DEVELOPMENT OF A LECTIN BINDING ASSAY FOR MUCIN-TYPE *O*-GLYCOSYLATED SERUM PROTEINS: IDENTIFICATION OF POTENTIAL BIOMARKERS FOR SCREENING OF BREAST CANCER

LEE CHENG SIANG

FACULTY OF MEDICINE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2017

DEVELOPMENT OF A LECTIN BINDING ASSAY FOR MUCIN-TYPE *O*-GLYCOSYLATED SERUM PROTEINS: IDENTIFICATION OF POTENTIAL BIOMARKERS FOR SCREENING OF BREAST CANCER

LEE CHENG SIANG

THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

FACULTY OF MEDICINE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2017

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: Lee Cheng Siang

Matric No: MHA110058

Name of Degree: Doctor of Philosophy

Title of Thesis ("this Work"): Development of a Lectin Binding Assay for Mucin-Type *O*-glycosylated serum proteins: Identification of Potential Biomarkers for Screening of Breast Cancer

Field of Study: Molecular Medicine

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date:

Subscribed and solemnly declared before,

Witness's Signature Date:

Name:

Designation:

ABSTRACT

Mucin-type *O*-glycosylated proteins are known to be related to many pathological diseases, especially malignancy. Mucins such as CA 27.29, CA 125 and CA 19-9 have been actively used as biomarkers for cancer detection and monitoring. Complex biological samples like serum contain a myriad of proteins with vast concentration differences. The broad dynamic range of the proteins is likely to cause masking of the mucin-type *O*-glycosylated proteins which are of low abundances, and hence, their failure of being identified by any typical detection method. Further, a reliable detection assay is currently unavailable to measure total mucin-type *O*-glycosylated proteins of biological samples, which may contain one or more of these macromolecules of unknown structures. Therefore, the present study was generally aimed at the development of a microassay to estimate mucin-type *O*-glycosylated proteins that is based on binding with plant lectins. In the present study, a microassay was developed based on binding of plant lectins to the typical GalNAc-Ser/Thr structural feature of mucin-type *O*-glycans. Since the sugar-amino acid linkage in the mucin-type *O*-glycosylated protein is invariably cryptic, the heterogeneous peripheral and core saccharides of model glycoconjugates were chemically removed before optimising conditions for using an enzyme-linked lectin assay (ELLA). Desialylation of the model glycoconjugates led to maximal binding of the lectins but additional treatments such as Smith degradation did not result in increased binding. Of the lectins tested for their ability to probe the desialylated *O*-glycans, jacalin showed the highest sensitivity followed by champedak galactose binding (CGB) lectin and *Vicia villosa* agglutinin (VVA). Further improvement in the sensitivity of ELLA was achieved by using microtiter plates that were pre-coated with the CGB lectin, which increased the specificity of the assay to mucin-type *O*-glycosylated proteins. Finally, the applicability

of the developed sandwich ELLA to crude samples was demonstrated by estimating trace quantities of the mucin-type *O*-glycosylated proteins in the human serum. In the second part of the study, perchloric acid was used to enhance detection of mucin-type *O*-glycosylated proteins in human serum. Sensitivity and specificity of the earlier developed sandwich enzyme-linked lectin assay were significantly improved with the use of serum perchloric acid isolates. When a pilot case-control study was performed using the serum perchloric acid isolates of normal participants (n = 105) and patients with stage 0 (n = 31) and stage I (n = 48) breast cancer, higher levels of total mucin-type *O*-glycosylated proteins in sera of both groups of early stage breast cancer patients compared to the normal control women were demonstrated. Analysis of CGB lectin bound serum perchloric acid isolates proteins found proteoglycan 4 to vary inversely with plasma protease C1 inhibitor in both the early stages of breast cancer patients compared to the controls. The data suggest that the ratio of serum proteoglycan 4 to protease C1 inhibitor may be used for screening of early breast cancer although this requires further validation in clinically representative populations.

ABSTRAK

Protein berglikosilat-*O* jenis musin diketahui berkait dengan pelbagai penyakit patologi, terutamanya kanser. Musin seperti CA 27.29, CA 125 dan CA 19-9 kerap digunakan sebagai penanda biologi untuk mengesan dan memantau kanser. Sampel biologi kompleks seperti serum mengandungi pelbagai protein dengan kepekatan yang amat berbeza-beza. Julat dinamik protein yang luas mempunyai kemungkinan tinggi untuk melindungi kewujudan protein berglikosilat-*O* jenis musin yang rendah timbunannya, dan oleh itu, menyebabkan kegagalannya untuk dikenali oleh sebarang kaedah pengesanan yang lazim. Selain itu, asai pengesan untuk mengukur jumlah protein terglikosilat-*O* jenis musin di dalam sampel biologi, yang mungkin mengandungi satu atau lebih makromolekul dengan struktur yang tidak diketahui, tidak wujud. Oleh itu, kajian ini secara umumnya bertujuan untuk menghasilkan mikroasai yang dapat menganggar protein berglikosilat-*O* jenis musin, berasaskan kaedah pengikatan dengan lektin tumbuhan. Dalam kajian ini, satu kaedah mikroasai telah dihasilkan berasaskan pengikatan lektin tumbuhan dengan struktur *O*-glikan jenis musin yang lazim, GalNAc-Ser/Thr. Oleh kerana pengikatan gula-asid amino pada protein berglikosilat-*O* jenis musin adalah kurang seragam, pelbagai sakarida periferal dan teras pada model glikokonjugat telah disingkirkan secara kimia sebelum interaksi mereka diuji menggunakan asai enzim berangkai lektin (ELLA). Penyahsialan model glikokonjugat menyebabkan pengikatan maksimum oleh lektin, tetapi rawatan tambahan seperti degradasi Smith tidak meningkatkan ikatan. Di antara lektin yang diuji untuk kebolehan sebagai prob bagi glikan-*O* tanpa asid sialik, jacalin menunjukkan sensitiviti yang paling tinggi, diikuti dengan lektin pengikat galaktosa cempedak (CGB) dan *Vicia villosa* aglutinin (VVA). Penambahbaikan sensitiviti ELLA yang seterusnya diperoleh dengan menggunakan plat mikrotiter yang diselaputi lektin CGB, yang menambahkan

kekhususan asai ini terhadap protein berglikosilat-*O* jenis musin. Akhir sekali, penggunaan ELLA sandwic yang dihasilkan ke atas sampel mentah telah ditunjukkan dengan menganggar kuantiti kecil protein berglikosilat-*O* jenis musin di dalam serum manusia. Dalam bahagian kedua kajian ini, asid perklorik telah digunakan untuk mempertingkatkan pengesanan protein berglikosilat-*O* jenis musin di dalam serum bagi mencari penanda biologi baru untuk pesakit kanser payudara peringkat awal. Tahap sensitiviti dan kekhususan asai sandwic yang dihasilkan pada bahagian awal kajian ini didapati meningkat dengan signifikan dengan penggunaan isolat asid perklorik serum. Apabila kajian perintis kes kawalan dilaksanakan dengan menggunakan isolat asid perklorik serum dari subjek normal (n=105) dan pesakit kanser payudara tahap 0 (n=31) dan tahap 1 (n=48), paras protein berglikosilat-*O* jenis musin di dalam serum adalah lebih tinggi bagi kedua-dua kumpulan pesakit kanser berbanding wanita kawalan normal. Analisis protein isolat asid perklorik yang berikat lektin CGB menunjukkan perubahan songsang timbunan proteoglikan 4 dan perencat plasma protease C1 yang signifikan dalam kedua-dua pesakit kanser payudara tahap awal berbanding kawalan. Data ini mencadangkan bahawa nisbah proteoglikan 4 serum kepada perencat plasma protease C1 boleh digunakan untuk menyaring kes-kes kanser payudara peringkat awal, walaupun ianya memerlukan pengesahsahihan seterusnya dalam populasi wakilan klinikal.

ACKNOWLEDGEMENTS

In all these years, further study in research had never been easier. There are times where I have experienced countless emotional break down due to my work and life that I am struggling to go through. Fortunately, I have met my beloved wife, **Lily Tan Cheng Pei**, who always support me and provide advice so that I would never give up on my work easily. Because of her, I have managed to go through all this and I am truly grateful.

This thesis would not be possible without the help from my supervisor, **Prof. Dr. Onn Haji Hashim**, who willing to spend his time and patient on guiding me throughout the entire process of this study. His knowledgeable input has inspired me to explore new ideas, which greatly diversified the research project. For these, I am truly thankful.

I wish to thank my second supervisor, **Assoc. Prof. Dr. Puteri Shafinaz Abdul Rahman**, for her warm encouragement and support during my stay in the laboratory.

I would also like to thank my ex-supervisor, **Prof. Emeritus Veer Bhavanandan**, for introduce me to this field of research, which inspire me to pursue further up until this level.

I like to express my appreciation to **Prof. Dr. Nur Aishah Mohd Taib** and all her supporting staff members for providing me the clinical samples.

My special thanks to all lab mates (Ann, Aizat, Thanes, Izlina, Esther, Ramarao, Jessie, Jaime, Mardiaty and Daniel Lim) and closest department friends (Dennis Gan, Kong, Ryan and Francis) for all your support and accompany especially during lunch hour.

I like to express my appreciation to all the staff members and postgraduate students of Department of Molecular Medicine for direct or indirect contribution on both administrative and laboratory work.

This research has made successful by the contribution from UMRG (RG210/10HTM) and HIR-MoHE (UM.0000069.HIR.C1) research grants.

Last but not least, I owe my deepest gratitude to all my family members and friends for their moral support and motivation that keeps me going throughout the entire study.

TABLE OF CONTENTS

Abstract iii

Abstrak v

Acknowledgements vii

Table of Contents viii

List of Figures xiii

List of Tables xv

List of Symbols and Abbreviations xvi

List of Appendices xx

CHAPTER 1: INTRODUCTION 1

1.1 Study objectives 5

CHAPTER 2: LITERATURE REVIEW 6

2.1 Protein glycosylation 6

2.1.1 Mucins are *O*-glycosylated proteins 6

2.1.2 Structure and biosynthesis of *O*-glycans 10

2.1.3 *O*-linked glycans and diseases 12

2.1.4 Analysis of *O*-glycans 16

2.2 Lectins 17

2.2.1 Plant lectins 19

2.2.2 Applications of plant lectins 21

2.2.3 Champedak galactose binding lectins 23

2.2.4 Jacalin 26

2.2.5 *Vicia villosa* agglutinin 28

2.3 Breast cancer 29

2.3.1 Types of breast cancer 31

2.3.2 Diagnosis and classification 32

2.3.3 Mucins and breast cancer 36

CHAPTER 3: MATERIALS AND METHODS 38

3.1 Materials 38

3.2 Samples 43

3.2.1 Test glycoproteins 43

3.2.2 Human blood samples 43

3.2.3 Serum samples preparation 44

3.3 Purification of CGB lectin 46

3.3.1 Preparation of crude extracts of champedak seeds 46

3.3.2 Preparation of galactose affinity column 48

3.3.3 Isolation of CGB lectin 49

3.4 Quality assessment of isolated CGB lectin 50

3.4.1 Quantitation of CGB lectin 50

3.4.2 SDS-PAGE 50

3.4.3 Silver staining 54

3.4.4 MALDI-TOF/TOF intact protein analysis 56

3.4.5 Lectin radial diffusion 56

3.5 Development of *O*-glycosylated protein assay 58

3.5.1 Modification of *O*-glycans 58

3.5.2 Analysis of monosaccharide of native and modified glycoproteins 59

3.5.3 Protein biotinylation 60

3.5.4 Microtiter plate lectin binding assay 60

3.6 Enrichment of serum glycoproteins using perchloric acid 62
3.6.1 Perchloric acid treatment of serum samples 62
3.6.2 Two-dimensional gel electrophoresis 63
3.6.3 In-gel trypsin digestion 68
3.6.4 MALDI-TOF/TOF analysis 70
3.6.5 2-DE CGB lectin blotting 71
3.6.6 In-solution trypsin digestion and mass spectrometry 73
3.6.7 Biotinylated-asialoBSM spiking experiment 73
3.7 Profiling of *O*-glycosylated protein in breast cancer serum perchloric acid isolates.. 74
3.7.1 Sepharose conjugation of CGB lectin 74
3.7.2 Sepharose-CGB lectin precipitation 76
3.7.3 Western blot 77
3.8 Statistical Analysis 77

CHAPTER 4: RESULTS 78

4.1 Development of assay to detect *O*-glycosylated proteins 78
4.1.1 Purification of CGB lectin 78
4.1.2 CGB lectin purity 80
4.1.3 Binding properties of CGB lectin 82
4.1.4 Concentrations of monosaccharides in PGM after chemical treatment ... 85
4.1.5 Lectin binding of native and chemically-treated glycoproteins 85
4.1.6 Direct versus sandwich ELLA 90
4.1.7 Effect of *N*-glycosylated protein on sandwich ELLA for *O*-glycosylated protein 91
4.1.8 Detection of *O*-glycosylated proteins in serum samples 94
4.1.9 Assay reproducibility 97

4.2 Enrichment of *O*-glycosylated proteins using perchloric acid 100

4.2.1 Perchloric acid treatment of human serum sample..... 100

protein recovered from serum after 0.6 N of perchloric acid treatment was calculated as approximately 1%. 102

4.2.2 Proteomic analysis of serum perchloric acid isolates..... 102

4.2.3 LC/MS analysis of proteins in serum perchloric acid isolates 106

4.2.4 Protein glycosylation of serum perchloric acid isolates..... 106

4.2.5 CGB lectin blotting analysis..... 109

4.2.6 Determination of total *O*-glycosylated proteins in serum perchloric acid isolates 109

4.2.7 Recovery of *O*-glycosylated proteins from serum after perchloric acid treatment analysis 112

4.3 Application of developed assay on early stages of breast cancer..... 112

4.3.1 Analysis of *O*-glycosylated proteins in sera of patients with stage 0 and stage I breast cancer..... 114

4.3.2 Effect of perchloric acid treatment on performance of sandwich ELLA 114

4.3.3 CGB lectin precipitation of serum perchloric acid isolates..... 116

4.3.4 Interaction of breast cancer serum perchloric acid isolates with CGB lectin 119

4.3.5 Identification of perchloric acid isolated proteins of altered abundance 124

4.3.6 Ratio of proteoglycan 4 to plasma protease C1 inhibitor..... 124

4.3.7 Western blot analysis..... 128

CHAPTER 5: DISCUSSION 130

5.1 Isolation of CGB lectin..... 130

5.2 Development of *O*-glycosylated protein assay..... 133

5.3 Application of sandwich ELLA on normal human serum samples..... 137

5.4 Enrichment of serum *O*-glycosylated proteins using perchloric acid 140

5.5 Levels of *O*-glycosylated proteins in early breast cancer 146

CHAPTER 6: CONCLUSION 153

References 155

List of Publications and Papers Presented 177

Appendix 181

LIST OF FIGURES

Figure 2.1: Examples of *O*- and *N*-glycans. 7

Figure 2.2: Typical structural examples of transmembrane and secreted mucins. 9

Figure 2.3: Common core structures of *O*-glycans. 11

Figure 2.4: Biosynthesis of core 1-4 structures of *O*-glycans. 13

Figure 2.5: 3D crystal structure of CGB lectin. 25

Figure 2.6: 3D crystal structure of Jacalin. 27

Figure 2.7: 3D crystal structure of VVA. 30

Figure 3.1: Template of wells for the lectin radial diffusion method. 57

Figure 4.1: Isolation of CGB lectin using galactose-conjugated Sepharose column. 79

Figure 4.2: Analysis of CGB lectin, jacalin and VVA by SDS-PAGE. 81

Figure 4.3: Linear mode MALDI-TOF analysis of CGB lectin. 83

Figure 4.4: Radial immunodiffusion assay using CGB lectin with test glycoproteins. 84

Figure 4.5: Binding of serially diluted biotinylated-jacalin to various glycoconjugates. 87

Figure 4.6: Binding of serially diluted biotinylated-CGB lectin to various glycoconjugates. 88

Figure 4.7: Binding of serially diluted biotinylated-VVA to various glycoconjugates. 89

Figure 4.8: Development of ELLA using microtiter plates precoated with different lectins. 92

Figure 4.9: Influence of *N*-glycosylated proteins on the sandwich ELLA for *O*-glycosylated proteins. 93

Figure 4.10: Effect of mild acid hydrolysis on test serum sample. 95

Figure 4.11: Development of ELLA using microtiter plates precoated with different lectins on serum. 96

Figure 4.12: Estimation of *O*-glycosylated protein in human serum samples subjected to mild acid treatment. 98

Figure 4.13: Concentrations of protein in serum treated with different normalities of perchloric acid. 101

Figure 4.14: Typical 2-DE profiles of serum perchloric acid isolates and neat serum. 103

Figure 4.15: Summary of proteins detected in serum perchloric acid isolates using 2-DE and QTOF-LC/MS. 108

Figure 4.16: CGB lectin blot of pooled serum perchloric acid isolates. 110

Figure 4.17: Levels of *O*-glycosylated proteins in serum perchloric acid isolates. 111

Figure 4.18: Recovery analysis of biotinylated-asialoBSM after perchloric acid treatment on serum. 113

Figure 4.19: Estimation of total *O*-glycosylated protein in serum samples of cancer negative women (control) and patients with early stages of breast cancer. 115

Figure 4.20: Sensitivity and specificity of sandwich ELLA to detect early breast cancer. 117

Figure 4.21: Optimization of CGB lectin precipitation on serum perchloric acid isolates. 118

Figure 4.22: Release of *O*-glycosylated proteins from CGB lectin beads with and without heat treatment. 120

Figure 4.23: Unspecific binding of serum perchloric acid isolates with CGB-Sepharose beads. 121

Figure 4.24: Profiles of CGB lectin bound proteins of cancer negative women and breast cancer patients. 122

Figure 4.25: Levels and ratio of CGB lectin bound proteins in cancer negative women and breast cancer patients 126

Figure 4.26: Ratio of percentage volume of proteoglycan 4 to plasma protease C1 inhibitor. 127

Figure 4.27: Western blot analysis of protease C1 inhibitor on samples of early breast cancer. 129

Figure 5.1: Concept of *O*-glycan modification using chemical methods 135

Figure 5.2: CGB lectin-Jacalin sandwich ELLA 139

LIST OF TABLES

Table 2.1: List of commercially available lectins 20

Table 2.2: Characteristics of stages of breast cancer 34

Table 3.1: Demographics of breast cancer patients and healthy participants. 45

Table 3.2: Gel preparation for SDS-PAGE. 53

Table 3.3: Gel preparation for 8 – 18% gradient gel. 67

Table 4.1: Carbohydrate composition of pig gastric mucin (PGM) after different chemical treatments to modify the saccharide chains. 86

Table 4.2: Reproducibility of the developed sandwich ELLA 99

Table 4.3: Proteins from 2-Dimentional Electrophoresis identified by mass spectrometry analysis 104

Table 4.4: QTOF LC/MS detected proteins in serum perchloric acid isolate. 107

Table 4.5: Intensity of peptide bands from CGB lectin precipitation of serum perchloric acid isolates. 123

Table 4.6: Protein identity of CGB lectin isolated proteins of altered abundance. 125

LIST OF SYMBOLS AND ABBREVIATIONS

2-DE	:	2-Dimensional gel electrophoresis
A1AG	:	Alpha-1-acid glycoprotein
A1AT	:	Alpha-1-antitrypsin
$A_{280\ nm}$	:	Absorbance at 280 nm
$A_{415\ nm}$	:	Absorbance at 415 nm
AACT	:	Alpha-1-antichymotrypsin
ALBU	:	Serum albumin
AMBP	:	Alpha-1-microglobulin/bikunin precursor
ANR42	:	Ankyrin repeat domain-containing protein 42
ANT3	:	Antithrombin III
APOA1	:	Apolipoprotein A-I
APOA4	:	Apolipoprotein A-IV
APOH	:	Beta-2-glycoprotein 1
APS	:	Ammonium persulphate
Asn	:	Asparagine
BCA	:	Bicinchoninic acid
BSM	:	Bovine submaxialliary mucin
CA	:	Cancer antigen
Ca	:	Calcium
CBG	:	Corticosteroid-binding globulin
CGB	:	Champedak galactose binding

CID	:	Collision-induced dissociation
CLUS	:	Clusterin
CMB	:	Champedak mannose binding
CO4A/CO4B	:	Complement C4-A or Complement C4-B
CXCL7	:	Platelet basic protein
DCIS	:	Ductal carcinomas *in situ*
ddH_2O	:	Double distilled water
DTT	:	Dithiothreitol
DVS	:	Divinyl Sulfone
ECD	:	Electron capture dissociation
EDTA	:	Ethylenediaminetetraacetic acid
EGF	:	Epidermal-growth-factor
ELLA	:	Enzyme-linked lectin assay
ER	:	Estrogen receptors
ESI	:	Electrospray ionisation
ETD	:	Electron transfer dissociation
FA5	:	Coagulation factor V
FETUA	:	Alpha-2-HS glycoprotein
Fuc	:	Fucose
Gal	:	Galactose
GalNAc	:	*N*-acetylgalactosamine
GCP2	:	Gamma-tubulin complex component 2

GlcNAc	:	*N*-acetylglucosamine
HEMO	:	Hemopexin
HER2	:	Human epidermal growth factor receptor 2
HPT	:	Haptoglobin
IC1	:	Plasma protease C1 inhibitor
Ig	:	Immunoglobulin
IGKC	:	Immunoglobulin kappa chain C region
IPG	:	Immobiline pH gradient
ITIH4	:	Inter-α-trypsin inhibitor heavy chain H4
JAC	:	Jacalin
Lac	:	Lactose
LCIS	:	Lobular carcinomas *In Situ*
Lys	:	Lysine
MALDI	:	Matrix-assisted laser desorption/ionisation
Neu5Ac	:	*N*-acetylneuraminic acid
PBS	:	Phosphate buffered saline
PCA	:	Perchloric acid
PGM	:	Pig gastric mucin
ppGalNAct	:	Polypeptide-*N*-Acetylgalactosaminyltransferase
pI	:	Isoelectric focusing point
PR	:	Progesterone receptors
PSM	:	Pig submaxialliary mucin

PTGDS	:	Prostaglandin-H2 D-isomerase
PTS	:	Proline, threonine, and serine
QTOF LC/MS	:	Quadrupole time of flight liquid chromatography mass spectrometry
SDS-PAGE	:	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SEA	:	Sperm-protein-enterokinase-agrin
Ser	:	Serine
TEMED	:	Tetramethylethylenediamine
Thr	:	Threonine
THRB	:	Prothrombin
TOF	:	Time of flight
TTHY	:	Transthyretin
UDP	:	Uridine diphosphate
UMMC	:	University of Malaya Medical Centre
UV	:	Ultraviolet
VVA	:	*Vicia villosa* agglutinin
VWF	:	Von Willebrand factor
IgHG	:	Immunoglobulin gamma heavy chain
IgLC	:	Immunoglobulin light chain

LIST OF APPENDICES

Appendix A: Ethical approval from Medical Ethics Committee University of Malaya Medical Centre…….. 181

Appendix B: Solid ammonium sulphate required for precipitation…….. 182

CHAPTER 1: INTRODUCTION

Glycobiology involves studies of the structures, biosynthesis, and biology of saccharides (sugar chains) that are widely distributed in nature (Varki & Sharon, 2009). Sugar chains or glycans are often conjugated to non-carbohydrates such as lipids and proteins to form glycolipids and glycoproteins, respectively. The process of glycan conjugation is also termed as glycosylation. It is the most common post-translational modification of proteins. Sugars that are conjugated to Ser/Thr or Asn residues of protein polypeptides are referred to as *O*- or *N*-glycans, respectively. The *O*-glycosylated proteins, also termed as mucin type glycoproteins, are mostly present at the apical surface of epithelial cells.

Mucins are one of the major constituents of mucus, which provides lubrication to the cell surface of the gastrointestinal tract and protects it against pathogens (McGuckin *et al.*, 2011). They are large glycoproteins which contain multiple tandem repeats that are rich in Ser and Thr residues which are mostly linked with *O*-glycans (*O*-GalNAc) (Argueso & Guzman-Aranguez, 2010). High levels of circulating mucins in the blood plasma may be associated with many diseases, particularly malignancy (Bafna *et al.*, 2010; Brockhausen, 2006; Storr *et al.*, 2008). Overexpression of mucins in the plasma is mostly related to their high secretion or released by membrane cell shedding (Thathiah *et al.*, 2003; Thathiah & Carson, 2004). Mucins such as MUC1 (CA15-3) (Duffy *et al.*, 2010) and MUC16 (CA125) (Bast & Spriggs, 2011) are commonly used as tumor markers for breast and ovarian cancers, respectively.

Many studies have been focused on the specific levels of glycoproteins that are secreted into the blood stream so that they can be used as potential biomarkers (Drake *et al.*, 2006; Mohamed *et al.*, 2008; Seriramalu *et al.*, 2010). Aside from changes in the levels of the glycoproteins, cancer may also be associated with altered glycosylation of

proteins. Modifications of glycan structures that are commonly reported in cancer include: increased branching of glycans, premature sialic acid termination, which causes shorter glycan structures, and over expression of certain glycan structures (Dall'olio, 1996). For example, sialylated Lewis^{a} (Neu5Acα2,3Galβ1,3(Fucα1,4)GalNac-R), which is detected by the serological assay CA19‑9, was found to be associated with pancreatic colorectal, gastric or biliary cancer, and it is now commonly used to monitor clinical response to therapy (Sipponen & Lindgren, 1986; Ugorski & Laskowska, 2002).

Most tumor markers that have been discovered are glycoproteins, and they are detected based on immunological methods. These detection methods often have narrow specificity for a single marker. However, they are still unable to efficiently diagnose cancer specifically. Information on glycan changes could also be possibly used for disease prognosis and may further improve the specificity of cancer markers. Currently, there is no reliable assay that could detect total *O*-glycosylated protein content in a bodily fluid sample, which may contain one or more *O*-glycans of unknown structures. Therefore, a simple and reliable assay to detect *O*-glycosylated proteins in biological fluids of the human body is necessary.

The aim of the present study was to develop a sensitive assay to determine the levels of *O*-glycosylated proteins in serum samples. This task was complicated by the fact that the saccharides in this class of glycoconjugates are highly heterogeneous (Roussel *et al.*, 1988; Strous & Dekker, 1992). Firstly, the saccharides in *O*-glycans are formed by chain elongation, branching and chain terminations to generate a bewildering array of structures (Bhavanandan & Furukawa, 1995). Secondly, the total *O*-glycans content of a biological sample is likely to be constituted of a mixture of molecules of unknown saccharide structures. However, despite the structural variability, *O*-glycans are also known to have eight different core structures, all these eight core structures

have a unique structural feature that distinguishes them from other classes of glycoconjugates (*N*-glycans, glycolipids and proteoglycans/glycosaminoglycans). The unique structural feature that is present in all *O*-glycans is the GalNAc residue at the reducing termini of the saccharides, which links them to specific serine or threonine residues of *O*-glycosylated proteins. In fact, the sensitivity of these GalNAc-Ser/Thr linkages to mild alkaline conditions has been very valuable in the characterization of *O*-glycans (Carlson, 1968; Edge & Spiro, 1987) and for their colorimetric estimation (Bhavanandan *et al.*, 1990). Previously, a method that was based on the specific chemistry of the GalNAc-Ser/Thr linkage had been developed for the estimation of *O*-glycans (Bhavanandan *et al.*, 1990). However, the sensitivity of this colorimetric assay was very limited. Therefore, in this study, the use of lectins to estimate trace amounts of *O*-glycosylated proteins was explored.

The strategy used in the development of the assay was firstly to maximise exposure of the cryptic sugar–amino acid linkage by removal of the heterogeneous peripheral and core saccharides. Secondly, the best lectin to quantitate the exposed linkage regions of *O*-glycans by Enzyme-Linked Lectin Assay (ELLA) was determined. Towards this objective, model glycoconjugates (PGM, BSM, PSM, and fetuin) were chemically treated to remove peripheral saccharides. The untreated and treated glycoconjugates were then tested for their interaction with three selected lectins, *i.e.*, jacalin, *Vicia villosa* agglutinin (VVA) and champedak galactose binding (CGB) lectin that are known to bind *O*-glycan moieties. Whilst jacalin and CGB lectin are known to bind to Galβ1,3GalNAcα-Ser/Thr (T-antigen) (Ahmed & Chatterjee, 1989; Hashim *et al.*, 1991), VVA has been shown to selectively interact with GalNAcαSer/Thr (Tn-antigen) structure (Tollefsen & Kornfeld, 1983b). The modifications and lectin combination that gave the highest sensitivity were then used to develop the assay and it

was then evaluated for its ability to quantitate trace quantities of *O*-glycosylated proteins in serum samples.

The assay that was developed in the present study utilises pre-coated champedak galactose binding (CGB) lectin to capture *O*-glycosylated proteins from serum samples, followed by binding of biotinylated-jacalin as a lectin probe. When used to assay for *O*-glycosylated proteins in serum samples, it faces the challenge of IgA1, which is highly abundant in the human serum. IgA1 is an *O*-glycosylated protein that strongly interacts with both the CGB lectin and jacalin (Hashim *et al.*, 1991; Jagtap & Bapat, 2010). In its presence, the immunoglobulin would occupy the majority of the binding sites of the lectins, and this may limit the lectins' interaction of other *O*-glycosylated proteins of low abundance in the serum. Therefore, in the second part of the study, perchloric acid was firstly used to enrich heavily glycosylated proteins from serum samples before being subjected to the enzyme-linked lectin assay. Perchloric acid treatment of serum is known to precipitate non- and low glycosylated proteins, leaving those heavily glycosylated proteins in the supernatant (Winzler *et al.*, 1948). In the present study, proteomics analysis was initially applied to detect proteins that were solubilised in the acid. Through these methods, the ability of perchloric acid to precipitate IgA1 from serum was able to be confirmed. Further analysis was also performed to determine the applicability of perchloric acid to enrich substantial amount of *O*-glycosylated proteins from serum samples.

In the final part of the study, total levels of *O*-glycosylated proteins in serum samples of patients with stage 0 and stage I breast cancer as well as those from controls or cancer free women was determined using the developed assay. The use of perchloric acid to improve detection of heavily *O*-glycosylated proteins in early breast cancer serum was demonstrated as well. A method to discover the binding profile of CGB lectin to *O*-glycosylated protein of early breast cancer serum perchloric acid isolates

was developed based on lectin precipitation and gel-based proteomics. By comparing the perchloric acid-enriched glycoprotein profiles of patients with pre-invasive and invasive breast cancer with those of the controls, glycoproteins of altered abundance that may be used as potential biomarkers for screening of early breast cancer were finally identified.

1.1 Study objectives

The specific objectives of the present study were:

i. To develop a sensitive assay for estimation of total *O*-glycosylated proteins in serum samples using plant lectins.

ii. To profile the *O*-glycosylated proteins that were solubilised in serum perchloric acid isolates using proteomics.

iii. To evaluate the levels of *O*-glycosylated proteins in serum samples of patients with early stages of breast cancer.

iv. To identify specific proteins of altered abundance in patients with early breast cancer.

CHAPTER 2: LITERATURE REVIEW

2.1 Protein glycosylation

Proteins are commonly modified before they can be fully functional. The most common post-translational modification of proteins, which occurs in the rough endoplasmic reticulum, is glycosylation. It is an enzymatic reaction that covalently attaches sugar molecules to an amino acid residue of a protein (Varki & Sharon, 2009). These sugar molecules, or glycans, are typically diverse and complex in nature. Glycans on proteins may exist in the form of a simple monosaccharide to a multiple branched form of complex tetra-antennary structure. *O*-glycans refer to carbohydrate moieties that are covalently attached to the hydroxyl group of serine/threonine residue via an *O*-glycosidic bond (Brockhausen *et al.*, 2009), whereas *N*-glycans are those being conjugated to asparagine residue through *N*-glycosidic bond (Stanley *et al.*, 2009). *O*-glycans and *N*-glycans do not share similarity in terms of structure. Both of the glycans have unique features and different sets of sugars. For example, *O*-glycans contain *N*-acetylgalactosamine (GalNAc) as the first sugar to be linked with serine or threonine residue, and galactose may be present most of the time (Brockhausen *et al.*, 2009). In contrast, *N*-glycans commonly have core sugar containing mannoses, with *N*-acetylglucosamine directly linked to the asparagine residue (Stanley *et al.*, 2009). Figure 2.1 illustrates typical examples of *O*-glycans and *N*-glycans.

2.1.1 Mucins are *O*-glycosylated proteins

Mucins contain multiple *O*-glycosylation sites that encompass 80% of its total molecular weight. These are large glycoproteins with molecular mass of 0.5 to 20 MDa. Mucins are mucosal proteins that are present at many epithelial surfaces of the

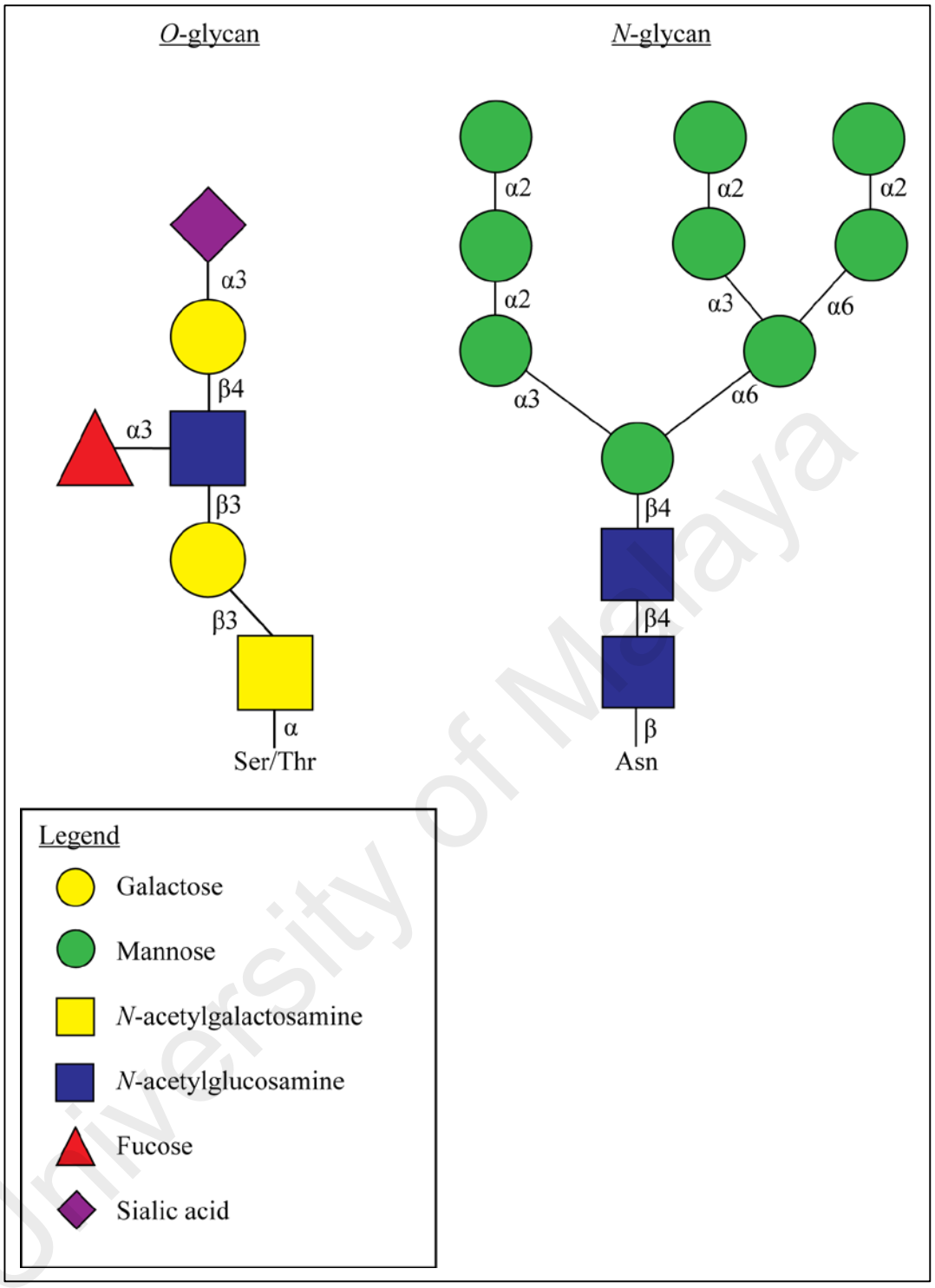


Figure 2.1: Examples of *O*- and *N*-glycans.

The common feature of an *O*-glycan is the *N*-acetylgalactosamine being the first saccharide to attach on serine or threonine residues of a polypeptide. Subsequent addition of saccharide on the *N*-acetylgalactosamine further elongates the glycan, which forms a wide heterogeneous structure. *N*-glycan contains *N*-acetylglucosamine as the first sugar to link with arginine residue and has complex branches of core mannose structure.

body, including the gastrointestinal, genitourinary, and respiratory tracts (Brockhausen *et al.*, 2009). The main role of mucins is to shield the epithelial surfaces of the tracts from physical and chemical damages as well as infections by pathogens (Brockhausen *et al.*, 2009). Extensive glycosylation of mucins are responsible for the viscosity and gel-like properties of mucus.

The genome of human mucins family has been identified and designated as MUC1 to MUC21 (Kufe, 2009). Human mucins are further classified into two different types, which are membrane bound mucins and secreted mucins. Both the membrane bound and secreted mucins contain heavily *O*-glycosylated site regions that are rich in proline (P), threonine (T) and serine (S) residues (Bansil & Turner, 2006). This region is commonly known as PTS region due to the fact that it contains three of these amino acids. In addition to the PTS region, the membrane bound mucins, such as MUC1, MUC13 and MUC16 are characterised to contain transmembrane domain, sea-urchin-sperm-protein-enterokinase-agrin (SEA) domain and/or one or two epidermal-growth-factor (EGF)-like domain (Dekker *et al.*, 2002; Kufe, 2009). On the other hand, secreted mucins (MUC2, MUC5AC, MUC5B and MUC6) contain several cysteine rich regions that are present at either side of the amino and carboxy terminals of the mucin peptide core. This region shares sequence resemblance to the C and D domains of von Willebrand factor, and the C-terminal cystine knot domains (Bansil & Turner, 2006). Minimal *N*-glycosylation sites may be found at this region, as well as the SEA domain of transmembrane mucins. Figure 2.2 illustrates the typical structure of mucins.

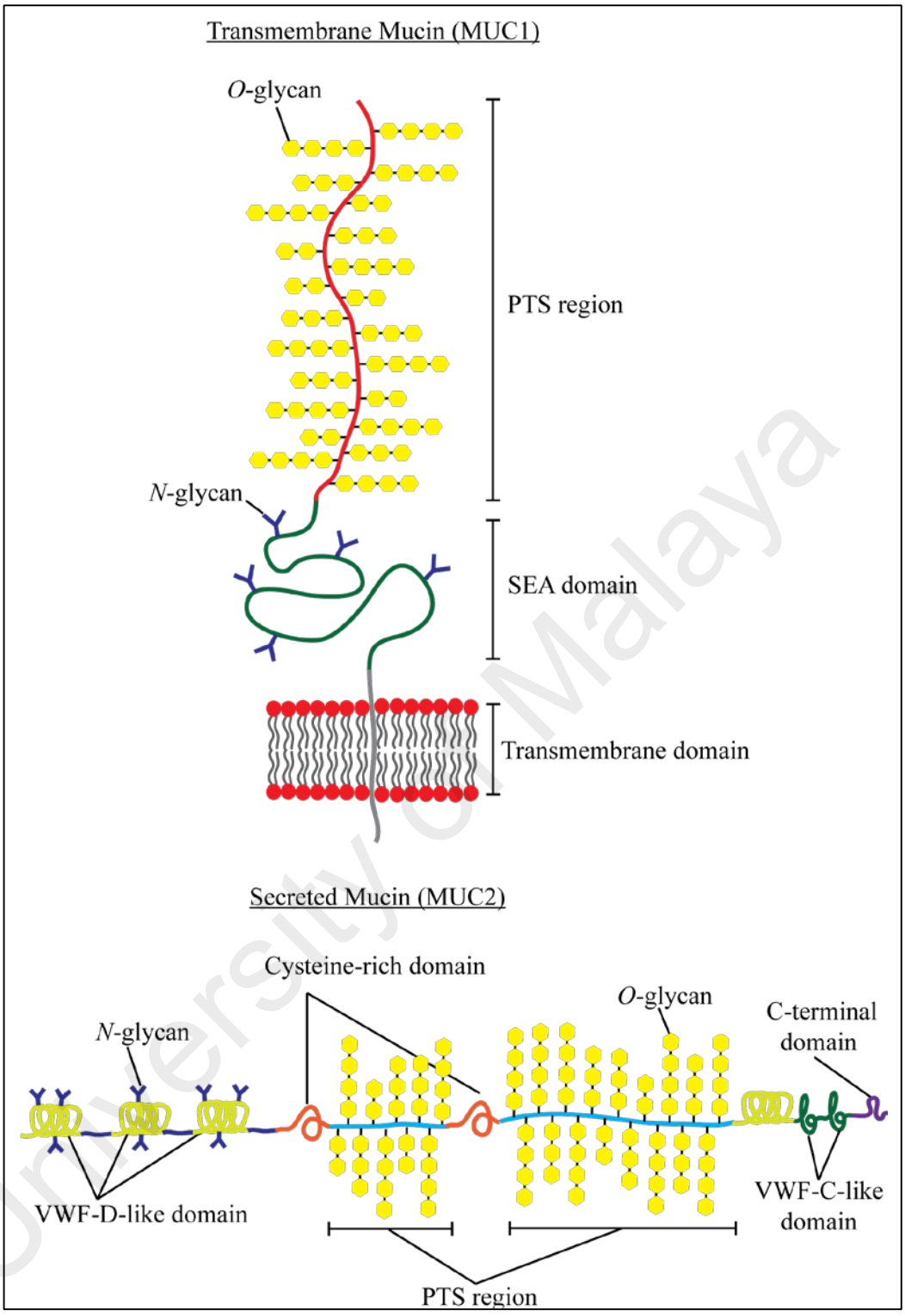


Figure 2.2: Typical structural examples of transmembrane and secreted mucins.

MUC1 and MUC2 are very well characterised transmembrane and secreted mucins, respectively. Both type of mucins share structural similarity of having PTS region that are heavily glycosylated. These mucins are distinguishable by the different structures of their domains.

2.1.2 Structure and biosynthesis of *O*-glycans

The biosynthesis of mucin type *O*-glycans begins with the transfer of a single *N*-acetylgalactosamine (GalNAc) from UDP-GalNAc to the serine or threonine residue, forming a structure of GalNAc-Ser/Thr (Brockhausen *et al.*, 2009). GalNAc-Ser/Thr may often be referred to as the Tn antigen. This transfer reaction is catalysed by a family of UDP-GalNAc:polypeptide-N-acetyl-galactosaminyltransferases (ppGalNAcTs), which are encoded by at least 21 unique genes (Ten Hagen *et al.*, 2003; Tian & Ten Hagen, 2009). The reason of having multiple enzymes for such a simple reaction is still unclear. However, there has been a suggestion that multiple enzymes are required to ensure high specificity of glycosylation on the most exact serine or threonine residue that is to be glycosylated on a polypeptide (Brockhausen *et al.*, 2009; Gerken *et al.*, 2008; Tabak, 2010). Although the specificities of the enzymes have not been elucidated, two classes of ppGalNAcTs have generally been identified. One of which catalyses the reaction only on an unglycosylated peptide, whilst the other will only react with a glycopeptide (glycosylated peptide) (Tabak, 2010).

Subsequent to the attachment of GalNAc, more sugar residues are added to the Tn structure (GalNAc-Ser/Thr), which in turn forming eight possible variable Core structures (Figure 2.3) (Brockhausen *et al.*, 2009; Tabak, 2010). Cores 1 to 4 structures are more commonly occurring in mucins of normal cells, whilst cores 5 to 8 are less commonly found in human cells. Each addition of sugar to the Tn antigen requires a series of specific glycosyltransferases. For example, the addition of galactose (Gal) to GalNAc is facilitated by the Core β1,3-galactosyltransferase, resulting in the formation of Core 1 structure or also known as the T-antigen (Galβ1,3GalNAcα1-*O*-Ser/Thr). Following this step, β1,6-*N*-acetylglucosaminlytransferase may add *N*-acetylglucosamine (GlcNAc) onto the Core 1 structure to form the Core 2 structure, Galβ1,3[GlcNAcβ1,6]GalNAcα1-*O*-Ser/Thr. Core 3 structure (GlcNAcβ1,3GalNAcα1-

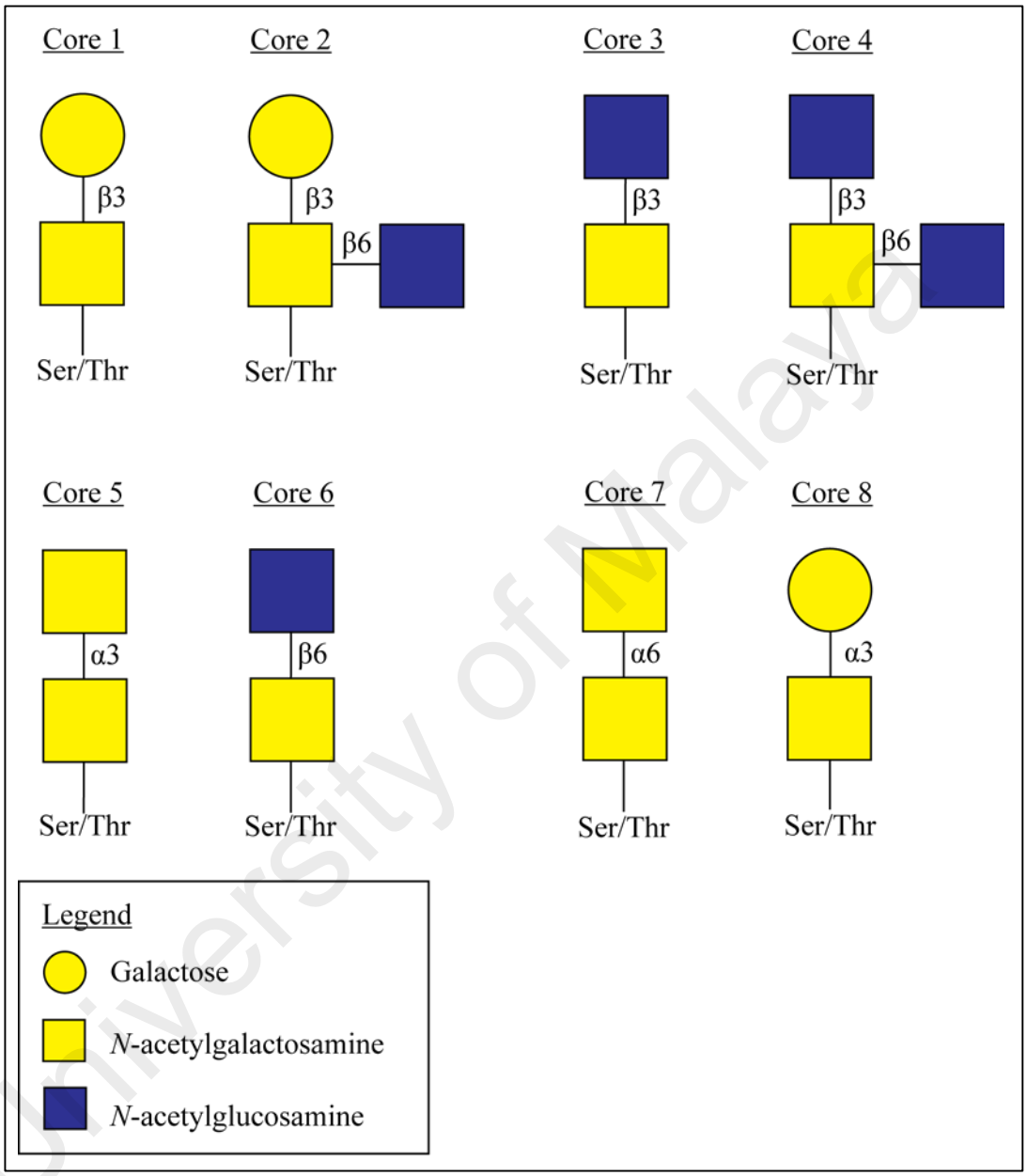


Figure 2.3: Common core structures of *O*-glycans.

