



2014

## METHODS DEVELOPMENT IN BIOLOGICAL MASS SPECTROMETRY: APPLICATION IN GLYCOPROTEOMICS

Sanja Trajkovic

University of Kentucky, sale9yu@yahoo.co.uk

[Right click to open a feedback form in a new tab to let us know how this document benefits you.](#)

---

### Recommended Citation

Trajkovic, Sanja, "METHODS DEVELOPMENT IN BIOLOGICAL MASS SPECTROMETRY: APPLICATION IN GLYCOPROTEOMICS" (2014). *Theses and Dissertations--Chemistry*. 36.  
[https://uknowledge.uky.edu/chemistry\\_etds/36](https://uknowledge.uky.edu/chemistry_etds/36)

This Doctoral Dissertation is brought to you for free and open access by the Chemistry at UKnowledge. It has been accepted for inclusion in Theses and Dissertations--Chemistry by an authorized administrator of UKnowledge. For more information, please contact [UKnowledge@lsv.uky.edu](mailto:UKnowledge@lsv.uky.edu).

## **STUDENT AGREEMENT:**

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

## **REVIEW, APPROVAL AND ACCEPTANCE**

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Sanja Trajkovic, Student

Dr. Bert C. Lynn, Major Professor

Dr. Dong-Sheng Yang, Director of Graduate Studies

METHODS DEVELOPMENT IN BIOLOGICAL MASS SPECTROMETRY:  
APPLICATION IN GLYCOPROTEOMICS

---

DISSERTATION

---

A dissertation submitted in partial fulfillment of the  
Requirements for the degree of Doctor of Philosophy in the  
College of Chemistry  
At the University of Kentucky

By

Sanja Trajkovic

Lexington, Kentucky

Director: Dr. Bert C. Lynn, Professor of Chemistry

Lexington, Kentucky

2014

Copyright@Sanja Trajkovic 2014

## ABSTRACT

### METHODS DEVELOPMENT IN BIOLOGICAL MASS SPECTROMETRY: APPLICATION IN GLYCOPROTEOMICS

Proteomics refers to global characterization of the full set of proteins present in a biological sample. Various analytical disciplines contribute to proteomics but mass spectrometry became method of choice for analysis of complex protein samples. Mass spectrometry allows for high throughput analysis of the proteome but, moreover, it has the ability to acquire higher-order information such as post-translational modifications (PTM). Glycosylation is the most abundant PTM on eukaryotic proteins.

This dissertation will focus on method development for structural proteomics that will be utilized to explain the glycoproteome of obligate intracellular protozoan parasite *Toxoplasma gondii* as a model system.

Optimization of sample preparation is addressed in the first part of this dissertation. Sample preparation for mass spectrometry analysis is a critical step in the proteomics workflow because the quality and reproducibility of sample extraction and preparation significantly impacts the separation and identification capabilities of mass spectrometers. Also, there are problems unique to intracellular parasites as limited amount, host cell impurity and choice of the host. The additional obstacle is to extract only glycosylated proteins for which there is no one standard method. Here we report the optimal sample preparation method utilizing agarose bound Concanavalin A (Con A) beads to efficiently pull down glycoproteins, dialyze and analyze them using MuDPIT. This method was further enhanced by passing the non-retained protein fraction (first flow-through) through a second Con A column and then passing the second non-retained protein fraction (second flow-through) through the third Con A column (3 sequential pull-downs) yielding 394 benchmark proteins.

Glycoproteome of *Toxoplasma gondii* is not yet fully understood. However, evidence suggests that glycosylation could be essential for cyst formation and maintenance which is characteristic of chronic stage of disease. The focus of the second part of dissertation is to better understand the differences in glycoproteomes of tachyzoites and tissue cysts. Cyst proteins pulled down using optimized sample preparation method that do not appear in the tachyzoites pulldowns could be critical elements in the structural stability of the tissue cyst.

KEYWORDS: Mass spectrometry, glycoproteomics, lectin affinity purification, MuDPIT, *Toxoplasma gondii*.

Sanja Trajkovic  
Student's Signature

May 6, 2014  
Date

METHODS DEVELOPMENT IN BIOLOGICAL MASS SPECTROMETRY:  
APPLICATION IN GLYCOPROTEOMICS

By

Sanja Trajkovic

Dr. Bert C. Lynn  
Director of Dissertation

Dr. Dong-Sheng Yang  
Director of Graduate Studies

May 6, 2014

To my Mother, always.

## ACKNOWLEDGEMENTS

The experience of the Ph.D. program profoundly depends on the choice of advisor and I consider myself very lucky to be part of the research group of Dr. Bert Lynn. His inspiring guidance, enthusiastic encouragement, patience and support will always be remembered and deeply appreciated.

I am very grateful for the support and guidance provided by my advisory committee because this work greatly benefited from their support and suggestions.

Many thanks to Dr. Anthony Sinai, professor of Microbiology, Immunology and Molecular Genetics, for teaching me skills necessary to work with *T. gondii* and generously allowing me to use resources of his laboratory, which, in turn, made it possible for me to complete this project.

I would also like to thank Dr. Jack Goodman, manager of the University of Kentucky Mass Spectrometry Facility, for all his help and advice.

Daily dynamics, positive and relaxed atmosphere in the Lynn Laboratory are just some of the things for which I am grateful to all my group colleagues. I will always remember the time we spent together with a smile on my face.

Last but not least, I would like to thank my mother and father for their unconditional love and support. My life's achievements would never be realized without their never-ending belief in me.



## TABLE OF CONTENTS

ACKNOWLEDGEMENTS .....	iii
LIST OF FIGURES .....	vii
LIST OF TABLES .....	ix
CHAPTER 1 STATEMENT OF RESEARCH .....	1
CHAPTER 2 INTRODUCTION .....	3
Proteomics.....	3
Proteomics vs genomics .....	4
Proteomics Strategies .....	5
Separation Technologies.....	6
High Performance Liquid Chromatography (HPLC).....	6
Reversed Phase (RP) Chromatography .....	7
Liquid Chromatography – Mass Spectrometry (LC-MS) for peptides .....	10
Strong Cation Exchange Chromatography.....	11
Multidimensional Protein Identification Technology (MuDPIT) .....	12
Electrospray Ionization - Ion Trap Mass Spectrometry .....	15
Electrospray ionization (ESI) – Making Droplets.....	15
The Path of Ion .....	17
Tandem Mass Spectrometry.....	22
Bioinformatics for Proteomics.....	23
CHAPTER 3 SAMPLE PREPARATION FOR	
MASS SPECTROMETRY OF GLYCOPROTEINS .....	26
Posttranslational modifications of proteins.....	26
Protein Glycosylation.....	27
Glycans .....	28
Lectins.....	32
Concanavalin A (Con A).....	32
Lectin Affinity Chromatography .....	35
Cell disruption and Protein Solubilization.....	36

Protein Purification Techniques.....	37
Gel Electrophoresis .....	38
Three-Layer Sandwich Gel Electrophoresis.....	40
Dialysis.....	41
Experimental.....	45
Materials .....	45
Equipment.....	45
Instrument Optimization – Effect of Inlet Capillary Temperature .....	46
Cell Culture Growth, Maintenance and Lysis.....	47
Lectin Pull-down.....	48
Cell Lysate Preparation .....	48
Con A Column Preparation .....	48
The Pull-down Procedure .....	48
Dialysis .....	50
The Three-layer Sandwich Gel Electrophoresis .....	51
Sample Preparation for Mass Spectrometric Analysis.....	52
Mass Spectrometry Analysis.....	53
Protein Data Analysis .....	55
Results and Discussion .....	56
Instrument Optimization – Effect of Inlet Capillary Temperature.....	56
Cell Disruption method and Protein Denaturation.....	58
Elution Buffer Evaluation .....	60
Development of glycoprotein purification method.....	62
Comparison of three layer sandwich gel and dialysis for purification of glycoproteins.....	62
Three layer sandwich gel electrophoresis.....	63
Dialysis .....	66
Sequential Pull-down Method.....	67
Conclusion .....	73
CHAPTER 4 GLYCOPROTEOMES OF TACHYZOITES AND	
TISSUE CYSTS AND THEIR DIFFERENCIES .....	75
<i>Toxoplasma gondii</i> .....	75

Structure of <i>Toxoplasma gondii</i> .....	76
Glycoproteome of <i>Toxoplasma gondii</i> .....	77
Experimental.....	79
Materials .....	79
Equipment.....	79
Cell Culture Growth and Maintenance .....	79
Lectin Pull-down.....	80
Cell Lysate Preparation .....	80
Con A Column Preparation.....	81
Sample Preparation for Mass Spectrometric Analysis.....	83
Mass Spectrometry Analysis.....	83
Protein Data Analysis .....	85
Results and Discussion .....	86
Invasion related glycosylated proteins .....	88
Energy related glycosylated proteins .....	90
DNA and translational machinery related glycosylated proteins.....	90
Other identified glycosylated proteins .....	91
Bradyzoites related glycosylated proteins.....	97
Conclusion.....	99
CHAPTER 5 CONCLUSION.....	102
APPENDIX.....	105
REFERENCES .....	127
VITA.....	137

## LIST OF FIGURES

Figure 2.1. Principles of reversed-phase liquid chromatography .....	8
Figure 2.2. Principles of the strong cation exchange liquid chromatography .....	12
Figure 2.3. A. Finnigan LCQ Deca ion trap mass spectrometer .....	14
Figure 2.4. ESI chamber, Finnigan LCQ Deca ion trap mass spectrometer .....	16
Figure 2.5. Electrospray ionization (ESI) process .....	18
Figure 2.6. Ion trap elements: two end-caps and one ring electrode .....	19
Figure 2.7. Mathieu stability diagram for a 3D quadrupole ion trap .....	21
Figure 2.8. Peptide fragmentation notation.....	23
Figure 3.1. Example of $\alpha$ -glycosidic bond.....	29
Figure 3.2. The core sequence of N-linked glycans.....	31
Figure 3.3. Example of glycans that Con A binds. The parts of the glycan required for binding are in the box. ....	33
Figure 3.4. A. D-mannose, B. D-glucose, C. methyl- $\alpha$ -D-mannoside and D. methyl- $\alpha$ -D-glycoside, Haworth projection .....	34
Figure 3.5. Affinity chromatography principle.....	35
Figure 3.6. The pull-down procedure. ....	49
Figure 3.7. A. Dialysis set-up. B. Closer look at the regenerated cellulose membrane...	51
Figure 3.8. Inlet capillary temperature optimization with infusion .....	57
Figure 3.9. Inlet capillary temperature optimization with chromatographic separation..	57
Figure 3.10. Elution profile of mannose-BSA .....	58
Figure 3.11. Effect of cell lysate heating to the qualitative identification of the proteins .....	60
Figure 3.12. The effect of the composition of elution solution on the qualitative identification of proteins .....	61
Figure 3.13. Sandwich gel with dye in different concentrations of ES-1100 or Tris + $\alpha$ -methyl- mannoside.....	65
Figure 3.14. Sandwich gel with dye in different concentrations of ES-1100 or Tris + $\alpha$ -methyl-mannoside.....	66
Figure 3.15. Sequential pull-down.....	68

Figure 3.16. Mass spectrum .....	71
Figure 3.17. Chromatogram .....	72
Figure 4.1. Staining of the tissue cyst .....	78
Figure 4.2. Staining of the tissue cyst .....	78
Figure 4.3. <i>Toxoplasma gondii</i> ME49, tachyzoites, pull-down triplicate. ....	86
Figure 4.4. Pull-down method dependence on molecular weight of the protein .....	87
Figure 4.5. Distribution of identified glycoproteins according to function .....	92

## LIST OF TABLES

Table 3.1. Glycan building monosaccharides .....	30
Table 3.2. Results of the triplicate sample analysis .....	69
Table 3.3. List of expected fragment ions of the ion 638.75 m/z .....	72
Table 4.1. Proteins identified in three out of three replicates .....	93
Table 4.2. Proteins identified in tissue cysts .....	100
Table A.1. Proteins identified in two out of three replicate .....	105
Table A.2. Proteins identified in one out of three replicates .....	109

## CHAPTER 1

### STATEMENT OF RESEARCH

*Toxoplasma gondii* is obligate intracellular protozoan parasite that causes the disease called toxoplasmosis. This parasite undergoes three stages in its lifecycle. In the primary host the parasite is in the stage of oocysts. Shortly upon ingestion by the secondary host it transforms into rapidly growing tachyzoites causing the acute stage of toxoplasmosis. Tachyzoites localize in skeletal muscle, myocardium, brain, and eyes and develop into slow growing bradyzoites contained in tissue cysts. These cysts remain in the host throughout the life causing the chronic toxoplasmosis.

The logical way to prevent the chronic toxoplasmosis is to prevent the cyst formation. However, that mechanism of parasite's transformation from the tachyzoite to bradyzoite form and the formation of tissue cysts is yet unknown. According to current knowledge, there is a strong indication that the differences in protein glycosylation between tachyzoites and tissue cysts may provide a better insight into this process and, therefore, suggest the possible ways to prevent it.

Estimates suggest that over 30% of global human population is infected and suffers from toxoplasmosis. According to Centers for Disease Control and Prevention (CDC), the number of infected men, women and children in the USA only is around 60 million. Currently, only palliative care can be utilized but the cure is yet to be discovered.

Despite such a wide spread infection that sparks scientific interest, this parasite's weaknesses and, therefore, potential drug targets, remain largely unknown. One of the reasons is the lack of methodologies to study the parasite. This dissertation represents the efforts toward developing a method to study the glycoproteome of intracellular organisms and consequently increase the knowledge base of *Toxoplasma gondii*.

The dissertation is organized in three main parts. The first chapter outlines the current knowledge and background of only a relevant subset of methodologies applied in proteomics. In addition to the definition of proteomics and bottom-up proteomics, the discussion focuses on separation methods as well as ionization techniques, mass

spectrometry and data analysis. These methodologies provide valuable tools utilized to study the glycoproteome of *Toxoplasma gondii*.

Second chapter describes the steps in developing the workflow to study the glycoproteome. Methods are validated utilizing a whole cell lysate of HeLa cells as a benchmark. The workflow is crafted by choosing the best approach for each step in the method.

The established workflow developed in Chapter 2 is applied to study glycoproteomes of tachyzoite and cyst form of *Toxoplasma gondii* is discussed in the third chapter. Analyzing the differences between the two proteomes provides the direction for further research that is beyond the scope of this dissertation.



## CHAPTER 2

### INTRODUCTION

#### Proteomics

Proteomics is an ever-growing set of qualitative and quantitative methodologies used to define the structure and function of all proteins. [1, 2] Considering the role of proteins and their interactions in cellular function, proteomics has become a leading technology to further unravel biological processes. [3]

Various analytical disciplines contribute to proteomics but mass spectrometry became the method of choice for analysis of complex protein samples. Three pillars of mass spectrometry based proteomics are protein separations science, mass spectrometry (MS) and bioinformatics. [4]

Protein separation science focuses on enrichment of a subset of proteome from a highly complex proteome using a diverse set of strategies such as depletion of highly abundant proteins and sample fractionation attempting to decrease sample complexity. These strategies contribute to the overall analysis by reducing competition for ionization and improving duty cycle thus extending dynamic range of proteins identified by mass spectrometry.

Mass spectrometry allows for high throughput analysis of proteome but, moreover, it has the ability to acquire higher-order information such as protein localization, protein-protein interactions and protein post-translational modifications (PTM). [5]

High throughput mass spectrometry generates extensive amounts of data that would be impossible to analyze without bioinformatics tools. Utilizing various software, organizing and analyzing overwhelming amounts of biological data becomes a swift task enabling substantial contribution to the knowledge base.

Proteomics, as a large-scale study of proteins, can generally be divided in three categories: structural proteomics that analyzes protein structure, functional proteomics

that analyzes protein-protein interactions and expression profiling proteomics that analyzes the expression of proteins.

The global analysis of expressed proteins can identify particular proteins affected by a certain treatment or disease thus helping in drug target or biomarker discovery. This dissertation will focus on this category utilizing structural proteomics to explain the glycoproteome of *Toxoplasma gondii* as a model system.

### Proteomics vs genomics

Genome refers to the genetic material of an organism including both genes and non-coding sequences of DNA and RNA. Genome of a certain organism is the same throughout all the cells comprising that organism. After transcription of DNA to RNA, proteins are expressed in the process called translation. However, a large portion of the genome will be silenced until environmental stress or other extracellular signal stimulates the expression of a certain protein. Therefore, regardless of technique, it will not be possible to visualize all proteins that genome can code for. [6]

While the genome is the same from cell to cell, the proteome is very different depending on the cell type and function. Also, as a consequence to stimuli genes can alter protein expression or transform their function. Therefore, information obtained from genome sequencing does not depict currently expressed proteome.

Most proteins are posttranslationally modified that further contributes to complexity. There are many types of posttranslational modifications (PTM) such as glycosylation, phosphorylation, acetylation, methylation and many other. These modifications usually regulate protein function and make the whole proteome very complex and dynamic. As the name suggests, PTMs take place after translation, and therefore, cannot be fully understood and studied at the genome level.

However, genomics provided an enormous benefit to proteomics. Those two technologies are complementary and some scientific questions are better addressed by one over the other approach. [3] Simply, the genome provides possibility while the proteome is your reality.

Genomics alone cannot determine the nature of functional proteins. Hence, proteomics, as a study of the proteome, is necessary to analyze the current state of any organism, in health or disease, at molecular level.

### Proteomics Strategies

Two major strategies used for protein identification are top-down and bottom-up proteomics. The top-down approach analyzes intact proteins while bottom-up analyzes chemically or enzymatically produced peptides. Even though the top-down strategy has better sequence coverage and characterization of posttranslational modifications, front-end separation of intact proteins is arduous and it requires higher mass accuracy instruments. The bottom-up approach is much more suitable for complex protein mixtures and large-scale analysis. [1, 7] Therefore, bottom-up proteomics will be discussed in more details.

Bottom-up proteomics is based on proteolytic digestion of proteins prior to mass analysis. Among many available proteases trypsin is the most widely used for generating peptides for mass spectrometry analysis. Trypsin is a serine protease that cuts the protein sequence on the C-terminal side of lysine and arginine unless they are followed by proline. This results in peptides with at least two protonation sites, C-terminal basic residue and N-terminal amino group which is a minimum requirement for tandem mass spectrometry. [4, 8-10]

Tandem mass spectrometry is a multi-stage mass analysis that provides a unique capability to obtain structural information that can be useful in structure elucidation. It is the key technique for protein or peptide sequencing and posttranslational modification analysis. Collision-induced dissociation (CID) produces peptide fragments and provides information about peptide sequence. Multiple collisions with gas atoms internally heat the peptide cations which results in peptide bond fragmentation. However, this process also leads to losses of water,  $\text{NH}_3$ , and labile posttranslational modifications as sulfation or nitration of tyrosine. Consequently, sequence information of large peptides is limited. [5, 11]

The result of peptide sequence analysis is then searched against database of in silico generated peptide fragmentation patterns where algorithms are utilized to reconstruct the protein sequence. However, there is an overwhelming number of proteolytic peptides in a sample, and only a small subset of all peptides in a sample can be analyzed in a single MS run. Mass spectrometers only sample a small percentage of the total number of peptides in a sample which limits the number of proteins in a sample that can be identified. [7]

Overall, the bottom-up strategy, with good front-end separation is well suited for large scale protein identification and analysis of posttranslational modification, therefore, it is utilized in method development for glycoproteomics.

## Separation Technologies

### High Performance Liquid Chromatography (HPLC)

Prof. Csaba Horváth first mentioned the HPLC acronym in 1970 describing high pressure liquid chromatography where high pressure of maximum 500 psi was used to generate the flow required for liquid chromatography in packed columns. Soon after, new pumps were developed with maximum pressure of up to 6,000 psi that together with improved injectors, detectors, and advanced packing material gave a new name to the old acronym, high performance liquid chromatography. [12]

Today HPLC is one of the most powerful tools in analytical chemistry with the ability to separate, identify, and quantify compounds that are present in any sample dissolved in a liquid.

Components of HPLC system are: injector, pump, separation column and detector. Appropriate mixture of solvents, mobile phase, is continuously running through the system under high pressure. After injection the analyte, now within the mobile phase, is forced through the separation column. Column is packed with the stationary phase comprised of small particles (3 – 10  $\mu\text{m}$ ) with specific residue that have different affinity

toward components of the passing analyte. Considering the affinity of the residue, components of the analyte are separated by how long they are retained on the residue. The least retained components are eluted from the column first with the short retention time followed by strongly retained ones with the long retention time. From the column the flow is directed toward the detector which in our case is mass spectrometer.

The choice of liquid chromatography technique depends on the type of sample. Reversed-phase chromatography and ion exchange chromatography together with mass spectrometry analysis are typically used for peptide separation in proteomics.

### Reversed Phase (RP) Chromatography

Reversed phase liquid chromatography (Figure 2.1) is based on interaction between the analyte within a polar mobile phase and hydrophobic stationary phase. Nonpolar stationary phases are often made of spherical silica particles with surface modified with hydrocarbon chains called the bonded phase. Particles are made porous in order to increase their surface area and consequently the number of hydrocarbon chains that will bind the analyte.

In highly polar (aqueous) mobile phase analytes bind to the stationary phase. Using organic solvent gradient, when the non-polarity of the mobile phase matches or exceeds non-polarity of the analyte, it will desorb from the stationary phase into the mobile phase. Therefore, analytes are eluted in the order of increasing hydrophobicity.

Analyte interacts with the stationary phase via weak van der Waals or dispersive interactions that are the result of overlap of the outer electron clouds between the analyte and stationary phase.

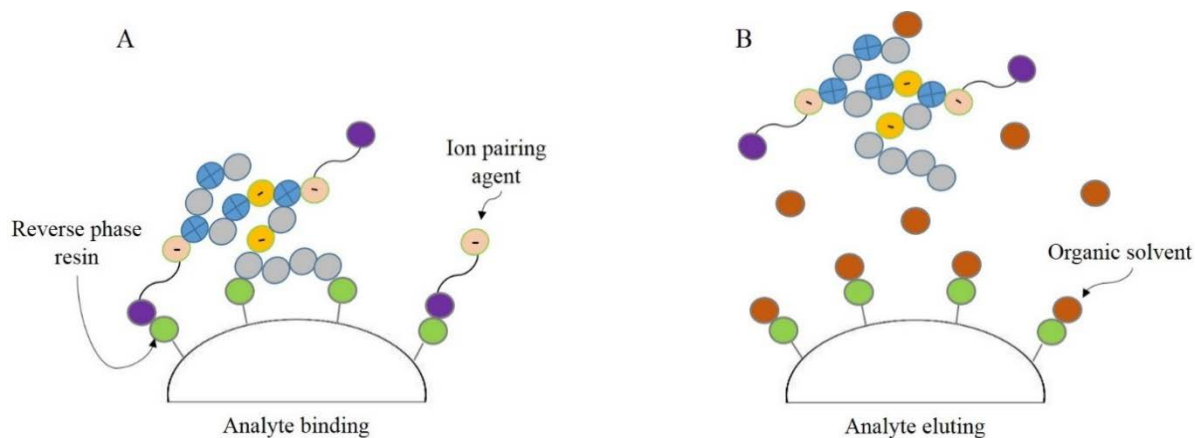
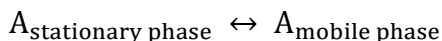


Figure 2.1. Principles of reversed-phase liquid chromatography. A. Analyte bound to the stationary phase. B. Analyte eluted from the stationary phase due to presence of organic solvent.

