Approaches to Construct Glycan Library & Structure Based Design, Synthesis and Activity Studies of Small Hybrid Molecules as HDAC and G9A Dual Inhibitors

Qing Zhang

Follow this and additional works at: https://scholarworks.gsu.edu/chemistry_diss

Recommended Citation
Zhang, Qing, "Approaches to Construct Glycan Library & Structure Based Design, Synthesis and Activity Studies of Small Hybrid Molecules as HDAC and G9A Dual Inhibitors." Dissertation, Georgia State University, 2018.
doi: https://doi.org/10.57709/13413100
O-glycosylation is a common and complex post-translational modification of many biomolecules. While this dissertation will focus on the glycosylation with proteins, other types of biomolecules are also known to couple with glycans, most notably lipids. Protein glycosylation is usually divided into three different kinds of groups based on the first atom which links the glycan with the protein: N-type, O-type and the rare C-type. O-Glycosylation, the subject of this dissertation, is the glycans attached to the oxygen of the hydroxyl group on serine (Ser) or threonine (Thr) residue.
Chapter 1 describes a strategy for the synthesis of O-mannosyl glycan. O-Mannosylation is a vital protein modification involved in brain and muscle development, whereas the biological relevance of O-mannosyl glycans remains largely unknown, due to the lack of structurally defined glycoforms. An efficient scaffold synthesis/enzymatic extension (SSEE) strategy was thus developed to prepare such structures, by combining gram-scale convergent chemical synthesis of 3 scaffolds and strictly controlled sequential enzymatic extension catalyzed by glycosyltransferases. Totally, 45 O-mannosyl glycans were obtained, covering majority of identified mammalian structures. Subsequent glycan microarray analysis revealed fine specificities of glycan-binding proteins and specific antisera.

In chapter 2, we try to develop a new method to achieve O-GalNAc glycan from cell. Protein O-glycosylation is a universal post-translational modification that plays an essential role in many biological regulations. Recently we reported a technology termed Cellular O-Glycome Reporter/Amplification (CORA) to amplify and profile mucin-type O-glycans from living cells. However, the application and development of the CORA method is limited by the precursor function. Here we described a rapid parallel synthesis of cellular O-glycome precursors via microwave assisted reaction. In total, 26 Ac3GalNAc-α-Bn derivatives, including fluorescent and other reactive functional groups, were successfully synthesized. Furthermore, subsequent activity screening and evaluation of these precursor toward living cell and T-synthase were performed.

In chapter 3, the first small molecule capable of acting as a dual inhibitor targeting both G9a and HDAC was reported. Aberrant enzymatic activities or expression profiles of epigenetic regulations are therapeutic targets for cancers. Among these, histone 3 lysine 9 methylation (H3K9Me2) and global de-acetylation on histone proteins are associated with multiple cancer phenotypes including leukemia, prostatic carcinoma, hepatocellular carcinoma and pulmonary...
carcinoma. In this study, we report the discovery of the first small molecule capable of acting as a dual inhibitor targeting both G9a and HDAC. Our structure based design, synthesis, and screening for the dual activity of the small molecules led to the discovery of compound 14 which displays promising inhibition of both G9a and HDAC in a low micro-molar range in cell based assays.

INDEX WORDS: O-glycan, chemoenzymatic synthesis, microarrays, O-mannosyl glycans, CORA, O-GalNAc glycan, epigenetics, dual inhibitors, G9a inhibitors, HDAC inhibitors
APPROACHES TO CONSTRUCT GLYCAN LIBRARY &
STRUCTURE BASED DESIGN, SYNTHESIS AND ACTIVITY STUDIES OF SMALL
HYBRID MOLECULES AS HDAC AND G9A DUAL INHIBITORS

by

QING ZHANG

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy
in the College of the Arts
Georgia State University
2018
APPROACHES TO CONSTRUCT GLYCAN LIBRARY &
STRUCTURE BASED DESIGN, SYNTHESIS AND ACTIVITY STUDIES OF SMALL
HYBRID MOLECULES AS HDAC AND G9A DUAL INHIBITORS

by

QING ZHANG

Committee Chair: Peng George Wang
Committee: Jun Yin
Gangli Wang

Electronic Version Approved:
Office of Graduate Studies
College of the Arts
Georgia State University
December 2018
DEDICATION

Dedicated to my wife Ying Luo, and my parents, for their love, understanding and support.
ACKNOWLEDGEMENTS

I really appreciate my research advisor Dr. Peng George Wang for his financial support and incessant guidance in the past five years. Dr. Wang’s enthusiasm for sciences and intelligence inspired my research careers. Dr. Jun Yin and Dr. Gangli Wang deserve recognition for their willingness to serve on my Defense Committees. Many thanks to Dr. Tongzhong Ju in Emory for the collaborative projects.

It is my fortune to work with the Wang Group, deserving to mention here. Especially, I need to thank Dr. Zhongying Xiao who introduced me to the art of organic synthesis during my first year. I also want to thank Dr. Tiehai Li and Dr. Yunpeng Liu for their patient help and guidance for my research. Many thanks to Dr. Zhonghua Li, Dr. Lei Li, Dr. Xuefeng Cao, Dr. Varma Saikam and Dr. Gaolan Zhang for providing meaningful suggestions for my projects. Thanks for Shukkoor Kondangaden and Lanlan Zang for dual inhibitor project, Shuaishuai Wang for O-mannosyl glycan project and other members for my research career.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS .................................................................................................................. V  

LIST OF TABLES .......................................................................................................................... IX  

LIST OF FIGURES ........................................................................................................................ X  

LIST OF ABBREVIATIONS .............................................................................................................. XIII  

1 FACILE CHEMOENZYMATIC SYNTHESIS OF O-MANNOSYL GLYCAN  

1.1 Introduction .......................................................................................................................... 1  

1.1.1 Biosynthesis of O-mannosyl Glycan ................................................................................. 1  

1.1.2 Functional effect of O-mannosylation .............................................................................. 4  

1.2 Results and discussion ........................................................................................................... 7  

1.2.1 Convergent synthesis of scaffold structures ................................................................. 7  

1.2.2 Enzymatic extension of scaffold structures ...................................................................... 9  

1.2.3 Microarray analysis .......................................................................................................... 14  

1.3 Conclusion ............................................................................................................................ 16  

1.4 Experiment Section .............................................................................................................. 16  

1.4.1 Chemical synthesis .......................................................................................................... 16  

1.4.2 Enzymatic extension ........................................................................................................ 35  

1.4.3 Glycan Microarray preparation, assay and data .............................................................. 40  

1.4.4 NMR and MS analysis of purified O-mannosyl glycan................................................... 44
2 DESIGN, SYNTHESIS AND ACTIVITY STUDIES OF CELLULAR O-
GLYCOME PRECURSOR .................................................................................. 64

2.1 Introduction ........................................................................................................ 64

2.1.1 Biosynthesis of O-GalNAc protein ................................................................. 64

2.1.2 Functional effects of O-GalNAc glycosylation ............................................. 66

2.1.3 CORA technology .......................................................................................... 70

2.2 Result and discussion .......................................................................................... 72

2.2.1 Design and Synthesis of CORA precursor candidates. ............................... 72

2.2.2 CORA activity and T-synthase activity assays ............................................ 73

2.2.3 SAR study and evaluation of the precursor activity ..................................... 77

2.3 Conclusion .......................................................................................................... 78

2.4 Experiment section ............................................................................................. 78

2.4.1 Chemical synthesis ......................................................................................... 78

3 STRUCTURE BASED DESIGN, SYNTHESIS AND ACTIVITY STUDIES
OF SMALL HYBRID MOLECULES AS HDAC AND G9A DUAL INHIBITORS ...... 95

3.1 Introduction ........................................................................................................ 95

3.2 Result and discussion .......................................................................................... 98

3.2.1 Compound design ......................................................................................... 100

3.2.2 Synthesis ....................................................................................................... 101

3.2.3 Functional potency evaluation for G9a inhibition ...................................... 105
3.2.4 Functional potency evaluation for HDAC inhibition.......................... 106

3.2.5 Molecular Docking Analysis ............................................................ 107

3.2.6 Cell Anti-proliferation Assay............................................................ 109

3.2.7 ADME Prediction Studies................................................................. 111

3.3 Conclusions .......................................................................................... 112

3.4 Experiment Sections.............................................................................. 113

3.4.1 Synthesis ............................................................................................ 114

3.4.2 Molecular Docking Analysis ............................................................... 129

3.4.3 Cloning, Protein Expression and Purification ..................................... 130

3.4.4 MALDI-TOF-MS................................................................................ 131

3.4.5 Cell Based Assays................................................................................ 131

3.4.6 HDAC Activity Assay........................................................................... 132

3.4.7 G9a H3K9me2 Cellular Assay............................................................. 133

3.4.8 Toxicity Assay...................................................................................... 133

APPENDICES: $^1$H AND $^{13}$C SPECTRA OF COMPOUNDS............................... 135

REFERENCES.............................................................................................. 203
LIST OF TABLES

Table 3.1 Combination study of BIX-01294 and SAHA .................................................. 99

Table 3.2 H3K9Me2 cell immunofluorescence In-Cell Western (ICW) assay results .......... 105

Table 3.3 Results of homogeneous cellular histone deacetylase assay ................................ 107

Table 3.4 GLIDE docking results for MS-344 and compound 14 at the catalytic site of HDAC8
(PDB ID: 1T67) ................................................................................................................. 108

Table 3.5 GLIDE docking results for BIX-01294 and compound 14 at the catalytic site of G9a
(PDB ID: 3FPD) ................................................................................................................. 108

Table 3.6 Inhibition of compounds 13 and 14 on the growth of cancer cells and normal cells. 111

Table 3.7 ADME prediction results ..................................................................................... 111
LIST OF FIGURES

Figure 1.1 (A) The 4 cores of mammalian O-mannosyl glycans; (B) identified extensions on core M1 and M2........................................................................................................................................................................... 2

Figure 1.2 Biosynthesis of representative O-mannosyl glycan in the ER and Golgi apparatus. ... 3

Figure 1.3 The 45 core M1 and M2 O-mannosyl glycans prepared in this study....................... 6

Figure 1.4 (A) The four building blocks; (B) The synthetic scheme for the assembly of the three O-mannosyl scaffold structures. .................................................................................................................................................. 8

Figure 1.5 Enzymatic extension of core M1 and core M2 structures. ........................................ 10

Figure 1.6 Enzymatic extension of M301 to generate core M2 O-mannosyl glycans................. 13

Figure 1.7 Microarray analysis and evaluation of binding to spotted O-mannosyl glycans....... 15

Figure 1.8 Microarray analysis towards Aleuria aurantia lectin................................................ 42

Figure 1.9 Microarray analysis towards Concanavalin A from Canavalia ensiformis............. 43

Figure 1.10 Microarray analysis and binding profile of O-mannosyl glycans towards rabbit 26559.................................................................................................................................................................................................................. 43

Figure 1.11 Microarray analysis towards mouse I1H6 antibody................................................. 44

Figure 2.1 Biosynthesis of representative O-GalNAc protein in the ER and Golgi apparatus..... 66

Figure 2.2 Principle of CORA technology.................................................................................. 71

Figure 2.3 Structure of designed Ac3GalNAc-α-Bn derivatives............................................... 72

Figure 2.4 Microwave assisted synthesis of Ac3GalNAc-α-Bn derivatives......................... 73

Figure 2.5 Synthesis of Ac3GalNAc-α-Bn derivatives with azido group................................. 73

Figure 2.6 CORA activity and T-synthase activity of designed precursors.............................. 77

Figure 3.1 Examples for known HDAC inhibitors and G9a Inhibitors. .................................. 96
Figure 3.2 A) Design for Hybrid Molecules, B) Structural representation of designed compounds. ........................................................................................................................................ 101

Figure 3.3 Synthesis scheme of dimethoxy analogs................................................................................................................ 102

Figure 3.4 Synthesis scheme for monomethoxy analogs........................................................................................................ 104

Figure 3.5 Synthesis scheme of compound 19-22. ................................................................................................................. 105

Figure 3.6 Molecular docking study results.......................................................................................................................... 110
This dissertation is written based on following published paper:

1. Shuaishuai Wang, Qing Zhang, CongCong Chen, Yuxi Guo, Madhusudhan Reddy Gadi, Jin Yu, Ulrika Westerlind, Yunpeng Liu, Xuefeng Cao, Peng George Wang* and Lei Li*, Facile chemoenzymatic synthesis of O-mannosyl glycans. Angew. Chem. Int. Ed. 2018, 57, 9268-9273 (First two authors contributed equally to this work.)

2. Lanlan Zang, Shukkoor M. Kondengaden, Qing Zhang, Xiaobo Li, Dilep K. Sigalapalli, Shameer M. Kondengadan, Kenneth Huang, Keqin Kathy Li, Shanshan Li, Zhongying Xiao, Liuqing Wen, Hailiang Zhu, Bathini N. Babu, Lijuan Wang, Fengyuan Che and Peng George Wang*, Structure based design, synthesis and activity studies of small hybrid molecules as HDAC and G9a dual inhibitors. Oncotarget. 2017, 8, 63187-63207. (First three authors contributed equally to this work.)
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>alpha</td>
</tr>
<tr>
<td>Ac</td>
<td>acetyl</td>
</tr>
<tr>
<td>br</td>
<td>broad (IR and NMR)</td>
</tr>
<tr>
<td>$\beta$</td>
<td>beta</td>
</tr>
<tr>
<td>n-Bu</td>
<td>normal-butyl</td>
</tr>
<tr>
<td>t-Bu</td>
<td>tert-butyl</td>
</tr>
<tr>
<td>Bz</td>
<td>benzoyl</td>
</tr>
<tr>
<td>$^\circ$C</td>
<td>degrees Celsius</td>
</tr>
<tr>
<td>calcd</td>
<td>calculated</td>
</tr>
<tr>
<td>$\delta$</td>
<td>chemical shift in parts per million downfield from tetramethylsilane</td>
</tr>
<tr>
<td>d</td>
<td>doublet (spectra); day(s)</td>
</tr>
<tr>
<td>DBU</td>
<td>1,8-diazabicyclo[5.4.0]undec-7-ene</td>
</tr>
<tr>
<td>DMAP</td>
<td>4-(N,N-dimethylamino)pyridine</td>
</tr>
<tr>
<td>DMF</td>
<td>N,N-dimethylformamide</td>
</tr>
<tr>
<td>DMSO</td>
<td>dimethylsulfoxide</td>
</tr>
<tr>
<td>EDCI</td>
<td>N-(3-dimethylaminopropyl)-N'-'ethylcarbodiimide hydrochloride</td>
</tr>
<tr>
<td>eq.</td>
<td>equivalent</td>
</tr>
<tr>
<td>Et</td>
<td>ethyl</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>gamma</td>
</tr>
<tr>
<td>g</td>
<td>gram(s)</td>
</tr>
<tr>
<td>h</td>
<td>hour(s)</td>
</tr>
<tr>
<td>J</td>
<td>coupling constant in Hz (NMR)</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Chemical Name</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------------------------------------------</td>
</tr>
<tr>
<td>TBAF</td>
<td>tetrabutylammonium fluoride</td>
</tr>
<tr>
<td>TBAI</td>
<td>tetrabutylammonium iodide</td>
</tr>
<tr>
<td>TBS</td>
<td>t-butyldimethylsilyl</td>
</tr>
<tr>
<td>Tf</td>
<td>trifluoromethanesulfonyl</td>
</tr>
<tr>
<td>THF</td>
<td>tetrahydrofuran</td>
</tr>
<tr>
<td>TLC</td>
<td>thin layer chromatography</td>
</tr>
<tr>
<td>TMS</td>
<td>trimethylsilyl</td>
</tr>
<tr>
<td>TSTU</td>
<td>N,N,N′,N-tetramethyl (succinimido) uronium tetrafluoroborate</td>
</tr>
</tbody>
</table>
1 FACILE CHEMOENZYMATIC SYNTHESIS OF O-MANNOSYL GLYCANS

O-Mannosylation is a vital protein modification involved in brain and muscle development, whereas the biological relevance of O-mannosyl glycans remains largely unknown, due to the lack of structurally defined glycoforms. An efficient scaffold synthesis/ enzymatic extension (SSEE) strategy was thus developed to prepare such structures, by combining gram-scale convergent chemical synthesis of 3 scaffolds and strictly controlled sequential enzymatic extension catalyzed by glycosyltransferases. Totally, 45 O-mannosyl glycans were obtained, covering majority of identified mammalian structures. Subsequent glycan microarray analysis revealed fine specificities of glycan-binding proteins and specific anti-sera.

1.1 Introduction

1.1.1 Biosynthesis of O-mannosyl Glycan

O-Mannosylation is a vital protein post-translational modification that plays essential roles in brain and muscle development and normal tissue function.\textsuperscript{[1]} Alpha-dystroglycan (α-DG) is the most studied O-mannosyl protein that is widely distributed in muscle and brain tissues. Abnormal O-mannosylation on α-DG disrupts the receptor function of dystroglycan and leads to congenital muscular dystrophy.\textsuperscript{[2]} In addition, O-mannosyl glycans have been reported to be involved in other human diseases including arenaviral Infection, cancer and metastasis.\textsuperscript{[3]}

In mammalian brain tissue, O-mannosyl glycans account for up to 30% of all O-linked glycans.\textsuperscript{[4]} Regardless of the abundance, studies on mammalian O-mannosylation have been mainly focused on α-DG. Recent advances in glycoproteomics enabled discovery of an increased number of O-mannosylated proteins and various O-mannosyl glycans.\textsuperscript{[1b, 5]} To date, more than 20 O-mannosyl glycans have been identified, which are classified into four core types.\textsuperscript{[5]} As illustrated in Figure 1.1A, core M0, identified in cadherin family proteins,\textsuperscript{[6]} represents a single structure of
one mannose (Man) residue α-linked to serine (Ser) or threonine (Thr). Core M1 contains a linear β1,2-linked N-acetylglucosamine (GlcNAc), whereas the branched core M2 contains both β1,2- and β1,6-linked GlcNAc-residue. Core M3 is a phosphorylated trisaccharide, to which glycosaminoglycan-like polysaccharides usually are attached.\textsuperscript{[5]}

**Figure 1.1** (A) The 4 cores of mammalian O-mannosyl glycans; (B) identified extensions on core M1 and M2.

All O-Man glycans are initiated by addition of mannose by a pair of mammalian protein O-mannosyl-transferases: POMT1 and POMT2, which appear to act as a functional heterodimer in vivo.\textsuperscript{[7]} These enzymes have different expression patterns across tissues in humans, and different tissue-specific isoforms formed through alternative gene splicing.

Like most other O-glycan-transferases, both POMT1 and POMT2 are found in the ER (Fig 1.1), but interestingly they use a unique sugar donor: dolichol phosphate mannose (Dol–P–Man), unlike other glycosyltransferases which utilize nucleotide sugars.\textsuperscript{[1a]} After the attachment of the first mannosyl residue, glycoproteins are transported to the Golgi apparatus. Both Core M1 and Core M2 structures are then extended by addition of GlcNAc through a β1–2 linkage by the UDP–GlcNAc protein O-mannose β1-2 N-acetylgalactosaminyltransferase 1 (POMGNT1).\textsuperscript{[8]} POMGNT1 works in the Golgi, and appears to be completely independent of the enzymes that catalyze
formation of the GlcNAcβ1–2Man linkages in N-glycans, since those two enzymes (GnT-I and GnT-II) show no reactivity towards O-mannosyl peptides in vitro.\[^9\] The core M1 structure is then further elongated by Golgi-resident glycosyltransferases, including galactosyltransferases (GALTs), sialyltransferases (SIATs), glucuronyltransferases (GLCATs), sulfotransferase HNK-1ST and α1-3 fucosyltransferase 9 (FUT9), to give a variety of end structures, but detailed mechanisms of how these steps are regulated are not well-understood.\[^1a\] Core M2 glycans, thought to exist only in the brain, have a branched trisaccharide core: GlcNAcβ1–6(GlcNAcβ1–2)Man–Ser/Thr. After synthesis of the Core M1 dimer, a branching GlcNAcβ1–6 can be added to form Core M2 structures. This is done by the brain-specific N-acetylglucosaminyltransferase IX (GnT-IX) enzyme,\[^10\] which is also involved in the synthesis of highly branched complex-type N-glycans.\[^11\]

\[\text{Figure 1.2 Biosynthesis of representative O-mannosyl glycan in the ER and Golgi apparatus.}\]
1.1.2 Functional effect of O-mannosylation

Understanding the works of O-mannosylation on the biophysical and biological properties on a protein or peptide is an important research area as well. Given the well documented structural effects of O-GalNAc type glycans on mucin domains in other systems, it was of significant interest to investigate how the presence of both O-Man and O-GalNAc glycans on the mucin domain of α-DG might affect the structure of the protein. As with O-GalNAc glycosylation, biophysical and structural studies of synthetic model systems allowed for quick and robust conclusions to be drawn. For example, by synthesizing a series of glycopeptides derived from the α-DG mucin region and containing up to four consecutive O-Man residues, it was possible to conclude that the glycopeptides existed in a mostly unordered structure.\[^{[12]}\] Interestingly, when the glycans were switched to O-GalNAc, the same sequence was considerably more ordered, as would be expected given the previous work on mucin-domain structures.\[^{[13]}\] This comparison held across three different biophysical characterization methods: circular dichroism (CD), NMR and hydrogen–deuterium exchange (HDX).\[^{[12]}\] Combined with the well-documented importance of O-Man glycans in the functioning of α-DG,\[^{[1]}\] these observations led to a proposed model where the O-GalNAc glycans provide the mucin domain with a defined, rigid structure that supports optimal display of the functionally important end-groups of the O-Man glycans.\[^{[12, 14]}\] The results of structural studies of the glycopeptide bearing a negatively charged 6-O-phosphomannose residue also supported this assertion. As with previous work on the structural effects of O-mannosylation, the authors found little structural difference between the unglycosylated peptide and the phosphomannose-containing glycopeptide.\[^{[15]}\] Replacing the mannose derivative with an O-GalNAc, however, resulted in significant changes in the NMR spectra of the glycopeptide.
In addition to the studies of the effects of O-glycosylation on unstructured mucin domains, many groups, including our own, have undertaken research programs to get this knowledge with respect to folded peptides and proteins. One example of this type of research is the work done by the Tan group on a Family 1 carbohydrate-binding module (CBM) of a fungal cellulase. The CBM is a small, 36-residue peptide domain and is responsible for recognition and binding of cellulose. It is naturally glycosylated by several O-glycans, which have been implicated as critical in the functions of this domain.\textsuperscript{16} In order to characterize the specific effects of O-mannosylation on the CBM, a library of CBM glycoforms bearing mono- and oligo-mannose chains at each of the three possible glycosylation sites were chemically synthesized. By comparing the individual members of this glycoform library, the Tan group was able to show that O-mannosylated CBM was more stable and bound tighter to cellulose.\textsuperscript{17} Particularly intriguingly, the authors found that glycosylation at Ser3, more so than at either of the other two sites, was the single most important factor in the thermostability of the glycoforms.

The core M1 and M2 are typically extended and have been identified with structures including N-acetyllactosamine (Gal- β1,4-GlcNAc, LN), 3-sialyl-N-acetyllactosamine (Neu5Ac-α2,3-Gal-β1,4-GlcNAc, 3SLN), Lewis X [Gal-β1,4-(Fuc-α1,3-)GlcNAc, LeX], sialyl-LeX (sLeX), and the human natural killer-1 (3-sulfo-GlcA-β1,3-Gal-β1,4-Glc, HNK-1) epitopes (\textbf{Fig 1.1B}).\textsuperscript{11b} These terminal epitopes largely diversify core M1 and M2 based glycans and in fact, the majority of identified structures belong to these two core types.\textsuperscript{11b} Additionally, it was found that the structures centered on core M1 and M2 account for 15% and 5% of all brain O-linked glycans, respectively.\textsuperscript{4} Despite the abundance of core M1 and M2 structures, very little is known about their functional relevance.\textsuperscript{5} Therefore, structurally well-defined O-mannosyl glycans not only
serve as ideal standards to identify and characterize such glycoforms, but also provide unique probes to uncover their biological roles.

To date, only a few O-mannosyl glycans were prepared. A core M1 tetrasaccharide (Fig 1.2, M102) was firstly synthesized in 1999, and later two lower homologs (M100, M101). Most recently, Cao and co-workers prepared six core M1 structures (M100, M101, M102, M104, M105, and M102G) employing a chemoenzymatic approach. Glycopeptides harboring basic O-mannosyl structures (e.g., M101, M102, core M0, M000, or core M3) were also generated. Nevertheless, the majority of O-mannosyl glycans (especially core M2 branched structures) are still not accessible, which hampered in-depth understanding of O-mannosylation. Here an efficient strategy was developed employing scaffold synthesis/enzymatic extension (SSEE) to prepare 45 structurally well-defined core M1 and M2 based O-mannosyl glycans (Fig 2), covering all identified glycoforms with the exception of the HNK-1 epitope.

**Figure 1.3** The 45 core M1 and M2 O-mannosyl glycans prepared in this study.
1.2 Results and discussion

1.2.1 Convergent synthesis of scaffold structures

Considering the structural signatures of O-mannosyl glycans, it is concluded that all core M1 and M2 structures can be enzymatically elaborated from the following three scaffolds, M100, M201 and M301 (Fig 1.2). We envisioned that the protected Man derivative 1 (Fig 1.3A) would serve as a versatile precursor for the synthesis of these scaffolds. It has a free hydroxyl group at the C-2 position and the 4,6-hydroxyl groups are protected as the benzylidene acetal. This arrangement allows performing chemical glycosylation at the C-2 position followed by the C-6 position with the deprotection of the acetal protecting group.

As shown in Figure 1.3, the three scaffolds required for the enzymatic extension were synthesized in a convergent approach and were assembled by three simple building blocks, 1, 2 and 3 and the Fmoc-Thr(OH)-OtBu 4. To prepare the glyco-amino acid M100, donor 2 was chemically glycosylated to the glycosyl acceptor 1 at the C-2 position to obtain the disaccharide 5 in 85% yield. An N-phenyltrifluoroacetimidate donor was used instead of the typical trichloroacetimidate, as the latter can generate over 50% of a glucosamine thioether by-product through thioether migration. Typical deprotection and protection strategies were employed to convert the benzylidene acetal, 4-methoxybenzyl ether (PMB), and trichloroethoxycarbonyl (Troc) protecting groups on 5 to acetyl (Ac) groups to afford 6 over three steps, with a total yield of 80%. The obtained thiophenyl glycosyl donor 6 was coupled to the protected threonine amino acid 4 at 0 °C in the presence of N-iodosuccinimide (NIS) and AgOTf to obtain the protected glyco-amino acid derivative 7 in excellent yield. The t-butyl (tBu) and Ac groups were then removed successively, followed by reintroduction of the partially cleaved
fluorenylmethyloxycarbonyl (Fmoc) group to provide the final product 8 (M100) in gram-amounts with a total yield of 90% over three steps.

Figure 1.4 (A) The four building blocks; (B) The synthetic scheme for the assembly of the three O-mannosyl scaffold structures.

Reagents and conditions: (a) TMSOTf, DCM, -60 °C, 5: 85%; 10: 79%; 14: 75%; 16: 82%; (b) (i) TFA, DCM; (ii) Zn, AcOH, 80% over three steps; (c) NIS, AgOTf, DCM/Et2O=1:1, 0 °C, 7: 93%; 12: 89%; 18: 83%; (d) (i) TFA, DCM; (ii) NaOMe, MeOH; (iii) FmocOSu, NaHCO3, Acetone/H2O=3:1, 8: 90%; 13: 85%; 19: 87%; (e) (i) DDQ, DCM/PBS Buffer=9:1; (ii) Ac2O, Py; (iii) EtSH, TsOH, DCM, 9: 77%; 15: 81%; (f) (i) Zn, AcOH; (ii) Ac2O, Py, 11: 87%; 17: 82%.
To synthesize **M301**, the PMB group on the disaccharide 5 was initially converted to Ac, followed by deprotection of the benzylidene acetal to achieve the diol 9. Glycosylation of 9 with the glycosyl donor 3 yielded the protected target tetrasaccharide 10 in 79% yield. Successively, the Troc protecting groups were converted to Ac groups and provided compound 11, which was used as glycosyl donor for coupling with the amino acid 4 to obtain the tetrasaccharide 12 in 89% yield. Complete deprotections of tBu and Ac groups were performed under standard reaction conditions and reprotction of the partially Fmoc deprotected amine afforded gram-amounts of compound 13 (M301), in a total yield of 85% over three steps. Similarly, 19 (M201) was prepared, by coupling of disaccharide donor 3 to the C-2 position of acceptor 1 followed by coupling of 2 to the C-6 position., Then coupling of the obtained oligosaccharide thiosglycoside 18 with amino acid 4 gave the desired glycosylated-amino acid 19 in a total yield of 29% over 11 steps.

**1.2.2 Enzymatic extension of scaffold structures**

To note the Fmoc was reintroduced to facilitate product detection and purification, as Fmoc has both UV (260 nm) and fluorescence (Ex. 260 nm, Em. 310 nm) absorbance. The bulky hydrophobic group also enhanced retention and separation of hydrophilic O-mannosyl glycans on reversed phase chromatography. Herein, HPLC was used to detect, purify and quantify enzymatically extended O-mannosyl glycans.
Figure 1.5 Enzymatic extension of core M1 and core M2 structures.
Reagents and conditions: (a) NmLgtB, UDP-Gal, Mg\textsuperscript{2+}; (b) PmST1-M144D, NmCSS, Neu5Ac, CTP, Mg\textsuperscript{2+}; (c) Pd26ST, NmCSS, Neu5Ac, CTP, Mg\textsuperscript{2+}; (d) Hp3FT, GDP-Fuc, Mg\textsuperscript{2+}; (e) PmST1-M144D, NmCSS, Neu5Gc, CTP, Mg\textsuperscript{2+}; (f) Pd26ST, NmCSS, Neu5Gc, CTP, Mg\textsuperscript{2+}; (g) GlcAT-P, UDP-GlcA, Mg\textsuperscript{2+}; (h) PmST1m, NmCSS, Neu5Ac, CTP, Mg\textsuperscript{2+}; (i) βGalD.

Four robust bacterial GTs and a human β1,3-glucuronyltransferase (GlcAT-P) were used to extend the scaffold M100 to eight core M1 O-mannosyl glycans (Fig 1.4A). The bacterial GTs include, β1,4-galactosyltransferase from Neisseria meningitidis (NmLgtB),\textsuperscript{[24]} mutant M144D of α2,3-sialyltransferase 1 from Pasteurella multocida (PmST1-M144D) with decreased donor hydrolysis plus reduced sialidase activity,\textsuperscript{[25]} α2,6-sialyltransferase from Photobacterium damselae (Pd26ST),\textsuperscript{[26]} and C-terminal 66 amino acids truncated α1,3-fucosyltransferase from Helicobacter pylori (Hp3FT) with increased solubility.\textsuperscript{[27]} These GTs are all highly active and recognize minimal motifs (e.g., GlcNAc for NmLgtB, LN for Pd26ST and Hp3FT, LN or Le\textsuperscript{X} for
PmST1-M144D) on oligosaccharide acceptors, and were previously applied for preparing complex glycans.\(^{[28]}\)

As illustrated in Figure 1.4, M101 was prepared by the NmLgtB-catalyzed reaction, containing M100 (10 mM), uridine 5'-diphosphogalactose (UDP-Gal) (15 mM), MgCl\(_2\) (10 mM), and an appropriate amount of NmLgtB. After overnight incubation, an \(m/z\) peak of 867.2972 was observed on the ESI mass spectrum, corresponding to M101 [M-H]\(^-\). Meanwhile, on the HPLC-UV\(_{260\text{ nm}}\) profile, a new peak (\(T_R = 21.71\) min) was observed, of which the area underneath increased while the peak corresponding to M100 (\(T_R = 23.32\) min) became smaller over time. After over 90\% conversion (monitored by HPLC-UV\(_{260\text{ nm}}\) with a 4.6 × 250 mm Inertsil ODS-4 column), M101 was isolated by HPLC using a semi-preparative column (Inertsil ODS-4, 20 × 250 mm). The purified M101 was then lyophilized, and further extended to afford M102 and M104, catalyzed by PmST1-M144D and Hp3FT, respectively (Fig 1.4). To make glycosylation reactions more efficient and less costly, the one-pot two-enzyme (OPTE)\(^{[29]}\) approach was adopted when adding sialic acid residues. For example, in the PmST1-M144D-catalyzed reaction to generate M102, Neu5Ac, cytidine 5'-triphosphate (CTP), MgCl\(_2\), and \(N.\) meningitidis CMP-sialic acid synthetase (NmCSS)\(^{[30]}\) were added to achieve in situ generation of the sugar donor CMP-Neu5Ac. Finally, M105 was prepared from M104 via the same OPTE approach. Employing a bacterial \(\alpha\)2,6-sialyltransferase (Pd26ST), we were also able to generate a not yet identified structure, M103, harboring the 6-sialyl-N-acetyllactosamine (Neu5Ac-\(\alpha\)2,6-Gal-\(\beta\)1,4-GlcNAc, 6SLN) epitope (Fig 4), which may serve as an ideal standard for mining such structures.

Core M1 glycans terminated with \(N\)-glycolylneuraminic acid (Neu5Gc) (e.g. 3S\(_{\text{Gc}}\)LN, Fig 1.1) were also observed on animal \(\alpha\)-DG.\(^{[4, 31]}\) Structures with a Neu5Gc residue (M102G and M103G) were thus prepared, using PmST1-M144D- and Pd26ST-catalyzed OPTE system similar
to the synthesis of **M102** and **M103**. The HNK-1 epitope is a unique sulfated trisaccharide that highly expressed in the nervous system and plays critical roles in neuronal plasticity and diseases.\[^{32}\] We optimized the codon and cloned the $\beta$-glucuronosyltransferase gene (GlcAT-P) for heterogeneous expression in *E. coli*. Even though pure soluble proteins were not obtained, we were able to synthesize a precursor structure of the HNK-1 epitope on core M1 (**Fig 1.4, M106**) using GlcAT-P-containing cell lysate. Research in obtaining a 3-sulfotransferase for generating HNK-1 epitope is undergoing.

Core M2 O-mannosyl glycans can be classified into symmetric (with same motifs on both $\beta1,2$- and $\beta1,6$-branches, e.g., **M0X0**) and asymmetric (with different motifs on the branches, e.g., **M2XX** and **M3XX**) structures. To generate symmetric structures, chemically prepared **M201** was firstly galactosylated by NmLgtB to afford **M010**. Similar as described for the synthesis of core M1 glycans, **M020**, **M030** and **M040** were then prepared starting from **M010** via reactions catalyzed by PmST1-M144D, Pd26ST, and Hp3FT, respectively. On the other hand, **M050** was prepared starting from **M040** via the PmST1-M144D-catalyzed reaction (**Fig 1.4B**). Specifically, **M000** was generated by treating **M201** with a $\beta$-galactosidase from *Streptococcus pneumoniae* ($\beta\text{GalD}$).\[^{31}\]

The synthesis of asymmetric core M2 O-mannosyl glycans was performed starting with **M201** or **M301**, by enzymatic extension of the Gal-containing branch first and then the other, in a
Figure 1.6 Enzymatic extension of M301 to generate core M2 O-mannosyl glycans. Reagents and conditions: (a) NmLgtB, UDP-Gal, Mg\(^{2+}\); (b) PmST1-M144D, NmCSS, Neu5Ac, CTP, Mg\(^{2+}\); (c) Pd26ST, NmCSS, Neu5Ac, CTP, Mg\(^{2+}\); (d) Hp3FT, GDP-Fuc, Mg\(^{2+}\); (h) PmST1m, NmCSS, Neu5Ac, CTP, Mg\(^{2+}\) strictly controlled sequential manner. The synthesis of M2XX is illustrated in Figure 1.4B. Such synthetic routes were designed according to substrate specificities of corresponding GTs to avoid undesirable glycosylation. With M234 as an example, the scaffold M201 was first fucosylated to form M204, followed by galactosylation on the β1,6-branch to provide M214. Such a synthetic sequence eliminates the addition of Fuc onto the β1,6-branch, as Hp3FT requires an LN disaccharide motif for its activity. Lastly, M234 was formed by the Pd26ST-catalyzed reaction, which attaches a Neu5Ac residue selectively onto the terminal Gal on the β1,6-branch, as α1,3-fucosylation prevents Pd26ST-catalyzed α2,6-sialylation on the β1,2-branch. The synthetic scheme of M3XX is shown in Figure 1.5. It should be noted that M224 and M324 were synthesized from M214 and M314 (Fig 1.4B, Fig 1.5) by using mutant E271F/R313Y of PmST1.
(PmST1m) instead of PmST1-M144D, which greatly prefers the LN disaccharide over the LeX trisaccharide,\textsuperscript{[34]} to avoid undesired sialylation. Collectively, all 36 possible combinations of core M2 glycans that harboring LN, 3SLN, 6SLN, LeX or sLeX motifs were prepared. These glycans were purified by HPLC, and characterized by ESI/MALDI-MS & NMR to confirm the structures.

1.2.3 Microarray analysis

The synthesized O-mannosyl glycans are well-defined and closely related glycoforms (Fig 1.2), providing unique probes for mining fine specificities of glycan-binding proteins (GBPs). As shown in Figure 1.6A, microarray analysis showed that both Ricinus communis lectin I (RCA-I) and Erythrina crystagalli lectin (ECA) strongly bound to M010, consistent with its primary specificity towards terminal LN epitope.\textsuperscript{[35]} Moreover, ECA exhibited a broader specificity towards all O-mannosyl glycans harboring a free terminal LN (M101, M201, M21X, M301, and M31X), whereas RCA-I seems to prefer terminal LN on the β1,6-branch (M301 and M21X) over the β1,2-branch (M101, M201, M314 and M215) (Fig 1.6A). Such a branch preference was also found for the anti-CD15s antibody (specific to sLeX epitope), which bound to glycans that contain sLeX on the β1,6-branch (M050, M3X5) but not to that with sLeX only on the β1,2-branch (M105, M2X5) (Fig 1.6B). On the other hand, Aleuria aurantia lectin (AAL, specific to α-Fuc) exhibited a preference towards the β1,2-branch, as well as other fine specificities (Fig 1.7).

ConA, an α-Man specific lectin commonly used for enriching tryptic O-mannosyl peptides, strongly bound to M100 but not to other natural core M1 or any core M2 structures. It thus can be speculated that detection or enrichment of O-mannosyl glycans using ConA may miss a substantial amount of complex structures (Fig 1.8). Interestingly, weak bindings of ConA to unnatural α2,6-sialylated core M1 glycans (M103, M103G) was observed (Fig 1.8), suggesting that such modification may result in conformational changes that facilitated ConA binding.
Anti-glycan antibodies, on the other hand, may present as better detection tools. For example, anti-serum from two rabbits (26559, 26560) that immunized with a core M2 glycan (M000) conjugate bound to a broad range of core M2-containing glycopeptides regardless of varied peptide sequences.\textsuperscript{[21]} We further evaluated the anti-sera toward synthesized O-mannosyl glycans. Our results showed that both sera exhibit binding specificities towards core M2 O-mannosyl glycans that containing at least one free terminal GlcNAc residue (M000, M20X, M30X) as well as the core M1 disaccharide M100 (\textbf{Fig 1.6C, Fig 1.9}). Moreover, antisera from rabbit 26560 exhibited comparable bindings to all those glycans, whereas anti-sera from rabbit 26559 showed stronger binding to glycans contain a free GlcNAc on the β1,6-branch (M20X)
compared with binding to the β1,2-branch (M100, M30X) (Fig 1.9). The results imply that antibodies generated from different hosts may possess certain individual heterogeneity (or individual difference). Nevertheless, our results revealed fine specificities of the anti-sera towards O-mannosyl glycans. Comparing with ConA, these anti-sera are advantageous in the detection of branched O-mannosyl glycans.