Cores 1 to 4 are the most commonly found structure in mucins, whilst cores 5 to 8 are the least commonly occur glycans found in human cells. Some have suggested that these rare glycans may appear in some adenocarcinoma tissue (Brockhausen *et al.*, 2009).

O-Ser/Thr) is formed by addition of GlcNAc to the Tn antigen, and subsequent addition of another GlcNAc forms the Core 4 structure, GlcNAcβ1,3[GlcNAcβ1,6]GalNAcα1-*O*-Ser/Thr. Figure 2.4 summarises the biosynthesis of the *O*-glycans.

The *O*-glycan core structure can proceed to elongation and/or ends with termination. During the elongation process, common repeating units of Galβ1-3/4GlcNAcβ1-3 (poly-*N*-acetyllactosamines), which are also known as i-antigen, are linked to the core structure that will become the backbone of the *O*-glycans (Brockhausen, 1999). This reaction is catalysed by β4-Gal-transferase and iβ3-GlcNAc-transferase in alternate fashion. The reaction may continue further and resulting in a longer repeating unit of linear poly-*N*-acetyllactosamines [Galβ1-4GlcNAcβ1-3]$_n$. Poly-*N*-acetyllactosamines can be linked through β1-6 to form a branched glycan structure, resulting in large I-antigen structure (Brockhausen *et al.*, 2009). Less common disaccharides, such as GalNAcβ1-4GlcNAc- (LacdiNAc) and Galβ1-3GlcNAc- sequences may also be found in certain *O*-glycosylated proteins. The glycosylation process is terminated by addition of the ABO blood type antigens, Lewis antigens, sialic acid, or sulphate. Some of this termination may lead to alteration of the physiology of the cell and are often associated with development of diseases such as cancer.

2.1.3 *O*-linked glycans and diseases

Structural changes due to atypical biosynthesis of glycans or altered expression of *O*-glycosylated proteins have been associated with several diseases. *O*-glycosylation associated diseases can be developed because of either gene inheritance or acquisition (non-inherited). Mutations in the *GALNT3*, which encodes the ppGalNAcT-3 enzyme, causes inability to initiate *O*-glycosylation on a potent phosphaturic protein (FGF23)

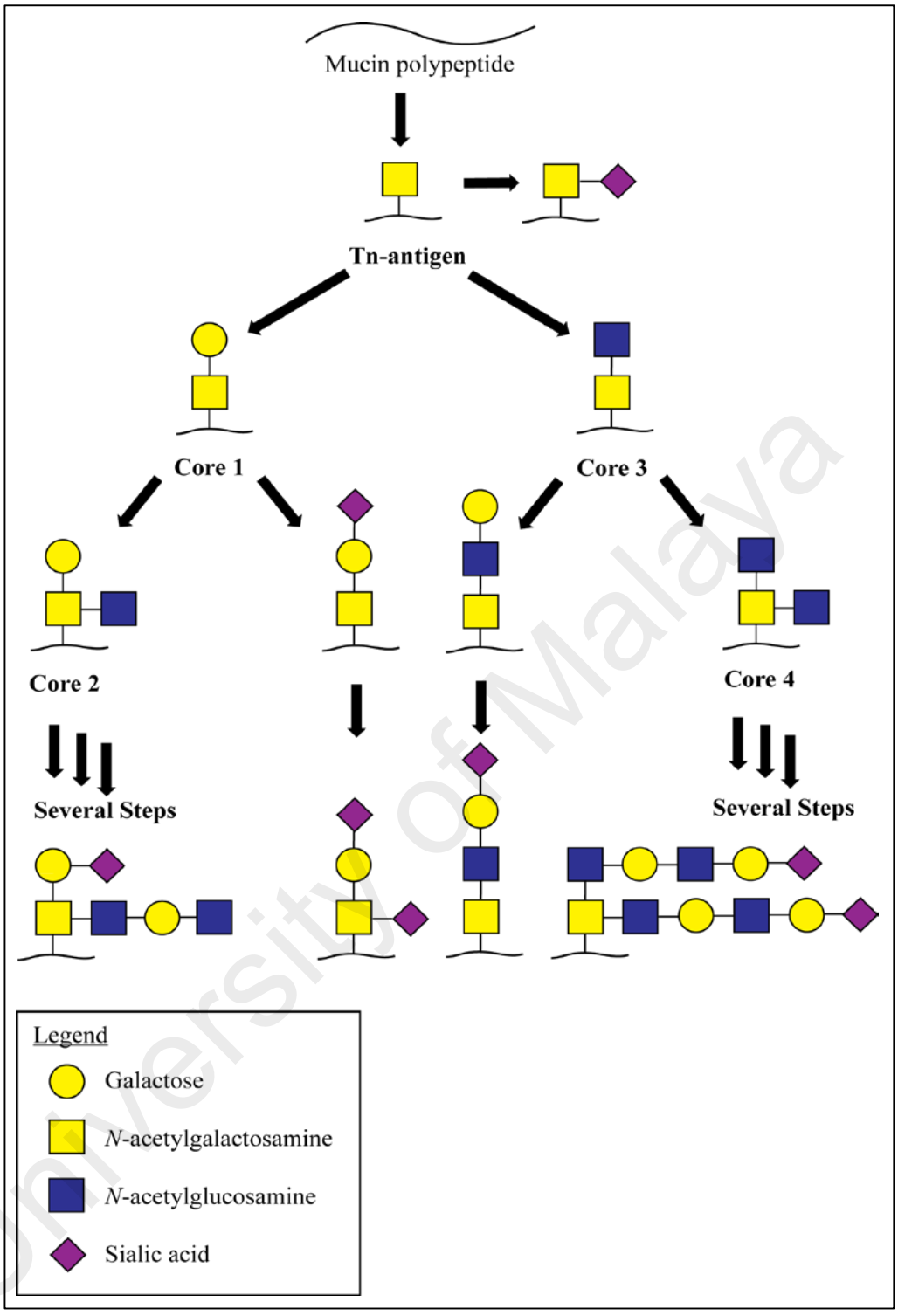


Figure 2.4: Biosynthesis of core 1-4 structures of *O*-glycans.

The formation of core 1 structure is catalysed by the Core β1,3-galactosyltransferase. Addition of GlcNAc onto the core 1 structure is catalysed by β1,6-*N*-acetylglucosaminlytransferase and results in the formation of the core 2 structure. Addition of GlcNAc to the Tn antigen forms the core 3 structure, and subsequent addition of another GlcNAc forms a core 4 structure. The process of elongation involves multiple steps and enzymes. Glycosylation is often terminated by addition of the ABO blood type antigens, Lewis antigens, sialic acid, or sulphate.

and eventually leads to familial tumoral calcinosis disease (Chefetz & Sprecher, 2009; Topaz *et al.*, 2004). This is a rare inherited disease that is characterised by hyperphosphatemia and massive calcium deposits in the skin and subcutaneous tissues. The aberrant *O*-glycosylation of the FGF23 protein causes proteolytic inactivation and hinders the function of phosphate homeostasis regulation.

Another example is IgA nephropathy, which is manifested by the deposition of IgA in the glomerulus. The kidney disease is also known to be associated with aberrant *O*-glycosylation of IgA1. In normal subjects, IgA1 is heavily *O*-glycosylated in the hinge region and contains core 1 structure with some degree of sialylation (Barratt *et al.*, 2007; Narita & Gejyo, 2008). Conversely, patients with IgA nephropathy have altered glycosylation, which causes the deficiency of galactose in the core structure of IgA1 glycans (Allen *et al.*, 1995; Tomana *et al.*, 1997). As a result, this underglycosylation is believed to be one of the causes for IgA1 being deposited in the glomerulus (Hiki *et al.*, 1999).

Many studies have claimed that the altered expression of mucins is closely related to cancer (Bhavanandan, 1991; Brockhausen, 1999). As an example, the expression of MUC2, MUC4 and MUC5 were found to be significantly elevated in pancreas tissues of patients with pancreatic cancer (Balague *et al.*, 1994). Up until now, efforts to discover the relationship of mucins with cancer are still actively on-going. The association of over-expression of MUC1 mucin and under-expression of MUC4 mucin with the urinary bladder cancer has also been suggested in a recent study (Kaur *et al.*, 2014).

Progression of cancer is also associated with changes in the *O*-glycan structures. Early termination by sialic acid on Tn antigen (GalNAc-Ser/Thr), which is mediated by ST6GalNAc 1 (sialyltransferase), prevents further elongation of glycans

and causes the glycan structure to be truncated (Julien *et al.*, 2012). Sialyl-Tn antigen was been found to be over-expressed in breast (Leivonen *et al.*, 2001), colon (Itzkowitz *et al.*, 1989), endometrium (Inoue *et al.*, 1991), oesophagus (Ikeda *et al.*, 1993), ovary (Numa *et al.*, 1995), pancreas (Kim *et al.*, 2002) and stomach (David *et al.*, 1996) cancer. The over-expression of sialyl-Tn antigen may be attributed to a mutation in the Cosmc gene that encodes a chaperone that is essential for the expression of β1-3 galactosyltransferase (Ju *et al.*, 2008). Consequently, the interrupted expression of β1-3 galactosyltransferase causes the inability of forming the core 1 structure, leaving only the Tn antigen structure. The truncated glycan is likely involved in the progression of tumour cells that promotes metastasis (Hauselmann & Borsig, 2014).

Another example of altered glycosylation of *O*-glycans in cancer is the over-expression of sialyl Lewis antigen (Varki *et al.*, 2009b). Lewis antigen is known to be highly expressed in cancer cells (Christiansen *et al.*, 2014). One of the typical example is the sialyl Lewis^{a} antigen, which has been widely used as a tumour marker for the management of pancreatic cancer (Tempero *et al.*, 1987). In some tissues, particularly in gastrointestinal tissues, disialyl Lewis^{a} is commonly found in normal cells. However, in malignant cells, the expression of disialyl Lewis^{a} tend to be significantly reduced, whilst the sialyl Lewis^{a} is highly expressed (Itai *et al.*, 1988; Itai *et al.*, 1991). The consequences of losing a single sialic acid on the Lewis^{a} antigen are diminishing cell to cell recognition between the mucosal epithelial cells and lymphoid cells and promotion of E-selectin binding activity (Miyazaki *et al.*, 2004). Tumour angiogenesis was found to be closely associated with E-selectin interaction (Tei *et al.*, 2002).

2.1.4 Analysis of *O*-glycans

Glycosylation has been discovered ever since the 1870s (Varki & Sharon, 2009). The analytical techniques of glycoconjugates have greatly evolved since then. Glycosylated protein in complex biological sample is commonly determined using 2-dimensional gel electrophoresis (2-DE) followed by protein transfer and lectin detection (Jayapalan *et al.*, 2012; Jayapalan *et al.*, 2013; Mohamed *et al.*, 2008). This technique is able to determine the present of glycosylation on proteins but not the glycan structure and the specific site of glycosylation. To study the specific site of glycosylation and the *O*-glycan structures of glycosylated proteins, more advance and complicated techniques have to be applied. This involves multiple steps, including purification of glycosylated proteins from a complex sample, glycan release from peptides, several steps of enzymatic reactions and mass spectrometry detection.

Intact *O*-glycans are usually released from peptides using chemical treatment. A reliable enzyme for removal of *O*-glycans from peptides is currently unavailable. In contrast, removal of *N*-glycans is commonly performed using an *N*-glycosidase rather than a chemical method due to the fact that enzymatic approach preserves the overall structure of both the glycans and the peptides (Novotny *et al.*, 2013). Successful analysis of *O*-glycans is very dependent on their quality. Therefore, release of *O*-glycans from peptides, prior to any analysis, still remains a challenge. One of the most popular chemical techniques of releasing *O*-glycans is the reductive β-elimination (Carlson, 1968). This technique involves incubation of glycoproteins in 0.05 M sodium hydroxide and 1.0 M sodium borohydride at 45°C for 16 – 18 hours. The role of sodium hydroxide is to create an alkaline condition that cleaves *O*-glycans from glycoproteins. Sodium borohydride immediately reduces the terminal sugar to form an alditol that prevents further degradation by the alkali (Zauner *et al.*, 2012). However, the reduction restricts labelling on the terminal sugar by fluorophore or chromophore for

fluorescence or UV detection, respectively. Therefore, glycans released by reductive β-elimination are only able to be analysed using mass spectrometry or high pH anion exchange chromatography (HPAEC) coupled with pulsed amperometric detection (Kotani & Takasaki, 1997; Lloyd *et al.*, 1996).

Mass spectrometry system such as matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) are most commonly used instruments to predict the structure of glycans (Zauner *et al.*, 2012). These systems are often coupled with other mass spectrometric fragmentation systems, such as collision-induced dissociation (CID), electron capture dissociation (ECD) or electron transfer dissociation (ETD), which greatly enhance the potential of glycan analysis (Pitteri *et al.*, 2005; Zubarev *et al.*, 2000). The fragmentation technique is able to determine the specific site of glycosidic linkage between saccharides. Although structural determination of glycans using mass spectrometry is very reliable, the technique is limited to structural prediction based on molecular size. Hexoses that share similar molecular weights but differ in stereoisomeric structures cannot be differentiated by mass spectrometry. To date, several software such as GlycoWorkbench (Damerell *et al.*, 2015) and Glyco-peakfinder (Maass *et al.*, 2007) as well as glycan database (Campbell *et al.*, 2008) have been developed to assist with the prediction of glycan structures.

2.2 Lectins

Lectins are carbohydrate binding proteins, which are found ubiquitously in nature. The earliest lectin discovered was ricin, which was extracted from the seeds of castor bean by Peter Hermann Stillmark in 1888 (Varki *et al.*, 2009a). Stillmark demonstrated that the castor bean seeds extract was capable of agglutinating red blood cells from different animals (Rudiger & Gabius, 2001; Varki *et al.*, 2009a). Due to the

fact that lectins can agglutinate red blood cells, the carbohydrate binding protein is used to be recognised as agglutinin. Later, since it was known that not all carbohydrate binding proteins possess the ability to agglutinate red blood cells, the term lectin was used instead. The word "lectin" is derived from the Latin word "legere", meaning "to select" or "to choose" (Sharon & Lis, 2004). The selectivity of carbohydrate to be recognised by a lectin can be highly specific. For example, concanavalin A specifically binds to α-D-mannosyl and α-D-glucosyl groups (Van Damme *et al.*, 1998), whilst peanut agglutinin specifically recognises Gal-β(1-3)-GalNAc (Van Damme *et al.*, 1998). However, the interaction between a lectin and a single sugar molecule is rather weak with the dissociation constant ranging from micromolar to milimolar (Bouckaert *et al.*, 2005; Rabinovich *et al.*, 2007). As a result, lectins are often multivalent in order to raise their binding affinity (Cummings & Esko, 2009).

Initially, lectins are often classified based on which glycan structure that the protein binds. With the emergence of molecular cloning technology, the classification of lectin becomes more consistent and diverse. At present, lectins can be classified based on the proteins structure homology and evolutionary relatedness (Varki *et al.*, 2009a). The highly conserved amino acid sequences of carbohydrate recognition domain on lectins are often targeted to characterise lectins into related families or subfamilies (Ghazarian *et al.*, 2011). For example, the C-type lectins are one of the most abundant lectins found in animals and require Ca^{2+} ion for carbohydrate recognition (Ghazarian *et al.*, 2011). Galectins (formerly known as S-type lectins) do not require Ca^{2+} ion to function but free thiols are necessary for their stability (Varki *et al.*, 2009a). On the other hand, I-type lectins share similar structures with immunoglobulins but they are not considered as antibodies (Angata & Brinkman-Van der Linden, 2002; Varki & Crocker, 2009). Siglecs belong to the I-type lectins families but they specifically recognise sialic acids (Angata & Brinkman-Van der Linden, 2002).

Other families of lectins are Calnexin, Eglectins, Eel agglutinins, Hyaluronan-binding proteins, Ficolins, Pentraxins, L-type, P-type, and R-type lectins (Varki *et al.*, 2009a).

2.2.1 Plant lectins

As mentioned earlier, ricin was the earliest lectin discovered from a highly toxic plant known as *Ricinus communis* (Varki *et al.*, 2009a). Paradoxically, research on lectin only took off several decades after ricin's discovery when James B. Sumner successfully purified and crystallised an agglutinin from Jack bean (*Canavalia ensiformis*) in 1919. The lectin, which is now known as concavanalin A (Sharon & Lis, 2004), was also used for the first time to demonstrate the binding of lectins with sugar (Sumner & Howell, 1936). Since then, lectins are known to interact with carbohydrates and as a result, this led to further understanding of the carbohydrate structures of blood type antigens (Sharon & Lis, 2004). Table 2.1 shows a list of commercially available lectins that are most commonly used.

Lectins may be found in various types of plants but most abundantly detected in Leguminoseae and Euphorbiaceae (Dias *et al.*, 2015). Although lectins are found in various tissues of plants such as fruit, seed, root, latex, bark and flowers, the actual physiological role of these lectins are currently unclear (Dias *et al.*, 2015). Some evidences have suggested that lectins play an important role in plant defence against the invasion of virus, bacteria, fungi or even eukaryote organism (Dias *et al.*, 2015; Peumans & Van Damme, 1995). Apart from being involved in plant immunity, lectins are believed to mediate symbiosis relationship between plants and microorganisms (De Hoff *et al.*, 2009). Some have also suggested that lectins may be involved in plant cell regulatory and signalling pathways (Chen *et al.*, 2002).

Table 2.1: List of commercially available lectins

No.	Nomenclature	Source	Carbohydrate specificity
1	LTL	*Lotus tetragonolobus*	Fucα1-3(Galβ1-4)GlcNAc, Fucα1-2Galβ1-4GlcNAc
2	PSA	*Pisum sativum*	Fucα1-6GlcNAc, α-D-Glc, α-D-Man
3	LCA	*Lens culinaris*	Fucα1-6GlcNAc, α-D-Glc, α-D-Man
4	MAL	*Maackia amurensis*	Siaα2-3Galβ1-4GlcNAc
5	SNA	*Sambucus nigra*	Siaα2-6Gal/GalNAc
6	SSA	*Sambucus sieboldiana*	Siaα2-6Gal/GalNAc
7	PHA-L	*Phaseolus vulgaris*	Tri/tetra-antennary complex-type *N*-glycan
8	RCA120	*Ricinus communis*	Galβ1-4GlcNAc
9	NPA	*Narcissus pseudonarcissus*	High-Man, Manα1-6Man
10	ConA	*Canavalia ensiformis*	High-Man, Manα1-6(Manα1-3)Man
11	GNA	*Galanthus nivalis*	High-Man, Manα1-3Man
12	Jacalin	*Artocarpus integrifolia*	Galβ1-3GalNAc, GalNAc
13	PNA	*Arachis hypogaea*	Galβ1-3GalNAc
14	VVA	*Vicia villosa*	α-Linked terminal GalNAc, GalNAcα1-3Gal
15	WGA	*Triticum aestivum*	Chitin oligomers, NeuAc

Data obtained from (Hirabayashi *et al.*, 2011)

Based on the protein structure, plant lectins are divided into four classes, namely merolectins, hololectins, chimerolectins and superlectins (Macedo *et al.*, 2015; Peumans & Van Damme, 1995; Van Damme *et al.*, 1998). Merolectins are the simplest class of lectins, which contains only a single carbohydrate binding domain. Lectins in this class are small and made of a single polypeptide. Because of the single binding site, these lectins do not have agglutinin activity. Hololectins, in contrast, are capable of agglutinating cells as they are exclusively built-up of two or more similar carbohydrate binding domains, and most plant lectins fall under this class. Similar to the hololectins is an independent class of lectins known as superlectins, which consist of multiple carbohydrate binding domains. However, the domains of superlectins recognise structurally unrelated sugars. On the other hand, chimerolectins possess a carbohydrate binding domain and additional domain that act as the catalytic domain or provides other biological activities. A more detailed characterisation of lectins, which was based on the carbohydrate binding domain specificity and structure homology, has been proposed by Vann Damme and co-workers in 1998 (Van Damme *et al.*, 1998). To date, there are a total of 12 distinctive lectin families, which include *Agaricus bisporus* agglutinin homologues, amaranthins, class V chitinase homologues with lectin activity, the cyanovirin family, EEA family, GNA family, proteins with hevein domains, jacalins, proteins with legume lectin domains, LysM domain, nictaba family (formerly cucurbitaceae phloem lectins) and the ricin-B family (Van Damme *et al.*, 1998).

2.2.2 Applications of plant lectins

Due to the fact that lectins possess glycan binding properties, these proteins have become an important tool for research in the fields of medicine and basic sciences. Blood typing is one of the typical examples of lectin application. It was first

demonstrated by William C. Boyd and Karl O., independently, in 1940s (Sharon & Lis, 2004). These researchers found that crude extracts from the lima bean (*Phaseolus limensis*) and tufted vetch (*Vicia cracca*) agglutinate blood type A but not blood type B and O cells. However, extracts from the asparagus pea were discovered to interact with red cells of blood type O. Years later, the extract from *Griffonia simplicifolia* was shown to have exclusive interaction with erythrocytes type B (Khan *et al.*, 2002) and lectins with higher specificity to the blood type A_1 rather than the blood type A_2 were discovered in the extract of *Dolichos biflorus* (Sharon & Lis, 2004).

Lectins have been used in histochemistry technique to detect glycosylation changes at the cellular level. Barkhordari *et al.* (2004) had used a total of 27 lectins to screen normal lung cells and further proposed that a few of the lectins used were able to specifically characterise different types of lung cells. In a more recent study, increased binding of lectins was observed on higher grades of colon carcinoma tissues (Hagerbaumer *et al.*, 2015). These findings suggest that the glycosylation changes in colon tissue may have been involved in cancer cells progression. However, the application of lectins in histochemistry usually comes with the disadvantage of the limited information about the proteins they bind (Tang *et al.*, 2015).

Glycoconjugates can be detected in biological samples using a technique known as enzyme linked lectin assay (ELLA) (Gornik & Lauc, 2007). In this method, lectins are conjugated to enzymes and allowed to interact with the samples containing glycoconjugates. The levels of lectins bound correspond to the intensity of the substrate, which the enzyme catalysed. This method is simple to perform and robust. A more advance technique that utilises lectin binding to detect levels of glycoconjugates in samples is known as lectin array (Hirabayashi *et al.*, 2011). In this technique, multiple lectins are immobilised onto a chip and samples that were labelled with chromophore or fluorophore were allowed to interact with the lectins (Hirabayashi *et al.*, 2011; Hu &

Wong, 2009). Lectin array allows multiplexing lectins detection on a single sample, whereas standard ELLA method only allows single lectin detection at a time. The use of lectin array has been demonstrated on multiple types of cancer which generally showed different characteristic glycosylation in distinctive cancers (Huang *et al.*, 2014; Li *et al.*, 2013; Nakajima *et al.*, 2015; Nishijima *et al.*, 2012).

Another common use of lectins is to enrich glycoproteins from complex samples prior to an analysis, particularly those in proteomics. With the used of immobilised lectins in a typical affinity chromatography separation, profiling of *O*- or *N*-glycoproteome of biological samples are made possible (Jayapalan *et al.*, 2012). Enriched glycoproteins are then subjected to proteomic analysis and mass spectrometry to determine the glycoproteins profile. Alternatively, the lectin may also be conjugated to magnetic beads for easier isolation (Ahn *et al.*, 2014).

2.2.3 Champedak galactose binding lectins

Artocarpus integer, widely known as champedak, is a species of a fruit tree in the family Moraceae, which also share the same genus as jackfruit (*i.e. Artocarpus integrifolia*). Champedak is most commonly found in Southeast Asian region. The seeds of champedak are rich in lectins, which were firstly isolated in the early 90s (Hashim *et al.*, 1991). The extract from the champedak seeds was demonstrated to be able to agglutinate erythrocytes of blood groups A, B, AB and O and also shown to be able to precipitate IgA1 but not IgA2, IgD, IgG and IgM in the same study (Hashim *et al.*, 1991). In addition, several galactose derivatives were found to inhibit the binding of the champedak lectins. These findings suggested that the binding motif of champedak lectin was towards galactose as well as the T-antigen. Several years later, a mannose binding lectin which is mitogenic to T cell proliferation, was successfully isolated from

the crude extract of the champedak seeds (Lim *et al.*, 1997). The mannose binding champedak lectin was shown to have strong interaction towards human IgM and IgE and binding preference towards Manα1-3Man (Lim *et al.*, 1997). Hence, the two champedak seeds lectins are presently termed as champedak galactose binding (CGB) and champedak mannose binding (CMB) lectins, based on the sugars that they bind.

The CGB lectin is a homotetrameric protein composed of four identical subunits (Figure 2.5) (Gabrielsen *et al.*, 2014). Each subunit has a heavy α-chain and a light β-chain (Gabrielsen *et al.*, 2014). The α-chain comprises 133 amino acid residues, while the β-chain consists of 21 amino acid residues (Abdul-Rahman *et al.*, 2002; Gabrielsen *et al.*, 2014). The polypeptide of CGB lectin forms a total of 12 β-strands, 11 of which are in the heavy α-chain and the remaining β-strand is in the light β-chain (Gabrielsen *et al.*, 2014). The β-strands are arranged into three antiparallel β-sheets with each sheet contains four β-strands, forming a β-prism structure (Gabrielsen *et al.*, 2014). The β-prism configuration is stabilised by hydrogen bonds. CGB lectin has been shown to have high binding specificity to galactose but not to mannose (Gabrielsen *et al.*, 2014).

CGB lectin has been used in considerable studies, particularly in cancer research. Abdul-Rahman *et al.* (2007) have demonstrated the use of CGB lectin to detect expression of *O*-glycosylated proteins in serum samples of patients with endometrial adenocarcinoma, squamous cell cervical carcinoma and cervical adenocarcinoma that were subjected to Western blotting. They have discovered a cluster of glycoproteins that was overexpressed in patients with endometrial adenocarcinoma, which was not possibly identified using conventional gel-based separation technique. This protein was later identified as a fragment of inter-α-trypsin inhibitor heavy chain H4 (ITIH4) and found to be highly expressed in cancers

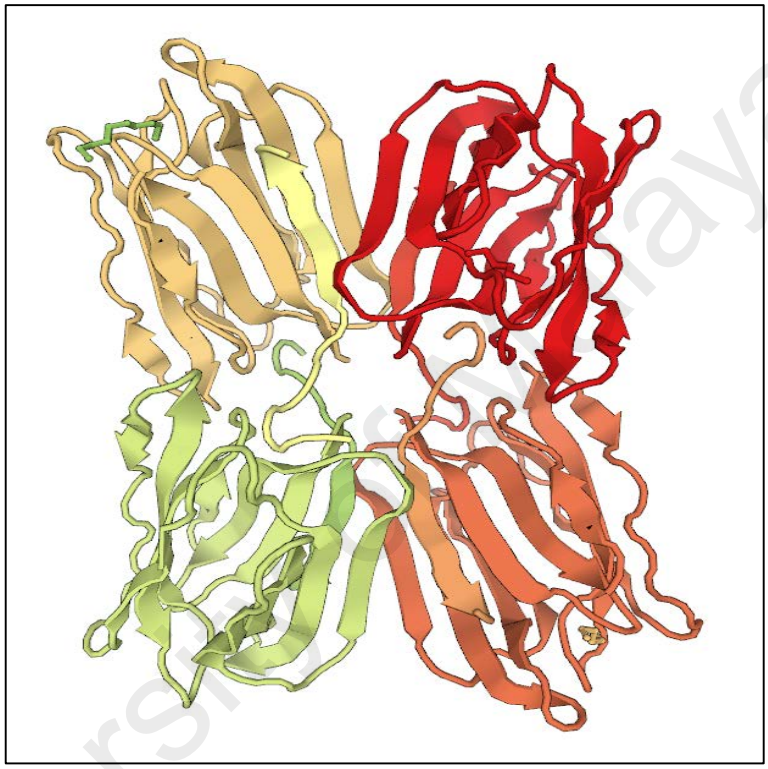


Figure 2.5: 3D crystal structure of CGB lectin.

CGB lectin is a homotetrameric protein with each subunit consists of a heavy α-chain and a light β-chain. In each subunit, the polypeptide forms a total of 12 β-strands, in which 11 of these strands are the heavy α-chain and the last β-strand is the light β-chain.

Source: http://www.rcsb.org/pdb/pv/pv.do?pdbid=4AK4&bionumber=1

associated with high oestrogen levels (Mohamed *et al.*, 2008). Further investigations revealed the ITIH4 fragment was also overexpressed in serum samples of patients with hydatidiform mole as well as normal pregnant women, which dispute the early belief of the association of the peptide with cancer (Mohamed *et al.*, 2013). In another study using similar approaches, the ITIH4 fragment was shown to be increased in the serum of patients with benign prostate hyperplasia but not those with prostate cancer (Jayapalan *et al.*, 2013). Interestingly, when the urine of patients with benign prostate hyperplasia and prostate cancer patients was analysed, increased levels of ITIH4 fragment was apparently detected in the prostate cancer patients but not in those with benign prostate hyperplasia (Jayapalan *et al.*, 2013).

2.2.4 Jacalin

Jacalin is a galactose binding lectin, found in abundant in the seeds of jackfruit (*Artocarpus integrifolia*). Jacalin and CGB lectin share sequence homology in which at least 97% of the amino acid sequences are identical (Gabrielsen *et al.*, 2014). Similar to CGB lectin, the three dimensional structure of jacalin was also characterized as a homotetrameric protein, with each monomer containing heavy α-chain and light β-chain (Figure 2.6) (Sankaranarayanan *et al.*, 1996). The monomer is made up of three anti-parallel β-sheets with four β-strands in each sheet forming a β-prism structure. Despite sharing structural homology with the CGB lectin, both lectins apparently displayed different binding motifs (Gabrielsen *et al.*, 2014). Unlike the CGB lectin, jacalin has been reported to bind with other sugars such as mannose, glucose, *N*-acetylneuraminic acid, and *N*-acetylmuramic acid (Bourne *et al.*, 2002). The study also revealed that jacalin bound to all these sugars using the same binding site. This showed that the β-prism structure of jacalin was flexible enough to accommodate a variety of different

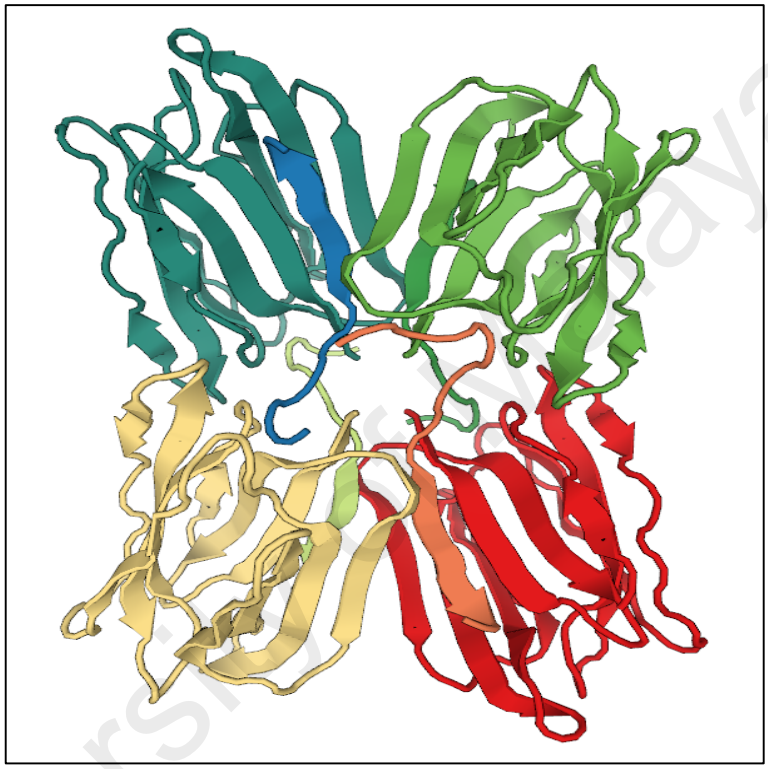


Figure 2.6: 3D crystal structure of Jacalin.

Jacalin, which is similar to CGB lectin, is a homotetrameric protein. Each subunit is made of a heavy α-chain and a light β-chain. The polypeptides arrange itself into three anti-parallel β-sheets with four β-strands in each sheet forming a β-prism structure.

Source: http://www.rcsb.org/pdb/pv/pv.do?pdbid=1KU8&bionumber=1

sugars (Bourne *et al.*, 2002).

Jacalin has been shown to precipitate human IgA since late 1970s (Jagtap & Bapat, 2010). The lectin, which was first isolated by Chatterjee *et al.* (1979) shows strong interaction with T antigen in serum samples. Jacalin first came to the attention of immunologist because of its mitogenic properties (Bunn-Moreno & Campos-Neto, 1981; Saxon *et al.*, 1987). Jacalin, which is presently commercially available, may be considered as one of the most frequently used galactose binding lectin. Research using jacalin is diverse. For example, a study has demonstrated the inhibitory effect of jacalin towards HIV-1 infection on CD4+ cells (Corbeau *et al.*, 1994). Recently, jacalin was included in a lectin array chip to discover potential biomarkers for subfertility (Xin *et al.*, 2016). Jacalin is also commonly used to isolate *O*-glycosylated proteins from complex biological samples (Kabir, 1998).

2.2.5 *Vicia villosa* agglutinin

Vicia villosa agglutinin (VVA) is known to specifically interact with *N*-acetylgalactosamine (GalNAc). This lectin may be extracted from the seeds of *Vicia villosa* or commonly known as hairy vetch (Tollefsen & Kornfeld, 1983a). Hairy vetch belongs to the legume family and is usually found in Europe and western Asia. This predominant lectin in the seeds of *Vicia villosa* does not agglutinate erythrocytes type A, B and O (Tollefsen & Kornfeld, 1983b). However, a small amount of the lectin that was separated using immobilised porcine blood substances and eluted using 1 mM *N*-acetylgalactosamine was characterised to possess haemagglutinating activity (Kimura *et al.*, 1979; Tollefsen & Kornfeld, 1983a).

The crystal structure of VVA, as well as the binding mechanism of the lectin with Tn-antigen (GalNAc-Ser/Thr), have been elucidated (Babino *et al.*, 2003). The

molecular structure of VVA contains four similar monomers in which each monomer consists of 233 amino acid residues (Figure 2.7) (Babino *et al.*, 2003). VVA shares amino acid sequence similarity with other lectins from the legumes family such as soy bean, scotch broom and *E. corallodendron* (Osinaga *et al.*, 1997). The tetrameric structure of VVA is displayed as two 'canonical' dimer packed face to face in parallel fashion (Babino *et al.*, 2003). As a result, this forms a large channel in the middle of the structure. Despite being able to interact with Tn-antigen, VVA is able to bind to GlcNAc and fucose (Babino *et al.*, 2003). Evidences have shown that VVA may contain potential glycosylation (Osinaga *et al.*, 1997; Tollefsen & Kornfeld, 1983b). Because of its binding specificity to Tn-antigen, VVA has been used in many studies, particularly those involved in elucidating the levels of Tn-antigen in cancer (Kawaguchi *et al.*, 2006; Konska *et al.*, 2006; Nishiyama *et al.*, 1987).

2.3 Breast cancer

Cancer is a disease that is associated with abnormal cell growth with the potential of spreading to other parts of the body. Breast cancer is described as cancer that begins from the breast area. Although this cancer mainly occurs in women, some men may be affected as well. Breast cancer mostly begin from the cells that line the ducts (ductal carcinoma) and in some cases in the lobular cells (lobular carcinoma) (Saunders & Jassal, 2009).

For many years, breast cancer has been recognized as the most clinically diagnosed cancer in the world (Torre *et al.*, 2015). Epidemiological data shows that over 1.7 million cases have been reported, with total deaths of 521,900 recorded in 2012 (Torre *et al.*, 2015). Breast cancer alone accounts for 25% of all cancer cases and 15% of all cancer deaths amongst females in the world. According to the recent National

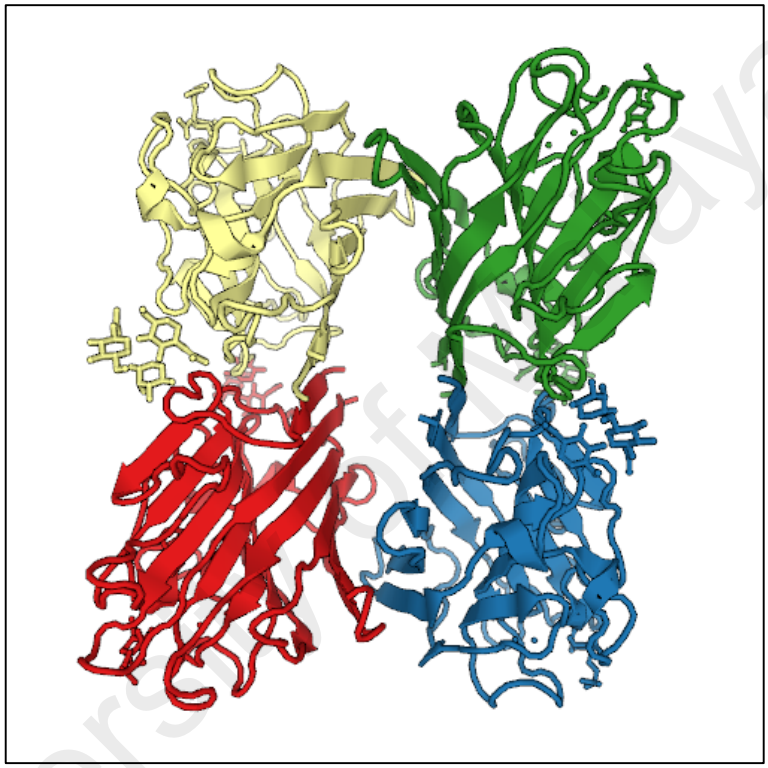


Figure 2.7: 3D crystal structure of VVA.

VVA is made of four similar monomers in which each monomer consist of 233 amino acids. The two canonical dimers packed face to face in parallel fashion, forming a large channel in the middle of the structure. VVA is known to be glycosylated.

Source: http://www.rcsb.org/pdb/pv/pv.do?pdbid=1N47&bionumber=1

Cancer Registry statistics report by the Ministry of Health, Malaysia, breast cancer is the most common cancer diagnosed in Malaysia with total prevalence of 18.1% (Omar & Tamin, 2011). Fortunately, the advancement of diagnostic tools and treatment has greatly improved the survival rates of patients with breast cancer (Cedolini *et al.*, 2014; Kaplan *et al.*, 2015).

2.3.1 Types of breast cancer

There are several different types of breast cancer which can be identified depending on the location and morphology of the cancer cells. Breast cancer is classified into two categories, which are the non-invasive and invasive cancers. Breast carcinomas *in situ* are considered as non-invasive cancer (Felix, 1999). *In situ* carcinoma is a precancerous state whereby cancer cells have not invaded outside of the ductal or lobular area. *In situ* carcinomas can be further classified into two categories known as ductal carcinomas *in situ* (DCIS) and lobular carcinomas *in situ* (LCIS). DCIS tumour cells are larger and more irregular that are found in the lining of the milk duct (Felix, 1999). LCIS has similar morphology with the lining cells of the acinar lobules and consists of small uniform cells (Felix, 1999). LCIS is difficult to detect through mammogram and typically detected incidentally from biopsies for other reasons.

The most common invasive breast cancer is the invasive carcinoma of no special type, which previously was known as invasive ductal carcinoma (Felix, 1999; Sinn & Kreipe, 2013). This type of cancer begins at the ductal cell lining and the cancer cells may continue to grow towards the fatty tissue area or the lymphatic system of the breast. At this point, cancer cells may spread towards the other parts of the body. Presently, invasive breast cancers have been categorised into several different subtypes.

These subtypes include invasive lobular carcinoma, tubular carcinoma, carcinomas with medullary features, metaplastic carcinoma, carcinomas with apocrine differentiation, adenoid cystic carcinoma, mucinous carcinomas, invasive mucinous carcinoma, carcinomas with neuroendocrine features, invasive papillary carcinoma, and some other rare situation that do not fit into any of the categories (Sinn & Kreipe, 2013).

2.3.2 Diagnosis and classification

Breast cancer is commonly diagnosed through physical examination and imaging technique such as mammography. Mammography is the most commonly used imaging technique that utilizes X-ray to detect anomalies in the breast (Joy *et al.*, 2005). Although the detection sensitivity and specificity of mammogram are high, breast density can affect the mammogram detection sensitivity (Joy *et al.*, 2005; Olson *et al.*, 2012). Higher breast density may result in poor image capture and lead to inconclusive findings. Non-X-ray based method, such as ultrasound and magnetic resonance imaging may be used to obtain better imagery as they are less affected by high breast density (Joy *et al.*, 2005).

For more conclusive diagnosis, anomalies detected by imaging techniques are subsequently verified by biopsy. Depending on the situation, breast tissue samples for biopsy can be obtained via several different procedures such as fine needle aspiration cytology, core needle biopsy (Ahn *et al.*, 2013; Liikanen *et al.*, 2016) and vacuum-assisted breast biopsy (Yu *et al.*, 2010). These biopsy procedures are considerably less invasive as they do not need an open surgery to obtain the tissue samples. Instead, sample tissues are collected by utilising a needle that may be guided by sonography imaging system to the abnormal breast area. Although the accuracy of fine needle aspiration cytology and core needle biopsy procedures is considerably high (Ahn *et al.*,

2013), both the procedures suffer from some limitations. Fine needle aspiration cytology limits the ability to determine ductal carcinoma *in situ* and the volume of sample obtained is low (Berner *et al.*, 2003). A study using an animal model has shown that the core needle biopsy may increase the risk of distant metastases (Mathenge *et al.*, 2014). Therefore, vacuum-assisted breast biopsy was introduced to overcome these limitations (Berner *et al.*, 2003; Burbank *et al.*, 1996).

Tissue samples obtained from biopsies are further evaluated under microscopy or other molecular analysis to specifically identify each subtypes of breast cancer. This would affect the decision of treatment on patients and allow more effective prognosis. The classification of breast cancers are affected by several factors, which are staging, histopathology features, grading, receptor status, and DNA assays findings (Ellis *et al.*, 2003). Breast cancer staging is based on the TNM system that evaluates the size of the tumour (T), whether the tumour has spread to lymph nodes (N) or metastasised (M). There are five main stages of breast cancer ranging from stage 0 to stage IV. The characteristics of breast cancer stages are summarised in Table 2.2. Stage 0 is considered as non-invasive breast cancer. DCIS and LCIS are categorised under stage 0, whilst Stage I to stage IV breast cancers are the invasive cancer. Stage I and stage II breast cancer are divided into two subcategories, whereas stage III is divided into three subcategories. Each category has different characteristics based on the size of tumour, the presence of cancer cells in the lymph nodes and the metastasis of cancer cells. Patients with early stage of breast cancer have higher survivability compared to those at later stage of breast cancer (American Cancer Society, 2015).

Breast cancers are generally characterised based on the histological features of the biopsy samples using microscopy. Different types of breast cancers are generally distinguished based on the histological features of the tumour cells (Ellis *et al.*, 1992).

Table 2.2: Characteristics of stages of breast cancer

Stage	Characteristic
Stage 0	• Non-invasive breast cancer • Cancerous cells are resided within the duct and the lobule. • No evidence of cancerous cells invading to other part of the breast.
Stage I	
Stage IA	• Tumour size up to 2 cm or smaller and has not spread outside the breast.
Stage IB	• No tumour in the breast. Cancer cells in lymph nodes larger than 0.2 mm but smaller than 2 mm OR • Tumour in the breast smaller than 2 cm. Cancer cells in lymph nodes larger than 0.2 mm but smaller than 2 mm.
Stage II	
Stage IIA	• No tumour in the breast. Cancer cells larger than 2 mm found in 1 to 3 axillary lymph nodes or in the lymph nodes near the breastbone OR • Tumour 2 cm or smaller. Spread to axillary lymph nodes OR • Tumour larger than 2 cm but smaller than 5 cm. Did not spread to axillary lymph nodes.
Stage IIB	• Tumour larger than 2 cm but smaller than 5 cm. Cancer cells in lymph nodes larger than 0.2 mm but smaller than 2 mm OR • Tumour larger than 2 cm but smaller than 5 cm. Cancer cells spread to 1 to 3 axillary lymph nodes or in the lymph nodes near the breastbone OR • Tumour is larger than 5 cm but has not spread to the lymph nodes

(Table 2.2, continued)

Stage III	
Stage IIIA	• No tumour in the breast or may present in any size. Cancer found in 4 to 9 axillary lymph nodes or in the lymph nodes near the breastbone OR • Tumour is larger than 5 cm. Cancer cells in lymph nodes larger than 0.2 mm but smaller than 2 mm OR • Tumour is larger than 5 cm. Cancer cells spread to 1 to 3 axillary lymph nodes or the lymph nodes near the breastbone.
Stage IIIB	• Tumour may be any size. Spread to chest wall and/or skin of the breast AND • Spread to 9 axillary lymph nodes OR • Spread to lymph nodes near the breastbone
Stage IIIC	• No tumour in the breast or may present in any size. Spread to chest wall and/or the skin of the breast AND • Spread to 10 or more axillar lymph nodes OR • Spread to lymph nodes above or below the collarbone OR • Spread to axillary lymph node or lymph nodes near the breastbone
Stage IV	• Tumour has spread to other parts of the body, such as lungs, distant lymph nodes, skin, bones, liver, or brain.

source: http://www.breastcancer.org/symptoms/diagnosis/staging

The grading system evaluates the differentiation of cancer cells compared to the normal breast cells (Elston & Ellis, 1991). Cancer cells are graded as differentiated (low-grade), moderately differentiated (intermediate-grade), and poorly differentiated (high-grade). Typically, patients with lower grade have higher survivability compared to the higher grade patients (Elston & Ellis, 1991). Receptor status determined the presence or absence of estrogen receptors (ER), progesterone receptors (PR) and human epidermal growth factor receptor 2 (HER2) using immunohistochemistry. Breast cancer cells that are absent of all three of these receptors are known as triple negative and they are usually more difficult to treat (Dent *et al.*, 2007). DNA assay can be used to compare the gene expression patterns of the cancer cells with normal cells for more efficient characterisation (Ross *et al.*, 2008).