Mechanism of the analyte transfer from the mobile phase to or from the stationary phase can be explained by adsorption and desorption of the analyte from the stationary phase. In both cases, transfer of the analyte is explained by equilibrium process of the environment exchange, from mobile phase to stationary phase and back to mobile phase, as described by the equation. The stage when the equilibrium has been established is called theoretical plate height. [13, 14]



Stationary phase is packed in the column. Properties of chromatographic column such as retention time, resolution and selectivity depend upon particle size, pore size, functionality and chain length of the bonded phase and column length. They also depend on mobile phase composition, gradient profile as well as the length and internal diameter of connecting tubing. Resolution equation explains the relationship between retention ( $k$ ), efficiency ( $N$ ), and selectivity ( $\alpha$ ):

$$R_s = \frac{1}{4} N_{1/2} \left( \frac{\alpha - 1}{\alpha} \right) \left( \frac{k}{k - 1} \right)$$

While the smaller particle size improve efficiency (N), retention (k) depend on duration of the whole analysis. However, selectivity has the greatest impact on resolution. Selectivity can be explained as the distance between adjacent peaks and it increases with increasing distance.

Column efficiency is determined by the plate count (N) and the theoretical plate height (H).

$$N = \frac{L}{H}$$

L – length of the separation column

H – theoretical plate height

Van Deemter equation describes the relationship between the theoretical plate height and linear velocity.

$$H = A + \frac{B}{u} + Cu$$

u – linear velocity of the mobile phase

A – Eddy diffusion

B – longitudinal diffusion coefficient

C – resistance to mass transfer in the stationary (Cs) and mobile (Cm) phases

Eddy diffusion term (A) describes different paths analyte can have while passing through the column and it depends on the size of particles and their uniformity.

Longitudinal diffusion term (B/u) depends on molecular diffusion in the axial direction. Cu describes the mass transfer time needed to establish the equilibrium between mobile and stationary phase.

The column efficiency increases with lower value of theoretical plate height. Higher values of theoretical plate height indicated longer time to establish the equilibrium and, therefore, broader chromatographic peaks. [15]

## Liquid Chromatography – Mass Spectrometry (LC-MS) for peptides

Peptides and proteins are usually analyzed by reversed-phase HPLC with mass spectrometric detection. Specificity and selectivity of mass spectrometry lower requirements for separation selectivity but put constraints on method development by limiting the number of workable reagents.

HPLC delivers the analyte in the liquid form but mass spectrometers analyze ions in the gas phase. One of techniques that can bridge the gap and transfer peptides to the gas phase without degradation is electrospray ionization (ESI). This technique is affected by the surface charge and tension, the size of the liquid drop delivered from the HPLC system, solvent volatility and solvation strength. In turn, these parameters are affected by the internal diameter and the length of the column and the post-column tubing, mobile phase composition, pH, modifiers, etc. [16]

Concentration of the analyte reaching ESI at the certain moment depends on the internal diameter of the HPLC column. Analytes are less diluted in columns with smaller diameter thus yielding higher signal on the detector. Typical internal diameters are between 180 and 360  $\mu\text{m}$  for flow rates of 1 to 10  $\mu\text{L}/\text{min}$ .

However, after the analyte exits the separation column it passes through the connecting tubing until it reaches the ESI probe. During that time dispersion of the analyte occurs leading to the peak broadening and decreased sensitivity. For that reason the post-column tubing should be minimized.

Composition and properties of the mobile phase have a major impact on ionization efficiency. The usual composition of the mobile phase for HPLC systems with mass spectrometry detection is a mix of water and organic modifier, such as methanol or acetonitrile, as well as ionic modifier as formic acid ( $\text{pK}_a = 3.77$ ) that is a preferred ionic modifier for LC-MS systems. Ionic modifier is added to reversed-phase HPLC as an ion pairing agent that increases retention time, prevents ionization of carboxyl groups and protonates amine and silanol groups. [13, 17]



The purpose of the organic modifier is to compete for the analyte against the stationary phase. Most commonly used organic modifier is acetonitrile due to lower surface tension of the mixture resulting in better electrospray ionization.

Peptides can carry one or more charges, on carboxylate (-COOH) and amine (-NH<sub>2</sub>) groups as well as on amino acid residues. Number of charges depends on the pH of the mobile phase and pK<sub>a</sub> value of each residue.

### Strong Cation Exchange Chromatography

As stated before, peptides can carry one or more charges, on carboxylate (-COOH) and amine (-NH<sub>2</sub>) groups as well as on amino acid residues. The overall charge of the peptide depends on its amino acid sequence. Ion exchange chromatography separates peptide on the type of their overall charge (positive or negative) and on the relative charge strength.

The premise of ion exchange chromatography is that ions can be exchanged with ions of the same type regardless of their mass. Ions from the solution reversibly bind to ions of the opposite charge that are bound to the stationary phase called ion exchanger. Desorption of the analyte depends upon ionic strength of the mobile phase. Analytes with higher net charge require higher ionic strength to desorb. [18]

Ion exchange stationary phases are classified in two main groups: cation (separate cations) and anion (separate anions) exchangers. Cation exchangers utilize acidic groups. If those groups are derived from the strong acid such as sulfonic acid, they are called strong cation exchangers. [4]

One of the factors affecting the elution is exchange capacity, or the concentration of ion exchange groups on the surface of the particle.

Another factor to consider is the pH of the mobile phase. The analyte will not be retained on the column unless it is ionized at the particular pH. Also, if the ion exchanger is neutralized that site is no longer available for binding.

The nature and concentration of the salt controls its eluting strength. NH<sub>4</sub><sup>+</sup> is commonly used in strong cation exchange chromatography.

The separation of components of complex mixtures is achieved by utilizing salt gradients that gradually increase ionic strength of the mobile phase thus eluting components in the order of increasing net charge (Figure 2.2). [4, 18]

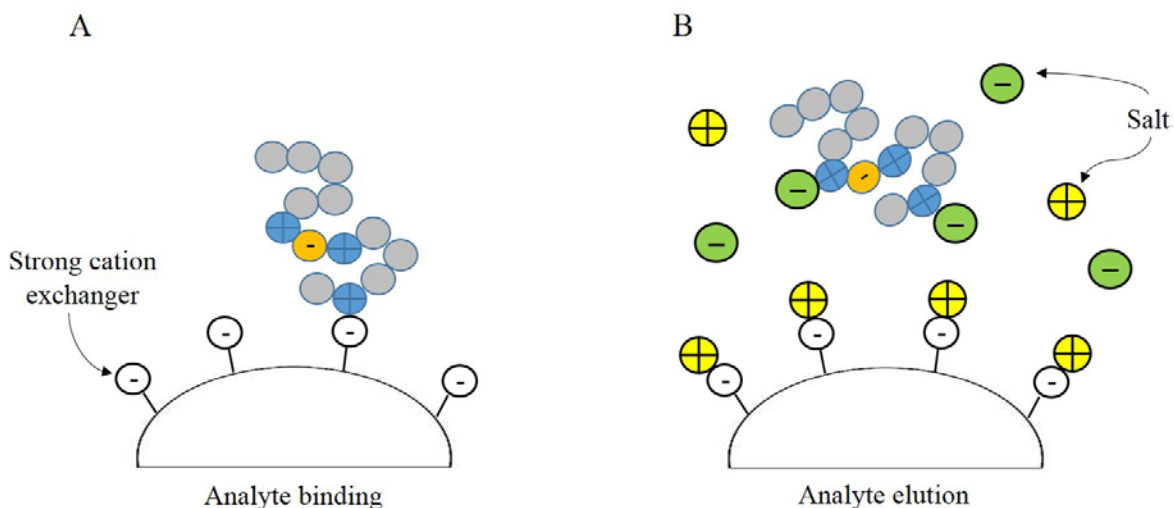


Figure 2.2. Principles of the strong cation exchange liquid chromatography. A. Analyte bound to the stationary phase. B. Analyte eluted from the stationary phase due to presence of ions in the solution.

### Multidimensional Protein Identification Technology (MuDPIT)

Regardless of the species, the whole cell lysate contains thousands of proteins. This complex mixture is further complicated utilizing digestion to produce peptides required for mass spectrometric analysis.

On the other hand, mass spectrometry can analyze a limited number of peptides at any given moment. Additionally, limited duty cycle of mass spectrometers may result in undersampling of complex proteomic samples. Low abundant peptides are at higher risk of being lost considering that only three most abundant peptides undergo fragmentation.

For all these reasons, complex samples like whole cell lysate have to be separated in multiple less complex samples before mass spectrometric analysis to increase the resolution and peak capacity that are crucial to the success of the analysis.

Solution lies in the reduction of sample complexity which is achieved by sample purification and fractionation. If the experiment aims to study a subset of proteins, the complexity of the sample can be attained by removal of high abundance proteins utilizing, for example, affinity chromatography. Fractionation of the sample based on a certain physico-chemical property and performing multiple HPLC runs further decreases sample complexity and reduces the number of coeluting peptides.

The most commonly used fractionation method is multidimensional protein identification technique (MuDPIT). The strength of this method lies in combination of two orthogonal types of separation (Figure 2.3). First, peptides are separated based on differences in their accessible surface charges utilizing strong cation exchange (SCX) chromatography. Peptides in each subgroup are then separated based on differences in their hydrophobicity utilizing reversed-phase HPLC chromatography.[19]

For MuDPIT analysis strong cation exchange resin is packed in tandem to reversed-phase C<sub>18</sub> resin. Separation starts when strong cation exchange resin binds positively charged peptides. Peptides are eluted off the strong cation exchange column by increasing ionic strength of the mobile phase in a stepwise manner from low to high salt concentration. A portion of peptides is released in each salt step depending on the salt concentration. Those peptides proceed to reversed-phase C<sub>18</sub> where they bind based on their hydrophobicity. Using organic gradient, peptides are eluted off the C<sub>18</sub> resin and continue toward the mass spectrometer. Once the C<sub>18</sub> resin is equilibrated back to high aqueous conditions, the next cycle starts with the higher salt concentration. [20-23]

This method is customizable in terms that complexity of the sample dictates the number of the salt steps and the slope of the organic gradient.

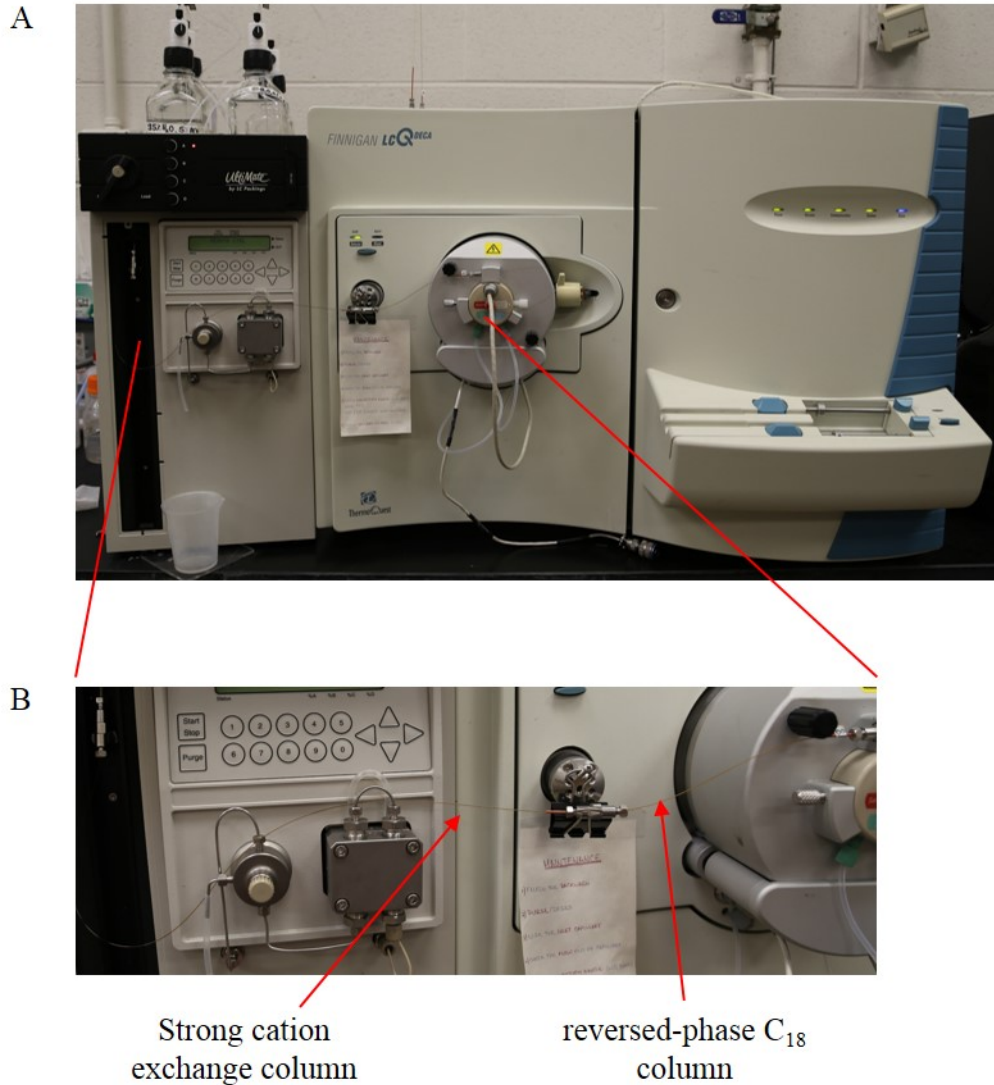


Figure 2.3. A. Finnigan LCQ Deca ion trap mass spectrometer coupled with LC Packings Ultimate 47 quaternary capillary LC system. B. Strong cation exchange and reversed-phase C<sub>18</sub> columns in series.

Separation can be done online as well as offline. In offline version, elutions from strong cation exchange resin are collected and then injected separately. However, the online version prevents sample loss rendering increased sensitivity and greater dynamic range of MuDPIT analysis than some other methods, such as an offline ion exchange approach.

However, regardless of how well is the sample separated or purified there will always be a portion of the proteome such as low-abundant proteins, proteins with extreme

pI and molecular weight and membrane proteins that will be lost during sample processing. [22][23]

### Electrospray Ionization - Ion Trap Mass Spectrometry

Peptides are polar, nonvolatile and thermally unstable species that are eluted from the liquid chromatography system in the solution, and hence incompatible with the mass spectrometry analysis, that requires ions in the gas phase. At first glance, mass spectrometers would not be as suitable detectors for HPLC technique as they are for gas chromatography. However, an appropriate sample introduction technique can bridge the gap. One of techniques that transfers peptides to the gas phase without degradation is electrospray ionization (ESI) particularly useful for the analysis of polar compounds. [24, 25]

### Electrospray ionization (ESI) – Making Droplets

Analyte is eluted from the HPLC column in a liquid mobile phase and must be converted into gas phase ions that can then be analyzed by the mass spectrometer. Ionization of the analyte occurs in the solution within the charged mobile phase and the next challenge is how to transfer it to the gas phase. Electrospray ionization (ESI) is particularly popular because the process of gas-phase ion formation takes place at the atmospheric pressure and allows ion formation directly from the solution. Since peptides are not fragmented during this process ESI is considered to be a “soft ionization” technique. This technique primarily generates molecular ions ( $M^+$ ), protonated molecules ( $[M + H]^+$ ), simple adduct ions ( $[M + Na]^+$ ) and ions with neutral losses ( $[M + H - H_2O]^+$ ).

The electrospray ionization process occurs at the ESI probe whose purpose is to produce charged aerosol droplets that contain sample ions. The whole process involves

three steps: 1) nebulization of sample droplets into smaller charged droplets, 2) desolvation of droplets and 3) ion ejection.

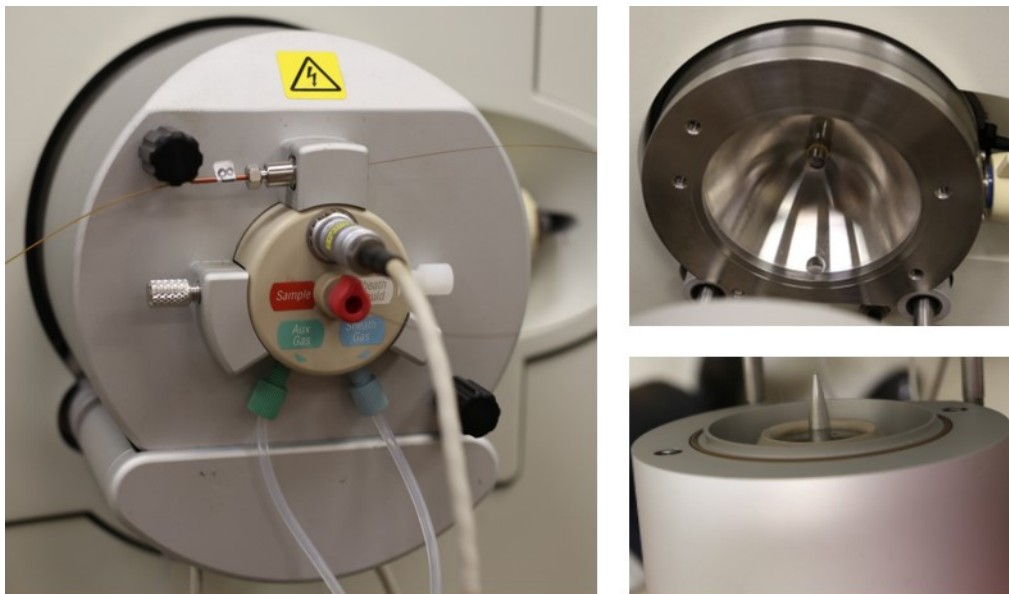


Figure 2.4. ESI chamber, Finnigan LCQ Deca ion trap mass spectrometer.

HPLC column is directly attached to ESI probe (Figure 2.4). The sample is fed in to a fused silica capillary that is, together with spray needle and spray nozzle, part of the ESI probe. High positive voltage (+500 V to +4.5 kV) applied to the tip of the spray needle induces charge separation within the drop by pulling positive charge towards the liquid front and forming the Taylor cone. Negative charge is removed by discharge against the metal wall of the spray needle. [26] A small highly positively charged droplet leaves the surface of the liquid when electrostatic repulsion overcomes surface tension. Opposite from the ESI probe is a heated capillary. Potential between the spray needle and heated capillary drives the droplet through the surrounding gas toward the entrance to mass analyzer through ion optics. Assisted by nitrogen gas, applied through the spray nozzle, solvent molecules leave the droplet that in turn decreases in size. According to ion evaporation model (IEM) the charge density at the droplet surface increases until it reaches about 80% of the Rayleigh limit when electrostatic repulsion becomes stronger than the surface tension and the droplet undergoes Coulomb explosion into smaller droplets. [27-30] This process continues until droplets are so small that ions desorb into

the gas phase and are sampled by the mass analyzer. After ejection and before they reach the mass analyzer, ions can undergo some gas-phase reactions like proton transfer and charge exchange. [31]

Several features affect the ESI process like solvent volatility, droplet size and liquid surface tension. Smaller droplets are formed with lower flow rates which is achieved with splitter. Another factor influencing qualitative and quantitative performance of mass spectrometer is variability of analyte response known as ion suppression. The main cause of this process is presence of less volatile compounds in a solution phase that can change the droplet solution properties. But other processes as precipitation and ion pairing can also be responsible for ion suppression. All these in turn affect the droplet formation and evaporation thus affecting the amount of charged ions reaching the detector. [17, 32]

Also, pH of the mobile phase should be chosen so it keeps the analyte charged. Mobile phases with low pH are utilized for the analysis of basic analytes in the positive electrospray mode. Weak acid, like formic, is added to adjust the pH while salts, strong acids and bases are very detrimental due to solid deposits. The concentration of the added acid should be low to avoid the competition with the analyte for the proton because that can also decrease analyte response. [33]

### The Path of Ion

Droplets created in ESI process travel toward the heated capillary that further helps desolvation process. Temperature of the heated capillary can be anywhere from 120 °C to 350 °C depending on the flow rate, type of the analyte and the mobile phase. Positive ions are transported through the heated capillary utilizing pressure gradient from the atmospheric region to capillary-skimmer region (~ 1 torr) assisted by a potential of 0 to +10 V and enter the tube lens. The purpose of the tube lens is to focus ions and stop the flow of ions into the mass analyzer during detection. Potential applied to the tube lens focuses ions toward the opening of the skimmer. Tube lens offset voltage is additional potential (0 to +40 V) applied during ion collection that accelerates ions that

collide with the background gas further helping with desolvation process and thus increasing sensitivity. However, if collisions are too energetic sensitivity decreases due to fragmentation (ion source collision induced dissociation). Tuning procedure optimizes sensitivity of the tube lens offset voltage. During detection of ions that already entered mass analyzer, the voltage applied to the tube lens is -200 V thus deflecting all positive ions away from the opening of the analyzer. [33]

Passing the skimmer, ions enter lower pressure ion optics region at  $10^{-3}$  torr on their way to mass analyzer. Ion optics have three parts: quadrupole, octapole and interoctapole lens. Electric field made of RF (2.45 MHz, 400 V) and dc offset voltage (-10 to +10 V) applied in this region guides and focuses ions that are now entering the vacuum region of the mass analyzer. [33]

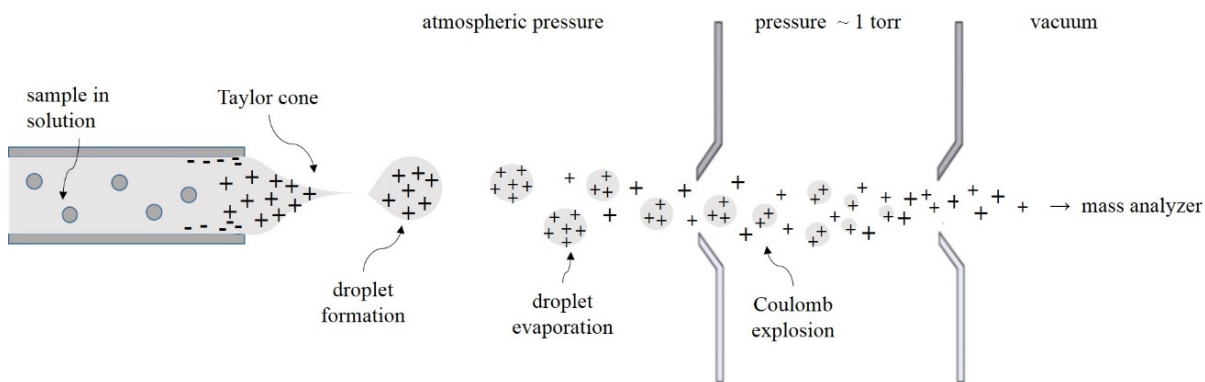


Figure 2.5. Electrospray ionization (ESI) process.