1.3 Conclusion

In summary, by combining convergent chemical synthesis with a strictly programmed enzymatic synthesis in a stepwise manner, an efficient scaffold synthesis/ enzymatic extension (SSEE) strategy was developed to access 45 mammalian O-mannosyl glycans. Such unique glycoforms provide not only standards for identifying O-mannosyl glycans and revealing their biological roles, but also ideal probes for mining fine details of protein-glycan interactions.

1.4 Experiment Section

1.4.1 Chemical synthesis

All chemicals were purchased as reagent grade and used without further purification. Anhydrous dichloromethane (CH₂Cl₂), acetonitrile (CH₃CN), tetrahydrofuran (THF), N,N-dimethyl formamide (DMF), diethyl ether (Et₂O), toluene, and methanol (MeOH) were purchased from a commercial source without further distillation. Pulverized Molecular Sieves MS-4 Å (Aldrich) for glycosylation was activated by heating at 350 °C for 3 h. Reactions were monitored by analytical thin-layer chromatography (TLC) in EM silica gel 60 F254 plates and visualized under UV (254 nm) and/or by staining with acidic ceric ammonium molybdate or p-anisadehyde. Flash chromatography was performed on silica gel (Merck) of 40-63μm particle size and P2 gel (Biorad). ¹H and ¹³C NMR spectra were recorded on a Bruker AVANCE 400 (400 MHz), and
Bruker AVANCE 600 (600 MHz) spectrometer at 25 °C. All $^1$H Chemical shifts (in ppm) were assigned according to CDCl$_3$ (δ= 7.24 ppm) and D$_2$O (δ = 4.79 ppm) and all $^{13}$C NMR was calibrated with CDCl$_3$ (δ = 77.00 ppm). Coupling constants (J) are reported in hertz (Hz). Splitting patterns are described using the following abbreviations: s, singlet; brs, broad singlet; d, doublet; t, triplet; q, quartet; dd, doublet of doublet; m, multiplet. $^1$H NMR spectra are reported in the following order: chemical shift, multiplicity, coupling constant(s), and number(s) of protons. All NMR signals were assigned on the basis of $^1$H NMR, COSY, HSQC, HMQC, and $^{13}$C NMR experiments. HPLC-MS experiments were performed on an LTQ-Orbitrap Elite mass spectrometer (Thermo Fisher) equipped with EASY-spray source and nano-LC UltiMate 3000 high-performance liquid chromatography system (Thermo Fisher). Samples were transmitted into MS with a silica column. LTQ-Orbitrap Elite mass spectrometer was operated in the data-dependent mode. A full-scan survey MS experiment (m/z range was set according to the molecular weight of O-mannose glycan; automatic gain control target, 1,000,000 ions; resolution at 400 m/z, 240,000; maximum ion accumulation time, 200 ms) was acquired by the Orbitrap mass spectrometer. MALDI-TOF MS analyses were performed on UltrafleXtreme MALDI TOF/TOF Mass Spectrometer (Bruker). Scan range of MS was set according to the molecular weight of O-mannose glycans, and reflector mode was used for O-mannose glycan analysis. Mass spectra were obtained in both positive and negative extraction mode with the following voltage settings: ion source 1 (19.0 kV), ion source 2 (15.9 kV), and lens (9.3 kV). The reflector voltage was set to 20 kV. The laser was pulsed at 7 Hz and the pulsed ion extraction time was set at 400 ns. The laser power was kept in the range of 40–90%. Fmoc-Thr(OH)-OtBu (4) was purchased from Sigma-Aldrich.
**Glycosylation of N-phenyltrifluoroacetimidate donor (A)**

A mixture of donor (1.2 mmol), acceptor (1 mmol) and 4 Å molecular sieves in dry CH$_2$Cl$_2$ was stirred at room temperature under argon for 1 h. TMSOTf (0.2 mmol) was added at -60 °C. The reaction mixture was stirred at -60 °C for 1 h before it was quenched with a few drops of triethylamine. The resulting mixture was filtered. The filtrate was concentrated *in vacuo* and purified on a silica gel column to produce the product.

**Glycosylation of thioether donor procedure (B)**

A mixture of thioether donor (1 mmol), amino acid acceptor 4 (1.5-2 mmol) and 4 Å molecular sieves in dry Et$_2$O/CH$_2$Cl$_2$ (1:1, v/v) were stirred at room temperature under argon for 1 h. NIS (1.5 mmol) and AgOTf (0.2 mmol) were added at 0 °C. The reaction mixture was stirred for 10 h before it was quenched with a few drops of triethylamine. The resulting mixture was filtered. The filtrate was diluted with CH$_2$Cl$_2$ and washed with 5% aqueous Na$_2$S$_2$O$_3$, saturated aqueous NaHCO$_3$, brine, dried over Na$_2$SO$_4$, and concentrated *in vacuo*. The residue was purified on a silica gel column to produce the product.

**Global deprotection of Ac and tBu and reintroducing of Fmoc (C)**

Oligosaccharide glycosyl amino acid derivative (1 mmol) was dissolved in TFA/CH$_2$Cl$_2$ (1:1, v/v) and stirred at room temperature under argon for 4 h. The mixture was concentrated *in vacuo* and the residue was dissolved in MeOH, and NaOMe in MeOH was added until pH was 10. After stirring at room temperature for 2 h, the solution was neutralized with ion-exchange resin (H+), then filtered and concentrated *in vacuo*. The crude product, NaHCO$_3$ (4 mmol) and 9-fluorenymethyl-Nsuccimidylcarbonate (3 mmol) were dissolved in H$_2$O/acetone (1:1, v/v) and
this mixture was stirred at room temperature. After 2 h, the mixture was concentrated \textit{in vacuo} and purified on a silica gel column to afford the product.

\textbf{Transformation of PMB to Ac and cleavage of benzylidene (D)}

A solution of oligosaccharide (1 mmol) in a mixture of CH$_2$Cl$_2$/H$_2$O (30:1, v/v) was treated with DDQ (1.2 mmol) in ice bath and stirred at 25 °C for 3 h. Triethylamine was added and the solvent was removed \textit{in vacuo}. The residue was dissolved in CH$_2$Cl$_2$ was cooled down to 0 °C, followed by slow addition of acetic anhydride (3 mmol) and TEA (5 mmol). The reaction mixture was stirred at room temperature overnight and concentrated \textit{in vacuo}. The crude product was dissolved in anhydrous MeOH, followed by addition of TsOH (0.1 mmol) and EtSH (6 mmol). The reaction mixture was stirred at rt. for 6 h and then quenched with triethylamine and concentrated \textit{in vacuo}. The mixture was purified with silica column to get the product.

\textbf{Transformation of NHTroc to NHAc (E)}

N-Troc protected oligosaccharide (1 mmol) was dissolved in AcOH at room temperature, followed by addition of Zn dust (<10 micron, 10 mmol). After being stirred at 40 °C for 24 h, the mixture was concentrated \textit{in vacuo} to give a residue for the next step without further purification. To a solution of the residue in pyridine was added Ac$_2$O. After being stirred at room temperature for 12 h, the solution was diluted with ethyl acetate and washed with aqueous HCl (1 M), saturated aqueous NaHCO$_3$, and brine solution. The organic layer was dried over Na$_2$SO$_4$, filtered, and concentrated \textit{in vacuo} to give a residue, which was purified by silica gel column chromatography to generate NHAc containing compound.
Preparation of the versatile precursor 1.

4,6-O-benzylidene-1-thio-α-D-Mannopyranoside\textsuperscript{[36]} (13.40 g, 37.2 mmol), dibutyltin oxide (10.23 g, 40.9 mmol) and toluene (250 mL) were refluxed for 3 h, followed by concentration \textit{in vacuo}. The residue, tetrabutylammonium iodide (15.09 g, 40.9 mmol), 4-Methoxybenzyl chloride (6.39 mL, 40.9 mmol), and toluene (250 mL) were stirred at 70 °C overnight. Toluene was removed by concentration \textit{in vacuo} and the residue was diluted with ethyl acetate (300 mL). The resulting solution was washed with water and saturated aqueous NaHCO\textsubscript{3}, dried over Na\textsubscript{2}SO\textsubscript{4}, filtered, and concentrated \textit{in vacuo}. The resulting residue was purified by silica gel column chromatography (hexane/ethyl acetate, 4:1, v/v) to afford precursor 1 (16.26 g, 91%) as white foams. R\textsubscript{f} = 0.28 (hexane/ethyl acetate, 4:1, v/v); \textsuperscript{1}H NMR (400 MHz, CDCl\textsubscript{3}) δ 7.51 (dd, J = 7.6, 1.8 Hz, 2H), 7.46 – 7.34 (m, 5H), 7.35 – 7.25 (m, 5H), 6.88 (d, J = 8.7 Hz, 2H), 5.61 (s, 1H), 5.57 (s, 1H), 4.81 (d, J = 11.4 Hz, 1H), 4.66 (d, J = 11.4 Hz, 1H), 4.33 (td, J = 9.8, 4.9 Hz, 1H), 4.25 – 4.11 (m, 3H), 3.93 (dd, J = 9.5, 3.4 Hz, 1H), 3.84 (t, J = 10.3 Hz, 1H), 3.80 (s, 3H), 2.95 (d, J = 1.2 Hz, 1H); \textsuperscript{13}C NMR (100 MHz, CDCl\textsubscript{3}) δ 159.47, 137.42, 133.27, 131.65, 129.74, 129.61, 129.10, 128.94, 128.20, 127.63, 126.05, 113.90, 101.57, 87.76, 78.92, 75.35, 72.87, 71.32, 68.48, 64.57, 55.23; MALDI-MS: [M+Na]\textsuperscript{+} C\textsubscript{27}H\textsubscript{28}O\textsubscript{6}SNa calculated for 503.1504, found 503.1515.

Preparation of N-phenyltrifluoroacetimidate donor 2.
1,3,4,6-tri-O-acetyl-2-deoxy-2-(2,2,2-trichloroethoxycarbonylamino)-β-D-glucopyranose (5.70 g, 10.9 mmol) was dissolved in THF (100 mL) followed by the addition of benzylamine (2.10 mL, 19.6 mmol). After the reaction mixture was stirred at room temperature for 8 h under dry atmosphere, it was evaporated and the residue was diluted with ethyl acetate and washed successively with 1 M aqueous HCl, saturated aqueous NaHCO₃, and brine. The organic layer was dried over MgSO₄, filtrated and concentrated under high vacuum. The crude product was dissolved in CH₂Cl₂ (200 mL), then 2,2,2-Trifluoro-N-phenylacetimidoyl Chloride (3.42 mL, 21.8 mmol) and 1,8-diazabicyclo-[5,4,0]-7-undecene (2.44 mL, 16.4 mmol) were added at 0 °C under dry atmosphere. After the reaction mixture was stirred at room temperature for 2 h, it was evaporated followed by purification by silica gel column chromatography (hexane/ethyl acetate 4/1) to afford imidate donor 2 (5.50 g, 77% over two steps) as a white foams, which were found unstable and used immediately in the subsequent glycosylation without further characterization.

Preparation of N-phenyltrifluoroacetimidate donor 3.

Phenyl 2,3,4,6-tetra-O-acetyl-β-D-galactopyranosyl-(1→4)-3,6-di-O-acetyl-2-deoxy-2-(2,2,2-trichloroethoxycarbonylamino)-β-D-glucopyranoside[37] (5.10 g, 5.9 mmol) was dissolved in Actone/H₂O (10:1, v/v, 60 mL) at -30 °C and N-Bromosuccinimide (5.25 g, 29.5 mmol) was added to the solution. After stirred at -30 °C for 2 h, the reaction was quenched with 5% aqueous Na₂S₂O₃ (10 mL) and saturated aqueous NaHCO₃ (10 mL). The mixture was diluted with ethyl acetate and washed with brine. The organic layer was dried over NaSO₄, filtered, and concentrated. The residue was dissolved in CH₂Cl₂ (100 mL), then 2,2,2-Trifluoro-N-phenylacetimidoyl
Chloride (1.85 mL, 11.8 mmol) and 1,8-diazabicyclo-[5,4,0]-7-undecene (1.32 mL, 8.9 mmol) were added at 0 °C under dry atmosphere. After the reaction mixture was stirred at room temperature for 2 h, it was evaporated followed by purification by silica gel column chromatography (hexane/aceton 4/1) to afford imidate donor 3 (4.29 g, 75% over two steps) as a white foams, which were found instable and used immediately in the subsequent glycosylation without further characterization.

![Chemical structure](image)

**Phenyl 3,4,6-tri-O-acetyl-2-deoxy-2-(2,2,2-trichloroethoxycarbonylamino)-β-D-glucopyranosyl-(1→2)-4,6-O-benzylidene-3-(4-Methoxybenzyl)-1-thio-α-D-Mannopyranoside (5)**

Compound 1 (3.37 g, 7.0 mmol) was glycosylated with fresh-made donor 2 (5.50 g, 8.43 mmol) by following general procedure A to get the desired compound 5 (5.63 g, 85%) as white foam. R_f = 0.30 (hexane/ethyl acetate, 3:1, v/v); ^1^H NMR (400 MHz, CDCl_3) δ 7.60 – 7.49 (m, 2H), 7.49 – 7.37 (m, 5H), 7.37 – 7.30 (m, 5H), 6.94 – 6.84 (m, 2H), 5.64 (s, 1H), 5.54 (dd, J = 10.7, 9.3 Hz, 1H), 5.48 (d, J = 1.4 Hz, 1H), 5.37 – 5.30 (m, 1H), 5.14 – 5.01 (m, 2H), 4.78 (d, J = 11.2 Hz, 1H), 4.74 – 4.63 (m, 2H), 4.52 (d, J = 12.5 Hz, 1H), 4.40 (dd, J = 3.3, 1.6 Hz, 1H), 4.35 – 4.11 (m, 5H), 3.96 (dd, J = 9.8, 3.2 Hz, 1H), 3.88 – 3.78 (m, 4H), 3.77 – 3.70 (m, 1H), 3.44 (dt, J = 10.9, 8.1 Hz, 1H), 2.05 (s, 3H), 2.04 (s, 3H), 2.02 (s, 3H); ^13^C NMR (100 MHz, CDCl_3) δ 170.63, 170.27, 169.58, 159.40, 137.50, 133.42, 131.72, 129.87, 129.66, 129.23, 128.94, 128.25, 127.91, 126.06, 113.87, 101.55, 87.40, 78.75, 76.40, 74.40, 72.49, 72.09, 68.80, 68.49, 65.32,
62.16, 60.42, 55.26, 21.06, 20.71, 20.65, 20.58; MALDI-MS: [M+Na]+ C_{42}H_{46}Cl_{3}NO_{15}SNa calculated for 964.1552, found 964.1538.

Phenyl 2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-β-D-glucopyranosyl-(1→2)-3,4,6-tri-O-acetyl-1-thio-α-D-Mannopyranoside (6)

Compound 5 (2.83 g, 3.0 mmol) was dissolved in TFA/CH$_2$Cl$_2$ (10:1, v/v, 50 mL) and stirred at room temperature under argon for 1 h. CH$_2$Cl$_2$ (100 mL) was added to dilute the reaction and then neutralized with saturated aqueous NaHCO$_3$. The organic layer was separated and washed with brine. After dried over anhydrous Na$_2$SO$_4$ and concentrated, the residue was treated by following general procedure E to afford compound 6 (1.74 g, 80%) as white foam. R$_f$ = 0.30 (hexane/acetone, 3:1, v/v); $^1$H NMR (400 MHz, CDCl$_3$) δ 7.58 – 7.42 (m, 2H), 7.41 – 7.24 (m, 3H), 5.73 (d, J = 7.9 Hz, 1H), 5.51 (t, J = 9.9 Hz, 1H), 5.45 (s, 1H), 5.30 (t, J = 10.0 Hz, 1H), 5.12 – 4.95 (m, 2H), 4.53 – 4.41 (m, 2H), 4.33 – 4.20 (m, 2H), 4.11 – 3.97 (m, 2H), 3.72 (dd, J = 10.2, 5.1 Hz, 1H), 3.59 (dt, J = 11.5, 8.1 Hz, 1H), 2.11 – 2.07 (m, 6H), 2.05 (s, 3H), 2.03 (s, 3H), 2.02 – 1.99 (m, 6H), 1.86 (s, 3H); $^{13}$C NMR (100 MHz, CDCl$_3$) δ 170.78, 170.67, 170.62, 170.50, 170.35, 169.56, 169.51, 133.14, 131.55, 129.18, 127.87, 98.40, 85.07, 75.47, 71.84, 71.47, 70.46, 69.53, 68.83, 66.20, 62.71, 62.05, 55.63, 23.19, 20.75, 20.72, 20.70, 20.67, 20.63; [M+Na]$^+$ C$_{32}$H$_{41}$NO$_{16}$SNa calculated for 750.2044, found 750.2032.
Compound 6 (1.74 g, 2.4 mmol) was glycosylated with compound 4 (1.43 g, 3.6 mmol) by following general procedure B to get the desired compound 7 (2.27 g, 93%) as white foam. R_f = 0.29 (hexane/acetone, 3:1, v/v); ¹H NMR (400 MHz, CDCl₃) δ 7.78 (d, J = 7.4 Hz, 2H), 7.72 – 7.54 (m, 2H), 7.47 – 7.29 (m, 4H), 5.76 (d, J = 7.9 Hz, 1H), 5.56 (d, J = 9.0 Hz, 1H), 5.43 (t, J = 9.8 Hz, 1H), 5.23 (t, J = 9.7 Hz, 1H), 5.10 – 4.94 (m, 2H), 4.91 – 4.80 (m, 2H), 4.58 – 4.14 (m, 8H), 4.14 – 4.05 (m, 2H), 4.05 – 3.88 (m, 2H), 3.72 – 3.51 (m, 2H), 2.08 (s, 6H), 2.07 (s, 3H), 2.04 (s, 3H), 2.03 (s, 3H), 2.02 (s, 3H), 1.95 (s, 3H), 1.53 (s, 9H), 1.32 (d, J = 5.6 Hz, 3H); ¹³C NMR (100 MHz, CDCl₃) δ 170.65, 170.60, 170.56, 169.47, 169.32, 156.52, 143.84, 143.74, 141.31, 127.78, 127.11, 125.18, 120.04, 99.00, 98.89, 82.67, 74.15, 71.86, 71.65, 69.95, 69.21, 68.79, 67.36, 66.09, 62.89, 62.10, 58.92, 55.31, 47.17, 28.03, 23.24, 20.75, 20.71, 20.64, 17.88; [M+Na]^+ C₄₉H₆₂N₂O₂₁Na calculated for 1037.3743, found 1037.3752.
**Nα-9-Fluorenylmethyloxycarbonyl-O-[2-acetamido-2-deoxy-β-D-glucopyranosyl-(1→2)-α-D-mannopyranosyl]-L-threonine (8, M100)**

Compound 7 (2.13 g, 2.1 mmol) was treated by following general procedure C to afford compound 8 (1.34 g, 90%) as white foam. \( R_f = 0.40 \) (ethyl acetate/methanol/H\( \text{H}_2\text{O} \), 4:2:0.5, v/v); \(^1\)H NMR (400 MHz, D\( \text{D}_2\text{O} \)) δ 7.74 – 7.60 (m, 2H), 7.60 – 7.40 (m, 2H), 7.40 – 7.18 (m, 4H), 4.60 (dd, \( J = 10.7, 4.9 \) Hz, 1H), 4.42 – 4.27 (m, 2H), 4.23 (d, \( J = 6.3 \) Hz, 1H), 4.12 – 3.96 (m, 1H), 3.83 – 3.71 (m, 4H), 3.71 – 3.42 (m, 7H), 3.42 – 3.30 (m, 3H), 3.27 (s, 1H), 1.96 (s, 3H), 0.94 (d, \( J = 6.1 \) Hz, 3H); \(^{13}\)C NMR (100 MHz, D\( \text{D}_2\text{O} \)) δ 174.87, 158.05, 143.91, 143.49, 140.96, 140.91, 127.96, 127.41, 124.90, 124.78, 120.09, 100.01, 98.04, 77.19, 75.59, 73.28, 73.05, 69.84, 69.26, 67.22, 65.92, 61.60, 60.52, 55.26, 47.10, 22.35, 18.33; [M-H] \(+\) C\(_{33}\)H\(_{41}\)N\(_2\)O\(_{15}\) calculated for 705.2585, found 705.2470.

![Structural formula of 8](image)

**Phenyl 3,4,6-tri-O-acetyl-2-deoxy-2-(2,2,2-trichloroethoxycarbonylamino)-β-D-glucopyranosyl-(1→2)-3-O-acetyl-1-thio-α-D-Mannopyranoside (9)**

Compound 5 (3.58 g, 3.8 mmol) was treated by following general procedure C to afford compound 9 (2.27 g, 77%) as white foam. \( R_f = 0.33 \) (hexane/acetone, 4:1, v/v); \(^1\)H NMR (400 MHz, CDCl\(_3\)) δ 7.42 (d, \( J = 6.5 \) Hz, 2H), 7.36 – 7.19 (m, 3H), 6.00 (d, \( J = 8.2 \) Hz, 1H), 5.35 (s, 1H), 5.25 (t, \( J = 9.9 \) Hz, 1H), 4.99 (t, \( J = 9.5 \) Hz, 1H), 4.88 (d, \( J = 9.6 \) Hz, 1H), 4.71 – 4.56 (m, 2H), 4.52 – 4.37 (m, 2H), 4.33 – 4.15 (m, 2H), 4.15 – 3.97 (m, 2H), 3.87 (t, \( J = 10.5 \) Hz, 1H), 3.80 – 3.45 (m, 4H), 2.93 (d, \( J = 8.7 \) Hz, 1H), 2.10 (s, 3H), 2.08 (s, 3H), 2.00 (s, 3H), 1.98 (s, 3H); \(^{13}\)C NMR (100 MHz, CDCl\(_3\)) δ 170.99, 170.89, 170.77, 169.52, 154.57, 133.42, 131.69, 129.27,
127.94, 100.41, 95.00, 85.81, 76.94, 74.56, 73.42, 73.20, 71.77, 71.35, 68.58, 64.12, 62.12, 61.50, 56.05, 20.91, 20.68, 20.59; [M+Na]$^+$ C$_{29}$H$_{36}$Cl$_3$NO$_{15}$SNa calculated for 798.0769, found 798.0778.

Phenyl 3,4,6-tri-O-acetyl-2-deoxy-2-(2,2,2-trichloroethoxycarbonylamino)-β-glucopyranosyl-(1→2)-[2,3,4,6-tetra-O-acetyl-β-D-galactopyranosyl-(1→4)-3,6-di-O-acetyl-2-deoxy-2-(2,2,2-trichloroethoxycarbonylamino)-β-D-glucopyranosyl]-(1→6)-3-O-acetyl-1-thio-α-D-mannopyranoside (10)

Compound 9 (2.10 g, 2.7 mmol) was glycosylated with fresh-made donor 3 (3.05 g, 3.2 mmol) by following general procedure A to get the desired compound 10 (3.26 g, 79%) as white foam. $R_f$ = 0.35 (hexane/acetone, 2:1, v/v); $^1$H NMR (400 MHz, CDCl$_3$) δ 7.43 (d, $J$ = 6.7 Hz, 2H), 7.39 – 7.24 (m, 3H), 6.32 (d, $J$ = 8.4 Hz, 1H), 6.05 (d, $J$ = 8.4 Hz, 1H), 5.45 (s, 1H), 5.43 – 5.31 (m, 3H), 5.09 (dd, $J$ = 10.3, 7.9 Hz, 1H), 5.05 – 4.91 (m, 3H), 4.86 (dd, $J$ = 9.8, 2.6 Hz, 1H), 4.78 (d, $J$ = 8.2 Hz, 1H), 4.76 – 4.62 (m, 3H), 4.59 – 4.44 (m, 3H), 4.41 (s, 1H), 4.26 (dd, $J$ = 12.3, 5.4 Hz, 1H), 4.20 – 4.00 (m, 6H), 3.96 (d, $J$ = 10.9 Hz, 1H), 3.92 – 3.81 (m, 2H), 3.79 – 3.65 (m, 2H), 3.61 (dd, $J$ = 8.7, 3.8 Hz, 1H), 3.49 (dq, $J$ = 17.4, 8.7 Hz, 2H), 3.00 (s, 1H), 2.15 (s, 3H), 2.13 (s, 6H), 2.09 (s, 3H), 2.08 (s, 3H), 2.06 (s, 3H), 2.04 (s, 3H), 2.01 (s, 3H), 1.97 (s, 3H); $^{13}$C NMR (100 MHz, CDCl$_3$) δ 170.77, 170.68, 170.50, 170.40, 170.19, 170.12, 169.52, 169.11, 155.08, 154.61, 133.76, 130.97, 129.19, 127.63, 101.24, 100.97, 99.16, 95.48, 95.35, 85.16, 76.47, 74.73, 74.52, 73.30, 72.72, 72.47, 71.97, 71.14, 70.99, 70.74, 69.15, 68.88, 67.77, 66.67, 63.87, 62.07, 60.93, 56.75, 56.14, 21.03, 20.94, 20.88, 20.72, 20.69, 20.63, 20.60, 20.50; [M+Na]$^+$ C$_{56}$H$_{70}$Cl$_6$N$_2$O$_{32}$SNa calculated for 1547.1661, found 1547.1603.
Phenyl 2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-β-glucopyranosyl-(1→2)-[2,3,4,6-tetra-O-acetyl-β-D-galactopyranosyl-(1→4)-2-acetamido-3,6-di-O-acetyl-2-deoxy-β-D-glucopyranosyl-(1→6)]-3,4-di-O-acetyl-1-thio-α-D-mannopyranoside (11)

Compound 10 (3.10 g, 2.0 mmol) was treated by following general procedure C to afford compound 11 (2.30 g, 87%) as white foam. Rf = 0.22 (hexane/acetone, 3:2, v/v); 1H NMR (400 MHz, CDCl3) δ 7.46 – 7.36 (m, 2H), 7.35 – 7.26 (m, 3H), 7.26 – 7.20 (m, 1H), 6.46 (d, J = 10.0 Hz, 1H), 5.89 (dd, J = 10.7, 9.2 Hz, 1H), 5.68 (d, J = 1.1 Hz, 1H), 5.50 – 5.37 (m, 2H), 5.34 (d, J = 2.8 Hz, 1H), 5.14 – 4.99 (m, 3H), 4.99 – 4.82 (m, 2H), 4.54 – 4.36 (m, 3H), 4.26 (dd, J = 12.2, 5.9 Hz, 1H), 4.22 – 4.11 (m, 3H), 4.11 – 4.00 (m, 4H), 3.95 (dd, J = 12.1, 2.2 Hz, 1H), 3.85 (t, J = 7.1 Hz, 1H), 3.80 – 3.66 (m, 2H), 3.53 (ddd, J = 9.7, 5.0, 1.8 Hz, 1H), 3.15 (d, J = 10.8 Hz, 1H), 3.02 – 2.89 (m, 1H), 2.14 (s, 3H), 2.11 (s, 3H), 2.08 (s, 6H), 2.07 (s, 3H), 2.05 (s, 3H), 2.04 (d, J = 3.8 Hz, 3H), 2.02 (d, J = 4.8 Hz, 6H), 1.99 (s, 3H), 1.98 (s, 3H), 1.95 (s, 3H), 1.88 (s, 3H); 13C NMR (100 MHz, CDCl3) δ 172.69, 172.18, 170.86, 170.60, 170.31, 170.29, 170.08, 170.00, 169.93, 169.78, 169.08, 133.09, 130.04, 129.24, 127.24, 102.96, 101.07, 96.29, 83.45, 77.41, 77.09, 76.78, 76.15, 74.31, 72.83, 72.75, 71.88, 70.88, 70.62, 70.55, 70.43, 69.98, 69.53, 69.15, 68.44, 66.51, 65.77, 62.35, 62.07, 60.57, 56.92, 53.49, 23.28, 23.25, 21.02, 20.89, 20.80, 20.67, 20.64, 20.61, 20.58, 20.52, 20.48; [M+Na]+ C56H74N2O31SNa calculated for 1325.3894, found 1325.3935.
\[ N^\alpha-9\text{-Fluorenylmethyloxycarbonyl-}O\{-2\text{-acetamido-}3,4,6\text{-tri-O-acetyl-}2\text{-deoxy-}\beta\text{-glucopyranosyl-(1→2)-}[2,3,4,6\text{-tetra-O-acetyl-}\beta\text{-D-galactopyranosyl-(1→4)-}2\text{-acetamido-}3,6\text{-di-O-acetyl-}2\text{-deoxy-}\beta\text{-D-glucopyranosyl]-(1→6)-}3,4\text{-di-O-acetyl-}\alpha\text{-D-mannopyranosyl}-L\text{-threonine tertbutyl ester (12)} \]

Compound 11 (2.00 g, 1.5 mmol) was glycosylated with compound 4 (1.19 g, 3.0 mmol) by following general procedure B to get the desired compound 12 (2.12 g, 89%) as white foam. \( R_f = 0.24 \) (hexane/acetone, 3:2, v/v); \(^1\)H NMR (600 MHz, CDCl\(_3\)) \( \delta \) 7.80 – 7.74 (m, 2H), 7.69 – 7.52 (m, 2H), 7.44 – 7.37 (m, 2H), 7.37 – 7.30 (m, 3H), 6.44 (d, \( J = 10.0 \) Hz, 1H), 5.95 – 5.85 (m, 1H), 5.60 (d, \( J = 9.3 \) Hz, 1H), 5.43 – 5.31 (m, 3H), 5.17 – 5.01 (m, 3H), 4.96 (dd, \( J = 10.4, 3.4 \) Hz, 1H), 4.94 – 4.83 (m, 2H), 4.54 – 4.42 (m, 3H), 4.42 – 4.23 (m, 5H), 4.23 – 4.02 (m, 7H), 3.97 – 3.83 (m, 2H), 3.83 – 3.70 (m, 2H), 3.70 – 3.60 (m, 1H), 3.60 – 3.52 (m, 1H), 3.14 (d, \( J = 11.3 \) Hz, 1H), 2.91 (dd, \( J = 17.8, 7.9 \) Hz, 1H), 2.16 (s, 3H), 2.12 (s, 3H), 2.11 (s, 3H), 2.09 (s, 3H), 2.09 (s, 3H), 2.07 (t, \( J = 2.7 \) Hz, 6H), 2.05 (s, 3H), 2.04 (s, 3H), 2.02 (s, 3H), 1.99 (s, 3H), 1.97 (s, 3H), 1.95 (s, 3H), 1.52 (s, 9H), 1.31 – 1.24 (m, 3H); \(^{13}\)C NMR (150 MHz, CDCl\(_3\)) \( \delta \) 172.57, 172.20, 170.87, 170.63, 170.32, 170.11, 170.04, 169.97, 169.93, 169.89, 169.10, 169.03, 156.59, 143.95, 143.79, 141.29, 127.71, 127.11, 127.08, 125.25, 119.98, 119.95, 102.90, 101.11, 98.35, 96.47, 82.46, 78.01, 76.15, 72.90, 72.80, 72.71, 71.80, 70.91, 70.64, 70.62, 70.07, 69.85, 69.17, 69.09, 68.32, 67.25, 66.52, 65.79, 62.39, 61.97, 60.57, 59.03, 56.99, 53.51, 47.20, 28.11, 23.30, 23.28, 21.04,
20.93, 20.85, 20.81, 20.69, 20.63, 20.61, 20.58, 20.50, 18.06; [M+Na]$^+$ C$_{73}$H$_{95}$N$_3$O$_{36}$Na calculated for 1612.5593, found 1612.5678.

N$^\text{α}$-9-Fluorenylmethyloxycarbonyl-O-{2-acetamido-2-deoxy-β-glucopyranosyl-(1→2)}-[β-D-galactopyranosyl-(1→4)-2-acetamido-2-deoxy-β-D-glucopyranosyl-(1→6)-α-D-mannopyranosyl]-L-threonine (13, M301)

Compound 12 (1.90 g, 1.2 mmol) was treated by following general procedure C to afford compound 13 (1.11 g, 85%) as white foam. $R_f = 0.30$ (ethyl acetate/methanol/H$_2$O, 4:2:0.5, v/v); $^1$H NMR (400 MHz, MeOD) $\delta$ 7.79 (d, $J = 7.5$ Hz, 2H), 7.74 – 7.58 (m, 2H), 7.44 – 7.27 (m, 4H), 4.55 (d, $J = 7.7$ Hz, 1H), 4.50 – 4.31 (m, 5H), 4.23 (t, $J = 6.7$ Hz, 1H), 4.18 – 4.02 (m, 2H), 4.01 – 3.84 (m, 5H), 3.84 – 3.51 (m, 14H), 3.51 – 3.30 (m, 6H), 2.08 (s, 3H), 2.06 (s, 3H), 1.22 (d, $J = 5.9$ Hz, 2H); $^{13}$C NMR (100 MHz, MeOD) $\delta$ 173.40, 172.73, 157.41, 144.06, 143.80, 141.18, 127.42, 126.85, 124.88, 119.55, 103.81, 101.78, 98.66, 79.69, 79.33, 76.38, 75.74, 75.04, 73.89, 73.41, 72.46, 71.20, 70.54, 69.85, 68.93, 68.64, 67.77, 66.59, 61.16, 60.88, 60.39, 55.28, 47.08, 22.35, 22.12, 18.44; [M-H]$^-$ C$_{47}$H$_{64}$N$_3$O$_{25}$ calculated for 1071.3907, found 1071.3741.
Phenyl 2,3,4,6-tetra-O-acetyl-β-D-galactopyranosyl-(1→4)-3,6-di-O-acetyl-2-deoxy-2-(2,2,2-trichloroethoxycarbonylamino)-β-D-glucopyranosyl-(1→2)-4,6-O-benzylidene-3-(4-Methoxybenzyl)-1-thio-α-D-Mannopyranoside (14)

Compound 1 (1.8 g, 3.8 mmol) was glycosylated with fresh-made donor 3 (4.23 g, 4.5 mmol) by following general procedure A to get the desired compound 14 (3.46 g, 75%) as white foam. R$_f$ = 0.33 (hexane/acetone, 3:1, v/v); $^1$H NMR (400 MHz, CDCl$_3$) δ 7.51 (dd, $J$ = 7.6, 1.8 Hz, 2H), 7.46 – 7.25 (m, 10H), 6.87 (d, $J$ = 8.7 Hz, 2H), 5.62 (s, 1H), 5.48 (s, 1H), 5.43 – 5.23 (m, 3H), 5.12 (dd, $J$ = 10.4, 7.9 Hz, 1H), 4.97 (dd, $J$ = 10.4, 3.4 Hz, 1H), 4.81 (d, $J$ = 8.2 Hz, 1H), 4.75 – 4.62 (m, 3H), 4.59 – 4.45 (m, 3H), 4.34 (d, $J$ = 1.7 Hz, 1H), 4.30 – 4.03 (m, 6H), 3.94 – 3.86 (m, 2H), 3.85 – 3.72 (m, 5H), 3.65 – 3.51 (m, 2H), 2.15 (s, 3H), 2.06 (s, 3H), 2.06 (s, 6H), 2.04 (s, 3H), 1.97 (s, 3H); $^{13}$C NMR (100 MHz, CDCl$_3$) δ 170.40, 170.33, 170.15, 170.08, 169.09, 159.30, 153.99, 137.55, 133.40, 131.49, 130.00, 129.44, 129.24, 128.90, 128.23, 127.84, 126.10, 113.80, 101.53, 100.98, 99.17, 95.32, 86.79, 78.47, 76.24, 74.48, 74.23, 73.00, 71.90, 71.38, 70.97, 70.75, 69.13, 68.46, 66.69, 65.30, 62.24, 60.94, 55.98, 55.26, 20.82, 20.79, 20.66, 20.63, 20.61, 20.52; [M+Na]$^+$ C$_{54}$H$_{62}$Cl$_3$NO$_{23}$SNa calculated for 1252.2397, found 1252.2444.

Phenyl 2,3,4,6-tetra-O-acetyl-β-D-galactopyranosyl-(1→4)-3,6-di-O-acetyl-2-deoxy-2-(2,2,2-trichloroethoxycarbonylamino)-β-D-glucopyranosyl-(1→2)-3-O-acetyl-1-thio-α-D-Mannopyranoside (15)

Compound 14 (3.3 g, 2.7 mmol) was treated by following general procedure C to afford compound 15 (2.31 g, 81%) as white foam. R$_f$ = 0.22 (hexane/acetone, 2:1, v/v); $^1$H NMR (400
MHz, CDCl$_3$) $\delta$ 7.40 (d, $J = 6.6$ Hz, 2H), 7.34 – 7.21 (m, 3H), 6.07 (d, $J = 9.1$ Hz, 1H), 5.35 (s, 1H), 5.32 (d, $J = 3.1$ Hz, 1H), 5.10 – 4.99 (m, 2H), 4.94 (dd, $J = 10.4$, 3.3 Hz, 1H), 4.87 (dd, $J = 9.9$, 2.6 Hz, 1H), 4.72 (d, $J = 11.9$ Hz, 1H), 4.52 – 4.42 (m, 2H), 4.38 (d, $J = 11.7$ Hz, 3H), 4.31 – 4.19 (m, 1H), 4.15 – 3.98 (m, 4H), 3.94 – 3.78 (m, 2H), 3.78 – 3.63 (m, 3H), 3.63 – 3.53 (m, 1H), 3.43 (d, $J = 3.9$ Hz, 1H), 2.97 (d, $J = 8.7$ Hz, 1H), 2.11 (s, 3H), 2.09 (s, 3H), 2.06 (s, 3H), 2.03 (s, 3H), 2.02 (s, 3H), 1.98 (s, 3H), 1.94 (s, 3H); $^{13}$C NMR (100 MHz, CDCl$_3$) $\delta$ 171.12, 170.46, 170.42, 170.15, 170.13, 169.44, 154.88, 133.34, 131.56, 129.27, 127.88, 101.17, 100.94, 95.06, 85.69, 76.09, 74.58, 73.61, 73.28, 72.73, 71.63, 70.83, 70.69, 69.19, 66.66, 63.99, 62.23, 61.48, 60.91, 55.97, 20.87, 20.78, 20.75, 20.64, 20.58, 20.47; [M+Na]$^+$ C$_{41}$H$_{52}$Cl$_3$NO$_{23}$SNa calculated for 1086.1614, found 1086.1658.

