PROTEIN AND GLYCAN PROFILING OF NATURAL RUBBER LATEX FROM MALAYSIAN HEVEA CLONE, RRIM 2025

NUR SYAFIQAH BINTI MOHD AFANDI

UNIVERSITI SAINS MALAYSIA

2024

PROTEIN AND GLYCAN PROFILING OF NATURAL RUBBER LATEX FROM MALAYSIAN HEVEA CLONE, RRIM 2025

by

NUR SYAFIQAH BINTI MOHD AFANDI

Thesis submitted in fulfillment of the requirements
for the degree of
Master of Science

October 2024

ACKNOWLEDGEMENT

بِسْـــــــــــــمِ اللهِ الرَّحْمَنِ الرَّحِيْمِ

First and foremost, I would like to express appreciation towards my supervisor, Assoc. Prof. Dr Mohd Nazri Ismail, for his utmost guidance, and patience. I would like to thank the hands-on learning on mass spectrometry instruments, academics and industrial exposures, and opportunities that were given to me by Dr Nazri during my MSc study. These generous opportunities opened up so many doors to my future endeavours.

Secondly, I am grateful towards the financial support that I've received from the Fundamental Research Grant Scheme FRGS/1/2018/STG04/ USM/02/1 under the Ministry of Higher Education (MoHE) and Graduate Excellence Programme (GrEP) Majlis Amanah Rakyat (MARA) 2022. Special thanks to Encik Azmin Hassan, the owner of the small rubber plantation, for contributing samples and helping identify RRIM 2025 clones.

My sifu senior, Mohd Afiq Hazlami Habib, has given me relentless guidance and diluted down the bioinformatics and alienated computer language terms for me, who has no idea how challenging it is to learn everything from scratch. This experience was life-changing, along with Dasi Ong Yaqut Ong, the infamous laboratory dweller and a zoologist lady, who helped me immensely with my lab work. My InterVenn interns, Nurul Hakimah Abd Salim, Datie Md. Amin, and Gayathri A/P Kumar. These people always provide me emotional support with various crash courses, making my learning journey fun and satiable. Along the journey, I came to love STEM unconditionally. Thank you, my Pakistani friends, Rozina Kakar, Rabia Tasaduq,

Munaza Api, our only Iraqi, Huda, the only Indonesian in the group, Putrie, and then, Shakirah for always making a gathering showered with many comfort foods. To my housemates, Fatimah, Siti, Wani, Farehah, and Athirah, Aisyah, thank you for making my life in Penang easier and entertaining.

To my mentors, those who I consider my lifelong teachers, and startup partners, Dr. Roslan Abd. Jelani, Encik Md. Asham Ahmad, Ahmad Haikal Ahmad Mochtar, Ahmad Afif Ahmad Mohtari Yasim, Tengku Nur Nabilah Putri Tengku Alam Shah, Nurshafira Noh, Siti Hawa Mohd, your pearls of wisdom and visions have shaped who I am today. To my postgrads fellas, Mursyida, Saadah, Anis Muneerah, kak Farah Alwani, Ching JJ, and my confidantes that live miles away, Amni Azharan, Najihah Zalami, Syifaa Hassan, Ilyana Shukri, Amirah Mardhiah.

To my family, my parents, Mohd Afandi Ismail and Sanisah Yaacob, and in-laws thank you for bearing with me due to so many uncertainties in my life. Without their blessings and prayers, I would not be able to muster up my courage to finish this study. To my husband, Muhammad Nazirul Naim Zaharuddin, thank you for showing me great virtues as examples, your unwavering support, unconditional love, and the consistency that you have shown me from day 1.

NUR SYAFIQAH BINTI MOHD AFANDI
Desasiswa Bakti H09, 113
13 December, 2023

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF SYMBOLS AND ABBREVIATIONS

%	Percent
℃	Degree celcius
ABrC	Analytical Biochemistry Research Centre
ACBP1	Acyl-CoA-binding domain-containing protein 1
ACN	Acetonitrile
Asn	Asparagine
ATP	Adenosine triphosphate
BSA	Bovine serum albumin
CCDs	Cross-reactive carbohydrate determinants
CIB	Cornell Institute of Biotechnology
CID	Collision-induced dissociation
CPT	Cis-prenyltransferase
Cys	Two cysteines
Da	Dalton
DDA	Data-dependent acquisition
dH2O	Deionised water
DHB	Dihydroxybenzoic acid
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
ER	Endoplasmic reticulum
ESI	Electrospray ionisation
ETD	Electron Transfer Dissociation
FA	Ford Acre
FA	Formic acid

FDR	False Discovery Rate
FELDA	Federal Land Development Authority
FRGS	Fundamental Research Grant Scheme
Fuc	Fucose
g	Gram
Gal	Galactose
GalNAc	N-Acetylgalactosamine
GlcNAc	β1,2-linked N-acetylglucosamine
GO	Gene ontology
GrEP	Graduate Excellence Programme
GT	Godang Tapen
h	Hour
HCD	High Collision Dissociation
HILIC	Hydrophilic interaction chromatography
HMGR	HMG-CoA reductase
HMGS	Hydroxymethylglutaryl-CoA synthase
IAA	Iodoacetamide
IAN	Instituto Agronômico do Norte
IEF	Isoelectric focusing
IgE	Immunoglobulin E
kDa	Kilodalton
KEGG	Kyoto Encyclopedia of Genes and Genomes
KH2PO4	Potassium dihydrogen phosphate
LC-MS/MS	Liquid chromatography-tandem mass spectrometry
LTC	Latex timber clone
LTQ	Linear trap quadrupole
M	Molarity
m/z	Mass/charge

MALDI	Matrix-assisted laser desorption/ionisation
MALDI-TOF/TOF	MALDI time-of-flight mass spectrometry
MARA	Majlis Amanah Rakyat
MEP	Methylerythritol
mg/ml.	Miligram/mililitre
min	Minute
mM	Milimolar
MoHE	Ministry of Higher Education
MRB	Malaysian Rubber Board
MS/MS	Tandem mass spectrometry
MVA	Mevalonate
NaOH	Sodium hydroxide
NCBI	National Center for Biotechnology Information
Neu5Ac/NANA	N-acetylneuraminic acid
NH4AcO	Ammonium acetate
nm	Nanometre
nr	Non-redundant protein sequences
NRL	Natural rubber latex
PB	Prang Besar
PBS	Phosphate buffer saline
perMe	Permethylation
PMF	Peptide mass fingerprints
PPAR	Peroxisome proliferator-activated receptors
ppm	Parts per Million
PR	Pathogenesis-related
Pro	Proline
Proteome Hb	*Hevea brasiliensis* protein database
PSM	Peptide-spectrum matches
PTMs	Post-translational modifications