### Mass Analyzer - Ion Trap

The heart of every mass spectrometer is mass analyzer where ions are separated based on their mass to charge ratios ( $m/z$ ) and sent to the detector. There are four general types of mass analyzers: quadrupole, time of flight, sector and ion trap. Our instrument is equipped with ion trap that will be the focus of further discussion.

An ion trap consists of three stainless steel electrodes, two end-caps and one ring electrode, of hyperbolical geometry (Figure 2.6). They together form the cavity with dimensions of  $r_0$  and  $z_0$  as shown in the picture. Both end caps have holes in their centers



where ions enter the cavity of the ion trap and exit through the opposite end cap after they have been separated according to their mass to charge ratio. [31]



Figure 2.6. Ion trap elements: two end-caps and one ring electrode.

As the name says, an ion trap mass analyzer traps ions in a three dimensional electric field. Ions are stored, separated according to their mass to charge ratio and then selectively ejected which increases the sensitivity of the device. [34]

Gating lens controls the entry of the ions into the trap. Negative potential is applied to it to enable positive ions to enter the trap. The time of ion accumulation is optimized to allow for maximal signal and minimal space charge. Excess ions in the trap are detrimental due to changes in electrical fields they cause, while too few ions lead to loss of sensitivity. Once the optimal amount of ions entered the trap the potential on the gating lens is changed to positive to prevent further entrance of positive ions. Ion trap can analyze negative ions as well. In that case the potential applied to the gating lens is reversed. [35]

Inside the ion trap RF oscillating potential is applied to the ring electrode and focuses ions to the center of the trap. Ions move in complex pattern and high kinetic energy. Damping gas (helium, 1 mTorr) is used to prevent the premature ejection by colliding with the ions and reducing their kinetic energy. This results in focusing ions toward the center of the trap. [33, 35]

Potential applied to the ions in the ion trap is described by the following equation:

$$\Phi_{r,z} = \frac{(U - V \cos \Omega t)}{2} \left[ \frac{r^2 - 2z^2}{r_0^2} \right] + \frac{(U - V \cos \Omega t)}{2}$$

U – dc potential  
 V – Rf amplitude applied to ring electrode  
 $\Omega$  – angular frequency of RF (V), typically 1.1 MHz  
 $r_0$  – internal radius the ring electrode  
 r – distance between the ion and the ring electrode  
 z – distance between the ion and the end-cap electrode

Solutions are given by the second order differential equation named the Mathieu equation. Ion stability in the ion trap is described with two dimensionless Mathieu parameters  $a_z$  and  $q_z$ ,

$$q_z = \frac{2eV}{mr_0^2\Omega^2}$$

$$a_z = \frac{4eU}{mr_0^2\Omega^2}$$

Dimensions of the ion trap, angular frequency of RF potential, mass and charge of the ions are constant values. Therefore, dc potential (U) and RF amplitude (V) control ion's motion described by the solutions of Mathieu's equation or graphically presented with the plot of  $a_z$  vs.  $q_z$  named Mathieu stability diagram (Figure 2.7). [36] There are two types of solutions of Mathieu's equation:

- a) Periodic and unstable – describe boundaries of unstable region of the stability diagram,
- b) Periodic and stable – describe motion of ions within the ion trap.

Ions are trapped if the intersection of  $a_z$  and  $q_z$  falls within boundaries of the stability diagram. However, if the intersection value falls outside of boundaries of the stability diagram, the ion is ejected.

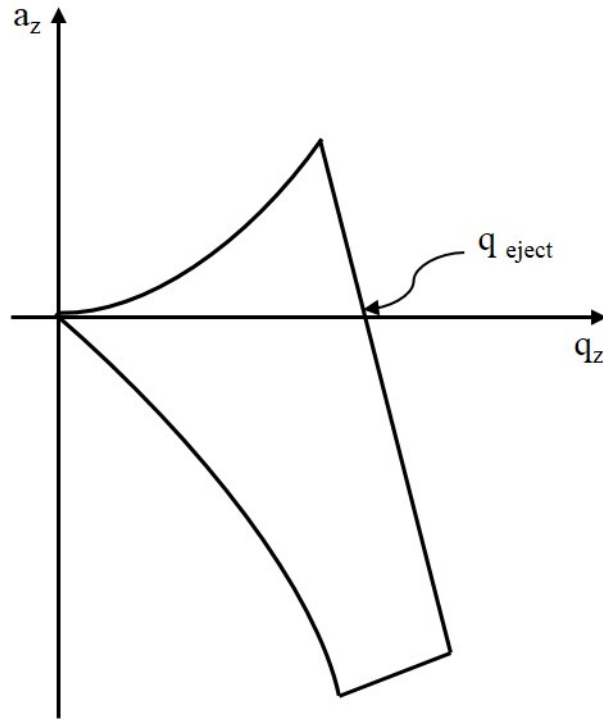


Figure 2.7. Mathieu stability diagram for a 3D quadrupole ion trap

In commercial ion traps dc voltage is constant and set to give radial stability to all ions. Their stability and ejection is controlled in the axial direction.

When dc voltage is set to zero stability diagram collapses to  $q_z$  axis and ion trap operates in the “mass selective instability scan” mode. Rising RF potential applied to the ring electrode moves the ions along the  $q_z$  axis. Stable ion trajectories have the  $q_z$  value between 0 and 0.908. As the  $q_z$  value approaches 0.908, ions gain the kinetic energy, reach the boundary of the stability diagram, become unstable and are ejected from the ion trap through holes in the end-cap. This is the low-mass cutoff limit. Mass selective instability scan mode of operation generates the full scan mass spectrum by sequentially ejecting ions from low  $m/z$  to high  $m/z$ . [31, 35] However, this approach does not allow the selection of ions.

Resonance ejection is one of the methods for ion selection that induces instability of a particular ion utilizing ac voltage of high amplitude or supplementary RF potential on the end-caps that changes secular frequency of the ion resulting in its ejection.

An alternative method is axial modulation where RF frequency is constant but the amplitude is ramped which can selectively eject several or all ions.

After ejection, ions are accelerated by the conversion dynode and sent to the channel electron multiplier for detection. [31] The number of ions reaching the detector depend on ESI efficiency, ion sampling efficiency into the vacuum and ion transmission efficiency through ion optics and the ion trap. [27]

### Tandem Mass Spectrometry

Tandem mass spectrometry (MS/MS) is a multiple stage mass spectrometry that acquires data in a data dependent mode. One of the advantages of an ion trap mass analyzer is the ability to perform tandem mass spectrometry in time without the need for an additional mass analyzer.

The first step in data-dependent mode is to acquire a full scan mass spectra. When one of the ions reaches signal threshold, it is isolated by applying the appropriate RF wave form that will eject all other ions. Then, “tickle” potential is applied across the end-caps to resonantly excite the ion of interest. This is similar to axial modulation. However, the amplitude of “tickle” potential is chosen in such manner not to eject the ion but to pull it away from the center of the trap. Ion gains energy from the RF drive potential and collides with helium experiencing collision induced dissociation (CID). Resulting fragments are analyzed by increasing amplitude of RF potential and ejected from low  $m/z$  to high  $m/z$ . After obtaining the MS/MS spectrum analyzed ion is put on the exclusion list and not analyzed during selected time frame regardless of the signal intensity. During that time the ion trap fills again with all ions and instrument analyzes other ions that reached signal threshold.

To enhance the statistics three micro-scans are averaged to generate the data dependent full scan spectrum. Three most intense ions are then subjected to tandem mass spectrometry with five spectra averaged to produce one MS/MS spectrum.

CID of the peptide cleaves the amide bond and generates fragments characteristic of the specific amino acid sequence (Figure 2.8). Charge can be retained on N-terminus yielding a, b and c ions or on C-terminus yielding x, y and z.[37] However, b and y ions are most abundant when low energy CID is employed.

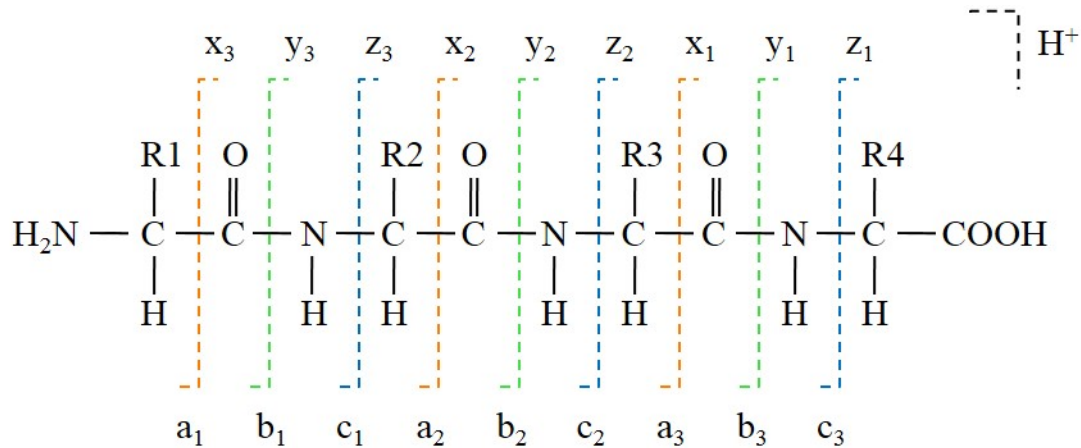


Figure 2.8. Peptide fragmentation notation.

The first step in proteomic data processing is to assign the spectrum of a fragment to a primary protein sequence. To accomplish that an overwhelming number of computational approaches and tools have been developed. [38] In general, experimental mass values are compared with calculated peptide masses and fragment ion mass values. Calculated masses are obtained when the set of cleavage rules is applied to the protein primary sequence database. Scoring algorithms are used to identify the closest matches thus identifying the unknown protein from the sample. Considering proteomes of different species are not yet fully elucidated if the sequence database does not contain the unknown protein then the algorithm matches it to the protein of the closest homology. [39]

Mascot is a powerful search engine that provides search method for MS/MS ion search, and therefore, is utilized for the purpose of qualitative proteomics.

Mascot search engine converts MS/MS spectra into mascot generic format (MGF), which is a list of pairs of mass and intensity values. MGF files are searched against the database of calculated peptide or fragment ion masses. [40]

A set of search parameters, such as choice of proteolytic enzyme, number of missed cleavages, peptide mass tolerance, choice of search masses and protein molecular weight, provide versatility and the ability to identify proteins with maximum discrimination and the highest score.

Mascot search engine is based on Mowse algorithm but it also adds probability-based algorithm. [41] Proteins are identified based on the probability (P) that the match between the experimental data and theoretical value is a random event giving the lowest probability to the best match. The score is calculated using the equation below.

$$\text{score} = -10 \times \log_{10} P$$

These values are converted into Mowse scores assigning the highest score to the best match. [42]

Mowse (MOlecular Weight SEarch) method identifies the proteins based on molecular weight of their respective peptides measured by mass spectrometry. To achieve that, experimental data are first compared with calculated peptide masses in the sequence database and only entries that fall within a given mass tolerance are accepted as a match. Each match is assigned a statistical weight that is determined utilizing empirical factors such as frequency (F).

A frequency is calculated using matrices where intervals of 100 Da of peptide mass were put in the rows (i) and intervals of 10 kDa of intact protein mass were put in columns (j). Each experimentally obtained ion falls into the appropriate matrix element  $f_{i,j}$ . The frequency of occurrence gives the size distribution of peptide masses as a function of protein mass. The frequency is normalized by dividing the elements of each 10 kDa column by the largest value in that column. This process gives the factor m according to the equation below.

$$m_{i,j} = \frac{f_{i,j}}{f_{i,j}_{\max \text{ in column } j}}$$

Factors  $m$  for each entry are then multiplied ( $\prod_n m_{i,j}$ ) and used together with the molecular weight of the entry ( $MW_{\text{prot}}$ ) to calculate the score utilizing the following equation

$$\text{score} = \frac{50,000}{MW_{\text{prot}} \times \prod_n m_{i,j}}$$

## CHAPTER 3

### SAMPLE PREPARATION FOR MASS SPECTROMETRY OF GLYCOPROTEINS

Following translation, the next step in protein biosynthesis are various posttranslational modifications (PTM) such as glycosylation, methylation, phosphorylation and many others. Considering the critical role of PTMs as regulators of protein activity and function it is essential to expand the current body of knowledge on this topic. However, very low levels of PTMs within complex biological system poses an impediment to researchers. Therefore, development of new enrichment techniques are necessary to prevail in this challenge.

The purpose of this chapter of dissertation is to describe the effort put to develop the new technique for enrichment of glycoproteins from the complex matrix such as whole cell lysate.

#### Posttranslational modifications of proteins

Posttranslational modifications (PTM) are ubiquitous and denote the attachment of various functional groups or structural changes of the proteins after translation further diversifying protein structure beyond what is acquired by gene transcription. These modifications play a key role in many biological events like gene expression, signal transduction, protein-protein interaction and many more because they regulate protein activity by changing its surface or introducing new functional groups. In turn, this changes thermodynamic and kinetic features of proteins as folding rate and stability. [43, 44]

There are many PTMs such as glycosylation, phosphorylation, methylation, ubiquitination, lipidation, nitrosylation, acetylation and proteolysis. Glycosylation has been recognized as one of the major PTMs. [44] Since glycoproteins are the central theme of this dissertation glycosylation is discussed in more details.



## Protein Glycosylation

In general, glycosylation is the process of covalent binding of glycans and proteins catalyzed by a series of enzymes. Glycosylation can be discussed from the perspective of protein or of glycan. From the protein perspective, benefits of glycosylation are enhanced thermal stability and folding kinetics of the protein, increased solubility and resistance to proteolysis. Also, the process of glycosylation itself depends on the number, location and surroundings of the glycosylation sites on the protein. [45] From the glycan perspective, properties of the protein depend on the size of the glycan, its flexibility and structure. [43, 45]

The formation of the bond between a saccharide and the amino acid is the first step in biosynthesis of glycan parts of glycoproteins. This bond initiates a series of enzymatic steps that yield an elaborate oligosaccharide structures on the protein. It is known up today that 13 monosaccharides and 8 amino acids are involved in at least 41 glycoprotein linkages divided in the following five types of glycosylation:

1. O-linked glycosylation – through hydroxyl oxygen of serine, threonine or tyrosine,
2. N-linked glycosylation – through nitrogen of asparagine side chain if in the sequence N-X-(S/T) as standard consensus sequence or N-X-C, N-G and N-X-V as non-standard consensus sequences, where X cannot be proline,
3. C-linked mannosylation – mannose linked to the tryptophan side chain if in the W-X-X-W consensus sequence,
4. glypiation – through glycosylphosphatidylinositol (GPI) anchor that links proteins and lipids.
5. phosphoglycosylation – mannose, fucose, xylose or N-acetyl-D-glucosamine linked to serine. [46]

Sites of glycosylation vary throughout the cell and include Golgi complex, endoplasmatic reticulum, nucleus and cytosol. [47] However, the most abundant glycosylations in eukaryotic proteins are N-linked and O-linked protein glycosylations and they take place in cytoplasm, endoplasmic reticulum (ER) and the Golgi complex.

Large part of the proteome of the cell is synthesized and folded in endoplasmic reticulum (ER) and then transported to the Golgi complex following a conventional ER–Golgi secretory route. One of the key functions of the ER is production of glycans. Mannose rich glycans are synthesized on the cytosolic face of the ER. These precursors are then moved into the lumen of ER where they are further branched by adding more mannose and glucose. While within ER precursor glycans are transferred to the asparagine residue onto still unfolded protein by glycosyltransferases thus forming N-linked precursor glycans. Once they reach Golgi complex, glycans are edited by glycosidases to form mature N-linked glycans. O-linked glycosylation happens during passage through the Golgi complex. [44, 48-52]

After translation, newly synthesized proteins are transported from endoplasmic reticulum to Golgi complex. Along this path they are modified by the attachment of variety of different functional groups such as carbohydrates and lipids. These groups change the structure and properties of proteins reversibly or irreversibly. [44, 47]

Golgi complex and endoplasmic reticulum are not the only places where posttranslational modifications happen, however, those are sites of glycosylation.

## Glycans

According to the IUPAC Gold book, glycans are defined as "compounds consisting of a large number of monosaccharides linked glycosidically". [53]

Glycosidic bond is the bond between the hemiacetal group of the saccharide and hydroxyl or amine group of another organic molecule (Figure 3.1).

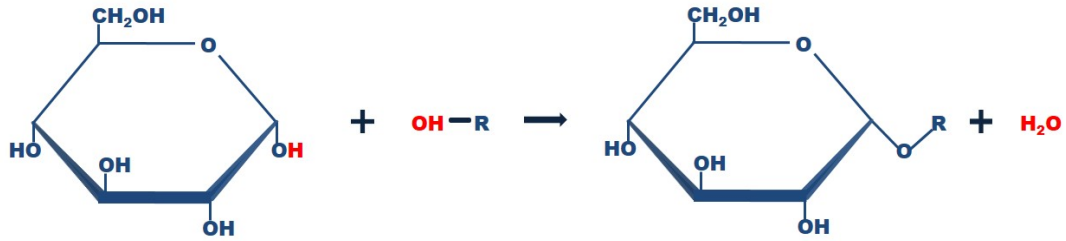
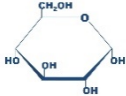

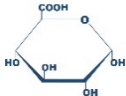

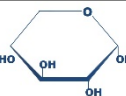

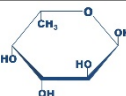

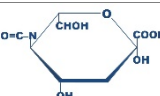

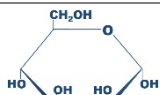

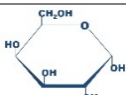

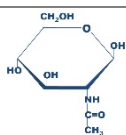

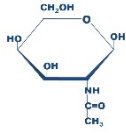



Figure 3.1. Example of  $\alpha$ -glycosidic bond.

Glycans are built from 9 different monosaccharides joined together with glycosidic bond to form oligosaccharides (2 – 20 monosaccharides) or polysaccharides ( $20 <$  monosaccharides). Structures of polysaccharides are more complex than proteins since they can form various topologies. They contribute to protein diversity not only between different organisms but different glycans can also be present on one type of the protein. Even microheterogeneous glycans that differ in the structure of only one or more of its monosaccharides can greatly alter thermodynamic properties of a certain protein.

[54]

Table 3.1. Glycan building monosaccharides.

Monosaccharide	Haworth projection	Abbreviation	Symbol
D-glucose		Glc	
D-Glucuronic acid		GlcA	
D-Xylose		Xyl	
L-Fucose		Fuc	
N-acetylneuraminic acid		NeuAc	
D-Mannose		Man	
D-Galactose		Gal	
N-acetyl-D-glucosamine		GlcNAc	
N-acetyl-D-galactosamine		GalNAc	

Both N- and O- linkages are through N-acetylglucosamine. However, N-linked glycans are bonded to asparagine, while O-linked glycans to serine or threonine. Unlike O-linked glycans, N-linked glycans have standard and non-standard consensus sequence. The standard consensus sequence involves three amino acids. It starts with asparagine

linked to any amino acid but proline and ends with serine or threonine (N-X-(S/T)). Additionally, there are two non-standard consensus sequences: asparagine – X – cysteine (N-X-C) and asparagine – X – valine (N-X-V) where X cannot be proline also. [46]

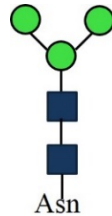


Figure 3.2. The core sequence of N-linked glycans.

All N-linked glycans have the same core sequence as shown in the (Figure 3.2). They are divided into three classes according to the monosaccharides attached to the core structure:

1. Oligomannose – only mannose residues attached,
2. Complex – each branch starts with N-acetylglucosamine (GlcNAc) and develops further,
3. Hybrid – only mannose residues are attached to one of the core's branches and the variety of monosaccharides on the other. [55]

O-linked glycans are covalently attached to hydroxyl group of serine or threonine by an O-glycosidic bond. There are several types of O-linked glycans:

1. O-N-acetylgalactosamine (O-GalNAc) – starts with N-acetylgalactosamine followed by galactose, N-acetylglucosamine, fucose, or sialic acid, but not mannose, glucose, or xylose,
2. O-fucose – that can be elongated to a tetrasaccharide,
3. O-glucose – typically exists as trisaccharide comprised of glucose and two xyloses, although it is also found as a monosaccharide,
4. O-N-acetylglucosamine (O-GlcNAc) – starts with N-acetylglucosamine,
5. O-mannose – very complex structures and common to both prokaryotes and eukaryotes. [56]

## Lectins

Lectins are saccharide binding proteins ubiquitous across all species. They also may have saccharides on their surface, which classifies them as glycoproteins, but what defines them as lectins is their ability to bind saccharides.

Lectins' major role is to facilitate cell to cell contact. This contact is provided by interaction of lectin displayed on the surface of one cell with saccharide displayed on the surface of another cell. A molecule of lectin usually has more than one saccharide binding sites. Each site forms a weak bond with the saccharide, however, synergistic effect of multiple binding sites provides a strong interaction. [51]

There are two types of interactions between lectin and saccharide. First type is when lectin with two binding sites binds two ligands. These complexes are linear and flexible and ligands can vary in structure. The second type is formed when either lectin or ligand has more than two binding sites yielding complex, cross-linked structures. The ligand composition in this case has to be the same to be able to form the lattice. [30, 57-59]

Lectins can be used for the purpose of studying the glycoproteome for substantial reduction of the sample complexity. They are utilized for the extraction of glycoproteins from the whole cell lysate for the affinity chromatography technique.

Concanavalin A (Con A) is one of the most widely used lectins. Perhaps one of the reasons is its low affinity toward non-glycosylated proteins.

Con A was used for the development and conformation of the method described in this dissertation to extract glycoproteins from the whole cell lysate of HeLa cells (benchmark) as well as from whole cell lysate of tachyzoites and tissue cysts of *Toxoplasma gondii*. For that reason Con A is discussed in more details.

## Concanavalin A (Con A)

Concanavalin A (Con A) is a lectin isolated from jack bean *Canavalia ensiformis*. At pH 7.5 Con A is a tetramer but at the pH 5 and lower it reversibly dissociates into dimers. Each subunit is comprised of 237 amino acids and has a molecular weight of 26 kDa. They fold in the formation of two  $\beta$  sheets that create  $\beta$  sandwich. [60]

Also, each subunit has one saccharide binding site and two metal binding sites: one for a transition metal ion and one for calcium ion. If metal ions are removed in acidic conditions Con A can no longer bind saccharides. However, replenishing metal ions at the pH 7 restores saccharide-binding activity. This proves that the role of metal ions is to pull the amino acid residues into proper conformation required for binding. [61-63]

ConA binds  $\alpha$ -mannose and  $\alpha$ -glucose sugars and N-glycans but it does not bind O-glycans on animal cell glycoproteins. The binding affinity is much higher for oligomannose-type N-glycans than complex-type biantennary N-glycans, while it does not bind highly branched complex-type N-glycans. Examples of glycans that Con A binds are shown in the Figure 3.3. Circled portions of glycans are required for binding. [64]

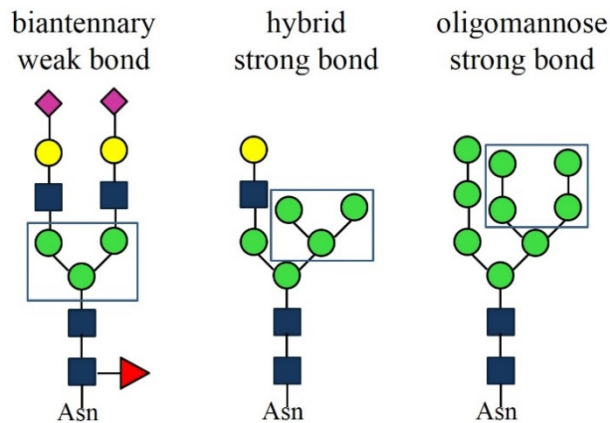


Figure 3.3. Example of glycans that Con A binds. The parts of the glycan required for binding are in the box.

