4-30-2018

SMALL MOLECULE INHIBITORS OF G9a AND HDAC FOR TARGETED CANCER THERAPY AND DNA ENCODED GLYCAN LIBRARY (DEGL) FOR EARLY DETECTION OF CANCER

Shukkoor Muhammed kondengaden
Georgia State University

SMALL MOLECULE INHIBITORS OF G9a AND HDAC FOR TARGETED CANCER THERAPY AND DNA ENCODED GLYCAN LIBRARY (DEGL) FOR EARLY DETECTION OF CANCER

by

MUHAMMED SHUKKOOR KONDENGADEN

Under the Direction of Peng George Wang, PhD

ABSTRACT

This dissertation focuses on two aspects, targeted cancer therapy based on epigenetics and cancer diagnostics using DNA encoded glycan libraries. The first part addresses cancer therapy, wherein we discuss the discovery of a new class of G9a inhibitors and development of dual inhibitors targeting epigenetic enzymes G9a and Histone Deacetylases (HDAC). Aberrant enzymatic activities or expression profiles of epigenetic regulations are therapeutic targets for cancers. Among these, histone 3 lysine 9 methylation (H3K9Me2) marked by methyltransferase enzyme G9a and global de-acetylation on histone proteins by deacetylation enzyme HDAC typically associated with multiple cancer phenotypes including leukemia, prostatic carcinoma, hepatocellular carcinoma and pulmonary carcinoma. Our structure-based virtual screening

identified a series of compounds as potential inhibitors of G9a. Among these new class of G9a inhibitors, DCG066 was confirmed by *in vitro* biochemical, and cell-based enzyme assays. In another effort, several compounds featuring both G9a and HDAC pharmacophore was synthesized and screened for their dual activity and presented their detailed structure-activity relationship studies.

Secondly, this dissertation elaborates a new method for the detection of cancer. Tumor-associated antigens like Globo H, are found in many cancers including breast cancer and ovarian cancer. Applying principles of DNA encoding and screening, we developed a DNA encoded glycan library (DEGL) of these tumor specific glycans. Then, DEGL is used to detect cancer by screening the serum samples for anti-Globo H antibodies.

INDEX WORDS: Epigenetics, Lysine methyltransferase; G9a inhibitor, HDAC inhibitors, Dual inhibitors, MALDI-TOF assay, DNA encoded glycan library, Glyco-PCR, Carbohydrate oligonucleotide conjugates.

SMALL MOLECULE INHIBITORS OF G9a AND HDAC FOR TARGETED CANCER THERAPY AND DNA ENCODED GLYCAN LIBRARY (DEGL) FOR EARLY DETECTION OF CANCER

by

MUHAMMED SHUKKOOR KONDENGADEN

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2018

SMALL MOLECULE INHIBITORS OF G9a AND HDAC FOR TARGETED CANCER THERAPY AND DNA ENCODED GLYCAN LIBRARY (DEGL) FOR EARLY DETECTION OF CANCER

by

MUHAMMED SHUKKOOR KONDENGADEN

Committee Chair: Peng George Wang

Committee: Maged Henary

Gangli Wang

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

May 2018

DEDICATION

I dedicate this work to my parents *Sh. Muhammed K and Smt. Fathima C* and all my teachers, without your help and guidance I could not be even closer to achieving this milestone. I also dedicate this work to my wife Akhila, who came long way to share the excitement and stress of each steps in this journey. My friends Dr. Pankaj Kumar and Harshad Kamble, both of you deserve a piece of this cake too, I know we started this journey way back from our days back in NIPER.

ACKNOWLEDGEMENTS

I sincerely acknowledge my Ph.D. advisor *Prof. Peng George Wang* for giving me this wonderful opportunity, his energy and dedication to science is matchless and inspired me throughout the years. He trusted me and gave me the freedom to do what I liked to do, and later gave me the exact direction and guidance whenever I needed them most. I also indebted to his honest remarks which have molded me into a better researcher and person. I thank him from my deep heart for all the good things he has done for me. My PhD committee members Dr. Maged Henary and Dr. Gangly Wang also helped to shape this work. Advises during the annual meeting and oral exam were remarkable and helped me to the smooth run to finish line. I also wish to acknowledge *Dr. Kathy Li, Dr. Abasaheb, Dr. Bharat Gurale, Dr. Yunpeng Liu, Dr. Zhongying Xiao and Zaikuan Yu* who all helped me to settle in the lab. Finally, our team, *Qing Zang, Lanlan Zang, Liuqing Wen, Aishwarya Parameswaran, Huajie Zhang, Kenneth Huang, Shanshan Li, Hailiang Zhu and Xu Li*, who significantly contributed to the successful completion of this work, and all other lab members who helped me and supported me throughout the journey. I always enjoyed my stay at the group and I sincerely thank all the Wang's group members for those beautiful moments.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Chronic myelogenous leukemia (CML)

Histone lysine methyltransferases (KMTs)

Histone acetyltransferases (HATs)

DNA methyltransferases (DNMTs)

Methyl-CpG binding domain (MBD)

Histone demethylases (KDMs),

Histone deacetylases (HDACs)

Ten-eleven translocation (TET)

Plant homeodomain (PHD)

Histone 3 lysine 9 methylation (H3K9Me2)

Prostate Specific Antigen (PSA)

Tumor associated glygans (TAG)

Colorectal cancer (CRC)

DNA encoded glycan libraries (DEGL).

Acute myeloid leukemia (AML)

H4 lysine 20 (H4K20),

H3 lysine 9 (H3K9),

H3 lysine 27 (H3K27)

H3 lysine 4 (H3K4)

H3 lysine 36 (H3K36)

H3 lysine 79 (H3K79)

GLP (G9a like protein)

Inducible pluripotent stem cells (iPSCs)

S-adenosyl-methionine (SAM)

Leukemia stem cells (LSCs)

Structure-activity relationship (SAR)

Structure-based virtual screenings (SBVS)

Root mean-square deviation (RMSD)

Surface plasmon resonance (SPR)

Gene Expression across Normal and Tumor tissues (GENT)

Hematopoietic stem cells (HSCs)

Isopropyl-1-thio-D-galactopyranoside (IPTG)

Fetal bovine serum (FBS)

Dulbecco's modified Eagle's medium (DMEM)

Quantitative real-time PCR (qRT-PCR)

7-amino-actinomycin (7AAD)

Protein lysine methyltransferases enzymes (PKMT),

Protein arginine methyltransferases (PRMT)

Histone 3 lysine 9 (H3K9)

In-Cell Western (ICW)

7-amino-4-methylcoumarin (AMC)

High resolution mass spectrum (HRMS)

Multiplex glycan bead array (MGBA)

DNA encoded library (DEL)

Glycan-binding proteins (GBP)

Next generation Sequencing (NGS)

Consortium of Functional Glycomics (CFG)

N,N-Diisopropylethylamine (DIPEA)

1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC)

Hydroxybenzotriazole (HOBt)

1 INTRODUCTION

1.1 Cancer therapy

Cancer is one of the leading causes of death across globe and is the second leading cause of death in the United States.[1] More than 14 million new cases and 8.2 million deaths reported were related to cancer in 2012. Cancer cases are projected to rise over 22 million by next two decades.[2] Even though with these continued increase in reported cancer cases, death rate related to the cancer is recently in decline largely contributed by the advancement in cancer therapy. Traditionally cancer is being treated by chemotherapies, irradiation and surgery and more recently by targeted therapies and cancer immunotherapies. Traditional chemotherapy act on rapidly dividing cancer cells, but it also affects rapidly dividing normal tissues like hair, gastrointestinal epithelium and bone marrows. Unlike this, targeted therapies aim a specific pathway in cancer proliferation that is either exclusively used by the cancer tissues or overexpressed compared to the normal cells.[3, 4]

Targeted therapies, which refers to monoclonal antibodies and small molecule inhibitors are the major class of cancer therapy today. It span to all different sort of malignancies and a major drug component in many of the cancer treatments including breast, lung, pancreatic, and colorectal cancers, as well as lymphoma, leukemia, and multiple myeloma.[3] In this dissertation, we focus on small molecules based therapies for cancer. Small molecules have several advantages over the monoclonal antibodies, they are administered orally, and their chemical synthesis is much cheaper than the bioengineered antibodies. Small molecule therapies rose to prominence after the discovery of the Imatinib, a tyrosin kinase inhibitor with its tremendous success in treating chronic myelogenous leukemia (CML). The success of Imatinib is attributed to the single causal mutation

of the CML and a similar success was not seen in other cancers because they are often associated with multi factorial origin.[5]

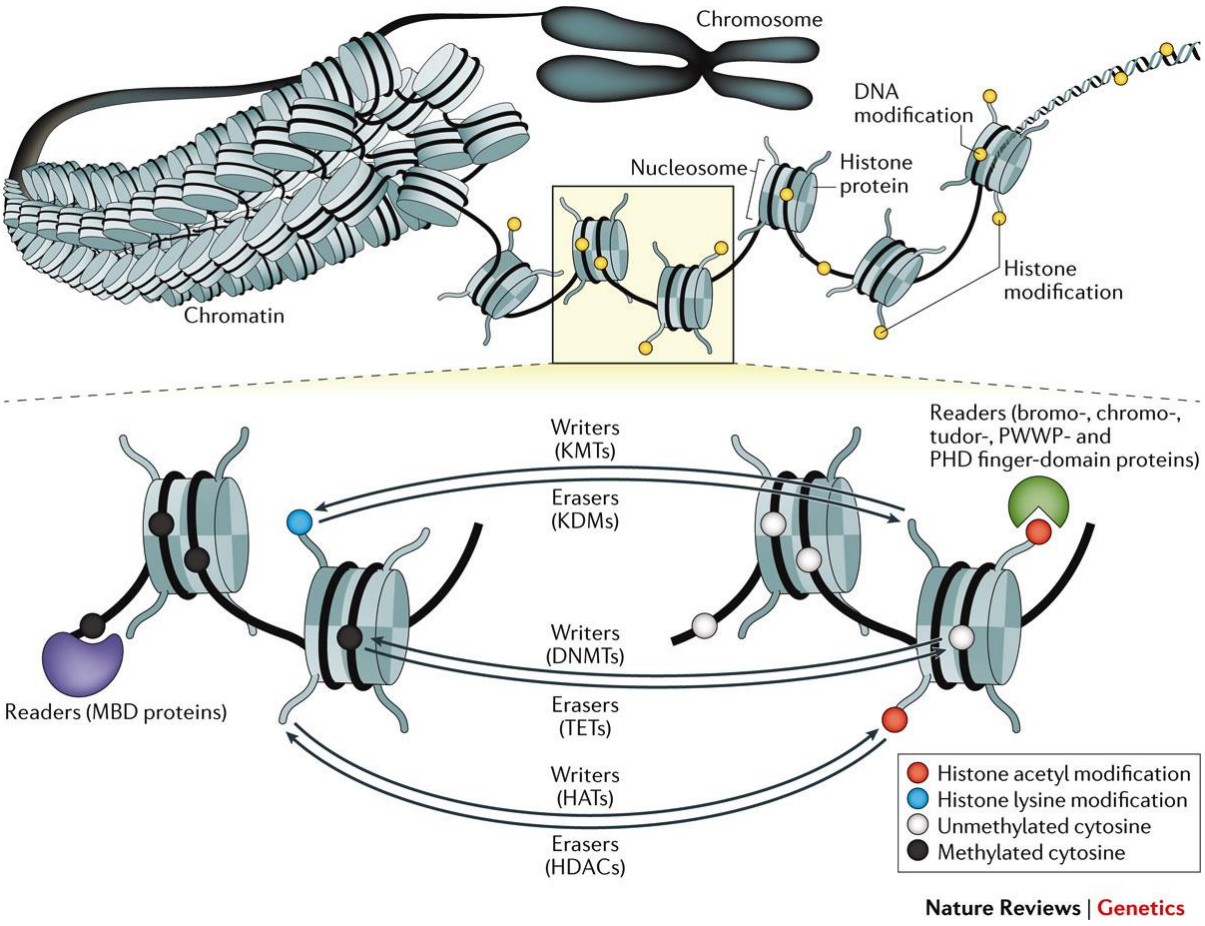


Figure 1.1. Modulation of covalent modifications on chromatin.
A 147 bp sequence of DNA is wrapped around a core of eight histone proteins to compact the genome into nucleosomes and then into chromatin and chromosomes. A subset of covalent modifications (yellow circles) on both the histone and DNA components, which control accessibility of DNA to transcription factors and other regulators, are shown. Covalent marks are established by 'writers', such as histone lysine methyltransferases (KMTs), histone acetyltransferases (HATs) and DNA methyltransferases (DNMTs). These modifications are interpreted by 'readers', including methyl-CpG binding domain (MBD) proteins on the DNA and multiple proteins for the histone marks as shown. Progress over the past decade has shown that almost all of the marks can be removed by 'erasers', such as histone demethylases (KDMs), histone deacetylases (HDACs) and by the ten-eleven translocation (TET) family of 5-methylcytosine oxidases. The interplay between these enzymes helps to establish and maintain cellular identity in addition to the central role of transcription factors by regulating access to the DNA sequence. PHD, plant homeodomain.[6]

More recently epigenetics has taken center stage in cancer therapy. Epigenetics refers study of heritable changes that regulate gene expression but do not involve changes to the underlying DNA sequence. Epigenetic changes governed by the covalent modifications occurred to the chromatin components, which includes DNA, RNA and histone proteins (Figure 1.1). The most common epigenetic modifications also called as post translational modifications includes DNA methylation, histone methylation, acetylation, phosphorylation, glycosylation and ubiquitylation.[7] Conversely to genetic alterations, epigenetic modifications are reversible and controlled by three sets of enzymes called 'Erasers' who mark the modification, 'Reader' who read the modifications and respond with either gene activation or deactivation, and 'Erasers' who reverse the modifications (Figure 1.1).[6] The entire process is significantly affected by the external factors like environmental exposure to carcinogens, life style, smoking and widely altered in cancer. Tumor cells use epigenetic process to escape from the immune surveillance and chemotherapy. Hence, these modifications and their corresponding enzymes are fiercely pursued as drug discovery targets. Most recent status of epigenetic drug development is listed in Table 1, drugs targeting DNA methylation writers and histone acetylation erasers are already approved for therapy and several more drugs are currently under clinical trial.[8]

In this study, we targeted two epigenetic enzymes G9a and HDAC. Histone 3 lysine 9 methylation (H3K9Me2) and global de-acetylation at histone proteins are associated with multiple cancer phenotypes including leukemia, prostate carcinoma, hepatocellular carcinoma and lung cancer.[9-11] Cancer is a disease with difficult treatment options due to the multifactorial basis of initiation, and progression; a treatment targeting multiple components instead of a single component would therefore be of interest in cancer therapeutics.[12-14] G9a and HDAC are proven targets for cancer and several groups published many inhibitors for each class with

crystallography data suggesting the specific binding and related pharmacological action.[11, 15, 16] Our Structure based design, synthesis, and activity studies discovered the first small molecule as a hybrid targeting HDAC and G9a. In another work, we discuss the structure based virtual screening of a library of 200,000 small molecules for the discovery of G9a specific inhibitors. This two works are elaborated in the first half of the dissertation.

Table 1.1. *Epigenetics drugs in clinical trials*

Compound	Target	Status
Decitabine	DNMT1	FDA approved
Azacitidine	DNMT1	FDA approved
Guadecitabine	DNMT1	Phase III
4-Thio-2-deoxycytidine	DNMT1	Phase I
ASTX727	DNMT1	Phase II
AG-120	IDH1	Phase III
AG-221	IDH2	Phase III
AG-881	IDH1 and IDH2	Phase I
IDH305	IDH1	Phase II
FT-2102	IDH1	Phase I
BAY1436032	IDH1	Phase I
IDH1 peptide vaccine	IDH1	Phase I
Belinostat	HDACs	FDA approved
Vorinostat	HDACs	FDA approved
Romidepsin	HDACs	FDA approved
Panobinostat	HDACs	FDA approved
Valproic acid	HDACs	FDA approved
KA2507	HDAC6	Phase I
ACY-241	HDACs	Phase I
Tucidinostat	HDACs	Phase III
CKD-581	HDACs	Phase I
CUDC-907	HDACs	Phase I
CXD101	HDACs	Phase I
EDO-S101	HDACs	Phase I
Givinostat	HDACs	Phase II
MPT0E028	HDACs	Phase I
OBP-801	HDACs	Phase I
Resminostat	HDACs	Phase II
Tefinostat	HDACs	Phase II
AR-42	HDACs	Phase I
Entinostat	HDACs	Phase III
Mocetinostat	HDACs	Phase II

Ricolinostat	HDACs	Phase II
ABBV-075	BETs	Phase I
BMS-986158	BETs	Phase II
GSK2820151	BETs	Phase I
INCB054329	BETs	Phase II
INCB057643	BETs	Phase II
TEN-010	BETs	Phase I
ZEN003694	BETs	Phase I
OTX015	BETs	Phase I
CPI-0610	BETs	Phase II
FT-1101	BETs	Phase I
GSK525762	BETs	Phase II
MAK683	EED	Phase II
Tazemetostat	EZH2	Phase II
CPI-1205	EZH2	Phase I
DS-3201	EZH2, EZH1	Phase I
GSK2816126	EZH2	Phase I
Pinometostat	DOT1L	Phase I
Tranylcypromine	KDM1A	FDA approved
GSK2879552	KDM1A	Phase II
INCB059872	KDM1A	Phase II

1.2. Cancer diagnosis

Early detection of cancer is one of the most significant part in the success of cancer therapies. Cancer diagnosis has advanced tremendously in recent years, yet they rely predominantly on invasive (*i.e.* random biopsies and surgery) methods, and the noninvasive methods available are limited to certain cancer types. More often noninvasive methods are not sensitive enough to distinguish aggressive tumors from indolent ones like in the case of Prostate Specific Antigen (PSA) test in prostate cancer diagnosis.[17] Hence, new methods for the early detection of cancer with high sensitivity, specificity and throughput are equally essential with the advancement in cancer therapeutics.

In past two decades, cancer biology specifically cancer glycobiology has moved substantially. Understanding the role of glycans in cancer related mechanism like cell-cell

recognition, cell adhesion, metastasis formation and immune modulation etc. greatly facilitated therapeutic intervention of various types of cancers.[18] This, understanding how the glycan structure are modified in the tumor cells and how they behave to the tumor microenvironment is of critical importance. The alterations of glycosylation or glycan presentation on tumor cells are cell specific, protein specific and site specific, hence considered as tumor biomarkers.[18-22]

Several alloantibodies are generated against the tumor associated glygans (TAG) and detection of this antibodies can correlate with the cancer. Patient sera has the TAG specific antibodies way before the aggressive cancer is diagnosed, hence the early detection is possible by the screening of patient sera for the cancer specific antibodies. Aberrant glycosylation of MUC-1 is observed in colorectal cancer and detection of autoantibodies specific to this glycan epitope has predicted the CRC with more than 95% specificity.[23] Another cancer associated carbohydrate antigen globo-H, which is an hexasaccharide present in cell surface as a glycolipid is a proven biomarker, and screening of antibodies against this glycan is found to be effective in diagnosing breast cancers, ovarian cancer, and thyroid cancer.[24-26] In the second part of the dissertation, we describe a novel method for the ultra-sensitive detection of glycan specific antibodies using DNA encoded glycan libraries (DEGL).

2 SMALL MOLECULE INHIBITORS OF G9A AND HDAC FOR TARGETED CANCER THERAPY

2.1 Discovery of novel small molecule inhibitors of lysine methyltransferase G9a and their mechanism in leukemia cell lines

The work presented in this chapter is reproduced with some modification as to fit the format with permission from the journal European journal of medicinal chemistry. My contributions as first author were the design, chemical synthesis and figure/manuscript preparation.

Kondengaden, Shukkoor M., Liu-fei Luo, Kenneth Huang, Mengyuan Zhu, Lanlan Zang, Eudoxie Bataba, Runling Wang et al. "Discovery of novel small molecule inhibitors of lysine methyltransferase G9a and their mechanism in leukemia cell lines." *European journal of medicinal chemistry* 122 (2016): 382-393.

ABSTRACT: Lysine methyltransferase G9a regulates the transcription of multiple genes by primarily catalyzing mono- and di-methylation of histone H3 lysine 9, as well as several non-histone lysine sites. An attractive therapeutic target in treating leukemia, knockout studies of G9a in mice have found dramatically slowed proliferation and self-renewal of acute myeloid leukemia (AML) cells due to the attenuation of HoxA9-dependent transcription. In this study, a series of compounds were identified as potential inhibitors through structure-based virtual screening. Among these compounds, a new G9a inhibitor, DCG066, was confirmed by *in vitro* biochemical, and cell based enzyme assays. DCG066 has a novel molecular scaffold unlike other G9a inhibitors presently available. Like G9a's histone substrate, DCG066 can bind directly to G9a and inhibit methyltransferase activity *in vitro*. In addition to suppressing G9a methyltransferase activity and reducing histone H3 methylation levels, DCG066 displays low cytotoxicity in leukemia cell lines with elevated levels of G9a expression, including K562. This work presents DCG066 as an

inhibitor of G9a with a novel structure, providing both a lead in G9a inhibitor design and a means for probing the functionality of G9a.

2.1.1 Introduction

Epigenetic regulation plays a vital role in gene transcription and expression, occurring not through alterations in the genetic code, but rather by modifications to histones or other vital proteins. Among the regulatory processes utilized in posttranslational modifications, histone modification is crucial.[27] Comprised of at least eight variations, including acetylation, methylation, phosphorylation, sumoylation and so forth, the causal relation between histone methylation in gene expression and histone lysine residues is unusual.[28] Frequently, the methylation of H4 lysine 20 (H4K20), H3 lysine 9 (H3K9), and H3 lysine 27 (H3K27) are associated with transcription suppression, whereas methylation of H3 lysine 4 (H3K4), H3 lysine 36 (H3K36) and H3 lysine 79 (H3K79) tends lead to transcription activation.[28-34] Of interest, H3K9-methylation correlates with transcription suppression, gene silencing and genomic stability. H3K9-methylation is regulated by histone lysine *N*-methyltransferase EHMT2 (G9a) and analogous protein GLP (G9a like protein). G9a is responsible for catalyzing the mono- and di-methylation of H3K9 in euchromatic regions. G9a is also capable of methylating H3K27, lysine 373 of p53 and other gene expression regulators. However in several cancers, particularly chronic myeloid leukemia, G9a and homologous protein GLP are found to be overexpressed with the proliferation of cancer cells suppressed through knocking out G9a.[35, 36] Because G9a and GLP also play a significant role in gametogenesis and embryonic development,[37-40] G9a/GLP inhibitors have been used with success in cell reprograming studies and in producing inducible pluripotent stem cells (iPSCs) by various groups.[41-44] As displayed in Table 1, current G9a inhibitors can be classified into three overall categories: type I with BIX-01294 and its

derivatives,[45, 46] type II with BIX-0138 and its derivatives,[47] and type III with fungal metabolite chaetocin.[48, 49] Reported in 2007, BIX-01294 was the first G9a selective inhibitor that did not target the binding pocket of cofactor *S*-adenosyl-methionine (SAM).[50] Since the discovery of BIX-01294, many G9a inhibitors including UNC0224, UNC0638, UNC0642, E72 and A-366 have emerged based on the pharmacophore of BIX-01294 (Figure 2.1).[45, 47, 51-54] All of the current G9a inhibitors except chaetocin possess the same quinazoline core and display similar biological activity, performing well in enzyme and cellular assays.

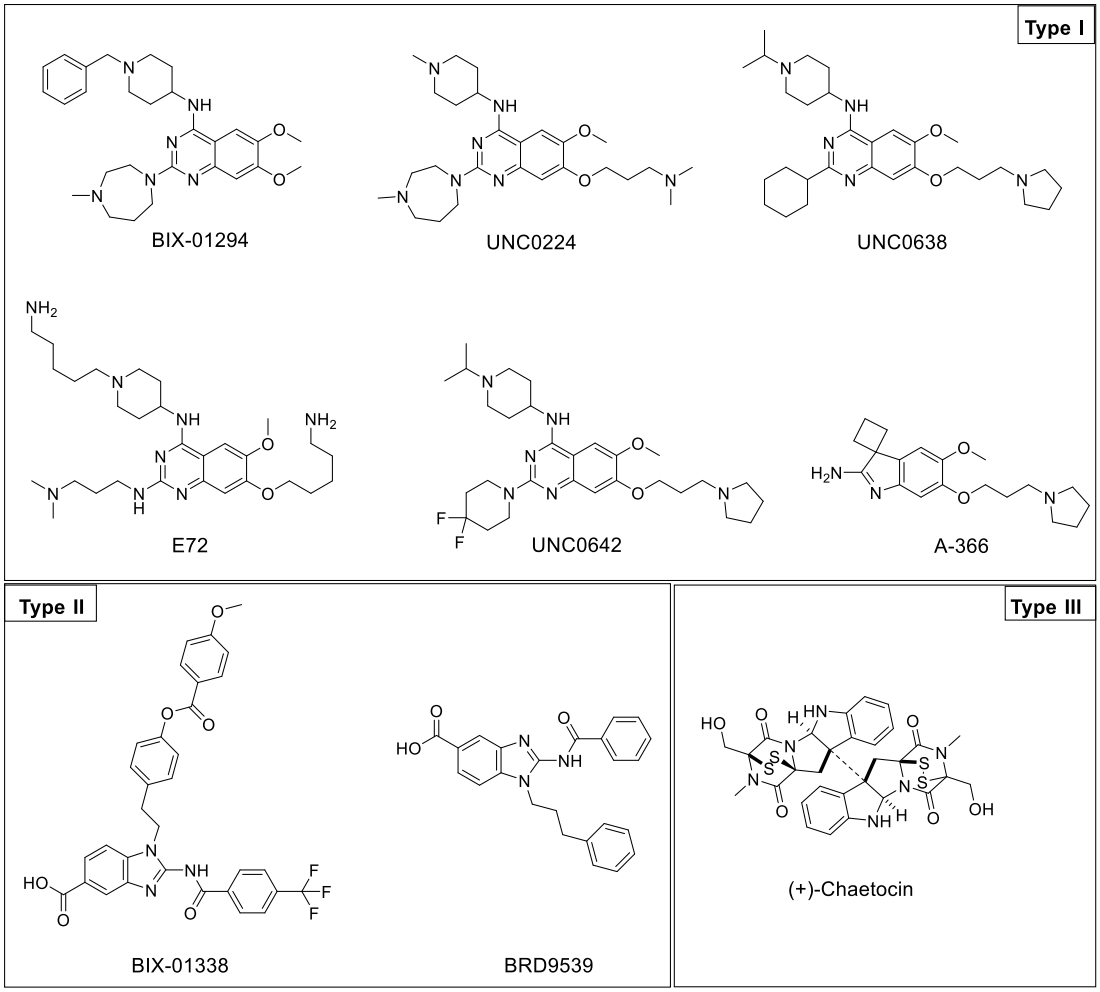


Figure 2.1. Structures of known G9a inhibitors.

Of particular interest is the inhibitors belonging to the type I and III categories, as clinical trials have indicated their high efficacy in inhibiting the proliferation of leukemia cells. By attenuating HoxA9-dependent transcription, G9a inhibitors are able to reduce the self-renewal and proliferation of acute myeloid leukemia (AML) cells.[46] Previous studies have indicated that mutations involved in DNA methylation could be responsible for mediating abnormal self-renewal and differentiation in leukemia stem cells (LSCs) and, discovery of targeted drugs for epigenetic enzymes could provide a new route for leukemia treatment.[55, 56] G9a has emerged as an attractive drug target for leukemia. However, current inhibitors of G9a are derivatives of parent compound BIX-01294 generated through manual or structure-activity relationship (SAR) analysis, offering limited design space for new inhibitors. Thus, there is an interest in searching for new scaffolds capable of inhibiting its activity. This study aims to search for new scaffolds for G9a inhibition through a combination of virtual and enzymatic screening.

High-throughput and structure-based virtual screenings (SBVS) are widely utilized methods in the discovery of lead structures. Through SBVS, DCG066 was identified as an inhibitor with binding affinity comparable to that of the native substrates. *In vitro* biochemical and cell based assays verified that DCG066 had a level of activity comparable to that of BIX-01294, in addition to induction of cell apoptosis. In studying inhibitors for G9a, the expression levels of G9a in varying cancer cell lines were also examined to determine which cell line would be most appropriate for *in vitro* assays. From these observations, a trend emerged; the expression levels of G9a in most hematologic cancer cells were much higher than in normal blood cells.

Since the overexpression of G9a is related to several cancers,[57] we analyzed the G9a mRNA and protein level using quantitative real-time PCR and western blotting, respectively, in order to quantify expression levels across different cell lines. K562, a leukemia cell line, was found to have

G9a expression levels higher than all solid tumor cells, and thus was used in further *in vitro* cell assays. DCG066 decreases di-methylation levels of histone H3 lysine 9 (H3K9Me2), inhibits cell proliferation and induces cell apoptosis. Displaying comparable activity to BIX-01294 and lower cytotoxicity in leukemia cells, DCG066 offers possibilities as both a probe for the functions of G9a and as a novel lead in G9a inhibitor design.

2.1.2 Results and Discussion

2.1.2.1 Virtual screening

A schematic representation of the overall procedure is presented in Figure 2.2. The crystal structure of G9a in complex with UNC0638 (PDB entry: 3RJW) was selected to be the binding pocket in docking studies. Chain A of the protein and the corresponding UNC0683 were both extracted from 3RJW to serve as the template for the binding pocket model. In generating the library of compounds to screen, we searched for a public database containing enormous number of small molecule compounds that would be available for subsequent *in vitro* screening. The SPECS database (http://www.specs.net/), containing 200,000 small molecule compounds, was found to meet the criteria. The database was then filtered for compounds with log P value smaller than 5 by Pipeline Pilot 7.5, leaving about 90,000 compounds that would likely be soluble in aqueous solution to a certain degree. In evaluating which of the two available docking methods (Glide 5.5, Gold 5.0) were to be used, two criteria were employed: the root mean-square deviation (RMSD) value and the enrichment factor. The RMSD value of UNC0683 between the best-ranked pose by the docking methods and the initial pose would provide a measure for the accuracy of the docking algorithm in comparison with experimentally determined results. The enrichment factor would measure the fraction of active compounds found in a certain percentage divided by the fraction of the screened library. Between these two criteria, Glide 5.5 was found to be the best and

was used in docking studies. The Glide XP mode with default settings was used, following standard procedure. The hit compounds were ranked based on G-score and the top 1000 compounds were selected for scaffold diversity analysis by Pipeline Pilot 7.5 and visual inspection. Afterwards, candidate molecules were chosen based on the following criteria: (1) molecular weight of the ligand is lower than 600 Da, which would benefit from further structural optimization, (2) both geometric and chemical features correspond with the active site of G9a, and (3) the binding poses and chemical structures were reasonable. From this, 125 candidate molecules were identified for purchase and further G9a inhibition activity assessment.

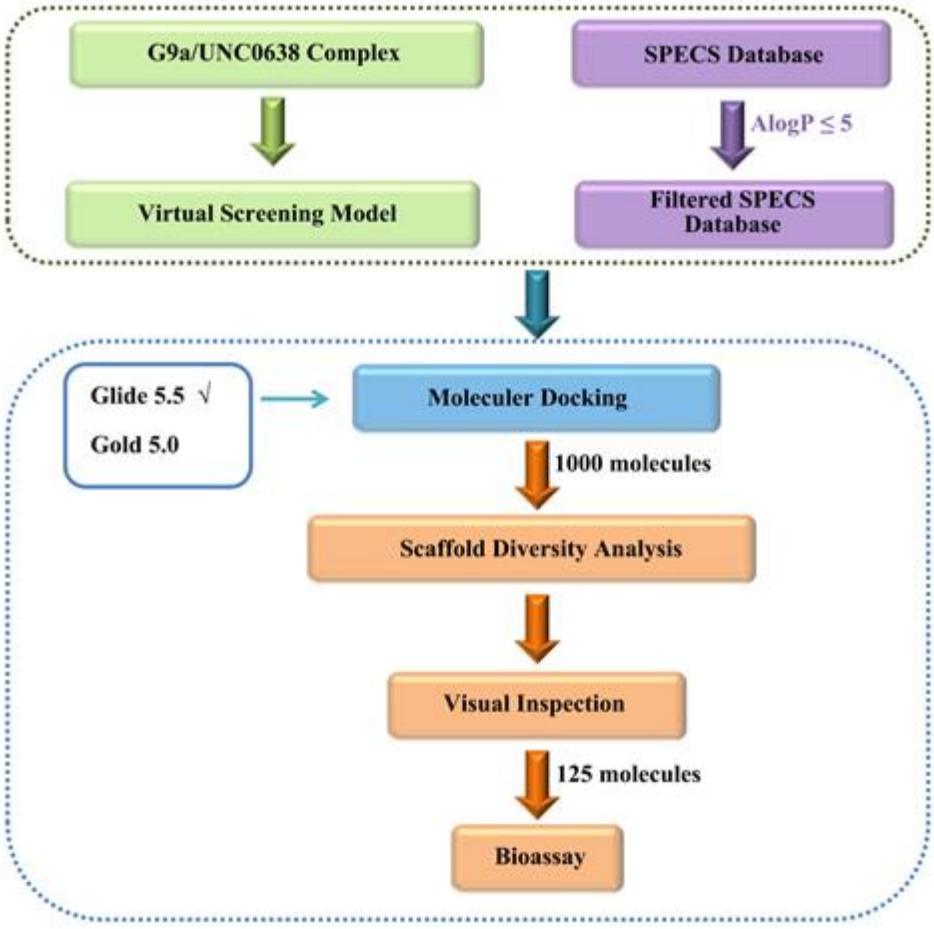


Figure 2.2. Flow chart for the virtual screening

2.1.2.2 In vitro screening for direct binding and enzymatic inhibition

To verify whether the 125 selected compounds could inhibit G9a activity, two different approaches were used to determine the direct binding and the enzymatic inhibition. Surface plasmon resonance (SPR) biosensor (Biacore 3000) was used to test for binding affinity; this was followed by radioactive methylation assay to test the inhibition of enzymatic activity. As SPR is one of the best methods for studying label-free binding affinity,[58, 59] it is frequently used in drug discovery. To verify whether the 125 compounds could potentially inhibit G9a activity through direct binding to the enzyme, we carried out the experiments by using Biacore 3000. G9a was immobilized on the sensor chip; then the ligand and the candidate compounds sequentially flowed over the chip surface in the concentration range of 10-100 μM. Any interactions between the compounds and G9a would be detected as changes of response units (RU) in the sensorgram. Among the 125 candidates, 14 were found to interact with G9a (Fig. 2.3). The sensorgram response to the candidate compounds ranged from 10RU to approximately 200RU. Compounds that failed to display any interactions with G9a were excluded from subsequent testing.

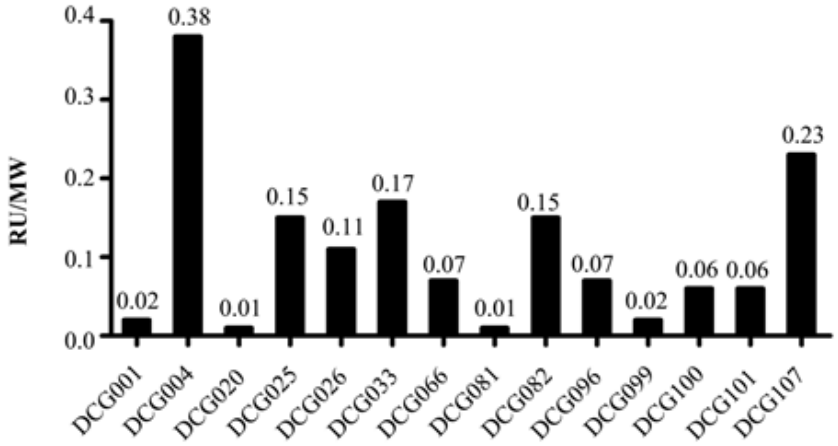
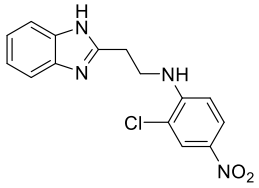
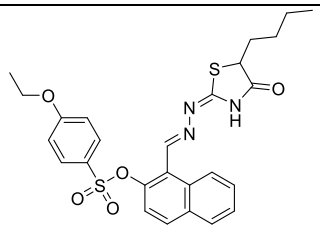
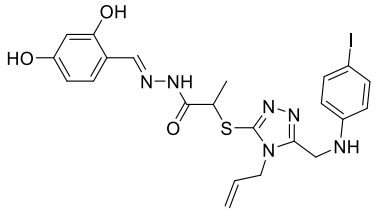
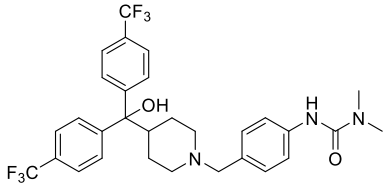
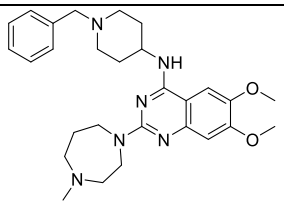


Figure 2.3. Small molecules binding screen by SPR (Biacore 3000).
Through small molecule-protein binding assay using SPR, out of the 125 compounds obtained from virtual screening 14 were found to have the capacity to interact with the G9a catalytic domain. The compounds concentration is 100uM dissolved in the analysis buffer (PBS added P20 and 5% DMSO).

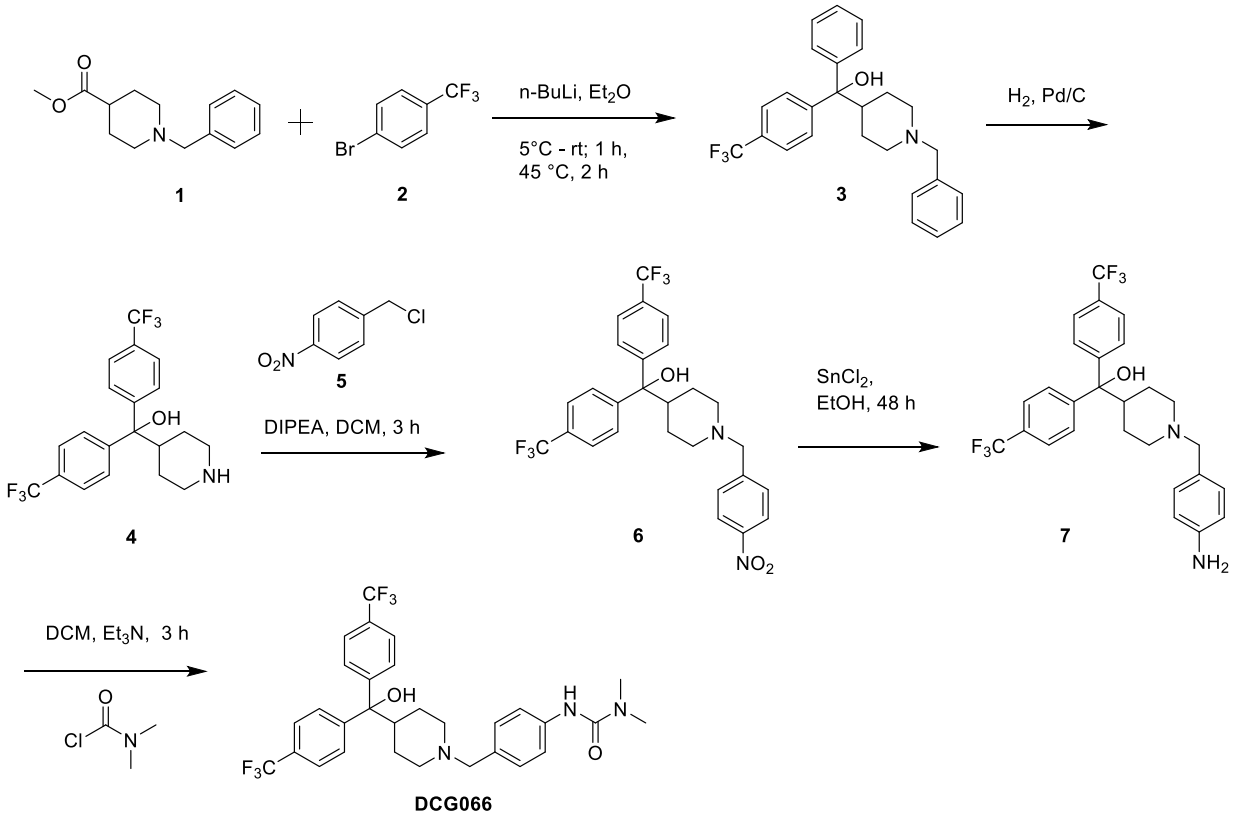
Table 2.1. *Chemical structures of identified small molecules and their inhibition potency.*

Hypothesis	SPECS ID	Compound structure	Inhibition Ratio* %	Cell Inhibition * %
DCG001	AH-034/32473019		61	34
DCG020	AE-848/11486036		74	65
DCG033	AK-968/40335567		74	91
DCG066	AK-105/40833871		98	97
BIX-01294	-		88	98

*Compounds used in both G9a enzyme inhibition assay and cell proliferation inhibition assay had a concentration of 10 μM.

Having recognized that 14 of the compounds screened showed significant binding with G9a, we employed a tritium (H^3)-labeled radioactive methylation assay to test the inhibition of enzymatic activity.[60] 10 µM solutions of the compounds were used for radioactive assays and four of these compounds displayed more than 50% inhibition at this concentration. Of these four compounds, DCG066 displayed inhibition comparable to that of BIX-01294 (Table 2.1). Moreover, all four compounds possess molecular structures that do not fall into any of the three categories of existing G9a inhibitors. Tentatively, the newly discovered DCG066 could be classified as a type IV inhibitor.

2.1.2.3 Synthesis



Scheme 2.1. Synthesis of DCG066

For confirmation and further detailed mechanistic studies, we synthesized DCG066 (Scheme I). Commercially available methyl 1-benzylpiperidine-4-carboxylate (**1**) was used as the starting material to synthesize **3** upon reaction with 1-bromo-4-(trifluoromethyl)benzene (**2**) after lithiation. Subsequent hydrogenolysis afforded a free secondary amine (**4**), which was converted to **6** upon nucleophilic substitution with *para*-nitrobenzylchloride (**5**). Reduction of the nitro group to amine **7** followed by reaction with dimethylcarbamic chloride yielded the targeted compound DCG066.

2.1.2.4 In-vitro biochemical assays

To confirm the result of the radioactive methylation assay, we did two more assays. To examine the interactions between DCG066 and G9a, a kinetic assay by SPR was used with concentrations of DCG066 ranging from 25 µM to 300 µM. The test was performed using Biacore 3000 in Wizard mode. G9a protein was immobilized on the CM5 sensor chip as ligand by amine coupling. The concentration of DCG066 was from 25 µM to 300µM. Since the sensogram of each concentration can reach equilibrium, we choose steady-state analysis to calculate the affinity constant (K_d) of DCG066 (K_d=18.05±2.48 µM, Figure 2.4 A and 2.4 B). MALDI-TOF mass spectrum was used to quantitatively determine the methylation of the histone H3 peptides catalyzed by G9a enzyme with and without DCG066. The assay was performed ass followed, 100 nM of protein G9a, 2 µM of substrate peptide and 1 µM S-adenosyl methionine (SAM) were added to 50 mM HEPES pH 8.0, 5 µg/ml BSA and 0.1% β-mercaptoethanol with DCG066 at a final concentration of 10 µM. After 30 minutes of reaction, TFA was added to the buffer to terminate the reaction. 1 µL of the reaction solution was used to detect the modifications of histone H3 (1-24). The relative intensity of H3K9me2 in the sample after the addition of DCG066 was found to be lower than that without

DCG066, indicating the ability for DCG066 to inhibit H3K9-dimethylation (Figure 2.4 C & 2.4 D).

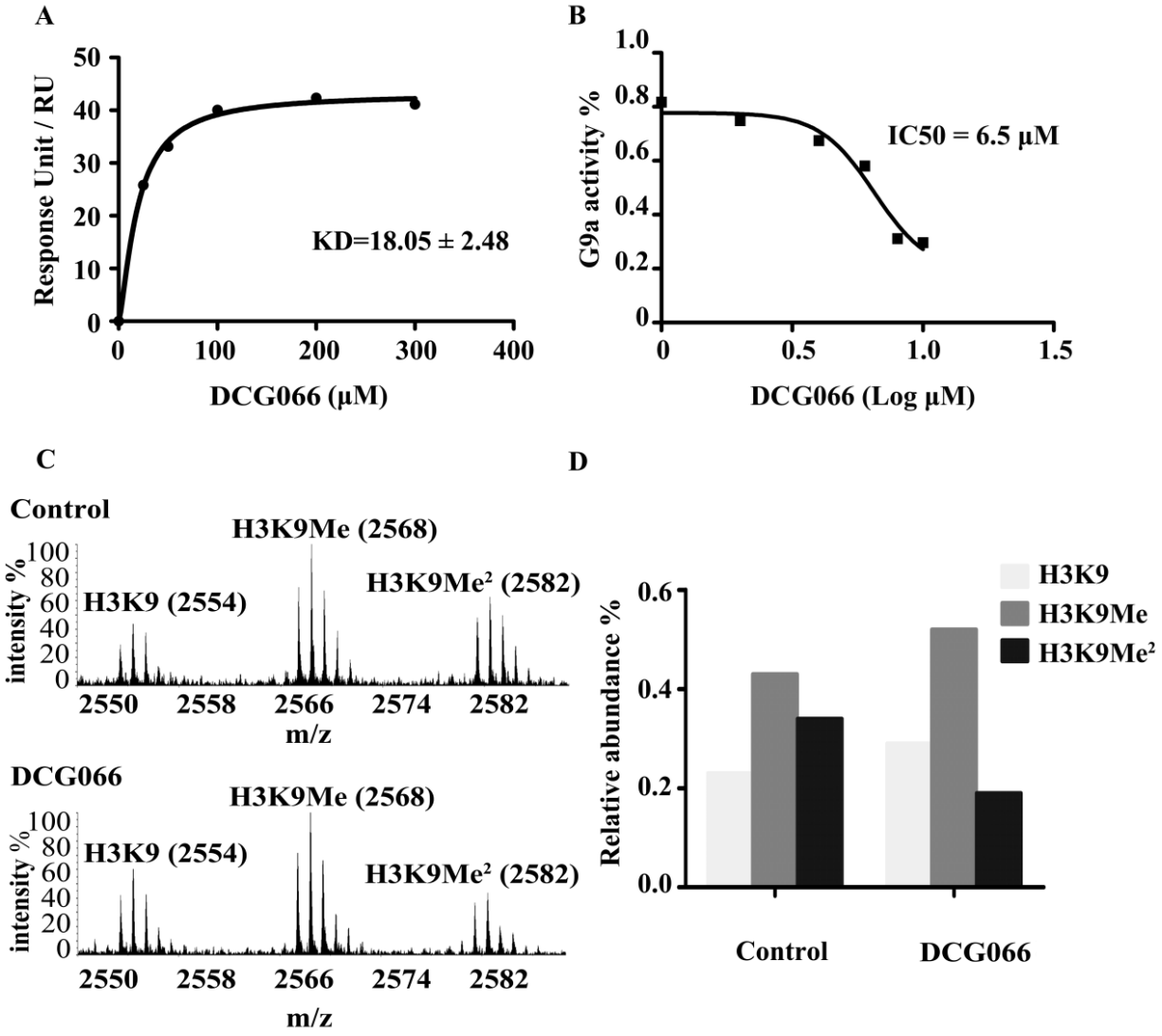


Figure 2.4. In vitro kinetic assay and enzyme activity inhibition of DCG066. **(A)** Binding assay of DCG066 and recombinant G9a catalytic domain by Biacore 3000. The K_d is approximately 18 μM calculated by nonlinear least-square (NLLSQ) fitting. **(B)** DCG066 can inhibit G9a methyltransferase activity with an IC50 of 6.5 μM, which is like that of BIX-01294. **(C, D)** In MADI-TOF-MS, dimethyl level of synthetic peptide of histone H3 (1-20) decreased with the addition of DCG066, while the monomethyl level increased.

2.1.2.5 DCG066 inhibits cancer cell proliferation, decreases histone H3 lysine 9 (H3K9) di-methyl levels, blocks cell cycle and induces apoptosis

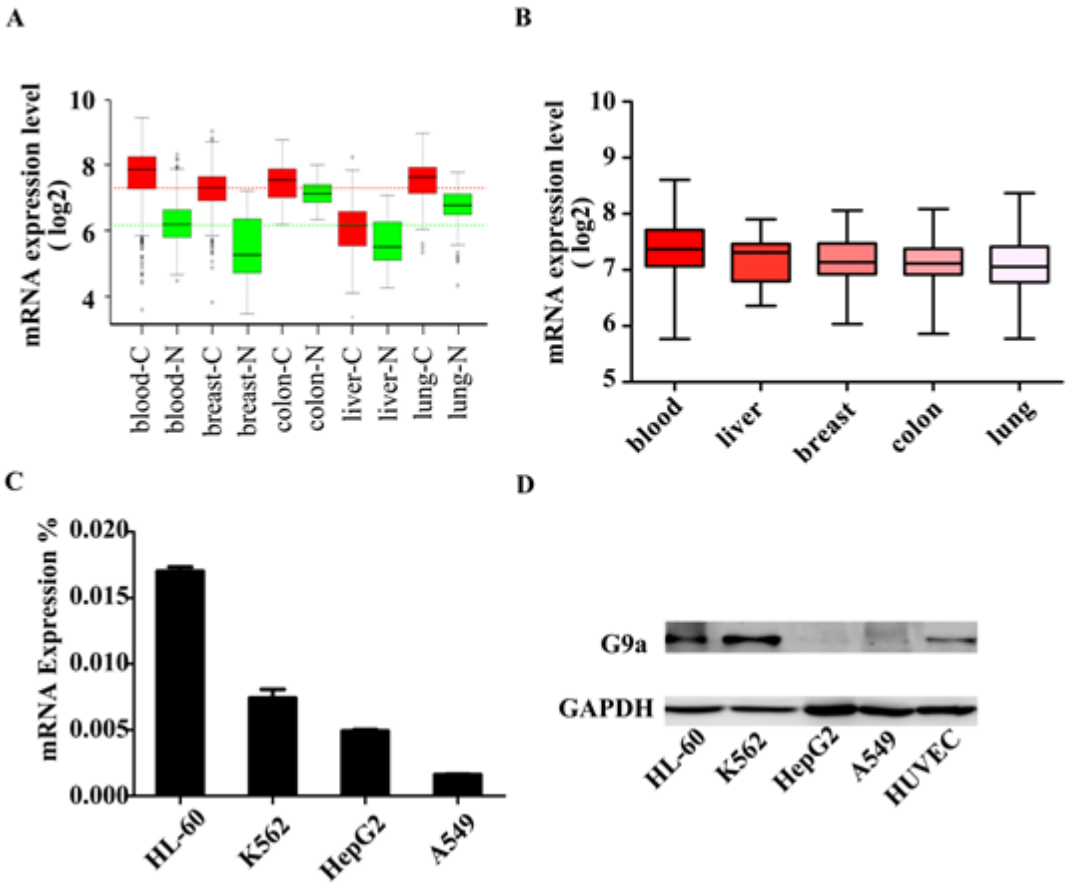


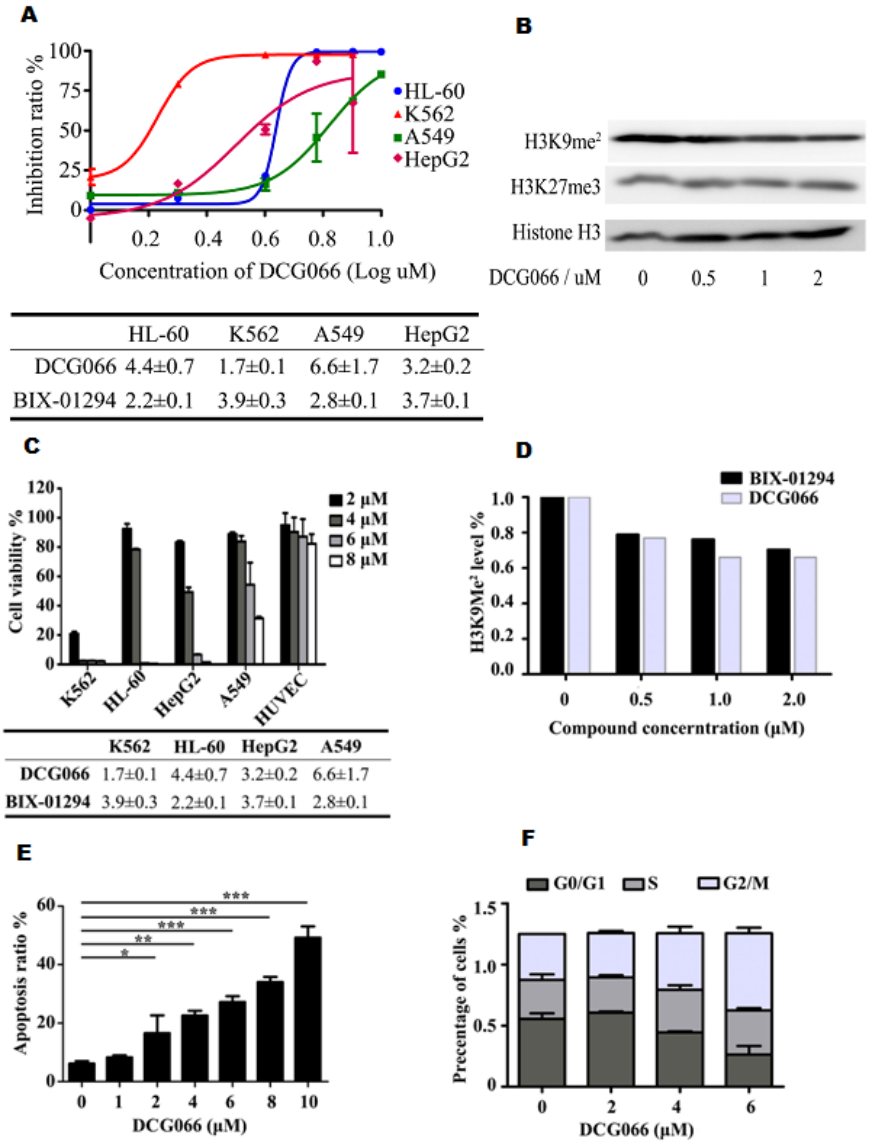
Figure 2.5. G9a expression in human cells and cancer cells.
(A) Pattern of G9a expression across diverse normal and tumor tissues was obtained from GENT (http://medical-genome.kribb.re.kr/GENT/). The average G9a expression in cancer tissue is higher than normal tissues, particularly for hematology system. **(B)** Analysis of G9a expression level in different cancer cell lines. The statistical data was obtained from CCLE (http: // www. broadinstitute. org/ccle/home). G9a is expressed higher in leukemia cell lines than other cancer cells. **(C, D)** The expression of G9a in different cancer cells (K562, HL-60, HepG2, A549 and HUVEC) was detected in both mRNA level (C) and protein level (D). Expression of G9a in leukemia cells is higher than other cancer cells (A549 and HepG2) and normal cells (HUVEC).

In order to find a good system to test the G9a inhibitor's inhibition activity in cell, the distinction of G9a gene expression between cancer tissues and normal tissue were analyzed by using GENT[61, 62] (*Gene Expression across Normal and Tumor tissues)* database (http://medical-

genome.kribb.re.kr/GENT/). Meanwhile, the G9a gene expression levels of all available cancer cell lines were analyzed by using the databased of *Broad-Novartis Cancer Cell Line Encyclopedia* (http: // www. broadinstitute. org/ccle/home). Compared to normal cells, G9a expression levels of cancer cells are higher, especially in malignant hematologic cancer cell lines (Figure 2.5 A and 2.5 B). To ensure that the cells lines used in our following experiments have appropriate levels of G9a expressed, the G9a mRNA expression levels were tested by using real time-PCR (Figure 2.5 C) and G9a protein expression levels were checked by using western blot (Figure 2.5 D). mRNA and protein levels of G9a in the tested leukemia cells (HL60 and K562 cell lines) are higher than the levels in solid cancer cells (A549 for lung cancer and HepG2 for liver cancer). Although normal cells (HUVEC) has detectable G9a protein expressed, it is obviously lower than that in leukemia cells.

Throughout the *in vitro* enzymatic activity assay, DCG066 is the best compound in inhibiting G9a activity among the 125 tested compounds. To verify the ability of these compounds to inhibit cell proliferation, DCG066 was tested at varying concentrations in HL-60, K562, A549 and HepG2 cell lines, and BIX-01294 was used as a control (Figure 2.6 C). The half maximal inhibitory concentrations (IC_{50}) of DCG066 were roughly equivalent to BIX-01294 across the cell lines, with K562 displaying the greatest sensitivity to DCG066 (IC_{50} =1.7±0.3 μM) (Figure 2.6 A).

DCG066 was also considered for its ability to decrease the level of dimethylated histone substrate (H3K9me2). After treating K562 cells with DCG066 at varying concentrations (0.5 μM to 2 μM) and incubating for 72 hours, histones were extracted from cell. The quantified total histones were measured through immuno-blotting by using anti-histone H3 antibody (Cell Signaling Technology) and anti-histone H3 methylation antibodies (Abcam) to detect expression levels of the relevant histone methylation levels. DCG066 reduced histone H3 lysine 9-



	HL-60	K562	A549	HepG2
DCG066	4.4±0.7	1.7±0.1	6.6±1.7	3.2±0.2
BIX-01294	2.2±0.1	3.9±0.3	2.8±0.1	3.7±0.1

	K562	HL-60	HepG2	A549
DCG066	1.7±0.1	4.4±0.7	3.2±0.2	6.6±1.7
BIX-01294	3.9±0.3	2.2±0.1	3.7±0.1	2.8±0.1

Figure 2.6. The anti-cancer effect of DCG066 in cell

(A) IC50 of the DCG066 on inhibition of cells proliferation, similar with BIX-01294, the effect acted on leukemia cells maybe much better than it on adherent cells (A549 for lung cancer and HepG2 for liver cancer). (B) The test of histone methylation level in K562 with the treatment of DCG066. DCG066 can reduce histone H3 lysine 9 dimethylation (H3K9me2) levels in K562 cell but can't influent H3K27me3 level. (C) DCG066 can inhibit the proliferation of several cancer cells, but effects in healthy cells are small. The cell viability rates shown as bar, the half maximal inhibitory concentrations (IC50s) are shown as table under the bar Fig. 4A. The IC50 of DCG066 is like BIX-01294 and may act better on leukemia cells compared to adherent cells, such as A549 and HepG2. (D) The test of histone methylation level in K562 with the treatment of DCG066. DCG066 can reduce histone H3 lysine 9 dimethylation (H3K9Me2) levels in a lower concentration than BIX-01294. (E, F) DCG066 can induce apoptosis of K562 cell, with the treatment of 2 μM DCG066 for 24 hours. Meanwhile, DCG066 can block cell cycle in G2/M stage in a low concentration. All these procedures were taken by flow cytometry assay.

dimethylation (H3K9Me2) levels in a comparable manner as BIX-01294 (Figure 2.6 D). 1.0 -2.0 µM DCG066 decreased H3K9-dimethylation in K562 cells but displayed no changes in H3K27 tri-methylation (Figure 2.6 B). Based on these findings, it can be concluded that DCG066 is a G9a selective inhibitor. Upon confirming DCG066's ability to inhibit K562 cell proliferation, the potential impact of DCG066 on the cell cycle and apoptosis was investigated. After treating K562 cells with DCG066 and incubating for 24 hours, cell cycle and apoptosis were detected using flow cytometry. DCG066 was found to block the cell cycle at stage G2/M in the K562 cells, potentially leading to cell apoptosis (Figure 2.6 E, 2.6 F).

2.1.2.6 DCG066 binding pocket similarity to BIX-01294 binding pocket

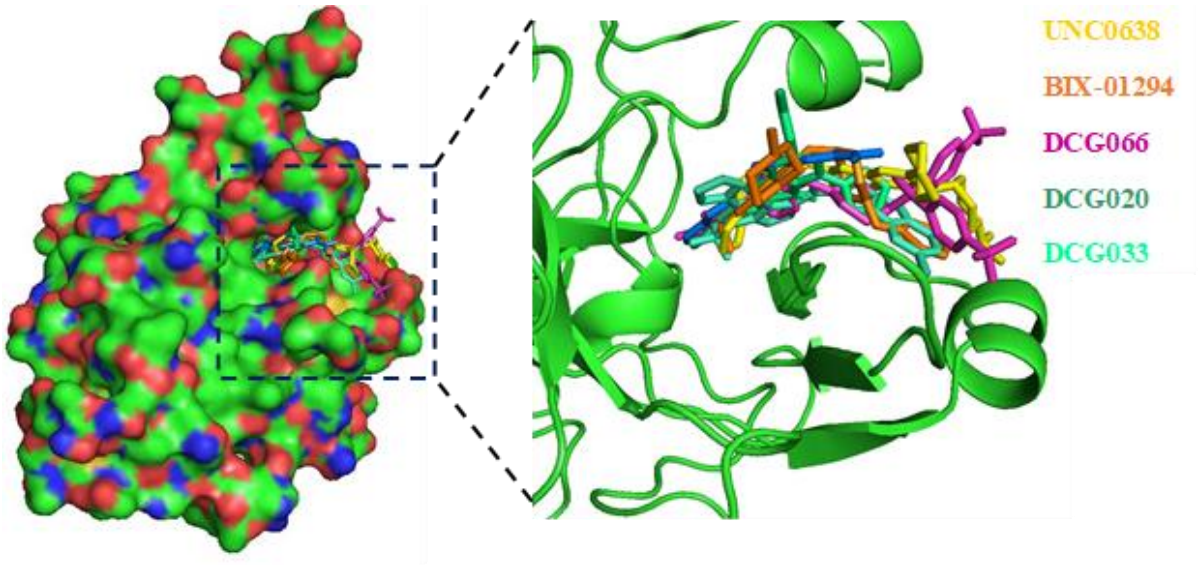


Figure 2.7. Docking images selected from select molecules.
Two positive controls (UNC0638 and BIX-01294) and four hit compounds (Molecule 1 to 4) were docked into G9a binding pocket.

To better understand the binding mechanism, we conducted molecular docking studies with the two positive controls (UNC0638 and BIX-01294) and the selected 14 molecules from virtual screening. Because UNC0638 is also the co-crystallized ligand, we calculated RMSD between the re-docked ligand and co-crystalized ligand. Our result showed that the RMSD value is 0.995 A°,

suggesting our re-docking procedure is reliable. All 16 re-docked molecules are ranked by "Docking Score" and "Enzymatic Inhibit Activity" (Table 2.2). Considering the overall ranking, DCG066 has the highest performance, therefore DCG066 was chosen for the detailed study (Table 2.2). Selected docking result images were rendered in Figure 2.7.

Table 2.2. *Docking score and enzymatic inhibition activity*

Molecules		Docking Score	Enzymatic Inhibit Activity (%)	SPR Response (RU/MW)	Cell Growth Inhibition Activity (%)
Positive Control	UNC0638	8.7093	100%	Not Test	100%
Molecule 1	**DCG066**	**7.6982**	**98%**	**0.07**	**97%**
Positive Control	BIX01294	7.593	88%	Not Test	98%
Molecule 2	DCG020	5.6658	74%	0.01	65%
Molecule 3	DCG033	4.7977	74%	0.17	91%
Molecule 4	DCG001	6.7903	61%	0.02	34%
Molecule 5	DCG025	7.0443	less than 50%	0.15	Not Test
Molecule 6	DCG026	6.7541	less than 50%	0.11	Not Test
Molecule 7	DCG081	6.4155	less than 50%	0.01	Not Test
Molecule 8	DCG082	6.0256	less than 50%	0.15	Not Test
Molecule 9	DCG096	5.8841	less than 50%	0.07	Not Test
Molecule 10	DCG099	5.6738	less than 50%	0.02	Not Test
Molecule 11	DCG100	5.4047	less than 50%	0.06	Not Test
Molecule 12	DCG101	5.1079	less than 50%	0.06	Not Test
Molecule 13	DCG107	4.9077	less than 50%	0.23	Not Test
Molecule 14	DCG004	4.3035	less than 50%	0.38	Not Test

We also studied the interactions between DCG066 and G9a compared to BIX-01294 and UNC0638 (Figure 2.8). Top scored conformations were chosen for analysis. Results indicated that UNC0683, BIX-01294 and DCG066 all bind to the peptide substrate pocket (Figure 2.8 A & 2.8 B). From the model, BIX and DCG066 were shown to form hydrophobic contacts with residues Y1085, Y1067, F1087, F1152, Y1154, R1157 and F1158 (Figure 2.8 C-2.8 F). In addition to these, DCG066 also forms extra hydrogen bonds with Y1154 and R1157 (Figure 2.8 F), offering a potential explanation of its binding.

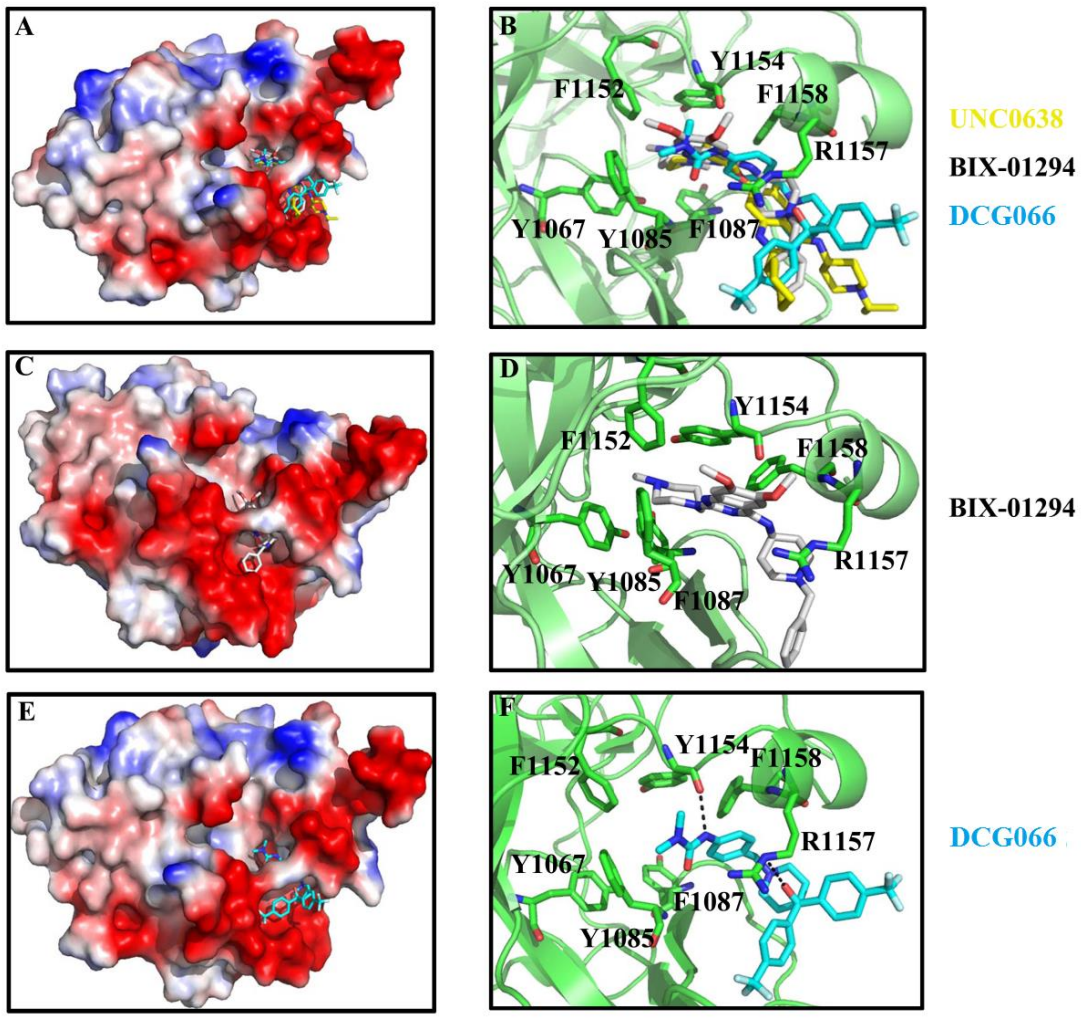


Figure 2.8. Complex models of G9a with superimposed inhibitors (UNC0683, BIX-01294, DCG066).
Left panel (A, C, E) shows the structural electrostatic surfaces (blue indicates positive charge and red indicates negative charge) and small molecules shows as sticks. Right panel (B, D, F) shows

detailed interactions. G9a shows as cartoon in green color and residues, which could form interactions with UNC0683, BIX-01294 or DCG66 are shown as green sticks. (A, B) The binding pocket of crystal structure of G9a-UNC0683 (PDB: 3RJW). [45] (C, D) The computational model of G9a-BIX-01294. (E, F) The computer mimic model of G9a-DCG066. In these models, the interaction surfaces of all these compounds are very similar. Both BIX-01294 and DCG066 can form hydrophobic interactions with Tyr1085, Tyr1067, Phe1087, Phe1152, Tyr1154, Arg1157 and Phe1158. DCG066 can also form hydrogen bond with Tyr1154 and Arg1157 (F).

2.1.3 Conclusion

H3K9 methylation is a crucial epigenetic regulator in gene transcription and embryonic cell differentiation. Thus, G9a is emerging as a drug target for leukemia treatments. Recent studies have shown that G9a may selectively regulate the fast proliferating myeloid progenitors but displays no effects on hematopoietic stem cells (HSCs). Primary acute myeloid leukemia (AML) cells from patients are found to be sensitive to G9a inhibition, causing reduction in proliferation and self-renewal through attenuating HoxA9-dependent transcription.[46]

Many of the current G9a selective inhibitors can be sorted into three categories, all of which are derivatives of BIX-01294. In this study, several small molecules from the SPECS Database were screened using virtual screening to discover potential G9a inhibitors. Virtual screening followed by bioassays allowed us to identify a new scaffold embedded in DCG066 for further inhibitor development. Initially we investigated various cancer cell lines for their G9a expression profile using immune assay and found leukemia cell lines to have the highest expression levels of target protein, hence this cell line were used for the cell based assay. Inhibition of cancer cell growth was tested by CCK8 and the IC_{50} of DCG066 found to be comparable to BIX-01294 (1.7±0.1 and 3.9±0.3 respectively in K562 cell line). DCG066 also demonstrated the ability to induce apoptosis through blocking the cell cycle at stage G2/M in the K562 cells. Further experiments with DCG066 showed that compound decreased H3K9 dimethylation levels in K562 cells.

To examine the interactions between DCG066 and the catalytic domain of G9a more closely, a binding assay was performed by SPR and the structure-function relationship was modeled using GLIDE. DCG066 can bind to G9a with low dissociation constant (K_d =18 μM) as measured by SPR. While DCG066 binds in the peptide substrate pocket in G9a with a pose similar to BIX-01294, DCG066 forms extra hydrogen bonds with residues Y1154 and R1157, suggesting an explanation for the relative activity between the two molecules. In this study, a new lead compound for G9a inhibition with a similar activity to BIX-01294 and a novel structure, DCG066, was identified from a database of small molecules. This discovery provides both new avenues in G9a inhibitor design and a potential probe for G9a. In future studies, we plan to examine the structure activity relationship (SAR) of DCG066 analogs to G9a in an effort to improve potency and reduce toxicity.

2.1.4 Experimental Section

2.1.4.1 Materials and physical measurements

All reactions were carried out under an atmosphere of dry nitrogen. All reagents and anhydrous solvents were obtained commercially and used without further purification. ^{1}H NMR spectra were recorded at 400 MHz and ^{13}C NMR spectra were recorded at 100 MHz using Me_4Si as an internal standard. High-resolution mass spectra (MS) data were acquired using a Thermo-Fisher Orbitrap Elite mass spectrometer with an electrospray ionization (ESI) source. All the samples were run under FT control at 600,000 resolution. All chromatographic purification was performed on Biotage Isolera One using normal phase silica gel 60. HPLC chromatogram for DCG066 compound was acquired using an Agilent 6110 Series system with UV detector set to 254 nm. Samples were injected (5 μL) onto an Agilent Eclipse Plus 4.6 × 250 mm, 3.5 μM, C18 column at room temperature. Sample was eluted using a linear gradient of 5% to 95% B (ACN +

0.1% TFA) in 30 min with A being H_2O + 0.1% TFA. DCG066 had >95% purity using the HPLC method described above.

2.1.4.2 Chemical synthesis

(1-Benzylpiperidin-4-yl)bis(4-(trifluoromethyl)phenyl)methanol: Synthesis of the intermediate compound **6** was carried out according to the previously reported procedure.[63] *n*-Butyllitium solution (19 mL, 1.6 M in hexane) was added to 6.8 g of 4-bromobenzotrifluride **2** (30 mmol) in diethyl ether. To this mixture 2.4 g (10 mmol) of methyl-1-benzylpiperidine-4-carboxylate **1** were added dropwise at 5 ^oC. The reaction mixture was stirred at room temperature for one hour and at 45 ^oC for 2 hours. The crude product was obtained according to the reported procedures, and purified by silica gel flash column chromatography using 0-50% ethyl acetate/hexane as eluent to yield a brown solid. (4.4 g, 77% yield). ^1H NMR ($CDCl_3$): δ 7.64 (m, 8H, 2 x trifluromethylphenyl Ar*H)*, 7.33-7.30 (m, 5H, Benzyl Ar*H*) 3.56 (s, 2H, Benzylic C*H*$_2$), 3.00-2.97 (m, 2H, N-CH_2-C*H*$_2$), 2.54 (t, J=11.6Hz, 1H, N-CH_2-CH_2-C*H*), 2.10-2.07 (m, 2H, N-CH_2-C*H*$_2$), 1.60-1.45 (m, 4H, 2 x N-C*H*$_2$). 13C NMR ($CDCl_3$) δ 150.9, 149.6, 139.4, 128.4 (2C), 128.1 (4C), 127.3 (2C), 126.1 (2C), 125.8 (2C), 125.3 (3C), 125.2 (2C), 81.4, 62.7, 53.8 (2C), 43.8, 26.2 (2C). HRMS (ESI) m/z: calcd for $C_{27}H_{25}F_6NO$ $[M + H]^+$ 494.1919; found 494.1883.

(1-(4-Nitrobenzyl)piperidin-4-yl)bis(4-(trifluoromethyl)phenyl)methanol: The mixture of compound **3** (2 g, 4.04 mmol) and 5 wt% Pd/C (500 mg) in ethanol (200 mL) was stirred for 24 hours at room temperature under hydrogen balloon (1 atm). The reaction mixture was filtered and concentrated to a sticky solid. HRMS (ESI) m/z: calcd for $C_{20}H_{19}F_6NO$ $[M + H]^+$ 404.1449 ; found 404.1426. Compound **4** was used for next step without further purification. The mixture of compound **4** (1.2 g, 3 mmol) and DIPEA (1.2 mL) in 10 mL of methylene chloride were added into the solution of 4-nitrobenzylchloride **5** (600 mg, 3.5 mmol) in 20 mL of methylene chloride.