Phenyl 2,3,4,6-tetra-O-acetyl-β-D-galactopyranosyl-(1→4)-3,6-di-O-acetyl-2-deoxy-2-(2,2,2-trichloroethoxycarbonylamino)-β-D-glucopyranosyl-(1→2)-[3,4,6-tri-O-acetyl-2-deoxy-2-(2,2,2-trichloroethoxycarbonylamino)-β-glucopyranosyl-(1→6)]-3-O-acetyl-1-thio-α-D-Mannopyranosyl (16)

Compound 15 (2.10 g, 2.0 mmol) was glycosylated with fresh-made donor 2 (1.54 g, 2.4 mmol) by following general procedure A to get the desired compound 16 (2.75 g, 82%) as white foam. R$_f$ = 0.25 (hexane/acetone, 2:1, v/v); $^1$H NMR (400 MHz, CDCl$_3$) $\delta$ 7.39 (dd, $J = 7.5$, 1.7 Hz, 2H), 7.34 – 7.21 (m, 3H), 6.96 (d, $J = 7.0$ Hz, 1H), 5.93 (d, $J = 9.7$ Hz, 1H), 5.73 (t, $J = 10.0$ Hz, 1H), 5.43 (s, 1H), 5.33 (d, $J = 3.2$ Hz, 1H), 5.18 – 4.89 (m, 6H), 4.81 (dd, $J = 10.0$, 2.7 Hz, 1H), 4.72 – 4.53 (m, 3H), 4.50 – 4.38 (m, 3H), 4.36 (s, 1H), 4.24 (dd, $J = 12.3$, 4.4 Hz, 1H), 4.16
- 4.03 (m, 4H), 4.03 – 3.89 (m, 4H), 3.85 (t, $J = 6.6$ Hz, 1H), 3.83 – 3.65 (m, 3H), 3.65 – 3.56 (m, 1H), 3.17 (dd, $J = 18.1$, 8.0 Hz, 1H), 2.12 (s, 3H), 2.11 (s, 3H), 2.09 (s, 3H), 2.05 (s, 3H), 2.04 (s, 6H), 2.01 (s, 3H), 2.01 (s, 3H), 1.95 (s, 3H), 1.94 (s, 3H); $^{13}$C NMR (100 MHz, CDCl$_3$) δ 170.80, 170.59, 170.51, 170.42, 170.14, 170.10, 169.76, 169.29, 154.95, 133.57, 130.98, 129.18, 127.65, 101.23, 100.94, 100.49, 96.05, 94.96, 85.58, 77.60, 76.27, 75.19, 74.19, 73.55, 72.93, 72.63, 71.62, 71.44, 70.96, 70.87, 70.45, 69.34, 69.09, 67.65, 66.80, 63.26, 62.51, 62.13, 61.26, 57.10, 55.69, 21.02, 20.88, 20.81, 20.75, 20.64, 20.60, 20.57, 20.55, 20.48; [M+Na]$^+$ C$_{56}$H$_{70}$Cl$_6$N$_2$O$_{32}$SNa calculated for 1547.1661, found 1547.1738.

Phenyl 2,3,4,6-tetra-O-acetyl-β-D-galactopyranosyl-(1→4)-2-acetamido-3,6-di-O-acetyl-2-deoxy-β-D-glucopyranosyl-(1→2)-[2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-β-glucopyranosyl-(1→6)]-3,4-di-O-acetyl-1-thio-α-D-Mannopyranoside (17)

Compound 16 (2.50 g, 1.6 mmol) was treated by following general procedure C to afford compound 17 (1.75 g, 82%) as white foam. $R_f = 0.22$ (hexane/acetone, 3:2, v/v); $^1$H NMR (400 MHz, CDCl$_3$) δ 7.46 – 7.36 (m, 2H), 7.36 – 7.26 (m, 2H), 7.26 – 7.19 (m, 1H), 7.10 (d, $J = 6.9$ Hz, 1H), 6.55 (d, $J = 9.9$ Hz, 1H), 5.84 (dd, $J = 10.6$, 8.9 Hz, 1H), 5.74 (s, 1H), 5.50 – 5.38 (m, 2H), 5.33 (d, $J = 2.8$ Hz, 1H), 5.15 – 4.98 (m, 4H), 4.94 (dd, $J = 10.4$, 3.4 Hz, 1H), 4.48 (d, $J = 7.9$ Hz, 1H), 4.37 – 3.97 (m, 11H), 3.85 (t, $J = 7.0$ Hz, 1H), 3.77 – 3.55 (m, 3H), 3.19 (d, $J = 10.9$ Hz, 1H), 2.90 (dt, $J = 10.5$, 8.1 Hz, 1H), 2.15 (s, 3H), 2.12 (s, 3H), 2.10 (s, 3H), 2.09 (s, 3H), 2.08 (s, 3H), 2.05 (s, 3H), 2.03 (s, 6H), 2.02 (s, 3H), 2.01 (s, 3H), 1.98 (s, 3H), 1.95 (s, 3H), 1.91 (s, 3H); $^{13}$C NMR (100 MHz, CDCl$_3$) δ 172.73, 172.09, 171.03, 170.70, 170.63, 170.53, 170.35, 170.26,
Compound 17 (1.57 g, 1.2 mmol) was glycosylated with compound 4 (0.98 g, 2.4 mmol) by following general procedure B to get the desired compound 18 (1.56 g, 83%) as white foam. Rf = 0.23 (hexane/acetone, 3:2, v/v); 1H NMR (400 MHz, CDCl3) δ 7.76 (d, J = 7.4 Hz, 2H), 7.71 – 7.51 (m, 2H), 7.49 – 7.25 (m, 5H), 7.05 (d, J = 7.0 Hz, 1H), 6.51 (d, J = 9.6 Hz, 1H), 5.87 – 5.72 (m, 1H), 5.57 (d, J = 9.3 Hz, 1H), 5.41 – 5.28 (m, 3H), 5.19 – 5.01 (m, 4H), 5.01 – 4.87 (m, 2H), 4.59 – 4.41 (m, 3H), 4.41 – 3.96 (m, 16H), 3.90 – 3.81 (m, 1H), 3.81 – 3.70 (m, 1H), 3.70 – 3.55 (m, 3H), 3.17 (d, J = 11.2 Hz, 1H), 2.88 (dd, J = 17.5, 8.1 Hz, 1H), 2.18 (s, 3H), 2.16 (s, 3H), 2.12 (s, 3H), 2.12 (s, 3H), 2.10 (s, 3H), 2.06 (s, 3H), 2.05 (s, 3H), 2.05 (s, 3H), 2.04 (s, 6H), 2.02 (s, 3H), 1.97 (s, 3H), 1.94 (s, 3H), 1.52 (s, 9H), 1.27 (d, J = 8.5 Hz, 3H); 13C NMR (100 MHz, CDCl3) δ 172.45, 172.04, 171.00, 170.70, 170.62, 170.57, 170.35, 170.22, 170.12, 169.82, 169.38, 169.01,
156.57, 143.95, 143.78, 141.28, 127.70, 127.10, 125.25, 119.98, 102.88, 101.06, 98.03, 96.56, 82.47, 78.21, 77.93, 72.88, 72.79, 71.91, 71.03, 70.88, 70.52, 69.85, 69.50, 69.22, 69.11, 68.61, 68.38, 67.25, 66.53, 65.93, 62.51, 60.61, 58.98, 57.02, 53.84, 53.48, 47.19, 29.28, 28.13, 23.28, 23.18, 21.09, 20.72, 20.68, 20.63, 20.59, 20.53, 17.95; \([M+Na]^+\)C_{73}H_{95}N_{3}O_{36}Na calculated for 1612.5593, found 1612.5663.

\[\text{N}^\alpha\text{-9-Fluorenylmethyloxycarbonyl-O-}\{\beta\text{-D-galactopyranosyl-(1→4)}\text{-2-acetamido-2-deoxy-}\beta\text{-D-glucopyranosyl-(1→2)}\text{-[2-acetamido-2-deoxy-}\beta\text{-glucopyranosyl-(1→6)}\text{-}\alpha\text{-D-Mannopyranosyl]-L-threonine (19, M201)}\]

Compound 18 (1.90 g, 1.2 mmol) was treated by following general procedure C to afford compound 19 (1.09 g, 87%) as white foam. \(R_f = 0.30\) (ethyl acetate/methanol/H\(_2\)O, 4:2:0.5, v/v); \(^1\)H NMR (400 MHz, D\(_2\)O) \(\delta\) 7.68 – 7.56 (m, 2H), 7.56 – 7.36 (m, 2H), 7.36 – 7.16 (m, 5H), 4.59 – 4.51 (m, 1H), 4.47 (d, \(J = 8.4\) Hz, 1H), 4.43 – 4.23 (m, 4H), 4.18 (d, \(J = 6.4\) Hz, 1H), 4.09 (d, \(J = 10.3\) Hz, 1H), 3.98 (d, \(J = 4.8\) Hz, 1H), 3.90 – 3.80 (m, 4H), 3.80 – 3.55 (m, 16H), 3.55 – 3.42 (m, 5H), 3.41 – 3.24 (m, 5H), 1.97 (s, 3H), 1.95 (s, 4H), 0.95 (d, \(J = 6.1\) Hz, 3H); \(^{13}\)C NMR (100 MHz, D\(_2\)O) \(\delta\) 174.59, 174.15, 158.00, 143.88, 143.45, 140.96, 140.90, 127.97, 127.43, 124.80, 120.11, 102.90, 101.16, 99.84, 98.22, 78.31, 77.18, 76.85, 75.90, 75.31, 74.56, 73.93, 72.48, 71.95, 70.94, 70.06, 69.21, 68.56, 67.59, 65.98, 61.02, 60.87, 59.88, 55.57, 54.82, 48.89, 47.09, 22.50, 22.33, 18.49; \([M+Na]^+\)C_{47}H_{65}N_{3}O_{25}Na calculated for 1094.3805, found 1094.3876.
1.4.2 Enzymatic extension

Materials and enzymes

N-Acetylneuraminic acid (Neu5Ac), N-Glycolylneuraminic acid (Neu5Gc) and cytidine 5′-triphosphate (CTP) were purchased from Carbosynth Limited. Sugar nucleotides uridine 5′-diphospho-galactose (UDP-Gal)\textsuperscript{38}, guanosine 5′-diphospho-L-fucose (GDP-Fuc)\textsuperscript{39}, uridine 5′-diphosphoglucuronic acid (UDP-GlcA)\textsuperscript{40} were prepared as described previously reported. Enzymes including CMP-sialic acid synthetase from Neisseria meningitidis (NmCSS)\textsuperscript{30} mutant M144D of α2,3-sialyltransferase 1 from Pasteurella multocida (PmST1-M144D)\textsuperscript{25} mutant E271F/R313Y of PmST1 (PmST1m)\textsuperscript{34} α2,6-sialyltransferase from Photobacterium damselae (Pd2,6ST)\textsuperscript{26} β1,4-galactosyltransferase from N. meningitidis (NmLgtB)\textsuperscript{24} and C-terminal 66 amino acid truncated α1,3-fucosyltransferase from Helicobacter pylori (Hp3FT)\textsuperscript{41} were expressed and purified as previously described. All enzymes were desalted against 50 mM Tris-HCl, and 20% glycerol, and stored at -20 °C for long-term use. β-galactosidase (βGalD) from Streptococcus pneumoniae was purchased from Prozyme, CA.

Cloning and expression of human GlcAT-P

The human β1,3-galacturonyltransferase gene (GlcAT-P) gene (GenBank: AB029396.1) was codon optimized, synthesized and cloned into vector pET15b vector (Genescript, NJ). The recombinant plasmid harboring GlcAT-P was transformed into E. coli BL21 (DE3) for heterogenous expression. The recombinant strain was cultured in LB medium at 37 °C with brief shaking (180 rpm) until OD\textsubscript{600nm} reached 0.6 to 0.8, followed by addition of isopropyl-b-D-thiogalactopyranoside (IPTG) to a final concentration of 0.2 mM. After 20 h induction at 16 °C,
the cells were harvested by centrifuging at 4000 rpm for 30 min. The target protein was expressed as inclusion body. *E. coli* cell lysate was used to in GlcAT-P-catalyzed reactions.

**Optimized GlcAT gene sequence**

```
CATCAATCCACCCTGGCACCCTGCTGGCAGTCCATAAGGACGAAGGTTCCG
ATCCCGCTCGTGAGACCCCGCCGGGCAGACCCCGCTGAGTGACTGCACCAGCGAT
CGTGACATCGTGAAAGTGGTGCAGTACCCGACCCTACAGCCTACAGGGCTCCGG
AGCGATACCCCTGCGAGACCCTTCACCCTACAGCCTACAGGGCTCCGG
AAGGCTGAGCTGACCGTGATATGGCTAACCCTGCTGACGTTCCGAACCTGGA
CTGGTTGTGGAAAGATGCACCGCTGCTAGCCCGCTGACGGCTGACGGCTGAC
ACCCTGCTGACGTTCCGAACCTGGA
```

**General HPLC methods**

HPLC method for monitoring reactions and purity analysis of final products: An analytical GL Science Inertsil ODS-4 column (100 Å, 5 μm, 4.6 mm × 250 mm) was used for separation,
monitored by a UV detector (260 nm) or fluorescent detector (Ex 260nm, Em 310 nm).[23] The running solvents are solvent A(H2O with 0.1% TFA) and solvent B (acetonitrile with 0.1% TFA). The running condition is gradient elution with solvent B% linear increased from 20% to 40% within 25 mins, with a total flow rate of 1 mL/min.

HPLC method for purifying final products: An analytical GL Science Inertsil ODS-4 column (100 Å, 5 μm, 4.6 mm × 250 mm) was used for separating small reactions with 3 mg or less products, and a semipreparative Inertsil ODS-4 column (100 Å, 5 μm, 20 mm × 250 mm) was used for separating products with over 5 mgs. The method for using the analytical column is same as above and monitored by a UV detector (260 nm). The method for using the semipreparative column is similar as that for the analytical column, with the only difference of flow rate at 18.9 mL/min instead. Commonly, the analytical column enabled up to 1 mg product separation per run, while the semipreparative one enabled up to 15 mg product separation per run.

**HPLC quantification of purified glycans**

Firstly, 10 mg M201 was weighted out accurately and diluted to 1 μM as standard solution. Then different volumes of standard solution, 1 μL, 2 μL, 3 μL, 4 μL, 6 μL, 8 μL, 10 μL, 20 μL, 30 μL, 40 μL, 60 μL, 80 μL, were injected into the HPLC by the method mentioned above using fluorescent (Ex. 260 nm, Em. 310 nm) in three replicates. The peak area of each injection was recorded and calculated to make the standard curve. All the purified glycans was quantified by the same condition as built the standard curve.

**β1,4-galactosylation by NmLgtB**
Reaction mixtures contain Tris-HCl (100 mM, pH 7.5), an acceptor O-mannosyl glycan (10 mM), UDP-Gal (15 mM), MgCl₂ (10 mM), and an appropriate amount of NmLgtB. Reactions were incubated at 37 °C overnight and monitored by HPLC and/or MALDI-MS. After over 90% acceptor was converted, the reaction was quenched by freezing at -80 °C for 30 min, thaw and brief centrifugation to remove protein precipitants. The sample was then concentrated and subject for HPLC separation, product-containing fractions were pooled and lyophilized for characterization and next step use. The purity and quantification of each glycan was confirmed by HPLC as described above. NmLgtB-catalyzed reactions give 90-96% yields (HPLC purified glycan product/starting glycan substrate × 100%).

**One-pot two-enzyme α2,3-sialylation catalyzed by PmST1-M144D or PmST1m**

Reaction mixtures contain Tris-HCl (100 mM, pH 7.5), an acceptor glycan (10 mM), CTP (15 mM), Neu5Ac or Neu5Gc (15 mM), MgCl₂ (10 mM), and appropriate amount of NmCSS and PmST1-M144D (or PmST1m). PmST1-M144D-catalyzed reactions were incubated at 37 °C overnight and monitored by HPLC and/or MALDI-TOF. After over 90% acceptor was converted, the reaction was quenched, concentrated and subject for HPLC separation. PmST1m-catalyzed reactions were incubated at 37 °C for 30 min, quenched, and concentrated for HPLC separation. Product-containing fractions were pooled and lyophilized for characterization and next step use. PmST1-catalyzed reactions give 85-92% yields (HPLC purified glycan product/starting glycan substrate × 100%).

**One-pot two-enzyme α2,6-sialylation catalyzed by Pd26ST**
Reaction mixtures contain Tris-HCl (100 mM, pH 7.5), an acceptor glycan (10 mM), CTP (15 mM), Neu5Ac or Neu5Gc (15 mM), MgCl₂ (10 mM), and appropriate amount of NmCSS and Pd26ST. Reactions were incubated at 37 °C overnight and monitored by HPLC and/or MALDI-TOF. After over 90% acceptor was converted, reactions were quenched, concentrated before HPLC separation. Product-containing fractions were pooled and lyophilized for characterization and next step use. Pd26ST-catalyzed reactions give 89-94% yields (HPLC purified glycan product/starting glycan substrate × 100%).

**α1,3-fucosylation catalyzed by Hp3FT**

Reaction mixtures contain Tris-HCl (100 mM, pH 7.5), an acceptor glycan (10 mM), GDP-Fuc (15 mM), MgCl₂ (10 mM), and appropriate amount of Hp3FT. Reactions were incubated at 37 °C overnight and were monitored by HPLC and/or MALDI-TOF. After over 90% acceptor was converted, reactions were quenched, concentrated before HPLC separation. Product-containing fractions were pooled and lyophilized for characterization and next step use. Hp3FT-catalyzed reactions give 85-86% yields (HPLC purified glycan product/starting glycan substrate × 100%).

**GlcAT-P-catalyzed addition of GlcA**

The reaction mixture contains Tris-HCl (100 mM, pH 7.5), M101 (10 mM), UDP-GlcA (15 mM), MgCl₂ (10 mM), and appropriate amount of GlcAT-P cell lysis. Reaction was incubated at 37 °C overnight. The reaction was monitored by HPLC until 90% M101 was converted. The reaction was then quenched and concentrated for HPLC separation. Product-containing fractions were pooled and lyophilized for characterization. The reaction gives a yield of 90% (HPLC purified glycan product/starting glycan substrate × 100%).
**β-galactosidase-catalyzed reaction**

The reaction was performed according to the instruction using commercial β-galactosidase in the presence of M201. The reaction was monitored and product was purified by HPLC. The reaction gives a yield of 96% (HPLC purified glycan product/starting glycan substrate × 100%)

### 1.4.3 Glycan Microarray preparation, assay and data

**Method for removing Fmoc**

O-mannosyl glycans (50 μg) were dissolved in 200 μL H₂O, and 30 μL triethylamine was added to remove the Fmoc group at room temperature for 4 h. The reactions were then lyophilized and hexane extraction was used to remove free Fmoc.

**Method for microarray preparation**

O-mannosyl glycans microarray was prepared as previously reported,[42] briefly, 8 subarrays was printed on N-hydroxysuccinimide (NHS)-derivatized slides by Z Biotech (Aurora, CO, USA). Within each subarray, each glycan was printed in six replicates with print buffer, and print buffer was also printed as a negative control. In addition, Biotin-PEG2-Amine (0.01 mg/mL) (positive control 1), Rabbit IgG (0.1 mg/mL) (positive control 2) were printed in six replicates with print buffer to serve as a positive control. A marker containing human IgG conjugate with Cy3 (0.01 mg/mL) and human IgG conjugate with Alexa 647 (0.01 mg/mL) was also printed in the replicates of six.

**Method for microarray assay**
**Materials:** All biotinylated lectins were purchased from EYLabs (San Mateo, CA). Cy5-streptavidin, Cy3-streptavidin, goat anti-rabbit IgG-Alexa Fluor 647 conjugate, goat anti-mouse IgG-Alexa Fluor 647 conjugate were purchased from ThermoFisher Scientific (Waltham, MA). Mouse anti-human CD15s (sialyl-lewis X) antibody was purchased from BD Biosciences (Franklin Lakers, NJ). Mab(IIH6) was a kind gift from Dr. Kevin Campbell (HHMI, University of Iowa).

Procedures: Microarray slides were rehydrated for 30 min in blocking buffer (50 mM ethanolamine in 50 mM sodium borate, pH 9.2) and wash with H$_2$O before assay. All assay were performed as previously reported. Plant lectins, including Concanavalin A from *Canavalia ensiformis* (Con A, 10 µg/mL), *Aleuria aurantia* lectin (AAL, 10 µg/mL), *Ricinus Communis* lectin I (RCA-I, 10 µg/mL), *Erythrina cristagalli* Lectin (ECA, 10 µg/mL) were applied with appropriate concentrations, and detected by Cy3-streptavidin or Cy5-streptavidin (1 µg/mL). Anti-CD15s antibody (10 µg/mL), and IIH6 antibody (1:200 dilution) antibody were also tested. The primary antibodies were bound by goat anti-mouse IgG-Alexa Fluor 647 conjugate (5 µg/mL). Twelve subarrays were assayed with rabbit serums from Dr. Ulrika Westerlind. Goat anti-rabbit IgG-Alexa Fluor 647 conjugate (5 µg/mL) was used to bound with the primary antibodies. After binding, the slides were scanned with a microarray scanner (GenePix 4000B).

**Results and analysis**

Firstly, binding specificities *Aleuria aurantia* lectin (AAL, specific to $\alpha$-linked Fuc) was profiled. As shown in Figure S2, AAL bound to all O-mannosyl glycans with an $\alpha$1,3-linked Fuc residue (Fig 6A). Meanwhile, weaker bindings were observed towards M3X4 and M3X5 (contain one Fuc on the $\beta$1,6-branch) than M2X4 and M2X5 (contain one Fuc on the $\beta$1,2-branch) with the
exception of M345 and M245 (contain Fuc residues on both branches), indicating AAL possesses a branch preference towards the $\beta_{1,2}$-branch. In addition, AAL showed higher bindings to glycans that contain the Le$^X$ epitope (e.g., M104, M040, M204, M214, M304, M314) than those contain the sLe$^X$ epitope (e.g., M105, M050, M205, M215, M305, M315), suggesting a same-chain glycosylation ($\alpha_{2,3}$-sialylation) influence. Besides ECA and RCA-I’s branch preferences, When comparing binding specificities of RCA-I towards M212, M213, M214 and M215, a series of glycans with the same $\beta_{1,6}$-branch (LN disaccharide) but differs in the $\beta_{1,2}$-branch, an apparent preference of M213>M214>M215>M212 was observed, suggesting a side-chain influence on RCA-I binding. Such influences were also found for ECA (Fig 1.6A) and AAL (varied bindings to M104, M204, M214, M224, M234), but in different orders.

As expected, a $\alpha$-DG antibody (IIH6) failed to bind any core M1 and branched core M2 structures, as it is specific for core M3 glycans.$^{[43]}$

![Figure 1.8 Microarray analysis towards Aleuria aurantia lectin.](image)
Figure 1.9 Microarray analysis towards Concanavalin A from Canavalia ensiformis

Figure 1.10 Microarray analysis and binding profile of O-mannosyl glycans towards rabbit 26559
1.4.4 NMR and MS analysis of purified O-mannosyl glycan

**M100**

$^1$H NMR (400 MHz, D$_2$O) $\delta$ 7.74 – 7.60 (m, 2H), 7.60 – 7.40 (m, 2H), 7.40 – 7.18 (m, 4H), 4.60 (dd, $J = 10.7$, 4.9 Hz, 1H), 4.42 – 4.27 (m, 2H), 4.23 (d, $J = 6.3$ Hz, 1H), 4.12 – 3.96 (m, 1H), 3.83 – 3.71 (m, 4H), 3.71 – 3.42 (m, 7H), 3.42 – 3.30 (m, 3H), 3.27 (s, 1H), 1.96 (s, 3H), 0.94 (d, $J = 6.1$ Hz, 3H); $^{13}$C NMR (100 MHz, D$_2$O) $\delta$ 174.87, 158.05, 143.91, 143.49, 140.96, 140.91, 127.96, 127.41, 124.90, 124.78, 120.09, 100.01, 98.04, 77.19, 75.59, 73.28, 73.05, 69.84, 69.26, 67.22, 65.92, 61.60, 60.52, 55.26, 47.10, 22.35, 18.33; ESI-MS, calculated: 706.2585; found [M-H]$^-$ 705.2470; MALDI-MS: calculated for 729.2483, found [M+Na]$^+$ 729.2478.

**M101**

$^1$H NMR (600 MHz, D$_2$O) $\delta$ 7.71 – 7.55 (m, 2H), 7.71 – 7.55 (m, 2H), 7.51 – 7.29 (m, 4H), 4.77 (d, $J = 4.9$ Hz, 1H), 4.75 (d, $J = 4.8$ Hz, 1H), 4.65 (m, 1H), 4.63 (m, 1H), 4.57 (dd, $J = 11.0$, 4.6 Hz, 1H), 4.38 (d, $J = 7.8$ Hz, 1H), 4.35 (d, $J = 5.5$ Hz, 1H), 4.29 – 4.20 (m, 2H), 3.95 (d, $J = 3.9$ Hz, 1H), 3.83 – 3.71 (m, 4H), 3.71 – 3.42 (m, 7H), 3.42 – 3.30 (m, 3H), 3.27 (s, 1H), 1.96 (s, 3H), 0.94 (d, $J = 6.1$ Hz, 3H); $^{13}$C NMR (100 MHz, D$_2$O) $\delta$ 174.87, 158.05, 143.91, 143.49, 140.96, 140.91, 127.96, 127.41, 124.90, 124.78, 120.09, 100.01, 98.04, 77.19, 75.59, 73.28, 73.05, 69.84, 69.26, 67.22, 65.92, 61.60, 60.52, 55.26, 47.10, 22.35, 18.33; ESI-MS, calculated: 706.2585; found [M-H]$^-$ 705.2470; MALDI-MS: calculated for 729.2483, found [M+Na]$^+$ 729.2478.
2.1 Hz, 1H), 3.86 (m, 2H), 3.81 – 3.57 (m, 10H), 3.57 – 3.42 (m, 4H), 3.38 (t, J = 9.7 Hz, 1H), 1.96 (s, 3H), 0.96 (d, J = 6.4 Hz, 3H). ESI-MS, calculated: 868.3113; found [M-H]− 867.2972; MALDI-MS, found [M+Na]+ 891.305, [M+K]+ 907.279.

1H NMR (600 MHz, D2O) δ 7.90 – 7.73 (m, 2H), 7.71 – 7.50 (m, 2H), 7.45 – 7.25 (m, 4H), 4.65 (s, 1H), 4.59 (s, 1H), 4.54 (dd, J = 11.0, 4.6 Hz, 1H), 4.41 (d, J = 7.8 Hz, 1H), 4.30 (d, J = 7.1 Hz, 1H), 4.23 (s, 1H), 4.17 (d, J = 5.6 Hz, 1H), 4.00 (d, J = 12.5 Hz, 1H), 3.89 – 3.66 (m, 8H), 3.65 – 3.36 (m, 16H), 3.32 (t, J = 9.6 Hz, 1H), 2.64 (dd, J = 12.4, 4.4 Hz, 1H), 1.91 (s, 3H), 1.90 (s, 3H), 1.69 (t, J = 12.1 Hz, 1H), 0.91 (d, J = 6.2 Hz, 3H). ESI-MS, calculated: 1159.4068; found [M-H]− 1158.3877.
**M102G**

$^1$H NMR (600 MHz, D$_2$O) $\delta$ 7.86 – 7.73 (m, 2H), 7.67 – 7.49 (m, 2H), 7.41 – 7.25 (m, 4H), 4.64 – 4.60 (m, 1H), 4.57 – 4.50 (m, 1H), 4.40 (d, $J$ = 8.0 Hz, 1H), 4.28 (d, $J$ = 6.7 Hz, 1H), 4.23 (s, 2H), 4.17 (d, $J$ = 6.7 Hz, 1H), 4.03 – 3.97 (m, 3H), 3.94 (s, 1H), 3.86 – 3.35 (m, 24H), 3.31 (t, $J$ = 9.7 Hz, 1H), 2.64 (d, $J$ = 8.4 Hz, 1H), 1.88 (s, 3H), 1.69 (t, $J$ = 12.1 Hz, 1H), 0.90 (d, $J$ = 6.2 Hz, 3H). ESI-MS, calculated: 1175.4017; found [M-H]$^-$ 1174.3850; MALDI-MS, found [M+Na]$^+$ 1198.394, [M+K]$^+$ 1214.375.

**M103**

$^1$H NMR (600 MHz, D$_2$O) $\delta$ 7.83 – 7.72 (m, 2H), 7.63 – 7.49 (m, 2H), 7.42 – 7.25 (m, 4H), 4.69 (s,1H), 4.68 (s,1H), 4.53 (dd, $J$ = 11.0, 3.4 Hz, 1H), 4.35 – 4.25 (m, 2H), 4.20 (s, 2H), 4.02 (s, 1H), 3.89 (t, $J$ = 9.4 Hz, 1H), 3.86 – 3.67 (m, 7H), 3.67 – 3.37 (m, 15H), 3.34 (t, $J$ = 9.4 Hz, 1H), 2.53 (d, $J$ = 8.7 Hz, 1H), 1.92 (s, 3H), 1.91 (s, 3H), 1.64 (t, $J$ = 12.2 Hz, 1H), 0.92 (d, $J$ = 6.0 Hz, 3H). ESI-MS, calculated: 1159.4068; found [M-H]$^-$ 1158.3878.

**M103G**

$^1$H NMR (600 MHz, D$_2$O) $\delta$ 7.89 – 7.74 (m, 2H), 7.71 – 7.53 (m, 2H), 7.49 – 7.29 (m, 4H), 4.74 (s, 1H), 4.67 – 4.60 (m, 1H), 4.62 – 4.54 (m, 1H), 4.40 – 4.30 (m, 2H), 4.24 (d, $J$ = 5.1 Hz, 2H), 4.08 (s, 1H), 4.04 (s, 2H), 3.95 (t, $J$ = 9.5 Hz, 1H), 3.89 – 3.71 (m, 10H), 3.71 – 3.34 (m, 14H), 2.59 (dd, $J$ = 12.5, 4.4 Hz, 1H), 2.03 – 1.93 (m, 3H), 1.70 (t, $J$ = 12.2 Hz, 1H), 0.97 (d, $J$ = 6.4 Hz, 3H). ESI-MS, calculated: 1175.4017; found [M-H]$^-$ 1174.3792MALDI-MS, found [M+Na]$^+$ 1198.391, [M+K]$^+$ 1214.372.
M104

$^1$H NMR (600 MHz, D$_2$O) δ 7.83 – 7.60 (m, 2H), 7.60 – 7.40 (m, 2H), 7.38 – 7.18 (m, 4H), 5.04 (d, $J = 3.8$ Hz, 1H), 4.76 (d, $J = 6.8$ Hz, 1H), 4.64 – 4.54 (m, 1H), 4.43 (s, 1H), 4.36 (d, $J = 7.6$ Hz, 2H), 4.25 (s, 1H), 4.12 (s, 1H), 4.05 (s, 1H), 3.90 – 3.70 (m, 9H), 3.70 – 3.32 (m, 11H), 1.94 (s, 3H), 1.11 (d, $J = 6.2$ Hz, 3H), 0.98 (d, $J = 5.8$ Hz, 3H). ESI-MS, calculated: 1014.3693; found [M-H]$^-$ 1013.3524.

M105

$^1$H NMR (600 MHz, D$_2$O) δ 7.85 – 7.73 (m, 2H), 7.68 – 7.48 (m, 2H), 7.42 – 7.23 (m, 4H), 4.96 (d, $J = 3.2$ Hz, 1H), 4.65 (s, 1H), 4.64 (s, 1H), 4.53 (dd, $J = 10.9$, 4.5 Hz, 2H), 4.36 (d, $J = 7.9$ Hz, 1H), 4.28 (d, $J = 8.0$ Hz, 1H), 4.25 – 4.14 (m, 2H), 4.00 – 3.88 (m, 2H), 3.87 – 3.60 (m, 13H), 3.60 – 3.34 (m, 14H), 3.28 (t, $J = 9.6$ Hz, 1H), 2.62 (d, $J = 8.0$ Hz, 1H), 1.89 (s, 3H), 1.88 (s, 3H), 1.67 (t, $J = 12.1$ Hz, 1H), 1.03 (d, $J = 6.2$ Hz, 3H), 0.90 (d, $J = 6.1$ Hz, 3H). ESI-MS, calculated: 1305.4647; found [M-H]$^-$ 1304.4440.

M106

$^1$H NMR (600 MHz, D$_2$O) δ 7.84 – 7.64 (m, 2H), 7.64 – 7.44 (m, 2H), 7.41 – 7.19 (m, 4H), 4.65 (s, 1H), 4.58 (d, $J = 7.9$ Hz, 1H), 4.50 (dd, $J = 11.0$, 4.7 Hz, 1H), 4.36 (d, $J = 7.8$ Hz, 1H), 4.32 – 4.23 (m, 1H), 4.22 – 4.12 (m, 2H), 4.02 (d, $J = 2.9$ Hz, 1H), 3.92 (d, $J = 2.1$ Hz, 1H), 3.83
- 3.51 (m, 15H), 3.51 - 3.36 (m, 5H), 3.36 - 3.24 (m, 2H), 1.88 (s, 3H), 0.90 (d, J = 6.3 Hz, 3H).

ESI-MS, calculated: 1044.3434; found [M-H]^- 1043.3264.

\[ \text{M000} \]

^1H NMR (600 MHz, D2O) δ 7.94 - 7.79 (m, 2H), 7.71 - 7.54 (m, 2H), 7.50 - 7.29 (m, 4H),
4.76 (dd, J = 10.8, 4.9 Hz, 1H), 4.67 (s, 1H), 4.56 (dd, J = 11.0, 4.7 Hz, 1H), 4.47 (d, J = 8.5 Hz, 1H), 4.34 (d, J = 8.3 Hz, 1H), 4.27 (s, 1H), 4.18 (d, J = 6.1 Hz, 1H), 4.12 (d, J = 10.4 Hz, 1H),
4.01 (s, 1H), 3.86 (d, J = 12.3 Hz, 1H), 3.80 (d, J = 10.9 Hz, 1H), 3.75 - 3.43 (m, 10H), 3.43 -
3.24 (m, 5H), 1.96 (s, 3H), 1.94 (s, 3H), 0.96 (d, J = 6.4 Hz, 3H). ESI-MS, calculated: 909.3379;

\[ \text{M010} \]

^1H NMR (600 MHz, D2O) δ 7.91 - 7.72 (m, 2H), 7.72 - 7.49 (m, 2H), 7.49 - 7.27 (m, 4H),
4.74 (d, J = 3.3 Hz, 1H), 4.66 (d, J = 4.5 Hz, 1H), 4.60 - 4.51 (m, 1H), 4.51 - 4.43 (m, 2H), 4.41 -
4.35 (m, 2H), 4.34 - 4.24 (m, 2H), 4.20 (d, J = 5.8 Hz, 2H), 4.11 (d, J = 10.8 Hz, 1H), 4.03 (d, J =
14.8 Hz, 1H), 4.00 - 3.79 (m, 4H), 3.79 - 3.39 (m, 21H), 3.31 (t, J = 10.0 Hz, 1H), 1.96 (s, 3H),
1.94 (s, 3H), 0.99 (d, J = 4.8 Hz, 3H). ESI-MS, calculated: 1233.4435; found [M+H]^+ 1234.4497,

\[ \text{M020} \]
\( ^1H \) NMR (600 MHz, D\(_2\)O) \( \delta \) 7.92 – 7.78 (m, 2H), 7.72 – 7.54 (m, 2H), 7.52 – 7.29 (m, 4H), 4.76 – 4.74 (m, 1H), 4.68 – 4.66 (m, 1H), 4.59 – 4.50 (m, 2H), 4.52 – 4.44 (m, 2H), 4.40 (d, \( J = 7.7 \) Hz, 1H), 4.35 (d, \( J = 7.5 \) Hz, 1H), 4.28 (s, 2H), 4.20 (d, \( J = 6.4 \) Hz, 1H), 4.17 – 4.00 (m, 4H), 3.98 – 3.39 (m, 38H), 3.30 (t, \( J = 10.3 \) Hz, 1H), 2.68 (dd, \( J = 12.6, 4.0 \) Hz, 2H), 1.96 (s, 9xH), 1.92 (s, 3H), 1.78 (t, \( J = 12.5 \) Hz, 2H), 0.99 (d, \( J = 6.1 \) Hz, 3H). ESI-MS, calculated: 1815.6344; found [M-H]\(^-\) 1814.6297, [M-2H]\(^-\) 906.8041.

\( ^1H \) NMR (600 MHz, D\(_2\)O) \( \delta \) 7.90 – 7.78 (m, 2H), 7.71 – 7.54 (m, 2H), 7.50 – 7.29 (m, 4H), 4.62 – 4.53 (m, 1H), 4.50 (d, \( J = 8.4 \) Hz, 2H), 4.42 – 4.33 (m, 2H), 4.33 – 4.25 (m, 2H), 4.22 (d, \( J = 4.7 \) Hz, 1H), 4.17 – 4.05 (m, 2H), 4.00 – 3.42 (m, 40H), 3.31 (t, \( J = 9.8 \) Hz, 1H), 2.62 – 2.56 (m, 2H), 1.98 (s, 3H), 1.96 (s, 3H), 1.96 (s, 3H), 1.95 (s, 3H), 1.68 (t, \( J = 12.2 \) Hz, 2H), 1.00 (d, \( J = 6.4 \) Hz, 3H). ESI-MS, calculated: 1815.6344; found [M-H]\(^-\) 1814.6103, [M-2H]\(^-\) 906.7999.

\( ^1H \) NMR (600 MHz, D\(_2\)O) \( \delta \) 7.71 – 7.54 (m, 2H), 7.54 – 7.32 (m, 2H), 7.23 (m, \( J = 31.1 \) Hz, 4H), 4.99 – 4.92 (m, 2H), 4.61 (d, \( J = 7.0 \) Hz, 1H), 4.48 (d, \( J = 9.1 \) Hz, 1H), 4.44 – 4.35 (m, 1H), 4.35 – 4.21 (m, 3H), 4.18 (d, \( J = 7.1 \) Hz, 1H), 4.14 (d, \( J = 5.5 \) Hz, 1H), 4.03 (d, \( J = 34.7 \) Hz, 3H), 3.90 – 3.16 (m, 35H), 1.88 (s, 3H), 1.85 (s, 3H), 1.08 – 0.98 (m, 6H), 0.93 (d, \( J = 5.0 \) Hz,

**M050**

$^1$H NMR (600 MHz, D$_2$O) δ 7.87 – 7.72 (m, 2H), 7.67 – 7.48 (m, 2H), 7.44 – 7.26 (m, 4H), 5.00 – 4.91 (m, 2H), 4.64 – 4.63 (m, 1H), 4.62 – 4.58 (m, 2H), 4.54 – 4.46 (m, 2H), 4.46 – 4.35 (m, 3H), 4.35 – 4.27 (m, 2H), 4.23 (s, 2H), 4.14 (d, $J = 5.5$ Hz, 1H), 4.10 – 3.91 (m, 5H), 3.91 – 3.61 (m, 21H), 3.61 – 3.31 (m, 22H), 3.23 (t, $J = 9.9$ Hz, 1H), 2.66 – 2.56 (m, 2H), 1.89 (s, 6H), 1.88 (s, 3H), 1.85 (s, 3H), 1.68 (t, $J = 12.0$ Hz, 2H), 1.08 – 0.99 (m, 6H), 0.91 (d, $J = 5.7$ Hz, 3H).