Once Con A binds a glycan and after the purification procedure, it is important to select a competitive inhibitor that will provide elution of the glycoproteins. Therefore, it is important to understand the mechanism of binding.

Con A binds the saccharide by forming hydrogen bonds with almost all hydroxyl groups on the saccharide. The strongest bond is formed with D-mannose, then D-fructose and lastly D-glucose which still contains at least a part of configuration required for binding.

Con A binds the saccharide in the chair conformation with C4 deepest in the binding site and C1 closest to the surface. Hydroxyl group on C1 is extending into the solvent and is not essential for binding with the protein, while the one on C2 has some influence on binding which is proved by the higher affinity of binding of D-mannose than D-glucose. It seems that in D-mannose C2 hydroxyl group better approaches the surface of the protein. However, hydroxyl groups on C3, C4 and C6 stabilize the conformation through hydrogen bonding with the protein and are essential for binding.

The type of the bond and the substituents also play a role in the binding affinity. Con A has much higher affinity for  $\alpha$  glycosidic bond than  $\beta$  glycosidic bond. Additionally, substituting hydrogen for the methyl group in  $\alpha$ -methyl glycosidic bond increases affinity several fold. As a consequence, methyl- $\alpha$ -D-mannopyranoside is the best binding inhibitor than any other saccharide, followed by methyl- $\alpha$ -D-glycopyranoside (Figure 3.4). [30, 64]

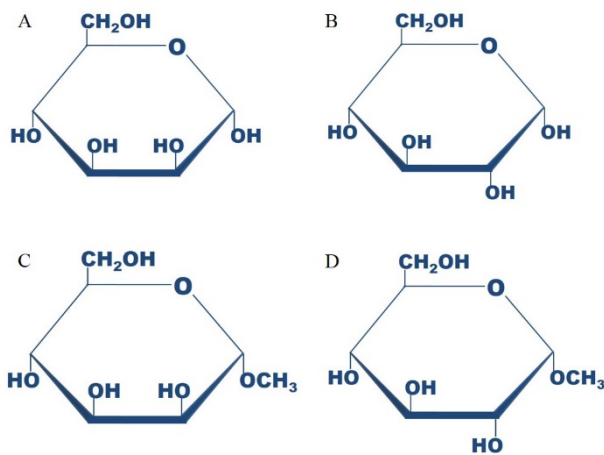


Figure 3.4. A. D-mannose, B. D-glucose, C. methyl- $\alpha$ -D-mannoside and D. methyl- $\alpha$ -D-glycoside, Haworth projection.

For the purpose of this project we utilized Con A for the separation of glycoproteins from the whole cell lysis.



## Lectin Affinity Chromatography

Affinity chromatography is based on specific interaction between ligand chemically bound to the solid support and the target molecule that is most often a biological molecule (Figure 3.5). Therefore, molecules are separated based on their biospecific binding to the immobilized ligand. The principle of affinity chromatography is shown in the figure below.

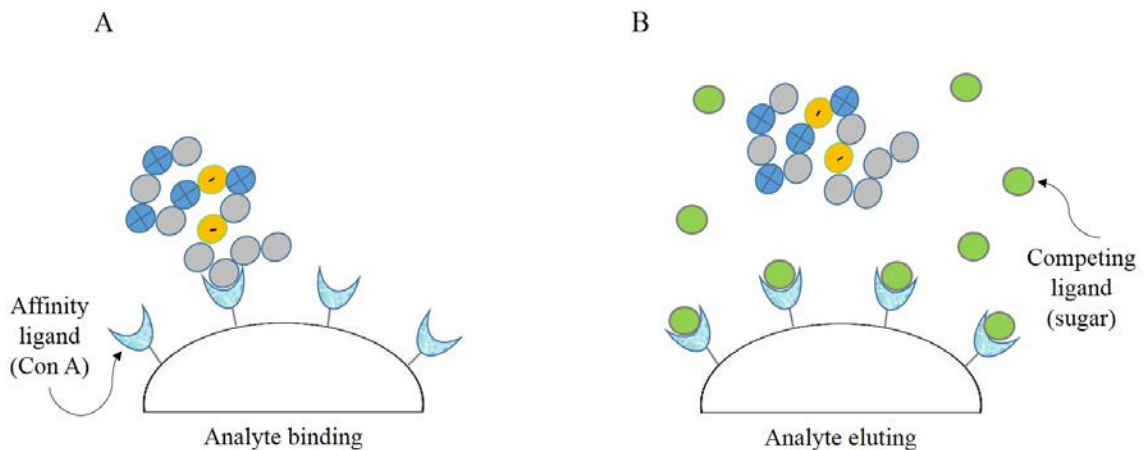


Figure 3.5. Affinity chromatography principle. A. Ligand binds the glycoprotein. B. Elution of the glycoprotein utilizing saccharide as a competing ligand.

Fundamental principles of binding are based on hydrophobicity and ion-ion interactions as for the other types of chromatography but also hydrogen bond formation. The difference is in biospecificity. There is usually more than one binding site on both ligand and target molecule but they are arranged in unique orientation so they fit as lock and key. Similar but less specific molecules can have structure that fits closely and they will form a weaker bond with the immobilized ligand. [65, 66]

Due to biospecificity, the process of protein separation and purification has to be optimized for each ligand and the target molecule but there are some general rules applicable to all affinity chromatography procedures. For example, there is a limited

number of buffer systems that can be used in order to preserve the binding properties of the immobilized ligand. Also, detergents, pH and salt concentration are crucial for successful binding event.

The outcome of affinity chromatography depends on several factors like ligand selectivity, ligand stability, recovery, etc. Ligand stability determines durability and versatility of the method. If, for example, the ligand is stable in different buffer systems and under the wide pH range, the method will allow for more options on sample preparation side. Recovery is determined by the amount of the target molecule efficiently eluted from the immobilized ligand. However, ligand selectivity is perhaps the most important factor that determines efficient separation and purification. The immobilized ligand is also a protein with its specific amino acid sequence folded in the specific 3D structure. The actual site of affinity binding is the binding pocket of that protein. Only the molecules that can at least loosely fit into the binding pocket will be retained and separated from the matrix.

#### Cell disruption and Protein Solubilization

In a biological system, proteins are found within the cell, at the cell membrane as well as in the extracellular space. To that end, cell disruption is a first step in analysis of the whole proteome inevitable to access all proteins. Cells can be disrupted chemically, utilizing buffers with various combinations of detergents, or mechanically, utilizing ultrasonication, homogenization or mechanical shearing.

Buffers, as radioimmunoprecipitation assay (RIPA) and nonidet P40, have been used for the lysis of *Toxoplasma gondii* and subsequent lectin pull-down. [66] However, those detergents pose the risk of decreased lectin viability or even denaturation. [67]

Alternative option is mechanical shearing of cells passing them through a narrow valve under high pressure in the buffer with no detergents added. This process disrupts the membrane of the cells but detergents are still required to release membrane- or cytoskeleton- bound proteins.

A natural surroundings of a native protein even in the whole cell lysate is a very complex matrix material composed of other proteins, lipids, carbohydrates, salts and other biological macromolecules. Also, proteins are often parts of protein complexes or integrated into organelles or plasma membranes and generally insoluble in their native state. Therefore, they have to be denatured to solubilize which means the disruption of all disulfide, hydrogen, dipole-dipole and van der Waals bonds as well as electrostatic and hydrophobic interactions. This is achieved by addition of salts, charged detergents, chaotropes, and reduction and alkylation reagents to lysis buffers. However, they are very different in their ability to disrupt cells and solubilize proteins and often are used together. Criteria to consider when choosing the lysis buffer: pH, ionic strength, detergents and denaturants, proteolysis preventers and protein stability enhancers. [68]

Ionic detergents, like SDS, contain a head group with either positive or negative charge and are useful for breaking protein-protein interactions. Micelle size is determined by combining the repulsion of the ionic head group with the hydrophobic attraction of the side chain. The size can increase if the charge on the head group is neutralized. Therefore, increasing ionic strength of the solution reduces critical micelle concentration (CMC), but temperature change has no effect.

Non-ionic detergents, like triton X-100, have uncharged hydrophilic group and are used for breaking lipid-protein or lipid-lipid interactions. They gently solubilize proteins, retaining their native structure and function. Unlike ionic detergent, critical micelle concentration of non-ionic detergent increases with increased temperature but ionic strength of the solution has no effect.

Zwitterionic detergents, like CHAPS, combine properties of both ionic and non-ionic detergents. They are best used for breaking protein-protein interactions. [69, 70]

For all of these reasons, when the downstream processing can tolerate the presence of detergents, the best approach is to utilize the strengths of both mechanical and chemical cell disruption.

Protein Purification Techniques

After elution from the Con A column, proteins are obtained in an elution buffer whose composition may not be compatible with the next step of the sample preparation. For that reason, two most widely used techniques for protein purification, gel electrophoresis and dialysis, are discussed below.

## Gel Electrophoresis

Ever since its introduction in late 1950' the polyacrylamide gel electrophoresis technique (PAGE) is the most widely used method and the golden standard for separation and analysis of protein mixtures as well as buffer exchange. The premise of this separation technique is that proteins with different molecular weights and the same charge will assume different electrophoretic motilities. However, proteins are vastly diverse with respect to charge as well as 3D structure. To accomplish separation based solely on molecular weight differences, they have to be denatured and saturated with the charge which is achieved utilizing sodium dodecyl sulfate (SDS).

As an anionic detergent, SDS binds with its hydrophobic tail to hydrophobic portions of a protein in a proportion of one SDS anion for every two amino acids. The specific mass ratio is 1.4:1. This yields an equal charge density per unit length and disrupts the three dimensional structure of the protein. The polar head of this detergent interacts with the hydrophilic solution and allows the protein to exist stably in solution in its extended conformation. The outcome of this interaction are denatured and highly negatively charged proteins whose migration is not determined by intrinsic charge but by molecular weight. [71] [69]

In the electric field, applied across the polyacrylamide gel, negatively charged SDS-protein complexes move from cathode to anode. Since they are introduced to the polyacrylamide gel from the cathode side they are forced to move through pores of the gel. The rate at which they travel through the gel depends on the strength of the field, size and shape of the SDS-protein complex. However, these are not the only determinants in this technique. Solvent, with its temperature and ionic strength, as well

as pH and the counterions with their shielding effect can greatly impact the separation. [71, 72]

Polyacrylamide gel is obtained by polymerization of the acrylamide monomer into long chains and crosslinked with N,N'-methylene-bis-acrylamide (bis) into three-dimensional structures. Polymerization starts with the addition of ammonium persulfate (APS) and the base N,N,N',N'-tetramethylethylenediamine (TEMED) that catalyzes the decomposition of persulfate ion to produce a free radical. This free radical then reacts with the acrylamide monomer transferring the charge to each subsequent monomer upon binding thus creating a linear polymer. The occasional bis-acrylamide crosslinks the linear polymers creating the sieve-like network whose pore sizes depend on concentrations of monomer and the crosslinker. [71, 72]

The final gel used for protein separation is a composite of two polyacrylamide gels, stacking and resolving that differ in the pore size and pH. Sample is loaded onto the stacking gel that has larger pore size and pH of 6.8. The purpose of this portion of the gel is to focus all proteins into an infinitely thin band before they enter the resolving gel. This process utilizes negatively charged chloride and glycine ions from the buffer whose pH is 8.3.

Chloride ions create the cloud around the proteins preventing proteins to see the field. As the chloride ions move through the gel fast, the proteins further from the entrance to the resolving gel lose chloride ions earlier than the proteins closer to the resolving gel and start to move toward the anode

Glycine is a weak acid and, therefore, exists as a zwitterion at low pH glycinate anion at high pH. Once the electric field is established, glycine moves from the buffer (pH 8.3) to the stacking gel (pH 6.8) towards the anode. The decrease of pH in the stacking gel causes the loss of a portion of the charge on glycine that, consequently, moves slower. The zone between chloride and glycine ions creates a high strength field that contains all proteins. This very thin band moves in this arrangement toward the resolving gel. Once the glycine reaches the resolving gel the increase of pH causes it to move faster and it outruns the proteins allowing them to see the electric field.

As the result of this process all proteins are focused at the interface of stacking and resolving portion of the gel providing separation within the resolving gel based only on their mass differences.

Once the negatively charged proteins reach the resolving gel they are driven by the electric field. However, smaller proteins move faster while the larger proteins have more difficulty moving through the pores of the gel. That is the basic principle of protein separation on the polyacrylamide gel. [71]

SDS polyacrylamide gel electrophoresis (SDS-PAGE) can be utilized for separation as well as purification and concentration. For separation purposes the gel is usually cast between two 7 x 8 cm glass plates separated by 0.5 to 1.5 mm spacer. Looking from the top of the gel the stacking portion is about 3 cm long followed by about 5 cm of the resolving gel where the separation takes place. However, the benefits of the protein stacking process can be used as a buffer exchanger to purify the sample of buffer components incompatible with the downstream analysis and to concentrate the proteins into finite band. This technique is called three-layer sandwich gel electrophoresis and requires different layout of the gel.

### Three-Layer Sandwich Gel Electrophoresis

The buffers required for sample preparation and extraction from the cell culture are incompatible with the downstream mass spectrometry analysis. Also, sample loss and dilution becomes a significant factor when multiple preparation steps are involved. Three-layer sandwich gel electrophoresis (TSGE) is developed with the purpose to concentrate the proteins and to serve as a buffer exchanger thus increasing the efficiency of the proteomic analysis.

TSGE protocol combines the properties of agarose and polyacrylamide gels. As the name says, the gel is comprised of three parts. On the bottom of the cylindrical cartridge is 40 % polyacrylamide gel (~ 0.5 cm) that serves as a sealing layer. It is covered with the 12 % polyacrylamide gel (~ 2 mm) where the proteins are concentrated.

The top layer is made of 2% agarose gel (~ 3 cm) that is used to immobilize the proteins from the sample.

The combination of pH throughout the gel plays a significant role in concentrating the proteins. The pH ranges from 8.3 in the running buffer through around 7 in agarose gel and 6.8 in the concentration layer to 11 in the sealing layer.

When the current is applied these pH changes influence glycine present in running buffer as well as all three layers of the gel. When glycinate ions enter the agarose gel from the running buffer, they lose a portion of charge and move slower. On the other hand, chloride anions move rapidly toward the anode. This creates the zone of high field that contains the proteins and move toward the anode reaching the concentration portion of the gel as a narrow band. At the interface of concentration and sealing layer the pH changes from 6.8 to 11 that results in increase of negative charge on glycinate ions that now move rapidly toward anode leaving the proteins out of high strength field. Facing the very small pores of a sealing layer, proteins are stopped and accumulate at the interface of the concentration and sealing layer.

After overnight electrophoresis, the whole structure can be pushed out of the cartridge to access the concentration layer and prepare it for in-gel digestion. [73]

In comparison with other purification techniques TSGE proved to significantly improve the protein recoveries and the downstream analysis. [73] However, this method was tested on buffers containing high concentrations of salts and detergents but not on the buffers that contain high concentrations of sugar as elution buffers utilized in lectin affinity chromatography.

## Dialysis

Dialysis is one of the most widely used methods for sample purification and buffer exchange by removal of low molecular weight contaminants or solution components.

The separation is driven by concentration gradient. It is spontaneous separation process in solution, based on the diffusion through a semipermeable membrane with the definite pore size allowing only the molecules smaller than the pores to pass through.

In practice, the two solutions are placed in two different compartments separated only by a semipermeable membrane. The membrane has a definite pore size and only molecules below a certain molecular weight can pass through, defining the molecular weight cutoff limit (MWCO). The molecules below MWCO can freely pass through the membrane in both directions while larger molecules are retained. The solution containing the analyte of interest and undesired small molecules is in several orders of magnitude smaller volume than the target buffer solution. The purpose of such a difference in volume is to maintain the concentration gradient. Due to the concentration gradient across the membrane molecules diffuse from regions of high concentration to regions of low concentration.

Also, the target buffer is stirred in order to increase the diffusion rate. After entering the low concentration buffer the small molecules form a Nernst diffusion layer where their concentration is higher in comparison with the bulk volume of the target buffer. This slows down the dialysis due to low local concentration gradient. However, stirring efficiently overcomes this obstacle and maintains high concentration gradient.

[74]

The diffusion phenomenon is governed by the Fick's law:

$$J = -D \frac{\partial \phi}{\partial x}$$

J – diffusion flux [ $\frac{\text{mol}}{\text{m}^2 \cdot \text{s}}$ ],

D – diffusion coefficient or diffusivity [ $\frac{\text{m}^2}{\text{s}}$ ],

$\phi$  – concentration [ $\frac{\text{mol}}{\text{m}^3}$ ], and

x – position [m].



Diffusion flux (J) is the measure of the amount of substance passing through the unit area per unit time and depends on the diffusion coefficient and the concentration gradient.

Diffusion coefficient (D) is given by Stokes–Einstein equation:

$$D = \frac{k_B T}{6\pi\eta r}$$

$k_B$  – Boltzman's constant  $1.38 \times 10^{-23} \left[\frac{J}{K}\right]$ ,

T – absolute temperature [K],

$\eta$  – viscosity, and

r – radius of spherical particle.

D is directly proportional to the temperature but inversely proportional to viscosity of the fluid and the size of the particles. For biological molecules  $10^{-11} < D < 10^{-10} \text{ m}^2/\text{s}$ .

As the concentration gradient ( $\partial\phi / \partial x$ ) decreases over time, the system reaches equilibrium and concentrations of small molecules in both compartments become equal. However, this is avoided by frequent change of the target buffer replenishing the system with the new target buffer that is free of the undesired molecules.

Therefore, the best diffusion is achieved in solutions with high concentration gradient, for small molecules in low density solutions at higher temperatures. [74]

Dialysis is a simple, low cost technique that depends on several factors as the time required for buffer exchange, membrane chemistry and its morphological properties as well as the shape of the protein molecules that should not be exchanged between the two compartments. [75]

Time required to complete the buffer exchange depends on the membrane area and temperature. Larger areas will allow for faster exchange but will also increase the loss of protein due to adsorption on the membrane. The increase in temperature will lead

to the faster transport through the membrane due to increased motion of molecules but can also degrade the sample. [75]

Also, considering that buffer molecules can go through the membrane in both directions dialysis can increase the sample volume thus diluting it. For that reason, the samples with low protein concentrations but high salt and detergent concentrations can experience significant sample dilution. This can partially be avoided by stepwise dialysis, where the initial target buffer contains a certain amount of the salt and detergent to decrease the concentration gradient and prevent the sample dilution. [75]

Even though the mechanism of the dialysis remains the same, the technique is significantly improved. The enhanced membrane morphology provides reduced protein adsorption and consequential loss. Also, the commercially available systems allow for a wide range of sample volume and convenience for sample handling. [75]

Glycosylation of the proteins has been recognized as one of the major posttranslational modifications responsible for regulation of protein function. But better understanding of processes involving glycoproteins is hampered due to low abundance of glycoproteins in the biological systems. The techniques described in the introductory portion of this chapter are tested in attempt to find the best approach for isolation and high yield of glycoproteins.

## Experimental

### Materials

Neurotensin and vasopressin peptide standards were purchased from Sigma-Aldrich (St. Louis, MO).

Tris base, SDS, acrylamide, N,N-methylene-bis-acrylamide (Bis), Laemmli sample buffer, ammonium persulfate (APS), N,N,N,N-tetramethylethylenediamine (TEMED), glycine, agarose and Coomassie Brilliant Blue G-250 were obtained from BioRad (Hercules, CA).

Trypsin-EDTA (phenol red, 0.05 %) was obtained from Life Technologies (Grand Island, NY). Protease inhibitor cOmplete was purchased from Roche Diagnostics (Indianapolis, IN).

Slide-A-Lyzer MINI Dialysis Device, 3.5K MWCO, 2mL and bicinchoninic acid assay (BCA) were purchased from Thermo Scientific (Rockford, IL). Agarose bound Concanavalin A (Con A) resin, mannose-BSA, glucose-BSA and glycoprotein eluting solution for mannose- and glucose-binding lectins were purchased from Vector Laboratories (Burlingame, CA).

Ammonium bicarbonate, ammonium citrate, formic acid, methanol, Optima grade acetonitrile, Optima grade water, calcium chloride dehydrate and manganese chloride tetrahydrate were purchased from Fisher Scientific (Chicago, IL). Sequencing grade modified trypsin was supplied by Promega (Madison, WI).

### Equipment

Electro-Eluter model 422 (BioRad) coupled with a power PAC 1000 power supply (BioRad) was used for three-layer sandwich gel electrophoresis. A Finnigan LCQ Deca ion trap mass spectrometer (San Jose, CA) coupled with LC Packings Ultimate 47

quaternary capillary LC system (San Francisco, CA) was used for mass spectrometric analysis.

#### Instrument Optimization – Effect of Inlet Capillary Temperature

The purpose of the following procedure is to determine the most optimal inlet capillary temperature. Vasopressin and neurotensin were dissolved in mobile phase made of 94.9% water, 5% of acetonitrile and 0.1% of formic acid and used as standard peptides. Single charged ions of vasopressin and neurotensin are 1085.4 and 1030.5 respectively, while their double charged ions are 542.7 and 515.7 respectively.

The instrument was set up for the syringe infusion and tuned. The ratio of single to double charged ions at different temperatures was measured for both peptides using that tuning file while constantly infusing the standard peptide solution and acquiring data. Intensity of each ion was averaged over 10 minute time frame. The temperature was changed in increments. The temperature at which the double charged ions are favored is chosen for semi-automatic retuning. The new tuning file and the inlet capillary temperature were used for mass spectrometry analysis.

The effect of the inlet capillary temperature was also examined when a chromatographic separation was included. For that reason neurotensin was analyzed at 2 temperatures (150 °C and 220 °C) using the tuning files for each temperature, chromatographic separation utilizing gradient elution (load method of MudPit) and C18 column. Gradient elution was 110 minutes long.

The mobile phase composition was first held at the appropriate mix of solvents C and D for 15 min (0 to 15 min), and then was switched to solvent A for additional 2 min (15 to 17 min). The gradient was linearly changed from 100% to 50% over 48 min (17 to 65 min) and from 50% to 20% over 20 min (65 to 85 min). The mobile phase composition was returned to 100% of solvent A in 5 min (85 to 90 min) and held at 100% for the next 20 min (90 to 110 min).