2.3.3 Mucins and breast cancer

MUC1 is one of the earliest discovered human mucin and the most extensively studied mucin. MUC1 has been shown to have an association with the development of breast cancer (Brockhausen *et al.*, 1995; Ghosh *et al.*, 2013; Iizuka *et al.*, 2015; Rakha *et al.*, 2005). In the normal breast epithelial cells, MUC1 is usually present mainly on the surface of the apical plasma membrane (Duffy *et al.*, 2010; Gendler, 2001). However, in malignant cells, MUC1 may be highly expressed in the cytoplasm and the entire plasma membrane (Rahn *et al.*, 2001). This mucin is released into the blood stream through secretion or shedding from the malignant cells, causing the increased levels of MUC1 in blood. Tumour antigens CA 27.29 and CA 15-3, that are derived from MUC1, are used as a diagnostic and prognostic marker in most clinical practice (Duffy *et al.*, 2000). Unfortunately, these markers are proven to have less sensitivity in detecting early stages of breast cancer (Duffy *et al.*, 2010). Furthermore, some studies

have found overexpression of MUC1 in other carcinomas such as ovarian, lung, colon and pancreatic cancer (Gendler, 2001), resulting in the lack of its specificity. Other mucins such as the MUC3 and MUC4 are also found to be highly elevated in the epithelial cells of breast cancer patients (Rakha *et al.*, 2005).

Despite being overly expressed, altered glycosylation in MUC1 has also been detected in the breast cancer tissues (Geng *et al.*, 2013). Ghosh and co-workers have observed an underglycosylated form of MUC1, which was highly expressed in the apical membrane of lower grade of breast tumour (Ghosh *et al.*, 2013). Early termination by sialylation of α2,6-linked sialic acid on cell mucins glycan could reduce the tumour cell-cell interaction and thus, promote metastasis (Lin *et al.*, 2002). Such altered glycosylation has long been recognised as a tumour development in breast cancer. However, recent studies in an animal model has uncovered that the deficiency of core 1 *O*-glycan structure apparently decreases the progression of breast cancer (Song *et al.*, 2015).

CHAPTER 3: MATERIALS AND METHODS

3.1 Materials

All chemicals and consumables used in this study are as listed below**:**

Chemicals	Source
3,3'-diaminobenzidine tetrahydrochloride	Thermo Fisher Scientific *Rockford, IL, USA*
Acetic acid	JT Baker *Philadelphia, USA*
Acetonitrile	Merck *Darmstadt, Germany*
Acrylamide	GE Healthcare Life Sciences *Uppsala, Sweden*
Agarose	Thermo Fisher Scientific *Rockford, IL, USA*
Amberlite® XAD4	Sigma-Aldrich *St. Louis, MO, USA*
Ammonium bicarbonate	Sigma-Aldrich *St. Louis, MO, USA*
Ammonium persulphate	Sigma-Aldrich *St. Louis, MO, USA*
Ammonium sulphate	Sigma-Aldrich *St. Louis, MO, USA*
Barium carbonate	Sigma-Aldrich *St. Louis, MO, USA*
Boric acid	Sigma-Aldrich *St. Louis, MO, USA*
Bromophenol blue	GE Healthcare Life Sciences *Uppsala, Sweden*
D-galactose	Sigma-Aldrich

	St. Louis, MO, USA
Dithiothreitol	GE Healthcare Life Sciences *Uppsala, Sweden*
Divinylsulfone	Merck *Darmstadt, Germany*
Drystrip cover fluid	GE Healthcare Life Sciences *Uppsala, Sweden*
Ethanol	John Kollin Co. *Midlothian, UK*
Ethylenediaminetetraacetic acid disodium salt	Systerm *Selangor, Malaysia*
Formic acid	Sigma-Aldrich *St. Louis, MO, USA*
Glycerol	GE Healthcare Life Sciences *Uppsala, Sweden*
Glycine	Merck *Darmstadt, Germany*
Hydrochloric acid	Merck *Darmstadt, Germany*
Hydrogen peroxide	Sigma-Aldrich *St. Louis, MO, USA*
Iodoacetamide	GE Healthcare Life Sciences *Uppsala, Sweden*
IPG Buffer (pH 3-10)	GE Healthcare Life Sciences *Uppsala, Sweden*
Magnesium chloride	Sigma-Aldrich *St. Louis, MO, USA*
N, *N*'-methylenebisacrylamide	GE Healthcare Life Sciences *Uppsala, Sweden*
Orange G	Sigma-Aldrich *St. Louis, MO, USA*

Perchloric acid	Sigma-Aldrich *St. Louis, MO, USA*
p-nitrophenyl phosphate	Sigma-Aldrich *St. Louis, MO, USA*
Potassium Chloride	Systerm *Selangor, Malaysia*
Potassium hydroxide	Merk *Darmstadt, Germany*
Potassium phosphate monobasic	Systerm *Selangor, Malaysia*
SIGMAFAST™ BCIP®/NBT tablet	Sigma-Aldrich *St. Louis, MO, USA*
Silver nitrate	Merck *Darmstadt, Germany*
Sodium acetate, trihydrate	Merck *Darmstadt, Germany*
Sodium azide	Sigma-Aldrich *St. Louis, MO, USA*
Sodium bicarbonate	Merck *Darmstadt, Germany*
Sodium borohydride	Sigma-Aldrich *St. Louis, MO, USA*
Sodium carbonate	Merck *Darmstadt, Germany*
Sodium chloride	Merck *Darmstadt, Germany*
Sodium dodecyl sulphate	Fisher Scientific *Pittsburgh, PA, USA*
Sodium hydroxide	Systerm *Selangor, Malaysia*
Sodium periodate	Sigma-Aldrich *St. Louis, MO, USA*

Sodium phosphate dibasic	ICN Biomedicals, Inc. *Irvine, CA, USA*
Sodium tetraborate, decahydrate	Sigma-Aldrich *St. Louis, MO, USA*
Sodium thiosulphate	Sigma-Aldrich *St. Louis, MO, USA*
Sulphuric acid	Fisher Scientific *Pittsburgh, PA, USA*
Tetramethylethylenediamine	Sigma-Aldrich *St. Louis, MO, USA*
Trifluoroacetic acid	Sigma-Aldrich *St. Louis, MO, USA*
Tris	Sigma-Aldrich *St. Louis, MO, USA*
Triton X-100	Sigma-Aldrich *St. Louis, MO, USA*
Tween-20	Merck *Darmstadt, Germany*
Urea	Merck *Darmstadt, Germany*
α-cyano-hydroxycinnamic acid	Sigma-Aldrich *St. Louis, MO, USA*
β-mercaptoethanol	Sigma-Aldrich *St. Louis, MO, USA*

Consumables	Source
Biotinylated-*Vicia villosa* agglutinin	E-Y Labs *San Mateo, CA, USA*
Biotinylated-jacalin	E-Y Labs

	San Mateo, CA, USA
C_{18} ZipTip	Milipore *Billerica, MA, USA*
Cellulose membrane dialysis tubing	Sigma-Aldrich *St. Louis, MO, USA*
Chromatography paper (3 mm)	Whatman *Maidstone, UK*
ELISA immunoassay plates (medium binding)	Jet-Biofill *Guangzhou, China*
EZ-Link sulfo-NHS biotin	Thermo Fisher Scientific *Rockford, IL, USA*
Horse radish peroxidase-goat anti-rabbit	Abcam *CBG, UK*
Immobilised pH gradient strip, pH 3-10, 11 cm	GE Healthcare Life Sciences *Uppsala, Sweden*
Jacalin	E-Y Labs *San Mateo, CA, USA*
Lysozyme	Sigma-Aldrich *St. Louis, MO, USA*
Nitrocellulose membrane	Sartorius *Goettingen, Germany*
Pierce bicinchoninic acid protein assay kit	Thermo Fisher Scientific *Rockford, IL, USA*
Polypropylene column	Bio-Rad Laboratories *Berkeley, CA, USA*
Rabbit anti-plasma C1 inhibitor	Abcam *CBG, UK*
Sepharose CL-4B	Sigma-Aldrich *St. Louis, MO, USA*
Spectra™ Multicolor High Range Protein Ladder	Thermo Fisher Scientific *Rockford, IL, USA*
Spectra™ Multicolor Low Range Protein Ladder	Thermo Fisher Scientific

	Rockford, IL, USA
Streptavidin-alkaline phosphatase	Sigma-Aldrich *St. Louis, MO, USA*
Trypsin Gold	Promega *Madison, USA*
Vicia villosa agglutinin	E-Y Labs *San Mateo, CA, USA*

3.2 Samples

3.2.1 Test glycoproteins

Porcine gastric mucin (PGM) was partially purified from commercial preparations (Sigma-Aldrich, St. Louis, MO, USA) as described by (Variyam & Hoskins, 1981). Bovine submaxillary mucin (BSM) and porcine submaxillary mucin (PSM) were prepared from fresh glands as previously described by (Bhavanandan & Hegarty, 1987). Fetuin, alpha1-acid glycoprotein (AGP), and ovomucoid (trypsin inhibitor) were obtained from Sigma-Aldrich, St. Louis, MO, USA.

3.2.2 Human blood samples

A case control study was performed by matching cases to controls by age and ethnicity. Cases were recruited consecutively between April 2013 and December 2014. Clinical data and bio-specimens were collected prospectively as part of the University Malaya Medical Centre (UMMC) breast cancer clinical and biospecimen repository. Table 3.1 shows the demographic of the participated subjects. A total of 79 consecutive pre-treatment blood samples of stage 0 and I breast cancer patients were recruited into the study. Among these, 31 were pathologically confirmed as stage 0 (carcinoma *in*

situ) while 48 were diagnosed with stage I breast cancer. The controls were recruited from healthy women attending opportunistic breast mammogram at the Department of Biomedical Imaging in which their mammogram showed no indications of abnormalities (BIRADS 1) from April 2014 to Dec 2014. Written informed consent was obtained from all subjects who participated in the study prior to sample collection. This study and its consent procedure were approved by the Medical Ethics Committee of the University of Malaya Medical Centre (MEC No. 1004.17, Appendix A).

3.2.3 Serum samples preparation

All blood samples were collected into serum BD Vacutainer tubes (Becton, Dickinson and Co., Franklin Lakes, NJ, USA) and allowed to coagulate at room temperature (23°C) for 30 min. Serum was separated from blood clot by centrifugation at 1,300 x g for 10 min. Serum was stored as 1 ml of aliquots in -80°C freezer.

Table 3.1: Demographics of breast cancer patients and healthy participants.

	Control (n = 105)	Stage 0 (n = 31)	Stage I (n = 48)	Significance
Age (Years)	54.84 ± 1.11	54.00 ± 2.31	59.96 ± 1.40	n.s.
Ethnicity (M/C/I)*	35/49/20	5/20/6	13/29/6	n.s.

*M, Malay; C, Chinese; I, Indian

n.s., not significant

3.3 Purification of CGB lectin

3.3.1 Preparation of crude extracts of champedak seeds

Solutions	Final concentration	Amount
Phosphate Buffered Saline:		
Sodium chloride	137 mM	8 g
Potassium chloride	2.7 mM	0.2 g
Disodium phosphate	10 mM	1.42 g
Potassium phosphate monobasic	1.8 mM	0.24 g
Distilled water	-	800 ml
HCl	-	adjusted to pH 7.4
Distilled water	-	topped up to 1,000 ml
Sodium carbonate buffer:		
Sodium carbonate	0.5 M	26.5 g
Distilled water	-	topped up to 500 ml
Sodium bicarbonate buffer:		
Sodium bicarbonate	0.5 M	105 g
Distilled water	-	2,500 ml
20% galactose in sodium carbonate buffer:		
Galactose	20% (w/v)	40 g
Sodium carbonate buffer	0.5 M	topped up to 200 ml

0.8 M galactose in PBS:

Galactose	0.8 M	72.1 g
PBS	-	topped up to 500 ml

Champedak fruits were purchased from a local market. Crude extract from champedak seeds was prepared according to the protocol established by Hashim *et al.* (1991). Champedak seeds were removed from the flesh and thoroughly cleaned using distilled water. The seed coats were peeled, leaving only the endosperm. The uncoated seeds were cut into small pieces and subjected to drying in an oven at 37°C for two days. Dried seeds were ground to powder using an electrical blender. About 100 g of the powdered seeds were suspended in 1,000 ml of phosphate buffered saline (PBS) with constant stirring at 4°C for overnight. The crude extract was separated from the suspension by centrifugation at 8,000 x g for 15 min at 4°C. The resulting supernatant was collected and subjected to ammonium sulphate precipitation to remove impurities. The amount of ammonium sulphate required was determined using the ammonium sulphate precipitation table (Appendix A). Ammonium sulphate was added gradually to reach 60% saturation at 4°C and the suspension was left to stir for two to three hours. The solution was later centrifuged at 8,000 x g for 15 min at 4°C. The resulting pellet was recovered whilst the supernatant was discarded. The pellet was reconstituted with PBS and residual ammonium sulphate was removed by extensive dialysis using cellulose membrane dialysis tubing. The solution was dialysed against PBS at 4°C for 48 hours with four changes of the buffer. After dialysis, the crude extract of champedak seeds was stored at -20°C until further use.

3.3.2 Preparation of galactose affinity column

The preparation of galactose affinity column was begun by activating the Sepharose 4B gel by using Divinylsulfone (DVS). The activation of Sepharose 4B was based on the procedure performed by Hermanson *et al.* (1992). One hundred ml of Sepharose 4B was washed with distilled water in a sintered glass funnel while excess water was gradually removed using vacuum suction. After washing, vacuum suction was continued until the beads became a wet cake form. The moist gel was transferred into a 500 ml beaker containing 100 ml of 0.5 M sodium carbonate and stirred gently using a glass rod. While stirring, 10 ml of DVS was added dropwise in a fume hood for a period of 15 min. The slurry was stirred continuously for an hour at room temperature (23°C). The gel was washed extensively with distilled water, until the filtrate was no longer acidic.

The DVS-activated Sepharose 4B gel was transferred into an equal volume of 20% (w/v) D-galactose in 0.5 M sodium carbonate solution followed by gentle stirring at room temperature (23°C) for 24 hours. Successively, it was washed with 2 L of distilled water and 2 L of 0.5 M sodium bicarbonate prior to re-suspension in 100 ml of 0.5 M sodium bicarbonate containing 5 ml of β-mercaptoethanol. The slurry was continuously stirred for two hours in a fume hood to block excess vinyl reactive groups. The gel was washed with 2 L of distilled water followed by 2 L of PBS. Column packing was performed by slowly transferring the galactose-Sepharose 4B slurry into a polypropylene column (2.8 cm in diameter), which was half-filled with PBS. The buffer was released while subsequently adding more buffer into the column until the gel was packed. The column was stored in 4°C until use.

3.3.3 Isolation of CGB lectin

Packed galactose-Sepharose 4B column was washed extensively by passing through substantial volume of PBS buffer. Fractions of 10 ml were collected from the flow through and possible contaminant was detected by reading absorbance at 280 nm. Crude extract was introduced only after the absorbance reading reached a baseline level of close to zero. Ten ml of champedak crude extract was introduced into the column and PBS was subsequently added to remove all unbound materials. The absorbance of each collected fraction was again recorded. Bound CGB lectin was eluted using 0.8 M galactose in PBS after the absorbance reading of collected fractions had reached near to zero. The galactose solution was added continuously until the absorbance reading of the eluate showed zero. The column was flushed with excessive volume of PBS to remove all the galactose solution before storing the column in 4°C. Five to ten fractions of the eluate with highest absorbance value and contained the CGB lectin, were pooled and dialysed against 0.5X PBS in cellulose membrane tubing. Dialysis was performed at 4°C for two days with four changes of fresh buffer. After dialysis, the lectin solution was subjected to lyophilisation using freeze dryer equipment (Labconco Corporation, Kansas City, Missouri). Ten mg of dried CGB lectin powder was weighed and reconstituted in 1 ml of PBS. The concentration of lectin was determined using bicinchoninic acid (BCA) protein assay. Purity of the CGB lectin was assessed using sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) (Section 3.4.2) and MALDI-TOF/TOF in linear mode analysis (Section 3.4.4). The lectin was electrophoresed on 18% SDS-polyacrylamide gel and visualised using silver stain. Binding activity was evaluated using lectin radial diffusion (Section 3.4.5).

3.4 Quality assessment of isolated CGB lectin

3.4.1 Quantitation of CGB lectin

The concentration of isolated CGB lectin was estimated using Pierce BCA protein assay kit from Thermofisher Scientific (Rockford, IL, USA). The protein assay was performed based on the microassay protocol using 96 wells plate. Bovine serum albumin standard was serially diluted from the highest concentration of 1,000 µg/ml to the lowest concentration of 15.63 µg/ml. Samples were diluted accordingly. Twenty five µl of standard and samples were added into the wells. Working solution of BCA was prepared by combining one part of Solution B with 50 parts of Solution A and mixed by vortexing. Two hundred µl of working solution was added into each well containing samples and standard. The plate was incubated at 37°C for 30 min and the absorbance reading was obtained at 570 nm.

3.4.2 SDS-PAGE

Solutions	Final concentration	Amount
Solution A (monomer):		
Acrylamide	30%	60 g
N, N'-methylenebisacrylamide	0.8%	1.6 g
Double-distilled water	-	topped up to 200 ml
Amberlite® XAD4		2 g

The solution was stirred in dark at 4°C overnight. The solution was filtered and stored in dark bottle at 4°C.

Solution B (4X resolving buffer):

Tris	1.5 M	90.75 g
Double-distilled water	-	300 ml
HCl	-	adjusted to pH 8.8
Double-distilled water	-	topped up to 500 ml

The solution was stored at 4°C.

Solution C (10% (w/v) SDS):

SDS	10% (w/v)	10 g
Double-distilled water	-	topped up to 100 ml

The solution was stored at room temperature (23°C).

Solution D (10% (w/v) Ammonium persulphate (APS)):

APS	10% (w/v)	1 g
Double-distilled water	-	topped up to 10 ml

APS was kept in aliquots of 100 µl at -20°C.

Solution E (N, N, N', N'-tetramethylethylenediamine (TEMED))

Solution F (Stacking buffer):

Tris	0.5 M	6.1 g
Double-distilled water	-	50 ml
HCl	-	adjusted to pH 6.8
Double-distilled water	-	topped up to 100 ml

The solution was stored at 4°C.

4X Laemmli buffer:

Solution F	62.5 mM	2.5 ml
Glycerol	10% (v/v)	2.0 ml
SDS	2% (w/v)	400 mg
DTT	1% (w/v)	200 mg
Bromophenol blue	0.01% (w/v)	few grains
Double-distilled water	-	topped up to 20 ml

The buffer was stored in aliquots of 1 ml at -20°C.

SDS-Running buffer:

Tris	25 mM	3.03 g
Glycine	192 mM	14.4 g
SDS	0.1% (w/v)	1.0 g
Double-distilled water	-	topped up to 1 L

The buffer was stored at 4°C.

Electrophoresis was performed using Mini-Protean® Electrophoresis System coupled with PowerPac™ Basic (Bio-Rad Laboratories, Berkeley, CA, USA). Clean glass plates were assembled on casting frame and casting stand according to the manufacturer's protocol. Spacer plate with 1.0 mm of thickness was used. The resolving gel mixture of the required percentage was prepared according to Table 3.2. The gel mixture was gently pipetted onto the glass plates until it reached 2 cm before the short plate. Distilled water was added to prevent uneven gel surface. The gel was left to polymerise for at least one hour at room temperature (23°C). Distilled water was removed from the gel and stacking gel mixture was added. A gel comb was inserted into the glass plates and was left to polymerise.

Table 3.2: Gel preparation for SDS-PAGE.

Solution	Resolving gel		Stacking gel
	8%	18%	
Solution A (µl)	3,200	7,200	650
Solution B (µl)	3,000	3,000	-
Solution C (µl)	120	120	50
Solution F (µl)	-	-	1,250
Double distilled water (µl)	5,680	1,680	3,050
[#] Total volume (µl)	12,000	12,000	5,000
* Solution D (µl)	60	60	25
* Solution E (µl)	12	12	5

Total volumes of the solutions were sufficient for preparation of two gels.

* Added prior to use.

Samples were mixed with 4X Laemmli buffer at the ratio of 1:1. The mixture was heated at 95°C for 5 min. The polymerised gels were removed from the casting frame and they were assembled on the electrode assembly unit. The unit was then placed into a mini tank. SDS-running buffer was filled into the electrode chamber as well as into the tank, up until to the designated levels. Gel comb was removed from each gel. Samples were loaded into each well together with an appropriate protein ladder. Spectra™ Multicolor Low Range Protein Ladder and High Range Protein Ladder were used for 18% gel and 8% gel, respectively. The electrophoresis was started at a constant voltage of 65 V for 30 min before being increased to 100 V. The run was halted when the dye front had reached 0.5 cm from the bottom of the gel. The gels were removed from the glass plates and were immediately transferred into plastic containers containing fixing solution for silver staining.

3.4.3 Silver staining

Solutions	Final concentration	Amount
Fixing solution:		
Ethanol	40% (v/v)	400 ml
Acetic acid	10% (v/v)	100 ml
Double-distilled water	-	topped up to 1 L
Sensitising solution:		
Ethanol	30% (v/v)	300 ml
Sodium acetate, trihydrate	500 mM	68 g
Sodium thiosulphate	0.2% (w/v)	2 g
Double-distilled water	-	topped up to 1 L

Silver solution:

Silver nitrate	0.25% (w/v)	2.5 g
Double-distilled water	-	topped up to 1 L

Developing solution:

Sodium carbonate	2.5% (w/v)	25 g
Distilled water	-	topped up to 1 L
*Formaldehyde	0.0148% (v/v)	400 µl

* Added prior to use.

Stop solution:

EDTA	39.2 mM	14.6 g
Distilled water	-	topped up to 1 L

Depending on the need, the total volume of each solution was adjusted accordingly. All incubation procedures were done at room temperature (23°C) with gentle shaking on an Orbital shaker (Daiki Scientific, Korea). The silver stained protocol employed was of that compatible for mass spectrometry analysis. Gels were incubated in fixing solution for 30 min. Fixing solution was removed and sensitising solution was added. The gels were incubated in the solution for 30 min. Sensitising solution was removed and the gels were washed thrice with distilled water for 5 min in each step. Silver solution was then added and the gels were left incubated for 20 min. The gels were developed with developing solution after the silver solution had been removed. After incubation for 10-15 mins or once all the desired spots had appeared, the developing solution was replaced by stop solution. The gel images were scanned using Image Scanner III (GE Healthcare, Uppsala, Sweden).

3.4.4 MALDI-TOF/TOF intact protein analysis

CGB lectin was diluted in 70% acetonitrile containing 0.1% formic acid to a final concentration of 100 µg/ml. Two µl of the diluted lectin was mixed with 2 µl of 10 mg/ml alpha-cyano-hydroxycinnamic acid matrix in 70% acetonitrile/0.1% formic acid buffer. About 0.7 µl of the mixture was spotted onto a MALDI plate and allowed to crystallise at room temperature (23°C). The plate was inserted into the AB SCIEX MALDI-TOF/TOF 5800 analyser (Framingham, MA, USA). The system was set to linear mode with 5500 of laser shots. Calibration was performed using 80 µg lysozyme. Mass spectra were retrieved and viewed on Data Explorer software (Applied Biosystems, CA, USA).

3.4.5 Lectin radial diffusion

Agarose solution was prepared by adding 0.5 g of agarose powder into 100 ml of PBS. The solution was boiled in microwave for 1.5 min and was allowed to cool to about 60°C. The agarose solution was pipetted onto glass slides using plastic Pasteur pipettes and was allowed to harden at room temperature (23°C). Five tiny wells with the diameter of 2 mm were made according to the template shown in Figure 3.1 using plastic Pasteur pipettes. Two µl of 5 mg/ml test *O*-glycosylated proteins (bovine submaxillary mucin and pig gastric mucin) and *N*-glycosylated proteins (Alpha-1-acid glycoprotein and ovomucoid) were loaded into each well as indicated in Figure 3.1. Two µl of 2 mg/ml of CGB lectin was loaded in the centre well. The slide was placed in a container containing wet tissue to maintain moisture. The samples were allowed to diffuse at room temperature (23°C) for 24 hours.

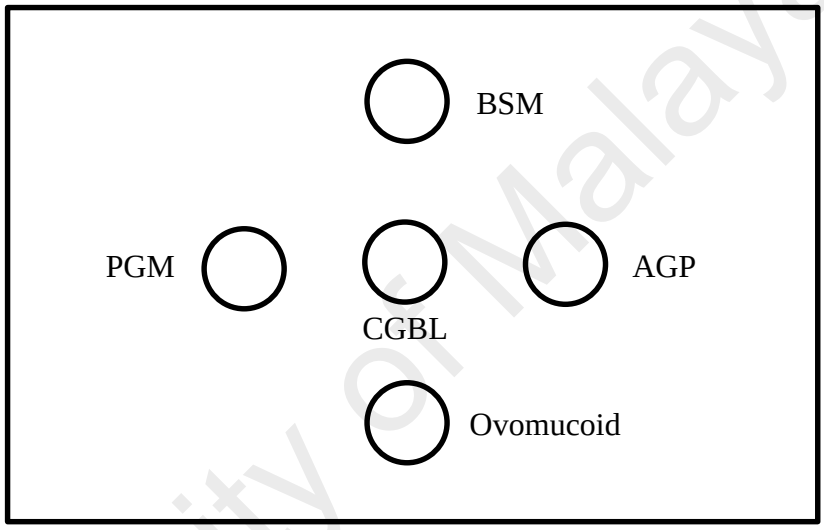


Figure 3.1: Template of wells for the lectin radial diffusion method.

BSM; Bovine submaxilliary mucin, PGM; Pig gastric mucin, AGP; Alpha-1-acid glycoprotein, CGBL; Champedak galactose binding lectin.

3.5 Development of *O*-glycosylated protein assay

3.5.1 Modification of *O*-glycans

Solutions	Final concentration	Amount
1 N Sulphuric acid:		
Sulphuric acid, 18 M	1 N	2.78 ml
Distilled water	-	topped up to 100 ml
Sodium acetate buffer:		
Sodium acetate, trihydrate	0.032 M	2.18 g
Acetic acid, 1 M	0.018 M	9 ml
Distilled water	-	topped up to 500 ml
pH was at 4.5.		
Sodium periodate solution:		
Sodium periodate	0.05 M	1.07 g
Sodium acetate buffer	-	topped up to 100 ml
Sodium borate buffer:		
Sodium tetra-borate, 0.05 M	0.015 M	300 ml
Boric acid, 0.2 M	0.14 M	700 ml
pH was at 8.0.		
Sodium borohydride solution:		
Sodium borohydride	0.3 M	1.13 g
Sodium borate buffer	-	100 ml

The solution was prepared prior to use.

Chemical desialylation was performed by treatment with mild acid hydrolysis (0.1 N H_2SO_4, 80ºC, 60 min) and the asialo product recovered by neutralisation with sodium hydroxide followed by dialysis against distilled water and lyophilisation. Lyophilised asialo sample was reconstituted in ultra-pure water. Protein concentration was determined using the BCA microassay.

Native as well as desialylated mucins, were subjected to Smith degradation (Andrews *et al.*, 1969; Perlin, 2006), which involved mild periodate oxidation of 100 mg sample with 20 ml of 0.05 M sodium periodate in 0.05 M sodium acetate, pH 4.5 buffer at 4ºC for 18 hours in dark. The solution was transferred to dialysis tubing and dialysed against distilled water for 48 hours with two changes of water at 4°C. Dialysis was continued with sodium borate buffer. The aldehyde groups generated on the susceptible saccharides were reduced by treatment with an equal volume of 0.3 M sodium borohydride in 0.2 M sodium borate buffer. After incubation for three hours at room temperature (23°C), the reaction was terminated by neutralization with acetic acid. The samples were finally subjected to mild acid hydrolysis (0.025 M H_2SO_4 at 80ºC for 60 min) and neutralised with 1.0 N sodium hydroxide. The sample was recovered by dialysis against distilled water for 48 hours at 4°C and following that they were lyophilised. The Smith-degraded samples were reconstituted and the protein concentration determined by the BCA microassay (Section 3.4.1).

3.5.2 Analysis of monosaccharide of native and modified glycoproteins

For neutral sugars and hexosamines, the samples were hydrolysed with 2.0 N trifluoroacetic acid at 100°C for six hours. The hydrolysates were dried in a vacuum centrifuge (Speed-Vac) (Labconco Corporation, Kansas City, Missouri), reconstituted in HPLC grade water and analysed on a Carbopac PA10 high pH anion-exchange column

(HPAEC) using a Dionex BioLC HPLC by isocratic elution with 16 mM sodium hydroxide and pulsed amperometric detection (Lee, 1990). Sialic acids were determined after hydrolysis with 0.1 N sulfuric acid at 80°C for one hour. The hydrolysates was neutralized with barium carbonate, centrifuged and the supernatant dried, reconstituted as above and analysed by HPAEC as described by Manzi and co-workers (Manzi *et al.*, 1990). The response factors for the monosaccharides were determined using standard sugars.

3.5.3 Protein biotinylation

Biotinylation of CGB lectin and asialoBSM were performed according to the manufacturer's instruction. Biotinylated protein was prepared by adding sufficient volume of 10 mg/ml EZ-Link sulfo-NHS-biotin reagent to a solution of the respective protein (5 mg/ml in PBS). The solution was mixed gently at room temperature (23°C) for 60 min. The reaction mixture was dialysed for 48 hours with two changes of distilled water. The amount of biotinylated protein recovered in the retentate was determined by using BCA protein microassay (Section 3.4.1).

3.5.4 Microtiter plate lectin binding assay

Solutions	Final concentration	Amount
Coating buffer:		
Sodium carbonate	0.05 M	2.65 g
Sodium azide	0.05%	0.25 g
Distilled water	-	300 ml
HCl	-	adjusted to pH 9.6

Distilled water	-	topped up to 500 ml

Wash buffer:

Tris	0.01 M	1.21 g
Sodium azide	0.05%	0.5 g
Distilled water	-	800 ml
Tween-20	0.05%	0.5 ml
HCl	-	adjusted to pH 8.0
Distilled water	-	topped up to 1,000 ml

Substrate buffer:

Magnesium Chloride	1 mM	19 mg
Coating buffer	0.05 M	topped up to 200 ml

p-nitrophenyl phosphate in substrate buffer:

p-nitrophenyl phosphate	1 mg/ml	10 mg
Substrate buffer	-	topped up to 10 ml

ELLA was performed on samples either directly coated onto 96 well microtiter plates or on samples bound to lectin pre-coated plates. For coating of plates, 50 µl of the sample or lectin diluted in coating buffer (0.05 M sodium carbonate, pH 9.6) was added to each well and incubated overnight at 4ºC. The wells were washed thrice with TBS-T (0.01 M Tris-HCl, pH 8.0, 0.05% Tween-20 and 0.05% sodium azide) and blocked by incubating with the wash buffer for 30 min. In the case of the lectin pre-coated plate, 50 µl test sample in phosphate buffered saline (PBS) was added to each well, incubated for 90 min at 20ºC and unbound material removed by washing as above.

In both cases, the quantity of *O*-glycosylated protein in the test sample bound to the plate was determined as follow: 50 µl of diluted biotinylated lectin in TBS-T was added to each well and incubated for 90 min. The solution was removed and the wells washed three times. Fifty µl of avidin-alkaline phosphatase (1:10,000) in TBS-T was then added to the wells and incubated for 90 min. The plate was washed thrice before adding 50 µl of 1 mg/ml p-nitrophenyl phosphate in substrate buffer (0.05 M sodium carbonate, pH 9.6, 1 mM $MgCl_2$). After appropriate incubation time, typically 20-30 min at 20°C, the absorbance was read at 415 nm on a microplate reader (Bio-Rad, USA). Checkerboard assays were performed with serial dilutions of each ligand and lectin to determine optimal concentrations (Duk *et al.*, 1994; McCoy *et al.*, 1983). All ELLAs were done in triplicates and the mean values plotted. Typically, the variation (standard deviation) between the replicates was between 5-10 % of the mean values.

3.6 Enrichment of serum glycoproteins using perchloric acid

3.6.1 Perchloric acid treatment of serum samples

Enrichment of glycoproteins was performed by treating the serum samples with perchloric acid at various concentrations to optimise the most suitable concentration. Fifty µl of serum, unless otherwise stated, was incubated with an equal volume of cold perchloric acid for 20 min. Supernatant was collected after centrifugation at 10,000 x g for 10 min and neutralised by adding sufficient amount of 1.0 N potassium hydroxide, with phenol red as the pH indicator. Perchlorate salt was removed by centrifugation at 10,000 x g for 10 min. The supernatant was collected and protein concentration was estimated using the BCA protein microassay. Serum perchloric acid isolates obtained were finally subjected to dialysis and lyophilisation prior to their subsequent analyses.

3.6.2 Two-dimensional gel electrophoresis

Solutions	Final concentration	Amount
Sample buffer:		
Urea	9 M	13.5 g
DTT	60 mM	250 mg
IPG Buffer (pH 3-10)	2% (v/v)	0.5 ml
Triton X-100	0.5% (v/v)	0.13 ml
Double-distilled water	-	topped up to 25 ml

Sample buffer was stored as 250 µl aliquot at -20°C.

Rehydration buffer:		
Urea	8 M	12 g
IPG buffer (pH 3-10)	0.5% (v/v)	0.13 ml
Triton X-100	0.5 % (v/v)	0.13 ml
Orange G	-	Few grains
Double-distilled water	-	topped up to 25 ml

Rehydration buffer was stored as 1 ml aliquot at -20°C. About 2 mg of DTT was added into 1 ml of the buffer prior to use.

SDS Equilibration buffer:		
Urea	6 M	72.07 g
Tris-HCl, pH 8.8 (Solution B)	75 mM	10 ml
Glycerol	87% (w/v)	69 ml
SDS	2% (w/v)	4 g
Double-distilled water	-	topped up to 200 ml

The buffer was stored at room temperature (23°C).

First equilibration: 100 mg of DTT was added into 10 ml of equilibration buffer prior to use. Second equilibration: 450 mg of iodoacetamide was added into 10 ml of equilibration buffer prior to use.

Cathode buffer:

Tris	25 mM	3.03 g
Glycine	192 mM	14.4 g
SDS	0.1% (w/v)	1.0 g
Distilled water	-	topped up to 1 L

The buffer was freshly prepared before use.

Anode buffer:

Tris	0.37 M	136.2 g
HCl	-	Adjusted to pH 8.8
Distilled water	-	topped up to 3 L

The buffer was stored in 4°C.

Agarose solution:

Agarose	0.5% (w/v)	0.5 g
SDS-Running buffer	-	top up to 100 ml
Bromophenol blue	-	few grains

The solution was heated in microwave until the agarose was dissolved.

Samples of lyophilised normal human serum perchloric acid isolates as well as neat human serum were subjected to two-dimensional gel electrophoresis (2-DE).

Lyophilised serum perchloric acid isolates was first reconstituted into 20 µl of ddH_2O. The sample was allowed to sit at room temperature (23°C) for 30 min before adding 60 µl of sample buffer. For neat human serum samples, 100 µg of serum protein was added into appropriate volumes of sample buffer, in which about three times volume of the sample. Samples in sample buffer were left to incubate for 30 min at room temperature (23°C). Next, 2 mg of dithiothreitol (DTT) was added into 1 ml of rehydration buffer. Appropriate volume of the buffer was added into the sample mixtures to a final volume of 200 µl. Again, the sample mixtures were allowed to incubate at room temperature (23°C) for another 30 min. The sample mixtures were pipetted into each channel, followed by gently sliding down the pH 3-10 Immobilized pH gradient (IPG) dry strips (11 cm) with the gel surface facing down into the solution. Three ml of drystrip cover fluid was added into each channel to avoid sample evaporation and urea crystallisation. The strips were rehydrated in the solutions for 18 hours at room temperature (23°C). Isoelectric focusing was performed on Multiphor™ II Flatbed electrophoresis system coupled with thermostatic circulator (Grant Instrument Ltd., Cambridge, UK) and EPS-3501 XL power supply (GE Healthcare, Uppsala, Sweden).

The system apparatus was assembled according to the manufacturer's protocol. The Multiphor™ unit was conditioned to 18°C. Sufficient drystrip cover fluid was poured on the cooling plate and the immobiline drystrip tray was positioned on top of the plate, following the exact location where the anode electrode sat. Both the wires from the drystrip tray were connected accordingly to the tank. About 10 ml of drystrip cover fluid was added into the drystrip tray and an immobiline drystrip aligner with 12-groove-side-up was placed on top of the fluid without introducing any bubbles. The rehydrated IPG strips were placed on the drystrip aligner with the gel facing upward. Electrofocusing strips were moistened with distilled water and subsequently, placed on

top of the IPG strips at both ends. The anode and the cathode electrodes were place on the electrode strips accordingly. Electrophoresis was started at 300 V, 2 mA, 5 W for 30 min and continued at 3500 V, 2 mA, 5 W and 12 kV/h.

Second dimension electrophoresis was performed on an 8 – 18% gradient gel. The glass plates (16 x 18 cm) were assembled according to the manufacturer's instruction. One mm thick spacers were used in the assembly. The gradient gel solutions were prepared according to Section 3.4.2 and Table 3.3. Gradient maker (Model SG 30, Hoefer, USA) coupled with a peristaltic pump (EYELA, Tokyo Rikakikai Co., Tokyo, Japan) were used to assist on gradient gel casting. Twelve ml of the heavy solution and the light solution was added into the right chamber and the left chamber of the gradient maker, respectively. Both solutions were continuously stirred using magnetic stirrers. Forty µl of APS solution and 4 µl of TEMED were added into each chamber. While stirring, the heavy solution in the right chamber was released and peristaltic pump was used to draw the solution into the sandwich glass plates. When about 2 ml of the heavy solution was released, the light solution was then released to allow the solution to mix with the heavy solution. The mixture was continuously drawn to the sandwich glass plates until it reached about 14 cm from the bottom of the plates. Immediately, the gel was overlaid with distilled water. The gels were left to polymerise overnight at room temperature (23°C).

The second dimension run was performed on SE 600 Ruby electrophoresis system (Hoefer, Holliston, MA, USA) coupled with a thermostatic circulator (Grant Instrument Ltd. Cambridge, UK) and an EPS601 power supply (GE Healthcare, Uppsala, Sweden). The electrophoresed IPG strips were equilibrated in equilibration buffer containing 10 mg/ml DTT for 15 min in 13 cm screw cap tubes. The buffer was discarded and was replaced with another equilibration buffer containing 45 mg/ml

Table 3.3: Gel preparation for 8 – 18% gradient gel.

Solution	Light solution (8%)	Heavy solution (18%)
Solution A (ml)	6.4	14.4
Solution B (ml)	6	6
Solution C (µl)	240	240
Sucrose (g)	-	3.6
Double distilled water (ml)	11.36	Topped up to 24 ml
# Total volume (ml)	24	24
* Solution D (µl)	40	40
* Solution E (µl)	4	4

Total volumes of the solutions were sufficient for preparation of two gels.

* Added directly into the gradient chamber.

iodoacetamide. The strips were equilibrated for another 15 min. Next, equilibrated strip was aligned on the gradient gel and was sealed with 0.5% of agarose. The agarose was allowed to harden before the run. The upper chamber was assembled on top of the plates accordingly and the whole unit was placed into the electrophoresis tank. The tank was filled with anode buffer while the upper chamber was filled with cathode buffer. First phase of the electrophoresis was started at 50 V, 40 mA, 15 W/gel for 30 min, while the second phase was continued at 600 V, 40 mA, 25, 15 W/gel until the dye front reached the bottom of the gel. Protein spots on the gel were visualised by performing the typical silver stain protocol as described (Section 3.4.4). Silver stained gel images were captured using Image Scanner III (GE Healthcare, Uppsala, Sweden).

3.6.3 In-gel trypsin digestion

Solutions	Final concentration	Amount
Destain:		
Potassium ferricyanide	15 mM	24.7 mg
Sodium thiosulphate	50 mM	39.6 mg
Double-distilled water	-	5 ml
Ammonium bicarbonate:		
Ammonium bicarbonate	100 mM	316 mg
Double-distilled water	-	40 ml
**Reduce:*		
Dithiothreitol	10 mM	7.7 mg
Ammonium bicarbonate solution	100 mM	5 ml

**Alkylate:*

Iodoacetamide	55 mM	50.9 mg
Ammonium bicarbonate solution	100 mM	5 ml

Wash buffer:

Acetonitrile	50%	10 ml
Ammonium bicarbonate solution	50 mM	10 ml

Digestion solution:

Ammonium bicarbonate solution	40 mM	400 µl
Acetonitrile	10%	100 µl
Double-distilled water	-	500 µl

Trypsin solution:

Trypsin stock (1 µg/µl)	20 ng/µl	5 µl
Digestion solution	-	245 µl

* Prepared prior to use.

Silver stained 2-DE gels were rinsed with ultra-pure water for several times. Spots from the 2-DE gel were picked using pre-cut pipette tips. The gel plugs were transferred into microcentrifuge tubes containing double-distilled water and stored in -20°C until further use. One hundred µl of destain solution was added after several washes with double-distilled water on tube shaker (LMS Co., Tokyo, Japan). Constant agitation was applied until the gel plugs become translucent. Destain solution was removed and 150 µl of reducing solution was added into the tube. Peptides in the gel plugs were reduced at 60°C for 30 min. After cooling at room temperature (23°C), the reduced solution was replaced with 150 µl of alkylate solution and left incubated in dark for 20 min. The plugs were washed thrice with 500 µl of 50% acetonitrile in 50 mM ammonium bicarbonate for 20 min in each step. Fifty µl of 100% acetonitrile was added into the tube after the wash solution has been removed. The tubes were shaken

for 15 min until the gels are completely dehydrate. Acetonitrile was removed and the gel plugs were dried using speed vacuum (Labconco Corporation, Kansas City, Missouri). Twenty µl of 20 ng/µl trypsin solution was added into the tube and incubated at room temperature (23°C) for one hour. Sufficient volume of 10% acetonitrile in 40 mM ammonium bicarbonate was added until the buffer covered all the gels. The tubes were then incubated for overnight at 37°C. Peptides from the gels were extracted by adding 50 µl of 50% acetonitrile with constant agitation for 15 min. The extracted peptides were transferred into a new tube. For higher yield, the extraction was repeated by adding 100% acetonitrile. The peptide solution was dried using speed vacuum.

3.6.4 MALDI-TOF/TOF analysis

Lyophilised peptides were desalted according to manufacturer's instruction for C_{18} reversed phase media ZipTip column (Milipore, Billerica, USA). Lyophilised peptides were reconstituted with 1% formic acid in 2% acetonitrile. ZipTip resin was first equilibrated with 100% acetonitrile by pipetting and discarding 10 µl of the solution for three times. Equilibration was repeated using 0.1% formic acid. Peptides were loaded into the resin by pipetting up and down for a total of 10 times. The tip was washed by pipetting and discarding 10 µl of 0.1% formic acid for six times. The peptides were eluted by pipetting 4 µl of 60% acetonitrile in 0.1% formic acid for ten times in a fresh tube.

An equal volume of 10 mg/ml alpha-cyano-hydroxycinnamic acid matrix in 70% acetonitrile/0.1% formic acid buffer was added into the desalted peptides solution. About 0.7 µl of the peptide mixtures was spotted onto the MALDI plate and allowed to

crystallise at room temperature (23°C). After the sample has crystallised, additional volume of sample was added to increase the amount of peptides on a single spot. The plate was inserted into the AB SCIEX MALDI-TOF/TOF 4800 analyser (Framingham, MA, USA). Proteins were identified using MASCOT search engine (Matrix Science, Boston, MA, USA) with Swiss-Prot selected as database. Searches were done with the following parameters: *Homo sapiens* was selected as the taxon. Carbamidomethylation and methionine oxidation were selected for variable modification. Trypsin was selected as the cleaving agent and missed cleavage was set to one. Ion mass tolerance for the precursor ion and MS/MS mode were set to 100 ppm and 0.2 Da, respectively. Only monoisotopic masses were included. Score greater than 43 indicates protein identity and extensive homology ($P < 0.05$).

3.6.5 2-DE CGB lectin blotting

Solutions	Final concentration	Amount
Transfer buffer:		
Tris	25 mM	3.03 g
Glycine	192 mM	14.4 g
Distilled water	-	500 ml
Methanol	20% (v/v)	200 ml
SDS, 10%	0.1%	10 ml
Distilled water	-	topped up to 1 L
TBS-T:		
Tris	20 mM	4.48 g
NaCl	500 mM	58.48 g
Distilled water	-	1.5 L
HCl	-	adjusted to pH 7.5
Tween 20	0.05%	1 ml
Distilled water	-	topped up to 2 L

Gelatin solution:

Gelatin	3% (w/v)	15 g
TBS-T	-	500 ml

Substrate solution:

SIGMAFAST™ BCIP®/NBT tablet	-	1 tablet
Distilled water	-	10 ml

Serum perchloric acid protein isolates that were separated by 2-DE were transferred onto nitrocellulose membrane using a semidry transfer system (Novablot kit for the Multiphor II Electrophoresis system, GE Healthcare Biosciences, Uppsala, Sweden). The graphite anode plate was wet with distilled water and assembled on the Multiphor system accordingly. Six pieces of chromatography papers were cut according to the size of the 2-DE gel. The chromatography papers, the 2-DE gel and a nitrocellulose membrane were soaked in transfer buffer. Three pieces of the chromatography papers were arranged on the anode plate, followed by the nitrocellulose membrane on top of the papers. The gel was then aligned on top of the nitrocellulose membrane and the sandwich was completed by laying another three sheets of chromatography papers on top of the gel. Air bubbles were removed by rolling the sandwich with a roller. The cathode plate was placed on top of the sandwich and electrophoresis was performed for two hours at a constant current of 0.8 mA/cm^2.

The nitrocellulose membrane was removed from the sandwich and subsequently, blocked by incubating with 50 ml of 3% gelatin in TBS-T for one hour at room temperature (23°C) with constant shaking. The membrane was washed thrice with TBS-T for 10 min each, followed by 50 ml of 20 μg/ml biotinylated-CGB lectin in TBS-T at 4°C, overnight. Fifty ml of 1 μg/ml streptavidin-alkaline phosphatase in TBS-

T was added after three rounds of washing. The membrane was left incubated at room temperature (23°C) for 1 hour, followed by three times washing. The membrane was developed by adding 10 ml of BCIP/NBT substrate solution with constant shaking. The development reaction was stopped by rinsing with distilled water and air dried before image scanning using Image Scanner III (GE Healthcare, Uppsala, Sweden).

3.6.6 In-solution trypsin digestion and mass spectrometry

Lyophilised protein samples of serum perchloric acid isolates were reconstituted in 37 µl of 4 mM DTT in 50 mM ammonium bicarbonate. The samples were reduced at 95ºC for 5 min. Alkylation was performed by incubating 3 µl of 100 mM iodoacetamide at room temperature (23°C) in dark for 20 min. The protein sample was digested by adding 2 µl of 0.1 µg/µl Trypsin (Promega, Madison, WI, USA) for 16 hours at 37ºC. The reaction was quenched by freezing the sample. Digested peptides were desalted using µ-C18 ZipTips (Billerica, MA, USA) and eluted in 50% acetonitrile containing 1% formic acid (Section 3.6.4). The peptides were subjected to QTOF LC/MS analysis using Agilent 6550 iFunnel QTOF LC/MS system (Agilent, Santa Clara, CA, USA) for peptides identification. Tryptic peptides were separated using the C18 HPLC-Chip with a linear gradient of water/acetonitrile. Spectra were analysed to identify proteins of interest according to the UNIPROT database.