ESI-MS, calculated: 2107.7502; found [M-2H]⁻ 1052.8568.

**M201**

$^1$H NMR (400 MHz, D$_2$O) δ 7.68 – 7.56 (m, 2H), 7.56 – 7.36 (m, 2H), 7.36 – 7.16 (m, 5H), 4.59 – 4.51 (m, 1H), 4.47 (d, $J = 8.4$ Hz, 1H), 4.43 – 4.23 (m, 4H), 4.18 (d, $J = 6.4$ Hz, 1H), 4.09 (d, $J = 10.3$ Hz, 1H), 3.98 (d, $J = 4.8$ Hz, 1H), 3.90 – 3.80 (m, 4H), 3.80 – 3.55 (m, 16H), 3.55 – 3.42 (m, 5H), 3.41 – 3.24 (m, 5H), 1.97 (s, 3H), 1.95 (s, 4H), 0.95 (d, $J = 6.1$ Hz, 3H); $^{13}$C NMR (100 MHz, D$_2$O) δ 174.59, 174.15, 158.00, 143.88, 143.45, 140.96, 140.90, 127.97, 127.43, 124.80, 120.11, 102.90, 101.16, 99.84, 98.22, 78.31, 77.18, 76.85, 75.90, 75.31, 74.56, 73.93, 72.48, 71.95, 70.94, 70.06, 69.21, 68.56, 67.59, 65.98, 61.02, 60.87, 59.88, 55.57, 54.82, 48.89,

\[ \text{M202} \]

\[ ^1 \text{H NMR (600 MHz, D}_2\text{O)} \delta 7.63 – 7.47 (m, 2H), 7.47 – 7.26 (m, 2H), 7.26 – 7.07 (m, 4H), 4.63 (s, 1H), 4.50 – 4.34 (m, 3H), 4.31 – 4.19 (m, 2H), 4.15 (s, 1H), 4.08 – 3.97 (m, 3H), 3.93 (d, } J = 11.3 \text{ Hz, 1H)}, 3.90 – 3.50 (m, 21H), 3.50 – 3.35 (m, 5H), 3.35 – 3.20 (m, 3H), 2.63 (d, } J = 8.5 \text{ Hz, 1H}), 1.90 (s, 6H), 1.87 (s, 3H), 1.75 (t, } J = 12.2 \text{ Hz, 1H)}, 0.93 (d, } J = 5.5 \text{ Hz, 3H). ESI-MS, calculated: 1362.4861; found [M-H]- 1361.4659, [M-2H]^{2-} 680.2284; MALDI-MS; found [M+Na]+ 1385.436, [M+K]+ 1401.414. \]

\[ \text{M203} \]

\[ ^1 \text{H NMR (600 MHz, D}_2\text{O)} \delta 7.92 – 7.78 (m, 2H), 7.72 – 7.54 (m, 2H), 7.49 – 7.30 (m, 4H), 4.63 – 4.55 (m, 2H), 4.47 (d, } J = 8.6 \text{ Hz, 1H)}, 4.37 (m, 2H), 4.29 (m, 1H), 4.20 (m, 1H), 4.14 (d, } J = 10.0 \text{ Hz, 1H)}, 4.07 (s, 1H), 3.94 (m, 2H), 3.91 – 3.43 (m, 33H), 3.43 – 3.27 (m, 4H), 2.59 (dd, } J = 12.5, 4.5 \text{ Hz, 1H}), 1.98 (s, 3H), 1.96 (s, 3H), 1.94 (s, 3H), 1.68 (t, } J = 12.2 \text{ Hz, 1H)}, 0.97 (d, } J = 6.4 \text{ Hz, 3H). ESI-MS, calculated: 1362.4861; found [M-H]- 1361.4653, [M-2H]^{2-} 680.2283; MALDI-MS; found [M-H]- 1361.438. \]
**M204**

\(^1\)H NMR (600 MHz, D\(_2\)O) δ 7.90 – 7.68 (m, 2H), 7.67 – 7.50 (m, 2H), 7.46 – 7.26 (m, 4H), 5.03 (d, J = 3.5 Hz, 1H), 4.66 (s, 1H), 4.59 (d, J = 11.8 Hz, 1H), 4.54 – 4.42 (m, 2H), 4.36 (d, J = 7.8 Hz, 2H), 4.27 – 4.16 (m, 2H), 4.11 (d, J = 10.2 Hz, 2H), 3.99 (s, 1H), 3.94 – 3.55 (m, 18H), 3.99 – 3.55 (m, 8H), 1.96 (s, 3H), 1.94 (s, 3H), 1.10 (d, J = 6.4 Hz, 3H), 0.96 (d, J = 6.1 Hz, 3H).


**M205**

\(^1\)H NMR (600 MHz, D\(_2\)O) δ 7.84 – 7.72 (m, 2H), 7.66 – 7.48 (m, 2H), 7.43 – 7.26 (m, 4H), 4.98 (d, J = 3.8 Hz, 1H), 4.61 (s, 1H), 4.56 – 4.50 (m, 1H), 4.45 – 4.35 (m, 2H), 4.31 (d, J = 8.0 Hz, 1H), 4.24 – 4.17 (m, 1H), 4.14 (dd, J = 6.4, 2.1 Hz, 1H), 4.07 (d, J = 10.1 Hz, 1H), 4.02 (s, 1H), 3.98 (dd, J = 9.8, 2.8 Hz, 1H), 3.88 – 3.68 (m, 11H), 3.68 – 3.64 (m, 2H), 3.64 – 3.50 (m, 10H), 3.50 – 3.37 (m, 7H), 3.37 – 3.22 (m, 4H), 2.65 (dd, J = 12.4, 4.4 Hz, 1H), 1.91 (s, 3H), 1.90 (s, 3H), 1.88 (s, 3H), 1.71 (t, J = 12.2 Hz, 1H), 1.05 (d, J = 6.5 Hz, 3H), 0.92 (d, J = 6.4 Hz, 3H).

ESI-MS, calculated: 1508.5440; found [M-H] \(^{-}\) 1507.5277.

**M212**
$^1$H NMR (600 MHz, D$_2$O) δ 7.93 – 7.79 (m, 2H), 7.72 – 7.55 (m, 2H), 7.50 – 7.30 (m, 4H), 4.78 (d, $J = 5.7$ Hz, 1H), 4.64 (d, $J = 3.2$ Hz, 1H), 4.55 (d, $J = 10.7$ Hz, 2H), 4.52 – 4.43 (m, 2H), 4.43 – 4.3 (m, 2H), 4.33 – 4.24 (m, 2H), 4.20 (d, $J = 4.9$ Hz, 1H), 4.16 – 3.98 (m, 3H), 3.98 – 3.8 (m, 32H), 3.30 (t, $J = 9.8$ Hz, 1H), 2.68 (dd, $J = 12.5$, 4.4 Hz, 1H), 1.96 (s, 6H), 1.93 (s, 3H), 1.75 (t, $J = 12.2$ Hz, 1H), 1.00 (d, $J = 6.0$ Hz, 3H). ESI-MS, calculated: 1524.5390; found [M-H]$^-$ 1523.5161, [M-2H]$^2-$ 761.2533; MALDI-MS; found [M+Na]$^+$ 1547.593.

![Diagram of M213](image)

$^1$H NMR (600 MHz, D$_2$O) δ 7.88 – 7.75 (m, 2H), 7.68 – 7.51 (m, 2H), 7.49 – 7.29 (m, 4H), 4.78 – 4.72 (m, 1H), 4.70 – 4.65 (m, 1H), 4.61 – 4.51 (m, 1H), 4.47 (d, $J = 10.6$ Hz, 1H), 4.42 – 4.33 (m, 2H), 4.30 (d, $J = 7.7$ Hz, 1H), 4.28 – 4.18 (m, 2H), 4.12 (s, 2H), 3.99 – 3.8 (m, 35H), 3.31 (t, $J = 9.9$ Hz, 1H), 2.57 (dd, $J = 12.7$, 4.3 Hz, 1H), 1.98 (s, 3H), 1.96 (s, 3H), 1.93 (s, 3H), 1.73 (t, $J = 12.2$ Hz, 1H), 1.00 (d, $J = 6.2$ Hz, 3H). ESI-MS, calculated: 1524.5390; found [M-H]$^-$ 1523.5166, [M-2H]$^2-$ 761.2534; MALDI-MS; found [M-H]$^-$ 1523.553.

![Diagram of M214](image)

$^1$H NMR (600 MHz, D$_2$O) δ 7.94 – 7.78 (m, 2H), 7.72 – 7.54 (m, 2H), 7.54 – 7.30 (m, 4H), 5.03 (d, $J = 3.8$ Hz, 1H), 4.77 – 4.75 (m, 1H), 4.67 (s, 1H), 4.49 (d, $J = 7.9$ Hz, 1H), 4.36 (d, $J = 7.8$ Hz, 2H), 4.31 (d, $J = 7.7$ Hz, 1H), 4.27 (s, 1H), 4.21 (d, $J = 5.5$ Hz, 1H), 4.11 (d, $J = 11.3$ Hz, 1H), 4.03 (s, 1H), 3.97 – 3.38 (m, 33H), 3.30 (t, $J = 10.0$ Hz, 1H), 1.96 (s, 3H), 1.94 (s, 3H), 1.11
(d, \(J = 6.2\) Hz, 3H), 0.99 (d, \(J = 6.3\) Hz, 3H). ESI-MS, calculated: 1379.5014; found [M-H]⁻ 1378.4804.

**M215**

\(^1\)H NMR (600 MHz, D₂O) \(\delta\) 7.91 – 7.80 (m, 2H), 7.71 – 7.55 (m, 2H), 7.49 – 7.31 (m, 4H), 5.03 (d, \(J = 3.9\) Hz, 1H), 4.77 – 4.75 (m, 1H), 4.68 – 4.65 (m, 1H), 4.56 (dd, \(J = 10.9, 5.0\) Hz, 1H), 4.51 – 4.41 (m, 2H), 4.38 (d, \(J = 8.1\) Hz, 1H), 4.34 – 4.25 (m, 2H), 4.21 (dd, \(J = 6.3, 2.2\) Hz, 1H), 4.15 – 4.06 (m, 2H), 4.03 (dd, \(J = 9.8, 3.0\) Hz, 1H), 3.96 – 3.76 (m, 12H), 3.76 – 3.42 (m, 26H), 3.30 (t, \(J = 9.8\) Hz, 1H), 2.70 (dd, \(J = 12.4, 4.6\) Hz, 1H), 1.96 (s, 3H), 1.95 (s, 3H), 1.93 (s, 3H), 1.75 (t, \(J = 12.1\) Hz, 1H), 1.10 (d, \(J = 6.6\) Hz, 3H), 1.00 (d, \(J = 6.3\) Hz, 3H). ESI-MS, calculated: 1670.5969; found [M-H]⁻ 1669.5942, [M-2H]⁻ 834.2878.

**M223**

\(^1\)H NMR (600 MHz, D₂O) \(\delta\) 7.91 – 7.77 (m, 2H), 7.63 (ddd, \(J = 28.7, 16.5, 7.2\) Hz, 2H), 7.54 – 7.27 (m, 4H), 4.57 (dd, \(J = 11.0, 5.1\) Hz, 1H), 4.48 (d, \(J = 8.4\) Hz, 1H), 4.42 (d, \(J = 7.9\) Hz, 1H), 4.40 – 4.34 (m, 2H), 4.34 – 4.25 (m, 1H), 4.25 – 4.18 (m, 1H), 4.17 – 4.00 (m, 3H), 4.00 – 3.41 (m, 40H), 3.31 (t, \(J = 9.8\) Hz, 1H), 2.69 (dd, \(J = 12.5, 4.5\) Hz, 1H), 2.59 (dd, \(J = 12.5, 4.5\) Hz, 1H), 1.98 (s, 3H), 1.96 (s, 6H), 1.93 (s, 3H), 1.76 (t, \(J = 12.2\) Hz, 1H), 1.69 (t, \(J = 12.2\) Hz, 1H), 1.00 (d, \(J = 6.3\) Hz, 3H). ESI-MS, calculated: 1815.6344; found [M-H]⁻ 1814.6084, [M-2H]⁻ 906.7991.
M224

$^1$H NMR (600 MHz, D$_2$O) δ 7.94 – 7.81 (m, 2H), 7.78 – 7.56 (m, 2H), 7.50 – 7.30 (m, 4H), 5.07 – 4.99 (m, 2H), 4.54 – 4.46 (m, 2H), 4.46 – 4.33 (m, 2H), 4.29 (s, 1H), 4.26 – 4.14 (m, 1H), 4.08 (d, $J = 10.8$ Hz, 1H), 4.05 – 3.98 (m, 1H), 3.92 (t, $J = 12.1$ Hz, 1H), 3.89 – 3.55 (m, 28H), 3.55 – 3.38 (m, 9H), 3.28 (t, $J = 9.8$ Hz, 1H), 2.69 (dd, $J = 12.3$, 4.4 Hz, 1H), 2.00 – 1.86 (m, 9H), 1.73 (t, $J = 12.1$ Hz, 1H), 1.15 – 1.02 (m, 6H), 0.94 (d, $J = 6.3$ Hz, 3H). ESI-MS, calculated: 1670.5969; found [M-H]$^-$ 1669.5733, [M-2H]$^{2-}$ 834.2812.

M225

$^1$H NMR (600 MHz, D$_2$O) δ 7.93 – 7.80 (m, 2H), 7.72 – 7.55 (m, 2H), 7.53 – 7.33 (m, 4H), 5.03 (d, $J = 4.0$ Hz, 1H), 4.61 – 4.53 (m, 2H), 4.52 – 4.35 (m, 5H), 4.30 (s, 2H), 4.24 – 4.15 (m, 1H), 4.05 (m, 4H), 3.98 – 3.38 (m, 39H), 3.37 – 3.25 (m, 2H), 3.23 – 3.15 (m, 1H), 2.74 – 2.65 (m, 2H), 1.96 (s, 6H), 1.95 (s, 3H), 1.93 (s, 3H), 1.92 (s, 3H), 1.74 (t, $J = 11.9$ Hz, 2H), 1.10 (d, $J = 6.5$ Hz, 3H), 0.99 (d, $J = 6.4$ Hz, 3H). ESI-MS, calculated: 1961.6923; found [M-2H]$^{2-}$ 979.8275.

M234

$^1$H NMR (600 MHz, D$_2$O) δ 7.92 – 7.78 (m, 2H), 7.72 – 7.55 (m, 2H), 7.53 – 7.32 (m, 4H), 5.03 (d, $J = 3.8$ Hz, 1H), 4.56 (d, $J = 12.5$ Hz, 1H), 4.51 (d, $J = 8.3$ Hz, 1H), 4.45 – 4.33 (m, 3H),
4.29 (s, 2H), 4.21 (d, \(J = 6.3\) Hz, 1H), 4.15 – 4.02 (m, 2H), 4.02 – 3.35 (m, 38H), 3.35 – 3.26 (m, 1H), 2.60 (dd, \(J = 12.2, 4.6\) Hz, 1H), 1.97 (s, 3H), 1.96 (s, 3H), 1.95 (s, 3H), 1.70 – 1.63 (m, 1H), 1.10 (d, \(J = 6.3\) Hz, 3H), 0.99 (d, \(J = 5.4\) Hz, 3H). ESI-MS, calculated: 1670.5969; found [M-H]⁻ 1669.5733, [M-2H]²⁻ 834.2816; MALDI-MS; found [M-H]⁻ 1669.520.

\[\text{M235}\]

\(^1\text{H NMR (600 MHz, D}_2\text{O)} \delta 7.92 – 7.79 (m, 2H), 7.78 – 7.54 (m, 2H), 7.51 – 7.32 (m, 4H), 5.03 (d, \(J = 4.0\) Hz, 1H), 4.62 – 4.56 (m, 2H), 4.50 (d, \(J = 8.1\) Hz, 1H), 4.49 – 4.44 (m, 1H), 4.43 (d, \(J = 8.0\) Hz, 1H), 4.41 – 4.34 (m, 1H), 4.32 – 4.27 (m, 2H), 4.21 (d, \(J = 5.2\) Hz, 1H), 4.15 – 4.05 (m, 2H), 4.02 (dd, \(J = 10.1, 2.7\) Hz, 1H), 3.99 – 3.39 (m, 46H), 3.35 – 3.27 (m, 1H), 2.69 (dd, \(J = 12.4, 4.5\) Hz, 1H), 2.60 (dd, \(J = 12.1, 4.3\) Hz, 1H), 1.96 (s, 3H), 1.96 (s, 3H), 1.92 (s, 3H), 1.75 (t, \(J = 12.3\) Hz, 1H), 1.71 – 1.64 (m, 1H), 1.10 (d, \(J = 6.5\) Hz, 3H), 1.00 (d, \(J = 6.0\) Hz, 3H). ESI-MS, calculated: 1961.6923; found [M-2H]²⁻ 979.8283.

\[\text{M245}\]

\(^1\text{H NMR (600 MHz, D}_2\text{O)} \delta 7.87 – 7.75 (m, 2H), 7.68 – 7.48 (m, 2H), 7.46 – 7.26 (m, 4H), 4.96 (d, \(J = 3.2\) Hz, 2H), 4.60 – 4.58 (m, 2H), 4.48 – 4.39 (m, 3H), 4.36 (d, \(J = 7.6\) Hz, 2H), 4.31 (d, \(J = 9.5\) Hz, 2H), 4.22 (d, \(J = 5.9\) Hz, 2H), 4.16 – 4.15 (m, 1H), 4.13 (d, \(J = 5.7\) Hz, 1H), 4.02 (d, \(J = 10.4\) Hz, 1H), 3.95 (d, \(J = 9.9\) Hz, 2H), 3.88 – 3.61 (m, 18H), 3.61 – 3.49 (m, 9H), 3.49 – 3.29 (m, 9H), 3.27 – 3.17 (m, 2H), 2.62 (dd, \(J = 12.2, 3.7\) Hz, 1H), 1.89 (s, 6H), 1.87 (s, 3H), 1.66
(t, J = 12.3 Hz, 1H), 1.03 (d, J = 6.0 Hz, 6H), 0.91 (d, J = 6.2 Hz, 3H). ESI-MS, calculated: 1816.6548; found [M-H]− 1815.6307, [M-2H]2− 907.3100.

M301

$^1$H NMR (400 MHz, MeOD) δ 7.79 (d, J = 7.5 Hz, 2H), 7.74 – 7.58 (m, 2H), 7.44 – 7.27 (m, 4H), 4.55 (d, J = 7.7 Hz, 1H), 4.50 – 4.31 (m, 5H), 4.23 (t, J = 6.7 Hz, 1H), 4.18 – 4.02 (m, 2H), 4.01 – 3.84 (m, 5H), 3.84 – 3.51 (m, 14H), 3.51 – 3.30 (m, 6H), 2.08 (s, 3H), 2.06 (s, 3H), 1.22 (d, J = 5.9 Hz, 2H); $^{13}$C NMR (100 MHz, MeOD) δ 173.40, 172.73, 157.41, 144.06, 143.80, 141.18, 127.42, 126.85, 124.88, 119.55, 103.81, 101.78, 98.66, 79.69, 79.33, 76.38, 75.74, 75.04, 73.89, 73.41, 72.46, 71.20, 70.54, 69.85, 68.93, 68.64, 67.77, 66.59, 61.16, 60.88, 60.39, 55.28, 47.08, 22.35, 22.12, 18.44. ESI-MS, calculated: 1071.3907; found [M-H]− 1070.3777; MALDI-MS; found [M+Na]$^+$ 1094.388.

M302

$^1$H NMR (600 MHz, D$_2$O) δ 7.79 – 7.66 (m, 2H), 7.61 – 7.42 (m, 2H), 7.38 – 7.22 (m, 4H), 4.68 – 4.58 (m, 1H), 4.58 – 4.47 (m, 1H), 4.47 – 4.36 (m, 1H), 4.33 (d, J = 7.8 Hz, 1H), 4.29 (t, J = 7.7 Hz, 1H), 4.22 (d, J = 7.6 Hz, 1H), 4.20 – 3.96 (m, 3H), 3.93 – 3.80 (m, 1H), 3.80 – 3.34 (m, 3H), 3.34 – 3.20 (m, 3H), 2.63 (dd, J = 12.4, 4.1 Hz, 1H), 1.91 (s, 6H), 1.87 (s, 3H), 1.74 (t, J = 12.2 Hz, 1H), 0.94 (d, J = 6.2 Hz, 3H). ESI-MS, calculated: 1362.4861; found [M-H]$^-$ 1361.4634, [M-2H]$^{2-}$ 680.2272; MALDI-MS; found [M+H]$^+$ 1363.4380, [M+Na]$^+$ 1385.418.
M303

$^1$H NMR (600 MHz, D$_2$O) $\delta$ 7.92 – 7.73 (m, 2H), 7.73 – 7.49 (m, 2H), 7.49 – 7.29 (m, 4H), 4.68 (s, 1H), 4.60 – 4.43 (m, 2H), 4.36 (d, $J = 7.9$ Hz, 1H), 4.29 (d, $J = 6.5$ Hz, 1H), 4.28 – 4.18 (m, 3H), 4.18 – 4.05 (m, 3H), 4.04 – 3.88 (m, 4H), 3.88 – 3.41 (m, 21H), 3.41 – 3.26 (m, 4H), 2.59 (dd, $J = 12.4$, 3.8 Hz, 1H), 1.99 (s, 3H), 1.98 (s, 3H), 1.96 (s, 3H), 1.75 (t, $J = 12.1$ Hz, 1H), 1.02 (d, $J = 6.3$ Hz, 3H). ESI-MS, calculated: 1362.4861; found [M-H]$^- 1361.4637$, [M-2H]$^{2-} 680.2274$; MALDI-MS; found [M-H]$^- 1361.427$.

M304

$^1$H NMR (600 MHz, D$_2$O) $\delta$ 7.87 – 7.73 (m, 2H), 7.74 – 7.51 (m, 2H), 7.51 – 7.29 (m, 4H), 5.02 (d, $J = 3.9$ Hz, 1H), 4.75 (d, $J = 7.1$ Hz, 1H), 4.65 – 4.54 (m, 1H), 4.54 – 4.44 (m, 2H), 4.41 – 4.32 (m, 1H), 4.27 (d, $J = 7.8$ Hz, 1H), 4.21 (d, $J = 4.7$ Hz, 2H), 4.16 – 4.02 (m, 2H), 3.90 (d, $J = 10.4$ Hz, 1H), 3.89 – 3.29 (m, 26H), 1.96 (s, 3H), 1.93 (s, 3H), 1.10 (d, $J = 6.5$ Hz, 3H), 1.00 (d, $J = 6.3$ Hz, 3H). ESI-MS, calculated: 1217.4486; found [M+H]$^+$ 1218.4589, [M+Na]$^+$ 1240.4386; MALDI-MS, found [M-H]$^- 1216.436$.

M305

$^1$H NMR (600 MHz, D$_2$O) $\delta$ 7.77 – 7.59 (m, 2H), 7.59 – 7.37 (m, 2H), 7.37 – 7.15 (m, 4H), 4.98 (d, $J = 3.4$ Hz, 1H), 4.66 (s, 1H), 4.54 (s, 1H), 4.51 – 4.39 (m, 2H), 4.39 – 4.26 (m, 3H), 4.26
– 3.95 (m, 5H), 3.95 – 3.20 (m, 32H), 2.66 (dd, \( J = 12.6, 4.4 \) Hz, 1H), 1.93 (s, 6H), 1.89 (s, 3H), 1.80 (t, \( J = 12.2 \) Hz, 1H), 1.06 (d, \( J = 6.2 \) Hz, 3H), 0.97 (d, \( J = 6.2 \) Hz, 3H). ESI-MS, calculated: 1508.544; found [M-H]\(^{-}\) 1507.5194, [M-2H]\(^{2-}\) 753.2551.

**M312**

\(^1\)H NMR (600 MHz, D\(_2\)O) \( \delta 7.92 – 7.73 \) (m, 2H), 7.73 – 7.56 (m, 2H), 7.54 – 7.33 (m, 4H), 4.80 – 4.74 (m, 1H), 4.65 – 4.58 (m, 1H), 4.59 – 4.46 (m, 4H), 4.38 (d, \( J = 7.8 \) Hz, 2H), 4.34 – 4.23 (m, 2H), 4.19 (d, \( J = 5.0 \) Hz, 1H), 4.11 (d, \( J = 10.3 \) Hz, 1H), 3.99 (s, 1H), 3.95 – 3.78 (m, 7H), 3.78 – 3.42 (m, 26H), 3.39 – 3.26 (m, 1H), 2.61 (dd, \( J = 12.4, 4.4 \) Hz, 1H), 1.97 (s, 6H), 1.96 (s, 3H), 1.66 (t, \( J = 11.9 \) Hz, 1H), 0.99 (d, \( J = 6.2 \) Hz, 3H). ESI-MS, calculated: 1524.5390; found [M-H]\(^{-}\) 1523.5249, [M-2H]\(^{2-}\) 761.2589.

**M313**

\(^1\)H NMR (600 MHz, D\(_2\)O) \( \delta 7.94 – 7.82 \) (m, 2H), 7.75 – 7.57 (m, 2H), 7.52 – 7.31 (m, 4H), 4.75 – 4.73 (m, 1H), 4.69 – 4.67 (m, 1H), 4.52 (d, \( J = 7.7 \) Hz, 2H), 4.44 – 4.35 (m, 2H), 4.35 – 4.25 (m, 2H), 4.18 (d, \( J = 6.4 \) Hz, 1H), 4.11 (d, \( J = 10.4 \) Hz, 1H), 3.98 – 3.89 (m, 2H), 3.89 – 3.78 (m, 6H), 3.78 – 3.42 (m, 28H), 3.33 – 3.26 (m, 1H), 2.61 (dd, \( J = 12.8, 4.3 \) Hz, 1H), 1.98 (s, 6H), 1.96 (s, 3H), 1.65 (t, \( J = 11.9 \) Hz, 1H), 0.96 (d, \( J = 5.7 \) Hz, 3H). ESI-MS, calculated: 1524.5390; found [M-H]\(^{-}\) 1523.5148, [M-2H]\(^{2-}\) 761.2527; MALDI-MS; found [M+H]\(^{+}\) 1525.527, [M+Na]\(^{+}\) 1547.527.
M314

$^1$H NMR (600 MHz, D$_2$O) $\delta$ 7.89 – 7.78 (m, 2H), 7.70 – 7.53 (m, 2H), 7.49 – 7.28 (m, 4H), 5.02 (d, $J = 3.8$ Hz, 1H), 4.76 – 4.74 (m, 1H), 4.69 – 4.67 (m, 1H), 4.54 – 4.46 (m, 2H), 4.35 (d, $J = 8.2$ Hz, 2H), 4.32 – 4.17 (m, 4H), 4.09 (d, $J = 10.0$ Hz, 1H), 4.04 (s, 1H), 3.91 (d, $J = 11.2$ Hz, 2H), 3.88 – 3.69 (m, 10H), 3.69 – 3.46 (m, 14H), 3.46 – 3.26 (m, 6H), 1.96 (s, 3H), 1.93 (s, 3H), 1.10 (d, $J = 6.5$ Hz, 3H), 0.99 (d, $J = 6.3$ Hz, 3H). ESI-MS, calculated: 1379.5014; found [M-H]$^-$ 1378.4948.

M315

$^1$H NMR (600 MHz, D$_2$O) $\delta$ 7.88 – 7.74 (m, 2H), 7.69 – 7.50 (m, 2H), 7.47 – 7.30 (m, 4H), 5.01 (d, $J = 3.4$ Hz, 1H), 4.83 – 4.79 (m, 1H), 4.65 – 4.63 (m, 1H), 4.53 – 4.31 (m, 5H), 4.29 – 4.17 (m, 3H), 4.16 – 4.00 (m, 3H), 3.98 – 3.37 (m, 37H), 3.31 (t, $J = 10.4$ Hz, 1H), 2.69 (dd, $J = 12.7$, 4.4 Hz, 1H), 1.96 (s, 3H), 1.95 (s, 3H), 1.91 (s, 3H), 1.81 (t, $J = 12.3$ Hz, 1H), 1.09 (d, $J = 6.0$ Hz, 3H), 0.99 (d, $J = 6.3$ Hz, 3H). ESI-MS, calculated: 1670.5969; found [M-H]$^-$ 1669.5717, [M-2H]$^{2-}$ 834.2809; MALDI-MS; found [M+H]$^+$ 1669.519.

M323

$^1$H NMR (600 MHz, D$_2$O) $\delta$ 7.90 – 7.76 (m, 2H), 7.67 – 7.52 (m, 2H), 7.48 – 7.29 (m, 4H), 4.74 – 4.72 (m, 1H), 4.69 – 4.67 (m, 1H), 4.54 (d, $J = 14.9$ Hz, 2H), 4.50 (d, $J = 8.1$ Hz, 1H), 4.46
(d, J = 7.8 Hz, 1H), 4.36 (d, J = 7.8 Hz, 1H), 4.32 – 4.18 (m, 3H), 4.16 – 4.03 (m, 3H), 4.05 – 3.36 (m, 40H), 3.31 (t, J = 9.8 Hz, 1H), 2.80 – 2.63 (m, 1H), 2.62 – 2.53 (m, 1H), 1.96 (s, 9H), 1.95 (s, 3H), 1.80 (t, J = 12.1 Hz, 1H), 1.77 – 1.68 (m, 1H), 1.01 (d, J = 6.2 Hz, 3H). ESI-MS, calculated: 1815.6344; found [M-H]⁻ 1814.6059, [M-2H]²⁻ 906.7983.

**M324**

^1^H NMR (600 MHz, D₂O) δ 7.90 – 7.78 (m, 2H), 7.71 – 7.54 (m, 2H), 7.52 – 7.30 (m, 4H), 5.01 (d, J = 3.8 Hz, 1H), 4.58 – 4.52 (m, 2H), 4.49 (d, J = 8.0 Hz, 1H), 4.45 (d, J = 7.7 Hz, 2H), 4.35 (d, J = 8.1 Hz, 2H), 4.27 (d, J = 6.8 Hz, 2H), 4.20 (d, J = 5.5 Hz, 1H), 4.14 – 4.02 (m, 3H), 3.99 – 3.55 (m, 28H), 3.54 – 3.44 (m, 5H), 3.44 – 3.36 (m, 2H), 3.31 (t, J = 9.7 Hz, 1H), 2.68 (dd, J = 11.9, 3.4 Hz, 1H), 1.96 (s, 3H), 1.95 (s, 3H), 1.92 (s, 3H), 1.80 (t, J = 12.5 Hz, 1H), 1.09 (d, J = 6.5 Hz, 3H), 0.99 (d, J = 6.2 Hz, 3H). ESI-MS, calculated: 1670.5969; found [M-H]⁻ 1669.5715, [M-2H]²⁻ 834.2804; MALDI-MS, found [M+Na]^+ 1693.566, [M+K]^+ 1709.538.

**M325**

^1^H NMR (600 MHz, D₂O) δ 7.92 – 7.78 (m, 2H), 7.74 – 7.53 (m, 2H), 7.48 – 7.30 (m, 4H), 5.02 (d, J = 4.0 Hz, 1H), 4.55 (dd, J = 8.8, 4.4 Hz, 1H), 4.53 – 4.42 (m, 3H), 4.37 (dd, J = 17.1, 7.9 Hz, 2H), 4.34 – 4.24 (m, 2H), 4.24 – 4.15 (m, 1H), 4.15 – 3.99 (m, 4H), 3.99 – 3.39 (m, 42H), 3.31 (t, J = 9.7 Hz, 1H), 2.75 – 2.62 (m, 2H), 1.96 (s, 3H), 1.95 (s, 3H), 1.92 (s, 6H), 1.77 (t, J =
12.4 Hz, 2H), 1.09 (d, J = 6.3 Hz, 3H), 0.99 (d, J = 6.2 Hz, 3H). ESI-MS, calculated: 1961.6923; found [M-H]⁻ 1960.6640, [M-2H]²⁻ 979.8271.

M334

¹H NMR (600 MHz, D₂O) δ 7.91 – 7.78 (m, 2H), 7.72 – 7.53 (m, 2H), 7.52 – 7.29 (m, 4H), 5.02 (d, J = 3.7 Hz, 1H), 4.58 – 4.51 (m, 2H), 4.49 (d, J = 7.2 Hz, 1H), 4.41 – 4.33 (m, 2H), 4.31 – 4.17 (m, 3H), 4.16 – 4.05 (m, 2H), 4.01 – 3.44 (m, 36H), 3.44 – 3.37 (m, 2H), 3.31 (t, J = 9.9 Hz, 1H), 2.57 (dd, J = 12.7, 2.6 Hz, 1H), 1.97 (s, 3H), 1.96 (s, 3H), 1.92 (s, 3H), 1.73 (t, J = 12.3 Hz, 1H), 1.09 (d, J = 6.4 Hz, 3H), 1.00 (d, J = 6.3 Hz, 3H). ESI-MS, calculated: 1670.5969; found [M-H]⁻ 1669.5713, [M-2H]²⁻ 834.2806; MALDI-MS, found [M+Na]⁺ 1693.567, [M+K]⁺ 1709.894.

M335

¹H NMR (600 MHz, D₂O) δ 7.93 – 7.79 (m, 2H), 7.74 – 7.56 (m, 2H), 7.51 – 7.31 (m, 4H), 5.02 (d, J = 3.7 Hz, 1H), 4.60 – 4.53 (m, 1H), 4.53 – 4.43 (m, 2H), 4.43 – 4.34 (m, 2H), 4.34 – 4.25 (m, 1H), 4.21 (d, J = 2.7 Hz, 1H), 4.17 – 4.06 (m, 2H), 4.01 (d, J = 9.6 Hz, 1H), 3.93 (d, J = 10.0 Hz, 2H), 3.90 – 3.38 (m, 46H), 3.31 (t, J = 9.8 Hz, 1H), 2.70 (dd, J = 12.3, 4.5 Hz, 1H), 2.59 (dd, J = 12.4, 4.4 Hz, 1H), 1.98 (s, 3H), 1.96 (s, 6H), 1.92 (s, 3H), 1.76 (t, J = 11.9 Hz, 1H), 1.72 – 1.64 (m, 1H), 1.09 (d, J = 6.5 Hz, 3H), 0.99 (d, J = 6.2 Hz, 3H). ESI-MS, calculated: 1961.6923; found [M-H]⁻ 1960.6671, [M-2H]²⁻ 979.8270.
\(^1\text{H NMR (600 MHz, D}_2\text{O)} \delta 7.94 – 7.79 (m, 2H), 7.77 – 7.54 (m, 2H), 7.51 – 7.30 (m, 4H), 5.03 (d, \(J = 4.1\) Hz, 2H), 4.57 – 4.52 (m, 1H), 4.52 – 4.43 (m, 1H), 4.37 (dd, \(J = 14.6, 7.7\) Hz, 2H), 4.33 – 4.25 (m, 1H), 4.21 (dd, \(J = 6.3, 2.3\) Hz, 1H), 4.13 – 4.07 (m, 1H), 4.07 – 3.98 (m, 1H), 3.93 (d, \(J = 10.7\) Hz, 1H), 3.90 – 3.54 (m, 32H), 3.54 – 3.38 (m, 10H), 3.31 (t, \(J = 9.7\) Hz, 1H), 2.70 (dd, \(J = 12.5, 4.5\) Hz, 1H), 1.96 (s, 3H), 1.95 (s, 3H), 1.92 (s, 3H), 1.77 (t, \(J = 12.1\) Hz, 1H), 1.15 – 1.05 (m, 6H), 0.99 (d, \(J = 6.3\) Hz, 3H). ESI-MS, calculated: 1816.6548; found [M-H]\(^-\) 1815.6285, [M-2H]\(^2-\) 907.3094.
2 DESIGN, SYNTHESIS AND ACTIVITY STUDIES OF CELLULAR O-GLYCOME PRECURSOR

Protein O-glycosylation is a universal post-translational modification that plays an essential role in many biological regulations. Recently we reported a technology termed Cellular O-Glycome Reporter/Amplification (CORA) to amplify and profile mucin-type O-glycans from living cells. However, the application and development of the CORA method is limited by the precursor function. Here we discussed a rapid parallel synthesis of cellular O-glycome precursors via microwave assisted reaction. In total, 26 Ac₃GalNAc-α-Bn derivatives, including fluorescent and other reactive functional groups, were successfully synthesized. Furthermore, subsequent activity screening and evaluation of these precursors toward living cell and T-synthase were performed.