## Cell Culture Growth, Maintenance and Lysis

HeLa cell lines were maintained in  $\alpha$ -Minimal Essential Medium ( $\alpha$ -MEM) supplemented with 7% dialysed fetal bovine serum, 2 mM L-glutamine, 100 U/mL penicillin, and 100  $\mu$ g/mL streptomycin.

Since the tissue culture media contains  $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$  ions and proteins from fetal calf serum (FBS) that can inhibit trypsin, the plates containing confluent layer of HeLa cells were washed with sterile PBS then covered with 0.05 % trypsin solution and incubated for 4 minutes at 37 °C to release them from substratum.

Cells were transferred to 14 mL test tube and centrifuged for 5 minutes at 2000 rpm. The media was removed and the cell pellet was washed with PBS and centrifuged again. The washing procedure was performed 3 times in total.

The cell lysis was performed in two different buffers. One of the samples was dissolved in 1 mL of RIPA buffer (10 mM tris-HCl (pH 7.5), 140 mM NaCl, 0.5% (w/v) sodium deoxycholate, 0.1% (w/v) SDS, 0.025% (w/v) sodium azide, 2 % (v/v) TritonX-100). Cells were lysed by passage through a 27 ½ gauge needle. Whole cell lysates were centrifuged (2000 rpm, 5 min) to remove unlysed cells and cell debris. Supernatants of multiple lysates were merged together and taken for further analysis.

The other cell lysis sample was dissolved in 1 mL of 10 mM tris buffer (pH 7.5) with 150 mM NaCl. Cells were lysed by passage through a 27 ½ gauge needle. Whole cell lysates were centrifuged as described above. Supernatants of multiple lysates were merged together and taken for further analysis.

Protein concentration was determined by bicinchoninic acid assay (BCA) according to manufacturer's instructions.

## Lectin Pull-down

### Cell Lysate Preparation

To the HeLa cell lysate calcium chloride and manganese chloride were added to the final concentration of 1 mM. Also, protease inhibitor as well as SDS to the final concentration of 0.05 % were added to the volume containing 1 mg of protein, determined by BCA.

### Con A Column Preparation

Three Con A columns were freshly prepared for each sample. Two milliliters of settled Con A resin were placed in each 5 mL plastic filter tube with the frit at the bottom. Each resin was washed with 15 mL of the column wash buffer (10 mM Tris, pH 7.5, 150 mM NaCl, 1 mM CaCl<sub>2</sub> and 1mM MnCl<sub>2</sub>). Immediately after preparation the sample was applied to the column.

### The Pull-down Procedure

Depending on the protein concentration, 3-4 mL of the whole cell lysate were applied to the first Con A column and incubated over night at 4 °C on the rotator (Figure 3.6). The following day the flow through fraction from the first column was applied directly to the freshly prepared second Con A column. To ensure all the flow through exited the first column, 1.5 mL of column wash buffer was applied to the first column and the flow through was caught in the second column until the flow stopped. The second column is then placed on the rotator and incubated over night at 4 °C.

The first column was washed with 15 mL of column wash buffer. Four milliliters of commercially obtained Vector Labs elution solution were applied to the first column

and incubated for 4 hours at room temperature on the rotator. Glycoproteins were eluted from the column directly into two dialysis cups. To ensure all the Vector Labs elution solution exited the first column, 1.5 mL of 25 mM ammonium bicarbonate buffer was applied to the column and eluate was caught until the flow stopped. The eluate, divided in two dialysis cups, was dialyzed for 24 hours against 9 L of 25 mM ammonium bicarbonate buffer.

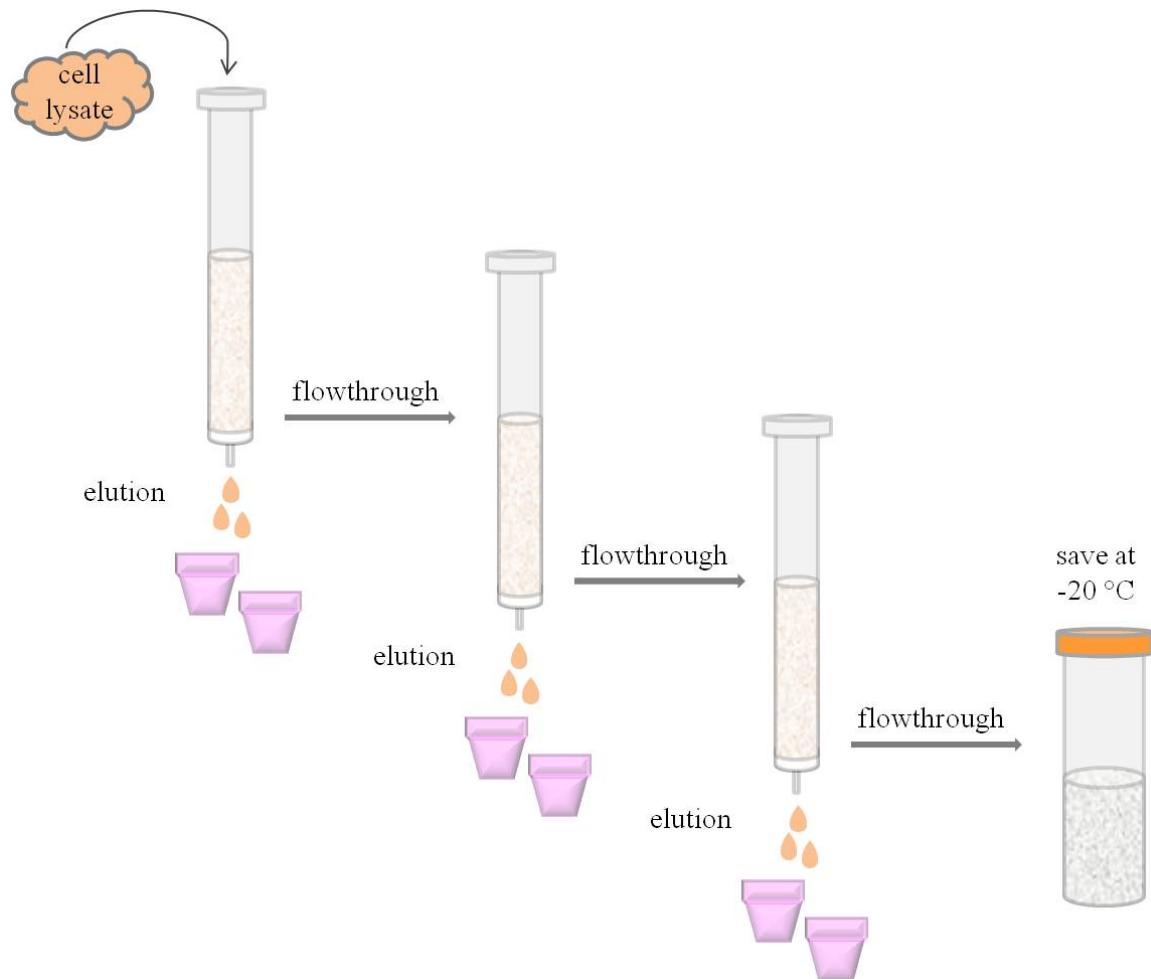


Figure 3.6. The pull-down procedure.

The following day, the procedure was repeated for the second and third column. The flow through fraction from the second column was applied directly to the freshly prepared third Con A column. To ensure all the flow through exited the second column, 1.5 mL of column wash buffer was applied to the second column and the flow through

was caught in the third column until the flow stopped. The third column is then placed on the rotator and incubated over night at 4 °C.

The second column was washed with 15 mL of the column wash buffer. Four milliliters of Vector Labs elution solution were applied to the second column and incubated for 4 hours at room temperature on the rotator. Glycoproteins were eluted from the column directly into two dialysis cups. To ensure all the Vector Labs elution solution exited the second column, 1.5 mL of 25 mM ammonium bicarbonate buffer was applied to the column and eluate was caught until the flow stopped. The eluate, divided in two dialysis cups, was dialyzed for 24 hours against 9 L of 25 mM ammonium bicarbonate buffer.

The final day of the sample preparation, glycoproteins from the third Con A column were eluted. The flow through fraction was stored at -20 °C and the column was washed with 15 mL of the column wash buffer. Four milliliters of Vector Labs elution solution were applied to the third column and incubated for 4 hours at room temperature on the rotator. Glycoproteins were eluted from the column directly into two dialysis cups. To ensure all the Vector Labs elution solution exited the third column, 1.5 mL of 25 mM ammonium bicarbonate buffer was applied to the column and eluate was caught until the flow stopped. The eluate, divided in two dialysis cups, was dialyzed for 24 hours against 9 L of 25 mM ammonium bicarbonate buffer.

The columns for all samples were made of fresh Con A resin to avoid sample contamination.

## Dialysis

Proteins were eluted from the Con A column with 4 mL of Vector Labs elution solution. Since the capacity of the dialysis cup was only 2 mL, the eluate was placed into 2 dialysis cups with molecular weight cutoff limit of 3.5 kDa. The cups were placed on the holder and the bottom was submerged in 1.5 L of stirred 25 mM ammonium bicarbonate pH 8, as shown in the (Figure 3.7). Ammonium bicarbonate solution was



replaced every 2 hours for four times and after 4 hours the fifth time. The sixth and final 1.5 L of ammonium bicarbonate was left over night.

The following day the membrane on the bottom of the dialysis cups was sealed with parafilm to prevent leakage and sweating of the regenerated cellulose membrane and the cups were placed in their 50 mL test tube. The samples were now ready for digestion.

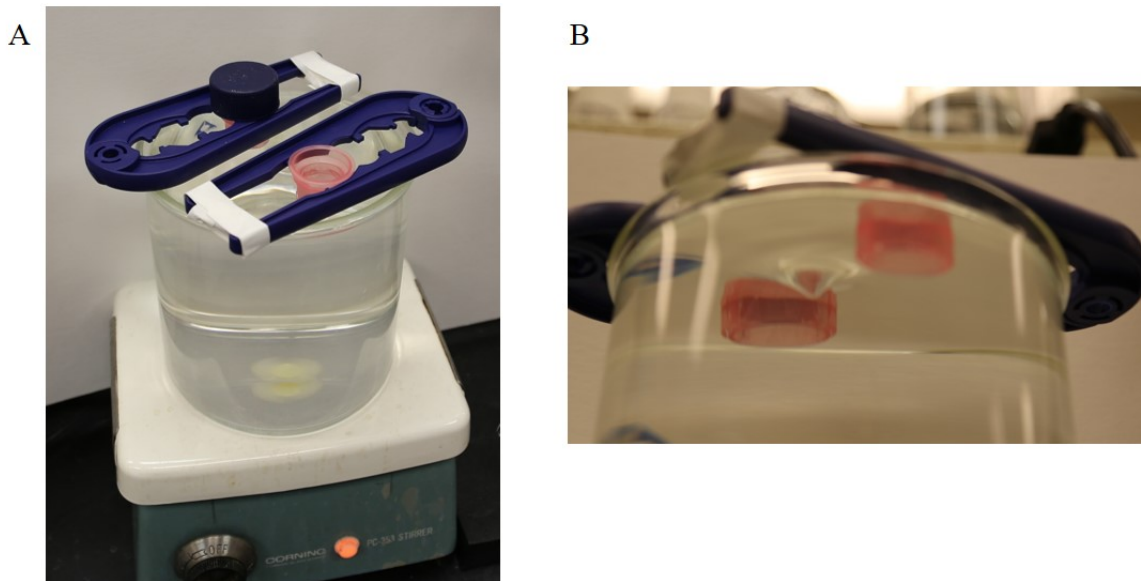


Figure 3.7. A. Dialysis set-up. B. Closer look at the regenerated cellulose membrane.

### The Three-layer Sandwich Gel Electrophoresis

A three-layer sandwich gel was prepared in 5 mL plastic filtration tube with the frit at the bottom. The gel was consisted of an acrylamide sealing layer at the bottom, an acrylamide concentration layer in the middle and an agarose loading layer on the top.

The tube was sealed with the stopper during the preparation process to prevent liquid leakage before the polymerization completion.

The gel and the sample were made according to the recipes and procedures outlined in the publication by Liu et. al. [73]

Briefly, the monomer solution for the sealing layer was loaded to the tube, covered by 200  $\mu$ L of a 1-butanol and allowed to polymerize for 30 minutes after which

the monomer solution for the concentration layer was loaded to the tube, covered by 200  $\mu\text{L}$  of a 1-butanol and also allowed to polymerize for 30 minutes.

The solution of 10 % SDS was added to the sample and incubated at 60°C for 10 minutes. Agarose loading layer, containing the sample, was prepared according to the recipe, quickly transferred to the tube and allowed to cool at 4°C for 10 min. At the end, 10  $\mu\text{L}$  of 0.5% agarose with 1% Bromphenol Blue was overlaid on the gel to monitor electrophoresis. Running buffer was prepared according to the Laemmli's protocol. [76] The gel was run at 15V for 30min, 5V overnight and at gradually increased voltage to 150V until the dye front was at the interface of the concentration and the sealing layer.

### Sample Preparation for Mass Spectrometric Analysis

Digestion of proteins for the dialyzed samples was performed by adding a trypsin solution directly to the dialysis cup and incubating for 18 hours at 37 °C water bath. The amount of added trypsin varied depending on the elution. To the eluate from the first column 10  $\mu\text{g}$  of trypsin was added, while 5  $\mu\text{g}$  each were added to the second and third eluate. After digestion, the peptides were transferred to 1.5 ml test tube and dried using vacuum centrifuge.

Digestion of proteins obtained by TSGE method started with excising the concentration layer of the gel and cutting it to the smallest pieces. These pieces were placed in the 1.5 mL test tube and covered with 50 mM ammonium bicarbonate buffer. After 10 minutes incubation at room temperature the buffer was removed and the pieces were covered with the buffer consisted of equal amounts of acetonitrile and 50 mM ammonium bicarbonate. After 10 minutes incubation at room temperature that buffer was replaced with the pure acetonitrile and incubated for additional 10 minutes at room temperature yielding dried gel pieces. After removal of acetonitrile the pieces were covered with the solution of trypsin in 25 mM ammonium bicarbonate and incubated at 4 °C for 40 minutes. During that time the gel pieces swelled by absorbing the trypsin solution. The excess solution was removed and enough of 25 mM of ammonium bicarbonate was added to cover the gel pieces. After this, the gel pieces, closed in the 1.5

mL test tube were incubated at 37 °C overnight. The following day peptides were recovered by transferring the buffer solution into the new 1.5 mL test tube. Remaining gel pieces were further extracted with 100% acetonitrile and transferred to the buffer tube. The resulting peptide solution was dried using vacuum centrifuge.

Dried peptides obtained from any of the two procedures described above were reconstituted in 20 µL of a mobile phase containing 94.9 % optima water, 5 % of optima acetonitrile and 0.1 % formic acid and analyzed by Finnigan LCQ Deca ion trap mass spectrometer utilizing MuDPIT method.

### Mass Spectrometry Analysis

Finnigan LCQ Deca ion trap mass spectrometer coupled with a LC Packings Ultimate quaternary capillary LC system was used for the mass spectrometry analysis of peptides.

Twenty microliters of digested glycoproteins obtained from the whole HeLa cell lysate were injected onto a laboratory fabricated fused silica capillary strong cation exchange (Partisil SCX) column (350 µm I.D. x 5 cm, 10 µm particles). This column was directly connected to a laboratory fabricated fused silica capillary C18 column (350 µm I.D. x 15 cm, Microsphere 3.5 µm particles). Columns were packed using a stainless steel packing cell. LC Packings Ultimate quaternary capillary LC system with a 20 µL injection loop was used for the sample and mobile phase delivery at 4 µL/min.

Mobile phase A was comprised of 94.9% optima grade water, 5% optima grade acetonitrile and 0.1% formic acid. Mobile phase B was comprised of 94.9% optima grade acetonitrile, 5% optima grade water and 0.1% formic acid. Mobile phase C was comprised of 300 mM ammonium acetate in optima grade water and mobile phase D was the same composition as mobile phase A.

Two-dimensional LC separation was performed with 12 isocratic salt elution where the following salt concentrations of 0 mM, 5 mM, 10 mM, 15 mM, 20 mM, 30 mM, 40 mM, 50 mM, 100 mM, 150 mM 225 mM and 300 mM were obtained by

appropriately mixing mobile phase C with mobile phase D. The 13<sup>th</sup> salt step was obtained by direct injection of 20  $\mu$ L of 3 M ammonium acetate.

Each salt elution was followed by a 110 min C18 reversed-phase gradient elution. The mobile phase composition was first held at the appropriate mix of solvents C and D for 15 min (0 to 15 min), then was switched to solvent A for additional 2 min (15 to 17 min). The gradient was linearly changed from 100% to 50% over 48 min (17 to 65 min) and from 50% to 20% over 20 min (65 to 85 min). Mobile phase composition was returned to 100% of solvent A in 5 min (85 to 90 min) and held at 100% for the next 20 min (90 to 110 min).

The voltage applied to electrospray source was 3.5 kV and 20 units of sheath gas flow. I Inlet capillary was held at 35.0 V and 150 °C. Other parameters were as follows: tube lens offset at -15.0 V, multipole 1 offset at -7.5 V, lens voltage at -50.0 V, multipole 2 offset at -13.0 V, multipole RF amplitude peak-peak at 400 V, entrance lens at -50.0 V and trap DC offset at -10.0 V.

Mass spectrometer was operated in a data-dependent mode. The full scan from  $m/z$  200 to 2000 was acquired first. The most intense precursor ion was selected from the previous full MS scan and submitted to collisional induced dissociation (CID) with an activation Q of 0.250, activation time of 45 ms, and 35% normalized collision energy (NCE). This was followed by a MS/MS scan between  $m/z$  200 and 2000 of the most intense ion of the previous full MS scan. The same procedure was performed for three most intense ions that were excluded from further tandem experiments for one minutes.

Each analytical full scan was constructed as an average of 3 full scans, while each analytical MS/MS scan was constructed as an average of 5 MS/MS scans.

Maximum ion trap injection time of 300 ms was used for the full scans and 500 ms was used for MS/MS scans. Automatic gain control (AGC) target for the full scan was set to  $2 \times 10^7$  and for MS/MS to  $5 \times 10^7$ .

## Protein Data Analysis

All 13 RAW files of a MudPIT run were uploaded to the MASCOT daemon software, combined as one file and searched against SwissProt, human taxonomy. Parameters were set as followed: cleavage enzyme was trypsin with three missed cleavage allowed. Peptide tolerance and the MS/MS tolerance were set to 1.0 Da. Oxidized methionine was chosen as a variable modification.

MudPIT scoring was used and the requirement for positive protein identification was bold red. This requires the protein hit to include at least one bold red peptide match indicating that the peptide is the highest scoring match to the MS/MS spectra.

## Results and Discussion

### Instrument Optimization – Effect of Inlet Capillary Temperature

C-terminus of peptides digested with trypsin have arginines (R) or lysines (K), two most basic or hydrophilic amino acids. Their carboxylic groups are neutral (COOH) in the solution due to free formic acid in the mobile phase. The first proton will reside at the residue of R or K making  $\text{NH}_3^+$  of initial  $\text{NH}_2$  because that is the most nucleophilic group. This proton will not change its place in the process of excitation and fragmentation. This would lead to loss of information about the sequence. However, the second proton can be anywhere along the backbone and reside at any carbonyl group. Upon excitation, the peptide can fragment anywhere along the peptide bond yielding two charged fragments permitting detection of both. In a population of a certain peptide there will be a statistical distribution of proton locations that will give different fragments allowing us to determine the peptide sequence. For all these reasons it is imperative to have at least 2+ precursor ions for the MS/MS analysis.

Increased temperature of the inlet capillary increases the statistical probability of singly charged precursor ions because the high kinetic energy increases the breakage of hydrogen bonded water that can take the proton along with it. Therefore, it is necessary to determine the optimal inlet capillary temperature.

Two commercially available peptides, neurotensin and vasopressin, were separately infused into mass spectrometer. Data were acquired continuously but intensity of ions was averaged over 10 minute time frame when the temperature of the capillary was equilibrated. The temperature was changed incrementally and plotted against the average ratio of intensity of single-to-double charged ions for a given temperature (Figure 3.8). The higher abundance of double charged ions was confirmed for the temperature of 150 °C.

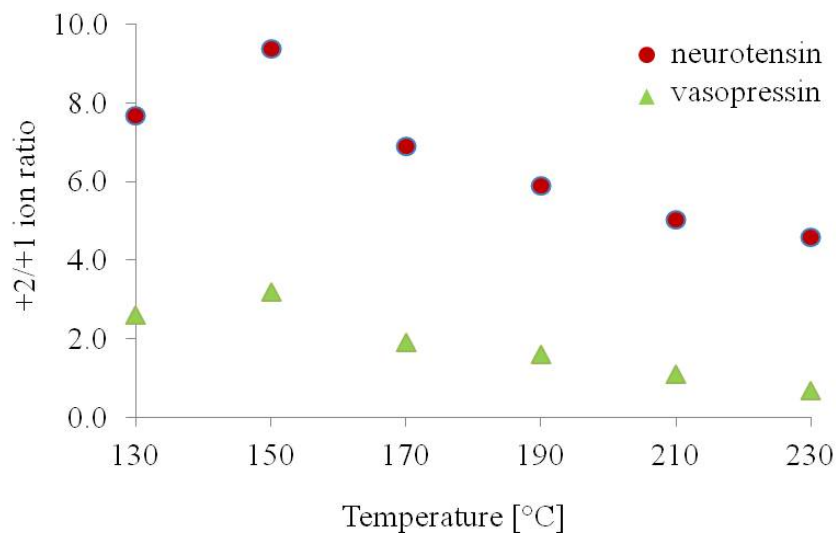


Figure 3.8. Inlet capillary temperature optimization with infusion for neurotensin and vasopressin.

The ratio of single and double charged ions was also examined with chromatographic separation utilizing neurotensin as standard peptide. Results show higher abundance of double charged ions at 150 °C (Figure 3.9) and that temperature was chosen as optimal for further work.

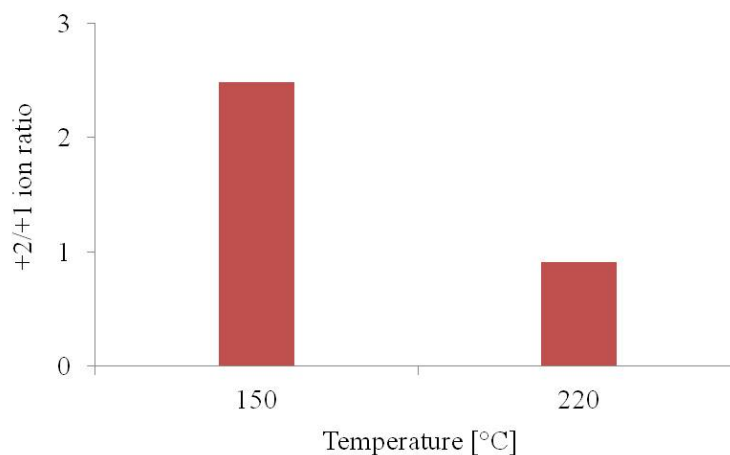


Figure 3.9. Inlet capillary temperature optimization with chromatographic separation for neurotensin.