R&D	Research and Development
REF	Rubber elongation factor
RNA	Ribonucleic acid
RP-LC	Reverse-phase liquid chromatography
RP-SPE	Reversed-phase of solid-phase
RRII	Rubber Research Institute of India
RRIM	Rubber Research Institute of Malaysia
SAX	Strong anion exchange
SCX	Strong cation exchange
SDS-PAGE	Sodium dodecyl-sulfate polyacrylamide gel
Ser	Serionine
SH-	Sulfhydryl group
SRPP	Small rubber particle
SSS	Sequence Similarity Search
TFA	Trifluoroacetic acid
Thr	Threonine
TiO_2	Titanium dioxide
TIP	Tonoplast intrinsic protein
Tjir	Tjirandji
TOF	Time of flight
uL	Microlitre
USA	United States
USM	Universiti Sains Malaysia
v/v	Volume/volume
W/cm2	Watts per square centimeter
w/v	Weight/volume
w/w dwt	Weight by weight of deadweight tonnage
αCHCA	α-Cyano-4-hydroxycinnamic acid

LIST OF APPENDICES

PEMPROFILAN PROTEIN DAN GLIKAN LATEKS GETAH SEMULAJADI DARIPADA KLON HEVEA MALAYSIA, RRIM 2025

ABSTRAK

Hevea brasiliensis merupakan pokok getah asli berasal dari Amerika Selatan yang majoritinya ditanam di Asia Tenggara, termasuk Malaysia, pengeksport keenam terbesar dunia. Spesis ini ditanam untuk lateks dan kayunya untuk sektor kepenggunaan seperti tayar, sarung tangan, dan perabot, yang menyumbang kepada pendapatan eksport Malaysia sebanyak RM17.8 bilion pada 2019. *H. brasiliensis* merupakan tumbuhan penghasil getah paling produktif dengan kandungan getah tertinggi (%) dengan 15-30 (w/w dwt) berbanding jenis tumbuhan alternatif lain seperti guayule, dan dandelion. Lateks getah semulajadi (LGS) *H. brasiliensis* terdiri dari 5 % spesis non-isoprena yang mengandungi banyak protein unik, dan dikaji secara meluas kerana kealergenikannya. Fungsi protein sangat ditentukan oleh modifikasi pasca-translasi (PTM) di mana salah satu PTM paling signifikan di dalam LGS adalah glikosilasi. Glikosilasi menambahbaik keterlarutan, pelipatan, kestabilan, perhimpunan, dan pengangkutan intrasel bagi peptida nasen. Walaupun diketahui kepentingannya dalam proses-proses biologi, aspek kualitatif protein dan glikan LGS kurang diketahui kerana kekompleksan makro- dan mikroheterogenitinya. Jadi, bagi memahami aspek proteom dan glikom dalam LGS *H. brasiliensis,* analisis-analisis jisim spektrometri telah dilakukan bagi klon lateks *Hevea* RRIM 2025 yang dikenali dengan hasil getah yang banyak dan lilitan besar. Tiga kaedah pengayaan telah digunakan untuk meningkatkan kebolehkesanan glikopeptida dinamakan *Strong anion*

exchange (SAX), Hydrophilic interaction chromatography (HILIC), and *Titanium dioxide (TiO_2),* diikuti dengan penjujukan menggunakan LTQ-Orbitrap Velos Pro LC-MS/MS. Data mentah telah dianotasi pada PEAKS7.5 dengan satu set parameter glikan untuk PTM dan hasil carian menunjukkan 29 protein telah dipadankan kepada Proteome Hb. Tujuh daripada 15 alergen telah dikenalpasti: Hev b 1, 2, 3, 5, 7, 14 dan 15. Analisis ontologi gen dan laluan KEGG telah dilakukan bagi memahami aspek fungsi protein di mana hasil menunjukkan kebanyakan protein-protein terlibat di dalam komponen-komponen biologi dan sel. Hanya 27.6 %, 14.8 %, 7.7 %, dan 0 % telah teranotasi di dalam KEGG bagi sampel kawalan, SAX, HILIC, dan TiO_2. Terdapat tiga tempat glikosilasi yang telah dikenalpasti pada patatin; Asn-284, 351, dan 352, manakala *rubber elongation factor (REF)* pada Asn-7. Manakala, sampel-sampel terhadam tripsin telah dipisahkan menggunakan PNGase A untuk pemprofilan glikomik menggunakan MALDI-TOF/TOF. Spektra kemudiannya disemi-anotasikan menggunakan GlycoWorkBench 2.0. Glikan yang paling banyak adalah Hex3HexNAc2Pen1dHex1, *m/z* 1505.7 (98 %), Hex5HexNAc4Pen1dHex3, *m/z* 2752.3 (75 %), and Hex3HexNAc4Pen1dHex1, *m/z* 1996.0 (55 %). Nisbah keamatan ion glikan bagi RRIM 2025 menunjukkan jenis terpangkas, diikuti dengan hibrid, dan manosa tinggi, merupakan jenis glikan terbanyak. Kesimpulannya, kaedah-kaedah pengayaan telah membantu untuk mencirikan protein dengan penemuan glikoprotein tambahan secara komprehensif, dan penemuan struktur-struktur glikan yang pertama kali direkodkan dalam getah lateks RRIM 2025.

PROTEIN AND GLYCAN PROFILING OF NATURAL RUBBER LATEX FROM MALAYSIAN HEVEA CLONE, RRIM 2025

ABSTRACT

Hevea brasiliensis is an original rubber crop from South America that is mostly cultivated in Southeast Asia, including Malaysia, the sixth largest producer in the world. This species is grown for its rubber latex and wood for consumer sectors such as tyres, gloves, and furniture, contributing to Malaysia's export earnings of RM17.8 billion in 2019. *H. brasiliensis* is the most productive rubber-producing crop with the highest rubber content (%) with 15-30 (w/w dwt) compared to the other alternative types, such as guayule, and dandelion. Natural rubber latex (NRL) of *H. brasiliensis* is made of 5% non-isoprene species, comprising many unique proteins widely investigated on their allergenicity. Protein function is heavily determined by its post-translational modifications (PTMs), with one of the most significant PTMs being glycosylation. Despite its importance in biological processes, NRL's qualitative aspect of protein, especially glycan, is poorly known because of its macro- and microheterogeneity complexes. Hence, to explore the proteome and glycoproteome of the NRL of *H. brasiliensis*, mass spectrometry analyses were conducted on the *Hevea* clone RRIM 2025 which is known for high rubber yield and large girth. Three enrichment methods were used to increase the detectability of glycopeptides, namely Strong anion exchange (SAX), Hydrophilic interaction chromatography (HILIC), and Titanium dioxide (TiO_2), prior to LTQ-Orbitrap Velos Pro LC-MS/MS sequencing. Raw data were annotated on PEAKS7.5 with a set of glycan parameters for PTMs, and

the search result showed 29 total proteins matched to Proteome Hb. Seven out of 15 allergens were identified: Hev b 1, 2, 3, 5, 7, 14, and 15. Gene ontology and KEGG pathway analyses were conducted to determine the protein functions, most of which were involved in biological and cellular components. Only 27.6%, 14.8%, 7.7%, and 0% were annotated in KEGG for control, SAX, HILIC, and TiO_2. Three glycosylation sites were identified on patatin: Asn-284, 351, and 352, while rubber elongation factor (REF) at Asn-7. Meanwhile, the tryptic-digested samples were cleaved with PNGase A for glycomic profiling using MALDI-TOF/TOF. The spectra were then semi-annotated using GlycoWorkBench2.0. The most abundant glycans were Hex3HexNAc2Pen1dHex1, *m/z* 1505.7 (98 %), Hex5HexNAc4Pen1dHex3, *m/z* 2752.3 (75%), and Hex3HexNAc4Pen1dHex1, *m/z* 1996.0 (55%). The ion intensities showed that truncated, hybrid, and high mannose were the most abundant glycans. To conclude, the enrichment methods helped characterise the proteins comprehensively with additional glycoprotein findings and the first discovered glycan structures recorded in the rubber latex of RRIM 2025.