2.1 Introduction

2.1.1 Biosynthesis of O-GalNAc protein

Protein glycosylation is a widespread post-translational modification that play a key role in diverse fundamental cellular processes[44]. The highly regulated repertoire of cellular glycans, by binding to lectins and sterically modulating molecular interactions, were involved in multiple cellular mechanisms that contribute to health and disease. Because of the poor sensitivity of existing glycomics approaches and the lack of efficient and unbiased strategies for releasing glycans from complex samples, current technologies for evaluating glycans require relatively large amounts of biological samples for detailed structural analyses, which limits many potential applications. Hence, there is an urgent need to develop a simple and sensitive method for the analyzis of all the glycans synthesized by cells (the cellular glycome).[45]

GalNAc is linked to Ser or Thr residues with an α-linkage by a large family of over 20 different polypeptide N-acetylglactosamine transferases (ppGalNAcTs). Unlike most
glycosylation, which happens in the ER, GalNAc O-glycosylation occurs in the Golgi after protein folding.[46] Each ppGalNAcT has two domains: a catalytic glycosyltransferase domain and ricin-homology lectin domain. This lectin domain has been implicated in governing the selectivity and activity of several members of the family.[47] Unlike N-glycosylation and several other types of O-glycosylation, to date no consensus sequence has been identified for the addition of O-GalNAc; it has even been suggested that no such sequence exists and that ppGalNAcTs recognize and modify protein secondary structures, like the β-turn, instead of specific primary sequences.[48] What is known, as demonstrated by several studies discussed in greater detail later on, is that each ppGalNAcT homolog has a slightly different activity profile for peptide and/or glycopeptide substrates. Additionally, the expression of each homolog varies spatially across tissues and temporally during development.[46, 49] Such a high-level regulation during cell differentiation and organismal development is indicative of critical roles for these enzymes and the O-GalNAc modifications they impart. Interestingly, the deletion of the ppGalNAcT-1, T-4, T-5, or T-13 genes in mice fails to result in any observable phenotype.[50] This might indicate a high level of redundancy and overlap between the functions of each family member. However this situation does not hold for all systems studied, as the deletion of any one of five different Drosophila melanogaster O-GalNAc transferases is lethal[49] and inactivating mutations of ppGalNAcT-3 in humans cause the rare autosomal recessive metabolic disorder familial tumoral calcinosis.[51] Clearly, in certain cellular contexts, a single O-GalNAc transferase can have unique and critical functions. After the initial O-GalNAc transfer, further elaboration is accomplished in the Golgi apparatus by a suite of up to 30 glycosyltransferases to yield a huge variety of mature “mucin-type” O-GalNAc glycans (Fig 2.1).
2.1.2 Functional effects of O-GalNAc glycosylation

O-GalNAc glycosylation is the most common form of O-glycosylation in mammals, occurring in an estimated 10% of all mammalian proteins and half of all proteins passing through the secretory system. The most well-known examples of proteins carrying this type of glycosylation are the mucins, and in fact O-GalNAc glycosylation is referred to as “mucin-type” glycosylation in the literature due to the strength of this association. Mucins can be divided into two categories: secreted mucins that form extensive oligomers and result in a viscous, mucosal layer around tissues; and membrane-bound mucins that are monomeric and form a significant amount of the glycocalyx that surrounds and protects many cells. The most relevant feature of mucin proteins is the so-called “mucin-domain” which consists of a large number of repeats of a mucin-subtype-specific sequence. These repeated sequences, often referred to as “tandem repeats”, are heavily glycosylated with densely clustered O-GalNAc glycans. Additionally, there are a number of non-mucin proteins contain domains that are structurally very similar to the...
heavily glycosylated, tandem-repeat regions of mucins, and these are also referred to as “mucin domains”.\textsuperscript{[55a, 56]}

O-GalNAc glycans in these mucin domains are frequently acknowledged for their ability to prevent proteolytic degradation.\textsuperscript{[57]} Early light microscopy data indicated that mucin domains adopt a rigid, extended conformation in solution. A multitude of studies since then have established that this structure is a general feature of mucin domains and that this rod-like structure is dependent on the presence of O-GalNAc glycan clusters throughout the sequence.\textsuperscript{[58]} Therefore, the function of O-GalNAc glycosylation in these domains has traditionally been thought of as mainly structural: producing a stable, extended structure on the cell surface for display of other, more functionally important, protein domains, with little importance placed on the actual glycosylation pattern(s) present.\textsuperscript{[47a]} For example, glycans on extracellular proteins are known to inhibit oligomerization through what is probably the nonspecific charge–charge repulsion of negatively charged end groups.\textsuperscript{[59]} The fact that the sequence of mucins and mucin domains has been only lightly conserved during evolution has also been used as evidence for this assumption.\textsuperscript{[60]}

While there is undeniably an important structural aspect to mucin glycosylation, numerous examples have emerged of glycans acting in a much more specific manner, for example, it has been suggested that they are integral to the molecular code used in everyday cellular communication and recognition.\textsuperscript{[61]} As evidence of these more specific roles, the glycosylation patterns of many proteins appears to be tightly regulated during cellular maturation,\textsuperscript{[62]} and both temporally and spatially during development.\textsuperscript{[63]} Glycan patterns on mucins have also been shown to vary in a region-specific way in several different body systems.\textsuperscript{[56, 64]} The intestinal tract is a prime example, where different core structures are specific to each region and sialic acids appear in a well-defined gradient increasing from the ileum to the colon.\textsuperscript{[65]} Mucin-type glycosylation is
also known to be important in governing selectivity in cell–cell communication and signaling in several different systems including T cell maturation,[66] Notch signaling,[67] leukocyte migration[68] and mucin signaling.[69] This type of glycosylation is critical to proper processing of many proteins from the proprotein form to maturity[70] and important in regulating cleavage by a disintegrin and metalloprotease (ADAM) proteases and subsequent shedding of the extracellular domains of many diverse membrane proteins including Notch[67] and TNF-α.[71] Interestingly, this regulation has been observed to be both positive and negative, so it is not a simple steric-interference mechanism and may involve specific binding interactions.[71] Cytokines, secreted soluble proteins that are involved in cell communication, are also known to be regulated through several types of glycosylation, including O-glycosylation with mucin-type glycans.[72] Chemical biology has clearly aided in the studies of O-GalNAc glycosylation. As discussed below, much of the early work on structural and functional effects of O-GalNAc glycosylation focused on small glycopeptides. The readily available and pure homogenous glycoforms made it possible to draw strong conclusions about the different substrate specificities of ppGalNAcTs and the ability of O-GalNAc glycans to affect the structural and functional properties of proteins.

Chemical biology has also contributed to the study of functional aspects of mucin-type glycans in biological systems. As mentioned earlier, glycosylation patterns on extracellular membrane proteins are very important for cellular communication and adhesion. For example, during inflammation, leukocytes adhere to endothelial cells near the site of injury before migration towards the injury site.[73] This adherence is the result of P-selectin on the endothelial cells binding specific ligands on leukocytes, the most important of which is known as P-selectin glycoprotein ligand-1 (PSGL-1). P-Selectin binds to the extreme terminus of PSGL-1, and both tyrosine sulfation and O-glycosylation of PSGL-1 in this region have been shown to be critical for high-
affinity binding in studies of genetic knockout models.\textsuperscript{[73]} To identify the particular modifications required for binding in this context, the Cummings group chemo-enzymatically synthesized a 23-mer glycosulfopeptide based on the N-terminal sequence of PSGL-1 that contain three sulfotyrosines and a sialyl Lewis x (sLex) hexasaccharide.\textsuperscript{[74]} To avoid problems with site-specificity of ppGalNAcTs, the peptide was synthesized on solid phase with the first GalNAc at position Thr 57 in place. Enzymatic transformations by a series of five glycosyltransferases constructed the full hexasaccharide, and lastly, addition of the sulfate groups was also accomplished enzymatically. It was found that the sulfotransferase was able to add sulfates to all three tyrosines in the glycopeptide simultaneously.\textsuperscript{[74a]} What’s more\textsuperscript{[74b]}, the authors synthesized aglycosulfopeptide with a core-1 sLeX-containing hexasaccharide and the same three sulfotyrosines through a very similar chemoenzymatic approach. The authors then tested the two final glycosulfopeptides for binding affinity to P-selectin and found that only one which construct with the core-2 glycan structure was bound to a measurable degree.\textsuperscript{[74]} In addition, when the terminal sialic acid was removed from the glycan, binding was abolished. Moreover, they found that the synthetic intermediate carrying the full-length glycan, but not the sulfates, did not bind, nor did the isolated hexasaccharide bind. The results of this study indicate that both modifications, sulfation and glycosylation, are critical to high-affinity binding between PSGL-1 and P-selectin. However, the authors were unable to chemo-enzymatically prepare any partially-sulfated glycopeptide products, and they could not address the question of whether or not specific sulfotyrosines were more important to binding than others. This question was solved in a follow-up study by the same group that used chemical synthesis to introduce the sulfo-tyrosine residues in a completely controlled manner.\textsuperscript{[75]} In this study they synthesized a large library of glycosulfopeptides containing each possible combination of 0–3 sulfate groups and several
different glycan structures. Characterization of each construction and comparison of the results revealed that the most important determinant of binding was the presence of a fucose in the sLeX epitope on the peptide, although the number and pattern of sulfate groups as well as the terminal sialic acid had strong effects on binding. Also noteworthy is the site-specific nature of this glycan dependence, moving the full-length glycan from the natural glycosylation site to a nearby site (both threonines) resulted in complete loss of binding.\textsuperscript{[75]}

2.1.3 CORA technology

Present on over 80% of proteins\textsuperscript{[76]} that traverse the secretory apparatus, mucin type O-glycan with O-GalNAc-linkage to Ser/Thr/Tyr residues is one of the most common types of protein glycosylation and is important in many normal and pathologic settings\textsuperscript{[77]}. In contrast to N-glycans, which can be released enzymatically, O-glycans require chemical strategies for their release—primarily alkaline $\beta$-elimination, which is inefficient and may result in O-glycan degradation. To address these problems, we recently developed a technology termed Cellular O-glycome Reporter/Amplification (CORA)$^{[78]}$, which used Ac$_3$GalNAc-$\alpha$-Bn as precursor to amplify and profile the mucin-type O-glycome in living cell, resulting 100-1000x increased sensitivity over traditional methods (Fig 2.2). Isotope-Cellular O-glycome Reporter Amplification (ICORA) was developed later to analyze the O-glycome from cells by isotope labeling\textsuperscript{[79]}. We noticed that CORA was showed effective at concentrations <250 $\mu$M and short incubation times with no observed side effect on cellular properties or glycosylation. Prior studies showed that treating cells with high concentrations of Bn-$\alpha$-GalNAc for O-glycan inhibition produced mainly short Gal-$\beta$3-GalNAc-$\alpha$1-O-Bn and sialyl-Gal-$\beta$3-GalNAc-$\alpha$1-O-Bn derivatives.\textsuperscript{[80]} However, our success with CORA using low concentrations of this Ac$_3$GalNAc-$\alpha$-Bn derivative as a precursor indicates that it is well taken up by live cells, converted to Bn-$\alpha$-GalNAc, efficiently by T-synthase and further accessed
by a wide range of enzymes in the secretory pathway, including the most terminal types of glycan modifications and extensions. But in the former work the precursor is limited to Ac$_3$GalNAc-α-Bn, which hindered many potential applications of CORA.

**Figure 2.2 Principle of CORA technology**

Herein, we described a rapid parallel synthesis of cellular O-glycome precursors via microwave assisted reaction. By this synthesis method, a library of Ac$_3$GalNAc-α-Bn derivatives was successfully synthesized, including fluorescent and other well-known functional groups. Subsequent screening of CORA activity on living cell and enzymatic activity on T-synthase were carried out to evaluate of these derivatives. With these new precursors with functional groups in hand, the application of CORA is widespread, such as analysis and isolation of glycan on HPLC, interrogation by glycan-binding proteins on microarray, secretory biological pathway study by fluorescent labeling, identify structure by IsoTaG and so on.
2.2 Result and discussion

2.2.1 Design and Synthesis of CORA precursor candidates.

In former work, the application of CORA was limited by the only precursor Ac$_3$GalNAc-α-Bn with weak UV and no active functional group. In this work, we structure-based designed 26 Ac$_3$GalNAc-α-Bn derivatives with active functional groups (Figure 2.3). The template of Ac$_3$GalNAc-α-Bn was reserved to keep the CORA activity to the greatest extent. These functional groups on the derivatives is designed to be used by well-established methods such as fluorescent labeling strategy (fluorescent group; 20 and 21), copper-catalyzed azide–alkyne cycloaddition (azide and alkyne groups; 18, 23, 24, 25, 26 and 27), Raman reporter strategy (nitrile group; 12), IsoTaG strategy (bromide group; 2, 3, 4, 5, 7, 8, 10 and 22), and Diels–Alder reaction (alkene group; 18).

![Figure 2.3 Structure of designed Ac$_3$GalNAc-α-Bn derivatives.](image)

As shown in scheme 1, compound 2-22 were synthesized by microwave reaction described by Beau and co-workers to achieve α-glycosidation from glycosamine pentaacetates$^{[81]}$. This one
step reaction avoids the use of 2-azido-2-deoxy or other nonparticipating temporary groups at C-2. By controlling the temperature and reaction time, we got desired product in acceptable yield. We also tried the reflux method developed by Du and co-workers\cite{82}, but we got complex mixture of decomposition products for some compounds. Microwave assisted reaction also showed shorter reaction time and better α:β ratio due to high reaction temperature. Compounds with azido group cannot be achieved by Fig 2.4 directly, due to reaction between azido group and Cu(OTf)$_2$. Compound 23-27 were prepared as outline in Fig 2.5. Compound 11 was treated with NaN$_3$ in MeCN to give compound 23. The nitro group on compound 13-16 were modified to azido group to produce Compound 24-27.

**Figure 2.4** Microwave assisted synthesis of Ac$_3$GalNAc-α-Bn derivatives.

**Figure 2.5** Synthesis of Ac$_3$GalNAc-α-Bn derivatives with azido group.

### 2.2.2 CORA activity and T-synthase activity assays.

Having these Ac$_3$GalNAc-α-Bn derivatives in hand, we first would like to determine if they are the substrate for T-synthase, since T-synthase activity is necessary for producing Bn-O-
<table>
<thead>
<tr>
<th>Compounds</th>
<th>Relative acceptor activity (RAA)</th>
<th>Relative CORA precursor activity (RCPA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1.81</td>
<td>2.17</td>
</tr>
<tr>
<td>3</td>
<td>1.93</td>
<td>2.19</td>
</tr>
<tr>
<td>4</td>
<td>1.99</td>
<td>2.85</td>
</tr>
<tr>
<td>5</td>
<td>1.55</td>
<td>2.34</td>
</tr>
<tr>
<td>6</td>
<td>1.19</td>
<td>2.53</td>
</tr>
<tr>
<td>7</td>
<td>1.67</td>
<td>2.89</td>
</tr>
<tr>
<td>8</td>
<td>1.96</td>
<td>2.09</td>
</tr>
<tr>
<td>9</td>
<td>1.92</td>
<td>2.65</td>
</tr>
<tr>
<td>10</td>
<td>2.17</td>
<td>0.05</td>
</tr>
<tr>
<td>11</td>
<td>0.59</td>
<td>0.05</td>
</tr>
<tr>
<td>12</td>
<td>1.63</td>
<td>0.05</td>
</tr>
<tr>
<td>13</td>
<td>4.69</td>
<td>1.30</td>
</tr>
<tr>
<td>14</td>
<td>2.26</td>
<td>0.23</td>
</tr>
<tr>
<td>15</td>
<td>3.01</td>
<td>0.45</td>
</tr>
<tr>
<td>16</td>
<td>3.01</td>
<td>0.21</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>1.64</td>
<td>1.86</td>
</tr>
<tr>
<td>19</td>
<td>11.00</td>
<td>0.05</td>
</tr>
<tr>
<td>20</td>
<td>13.38</td>
<td>0.05</td>
</tr>
<tr>
<td>21</td>
<td>1.91</td>
<td>1.41</td>
</tr>
<tr>
<td>22</td>
<td>3.85</td>
<td>4.29</td>
</tr>
<tr>
<td>23</td>
<td>1.34</td>
<td>0.97</td>
</tr>
<tr>
<td>24</td>
<td>1.59</td>
<td>0.79</td>
</tr>
<tr>
<td>25</td>
<td>1.29</td>
<td>1.58</td>
</tr>
<tr>
<td>26</td>
<td>1.28</td>
<td>1.68</td>
</tr>
</tbody>
</table>
glycans in CORA procedure, and cell lines without functional Cosmc cannot secret Bn-O-glycans. Therefore, we assayed the acceptor activity of all compounds for T-synthase by using HPLC. The procedure was as follows: 0.1 mM compound and 0.5 mM UDP-Galactose were catalyzed by 0.4 ng/mL T-synthase (provided by Dr. cummings’ lab) in 20 mM MES buffer pH 6.8 (containing 20 mM MnCl, 0.5% Triton X-100) at 37 °C for 18 h, then the reaction was quenched by equal volume of methanol, and stored in -80 °C for overnight. The freezed samples were unfroze and centrifuged at 14,000 rpm for 5 min, then the supernatants were loaded onto C18 (XBridge TM, 35 mm, 30x150 mm) column. SHIMADZU HPLC system was used, and mobile phase A consisted of 0.1% TFA in Acetonitrile and mobile phase B was 0.1% TFA in water. A linear gradient from 5 % to 60% A over 30 minutes at a flow rate of 0.4 mL/min was applied. Soft-ware EZStart 7.4 was used for data analysis. The acceptor activity of Bn-α-GalNAc was set as 1, the RAA (relative acceptor activity) of compound 2-27 was calculated and listed in Table 2.1. The data showed that most of the compounds were better accepter than Bn-α-GalNAc for T-synthase, especially compounds 19 and 20, which represented more than 10-fold accepter activity than compound 1. However, compound 11 of which RAA value was 0.59 was not as good as Bn-α-GalNAc, and compound 17 whose RAA value was zero was not an ac-ceptor for T-synthase.

Due to the promising RAA of most compounds, we had high expectations for their CORA precursor activity. Therefore, we further tested their CORA precursor activity in living cell. Firstly, we treated A549 Cell with 50 µM compound 2-27 to see if they can produce corresponding O-glycans like the known CORA precursor Bn-α-GalNAc. The result showed that compound 2-9, 13-16, 18, 21-27 can produce the corre-sponding O-glycan peaks like compound 1.Taken
compound 22 as an example, it produced a similar MS profile to Bn-α-GalNAc. As expected, compound 17 which is not the acceptor for T-synthase cannot work as CORA pre-cursor, and compound 11 whose RAA is less than 1 can only produce a few O-glycans in low MS intensity. Interestingly, compound 10, 12, 19 and 20 of which RAA are more efficient than Bn-α-GalNAc can only make little and weak O-glycan peaks in MS profile, particularly compound 19 and 20, which are the best two acceptors for T-synthase, can only produce core 1 and sialyl core 1 O-glycans. These data indicate that these compounds cannot be processed further in biological system. In order to evaluate the exact CORA precursor activity, we developed the following experiment. A549 cells were incubated with compound & Bn-α-GalNAc mixture at the ratio of 1:1 (50 mM: 50 mM) for 3 days, then the media was collected and went through the Amicon OR 10K centrifugal filter (Millipore), then purified by 50 mg Sep-PakOR C18 cartridge (Waters), and analyzed by MALDI-MS-TOF. The CORA precursor activity of Bn-α-GalNAc was set as 1, the relative CORA precursor activity (RCPA) was calculated by formulation: RCPA of compound = the total MS intensity of all derivatized-O-Glycan peaks/ the total MS intensity of all Bn-O-Glycan peaks. The MS intensity ratio of the pure compound to Bn-α-GalNAc was used for normalization. Every experiment was repeated for 3 times. The RCAP data in table 1 showed that compound 2-9, 13, 18, 21, 22, 25-27 performed stronger CORA precursor activity than standard compound 1,
while compound 14-16, 24 were less efficient than Bn-α-GalNAc, compound 10-12, 19, 20 barely had the CORA precursor activity.

**Figure 2.6** CORA activity and T-synthase activity of designed precursors.

### 2.2.3 **SAR study and evaluation of the precursor activity.**

The results showed that compound 2-9 with halogen substitution on aromatic ring performed stronger CORA activity than standard compound 1, which suggested that halogen on the aromatic ring contributes to improve the binding with glycosyltransferase. The compound with Br may be used for IsoTaG glycoproteomics platform probe. Derivatives 10, 11 and 12 don’t have CORA activity but showed substrate activity for T-synthase, which might result from metabolism during the secretory pathway in the cell. Precursor 13, 14, 15 and 16 bearing nitro group on phenyl ring showed better results from enzymatic substrate activity than CORA activity, which due to kind of prevention of nitro group during the secretory pathway. Compound 17 was not recognized by T-synthase which certainly led to no CORA activity. Most interestingly, Candidates 18, 23, 24, 25, 26 and 27 bearing with alkyne or azido group showed CORA activity and some of them even better than standard precursor compound 1, with which we can add a fluorescent tag by click reaction in the future. Compound 19 and 20 showed no CORA activity but very good substrate
activity, which suggested that phenol group got blocked during the secretory process. However, compound 19 and 20 with strong substrate activity may act as potent and competitive inhibitor of T-synthase, which is never reported before. Precursor 21 with fluorescence itself also showed CORA activity, which could be an ideal fluorescent label precursor without any additional reaction. Candidates 22 and 27 showed better CORA activity and substrate activity than candidates 4 and 26, which suggested longer linker between GalNAc and aromatic ring is good for enzyme to take up substrate. For the compounds with T-synthase activity but no or less CORA activity, we assumed they were metabolized in the cell since they should get through the cell membrane and go into Golgi easily.

2.3 Conclusion

In summary, we designed a rapid parallel synthesis of a library of cellular O-glycome precursor via microwave assisted reaction. A total number of 26 Ac$_3$GalNAc-$\alpha$-Bn derivatives, containing fluorescent and other popular functional groups, were successfully synthesized by this microwave assisted method. In addition, subsequent assays of CORA activity on living cell and enzymatic activity on T-synthase were tested to evaluate of these derivatives. With these new functional precursors in hand, the widespread application of CORA on many of areas is predicted, such as analysis and isolation of glycan on HPLC, interrogation by glycan-binding proteins on microarray, secretory biological pathway study by fluorescent labeling, identify structure by IsoTaG and so on.

2.4 Experiment section

2.4.1 Chemical synthesis

Materials:
All chemicals were purchased as reagent grade and used without further purification. Anhydrous dichloromethane (CH$_2$Cl$_2$), acetonitrile (CH$_3$CN), tetrahydrofuran (THF), $N$,$N$-dimethyl formamide (DMF), diethyl ether (Et$_2$O), toluene, and methanol (MeOH) were purchased from a commercial source without further distillation. Pulverized Molecular Sieves MS-4 Å (Aldrich) for glycosylation was activated by heating at 350 °C for 3 h. Reactions were monitored by analytical thin-layer chromatography (TLC) in EM silica gel 60 F254 plates and visualized under UV (254 nm) and/or by staining with acidic ceric ammonium molybdate or $p$-anisadehyde. Flash chromatography was performed on silica gel (Merck) of 40-63μm particle size and P2 gel (Biorad). $^1$H and $^{13}$C NMR spectra were recorded on a Bruker AVANCE 400 (400 MHz), and Bruker AVANCE 600 (600 MHz) spectrometer at 25 °C. All $^1$H Chemical shifts (in ppm) were assigned according to CDCl$_3$ ($\delta$= 7.24 ppm) and D$_2$O ($\delta$ = 4.79 ppm) and all $^{13}$C NMR was calibrated with CDCl$_3$ ($\delta$= 77.00 ppm). Coupling constants ($J$) are reported in hertz (Hz). Splitting patterns are described using the following abbreviations: s, singlet; brs, broad singlet; d, doublet; t, triplet; q, quartet; dd, doublet of doublet; m, multiplet. $^1$H NMR spectra are reported in the following order: chemical shift, multiplicity, coupling constant(s), and number(s) of protons. All NMR signals were assigned on the basis of $^1$H NMR, COSY, HSQC, HMQC, and $^{13}$C NMR experiments. HPLC-MS experiments were performed on an LTQ-Orbitrap Elite mass spectrometer (Thermo Fisher) equipped with EASY-spray source and nano-LC UltiMate 3000 high-performance liquid chromatography system (Thermo Fisher). Samples were transmitted into MS with a silica column. LTQ-Orbitrap Elite mass spectrometer was operated in the data-dependent mode. A full-scan survey MS experiment (m/z range was set according to the molecular weight of O-mannose glycan; automatic gain control target, 1,000,000 ions; resolution at 400 m/z, 240,000; maximum ion accumulation time, 200 ms) was acquired by the Orbitrap mass
MALDI-TOF MS analyses were performed on UltrafleXtreme MALDI TOF/TOF Mass Spectrometer (Bruker). Scan range of MS was set according to the molecular weight of O-mannose glycans, and reflector mode was used for O-mannose glycan analysis. Mass spectra were obtained in both positive and negative extraction mode with the following voltage settings: ion source 1 (19.0 kV), ion source 2 (15.9 kV), and lens (9.3 kV). The reflector voltage was set to 20 kV. The laser was pulsed at 7 Hz and the pulsed ion extraction time was set at 400 ns. The laser power was kept in the range of 40–90%.

**General Procedures**

**A) Microwave assisted synthesis of Ac3GalNAc-α-Bn derivatives**

A mixture of peracetylated galactosamine (1 mmol), acceptor (4 mmol) and copper(II) trifluoromethanesulfonate (0.15 mmol) in dry 1,2-Dichloroethane was stirred at 130 °C in microwave reactor for 25 min. The resulting solution was washed with saturated aqueous NaHCO₃ and brine solution, dried over Na₂SO₄, filtered, and concentrated in vacuo. The resulting residue purified on a silica gel column to get the final product. The alcohol acceptors and corresponding yield were shown in Table S1.

**B) Synthesis of precursor with azidobenzene group**

The precursor with nitrobenzene group (1 mmol) was dissolved in THF at room temperature, followed by addition of ammonium chloride (3 mmol) and Zn dust (<10 micron, 10 mmol). After being stirred overnight, the mixture was concentrated in vacuo to give a residue for the next step without further purification.

**C) Deacetylation reaction**
The peracetylated compound was dissolved in MeOH, and NaOMe in MeOH was added until pH was 10. After stirring at room temperature for 3 h, the solution was neutralized with ion-exchange resin (H+), then filtered and concentrated in vacuo to produce the product.

2-Bromobenzyl 2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-\(\alpha\)-D-galactopyranoside (2)

\(^1\)H NMR (400 MHz, CDCl\(_3\)) \(\delta 7.61\) (d, \(J = 7.8\) Hz, 1H), \(7.42 - 7.31\) (m, 2H), \(7.31 - 7.20\) (m, 1H), \(5.70\) (d, \(J = 9.7\) Hz, 1H), \(5.39\) (d, \(J = 2.5\) Hz, 1H), \(5.18\) (dd, \(J = 11.3, 3.3\) Hz, 1H), \(4.99\) (d, \(J = 3.7\) Hz, 1H), \(4.78\) (d, \(J = 11.7\) Hz, 1H), \(4.66 - 4.53\) (m, 2H), \(4.27\) (t, \(J = 6.7\) Hz, 1H), \(4.12\) (dd, \(J = 6.5, 2.4\) Hz, 2H), \(2.17\) (s, 3H), \(2.07\) (s, 3H), \(1.99\) (s, 3H), \(1.94\) (s, 3H); \(^{13}\)C NMR (101 MHz, CDCl\(_3\)) \(\delta 171.05, 170.57, 170.48, 170.17, 135.76, 133.28, 130.98, 130.34, 127.83, 124.26, 97.12, 69.92, 68.64, 67.46, 67.23, 62.07, 47.73, 23.44, 20.89.

3-Bromobenzyl 2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-\(\alpha\)-D-galactopyranoside (3)

\(^1\)H NMR (400 MHz, CDCl\(_3\)) \(\delta 7.55 - 7.41\) (m, 2H), \(7.25\) (dd, \(J = 4.7, 2.4\) Hz, 2H), \(5.80\) (d, \(J = 9.5\) Hz, 1H), \(5.39\) (d, \(J = 2.3\) Hz, 1H), \(5.18\) (dd, \(J = 11.4, 3.2\) Hz, 1H), \(4.98\) (d, \(J = 3.7\) Hz, 1H), \(4.69\) (d, \(J = 12.1\) Hz, 1H), \(4.60\) (ddd, \(J = 11.4, 9.6, 3.7\) Hz, 1H), \(4.49\) (d, \(J = 12.1\) Hz, 1H), \(4.20\) (t, \(J = 6.5\) Hz, 1H), \(4.16 - 4.02\) (m, 3H), \(2.16\) (s, 3H), \(2.06\) (s, 4H), \(1.99\) (s, 3H), \(1.96\) (s, 3H); \(^{13}\)C NMR (101 MHz, CDCl\(_3\)) \(\delta 170.92, 170.45, 170.35, 170.13, 138.98, 131.41, 131.06, 130.29, 126.68, 122.75, 97.06, 69.07, 68.25, 67.34, 67.04, 61.95, 47.74, 23.27, 20.77, 20.75.
4-Bromobenzyl 2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-α-D-galactopyranoside (4)

$^1$H NMR (400 MHz, CDCl$_3$) δ 7.52 (d, J = 8.3 Hz, 2H), 7.19 (d, J = 8.3 Hz, 2H), 5.55 (d, J = 9.7 Hz, 1H), 5.38 (d, J = 2.5 Hz, 1H), 5.17 (dd, J = 11.4, 3.2 Hz, 1H), 4.97 (d, J = 3.7 Hz, 1H), 4.66 (d, J = 11.9 Hz, 1H), 4.60 (ddd, J = 11.3, 9.8, 3.7 Hz, 1H), 4.46 (d, J = 11.9 Hz, 1H), 4.18 (t, J = 6.5 Hz, 1H), 4.16 – 4.03 (m, 2H), 2.17 (s, 3H), 2.06 (s, 3H), 2.00 (s, 3H), 1.93 (s, 3H); $^{13}$C NMR (101 MHz, CDCl$_3$) δ 171.08, 170.52, 170.43, 170.07, 135.68, 131.97, 129.93, 122.49, 97.20, 69.44, 68.40, 67.41, 67.13, 61.99, 47.87, 23.39, 20.87, 20.83.

4-Fluorobenzyl 2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-α-D-galactopyranoside (5)

$^1$H NMR (400 MHz, CDCl$_3$) δ 7.30 (dd, J = 8.6, 5.5 Hz, 2H), 7.07 (t, J = 8.6 Hz, 2H), 5.62 (d, J = 9.6 Hz, 1H), 5.38 (d, J = 2.8 Hz, 1H), 5.17 (dd, J = 11.4, 3.2 Hz, 1H), 4.98 (d, J = 3.7 Hz, 1H), 4.68 (d, J = 11.6 Hz, 1H), 4.59 (ddd, J = 11.3, 9.8, 3.7 Hz, 1H), 4.48 (d, J = 11.6 Hz, 1H), 4.20 (t, J = 6.5 Hz, 1H), 4.17 – 4.03 (m, 2H), 2.16 (s, 3H), 2.06 (s, 4H), 1.99 (s, 3H), 1.92 (s, 3H); $^{13}$C NMR (101 MHz, CDCl$_3$) δ 171.06, 170.52, 170.44, 170.08, 135.68, 131.97, 129.93, 122.49, 97.20, 69.47, 68.44, 67.43, 67.13, 61.99, 47.85, 23.34, 20.83, 20.80.

4-Chlorobenzyl 2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-α-D-galactopyranoside (6)
$^1$H NMR (400 MHz, CDCl$_3$) δ 7.47 – 7.32 (m, 2H), 7.32 – 7.21 (m, 2H), 5.58 (d, J = 9.6 Hz, 1H), 5.38 (d, J = 2.4 Hz, 1H), 5.18 (dd, J = 11.4, 3.3 Hz, 1H), 4.98 (d, J = 3.7 Hz, 1H), 4.68 (d, J = 11.8 Hz, 1H), 4.60 (ddd, J = 11.3, 9.7, 3.7 Hz, 1H), 4.48 (d, J = 11.8 Hz, 1H), 4.19 (t, J = 6.5 Hz, 1H), 4.16 – 4.02 (m, 2H), 2.17 (s, 3H), 2.06 (s, 3H), 1.99 (s, 3H), 1.93 (s, 3H); $^{13}$C NMR (101 MHz, CDCl$_3$) δ 171.10, 170.53, 170.45, 170.08, 135.18, 134.40, 129.65, 129.02, 97.19, 69.42, 68.42, 67.43, 67.14, 62.00, 47.89, 23.40, 20.87, 20.83.

3-Bromo-4-fluorobenzyl

2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-$\alpha$-D-galactopyranoside (7)

$^1$H NMR (400 MHz, CDCl$_3$) δ 7.54 (dd, J = 6.5, 2.1 Hz, 1H), 7.24 (ddd, J = 8.3, 4.6, 2.1 Hz, 1H), 7.13 (t, J = 8.3 Hz, 1H), 5.79 (d, J = 9.5 Hz, 1H), 5.39 (d, J = 2.5 Hz, 1H), 5.18 (dd, J = 11.4, 3.3 Hz, 1H), 4.98 (d, J = 3.7 Hz, 1H), 4.66 (d, J = 12.0 Hz, 1H), 4.59 (ddd, J = 11.4, 9.5, 3.7 Hz, 1H), 4.47 (d, J = 12.0 Hz, 1H), 4.19 (t, J = 6.6 Hz, 1H), 4.16 – 4.03 (m, 2H), 2.16 (s, 3H), 2.06 (s, 4H), 2.00 (s, 3H), 1.96 (s, 3H); $^{13}$C NMR (101 MHz, CDCl$_3$) δ 170.94, 170.43, 170.33, 170.11, 160.16, 157.69, 134.23, 134.19, 133.28, 128.86, 128.78, 116.79, 116.57, 109.42, 109.21, 97.02, 68.52, 68.20, 67.31, 67.07, 61.95, 47.79, 23.24, 20.76, 20.73.

4-Bromo-2-fluorobenzyl

2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-$\alpha$-D-galactopyranoside (8)
$^1$H NMR (400 MHz, CDCl$_3$) $\delta$ 7.40 – 7.27 (m, 2H), 7.22 (t, $J = 7.9$ Hz, 1H), 5.67 (d, $J = 9.6$ Hz, 1H), 5.38 (d, $J = 2.4$ Hz, 1H), 5.14 (dd, $J = 11.4$, 3.3 Hz, 1H), 4.99 (d, $J = 3.7$ Hz, 1H), 4.74 (d, $J = 11.7$ Hz, 1H), 4.59 (ddd, $J = 11.3$, 9.7, 3.7 Hz, 1H), 4.51 (d, $J = 11.7$ Hz, 1H), 4.21 (t, $J = 6.4$ Hz, 1H), 4.17 – 4.04 (m, 2H), 2.16 (s, 3H), 2.06 (s, 4H), 1.99 (s, 3H), 1.93 (s, 3H); $^{13}$C NMR (101 MHz, CDCl$_3$) $\delta$ 170.99, 170.51, 170.40, 170.14, 162.33, 159.82, 131.91, 131.86, 127.79, 127.75, 123.02, 122.87, 119.64, 119.40, 97.34, 68.36, 67.38, 67.16, 63.98, 61.97, 47.75, 23.29, 20.82, 20.77.

![Chemical structure of 3-(Trifluoromethyl)benzyl 2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-α-D-galactopyranoside (9)](attachment)

$^1$H NMR (400 MHz, CDCl$_3$) $\delta$ 7.66 – 7.57 (m, 2H), 7.51 (dd, $J = 4.8$, 1.9 Hz, 2H), 5.68 (d, $J = 9.5$ Hz, 1H), 5.40 (d, $J = 2.3$ Hz, 1H), 5.19 (dd, $J = 11.4$, 3.3 Hz, 1H), 5.00 (d, $J = 3.7$ Hz, 1H), 4.77 (d, $J = 12.2$ Hz, 1H), 4.68 – 4.54 (m, 2H), 4.21 (t, $J = 6.8$ Hz, 1H), 4.10 (qd, $J = 11.2$, 6.8 Hz, 2H), 2.17 (s, 3H), 2.06 (s, 4H), 2.00 (s, 3H), 1.94 (s, 3H); $^{13}$C NMR (101 MHz, CDCl$_3$) $\delta$ 171.07, 170.51, 170.41, 170.14, 137.75, 131.49, 129.34, 125.29, 125.25, 124.85, 124.81, 97.24, 69.27, 68.32, 67.38, 67.18, 61.97, 47.89, 23.24, 20.83, 20.77.

![Chemical structure of 4-(Bromomethyl)benzyl 2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-α-D-galactopyranoside (10)](attachment)
$^1$H NMR (400 MHz, CDCl$_3$) $\delta$ 7.41 (d, $J = 8.1$ Hz, 2H), 7.30 (d, $J = 8.0$ Hz, 2H), 5.59 (d, $J = 9.7$ Hz, 1H), 5.38 (d, $J = 2.4$ Hz, 1H), 5.18 (dd, $J = 11.4$, 3.3 Hz, 1H), 4.98 (d, $J = 3.7$ Hz, 1H), 4.71 (d, $J = 11.9$ Hz, 1H), 4.60 (ddd, $J = 11.4$, 9.8, 3.7 Hz, 1H), 4.53 – 4.44 (m, 3H), 4.20 (t, $J = 6.8$ Hz, 1H), 4.14 – 4.01 (m, 2H), 2.16 (s, 3H), 2.05 (s, 3H), 1.99 (s, 3H), 1.92 (s, 3H); $^{13}$C NMR (151 MHz, CDCl$_3$) $\delta$ 171.04, 170.51, 170.43, 170.08, 170.43, 170.08, 138.13, 136.99, 129.48, 128.69, 97.23, 69.73, 68.46, 67.43, 67.09, 61.97, 47.86, 32.97, 23.40, 20.86, 20.83.

4-(Chloromethyl)benzyl 2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-α-D-galactopyranoside (11)

$^1$H NMR (400 MHz, CDCl$_3$) $\delta$ 7.41 (d, $J = 8.1$ Hz, 2H), 7.32 (d, $J = 8.1$ Hz, 2H), 5.58 (d, $J = 9.7$ Hz, 1H), 5.39 (d, $J = 2.5$ Hz, 1H), 5.18 (dd, $J = 11.4$, 3.3 Hz, 1H), 4.98 (d, $J = 3.7$ Hz, 1H), 4.72 (d, $J = 11.8$ Hz, 1H), 4.66 – 4.55 (m, 3H), 4.51 (d, $J = 11.8$ Hz, 1H), 4.20 (t, $J = 6.8$ Hz, 1H), 4.17 – 4.01 (m, 2H), 2.17 (s, 3H), 2.06 (s, 3H), 1.99 (s, 3H), 1.92 (s, 3H). $^{13}$C NMR (101 MHz, CDCl$_3$) $\delta$ 171.08, 170.54, 170.47, 170.11, 137.84, 136.98, 129.07, 128.67, 97.21, 69.77, 68.49, 67.45, 67.11, 61.99, 47.88, 45.86, 23.42, 20.89, 20.85.

4-Cyanobenzyl 2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-α-D-galactopyranoside (12)

$^1$H NMR (400 MHz, CDCl$_3$) $\delta$ 7.68 (d, $J = 8.2$ Hz, 2H), 7.44 (d, $J = 8.2$ Hz, 2H), 5.67 (d, $J = 9.4$ Hz, 1H), 5.40 (d, $J = 2.5$ Hz, 1H), 5.21 (dd, $J = 11.4$, 3.2 Hz, 1H), 5.03 (d, $J = 3.6$ Hz, 1H),
4.77 (d, J = 12.6 Hz, 1H), 4.70 – 4.53 (m, 2H), 4.18 (t, J = 6.4 Hz, 1H), 4.15 – 4.03 (m, 2H), 2.17 (s, 3H), 2.05 (s, 3H), 2.01 (s, 3H), 1.95 (s, 3H); $^{13}$C NMR (101 MHz, CDCl$_3$) δ 171.14, 170.48, 170.38, 170.13, 142.05, 132.60, 128.37, 118.55, 112.23, 97.52, 69.11, 68.20, 67.27, 67.20, 61.87, 47.97, 23.37, 20.86, 20.83, 20.80.

2-Nitrobenzyl 2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-\(\alpha\)-D-galactopyranoside (13)

$^1$H NMR (400 MHz, CDCl$_3$) δ 7.99 (d, J = 8.2 Hz, 1H), 7.71 – 7.61 (m, 1H), 7.54 (dd, J = 11.8, 4.4 Hz, 2H), 5.67 (d, J = 9.6 Hz, 1H), 5.38 (d, J = 2.6 Hz, 1H), 5.15 (dd, J = 11.4, 3.2 Hz, 1H), 5.02 – 4.94 (m, 2H), 4.89 (d, J = 12.9 Hz, 1H), 4.63 (ddd, J = 11.4, 9.7, 3.6 Hz, 1H), 4.19 (t, J = 6.6 Hz, 1H), 4.12 (d, J = 5.9 Hz, 2H), 2.17 (s, 3H), 2.06 (s, 3H), 2.01 (s, 3H), 2.00 (s, 3H); $^{13}$C NMR (101 MHz, CDCl$_3$) δ 171.04, 170.56, 170.49, 170.43, 148.69, 133.43, 131.77, 130.10, 129.45, 124.91, 97.65, 68.29, 67.33, 67.25, 67.06, 61.99, 47.76, 23.30, 20.84, 20.82, 20.78.

