

8-8-2017

DNA Encoded Libraries (DEGL) of Glycan Antigens to Detect Antibodies: An Approach Towards Next Generation Functional Glycomics

Aishwarya Parameswaran
aparameswaran1@student.gsu.edu

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

Recommended Citation

Parameswaran, Aishwarya, "DNA Encoded Libraries (DEGL) of Glycan Antigens to Detect Antibodies: An Approach Towards Next Generation Functional Glycomics." Thesis, Georgia State University, 2017.
https://scholarworks.gsu.edu/chemistry_theses/101

This Thesis is brought to you for free and open access by the Department of Chemistry at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Chemistry Theses by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

DNA ENCODED LIBRARIES (DEGL) OF GLYCAN ANTIGENS TO DETECT
ANTIBODIES: AN APPROACH TOWARDS NEXT GENERATION FUNCTIONAL
GLYCOMICS

by

AISHWARYA PARAMESWARAN

Under the Direction of Peng George Wang, PhD

ABSTRACT

Structure and functional study of glycans are highly challenging due to the difficulties in analyzing glycans and limited availability of samples for study. These limitations could be resolved by attaching DNA barcode to the glycan, which virtually represent glycan in further application, by increasing the sensitivity of detection by polymerase chain reaction (PCR), requiring minimal samples for analysis. Assuming bigger arena of DNA Encoded Glycan Libraries (DEGL) in future, we propose here a method for uniquely coding all glycans using computer program that can convert the structural information of glycans to DNA barcode. A unique and universal coding for glycans will benefit both synthesis and analysis of DEGLs. As a proof of principle study, a small DNA Encoded Glycan Library (DEGL) of blood and globo series glycan antigen and its application was demonstrated in detecting blood group and breast cancer from plasma.

INDEX WORDS: DNA Encoded Glycan Library (DEGL), Glyco-PCR, Inverse Blood typing,
Blood antigens, Globo Series Glycans, Click Chemistry

DNA ENCODED LIBRARIES (DEGL) OF GLYCAN ANTIGENS TO DETECT
ANTIBODIES: AN APPROACH TOWARDS NEXT GENERATION FUNCTIONAL
GLYCOMICS

by

AISHWARYA PARAMESWARAN

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

Master of Science

in the College of Arts and Sciences

Georgia State University

2017

Copyright by
Aishwarya Parameswaran
2017

DNA ENCODED LIBRARIES (DEGL) OF GLYCAN ANTIGENS TO DETECT
ANTIBODIES: AN APPROACH TOWARDS NEXT GENERATION FUNCTIONAL
GLYCOMICS

by

AISHWARYA PARAMESWARAN

Committee Chair: Peng George Wang

Committee: Suri Saranathan. Iyer

Gregory Poon

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

August 2017

DEDICATION

I would like to thank to my family, for their intense support throughout the journey of my life so far. To my parents for their love, suggestions, and continuous support as always. They give me great comfort and encouragement whenever I'm depressed.

To my husband for his love, suggestions, and financial support throughout this journey. I would like to thank him a lot giving me immense encouragement and for bringing me back to normal, whenever I'm depressed and tensed.

To all other members of my family for their continued support, encouragement and love.

To all my friends and classmates, especially to Sunitha, who helped me lot to study and accompany during my setbacks.

To Arya, who helped me acquire knowledge and information, to enhance my skills and advance towards and set goals for my career.

Finally, I would like to thank to all the people whom I love and those who loves me

ACKNOWLEDGEMENTS

I would like to express my intense and special thanks to Dr. Peng George Wang, who gave me such a wonderful opportunity to study and do research in his big and energetic group,

which gave me a great opportunity to learn so many new things. His enthusiasm for academics, positive attitude towards everything inspires me a lot, which motivates and inspires me a lot for my future.

I sincerely thank Dr. Suri S. Iyer and Dr. Gregory Poon for their valuable suggestions and being a part of my committee.

I would like to extend my thanks to Shukkoor M. Kondengaden and Dr. Liuqing Wen, who helped me a lot in finishing my project on time during my master program.

I would also like to take this opportunity to thank all my lab-mates of Dr. Peng George Wang's group for their full support and advices.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	V
LIST OF TABLES	VIII
LIST OF FIGURES	IX
LIST OF ABBREVIATIONS	XI
1 INTRODUCTION	1
1.1 Purpose of the Study.....	4
<i>1.1.1 Click Chemistry</i>	<i>5</i>
<i>1.1.2 PCR.....</i>	<i>7</i>
<i>1.1.3 ABO Blood Group System</i>	<i>11</i>
<i>1.1.4 Globo-H</i>	<i>14</i>
2 EXPERIMENT	16
2.1 Systematic Coding of Sugars.....	16
2.2 Coding and Decoding of glycans through Program.....	19
<i>2.2.1 Coding.....</i>	<i>19</i>
<i>2.2.2 Decoding.....</i>	<i>19</i>
2.3 Materials	22
2.4 Gradient PCR and T_m Determination.....	22
2.5 Synthesis of Glycan(Antigen) – DNA Conjugates	23
<i>2.5.1 Click reaction procedure.....</i>	<i>23</i>

2.6	General protocol for qPCR	24
2.6.1	<i>For a typical 12.5 μL reaction</i>	24
2.7	Filtration of Antibody-Antigen DNA Conjugates	25
2.8	Immunoprecipitation of Antibody-Antigen-DNA Conjugates.....	25
2.9	Blood Collection.....	26
3	RESULTS AND DISSCUSSION.....	27
3.1	DNA sequences	27
3.2	T _m Determination	28
3.3	Glycan (Antigen) – DNA Conjugates	28
3.4	qPCR Reactions and Standard Curve.....	30
3.5	Blood Glycan – Antibody Sensitivity study: Filtration.....	32
3.6	Blood Glycan – Antibody Interaction Comparison: Filtration and Immunoprecipitation	36
3.7	DEGL: Real Life experiments – Inverse Blood Typing.....	38
3.8	Globo Series Library.....	40
3.9	Detection of glycan binding antibodies present in the cancer plasma.....	42
4	CONCLUSION.....	45
	REFERENCES.....	47

LIST OF TABLES

Table 1.1 Possible Blood types in offsprings.....	12
Table 1.2 Antibodies and Antigen present in ABO Blood Typing System.....	12
Table 1.3 Distribution of various Blood Groups among Population	14
Table 2.1 Glycan Representation (CFG)	16
Table 2.2 Glycan Representation with DNA Codes	18
Table 2.3 Long and Short DNA codes of the Glycan.....	18
Table 2.4 DNA code for short glycans	19
Table 2.5 Reaction mix for gradient PCR.....	23
Table 2.6 Reaction mix for qPCR.....	25
Table 3.1 DNA sequences.....	27
Table 3.2 Blood Group Antigens and Antibodies.....	39

LIST OF FIGURES

Figure 1.1 Staudinger Ligation	6
Figure 1.2 Copper-Catalyzed Click Chemistry (CuAAC)	7
Figure 1.3 Key Components Of PCR	8
Figure 1.4 Phases of qPCR.	10
Figure 1.5 ABO Blood antigen structures.....	13
Figure 1.6 Globo-H Structure	15
Figure 2.1 Flowchart describing Coding	20
Figure 2.2 Flowchart describing decoding.....	21
Figure 2.3 CFG representation and DNA code for blood antigen A	22
Figure 2.4 Click Conjugation of 5'-Hexynyl DNA and azido modified glycans	24
Figure 3.1 T _m Determination	28
Figure 3.2 PCR Comparison of DNA _A and G+DNA _A	29
Figure 3.3 DNA and Glycan- DNA Conjugate IP Analysis	30
Figure 3.4 qPCR limit of detection comparison of glycan DNA conjugate and pure DNA.....	31
Figure 3.5 Standard Curve Plot.....	32
Figure 3.6 Blood Group ABO antigen- DNA conjugate structures.....	33
Figure 3.7 Comparison of Filtration	35
Figure 3.8 Filtration study with G+DNA conjugates and Antibody.....	36
Figure 3.9 Ct value comparison	37
Figure 3.10 Amplification plot comparison.....	38
Figure 3.11 Amplification plot and Ct Value comparison of Blood Typing	40
Figure 3.12 Globo glycan structures and Globo-glycan conjugates	41

Figure 3.13 Amplification plot and Ct value Globo-H Vs VK-9 A) Amplification Plot and B) Ct Value plot.....	42
Figure 3.14 Ct Value Plot of different Globo glycan conjugate series with breast cancer serum (BC1 and BC2), Healthy Serum (HS) and No Serum (NS).....	43
Figure 3.15 Amplification Plot of different Globo glycan conjugate series with breast cancer serum (BC1 and BC2), Healthy serum (HS) and No serum (NS)	44

LIST OF ABBREVIATIONS

GlcNAc	N-acetylglucosamine
Glc	Glucose
Gal	Galactose
Man	Mannose
Neu5Ac	N-acetylneuraminic acid
Fuc	Fucose
GalNAc	N-acetylgalactosamine
Xyl	Xylose
GlcA	Glucuronic acid
Ido	Idose
DEGL	DNA Encoded Glycan Library
DNA	Deoxyribonucleic Acid
RNA	Ribonucleic Acid
GBP	Glycan Binding Proteins
PCR	Polymerase Chain Reaction
qPCR	quantitative Polymerase Chain Reaction
RT-PCR	Real Time Polymerase Chain Reaction
GAG	Glycosaminoglycans
PG	Proteoglycans
CHO	Chinese Hamster Ovary
GP	Glycoprotein
NGS	Next Generation Sequencing
EDC	1-Ethyl-3-(3-dimethylaminopropyl)-carbodiimide
NHS	N-hydroxysuccinimide
BoNT	Botulinum neurotoxin
CuAAC	Copper-catalyzed azide-alkyne Cycloaddition
SPAAC	Strain promoted alkyne-azide cycloaddition
A	Adenine

T	Thymine
C	Cytosine
G	Guanine
C _q	Quantification cycle
C _t	Threshold cycle
RBC	Red Blood Cells
a	alpha
b	beta
IUPAC	International Union of Pure and Applied Chemistry
μg	Microgram
μL	Microliter
μM	Micromolar
nM	Nano molar
pM	Picomolar
mM	Millimolar
T _m	Melting Temperature
°C	Degree Celsius
TBST	Tris Buffered Saline- Tween 20
kDa	kilodalton
RPM	Revolution Per Minute
G+DNA	Glycan + DNA
NTC	No Template Control
Ab	Antibody
Ag	Antigen
s	seconds
G	Glycan
HS	Healthy Serum
NS	No Serum
BC	Breast Cancer

1 INTRODUCTION

Glycans are important biomolecules along with proteins, lipids, and nucleic acids. Glycans can be divided into three main categories: O-linked glycans, N-linked glycans and glycolipids.² Majority of glycans are made up of N-acetylglucosamine (GlcNAc), Glucose (Glc), Galactose (Gal), Mannose (Man), N-acetylneuraminic acid (Neu5Ac), Fucose (Fuc), N-acetylgalactosamine (GalNAc), Xylose (Xyl), Glucuronic acid (GlcA), Idose (Ido).¹

Each cell in a multicellular organism is covered with dense and complex glycans. Most cellular secretions and extracellular matrices are found to be high in glycans and glycosylated. It is even seen that an envelope of a virus that bud from an infected cell to carry the glycosylation pattern of the host cell.³

Glycans and glycoconjugates play a very important role in many areas of biological processes including cell- cell interactions, pathogenicity, mediating immunological processes, stabilizing proteins and cancer.⁴⁻⁶ Expression patterns of glycans are highly altered in diseases like cancer, atherosclerosis, diabetes, retrovirus infection, thrombosis, arthritis, neurodegeneration and many other diseases.⁸ Glycan structures are the important part of the antigenic determinants during microbial infections to which the host immune cells recognizes and mount immune responses against the infection.⁹ For example, breast cancer can be detected by Globo-H glycan, an important biomarker seen in breast cancer.⁷ Therefore, glycans have a pivotal role in molecular recognition and signaling, making sugar-receptor interaction at the structural level highly necessary; and this interaction of polysaccharide with receptors are major area of focus in academic and industrial research.^{8,10}

Functional Glycomics; a systematic study of the structures and functions of carbohydrates is yet to reach the level attained by the genomics and proteomics despite so much importance of

glycans. Template driven biosynthesis of DNA-RNA-Protein by nature itself also known as the Central Dogma Theory of Molecular Biology made the researchers apply and adopt this knowledge in the synthesis and analysis of these molecules. Apart from this, the easy availability of experimentation methods and materials for the synthesis and analysis of DNA, RNA, and proteins made the researchers focus on genomics and proteomics and take up those areas to what they are today. However, such free-flowing information is not available for carbohydrate making functional glycomics exceptionally challenging. Since glycans are less organized without a template driven machinery, it is necessary to develop a specific method for synthesis and analysis for each glycan. Although with the recent advancement in technology allowed the development of solid phase automated and enzymatic synthesis, which reduced the difficulties in obtaining the glycans, but the materials and experimentations are still extremely difficult and complex especially for complex glycans. Therefore, there is a need for a development of versatile and ultra-sensitive detection technology as it is of high importance in the field of carbohydrate research for detection and characterization of glycan-protein interaction.

The methods now used for studying glycan-protein interaction are mainly carbohydrate microarrays,^{11,12} surface plasmon resonance,¹³ enzyme linked lectin assay,¹² isothermal calorimetry,¹⁴ glycosylated conductive polymer,¹⁵ and crystallographic studies.¹⁶ In carbohydrate research, widely used method is microarray with carbohydrate derivatives by various means of immobilization technology.^{11,17-18} However, all the above methods have its own limitations, mainly because these techniques uses immobilized glycans that do not mimic the biological environment of binding. The detection also requires tagging of the target, glycan modifications and specific binding of antibodies, in addition, the interaction of glycans with targets are also very

weak, which makes the generation sugar library having a glycodendrimer on a slide with optimum distance for specific binding highly challenging.

DNA-encoded glycan libraries (DEGL) can be considered as a potential alternative or adjunct to the prevailing glycan detecting techniques. DEGL adopt DNA sequences specifically for each glycan in the library, which enable sensitive detection by signal amplification PCR and identifying by DNA sequencing. For functional glycomics study, DEGL of interest is constructed and interacted with specific proteins. Unbound glycans are removed by washing, and bound ones are amplified using quantitative Polymerase Chain Reaction (qPCR), also known as real time PCR (RT-PCR) and sequenced to reveal identities.

Next generation technologies in genomics extended the applications of DNA beyond genetics, the most prominent one being DNA encoding. The concept of DNA encoding was introduced in 1992, but it gained momentum recently with the advancement of Next Generation Sequencing (NGS) technologies.²⁰⁻²² The major advantage of the technology is its efficiency in handling thousands of molecules with uniquely generated sequences and their high throughput analysis. The DNA tagging concept was initially applied in glycomics by Kwon and his coworkers, demonstrating Glyco-PCR, a method to detect Glycan-Binding Proteins (GBP) using DNA-conjugated glycan with high sensitivity.²³⁻²⁴ In one of the previous work reported by Kwon and his coworkers, they have shown glyco-quantitative polymerase chain reaction (Glyco-qPCR) assay platform that allows the ultrasensitive detection and quantification of glycan in biological samples. They used therapeutic carbohydrates like chondroitin sulfate (a GAG), and Proteoglycans (PG) from Chinese hamster ovary cells (CHO) separated by capillary electrophoresis, Glycoprotein obtained from PNGaseF treatment of decorin core Glycoprotein (GP). They used EDC/NHS chemistry to covalently link the DNA tag and glycan, and unbound DNA was washed using

streptavidin-coated magnetic beads as the glycans they used were biotinylated. They also did a proof of concept study, in GAG-binding proteins, where interaction was tested between Heparin-Antithrombin and Chondroitin-Antithrombin. With this work, they proved that glyco-qPCR can be used to analyze the glyco-interactome and by applying PCR amplification glycan as low as 1 zettamoles can be detected.²³

Later on, they also demonstrated glyco-PCR in the detection of active botulinum neurotoxin (BoNT). They used Sialyllactose, which interacts with binding-domain of BoNTs, sialyllactose-DNA conjugate was used as binding-probe for active BoNT and recovered through BoNT-immunoprecipitation. Glyco-PCR analysis of the bound sialyllactose-DNA is then used to detect low attomolar concentrations of BoNT and attomolar to femtomolar concentrations of BoNT in honey. They also did a comparative study through the dual binding of antibodies and monovalent or multivalent SL-DNA conjugates to toxins.²⁴

But these studies were limited to certain types of glycan and there was no rationale behind the selection of DNA sequences. Thus, developing a general platform for the DNA encoding of glycans and subsequent detection has a significance role to play in the functional glycomics. In this work, we demonstrated three fundamental aspects of DNA encoding of glycans, a universal methodology for the coding of glycans based on the structure of the glycan, an easy and efficient chemistry for the coupling of the DNA with the glycan and finally PCR and qPCR-based methods for the quantitative detection of the glycans.

