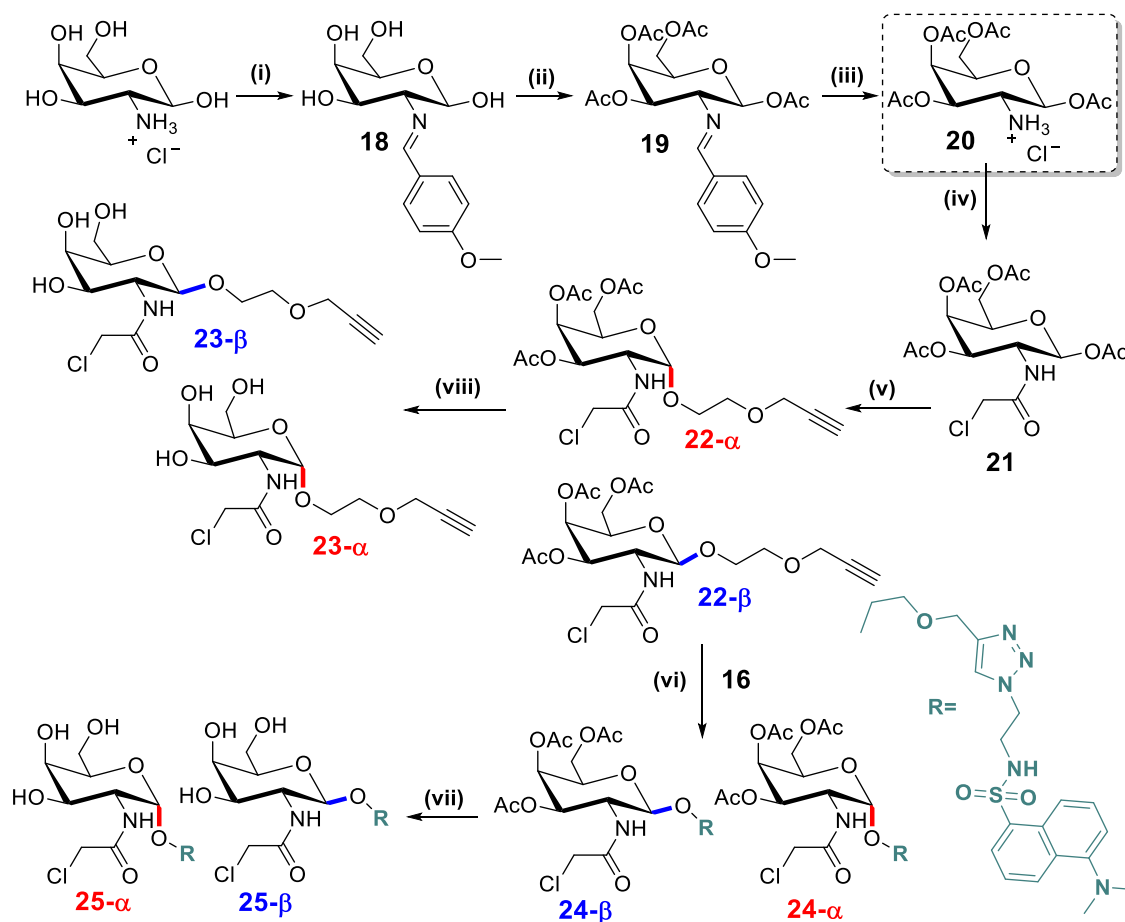


Figure 5.5. Different strategies for chemical proteomic studies. (A) Study of the labelling ability of different probes on one strain. (B) Study of the labelling ability of one probe on two different strains.

5.3. Chemical synthesis of galactosamine-based probes

Initially, this synthesis was performed by Hayley Wootton, MSci student working under my supervision. I used her protocol to repeat the synthesis for my own project. Galactosamine-based probe **25** was synthesised in seven steps from the commercially available galactosamine hydrochloride salt in 15% overall yield (8% for **25- α** and 7% for **25- β**) (Scheme 5.1). Unlike its glucosamine-based analogue, sugar protected galactosamine **20** was not commercially available. It was therefore synthesised in three steps from the non-protected galactosamine hydrochloride salt *via* an orthogonal protection/deprotection strategy. Following the method developed by Wulffen *et al.*,²³² the amine was first protected *via* imine formation using p-anisaldehyde to yield compound **18** (¹H NMR signal of N=CH observed at 8.13 ppm, see Appendix 1 Fig A1.38). The following selective and quantitative *O*-acetylation gave compound **19** (¹H NMR signals of 4 x RCOCH₃ observed between 1.82-2.13 ppm, see Appendix 1 Fig A1.40). Finally, subsequent imine hydrolysis under acidic conditions regenerated the free amine salt **20** (¹H NMR signal of NH₃⁺

observed at 8.71 ppm, see Appendix 1 Fig A1.42). From this point onwards, the synthesis of probe **25** is directly based on the synthesis developed for probe **15** (see 4.4). Following the attachment of the electrophilic WH (cmpd **21**), the linker was attached to the anomeric position to yield cmpd **22**, whose anomers were separated by normal phase chromatography (see 4.4) (alkyne proton signals detected at 2.47 and 2.45 ppm on ¹H NMR spectrum of **22- α** and **22- β** respectively, see Appendix 1 Fig A1.46 & 48). Each individual anomer was either reacted in a 1,3-dipolar cycloaddition with fluorophore **16** (triazole proton signal detected at 7.61 and 7.67 ppm on ¹H NMR spectrum of **24- α** and **24- β** respectively, see Appendix 1 Fig A1.54 & 56) or de-acetylated to afford the sugar-free partial probes **23- α** and **23- β** . The individual acetylated anomers of probe **24** were then hydrolysed to generate the probes **25** (refer to 4.4 for more details regarding this synthesis).



Scheme 5.1. Synthetic route of fluorescent probes **25α/β**. *Reagents and conditions:* (i) p-anisaldehyde (1 eq), NaOH (1 M), 0°C; 71% (ii) acetic anhydride (8 eq), pyridine (XS), rt, 16 h; 99% (iii) acetone, HCl (4 M, 1.2 eq), 56 °C, 5 min; 88% (iv) Et₃N (2.3 eq), chloroacetyl chloride (3 eq), rt, 1.5 h, 80% (v) propynol ethoxylate (3 eq), BF₃Et₂O (4 eq), DCM, 40°C, 8 hours, 70% (vi) **16** (1 eq), CuSO₄·5H₂O (1 eq), sodium ascorbate (1.5 eq), DIPEA (3 eq), THF : water 10:1, rt, 3hours, 74% (vii) MeONa (4 eq), MeOH, 10 mins, 0°C, 41%, (viii) MeONa (4 eq), MeOH, 10 mins, 0°C, 57%.

The identification of both anomers of galactosamine-based probe **24** was successfully performed *via* NMR analysis as described in chapter 4 (see 4.5.2). The weak orbital overlap between H₁ and H₂ was associated with a small coupling constant for **24-α** ($J_{1-2} = 3.5$ Hz) while the opposite was observed for **24-β** (Figure 5.6). Similarly, **24-β** is characterised by the presence of H₁ – H₃ and H₁ – H₅ correlations on the ROESY spectrum, which are absent for **24-α**.

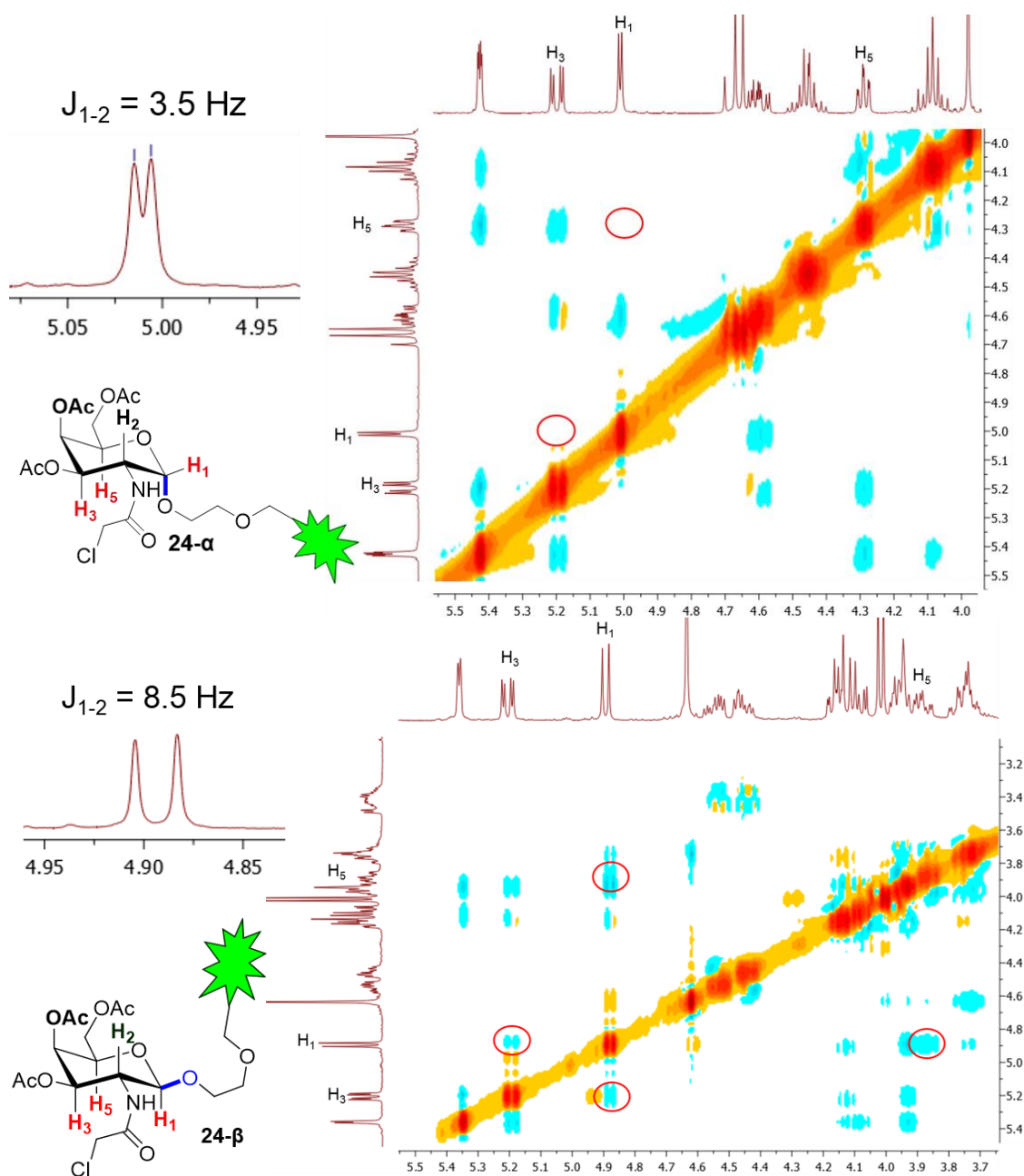


Figure 5.6. Identification of compound **24** anomers by ^1H NMR and 2D ROESY NMR, in CDCl_3 . (A) **24- α** (top) **24- β** (bottom).

5.4. Structural insight into the sugar scaffold for both probes

A similar NMR-based analysis was used to show the structural differences between the sugar backbones of both probes. Glucosamine and galactosamine are diastereoisomers. The structural difference between them relies exclusively on the stereochemistry at carbon 4, with an (*S*) configuration for glucosamine and an (*R*) stereocentre for

galactosamine. After full characterisation of the probes, the NMR data can be used to confirm the nature of the sugar backbone for each probe series.

In galactosamine-based probe **24**, proton H₄ is in equatorial position while both H₃ and H₅ are in axial positions (Figure 5.7, top). Weak coupling, associated with a weak orbital overlap, is therefore expected between H₄ and both H₃ and H₅. In the ¹H NMR spectrum of probe **24**, proton signals corresponding to H₃ (dd), H₄ (dd) and H₅ (td) are well defined and a small coupling constant is indeed observed between H₄ and both H₃ and H₅ ($J_{4-5} = 1.4\text{Hz}$, $J_{3-4} = 3.3\text{Hz}$) (Figure 5.7, top).

On the other hand, for glucosamine-based probe **14**, the H₄ proton is in axial position (Figure 5.7, bottom), strong coupling between H₄ and both H₃ and H₅ is therefore expected in this case. However, in the ¹H NMR of probe **14**, H₄ and H₅ are not well-defined peaks. Luckily, the analysis of the H₂ (ddd) and H₃ (dd) proton signals has enabled the indirect determination of the coupling constant between H₃ and H₄ (J_{3-4}). It is found to be 9.5Hz, which corresponds to a strong orbital overlap and strong coupling (Figure 5.7, bottom).

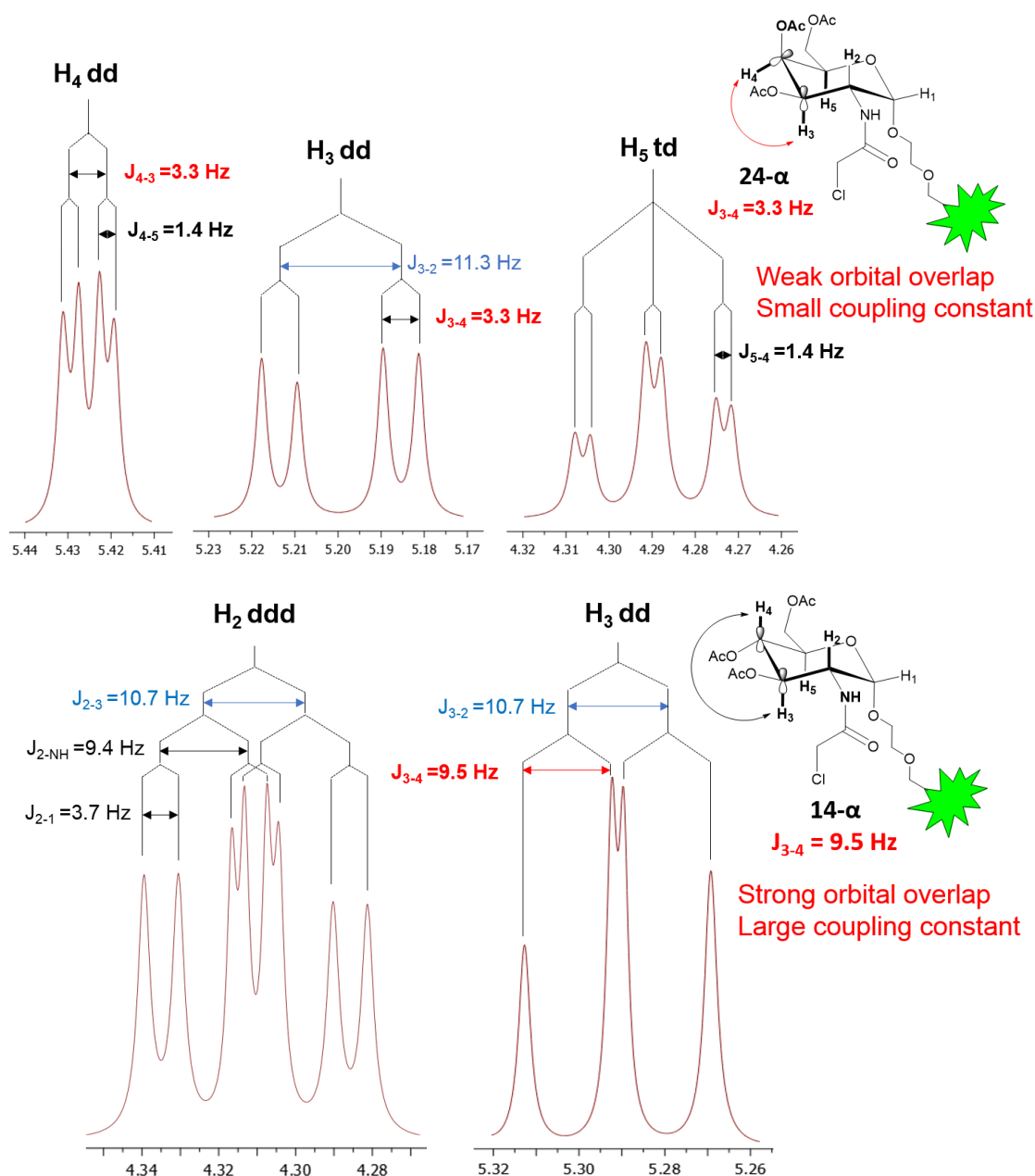


Figure 5.7. Top: Proton signals of H₂ (ddd) and H₃ (dd) in ¹H NMR (CDCl₃) of probe **14-α** (Glc-probe). Bottom: Proton signals of H₃ (dd), H₄ (dd) and H₅ (td) in ¹H NMR (CDCl₃) of probe **24-α**. Coupling constant allows for confirmation of sugar backbone.

Additionally, more information can be extracted from the ROESY spectrum of both probes.

As expected, the clear 1,3-diaxial interaction observed between H₄ and H₂ for the glucosamine-based probe **14**, is absent in the galactosamine-based probe **24** (Figure 5.8).

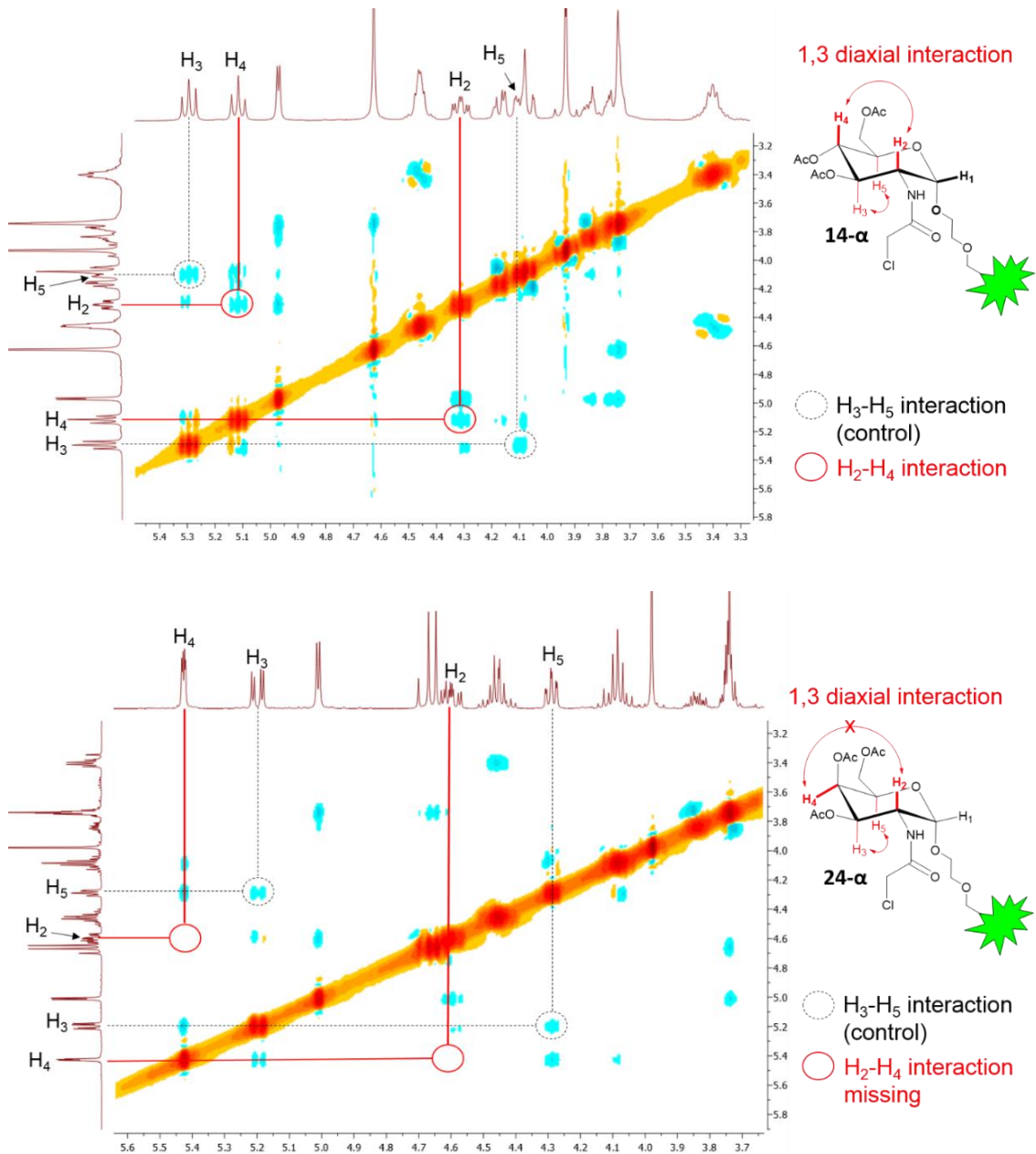


Figure 5.8. ROESY NMR of *Top*: probe **14- α** (CDCl₃). Presence of correlation between H₂ and H₄ resulting from a 1,3 diaxial interaction. *Bottom*: probe **24- α** (CDCl₃). Absence of correlation between H₂ and H₄ indicating no 1,3 diaxial interaction between them.

5.5. Chemical proteomic study: bacterial protein profiling

All four characterised probes were tested in whole cell protein labelling experiments on pathogenic *H. influenzae* strain R2866 with the aim of potentially labelling LgtC and / or other carbohydrate-binding proteins. This was performed at London School of Hygiene

and Tropical Medicine in collaboration with the group of Prof Brendan Wren. We reasoned that acetylated precursors, exhibiting a greater lipophilicity, were more likely to penetrate the cell membrane and were therefore more suited for whole-cell labelling work. The labile acetyl group may be cleaved off by bacterial esterases inside the cell. Four fluorescent probes were tested: two glucosamine-based probes **14- α** and **14- β** and two galactosamine-based probes **24- α** and **24- β** (Figure 5.9). As described in Chapter 4, molecules **16** and **17** were used as fluorophore controls for the assessment of the non-specific labelling.

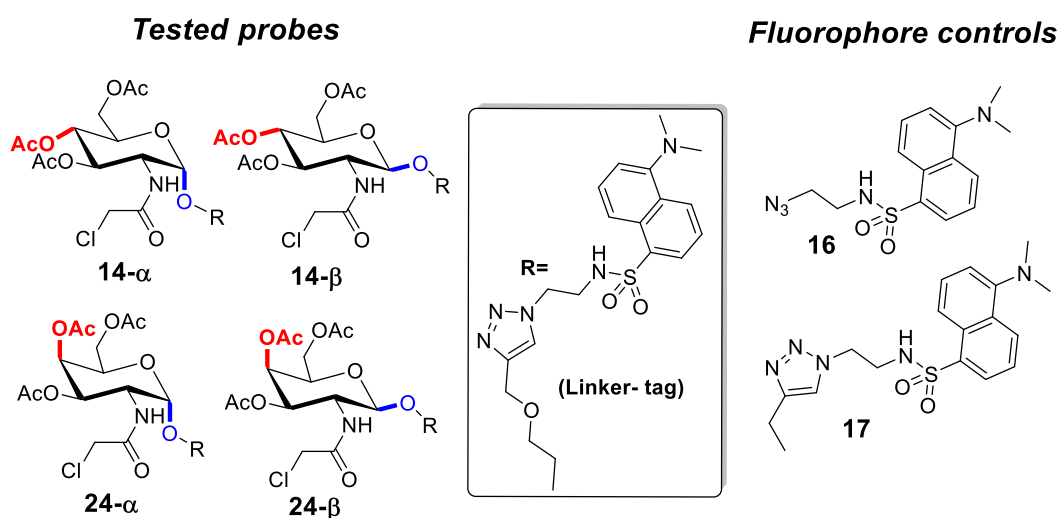


Figure 5.9. Fluorescent probes and control molecules tested in bacterial proteomic work

5.5.1. Bacterial growth for *H. influenzae* R2866

To identify the optimal growth period for the labelling experiments, growth kinetics of *H. influenzae* R2866 were first investigated. The growth of *H. influenzae* R2866 was monitored in supplemented Brain Heart infusion (sBHI) over 24 hours at 37°C and OD₆₀₀ was recorded at various time points. As shown in Figure 5.10, stationary growth phase was reached after 8-9 hours. The exponential growth phase, indicating healthy and performant cells, occurred over a period of two hours (between 5 and 7 hours after inoculation). From this, it can be concluded that the labelling experiment can be

performed on cell cultures of OD measurement between 0.6 and 1.4 to ensure fully alive and functioning cells.

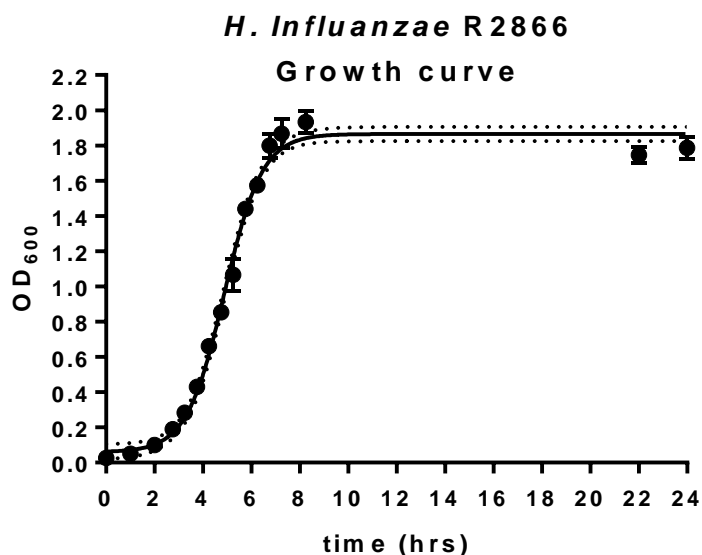


Figure 5.10. Growth curve of *H. influenzae* R2866. Cells are grown in sBHI broth over 24 hours. OD₆₀₀ is measured at specific time points. The experiment was carried out in triplicate (technical). Error bars represent standard deviation.

5.5.2. Through-cells protein labelling

Cell cultures of *H. influenzae* R2866 were grown in 20mL BHI broth for 5 hours (until OD₆₀₀=0.9) at which point the cells were harvested. Pellets equivalent to 4mL of cell cultures were incubated for 2 hours with all four probes (1mM in PBS (10% DMSO)) as well as with DMSO and fluorophore controls (cmpd **16** / **17**) (Figure 5.9). After incubation, the excess of un-bound probe was washed off (5% DMSO in PBS). The cells were subsequently lysed (Bug buster) and cell lysates separated on a 4-12% SDS-page gel visualised by fluorescence read-out and Coomassie staining. (Figure 5.11).

Interestingly, observation of the cell pellets under UV (365nm) after successful removal of all unbound probe (achieved after 4 washes) (Figure 5.11), demonstrated that those incubated with all four probes and fluorophore controls remain fluorescent while the control cells (incubated with DMSO) have no visible fluorescence. (Figure 5.11). This

indicates that some probe penetrates the cells and/or binds to the surface of the cells. Subsequent cell lysis shows that unlike the DMSO cell lysates, the probe and fluorophore control lysates were fluorescent. This suggests that, as seen for *E. coli* (chapter 4), the probe had penetrated the periplasm and/or cytoplasm of *H. influenzae* R2866 (Figure 5.11). This is of great interest as it indicates that beyond non-pathogenic *E. coli* lab strains, the designed probes also have the ability to penetrate the outer-membrane of pathogenic bacteria, known to be the first line of defence against foreign molecules. This therefore indicates that our probes have not been recognised as a “threat” by the pathogen, which was confirmed by viability assays performed on chocolate agar plate, indicating that the cells were still viable after incubation with the probes (Appendix 5 Fig A5.1). The lack of anti-microbial activity means that the probes are well suited for protein profiling studies on intact cells.

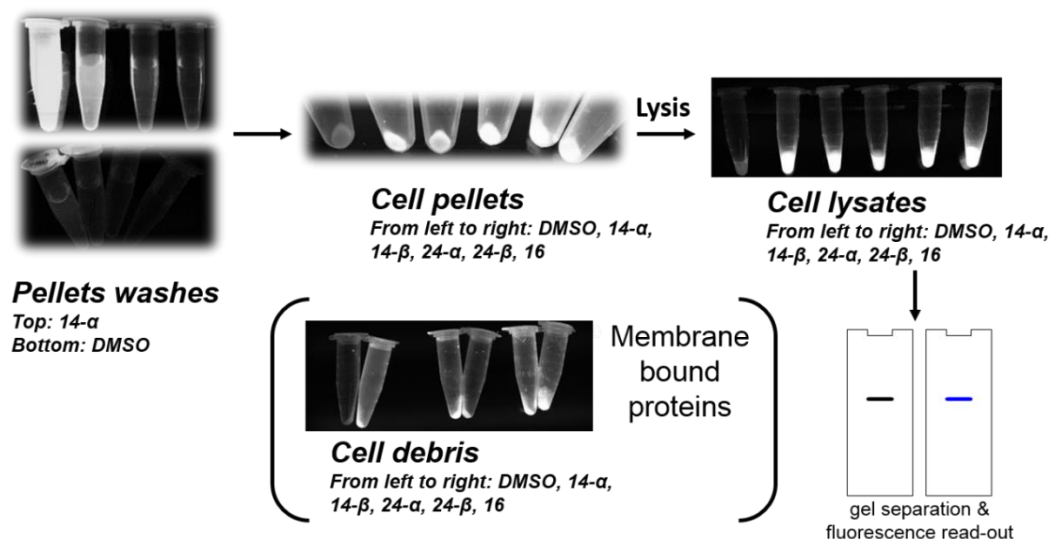


Figure 5.11. Workflow of the protein labelling in intact cells. Cell pellets and lysates obtained at different stages of the protocol are shown.

Interestingly, no significant difference in fluorescence was observed between all labelled lysates (Figure 5.11). Because of the high level of structural similarities between the tested probes, it is not unexpected that their degree of cell permeability is equivalent. Additionally, the cell debris recovered after cell lysis also had some residual fluorescence, indicating two possibilities: the probe may tag membrane bound proteins or it may

somehow be attached to the bacterial cell surface. The probes are designed to bind carbohydrate-binding enzymes which often are transmembrane proteins located in the bacterial cell wall, therefore the former hypothesis is not unexpected. Unfortunately, due to the high degree of “stickiness” of the post-lysis pellets (partially due to the bacterial DNA present in cell debris), no successful protocol for the profiling of these membrane-based labelled proteins has yet been found. The resuspension of the cell pellet in aqueous (PBS) or organic (DMSO) solvent or the use of DNase to break down DNA, have so far not been successful to achieve the above. This means that the significant proportion of carbohydrate-active enzymes located on the cell membrane may unfortunately not be investigated in this study. To account for the whole proteome further optimisation are required. However, the profiling of tagged soluble proteins has been successful using the described protocol and the results are described below.

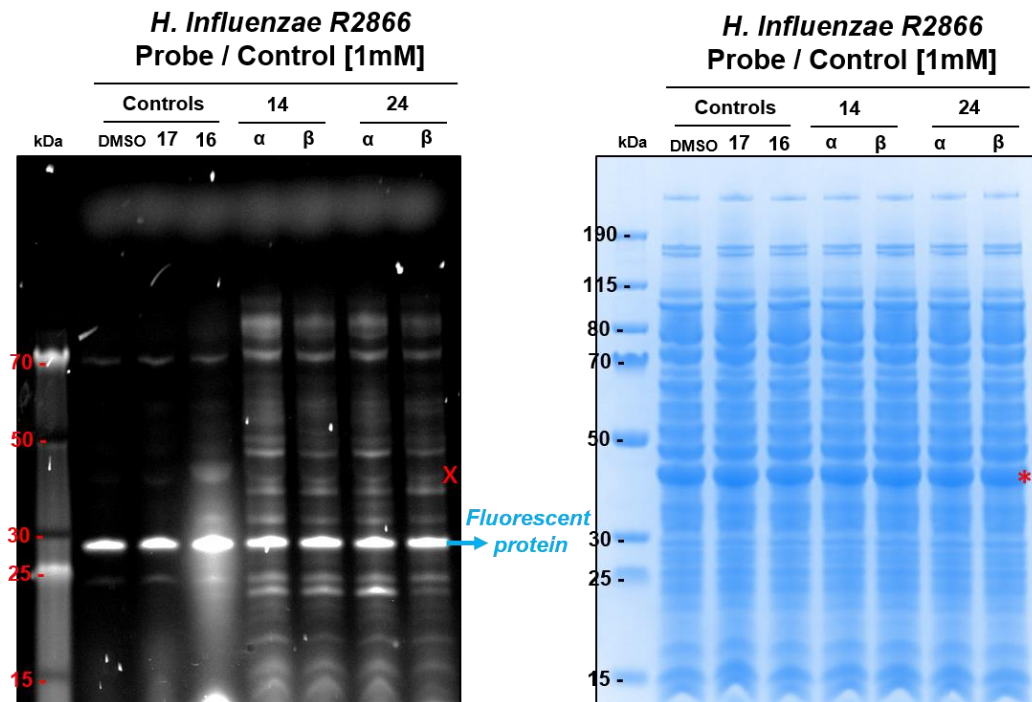


Figure 5.12. *H. influenzae* R2866 whole cells protein labelling profile performed by all four probes. *General conditions:* The cells are grown to $OD_{600} = 0.9$, pellets are suspended in PBS (900 μ L) and incubated with **24- α/β** / **14- α/β** / controls (100 μ L, 10mM stock) for two hours at 30°C. After being washed (5% DMSO in PBS) the cells are lysed with bug buster (100 μ L, X1 in PBS) and shaken for 30mins at rt. Loading buffer (2.5 μ L) is added to the clear supernatant (10 μ L) and loaded onto a 4-12% SDS page gel and run MES running buffer at 150V for 70 mins Fluorescence scanning (*left*), Coomassie stained (*right*). High abundance protein highlighted in red is not labelled by the probes.

While no specific labelling was observed for the fluorophore controls (cmpds **16** and **17**) the labelling of soluble proteins has been achieved by all four tested probes (Figure 5.12). One protein (~28kDa) was intrinsically fluorescent as is was observed under fluorescence scanning in all three control wells (DMSO, **16**, **17**). It was therefore not counted as part of the specifically labelled proteins. Additionally, while no difference was observed between DMSO and **17** controls, some background fluorescence was observed for **16**, most likely due to the reactive azide group (Figure 5.12). Therefore, for the rest of the work, control **16** was no longer used. Interestingly, both in term of potency and selectivity, no significant difference was observed in the labelling profile of all four probes (Figure 5.12). This may indicate that the fluorophore, a common motif to all probes, may play a key role in binding as was the case for the labelling of LgtC (Chapter 4).

The probes labelled many, but not all proteins, over a wide range of MW (from ~13 to 100kDa). Interestingly, comparison with the Coomassie stained gel indicates that the labelled targets are in relatively low abundance. Additionally, the high abundance protein (~45kDa, highlighted in red in Figure 5.12) was not labelled by our probes which suggests that the labelling process is directed more by protein specificity than by protein quantity. Subsequently, the effect of probe concentration and incubation time on the observed labelling profile was investigated.

5.5.2.1. Effect of probe concentration on labelling profile

Increasing concentrations of glucosamine-based probe **14- α** (0 - 1mM 10% DMSO in PBS) were used for the investigation of the probe concentration dependency while the incubation time remained unchanged (2 hours). A concentration dependant labelling was observed with weak protein tagging at 125 μ M of **14- α** with only four protein bands visible under fluorescence scanning (highlighted with a yellow star in Figure 5.13.A).

As the concentration of probe increases, the labelling of these four proteins becomes stronger and new bands appear, indicating the tagging of additional proteins. This also suggests a difference in affinity of the probe for its labelled targets. At 1mM of incubated probe, the labelling profile obtained during these investigations is highly comparable to the one presented in Figure 5.12 during initial testing, indicating the reproducibility of our labelling protocol. Unsurprisingly, as the concentration of probe increases, the cellular uptake of **14- α** increases accordingly, which is indicated by a greater amount of un-bound probe observed (Figure 5.13.B).

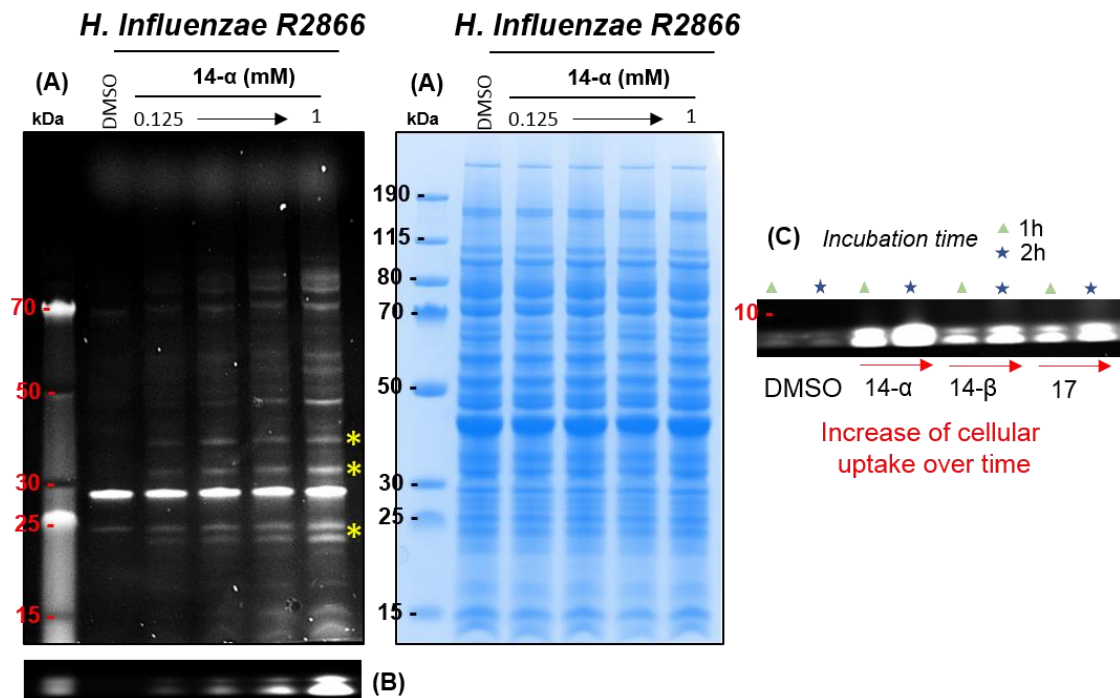


Figure 5.13. (A) Effect of probe concentration (**14- α**) on labelling profile of *H. influenzae* R2866 proteins in a whole cell labelling experiment. See Fig 5.12 for general conditions. Alterations: **14- α** (100 μ L, 1.25 – 10mM stock) The proteins detected at 125 μ M of probe are highlighted in yellow (B) Gel front, an increase in probe intake is observed. (C) Effect of incubation time on labelling profile of *H. influenzae* R2866 proteins in a whole cell labelling experiment. See Fig 5.12 for general conditions. Alterations: incubation time (1-2hrs) (only the fluorescence scanning is shown).

5.5.2.2. Effect of incubation time on labelling profile

The labelling of glucosamine-based probes (1mM in PBS (10% DMSO)) was investigated after 1 and 2 hours of incubation with *H. influenzae* R2866 strain cells. Interestingly, an increase in cellular uptake was observed upon increasing incubation time, as more un-

bound probe was found at the loading front of the gel after 2 hours than after 1 hour for both anomers and **17** control (Figure 5.13.C). This indicates that the transport mechanism of the probe through the cell membrane is time dependant and not immediate. However, this did not result in a stronger protein labelling as no significant difference in the labelling profile was observed after longer incubation time (not shown). This can be explained by two different factors:

- The target binding and modification is fast. After 1-hour, full target labelling is achieved therefore more probe entering the cell just remains un-bound.
- The target binding and modification is slow. Although more probe has penetrated the cell, our two time points (1 and 2-hours) are too close to each other for a difference to be observed.

In order to fully understand the labelling process, the investigation of the labelling profile after shorter (10 - 30mins) and longer incubation times (6 - 12hours) needs to be performed. Either way, cell penetration does not appear to be a limiting factor in this labelling process, as in-cell unbound probe is observed in all cases.

5.5.3. Bacterial cell lysate labelling

Next, the labelling of the cell lysates of *H. influenzae* R2866 was investigated. Cell pellets were lysed with Bug Buster prior to probe incubation and thereby generated cell lysates were treated with the four probes (Figure 5.9). For this experiment, deAc probes (**15** and **25**) were initially used as no cellular membrane needs to be crossed. However, for reasons that remain unclear to date, this was unsuccessful. The probe did not seem to label proteins with a strong enough intensity and the excessively large amounts of unbound probe made the gel imaging particularly challenging (Appendix 5 Fig A5.2). Different conditions were tried varying the probe concentration, the gel percentage and the running buffer. However, no attempted conditions resulted in the successful labelling of the cell

lysates by probe **15** and **25**. We hypothesise that this observation may be the result of the low reactivity of the deAc probe as discussed in chapter 3 (see Figure 3.15).

The labelling of bacterial cell lysates was therefore conducted with acetylated probes **14** and **24**. This experiment was initially performed with 1mM of probes for direct comparison with the intact cell labelling work previously described. However, this resulted in a large amount of un-bound probe making the imaging of the gel challenging. Therefore, as previously performed on *E. coli* lysates (Chapter 4), the labelling of *H. influenzae* lysates was performed with 0.5mM of probe.

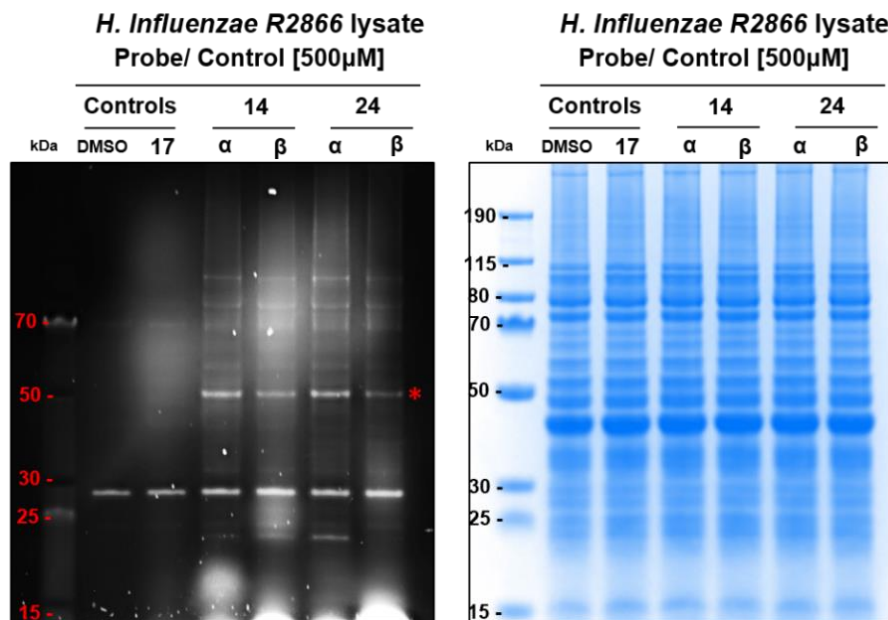


Figure 5.14. *H. influenzae* R2866 cell lysate protein labelling performed by all four probes. *General conditions:* The cells are grown to $OD_{600} = 0.9$ and lysed with bug buster (100 μ L, X1 in PBS, for pellet of 4mL of cell culture) and mixed for 30mins at rt. Resulting cell lysates (9 μ L) are incubated with **24- α/β** / **14- α/β** / controls (1 μ L, 5mM stock) for 1 hour at 30°C. See Fig 5.12 for further sample treatment. A protein labelled with a high intensity via this method is highlighted in red.

As observed for the whole cell work, a protein of approximately 28 kDa is intrinsically fluorescent while other proteins over a wide range of MW (20 – 100 kDa) are specifically labelled by the probes. However, once again, the observed labelling profiles are similar for all four probes. Additionally, it is worth noting that one band (~50kDa, highlighted in red in Figure 5.14) appeared significantly more fluorescent than the others, with a

comparable intensity to the intrinsically fluorescent protein. This is unlike what was observed during the whole cell work in which the intrinsically fluorescent protein was (by far) the most intense band.

Cell lysates were then incubated with decreasing concentration of glucosamine-based probe **14- α** / control **17** (1-hour incubation). While no background labelling was observed for the fluorophore control **17**, the labelling of proteins in bacterial cell lysates was concentration dependent, as observed for the intact cell work (Figure 5.15). The detection limit for tagged proteins under these conditions was $>100\mu\text{M}$ of probe. In this experiment, the presence of a strongly labelled protein in cell lysate (50 kDa) which seemed absent in intact cell labelling was also observed. We therefore conducted an experiment with both labelling techniques combined in a single gel for direct comparison.

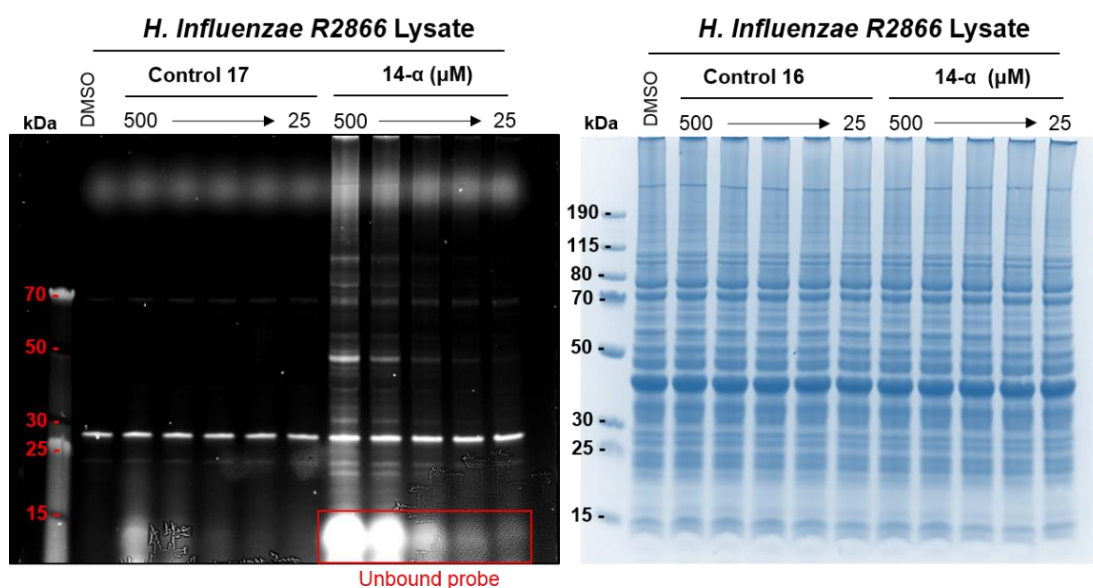


Figure 5.15. Effect of probe (**14- α**) concentration the labelling of *H. influenzae* R2866 protein in cell lysates. See Fig 5.14 for general conditions. Alterations: **14- α** (1 μL , 5 - 0.25mM stock).

5.5.4. Whole cell versus cell lysates labelling

For a direct comparison of the whole cell versus cell lysate labelling, an experiment in which both labelling protocols are conducted in parallel from the same cell cultures was performed. A bacterial culture was grown as previously described, half of which was used

for protein labelling through cells (1mM of **14**, 2 hours) while the other half was lysed and subsequently incubated with the glucosamine-based probe (0.5mM, 1 hour). The workflow of this experiment is shown in Figure 5.16.A.

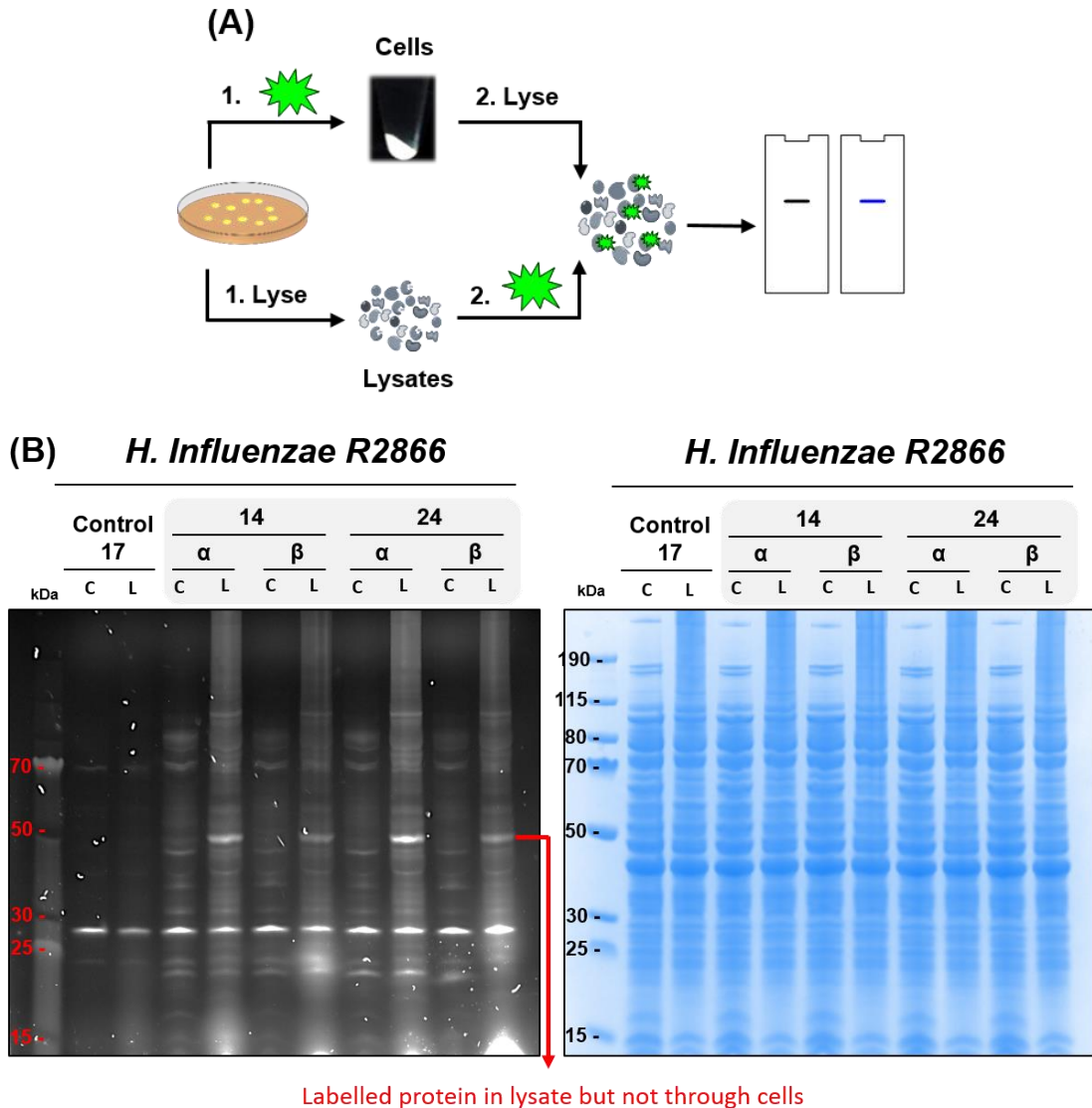


Figure 5.16. Direct comparison of intact *H. influenzae* R2866 cell labelling vs cell lysate labelling performed by all four probes. (A) Workflow. (B) Results. See Fig 5.12 and 5.14 for general conditions. C= intact cell labelling, L= cell lysate labelling.

The lysate samples resulting from both experiments were loaded on a single 4- 12% SDS-page page to enable the direct visual comparison of the labelling profiles. Interestingly, the labelling profile obtained *via* the two distinct protocols showed some significant differences, with some targets labelled exclusively though one strategy but not the other

(Figure 5.16.B). The most notable difference was observed for a target of MW approaching 50 kDa, labelled exclusively in bacterial cell lysate but not through intact *H. influenzae* cells (Figure 5.16.B), in line with previous observations.

A potential explanation for the above observations is that our probes penetrate exclusively in the periplasm environment of *H. influenzae* R2866 without crossing the inner membrane and reaching the cytoplasmic space. In an intact cell labelling experiment, only periplasmic proteins would therefore be labelled by the probes. However, when the cells are lysed, both periplasmic and cytoplasmic proteins are released, and both can be labelled in a cell lysate labelling experiment. Therefore, we hypothesise that the protein labelled exclusively in bacterial cell lysate may be a cytoplasmic protein. As a replacement for Bug Buster-induced cell lysis, a periplasmic extraction followed by cytoplasmic release was attempted with the aim of providing answers to our proposed hypothesis. Using a concentrated solution of sucrose in Tris buffer supplemented with EDTA²³³, the outer membrane extract was released as the soluble fraction. With repeated freeze and thaw cycles of the resulting cell pellet, the cytoplasmic cell content was subsequently released. However, the amount of proteins generated *via* this strategy was too low and could not be detected by gel electrophoresis. Therefore, the initial bacterial culture of *H. influenzae* R2866 was scaled up and the experiment was repeated with a larger cell pellet to address the problem. Although improved, the protein content of each fraction was still too low for conclusions to be drawn with confidence. This merits further optimisation.

5.5.1. Two-step labelling

Following this, a 2-step labelling experiment on both intact cells and cell lysates was performed with the aim of assessing the role of the sugar in target recognition and suppressing any interference related to the fluorophore scaffold. Acetylated probe

precursors equipped with an alkyne handle (**12- α/β** , **22- α/β**), were incubated with intact cells (2 hours, 1mM) or cell lysate (1 hour, 0.5mM). After incubation the cells were washed to remove unbound probe and lysed according to the protocol previously described. The resulting lysates were then submitted to *in situ* click ligation conditions in the presence of CuSO₄ sodium ascorbate and the Dansyl azide (cmpd **16**, 0.5mM) for 1 hours at 30°C, for the biorthogonal attachment of the fluorophore reporter group (Figure 5.17).

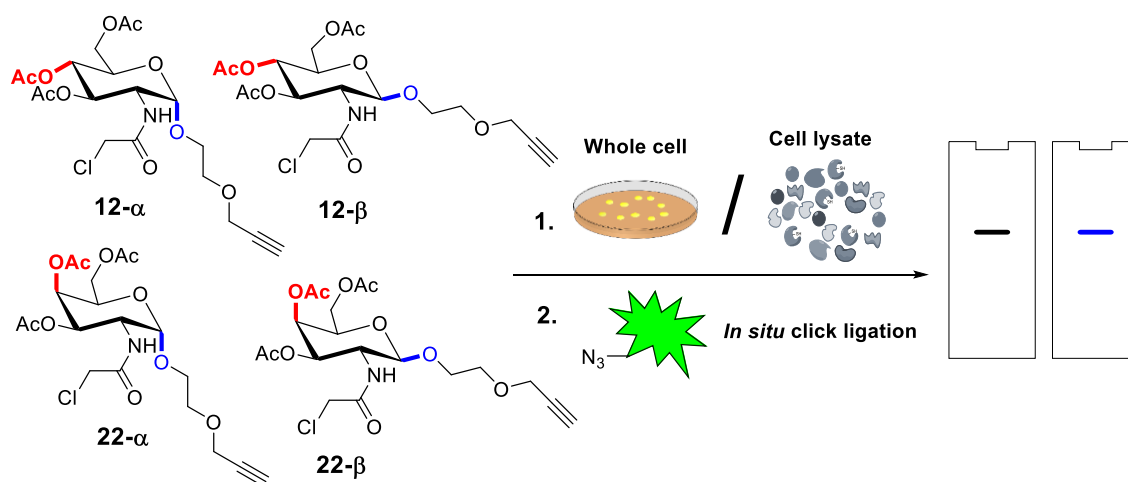


Figure 5.17. Two-step labelling protocol, workflow

No labelling was observed whether the precursor probe had been incubated on intact cells (see Appendix 5, Fig A5.3) or in cell lysates (not shown). Because the outcome was the same for both strategies, the absence of labelling could not be explained by an inability of the probe precursors to permeate the cell membrane. However, as the Coomassie revealed a gel with a much lower protein abundance than observed for the standard one-step labelling protocol, the above observation was initially thought to be the result of a dilution of the lysates following the *in situ* click ligation step (see Appendix 5, Fig A5.3). The two-step labelling protocol was then repeated with more initial cell pellet (8mL of cell culture). Additionally, *in situ* click ligation was performed using 10X more concentrated solutions of all reaction components to lower the subsequent dilution factor of the lysates. Although successful for the retrieval of protein abundance to levels comparable to the one-step protocol, this strategy did not have any effect on the outcome of the experiment as no

labelling was observed (see Appendix 5, Fig A5.4). This can be explained by two factors: (1) the *in situ* click ligation is unsuccessful (2) the probe precursors do not bind to their protein target.

(1) The former could be due to the reaction between the free alkyne and the azide not proceeding or the inaccessibility of the free alkyne following binding. The presence of DTT or related thiols can drastically inhibit the Click reaction and such components may be present in the commercial lysis buffer used. Therefore, the two-step labelling needs to be repeated cell on lysates obtained *via* a different method, for which the lysis buffer employed does not contain a reducing agent.

(2) The results gathered from previous experiments (one-step labelling) suggest that the modulation of the sugar scaffold has no significant effect on the observed labelling. This indicates that the sugar head does not act as the targeting motif for labelling which suggests that factor (2) may be responsible for the absence of labelling *via* a two-step protocol.

5.5.2. Mass spectrometry analysis

The above dataset suggests that the probes have the ability to label some specific *H. influenzae* proteins. However, the lack of clear differences observed in the labelling profile of all four probes combined with the absence of labelling observed in a two-step labelling protocol strongly suggests that the sugar head may not act as the recognition motif. On the contrary, the data indicates that the targeting group for protein recognition is a motif common among all four probes which is absent in a two-step labelling process: the Dansyl fluorophore. Should this hypothesis be correct, the nature of the tagged proteins may therefore be very different from anticipated, with very few to none being carbohydrate-binding enzymes.

In order to provide answers, seven labelled bands were extracted from the gel and submitted to mass spectrometry analysis at the King's CEMS Proteomics Facility (Figure 5.18). The bands highlighted in red in Figure 5.18 were not submitted for MS analysis because of cross-contamination. Shortly before submission of this thesis, the MS data was received. Therefore, the following section constitutes a preliminary analysis of the results.

The sequence of the peptides, obtained after trypsin digestion and LC-MS/MS analysis, were search against the current *H. influenzae* taxonomy reported on the Uniprot database. The raw data was searched at a stringency threshold of 5% false discovery rate for protein and peptide with a minimum of one peptide per protein. Any protein that is above this identity threshold is therefore deemed significant.

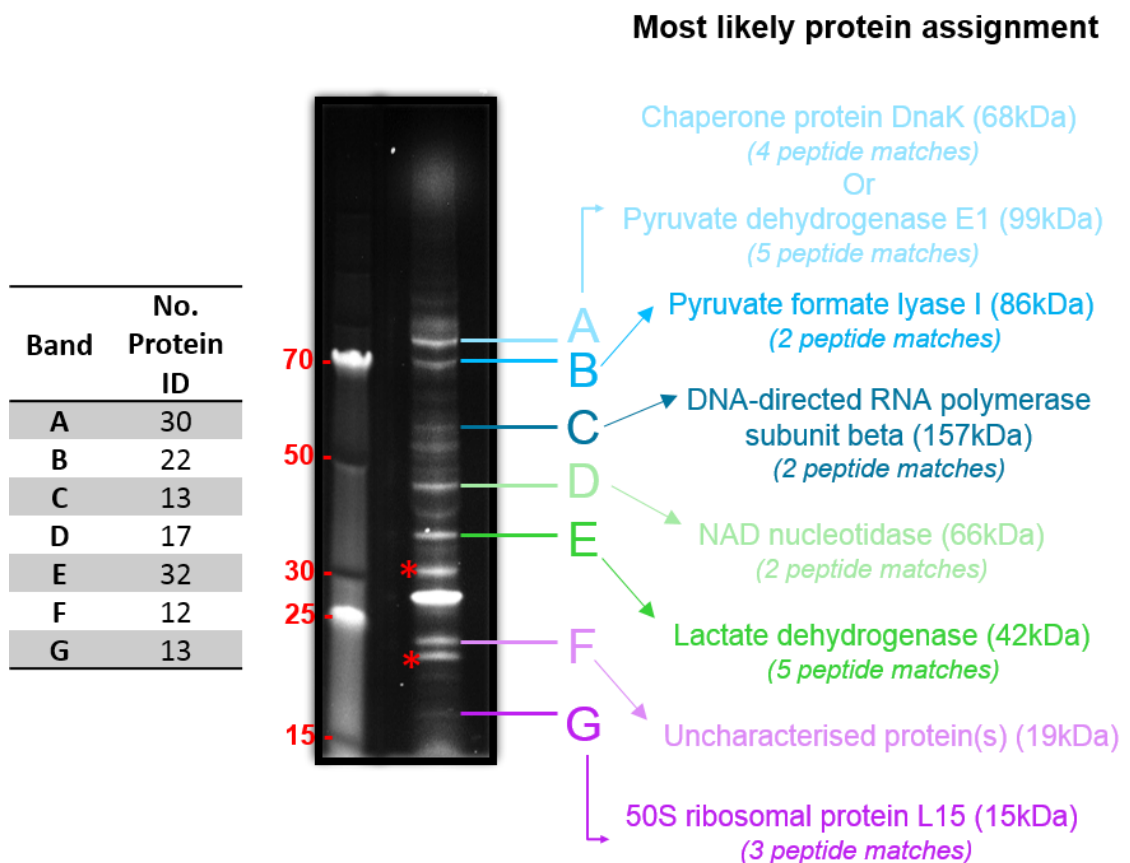


Figure 5.18. Fluorescent bands cut out and submitted to MS analysis for protein identification. Highlighted bands (red) were not submitted because of contamination. The gel was run for probe **14- α** (1mM), 2 hours incubation with *H. influenzae* cells (see chapter 7 for experimental detail)

After searching against the *HI* Taxonomy database, a total of 102 proteins were identified at this stringency, some with a poor peptide match (1-2) and some with a greater peptide match (4-5). For all seven bands, and accounting for all peptide matches, 139 proteins were identified. This means that 37 proteins have been identified as potential protein IDs for more than one band. Because of the great sensitivity of mass spectrometry, it is common to match more than one potential protein, however, in our case this was a rather large number of potential matches. This can be explained by the fact that the bands were cut out of the gel manually which is inevitably associated with a significant degree of contamination, especially when the amount of protein present is so abundant (see Coomassie stained gels above). Therefore, it is highly likely that one sample contains more than one protein. For each extracted band, we therefore focussed on the assignment which was associated with the greater number of peptide matches. The suggested protein ID for each band *via* this strategy is presented in Figure 5.18. For each band, the peptide match corresponded to 3 – 6% of protein sequence coverage and the position of the band on the gel in relation to the ladder corroborates with the MW of the proposed protein ID. Although in some case the observed band may be a fragment of the identified protein (band C for example), this is a first indication that the assignment may be robust. However, the subcellular location of identified proteins varies between cell inner membrane (for lactate dehydrogenase) and cytoplasm (for chaperone Dnak) and in other cases, is not reported on the Uniprot database. Therefore, robust conclusions regarding the subcellular location of the tagged proteins *via* intact cell labelling cannot be drawn at this stage. Additionally, these are preliminary results based on a single experiment and the band analysis should be repeated to check for consistency.

From these initial results, neither LgtC nor any relevant carbohydrate binding proteins have been identified as potential protein ID for the extracted bands. This is the case, not only for the proteins associated with the greater peptide matches, but for most of the 102 proteins identified. This observation is in line with the hypothesis made earlier regarding

the probability that the labelling may not be targeted by the sugar head group. However, it could also be partially due to the potential absence of membrane bound proteins analysed with the current labelling protocol. Among the 102 identified target proteins, only two are related to carbohydrate binding enzymes as they bind sugar phosphates: glucose-6-phosphate 1-dehydrogenase (56kDa) shares 1 peptide with extracted band B, and fructose-1,6 biphosphatase (36kDa) also has 1 peptide match with extracted band E. Interestingly, a peptidoglycan associated protein (16kDa) was identified has a potential candidate for extracted band G, as the result of 1 peptide match. This is an outer membrane-bound protein which may suggest that, unlike originally thought, some membrane bound proteins may be extracted with Bug buster.

On the other hand, many nucleotide-binding proteins have been identified. Among others, are the proteins assigned for three of the extracted bands: the chaperone protein DnaK identified for band A (Uniprot, P43736 (DNAK_HAEIN)), NAD nucleotidase for band D (Uniprot, Q4QNY4_HAEI8) and DNA-directed RNA polymerase subunit beta (a nucleotidyl-transferase) for band C. This indicates that the recognition motif, believed to be the fluorophore, may be a good nucleotide mimic. In chapter 4, it was concluded that the probe binds in the active site of LgtC with the fluorophore occupying the binding site of the donor substrate: UDP-Gal, a sugar nucleotide. Although these results do not support our original hypothesis for the labelling of carbohydrate-binding proteins, they are highly interesting as they corroborate previous observations regarding the binding mode of the probe.

The band previously labelled in *HI* cell lysate exclusively (see Figure 5.16) was also analysed by LC-MS/MS following the same protocol as the intact cell bands. Interestingly, for this specific band, the number of peptide matches with suggested target proteins was greater than for the intact cell work. We hypothesised that due to the more intense labelling of this band, less error was associated with the band extraction process. 11

peptides detected in the sample matched the elongation factor Tu (thermo unstable) (43kDa) which corresponded to 35% sequence coverage of the whole protein. This was about 10-fold greater than the coverage obtained for any band isolated from the intact cell work. Cell division protein FtsZ was also identified as a potential target in this sample with 11 peptide matches corresponding to 20% of whole sequence coverage. Both proteins are nucleotide binding enzymes (recognising GTP) and both are located in the cytoplasm of *HI*. This suggests that, as hypothesised, the protein labelled exclusively in cell lysate is a cytoplasmic protein. However, because of the absence of robust data regarding the subcellular localisation of the proteins tagged during intact cell labelling, no definite conclusion regarding the extent of the probe's cell penetration can yet be drawn.