The reaction mixture was stirred for 3 hours at room temperature; then saturated $NaHCO_3$ was added. The mixture was extracted with ethyl acetate, washed with water, dried and concentrated to give the crude product, which was purified by column chromatography using 0 -40% ethyl acetate/hexane as an eluent to yield a pale yellow solid. (1.1 g; 67% yield). ^{1}H NMR ($CDCl_3$): δ 8.17 (d, *J*=8.7Hz, 2H, Benzyl, *m*- Ar*H*), 7.61 (m, 8H, 2 x triflurometh ylphenyl Ar*H*), 7.50 (d, *J*=8.7Hz, 2H, Benzyl, *o*- Ar*H*), 3.60 (s, 2H, Benzylic CH_2), 2.92 (m, 2H, N-CH_2-CH_2), 2.52 (m, 1H, N-CH_2-CH_2-C*H*), 2.12-2.09 (m, 2H, N-CH_2-CH_2), 1.59-1.45 (m, 4H, 2 x N-CH_2). 13C NMR (CDCl3) δ 149.0 (2C), 147.1, 146.3, 129.4 (4C), 129.0, 126.0 (6C), 125.4 (2C), 125.3, 123.5 (2C), 79.3, 62.1, 53.7 (2C), 43.8, 23.2 (2C). HRMS (ESI) m/z: calcd for $C_{27}H_{24}F_6N_2O_3$ $[M + H]^+$ 539.1769; found 539.1737.

(1-(4-Aminobenzyl)piperidin-4-yl)bis(4-(trifluoromethyl)phenyl)methanol: $SnCl_2.2H_2O$ (900 mg, 4.0 mmol) was added into the solution of compound **6** (1.1 g, 2.0 mmol) in EtOH (20 mL) at r.t., a mild exotherm was observed. The resulting yellow solution was stirred for 2 days, and then the solvent was removed *in vacuo*. The residue was made basic with a 2N NaOH solution and the aqueous layer was extracted with DCM (2x25 mL). The combined organic layers were dried over Na_2SO_4 and then filtered. The solvent was removed *in vacuo* to yield a pale yellow solid (0.65 g, 65% yield), which was used in the next step without further purification.[64] ^{1}H NMR ($CDCl_3$) δ 7.59 (m, 8H, 2 x triflurometh ylphenyl Ar*H*), 7.07 (d, *J* = 7.5 Hz, 2H, benzyl, *m*- Ar*H*), 6.63 (d, *J* = 7.5 Hz, 2H, benzyl, *o*- Ar*H*), 3.62 (s, 1H, O*H*), 3.45 (s, 2H, benzylic CH_2), 3.27 (s, 2H, NH_2), 2.97 (d, *J* = 11.2 Hz, 2H, N-CH_2-CH_2), 2.49 (t, *J* = 11.6 Hz, 1H, N-CH_2-CH_2-C*H*), 2.03 (t, *J* = 11.4 Hz, 2H, N-CH_2-CH_2), 1.69 – 1.50 (m, 2H, N-CH_2), 1.43 (d, *J* = 12.6 Hz, 2H, N-CH_2). ^{13}C NMR ($CDCl_3$) δ 149.4 (2C), 145.6, 130.7 (4C), 129.1, 128.8, 126.7, 126.1 (2C), 125.4 (4C), 125.2

(2C), 114.9 (2C), 79.28, 62.5, 53.1 (2C), 43.8, 25.7 (2C). HRMS (ESI) m/z: calcd for $C_{27}H_{26}F_6N_2O$ $[M + H]^+$ 509.1983; found 509.2000.

3-(4-((4-(Hydroxybis(4-(trifluoromethyl)phenyl)methyl)piperidin-1-yl)methyl)phenyl)-1,1-dimethylurea (DCG066): Compound **7** (254 mg, 0.5 mmol) was dissolved in 20 mL of methylene chloride. To this resulting solution was then added DIPEA (0.18 ml, 1 mmol). After stirring for 30 min at 0 ^oC, dimethyl carbamoyl chloride (54 µl, 0.6 mmol) was added and the reaction was stirred for another 3 hours. Saturated $NaHCO_3$ solution was added and the mixture was extracted with DCM (3x20 mL). The combined organic phase was dried over Na_2SO_4 and concentrated under reduced pressure. The residue was purified on silica gel column, eluting with 5% MeOH in DCM (containing 0.5% Et_3N) to give a yellow solid (180 mg, 62% yield). ^1H NMR ($CDCl_3$): δ 7.57 (m, 8H, 2 x trifluromethylphenyl Ar*H*), 7.30 (d, *J*=8.4Hz, 2H, benzyl, *m*- Ar*H*), 7.16 (d, *J*=8.4Hz, 2H, benzyl, *o*- Ar*H*), 6.57 (s, 1H, OH), 3.47 (s, 2H, benzylic CH_2), 2.99 (s, 6H, N(CH_3)$_2$), 2.91 (d, *J*=11.4Hz, 2H, N-CH_2-CH_2), 2.47 (t, *J*=11.6Hz, 1H, N-CH_2-CH_2-C*H*), 2.03 (t, *J*=11.0Hz, 2H, N-CH_2-CH_2), 1.60-1.52 (m, 2H, N-CH_2), 1.38 (d, *J*=12.3Hz, 2H, N-CH_2). ^{13}C NMR ($CDCl_3$) δ 155.9, 149.5 (2C), 138.2, 129.8 (2C), 128.9, 128.7, 126.2 (5C), 125.4 (2C), 125.2 (2C), 122.7 (2C), 119.9 (2C), 79.2, 62.3, 53.2 (2C), 43.8, 36.4 (2C), 25.8 (2C). HRMS (ESI) m/z: calcd for $C_{30}H_{31}F_6N_3O_2$ $[M + H]^+$ 580.2354; found 580.2373.

2.1.4.3 Pharmacophore-Based Screening and Computational model.

Pharmacophore models were automatically constructed using LigandStout for the further screening. The SPECS database was filtered with log S > −4 to construct a database of theoretically soluble compounds (log P < 5). The compounds were then docked into G9a (PDB entry: 3RJW) at the UNC0638 binding pocket using the GLIDE program in standard precision mode.[65] The compounds with the highest binding characteristics were then purchased from SPECS Corp (The

Netherlands). The computer model of compound DCG066 and BIX-01294 were docked into G9a (PDB entry: 3RJW) using GLIDE in the XP mode. The final graphs were drawn by Pymol software. Docking validation of 16 molecules was done using Surflex-Dock 2.1[66] by following our previous procedures.[67] RMSD value was calculated using Openeye Toolkit.[68]

2.1.4.4 Biological evaluations

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 ℃.

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), 500mM NaCl, 0.1% β-mercaptoethanol, and 1mM PMSF. Cells were lysed by sonicating for 15 seconds with 6-second intervals for a total of 15 minutes on an ice bath. The supernatant of cell lysate was loaded onto a Ni^+ affinity column (Invitrogen) and eluted with buffer (20mM 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 ℃ for 12 hours. 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).

2.1.4.4.1 Tritium-labeled radioactive methylation assay

The tritium labeled radioactive methylation assay was used to test the inhibitory effects of the compounds on enzyme activity.[60] In the methylation assay, 2 μM biotin labeled peptide

substrate, 5μM [methyl-^{3}H]-SAM (78Ci/mmol, PerkinElmer), and varied concentrations of inhibitor were pre-incubated in the reaction buffer (50 mM HEPES pH 8.0, 10 mM NaCl, 1 mM DTT) for 30 minutes at room temperature. The reaction was initiated by adding recombination mG9a (969-1263) for a final concentration of 2ug/ml. 2uL of the reaction mixture was transferred to a 96 well plate coated with avidin (Corning), incubated for 30 minutes in PBST with 5 mM of unlabeled SAM. The plate was washed 3 times with 200 μl PBST per well to remove the remaining ^{3}H-SAM. Later, 50 mM HCl was used to elute the avidin binding peptide. Finally the elution buffer mixed with 200 μL scintillator fluid was analyzed by 1450 Microbeta scintillation counter (PerkinElmer).

2.1.4.4.2 MALDI-TOF-MS

The *in vitro* inhibition of G9a by the compounds was measured by MALDI-TOF mass spectrum (4800 Plus MALDI TOF/TOF Analyzer, ABI). 100 nM purified mG9a, 2 μM synthesized histone H3 (1-24) and 1 μM non-radioactive *S*-adenosyl methionine (Sigma) were added into a reaction buffer (50 mM HEPES pH 8.0, 5 μg/ml BSA and 0.1% β-mercaptoethanol) with or without an inhibitor for a final concentration of 10 μM. The reaction was incubated at room temperature for an hour, and stopped by TFA addition.

The result of mass spectrum was analyzed using the Data Explorer (TM) software, providing peak area scores while a statistical graph was drawn by Graphpad Prism 5.0.

2.1.4.4.3 Surface Plasmon Resonance

The ability of the small molecular compounds to bind to the target enzyme was tested by using SPR (Biacore 3000, GE). Approximately 5000RU of the target protein was immobilized on a CM5 sensor chip. Test compounds were dissolved in running buffer (PBS + P20 + 5% DMSO) at a concentration of 100 μM.

To screen the compounds, solutions were sequentially injected for 1 minute at the associated stage, dissociated in running buffer for 1 minute, and the sensor chip was then regenerated by running buffer for 1 minute. In this test, an empty cycle per every 5 cycles was run. From the Biacore evaluation software (BIA evaluation version 4.1), the maximum binding values of every compound were obtained.

For the G9a catalytic domain kinetics assay, five concentrations of the compounds were prepared and DMSO contents in these samples were equivalent with 5%. The test was performed in the Wizard mode; and the injection time and dissociation times were recorded. The results were analyzed using the static analysis option in the BIA evaluation software.

2.1.4.4.4 Cell culture

All leukemia cells such as HL-60, K562, U937 and Kasumi-1 were grown in RPMI 1640 with 10% fetal bovine serum (FBS), while adherent cell, including A549, HepG2, HCT116, SW1990, and MDA-MB-231 were grown in Dulbecco's modified Eagle's medium (DMEM) with 10% FBS.

2.1.4.4.5 Human G9a expression level analysis

In this study, G9a expression levels were analyzed both in cancer cell lines and human tissues. CCLE (Cancer Cell Line Encyclopedia, www.broadinstitute. org/ccle/home) and GENT (Gene Expression database of Normal and Tumor tissues, http://medical-genome.kribb.re.kr/GENT/) were the databases that provided information regarding gene expression in cancer cell lines and human tissues, respectively.

2.1.4.4.6 Quantitative Real-Time PCR

For quantitative real-time PCR (qRT-PCR), total RNA was extracted and reverse transcribed into cDNA. qRT-PCR was performed in ABI PRISM 7500 using SYBR Green PCR

Master mix (Takara). The mRNA expression level of G9a was normalized to the transcript level of GAPDH.

2.1.4.4.7 Cell viability analysis

Cell growth inhibition was detected by CCK8 assay (Dojindo). 1×10^4 cell suspensions were dispensed on a 96-well plate, and then pre-incubated for 4 hours in a humidified incubator. Compounds were dissolved in DMSO, diluted in DMEM culture medium for a concentration of 10× stock solution, and added as 10 ul per well. After incubation for 48 hours, 10 uL of CCK8 solution was added into the wells and the solution was allowed to incubate for 4 hours in a CO2 incubator. Afterwards, the absorbance was measured at 450 nm.

2.1.4.4.8 Flow cytometry assays

Cell apoptosis was determined by dual staining with annexin V conjugated to phycoerythrin (PE) and 7-amino-actinomycin (7AAD). K562 cells were treated with 0, 2, 4, 6, 8, 10 µM of DCG066. After incubation for 24 hours, cells were collected and stained with annexin V-PE and 7AAD (BD Pharmingen) for 15 minutes in the dark and analyzed by flow cytometry. Cells undergoing apotosis were identified as annexin V+ and/or 7AAD+ cells.

Additionally, cells (1×10^6) were treated with DCG066 (0, 1, 2, 4 µM) or DMSO for cell cycle analysis. After 24 hours of incubation, cells were collected and washed with cold PBS twice, and then suspended in 300 ul PBS, fixed by 700 ul ethanol. The fixed cells were washed by PBS twice and re-suspeneded in PI/RNase Staining Buffer (BD Pharmingen) and then incubated for 15 minutes before analysis. Flow cytometry experiments were performed using a LSR II cytometer (BD Pharmingen), and data were analyzed by using the FlowJo 7.6.1 software.

2.1.4.4.9 Histone extraction and Western blot

Histones from cell lysates were extracted using trichloroacetic acid precipitation as described previously.[69] 1×10^7 cells were harvested by centrifugation at 800×g for 5 minutes, and suspended on ice in hypotonic buffer (10mM Tris-HCl pH 8.0, 1mM KCl, 1.5mM MgCl2, 1mM DTT, 1mM PMSF) for 30 minutes. The cell lysate was centrifuged at 13000×g for 10 minutes, and the supernatant was discarded. The lysate pellets were re-suspended using 0.2N H_2SO_4, mixing at rotor at 4 ℃ for 30 minutes, centrifuged at 13000×g for 10 minutes at 4 ℃ with the precipitation discarded. 50% trichloroacetic acid (TCA) was added to the supernatant dropwise, and then the resulting solution was centrifuged for 20 minutes. The pellets were washed twice in ice-cold acetone and then dissolved by adding H_2O. Histone content was quantified by a Bradford assay and purity was tested by Coomassie blue staining.

Extracted histone was denatured by SDS loading buffer. Anti-histone H3 antibody (rabbit), anti-histone H3 lysine 9 trimethyl antibody (mouse), anti-histone H3 lysine 9 trimethyl antibody (rabbit), and mono- and di-methyl arginine antibody (mouse) were purchased from Abcam. Anti-histone H3 lysine 27 trimethyl antibody (rabbit) was purchased from Millipore.

2.2 Structure based design, synthesis and activity studies of small hybrid molecules as HDAC and G9a dual inhibitors

The work presented in this chapter is reproduced with some modification as to fit the format with permission from the journal *Oncotarget*. My contributions as first author were the design, chemical synthesis of hybrid molecules, MALDI assay, and figure/manuscript preparation.

Zang, Lanlan, Shukkoor M. Kondengaden, Qing Zhang, Xiaobo Li, Dilep K. Sigalapalli, Shameer M. Kondengadan, Kenneth Huang et al. "Structure based design, synthesis and activity studies of small hybrid molecules as HDAC and G9a dual inhibitors." *Oncotarget* 8, no. 38 (2017): 63187.

ABSTRACT: 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. Here, 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.

2.2.1 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.[70-74] 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.[75-77] 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.[9, 74, 75, 78] 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.[79-84] To meet this need, developing new lead molecules which target cancer from various stages of disease development from either known inhibitors or *de novo* is essential for improving effectiveness and side effects/toxicity profile.[14, 85-89] 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.

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).[90] However, despite the name, histone deacetylases have a wide range of substrates included but not limited to strictly histones.[77] Aberrant activity of HDACs have been well documented in several cancer phenotypes, with HDAC inhibitors (HDACIs, Figure 2.9) 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.[75, 91] HDACIs have been shown to induce apoptosis in both

solid and hematopoietic malignancies using both transcription dependent and transcription independent mechanisms.[92-94] 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.[70, 95] 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.[74, 96, 97] Presently, there are number of small molecules with different structural cores that have been found to inhibit G9a (Figure 2.9), which are also under consideration in clinical trials.[74, 98] 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.[99]

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.[100, 101] However, methylation of H3K9 and H3K27 are associated with gene silencing and repressive marks.[102]

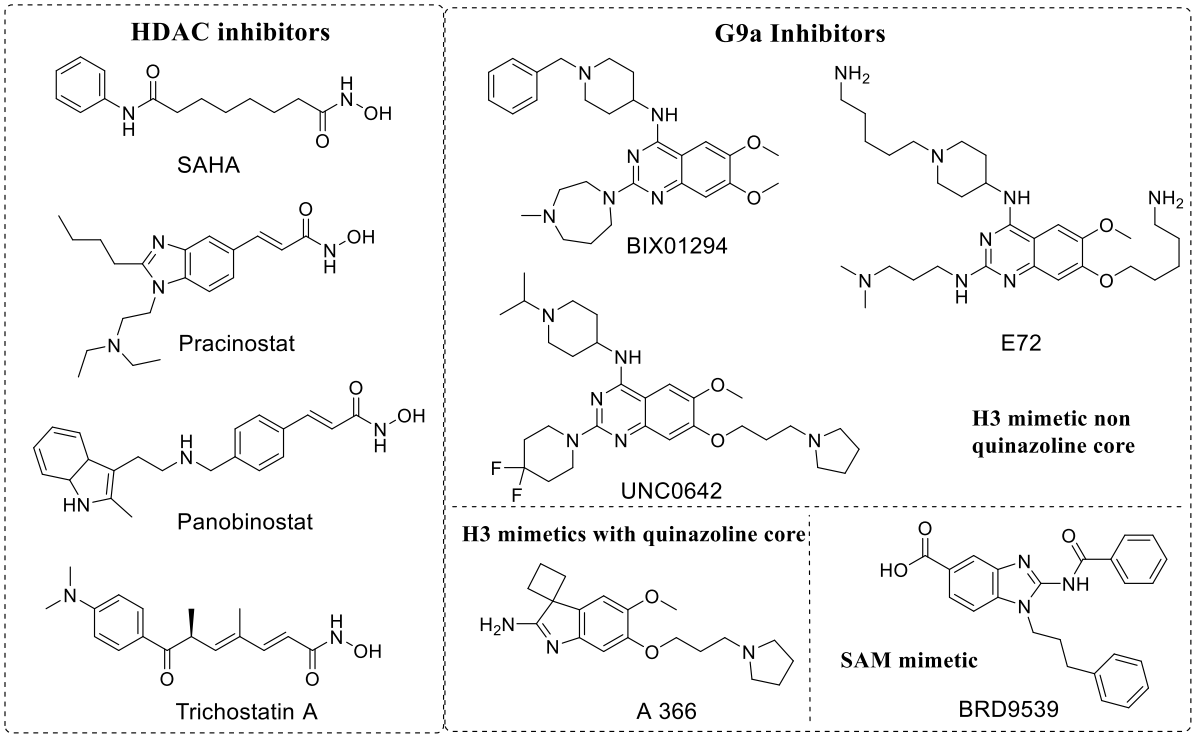


Figure 2.9. Examples for known HDAC inhibitors and G9a Inhibitors. [74, 77, 98, 103]

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.[14, 85, 104] 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.[105]

2.2.2 Results 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 goal, 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 Figure 2.10 and table 2.3, 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.

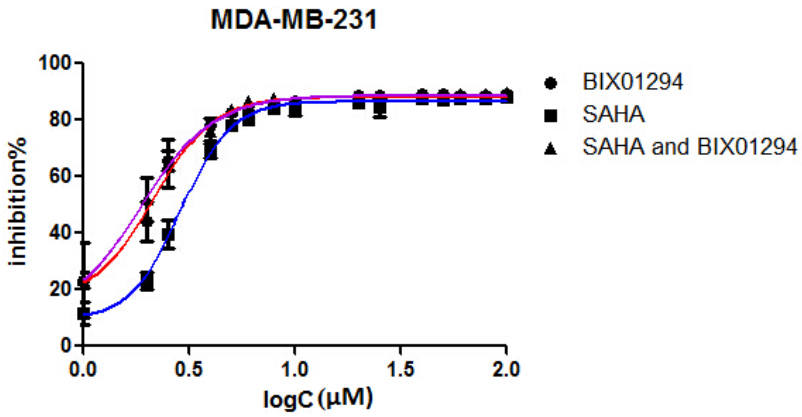


Figure 2.10. EC^{50} plot of BIX and SAHA combination study.

Table 2.3. *Combination study of BIX-01294 and SAHA*

Inhibitor (10 μM)	MDA-MB-231	MCF-7
SAHA	2.874±0.84	8.124±4.98
BIX-01294	2.155±0.88	8.103±1.99
SAHA+BIX-01294	1.891±0.56	8.564±2.17

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 of the formulation to ensure the desired bioavailability.[79, 82, 87, 106, 107] 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.[88] Currently available hybrid drugs target different stages of disease development despite their precursors targeting the same diseases.[85, 87, 107, 108] 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.

2.2.2.1 Compound design and synthesis

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.[13, 14, 79, 88, 109] 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 (Figure 2.11 A).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. Various analogs with different linker lengths and diverse groups at C6 and C4 cyclohexylamine positions were also designed. In Figure 2.11 B, quinazoline core is split to four quarters to discuss the SAR studies.

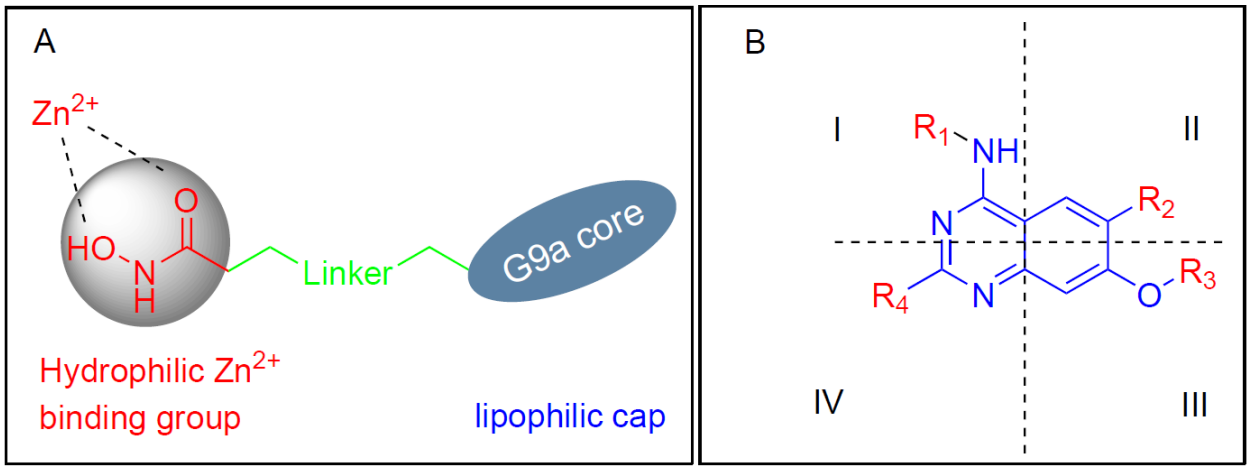
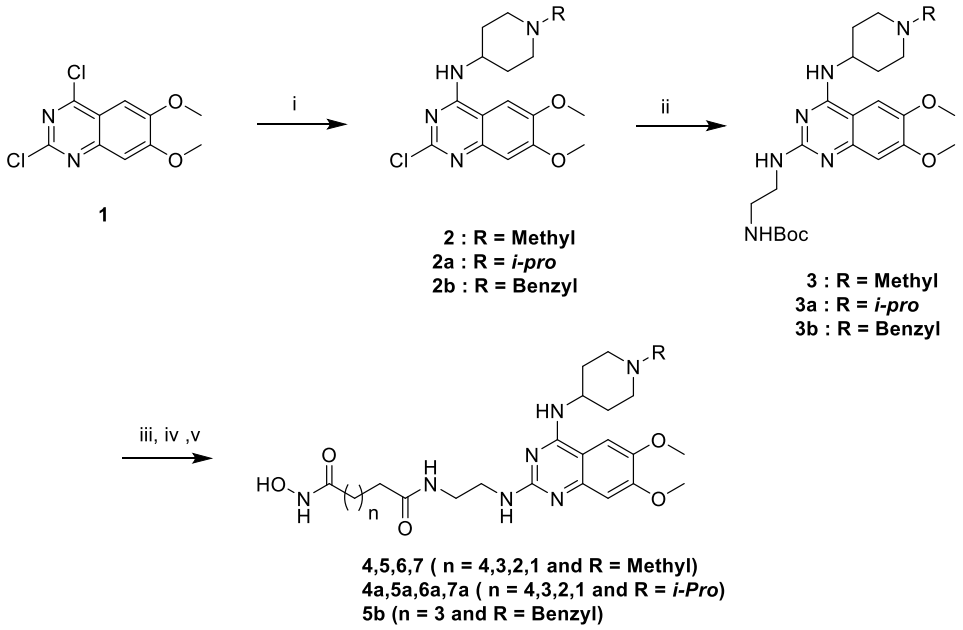


Figure 2.11. A) Design for Hybrid Molecules B) Structural representation of designed compounds (R = Linker + Hydroxamic acid).

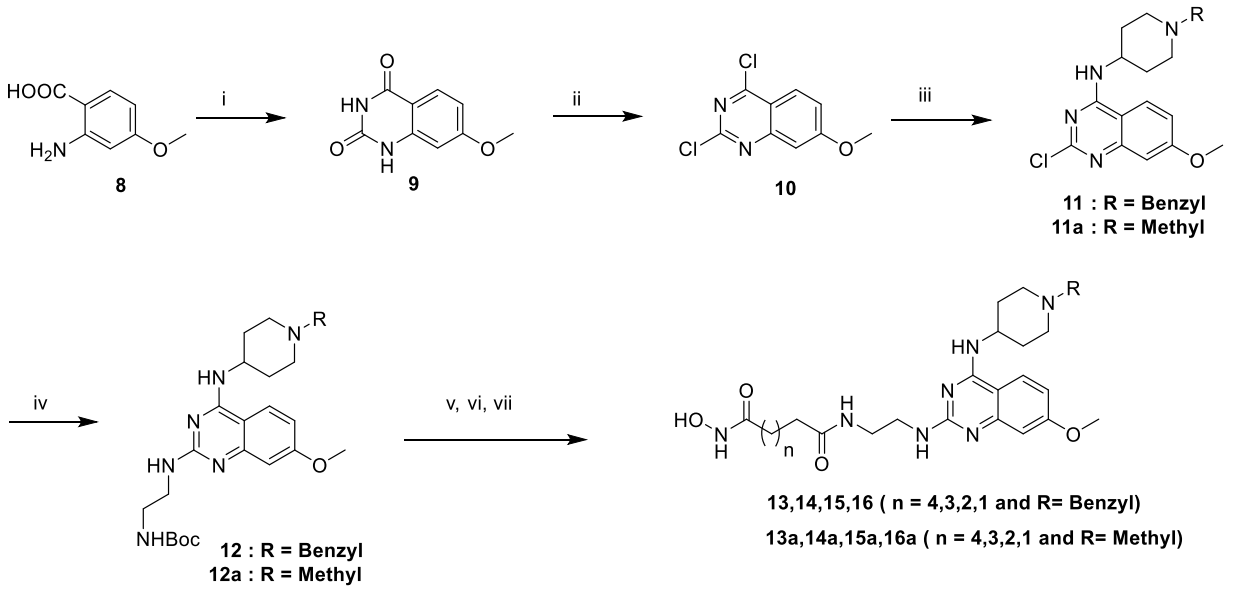
2.2.2.2 Synthesis

The designed analogs were synthesized from two building blocks; commercially available 2,4-dichloro-6,7-dimethoxyquinazoline (**1**) was used for the dimethoxyanalogs (Scheme 2.2) and, 2-amino-4-methoxybenzoic acid (**8**) used for monomethoxyanalogs (Scheme 2.3). 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, Scheme 2.2). While investigating the binding characteristics of known G9a/GLP inhibitors, the C6 methoxy group of quinazoline ring was hypothesized not to

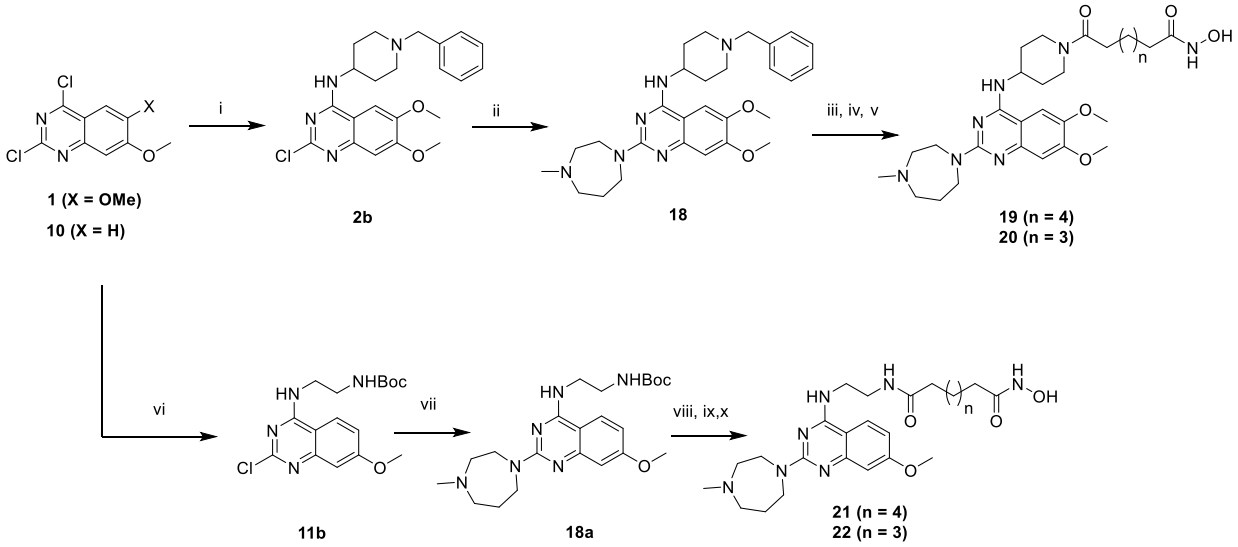
contribute significantly to ligand-receptor interactions. Therefore, the methoxy at C6 was eliminated to find a balance for HDAC activity. Compounds in class II designed from this rationale. Compounds **13-17** have a 4-aminobenzyl piperidine at C4, while **13a-17a** possess an isopropylpiperidin-4-amine (Scheme 2.3). Recent studies of UNC0965 wherein a biotin tag was applied to UNC0638 found that UNC0965 retained biological activity,[110] 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 (Scheme 2.4).



Scheme 2.2. 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 °CMicrowave, 10 min, 60-66%, (iii) TFA/DCM, 3 h, (iv) Monomethyl Suberate, EDCl, HOBt, 8 h, (v) 50% NH_2OH in water, MeOH, 60°C, 8 h , 30-38% over two steps.



Scheme 2.3. 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, EDCl, HOBt, 8 h, 70% in two steps, (vii) 50% NH_2OH in water, MeOH, 60 °C, 8 h , 30-40%.



Scheme 2.4. Reagents and conditions: (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, H_2, 8 h, (iv) Monomethylsuberate/monomethylpimelate, EDCl, HOBt, 8 h, (v) 50% NH_2OH in Water, MeOH, 60 °C, 8 h, 44% and 45%, (vi) NHBoc-ethylinediamine, 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% NH_2OH in water, MeOH, 60 °C, 8 h , 29%-36%.

Scheme 2.2 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.[111] Boc-protected ethylene diamine was treated with compound **2** at 160 ^{o}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 Scheme 2.3 requires the C6 demethoxy core, and this was synthesized according to the reported procedure to yield **10**.[112] Appropriate displacement and coupling reactions in this core as demonstrated in Scheme 2.3 afforded compounds **13-17** and **13a-17a.** Compounds with the HDAC pharmacophore at the C4 position (Figure 2B) were synthesized from the starting material **1**. As in Scheme 2.4, Pd/C hydragenolysis 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.[113] 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.

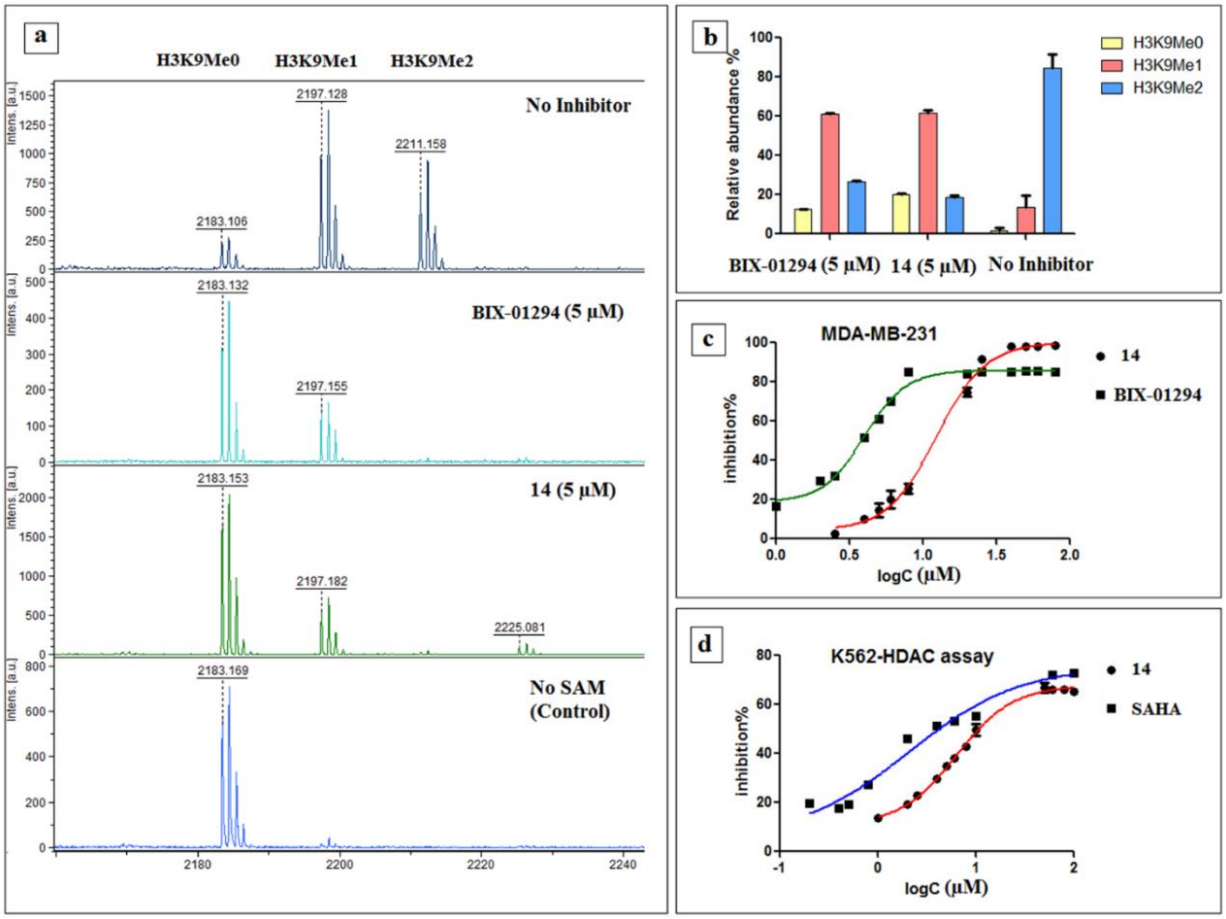


Figure 2.12. Results of in vitro biochemical and cell-based assays.
a) Methylation pattern observed via MALDI-TOF after incubating with inhibitor **14** and BIX-01294 for 30 min, **b)** % ratio of the H3K9Me0, H3K9Me1 and H3K9Me2 after incubating 30 min with **14** and BIX-01294 versus no inhibitor, **c)** In-Cell Western (ICW) assay of **14** and BIX-01294 in MDA-MB 231 cell lines, **d)** Result of homogenous histone deacetylase assay of **14** alongside SAHA in K562 cell lines.

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 (Figure

2.12, see appendix B.1. Figure 0.2, Table 0.1). 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.[113] 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.

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 2.4. 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.

Functional potency evaluation for HDAC inhibition: The enzymatic activity of HDAC was measured in intact cells using the homogeneous cellular assay method.[91] 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 2.5, see appendix B.1.

Table 2.4. *H3K9Me2 cell immunofluorescence In-Cell Western (ICW) assay results (MDA-MB-231 cell line)*

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

Table 2.5. *Results of homogeneous cellular histone deacetylase assay*

Compound	Hela IC_{50} (μM)	K562 IC_{50} (μM)
13	15.33±0.79	27.75±0.59
14	13.80±1.22	5.735±1.23
SAHA	5.044±0.53	2.056±0.59

Table 0.2 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.** Loss of activity in **5b** 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-aminopiperidin ring (R1) along with no substitution at C6.

Molecular Docking Analysis

Molecular Docking analysis and molecular dynamics simulations were widely used to estimate the interactions between protein and small molecular compounds theoretically[114, 115]. 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.[116] 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.[113, 117] Initially we chose HDAC8 protein structure (PDB ID:1T69) for the docking study because it has SAHA (which we used as the control in cell based assays) as

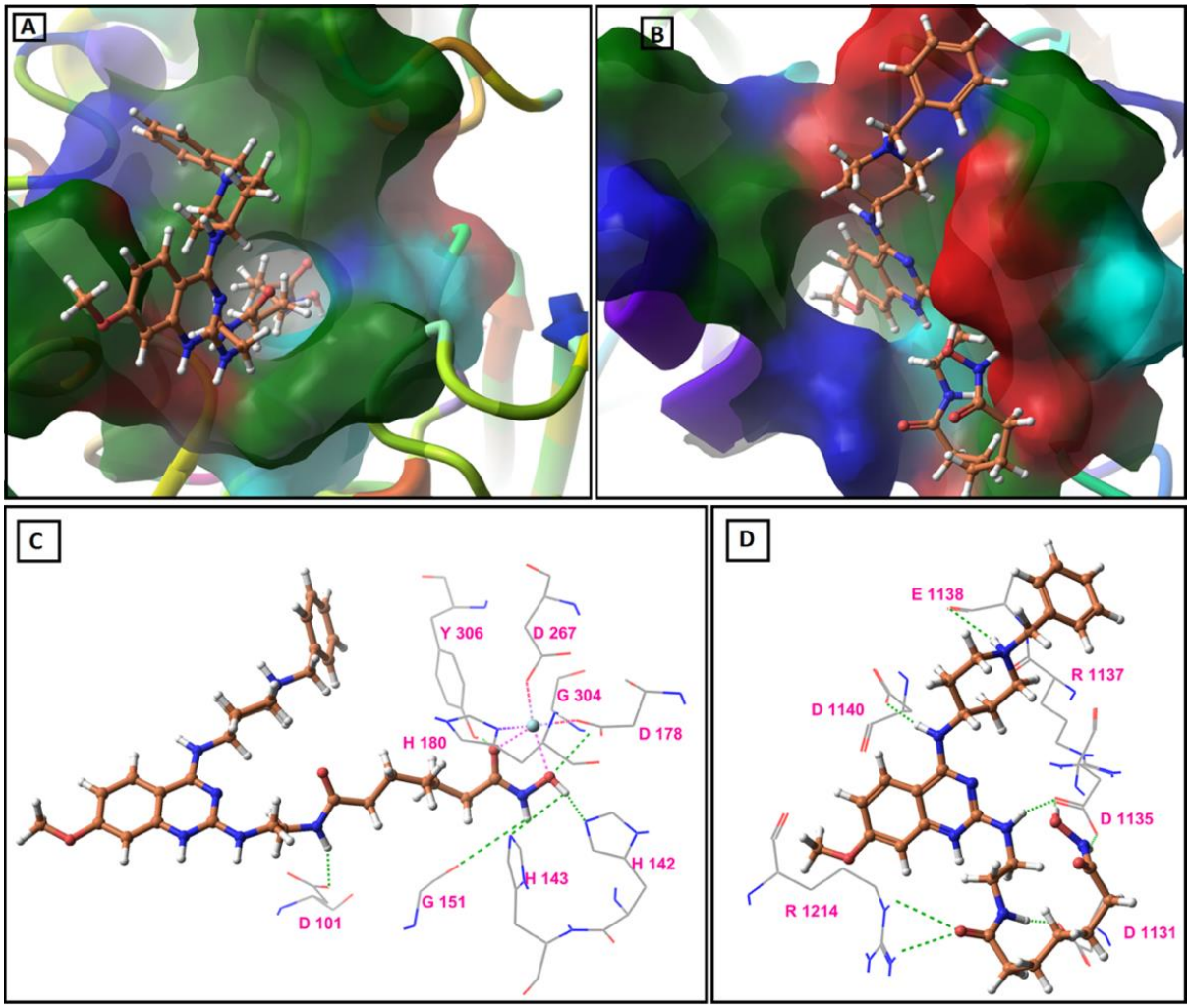


Figure 2.13. Molecular docking study results.
A) Predicted binding mode of compound 14 on HDAC8 (PDB ID: 1T67), B) Predicted binding mode of compound 14 on G9a (PDB ID: 3FPD), C) Binding model of active compound 14 (orange) as revealed from GLIDE docking in the HDAC8 catalytic site (PDB ID: 1T67). The green dashed lines represent hydrogen bonds. The turquoise sphere represents Zn cation and with a trigonal bipyramidal coordination geometry. The pink dashed lines indicate 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).

the co-crystallized ligand, but our study revealed a lower GLIDE score and docking score than the expected (Appendix B.2.), so 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 2.6 and 2.7 shows the results of docking along with the prominent interactions for compound 14 with HDAC8 and G9a. Figure 2.13 illustrated the predicted binding modes and the detailed protein−inhibitor interactions of HDAC8 and G9a with compound 14, respectively. Figure 2.13 A 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 tunnel. The zinc ion displayed a trigonal bipyramidal geometry and two points contacted with the ligand (Figure 2.13 A and 2.3 C). Docking studies suggested important structural/catalytic roles for Gly140, His142, Gly151 and Gly304 in the active site and extending to Tyr306, Asp101. Moreover, compound14occupies the binding pocket in a similar posture as MS-344 in the catalytic site of HDAC8 (Appendix B.2).

Table 2.6. *GLIDE docking results for MS-344 and compound 14 at the catalytic site of HDAC8 (PDB ID:1T67)*

S. No.	Ligand ID	Docking Score	GLIDE score	Interactions		
				Hydrogen Bonds		Interaction with Zn^{2+}
				Backbone	Side Chain	
1	MS-344	-7.931	-7.931	His142, His143, Gly151, Gly304;	Asp101, Tyr306;	Ionic interaction
2	14	-7.934	-8.369	Gly140, His142, Gly151, Gly304;	Asp101, Tyr306;	Ionic interaction

Table 2.7. *GLIDE docking results for BIX-01294 and compound 14 at the catalytic site of G9a (PDB ID: 3FPD)*

S.No.	Ligand ID	Docking Score	GLIDE Score	Interactions		
				Hydrogen Bonds		Interaction with Zn^{2+}
				Backbone	Side Chain	
1	BIX-01294	-7.664	-8.134	Ala1134	Asp1131, Asp1135, Asp1140;	NA
2	14	-7.321	-7.52	Arg1137, Glu1138	Asp1131, Asp1135, Asp1140, Arg1214;	NA

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 (Appendix B.2). 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 (Figure 2.13 B and 2.13 D).

Cell antiproliferation Assay

Table 2.8*. Inhibition of compounds 13 and 14 on the growth of cancer cells and normal cells*

Compound	EC_{50} (μM)			
	MDA-MB-231[b]	MCF-7[c]	A549[d]	HEK293[e]
13	89.33±1.23	79.43±2.72	>100[a]	56.96±1.12
14	10.02±1.66	37.36±2.20	36.24±1.76	19.95±0.19
SAHA	2.874±0.84	8.124±4.98	19.31±1.26	2.482±1.13
BIX-01294	2.155±0.88	8.103±1.99	21.74±2.73	2.048±0.98

[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.

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 2.8, and appendix B.3). 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**.

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.[118, 119] ADMET module of Discovery Studio 3.1 was used to predict physical properties. Using Lipinski's rule of five,[120] the octanol–water partition coefficient (AlogP98) should be less than 5. As seen in Table 2.9, 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 2.9. *ADME prediction results*

Entry	M.W	QPlogS[c]	PSA	PSA-2D[b]	AlogP98[a]
14	515.654	-3.702	161.25	141.462	2.511
SAHA	264.324	-2.139	102.256	81.037	1.838
BIX-01294	476.62	-6.792	50.675	63.249	4.189

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

2.2.3 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 to 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.

2.2.4 Experimental Section

Materials and Methods

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_3OD. 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.

Chemical Synthesis

H_3 (1-20, ARTKQTARKSTGGKAPRKQL): 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/H_2O (92.5:2.5:2.5:2.5) was used to cleave peptides off theresin. 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 were 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 ºC 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 $NaHCO_3$ solution. The combined organic phase was dried over Na_2SO_4 and concentrated under reduced pressure. The residue was purified on silica gel column, eluting with 5% MeOH in DCM (containing 0.5% Et_3N) 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 ºC, ^{1}H NMR (400 MHz, $CDCl_3$) 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). ^{13}C NMR (100 MHz, $CDCl_3$) 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 $C_{23}H_{36}N_6O_4$ $[M + H]^+$, 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 ºC, ^{1}H NMR (400 MHz, $CDCl_3$) 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).^{13}C NMR (100 MHz, $CDCl_3$) 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 $C_{25}H_{40}N_6O_4$ $[M + H]^+$, 489.3145; found, 489.3156.

N2-(2-aminoethyl)-6,7-dimethoxy-N4-(1-benzylylpiperidin-4-yl)quinazoline-2,4-diamine (3b): Brown solid, 2.1 g, 64% yield. M.P. 130-132 ºC, ^{1}H NMR (400 MHz, $CDCl_3$) 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).^{13}C NMR (100 MHz, $CDCl_3$) 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 $C_{29}H_{40}N_6O_4$ $[M + H]^+$, 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 CH_2Cl_2 (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 CH_2Cl_2 (2 ml) was added drop-wise at 0 °C. The mixture was allowed to stir at room temperature and monitored by TLC. Upon completion, the organic layer was washed with saturated aqueous $NaHCO_3$ solution followed by brine. The organic extracts were dried over Na_2SO_4, 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 $C_{27}H_{42}N_6O_5$ $[M + H]^+$, 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 H_2O (0.1% HCOOH)/CH_3CN (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. ^{1}H 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). ^{13}C NMR (100 MHz, MeOD) δ 183.7, 167.5, 148.4, 128.3, 124.8, 124.0, 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 $C_{26}H_{41}N_7O_5$ $[M + H]^+$, 532.3247; found, 532.3248. HPLC purity 95.45% ; t_R= 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. ^{1}H 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). ^{13}C 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 $C_{28}H_{45}N_7O_5$ $[M + H]^+$, 560.3560; found, 560.3554. HPLC purity 95.12% ; t_R= 14.820.

N1-(2-((6,7-dimethoxy-4-((1-methylpiperidin-4-yl)amino)quinazolin-2-yl)amino)ethyl)-N7-hydroxyheptanediamide (5): 38 mg, 30% yield. ^{1}H 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). ^{13}C 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 $C_{25}H_{39}N_7O_5$ $[M + H]^+$, 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)heptanediamide (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 $C_{27}H_{43}N_7O_5$ $[M + H]^+$, 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-hydroxyheptanediamide (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). 13C 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 $C_{31}H_{43}N_7O_5$ $[M + H]^+$, 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 $C_{24}H_{37}N_7O_5$ $[M + H]^+$, 504.2934; found, 504.2911. HPLC purity 96.81% ; t_R= 13.374

N1-hydroxy-N6-(2-((4-((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 $C_{26}H_{41}N_7O_5$ $[M + H]^+$, 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 $C_{23}H_{35}N_7O_5$ $[M + H]^+$, 490.2778; found, 490.2756. HPLC purity 95.61% ; t_R= 12.751

N1-hydroxy-N5-(2-((4-((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 $C_{25}H_{39}N_7O_5$ $[M + H]^+$, 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,[112] 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 $C_9H_8N_2O_3$ was found at 192.0773. Crude quinazoline-2,4-dione and 2.4 g of N,N-diethylaniline 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_2SO_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 $C_9H_6Cl_2N_2$ $[M + H]^+$, 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 ^oC,^1H NMR ($CDCl_3$, 400 MHz) δ 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) δ 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%.^1H NMR (400 MHz, $CDCl_3$) 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%. ^{1}H NMR (400 MHz, $CDCl_3$) 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* = 9.7 Hz, 4H), 3.26 (s, 2H), 2.99 (dd, *J* = 14.9, 7.4 Hz, 1H), 2.90 (s, 2H), 2.09 (d, *J* = 15.0 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%. ^{1}H NMR (400 MHz, $CDCl_3$) 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 (d, 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* = 13.6 Hz, 5H), 6.83 (d, *J* = 11.2 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 $C_{31}H_{43}N_7O_4$ [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, 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 $C_{25}H_{39}N_7O_4$ [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 $C_{30}H_{41}N_7O_4$ [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) heptanediamide (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* = 12.4 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 $C_{24}H_{37}N_7O_4$ $[M + H]^+$, 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* = 10.3 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 $C_{24}H_{37}N_7O_4$ $[M + H]^+$, 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* = 13.0 Hz, 2H), 4.68 (s, 2H), 3.88 (d, *J* = 12.3 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, 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 $C_{23}H_{35}N_7O_4$ $[M + H]^+$, 474.2829; found, 474.2807. HPLC purity 96.40%; t_R= 12.879.

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) δ 8.01 (s, 1H), 7.51 (d, *J* = 18.6 Hz, 5H), 6.76 (d, *J* = 17.4 Hz, 2H), 4.49 (d, *J* = 13.1 Hz, 3H), 3.86 (s, 3H), 3.53 (d, *J*

= 9.5 Hz, 5H), 3.22 (s, 3H), 2.23 (m, 6H), 1.93 (s, 4H). ^{13}C NMR (100 MHz, MeOD) δ 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 $C_{23}H_{35}N_7O_4$ $[M + H]^+$, 536.2985; found, 536.2998. HPLC purity 94.34% ; t_R= 16.051.

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) δ 8.10 (s, 1H), 6.87 (d, *J* = 11.2 Hz, 2H), 4.70 (s, 1H), 3.92 (s, 3H), 3.73 (dd, *J* = 12.9, 6.5 Hz, 1H), 3.66 (m, 3H), 3.46 (s, 2H), 3.35 (d, *J* = 15.5 Hz, 2H), 3.23 (m, 1H), 2.88 (d, *J* = 14.5 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,) δ 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 $C_{22}H_{33}N_7O_4$ $[M + H]^+$, 460.2672; found, 460.2648. HPLC purity 95.90%; t_R= 12.615.

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,[111] which was treated with Pd/C under H_2 gas to get the free amine. HRMS (ESI): *m/z*calcd for $C_{21}H_{32}N_6O_2$ $[M + H]^+$, 401.2264; found, 401.2642. This amine was directly used in procedure C while using monomethyl suberate ester to get 19 and monomethyl pimelate 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* = 9.7 Hz, 3H), 2.83 (d, *J* = 7.7 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 (100 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 $C_{29}H_{45}N_7O_5$[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* = 9.9 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 $C_{28}H_{43}N_7O_5$[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 ethyline diamine 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 monomethyl suberate or monomethyl pimelate 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, $CDCl_3$) δ 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 $C_{16}H_{21}ClN_4O_3$+ [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, $CDCl_3$) δ 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 $C_{22}H_{34}N_6O_3[M + H]^+$, 431.2771; found, 431.2746.

N1-hydroxy-N8-(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) δ 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) δ 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 $C_{25}H_{39}N_7O_4[M + H]^+$, 502.3142; found, 502.3128. HPLC purity 96.21%; t_R= 13.810.

N1-hydroxy-N7-(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) δ 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) δ 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 $C_{24}H_{37}N_7O_4[M + H]^+$, 488.2985; found, 488.2960. HPLC purity 96.80%; t_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 benzyl group was removed using Pd catalyzed hydrogenolysis, mixture of compound 24 (600 mg, 1.2 mmol) and 10 wt% $Pd(OH)_2/C$ (90 mg) in ethanol (100 ml) was stirred

for 40 h 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_2CO_3 (280 mg, 2 mmol), reaction mixture was warmed to room temperature and then at 60 ^oC. 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_2O to get the intermediate ester, which was then dissolved in 2 ml of MeOH and treated with 50% NH_2OH/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_2O (0.1% HCOOH) as eluent. 94 mg, 40% overall yield. ^{1}H NMR (400 MHz, MeOD) δ 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 = 14.1 Hz, 3H), 1.63 (d, J = 8.4 Hz, 4H), 1.48-1.42 (m, 3H), 1.40-1.31 (m, 6H). ^{13}C NMR (100 MHz, MeOD) δ 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 $C_{30}H_{49}N_7O_4$ $[M + H]^+$, 572.3924; found, 572.3925. HPLC purity 96.80%; t_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-(benzyloxy)-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. ^{1}H NMR (400 MHz,

$CDCl_3$) 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). ^{13}C NMR (100 MHz, $CDCl_3$) 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 $C_{27}H_{36}N_6O$ $[M + H]^+$ is 461.9.

4-((1-benzylpiperidin-4-yl)amino)-2-(4-methyl-1,4-diazepan-1-yl)quinazolin-7-ol (28): Aryl demethylation using BBr3 was employed.[121] BBr3 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 room temperature 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 $NaHCO_3$, 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 $C_{26}H_{34}N_6O[M + H]^+$ 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 $C_{35}H_{50}N_6O_3[M + H]^+$, 603.4022; found, 603.4031. Subsequently, compound 29 was dissolved in 2 ml of MeOH and treated with 50% NH_2OH/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_2O as eluent. Fractions collected were concentrated and lyophilized to get brown powder, 54 mg, 23% yield over two steps. ^1H NMR (400 MHz, MeOD) δ 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* = 9.0 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) δ 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 $C_{33}H_{47}N_7O_3$ $[M + H]^+$, 590.3819; found, 590.3832. HPLC purity 96.26%; t_R= 14.192.

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 based on the GLIDE docking score and molecular recognition interactions. All the 3D figures were obtained using Schrödinger Suite 2014-3.

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).

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 minutes 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, a detailed description of data processing is provided in appendix B.1.

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% CO_2 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% CO_2 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% CO_2 in Eagle's Minimum Essential Medium supplemented with 10% fetal bovine serum and 0.5% antibiotic-antimycotic solution.

HDAC Activity Assay

The manual assay was developed by Thomas's group.[122] 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% CO_2). 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 mM Tris–HCl, pH 8.0, 137 mM NaCl, 2.7 mM KCl, 1 mM $MgCl_2$, 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. IC_{50}s were calculated using GraphPad Prizm statistical package with sigmoidal variable slope dose response curve fit.

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 room temperature 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 room temperature, protect plate from light during incubation). After 5 washes 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.

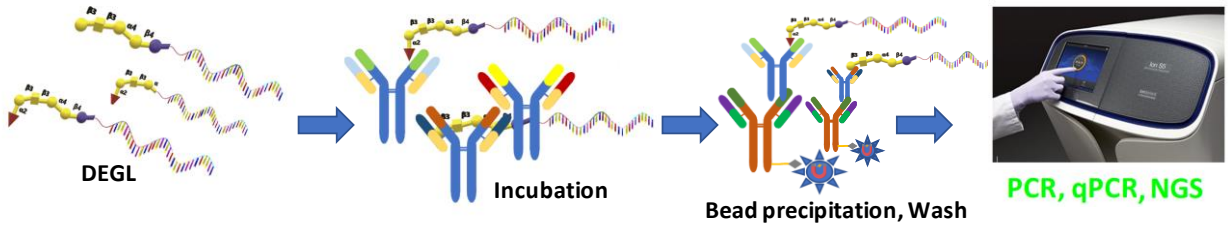
Toxicity Assay

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 ^{3}V 1420 multi label plate reader (Perkin–Elmer, Wellesley, USA). EC_{50}s were calculated using GraphPad Prizm statistical package with sigmoidal variable slope dose response curve fit.

3 CANCER DIAGNOSTICS

3.1. DNA encoded library of Globo series glycans for early stage detection of cancer

ABSTRACT: DNA encoded libraries (DEL) are used with immense success in screening small molecules for drug discovery. The concept of DEL is even more beneficial for glycans since these complex molecules are extremely difficult to detect itself and often require secondary methods. However, the split and pool method used in DEL is not adaptable in glycans because of the synthetic challenges. Alternatively, we propose here a single step encoding of glycans by click chemistry using predefined structure-based DNA code specific to each glycan. DNA-Glycan conjugates then pooled to get DNA encoded glycan library (DEGL) and used in multiplex detection of glycan binding proteins. As a proof of principle study, we synthesized blood glycans and globo glycans and demonstrated glycan specific DNA encoding, conjugation chemistry, selection and detection using PCR, qPCR and Next Generation Sequencing. Selectivity, specificity and sensitivity of the method were also evaluated and compared with the conventional methods. Overall, DEGL provide a highly sensitive solution phase assay, uses femtomoles of sample, and unprecedented throughput and sample capacity when Next Generation Sequencing platforms are used.



Key Words: DNA encoded glycan library, Glyco-PCR, Carbohydrate oligonucleotide conjugates.

3.1.1. Introduction:

Glycans and glycoconjugates are one of the most abundant biomolecules, playing important roles in several biological events such as molecular recognition, adhesion, pathogenesis, and inflammation.[123] Expression patterns of complex carbohydrates are widely altered in cancer, retrovirus infection, atherosclerosis, thrombosis, diabetes, neurodegeneration, arthritis and other diseases.[124] Glycan structures also play a critical role in host defense mechanism as they form a major portion of the antigenic structures, along with other microbe-derived molecules, which are recognized by host immune cells and mount an immune response.[125] Hence, the structural consequences of the interactions of oligosaccharides with receptors of interest are one of the major focuses of pharmaceutical research now. [124, 126]

Nature has provided a systematic, template driven, coded system that has greatly facilitated scientific understanding of nucleic acids and proteins. Similar natural organizing principles do not exist for the incredibly complex myriad carbohydrates (glycans), which also lack UV/ fluorescent properties hence fallen behind. Furthermore, even with the recent advances in solid phase automated synthesis and enzymatic synthesis which eased the difficulties in obtaining a variety of glycans, materials available for experimentation remain extremely precious, especially for complex glycans. Hence, the development of a highly versatile and ultra-sensitive detection technology is particularly important in the field of carbohydrate research for the characterization of glycan-protein interactions.

Microarrays and ELISA are the most used methods in glycomics now although with their limitations; microarray has the disadvantage of requiring specific instrumentations and not mimicking the biological environment while ELISA is not suitable for high throughput analysis.[125, 127-129] In another pioneering work named multiplex glycan bead array (MGBA)

by Jin-Xiong She's group in collaboration with our group, recently reported the use of colored Luminex beads for the multiplex detection of glycan binding proteins,[130] which is certainly improved in magnitude and throughput, but still require instrumentation specifically for this use. Alternatively, we assumed DNA-encoded glycan libraries (DEGLs) as a potential replacement or adjunct to the existing technologies. Encoding a small molecule with a piece of DNA enhances the detection limit of the small molecules many-fold by polymerase chain reaction (PCR).[131] Although the concept of DNA encoding was introduced in 1992, it has gained momentum recently with the advancement in DNA next generation sequencing (NGS) technologies and involved in many of the discovery of new lead molecules.[132-134] The technology also referred as DEL, generally adopt split and pool library synthesis, which is practically suited for the construction of libraries spanning millions to billions of small molecules.[135] However, the split and pool synthesis which heavily depend on the aqueous phase coupling reactions is not suitable in the highly complicated glycan synthesis. Hence, we need an alternate approach for the affordable use of DNA encoding in glycomics. In this work, we describe an affordable approach to use the enormous potential of DNA encoding in functional glycomics via structure-based DNA coding of glycans and single step click conjugation of DNA code and glycans.

Kwon and coworkers showed the feasibility of using DNA tag in glycomics; their work coined Glyco-PCR, successfully demonstrated detection glycan-binding proteins (GBP) using DNA conjugated glycans with high sensitivity.[136, 137] Although, the method asserted the ultra-sensitive detection by DNA amplification, it hasn't addressed how a large pool of glycan is analyzed with an affordable cost. More recently, DNA encoded library involving glycan were synthesized by biocatalysis and split and pool,[138] again it is limited to very limited number of glycans. However, the glycosyltransferases enzymes available to facilitate such an approach is

limited, which means the DNA encoding of glycan via split and pool will be restricted to the enzyme specificity. To overcome this and enhance the scope of DNA Encoded Glycan Library (DEGL), single step tagging of glycans previously synthesized by enzymatic or chemical methods with a relatively big DNA fragment remains the only option. Though custom synthesis of such DNA sequences is possible and pragmatic for small DEGLs, it would be a cost-prohibitive option when dealing with hundreds of sugars, and hence compromising on the enormous potential of high throughput analysis. These challenges are certainly inevitable but can be minimized if we could develop a broadly acceptable protocol exclusively for glycans, from the synthesis of DEGL to final data analysis. A universal protocol will allow manufacturers to catalog DNA codes for different glycans and sell at a lower price. This is particularly reasonable in glycans since the number of biologically significant glycans characterized so far is less than twenty thousand.[139] It is worthy to note here that, cataloging DNA codes for glycans can also achieved by selecting random DNA codes, but we wanted to use the full potential of what DNA manipulation has to offer in functional glycomics; which is only possible when DNA code is somehow corelated to the glycan in question. For instance, structurally defined DNA code will ensure minimum errors in selection and sequencing, since it is more likely to have same glycan epitopes selected against a target than a totally different glycan structure. The same principle can also be applied to other closely related technologies like probe-based assays and DNA arrays. In order to the ubiquitous use of DNA codes, we require systematic selection of DNA codes, highly feasible DNA-glycan conjugation chemistry, and well-characterized detection procedures. Here, we address all three issues, firstly, an in-house software program was developed (available at http://131.96.145.142:8000/cgi-bin/form.py) for generating a DNA code specifically for each glycan. Next, the DNA code amended with primers and 5’ hexynyl modification was 'click' conjugated with the corresponding azido-glycan,

and finally, selection and detection methods were optimized. As a proof of principle application study, DEGL of Globo series glycans were synthesized and applied in detecting cancer-specific antiglycan antibody VK9. Certain Globo-series glycans are overexpressed in some cancers and proven as biomarkers for the early detection of breast cancer and ovarian cancer.[25, 140-142] Recent developments in the Globo-H based immunotherapy also require highly sensitive methods to quantitatively detect the antiglobo H antibodies for monitoring therapy and in clinical trials.[143-145]

3.1.2. Results and Discussion

3.1.2.1. Systematic coding of glycans

Our first aim was to develop a unique DNA coding method that would be consistent with the existing representation of glycans, so that any glycans known or yet to be characterized can get a unique name in DNA terms. Unlike peptides, proteins, and genetic materials, representation of carbohydrates brings many challenges. A detailed and accurate depiction of glycans should feature composition, sequence, branching position, possible modifications, and anomeric configuration. Kornfeld et al. first proposed the use of symbols to depict carbohydrates,[146] and this approach has since replaced the IUPAC naming,[147] with modifications adapted to fit the entire glycome.[148, 149] All these approaches are based on different shapes and color codes to represent the building blocks (Figure 3.1), which gave us the idea of using nucleotide bases to represent the building blocks. The number of shapes and colors we could use for the representation are infinite, but only four nucleotide bases A, T, G, and C available for DNA coding. Hence, we depended on permutation and combination approach to cover the whole glycome. The number of sugar monomers are more than one hundred, and they can be incorporated into the oligosaccharide in multiple ways, making glycan structural information more complex. These inherent difficulties

prompted us to develop a dedicated program that could sort the structural information to coding vocabulary and transform IUPAC names to single-stranded DNA code. For natural adaptation of the codes into functional glycomics, the codes should be of minimum length but still code all the information and enabling PCR amplification.

Structure-based coding must have a controlled vocabulary of the structural components. For attaining this goal, similar to the Consortium of Functional Glycomics (CFG) representations, initially, we split the structural information of the carbohydrate into the components and the structure (Figure 3.1). Then we gathered all monomers into library A and different linkages to library B. Next, the monomers in the library A assigned with a unique four-letter code. There are 256 possible combinations with the four letters A, T, G, and C. More than 70 monomers were-

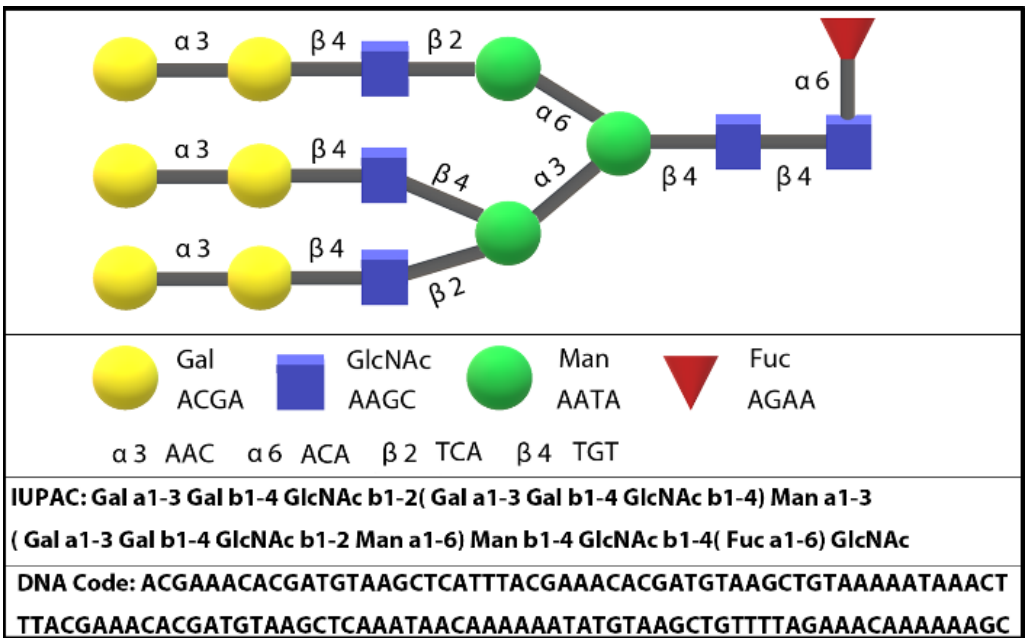


Figure 3.1. Different representation of glycans

added to this library (Appendix C.1, Table 0.5), effectively using about 70 randomly-generated codes. Similarly, all linkages library B were coded with 3 letters. Thirty-six linkages are currently in this library with further space for future modifications (Appendix C.1, Table 0.6). Many glycans

feature functional group modifications as part of the structure, and to add this information we added a third library (library C, Appendix C.1, Table 0.7). More than 100 modifications were identified and added to the library and assigned four letter codes which were not used in the library A. Any special characters like '(' used in the IUPAC input were also incorporated with specific codes. Finally, we consolidated these libraries into a Python-based program, and several rounds of evaluations performed with diverse glycan structures. A sample library of 100 glycans was tested, and the program successfully delivered codes for all the glycans tested (Appendix C.1, Table 0.9). A sequence length between 25 to 120 bases ideally fit the PCR applications and, among the 100 glycans randomly selected most of them fell in this limit, but few glycans had sequence length smaller or higher than the range specified. Through the careful observation of the sample pool, we identified many of the repeating fragments in the glycans (e.g., six sugar core structure of N-glycans). Hence, we assumed taking this as a single block would reduce the overall length of the codes of long glycans, and a fourth library (library D, Appendix C.1, Table 0.8) was added to the program featuring many of the most common building blocks found in the glycans. For the glycans possessing a code smaller than the 25 bases, a random sequence of 20-base DNA was added to the code with a special mention in the program to avoid misinterpretation. These two modifications allowed us to keep the sequence length within the desired limits.

3.1.2.2. Glycan codes to functional glycomics

Once a reliable DNA code was established, we synthesized the blood group antigens A, B and O with an azide linker by adopting chemo-enzymatic strategies. These glycans then conjugated to the 5'-hexyne modified DNA codes (IDT, Coralville, IA) corresponding to the antigens via click chemistry to get the glycan-DNA conjugates for the initial studies (Figure 3.2a, Experimental section Figure 3.8). Blood group antigens were chosen because of two reasons, the abundant

antibody presence, and any specificity and selectivity studies could be achieved easily with these structurally similar glycans.

The success of the DEGL lies in the utility of applying the DNA codes to the standard PCR and qPCR protocols. It is crucial to have the glycan-coupled DNA (G+DNA) achieve similar PCR efficiency to the native DNA, so we compared the G+DNA conjugates with the corresponding native DNA strands to verify the efficiency of amplification in both PCR and qPCR. PCR was performed on DNA and G+DNA templates using standard thermal cycles with varying annealing temperatures. Gel electrophoresis analysis implied both DNA and G+DNA amplified well in the range of 52-60 ºC (Experimental section, Figure 3.7). Next, we examined the qPCR detection limit of the G+DNA conjugates; both DNA and G+DNA conjugates were serially diluted to provide different concentrations of templates, with final concentrations ranging from 2.4 nM to 16 pM. Standard curve qPCR was carried out to obtain the corresponding critical threshold value (Ct), which is used to compare the concentration of the templates in the qPCR reaction mixture, with a low Ct indicating high template concentration. Ct values of the conjugates from 2.4 nM to 16 pM concentrations were observed in the range of 5-25, while the negative control (no template control, NTC) gave a Ct value above 30 (Figure 3.2b). A standard plot of the Ct value *vs* log concentration of the DNA and G+DNA conjugate fit well within a linear relationship indicating the successful application of G+DNA for quantitative detection up to *pico* molar level (Experimental section 3.13).

Encouraged by the result obviously suggesting the detection of glycans from even a picomolar level, we proceeded to the final goal, screening of glycans against glycan binding proteins. DEGs can be used either as a screening tool against the known concentration of proteins or a detection tool to identify glycan-binding molecules. We did two sets of experiments for validating these two

aspects of DEGs; initially, we demonstrated the detection of specific glycan binding proteins by using the familiar blood glycan-antibody interactions. All three G+DNA conjugates (G1+A DNA, G2+B DNA, and G3+O DNA, with G1, G2 and G3 representing A, B, and O glycan antigens, respectively) were interrogated with the commercially-procured human blood antibodies (IgM). Concentrations of 5 µM, 2.5 µM, and 1 µM of the G+DNA were incubated with 2 µL of the antibodies; unbound DNA was eliminated either by filtration with 100k cut off centrifuge tubes or precipitation with magnetic protein L beads. The filtrate/eluent hence obtained then subjected to qPCR assay, and Ct values of the G+DNA then compared with the Ct values of negative control wherein antibody was omitted from the incubation. Ct difference of more than 10 points was observed at all three G+DNA concentrations (Experimental section, Figures 3.14-3.17), indicating the successful application of DEG in screening.

In another experiment, we evaluated the use of DEGs in the detection of glycan-binding proteins, a mixture of all three antibodies (1:1:1) interrogated with each of the G+DNA conjugates (2.5 uM) separately. As seen in Figure 3.2c, Ct values of 8.4 and 7.7 respectively for G1+A DNA and G2+B DNA infer the high affinity binding of A and B glycans with their corresponding antibodies Ab A and Ab B. But, a similarly tested G3+O DNA showed relatively higher Ct value but still less than that of the no antibody controls indicating weak interaction of antigens with the Ab O (Figure 3.2c). This result implies that DEGs are ideal for the selective detection even from a mixture of closely related target proteins. We also tested the specificity of the method, for this aim, we investigated a single G+DNA (G2+B DNA) against the three antibodies, and a clear amplification was visible only when both Ab B and G2+B DNA were in the mix. All the remaining tests were poorly amplified (Figure 3.2d). This result proved that the amplification was indeed from the target

specific binding of glycan, and there was no significant interference of nonspecific binding of DNA and target proteins.

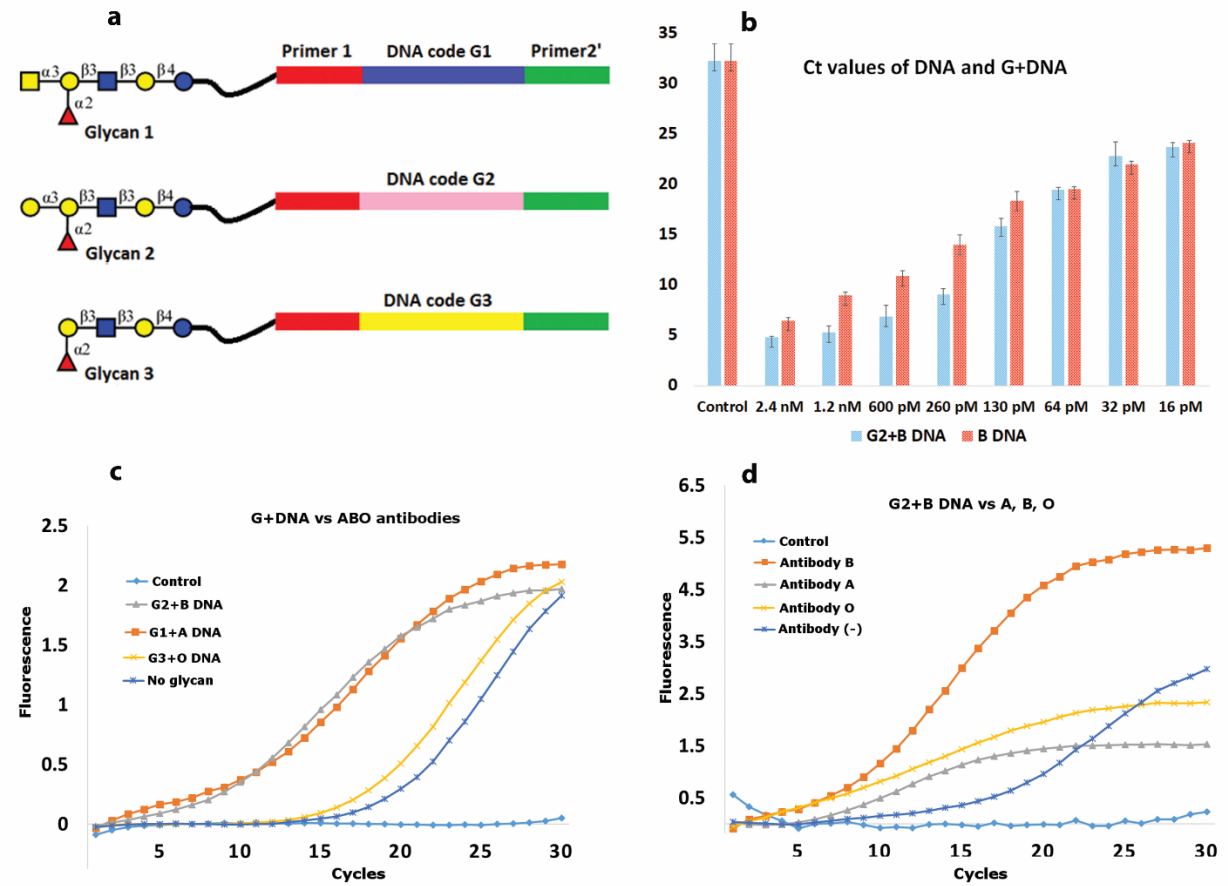


Figure 3.2. DNA encoded blood glycans and their detection by qPCR.
a) DNA Encoded Blood Glycans, b) qPCR limit of detection comparison of glycan DNA conjugate and pure DNA, c) Each G+DNA conjugate selected against the mixture of antibodies composing equal concentration of Ab A, Ab B and Ab O, d) G2+B DNA selected against each antibody separately.

3.1.2.3 DNA Encoded Glycan Library of globo series glycans and multiplex analysis using Next Generation Sequencing

After successfully demonstrating blood antigens selection and amplification by PCR and qPCR, we wanted to apply the technology to the next generation sequencing platforms. Hence, we designed a DNA Encoded Glycan Library (DEGL) of globo series glycans (Figure 3.3). Globo-

glycans are characterized as glycosphingolipids, and their expression pattern is well correlated with cancer metastasis and progression. Aberrant expression of immunogenic globo glycans triggers the production of the anti-glycan antibodies targeting these epitopes. These antibodies are generally observed in the sera even before cancer progresses to the late stages and effectively act as biomarkers for certain types of cancer including breast, lung, prostate and ovarian cancer.[26, 150, 151] Globo-H and its truncated analogs Gb5, Gb4, Gb3, and Gb2 were synthesized from the enzymatic extension of chemically synthesized Gb-2. To test the terminal epitope specificity alone, Bb4, Bb3 and Bb2 were also synthesized via chemo-enzymatic strategy (Procedure is described in experimental section). All these glycans were then conjugated to the corresponding DNA codes via click chemistry as described previously.