3.6.7 Biotinylated-asialoBSM spiking experiment

Five µg of biotinylated-asialoBSM was spiked into serum samples, followed by perchloric acid treatment as earlier described (Section 3.6.1). A theoretical control was prepared by addition of 5 µg of biotinylated-asialoBSM into serum after perchloric acid

treatment. Levels of biotinylated-asialoBSM were measured using the following protocol: Microtiter plate (Jet-Biofill, Guangzhou, China) was coated overnight with 50 µl of 1 µg CGB lectin per well. The plate was washed thrice with washing buffer containing 0.01 M Tris-HCl, pH 8.0, 0.05% Tween-20 and 0.05% sodium azide. Biotinylated-asialoBSM spiked samples were added to the precoated wells at dilutions ranging from 1:5,000 to 1:160,000. Biotinylated-asialoBSM, at concentrations ranging from 10 to 0.156 ng/ml, was used as standard.

3.7 Profiling of *O*-glycosylated protein in breast cancer serum perchloric acid isolates

3.7.1 Sepharose conjugation of CGB lectin

Solutions	Final concentration	Amount
Coupling buffer:		
Sodium bicarbonate	0.1 M	8.4 g
Sodium chloride	0.5 M	29.22 g
Distilled water	-	800 ml
HCl	-	adjusted to pH 8.3
Distilled water	-	topped up to 1 L
Glycine solution:		
Glycine	0.2 M	1.5 g
Distilled water	-	100 ml

Acetate buffer:

Sodium acetate, trihydrate	0.1 M	6.8 g
Sodium chloride	0.5 M	14.61 g
Distilled water	-	300 ml
HCl	-	adjusted to pH 4
Distilled water	-	topped up to 500 ml

Storage buffer:

Sodium chloride	1.0 M	5.8 g
Sodium azide	0.05%	0.05 g
Distilled water	-	100 ml

Purified CGB lectin was conjugated to cyanogen-bromide activated Sepharose 4B beads according to the manufacturer's protocol (Sigma Chemical, St. Louis, MO, USA). One g of the Sepharose beads was used, which resulted in 3.5 ml of total gel volume. About 17.6 mg (5 mg protein per ml of gel) of CGB lectin was dissolved in coupling buffer (0.1 M $NaHCO_3$, 0.5 M NaCl, pH 8.3). The beads were washed and swelled in 50 ml tube with cold 1 mM HCl solution for at least 30 min. The washing step was performed by adding a total of 200 ml HCl in several aliquots with gentle agitation and brief centrifugation to remove the supernatant. The Sepharose was washed with five to ten column volumes of distilled water followed by washing with coupling buffer (5 ml per gram dry gel). The CGB lectin solution was immediately transferred to the slurry and was allowed to mix at 4°C for overnight using a 50 ml tube rotary revolver (Labnet International Inc. Edison, NJ, USA). Unreacted lectin was washed using coupling buffer for several times. Unreacted groups of the beads were blocked with 0.2 M glycine for two hours at room temperature (23°C). Extensive

washing was performed to remove the blocking solution by alternating coupling buffer and acetate buffer (0.1 M acetate buffer, 0.5 M NaCl, pH 4) for five cycles. The Sepharose conjugated CGB lectin was stored in 1.0 M NaCl containing 0.05% NaN_3.

3.7.2 Sepharose-CGB lectin precipitation

About 5 µl of Sepharose conjugated CGB lectin was transferred into a 1.5 ml microcentrifuge tube. The beads were washed three times by adding 1 ml of wash buffer (0.01 M Tris-HCl, pH 8.0, 0.05% Tween-20 and 0.05% sodium azide) with continuous vortex for 10 min. The wash buffer was removed by brief centrifugation. The beads were blocked with wash buffer for at least 30 min with continuous vortex prior to use.

Two hundred and fifty µl of serum perchloric acid isolate (120 µg/ml) was added to the Sepharose conjugated CGB lectin beads and incubated for 90 min at room temperature (23°C) with constant vortex. Unbound proteins were removed by centrifugation and the beads were washed as earlier described. The CGB lectin bound proteins were released by adding 25 µl of Laemmli buffer and incubated at 95°C for 5 min. Released peptides were resolved in 8% SDS-polyacrylamide gel (Biorad, Hercules, CA, USA) (Section 3.4.2). Peptide bands were visualized by silver staining. The developed gel was scanned using ImageScanner III (GE Healthcare, Little Chalfont, BU, UK) and image analysis was performed using GelAnalyzer software (http://www.gelanalyzer.com/). Each of the peptide bands was excised and subjected to typical in-gel trypsin digestion (Section 3.6.3). The extracted peptides were analysed using Agilent 6540 mass spectrometer (Agilent, Santa Clara, CA, USA). Proteins were identified by subjecting the spectra to Mascot sequence matching software (Matrix Science, Boston, MA, USA) analysis using the Ludwig NR database.

3.7.3 Western blot

Proteins from a duplicate SDS-polyacrylamide gel were transferred onto nitrocellulose membrane using a mini trans-blot system (Biorad, CA, USA) for 1 hour at a constant voltage of 100 V. Standard Towbin buffer without SDS (25 mM Tris, 192 mM glycine, 20% methanol) was used as transfer buffer. The membrane was blocked with 5% skim milk in TBS-T (20 mM Tris, 500 mM NaCl, 0.05 % Tween-20, pH 7.5) for 30 min and washed at least thrice with TBS-T. Diluted rabbit anti-plasma C1 inhibitor (Abcam, CBG, UK) at 1:1,000 ratio was added and allowed to bind at 4°C overnight. Horse radish peroxidase conjugated secondary goat anti-rabbit (Abcam, CBG, UK) (diluted at 1:2,000) was added after washing with TBS-T and incubated for one hour. The membrane was washed thrice before being developed using 1 mg/ml 3,3'-diaminobenzidine (Thermofisher, MA, USA) in TBST containing 0.02% hydrogen peroxide.

3.8 Statistical Analysis

All data was expressed in mean ± SEM, unless otherwise stated. Statistical analysis was performed using IBM SPSS statistical software version 21 (IBM, New York, USA). Student t-test was used for comparison between groups. Receiver operative characteristic (ROC) was performed to evaluate the ELLA performance to detect stage 0 and stage I breast cancer. P values of less than 0.05 were considered significant.

CHAPTER 4: RESULTS

4.1 Development of assay to detect *O*-glycosylated proteins

Altered abundance of *O*-glycosylated proteins are commonly observed in many malignancies (Bhavanandan, 1991; Mukhopadhyay *et al.*, 2011; Tian & Ten Hagen, 2009). Currently, there is no reliable assay to estimate total *O*-glycosylated proteins in complex biological samples such as serum. Immunological based method has narrow specificities, which is not suitable for detection of total *O*-glycosylated proteins. Hence, in this part of the study, a method to estimate levels of *O*-glycosylated proteins in biological samples was designed based on binding of lectins to the core structure of *O*-glycans. The initial strategy employed was removal of the heterogeneous peripheral and core saccharides using chemical treatments to expose the cryptic sugar-amino acid linkage. This was then followed by selection of suitable lectins for coating and capturing of *O*-glycosylated proteins in the developed Enzyme-linked Lectin Assay (ELLA).

4.1.1 Purification of CGB lectin

CGB lectin was purified in-house. The lectin was isolated from crude extract of the seeds of champedak using galactose conjugated Sepharose affinity chromatography (Section 3.3.3). Typically, each fraction of 10 ml was collected and the protein content was estimated by reading its absorbance at 280 nm. The crude extract and the elution buffer were added into the column at the fraction numbers as indicated in Figure 4.1. A baseline absorbance value of close to zero was necessary before addition of the crude extract and elution buffer.

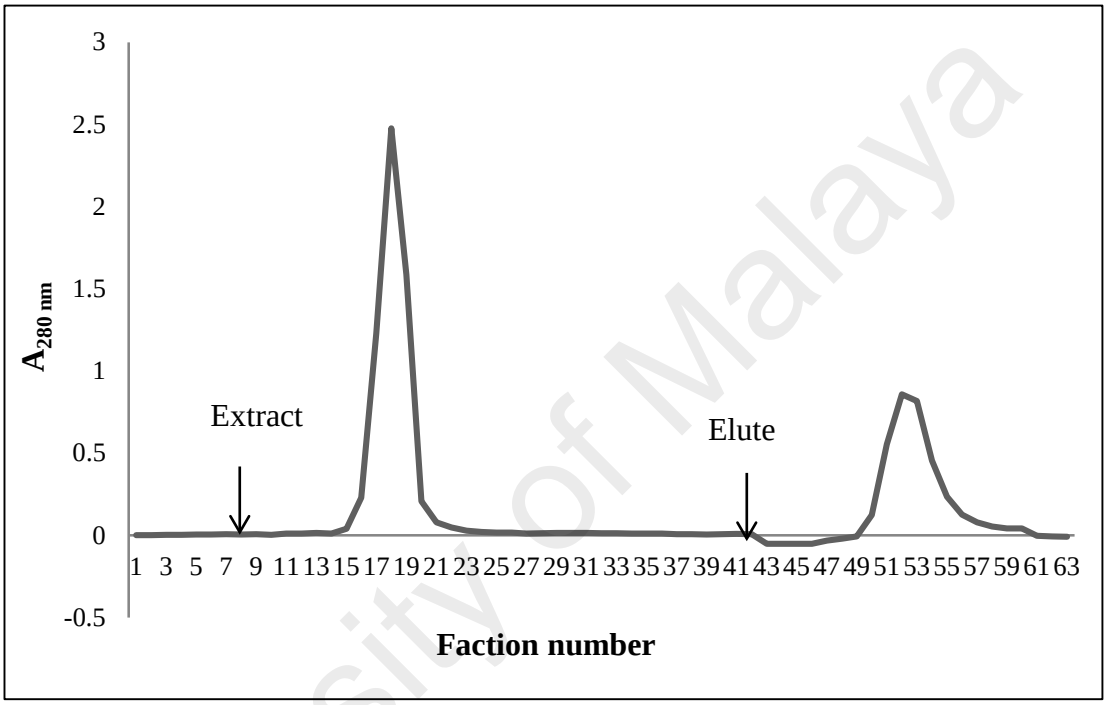


Figure 4.1: Isolation of CGB lectin using galactose-conjugated Sepharose column.

The crude extract of the seeds of champedak was loaded into an immobilised galactose column after a series of washing. Each fraction of 10 ml was collected and the protein content was measured for its absorbance at 280 nm. Unbound material was removed by washing with PBS until a baseline level of near zero was achieved. CGB lectin was dissociated from the immobilised galactose column using 0.8 M of D-galactose in PBS solution.

A slight modification to the previously established protocol for isolation of CGB lectin (Hashim *et al.*, 1991) was made to achieve greater efficiency and higher yield. To avoid contamination, the column was firstly washed with at least five bed column volumes (approximately 500 ml) of PBS before introduction of the crude extract. To ensure that no contaminant was present, 1 ml fractions were collected after washing and absorbance at 280 nm was measured. Soon after the crude extract was introduced, all unbound materials were discarded. The column was again washed with another five bed column volumes of PBS to remove all unbound materials. D-galactose (0.8 M) solution was added to elute the CGB lectin from the column. The first 50 ml of eluate was discarded as it only contained PBS. The subsequent 90 ml of eluate was collected, with presence of the lectin. The eluted CGB lectin was further dialysed against 0.5X PBS and concentrated using freeze drying technique. After lyophilisation, the powdered CGB lectin was reconstituted in minimal volume of dH_2O and further dialysed against 0.5X PBS for two more days. This apparently increased the lectin concentration, which successfully yielded about 5 mg of CGB lectin per 10 mg of dried powder.

4.1.2 CGB lectin purity

To determine for its purity, SDS-PAGE analysis of the purified CGB lectin was performed. In this experiment, Jacalin and VVA lectin, which were sourced commercially, were also resolved using the SDS-PAGE for comparison. Figure 4.2 demonstrates the result of the experiment, with the CGB lectin having the highest purity without any apparent contamination. In contrast, both jacalin and VVA showed presence of multiple non-lectin peptide bands, and are therefore of lesser purity despite that they were commercially prepared. This clearly shows that the lectin isolation

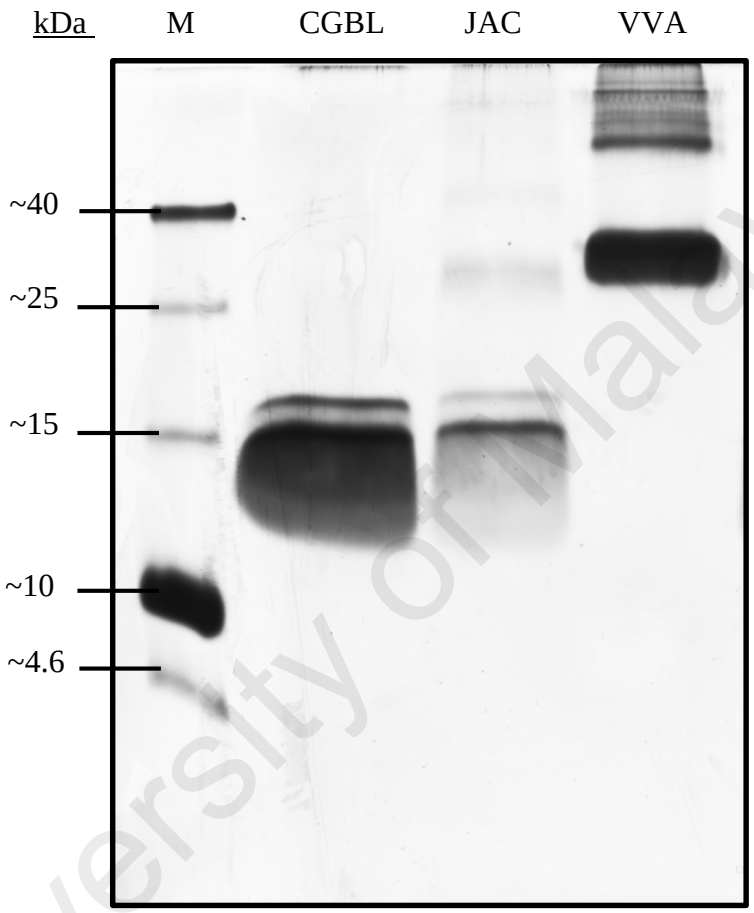


Figure 4.2: Analysis of CGB lectin, jacalin and VVA by SDS-PAGE.

Two μg of lectins were resolved by electrophoresis in 18% polyacrylamide gel. M refers to low range protein markers. CGBL; CGB lectin, JAC; Jacalin, VVA; *Vicia villosa* agglutinin.

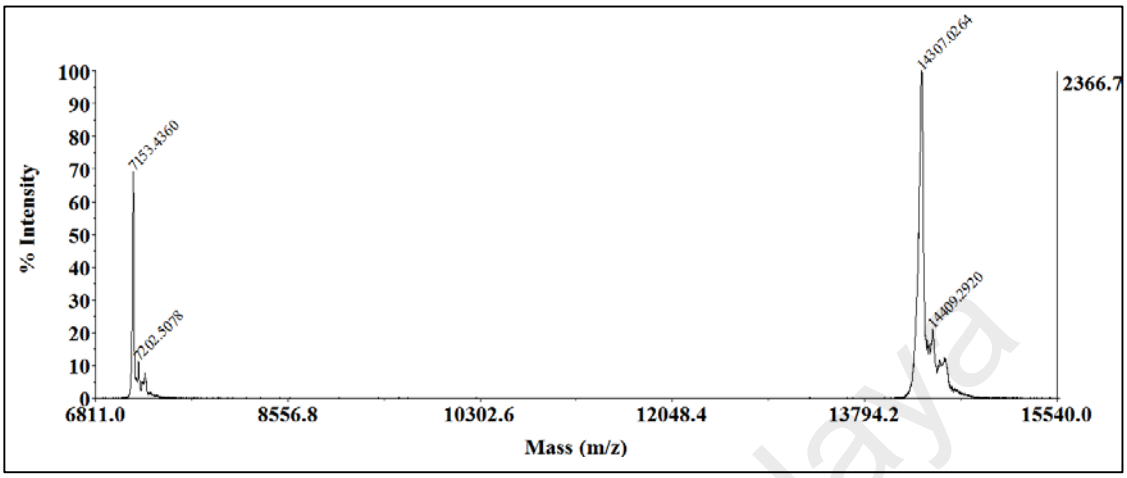
method that was adopted in the present study was able to provide a pure and almost contaminant free lectin. Both the CGB lectin and Jacalin were seen sharing similar 16 kDa and 14 kDa subunits.

CGB lectin was subjected to linear mode MALDI-MS analysis for further detection of possible contamination. The mass spectrometry system was first calibrated using commercial lysozyme with molecular weight of 14,307 Da. Figure 4.3a shows the mass spectrum of lysozyme protein. Only two clustered peaks were generated showing high purity of the protein and that the instrument was functioning correctly. The CGB lectin showed two clustered peaks at 14.6 kDa and 15.8 kDa with their respective doubly charge ion peaks at 7.3 kDa and 7.9 kDa (Figure 4.3b). No other apparent peak was detected, which indicate a high purity of the isolated lectin.

4.1.3 Binding properties of CGB lectin

To assess the binding properties of the purified CGB lectin, a method similar to radial immuno diffusion technique was performed using the purified CGB lectin. Two known *O*-glycosylated proteins (BSM and PGM) and two *N*-glycosylated proteins (AGP and Ovomucoid) were used as test glycoproteins. As expected, the results showed that the CGB lectin had strongly interacted with both the BSM and PGM (Figure 4.4). These were seen as dark gray precipitin lines between the lectin and the test glycoproteins. No interaction was observed between the CGB lectin and both the test *N*-glycosylated proteins.

a)



b)

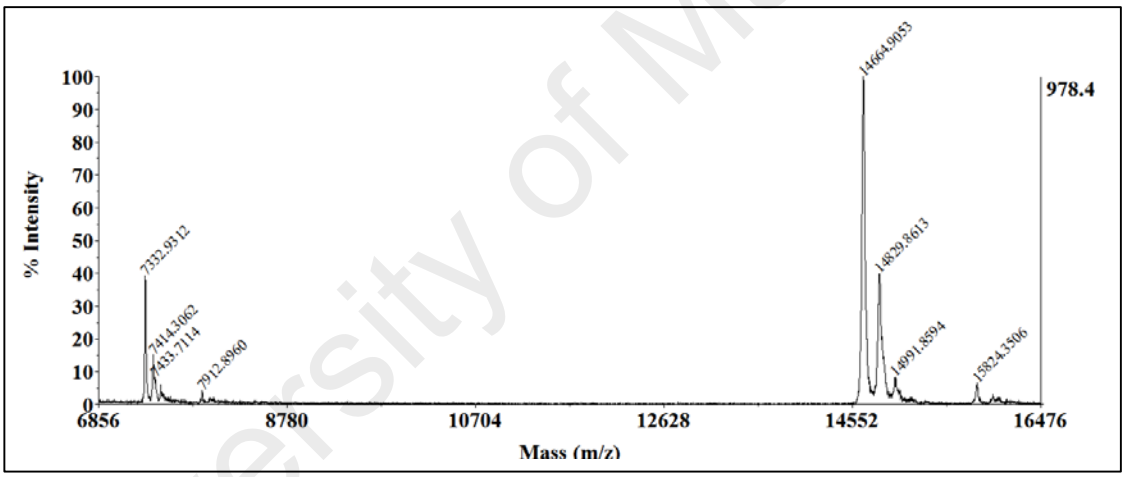


Figure 4.3: Linear mode MALDI-TOF analysis of CGB lectin.

The instrument was first calibrated with lysozyme (14,307 Da) (panel a). The laser intensity was set at 5500 and mass range was set at 7000 Da – 30,000 Da. Panel b shows the mass spectrum of CGB lectin (14.6 kDa and 15.8 kDa).

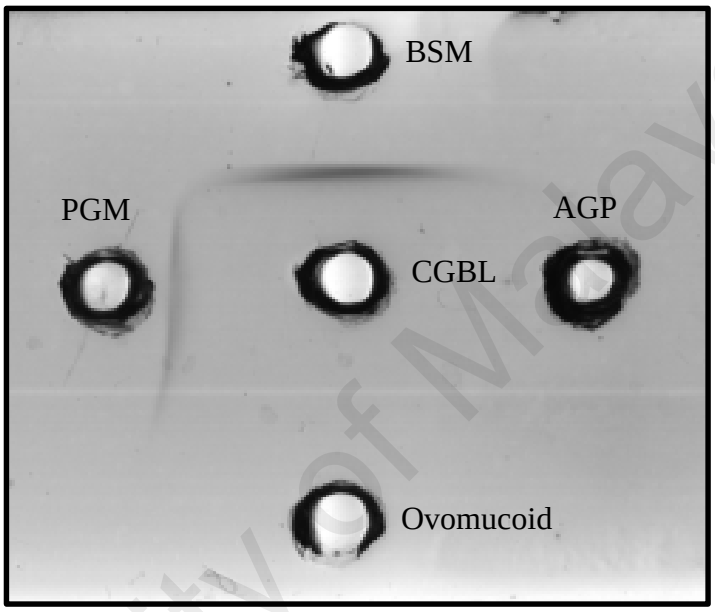


Figure 4.4: Radial immunodiffusion assay using CGB lectin with test glycoproteins.

Lectin radial diffusion assay was performed using 0.5% agarose gel. Five µg of sample was added into each well as indicated. Positive interaction of CGB lectin (CGBL) with test glycoproteins was determined as a gray precipitin line which appeared between the CGB lectin and test glycoprotein. BSM; Bovine submaxillary mucin, PGM; Pig gastric mucin, AGP; Alpha-1-acid glycoprotein.

4.1.4 Concentrations of monosaccharides in PGM after chemical treatment

The concentrations of the monosaccharides in the glycans of PGM after acid hydrolysis and Smith degradation were determined using HPAEC-PAD system. The results of the analysis, expressed as molar ratios relative to GalNAc are summarized in Table 4.1. As expected, mild acid hydrolysis removed all of the sialic acids from the glycans of PGM. This treatment also removed about 38% of fucose (Table 4.1). The GlcNAc, galactose and GalNAc contents were not significantly affected by mild acid hydrolysis (Table 4.1). Subjecting the desialylated samples to Smith degradation resulted in removal of varying amounts of the now exposed penultimate saccharides. Smith-degradation of asialoPGM resulted in loss of 100% fucose, 58% galactose and 9% GlcNAc (Table 4.1). About 100% of fucose, 49% of galactose, 8% of GlcNAc, and 33% of sialic acid were lost when PGM was directly subjected to Smith degradation. While all of the terminal fucose appears to be susceptible to periodate cleavage, about two-third of sialic acid residues seemed to be protected. This may be due to *O*-acetylation of the sialic acid residues (Varki, 1992).

4.1.5 Lectin binding of native and chemically-treated glycoproteins

The ability of biotinylated jacalin, VVA and CGB lectin to interact with native and chemically modified glycans of PGM, BSM, PSM and fetuin was evaluated by performing direct ELLA. In all cases, 50 µl of the test glycoconjugates at 1 µg protein/ml in coating buffer was used to coat the well. Biotinylated lectins were serially diluted at the range of 10 – 0.16 µg/ml for jacalin, and 20 – 0.31 µg/ml for VVA and CGB lectin. Figures 4.5, 4.6 and 4.7 illustrate the results of the experiments, which involved 12 different combinations of lectins and chemical treatments. Of the three lectins evaluated, jacalin demonstrated the highest detection sensitivity for all the native

Table 4.1: Carbohydrate composition of pig gastric mucin (PGM) after different chemical treatments to modify the saccharide chains.

	Molar Ratio of Monosaccharides (with GalNAc as 1.00)				
	GalNAc	**GlcNAc**	**Gal**	**Fuc**	**Sialic acid**
PGM	1.00	2.03	1.25	0.65	0.15
asialoPGM	1.00	1.87	1.23	0.40	0.00
[a] Smith-degraded PGM	1.00	1.86	0.64	0.00	0.10
[a] Smith-degraded asialoPGM	1.00	1.71	0.52	0.00	0.00

[a] Sample subjected to Smith Degradation as described in Section 3.5.1.

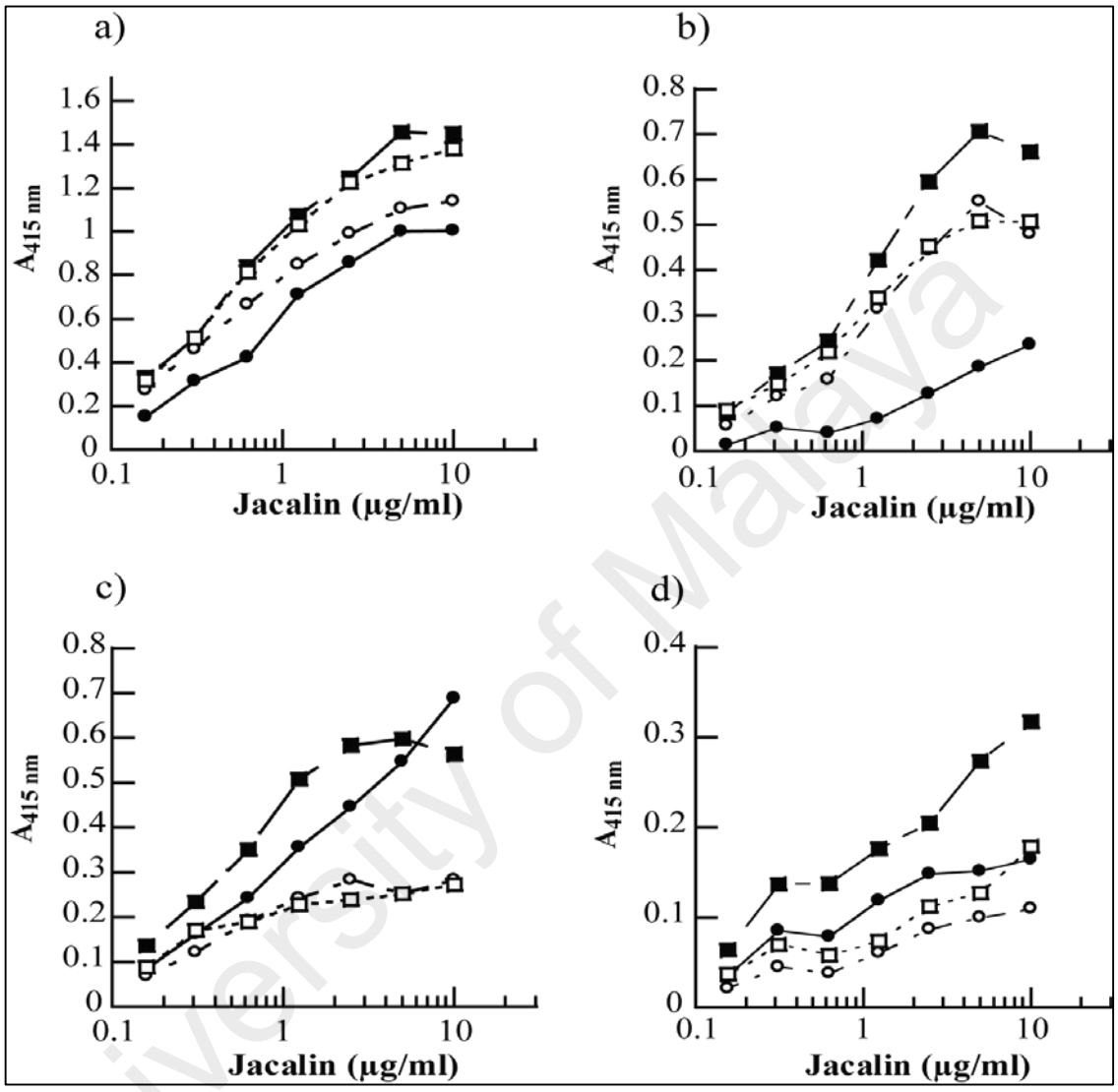


Figure 4.5: Binding of serially diluted biotinylated-jacalin to various glycoconjugates.

Panels a, b, c and d refer to PGM, BSM, PSM and fetuin, before (●) and after chemical treatment to modify the saccharide chains. The glycoconjugates were subjected to desialylation by mild acid hydrolysis (■), Smith degradation (○) and both desialylation and Smith degradation (□).

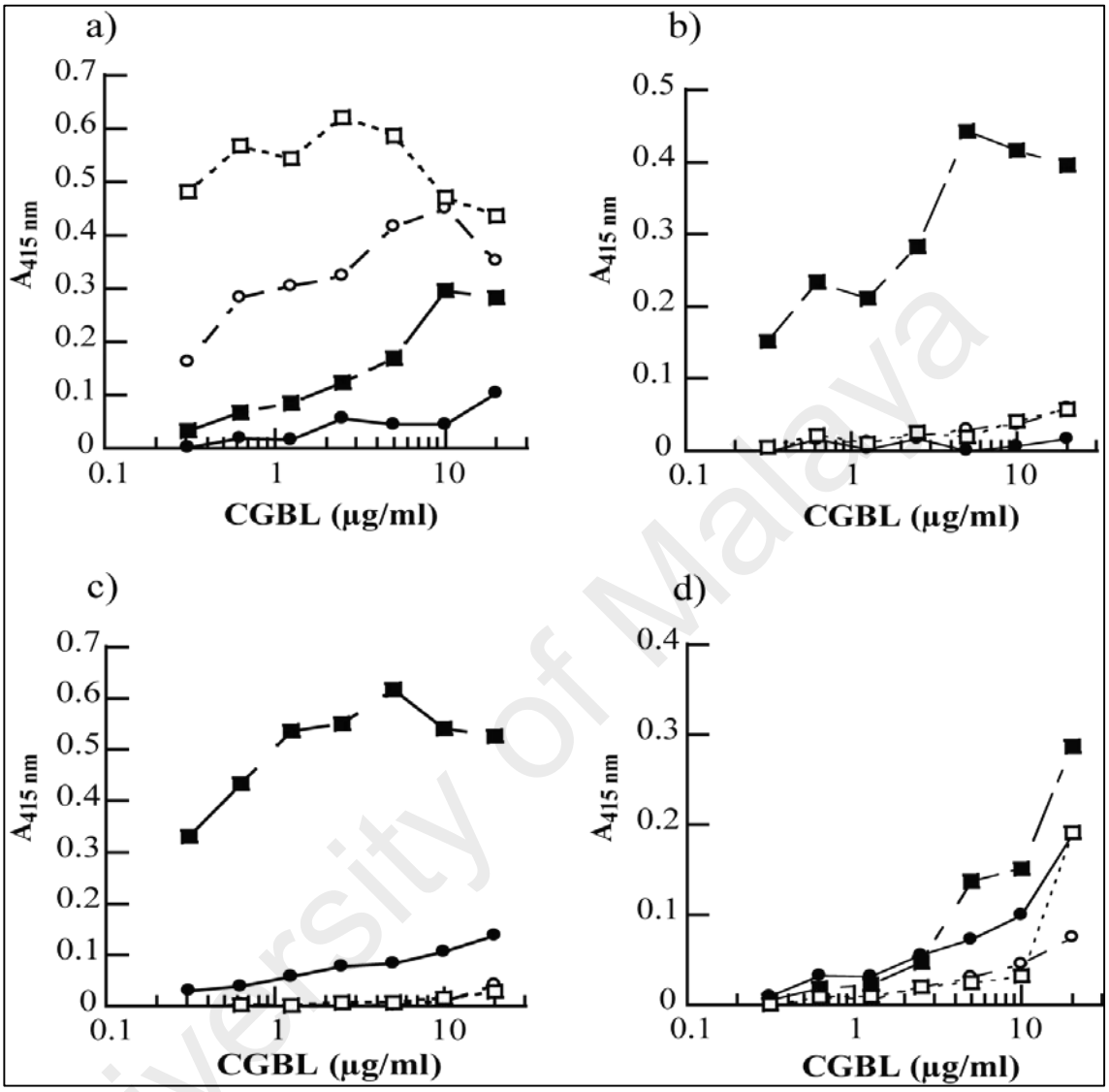


Figure 4.6: Binding of serially diluted biotinylated-CGB lectin to various glycoconjugates.

Panels a, b, c and d refer to PGM, BSM, PSM and fetuin, before (●) and after chemical treatment to modify the saccharide chains. The glycoconjugates were subjected to desialylation by mild acid hydrolysis (■), Smith degradation (○) and both desialylation and Smith degradation (□).

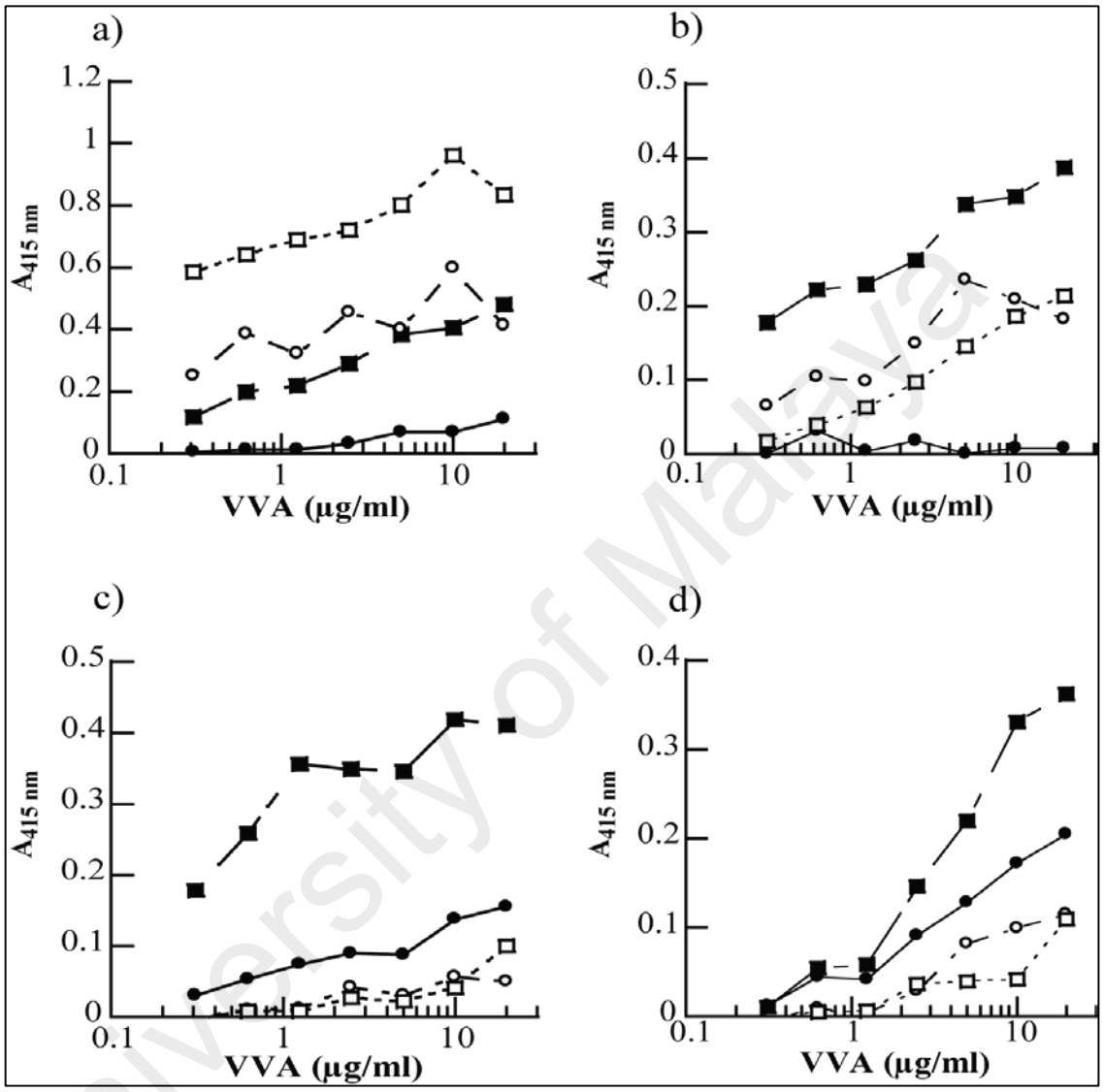


Figure 4.7: Binding of serially diluted biotinylated-VVA to various glycoconjugates.

Panels a, b, c and d refer to PGM, BSM, PSM and fetuin, before (●) and after chemical treatment to modify the saccharide chains. The glycoconjugates were subjected to desialylation by mild acid hydrolysis (■), Smith degradation (○) and both desialylation and Smith degradation (□).

glycoproteins tested. Binding of all three lectins was enhanced after desialylation by mild acid hydrolysis. For example, substantially increased binding of VVA to asialoBSM compared to BSM and CGB lectin to asialoPSM compared to PSM can be noted in Figures 4.7b and 4.6c, respectively. The influence of Smith-degradation on the binding of lectins to glycoproteins was highly variable. Smith-degradation improved the binding of all lectins to PGM and BSM (Figures 4.5a, 4.5b, 4.6a, 4.6b, 4.7a and 4.7b) but their binding to PSM and fetuin was drastically reduced (Figures 4.5c, 4.5d, 4.6c, 4.6d, 4.7c and 4.7d). The influence of Smith-degradation on the ability of the desialylated glycoprotein to interact with the three lectins was evaluated. Prior desialylation did not further improve the binding of jacalin to Smith degraded glycoproteins (Figure 4.5). In contrast, the binding of CGB lectin and VVA to asialoPGM was significantly increased after Smith degradation (Figures 4.6a and 4.7a). In almost all cases, binding of the lectins to the Smith-degraded glycoproteins were lower compared to the desialylated glycoproteins. Based on these results, further studies were carried out primarily using asialoPGM and asialoBSM as ligands and biotinylated jacalin as the probe lectin.

4.1.6 Direct versus sandwich ELLA

High levels of glycosylation in proteins are known to bind to plastic microtiter plates rather poorly. Thus, we performed a series of experiments to compare the binding of biotinylated jacalin to the test glycoproteins that were directly coated onto microtiter plates versus those that interacted with pre-coated lectins (*i.e.* direct ELLA versus sandwich ELLA). The optimal concentration of lectin for coating the plates was first determined. In a typical experiment, 0.25, 0.50, 1.00 and 2.00 µg of jacalin in 50 µl of coating buffer were coated onto the wells of the microtiter plates. After washing

to remove unbound lectin, the wells were incubated for 90 minutes with 50 µl of asialoPGM (1 µg protein/ml PBS). The plates were washed, and following that, incubated with a serial dilution of biotinylated jacalin at a starting concentration of 5 µg/ml and further processed as described in section 3.5.4 of Materials and Methods. The results revealed that absorbance increased by 47.0, 92.1 and 110.5% when the coating concentration of jacalin was increased from 0.25 to 0.50, 1.00 and 2.00 µg/well, respectively. Based on this, 1 µg lectin per well was chosen as the coating concentration for all subsequent experiments.

The binding of biotinylated jacalin to directly coated asialoBSM and asialoPGM was compared to those interacted with 1.0 µg of pre-coated jacalin, CGB lectin or VVA. Our results showed that pre-coating of wells with a lectin substantially increased the binding of sensitivity of the ELLA for the ligands tested (Figure 4.8). Thus, the binding of biotinylated jacalin to asialoBSM and asialoPGM was substantially increased when they were bound to the lectin pre-coated wells compared to the uncoated wells (Figures 4.8a and 4.8b). For both cases of asialoBSM and asialoPGM, CGB lectin coated wells gave the highest reading followed by jacalin. When VVA was used as the coating lectin, increased sensitivity was obtained with asialoBSM but not asialoPGM.

4.1.7 Effect of *N*-glycosylated protein on sandwich ELLA for *O*-glycosylated protein

The interference of *N*-glycosylated proteins in the developed sandwich ELLA was tested using asialoAGP and ovomucoid. These *N*-glycosylated proteins did not show significant interaction in the CGB lectin-jacalin (Figure 4.9a) and jacalin-jacalin sandwich ELLA (Figure 4.9b) even at a concentration of 1000 ng protein/ml. The

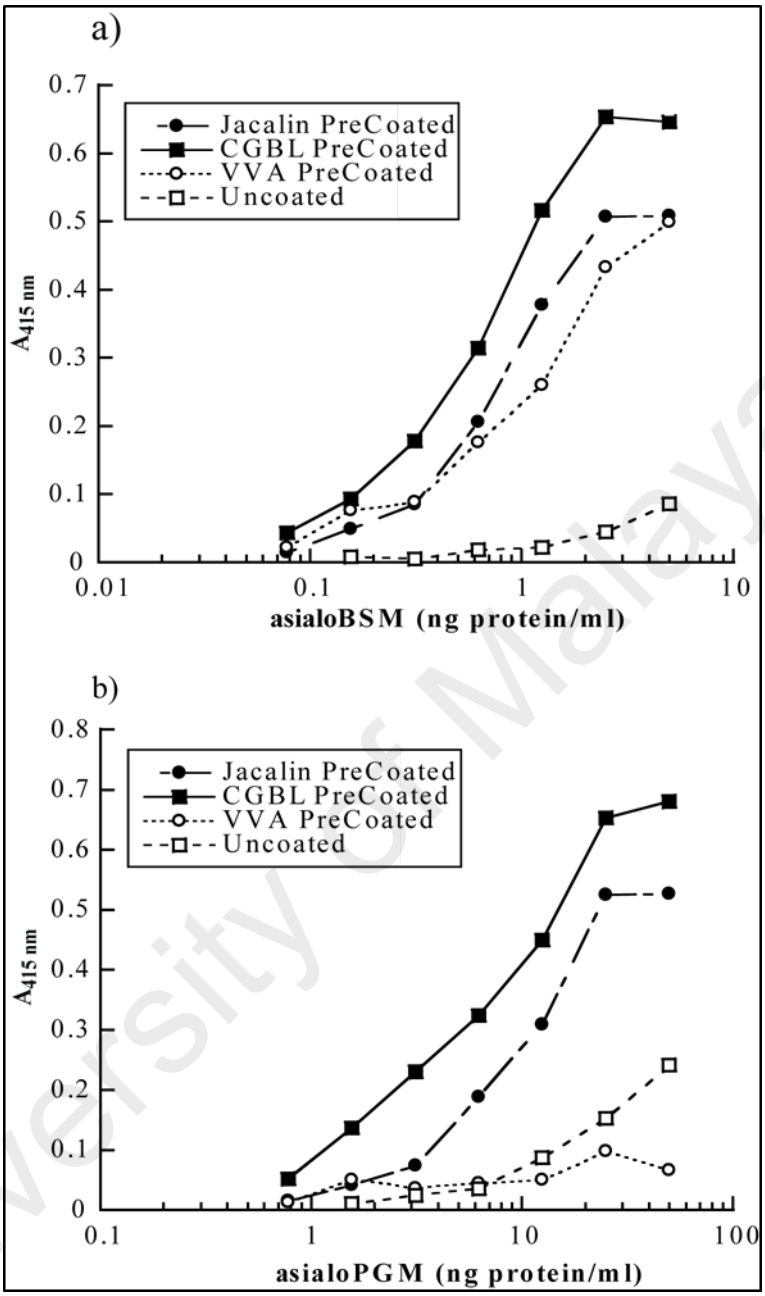


Figure 4.8: Development of ELLA using microtiter plates precoated with different lectins.

Binding of serially diluted biotinylated jacalin to desialylated BSM (panel a) and PGM (panel b) was tested using microtiter plates that were either not precoated or precoated with jacalin, CGB lectin (CGBL) and VVA.

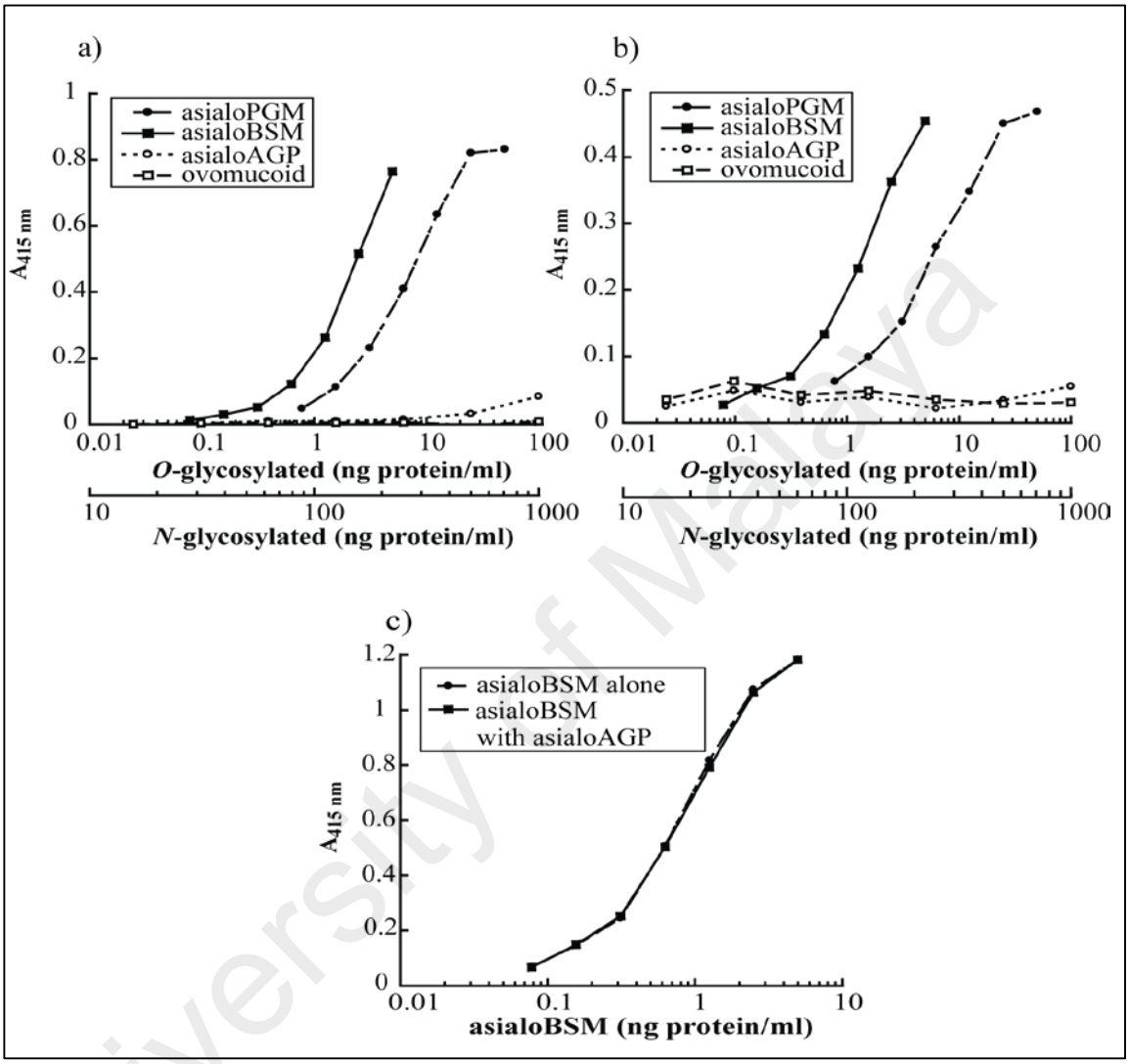


Figure 4.9: Influence of *N*-glycosylated proteins on the sandwich ELLA for *O*-glycosylated proteins.

AsialoAGP and ovomucoid show minimum binding in CGB lectin-jacalin (panel a) and jacalin-jacalin (panel b) sandwich ELLA for asialoPGM and asialoBSM. Panel c illustrates the results of CGB lectin-jacalin sandwich ELLA on serially diluted solution of asialoBSM (5 ng/ml) and a mixture of asialoBSM (5 ng protein/ml) and asialoAGP (500 ng protein/ml).

presence of 500 ng protein/ml of asialoAGP during CGB-jacalin sandwich ELLA of 5 ng protein/ml of asialoBSM was evaluated. The results presented in Figure 4.9c clearly showed that the presence of a 100-fold excess of an *N*-glycosylated protein had hardly any effect in the CGB lectin-jacalin sandwich ELLA of an *O*-glycosylated protein.