1.1 Purpose of the Study

Glycans play a very important role in many biological processes including cell-cell responses, pathogenicity and immune responses etc. Despite so much importance, there is still a need for an efficient method for detection and analyzing glycan-protein interactions in the field of

carbohydrate research. Currently, carbohydrate microarray is successfully being used with the help of immobilization technology. But this also has its own limitations, most important one is its efficiency of mimicking the biological environment of the binding. Hence, we envision here developing a successful method for DNA-encoded Glycan Libraries (DEGLs).

There has already been reported evidence,²³⁻²⁴ for demonstrating Glyco-PCR, as a method to detect Glycan- Binding Protein (GBP). But these studies do not have any rationale behind the selection of DNA tags. Our aim is to develop a broadly accepted general platform for DNA encoding in glycomics. This could be achieved by proper selection of DNA codes, DNA-Glycan conjugation by click chemistry and method of detection of DNA-Glycan conjugates using qPCR.

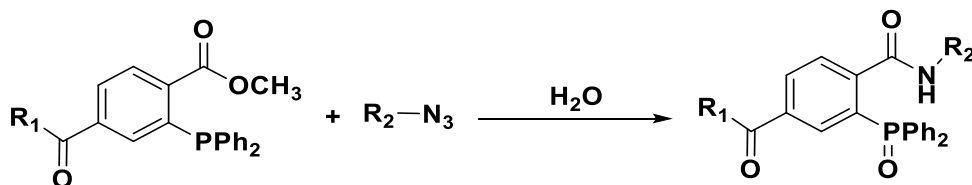
This work attempted to address all three steps generating a DNA code specifically for each glycan by a developing a software, which could be used in future for any DEGL containing any specific glycans, generation of DNA tags for each element in the library by adding forward and reverse primer binding on each end for PCR amplification, and 5' terminal alkyne modification to offer coupling with the glycan using click chemistry (CuAAC) and to develop a method of detection of DNA-Glycan conjugates using qPCR as a proof principle application study.

Generating a DNA code specific for each glycan-based on its structure was attained by developing a software using basic computing principles in a computer language Python. The development and the logic underlying in generating a DNA code specific for each glycan is described elaborately under experiment section later.

1.1.1 Click Chemistry

Lack of abundant samples to play with demands a near perfect conjugation chemistry for the synthesis of DNA glycan conjugates. We tested few of the well-known bioorthogonal reactions widely used in chemical biology, glycan analysis and detection many cell surface glycans.²⁵⁻²⁶

Well developed bioorthogonal reactions include Staudinger Ligation,²⁷ Copper-catalyzed azide-alkyne Cycloaddition (also called as CuAAC),²⁸ strain promoted alkyne-azide cycloaddition (also called as Cu-free click chemistry or SPAAC).²⁹

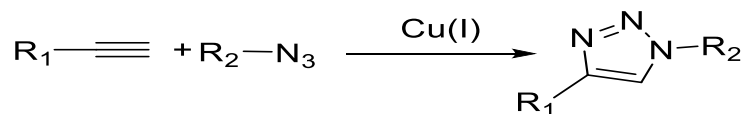


R1,R2 = Biomolecules, Bioactive Molecules, Fluorophores, Affinity Tags etc.

Figure 1.1 Staudinger Ligation

Staudinger ligation was the first bioorthogonal reaction reported by Staudinger and co-workers in 1919.²⁷ This is bioorthogonal ligation that transforms azides to primary amines by phosphines.³⁰ Later some modifications were made to this reaction, such as using modified reagent. Accompanied by those changes, Staudinger ligation suffered from slow kinetics and easily oxidized phosphine reagents during the reaction process, thus making the reaction not a perfect candidate for glycan analysis and detection.³¹

The most popular and common bioorthogonal method is copper-catalyzed click chemistry.^{28,32} This is a copper-catalyzed cycloaddition between alkyne and azide to form a 5-membered ring.³³ In early 2000's, Sharpless group and Meldel group discovered that this reaction could be accelerated by copper dramatically, and this reaction can take place in an aqueous system.^{27,31} This copper-catalyzed azide-alkyne cycloaddition (CuAAC) is used widely in chemical biology, cell biology, cell surface glycan analysis and detection. This reaction being very simple and easy to perform it is widely been used by researchers both in industry and academia. During our study, we found most satisfying results with the click reactions and we adopted this chemistry for all our future work.



R1,R2 = Biomolecules, Bioactive Molecules, Fluorophores, Affinity Tags etc.

Figure 1.2 Copper-Catalyzed Click Chemistry (CuAAC)

1.1.2 PCR

Another important part of our study is PCR, which is highly essential for the detection of DNA that is coupled with glycan which in turn symbolizes the glycan presence. The advent of the Polymerase Chain Reaction (PCR) radically transformed biological science from the time it was first discovered by Mullis in 1990. PCR based strategies have propelled huge scientific endeavors such as Human Genome Project. The technique is currently used by clinicians and researchers to diagnose diseases, clone and sequence genes, and carry out sophisticated quantitative and genomic studies in rapid and very sensitive manner. One of the most important medical applications of PCR method is the detection of pathogens. In addition, it is also widely used in forensics for the identification of criminals.⁴⁶

PCR can be performed using source DNA from a variety of tissues and organisms, including peripheral blood, skin, hair, saliva, and microbes. Only trace amounts of DNA are needed for PCR to generate enough copies to be analyzed using conventional laboratory methods. For this reason, PCR is a sensitive assay. Each PCR assay requires the presence of template DNA, primers, nucleotides, and DNA polymerase. The DNA polymerase is the key enzyme that links individual nucleotides together to form the PCR product. The nucleotides include the four bases – adenine, thymine, cytosine, and guanine (A, T, C, G) – that are found in DNA. These act as the building blocks that are used by the DNA polymerase to create the resultant PCR product. The primers in the reaction specify the exact DNA product to be amplified. The primers are short DNA fragments with a defined sequence complementary to the target DNA that is to be detected and

amplified. These serve as an extension point for the DNA polymerase to build on. The above-mentioned components are mixed in a test tube or 96-well plate and then placed in a thermal cycler that allows repeated cycles of DNA amplification to occur in three basic steps. The reaction solution is first heated above the melting point of the two complementary DNA strands of the target DNA, which allows the strands to separate, a process called denaturation. The temperature is then lowered to allow the specific primers to bind to the target DNA segments, a process known as hybridization or annealing. Annealing between primers and the target DNA occurs only if they are complementary in sequence (e.g. A binding to G). The temperature is raised again, at which time the DNA polymerase is able to extend the primers by adding nucleotides to the developing DNA strand. With each repetition of these three steps, the number of copied DNA molecules increases exponentially.⁴⁶ The PCR product is analyzed using agarose gel electrophoresis, which separates the DNA product based on the size and charge. The two main methods used for visualization of DNA are staining the product using ethidium bromide dye, which intercalates between the two strands of the DNA. However, due to the prevailing carcinogenic and serious health effects of ethidium bromide, its use has been reduced and replaced by SYBR™, a safe DNA gel stain.

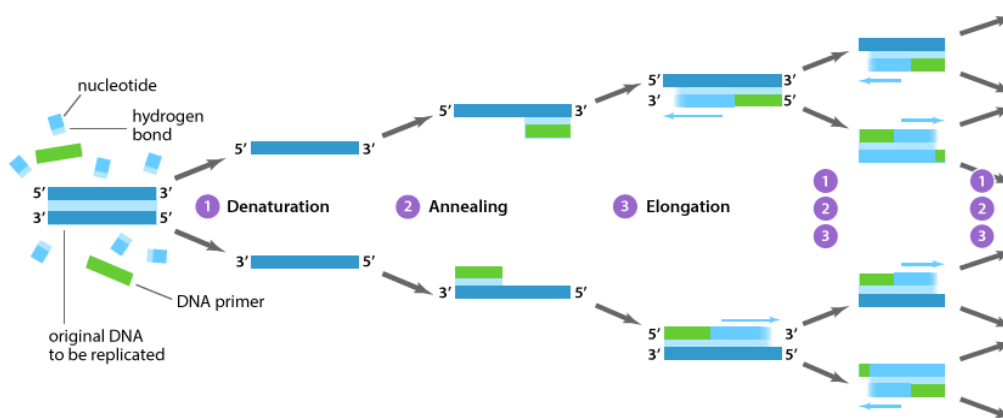


Figure 1.3 Key Components Of PCR

1) Denaturation: the reaction is heated to 95°C-98°C for 20-30 secs to break the hydrogen bonds; 2) Annealing: The reaction temperature is lowered to 50°C- 65°C for 20-40 secs (preferably 2°C-4°C less than the T_m of primers) to allow the primers to anneal to the template strands; 3) Elongation: The temperature is increased (optimum temperature of DNA polymerase preferably 72°C-78°C) to allow the addition of dNTP's. At end of each cycle, the target sequence is doubled.

The next generation advancement of the PCR technology was the development of RT-qPCR. While in end point, PCR requires that PCR products were detected and quantified by gel electrophoresis after completion of the reaction, real-time qPCR technology allows quantification of PCR products in “real time” during each PCR cycle, yielding a quantitative measurement of PCR products accumulated during the course of the reaction. Real-time reactions are carried out in a thermocycler that permits measurement of a fluorescent detector molecule, which decreases post-processing steps and minimizes experimental error. This is most commonly achieved through the use of fluorescence based technologies. Three main fluorescence based technologies are, probe sequences that fluoresce upon (i) hydrolysis (TaqMan; Applied Biosystems, Foster City, CA, USA) or hybridization (LightCycler; Roche, Indianapolis, IN, USA); (ii) fluorescent hairpins; or (iii) intercalating dyes (SYBR Green). As SYBR™ Green binding is not specific for a target sequence this system can be readily used for different gene assays, is flexible, inexpensive, and accurate results can be obtained provided validation of the specificity by melt curve (or dissociation curve) analysis. The TaqMan™ chemistry is more expensive than DNA binding dye assays, but the presence of the hydrolysis probe ensures that only specific amplicons are measured. In addition, multiplexing reactions are possible, although their set-up requires an important optimization phase.^{47,48}

In brief, similar to end-point PCR, qPCR consists a succession of amplification cycles in which the template nucleic acid is denatured, annealed with specific oligonucleotide primers, and extended to generate a complementary strand using a thermostable DNA polymerase. This results in an exponential increase of amplicons (amplification products) that, in contrast with end-point

PCR, can be monitored at every cycle (in real time) using a fluorescent reporter. The increase in fluorescence is plotted against the cycle number to generate the amplification curve, from which a quantification cycle C_q (often described as C_t for cycle threshold) value can be determined. C_q corresponds to the number of cycles for which the amount of fluorescence (hence, of the template) is significantly higher than the background fluorescence. Therefore, the C_q value can be linked to the initial concentration of target nucleic acid and serves as a basis for absolute or relative template quantification. All steps of (RT-)qPCR may introduce experimental errors. qPCR is a robust technique, but due to its high sensitivity, very small variations can induce non-negligible differences in the results. To measure intra-assay variability, which follows a statistical distribution, RT-qPCR can be performed in triplicate (experimental replicates). Inter-assay variability can be estimated using a “reference” sample that will be included in each experiment.⁴⁸

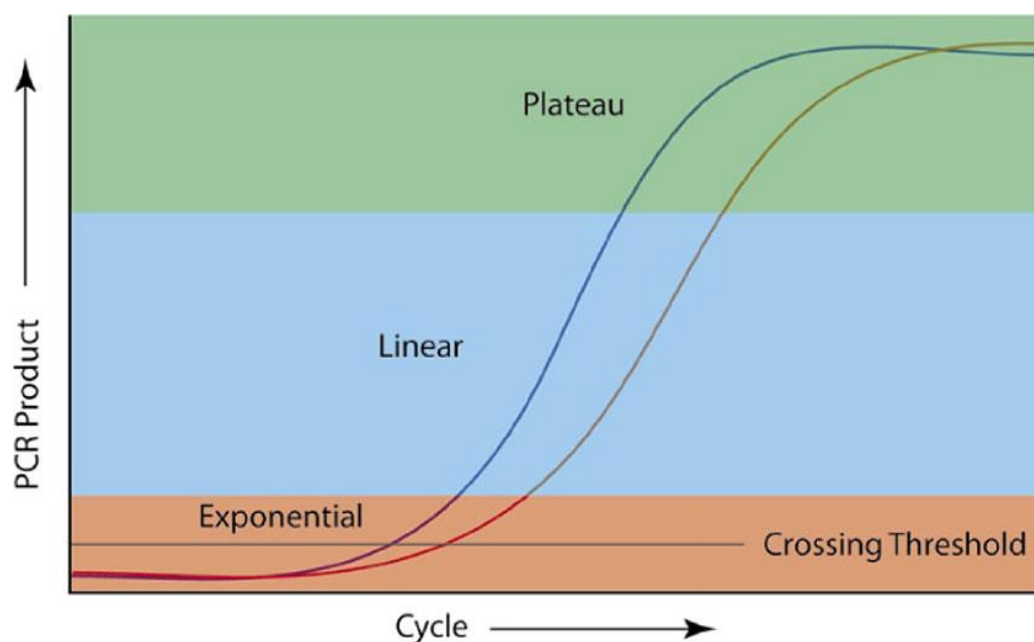


Figure 1.4 Phases of qPCR.

The PCR process goes through three main phases as the number of cycles and the amount of product generated increase. Initially, when the amount of product is small and enzyme and reagents are not limiting, product generation is exponential and the reaction is closest to 100% efficiency. This exponential growth is hard to detect initially through real-time fluorescence because the amount of product is small. During the linear phase, products continue to accumulate, but the reaction efficiency begins to fall and reagents

*become limiting. Finally, in the plateau phase of the reaction, accumulation of product ceases as the reaction is exhausted for a number of different reasons.*⁴⁷

For overall method development we adopted two systems, initially, we used the most known blood glycan antibody system (ABO) for the early stage method optimizations and, later we used the Globo-series glycan analogs as a small DNA encoded library for the proof of principle demonstration via detecting the breast cancer from patient plasma.

1.1.3 ABO Blood Group System

The discovery of the ABO blood group, over 100 years ago, caused great excitement. Until then, all blood had been assumed to be the same, and the often-tragic consequences of blood transfusions were not understood. The modern basis of the ABO blood group system was coined by Karl Landsteiner in 1900 when he discovered that the blood group of certain people agglutinated red cells present in other individuals.³⁴ Ever since it has been studied extensively and well characterized with the structures of antigens and antibodies involved. The terminal carbohydrate structures present on the red blood cells and their antibodies present in the blood plasma determine the blood group.

The ABO blood group antigens remain of prime importance in transfusion medicine—they are the most immunogenic of all the blood group antigens. The ABO blood group antigens also appear to have been important throughout our evolution because the frequencies of different ABO blood types vary among different populations, suggesting that a particular blood type conferred a selection advantage (e.g., resistance against an infectious disease.)

The ABO blood group antigens are encoded by one genetic locus, the ABO locus, which has three alternative (allelic) forms—A, B, and O. A child receives one of the three alleles from each parent, giving rise to six possible genotypes and four possible blood types (phenotypes).

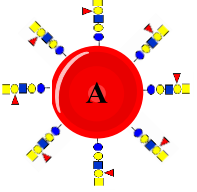
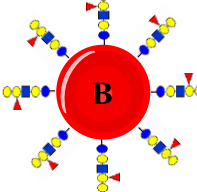
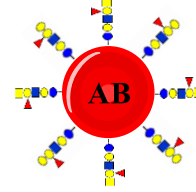
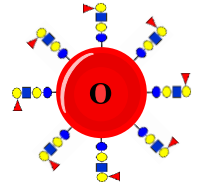



Table 1.1 Possible Blood types in offsprings

		ABO Alleles Inherited from Mother			
		A	B	O	
ABO Inherited from Father	A	A	AB	A	
	B	AB	B	B	
	O	A	B	O	

The four basic ABO phenotypes are O, A, B, and AB. After it was found that blood group A RBCs reacted differently to a particular antibody (later called anti-A1), the blood group was divided into two phenotypes, A1 and A2. RBCs with the A1 phenotype react with anti-A1 and makeup about 80% of blood type A. RBCs with the A2 phenotype do not react with anti-A1 and they make up about 20% of blood type A. A1 red cells express about 5 times more A antigen than A2 red cells, but both types of red cell react with anti-A, and A1 and A2 blood groups are interchangeable.