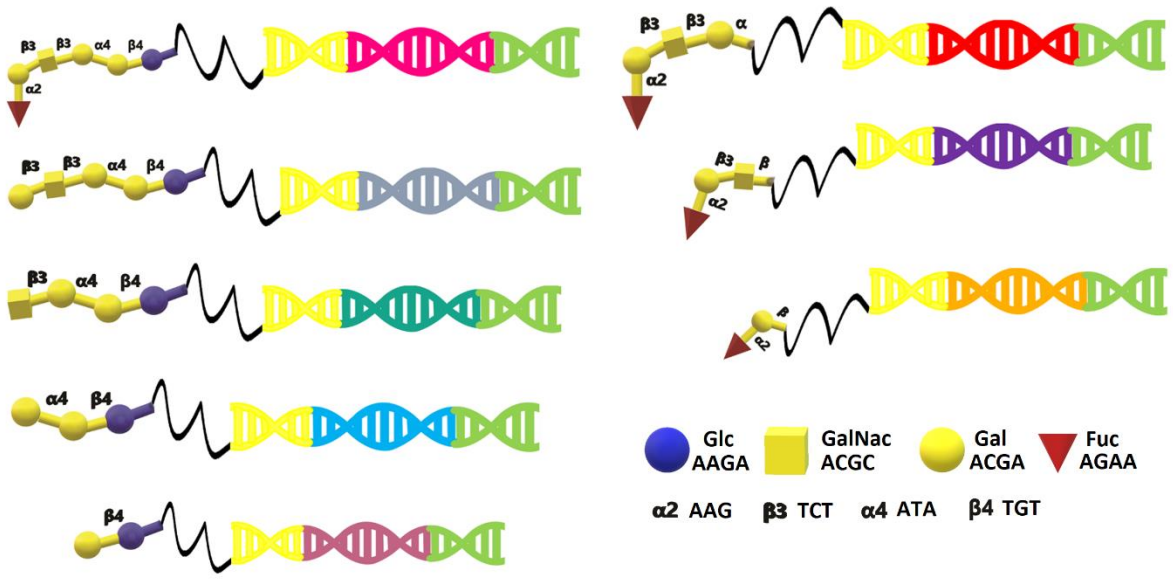


Figure 3.3. Globo glycan structure and representation of globo-glycan conjugates.

ELISA of globo series glycans and DNA Encoded Glycan Library of globo series glycans

Initially we validated the selected system with the conventional ELISA to confirm the antibody binding capacity and specificity of the glycans. Since sugars have low binding affinity

for unmodified plastic surfaces, the binding protein, streptavidin, was used to capture biotinylated sugars (synthesis is given in experimental section). Simultaneously, we also tested DEGL using ELISA to see the effect of different presentation of glycans. For this assay, we coated the DEGL onto plastic surfaces directly by the help of Pierce DNA coating solution. Both naked glycan and DEGL was tested against VK9, which is a mouse IgG anti-Globo H monoclonal antibody. In the ELISA of biotinylated sugars, VK9 has very high binding intensities to GbH and Bb4 and very low binding to Bb3 (Figure 3.4a). However, DEGL ELISA also showed week binding for Bb2 along with the strong binding with GbH and Bb4 (Figure 3.4b). This result bolstered our earlier observation that immobilized glycans on a surface may not provide accurate information of the binding. We rationalize this difference in pattern to the different presentation of glycan epitope, all of the four glycans have the same terminal epitope of Fuc a1-2 Gal(b1-3 GalNAc), which is recognized by VK9 antibodies,[152] but a long DNA tag on the DEGL will give more space to the glycan antigens binding with the antibody, while normal ELISA would block the binding of the short glycans of Bb3 and Bb2. This kind of interferences will be further negligible when a DNA encoded glycan library is used, because the antigen-antibody interaction takes place in a more biologically significant solution phase.

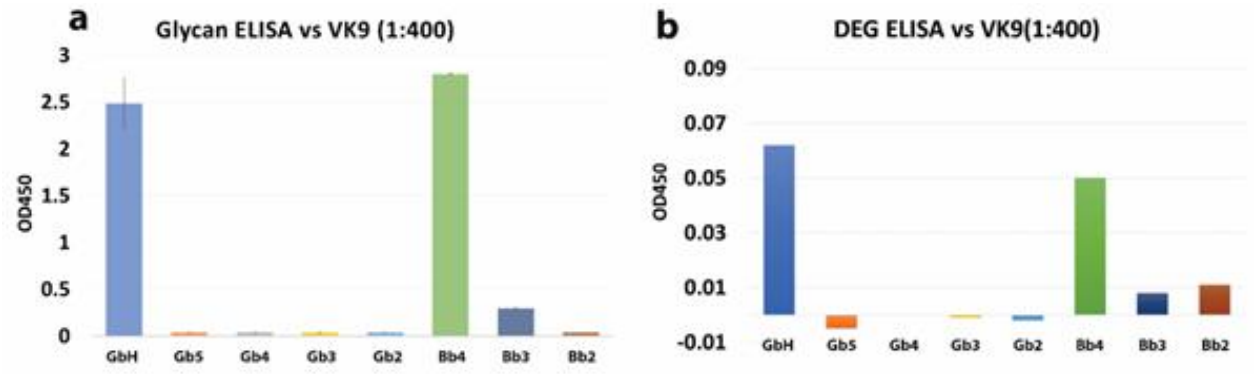


Figure 3.4. ELISA detection of glycans and glycan-DNA conjugates.
a) ELISA of glycans conjugated with biotin immobilized on avidin coated surface against VK9.
b) ELISA of glycans+DNA codes immobilized via pierce DNA coating solution against VK9.

qPCR detection GbH against VK9 antibody

After validating the DEGL, we tested the GbH+DNA with the monoclonal anti globo-H antibody VK9 using the protein A/G protocol, and the qPCR results for two different concentrations (2.5 µM and 5 µM of GbH+DNA incubated with 2 µL of VK-9 antibody) are given in the Figure 3.5. As expected, a concentration dependent binding was observed as indicated by the Ct values (7 and 19 for 5 µM and 14 and 21 for 2.5 µM).

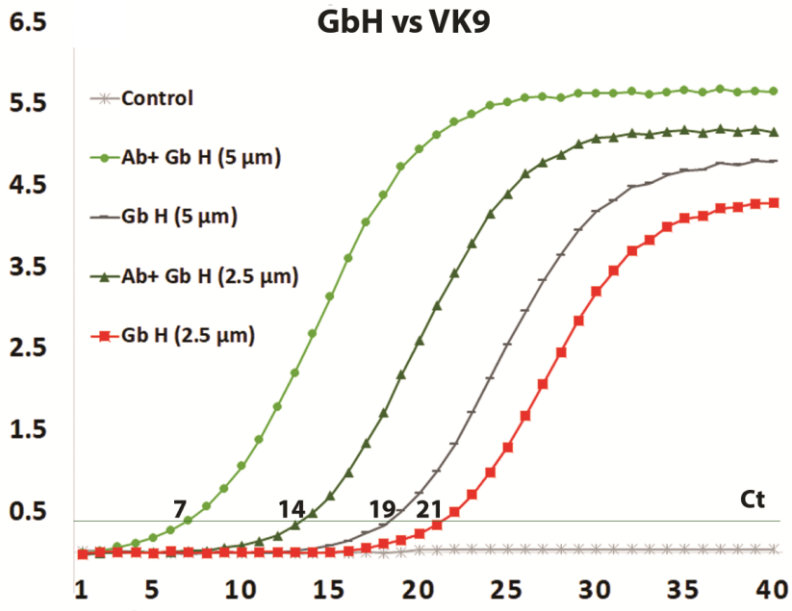


Figure 3.5. qPCR detection GbH against VK9 antibody.

Detection of glycan binding antibodies present in the cancer plasma

After gaining the confidence from the globo-H and VK-9 interrogation studies, we conducted an elaborate study using all the globo series glycan against two breast cancer plasma samples (sanguine bioscience BC1 and BC2) and compared the Ct values of each glycan conjugates with the sample treated with both healthy serum (HS) and no serum (NS). The results show that only GbH, Bb4 and Bb3 had significant amplification with the Ct values less than 8,

while other glycan conjugates show relatively less amplification (Figure 3.6). These results are consistent with the micro array based methods reported by Chi-Huey Wong group.[26] Though this approach accurately differentiated the cancer plasma from the healthy plasma based on the presence of glycan binding antibodies, it is not convenient to test every glycan conjugate separately and it is not practical when several plasma samples are needed to be tested. But these difficulties can be easily overcome by employing next generation sequencing (NGS), which allow multiplex detection of the glycan conjugate from a large cohort of plasma samples with one additional step of barcode indexing. Hence NGS technology allow the high throughput detection of hundreds of plasma samples with even thousands of glycans in the DEGL.

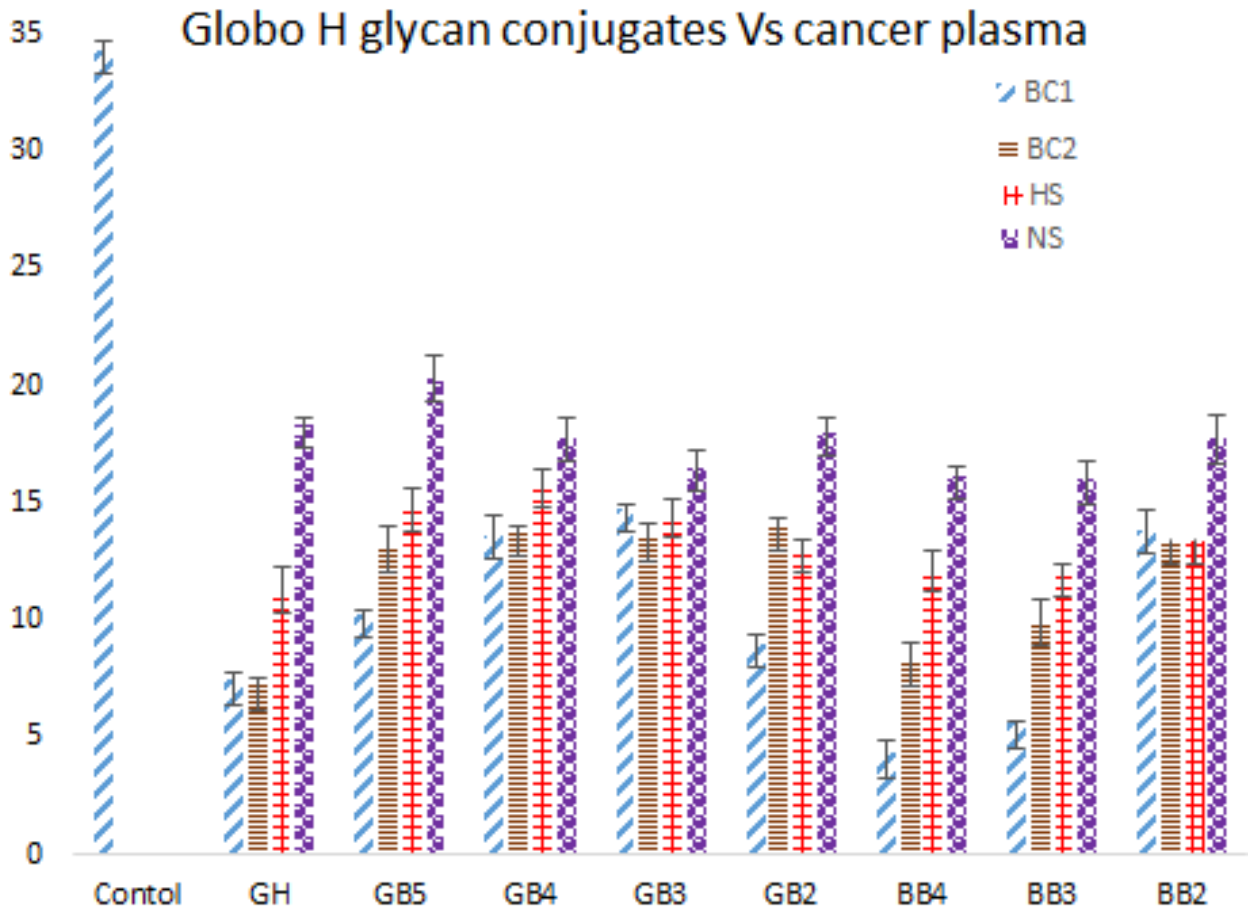


Figure 3.6. Ct value plot of different globo series glycan conjugates interrogated with breast cancer plasma (BC1 and BC2), health serum (HS) and no serum (NS).

NGS experiments:

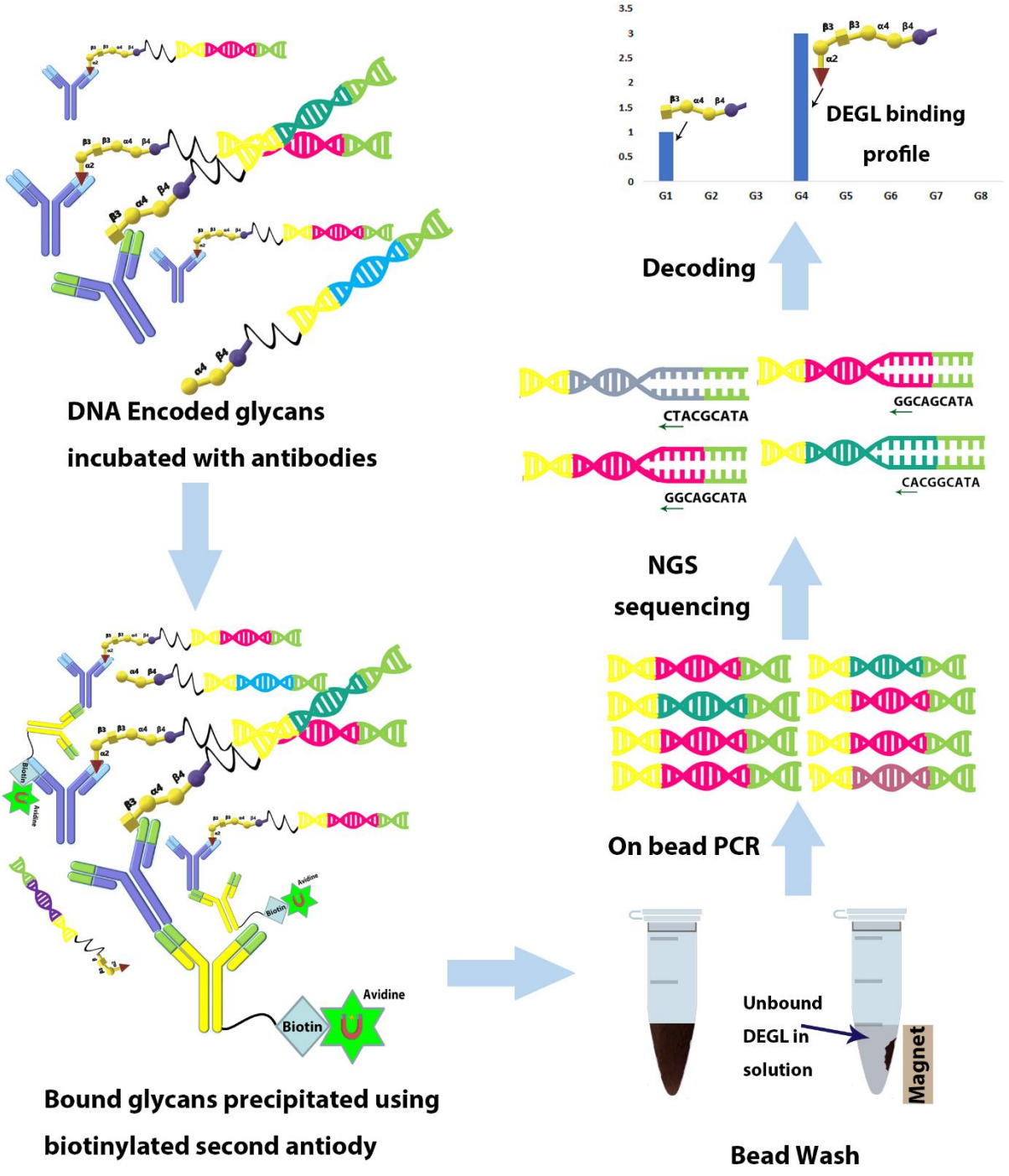


Figure 3.7. Protocol for the multiplex detection of glycan binding to the target.

Next, we wanted to translate the results to more promising NGS based parallel screening. Initially we tested the different concentrations of DEGL with different concentrations of VK9 antibody using the PCR and qPCR assays. We incubated 1 uM, 500 nM, 100 nM and 10 nM samples with 0.5 µL, 1 µL, and 2 µL of the VK9 antibody, and unbound glycans were washed out after precipitating the bound glycan-VK9 complex on protein A/G beads. Each assay was compared to the same amount of DEGL with no VK9 negative control. Then, washed beads were used as the template for the NGS analysis. Briefly, beads were diluted to 20 µL, and from this, 2 µL was used for the initial PCR amplification (18 cycles). PCR products were then purified using AM pure beads (Beckman Coulter) according to the manufacturer's protocol and used for the NGS fusion PCR to incorporate the index codes and ion-torrent NGS adapters. Now about 150 bp DNA codes, it was again purified using AM pure beads, and all barcoded samples were run through an Agilent Bioanalyzer for quality control. Samples were then pooled in equimolar concentration and adjusted to the desired 26 pM for the NGS, analyzed using ion-PGM semiconductor sequencer. Data were analyzed using 'FASTAPTAMER' software, which is a custom designed software for the analysis of sequence count and read per million of aptamers after selection.[153] Unfortunately, the sequencing results (note provided) were not good enough to discriminate the antibody samples with the no antibody samples. We assumed this could be due to the high sensitivity of the sequencing method, and even the smallest of nonspecific DNA binding to the protein A/G beads (Ct values from qPCR suggest minor nonspecific binding) was affecting the overall performance of the assay. Hence, we decided to skip the procedure involving the beads and instead used biotin tagged secondary antibody specific to VK9. After incubation, the complex of DEGL+VK9+anti-VK9-biotin was separated using the magnetic streptavidin beads and washed several times to remove any unbound DEGL from the mixture (Figure 3.7). Downstream processing of beads was

performed as described above, and we compared the relative enrichment of all the library members in comparison with the no VK9 negative control. As shown in Figure 3.8, only Globo-H, the antigen specific to the monoclonal antibody VK9 and its truncated analog Bb4 is enriched compared to the rest of glycans indicating the efficiency of NGS method in detecting the glycans from a mixture. Even though the results were consistent over several runs, there was some random amplification in no serum controls as seen in the case of BB2 in figure 3.8. We assumed this might be due to the nonspecific binding of DNA to any of the components in the selection procedure, for instance binding to the secondary antibody or magnetic beads.

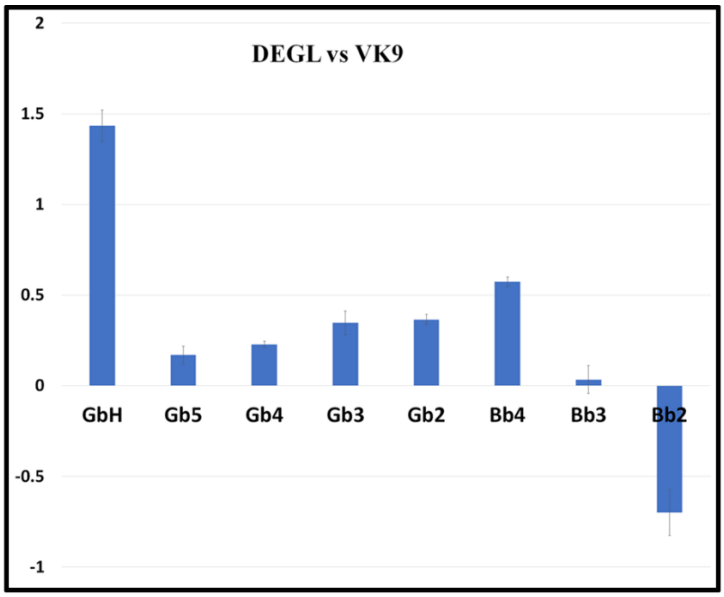


Figure 3.8. Multiplex detection of DEGL against VK9 antibody using NGS method.

Single stranded DNAs has the tendency to form the secondary structure, which may be the cause for binding like aptamers. To minimize this, we then hybridized the DNA conjugated glycans with their complimentary strand ssDNA to get the dsDNA, double strand DNA forms the helix and generally void of any binding. Next, dsDNA glycan library is incubated against the VK9 and selection and sequencing has done according to the same procedure. Yet, the random

amplification in the negative control was still observed but this time the amplified DNA was Gb3 DNA (result not shown). The random amplification of two different sequences in two experiments indicate that DNA might be selectively amplified after the selection due to any PCR bias and not because of nonspecific binding we initially assumed. To solve this, we added a control sequence immediately after the selection to all the samples, which will act both act as an internal standard and a supplemented template in negative control. With this modified method we further carried out the selection and sequencing of Globo-H series DEGL against the VK9 antibody. The results shown in the figure 3.9 clearly indicate the selective enrichment of GbH glycan along with the BB4 glycan in a lesser extent.

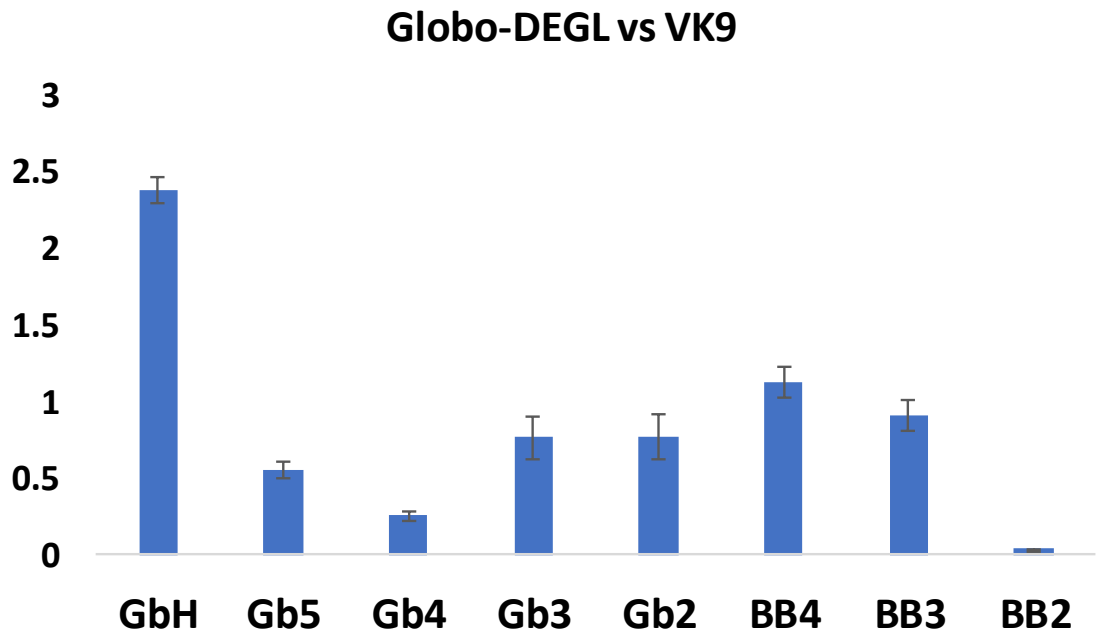


Figure 3.9. Selection of globo series glycan library against VK9 antibody. Graph is plotted as the ratio of enrichment of each DNA sequence of DEGL treated with antibody against no antibody control in triplicate (Error bar is of standard error).

Detection of natural antiglycan antibodies

Antiglycan antibodies are one of the major class of naturally occurring antibodies in the serum. Several antiglycan antibodies are known against the tumor specific glycans and glycans associated with autoimmune diseases, and there is a notable interest in pursuing them as potential biomarkers for such diseases.[154-156] To demonstrate the feasibility of detecting naturally occurring antiglycan antibodies via DEGL and NGS method, we further choose the blood glycan system which we earlier validated using PCR and qPCR. DEGL library was expanded with the addition of blood A and blood B glycans and selected the 10-glycan library against various blood groups. Figure 3.10, shows the selective enrichment of blood A antigen signal because sera of people having B blood group comprises substantial number of natural antibodies against the A antigen but no antibodies against the antigen B.

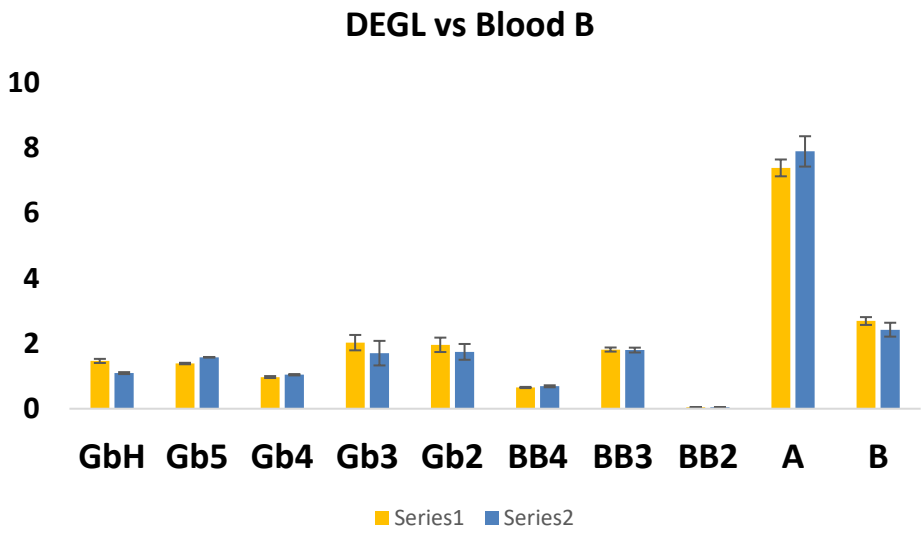


Figure 3.10. Multiplex detection of DEGL against natural antiglycan antibody of blood B antibody using NGS method. Graph is plotted as the ratio of enrichment of each DNA sequence of DEGL treated with antibody against no antibody control in triplicate (Error bar is of standard error).

Comparing the different methods available to study the functions of glycans are not easy, they are largely dependent on assay platform, glycans and GBPs used in the assays and several other parameters like glycan presentation, incubation time, buffers and wash time etc.[129] Even the variability is significant in same platform technology, as of the glycan arrays from different group with different fabrication and assay protocol.[157] Our findings with the DEGL involving PCR, qPCR and NGS were positively corelated with the existing methods in selectivity, specificity and sensitivity (Table 3.1). DEGL requires femtomoles of glycans for the multiplex detection of glycan specific antibodies in the NGS assays, which is almost same as the glycans required in the microarray method which uses the minimum sample among the various methods.[129] Our method's relative low sensitivity to the other methods in the table is most likely due to the monovalent presentation of glycan to the binding proteins unlike the multivalent presentations in ELISA, glycan arrays and MGBA. But, such presentations of glycan can be easily attained with the DEGL and hence a higher sensitivity could be attained.

Most notable feature to highlight with the DEGL is its potential as method of high throughput and high capacity in sample handling. Moreover, unlike the ELISA, Glycan array or MGBA which are mono-dimensional and reads the optical density or florescence of the dye conjugated secondary antibody for the glycan detection, DEGL is multi-dimensional in glycan detection. DNA encoding brings tremendous possibility in detection and selection procedures as evidenced by the diverse methods opted by various research groups for the selection of small molecules.[158] Apart from the simple PCR, qPCR and NGS approaches, many groups have developed technologies that can detect binders specifically for each proteins from a mixture of targets or even from the cell lysates.[159, 160] Billions and trillions of compounds are screened against target proteins for drug development purpose, hence the number of glycans in the DEGL

is virtually unlimited.[135] On another side, with the proper selections of barcoding and sequencing platforms it is possible to screen about 800 assays in a single run. Also, fabrication of DEGL is simple chemical reaction widely used by chemist and biologists and very flexible to incorporate the multivalent presentation of glycans, which may be critical for many weak or medium binding glycans. One drawback of the assay is the time taken from the incubation to the sequencing result; it requires a full day of preparation before placing the sample for sequencing, which take almost five hours for the delivery of data. All together the DEGL screening take about two days to complete and it may be not suited for the quickest results, especially in certain clinical application. However, we think our glycan specific coding based on the structure could facilitate the use of multiplex qPCR using several probes for the diagnostic and vaccine monitoring.

Table 3.1 *Comparison of different platform technologies[26, 129, 154]*

	DEGL	ELISA	MGBA	Glycan Arrays
Glycan*	100 fmol	2500 fmol	1000 fmol	60 fmol
Antibody**	500 ng	10 fmol/mL	250 fmol/mL	700 fmol/mL
Serum Dilution***	1:20	1:400	1:500	1:500

*from references 26 and 129, ** reference 129, *** from references 26 and 154

3.1.3. Conclusion

Advancement of the technologies in DNA synthesis and sequencing has extended the scope of DNA beyond genetics. For example, DNA encoded libraries of small molecules in drug screening and lead optimization is a priority of research in major pharmaceutical companies. We recognized the importance of applying DEL technologies to functional glycomics, because this field is otherwise restricted by low sensitivity detection techniques and extreme difficulty in obtaining samples. Applying the highly sensitive PCR based detection will enhance the reach of functional

glycomics to sub-nanogram samples and allow tandem high throughput screening. For the encoding purpose, we described here a unique method to code each glycan, analogous to the codons of protein synthesis, which accounted for extensive structural information. The program which is available (http://131.96.145.142:8000/cgi-bin/form.py) for the use of the research community is aimed at providing a uniformity in the codes for every single glycan in future DEGL applications. A structure specific code also helps researchers to develop target specific qPCR probe to analyze the binding of specific glycan or a group of glycans having the same structural motifs without requiring a NGS. We also elaborated methods for the DNA-Glycan conjugation and various selection protocols and information retrieval. Finally, as a proof of principle application study, we synthesized DEGL consisting of the globo series glycans and demonstrated their application in detecting the antiglycan antibodies. Looking forward to the future, we believe, DEGL platform described here would give access to many researchers across the globe to use DNA code to decipher plethora of information regarding how glycans talk in biological significance. To lead the effort, we are extending the scope of our collection of DNA encoded glycans to several important classes and make them available for other researchers. This way with the contributions from several groups, it would probably pave the way for next generation of functional glycomics.

3.1.4. Experimental section

Materials

All chemicals and biological reagents were purchased from Thermo Fisher unless otherwise mentioned. Maxima SYBR Green/ROX qPCR Master Mix, Platinum Pfx DNA Polymerase were purchased from Life Technologies (Carlsbad, CA). Micro Bio-Gel P-30 Chromatography Columns were purchased from Bio-Rad (Hercules, CA). Tris[(1-Benzyl-1H-

1,2,3-Triazol-4-yl) methyl] amime (TBTA) Click chemistry Ligand were purchased from TCl (Tokyo,Japan). MicroAmp 96 well Fast PCR Reaction Plate, MicroAmp Optical Adhesive Film, MicroAmp Fast Reaction Tubes, Strips were purchased from Applied Biosystems (Foster City, CA). All oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA). UDP-GlcNAc, UDP-GalNAc, UDP-Gal, and GDP-Fucose were prepared using one-pot multienzyme system as reported previously.[161, 162] β1–3-*N*-acetylglucosaminyltransferase (LgtA) and β1–4-galactosyltransferase (LgtB) from *Neisseria meningitides* were prepared as reported previously.[163, 164] β1–3-*N*-acetylglucosaminyltransferase (LgtD) and α1–4-galactosyltransferase (LgtC) were prepared as reported previously.[165] α1,2-FucT (HmFucT) from *Helicobacter mustelae*[166] was cloned into pET-28a vector and expressed in *E.coli* BL21 (DE3). β1-3-*N*-acetylgalactosaminyltransferase (GTA) and β 1-3-galactosyltransferase (GTB) from human[167] were cloned into pET-28a vector and expressed in *E.coli* BL21 (DE3). The purified proteins were concentrated and desalted with 10 kDa molecular weight cut-off (Millipore, MWCO) spin filters for further use. GloboH-related oligosaccharides were prepared by following the previous protocol with minor changes.[26] Primary antibodies used is mouse anti-Globo H monoclonal antibody VK-9 (IgG; eBioscience™, Catalog #:14-9700-82). The secondary antibodies used were Biotin-conjuagted Goat anti-mouse IgG(H+L) and HRP-conjugated goat anti-mouse IgG(H+L) (ThermoScientific, Catalog #:A16066). Beads used were streptavidin conjugated magnetic beads (Dynabeads™ MyOne™ Streptavidin T1, Invitrogen, Catalog #:65601), Pierce Protein A/G Magnetic Beads (ThermoScientific, Catalog #:88803) and Agencourt AMPure Beads (Beckman Coulter). Taq DNA polymerase, NHS-PEG4-Biotin and Pierce DNA Coating Solution were all purchased from ThermoScientific.

DNA Sequences

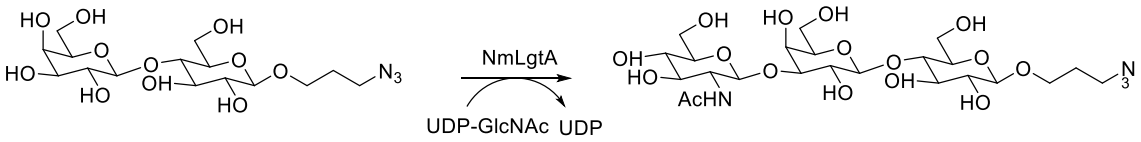
Table 3.2. *DNA codes used in the study*

Name	5'Mod	Sequence	3'Mod
Oligo O	5'Hexynyl	**AATGATACGGCGACCACCGAA**GAAAAGACGATCTAA GCTCTACGATGTAAGA***CACGTCTGAACTCCAGTCAC***	None
Oligo A	5'Hexynyl	**AATGATACGGCGACCACCGA**ACGCAACTTTAGAAAA GAAAACGATCTAAGATCTACGATGTAA***CACGTCTGAAC TCCAGTCAC***	None
Oligo B	5'Hexynyl	**AATGATACGGCGACCACCGA**ACGAAACTTTAGAAAA GAAAACGATCTAAGCTCTACGATGTAAGA***CACGTCTGA ACTCCAGTCAC***	None
Gb-H	5'Hexynyl	**AATGATACGGCGACCACCGAA**AGAAAAGACGATCTA CGCTCTACGAATAACGATGTAAGA***CACGTCTGAACTCC AGTCAC***	None
Gb-5	5'Hexynyl	**AATGATACGGCGACCACCGAA**ACGATCTACGCTCTAC GAATAACGATGTAAGA***CACGTCTGAACTCCAGTCAC***	None
Gb-4	5'Hexynyl	**AATGATACGGCGACCACCGA**ACGCTCTACGAATAACG ATGTAAGA***CACGTCTGAACTCCAGTCAC***	None
Gb-3	5'Hexynyl	**AATGATACGGCGACCACCGAA**ACGAATAACGATGTA AGACCCCAGTCAGGCCTAACGTA***CACGTCTGAACTCCA GTCAC***	None
Gb-2	5'Hexynyl	**AATGATACGGCGACCACCGAA**ACGATGTAAGACCCC AGTCAGGCCTAACGTA***CACGTCTGAACTCCAGTCAC***	None

Bb-4	5'Hexynyl	**AATGATACGGCGACCACCGAA**AGAAAAGACGATCTA CGCTCTACGA ***CACGTCTGAACTCCAGTCAC***	None
Bb-3	5'Hexynyl	**AATGATACGGCGACCACCGAA**AGAAAAGACGATCTA CGCCCCCAGTCAGGCCTAACGTA***CACGTCTGAACTCCA GTCAC***	None
Bb-2	5'Hexynyl	**AATGATACGGCGACCACCGAA**AGAAAAGACGACCCC AGTCAGGCCTAACGTA***CACGTCTGAACTCCAGTCAC***	None
FW Primer	None	**AATGATACGGCGACCACCGAA**	None
RV Primer	None	***GTGACTGGAGTTCAGACGTG***	None

***** For all 5' to 3' DNA Sequences **BOLD** is the forward primer and ***BOLD ITALIC*** is the reverse primer binding site.

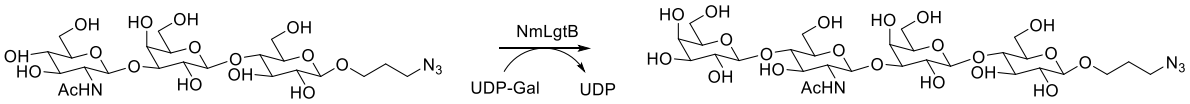
Preparative scale synthesis of blood antigen ABO and globo series glycans



1 2

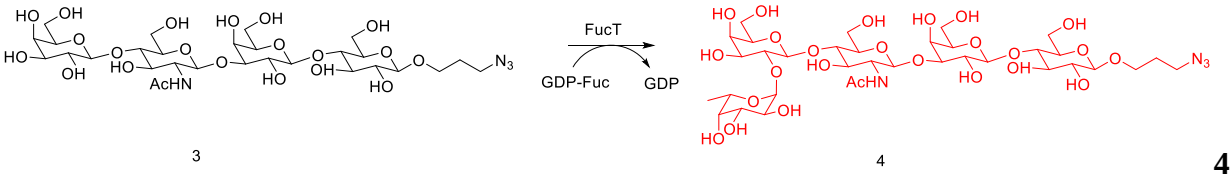
2 was prepared from **1** using LgtA from *Neisseria meningitides*. In detail, a reaction mixture in a final volume of 50 ml containing 50 mM Tris-HCl, 425 mg of 1 (1 mmol), 607 mg of UDP-GlcNAc (1 mmol), 5 mM of Mg^{2+}, 8 mg of LgtA was carefully shaken at 37°C to allow the formation of **2.** The reaction was monitored by TLC (EtOAc/MeOH/H2O/HOAc=5:2:1.4:0.4). Once the reaction was no longer move forward, equal volume ethanol was added to remove

proteins and the solution was concentrated *in vacuo*. After purification by using Bio-Gel P-2 column. 530 mg of **2** was obtained in 85% yield regarding 1.



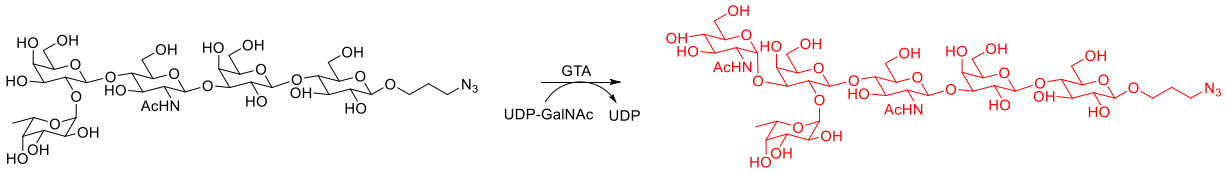
2 3

3 was prepared from **2** using LgtB from *Neisseria meningitides*. In detail, 503 mg of 2 was dissolved in a 80 ml of reaction mixture containing 50 mM Tris-HCl, 12 mM of UDP-Gal, 5 mM of Mg^{2+}, 10 mg of LgtB. The reaction was carefully shaken at 37°C for overnight to allow the formation of **3.** The reaction was monitored by TLC (EtOAc/MeOH/H2O/HOAc=5:2:1.4:0.4). Once reaction finished, equal volume ethanol was added to precipitate proteins and the supernatant was concentrated *in vacuo*. After purification by using Bio-Gel P-2 column. 432 mg of **3** was obtained in 76% yield regarding 2.



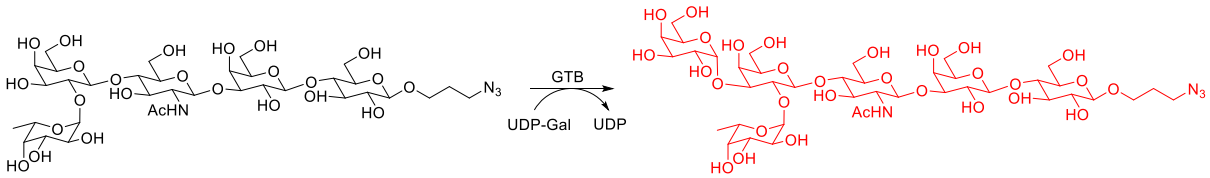
3 4 **4**

(blood O-antigen) was prepared **3** using FucT from *Helicobacter mustelae*. In detail, reaction was performed in a final volume of 35 ml mixture containing 50 mM Tris-HCl (pH 8.0), 277 mg of 3 (10 mM), 5 mM of Mg^{2+}, 15 mM of GDP-Fucose, and 4 mg of FucT. 20 units of alkaline phosphatase were added into reaction system to hydrolyze the newly formed GDP to improve the reaction yield. The reaction was carefully shaken at 37°C for overnight and monitored by TLC (EtOAc/MeOH/H2O/HOAc=5:2:1.4:0.4). Once reaction finished, the product was purified by using Bio-Gel P-2 column. 211 mg of 4 was obtained in 64 % isolated yield regarding 3. Product was confirmed by MALDI-TOF-MS analysis.



4 5

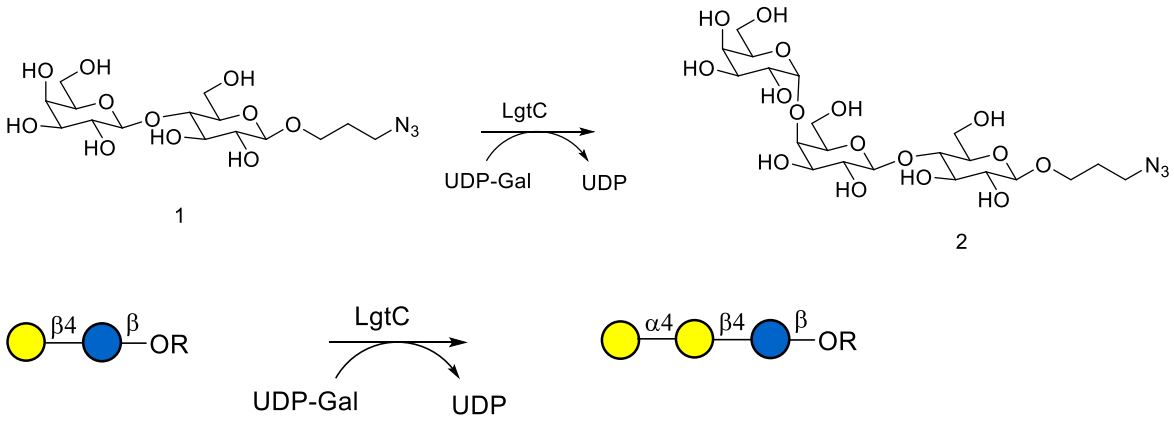
5 (blood A-antigen) was prepared from blood O-antigen using GTA from human. In detail, reaction was performed in a final volume of 14 ml mixture containing 20 mM Tris-HCl (pH 8.0), 66 mg of 4 (5 mM), 3 mM of Mg^{2+}, 7 mM of UDP-GalNAc, and 5 mg of GTA. The reaction was carefully shaken at 37°C for overnight for the formation of 5. The reaction was monitored by TLC (isopropanol/NH4OH/H2O = 7:3:2). Once reaction finished, the product was purified by using Bio-Gel P-2 column. 45 mg of 5 was obtained in 56 % isolated yield regarding 4. Product was confirmed by MALDI-TOF-MS analysis.



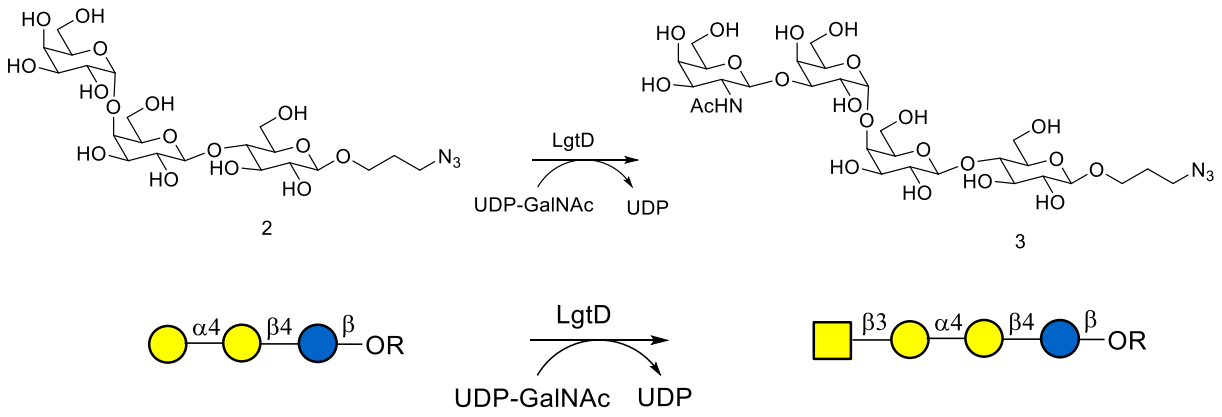
4 6

6 (blood B-antigen) was prepared from blood O-antigen using GTB from human. In detail, reaction was performed in a final volume of 14 ml mixture containing 20 mM Tris-HCl (pH 8.0), 66 mg of 4 (5 mM), 3 mM of Mg^{2+}, 7 mM of UDP-Gal, and 5 mg of GTB. The reaction was carefully shaken at 37°C for overnight for the formation of 6. The reaction was monitored by TLC (isopropanol/NH4OH/H2O = 7:3:2). Once reaction finished, the product was purified by using Bio-Gel P-2 column. 39 mg of 6 was obtained in 51 % isolated yield regarding 4. Product was confirmed by MALDI-TOF-MS analysis.

Synthesis of Globo-H and related glycans

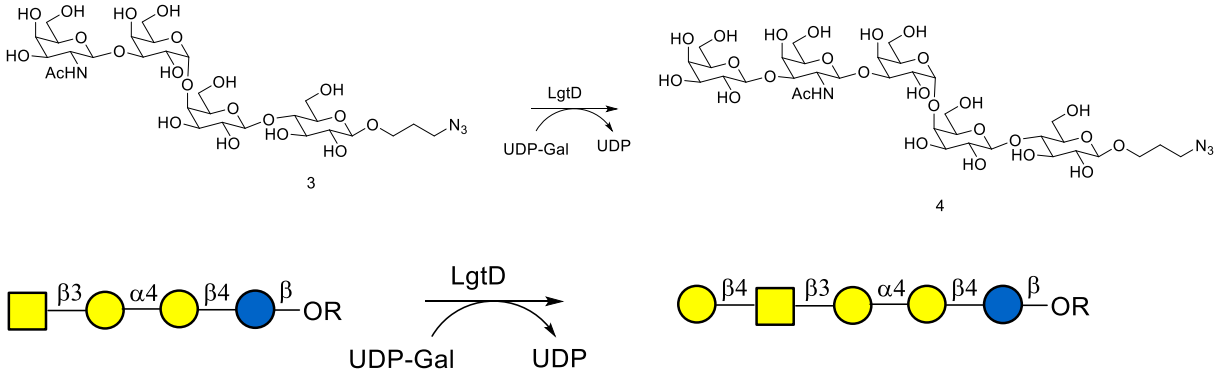


2 was prepared from **1**. In detail, a reaction mixture in a final volume of 10 ml containing 50 mM Tris-HCl, 85 mg of 1 (2 mmol), 153 mg of UDP-Gal (2.5 mmol), 5 mM of Mg^{2+}, 2 mg of LgtC was carefully shaken at 37°C to allow the formation of **2.** The reaction was monitored by TLC (EtOAc/MeOH/H2O/HOAc=5:2:1.4:0.4). Once the reaction was no longer move forward, equal volume ethanol was added to remove proteins and the solution was concentrated *in vacuo*. After purification by using Bio-Gel P-2 column. 99 mg of **2** was obtained in 86% yield regarding **1**.

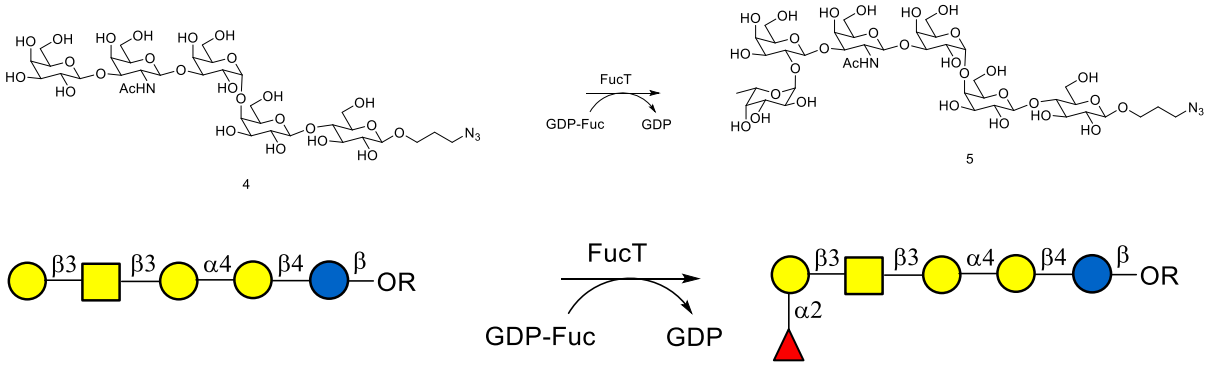


In a reaction mixture of 5 ml volume, 50 mM Tris-HCl, 59 mg of 2 (1 mmol), 5 mM of Mg^{2+}, 98 mg of UDP-GalNAc (1.5 mmol), 2 mg of LgtD were added. the reaction was carefully shaken at 37°C overnight to allow the formation of **3.** Once the reaction finished, equal volume ethanol

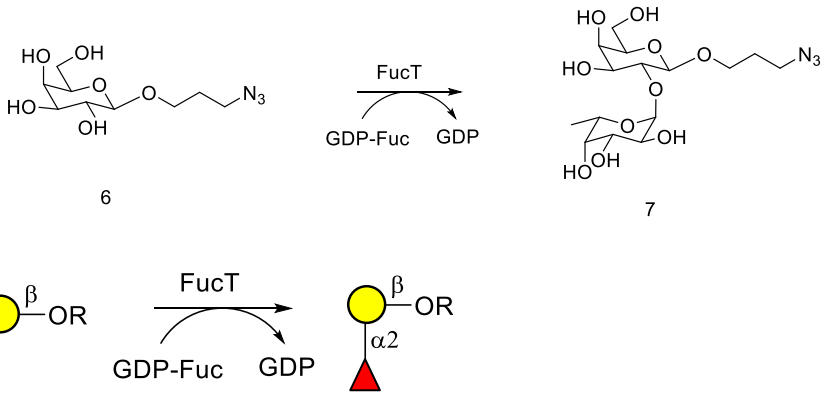
was added to remove proteins and the solution was concentrated *in vacuo*. After purification by using Bio-Gel P-2 column, 63 mg of **3** was obtained in 80% yield regarding **2**.



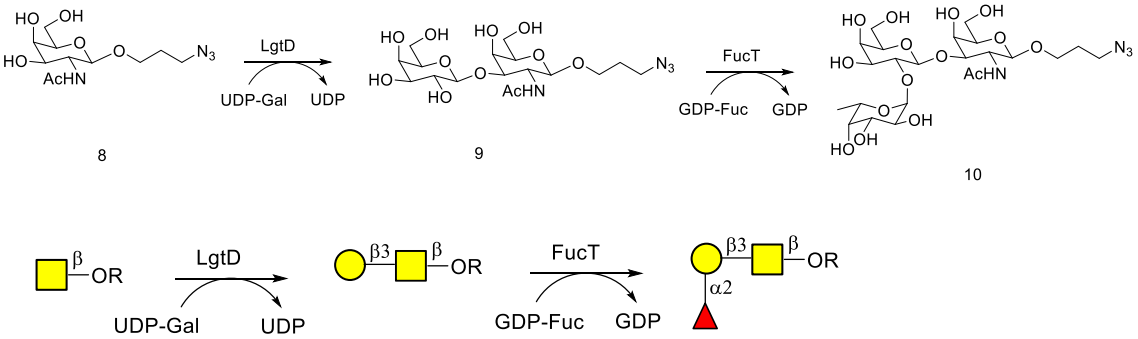
4 was prepared from **3**. A 5 ml of reaction mixture containing 50 mM Tris-HCl, 5 mM of Mg^{2+}, 47 mg of **3** (0.6 mmol), 65 mg of UDP-Gal (1 mmol), and 5 mg of LgtD was carefully shaken at 37°C overnight to allow the formation of **3.** 5 unit of alkaline phosphatase was added to hydrolyze the formation of UDP to improve the reaction yield. After purification by using Bio-Gel P-2 column, 36 mg of **4** was obtained in 62% yield.



GloboH (**5**) was prepared from **4**. A reaction mixture containing 50 mM Tris-HCl, 5 mM of Mg^{2+}, 29 mg of **4** (0.3 mmol), 32 mg of GDP-Fuc (1 mmol), and 2 mg of FucT was carefully shaken at 37°C overnight to allow the formation of GloboH. 5 unit of alkaline phosphatase was added to hydrolyze the byproduct GDP to improve the reaction yield. After purification by using Bio-Gel P-2 column, 24 mg of GloboH was obtained in 71% yield.

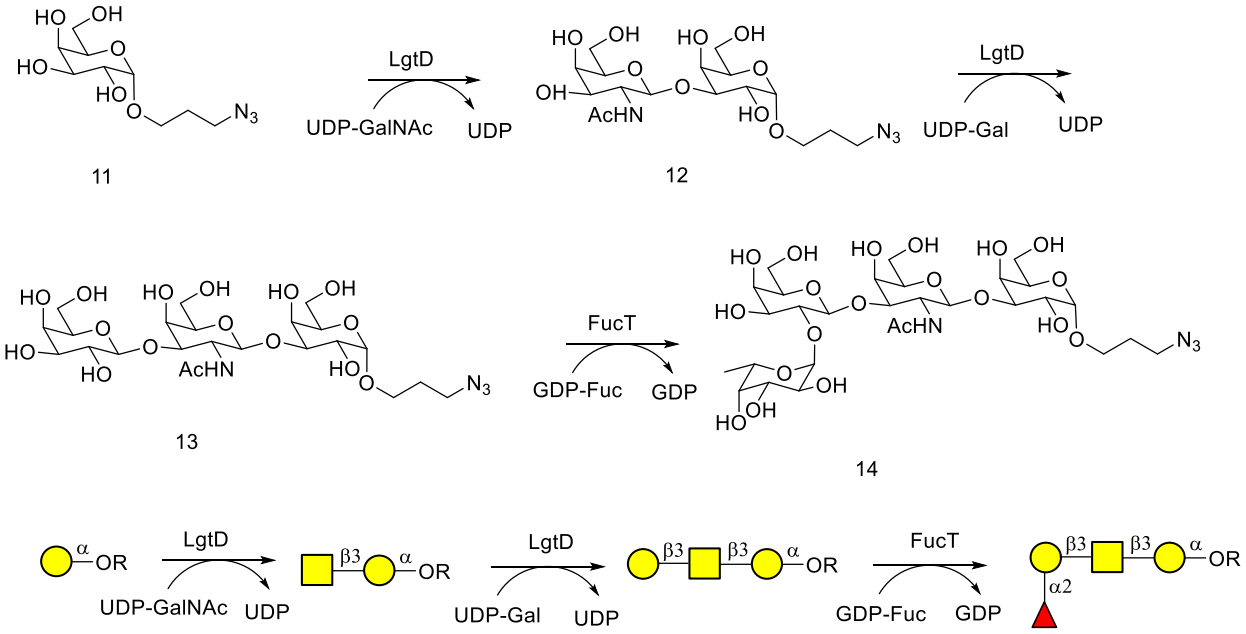


In a 10 ml of reaction mixture, 50 mM Tris-HCl, 53 mg of **6** (2 mmol), 5 mM of Mg^{2+}, 158 mg of GDP-Fucose (2.5 mmol), 3 mg of FucT were added. the reaction was carefully shaken at 37°C overnight to allow the formation of **7.** Once the reaction finished, equal volume ethanol was added to remove proteins and the solution was concentrated *in vacuo*. After purification by using Bio-Gel P-2 column, 61 mg of 7 was obtained in 75% yield.



10 was synthesized from 8 by two reaction steps. In a 5 ml of reaction mixture, 50 mM Tris-HCl, 5 mM of Mg^{2+}, 61 mg of **8** (2 mmol), 153 mg of UDP-Gal (2.5 mmol), 4 mg of LgtD were added. the reaction was carefully shaken at 37°C overnight to allow the formation of **9.** Once the reaction finished, equal volume ethanol was added to remove proteins and the solution was concentrated *in vacuo*. After purification by using Bio-Gel P-2 column, 77 mg of **9** was obtained in 83% yield regarding **8**. In second reaction step, 77 mg of **9** was incubated in a 5 ml of 50 mM

Tris-HCl with 5 mM of Mg^{2+}, 47 mg of **9** (1 mmol), 127 mg of GDP-Fucose (2 mmol), and 3 mg of FucT. The reaction was carefully shaken at 37°C overnight to allow the formation of **10.** Once the reaction finished, equal volume ethanol was added to remove proteins and the solution was concentrated *in vacuo*. After purification by using Bio-Gel P-2 column, 69 mg of **10** was obtained in 57% yield regarding to **8**.



14 was prepared from 11 stepwisely. In a 10 ml of reaction system, 50 mM Tris-HCl, 5 mM of Mg^{2+}, 53 mg of **11** (2 mmol), 163 mg of UDP-GalNAc (2.5 mmol), and 2 mg of LgtD were added. The reaction was performed at 37°C overnight to allow the formation of **12.** 12 was purified by Bio-Gel P-2 column (75 mg, 81% yield). Then, 75 mg of 12 was dissolved in 10 ml water containing 50 mM Tris-HCl, 5 mM of Mg^{2+}. 122 mg of UDP-Gal (2 mmol) and 7 mg of LgtD were added. In addition, 5 units of alkaline phosphatase was added to hydrolyze the byproduct GDP to improve the reaction yield. The reaction was performed at 37°C. **13** was obtained after P-2 gel purification. In last reaction step, 32 mg of 13(0.5 mmol) was dissolved in 5 ml of water

containing 50 mM Tris-HCl, 5 mM of Mg^{2+}. 1.5 mg of FucT and 48 mg of GDP-Fuc(0.75 mmol) was added to start the reaction at 37°C. 2 units of alkaline phosphatase was added to hydrolyze the byproduct GDP to improve the reaction yield. 14 was purified by P-2 column (28 mg, 73% yield).

Gradient PCR and Tm Determination

Initially a gradient PCR was performed to figure out the optimal melting temperature to be used in the further quantitation experiments. Platinum Pfx DNA Polymerase was used for the PCR, approximately 26 µg of template, 1 µL of primers (10 µM), 1mm dNTP, 10X amplifying buffer, $MgSO_4$, Polymerase and nuclease free water was used for a typical 25 µL reaction. Standard thermocycling condition were used with different temperature in the melting phase (95°C for 10 min, X°C for 30s, 95°C for 15s, 13 cycles) wherein X represents the melting phase and temperature gradient from 50 ºC - 60 ºC were tested during this phase. The amplified PCR products were analyzed using 4% agarose gel electrophoresis. T_m determined using this method was used for qPCR.

Component	25 µl	50 µl	Final Concentration
10X Amplifying Buffer	5	10	-
MgSO4	1	2	-
dNTP(10mM)	0.75	1.5	0.3mM
Forward Primer(10uM)	0.75	1.5	0.3uM
Reverse Primer(10uM)	0.75	1.5	0.3uM
Template(1uM)	1	2	0.04
Polymerase	0.3	0.5	-
Nuclease Free Water	15.4	31	-

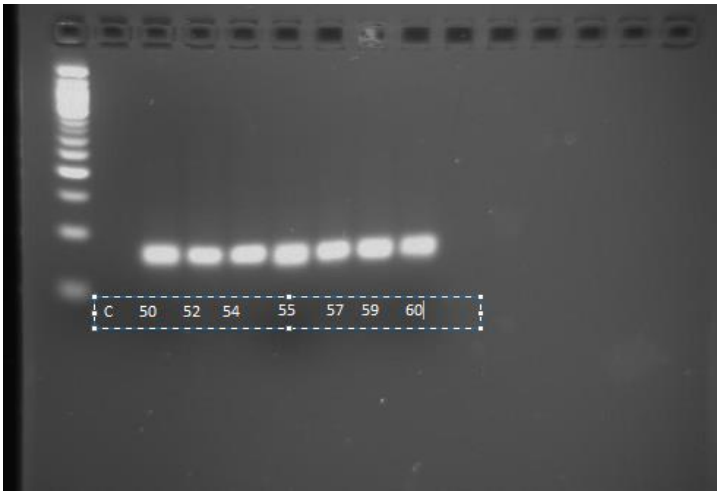


Figure 3.11. Gradient PCR for Tm determination.

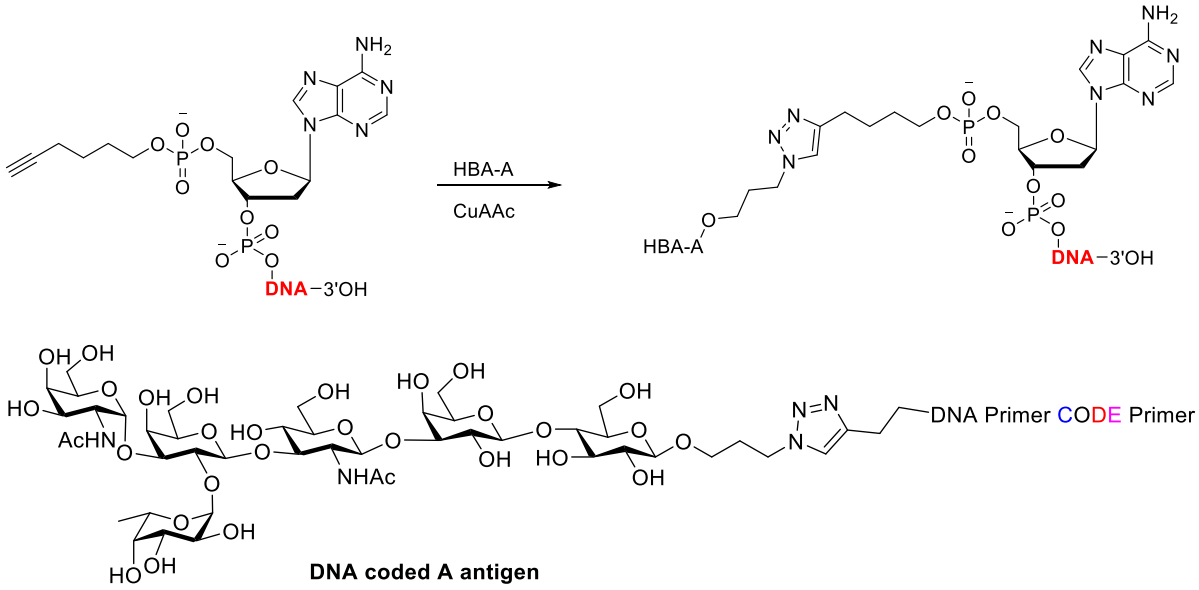


Figure 3.12. Click conjugation of 5'-hexynyl DNA and azido modified glycans.

Synthesis of glycan DNA conjugates: Glycan(Antigen)- DNA Conjugates were synthesized using Azido-Alkyne cycloaddition click reaction. The 5'-/5Hexynyl-terminated DNA was procured from the commercial suppliers (IDT) with standard desalting purified. Desalted DNA was HPLC purified before click conjugation. All glycans were synthesized via chemo enzymatic method with the azido propyl linker at the reducing end as described above.

Click reaction procedure: 5'-Hexynyl-terminated DNA (500 µM, 20 µL), azido glycan (1.5 mM, 30 µL), 2M TEA buffer (pH 7, 20 µL, final concentration of 0.2 M), 5 mM of freshly prepared Ascorbic acid solution (20 µL, final concentration 0.5mm), 10 mm of copper-TBTA in 55% DMSO (10 µl, final concentration 0.5 mm) and 50 volume % DMSO (100 µl) were mixed together in a tube. The reaction mixture was vortexed and kept at room temperature for overnight. The reaction mixture was purified using Micro Bio-Gel P-30 Chromatography Columns (20 base pair cut-off). The concentration of the Glycan (Antigen)-DNA Conjugates was determined by absorbance at 260/280 using Nanodrop (ThermoFisher).

Any presence of residual unreacted DNA was further tested by PCR amplification and gel electrophoresis. Recovered reaction mix was incubated with the specific antibody for two hours and washed with the washing buffer, and later eluted with the elution buffer. Both wash and elution buffers were collected and performed a PCR. Gel analysis of the PCR product did not show a band in the wash buffer (residual unreacted DNA) but a strong band was seen in elution buffer indicating all DNA is coupled with the azido sugar (Figure 3.11)

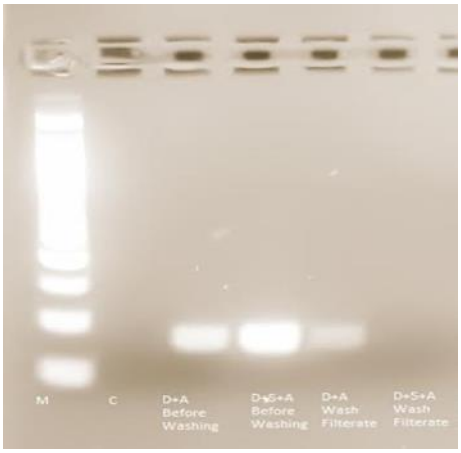


Figure 3.13. Gel analysis of 'clicked' DNA.

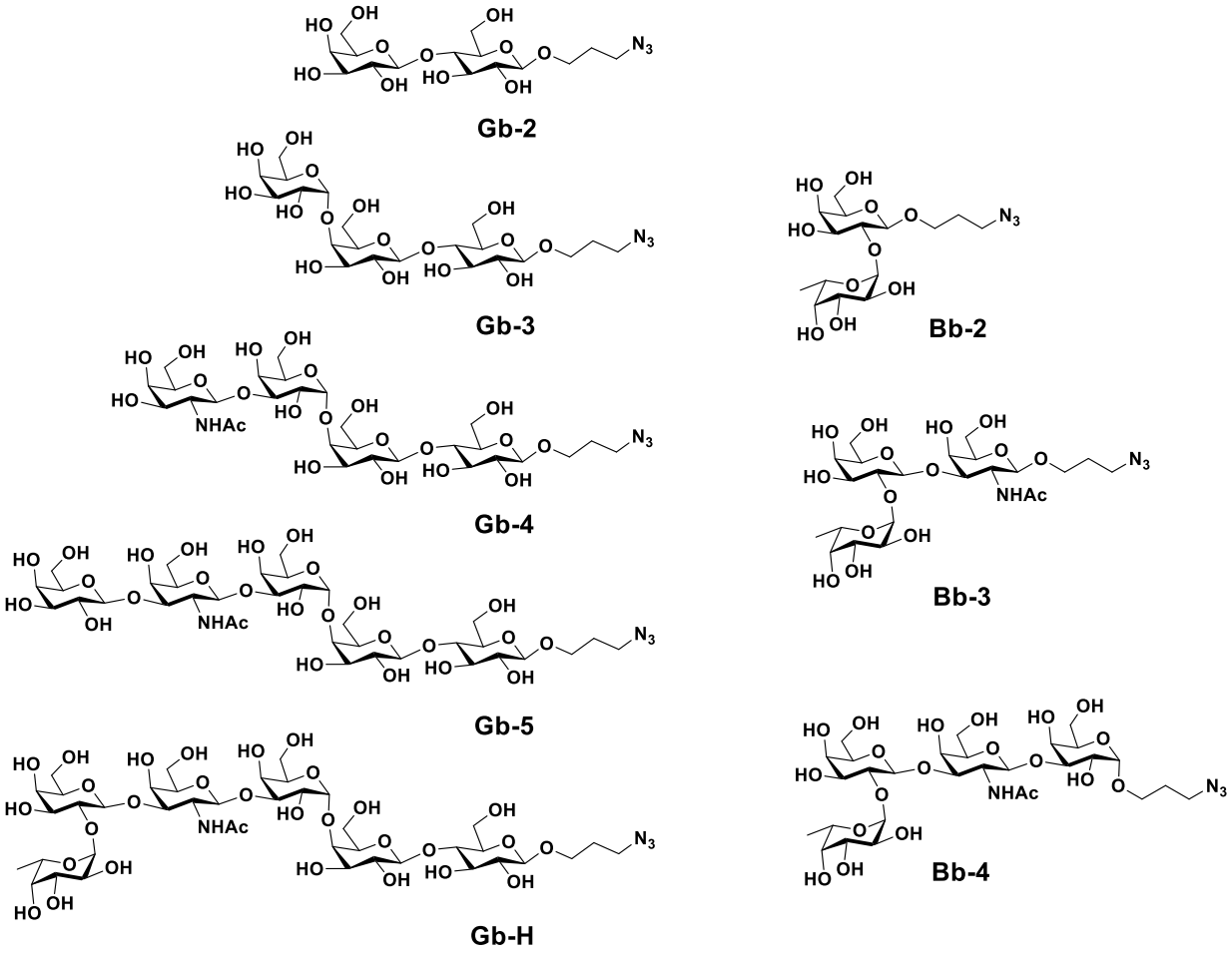


Figure 3.14. Glycan structures used for the conjugation with alkyne DNA and alkyne biotin

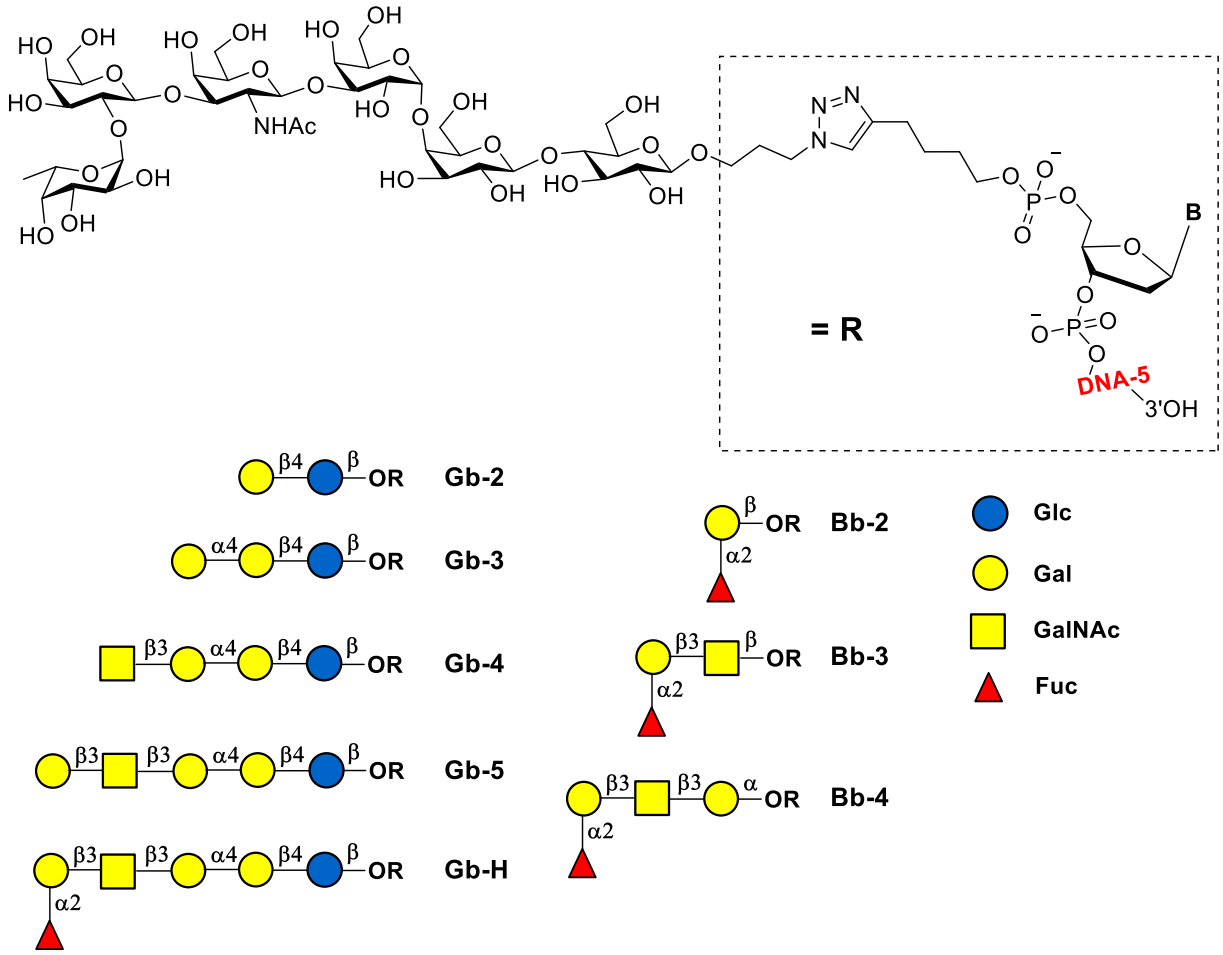


Figure 3.15. Structure of Glycan DNA conjugates.

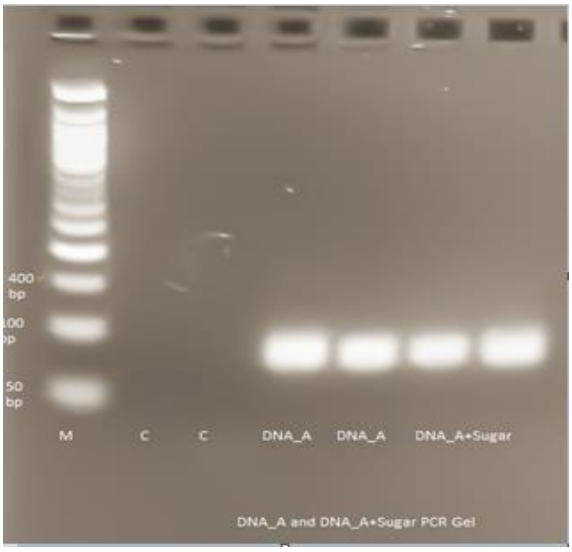
Characterization of DNA+Glycan conjugates

DNA codes for each glycans were purified by HPLC before used in click reaction, Preparative HPLC purification was achieved using following conditions; Agilent 1200 HPLC system, Xbridge semi prep C18 column, 5 uM, 10X 150 mm; Solvent A : pH 7.0 TEAA buffer 0.1 M, Solvent B: 20% CAN in solvent A; 5% B to 95% B over 30 minutes with a flow rate of 1.2 mL/minute. After click reaction, products were purified and concentrated and analyzed by HPLC. Analytical method: Agilent 1200 HPLC or Shimadzu LC20AT, Eclipse Plus C18, 3.5 uM, 4.6 X

100 mm column; Solvent A : pH 7.0 TEAA buffer 0.1 M, Solvent B: 20% CAN in solvent A; 5% B to 95% B over 30 minutes with a flow rate of 0.6 mL/minute.

PCR comparison of DNA and G+DNA

A PCR comparison was conducted between pure DNA and G+DNA to make sure that the DNA can be used for further PCR reaction after conjugation with glycan. PCR was conducted under the same conditions as stated under Gradient PCR and Tm Determination and analyzed using 4% agarose gel electrophoresis.



A : DNA 100 Ladder

B: Control

C : Control

D : DNA_A

E : DNA_A

F: G+DNA_A

G: G+DNA_A

Figure 3.16. PCR comparison of pure DNA and glycan conjugated DNA.

qPCR Methods

General qPCR protocol for a typical 12.5 µL reaction: The glycan(Antigen)- DNA Conjugates were added to 6.25 µL 2X Maxima SYBR Green/ ROX qPCR Master Mix with 1 µL primers (Forward Primer – 500 nM and Reverse Primer- 300 nm). qPCR was performed with Applied Biosystems Stepone System (50 ^{o}C for 2 mins (Holding), 95^{o}C for 10 mins (Holding), 95 ^{o}C for 15 s, 60^{o}C for 10 s ,72 ^{O}C for 10 s for 40 cycles).