At this stage, it is important to note that the above discussion is based on preliminary data, obtained in the absence of protein enrichment. This, combined with the fact that the whole bacterial proteome is present on the gel and the significant human error associated with cutting out the bands, indicate that the resulting MS data may just correspond to abundant proteins. Although it is interesting to note a pattern in the pull-out of nucleotide binding proteins, it remains relatively unlikely that all low abundance proteins labelled by the probe correspond to the above identified targets (Figure 5.18). In order to improve the reliability of this analysis, a protein pull-down based on the Dansyl scaffold could be employed *via* the use of anti-Dansyl antibody available commercially (A-6398 Thermo Fisher Scientific).

5.6. Summary and conclusions

In this chapter, carbohydrate-based fluorescent probes equipped with an electrophilic warhead were applied to a chemical proteomic study on non-typeable *H. influenzae* R2866 strain. The aim of the study was to assess the ability of the probes to recognise and covalently label carbohydrate binding proteins in pathogenic bacteria. In order to gain

insight into the role of the sugar head for recognition and to exploit the wide diversity of sugar scaffolds utilised by bacteria, a series of galactosamine-based probes was synthesised in addition to the glucosamine-based probes described in Chapter 4. Their synthesis was directly based on the route developed in the previous chapter with an additional three steps to generate the non-commercially available peracetylated galactosamine hydrochloride salt. Probe **24** was obtained with a 26% yield over 6 steps (14% for **24- α** and 12% for **24- β**), and NMR analysis enabled the unequivocal identification of both anomers.

The four fluorescent probes (**14- α/β** and **24- α/β**) were tested for their labelling ability in two types of experiments: (1) an intact cell labelling approach during which live cells were incubated with the probes and subsequently lysed. (2) a cell lysate labelling strategy during which the cells were lysed prior to incubation with fluorescent probes. The resulting cell lysates from both methods were separated on an SDS-page gel and visualised under fluorescence read-out. The abundance of proteins was detected by Coomassie staining. For both strategies, the probes specifically labelled bacterial proteins in a concentration dependent manner. Interestingly, for a single probe, differences were observed in the labelling profiles resulting from both strategies. We hypothesised that this may be the result of the probe not penetrating the cytoplasm in the whole cell labelling experiment, in which case the labelling of cytoplasmic proteins is achieved exclusively in a cell lysate labelling experiment. On the other hand, for one labelling strategy, the observed labelling profile for all four probes is not significantly different, which suggests that protein recognition is initiated by a motif common to all probes (fluorophore/linker group). The absence of fluorescent bands on fluorophore control lanes, indicates that the sugar head, bearing the electrophilic moiety, is responsible for protein modification. However, the above observations suggest that it has a minimal role in protein recognition. This hypothesis was corroborated by the lack of tagged proteins as a result of a two-step labelling experiment, indicating that in the absence of the

fluorophore, the probe does not bind to proteins. However, at this stage, care should be taken not to over-interpret this result as the presence of reducing agent in lysis buffer may drastically inhibit the Click reaction.

Subsequent preliminary proteomics MS analysis indicated that the tagged proteins may not be carbohydrate binding proteins. However, a significant number of nucleotide-binding proteins were identified as potential target candidates which suggests that the Dansyl motif may be a good nucleotide mimic. Although this is in line with observations made in Chapter 4 (Dansyl moiety of the probe occupying the UDP-Gal binding pocket of LgtC), this preliminary MS data should not be over interpreted as, in the absence of protein enrichment, the identified targets may correspond the most abundant proteins in the sample.

Chapter 6

Conclusions and outlook

Antibiotic resistance is one of the greatest current medical challenges. Towards developing novel strategies to tackle this major issue, the development of molecular tools to study biological targets involved in virulence is of great interest. In bacteria glycosyltransferases (GTs) are responsible for the biosynthesis and alterations of surface glycoconjugates involved in bacterial adhesion to host's cells and primary infection mechanisms. In this thesis, we have developed small chemical tools in the form of inhibitors and fluorescent probes for the study of LgtC, a bacterial galactosyltransferase involved in virulence and infection in pathogenic Gram-negative bacteria. Inspired by previous findings suggesting the presence of an accessible cysteine residue in the active site of LgtC, and willing to exploit the great potential of resurging covalent modifiers, our strategy was to combine an electrophilic warhead and a carbohydrate scaffold for the design of small molecular tools for LgtC and other carbohydrate binding enzymes. Interestingly, except for the recent example of the epoxide-containing galactose developed by Titz *et al.* for the study of bacterial lectins, this strategy has not been commonly explored. The challenges we faced during this research may provide an explanation as to why that may be the case.

In Chapter 2, we designed disaccharide lactosamine-based molecules, equipped with an electrophilic warhead at a specific position of the scaffold, for the targeting of cys246 in the active site of LgtC. We found that, beyond showing inhibitory activity, this class of inhibitors also exhibited some residual substrate properties. The biochemical assay we use to perform inhibition studies relies on the detection of the secondary product of the LgtC reaction, which makes the dissociation between both activities impossible in that case. However, using an assay which detects the formation of primary enzymatic product or the depletion of the substrate would enable a better characterisation of this disaccharide series. Although it makes them great synthetic tools for the biosynthesis of complex oligosaccharides, the reported substrate promiscuity of LgtC and other bacterial GTs renders the rational design of carbohydrate based molecular tools challenging.

In Chapter 3, based on the reported weak substrate activity of monosaccharides for LgtC, we shifted our focus to the development of glucosamine-based inhibitors. As these inhibitors have no residual substrate properties, they were useful tools to investigate the putative covalent mode of action of this series of monosaccharides. It was with surprise that a non-covalent inhibition mode was elucidated for these small inhibitors. Initial substrate competition experiments indicated a non-competitive inhibition for both of LgtC's substrates which suggested an unprecedented allosteric mode of inhibition for the electrophilic monosaccharides. This may explain the absence of covalent inhibition for lack of nucleophilic residue in the vicinity.

In Chapter 4, a fluorescent probe (**15**) derived from the scaffold of the most potent monosaccharide inhibitor (**11**) was synthesised. Robust data suggested that, unlike inhibitor **11**, the probe has the ability to covalently label recombinant LgtC. In addition to this, the observed absence of competitive binding between the probe and the inhibitor combined with opposing activity towards a LgtC mutant strongly indicate a different binding mode for these two molecules. Because our results indicate with confidence that the probe binds in the active site of the enzyme, the data gathered in chapter 4 support the hypothesis suggested in Chapter 3 regarding an allosteric binding mode for the inhibitor. Allosteric binding of a small sugar scaffold has never been reported before for LgtC and to the best of my knowledge neither has it been for other bacterial GTs. To date, this remains unproven as a greater evidence base is required to conclude with confidence that **11** is an allosteric inhibitor of LgtC. However, the dataset obtained during this research supports this mode of action in its entirety. Crystallography work has been initiated with the aim of solving the structure of LgtC in complex with inhibitor **11** to provide a definite answer regarding the binding of the monosaccharide in the enzyme. Some promising conditions have been identified (see Table 3.2) which merit further investigation.

In Chapter 4 and 5, fluorescent carbohydrate-based covalent probes have been synthesised and applied to the labelling of recombinant LgtC and other proteins in non-pathogenic and pathogenic bacteria. In both cases, investigation into the probes' mode of action indicates that the sugar head, originally designed as the targeting motif for the labelling of sugar binding proteins, plays little to no role in protein recognition. On the contrary, our data suggest with a high degree of confidence that the target recognition is controlled by the fluorophore moiety which is believed to act a nucleotide mimic. This suggestion is based on two findings: (1) for the labelling of recombinant LgtC, the fluorophore binds in the binding pocket of the donor substrate: a sugar nucleotide (2) preliminary MS analysis suggests a number of nucleotide-binding proteins as the potential identity of the target proteins labelled in pathogenic bacteria. To the best of my knowledge, the Dansyl fluorophore scaffold has never been reported as a nucleotide mimic before.

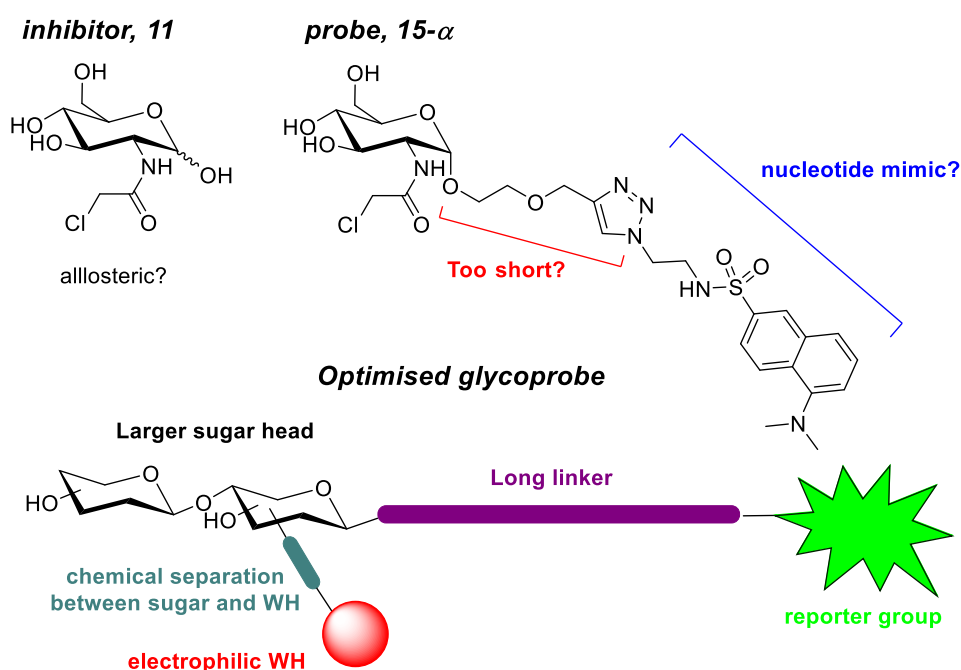


Figure 6.1. The research undertaken during this PhD: conclusions, hypothesis

We hypothesise that the lack of recognition motif properties observed for the sugar head is partially due to an unsuitable chemical linker between the sugar and the fluorophore.

After reflection on the challenges faced during this research, it is our belief that the linker is too short in length to provide enough spatial separation between the two entities. Therefore, the putative nucleotide mimic properties of the larger Dansyl motif take predominance over the small sugar group. Structural optimisation of the current probe with a larger carbohydrate scaffold and a longer linker may enable the development of highly efficient probes for the labelling of carbohydrate binding proteins in bacteria.

Finally, our dataset raises important questions regarding the design of carbohydrate-based covalent inhibitors or probes. During this research, it has become clear that an electrophilic warhead attached directly onto a carbohydrate scaffold may not be an ideal strategy for the design of covalent inhibitors and probes. Although it has been reported once in the literature¹¹⁶, our research has provided evidence of partial intramolecular quenching of WH reactivity by the electron donating effect of a free sugar scaffold. This is supported by the fact that the acetylated sugar probes developed in this work were significantly more reactive than the free sugar analogues. We therefore suggest that for the design of covalent glycoprobes and inhibitors, sufficient chemical separation should be provided between the sugar recognition motif and the electrophilic WH to ensure full reactivity towards nucleophilic residues.

A number of areas for further investigation have arisen, in order to provide greater insight into some of the key questions which emerged from this work. These include:

- The development of a MS-based biochemical assay for the detection of the trisaccharide primary product of the LgtC reaction. This would allow a thorough investigation into the mode of action of the disaccharide candidates **1** and **2** and separation of their substrate from inhibitory activities, which is not possible with the current phosphatase-coupled colorimetric assay.
- The elucidation of the structure of LgtC in complex with monosaccharide inhibitor **11** by X-ray crystallography *via* the optimisation of the crystallisation conditions

identified herein. This will provide irrefutable evidence regarding the binding mode of the inhibitor which, with the current dataset, is proposed to be allosteric.

- The investigation into the use of longer PEG chemical spacers between: (1) the sugar and the electrophilic WH in order to test the impact on WH reactivity, (2) the sugar and the reporter group in order to test effect on binding and the sugar's ability to act as a targeting group.
- Confirmation of the Dansyl group's nucleotide mimicking properties by chemical attachment of an electrophilic WH on the fluorophore's scaffold and the use of anti-Danzyl antibody for subsequent enrichment.
- Synthesis of a series of probe analogues incorporating alternative reporter groups in order to assess the effects of removing the putative nucleotide mimic on binding. This may include a variety of fluorophores such as coumarin or fluorescein, or a different detection strategy such as the use of biotin which allows subsequent enrichment and therefore more accurate identification of target proteins.

In conclusion, the work carried during this PhD has answered different but equally interesting questions than the ones initially posed. The challenges encountered may explain why carbohydrate-based covalent probes remain rarely used. They are synthetically challenging, may not exhibit the expected target binding and may possess a very low reactivity towards nucleophiles. However, this research provides a deep enough understanding of these issues to propose with confidence a viable potential methodology for the efficient design of glycoprobes (Figure 6.1). Such optimised probes may be ideal tools to enable the identification of new virulence factors for the development of novel antimicrobial strategies.

Chapter 7

Material and methods

7.1. Covalent docking

CovalentDock Cloud program was used in order to carry out covalent docking studies. <http://docking.sce.ntu.edu.sg/>. .mrv file were generated using MarvinSketch 16.7.25.0 and converted to .mol2 file using Open Babel 2.3.2. The protein and ligand complex (.dlg files) were visualised using PyMol Molecular Graphic System Version 1.7.4.5 Edu.

LgtC 1GA8 was retrieved from the Protein Data Bank and modified in WordPad for the elimination of all non-ATOM character. On PyMol makes all hydrogen atoms explicit.

The ligand was drawn in 3D with explicit hydrogen atom on MarvinSketch and then uploaded onto the CovalentDock Cloud program as a .mol2 file.

7.2. Chemical synthesis: protocols and compound characterisation

General. All chemical reagents were purchased from commercial source and were used as received, with exception of propynol ethoxylate which was distilled prior to use. Normal phase chromatography was performed on silica gel (particle size 40 – 63 μm). Reverse phase chromatography was performed on Gradifrac BIO-RAD BioLogic LP model 2110 or preparative HPLC Agilent Technology 1260 Infinity fitted with ZORBAX eclipse XDB-C18 PrepHT column (21.2 cm x 100 mm, particle size 5 μm). Semi-preparative chromatography was performed using an Agilent 1260 Infinity machine fitted with a ZORBAX 300SB-C18 column (25 cm x 9.4 mm, particle size 5 μm). LC-MS analysis was performed on Agilent eclipse XDB-C8 column (4.6 x 150 mm, 5 μm) on Agilent Technologies 1260 Infinity II LC system coupled with expression^L Advion MS system. Solvents used for reverse phase purification were all HPLC grade. Thin layer chromatography (TLC) was performed on precoated aluminium plates (Silica Gel 60 F254, Merck). Compounds were visualised by TLC stain *p*-anisaldehyde and/or exposure to UV light (365 nm). All compounds are characterized by ¹H-NMR, ¹³C-NMR, and (HR)MS. ¹H-NMR spectra were recorded at 298K using a Bruker Ascend™ 400 spectrometer at 400

MHz. ^{13}C -NMR spectra were recorded at 298K on either a Bruker Ascend™ 400 spectrometer at 100 MHz or Bruker Ascend™ 600 spectrometer at 150 MHz. Chemical shifts (δ) are reported in ppm, referenced to residual solvent peaks. Full peak assignment was achieved with the aid of 2D NMR, namely COSY, HSQC, HMBC and DEPT (also generated on spectrometers detailed prior). HRMS spectra were recorded at the EPSRC National Mass Spectrometry Service Centre, Swansea. In house, reported masses were obtained using an Advion expression^L CMS mass spectrometer, in combination with an Agilent 1260 Infinity LC system fitted with a ZORBAX Eclipse XDB-C8 column (15.0 cm x 4.6 mm, particle size 5 μm) (0.1% formic acid in water/methanol, flow rate: 1 mL/min, detection wavelength: 210/254 nm).

7.2.1. Chapter 2 compounds: disaccharides

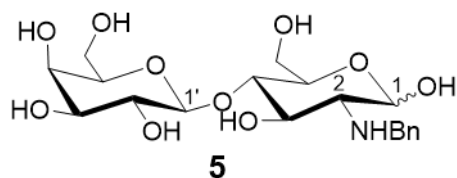
1-N-benzyl-1-deoxy-lactulose (3)

Lactulose (10 g) was reacted with an excess of benzylamine (13 mL). The reaction mixture was stirred at 40°C for 30 hours. Portions of cold acetone (10 mL) were added to the residue and the syrup-like precipitate formed was filtered. The filtrate was concentrated under vacuum and cold diethyl ether was added to the residue. The white crystalline precipitate was filtered and combined with the first syrup-like precipitate (by dissolving in methanol). The residue was taken as is through the next stage.

2-lactosamine-N-benzyl (5)

3 previously obtained was dissolved in 40 mL of MeOH and 4 mL of AcOH were added to the reaction mixture. The solution was stirred at room temperature for 2 hours. **5** was separated from the bis-benzylated by-product **4** and remaining lactulose by normal phase chromatography (gradient 0%-25% MeOH in DCM). **5** was isolated as a mixture of anomers. %Yield= 51%.

¹H NMR (400 MHz, D₂O) δ 2.35 – 2.39 (m, 0.7H, H_{2β}), 2.51 (dd, J_{2α-3α}=10.2 Hz, J_{2α-1α}= 3.5 Hz, 1H, H_{2α}), 3.27 – 3.97 (m, 20.7H, H_{α+β}), 4.20 – 4.23 (m,

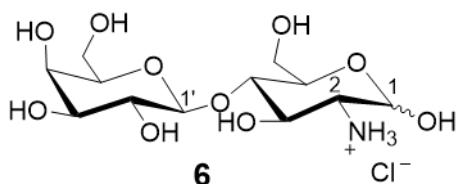


1.7H, H_{1β}), 4.51 (d, J_{1β-2β}= 8.1 Hz, 0.7H, H_{1β}), 5.08 (d, J_{1α-2α}= 3.4 Hz, 1H, H_{1α}), 7.08 – 7.13 (m, 1.7H, Bn), 7.15 – 7.26 (m, 7.3H, Bn). **HR-MS** m/z= 432.1866 [M+H]⁺ (calculated 432.1864)

2-lactosamine ammonium chloride (**6**)

5 (600 mg) was dissolved in water (10 mL) and dioxane (5 mL). Pearlmans's catalyst (0.3 eq) was added to the reaction mixture alongside a drop of concentrated HCl. The reaction mixture was set up on the hydrogenation apparatus stirred at room temperature under 5 bars of hydrogen for 24 hours. After evaporation of the solvent, the residue was purified by reverse phase chromatography (gradient 2%-10% MeOH in water). %Yield= 90 %. **6** was isolated as a mixture of anomers.

¹H NMR (400 MHz, D₂O) δ 3.01 (dd, J_{2β-3β}= 10.7 Hz, J_{2β-1β} = 8.3 Hz, 0.5H, H_{2β}), 3.29 (d, J_{2α-1α}= 3.6 Hz, 0.5H, H_{2α}), 3.49 – 4.03 (m, 16H, H_{α+β}), 4.42 – 4.45



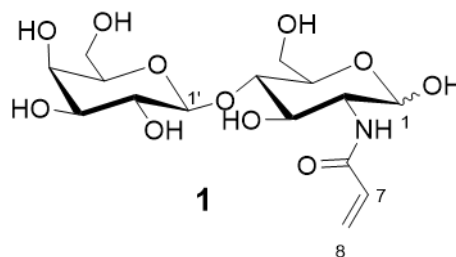
(m, 1.5H, H_{1β}), 4.94 (d, J_{1β-2β}= 8.4 Hz, 0.5H, H_{1β}), 5.41 (d, J_{1α-2α}= 3.6 Hz, 0.5H, H_{1α}). **HR-MS** m/z= 342.1395 [M+H]⁺ (calculated 342.1395)

2-deoxy-N-acryloyl-2-lactosamide (**1**) 2-deoxy-2-acetylchloridelactosamide (**2**)

6 was dissolved into MeOH (2mL), TEA (2 eq) was added and the reaction mixture was stirred for 2 mins before being cooled down to 0°C. The desired acid anhydride (3eq) was added to the reaction mixture at 0°C and which was stirred until completion (monitored by TLC, 20% MeOH in DCM). The reaction mixture was the passed through activated Amberlite IR- 120 (H⁺) resin (cation exchange column), the desired product was recovered while residual **6** and triethylammonium salt remained attached on the resin. The resin was washed two time with MeOH. The product was then purified by reversed

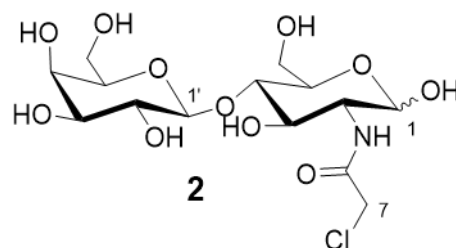
phase preparative HPLC (C3 column, H₂O, MeOH gradient) and solvent was evaporated on freeze dryer to yield a white powder. Yield 20-25%.

¹³C NMR (100MHz, D₂O) 168.8 (C=O_β), 168.7 (C=O_α), 129.8, 129.5, 128.0, 127.9 (C_{7α}, C_{7β}, C_{8α}, C_{8β}) 102.9 (C_{1'α}), 102.8 (C_{1'β}), 94.8, 90.5 (C_{1α} C_{1β}), 78.6, 78.2, 75.3, 74.8, 72.4, 72.4, 70.9, 70.9, 70.2, 70.2,



69.2, 69.2, 68.5, 68.5, 61.0, 61.0, 59.8, 56.2, 53.7, 53.7. **HR-MS** m/z= 394.1353 [M-H]⁻ (calculated 394.1355)

¹³C NMR (100MHz, D₂O) 170.2 (C=O_β), 169.9 (C=O_α), 102.8, 102.8 (C_{1'α}, C_{1'β}), 94.5 (C_{1β}), 90.3 (C_{1α}) 78.6, 78.2, 75.3, 74.8, 72.4, 72.2, 70.9, 70.7, 70.7, 70.2, 69.2, 69.2, 68.5, 68.5, 61.0, 60.9, 59.9,



59.8, 56.6, 54.0, 42.3 (C_{7α}), 42.2 (C_{7β}). **HR-MS** m/z= 416.0969 [M-H]⁻ (calculated 416.0965)

7.2.2. Chapter 3 compounds: monosaccharides

1,3,4,6-Tetra-O-acetyl-2-deoxy-N-acryloyl-2-glucosamide (**8**) *1,3,4,6-Tetra-O-acetyl-2-deoxy-N-acetylchloride-2-glucosamide* (**9**) *(3R,4R,5S,6R)-6-(acetoxymethyl)-3-propionamidotetrahydro-2H-pyran-2,4,5-triyl triacetate* (**8s**)

To a suspension of 1,3,4,6-Tetra-O-acetyl-2-amino-2-deoxy-beta-D-glucopyranose hydrochloride (0.4 mmol) in DCM (2 mL), triethylamine (2 eq) was added. The reaction mixture was allowed to stir at room temperature until all the solid is dissolved (1-2 mins). The solution was cooled down to 0°C and acryloyl chloride/ chloroacetyl chloride (3 eq) was added dropwise to the reaction mixture. The reaction mixture was stirred at 0°C for 10 mins and at room temperature until TLC showed full conversion (1hr). After completion, the reaction mixture was quenched by addition of saturated NaHCO₃ solution

(5 mL). DCM (10 mL) was added to the reaction mixture and the two layers were separated. The aqueous phase was washed with DCM (3 x 10 mL). Organic layers were combined and dried over Na₂SO₄. After filtration the mixture was concentrated under vacuum. The residue was purified by normal phase column chromatography (EtOAc/Hexane). The pure product was yielded as an off-white crystalline product (%yield= 70%-90%).

¹H NMR (400 MHz, CDCl₃ + 1 drop of DMSO-d₆) δ 1.88, 1.92, 1.97,

1.98 (4s, 12H, 4 CH₃CO), 3.74 (d, J= 8.1 Hz, 1H, H₅), 4.00 (d, J₆₋₆=

12.3Hz, 1H, H₆), 4.17 (dd, J₆₋₆= 12.3 Hz, J₆₋₅= 4.0 Hz, 1H, H₆), 4.20

- 4.32 (m, 1H, H₂), 4.99 (t, J₄₋₃= 9.6 Hz, 1H, H₄), 5.15 (t, J₃₋₄= 9.9 Hz,

1H, H₃), 5.51 (d, J₉₋₇= 10.3 Hz, 1H, H₉'), 5.66 (d, J₁₋₂= 8.7 Hz, 1H,

H₁), 5.96 (dd, J₇₋₉= 17.0 Hz, J₇₋₉'= 10.30 Hz, 1H, H₇), 6.13 (d, J₉₋₇= 17.1 Hz, 1H, H₉), 7.34 (d,

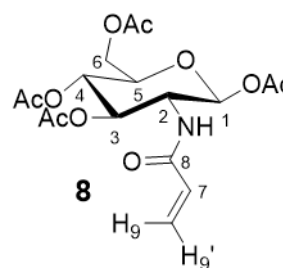
1H, J_{NH-2}=9.4 Hz, NH_{amide}). **¹³C NMR** (100 MHz, CDCl₃) δ 20.5, 20.5, 20.6, 20.8 (4C, CH₃CO),

52.3 (C₂), 61.6 (C₆), 68.1 (C₄), 72.5, 72.6 (C₅, C₃), 92.4 (C₁), 126.6 (C₉), 130.6 (C₇), 165.7 (C₈),

169.3, 169.3, 170.5, 170.5 (4C, CH₃CO). **HR-MS** m/z= 402.1395 [M+H]⁺ (calculated

402.1395), m/z= 419.1659 [M+NH₄]⁺ (calculated 419.1660), m/z= 424.1211 [M+Na]⁺

(calculated 424.1214)



¹H NMR (400 MHz, CDCl₃) δ 2.04, 2.05, 2.09, 2.12 (4s, 12H,

CH₃CO), 3.84 (ddd, J₅₋₄= 9.8 Hz, J₅₋₆= 4.6 Hz, J₅₋₆'=2.3 Hz, 1H, H₅),

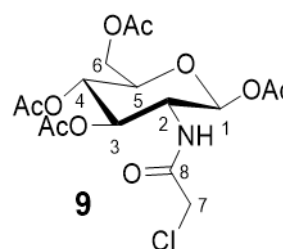
3.98 (d, J= 0.9 Hz, 1H, H₇), 4.13 (dd, J₆₋₅= 2.3 Hz, J₆₋₆=12.5 Hz, 1H,

H₆), 4.22 (dt, J₂₋₃= 10.4 Hz, J_{2-NH}= 9.0, 1H, H₂), 4.29 (dd, 1H, J₆₋₆=

12.5 Hz, J₆₋₅= 4.6 Hz, 1H, H₆), 5.15 (t, J= 9.5 Hz, 1H, H₄), 5.27 (dd, J₃₋₂= 10.4 Hz, J₃₋₄= 9.3 Hz,

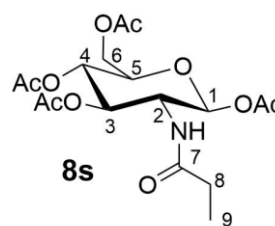
1H, H₃), 5.81 (d, J₁₋₂= 8.6 Hz, 1H, H₁), 6.57 (d, J_{NH-2}= 9.2 Hz, 1H, NH). **¹³C NMR** (100 MHz,

CD₃Cl₃) δ 20.6, 20.7, 20.9, 21.0 (4C, CH₃CO), 42.4 (C₇), 53.7 (C₂), 61.7 (C₆), 67.8 (C₄), 72.0



(C₃), 73.1 (C₅), 92.3 (C₁), 166.6 (C₈), 169.4, 169.5, 170.8, 171.0 (4C, CH₃CO). **HR-MS** m/z= 441.1270 [M+NH₄]⁺ (calculated 441.1270)

¹H NMR (400 MHz, CDCl₃) δ 1.09 (t, J₉₋₈= 7.6 Hz, 3H, H₉), 2.03, 2.03, 2.08, 2.10 (4s, 12H, CH₃CO), 2.13 (q, J₈₋₉= 7.6 Hz, 2H, H₈), 3.80 (ddd, J= 12.9 Hz, 7.7 Hz, 5.4 Hz, 1H, H₅), 4.09 – 4.15 (m, 1H, H₆), 4.26 (dd, J= 12.5 Hz, 4.5 Hz, 1H, H₆), 4.32 (dt, J= 14.1 Hz, 4.5

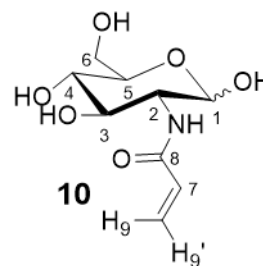


Hz, 1H, H₂), 5.09 – 5.19 (m, 2H, H₄, H₃), 5.62 (d, J_{NH-2}= 9.5 Hz, 1H, NH), 5.69 (d, J₁₋₂= 8.8 Hz, 1H, H₁). **¹³C NMR** (100 MHz, CD₃Cl₃) δ 9.96 (C₉), 20.7, 20.8, 20.9, 21.0 (4C, CH₃CO), 29.9 (C₈), 53.0 (C₂), 61.8 (C₆), 67.9, 72.7, 73.1 (C₃, C₄, C₅), 92.8 (C₁), 169.4, 169.7, 170.8, 170.4, (4C, CH₃CO), 174.0 (C₇).

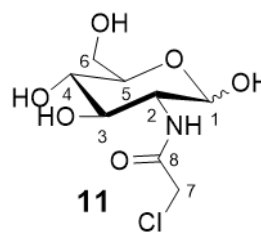
2-deoxy-N-acryloyl-2-glucosamide (**10**) *2-deoxy-N-acetylchloride-2-glucosamide* (**11**) *2-deoxy-N-ethyl-2-glucosamide* (**10s**)

To a solution of **8/9** (0.28 mmol) in MeOH (4 mL), sodium methoxide in MeOH (0.5 M) (4 eq) was added at 0 °C. The reaction was stirred at 0 °C for 10 mins and quenched with Amberlite IR- 120 (H+). Amberlite was removed through filtration, and the filtrate was concentrated in vacuo. The the fully deprotected compound was purified by reversed phase preparative HPLC (C3 column, H₂O, MeOH gradient) and solvent was evaporated on lyophilisator to yield a white powder in quantitative yield.

¹H NMR (400 MHz, D₂O) δ 3.40 - 3.59 (m, 2H_{α+β}), 3.68-3.91 (m, 6H_{α+β}), 3.95 (dd, J₂₋₃= 10.7 Hz, J₂₋₁= 3.5 Hz, 1H, H_{2α}), 4.74 (d, J₁₋₂= 8.4 Hz, H_{1β}), 5.21 (d, J₁₋₂= 3.5 Hz, H_{1α}), 5.78 (dd, J_{9'-7}= 10.0 Hz, J= 1.4 Hz, H_{9'}), 6.15 - 6.36 (m, H₇, H_{8'}_{α+β}). **¹³C NMR** (100 MHz, D₂O) δ 54.2 (C_{2α}), 56.8 (C_{2β}), 60.6 (C_{6α}), 60.7 (C_{6β}), 69.8, 70.0, 70.7, 71.6, 73.9, 75.9, 90.8 (C_{1α}), 94.9 (C_{1β}), 127.9 (C_{9β}), 128.0 (C_{9α}), 129.6 (C_{7α}), 129.9 (C_{7β}), 168.8 (C₈), 169.1 (C₈). **HR-MS** m/z= 234.0972 [M+H]⁺ (calculated 234.0972)

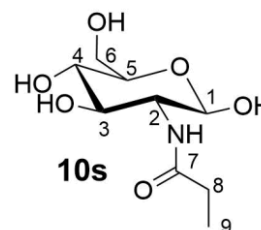


¹H NMR (400 MHz, MeOD) δ 3.33 - 3.37 (m, 1.5H $\alpha+\beta$), 3.47 - 3.51 (m, 0.5H β), 3.58 - 3.86 (m, 7H $\alpha+\beta$), 4.01-4.15 (m, 3H, H₇ $\alpha+\beta$), 4.63 (d, J₁₋₂=8.3 Hz, H_{1 β}), 5.08 (d, J₁₋₂=3.4 Hz, H_{1 α}). **¹³C NMR** (100 MHz, MeOD) δ 43.3 (C_{7 α}), 43.4 (C_{7 β}), 56.1, 59.1, 62.7, 62.8, 72.1, 72.3,



72.7, 73.1, 75.6, 78.0 (C_{2 α - β} , C_{6 α - β} , C_{4 α - β} , C_{3 α - β} , C_{5 α - β}), 92.4 (C_{1 α}), 96.6 (C_{1 β}), 169.4 (C₈), 169.9 (C₈). **HR-MS** m/z=256.0583 [M+H]⁺ (calculated 256.0582)

¹H NMR (400 MHz, D₂O) δ 1.13 (td, J= 7.6 Hz, 1.7 Hz, 4.7H, H_{9 $\alpha+\beta$}), 2.32 (qd, J= 7.6 Hz, 2.9 Hz, 3.2H, H_{8 $\alpha+\beta$}), 3.43 - 3.58 (m, 2.8 H $\alpha+\beta$), 3.69 (dd, J_{2-NH}= 10.1 Hz, J₂₋₁= 8.5 Hz, 0.6H, H_{2 β}), 3.72 - 3.98 (m, 6.7H, H $\alpha+\beta$), 4.72 (d, J₁₋₂= 8.4 Hz, H_{1 β}), 5.20 (d, J₁₋₂= 3.5 Hz, H_{1 α}). **¹³C**



NMR (100 MHz, D₂O) δ 9.43 (C_{9 β}), 9.48 (C_{9 α}), 29.0 (C_{8 α}), 29.3 (C_{8 β}), 53.9 (C_{2 α}), 56.6 (C_{2 β}), 60.6 (C_{6 α}), 60.7 (C_{6 β}), 69.8, 70.0, 70.6, 71.5, 73.8, 75.9, (C₃, C₄, C₅ $\alpha+\beta$) 90.8 (C_{1 α}), 94.9 (C_{1 β}), 178.5 (C₈), 178.7 (C₇).

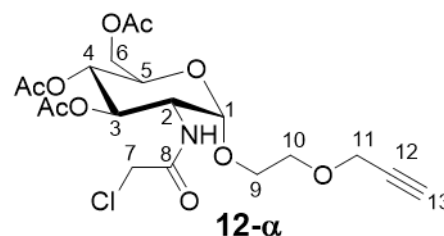
7.2.3. Chapter 4 compounds: glucosamine-based probes and fluorophores

(2R,3S,4R,5R,6S)-2-(acetoxymethyl)-5-(2-chloroacetamido)-6-(2-(prop-2-yn-1-yloxy)ethoxy) tetrahydro-2H-pyran-3,4-diyl diacetate (**12- α**) (2R,3S,4R,5R,6R)-2-(acetoxymethyl)-5-(2-chloroacetamido)-6-(2-(prop-2-yn-1-yloxy)ethoxy)tetrahydro-2H-pyran-3,4-diyl diacetate (**12- β**)

9 (1.06 mmol) was dissolved in DCM (3 mL) and cooled to 0°C. Boron trifluoride diethyl etherate (4 eq) and propynol ethoxylate (3 eq) were added dropwise, the solution heated to 40°C and stirred for 8 h. The reaction mixture was diluted with DCM (7 mL) and washed with K₂CO₃ (2 x 10 mL), H₂O (2 x 10 mL) and NaCl (2 x 10 mL). The organic layer was retrieved and dried over MgSO₄. The solvent was removed under vacuum and the crude compound purified with flash column chromatography (solvent system: 0-20% acetone

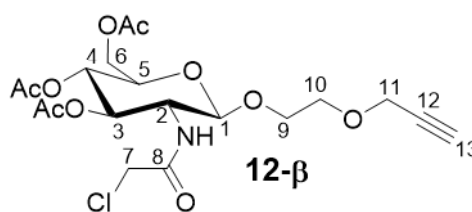
in toluene) yielding two products. Each was further purified with flash column chromatography (solvent system: 10-30% EtOAc in Hexane) to yield **12- α** and **12- β** as pale-yellow oils. Overall % Yield= 72 (35% for **22- α** , 37% for **22- β**)

$^1\text{H NMR}$ (400 MHz, CDCl_3) δ 2.00, 2.03, 2.09 (3s, 9H, 4 CH_3CO), 2.48 (t, $J_{13-11}= 2.4$ Hz, 1H, H_{13}), 3.66 – 3.74 (m, 3H, H_9 , H_{10}), 3.82 – 3.89 (m, 1H, H_9), 3.99 (d, $J=$



$J_{11-13}= 2.4$ Hz, 2H, H_{11}), 4.25 (dd, $J_{6-6}= 12.2$ Hz, $J_{6-5}= 4.2$ Hz, 1H, H_6), 4.33 (ddd, $J_{2-3}= 10.6$ Hz, $J_{2-\text{NH}}= 9.5$ Hz, $J_{2-1}= 3.7$ Hz, 1H, H_2), 4.92 (d, $J_{1-2}= 3.7$ Hz, 1H, H_1), 5.13 (t, $J= 9.7$ Hz, 1H, H_4), 5.30 (dd, $J_{3-2}= 10.5$ Hz, $J_{3-4}= 9.6$ Hz, 1H, H_3), 6.83 (d, $J_{\text{NH}-2}= 9.3$ Hz, 1H, NH). **$^{13}\text{C NMR}$** (100 MHz, CDCl_3) δ 20.8, 20.8, 20.9 (3C, CH_3CO), 42.5 (C_7), 52.3 (C_2), 58.5 (C_{11}), 62.0 (C_6), 67.7 (C_9), 68.0 (C_5), 68.2 (C_4), 68.6 (C_{10}), 71.2 (C_3), 75.1 (C_{13}), 79.5 (C_{12}), 97.4 (C_1), 166.4 (C_8), 169.5, 170.9, 171.2 (3C, CH_3CO). **LC-MS** (m/z) [$\text{M} + \text{Na}^+$]: calculated 486.1; found 486.3

$^1\text{H NMR}$ (400 MHz, CDCl_3) δ 2.01, 2.01, 2.07 (3s, 9H, 4 CH_3CO), 2.45 (t, $J_{13-11}= 2.3$ Hz, 1H, H_{13}), 3.61 – 3.81 (m, 4H, H_5 , H_9 , H_{10}), 3.84 – 3.93 (m, 1H, H_2), 3.93 – 4.06 (m, 3H, H_9 , H_7), 4.09 – 4.21 (m, 3H, H_{11} ,



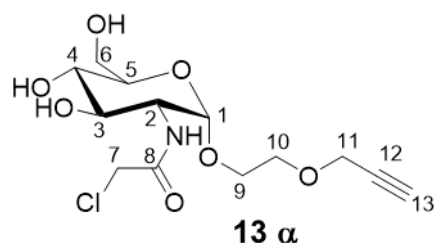
H_6), 4.25 (dd, $J_{6-6}= 12.3$ Hz, $J_{6-5}= 4.7$ Hz, 1H, H_6), 4.83 (d, $J_{1-2}= 8.3$ Hz, 1H, H_1), 5.06 (t, $J= 9.6$ Hz, 1H, H_4), 5.33 (t, $J= 9.9$ Hz, 1H, H_3), 6.67 (d, $J_{\text{NH}-2}= 8.5$ Hz, 1H, NH). **$^{13}\text{C NMR}$** (100 MHz, CDCl_3) δ 20.8, 20.8, 20.9 (3C, CH_3CO), 42.6 (C_7), 55.1 (C_2), 58.6 (C_{11}), 62.2 (C_6), 68.6 (C_4), 68.9 (C_9), 69.3 (C_{10}), 72.0 (C_3), 72.0 (C_5), 74.9 (C_{13}), 79.6 (C_{12}), 100.8 (C_1), 166.6 (C_8), 169.5, 170.8, 170.8 (3C, CH_3CO). **LC-MS** (m/z) [$\text{M} + \text{Na}^+$]: calculated 486.1; found 486.2

2-chloro-N-((2S,3R,4R,5S,6R)-4,5-dihydroxy-6-(hydroxymethyl)-2-(2-(prop-2-yn-1-yloxy)ethoxy)tetrahydro-2H-pyran-3-yl)acetamide (**13- α**) *2-chloro-N-((2R,3R,4R,5S,6R)-4,5-*

dihydroxy-6-(hydroxymethyl)-2-(2-(prop-2-yn-1-yloxy)ethoxy)tetrahydro-2H-pyran-3-yl)acetamide (13-β)

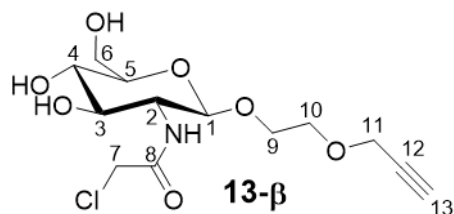
12-α(β) (0.10 mmol) was added to MeOH (1.03 mL) and cooled to 0°C. NaOMe (0.5 M in MeOH, 3 eq) was added dropwise and the reaction mixture stirred at 0°C for 10mins. The reaction was quenched with addition of acidified H₂O (2 mL) and the solvent removed under vacuum. The crude compound was purified with flash column chromatography (solvent system: 0-5% MeOH in DCM) and lyophilised to yield **13-α(β)** as a white powder. %Yield= 63 (α), 59 (β)

¹H NMR (400 MHz, D₂O) δ 3.43 – 3.50 (m, 1H, C₄), 3.61 – 3.73 (m, 2H, H₅, H₉), 3.73 – 3.80 (m, 4H, H₃, H₆, H₁₀), 3.81 – 3.87 (m, 2H, H₆, H₉), 3.93 – 3.98 (m, 1H, H₂), 4.15 (s, 2H, H₇), 4.23 (s, 2H, H₁₁), 4.90 (d, J₁₋₂= 3.6



Hz, 1H, H₁). **¹³C NMR** (100 MHz, D₂O) δ 42.2 (C₇), 53.9 (C₂), 57.9 (C₁₁), 60.4 (C₆), 66.5 (C₉), 68.8 (C₁₀), 69.8 (C₄), 70.9 (C₃), 71.9 (C₅), 96.8 (C₁), 170.0 (C₈)

¹H NMR (400 MHz, D₂O) δ 3.40 – 3.48 (m, 2H, H₃, H₄), 3.56 – 3.65 (m, 1H, H₅), 3.66 – 3.81 (m, 5H, H₂, H₆, H₉, H₁₀), 3.88 – 3.93 (m, 1H, H₆), 4.00 (d, 1H, J₉= 11.7 Hz, J=5.4 Hz, J=2.8 Hz, H₉), 4.17 (s, 2H, H₇),



4.21 (s, 2H, H₁₁), 4.61 (d, J₁₋₂= 8.4 Hz, H₁). **¹³C NMR** (100 MHz, D₂O) δ 42.7 (C₇), 55.9 (C₂), 57.9 (C₁₁), 60.6 (C₆), 68.6, 68.8 (C₉, C₁₀), 69.8 (C₄), 73.4 (C₅), 75.8 (C₃), 100.8 (C₁), 170.2 (C₈)

N.B: As mentioned in Chapter 4 (p.141) ¹H₁₃ has exchanged for ²D₁₃ in D₂O and is therefore not visible on the NMR spectra.

(2R,3S,4R,5R,6S)-2-(acetoxymethyl)-5-(2-chloroacetamido)-6-(2-((1-(2-((5-(dimethylamino)naphthalene)-1-sulfonamido)ethyl)-1H-1,2,3-triazol-4-yl)methoxy)ethoxy)tetrahydro-2H-pyran -3,4-diyl diacetate (14-α) (2R,3S,4R,5R,6R)-2-

(acetoxymethyl)-5-(2-chloroacetamido)-6-(2-((1-(2-((5-(dimethylamino)naphthalene)-1-sulfonamido)ethyl)-1H-1,2,3-triazol-4-yl)methoxy) ethoxy)tetrahydro-2H-pyran-3,4-diy) diacetate (**14-β**)

12-α(β) (0.16 mmol) and **16** (1 eq) were dissolved in THF (1 mL) and the mixture stirred. Sodium ascorbate (2 eq) and copper sulfate pentahydrate (1.2 eq) were dissolved in H₂O (110 μL) and added to the mixture. DIPEA (5 eq) was added and the mixture covered with aluminium foil and stirred at room temperature for 3 h. The mixture was extracted with EtOAc (15 mL), washed with H₂O (5 mL) and brine (5 mL), and concentrated under vacuum. The crude residue was purified first with flash column chromatography (solvent system 0-100% EtOAc in Hexane) and second with flash column chromatography (solvent system 0-70% acetone in toluene) to yield **14-α(β)** as a green powder. % Yield= 73 (α), 81 (β)

¹H NMR (400 MHz, CDCl₃) δ

2.01, 2.02, 2.05 (3s, 9H, 4

CH₃CO), 2.88 (s, 6H, H₂₆), 3.32 –

3.48 (m, 2H, H₁₅), 3.70 – 3.76 (m,

2H, H₁₀), 3.76 – 3.80 (m, 1H, H₉),

3.81 – 3.89 (m, 1H, H₉), 3.93 (d,

J= 1.9 Hz, 2H, H₇), 4.03 – 4.13 (m,

2H, H₅, H₆), 4.13 – 4.20 (m, 1H,

H₆), 4.31 (ddd, J₂₋₃= 10.7 Hz, J_{2-NH}= 9.4 Hz, J₂₋₁= 3.7 Hz, 1H, H₂), 4.40 – 4.52 (m, 2H, H₁₄), 4.63

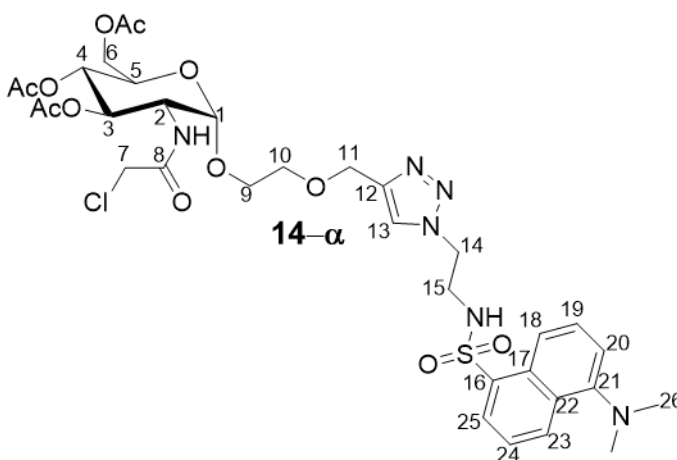
(s, 2H, H₁₁), 4.97 (d, J₁₋₂= 3.6 Hz, 1H, H₁), 5.12 (t, J= 9.7 Hz, 1H, H₄), 5.29 (dd, J₃₋₂= 10.7 Hz,

J₃₋₄= 9.5 Hz, 1H, H₃), 6.22 (t, J= 6.2 Hz, 1H, NH_{sulfonamide}), 7.06 (d, J_{NH-2}= 9.3 Hz, 1H, NH_{amide}),

7.18 (d, J= 7.5 Hz, 1H, H₂₀), 7.49 – 7.56 (m, 2H, H₁₉, H₂₄), 7.59 (s, 1H, H₁₃), 8.17 – 8.26 (m,

2H, H₁₈, H₂₅), 8.55 (d, J= 8.5 Hz, 1H, H₂₃). **¹³C NMR** (100 MHz, CDCl₃) δ 20.8, 20.8, 20.9 (3C,

CH₃CO), 42.5 (C₇), 43.0 (C₁₅), 45.5 (C₂₆), 50.6 (C₁₄), 52.4 (C₂), 62.1 (C₆), 64.5 (C₁₁), 67.7 (C₉),



67.9 (C₅), 68.3 (C₄), 69.6 (C₁₀), 71.2 (C₃), 94.4 (C₁), 115.5 (C₂₀), 118.8 (C₁₈), 123.3 (C₂₄), 124.2 (C₁₃), 128.6 (C₁₉), 129.6, 129.6 (C₂₅, C₂₂), 130.1 (C₁₇), 130.9 (C₂₃), 134.6 (C₁₆), 144.7 (C₁₂), 152.2 (C₂₁), 166.8 (C₈), 169.6, 171.1, 171.5 (3C, CH₃CO). **LC-MS** (m/z) [M + H⁺]: calculated 784.2; found 784.3

N.B: Traces of DMSO present (2.61ppm on ¹H NMR and 41.1pm ¹³C NMR)

¹H NMR (400 MHz, CDCl₃) δ

1.96, 2.02, 2.06 (3s, 9H, 4

CH₃CO), 2.89 (s, 6H, H₂₆),

3.35 – 3.47 (m, 2H, H₁₅), 3.67

– 3.77 (m, 3H, H₅, H₁₀), 3.80 –

3.88 (m, 1H, H₉), 3.88 – 4.05

(m, 4H, H₂, H₉, H₇), 4.10 (dd,

J₆₋₆ = 12.2 Hz, J₆₋₅ = 2.4 Hz, H₆), 4.24 (dd, J₆₋₆ = 12.2 Hz, J₆₋₅ = 2.4 Hz, H₆), 4.42 – 4.56 (m, 2H,

H₁₄), 4.63 (s, 2H, H₁₁), 4.91 (d, J₁₋₂ = 8.4 Hz, 1H, H₁), 5.07 (t, J = 9.6 Hz, 1H, H₄), 5.28 – 5.35

(m, 1H, H₃), 6.06 (t, J = 6.2 Hz, 1H, NH_{sulphonamide}), 7.12 (d, J_{NH-2} = 8.5 Hz, 1H, NH_{amide}), 7.19

(d, J = 7.5 Hz, 1H, H₂₀), 7.51- 7.59 (m, 2H, H₁₉, H₂₄), 7.66 (s, 1H, H₁₃), 8.20 – 8.26 (m, 2H, H₁₈,

H₂₅), 8.56 (d, J = 8.5 Hz, 1H, H₂₃). **¹³C NMR** (100 MHz, CDCl₃) δ 20.8, 20.8, 20.9 (3C, CH₃CO),

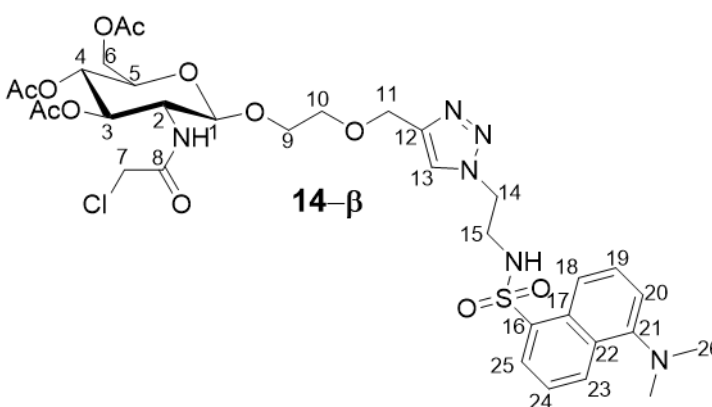
42.7 (C₇), 42.9 (C₁₅), 45.6 (C₂₆), 50.4 (C₁₄), 55.2 (C₂), 62.2 (C₆), 64.5 (C₁₁), 68.7 (C₄), 68.7

(C₉), 70.2 (C₁₀), 71.9 (C₅), 72.2 (C₃), 100.8 (C₁), 115.5 (C₂₀), 118.8 (C₁₈), 123.3 (C₂₄), 124.5

(C₁₃), 128.7 (C₁₉), 129.7, 129.7 (C₁₇, C₂₅), 130.1 (C₂₂), 131.0 (C₂₃), 134.4 (C₁₆), 144.7 (C₁₂),

152.3 (C₂₁), 167.2 (C₈), 169.6, 170.9, 170.9 (3C, CH₃CO). **LC-MS** (m/z) [M + H⁺]: calculated

784.2; found 784.3



2-chloro-N-((2S,3R,4R,5S,6R)-2-(2-((1-(2-((5-(dimethylamino)naphthalene)-1-

sulfonamido) ethyl)-1H-1,2,3-triazol-4-yl)methoxy)ethoxy)-4,5-dihydroxy-6-

(hydroxymethyl)tetrahydro-2H-pyran-3-yl)acetamide (15-α) 2-chloro-N-

((2*R*,3*R*,4*R*,5*S*,6*R*)-2-(2-((1-(2-((5-(dimethylamino) naphthalene)-1-sulfonamido)ethyl)-1*H*-1,2,3-triazol-4-yl)methoxy)ethoxy)-4,5-dihydroxy-6-(hydroxymethyl)tetrahydro-2*H*-pyran-3-yl)acetamide (**15-β**)

14-α(β) (0.074 mmol) was dissolved in MeOH (790 μL) and cooled to 0°C. NaOMe (0.5 M in MeOH, 3 eq) was added dropwise and the solution stirred at 0°C for 10 mins. The reaction was quenched with acidified H₂O (2 mL) and the solvent removed under vacuum. The crude compound was first purified with flash column chromatography (solvent system: 0-5% MeOH in DCM) and the solvent again removed under vacuum. The obtained solid was dissolved in H₂O (15 mL), filtered, and purified with preparative chromatography (2 x 7.5 mL), gradient 20 – 55% CH₃CN against H₂O over 7 min (flow rate: 20 mL/min), detecting at 325 nm. The solvent was removed under vacuum, the purified compound re-dissolved in H₂O (3 mL) and lyophilised to yield **15-α(β)** as a green powder. %Yield= 55 (α), 52 (β)

¹H NMR (400 MHz, D₂O) δ 2.84

(s, 6H, H₂₆), 3.41 (dd, J= 10.0 Hz,

J= 9.1 Hz, 1H, H₅), 3.50 – 3.59 (m,

6H, H₄, H₉, H₁₀, H₁₅), 3.64 – 3.75

(m, 4H, H₃, H₆, H₉), 3.88 (dd, J₂₋₃=

10.6 Hz, J₂₋₁= 3.6 Hz, 1H, H₂), 4.02

(s, 2H, H₇), 4.07 (s, 2H, H₁₁), 4.29

(dd, J= 6.4 Hz, J= 4.3 Hz, 2H, H₁₄),

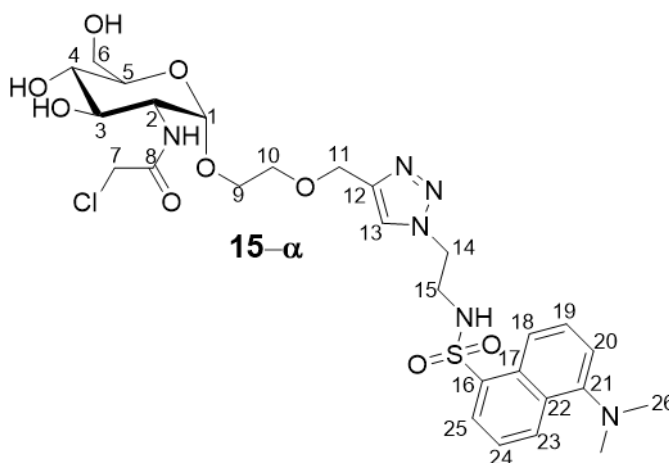
4.78 (s, 1H, H₁*), 7.36 – 7.33 (m, 2H, H₁₃, H₂₀), 7.56 – 7.51 (m, 1H, H₁₉), 7.58 (dd, J₂₄₋₂₃= 8.6

Hz, J₂₄₋₂₅= 7.5 Hz, 1H, H₂₄), 7.92 (d, J₁₈₋₁₉= 8.7 Hz, 1H, H₁₈), 8.10 (d, J₂₅₋₂₄= 7.3 Hz, 1H, H₂₅),

8.40 (d, J₂₃₋₂₄= 8.6 Hz, 1H, H₂₃), 60.3 (C₆). ¹³C NMR (100 MHz, D₂O): δ 42.0 (C₁₅), 42.1 (C₇),

45.0 (C₂₆), 49.3 (C₁₄), 53.8 (C₂), 62.5 (C₁₁), 66.4 (C₉), 68.8 (C₁₀), 69.7 (C₅), 70.9 (C₃), 71.8

(C₄), 96.8 (C₁), 115.8 (C₂₀), 119.0 (C₁₈), 123.8 (C₂₄), 124.9 (C₁₃), 128.4 (C₁₇), 128.4 (C₁₉),



128.7 (C₂₂), 129.5 (C₁₇), 130.1 (C₁₉), 133.4 (C₁₆), 142.6 (C₁₂), 150.4 (C₂₁), 169.8 (C₈). **LC-MS** (m/z) [M + H]⁺: calculated 657.2; found 657.3. **HR-MS** m/z= 657.2110 [M+H]⁺ (calculated 657.2031)

* anomeric proton H₁ is under the water peak

¹H NMR (400 MHz, D₂O) δ

2.89 (s, 6H, H₂₆), 3.42 – 3.49

(m, 2H, H₄, H₅), 3.51 – 3.58 (m,

2H, H₁₀), 3.57 – 3.65 (m, 3H,

H₃ H₁₅), 3.67 – 3.80 (m, 3H, H₂,

H₆, H₉), 3.88 – 3.98 (m, 2H, H₆,

H₉), 4.08 (s, 2H, H₇), 4.15 (s,

2H, H₁₁), 4.32 – 4.39 (m, 2H, H₁₄), 4.58 (d, J₁₋₂= 8.4 Hz, 1H, H₁), 7.35 – 7.42 (m, 2H, H₁₃, H₂₀),

7.55 – 7.67 (m, 2H, H₁₉, H₂₄), 7.99 (d, J₁₈₋₁₉= 8.7 Hz, 1H, H₁₈), 8.15 (d, J₂₅₋₂₄= 7.3 Hz, 1H, H₂₅),

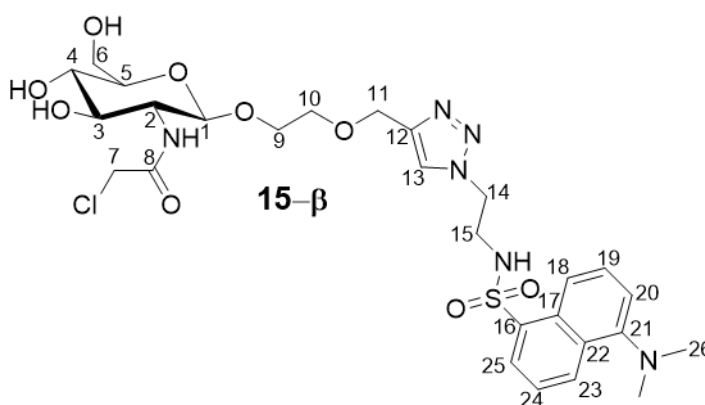
8.46 (d, J₂₃₋₂₄= 8.6 Hz, 1H, H₂₃). **¹³C NMR** (100 MHz, D₂O): δ 42.0 (C₁₅), 42.3 (C₇), 45.0 (C₂₆),

49.3 (C₁₄), 55.9 (C₂), 60.7 (C₆), 62.7 (C₁₁), 68.7 (C₉), 68.8 (C₁₀), 69.9 (C₄), 73.5 (C₃), 75.9 (C₅),

100.8 (C₁), 115.9 (C₂₀), 118.9 (C₁₈), 123.8 (C₂₄), 124.9 (C₁₃), 128.5 (C₁₇), 128.6 (C₁₉), 128.8

(C₂₂), 129.6 (C₂₅), 130.2 (C₂₃), 133.4 (C₁₆), 150.6 (C₂₁), 170.0 (C₈). **LC-MS** (m/z) [M + H]⁺:

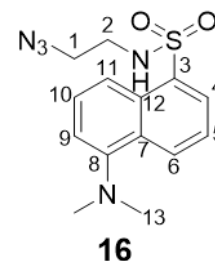
calculated 657.2; found 657.3. **HR-MS** m/z= 657.2111 [M+H]⁺ (calculated 657.2031)



N-(2-azidoethyl)-5-(dimethylamino)naphthalene-1-sulfonamide (**16**)

Dansyl chloride (5.00 mmol), 2-bromoethylamine hydrobromide (1 eq) and triethylamine (2 eq) were added to DCM (25 mL) and stirred at room temperature for 4 h. The solvent was removed under vacuum. The resulting residue was dissolved in MeCN (50 mL), NaN₃ (2.5 eq) was added and the mixture refluxed overnight. The solvent was removed under vacuum and the crude product purified by flash column chromatography (solvent system: 0-50% EtOAc in Hexane) to yield **16** as a green oil. % Yield= 87.

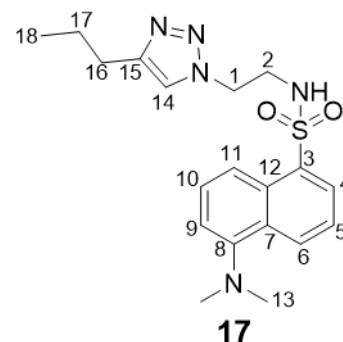
¹H NMR (400 MHz, CDCl₃) δ 2.89 (s, 6H, H₁₃), 3.03 – 3.08 (m, 2H, H₂), 3.29 – 3.33 (m, 2H, H₁), 4.99 (t, J_{NH, H2} = 5.9 Hz, 1H, NH _{sulphonamide}), 7.21 (d, J₉₋₁₀ = 7.6 Hz, 1H, H₉), 7.53 (dd, J₅₋₆ = 8.5 Hz, J₅₋₄ = 7.4 Hz, 1H, H₅), 7.60 (dd, J₁₀₋₁₁ = 8.6 Hz; J₁₀₋₉ = 7.7 Hz, 1H, H₁₀), 8.24 – 8.26 (m, 2H, H₄), 8.26 – 8.28 (m, 2H, H₁₁), 8.57 (d, J₆₋₅ = 8.5 Hz, 1H, H₆). **¹³C NMR** (100 MHz, CDCl₃) δ 42.5 (C₂), 45.6 (C₁₃), 51.1 (C₁), 115.5 (C₉), 118.6 (C₁₁), 123.3 (C₅), 128.8 (C₁₀), 129.6 (C₁₂), 129.8 (C₄), 130.1 (C₇), 131.0 (C₆), 134.6 (C₃), 152.3 (C₈). **HR-MS** m/z = 320.1182 [M+H]⁺ (calculated 320.1181)



5-(dimethylamino)-N-(2-(4-propyl-1H-1,2,3-triazol-1-yl)ethyl)naphthalene-1-sulfonamide
(17)

Titled compound was obtained from **16** and 1-pentyne under the click reaction conditions described for **12**. % Yield = 87

¹H NMR (400 MHz, CDCl₃) 0.95 (t, J₁₈₋₁₇ = 7.4 Hz, 3H, H₁₈), 1.58 – 1.68 (m, 2H, H₁₇), 2.58 – 2.63 (m, 2H, H₁₆), 3.44 (dd, J = 11.2 Hz, J = 6.1 Hz, 2H, H₂), 4.30 – 4.36 (m, 2H, H₁), 4.41 – 4.48 (m, 1H, NH), 7.11 (s, 1H, H₁₄), 7.19 (d, J₉₋₁₀ = 7.1 Hz, H₉), 7.51 – 7.57 (m, 2H, H₅, H₁₀), 8.18 (dt, J = 8.7 Hz, J = 1 Hz, 1H, H₁₁) 8.25 (dd, J = 7.3 Hz, J = 1.2 Hz, 1H, H₄), 8.54 – 8.58 (m, 1H, H₆). **¹³C NMR** (100 MHz, CDCl₃) δ 14.0 (C₁₈), 22.8 (C₁₇), 27.7 (C₁₆), 42.9 (C₂), 45.6 (C₁₃), 50.0 (C₁), 115.5 (C₉), 118.6 (C₁₁), 122.1 (C₁₄), 123.3 (C₅), 128.9 (C₁₀), 129.6 (C₁₂), 129.7 (C₄), 130.1 (C₇), 131.0 (C₆), 134.5 (C₃), 148.3 (C₁₅), 152.2 (C₈)

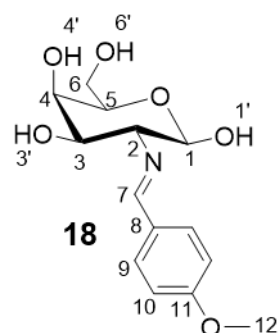


7.2.4. Chapter 5 compounds: galactosamine-based probes

(2R,3R,4R,5R,6R)-6-(hydroxymethyl)-3-(((E)-4-methoxybenzylidene)amino)tetrahydro-2H-pyran-2,4,5-triol (**18**)

Galactosamine hydrochloride (9.25 mmol) was dissolved in 1M NaOH (9.30 mL) at 0°C. *p*-anisaldehyde (1 eq) was added dropwise, the mixture shaken vigorously until formation of a white solid and placed in the freezer (-20°C) for 1 h for complete precipitation. Formed solid was collected by filtration, washed dropwise with ice-cold H₂O (10 mL), ice-cold 1 : 1 EtOH : Et₂O (10 mL) and dried under vacuum to yield **18** as a white powder. % Yield= 56.