The immune system forms antibodies against whichever ABO blood group antigens are not found on the individual's RBCs. Thus, a group A individual will have anti-B antibodies and a group B individual will have anti-A antibodies. Blood group O is common, and individuals with this blood type will have both anti-A and anti-B in their serum. Blood group AB is the least common, and these individuals will have neither anti-A nor anti-B in their serum.

Table 1.2 Antibodies and Antigen present in ABO Blood Typing System

RBC of different blood group				
Antibodies in Serum	 Ab B	 Ab A	None	 Ab A Ab B
Antigen in RBC surface	A	B	A and B	None
Genotype	AO or AA	BO or BB	AB	OO

The ABO locus is located on chromosome 9 and has three main allelic forms: A, B, and O. The A allele encodes for a glycosyltransferase that produces the A antigen (N-acetylgalactosamine is its immunodominant sugar), and the B allele encodes for a glycosyltransferase that creates the B antigen (D-galactose is its immunodominant sugar). The O allele encodes for an enzyme with no function, and therefore neither A or B antigen is produced, leaving the underlying precursor (the H antigen) unchanged. These antigens are incorporated into one of four types of oligosaccharide chain, type 2 being the most common in the antigen-carrying molecules in RBC membranes.

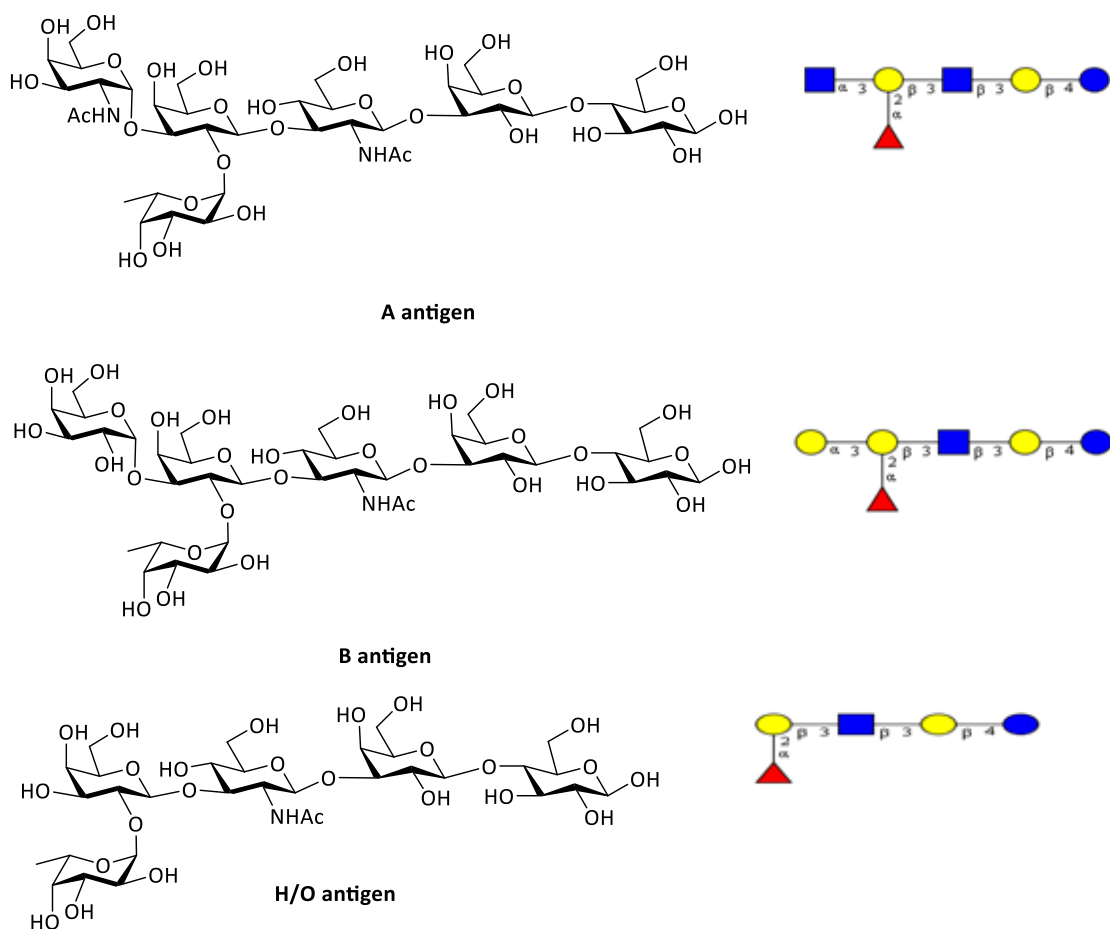


Figure 1.5 ABO Blood antigen structures.

For A/B antigen synthesis to occur, a precursor called the H antigen must be present. In RBCs, the enzyme that synthesizes the H antigen is encoded by the H locus (FUT1). In saliva and other bodily secretions, the enzyme that synthesizes the H antigen is encoded by the Se locus

(FUT2). The H locus is located on chromosome 19 and it encodes a fucosyltransferase that produces the H antigen on RBCs.⁴⁹

Table 1.3 Distribution of various Blood Groups among Population

Population	Blood Group 'O'	Blood Group 'A1'	Blood Group 'A2'	Blood Group 'B'	Blood Group 'A1B'	Blood Group 'A2B'
Caucasians	44%	33%	10%	9%	3%	1%
Africans	49%	19%	8%	20%	3%	1%
Asians	43%	27%	Rare	25%	5%	Rare

1.1.4 *Globo-H*

Similarly, eukaryotic cells are covered by a glycocalyx: an extensive network of polysaccharides. The glycocalyx serves as a rich source of binding sites for receptors and ligands, as well as pathogens and toxins. The mammalian glycome is estimated to consist of a few hundred unique glycan structures on glycoproteins and glycolipids. One such glycan is Globo H.⁵⁰

Some of the Globo-series glycans are overexpressed in certain cancers and are proven as biomarkers for early detection of breast cancer and ovarian cancer.^{7,35-37} Recent developments in the Globo-H immunotherapy also needs high sensitive methods to quantitatively detect the antiglobo-H antibodies for monitoring the therapy and in clinical trials.³⁸⁻⁴⁰

Globo H is a glycosphingolipid of the Globo series with a sugar terminus resembling the blood group antigen H determinant. First identified in human teratocarcinoma and breast cancer cells, Globo H was found to be overexpressed on the cell surface of several epithelial cancers such as breast, colon, endometrial, gastric, pancreatic, lung, and prostate cancers. It is also moderately expressed in normal epithelial tissues (lung, breast, prostate, stomach, pancreas, and ovary), but its distribution is restricted to apical epithelial cells at lumen borders.³⁶ In breast cancer Globo H expression was observed in >60% of ductal, lobular, and tubular carcinoma, but not in nonepithelial breast tumors. Globo H is not expressed in normal tissue except for weak expression

in the apical epithelial cells at lumen borders, a site that appears to be inaccessible to the immune system. Thus, Globo H has been considered as an ideal target for immunotherapy of many epithelial cancers and indeed two phase I trials of a Globo H-based vaccine in breast and prostate cancer, respectively, have shown promising results.⁴⁵

Functionally, Globo H has been associated with tumor stem cells, to be a potent inducer of angiogenesis, and an immunosuppressor through Notch signaling. The high Globo H expression by only cancer and cancer stem cells made it an attractive target for the generation of therapeutic cancer vaccines. These vaccines underwent the long history of development and improvement and are currently being tested in clinical trials for treatment. Globo H has therefore been considered as one promising tumor associated glycan biomarker, in particular for breast cancer.³⁶

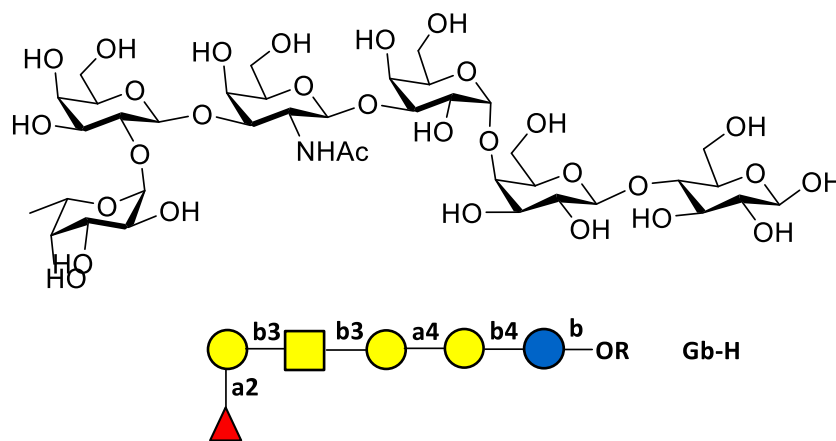


Figure 1.6 Globo-H Structure

2 EXPERIMENT

2.1 Systematic Coding of Sugars

The first aim was to develop a unique coding method that is consistent with the representation of glycans, so it can be used in future for any DEGL with ease and without any uncertainty in future glycomics. Unlike peptides, proteins and DNA/RNA, representation of carbohydrate brings many challenges. A detailed representation should feature composition, sequence, branching position, modifications, and anomeric configuration. The use of symbols to represent the glycans was first proposed by Kornfield et.al.⁴¹ This approach has replaced the use of IUPAC naming for representation of glycan, with modifications adapted timely to fit the entire glycome.⁴²⁻⁴³ All these representations are based on symbols and color codes to represent the building blocks.

Table 2.1 Glycan Representation (CFG)

Symbolic Representation (Cartoon Representation CFG)	IUPAC Naming (Linear Code)
	GalNAc b1-4 (Fuc a1-2) GlcNAc b1-4 Man a1-3 (GalNAc b1-4 (Fuc a1-2) GlcNAc b1-4 Man a1-6) Man b1-4 GlcNAc b1-4 (Fuc a1-6) GlcNAc

With the above approach, it is pragmatic to use nucleotide bases to represent the glycans. But this also brings prevailing challenges for using nucleotide bases for glycans as there are many symbols and color codes are available for representation, but there are only 4 nucleotides bases A, T, G, and C are available for coding. To have enough nucleotide codes to cover the entire the glycome we had to depend on permutation and combination approach. There are many sugar monomers which are incorporated in many ways in a glycan, making glycan structural information

very complex. These inherit difficulties lead to the development of a dedicated computer program, which can manage the structural information of the glycan to coding dictionaries and transform the IUPAC name of glycan to single stranded DNA code. For easy acceptance of the codes, the codes should be of minimal length storing all the glycan information and enable both PCR and qPCR amplification.

For a perfect structure based coding, there must be a controlled vocabulary for all the structural components. For attaining this goal, the structural components of the carbohydrate were split into the structural components monosaccharides, linkages, and modifications etc. Three independent libraries were generated namely library A for monomer, library B for linkage and library C for modification. Next, all the monomers were gathered in library A and assigned each one with unique four-letter code. There are about 256 combinations generated using A, T, G, and C. About 70 monomers are added to this library as of now using the randomly generated codes. In a similar way, all possible linkages were accounted for library B, using codes based on 3 letters. Currently, there are 36 linkages in this library. For library C about 100 modifications were identified and incorporated in the library C and assigned with 4 letter codes which were not used in the library A. Special characters like '(' in the IUPAC naming for branching were assigned special codes. All these libraries were then incorporated into a Python-based program and the program was tested for about 100 glycans, for which the program successfully delivered the codes. The libraries can be extended and modified for addition as and when needed, to make this program a universal one and available for research community.

Table 2.2 Glycan Representation with DNA Codes

Symbolic Representation (Cartoon Representation CFG)	IUPAC Naming (Linear Code)	DNA Code obtained through the program
	GalNAc b1-4 (Fuc a1-2) GlcNAc b1-4 Man a1-3 (GalNAc b1-4 (Fuc a1-2) GlcNAc b1-4 Man a1-6) Man b1-4 4 GlcNAc b1-4	ACGCTGTTTT AGAAAAGAA AAAGCTGTAA TAACTTTAC GCTGTTTTAG AAAAGAAAA AGCTGTAATA ACAAAAAAT

A sequence length between 50 -100 is best for PCR applications. Among 100 glycans tested most fell within this limit, but some exceeded this. Through careful observation of the sample library, there were many repeating fragments in the glycans (e.g. Six sugar core of N-glycans). Hence, by taking this as a single block would reduce the length overall length of the codes of the long glycans, so a fourth library (Library D) was included in the program containing most of the common building blocks found in the glycans. Similar to other libraries, library D can also be extended, so as to cover the entire glycome and make it a universal one.

Table 2.3 Long and Short DNA codes of the Glycan

IUPAC Naming (Linear Code)	Long DNA Code obtained through the program	Short DNA Code obtained through the program
GalNAc b1-4 (Fuc a1-2) GlcNAc b1-4 Man a1-3 (GalNAc b1-4 (Fuc a1-2) GlcNAc b1-4 Man a1-6) Man b1-4 GlcNAc b1-4 (Fuc a1-6) GlcNAc	ACGCTGTTTTAGAAAAGAAA AAGCTGTAATAAACTTTACG CTGTTTTAGAAAAGAAAAG CTGTAATAACAAAAAATATG TAAGCTGTTTTAGAAACAAA AAAGC Length – 105	ACGCTGTTTTAGAAA AGAAAAGCTGTAA TAACTTTACGCTGT TTAGAAAAGAAA AGCTGTAATAACAA AAGTACA Length - 79

Another probable issue encountered was with short glycans, which gives small codes of length less than 25. This will face practical difficulty while doing qPCR. To resolve this issue a random DNA sequence of length 20 was added to the code with a mention in the program to avoid

the misreading. These two modifications made sure to keep the sequences within the desired limits to be used in the qPCR.

Table 2.4 DNA code for short glycans

IUPAC Naming (Linear Code)	DNA Code obtained through the program	Expanded DNA Code obtained through the program
Rha	ATAG Length – 4	ATAGCCCCAGTCAGGC CTAACGTA Length - 24

2.2 Coding and Decoding of glycans through Program

2.2.1 Coding

For getting nucleotide sequence, glycan sequence is given in the input box with spaces in between each character of the glycan sequence. When a glycan sequence is inputted the program looks for each character from left to right. It checks for match in monomer dictionary (Library A), linkage dictionary (Library B) and modification dictionary (Library C). If a match is found, the corresponding nucleotide is placed under the name ‘**nuc**’.

If the length of the nucleotide is less than 20, then an R nucleotide sequence is added to the ‘**nuc**’ to increase the length the nucleotide sequence and output is published.

If length of ‘**nuc**’ is more than 20 then the whole length of ‘**nuc**’ is searched for any matching sequences in the core dictionary (Library D). If there is match then a lengthy nucleotide core is replaced by a short 5 length sequence. This is done to reduce a lengthy nucleotide to short one for cost effectiveness. Reduced ‘**nuc**’ output is published.

2.2.2 Decoding

For decoding the nucleotide sequence back to glycan sequence, in the input box nucleotide sequence is given. The nucleotide sequence is searched for any matching five-length sequences in

the core dictionary (Library D). If there is any matching sequence then the length of nucleotide is extended to the original lengthy nucleotide sequence.