4.1.8 Detection of *O*-glycosylated proteins in serum samples

To evaluate the applicability of the developed sandwich ELLA for estimation of *O*-glycosylated proteins in crude samples, a few preliminary experiments using human serum was carried out. Since desialylation increased binding of the lectins to all the test glycoconjugates, the serum samples were firstly subjected to mild acid hydrolysis. The effect of desialylation on serum samples was demonstrated using CGB lectin-jacalin sandwich ELLA (Figure 4.10). The result clearly showed that desialylation with acid hydrolysis improved the detection of *O*-glycosylated proteins. The desialylated serum samples were then subjected to direct ELLA and sandwich ELLA using biotinylated jacalin as the probe lectin and CGB lectin, jacalin and VVA as the coating lectins. Typical results of the binding of biotinylated jacalin to the desialylated *O*-glycosylated proteins in serum are illustrated in Figure 4.11. It can be observed that the absorbance readings are highest when wells were precoated with CGB lectin. Relatively high absorbance readings were also obtained with wells coated with jacalin but the readings were very low when wells were uncoated or coated with VVA.

Next, three human serum samples (N1, N2 and N3) were desialylated and subjected to sandwich ELLA using biotinylated jacalin as the probe lectin, and jacalin and CGB lectin as coating lectins. In each assay, asialoBSM in the range of 0.08 to 5.0 µg protein/mL was included as the standard. Typical results obtained for one serum

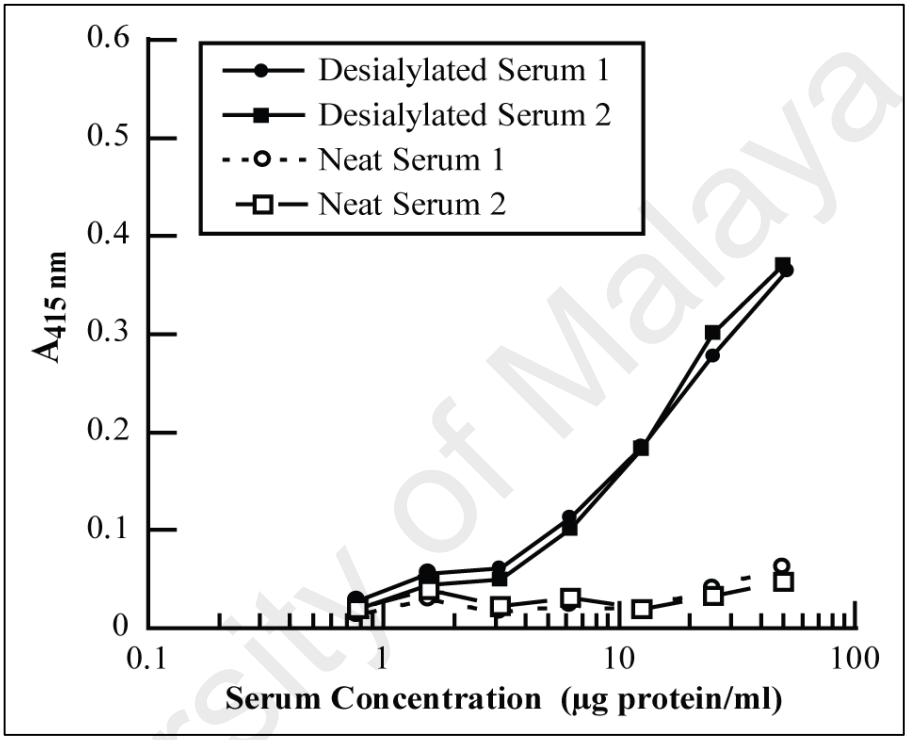


Figure 4.10: Effect of mild acid hydrolysis on test serum sample.

Serum sample was diluted and subjected to mild acid hydrolysis treatment. Neat serum and acid hydrolysed serum (Desialylated serum) were tested for total *O*-glycosylated protein estimation using CGB lectin-jacalin sandwich ELLA. Mild acid hydrolysis substantially increases the detection of *O*-glycosylated protein (——).

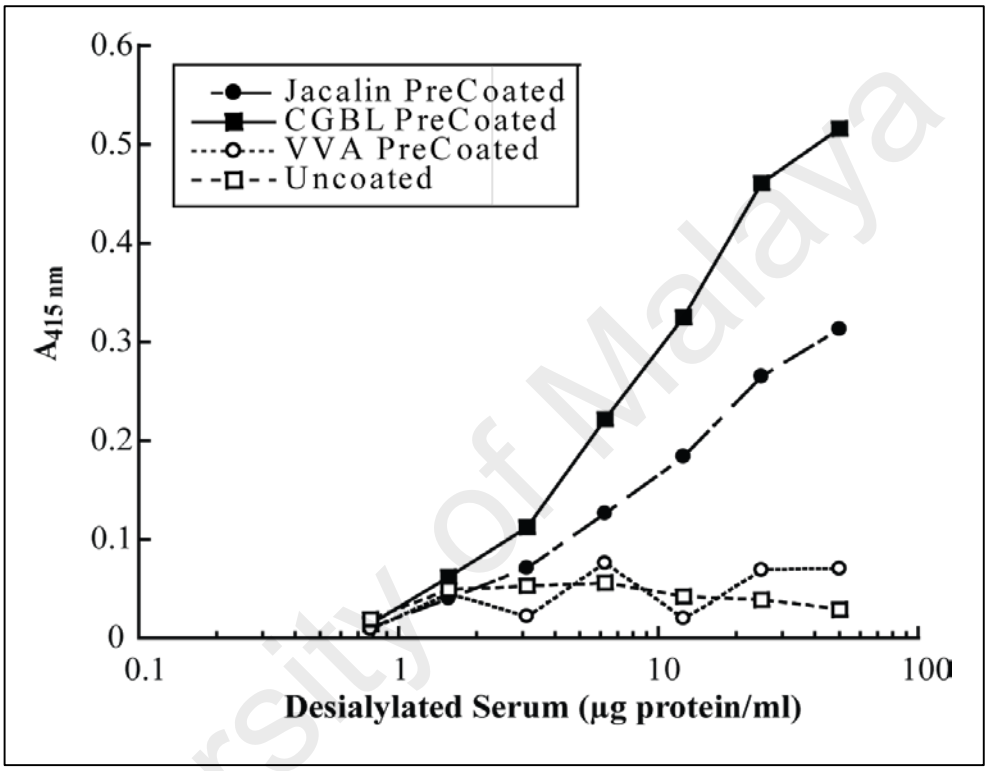


Figure 4.11: Development of ELLA using microtiter plates precoated with different lectins on serum.

ELLA was tested on normal human serum samples using microtiter plates that were either not precoated or precoated with jacalin, CGB lectin (CGBL) and VVA.

sample (N2) and the asialoBSM standard by jacalin-jacalin sandwich ELLA and CGB lectin-jacalin sandwich ELLA are illustrated in Figure 4.12. Using the results of the jacalin-jacalin sandwich ELLA (Figure 4.12a and 4.12b) the *O*-glycosylated protein content of N1, N2 and N3 were calculated to be 36.0, 36.1 and 36.8 pg asialoBSM protein/µg serum protein, respectively. The corresponding values calculated from the results of the CGB lectin-jacalin sandwich ELLA (Figure 4.12c and Figure 4.12d) were 27.7, 30.3 and 31.1 pg asialoBSM protein/µg serum protein, respectively. Since the CGB lectin had the highest purity (Figure 4.2) and greater specificity compared to jacalin, the CGB lectin-jacalin sandwich ELLA method appeared more preferable to estimate the levels of *O*-glycosylated proteins in serum. For further downstream applications, the CGB lectin-jacalin sandwich ELLA was therefore used.

4.1.9 Assay reproducibility

A random selected serum sample was used to determine the assay variation. Eight replicates of the assay were performed on the same day (intra-assay) as well as on different days (inter-assay). AsialoBSM was used as *O*-glycosylated protein standard. The percentage of coefficient of variant (CV) was less than 6% for intra-assay, while the inter-assay variation was less than 10% (Table 4.2). This indicates that the developed assay was reproducible and reliable to measure *O*-glycosylated proteins in serum samples.

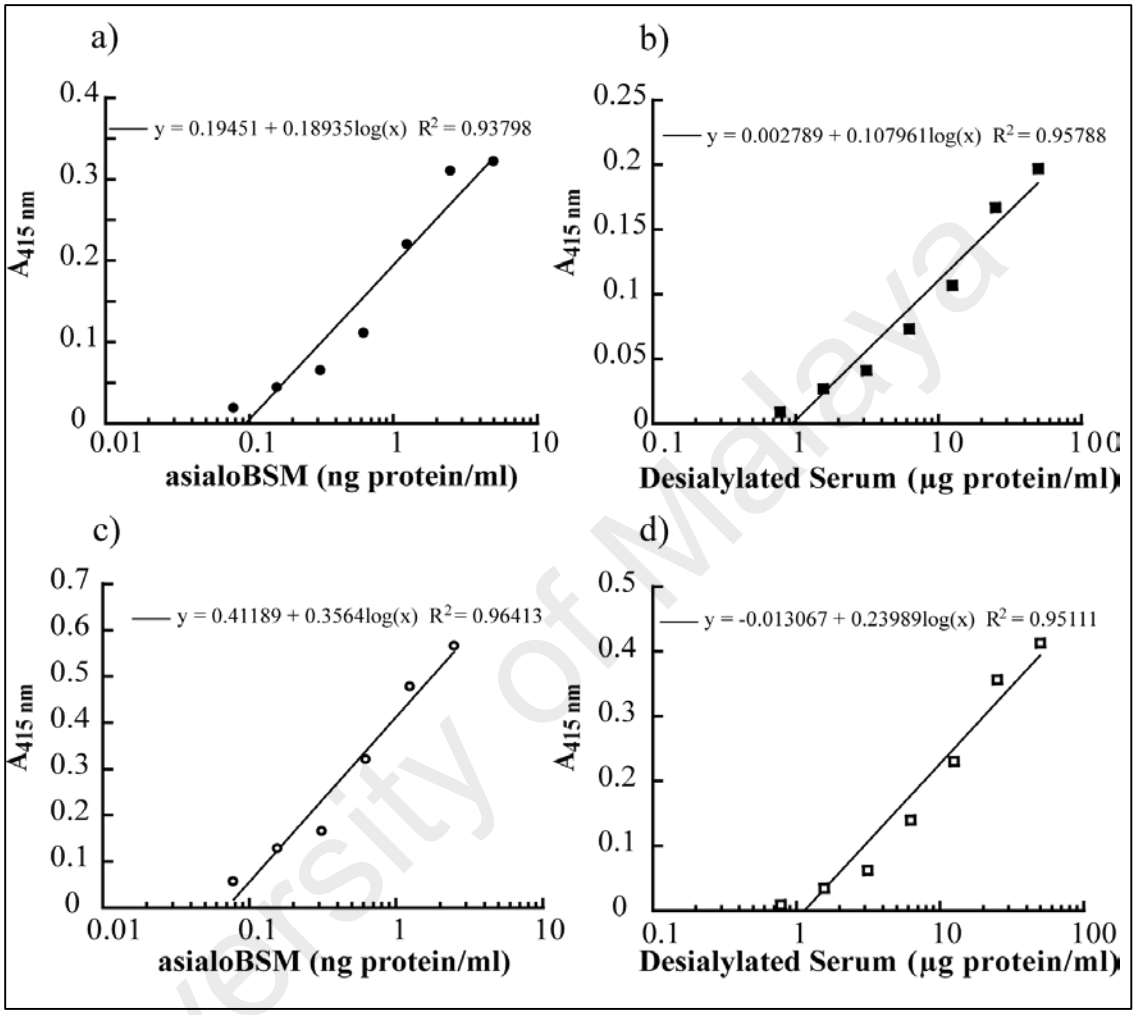


Figure 4.12: Estimation of *O*-glycosylated protein in human serum samples subjected to mild acid treatment.

Sandwich ELLA was performed using asialoBSM as the standard *O*-glycosylated protein (panels a and c). The *O*-glycosylated protein of desialylated serum samples were estimated using jacalin (panels a and b) or CGB lectin (panels c and d) as the coating lectin and biotinylated jacalin as the probe lectin.

Table 4.2: Reproducibility of the developed sandwich ELLA

	Intra-assay			Inter-assay		
	Mean (pg mucin/µg serum)	**SD**	**CV (%)**	**Mean (pg mucin/µg serum)**	**SD**	**CV (%)**
Serum *O*-glycosylated protein	60.81	3.59	5.90	69.34	6.28	9.06

SD, Standard Deviation; CV, Coefficient of Variation

4.2 Enrichment of *O*-glycosylated proteins using perchloric acid

In the previous section, a method to estimate total *O*-glycosylated proteins in serum was successfully designed. The method was based on the use of CGB lectin as pre-capture lectin to enhance detection. However, samples such as serum contain significant amounts of IgA1, an *O*-glycosylated protein, which may interfere the binding of other low abundance *O*-glycosylated proteins to interact with CGB lectin. Therefore, a strategy to enrich heavily *O*-glycosylated proteins using perchloric acid was employed. The use of perchloric acid was first demonstrated by Winzler *et al.* (1948), which successfully isolated heavily glycosylated proteins from serum samples. The following part of the study demonstrated the efficiency of perchloric acid to isolate substantial amount of glycoproteins from serum that do not include IgA1 and thus, allowing the discovery of proteins that may not be easily detected using typical proteomic platforms.

4.2.1 Perchloric acid treatment of human serum sample

Human serum was treated with four different normalities of perchloric acid (2 N, 1 N, 0.5 N and 0.25 N) to determine the optimal concentration for isolation of proteins. Figure 4.13 shows the amounts of protein obtained after serum was treated with different normalities of perchloric acid. Serum samples treated with 0.25 N and 0.5 N of perchloric acid produce relatively higher concentrations of protein, which appeared constant when higher normalities of the acid were used. The optimal normality of perchloric acid that was needed to solubilise a practical amount of protein from serum was determined as 0.6 N. This same perchloric acid normality has been previously used in studies of mucoproteins in serum samples (Winzler *et al.*, 1948).

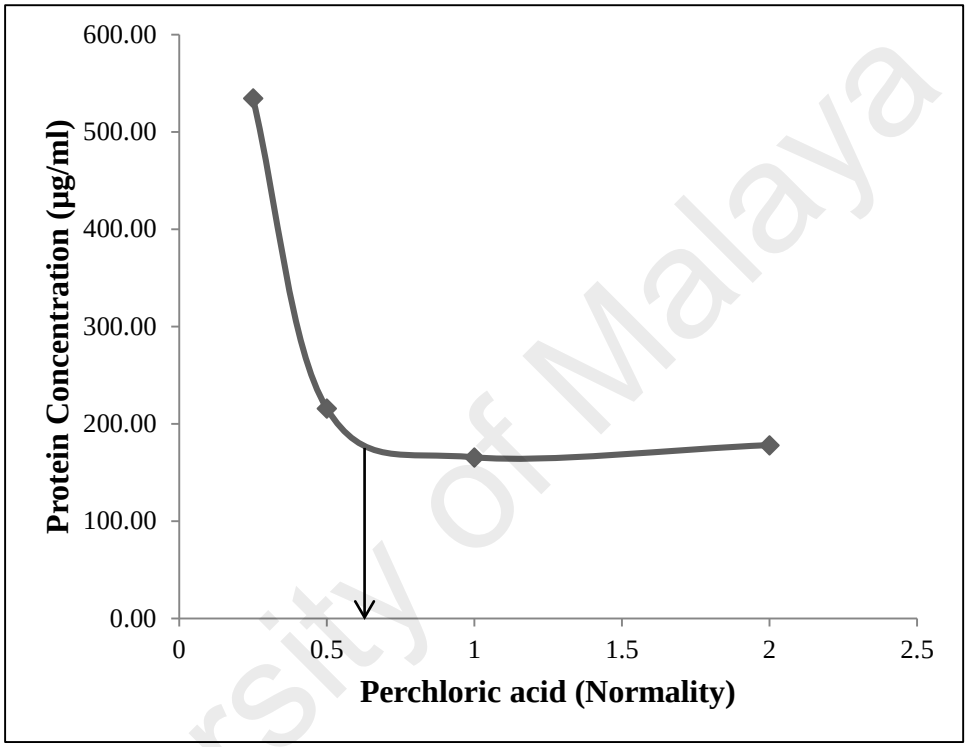


Figure 4.13: Concentrations of protein in serum treated with different normalities of perchloric acid.

Fifty µl of serum sample was subjected to four different normalities of perchloric acid precipitation. The concentrations of protein in the supernatant were estimated using standard BCA assay.

The amount of protein recovered from serum after 0.6 N of perchloric acid treatment was calculated as approximately 1%.

4.2.2 Proteomic analysis of serum perchloric acid isolates

About 600 µl of pooled serum from healthy donors was subjected to perchloric acid precipitation. Protein from the supernatant was purified by dialysis and lyophilisation. Dried protein sample was then reconstituted into 20 µl of ddH_2O. Each 600 µl of neat serum yielded about 100 µg of protein. Isoelectric focusing was performed on the 11 cm IPG strip of pI 3 – 10 and further resolved on 8 – 18 % SDS-polyacrylamide gel. Figure 4.14a depicts a typical silver stained 2-DE gel profile of the serum perchloric acid protein isolates. As expected, these proteins, which were solubilized in perchloric acid, were resolved at the pH 3 to 7 region. In comparison with the 2-DE gel profile of neat serum (Figure 4.14b), minimal immunoglobulins appeared to be present in the serum perchloric acid isolate sample. The analysis shows that perchloric acid treatment was able to reduce a significant amount of immunoglobulins from the serum sample.

Peptide spots from serum perchloric acid isolate 2DE gel were excised and submitted to mass spectrometry analysis for protein identification. A total of 31 spots were successfully identified. Details of all the identified proteins are listed in Table 4.3. Proteins having different isoforms, such as the beta-2-glycoprotein, were often detected as multiple peptide spots aligned in a train-like manner (spot numbers 12-17). Although 31 peptide spots were successfully identified, some appeared to be different isoforms or species of the same protein. Therefore, a total of 18 distinctive proteins was finally identified in the serum perchloric acid isolate that was resolved by 2DE and subjected to mass spectrometry analysis and database query.

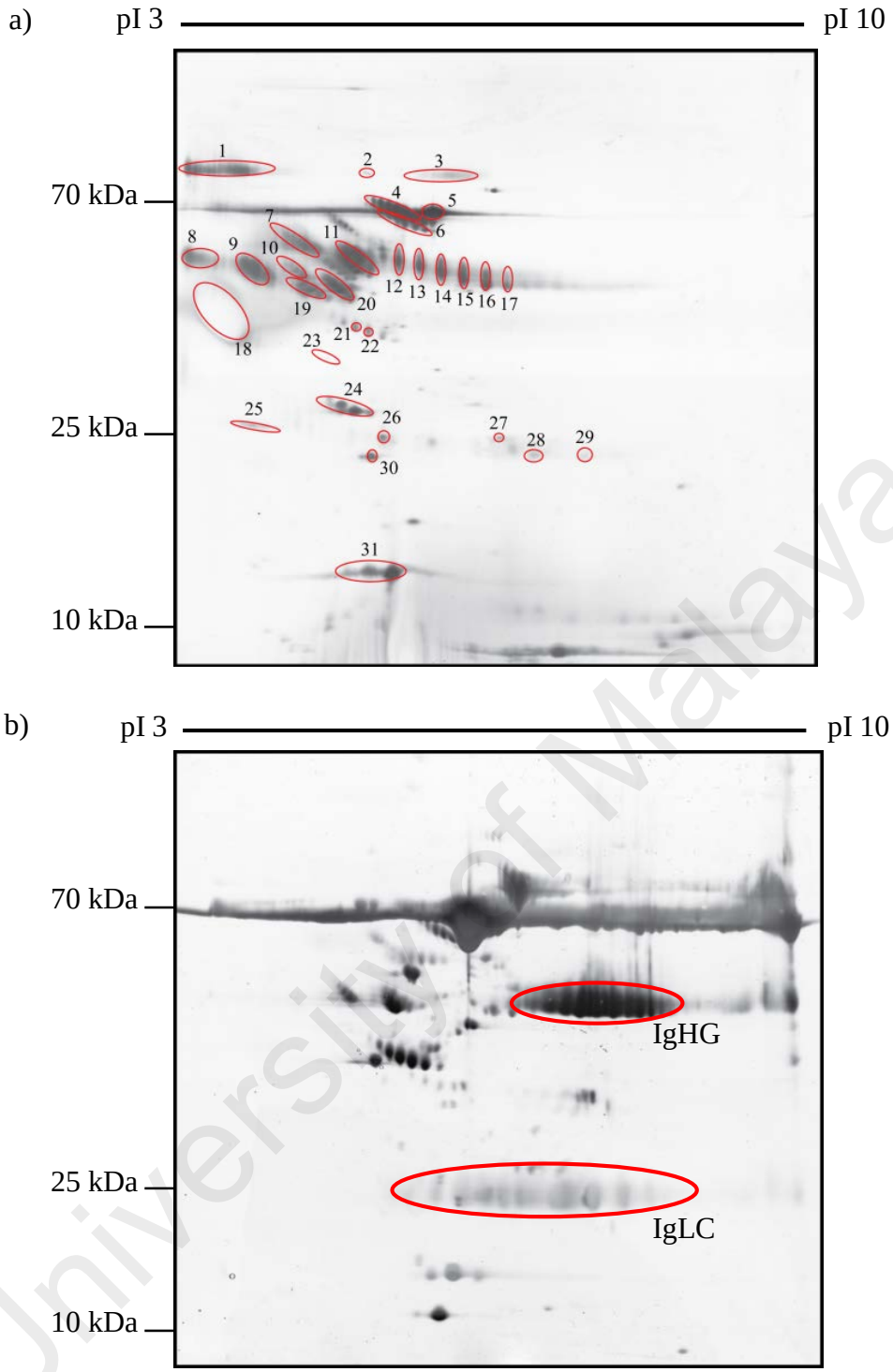


Figure 4.14: Typical 2-DE profiles of serum perchloric acid isolates and neat serum.

Serum perchloric acid protein isolates (100 µg) was focused on a pH 3-10 IPG strips (11 cm) followed by separation on 8-18% SDS-polyacrylamide gel (panel a). Spots (numbered 1-31) were picked and subjected to MALDI TOF/TOF analysis. Identities of the proteins are listed in Table 3.3. For comparison, 100 µg of normal serum proteins were subjected to similar 2-DE experiment (panel b). Immunoglobulin spots are circled in red. IgHG: Immunoglobulin gamma heavy chain, IgLC: Immunoglobulin light chain.

Table 4.3: Proteins from 2-Dimentional Electrophoresis identified by mass spectrometry analysis

Spot No.	Accession No.	Abbreviation	Protein Name	Mascot score	Sequence coverage	Theoretical Mass	Theoretical pI	Glycan	*Ref
1	P02763	A1AG	Alpha-1-acid glycoprotein	92	11%	23497	4.93	*N*-	[1-4]
2	Q9BSJ2	GCP2	Gamma-tubulin complex component 2	45	23%	102469	6.38	-	-
3	Q8N9B4	ANR42	Ankyrin repeat domain-containing protein 42	49	34%	43024	6.01	-	-
4	P02790	HEMO	Hemopexin	161	17%	51643	6.55	*N*-, *O*-	[5, 6]
5	P02768	ALBU	Serum Albumin	561	48%	69321	5.92	-	-
6	P02790	HEMO	Hemopexin	493	44%	51643	6.55	*N*-, *O*-	[5, 6]
7	P08185	CBG	Corticosteroid-binding globulin	56	28%	45112	5.64	*N*-	[3, 4]
8	P02763	A1AG	Alpha-1-acid glycoprotein	90	29%	23588	5.03	*N*-	[1-4]
9	P01011	AACT	Alpha-1-antichymotrypsin	76	31%	47621	5.33	*N*-	[2, 7]
10	P02765	FETUA	Alpha-2-HS glycoprotein	83	21%	39300	5.43	*N*-, *O*-	[1-3]
11	P01008	ANT3	Antithrombin III	340	39%	52605	6.32	*N*-	[2, 3]
12	P02749	APOH	Beta-2-glycoprotein 1	72	46%	38273	8.34	*N*-	[2, 3]
13	P02749	APOH	Beta-2-glycoprotein 1	92	38%	38273	8.34	*N*-	[2, 3]
14	P02749	APOH	Beta-2-glycoprotein 1	146	26%	38273	8.34	*N*-	[2, 3]

(Table 4.3, continued)

15	P02749	APOH	Beta-2-glycoprotein 1	228	46%	38273	8.34	*N*-	[2, 3]
16	P02749	APOH	Beta-2-glycoprotein 1	71	33%	38273	8.34	*N*-	[2, 3]
17	P02749	APOH	Beta-2-glycoprotein 1	130	21%	38273	8.34	*N*-	[2, 3]
18	P02763	A1AG	Alpha-1-acid glycoprotein	296	34%	23497	4.93	*N*-	[1-4]
19	P01009	A1AT	Alpha-1-antitrypsin	20	7%	46707	5.37	*N*-	[1-3]
20	P01009	A1AT	Alpha-1-antitrypsin	106	11%	46707	5.37	*N*-	[1-3]
21	P00738	HPT	Haptoglobin	111	10%	45177	6.13	*N*-	[6, 5]
22	P00738	HPT	Haptoglobin	110	14%	45177	6.13	*N*-	[6, 5]
23	P10909	CLUS	Clusterin	107	14%	52461	5.89	*N*-	[5, 8]
24	P02760	AMBP	Protein AMBP	88	4%	38974	5.95	*N*-, *O*-	[1]
25	P00734	THRB	Prothrombin	300	12%	69992	5.64	*N*-	[1-3]
26	P41222	PTGDS	Prostaglandin-H2 D-isomerase	59	23%	21015	7.66	*N*-, *O*-	[2, 8]
27	P01834	IGKC	Ig Kappa chain C region	76	15%	11602	5.58	-	-
28	P01834	IGKC	Ig Kappa chain C region	93	32%	11602	5.58	-	-
29	P01834	IGKC	Ig Kappa chain C region	92	32%	11602	5.58	-	-
30	P02647	APOA1	Apolipoprotein A-I	100	7%	30759	5.56	-	-
31	P02766	TTHY	Transthyretin	703	77%	15877	5.52	*N*-	[5]

* References of protein glycosylations: [1.] (Halim *et al.*, 2012) [2.] (Liu *et al.*, 2005b) [3.] (Bunkenborg *et al.*, 2004) [4.] (Treuheit *et al.*, 1992) [5.] (Halim *et al.*, 2013) [6.] (Nilsson *et al.*, 2009) [7.] (Zhang *et al.*, 2003) [8.] (Chen *et al.*, 2009)

4.2.3 LC/MS analysis of proteins in serum perchloric acid isolates

In addition to 2-DE, QTOF LC/MS analysis was also performed to identify proteins that may not have been detected by the electrophoretic separation. Fourteen proteins were successfully identified when serum perchloric acid isolates were analysed by QTOF LC/MS analysis. These isolates were initially subjected to dialysis and lyophilisation, followed by in-solution trypsin digestion. The list of proteins identified by QTOF LC/MS analysis is shown in Table 4.4. A total of 14 proteins were identified using this system. When compared to the 2-DE, eight proteins were similarly identified using both methods while five other proteins were exclusively identified by QTOF LC/MS (Figure 4.15). Ten proteins that were identified using 2-DE method were apparently not detected by QTOF LC/MS (Figure 4.15).

4.2.4 Protein glycosylation of serum perchloric acid isolates

To determine the types of glycosylation, all proteins which were identified using 2DE and/or QTOF LC/MS were matched with the UNIPROT database. Presence of glycosylation was identified and the types of glycosylation in the proteins are listed in Table 4.3 and Table 4.4. Among the detected proteins, almost all are known to be glycosylated. Plasma protease C1 inhibitor, hemopexin, alpha-2-HS glycoprotein, prostaglandin-H2, D-isomerase protein, AMBP, and complement C4-A were identified as those that are *O*-glycosylated whilst seventeen other identified proteins were documented to be *N*-glycosylated.

Table 4.4: QTOF LC/MS detected proteins in serum perchloric acid isolate.

Accession No.	Abbreviation	Protein Name	MS/MS search score	Number of Matched Sequence	Sequence Coverage (%)	Theoretical Mass (Da)	Glycan	*Ref.
P05155	IC1	Plasma protease C1 inhibitor	194.37	10	26.6	55381.4	*N*-, *O*-	[1-3]
P01009	A1AT	Alpha-1-antitrypsin	174.65	9	31.1	46906.8	*N*-	[4-6]
P02763	A1AG	Alpha-1-acid glycoprotein 1	104.45	5	27.8	23739.3	*N*-	[4-7]
P02749	APOH	Beta-2-glycoprotein 1	96.15	5	27.5	39609.7	*N*-	[5, 6]
P00734	THRB	Prothrombin	81.88	4	10.4	71519.0	*N*-	[4-6]
P02790	HEMO	Hemopexin	79.74	4	12.7	52417.1	*N*-, *O*-	[2, 8]
P02766	TTHY	Transthyretin	76.93	4	48.2	16000.8	*N*-	[5]
P02760	AMBP	Protein AMBP	55.81	3	11.3	39911.7	*N*-, *O*-	[4]
P06727	APOA4	Apolipoprotein A-IV	45.18	2	7.3	45398.2	-	-
P19652	A1AG	Alpha-1-acid glycoprotein 2	40.12	2	14.4	23887.5	*N*-	[5-7]
P0C0L4/ P0C0L5	CO4A/CO4B	Complement C4-A or Complement C4-B	34.82	2	0.8	194379.5	*N*-, *O*-	[2, 5, 6]
P08185	CBG	Corticosteroid-binding globulin	19.52	1	3.7	45311.2	*N*-	[5, 6]
P12259	FA5	Coagulation factor V	18.51	1	0.5	252839.9	*N*-	[5]
P02775	CXCL7	Platelet basic protein	17.09	1	7	14179.2	-	-

* References of protein glycosylations: [1.] (Bock *et al.*, 1986) [2.] (Halim *et al.*, 2013) [3.] (Liu *et al.*, 2005a) [4.] (Halim *et al.*, 2012) [5.] (Liu *et al.*, 2005b) [6.] (Bunkenborg *et al.*, 2004) [7.] (Treuheit *et al.*, 1992) [8.] (Nilsson *et al.*, 2009)

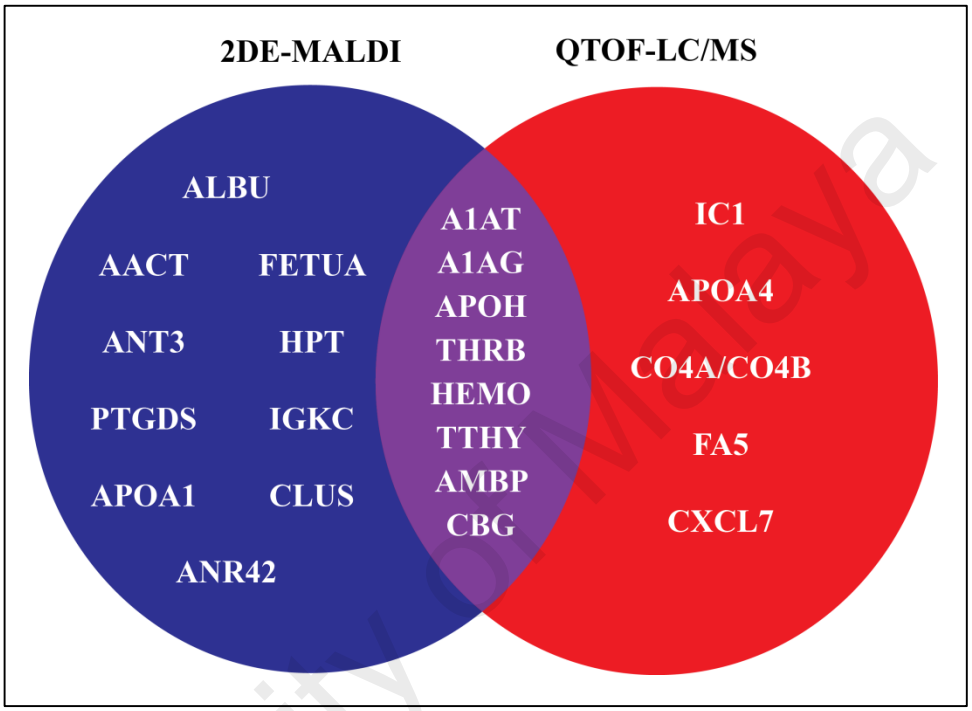


Figure 4.15: Summary of proteins detected in serum perchloric acid isolates using 2-DE and QTOF-LC/MS.

Proteins that were identified using 2-DE-MALDI are summarised in the blue circle while the red circle represents proteins that were identified using QTOF-LC/MS. The overlapped area of the two circles, coloured in purple, represents proteins identified by both platforms. Names of abbreviated proteins are listed in Table 3.3 and Table 3.4.

4.2.5 CGB lectin blotting analysis

Aside from determining the presence of *O*-glycosylated protein in serum perchloric acid isolates based on previous studies, CGB lectin blotting analysis was also carried out. 2-DE was performed on pooled serum perchloric acid isolates, followed by protein transfer and lectin blotting as described in section 3.6.5. The profile of *O*-glycosylated proteins in the serum perchloric acid isolate is shown in Figure 4.16. There were about 10 noticeable spots detected visually. Since attempts made to identify the proteins which were on membranes using mass spectrometry were not successful, effort was made by matching the spots with the earlier generated 2-DE gel of serum perchloric acid isolate (Figure 4.14a). For better prediction accuracy, both the 2-DE gel and CGB lectin blot image were aligned using Progenesis software. Five proteins that may contain *O*-glycans were identified in this manner. They were hemopexin, beta-2-glycoprotein 1, corticosteroid-binding globulin, alpha-HS-glycoprotien and AMBP. Evidence of *O*-glycosylation of these proteins had been previously documented except for corticosteroid-binding globulin, which was reported to have *N*-glycans (Table 4.3).

4.2.6 Determination of total *O*-glycosylated proteins in serum perchloric acid isolates

To analyse the capability of perchloric acid in enriching *O*-glycosylated proteins from human serum, levels of the *O*-glycosylated proteins were determined using a sandwich ELLA. The lectin assay was performed on ten random normal human serum samples that were both treated and null treated with perchloric acid, and the levels of the *O*-glycosylated proteins were expressed in ng asialoMucin/µg protein (Figure 4.17). The data suggests that perchloric acid was capable of enriching approximately 14.6-fold amounts of *O*-glycosylated proteins from the serum samples.

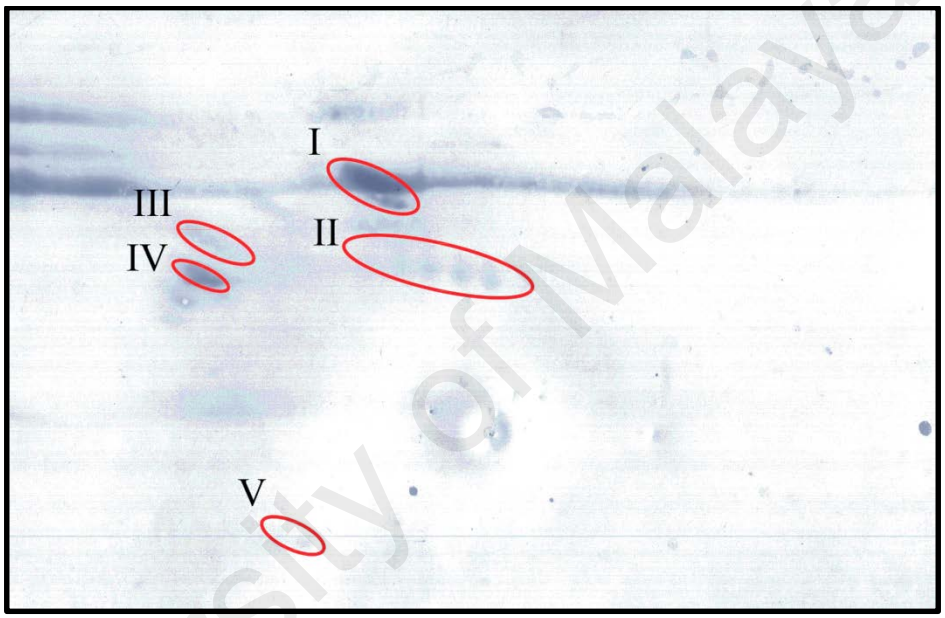


Figure 4.16: CGB lectin blot of pooled serum perchloric acid isolates.

Proteins from 2-DE gel were transferred onto a nitrocellulose membrane and blotted with CGB lectin. Protein spots were compared with Figure 3.13a for prediction of their identities. I; Hemopexin, II; Beta-2-glycoprotein 1, III; Corticosteroid-binding globulin, IV; Alpha-2-HS glycoprotein, V; Protein AMBP.

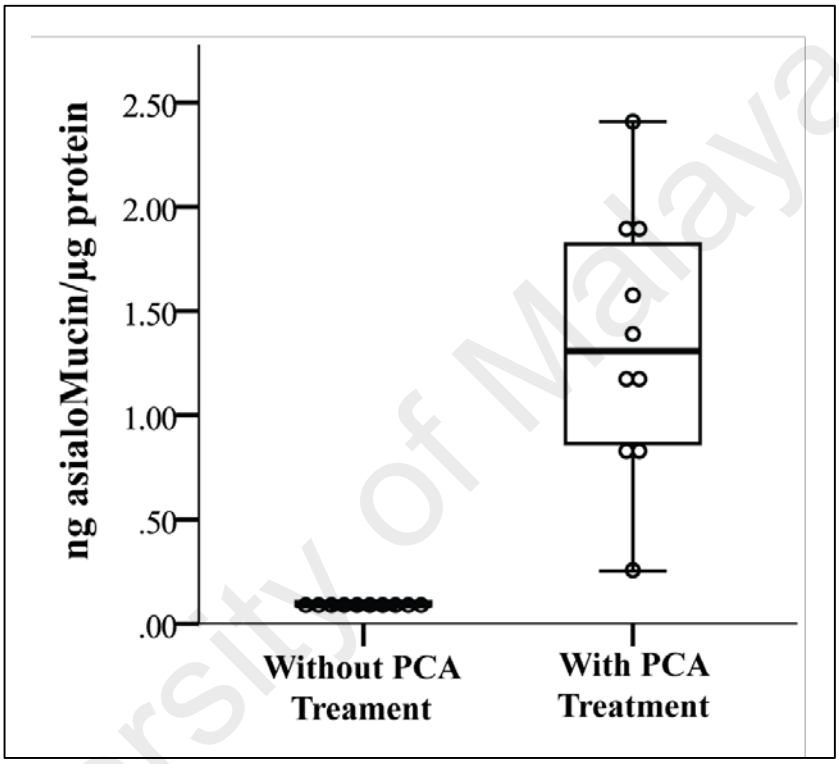


Figure 4.17: Levels of *O*-glycosylated proteins in serum perchloric acid isolates.

Serum samples were prepared with or without perchloric acid (PCA) treatment. The levels of *O*-glycosylated proteins were estimated using sandwich ELLA with asialoBSM as the standard *O*-glycosylated protein.

4.2.7 Recovery of *O*-glycosylated proteins from serum after perchloric acid treatment analysis

In another related experiment, whether perchloric acid would have any effect on the recovery of *O*-glycosylated- and biotinylated-asialoBSM that was spiked into serum were assessed. In this experiment, 5 µg of biotinylated-asialoBSM was spiked into 50 µl of serum samples before and after perchloric acid treatment. To detect the spiked biotinylated-asialo-BSM, a technique similar to the earlier described sandwich ELLA in which serum samples were first allowed to be captured by CGB lectin that was precoated onto plates was employed. This was followed by binding of streptavidin-conjugated alkaline phosphatase. Due to presence of other *O*-glycosylated proteins in the serum perchloric acid isolates, not all asialoBSM was expected to bind to the CGB lectin. Figure 4.18 shows the amounts of spiked biotinylated-asialoBSM recovered from serum prior to and after treatment with perchloric acid, which were not significantly different. Results generated also indicated that about 92.1% of biotinylated-asialoBSM was successfully retained. This suggests that perchloric acid was highly effective in enriching the heavily glycosylated proteins from serum.

4.3 Application of developed assay on early stages of breast cancer

In previous parts of this study, the development of a sandwich ELLA to detect total *O*-glycosylated proteins and the proteomic profiles of serum perchloric acid isolates were described. Here, sandwich ELLA and perchloric acid treatment were applied on serum samples of early breast cancer patients to determine the levels of *O*-glycosylated protein. In addition, the combination technique of perchloric acid treatment and CGB lectin isolation was developed and applied on the serum samples in order to investigate the specific glycoproteins that contribute to the altered levels of *O*-

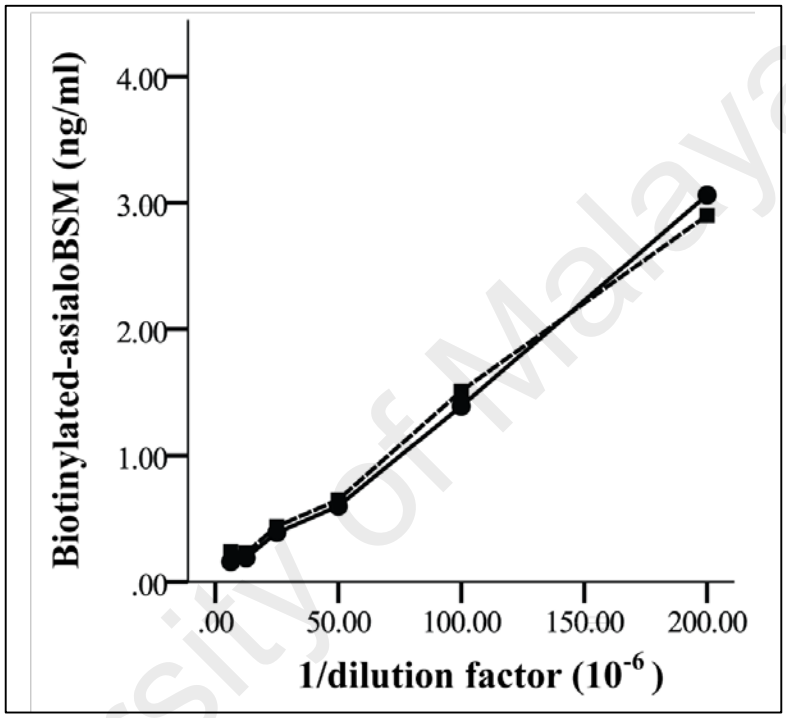


Figure 4.18: Recovery analysis of biotinylated-asialoBSM after perchloric acid treatment on serum.

O-glycosylated proteins recovery assay was performed by spiking biotinylated-asialoBSM into serum samples and followed by perchloric acid treatment. (●) Biotinylated-asialoBSM spiked into serum samples followed by perchloric acid treatment; (■) biotinylated-asialoBSM spiked on serum perchloric acid isolates.

glycosylated proteins. Thus, such glycoproteins may become the potential candidate of early breast cancer detection.

4.3.1 Analysis of *O*-glycosylated proteins in sera of patients with stage 0 and stage I breast cancer

Using the earlier described sandwich ELLA, the levels of *O*-glycosylated proteins in serum samples of patients with stage 0 (n = 31) and stage I (n = 48) breast cancer as well as those from the controls (n = 105) were evaluated. There was no statistical difference in the distribution of age and ethnic groups in the three groups. Serum samples of breast cancer patients of both stages showed significantly higher levels of *O*-glycosylated proteins compared to those of normal subjects (Figure 4.19a). When the experiment was repeated on the same serum samples but were subjected to perchloric acid treatment, relatively higher levels of *O*-glycosylated proteins were detected in all subjects. Similar differences in the levels of *O*-glycosylated proteins between controls and patients with stage 0 and stage I breast cancer were also observed (Figure 4.19b).

4.3.2 Effect of perchloric acid treatment on performance of sandwich ELLA

A receiver operative characteristic (ROC) analysis was subsequently carried out to evaluate performance of the assay in detecting the *O*-glycosylated proteins in serum samples of the breast cancer patients. Assay performed on non-perchloric acid treated stage 0 serum samples had an overall accuracy of 0.760 area under the curve (AUC) whilst increased accuracy of detection of up to 0.869 AUC was observed in assay performed using the same serum samples that were treated with perchloric acid (Figure

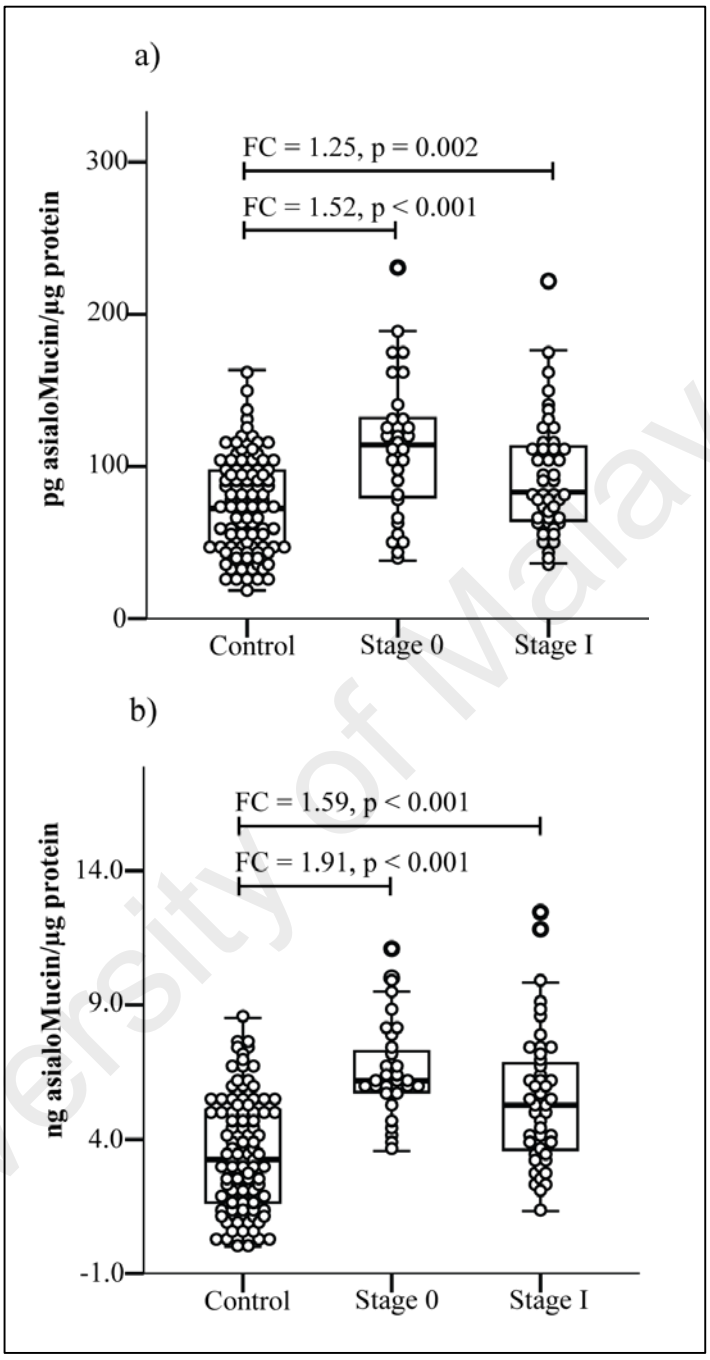


Figure 4.19: Estimation of total *O*-glycosylated protein in serum samples of cancer negative women (control) and patients with early stages of breast cancer.