¹H NMR (400 MHz, DMSO-d₆) δ 3.09 (dd, J= 9.5 Hz, J= 7.7 Hz, 1H, H₂), 3.46 (t, J = 6.2 Hz, 1H, H₅), 3.50 – 3.62 (m, 3H, H₃, H₆), 3.67 (t, J = 3.5 Hz, 1H, H₄), 3.80 (s, 3H, H₁₂), 4.39 (d, J_{H4', H4} = 4.4 Hz, 1H, H_{4'}), 4.50 (d, J_{3'-3} = 7.0 Hz, 1H, H_{3'}), 4.58 – 4.64 (m, 2H, H₁, H_{6'}), 6.42 (d, J_{1'-1} = 6.3 Hz, 1H, H_{1'}), 6.98 (d, J₁₀₋₉ = 8.8 Hz, 2H, H₁₀), 7.67 (d, J₉₋₁₀ = 8.8 Hz, 2H, H₉), 8.13 (s, 1H, H₇). **¹³C NMR** (100 MHz, DMSO-d₆) δ

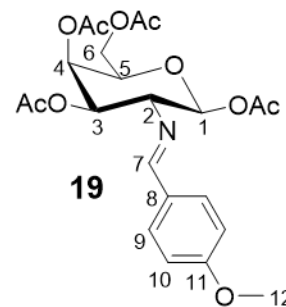


55.3 (C₁₂), 60.7 (C₆), 67.2 (C₄), 71.6 (C₃), 74.5 (C₂), 75.1 (C₅), 96.1 (C₁), 113.9 (C₁₀), 129.2 (C₈), 129.5 (C₉), 161.0 (C₁₁), 161.2 (C₇),

(2S,3R,4R,5R,6R)-6-(acetoxymethyl)-3-(((*E*)-4-methoxybenzylidene)amino)tetrahydro-2H-pyran-2,4,5-triyl triacetate (**19**)

18 (3.20 mmol) was dissolved in pyridine (10 mL) at 0°C. Acetic anhydride (8 eq) was added dropwise, the reaction warmed to room temperature and stirred overnight (16 h). The reaction mixture was pooled into ice (50 mL) and stirred until this melted. The formed precipitate was collected by filtration, washed with ice-cold H₂O (200 mL), ice-cold EtOH (2 mL) and dried under vacuum to yield **19** as a white powder. %Yield=85.

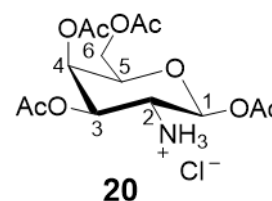
¹H NMR (400 MHz, DMSO-*d*₆) δ 1.82, 1.97, 2.00, 2.12 (4s, 12H, CH₃CO), 3.52 (dd, *J*₂₋₃= 10.3, *J*₂₋₁= 8.2 Hz, 1H, H₂), 3.79 (s, 3H, H₁₂), 4.04 – 4.09 (m, 2H, H₆), 4.46 (t, *J*₅₋₆= 6.5 Hz, 1H, H₅), 5.27 (d, *J*₄₋₃= 3.3 Hz, 1H, H₄), 5.35 (dd, *J*₃₋₂= 10.4 Hz, *J*₃₋₄= 3.4 Hz, 1H, H₃), 5.98 (d, *J*₁₋₂= 8.2 Hz, 1H, H₁), 6.99 (d, *J*₁₀₋₉= 8.8 Hz, 2H, H₁₀), 7.67 (d, *J*₉₋₁₀= 8.8 Hz, 2H, H₉), 8.31 (s, 1H, H₇). **¹³C NMR** (100 MHz, DMSO-*d*₆) δ 20.3, 20.4, 20.5, 20.5, (4C, CH₃CO), 55.4 (C₁₂), 61.4 (C₆), 66.0 (C₄), 68.4 (C₂), 70.9 (C₅), 70.9 (C₃), 92.8 (C₁), 114.2 (C₁₀), 128.3 (C₈), 129.9 (C₉), 161.8 (C₁₁), 164.7 (C₇), 168.6, 169.2, 169.9, 170.0 (4C, CH₃CO)



(2S,3R,4R,5R,6R)-2,4,5-triacetoxy-6-(acetoxymethyl)tetrahydro-2H-pyran-3-aminium chloride (**20**)

19 (2.70 mmol) was dissolved in acetone (11 mL) and heated to reflux. 4M HCl (1.2 eq) was added dropwise, the solution stirred to formation of a white precipitate (5 min), at which point heating was continued for a further 2 mins. The formed precipitate was collected by filtration and washed with ice-cold acetone (50 mL) to yield **20** as a white powder. % Yield= 88.

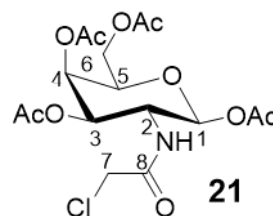
¹H NMR (400 MHz, DMSO-*d*₆) δ 1.99, 2.00, 2.12, 2.16 (4s, 12H, CH₃CO), 3.39 (dd, *J*₂₋₁= 8.2 Hz, *J*₂₋₃= 6.2 Hz, 1H, H₂), 4.98 – 4.09 (m, 2H, H₆), 4.29 (t, *J*= 6.3 Hz, 1H, H₅), 5.24 – 5.32 (m, 2H, H₃, H₄), 5.89 (d, *J*₁₋₂= 8.7 Hz, 1H, H₁), 8.71 (s, 3H, NH). **¹³C NMR** (100 MHz, DMSO-*d*₆) δ 20.3, 20.5, 20.7, 20.8 (4C, CH₃CO), 49.4 (C₂), 61.2 (C₆), 68.8, 65.8 (C₃, C₄), 71.1 (C₅), 90.3 (C₁), 168.6, 169.3, 169.9, 169.9 (4C, CH₃CO).



(2S,3R,4R,5R,6R)-6-(acetoxymethyl)-3-(2-chloroacetamido)tetrahydro-2H-pyran-2,4,5-triyl triacetate (**21**)

20 (1.5 mmol) was added to DCM (12 mL). Triethylamine (2.5 eq) was added dropwise and the mixture stirred to a clear solution before being cooled to 0°C. Chloroacetyl chloride (3 eq) was added dropwise and the solution stirred at room temperature to reaction completion (1.5 h, monitored by TLC 1:1 hexane: EtOAc). The solution was diluted with DCM (8 mL) and washed with acidified H₂O (2 x 20 mL), K₂CO₃ (2 x 20 mL) and NaCl (2 x 20 mL). The organic layer was dried over MgSO₄, solvent removed under vacuum and the crude compound purified with normal phase flash column chromatography (solvent system: 0-50% EtOAc in Hexane) to yield **21** as a pale-yellow powder. % Yield= 92%

¹H NMR (400 MHz, CDCl₃) δ 2.02, 2.05, 2.13, 2.17 (4s, 12H, CH₃CO), 4.00 (s, 2H, H₇), 4.06 (td, J=6.5 Hz, J= 1.0Hz, 1H, H₅), 4.09 – 4.20 (m, 2H, H₆), 4.38 (dt, J₂₋₃= 11.3 Hz, J_{2-NH}= 9.0 Hz, 1H, H₂), 5.24 (dd, J₃₋₂= 11.3 Hz, J_{3,4}= 3.3 Hz, 1H, H₃), 5.38 - 5.42 (m, 1H, H₄), 5.82

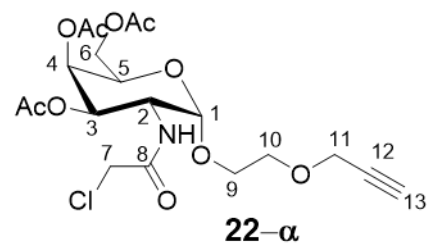


(d, J₁₋₂= 8.8 Hz, 1H, H₁), 6.48 (d, J=9.2 Hz, 1H, NH). **¹³C NMR** (100 MHz, CDCl₃): δ 20.7, 20.7, 20.8, 21.0 (4C, CH₃CO), 42.5 (C₇), 50.5 (C₂), 61.5 (C₆), 66.6 (C₄), 69.9 (C₃), 71.9 (C₅), 92.5 (C₁), 166.9 (C₈), 169.5, 170.2, 170.6, 170.6 (CH₃CO). **LC-MS** (m/z) [M + Na⁺]: calculated 446.2; found 446.3

(2R,3R,4R,5R,6S)-2-(acetoxymethyl)-5-(2-chloroacetamido)-6-(2-(prop-2-yn-1-yloxy)ethoxy) tetrahydro-2H-pyran-3,4-diyl diacetate (**22-α**) *(2R,3R,4R,5R,6R)*-2-(acetoxymethyl)-5-(2-chloroacetamido)-6-(2-(prop-2-yn-1-yloxy)ethoxy)tetrahydro-2H-pyran-3,4-diyl diacetate (**22-β**)

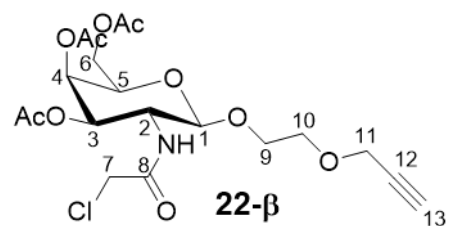
Titled compounds were obtained from compd **21** under the glycosylation reaction described for **12-α(β)**. Overall % Yield= 70 (34% for **22-α**, 36% for **22-β**)

¹H NMR (400 MHz, CDCl₃) δ 1.99, 2.05, 2.16 (3s, 9H, CH₃CO), 2.47 (t, J = 2.4 Hz, 1H, H₁₃), 3.67 – 3.74 (m, 3H, H₉, H₁₀), 3.83 – 3.90 (m, 1H, H₉), 4.01 (d, J = 3.2 Hz, 2H, H₇), 4.11 (m, 2H, H₆), 4.19 (d, J = 2.4 Hz, 1H, H₁₁), 4.28



(t, J₅₋₆ = 6.6 Hz, 1H, H₅), 4.57 (ddd, J₂₋₃ = 11.2 Hz, J_{2-NH} = 9.7 Hz, J₂₋₁ = 3.7 Hz, 1H, H₂), 4.95 (d, J_{H1, H2} = 3.7 Hz, 1H, H₁), 5.24 (dd, J₃₋₂ = 11.3 Hz, J₃₋₄ = 3.3 Hz, 1H, H₃), 5.41 (dd, J₄₋₃ = 3.2 Hz, J₄₋₅ = 1.2 Hz, 1H, H₄), 6.71 (d, J_{NH, H2} = 9.7 Hz, 1H, NH). **¹³C NMR** (100 MHz, CDCl₃) δ 20.9 (3 x CH₃CO), 42.6 (C₇), 48.3 (C₂), 58.5 (C₁₁), 62.0 (C₆), 67.0 (C₅), 67.5 (C₄), 67.7 (C₉), 68.6 (C₃), 68.6 (C₁₀), 75.1 (C₁₃), 79.5 (C₁₂), 97.9 (C₁), 166.4 (C₈), 170.4, 170.6, 170.8 (3C, CH₃CO). **LC-MS** (m/z) [M + Na⁺]: calculated 486.1; found 486.3

¹H NMR (400 MHz, CDCl₃) δ 2.01, 2.05, 2.15 (3s, 9H, CH₃CO), 2.45 (t, J = 2.3 Hz, 1H, H₁₃), 3.65 – 3.73 (m, 2H, H₁₀), 3.76 – 3.84 (m, 1H, H₉), 3.93 (t, J = 6.7 Hz, 1H, H₅), 3.97 – 4.08 (m, 4H, H₂, H₉, H₇), 4.12 – 4.23

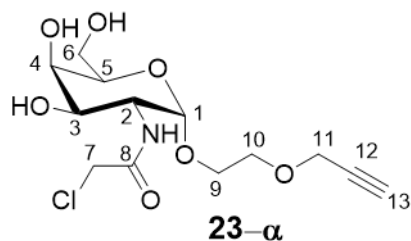


(m, 4H, H₆, H₁₁), 4.84 (d, J₁₋₂ = 8.4 Hz, 1H, H₁), 5.30 (dd, J₃₋₂ = 11.3 Hz, J₃₋₄ = 3.4 Hz, 1H, H₃), 5.38 (d, J₄₋₃ = 3.4 Hz, 1H, H₄), 6.49 (d, J_{NH, H2} = 8.6 Hz, 1H, NH). **¹³C NMR** (100 MHz, CDCl₃) δ 20.8, 20.8, 20.8 (3C, CH₃CO), 42.7 (C₇), 52.1 (C₂), 58.6 (C₁₁), 61.6 (C₆), 66.8 (C₄), 68.9 (C₉), 69.3 (C₁₀), 69.8 (C₃), 71.0 (C₅), 74.9 (C₁₃), 79.6 (C₁₂), 101.2 (C₁), 166.6 (C₈), 170.3, 170.5, 170.6 (3C, CH₃CO). **LC-MS** (m/z) [M + Na⁺]: calculated 486.1; found 486.3

2-chloro-N-((2S,3R,4R,5R,6R)-4,5-dihydroxy-6-(hydroxymethyl)-2-(2-(prop-2-yn-1-yloxy)ethoxy)tetrahydro-2H-pyran-3-yl)acetamide (**13-α**) *2-chloro-N-((2R,3R,4R,5R,6R)-4,5-dihydroxy-6-(hydroxymethyl)-2-(2-(prop-2-yn-1-yloxy)ethoxy)tetrahydro-2H-pyran-3-yl)acetamide* (**13-β**)

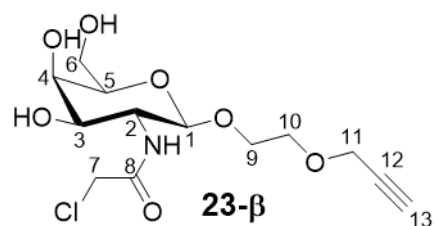
Titled compounds were obtained from compd **22-α** and **22-β** respectively under the deacetylation conditions described for **13-α(β)**. Average % Yield = 57 (α), 52 (β)

¹H NMR (400 MHz, D₂O) δ 3.69 – 3.75 (m, 1H, H₉), 3.78 – 3.82 (m, 2H, H₆), 3.82 – 3.86 (m, 2H, H₁₀), 3.88 – 3.94 (m, 1H, H₉), 3.98 – 4.06 (m, 3H, H₃, H₄, H₅), 4.22 (d, J = 0.5 Hz, 2H, H₇), 4.27 (dd, J₂₋₃ = 10.6 Hz, J₂₋₁ = 3.8 Hz, 1H,



H₂), 4.30 (s, 2H, H₁₁), 5.00 (d, J₁₋₂ = 3.8 Hz, 1H, H₁). **¹³C NMR** (100 MHz, D₂O) δ 42.3 (C₇), 50.3 (C₂), 57.9 (C₁₁), 61.2 (C₆), 66.5 (C₉), 67.7 (C₃), 68.5 (C₄), 68.8 (C₁₀), 71.1 (C₅), 75.8 (*t* C₁₃), 78.9 (*t*, C₁₂)*, 97.0 (C₁), 170.1 (C₈)

¹H NMR (400 MHz, D₂O) δ 3.74 – 3.71 (m, 1H, H₅), 3.75 – 3.80 (m, 2H, H₁₀), 3.80 – 3.88 (m, 4H, H₃, 2xH₆, H₉), 4.95 – 4.02 (m, 2H, H₂, H₄), 4.06 (ddd, J = 11.7 Hz, J = 5.4 Hz, J = 2.9 Hz, 1H, H₉), 4.23 (s, 2H, H₇), 4.26 (s,



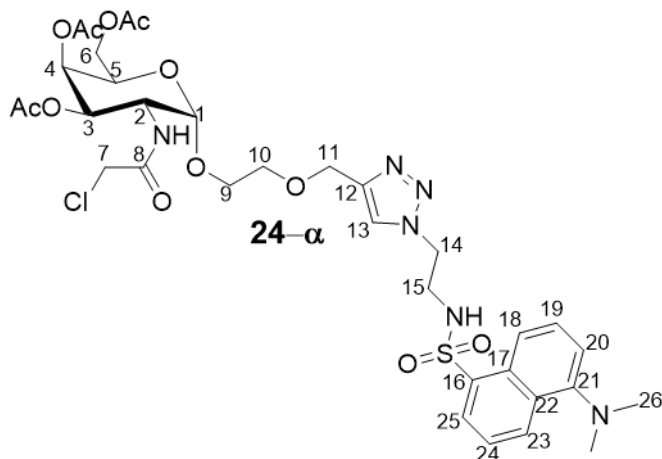
2H, H₁₁), 4.61 (d, J₁₋₂ = 8.4 Hz, 1H, H₁). **¹³C NMR** (100 MHz, D₂O) δ 42.4 (C₇), 53.0 (C₂), 57.9 (C₁₁), 61.0 (C₆), 67.9 (C₄), 68.7, 68.7 (C₉, C₁₀), 70.7 (C₃), 75.2 (C₅), 75.7 (*t*, C₁₃), 78.9 (*t*, C₁₂) *, 101.3 (C₁), 170.4 (C₈)

*¹H₁₃ is exchanged for ²D₁₃ in D₂O which split C₁₃ and C₁₂ signals in ¹³C NMR

(2*R*,3*R*,4*R*,5*R*,6*S*)-2-(acetoxymethyl)-5-(2-chloroacetamido)-6-(2-((1-(2-((5-(dimethylamino)naphthalene)-1-sulfonamido)ethyl)-1*H*-1,2,3-triazol-4-yl)methoxy)ethoxy)tetrahydro-2*H*-pyran -3,4-diyl diacetate (**14-α**) (2*R*,3*R*,4*R*,5*R*,6*R*)-2-(acetoxymethyl)-5-(2-chloroacetamido)-6-(2-((1-(2-((5-(dimethylamino)naphthalene)-1-sulfonamido)ethyl)-1*H*-1,2,3-triazol-4-yl)methoxy) ethoxy)tetrahydro-2*H*-pyran-3,4-diyl diacetate (**14-β**)

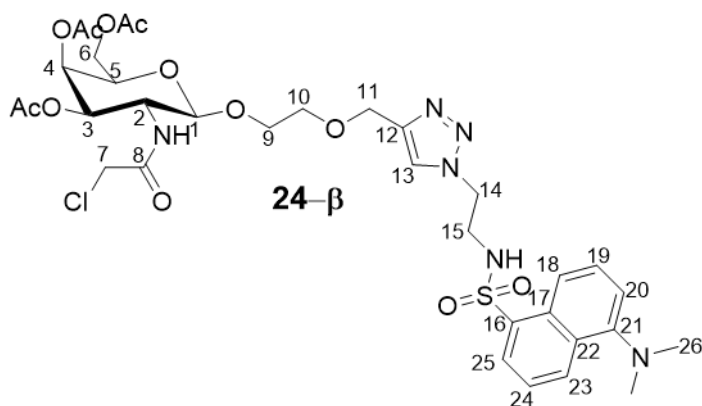
Titled compounds were obtained from compd **22-α** and **22-β** respectively under the click reaction conditions described for **14-α(β)**. % Yield= 74 (α), 82 (β)

¹H NMR (400 MHz, CDCl₃) δ 1.98, 2.00, 2.11 (3s, 9H, CH₃CO), 2.89 (s, 6H, H₂₆), 3.37 – 3.43 (m, 2H, H₁₅), 3.70 – 3.76 (m, 3H, H₉, H₁₀), 3.80 – 3.88 (m, 1H, H₉), 3.98 (s, 2H, H₇), 4.03 – 4.15 (m, 2H, H₆), 4.29 (td, J₅₋₆ = 6.4 Hz, J₅₋₄ = 1.4 Hz, 1H, H₅), 4.39 – 4.52 (m, 2H, H₁₄), 4.56 – 4.70 (m,



3H, H₂, H₁₁), 5.01 (d, J₁₋₂ = 3.6 Hz, 1H, H₁), 5.20 (dd, J₃₋₂ = 11.3 Hz, J₃₋₄ = 3.3 Hz, 1H, H₃), 5.43 (dd, J₄₋₃ = 3.3 Hz, J₄₋₅ = 1.4 Hz, 1H, H₄), 6.14 (t, J_{NH-15} = 5.6 Hz, 1H, NH_{sulfonamide}), 7.14 (d, J_{NH-2} = 9.6 Hz, 1H, NH_{amide}), 7.18 (d, J₂₀₋₁₉ = 7.4 Hz, 1H, H₂₀), 7.50 – 7.57 (m, 2H, H₁₉, H₂₄), 7.61 (s, 1H, H₁₃), 8.19 – 8.26 (m, 2H, H₁₈, H₂₅), 8.56 (d, J₂₃₋₂₄ = 8.5 Hz, 1H, H₂₃). **¹³C NMR** (100 MHz, CDCl₃) δ 20.8, 20.9, 21 (3C, CH₃CO), 42.7 (C₇), 42.8 (C₁₅), 45.6 (C₂₆), 48.3 (C₂), 50.4 (C₁₄), 62.0 (C₆), 64.4 (C₁₁), 66.9 (C₅), 67.6, 67.6 (C₄, C₉), 68.7 (C₃), 69.3 (C₁₀), 98.0 (C₁), 115.5 (C₂₀), 118.9 (C₁₈), 123.3 (C₂₄), 124.2 (C₁₃), 128.6 (C₁₉), 129.6 (C₂₅), 129.6 (C₁₇), 130.1 (C₂₂), 130.8 (C₂₃), 134.7 (C₁₆), 144.8 (C₁₂), 152.2 (C₂₁), 166.9 (C₈), 170.6, 170.7, 171.3 (3C, CH₃CO). **LC-MS** (m/z) [M + H⁺]: calculated 784.2; found 784.3

¹H NMR (400 MHz, CDCl₃) δ 1.97, 1.99, 2.14 (3s, 9H, CH₃CO), 2.89 (s, 6H, H₂₆), 3.33 – 3.50 (m, 2H, H₁₅), 3.67 – 3.81 (m, 2H, H₁₀), 3.84 – 3.99 (m, 3H, H₅, H₉), 4.02 (d, J = 6.9 Hz, 2H, H₇), 4.05 – 4.20 (m, 3H, H₂, H₆),



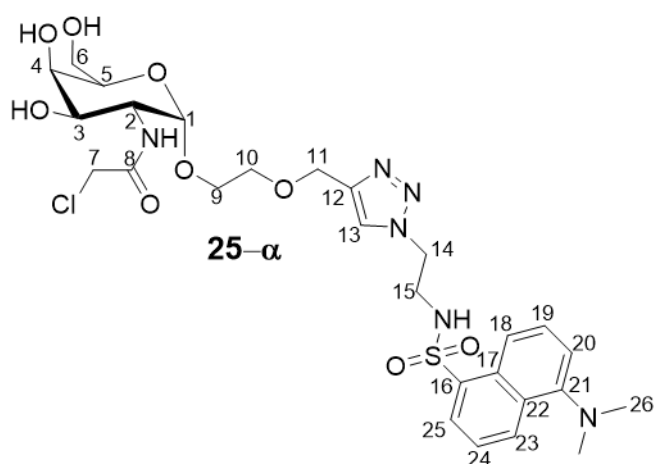
4.40 – 4.49 (m, 1H, H₁₄), 4.50 – 4.59 (m, 1H, H₁₄), 4.63 (s, 2H, H₁₁), 4.89 (d, J₁₋₂ = 8.5 Hz, 1H, H₁), 5.21 (dd, J₃₋₂ = 11.2 Hz, J₃₋₄ = 3.4 Hz, 1H, H₃), 5.36 (d, J = 2.9 Hz, 1H, H₄), 6.07 (t, J_{NH-15} =

5.8 Hz, 1H, NH_{sulfonamide}), 7.09 (d, J_{NH-2} = 8.8 Hz, 1H, NH_{amide}), 7.19 (d, J₂₀₋₁₉ = 7.6 Hz, 1H, H₂₀), 7.61 – 7.50 (m, 2H, H₁₉, H₂₄), 7.67 (s, 1H, H₁₃), 8.19 (d, J₁₈₋₁₉ = 8.7 Hz, 1H, H₁₈), 8.24 (dd, J₂₅₋₂₄ = 7.3, 1.1 Hz, 1H, H₂₅), 8.56 (d, J₂₃₋₂₄ = 8.5 Hz, 1H, H₂₃). ¹³C NMR (100 MHz, CDCl₃) δ 20.8, 20.8, 20.9 (3C, CH₃CO), 42.8 (C₇), 42.8 (C₁₅), 45.6 (C₂₆), 50.4 (C₁₄), 51.8 (C₂), 61.6 (C₆), 64.5 (C₁₁), 67.0 (C₄), 68.6 (C₉), 70.4 (C₃), 70.7 (C₁₀), 70.8 (C₅), 101.3 (C₁), 115.6 (C₂₀), 118.8 (C₁₈), 123.6 (C₂₄), 124.6 (C₁₃), 128.8 (C₁₉), 129.6 (C₁₇), 129.7 (C₂₅), 130.2 (C₂₂), 131.0 (C₂₃), 134.3 (C₁₆), 144.8 (C₁₂), 152.2 (C₂₁), 167.3 (C₈), 170.5, 170.6, 170.8 (3C, CH₃CO). **LC-MS** (m/z) [M + H⁺]: calculated 784.2; found 784.3

2-chloro-N-((2S,3R,4R,5R,6R)-2-(2-((1-(2-((5-(dimethylamino)naphthalene)-1-sulfonamido)ethyl)-1H-1,2,3-triazol-4-yl)methoxy)ethoxy)-4,5-dihydroxy-6-(hydroxymethyl)tetrahydro-2H-pyran-3-yl)acetamide (**25-α**) *2-chloro-N-((2R,3R,4R,5R,6R)-2-(2-((1-(2-((5-(dimethylamino)naphthalene)-1-sulfonamido)ethyl)-1H-1,2,3-triazol-4-yl)methoxy)ethoxy)-4,5-dihydroxy-6-(hydroxymethyl)tetrahydro-2H-pyran-3-yl)acetamide* (**25-β**)

Titled compounds were obtained from compd **24-α** and **24-β** respectively under the deacetylation conditions described for **15-α(β)**. % Yield= 41 (α), 49 (β)

¹H NMR (400 MHz, D₂O) δ 2.88 (s, 6H, H₂₆) 3.54 – 3.65 (m, 5H, H₉, H₁₀, H₁₅), 3.70 – 3.82 (m, 3H, H₆, H₆, H₉), 3.87 – 3.93 (m, H, H₃, H₅), 3.96 (d, J = 3.1 Hz, 1H, H₄), 4.10 (s, 2H, H₇), 4.17 – 4.23 (m, 3H, H₂, H₁₁), 4.33 – 4.39 (m, 2H, H₁₄), 4.89 (d, J₁₋₂ = 3.7 Hz, 1H, H₁), 7.37 (d, J = 7.6 Hz, 1H,



H₂₀), 7.43 (s, 1H, H₁₃), 7.55 – 7.65 (m, 2H, H₁₉, H₂₄), 8.01 (d, J=8.6 Hz, 1H, H₁₈), 8.13 (d, J=7.2 Hz, 1H, H₂₅), 8.44 (d, J= 8.6 Hz, 1H, H₂₃). ¹³C NMR (100 MHz, D₂O) δ 42.2 (C₇), 42.3 (C₁₅), 45.0 (C₂₆), 49.5 (C₁₄), 50.3 (C₂), 61.2 (C₆), 62.6 (C₁₁), 66.4 (C₉), 67.7 (C₃ / C₅), 68.5 (C₄), 68.8 (C₁₀), 71.0 (C₃ / C₅), 96.9 (C₁), 115.8 (C₂₀), 119.0 (C₁₈), 123.8 (C₂₄), 125.0 (C₁₃), 128.4 (C₁₉), 128.6 (C₁₇), 128.8 (C₂₂), 129.4 (C₂₅), 130.0 (C₂₃), 133.8 (C₁₆), 142.8 (C₁₂), 150.6 (C₂₁), 169.9 (C₈). **LC-MS** (m/z) [M + H⁺]: calculated 657.2; found 657.3

¹H NMR (400 MHz, D₂O): δ 2.88

(s, 6H, H₂₆) 3.51 – 3.58 (m, 2H,

H₁₀), 3.58 – 3.63 (m, 2H, H₁₅),

3.63 – 3.64 (m, 2H, H₅, H₉), 3.75

– 3.84 (m, 3H, H₄, H₆, H₆), 3.91 –

3.99 (m, 3H, H₂, H₃, H₉), 4.07 (d,

J= 2.2 Hz, 2H, H₇), 4.15 (s, 2H,

H₁₁), 4.35 (dd, J= 6.3 Hz, J= 4.5 Hz, 2H, H₁₄), 4.52 (d, J₁₋₂= 8.5 Hz, 1H, H₁), 7.37 (d, J₂₀₋₁₉= 7.5

Hz, 1H, H₂₀), 7.40 (s, 1H, H₁₃), 7.64 – 7.56 (m, 2H, H₁₉, H₂₄), 8.00 (d, J₁₈₋₁₉= 8.7 Hz, 1H, H₁₈),

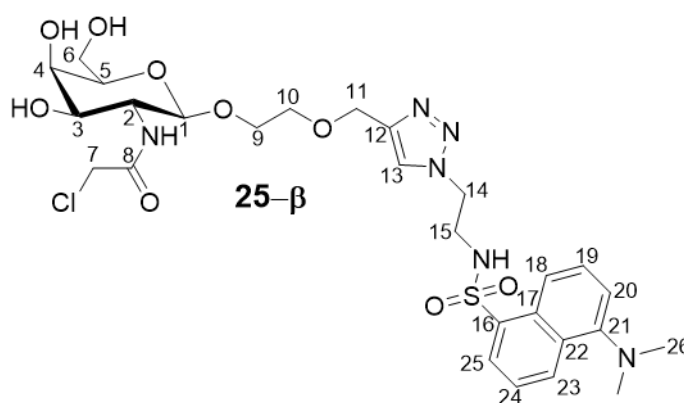
8.14 (dd, J₂₅₋₂₄= 7.4, 1.0 Hz, 1H, H₂₅), 8.44 (d, J₂₃₋₂₄= 8.6 Hz, 1H, H₂₃). ¹³C NMR (100 MHz,

D₂O) δ 42.1 (C₁₅), 42.4 (C₇), 45.0 (C₂₆), 49.4 (C₁₄), 52.9 (C₂), 60.9 (C₆), 62.7 (C₁₁), 67.8 (C₃),

68.6 (C₉), 68.9 (C₁₀), 70.7 (C₄), 75.1 (C₅), 101.3 (C₁), 115.8 (C₂₀), 119.0 (C₁₈), 123.8 (C₂₄),

124.9 (C₁₃), 128.5 (C₁₇, C₂₂), 128.8 (C₁₉), 129.5 (C₂₅), 130.1 (C₂₃), 133.5 (C₁₆), 142.9 (C₁₂),

150.6 (C₂₁), 170.2 (C₈). **LC-MS** (m/z) [M + H⁺]: calculated 657.2; found 657.3



7.3. Expression and purification of LgtC

E. coli clone NMC-41, previously transformed with plasmid containing sequences encoding His-tagged LgtC and ampicillin resistance, was used for the expression of LgtC. (Wakarchuk, Cunningham, Watson, Young, 1998).

Materials: 1) Nutrient agar plate containing 100 µg/mL ampicillin 2) TB + M9 salts media; 3) M9 supplement + ampicillin; 4) 0.5 M IPTG; 5) Disruption buffer.

Equipment: 1) Sterile working conditions / tips / plates / Falcons 2) Erlenmeyer flasks 3) Incubator 4) Plastic spectrophotometry cuvettes 5) 500 mL centrifuge tubes used on Beckman Coulter J6-M1 Centrifuge 6) 50 mL centrifuge tubes used on Beckman Coulter Avanti J-26-XP Centrifuge

Procedures: To 100 mL of sterile TB+M9 ampicillin liquid media in a 500 mL baffled Erlenmeyer flask, single colony of the freshly grown *E. coli* from the agar plate is added. The starter culture is grown overnight in incubator Thermo Scientific MaxQ 8000 set to 37°C and 200 rpm. In the morning, the starter culture is diluted to OD₆₀₀ of 0.1-0.2 with addition of 500 mL of sterile TB+M9 ampicillin liquid media and transferred into 2000 mL baffled Erlenmeyer flasks. Then the main culture is grown at 37°C, 200 rpm until OD₆₀₀ reaches 0.6-1.0. The cells are then inducted with IPTG (0.5 mM) to start the expression of LgtC. Cells are then grown at 37°C, 200 rpm overnight. In the morning, cells are harvested by centrifugation at 4000 rpm for 10 min and washed with disruption buffer. The cell pellet can now be stored by freezing at -80°C or used directly in the next step (cell disruption). The frozen cell pellet is re-suspended in disruption buffer, using around 10mL of buffer per gram (wet cell weight) of cell. Cells are disrupted using cell disruptor, followed with centrifugation at 4°C, 20,000 rpm for 1.5 h. Supernatant is collected and filtered through 0.2µm filter tips. The protein is purified by nickel column (HisTrap™ HP) on AKTA start. Pure protein is collected in 15mL falcon tubes pre-filled with 6mL of 10mM MnCl₂ solution. After the purification, protein is dialyzed overnight. Any precipitate is filtered, the protein is aliquoted, flash frozen in liquid nitrogen and stored at -80°C.

Recipes:

- 1) TB + M9 salts liquid media (1100 mL): 13.2 g tryptone, 26.3 g yeast extract, 4.4 mL glycerol and 11 g M9 salts are dissolved in 1100 mL of deionized water. The

media is then transferred to Duran bottles and autoclaved. Under sterile conditions, desired volume of the liquid media is added into sterile Erlenmeyer flasks, followed by addition of M9 salt supplements.

- 2) M9 salt supplements (for 100mL of media): 1M MgCl₂ (0.1mL), 0.1M CaCl₂ (0.1mL), 10mg/ml vitamin B1 (0.05mL), 20% w/w glucose (1mL), 20% w/w casamino acids (2mL), 100mg/ml ampicillin salt (0.1mL).
- 3) Disruption buffer (500 mL): 77 mg DTT (10 mM), 2.09 g MOPS (20 mM), 14.6 g NaCl (500 mM), 170 mg imidazole (5 mM) and 1 protease inhibitor tablet (without EDTA) are dissolved into 500 mL deionized water. The pH is adjusted to 7.0 with 1 M NaOH.
- 4) Nickel column purification buffers: *Loading*: MOPS (20mM), NaCl (500mM), Imidazole (5mM), pH=7.5 (adj. w/ 1MNaOH). *Washing*: MOPS (20mM), NaCl (500mM), Imidazole (20mM), pH=7.5 (adj. w/ 1MNaOH). *Eluting*: MOPS (100mM), NaCl (500mM), Imidazole (500mM), (pH no adjustment)
- 5) Dialysis buffers: *Dialysis 1 (2L, 1hr)*: Imidazole (250mM), MOPS (50mM), NaCl (250mM), MnCl₂ (5 mM). *Dialysis 2 (2L, 1hr)*: Imidazole (100mM), MOPS (20mM), NaCl (100mM), MnCl₂ (5 mM). *Dialysis 3 (2L, overnight)*: MOPS (20mM), MnCl₂ (5 mM).

7.4. Biochemical colorimetric assay: general protocols

General. All reagents were obtained commercially and used as received, unless otherwise stated. The malachite green reagents were prepared according to the method of Veldhoven *et al.* Absorbance measurements were carried out on a BMG Labtech POLARstar Optima multiplate reader. The colorimetric glycosyltransferase assay as

adapted from Wu *et al.* All assays were carried out in Nunc clear, flat-bottom 96-well plates.

7.4.1. General assay procedure

LgtC activity was adjusted to 20-50% turnover of UDP-Gal donor. We have previously shown that within this turnover range IC_{50} values are obtained reproducibly. All concentrations given for the assay components are final concentrations. LgtC was activated with a solution of 5 mM of DDT for 30 mins at 30 °C (ratio 1:1). The total volume per well was 150 μ L and comprised of buffer (13 mM HEPES, pH 7.0), $MnCl_2$ (5 mM), chicken egg-white lysozyme (CEL) (1 $mg \cdot mL^{-1}$), calf-intestinal phosphatase (10 $U \cdot mL^{-1}$), Triton X-100 (0.01%), UDP-Gal (28 μ M) lactose (2 mM), LgtC, and either inhibitor in buffer (10% v/v) or buffer only. The concentrations of LgtC, lactose and inhibitor were varied depending on the type of assay. On each plate, a UDP calibration curve (0–12.5 μ M) was included. Wells for the calibration curve included all components of the standard reaction except for Lactose and inhibitor (replaced by HEPES buffer) and UDP-Gal (replaced by UDP). The reaction was incubated for 20 min at 30 °C and stopped by the addition of malachite reagent A (30 μ L). After thorough shaking, malachite reagent B was added (30 μ L) and the absorbance at 620 nm of each well was recorded over 14 cycles. See short description and corresponding table for individual experiment protocols.

7.4.2. Standard procedure for the collection and analysis data

Absorbance at 620 nm (AU) recorded at cycle 9 was converted to the concentration of UDP (mM) formed during the reaction using linear regression from a calibration curve (0–12.5 μ M) constructed for every experiment. Averages and standard deviations were calculated in Microsoft Excel.

Activity assay: Both background and assay values are corrected to zero when [LgtC] =0. The assay is then corrected for the background and each processed [UDP] value is divided by the reaction time (20 mins) to obtain the velocity of the enzyme at different LgtC concentration. The average velocity is then plotted against the dilution factor of the enzyme and a linear curve is obtained.

Inhibition assay: A negative control (0 μ M inhibitor) and a blank (0 mM inhibitor, 0 mM acceptor) is included on each plate in triplicate. After linear regression, subtracting the blank from the negative control affords the assay window. The background value for each inhibitor concentration (no acceptor, but otherwise identical components) is subtracted from each inhibitor concentration data point. Once corrected for the background, the absorbance in the presence of each inhibitor concentration is divided by the assay window and represented as a percentage. This percentage is plotted against log [inhibitor] and analysed using Graph-Pad Prism 6 software to afford the relative IC₅₀ values if the data represented a sigmoidal curve.

Substrate assay: Both control and assay are corrected for the background and the average of each duplicate is plotted against the concentration of acceptor substrates. The [UDP formed] was plotted against the [substrate] using Graph-Pad Prism 6.

7.4.3. Activity assay

Aliquots (15 μ L) of MnCl₂, CEL, CIP with lactose (30 μ L) (replaced with HEPES buffer in columns 2-5) and HEPES buffer (45 μ L) (stock solutions: 5x concentration stated in general assay protocol for lactose, 10x for all other components). Six different known enzyme concentrations (15 μ L) were added in columns 4 to 7. Reaction was started with the addition of UDP-Gal (15 μ L, 28 μ M) (stock solution 280 μ M) in columns 4 to 7. Reaction mixture (150 μ L) is then treated as stated in general assay procedure.

Table 7.1. Well map for activity assay

	2	3	4	5	6	7	increasing dilution factor of the enzyme ↓ 0
	Calibration curve		Background		Assay		
	[UDP] (μM)		UDP-Gal ($28 \mu\text{M}$)				
B	12.5	12.5	- Lac	- Lac	+ Lac	+ Lac	
C	6.25	6.25	- Lac	- Lac	+ Lac	+ Lac	
D	3.13	3.13	- Lac	- Lac	+ Lac	+ Lac	
E	1.56	1.56	- Lac	- Lac	+ Lac	+ Lac	
F	0.78	0.78	- Lac	- Lac	+ Lac	+ Lac	
G	0	0	- Lac	- Lac	+ Lac	+ Lac	

7.4.4. Inhibition assay

Aliquots (15 μL) of DTT activated LgtC (concentration previously determined regarding the results of the activity assay, to achieve 20-50% turnover), Triton X-100, MnCl_2 , CEL, CIP, buffer and inhibitor (of increasing concentration) were added in columns 4 to 11 (replaced by 15 μL of buffer in columns 2 and 3). Aliquot of Lactose acceptor (30 μL) was also added (replaced with buffer in all background and calibration curve wells) (stock solutions: 5x concentration stated in general assay protocol for lactose, 10x for all other components). Reaction was started with the addition of UDP-Gal (15 μL , 28 μM) (stock solution 280 μM) in columns 4 to 11. Reaction mixture (150 μL) is then treated as stated in general assay procedure.

* For the enzyme-inhibitor pre-incubation assay, UDP-Gal (15 μL) is added at the same time as the other components and the acceptor/buffer is added after desired incubation time, in order to start the enzymatic reaction.

Table 7.2. Well map for inhibition assay

	2	3	4	5	6	7	8	9	10	11	
	Calibration curve [UDP] (μM)		UDP-Gal, (28 μM)								
			0	Increasing [inhibitor] \longrightarrow							
B	12.5	12.5	+ Lac	+ Lac	+ Lac	+ Lac	+ Lac	+ Lac	+ Lac	+ Lac	Assay
C	6.25	6.25	+ Lac	+ Lac	+ Lac	+ Lac	+ Lac	+ Lac	+ Lac	+ Lac	
D	3.13	3.13	+ Lac	+ Lac	+ Lac	+ Lac	+ Lac	+ Lac	+ Lac	+ Lac	
E	1.56	1.56	- Lac	- Lac	- Lac	- Lac	- Lac	- Lac	- Lac	- Lac	BG
F	0.78	0.78	- Lac	- Lac	- Lac	- Lac	- Lac	- Lac	- Lac	- Lac	
G	0	0	- Lac	- Lac	- Lac	- Lac	- Lac	- Lac	- Lac	- Lac	

7.4.5. CIP control experiment

Aliquots (15 μL) of MnCl_2 , CEL, Triton X-100 diluted in HEPES buffer (60 μL) were combined with CIP (15 μL) (replaced with buffer in background wells). Increasing concentration of inhibitor in HEPES buffer (15 μL) or buffer only (15 μL , control) was added. The reaction was started by addition of UDP (15 μL , 10 μM) or HEPES buffer (15 μL , background). Reaction mixture (150 μL) is then treated as stated in general assay procedure.

Table 7.3. Well map for inhibition of CIP (control)

	2	3	4	5	6	7	8	9	10	11	
	Calibration curve [UDP] (μM)		UDP (10 μM)								
			0	Increasing [inhibitor] \longrightarrow							
B	12.5	12.5	+ CIP	+ CIP	+ CIP	+ CIP	+ CIP	+ CIP	+ CIP	+ CIP	Assay
C	6.25	6.25	+ CIP	+ CIP	+ CIP	+ CIP	+ CIP	+ CIP	+ CIP	+ CIP	
D	3.13	3.13	+ CIP	+ CIP	+ CIP	+ CIP	+ CIP	+ CIP	+ CIP	+ CIP	
E	1.56	1.56	- CIP	- CIP	- CIP	- CIP	- CIP	- CIP	- CIP	- CIP	BG
F	0.78	0.78	- CIP	- CIP	- CIP	- CIP	- CIP	- CIP	- CIP	- CIP	
G	0	0	- CIP	- CIP	- CIP	- CIP	- CIP	- CIP	- CIP	- CIP	

7.4.6. Substrate assay

Aliquots (15 μL) of DTT activated LgtC (concentration previously determined regarding the results of the activity assay, to achieve 20-50% turnover), MnCl_2 , CEL and CIP (stock solutions: 10x concentration detailed in general assay protocol) and HEPES buffer (45 μL) were added in columns 4 to 9. Increasing concentration acceptor substrates (30 μL) were

added: natural lactose in column 6 and 7 (positive control, duplicate), substrate candidate in column 8 and 9 (assay, duplicate) and HEPES buffer in column 4 and 5 (negative control, background). To start the reaction, UDP-Gal (15 μ L, 28 μ M) (stock solution 280 μ M) was added to columns 4 to 9. Reaction mixture (150 μ L) is then treated as stated in general assay procedure.

Table 7.4. Well map for substrate assay

	2	3	4	5	6	7	8	9
	Calibration curve		Background		Control		Assay	
	[UDP] (μ M)		UDP-Gal (28 μ M)					
					+ Lactose	+ Lactose	+	+
B	12.5	12.5	x	x	2 mM	2 mM	↑	↑
C	6.25	6.25	x	x	1 mM	1 mM		
D	3.13	3.13	x	x	0.5 mM	0.5 mM		
E	1.56	1.56	x	x	0.25 mM	0.25 mM		
F	0.78	0.78	x	x	0.125 mM	0.125 mM		
G	0	0	x	x	0 mM	0 mM	0	0

7.4.7. Dialysis

LgtC (activated with 5 mM DTT) was pre-incubated with **11** (1 mM in HEPES buffer 10% v/v) or HEPES buffer (control), in the presence of UDP-Gal (28 μ M), for 10 min at 30 °C. Lactose (2 mM) and others assay components (MnCl₂, CIP, CEL, Triton) were added, and the reactions were incubated for 20 min at 30 °C. These samples were set up in duplicates for (i) the determination of enzyme activity before diafiltration under standard assay conditions (ii) the determination of enzyme activity after diafiltration. Reaction mixture for (ii) (3mL) was collected in Vivaspin concentrator tubes (pore size 10,000 MWCO) and concentrated (centrifugation at 4 °C, 4000 rpm, 1 hour). The residual volume (350 μ L) was washed with HEPES buffer (13mM, pH 7.0, 2 mL). Concentration and washes were repeated 2 more times. The final residual volume (230 μ L) was diluted to 420 μ L with HEPES buffer (13mM, pH 7.0), and enzyme activity after diafiltration was determined under standard assay conditions. Each experiment was carried out in triplicate (100% activity = complete conversion of UDP-Gal).

7.4.8. Substrate kinetics

Activated LgtC was incubated with **11** (0-1 mM), MnCl₂ (5 mM), CIP (10 U/ml), CEL (1 mg/ml), Triton (0.01%) fixed concentration of partner substrate and increasing concentration of studied substrate at 30 °C in HEPES buffer (13mM, pH 7.0). The velocity was calculated over three times points 0, 10 and 20 mins and corresponds to the gradient $d[\text{UDP}]_{\text{formed}}/dt$ after background subtraction. Malachite A was added after 0, 10 or 20 mins of incubation in corresponding wells to stop the reaction and malachite B was added in all wells after 20 mins. The absorbance at 620 nm of each well was then recorded over 14 cycles. UDP-Gal turnover was kept <10%. Data were fitted to a Michaelis Menten plot with GraphPad Prism6.

Table 7.5. Competition for Lactose (acceptor): [UDP-Gal] = 100µM, [lactose acceptor] = 1-15 mM (each plate corresponds to one concentration of acceptor substrate tested)

	2	3	4	5	6	7	8	9	10	11	12	
	Calibration curve		BG	Assay			BG	Assay		BG	Assay	[11]
	[UDP] (µM)	UDP-Gal (100 µM)										
B	12.5	12.5	-Acc	+Acc	+Acc	-Acc	+Acc	+Acc	-Acc	+Acc	+Acc	↑
C	6.25	6.25	-Acc	+Acc	+Acc	-Acc	+Acc	+Acc	-Acc	+Acc	+Acc	
D	3.13	3.13	-Acc	+Acc	+Acc	-Acc	+Acc	+Acc	-Acc	+Acc	+Acc	
E	1.56	1.56	-Acc	+Acc	+Acc	-Acc	+Acc	+Acc	-Acc	+Acc	+Acc	
F	0.78	0.78	-Acc	+Acc	+Acc	-Acc	+Acc	+Acc	-Acc	+Acc	+Acc	
G	0	0	-Acc	+Acc	+Acc	-Acc	+Acc	+Acc	-Acc	+Acc	+Acc	
			0 mins			10 mins			20 mins			

Table 7.6. Competition for UDP-Gal (donor): [Lactose] = 5mM, [UDP-Gal] = 0-100 mM (each plate corresponds to one concentration of acceptor substrate tested)

	2	3	4	5	6	7	8	9	10	11	12	
	Calibration curve		BG	Assay			BG	Assay		BG	Assay	[11]
	[UDP] (µM)	Lactose (5mM)										
B	12.5	12.5	-Don	+Don	+Don	-Don	+Don	+Don	-Don	+Don	+Don	↑
C	6.25	6.25	-Don	+Don	+Don	-Don	+Don	+Don	-Don	+Don	+Don	
D	3.13	3.13	-Don	+Don	+Don	-Don	+Don	+Don	-Don	+Don	+Don	
E	1.56	1.56	-Don	+Don	+Don	-Don	+Don	+Don	-Don	+Don	+Don	
F	0.78	0.78	-Don	+Don	+Don	-Don	+Don	+Don	-Don	+Don	+Don	
G	0	0	-Don	+Don	+Don	-Don	+Don	+Don	-Don	+Don	+Don	
			0 mins			10 mins			20 mins			

7.4.9. WH reactivity monitoring by ^1H NMR

A solution of L-cysteine and the requisite WH (full inhibitor or model) (ratio 1:1, 80mM potassium phosphate buffer/2MM DSS- d_4 D_2O , pH 7) was made up in a glass NMR tube. The solution was thoroughly mixed by sonication and the ^1H NMR spectrum of the reaction mixture was recorded at 298K on a Bruker BioSpin machine at 400MHz at various time points (0-16hrs). The ^1H NMR spectrum of L-cysteine and each electrophile (full inhibitor or model WH) alone was also recorded under the same conditions as controls.

7.5. Crystallography

(1) Frozen LgtC batch (2mg/mL, 5mL) was dialysed in 2L crystallisation buffer for 8 hours and submitted to the following treatments.

(2) Freshly expressed LgtC purified on Nickel column was purified size exclusion chromatography on Superdex 74pg Prepacked 16x600mm column eluted with crystallisation buffer hours and submitted to the following treatments.

The enzyme was then concentrated in Vivaspin concentrator tubes (pore size 10,000 MWCO) (4krpm, 4°C) until concentration reached 20mg/mL (on LabTech nanodrop ND800). A stock of inhibitor **11** (40mM) was made in crystallisation buffer. LgtC and **11** were incubated at rt for 1 hour in a 1:1 ratio. The mixture was re concentrated in Vivaspin concentrator tubes (pore size 10,000 MWCO) (4krpm, 4°C) for 10 mins (protein concentration=15mg/mL on nanodrop). On a 24 well hanging drop crystallization plate, reservoir solutions were prepared as in Table 7.7. Droplets (2 μL) were allowed to equilibrate against the reservoir well solutions (2 μL) as hanging drops and were streak seeded. Plate were stored at 16°C.

Table 7.7. Fine screens on reported crystallisation conditions. PEG: polyethylene glycol 2000 monomethyl ether. Solvent: 0.1 M Sodium acetate pH in all wells

5 % PEG pH 4.0	10 % PEG pH 4.0	15 % PEG pH 4.0	20 % PEG pH 4.0	25 % PEG pH 4.0	30 % PEG pH 4.0
5 % PEG pH 4.5	10 % PEG pH 4.5	15 % PEG pH 4.5	20 % PEG pH 4.5	25 % PEG pH 4.5	30 % PEG pH 4.5
5 % PEG pH 5.0	10 % PEG pH 5.0	15 % PEG pH 5.0	20 % PEG pH 5.0	25 % PEG pH 5.0	30 % PEG pH 5.0
5 % PEG pH 5.5	10 % PEG pH 5.5	15 % PEG pH 5.5	20 % PEG pH 5.5	25 % PEG pH 5.5	30 % PEG pH 5.5

On a 96 well seating drops crystallisation plate (one for each of the commercially available screening kits INDEX, PACT, JSCG and SALTRX) droplets (150nL) were allowed to equilibrate against the reservoir well solutions (150nL) as seating drops. This was operated on LabTech Mosquito LCP Robot. Plates were stored at 16°C.

Crystallisation buffer: 3mM TCEP, 5mM MnCl₂, 50mM NH₄OAc, pH7

7.6. Monitoring of attachment of linker on *cmpd 9* by HPLC

To determine their retention times, purified samples of **12- α** and **12- β** were analysed by analytical HPLC (SB300-C18 column, injection volume: 10 μ L, gradient: 10-70% MeOH (0.1% TFA) against H₂O (0.1% TFA) over 10 min, flow rate 1 mL/min, detecting at 210 nm).

In duplicate, **9** (51.1 mg, 0.12 mmol, 1eq) was dissolved in DCM (400 μ L), cooled to 0°C and boron trifluoride diethyl etherate (58.3 μ L, 0.47 mmol, 4eq) and propynol ethoxylate (34.2 μ L, 0.35 mmol, 3eq) added dropwise. One reaction was stirred at room temperature while the other was stirred at 40°C. For each reaction, at pre-determined time intervals (0.3, 3, 8, 16, 24 and 75 h) an aliquot (20 μ L) was taken and extracted in EtOAc (180 μ L) with H₂O (1 mL). 10 μ L of each organic layer was removed and diluted in EtOAc (40 μ L). Each sample was analysed by HPLC (method as above) to determine relative amounts of **12- α** and **12- β** present.

7.7. Fluorescence scanning

Emission spectra were recorded on a Varian Cary Eclipse Spectrometer using a Type 3-Q-10mm fluorimeter rectangular quartz cell.

HEPES buffer titration: Increasing amounts of HEPES buffer were added to a solution of **15 α / β** in DMSO at fixed concentration (10 μ M) (as per Table 7.8). Total volume = 2mL. Emission spectra was generated as previously described.

Table 7.8. HEPES buffer titration for fluorescence measurements

Sample	15 α / β (μ L) (0.1mM in DMSO)	DMSO (μ L)	HEPES buffer (μ L)	% HEPES buffer
1	200	1800	0	0
2	200	1600	200	10
3	200	1300	500	25
4	200	800	1000	50
5	200	300	1500	75
6	200	0	1800	90

7.8. LgtC-labelling experiments

Equipment. SDS-PAGE electrophoresis was performed using pre-cast Novex™ WedgeWell™ 12% Tris-Glycine gels (1.0mm x 10 well), with prepared 1X Tris-Glycine SDS running buffer at 175V, against Fisher BioReagents™ EZ-Run™ Prestained *Rec* Protein Ladder (10-170 kDa). In-gel UV scanning for gel visualisation was performed using a Bio-Rad ChemiDoc™ MP Imaging system. Coomassie staining was achieved using Generon Quick Coomassie stain.

Solutions

- Stock solutions (5mM) of **15- α** , **15- β** , **17** and **13- α** were prepared in HEPES buffer (pH 7), aliquoted (10 μ L) and stored in -20°C.
- 10X Tris-Glycine SDS running buffer: 10.0 g SDS, 30.3 g Tris and 144.0 g glycine were dissolved in 1 L deionised H₂O.

- 1X Tris-Glycine SDS running buffer (electrophoresis buffer): 100 mL 10X Tris-Glycine SDS running buffer was diluted in 900 mL deionised H₂O.
 - Loading buffer: 0.8 g SDS, 4.0 mL glycerol and 2.4 mL Tris buffer (1M, pH 6.8) was dissolved in 3.6 mL deionised H₂O.
 - HEPES buffer: 1.55 g HEPES was dissolved in 400 mL deionised H₂O.
- pH 3: Above adjusted to pH3 by addition of 1M HCl solution and made to 500mL total volume with deionised H₂O.
- pH 7, 9, 11 and 13: Above adjusted to correct pH by addition of 1M NaOH solution and made to 500mL total volume with deionised H₂O.

7.8.1. General procedure for labelling of recombinant LgtC

1-step labelling of recombinant LgtC. Solutions are kept on ice at all time. Reaction mixture (10 μ L) is made with **15- α / β** (1 μ L of 5mM stock) (replaced by H₂O in control lane), LgtC (1 μ L of 300/150 μ M stock), potential additives depending on experiment (1 μ L of 10x stock solutions) and made up to 10 μ L with HEPES buffer. Once prepared, the samples are shaken gently and incubated for 30 mins at 30°C, with gentle shaking at 5 min intervals. Loading buffer (2.5 μ L) * is added, reaction mixture thoroughly mixed, and the samples incubated for a further 15 min at 30°C. Samples were analysed by SDS-PAGE electrophoresis under the conditions and using the equipment mentioned above. Lane 1 always contains the protein ladder (5 μ L).

The gel was rinsed with H₂O and imaged by in-gel UV scanning, with 5 minutes activation prior to exposure aiding visualisation of bands. Sequentially, gels are stained with Coomassie Blue via soaking in Quick Coomassie stain on a platform rocker for 2 hours.

* In Figure 4.11.C LgtC in first incubated with loading buffer for 15 min at 30°C and then incubated with probe/water for 30 mins at 30°C.

2-step labelling of recombinant LgtC. Recombinant LgtC was expressed and purified as previously discussed. Four samples were prepared according to Table 7.9. This protocol contains two main steps: (A) incubation of partial probe **13- α** (or HEPES buffer for control) with LgtC to achieve the covalent modification of the enzyme. (B) *In situ* click ligation for the attachment of the reporter group.

Table 7.9. Samples for 2-step labelling of recombinant LgtC

	STEP (A)			STEP (B)	
Sample A	LgtC	+	HEPES buffer	DMSO	CONTROLS
Sample B	LgtC	+	HEPES buffer	16	
Sample C	LgtC	+	HEPES buffer	17	
Sample D	LgtC	+	13-α	16	ASSAY

Firstly, the 4 samples for step (A) were prepared as shown in Table 7.10. They are then incubated at 30°C for 30 mins with gentle shake every 5 min. Upon completion, the mixture was transferred into Amicon ® Ultra - 0.5mL centrifugal filters (MW cut-off: 10 kDa) and the samples were centrifuged at 4°C, 4000 rpm for 10mins. The residue was washed three times with HEPES buffer (0.5mL, 13 mM, pH 7.0). The whole centrifugation step took about 45 mins and a 50-55 μ L residue was retrieved and put through step (B) conditions (see Table 7.11 and Table 7.12). Equal volumes of the THTPA and CuSO₄ solutions were premixed together (becomes blue). The Click reaction was started by addition of 50 μ L sodium ascorbate. Subsequently, the reactions were incubated at 30°C for 1 hour and gently shaken every 10 mins. Once finished, samples were transferred into four new Amicon ® Ultra - 0.5mL centrifugal filters (MW cut-off: 10 kDa) and centrifuged at 4°C, 4000 rpm for 10 mins. Samples were then washed with HEPES - EDTA (HEPES 13 mM, EDTA 5 mM, pH 7.0) buffer for three times. The remainder (40 μ L) was carefully collected in 0.5 mL Eppendorf tubes. Samples were denatured by adding loading buffer (10 μ L), followed by 15 min incubation at 30°C. Denatured samples were loaded on SDS page gel, developed and analysed as described in the equipment section.

Table 7.10. (A) Covalent modification of LgtC

	LgtC	13-α (alkyne) 10mM stock	HEPES buffer
A	100 μ L	-	400 μ L
B	100 μ L	-	400 μ L
C	100 μ L	-	400 μ L
D	100 μ L	25 μ L	375 μ L

Table 7.11. (B) Click ligation *in situ*

	Mixture from (A)	HEPES buffer	DMSO	17	16	CuSO₄/THPTA	Sodium ascorbate
A	50 μ L	275 μ L	25 μ L	-	-	100 μ L	50 μ L
B	55 μ L	270 μ L	-	-	25 μ L	100 μ L	50 μ L
C	50 μ L	275 μ L	-	25 μ L	-	100 μ L	50 μ L
D	55 μ L	270 μ L	-	-	25 μ L	100 μ L	50 μ L

Table 7.12. Reagents preparation for 2-step labelling of recombinant LgtC

Reagents	Stock concentration	Final concentration	Solvent
13-α	10mM	500 μ M	HEPES buffer
THTPA	12.5mM	1.25mM	HEPES buffer
CuSO ₄ ·5 H ₂ O	2.5mM	250 μ M	H ₂ O
16 / 17	10mM	500 μ M	DMSO
Sodium ascorbate	50mM	5mM	HEPES buffer

7.8.2. LC-MS/MS analysis for identification of target residue

Enzymatic Digestion: The gel band containing the LgtC protein and probe was not subjected to cysteine residue reduction and alkylation prior to digestion with trypsin. Trypsin digestion (0.8mg; Bovine from Sigma Aldrich) was carried out overnight at room temperature after initial incubation at 37°C for 2 hours.

LC-MS/MS: Peptides were extracted from the gel pieces by a series of acetonitrile and aqueous washes. The extract was pooled with the initial supernatant and lyophilised. The sample was then resuspended in 40 μ l of resuspension buffer (2% ACN in 0.05% FA) and analysed by LC-MS/MS (10 μ l). Chromatographic separation was performed using a

U3000 UHPLC NanoLC system (ThermoFisherScientific, UK). Peptides were resolved by reversed phase chromatography on a 75µm C18 column (50cm length) using a three-step linear gradient of 80% acetonitrile in 0.1% formic acid. The gradient was delivered to elute the peptides at a flow rate of 250nl/min over 60 min. The eluate was ionised by electrospray ionisation using an Orbitrap Fusion Lumos (ThermoFisherScientific, UK) operating under Xcalibur v4.1.5. The instrument was programmed to acquire in automated data-dependent switching mode, selecting precursor ions based on their intensity for sequencing by collision-induced fragmentation using a TopN CID method. The MS/MS analyses were conducted using collision energy profiles that were chosen based on the mass-to-charge ratio (m/z) and the charge state of the peptide.

Database Searching: Raw mass spectrometry data were processed into peak list files using Proteome Discoverer (ThermoScientific; v2.2). The raw data file was processed and searched using the Sequest search algorithm (Eng *et al.*; PMID 24226387) against the sequence of LgtC protein retrieved from Uniprot (Q93EK7_NEIME). This increased the chance of correct assignments of the fragmentation spectra and allowed for the reduction of false positive identifications.

7.8.3. Labelling of non-recombinant LgtC: through cells or in cell lysates

Labelling of intact *E. coli* cells overexpressing recombinant LgtC. *E. coli* cells were grown as previous described. Three main cultures (450mL) were inoculated from a single starter culture and were shaken at 37°C for 2 hours. Culture 1 (OD= 0.7) was harvested without being IPTG-induced (control culture) while the other two were induced with IPTG (0.5mM). Culture 2 was harvested 3 hours after induction (OD= 4.4) and culture 3, 8hours after induction (OD= 11). Pellet (850mg, 4g, 6g respectively) was suspended in HEPES buffer (6mL, 28mL, and 35mL respectively). The suspension was then transferred into 1.5mL Eppendorfs, samples were centrifuged, supernatant discarded and cell pellet

flash frozen and store in -80°C . (mass of cells in each sample =100mg approx.). Frozen cells (1 sample divided in 4) were re-suspended in HEPES buffer (13 mM, pH 7.0, OD_{600} =10-15). Bacterial cells (900 μL , 25mg of cells) are incubated with probe/control (100 μL of 10mM stock solution, 10% DMSO) at 30°C for 3h (or chosen incubation time when studying its impact on labelling, see chapter 4). Samples were centrifuged at 4,000 rpm 4°C , and cell pellets were washed with 10% DMSO HEPES buffer once and HEPES buffer twice (until all the free probe has been washed away, detected under UV lamp, 365nm). 300 μL of 1x Bug Buster (diluted from 10X Bug Buster stock with HEPES buffer) was added into each sample. The cell suspension was incubated on a rotating mixer at a slow setting for 20 mins at rt. Next, samples were centrifuged at 12,000 rpm, 4°C for 10 mins. Clear supernatant was collected, loading buffer (2.5 μL) was added to a 10 μL aliquot of each sample for protein denaturation followed by 15 min incubation at 30°C (the rest of the sample was kept at -20°C for potential further use). Denatured samples were loaded on SDS page gel, developed and analysed as described in the equipment section.

For the labelling of LgtC overexpressing *E. coli*

cell lysates. Frozen cells generated as previously described were re-suspended in HEPES buffer (13 mM, pH 7.0, OD_{600} =10-15) and divided in 8 samples to which was added 150 μL of 1x Bug Buster (diluted from 10X Bug Buster stock with HEPES buffer). The cell suspension was incubated on a rotating mixer at a slow setting for 20 mins at rt. Next, samples were centrifuged at 12,000 rpm, 4°C for 10 mins. Clear supernatant was collected and was incubated (9 μL) with probe/control (1 μL of 10x stock solution) for 1 hour at 30°C (the rest of the lysate was flash frozen and stored at -80°C for potential further use). Loading buffer (2.5 μL) was added for protein denaturation followed by 15 min incubation at 30°C . Denatured samples were loaded on SDS page gel, developed and analysed as described in the equipment section.

7.9. Bacterial growth curve

H. influenzae R2866 (from glycerol stock stored at -80°C) was grown overnight into 10mL BHI broth (supplemented Brain heart infusion broth: 15 g Brain heart infusion in 400 mL deionised water, 0.4µL/mL NAD, and 10µL/mL Hemin) at 37°C 180 rpm. In the morning 300µL of overgrown culture (OD₅₉₀=2) was inoculated in fresh BHI (10mL) broth and incubated at 37°C 180 rpm. OD₅₉₀ was recorded at various time points. The data were tested in triplicate and presented as average.

7.10. Labelling of *H. influenzae* R2866 proteins

Equipment:

- NuPAGE™ 4-12% Bis-Tris Protein Gels, 1.0 mm, 12-well
- PageRuler™ Prestained Protein Ladder 15 – 190 kDa
- NuPAGE™ MES SDS Running Buffer (20X)
- Invitrogen PowerEase® 500 Power Supply

7.10.1. Protein labelling in intact cells

Cell cultures of *H. influenzae* R2866 (from glycerol stock stored at -80°C) were grown in 5x10mL sBHI broth overnight. In the morning 300µL of overgrown culture (OD₅₉₀=2) was inoculated in fresh sBHI (10mL) broth and incubated at 37°C 180 rpm. When OD₆₀₀=0.9, 4x1mL of cell cultures were transferred into 7 eppendorfs (1.5mL) (3 controls, 4 probes). The supernatant (media) was discarded and the cell pellets resuspended into 900µL of PBS and 100µL of probe/control/DMSO solution (10x desired concentration in DMSO). The samples were incubated for 1 / 2 hours at 30°C with regular mixing. The samples were then centrifuged (2 mins, 12 krpm), the supernatant was separated and kept for imaging. The cell pellets were washed with a solution of 5% DMSO in PBS, pellets were

resuspended, centrifuged, supernatant was separated. This was repeated four times. The resulting pellets and washes were imaged under UV (365nm). Bug buster (1x in PBB, 100µL) was added to each pellet, the resulting suspension was shaken for 30 mins at rt. The samples were centrifuged (10 mins, 12 krpm) and the supernatant containing the soluble proteins was recovered. Both pellets and soluble fractions were imaged under UV (365nm). Loading buffer (2.5µL) was added to cell lysates (10µL) and incubated for 10 mins at 50°C. The samples (10µL) were then loaded on pre cast SDS-page gel against protein ladder and run in 1X MES running buffer at 150V for 70mins. The gel was then visualised by fluorescence read-out and Coomassie staining.