## Cell Disruption method and Protein Denaturation

The choice of the cell lysis and protein denaturation method depends on compatibility with downstream analysis. For the analysis of glycoproteins, cell lysate is applied to the Con A column, therefore, the buffer containing the cell lysate has to be chosen so that Con A retains its binding activity. Con A is found to tolerate up to 0.05% of SDS, 1% of Triton-X and NP-40. On the other hand, RIPA buffer is one of the most efficient buffers used for the cell lysis. Hence, the stability of the Con A column was tested for the cell lysate in RIPA buffer.

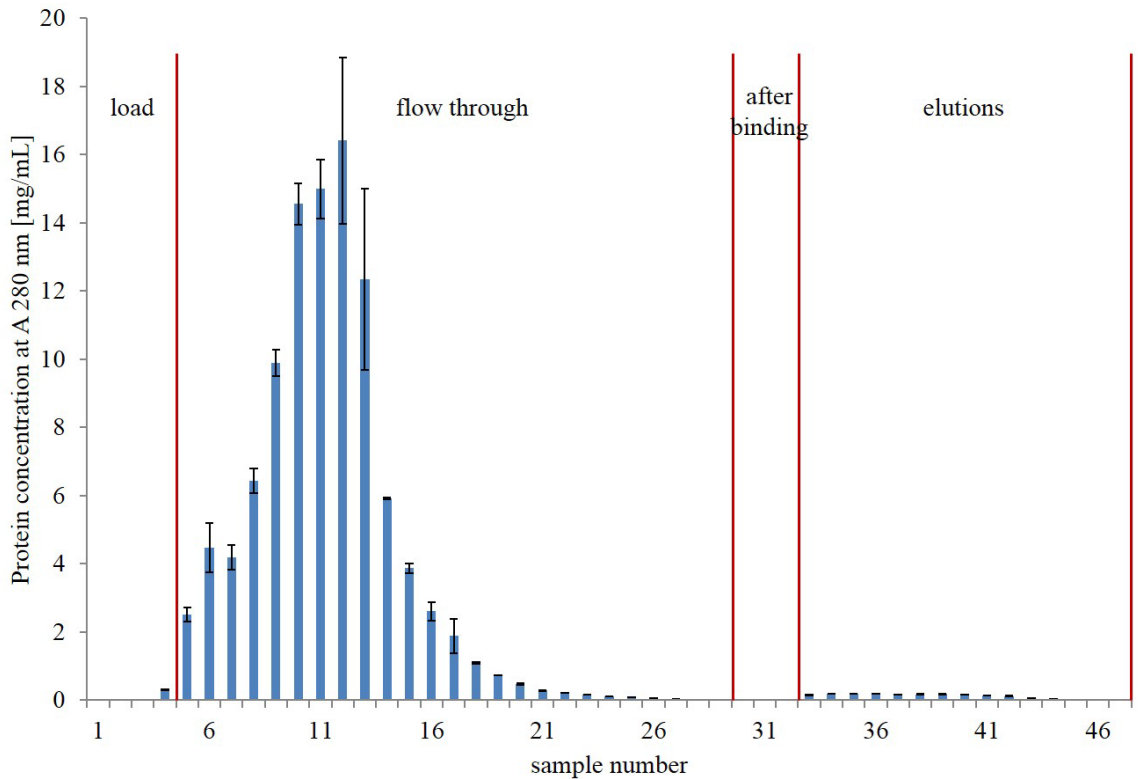


Figure 3.10. Elution profile of mannose-BSA (1 mg/mL) in RIPA buffer.

Results, shown in the Figure 3.10, represent the protein concentration in 250  $\mu$ L fractions calculated by measuring the absorbance at 280 nm. Fractions were caught from the gravity flow Con A column. Mannose-BSA was used as a standard protein to bind to the Con A and was eluted with the Vector Labs elution solution.



The highest protein concentrations were detected in the flow through fractions confirming that Con A lost its binding activity thus eliminating RIPA buffer as a choice for the cell lysis.

Mechanical shearing is alternative approach that disrupts the membrane of the cells but detergents are still required to release membrane- or cytoskeleton- bound proteins. Also, proteins retain their native structure preventing efficient digestion downstream.

Following the elution, digestion is the next step in sample preparation that needs to be considered when choosing cell lysis and denaturation method. Trypsin is the most widely used protease that specifically cleaves proteins at carboxyl terminus of arginine or lysine unless there is adjacent proline. Peptide fragments with one missed cleavage are common due to blocked or slowed digestion, for example in the case of multiple adjacent cleavage sites or if an acidic residue is on either side of the cleavage site. Also when the protein is in its natural folded state not all arginines and lysines are accessible for digestion. Therefore, detergents are used to unfold and denature the proteins allowing for more efficient digestion. On the other hand, trypsin is also a protein, and therefore, sensitive to the presence of detergents. Trypsin retains most of its activity in 0.1% (w/v) SDS solution.

All things considered, heating the cell lysate containing low amount of SDS at 95 °C for 5 minutes was reported as viable alternative. [77] The increase in temperature increases the kinetic energy of the protein, which disrupts the weak bonds and allows for equally probable multiple microstates of the protein to exist. Consequently, entropy raises and protein's random structure is favored, which facilitates digestions due to more accessible cleavage sites.

To explore this approach, the new batch of HeLa cells was lysed in tris buffer and 0.05% SDS was added post lysis obtaining aliquots for 2 samples. Samples were treated exactly the same except one of them was heated 95 °C for 5 minutes and cooled to the room temperature before applying on to Con A column. The cooling to the room temperature was necessary because the sample was applied to the Con A, which is also a protein and, therefore, prone to denaturation at high temperatures.

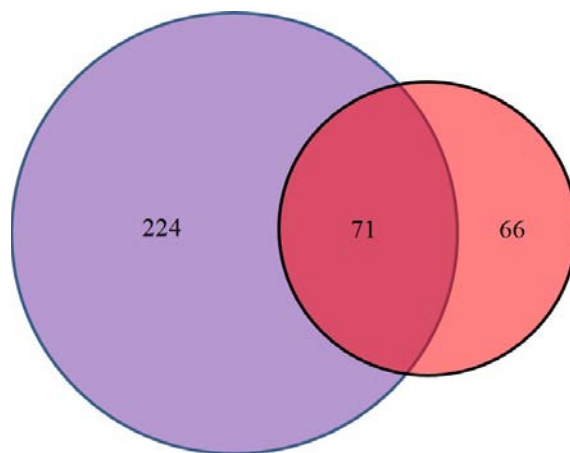


Figure 3.11. The effect of cell lysate heating to the qualitative identification of the proteins. Purple circle: no heat involved. Red circle: heated sample.

Results, shown in the Figure 3.11, suggest that heating the sample does not increase the number of identified proteins. As the increase in temperature results in protein unfolding, cooling process allows for random refolding of denatured proteins decreasing the number of accessible cleavage sites. This results in long peptides with multiple missed cleavages that fall outside of the analytical window of the mass spectrometry method and thus lower the number of identified proteins. However, due to randomness some cleavage sites remain accessible allowing for additional 66 proteins from the heated sample to be identified. The heating approach was abandoned since it yielded overall less proteins.

In conclusion, the cells were lysed in 10 mM tris buffer, pH 7.5, 0.15 M NaCl and protease inhibitor. To prevent foam formation 0.05% SDS was added after lysis as well as 1 mM  $\text{CaCl}_2$  and 1 mM  $\text{MnCl}_2$  required for subsequent binding to Con A.

#### Elution Buffer Evaluation

Con A binds glycoproteins containing  $\alpha$ -mannose,  $\alpha$ -glucose and N-glycans on their surface when the whole cell lysate at the neutral pH containing calcium and manganese is applied to ConA column. Applying binding inhibitors, such as methyl- $\alpha$ -D-mannopyranoside, will release glycoproteins from binding pocket of Con A. The

release is further enhanced by removal of metal ions and decrease of the pH when Con A reversibly loses its binding ability. However, low pH is not suitable for downstream protein digestion due to reversible inactivation of trypsin. Therefore, the solution of binding inhibitor at neutral pH is preferred because it eliminates the need for subsequent buffer exchange.

Achieving a complete elution with the in-house prepared binding inhibitor solution can be difficult while commercially available solutions promise a complete elution but are a proprietary blend of unknown composition thus requiring a subsequent buffer exchange. For that reason we tested two elution solutions. The first was in-house made tris buffer without added metal ions at pH 7.5 containing 0.2 M methyl- $\alpha$ -D-mannopyranoside. The second elution solution (Vector Labs elution solution, ES 1100) with pH 3.0 and undisclosed composition was purchased.

Two samples, of the same protein concentration and otherwise treated the same, were eluted from the Con A column using either tris buffer at pH 7.5 or commercially obtained solution at pH 3.0.

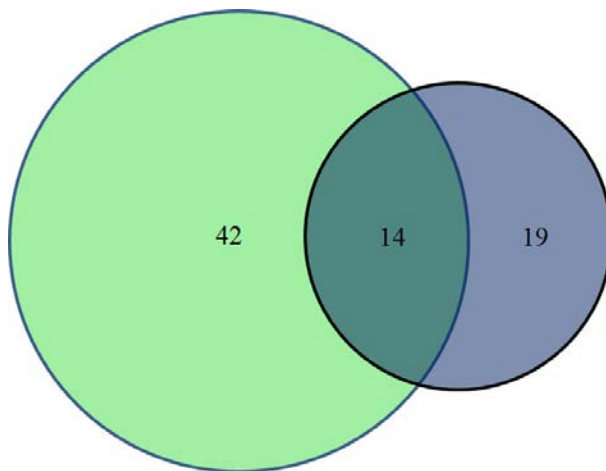


Figure 3.12. The effect of the composition of elution solution on the qualitative identification of proteins. Green circle: Vector Labs elution solution. Blue circle: 10 mM tris pH 7.5 with 0.2 M  $\alpha$ -methyl-mannoside and 0.2 M  $\alpha$ -methyl-glucoside.

Results, shown in the Figure 3.12, prove more efficient elution utilizing commercially obtained solution. However, not all proteins identified utilizing tris buffer

are also identified using the Vector Labs elution solution probably due to pH sensitivity of those proteins. Nevertheless, the Vector Labs elution solution was chosen for further sample processing.

#### Development of glycoprotein purification method

The purification method set up is outlined in the Figure 14. In brief, the whole cell lysate was applied to the first Con A column. Non-retained protein fraction (first flow-through) was passed through a second Con A column and then the second non-retained protein fraction (second flow-through) was passed through the third Con A column yielding 3 sequential pull-downs.

Glycoprotein eluting solution was applied to each column to elute the retained protein fraction. Glycoproteins were eluted from the column directly into two dialysis cups and dialysed against 25 mM ammonium bicarbonate buffer.

After dialysis the bottom of the dialysis cup was sealed and proteins were digested directly in the cup by adding a trypsin solution. After digestion, peptides were dried using vacuum centrifuge and reconstituted in a mobile phase appropriate for the mass spectrometry analysis utilizing MuDPIT method.

To achieve the best yield of this method development the steps of this protocol were individually assessed and optimized. For that purpose HeLa whole cell lysate was utilized as benchmark.

#### Comparison of three layer sandwich gel and dialysis for purification of glycoproteins

Glycoproteins are eluted in commercially obtained, very viscous Vector Labs elution solution incompatible with the downstream tryptic digestion due to low pH of 3.0. Adjusting the pH of the blank Vector Labs elution solution (no proteins) with ammonium hydroxide invoked precipitation. For that reason, the solution containing eluted proteins had to be exchanged with 25 mM ammonium bicarbonate pH 8 that is preferred solution

for the tryptic digestion. Thus two methods were evaluated: three layer sandwich gel and dialysis.

### Three layer sandwich gel electrophoresis

As described in introductory part of this chapter, three layer sandwich gel electrophoresis (TSGE) proved to be superior clean-up method for the protein solutions at near neutral pH, containing high concentrations of salts and detergents. [73] However, protein solution obtained after elution from the Con A column contains high saccharide concentration at pH 3.0 in proprietary solution due to elution with the Vector Labs elution solution. Therefore, TSGE has to be evaluated for this settings as well.

To that end, BSA coupled with glucose (g-BSA) was commercially obtained and used as a standard. The standard sample was prepared by dissolving 400  $\mu\text{g}$  (10  $\mu\text{L}$  stock) of g-BSA in 390  $\mu\text{L}$  of Vector Labs elution solution. This standard sample was loaded and TSGE was run. In spite of running the gel for 28 hours the bromophenol blue, considered as zero molecular weight marker, never reached the concentration layer but was observed in the middle of the agarose gel. Nevertheless, the concentration layer was excised and treated as if proteins were prepared for the mass spectrometry analysis. However, the mass spectrometry analysis showed no significant protein hits raising the question if the problem was Vector Labs elution solution or the protein glycosylation.

These two questions were obvious because the Vector Labs elution solution already showed precipitation when attempting to adjust the pH from 3.0 to 7.5. On the other hand, saccharides are hydrophilic but uncharged compounds. Therefore, glycosylated proteins extensively decorated with saccharides may not be able to see the electric field due to the shielding effect of saccharides on the protein surface. To give an answer to this question two experiments were performed.

The first experiment involved comparison of TSGE of glycosylated protein standards versus non-glycosylated protein standards, both diluted with Vector Labs elution solution to the same concentration yielding 2 samples. Both samples, g-BSA and non-glycosylated BSA, of the same concentrations were prepared from their respective

standard stock solutions by dilution in Vector Labs elution solution and TSGE was run. Both samples yielded same results with the bromophenol blue in the middle of agarose gel and no significant protein hits by mass spectrometry.

To exclude protein glycosylation as a possible impediment for TSGE methodology, the next set of two samples were prepared exactly the same as described in the paragraph above except the protein stock solutions were dissolved in solution containing 10 mM tris buffer, pH 7.5, 0.15 M NaCl and 0.2 M  $\alpha$ -methyl-mannoside as a binding inhibitor. In both samples the bromophenol blue was found at the bottom of the concentration layer after the overnight run. The concentration layer was excised and samples were prepared for mass spectrometry analysis. BSA was identified in both samples confirming that protein glycosylation does not pose an obstacle for TSGE methodology.

The next experiment addressed the question of the compatibility of Vector Labs elution solution with TSGE and for that reason no proteins were included in this experiment. To that end four samples with the decreasing concentrations of the commercially obtained eluting buffer (100%, 75%, 50%, 25% and 0% in tris buffer) were prepared according to the recipe for TSGE sample preparation. Additionally, two samples containing decreasing concentrations of binding inhibitor ( $\alpha$ -methyl-mannoside) were also prepared in tris buffer, pH 7.5, with 0.15 M NaCl to examine effects of the monosaccharide on the TSGE gel. Gels were run over night as outlined in the protocol for TSGE. [73] Results are shown in the Figure 21.

Bromophenol blue dye has a molecular weight of 670 g/mol and is negatively charged at neutral pH thus migrating at the same direction as proteins in the electric field. For that reason, besides using the color of bromophenol blue to visualize the progress of the experiment, this dye can be used as a “zero molecular weight” marker, which was exactly its purpose in this experiment.

It is obvious from the Figure 2.13 that the movement of bromophenol blue depends on the concentration of Vector Labs elution solution but not on the concentration of the binding inhibitor; the higher the concentration of the Vector Labs elution solution the less movement of the dye. Dye moved 22.2% further down the agarose gel when the concentration was decreased by 25%. However, this is not a linear dependence because

the dye moved 77.8% further down the agarose gel when the concentration was decreased by 50% of its initial concentration. It is evident that diluting the Vector Labs elution solution to 25% of its initial concentration leads to bromophenol blue reaching the concentration layer. But that would not be feasible when eluting the glycoproteins with 4 mL of Vector Labs elution solution because the dilution would yield 16 mL of much diluted sample known to suffer of high protein loss in subsequent sample treatment.

Considering gels containing the binding inhibitor in tris buffer, the concentration of the binding inhibitor has no effect on the bromophenol blue movement in the electric field. Therefore, some other component of proprietary Vector Labs elution solution is causing the problem.

As previously stated, the pH adjustment of the Vector Labs elution solution led to precipitation. Out of curiosity if the precipitate formed within the gel as well, gels were stored at 4 °C for 2 days to allow the dye to diffuse and then visualize the site of the dye build-up (Figure 3.14).

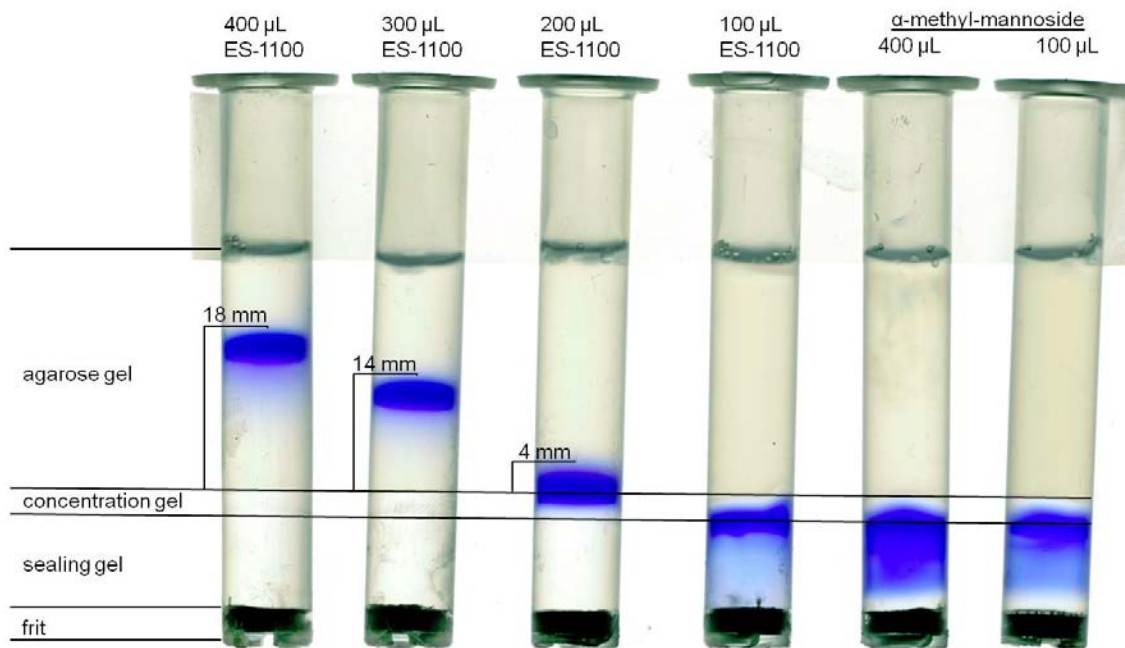


Figure 3.13. Sandwich gel with dye in different concentrations of ES-1100 or Tris +  $\alpha$ -methyl-mannoside elution solution.

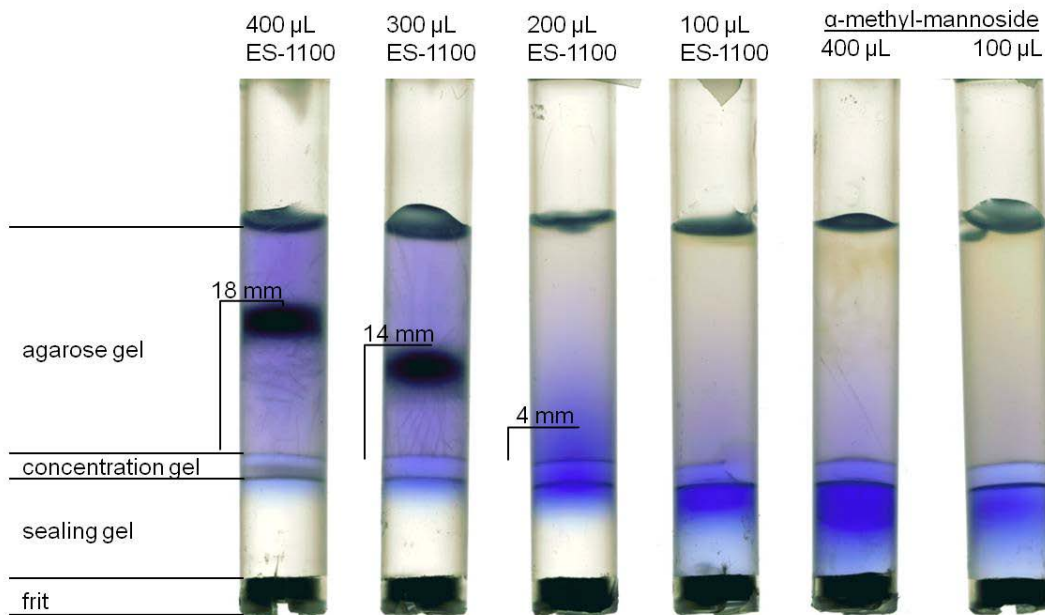


Figure 3.14. Sandwich gel with dye in different concentrations of ES-1100 or Tris +  $\alpha$ -methyl-mannoside elution solution after 2 days in 4°C.

Even after the dye diffused throughout the gel the white precipitate was obvious in two gels with the highest concentration of Vector Labs elution solution but less so in the third and fourth gel where the concentrations were 50% and 25% respectively (Figure 3.14).

These experiments led to conclusion that TSGE method is not suitable for use with Vector Labs elution solution (ES-1100) due to precipitation within the agarose gel. For that reason another method had to be considered and dialysis was the next logical option.

## Dialysis

Dialysis is one of the most widely used methods for sample purification and buffer exchange by concentration gradient driven separation based on the diffusion through a semipermeable membrane. However, one of the draw-backs of dialysis is loss of proteins because they tend to stick to the membrane as well as the wall of the plastic



cup. For that reason digestion was conducted within the cup allowing for parts of the stuck proteins to be digested. Even though digestion of those proteins would not be complete, available peptides would contribute to their identification consequently increasing the number of the identified proteins.

In practice, 3 mg of whole cell lysate was applied to the Con A column and incubated overnight at 4 °C. Retained proteins were eluted with 4 mL of Vector Labs elution solution into two dialysis cups and dialyzed against 25 mM ammonium bicarbonate for 24 hours. The following day trypsin was added directly to the dialysis cups and proteins were digested overnight. After the sample was dried and reconstituted in the mobile phase, mass spectrometry analysis was performed utilizing MuDPIT separation.

As a result 295 proteins were identified with false discovery rate of 17.03% and highest score of 964 making this procedure a method of choice for buffer exchange.

### Sequential Pull-down Method

Binding of glycoproteins to a lectin depends on how well does the exposed glycan fit into the binding pocket of the lectin. Glycan that fits better will replace the weakly bound ones due to increase of the bond strength and decrease of the free energy of the bound state. Therefore, weakly bound proteins can be replaced by proteins that bind more tightly.