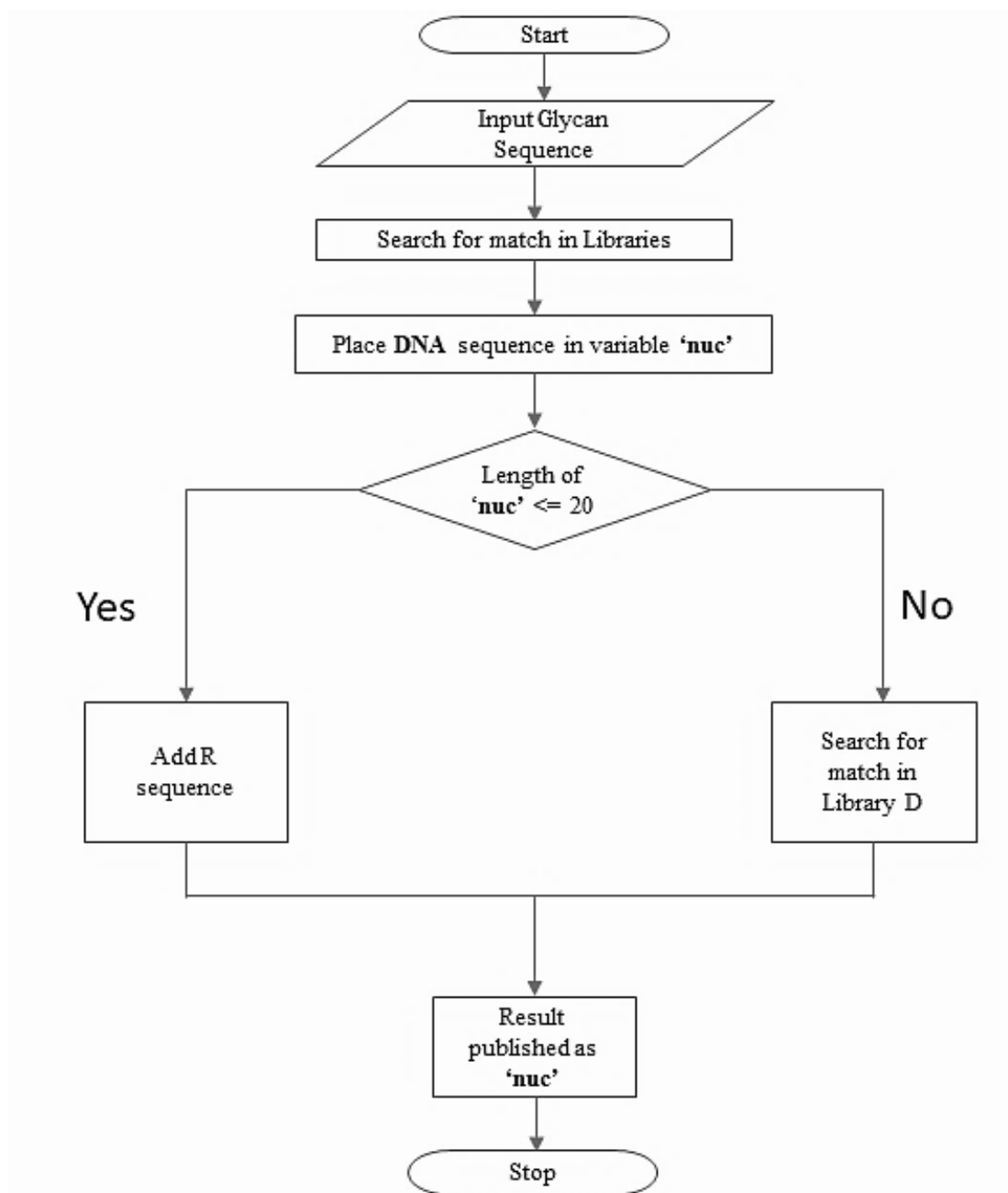


Figure 2.1 Flowchart describing Coding

If length of nucleotide is more than 0; first four nucleotide sequence are taken and match is found in the carb dictionary (Library A) and corresponding carb is taken for the output. Next

three nucleotide sequences are taken and searched for a match in linkage dictionary (Library B); if there are no match then next four nucleotides are taken and searched in modification dictionary (Library C) and next three nucleotides are again searched for a match in linkage dictionary (Library C). Then next three nucleotides are taken and seen if it is 'AAA' or 'TTT', if it is so then a corresponding branching is taken for output. Then next three nucleotides are taken and seen if it is 'AAA' or 'TTT', if it is so then a corresponding branching is taken for output. This is done to cover the double branching cases. This continues for the complete length of the nucleotide sequence, before publishing the output.

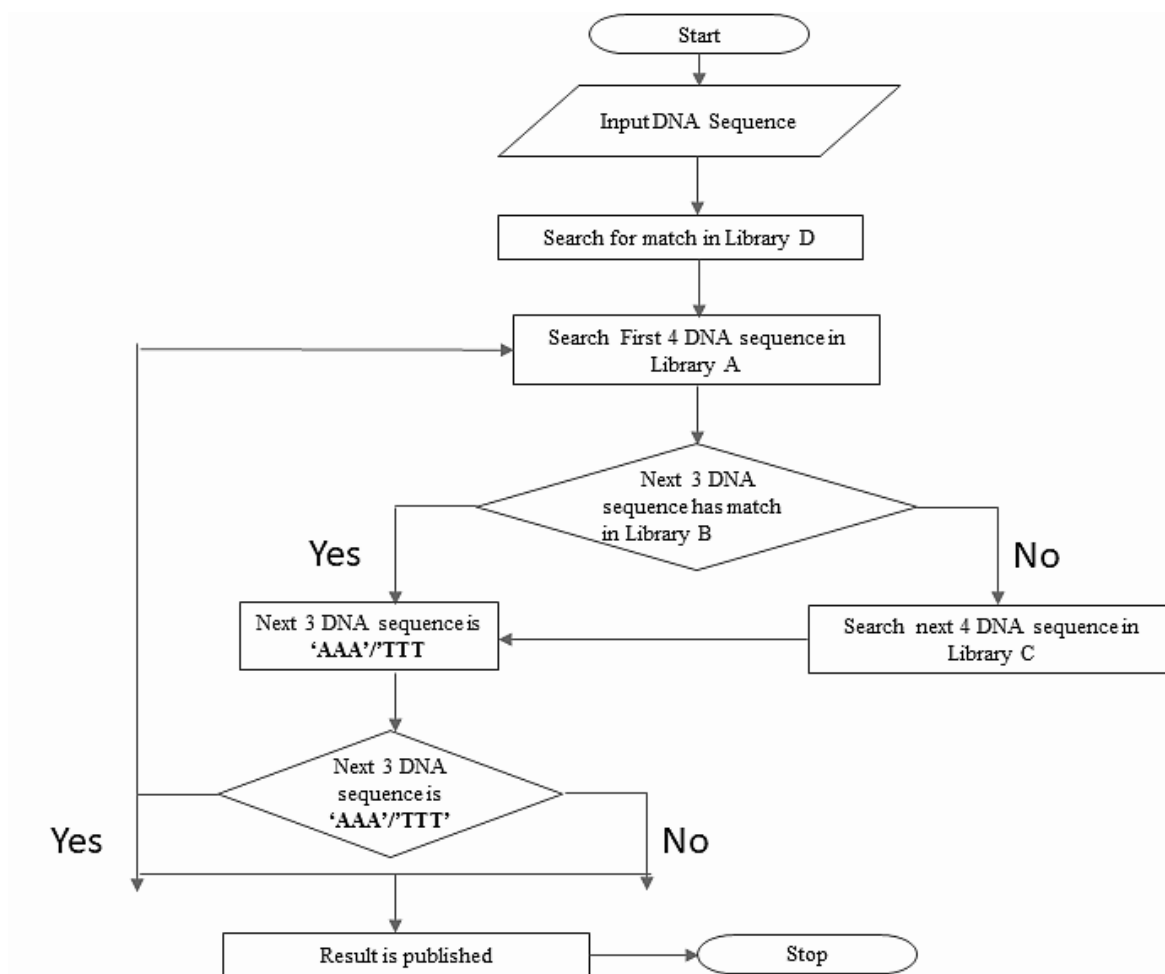


Figure 2.2 Flowchart describing decoding

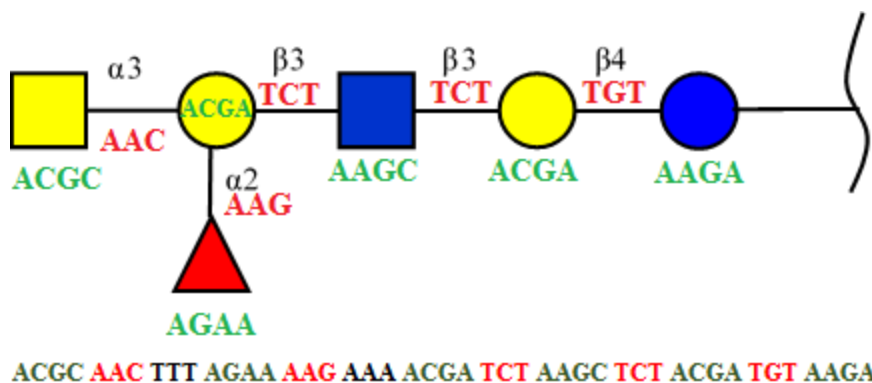


Figure 2.3 CF representation and DNA code for blood antigen A

2.3 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] amine (TBTA) Click chemistry Ligand were purchased from TCI (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). 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) were purchased from Thermo Fisher.

2.4 Gradient PCR and T_m Determination

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 during annealing phase of the cycle (95°C for 10 min, X°C for 30s, 95°C for 15s, 13 cycles) wherein X represents the annealing phase and temperature gradient from 50 °C - 60 °C 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.

Table 2.5 Reaction mix for gradient PCR

Component	25 uL	50 uL	Final Concentration
10X Amplifying Buffer	5 uL	10 uL	-
MgSO ₄	1 uL	2 uL	-
dNTP(10mM)	0.75 uL	1.5 uL	0.3mM
Forward Primer(10uM)	0.75 uL	1.5 uL	0.3uM
Reverse Primer(10uM)	0.75 uL	1.5 uL	0.3uM
Template(1uM)	1 uL	2 uL	0.04
Polymerase	0.3 uL	0.5 uL	-
Nuclease-Free Water	15.4 uL	31 uL	-

2.5 Synthesis of Glycan(Antigen) – 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. All glycans were synthesized via chemo enzymatic method with the azido propyl linker at the reducing end.

2.5.1 Click reaction procedure

5’-Hexynyl-terminated DNA (500µM, 20 µL, final concentration of 50 µM), azido glycan (500µM, 30 µL, final concentration of 75µM), 2M TEA buffer (pH 7, 20 µL, final concentration of 0.2M), 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).

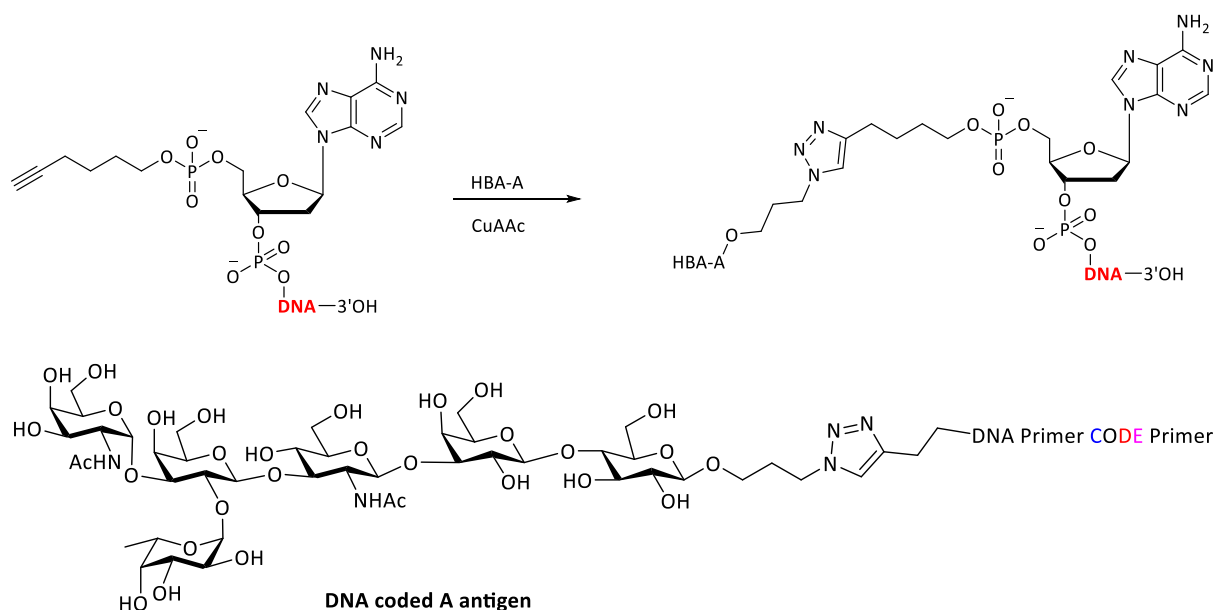


Figure 2.4 Click Conjugation of 5'-Hexynyl DNA and azido modified glycans

2.6 General protocol for qPCR

2.6.1 For a typical 12.5 μ L reaction

The glycan(Antigen)- DNA Conjugates were added to 6.25 μ L 2X Maxima SYBR Green/ROX qPCR Master Mix with 1 μ L primers (Forward Primer – 500 nm and Reverse Primer- 300 nm). qPCR was performed with Applied Biosystems Stepone System (50 $^{\circ}$ C for 2 mins (Holding), 95 $^{\circ}$ C for 10 mins (Holding), 95 $^{\circ}$ C for 15 s, 60 $^{\circ}$ C for 10 s, 72 $^{\circ}$ 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 a large amount of DNA will have a lower Ct value compared to those samples containing relatively less amount of DNA.

Table 2.6 Reaction mix for qPCR

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

2.7 Filtration of Antibody-Antigen DNA Conjugates

10 μ L of glycan(antigen)-DNA Conjugate (5 μ M) was incubated with 2 μ L of antibody (if plasma is used then 2 μ L of plasma is diluted to 8 μ L using water) for 2 hours at room temperature. After two hours, the reaction volume was made up to 100 μ L using TBST buffer. The reaction solution was washed using 100 μ L TBST buffer for 7 times in 100kda cutoff centrifugation filtration filter tubes at 5000g for 10 mins. After washing the filtered and concentrated solution was used as a template for qPCR analysis.

2.8 Immunoprecipitation of Antibody-Antigen-DNA Conjugates

10 μ L of glycan(antigen)- DNA Conjugate (5 μ M) was incubated with 2 μ L of antibody (if plasma is used then 2 μ L of plasma is diluted to 8 μ L using water) for 2 hours at room temperature. After two hours, the reaction volume was made up to 100 μ L using TBST buffer and added to 20 μ L 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 μ L of TBST buffer. The beads were directly used a template for qPCR analysis.

2.9 Blood Collection

Blood was collected from 10 volunteers by finger pricking, and collected to heparin-coated blood collection tubes. The collected blood was centrifuged at 5000 rpm for about 5 mins for obtaining plasma from blood. The collected plasma is stored at -20°C until use.

3 RESULTS AND DISSCUSSION

In this study, I have developed a program which can generate DNA code specifically for each glycan (described in systematic coding of sugars under experimental section). Next steps were to successfully couple DNA and glycan using click chemistry, develop methods for glycan and protein binding, and analyze the glycan using qPCR.

3.1 DNA sequences

Using the developed Python-based program DNA sequences for blood group antigens and Globo glycan series were obtained.

Table 3.1 DNA sequences

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

Name	5'Mod	Sequence	3'Mod
O DNA	5Hexynyl	AATGATACGGCGACCACCGAAGAAAAGACGATCTAAGCTCT ACGATGTAAGACACGTCTGAACTCCAGTCAC	None
A DNA	5Hexynyl	AATGATACGGCGACCACCGAACGCAACTTTAGAAAAGAAAA CGATCTAAGATCTACGATGTAACACGTCTGAACTCCAGTCAC	None
B DNA	5Hexynyl	AATGATACGGCGACCACCGAACGAACTTTAGAAAAGAAAA CGATCTAAGCTCTACGATGTAAGACACGTCTGAACTCCAGTCAC	None
Gb-H	5Hexynyl	AATGATACGGCGACCACCGAAAGAAAAGACGATCTACGCTC TACGAATAACGATGTAAGACACGTCTGAACTCCAGTCAC	None
Gb-5	5Hexynyl	AATGATACGGCGACCACCGAACGATCTACGCTCTACGAATA ACGATGTAAGA CACGTCTGAACTCCAGTCAC	None
Gb-4	5Hexynyl	AATGATACGGCGACCACCGAACGCTCTACGAATAACGATGTA AGACACGTCTGAACTCCAGTCAC	None
Gb-3	5Hexynyl	AATGATACGGCGACCACCGAACGAATAACGATGTAAGACC CCAGTCAGGCCTAACGTACACGTCTGAACTCCAGTCAC	None
Gb-2	5Hexynyl	AATGATACGGCGACCACCGAACGATGTAAGACCCAGTCA GGCCTAACGTA CACGTCTGAACTCCAGTCAC	None
Bb-4	5Hexynyl	AATGATACGGCGACCACCGAAAGAAAAGACGATCTACGCTC TACGACACGTCTGAACTCCAGTCAC	None
Bb-3	5Hexynyl	AATGATACGGCGACCACCGAAAGAAAAGACGATCTACGCC CCAGTCAGGCCTAACGTACACGTCTGAACTCCAGTCAC	None
Bb-2	5Hexynyl	AATGATACGGCGACCACCGAAAGAAAAGACGACCCAGTCA GGCCTAACGTACACGTCTGAACTCCAGTCAC	None
FW Primer	None	AATGATACGGCGACCACCGAA	None
RV Primer	None	GTGACTGGAGTTCAGACGTG	None

The obtained DNA sequences were flanked with primer binding domains and 5' Hexynyl modification were procured from IDT.