Viability assays were performed on bacitracin chocolate agar plate (chocolate agar: Columbia blood agar with 10% lysed horse blood) by addition of a 1/10,000 cell suspension (after incubation with probe/control) in sBHI broth (50µL). The cell suspension was streaked on the plate and incubated at 37°C for 48 hours.

7.10.2. Protein labelling in cell lysates

Cell cultures of *H. influenzae* R2866 were grown as described in 7.10.1. The pellets were lysed by addition of Bug buster (1x in PBB, 100µL) as described, prior to any contact with the probes. The resulting lysates were kept on ice and the cell debris was discarded. Lysates (9µL) was incubated with solutions of probes/controls/DMSO (1µL, 10x desired concentration for 1 hour at 30°C with regular mixing. Loading buffer (2.5µL) was added to cell lysates (10µL) and incubated for 10 mins at 50°C. The samples (10µL) were then loaded on pre cast SDS-page gel against protein ladder and run in 1X MES running buffer at 150V for 70mins. The gel was then visualised by fluorescence read-out and Coomassie staining.

7.10.3. Periplasmic extraction

Extraction Buffer: 100g sucrose, 0.1mol/L Tris-base, 0.1mol/L EDTA, 500mL water

Cell cultures of *H. influenzae* R2866 were grown and pellets were incubated with the probes/controls/DMSO and subsequently washed as described in 7.10.1. A. Cell pellets were resuspended in 490 μ L of extraction buffer and 10 μ L of lysosome (50mg/mL solution) and incubated on ice for 20 mins with regular inversion. Cells were pelleted at 12,000g for 25 mins. The supernatant contained the periplasmic fraction. Pellets were resuspended in extraction buffer and subjected to 3 freeze/thaw cycles to liberate the cytoplasm. The lysates were treated, and proteins were separated as previously described.

7.10.4. Protein identification by MS

Enzymatic Digestion: In-gel reduction, alkylation and digestion with trypsin was performed on the seven gel bands prior to subsequent analysis by mass spectrometry. Cysteine residues were reduced with dithiothreitol and derivatised by treatment with iodoacetamide to form stable carbamidomethyl derivatives. Trypsin digestion was carried out overnight at room temperature after initial incubation at 37°C for 2 hours.

LC-MS/MS: Peptides were extracted from the gel pieces by a series of acetonitrile and aqueous washes. The extract was pooled with the initial supernatant and lyophilised. The sample was then resuspended in 40 μ L of resuspension buffer (2% ACN in 0.05% FA) and analysed by LC-MS/MS (10 μ L). Chromatographic separation was performed using a U3000 UHPLC NanoLC system (ThermoFisherScientific, UK). Peptides were resolved by reversed phase chromatography on a 75 μ m C18 column (50cm length) using a three-step linear gradient of 80% acetonitrile in 0.1% formic acid. The gradient was delivered to elute the peptides at a flow rate of 250nl/min over 60 min. The eluate was ionised by electrospray ionisation using an Orbitrap Fusion Lumos (ThermoFisherScientific, UK) operating under Xcalibur v4.1.5. The instrument was programmed to acquire in automated data-dependent switching mode, selecting precursor ions based on their

intensity for sequencing by collision-induced fragmentation using a TopN CID method. The MS/MS analyses were conducted using collision energy profiles that were chosen based on the mass-to-charge ratio (m/z) and the charge state of the peptide.

Database Searching: Raw mass spectrometry data were processed into peak list files using Proteome Discoverer (ThermoScientific; v2.2). The raw data file was processed and searched using the Sequest search algorithm (Eng *et al.*; PMID 24226387) against the current Haemophilus influenzae database from Uniprot (HI; 4957 reviewed entries).

References

1. Fleming, A. & Wong, J. Giants in the Field of Microbiology Series and the Discovery of Penicillin. **10**, 124–126 (2003).
2. Clatworthy, A. E., Pierson, E. & Hung, D. T. Targeting virulence: a new paradigm for antimicrobial therapy. *Nat. Chem. Biol.* **3**, 541–548 (2007).
3. Neu, H. C. The crisis in antibiotic resistance. *Science* **257**, 1064–1073 (1992).
4. O'Neill, J. Antimicrobial Resistance : Tackling a crisis for the health and wealth of nations. *Rev. Antimicrob. Resist.* 1–16 (2014).
5. Tally, F. P. & DeBruin, M. F. Development of daptomycin for Gram-positive infections. *J. Antimicrob. Chemother.* **46**, 523–526 (2000).
6. Leach, K. L., Brickner, S. J., Noe, M. C. & Miller, P. F. Linezolid, the first oxazolidinone antibacterial agent. *Ann. N. Y. Acad. Sci.* **1222**, 49–54 (2011).
7. (EARS-Net), E. A. R. S. N. *Antimicrobial resistance surveillance in Europe. European centre for disease prevention and control* 1–99 (2015).
8. Mulani, M. S., Kamble, E. E., Kumkar, S. N., Tawre, M. S. & Pardesi, K. R. Emerging strategies to combat ESKAPE pathogens in the era of antimicrobial resistance: A review. *Front. Microbiol.* **10**, (2019).
9. Erwin, A. L. *et al.* Role of IgtC in resistance of nontypeable Haemophilus influenzae strain R2866 to human serum. *Infect. Immun.* **74**, 6226–6235 (2006).
10. Dalal, J. *et al.* Development and pre-clinical evaluation of a synthetic oligosaccharide-protein conjugate vaccine against Neisseria meningitidis serogroup C. *Vaccine* **37**, 5297–5306 (2019).
11. Turk, D. C. The pathogenicity of Haemophilus influenzae. *J. Med. Microbiol.* **18**, 1–16 (1984).
12. Yih-Ling Tzeng, D. S. S. Epidemiology and pathogenesis of Neisseria meningitidis. *Microbes Infect.* **2**, 687–700 (2000).
13. P., B. *et al.* Neisseria meningitidis lipopolysaccharides in human pathology. *J. Endotoxin Res.* **7**, 401–420 (2001).
14. Gorla, M. C. *et al.* Emergence of resistance to ciprofloxacin in neisseria meningitidis in Brazil. *J. Med. Microbiol.* **67**, 286–288 (2018).
15. Campos, J. Haemophilus influenzae: From the post-vaccination era to antibiotic resistance. *Clin. Microbiol. Infect.* **7**, 287–290 (2001).
16. Tristram, S., Jacobs, M. R. & Appelbaum, P. C. Antimicrobial resistance in Haemophilus influenzae. *Clin. Microbiol. Rev.* **20**, 368–389 (2007).
17. Chang, C. M. *et al.* Colonisation of fluoroquinolone-resistant Haemophilus influenzae among nursing home residents in southern Taiwan. *J. Hosp. Infect.* **75**, 304–308 (2010).
18. Tomeh, M. O., Starr, S. E., McGowan, J. E., Terry, P. M. & Nahmias, A. J. Ampicillin-Resistant Haemophilus influenzae Type B Infection. *JAMA J. Am. Med. Assoc.* **229**, 295–297 (1974).

19. Dcosta, V. M. *et al.* Antibiotic resistance is ancient. *Nature* **477**, 457–461 (2011).
20. Exner, M. *et al.* Antibiotic resistance: What is so special about multidrug-resistant Gram-negative bacteria? *GMS Hyg. Infect. Control* **12**, Doc05 (2017).
21. Waglechner, N. & Wright, G. D. Antibiotic resistance: It's bad, but why isn't it worse? *BMC Biol.* **15**, 1–8 (2017).
22. Lowe, J. Mechanisms of Antibiotic Resistance. *Microbiol. Spectr.* **4**, 1–24 (2015).
23. Coussens, N. P. *et al.* Better living through chemistry: Addressing emerging antibiotic resistance. *Exp. Biol. Med.* **243**, 538–553 (2018).
24. Schroeder, M., Brooks, B. D. & Brooks, A. E. The complex relationship between virulence and antibiotic resistance. *Genes (Basel)*. **8**, (2017).
25. Rasko, D. A. & Sperandio, V. Anti-virulence strategies to combat bacteria-mediated disease. *Nat. Rev. Drug Discov.* **9**, 117–128 (2010).
26. Calvert, M. B., Jumde, V. R. & Titz, A. Pathoblockers or antivirulence drugs as a new option for the treatment of bacterial infections. *Beilstein J. Org. Chem.* **14**, 2607–2617 (2018).
27. Shoop, W. L. *et al.* Anthrax lethal factor inhibition. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 7958–7963 (2005).
28. Muschiol, S. *et al.* A small-molecule inhibitor of type III secretion inhibits different stages of the infectious cycle of *Chlamydia trachomatis*. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 14566–14571 (2006).
29. Hentzer, M. *et al.* Attenuation of *Pseudomonas aeruginosa* virulence by quorum sensing inhibitors. *EMBO J.* **22**, 3803–3815 (2003).
30. Pinkner, J. S. *et al.* Rationally designed small compounds inhibit pilus biogenesis in uropathogenic bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 17897–17902 (2006).
31. Beceiro, A., Tomás, M. & Bou, G. Antimicrobial resistance and virulence: A successful or deleterious association in the bacterial world? *Clin. Microbiol. Rev.* **26**, 185–230 (2013).
32. Ternent, L., Dyson, R. J., Krachler, A. M. & Jabbari, S. Bacterial fitness shapes the population dynamics of antibiotic-resistant and -susceptible bacteria in a model of combined antibiotic and anti-virulence treatment. *J. Theor. Biol.* **372**, 1–11 (2015).
33. Wang, T. *et al.* Quorum-sensing contributes to virulence, twitching motility, seed attachment and biofilm formation in the wild type strain Aac-5 of *Acidovorax citrulli*. *Microb. Pathog.* **100**, 133–140 (2016).
34. Yeung, A. T. Y., Bains, M. & Hancock, R. E. W. The sensor kinase CbrA is a global regulator that modulates metabolism, virulence, and antibiotic resistance in *Pseudomonas aeruginosa*. *J. Bacteriol.* **193**, 918–931 (2011).
35. Szczepanowski, R. *et al.* The 120 592 bp IncF plasmid pRSB107 isolated from a sewage-treatment plant encodes nine different antiobiotic-resistance determinants, two iron-acquisition systems and other putative virulence-associated functions. *Microbiology* **151**, 1095–1111 (2005).
36. Tytgat, H. L. P. & Lebeer, S. The Sweet Tooth of Bacteria: Common Themes in Bacterial Glycoconjugates. *Microbiol. Mol. Biol. Rev.* **78**, 372–417 (2014).

37. Kay, E. *et al.* Systems analysis of bacterial glycomes. *Biochem. Soc. Trans.* **38**, 1290–1293 (2010).
38. Preston, A., Mandrell, R. E., Gibson, B. W. & Apicella, M. A. The lipooligosaccharides of pathogenic gram-negative bacteria. *Crit. Rev. Microbiol.* **22**, 139–180 (1996).
39. Lukacova, M., Barák, I. & Kazár, J. Role of structural variations of polysaccharide antigens in the pathogenicity of Gram-negative bacteria. *Clin. Microbiol. Infect.* **14**, 200–206 (2008).
40. Moran, A. P., Prendergast, M. M. & Appelmelk, B. J. Molecular mimicry of host structures by bacterial lipopolysaccharides and its contribution to disease. *FEMS Immunol. Med. Microbiol.* **16**, 105–115 (1996).
41. Pupo, E. & Hardy, E. Complexity and solutions to the isolation problem of Gram negative lipopolysaccharides' bacteria molecular species. *Biotechnol. Appl.* **26**, 1–15 (2009).
42. Lairson, L. L., Henrissat, B., Davies, G. J. & Withers, S. G. Glycosyltransferases: Structures, Functions, and Mechanisms. *Annu. Rev. Biochem.* **77**, 521–555 (2008).
43. Breton, C., Fournel-Gigleux, S. & Palcic, M. M. Recent structures, evolution and mechanisms of glycosyltransferases. *Curr. Opin. Struct. Biol.* **22**, 540–549 (2012).
44. Wagner, G. K. & Pesnot, T. Glycosyltransferases and their Assays. *ChemBioChem* **11**, 1939–1949 (2010).
45. Krupička, M. & Tvaroška, I. Hybrid quantum mechanical/molecular mechanical investigation of the β -1,4-galactosyltransferase-I mechanism. *J. Phys. Chem. B* **113**, 11314–11319 (2009).
46. 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).
47. Liang, D.-M. *et al.* Glycosyltransferases: mechanisms and applications in natural product development. *Chem. Soc. Rev.* **44**, 8350–74 (2015).
48. Breton, C., Šnajdrová, L., Jeanneau, C., Koča, J. & Imberty, A. Structures and mechanisms of glycosyltransferases. *Glycobiology* **16**, 29–37 (2006).
49. 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).
50. Gloster, T. M. Advances in understanding glycosyltransferases from a structural perspective. *Curr. Opin. Struct. Biol.* **28**, 131–141 (2014).
51. Liu, J. & Mushegian, A. Three monophyletic superfamilies account for the majority of the known glycosyltransferases. *Protein Sci.* **12**, 1418–1431 (2003).
52. Lovering, A. L., De Castro, L. H., Lim, D. & Strynadka, N. C. J. Structural insight into the transglycosylation step of bacterial cell-wall biosynthesis. *Science (80-.)*. **315**, 1402–1405 (2007).
53. Griffin, R. *et al.* Digalactoside expression in the lipopolysaccharide of *Haemophilus influenzae* and its role in intravascular survival. *Infect. Immun.* **73**, 7022–7026 (2005).
54. Persson, K. *et al.* Crystal structure of the retaining galactosyltransferase LgtC from

- Neisseria meningitidis* in complex with donor and acceptor sugar analogs. *Nat. Struct. Biol.* **8**, 166–175 (2001).
55. Bayliss, C. D., Field, D. & Moxon, E. R. The simple sequence contingency loci of *Haemophilus influenzae* and *Neisseria meningitidis*. *J. Clin. Invest.* **107**, 657–662 (2001).
 56. Van Der Woude, M. W. & Bäumlér, A. J. Phase and antigenic variation in bacteria. *Clin. Microbiol. Rev.* **17**, 581–611 (2004).
 57. Zhu, P. *et al.* Genetic diversity of three lgt loci for biosynthesis of lipooligosaccharide (LOS) in *Neisseria* species. *Microbiology* **148**, 1833–1844 (2002).
 58. Wakarchuk, W. W., Cunningham, A., Watson, D. C. & Young, N. M. Role of paired basic residues in the expression of active recombinant galactosyltransferases from the bacterial pathogen *Neisseria meningitidis*. *Protein Eng.* **11**, 295–302 (1998).
 59. Ly, H. D., Loughéed, B., Wakarchuk, W. W. & Withers, S. G. Mechanistic studies of a retaining alpha-galactosyltransferase from *Neisseria meningitidis*. *Biochemistry* **41**, 5075–5085 (2002).
 60. Tvaroška, I. Molecular modeling insights into the catalytic mechanism of the retaining galactosyltransferase LgtC. *Carbohydr. Res.* **339**, 1007–1014 (2004).
 61. Chan, P. H. W., Weissbach, S., Okon, M., Withers, S. G. & McIntosh, L. P. Nuclear Magnetic Resonance Spectral Assignments of alpha-1,4-Galactosyltransferase LgtC from *Neisseria meningitidis*: Substrate Binding and Multiple Conformational States. *Biochemistry* **51**, 8278–8292 (2012).
 62. Chan, P. H. *et al.* Investigating the Structural Dynamics of the alpha-1,4-Galactosyltransferase LgtC from *Neisseria meningitidis* by NMR Spectroscopy. *Biochemistry* (2012).
 63. Gómez, H., Polyak, I., Thiel, W., Lluch, J. M. & Masgrau, L. Retaining glycosyltransferase mechanism studied by QM/MM methods: Lipopolysaccharyl- α -1,4-galactosyltransferase C transfers α -galactose via an oxocarbenium ion-like transition state. *J. Am. Chem. Soc.* **134**, 4743–4752 (2012).
 64. Kajimoto, T. & Node, M. Synthesis of glycosyltransferase inhibitors. *Synthesis (Stuttg.)* **4**, 3179–3210 (2009).
 65. Takayama, S. *et al.* Selective Inhibition of -1, 4- and -1, 3-Galactosyltransferases : Donor Sugar-Nucleotide Based Approach. **7**, 401–409 (1999).
 66. Compain, P. & Martin, O. R. Carbohydrate mimetics-based glycosyltransferase inhibitors. *Bioorganic Med. Chem.* **9**, 3077–3092 (2001).
 67. Wang, R. *et al.* A search for pyrophosphate mimics for the development of substrates and inhibitors of glycosyltransferases. *Bioorg. Med. Chem.* **5**, 661–672 (1997).
 68. Wang, S., Cuesta-Seijo, J. a, Lafont, D., Palcic, M. M. & Vidal, S. Design of glycosyltransferase inhibitors: pyridine as a pyrophosphate surrogate. *Chemistry* **19**, 15346–57 (2013).
 69. Wang, S. *et al.* Design of Glycosyltransferase Inhibitors: Serine Analogues as Pyrophosphate Surrogates? *Chempluschem* **80**, 1525–1532 (2015).

70. Pesnot, T., Jørgensen, R., Palcic, M. M. & Wagner, G. K. Structural and mechanistic basis for a new mode of glycosyltransferase inhibition. *Nat. Chem. Biol.* **6**, 321–323 (2010).
71. Descroix, K. *et al.* Inhibition of galactosyltransferases by a novel class of donor analogues. *J. Med. Chem.* **55**, 2015–24 (2012).
72. Brown, J. R., Crawford, B. E. & Esko, J. D. *Glycan antagonists and inhibitors: A fount for drug discovery. Critical Reviews in Biochemistry and Molecular Biology* **42**, (2007).
73. Khan, S. H., Crawley, S. C., Kanie, O. & Hindsgaul, O. A trisaccharide acceptor analog for N-acetylglucosaminyltransferase V which binds to the enzyme but sterically precludes the transfer reaction. *J. Biol. Chem.* **268**, 2468–2473 (1993).
74. Tedaldi, L. & Wagner, G. K. Beyond substrate analogues: new inhibitor chemotypes for glycosyltransferases. *Medchemcomm* **5**, 1106 (2014).
75. Xu, Y. *et al.* Bioorganic & Medicinal Chemistry Covalent inhibitors of LgtC : A blueprint for the discovery of non-substrate-like inhibitors for bacterial glycosyltransferases. *Bioorg. Med. Chem.* **25**, 3182–3194 (2017).
76. Xu, Y. *et al.* Structure-activity relationships in a new class of non-substrate-like covalent inhibitors of the bacterial glycosyltransferase LgtC. *Bioorganic Med. Chem.* **26**, 2973–2983 (2018).
77. Strelow, J. M. A Perspective on the Kinetics of Covalent and Irreversible Inhibition. *J. Biomol. Screen.* **22**, 3–20 (2017).
78. Singh, J., Petter, R. C., Baillie, T. a & Whitty, A. The resurgence of covalent drugs. *Nat. Rev. Drug Discov.* **10**, 307–317 (2011).
79. Robertson, J. G. Current Topics Mechanistic Basis of Enzyme-Targeted Drugs. **44**, (2005).
80. Potashman, M. H. & Duggan, M. E. Covalent Modifiers : An Orthogonal Approach to Drug Design. *J. Med. Chem.* **52**, (2009).
81. London, N. *et al.* Covalent docking of large libraries for the discovery of chemical probes. *Nat. Chem. Biol.* **10**, 1066–72 (2014).
82. Ouyang, X. *et al.* CovalentDock: Automated covalent docking with parameterized covalent linkage energy estimation and molecular geometry constraints. *J. Comput. Chem.* **34**, 326–336 (2013).
83. Bauer, R. A. Covalent inhibitors in drug discovery: From accidental discoveries to avoided liabilities and designed therapies. *Drug Discov. Today* **20**, 1061–1073 (2015).
84. Ghosh, A. K., Samanta, I., Mondal, A. & Liu, W. R. Covalent Inhibition in Drug Discovery. *ChemMedChem* **14**, 889–906 (2019).
85. Lonsdale, R. & Ward, R. A. Structure-based design of targeted covalent inhibitors. *Chem. Soc. Rev.* **47**, 3816–3830 (2018).
86. Baillie, T. A. Targeted Covalent Inhibitors for Drug Design. *Angew. Chemie - Int. Ed.* **55**, 13408–13421 (2016).
87. Ray, S. & Murkin, A. S. New Electrophiles and Strategies for Mechanism-Based and Targeted Covalent Inhibitor Design. *Biochemistry* (2019).

88. Hagel, M. *et al.* Selective irreversible inhibition of a protease by targeting a noncatalytic cysteine. *Nat. Chem. Biol.* **7**, 22–24 (2011).
89. Hagel, M. *et al.* Selective irreversible inhibition of a protease by targeting a noncatalytic cysteine. *Nat. Chem. Biol.* **7**, 22–24 (2011).
90. Serafimova, I. M. *et al.* Reversible targeting of noncatalytic cysteines with chemically tuned electrophiles. *Nat. Chem. Biol.* **8**, 471–476 (2012).
91. Jöst, C., Nitsche, C., Scholz, T., Roux, L. & Klein, C. D. Promiscuity and selectivity in covalent enzyme inhibition: A systematic study of electrophilic fragments. *J. Med. Chem.* **57**, 7590–7599 (2014).
92. Jiang, J., Lazarus, M. B., Pasquina, L., Sliz, P. & Walker, S. A neutral diphosphate mimic crosslinks the active site of human O-GlcNAc transferase. *Nat. Chem. Biol.* **8**, 72–77 (2012).
93. Yang, P. & Liu, K. Activity-based protein profiling: Recent advances in probe development and applications. *ChemBioChem* **16**, 712–724 (2015).
94. Marko, F. & Matthew, B. Activity Based Probes as a tool for Functional Proteomic Analysis of Proteases. *Expert Rev Proteomics* **71**, 233–236 (2013).
95. Backus, K. M. *et al.* Proteome-wide covalent ligand discovery in native biological systems. *Nature* **534**, 570–4 (2016).
96. Sleno, L. & Emili, A. Proteomic methods for drug target discovery. *Curr. Opin. Chem. Biol.* **12**, 46–54 (2008).
97. Ziegler, S., Pries, V., Hedberg, C. & Waldmann, H. Target identification for small bioactive molecules: Finding the needle in the haystack. *Angew. Chemie - Int. Ed.* **52**, 2744–2792 (2013).
98. Ovaa, H. *et al.* Chemistry in living cells: Detection of active proteasomes by a two-step labeling strategy. *Angew. Chemie - Int. Ed.* **42**, 3626–3629 (2003).
99. Heal, W. P., Dang, T. H. T. & Tate, E. W. Activity-based probes: Discovering new biology and new drug targets. *Chem. Soc. Rev.* **40**, 246–257 (2011).
100. Aaron W. Puri, M. B. Using Small Molecules to Dissect Mechanisms of Microbial Pathogenesis. *ACS Chem. Biol.* **4**, 603–616 (2009).
101. Anthouard, R. & DiRita, V. J. Chemical biology applied to the study of bacterial pathogens. *Infect. Immun.* **83**, 456–469 (2015).
102. Flannery, A., Gerlach, J., Joshi, L. & Kilcoyne, M. Assessing Bacterial Interactions Using Carbohydrate-Based Microarrays. *Microarrays* **4**, 690–713 (2015).
103. Willems, L. I. *et al.* From covalent glycosidase inhibitors to activity-based glycosidase probes. *Chemistry* **20**, 10864–10872 (2014).
104. Albesa-Jové, D. & Guerin, M. E. The conformational plasticity of glycosyltransferases. *Curr. Opin. Struct. Biol.* **40**, 23–32 (2016).
105. Ouyang, X., Zhou, S., Ge, Z., Li, R. & Kwoh, C. K. CovalentDock Cloud: a web server for automated covalent docking. *Nucleic Acids Res.* **41**, 329–332 (2013).
106. Shan, Y., Oulaidi, F. & Lahmann, M. Lactosamine from lactulose via the Heyns rearrangement: A practical protocol. *Tetrahedron Lett.* **54**, 3960–3961 (2013).
107. Tedaldi, L., Evitt, A., Göös, N., Jiang, J. & Wagner, G. K. A practical

- glycosyltransferase assay for the identification of new inhibitor chemotypes. *Medchemcomm* **5**, 1193 (2014).
108. Ostash, B. & Walker, S. Moenomycin family antibiotics: Chemical synthesis, biosynthesis, and biological activity. *Nat. Prod. Rep.* **27**, 1594–1617 (2010).
 109. Derouaux, A., Sauvage, E. & Terrak, M. Peptidoglycan glycosyltransferase substrate mimics as templates for the design of new antibacterial drugs. *Front. Immunol.* **4**, 1–6 (2013).
 110. Zuegg, J. *et al.* Carbohydrate scaffolds as glycosyltransferase inhibitors with in vivo antibacterial activity. *Nat. Commun.* **6**, 1–11 (2015).
 111. Baizman, E. R. *et al.* Antibacterial activity of synthetic analogues based on the disaccharide structure of moenomycin, an inhibitor of bacterial transglycosylase. *Microbiology* **146**, 3129–3140 (2000).
 112. Cipolla, L., Gabrielli, L., Bini, D., Russo, L. & Shaikh, N. Kdo: A critical monosaccharide for bacteria viability. *Nat. Prod. Rep.* **27**, 1618–1629 (2010).
 113. Kondo, K. I., Doi, H., Adachi, H. & Nishimura, Y. Synergistic effect of CMP/KDO synthase inhibitors with antimicrobial agents on inhibition of production and release of vero toxin by enterohaemorrhagic *Escherichia coli* O157:H7. *Bioorganic Med. Chem. Lett.* **14**, 467–470 (2004).
 114. Riedl, B. & Schmid, W. A concise route to access C-glycosidic tetrazolyl analogues of Kdo as bioisosteres. *Carbohydr. Res.* **456**, 30–34 (2018).
 115. Titz, A. *Carbohydrates as Drugs. Topics in Medicinal Chemistry* (2014). doi:0
 116. Wagner, S. *et al.* Covalent Lectin Inhibition and Application in Bacterial Biofilm Imaging. *Angew. Chemie - Int. Ed.* 16559–16564 (2017).
 117. Sommer, R. *et al.* Glycomimetic, Orally Bioavailable LecB Inhibitors Block Biofilm Formation of *Pseudomonas aeruginosa*. *J. Am. Chem. Soc.* **140**, 2537–2545 (2018).
 118. Tielker, D. *et al.* *Pseudomonas aeruginosa* lectin LecB is located in the outer membrane and is involved in biofilm formation. *Microbiology* **151**, 1313–1323 (2005).
 119. De Cesco, S., Kurian, J., Dufresne, C., Mittermaier, A. K. & Moitessier, N. Covalent inhibitors design and discovery. *Eur. J. Med. Chem.* **138**, 96–114 (2017).
 120. Blake, L. & Soliman, M. E. S. Identification of irreversible protein splicing inhibitors as potential anti-TB drugs: Insight from hybrid non-covalent/covalent docking virtual screening and molecular dynamics simulations. *Med. Chem. Res.* **23**, 2312–2323 (2014).
 121. Amara, N. *et al.* Covalent Inhibition of Bacterial Quorum Sensing. *J. Am. Chem. Soc.* **131**, 10610–10619 (2009).
 122. Morkunas, B. *et al.* Inhibition of the production of the *Pseudomonas aeruginosa* virulence factor pyocyanin in wild-type cells by quorum sensing autoinducer-mimics. *Org. Biomol. Chem.* **10**, 8452–8464 (2012).
 123. Amara, N. *et al.* Fine-Tuning Covalent Inhibition of Bacterial Quorum Sensing. *ChemBioChem* **17**, 825–835 (2016).
 124. Carson, J. F. The reaction of fructose with isopropylamine and cyclohexylamine. *J. Am. Chem. Soc.* **77**, 1881–1884 (1955).

125. Heyns, K. & Meinecke, K. Über Bildung und Darstellung von d-Glucosamin aus Fructose und Ammoniak. *Chem Ber* **352**, 1453–1462 (1953).
126. Heyns, K. & Koch, W. Über die Bildung eines Aminozuckers aus d-Fructose und Ammoniak. *Z. Naturforsch.* **7b**, 486 (1952).
127. Wrodnigg, T.M.; Stütz, A. E. The Heyns rearrangement revisited: an exceptionally simple two-step chemical synthesis of D-lactosamine from lactulose. *Angew. Chem. Int. Ed* **38**, 827–828 (1999).
128. Sánchez-Viesca, F. & Gómez, R. On the Regiochemistry in the Heyns Rearrangement. *Am. J. Chem.* **5**, 86–89 (2015).
129. Ágoston, K., Dékány, G., Bajza, I. & Hederos, M. Synthesis of lactosamine from lactulose: scalable approach for the Heyns rearrangement. *Tetrahedron Lett.* **57**, 2595–2597 (2016).
130. Wang, Z. Pearlman's Catalyst. *Compr. Org. Name React. Reagents* 2143–2146 (2010).
131. Ji, H., Jing, Q., Huang, J. & Silverman, R. B. Acid-facilitated debenzoylation of N-Boc, N-benzyl double protected 2-aminopyridinomethyl pyrrolidine derivatives. *Tetrahedron* **68**, 1359–1366 (2012).
132. Chandra, T. & Zebrowski, J. P. Hazards associated with laboratory scale hydrogenations. *J. Chem. Heal. Saf.* **23**, 16–25 (2016).
133. Wu, Z. L., Ethen, C. M., Prather, B., MacHacek, M. & Jiang, W. Universal phosphatase-coupled glycosyltransferase assay. *Glycobiology* **21**, 727–733 (2011).
134. Itaya, K. & Ui, M. A new micromethod for the colorimetric determination of inorganic phosphate. *Clin. Chim. Acta.* **14**, 361–366 (1966).
135. Lairson, L. L., Watts, A. G., Wakarchuk, W. W. & Withers, S. G. Using substrate engineering to harness enzymatic promiscuity and expand biological catalysis. *Nat. Chem. Biol.* **2**, 724–728 (2006).
136. Lairson, L. L., Wakarchuk, W. W. & Withers, S. G. Alternative donor substrates for inverting and retaining glycosyltransferases. *Chem. Commun.* 365–367 (2007).
137. Morgen, M. *et al.* Spiroepoxytriazoles Are Fumagillin-like Irreversible Inhibitors of MetAP2 with Potent Cellular Activity. *ACS Chem. Biol.* **11**, 1001–1011 (2016).
138. Kim, K. B. & Crews, C. M. From epoxomicin to carfilzomib: Chemistry, biology, and medical outcomes. *Nat. Prod. Rep.* **30**, 600–604 (2013).
139. Cohen, M. S., Zhang, C., Shokat, K. M. & Taunton, J. Biochemistry: Structural bioinformatics-based design of selective, irreversible kinase inhibitors. *Science (80-)*. **308**, 1318–1321 (2005).
140. Leproult, E., Barluenga, S., Moras, D., Wurtz, J. M. & Winssinger, N. Cysteine mapping in conformationally distinct kinase nucleotide binding sites: Application to the design of selective covalent inhibitors. *J. Med. Chem.* **54**, 1347–1355 (2011).
141. Lim, S. M. *et al.* Therapeutic targeting of oncogenic K-ras by a covalent catalytic site inhibitor. *Angew. Chemie - Int. Ed.* **53**, 199–204 (2014).
142. Cabrales, P. RRx-001 Acts as a Dual Small Molecule Checkpoint Inhibitor by Downregulating CD47 on Cancer Cells and SIRP- α on Monocytes/Macrophages.

- Transl. Oncol.* **12**, 626–632 (2019).
143. Lagoutte, R., Patouret, R. & Winssinger, N. Covalent inhibitors: an opportunity for rational target selectivity. *Curr. Opin. Chem. Biol.* **39**, 54–63 (2017).
 144. Pettinger, J. *et al.* An Irreversible Inhibitor of HSP72 that Unexpectedly Targets Lysine-56. *Angew. Chemie - Int. Ed.* **56**, 3536–3540 (2017).
 145. Mukherjee, H. & Grimster, N. P. Beyond cysteine: recent developments in the area of targeted covalent inhibition. *Curr. Opin. Chem. Biol.* **44**, 30–38 (2018).
 146. Narayanan, A. & Jones, L. H. Sulfonyl fluorides as privileged warheads in chemical biology. *Chem. Sci.* **6**, 2650–2659 (2015).
 147. Kamps, M. P., Taylor, S. S. & Sefton, B. M. Direct evidence that oncogenic tyrosine kinases and cyclic AMP-dependent protein kinase have homologous ATP-binding sites. *Nature* **310**, 589–592 (1984).
 148. Manvar, D., Singh, K. & Pandey, V. N. Affinity labeling of hepatitis C virus replicase with a nucleotide analogue: Identification of binding site. *Biochemistry* **52**, 432–444 (2013).
 149. Switzerg, R. L. Chemical Modification of *Salmonella typhimurium*. **265**, 5487–5493 (1990).
 150. Lee, C. U. & Grossmann, T. N. Reversible covalent inhibition of a protein target. *Angew. Chemie - Int. Ed.* **51**, 8699–8700 (2012).
 151. Accardi, F. *et al.* Mechanism of action of bortezomib and the new proteasome inhibitors on myeloma cells and the bone microenvironment: Impact on myeloma-induced alterations of bone remodeling. *Biomed Res. Int.* **2015**, (2015).
 152. Akçay, G. *et al.* Inhibition of Mcl-1 through covalent modification of a noncatalytic lysine side chain. *Nat. Chem. Biol.* **12**, 931–936 (2016).
 153. Dhillon, S. & Weber, J. Saxagliptin. *Drugs* **69**, 2103–2114 (2009).
 154. Bradshaw¹, J. M. *et al.* Prolonged and tunable residence time using reversible covalent kinase inhibitors. *Nat. Chem. Biol.* **11**, 525–531 (2015).
 155. Shah, S. U. A., Jawed, H., Awan, S. I., Anjum, S. & Simjee, S. U. The anti-arthritic and immune-modulatory effects of NHAG: A novel glucosamine analogue in adjuvant-induced arthritis. *Biomed Res. Int.* **2013**, (2013).
 156. Kim, H. K. & Park, T. G. Synthesis and characterization of thermally reversible bioconjugates composed of α -chymotrypsin and poly(N-isopropylacrylamide-co-acrylamido-2-deoxy-D-glucose). *Enzyme Microb. Technol.* **25**, 31–37 (1999).
 157. Zhang, K. *et al.* 'Bitter-Sweet' Polymeric Micelles Formed by Block Copolymers from Glucosamine and Cholic Acid. *Biomacromolecules* **18**, 778–786 (2017).
 158. Masuko, S. *et al.* Chemoenzymatic synthesis of uridine diphosphate-GlcNAc and uridine diphosphate-GalNAc analogs for the preparation of unnatural glycosaminoglycans. *J. Org. Chem.* **77**, 1449–1456 (2012).
 159. Lonsdale, R. *et al.* Expanding the Armory: Predicting and Tuning Covalent Warhead Reactivity. *J. Chem. Inf. Model.* **57**, 3124–3137 (2017).
 160. Shannon, D. A. & Weerapana, E. Covalent protein modification: The current landscape of residue-specific electrophiles. *Curr. Opin. Chem. Biol.* **24**, 18–26

- (2015).
161. Martin, J. S., MacKenzie, C. J., Fletcher, D. & Gilbert, I. H. Characterising covalent warhead reactivity. *Bioorganic Med. Chem.* **27**, 2066–2074 (2019).
 162. Punthasee, P. *et al.* Covalent Allosteric Inactivation of Protein Tyrosine Phosphatase 1B (PTP1B) by an Inhibitor-Electrophile Conjugate. *Biochemistry* **56**, 2051–2060 (2017).
 163. Hayden Peacock, Radhika Bachu, and P. A. B. Covalent Stabilization of a Small Molecule-RNA Complex. *Bioorg Med Chem Lett.* **21**, 5002–5005 (2011).
 164. Lee, H. H. *et al.* Expression, purification and crystallization of CTB-MPR, a candidate mucosal vaccine component against HIV-1. *IUCrj* **1**, 305–317 (2014).
 165. Xu, Y., Uddin, N. & Wagner, G. K. *Covalent Probes for Carbohydrate-Active Enzymes: From Glycosidases to Glycosyltransferases. Methods in Enzymology* **598**, (Elsevier Inc., 2018).
 166. Tsai, C. S., Li, Y. K. & Lo, L. C. Design and synthesis of activity probes for glycosidases. *Org. Lett.* **4**, 3607–3610 (2002).
 167. Rempel, B. P. & Withers, S. G. Covalent inhibitors of glycosidases and their applications in biochemistry and biology. *Glycobiology* **18**, 570–586 (2008).
 168. Wu, L. *et al.* An overview of activity-based probes for glycosidases. *Curr. Opin. Chem. Biol.* **53**, 25–36 (2019).
 169. Sakurai, K., Yamaguchi, T. & Mizuno, S. Design and synthesis of fluorescent glycolipid photoaffinity probes and their photoreactivity. *Bioorganic Med. Chem. Lett.* **26**, 5110–5115 (2016).
 170. Hsu, Y. L. *et al.* Development of Activity-Based Probes for Imaging Human α -L - Fucosidases in Cells. *J. Org. Chem.* **80**, 8458–8463 (2015).
 171. Phenix, C. P. *et al.* Imaging of enzyme replacement therapy using PET. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 10842–10847 (2010).
 172. Witte, M. D. *et al.* Ultrasensitive in situ visualization of active glucocerebrosidase molecules. *Nat. Chem. Biol.* **6**, 907–913 (2010).
 173. Li, K. Y. *et al.* Synthesis of cyclophellitol, cyclophellitol aziridine, and their tagged derivatives. *European J. Org. Chem.* **2014**, 6030–6043 (2014).
 174. Anderson, L. N. *et al.* Activity-based protein profiling of secreted cellulolytic enzyme activity dynamics in *Trichoderma reesei* QM6a, NG14, and RUT-C30. *Mol. Biosyst.* **9**, 2992–3000 (2013).
 175. Vocadlo, D. J. & Bertozzi, C. R. A strategy for functional proteomic analysis of glycosidase activity from cell lysates. *Angew. Chemie - Int. Ed.* **43**, 5338–5342 (2004).
 176. Drake, R. R. & Elbein, A. D. Photoaffinity labelling of glycosyltransferases. *Glycobiology* **2**, 279–284 (1992).
 177. Drake, R. R., Evans, R. K., Wolf, M. J. & Haleys, B. E. Development of Photoaffinity Probes. **264**, 11928–11933 (1989).
 178. Zeng, Y. *et al.* Synthesis of aryl azide derivatives of UDP-GlcNAc and UDP-GalNAc and their use for the affinity labeling of glycosyltransferases and the UDP-

- HexNAc pyrophosphorylase. *Anal. Biochem.* **239**, 99–106 (1996).
179. Takaya, K. *et al.* Rational design, synthesis, and characterization of novel inhibitors for human β 1,4-galactosyltransferase. *J. Med. Chem.* **48**, 6054–6065 (2005).
 180. Rostovtsev, V. V., Green, L. G., Fokin, V. V. & Sharpless, K. B. A stepwise Huisgen cycloaddition process: Copper(I)-catalyzed regioselective 'ligation' of azides and terminal alkynes. *Angew. Chemie - Int. Ed.* **41**, 2596–2599 (2002).
 181. Nwe, K. & Brechbiel, M. W. Growing Applications of "Click Chemistry" for Bioconjugation in Contemporary Biomedical Research. *Cancer Biother. Radiopharm.* **24**, 289–302 (2009).
 182. Liang, L. & Astruc, D. The copper(I)-catalyzed alkyne-azide cycloaddition (CuAAC) 'click' reaction and its applications. An overview. *Coord. Chem. Rev.* **255**, 2933–2945 (2011).
 183. Kolb, H. C., Finn, M. G. & Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chemie - Int. Ed.* **40**, 2004–2021 (2001).
 184. B. T. Worrell, J. A. Malik, V. V. F. Direct Evidence of a Dinuclear Copper Intermediate in Cu(I)-Catalyzed Azide-Alkyne Cycloadditions. **340**, 457–461 (2013).
 185. He, X. P. *et al.* Carbohydrate CuAAC click chemistry for therapy and diagnosis. *Carbohydr. Res.* **429**, 1–22 (2016).
 186. Dedola, S., Nepogodiev, S. A. & Field, R. A. Recent applications of the Cu^I-catalysed Huisgen azide-alkyne 1,3-dipolar cycloaddition reaction in carbohydrate chemistry. *Org. Biomol. Chem.* **5**, 1006–1017 (2007).
 187. Witczak, Z. J. & Bielski, R. *Click Chemistry in Glycoscience: New Developments and Strategies.* (Wiley Online Books, 2013).
 188. Mahon, B. P. *et al.* Saccharin: A lead compound for structure-based drug design of carbonic anhydrase IX inhibitors. *Bioorganic Med. Chem.* **23**, 849–854 (2015).
 189. Fraser, B. H. *et al.* Synthesis of 1,4-triazole linked zanamivir dimers as highly potent inhibitors of influenza A and B. *Medchemcomm* **4**, 383–386 (2013).
 190. Lo Conte, M. *et al.* Modular approach to triazole-linked 1,6- α -D-oligomannosides to the discovery of inhibitors of mycobacterium tuberculosis cell wall synthetase. *J. Org. Chem.* **75**, 6326–6336 (2010).
 191. Herzog, I. M., Feldman, M., Eldar-Boock, A., Satchi-Fainaro, R. & Fridman, M. Design of membrane targeting tobramycin-based cationic amphiphiles with reduced hemolytic activity. *Medchemcomm* **4**, 120–124 (2013).
 192. Hu, X. Le *et al.* Triazole-Linked Glycolipids Enhance the Susceptibility of MRSA to β -Lactam Antibiotics. *ACS Med. Chem. Lett.* **6**, 793–797 (2015).
 193. Peng, H. *et al.* A fluorescent probe for fast and quantitative detection of hydrogen sulfide in blood. *Angew. Chemie - Int. Ed.* **50**, 9672–9675 (2011).
 194. Yamada, T., Park, K., Monguchi, Y., Sawama, Y. & Sajiki, H. Mild deuteration method of terminal alkynes in heavy water using reusable basic resin. *RSC Adv.* **5**, 92954–92957 (2015).

195. Dobretsov, G. E., Syrejschikova, T. I. & Smolina, N. V. On mechanisms of fluorescence quenching by water. *Biophysics (Oxf)*. **59**, 183–188 (2014).
196. Hoch, D. G., Abegg, D. & Adibekian, A. Cysteine-reactive probes and their use in chemical proteomics. *Chem. Commun.* **54**, 4501–4512 (2018).
197. Donkor, E. S. Sequencing of bacterial genomes: Principles and insights into pathogenesis and development of antibiotics. *Genes (Basel)*. **4**, 556–572 (2013).
198. Brötz-Oesterhelt, H., Bandow, J. E. & Labischinski, H. Bacterial proteomics and its role in antibacterial drug discovery. *Mass Spectrom. Rev.* **24**, 549–565 (2005).
199. Saleh, S., Staes, A., Deborggraeve, S. & Gevaert, K. Targeted Proteomics for Studying Pathogenic Bacteria. *Proteomics* **1800435**, 1800435 (2019).
200. Pérez-Llarena, F. J. & Bou, G. Proteomics as a tool for studying bacterial virulence and antimicrobial resistance. *Front. Microbiol.* **7**, 1–21 (2016).
201. Vranakis, I. *et al.* Proteome studies of bacterial antibiotic resistance mechanisms. *J. Proteomics* **97**, 88–99 (2014).
202. Rix, U. & Superti-Furga, G. Target profiling of small molecules by chemical proteomics. *Nat. Chem. Biol.* **5**, 616–624 (2009).
203. Jeffery, D. A. & Bogyo, M. Chemical proteomics and its application to drug discovery. *Curr. Opin. Biotechnol.* **14**, 87–95 (2003).
204. Wright, M. H. & Sieber, S. A. Chemical proteomics approaches for identifying the cellular targets of natural products. *Nat. Prod. Rep.* **33**, 681–708 (2016).
205. Wright, M. H. Chemical Proteomics of Host–Microbe Interactions. *Proteomics* **18**, 1–10 (2018).
206. Zhu, H., Tamura, T. & Hamachi, I. Chemical proteomics for subcellular proteome analysis. *Curr. Opin. Chem. Biol.* **48**, 1–7 (2019).
207. Staub, I. & Sieber, S. A. B-Lactams As Selective Chemical Probes for the in Vivo Labeling of Bacterial Enzymes Involved in Cell Wall Biosynthesis, Antibiotic Resistance, and Virulence. *J. Am. Chem. Soc.* **130**, 13400–13409 (2008).
208. Staub, I. & Sieber, S. A. β -Lactam probes as selective chemical-proteomic tools for the identification and functional characterization of resistance associated enzymes in MRSA. *J. Am. Chem. Soc.* **131**, 6271–6276 (2009).
209. Wright, M. H., Fetzer, C. & Sieber, S. A. Chemical Probes Unravel an Antimicrobial Defense Response Triggered by Binding of the Human Opioid Dynorphin to a Bacterial Sensor Kinase. *J. Am. Chem. Soc.* **139**, 6152–6159 (2017).
210. Garron, M. L. & Cygler, M. Structural and mechanistic classification of uronic acid-containing polysaccharide lyases. *Glycobiology* **20**, 1547–1573 (2010).
211. Esposito, D. *et al.* Structural basis for the glycosyltransferase activity of the Salmonella effector SseK3. *J. Biol. Chem.* **293**, 5064–5078 (2018).
212. Sjögren, J. & Collin, M. Bacterial glycosidases in pathogenesis and glycoengineering. *Future Microbiol.* **9**, 1039–1051 (2014).
213. Blair, D. E., Schüttelkopf, A. W., MacRae, J. I. & Van Aalten, D. M. F. Structure and metal-dependent mechanism of peptidoglycan deacetylase, a streptococcal virulence factor. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15429–15434 (2005).

214. Watanabe, S., Shimomura, Y., Ubukata, K., Kirikae, T. & Miyoshi-Akiyama, T. Concomitant regulation of host tissue-destroying virulence factors and carbohydrate metabolism during invasive diseases induced by group G streptococci. *J. Infect. Dis.* **208**, 1482–1493 (2013).
215. Katherine Rainey, Suzanne M. Michalek, Zezhang T. Wen, H. W. Glycosyltransferase-Mediated Biofilm Matrix Dynamics and Virulence of *Streptococcus mutans*. *Am. Soc. Microbiol.* **85**, 1–15 (2019).
216. Lu, Q., Li, S. & Shao, F. Sweet Talk: Protein Glycosylation in Bacterial Interaction With the Host. *Trends Microbiol.* **23**, 630–641 (2015).
217. Valguarnera, E., Kinsella, R. L. & Feldman, M. F. Sugar and Spice Make Bacteria Not Nice: Protein Glycosylation and Its Influence in Pathogenesis. *J. Mol. Biol.* **428**, 3206–3220 (2016).
218. Stimson, E. *et al.* Meningococcal pilin: a glycoprotein substituted with digalactosyl 2,4-diacetamido-2,4,6-trideoxyhexose. *Mol. Microbiol.* **17**, 1201–1214 (1995).
219. Grass, S. *et al.* The *Haemophilus influenzae* HMW1 adhesin is glycosylated in a process that requires HMW1C and phosphoglucomutase, an enzyme involved in lipooligosaccharide biosynthesis. *Mol. Microbiol.* **48**, 737–751 (2003).
220. Jank, T., Belyi, Y. & Aktories, K. Bacterial glycosyltransferase toxins. *Cell. Microbiol.* **17**, 1752–1765 (2015).
221. Voth, D. E. & Ballard, J. D. *Clostridium difficile* Toxins: Mechanism of Action and Role in Disease. *Clin. Microbiol. Rev.* **18**, 247–263 (2005).
222. Allhorn, M., Olin, A. I., Nimmerjahn, F. & Colin, M. Human IgC/Fc γ R interactions are modulated by streptococcal IgG glycan hydrolysis. *PLoS One* **3**, (2008).
223. Pastoriza Gallego, M. & Hulen, C. Influence of sialic acid and bacterial sialidase on differential adhesion of *Pseudomonas aeruginosa* to epithelial cells. *Colloids Surfaces B Biointerfaces* **52**, 154–156 (2006).
224. Lis, H. & Sharon, N. Lectins: Carbohydrate-specific proteins that mediate cellular recognition. *Chem. Rev.* **98**, 637–674 (1998).
225. Sharon, N. Bacterial lectins, cell-cell recognition and infectious disease. *FEBS Lett.* **217**, 145–157 (1987).
226. Simon, P. M., Goode, P. L., Mobasser, A. & Zopf, D. Inhibition of *Helicobacter pylori* binding to gastrointestinal epithelial cells by sialic acid-containing oligosaccharides. *Infect. Immun.* **65**, 750–757 (1997).
227. Stephen P. Diggle *et al.* The galactophilic lectin, LecA, contributes to biofilm development in *Pseudomonas aeruginosa*. *Environ. Microbiol.* **8**, 1095–1104 (2006).
228. Stubbs, K. A. *et al.* Synthesis and use of mechanism-based protein-profiling probes for retaining β -D-glucosaminidases facilitate identification of *Pseudomonas aeruginosa* NagZ. *J. Am. Chem. Soc.* **130**, 327–335 (2008).
229. Stubbs, K. A. Activity-based proteomics probes for carbohydrate-processing enzymes: Current trends and future outlook. *Carbohydr. Res.* **390**, 9–19 (2014).
230. Palaniappan, K. K. & Bertozzi, C. R. Chemical Glycoproteomics. *Chem. Rev.* **116**, 14277–14306 (2016).

231. Nizet, V., Colina, K. F., Almquist, J. R., Rubens, C. E. & Smith, A. L. A virulent nonencapsulated *Haemophilus influenzae*. *J. Infect. Dis.* **173**, 180–186 (1996).
232. Wulffen, B., Rudolf, C., Pofahl, M. & Heckel, A. Caged glucosamine-6-phosphate for the light-control of riboswitch activity. *Photochem. Photobiol. Sci.* **11**, 489–492 (2012).
233. Shu Quan , Annie Hiniker , Jean-François Collet, and J. C. A. B. Chapter 22 Isolation of Bacteria Envelope Proteins. in *Methods in Molecular Biology* **966**, 359–366 (2013).

Appendices

Appendix 1 (Figures A1.1 – A1.61)

¹H NMR & ¹³C NMR spectra of all synthesised cmpds (Chapters 2-5)

Appendix 2 (Figures A2.1 – A2.2)

¹H NMR of cmpds whose synthesis was attempted

Appendix 3 (Figures A3.1 – A3.4)

Additional information relevant to Chapter 3

- MS analysis for detection of the **11** – LgtC covalent adduct
- Monitoring of the reaction of keto-based WH fragments with cysteine by ¹H NMR

Appendix 4 (Figures A4.1 – A4.2)

Additional information relevant to Chapter 4

- ¹H NMR spectra of collected fractions after distillation of Propynol ethoxylate
- Additional analysis on monosaccharide **11** (see 4.9)

Appendix 5 (Figure A5.1 – A5.4)