3-Nitrobenzyl 2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-\(\alpha\)-D-galactopyranoside (14)

$^1$H NMR (400 MHz, CDCl$_3$) δ 8.21 (dd, J = 4.1, 2.1 Hz, 2H), 7.70 – 7.53 (m, 2H), 5.72 (d, J = 9.4 Hz, 1H), 5.41 (d, J = 2.5 Hz, 1H), 5.21 (dd, J = 11.4, 3.2 Hz, 1H), 5.03 (d, J = 3.6 Hz, 1H), 4.82 (d, J = 12.5 Hz, 1H), 4.69 – 4.57 (m, 2H), 4.21 (t, J = 6.5 Hz, 1H), 4.11 (dd, J = 8.9, 2.9 Hz, 2H), 2.17 (s, 3H), 2.07 (s, 3H), 2.01 (s, 3H), 1.98 (s, 3H); $^{13}$C NMR (101 MHz, CDCl$_3$) δ 171.14,
4-Nitrobenzyl 2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-α-D-galactopyranoside (15)

$^1$H NMR (400 MHz, CDCl$_3$) $\delta$ 8.23 (d, J = 8.6 Hz, 2H), 7.50 (d, J = 8.6 Hz, 2H), 5.76 (d, J = 9.4 Hz, 1H), 5.41 (d, J = 2.6 Hz, 1H), 5.22 (dd, J = 11.4, 3.2 Hz, 1H), 5.05 (d, J = 3.5 Hz, 1H), 4.82 (d, J = 12.8 Hz, 1H), 4.65 (dd, J = 14.9, 8.3 Hz, 2H), 4.20 (t, J = 6.5 Hz, 1H), 4.16 – 4.05 (m, 3H), 2.18 (s, 3H), 2.05 (s, 4H), 2.02 (s, 3H), 1.97 (s, 3H); $^{13}$C NMR (101 MHz, CDCl$_3$) $\delta$ 171.12, 170.48, 170.37, 170.17, 147.80, 144.05, 128.38, 123.98, 97.58, 68.78, 68.17, 67.25, 67.22, 61.88, 47.97, 23.35, 20.85, 20.82, 20.79.

4-Nitrophenethoxy 2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-α-D-galactopyranoside (16)

$^1$H NMR (400 MHz, CDCl$_3$) $\delta$ 8.20 (d, J = 8.7 Hz, 2H), 7.43 (d, J = 8.7 Hz, 2H), 5.47 (d, J = 9.5 Hz, 1H), 5.30 (d, J = 2.2 Hz, 1H), 5.10 (dd, J = 11.4, 3.2 Hz, 1H), 4.89 (d, J = 3.6 Hz, 1H), 4.53 (ddd, J = 11.4, 9.6, 3.6 Hz, 1H), 4.01 (tdd, J = 11.9, 10.6, 6.7 Hz, 3H), 3.87 (t, J = 6.9 Hz, 1H), 3.74 (dt, J = 9.9, 6.5 Hz, 1H), 3.05 (t, J = 6.5 Hz, 2H), 2.15 (s, 3H), 2.03 (s, 3H), 2.00 (s, 3H),
1.85 (s, 3H); $^{13}$C NMR (101 MHz, CDCl$_3$) δ 171.09, 170.40, 170.32, 169.93, 146.91, 146.45, 129.77, 123.79, 97.65, 68.23, 67.89, 67.19, 66.89, 61.72, 47.90, 35.68, 23.23, 20.81, 20.76, 20.68.

3-Methoxy-4-nitrobenzyl

2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-α-D-galactopyranoside (17)

$^1$H NMR (400 MHz, CDCl$_3$) δ 7.87 (d, J = 8.2 Hz, 1H), 7.07 – 6.95 (m, 2H), 5.66 (d, J = 9.4 Hz, 1H), 5.41 (d, J = 2.3 Hz, 1H), 5.22 (dd, J = 11.4, 3.2 Hz, 1H), 5.04 (d, J = 3.6 Hz, 1H), 4.75 (d, J = 12.6 Hz, 1H), 4.63 (ddd, J = 11.5, 9.5, 3.6 Hz, 1H), 4.56 (d, J = 12.6 Hz, 1H), 4.19 (t, J = 6.8 Hz, 1H), 4.13 – 4.07 (m, 2H), 3.99 (s, 3H), 2.18 (s, 3H), 2.05 (s, 3H), 2.01 (s, 3H), 1.95 (s, 3H); $^{13}$C NMR (101 MHz, CDCl$_3$) δ 171.20, 170.52, 170.39, 170.12, 153.36, 143.77, 126.31, 119.37, 112.76, 97.56, 69.07, 68.22, 67.29, 61.88, 56.73, 48.06, 23.40, 20.88, 20.85.

4-Ethynylbenzyl 2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-α-D-galactopyranoside (18)

$^1$H NMR (400 MHz, CDCl$_3$) δ 7.51 (d, J = 8.2 Hz, 2H), 7.28 (d, J = 8.1 Hz, 2H), 5.63 (d, J = 9.6 Hz, 1H), 5.39 (d, J = 2.5 Hz, 1H), 5.18 (dd, J = 11.4, 3.3 Hz, 1H), 4.98 (d, J = 3.7 Hz, 1H), 4.71 (d, J = 12.0 Hz, 1H), 4.60 (ddd, J = 11.3, 9.7, 3.7 Hz, 1H), 4.51 (d, J = 12.0 Hz, 1H), 4.20 (t, J = 6.7 Hz, 1H), 4.16 – 4.03 (m, 2H), 3.12 (s, 1H), 2.17 (s, 3H), 2.05 (s, 4H), 1.99 (s, 3H), 1.93 (s, 3H); $^{13}$C NMR (101 MHz, CDCl$_3$) δ 171.06, 170.52, 170.44, 170.09, 137.39, 132.51, 128.07, 122.29, 97.27, 83.19, 69.69, 68.41, 67.42, 67.13, 61.99, 47.86, 23.37, 20.85, 20.82.
4-Nitrophenyl 2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-\(\alpha\)-D-galactopyranoside (19)

\(^1\)H NMR (400 MHz, CDCl\(_3\)) \(\delta\) 8.24 (d, \(J = 9.2\) Hz, 2H), 7.19 (d, \(J = 9.2\) Hz, 2H), 5.83 – 5.72 (m, 2H), 5.46 (d, \(J = 2.5\) Hz, 1H), 5.41 (dd, \(J = 11.4, 3.3\) Hz, 1H), 4.80 (ddd, \(J = 11.6, 9.2, 3.5\) Hz, 1H), 4.20 (t, \(J = 6.5\) Hz, 1H), 4.17 – 4.09 (m, 1H), 4.09 – 3.99 (m, 1H), 2.20 (s, 3H), 2.07 (s, 4H), 1.99 (s, 3H), 1.92 (s, 3H); \(^{13}\)C NMR (101 MHz, CDCl\(_3\)) \(\delta\) 171.27, 170.41, 170.33, 170.28, 160.80, 143.22, 126.00, 116.61, 96.41, 68.31, 67.79, 66.99, 61.60, 48.04, 23.37, 20.89, 20.81, 20.66.

4-Methylumbelliferyl 2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-\(\alpha\)-D-galactopyranoside (20)

\(^1\)H NMR (400 MHz, CDCl\(_3\)) \(\delta\) 7.56 (d, \(J = 8.8\) Hz, 1H), 7.09 (d, \(J = 2.4\) Hz, 1H), 7.03 (dd, \(J = 8.8, 2.4\) Hz, 1H), 6.20 (d, \(J = 1.1\) Hz, 1H), 5.92 (d, \(J = 9.3\) Hz, 1H), 5.70 (d, \(J = 3.5\) Hz, 1H), 5.46 (d, \(J = 2.5\) Hz, 1H), 5.41 (dd, \(J = 11.4, 3.3\) Hz, 1H), 4.79 (ddd, \(J = 11.5, 9.3, 3.5\) Hz, 1H), 4.24 (t, \(J = 6.6\) Hz, 1H), 4.17 – 4.09 (m, 1H), 4.05 (dd, \(J = 11.3, 7.1\) Hz, 1H), 2.42 (d, \(J = 1.0\) Hz, 3H), 2.20 (s, 3H), 2.06 (s, 4H), 2.00 (s, 3H), 1.94 (s, 3H); \(^{13}\)C NMR (101 MHz, CDCl\(_3\)) \(\delta\) 171.20, 170.43,
2-Methylnaphthyl 2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-α-D-galactopyranoside
(21)

$^1$H NMR (400 MHz, CDCl$_3$) δ 7.92 – 7.83 (m, 3H), 7.78 (s, 1H), 7.58 – 7.48 (m, 2H), 7.44 (dd, J = 8.4, 1.6 Hz, 1H), 5.70 (d, J = 9.7 Hz, 1H), 5.41 (d, J = 2.4 Hz, 1H), 5.23 (dd, J = 11.4, 3.3 Hz, 1H), 5.05 (d, J = 3.7 Hz, 1H), 4.90 (d, J = 11.8 Hz, 1H), 4.68 (d, J = 11.8 Hz, 1H), 4.62 (ddd, J = 11.3, 9.8, 3.6 Hz, 1H), 2.17 (s, 3H), 2.06 (s, 4H), 2.00 (s, 4H), 1.90 (s, 3H); $^{13}$C NMR (101 MHz, CDCl$_3$) δ 170.92, 170.45, 170.38, 170.01, 133.94, 133.17, 133.14, 128.60, 127.88, 127.78, 127.36, 126.55, 126.43, 125.81, 97.02, 77.42, 77.10, 76.78, 70.25, 68.45, 67.41, 67.02, 61.98, 47.74, 23.24, 20.77, 20.72.

4-Bromophenethoxy 2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-α-D-galactopyranoside
(22)

$^1$H NMR (400 MHz, CDCl$_3$) δ 7.45 (d, J = 8.3 Hz, 2H), 7.12 (d, J = 8.3 Hz, 2H), 5.36 (d, J = 9.7 Hz, 1H), 5.31 (d, J = 2.1 Hz, 1H), 5.09 (dd, J = 11.3, 3.2 Hz, 1H), 4.81 (d, J = 3.6 Hz, 1H), 4.51 (ddd, J = 11.3, 9.8, 3.6 Hz, 1H), 4.09 – 3.97 (m, 2H), 3.97 – 3.85 (m, 2H), 3.64 (dt, J = 9.7,
6.6 Hz, 1H), 2.88 (t, J = 6.1 Hz, 2H), 2.14 (s, 3H), 2.05 (s, 4H), 1.99 (s, 3H), 1.82 (s, 3H); $^{13}$C NMR (101 MHz, CDCl₃) δ 171.00, 170.43, 170.35, 169.97, 137.70, 131.64, 130.66, 120.48, 97.69, 68.42, 67.29, 66.81, 61.78, 47.78, 35.20, 23.22, 20.82, 20.77.

4-(Azidomethyl)benzyl 2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-α-D-galactopyranoside (23)

To a solution of compound 11 (100 mg, mmol) in acetonitrile was added sodium azide. After being stirred at 80 °C for 12 h, the solution was concentrated in vacuo, diluted with ethyl acetate and washed with water, and brine solution. The organic layer was dried over Na₂SO₄, filtered, and concentrated in vacuo to give a residue, which was purified by silica gel column chromatography to generate compound 23. $^1$H NMR (400 MHz, CDCl₃) δ 7.34 (s, 4H), 5.73 (d, J = 9.6 Hz, 1H), 5.39 (d, J = 2.3 Hz, 1H), 5.18 (dd, J = 11.4, 3.3 Hz, 1H), 5.00 (d, J = 3.7 Hz, 1H), 4.73 (d, J = 11.8 Hz, 1H), 4.60 (ddd, J = 11.3, 9.7, 3.7 Hz, 1H), 4.52 (d, J = 11.8 Hz, 1H), 4.36 (s, 2H), 4.21 (t, J = 6.5 Hz, 1H), 4.16 – 4.02 (m, 2H), 2.16 (s, 3H), 2.05 (s, 4H), 1.99 (s, 3H), 1.92 (s, 3H); $^{13}$C NMR (101 MHz, CDCl₃) δ 170.92, 170.44, 170.37, 170.05, 136.80, 135.63, 128.68, 128.55, 97.12, 69.66, 68.38, 67.36, 66.98, 61.91, 54.41, 47.76, 23.27, 20.77, 20.74.

2-Azidobenzyl 2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-α-D-galactopyranoside (24)
$^1$H NMR (400 MHz, CDCl$_3$) $\delta$ 7.42 (td, $J = 7.9$, 1.5 Hz, 1H), 7.32 (dd, $J = 7.5$, 1.3 Hz, 1H), 7.18 (ddd, $J = 11.2$, 8.4, 4.4 Hz, 2H), 5.65 (d, $J = 9.7$ Hz, 1H), 5.38 (d, $J = 2.5$ Hz, 1H), 5.16 (dd, $J = 11.3$, 3.3 Hz, 1H), 4.98 (d, $J = 3.7$ Hz, 1H), 4.69 (d, $J = 11.4$ Hz, 1H), 4.59 (ddd, $J = 11.2$, 9.8, 3.7 Hz, 1H), 4.45 (d, $J = 11.5$ Hz, 1H), 4.29 (t, $J = 6.7$ Hz, 1H), 4.11 (qd, $J = 11.2$, 6.6 Hz, 2H), 2.17 (s, 3H), 2.07 (s, 3H), 1.99 (s, 3H), 1.93 (s, 3H); $^{13}$C NMR (101 MHz, CDCl$_3$) $\delta$ 171.06, 170.56, 170.50, 170.06, 138.95, 130.95, 130.19, 127.79, 125.07, 118.55, 97.39, 68.67, 67.49, 67.10, 66.12, 62.02, 47.83, 23.42, 20.90, 20.84.

3-Azidobenzyl 2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-\(\alpha\)-D-galactopyranoside (25)

$^1$H NMR (400 MHz, CDCl$_3$) $\delta$ 7.37 (t, $J = 7.8$ Hz, 1H), 7.09 (d, $J = 7.7$ Hz, 1H), 7.03 (dd, $J = 8.0$, 1.6 Hz, 1H), 6.97 (d, $J = 1.7$ Hz, 1H), 5.60 (d, $J = 9.6$ Hz, 1H), 5.39 (d, $J = 2.4$ Hz, 1H), 5.19 (dd, $J = 11.4$, 3.3 Hz, 1H), 4.98 (d, $J = 3.7$ Hz, 1H), 4.70 (d, $J = 12.0$ Hz, 1H), 4.60 (ddd, $J = 11.4$, 9.7, 3.7 Hz, 1H), 4.50 (d, $J = 12.0$ Hz, 1H), 4.20 (t, $J = 6.8$ Hz, 1H), 4.16 – 4.06 (m, 2H), 2.17 (s, 3H), 2.07 (s, 3H), 2.00 (s, 3H), 1.95 (s, 3H); $^{13}$C NMR (101 MHz, CDCl$_3$) $\delta$ 171.10, 170.54, 170.45, 170.12, 140.67, 138.73, 130.25, 124.65, 118.95, 118.74, 97.11, 69.45, 68.41, 67.43, 67.15, 62.01, 47.90, 23.42, 20.89, 20.85.

4-Azidobenzyl 2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-\(\alpha\)-D-galactopyranoside (26)
\[ ^1H\text{NMR (400 MHz, CDCl}_3\text{) }\delta 7.31 (d, J = 8.5 Hz, 2H), 7.04 (d, J = 8.5 Hz, 2H), 5.58 (d, J = 9.6 Hz, 1H), 5.38 (d, J = 2.3 Hz, 1H), 5.17 (dd, J = 11.4, 3.3 Hz, 1H), 4.97 (d, J = 3.7 Hz, 1H), 4.68 (d, J = 11.7 Hz, 1H), 4.59 (ddd, J = 11.3, 9.7, 3.7 Hz, 1H), 4.48 (d, J = 11.7 Hz, 1H), 4.20 (t, J = 6.9 Hz, 1H), 4.16 - 4.03 (m, 2H), 2.17 (s, 3H), 2.06 (s, 4H), 1.99 (s, 3H), 1.92 (s, 3H); \]^1C NMR (101 MHz, CDCl\text{3})\delta 171.09, 170.53, 170.45, 170.06, 140.37, 133.33, 129.97, 119.38, 97.03, 69.54, 68.45, 67.43, 67.07, 61.99, 47.88, 23.41, 20.87, 20.84.

4-Azidophenethoxy 2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-\alpha-D-galactopyranoside (27)

\[ ^1H\text{NMR (400 MHz, CDCl}_3\text{) }\delta 7.22 (d, J = 8.4 Hz, 2H), 6.99 (d, J = 8.4 Hz, 2H), 5.40 – 5.28 (m, 2H), 5.10 (dd, J = 11.3, 3.3 Hz, 1H), 4.82 (d, J = 3.6 Hz, 1H), 4.52 (ddd, J = 11.2, 9.8, 3.6 Hz, 1H), 4.11 – 3.97 (m, 2H), 3.97 – 3.84 (m, 2H), 3.65 (dt, J = 9.7, 6.6 Hz, 1H), 2.90 (t, J = 6.4 Hz, 2H), 2.15 (s, 3H), 2.04 (s, 3H), 1.99 (s, 3H), 1.84 (s, 3H); \]^1C NMR (101 MHz, CDCl\text{3})\delta 171.15, 135.45, 130.34, 119.27, 97.76, 68.73, 68.56, 67.42, 66.92, 61.89, 47.92, 35.27, 23.36, 20.92, 20.88, 20.81.

2.4.2 Assays

A549 Cells were incubated with or without 50 uM Bn-\alpha-GalNAc or compound\& Bn-\alpha-GalNAc mixture at the ration of 1:1 (50 \uM: 50 \uM) for 3 days, then collected the media, purified them by C18 SPE column (waters), and analyzed them by MALDI. Relative activity as CORA precursor was calculated by formulation as follow: the intensity of the peaks of all
derivatized-O-Glycans/ the in-tensity of the peaks of all Bn-O-Glycans/(the intensity or area of the peaks of compound/ the intensity or area of the peaks of Bn-a-GalNAc). Every experiment was repeated for 3 times.

0.1 mM compounds and 0.5 mM UDP-Galactose were catalyzed by 0.4 ng/uL T-synthase (provided by Dr. Cummings’ lab) in 20 mM MES buffer pH 6.8 (containing 20 mM MnCl, 0.5% Triton X-100) at 37 °C for 18 h, then the reactions were quenched by equal volume of methanol, and stored in -80 °C for overnight. The frozen samples were un-froze and centrifuged at 14,000 rpm for 5 min, then the supernatants were loaded on-to C18 (XBridge TM, 35 μm, 30*150 mm) column. SHIMADZU HPLC system was used, and mobile phase A consisted of 0.1% TFA in Acetonitrile and mobile phase B was 0.1% TFA in water. A linear gradient from 5 % to 60% over 30 minutes at a flow rate of 0.4 mL/min was applied. Software EZStart 7.4 was used for data analysis. The disaccharide product peak HPLC condition: The Yields was calculated by formulation as follow: The acceptor activity = Area of disaccharide peak / Area of (disaccharide+ the left com-pound)*100. The acceptor activity of Bn-a-GalNAc was set as 100. The relative acceptor activity = (The acceptor activity of com-pound/ The acceptor activity of Bn-a-GalNAc )*100.
3 STRUCTURE BASED DESIGN, SYNTHESIS AND ACTIVITY STUDIES OF SMALL HYBRID MOLECULES AS HDAC AND G9A DUAL INHIBITORS

Aberrant enzymatic activities or expression profiles of epigenetic regulations are therapeutic targets for cancers. Among these, histone 3 lysine 9 methylation (H3K9Me2) and global deacetylation on histone proteins are associated with multiple cancer phenotypes including leukemia, prostatic carcinoma, hepatocellular carcinoma and pulmonary carcinoma. In this study, we report the discovery of the first small molecule capable of acting as a dual inhibitor targeting both G9a and HDAC. Our structure based design, synthesis, and screening for the dual activity of the small molecules led to the discovery of compound 14 which displays promising inhibition of both G9a and HDAC in low micro-molar range in cell based assays.

3.1 Introduction

Epigenetic modifications describe post-translational modifications that occur on the protein without lasting impact on the base genomic code. Likewise, epigenetic modifications are reversible due to the manner in which they occur, making restoration of the epigenome to its normal function a crucial target in many disease models.\[^{83}\] In particular, dimethylation of histone 3 at lysine 9 (H3K9Me2) and various acetylation marks on histones are directly correlated to the onset and advancement of multiple cancer phenotypes, including leukemia, prostatic carcinoma, hepatocellular carcinoma and pulmonary carcinoma.\[^{84}\] Recently, there has been much success in the development of small molecules targeting these post-translational modifications. Epigenetics is still a field much in its infancy considering the number of epigenetic targets.\[^{83a,84a,85}\] From this vast pool, epigenetic markers that related to leukemogenesis and tumorigenesis has shown to be a promising application of epigenetics.
Cancer is a disease with complicated treatment options due to the multifactorial basis of initiation and progression. Therefore a treatment targeting multiple components instead of a single component could be a particular interest in cancer therapeutics.\[86\] To meet this need, developing new lead molecules which target cancer from various stages of disease development from either known inhibitors or \textit{de novo} is essential for improving effectiveness and side effects/toxicity profile.\[87\] Herein we report the design, synthesis and extensive biological evaluation of a class of small molecules targeting the enzymes histone deacetylases (HDACs) and histone methyltransferases (G9a), both are key posttranslational enzymes in cancer development.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{known_HDAC_inhibitors_and_G9a_Inhibitors.png}
\caption{Examples for known HDAC inhibitors and G9a Inhibitors.}
\end{figure}

Histone deacetylases (HDACs) fall into the category of eraser enzymes, so termed due to their ability to reverse the acetylation modification employed by another enzyme histone acetyl transferases (HATs).\[88\] However, despite the name, histone deacetylases have a wide range of
substrates included but not limited to strictly histones.\cite{84c} Aberrant activity of HDACs have been well documented in several cancer phenotypes, with HDAC inhibitors (HDACIs, Fig 1) proven as antineoplastic agents. HDACIs have multiple cell type-specific effects in vitro and in vivo, such as growth arrest, cell differentiation and apoptosis in malignant cells.\cite{84a, 89} HDACIs have been shown to induce apoptosis in both solid and hematopoietic malignancies using both transcription dependent and transcription independent mechanisms.\cite{90} While histone acetylation regulated by HDAC and their corresponding writer enzymes HATs, protein methylation is similarly regulated by protein lysine methyltransferases enzymes (PKMT), protein arginine methyltransferases (PRMT) and their corresponding demethylase enzymes.\cite{83a, 91} Our particular interest is the PKMT G9a (also known as KMT1C, EHMT2), which is a histone 3 lysine 9 (H3K9) specific methyltransferase that is overexpressed in many cancers including leukemia, hepatocellular carcinoma and pulmonary carcinoma. G9a is notable for its roles in cancer cell proliferation, whereas knockdown of G9a in prostate, lung, and leukemia cancer cells resulted in the inhibition of cell growth.\cite{83e, 92} Presently, there are number of small molecules with different structural cores that have been found to inhibit G9a, which are also under consideration in clinical trials.\cite{83e, 93} In addition to catalyzing mono- and dimethylation of H3K9, G9a and its closely related protein GLP, also dimethylate lysine 373 of the tumor suppressor p53 to repress the transcriptional activity of p53.\cite{94}

Different modifications of chromatin associated with variable functions, with the extent of modifications and specific residue selected imparting different overall result. For example, hyper-acetylation of histone H3 and H4 both known to be activators associated with ongoing transcription.\cite{95} However, methylation of H3K9 and H3K27 are associated with gene silencing and repressive marks.\cite{96} Following these pathways of distinct and uncorrelated functionalities, we
proposed that concurrent inhibition of G9a and HDAC may help histone protein to retain the transcriptional activity via targeting two pathways. First affecting the reduced formation of corepressor marks (H3K9Me2 and p53K373Me2) as a downstream effect of G9a inhibition, and secondly with a relaxed and acetylated chromatin promoted by the HDAC inhibition. We also hypothesize a potential competition between the HDAC and G9a enzyme because they share common substrate (H3K9 and p53K373), a substrate in its acetylated state cannot be further methylated, thereby causing an indirect inhibition of G9a. It should also be noted that HDACIs show only a moderate and limited biological response when applied as a mono treatment, with therapeutic effect significantly improved in combination with other anticancer agents.[87a, 87c, 97] Hence, the optimal deployment of these molecules may be a combination with other epigenetic drugs, acting against the set of enzymes responsible for the setup and maintenance of epigenetic information.[98]

3.2 Result and discussion

Initial tests designed around assessing whether a synergistic effect would be observed in real when a combination of G9a inhibitor and HDAC inhibitor used in conjunction. Towards this, the MDA-MB-231 and MCF-7 cell lines treated with either SAHA (1-100 µM), BIX-01294 (1-100 µM), or a mixture of SAHA and BIX-01294 (1:1; 1-100 µM). As indicated in Table 1 (EC$_{50}$ graphs available in Supplemental Information, Figure S1), when applied in combination performance was enhanced towards MDA-MB-231 and comparable in MCF-7. Despite being two distinct molecules with different physiochemical properties, application of both displayed a significant improvement (approximately 34% lower EC$_{50}$ to SAHA, and 13% lower EC$_{50}$ to BIX-
01294 in MDA-MB 231). Effectively, this provided the basis for incorporating both SAHA and BIX-01294 into a single moiety capable of preserving inhibitory activity against both targets.

Table 3.1 Combination study of BIX-01294 and SAHA

<table>
<thead>
<tr>
<th>Inhibitor (10 µM)</th>
<th>MDA-MB-231</th>
<th>MCF-7</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAHA</td>
<td>2.874±0.84</td>
<td>8.124±4.98</td>
</tr>
<tr>
<td>BIX-01294</td>
<td>2.155±0.88</td>
<td>8.103±1.99</td>
</tr>
<tr>
<td>SAHA + BIX-01294</td>
<td>1.891±0.56</td>
<td>8.564±2.17</td>
</tr>
</tbody>
</table>

Multi-target treatments are typically utilized two approaches, formulating two active moieties as a cocktail, or hybridizing properly selected active moieties into a single molecule. The first method heavily relies on both compounds having comparable solubility, or else fine-tuning the formulation to ensure the desired bioavailability.\[86a, 86d, 87d, 99\] In contrast, a dual-target drug doesn't have such issue with different solubility or bioavailability. However, hybrid compounds are challenging to design due to the difficulties in optimizing a pharmacophore from two dissimilar compounds that can retain multiple functionalities inside the body. There have been some notable successes in the latter regard, with few drugs already on the market with superior performance to their cocktail counterparts.\[87c\] Currently available hybrid drugs target different stages of disease development despite their precursors targeting the same diseases.\[87a, 87d, 99b, 100\] In light of these studies, we propose that a hybrid drug that instead targets components belonging to the same scheme in disease progression or has otherwise interdependent functionality would yield an improved synergistic effect. Herein we report the design, synthesis, screening and biological study of a lead molecule for further development as a drug candidate utilizing this principle.
3.2.1 Compound design

From the available data on G9a and HDAC in the forms of known inhibitors and their respective X-ray crystal structures, we observed several details regarding their prospective ligands. For instance, all HDAC inhibitors are comprised of three parts- a lipophilic cap connected to a hydrophilic Zn$^{2+}$ binding group by an alkyl, or arylene linker, with the cap portion being different bulky groups. Likewise, current G9a inhibitors, except for the fungal metabolite chaetocin, are primarily derived from the parent compound BIX-01294. Since the discovery of BIX-01294, a clear majority of the G9a inhibitors to emerge based on the quinazoline core of BIX-01294. As the lipophilic quinazoline core resemble the lipophilic bulky cap for HDAC inhibitors, we reasoned that the G9a core could feasibly function as the core scaffold of a HDAC and G9a dual inhibitor. There have been notable success stories in regard of varying the cap or linker portion of the HDAC inhibitor, while the metal binding portion kept as a hydroxamic acid or an ortho-amino benzamide.\textsuperscript{[86a, 87c, 87e, 101]} Following this line of thinking, we added the linker and the hydroxamic acid at the C2, and C4 of quinazoline ring to obtain the desired hybrid molecules (Fig 2A). This design was inspired by the fact G9a have many inhibitors with bulky side chains as in the case of E72, HDACIs can also afford a reasonable variety of lipophilic cores (Fig 1). Various analogs with different linker lengths and diverse groups at C6 and C4 cyclohexylamine positions were also designed. All these compounds classified into three for ease of reference.
3.2.2 Synthesis

The designed analogs were synthesized from two building blocks; commercially available 2,4-dichloro-6,7-dimethoxyquinazoline (1) was used for the dimethoxy analogs (Fig 3) and, 2-amino-4-methoxybenzoic acid (8) used for monomethoxy analogs (Fig 4). Initially, we synthesized only a few analogs of class III to assess the effectiveness of the HDAC substitution while opening the piperazine ring originally present at the prototype BIX-01294, this class was particularly intended to explore the SAR of R1. The bulky seven-member ring was then replaced with an ethylene diamine and coupled with various esters of different chain lengths (three to seven carbons for investigating the optimal length for the HDAC inhibition) to produce compounds 4, 5, 6 and 7. To check the effect of bulky groups at the C4 position of the heterocyclic ring, an isopropyl group then introduced to the tertiary amine instead of the methyl group to produce the set of compounds 4a, 5a, 6a and 7a (Series IIIA, Fig 3). While investigating the binding characteristics of known G9a/GLP inhibitors, the C6 methoxy group of quinazoline ring was hypothesized not to contribute significantly to ligand-receptor interactions. Therefore, the methoxy at C6 was
eliminated to find a balance for HDAC activity. Compounds in class II designed from this rationale. Compounds 13-17 have a 4-aminobenzyl piperidine at C4, while 13a-17a possess an isopropylpiperidin-4-amine. Recent studies of UNC0965 wherein a biotin tag was applied to UNC0638 found that UNC0965 retained biological activity,\textsuperscript{[102]} following this lead, C4 of quinazoline ring was explored. Compounds with the HDAC pharmacophore on the C4 carbon of quinazoline core termed as class I, with analogs 19 and 20 retaining the C6 methoxy group, and 21 and 22 lacking the methoxy group.

**Figure 3.3** Synthesis cheme of dimethoxy analogs. 
Reagents and conditions: (i) 1-methylpiperidin-4-amine/1-isopropylpiperidin-4-amine, DIPEA, DMF, rt, 3 h, 80-86%, (ii) tert-butyl (2-aminoethyl)carbamate, DIPEA, 160 °C Microwave, 10 min, 60-66%, (iii) TFA/DCM, 3 h, (iv) MonomethylSuberate, EDCI, HOBT, 8 h, (v) 50% NH\textsubscript{2}OH in water, MeOH, 60°C, 8 h, 30-38% over two steps.

**Fig 3.3** compounds were synthesized from the commercially available starting material 1. An initial displacement reaction using a primary amine was used to introduce the C4 selective
substitution, with the second displacement to introduce the linker at the C2 position following the microwave assisted reaction previously reported.\textsuperscript{103} Boc-protected ethylene diamine was treated with compound 2 at 160 °C in a microwave for 10 min to yield product 3 with excellent yield. Afterward, the amine 3 was deprotected with TFA/DCM, and the free amine was treated with corresponding monomethyl esters (carbon chain 2-6) in the presence of coupling reagent EDCI and HOBT for about 8 hours to produce mono methyl ester substituted at the C2 position. The ester compounds further treated with hydroxylamine in water to get the corresponding hydroxamic acid derivatives, which were purified using reverse phase flash chromatography to obtain compounds 4-7 and 4a-7a in good yield. Synthesis of compounds in Fig 4 requires the C6 demethoxy core, and this was synthesized according to the reported procedure to yield 10.\textsuperscript{104} Appropriate displacement and coupling reactions in this core as demonstrated in Fig 4 afforded compounds 13-17 and 13a-17a. Compounds with the HDAC pharmacophore at the C4 position (Fig 2B) were synthesized from the starting material 1. As Figure 5 showed, Pd/C hydrogenolysis was used to eliminate the benzyl group and produce the free amine of 18 for the coupling of monomethyl esters to result in compounds 19-22. A parallel synthesis and testing strategy were used in establishing the primary SAR, with the clear rationalization of the structure and activity at each stage we could reduce the synthetic targets. Initially, we evaluated the G9a potential after the introduction of HDAC pharmacophore. A biochemical assay using MALDI-TOF was used to visualize the effects of the synthesized compounds on G9a enzymatic activity, we carried out a biochemical reaction involving target enzyme G9a, methyl donor SAM and substrate H3 peptide at a concentration of 400 nM, 10 μM and 5 μM respectively.\textsuperscript{105} After successfully optimizing the reaction conditions and reaction time to see at least 80% of the substrate converted to the methylated form (H3K9Me1
or H3K9Me2) with no tri-methylation, we tested BIX-01294 for an optimum level of inhibition and fixed the concentration as 5 µM for each inhibitor.

Figure 3.4 Synthesis scheme for monomethoxy analogs. 
Reagents and conditions: (i) Urea, 200°C, 2 h, (ii) POCl3, reflux 16 h, 40% in two steps, (iii) 4-aminobenzylpiperidin/1-isopropylpiperidin-4-amine, DIPEA, DMF, rt, 3 h, 74% and 86%, (iv) tert-butyl (2-aminoethyl)carbamate, DIPEA, 160 °C Microwave, 10 min, 64% and 68%, (v) TFA/DCM, 3 h, (vi) MonomethylSuberate, EDCI, HOBt, 8 h, 70% in two steps, (vii) 50% NH₂OH in water, MeOH, 60 °C, 8 h, 30-40%.

Most of the compounds retained G9a inhibition capabilities, indicated by the reduction in the ratio of the H3K9Me1 and H3K9Me2 peaks compared to the control reaction (Fig 3.6). While fractionally less potent than parent BIX-01294, retention of inhibitory capabilities was nonetheless verification of the initial hypothesis. These result corresponded to the MALDI-TOF study done by a previously reported procedure. With the knowledge that G9a activity preserved in the biochemical assay, next, we investigated the effect in the cell. H3K9Me2 cell immunofluorescence In-Cell Western (ICW) assay was used to assess G9a inhibition potential, and homogeneous cellular histone deacetylase assay used for measuring HDAC inhibition.
Figure 3.5 Synthesis scheme of compound 19-22.
Reagents and condition: (i) 4-aminobenzylpiperidin, DIPEA, DMF, rt, 3 h, 90%, (ii) 1-methyl-1,4-diazepane, DIPEA, 160 °C Microwave, 10 min, 74%, (iii) EtOH, Pd/C, H2, 8 h, (iv) Monomethylsuberate/monomethylpimelate, EDCl, HOBt, 8 h, (v) 50% NH2OH in Water, MeOH, 60 °C, 8 h, 44% and 45%, (vi) NHBoc-ethylendiamine, DIPEA, DMF, rt, 3 h, 78%, (vii) 1-methyl-1,4-diazepane, DIPEA, 160 °C Microwave, 10 min, 69%, (viii) TFA/DCM, 8 h, (ix) Monomethylsuberate/ monomethylpimelate, EDCl, HOBt, 8 h, (x) 50% NH2OH in water, MeOH, 60 °C, 8 h, 29%-36%.

3.2.3 Functional potency evaluation for G9a inhibition

For assessing the functional potency of the dual inhibitors, we evaluated all the compounds by H3K9Me2 cell immunofluorescence In-Cell Western (ICW) assays and the results shown in Table 2. The MDA-MB-231 cell line was used in this study as this cell line possesses robust H3K9Me2 levels.[5] Our results suggested that compounds belonging to the class IV (southwest directing HDAC) exert a G9a activity comparable to the parent compound BIX-01294, but all other classes were significantly less potent.

Table 3.2 H3K9Me2 cell immunofluorescence In-Cell Western (ICW) assay results.

<table>
<thead>
<tr>
<th>Compound</th>
<th>G9a IC$_{50}$ (µM)</th>
<th>Compound</th>
<th>G9a IC$_{50}$ (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>96.69±1.68</td>
<td>4a</td>
<td>74.21±1.94</td>
</tr>
<tr>
<td>5</td>
<td>&gt;100</td>
<td>5a</td>
<td>66.63±3.98</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>&gt;100</td>
<td>6a</td>
</tr>
<tr>
<td>-----</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>7</td>
<td>76.74±0.89</td>
<td>7a</td>
<td>54.55±3.05</td>
</tr>
<tr>
<td>13</td>
<td>37.79±2.80</td>
<td>13a</td>
<td>&gt;100</td>
</tr>
<tr>
<td>14</td>
<td>7.136±1.62</td>
<td>14a</td>
<td>46.83±1.97</td>
</tr>
<tr>
<td>15</td>
<td>72.10±1.37</td>
<td>15a</td>
<td>46.97±3.33</td>
</tr>
<tr>
<td>16</td>
<td>90.26±3.75</td>
<td>16a</td>
<td>45.71±1.76</td>
</tr>
<tr>
<td>19</td>
<td>99.63±3.13</td>
<td>21</td>
<td>ND</td>
</tr>
<tr>
<td>20</td>
<td>97.51±2.78</td>
<td>22</td>
<td>60.65±3.66</td>
</tr>
<tr>
<td>BIX-01294</td>
<td>4.563±1.2</td>
<td>5b</td>
<td>87.39±5.44</td>
</tr>
</tbody>
</table>

3.2.4 Functional potency evaluation for HDAC inhibition

The enzymatic activity of HDAC was measured in intact cells using the homogeneous cellular assay method. A cell-permeable peptide Boc-K(Ac)-AMC used as the HDAC substrate, after deacetylation it is cleaved by trypsin to release the fluorescent 7-amino-4-methylcoumarin (AMC) and further quantified. Each compound candidate tested in both Hela and K562 cell lines, two compounds (13 and 14) showed significant HDAC inhibition (Table 3, see Supplemental Information Table 2 for full results). An evaluation of these structures indicated that only class III compounds displayed the desired activity, it was inconclusive if the R1 and R2 substitutions were responsible. Compound 5b, synthesized from 3b with a benzyl group at R1 and a methoxy at R2, did not show decent inhibition of HDAC compared with Compound 14. Which indicated two important strategies for designing HDAC inhibitors with quinazoline core; an aromatic ring at R2 is essential for HDAC activity while the methoxy group at C6 of quinazoline core almost destroys HDAC inhibition. Linker lengths vary from Compound 11 to 14, in which, compounds with 5 or 6 methylene groups between the lipophilic core and Zn binding domain found to be the best choice. Similarly, compounds 19-22 with the R1 substituted with the HDAC chain did not show promising activity, leaving 13 and 14 as the candidates for further study. Closer examination of the tested
compounds indicated that HDAC activity is limited to compounds with a benzyl group at the 4-amino piperidin ring (R1) along with no substitution at C6.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Hela IC₅₀ (µM)</th>
<th>K562 IC₅₀ (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>15.33±0.79</td>
<td>27.75±0.59</td>
</tr>
<tr>
<td>14</td>
<td>13.80±1.22</td>
<td>5.735±1.23</td>
</tr>
<tr>
<td>SAHA</td>
<td>5.044±0.53</td>
<td>2.056±0.59</td>
</tr>
</tbody>
</table>

3.2.5 Molecular Docking Analysis

Molecular Docking analysis and molecular dynamics simulations were widely used to estimate the interactions between protein and small molecular compounds theoretically. Our assays found that 14 had good cellular potency for inhibition of both G9a and HDAC, so docking studies were used to examine the interactions of 14 with the target proteins compared to known ligands using Schrödinger Suite 2014-3. The crystal structure of human HDAC8 with MS-344 (PDB ID: 1T67) and human G9a with BIX-01294 (PDB ID: 3FPD) were employed as the templates for molecular docking studies. SP Glide algorithm was first validated by docking MS-344 and BIX-01294 from the complex back to the protein, ligand preparation was done using LigPrep with OPLS_2005. The search space was defined using Receptor Grid Generation in Glide, with the centroid of the ligand chosen to define the grid box. Standard precision mode was selected for validation docking, and default settings for scaling van der Waals radii were used. No constraints were defined for the docking runs. The docking pose with the highest score returned for MS-344 and BIX-01294 were compared with the starting protein complex. For subsequent molecular docking of compound 14 in the binding site of HDAC8 and G9a, LigPrep was used for energy minimizations of the molecule with the OPLS_2005 force field. Using the initial grids
generated for HDAC8 and G9a, the standard precision docking was repeated for compound 14 as described above.