3.2 T_m Determination

T_m was determined using the protocol for gradient PCR and T_m Determination (under experimental section) for further quantitation experiments. After PCR to determine the optimum temperature the PCR products were analyzed in using 4% agarose gel.

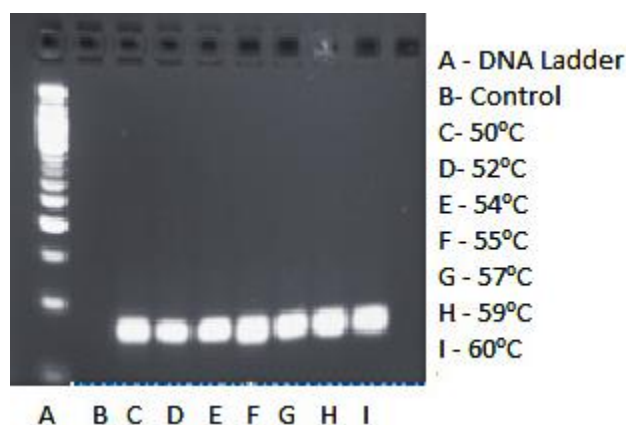


Figure 3.1 T_m Determination

Gel analysis of the PCR product (Figure 3.1) showed the optimum temperature to be in a range of 52°C-60°C. 60°C was used as annealing temperature for further experiments.

3.3 Glycan (Antigen) – DNA Conjugates

After T_m determination, the 5'-Hexynyl terminated DNA and azido glycan were coupled together using click chemistry according to the protocol mentioned under synthesis of Glycan(Antigen) – DNA Conjugates (under experimental section).

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 T_m Determination and analyzed using 4% agarose gel electrophoresis.

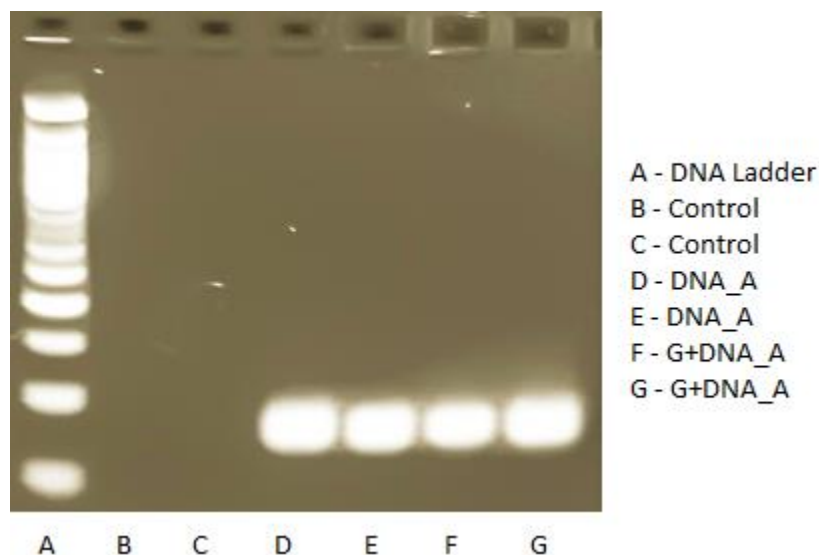


Figure 3.2 PCR Comparison of DNA_A and G+DNA_A

From the gel analysis of the PCR product (Figure 3.2) a clear concrete band was visible in both DNA and G+DNA, which made sure that the DNA can be used for the further analysis after conjugation with the glycan.

To make sure that the Glycan-DNA conjugate can be used for binding the protein with the glycan and DNA can be used for PCR analysis, any presence of residual unreacted DNA was further tested by PCR amplification and gel electrophoresis. Pure DNA_A and G+DNA_A 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 (Figure 3.3) did not show a band in the wash buffer (residual unreacted DNA) but a strong band was seen in elution buffer indicating near perfect efficiency of coupling protocol.

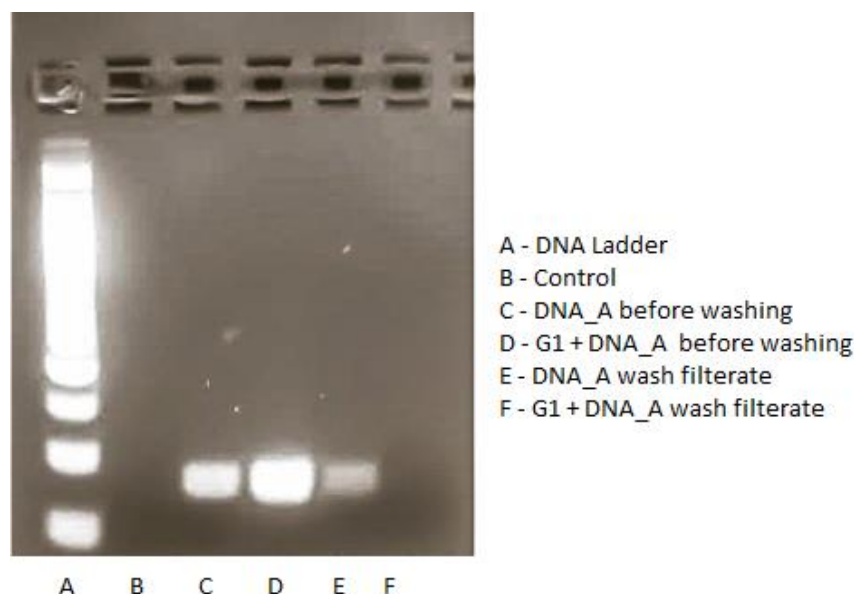


Figure 3.3 DNA and Glycan- DNA Conjugate IP Analysis

3.4 qPCR Reactions and Standard Curve

The success of DEGL lies in the readiness of applying the DNA codes to the standard PCR and qPCR protocols. It is important to have the glycan-coupled DNA (G+DNA) achieve similar PCR efficiency to the native DNA, so compared the G+DNA conjugates with the corresponding DNA strands to verify the efficiency of amplification in both PCR and qPCR. In all the above sections, PCR was performed on the both DNA and G+DNA using standard thermocycling with different annealing temperatures and Immunoprecipitation methods. Next, experiment was carried out to determine the qPCR detection limit of both pure DNA and G+DNA. Both pure DNA and G+DNA 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 Ct values, Critical threshold value (Ct) is generally used to compare the concentration of the templates in the qPCR reaction mixture, 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 3.4). The test was not conducted for further low concentrations worrying about the

interference of primer dimer formation, a potential challenge in amplifying templates smaller than 100 bases. Nevertheless, a standard plot of the Ct value vs log concentration of the DNA and G+DNA conjugate (Figure 3.5) did show the linear relation indicating the successful application of G+DNA for quantitative detection up to the picomolar level. Both DNA and DNA-Glycan conjugates gave similar amplification, assuring that the DNA tag can be an alternative to detect the glycan itself. Encouraged by the result clearly suggesting the detection of glycans from even a picomolar level, proceeded to the final goal, screening of glycans against glycan binding proteins.

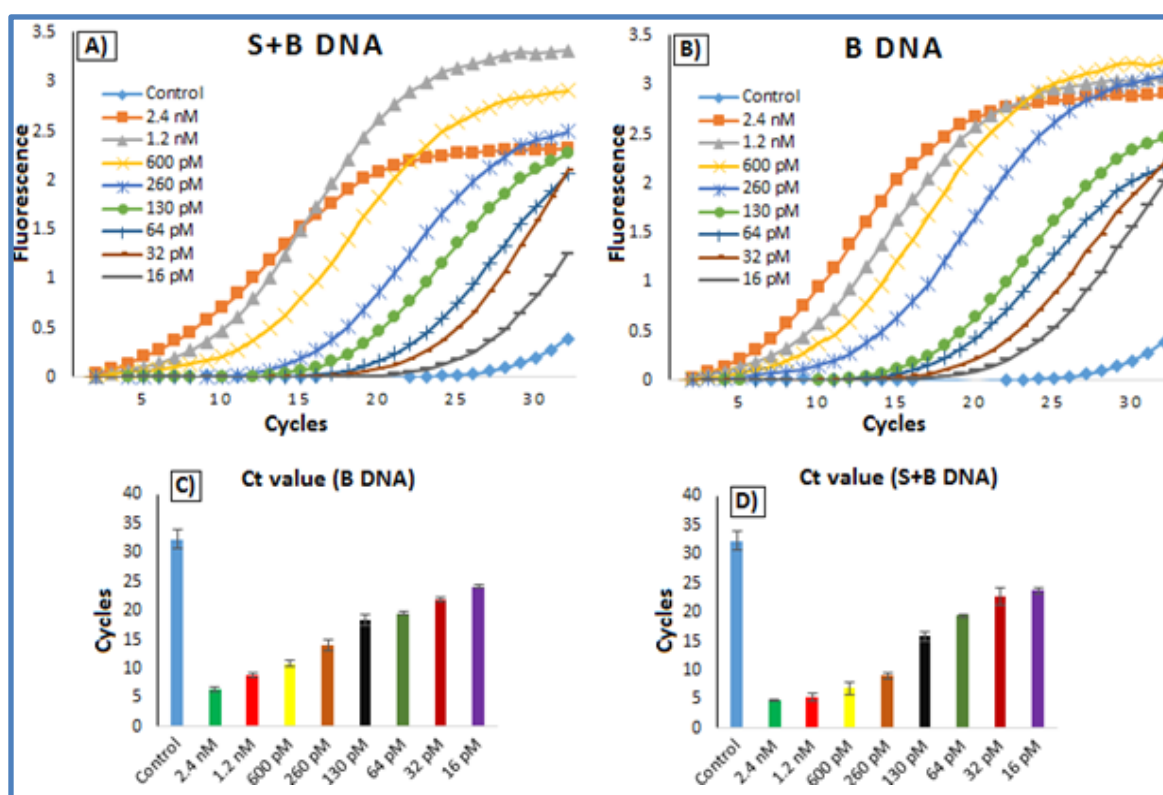


Figure 3.4 qPCR limit of detection comparison of glycan DNA conjugate and pure DNA
 A & C amplification plot and Ct value plot of glycan DNA conjugate; B & D amplification plot and Ct value plot of pure DNA

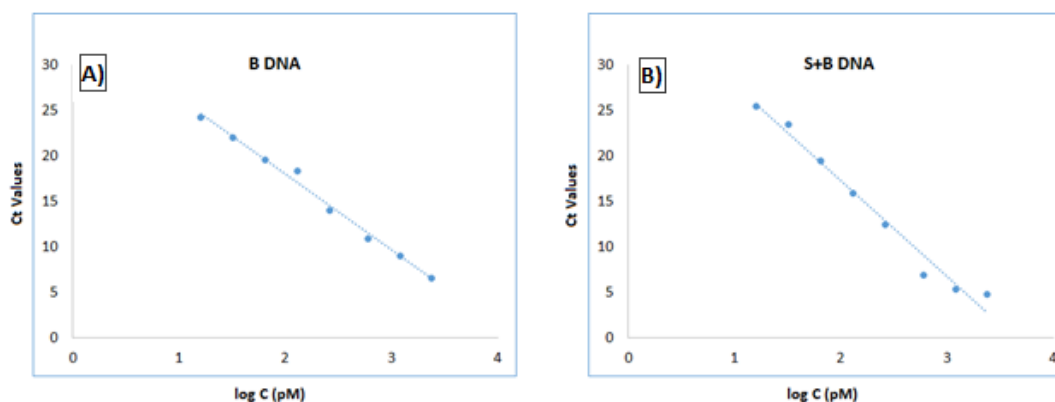


Figure 3.5 Standard Curve Plot

A) Standard Curve plot of pure DNA; B) Standard Curve plot of Glycan DNA conjugate

3.5 Blood Glycan – Antibody Sensitivity study: Filtration

Since G+DNA could be applied in both screening and detection, DEGLs enable interrogation of the glycans against target proteins in solution, while known glycan-DNA interactions could be used to detect a specific target with high sensitivity. Two set of experiments were done for validating these two aspects of DEGL; initially demonstrated the detection of specific glycan binding proteins by using the familiar blood glycan-antibody interactions. As a proof of principle study to demonstrate the significance of DNA Encoded Glycan Library (DEGL), the blood group antigen-antibody interactions were chosen.

Structures of the O (or H), A, and B antigens are well known (Figure 3.6). This system was selected for the study mainly due to two reasons, i) It is most studied and accepted among all the glycan interactions, and ii) They offer the opportunity to study both the sensitivity and specificity of the technique because of the structural similarities. Three blood group antigens A, B and O sugars were synthesized chemo-enzymatically. Synthesis of glycan-DNA conjugate was achieved via click chemistry.⁴⁴ A, B and O glycans were coupled with the corresponding 5' alkyne modified DNA (DNA code + Primers).

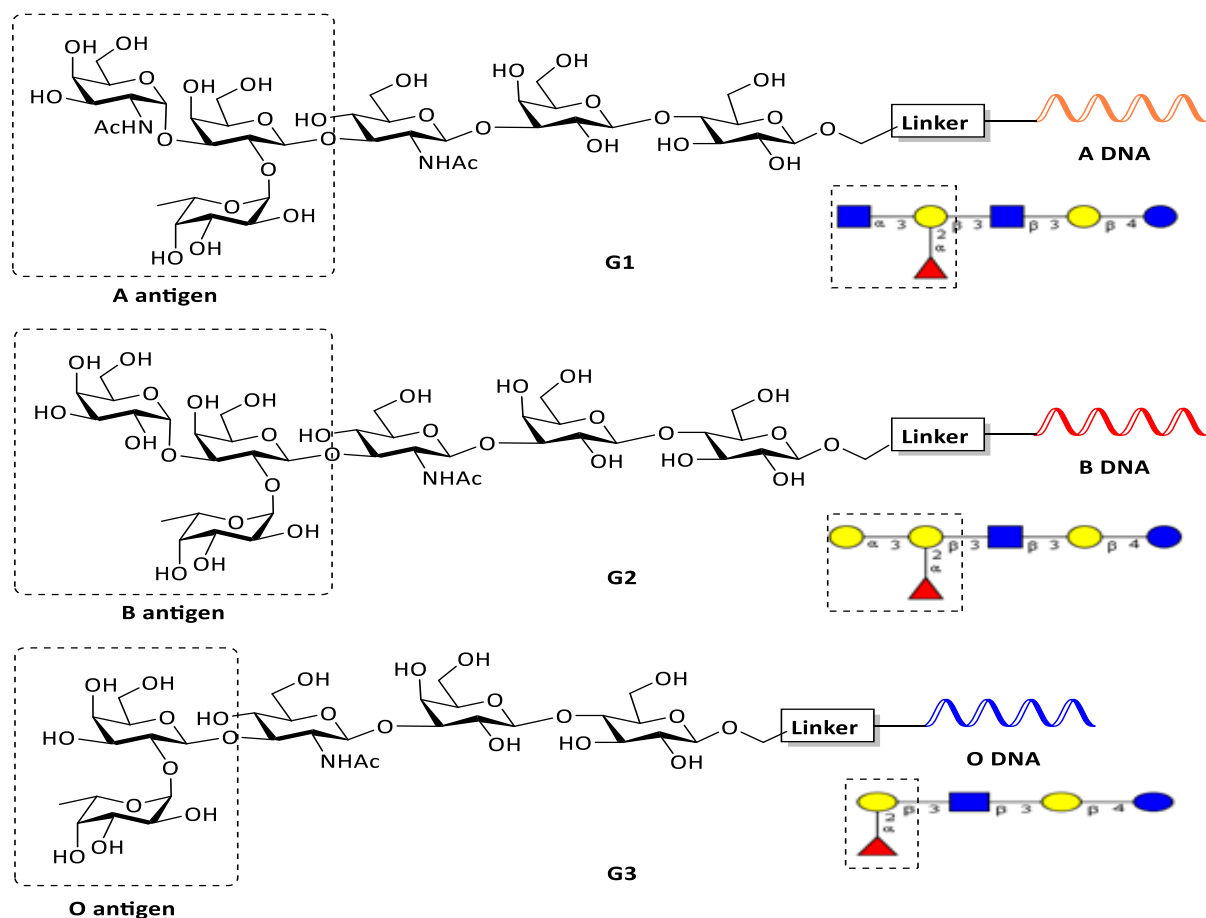


Figure 3.6 Blood Group ABO antigen- DNA conjugate structures

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, mouse). Concentrations of 5 μ M, 2.5 μ M, and 1 μ M of the G+DNA were incubated with 2 μ L of the antibodies; unbound DNA was eliminated by filtration with 100k cut off centrifuge tubes. The filtrate after incubating with the antibodies were subjected to qPCR assay, and Ct values of the G+DNA were compared with the Ct values of negative control where antibody was omitted from the incubation. A Ct difference of more than 10 points was observed at all three G+DNA concentrations indicating the requirement of low concentration of glycans for the accurate results. The filtration results were consistent with the concentration of antigens used, and were found to be more sensitive (Figure 3.7). The higher

sensitivity of the filtration method may have been due to the size difference of IgM antibodies and DNA strands (900 kDa and 25 kDa, respectively). From the above experiment, it was proven that DEGL can be used for detection of specific glycan binding proteins by using the familiar blood glycan-antibody interactions over a range of concentration (5 μ M, 2.5 μ M, 1 μ M). With the success of above experiment, 2.5 μ M was used for further demonstration of selectivity and specificity of the glycan binding proteins by using blood glycan-antibody interactions.