To analyze glycoproteins with the lower binding affinity sequential pull-down method was evaluated. In this method the sample was applied to the first Con A column and the non-retained protein fraction (first flow-through) was passed through a second Con A column and then the second non-retained protein fraction (second flow-through) was passed through the third Con A column (3 sequential pull-downs). Employing the sequential pull-down glycoproteins that bind with the highest affinity are retained in the first pull-down allowing for the weaker bound ones to be retained in the next pull-down and even weaker in the third one.

To that end, 3 mg of the whole cell lysate were applied to the Con A column and processed as described. Results, shown in the Figure 3.15, prove that flow through after the first and second pull-down still contain glycosylated proteins. Even though vendor claims 1 mL of settled Con A resin can capture 4 mg of ovalbumin, these results prove that 1.5 mL of settled Con A resin cannot capture all glycoproteins in 3 mg of the whole cell lysate. After the second pull-down, 39 additional proteins were identified and 32 after the third pull-down. In total, 366 proteins were detected, which is 19.4% more than with only the first pull-down.

As mentioned before, the whole cell lysate contains 0.05% SDS. Even though Con A is not denatured by such low concentration of SDS, there is still a possibility that a certain percentage of Con A molecules will lose their activity. Besides the binding strength explained above, this activity loss can be another reason for identifying unique proteins in the second and third pull-down. Moreover, proteins can bind in non-specific fashion in any of the three columns thus increasing the number of unique proteins identified in any of the three pull-downs.

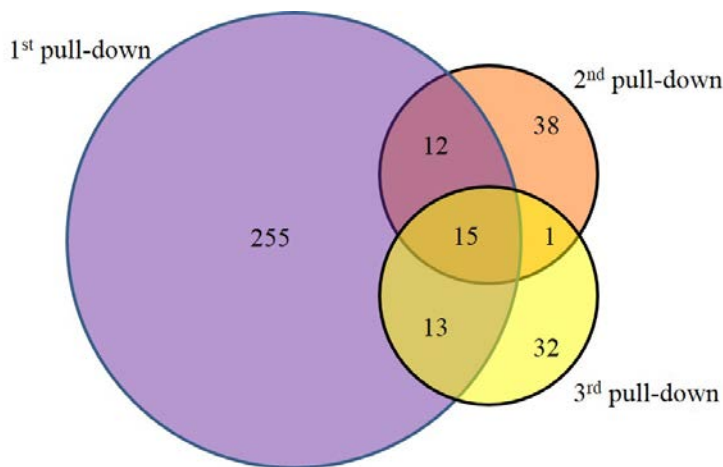


Figure 3.15. Sequential pull-down. Purple circle: first pull-down. Orange circle: second pull-down. Yellow circle: third pull-down.

Reproducibility of this approach was evaluated with the analysis of three samples of the same amount of protein (1.4 mg). Samples were all prepared from the same whole cell lysate and treated the same. Results of the total number of identified proteins in each

samples are presented in the table 3.2 below along with the highest protein score and false discovery rate (FDR).

Table 3.2. Results of the triplicate sample analysis.

<b>sample</b>	<b># protein</b>	<b>FDR</b>	<b>highest score</b>
<b>1</b>	159	16.35%	955
<b>2</b>	193	12.16%	1091
<b>3</b>	226	11.51%	1815

On average 193 proteins were identified with the standard deviation of 34.

However, utilization of mass spectrometry and bioinformatics tools provide deeper understanding of the obtained results, therefore, further validation of this method of isolation of glycoproteins can increase the confidence in database searching results. A better understanding of results starts with understanding of types of acquired peptides and their fragmentation products. Some peptides have N-glycosylation consensus sequence as a part of the amino acid sequence but some do not. Peptides with the consensus sequence may carry a glycan attached to it but not all of them do. The most abundant peptides that reach ion trap undergo fragmentation. The peptide fragmentation can be invoked anywhere along the peptide backbone breaking the peptide bonds. The fragment that carries the charge is the detected ion that can be identified by the search against the protein database for the particular organism. However, glycans attached to peptides can produce ions of larger mass or even sterically block trypsin to access the cleavage sites due to their elaborate structure and prevent the digestion of that site. [78] For those reasons many peptides will not be identified because their masses do not match the masses of the peptides in the database. If the glycan is very complex and introduces a much larger increase in mass or the peptide has multiple missed cleavages, the mass of that peptide may not fall within the detection range of the mass spectrometer used preventing the ion to even be detected. However, if the mass of the glycan is known and not too large, the mass of the expected ion can be calculated and the perceived peptide can be searched manually. But that does not mean that peptides with the consensus sequence cannot be identified automatically. About two thirds of proteins that contain

the consensus sequence are actually glycosylated. Remainders are identified as potential N-glycan sites. Therefore, it is possible to identify peptides with the consensus sequence but experimental proof of N-glycan is required. For all these reasons, some peptides with the consensus sequence will be detected but some will not. One example of such occurrence is endoplasmin that is used here as an example of the identified proteins and thereof analysis.

Endoplasmin precursor has a nominal mass of 92411, 803 amino acids sequence and calculated pI value of 4.76. In UniProtKB/Swiss-Prot database it is listed under the code P14625 (ENPL\_HUMAN) and under the name “endoplasmin”. This molecular chaperone is found in endoplasmatic reticulum and helps transport and process secreted proteins. It also serves in ER-associated degradation (ERAD) of terminally misfolded or unassembled proteins. [79]

Utilizing this method, endoplasmin was identified with the score of 368 with sequence coverage of 43%. The protein sequence is shown below with the identified peptides in red color and the ones that were not identified in black color.

```

1  MRALWVLGLC CVLLTFGSVR ADDEVDVDGT VEEDLGKSRE GSRTDDEVVQ
51  REEEAIQLDG LNASQIRELR EKSEKFAFQA EVNRMMKLI NSLYKNKEIF
101 LRELISNASD ALDKIRLISL TDENALSGNE ELTVKIKCDK EKNLLHVTDT
151 GVGMTREELV KNLGTIAKSG TSEFLNKMTE AQEDGQSTSE LIGQFGVGFY
201 SAFLVADKVI VTSKHNDTQ HIWESDSNEF SVIADPRGNT LGRGTTITLV
251 LKEEASDYLE LDTIKNLVKK YSQFINFPIY VWSSKTETVE EPMEEEEAAK
301 EEKEESDDEA AVEEEEEEEKK PKTKKVEKTV WDWELMNDIK PIWQRPSKEV
351 EEDEYKAFYK SFSKESDDPM AYIHFTAEGE VTFKSILFVP TSAPRGLFDE
401 YGSKKSDYIK LYVRRVITD DFHDMMPKYL NFVKGVVSD DLPLNVSRET
451 LQQHLLKVI RKKLVRKTLD MIKKIADDKY NDTFWKEFGT NIKLGVIEDH
501 SNRTRLAKLL RFQSSHPTD ITSLDQYVER MKEKQDKIYF MAGSSRKEAE
551 SSPFVERLLK KGYEVIYLTE PVDEYCIQAL PEFDGKRFQN VAKEGVKFDE
601 SEKTKESREA VEKEFEPLN WMKDKALKDK IEKAVVSQRL TESPCALVAS
651 QYGWSGNMER IMKAQAYQTG KDISTNYAS QKKTFEINPR HPLIRDMLRR
701 IKEDEDDKTV LDLAVVLFET ATLRSGYLLP DTKAYGDRIE RMLRLSLNID
751 PDAKVEEPEE EEPEETAEDT TEDTEQDEDE EMDVGTDEEE ETAKESTAEK
801 DEL

```

Considering consensus sequence for N-glycosylation there are 8 possible N-glycosylations sites (underlined) but only two, at Asn-217 and Asn-445, have been reported (underlined and bold italic). [80]

Even more details can be obtained on the peptide level. For example, let's consider the ion 638.75 m/z from the list of ions assigned to endoplasmic reticulum (Figure 3.16). Since 638.75 m/z ion is doubly charged, it was identified with the mass of 1275.48 m/z for the singly charged ion. Singly charged ion was matched with the neutral peptide with calculated monoisotopic mass of 1274.64 m/z. The fragmentation pattern of 638.75 m/z ion was matched with the sequence ELISNASDALDK. Out of 129 predicted ions shown in the table 3.3, 29 were identified (bold) by analysis of 36 most intense peaks shown in the mass spectrum below.

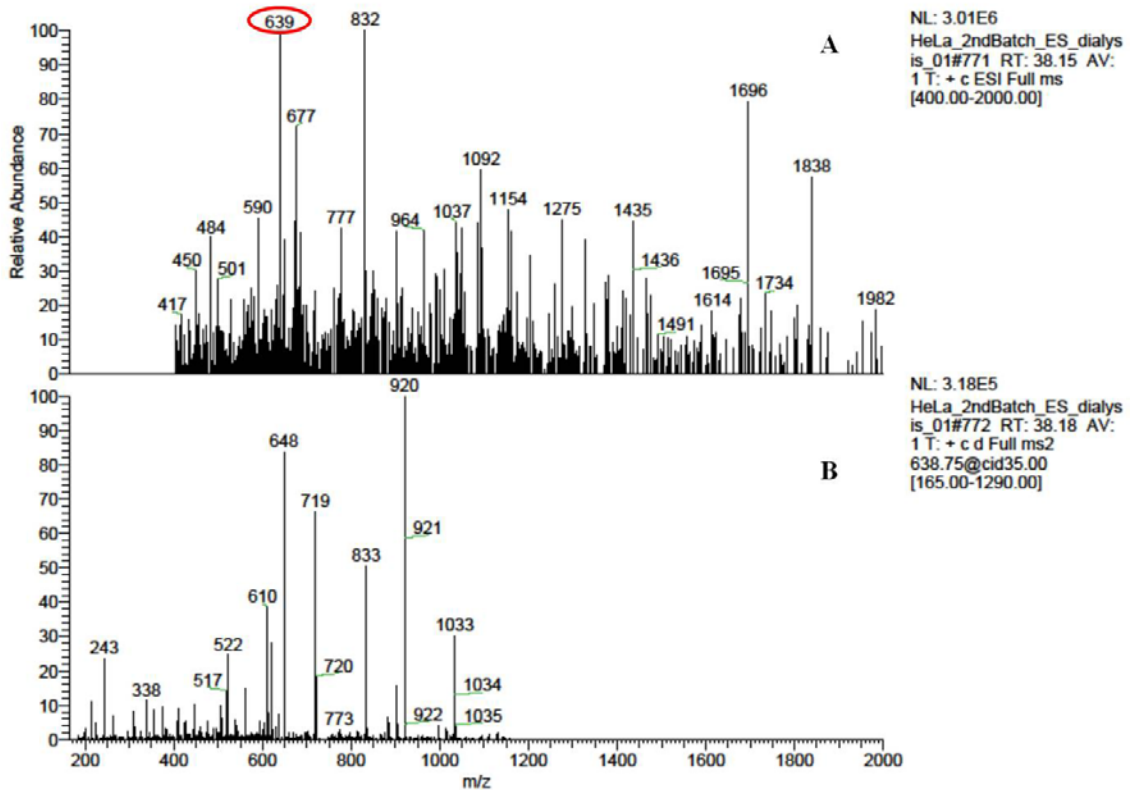


Figure 3.16. Mass spectrum: A. Full mass spectrum (MS), B. Mass spectrum of fragmented ion (MS/MS)

Table 3.3. List of expected fragment ions of the ion 638.75 m/z. Detected in bold.

#	b	b <sup>++</sup>	b*	b <sup>+++</sup>	b <sup>0</sup>	b <sup>0++</sup>	Seq.	y	y <sup>++</sup>	y*	y <sup>+++</sup>	y <sup>0</sup>	y <sup>0++</sup>	#
1	130.04	65.52			112.03	56.52	E							12
2	<b>243.13</b>	122.07			225.12	113.06	L	1146.60	573.80	<b>1129.57</b>	565.29	<b>1128.58</b>	564.79	11
3	<b>356.21</b>	178.61			<b>338.20</b>	169.60	I	<b>1033.51</b>	<b>517.26</b>	1016.48	<b>508.74</b>	1015.50	<b>508.25</b>	10
4	443.25	222.12			425.23	213.12	S	<b>920.43</b>	460.71	<b>903.40</b>	452.20	<b>902.42</b>	451.71	9
5	557.29	279.15	540.26	270.63	539.28	270.14	N	<b>833.39</b>	417.20	816.37	<b>408.69</b>	815.38	<b>408.19</b>	8
6	628.33	314.66	611.30	306.15	<b>610.31</b>	305.66	A	<b>719.35</b>	360.18	702.33	351.66	701.34	351.17	7
7	715.36	358.18	698.33	349.67	697.35	349.17	S	<b>648.31</b>	324.66	631.29	316.15	630.30	315.65	6
8	830.38	415.69	813.36	407.18	812.37	406.69	D	<b>561.28</b>	281.14	544.26	272.63	543.27	272.14	5
9	901.42	451.21	884.39	442.70	<b>883.41</b>	442.21	A	<b>446.26</b>	223.63	429.23	<b>215.12</b>	428.25	<b>214.62</b>	4
10	1014.51	<b>507.75</b>	997.48	499.24	<b>996.49</b>	498.75	L	<b>375.22</b>	188.11	358.19	179.60	357.21	179.11	3
11	<b>1129.53</b>	565.27	1112.51	556.75	<b>1111.52</b>	556.26	D	<b>262.13</b>	131.57	245.11	123.06	244.12	122.56	2
12							K	147.11	74.06	130.08	65.54			1

Considering chromatograms (Figure 3.17), if only total ion chromatogram at 38.15 minutes no discernable peaks are observed. However, if the chromatogram of only a single ion is extracted the peak is obvious.

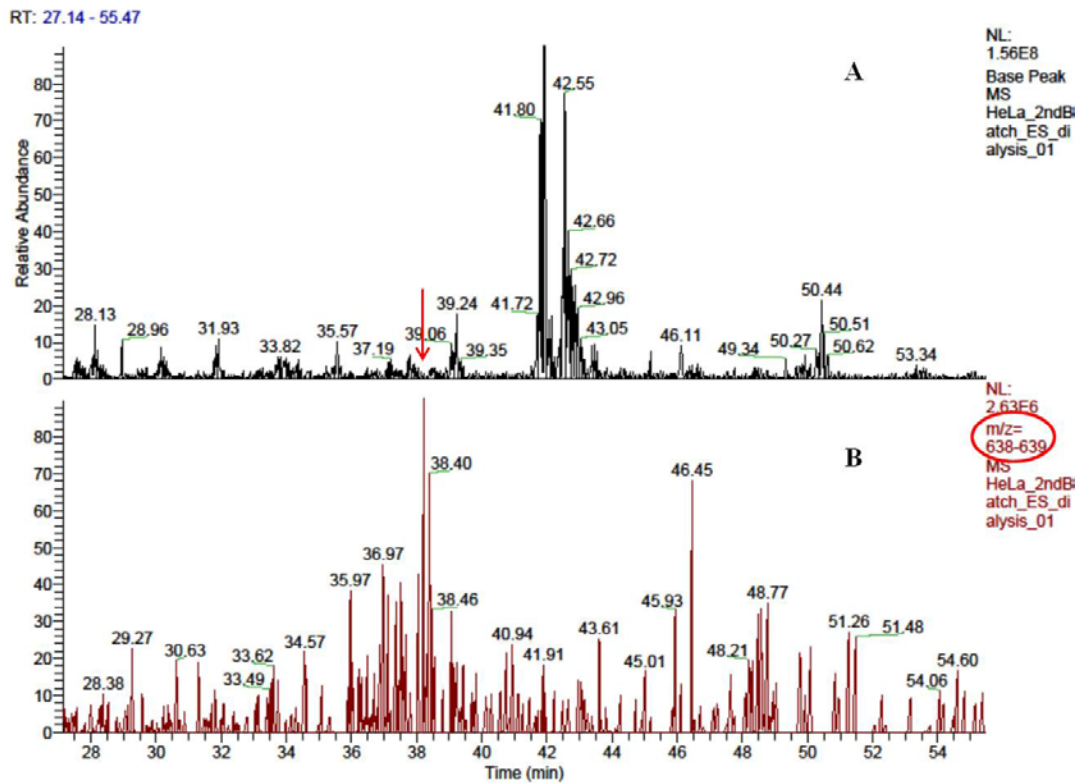


Figure 3.17. Chromatogram: A. Total ion chromatogram, B. Single ion chromatogram for the ion 638.75 m/z.

In light of the fast development of bioinformatics tools, one could argue that identified glycoproteins could be extracted from the list of identified proteins from the whole cell lysate acquiring only proteins with appropriate consensus sequence and hence significantly decrease the sample handling time. However, there are two obstacles to that simplified approach. First, a protein can have multiple consensus sequences but glycans will not be attached to all of them. Hence the glycosylation has to be confirmed and the lectin binding is a viable conformation. The second obstacle is the sample complexity. Glycosylation of the human proteome is at the level of 5%. Since only the most abundant proteins are identified, glycoproteins would compete for identification with a large number of other proteins in the mixture and thus would very likely not to be identified. Therefore, the reduction in sample complexity utilizing the lectin pull-down method provides a substantial advance in protein identification.

## Conclusion

Up to date, there is no one method capable of profiling all proteins present in the whole cell lysate. Therefore, sample separation and purification methods are developed as an answer to the particular question about the proteome. Sample preparation for mass spectrometry analysis is a critical step in the proteomics workflow because the quality and reproducibility of sample extraction and preparation significantly impacts the separation and identification capabilities of mass spectrometers. The additional obstacle is to extract only glycosylated proteins for which there is no one standard method.

The method discussed in this chapter was developed in an effort to identify as many as possible N-glycosylated proteins. This is achieved by application of the whole cell lysate to agarose bound Concanavalin A (Con A) beads to efficiently pull down glycoproteins, dialyze and analyze them using MuDPIT. This method was further enhanced by passing the non-retained protein fraction (first flow-through) through a second Con A column and then passing the second non-retained protein fraction (second flow-through) through the third Con A column (3 sequential pull-downs) yielding 366 N-

glycosylated proteins. Utilizing three sequential pull-downs rather than just one, the number of identified proteins was increased by 19.4%.

Considering the tendency of proteins to stick to the plastic surfaces, this method was crafted with minimal protein loss in mind. That was achieved by minimizing the number of steps in the procedure as well as the equipment. Although the use of more elaborate separation protocols could increase the depth of protein coverage, there will always be peptides that are out of the analytical window and will not be identified.



CHAPTER 4  
GLYCOPROTEOMES OF TACHYZOITES AND TISSUE CYSTS  
AND THEIR DIFFERENCIES

*Toxoplasma gondii*

*Toxoplasma gondii*, as a member of Apicomplexa phylum, is obligate intracellular protozoan parasite that causes the disease called toxoplasmosis. Even though humans are accidental hosts, some estimates suggest that over 30 % of human population is infected. Infected humans could have parasite colonies anywhere in brain or muscle tissue. Therefore, human symptoms of toxoplasmosis may depend on parasite's final location explaining, for example, the newly found link between schizophrenia and toxoplasmosis. [81]

Recent studies show that a *Toxoplasma gondii* found in the brain of mammals encodes the enzyme for producing dopamine and as consequence production and release of significantly higher than normal amount of dopamine in infected brain cells. These findings may change the way dopamine-related neurological disorders as schizophrenia, attention deficit hyperactivity disorder and Parkinson's disease are treated. [81]

Most infected people are healthy and do not exhibit any symptoms or develop sore lymph nodes, muscle pains and other minor flu-like symptoms that last for only several weeks. However, for those who are immune-suppressed, and particularly for fetus, there are significant health risks that can occasionally be fatal. [81]

Life cycle of *Toxoplasma gondii* starts with the sexual cycle in its feline definitive hosts. The cat sheds parasites in the form of oocysts that take 1 to 5 days to sporulate and become infective. Asexual cycle starts once the intermediate hosts such as birds, rodents, humans, etc. ingest the material contaminated with oocysts that shortly after transform into rapidly growing tachyzoites. Once tachyzoites are localized in brain or muscle tissue they develop into tissue cysts containing bradyzoites. [82]

Bradyzoites and tachyzoites are two interconvertable stages of *Toxoplasma gondii* in intermediate hosts. Transmissible bradyzoites are growing slowly making a dormant,

encysted form. Upon ingestion the cysts' wall is digested releasing bradyzoites. Being resistant to gastric proteases, they invade the small intestine where they convert into rapidly growing, infectious tachyzoites. This form replicates inside a parasitophorous vacuole (PV), egress, and then infects neighboring cells. Host's immune system eliminates most of parasites but enough of them survive and convert back into bradyzoites and wait until a member of a feline family ingests the intermediate host to finish their life cycle. [82]

Toxoplasmosis can be treated with combinations of antibiotics as pyrimethamine, which interfere with RNA and DNA synthesis, with either trisulfapyrimidines or sulfadiazine that stops folic acid production in parasite but *Toxoplasma gondii* cannot be eliminated completely. Parasites can remain within tissue cells in a less active stage (cyst) in locations difficult for medication to get to.

All this calls for better understanding of *Toxoplasma gondii* that will point toward possible cure. The whole proteome of *Toxoplasma gondii* is yet to be annotated and this dissertation represents the effort put toward developing a method to study the glycoproteome of intracellular organisms and consequently increase the knowledge base of *Toxoplasma gondii*.

### Structure of *Toxoplasma gondii*

*Toxoplasma gondii* is a part of Apicomplexa phylum named after distinct apical complex structure involved in penetrating a host's cell.

*Toxoplasma gondii* goes through three infectious stages: tachyzoites (rapidly multiplying), bradyzoites (slow growing) and sporozoites. [83] This dissertation focuses on tachyzoites and tissue cysts containing bradyzoites.

Tachyzoites and bradyzoites have a similar structure. They are both crescent shaped, 2 to 6  $\mu\text{m}$  long with pointed front end and rounded back end. They consist of triple membrane system called pellicle and subpellicular microtubules, apical rings, polar rings, conoid, rhoptries, micronemes, micropore, endoplasmic reticulum, Golgi complex,

ribosomes, rough and smooth endoplasmic reticulum, micropore, mitochondrion, nucleus, dense granules, amylopectin granules and apicoplast. [83]

However, there are some differences as the appearance, position and number of some organelles. The structure of rhoptries of tachyzoites look like labyrinth while bradyzoites have uniformly dense rhoptries with some of them folded back on themselves. Also the nucleus in bradyzoites is positioned toward the front end while more centrally located in tachyzoites. Higher numbers of micronemes and dense granules are reported in tachyzoites, while the higher number of amylopectin granules is observed in bradyzoites. Also, bradyzoites are less prone to proteolytic digestion than are tachyzoites. [83]

While tachyzoites grow in groups called clones, bradyzoites grow within tissue cysts. The wall of the tissue cyst is about 0.5  $\mu\text{m}$  thick and elastic and composed of the host cell and parasite materials. Tissue cysts vary in size depending on the cyst age, from 5  $\mu\text{m}$  for young cysts containing as few as two bradyzoites to 70  $\mu\text{m}$  for the older cysts that can contain several hundred bradyzoites. [83] .