Ct or threshold cycle, is a measurement of signal intensity for qPCR experiments. In a qPCR experiment, PCR is performed in presence of a fluorogenic intercalating dye (SYBR Green in our case). The dye intercalates with double-stranded DNA, as more double-stranded DNA is produced in each PCR cycle the dye increases the fluorescence intensity. Once the fluorescence intensity reaches a threshold level, the cycle number is recorded by the instrument as Ct value. Therefore, sample having large amount of DNA will have a lower Ct value compared to those samples containing relatively less amount of DNA.

Standard curve plot of DNA and G+DNA

To prove the consistency over concentration for both pure DNA and G+DNA, a qPCR standard curve reaction was conducted over a series of 8 different concentrations (2.4 nM, 1.2 nM, 600 nM, 206 nM, 130 nM, 64 pM, 32 pM, and 16 pM) with standard thermocycling conditions stated above.

Component	Volume(in uL)
2x Maxima SYBR Green/ROX qPCR master mix	6.25
Forward Primer(500nm)	1
Reverse Primer(300nm)	1
Template(Gradient as stated above)	1
Nuclease Free Water	3.25

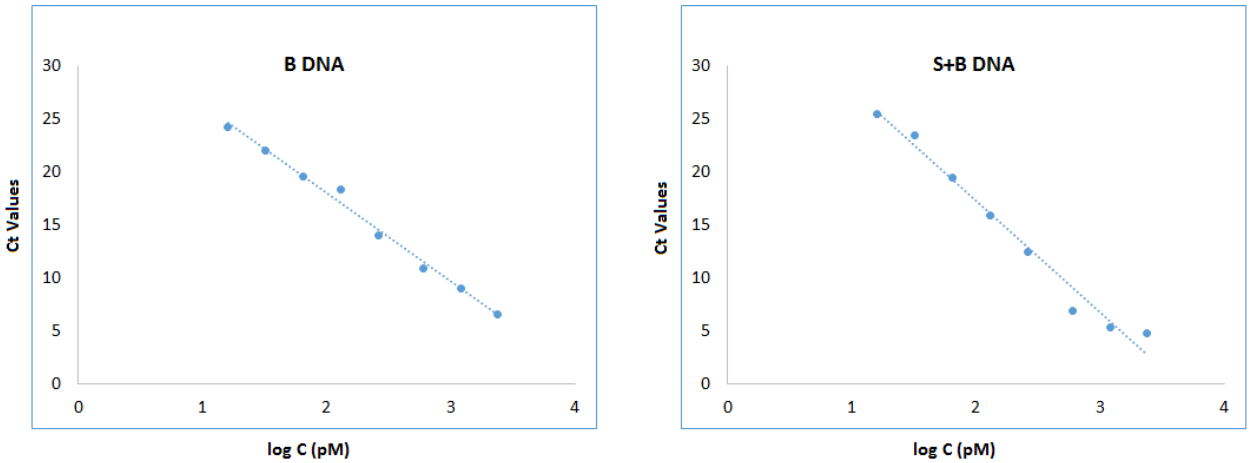


Figure 3.17. Standard curve plot of pure DNA (B DNA) and glycan conjugated B DNA.

Immunoprecipitation of Antibody-Antigen-DNA Conjugates

10 µL of glycan(antigen)-DNA Conjugate (5 µM) was incubated with 2 µL of antibody (if plasma is used then 2 µL of plasma is diluted to 8 µL using water) for 2 hours at room temperature. After two hours, the reaction volume was made up to 100 µL using TBST buffer and added to 20 µL of prewashed A/G protein bead and incubated for one hour at room temperature with occasional shaking at an interval of 15 mins. After one hour the beads were washed thoroughly for 7 times using TBST buffer. After washing the beads were reconstituted in 20 µL of TBST buffer. The beads (2 µL) were directly used as template for qPCR analysis.

DNA encoded glycans for the detection of glycan binding proteins

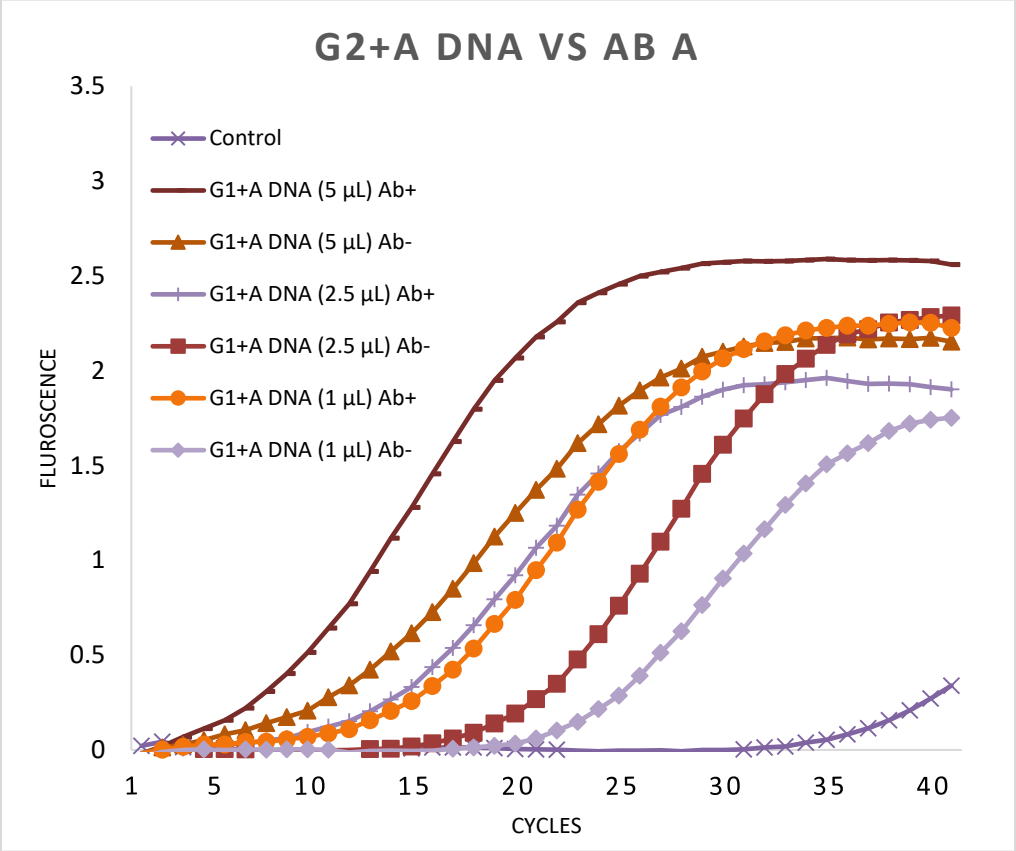


Figure 3.18. G1+A DNA for the detection of A antibodies.
10 µL of three different concentration (5 µM, 2.5 µM and 1 µM) incubated with 2 µL of Blood Group A Antigen Antibody HE-193 (MA1-19693). IP and qPCR was performed as described previously.

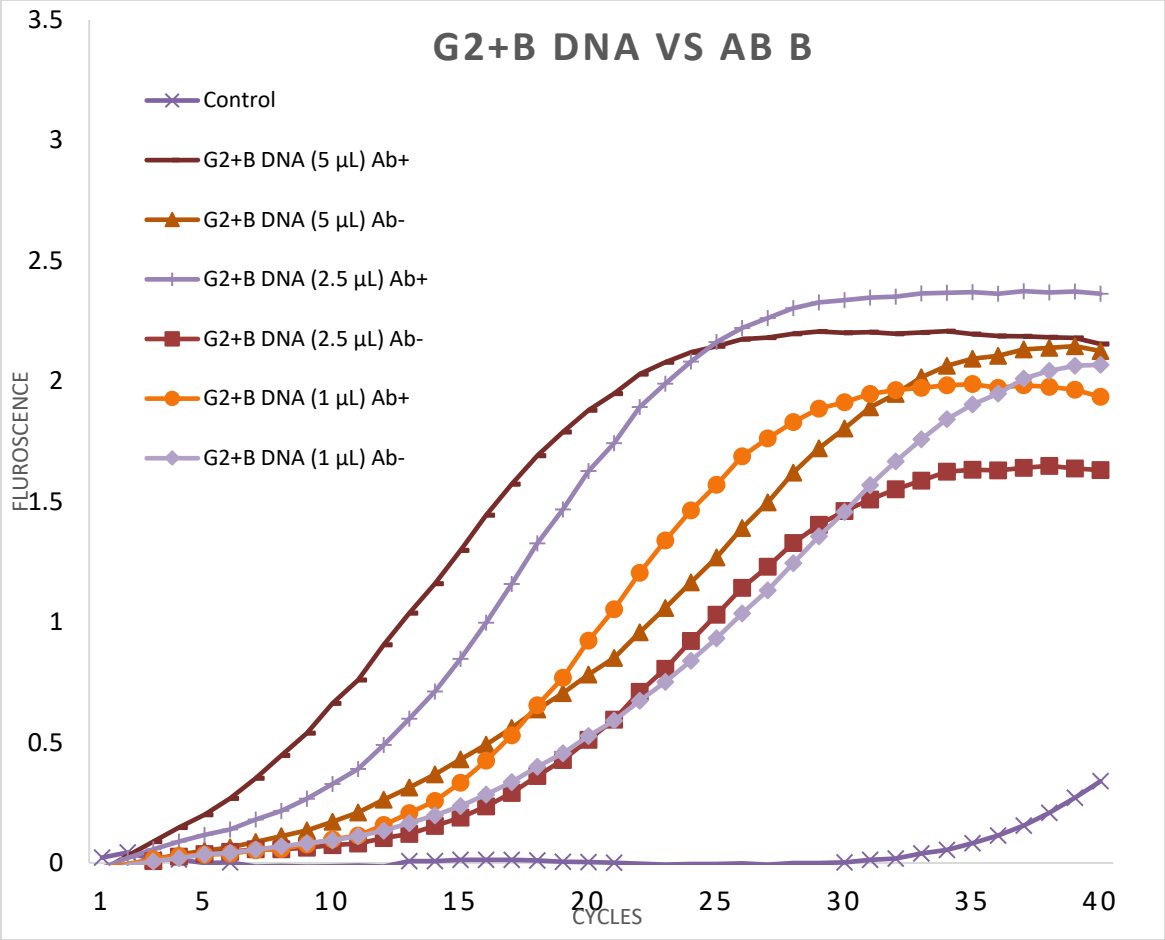


Figure 3.19. G2+A DNA for the detection of B antibodies.
10 µL of three different concentration (5 µM, 2.5 µM and 1 µM) incubated with 2 µL of Blood Group B Antigen Antibody HEB-29 (MA1-19691). IP and qPCR was performed as described previously.

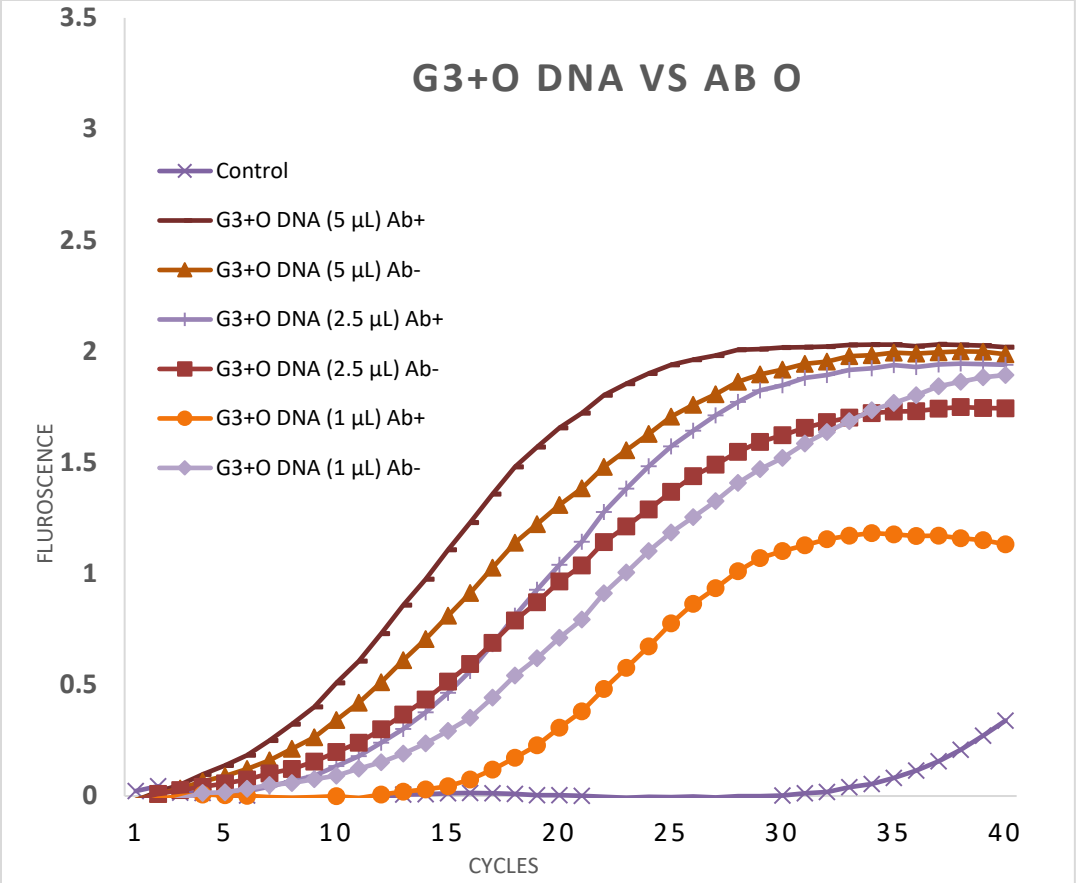


Figure 3.20. G3+O DNA for the detection of O antibodies.
10 µL of three different concentration (5 µM, 2.5 µM and 1 µM) incubated with 2 µL of Blood Group ABH Antigen Antibody HE-10 (MA1-19694). IP and qPCR was performed as described previously.

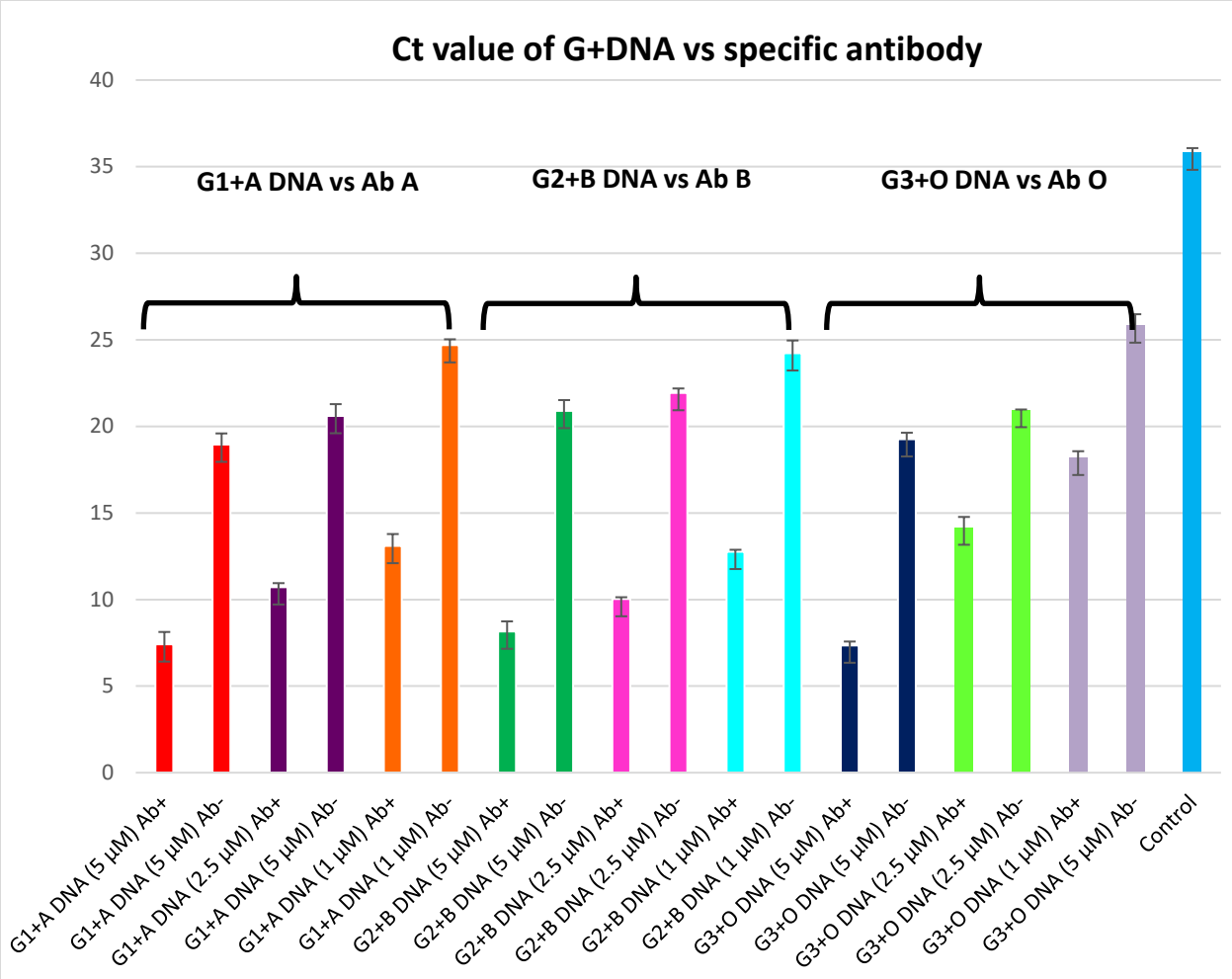


Figure 3.21. Ct value of different concentrations of G+DNA when incubated with the specific antibody.

Selectivity and specificity analysis of DEGL

Glycan-DNAs with the mixture of antibodies A, B and O (G+DNA Vs ABO Abs 1:1:1):

Antibody mixture was prepared by mixing 2 µL each of Blood Group A Antigen Antibody HE-193 (MA1-19693), Blood Group B Antigen Antibody HEB-29 (MA1-19691) and Blood Group ABH Antigen Antibody HE-10 (MA1-19694). Then, 10 µL of each glycan conjugates (2.5 µM) were incubated with 2 µL of antibody mix separately. Immunoprecipitation and qPCR

analysis was performed as described above. A no antigen control with none conjugated DNA (A DNA, 2.5 µM) is used in immunoprecipitation and a no template control was used in qPCR.

G2+B DNA vs A, B, and O antibodies: In this experiment, 10 µL of G2+B DNA (2.5 µM) was incubated with 2 µL of each antibody. A no antibody control is used in immunoprecipitation and a no template control was used in qPCR. Immunoprecipitation and qPCR analysis was performed as described above.

Globo-H vs VK-9: The immunoprecipitation procedure for Globo-H with VK9 were same as stated in Immunoprecipitation of Antibody-Antigen- DNA Conjugates. The volume of antibody used was 2 µL and Globo-H was 10 µL (2.5 µM).

ELISA experiment.

The 96-well plate (Costar Polystyrene High Binding Plate 3590) was coated with 100 ml of 10 ug/ml Streptavidin (Sigma) in 0.01 M PBS (pH 7.4) at 4oC overnight. The coated plate was then washed with 150 ul of 0.05% Tween-20/PBS buffer (pH 7.4) (PBST) for three times and then add 100ul Biotinylated sugar (synthesized via click reaction, Figure S1) for capturing. After 30 minutes, the plate was washed. For DNA-sugar ELISA, the only difference is to coat the DNA-sugar onto plastic surfaces directly at the help of Pierce DNA Coating Solution for overnight. Then the plate was blocked with 2% (w/v) BSA in PBST for 2 hrs. After washed with another 150 ul PBST for 3 times, 100 ul VK9 (a mouse IgG anti-Globo H monoclonal antibody) with a series of dilution was added for incubation at room temperature for 2 hrs. After washed with 200 ul PBST for 6 times, HRP-goat anti-mouse IgG(H+L) was added to the plate at a dilution as recommended for 1 hr at room temperature. After washed with another 200 ul PBST for 6 times, 100 ul TMB solution was added to the plate and then stopped by 100 ul 1 M phosphoric acid. The plate was

then read at OD450 with a plate reader (PerkinElmer, 2030 Multilabel Reader Victor). The result is calculated by minus OD450 value of control (no antibody).

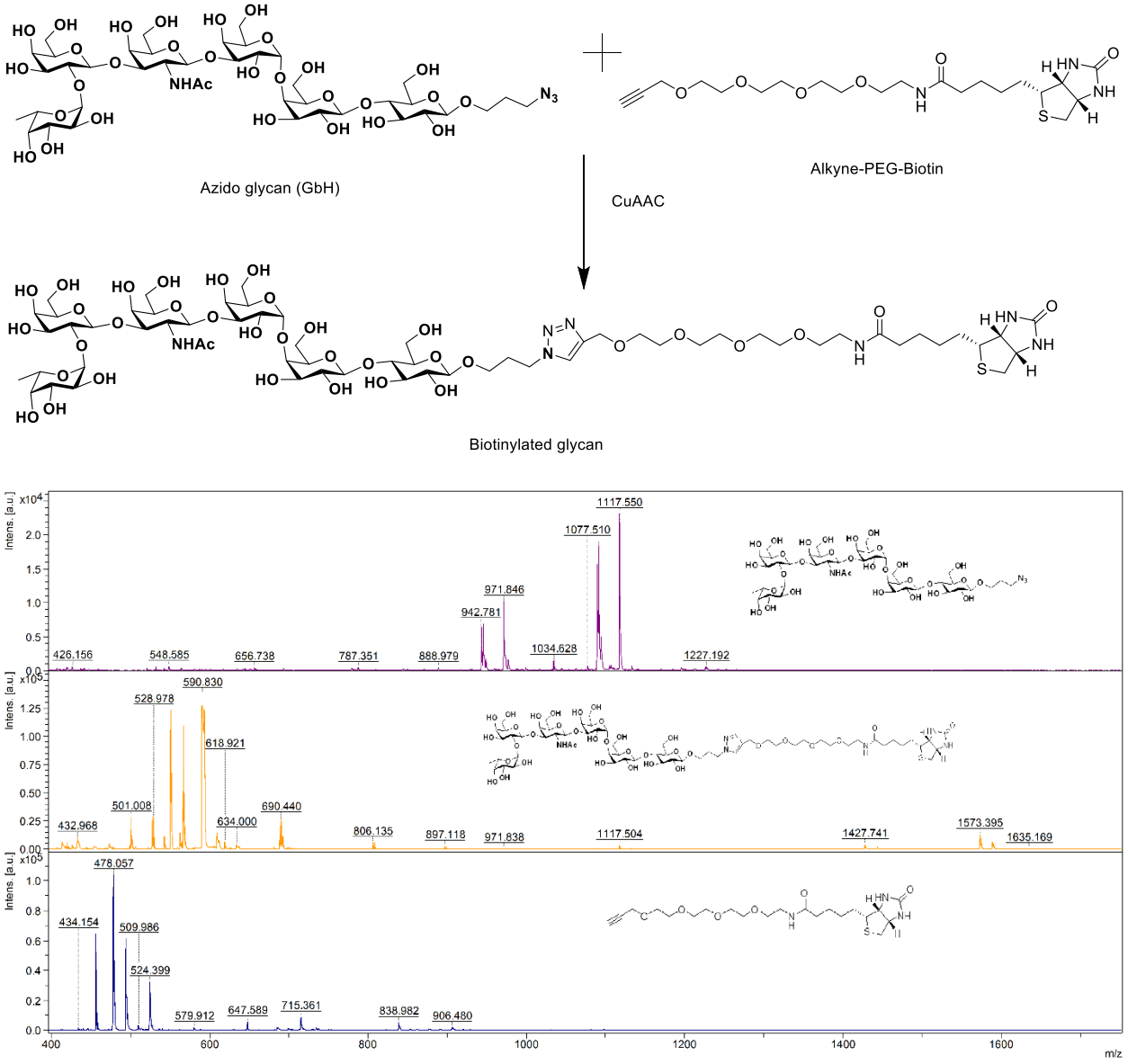


Figure 3.22. MALDI TOF analysis of GH and biotinylated-GbH.

NGS experiment

NGS Protocol

1. Selection: 10 ul of 100nM DEGL was incubated with 2ul (1ug) Vk-9 antibody (Globo H Monoclonal Antibody, eBioscience; Catalog Number: 14-9700-82) for 2 hrs at 37oC. After 2 hrs 0.2 ul of Biotin Conjugated Goat Anti-Mouse IgG Secondary Antibody (Thermofischer Scientific; Catalog Number:31800), 40 ul of Sheared Salmon Sperm DNA (Thermofischer Scientific; catalog Number: AM9680) and 50 ul of TBST buffer (0.1 % Tween 20, 0.1% BSA) was added and incubated at 37oC for 1 hr. 20ul of Streptavidin Magnetic beads (Dynabeads Myone Streptavidin T1; Catalog Number: 65601) was prewashed using PBS as per user instructions. After 1 hr the solution was incubated with prewashed Streptavidin beads for 30 mins at 37oC. The beads were washed 7 times using 100 ul TBST buffer. The washed beads were then diluted in 20 ul of TBST buffer.

2. Library Preparation: 1ul of washed streptavidin beads were used for amplification using template primers with Taq polymerase (Thermofischer Scientific; Catalog Number: EP0402) for 18 cycles. The PCR product was purified using Agencourt AMPure XP Magnetic beads (30% PEG) as per user instructions. The purified PCR product was amplified again using NGS adapters (30 Cycles) and purified using Agencourt AMPure XP Magnetic beads. The purified product's concentration was measured using Agilent 2100 Bioanalyzer. An equimolar solution of 26pM was prepared which was loaded in Ion Chef for the Chip Preparation and sequencing by Ion PGM.

4. CONCLUSION

Cancer is the most lethal and feared disease of this century. The disease has several causes including gene mutation, abnormal protein production and external factors like lifestyle and environmental factors.[168, 169] These multifactorial dispositions have delineated the overall lack of success in cancer therapy compared to other diseases like heart disease and diabetes. Traditionally cancer is treated with the chemo therapeutic agents and more recently with the targeted small molecules and immune therapy.[3, 4, 8, 170] Chemotherapeutic agents target rapidly dividing cells lacking any specificity, hence it eliminates cancer cells and other fast-growing tissues alike. Researchers quest to minimizing this off- target effect yielded the first targeted therapy for cancer in early 1990s.[3] Ever since, the war against cancer is led by the small molecules targeting several important proteins in cancer development and metastasis.[3, 6, 9-12, 78, 85, 87, 171-175] In another side, it is evident that cancer diagnosis at initial stages of disease development is critical for the therapy.[1, 20, 23, 176-178]. Diverse strategies are applied for the early detection of cancer by searching the cancer biomarkers in the body, yet a very conclusive non-invasive method for the detection is still not met.

Our search of inhibitors targeting epigenetic enzyme G9a yielded a series of compounds as potential inhibitors through structure based virtual screening. Among these compounds, a new G9a inhibitor, DCG066, was confirmed by *in vitro* biochemical and cell-based enzyme assays. DCG066 has a novel molecular scaffold unlike other G9a inhibitors presently available. Like G9a's histone substrate, DCG066 can bind directly to G9a and inhibit methyltransferase activity *in vitro*. DCG066 displays low cytotoxicity in leukemia cell lines with elevated levels of G9a expression, including K562. This work presents DCG066 as an inhibitor of G9a with a novel structure, providing both lead in G9a inhibitor design and a means for probing the functionality of

G9a. We also designed hybrid molecules targeting G9a and HDAC, two epigenetic enzymes involved in many cancers. Detailed SAR study of compounds featuring pharmacophores of both G9a inhibitors and HDAC inhibitors identified compound **13** and **14** with dual inhibition of targets in sub micromolar level. The activity of the compounds confirmed by cellular and biochemical assays and the molecules have drug like physico-chemical properties. Hence has the potential to developed as a lead molecule for targeted therapy of cancer.

We also elaborated here a robust method for the detection of glycan specific antibodies relevant to cancer prognosis. Glycans are an important class of cancer markers and not been used to the potential owing to the complexity in their detection and analysis. We solved this critical issue by conjugating the glycans with a structure-based DNA code and further analyzed the DNA glycan conjugate in the DNA way, i.e. by using the molecular biology methods PCR, qPCR and next generation sequencing. We demonstrated the detection of anti-glycan antibodies in cancer plasm using the DNA encoded glycan library (DEGL) of globo series glycans.

REFERENCES

1. Siegel, R.L., K.D. Miller, and A. Jemal, *Cancer statistics, 2018.* CA: a cancer journal for clinicians, 2018. **68**(1): p. 7-30.
2. Stewart, B. and C.P. Wild, *World cancer report 2014.* Health, 2017.
3. Gerber, D.E., *Targeted therapies: a new generation of cancer treatments.* American family physician, 2008. **77**(3).
4. Egbuchulam, J., *Ekwe Mgba*. viii, 80 pages.
5. Keefe, D.M. and E.H. Bateman, *Tumor control versus adverse events with targeted anticancer therapies.* Nature reviews Clinical oncology, 2012. **9**(2): p. 98.
6. Jones, P.A., J.-P.J. Issa, and S. Baylin, *Targeting the cancer epigenome for therapy.* Nature reviews Genetics, 2016. **17**(10): p. 630.
7. Miozzo, M., V. Vaira, and S.M. Sirchia, *Epigenetic alterations in cancer and personalized cancer treatment.* Future Oncology, 2015. **11**(2): p. 333-348.
8. Kelly, A.D. and J.-P.J. Issa, *The promise of epigenetic therapy: reprogramming the cancer epigenome.* Current opinion in genetics & development, 2017. **42**: p. 68-77.
9. Waldmann, T. and R. Schneider, *Targeting histone modifications—Epigenetics in cancer.* Current opinion in cell biology, 2013. **25**(2): p. 184-189.
10. Vecchio, L., et al., *Importance of epigenetic changes in cancer etiology, pathogenesis, clinical profiling, and treatment: What can be learned from hematologic malignancies?* Biochimica et Biophysica Acta (BBA)-Reviews on Cancer, 2013. **1836**(1): p. 90-104.
11. Kaniskan, H.U.m., M.L. Martini, and J. Jin, *Inhibitors of protein methyltransferases and demethylases.* Chemical reviews, 2017.

12. José-Enériz, E.S., et al., *Dual epigenetic modifiers for cancer therapy.* Molecular & cellular oncology, 2017. **4**(4): p. e1342748.

13. Zhang, X., et al., *The design and synthesis of a new class of RTK/HDAC dual-targeted inhibitors.* Molecules, 2013. **18**(6): p. 6491-503.

14. Seo, S.Y., *Multi-targeted hybrids based on HDAC inhibitors for anti-cancer drug discovery.* Arch Pharm Res, 2012. **35**(2): p. 197-200.

15. West, A.C. and R.W. Johnstone, *New and emerging HDAC inhibitors for cancer treatment.* The Journal of clinical investigation, 2014. **124**(1): p. 30-39.

16. Zhan, P., et al., *Medicinal chemistry insights into novel HDAC inhibitors: an updated patent review (2012-2016).* Recent patents on anti-cancer drug discovery, 2017. **12**(1): p. 16-34.

17. Ferrer-Batalle, M., et al., *Comparative Study of Blood-Based Biomarkers, alpha 2,3-Sialic Acid PSA and PHI, for High-Risk Prostate Cancer Detection.* International Journal of Molecular Sciences, 2017. **18**(4).

18. Pinho, S.S. and C.A. Reis, *Glycosylation in cancer: mechanisms and clinical implications.* Nature Reviews Cancer, 2015. **15**(9): p. 540.

19. Takahashi, M., et al., *Disease-associated glycans on cell surface proteins.* Molecular aspects of medicine, 2016. **51**: p. 56-70.

20. Adamczyk, B., T. Tharmalingam, and P.M. Rudd, *Glycans as cancer biomarkers.* Biochimica et Biophysica Acta (BBA)-General Subjects, 2012. **1820**(9): p. 1347-1353.

21. Miyoshi, E. and M. Nakano, *Fucosylated haptoglobin is a novel marker for pancreatic cancer: detailed analyses of oligosaccharide structures.* Proteomics, 2008. **8**(16): p. 3257-3262.

22. Taniguchi, N. and Y. Kizuka, *Glycans and cancer: role of N-glycans in cancer biomarker, progression and metastasis, and therapeutics*, in *Advances in cancer research*. 2015, Elsevier. p. 11-51.

23. Pedersen, J.W., et al., *Cancer-associated autoantibodies to MUC1 and MUC4—A blinded case–control study of colorectal cancer in UK collaborative trial of ovarian cancer screening.* International journal of cancer, 2014. **134**(9): p. 2180-2188.

24. Pochechueva, T., et al., *Naturally occurring anti-glycan antibodies binding to Globo H-expressing cells identify ovarian cancer patients.* J Ovarian Res, 2017. **10**(1): p. 8.

25. Cheng, S.P., et al., *Aberrant expression of tumor-associated carbohydrate antigen Globo H in thyroid carcinoma.* J Surg Oncol, 2016. **114**(7): p. 853-858.

26. Wang, C.C., et al., *Glycan microarray of Globo H and related structures for quantitative analysis of breast cancer.* Proc Natl Acad Sci U S A, 2008. **105**(33): p. 11661-6.

27. Kouzarides, T., *Chromatin modifications and their function.* Cell, 2007. **128**(4): p. 693-705.

28. Barski, A., et al., *High-resolution profiling of histone methylations in the human genome.* Cell, 2007. **129**(4): p. 823-37.

29. Cao, R., et al., *Role of histone H3 lysine 27 methylation in Polycomb-group silencing.* Science, 2002. **298**(5595): p. 1039-43.

30. Kirmizis, A., et al., *Silencing of human polycomb target genes is associated with methylation of histone H3 Lys 27.* Genes Dev, 2004. **18**(13): p. 1592-605.

31. Schotta, G., et al., *A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin.* Genes Dev, 2004. **18**(11): p. 1251-62.

32. Shilatifard, A., *Molecular implementation and physiological roles for histone H3 lysine 4*

(H3K4) methylation. Curr Opin Cell Biol, 2008. **20**(3): p. 341-8.

33. Ruthenburg, A.J., C.D. Allis, and J. Wysocka, *Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark.* Mol Cell, 2007. **25**(1): p. 15-30.

34. Evertts, A.G., et al., *H4K20 methylation regulates quiescence and chromatin compaction.* Mol Biol Cell, 2013. **24**(19): p. 3025-37.

35. Huang, J., et al., *G9a and Glp methylate lysine 373 in the tumor suppressor p53.* J Biol Chem, 2010. **285**(13): p. 9636-41.

36. Chen, M.W., et al., *H3K9 histone methyltransferase G9a promotes lung cancer invasion and metastasis by silencing the cell adhesion molecule Ep-CAM.* Cancer Res, 2010. **70**(20): p. 7830-40.

37. Russo, V., et al., *H3K9 trimethylation precedes DNA methylation during sheep oogenesis: HDAC1, SUV39H1, G9a, HP1, and Dnmts are involved in these epigenetic events.* J Histochem Cytochem, 2013. **61**(1): p. 75-89.

38. Feldman, N., et al., *G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis.* Nat Cell Biol, 2006. **8**(2): p. 188-94.

39. Epsztejn-Litman, S., et al., *De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes.* Nat Struct Mol Biol, 2008. **15**(11): p. 1176-83.

40. Ushijima, Y., et al., *Roles of histone H3K9 methyltransferases during Drosophila spermatogenesis.* Chromosome Res, 2012. **20**(3): p. 319-31.

41. Shi, Y., et al., *Induction of pluripotent stem cells from mouse embryonic fibroblasts by Oct4 and Klf4 with small-molecule compounds.* Cell Stem Cell, 2008. **3**(5): p. 568-74.

42. Shi, Y., et al., *A combined chemical and genetic approach for the generation of induced pluripotent stem cells.* Cell Stem Cell, 2008. **2**(6): p. 525-8.

43. Mezentseva, N.V., et al., *The histone methyltransferase inhibitor BIX01294 enhances the cardiac potential of bone marrow cells.* Stem Cells Dev, 2013. **22**(4): p. 654-67.

44. Feng, B., et al., *Molecules that promote or enhance reprogramming of somatic cells to induced pluripotent stem cells.* Cell stem cell, 2009. **4**(4): p. 301-312.

45. Vedadi, M., et al., *A chemical probe selectively inhibits G9a and GLP methyltransferase activity in cells.* Nat Chem Biol, 2011. **7**(8): p. 566-74.

46. Lehnertz, B., et al., *The methyltransferase G9a regulates HoxA9-dependent transcription in AML.* Genes Dev, 2014. **28**(4): p. 317-27.

47. Yuan, Y., et al., *A small-molecule probe of the histone methyltransferase G9a induces cellular senescence in pancreatic adenocarcinoma.* ACS chemical biology, 2012. **7**(7): p. 1152-1157.

48. Iwasa, E., et al., *Total synthesis of (+)-chaetocin and its analogues: their histone methyltransferase G9a inhibitory activity.* J Am Chem Soc, 2010. **132**(12): p. 4078-9.

49. Chaib, H., et al., *Anti-leukemia activity of chaetocin via death receptor-dependent apoptosis and dual modulation of the histone methyl-transferase SUV39H1.* Leukemia, 2012. **26**(4): p. 662-74.

50. Kubicek, S., et al., *Reversal of H3K9me2 by a small-molecule inhibitor for the G9a histone methyltransferase.* Mol Cell, 2007. **25**(3): p. 473-81.

51. Liu, F., et al., *Discovery of a 2,4-diamino-7-aminoalkoxyquinazoline as a potent and selective inhibitor of histone lysine methyltransferase G9a.* J Med Chem, 2009. **52**(24): p. 7950-3.

52. Liu, F., et al., *Optimization of cellular activity of G9a inhibitors 7-aminoalkoxy-quinazolines.* J Med Chem, 2011. **54**(17): p. 6139-50.

53. Liu, F., et al., *Discovery of an in Vivo Chemical Probe of the Lysine Methyltransferases G9a and GLP.* J Med Chem, 2013.

54. Sweis, R.F., et al., *Discovery and Development of Potent and Selective Inhibitors of Histone Methyltransferase G9a.* ACS Medicinal Chemistry Letters, 2014: p. 140110075408001.

55. Liu, Y., et al., *High throughput enzyme inhibitor screening by functionalized magnetic carbonaceous microspheres and graphene oxide-based MALDI-TOF-MS.* Journal of The American Society for Mass Spectrometry, 2011. **22**(12): p. 2188-2198.

56. Li, K.K., et al. *DNA methyltransferases in hematologic malignancies*. in *Seminars in hematology*. 2013. Elsevier.

57. Kunze, M.B., et al., *Loop interactions and dynamics tune the enzymatic activity of the human histone deacetylase 8.* J Am Chem Soc, 2013. **135**(47): p. 17862-8.

58. Karlsson, R., *SPR for molecular interaction analysis: a review of emerging application areas.* J Mol Recognit, 2004. **17**(3): p. 151-61.

59. Jun Matsui, K.A., Noriaki Hara, Daisuke Miyoshi, Hidemi Nawafune, Katsuyuki Tamaki, and Naoki Sugimoto, *SPR Sensor Chip for Detection of Small Molecules Using Molecularly Imprinted Polymer with Embedded Gold Nanoparticles.* Anal Chem, 2005. **77**(13): p. 4282-4285.

60. Wang, J., et al., *Pharmacophore-based virtual screening and biological evaluation of small molecule inhibitors for protein arginine methylation.* J Med Chem, 2012. **55**(18): p. 7978-87.

61. Barretina, J., et al., *The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity.* Nature, 2012. **483**(7391): p. 603-307.

62. Shin, G., et al., *GENT: gene expression database of normal and tumor tissues.* Cancer informatics, 2011. **10**: p. 149-57.

63. Farooq, S., S. Trah, and A. Jeanguenat, *Derivatives of (1-benzyl-piperidine-4-yl)-diphenyl-methanol and their use as pesticide*. 2003, Google Patents.

64. Lagu, B. and M. Wachter, *Quaternary salt CCR2 antagonists*. 2010, Google Patents.

65. Friesner, R.A., et al., *Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy.* J Med Chem, 2004. **47**(7): p. 1739-1749.

66. Jain, A.N., *Surflex-Dock 2.1: robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search.* J. Comput. Aided Mol. Des., 2007. **21**(5): p. 281-306.

67. Leng, T.D., et al., *Amiloride Analogs as ASIC1a Inhibitors.* CNS Neurosci. Ther., 2016.

68. OpenEye Scientific Software, I., Santa Fe, NM, USA, www.eyesopen.com, *OEChem*. 2014.

69. Andrew P. Butler$, C.V.B., and Thomas J. Slaga, *Phosphorylation of Histones Is Stimulated by Phorbol Esterisn Quiescent Reuber H35 Hepatoma Cells.* J Biol Chem, 1986. **Vol . 261**: p. 9421-9425.

70. Arrowsmith, C.H., et al., *Epigenetic protein families: a new frontier for drug discovery.* Nature reviews Drug discovery, 2012. **11**(5): p. 384-400.

71. Black, J.C., C. Van Rechem, and J.R. Whetstine, *Histone lysine methylation dynamics: establishment, regulation, and biological impact.* Molecular cell, 2012. **48**(4): p. 491-507.

72. Krishnan, S., S. Horowitz, and R.C. Trievel, *Structure and function of histone H3 lysine 9 methyltransferases and demethylases.* Chembiochem, 2011. **12**(2): p. 254-263.

73. Ohzeki, J.i., et al., *Breaking the HAC Barrier: Histone H3K9 acetyl/methyl balance regulates CENP-A assembly.* The EMBO journal, 2012. **31**(10): p. 2391-2402.

74. Liu, F., et al., *Discovery of an in vivo Chemical Probe of the Lysine Methyltransferases G9a and GLP.* Journal of medicinal chemistry, 2013. **56**(21): p. 8931-8942.

75. Dokmanovic, M., C. Clarke, and P.A. Marks, *Histone deacetylase inhibitors: overview and perspectives.* Molecular cancer research, 2007. **5**(10): p. 981-989.

76. Inoue, S., et al., *Apoptosis induced by histone deacetylase inhibitors in leukemic cells is mediated by Bim and Noxa.* Leukemia, 2007. **21**(8): p. 1773-1782.

77. Xu, W.S., R.B. Parmigiani, and P.A. Marks, *Histone deacetylase inhibitors: molecular mechanisms of action.* Oncogene, 2007. **26**(37): p. 5541-52.

78. Copeland, R., M. Moyer, and V. Richon, *Targeting genetic alterations in protein methyltransferases for personalized cancer therapeutics.* Oncogene, 2012. **32**(8): p. 939-946.

79. Chen, J.B., et al., *Design and synthesis of dual-action inhibitors targeting histone deacetylases and 3-hydroxy-3-methylglutaryl coenzyme A reductase for cancer treatment.* J Med Chem, 2013. **56**(9): p. 3645-55.

80. Medina-Franco, J.L., et al., *Shifting from the single to the multitarget paradigm in drug discovery.* Drug Discov Today, 2013. **18**(9-10): p. 495-501.

81. Morphy, R. and Z. Rankovic, *Designed multiple ligands. An emerging drug discovery paradigm.* Journal of medicinal chemistry, 2005. **48**(21): p. 6523-6543.

82. Zheng, H., M. Fridkin, and M. Youdim, *From single target to multitarget/network*

therapeutics in Alzheimer's therapy. Pharmaceuticals (Basel), 2014. **7**(2): p. 113-35.

83. Zimmermann, G.R., J. Lehar, and C.T. Keith, *Multi-target therapeutics: when the whole is greater than the sum of the parts.* Drug Discov Today, 2007. **12**(1-2): p. 34-42.

84. Liu, X., et al., *Gain-of-function mutations of Ptpn11 (Shp2) cause aberrant mitosis and increase susceptibility to DNA damage-induced malignancies.* Proc Natl Acad Sci U S A, 2016. **113**(4): p. 984-9.

85. Ko, K.S., et al., *Development of a chimeric c-Src kinase and HDAC inhibitor.* ACS Med Chem Lett, 2013. **4**(8): p. 779-783.

86. Olson, D.E., et al., *Discovery of the first histone deacetylase 6/8 dual inhibitors.* J Med Chem, 2013. **56**(11): p. 4816-20.

87. Beckers, T., et al., *Chimerically designed HDAC- and tyrosine kinase inhibitors. A series of erlotinib hybrids as dual-selective inhibitors of EGFR, HER2 and histone deacetylases.* MedChemComm, 2012. **3**(7): p. 829.

88. Cai, X., et al., *Discovery of 7-(4-(3-ethynylphenylamino)-7-methoxyquinazolin-6-yloxy)-N-hydroxyheptanamide (CUDc-101) as a potent multi-acting HDAC, EGFR, and HER2 inhibitor for the treatment of cancer.* J Med Chem, 2010. **53**(5): p. 2000-9.

89. Tran, H.T., et al., *Improved therapeutic effect against leukemia by a combination of the histone methyltransferase inhibitor chaetocin and the histone deacetylase inhibitor trichostatin A.* J Korean Med Sci, 2013. **28**(2): p. 237-46.

90. Batty, N., G.G. Malouf, and J.P. Issa, *Histone deacetylase inhibitors as anti-neoplastic agents.* Cancer Lett, 2009. **280**(2): p. 192-200.

91. Botrugno, O.A., F. Santoro, and S. Minucci, *Histone deacetylase inhibitors as a new weapon in the arsenal of differentiation therapies of cancer.* Cancer Lett, 2009. **280**(2): p.

134-44.

92. Duan, H., C.A. Heckman, and L.M. Boxer, *Histone deacetylase inhibitors down-regulate bcl-2 expression and induce apoptosis in t (14; 18) lymphomas.* Molecular and cellular biology, 2005. **25**(5): p. 1608-1619.

93. Lai, M.J., et al., *Synthesis and biological evaluation of 1-arylsulfonyl-5-(N-hydroxyacrylamide)indoles as potent histone deacetylase inhibitors with antitumor activity in vivo.* J Med Chem, 2012. **55**(8): p. 3777-91.

94. Luchenko, V.L., et al., *Histone deacetylase inhibitor-mediated cell death is distinct from its global effect on chromatin.* Mol Oncol, 2014.

95. Wu, H., et al., *Structural biology of human H3K9 methyltransferases.* PloS one, 2010. **5**(1): p. e8570.

96. Vedadi, M., et al., *A chemical probe selectively inhibits G9a and GLP methyltransferase activity in cells.* Nature chemical biology, 2011. **7**(8): p. 566-574.

97. Spannhoff, A., et al., *The emerging therapeutic potential of histone methyltransferase and demethylase inhibitors.* ChemMedChem, 2009. **4**(10): p. 1568-1582.

98. Sweis, R.F., et al., *Discovery and Development of Potent and Selective Inhibitors of Histone Methyltransferase G9a.* ACS Medicinal Chemistry Letters, 2014. **5**(2): p. 205-209.

99. Huang, J., et al., *G9a and Glp methylate lysine 373 in the tumor suppressor p53.* Journal of Biological Chemistry, 2010. **285**(13): p. 9636-9641.

100. Xu, W., R. Parmigiani, and P. Marks, *Histone deacetylase inhibitors: molecular mechanisms of action.* Oncogene, 2007. **26**(37): p. 5541-5552.

101. Wang, H., et al., *Discovery of (2E)-3-{2-butyl-1-[2-(diethylamino)ethyl]-1H-*

benzimidazol-5-yl}-N-hydroxyacrylami de (SB939), an orally active histone deacetylase inhibitor with a superior preclinical profile. J Med Chem, 2011. **54**(13): p. 4694-720.

102. Krejci, J., et al., *Genome-wide reduction in H3K9 acetylation during human embryonic stem cell differentiation.* J Cell Physiol, 2009. **219**(3): p. 677-87.

103. Hou, J., et al., *Discovery and extensive in vitro evaluations of NK-HDAC-1: a chiral histone deacetylase inhibitor as a promising lead.* J Med Chem, 2012. **55**(7): p. 3066-75.

104. Frew, A.J., R.W. Johnstone, and J.E. Bolden, *Enhancing the apoptotic and therapeutic effects of HDAC inhibitors.* Cancer letters, 2009. **280**(2): p. 125-133.

105. Ververis, K., et al., *Histone deacetylase inhibitors (HDACIs): multitargeted anticancer agents.* Biologics: targets & therapy, 2013. **7**: p. 47.

106. Meunier, B., *Hybrid Molecules with a Dual Mode of Action: Dream or Reality?†.* Accounts of chemical research, 2007. **41**(1): p. 69-77.

107. Wang, J., et al., *Potential advantages of CUDC-101, a multitargeted HDAC, EGFR, and HER2 inhibitor, in treating drug resistance and preventing cancer cell migration and invasion.* Mol Cancer Ther, 2013. **12**(6): p. 925-36.

108. Chen, L., et al., *Dual inhibitors of inosine monophosphate dehydrogenase and histone deacetylases for cancer treatment.* Journal of medicinal chemistry, 2007. **50**(26): p. 6685-6691.

109. Zang, L.-L., et al., *SAHA-based novel HDAC inhibitor design by core hopping method.* Journal of Molecular Graphics and Modelling, 2014. **54**: p. 10-18.

110. Konze, K.D., et al., *A Chemical Tool for In Vitro and In Vivo Precipitation of Lysine Methyltransferase G9a.* ChemMedChem, 2014. **9**(3): p. 549-553.

111. Liu, F., et al., *Discovery of a 2, 4-diamino-7-aminoalkoxyquinazoline as a potent and*

selective inhibitor of histone lysine methyltransferase G9a. Journal of medicinal chemistry, 2009. **52**(24): p. 7950-7953.

112. Van Horn, K.S., et al., *Antileishmanial activity of a series of N(2),N(4)-disubstituted quinazoline-2,4-diamines.* J Med Chem, 2014. **57**(12): p. 5141-56.

113. Chang, Y., et al., *Structural basis for G9a-like protein lysine methyltransferase inhibition by BIX-01294.* Nature structural & molecular biology, 2009. **16**(3): p. 312-317.

114. Li, X.B., et al., *Novel inhibitor design for hemagglutinin against H1N1 influenza virus by core hopping method.* PLoS One, 2011. **6**(11): p. e28111.

115. Li, X., et al., *Study of SHP-2 (PTPN11) allosterism on structural movement using solution perturbed molecular dynamics simulation.* Journal of Molecular Liquids, 2016. **223**: p. 509-515.

116. Suite, S., *Virtual Screening Workflow; Glide version 2014-3.* New York, NY, 2014.

117. Somoza, J.R., et al., *Structural snapshots of human HDAC8 provide insights into the class I histone deacetylases.* Structure, 2004. **12**(7): p. 1325-1334.

118. Zang, L.L., et al., *SAHA-based novel HDAC inhibitor design by core hopping method.* Journal of Molecular Graphics & Modelling, 2014. **54**: p. 10-18.

119. Sun, S.X., et al., *Design, synthesis, biological activity and molecular dynamics studies of specific protein tyrosine phosphatase 1B inhibitors over SHP-2.* Int J Mol Sci, 2013. **14**(6): p. 12661-74.

120. Lipinski, C.A., et al., *Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings.* Adv Drug Deliv Rev, 2001. **46**(1-3): p. 3-26.

121. McOmie, J.F., M. Watts, and D.E. West, *Demethylation of aryl methyl ethers by boron*

tribromide. Tetrahedron, 1968. **24**(5): p. 2289-2292.

122. Ciossek, T., et al., *A homogeneous cellular histone deacetylase assay suitable for compound profiling and robotic screening.* Anal Biochem, 2008. **372**(1): p. 72-81.

123. Chevolot, Y., et al., *DNA-based carbohydrate biochips: a platform for surface glyco-engineering.* Angew Chem Int Ed Engl, 2007. **46**(14): p. 2398-402.

124. Fernandez-Tejada, A., F.J. Canada, and J. Jimenez-Barbero, *Recent Developments in Synthetic Carbohydrate-Based Diagnostics, Vaccines, and Therapeutics.* Chemistry, 2015. **21**(30): p. 10616-28.

125. Wang, D., et al., *Carbohydrate microarrays for the recognition of cross-reactive molecular markers of microbes and host cells.* Nature biotechnology, 2002. **20**(3): p. 275-281.

126. Horlacher, T. and P.H. Seeberger, *Carbohydrate arrays as tools for research and diagnostics.* Chem Soc Rev, 2008. **37**(7): p. 1414-22.

127. Song, X., et al., *Glycan microarrays of fluorescently-tagged natural glycans.* Glycoconjugate journal, 2015. **32**(7): p. 465-473.

128. Pochechueva, T., et al., *Multiplex suspension array for human anti-carbohydrate antibody profiling.* Analyst, 2011. **136**(3): p. 560-569.

129. Pochechueva, T., et al., *Comparison of printed glycan array, suspension array and ELISA in the detection of human anti-glycan antibodies.* Glycoconj J, 2011. **28**(8-9): p. 507-17.

130. Purohit, S., et al., *Multiplex glycan bead array for high throughput and high content analyses of glycan binding proteins.* Nature Communications, 2018. **9**.

131. Brenner, S. and R.A. Lerner, *Encoded combinatorial chemistry.* Proceedings of the National Academy of Sciences, 1992. **89**(12): p. 5381-5383.

132. Litovchick, A., et al., *Encoded Library Synthesis Using Chemical Ligation and the Discovery of sEH Inhibitors from a 334-Million Member Library.* Sci Rep, 2015. **5**: p. 10916.

133. Kleiner, R.E., C.E. Dumelin, and D.R. Liu, *Small-molecule discovery from DNA-encoded chemical libraries.* Chemical Society Reviews, 2011. **40**(12): p. 5707.

134. Lerner, R.A. and S. Brenner, *DNA-Encoded Compound Libraries as Open Source: A Powerful Pathway to New Drugs.* Angewandte Chemie International Edition, 2017. **56**(5): p. 1164-1165.

135. Goodnow, R.A., C.E. Dumelin, and A.D. Keefe, *DNA-encoded chemistry: enabling the deeper sampling of chemical space.* Nature Reviews Drug Discovery, 2017. **16**(2): p. 131-147.

136. Kwon, S.J., et al., *Signal Amplification by Glyco-qPCR for Ultrasensitive Detection of Carbohydrates: Applications in Glycobiology.* Angewandte Chemie-International Edition, 2012. **51**(47): p. 11800-11804.

137. Kwon, S.J., et al., *High sensitivity detection of active botulinum neurotoxin by glyco-quantitative polymerase chain-reaction.* Anal Chem, 2014. **86**(5): p. 2279-84.

138. Thomas, B., et al., *Application of Biocatalysis to on-DNA Carbohydrate Library Synthesis.* Chembiochem, 2017. **18**(9): p. 858-863.

139. Shinmachi, D., et al., *Using GlyTouCan Version 1.0: The First International Glycan Structure Repository*, in *A Practical Guide to Using Glycomics Databases*. 2017, Springer. p. 41-73.

140. Huang, C.Y., et al., *Carbohydrate microarray for profiling the antibodies interacting with Globo H tumor antigen.* Proc Natl Acad Sci U S A, 2006. **103**(1): p. 15-20.

141. Wu, K., et al., *Electrophoretic deposition of poly [3-(3-N, N-diethylaminopropoxy) thiophene] and composite films.* Materials Chemistry and Physics, 2011. **125**(1): p. 210-218.

142. Liao, S.F., et al., *Immunization of fucose-containing polysaccharides from Reishi mushroom induces antibodies to tumor-associated Globo H-series epitopes.* Proc Natl Acad Sci U S A, 2013. **110**(34): p. 13809-14.

143. Danishefsky, S.J., et al., *Development of Globo-H Cancer Vaccine.* Accounts of Chemical Research, 2015. **48**(3): p. 643-652.

144. Zhou, Z., et al., *A Fully Synthetic Self-Adjuvanting Globo H-Based Vaccine Elicited Strong T Cell-Mediated Antitumor Immunity.* Chem Sci, 2015. **6**(12): p. 7112-7121.

145. O'Cearbhaill, R.E., et al., *A Phase I Study of Unimolecular Pentavalent (Globo-H-GM2-sTn-TF-Tn) Immunization of Patients with Epithelial Ovarian, Fallopian Tube, or Peritoneal Cancer in First Remission.* Cancers (Basel), 2016. **8**(4).

146. Kornfeld, S., E. Li, and I. Tabas, *The synthesis of complex-type oligosaccharides. II. Characterization of the processing intermediates in the synthesis of the complex oligosaccharide units of the vesicular stomatitis virus G protein.* Journal of Biological Chemistry, 1978. **253**(21): p. 7771-7778.

147. D-glucan, A., *Polysaccharide nomenclature.* Eur. J. Biochem, 1982. **126**: p. 439-441.

148. Harvey, D.J., et al., *Symbol nomenclature for representing glycan structures: Extension to cover different carbohydrate types.* Proteomics, 2011. **11**(22): p. 4291-4295.

149. Varki, A., et al., *Symbol nomenclature for glycan representation.* Proteomics, 2009. **9**(24): p. 5398-5399.

150. Chang, W.W., et al., *Expression of Globo H and SSEA3 in breast cancer stem cells and*

the involvement of fucosyl transferases 1 and 2 in Globo H synthesis. Proc Natl Acad Sci U S A, 2008. **105**(33): p. 11667-72.

151. Pochechueva, T., et al., *Naturally occurring anti-glycan antibodies binding to Globo H-expressing cells identify ovarian cancer patients.* Journal of Ovarian Research, 2017. **10**.

152. Kudryashov, V., et al., *Characterization of a mouse monoclonal IgG3 antibody to the tumor-associated globo H structure produced by immunization with a synthetic glycoconjugate.* Glycoconj J, 1998. **15**(3): p. 243-9.

153. Alam, K.K., J.L. Chang, and D.H. Burke, *FASTAptamer: A Bioinformatic Toolkit for High-throughput Sequence Analysis of Combinatorial Selections.* Mol Ther Nucleic Acids, 2015. **4**: p. e230.

154. Purohit, S., et al., *Multiplex glycan bead array for high throughput and high content analyses of glycan binding proteins.* Nature communications, 2018. **9**(1): p. 258.

155. Wu, C.S., et al., *Cancer-Associated Carbohydrate Antigens as Potential Biomarkers for Hepatocellular Carcinoma.* Plos One, 2012. **7**(7).

156. Jacob, F., et al., *Serum antiglycan antibody detection of nonmucinous ovarian cancers by using a printed glycan array.* International Journal of Cancer, 2012. **130**(1): p. 138-146.

157. Wang, L., et al., *Cross-platform comparison of glycan microarray formats.* Glycobiology, 2014. **24**(6): p. 507-517.

158. Chan, A.I., L.M. McGregor, and D.R. Liu, *Novel selection methods for DNA-encoded chemical libraries.* Current opinion in chemical biology, 2015. **26**: p. 55-61.

159. McGregor, L.M., T. Jain, and D.R. Liu, *Identification of ligand-target pairs from combined libraries of small molecules and unpurified protein targets in cell lysates.* J Am Chem Soc, 2014. **136**(8): p. 3264-70.

160. Li, G., et al., *Photoaffinity labeling of small-molecule-binding proteins by DNA-templated chemistry.* Angew Chem Int Ed Engl, 2013. **52**(36): p. 9544-9.

161. Zhao, G., et al., *Enzymatic route to preparative-scale synthesis of UDP-GlcNAc/GalNAc, their analogues and GDP-fucose.* Nat Protoc, 2010. **5**(4): p. 636-46.

162. Muthana, M.M., et al., *Efficient one-pot multienzyme synthesis of UDP-sugars using a promiscuous UDP-sugar pyrophosphorylase from Bifidobacterium longum (BLUSP).* Chemical Communications, 2012. **48**(21): p. 2728-2730.

163. Lau, K., et al., *Highly efficient chemoenzymatic synthesis of β1–4-linked galactosides with promiscuous bacterial β1–4-galactosyltransferases.* Chemical Communications, 2010. **46**(33): p. 6066-6068.

164. Li, Y., et al., *Donor substrate promiscuity of bacterial β1–3-N-acetylglucosaminyltransferases and acceptor substrate flexibility of β1–4-galactosyltransferases.* Bioorganic & medicinal chemistry, 2016. **24**(8): p. 1696-1705.

165. Su, D.M., et al., *Enzymatic synthesis of tumor-associated carbohydrate antigen Globo-H hexasaccharide.* Org Lett, 2008. **10**(5): p. 1009-12.

166. Heidtman, M.I., M. Merighi, and J.M. McCoy, *Alpha (1, 2) fucosyltransferases suitable for use in the production of fucosylated oligosaccharides*. 2015, Google Patents.

167. Seto, N.O., et al., *Enzymatic synthesis of blood group A and B trisaccharide analogues.* Carbohydr Res, 2000. **324**(3): p. 161-9.

168. Kaneda, A. *Cancer epigenome and its application to diagnosis and therapy*. in *CANCER SCIENCE*. 2018. WILEY 111 RIVER ST, HOBOKEN 07030-5774, NJ USA.

169. Weinstein, I.B. and A.K. Joe, *Mechanisms of disease: oncogene addiction—a rationale for molecular targeting in cancer therapy.* Nature Reviews Clinical Oncology, 2006.

3(8): p. 448.

170. Chari, R.V., *Targeted cancer therapy: conferring specificity to cytotoxic drugs.* Accounts of chemical research, 2007. **41**(1): p. 98-107.

171. Coussens, L.M., B. Fingleton, and L.M. Matrisian, *Matrix metalloproteinase inhibitors and cancer—trials and tribulations.* Science, 2002. **295**(5564): p. 2387-2392.

172. McArthur, H., *An overview of HER-targeted therapy with lapatinib in breast cancer.* Advances in Therapy, 2009. **26**(3): p. 263-271.

173. Shangary, S. and S. Wang, *Small-molecule inhibitors of the MDM2-p53 protein-protein interaction to reactivate p53 function: a novel approach for cancer therapy.* Annual review of pharmacology and toxicology, 2009. **49**: p. 223-241.

174. Zhang, J., P.L. Yang, and N.S. Gray, *Targeting cancer with small molecule kinase inhibitors.* Nature Reviews Cancer, 2009. **9**(1): p. 28.

175. You, J.S. and P.A. Jones, *Cancer genetics and epigenetics: two sides of the same coin?* Cancer cell, 2012. **22**(1): p. 9-20.

176. Salgado, R., et al., *Tumor-infiltrating lymphocytes and associations with pathological complete response and event-free survival in HER2-positive early-stage breast cancer treated with lapatinib and trastuzumab: a secondary analysis of the NeoALTTO trial.* JAMA oncology, 2015. **1**(4): p. 448-455.

177. Saadatmand, S., et al., *Influence of tumour stage at breast cancer detection on survival in modern times: population based study in 173 797 patients.* Bmj, 2015. **351**: p. h4901.

178. Moyer, V.A., *Screening for lung cancer: US Preventive Services Task Force recommendation statement.* Annals of internal medicine, 2014. **160**(5): p. 330-338.

APPENDICES

Appendix A Supporting information for Discovery of novel small molecule inhibitors of lysine methyltransferase G9a and their mechanism in leukemia cell lines

Appendix A.1. *Characterization of compounds*

HPLC (DCG066)

Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	17.930	MM T	0.1374	784.53436	95.15216	1.5957
2	19.566	MM T	0.2819	4.67274e4	2762.76855	95.0414
3	20.776	MM T	0.1540	1653.40002	178.89224	3.3629
Totals :				4.91653e4	3036.81296	

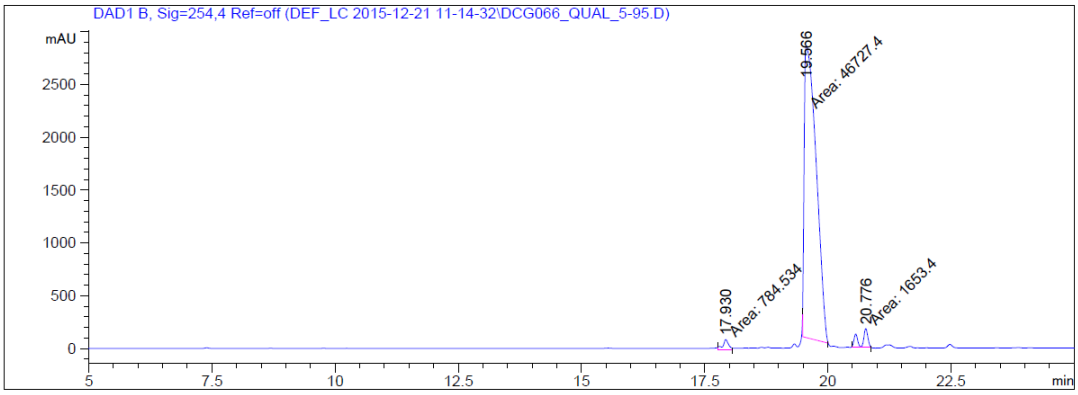
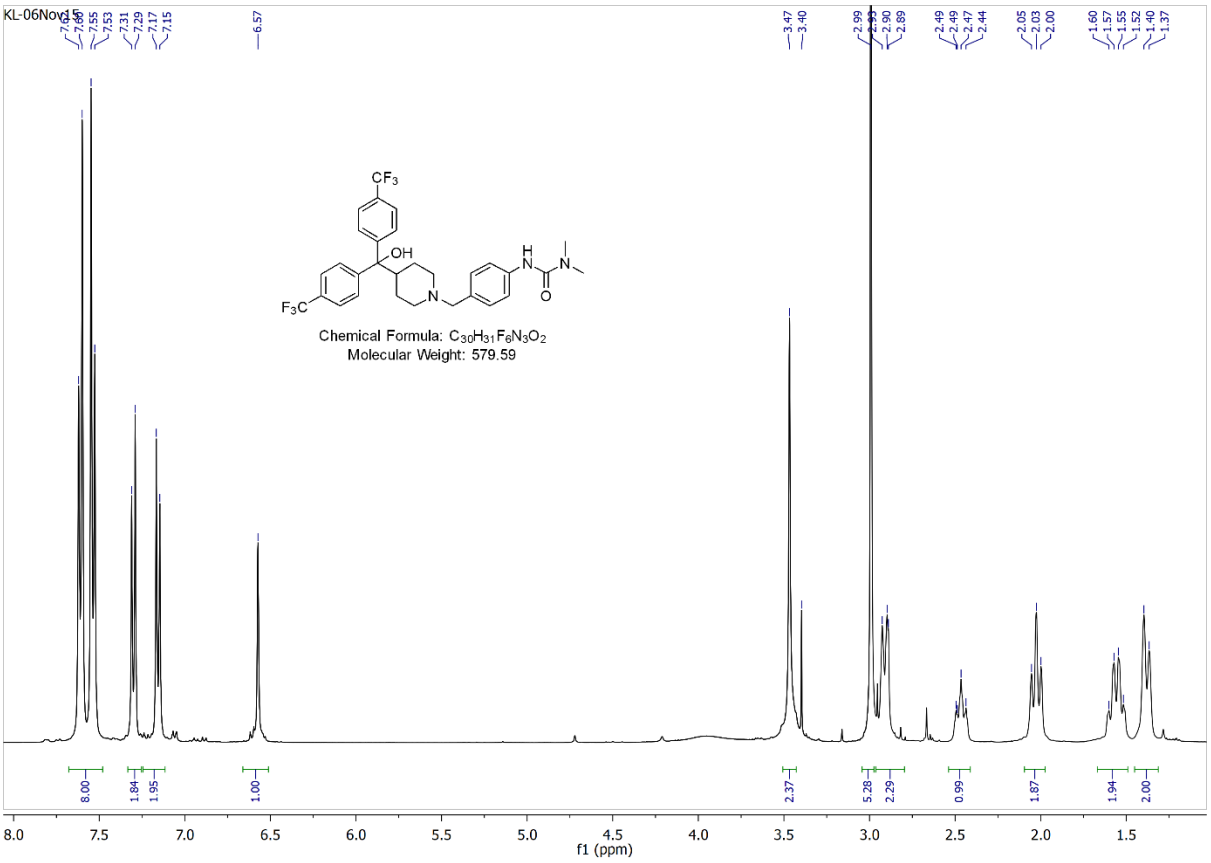
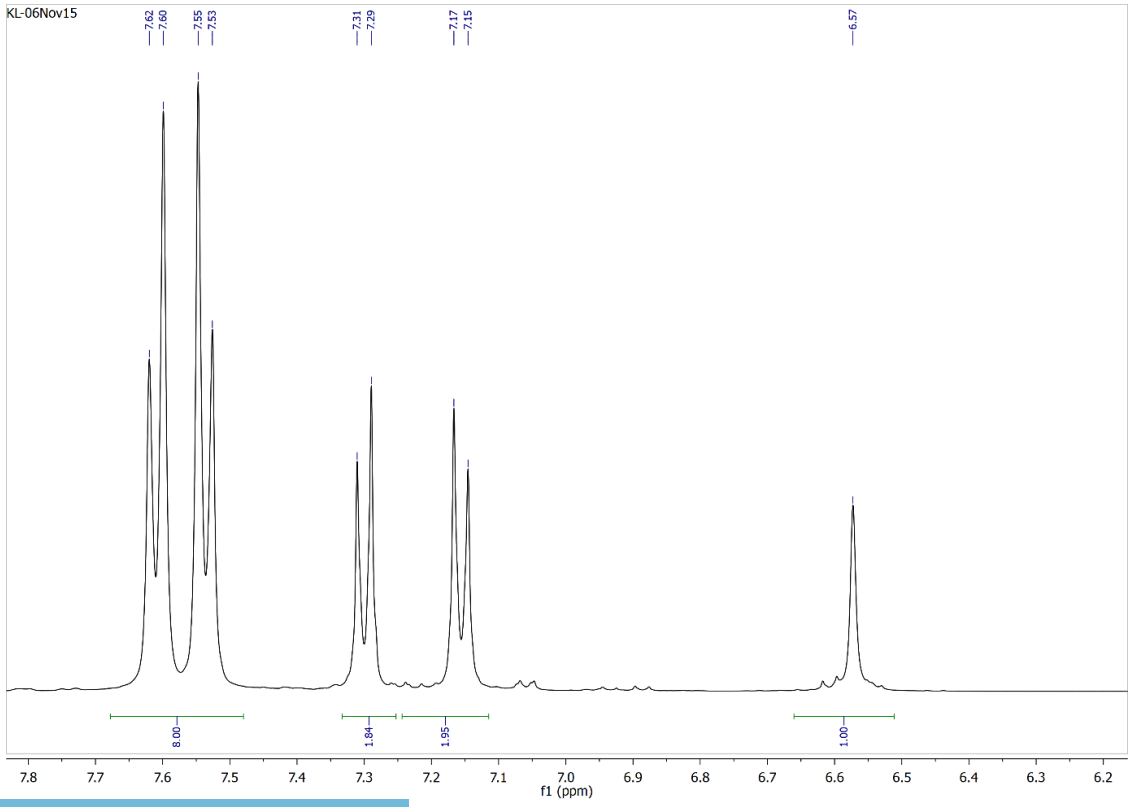
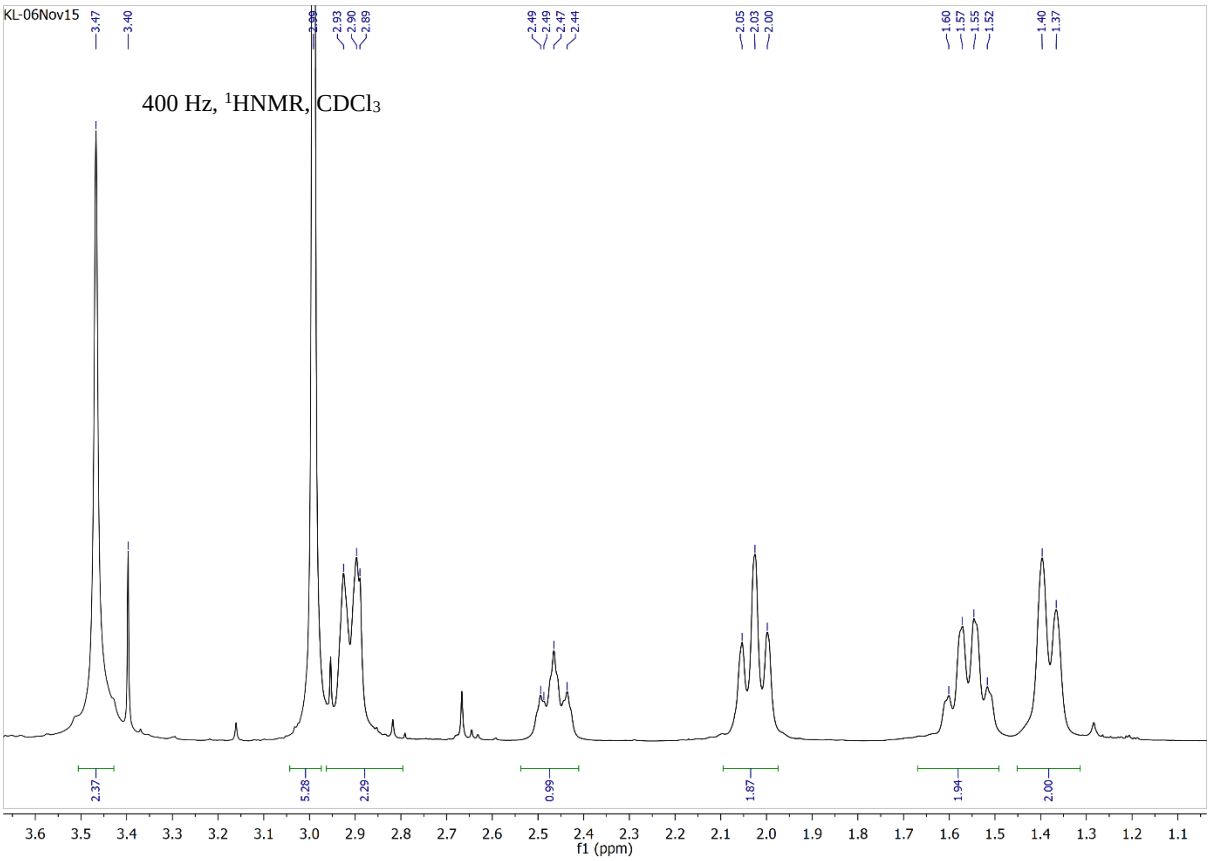