Additional information relevant to Chapter 5

- Viability test
- Labelling in cell lysates with deAc probes
- 2-step labelling

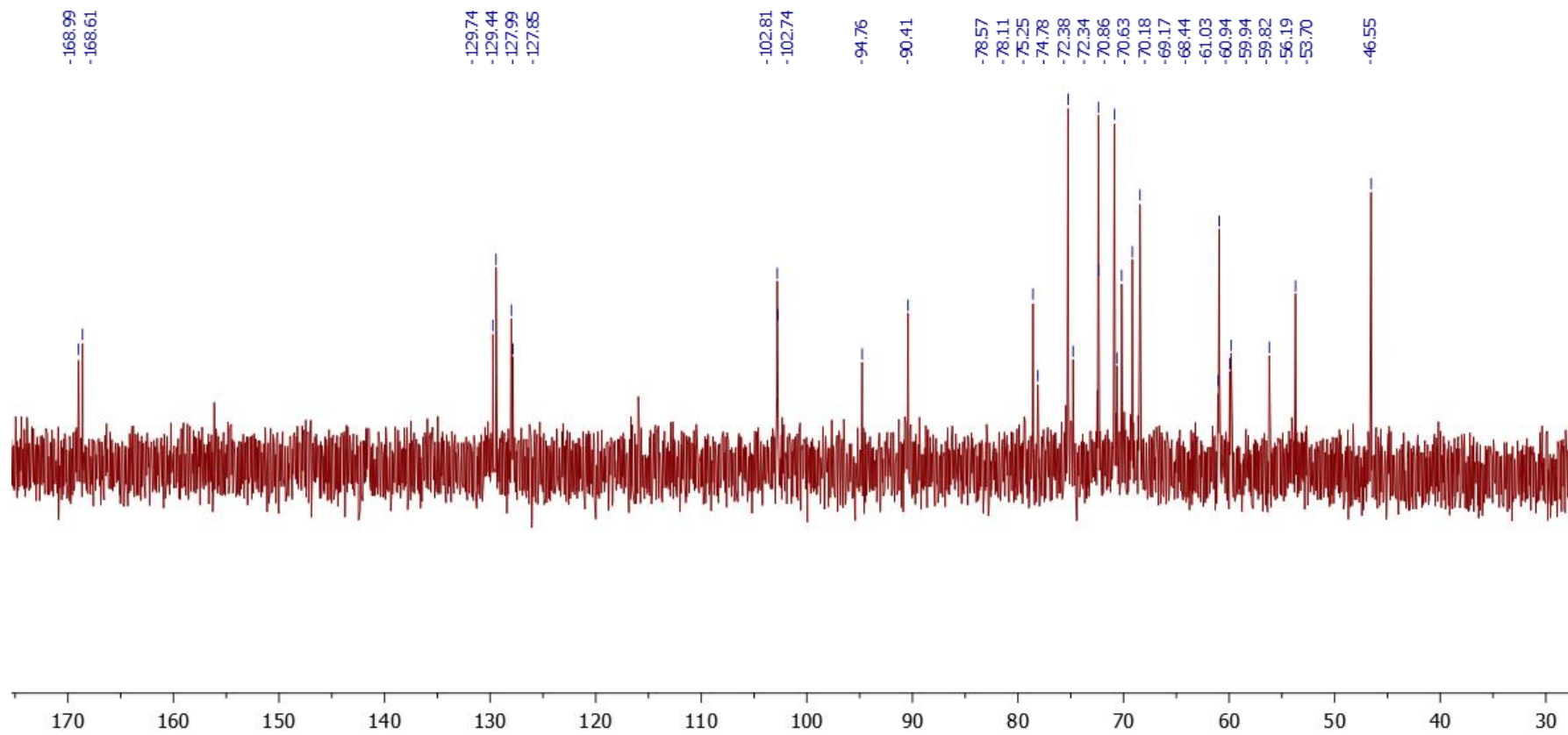


Figure A1.1. ^{13}C NMR of compd **1** in D_2O

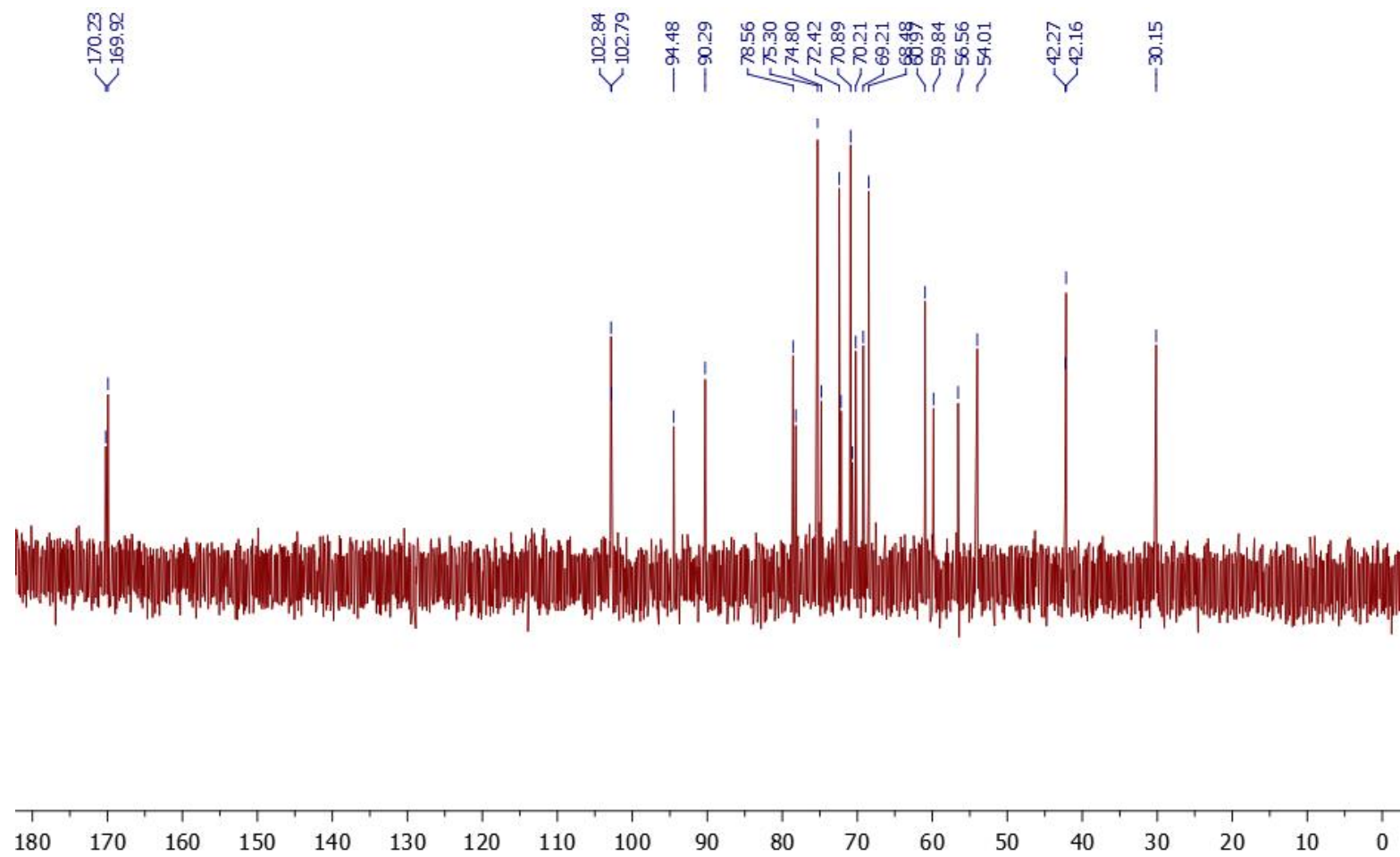


Figure A1.2. ^{13}C NMR of cmpd **2** in D_2O

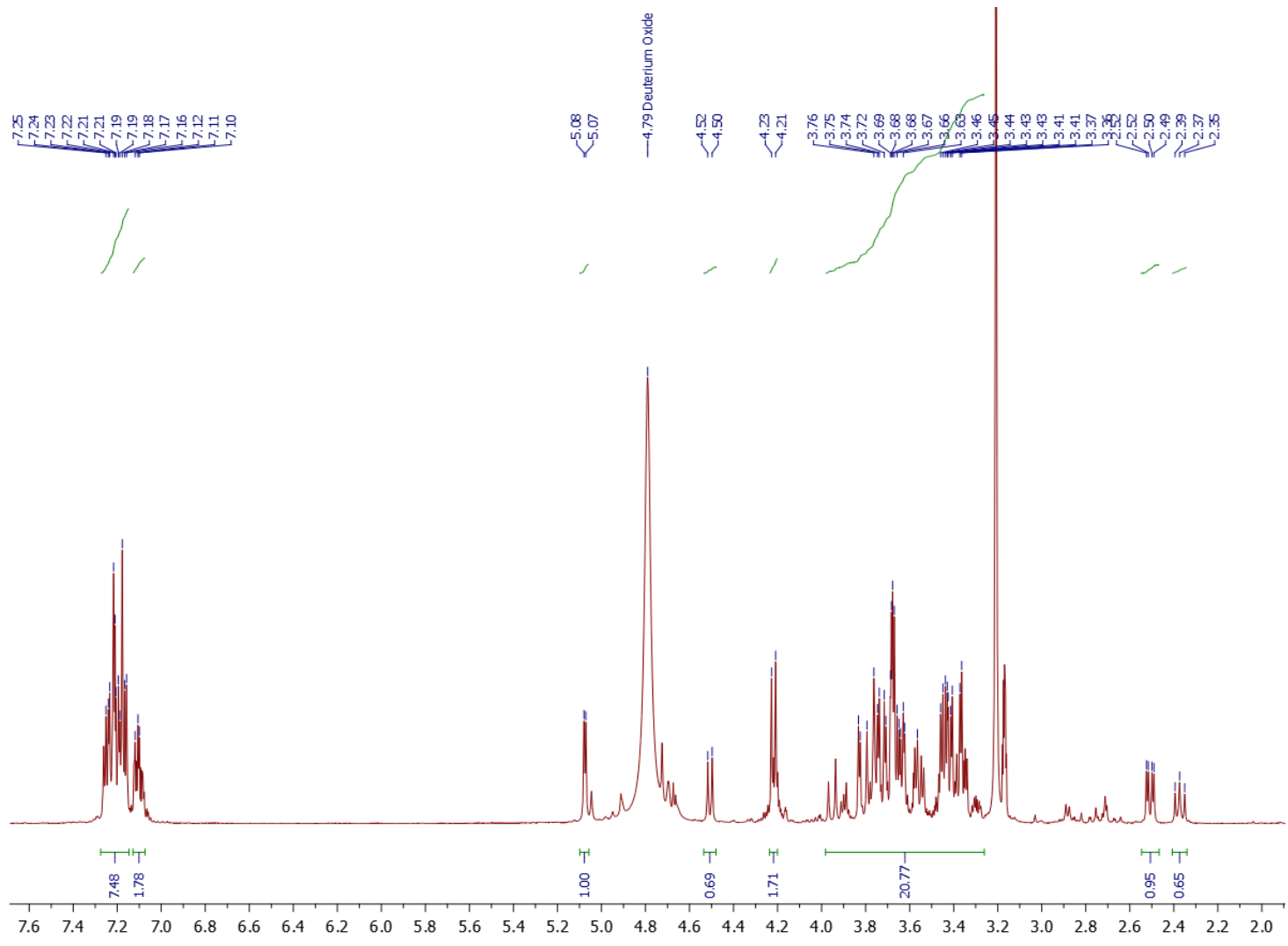


Figure A1.3. ^1H NMR of cmpd 5 in D_2O

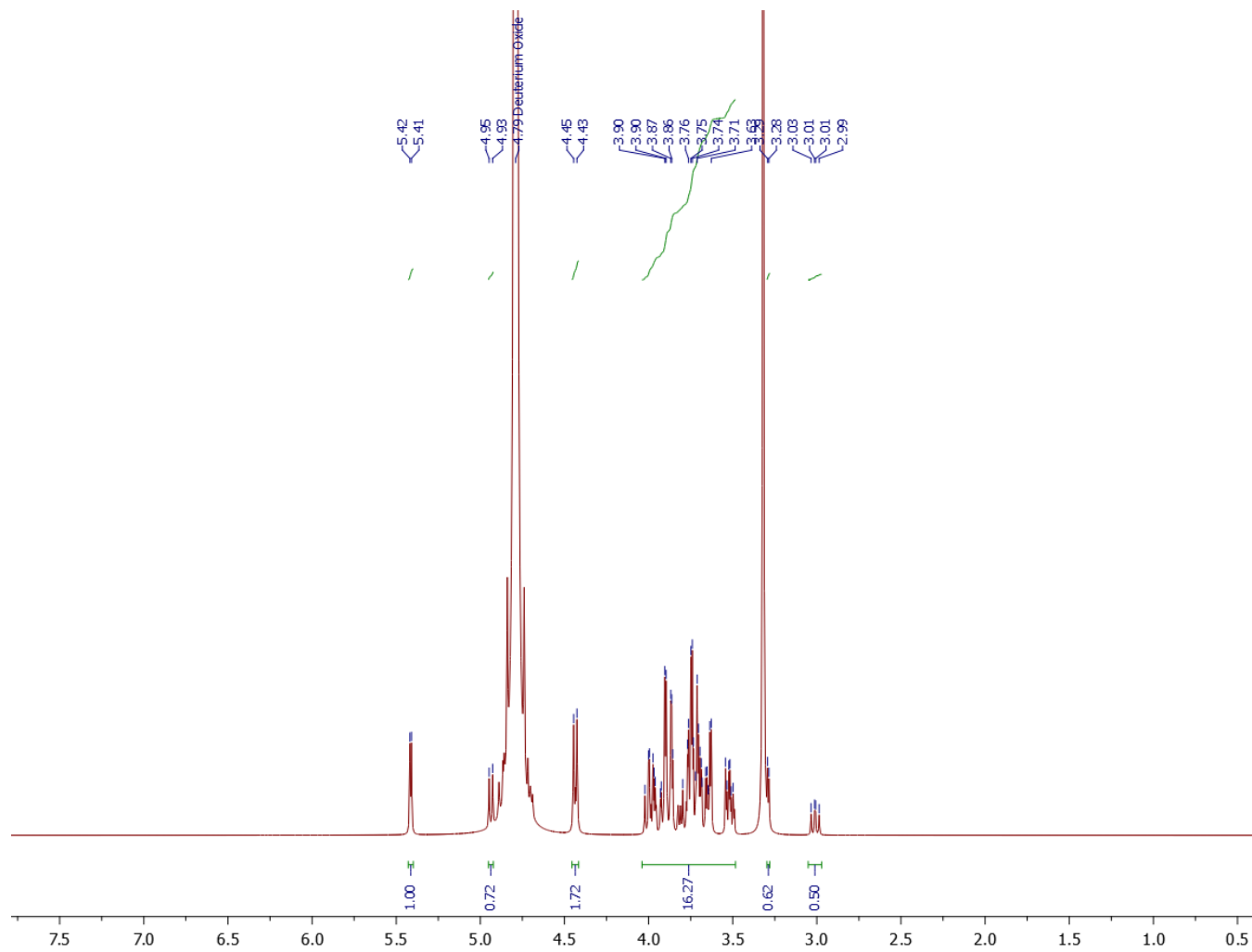


Figure A1.4. ¹H NMR of compd 6 in D₂O

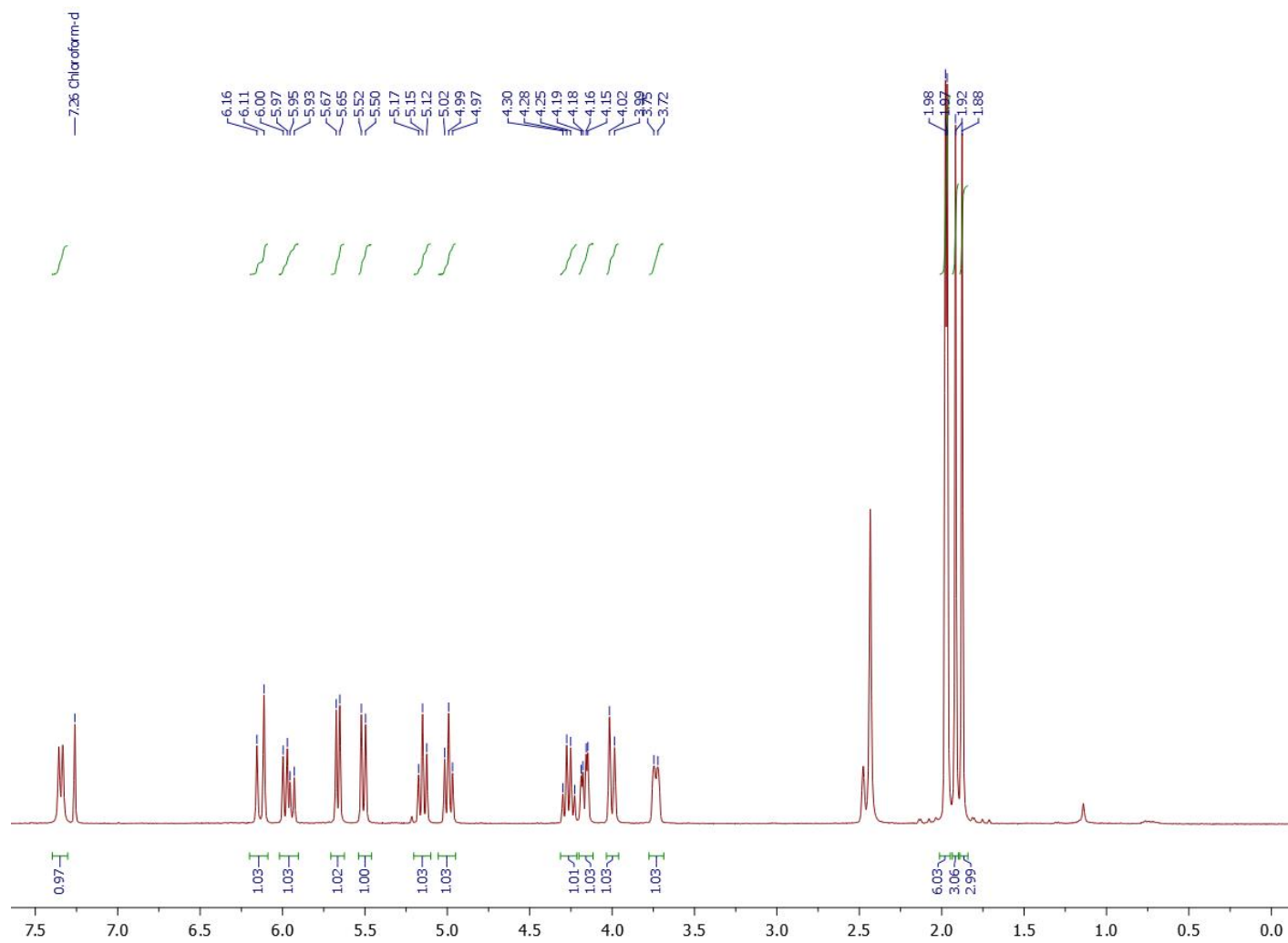


Figure A1.5. ^1H NMR of compd **8** in CDCl_3

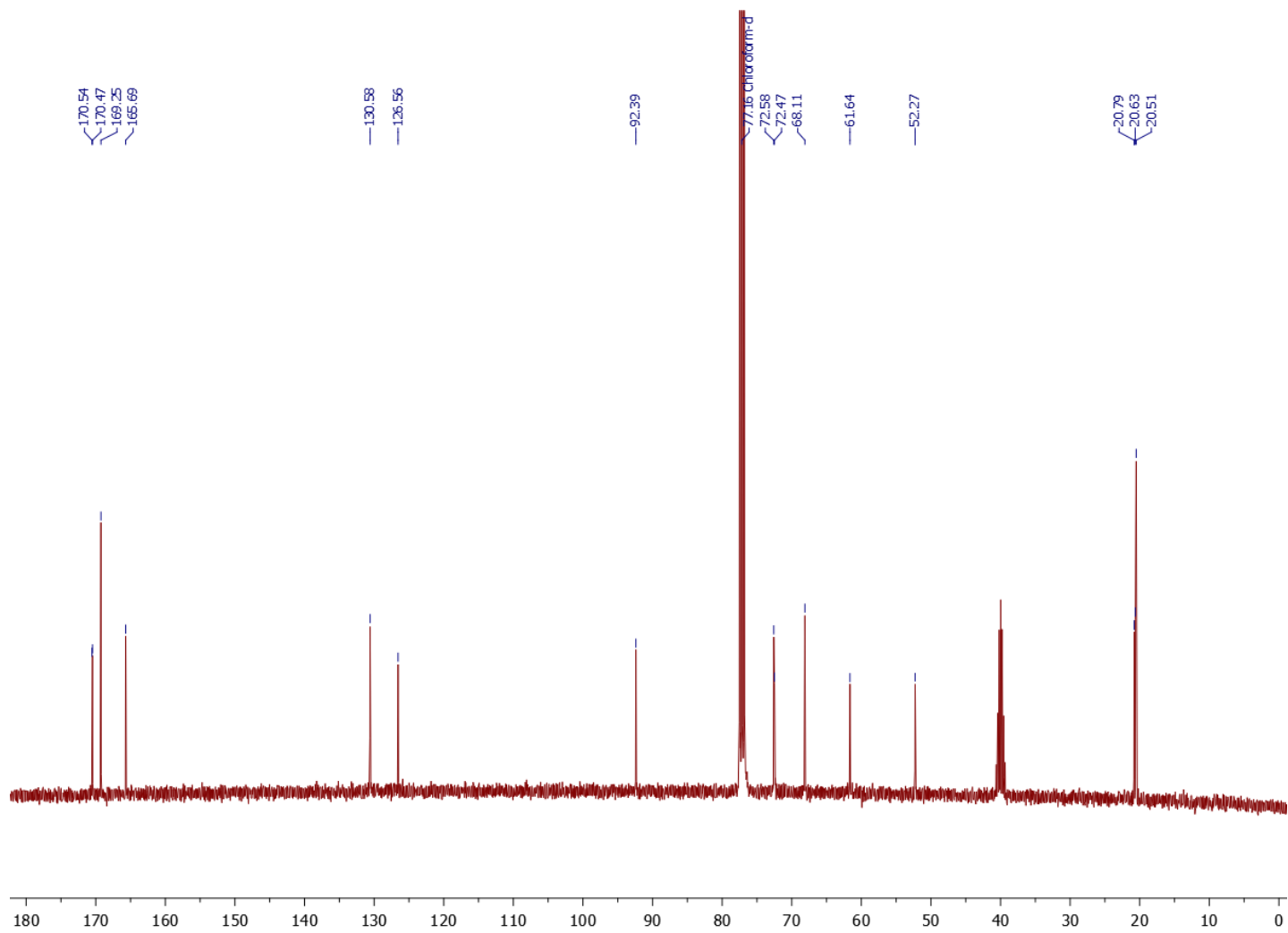


Figure A1.6. ^{13}C NMR of compd **8** in CDCl_3

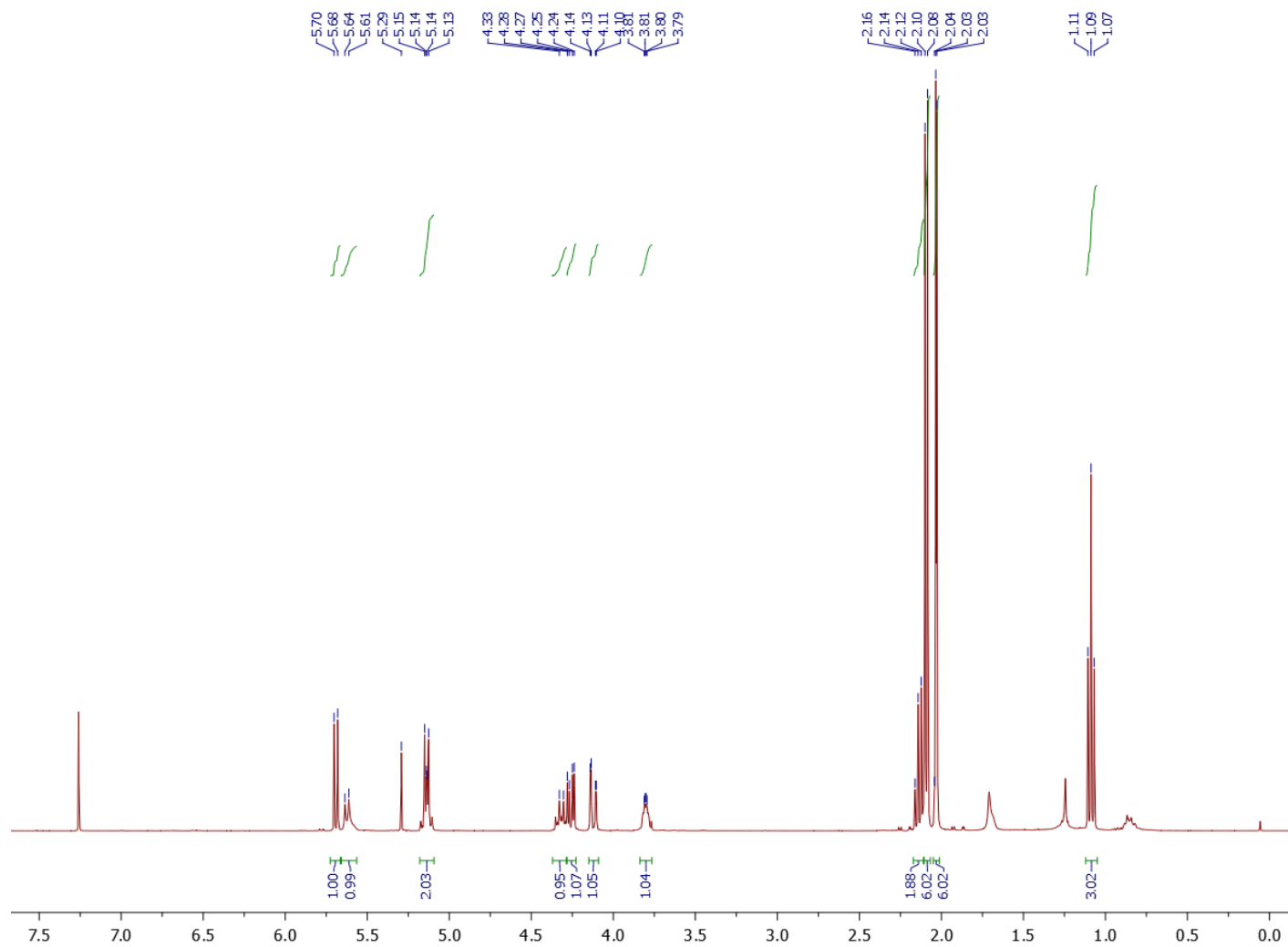


Figure A1.7. ^1H NMR of **cmpd 8s** in CDCl_3

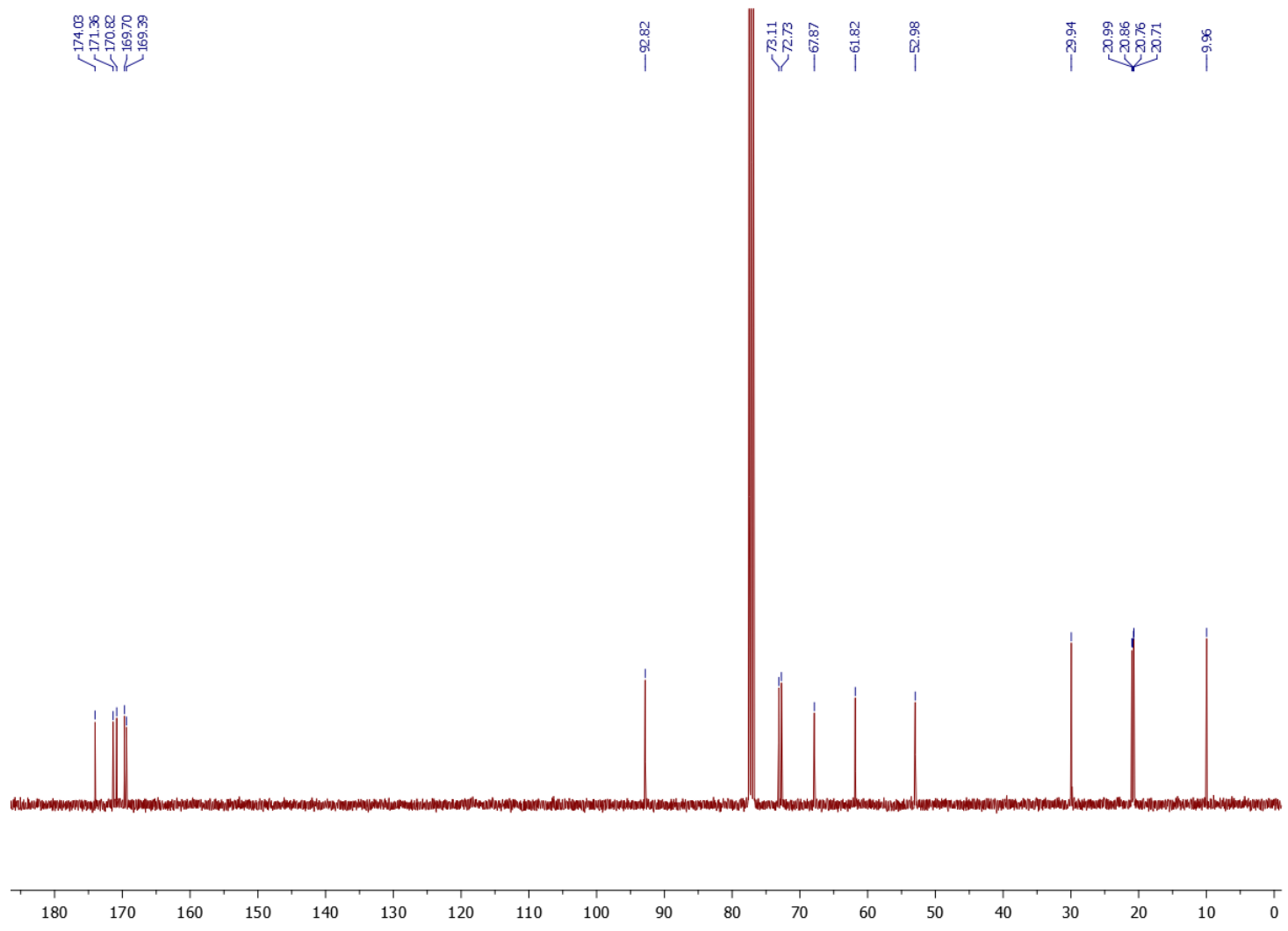


Figure A1.8. ^{13}C NMR of cmpd **8s** in CDCl_3

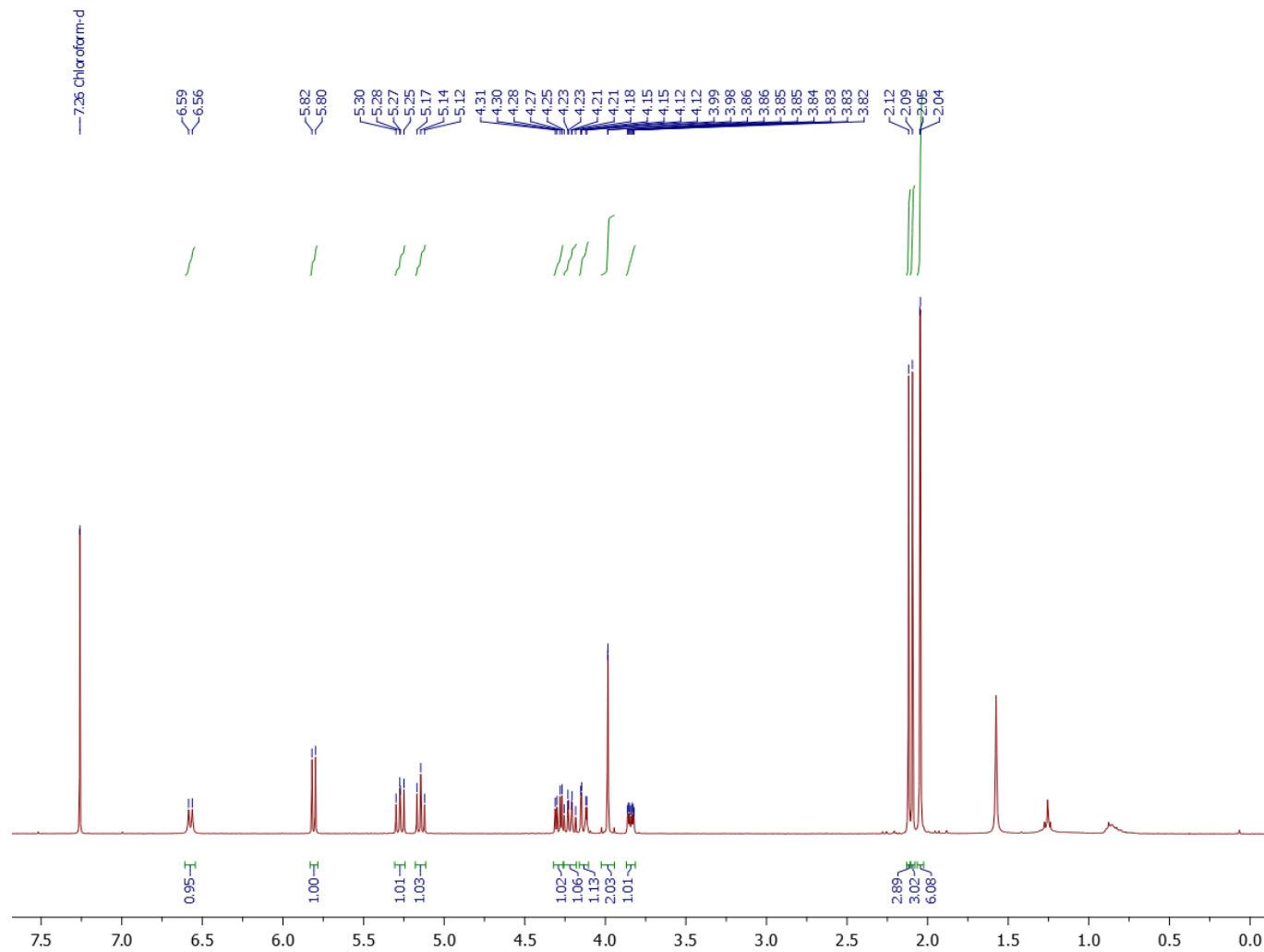


Figure A1.9. ^1H NMR of compd **9** in CDCl_3

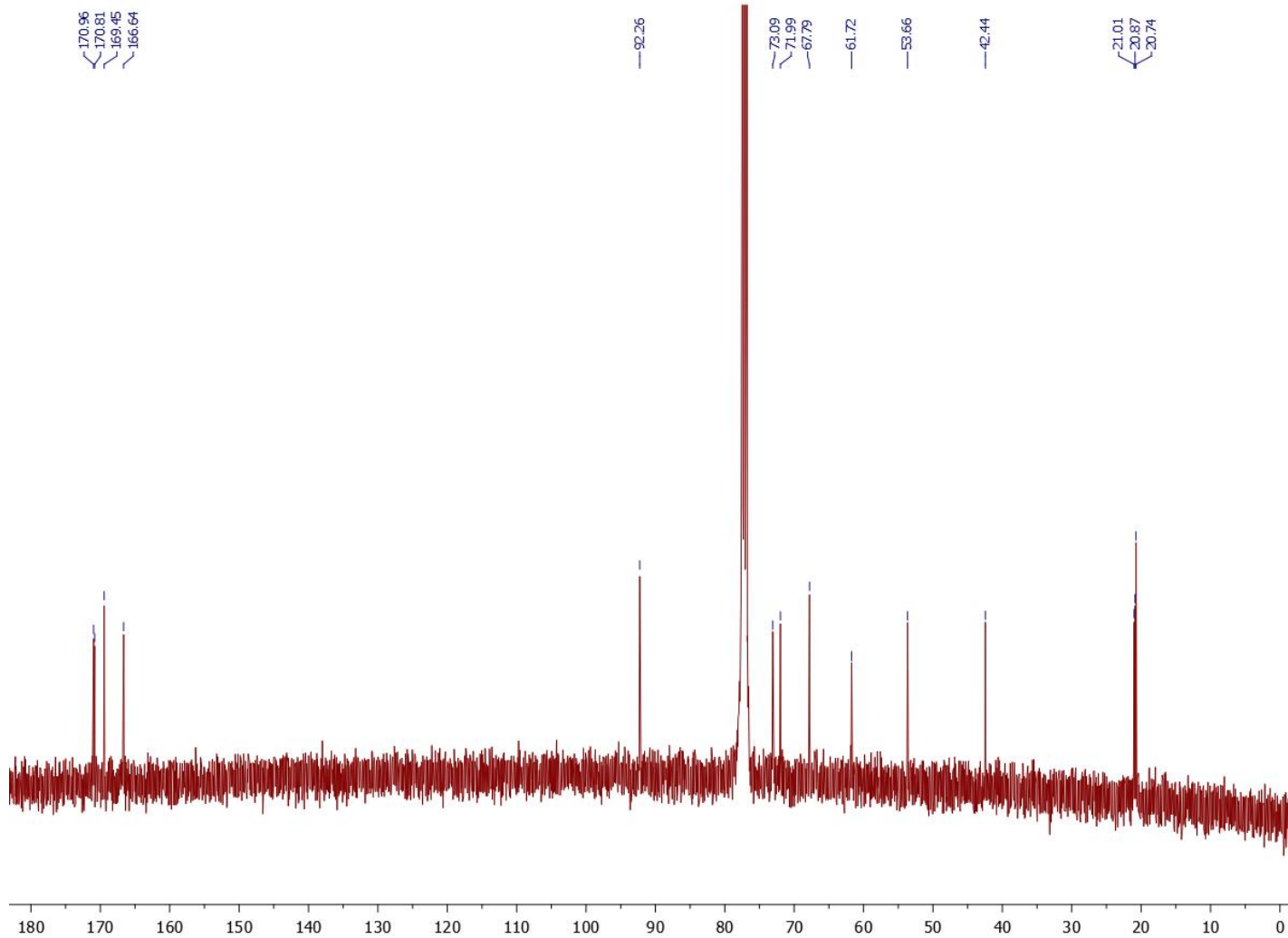


Figure A1.10. ^{13}C NMR of **cmpd 9** in CDCl_3

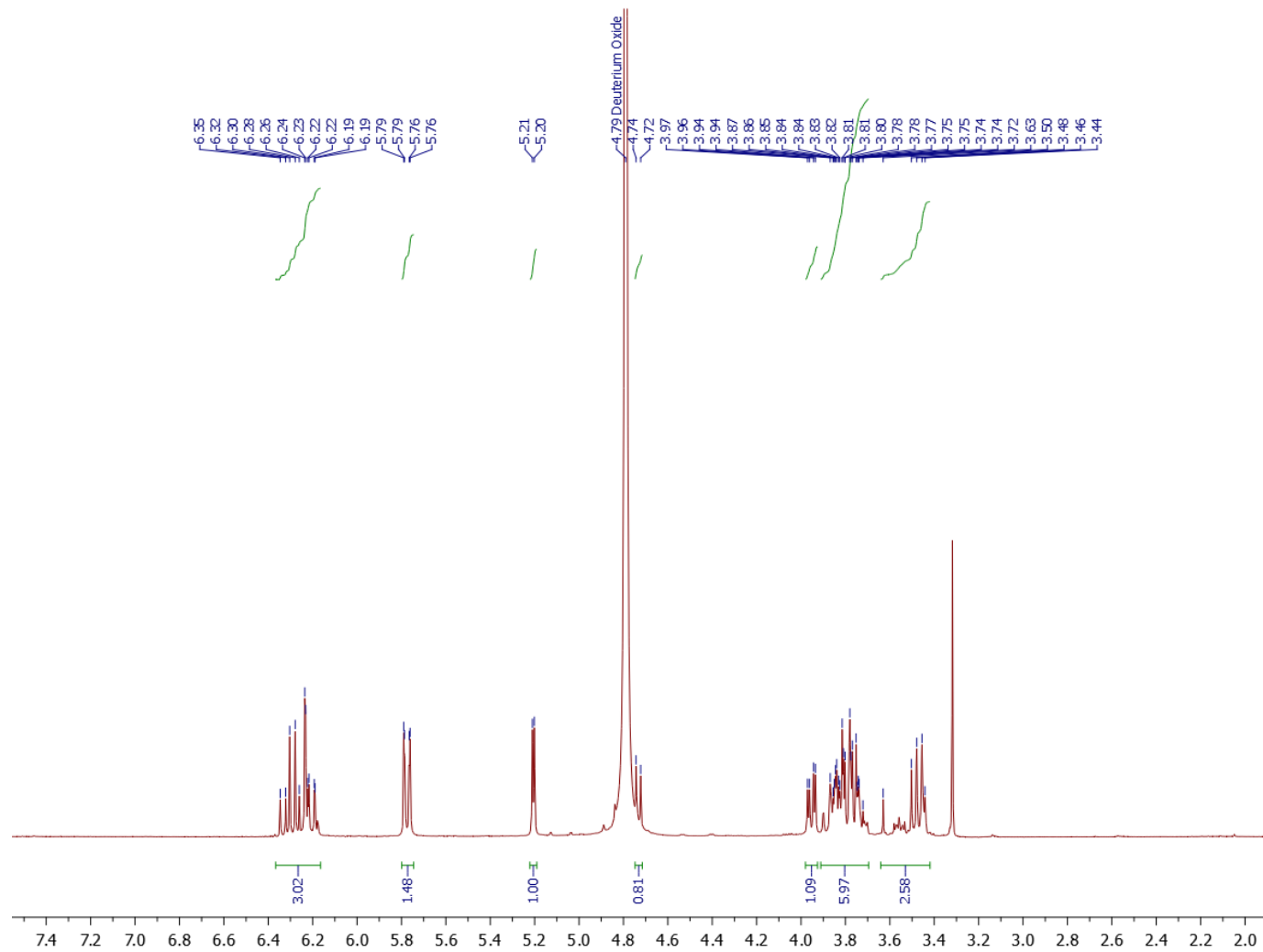


Figure A1.11. ^1H NMR of compd **10** in D_2O

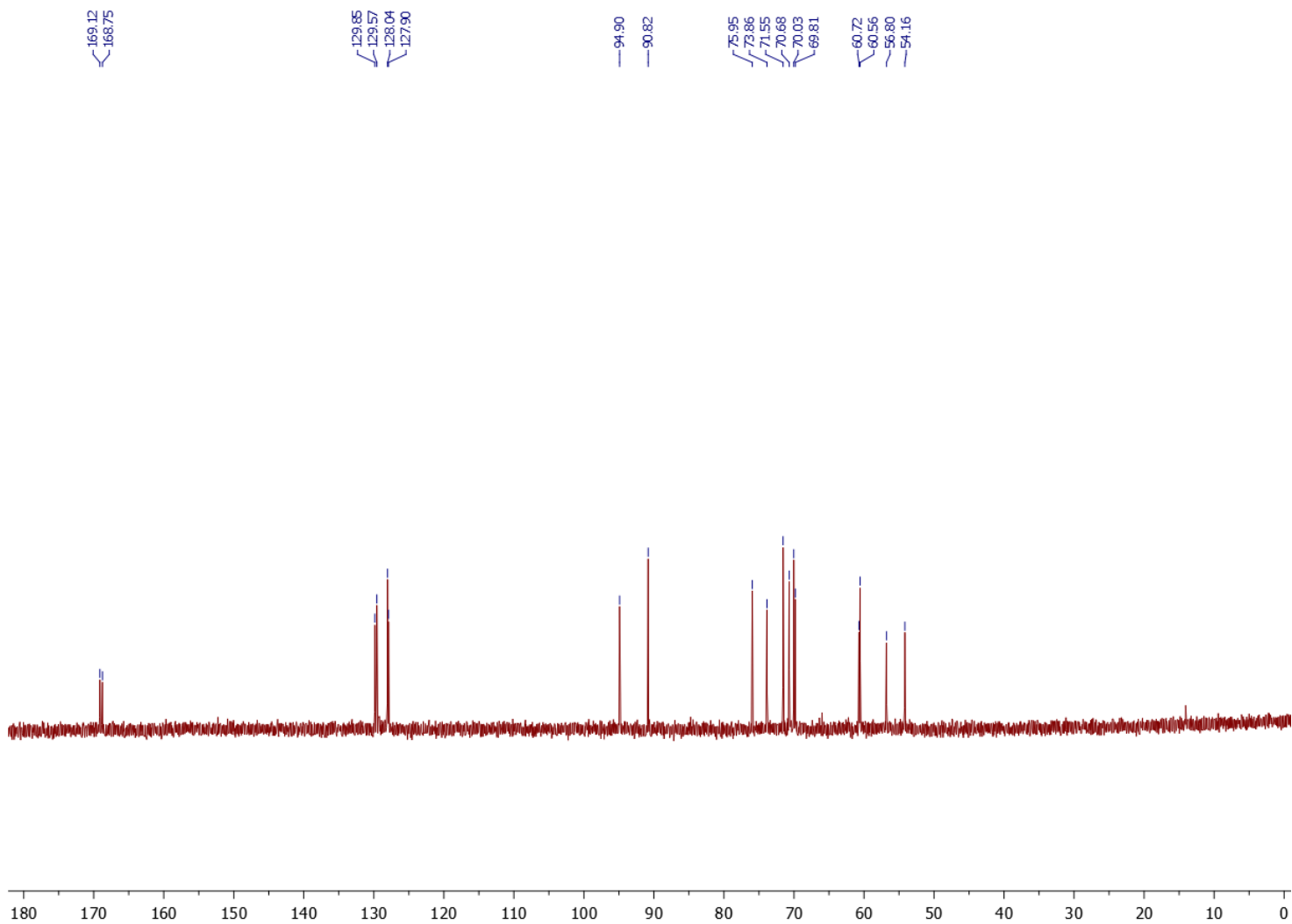


Figure A1.12. ^{13}C NMR of compd **10** in D_2O

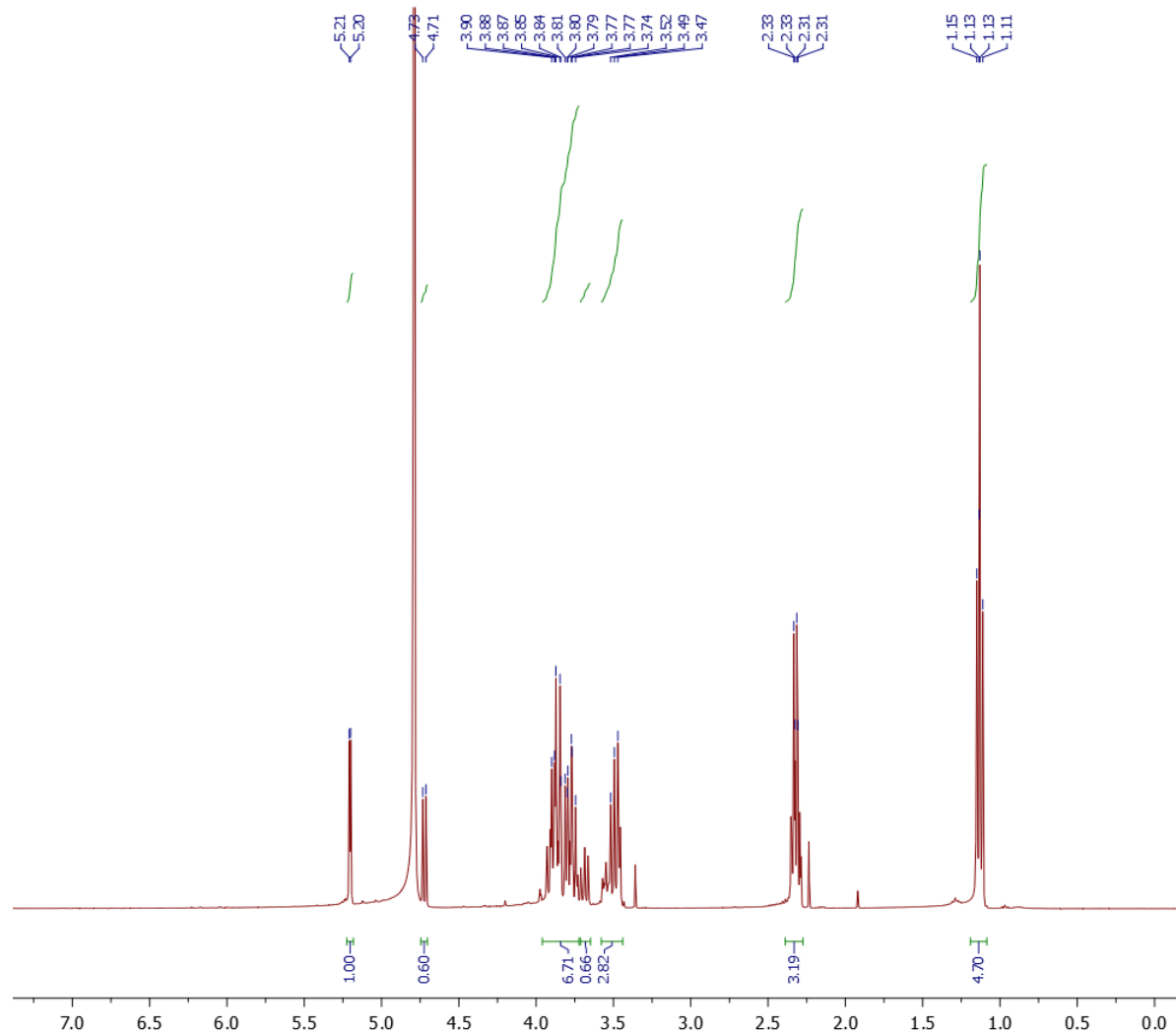


Figure A1.13. ^1H NMR of compd **10s** in D_2O

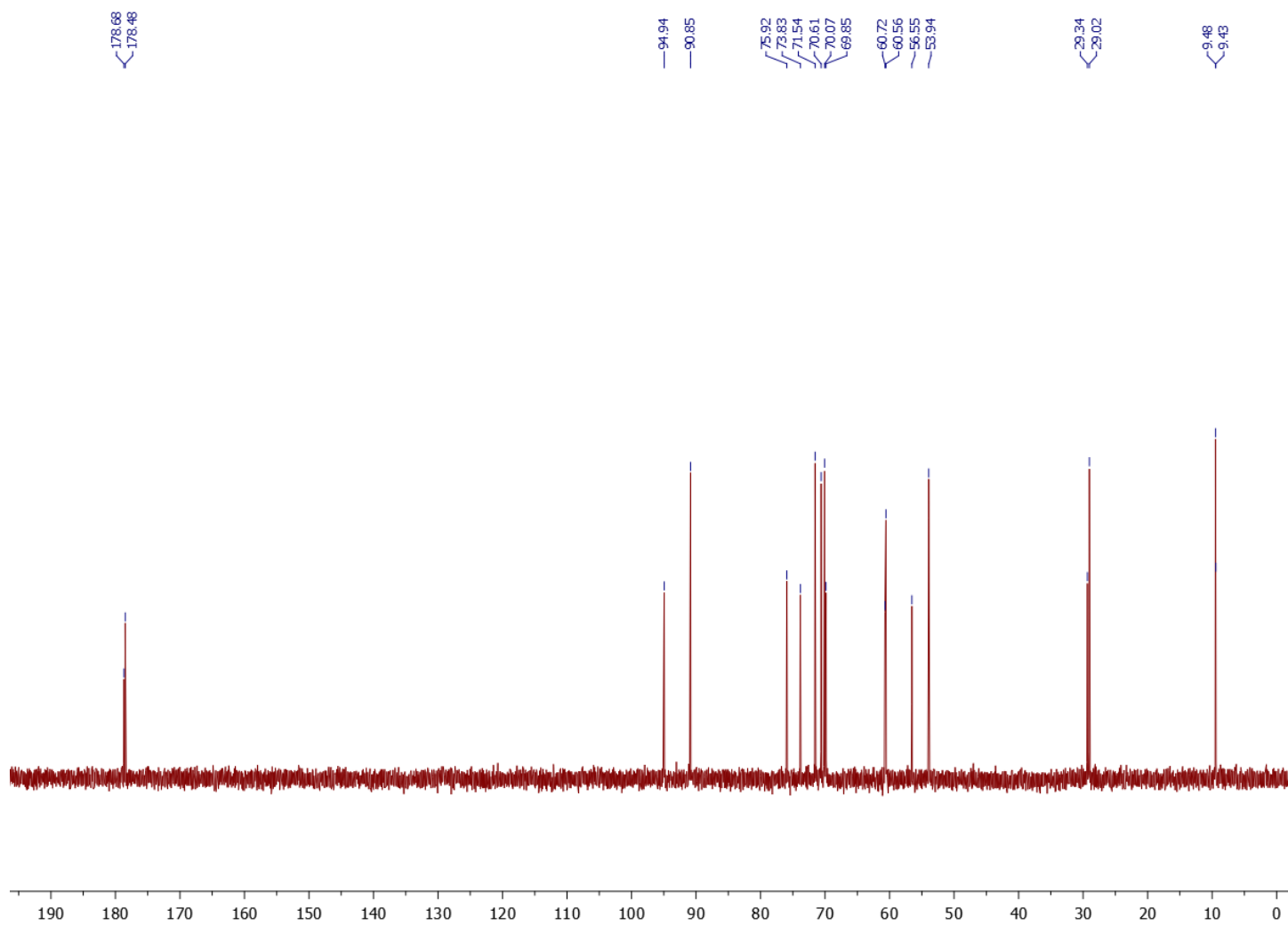


Figure A1.14. ^{13}C NMR of cmpd 10s in D_2O

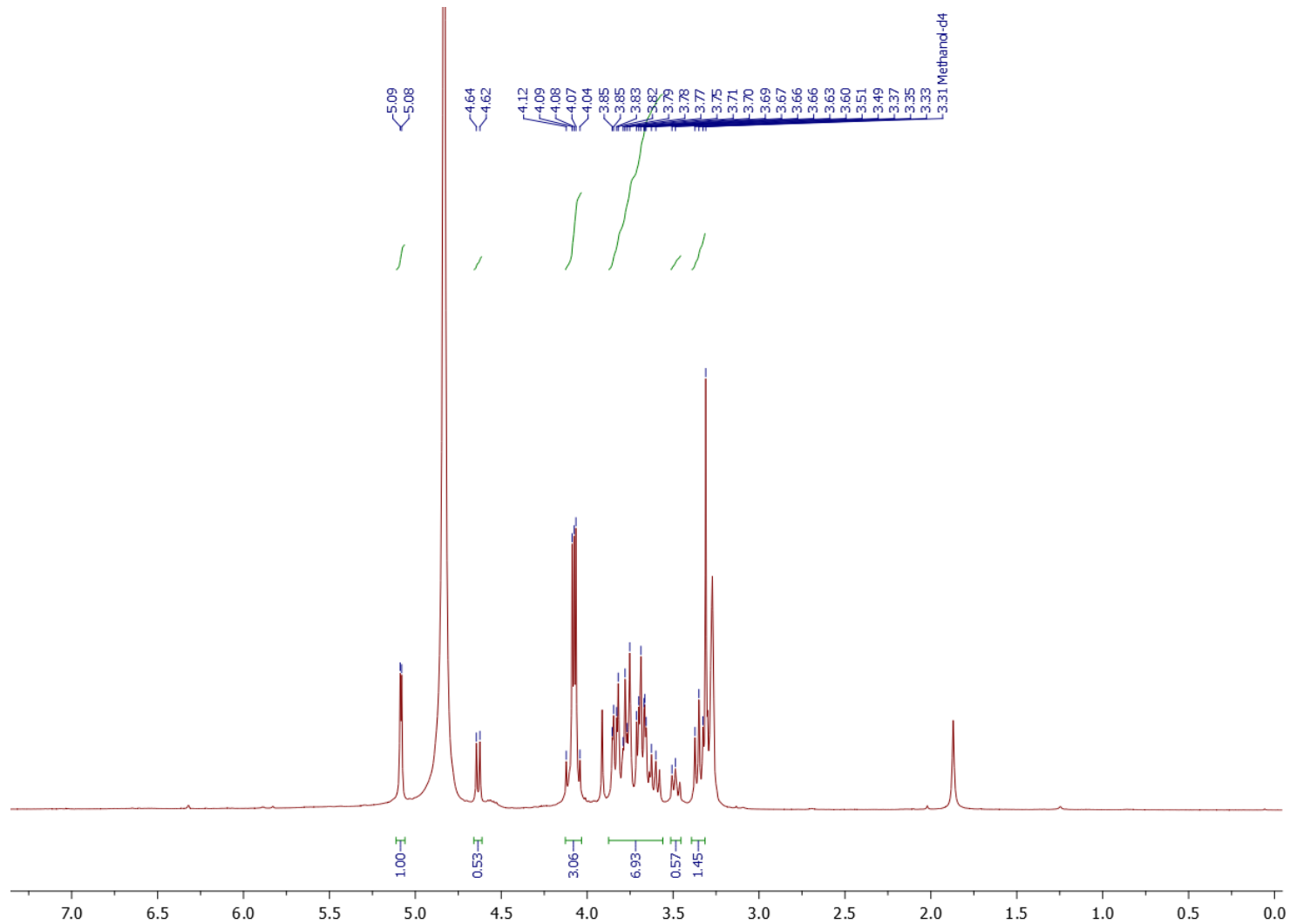


Figure A1.15. ^1H NMR of compd 11 in MeOD

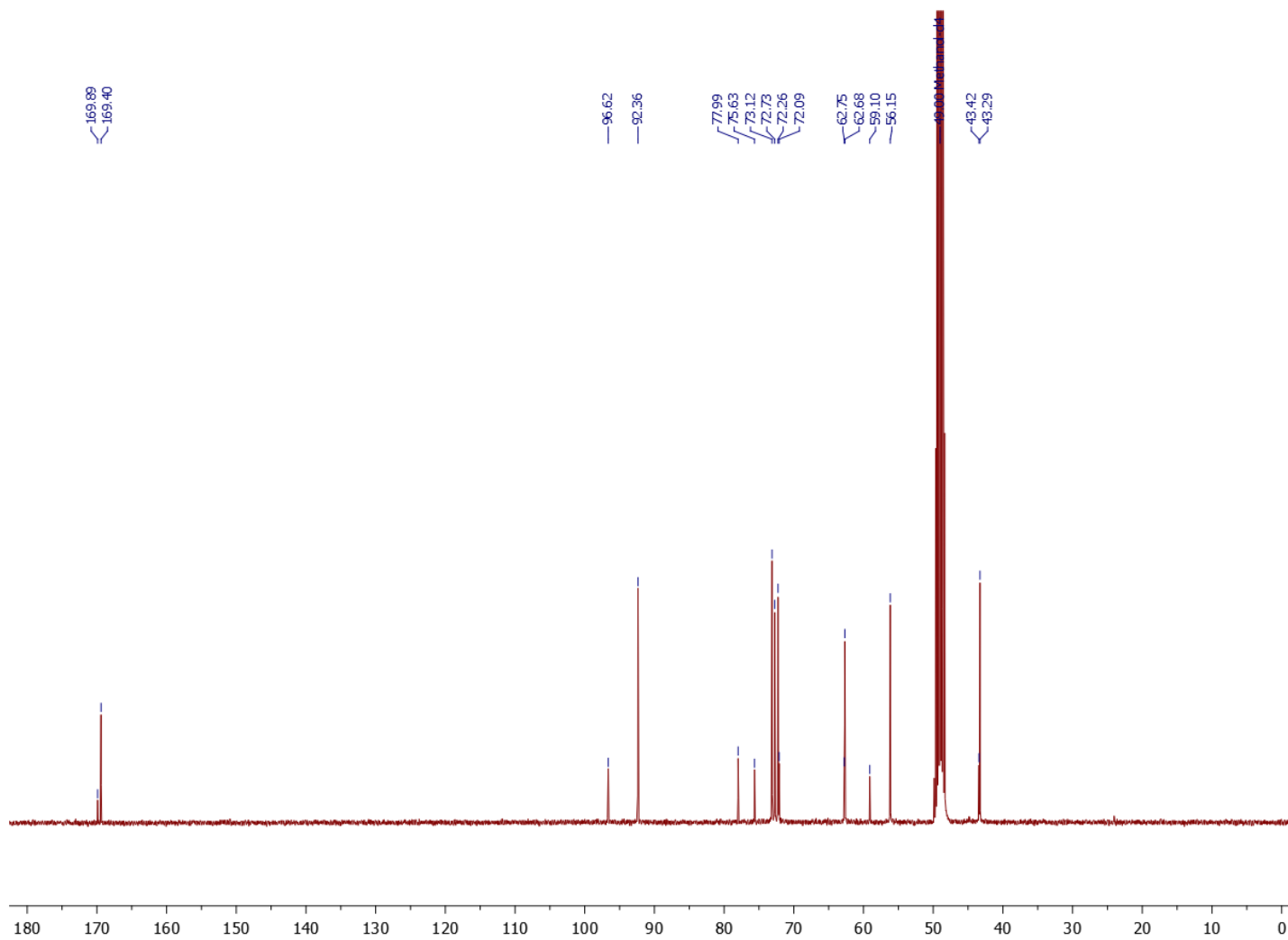


Figure A1.16. ^{13}C NMR of cmpd 11 in MeOD

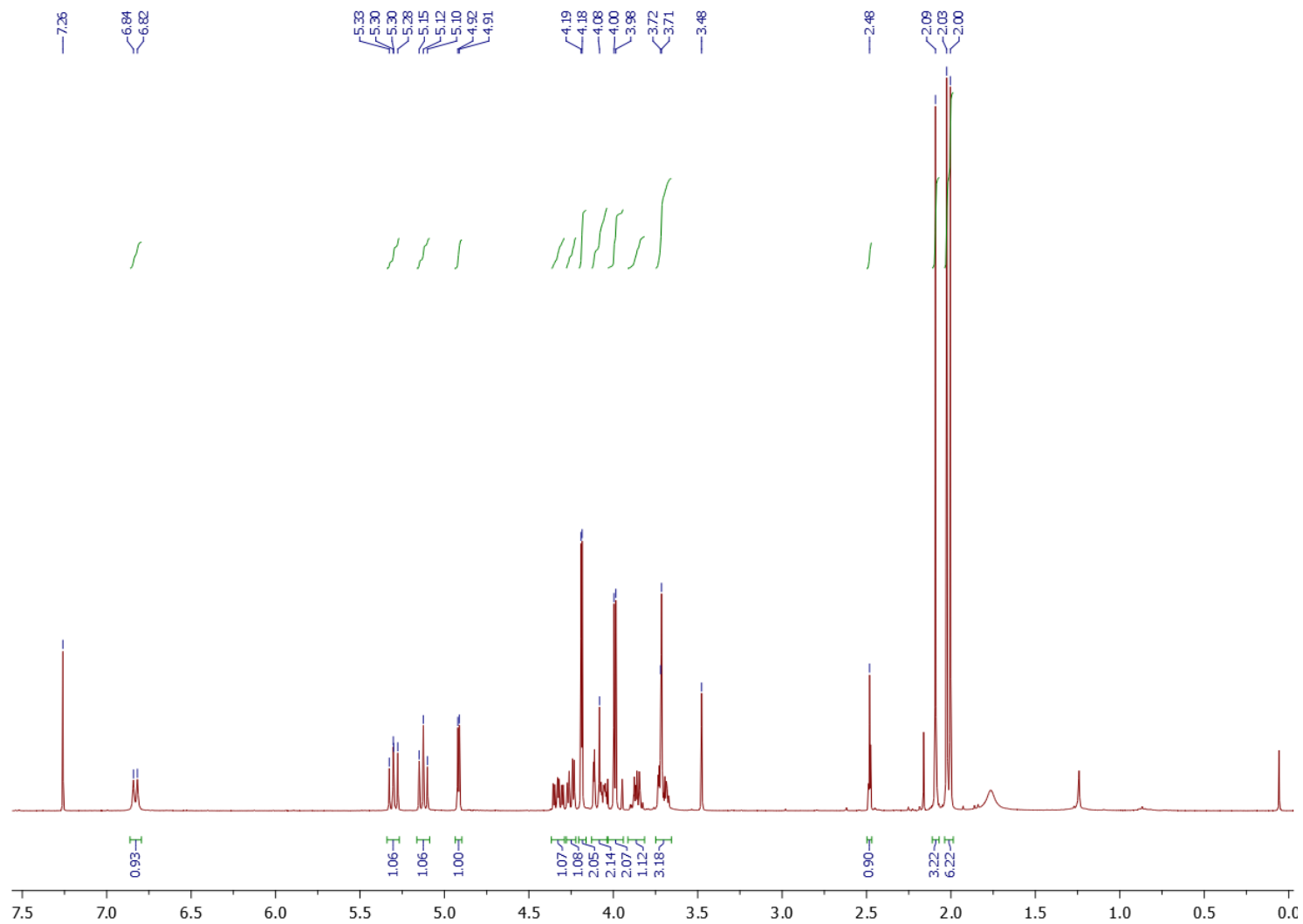


Figure A1.17. ^1H NMR of **cmpd 12- α** in CDCl_3

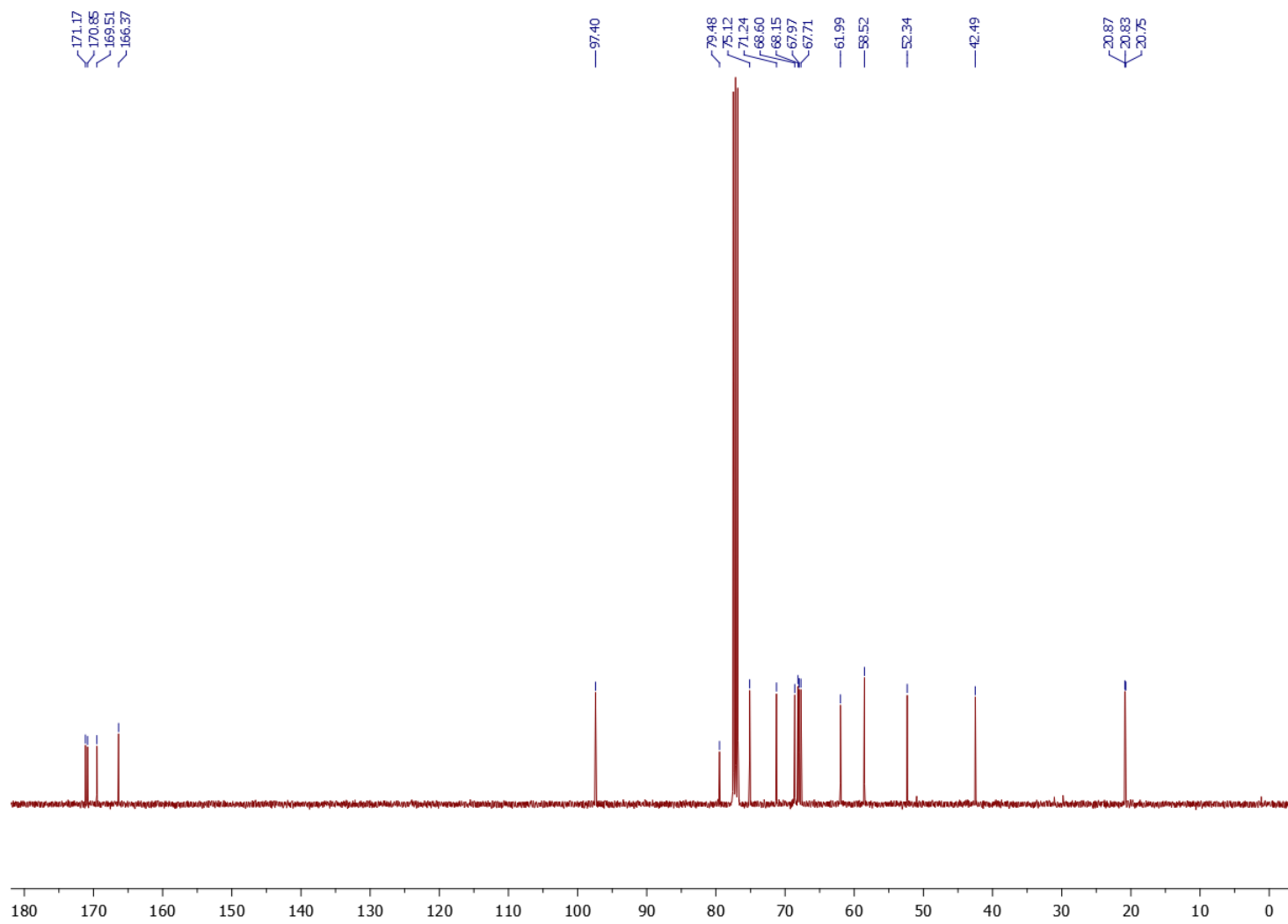


Figure A1.18. ^{13}C NMR of compd **12- α** in CDCl_3

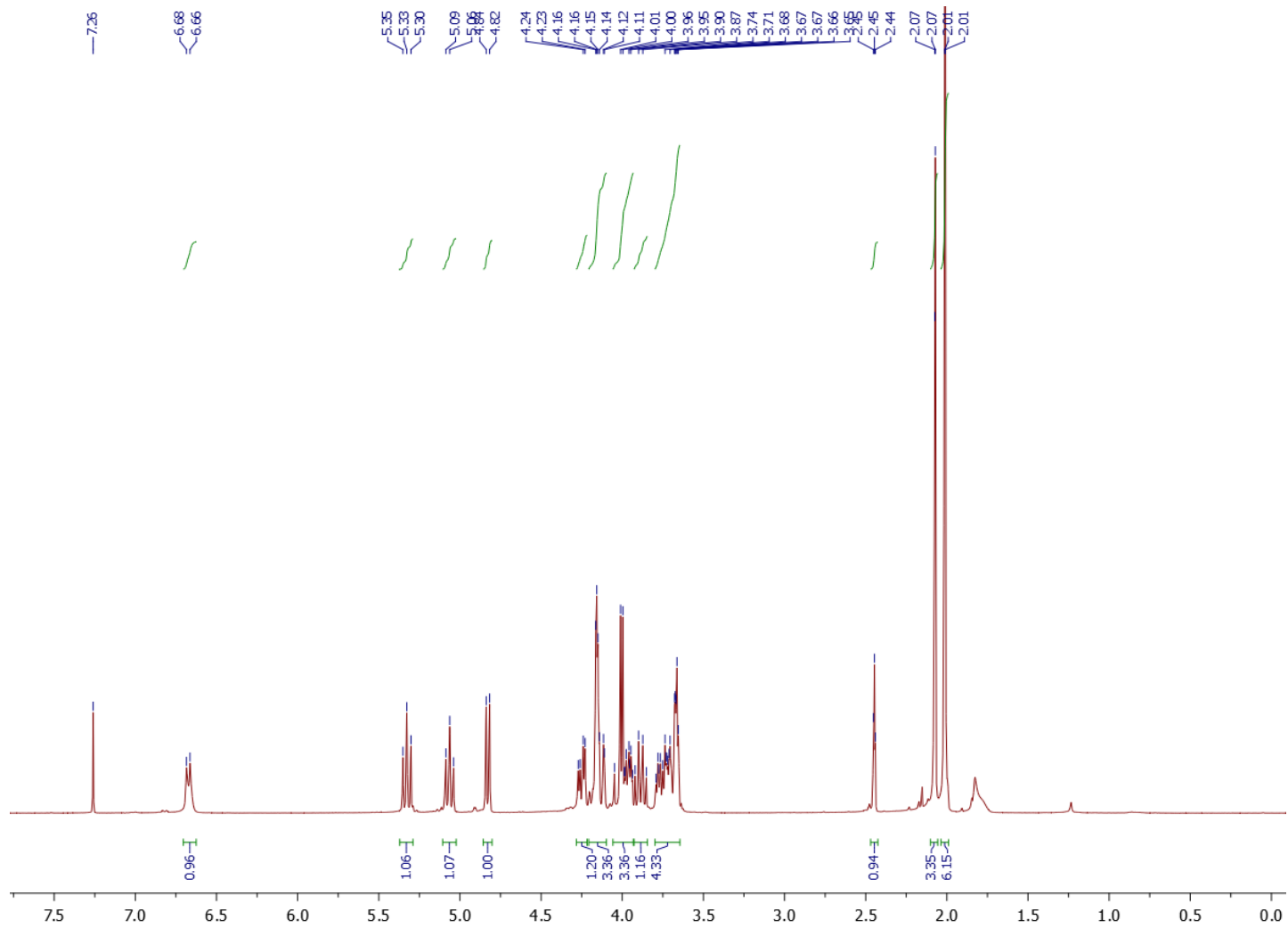


Figure A1.19. ^1H NMR of compd **12- β** in CDCl_3

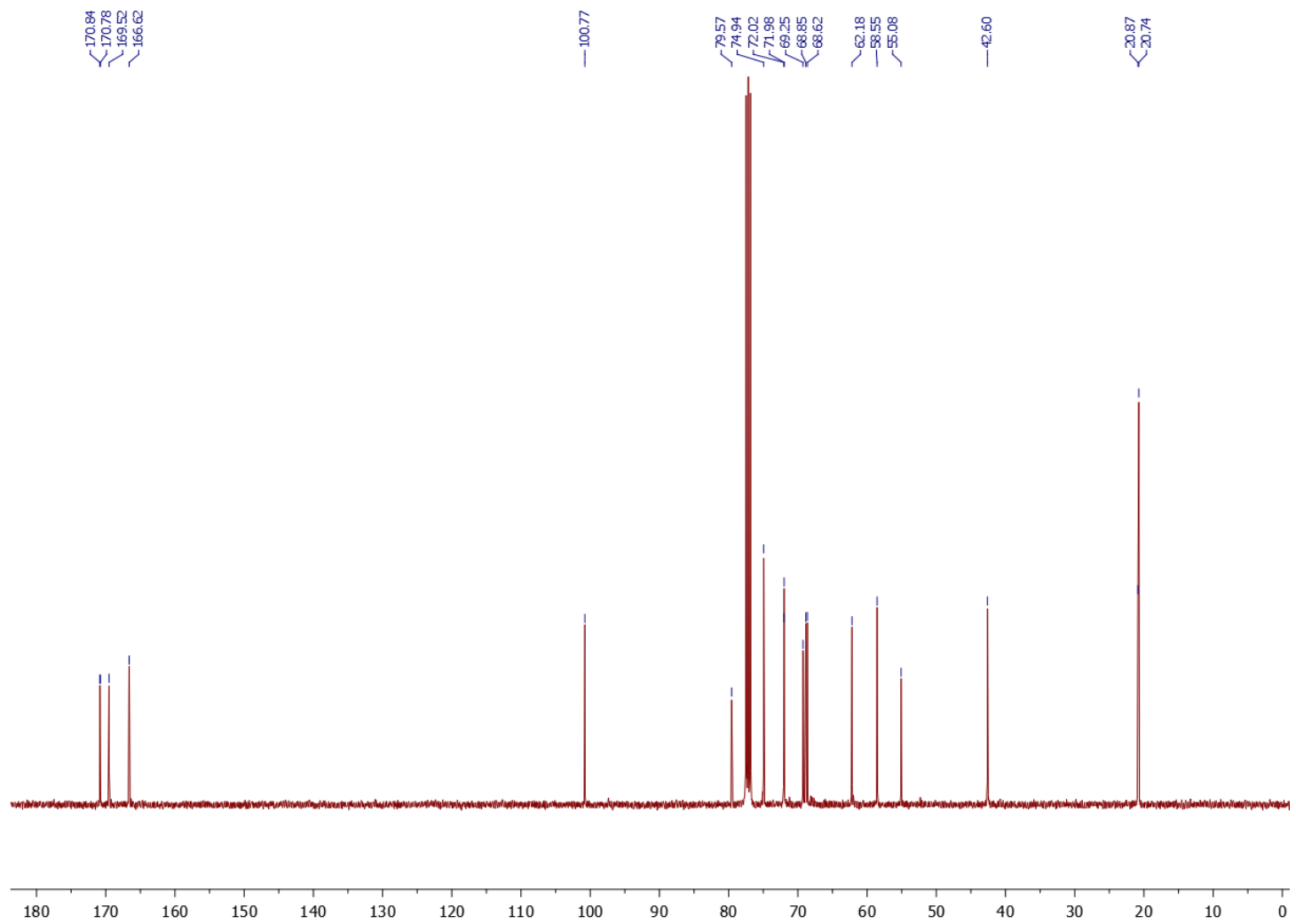


Figure A1.20. ^{13}C NMR of compd **12- β** in CDCl_3

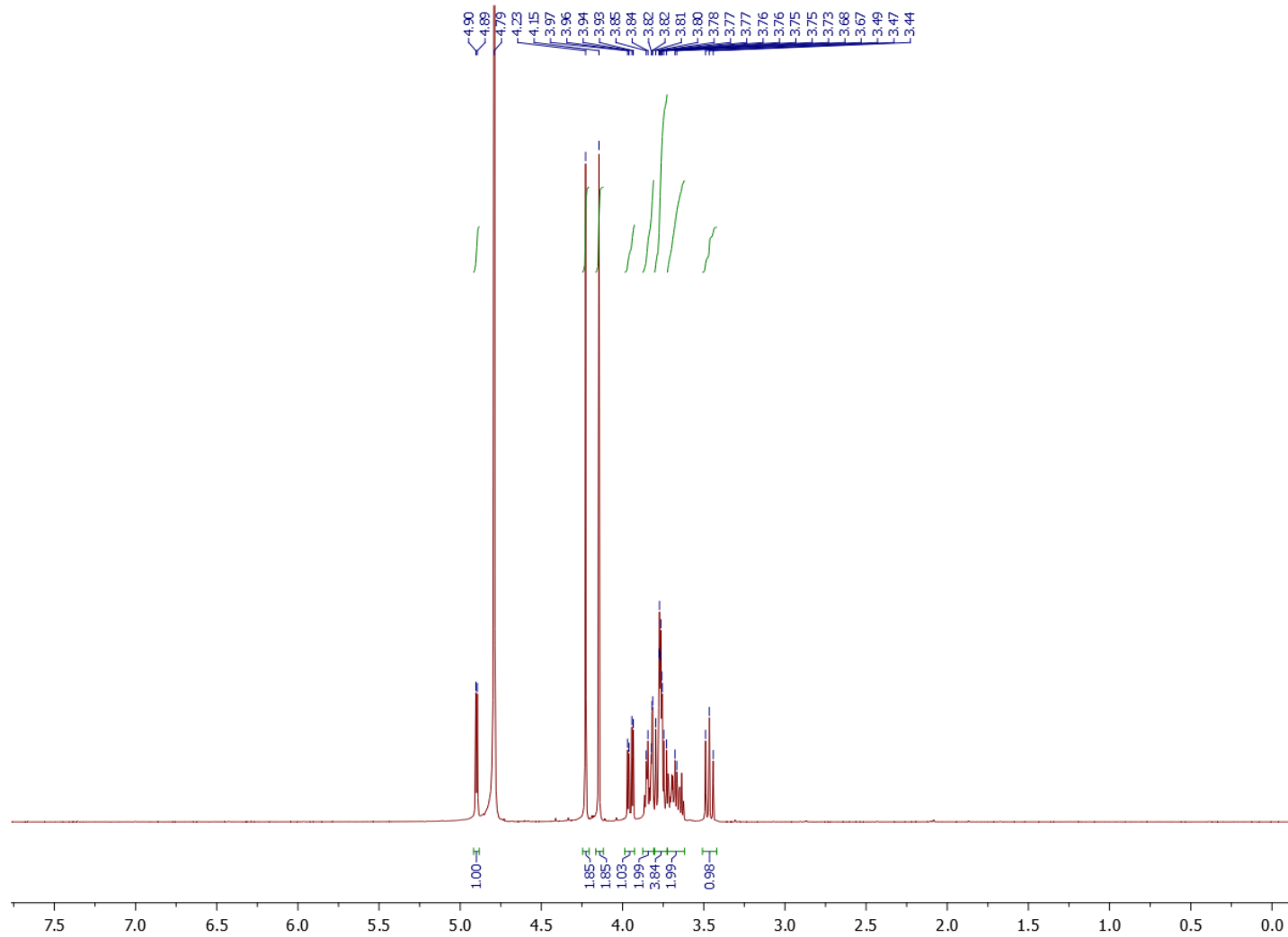


Figure A1.21. ^1H NMR of compd **13- α** in D_2O

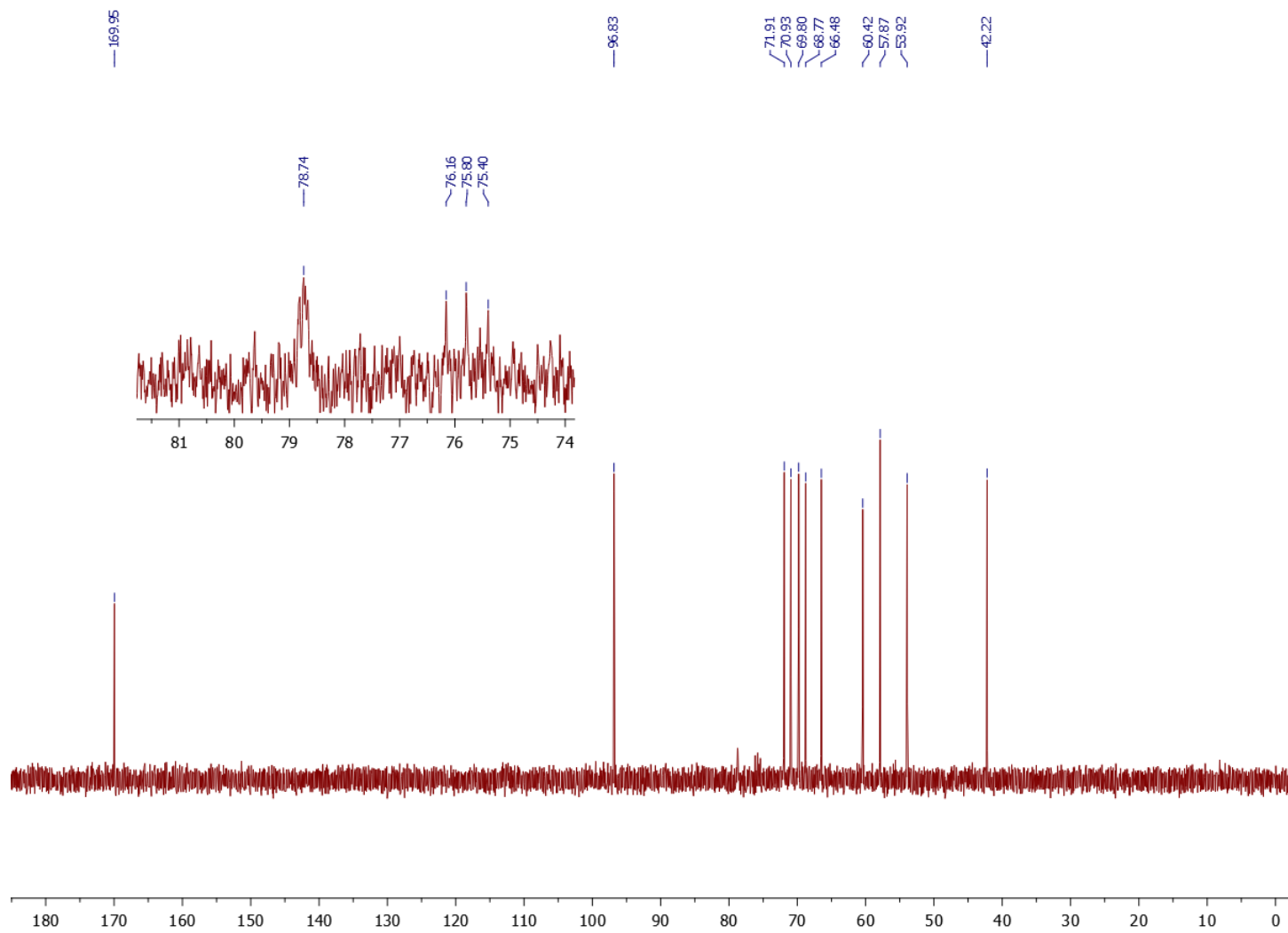


Figure A1.22. ^{13}C NMR of compd **13- α** in D_2O

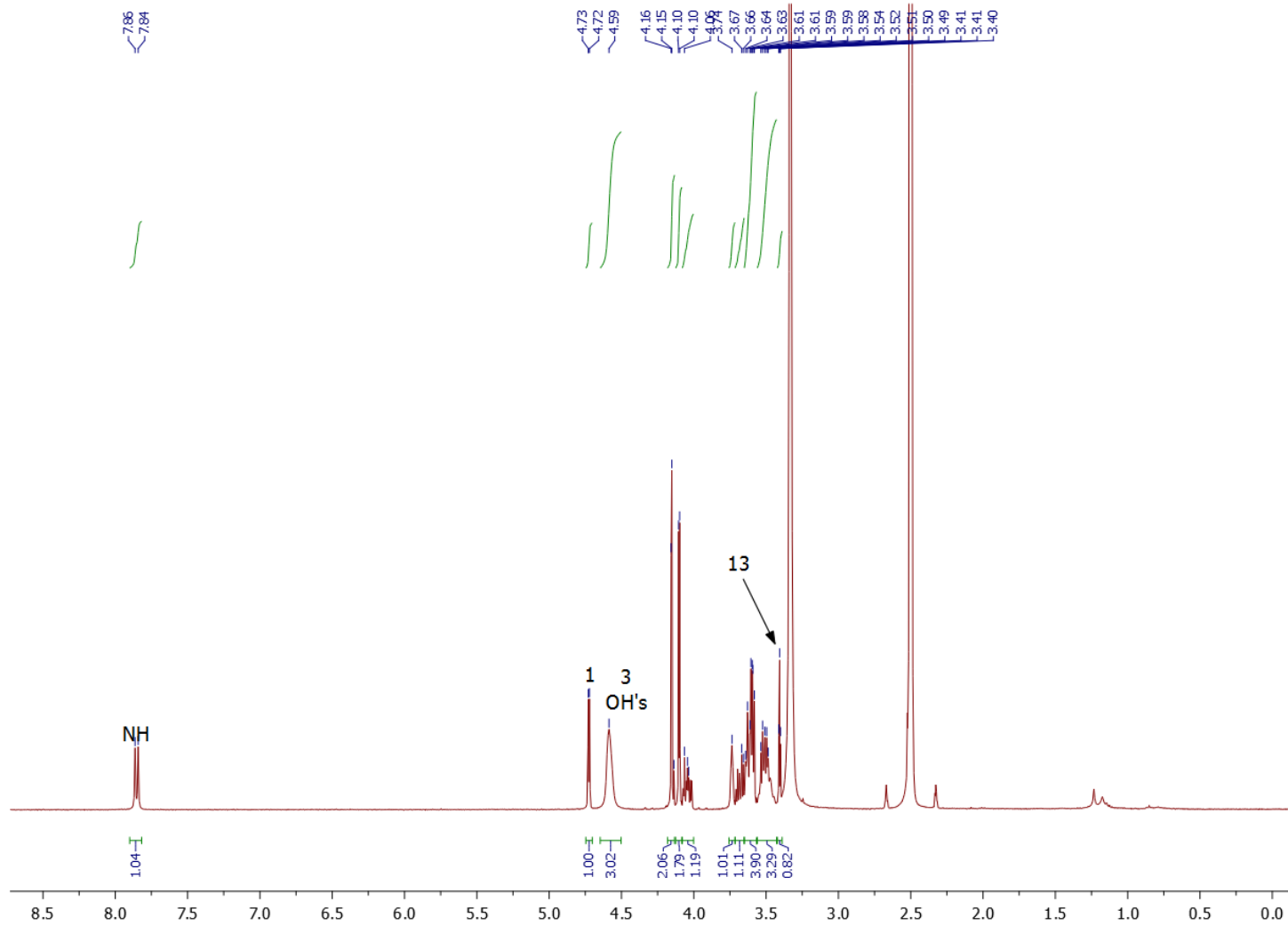


Figure A1.23. ^1H NMR of cmpd **13- α** in d_6 -DMSO. Terminal alkyne proton H_{13} is present at 3.41ppm (triplet)