Next aspect of DEGL is to screen for the glycan binding proteins and, for addressing this mixed all three antibodies (1:1:1) and interrogated this mixture with each of the G+DNA conjugates (2.5 μ M) separately.

As seen in figure 3.8a, Ct values of 8.4 and 7.7 respectively for G1+A DNA and G2+B DNA indicate 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 compared to both A and B conjugates but still less than that of the no antibody controls indicating weak interaction of antigens with the Ab-O.

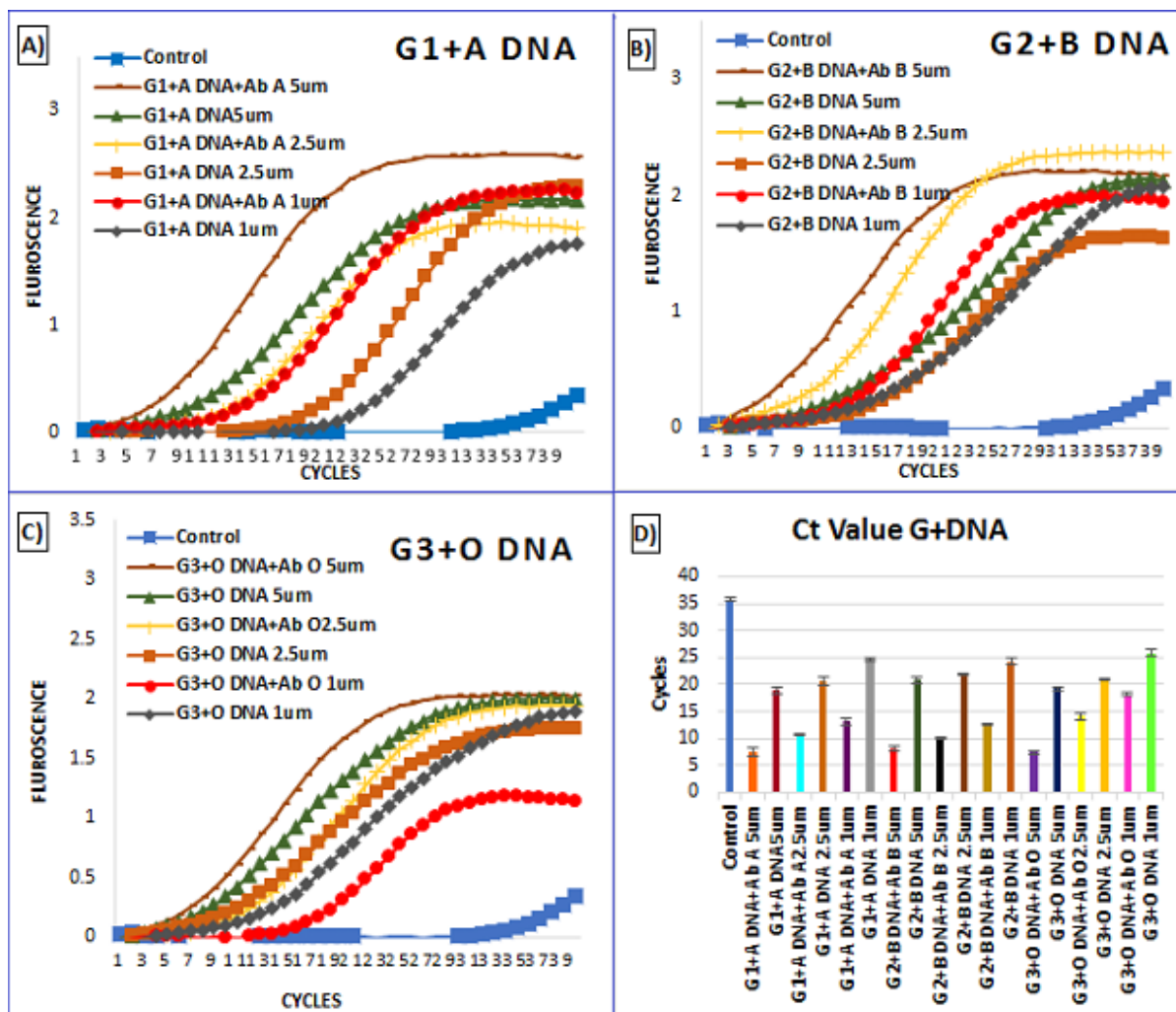


Figure 3.7 Comparison of Filtration

Sensitivity Study A) G1+A DNA against the antibody A at different concentrations (5 μM, 2.5 μM, 1 μM); B) G2+B DNA against the antibody B at different concentrations (5 μM, 2.5 μM, 1 μM); C) G3+O DNA against the antibody O at different concentrations (5 μM, 2.5 μM, 1 μM); D) Ct values plot for G+DNA at different concentrations (5 μM, 2.5 μM, 1 μM).

This result indicates that DEGL could also work with many closely related target proteins around without any loss of selectivity. The specificity of the method was also tested, for this investigated a single G+DNA (G2+B DNA) against the three antibodies, a clear amplification was visible only when both Ab B and G2+B DNA were in the mix and all the remaining tests were poorly amplified. This result clearly indicates that the amplification is indeed from the target

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

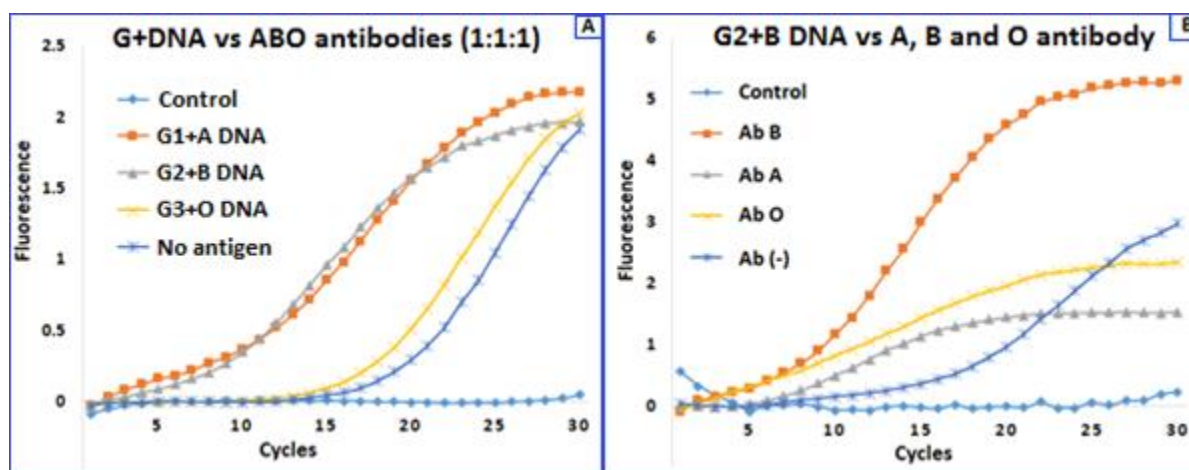


Figure 3.8 Filtration study with G+DNA conjugates and Antibody

A) G+DNA conjugates against the mixture of antibodies having equal concentration of Ab-A, Ab-B and Ab-O B) G2+B DNA against each antibody separately

3.6 Blood Glycan – Antibody Interaction Comparison: Filtration and

Immunoprecipitation

The next step was to try with natural blood serum, for this blood was collected using finger pricking as stated in blood collection (under experimental section). Since the natural blood has IgG antibodies, a comparison between filtration and immunoprecipitation was conducted to find the best method. For interrogating this aspect G+DNA at concentration of 2.5 μM was incubated with 2 μl of diluted plasma (2 μl of plasma was diluted with 8 μl water) with two different blood plasma (Blood Groups O and B). The unbound DNA was washed by both filtration and immunoprecipitation methods. The filtrate/eluent after incubating with the blood plasma (washed of unbound DNA) were subjected to qPCR assay, and Ct values of the G+DNA were compared with the Ct values of negative control where blood plasma was omitted from the incubation. A Ct difference of more than 10 points was observed at filtration results. The filtration results were consistent with the concentration of antigen used (figure 3.7), and were found to be more sensitive. The higher sensitivity of the filtration method may have been due to the size difference of IgM

antibodies and DNA strands (150 kDa and 25 kDa, respectively). A Ct difference of about 10 points was observed at immunoprecipitation results. The immunoprecipitation was consistent with the concentration of antigen used, and was found to be more accurate and specific from the amplification curve (Figure 3.10). The higher accuracy and specificity of the immunoprecipitation may have been due to the compatibility of beads with the qPCR. Ct value of 20.3 in G1+A DNA and 22.2 in G2+B DNA (Immunoprecipitation) and Ct value of 4.5 in G1+A DNA and 6.8 in G2+B DNA (Filtration) was obtained with blood group B (Figure 3.9).

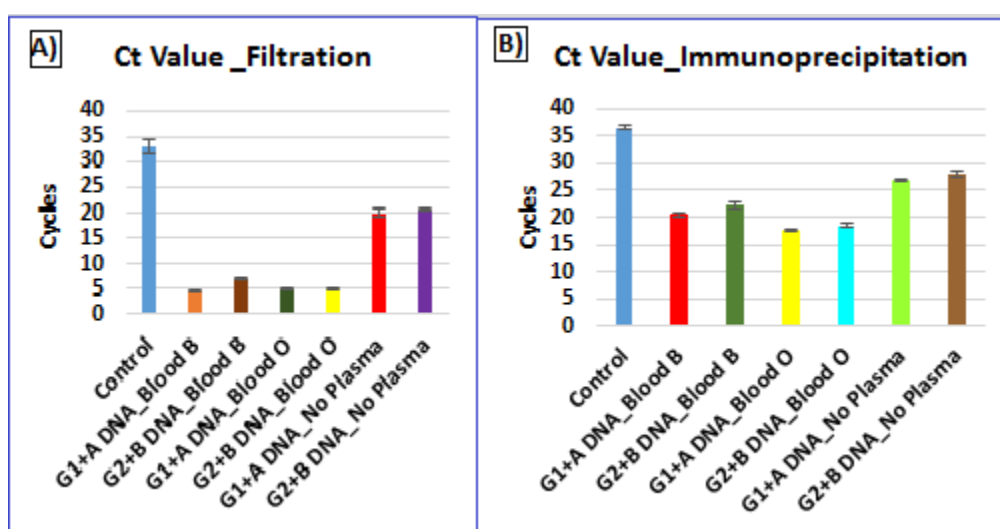


Figure 3.9 Ct value comparison

A) Ct value comparison Filtration; B) Ct value Comparison Immunoprecipitation

Similarly, Ct value of 17.6 in G1+A DNA and 18.4 in G2+ B DNA (Immunoprecipitation) and Ct value of 4.9 in G1+A DNA and 4.8 in G2+B DNA (Filtration) was obtained with blood group O. A lower Ct value in G1+ A DNA with Blood Group B and almost same Ct value in G1+A DNA and G2+B DNA with Blood Group O proved this method can be used for inverse blood group typing.

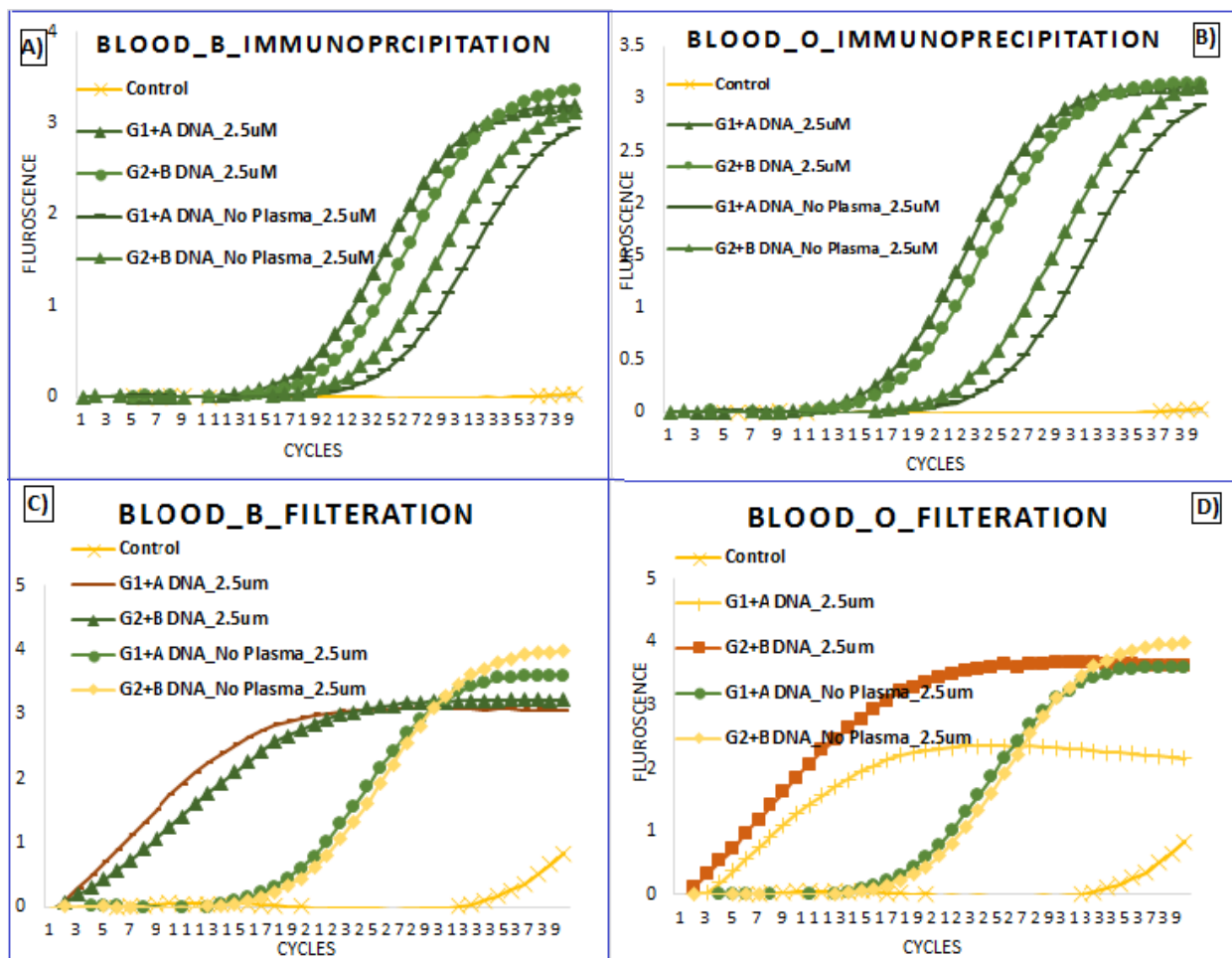


Figure 3.10 Amplification plot comparison

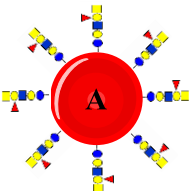
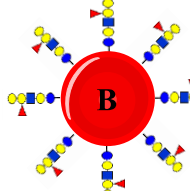
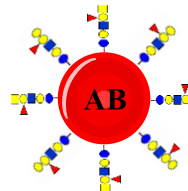
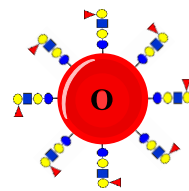



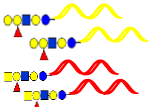
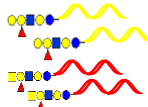
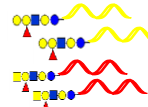
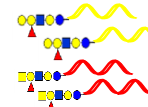




A) Amplification Plot Blood Group B Immunoprecipitation; B) Amplification Plot Blood Group O Immunoprecipitation; C) Amplification Plot Blood Group B Filtration; D) Amplification Plot Blood Group O Filtration.