Figure A.1. HPLC of DCG066

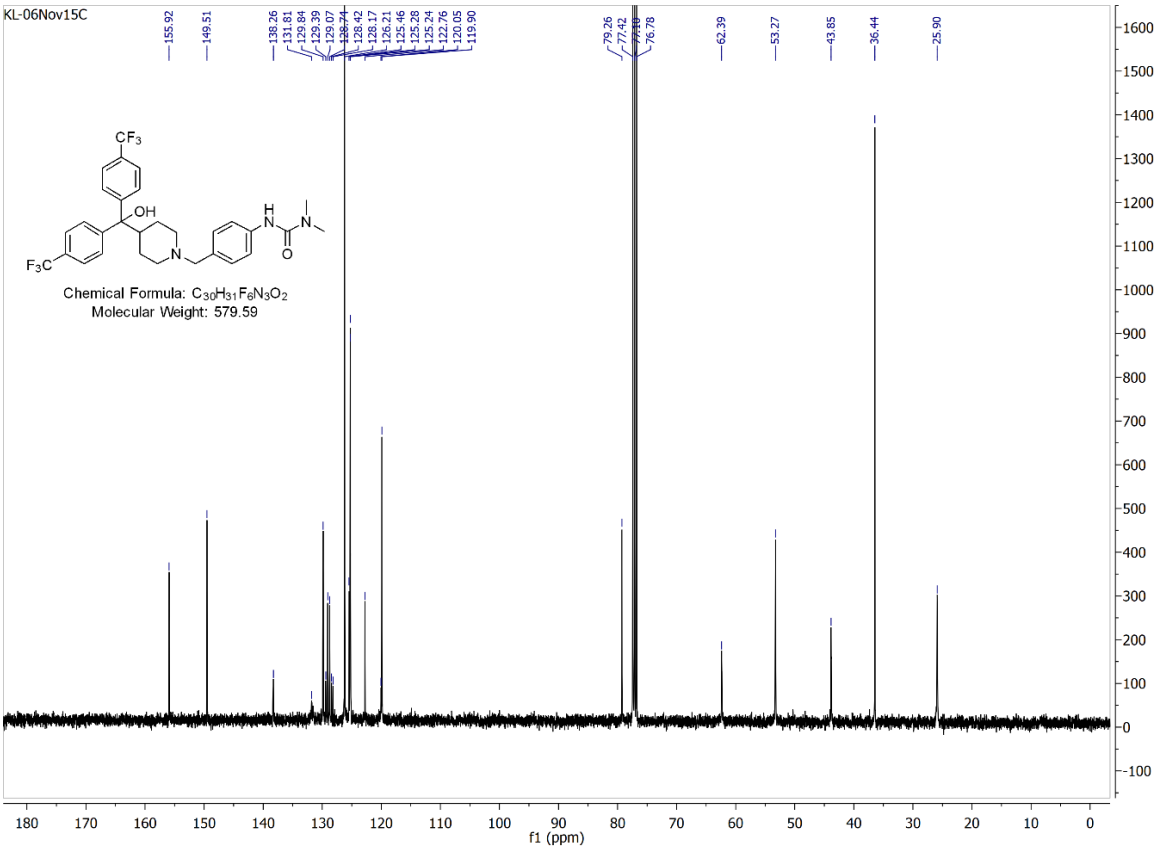
^{1}H NMR of DCG066

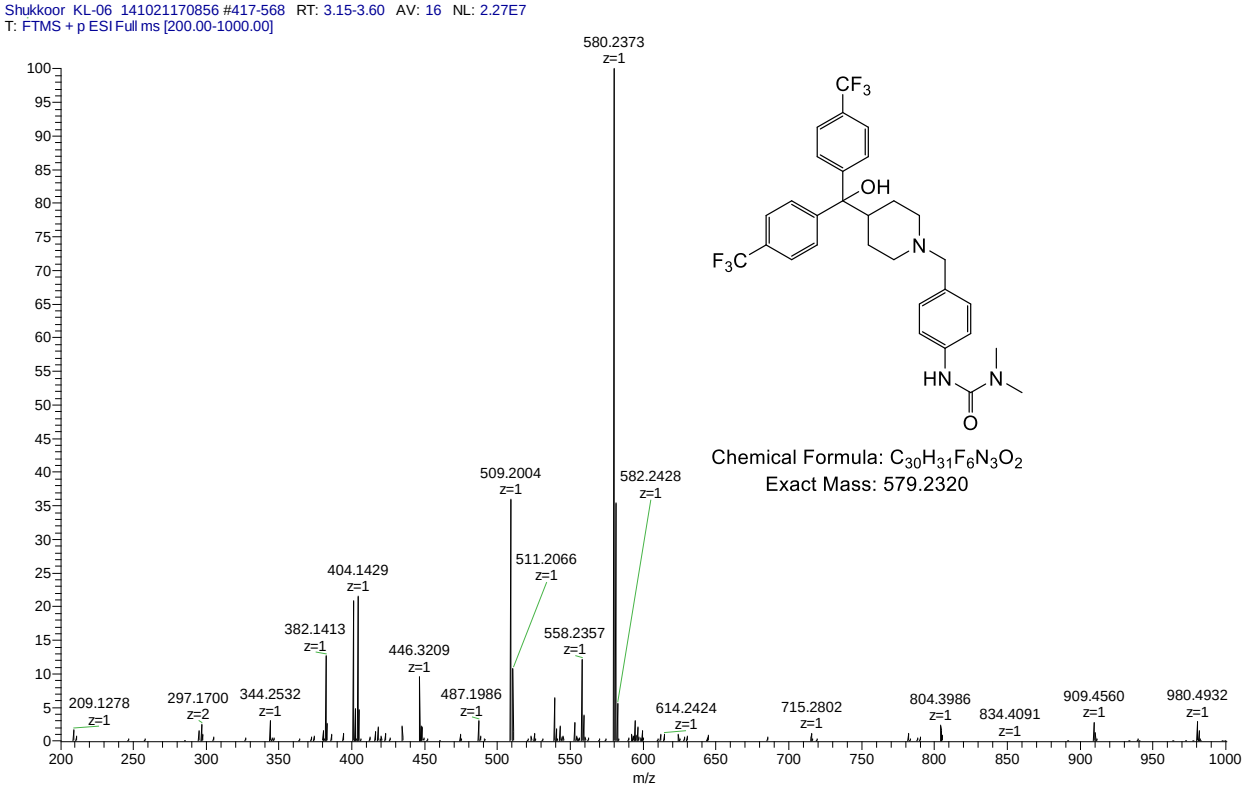






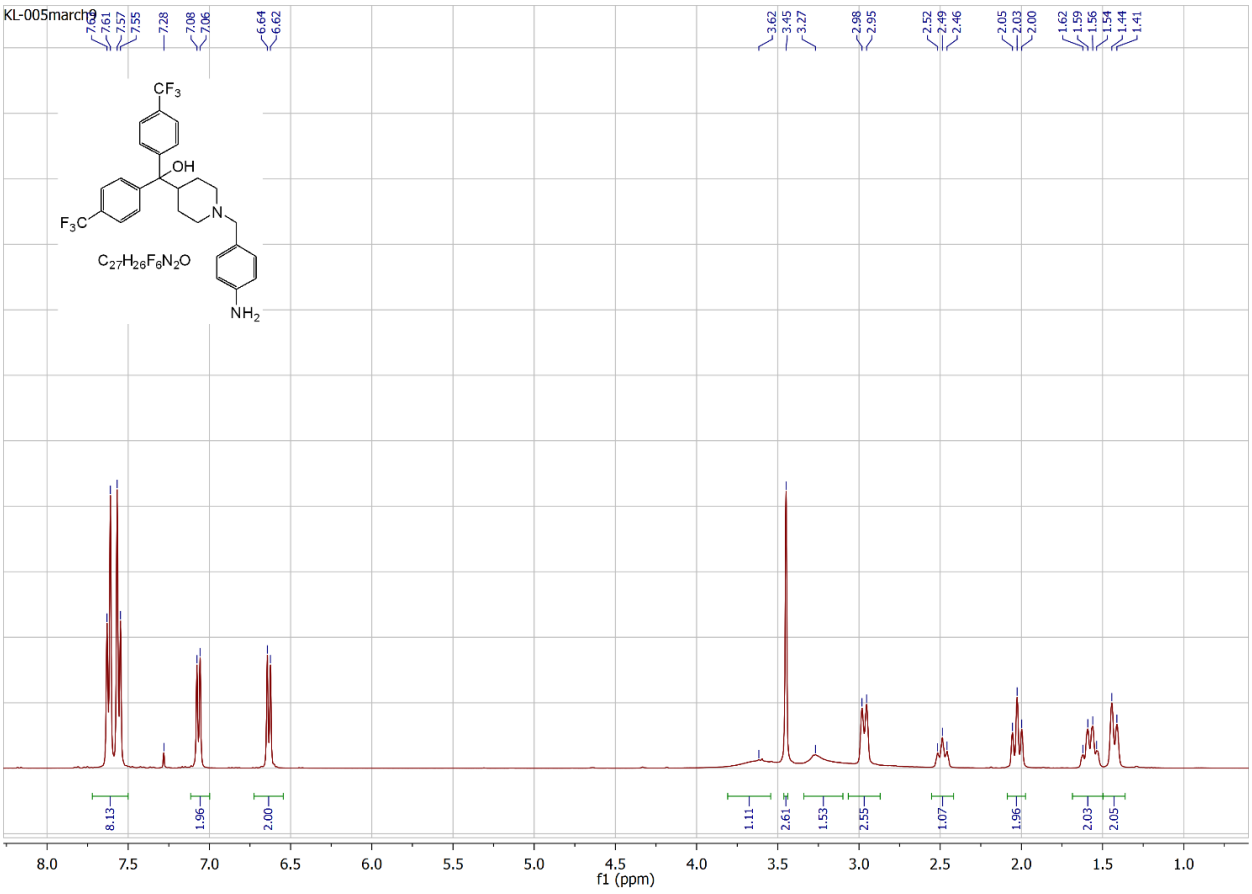
^{13}C NMR of DCG066



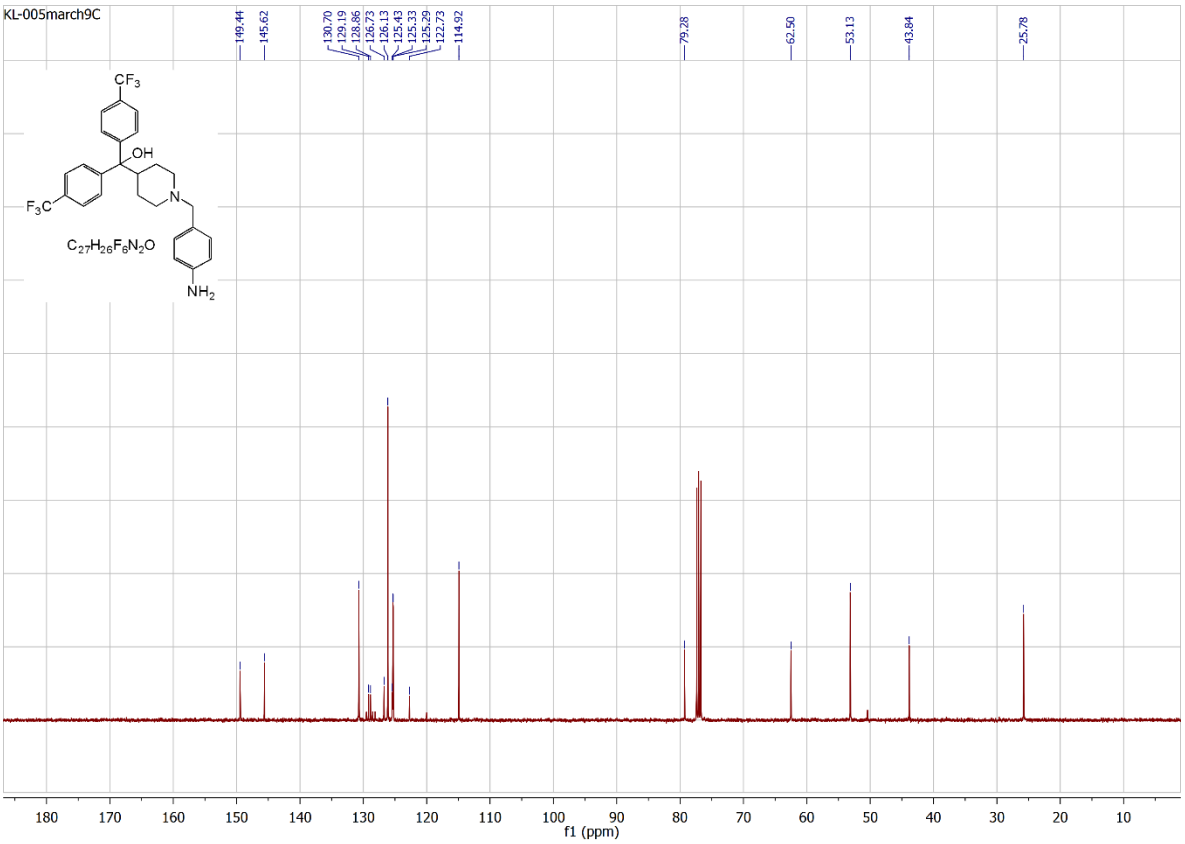


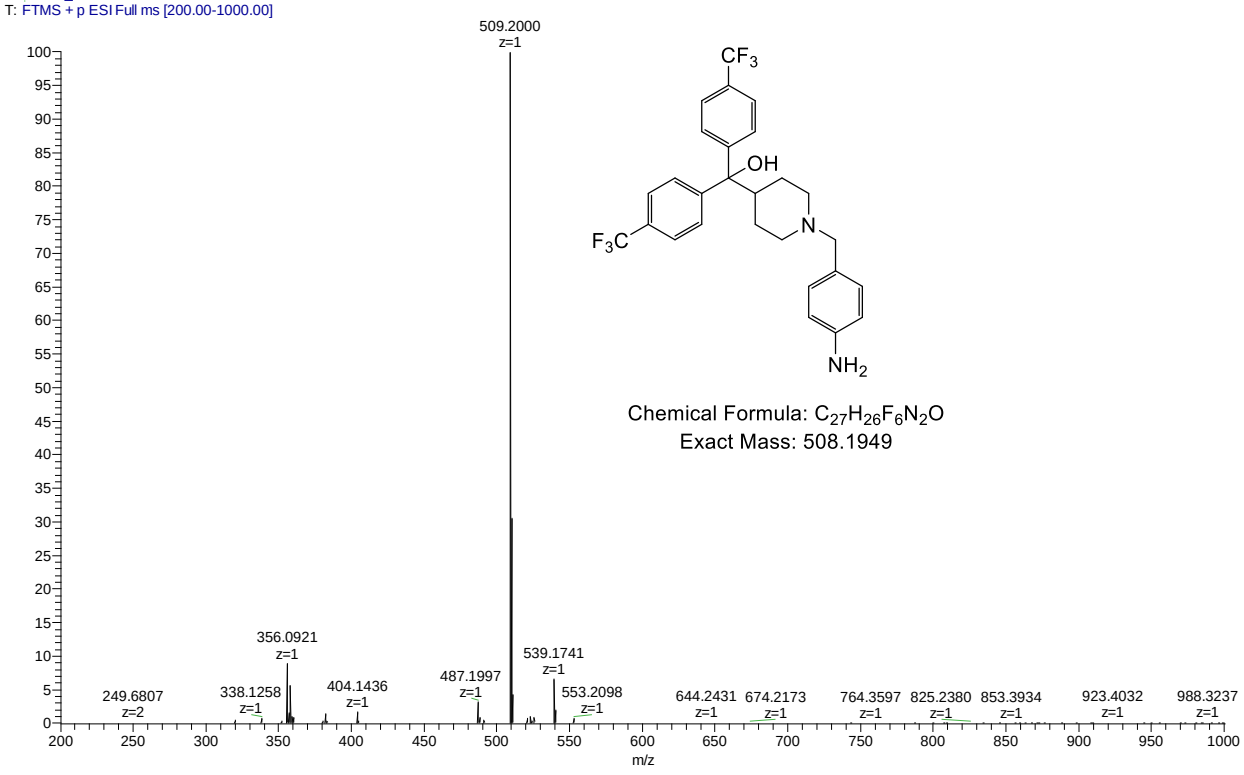
Spectral data of compound 7

^{1}H NMR



^{13}C NMR





Chemical Formula: $C_{27}H_{26}F_6N_2O$
Exact Mass: 508.1949

Appendix B. Supporting information for Structure based design, synthesis and activity studies of small hybrid molecules as HDAC and G9a dual inhibitors

Appendix B.1. Supporting information for assays

MALDI-TOF study of methylation

MALDI-TOF based experiment was performed according to the protocol developed by Chang etal.[113] MALDI spectrums were collected using Bruker flex control software and analyzed by flex analysis. After labelling each cluster peaks of H3K9Me0, H3K9Me1 and H3K9Me2 for all of the tested concentrations, area under the cluster (AUC) were extracted by using the same flex analysis software. % abundance of each peak was calculated by following formula,

A = % Abundance of (H3K9Me0) = area of H3K9Me0/ (area of H3K9Me0+area of H3K9Me1 + area of H3K9Me2)

B = % Abundance of (H3K9Me1) = area of H3K9Me1/ (area of H3K9Me0+area of H3K9Me1 + area of H3K9Me2)

C = % Abundance of (H3K9Me2) = area of H3K9Me2/ (area of H3K9Me0+area of H3K9Me1 + area of H3K9Me2).

This was repeated for each spectra (3 multiples for each samples).

G9a catalyze dimethylation of H3K9 and hence formation of H3K9Me2 was considered as the product formation and H3K9Me0 and H3K9Me1 is considered substrate not modified to the final product. Hence here % conversion to product is also C, from this to get the %maximal activity (%MA), C was compared to the % conversion when no inhibitor was used (D).

Finally % inhibition was found by subtracting $\%MA_{(i)}$ from 100

Example: % methylation levels and % inhibition when used 5 µM compound **14**

A = 13.09

B = 61.18

C = 25.72

% conversion when no inhibitor was used D = 84.43 (average of six measurements)

%MA = 100 * 25.72/84.43

= 30.46

Finally % inhibition = 100 – 30.46

= **69.54%**

Average of 3 values were reported in the Figure S1 and Table 1.

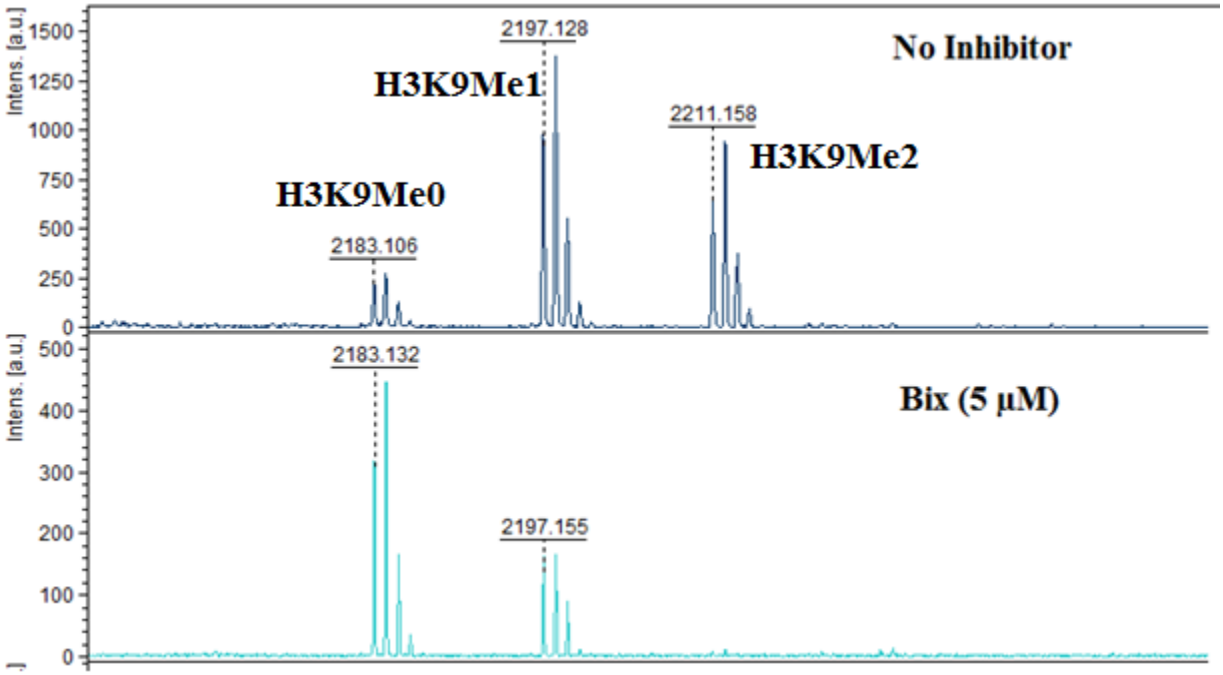


Figure B.1. A sample MALDI-TOF image used for the % inhibition analysis

Table B.1. *MALDI-TOF methylation study of inhibitors at 5 µM concentration for 30 min*

CPD#	Inhibition (%)	**CPD#**	Inhibition (%)
4	26.96±12.15	**4a**	11.22±3.68
5	65.02±1.95	**5a**	47.55±9.32
6	26.64±11.34	**6a**	50.65±16.51
7	19.58±20.95	**7a**	15.01±15.73
13	24.91±17.42	**13a**	62.00±14.44
14	29.68±23.32	**14a**	69.55±3.43
15	69.12±24.54	**15a**	47.85±12.88
16	40.32±13.81	**16a**	40.32±13.81
19	10.71±25.88	**20**	14.62±15.81
21	22.75±26.72	**22**	40.26±15.33
5b	38.01±4.99	**BIX-01294**	77.68±5.73

HDAC assay results

Table B.2. *Cell based homogenous HDAC assay results*

CPD#	IC_{50}-HDAC		
	Hela[c]	**A549[d]**	**K562[e]**
4	NA[a]	NA	NA
4a	NA	NA	NA
5	NA	NA	NA
5a	NA	NA	NA
6	NA	NA	NA
6a	NA	NA	NA
7	NA	NA	NA
7a	NA	NA	NA
13	15.33±0.79	>100	27.75±0.59
13a	>100	>100	>100
14	13.80±1.22	>100	5.735±1.23
14a	>100[b]	>100	>100
15	>100	>100	>100
15a	>100	>100	>100
16	>100	>100	>100
16a	>100	>100	>100
19	>100	>100	>100
20	>100	>100	>100
21	>100	>100	>100
22	>100	>100	>100
BIX	NA	NA	NA
SAHA	5.044±0.53	>100	2.056±0.59

NA[a] not active up to the highest concentration tested (the highest concentration of all compounds is 100 µM;

>100[b] in the cases where the IC_{50} did not reach at the highest tested concentration (100 µM);

[c]Hela: human cervical cancer cell line; [d]A549: human lung cancer cell line; [e]K562: human immortalized myelogenous leukemia cell line; SAHA was used as the positive control. Data are shown as mean ± SD of triplicate.

Appendix B.2. Molecular docking study results

HDAC (PDB ID:1T69) protein interactions study

Initially we chose HDAC8 protein structure (PDB ID:1T69) for the docking study because it has SAHA (which we used as the control in cell based assays) as the co-crystallized ligand, but our study revealed a lower GLIDE score and docking score than the expected, and so we did a similar study on another HDAC8 protein structure 1T67 and found a higher binding scores and chose this for later study.

Table B.3. *Glide docking study results for compound 14 and SAHA at the catalytic site of HDAC8 (PDB ID: 1T69)*

NO.	Ligand Id	Glide score	Docking score	INTERACTIONS			
				H- bonds		π-π	Interaction with Zn^{2+} atom
				Backbone	Side chain		
1	SAHA	-5.794	-5.794	His142, His143, Gly151, Gly304	Asp101, Tyr306	Phe152	+
2	**14**	-8.858	-8.471	Gly140, His142, Gly151, Gly206, Phe207, Pro209, Gly304	Asp101, Tyr306	-	+

Superimposed images of 14, MS-344 with HDAC8 (PDB ID: 1T67) and 14, BIX-01294 with G9a (PDB ID: 3FPD)

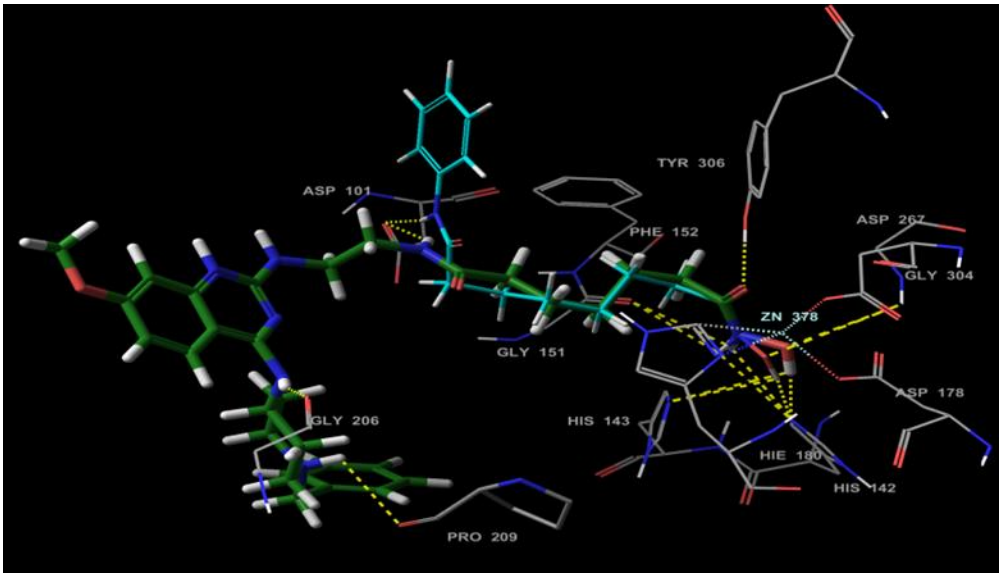
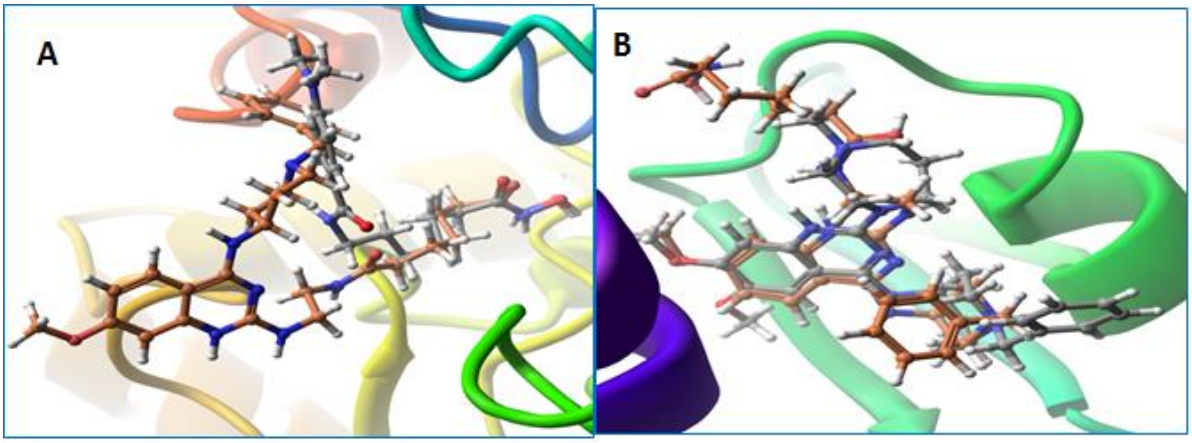


Figure B.2. Binding analysis image of compound 14 and SAHA with HDAC8 (PDB ID: 1T69).



FigureB.3. Super imposed ligands with their corresponding target enzyme.
A) Superimposition of cocrystallized MS-344 (gray), best docked pose of **14** (orange) in the catalytic site of HDAC8 (PDB ID: 1T67), **B)** Superimposition of BIX-01294 (gray), best docked pose of **14** (orange) in the catalytic site of G9a (PDB ID: 3FPD).

Appendix B.3. Cytotoxicity study results

***Table B.4.**. Detailed results of cytotoxicity study*

CPD#	EC_{50}(µM)		
	MDA-MB-231[c]	MCF-7[d]	A549[e]
4	>100[b]	>100	NA[a]
4a	>100	>100	NA
5	>100	>100	NA
5a	>100	>100	NA
6	>100	>100	NA
6a	>100	>100	NA
7	>100	>100	NA
7a	>100	>100	NA
13	89.33±1.23	79.43±2.72	>100
13a	>100	>100	NA
14	10.02±1.66	37.36±2.20	36.24±1.76
14a	82.32	>100	NA
15	95.15	>100	NA
15a	77.62	>100	NA
16	38.15	57.29	>100
16a	90.54	>100	NA
19	>100	>100	NA
20	>100	>100	NA
21	>100	>100	NA
22	>100	>100	NA
26	31.28±3.30	>100	NA
30	24.01±3.64	>100	NA
5b	12.29±3.27	74.57±1.81	NA
BIX-01294	2.155±0.88	8.103±1.99	21.74±2.73
SAHA	2.874±0.84	8.124±4.98	19.31±1.26

NA[a], not active up to the highest concentration tested (the highest concentration of all compounds is 100 µM.

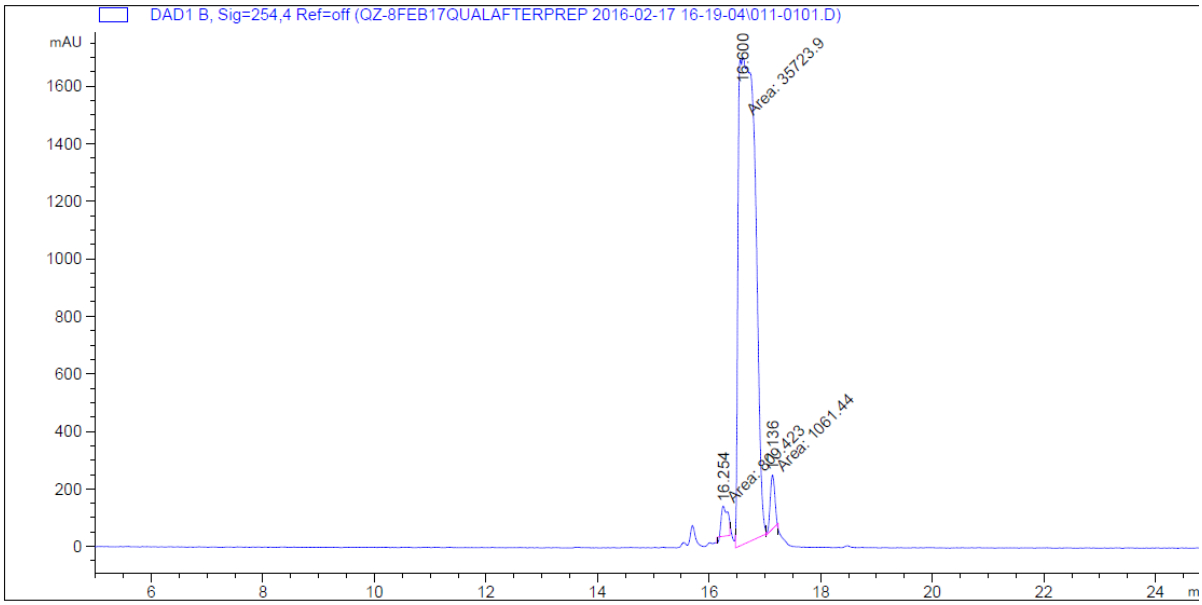
>100[b] in the cases where the IC_{50} did not reach at the highest tested concentration (100 µM).

[c]MDA-MB-231: breast cancer cell line; [d]MCF-7: breast cancer cell line; [e]A549: human lung cancer cell line; SAHA and BIX-01294 are used as the positive controls;

Cells were exposed to the different inhibitors with various concentrations for 72 h, Inhibition of cell growth by the listed compounds was determined by using CCK-8 kit. Data are shown as mean ± SD of triplicate.

Appendix B.4. Mass, NMR, HPLC spectra of compounds

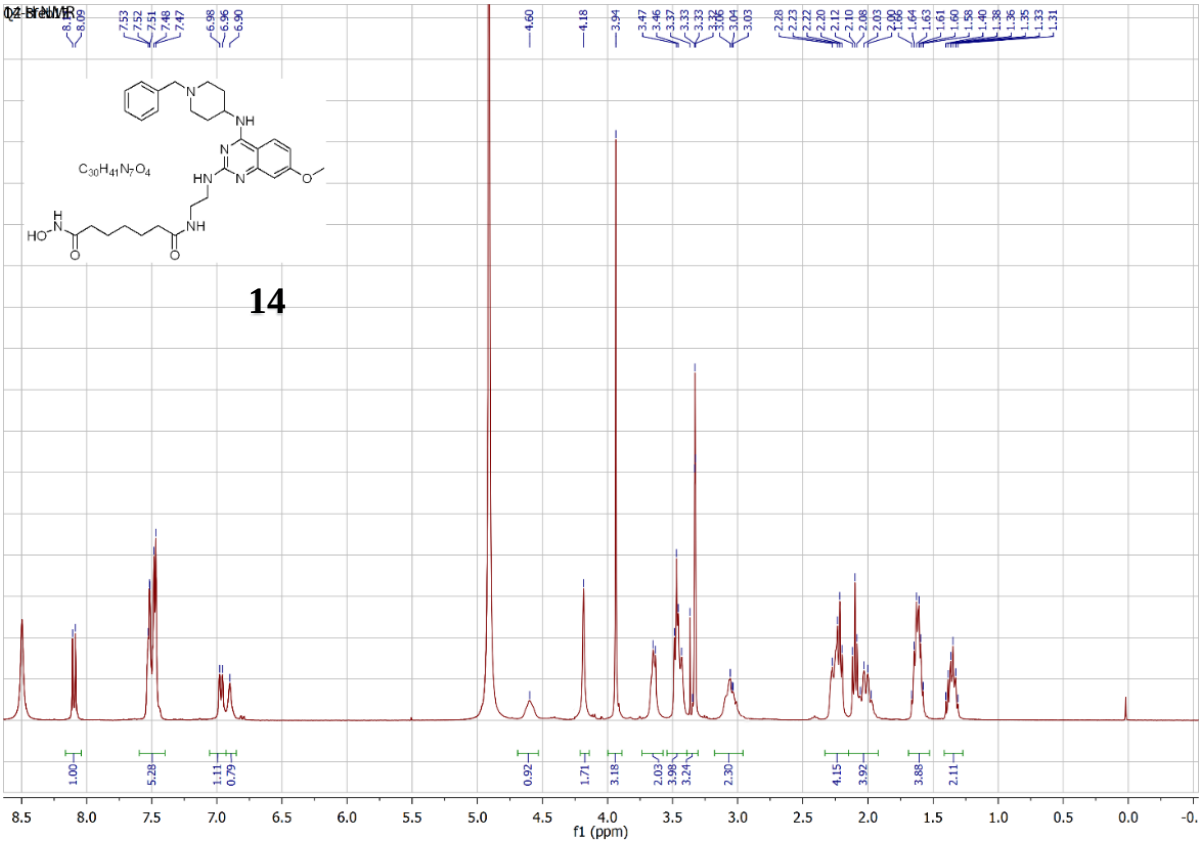
HPLC of compound **14**

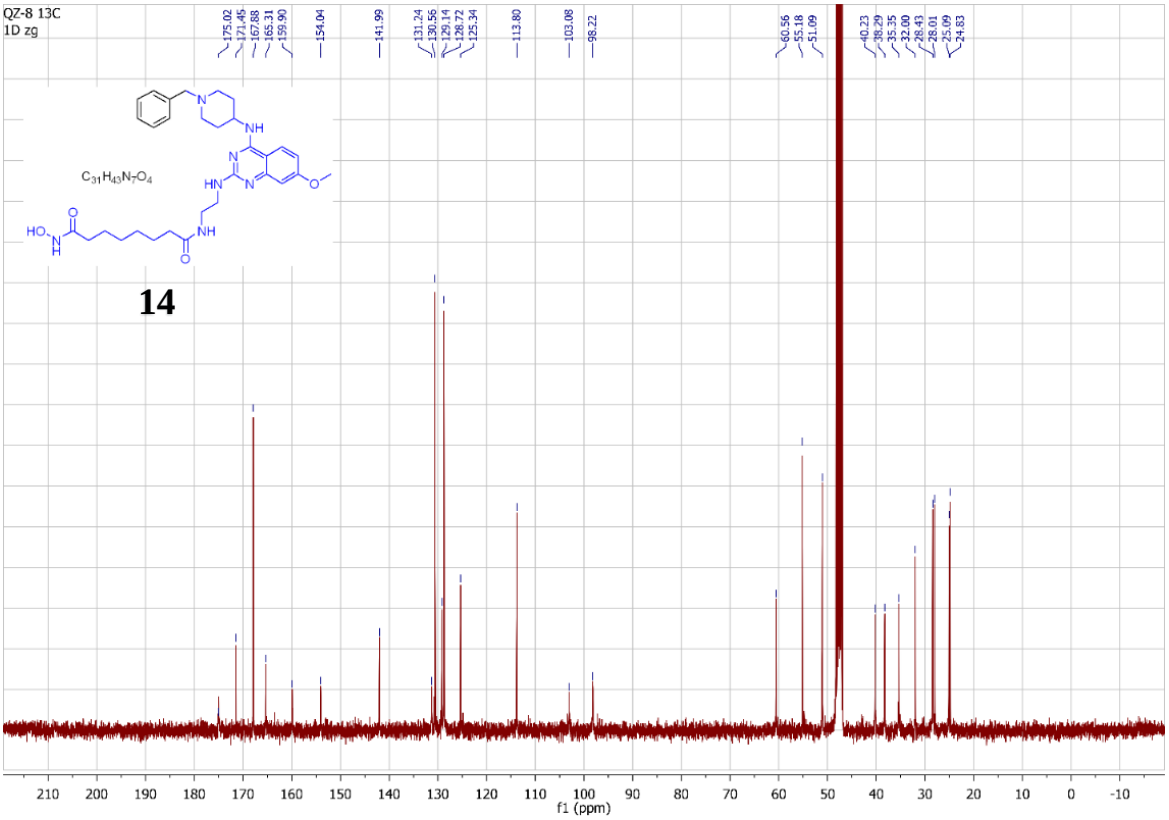


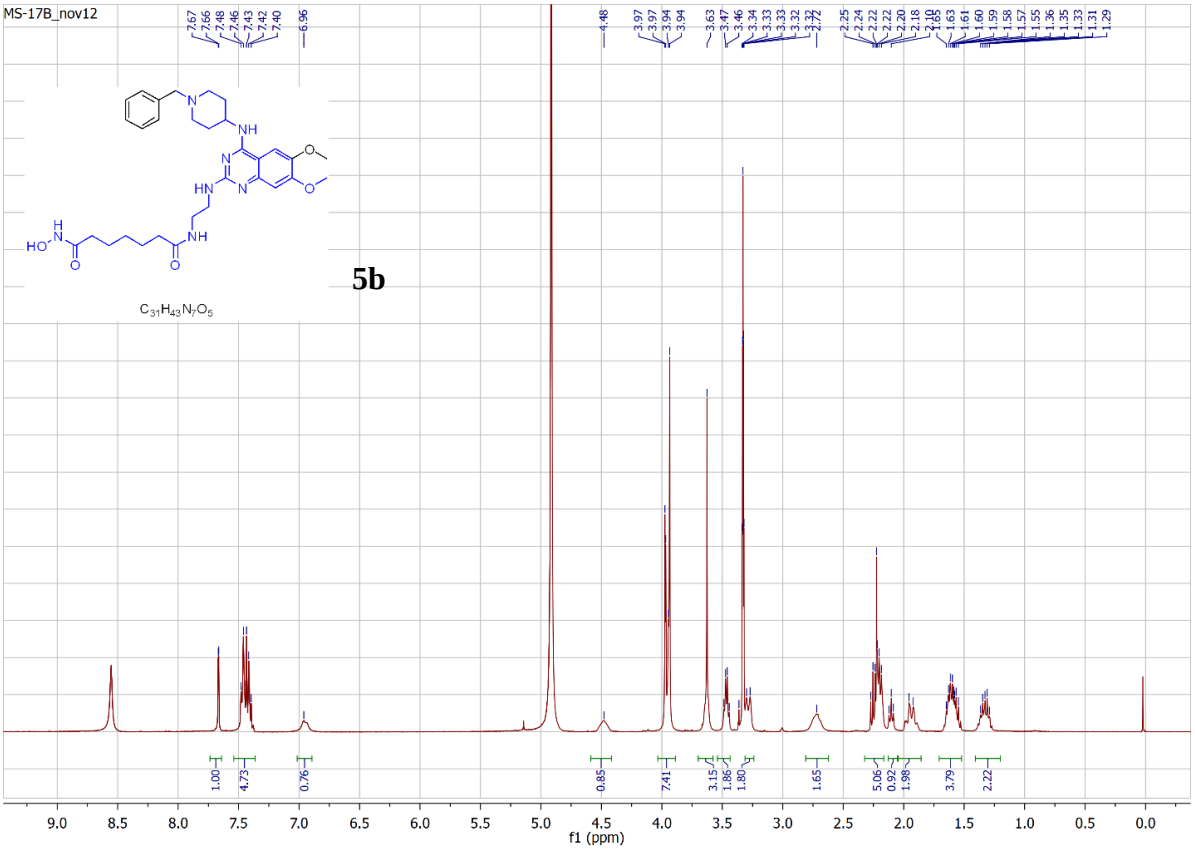
Signal 1: DAD1 B, Sig=254,4 Ref=off

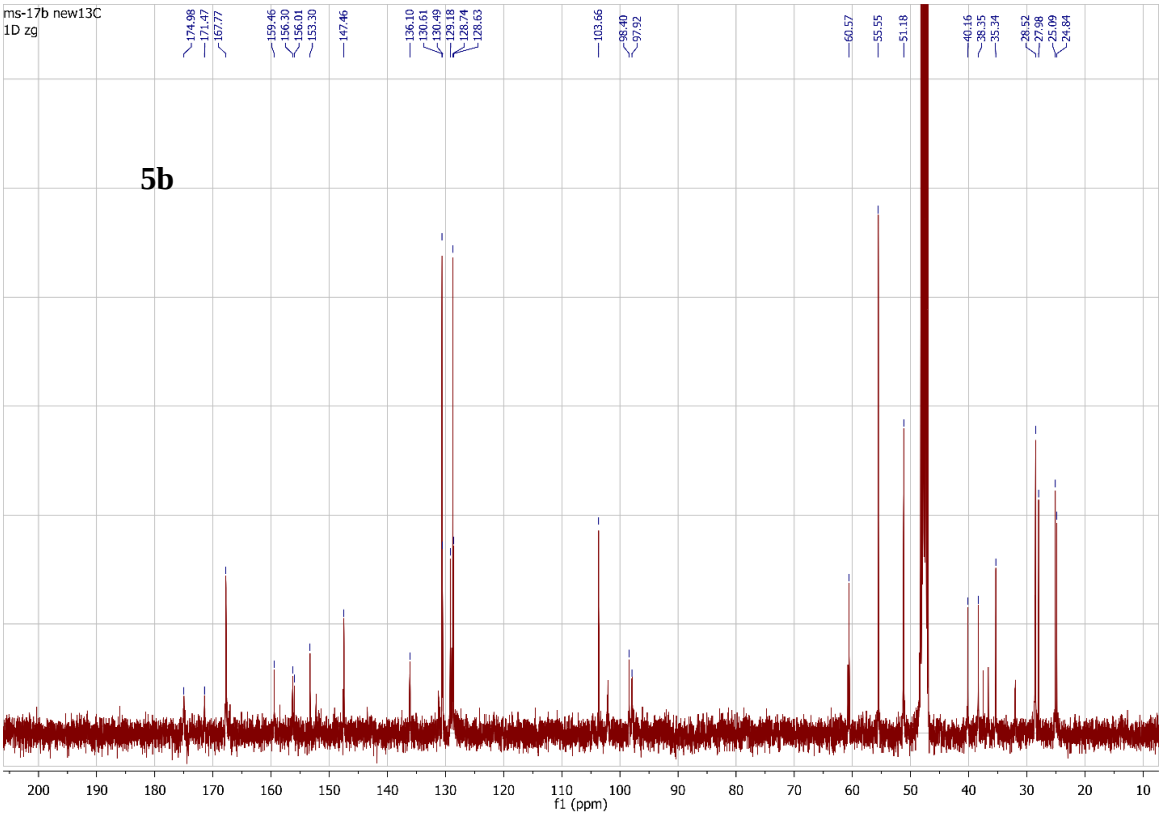
Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	16.254	MM T	0.1279	809.42334	105.45708	2.1530
2	16.600	MM T	0.3513	3.57239e4	1694.89990	95.0236
3	17.136	MM	0.0941	1061.43896	188.08916	2.8234
Totals :				3.75948e4	1988.44614	