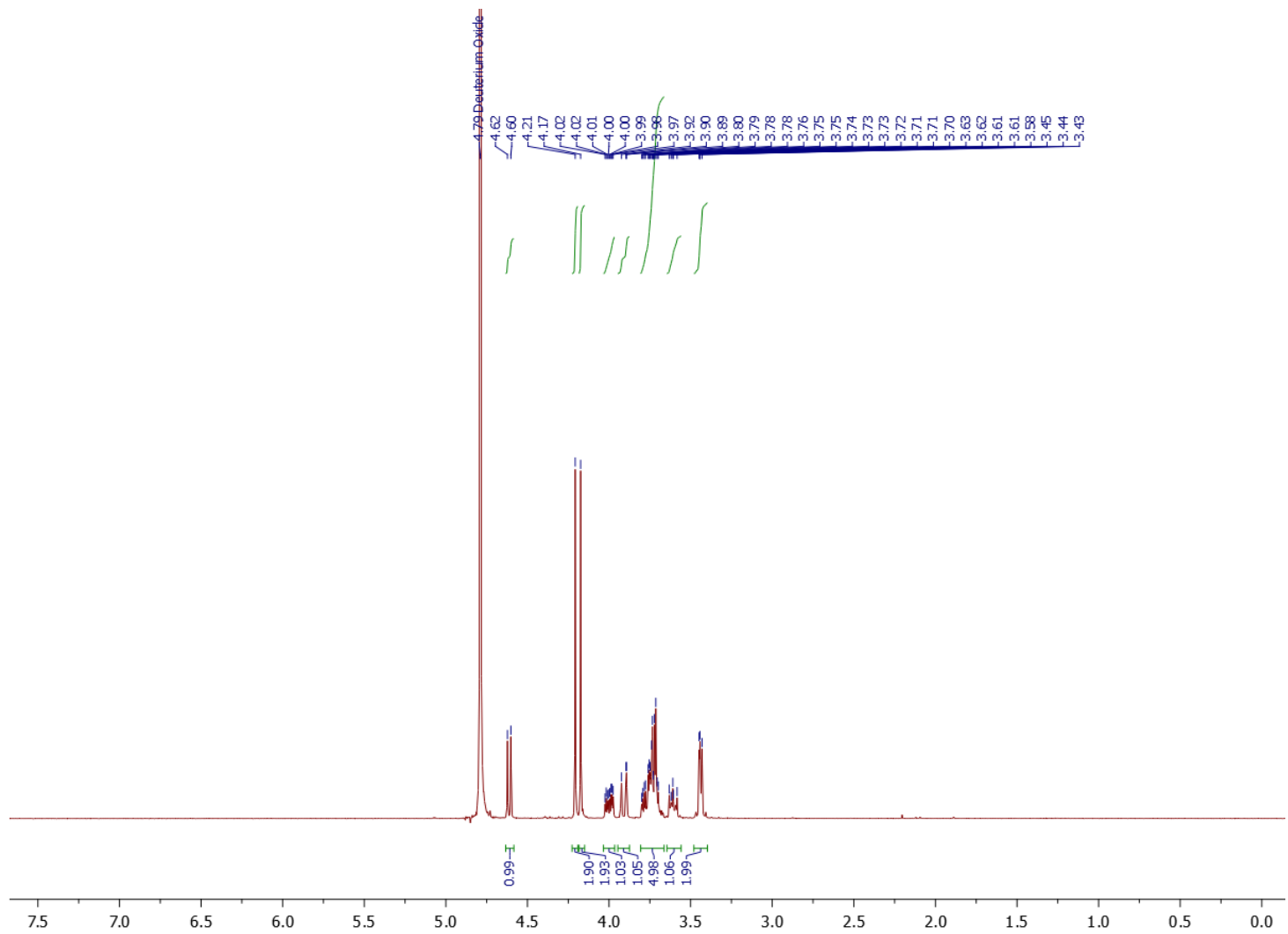


Figure A1.24. ^1H NMR of **cmpd 13- β** in D_2O

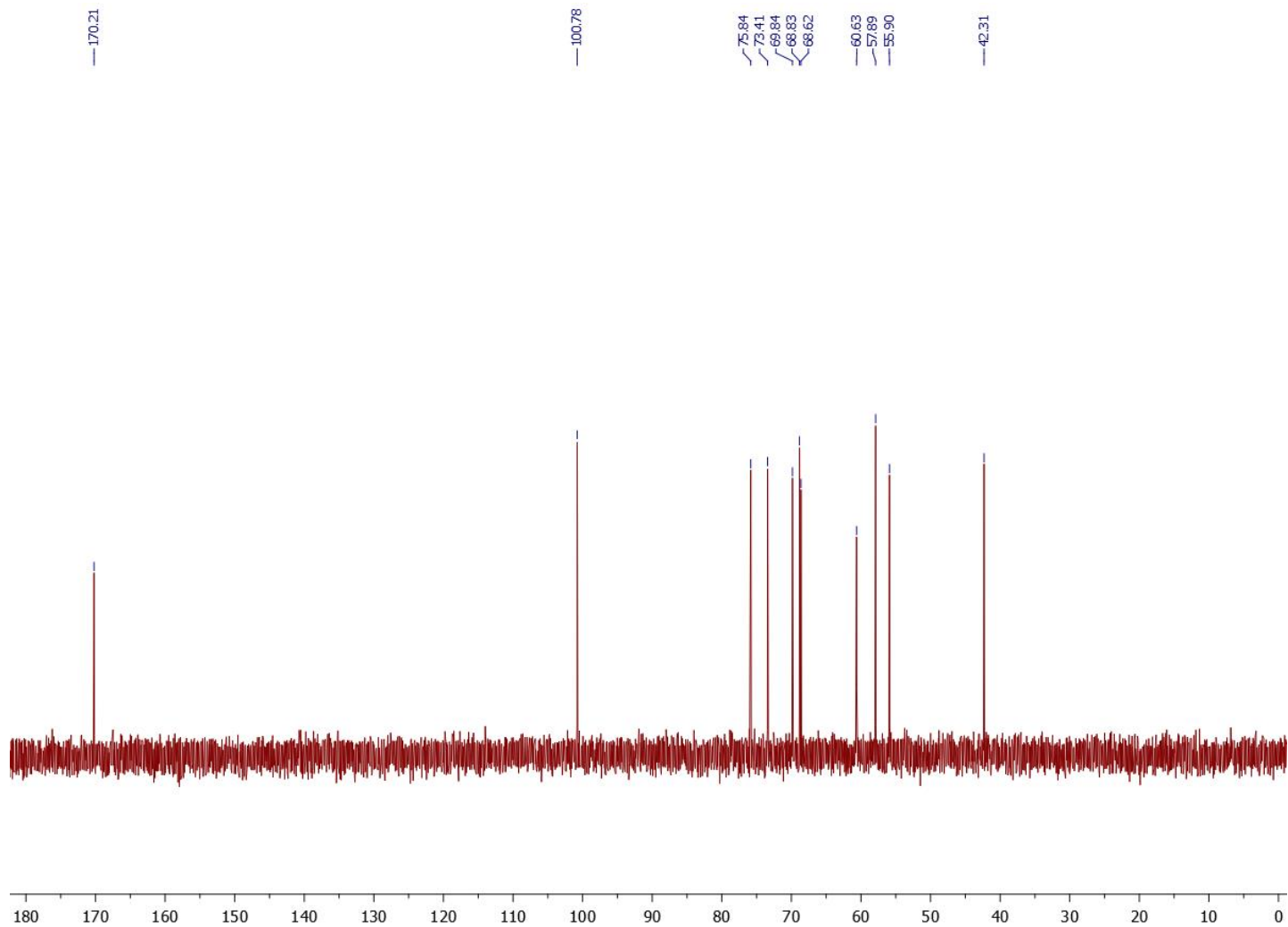


Figure A1.25. ^{13}C NMR of cmpd **13- β** in D_2O

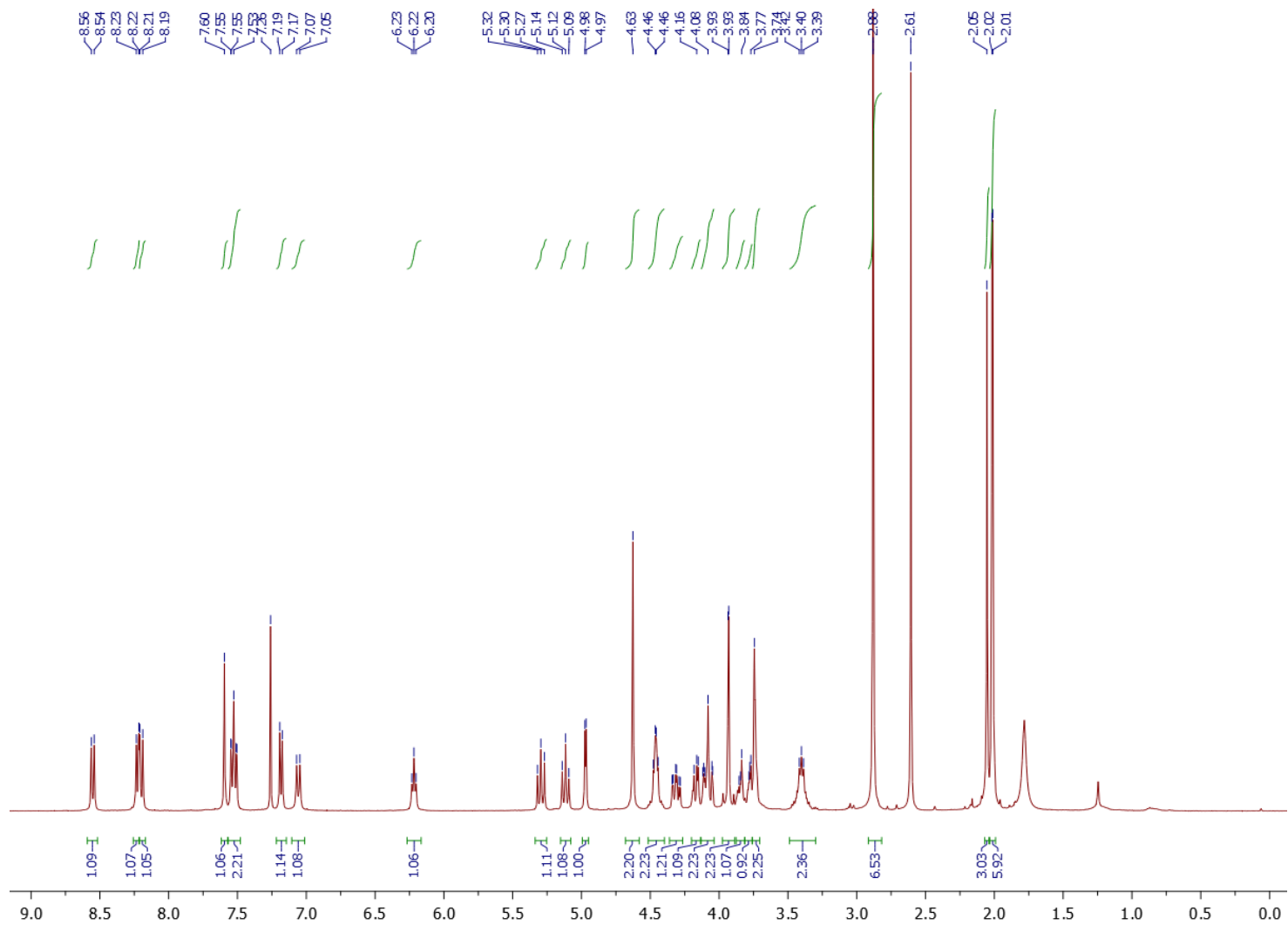


Figure A1.26. ^1H NMR of cmpd **14- α** in CDCl_3

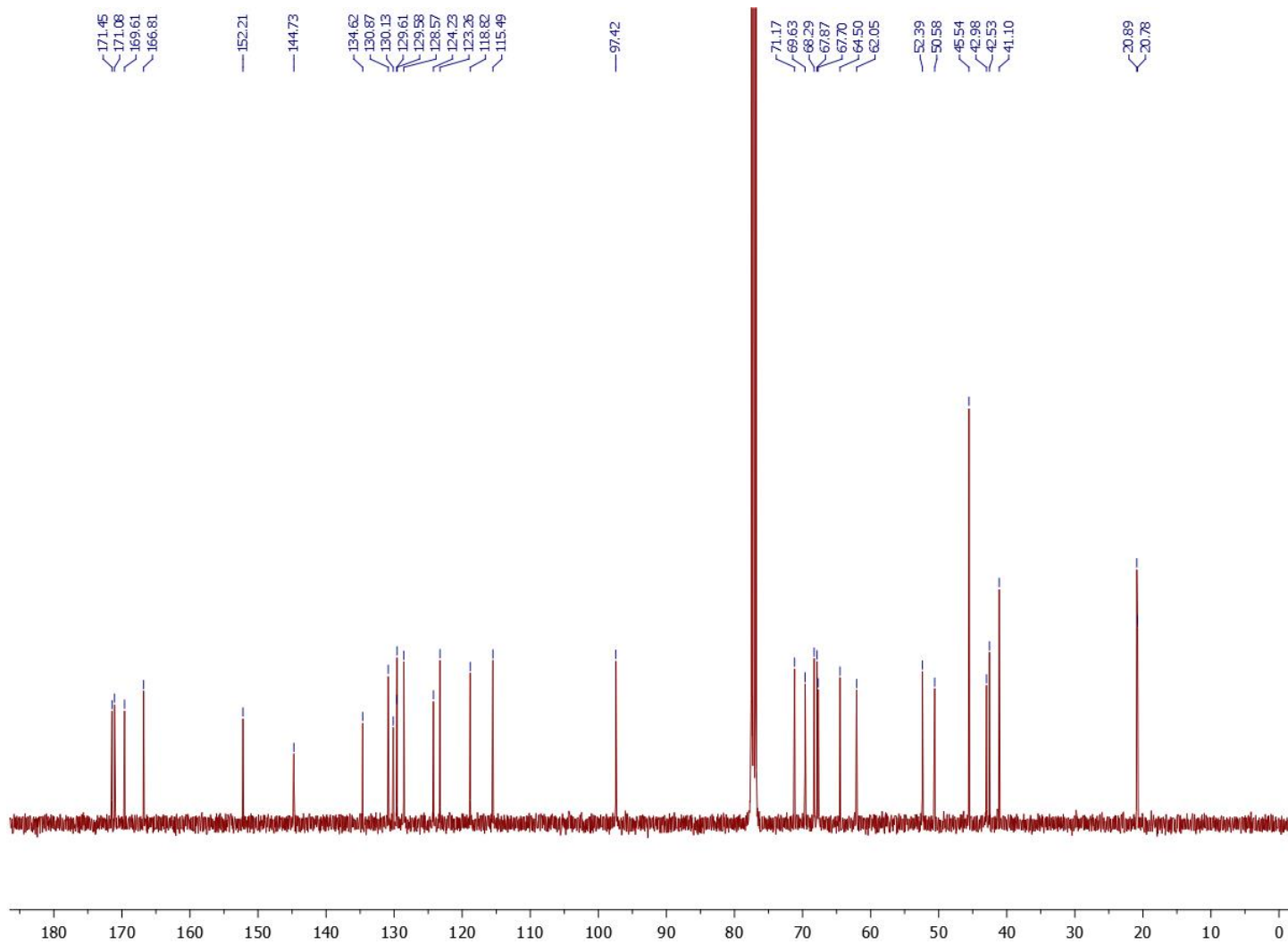


Figure A1.27. ^{13}C NMR of compd **14- α** in CDCl_3

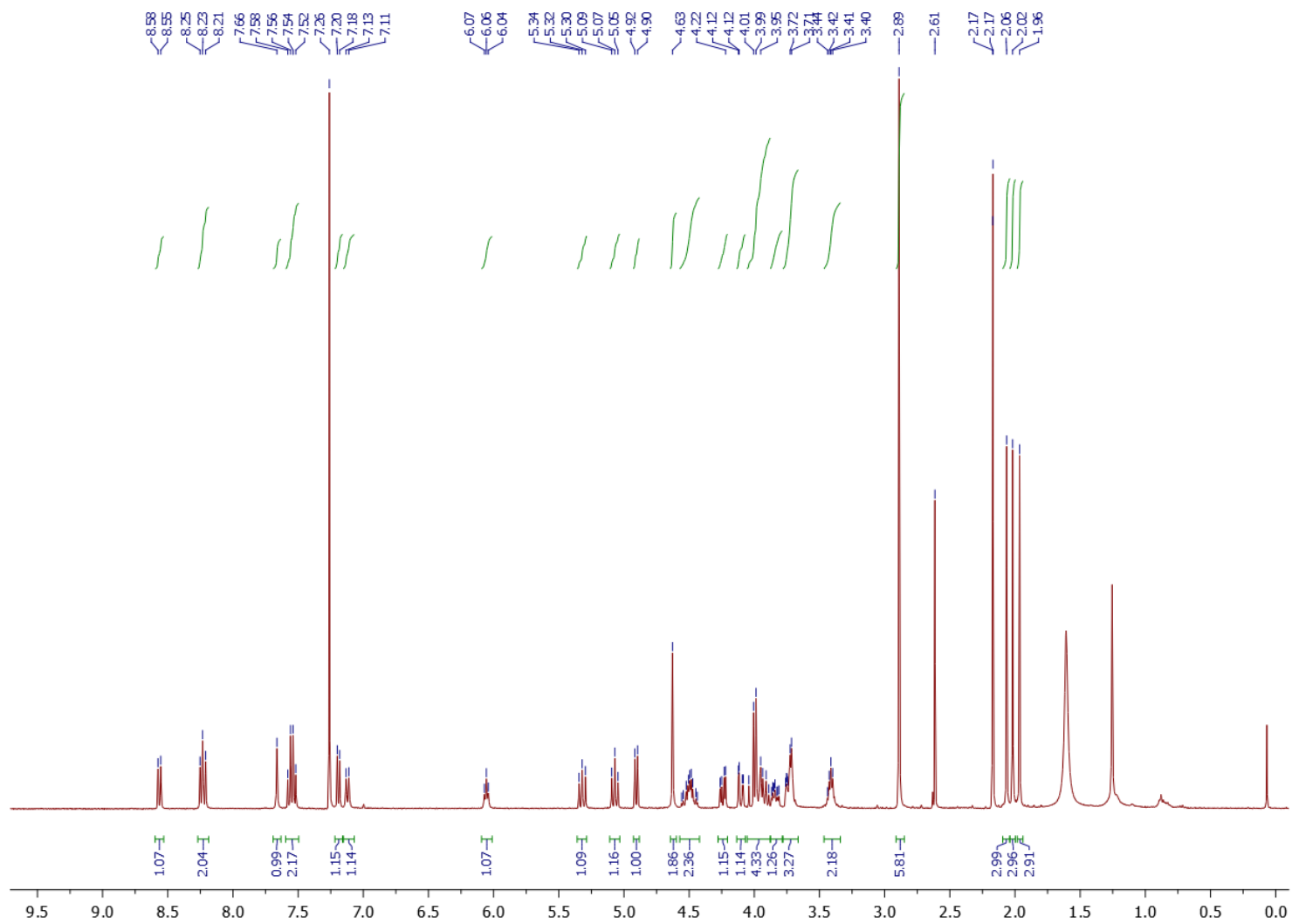


Figure A1.28. ^1H NMR of compd **14- β** in CDCl_3

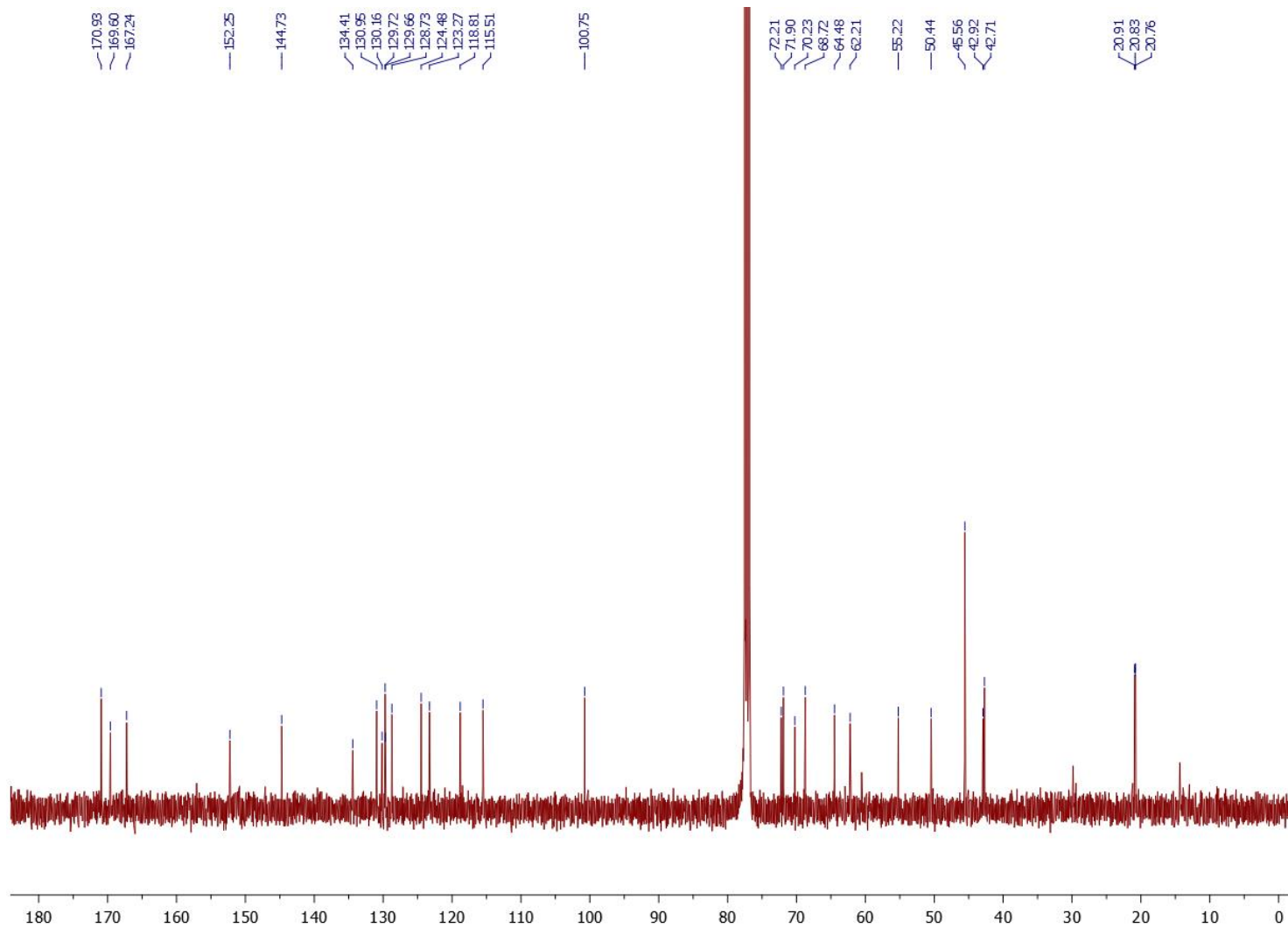


Figure A1.29. ¹³C NMR of compd 14-β in CDCl₃

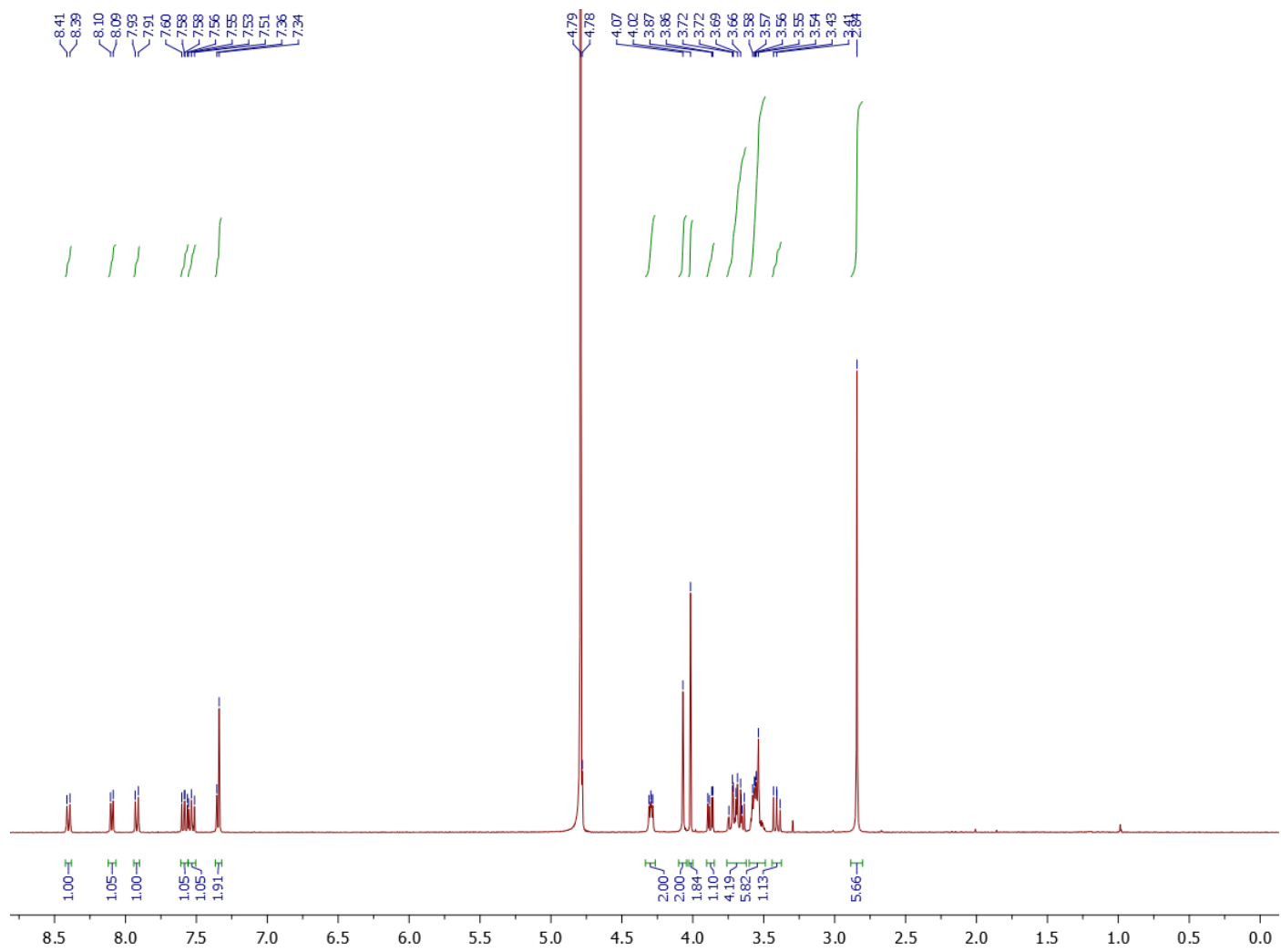


Figure A1.30. ^1H NMR of compd 15- α in D_2O

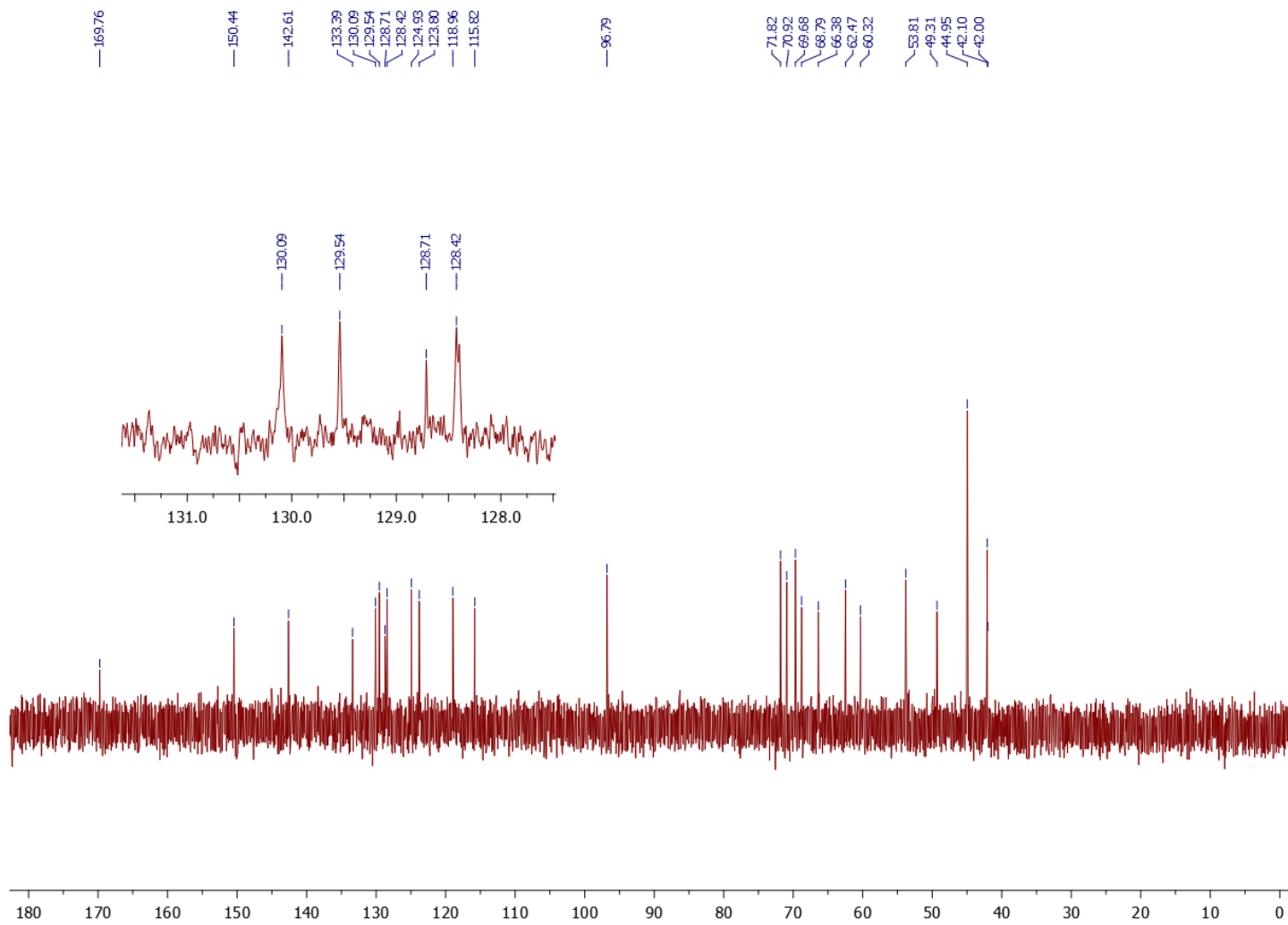


Figure A1.31. ^{13}C NMR of cmpd **15- α** in D_2O

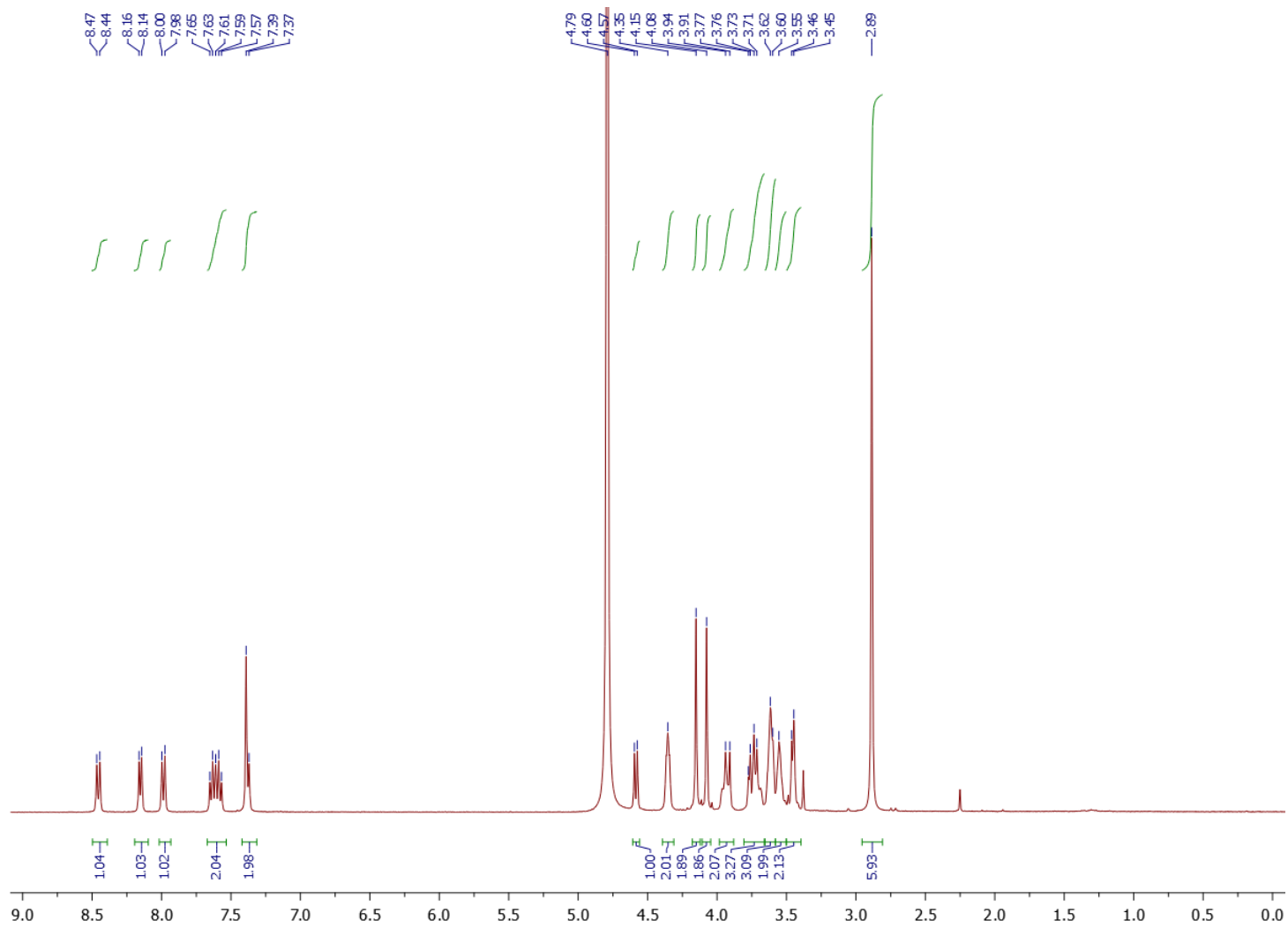


Figure A1.32. ^1H NMR of compd **15- β** in D_2O

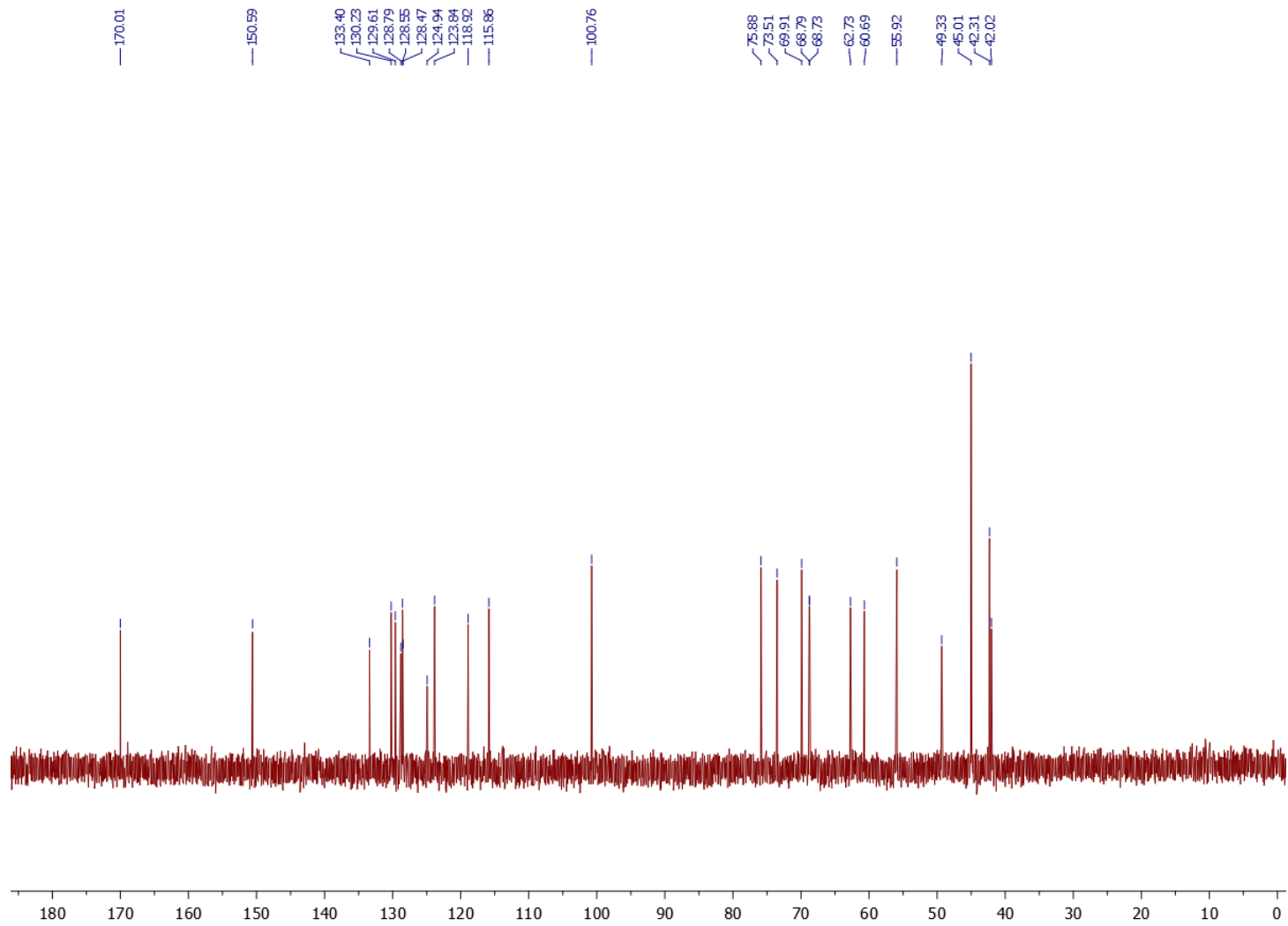


Figure A1.33. ^{13}C NMR of compd **15- β** in D_2O

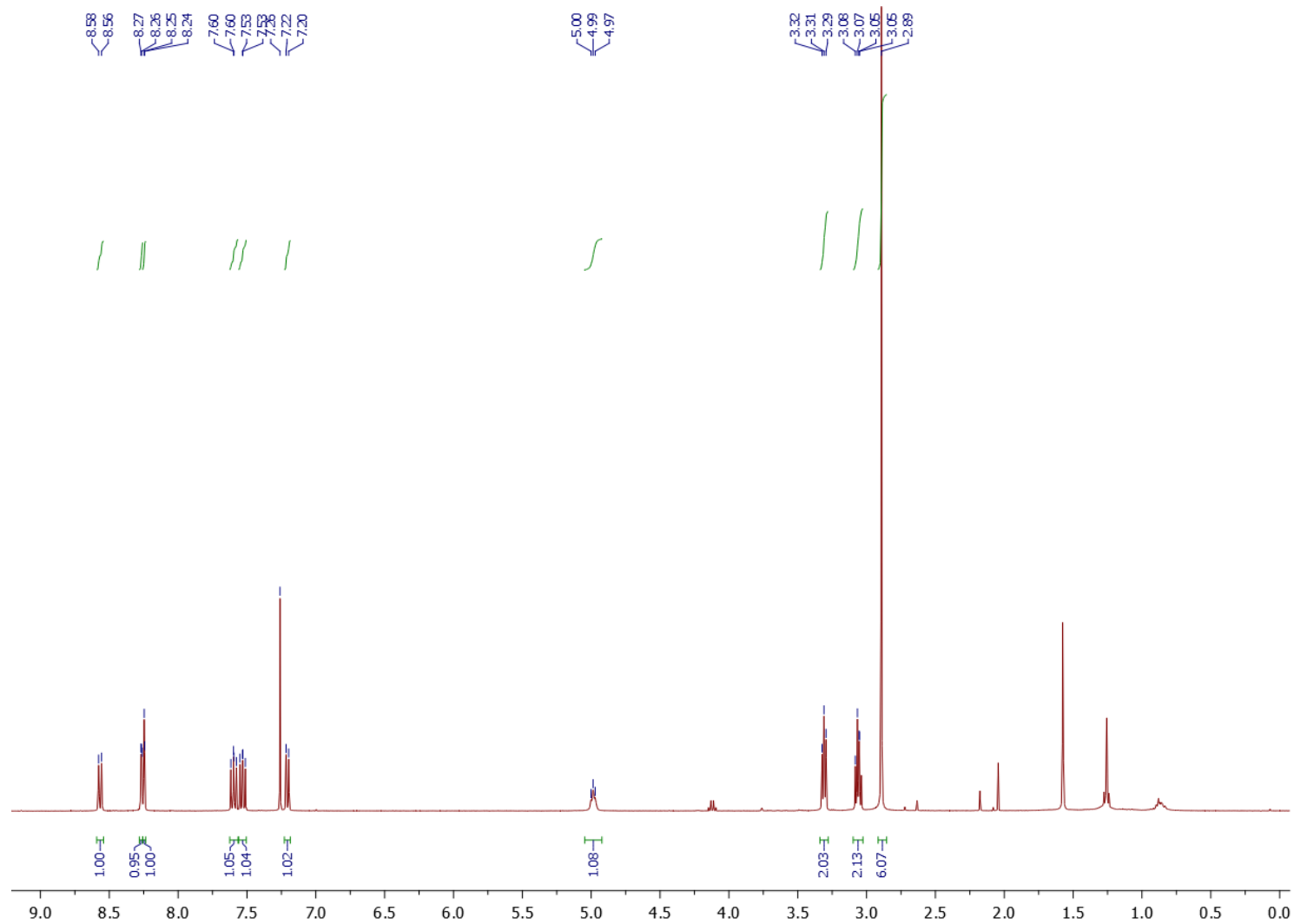


Figure A1.34. ^1H NMR of compd **16** in CDCl_3

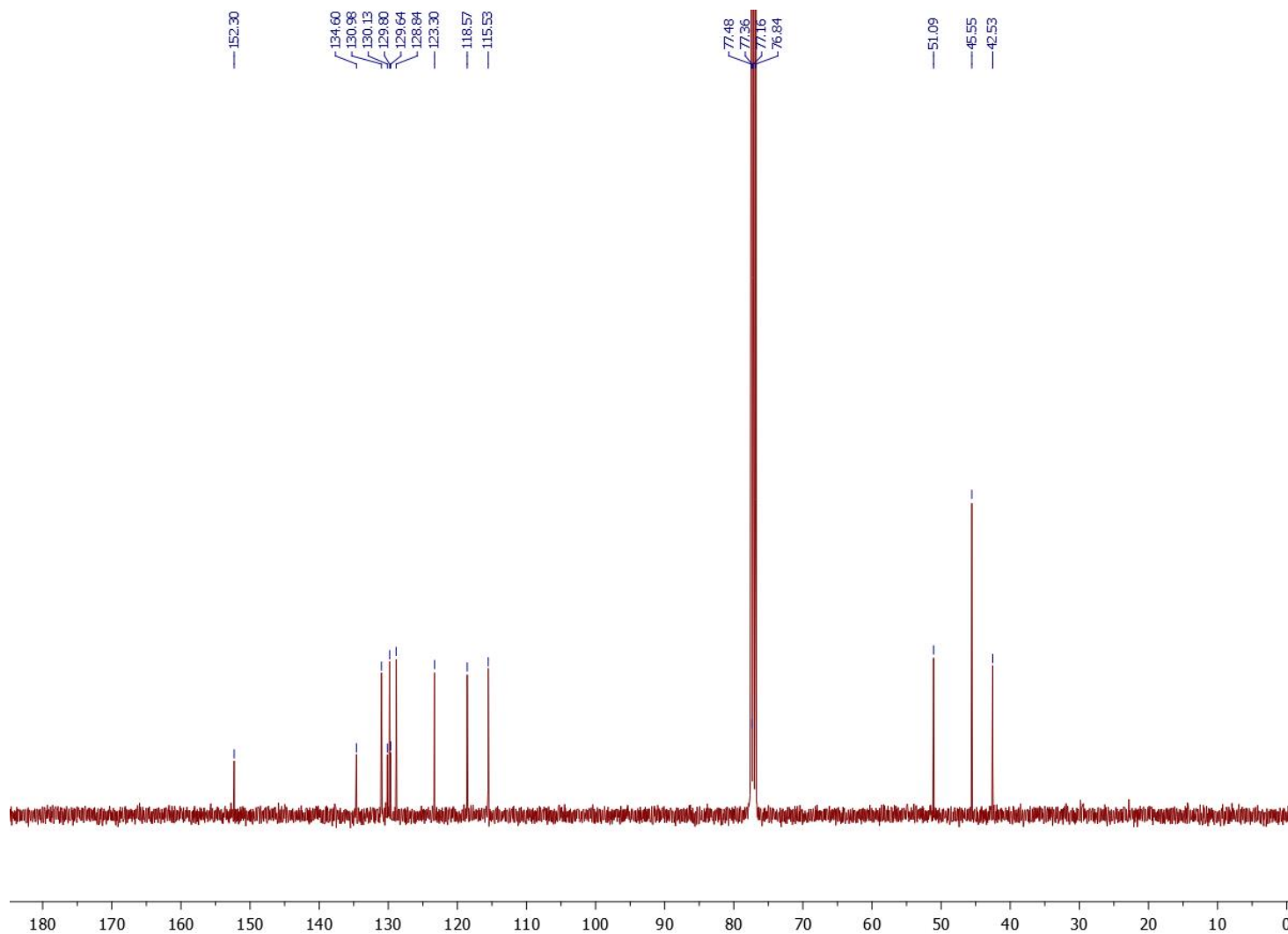


Figure A1.35. ^{13}C NMR of compd 16 in CDCl_3

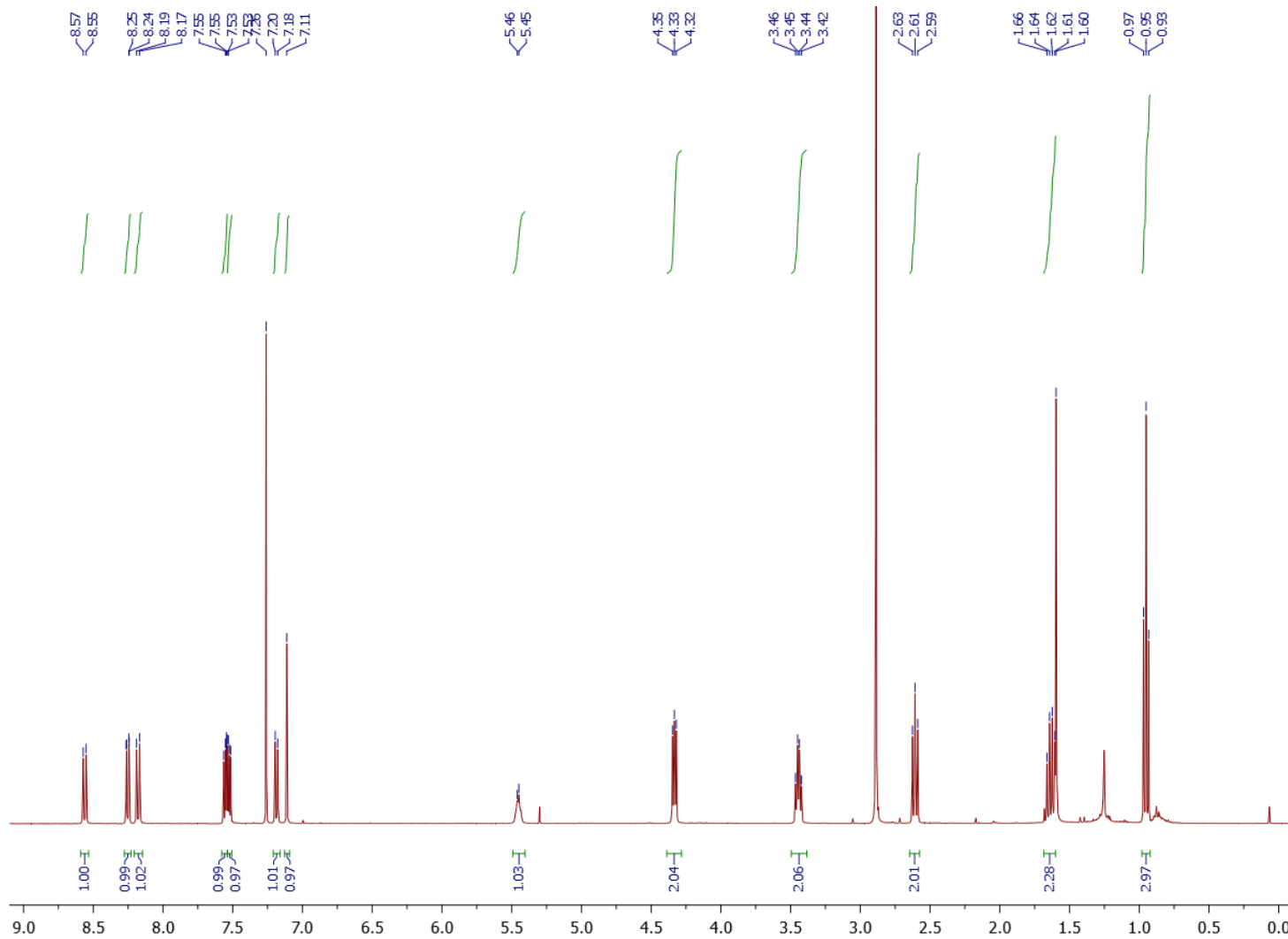


Figure A1.36. ^1H NMR of compd **17** in CDCl_3

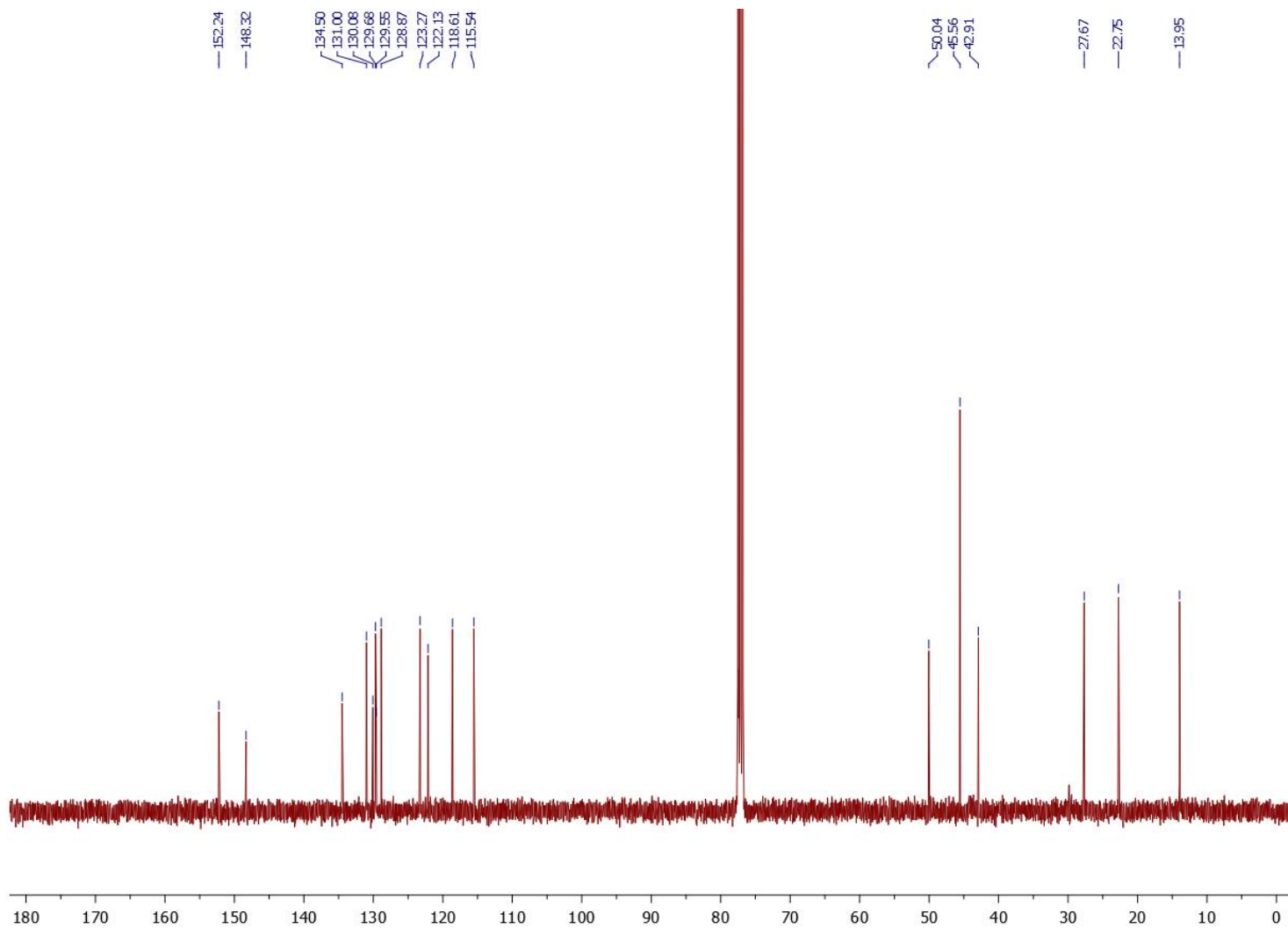


Figure A1.37. ^{13}C NMR of compd 17 in CDCl_3

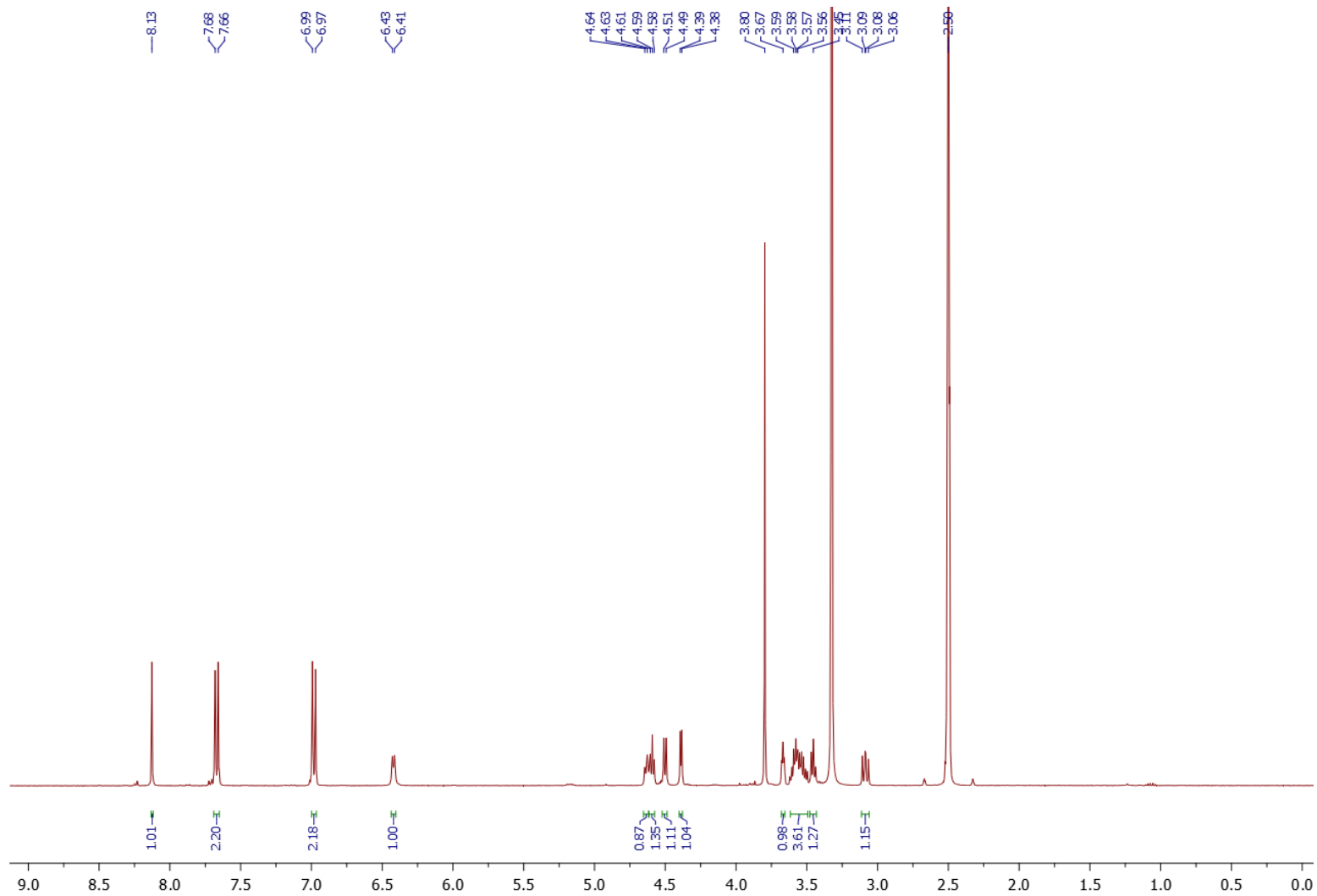


Figure A1.38. ¹H NMR of compd **18** in DMSO-d₆

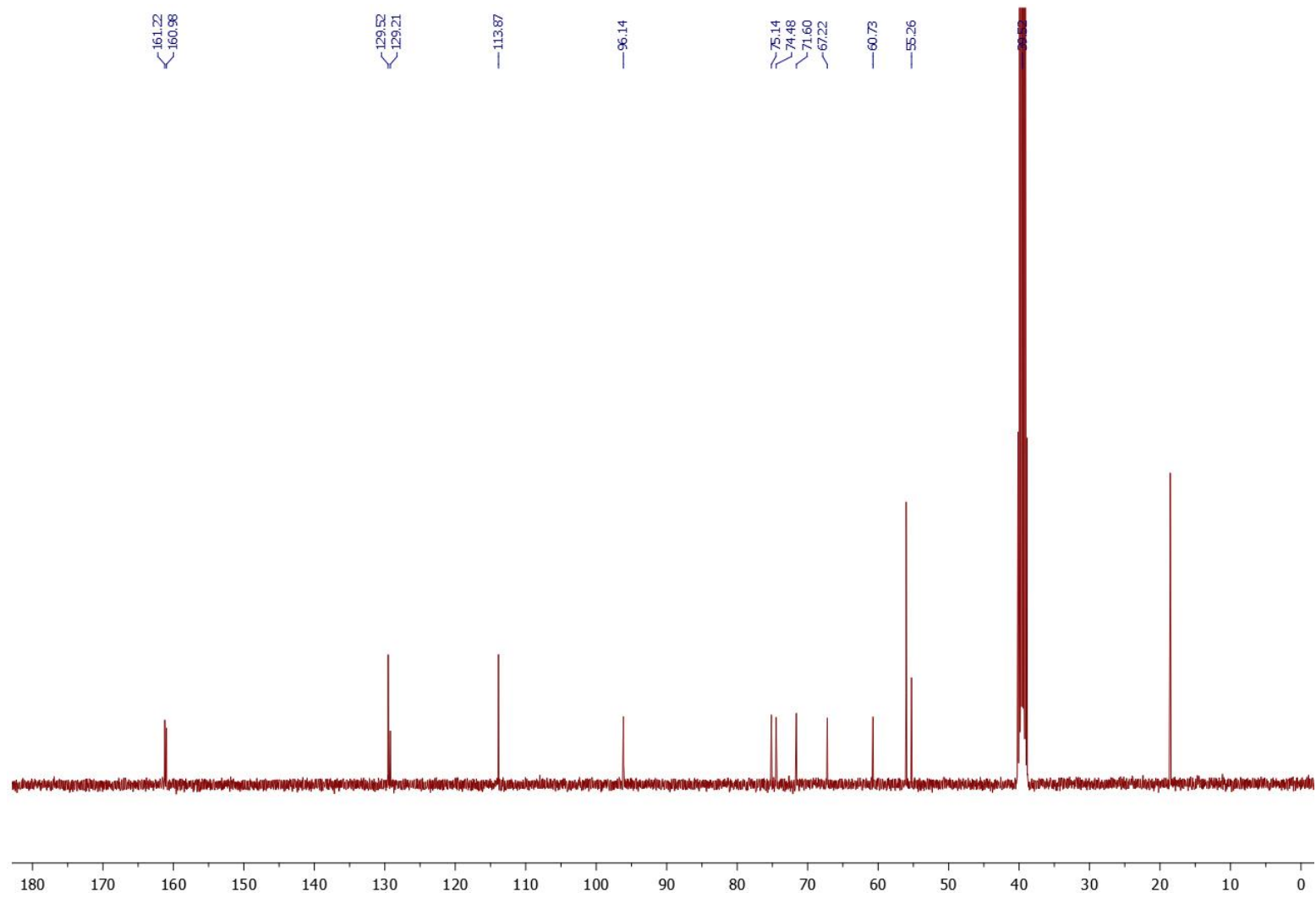


Figure 1A.39. ^{13}C NMR of **cmpd 18** in DMSO-d_6 (presence of ethanol)