3.7 DEGL: Real Life experiments – Inverse Blood Typing

From the successful results with two real blood experiments, blood typing experiments were conducted by collecting samples from healthy volunteers. Identifying the serum antibody present in the blood plasma will help to identify the blood group, person having the A antibodies alone in the blood have B antigen (group B), person with B antibodies alone are of A blood group, those who have both the antibodies belongs to the O and no antibodies in the blood indicating AB group (Table 3.2). This type of blood typing is called as inverse blood typing. However, the

conventional blood typing used widely is based on agglutination which is a direct blood typing using antigens (glycan antigen) on the surface of the RBC's.

Table 3.2 Blood Group Antigens and Antibodies

RBC of different blood group				
Antibodies in Plasma	 Ab B	 Ab A	Ab (-)	 Ab A Ab B
DEGL (G1+A DNA, G2+B DNA)				
Results (Ct values)				

Few microliters of blood were collected from 10 healthy volunteers using a finger prick, collected blood were centrifuged to separate the plasma and cellular components. The blood plasma was used to detect the blood group, 4 uL of the plasma was diluted to 20 uL and from this 2 uL each was incubated with A and B antibodies for 2 hours to allow the formation of G+DNA-Ab complexes. Further, G+DNA-Ab complexes were allowed to interact with the protein A/G magnetic beads for 1 hour, separated using a magnetic stand and washed seven times to discard any unbound G+DNA. After thoroughly washing the G+DNA-Ab-Bead complex, it was diluted to 10 uL with water and 1 uL was directly used as a template in the qPCR. Ct values of the G+DNA in the blood plasma was compared with the Ct value of G+DNA with no plasma. Among the 10 samples tested, two samples showed low Ct values for both G1+A DNA and G2+B DNA indicating the presence of both antibodies and hence belonged to O group. Five of the samples had a low Ct value for G2+B DNA means A group and three of the samples had a low Ct value for the

G1+A DNA implying B group (Figure 3.11). All the results were accorded with the conventional blood typing. The above results indicated that the DEGL can be used for inverse blood typing, which can be an alternative for conventional blood typing. However, using this method the rhesus group cannot be diagnosed as it is a protein antigen-antibody interaction.

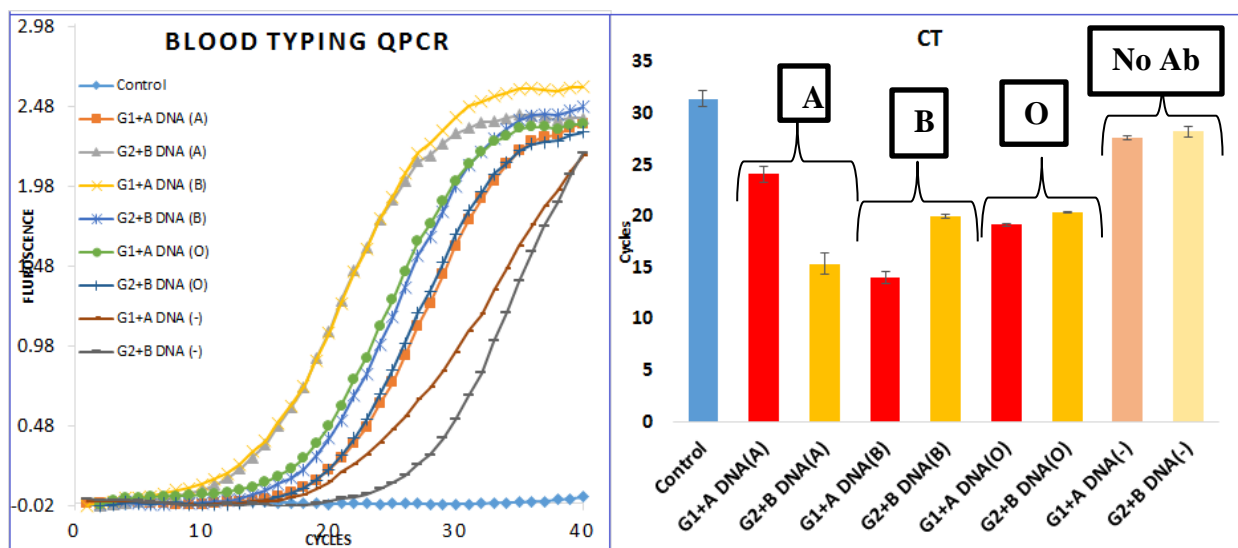


Figure 3.11 Amplification plot and Ct Value comparison of Blood Typing

3.8 Globo Series Library

After the successful completion of library synthesis and qPCR method development using blood glycans and inverse blood typing, next the concept was applied to Globo glycans. Here DEGL and glyco-PCR were used to profile the Globo-glycan binding antibodies and there by detecting the cancers expressing these antigens. Globo-glycans are part of the glycosphingolipids and their expression pattern is well correlated with the cancer metastasis and progression. Aberrant expression of Globo-glycans will also prompt the body to generate the anti-glycan antibodies targeting these epitopes. These antibodies are generally observed in the sera even before the cancer progression to the late stages and effectively act as biomarkers for certain types of cancer including breast, lung, prostate and ovarian cancer.^{7,36,45}

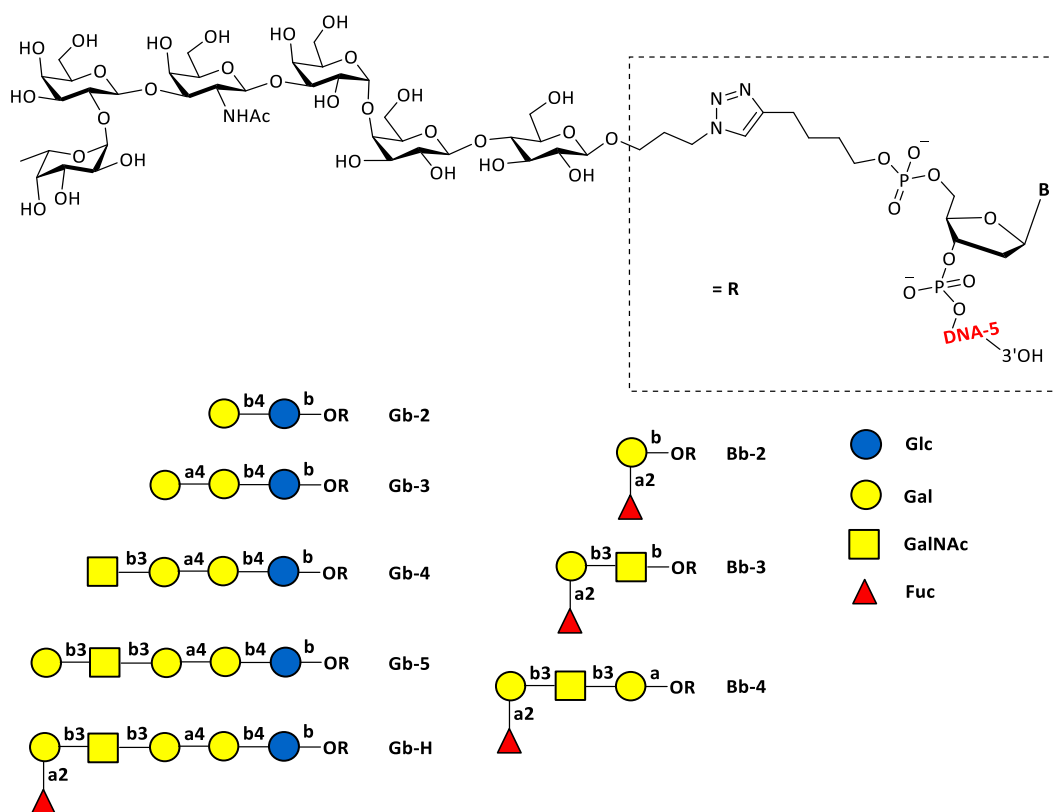


Figure 3.12 Globo glycan structures and Globo-glycan conjugates

Globo-H and its truncated analogs Gb-5, Gb-4, Gb-3, and Gb-2 were synthesized from the enzymatic extension of chemically synthesized Gb-2. To test the terminal epitope specificity alone; Bb-4, Bb-3 and Bb-2 were also synthesized via chemo-enzymatic strategy.

Initially, tested with the Globo-H glycan conjugate alone with the commercial anti Globo-H human antibody VK-9 in the same way as with the blood glycans, and the results are shown in the figure 3.13. Two concentrations of Globo-H glycan-DNA conjugate (2.5 μM and 5 μM) were incubated with 2 μL of VK-9 antibody and compared the Ct value with the no antibody control experiments. As expected, the amplification and Ct values were correlated with the concentration of the antigens present in the solution, a low Ct value of 7 and 19 was observed for 5 μM conjugate and comparatively higher Ct value of 14 and 21 was obtained for 2.5 μM conjugate.

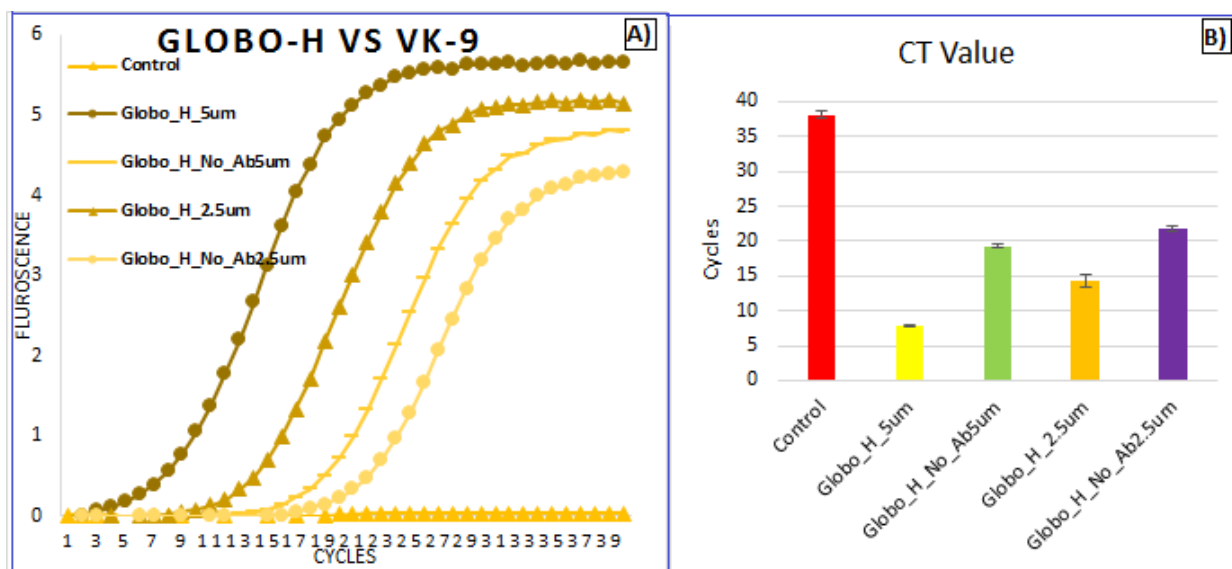


Figure 3.13 Amplification plot and Ct value Globo-H Vs VK-9 A) Amplification Plot and B) Ct Value plot

3.9 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 3.14 and 3.15).

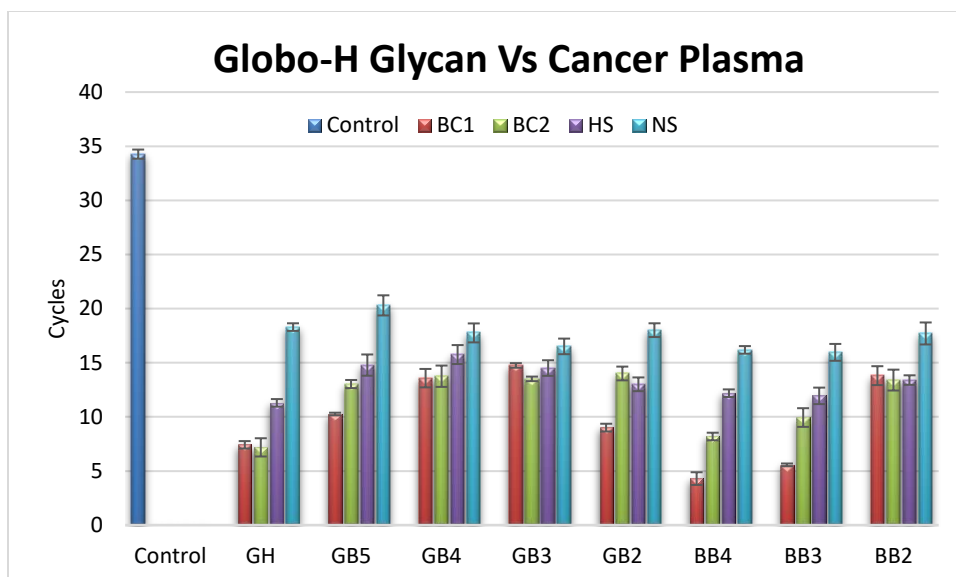


Figure 3.14 Ct Value Plot of different Globo glycan conjugate series with breast cancer serum (BC1 and BC2), Healthy Serum (HS) and No Serum (NS).

These results are consistent with the micro array based methods reported by Chi-Huey Wong group.⁷ 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 this difficulty can be easily overcome by employing next generation sequencing (NGS), which allow mixing of the glycan conjugate and mixing of plasma samples with one additional step of barcode indexing. Hence NGS technology allows the high throughput detection of hundreds of plasma samples with even thousands of glycans in the DEGL.