CHAPTER 1

INTRODUCTION

1.1 Background of the study

Hevea brasiliensis is an original rubber crop from Bolivia, Brazil, Colombia, Venezuela, and Peru, which is essential in the latex and rubber industry (Witt & Luke, 2017). At the end of the 19th century, most of the natural rubber latex (NRL) was exported from Southeast Asia, which covers almost 95 % of the world market (Balsiger, 2000; Lemos, 2023). According to Global Rubber Latex Thread Market 2018, in 2014, Malaysia was the largest contributor to the latex industry, making up 35 % of the world's supply (360 Market Updates, 2018). This species is mainly cultivated for its latex and its by-product, rubber wood.

According to the Rubber Research Institute of Nigeria, Malaysia has at least 10 clones of *Hevea brasiliensis,* known by their initials, RRIM- and PB- (Omokhafe et al., 2017), with RRIM 600 as the most established Malaysian clone. Each of the clones has a different rubber yield and stability and has been extensively studied for its latex composition, latex gene expression, and metabolic pathway of rubber biosynthesis. Although there is a stern demand for latex production, the quantitative aspect of rubber latex was not comprehensively understood, especially the critical post-translational modifications (PTMs) of protein. The proteomic analysis is integral in elucidating the dynamism of a protein structure relative to its function. Recently, PTMs have garnered massive attention in plant biology due to the growing evidence of their roles in plant-pathogen relationships (Liu et al., 2021).

Hevea brasiliensis is the most productive crop in rubber production since *H. brasiliensis* has the highest rubber content (%), 15-30 (w/w dwt) compared to the other rubber-producing plants such as guayule, *Parthenium argentatum*, dandelion, *Taraxacum* kok-saghyz, and the Ficus family (Cherian et al., 2019). NRL is a superior elastomer that has been widely produced by consumers. However, NRL has caused a great deal of significant immunoglobulin E (IgE)-mediated hypersensitivity in health workers (Parisi et al., 2021). The allergenicity is often associated with the plant's defense against biotic and abiotic threats, as latex itself is part of the mechanism (Wangorsch et al., 2022). The presence of pathogen-related (PR) proteins such as b-1-3 glucanases and chitinase might be responsible for triggering allergic reactions with some conjugate of N-linked glycosylation. The role of allergenic protein is frequently attributed to the presence of glycan epitope since the structural characteristics are stipulated to mediate the immune response. However, little is known regarding the relationship between proteins and glycan epitopes and their contribution to allergenic effects.

Glycoproteins are abundant in nature, however, their qualitative aspect is poorly known due to their macro- and microheterogeneity complexes. The low quantity of glycoproteins compared to the proteins and some analytical challenges, such as poor ionisation during the MS/MS analysis and yield loss during multiple stages of sample preparations, add more arduousness in retrieving glycopeptides. Most of the glycoprotein analyses in *Hevea* latex were conducted using protein separation techniques such as sodium dodecyl-sulfate polyacrylamide gel (SDS-PAGE) and western blotting (Magnelli et al., 2012; Tokhtaeva et al., 2017). However, these methods are only suitable for limited protein slots.

The nature of glycoproteins and analytical challenges can be handled through sample preparations and advanced instrumentation. For instance, enrichment methods paired with tandem mass spectrometry (MS/MS) can quantify low quantities of glycan moieties together with their peptide sequences. The MS/MS paired with bioinformatic tools will enable the structural and site elucidation of the glycoproteins. Examples of commonly used tandem MS are Liquid chromatography-tandem mass spectrometry (LC-MS/MS) and Matrix-assisted laser desorption/ionisation time-of-flight/time-of-flight mass spectrometry (MALDI-TOF/TOF MS) that enable the batch identification of proteome and glycome.

1.2 Problem statement

Hevea brasiliensis is widely studied for its biopolymer properties but lacks attention to its biochemistry nature, especially the proteins and their post-translational modifications (PTMs). Most of the proteomic study was limited to the few proteins that are already abundant in the rubber latex of *H. brasiliensis*, for example, rubber elongation factor (REF) and small rubber particle (SRPP) (Wadeesirisak et al., 2023). Instead, the only discovery study for comprehensive protein profiling in *H. brasiliensis* was done by Habib et al. in 2017 on the RRIM 600 clone, the already established clone. However, the Habib et al. studies did not report any glycan structures together with the glycosylation results. The lack of glycan structure information makes it more challenging to understand the rubber latex blueprints.

In *H. brasiliensis*, only a few of the 15 internationally recognised allergens are glycosylated (Berthelot et al., 2014). The glycosylated allergens were as fragmented glycan isoforms, as seen in isoform I of Hev b 2 based on the published data on GenBank (accession code ABN09655.1), where it lacks mannose (Rodríguez-Romero

et al., 2014) The published major latex allergen, Hev b 4 on UniProt (Q6T4P0_HEVBR), also still needs a review on its annotation score. Although there has been quite a progression in the characterisation of *H. brasiliensis* allergenic glycoproteins, there is the possibility of the undiscovered domain of other isoforms and glycoproteins from non-allergens. Many of the glycoprotein structures are not yet being published in the public domain and lack verification and substantial characterisations. For an instant, only 29 glycoproteins were associated with *H. brasiliensis* on the GlyCosmos Glycoproteins database (accessed date: 19/7/2024, https://glycosmos.org/search?keyword=hevea+brasiliensis&commit=Search) with mostly reported from non-Malaysian cultivars.

As early as 2003, the RRIM 2000 series has been recommended for cultivations. However, only almost a decade later, the Malaysian Rubber Board (MRB) made a specific distinction for RRIM 2025 and RRIM 2023 to be widely cultivated in Malaysian Rubber Budwood Centres (MRBC), stationed in one of the north provinces of Peninsular Malaysia (Jalani & Ramli, 2003; Syed Rashid Ali, 2011). As RRIM 2025 is easily accessible in the rubber plantation near Penang compared to its sister clone, RRIM 2025 was chosen due to its availability and for the other agronomic values such as high latex yield and large girth (Mayati & Hadafi, 2021). Furthermore, researchers mainly mention this clone for its wood superiority, while its proteome natural rubber latex remains unknown. All previously identified issues that conclude lack of characterisation have led to limited reliable and verified data from proteome and glycome databases. The absence of reliable data sources will hinder the bridging of translational with functional works.

It can be hypothesised that, with the recent transcriptome study on RRIM 2025 (Kamerkar & Othman, 2022), the batch identification of protein through

comprehensive profiling and potentially new sites of N-glycosylation will provide more insights for knowledge mapping. The insights obtained from this study are expected to be utilised to improve the in-silico database, i.e., verifying the current *H. brasiliensis* proteomic database, and algorithms of glycosylation predictor tools. As NRL yield and disease tolerance are two of the many important factors for industrial propagation, the protein list may facilitate the biomarker identification to produce new high-quality *H. brasiliensis* clones. The new clones are critical to keep up with the increasing demand and climate change.