Serum samples were analysed by sandwich ELLA without (panel a) and after perchloric acid treatment (panel b).

4.20a). Similarly, perchloric acid treatment of serum improved the assay detection of stage I breast cancer from 0.638 AUC to 0.730 AUC (Figure 4.20b). In case of stage 0, treatment with sera with perchloric acid improved the sensitivity and specificity of detection from 68% to 81% whilst for stage I breast cancer, the sensitivity and specificity of the assay rose from 56% to 65%.

4.3.3 CGB lectin precipitation of serum perchloric acid isolates

As attempts made to identify proteins on blot using mass spectrometry was not successful, a different approach to determine *O*-glycosylated proteins that were captured by the CGB lectin was developed. This method uses CGB lectin that was conjugated to Sepharose beads. The CGB-Sepharose beads were initially allowed to interact with serum perchloric acid isolates and the beads were then washed extensively with washing buffer to remove unbound proteins. Bound *O*-glycosylated proteins were subsequently released using Laemmli buffer and they were resolved by electrophoresis in 8% SDS-polyacrylamide gel. The optimal amount of the CGB lectin beads and serum perchloric acid isolate used were determined. About 5 µl, 10 µl and 20 µl of CGB lectin bead volumes were tested along with 30 µg of serum perchloric acid protein isolate. The results show that higher intensity of protein bands appeared to be generated with less volumes of CGB lectin beads used (Figure 4.21a). In contrast, gradual increase of intensity was observed when higher amounts of serum perchloric acid protein isolate were used (Figure 4.21b). Therefore, five µl of CGB lectin bead volume and 30 µg of serum perchloric acid protein isolate were used for the subsequent analysis.

The glycoproteins released from the CGB lectin conjugated beads with or without heat treatment were also compared in the present study. The results apparently showed that peptides with molecular weights greater than 70 kDa were only released

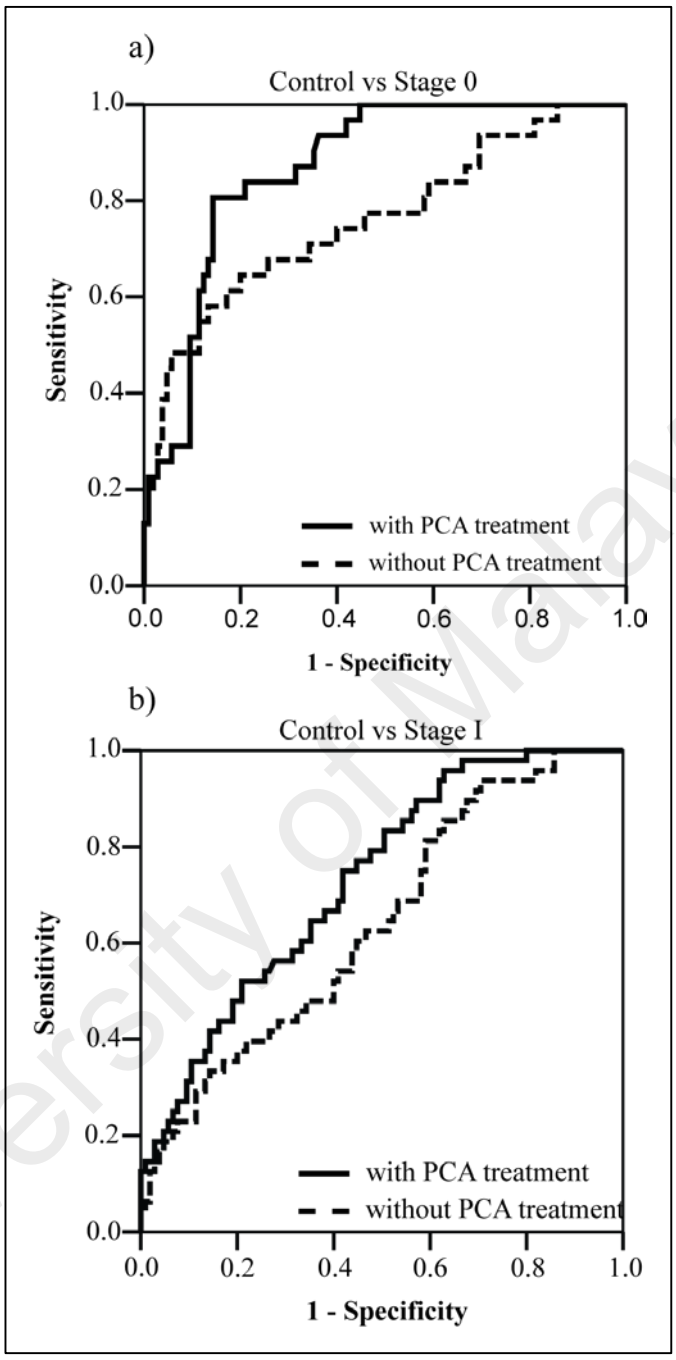


Figure 4.20: Sensitivity and specificity of sandwich ELLA to detect early breast cancer.

The assay performance was evaluated by performing Receiver Operative Characteristic analysis on the levels of *O*-glycosylated proteins detected before (----) and after serum treated with perchloric acid (—). Panels a and b refer to analysis of stage 0 and stage I serum samples, respectively.

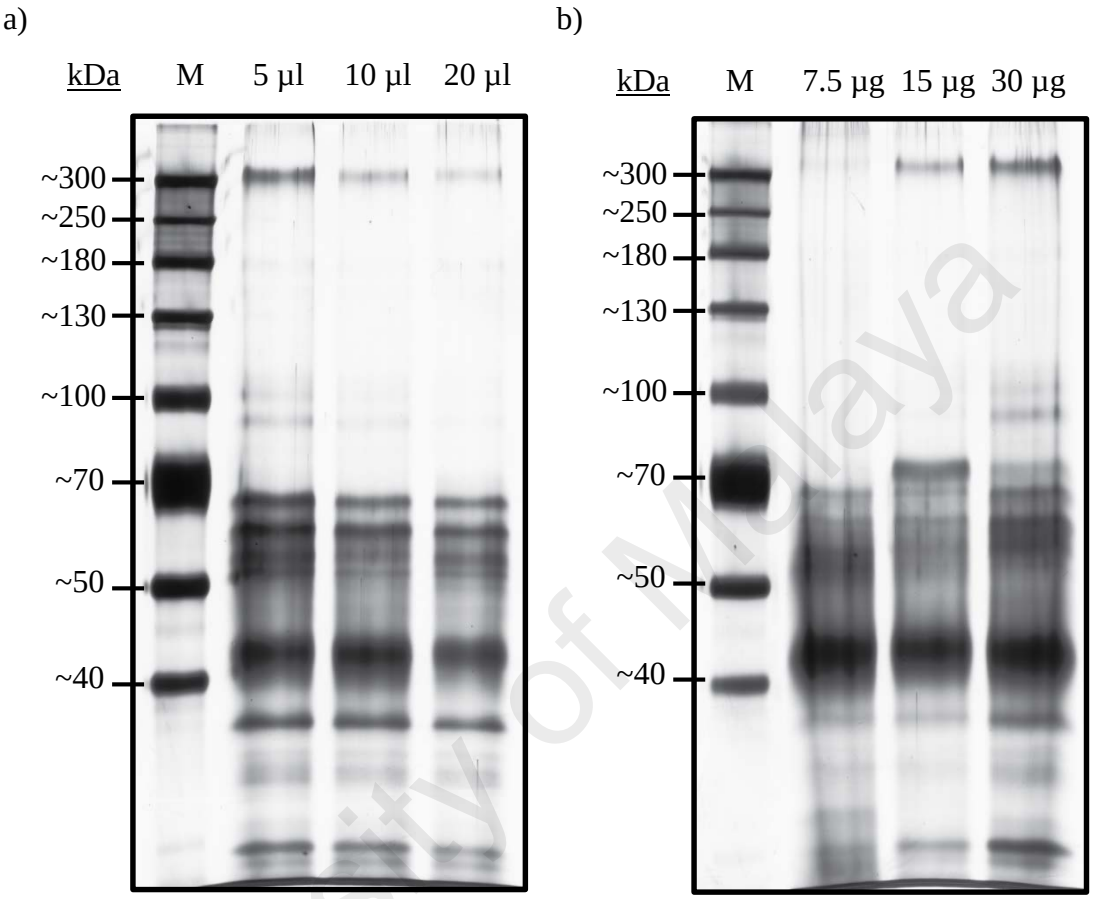


Figure 4.21: Optimization of CGB lectin precipitation on serum perchloric acid isolates.

Different volumes of beads ranging from 5 µl to 20 µl were tested on 30 µg protein of serum perchloric acid isolates sample (panel a). A similar experiment was performed on 5 µl beads with varies amount of serum perchloric acid isolates proteins (7.5, 15 and 30 µg protein) (panel b). Samples were run in 8% SDS-PAGE gel.

from the CGB lectin beads when heat was applied (Figure 4.22). However, the lower molecular weight proteins appeared to be released from the beads regardless of heat treatment. Since high molecular weight proteins were the proteins of interest in the present study, heat treatment was therefore mandatory so that all glycoproteins were released from the beads during the lectin affinity chromatography experiments.

To investigate the possibility of unspecific direct binding of proteins with the Sepharose beads, a separate experiment was repeated as previously described but using unconjugated Sepharose beads as well as CGB lectin-Sepharose without addition of any sample. The results show that there appear to be minor non-specific binding of proteins from the serum perchloric acid isolates with Sepharose beads was clearly noticeable at the range of 50 to 70 kDa molecular weight region (Figure 4.23). On the other hand, CGB lectin-Sepharose without addition of any protein sample shows no apparent peptide band.

4.3.4 Interaction of breast cancer serum perchloric acid isolates with CGB lectin

When the CGB lectin precipitation was performed on pooled serum perchloric acid isolates of patients with stage 0 (n = 31) and stage I (n = 48) breast cancer as well as those of controls (n = 105), different protein profiles were generated (Figure 4.24). Some of the protein bands appeared to demonstrate differences in intensities between pooled sera of both stage 0 and I breast cancer patients and those that were generated from controls. When the SDS-PAGE serum protein profiles of controls were subjected to densitometry and compared with those of patients with both stage 0 and stage I breast cancer, only protein bands of b1, b3, b4 and b14 appeared to demonstrate more than 1.5-fold difference of intensities (Table 4.5).

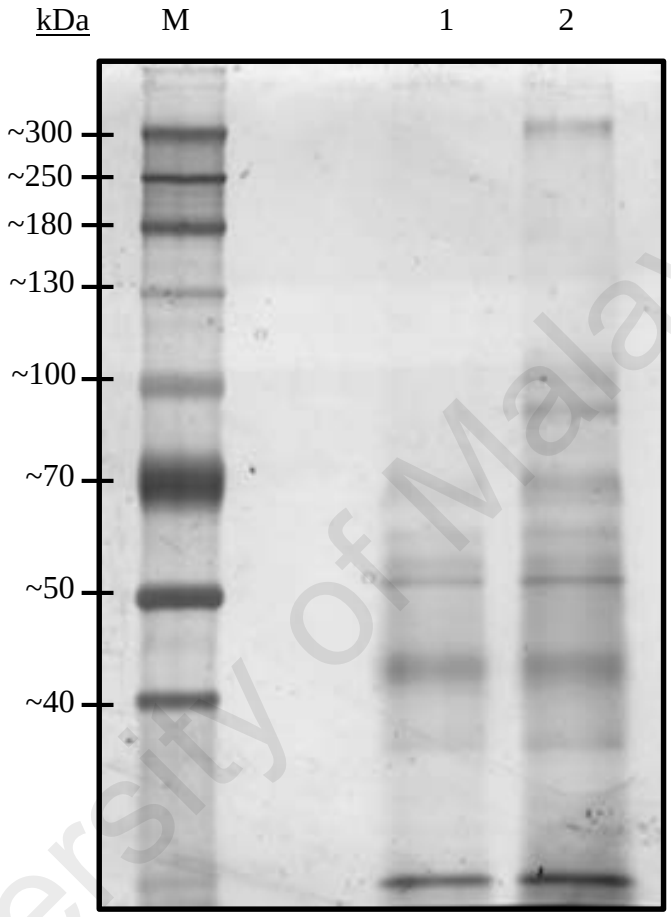


Figure 4.22: Release of *O*-glycosylated proteins from CGB lectin beads with and without heat treatment.

CGB lectin precipitation was performed on normal serum perchloric acid isolates. The bound glycoproteins were released by incubating the beads in Laemmlli buffer at room temperature (Lane 2) or at 95°C (Lane 1) for 5 minutes. The released samples were resolved in 8% SDS-polyacrylamide gel.

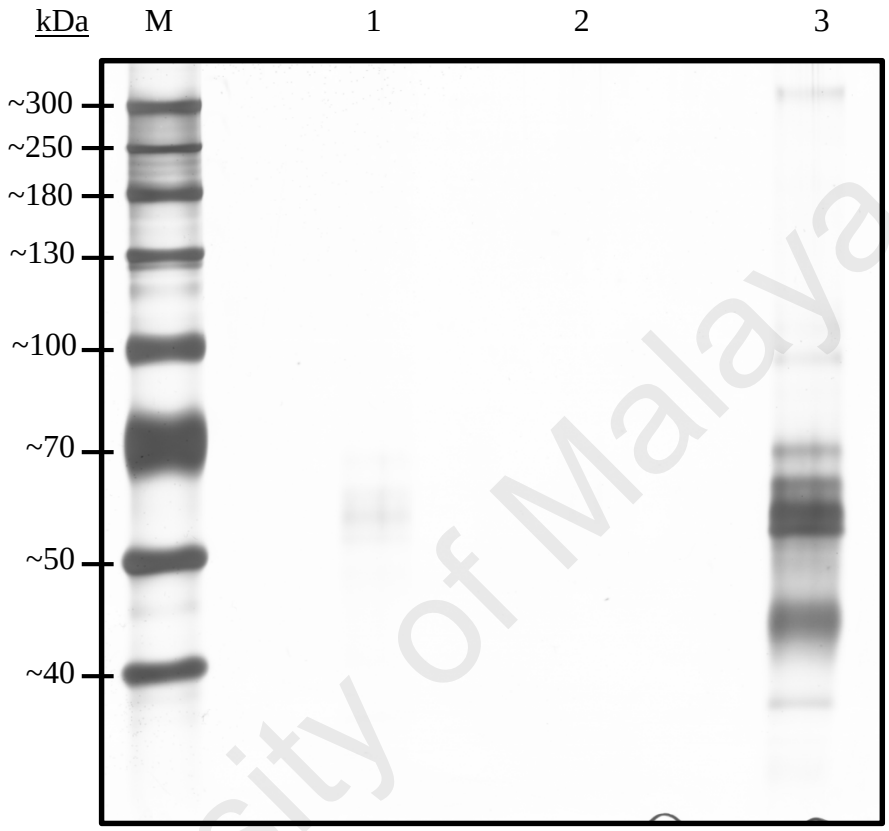


Figure 4.23: Unspecific binding of serum perchloric acid isolates with CGB-Sepharose beads.

Serum perchloric acid isolates was allowed to interact with unconjugated Sepharose beads (Lane 1). Lane 2 shows CGB-Sepharose beads without incubation with sample. The experiment was performed side by side with control serum perchloric acid isolates sample (Lane 3). Samples were run in 8% polyacrylamide gel.

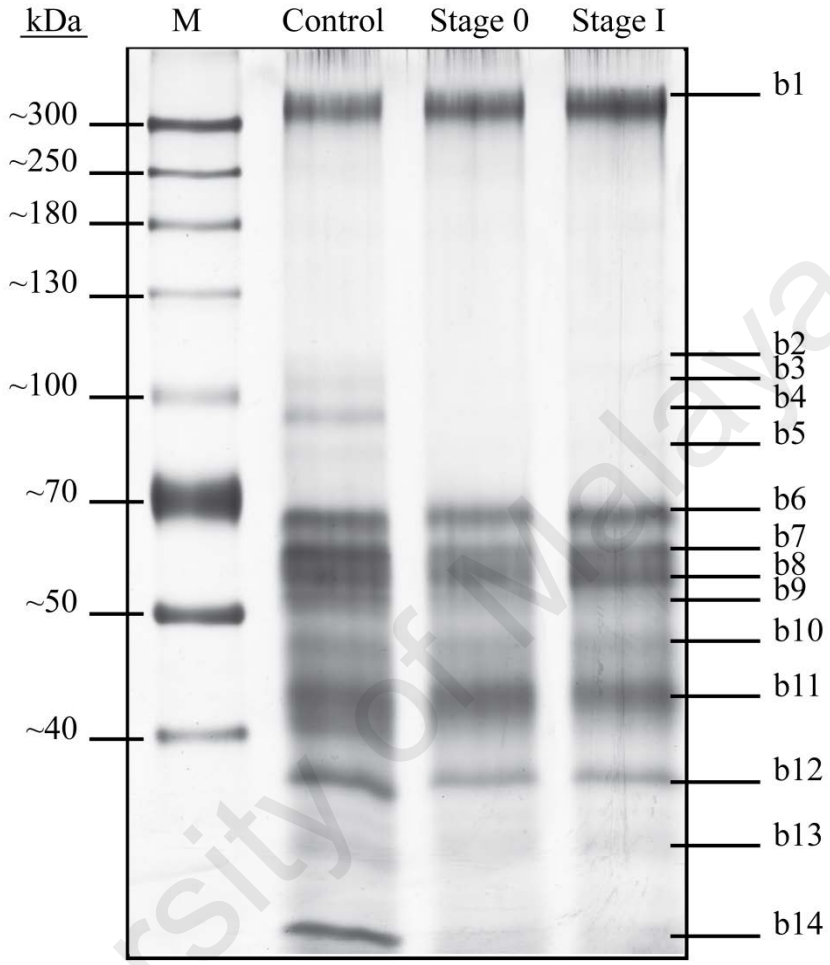


Figure 4.24: Profiles of CGB lectin bound proteins of cancer negative women and breast cancer patients.

CGB lectin conjugated to Sepharose was allowed to interact with serum perchloric acid isolates of cancer negative women (control) and patients with stage 0 and stage I breast cancer. Bound proteins were released with Laemmli buffer and resolved in 8% SDS-polyacrylamide gel, which was stained with silver. The gel image was analysed using GelAnalyzer software.

Table 4.5: Intensity of peptide bands from CGB lectin precipitation of serum perchloric acid isolates.

Band	Control (n = 4)	Stage 0 (n = 4)			Stage I (n = 4)		
	% Volume	% Volume	FC[b]	*p* value[c]	% Volume	FC[b]	*p* value[c]
b1	7.10±0.33[a]	14.57±2.10	+2.05	0.013	13.27±1.26	+1.87	0.013
b2	0.81±0.11	0.71±0.08	-1.14	n.s	0.79±0.04	-1.03	n.s
b3	1.44±0.15	0.83±0.06	-1.73	0.009	0.95±0.06	-1.52	0.022
b4	2.86±0.49	0.85±0.06	-3.36	0.007	0.95±0.09	-3.01	0.009
b5	1.36±0.10	0.97±0.06	-1.40	0.018	1.07±0.15	-1.27	n.s
b6	7.45±0.34	6.65±0.68	-1.12	n.s	6.62±0.72	-1.13	n.s
b7	7.74±0.34	5.55±1.08	-1.39	n.s	6.03±0.61	-1.39	n.s
b8	6.82±0.23	8.04±0.36	+1.18	0.030	8.52±0.13	+1.25	0.001
b9	5.45±0.69	4.65±0.51	-1.17	n.s	3.91±0.42	-1.39	n.s
b10	6.68±0.41	7.03±0.66	+1.05	n.s	6.54±0.64	-1.02	n.s
b11	19.28±0.61	19.14±0.79	-1.01	n.s	19.16±1.04	-1.01	n.s
b12	6.60±0.19	4.54±0.34	-1.45	0.002	4.44±0.52	-1.49	0.008
b13	3.41±0.16	2.75±0.22	-1.24	n.s	2.80±0.16	-1.22	0.037
b14	5.67±0.15	2.90±0.22	-1.96	0.001	3.18±0.48	-1.78	0.003

[a] Gel band intensity expressed in % Volume. Value in Mean±SEM from four technical replicates.

[b] FC; Fold Change of % volume value in relative with the control.

[c] *p* value obtained from student t-test by comparing the % volume value of stage 0 or stage 1 breast cancer subjects with control subjects.

4.3.5 Identification of perchloric acid isolated proteins of altered abundance

To identify these proteins of interest, the highly resolved bands (b1, b3, b4 and b14) were excised in accordance to those marked in Figure 4.24 and subjected to Q-TOF mass spectrometry analysis. Our database search identified the largest protein (b1) resolved by the SDS-PAGE as that of proteoglycan 4 (also known as lubricin), and two other protein bands that showed differences of abundance were those of plasma protease C1 inhibitor (Table 4.6). However, the fourth protein band of interest (b14) was not affirmatively identified probably due to presence of multiple proteins. When densitometry was reanalysed based on identities of the proteins, the total levels of plasma protease C1 inhibitor (b3 and b4) of stage 0 and stage I breast cancer patients were both significantly reduced compared to the controls (Figure 4.25b). On the contrary, the abundance of proteoglycan 4 (b1) was significantly higher in both stage 0 and stage I breast cancer patients (Figure 4.25a).

4.3.6 Ratio of proteoglycan 4 to plasma protease C1 inhibitor

Because of the reciprocal difference in levels of proteoglycan 4 and plasma protease C1 inhibitor between both groups of breast cancer patients compared to the controls, calculated ratios of abundances of the two serum proteins further amplified their differences, with more than 5-fold difference for stage 0 and 4.17-fold difference in case of stage I (Figure 4.26). This is suggestive of the strong potential application of the ratio of the two serum *O*-glycosylated proteins as biomarkers for detection of early breast cancer. However, the difference in fold changes between stage 0 and stage I (compared to the controls) was not statistically significant.

Table 4.6: Protein identity of CGB lectin isolated proteins of altered abundance.

Band‡	Protein Name	MS/MS Search score	Sequence Coverage (%)	Theoretical Mass (Da)	Observed Mass (Da)
b1	Proteoglycan 4	67	8	55460	317592
b3	Plasma Protease C1 Inhibitor	444	31	55734	95156
b4	Plasma Protease C1 Inhibitor	708	34	55734	85124
b14	Not identified	-	-	-	37715

‡Refers to excised electrophoretic bands as depicted in Figure 4.24

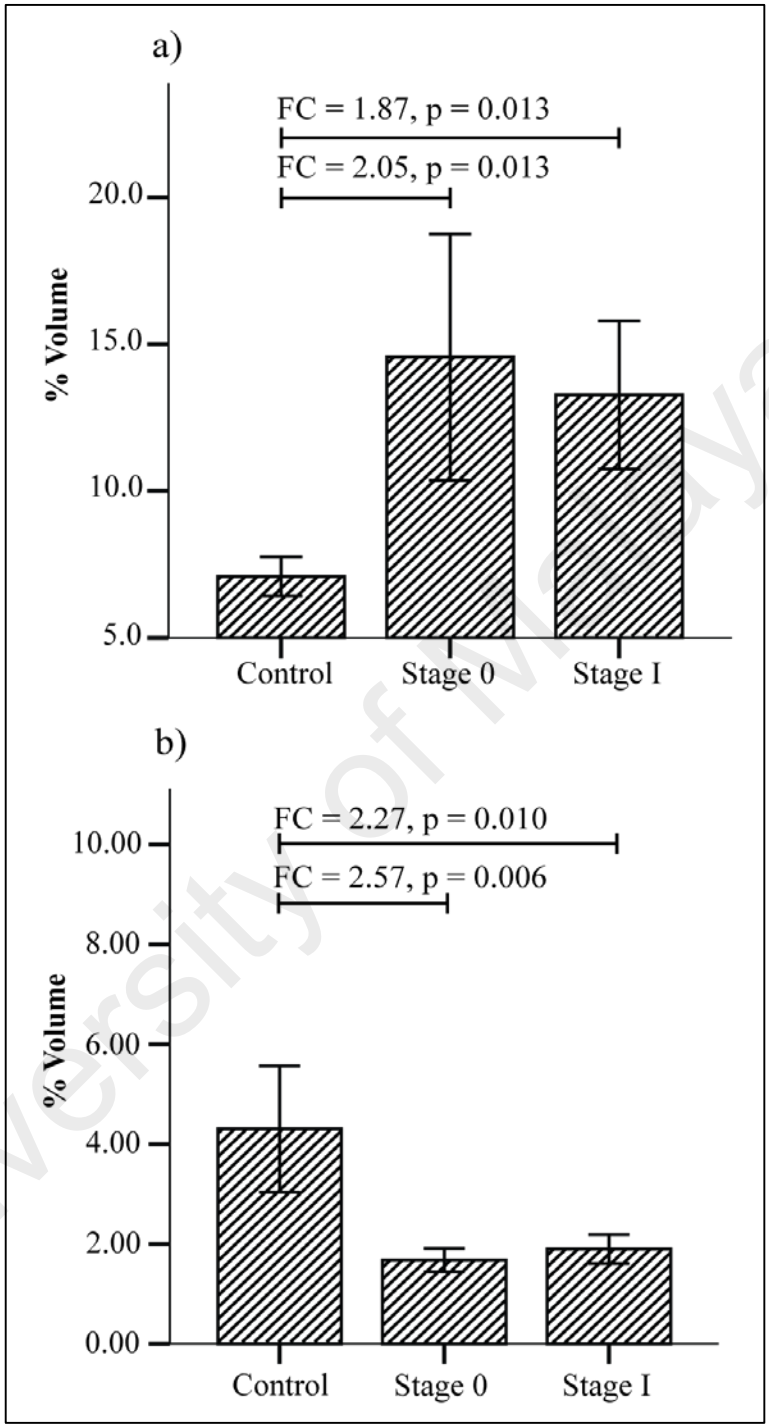


Figure 4.25: Levels and ratio of CGB lectin bound proteins in cancer negative women and breast cancer patients.

Percentage of volume was obtained using GelAnalyzer software. The percentage volume of each peptide band was compared using SPSS. Panels a and b refer to mean % volumes of lubricin and plasma protease C1 inhibitor, respectively.

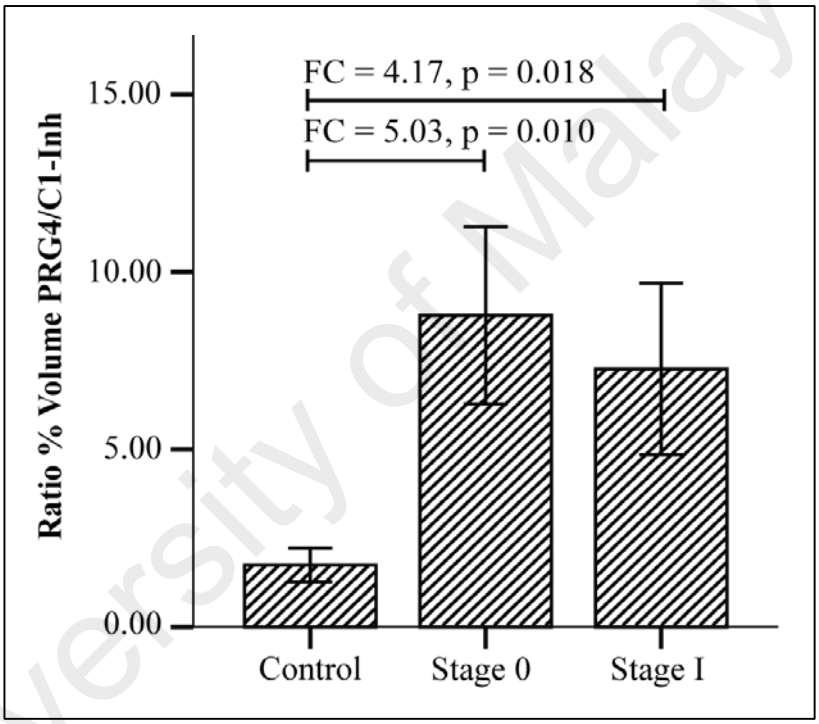


Figure 4.26: Ratio of percentage volume of proteoglycan 4 to plasma protease C1 inhibitor.

The ratio of proteoglycan 4 (lubricin)/plasma protease C1 inhibitor demonstrates higher fold-change difference between both the early stages of breast cancer compared to controls.

4.3.7 Western blot analysis

To validate the levels of protease C1 inhibitor in serum perchloric acid isolates collected after CGB lectin precipitation, Western blot analysis was performed using rabbit anti-plasma protease C1 inhibitor (Figure 4.27a). Similar differential intensities of protease C1 inhibitor bands, which correspond to the intensities of the peptide bands that were previously detected in SDS-PAGE (Figure 4.24) were observed in the blot. The levels of plasma protease C1 inhibitor from Western blot analysis were significantly lowered in stage 0 and stage I breast cancer patients compared to those of the controls (Figure 4.27b). When Western blot was repeated using neat serum and mild acid hydrolysed serum sample, levels of protease C1 inhibitor was not significantly different. Immuno detection using anti-sera against proteoglycan 4 was unfortunately not successful.

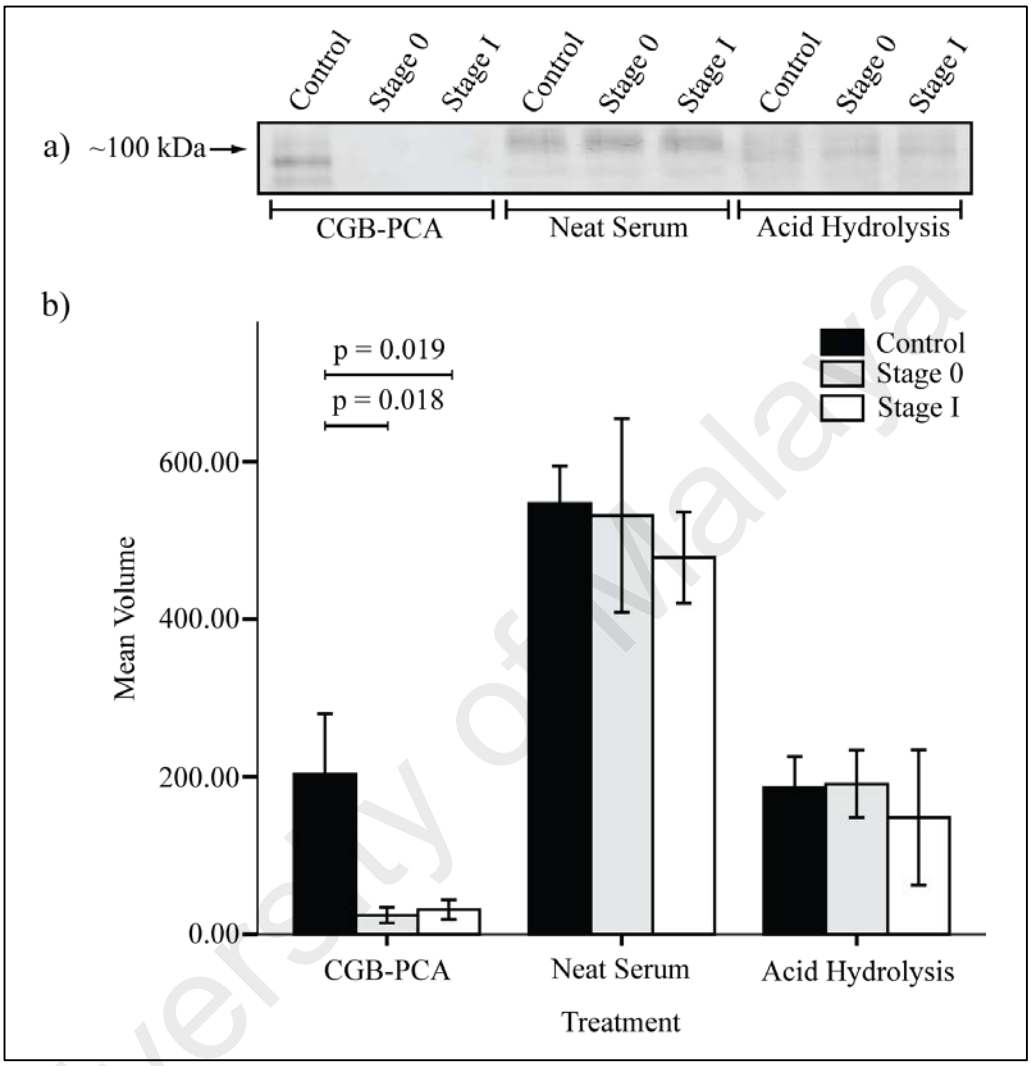


Figure 4.27: Western blot analysis of protease C1 inhibitor on samples of early breast cancer.

CGB lectin bound proteins of serum perchloric acid isolates, neat serum and acid hydrolysed serum samples were transferred onto nitrocellulose membrane and blotted using rabbit anti-plasma protease C1 inhibitor (panel a). Intensities of the bands were analysed using the GelAnalyzer software and volume intensity was compared across the three groups of samples analysed (panel b).

CHAPTER 5: DISCUSSION

5.1 Isolation of CGB lectin

Research performed more than two decades ago has shown that CGB lectins extracted from six different clones of champedak displayed uniform structural and binding properties (Hashim *et al.*, 1993). This is unlike the jackfruit lectin, jacalin, which has been shown to have a broader specificity and very much dependent on the source of the fruit. Whilst jacalins isolated from the seeds obtained from Brazil (Roque-Barreira *et al.*, 1986) and Philippine (Kondoh *et al.*, 1987) were shown to have specific interactions with *O*-glycosylated proteins such as IgA1, those that were isolated from seeds obtained from Okinawa prefecture in Japan showed additional non-specific interactions with *N*-glycosylated proteins such as IgA2, IgM, IgD and IgE (Kondoh *et al.*, 1987). The disparity in binding was observed again from recent studies using crystallized jacalin and CGB lectin (Gabrielsen *et al.*, 2014). Attempts to introduce mannose into CGB lectin crystals that was soaked with galactose or galactose-β1-3-*N*-acetyl-galactose-NAc were apparently unsuccessful. This was in contrast to jacalin crystals, where weak binding of mannose was seen; with the crystal structure of jacalin with bound methylated mannose in the sugar-binding site solved (Jeyaprakash *et al.*, 2004).

The CGB lectin that was used in the present study was freshly isolated and purified in the laboratory. Despite earlier reports on the structural and functional homogeneity of CGB lectins isolated from different clones of champedak, efforts were taken to ascertain that the lectin that was purified conforms to the standard structural and functional features of the CGB lectin, *i.e.*, specific interaction to *O*-glycans. In the present study, the procedure for isolation of CGB lectin was slightly modified for better efficiency. In this method, fractions were not collected; instead wash buffer was added

at fixed volumes based on the size of the column used. As such, the period for isolation of the lectin was cut short as it was no longer necessary to collect and monitor elution fractions.

The purity of the isolated CGB lectin was assessed using SDS-PAGE and visualised using silver staining. Silver stain has a sensitivity limit of 10-2 pmol/mm^2 (Switzer *et al.*, 1979) and therefore sensitive enough to show any possible contamination. In addition, the purity of the isolated CGB lectin was also analysed using the MALDI-TOF, in linear mode. In this experiment, aside from the peaks of the CGB lectin that appeared in the result spectrum, there was no other peak detected. This showed that the CGB lectin that was isolated in the present study was of the highest purity. The MALDI-TOF analysis is also crucial because the champedak seeds are known to also contain a mannose binding lectin (Seriramalu *et al.*, 2010) in addition to the CGB lectin. This lectin interacts with mannose residues, which make up the core structure of *N*-glycans (Stanley *et al.*, 2009). During the lectin isolation procedure, the mannose binding lectin may have been extracted as well. The MALDI-TOF analysis was performed to ensure that no mannose binding lectin was present in the isolated CGB lectin fraction. The mannose binding lectin, if present, will interact with *N*-glycosylated proteins and therefore leads to false positive results when the isolated CGB lectin solution is used. In the present study, the mannose binding lectin peak (16.4 kDa) was not seen in the analysis of the CGB lectin extract by MALDI-TOF.

CGB lectin is a homotetrameric protein (Gabrielsen *et al.*, 2014). Peptide bands of CGB lectin displayed on gel have molecular weights of approximately 14 kDa and 16 kDa. However, crystallography data shows that CGB lectin monomer comprises two chains; a heavy α-chain consisting of 133 amino acids and a light β-chain consisting of 21 amino acids (Gabrielsen *et al.*, 2014). The estimated molecular weights of both the heavy chain and light chain were 13,000 Da and 2,100 Da, respectively. Thus, the

heaviest peptide band appears on the gel may represent a monomer of the CGB lectin, whilst the second peptide band may represent the heavy α-chain.

In the present study, the binding properties of the purified CGB lectin were also assessed. This was to ensure that it was not lost and that it was also specific to *O*-glycans. Radial immunodiffusion, which is simple and easy to perform, is a traditionally used method to evaluate binding of specific antisera to antigens (Harrell & George, 1972). Similarly, the method may also be used to evaluate interactions of lectins. Aside from radial immunodiffusion, lectin binding may also be evaluated using blood agglutination assay (Rudiger & Gabius, 2001); however, it was not a preferred method as it requires fresh blood samples and unable to determine the binding specificity of the lectin. In the present study, radial immunodiffusion method was therefore employed and the CGB lectin was used instead of antisera. Radial immunodiffusion allows multiple antigens, in this case test glycoproteins, to be tested in a single run. The analysis of the purified CGB lectin was performed on two *O*-glycosylated proteins and two *N*-glycosylated proteins. The result revealed that the purified lectin had interacted with both the *O*-glycosylated proteins but not the *N*-glycosylated proteins. Hence, the data demonstrated that the binding activity of CGB lectin was maintained despite being subjected to multiple procedures in its purification, and that the lectin did not interact with *N*-glycosylated proteins. In any case, the CGB lectin is unlike those from *Nymphaea nouchali* tuber and jade beans which are calcium ion dependent. The binding activities of these lectins are easily lost in the absence of calcium ions and extra care is usually needed in order to extract such lectins (Cheung *et al.*, 2013; Kabir *et al.*, 2011).

5.2 Development of *O*-glycosylated protein assay

Currently, there is no simple and reliable assay to estimate the total content of *O*-glycosylated proteins in complex biological samples. This is important as far as cancer is concerned as the disease is known to be strongly associated with altered expression of mucin type *O*-glycosylated proteins in serum of cancer patient (Bhavanandan, 1991; Brockhausen, 1999). Changes of the glycan structures are commonly observed in tumour cells (Pinho & Reis, 2015). Overexpression of truncated GalNAc-type *O*-glycans is a common feature of many tumours. These structures, which are also called mucin type *O*-glycans, are frequently found in most transmembrane and secreted glycoproteins. Aberrant glycosylation may also occur in glycoproteins that display the truncated *O*-glycans. Common examples include the Thomsen-Friedenreich antigen (also known as core 1 or T antigen) and the Tn antigen (GalNAc) and their sialylated forms, ST and STn, which result from the incomplete synthesis of *O*-glycans (Kudelka *et al.*, 2015). This alteration usually involves early termination of sialic acid, which leads to the truncated forms of *O*-glycans (Julien *et al.*, 2012). Therefore, an assay to detect the overexpression of such *O*-glycans structures may be useful for cancer detection.

There have been considerable reports on attempts to target the carbohydrate structures of *O*-glycosylated proteins and quantitate their amounts using chemical methods. Bhavanandan *et al.* (1990) developed a colorimetric assay combined with the technique of β-elimination with Morgan-Elson reagent that reacts on *N*-acetylhexosamines. Kilcoyne *et al.* (2011) reported a similar approach but with the use of periodic acid-Schiff's reagent to optimise detection of carbohydrates on proteins that were applied into microtiter plates. These methods were shown to be highly effective on pure glycoprotein samples. However, the applicability of the assays on complex biological samples was not demonstrated. Complex biological sample such as serum

contains numerous types of molecules that may interact with the detection reagents and this may cause undesirable results. Moreover, colorimetric assay typically has low detection sensitivity.

In the present study, binding of lectins to the unique GalNAc-Ser/Thr linkage feature was targeted for the development of an assay to quantify *O*-glycosylated proteins in crude biological samples. Lectin has the capability to interact specifically with carbohydrate residues of an intact glycoprotein. This allows high detection sensitivity without the use of chemical reagents and purified glycoprotein. Since the GalNAc-Ser/Thr linkages are invariably cryptic in *O*-glycosylated proteins, it was necessary to first remove at least some of the peripheral saccharides. One approach to do this is by treatment with enzymes. However, this will require a supply of several highly purified glycosidases, which are very costly. Another approach is by using chemical treatment such as mild acid hydrolysis and Smith degradation, which involves mild periodate oxidation (Andrews *et al.*, 1969; Perlin, 2006). Figure 5.1 illustrates the concept of chemical treatment of proteins to generate uniform *O*-glycan structures bearing the unique GalNAc-Ser/Thr linkage feature that are recognised by the three lectins used in the present study.

Mild acid hydrolysis is a commonly used method for chemical desialylation, which involves sample incubation under low acidic condition at 80°C for 1 hour (Morelle *et al.*, 2006; Rohrer & Townsend, 1995). The method efficiently removes mostly sialic acid from glycans whilst other sugars remain intact. This was seen from previous findings (Morelle & Michalski, 2005) as well as in the present study (Section 4.1.4). The rationale for periodate oxidation was that it would cleave peripheral saccharides with exception of the GalNAc residue that is attached to Ser or Thr. The GalNAc-Ser/Thr will be preserved in the majority of the *O*-glycosylated proteins

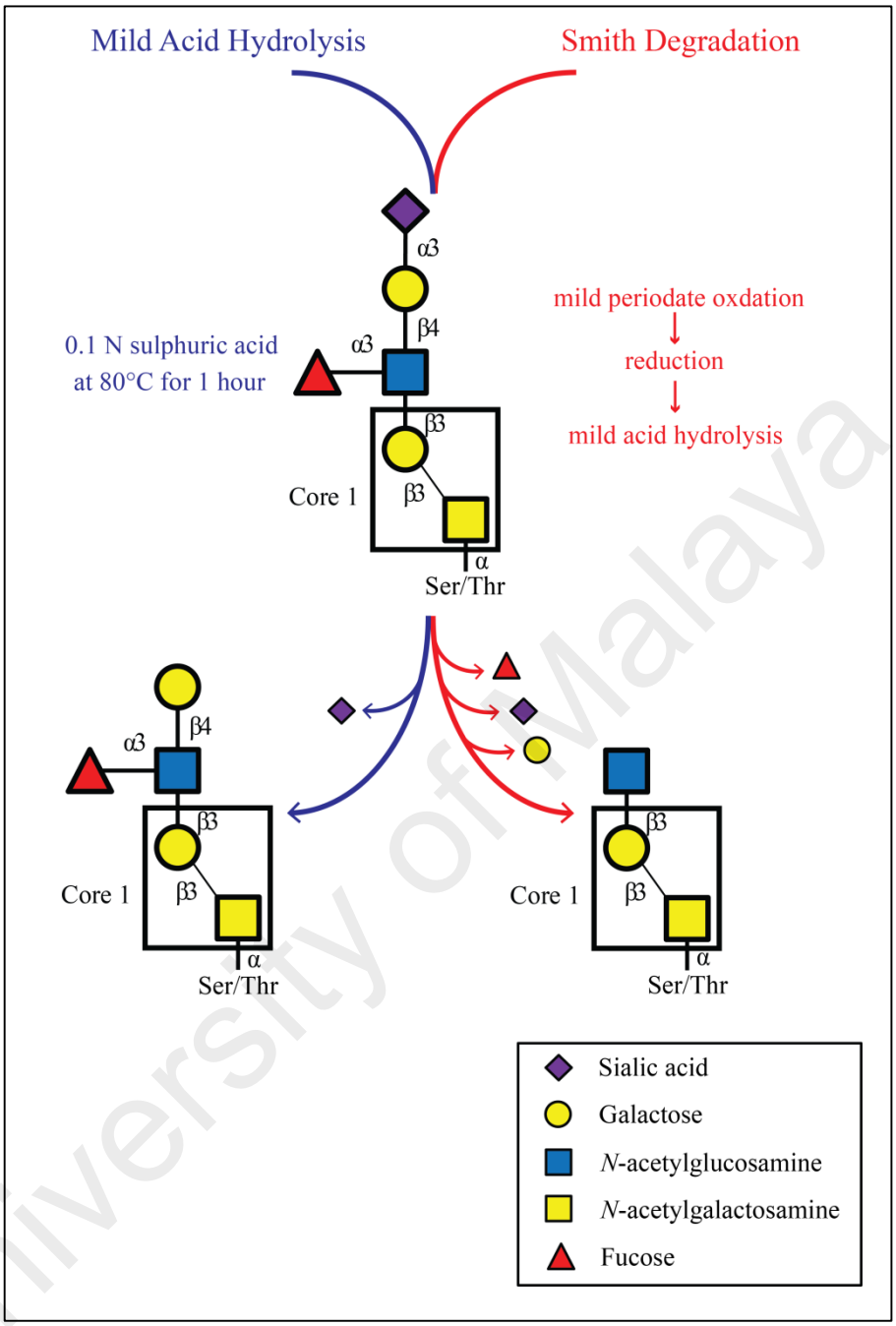


Figure 5.1: Concept of *O*-glycan modification using chemical methods.

The *O*-glycan displayed here is one example of many different types of glycans present in nature. Mild acid hydrolysis efficiently removes sialic acids from glycans, exposing the peripheral saccharides, which improves accessibility of lectin. Smith degradation removes most of the saccharides, leaving only the core structure of the glycan. Both the *N*-acetylglucosamine and *N*-acetylgalactosamine are less susceptible to the mild periodate oxidation due to presence of *N*-acetyl group and saccharide at carbon number three.

saccharides since they are based on core structures 1 to 5 in which position 3 of GalNAc is substituted making it resistant to periodate oxidation (Andrews *et al.*, 1969; Duk *et al.*, 1994; Perlin, 2006). Experiments with model glycoconjugates that were performed in the present study revealed that desialylation by mild acid hydrolysis alone was sufficient for maximal binding. Periodate oxidation apparently did not further increase binding of the lectins selected for this study. These experiments also showed that of the three lectins tested, jacalin provided the highest sensitivity of binding to the desialylated *O*-glycosylated proteins. This is probably due to the reported broad specificity of jacalin for *O*-glycosylated proteins (Bourne *et al.*, 2002; Strous & Dekker, 1992; Wu, 2001). Therefore, biotinylated jacalin was selected as the most suitable detecting reagent in all subsequent experiments.

As expected, the experiments of the present study showed that the sandwich binding approach in which a lectin was used to capture the *O*-glycosylated proteins is superior to direct coating of the sample onto the plastic microtiter plates. This is particularly relevant for crude samples having high levels of non-glycosylated proteins, such as serum, which are more likely to bind to plastic than the heavily glycosylated *O*-glycosylated proteins molecules that are of much lower concentrations. When different lectins were tested as the coating capture reagent, the best sensitivity in the developed sandwich ELLA was obtained using CGB lectin. This is in agreement with the earlier reported data suggesting that the CGB lectin and jacalin, despite having close structural similarities (Abdul-Rahman *et al.*, 2002; Gabrielsen *et al.*, 2014), demonstrated slightly distinct binding specificities (Hashim *et al.*, 1991).