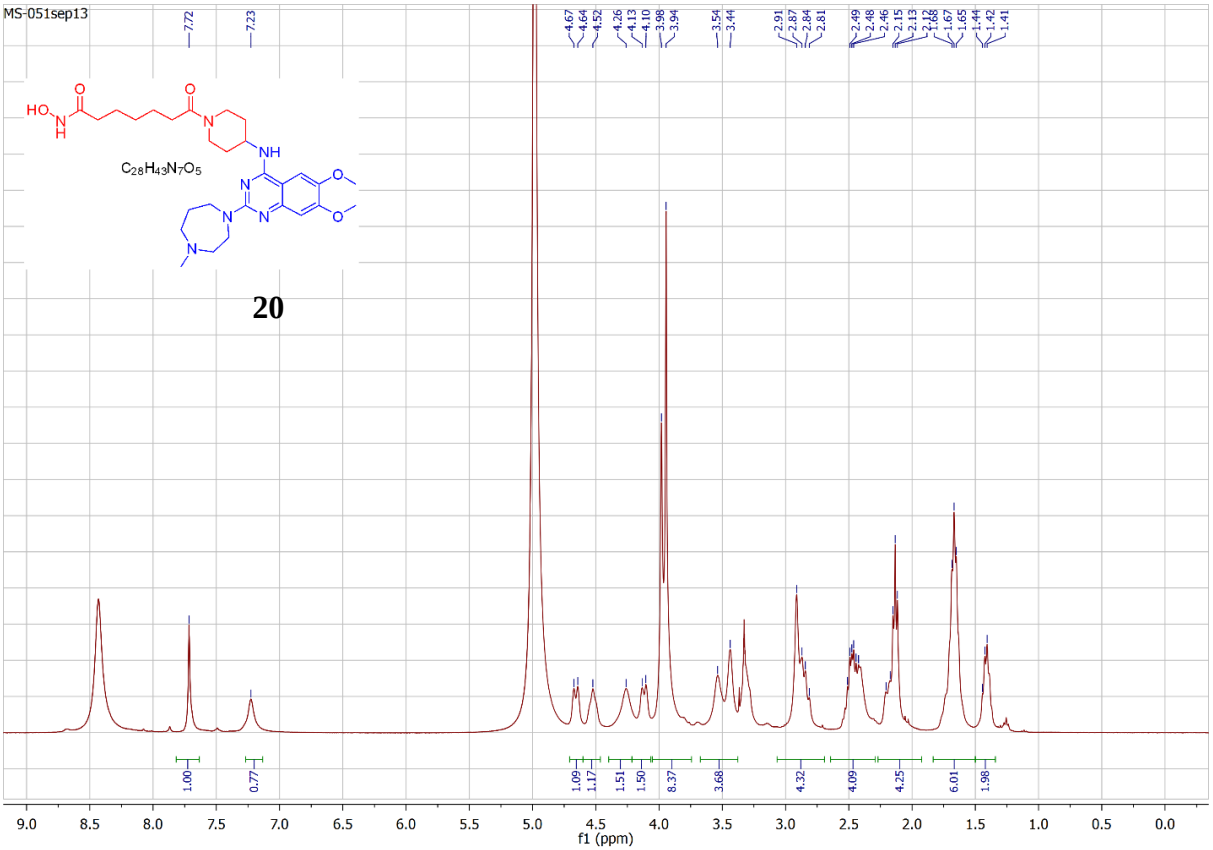
*** End of Report ***



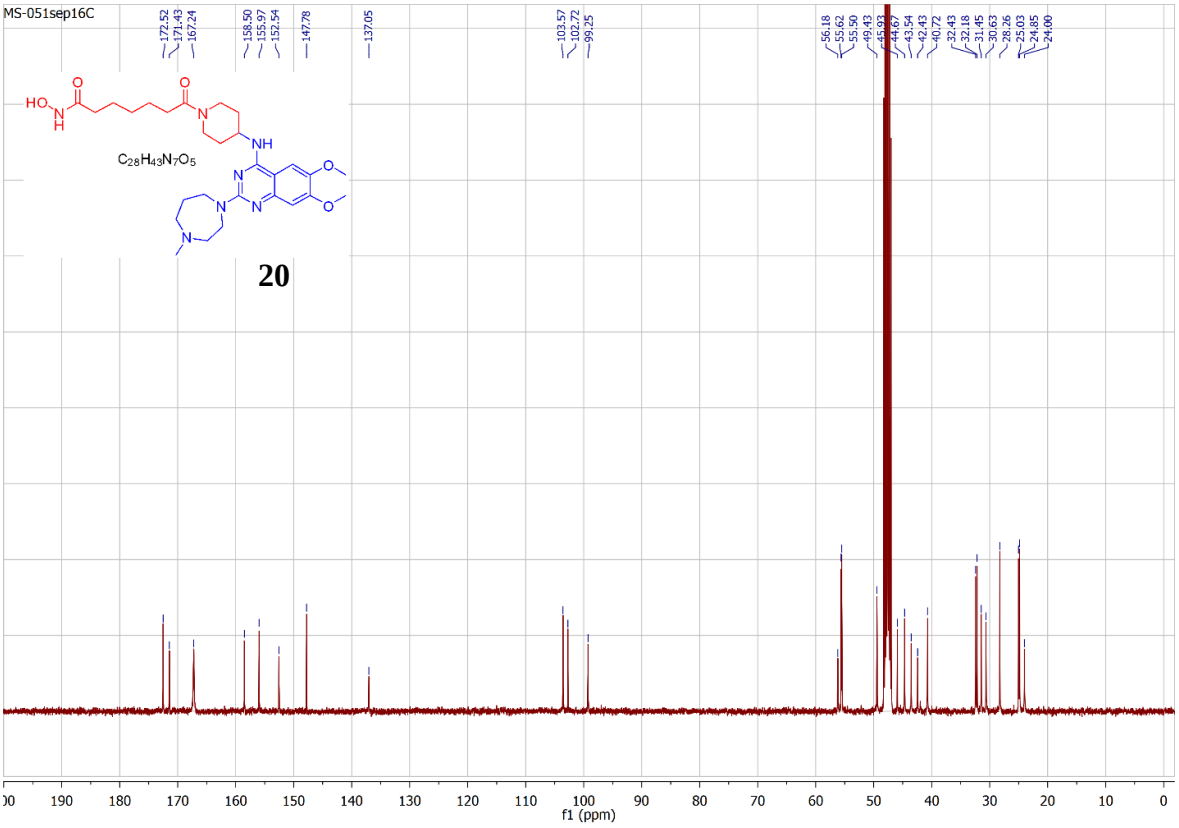




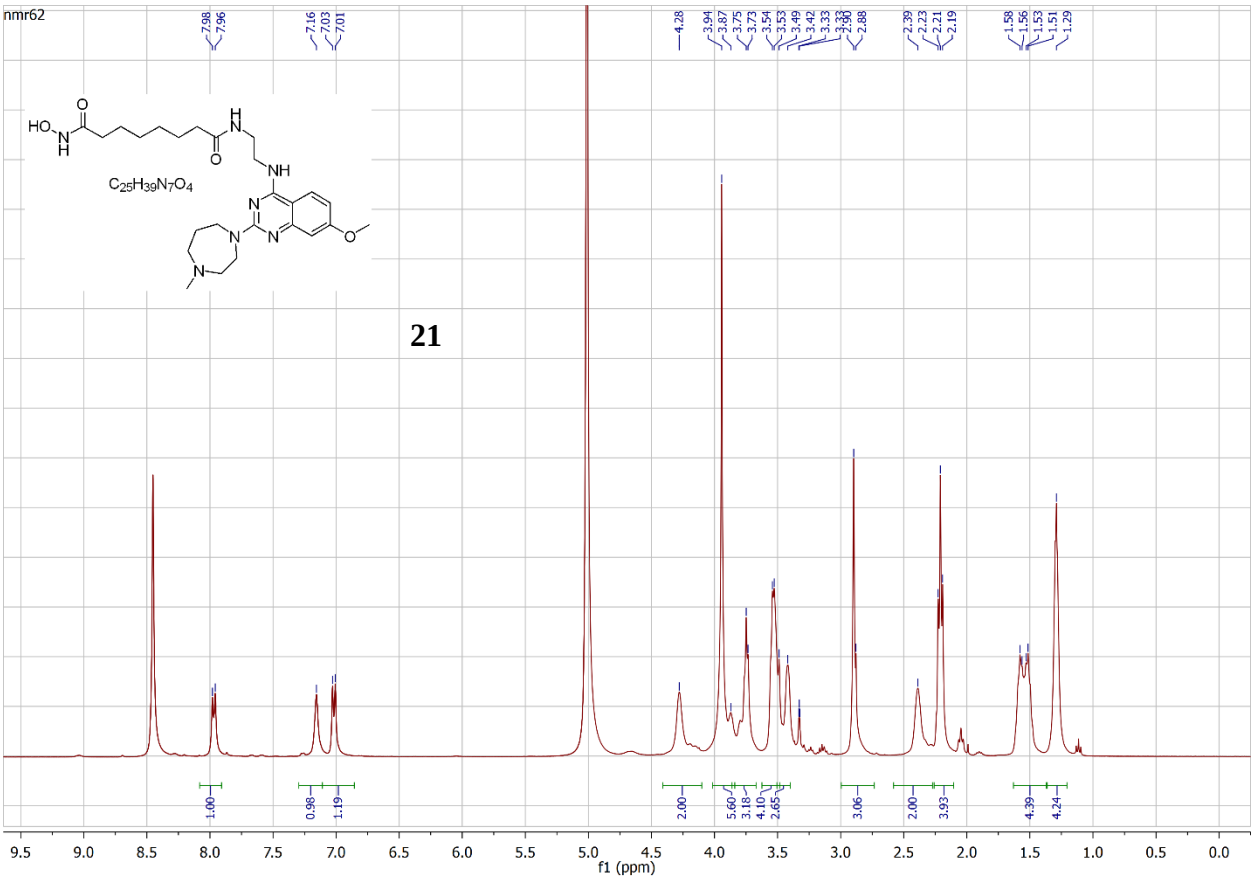


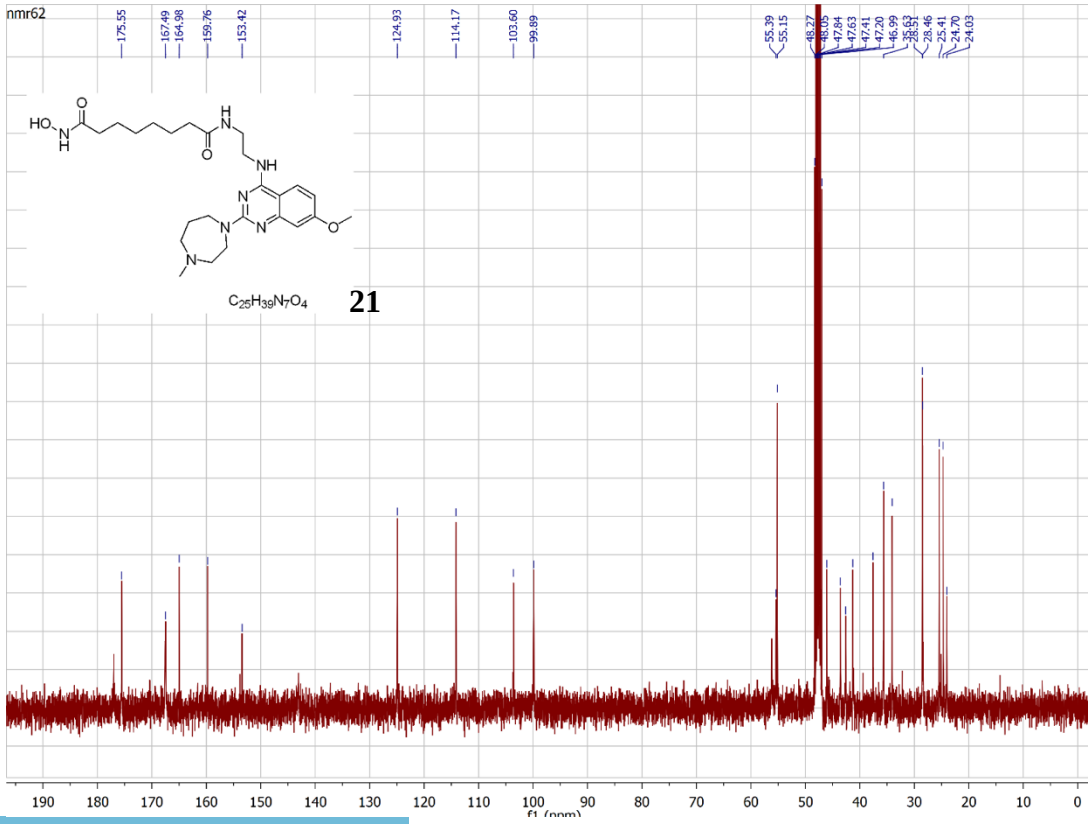


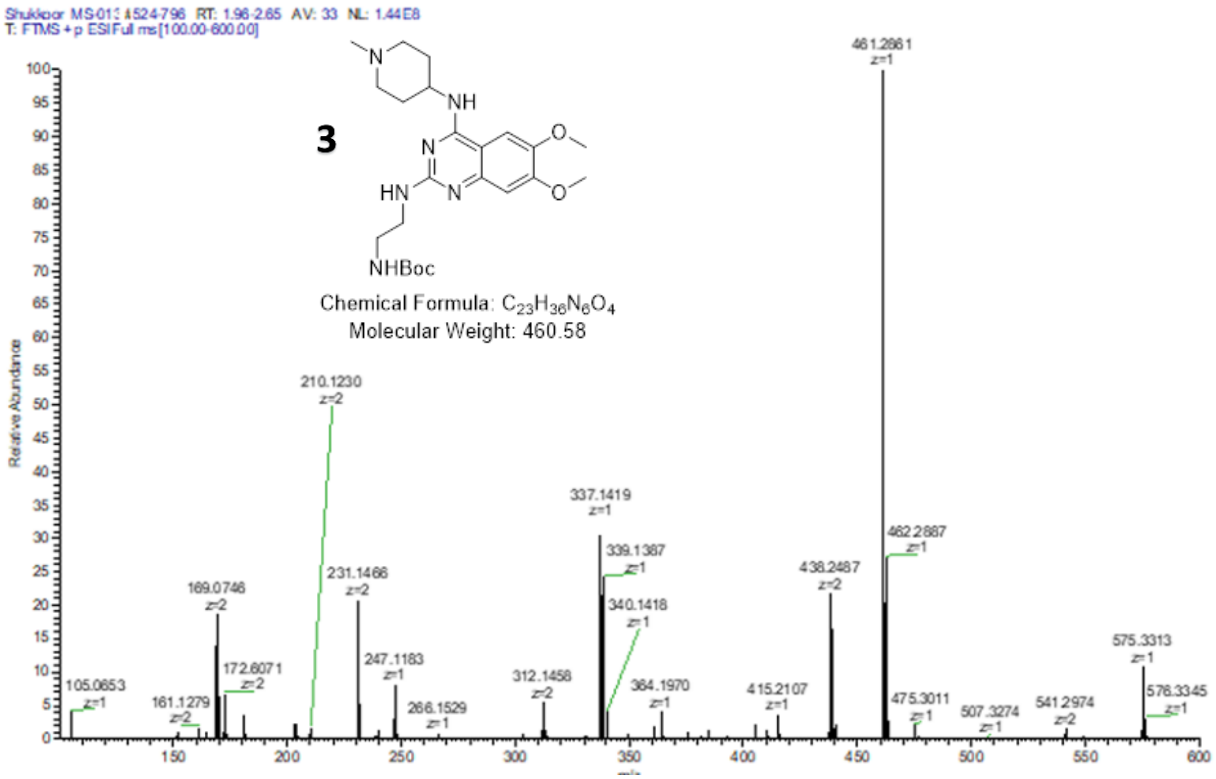
20

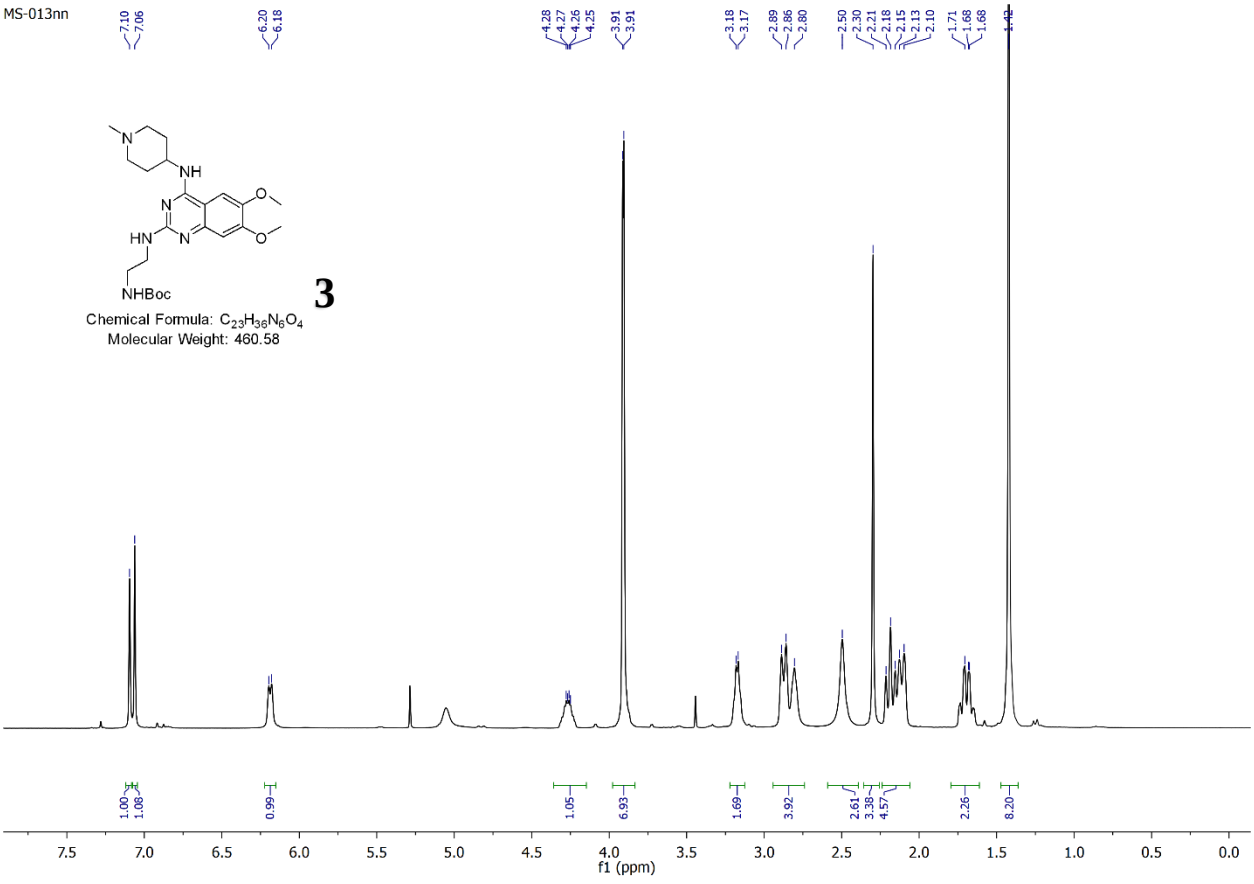


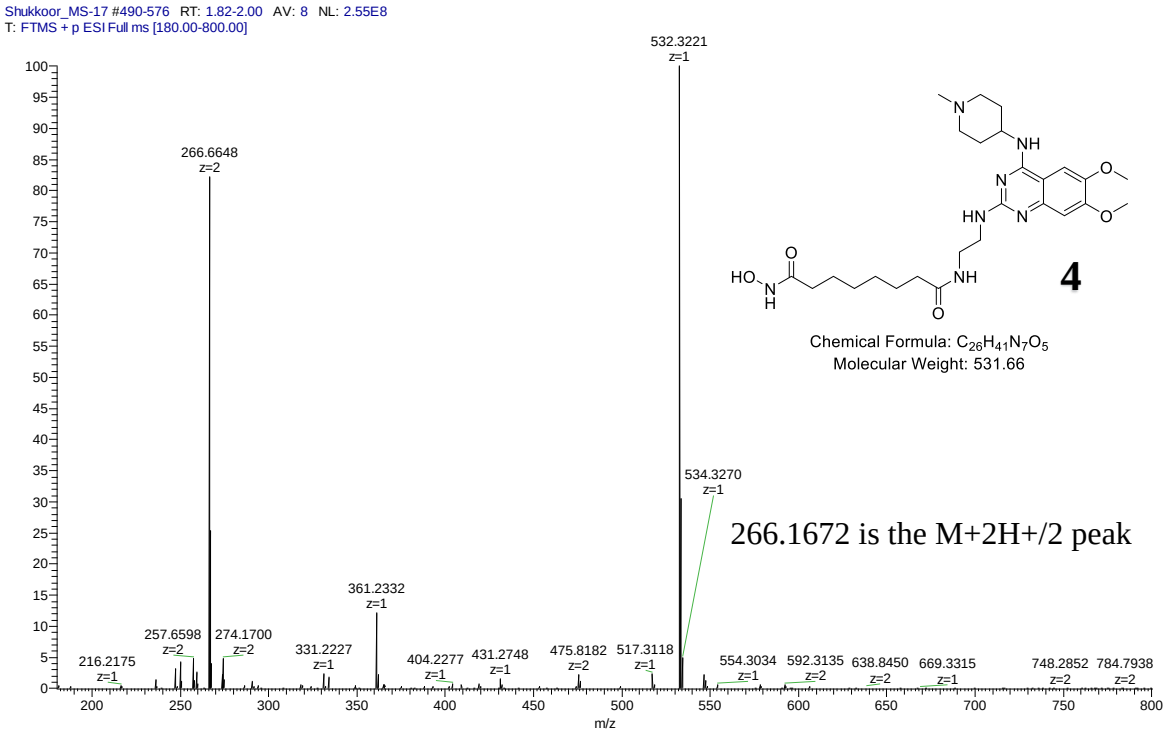
20

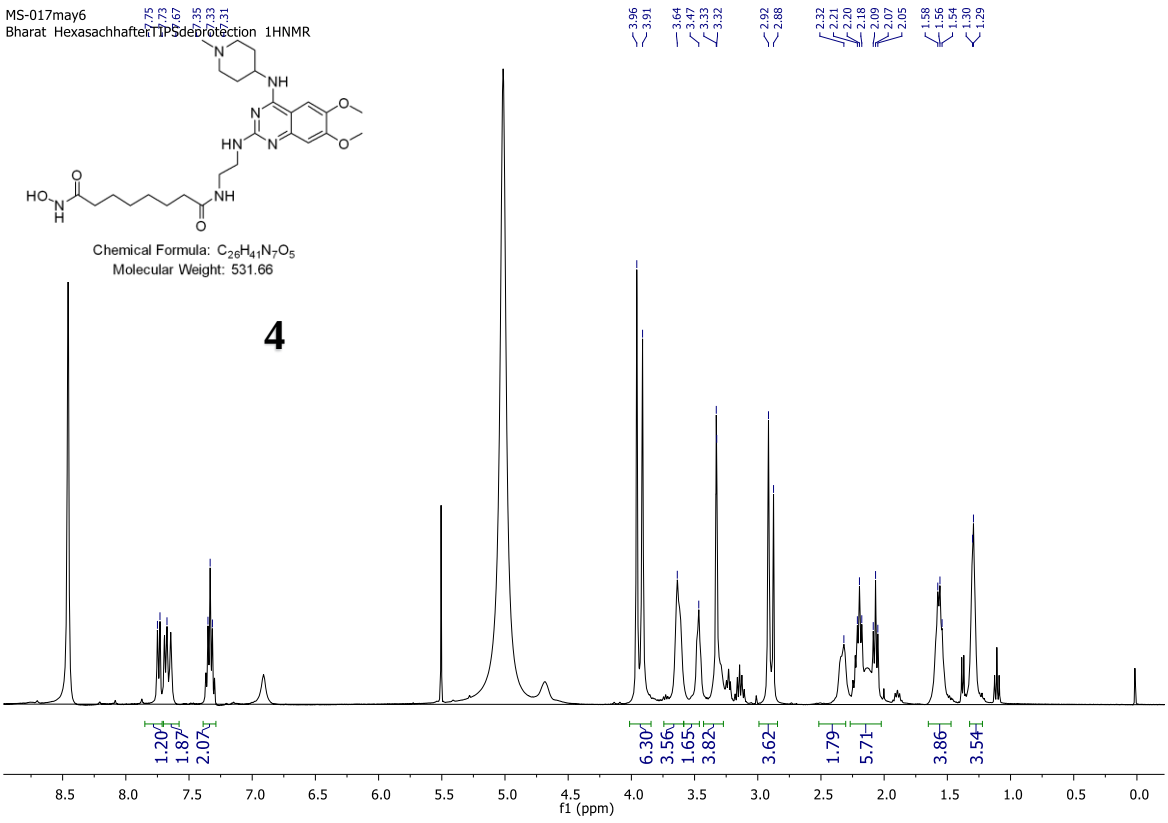


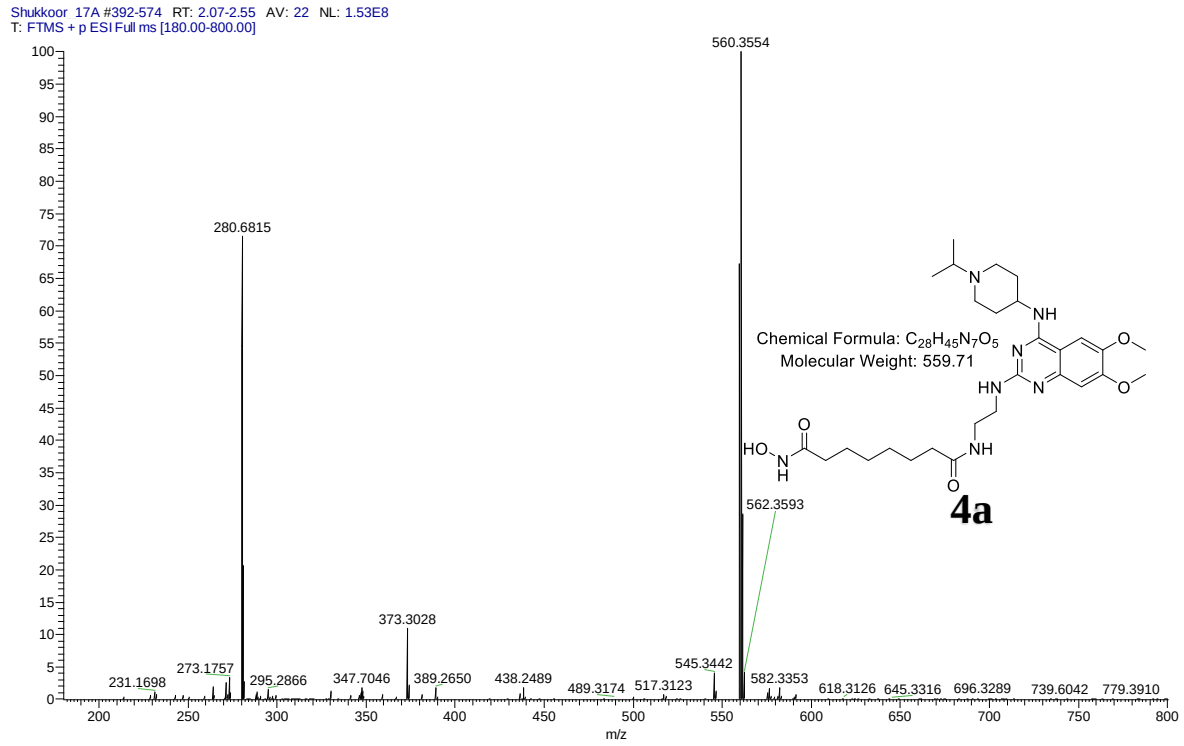


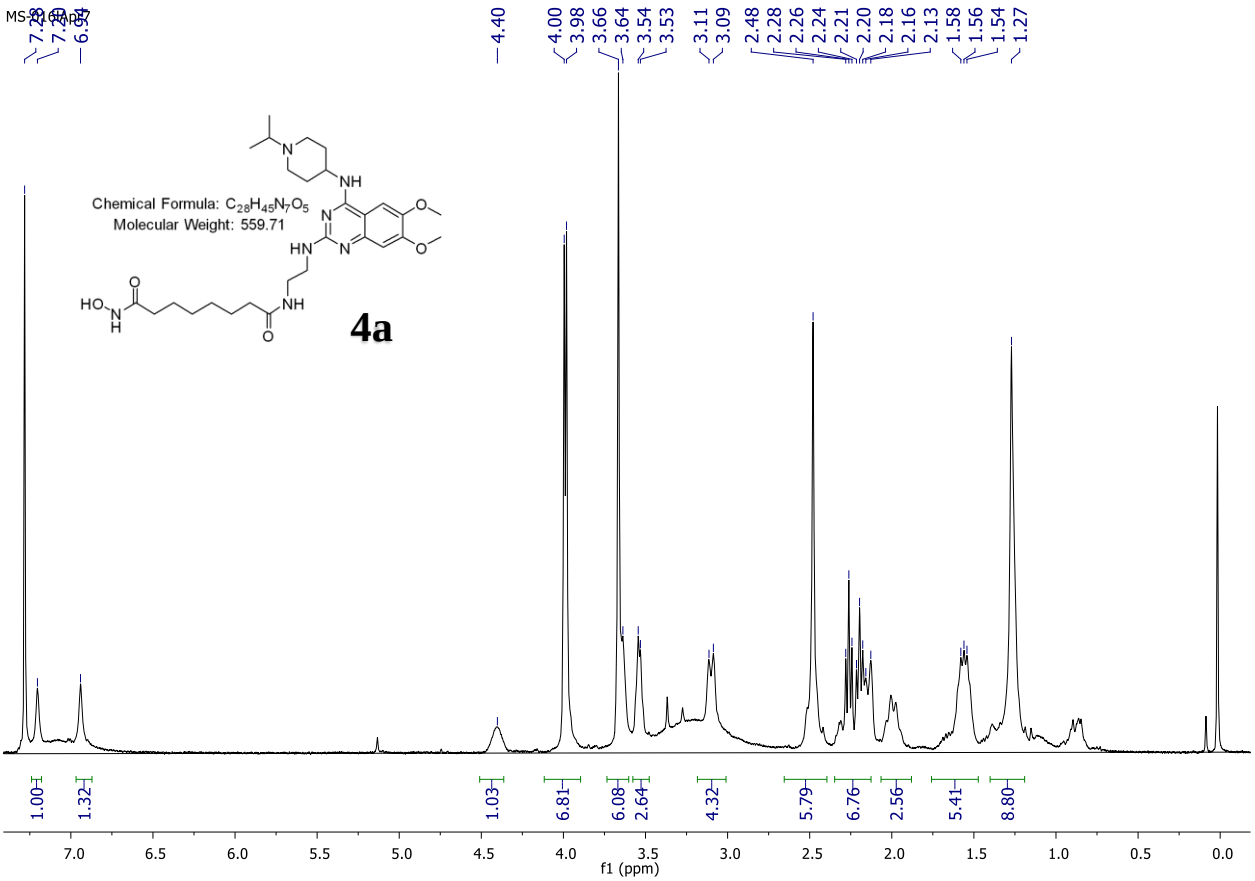


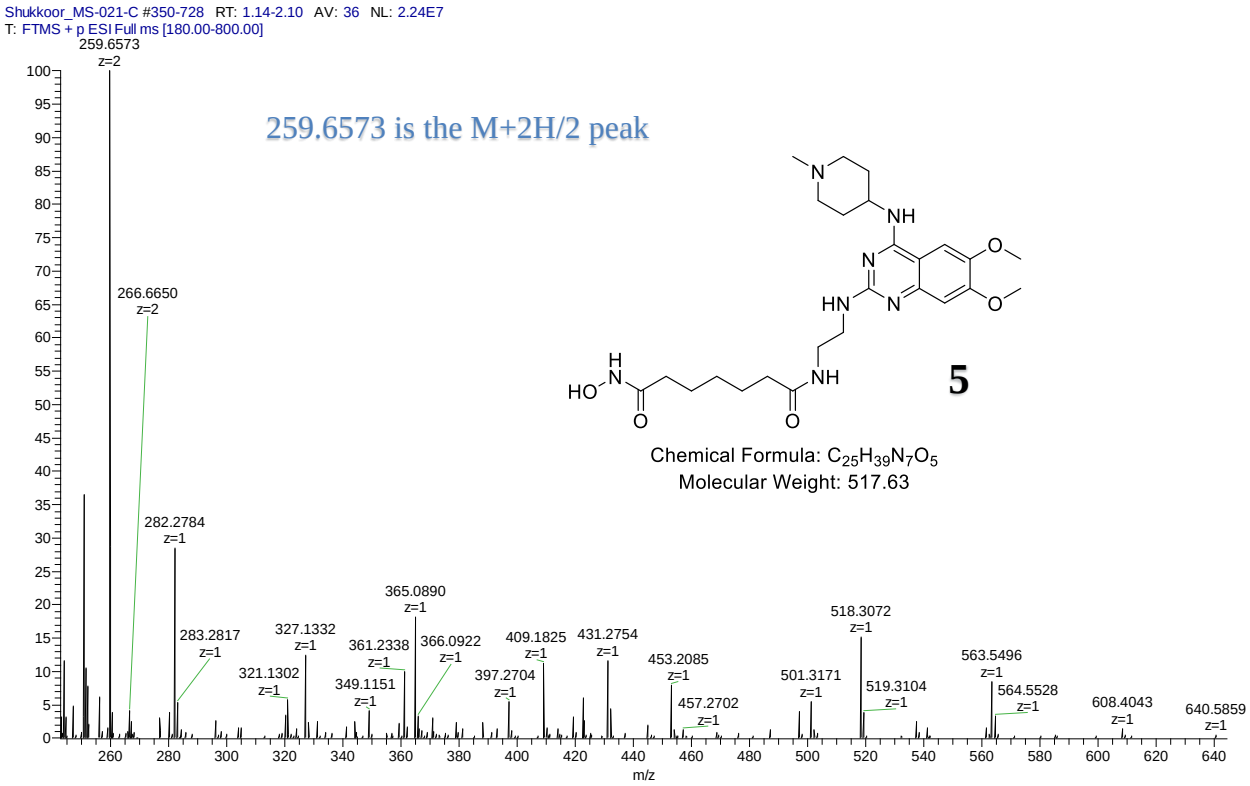


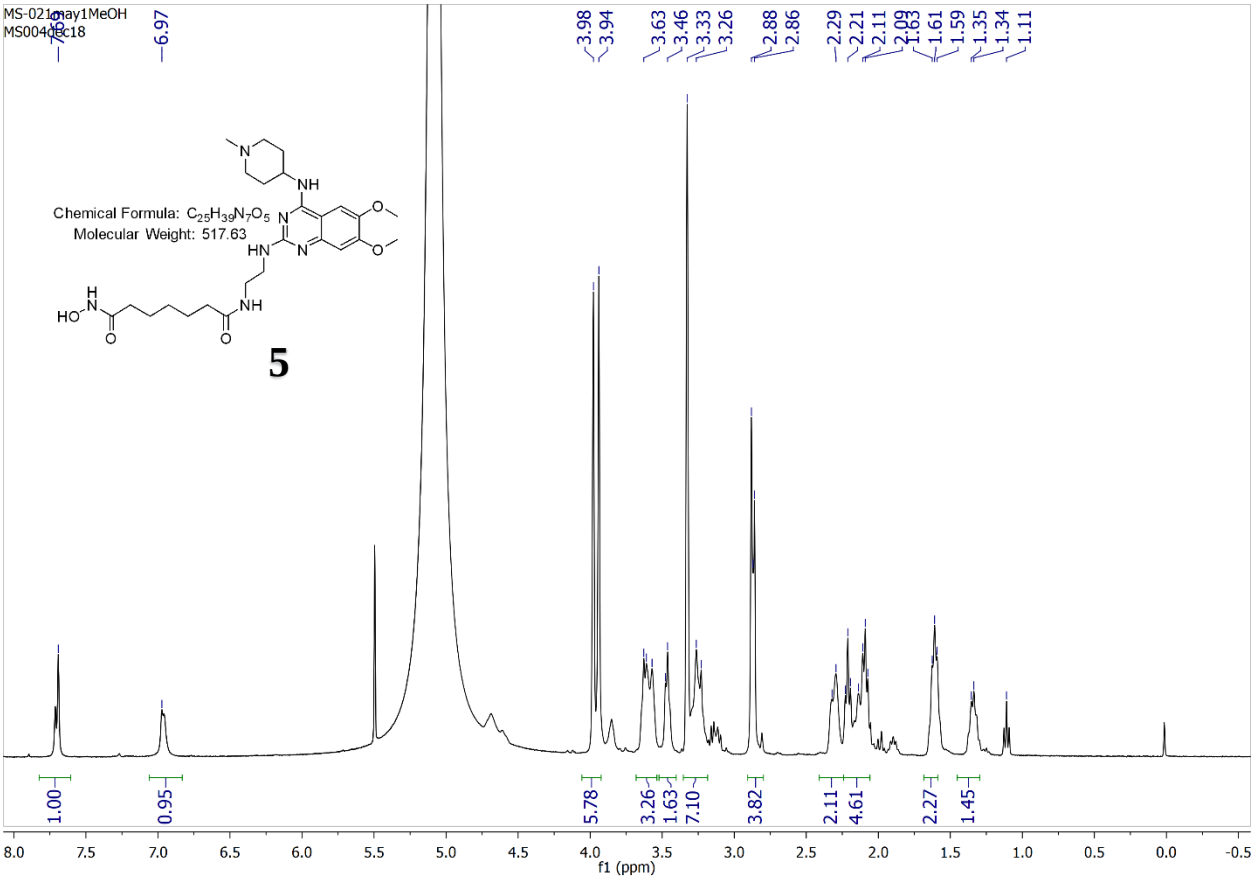


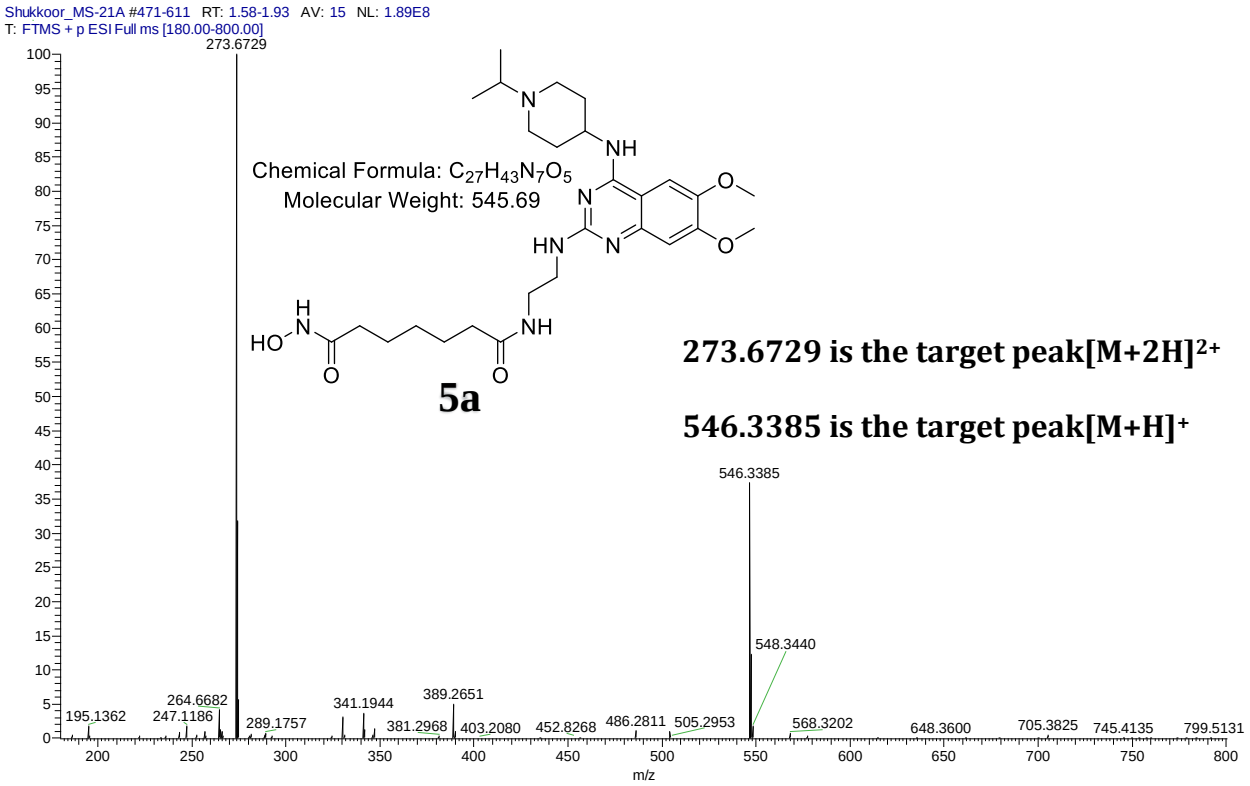


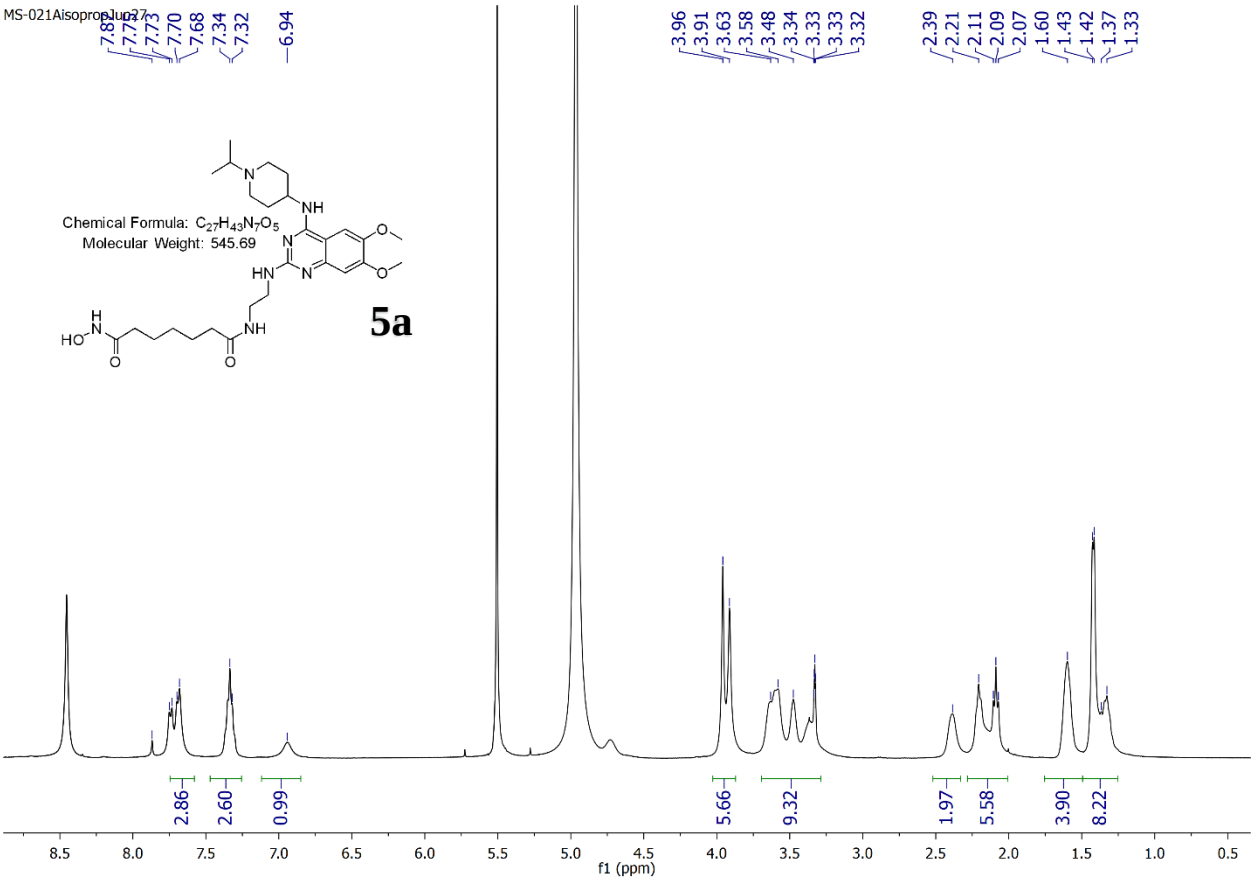


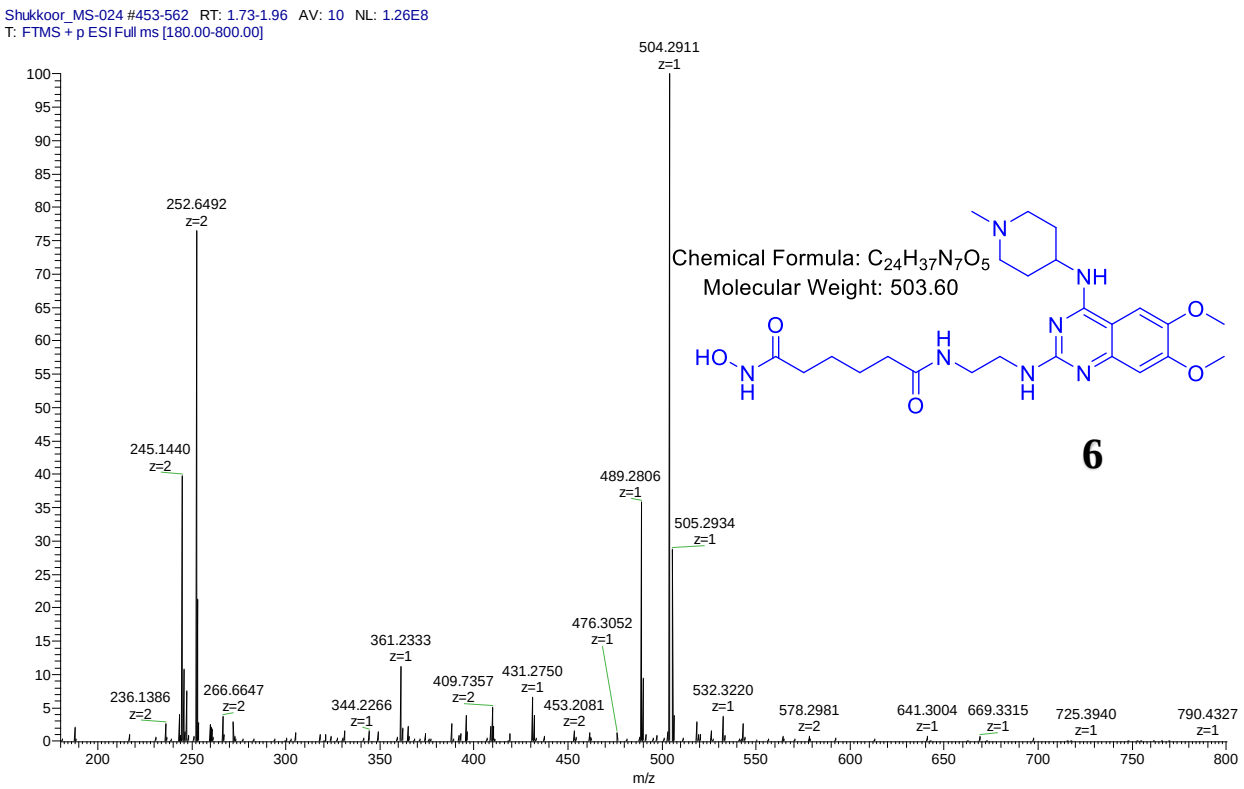


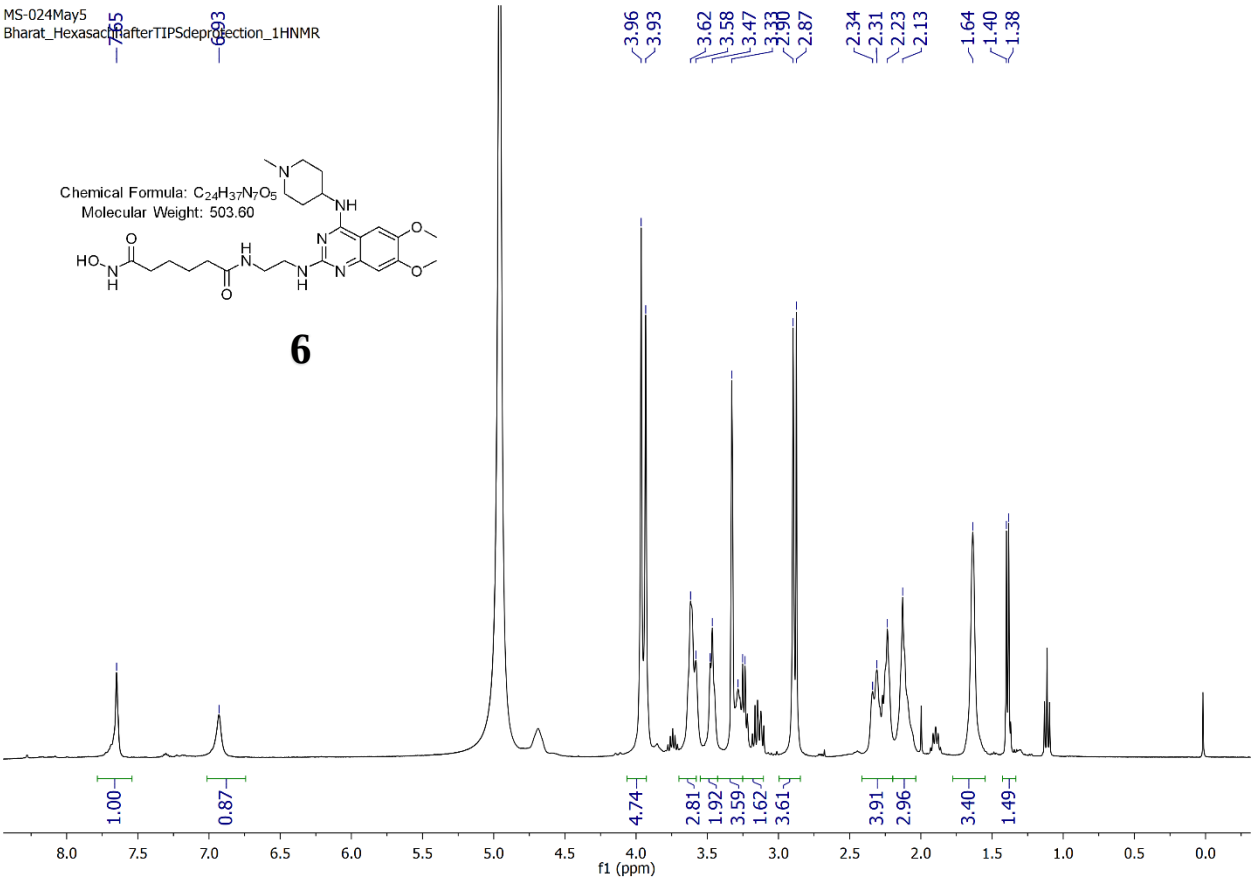


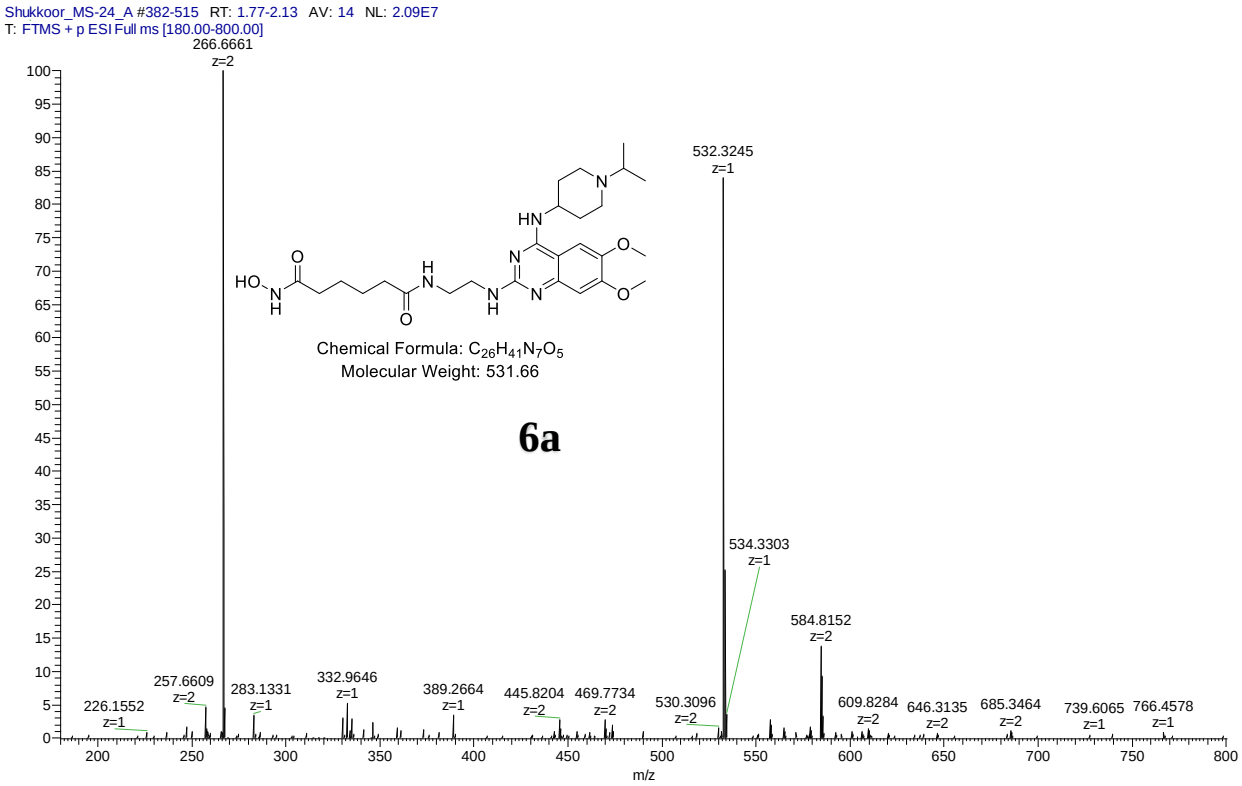






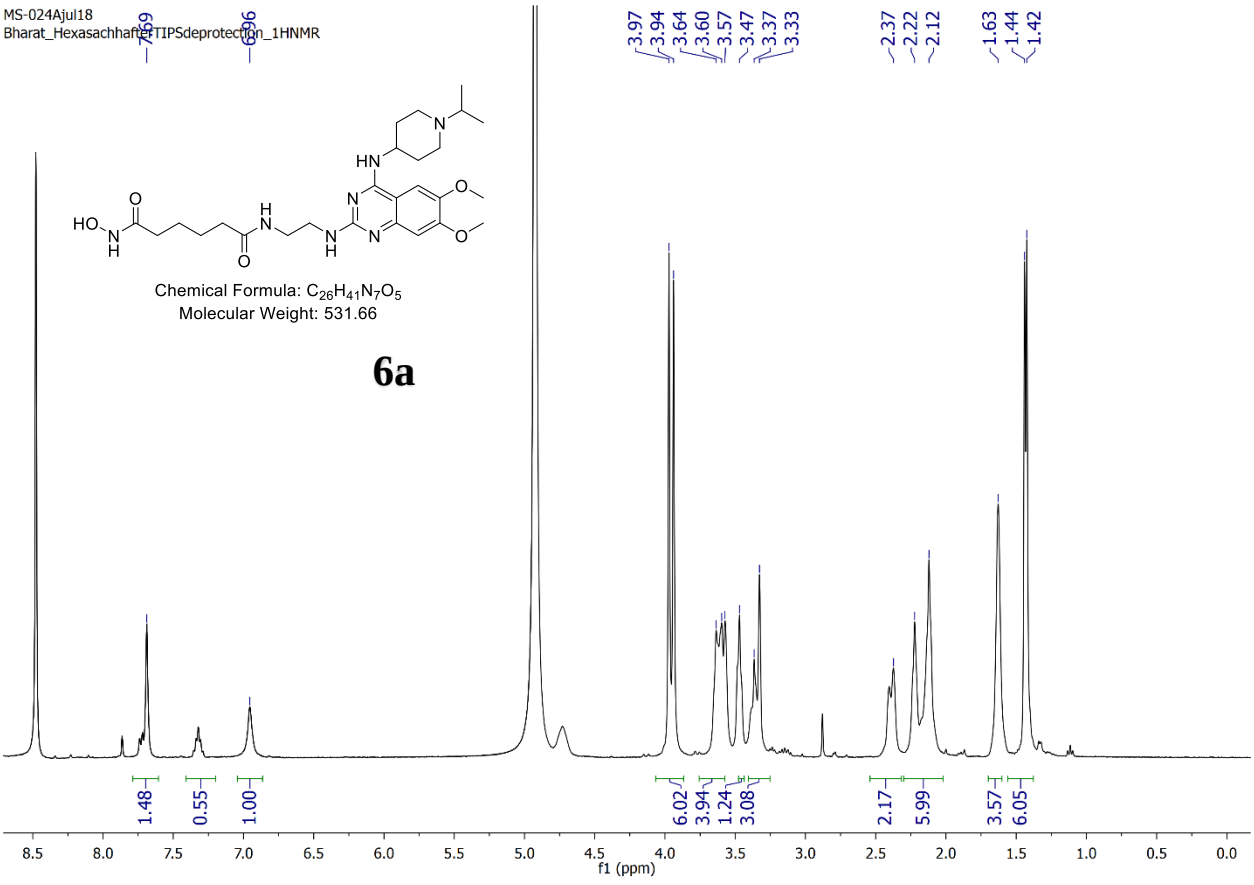






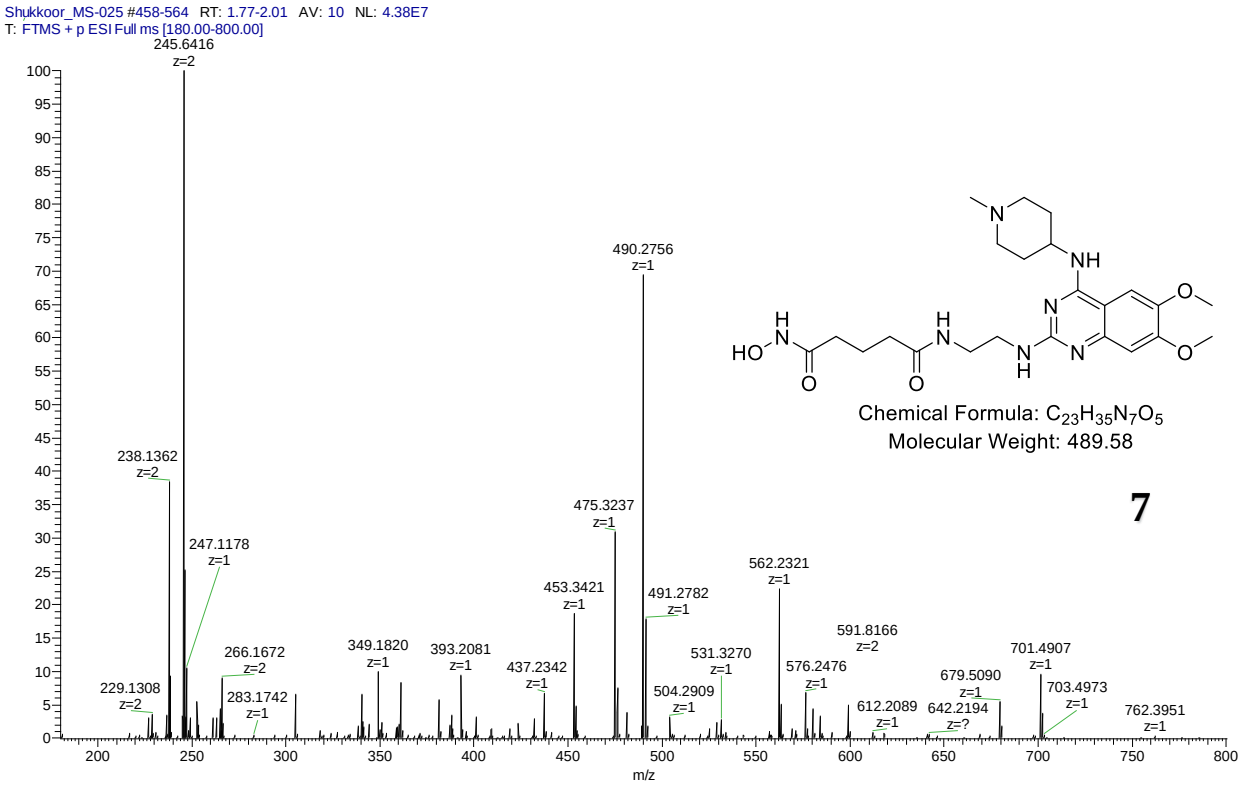
Chemical Formula: $C_{26}H_{41}N_7O_5$
Molecular Weight: 531.66

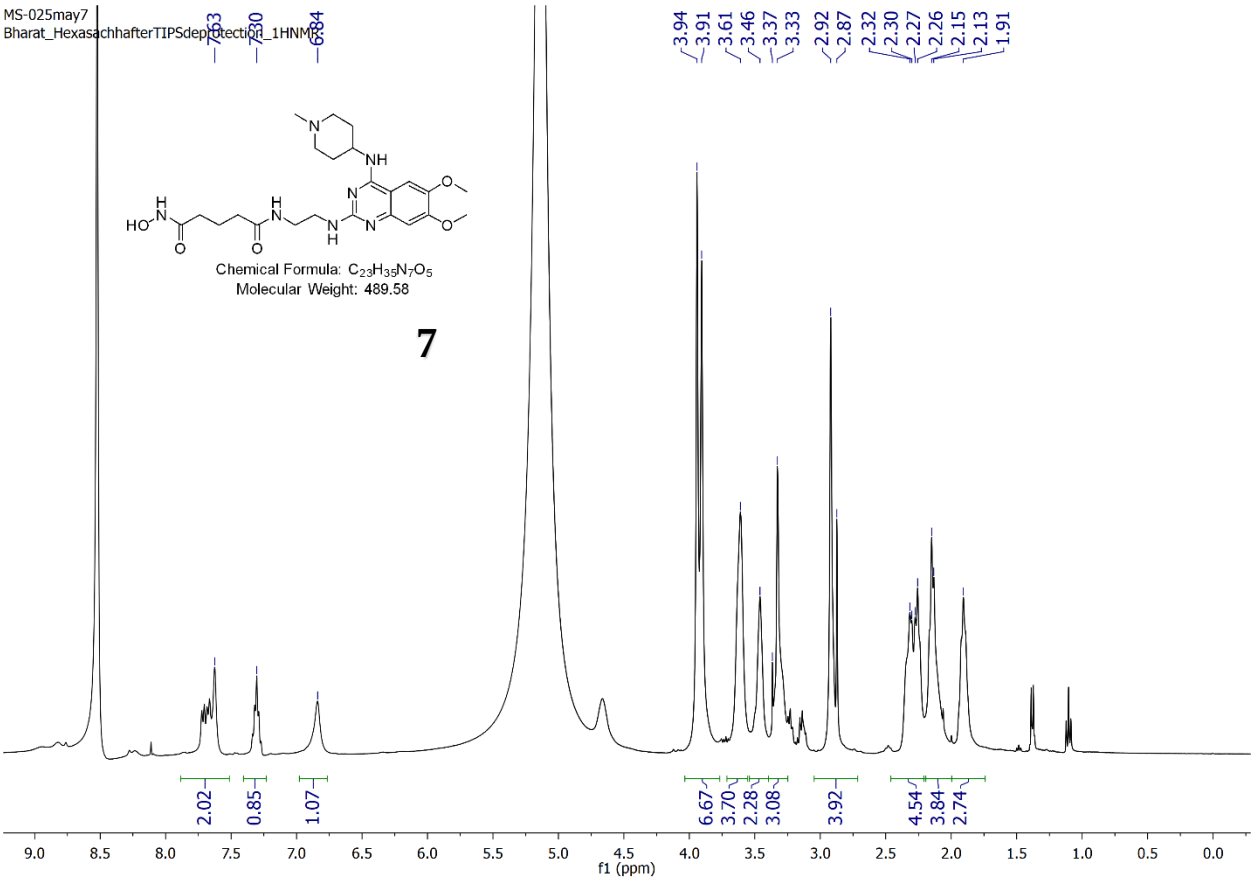
6a

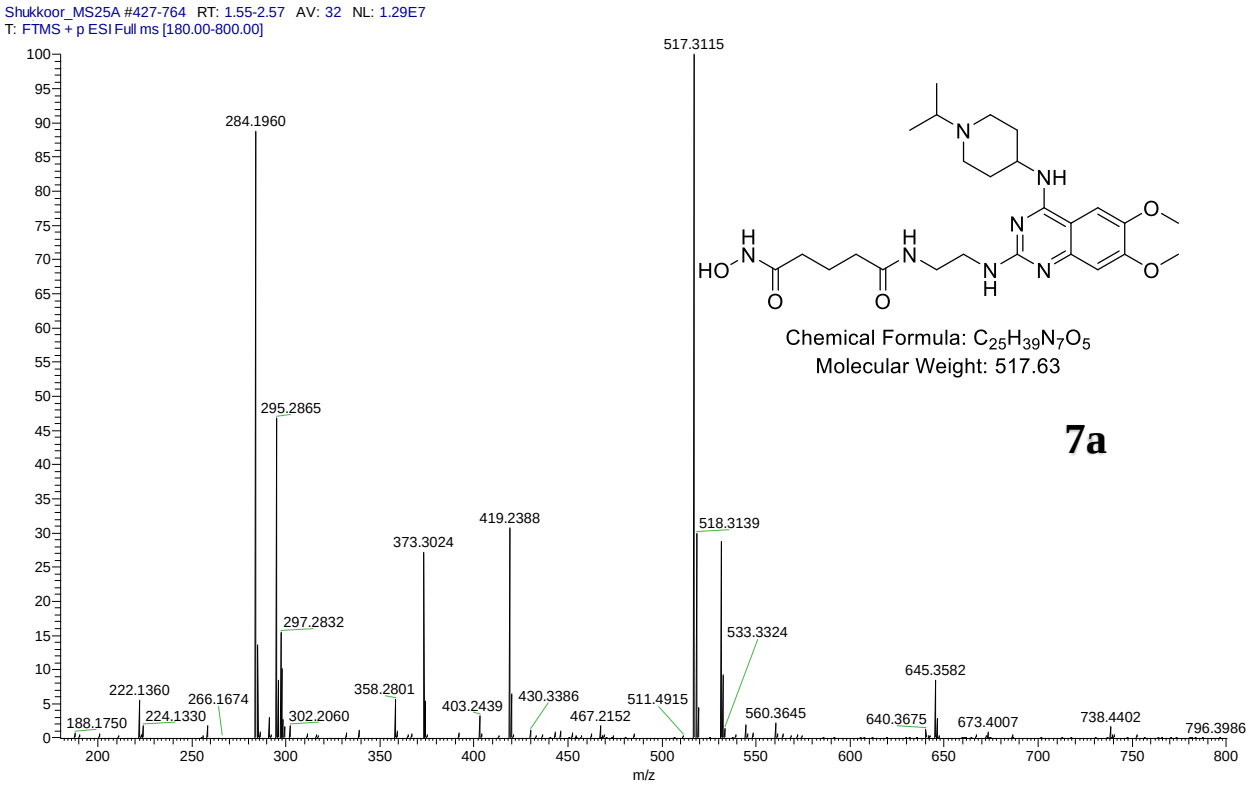


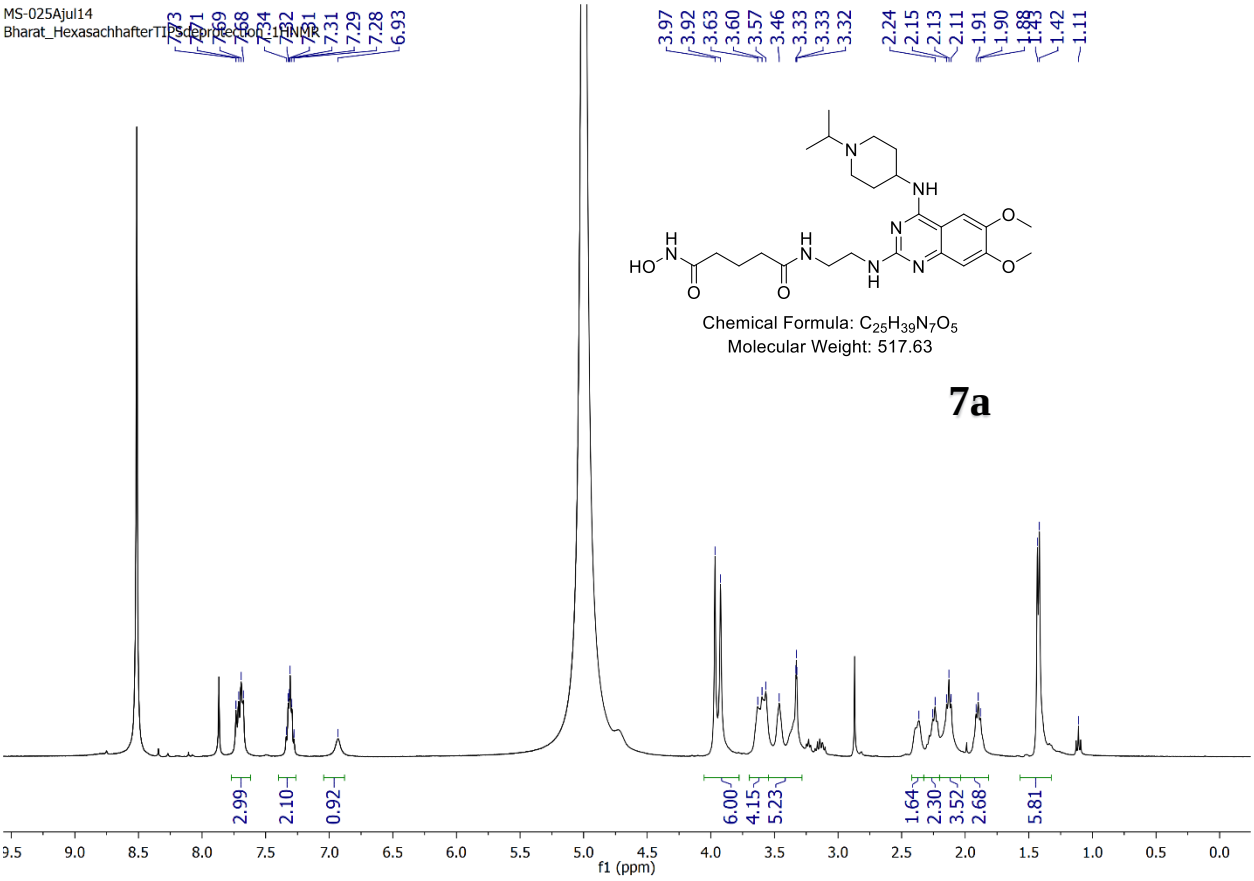
Chemical Formula: $C_{26}H_{41}N_7O_5$
Molecular Weight: 531.66

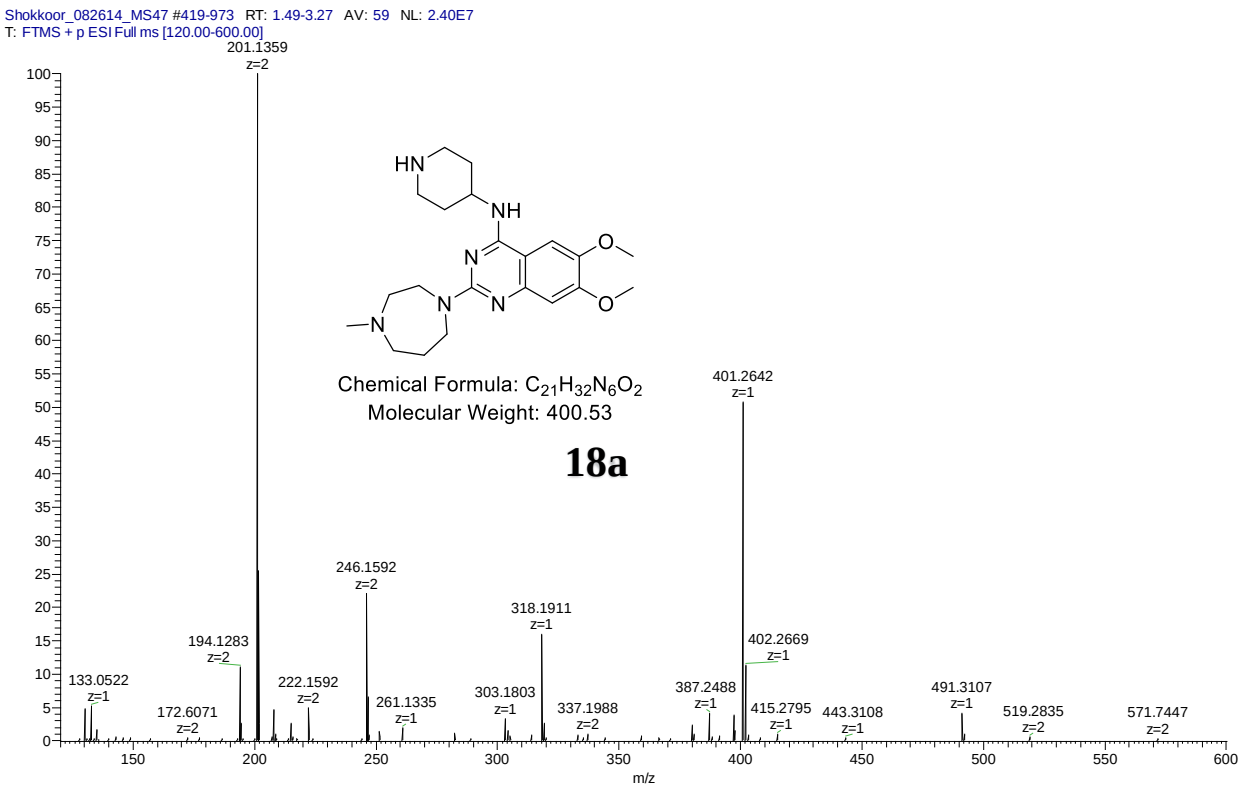
6a

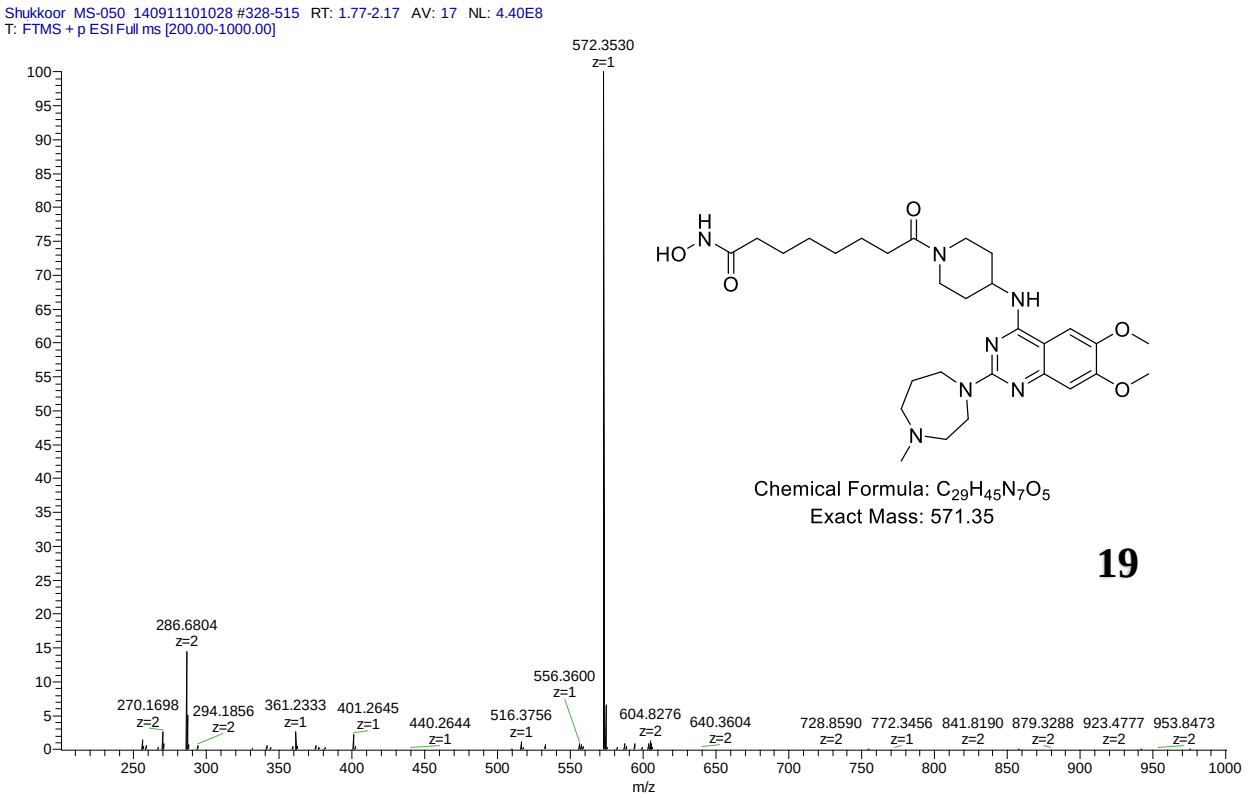






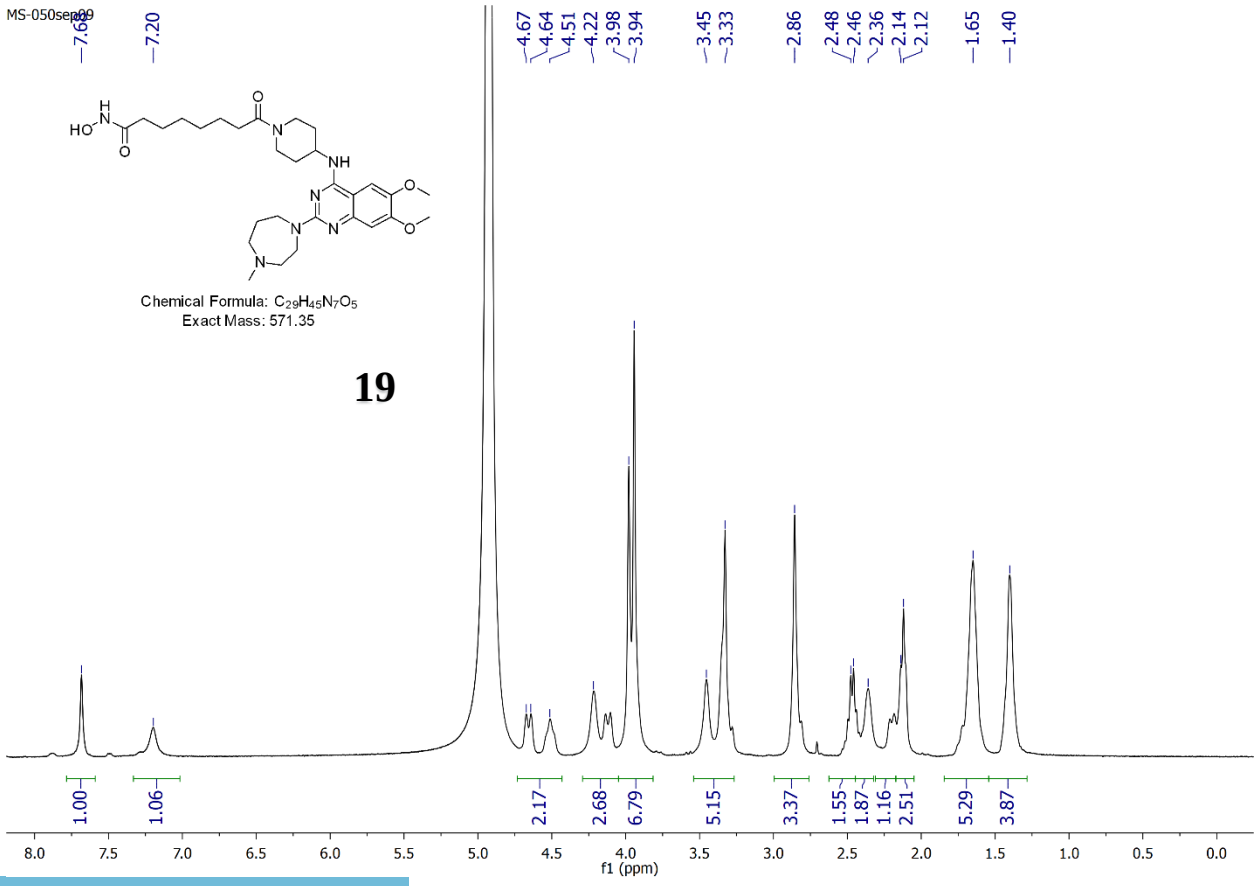






Chemical Formula: $C_{29}H_{45}N_7O_5$
Exact Mass: 571.35

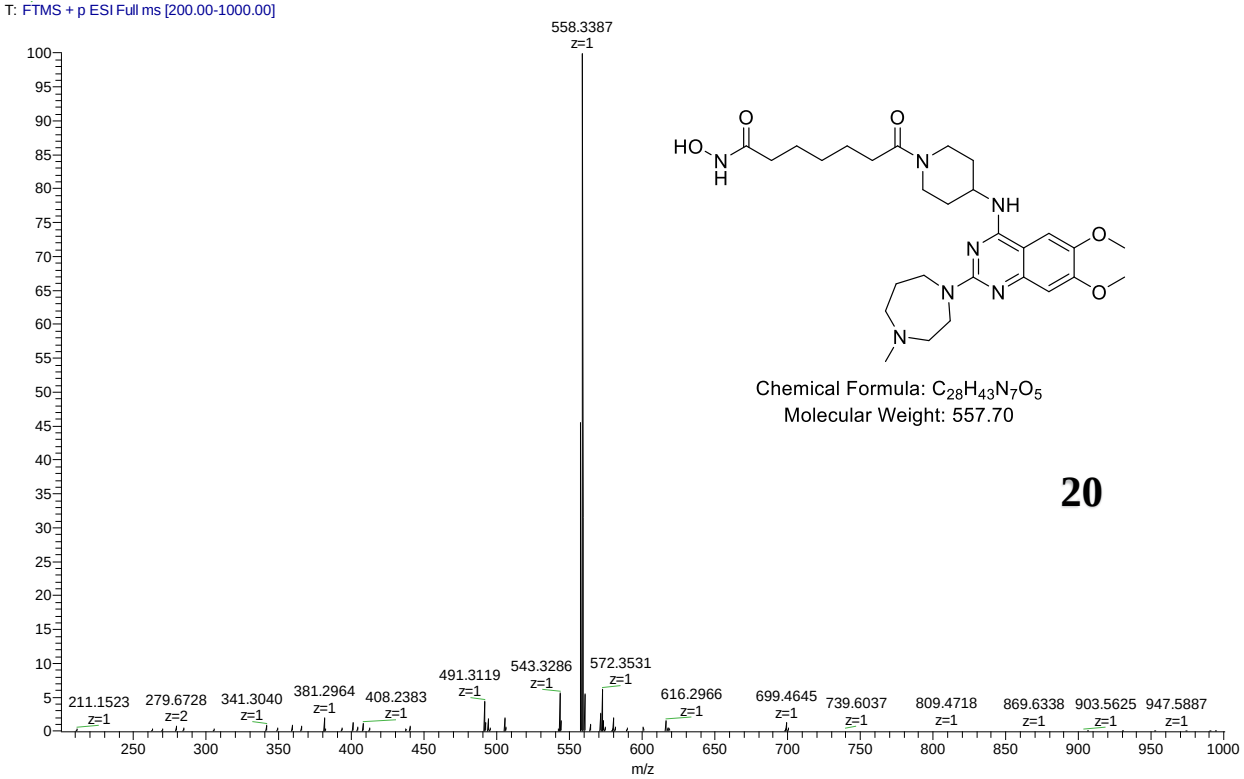
19



Chemical Formula: $C_{29}H_{45}N_7O_5$
Exact Mass: 571.35

19

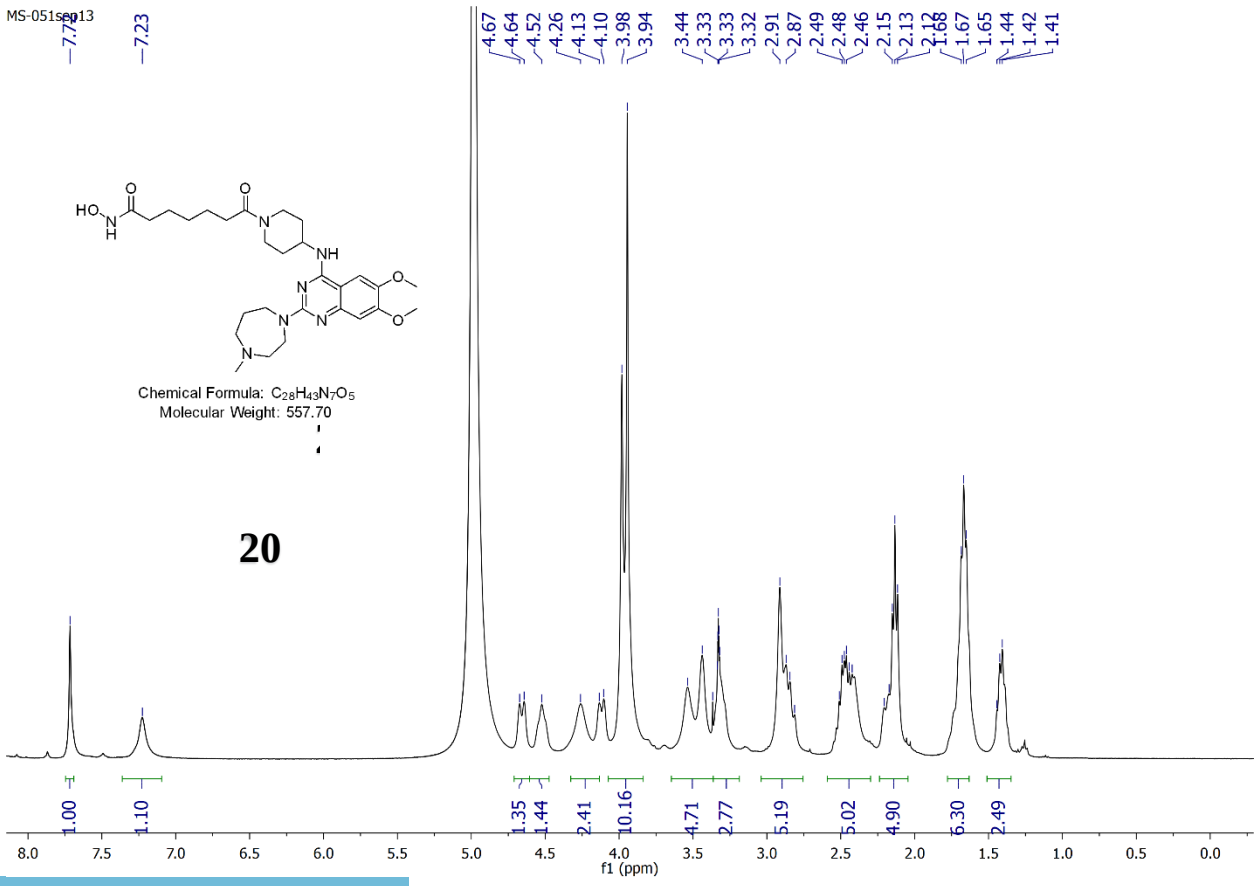
Shukkoor MS-051A #281-464 RT: 2.01-2.49 AV: 17 NL: 6.11E7
T: FTMS + p ESI Full ms [200.00-1000.00]



Chemical Formula: $C_{28}H_{43}N_7O_5$
Molecular Weight: 557.70

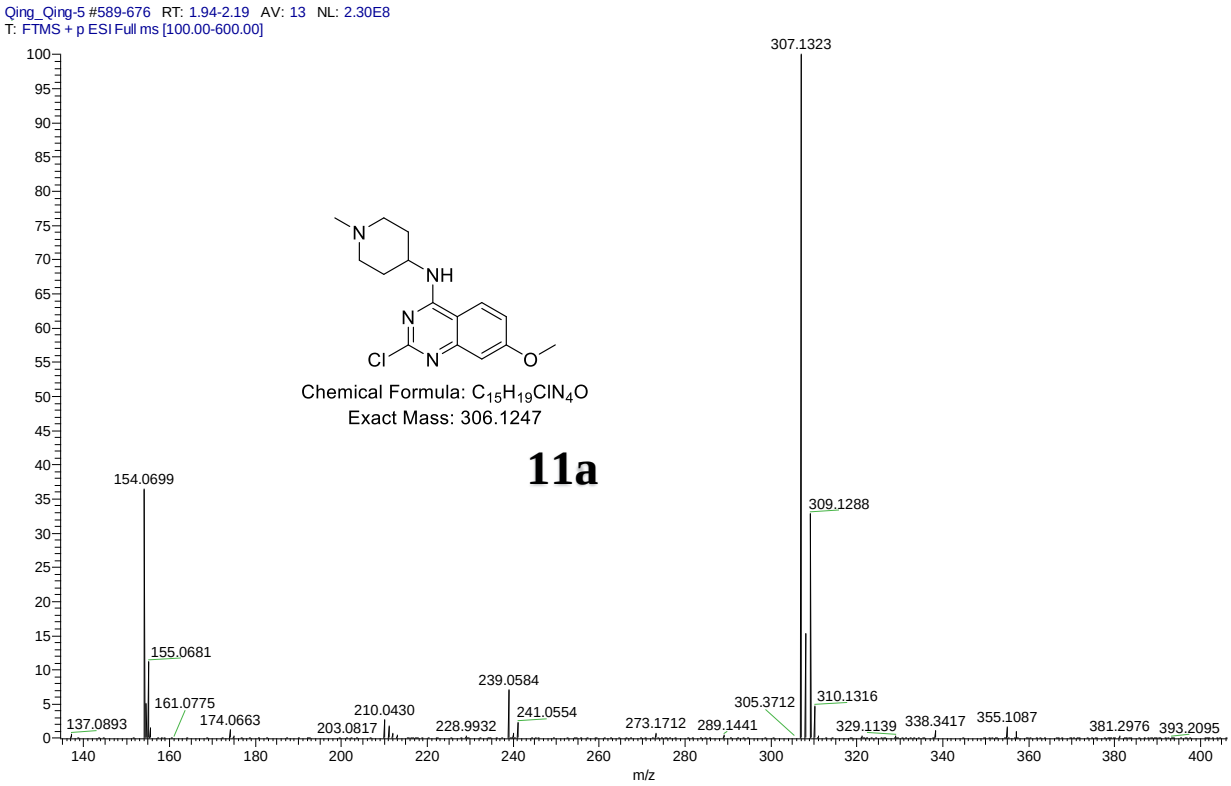
20

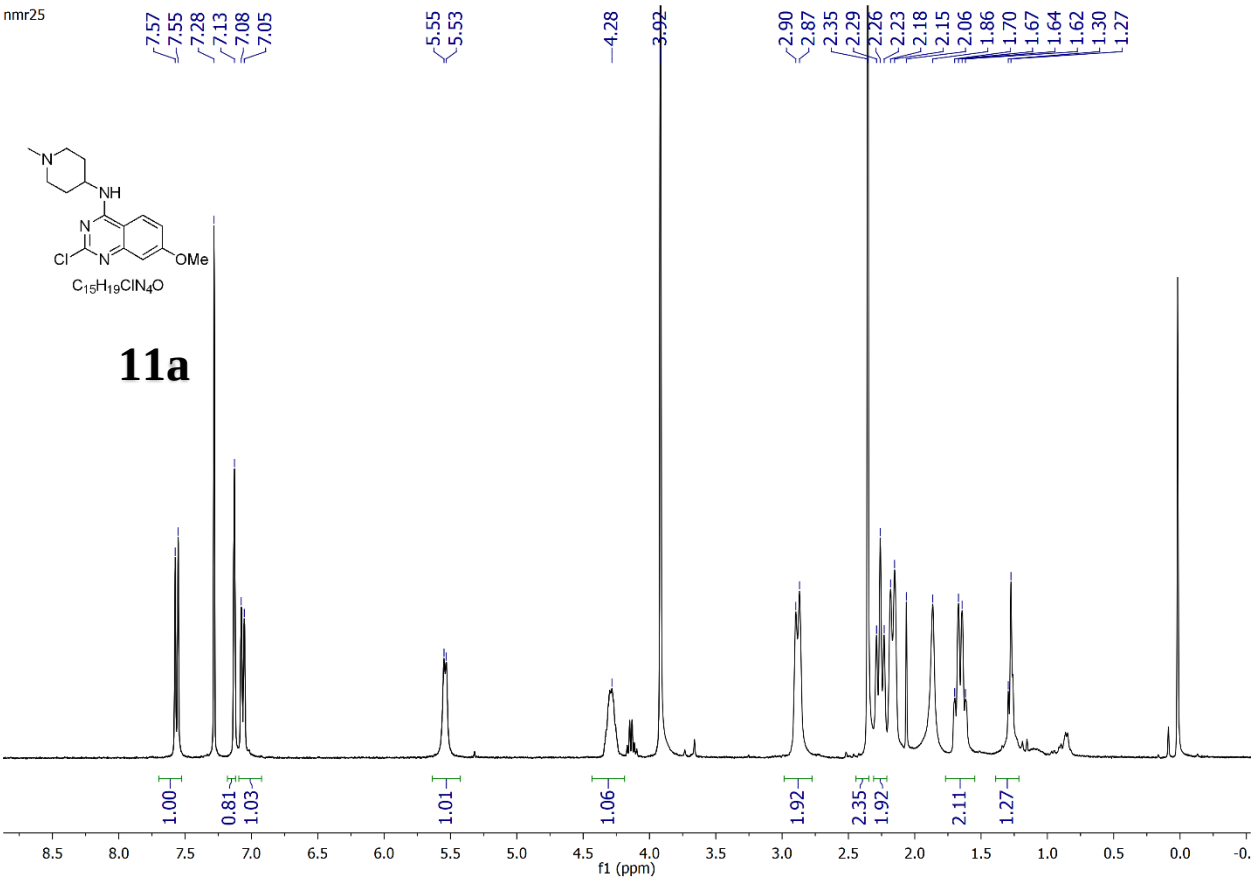
MS-051sep13

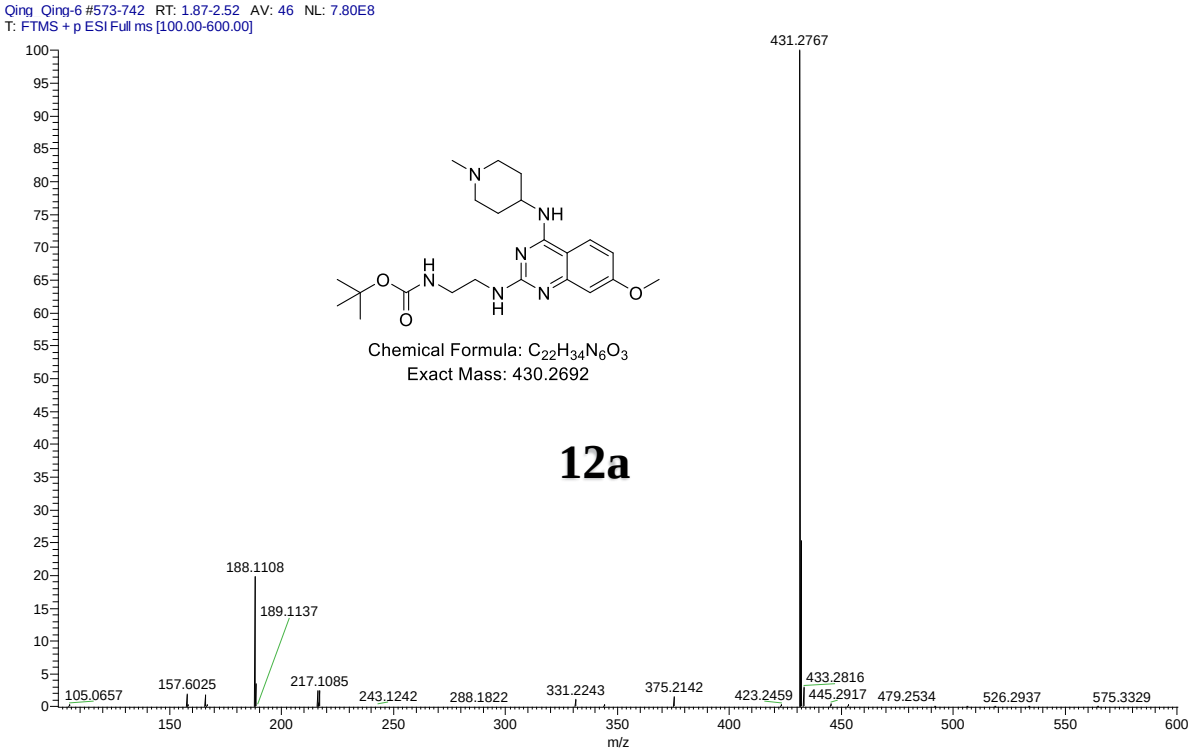


Chemical Formula: $C_{28}H_{43}N_7O_5$
Molecular Weight: 557.70

20

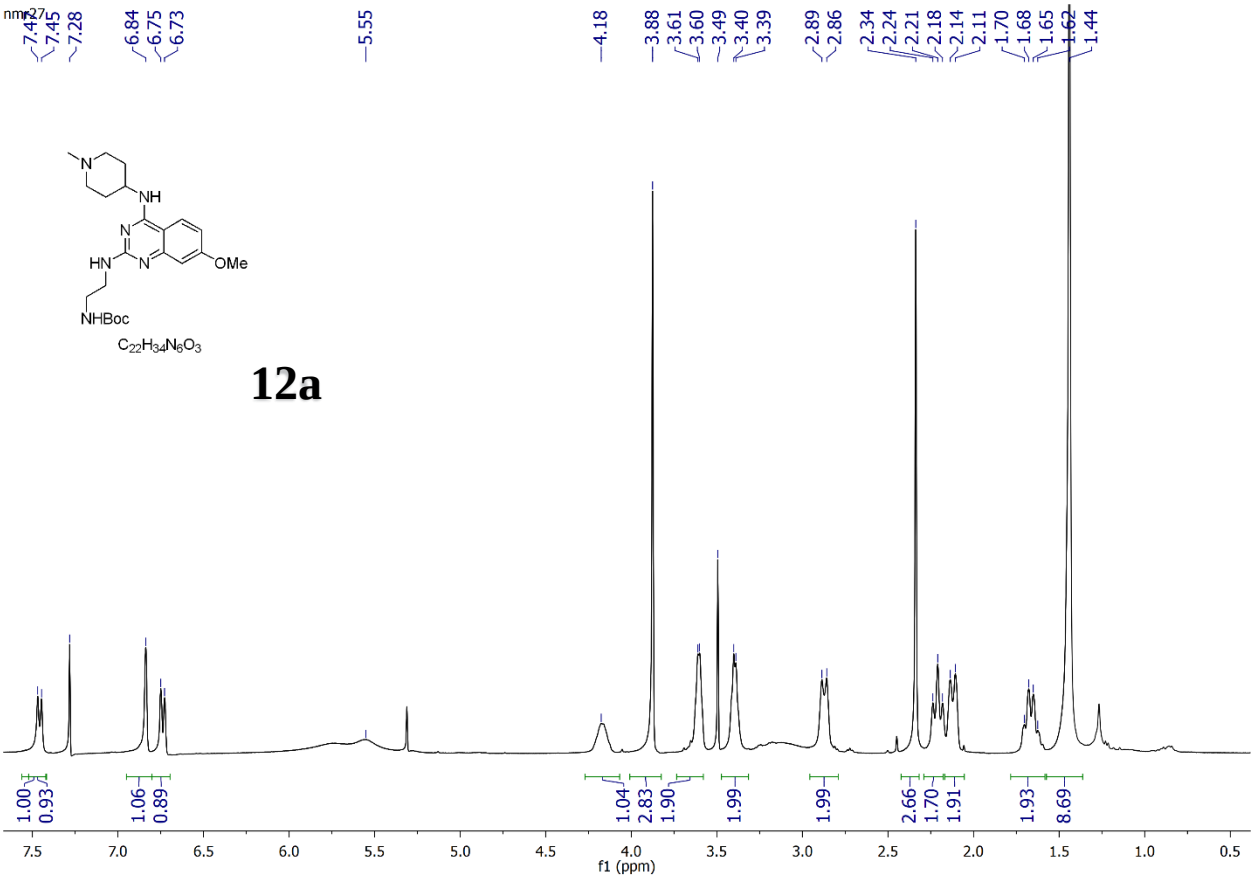






Chemical Formula: $C_{22}H_{34}N_6O_3$
Exact Mass: 430.2692

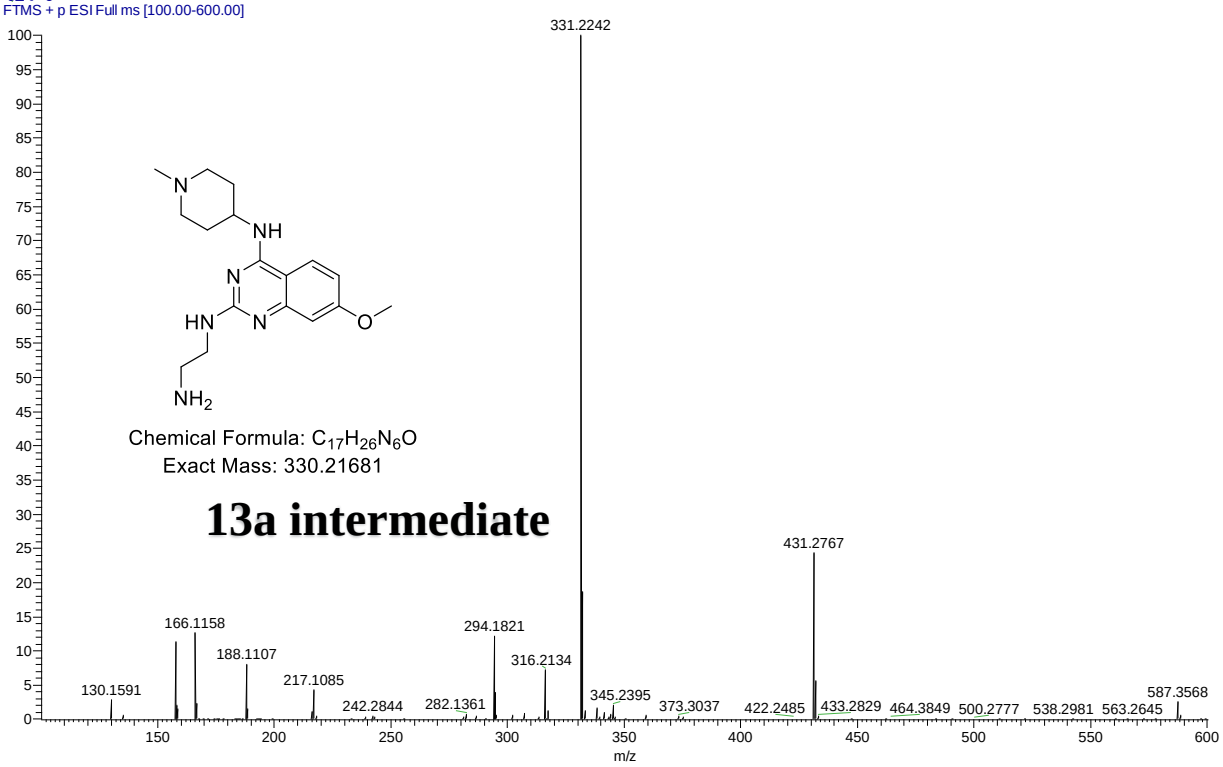
12a



$C_{22}H_{34}N_6O_3$

12a

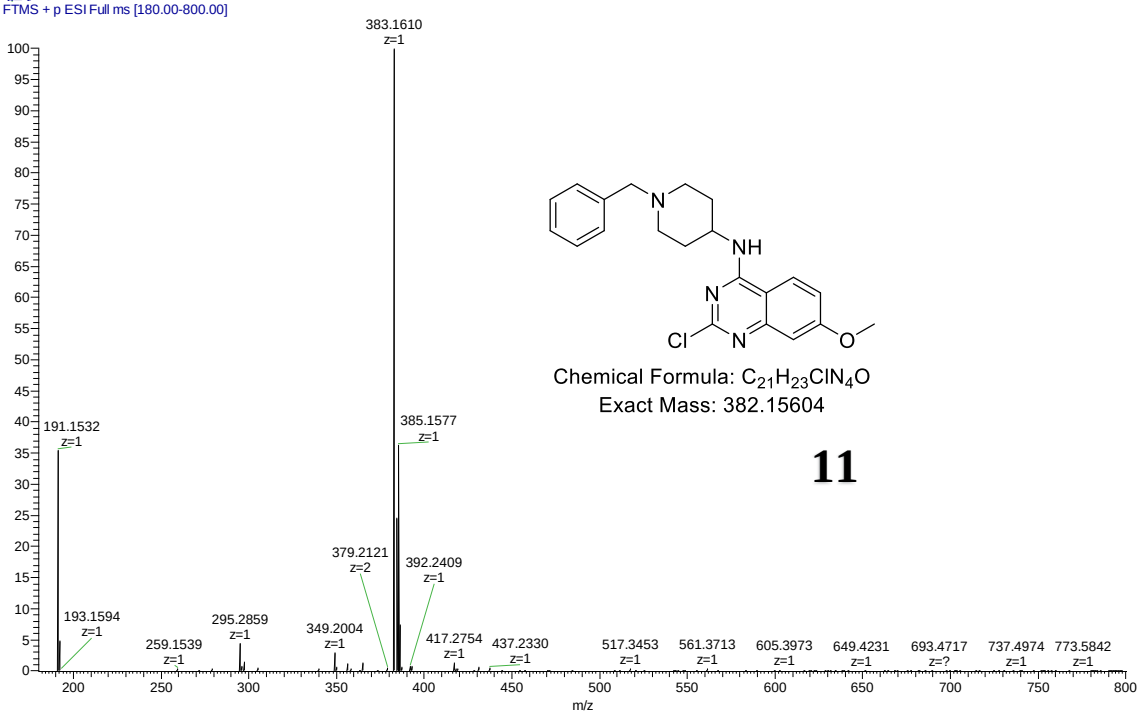
Qing_Qing-7 #672-863 RT: 2.17-2.72 AV: 29 NL: 2.01E8
T: FTMS + p ESI Full ms [100.00-600.00]



Chemical Formula: $C_{17}H_{26}N_6O$
Exact Mass: 330.21681

13a intermediate

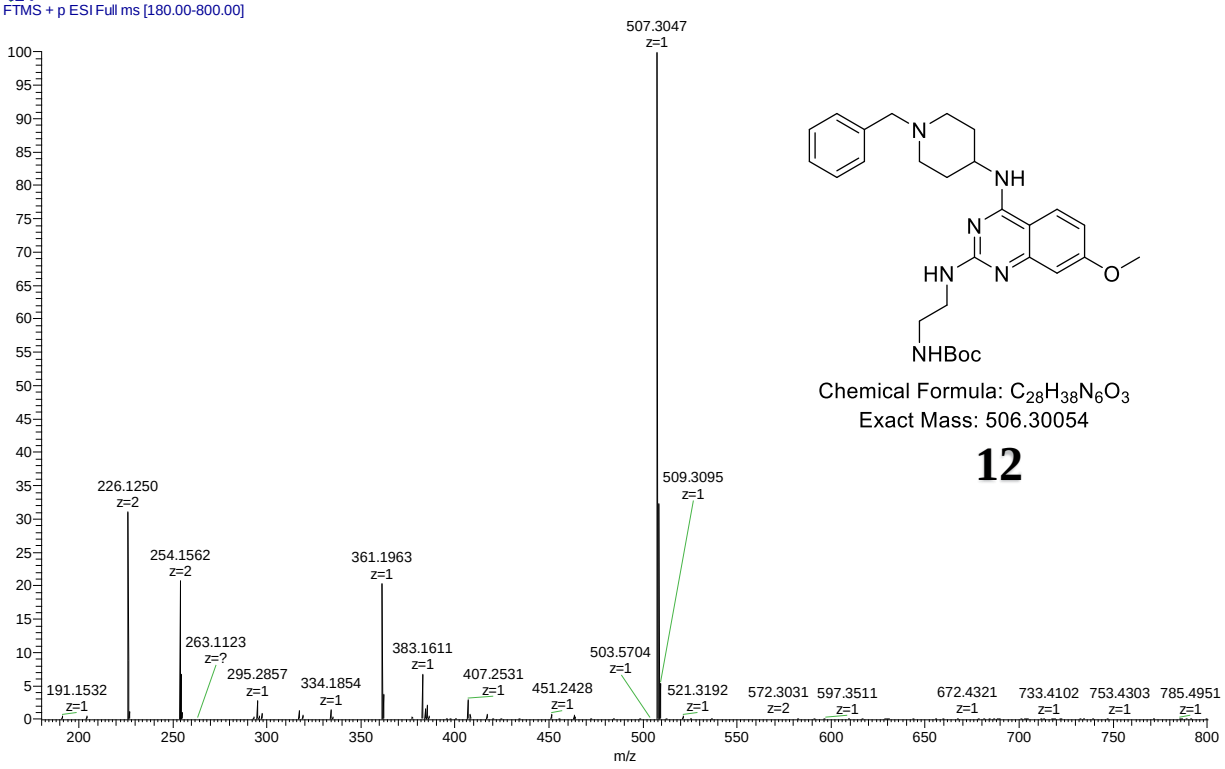
Qing_Q44 #487-635 RT: 1.97-2.34 AV: 19 NL: 4.50E8
T: FTMS + p ESI Full ms [180.00-800.00]