1.3 Objectives

The main challenge in this study is the complex nature of glycoproteins and glycans. Meanwhile, the second issue is the lack of discovery studies for proteome and glycome identifications. Hence, three objectives are sought for this study:

1. To evaluate and compare three glycopeptide enrichment methods: Hydrophilic interaction chromatography (HILIC), Strong anion exchange (SAX), and Titanium dioxide (TiO_2).
2. To perform the characterisation of proteins and glycoproteins from Malaysian *H. brasiliensis* (RRIM 2025) rubber latex using LC-MS/MS.
3. To profile the glycan from Malaysian *H. brasiliensis* (RRIM 2025) rubber latex using MALDI-MS and bioinformatics.

CHAPTER 2

LITERATURE REVIEW

2.1 *Hevea brasiliensis*

Hevea brasiliensis is a woody perennial crop member of the dioecious Euphorbiaceae family (Figure 2.1). Most euphorbs produce milky latex, which has whorled leaves and inflorescences containing a unisexual or/ bisexual reproductive system (Tomlinson, 2016). The *H. brasiliensis* latex carries global importance, and recently, for its timbers and power generation from logging residues (Hytönen et al., 2019). Southeast Asia is the home to latex commercialisation (Malaysia, Thailand, Indonesia, and Vietnam) which accounts for up to 90 % of global production (Sattayawaksakul & Choi, 2017). In Malaysia, it is heavily cultivated in the Southern (Kota Tinggi) and Central regions (Sungai Buloh), with two Malaysian Rubber Board (MRB) research centres stationed in the respective areas (Ali et al., 2020). Rubber plantation in Malaysia is the second-largest land occupation, with an estimated acreage of 1,137,14 ha in 2021(Malaysian Rubber Board, 2022).

Many crops were identified to produce natural rubber latex, including eight botanical families, 300 genera, and 1,800 species. Unfortunately, not many were found to produce higher molecular weight compared to the *Hevea* crop (Bushman et al., 2006). Among the nine genera of *Hevea*, *H. brasiliensis* yields latex the most (Carron et al., 1989). Besides its NRL as a consumer product, it also possesses biomedical uses in angiogenesis, wound healing, cell adhesion, and tissue formation. It is involved in

drug delivery systems to synthesise prostheses, sensors, and nanoparticles (Azevedo Borges et al., 2014).

Figure 2.1 Rubber tree (*Hevea brasiliensis)* in Felda Lubuk Merbau, Kedah

2.1.1 Natural rubber latex (NRL) of *H. brasiliensis*

Natural rubber latex (NRL) is a biopolymer considered an agricultural commodity (Budiman, 2003). It is mainly made of poly(cis-1,4-isoprene) coupled with another biological compound (Vaysse et al., 2012). This long-chain isoprene is found in the shell core of NRL particles, coated with phospholipids and proteins layer with a typical thickness of approximately 10-20 nm (Liao, 2017). Natural rubber biosynthesis in *H. brasiliensis* is produced through two pathways: methylerythritol (MEP) and mevalonate (MVA) pathway, where they were diligently studied to improve the latex yield (Cherian et al., 2019). In rubber biosynthesis pathways, rubber transferase catalyses the production of NRL through three stages: initiation, polymerisation, and termination (Men et al., 2019a). The latex of *H. brasiliensis* is

made up of three fractions: rubber, C-serum, and lutoid. Rubber particles are the most abundant in latex, with 26-51 %, followed by C-serum, 15-46 %, and lutoids, 28-43 % (Bottier, 2020a).

2.1.2 Clones of *Hevea brasiliensis*

Many types of clones were mainly named according to the place of origin or organisation. For instance, RRIM: Rubber Research Institute of Malaysia (Malaysia), PB: Prang Besar (Malaya), FA: Ford Acre (Brazil), IAN: Instituto Agronômico do Norte (Brazil), RRII: Rubber Research Institute of India (India), GT: Godang Tapen (Indonesia), Tjir: Tjirandji (Indonesia) (Gonçalves et al., 2003). According to the MRB, clones are classified into two groups: Group I (RRIM 928, RRIM 2001, RRIM 2002, PB 260, PB 350) and group II (RRIM 2007, RRIM 2023, RRIM 2024, RRIM 3001, PB 347, PB 373 and PB 374). Currently, RRIM 2025 (Fisol et al., 2021; Mohd Zin et al., 2021; Wan Mohd Zain et al., 2021) is attentively being studied due to its availability in many Malaysian plantations. However, there is yet a statistic about the availability of *H. brasiliensis* clones in rubber plantations, especially among the small rubber stakeholders in Malaysia, since small farmers contribute 93 % of the natural rubber collection (Hazir et al., 2020). A census of the type of clones cultivated on each plantation is essential to understanding the trend of clones' cultivation for strategic rubber planning. The RRIM 2025 clone of *H. brasiliensis* can be recognised from its physiological traits in leaf shape, stripes, and arrangement of leaves (Figure 2.2). Their leaflets have a huge gap between each other, cuneate base, and have a longer oval shape (obovate) and accuminate tip compared to PB 350 (Anjomshoae & Rahim, 2018). RRIM 2025 is the offspring of RRIM 803 and IAN 8073 and is specified as a latex timber clone (LTC) (Naji et al., 2014).

Figure 2.2 *Hevea brasiliensis* and leaflets of RRIM 2025

Approximately 63 % to 100 % of the proteins from the 27 biosynthetic pathways were effectively identified in RRIM 600. Additionally, the entire translated draft genome database (AnnHBM) yielded 1839 and 28 distinct proteins from the in-house annotation (Habib et al., 2017).

2.1.3 Protein and carbohydrate contents of *H. brasiliensis* NRL

The largest protein population in latex is the rubber elongation factor (REF), making up 10-60 % (Wang et al., 2010). Among the most abundant proteins are small rubber particles (SRPP), 1,3-beta-glucanase, hevamine, elicitor-responsive protein, and type II metacaspase (Abd-Rahman & Kamarrudin, 2018). A total carbohydrate study showed that the most abundant sugar in latex was sucrose and quebrachitol, followed by fructose, mannose, galactose, and sorbitol (Bellacicco et al., 2018). Minor sugar might possess the ability to discriminate a clone from its sibling clones for an instant through the significantly distinct amount of galactose (above 10 ppm) in PB 235 compared to RRIM 600, PB 217, PB 260, and GT 1.

2.1.4 Allergens

Allergens in *Hevea brasiliensis* have been extensively studied because they cause mild to severe allergic reactions through IgE-mediated response to NRL. Mild symptoms manifested in the forms of skin hives and itching, angioedema, and asthma could be as severe as anaphylaxis (Burke, 2006; Deval et al., 2008). This allergy is known to affect health workers and risk groups such as children with spina bifida and atopic individuals, especially those who are under medical care (Ghazal et al., 2009).