**Table 3.4 GLIDE docking results for MS-344 and compound 14 at the catalytic site of HDAC8 (PDB ID: 1T67)**

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Ligand ID</th>
<th>Docking Score</th>
<th>GLIDE score</th>
<th>Interactions</th>
<th>Hydrogen Bonds</th>
<th>Interaction with Zn²⁺</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MS-344</td>
<td>-7.931</td>
<td>-7.931</td>
<td></td>
<td>His142, His143, Gly151, Gly304; Asp101, Tyr306;</td>
<td>Ionic interaction</td>
</tr>
<tr>
<td>2</td>
<td>14</td>
<td>-7.934</td>
<td>-8.369</td>
<td></td>
<td>His142, Gly151, Gly304; Asp101, Tyr306;</td>
<td>Ionic interaction</td>
</tr>
</tbody>
</table>

Table 4 and 5 showed the results of docking along with the prominent interactions for compound 14 with HDAC8 and G9a. Fig 7 illustrated the predicted binding modes and the detailed protein–inhibitor interactions of HDAC8 and G9a with compound 14, respectively. Fig 7A showed that the catalytic tunnel of HDAC8 was occupied by the aliphatic side chain of the inhibitor, while the hydroxamate group chelated the zinc ion and hydrogen bonds with residues in the catalytic site.

**Table 3.5 GLIDE docking results for BIX-01294 and compound 14 at the catalytic site of G9a (PDB ID: 3FPD)**

<table>
<thead>
<tr>
<th>S.No.</th>
<th>Ligand ID</th>
<th>Docking Score</th>
<th>GLIDE score</th>
<th>Interactions</th>
<th>Hydrogen Bonds</th>
<th>Interaction with Zn²⁺</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>BIX-01294</td>
<td>-7.664</td>
<td>-8.134</td>
<td>Ala1134</td>
<td>Asp1131, Asp1135, Asp1140;</td>
<td>NA</td>
</tr>
<tr>
<td>2</td>
<td>14</td>
<td>-7.321</td>
<td>-7.52</td>
<td>Arg1137, Glu1138</td>
<td>Asp1131, Asp1135, Asp1140, Arg1214;</td>
<td>NA</td>
</tr>
</tbody>
</table>
tunnel. The zinc ion displayed a trigonal bipyramidal geometry and two points contacted with the ligand (Fig 7A and 7C). Docking studies suggested important structural/catalytic roles for Gly140, His142, Gly151 and Gly304 in the active site and extending to Tyr306, Asp101. Moreover, compound 14 occupies the binding pocket in a similar posture as MS-344 in the catalytic site of HDAC8 (Supplemental Information Fig S3).

A similar study was performed to establish the binding characteristics of compound 14 with G9a. The binding model of compound 14 showed that it shares common hydrogen bonding interactions with key residues of the catalytic domain in a mode comparable to BIX-01294 (Supplemental Information Fig S4). Most notably, the piperidine ring substituted at quinazolin-4-amine in compound 14 has hydrogen bonding interactions with Arg1137, Glu1138 residues, and the aliphatic chain was involved in some more hydrogen bonding interaction with the side chains of residues Asp1131, Asp1135, Asp1140 and Arg1214 (Fig 7B and 7D).

### 3.2.6 Cell Anti-proliferation Assay

Cell anti-proliferation assays were performed to determine the toxicity of these inhibitors. Several cell lines (MDA-MB-231, MCF-7 and A549) were incubated and then treated with varying concentrations of the inhibitors for 72 h, respectively. After the first cell culture screening, it was determined that the inhibitors were more effective with breast cell lines (MDA-MB 231 and MCF-7) compared to other cell lines, particularly compound 13 and 14 (Table 6, and Supplemental Information Table 4). These compounds were further evaluated against the control cell line HEK293 to test their toxicity with a non-cancerous cell line. As seen in Table 6, both SAHA and BIX-01294 appear to be toxic to cancer and normal cells, but compounds 13 and 14 displayed lower toxicity, particularly compound 14. 14 also showed improved anti-proliferation abilities in all cancer cell lines and reduced toxicity in normal cell line compared to 13.
Figure 3.6 Molecular docking study results.
A) Predicted binding mode of compound 14 on HDAC8 (PDB ID: 1T67), B) Predicted binding mode of compound 14 on G9a (PDB ID: 3FPD), C) Binding model of active compound 14 (orange) as revealed from GLIDE docking in the HDAC8 catalytic site (PDB ID: 1T67). The green dashed lines represent hydrogen bonds. The turquoise sphere represents Zn cation and with a trigonal bipyramidal coordination geometry. The pink dashed lines indicates the two contacts between the ligand and Zn cation, with the mixed dashed lines representing interaction between Zn cation and amino acid residues. H-bond distances (Å) between heteroatoms of ligand and amino acid residues are as follows: Asp101 (1.90), His142 (2.02), His143 (3.64), Gly151 (3.68), Gly304 (3.00), Tyr306 (2.17), D) Binding model of active compound 14 (orange) as revealed from GLIDE docking in the G9a catalytic site (PDB ID: 3FPD). The green dashed lines represent hydrogen bonds. H-bond distances (Å) between heteroatoms of ligand and amino acid residues
are as follows: Asp1131 (1.66), Asp1135 (1.75, 1.81), Arg1137 (3.33), Glu1138 (3.98), Asp1140 (1.77), Arg1214 (2.68, 2.90)

**Table 3.6 Inhibition of compounds 13 and 14 on the growth of cancer cells and normal cells.**

<table>
<thead>
<tr>
<th>compound</th>
<th>EC(_{50}) (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MDA-MB 231(^b)</td>
</tr>
<tr>
<td>13</td>
<td>89.33±1.23</td>
</tr>
<tr>
<td>14</td>
<td>10.02±1.66</td>
</tr>
<tr>
<td>SAHA</td>
<td>2.87±0.84</td>
</tr>
<tr>
<td>BIX-01294</td>
<td>2.15±0.88</td>
</tr>
</tbody>
</table>

\(^a\) >100 in the cases where the IC\(_{50}\) did not reach at the highest tested concentration (100 µM). \(^b\) MDA-MB-231: breast cancer cell line; \(^c\) MCF-7: breast cancer cell line; \(^d\) A549: human lung cancer cell line; \(^e\) HEK293 normal cell line; SAHA and BIX-01294 were used as the positive controls; Cells were exposed to the different inhibitors with various concentrations for 72h. Inhibition of cell growth by the listed compounds was determined by using CCK-8 kit. Data shown as mean ± SD of triplicates.

### 3.2.7 ADME Prediction Studies

The same procedures and principals from the earlier in silico physico-chemical evaluations of known HDACIs were applied here to evaluate these novel dual inhibitors.\(^{[109]}\) ADMET module of Discovery Studio 3.1 was used to predict physical properties. Using Lipinski’s rule of five,\(^{[110]}\) the octanol–water partition coefficient (AlogP98) should be less than 5. As seen in Table 7, the candidate compound 14 is well within accordance of the rule. In addition, other values also fell into the acceptable ranges of PSA-2D (7–200) and QplogS (−6.5 to 0.5), indicating 14 may possess good bioavailability. These parameters were also taken into consideration in identifying better inhibitors, suggesting that 14 has the characteristics desirable for a drug candidate.

**Table 3.7 ADME prediction results**

<table>
<thead>
<tr>
<th>Entry</th>
<th>M.W</th>
<th>QPlogS(^c)</th>
<th>PSA</th>
<th>PSA-2D(^b)</th>
<th>AlogP98(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>515.654</td>
<td>-3.702</td>
<td>161.25</td>
<td>141.462</td>
<td>2.511</td>
</tr>
<tr>
<td></td>
<td>264.324</td>
<td>-2.139</td>
<td>102.256</td>
<td>81.037</td>
<td>1.838</td>
</tr>
<tr>
<td>--------</td>
<td>---------</td>
<td>--------</td>
<td>---------</td>
<td>--------</td>
<td>-------</td>
</tr>
<tr>
<td>SAHA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIX-01294</td>
<td>476.62</td>
<td>-6.792</td>
<td>50.675</td>
<td>63.249</td>
<td>4.189</td>
</tr>
</tbody>
</table>

\(^a\) AlogP98 means atom-based LogP (octanol/water). \(^b\) PSA-2D means 2D fast polar surface area. \(^c\) QplogS means predicted aqueous solubility.

### 3.3 Conclusions

Considering the inherent deficiency of HDACIs as a monotherapy, in conjunction with the past success of incorporating the HDAC pharmacophore into many dual inhibitors, we hypothesized that the core metal ion binding hydrophilic segment could be coupled with the lipophilic core of G9a inhibitors to increase the effectiveness. Both G9a and HDACs are therapeutic targets for cancer therapy and are both capable of targeting identical substrates (H3K9 and lysine 373 of p53). In search of a lead molecule featuring both HDAC and G9a inhibition, the H3 mimicking quinazoline core of G9a inhibitors used as a base scaffold. Next, several modifications at different sites introduced to cover most of the possible chemical space related to the position and chain length (linker gap between the metal binding portion and G9a core). From this design, we synthesized more than 20 compounds and tested biochemically and in vitro to see if they displayed the desired dual activity. Our primary assessment of success was from MALDI-TOF evaluation of the H3K9 methylation profile, many of the compounds retained G9a inhibition potential. Cell-based assays for all the compounds against several cell lines were used to determine their inhibition potential, and we found that 13 and 14 displays the desired dual activities comparable to the controls SAHA and BIX-01294. Cell toxicity of these compounds was determined using CCK-8, showing that compound 14 was both more effective and less toxic compared to 13. The ADMET module of Discovery Studio 3.1 also predicted that the compound 14 has excellent physico-chemical properties, making it a viable drug candidate. Discovery of these small molecules with dual activity towards two epigenetic targets, HDACs and G9a, will
provide the route for developing similar compounds with high potency soon. It is also worth mentioning that compound 14 may also act as a valuable tool in investigating the multi-targeting strategy, and its possible impacts on epigenetic targets. As this is the first time this sort of work has been done in respect to these two targets, it may also provide the basis for understanding histone cross-talk among distinct epigenetic targets. With this broad prospective, we further plan for a detailed SAR study specifically to compound 14 to provide a more potent dual inhibitor in conjunction with a mechanistic reasoning to understand the synergistic effect of inhibiting both HDACs and G9a methyltransferases.

3.4 Experiment Sections

Reagents were purchased from commercial suppliers Sigma-Aldrich, Alfa Aesar, TCI and Acros, and were used without further purification unless otherwise indicated. Anhydrous solvents (e.g., DMF, DIPEA, MeOH, DCM) were purchased from Sigma-Aldrich. The synthetic progress was monitored using silica gel 60 F254 thin layer chromatography plates (Merck EMD Millipore). Microwave reactions were performed using Initiator for organic synthesis. Column chromatography purifications were performed on an Isolera one system using SNAP columns with KP-Sil silica or Zip Si columns with KP-Sil normal phase silica cartridges (unless otherwise stated). The nuclear magnetic resonance spectra were recorded on a 400 MHz spectrometer by Topspin 3.1 with solvents of CDCl$_3$ and CD$_3$OD. Chemical shifts described in ppm. Coupling constants, when reported, are described in hertz (Hz). High-resolution mass spectra (HRMS) data were acquired using orbitrap elite mass spectrometer with an electrospray ionization (ESI) source. All the samples were run under FT control at 600000 resolution. All temperatures are described in °C. The purity of all final compounds was confirmed by RP-HPLC analysis, was >95% or mentioned in the synthetic procedure. Analytical high performance liquid chromatography
(HPLC) was performed using a Waters Agilent 1260 infinity, column used was Agilent eclipse plus C18 3.5 µM reverse phase 150 mm×4.6 mm chromatography column. Samples were detected using a wavelength of 254 nm. All samples were analyzed using acetonitrile (0.1% TFA):water (0.1% TFA) 5-60% over 30 min and a flow rate of 0.4 ml/min. Preparative HPLC was performed using the XBridge prep C18, 5 µM, 10×150 mm column and a flow rate of 1 ml/min.

3.4.1 Synthesis

**H₃ (1-20, ARTKQTARKSTGGKAPRKL):** Peptide was synthesized through Fmoc-Strategy. Automated peptide synthesis was performed on Liberty Blue Peptide Synthesizer. Peptide were synthesized under microwave-assisted protocols on Wang resins. The deblock mixture was 20% piperidine in DMF. The following Fmoc-Lys(Boc)-Wang resin from Nova biochem were employed. The Fmoc-protected amino acids were purchased from Chempep. Cocktail of TFA/TIS/Dodt/H₂O (92.5:2.5:2.5:2.5) was used to cleave peptides off the resin. After cleavage, crude peptide was purified through a reverse phase C18 column (purchased from Agilent, Eclipse XDB-C18, 5 µm, 9.4*250mm).

**Procedure A:** General procedure for compounds 2, 2a and 2b. 4-amino-piperidines (18.01 mmol) were added to a solution of 2,4-dichloro-6,7-dimethoxyquinazoline (2.11 g, 8.14 mmol in DMF 20 ml), followed by the addition of N,N-diisopropylethylamine (1.5 ml, 8.62 mmol) and the resulting mixture was stirred at room temperature for 2 h until TLC showed that the starting material had disappeared. Water was added to the reaction mixture, and the resulting solution was extracted with ethyl acetate. The organic layer was washed with 0.5% acetic acid aqueous solution and brine, dried and concentrated to give the crude product, which was purified on flash column via eluting with hexane-ethyl acetate (20%) to get 3.0g of the desired compound, yield 80-86%. Spectral properties of the product was matched with the reported compounds.
**Procedure B**: General procedure for compounds 3, 3a and 3b. Compound 2 (6.0 mmol) was dissolved in 8 ml of isopropanol. To this solution was added tert-butyl (2-aminoethyl) carbamate (1.92 g, 12 mmol) and DIPEA (1.5 ml, 7.2 mmol). The resulting solution was placed inside a microwave at 160 °C for 10 min. After cooling, TLC indicated the reaction was completed. Solvent was removed under reduced pressure, the residue was dissolved in DCM, washed with the saturated NaHCO₃ solution. The combined organic phase was dried over Na₂SO₄ and concentrated under reduced pressure. The residue was purified on silica gel column, eluting with 5% MeOH in DCM (containing 0.5% Et₃N) to give 1.8 g of the Boc–protected amino compound as pale yellow solid, yield 60-66%.

**N2-(2-aminoethyl)-6,7-dimethoxy-N4-(1-methylpiperidin-4-yl)quinazoline-2,4-diamine (3)**: Brown solid, 1.8 g, 66% yield. M.P. 109-107 °C, ¹H NMR (400 MHz, CDCl₃) δ 7.10 (s, 1H), 7.06 (s, 1H), 6.19 (s, 1H), 4.33 – 4.20 (m, 1H), 3.91 (s, 6H), 3.17 (d, J = 5.4 Hz, 2H), 2.93 – 2.75 (m, 4H), 2.50 (s, 2H), 2.30 (s, 3H), 2.14 (m, 4H), 1.78 – 1.60 (m, 2H), 1.42 (s, 9H). ¹³C NMR (100 MHz, CDCl₃) δ 159.5, 155.9, 154.8, 153.2, 148.9, 147.9, 107.1, 106.7, 101.3, 80.2, 56.6, 56.1, 54.5, 46.8, 45.0, 39.2, 30.4, 27.8. HRMS (ESI): m/z calcd for C₂₃H₃₆N₆O₄ [M + H]⁺, 461.2876; found, 461.2862.

**N2-(2-aminoethyl)-6,7-dimethoxy-N4-(1-isopropylpiperidin-4-yl)quinazoline-2,4-diamine (3a)**: Brown solid, 1.7 g, 60% yield. M.P. 114-116 °C, ¹H NMR (400 MHz, CDCl₃) δ 7.23 (s, 1H), 7.19 (s, 1H), 4.13 (s, 1H), 3.84 (s, 6H), 3.58 – 3.44 (m, 4H), 3.39 (d, J = 4.9 Hz, 2H), 3.32 (d, J = 4.6 Hz, 2H), 2.96 – 2.83 (m, 2H), 2.11 (t, J = 11.3 Hz, 2H), 2.01 (d, J = 10.8 Hz, 2H), 1.74 – 1.59 (m, 2H), 1.40 (s, 9H), 1.01 (s, 6H). ¹³C NMR (100 MHz, CDCl₃) δ 165.1, 158.8, 156.3, 154.5, 145.6, 112.04, 108.9, 99.6, 80.8, 56.3, 55.9, 52.6, 49.9, 48.4, 41.4, 40.8, 31.7, 28.4. HRMS (ESI): m/z calcd for C₂₅H₄₀N₆O₄ [M + H]⁺, 489.3145; found, 489.3156.
N2-(2-aminoethyl)-6,7-dimethoxy-N4-(1-benzylpiperidin-4-yl)quinazoline-2,4-
diamine (3b): Brown solid, 2.1 g, 64% yield. M.P. 130-132 °C. 1H NMR (400 MHz, CDCl3) δ 7.36 – 7.14 (m, 5H), 6.79 (s, 1H), 6.58 (s, 1H), 6.01 (s, 2H), 5.44 (s, 1H), 4.11 (d, J = 12.4 Hz, 1H), 3.84 (m, 6H), 3.52 (s, 2H), 3.40 – 3.37 (m, 2H), 3.32 (d, J = 4.6 Hz, 2H), 2.93 – 2.84 (m, 2H), 2.11 (t, J = 11.3 Hz, 2H), 2.01 (d, J = 10.8 Hz, 2H), 1.75 – 1.59 (m, 2H), 1.38 (s, 9H). 13C NMR (100 MHz, CDCl3) δ 165.1, 158.8, 156.3, 154.5, 145.6, 138.1, 129.2, 128.2, 127.1, 112.01, 108.9, 102.3, 78.9, 63.0, 56.3, 55.9, 52.6, 49.9, 48.4, 41.4, 40.8, 31.7, 28.3. HRMS (ESI): m/z calcd for C29H40N6O4 [M + H]+, 537.3189; found, 537.3165.

NH-Boc protection was removed to get the free amines of 3, 3a and 3b using TFA/DCM overnight, dried amine was used directly in next step without further purification.

Procedure C: General procedure for compounds 4-7 and 4a-7a, to a stirred solution of corresponding monomethyl ester (0.25 mmol) in anhydrous CH2Cl2 (5 ml) was added EDCI (70 mg, 0.35 mmol) followed by HOBt (50 mg, 0.35 mmol) at 0 °C. After 30 min, a solution of compound 3 (138 mg, 0.3 mmol) and DIEPA (0.1 ml, 0.5 mmol) in CH2Cl2 (2 ml) was added drop-wise at 0 °C. The mixture was allowed to stir at rt and monitored by TLC. Upon completion, the organic layer was washed with saturated aqueous NaHCO3 solution followed by brine. The organic extracts were dried over Na2SO4, filtered, and concentrated under reduced pressure. The crude product was purified by flash chromatography (MeOH/DCM up to 20%) to afford desired compounds as colorless oily liquid. HRMS (ESI): m/z calcd for C27H42N6O5 [M + H]+, 531.3295; found, 531.3279. This intermediate in methanol (2.5 ml) was added a solution of hydroxylamine (1 ml, 50% in water). The resulting solution was stirred for 3 h at 60°C. Then solvent was removed under vacuum and the crude residue purified by flash chromatography using reverse phase silica
gel column using H₂O (0.1% HCOOH)/CH₃CN (0.1% HCOOH) as eluent (0-100 %). This afforded the expected derivatives as a yellow/brown sticky mass, 30-38% over 2 steps.

**N1-(2-((6,7-dimethoxy-4-((1-methylpiperidin-4-yl) amino) quinazolin-2-yl) amino) ethyl)-N8-hydroxyoctanediamide (4):** 44 mg, 33% yield. ¹H NMR (400 MHz, MeOD) δ 7.81 (s, 1H), 7.59 (s, 1H), 6.91 (s, 1H), 4.68 (s, 1H), 3.93 (s, 6H), 3.64-3.47 (m, 5H), 3.23 (d, J = 1.4 Hz, 4H), 3.14 (m, 1H), 2.90 (d, J = 16.0 Hz, 3H), 2.36-2.28 (m, 2H), 2.21-2.07 (m, 5H), 1.64 – 1.49 (m, 4H), 1.30 (s, 4H). ¹³C NMR (100 MHz, MeOD) δ 183.7, 167.5, 148.4, 128.3, 124.8, 117.1, 110.9, 104.0, 55.4, 54.7, 53.3, 52.8, 46.1, 42.1, 41.8, 38.2, 35.9, 32.3, 28.2, 25.4. HRMS (ESI): m/z calcd for C₂₆H₄₁N₇O₅ [M + H]⁺, 532.3247; found, 532.3248. HPLC purity 95.45% ; tₚ = 14.004

**N1-hydroxy-N8-(2-(((4-((1-isopropylpiperidin-4-yl)amino)-6,7-dimethoxyquinazolin-2-yl)amino)ethyl)octanediamide (4a):** 49 mg, 35% yield. ¹H NMR (400 MHz, MeOD) δ 7.76 (s, 1H), 7.70 (s, 1H), 6.90 (s, 1H), 4.72 (s, 1H), 3.92 (s, 6H), 3.60 (m, 5H), 3.48 - 3.38 (m, 4H), 3.33 (dd, J = 3.2, 1.6 Hz, 1H), 2.39-2.19 (m, 5H), 2.07 (t, J = 7.1 Hz, 2H), 1.56 (d, J = 5.5 Hz, 4H), 1.41-1.28 (m, 9H). ¹³C NMR (100 MHz, MeOD) δ 171.6, 166.3, 156.7, 153.1, 147.4, 147.2, 142.2, 135.7, 125.2, 124.7, 120.0, 117.3, 110.5, 104.1, 103.7, 57.9, 55.5, 45.0, 34.1, 32.1, 28.3, 25.0, 15.8. HRMS (ESI): m/z calcd for C₂₈H₄₅N₇O₅ [M + H]⁺, 560.3560; found, 560.3554. HPLC purity 95.12% ; tₚ = 14.820.

**N1-(2-((6,7-dimethoxy-4-((1-methylpiperidin-4-yl)amino)quinazolin-2-yl)amino)ethyl)-N7-hydroxyheptanediame (5):** 38 mg, 30% yield. ¹H NMR (400 MHz, MeOD) δ 7.70 (s, 1H), 6.97 (s, 1H), 4.69 (s, 2H), 3.96 (s, 6H), 3.85 (s, 1H), 3.67 – 3.49 (m, 4H), 3.47 (d, J = 5.7 Hz, 2H), 3.25 (d, J = 13.5 Hz, 2H), 3.15 (d, J = 7.4 Hz, 1H), 3.05 – 2.83 (m, 4H), 2.31 (d, J = 11.1 Hz, 2H), 2.23 – 2.06 (m, 5H), 1.64 – 1.52 (m, 3H), 1.35 (d, J = 6.9 Hz, 2H). ¹³C
NMR (100 MHz, MeOD) δ 174.8, 170.0, 163.8, 156.6, 156.0, 153.4, 147.3, 112.0, 108.9, 99.6, 56.8, 53.6, 46.0, 40.5, 38.7, 37.4, 34.4, 30.4, 28.3, 25.2. HRMS (ESI): m/z calcd for C_{25}H_{39}N_{7}O_{5} [M + H]^+; 518.3091; found, 518.3080. HPLC purity 95.21%; t_R = 14.402

N1-hydroxy-N7-(2-((4-((1-isopropylpiperidin-4-yl)amino)-6,7-dimethoxyquinazolin-2-yl)amino)ethyl)heptanediame (5a): 51 mg, 38% yield. \(^1\)H NMR (400 MHz, MeOD) δ 7.75 (s, 1H), 7.61 (s, 1H), 6.94 (s, 1H), 4.73 (s, 1H), 3.94 (s, 6H), 3.61 (m, 5H), 3.48 (s, 2H), 3.42 – 3.23 (m, 3H), 2.39-2.21 (s, 5H), 2.09 (t, J = 6.9 Hz, 3H), 1.60 (s, 4H), 1.38 (m, 8H). \(^13\)C NMR (100 MHz, MeOD) δ 170.8, 167.7, 164.5, 155.5, 154.2, 152.0, 147.3, 110.8, 104.2, 98.1, 55.4, 53.4, 52.6, 48.2, 47.1, 38.5, 37.1, 35.4, 31.9, 28.5, 25.6, 15.8. HRMS (ESI): m/z calcd for C_{27}H_{43}N_{7}O_{5} [M + H]^+; 546.3405; found, 546.3385. HPLC purity 93.80%; t_R = 14.991.

N1-(2-((4-((1-benzylpiperidin-4-yl)amino)-6,7-dimethoxyquinazolin-2-yl)amino)ethyl)-N7-hydroxyheptanediame (5b): 55 mg, 37% yield. \(^1\)H NMR (400 MHz, MeOD) δ 7.67 (d, J = 1.9 Hz, 1H), 7.44 (dt, J = 15.7, 7.9 Hz, 5H), 6.96 (s, 1H), 4.48 (s, 1H), 3.96 (dd, J = 12.2, 3.3 Hz, 7H), 3.63 (s, 3H), 3.54 – 3.44 (m, 2H), 3.28 (d, J = 11.8 Hz, 2H), 2.72 (s, 2H), 2.32 – 2.16 (m, 5H), 2.10 (t, J = 7.1 Hz, 1H), 1.94 (d, J = 13.8 Hz, 2H), 1.60 (ddd, J = 15.4, 12.7, 7.5 Hz, 4H), 1.41 – 1.20 (m, 2H). \(^13\)C NMR (100 MHz, MeOD) δ 174.9, 171.4, 167.7, 159.4, 156.3, 156.0, 153.3, 147.4, 136.1, 130.6, 130.4, 129.1, 128.7, 128.6, 103.6, 98.4, 97.9, 60.5, 55.5, 51.1, 40.1, 38.3, 35.3, 28.5, 27.9, 25.0, 24.8. HRMS (ESI): m/z calcd for C_{31}H_{43}N_{7}O_{5} [M + H]^+; 594.3440; found, 594.3460. HPLC purity 96.20%; t_R = 14.001.

N1-(2-((6,7-dimethoxy-4-((1-methylpiperidin-4-yl)amino)quinazolin-2-yl)amino)ethyl)-N6-hydroxyadipamide (6): 44 mg, 35% yield. \(^1\)H NMR (400 MHz, MeOD) δ 7.65 (s, 1H), 6.93 (s, 1H), 4.69 (s, 1H), 3.95 (s, 6H), 3.60 (d, J = 14.8 Hz, 4H), 3.21 (m, 3H), 2.89 (d, J = 9.2 Hz, 4H), 2.29 (m, 4H), 2.13 (s, 3H), 1.90 (dt, J = 13.6, 6.7 Hz, 1H), 1.64 (s, 4H), 1.39
(d, J = 6.6 Hz, 2H). $^{13}$C NMR (100 MHz, MeOD) δ 170.8, 168.2, 159.3, 156.5, 156.2, 153.3, 147.3, 111.2, 108.6, 103.0, 55.5, 52.9, 46.0, 41.8, 40.56, 38.2, 37.4, 31.8, 28.2, 24.6. HRMS (ESI): m/z calcd for C$_{24}$H$_{37}$N$_7$O$_5$ [M + H]$^+$, 504.2934; found, 504.2911. HPLC purity 96.81% ; t$_R$ = 13.374

N1-hydroxy-N6-(2-((1-isopropylpiperidin-4-yl)amino)-6,7-dimethoxyquinazolin-2-yl)amino)ethyl) adipamide (6a): 46 mg, 35% yield. $^1$H NMR (400 MHz, MeOD) δ 7.77 (s, 1H), 7.64 (s, 1H), 7.32 (s, 1H), 6.96 (s, 1H), 4.73 (s, 1H), 3.96 (s, 6H), 3.72 – 3.52 (m, 5H), 3.47 (s, 2H), 3.37 (d, J = 15.2 Hz, 2H), 2.39 (d, J = 12.1 Hz, 2H), 2.17 (m, 6H), 1.63 (s, 4H), 1.43 (d, J = 6.3 Hz, 6H). $^{13}$C NMR (100 MHz, MeOD) δ 174.5, 170.5, 168.0, 157.1, 156.0, 153.5, 147.3, 113.4, 108.4, 89.7, 57.2, 55.8, 49.2, 49.5, 40.8, 39.7, 38.0, 32.5, 28.2, 24.7, 15.5. HRMS (ESI): m/z calcd for C$_{26}$H$_{41}$N$_7$O$_5$ [M + H]$^+$, 532.3247; found, 532.3245. HPLC purity 96.95% ; t$_R$ = 14.164

N1-(2-((6,7-dimethoxy-4-((1-methylpiperidin-4-yl)amino)quinazolin-2-yl)amino)ethyl)-N5-hydroxyglutaramide (7): 40 mg, 33% yield. $^1$H NMR (400 MHz, MeOD) δ 7.68 (s, 1H), 7.63 (s, 1H), 7.30 (s, 1H), 6.84 (s, 1H), 4.66 (s, 1H), 3.92 (s, 6H), 3.61 (s, 4H), 3.46 (s, 2H), 3.35 (d, J = 15.4 Hz, 3H), 2.90 (m, 3H), 2.29 (m, 4H), 2.14 (d, J = 6.6 Hz, 4H), 1.91 (s, 2H). $^{13}$C NMR (100 MHz, MeOD) δ 170.8, 168.2, 159.4, 159.3, 156.5, 153.3, 147.3, 136.1, 122.1, 111.2, 108.6, 103.7, 55.5, 52.9, 46.1, 41.8, 40.0, 38.4, 38.2, 35.2, 31.8, 28.2, 24.6. HRMS (ESI): m/z calcd for C$_{25}$H$_{35}$N$_7$O$_5$ [M + H]$^+$, 490.2778; found, 490.2756. HPLC purity 95.61% ; t$_R$ = 12.751

N1-hydroxy-N5-(2-((1-isopropylpiperidin-4-yl)amino)-6,7-dimethoxyquinazolin-2-yl)amino)ethyl) glutaramide (7a): 50 mg, 38% yield. $^1$H NMR (400 MHz, MeOD) δ 7.67 (s, 1H), 6.93 (s, 1H), 4.72 (s, 2H), 3.94 (s, 6H), 3.59 (d, J = 12.4 Hz, 5H), 3.47 (s, 2H), 3.36 (d, J = 13.9 Hz, 2H), 2.37 (s, 2H), 2.22-2.12 (m, 6H), 1.63 (m, 4H), 1.44 (s, 6H). $^{13}$C NMR (100 MHz, MeOD) δ 173.1, 168.0, 158.5, 156.5, 153.3, 134.9, 147.5, 110.6, 103.7, 98.5, 57.8, 55.2, 40.1,
38.4, 35.7, 32.4, 31.7, 28.5, 16.2, 15.8. HRMS (ESI): \( m/z \) calcd for C_{25}H_{39}N_{7}O_{5} [M + H]^+; 518.3091; found, 518.3139. HPLC purity 93.38% ; \( t_R = 13.670 \).

**Compounds 13-16 and 13a-16a**

**2,4-dichloro-7-methoxyquinazoline:** Compound 10 was prepared according to the previously reported procedure,\(^{[104]}\) 3.4 g of anthranilic acid (20 mmol) and 3.5 equiv. of urea were finely powdered using mortar and pestle and heated to 200 °C in a round-bottom flask open to the atmosphere. After 2 h, the mixture was cooled, triturated with water, and filtered to give the product as crude. Product was dried and used in next step directly. Molecular ion peak for C_{9}H_{6}Cl_{2}N_{2}O_{3} was found at 192.0773. Crude quinazoline-2,4-dione and 2.4 g of N,N-diethyl aniline were mixed in 45 ml of phosphorus oxychloride, and the mixture was refluxed overnight under an argon atmosphere. The crude reaction mixture was concentrated, neutralized the excess of POCl_{3} using NaHCO_{3} and extracted to EA, dried on Na_{2}SO_{4} and evaporated, purified using flash column, eluting at 20% of EA/Hexane. White fluffy powder, 1.82 g, 40% overall yield. HRMS (ESI): \( m/z \) calcd for C_{9}H_{6}Cl_{2}N_{2} [M + H]^+; 228.9935; found, 228.9934.

**N-(1-benzylpiperidin-4-yl)-2-chloro-7-methoxyquinazolin-4-amine (11) and 2-chloro-N-(1-isopropylpiperidin-4-yl)-7-methoxyquinazolin-4-amine (11a).**

Compound 11 and 11a were prepared according to the procedure A, using 1-methylpiperidin-4-amine or 1-benzylpiperidin-4-amine. **11:** Yellow powder, 74%. M.P. 134-136 °C, \(^1\)H NMR (CDCl_{3}, 400 MHz) \( \delta \) ppm 7.54 (d, \( J = 9.0 \) Hz, 1H), 7.27-7.34(m, 5H), 7.10 (d, \( J = 2.4 \) Hz, 1H), 7.04 (dd, J1 = 9.0 Hz, J2 = 2.4 Hz, 1H), 5.61(d, \( J = 7.71 \) Hz, 1H), 4.23-4.33 (m, 1H), 3.88 (s, 3H), 3.57 (s, 2H), 2.91 (d, \( J = 11.9 \) Hz, 2H), 2.24-2.30 (m, 2H), 2.08-2.13 (m, 2H), 1.59-1.69 (m, 2H). \(^{13}\)C NMR (CDCl_{3}, 100 MHz) \( \delta \) ppm 163.7, 129.8, 158.3, 153.3, 137.7, 129.3, 128.3,
127.3, 122.0, 117.9, 107.2, 106.9, 62.9, 55.7, 52.0, 48.0, 31.9. HRMS (ESI): m/z calcd for C_{21}H_{23}ClN_4O [M + H]^+, 383.1639; found, 383.1610.

11a: Brownish yellow semi solid, 86\%. \textsuperscript{1}H NMR (400 MHz, CDCl\textsubscript{3}) δ 7.56 (d, J = 9.1 Hz, 1H), 7.13 (s, 1H), 7.07 (d, J = 9.1 Hz, 1H), 4.28 (s, 1H), 3.92 (s, 3H), 2.88 (d, J = 11.0 Hz, 2H), 2.35 (s, 3H), 2.26 (t, J = 11.6 Hz, 2H), 2.17 (d, J = 12.2 Hz, 2H), 1.86 (s, 1H), 1.66 (m, 2H). HRMS (ESI): m/z calcd for C_{15}H_{19}ClN_4O [M + H]^+, 306.1247; found, 307.1323.

N2-(2-aminoethyl)-N4-(1-benzylpiperidin-4-yl)-7-methoxyquinazoline-2,4-diamine (12) and N2-(2-aminoethyl)-7-methoxy-N4-(1-methylpiperidin-4-yl)quinazoline-2,4-diamine (12a)

Compounds 12 and 12a were obtained via Procedure B:

12: Brown solid, 1.94 g, 64\%. \textsuperscript{1}H NMR (400 MHz, CDCl\textsubscript{3}) δ 8.75 (s, 2H), 8.49 (d, J = 5.2 Hz, 1H), 7.96 (s, 1H), 7.37 – 7.14 (m, 4H), 6.62 (d, J = 8.9 Hz, 2H), 5.56 (s, 1H), 4.22 (s, 1H), 4.06 – 3.88 (m, 1H), 3.67 (d, J = 10.9 Hz, 3H), 3.51 (d, J = 19.7 Hz, 4H), 3.26 (s, 2H), 2.99 (dd, J = 14.9, 7.4 Hz, 1H), 2.90 (s, 2H), 2.09 (d, J = 35.0 Hz, 4H), 1.91 (d, J = 12.4 Hz, 3H), 1.28 (d, J = 10.5 Hz, 10H). HRMS (ESI): m/z calcd for C_{28}H_{38}N_6O_3 [M + H]^+, 507.3084; found, 507.3047.

12a: Brown solid, 1.75 g, 68\%. \textsuperscript{1}H NMR (400 MHz, CDCl\textsubscript{3}) δ 7.46 (d, J = 8.9 Hz, 1H), 6.84 (s, 1H), 6.74 (d, J = 8.8 Hz, 1H), 5.55 (s, 2H), 4.18 (s, 1H), 3.88 (s, 3H), 3.61 (d, J = 3.9 Hz, 2H), 3.39 (d, J = 4.8 Hz, 2H), 2.87 (d, J = 11.2 Hz, 2H), 2.34 (s, 3H), 2.21 (t, J = 11.3 Hz, 2H), 2.12 (d, J = 11.8 Hz, 2H), 1.66 (dd, J = 21.0, 10.4 Hz, 2H), 1.44 (s, 9H). HRMS (ESI): m/z calcd for C_{22}H_{34}N_6O_3 [M + H]^+, 431.2771; found, 431.2767.

Compounds 13-17 and 13a-17a were synthesized according to procedure C from the corresponding free amines, Yield varied from 30-40\%, yellow/brown sticky solids were obtained after purification.
N1-(2-((4-((1-benzylpiperidin-4-yl)amino)-7-methoxyquinazolin-2-yl)amino)ethyl)-
N8-hydroxyoctanediamide (13): 45 mg, 31% yield. $^1$H NMR (400 MHz, MeOD) δ 8.06 (s, 1H), 7.52 (d, $J = 33.6$ Hz, 5H), 6.83 (d, $J = 31.2$ Hz, 2H), 4.62 (s, 1H), 4.34 (s, 2H), 3.89 (s, 3H), 3.70 – 3.40 (m, 6H), 3.22 – 3.03 (m, 3H), 2.21 (m, 8H), 1.56 (s, 4H), 1.29 (s, 5H). $^{13}$C NMR (100 MHz, MeOD) δ 175.0, 171.5, 165.1, 159.8, 154.0, 151.3, 141.9, 130.9, 129.7, 129.6, 128.8, 125.4, 113.7, 102.9, 98.0, 59.9, 55.3, 50.7, 40.0, 38.3, 35.7, 32.2, 28.4, 28.3, 27.9, 25.4, 25.1. HRMS (ESI): m/z calcd for C$_{31}$H$_{43}$N$_7$O$_4$ [M + H]$^+$, 578.3455; found, 578.3444. HPLC purity 95.41%; $t_R =$ 16.756.