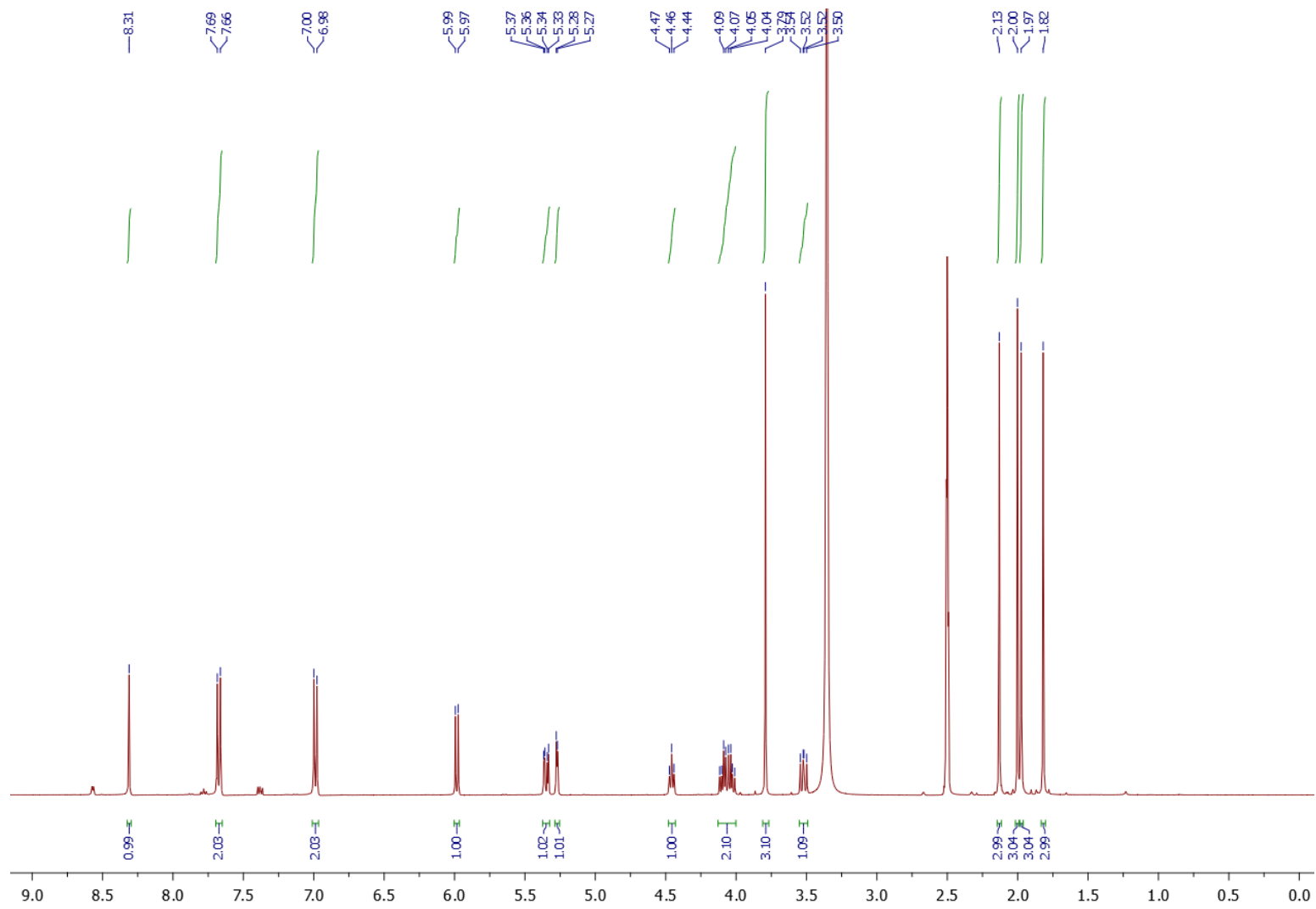


Figure A1.40. ^1H NMR of compd **19** in DMSO-d_6 (presence of traces of pyridine)

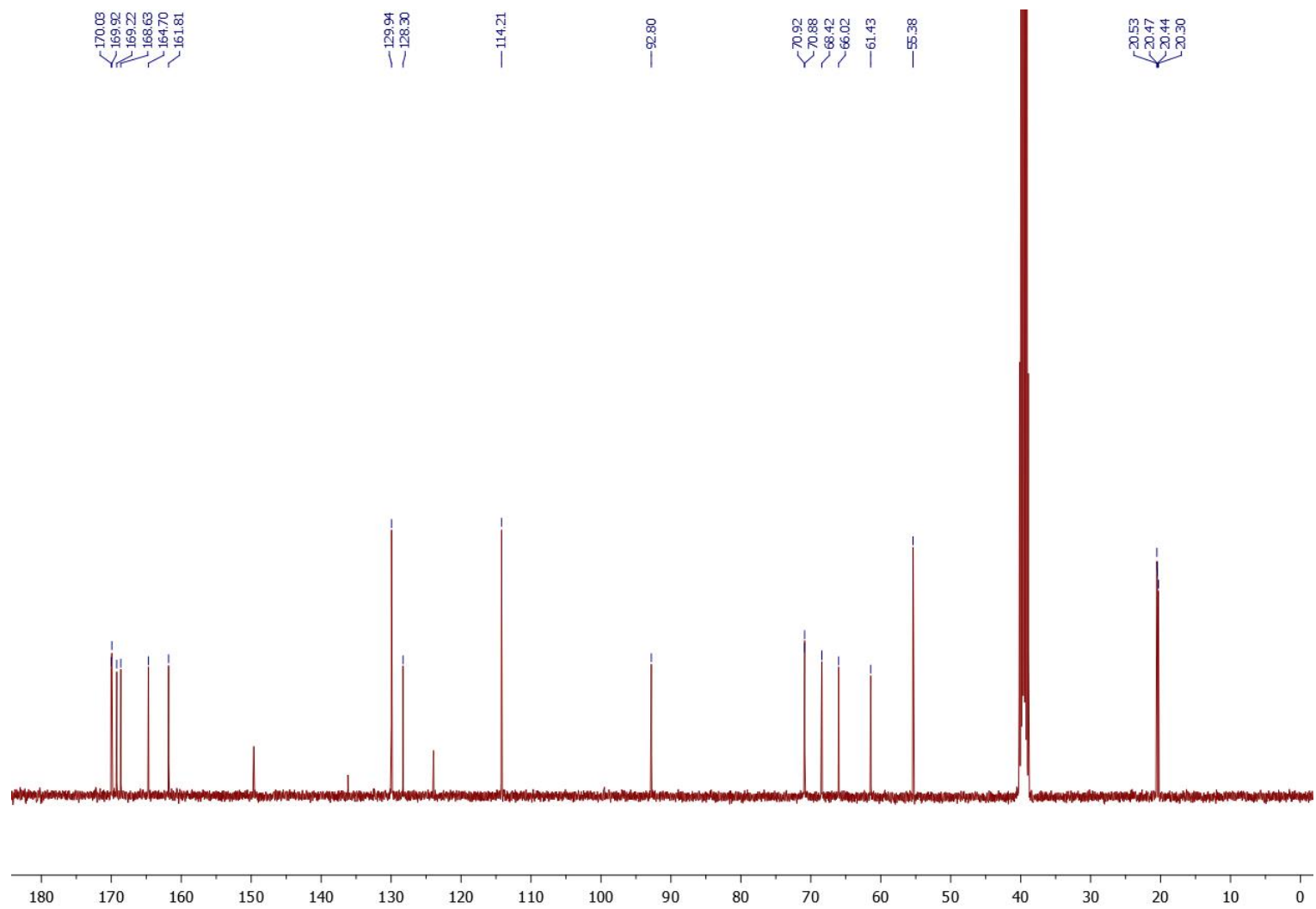


Figure A1.41. ^{13}C NMR of cmpd **19** in DMSO-d_6 (presence of traces of pyridine)

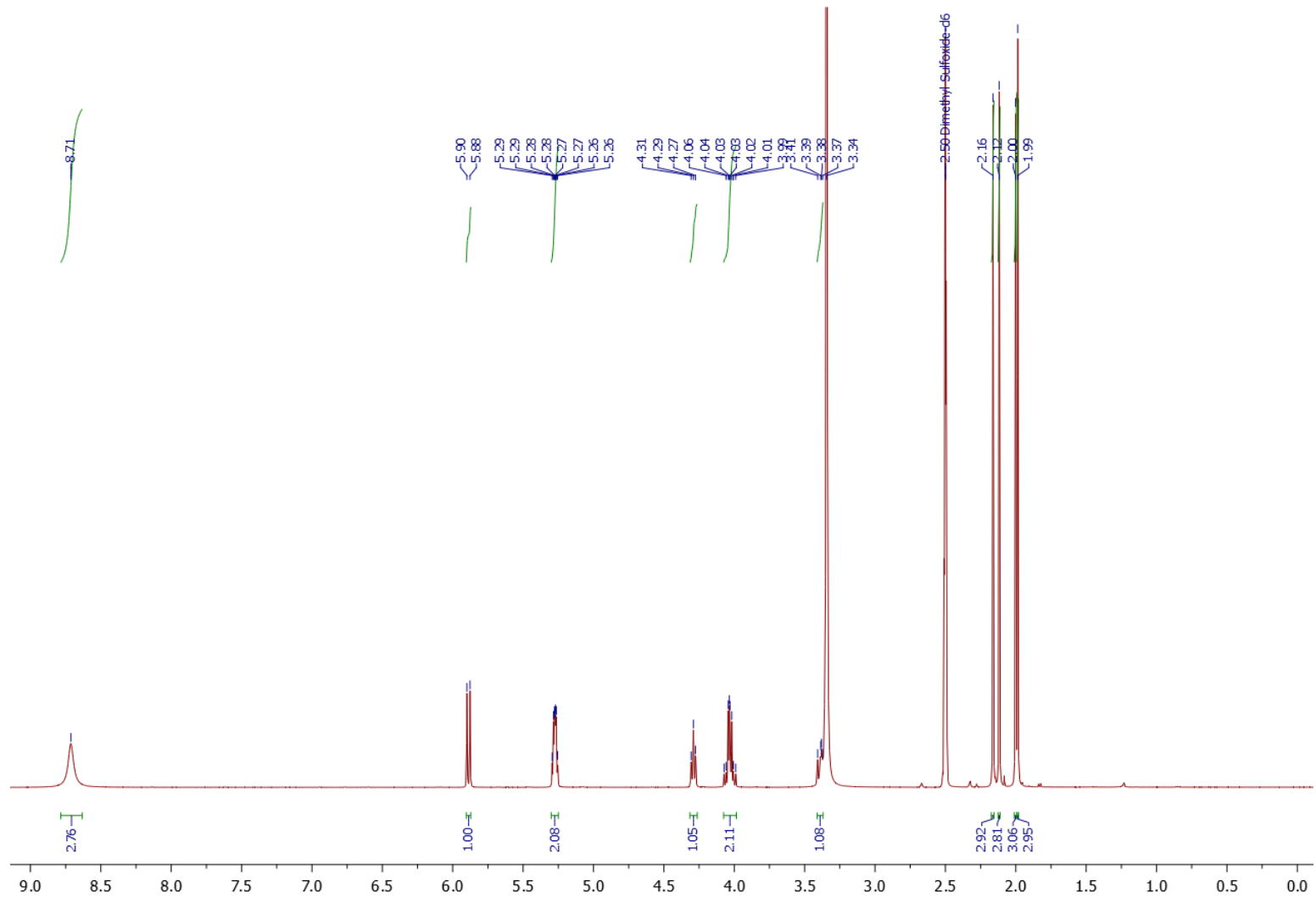


Figure A1.42. ^1H NMR of compd **20** in DMSO-d_6

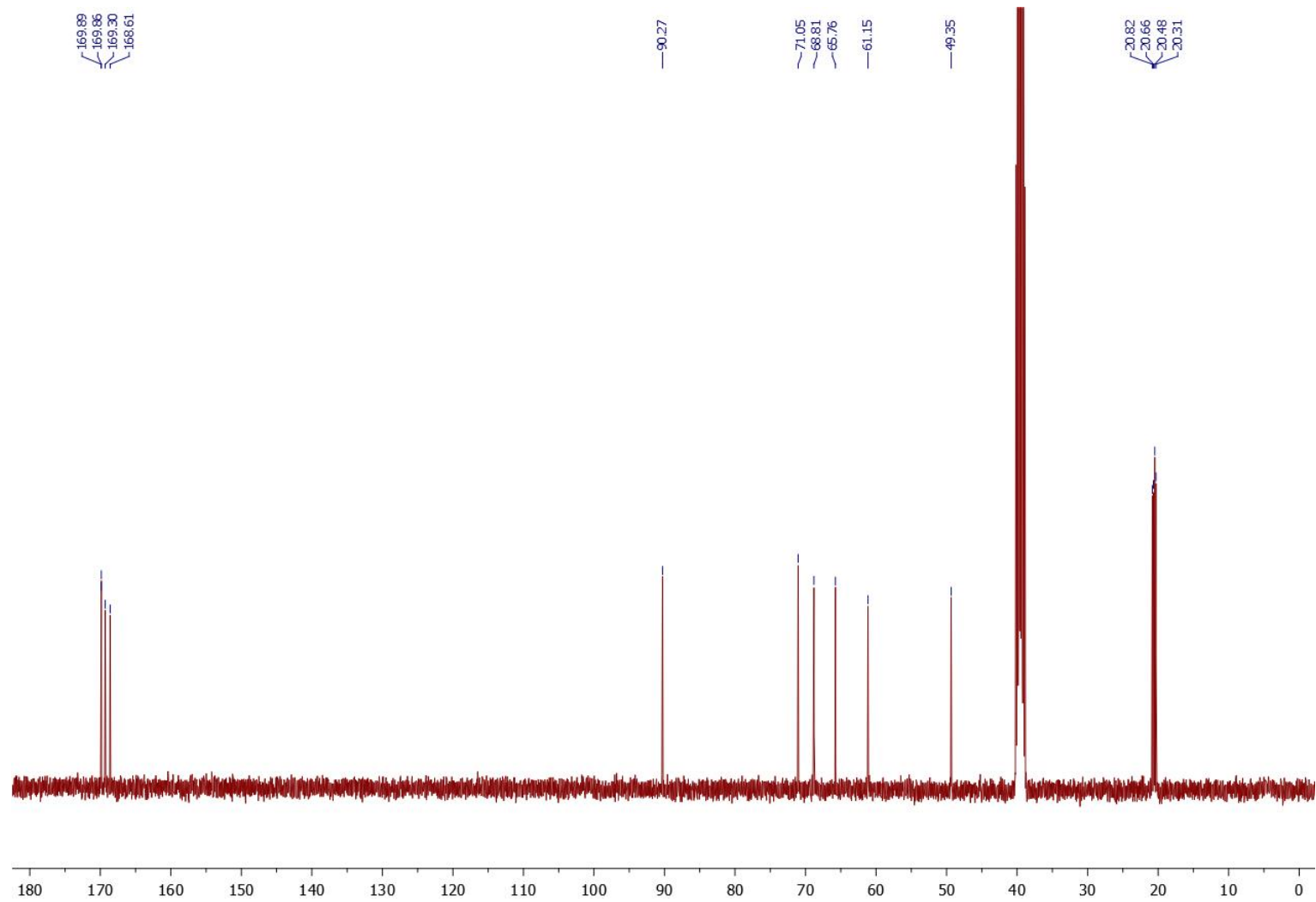


Figure A1.43. ^{13}C NMR of cmpd **20** in DMSO-d_6

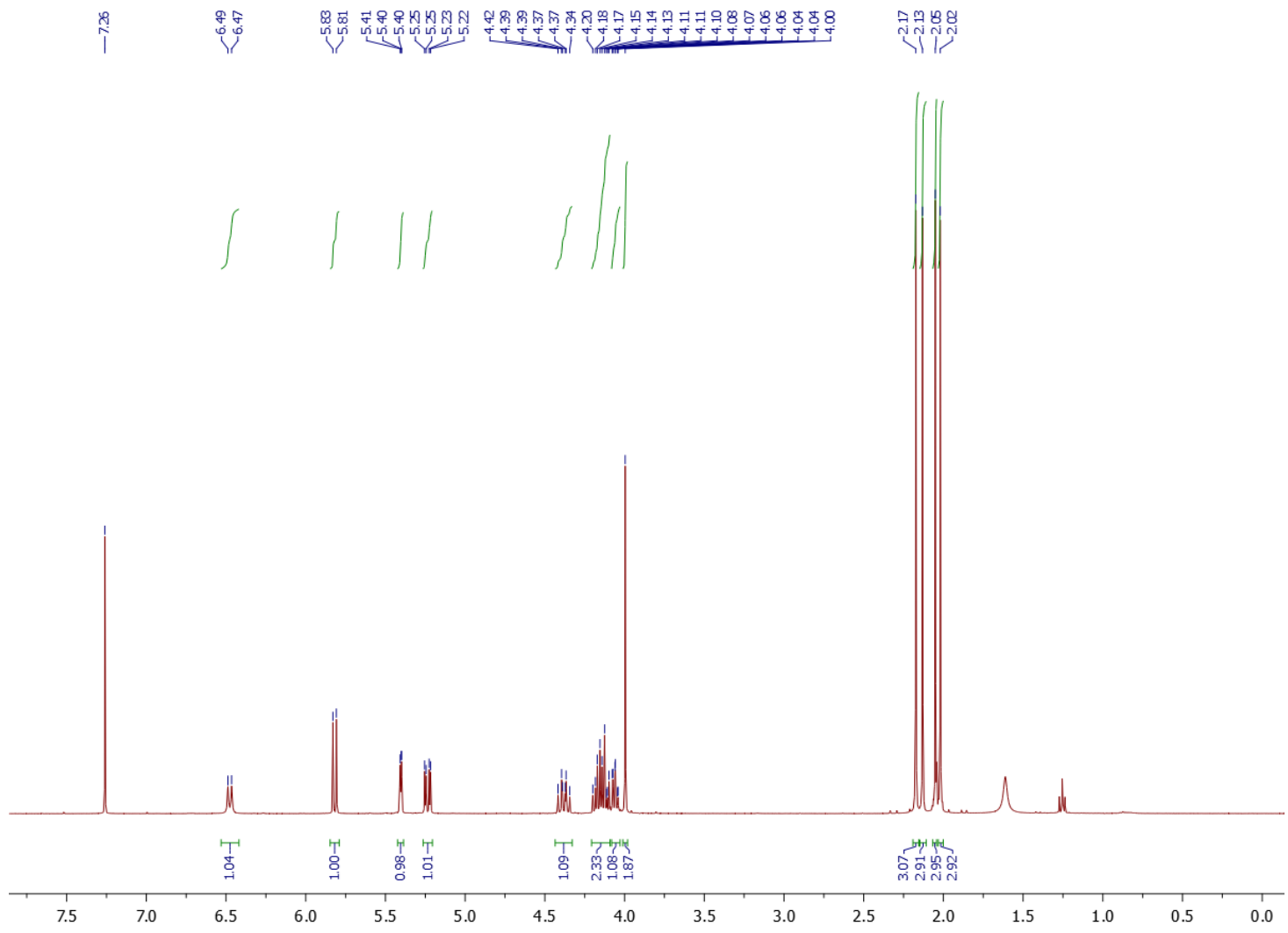


Figure A1.44. ¹H NMR of cmpd 21 in CDCl₃

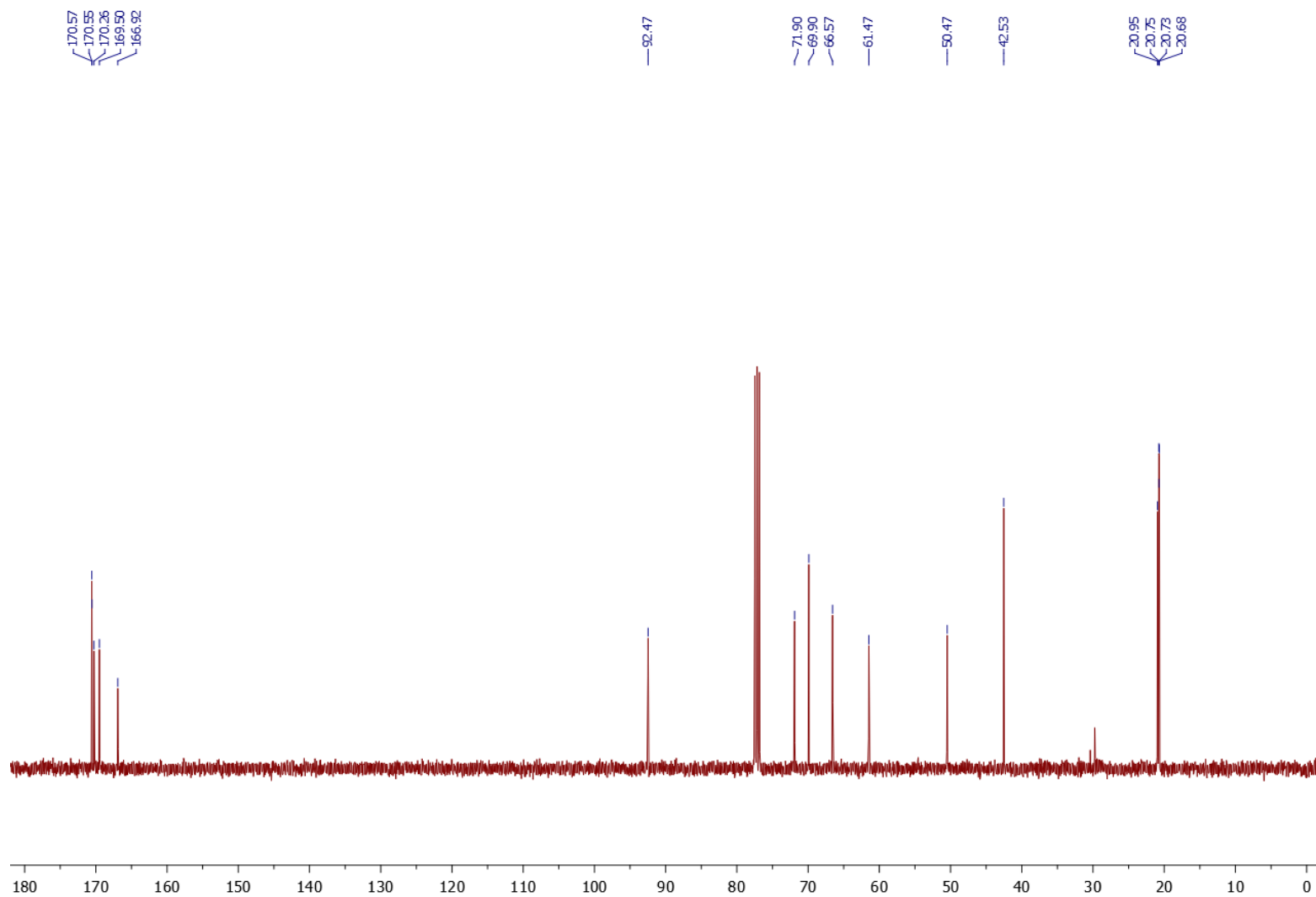


Figure A1.45. ^{13}C NMR of compd **21** in CDCl_3

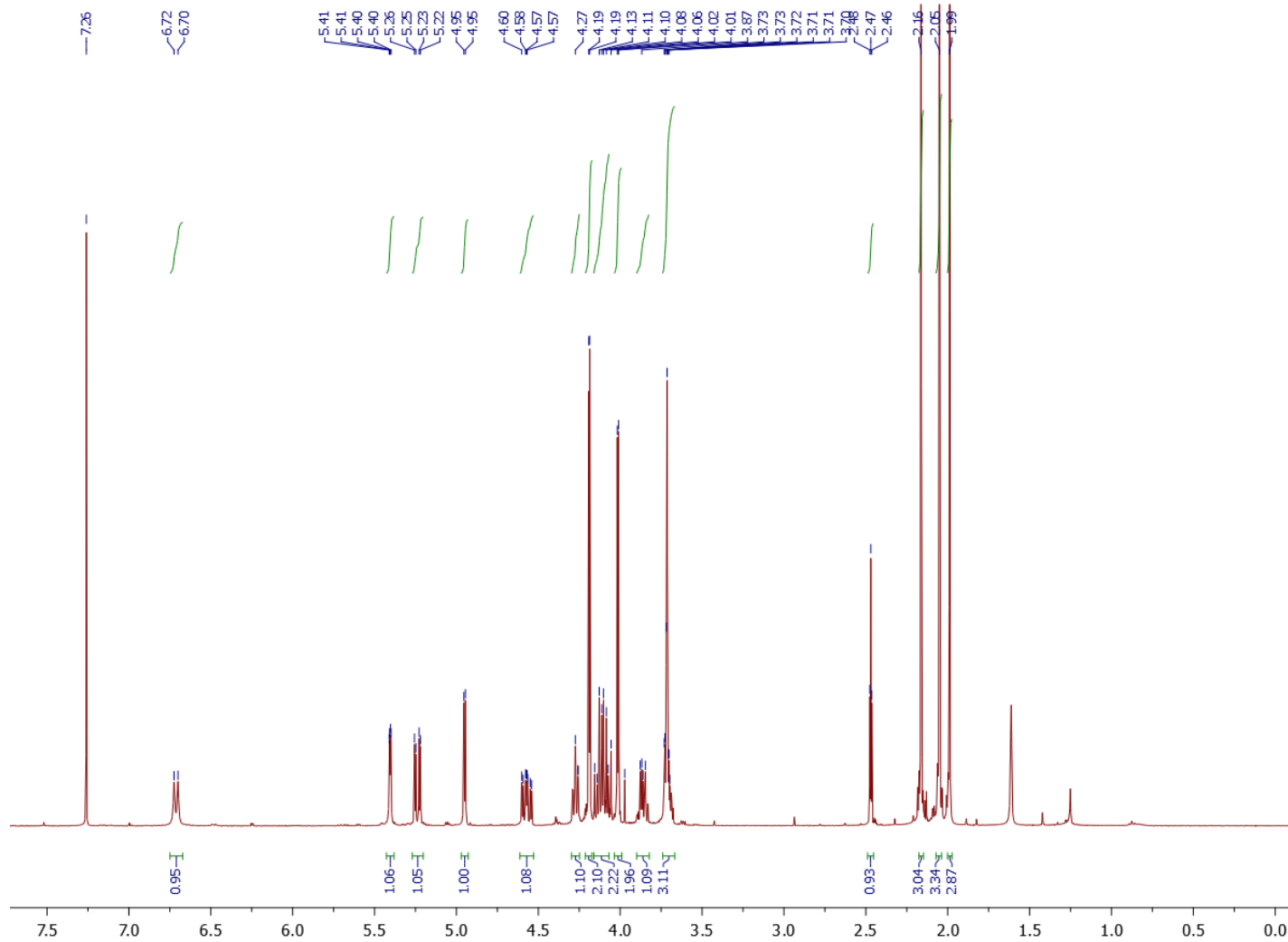


Figure A1.46. ^1H NMR of **cmpd 22- α** in CDCl_3

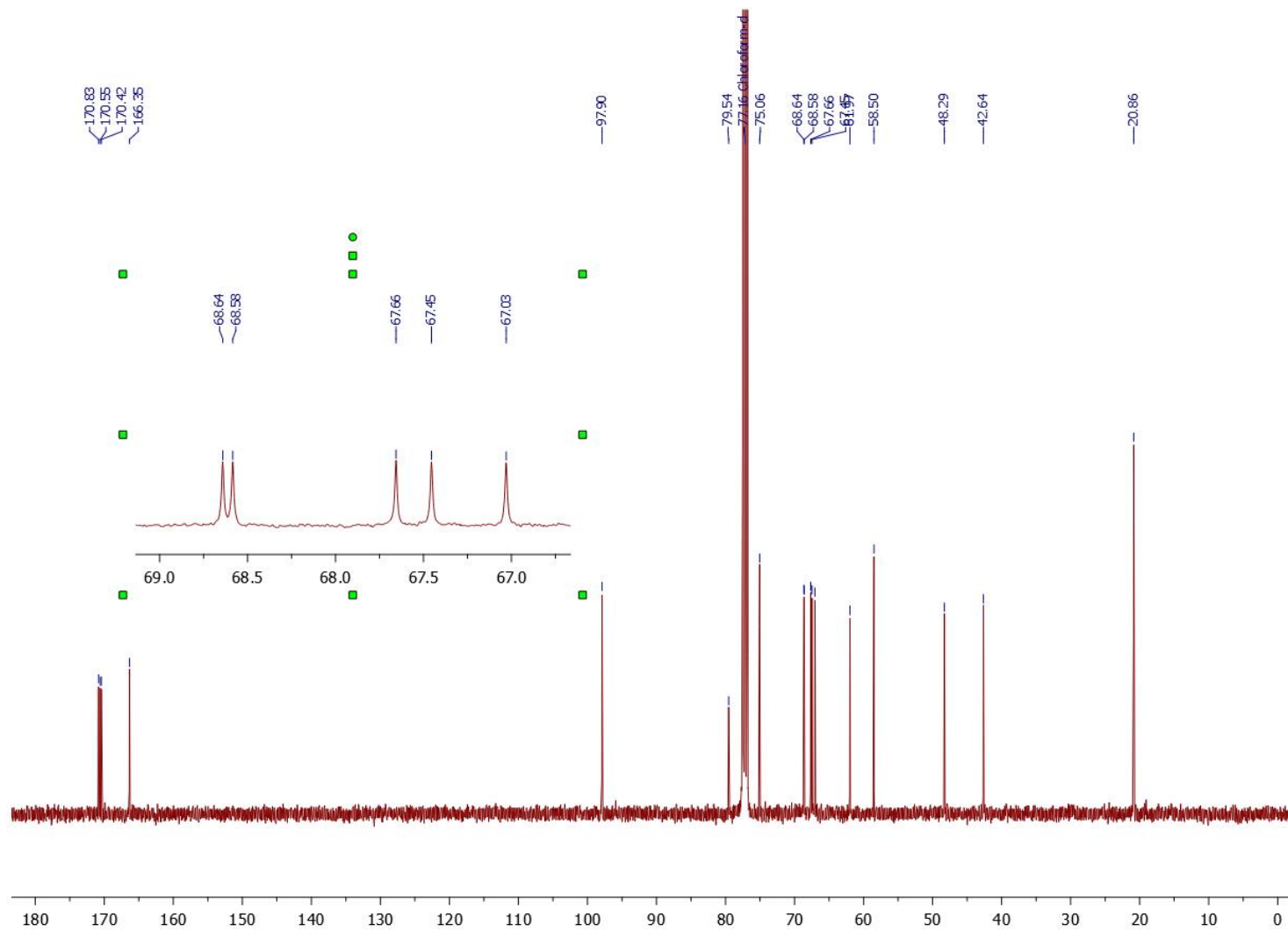


Figure A1.47. ^{13}C NMR of compd 22- α in CDCl_3

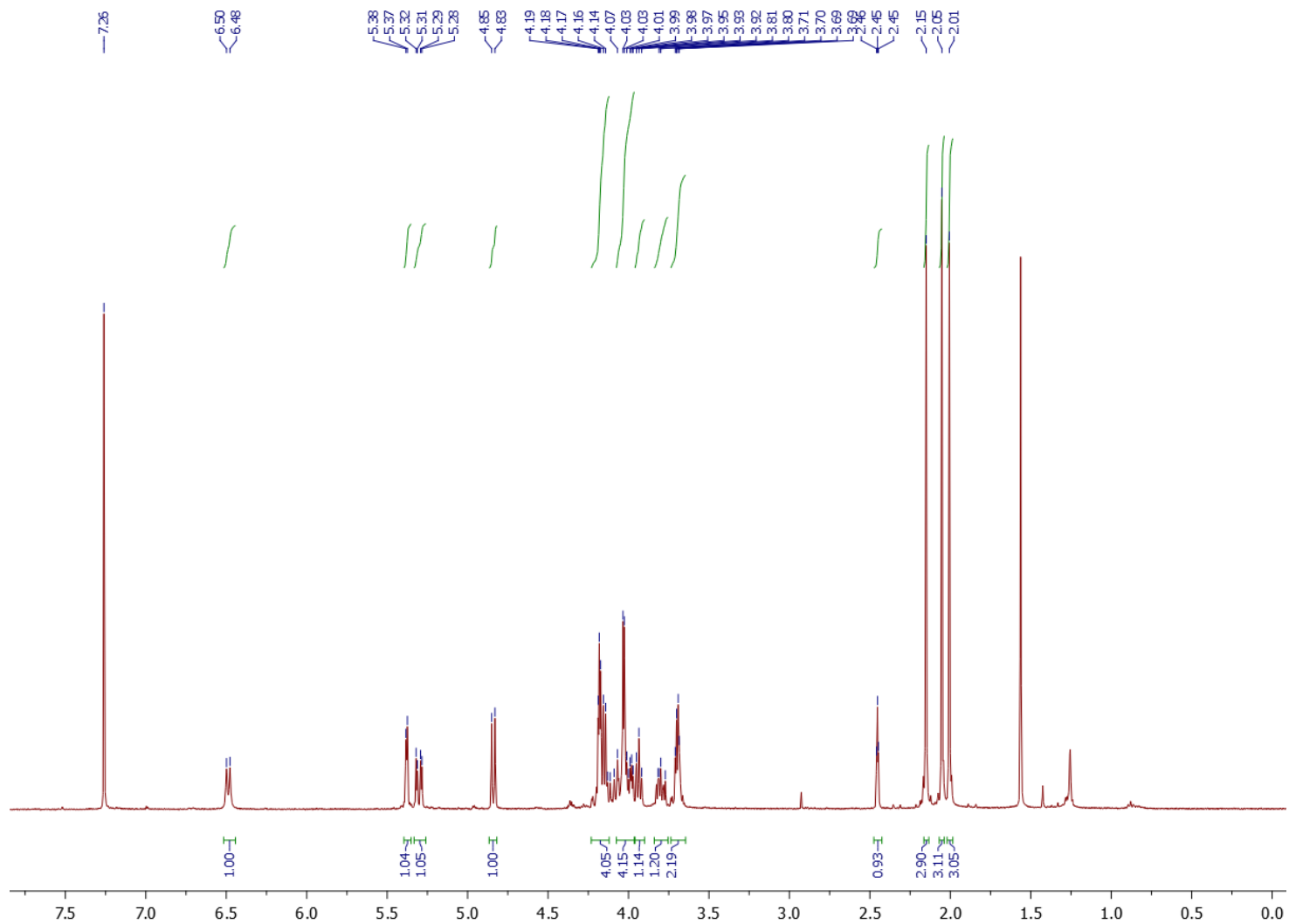


Figure A1.48. ^1H NMR of cmpd 22- β in CDCl_3

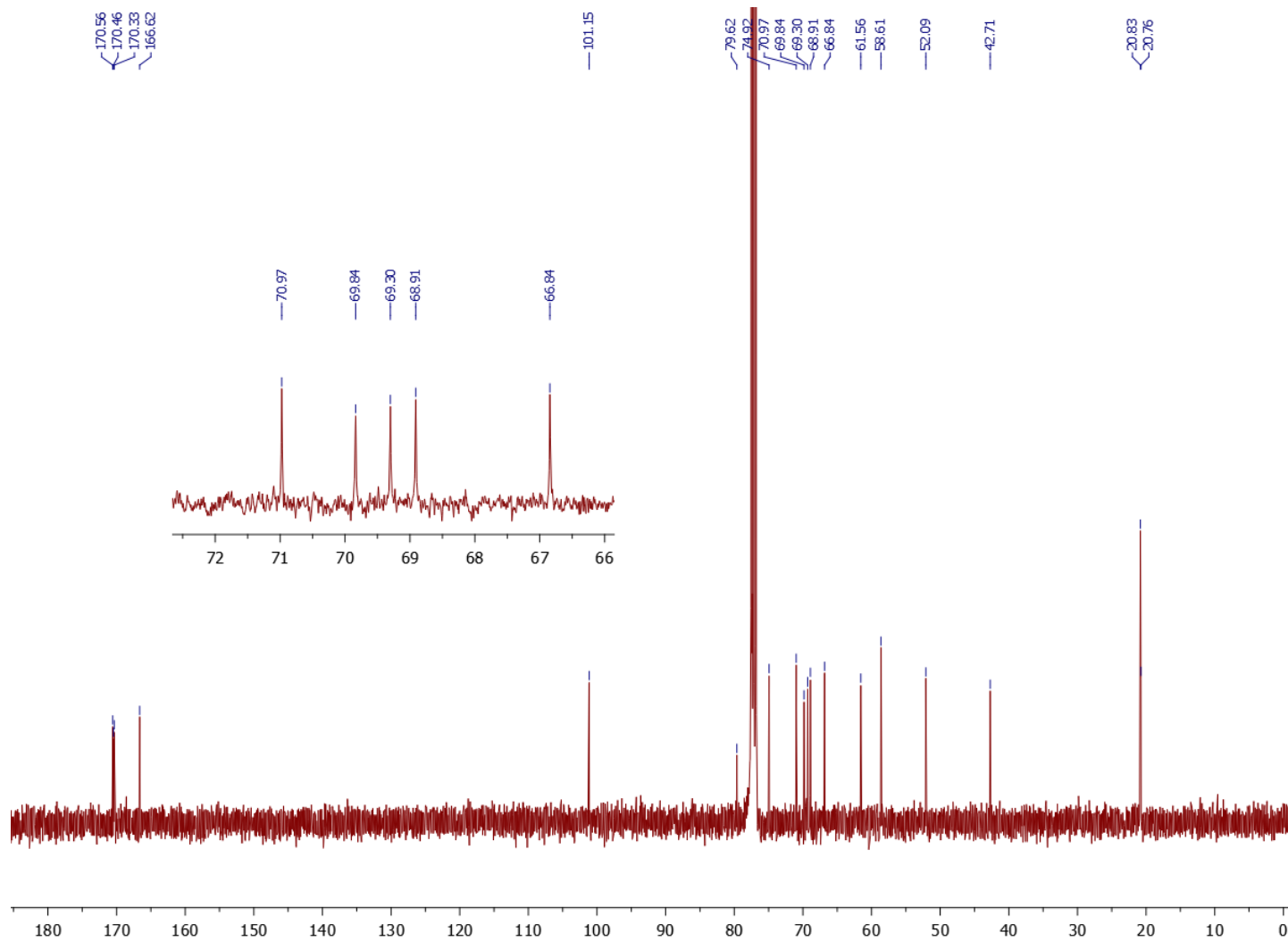


Figure A1.49. ^{13}C NMR of compd 22- β in CDCl_3

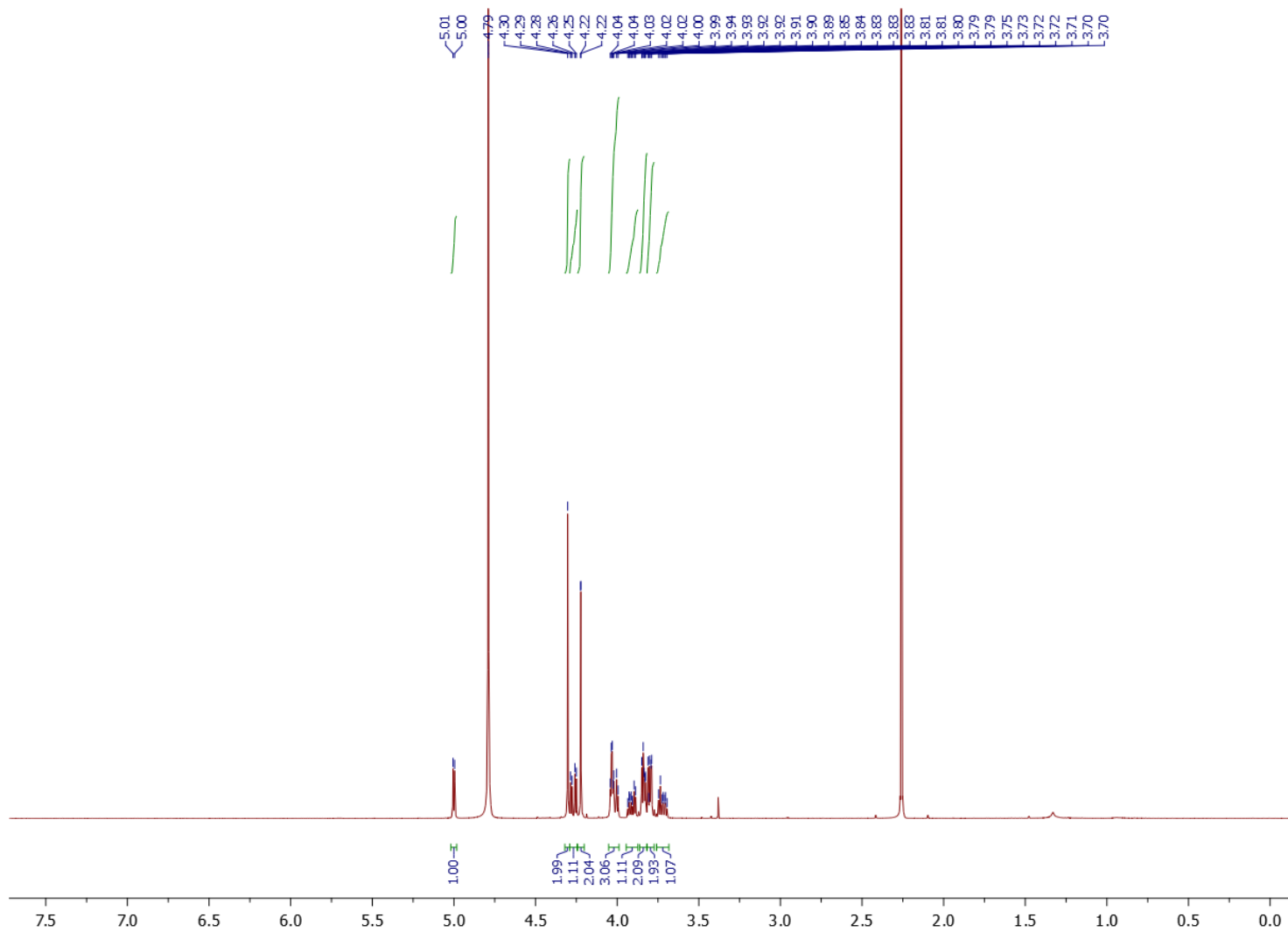


Figure A1.50. ^1H NMR of compd **23- α** in D_2O (presence of acetone, removed after lyophilisation)

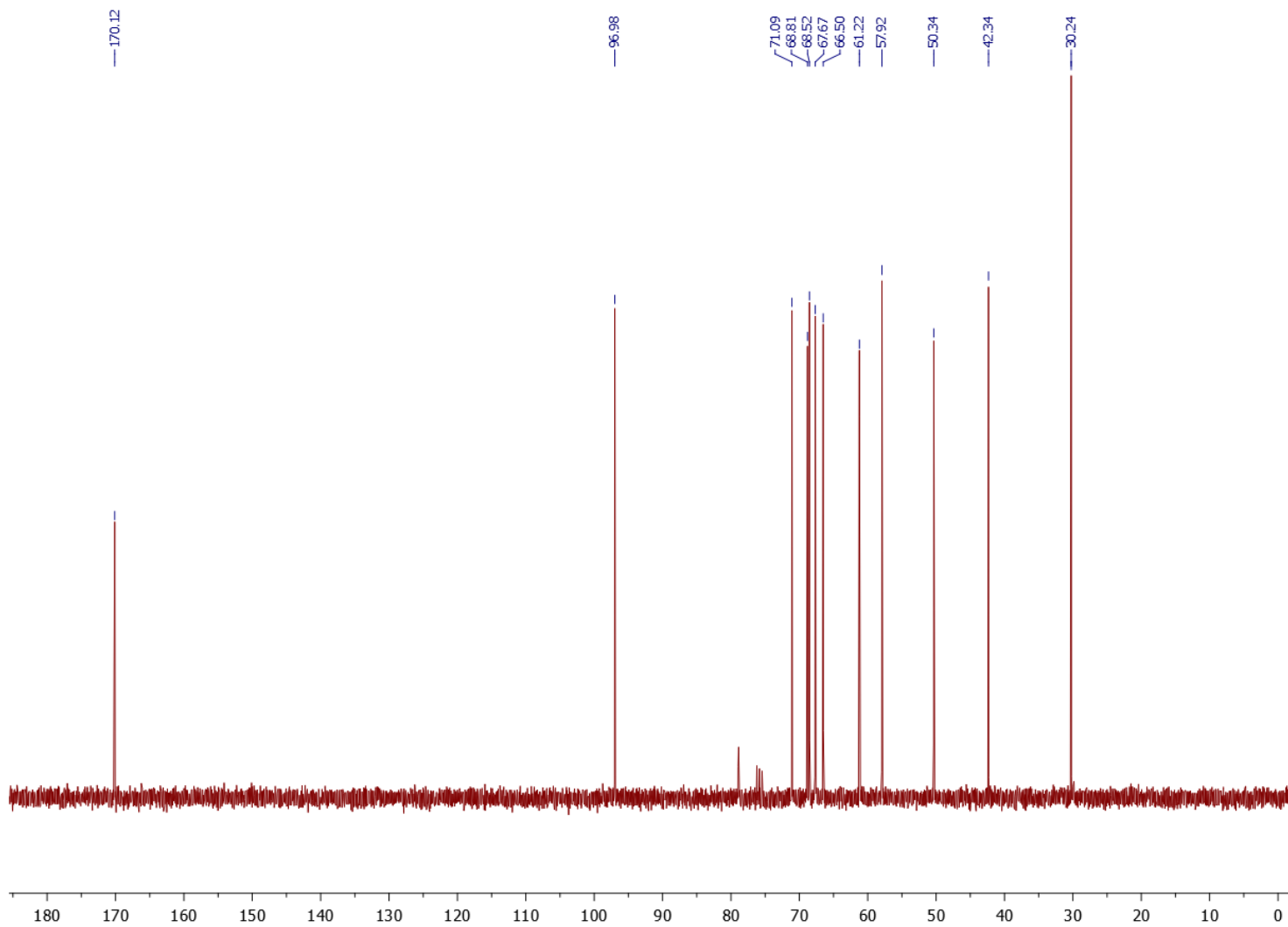


Figure A1.51. ^{13}C NMR of compd **23- α** in D_2O (presence of acetone, removed after lyophilisation)

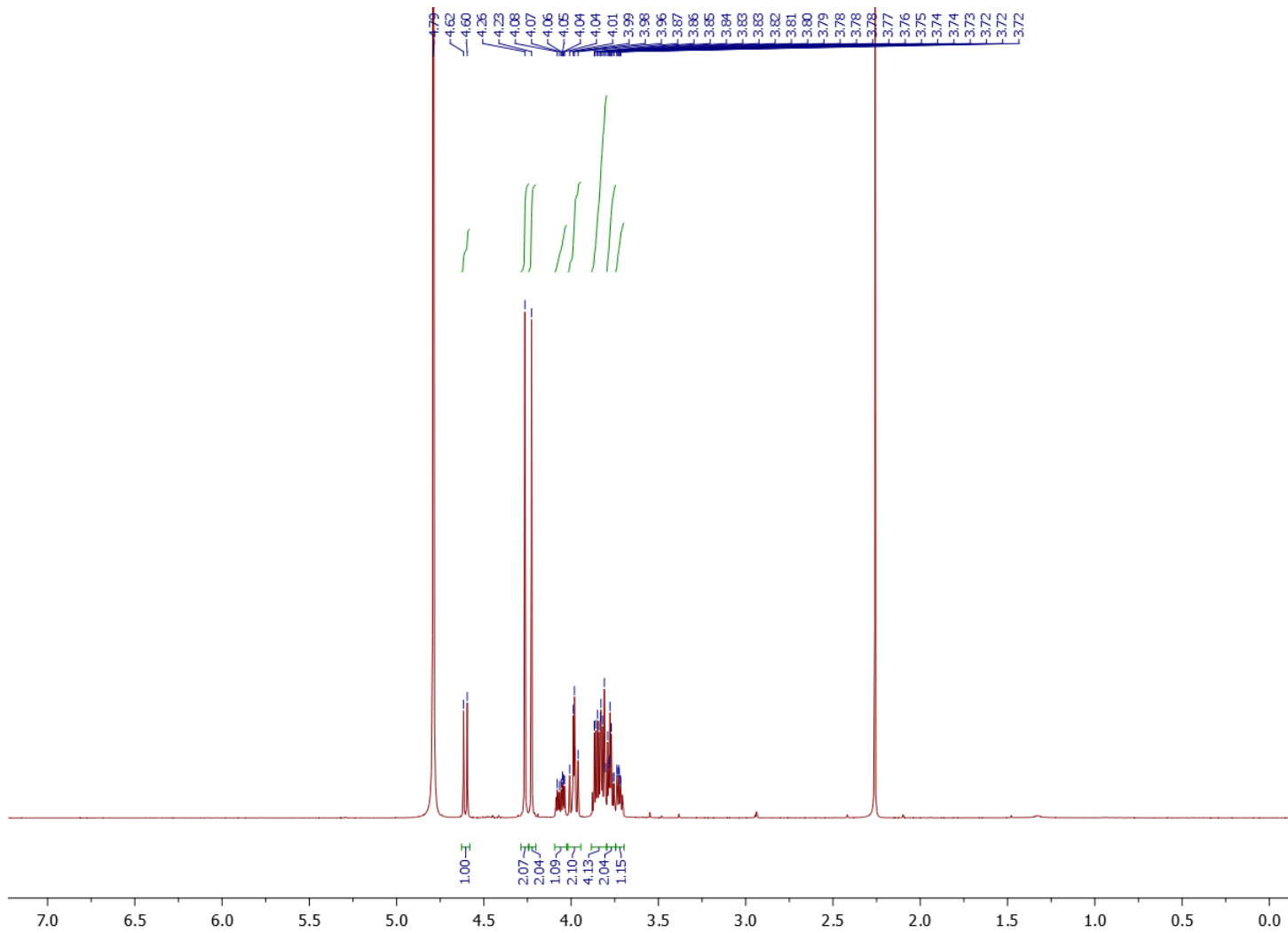


Figure A1.52. ^1H NMR of **cmpd 23- β** in D_2O (presence of acetone, removed after lyophilisation)

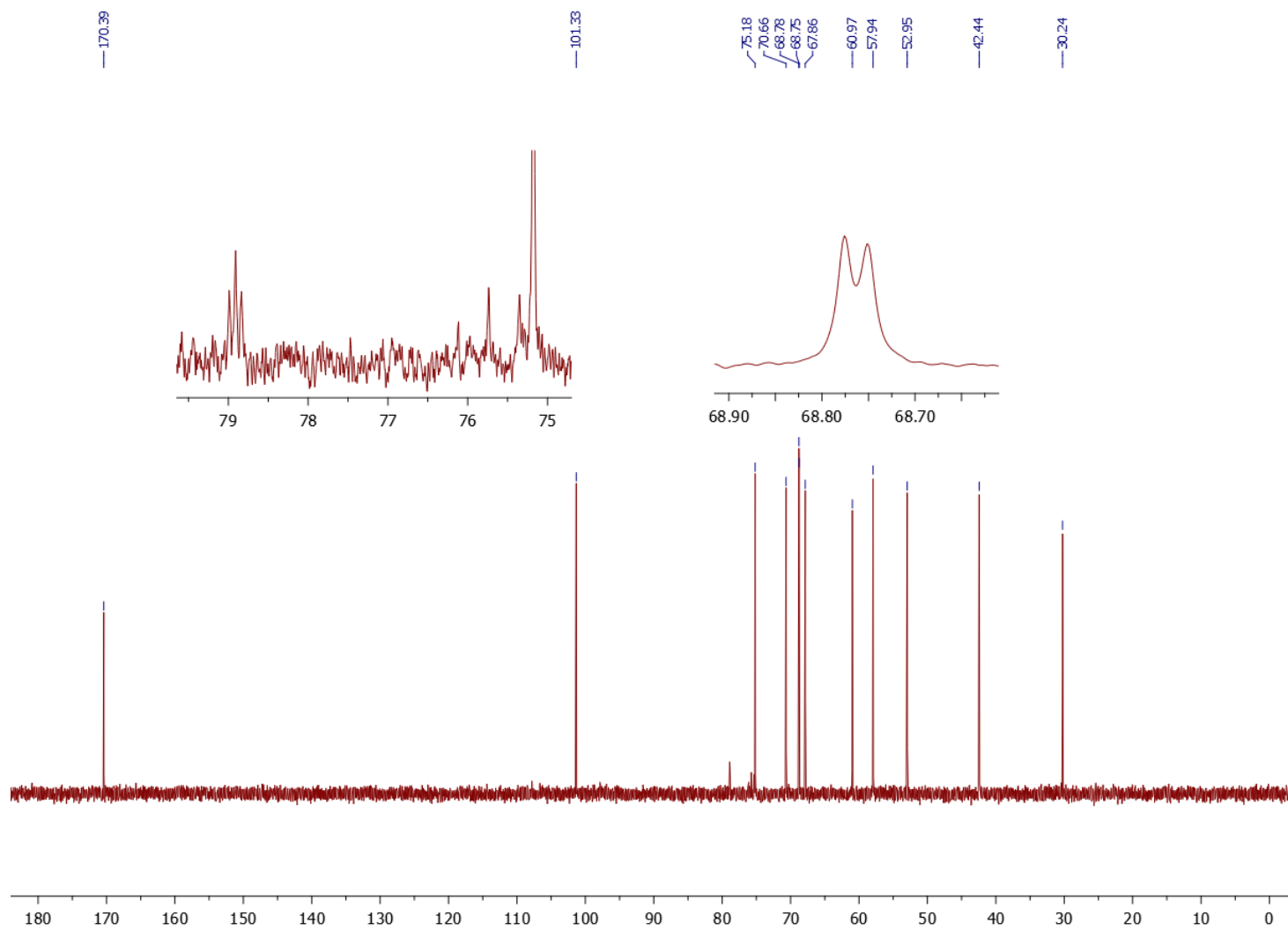


Figure A1.53. ^{13}C NMR of **cmpd 23- β** in D_2O (presence of acetone, removed after lyophilisation)