In *H. brasiliensis*, 15 latex proteins were recognised globally as an allergen, as shown in Figure 2.3 from the snapshot of the allergen database (Allergen Nomenclature, 2022). REF, beta 1,3-glucanase, SRPP, Hev b 5, and hevein precursor are classified as major allergens in natural rubber latex (Sussman et al., 2002) which later, in 2004, Yeang and colleagues found that Hev b 5 (Acidic protein) and Hev b 13 (Esterase) could be the biomarkers to estimate the allergenicity of latex gloves (Yeang et al., 2004).

Species	Allergen	Biochemical name	MW(SDS-PAGE)
Hevea brasiliensis (Para rubber tree (latex))			
	Hev b 1	Rubber elongation factor	14 kDa
	Hev b 2	beta-1,3-glucanase; glucan endo-1,3-beta-glucosidase, basic vacuolar isoform	35, 36.5 and 38 kDa
	Hev b 3	Small rubber particle protein (SRPP)	24 kDa
	Hev b 4	Lecithinase homologue; GDSL lipase/esterase family member	53-55 kDa
	Hev b 5	Acidic protein, unknown function	16 kDa
	Hev b 6	Hevein precursor	20 kDa
	Hev b 7	Patatin-like protein	42 kDa
	Hev b 8	Profilin	15 kDa
	Hev b 9	Enolase	51 kDa
	Hev b 10	Superoxide dismutase (Mn)	26 kDa
	Hev b 11	Class I chitinase	30 kDa
	Hev b 12	Non-specific lipid transfer protein type 1 (nsLTP1)	9 kDa
	Hev b 13	Esterase	42 kDa
	Hev b 14	Hevamine	30 kDa
	Hev b 15	Serine protease inhibitor	7.5 kDa

Figure 2.3 15 protein allergens in the NRL of *H. brasiliensis* (http://www.allergen.org/search.php?Species=Hevea%20brasiliensis)

Many of the suggested factors of allergenicity have been debated over the years as it may not be solely due to the protein epitope. Apart from the effect of glycosylation, some critical factors, such as resistance to proteolysis and enzymatic activity in the host, would disqualify the protein as an allergen (Huby et al., 2000). The glycosylated allergens have different mechanisms of allergenicity from protein, and the glycan motifs areylose and core-3-linked fucose (Altmann, 2007) In glycoprotein, its glycosyl group is involved in sensitisation due to the presence of sugar receptors on the antigens that improve the uptake of the proteins into the host (Ladics, 2018). However, Ladics

has also extensively demonstrated that carbohydrates are not necessarily the cross-reactive carbohydrate determinants (CCDs).

2.2 Post-translational modifications (PTMs)

PTMs are one of the response systems in plants to the stimuli that help plants adapt and cope with the environment. Phosphorylation, glycosylation, and ubiquitination are regarded as the three PTM types that have the most significant influence on how NRL-related enzymes ultimately operate (Yu et al., 2020).

2.2.1 Glycosylation

Glycosylation is where the glycan moieties or sugar residues are covalently attached to the protein with the involvement of glycosidases and glycosyltransferases (Kellokumpu et al., 2015). On a protein, it is possible to have multiple glycosylation sites. The multi-faceted glycosylation sites cause a galore of varieties in enzymes, substrates, and protein fluxes. This process plays a critical role in determining glycoproteins' structure and functional properties. Glycosylation improves nascent peptide solubility, folding, stability, assembly, and intracellular transport (Engel, 2015). In plants, complex N-glycan structures are vital to physiology, development, and stress tolerance (Strasser, 2022). The study of a collective set of glycan repertoires, glycome, is called glycomics.

Glycan is a polysaccharide set that exists typically in linear or branching structures. Glycan moieties are also usually found as conjugates, such as glycoprotein, proteoglycan, or glycolipid. The size and structure of glycans can vary depending on their enzymatic activities (Cohen & Varki, 2014). Their structures are not based on a rigid template, and hence, the complexity of profiling the glycans structures exceeds

that of deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) sequencing. Their macro- and microheterogeneity often make it challenging to handle, which requires further stabilisation methods to study the ideal glycan behaviour. Macroheterogeneity refers to the glycosylation sites of occupancy, while microheterogeneity refers to the diversity in the glycoforms attached to specific sites of glycosylation (Struwe & Robinson, 2019; Sumer-Bayraktar et al., 2012). The biological functions of glycan include cell proliferation, cell metastasis, cell-cell interaction, and immune response (Lin & Lubman, 2013). As a polysaccharide chain, glycan can influence an organism's physiological interaction.

The glycan repertoire is synthesised using glycosyltransferases by collecting many units of activated monosaccharides (equipped with adenosine triphosphate, ATP) within the cellular interaction. The *en bloc* attachment of glycans to the protein's surface (glycosylation) happens in the lumen of the endoplasmic reticulum (ER). After the process, they will be equipped with additional sugars to form many competent glycoproteins in the golgi (Stanley, 2011).

Plants carry different N-glycan types from mammals, distinguishable by the presence of these two glycan compositions: β1,2-xylose and core α1,3-fucose (Strasser et al., 2008). Generally, the β1,2-xylose non-reducing end in the plant lacks residues compared to mammalian. Three residues are rarely found in plants: (1) β1,2-linked *N*-acetylglucosamine (GlcNAc), *N*-acetylneuraminic acid (Neu5Ac) or also abbreviated as NANA, and galactose (Yoo et al., 2015). The following plant N-glycan found throughout the biosynthesis pathway process are presented in Figure 2.4:

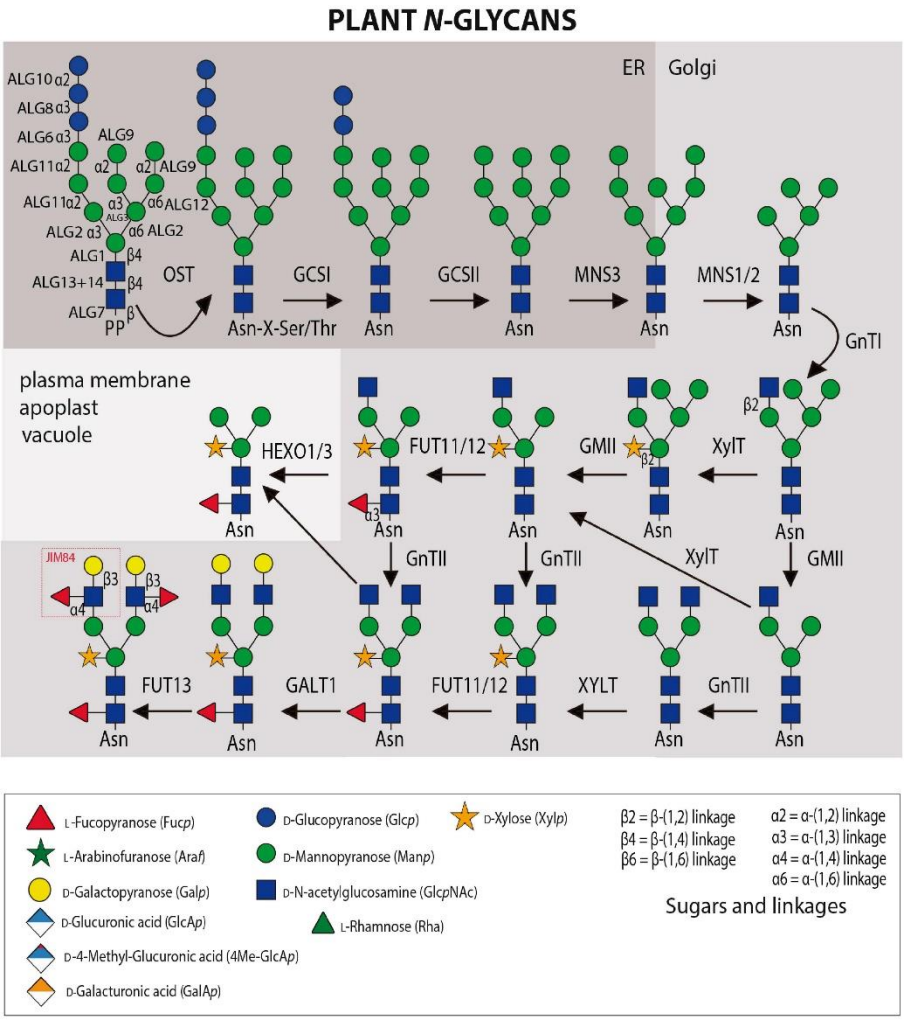

Figure 2.4 N-glycan plant biosynthesis pathway (Strasser et al., 2021)

N-linked glycan is the type attached to the nitrogen atom of amino acid, asparagine ($C_4H_8N_2O_3$), at the consensus sequence of Asn-X-Ser (or Thr) by an N-glycosidic bond where X can be any amino acid except Proline (Pro). The properties of glycoprotein are highly affected by N-glycan in terms of activity, conformation, antigenicity, solubility, and recognition by binding proteins in glycan—all N-glycan share a common core, $Man_3GlcNAc_2Asn$ (Stanley et al., 2009). According to a study in 2016, there are three types of N-glycans: (1) Oligomannose, where mannose acts as an extension; (2) Hybrid (Has an antenna extension above GlcNAc); (3) Complex (The most comprehensive glycan). As shown in Figure 2.5, they are classified based on the structural differences precedenting at mannose (Fisher & Ungar, 2016).

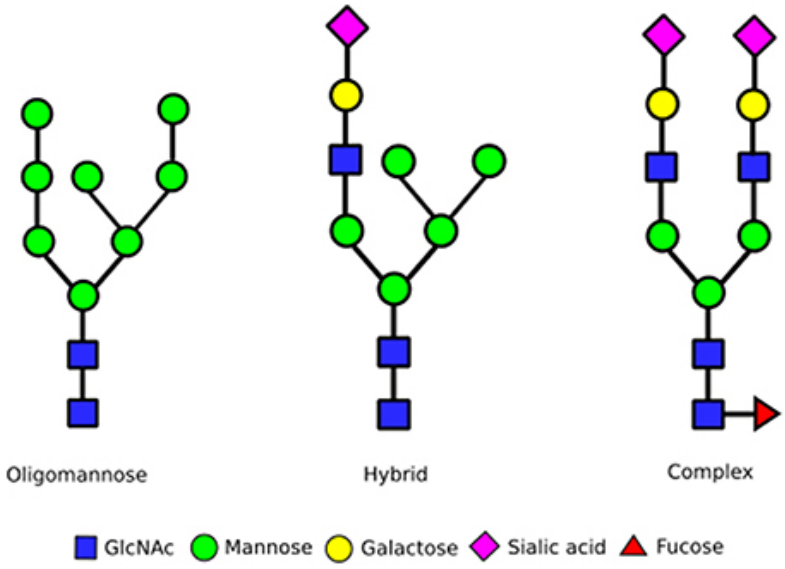

Figure 2.5 Types of N-glycan

O-glycan requires no specific sequon as the binding site, unlike the N-glycan. The O-glycan biosynthesis process is initiated with N-Acetylgalactosamine (GalNAc) transferase, a polypeptide enzyme, as a catalyst. It helps add monosaccharide GalNAc to the serine and threonine residues (Varki et al., 1999). They also consist of many clustering O-glycan groups due to an abundance of serine and threonine residues that could be found as mucin. The difference in the structure of N-glycan and O-glycan can be observed in Figure 2.6 as follows (Katie, 2020).

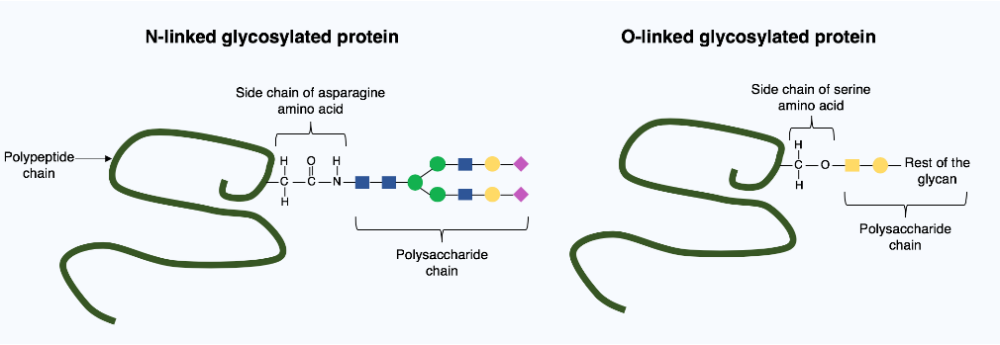

Figure 2.6 The structural difference between N- and O-glycan (Katie, 2020)

2.2.2 Glycoproteins in rubber latex of *H. Brasiliensis*

According to Habib et al., the total glycoproteins annotated through SwissProt and PlantGDB are 557 and 1025 in the RRIM 600 clone (Habib et al., 2017). Although it is already known that 15 latex allergens were collectively recognised in many *Hevea* clones, not all of them were glycosylated. So far, in the UniProt, only Hev b 2 and Hev b 13 were known to be glycosylated at the protein level, as shown in Table 2.1. While major latex allergen Hev b 4 (Q6T4P0), patatin (Hev b 7, accession id: O04008) and non-specific lipid-transfer protein (Hev b 12, accession id: Q8RYA8) were only found at transcript level which contradicted to the work of Berthelot et al., 2014 that claimed these proteins were glycosylated. Other than allergens, some biological roles of glycoproteins in NRL include the agglutination of rubber particles to maintain metabolite integrity and improve the affinity of REF for binding and insertion into the membrane (Berthelot et al., 2014; Chang et al., 2024).

Table 2.1 Glycosylated allergens in the latex of *H. brasiliensis* (https://www.uniprot.org/)

Uniprot id	Allergen	Protein	Sites of glycosylation
P52407	Hev b 2	Glucan endo-1,3-beta-glucosidase	Position 63, 350, 364 Asn 27 and Asn 314 (Rodríguez-Romero et al., 2014).
Q7Y1X1	Hev b 13	Esterase	Position 186, 193, 313 Asn163, Asn170, and Asn292-Phe-Ser (Rougé et al., 2010).

* Hev b 2 and Hev b 13 are major allergens in NRL of H. brasiliensis, which proved to be extensively glycosylated (Raulf-Heimsoth et al., 2007).