N1-hydroxy-N7-(2-((7-methoxy-4-((1-methylpiperidin-4-yl)amino)quinazolin-2-
yl)amino)ethyl)heptanediamide (13a): 44 mg, 33% yield. $^1$H NMR (400 MHz, CDCl$_3$) δ 7.46 (s, 1H), 6.85 – 6.49 (m, 2H), 4.20 (s, 1H), 3.87 (s, 3H), 3.70 – 3.46 (m, 4H), 2.95 (d, $J = 10.6$ Hz, 2H), 2.36 (s, 3H), 2.23 (dd, $J = 21.2$, 12.2 Hz, 6H), 2.15 – 2.00 (m, 6H), 1.81 (d, $J = 10.6$ Hz, 3H), 1.65 (s, 5H). $^{13}$C NMR (100 MHz, MeOD) δ 175.0, 168.6, 165.2, 153.8, 149.2, 134.8, 132.5, 125.4, 113.7, 108.7, 102.9, 98.0, 55.2, 52.8, 46.2, 42.2, 40.0, 38.2, 35.6, 32.2, 28.4, 28.3, 28.1, 25.3, 25.0. HRMS (ESI): m/z calcd for C$_{25}$H$_{39}$N$_7$O$_4$ [M + H]$^+$, 502.3142; found, 502.3143. HPLC purity 95.75%; $t_R =$ 13.767.

N1-(2-((4-((1-benzylpiperidin-4-yl)amino)-7-methoxyquinazolin-2-yl)amino)ethyl)-
N7-hydroxyheptanediamide (14): 49 mg, 35% yield. $^1$H NMR (400 MHz, MeOD) δ 8.10 (d, $J = 9.1$ Hz, 1H), 7.59 – 7.40 (m, 5H), 6.97 (d, $J = 9.1$ Hz, 1H), 6.90 (s, 1H), 4.60 (s, 1H), 4.18 (s, 2H), 3.94 (s, 3H), 3.64 (d, $J = 8.0$ Hz, 2H), 3.46 (dd, $J = 14.1$, 8.1 Hz, 4H), 3.39 – 3.31 (m, 2H), 3.18 – 2.96 (m, 2H), 2.23 (dd, $J = 19.2$, 11.7 Hz, 4H), 2.15 – 1.92 (m, 4H), 1.69 – 1.53 (m, 4H), 1.41 – 1.27 (m, 2H). $^{13}$C NMR (100 MHz, MeOD) δ 175.0, 171.4, 167.8, 165.3, 159.9, 154.0, 141.9, 131.2, 130.5, 129.1, 128.7, 125.3, 113.8, 103.0, 98.2, 60.5, 55.1, 51.0, 40.2, 38.2, 35.3, 32.0, 28.4,
28.0, 25.0, 24.8. HRMS (ESI): m/z calcd for C_{30}H_{41}N_{7}O_{4} [M + H]^{+}, 564.3298; found, 564.3307.
HPLC purity 95.02%; t_R= 16.600.

**N1-hydroxy-N7-(2-((7-methoxy-4-((1-methylpiperidin-4-yl)amino)quinazolin-2yl)amino)ethyl) heptanediadime (14a):** 49 mg, 40% yield. \(^1\)H NMR (400 MHz, MeOD) δ 8.10 (s, 1H), 7.17 (s, 1H), 6.94 (m, 1H), 4.62 (d, J = 62.4 Hz, 2H), 3.94 (s, 3H), 3.86 (s, 2H), 3.73 – 3.55 (m, 2H), 3.47 (s, 1H), 3.29 (m, 5H), 3.15 (dd, J = 14.9, 7.5 Hz, 2H), 2.97 – 2.77 (m, 4H), 2.45 – 2.17 (m, 3H), 2.24 – 2.02 (m, 3H), 1.61 (s, 2H), 1.42 – 1.22 (m, 2H). \(^{13}\)C NMR (100 MHz, CDCl\(_3\)) δ 178.5, 174.4, 166.0, 165.6, 159.2, 153.1, 129.78, 117.7, 115.8, 106.9, 59.6, 53.5, 50.7, 46.8, 42.1, 39.7, 39.3, 36.0, 32.0, 29.8, 28.4. HRMS (ESI): m/z calcd for C\(_{24}\)H\(_{37}\)N\(_7\)O\(_4\) [M + H]^{+}, 487.2907; found, 488.2962. HPLC purity 94.16%; t_R= 13.232.

**N1-(2-((4-((1-benzylpiperidin-4-yl)amino)-7-methoxyquinazolin-2-yl)amino)ethyl)-N6-hydroxyadipamide (15):** 52 mg, 38% yield. \(^1\)H NMR (400 MHz, MeOD) δ 8.03 (s, 1H), 7.52 (m, 6H), 6.78 (d, J = 40.3 Hz, 2H), 4.61 (s, 1H), 4.36 (s, 2H), 3.87 (s, 3H), 3.60 (m, 4H), 3.46 (m, 2H), 3.38 (m, 3H), 2.49 – 1.78 (m, 8H), 1.64 (s, 4H). \(^{13}\)C NMR (100 MHz, MeOD) δ 171.25, 167.79, 165.02, 159.78, 153.88, 139.89, 131.05, 129.70, 129.50, 128.93, 113.70, 111.3, 103.5, 66.8, 59.84, 55.36, 50.73, 41.6, 39.2, 38.4, 35.38, 32.02, 27.80. HRMS (ESI): m/z calcd for C\(_{24}\)H\(_{37}\)N\(_7\)O\(_4\) [M + H]^{+}, 550.3142; found, 550.3148. HPLC purity 96.48%; t_R= 16.262.

**N1-hydroxy-N6-(2-((7-methoxy-4-((1-methylpiperidin-4-yl)amino)quinazolin-2-yl)amino)ethyl)adipamide (15a):** 41 mg, 35% yield. \(^1\)H NMR (400 MHz, MeOD) δ 8.08 (s, 1H), 6.87 (d, J = 33.0 Hz, 2H), 4.68 (s, 2H), 3.88 (d, J = 22.3 Hz, 3H), 3.82 – 3.69 (m, 1H), 3.62 (s, 4H), 3.46 (s, 2H), 3.22 (m, 2H), 3.28 – 3.17 (m, 1H), 3.19 – 3.01 (m, 1H), 2.89 (d, J = 15.2 Hz, 3H), 2.42 – 2.19 (m, 3H), 2.09 (d, J = 29.5 Hz, 2H), 1.90 (s, 1H), 1.62 (s, 3H), 1.38 (d, J = 6.5 Hz, 2H). \(^{13}\)C NMR (100 MHz, MeOD) δ 173.7, 170.0, 167.9, 164.5, 159.2, 153.4, 141.4, 124.6, 113.0,
124

102.3, 97.4, 54.4, 53.6, 52.2, 46.2, 45.5, 41.6, 39.3, 37.5, 34.3, 34.0, 27.5. HRMS (ESI): \( m/z \) calcd for \( \text{C}_{23}\text{H}_{35}\text{N}_{7}\text{O}_{4} [\text{M} + \text{H}]^+ \), 474.2829; found, 474.2807. HPLC purity 96.40%; \( t_R = 12.879 \).

\textbf{N1-(2-((4-(1-benzylpiperidin-4-yl)amino)-7-methoxyquinazolin-2-yl)amino)ethyl)-N5-hydroxyglutaramide (16):} 52 mg, 39% yield. \(^1\)H NMR (400 MHz, MeOD) \( \delta \) 8.01 (s, 1H), 7.51 (d, \( J = 38.6 \text{ Hz} \), 5H), 6.76 (d, \( J = 47.4 \text{ Hz} \), 2H), 4.49 (d, \( J = 103.1 \text{ Hz} \), 3H), 3.86 (s, 3H), 3.53 (d, \( J = 59.5 \text{ Hz} \), 5H), 3.22 (s, 3H), 2.23 (m, 6H), 1.93 (s, 4H). \(^{13}\)C NMR (100 MHz, MeOD) \( \delta \) 174.1, 170.8, 167.9, 165.0, 159.7, 153.6, 141.7, 131.0, 129.0, 129.8, 128.3, 125.4, 115.4, 113.6, 102.8, 67.9, 59.9, 55.3, 50.7, 40.1, 38.3, 34.8, 31.7, 27.8, 21.6. HRMS (ESI): \( m/z \) calcd for \( \text{C}_{23}\text{H}_{35}\text{N}_{7}\text{O}_{4} [\text{M} + \text{H}]^+ \), 536.2985; found, 536.2998. HPLC purity 94.34%; \( t_R = 16.051 \).

\textbf{N1-hydroxy-N5-(2-((7-methoxy-4-((1-methylpiperidin-4-yl)amino)quinazolin-2-yl)amino)ethyl)glutaramide (16a):} 35 mg, 31% yield. \(^1\)H NMR (400 MHz, MeOD) \( \delta \) 8.10 (s, 1H), 6.87 (d, \( J = 31.2 \text{ Hz} \), 2H), 4.70 (s, 1H), 3.92 (s, 3H), 3.73 (dd, \( J = 12.9, 6.5 \text{ Hz} \), 1H), 3.66 (m, 3H), 3.46 (s, 2H), 3.35 (d, \( J = 15.5 \text{ Hz} \), 2H), 3.23 (m, 1H), 2.88 (d, \( J = 14.5 \text{ Hz} \), 2H), 2.46 – 2.21 (m, 4H), 2.15 (s, 2H), 1.90 (s, 2H), 1.37 (m, 4H). \(^{13}\)C NMR (100 MHz, MeOD) \( \delta \) 175.8, 171.2, 166.0, 165.2, 153.3, 151.3, 127.6, 113.2, 110.1, 103.0, 57.8, 54.6, 50.5, 46.9, 41.8, 39.2, 36.4, 33.6, 31.4, 19.0. HRMS (ESI): \( m/z \) calcd for \( \text{C}_{22}\text{H}_{33}\text{N}_{7}\text{O}_{4} [\text{M} + \text{H}]^+ \), 460.2672; found, 460.2648. HPLC purity 95.90%; \( t_R = 12.615 \).

\textbf{N-(1-benzylpiperidin-4-yl)-6,7-dimethoxy-2-(4-methyl-1,4-diazepan-1-yl)quinazolin-4-amine (18):} Compound 18 was synthesized according to the previously reported procedure,\(^{103}\) which was treated with Pd/C under \( \text{H}_2 \) gas to get the free amine. HRMS (ESI): \( m/z \) calcd for \( \text{C}_{21}\text{H}_{32}\text{N}_{6}\text{O}_{2} [\text{M} + \text{H}]^+ \), 401.2264; found, 401.2642. This amine was directly used in procedure C while using monomethylsuberate ester to get 19 and monomethylpimelate to obtain 20.
8-(4-((6,7-dimethoxy-2-(4-methyl-1,4-diazepan-1-yl)quinazolin-4-yl)amino)piperidin-1-yl)-N-hydroxy-8-oxooctanamide (19): 48 mg, 34% yield over 2 steps. $^1$H NMR (400 MHz, MeOD) δ 7.68 (s, 1H), 7.20 (s, 1H), 4.66 (d, $J = 12.0$ Hz, 1H), 4.51 (s, 1H), 4.22 (s, 2H), 4.12 (d, $J = 12.7$ Hz, 1H), 3.96 (m, 8H), 3.45 (s, 2H), 3.28 (d, $J = 19.7$ Hz, 3H), 2.83 (d, $J = 17.7$ Hz, 4H), 2.47 (dd, $J = 15.0, 7.3$ Hz, 2H), 2.36 (s, 2H), 2.16 (m, 4H), 1.65 (m, 6H), 1.40 (s, 5H). $^{13}$C NMR (101 MHz, MeOD) δ 172.6, 171.5, 167.5, 158.5, 155.8, 152.9, 147.6, 103.4, 102.7, 99.6, 56.3, 55.5, 55.4, 44.7, 43.7, 42.7, 40.7, 32.4, 32.2, 31.5, 30.6, 28.5, 28.3, 25.1, 25.0, 24.3. HRMS (ESI): $m/z$ calcd for C$_{29}$H$_{45}$N$_7$O$_{5}$[M + H]$^+$, 572.3516; found, 572.3530. HPLC purity 94.71%; $t_R= 15.347$.

7-(4-((6,7-dimethoxy-2-(4-methyl-1,4-diazepan-1-yl)quinazolin-4-yl)amino)piperidin-1-yl)-N-hydroxy-7-oxoheptanamide(20): 62 mg, 45% yield over two steps. $^1$H NMR (400 MHz, MeOD) δ 7.72 (s, 1H), 7.23 (s, 1H), 4.66 (d, $J = 12.1$ Hz, 1H), 4.52 (s, 1H), 4.26 (s, 2H), 4.12 (d, $J = 12.0$ Hz, 2H), 3.96 (d, $J = 14.8$ Hz, 8H), 3.49 (d, $J = 39.9$ Hz, 4H), 2.86 (dd, $J = 25.3, 14.7$ Hz, 4H), 2.64 – 2.29 (m, 4H), 2.27 – 1.92 (m, 4H), 1.83 – 1.50 (m, 6H), 1.50 – 1.34 (m, 2H). $^{13}$C NMR (100 MHz, MeOD) δ 172.5, 171.4, 167.2, 158.5, 155.9, 152.5, 147.7, 137.0, 103.5, 102.7, 99.2, 56.18, 55.6, 55.5, 49.4, 45.9, 44.6, 43.5, 42.4, 40.2, 32.4, 32.1, 31.4, 30.6, 28.2, 25.0, 24.8, 24.0. HRMS (ESI): $m/z$ calcd for C$_{28}$H$_{43}$N$_7$O$_5$[M + H]$^+$, 558.3404; found, 558.3387. HPLC purity 96.04%; $t_R= 14.311$.

Compounds 21 and 22

Compound 10 was treated with NHBoc ethyldiaminediamine as the procedure A to get the intermediate 11b, which was further treated with 1-methyl-1,4-diazepane in accordance to procedure B to yield 18a, followed by procedure C, using monomethylsuberate or monomethylpimelate to get 21 and 22.
tert-butyl (2-((2-chloro-7-methoxyquinazolin-4-yl)amino)ethyl)carbamate (11b): 78% yield. $^1$H NMR (400 MHz, CDCl$_3$) $\delta$ 7.70 (d, $J$ = 9.0 Hz, 2H), 7.08 – 6.88 (m, 2H), 5.39 (s, 1H), 3.85 (s, 3H), 3.67 (d, $J$ = 3.9 Hz, 2H), 3.57 – 3.37 (m, 2H), 1.40 (s, 9H). HRMS (ESI): $m/z$ calcd for C$_{16}$H$_{21}$ClN$_4$O$_3$+ [M + H]$^+$, 353.1380; found, 353.1372.

tert-butyl(2-((7-methoxy-2-(4-methyl-1,4-diazepan-1-yl)quinazolin-4-yl)amino)ethyl)carbamate (18a): 69% yield. $^1$H NMR (400 MHz, CDCl$_3$) $\delta$ 7.56 (d, $J$ = 8.9 Hz, 1H), 6.95 (s, 1H), 6.85 (s, 1H), 6.63 (s, 1H), 5.52 (s, 1H), 4.04 – 3.93 (m, 2H), 3.86 (s, 3H), 3.62 (d, $J$ = 4.9 Hz, 2H), 3.44 (d, $J$ = 4.5 Hz, 3H), 2.73 (s, 2H), 2.69 – 2.51 (m, 2H), 2.37 (s, 3H), 2.03 (s, 2H), 1.41 (s, 9H). HRMS (ESI): $m/z$ calcd for C$_{22}$H$_{34}$N$_6$O$_3$[M + H]$^+$, 431.2771; found, 431.2746.

$N_1$-hydroxy-$N_8$-(2-((7-methoxy-2-(4-methyl-1,4-diazepan-1-yl)quinazolin-4-yl)amino)ethyl)octanediamide (21): 36 mg, 29% yield over 2 steps. $^1$H NMR (400 MHz, MeOD) $\delta$ 7.97 (d, $J$ = 8.7 Hz, 1H), 7.16 (s, 1H), 7.02 (d, $J$ = 8.6 Hz, 1H), 4.28 (s, 2H), 3.94 (s, 3H), 3.87 (s, 2H), 3.74 (d, $J$ = 5.6 Hz, 3H), 3.60 – 3.48 (m, 4H), 3.42 (s, 2H), 2.89 (d, $J$ = 7.3 Hz, 3H), 2.39 (s, 2H), 2.21 (t, $J$ = 7.0 Hz, 4H), 1.55 (m, 4H), 1.29 (s, 4H). $^{13}$C NMR (100 MHz, MeOD) $\delta$ 175.5, 167.49 164.98 159.76 153.4, 124.9, 114.1, 103.6, 99.8, 55.3, 55.1, 48.2, 48.0, 47.8, 47.6, 47.4, 47.2, 46.9, 46.0, 43.5, 42.6, 41.3, 37.5, 35.6, 34.0, 28.5, 28.4, 25.4, 24.7, 24.0. HRMS (ESI): $m/z$ calcd for C$_{25}$H$_{39}$N$_7$O$_4$[M + H]$^+$, 502.3142; found, 502.3128. HPLC purity 96.21%; $t_R$ = 13.810.

$N_1$-hydroxy-$N_7$-(2-((7-methoxy-2-(4-methyl-1,4-diazepan-1-yl)quinazolin-4-yl)amino)ethyl)heptanediamide (22): 43 mg, 36% yield over 2 steps. $^1$H NMR (400 MHz, MeOD) $\delta$ 7.98 (d, $J$ = 9.1 Hz, 1H), 7.14 (d, $J$ = 1.9 Hz, 1H), 7.00 (m, 1H), 4.27 (s, 2H), 3.93 (s, 5H), 3.88 (s, 1H), 3.88 – 3.66 (m, 4H), 3.52 (m, 5H), 3.42 (s, 2H), 2.90 (s, 3H), 2.39 (s, 1H), 2.24 (dd, $J$ = 10.9, 7.1 Hz, 5H), 1.84 (dd, $J$ = 13.7, 6.5 Hz, 2H). $^{13}$C NMR (100 MHz, MeOD) $\delta$ 174.99, 171.13, 168.17, 164.87, 159.67, 153.28, 125.03, 114.13, 103.56, 99.82, 55.19, 43.67, 42.61, 41.15,
37.62, 35.26, 31.88, 24.93, 24.60. HRMS (ESI): m/z calcd for C\textsubscript{24}H\textsubscript{37}N\textsubscript{7}O\textsubscript{4}[M + H]\textsuperscript{+}, 488.2985; found, 488.2960. HPLC purity 96.80%; t\textsubscript{R} = 13.370.

7-(benzyloxy)-N-(1-isopropylpiperidin-4-yl)-6-methoxy-2-(4-methyl-1,4-diazepan-1-yl)quinazolin-4-amine (24) was synthesized from 1a, by following procedure A (23) and procedure B (24), then benzy group was removed using Pd catalyzed hydrogenolysis, mixture of compound 24 (600 mg, 1.2 mmol) and 10 wt% Pd(OH)\textsubscript{2}/C (90 mg) in ethanol (100 ml) was stirred for 40 hours at room temperature under hydrogen balloon. The reaction mixture was filtered and concentrated to provide the debenzylated product 4-((1-isopropylpiperidin-4-yl)amino)-6-methoxy-2-(4-methyl-1,4-diazepan-1-yl) quinazolin-7-ol (25) as brownish yellow solid, 90%.

N-hydroxy-7-((4-((1-isopropylpiperidin-4-yl)amino)-6-methoxy-2-(4-methyl-1,4-diazepan-1-yl)quinazolin-7-yl)oxy)heptanamide (26): Procedure D. Ethyl heptanoate (80 µl, 0.4 mmol) was added to the ice cold solution of compound 24 (200 mg, 0.4 mmol) in DMF and K\textsubscript{2}CO\textsubscript{3} (280 mg, 2 mmol), reaction mixture was warmed to room temperature and then at 60 °C. After 6 h reactions, mixture was evaporated to get the residue and dissolved in DCM and washed with brine, organic layer was vacuum dried and eluted in flash column using reverse phase silica at 40 % ACN/H\textsubscript{2}O to get the intermediate ester, which was then dissolved in 2 ml of MeOH and treated with 50% NH\textsubscript{2}OH/water mixture (1 ml) overnight to afford the targeted product. Reaction mixture was dried and purified using reverse column and further by HPLC using ACN (0.1% HCOOH) /H\textsubscript{2}O (0.1% HCOOH) as eluent. 94 mg, 40% overall yield. \textsuperscript{1}H NMR (400 MHz, MeOD) \(\delta\) 7.61 (s, 1H), 7.06 (s, 1H), 4.14 (s, 3H), 3.97 (d, J = 14.7 Hz, 6H), 3.50 (s, 4H), 3.15 (s, 3H), 3.05 (s, 3H), 2.74 – 2.57 (m, 3H), 2.38 (s, 3H), 2.22 – 2.03 (m, 3H), 1.90 (d, J = 24.1 Hz, 3H), 1.63 (d, J = 48.4 Hz, 4H), 1.48-1.42 (m, 3H), 1.40-1.31 (m, 6H). \textsuperscript{13}C NMR (100 MHz, MeOD) \(\delta\) 170.9, 164.2, 158.0, 155.5, 152.0, 146.6, 110.9, 108.5, 101.5, 70.3, 57.4, 56.8, 56.6, 53.1, 46.4, 44.3, 32.4,
29.8, 29.2, 28.8, 26.8, 25.1, 24.1, 20.3. HRMS (ESI): m/z calcd for C\textsubscript{36}H\textsubscript{49}N\textsubscript{7}O\textsubscript{4} [M + H]\textsuperscript{+}, 572.3924; found, 572.3925. HPLC purity 96.80%; t\textsubscript{R}= 13.370.

8-((4-((1-benzylpiperidin-4-yl)amino)-2-(4-methyl-1,4-diazepan-1-yl)quinazolin-7-yl)oxy)- N-hydroxyoctanamide (30): Targeted analog 30 was synthesized from the anthranillic acid starting material.

7-(benzyl(oxy))-N-(1-benzylpiperidin-4-yl)-2-(4-methyl-1,4-diazepan-1-yl)quinazolin-4-amine (27): prepared according to the procedure A and B.\textsuperscript{1}H NMR (400 MHz, CDCl\textsubscript{3}) δ 7.90 (d, J = 3.6 Hz, 1H), 7.26-7.18 (m, 5H), 7.12 (s, 1H), 6.98 (d, J = 3.9, 1H), 5.87 (s, 1H), 4.39-4.12 (m, 5H), 3.92 (s, 3H), 3.63 (s, 2H), 3.20 (d, J = 4.6 Hz, 2H), 3.04 (d, J = 2.4 Hz, 2H), 2.52 (d, J = 2.4 Hz, 2H), 2.28 (s, 3H), 1.82-1.66 (m, 6H). \textsuperscript{13}C NMR (100 MHz, CDCl\textsubscript{3}) δ 163.0, 159.0, 158.6, 138.4, 129.1, 128.2, 127.0, 122.2, 112.1, 104.8, 104.3, 63.1, 58.8, 57.2, 55.3, 52.4, 48.2, 46.6, 45.9, 45.8, 32.0, 27.6. MALDI-TOF: m/z for C\textsubscript{27}H\textsubscript{36}N\textsubscript{6}O [M + H]\textsuperscript{+} is 461.9.

4-((1-benzylpiperidin-4-yl)amino)-2-(4-methyl-1,4-diazepan-1-yl)quinazolin-7-ol (28): Aryl demethylation using BBr\textsubscript{3} was employed.\textsuperscript{[111]} BBr\textsubscript{3} solution in DCM (1 ml, 1M) was added to the ice cold solution of compound 27 (450 mg, 1 mmol), resulting solution was allowed to be in normal rt and stirred the reaction mixture under inert atmosphere. Reaction was monitored using mass spec (MALDI-TOF), after completion of the reaction at about 48 h, water was added to the mixture and basified with NaHCO\textsubscript{3}, extracted with DCM, washed with brine and dried to obtain a pale yellow solid and used for next steps without purification. MALDI-TOF: m/z for C\textsubscript{26}H\textsubscript{34}N\textsubscript{6}O[M + H]\textsuperscript{+} is 447.8, ratio of the product was over 90% to starting material.

8-((4-((1-benzylpiperidin-4-yl)amino)-2-(4-methyl-1,4-diazepan-1-yl)quinazolin-7-yl)oxy)- N-hydroxyoctanamide (30): Procedure D using 28 as the starting material, afforded the desired intermediate as colorless solid. HRMS (ESI): m/z calcd for C\textsubscript{35}H\textsubscript{50}N\textsubscript{8}O\textsubscript{3}[M + H]\textsuperscript{+}, 603.4022; found, 603.4031. Subsequently, compound 29 was dissolved in 2 ml of MeOH and
treated with 50% NH$_2$OH/water mixture (1 ml) overnight to afford the targeted product. Reaction mixture was dried and purified using reverse column and further by HPLC using ACN/H$_2$O as eluent. Fractions collected were concentrated and lyophilized to get brown powder, 54 mg, 23% yield over two steps. $^1$H NMR (400 MHz, MeOD) $\delta$ 8.10 (d, $J = 9.2$ Hz, 1H), 7.65 – 7.53 (m, 4H), 7.06 (d, $J = 2.2$ Hz, 1H), 6.94 (d, $J = 8.9$ Hz, 1H), 3.94 (d, $J = 19.0$ Hz, 2H), 3.80 (s, 2H), 3.72 – 3.59 (m, 4H), 3.50 (s, 2H), 3.22 (s, 3H), 2.48 (s, 2H), 2.34 (d, $J = 14.2$ Hz, 3H), 2.15 (s, 2H), 2.05 (s, 4H), 1.96 (s, 5H), 1.86 (s, 2H), 1.68 (s, 2H), 1.46 (s, 4H). $^{13}$C NMR (100 MHz, MeOD) $\delta$ 175.5, 167.4, 164.9, 159.7, 153.4, 140.3, 129.1, 128.2, 127.0, 124.9, 114.1, 109.0, 103.6, 68.7, 63.7, 55.3, 55.1, 46.0, 43.5, 42.6, 41.3, 34.0, 28.5, 28.4, 25.4, 24.7. HRMS (ESI): m/z calcd for C$_{33}$H$_{47}$N$_7$O$_3$ [M + H]$^+$, 590.3819; found, 590.3832. HPLC purity 96.26%; $t_R$ = 14.192.

### 3.4.2 Molecular Docking Analysis

**Protein Preparation and Grid Generation:** The coordinates for the HDAC8/MS-344 complex (PDB ID: 1T67) and G9a/BIX-01294 complex (PDB ID: 3FPD) were downloaded from the RCSB Protein Data Bank. In these structures, MS-344 and BIX-01294 are bound to HDAC8, G9a respectively. The PDB protein-ligand structures were processed with the Protein Preparation Wizard in the Schrödinger suite. The protein structure integrity was checked and adjusted, and missing residues and loop segments near the active site were added using Prime. The receptor was prepared for docking by the addition of hydrogen atoms and the removal of co-crystallized molecules except for Zn$^{2+}$, as it is near to the active site in the case HDAC. Active site water molecules outside 5.0 Å from the ligand were removed. The bound ligands were used to specify the active site. A 3D box was generated around each ligand to enclose the entire vicinity of active site. The receptor grid for each target was prepared with the help of OPLS_2005 force field. The
grid center was set to be the centroid of the co-crystallized ligand, and the cubic grid had a size of 20 Å.

**Ligand Preparation:** the 2D ligand structures were prepared using ChemBioDraw Ultra 12.0, and the 3D structures were generated by Schrödinger suite. Schrödinger’s LigPrep program was used to generate different conformations of ligands. All possible protomers and ionization states were enumerated for 14 and bound ligands using Ionizer at a pH of 7.4. Tautomeric states were generated for chemical groups with possible prototropic tautomerism.

**Molecular Docking:** molecular docking studies were performed by using a GLIDE docking module of Schrödinger suite. It performs grid-based ligand docking with energetics and searches for positive interactions between ligand molecules and a typically larger receptor molecule, usually a protein. Finally, prepared ligands were docked into the generated receptor grids using Glide SP docking precision. The results were analyzed on the basis of the GLIDE docking score and molecular recognition interactions. All the 3D figures were obtained using Schrödinger Suite 2014-3.

### 3.4.3 Cloning, Protein Expression and Purification

Mouse histone methyltransferase G9a (969-1263) cDNA was amplified from the cDNA of BALB/c mouse thymus, and the fragment was sub-cloned into a vector with a 6His-sumo tag. The mouse G9a (mG9a) was expressed in *Escherichia coli* BL21 (DE3) by the addition of 1 mM isopropyl-1-thio-D-galactopyranoside (IPTG) and incubated overnight at 16°C.

The 6His-sumo mG9a (969-1263) protein was purified using the following procedure: harvested cell pellet was re-suspended in 20 mM Tris (pH 8.0), 500 mM NaCl, 0.1% β-mercaptoethanol, and 1 mM PMSF. Cells were lysed by sonicating for 15 s with 6 s intervals for a total time of 15 min on an ice bath. The supernatant of cell lysate was loaded onto a Ni⁺ affinity
column (Invitrogen) then washed with buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 20 mM imidazole, 0.1% β-mercaptoethanol, and 1 mM PMSF). The 6His-sumo tag was cleaved from the column by adding Ubiquitin-like-specific protease 1 (ULP-1) at 4 °C for 12 h. Wash buffer was then run through the Ni⁺ column again and the elution buffer collected. Subsequently, advanced protein purification was done by HiTrap Q HP sequential Superdex 200 10/300 GL. Elute of every step was analyzed by SDS PAGE, stained by Coomassie brilliant blue (CBB).

### 3.4.4 MALDI-TOF-MS

The *in vitro* inhibition of G9a by the synthesized compounds were measured by MALDI-TOF mass spectrum (Bruker MALDI TOF/TOF Analyzer). 400 nM purified G9a, 5 µM synthesized histone H3 (1-21) and 10 µM non-radioactive S-adenosyl methionine (Sigma) were added in reaction buffer (50 mM HEPES pH 8.0, 5 µg/ml BSA and 0.1% β-Mercaptoethanol) with or without inhibitors (5 µM). The reaction was incubated at room temperature for 30 min, and stopped by TFA. 1 µl of the sample was mixed with CHCA matrix and m/z peaks were obtained at reflection positive mode. The results of mass spectrum were analyzed using the Bruker flex analysis software.

### 3.4.5 Cell Based Assays

Cell lines information:

MDA-MB-231 (breast cancer cell line), MCF-7 (breast cancer cell line), A549 (human lung cancer cell line), K562 (human immortalized myelogenous leukemia cell line), Hela (human cervical cancer cell line), HEK293 (normal cell line).

Reagents: CCK-8, Trichostatin A and trypsin were purchased from Sigma.
Cell line: MDA-MB-231, A549 cell lines were grown at 37°C/5% CO₂ in Dulbecco’s Modified Eagle’s Medium (from Sigma) supplemented with 10% fetal bovine serum and 2% 200 mM L-glutamine and 0.5% antibiotic-antimycotic solution (from Sigma).

Hela, K562 cell lines were grown at 37°C/5% CO₂ in RPMI 1640 medium (Gibco) supplemented with 10% fetal bovine serum and 0.5% antibiotic-antimycotic solution.

MCF-7 cell line was grown at 37°C/5% CO₂ in Eagle’s Minimum Essential Medium supplemented with 10% fetal bovine serum and 0.5% antibiotic-antimycotic solution.

3.4.6 HDAC Activity Assay

The manual assay was developed by Thomas’s group. HeLa cells were seeded into white 96-well cell culture plates (corning costar 3596) at a density of 8000-10000 cells/well (total volume 81 µl culture medium) and incubated under standard cell culture conditions(37°C, 5% CO₂). After 24 h, 9 µl inhibitors with different concentration were added to the HeLa cells and incubation was continued for 3 h under cell culture conditions. After this treatment period, 10 µl of a 2 mM stock solution of the substrate Boc-K(Ac)-AMC was added into the 96 well plates with HeLa cells and inhibitors. Cell culture plates were incubated under standard cell culture conditions for an additional 3 h before addition of 100 µl/well lysis/developer buffer mix (50 mMTris–HCl, pH 8.0, 137 mM NaCl, 2.7 mM KCl, 1 mM MgCl₂, 1 vol% Nonidet-P40, 2.0 mg/ml trypsin, 10 µM TSA). After final incubation for 3 h under cell culture conditions, fluorescence was measured at excitation of λex = 355 nm and emission of λem = 460 nm on the Perkin–Elmer Wallac Victor V 1420 multilabel plate reader (Perkin–Elmer, Wellesley, USA). A549 and K562 cell lines used the same method, respectively. IC₅₀s were calculated using GraphPad Prizm statistical package with sigmoidal variable slope dose response curve fit.
3.4.7 **G9a H3K9me2 Cellular Assay**

Cells were seeded at 8000-10000 cells (100 µl) in black-walled 96-well plates (Thermo 165305) and exposed to various inhibitor concentrations for 48 h. After the incubation, the media was removed and 100 µl fixation and permeabilization solution (2% formaldehyde in PBS) for fixation was added for 30 min. And then use 200 µl 0.1% Triton X100 in PBS washing solution to wash (allow wash to shake on a plate shaker for 5 min). After five washes, cells were blocked for 1 h with 150 µl blocking buffer to each well (1%BSA in PBS) (allow blocking at rt with moderate shaking on a plate shaker). After 1 h, remove the blocking buffer from the blocking step and add primary antibody in blocking buffer to cover the bottom of each well. (Three out of four replicates were exposed to the primary H3K9me2 antibody, Abcam no. 1220 at 1/500 dilution in 1% BSA, PBS for overnight, one replicate was reserved for the background control (only blocking buffer). The wells were washed five times with 0.1% Tween 20 in PBS, then secondary IR800 conjugated antibody (LiCor) and cell tag 700 stain added for 1 h. (incubate for 1 h with gentle shaking at rt, protect plate from light during incubation). After 5 wash with 0.1% Tween 20 in PBS, remove wash solution completely from wells. Turn the plate upside down and tap or blot gently on paper towels to remove traces of wash buffer. The plates were read on an Odyssey CL_x (LiCor) scanner at both 800 nm (H3K9me2 signal) and 700 nm (cell tag 700 stain signal) channels. IC_{50}s were calculated using GraphPad Prizm statistical package with sigmoidal variable slope dose response curve fit.

3.4.8 **Toxicity Assay**

A549, MDA-MB-231, MCF-7 and HEK293 cells were seeded at 8000-10000 cells (100 µl) in white 96-well plates and pre-incubate the plate for 24 h under standard cell culture conditions, respectively. And then the cells were exposed to the different inhibitors with various
concentrations for 72 h. Finally, 10 µl of CCK-8 kit solution was added to each well and incubated for 3-4 hours under standard cell culture conditions, and the 96 well plates were measured the absorbance at 450 nm using Perkin–Elmer Wallac Victor 3V 1420 multi label plate reader (Perkin–Elmer, Wellesley, USA). EC50s were calculated using GraphPad Prizm statistical package with sigmoidal variable slope dose response curve fit.
APPENDICES: $^1$H AND $^{13}$C SPECTRA OF COMPOUNDS

$^1$H NMR of Compound 1-1

$^{13}$C NMR of Compound 1-1
$^1$H NMR of Compound 1-5

$^{13}$C NMR of Compound 1-5
$^1$H NMR of Compound 1-6

$^{13}$C NMR of Compound 1-6
$^1$H NMR of Compound 1-7

$^{13}$C NMR of Compound 1-7
$^1$H NMR of Compound 1-8 (M100)

$^{13}$C NMR of Compound 1-8 (M100)
$^1$H NMR of Compound 1-9

$^{13}$C NMR of Compound 1-9
$^1$H NMR of Compound 1-10

$^{13}$C NMR of Compound 1-10
$^1$H NMR of Compound 1-11

$^{13}$C NMR of Compound 1-11
$^1$H NMR of Compound 1-12

$^{13}$C NMR of Compound 1-12
$^1$H NMR of Compound 1-13 (M301)

$^{13}$C NMR of Compound 1-13 (M301)
$^1$H NMR of Compound 1-14

$^{13}$C NMR of Compound 1-14
$^1$H NMR of Compound 1-15

$^{13}$C NMR of Compound 1-15
$^1$H NMR of Compound 1-16

$^{13}$C NMR of Compound 1-16
$^1$H NMR of Compound 1-17

$^{13}$C NMR of Compound 1-17
$^1$H NMR of Compound 1-18

$^{13}$C NMR of Compound 1-18
$^1$H NMR of Compound 1-19 (M201)

$^{13}$C NMR of Compound 1-19 (M201)
$^1$H NMR of M101

$^1$H NMR of M102
$^1$H NMR of M103

$^1$H NMR of M104
$^1$H NMR of M105

$^1$H NMR of M106
$^1$H NMR of M000

$^1$H NMR of M202
$^1$H NMR of M203

$^1$H NMR of M204
$^1$H NMR of M205

$^1$H NMR of M010
$^1$H NMR of M212

$^1$H NMR of M213
$^1$H NMR of M214

$^1$H NMR of M215
$^1$H NMR of M020

$^1$H NMR of M223
$^1$H NMR of M224

$^1$H NMR of M225
$^1$H NMR of M030

$^1$H NMR of M234
$^1$H NMR of M235

$^1$H NMR of M040
$^1$H NMR of M245

$^1$H NMR of M050
$^1$H NMR of M302

$^1$H NMR of M303
$^1$H NMR of M304

$^1$H NMR of M305
$^1$H NMR of M312

$^1$H NMR of M313
$^1$H NMR of M314

$^1$H NMR of M315
$^1$H NMR of M323

$^1$H NMR of M324
$^1$H NMR of M325

$^1$H NMR of M334
$^1$H NMR of M335

$^1$H NMR of M345
$^1$H NMR and $^{13}$C NMR of Compound 2-2
$^1$H NMR and $^{13}$C NMR of Compound 2-3
$^1$H NMR and $^{13}$C NMR of Compound 2-4
$^1\text{H}$ NMR and $^{13}\text{C}$ NMR of Compound 2-5
$^1$H NMR and $^{13}$C NMR of Compound 2-6
$^1$H NMR and $^{13}$C NMR of Compound 2-7
$^{1}H$ NMR and $^{13}C$ NMR of Compound 2-8
$^1$H NMR and $^{13}$C NMR of Compound 2-9
$^{1}$H NMR and $^{13}$C NMR of Compound 2-10
$^1$H NMR and $^{13}$C NMR of Compound 2-11
$^1$H NMR and $^{13}$C NMR of Compound 2-12
$^1$H NMR and $^{13}$C NMR of Compound 2-13
\(^{1}\text{H NMR and }^{13}\text{C NMR of Compound 2-14}\)
$^1$H NMR and $^{13}$C NMR of Compound 2-15
$^1$H NMR and $^{13}$C NMR of Compound 2-16
$^1$H NMR and $^{13}$C NMR of Compound 2-17
$^1$H NMR and $^{13}$C NMR of Compound 2-18
$^1$H NMR and $^{13}$C NMR of Compound 2-19
$^{1}$H NMR and $^{13}$C NMR of Compound 2-20
$^1$H NMR and $^{13}$C NMR of Compound 2-21
$^1$H NMR and $^{13}$C NMR of Compound 2-22
$^1$H NMR and $^{13}$C NMR of Compound 2-23
$^1$H NMR and $^{13}$C NMR of Compound 2-24
$^1$H NMR and $^{13}$C NMR of Compound 2-25
$^1$H NMR and $^{13}$C NMR of Compound 2-26
$^1$H NMR and $^{13}$C NMR of Compound 2-27$^{[1b][1b][1b][1b][1b][1b][1b][1b][1b][1b][1b][1b][1b][1b][1b][1b][1b][1b]}$
REFERENCES