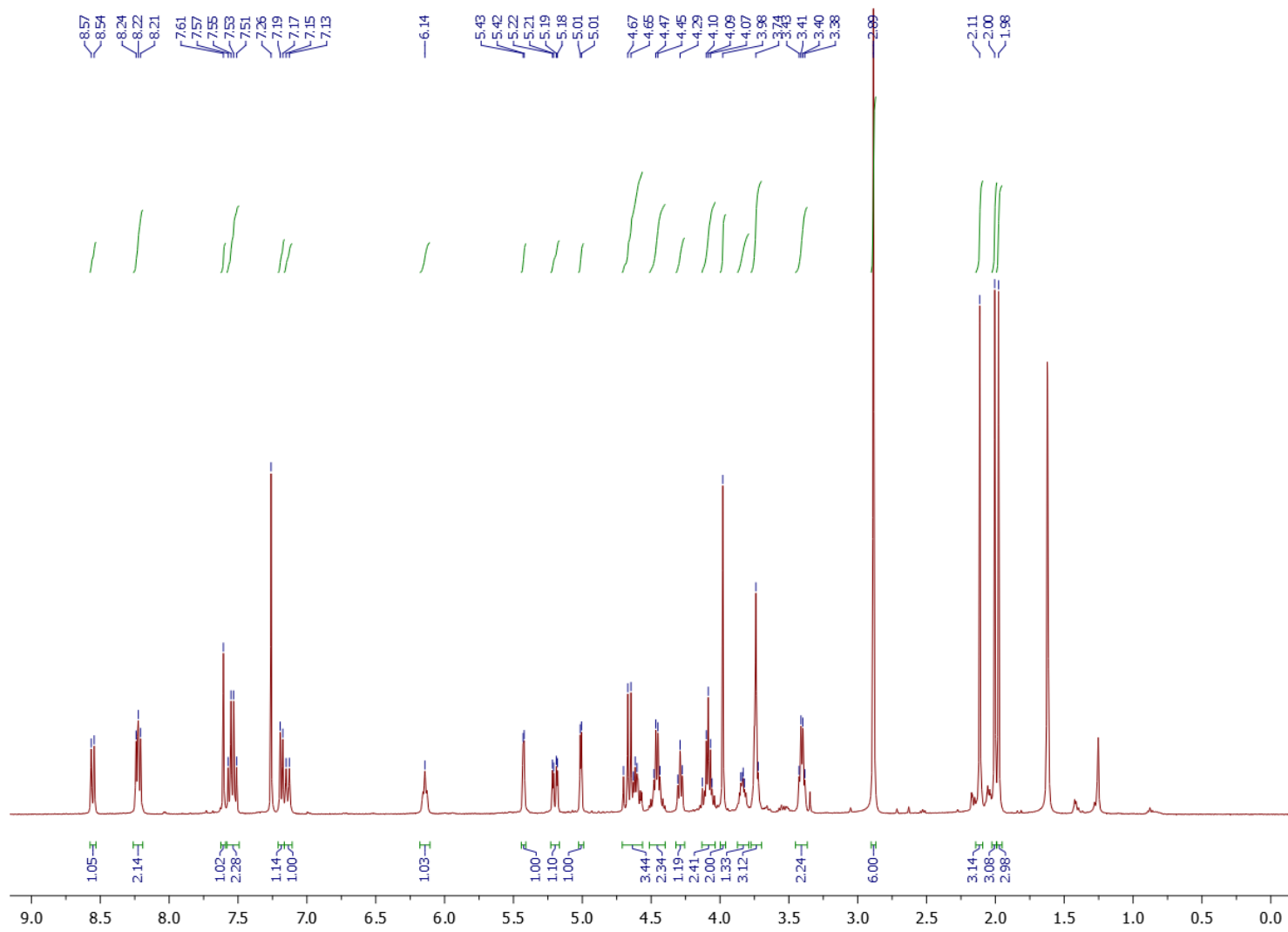


Figure A1.54. ^1H NMR of compd 24- α in CDCl_3

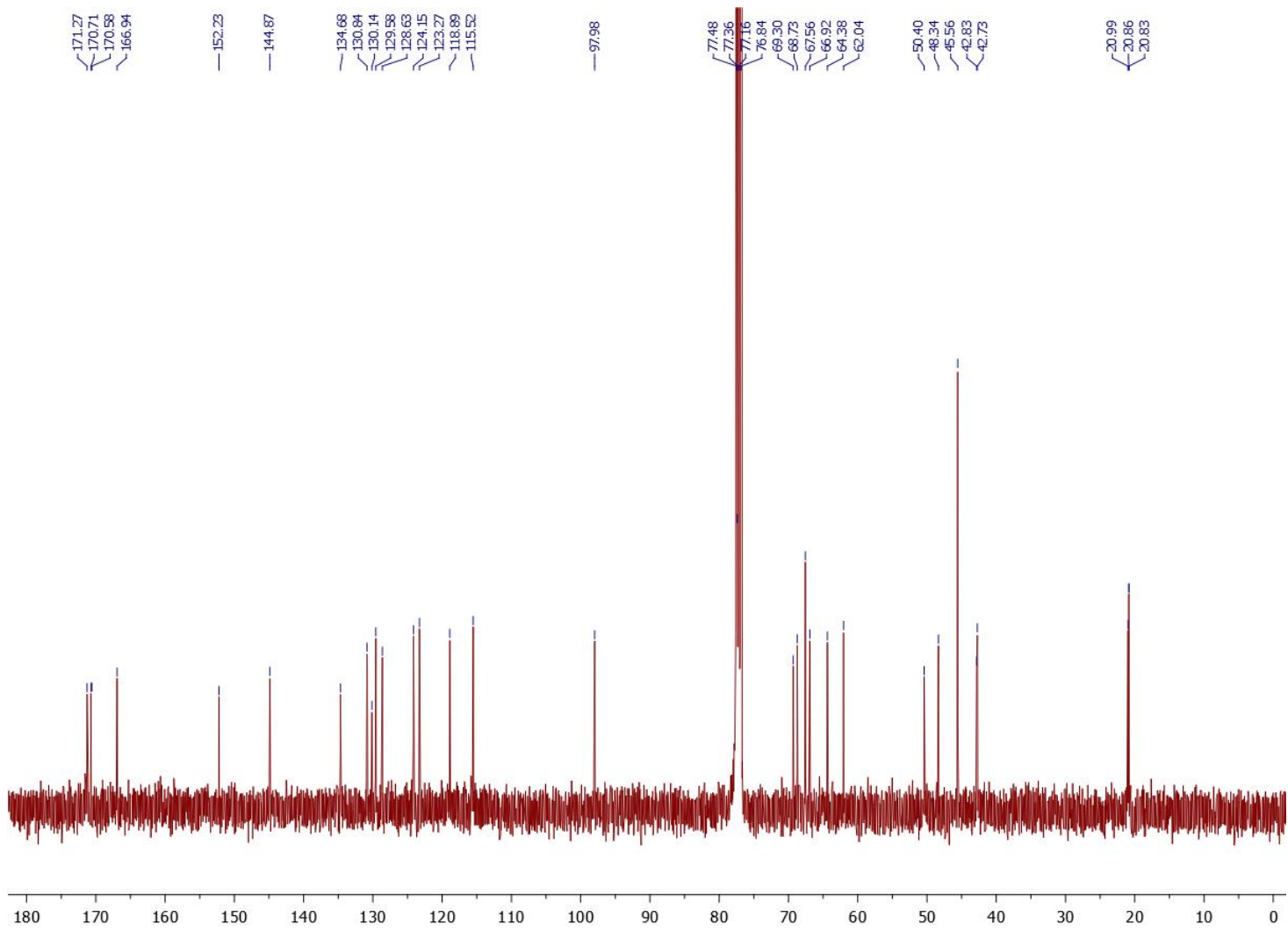


Figure A1.55. ¹³C NMR of compd 24- α in CDCl₃

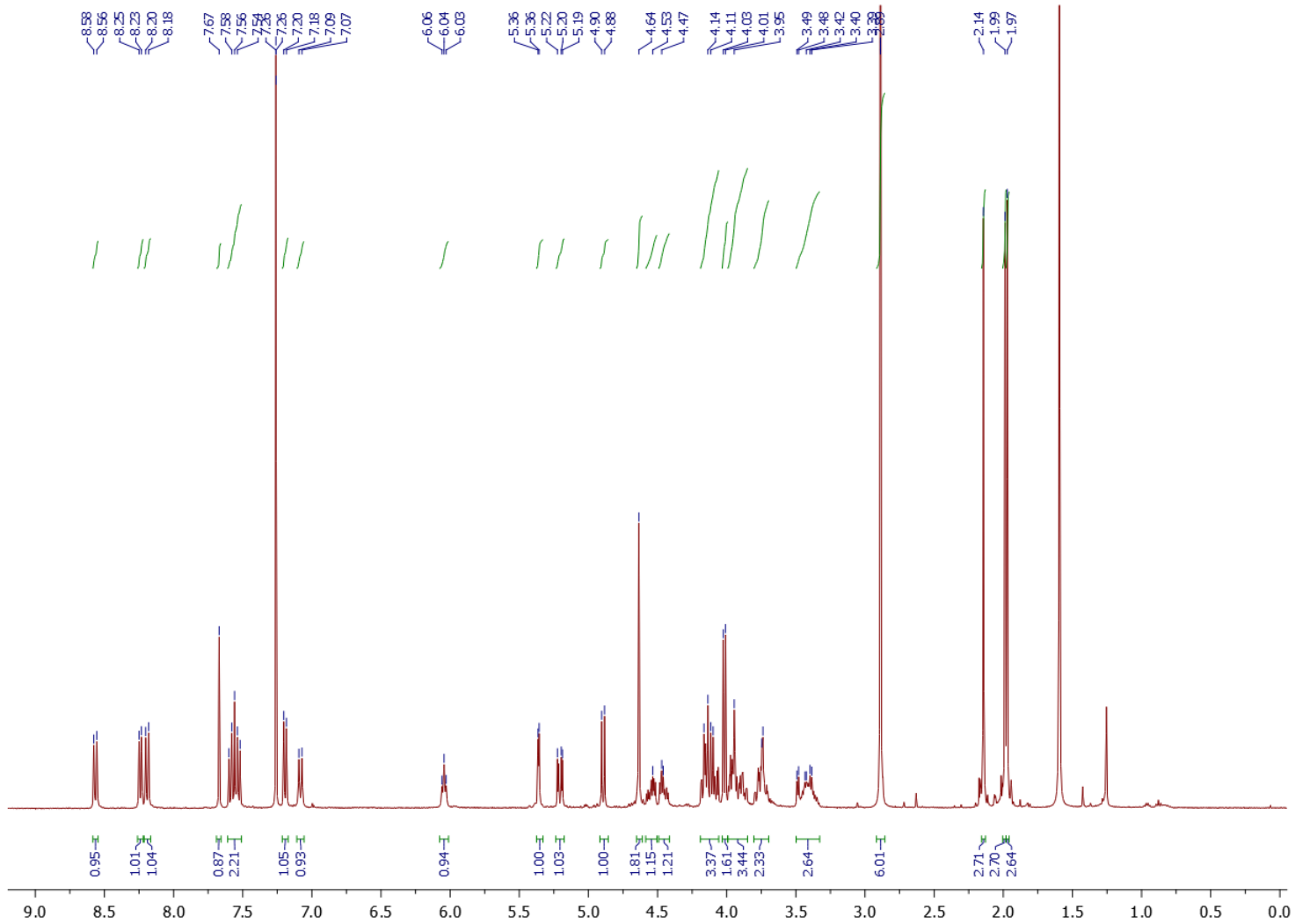


Figure A1.56. ¹H NMR of compd 24-β in CDCl₃

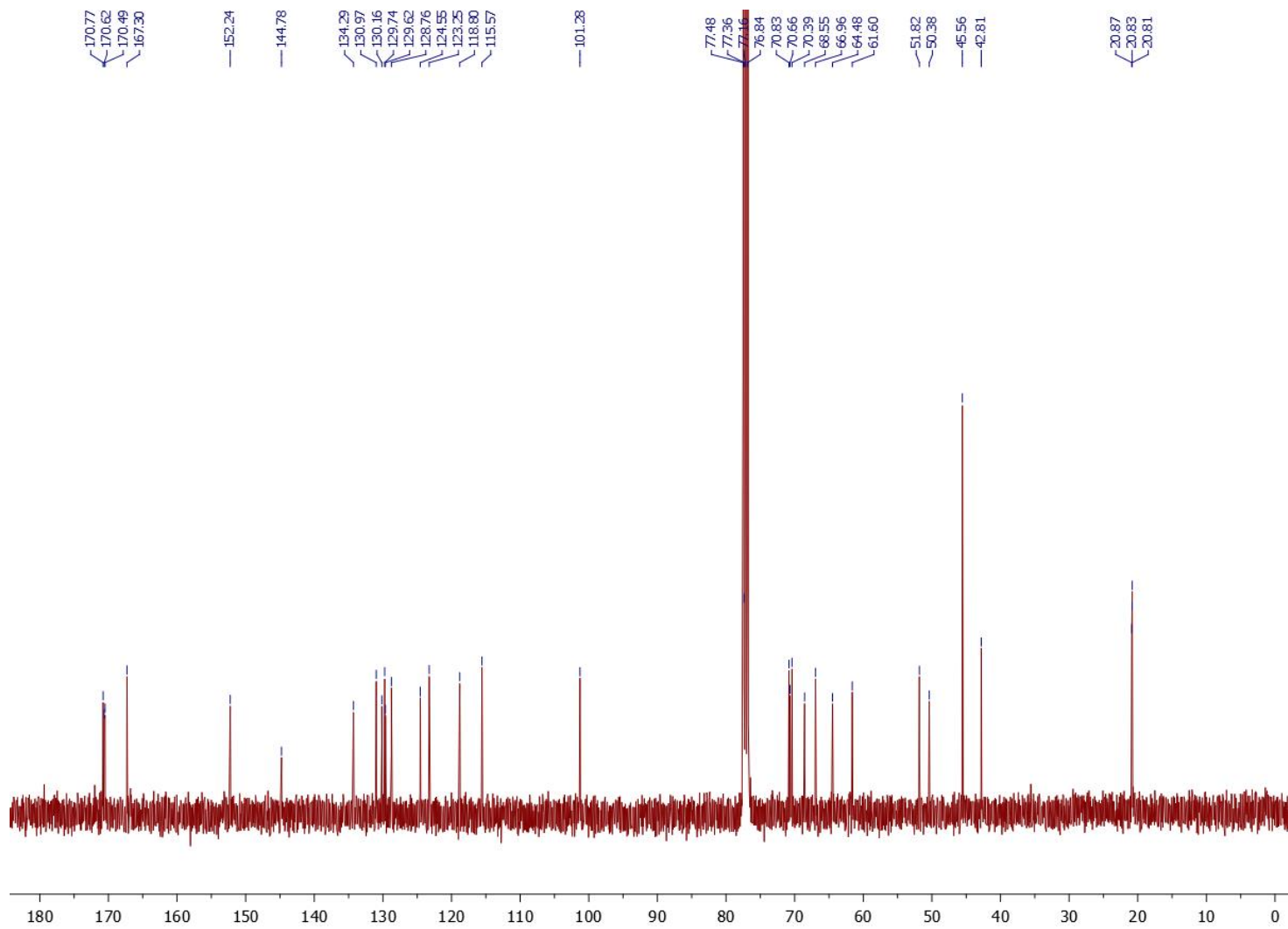


Figure A1.57. ^{13}C NMR of compd **24- β** in CDCl_3

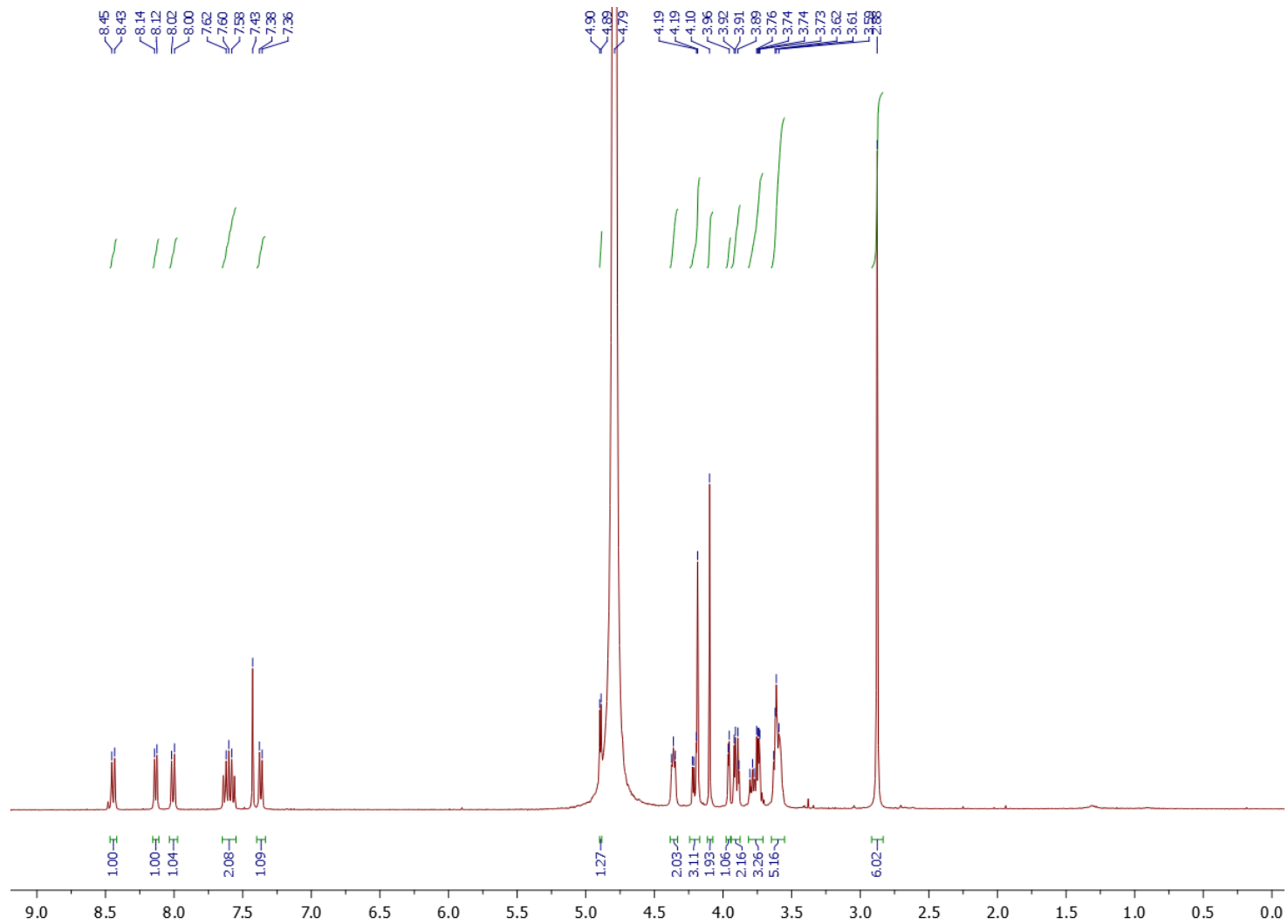


Figure A1.58. ¹H NMR of compd 25- α in D₂O

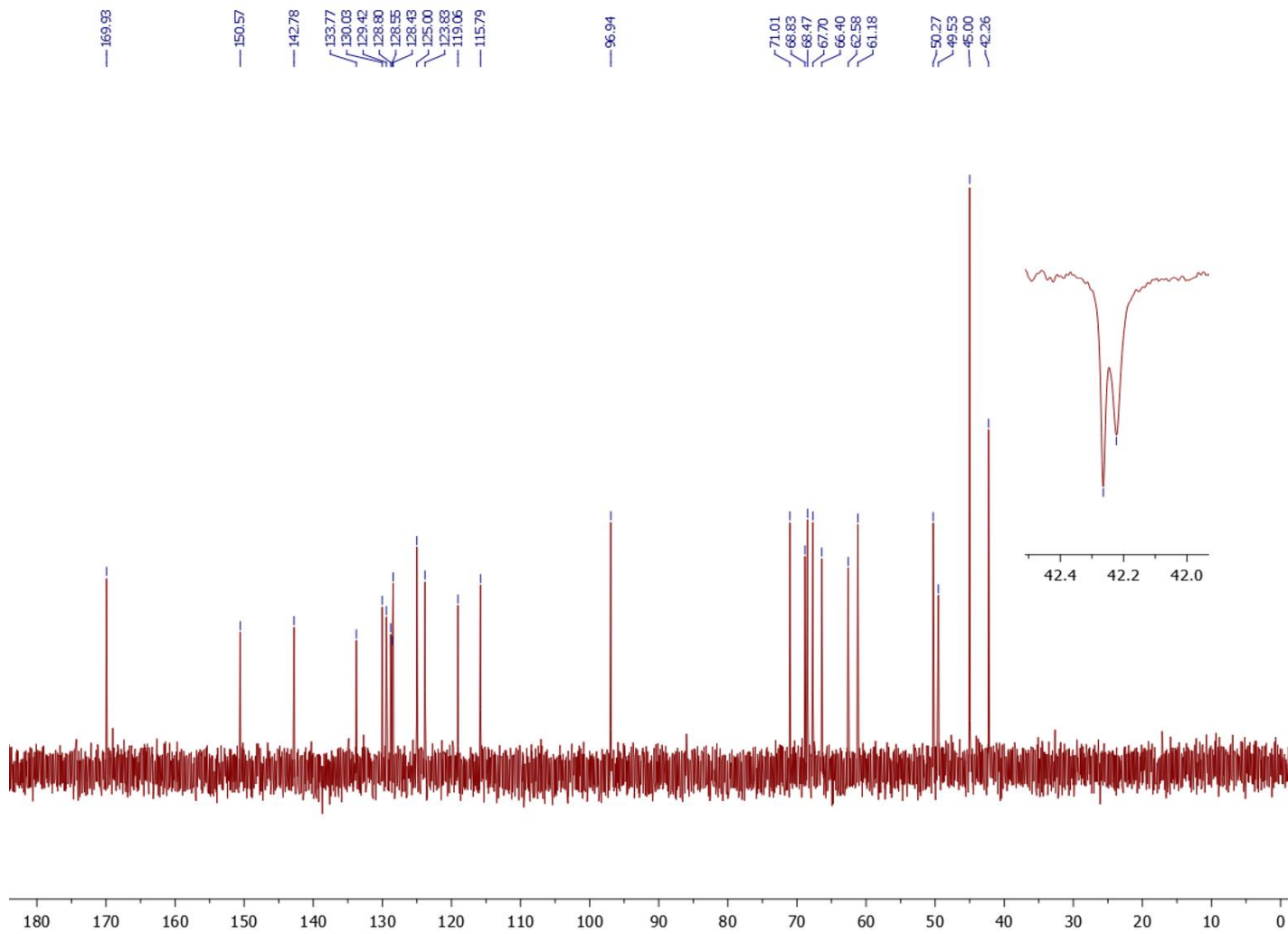


Figure A1.59. ^{13}C NMR of compd **25- α** in D_2O (region at 42.2 ppm has two peaks only visible on the DEPT135 spectrum)

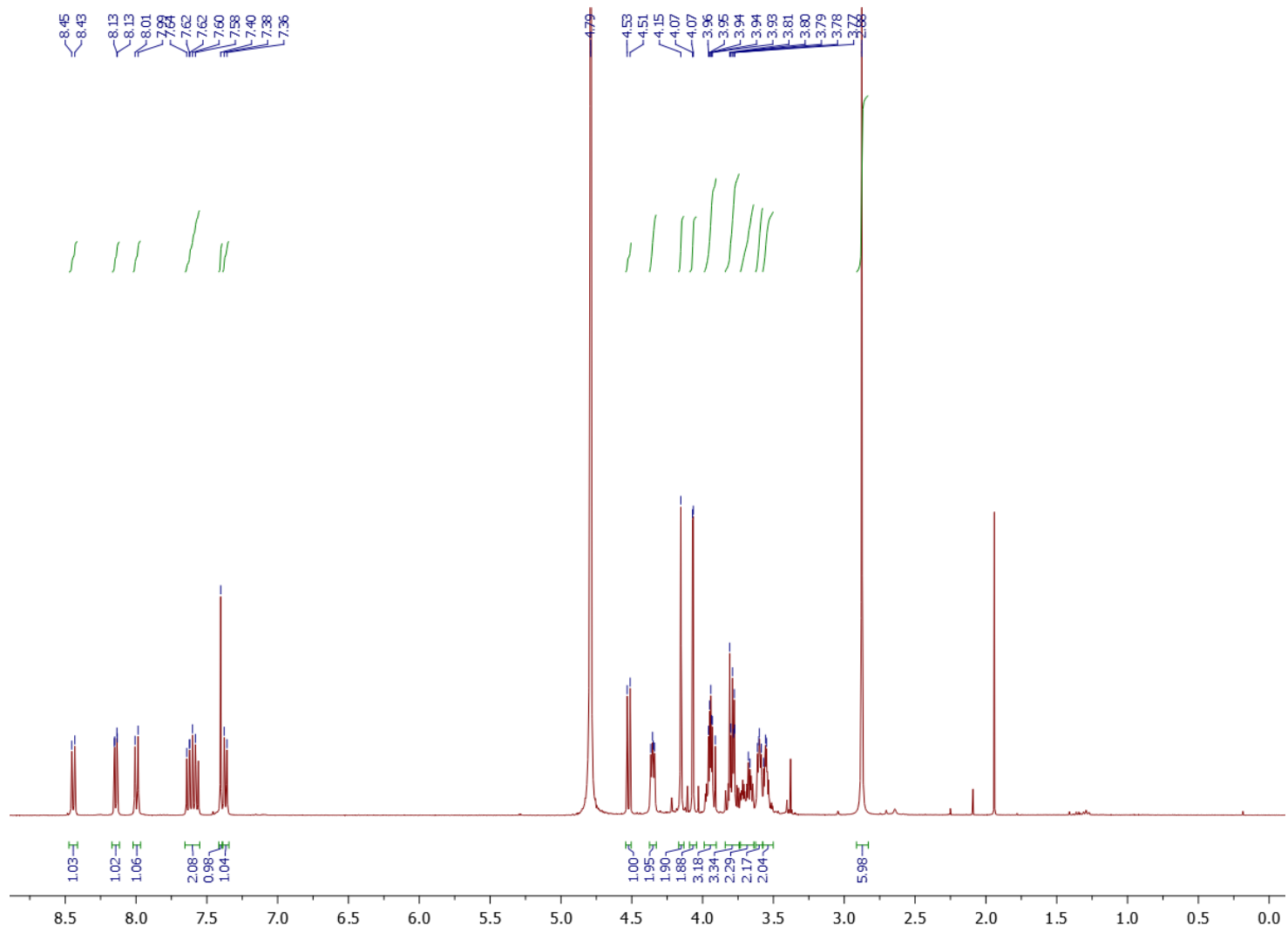


Figure A1.60. ^1H NMR of compd 25- β in D_2O

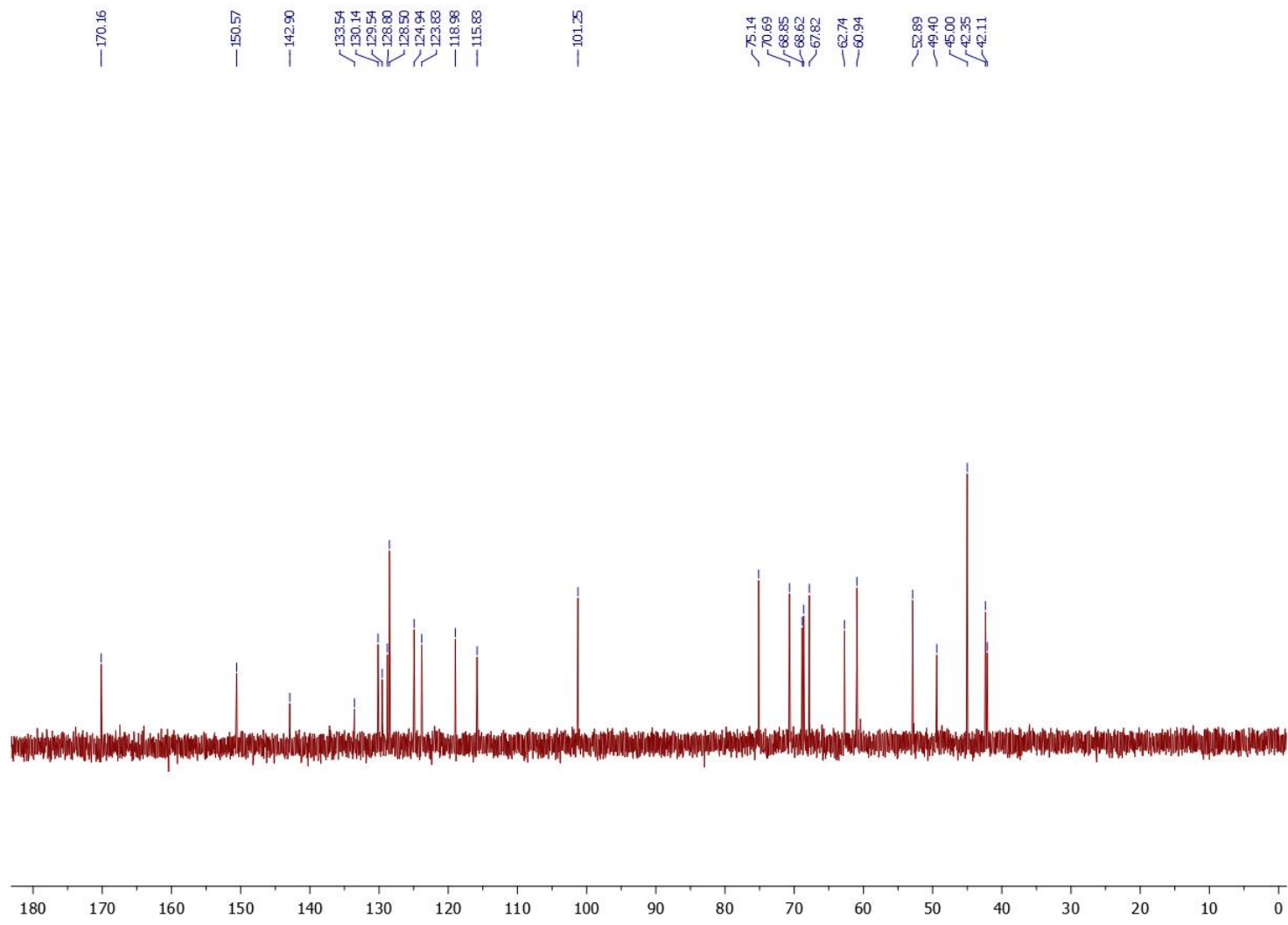


Figure A1.61. ^{13}C NMR of compd 25- β in D_2O

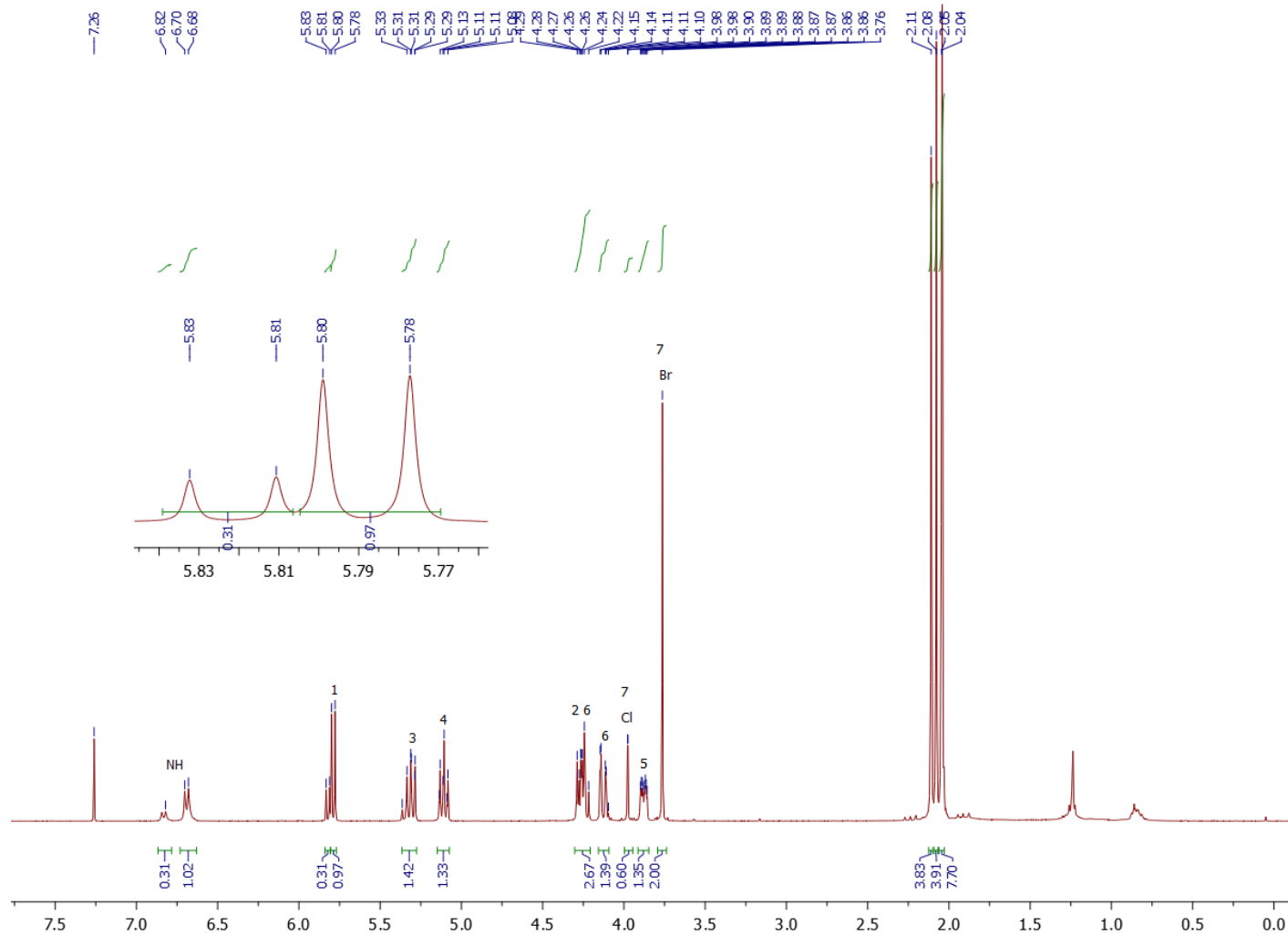


Figure A2.1. ¹H NMR of mixture of 9 and 9' in CDCl₃

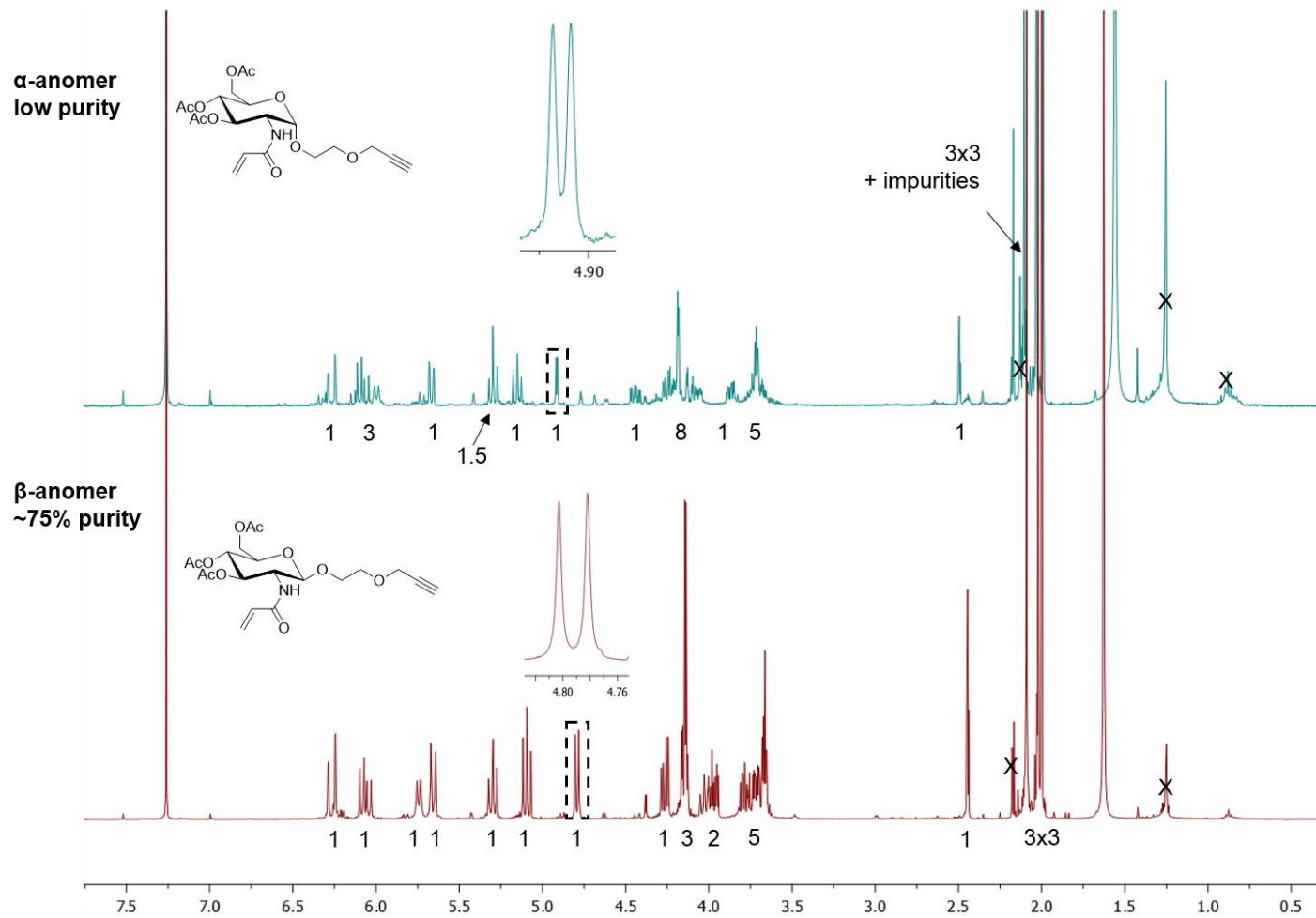


Figure A2.2. ^1H NMR of **12** analogues with MA WH in place (attempted synthesis)

1. LgtC control 2. LgtC + **11** 3. LgtC + DDT + **11** 30 mins incubation

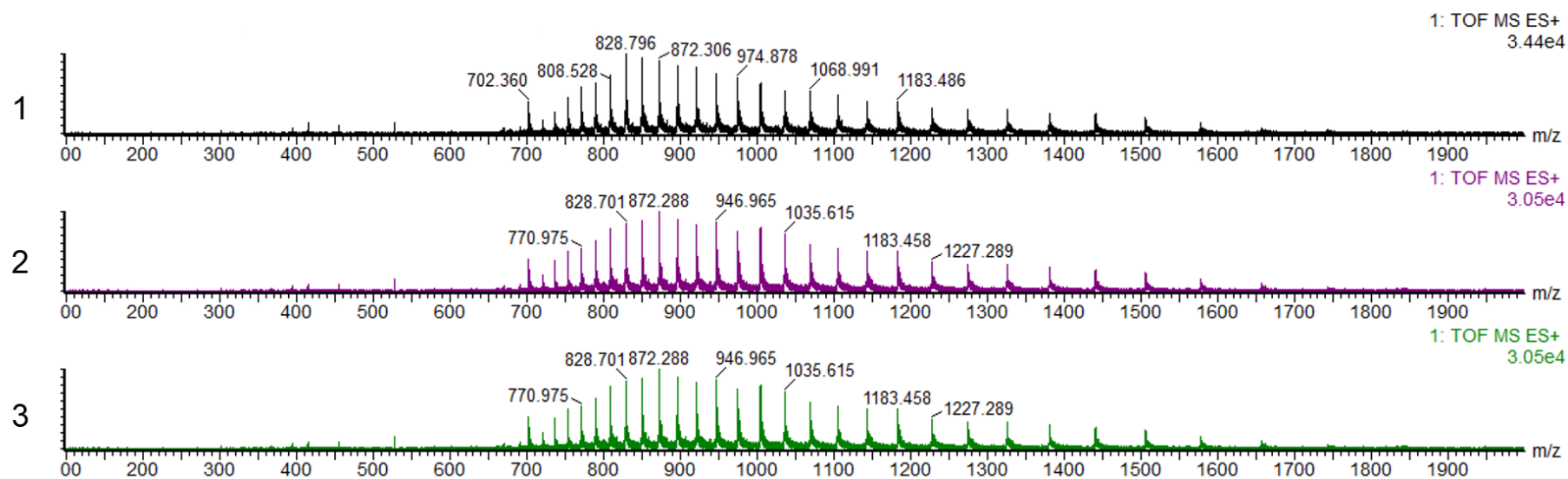


Figure A3.1. MS analysis for detection of potential covalent adduct between **11** and LgtC

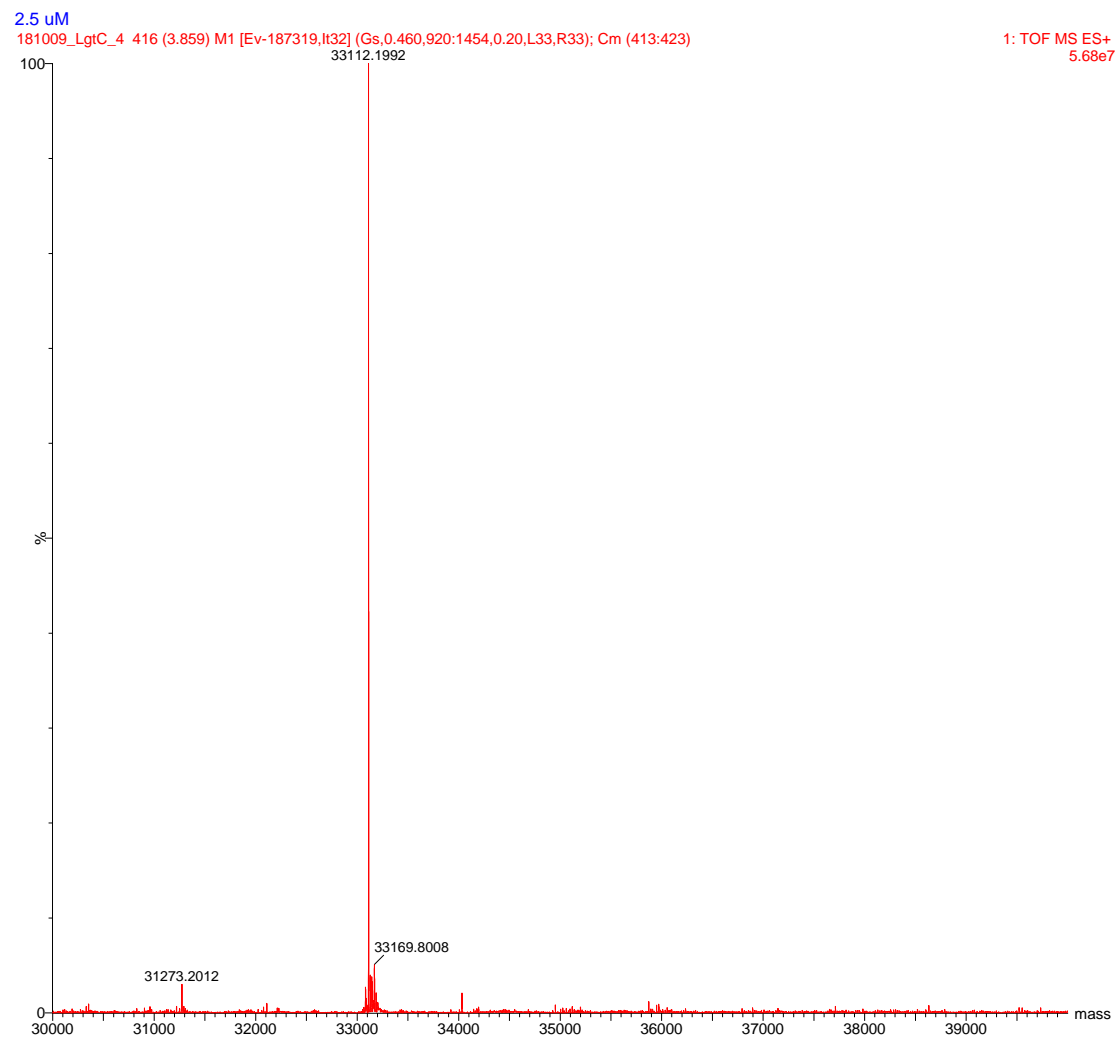


Figure A3.2. Deconvoluted MS spectrum: mass of LgtC non-modified (no covalent adduct detected)

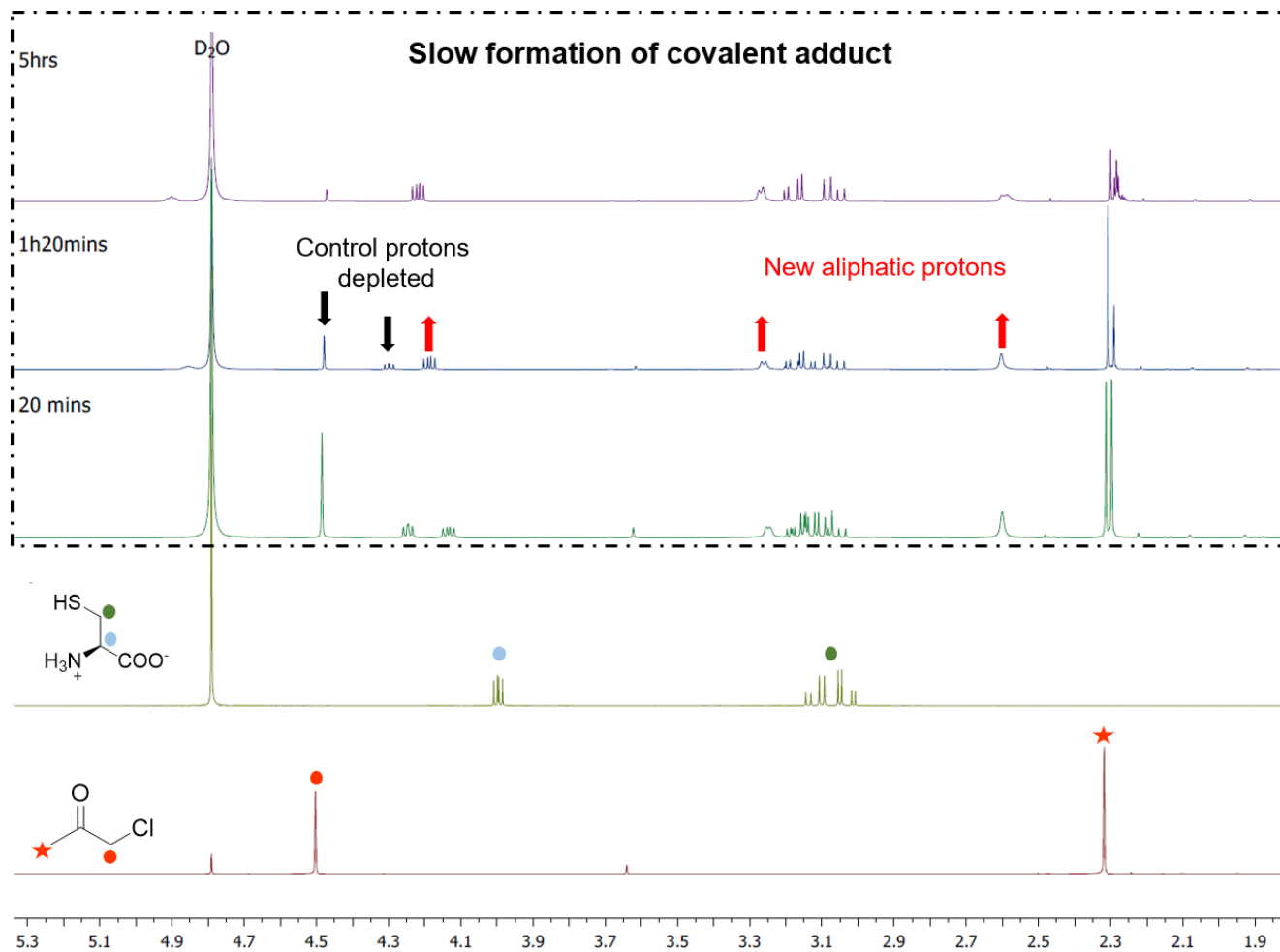


Figure A3.3. Reaction monitoring of chloroketone fragment with cysteine by ^1H NMR in buffered D_2O

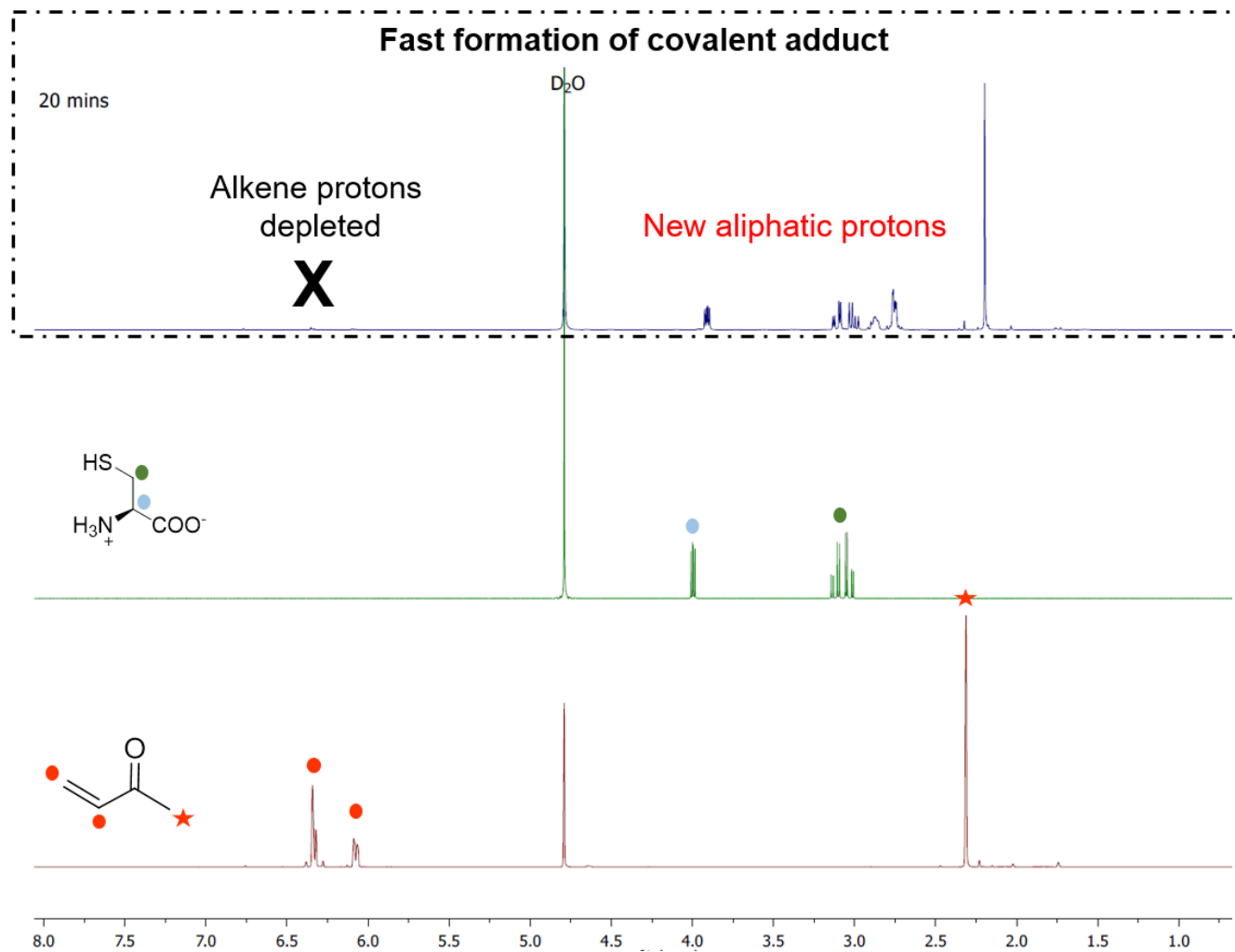


Figure A3.4. Reaction monitoring of enone fragment with cysteine by ^1H NMR in buffered D_2O

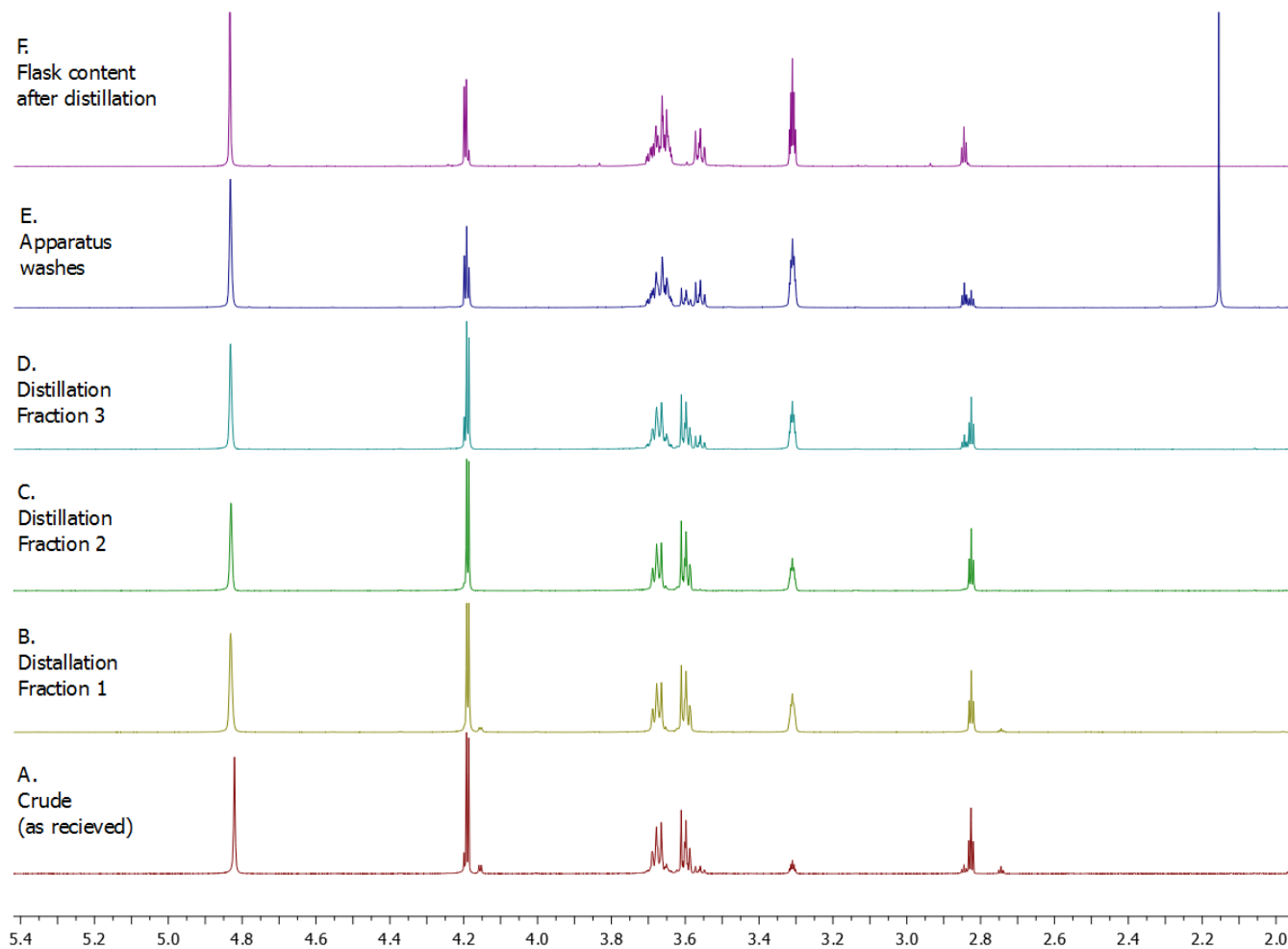


Figure A4.1. ^1H NMR of different collected fractions following the distillation of Propynol ethoxylate (in MeOD). A. Crude commercial product (10mL). B. C. and D. First, second and third fractions of 1mL, 4mL, and 2mL respectively collected from condensation (fraction 2 was the kept and used). E. washes of glassware with acetone (1mL, discarded). F. Content left in the flask after distillation (2mL, discarded)

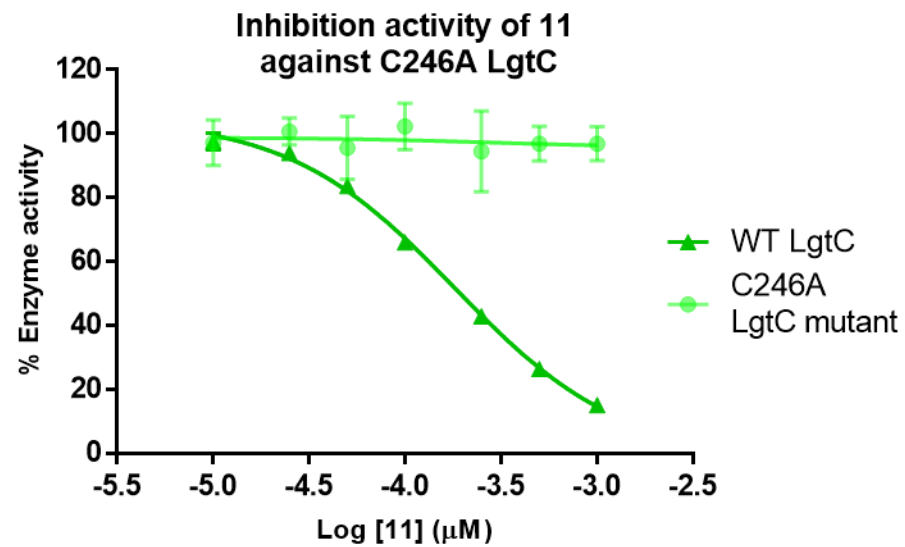
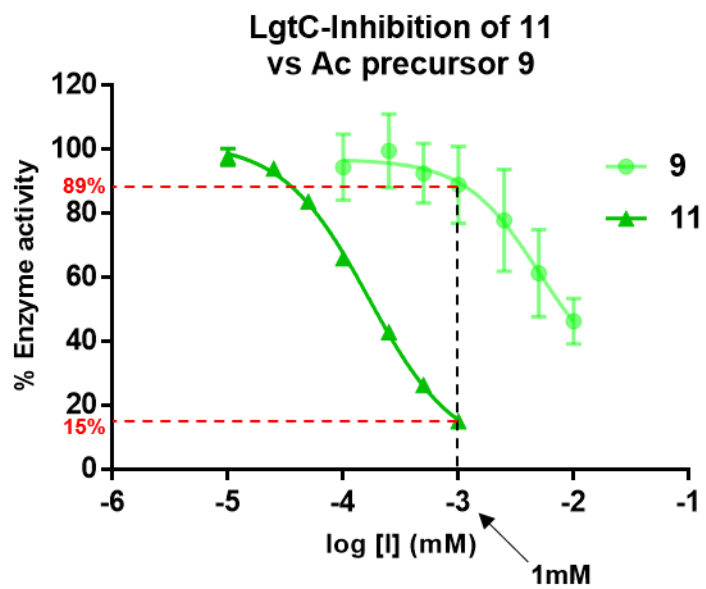


Figure A4.2. *Left:* Inhibition assay of 11 and 9 against WT LgtC. *Right:* Inhibition assay of 11 against WT and LgtC mutant. See Fig 2.11 for general conditions

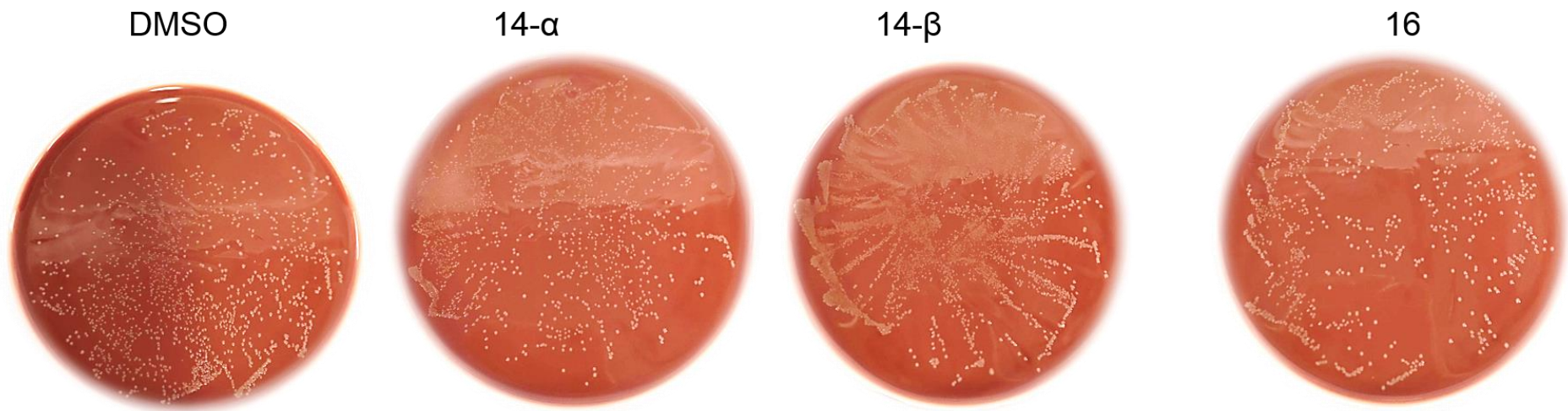


Figure A5.1. Viability test of *H. influenzae* R2866 after 2 hours incubation with glucosamine-based probes, azide 16 and DMSO performed on Agar chocolate plate (observed after 48hrs incubation at 37°C)

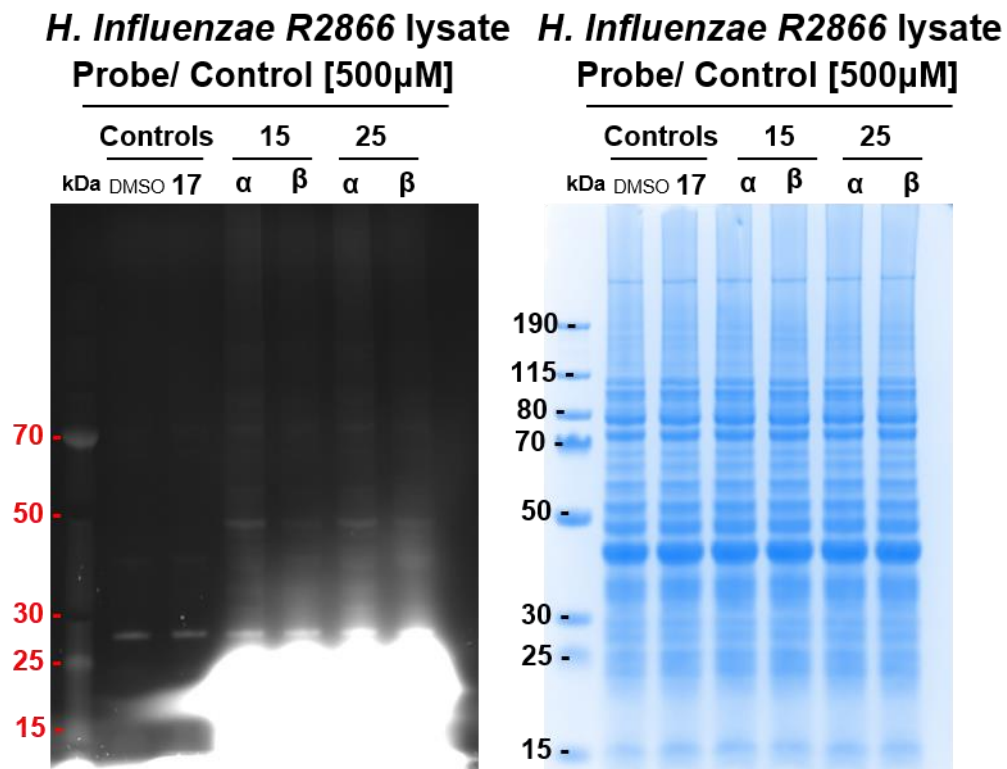


Figure A5.2. *H. influenzae* R2866 cell lysate protein labelling performed by all four deAc probes. *General conditions:* The cells are grown to OD₆₀₀= 0.9 and lysed with bug buster (100μL, X1 in PBS, for pellet of 4mL of cell culture) and mixed for 30mins at rt. Resulting cell lysates (9μL) are incubated with 25-α/β / 15-α/β / controls (1μL, 5mM stock) for 1 hour at 30°C. See Fig 5.12 for further sample treatment.

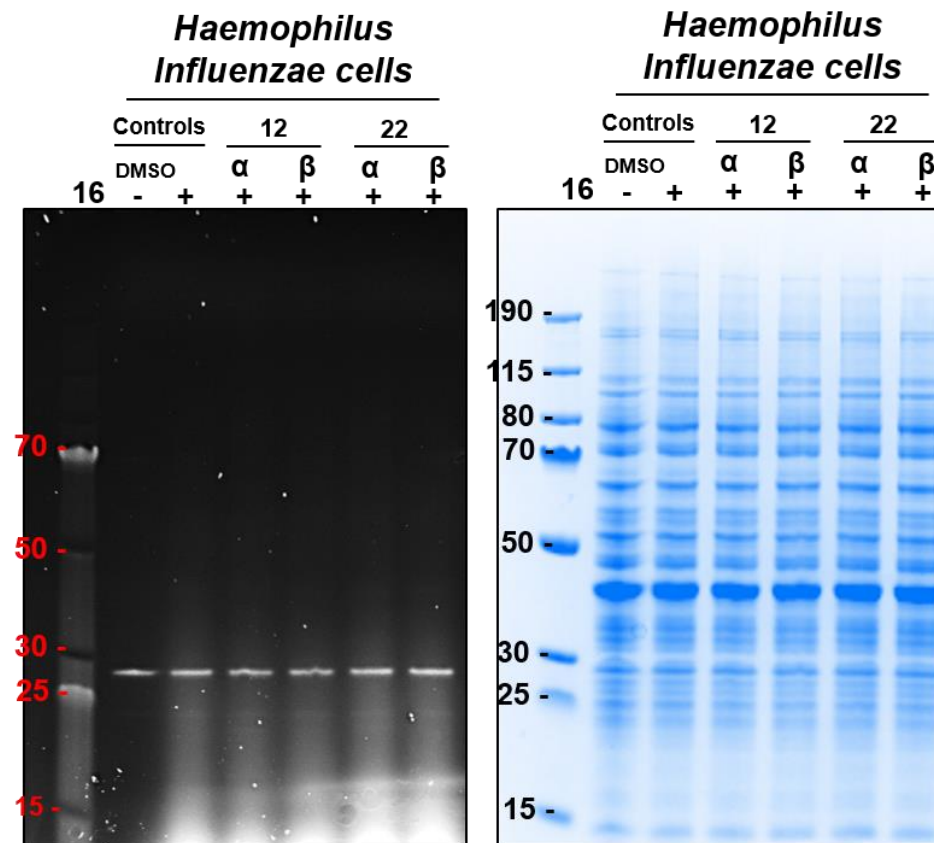


Figure A5.3. Two-step labelling of *H. influenzae* R2866 protein in a whole cell experiment performed by precursor probes. *General assay conditions:* **12- α/β / 22- α/β / DMSO** (10 μ L, 10mM stock in DMSO) are incubated with *H. influenzae* R2866 cell pellets (from 4mL of cell culture, suspended in 90 μ L of PBS) for 2 hours at 30°C. The cells are washed and lysed (100 μ L of Bug buster in PBS, 30 mins shaking at rt). Cell lysate are made up to 325 μ L with PBS and incubated with DMSO / 16 (25 μ L, 10mM stock), CuSO₄ (50 μ L, 2.5mM stock), THPTA (50 μ L, 12.5mM stock) and sodium ascorbate (50 μ L, 50mM stock) for 1 hour 30°C. The reaction mixture is washed in concentrators. See Fig 4.12 for following sample treatment. Fluorescence scanning (*left*) and Coomassie staining (*right*).

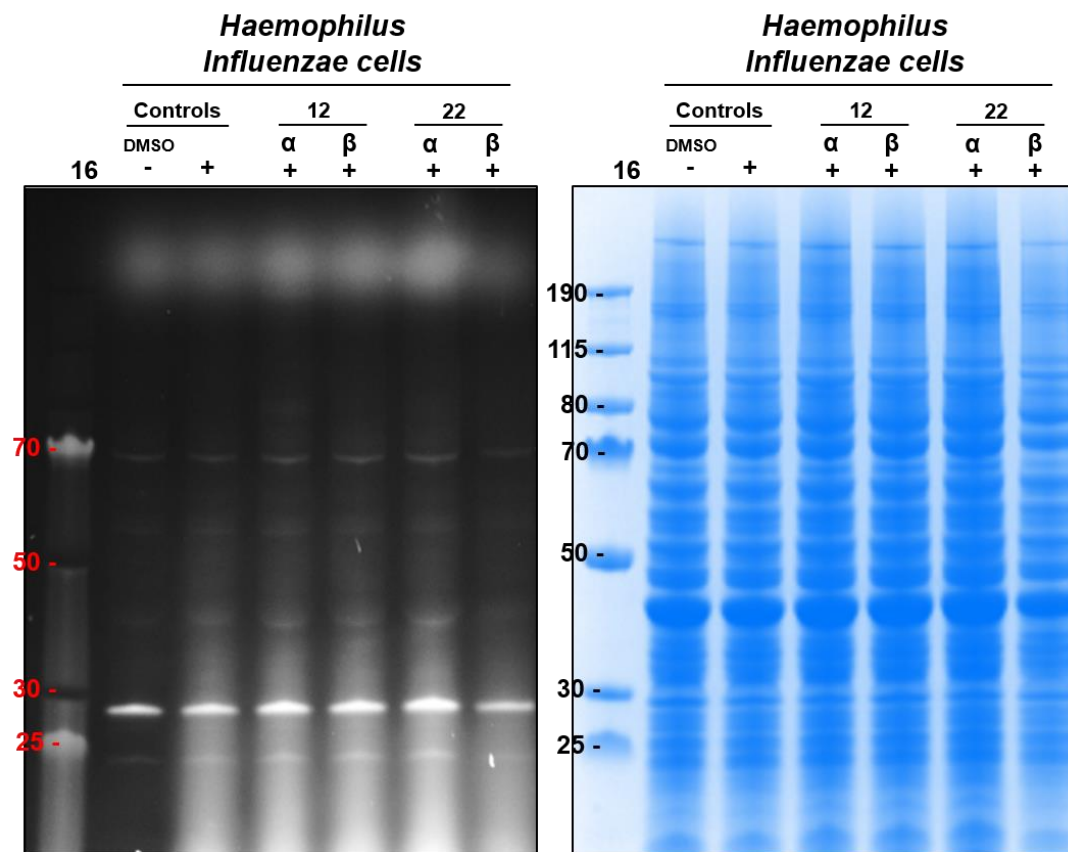


Figure A5.4. Two-step labelling of *H. influenzae* R2866 protein in a whole cell experiment performed by precursor probes. *General assay conditions:* **12- α/β / 22- α/β** / DMSO (10 μ L, 10mM stock in DMSO) are incubated with *H. influenzae* R2866 cell pellets (from 8mL of cell culture, suspended in 90 μ L of PBS) for 2 hours at 30°C. The cells are washed and lysed (100 μ L of Bug buster in PBS, 30 mins shaking at rt). Cell lysate are made up to 132.5 μ L with PBS and incubated with DMSO / 16 (2.5 μ L, 100mM stock), CuSO₄ (5 μ L, 25mM stock), THPTA (5 μ L, 125mM stock) and sodium ascorbate (5 μ L, 500mM stock) for 1 hour 30°C. The reaction mixture is washed in concentrators. See Fig 4.12 for following sample treatment. Fluorescence scanning (*left*) and Coomassie staining (*right*).