Many glycosylated proteins were predominantly present in these pathways, such as carbohydrate and energy metabolism, rubber latex metabolism degradation function, and cellular processes (Yu et al., 2020). Yu and colleagues also found that the glycoprotein isoforms of hydroxymethylglutaryl-CoA synthase (HMGS) were

responsive to the rubber latex yield stimulant, mainly from the rubber biosynthesis pathway of mevalonate (MVA) pathway. Alternately, glycoproteins produced from the methylerythritol (MEP) pathway are still unknown. In non-rubber isoprenoid pathways, RRIM 600 has a higher N-glycan biosynthesis expression than the Chinese clone, Reyan7-33-97 latex, on transcript level (Xin et al., 2021). Unfortunately, despite the obvious relevance of glycosylation in NRL, the impact of the N-glycan biosynthesis pathway on rubber biosynthesis and defence mechanism was not further examined.

2.3 Shotgun Proteomics

In global proteome identification, there are two types of mass spectrometry methods for proteome identification, which are bottom-up and top-down approaches, with the prior being the most common pipeline (Cassidy et al., 2021). Bottom-up proteomics, or its other name, shotgun proteomics, is initiated from the proteolytic digestion of protein instead of directly being analysed through MS (Suna & Mayr, 2018). The workflow is then followed by the separation of peptides in reverse-phase liquid chromatography (RP-LC) (Mikesh et al., 2006). The separated peptides are consecutively introduced into MS through electrospray ionisation (ESI), and ions are activated to initiate the fragmentation of peptides through tandem MS/MS. Throughout these steps, raw data from the amino acid sequence will be obtained from data acquisition software and annotated on the peptide database (Dyer et al., 2019; van Vliet, 2014).

2.3.1 In-solution digestion

In the ordinary proteomic study, two steps need to be done before the enrichment: the extraction and digestion of natural rubber latex. The typical method used for the digestion is all in-solution for cysteine-linked mass spectrometry analysis (Müller & Winter, 2017). The cysteine-linked MS analysis is popular due to the known function of cysteine linkage, which assists in protein folding and stabilisation (Feige et al., 2018). The example of cysteine linkage (protein inter- and intra-molecular disulfide bridges) is shown in Figure 2.7.

Before proteolytic digestion, the reduction and alkylation processes are standardised to ease protein identification. The disulfide linkage that causes protein folding can be reduced using dithiothreitol (DTT). The thiols in DTT will reduce the cystine to cysteine through the breakdown of the disulfide bonds (Weder & Belitz, 2003). To avoid the reformation of the disulfide bond, the free SH- (sulfhydryl or thiol) group is alkylated using iodoacetamide (IAA) (Mouchahoir & Schiel, 2018).

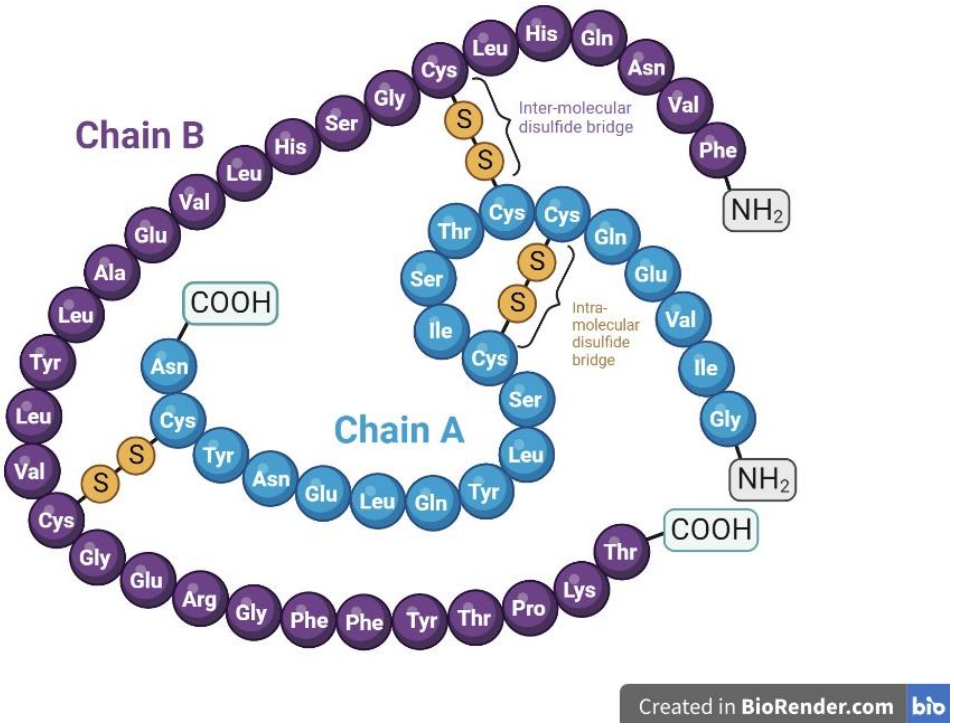

Figure 2.7 Human insulin inter- and intra-molecular disulfide bridges (Created with Biorender.com)

The subsequent step for fragmentation of large and intact protein, proteolytic digestion using trypsin, is selected for general use in mass spectrometry. This is because it exhibits a very high catalytic activity of protease and cleaves at the C-terminal of lysine and arginine sides (Manea et al., 2007). Through tryptic digestion, protein will be fragmented into small peptide sequences. Trypsin can produce water-soluble and charged peptides that can be well-separated using Strong cation exchange (SCX) and reversed-phase chromatography. The charged peptides can also be well-ionised through ESI (Drabovich et al., 2013). Due to the tryptic digestion efficiency, most of the deposited data in the global proteome database were approximately 96 % generated through this digestion (Dupree et al., 2020).

Through protease digestion, peptides are released, and some are bound with glycan moieties (glycopeptides). To study glycans, the glycan moieties may be hydrolysed from peptides. The PNGase F and A enzymes can cleave, specifically the N-linked glycopeptides. In contrast, O-glycan is commonly cleaved through chemical reactions such as β-elimination or hydrazine hydrolysis (Yang et al., 2018a).

2.3.2 Enrichment of glycopeptides

Glycosylation is known for its macro- and micro-heterogeneity and needs niche analytical consideration. Besides their known meta heterogeneity, non-glycosylated backgrounds in a biological sample create an abundance of signal noises in MS/MS. In addition, hydrophobic molecules are more likely to be ionised in the electrospray ionisation, causing the low proportions of hydrophilic glycopeptides to be suppressed during the process. Hence, the separation of glycopeptides from their

non-glycosylated species will improve the efficiency and sensitivity of the MS/MS. Three factors significantly influenced the enrichment methods: 1. The charges or hydrophilicity of glycopeptides, 2. Stationary and mobile phase, and 3. Bioorthogonal handles of samples and buffer/assay (Riley et al., 2021). For the separation, a few enrichment strategies are commonly deployed before the deglycosylation process. A multi-enrichment technique must be done to enrich glycosylated peptides comprehensively.