Possible interaction of *N*-glycosylated proteins with jacalin may occur since previous studies have shown jacalin interacts minimally with mannose and *N*-glycosylated proteins (Aucouturier *et al.*, 1988; Bourne *et al.*, 2002; Gabrielsen *et al.*, 2014). Therefore, the possibility of *N*-glycosylated proteins to interact with the

developed sandwich ELLA was also tested. Based on the result, there was no evidence of *N*-glycosylated proteins interfering with the sandwich ELLA. Hence, this concludes that the developed assay is highly specific towards *O*-glycosylated proteins and that there was no interaction with other proteins in complex samples.

5.3 Application of sandwich ELLA on normal human serum samples

In the subsequent analysis, a single type of glycoprotein standard was used in order to maintain result consistency. AsialoBSM was chosen as the *O*-glycosylated protein standard because this mucin displayed higher sensitivity compared to the other test glycoproteins. In addition, asialoBSM appeared to show better solubility in water compared to asialoPGM. Low protein solubility causes the concentration of the standard to be inconsistent and therefore affects the reproducibility of the sandwich ELLA.

As mentioned earlier, treatment of glycoproteins with mild acid hydrolysis was sufficient to increase the binding of lectins. Hence, mild acid hydrolysis was performed on serum samples before subjecting them to the sandwich ELLA to quantify total *O*-glycosylated proteins. As shown in the Results section (Section 4.1.8), prior mild acid hydrolysis on serum was able to improve the binding of lectins. Harsh conditions like acidic condition and high temperature may cause proteins to precipitate. In order to minimise protein precipitation, serum samples were first diluted at least 10 times from the stock concentration before being subjected to mild acid hydrolysis. It appeared that the developed sandwich ELLA was sensitive enough to measure *O*-glycosylated proteins even at low concentration and sample dilution did not seem to hinder the performance of the sandwich ELLA.

When the developed CGB lectin-jacalin and jacalin-jacalin sandwich ELLAs were used to estimate the *O*-glycosylated proteins content in three human serum samples, the values obtained using the two assays were slightly different. Clearly, the developed assays were not likely to yield absolute values for the *O*-glycosylated proteins content of test samples such as body fluids containing one or more *O*-glycosylated proteins of unknown structures. Aside from the type of *O*-glycosylated proteins used as the standard, the values obtained are also dependent on the lectins used in the sandwich ELLA. In the present study, the higher values obtained using the jacalin-jacalin sandwich ELLA compared to CGB lectin-jacalin sandwich ELLA may be due to the slight different binding specificities that were previously observed. But more importantly, the results obtained in the present study clearly demonstrates that the sandwich assay using CGB lectin as the coating lectin gave the highest sensitivity not only with model *O*-glycosylated proteins (BSM and PGM) but also with serum, a crude biological sample.

As a conclusion, the data of the present study appears to demonstrate that it is preferable to use two different lectins rather than the same lectin for both coating and probing purposes, in a sandwich assay. This was shown when the assay, using CGB lectin as the capture lectin and jacalin as the probe lectin was used in comparison with that using jacalin as both the capture and probe lectin to analyse total *O*-glycosylated proteins in the human serum. Figure 5.2 demonstrates the final concept of the ELLA that was developed in the present study. The application of this assay is further described in subsequent part of the thesis, in which the developed CGB lectin-jacalin sandwich ELLA was used for estimation of total *O*-glycosylated proteins and inter- and intra- variation of the sandwich assay of a serum sample was determined.

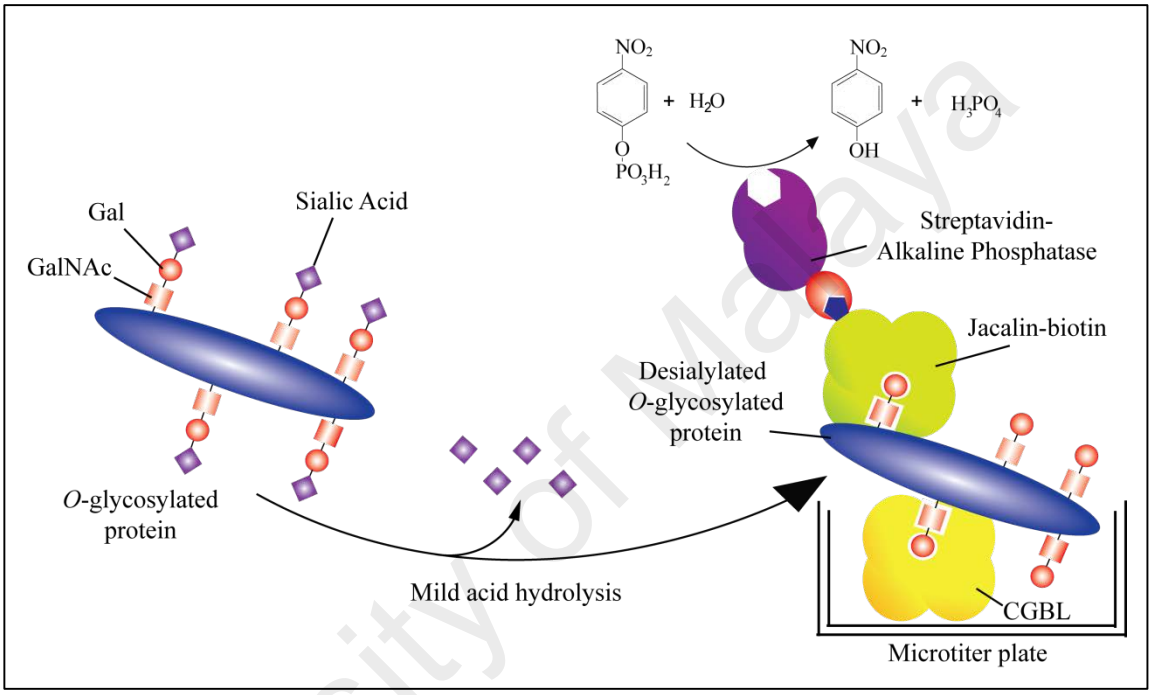


Figure 5.2: CGB lectin-Jacalin sandwich ELLA

Figure demonstrates the strategy involved in the newly developed CGB lectin-jacalin sandwich ELLA. Samples were initially desialylated by incubation in 0.1 N H_2SO_4 at 80°C before being subjected to CGB lectin-jacalin sandwich ELLA. The assay uses CGB lectin and biotinylated-jacalin as the capture and probe reagents, respectively. Streptavidin-alkaline phosphatase was allowed to interact with the probe lectin, which then converts p-nitrophenyl phosphate to a colored product that absorbs at 415 nm.

5.4 Enrichment of serum *O*-glycosylated proteins using perchloric acid

In the previous section, the ability of CGB lectin-jacalin sandwich ELLA to estimate *O*-glycosylated proteins in serum samples was demonstrated. Since the developed assay uses CGB lectin, which is known to interact with serum IgA1 to capture *O*-glycosylated proteins, mucin-type glycoproteins which are of lower abundances than the immunoglobulin, may have lesser chance to bind to the pre-coated lectin due to large amounts of IgA1 that are usually present in the serum. At present, affinity chromatography using antisera to deplete proteins of high abundance is one of the most popular techniques used for enrichment of other proteins. However, this has achieved little success as the proteins that are usually detected in samples like serum or plasma after being subjected to the immunodepletion technique are still considered to be highly abundant (Seriramalu *et al.*, 2010).

Alternatively, lectins can be immobilised onto beads, instead of antisera, in a similar concept of affinity chromatogarphy to target and isolate glycoproteins from biological samples, which can then be further analysed (Caragata *et al.*, 2016; Dong *et al.*, 2015; Song & Mechref, 2015). When the CGB lectin was immobilised and used as the stationary phase in affinity chromatography to separate neat human serum samples, only *O*-glycosylated proteins were selectively retained (Jayapalan *et al.*, 2012). However, this method was still unable to detect *O*-glycosylated proteins of low abundance as presence of high amounts of IgA1, which strongly binds to the lectin, hindered their interactions with the lectin. Therefore, the present study had explored a chemical method in an attempt to enrich the low abundance *O*-glycosylated proteins in complex biological sample such the human serum.

Treatment with perchloric acid, a traditional method which was first reported by Winzler *et al*. (1948) to isolate mucoproteins from plasma, precipitates non- and minimally-glycosylated proteins, leaving those that were highly glycosylated in the supernatant. In the present proteomics study, two different investigative approaches, *i.e.*, two-dimensional gel electrophoresis (2-DE) and QTOF LC/MS, which is a liquid chromatography-based mass spectrometry, were initially used to profile heavily glycosylated proteins that were solubilised when human serum samples were subjected to perchloric acid treatment. The two distinctive approaches are highly capable of detecting proteins in complex biological samples. However, both have unique advantages as well as limitations. For example, 2-DE, which is capable of separating large numbers of proteins in a complex biological sample, is known to have high protein separation ability with great resolution. In addition, the gel-based method has the advantage of being able to identify truncated proteins as well as their distinctive isoforms. However, it poorly detects proteins of low abundance and those that are highly acidic or basic, and incapable of separating hydrophobic proteins and those with high molecular weights. The separation of proteins by 2-DE is also labour-intensive and the method has lower reproducibility (Abdallah *et al.*, 2012). On the other hand, liquid chromatography-mass spectrometry may provide better reproducibility and is certainly less laborious. Proteins that are unable to be resolved using 2-DE, particularly those with larger molecular sizes, are likely to be detected using the non-gel based approach.

In the present study, eight proteins, including alpha-1-antitrypsin, alpha-1-acid glycoproteins, beta-2-glycoprotein 1, prothrombin, hemopexin, transthyretin, protein AMB, and corticosteroid-binding globulin, were consistently identified to be present in the serum perchloric acid isolates using both the 2-DE/MS and QTOF LC/MS techniques. Ten additional proteins, serum albumin, alpha-1-acid glycoprotein, alpha-2-

HS glycoprotein, antithrombin III, haptoglobin, prostaglandin-H2 D-isomerase, Ig kappa chain C region, apolipoprotein A-I, clusterin, and ankyrin repeat domain containing protein 42 were detected using 2-DE/MS, whilst five other proteins, plasma protease C1 inhibitor, apolipoprotein A-IV, complement C4-A/B, coagulation factor V, and platelet basic protein were identified using QTOF LC/MS. A Venn diagram to illustrate proteins that were present in the serum perchloric acid isolates that were identified using 2-DE/MS and/or QTOF LC/MS is available in section 4.2.3.

Despite the eight common proteins that were detected using both the 2-DE/MS and QTOF LC/MS techniques, the data of the present study generally showed a considerable number of other proteins that were only identified using either one of the methods used. For example, haptoglobin appeared to be only detected using 2-DE/MS, whereas plasma protease C1 inhibitor, which is relatively considered a large glycoprotein, was selectively detected using QTOF LC/MS. On the other hand, Zn-alpha-2-glycoprotein, which was earlier reported to be solubilised in perchloric acid (Anderson, 1965), was not identified using both the proteomics methods used. The different sets of proteins that were uniquely detected using the two different techniques of separation suggest that each platform has its own advantages as well as limitations for protein detection. As a matter of fact, this has been previously shown from studies comparing 2-DE/MS and LC/MS (Irar *et al.*, 2014; Pottiez *et al.*, 2010). Hence, there is no single proteomic platform that is capable of identifying all proteins in a complex sample such as serum (Abdallah *et al.*, 2012). Perhaps more importantly in this study is the fact that IgA1 was not detected in both proteomics analyses, and thus provides the evidence that perchloric acid treatment was capable of depleting the immunoglobulin from the serum samples.

Successful protein identification using mass spectrometry system is dependent on several limiting factors such as the availability of protein database and the quality of tryptic digested peptides. Although most of the spots were excised from the 2-DE gel and analysed, not all of them were successfully identified using the MALDI-TOF/TOF and database query. These may be due to the low concentrations of the peptides or lack of trypsin digestion sites. Some proteins, such as mucins, contain multiple repeating units of serine or threonine (Bansil & Turner, 2006) and therefore may have lesser tryptic digestion sites (lysine or arginine residues) that are required for protein identification. In QTOF LC/MS, in-solution digestion of a protein mixture sample was performed prior to the mass spectrometry analysis. However, it is practically impossible to maintain complete digestion of all proteins in a complex samples due to the limited enzymatic activity (Xie *et al.*, 2011). As a result, lesser numbers of proteins were detected using QTOF LC/MS as it compared to the profile generated by 2-DE/MS. Furthermore, perchloric acid treatment may have modified some of the peptides and as a result, rendering them unable to be matched with those in the protein database. Nevertheless, and despite such limitations, the proteomic analyses that were performed in this study were able to identify quite a substantial number of proteins from serum perchloric acid isolates, including those that have not been previously reported. The two profiles of serum proteins that were solubilised in perchloric acid that were established in the present study certainly offer a valuable reference for similar investigations in the future.

Amongst the heavily glycosylated serum proteins that were retained in perchloric acid, alpha-1-acid glycoprotein is most abundant. This serum glycoprotein, whose glycans structures are clearly defined (Nakano *et al.*, 2004), makes up about 70% of the total glycoproteins enriched using the method (Burston *et al.*, 1963). Aside from alpha-1-acid glycoprotein, other heavily glycosylated serum proteins such as

hemopexin, alpha-1-acid glycoprotein 2, Zn-alpha-2-glycoprotein and haptoglobin were also reported to be solubilized in perchloric acid (Anderson, 1965). In the present proteomics study, however, Zn-alpha-2-glycoprotein was not identified amongst the glycoproteins that were retained in perchloric acid although 13 other glycoproteins which were not earlier reported, *i.e.*, plasma protease C1 inhibitor, alpha-1-antitrypsin, beta-2-glycoprotein 1, prothrombin, transthyretin, protein AMB, apolipoprotein A-IV, complement C4-A/B, corticosteroid binding globulin, coagulation factor V, alpha-2-HS glycoprotein, antithrombin III and prostaglandin-H2 D-isomerase were additionally detected.

One of the main reasons for treating serum samples with perchloric acid in the present study was to enrich heavily *O*-glycosylated proteins. Whether the proteins that were earlier detected using 2-DE/MS and LC/MS were indeed glycosylated can be confirmed by performing two different analyses. Firstly, the proteins were investigated for any potential *O*-glycosylation based on the UNIPROT database. And secondly, CGB lectin blotting was performed to confirm presence of *O*-glycosylation.

In the present study, eight proteins, i.e, plasma protease C1 inhibitor, hemopexin, alpha-2-HS glycoprotein, beta-2-glycoprotein 1, prostaglandin-H2, D-isomerase protein, protein AMB, and complement C4-A, were found to be *O*-glycosylated based on the UNIPROT database. When blotted using the CGB lectin, five proteins, *i.e.*, hemopexin, beta-2-glycoprotein 1, corticosteroid-binding globulin, alpha-2-HS glycoprotein and protein AMB, showed strong interaction. Interestingly, whilst hemopexin, beta-2-glycoprotein 1, alpha-2-HS glycoprotein and protein AMB have been documented as having *O*-glycans in the UNIPROT database, corticosteroid-binding globulin was not. Hence, the interaction of CGB lectin with corticosteroid-binding globulin that was shown in the present study provides early indication of the

protein having *O*-glycans, although direct evidence using glycomic analysis on this is very much required.

The data of the present study, when taken in general, demonstrated the high capability of perchloric acid to enrich substantial amounts of *O*-glycosylated proteins in serum samples. In a separate experiment, perchloric acid treatment of serum was further shown to be able to enhance the detection limit of the earlier developed sandwich ELLA, an assay used to measure *O*-glycosylated proteins, by 14-fold. The data generated from the analysis of the recovery of asialoBSM demonstrated that perchloric acid was able to isolate over 90% of heavily glycosylated proteins in serum samples. Hence, this has allowed the earlier developed sandwich ELLA to detect other low-abundant *O*-glycosylated proteins in the serum.

Previous investigations had also reported the use of perchloric acid to enhance detectability of an assay. Perchloric acid has been commonly used as deproteinisation agent on serum samples to improve the detectability of pyruvate and lactate (Koshiishi *et al.,* 1998; Lu *et al.,* 2015; Marbach & Weil, 1967). The research has proven that the use of perchloric acid on serum could enhance detection of the metabolites. In another study, the use of perchloric acid to unmask presence of high mobility group box 1 protein from serum samples was demonstrated, and as a result, the efficiency of the conventional ELISA to detect the protein was greatly improved (Barnay-Verdier *et al.,* 2011). Clearly, these experimental findings, including those generated from the present study, indicate that perchloric acid is not only capable of enriching *O*-glycosylated serum proteins but also has the ability to unmask presence of the low abundant glycoproteins that are not usually detected using conventional protocols.

5.5 Levels of *O*-glycosylated proteins in early breast cancer

The cancer antigens CA 27.29 and CA 15-3, which are derived from a mucin known as MUC-1, are commonly used biomarkers for breast cancer (Duffy *et al.*, 2010; Pichon *et al.*, 2009). These biomarkers have low sensitivity and specificity to detect early stages of breast cancer (Duffy *et al.*, 2010) albeit they had officially received approval by the Food and Drug Administration in year 1997 (Fuzery *et al.*, 2013). Aside from these, there is no other reliable serum/plasma biomarker for breast cancer detection that receives recognition. Much biomarker research often focused on invasive cancer that was detected at the late stages of the disease (Mar-Aguilar *et al.*, 2013; Miller *et al.*, 2003; Sun *et al.*, 2007). This may eventually result in poor prognosis as cancer patients who are diagnosed at the early stages have higher rates of survival compared to those detected at the late stages (ACS, 2015). Therefore, the present study has put major emphasis on the non-invasive breast cancer and the earliest invasive breast cancer, which are often classified as stage 0 and stage I breast cancer, respectively (Ellis *et al.*, 2003). The discovery of biomarkers for these early stages of breast cancer would be more beneficial than looking at the later stages of the cancer. Furthermore, early detection of a non-invasive cancer would allow better control of the tumour from turning into an invasive cancer.

Tumour cells of cancer patients are known to actively release many different types of *O*-glycosylated proteins into the blood circulation (Gomes *et al.*, 2013; Pinho & Reis, 2015; Reis *et al.*, 2010). In the present study, sandwich ELLA performed on serum samples of patients with stage 0 and stage I breast cancer demonstrated highly elevated levels of total *O*-glycosylated proteins compared to those obtained from controls. When similar experiments were repeated using perchloric acid enriched fractions of serum samples of the same groups of subjects, the levels of *O*-glycosylated proteins also appeared highly elevated in both stage 0 and stage I breast cancer patients

despite that IgA1 had been removed from the supernatants. This is indicative of presence of additional heavily *O*-glycosylated proteins in the serum perchloric acid isolates that might have been captured by the sandwich ELLA in absence of IgA1.

The data of the present study is compatible with those earlier reported by Winzler and co-workers (Winzler *et al.*, 1948; Winzler & Smyth, 1948). In their study, the researchers demonstrated that perchloric acid was able to solubilise significantly higher amounts of mucoproteins from the plasma of cancer patients compared to the controls. However, the studies had analysed plasma samples from late stage cancer subjects as opposed to the present study, which analysed the serum samples of patients with the early stages of breast cancer. In the present study, substantial improvement of sensitivity and specificity of sandwich ELLA to detect both stage 0 and stage I breast cancer was also achieved using the perchloric acid-enriched serum samples. Hence, the elevated levels of total *O*-glycosylated proteins in breast cancer patients using sandwich ELLA were generally observed. The next step of this study was to determine the specific *O*-glycosylated proteins that might have contributed to the different levels observed.

Several methods, including lectin affinity chromatography (Jayapalan *et al.*, 2012) and lectin blotting (Abdul-Rahman *et al.*, 2002; Mohamed *et al.*, 2013), have been earlier used in studies identifying the composition of *O*-glycosylated proteins in sample fractions. However, these methods require substantial glycoprotein samples. In the present study, the amounts of proteins being retained in perchloric acid after its treatment were extremely low. Such an amount is not suitably used for analysis using both lectin affinity chromatography and the lectin blotting. In view of this, a method that utilised CGB-lectin-conjugated beads to capture the heavily *O*-glycosylated proteins in serum perchloric acid isolates prior to proteins separation through SDS-PAGE was adopted. The main advantage of this method is that it only requires minimal

amount of serum sample to be used. Secondly, it allows solubilisation of all *O*-glycosylated proteins that were bound to the lectin conjugated beads directly in Laemmli buffer before being resolved by SDS-PAGE. Using this method and by applying heat, high molecular weight proteins (>100 kDa) were able to be released from the lectin and these proteins were clearly resolved when analysed by SDS-PAGE. In contrast, the use of lectin affinity column chromatography would not able to release all proteins that were bound, especially those with high molecular weights. This is because the strong interaction between lectin and the high molecular weight proteins may not be easily removed by sugar elution. Although strong acids may be used to weaken the interaction, this may also disrupt the conjugated lectin causing it to be no longer effective. Therefore, the use of denaturing agents, such as Laemmli buffer, and heat treatment offers great advantage. Despite that the CGB lectin-conjugated-beads may also be disrupted, the amounts utilised were extremely minimal, and hence, reutilisation was not quite necessary.

When comparative analysis of the CGB lectin bound-perchloric acid enriched *O*-glycosylated proteins of the controls and patients with stage 0 and stage I breast cancer was performed, different electrophoretic profiles were clearly generated. Densitometry of the bands and analysis by mass spectrometry and database search demonstrated two identified serum proteins with more than 1.5-fold difference in abundance. Whilst, the intensity of proteoglycan 4 appeared significantly elevated in serum perchloric acid isolates of both stage 0 and stage I breast cancer samples, the bands of plasma protease C1 inhibitor proteins appeared significantly more intense in those of the controls. Hence, proteoglycan 4 may be one of the heavily *O*-glycosylated proteins whose levels were earlier shown to be elevated in the serum samples of patients with breast cancer. Interestingly, proteoglycan 4 was not identified in the earlier

analysis of glycoproteins that were solubilised in perchloric acid using both 2DE-MS and QTOF-LC/MS.

In view of the reciprocal trend of altered abundance of proteoglycan 4 and plasma protease C1 inhibitor between the breast cancer patients and cancer free women, the two serum *O*-glycosylated proteins stand to offer as strong complementary biomarker candidates for detection of early breast cancer although this requires validation in clinically representative populations using simpler high-throughput methods. At these early stages of breast cancer, treatment is generally more effective. In the case of stage 0 or carcinoma in situ, curative treatments precluding chemotherapy can be instituted.

Proteoglycan 4 is a secreted mucinous glycoprotein found highly in abundance in the synovial fluid of mammalian diarthrodal joints (Swann *et al.*, 1985). The glycoprotein, which is also known as lubricin, functions as a boundary lubricant at the articular cartilage surface (Jay, 1992; Swann *et al.*, 1977). However, proteoglycan 4 is not solely expressed in the articular tissues. Studies have shown that the glycoprotein is also expressed in tissues of the heart, liver, bone (Ikegawa *et al.*, 2000), eye (Schmidt *et al.*, 2013) and breast (Cheriyan *et al.*, 2012) as well as in the blood circulation (Jin *et al.*, 2012). This suggests that it may have other functions aside from lubrication of joints. One of the unique features of this protein is that it contains multiple *O*-linked oligosaccharide chains, which comprise mostly core 1 (Galβ1,3GalNAcα1-*O*-Ser/Thr) and core 2 (Galβ1,3[GlcNAcβ1,6]GalNAcα1-*O*-Ser/Thr) structures attached to a serine/threonine/proline rich (STP-rich) domain (Ali *et al.*, 2014). Interestingly, a recent study capitalizing on the advances in glycoproteomics using innovative cancer cell genetic engineering (Campos *et al.*, 2015b; Steentoft *et al.*, 2011) has clearly identified proteoglycan 4 as one of a variety of *O*-glycosylated proteins with the truncated STn (NeuAcα2-6GalNAcα1-*O*-Ser/Thr) glycoform in the sera of gastric

carcinoma patients (Campos *et al.*, 2015a). However, several attempts made to validate its altered levels in serum samples of breast cancer patients using different antisera in the present study had so far been unsuccessful.

Plasma protease C1 inhibitor, the second glycoprotein that was detected to be of altered abundance in stage 0 and stage I breast cancer patients, belongs to the serine proteinase inhibitor (serpin) family. Also known as serpin G1, it is active against complement C1s and C1r, kallikrein, and coagulation factor XIIa, and hence its roles as a major inhibitor of complement and in regulating physiological pathways such as blood coagulation, fibrinolysis and the generation of kinins (Cicardi *et al.*, 2005). Lower levels of plasma protease C1 inhibitor often relate to a rare genetic disease known as hereditary angioedema (Caliezi *et al.*, 2000; Triggianese *et al.*, 2015). Previously, the plasma protein has been shown to be elevated in 64 patients with several different malignancies using radial immunodiffusion (Starcevic *et al.*, 1991). In the present study, however, the intensity of plasma protease C1 inhibitor bands generated from serum perchloric acid isolates of normal control women appeared more intense than those generated from both stage 0 and stage I breast cancer patients. Although this generally reflects lower abundance of the protease inhibitor in the breast cancer patients, it could also be because of their differences in *O*-glycosylation and/or solubility in perchloric acid. Plasma protease C1 inhibitor is known to be heavily glycosylated with both the *N*- and *O*-linked oligosaccharide moieties (Bock *et al.*, 1986), and alteration of glycan structures of glycoproteins is also not an uncommon phenomenon in malignancies (Adamczyk *et al.*, 2012; Gomes *et al.*, 2013; Reis *et al.*, 2010).

In the present study, the altered abundances of the glycoproteins observed in breast cancer patients were also validated using Western blotting and ELISA. Western blot was preferable over ELISA as a validation method because this technique allows visual detection of the specific proteins and is also able to verify that the antisera used

bind specifically to the proteins of interest. In addition, identities of proteins that were derived from mass spectrometry can be further confirmed using Western blot. Although ELISA is a technique that is commonly used in many clinical diagnostics, it is not always possible to validate identities of proteins as ELISA is very much dependent on the quality of the assay and specificity of the antibody used.

When commercially developed antisera against proteoglycan 4 and plasma protease C1 inhibitor were used to detect the protein bands that were transferred onto nitrocellulose membrane, only the latter was successfully detected. The failure of the antiserum used to detect proteoglycan 4 may be attributed to a number of reasons. Firstly, perchloric acid treatment of the serum sample may have contributed to the structural changes on proteoglycan 4 and the antibody used in the experiment may no longer recognise the changed epitopes. Secondly, the commercial antisera that were used may not be able to interact with proteoglycan 4 that was transferred on blotting membrane. Although studies on proteoglycan 4 using the Western blot approach were fairly common in synovial fluid and purified protein samples (Ai *et al.*, 2015; Ali *et al.*, 2014; Estrella *et al.*, 2010), its detection is less commonly seen in the serum.

In the present study, the same reduced expression of protease C1 inhibitor in patients with early stages of breast cancer relative to the controls was further observed when CGB lectin precipitated serum perchloric acid isolates were allowed to interact with commercial antisera against the protease inhibitor in a Western blot experiment. However, when the Western blot experiment was repeated on neat serum samples that were subjected to acid hydrolysis, the same anti-protease C1 inhibitor used appeared to detect relatively similar intensities of the protein bands in the cancer-negative control samples as well as the stage 0 and stage I breast cancer samples. Contrary to the earlier speculation made, these results generally indicate the same levels of plasma protease C1 inhibitor that were present in all the samples that were analysed. Therefore, when taken

together with the earlier data obtained from CGB lectin precipitated serum perchloric acid isolates, it becomes apparent that the differential expression of protease C1 inhibitor that was earlier observed in patients with stage 0 and stage I breast cancer patients relative to the controls is more likely due to the altered glycosylation of the serum protein. The reduced interaction of CGB lectin with protease C1 inhibitor from patients with stage 0 and stage I breast cancer is suggestive of its under-galactosylation. However, direct evidences for this is very much required, and hence, this calls for further structural studies of the *O*-glycans of plasma protease C1 inhibitor in patients with early breast cancer in the future.

CHAPTER 6: CONCLUSION

In the present study, a sensitive method to detect mucin type *O*-glycosylated proteins using lectins was successfully developed. This sandwich lectin binding assay involved prior sample treatment with mild acid hydrolysis to remove sialic acids, and utilised CGB lectin and biotinylated-jacalin as capture and probe lectins, respectively. The developed method was demonstrated to be sensitive enough to estimate the levels of *O*-glycosylated proteins in the human serum, and showed high reproducibility. In addition, the assay did not display any non-specific binding towards other proteins, particularly the *N*-glycosylated proteins.

In the second part of the study, the protein composition of serum perchloric isolates were analysed using two proteomics platforms, *i.e.*, 2-DE/MS and the QTOF LC/MS. Amongst 23 proteins that were successfully identified using both the techniques, eight were found out to be *O*-glycosylated. Detection limit of the earlier developed assay was found to be significantly improved when serum samples were treated with perchloric acid prior to their estimation for total *O*-glycosylated proteins. The analysis further demonstrated the capability of perchloric acid to isolate substantial amounts of *O*-glycosylated proteins, whilst precipitated other non- or minimally glycosylated proteins, particularly IgA1, from the serum samples. When sandwich ELLA was used to analyse serum samples of patients with stage 0 and stage I breast cancer, significantly higher levels of total *O*-glycosylated proteins were detected compared to the control subjects. A much improved results was obtained when the analysis was repeated using serum perchloric acid isolate samples.

When the *O*-glycosylated proteins from serum perchloric isolates of breast cancer patients and control subjects were isolated using CGB lectin-conjugated beads and bound glycoproteins were subsequently analysed using SDS-PAGE, the

proteoglycan 4 bands appeared significantly elevated in serum perchloric acid isolates of both stage 0 and stage I breast cancer samples, relative to the controls. In contrast, the intensity of plasma protease C1 inhibitor bands was significantly lower compared to the serum perchloric acid isolates of the control subjects. However, further analysis by Western blot and immunodetection confirmed that the levels of the protease inhibitor were not significantly altered in the patients with both the early stages of breast cancer. Hence, this suggests that the reduced expression of protease C1 inhibitor observed in the CGB lectin bound protein of serum perchloric acid isolates was more likely due to structural changes of its *O*-glycan moieties.

REFERENCES

http://www.rcsb.org/pdb/pv/pv.do?pdbid=4AK4&bionumber=1

http://www.rcsb.org/pdb/pv/pv.do?pdbid=1KU8&bionumber=1

http://www.rcsb.org/pdb/pv/pv.do?pdbid=1N47&bionumber=1

http://www.breastcancer.org/symptoms/diagnosis/staging

http://www.gelanalyzer.com/).

Abdallah, C., Dumas-Gaudot, E., Renaut, J., & Sergeant, K. (2012). Gel-based and gel-free quantitative proteomics approaches at a glance. *Int J Plant Genomics, 2012*, 1-17.

Abdul-Rahman, M., Anuar Karsani, S., Othman, I., Abdul Rahman, P. S., & Hashim, O. H. (2002). Galactose-binding lectin from the seeds of champedak (Artocarpus integer): sequences of its subunits and interactions with human serum O-glycosylated glycoproteins. *Biochem Biophys Res Commun, 295*(4), 1007-1013.

Abdul-Rahman, P. S., Lim, B. K., & Hashim, O. H. (2007). Expression of high-abundance proteins in sera of patients with endometrial and cervical cancers: analysis using 2-DE with silver staining and lectin detection methods. *Electrophoresis, 28*(12), 1989-1996.

American Cancer Society (2015). *Breast Cancer*. American Cancer Society: Atlanta, GA.

Adamczyk, B., Tharmalingam, T., & Rudd, P. M. (2012). Glycans as cancer biomarkers. *Biochim Biophys Acta, Gen Subj, 1820*(9), 1347-1353.

Ahmed, H., & Chatterjee, B. P. (1989). Further characterization and immunochemical studies on the carbohydrate specificity of jackfruit (Artocarpus-Integrifolia) Lectin. *J Biol Chem, 264*(16), 9365-9372.

Ahn, H. S., Kim, S. M., Jang, M., La Yun, B., Kim, S. W., Kang, E., . . . Choi, H. Y. (2013). Comparison of sonography with sonographically guided fine-needle aspiration biopsy and core-needle biopsy for initial axillary staging of breast cancer. *J Ultrasound Med, 32*(12), 2177-2184.

Ahn, Y. H., Ji, E. S., Oh, N. R., Kim, Y. S., Ko, J. H., & Yoo, J. S. (2014). Differential proteomic approach for identification and verification of aberrantly glycosylated proteins in adenocarcinoma lung cancer (ADLC) plasmas by lectin-capturing and targeted mass spectrometry. *J Proteomics, 106*, 221-229.

Ai, M., Cui, Y., Sy, M. S., Lee, D. M., Zhang, L. X., Larson, K. M., . . . Warman, M. L. (2015). Anti-lubricin monoclonal antibodies created using lubricin-knockout mice immunodetect lubricin in several species and in patients with healthy and diseased joints. *PLoS One, 10*(2), e0116237.

Ali, L., Flowers, S. A., Jin, C., Bennet, E. P., Ekwall, A. K., & Karlsson, N. G. (2014). The O-glycomap of lubricin, a novel mucin responsible for joint lubrication, identified by site-specific glycopeptide analysis. *Mol Cell Proteomics, 13*(12), 3396-3409.

Allen, A. C., Harper, S. J., & Feehally, J. (1995). Galactosylation of N- and O-linked carbohydrate moieties of IgA1 and IgG in IgA nephropathy. *Clin Exp Immunol, 100*(3), 470-474.

Anderson, A. J. (1965). Factors affecting the amount and composition of the serum seromucoid fraction. *Nature, 208*(5009), 491-492.

Andrews, A. T., Herring, G. M., & Kent, P. W. (1969). The periodate oxidation of bovine bone sialoprotein, and some observations on its structure. *Biochem J, 111*(5), 621-627.

Angata, T., & Brinkman-Van der Linden, E. (2002). I-type lectins. *Biochim Biophys Acta, 1572*(2-3), 294-316.

Argueso, P., & Guzman-Aranguez, A. (2010). Structure and Biological Roles of Mucin-type O-glycans at the Ocular Surface. *Ocul Surf, 8*(1), 8-17.

Aucouturier, P., Duarte, F., Mihaesco, E., Pineau, N., & Preud'homme, J. L. (1988). Jacalin, the human IgA1 and IgD precipitating lectin, also binds IgA2 of both allotypes. *J Immunol Methods, 113*(2), 185-191.

Babino, A., Tello, D., Rojas, A., Bay, S., Osinaga, E., & Alzari, P. M. (2003). The crystal structure of a plant lectin in complex with the Tn antigen. *FEBS Lett, 536*(1-3), 106-110.

Bafna, S., Kaur, S., & Batra, S. K. (2010). Membrane-bound mucins: the mechanistic basis for alterations in the growth and survival of cancer cells. *Oncogene, 29*(20), 2893-2904.

Balague, C., Gambus, G., Carrato, C., Porchet, N., Aubert, J. P., Kim, Y. S., & Real, F. X. (1994). Altered expression of MUC2, MUC4, and MUC5 mucin genes in pancreas tissues and cancer cell lines. *Gastroenterology, 106*(4), 1054-1061.

Bansil, R., & Turner, B. S. (2006). Mucin structure, aggregation, physiological functions and biomedical applications. *Curr Opin Colloid Interface Sci, 11*(2-3), 164-170.

Barkhordari, A., Stoddart, R. W., McClure, S. F., & McClure, J. (2004). Lectin histochemistry of normal human lung. *J Mol Histol, 35*(2), 147-156.

Barnay-Verdier, S., Gaillard, C., Messmer, M., Borde, C., Gibot, S., & Marechal, V. (2011). PCA-ELISA: a sensitive method to quantify free and masked forms of HMGB1. *Cytokine, 55*(1), 4-7.

Barratt, J., Smith, A. C., & Feehally, J. (2007). The pathogenic role of IgA1 O-linked glycosylation in the pathogenesis of IgA nephropathy. *Nephrology, 12*(3), 275-284.

Bast, R. C., & Spriggs, D. R. (2011). More than a biomarker: CA125 may contribute to ovarian cancer pathogenesis. *Gynecol. Oncol., 121*(3), 429-430.

Berner, A., Davidson, B., Sigstad, E., & Risberg, B. (2003). Fine-needle aspiration cytology vs. core biopsy in the diagnosis of breast lesions. *Diagn Cytopathol, 29*(6), 344-348.

Bhavanandan, V. P. (1991). Cancer-associated mucins and mucin-type glycoproteins. *Glycobiology, 1*(5), 493-503.

Bhavanandan, V. P., & Furukawa, K. (1995). Biochemistry and oncology of sialoglycoproteins. In A. Rosenberg (Ed.), *Biology of the sialic acids*. New York: Plenum Press.

Bhavanandan, V. P., & Hegarty, J. D. (1987). Identification of the mucin core protein by cell-free translation of messenger RNA from bovine submaxillary glands. *J Biol Chem, 262*(12), 5913-5917.

Bhavanandan, V. P., Sheykhnazari, M., & Devaraj, H. (1990). Colorimetric determination of N-acetylhexosamine-terminating O-glycosidically linked saccharides in mucins and glycoproteins. *Anal Biochem, 188*(1), 142-148.

Bock, S. C., Skriver, K., Nielsen, E., Thogersen, H. C., Wiman, B., Donaldson, V. H., . . . et al. (1986). Human C1 inhibitor: primary structure, cDNA cloning, and chromosomal localization. *Biochemistry, 25*(15), 4292-4301.

Bouckaert, J., Berglund, J., Schembri, M., De Genst, E., Cools, L., Wuhrer, M., . . . De Greve, H. (2005). Receptor binding studies disclose a novel class of high-affinity inhibitors of the Escherichia coli FimH adhesin. *Mol Microbiol, 55*(2), 441-455.

Bourne, Y., Astoul, C. H., Zamboni, V., Peumans, W. J., Menu-Bouaouiche, L., Van Damme, E. J., . . . Rouge, P. (2002). Structural basis for the unusual carbohydrate-binding specificity of jacalin towards galactose and mannose. *Biochem J, 364*(Pt 1), 173-180.

Brockhausen, I. (1999). Pathways of O-glycan biosynthesis in cancer cells. *Biochim Biophys Acta, 1473*(1), 67-95.

Brockhausen, I. (2006). Mucin-type O-glycans in human colon and breast cancer: glycodynamics and functions. *Embo Rep, 7*(6), 599-604.

Brockhausen, I., Schachter, H., & Stanley, P. (2009). O-GalNAc Glycans. In A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart & M. E. Etzler (Eds.), *Essentials of Glycobiology* (2nd ed.). Cold Spring Harbor (NY).

Brockhausen, I., Yang, J. M., Burchell, J., Whitehouse, C., & Taylor-Papadimitriou, J. (1995). Mechanisms underlying aberrant glycosylation of MUC1 mucin in breast cancer cells. *Eur J Biochem, 233*(2), 607-617.

Bunkenborg, J., Pilch, B. J., Podtelejnikov, A. V., & Wisniewski, J. R. (2004). Screening for N-glycosylated proteins by liquid chromatography mass spectrometry. *Proteomics, 4*(2), 454-465.

Bunn-Moreno, M. M., & Campos-Neto, A. (1981). Lectin(s) extracted from seeds of Artocarpus integrifolia (jackfruit): potent and selective stimulator(s) of distinct human T and B cell functions. *J Immunol, 127*(2), 427-429.

Burbank, F., Parker, S. H., & Fogarty, T. J. (1996). Stereotactic breast biopsy: improved tissue harvesting with the Mammotome. *Am Surg, 62*(9), 738-744.

Burston, D., Tombs, M. P., Apsey, M. E., & Maclagan, N. F. (1963). The perchloric acid soluble basic and acidic proteins of the cytoplasm: variation in cancer. *Br. J. Cancer, 17*, 162-178.

Caliezi, C., Wuillemin, W. A., Zeerleder, S., Redondo, M., Eisele, B., & Hack, C. E. (2000). C1-Esterase inhibitor: an anti-inflammatory agent and its potential use in the treatment of diseases other than hereditary angioedema. *Pharmacol. Rev., 52*(1), 91-112.

Campbell, M. P., Royle, L., Radcliffe, C. M., Dwek, R. A., & Rudd, P. M. (2008). GlycoBase and autoGU: tools for HPLC-based glycan analysis. *Bioinformatics, 24*(9), 1214-1216.

Campos, D., Freitas, D., Gomes, J., Magalhaes, A., Steentoft, C., Gomes, C., . . . Reis, C. A. (2015a). Probing the O-glycoproteome of gastric cancer cell lines for biomarker discovery. *Mol Cell Proteomics, 14*(6), 1616-1629.

Campos, D., Freitas, D., Gomes, J., & Reis, C. A. (2015b). Glycoengineered cell models for the characterization of cancer O-glycoproteome: an innovative strategy for biomarker discovery. *Expert Rev Proteomics, 12*(4), 337-342.

Caragata, M., Shah, A. K., Schulz, B. L., Hill, M. M., & Punyadeera, C. (2016). Enrichment and identification of glycoproteins in human saliva using lectin magnetic bead arrays. *Anal Biochem, 497*, 76-82.

Carlson, D. M. (1968). Structures and immunochemical properties of oligosaccharides isolated from pig submaxillary mucins. *J Biol Chem, 243*(3), 616-626.

Cedolini, C., Bertozzi, S., Londero, A. P., Bernardi, S., Seriau, L., Concina, S., . . . Risaliti, A. (2014). Type of breast cancer diagnosis, screening, and survival. *Clin Breast Cancer, 14*(4), 235-240.

Chatterjee, B., Vaith, P., Chatterjee, S., Karduck, D., & Uhlenbruck, G. (1979). Comparative studies of new marker lectins for alkali-labile and alkali-stable carbohydrate chains in glycoproteins. *Int J Biochem, 10*(4), 321-327.

Chefetz, I., & Sprecher, E. (2009). Familial tumoral calcinosis and the role of O-glycosylation in the maintenance of phosphate homeostasis. *Biochim Biophys Acta, Mol Basis Dis, 1792*(9), 847-852.

Chen, R., Jiang, X., Sun, D., Han, G., Wang, F., Ye, M., . . . Zou, H. (2009). Glycoproteomics analysis of human liver tissue by combination of multiple enzyme digestion and hydrazide chemistry. *J Proteome Res, 8*(2), 651-661.

Chen, Y., Peumans, W. J., Hause, B., Bras, J., Kumar, M., Proost, P., . . . Van Damme, E. J. (2002). Jasmonic acid methyl ester induces the synthesis of a cytoplasmic/nuclear chito-oligosaccharide binding lectin in tobacco leaves. *FASEB J, 16*(8), 905-907.

Cheriyan, T., Guo, L., Orgill, D. P., Padera, R. F., Schmid, T. M., & Spector, M. (2012). Lubricin in human breast tissue expander capsules. *J Biomed Mater Res B Appl Biomater, 100*(7), 1961-1969.

Cheung, R. C., Leung, H. H., Pan, W. L., & Ng, T. B. (2013). A calcium ion-dependent dimeric bean lectin with antiproliferative activity toward human breast cancer MCF-7 cells. *Protein J, 32*(3), 208-215.

Christiansen, M. N., Chik, J., Lee, L., Anugraham, M., Abrahams, J. L., & Packer, N. H. (2014). Cell surface protein glycosylation in cancer. *Proteomics, 14*(4-5), 525-546.

Cicardi, M., Zingale, L., Zanichelli, A., Pappalardo, E., & Cicardi, B. (2005). C1 inhibitor: molecular and clinical aspects. *Springer Semin Immunopathol, 27*(3), 286-298.

Corbeau, P., Haran, M., Binz, H., & Devaux, C. (1994). Jacalin, a lectin with anti-HIV-1 properties, and HIV-1 gp120 envelope protein interact with distinct regions of the CD4 molecule. *Mol Immunol, 31*(8), 569-575.

Cummings, R. D., & Esko, J. D. (2009). Principles of glycan recognition. In A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart & M. E. Etzler (Eds.), *Essentials of Glycobiology* (2nd ed.). Cold Spring Harbor (NY).

Dall'olio, F. (1996). Protein glycosylation in cancer biology: an overview. *Clin Mol Pathol, 49*(3), 126-135.

Damerell, D., Ceroni, A., Maass, K., Ranzinger, R., Dell, A., & Haslam, S. M. (2015). Annotation of glycomics MS and MS/MS spectra using the GlycoWorkbench software tool. *Methods Mol Biol, 1273*, 3-15.

David, L., Carneiro, F., & SobrinhoSimoes, M. (1996). Sialosyl Tn antigen expression is associated with the prognosis of patients with advanced gastric cancer. *Cancer, 78*(1), 177-178.

De Hoff, P. L., Brill, L. M., & Hirsch, A. M. (2009). Plant lectins: the ties that bind in root symbiosis and plant defense. *Mol Genet Genomics, 282*(1), 1-15.

Dekker, J., Rossen, J. W., Buller, H. A., & Einerhand, A. W. (2002). The MUC family: an obituary. *Trends Biochem Sci, 27*(3), 126-131.

Dent, R., Trudeau, M., Pritchard, K. I., Hanna, W. M., Kahn, H. K., Sawka, C. A., . . . Narod, S. A. (2007). Triple-negative breast cancer: clinical features and patterns of recurrence. *Clin Cancer Res, 13*(15 Pt 1), 4429-4434.

Dias, R. D., Machado, L. D., Migliolo, L., & Franco, O. L. (2015). Insights into Animal and Plant Lectins with Antimicrobial Activities. *Molecules, 20*(1), 519-541.

Dong, L., Feng, S., Li, S., Song, P., & Wang, J. (2015). Preparation of concanavalin A-Chelating magnetic nanoparticles for selective enrichment of glycoproteins. *Anal Chem, 87*(13), 6849-6853.

Drake, R. R., Schwegler, E. E., Malik, G., Diaz, J., Block, T., Mehta, A., & Semmes, O. J. (2006). Lectin capture strategies combined with mass spectrometry for the discovery of serum glycoprotein biomarkers. *Mol Cell Proteomics, 5*(10), 1957-1967.

Duffy, M. J., Evoy, D., & McDermott, E. W. (2010). CA 15-3: Uses and limitation as a biomarker for breast cancer. *Clin Chim Acta, 411*, 1869-1874.

Duffy, M. J., Shering, S., Sherry, F., McDermott, E., & O'Higgins, N. (2000). CA 15-3: a prognostic marker in breast cancer. *Int J Biol Markers, 15*(4), 330-333.

Duk, M., Lisowska, E., Wu, J. H., & Wu, A. M. (1994). The Biotin/Avidin-Mediated Microtiter Plate Lectin Assay with the Use of Chemically-Modified Glycoprotein Ligand. *Anal Biochem, 221*(2), 266-272.

Edge, A. S. B., & Spiro, R. G. (1987). Presence of an O-glycosidically linked hexasaccharide in fetuin. *J Biol Chem, 262*(33), 16135-16141.

Ellis, I. O., Cornelisse, C. J., Schnitt, S. J., Sasco, A. J., Sastre-Garau, X., Kaaks, R., . . . Sapino, A. (2003). Tumours of the breast. In Fattaneh A. Tavassoli & P. Devilee (Eds.), *Pathology and Genetics of Tumours of the Breast and Female Genitals Organs*. Lyon: IARCPress.

Ellis, I. O., Galea, M., Broughton, N., Locker, A., Blamey, R. W., & Elston, C. W. (1992). Pathological prognostic factors in breast cancer. II. Histological type. Relationship with survival in a large study with long-term follow-up. *Histopathology, 20*(6), 479-489.

Elston, C. W., & Ellis, I. O. (1991). Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology, 19*(5), 403-410.