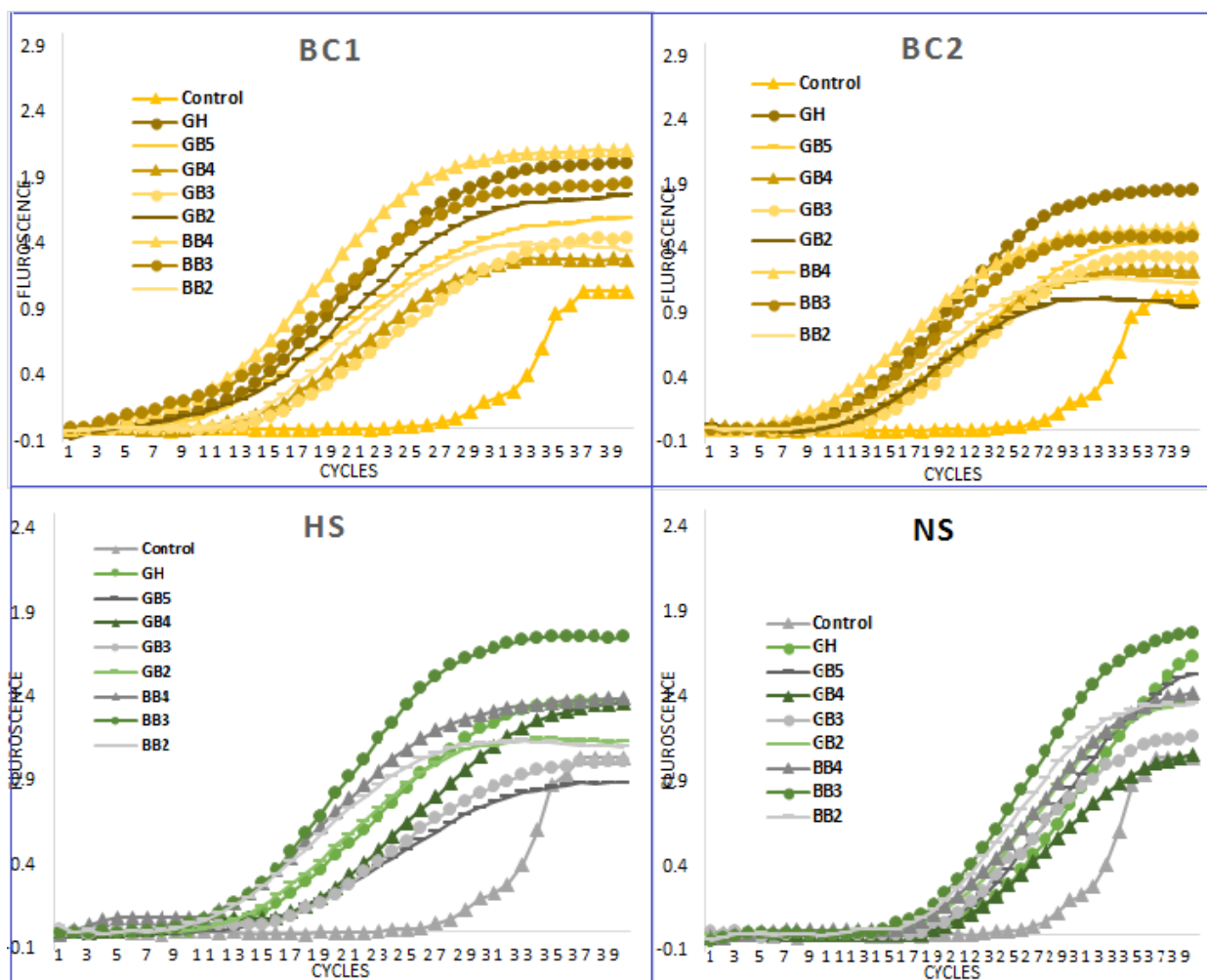


Figure 3.15 Amplification Plot of different Globo glycan conjugate series with breast cancer serum (BC1 and BC2), Healthy serum (HS) and No serum (NS)

4 CONCLUSION

With the advancement of the technologies in the DNA synthesis and sequencing have brought the enormous potential of DNA in the applications beyond genetics. DNA encoded libraries (DEL) are of the modern concept in drug screening and lead optimization. This study recognized the importance of DEL technologies in the functional glycomics (DEGL), which is otherwise limited to the low sensitive detection techniques and highly difficult to obtain samples. Applying the highly sensitive qPCR detection will enhance the reach of functional glycomics to low nano gram samples and high throughput screening. Here a unique method has been described to code each glycan, accounting their structural information very similar to how proteins coded with triplet codons. The program which is available for the use of research community is aimed to provide and keep a uniformity in the codes for every single glycan in the future DEGL technologies.

A strict library of DNA codes will not only reduce the confusion in the community but also bring down the synthetic expenses using optional bulk synthesis. Along with the unique DNA codes, it also produces an agreement on the coupling chemistry and primer selection to have the ready-made glycan codes for research use. Also identified copper assisted alkyne-azide coupling (CuAAC) chemistry, also known as 'click chemistry' for the best result in because most DNA suppliers readily offer azide and alkyne modifications on their library.

In a proof of principle study, generated a small DEGL consisting the common blood antigens and their detection in solution were demonstrated. With the same principle did a blood group analysis with the DEGL to confirm the use of DEGL in detecting the glycan binding antibodies. Although inverse blood typing can't replace the gold standard conventional direct blood typing based on the surface antigen, this study can be further extended to the next level

where the blood typing will be based on the analysis of P1PK, FORS, Se, Le and I blood glycan antigens. All the blood glycan antigens (ABO, P1PK, FORS, Se, Le and I) will form a complete blood glycan group typing. Some of them being highly rare antigens, this kind of analysis will make a place along with molecular blood typing and other rare antigen blood typing apart from the conventional ABO typing.

Finally, another DEGL consisting the Globo glycan series was generated and their detection in solution and breast cancer serum were demonstrated to confirm the use of DEGL in detecting the glycan binding antibodies which are of highly significant in the diagnostic applications. Here, Globo-glycans were analyzed in cancer serum which can be extended to several other diseases like Crohn's disease, Rheumatoid arthritis, cancers and infection where disease-specific glycan biomarkers and determinants can be found. This application of DEGL can be promising for the analysis of disease-specific glycan biomarkers and determinants for early diagnosis of the disease.

With the successful results from the proof of principle study namely inverse blood typing and detection of Globo-glycans in cancer samples, it has been proven that the DEGL will be an efficient method in future years to come. DEGL will find its own space along with other prevalent technologies available.

REFERENCES

1. Varki, A.; Sharon, N., Historical Background and Overview. In *Essentials of Glycobiology*, 2nd ed.; Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. E., Eds. Cold Spring Harbor (NY), 2009.
2. Bertozzi, C. R.; Rabuka, D., Structural Basis of Glycan Diversity. In *Essentials of Glycobiology*, 2nd ed.; Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. E., Eds. Cold Spring Harbor (NY), 2009.
3. Varki, A.; Esko, J.; Colley, K., Cellular Organization of Glycosylation. In *Essentials of Glycobiology*, 2nd ed.; Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. E., Eds. Cold Spring Harbor (NY), 2009.
4. Varki, A.; Lowe, J. B., Biological Roles of Glycans. In *Essentials of Glycobiology*, 2nd ed.; Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. E., Eds. Cold Spring Harbor (NY), 2009.
5. Pinho, S. S.; Reis, C. A., Glycosylation in cancer: mechanisms and clinical implications. *Nat Rev Cancer* 2015,15 (9), 540-55.
6. Dube, D. H.; Bertozzi, C. R., Glycans in cancer and inflammation--potential for therapeutics and diagnostics. *Nat Rev Drug Discov* 2005,4 (6), 477-88.
7. Wang, C.-C., Y.L Huang, C.-T. Ren, C.-W. Lin, J.-T. Hung, J.-C. Yu, A. L. Yu, C.-Y. Wu and C.-H. Wong (2008). "Glycan microarray of Globo H and related structures for quantitative analysis of breast cancer." *Proceedings of the National Academy of Sciences* 105(33): 11661-11666.
8. Mercey, E.; Sadir, R.; Maillart, E.; Roget, A.; Baleux, F.; Lortat-Jacob, H.; Livache, T., Polypyrrole oligosaccharide array and surface plasmon resonance imaging for the measurement of glycosaminoglycan binding interactions. *Analytical chemistry* 2008, 80 (9), 3476-3482.
9. Wang, D.; Liu, S.; Trummer, B. J.; Deng, C.; Wang, A., Carbohydrate microarrays for the recognition of cross-reactive molecular markers of microbes and host cells. *Nature biotechnology* 2002, 20 (3), 275-281.
10. Horlacher, T.; Seeberger, P. H., Carbohydrate arrays as tools for research and diagnostics. *Chemical Society reviews* 2008, 37 (7), 1414-22.

11. Wang, D.; Liu, S.; Trummer, B. J.; Deng, C.; Wang, A., Carbohydrate microarrays for the recognition of cross-reactive molecular markers of microbes and host cells. *Nature biotechnology* 2002, 20 (3), 275-281.
12. Song, X.; Heimbürg-Molinaro, J.; Smith, D. F.; Cummings, R. D., Glycan microarrays of fluorescently-tagged natural glycans. *Glycoconjugate J* 2015, 32 (7), 465-473.
13. Kornfeld, S.; Li, E.; Tabas, I., 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), 7771-7778.
14. Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Marth, J. D.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. E., Symbol nomenclature for glycan representation. *Proteomics* 2009, 9 (24), 5398-5399.
15. Zeng, X.; Qu, K.; Rehman, A., Glycosylated Conductive Polymer: A Multimodal Biointerface for Studying Carbohydrate-Protein Interactions. *Acc Chem Res* 2016, 49 (9), 1624-33.
16. Harvey, D. J.; Merry, A. H.; Royle, L.; Campbell, M. P.; Rudd, P. M., Symbol nomenclature for representing glycan structures: Extension to cover different carbohydrate types. *Proteomics* 2011, 11 (22), 4291-4295.
17. Citartan¹, M.; Tang¹, T.-H.; Tan, S.-C.; Hoe¹, C.-H.; Saini, R.; Tominaga⁴, J.; Gopinath⁴, S. C., Asymmetric PCR for good quality ssDNA generation towards DNA aptamer production. *Sonklanakarín Journal of Science and Technology* 2012, 34 (2), 125.
18. Sanchez, J. A.; Pierce, K. E.; Rice, J. E.; Wangh, L. J., Linear-After-The-Exponential (LATE)-PCR: An advanced method of asymmetric PCR and its uses in quantitative real-time analysis. *Proceedings of the National Academy of Sciences* 2004, 101 (7), 1933-1938.
19. Brenner, S.; Lerner, R. A., Encoded combinatorial chemistry. *Proceedings of the National Academy of Sciences* 1992, 89 (12), 5381-5383.
20. Litovchick, A.; Dumelin, C. E.; Habeshian, S.; Gikunju, D.; Guie, M. A.; Centrella, P.; Zhang, Y.; Sigel, E. A.; Cuzzo, J. W.; Keefe, A. D.; Clark, M. A., Encoded Library Synthesis Using Chemical Ligation and the Discovery of sEH Inhibitors from a 334-Million Member Library. *Scientific reports* 2015, 5, 10916.
21. Scheuermann, J.; Dumelin, C. E.; Melkko, S.; Neri, D., DNA-encoded chemical libraries. *Journal of biotechnology* 2006, 126 (4), 568-81.

22. Goodwin, S.; McPherson, J. D.; McCombie, W. R., Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* 2016, *17* (6), 333-351.
23. Kwon, S. J.; Lee, K. B.; Solakyildirim, K.; Masuko, S.; Ly, M.; Zhang, F. M.; Li, L. Y.; Dordick, J. S.; Linhardt, R. J., Signal Amplification by Glyco-qPCR for Ultrasensitive Detection of Carbohydrates: Applications in Glycobiology. *Angew Chem Int Edit* 2012, *51* (47), 11800-11804.
24. Kwon, S. J.; Jeong, E. J.; Yoo, Y. C.; Cai, C.; Yang, G. H.; Lee, J. C.; Dordick, J. S.; Linhardt, R. J.; Lee, K. B., High sensitivity detection of active botulinum neurotoxin by glyco-quantitative polymerase chain-reaction. *Anal Chem* 2014, *86* (5), 2279-84.
25. Baskin, J. M.; Bertozzi, C. R., Bioorthogonal click chemistry: Covalent labeling in living systems. *Qsar Comb Sci* 2007, *26* (11-12), 1211-1219.
26. Sletten, E. M.; Bertozzi, C. R., From mechanism to mouse: a tale of two bioorthogonal reactions. *Acc Chem Res* 2011, *44* (9), 666-76.
27. Saxon, E.; Bertozzi, C. R., Cell surface engineering by a modified Staudinger reaction. *Science* 2000, *287* (5460), 2007-10.
28. Kolb, H. C.; Finn, M. G.; Sharpless, K. B., Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew Chem Int Ed Engl* 2001, *40* (11), 2004-2021.
29. Jewett, J. C.; Bertozzi, C. R., Cu-free click cycloaddition reactions in chemical biology. *Chem Soc Rev* 2010, *39* (4), 1272-9.
30. Hangauer, M. J.; Bertozzi, C. R., Live cell Staudinger ligation: A FRET-based fluorogenic phosphine. *Abstr Pap Am Chem S* 2006, *232*, 767-767.
31. Lopez Aguilar, A.; Briard, J. G.; Yang, L.; Ovrzyn, B.; Macauley, M. S.; Wu, P., Tools for Studying Glycans: Recent Advances in Chemoenzymatic Glycan Labeling. *ACS Chem Biol* 2017, *12* (3), 611-621.
32. Sletten, E. M.; Bertozzi, C. R., Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. *Angew Chem Int Ed Engl* 2009, *48* (38), 6974-98.
33. McKay, C. S.; Finn, M. G., Click chemistry in complex mixtures: bioorthogonal bioconjugation. *Chem Biol* 2014, *21* (9), 1075-101.
34. Landsteiner, K., Zur Kenntnis der antifermentativen, lytischen und agglutinierenden Wirkungen des Blutserums und der Lymphe. *Zentralbl Bakteriol* 1900, *27*, 357-366.

35. Huang, C. Y.; Thayer, D. A.; Chang, A. Y.; Best, M. D.; Hoffmann, J.; Head, S.; Wong, C. H., Carbohydrate microarray for profiling the antibodies interacting with Globo H tumor antigen. *Proceedings of the National Academy of Sciences of the United States of America* 2006, *103* (1), 15-20.
36. Pochechueva, T.; Alam, S.; Schotzau, A.; Chinarev, A.; Bovin, N. V.; Hacker, N. F.; Jacob, F.; Heinzelmann-Schwarz, V., Naturally occurring anti-glycan antibodies binding to Globo H-expressing cells identify ovarian cancer patients. *Journal of ovarian research* 2017, *10* (1), 8.
37. Cheng, S. P.; Yang, P. S.; Chien, M. N.; Chen, M. J.; Lee, J. J.; Liu, C. L., Aberrant expression of tumor-associated carbohydrate antigen Globo H in thyroid carcinoma. *Journal of surgical oncology* 2016, *114* (7), 853-858.
38. Danishefsky, S. J.; Shue, Y. K.; Chang, M. N.; Wong, C. H., Development of Globo-H Cancer Vaccine. *Accounts of Chemical Research* 2015, *48* (3), 643-652.
39. Zhou, Z.; Liao, G.; Mandal, S. S.; Suryawanshi, S.; Guo, Z., A Fully Synthetic Self-Adjuvanting Globo H-Based Vaccine Elicited Strong T Cell-Mediated Antitumor Immunity. *Chemical science* 2015, *6* (12), 7112-7121.
40. O'Cearbhaill, R. E.; Ragupathi, G.; Zhu, J.; Wan, Q.; Mironov, S.; Yang, G.; Spassova, M. K.; Iasonos, A.; Kravetz, S.; Ouerfelli, O.; Spriggs, D. R.; Danishefsky, S. J.; Sabbatini, P. J., 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* 2016, *8* (4).
41. Kornfeld, S.; Li, E.; Tabas, I., 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), 7771-7778.
42. Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Marth, J. D.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. E., Symbol nomenclature for glycan representation. *Proteomics* 2009, *9* (24), 5398-5399.
43. Harvey, D. J.; Merry, A. H.; Royle, L.; Campbell, M. P.; Rudd, P. M., Symbol nomenclature for representing glycan structures: Extension to cover different carbohydrate types. *Proteomics* 2011, *11* (22), 4291-4295.

44. El-Sagheer, A. H.; Brown, T., Click chemistry with DNA. *Chemical Society reviews* 2010, 39 (4), 1388-405.
45. Chang, W. W.; Lee, C. H.; Lee, P.; Lin, J.; Hsu, C. W.; Hung, J. T.; Lin, J. J.; Yu, J. C.; Shao, L. E.; Yu, J.; Wong, C. H.; Yu, A. L., Expression of Globo H and SSEA3 in breast cancer stem cells and the involvement of fucosyl transferases 1 and 2 in Globo H synthesis. *Proceedings of the National Academy of Sciences of the United States of America* 2008, 105 (33), 11667-72.
46. Garibyan, L.; Avashia, N., Polymerase chain reaction. *J Invest Dermatol* 2013, 133 (3), e6.
47. VanGuilder, H. D.; Vrana, K. E.; Freeman, W. M., Twenty-five years of quantitative PCR for gene expression analysis. *Biotechniques* 2008, 44 (5), 619-26.
48. Postollec, F.; Falentin, H.; Pavan, S.; Combrisson, J.; Sohier, D., Recent advances in quantitative PCR (qPCR) applications in food microbiology. *Food Microbiol* 2011, 28 (5), 848-61.
49. Dean L. Blood Groups and Red Cell Antigens [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2005. Chapter 5, The ABO blood group.
50. Eller, C. H.; Chao, T. Y.; Singarapu, K. K.; Ouerfelli, O.; Yang, G. B.; Markley, J. L.; Danishefsky, S. J.; Raines, R. T., Human Cancer Antigen Globo H Is a Cell-Surface Ligand for Human Ribonuclease 1. *Acs Central Sci* 2015, 1 (4), 181-190.