Many glycopeptide enrichments have been done for global glycopeptide analysis, such as Hydrophilic interaction liquid interaction chromatography (HILIC), lectin, and hydrazide chemistry methods (Zhu et al., 2017). Among these three methods, the HILIC-based strategy was proven to employ the best results regarding glycopeptide coverage (Zacharias et al., 2016; C. Zhang et al., 2016). HILIC is the most favourable method for obtaining intact N- and O-linked glycopeptides (Yang et al., 2017). This method can simultaneously enhance the intactness of glycopeptides and glycan (Qing et al., 2020). Reversed-phase of solid-phase (RP-SPE) HILIC is the preferable type due to its simplicity, effectiveness, and cost-friendliness. In contrast to prior research, strong anion exchange (SAX) has garnered increased attention due to its superior ability to yield glycopeptides backbone in human serum compared to HILIC (Yang et al., 2017).. However, there is no clear mechanism on how SAX can yield glycopeptides better than HILIC. It is also unknown if SAX can improve the plant's glycopeptide capturing. SAX is a salt-based gradient method where the ionic strength is increased through consecutively added net charges in the pH buffer. When the pH of the buffer exceeds the pI of the protein, the separation occurs (Ahamed et al., 2007).

Meanwhile, Titanium dioxide (TiO_2) is a more specific method to enrich sialylated glycopeptides and phosphopeptides. In biomedicine, the sialylation of a glycopeptide is often associated with diseases (Nicholls et al., 2007). However, little is known regarding sialylated glycopeptides in the plant, not to mention in the latex of *H. brasiliensis* itself. Although they commonly could not be retrieved from the plant, trace amounts were observed in a few plants, including *Arabidopsis thaliana*, tobacco, banana, apple, mung bean sprouts, and yeast (Zeleny et al., 2006). Identifying small amounts of sialic acid would require using lower detection limits in the PEAKS7.5 database.

2.3.3 Deglycosylation

Deglycosylation is a crucial process to characterise glycans and their glycosylation sites in glycomics (Xiao et al., 2019). PNGase F is the most used enzyme to obtain broad and general N-glycan moieties by cleaving the asparagine-linked glycoprotein. They released almost all N-linked glycosylated proteins except for the N-glycan containing-(1,3)-linked core fucose (Freeze & Kranz, 2010). This α-(1,3)-linked core fucose can be found in plant and insect glycoproteins. To get a broader glycan substrate from plant and insect peptides, mixing PNGase A with PNGase F is recommended to assist the glycan cleavage. However, excessive PNGase A may reduce the deglycosylation activity of intact glycopeptides due to their huge molecular weight of 80 kDa (O'Neill, 1996).

2.4 Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS)

Liquid chromatography with tandem mass spectrometry (LC-MS/MS) is highly preferable and attributable to its integrated system, which has high selectivity and sensitivity. Due to their ability to eliminate the difficulty of traditional techniques, they are called shotgun proteome analysis. LC-MS/MS system recognises analytes from two physical properties, at the very least: (1) precursor and (2) product ion mass (Jannetto, 2015). Through this instrument, bioanalytical analysis can be done to quantify the targeted analytes concomitantly. They have higher sensitivity than the detection limits in ordinary LC-MS, which promises better results accompanied by the enrichment method applied to the sample. This feature enables the analysis of complex proteomes to be executed. One of the most prominent technologies of MS/MS is Orbitrap (Thermo Fischer Scientific, Massachusetts, USA). Apart from that, what makes tandem LC-MS/MS distinguishable from conventional LC-MS is the presence of ionisation techniques such as matrix-assisted laser desorption/ionisation (MALDI) and electrospray ionisation, ESI (Nadler et al., 2017).

2.4.1 Ionisation techniques: Electron Spray Ionization (ESI) and Matrix-assisted laser desorption/ionisation (MALDI)

ESI is an ion generative technique with high voltage applied to the liquid-producing charged aero spray. It is a common ionisation technique coupled to LC-MS/MS for proteomics (Jabbour & Snyder, 2014). In ESI, the liquid sample is injected into the metal-narrow capillary with a fixed voltage of 3-5 kV, where the droplets are transformed into charged droplets. The charged droplets are then evaporated into analytes with the application of dried nitrogen gas and adequate heat (Pitt, 2009). The evaporation happens when the kinetic energy is increased, increasing the droplets'

surface area and density. As a result, the droplet radius is reduced, and afterwards, the analytes are produced and pulled into the vacuum space of the mass spectrometry region (Ho et al., 2003).

The proteomic analysis using MALDI-TOF/TOF started with mixing the sample with a matrix (Singhal et al., 2015). The matrix solution is an organic compound that contains energy-absorbent. The mixture is ionised by shooting a laser beam on the sample plate. Throughout the process, singly protonated ions are generated through desorption and ionisation. These ions are then separated using a fixed acceleration based on their *m/z* ratio. The analytes are then detected and measured using a time of flight (TOF) analyser, where the time of analytes to travel along the flight tube and deflected to the detector is measured. A plethora of spectra called peptide mass fingerprints (PMF) is generated by detecting the signal. Four categories of compounds are commonly used as MALDI matrices: 1. Benzoic acid derivatives, 2. Cinnamic acid derivatives, 3. Heterocyclic matrices, and 4. Other matrices are acetophenone derivatives (Leopold et al., 2018). The most common matrix for peptides and oligonucleotides is α-cyano-4-hydroxycinnamic acid (αCHCA), while for oligosaccharides is dihydroxybenzoic acid (DHB) (Cutignano, 2019).

2.4.2 Mass spectrometry mode: Collision-induced dissociation (CID)

After the protonation of peptides, abundant peptide cations will be selected for further fragmentation in the MS region. In ESI, the fragmentation mode occurs within the ionisation chamber where the dissociation of ions takes place.

Collision-induced dissociation (CID) is one of the mainstream fragmentation methods in tandem mass spectrometry. This fragmentation occurs in the ion traps

(Murrell et al., 2003). The high-speed-protonated ions resulting from a strong electrical field collide with the neutral gases. The bond dissociation occurs when the high-speed ions increase the excitement level of ion internal energy. This conversion causes the release of two types of ions: b and y ions at the N- and C-terminal (Dudley & Bond, 2014). However, the drawback of CID is that labile PTMs, such as phosphorylation and O-GlcNAc at the serine/threonine residues, are naturally lost throughout the process (Gruhler et al., 2005). The phosphorylated group is labile due to the weak phosphodiester bond, the electronegativity of the phosphoryl group and low abundance (Tsiatsiani et al., 2015). In the tandem CID, it is typical to get glycopeptide information with less or no peptide backbone sequence (Huddleston et al., 2002).

2.4.3 Data acquisition

Thousands of mass spectra are produced during the chromatographic separation of a complex mixture of peptides derived from tryptic digestion, and sophisticated software is required to set the required parameters and to interpret the spectra patterns.

Data acquisition software is a parameter setting, recording, and reporting tool that quantifies peptide patterns by measuring their mass and intensities. The segregation and sorting of peptides are based on their mass/charge (*m/z*) (Coon et al., 2005). In tandem with MS/MS, there are two stages of mass spectrometry activity in data-dependent acquisition (DDA). During the MS^1 scan, ions which fit the parameter and criteria set will be screened for fragmentation in MS^2 (Defossez et al., 2021). While in MS^2, the scanning of accumulated product ions occurs.