Estrella, R. P., Whitelock, J. M., Packer, N. H., & Karlsson, N. G. (2010). The glycosylation of human synovial lubricin: implications for its role in inflammation. *Biochem J, 429*(2), 359-367.

Felix, J. C. (1999). Pathology of the female breast. In W. H. Hindle (Ed.), *Breast Care: A Clinical Guidebook for Women's Primary Health Care Providers*. Los Angeles, CA: Springer.

Fuzery, A. K., Levin, J., Chan, M. M., & Chan, D. W. (2013). Translation of proteomic biomarkers into FDA approved cancer diagnostics: issues and challenges. *Clin Proteomics, 10*(1), 13.

Gabrielsen, M., Abdul-Rahman, P. S., Othman, S., Hashim, O. H., & Cogdell, R. J. (2014). Structures and binding specificity of galactose- and mannose-binding lectins from champedak: differences from jackfruit lectins. *Acta Crystallogr F Struct Biol Commun, 70*(Pt 6), 709-716.

Gendler, S. J. (2001). MUC1, the renaissance molecule. *J Mammary Gland Biol Neoplasia, 6*(3), 339-353.

Geng, Y., Takatani, T., Yeh, K., Hsu, J. W., & King, M. R. (2013). Targeting Underglycosylated MUC1 for the Selective Capture of Highly Metastatic Breast Cancer Cells Under Flow. *Cell Mol Bioeng, 6*(2), 148-159.

Gerken, T. A., Ten Hagen, K. G., & Jamison, O. (2008). Conservation of peptide acceptor preferences between Drosophila and mammalian polypeptide-GalNAc transferase ortholog pairs. *Glycobiology, 18*(11), 861-870.

Ghazarian, H., Idoni, B., & Oppenheimer, S. B. (2011). A glycobiology review: carbohydrates, lectins and implications in cancer therapeutics. *Acta Histochem, 113*(3), 236-247.

Ghosh, S. K., Pantazopoulos, P., Medarova, Z., & Moore, A. (2013). Expression of underglycosylated MUC1 antigen in cancerous and adjacent normal breast tissues. *Clin Breast Cancer, 13*(2), 109-118.

Gomes, C., Almeida, A., Ferreira, J. A., Silva, L., Santos-Sousa, H., Pinto-de-Sousa, J., . . . Osorio, H. (2013). Glycoproteomic analysis of serum from patients with gastric precancerous lesions. *J Proteome Res, 12*(3), 1454-1466.

Gornik, O., & Lauc, G. (2007). Enzyme linked lectin assay (ELLA) for direct analysis of transferrin sialylation in serum samples. *Clin Biochem, 40*(9-10), 718-723.

Hagerbaumer, P., Vieth, M., Anders, M., & Schumacher, U. (2015). Lectin Histochemistry Shows WGA, PHA-L and HPA Binding Increases During Progression of Human Colorectal Cancer. *Anticancer Research, 35*(10), 5333-5339.

Halim, A., Nilsson, J., Ruetschi, U., Hesse, C., & Larson, G. (2012). Human urinary glycoproteomics; attachment site specific analysis of N- and O-linked glycosylations by CID and ECD. *Mol Cell Proteomics, 11*(4), M111 013649.

Halim, A., Ruetschi, U., Larson, G., & Nilsson, J. (2013). LC-MS/MS characterization of O-glycosylation sites and glycan structures of human cerebrospinal fluid glycoproteins. *J Proteome Res, 12*(2), 573-584.

Harrell, W. K., & George, J. R. (1972). Quantitative measurement of precipitating antibodies in streptococcal grouping antisera by the single radial immunodiffusion technique. *Appl Microbiol, 23*(6), 1047-1052.

Hashim, O. H., Gendeh, G. S., & Jaafar, M. I. (1993). Comparative analyses of IgA1 binding lectins from seeds of six distinct clones of Artocarpus integer. *Biochem Mol Biol Int, 29*(1), 69-76.

Hashim, O. H., Ng, C. L., Gendeh, S., & Nik Jaafar, M. I. (1991). IgA binding lectins isolated from distinct Artocarpus species demonstrate differential specificity. *Mol Immunol, 28*(4-5), 393-398.

Hauselmann, I., & Borsig, L. (2014). Altered tumor-cell glycosylation promotes metastasis. *Front Oncol, 4*, 28.

Hermanson, G. T., Mallia, A. K., & Smith, P. K. (1992). Immobilized Affinity Ligand Techniques: Academic Press.

Hiki, Y., Kokubo, T., Iwase, H., Masaki, Y., Sano, T., Tanaka, A., . . . Kobayashi, Y. (1999). Underglycosylation of IgA1 hinge plays a certain role for its glomerular deposition in IgA nephropathy. *J Am Soc Nephrol, 10*(4), 760-769.

Hirabayashi, J., Kuno, A., & Tateno, H. (2011). Lectin-based structural glycomics: A practical approach to complex glycans. *Electrophoresis, 32*(10), 1118-1128.

Hu, S., & Wong, D. T. (2009). Lectin microarray. *Proteomics Clin Appl, 3*(2), 148-154.

Huang, W. L., Li, Y. G., Lv, Y. C., Guan, X. H., Ji, H. F., & Chi, B. R. (2014). Use of lectin microarray to differentiate gastric cancer from gastric ulcer. *World J Gastroenterol, 20*(18), 5474-5482.

Iizuka, M., Nakanishi, Y., Fuchinoue, F., Maeda, T., Murakami, E., Obana, Y., . . . Masuda, S. (2015). Altered intracellular region of MUC1 and disrupted correlation of polarity-related molecules in breast cancer subtypes. *Cancer Sci, 106*(3), 307-314.

Ikeda, Y., Kuwano, H., Baba, K., Ikebe, M., Matushima, T., Adachi, Y., . . . Sugimachi, K. (1993). Expression of Sialyl-Tn antigens in normal squamous epithelium, dysplasia, and squamous cell carcinoma in the esophagus. *Cancer Res, 53*(7), 1706-1708.

Ikegawa, S., Sano, M., Koshizuka, Y., & Nakamura, Y. (2000). Isolation, characterization and mapping of the mouse and human PRG4 (proteoglycan 4) genes. *Cytogenet Cell Genet, 90*(3-4), 291-297.

Inoue, M., Ogawa, H., Tanizawa, O., Kobayashi, Y., Tsujimoto, M., & Tsujimura, T. (1991). Immunodetection of sialyl-Tn antigen in normal, hyperplastic and cancerous tissues of the uterine endometrium. *Virchows Arch A Pathol Anat Histopathol, 418*(2), 157-162.

Irar, S., Brini, F., Masmoudi, K., & Pages, M. (2014). Combination of 2DE and LC for plant proteomics analysis. *Methods Mol Biol, 1072*, 131-140.

Itai, S., Arii, S., Tobe, R., Kitahara, A., Kim, Y. C., Yamabe, H., . . . Kannagi, R. (1988). Significance of 2-3 and 2-6 sialylation of Lewis a antigen in pancreas cancer. *Cancer, 61*(4), 775-787.

Itai, S., Nishikata, J., Yoneda, T., Ohmori, K., Yamabe, H., Arii, S., . . . Kannagi, R. (1991). Tissue distribution of 2-3 and 2-6 sialyl Lewis A antigens and significance of the ratio of two antigens for the differential diagnosis of malignant and benign disorders of the digestive tract. *Cancer, 67*(6), 1576-1587.

Itzkowitz, S. H., Yuan, M., Montgomery, C. K., Kjeldsen, T., Takahashi, H. K., Bigbee, W. L., & Kim, Y. S. (1989). Expression of Tn, sialosyl-Tn, and T antigens in human colon cancer. *Cancer Res, 49*(1), 197-204.

Jagtap, U. B., & Bapat, V. A. (2010). Artocarpus: a review of its traditional uses, phytochemistry and pharmacology. *J Ethnopharmacol, 129*(2), 142-166.

Jay, G. D. (1992). Characterization of a bovine synovial fluid lubricating factor. I. Chemical, surface activity and lubricating properties. *Connect Tissue Res, 28*(1-2), 71-88.

Jayapalan, J. J., Ng, K. L., Razack, A. H. A., & Hashim, O. H. (2012). Identification of potential complementary serum biomarkers to differentiate prostate cancer from benign prostatic hyperplasia using gel- and lectin-based proteomics analyses. *Electrophoresis, 33*(12), 1855-1862.

Jayapalan, J. J., Ng, K. L., Shuib, A. S., Razack, A. H., & Hashim, O. H. (2013). Urine of patients with early prostate cancer contains lower levels of light chain fragments of inter-alpha-trypsin inhibitor and saposin B but increased expression

of an inter-alpha-trypsin inhibitor heavy chain 4 fragment. *Electrophoresis, 34*(11), 1663-1669.

Jeyaprakash, A. A., Srivastav, A., Surolia, A., & Vijayan, M. (2004). Structural basis for the carbohydrate specificities of artocarpin: variation in the length of a loop as a strategy for generating ligand specificity. *J Mol Biol, 338*(4), 757-770.

Jin, C., Ekwall, A. K., Bylund, J., Bjorkman, L., Estrella, R. P., Whitelock, J. M., . . . Karlsson, N. G. (2012). Human synovial lubricin expresses sialyl Lewis x determinant and has L-selectin ligand activity. *J Biol Chem, 287*(43), 35922-35933.

Joy, J. E., Penhoet, E. E., & Petitti, D. B. (2005). Benefits and Limitations of Mammography. In J. E. Joy, E. E. Penhoet & D. B. Petitti (Eds.), *Saving Women's Lives: Strategies for Improving Breast Cancer Detection and Diagnosis*. Washington (DC).

Ju, T. Z., Lanneau, G. S., Gautam, T., Wang, Y. C., Xia, B. Y., Stowell, S. R., . . . Cummings, R. D. (2008). Human tumor antigens to and sialyl Tn arise from mutations in Cosmc. *Cancer Res, 68*(6), 1636-1646.

Julien, S., Videira, P. A., & Delannoy, P. (2012). Sialyl-tn in cancer: (how) did we miss the target? *Biomolecules, 2*(4), 435-466.

Kabir, S. (1998). Jacalin: a jackfruit (Artocarpus heterophyllus) seed-derived lectin of versatile applications in immunobiological research. *J Immunol Methods, 212*(2), 193-211.

Kabir, S. R., Zubair, M. A., Nurujjaman, M., Haque, M. A., Hasan, I., Islam, M. F., . . . Absar, N. (2011). Purification and characterization of a Ca(2+)-dependent novel lectin from Nymphaea nouchali tuber with antiproliferative activities. *Biosci Rep, 31*(6), 465-475.

Kaplan, H. G., Malmgren, J. A., Atwood, M. K., & Calip, G. S. (2015). Effect of treatment and mammography detection on breast cancer survival over time: 1990-2007. *Cancer, 121*(15), 2553-2561.

Kaur, S., Momi, N., Chakraborty, S., Wagner, D. G., Horn, A. J., Lele, S. M., . . . Batra, S. K. (2014). Altered expression of transmembrane mucins, MUC1 and MUC4, in bladder cancer: pathological implications in diagnosis. *PLoS One, 9*(3), e92742.

Kawaguchi, T., Takazawa, H., Imai, S., Morimoto, J., Watanabe, T., Kanno, M., & Igarashi, S. (2006). Expression of Vicia villosa agglutinin (VVA)-binding glycoprotein in primary breast cancer cells in relation to lymphatic metastasis: is

atypical MUC1 bearing Tn antigen a receptor of VVA? *Breast Cancer Res Treat, 98*(1), 31-43.

Khan, F., Khan, R. H., Sherwani, A., Mohmood, S., & Azfer, M. A. (2002). Lectins as markers for blood grouping. *Med Sci Monit, 8*(12), RA293-300.

Kilcoyne, M., Gerlach, J. Q., Farrell, M. P., Bhavanandan, V. P., & Joshi, L. (2011). Periodic acid-Schiff's reagent assay for carbohydrates in a microtiter plate format. *Anal Biochem, 416*(1), 18-26.

Kim, G. E., Bae, H. I., Park, H. U., Kuan, S. F., Crawley, S. C., Ho, J. J., & Kim, Y. S. (2002). Aberrant expression of MUC5AC and MUC6 gastric mucins and sialyl Tn antigen in intraepithelial neoplasms of the pancreas. *Gastroenterology, 123*(4), 1052-1060.

Kimura, A., Wigzell, H., Holmquist, G., Ersson, B., & Carlsson, P. (1979). Selective affinity fractionation of murine cytotoxic T lymphocytes (CTL). Unique lectin specific binding of the CTL associated surface glycoprotein, T 145. *J Exp Med, 149*(2), 473-484.

Kondoh, H., Kobayashi, K., & Hagiwara, K. (1987). A simple procedure for the isolation of human secretory IgA of IgA1 and IgA2 subclass by a jackfruit lectin, jacalin, affinity chromatography. *Mol Immunol, 24*(11), 1219-1222.

Konska, G., Guerry, M., Caldefie-Chezet, F., De Latour, M., & Guillot, J. (2006). Study of the expression of Tn antigen in different types of human breast cancer cells using VVA-B4 lectin. *Oncol Rep, 15*(2), 305-310.

Koshiishi, I., Mamura, Y., Liu, J., & Imanari, T. (1998). Evaluation of an acidic deproteinization for the measurement of ascorbate and dehydroascorbate in plasma samples. *Clin Chem, 44*(4), 863-868.

Kotani, N., & Takasaki, S. (1997). Analysis of O-linked oligosaccharide alditols by high-pH anion-exchange chromatography with pulsed amperometric detection. *Anal Biochem, 252*(1), 40-47.

Kudelka, M. R., Ju, T., Heimburg-Molinaro, J., & Cummings, R. D. (2015). Simple sugars to complex disease--mucin-type O-glycans in cancer. *Adv Cancer Res, 126*, 53-135.

Kufe, D. W. (2009). Mucins in cancer: function, prognosis and therapy. *Nat Rev Cancer, 9*(12), 874-885.

Lee, Y. C. (1990). High-Performance Anion-Exchange Chromatography for Carbohydrate Analysis. *Anal Biochem, 189*(2), 151-162.

Leivonen, M., Nordling, S., Lundin, J., von Boguslawski, K., & Haglund, C. (2001). STn and prognosis in breast cancer. *Oncology, 61*(4), 299-305.

Li, Y. G., Wen, T., Zhu, M. Z., Li, L. X., Wei, J., Wu, X. L., . . . Yin, Z. N. (2013). Glycoproteomic analysis of tissues from patients with colon cancer using lectin microarrays and nanoLC-MS/MS. *Mol BioSyst, 9*(7), 1877-1887.

Liikanen, J., Leidenius, M., Joensuu, H., Vironen, J., Heikkila, P., & Meretoja, T. (2016). Breast cancer prognosis and isolated tumor cell findings in axillary lymph nodes after core needle biopsy and fine needle aspiration cytology: Biopsy method and breast cancer outcome. *Eur J Surg Oncol, 42*(1), 64-70.

Lim, S. B., Chua, C. T., & Hashim, O. H. (1997). Isolation of a mannose-binding and IgE- and IgM-reactive lectin from the seeds of Artocarpus integer. *J Immunol Methods, 209*(2), 177-186.

Lin, S., Kemmner, W., Grigull, S., & Schlag, P. M. (2002). Cell surface alpha 2,6 sialylation affects adhesion of breast carcinoma cells. *Exp Cell Res, 276*(1), 101-110.

Liu, D., Cramer, C. C., Scafidi, J., & Davis, A. E., 3rd. (2005a). N-linked glycosylation at Asn3 and the positively charged residues within the amino-terminal domain of the c1 inhibitor are required for interaction of the C1 Inhibitor with Salmonella enterica serovar typhimurium lipopolysaccharide and lipid A. *Infect Immun, 73*(8), 4478-4487.

Liu, T., Qian, W. J., Gritsenko, M. A., Camp, D. G., 2nd, Monroe, M. E., Moore, R. J., & Smith, R. D. (2005b). Human plasma N-glycoproteome analysis by immunoaffinity subtraction, hydrazide chemistry, and mass spectrometry. *J Proteome Res, 4*(6), 2070-2080.

Lloyd, K. O., Burchell, J., Kudryashov, V., Yin, B. W., & Taylor-Papadimitriou, J. (1996). Comparison of O-linked carbohydrate chains in MUC-1 mucin from normal breast epithelial cell lines and breast carcinoma cell lines. Demonstration of simpler and fewer glycan chains in tumor cells. *J Biol Chem, 271*(52), 33325-33334.

Lu, J., Pulsipher, B. S., & Grenache, D. G. (2015). Development of an enzymatic assay to measure lactate in perchloric acid-precipitated whole blood. *Clin Chim Acta, 444*, 208-211.

Maass, K., Ranzinger, R., Geyer, H., von der Lieth, C. W., & Geyer, R. (2007). "Glyco-peakfinder"--de novo composition analysis of glycoconjugates. *Proteomics, 7*(24), 4435-4444.

Macedo, M. L. R., Oliveira, C. F. R., & Oliveira, C. T. (2015). Insecticidal Activity of Plant Lectins and Potential Application in Crop Protection. *Molecules, 20*(2), 2014-2033.

Manzi, A. E., Diaz, S., & Varki, A. (1990). High-pressure liquid chromatography of sialic acids on a pellicular resin anion-exchange column with pulsed amperometric detection: a comparison with six other systems. *Anal Biochem, 188*(1), 20-32.

Mar-Aguilar, F., Mendoza-Ramirez, J. A., Malagon-Santiago, I., Espino-Silva, P. K., Santuario-Facio, S. K., Ruiz-Flores, P., . . . Resendez-Perez, D. (2013). Serum circulating microRNA profiling for identification of potential breast cancer biomarkers. *Dis Markers, 34*(3), 163-169.

Marbach, E. P., & Weil, M. H. (1967). Rapid enzymatic measurement of blood lactate and pyruvate. Use and significance of metaphosphoric acid as a common precipitant. *Clin Chem, 13*(4), 314-325.

Mathenge, E. G., Dean, C. A., Clements, D., Vaghar-Kashani, A., Photopoulos, S., Coyle, K. M., . . . Giacomantonio, C. A. (2014). Core needle biopsy of breast cancer tumors increases distant metastases in a mouse model. *Neoplasia, 16*(11), 950-960.

McCoy, J. P., Jr., Varani, J., & Goldstein, I. J. (1983). Enzyme-linked lectin assay (ELLA): use of alkaline phosphatase-conjugated Griffonia simplicifolia B4 isolectin for the detection of alpha-D-galactopyranosyl end groups. *Anal Biochem, 130*(2), 437-444.

McGuckin, M. A., Linden, S. K., Sutton, P., & Florin, T. H. (2011). Mucin dynamics and enteric pathogens. *Nat Rev Microbiol, 9*(4), 265-278.

Miller, J. C., Zhou, H., Kwekel, J., Cavallo, R., Burke, J., Butler, E. B., . . . Haab, B. B. (2003). Antibody microarray profiling of human prostate cancer sera: antibody screening and identification of potential biomarkers. *Proteomics, 3*(1), 56-63.

Miyazaki, K., Ohmori, K., Izawa, M., Koike, T., Kumamoto, K., Furukawa, K., . . . Kannagi, R. (2004). Loss of disialyl Lewis(a), the ligand for lymphocyte inhibitory receptor sialic acid-binding immunoglobulin-like lectin-7 (Siglec-7) associated with increased sialyl Lewis(a) expression on human colon cancers. *Cancer Res, 64*(13), 4498-4505.

Mohamed, E., Abdul-Rahman, P. S., Doustjaialil, S. R., Chen, Y., Lim, B. K., Omar, S. Z., . . . Hashim, O. H. (2008). Lectin-based electrophoretic analysis of the expression of the 35 kDa inter-alpha-trypsin inhibitor heavy chain H4 fragment in sera of patients with five different malignancies. *Electrophoresis, 29*(12), 2645-2650.

Mohamed, E., Jayapalan, J. J., Abdul-Rahman, P. S., Omar, S. Z., & Hashim, O. H. (2013). Enhanced expression of a 35 kDa fragment of inter-alpha-trypsin inhibitor H4 in sera of healthy pregnant women and patients with hydatidiform mole. *Biomark Res, 1*(1), 19.

Morelle, W., Flahaut, C., Michalski, J. C., Louvet, A., Mathurin, P., & Klein, A. (2006). Mass spectrometric approach for screening modifications of total serum N-glycome in human diseases: application to cirrhosis. *Glycobiology, 16*(4), 281-293.

Morelle, W., & Michalski, J. C. (2005). The mass spectrometric analysis of glycoproteins and their glycan structures. *Current Analytical Chemistry, 1*(1), 29-57.

Mukhopadhyay, P., Chakraborty, S., Ponnusamy, M. P., Lakshmanan, I., Jain, M., & Batra, S. K. (2011). Mucins in the pathogenesis of breast cancer: Implications in diagnosis, prognosis and therapy. *Biochim Biophys Acta, Rev Cancer, 1815*(2), 224-240.

Nakajima, K., Inomata, M., Iha, H., Hiratsuka, T., Etoh, T., Shiraishi, N., . . . Kitano, S. (2015). Establishment of new predictive markers for distant recurrence of colorectal cancer using lectin microarray analysis. *Cancer Med, 4*(2), 293-302.

Nakano, M., Kakehi, K., Tsai, M. H., & Lee, Y. C. (2004). Detailed structural features of glycan chains derived from alpha1-acid glycoproteins of several different animals: the presence of hypersialylated, O-acetylated sialic acids but not disialyl residues. *Glycobiology, 14*(5), 431-441.

Narita, I., & Gejyo, F. (2008). Pathogenetic significance of aberrant glycosylation of IgA1 in IgA nephropathy. *Clin Exp Nephrol, 12*(5), 332-338.

Nilsson, J., Ruetschi, U., Halim, A., Hesse, C., Carlsohn, E., Brinkmalm, G., & Larson, G. (2009). Enrichment of glycopeptides for glycan structure and attachment site identification. *Nat Methods, 6*(11), 809-811.

Nishijima, Y., Toyoda, M., Yamazaki-Inoue, M., Sugiyama, T., Miyazawa, M., Muramatsu, T., . . . Mikami, M. (2012). Glycan profiling of endometrial cancers using lectin microarray. *Genes Cells, 17*(10), 826-836.

Nishiyama, T., Matsumoto, Y., Watanabe, H., Fujiwara, M., & Sato, S. (1987). Detection of Tn antigen with Vicia villosa agglutinin in urinary bladder cancer: its relevance to the patient's clinical course. *J Natl Cancer Inst, 78*(6), 1113-1118.

Novotny, M. V., Alley, W. R., Jr., & Mann, B. F. (2013). Analytical glycobiology at high sensitivity: current approaches and directions. *Glycoconj J, 30*(2), 89-117.

Numa, F., Tsunaga, N., Michioka, T., Nawata, S., Ogata, H., & Kato, H. (1995). Tissue expression of Sialyl Tn antigen in gynecologic tumors. *J Obstet Gynaecol (Tokyo, Jpn), 21*(4), 385-389.

Olson, J. E., Sellers, T. A., Scott, C. G., Schueler, B. A., Brandt, K. R., Serie, D. J., . . . Vachon, C. M. (2012). The influence of mammogram acquisition on the mammographic density and breast cancer association in the Mayo Mammography Health Study cohort. *Breast Cancer Res, 14*(6), R147.

Omar, Z. A., & Tamin, N. S. I. (2011). *NCR Report 2007*. Malaysia: Ministry of Health.

Osinaga, E., Tello, D., Batthyany, C., Bianchet, M., Tavares, G., Duran, R., . . . Alzari, P. M. (1997). Amino acid sequence and three-dimensional structure of the Tn-specific isolectin B4 from Vicia villosa. *FEBS Lett, 412*(1), 190-196.

Perlin, A. S. (2006). Glycol-cleavage oxidation. *Adv Carbohydr Chem Biochem, 60*, 183-250.

Peumans, W. J., & Van Damme, E. J. (1995). Lectins as plant defense proteins. *Plant Physiol, 109*(2), 347-352.

Pichon, M. F., Le Brun, G., Hacene, K., Basuyau, J. P., Riedinger, J. M., Eche, N., . . . Charlier-Bret, N. (2009). Comparison of fifteen immunoassays for the measurement of serum MUC-1/CA 15-3 in breast cancer patients. *Clin Chem Lab Med, 47*(8), 985-992.

Pinho, S. S., & Reis, C. A. (2015). Glycosylation in cancer: mechanisms and clinical implications. *Nat Rev Cancer, 15*(9), 540-555.

Pitteri, S. J., Chrisman, P. A., Hogan, J. M., & McLuckey, S. A. (2005). Electron transfer ion/ion reactions in a three-dimensional quadrupole ion trap: reactions of doubly and triply protonated peptides with SO2*. *Anal Chem, 77*(6), 1831-1839.

Pottiez, G., Deracinois, B., Duban-Deweer, S., Cecchelli, R., Fenart, L., Karamanos, Y., & Flahaut, C. (2010). A large-scale electrophoresis- and chromatography-based

determination of gene expression profiles in bovine brain capillary endothelial cells after the re-induction of blood-brain barrier properties. *Proteome Sci, 8*, 57.

Rabinovich, G. A., Toscano, M. A., Jackson, S. S., & Vasta, G. R. (2007). Functions of cell surface galectin-glycoprotein lattices. *Curr Opin Struct Biol, 17*(5), 513-520.

Rahn, J. J., Dabbagh, L., Pasdar, M., & Hugh, J. C. (2001). The importance of MUC1 cellular localization in patients with breast carcinoma: an immunohistologic study of 71 patients and review of the literature. *Cancer, 91*(11), 1973-1982.

Rakha, E. A., Boyce, R. W., Abd El-Rehim, D., Kurien, T., Green, A. R., Paish, E. C., . . . Ellis, I. O. (2005). Expression of mucins (MUC1, MUC2, MUC3, MUC4, MUC5AC and MUC6) and their prognostic significance in human breast cancer. *Mod Pathol, 18*(10), 1295-1304.

Reis, C. A., Osorio, H., Silva, L., Gomes, C., & David, L. (2010). Alterations in glycosylation as biomarkers for cancer detection. *J Clin Pathol, 63*(4), 322-329.

Rohrer, J. S., & Townsend, R. R. (1995). Separation of partially desialylated branched oligosaccharide isomers containing alpha (2-->3)- and alpha (2-->6)-linked Neu5Ac. *Glycobiology, 5*(4), 391-395.

Roque-Barreira, M. C., Praz, F., Halbwachs-Mecarelli, L., Greene, L. J., & Campos-Neto, A. (1986). IgA-affinity purification and characterization of the lectin jacalin. *Braz J Med Biol Res, 19*(2), 149-157.

Ross, J. S., Hatzis, C., Symmans, W. F., Pusztai, L., & Hortobagyi, G. N. (2008). Commercialized multigene predictors of clinical outcome for breast cancer. *Oncologist, 13*(5), 477-493.

Roussel, P., Lamblin, G., Lhermitte, M., Houdret, N., Lafitte, J. J., Perini, J. M., . . . Scharfman, A. (1988). The complexity of mucins. *Biochimie, 70*(11), 1471-1482.

Rudiger, H., & Gabius, H. J. (2001). Plant lectins: occurrence, biochemistry, functions and applications. *Glycoconj J, 18*(8), 589-613.

Sankaranarayanan, R., Sekar, K., Banerjee, R., Sharma, V., Surolia, A., & Vijayan, M. (1996). A novel mode of carbohydrate recognition in jacalin, a Moraceae plant lectin with a beta-prism fold. *Nat Struct Biol, 3*(7), 596-603.

Saunders, C., & Jassal, S. (2009). *Breast Cancer*: Oxford University Press.

Saxon, A., Tsui, F., & Martinez-Maza, O. (1987). Jacalin, an IgA-binding lectin, inhibits differentiation of human B cells by both a direct effect and by activating T-suppressor cells. *Cell Immunol, 104*(1), 134-141.

Schmidt, T. A., Sullivan, D. A., Knop, E., Richards, S. M., Knop, N., Liu, S. H., . . . Sullivan, B. D. (2013). Transcription, Translation, and Function of Lubricin, a Boundary Lubricant, at the Ocular Surface. *Jama Ophthalmol., 131*(6), 766-776.

Seriramalu, R., Pang, W. W., Jayapalan, J. J., Mohamed, E., Abdul-Rahman, P. S., Bustam, A. Z., . . . Hashim, O. H. (2010). Application of champedak mannose-binding lectin in the glycoproteomic profiling of serum samples unmasks reduced expression of alpha-2 macroglobulin and complement factor B in patients with nasopharyngeal carcinoma. *Electrophoresis, 31*(14), 2388-2395.

Sharon, N., & Lis, H. (2004). History of lectins: from hemagglutinins to biological recognition molecules. *Glycobiology, 14*(11), 53R-62R.

Sinn, H. P., & Kreipe, H. (2013). A Brief Overview of the WHO Classification of Breast Tumors, 4th Edition, Focusing on Issues and Updates from the 3rd Edition. *Breast Care (Basel), 8*(2), 149-154.

Sipponen, P., & Lindgren, J. (1986). Sialylated Lewis determinant CA 19-9 in benign and malignant gastric tissue. *Acta. Pathol. Microbiol. Immunol. Scand. A., 94*(5), 305-311.

Song, E., & Mechref, Y. (2015). Defining glycoprotein cancer biomarkers by MS in conjunction with glycoprotein enrichment. *Biomark Med, 9*(9), 835-844.

Song, K., Herzog, B. H., Fu, J., Sheng, M., Bergstrom, K., McDaniel, J. M., . . . Xia, L. (2015). Loss of Core 1-derived O-Glycans Decreases Breast Cancer Development in Mice. *J Biol Chem, 290*(33), 20159-20166.

Stanley, P., Schachter, H., & Taniguchi, N. (2009). N-Glycans. In A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart & M. E. Etzler (Eds.), *Essentials of Glycobiology* (2nd ed.). Cold Spring Harbor (NY).

Starcevic, D., Jelic-Ivanovic, Z., & Kalimanovska, V. (1991). Plasma C1 inhibitor in malignant diseases: functional activity versus concentration. *Ann Clin Biochem, 28 (Pt 6)*, 595-598.

Steentoft, C., Vakhrushev, S. Y., Vester-Christensen, M. B., Schjoldager, K. T. B. G., Kong, Y., Bennett, E. P., . . . Clausen, H. (2011). Mining the O-glycoproteome using zinc-finger nuclease-glycoengineered SimpleCell lines. *Nat Methods, 8*(11), 977-982.

Storr, S. J., Royle, L., Chapman, C. J., Hamid, U. M., Robertson, J. F., Murray, A., . . . Rudd, P. M. (2008). The O-linked glycosylation of secretory/shed MUC1 from an advanced breast cancer patient's serum. *Glycobiology, 18*(6), 456-462.

Strous, G. J., & Dekker, J. (1992). Mucin-type glycoproteins. *Crit Rev Biochem Mol Biol, 27*(1-2), 57-92.

Sumner, J. B., & Howell, S. F. (1936). Identification of Hemagglutinin of Jack Bean with Concanavalin A. *J Bacteriol, 32*(2), 227-237.

Sun, Z. L., Zhu, Y., Wang, F. Q., Chen, R., Peng, T., Fan, Z. N., . . . Miao, Y. (2007). Serum proteomic-based analysis of pancreatic carcinoma for the identification of potential cancer biomarkers. *Biochim Biophys Acta, 1774*(6), 764-771.

Swann, D. A., Silver, F. H., Slayter, H. S., Stafford, W., & Shore, E. (1985). The molecular structure and lubricating activity of lubricin isolated from bovine and human synovial fluids. *Biochem. J., 225*(1), 195-201.

Swann, D. A., Sotman, S., Dixon, M., & Brooks, C. (1977). The isolation and partial characterization of the major glycoprotein (LGP-I) from the articular lubricating fraction from bovine synovial fluid. *Biochem J, 161*(3), 473-485.

Switzer, R. C., 3rd, Merril, C. R., & Shifrin, S. (1979). A highly sensitive silver stain for detecting proteins and peptides in polyacrylamide gels. *Anal Biochem, 98*(1), 231-237.

Tabak, L. A. (2010). The role of mucin-type O-glycans in eukaryotic development. *Semin Cell Dev Biol, 21*(6), 616-621.

Tang, H., Hsueh, P., Kletter, D., Bern, M., & Haab, B. (2015). The detection and discovery of glycan motifs in biological samples using lectins and antibodies: new methods and opportunities. *Adv Cancer Res, 126*, 167-202.

Tei, K., Kawakami-Kimura, N., Taguchi, O., Kumamoto, K., Higashiyama, S., Taniguchi, N., . . . Kannagi, R. (2002). Roles of cell adhesion molecules in tumor angiogenesis induced by cotransplantation of cancer and endothelial cells to nude rats. *Cancer Res, 62*(21), 6289-6296.

Tempero, M. A., Uchida, E., Takasaki, H., Burnett, D. A., Steplewski, Z., & Pour, P. M. (1987). Relationship of carbohydrate antigen 19-9 and Lewis antigens in pancreatic cancer. *Cancer Res, 47*(20), 5501-5503.

Ten Hagen, K. G., Fritz, T. A., & Tabak, L. A. (2003). All in the family: the UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases. *Glycobiology, 13*(1), 1R-16R.

Thathiah, A., Blobel, C. P., & Carson, D. D. (2003). Tumor necrosis factor-alpha converting enzyme/ADAM 17 mediates MUC1 shedding. *J. Biol. Chem., 278*(5), 3386-3394.

Thathiah, A., & Carson, D. D. (2004). MT1-MMP mediates MUC1 shedding independent of TACE/ADAM17. *Biochem J, 382*(1), 363-373.

Tian, E., & Ten Hagen, K. G. (2009). Recent insights into the biological roles of mucin-type O-glycosylation. *Glycoconjugate J, 26*(3), 325-334.

Tollefsen, S. E., & Kornfeld, R. (1983a). The B4 lectin from Vicia villosa seeds interacts with N-acetylgalactosamine residues alpha-linked to serine or threonine residues in cell surface glycoproteins. *J Biol Chem, 258*(8), 5172-5176.

Tollefsen, S. E., & Kornfeld, R. (1983b). Isolation and characterization of lectins from Vicia villosa. Two distinct carbohydrate binding activities are present in seed extracts. *J Biol Chem, 258*(8), 5165-5171.

Tomana, M., Matousovic, K., Julian, B. A., Radl, J., Konecny, K., & Mestecky, J. (1997). Galactose-deficient IgA1 in sera of IgA nephropathy patients is present in complexes with IgG. *Kidney Int, 52*(2), 509-516.

Topaz, O., Shurman, D. L., Bergman, R., Indelman, M., Ratajczak, P., Mizrachi, M., . . . Sprecher, E. (2004). Mutations in GALNT3, encoding a protein involved in O-linked glycosylation, cause familial tumoral calcinosis. *Nat Genet, 36*(6), 579-581.

Torre, L. A., Bray, F., Siegel, R. L., Ferlay, J., Lortet-Tieulent, J., & Jemal, A. (2015). Global cancer statistics, 2012. *CA Cancer J Clin, 65*(2), 87-108.

Treuheit, M. J., Costello, C. E., & Halsall, H. B. (1992). Analysis of the five glycosylation sites of human alpha 1-acid glycoprotein. *Biochem J, 283 (Pt 1),* 105-112.

Triggianese, P., Chimenti, M. S., Toubi, E., Ballanti, E., Guarino, M. D., Perricone, C., & Perricone, R. (2015). The autoimmune side of hereditary angioedema: Insights on the pathogenesis. *Autoimmun Rev, 14*(8), 665-669.

Ugorski, M., & Laskowska, A. (2002). Sialyl Lewis(a): a tumor-associated carbohydrate antigen involved in adhesion and metastatic potential of cancer cells. *Acta Biochim Pol, 49*(2), 303-311.

Van Damme, E. J. M., Peumans, W. J., Barre, A., & Rouge, P. (1998). Plant lectins: A composite of several distinct families of structurally and evolutionary related proteins with diverse biological roles. *Crit Rev Plant Sci, 17*(6), 575-692.

Variyam, E. P., & Hoskins, L. C. (1981). Mucin degradation in human colon ecosystems. Degradation of hog gastric mucin by fecal extracts and fecal cultures. *Gastroenterology, 81*(4), 751-758.

Varki, A. (1992). Diversity in the Sialic Acids. *Glycobiology, 2*(1), 25-40.

Varki, A., & Crocker, P. R. (2009). I-type Lectins. In A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart & M. E. Etzler (Eds.), *Essentials of Glycobiology* (2nd ed.). Cold Spring Harbor (NY).

Varki, A., Etzler, M. E., Cummings, R. D., & Esko, J. D. (2009a). Discovery and Classification of Glycan-Binding Proteins. In A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart & M. E. Etzler (Eds.), *Essentials of Glycobiology* (2nd ed.). Cold Spring Harbor (NY).

Varki, A., Kannagi, R., & Toole, B. P. (2009b). Glycosylation Changes in Cancer. In A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart & M. E. Etzler (Eds.), *Essentials of Glycobiology* (2nd ed.). Cold Spring Harbor (NY).

Varki, A., & Sharon, N. (2009). Historical background and overview. In A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart & M. E. Etzler (Eds.), *Essentials of Glycobiology* (2nd ed.). NY: Cold Spring Harbor Laboratory Press.

Winzler, R. J., Devor, A. W., Mehl, J. W., & Smyth, I. M. (1948). Studies on the Mucoproteins of Human Plasma. I. Determination and Isolation. *J Clin Invest, 27*(5), 609-616.

Winzler, R. J., & Smyth, I. M. (1948). Studies on the mucoproteins of human plasma; plasma mucoprotein levels in cancer patients. *J Clin Invest, 27*(5), 617-619.

Wu, A. M. (2001). Expression of binding properties of Gal/GalNAc reactive lectins by mammalian glycotopes (an updated report). *Adv Exp Med Biol, 491*, 55-64.

Xie, F., Liu, T., Qian, W. J., Petyuk, V. A., & Smith, R. D. (2011). Liquid chromatography-mass spectrometry-based quantitative proteomics. *J Biol Chem, 286*(29), 25443-25449.

Xin, A. J., Cheng, L., Diao, H., Wu, Y. C., Zhou, S. M., Shi, C. G., . . . Zhang, Y. L. (2016). Lectin binding of human sperm associates with DEFB126 mutation and serves as a potential biomarker for subfertility. *Sci Rep, 6*.

Yu, Y. H., Liang, C., & Yuan, X. Z. (2010). Diagnostic value of vacuum-assisted breast biopsy for breast carcinoma: a meta-analysis and systematic review. *Breast Cancer Res Treat, 120*(2), 469-479.

Zauner, G., Kozak, R. P., Gardner, R. A., Fernandes, D. L., Deelder, A. M., & Wuhrer, M. (2012). Protein O-glycosylation analysis. *Biol Chem, 393*(8), 687-708.

Zhang, H., Li, X. J., Martin, D. B., & Aebersold, R. (2003). Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotechnol, 21*(6), 660-666.

Zubarev, R. A., Horn, D. M., Fridriksson, E. K., Kelleher, N. L., Kruger, N. A., Lewis, M. A., . . . McLafferty, F. W. (2000). Electron capture dissociation for structural characterization of multiply charged protein cations. *Anal Chem, 72*(3), 563-573.

LIST OF PUBLICATIONS AND PAPERS PRESENTED

Publications:

1. Lee, C. S., Muthusamy, A., Abdul-Rahman, P. S., Bhavananthan, V. and Hashim, O. H. (2013). An improved lectin-based method for the detection of mucin-type O-glycans in biological samples. Analyst, 138, 3522-3529.

2. Lee, C. S., Mohd-Taib, N. A., Ashrafzadeh, A., Fadzli, F., Harun, F., Rahmat, K., Hoong, S. M., Abdul-Rahman, P. S. and Hashim, O. H. (2016). Unmasking heavily O-glycosylated serum proteins using perchloric acid: identification of serum proteoglycan 4 and protease C1 inhibitor as molecular indicators for screening of breast cancer. PLoS ONE, 11(2): e0149551.

Conference:

1. Lee C. S., Abdul-Rahman P. S., Mohd-Taib N. A., and Hashim O. H. (2014). Sandwich Enzyme-Linked Lectin Assay: A simple method for detection of total O-glycosylated proteins in biological samples. Poster presentation at the 15th International Conference on System Biology. Melbourne Convention and Exhibition Centre, Melbourne. 14th – 18th of September 2014.

Publication 1

Analyst

RSC Publishing

PAPER

View Article Online
View Journal | View Issue

An improved lectin-based method for the detection of mucin-type *O*-glycans in biological samples

Cite this: *Analyst*, 2013, **138**, 3522

Cheng-Siang Lee,[a] Arivalagan Muthusamy,[b] Puteri Shafinaz Abdul-Rahman,[ac] Veer P. Bhavanandan[a] and Onn Haji Hashim*[ac]

Mucins and mucin-type glycoproteins, collectively referred to as mucin-type *O*-glycans, are implicated in many important biological functions and pathological conditions, including malignancy. Presently, there is no reliable method to measure the total mucin-type *O*-glycans of a sample, which may contain one or more of these macromolecules of unknown structures. We report the development of an improved microassay that is based on the binding of lectins to the unique and constant GalNAc-Ser/Thr structural feature of mucin-type *O*-glycans. Since the sugar–amino acid linkage in the mucin-type *O*-glycans is invariably cryptic, we first chemically removed the heterogeneous peripheral and core saccharides of model glycoconjugates before examining for their interactions using an enzyme-linked lectin assay (ELLA). Desialylation of the model glycoconjugates led to maximal binding of the lectins but additional treatments such as Smith degradation did not result in increased binding. Of the lectins tested for their ability to probe the desialylated *O*-glycans, jacalin showed the highest sensitivity followed by champedak galactose binding (CGB) lectin and *Vicia villosa* agglutinin. Further improvement in the sensitivity of ELLA was achieved by using microtiter plates that were pre-coated with the CGB lectin, which increased the specificity of the assay to mucin-type *O*-glycans. Finally, the applicability of the developed sandwich ELLA to crude samples was demonstrated by estimating trace quantities of the mucin-type *O*-glycans in the human serum.

Received 6th September 2012
Accepted 21st April 2013

DOI: 10.1039/c3an36258b

www.rsc.org/analyst

Introduction

Mucins are large glycoproteins containing many clusters of glycosylated serines and threonines in tandem repeat regions.[1–3] Mucins and mucin-type glycoproteins, referred to as mucin-type *O*-glycans†, are implicated in many important biological functions and pathological conditions including malignancy.[4,5] In cancer, the over-expression and aberrant glycosylation of cell surface glycoconjugates are among the phenotypic changes that are widely recognized.[6] The tumor cells of cancer patients actively secrete and/or shed, during cell turnover, a repertoire of *O*-glycans into the circulation.[7,8] Hence, specific detection and quantitation of circulating tumor-associated *O*-glycans will be extremely beneficial for the early diagnosis and treatment of cancer. In fact, immunological methods are available for detection of cancer-associated *O*-glycans such as MUC1, CA125 and CA19-9.[9–11] However, these assays based on monoclonal antibody recognition of one specific epitope are of very narrow specificities and therefore not suitable for detection of the total *O*-glycan content of samples such as serum, saliva, urine or tissue extracts.

A problem in developing a single assay for *O*-glycans is the heterogeneity of their saccharides.[12] However, in contrast to the high variability of the peripheral saccharides, which are based on eight different core structures, the sugar–amino acid linkage in *O*-glycans is a unique and constant feature. Previously, one of us (VPB) had developed an assay for the estimation of *O*-glycans that was based on the specific chemistry of the GalNAc-Ser/Thr linkage.[13] Since the sensitivity of this colorimetric assay is very limited, in this study we explored the use of lectins to estimate trace amounts of *O*-glycans.

The aim of the present study was to develop a sensitive assay for the determination of the *O*-glycan content of a sample. This

[a]Department of Molecular Medicine, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia. E-mail: onnhashim@um.edu.my; Fax: 603 7967 4957
[b]Department of Orthopedics and Rehabilitation, Penn State University College of Medicine, Hershey, PA 17033, USA
[c]University of Malaya Centre for Proteomics Research, University of Malaya, 50603 Kuala Lumpur, Malaysia

† The nomenclature in this field is very confusing and keeps changing as new members of the glycoconjugates family are discovered. The term, *O*-glycan or *O*-linked glycoprotein, has been traditionally used to refer to glycoconjugates in which the saccharides are attached to the protein via an *O*-glycosidic linkage. The major class of *O*-glycans consists of those having saccharides linked via the reducing end of GalNAc to the hydroxyl group of serine and threonine. However, there are also glycoconjugates having saccharides, including GlcNAc-, mannose-, fucose-linked *O*-glycosidically to proteins. Therefore, the term "mucin-type *O*-glycan" has to be used to distinguish glycoconjugates having GalNAcα1-Ser/Thr linkage from other *O*-glycans. In this article the term *O*-glycans is used synonymously with mucin-type *O*-glycans.

Publication 2

PLOS | ONE

RESEARCH ARTICLE

Unmasking Heavily *O*-Glycosylated Serum Proteins Using Perchloric Acid: Identification of Serum Proteoglycan 4 and Protease C1 Inhibitor as Molecular Indicators for Screening of Breast Cancer

Cheng-Siang Lee[1], Nur Aishah Mohd Taib[2], Ali Ashrafzadeh[3], Farhana Fadzli[2], Faizah Harun[2], Kartini Rahmat[4], See Mee Hoong[2], Puteri Shafinaz Abdul-Rahman[1,5], Onn Haji Hashim[1,5]*

1 Department of Molecular Medicine, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia, 2 Department of Surgery, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia, 3 Medical Biotechnology Laboratory, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia, 4 Department of Biomedical Imaging, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia, 5 University of Malaya Centre for Proteomics Research, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia

* onnhashim@um.edu.my

OPEN ACCESS

Citation: Lee C-S, Taib NAM, Ashrafzadeh A, Fadzli F, Harun F, Rahmat K, et al. (2016) Unmasking Heavily *O*-Glycosylated Serum Proteins Using Perchloric Acid: Identification of Serum Proteoglycan 4 and Protease C1 Inhibitor as Molecular Indicators for Screening of Breast Cancer. PLoS ONE 11(2): e0149551. doi:10.1371/journal.pone.0149551

Editor: Roger Chammas, Universidade de São Paulo, BRAZIL

Received: December 3, 2015

Accepted: February 2, 2016

Published: February 18, 2016

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was funded by the HIR-MOHE H-20001-00-E000009 research grant from the University of Malaya. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Heavily glycosylated mucin glycopeptides such as CA 27.29 and CA 15–3 are currently being used as biomarkers for detection and monitoring of breast cancer. However, they are not well detected at the early stages of the cancer. In the present study, perchloric acid (PCA) was used to enhance detection of mucin-type *O*-glycosylated proteins in the serum in an attempt to identify new biomarkers for early stage breast cancer. Sensitivity and specificity of an earlier developed sandwich enzyme-linked lectin assay were significantly improved with the use of serum PCA isolates. When a pilot case-control study was performed using the serum PCA isolates of normal participants (n = 105) and patients with stage 0 (n = 31) and stage I (n = 48) breast cancer, higher levels of total *O*-glycosylated proteins in sera of both groups of early stage breast cancer patients compared to the normal control women were demonstrated. Further analysis by gel-based proteomics detected significant inverse altered abundance of proteoglycan 4 and plasma protease C1 inhibitor in both the early stages of breast cancer patients compared to the controls. Our data suggests that the ratio of serum proteoglycan 4 to protease C1 inhibitor may be used for screening of early breast cancer although this requires further validation in clinically representative populations.