Chemical Formula: $C_{21}H_{23}ClN_4O$
Exact Mass: 382.15604

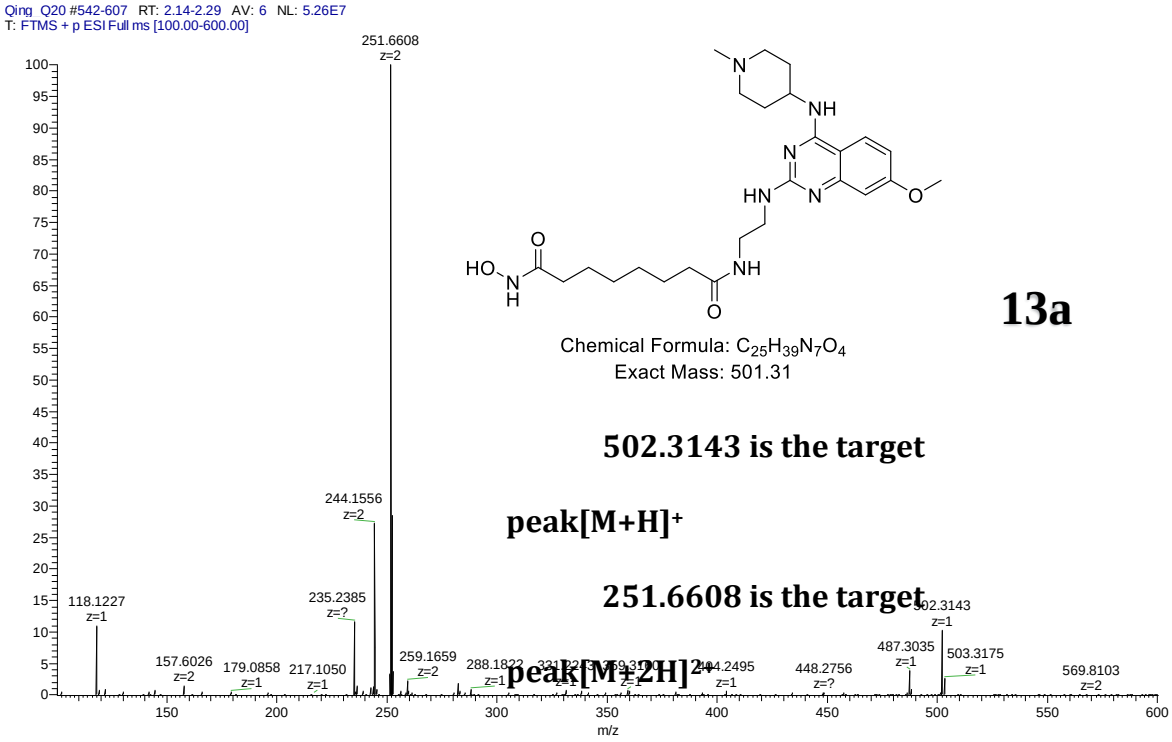
11

Qing_Q45 #543-712 RT: 2.02-2.46 AV: 21 NL: 3.40E8
T: FTMS + p ESI Full ms [180.00-800.00]



Chemical Formula: $C_{28}H_{38}N_6O_3$
Exact Mass: 506.30054

12

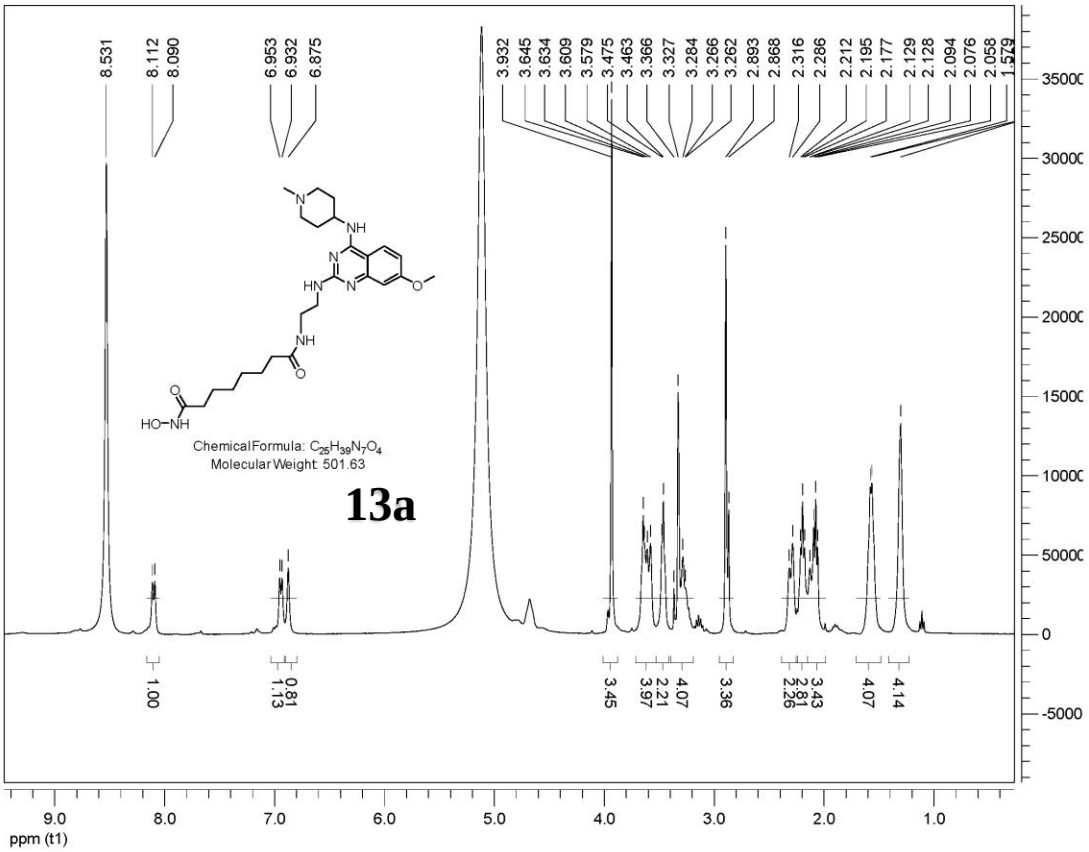


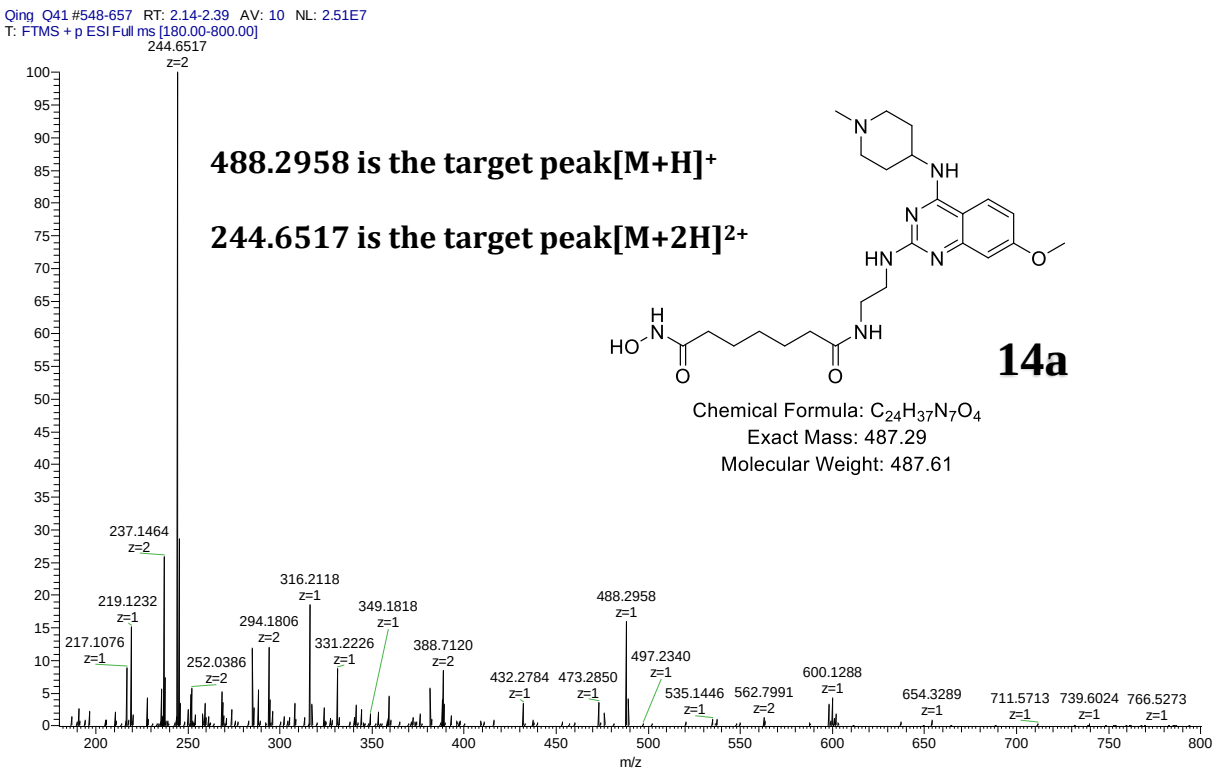
Chemical Formula: $C_{25}H_{39}N_7O_4$
Exact Mass: 501.31

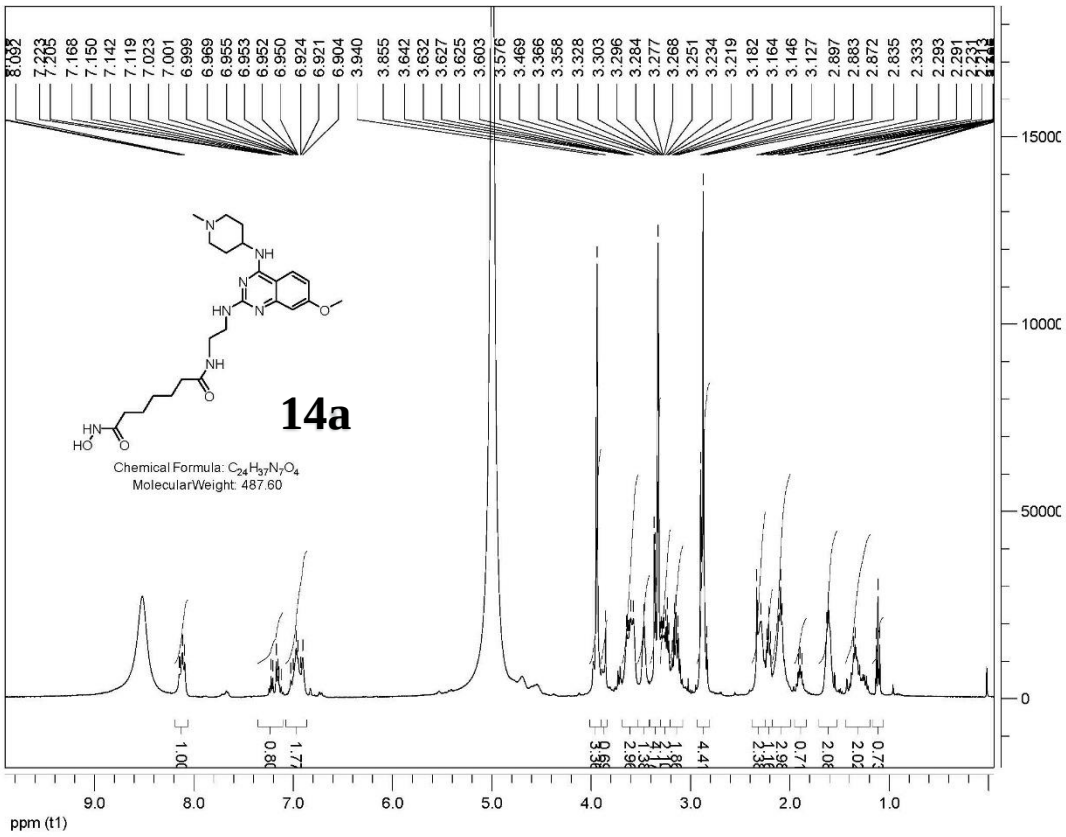
13a

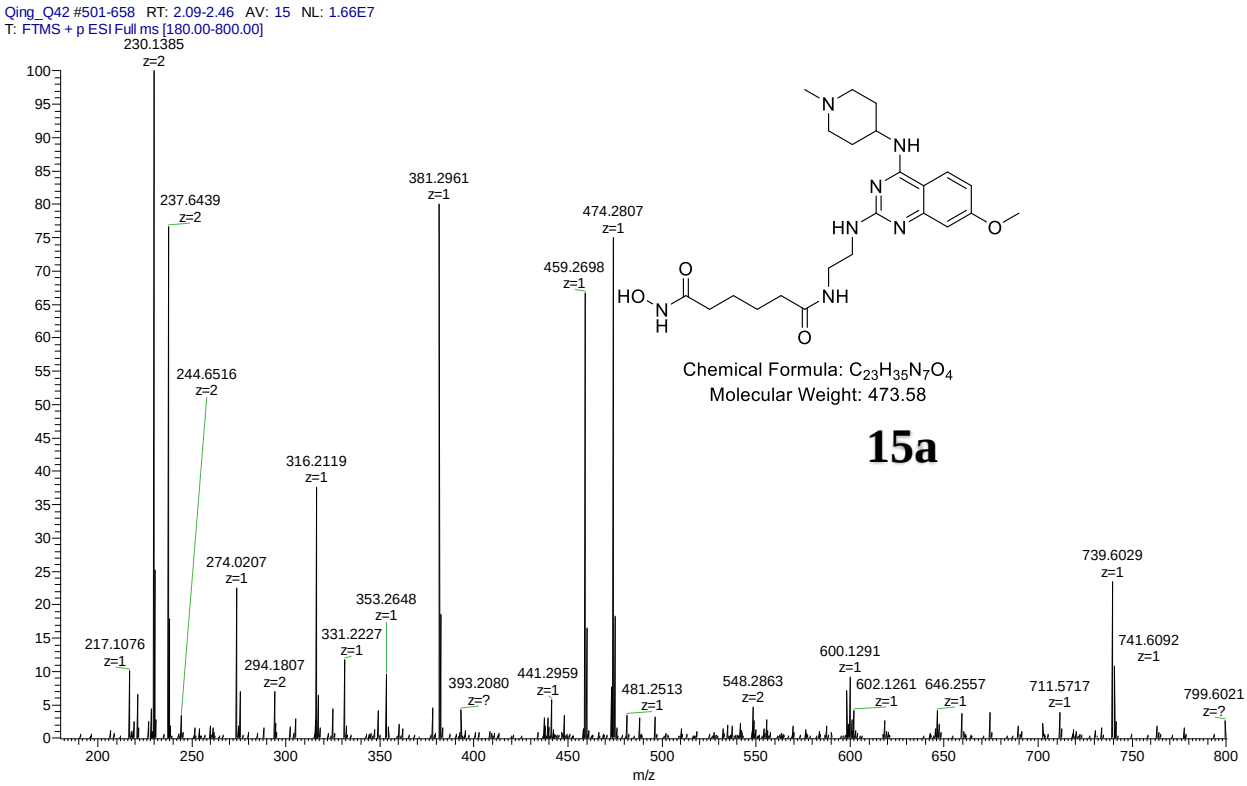
502.3143 is the target peak$[M+H]^+$

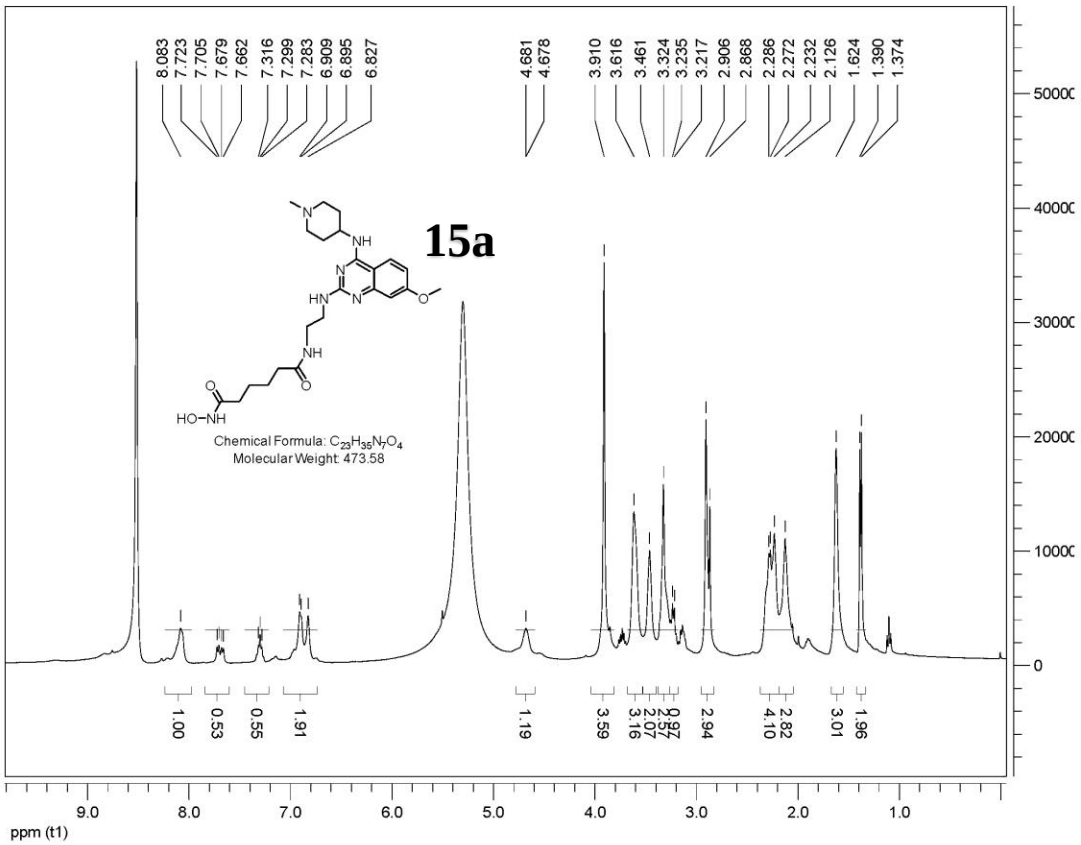
251.6608 is the target peak$[M+2H]^{2+}$

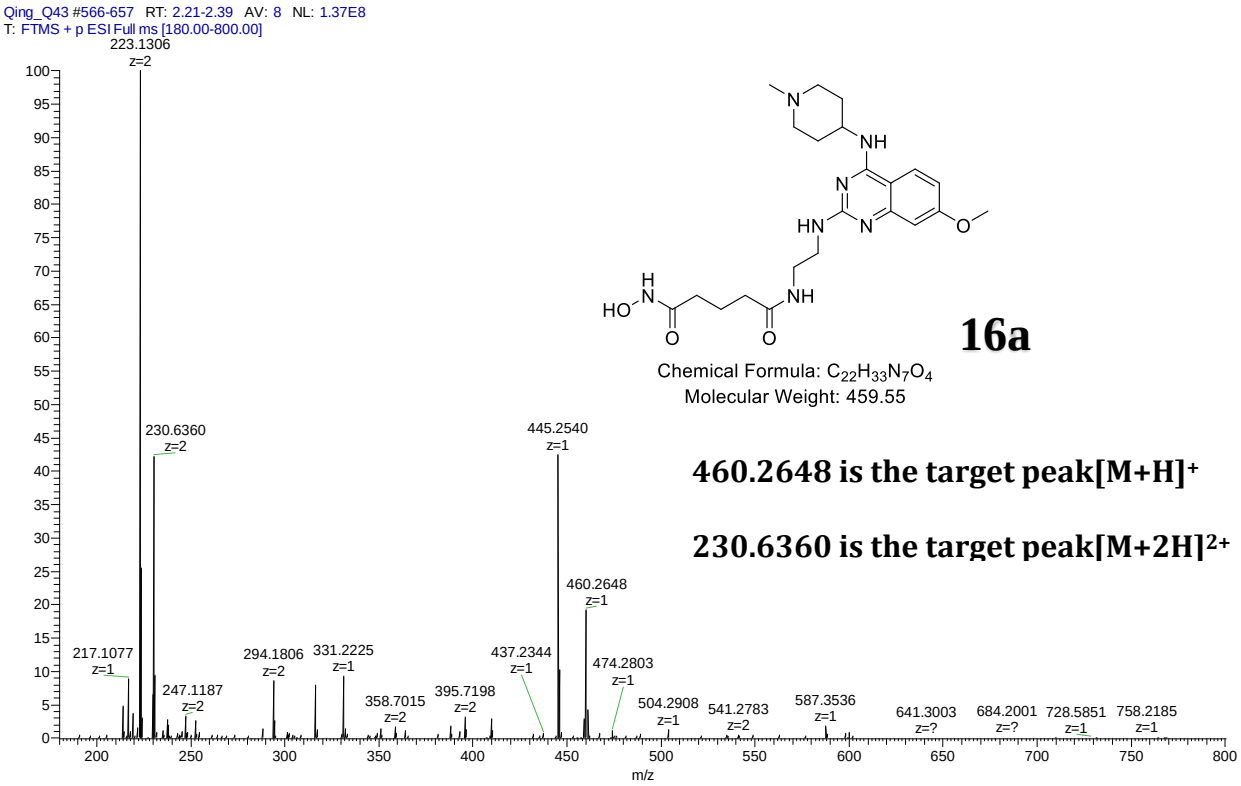










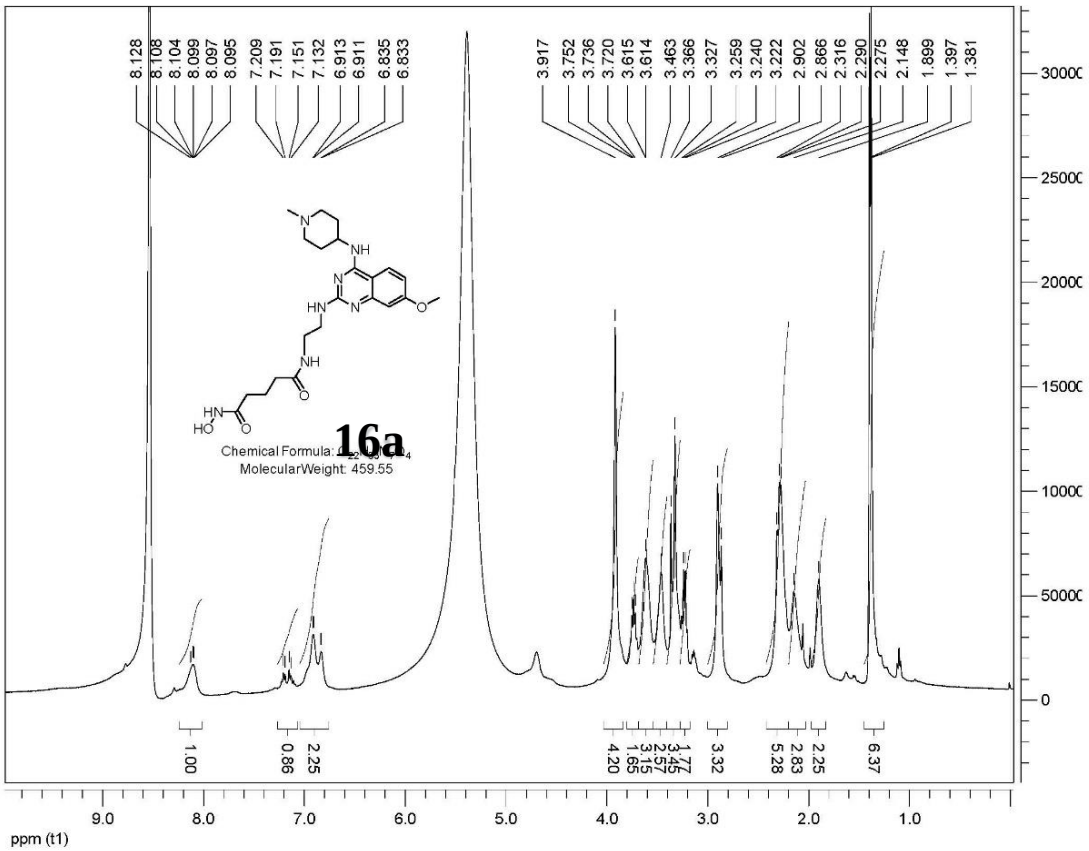


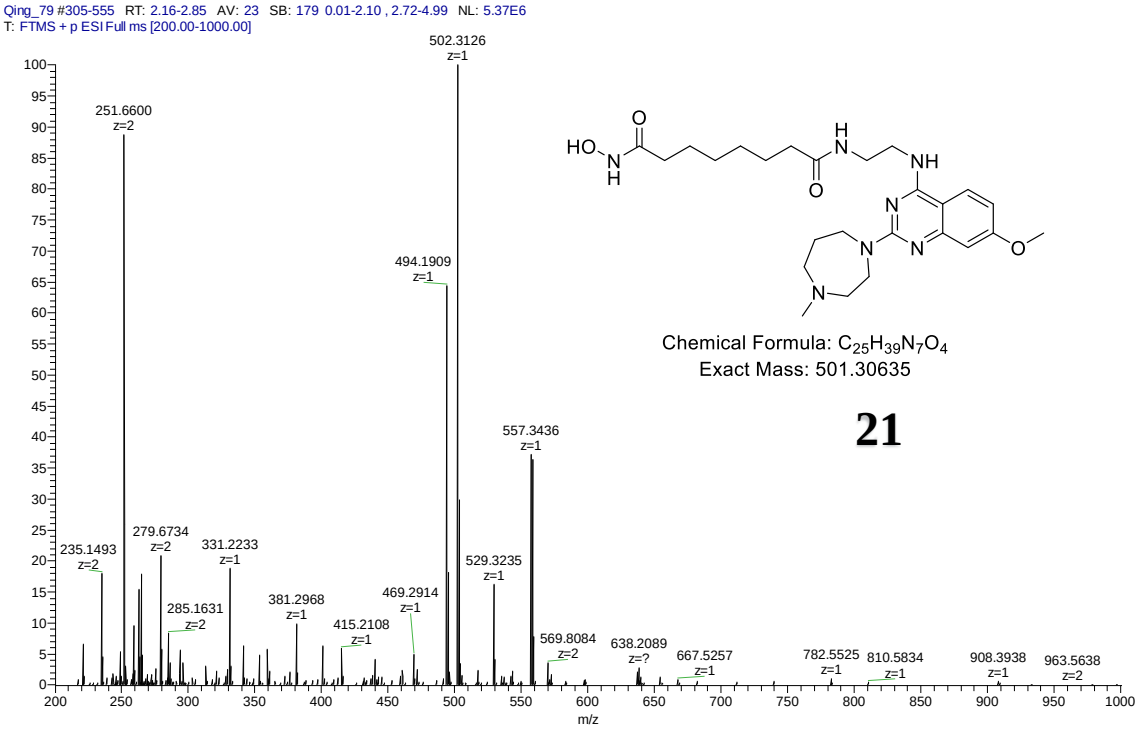
16a

Chemical Formula: $C_{22}H_{33}N_7O_4$
Molecular Weight: 459.55

460.2648 is the target peak$[M+H]^+$

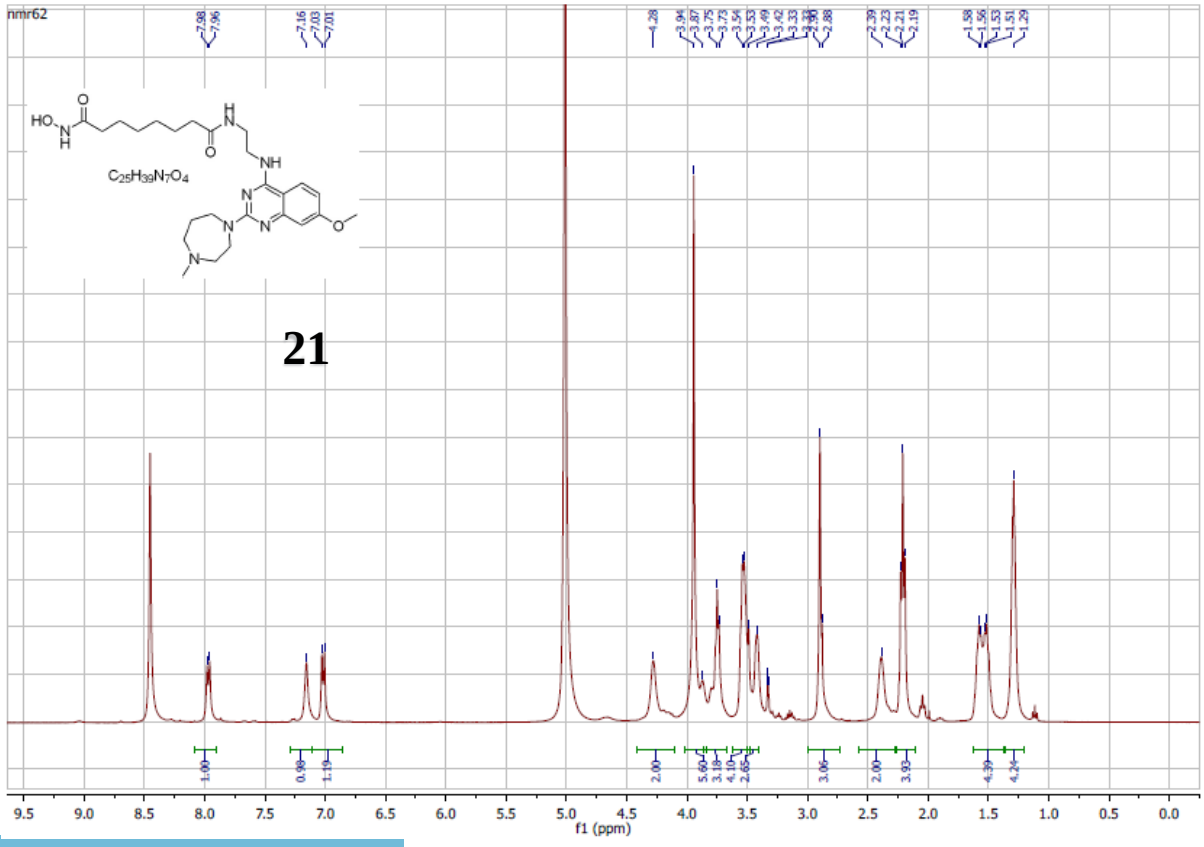
230.6360 is the target peak$[M+2H]^{2+}$

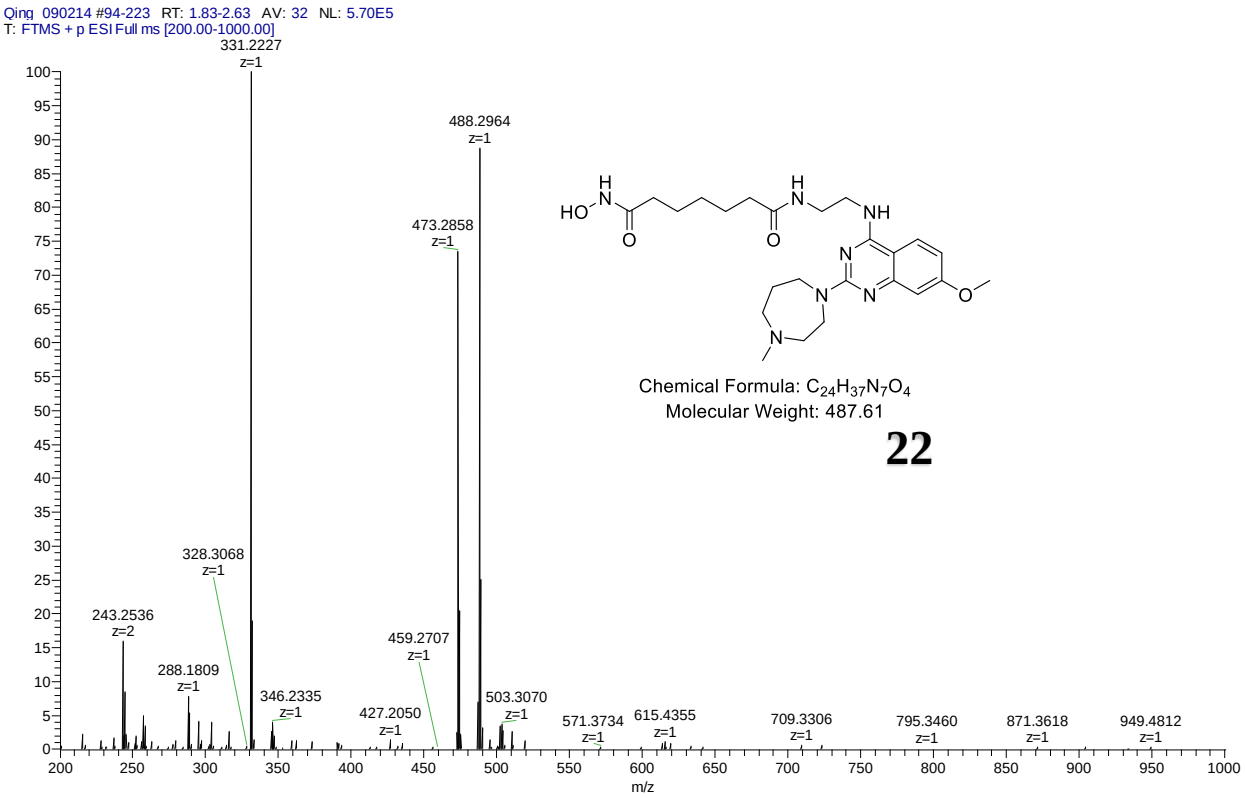




Chemical Formula: $C_{25}H_{39}N_7O_4$
Exact Mass: 501.30635

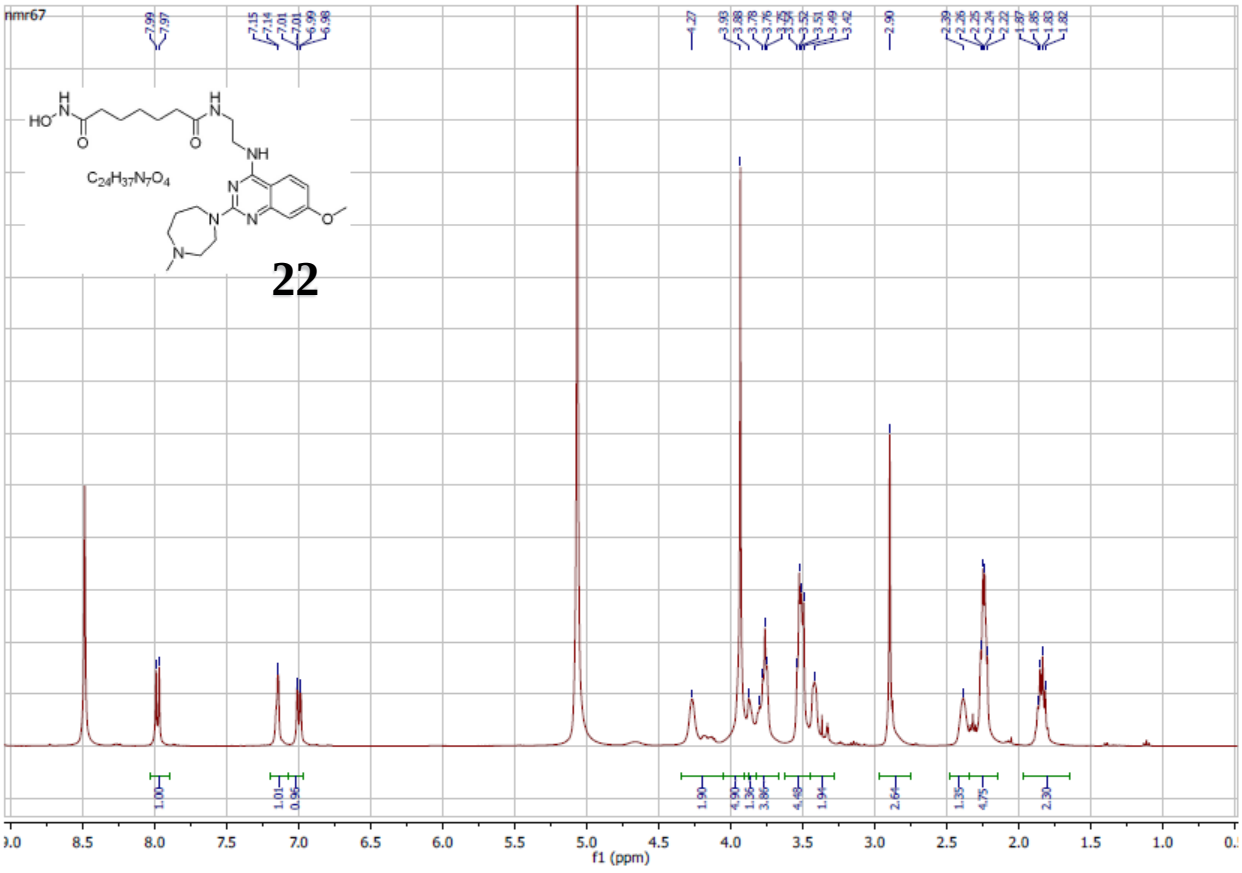
21





Chemical Formula: $C_{24}H_{37}N_7O_4$
Molecular Weight: 487.61

22

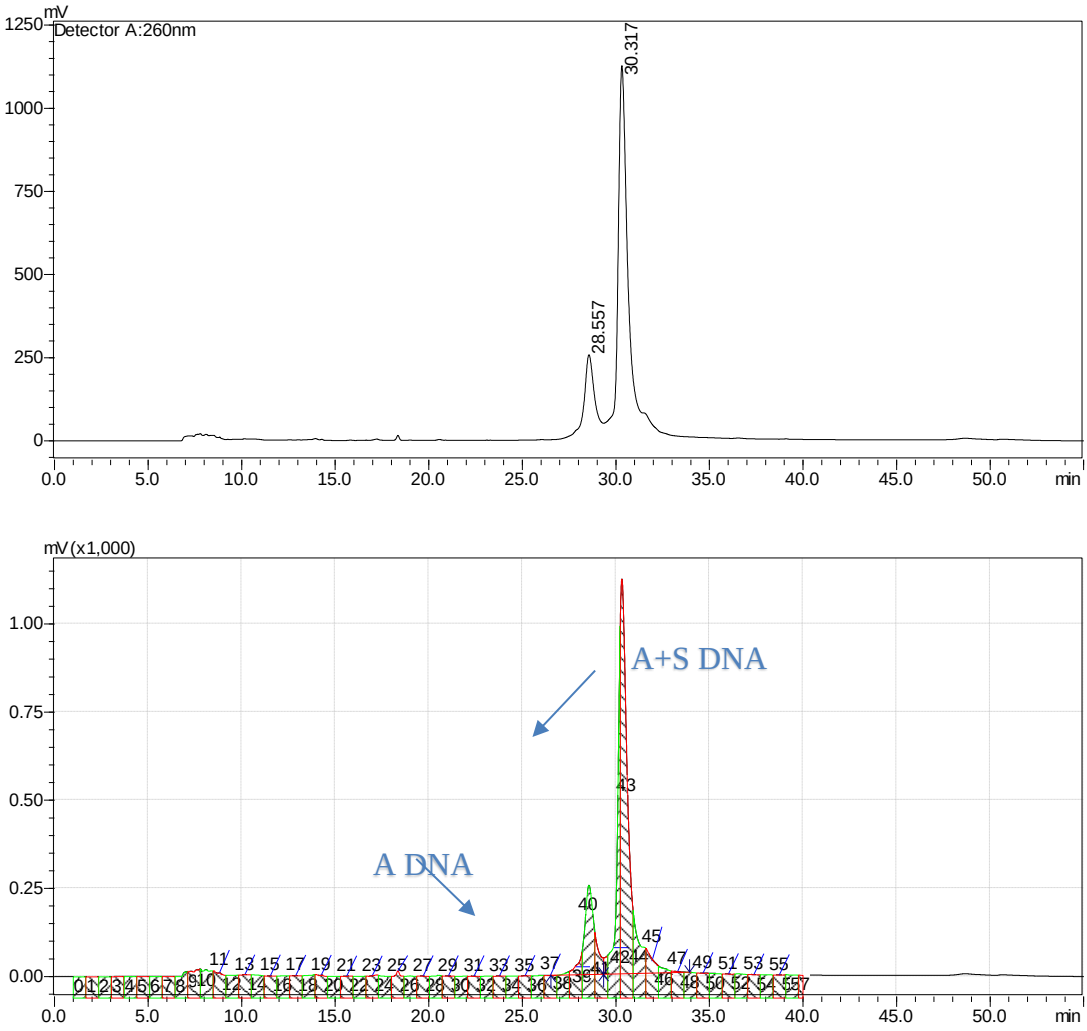


$C_{24}H_{37}N_7O_4$

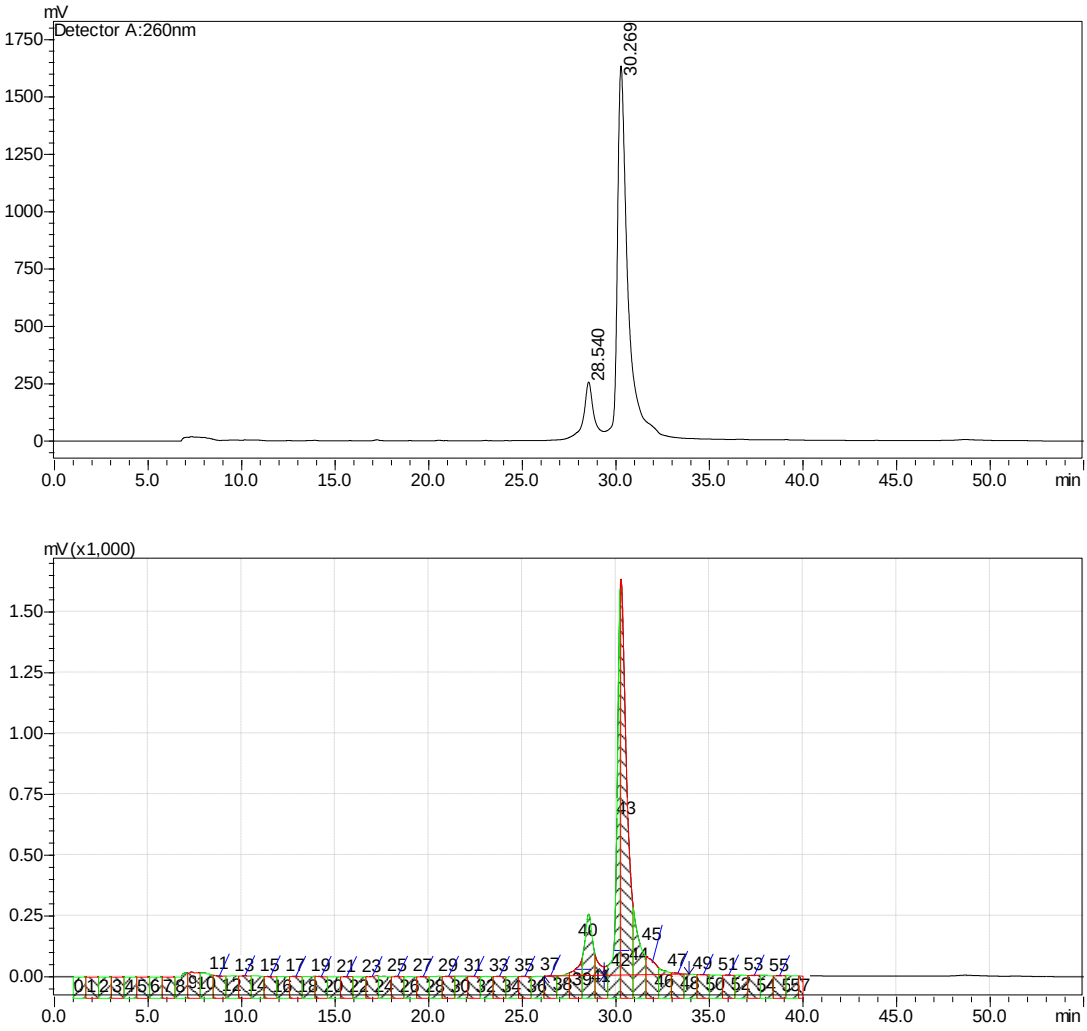
22

Appendix C Supporting information for DNA encoded library of Globo series glycans for early stage detection of cancer

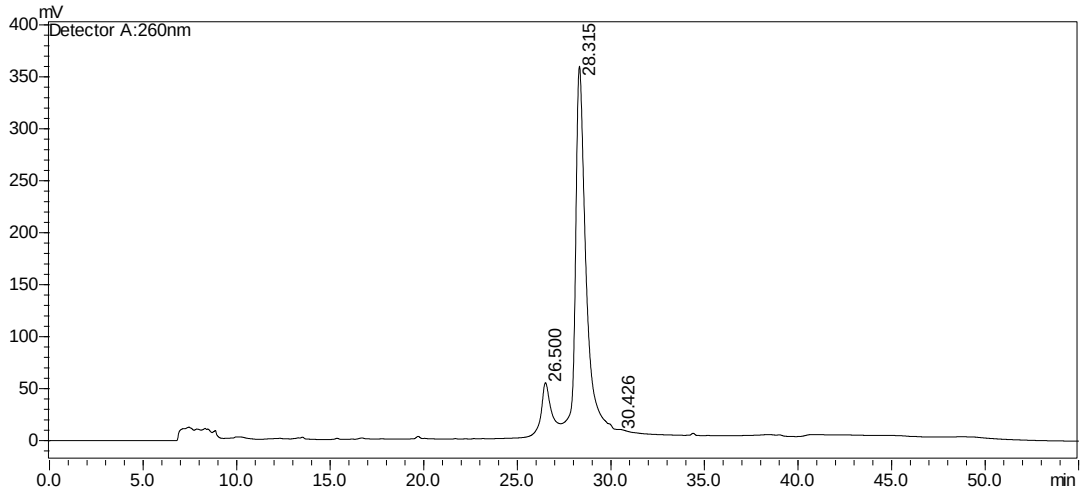
Appendix C.1. HPLC spectra of DNA and Glycan DNA conjugates



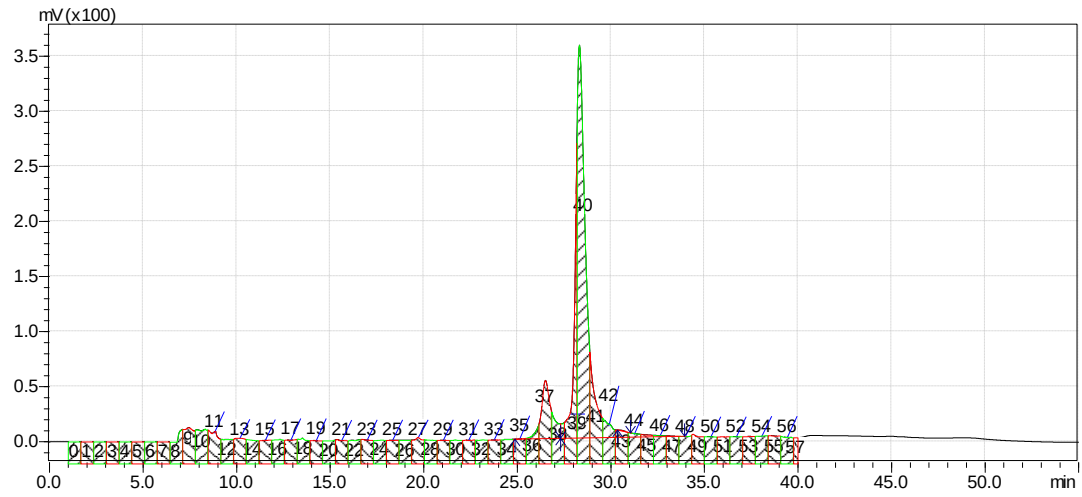
HPLC of A DNA+ Antigen A click reaction



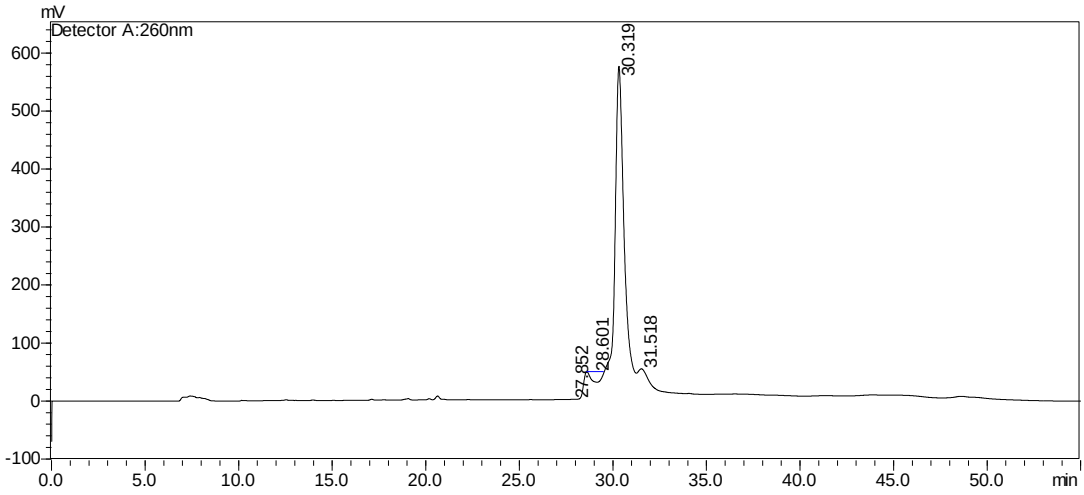
HPLC of B DNA+ Antigen B click reaction



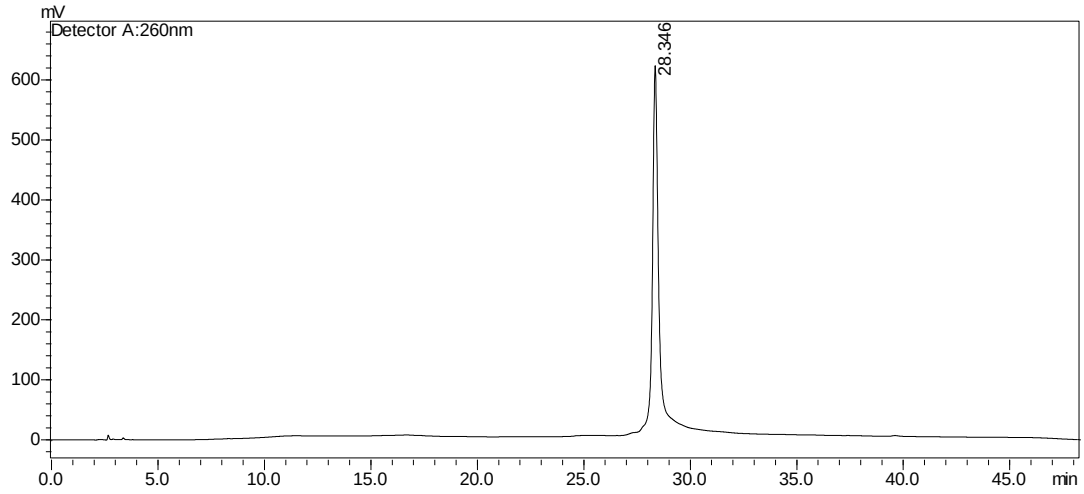
O+S



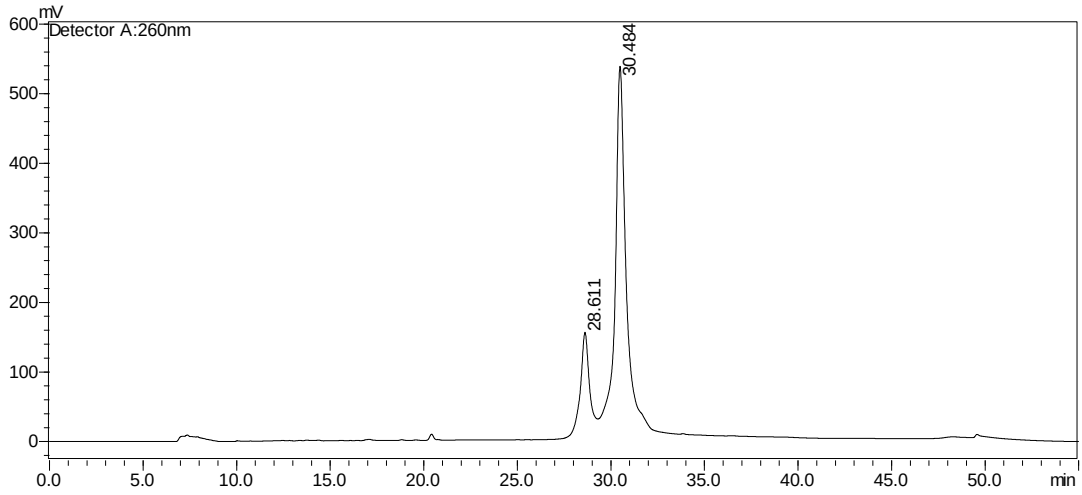
HPLC of O DNA+ Antigen O click reaction



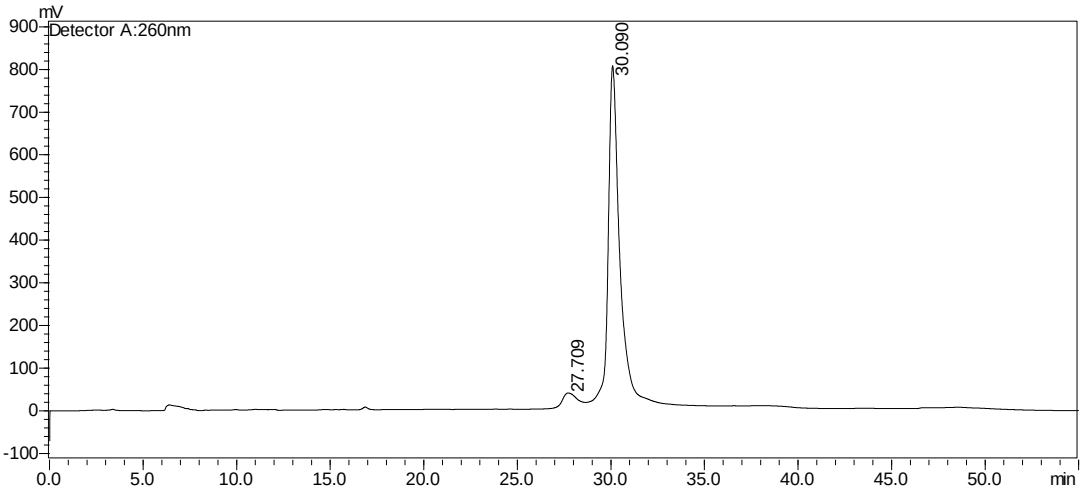
GbH+S



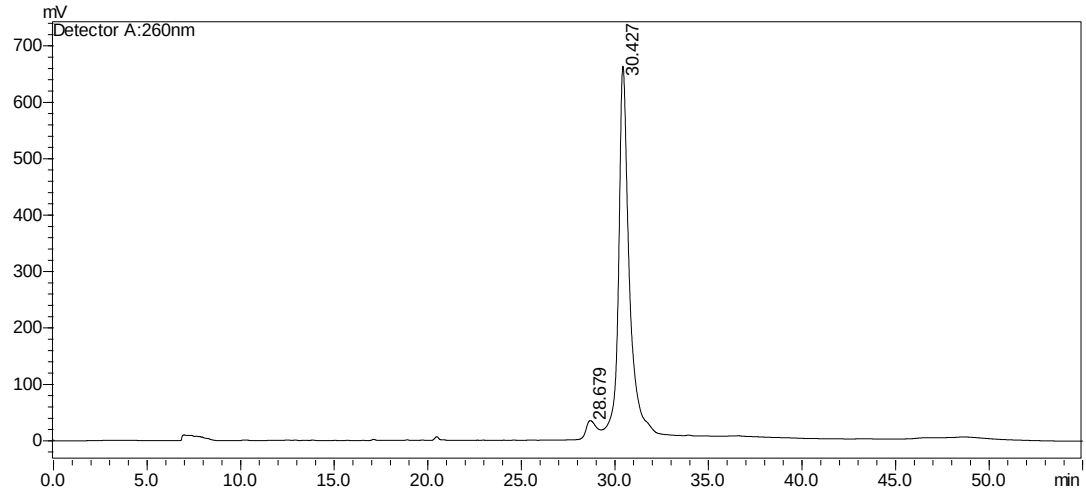
HPLC of GbH DNA used in click reaction after HPLC purification



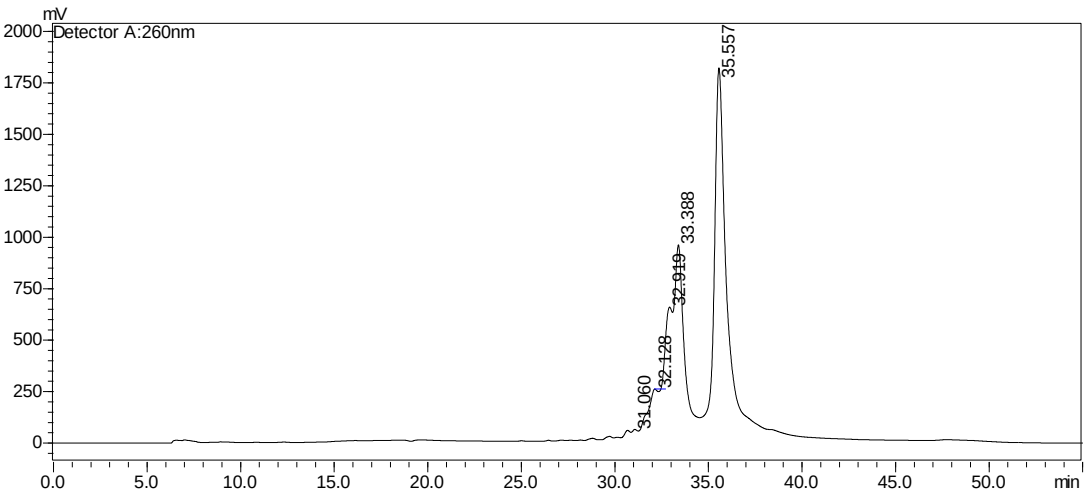
Gb5+S



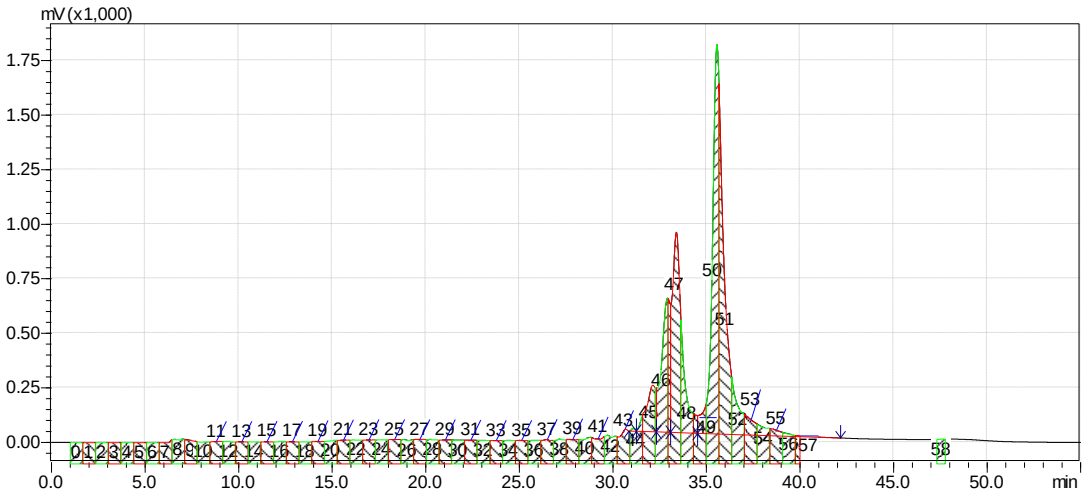
Gb4+S



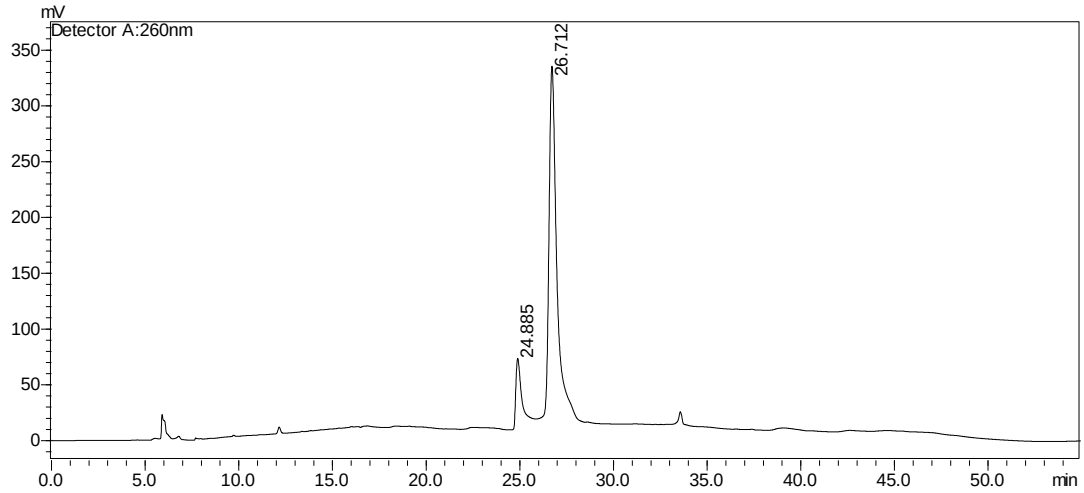
Gb3+S



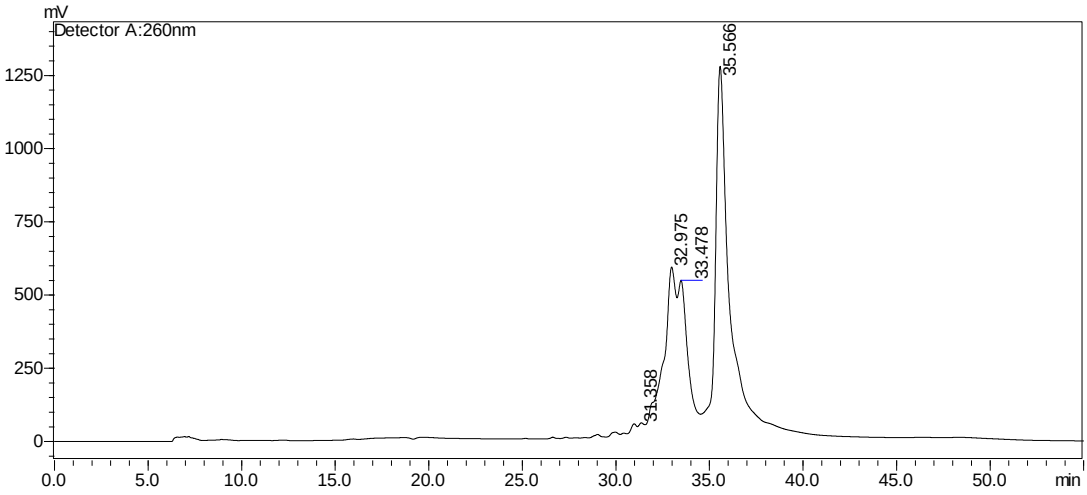
Gb2+S



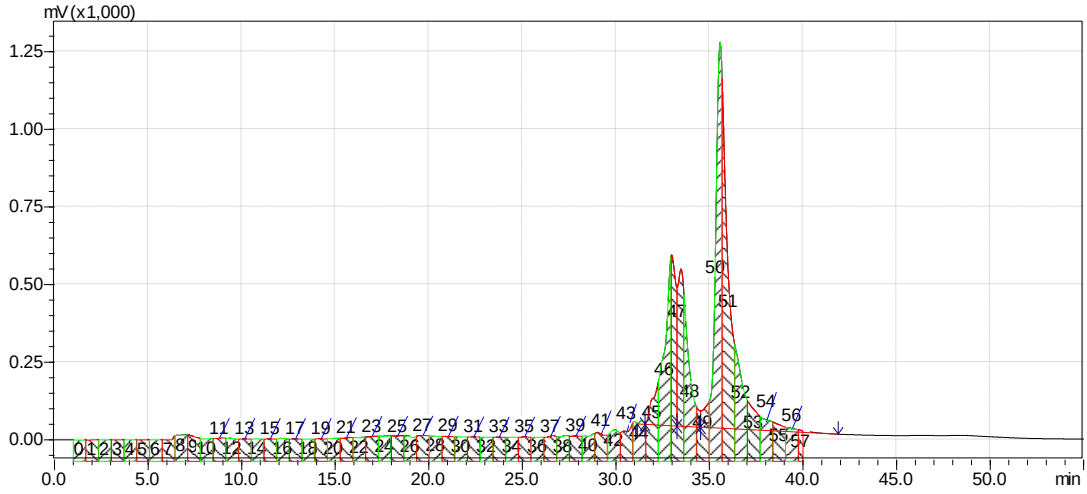
Gb2+S HPLC fraction collection (peak 2 is used in the study)



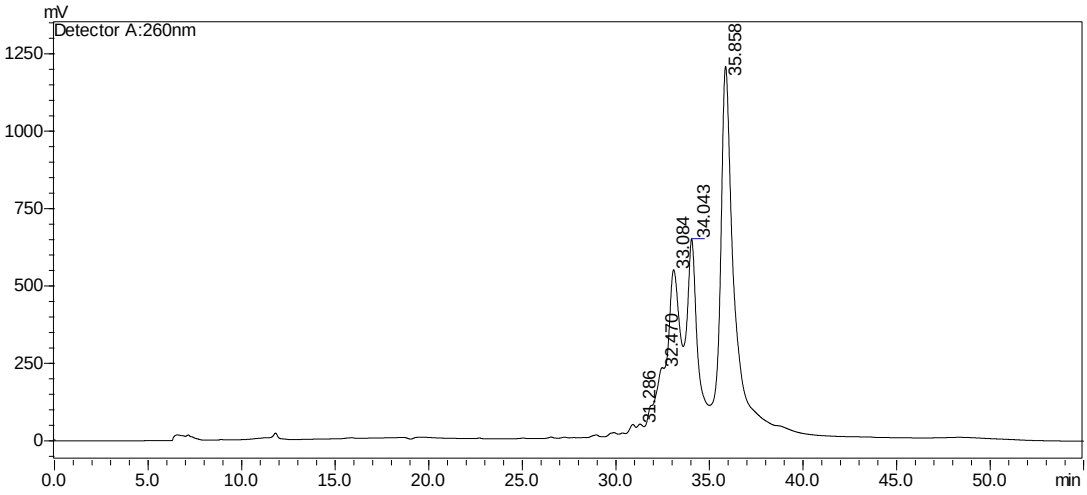
Bb4+S



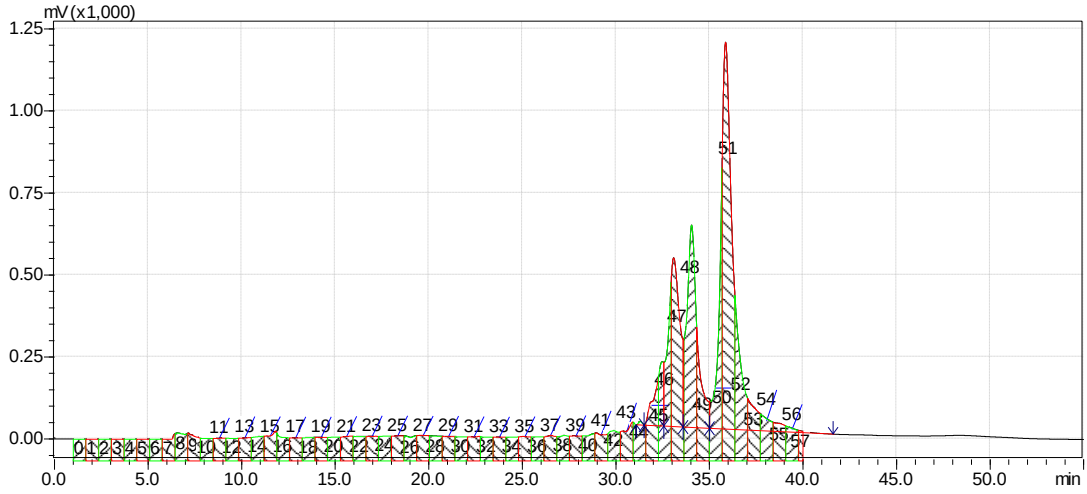
BB3+S



Bb3+S collection (peak two is used for study)

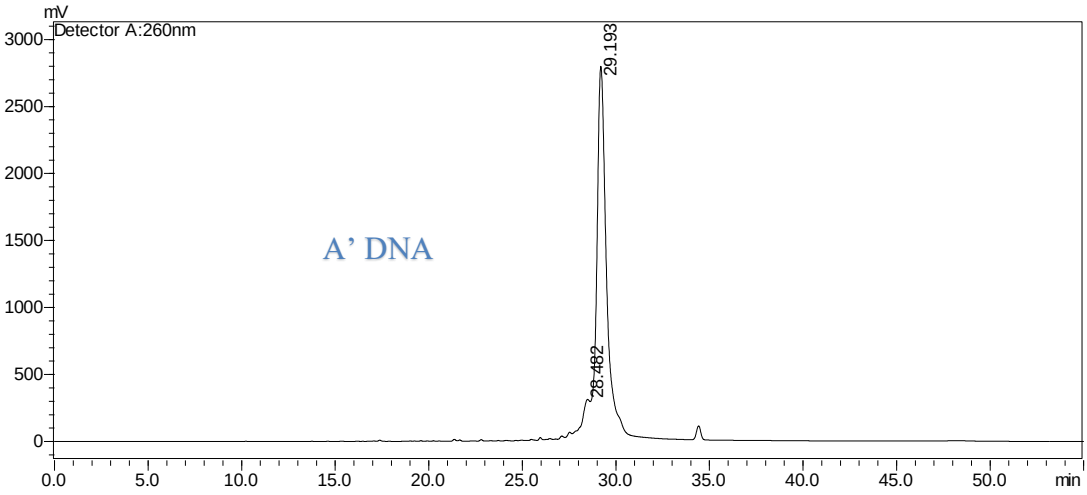


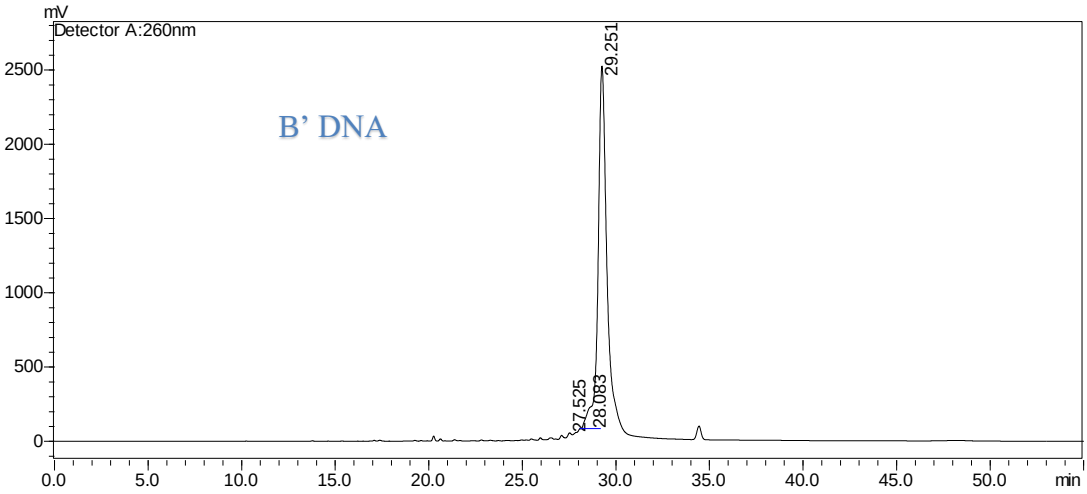
BB2+S

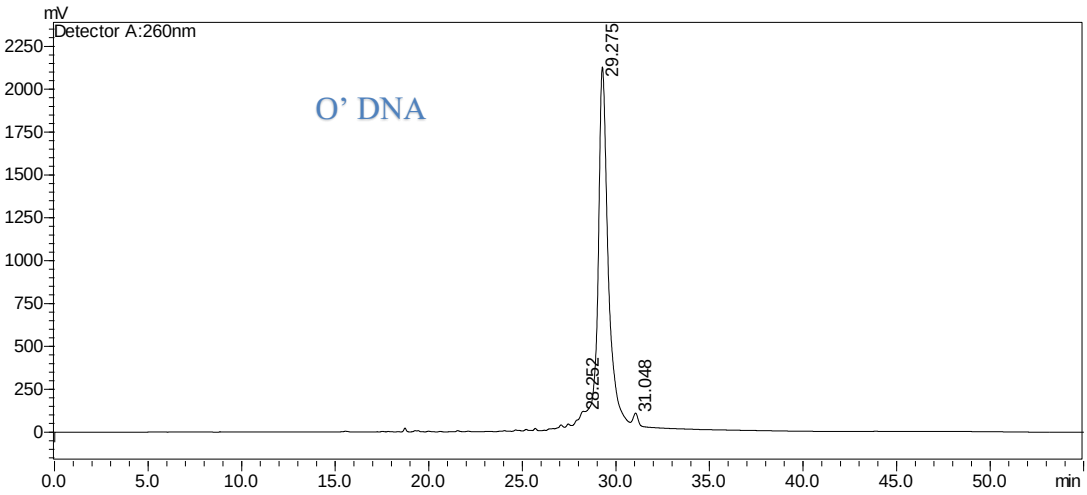


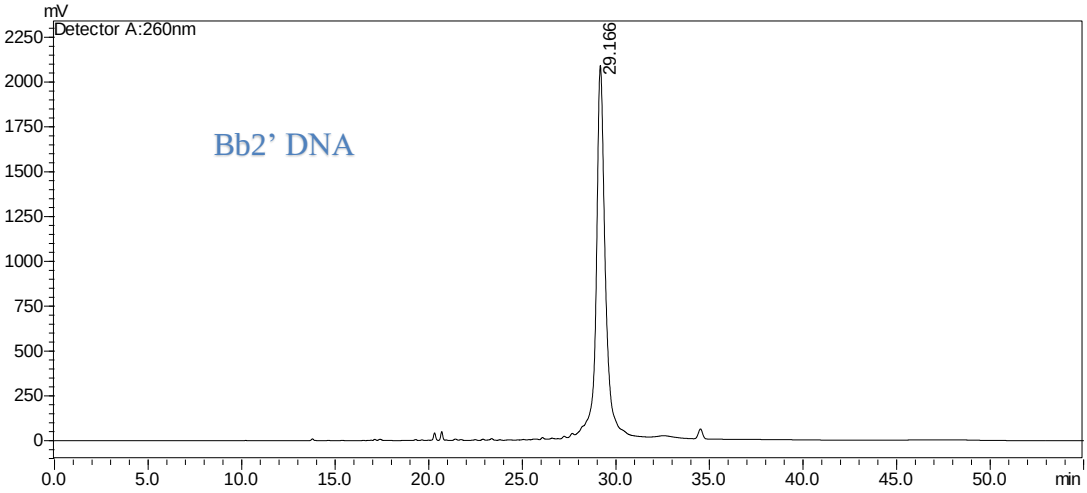
Bb2+S collection (peak two is used for study)

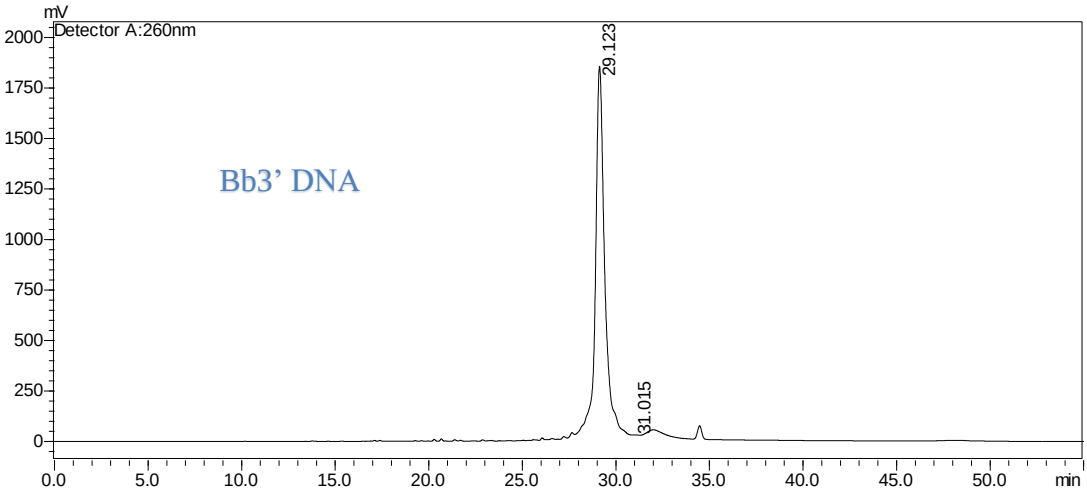
HPLC of complimentary DNA used in hybridization

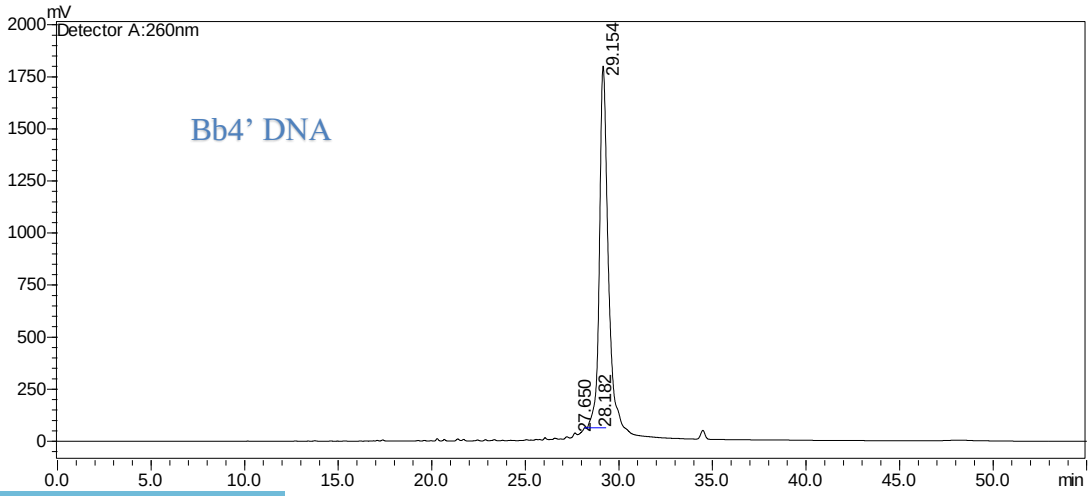


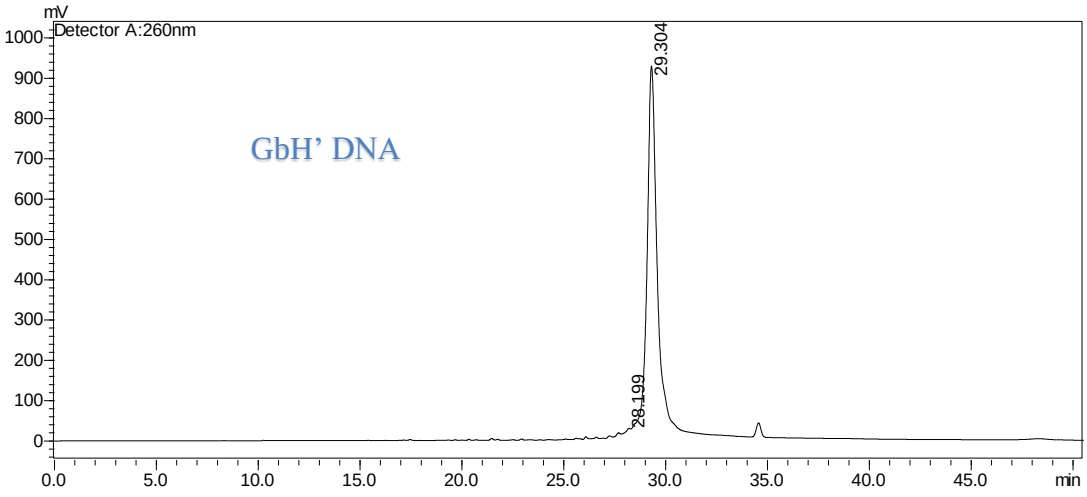


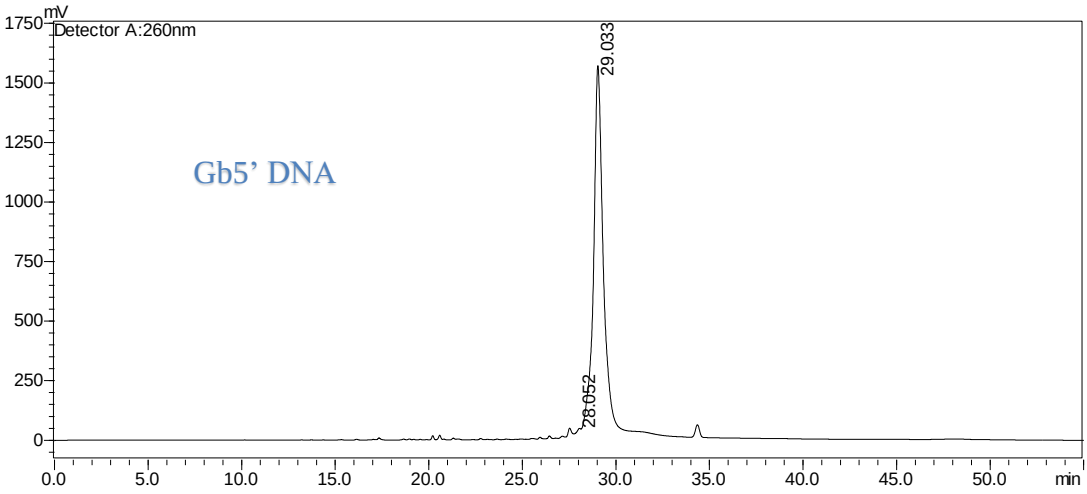


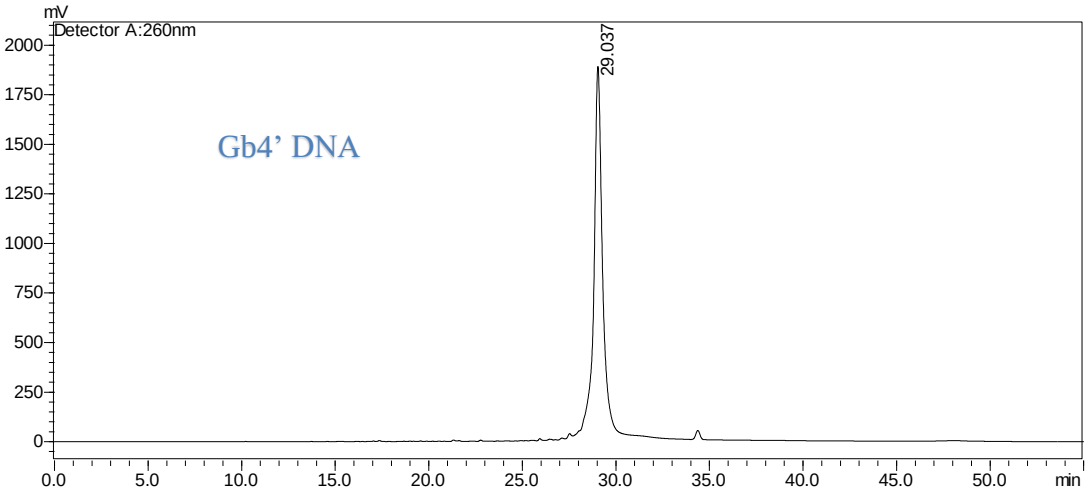


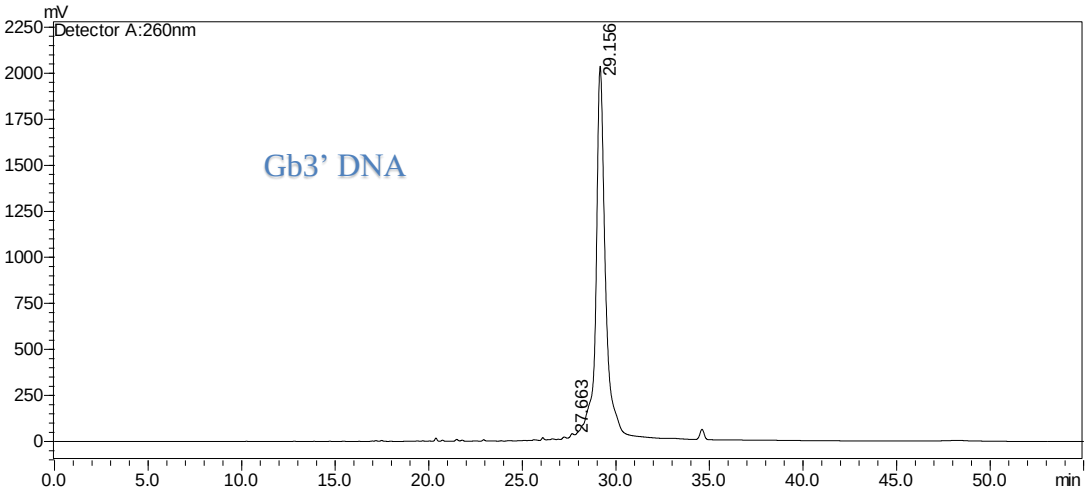


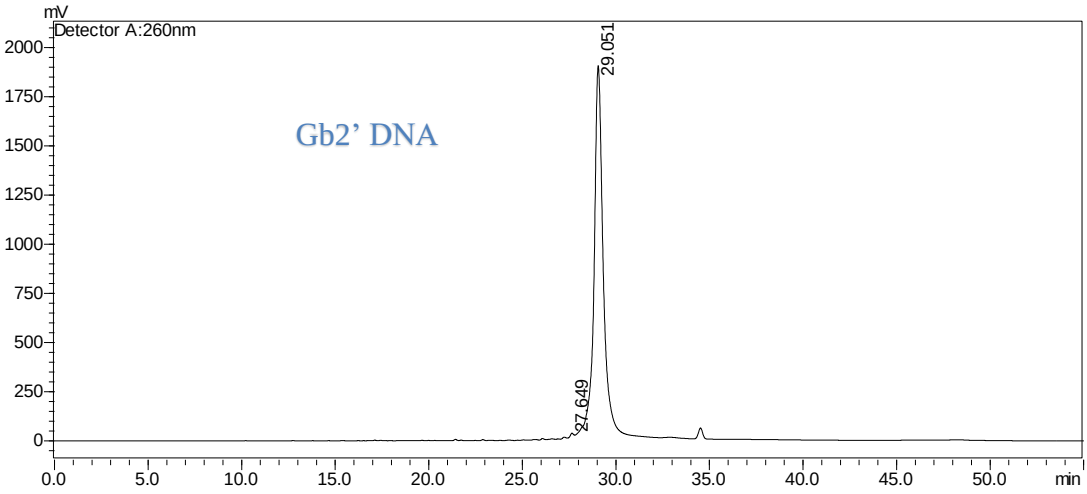












Appendix C.2. Coding Dictionaries and Examples

Sugar Dictionary (Carb Dict):

Table B.5. *Monosaccharides and their codes (library A)*

S.N	Sugar	Sugar Name	DNA Code(as in dictionary)
1.	All	Allose	AAAA
2.	AllNAc	N-Acetyl-	AAAC
3.	AllN	Allosamine	AAAG
4.	AllA	Alluronic	AAAT
5.	Alt	Altrose	AACA
6.	AltNAc	N-Acetyl-	AACC
7.	AltN	Alltrosamine	AACG
8.	AltA	Alturonic	AACT
9.	Glc	Glucose	AAGA
10.	GlcNAc	N-Acetyl-	AAGC
11.	GlcN	Glucosamine	AAGG
12.	GlcA	Glucoronic	AAGT
13.	Man	Mannose	AATA
14.	ManNAc	N-Acetyl-	AATC
15.	ManN	Mannosamin	AATG
16.	ManA	Mannuronic	AATT
17.	G µL	Gulose	ACAA
18.	GulNAc	N-Acetyl-	ACAC
19.	GulN	Gulosamine	ACAG
20.	GulA	Guluronic	ACAT
21.	Ido	Idose	ACCA
22.	IdoNAc	N-Acetyl-	ACCC
23.	IdoN	Idosamine	ACCG
24.	IdoA	Idouronic	ACCT
25.	Gal	Galactose	ACGA
26.	GalNAc	N-Acetyl-	ACGC
27.	GalN	Galactosami	ACGG
28.	GalA	Galacturonic	ACGT
29.	Tal	Talose	ACTA
30.	TalNAc	N-Acetyl-	ACTC
31.	TalN	Talosamine	ACTG
32.	TalA	Taluronic	ACTT
33.	Tag	Tagatose	ATCG
34.	TagA	Tagaturonic	ATCT
35.	Hep	Heptose	ATCG
36.	DDManH	D-Glycero-	ATGG
37.	Dha	3-deoxy-D-	ATGT

38.	LDManH	L-glycero-D-	ATTA
39.	Fuc	Fucose	AGAA
40.	FucNAc	*N*-acetyl-L-	AGAC
41.	FucN	Fucosamine	AGAG
42.	Qui	Quinovose	AGAT
43.	QuiNAc	*N*-acetyl-D-	ATAA
44.	QuiN	Quinovosami	ATAC
45.	Rha	Rhamnose	ATAG
46.	RhaNAc	*N*-acetyl-L-	ATAT
47.	RhaN	Rhamnosami	ATCA
48.	Tyv	Tyvelose	ATTC
49.	Oli	Olivose	ATTG
50.	Par	Paratose	ATTT
51.	Dig	Digitoxose	CAAA
52.	6dAlt	6-deoxy-L-	CAAC
53.	6dTal	6-deoxy-D-	CAAG
54.	Abe	Abequose	CAAT
55.	Api	Apiose	CACA
56.	Col	Colitose	CACC
57.	Neu5Ac	*N*-	AGGC
58.	Neu5GC	*N*-	AGGG
59.	Kdn	3-Deoxy-D-	AGGT
60.	Kdo	3-Deoxy-D-	ATCC
61.	Mur	Muramic	CACG
62.	MurNAc	*N*-	CACT
63.	MurNGc	*N*-	CCAA
64.	Sia	Sialic Acid	CCAG
65.	Neu	Neuraminic	CCAC
66.	Fru	Fructose	ATGA
67.	Sor	Sorbose	ATGC
68.	Psi	Psicose	ATAA
69.	Rib	Ribose	GGGG
70.	Ara	Arabinose	ATAT
71.	Xyl	Xylose	ATCA
72.	Lyx	Lyxose	ATCT
73.	Bac	Bacillosamin	CCAT
74.	R	For	CCCCAGTCAGGCCTAAC

Linkages Dictionary (Link Dict)

Table B.6. *Linkages/branching and their codes (library B)*

S.No	Linkage	DNA
1.	a1-1	AAT
2.	a1-2	AAG
3.	a1-3	AAC
4.	a1-4	ATA
5.	a1-5	AGA
6.	a1-6	ACA
7.	a1-7	ATT
8.	a1-8	ATC
9.	a2-1	GTA
10.	a2-2	GTC
11.	a2-3	GGT
12.	a2-4	GGC
13.	a2-5	GGA
14.	a2-6	GTT
15.	a2-7	GAC
16.	a2-8	GAT
17.	b1-1	ATG
18.	b1-2	TCA
19.	b1-3	TCT
20.	b1-4	TGT
21.	b1-5	TCC
22.	b1-6	TTA
23.	b1-7	TTC
24.	b1-8	GAA
25.	a?	GCC
26.	b?	CAA
27.	?1	CAC
28.	?2	CTC
29.	?3	CCA
30.	?4	GGG
31.	?5	CCT
32.	?6	AGG
33.	?7	CCG
34.	?8	CGC
35.	??	AGT
36.	)	AAA
37.	(	TTT

Modification Dictionary (Mod Dict):

Table B.7. *Modifications and their codes (library C)*

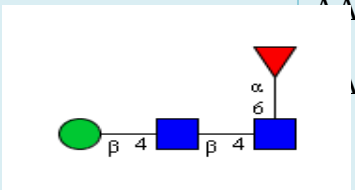
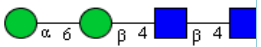
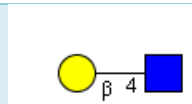
.No	Compound	Modification	NA	D Code
	Anhydro	[2Y]		AG
		[4Y]		A
		[7Y]		A
		[8Y]		A
		[?Y]		CC
	Hydroxyl	[2O		A
		[4O		A
		[7O		A
		[8O		A
		[?O		C
	Pyruvate	[2V]		C
		[4V]		C
		[7V]		C
		[8V]		C
		[?V]		C
	Sulphate	[2S]		C
		[4S]		C
		[7S]		C
		[8S]		C
		[?S]		C
	Phosphate	[2P]		C
		[4P]		C
		[7P]		C
		[8P]		C
		[?P]		C
	n-glycolyl	[2J]		C
		[4J]		C
		[7J]		C
		[8J]		C
		[?J]		CT
	n-acetyl	[2N		CT
		[4N		CT
		[7N		CT
		[8N		CT
		[?N		CT
	o-acetyl	[2Ac		CT
		[4Ac		CT

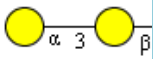
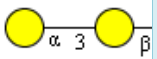
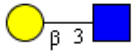
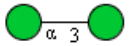
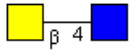
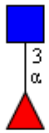
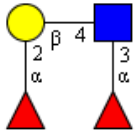
		[7Ac	CT
		[8Ac	CT
		[?Ac	CT
	Carboxylat	[2C	CT
		[4C	G
		[7C	G
		[8C	G
		[?C	G
	Inositol	[2IN	G
		[4IN	G
		[7IN	G
		[8IN	G
		[?IN	G
	pentyl	[2EE	G
		[4EE	G
		[7EE	G
		[8EE	G
		[?EE	G
	octyl	[2E	G
		[4E	G
		[7E	GT
		[8E	GT
		[?E	GT
	deactylated	[2Q]	GT
		[4Q]	TA
		[7Q]	TA
		[8Q]	TA
		[?Q]	TA
	N-Sulfate	[2Q	TA
		[4Q	TA
		[7Q	TA
		[8Q	TA
		[?QS	TA
	Pyruvate	[2P	TA
		[4P	TA
		[7P	TA
		[8P	TA
		[?PY	TA
	N-	[2E	TA
		[4E	TA
		[7E	TC
		[8E	TC
		[?EC	TC
	Phosphoch	[2PC	TC

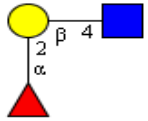
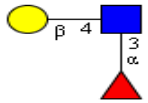
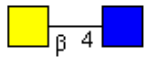
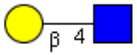
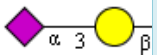
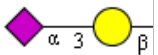
		[4PC	TG
		[7PC	TG
		[8PC	TG
		[?PC	TG
	Phosphoeth	[2PE	TG
		[4PE	TG
		[7PE	TG
		[8PE	TG
		[?PE	TG
	Methyl	[2M	TG
		[3M	TG
		[4M	TG
		[7M	TT
		[8M	TT
		[?M	TT
	Sulphate	[3S]	TT

Core Dictionary (Core Dict):

Table B.8. *Core structure and their codes (Library D)*

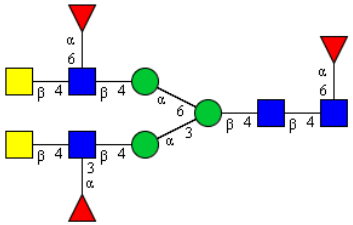
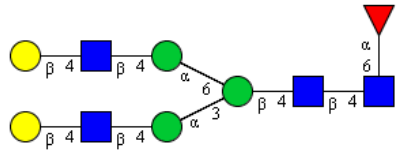
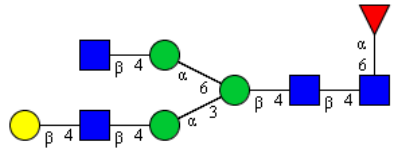
S.No	Sugar blocks for Secondary naming	CFG structure	2D	Long Nucleotide	Length	Assigned short DNA code
.	*Man* b1-4 *GlcNAc* b1-4(*Fuc* a1-6) *GlcNAc*			ATATGTAAGCTGTTTTAGAAAC AAAAAGC	31	GTACA
2.	*Man* a1-6*Man* b1-4 *GlcNAc* b1-4 *GlcNAc*			AATAACAAATATGTAAGCTGTAA GC	25	GTACT
3.	*Gal* b1-4 *GlcNAc* b1-4			ACGATGTAAGCTGT	14	GTCCT

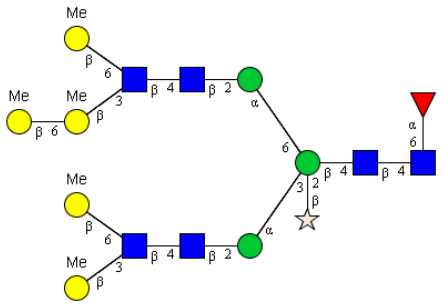
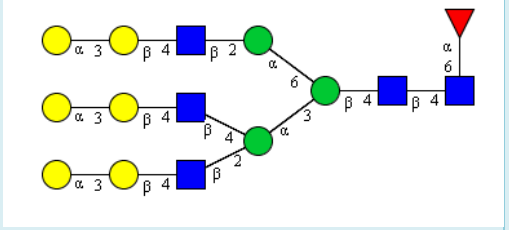
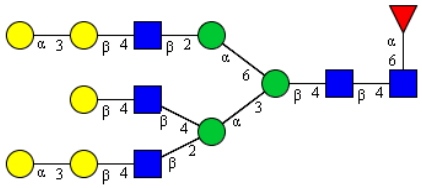
4.	*Gal* a1-3 *Gal* b1- 4 *GlcNAc* b1-2		ACGAAACACGATGTAAGC TCA	21	GTACC
5.	*Gal* a1-3 *Gal* b1- 4 *GlcNAc* b1-3		ACGAAACACGATGTAAGC TCT	21	GTTAT
6.	*Gal* b1-3 *GlcNAc* b1-4		ACGATCTAAGCTGT	14	GTTCA
7.	*Man* a1-3 *Man* a1-6		AATAAACAATAACA	14	GTTTG
8.	*GalNAc* b1-4 *GlcNAc* b1-2		ACGCTGTAAGCTCA	14	GTTTC
9.	*Fuc* a1-3 *GlcNAc* b1-2		AGAAAACAAGCTCA	14	GTTGT
10.	*Fuc* a1-2 *Gal* b1-4(*Fuc* a1- 3) *GlcNAc* b1-2		AGAAAAGACGATGTTTTAGAAA ACAAAAAGCTCA	34	GGATT

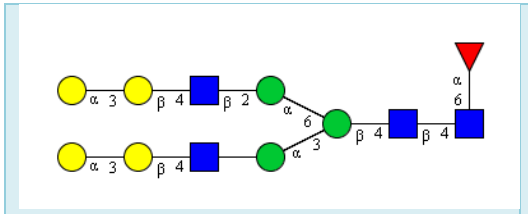
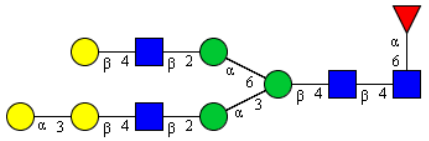
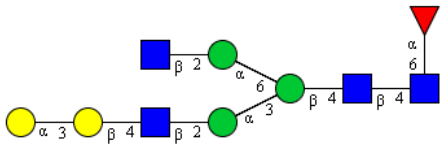
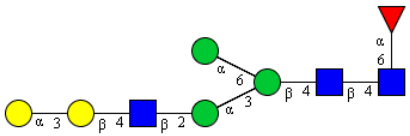
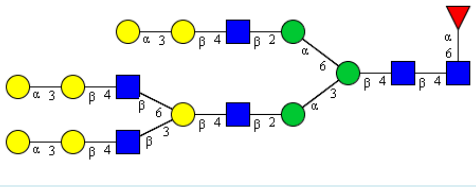
11.	*Fuc* a1-2 *Gal* b1- 4 *GlcNAc* b1-2			AGAAAAGACGATGTAAGCTCA	21	GTTGC
12.	*Gal* b1-4(*Fuc* a1- 3) *GlcNAc* b1-4			ACGATGTTTTAGAAAACAAAAAG CTGT	27	GTGAG
13.	*GalNAc* b1-4 *GlcNAc* b1-6			ACGCTGTAAGCTTA	14	GTGTC
14.	*Gal* b1-4 *GlcNAc* b1-6			ACGATGTAAGCTTA	14	GTGCG
15.	*NeuAc* a2-3 *Gal* b1- 4 *GlcNAc* b1-4			AGGCGGTACGATGTAAGCTGT	21	GTGGA
16.	*NeuAc* a2-3 *Gal* b1- 4 *GlcNAc* b1-6			AGGCGGTACGATGTAAGCTTA	21	GTGGT

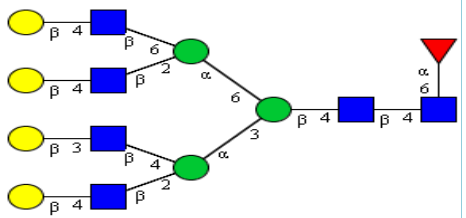
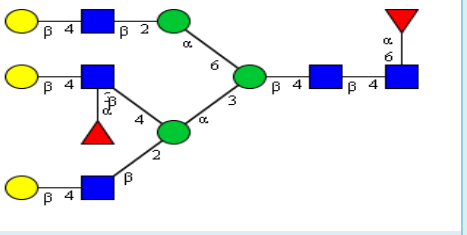
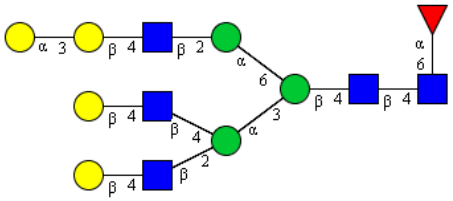
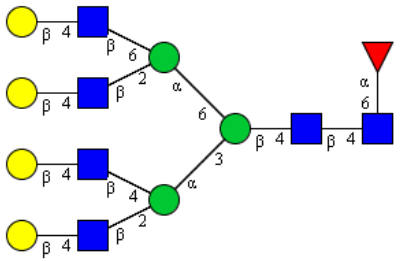
Coding Examples:

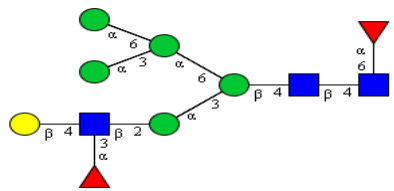
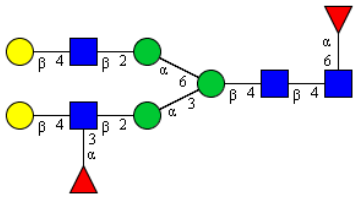
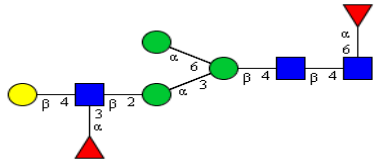
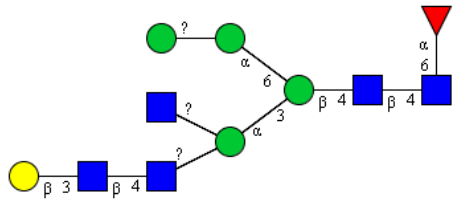
Table B.9. *Examples of DNA encoding of glycans using the program*

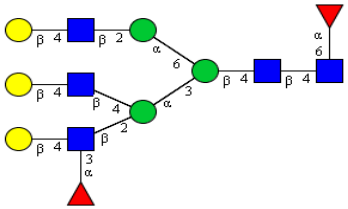
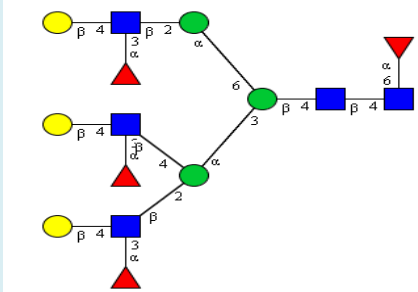
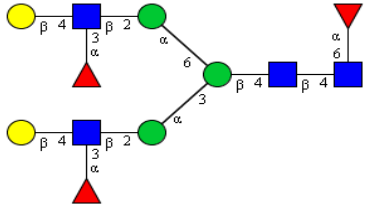
Structure	Long Nucleotide	Lengt	Short	Lengt
	acgctgttttagaaaagaaaaagctgtaataaactttacgct gttttagaaaagaaaaagctgtaataacaaaaaatatgtaa gctgttttagaaacaaaaaagc	105	acgctgttttagaaaagaaaa agctgtaataaactttacgctg ttttagaaaagaaaaagctgt aataacaaaagtaca	79
	acgatgtaagctgtaataaactttacgatgtaagctgtaata acaaaaaatatgtaagctgttttagaaacaaaaaagc	79	gtacgaataaactttacgatgt aagctgtaataacaaaagtac a	44
	acgatgtaagctgtaataaactttaagctgtaataacaaaa aatatgtaagctgttttagaaacaaaaaagc	72	gtacgaataaactttaagctgt aataacaaaagtaca	37

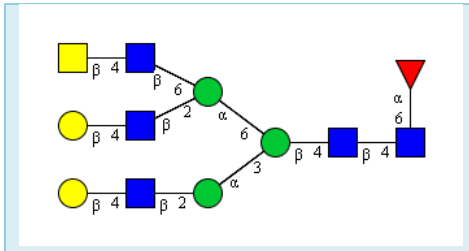
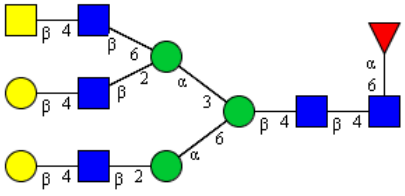
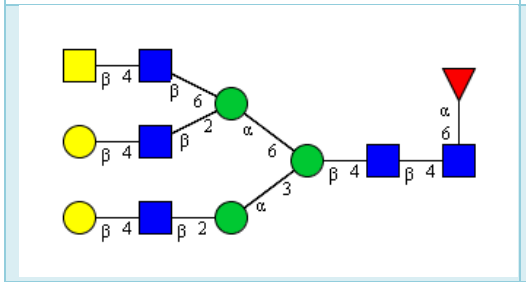
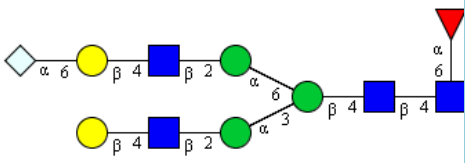
	acgatggttcttttacgatgggttaaaaaagctgtaagctca aataaactttacgatgggttaacgatgggtcttttacgatgg gttaaaaaagctgtaagctcaaataacaaaatttatcatcaa aaaatatgtaagctgttttagaaacaaaaaagc	159	acgatggttcttttacgatggg ttaaaaaagctgtaagctcaa ataaactttacgatgggttaac gatgggtcttttacgatgggtt aaaaaagctgtaagctcaaat aacaaaatttatcatcaaaagt	133
	acgaaacacgatgtaagctcatttacgaaacacgatgtaa gctgtaaaaataaactttacgaaacacgatgtaagctcaaa taacaaaaaatatgtaagctgttttagaaacaaaaaagc	120	acgaaacacgatgtaagctc atttacgaaacgtacgaaaaa taaactttgtaccaataacaaa agtaca	69
	acgaaacacgatgtaagctcatttacgatgtaagctgtaaa aataaactttacgaaacacgatgtaagctcaaataacaaaa aatatgtaagctgttttagaaacaaaaaagc	113	acgaaacacgatgtaagctc atttgtacgaaaaataaacttt gtaccaataacaaaagtaca	62

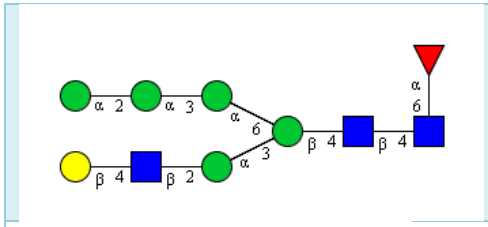
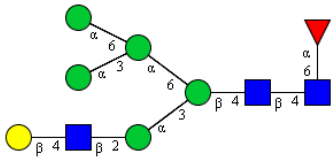
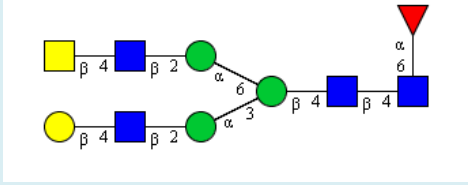
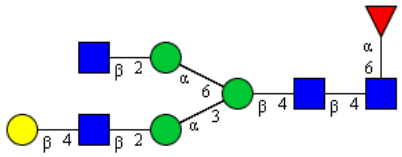
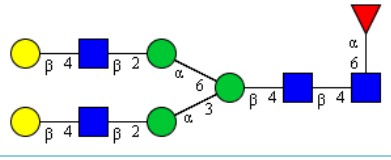
	acgaaacacgatgtaagctcaaataaactttacgaaacac gatgtaagctcaaataacaaaaaatatgtaagctgttttaga aacaaaaaagc	93	gtaccaataaactttacgaaa cacgatgtaagctcaaataac aaaagtaca	51
	acgaaacacgatgtaagctcaaataaactttacgatgtaag ctcaaataacaaaaaatatgtaagctgttttagaaacaaaa aagc	86	gtaccaataaactttacgatgt aagctcaaataacaaaagta ca	44
	acgaaacacgatgtaagctcaaataaactttaagctcaaat aacaaaaaatatgtaagctgttttagaaacaaaaaagc	79	gtaccaataaactt taagctcaaataacaaaagta ca	37
	acgaaacacgatgtaagctcaaataaactttaataacaaaa aatatgtaagctgttttagaaacaaaaaagc	72	gtaccaataaactttaataaca aaagtaca	30
	acgaaacacgatgtaagctcttttacgaaacacgatgtaa gcttaaaaacgatgtaagctcaaataaactttacgaaacac gatgtaagctcaaataacaaaaaatatgtaagctgttttaga aacaaaaaagc	134	gttattttacgaaac acgatgtaagcttaaaaacg atgtaagctcaaataaactttg taccaataacaaaagtaca	76

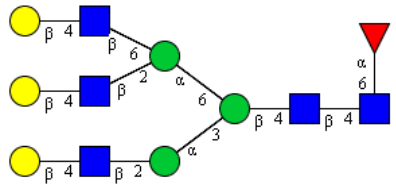
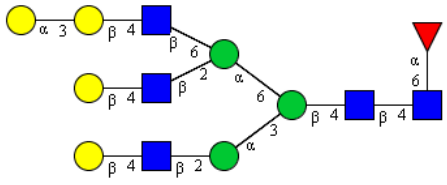
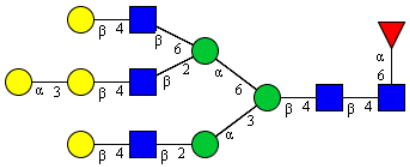
	acgatgtaagctcatttacgatctaagctgtaaaaataaact ttacgatgtaagctcatttacgatgtaagcttaaaaaataac aaaaaatatgtaagctgttttagaaacaaaaaagc	119	acgatgtaagctc atttgttcaaaaaataaacttta cgatgtaagctcatttacgat gtaagcttaaaaaataacaaa	84
	acgatgtaagctcatttacgatgttttagaaaacaaaaagct gtaaaaataaactttacgatgtaagctcaaataacaaaaaa tatgtaagctgttttagaaacaaaaaagc	112	acgatgtaagctcatttgtga gaaaaataaactttacgatgt aagctcaaataacaaaagta ca	64
	acgatgtaagctcatttacgatgtaagctgtaaaaataaact ttacgaaacacgatgtaagctcaaataacaaaaaatatgta agctgttttagaaacaaaaaagc	106	acgatgtaagctcatttgtac gaaaaataaactttgtaccaa taacaaaagtaca	55
	acgatgtaagctcatttacgatgtaagctgtaaaaataaact ttacgatgtaagctcatttacgatgtaagcttaaaaaataac aaaaaatatgtaagctgttttagaaacaaaaaagc	119	acgatgtaagctcatttgtac gaaaaataaactttacgatgt aagctcatttacgatgtaagct taaaaaataacaaaagtaca	84

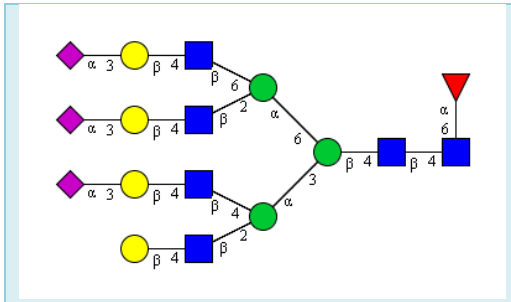
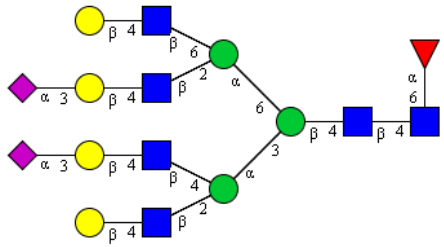
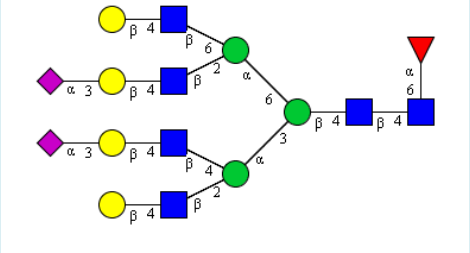
	acgatgttttagaaaacaaaaagctcaaataaactttaataa actttaataacaaaaaataacaaaaaatatgtaagctgtttta gaaacaaaaaagc	98	acgatgttttagaaaacaaaa agctcaaataaactttaataaa ctttaataacaaaaaataaca	72
	acgatgttttagaaaacaaaaagctcaaataaactttacgat gtaagctcaaataacaaaaaatatgtaagctgttttagaaa caaaaaagc	92	acgatgttttagaaaacaaaa agctcaaataaactttacgat gtaagctcaaataacaaaagt aca	66
	acgatgttttagaaaacaaaaagctcaaataaactttaataa caaaaaatatgtaagctgttttagaaacaaaaaagc	78	acgatgttttagaaaacaaaa agctcaaataaactttaataac aaaagtaca	52
	acgatctaagctgtaagcagttttaagcagtaaaaataaac tttaataagtaataaacaaaaatatgtaagctgttttagaaac aaaaaagc	92	gttcaaagcagttttaagcag taaaaataaactttaataagta ataaacaaagtaca	57

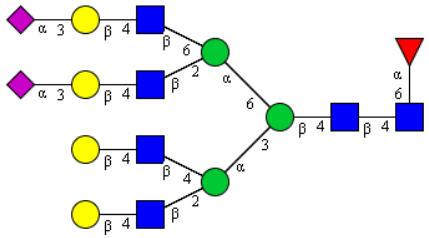
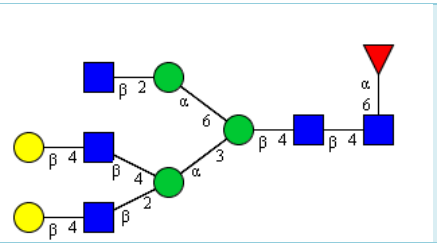
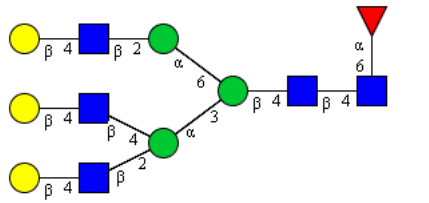
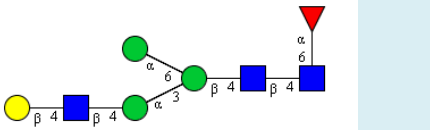
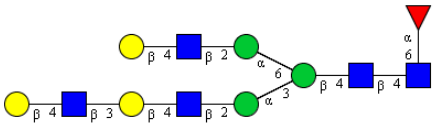
	acgatgttttagaaaacaaaaagctcatttacgatgtaagct gtaaaaataaactttacgatgtaagctcaaataacaaaaaa tatgtaagctgttttagaaacaaaaaagc	112	acgatgttttagaaaacaaaa agctcatttgtacgaaaaata aactttacgatgtaagctcaa ataacaaaagtaca	77
	acgatgttttagaaaacaaaaagctcatttacgatgttttag aaaacaaaaagctgtaaaaataaactttacgatgttttagaa aacaaaaagctcaaataacaaaaaatatgtaagctgtttta gaaacaaaaaagc	138	acgatgttttagaaaacaaaa agctcatttgtgagaaaaata aactttacgatgttttagaaaa caaaaagctcaaataacaaa agtaca	90
	acgatgttttagaaaacaaaaagctcaaataaactttacgat gttttagaaaacaaaaagctcaaataacaaaaaatatgtaa gctgttttagaaacaaaaaagc	105	acgatgttttagaaaacaaaa agctcaaataaactttacgat gttttagaaaacaaaaagctc aaataacaaaagtaca	79

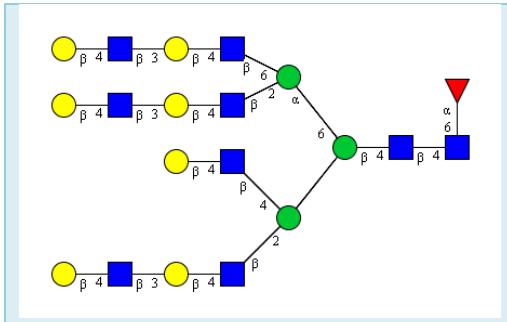
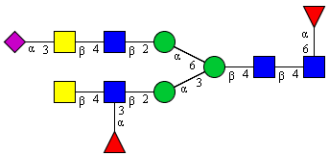
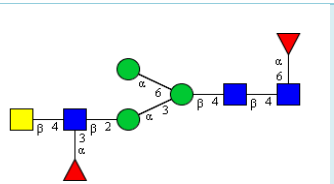
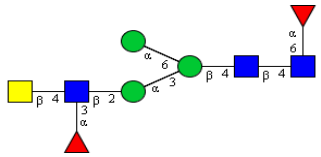
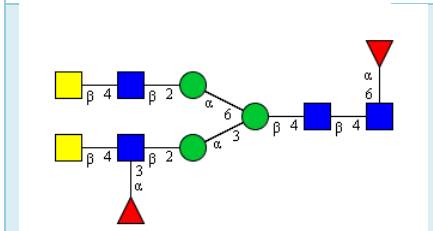
	acgatgtaagctcaaataaactttacgatgtaagctcattta cgctgtaagcttaaaaaataacaaaaaatatgtaagctgttt tagaaacaaaaaagc	99	acgatgtaagctcaaataaac tttacgatgtaagctcatttgtg tcaaaaataacaaaagtaca	64
	acgatgtaagctcaaataacatttacgatgtaagctcattta cgctgtaagcttaaaaaataaacaaaaaatatgtaagctgttt tagaaacaaaaaagc	99	acgatgtaagctcaaataaca tttacgatgtaagctcatttgtg tcaaaaataaacaaagtaca	64
	acgatgtaagctcaaataaactttacgctgtaagctcattta cgatgtaagcttaaaaaataacaaaaaatatgtaagctgttt tagaaacaaaaaagc	99	acgatgtaagctcaaataaac tttgtttctttacgatgtaagctt aaaaaataacaaaagtaca	64
	acgatgtaagctcaaataaactttaggggttacgatgtaag ctcaaataacaaaaaatatgtaagctgttttagaaacaaaa aagc	86	acgatgtaagctcaaataaac tttaggggttacgatgtaagct caaataacaaaagtaca	60

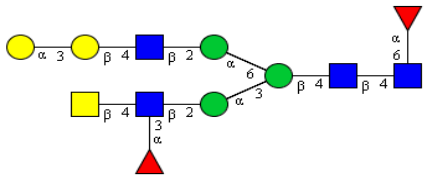
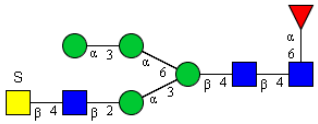
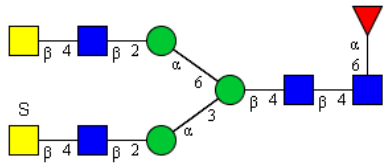
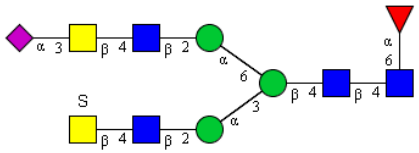
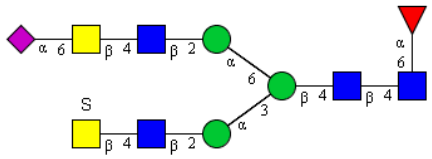
	acgatgtaagctcaaataaactttaataaagaataaacaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	79	acgatgtaagctcaaataaacttaataaagaataaacaataacaaagtaca	53
	acgatgtaagctcaaataaactttaataaactttaataacaaaaaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	85	acgatgtaagctcaaataaacttaataaactttaataacaaaaataacaaaagtaca	59
	acgatgtaagctcaaataaactttacgctgtaagctcaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	79	acgatgtaagctcaaataaacttttgtttcaataacaaaagtaca	44
	acgatgtaagctcaaataaactttaagctcaaataacaaaaatatgtaagctgttttagaaacaaaaaagc	72	acgatgtaagctcaaataaacttttaagctcaaataacaaaagtaca	46
	acgatgtaagctcaaataaactttacgatgtaagctcaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	79	acgatgtaagctcaaataaacttttacgatgtaagctcaaataacaaagtaca	53

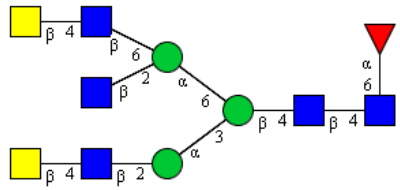
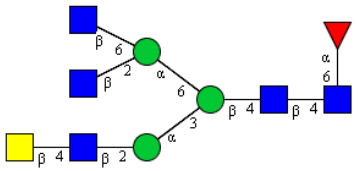
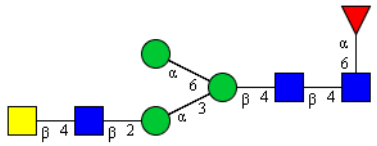
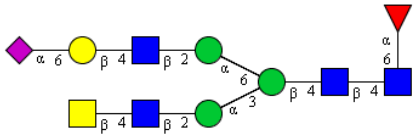
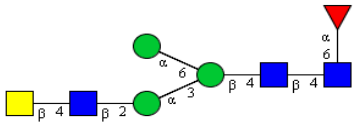
	acgatgtaagctcaaataaactttacgatgtaagctcattta cgatgtaagcttaaaaaataacaaaaaatatgtaagctgttt tagaaacaaaaaagc	99	acgatgtaagctcaaataaac tttacgatgtaagctcatttgtg cgaaaaataacaaaagtaca	64
	acgatgtaagctcaaataaactttacgatgtaagctcattta cgaaacacgatgtaagcttaaaaaataacaaaaaatatgt aagctgttttagaaacaaaaaagc	106	acgatgtaagctcaaataaac tttacgatgtaagctcatttac gaaacgtgcgaaaaataaca aaagtaca	71
	acgatgtaagctcaaataaactttacgaaacacgatgtaag ctcatttacgatgtaagcttaaaaaataacaaaaaatatgta agctgttttagaaacaaaaaagc	106	acgatgtaagctcaaataaac tttacgaaacacgatgtaagc tcatttgtgcgaaaaataaca aaagtaca	71

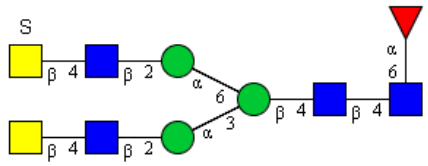
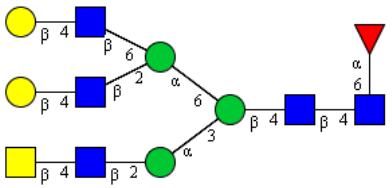
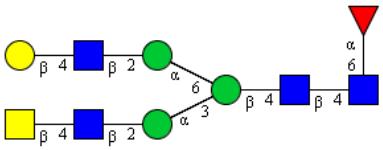
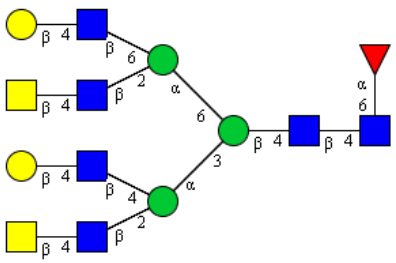
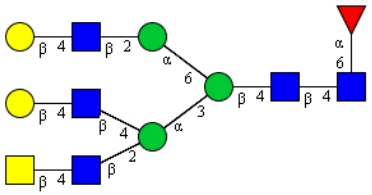
	acgatgtaagctcatttaggcggtacgatgtaagctgtaaa aataaactttaggcggtacgatgtaagctcatttaggcggt acgatgtaagcttaaaaaataacaaaaaatatgtaagctgt tttagaaacaaaaaagc	140	acgatgtaagctcatttaggc ggtacgatgtaagctgtaaaa ataaactttaggcggtacgat gtaagctcatttgtggtaaaa ataacaaaagtaca	98
	acgatgtaagctcatttaggcggtacgatgtaagctgtaaa aataaactttaggcggtacgatgtaagctcatttacgatgta agcttaaaaaataacaaaaaatatgtaagctgttttagaaac aaaaaagc	133	acgatgtaagctcatttgtgg aaaaaataaactttaggcggt acgatgtaagctcatttgtgc gaaaaataacaaaagtaca	82
	acgatgtaagctcatttaggcggtacgatgtaagctgtaaa aataaactttaggcggtacgatgtaagctcatttacgatgta agcttaaaaaataacaaaaaatatgtaagctgttttagaaac aaaaaagc	133	acgatgtaagctcatttgtgg aaaaaataaactttaggcggt acgatgtaagctcatttgtgc gaaaaataacaaaagtaca	82

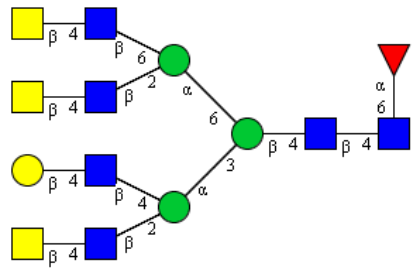
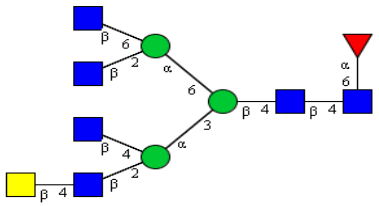
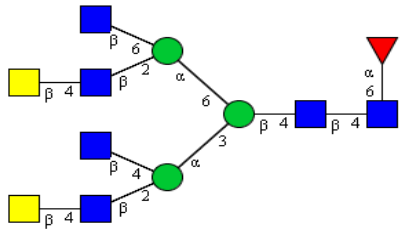
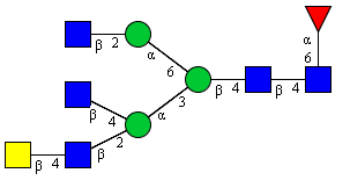
	acgatgtaagctcatttacgatgtaagctgtaaaaataaactttaggcggtacgatgtaagctcatttaggcggtacgatgtaagcttaaaaaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	133	acgatgtaagctcatttacgatgtaagctgtaaaaataaacttttaggcggtacgatgtaagctcatttgtggtaaaaataacaaaagtaca	91
	acgatgtaagctcatttacgatgtaagctgtaaaaataaactttaagctcaaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	92	acgatgtaagctcatttgtacgaaaaataaactttaagctcaaataacaaaagtaca	57
	acgatgtaagctcatttacgatgtaagctgtaaaaataaactttacgatgtaagctcaaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	99	acgatgtaagctcatttgtacgaaaaataaactttacgatgtaagctcaaataacaaaagta	64
	acgatgtaagctgtaataaactttaataacaaaaaatatgtagctgttttagaaacaaaaaagc	65	gtacgaataaactttaataacaaaagtaca	30
	acgatgtaagctctacgatgtaagctcaaataaactttacgatgtaagctcaaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	93	acgatgtaagctctacgatgtaagctcaaataaactttacgatgtaagctcaaataacaaaag	67

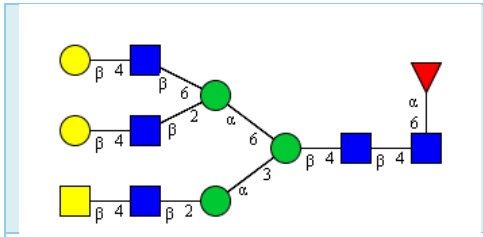
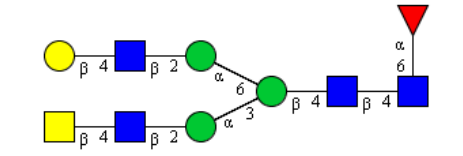
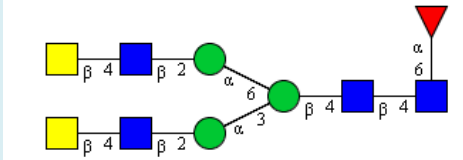
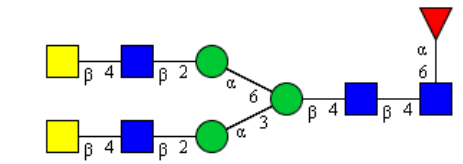
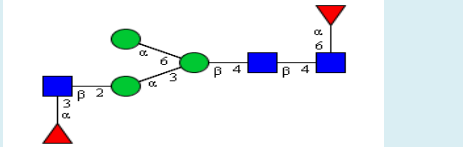
	acgatgtaagctctacgatgtaagctcatttacgatgtaag ctgtaaaaataaactttacgatgtaagctctacgatgtaagc tcatttacgatgtaagctctacgatgtaagcttaaaaaataa caaaaaatatgtaagctgttttagaaacaaaaaagc	161	acgatgtaagctctacgatgt aagctcatttacgatgtaagct gtaaaaataaactttacgatgt aagctctacgatgtaagctca tttacgatgtaagctctgtgcg	126
	acgctgttttagaaaacaaaaagctcaaataaactttaggc ggtacgctgtaagctcaaataacaaaaaatatgtaagctgt tttagaaacaaaaaagc	99	acgctgttttagaaaacaaaa agctcaaataaactttaggcg gtgtttcaataacaaaagtac a	64
	acgctgttttagaaaacaaaaagctcaaataaactttaataa caaaaaatatgtaagctgttttagaaacaaaaaagc	78	acgctgttttagaaaacaaaa agctcaaataaactttaataac aaaagtaca	52
	acgctgttttagaaaacaaaaagctcaaataaactttaataa caaaaaatatgtaagctgttttagaaacaaaaaagc	78	acgctgttttagaaaacaaaa agctcaaataaactttaataac aaaagtaca	52
	acgctgttttagaaaacaaaaagctcaaataaactttacgct gtaagctcaaataacaaaaaatatgtaagctgttttagaaa caaaaaagc	92	acgctgttttagaaaacaaaa agctcaaataaactttgtttca ataacaaaagtaca	57

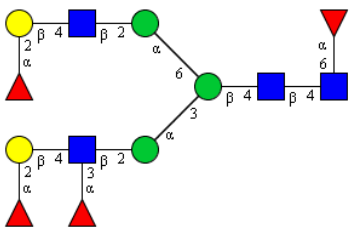
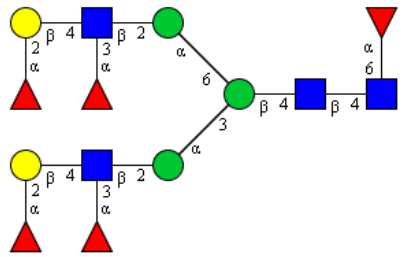
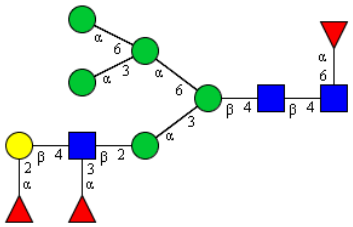
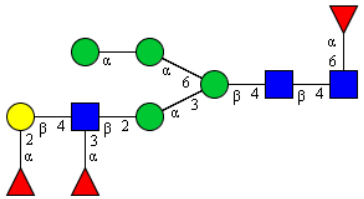
	acgctgttttagaaaacaaaaagctcaaataaactttacgaaacacgatgtaagctcaaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	99	acgctgttttagaaaacaaaaagctcaaataaactttgtaccaataacaaaagtaca	57
	acgccatttgtaagctcaaataaactttaataaacaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	76	acgccatttgtaagctcaaataaactttaataaacaataacaaaagtaca	50
	acgccatttgtaagctcaaataaactttacgctgtaagctcaaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	83	acgccatttgtaagctcaaataaactttgtttcaataacaaaagtaca	48
	acgccatttgtaagctcaaataaactttaggcggtacgatgtaagctcaaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	90	acgccatttgtaagctcaaataaactttaggcggtacgatgtaagctcaaataacaaaagta	64
	acgccatttgtaagctcaaataaactttaggcgttacgatgtaagctcaaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	90	acgccatttgtaagctcaaataaactttaggcgttacgatgtaagctcaaataacaa	64

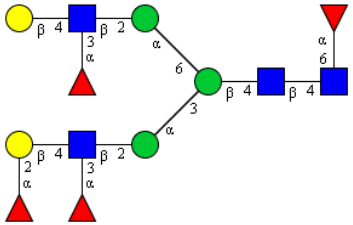
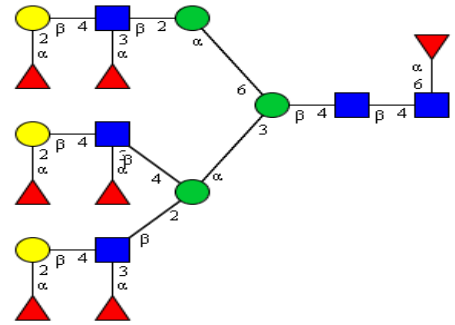
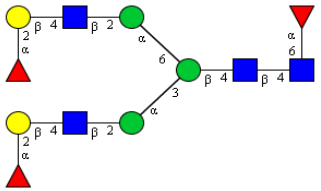
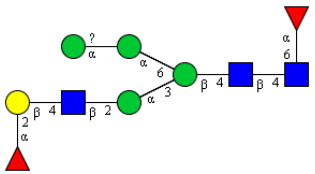
	acgctgtaagctcaaataaactttaagctcatttacgctgtaagcttaaaaaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	92	acgctgtaagctcaaataaacttttaagctcatttgtgtcaaaataacaaaagtaca	57
	acgctgtaagctcaaataaactttaagctcatttaagcttaaaaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	85	gtttcaataaactttaagctcatttaagcttaaaaaataacaaaagtaca	50
	acgctgtaagctcaaataaactttaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	65	gtttcaataaactttaataacaaaagtaca	30
	acgctgtaagctcaaataaactttaggcgttacgatgtaagctcaaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	86	gtttcaataaactttaggcgttacgatgtaagctcaaataacaaaagtaca	51
	acgctgtaagctcaaataaactttaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	65	gtttcaataaactttaataacaaaagtaca	30

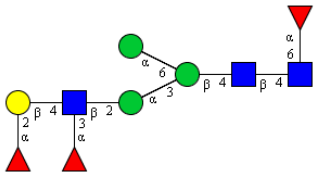
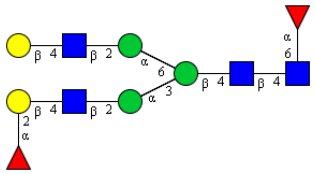
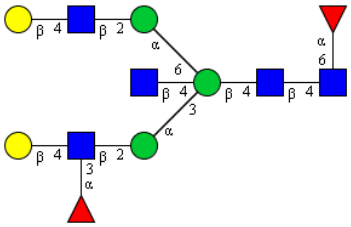
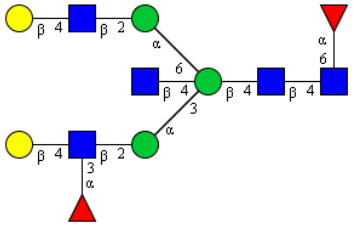
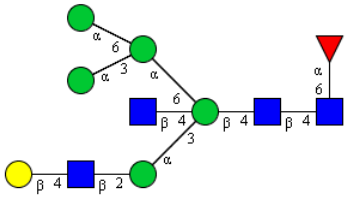
	acgctgtaagctcaaataaactttacgccatttgtaagctcaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	83	gtttcaataaactttacgccatttgtaagctcaaataacaaaagtaca	48
	acgctgtaagctcaaataaactttacgatgtaagctcatttacgatgtaagcttaaaaaataacaaaaaatatgtaagctgtttagaaacaaaaaagc	99	gtttcaataaactttacgatgtaagctcatttgtgcgaaaaataacaaaagtaca	55
	acgctgtaagctcaaataaactttacgatgtaagctcaaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	79	gtttcaataaactttacgatgtaagctcaaataacaaaagtaca	44
	acgctgtaagctcatttacgatgtaagctgtaaaaataaacttacgctgtaagctcatttacgatgtaagcttaaaaaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	119	gtttctttacgatgtaagctgtaaaaataaactttacgctgtaagctcatttgtgcgaaaaataacaaaagtaca	75
	acgctgtaagctcatttacgatgtaagctgtaaaaataaacttacgatgtaagctcaaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	99	gtttctttgtcctaaaaataaacttacgatgtaagctcaaataacaaaagtaca	55

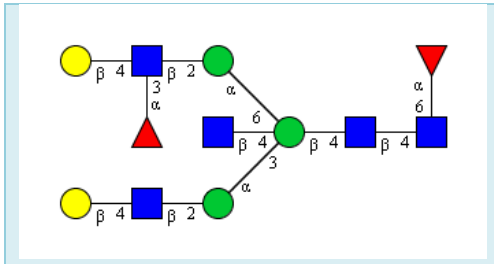
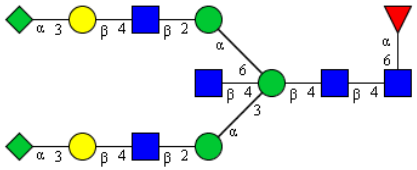
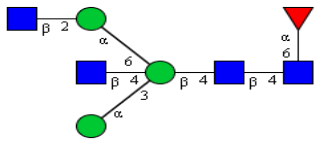
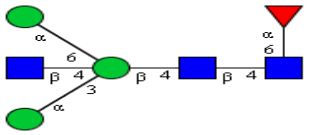
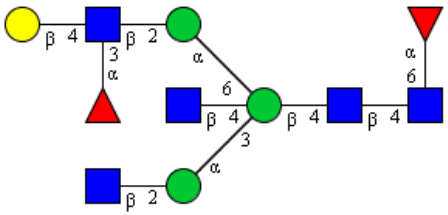
	acgctgtaagctcatttacgatgtaagctgtaaaaataaact ttacgctgtaagctcatttacgctgtaagcttaaaaaataac aaaaaatatgtaagctgttttagaaacaaaaaagc	119	acgctgtaagctcatttacgat gtaagctgtaaaaataaacttt acgctgtaagctcatttgtgtc aaaaataacaaaagtaca	84
	acgctgtaagctcatttaagctgtaaaaataaactttaagct catttaagcttaaaaaataacaaaaaatatgtaagctgtttta gaaacaaaaaagc	98	gttgaaataaactttaagctca tttaagcttaaaaaataacaaa agtaca	50
	acgctgtaagctcatttaagctgtaaaaataaactttacgct gtaagctcatttaagcttaaaaaataacaaaaaatatgtaag ctgttttagaaacaaaaaagc	105	gtttctttaagctgtaaaaata aactttacgctgtaagctcatt taagcttaaaaaataacaaaa gtaca	70
	acgctgtaagctcatttaagctgtaaaaataaactttaagct caaataacaaaaaatatgtaagctgttttagaaacaaaaaa gc	85	gtttctttaagctgtaaaaata aactttaagctcaaataacaa aagtaca	50

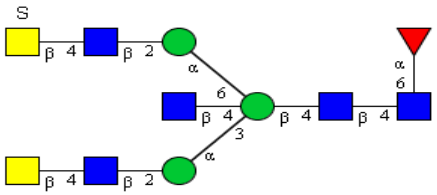
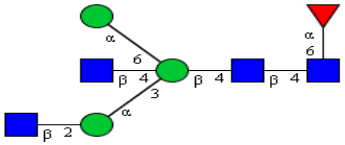
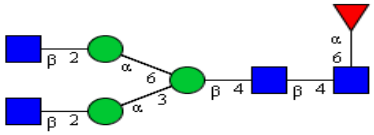
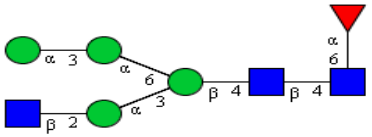
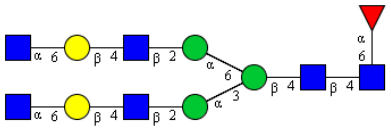
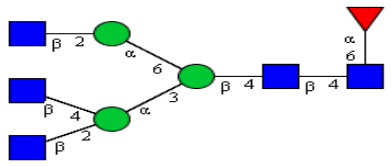
	acgctgtaagctcaaataaactttacgatgtaagctcattta cgatgtaagcttaaaaaataacaaaaaatatgtaagctgttt tagaaacaaaaaagc	99	gtttcaataaactttacgatgt aagctcatttgtgcgaaaaat aacaaaagtaca	55
	acgctgtaagctcaaataaactttacgatgtaagctcaaat aacaaaaaatatgtaagctgttttagaaacaaaaaagc	79	gtttcaataaactttacgatgt aagctcaaataacaaaagta ca	44
	acgctgtaagctcaaataaactttacgctgtaagctcaaat aacaaaaaatatgtaagctgttttagaaacaaaaaagc	79	gtttcaataaactttacgctgt aagctcaaataacaaaagta ca	44
	acgctgtaagctcaaataaactttacgctgtaagctcaaat aacaaaaaatatgtaagctgttttagaaacaaaaaagc	79	gtttcaataaactttacgctgt aagctcaaataacaaaagta ca	44
	agaaaacaagctcaaataaactttaataacaaaaaatatgt aagctgttttagaaacaaaaaagc	65	gttgtaataaactttaataaca aaagtaca	30

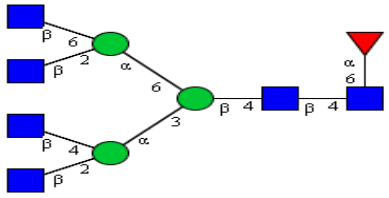
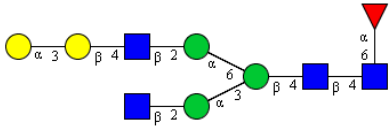
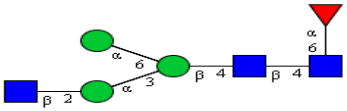
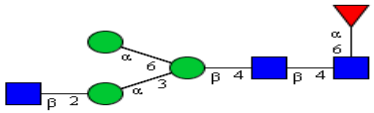
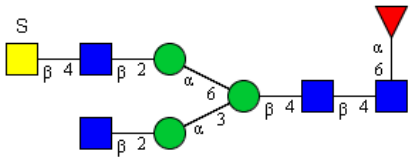
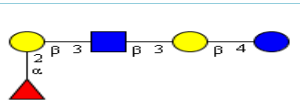
	agaaaagacgatgttttagaaaacaaaaagctcaaataaactttagaaaagacgatgtaagctcaaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	106	gttggaataaactttgttgcaataacaaaagtaca	35
	agaaaagacgatgttttagaaaacaaaaagctcaaataactttagaaaagacgatgttttagaaaacaaaaagctcaaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	119	gttggaataaactttagaaaagacgatgttttagaaaacaaaaagctcaaataacaaaagtaca	64
	agaaaagacgatgttttagaaaacaaaaagctcaaataactttaataaactttaataacaaaaaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	105	gttggaataaactttaataaacttttaataacaaaaaataacaaaagtaca	50
	agaaaagacgatgttttagaaaacaaaaagctcaaataactttaatagccaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	92	gttggaataaactttaatagccaataacaaaagtaca	37

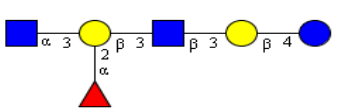
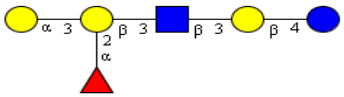
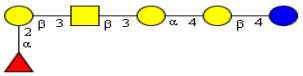
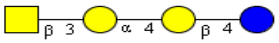
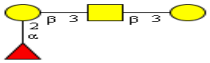
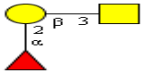
	agaaaagacgatgttttagaaaacaaaaagctcaaataaactttacgatgttttagaaaacaaaaagctcaaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	112	gttggaataaactttacgatgttttagaaaacaaaaagctcaaataacaaaagtaca	57
	agaaaagacgatgttttagaaaacaaaaagctcatttagaaaagacgatgttttagaaaacaaaaagctgtaaaaataaacttttagaaaagacgatgttttagaaaacaaaaagctcaaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	159	gttggtttagaaaaggtgagaaaaataaactttagaaaagacgatgttttagaaaacaaaaagctcaaataacaaaagtaca	82
	agaaaagacgatgtaagctcaaataaactttagaaaagacgatgtaagctcaaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	93	gttgcaataaactttagaaaagacgatgtaagctcaaataacaaaagtaca	51
	agaaaagacgatgtaagctcaaataaactttaatagccaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	79	gttgcaataaactttaatagccaataacaaaagtaca	37

	agaaaagacgatgttttagaaaacaaaaagctcaaataaactttaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	85	gttggaataaactttaataacaaaagtaca	30
	agaaaagacgatgtaagctcaaataaactttacgatgtaagctcaaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	86	gttgcaataaactttacgatgtaagctcaaataacaaaagtaca	44
	aagctgttttacgatgttttagaaaacaaaaagctcaaataaacaaatttacgatgtaagctcaaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	105	aagctgttttacgatgttttagaaaacaaaaagctcaaataacaaatttacgatgtaagctcaataacaaaagtaca	79
	aagctgttttacgatgttttagaaaacaaaaagctcaaataacaaatttacgatgtaagctcaaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	105	aagctgttttacgatgttttagaaaacaaaaagctcaaataacaaatttacgatgtaagctcaataacaaaagtaca	79
	aagctgttttacgatgtaagctcaaataaacaaatttaataaactttaataacaaaaaataacaaaaaatatgtaagctgtttagaaacaaaaaagc	98	aagctgttttacgatgtaagctcaaataaacaaatttaataactttaataacaaaaaataacaaaagtaca	72

	aagctgttttacgatgtaagctcaaataaacaaatttacgatgttttagaaaacaaaaagctcaaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	105	aagctgttttacgatgtaagctcaaataaacaaatttacgatgttttagaaaacaaaaagctcaaataacaaaagtaca	79
	aagctgttttaggtaacacgatgtaagctcaaataaacaaatttaggtaacacgatgtaagctcaaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	106	aagctgttttaggtaacacgatgtaagctcaaataaacaaattaggtaacacgatgtaagctcaaataacaaaagtaca	80
	aagctgttttaataaacaaatttaagctcaaataacaaaaatatgtaagctgttttagaaacaaaaaagc	71	aagctgttttaataaacaaatttaagctcaaataacaaaagtaca	45
	aagctgttttaataaacaaatttaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	64	aagctgttttaataaacaaatttaataacaaaagtaca	38
	aagctgttttaagctcaaataaacaaatttacgatgttttagaaaacaaaaagctcaaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	98	aagctgttttaagctcaaataacaaaatttacgatgttttagaaacaaaaagctcaaataacaaaagtaca	72

	aagctgttttacgctgtaagctcaaataaacaaatttacgccatttgtaagctcaaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	96	aagctgttttgtttcaataaacaaatttacgccatttgtaagctcaaataacaaaagtaca	61
	aagctgttttaagctcaaataaacaaatttaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	71	aagctgttttaagctcaaataaacaaatttaataacaaaagtaca	45
	aagctcaaataaactttaagctcaaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	65	aagctcaaataaactttaagctcaaataacaaaagtaca	39
	aagctcaaataaactttaataaacaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	65	aagctcaaataaactttaataaacaataacaaaagtaca	39
	aagcacaacgatgtaagctcaaataaactttaagcacaacgatgtaagctcaaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	93	aagcacaacgatgtaagctcaaataaactttaagcacaacgatgtaagctcaaataacaaaa	67
	aagctcatttaagctgtaaaaataaactttaagctcaaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	78	aagctcatttaagctgtaaaaataaactttaagctcaaataacaaaagtaca	52

	aagctcatttaagctgtaaaaataaactttaagctcatttaagcttaaaaaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	91	aagctcatttaagctgtaaaaataaactttaagctcatttaagcttaaaaaataacaaaagtac	65
	aagctcaaataaactttacgaaacacgatgtaagctcaaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	79	aagctcaaataaactttgtaccaataacaaaagtaca	37
	aagctcaaataaactttaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	58	aagctcaaataaactttaataacaaaagtaca	32
	aagctcaaataaactttaataacaaaaaatatgtaagctgttttagaaacaaaaaagc	58	aagctcaaataaactttaataacaaaagtaca	32
	aagctcaaataaactttacgccatttgtaagctcaaataacaaaaatatgtaagctgttttagaaacaaaaaagc	76	aagctcaaataaactttacgccatttgtaagctcaaataacaaaagtaca	50
	agaaaagacgatctaagctctacgatgtaaga	32		

	acgcaactttagaaaagaaaacgatctaagctctacgatgtaaga	45		
	acgaaactttagaaaagaaaacgatctaagctctacgatgtaaga	45		
	agaaaagacgatctacgctctacgaataacgatgtaaga	39		
	acgatctacgctctacgaataacgatgtaaga	32		
	acgctctacgaataacgatgtaaga	25		
	acgaataacgatgtaagaccccagtcaggcctaacgta	38		
	acgatgtaagaccccagtcaggcctaacgta	31		
	agaaaagacgatctacgctctacga	25		
	agaaaagacgatctacgcccccagtcaggcctaacgta	38		
	agaaaagacgaccccagtcaggcctaacgta	31		