#### Glycoproteome of *Toxoplasma gondii*

Unlike *Toxoplasma gondii*, majority of members of Apicomplexa phylum cannot form the tissue cysts. Those members are proved to lack enzymes required for the Asparagine Linked Glycan (ALG) pathway. However, the analysis of the genome of *Toxoplasma gondii* proves the ability of the parasite for N-linked glycosylation suggesting that N-linked glycosylation may play a key role in the establishment and maintenance of the tissue cyst. [84]

Glycosylation of tachyzoites of *Toxoplasma gondii* proteins has been a controversial issue and demonstrated only recently using lectin pull-down methods and mass spectrometry that showed the great diversity of glycoproteins in *Toxoplasma gondii*. [66, 77, 85, 86] However, the mechanism of transformation of tachyzoites to bradyzoites and the tissue cyst formation is yet unknown. Also, little is known about the glycoproteome of tissue cysts except their affinity to react with *Dolichos biflorus*

agglutinin (DBA) (Figure 4.1) and succinyl Wheat Germ Agglutinin (SWGA) lectins. [66] Combining DBA and Con A staining (Figure 4.2) demonstrates the presence of  $\alpha$ -linked mannose on the bradyzoites but not the cyst wall.

The evidence suggests that glycosylation could be essential for cyst formation and maintenance which is characteristic of chronic stage of disease. Therefore, further studies are required to better understand the interaction between *Toxoplasma gondii* and its host in order to develop new preventative and therapeutic strategies.

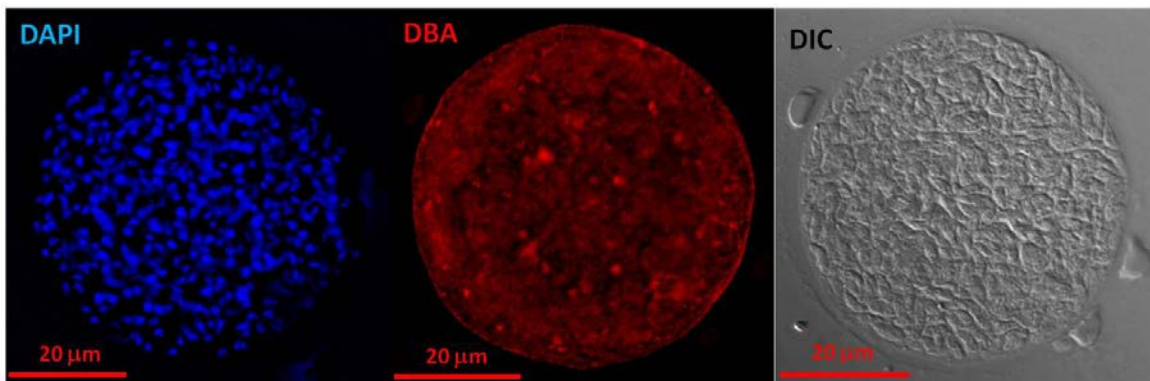


Figure 4.1. Staining of the tissue cyst. DAPI: Hoechst staining binds to DNA and each blue spot is an individual parasite. DBA: Dolichos biflorus agglutinin lectin binds to  $\alpha$ -linked N-acetylgalactosamines, which are abundantly present on the cyst wall. DIC: dif

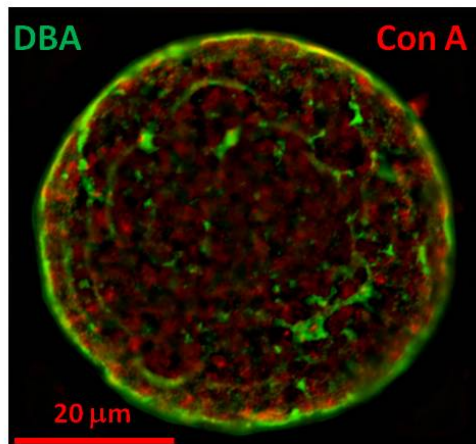


Figure 4.2. Staining of the tissue cyst. DBA, stained in green, highlights the cyst wall. Concanavalin A (Con A), stained in red, recognizes  $\alpha$ -linked mannose.

## Experimental

### Materials

Trypsin-EDTA (phenol red, 0.05 %) was obtained from Life Technologies (Grand Island, NY). Protease inhibitor cOmplete was purchased from Roche Diagnostics (Indianapolis, IN).

Slide-A-Lyzer MINI Dialysis Device, 3.5K MWCO, 2mL and bicinchoninic acid assay (BCA) were purchased from Thermo Scientific (Rockford, IL). Agarose bound Concanavalin A (Con A) resin and glycoprotein eluting solution for mannose- and glucose-binding lectins were purchased from Vector Laboratories (Burlingame, CA).

Ammonium bicarbonate, ammonium citrate, formic acid, methanol, Optima grade acetonitrile, Optima grade water, calcium chloride dehydrate and manganese chloride tetrahydrate were purchased from Fisher Scientific (Chicago, IL). Sequencing grade modified trypsin was supplied by Promega (Madison, WI).

### Equipment

A Finnigan LCQ Deca ion trap mass spectrometer (San Jose, CA) coupled with LC Packings Ultimate 47 quaternary capillary LC system (San Francisco, CA) was used for mass spectrometric analysis. Thermo French Pressure Cell Press (Rockford, IL) was utilized for cell lysis.

### Cell Culture Growth and Maintenance

Tachyzoites ME49 were grown in the Vero cell line. Vero cells were maintained in  $\alpha$ -Minimal Essential Medium ( $\alpha$ -MEM) supplemented with 7% heat-inactivated fetal bovine serum, 2 mM L-glutamine, 100 U/mL penicillin, and 100  $\mu$ g/mL streptomycin.

Plates with confluent Vero monolayers were infected with  $1 \times 10^6$  parasites and incubated at 37 °C until the majority of the Vero cells lysed out. The content of the plates was transferred to the 14 mL test tube and centrifuged at 2000 rpm for 10 minutes. Parasites and parasites containing cells were resuspended in 1 mL PBS and selectively lysed by passage through a 27 gauge needle. After centrifugation the obtained parasite pellet was washed three times by sequential centrifugation and resuspension in 10 mL sterile PBS. The final pellet was obtained after centrifugation at 2000 rpm for 10 minutes. Pellets from multiple plates were merged together and resuspended in 10 mM tris (pH 7.5) with 150 mM NaCl and protease inhibitor. French press at a pressure of 10,000 psi at medium setting was utilized for parasite lysis.

The purified brain cysts were obtained from Dr Sinai's laboratory. They were harvested from CBA/J mice around 3-4 weeks post infection and purified by percoll gradient purification. Purified cysts were prepared for the mass spectrometry analysis following the same protocol as used for tachyzoites.

#### Lectin Pull-down

#### Cell Lysate Preparation

Parasite pellets, tachyzoites or bradyzoites, were resuspended in 10 mM tris buffer (pH 7.5) with 150 mM NaCl and lysed utilizing French press at a pressure of 10,000 psi at medium setting. Protein concentration was determined by bicinchoninic acid assay (BCA) according to manufacturer's instructions. Calcium chloride and manganese chloride to the final concentration of 1 mM, protease inhibitor as well as SDS to the final concentration of 0.05 % were added to the cell lysate containing 1 mg of protein, determined by BCA.

## Con A Column Preparation

Three Con A columns were prepared for each sample. Two milliliters of settled Con A resin were placed in each plastic filter tube. Each resin was washed with 15 mL of buffer containing 10 mM Tris (pH 7.5), 150 mM NaCl, 1 mM CaCl<sub>2</sub> and 1mM MnCl<sub>2</sub>.

## The Pull-down Procedure

Depending on the protein concentration, 3-4 mL of the parasite lysate were applied to the first Con A column and incubated over night at 4 °C on the rotator. The following day the flow through fraction from the first column was applied directly to the freshly prepared second Con A column. To ensure all the flow through exited the first column, 1.5 mL of column wash buffer was applied to the first column and the flow through was caught in the second column until the flow stopped. The second column is then placed on the rotator and incubated over night at 4 °C.

The first column was washed with 15 mL of column wash buffer. Four milliliters of glycoprotein eluting solution were applied to the first column and incubated for 4 hours at room temperature on the rotator. Glycoproteins were eluted from the column directly into two dialysis cups. To ensure all the Vector Labs elution solution exited the first column, 1.5 mL of 25 mM ammonium bicarbonate buffer was applied to the column and eluate was caught until the flow stopped. The eluate, divided in two dialysis cups, was dialyzed for 24 hours against 9L of 25 mM ammonium bicarbonate buffer.

The following day the procedure was repeated for the second and third column. The flow through fraction from the second column was applied directly to the freshly prepared third Con A column. To ensure all the flow through exited the second column, 1.5 mL of column wash buffer was applied to the second column and the flow through was caught in the third column until the flow stopped. The third column is then placed on the rotator and incubated over night at 4 °C.

The second column was washed with 15 mL of the column wash buffer. Four milliliters of glycoprotein eluting solution were applied to the second column and

incubated for 4 hours at room temperature on the rotator. Glycoproteins were eluted from the column directly into two dialysis cups. To ensure all the Vector Labs elution solution exited the second column, 1.5 mL of 25 mM ammonium bicarbonate buffer was applied to the column and eluate was caught until the flow stopped. The eluate, divided in two dialysis cups, was dialyzed for 24 hours against 9 L of 25 mM ammonium bicarbonate buffer.

The final day of the sample preparation glycoproteins from the third Con A column were eluted. The flow through fraction was stored at -20 °C and the column was washed with 15 mL of the column wash buffer. Four milliliters of glycoprotein eluting solution were applied to the third column and incubated for 4 hours at room temperature on the rotator. Glycoproteins were eluted from the column directly into two dialysis cups. To ensure all the Vector Labs elution solution exited the third column, 1.5 mL of 25 mM ammonium bicarbonate buffer was applied to the column and eluate was caught until the flow stopped. The eluate, divided in two dialysis cups, was dialyzed for 24 hours against 9 L of 25 mM ammonium bicarbonate buffer.

Columns for all samples were made of fresh Con A resin to avoid sample contamination.

## Dialysis

Proteins were eluted from the Con A column with 4 mL of Vector Labs elution solution. Since the capacity of the dialysis cup was only 2 mL, the eluate was placed into 2 dialysis cups with molecular weight cutoff limit of 3.5 kDa. Cups were placed on the holder and the bottom was submerged in 1.5 L of stirred 25 mM ammonium bicarbonate pH 8. Ammonium bicarbonate solution was replaced every 2 hours for four times. The fifth solution was replaced after 4 hours. The sixth and final 1.5 L of ammonium bicarbonate were left over night.

The following day, membranes on the bottom of dialysis cups was sealed with parafilm to prevent leakage and sweating of the regenerated cellulose membrane and cups were placed in their 50 mL test tube. Samples were now ready for digestion.



## Sample Preparation for Mass Spectrometric Analysis

Digestion of proteins was performed by adding trypsin solution directly to the dialysis cup and incubating for 18 hours at 37 °C water bath. The amount of added trypsin varied depending on the elution. To the eluate from the first column 10 µg of trypsin was added, while 5 µg each were added to the second and third eluate.

After digestion, peptides were transferred to 1.5 ml Eppendorf tube and dried using vacuum centrifuge. Peptides were reconstituted in 20 µL of a mobile phase containing 94.9 % optima water, 5 % of optima acetonitrile and 0.1 % formic acid and analyzed by Finnigan LCQ Deca ion trap mass spectrometer utilizing MuDPIT method.

## Mass Spectrometry Analysis

Finnigan LCQ Deca ion trap mass spectrometer coupled with a LC Packings Ultimate quaternary capillary LC system was used for the mass spectrometry analysis of peptides.

Twenty microliters of digested glycoproteins obtained from the parasite lysate were injected onto a laboratory fabricated fused silica capillary strong cation exchange (Partisil SCX) column (350 µm I.D. x 5 cm, 10 µm particles). This column was directly connected to laboratory fabricated fused silica capillary C18 column (350 µm I.D. x 15 cm, Microsphere 3.5 µm particles). Columns were packed using a stainless steel packing cell. LC Packings Ultimate quaternary capillary LC system with a 20 µL injection loop was used for the sample and mobile phase delivery at 4 µL/min.

Mobile phase A was comprised of 94.9% optima grade water, 5% optima grade acetonitrile and 0.1% formic acid. Mobile phase B was comprised of 94.9% optima grade acetonitrile, 5% optima grade water and 0.1% formic acid. Mobile phase C was comprised of 300 mM ammonium acetate in optima grade water and mobile phase D was the same composition as mobile phase A.

Two-dimensional LC separation was performed with 12 isocratic salt elution where following salt concentrations of 0 mM, 5 mM, 10 mM, 15 mM, 20 mM, 30 mM, 40 mM, 50 mM, 100 mM, 150 mM 225 mM and 300 mM were obtained by appropriately mixing mobile phase C with mobile phase D. The 13<sup>th</sup> salt step was obtained by direct injection of 20  $\mu$ L of 3 M ammonium acetate.

Each salt elution was followed by a 110 min C18 reversed-phase gradient elution. Mobile phase composition was first held at the appropriate mix of solvents C and D for 15 min (0 to 15 min), then was switched to solvent A for additional 2 min (15 to 17 min). Gradient was linearly changed from 100% to 50% over 48 min (17 to 65 min) and from 50% to 20% over 20 min (65 to 85 min). Mobile phase composition was returned to 100% of solvent A in 5 min (85 to 90 min) and held at 100% for the next 20 min (90 to 110 min).

The voltage applied to electrospray source was 3.5 kV and 20 units of sheath gas flow. Inlet capillary was held at 35.0 V and 150 °C. Other parameters were as follows: tube lens offset at -15.0 V, multipole 1 offset at -7.5 V, lens voltage at -50.0 V, multipole 2 offset at -13.0 V, multipole RF amplitude peak-peak at 400 V, entrance lens at -50.0 V and trap DC offset at -10.0 V.

The mass spectrometer was operated in a data-dependent mode. Full scan from  $m/z$  200 to 2000 was acquired first. The most intense precursor ion was selected from the previous full MS scan and submitted to collisional induced dissociation (CID) with an activation  $Q$  of 0.250, activation time of 45 ms, and 35% normalized collision energy (NCE). This was followed by a MS/MS scan between  $m/z$  200 and 2000 of the most intense ion of the previous full MS scan. The same procedure was performed for three most intense ions that were excluded from further tandem experiments for one minutes.

Each analytical full scan was constructed as an average of 3 full scans, while each analytical MS/MS scan was constructed as an average of 5 MS/MS scans.

A maximum ion trap injection time of 300 ms was used for the full scans and 500 ms was used for MS/MS scans. The automatic gain control (AGC) target for the full scan was set to  $2 \times 10^7$  and for MS/MS to  $5 \times 10^7$ .

## Protein Data Analysis

All 13 RAW files of a MuDPIT run were uploaded to the MASCOT daemon software, combined as one file and searched against *Toxoplasma gondii* database (<http://toxodb.org/toxo/>). Parameters were set as followed: the cleavage enzyme was trypsin with three missed cleavage allowed. Peptide tolerance and the MS/MS tolerance were set to 1.0 Da. Oxidized methionine was chosen as a variable modification.

MudPIT scoring was used and the requirement for positive protein identification was bold red. This requires the protein hit to include at least one bold red peptide match indicating that the peptide is the highest scoring match to the MS/MS spectra.

## Results and Discussion

*Toxoplasma gondii* is intracellular parasite that has several life stages. Two stages that allow asexual expansion of the parasite are stages of tachyzoites and bradyzoites. Tachyzoite is a rapidly growing stage responsible for dissemination during acute infection while bradyzoite is a slow growing stage when tissue cysts are formed.

Utilizing methodology described in Chapter 2, a total of 394 N-glycosylated proteins, from tachyzoites ME49 strain, were identified with 57 proteins identified in all three replicates and 71 identified in at least two replicates (Figure 4.3).

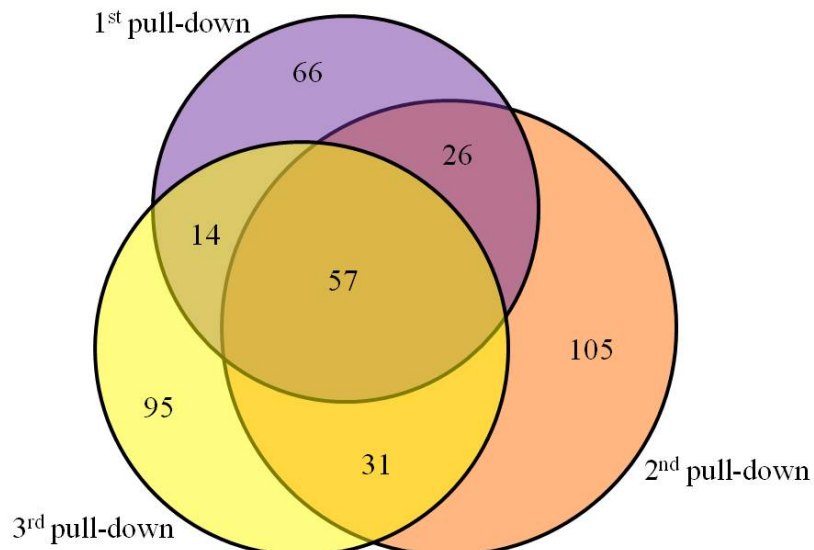


Figure 4.3. *Toxoplasma gondii* ME49, tachyzoites, pull-down triplicate. Purple circle: first pull-down replicate. Orange circle: second pull-down replicate. Yellow circle: third pull-down replicate.

Triplicate analysis shows that majority of proteins is identified in one out of three analyses. However, the closer look of the analyzed proteins reveals that the efficiency of the pull-down method depends on molecular weight of the protein (Figure 4.4). This was obtained by dividing all identified proteins in three groups: proteins identified in all three replicates, proteins identified in two out of three replicates and proteins identified in at least one replicate. Each group was further divided in three subgroups according to the molecular weight of the protein: MW < 30 kDa, 30kDa < MW < 100 kDa and 100 kDa <

MW. The number of proteins in each subgroup was normalized to the total number of proteins in each group and expressed as a percentage. The result of this analysis shows that proteins with molecular weight higher than 100 kDa are predominantly observed in one out of three replicates while proteins with molecular weight lower than 30 kDa are largely observed in all three replicates.

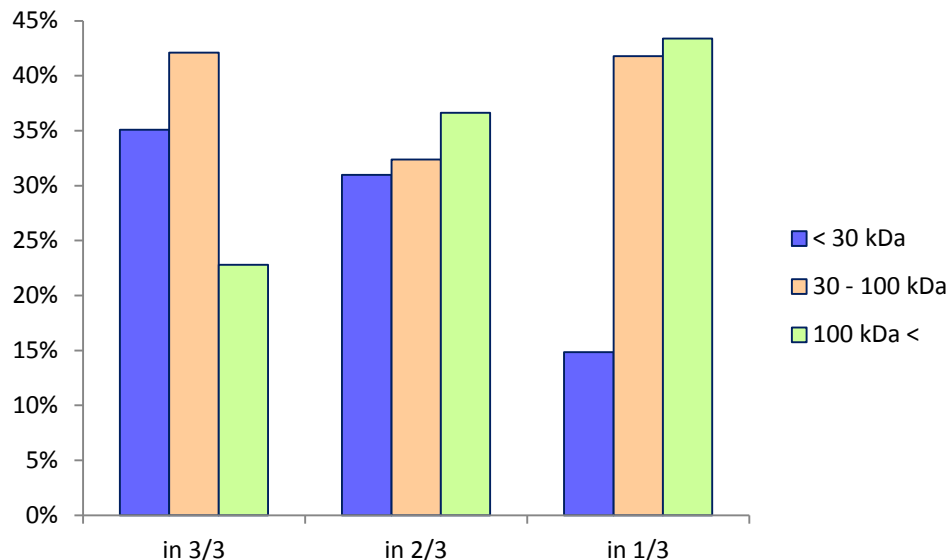


Figure 4.4. Pull-down method dependence on molecular weight of the protein. X-axis indicates in how many replicates were proteins of a certain molecular weight observed out of three replicates. Y-axis indicates the percentage of proteins of a certain molecular

This observation points out the possibility of pull-down method refinement in the future with the purpose to reduce the variability between replicates. To be more specific, after cell lysis proteins could be separated according to their molecular weight utilizing size-exclusion chromatography. However, discussion of results obtained in this dissertation is based on proteins identified in all three replicates.

In general, interpretation of data from a subproteome obtained by the enrichment, as in this case glycoproteome, will always suffer from some degree of contamination. For example, a glycoprotein could be interacting with the organelle or transported through the membrane at the time of extraction, and therefore, not be isolated by the enrichment procedure. Or a non-glycosylated protein could be interacting with the

glycoprotein, and therefore, be isolated with the enrichment procedure. For that reason, reproducible sample preparation and replicate analysis is crucial.

Therefore, the discussion of tachyzoite glycoproteome focuses on 57 proteins identified in all three replicates (table 3.1) indicating glycosylation as PTM. All other proteins can be found in the tables A.1 and A.2 in the Appendix.

#### Invasion related glycosylated proteins

Dense granules, rhoptry and micronemes are secretory organelles specialized for invasion and modification of the host cell. The entry of the parasite to the host cell starts with the adhesion proteins secreted by micronemes, which assist rhoptry proteins to form the moving junction resulting in parasitophorous vacuole.

Comparing the distribution of 57 identified proteins (table 4), obtained utilizing the methodology described in chapter 2, 8 are associated rhoptry (ROP 1, ROP 2A, ROP 7, ROP 8, ROP 11, ROP 15, ROP 40 and RON 4) and 4 with dense granule organelle (GRA 1, GRA 2, GRA 7, GRA 8). However, proteins associated with micronemes proteins were not identified.

Part of the moving junction macromolecular complex is rhoptry neck protein RON 4. RON 4 is exposed on the cytosolic side of the host cell during invasion and binds C-terminal region of  $\beta$ -tubulin of the host early in invasion. [87] All rhoptry proteins are associated with parasitophorous vacuole, either the ones injected through it into the host cell during invasion or responsible for establishment of parasitophorous vacuole.

Dense granule proteins also produce two types of proteins. One type is associated with the formation of specialized tubules through which the parasite acquires nutrients from the host. The other type is secreted into the host cell. [88]

Recently, a lot of effort has been put in evaluating vaccine antigens. [89, 90] Some of the candidates, such as GRA 1, GRA 7, ROP 1 and ROP 2A were identified in this analysis and listed in the table 4. All dense granule proteins induce a strong T and B cell responses of the host, but all infectious stages of *Toxoplasma gondii* express GRA 7

indicating it as a good candidate for the vaccine antigen. But even stronger immune response was obtained when GRA 7 was coupled with ROP 1. [89] ROP 1 is penetrating enhancing factor secreted into the nascent parasitophorous vacuole increasing the efficiency of invasion. Another couple for the vaccine antigen is ROP 2A and SAG 1. Glycosylation was found on ROP 2A, the protein that plays a role in association between the parasitophorous vacuole and host cell mitochondria. [91]

Out of 10 dense granule proteins, where all are associated with parasitophorous vacuole, GRA1, GRA2 and GRA8 were also identified utilizing method developed in chapter 2 of this dissertation. GRA 1 and GRA 2 are part of the tubulovesicular membranous network, the structure that allows small molecule trafficking. [92]

Serine-threonine phosphatase 2C (PP2C), found among glycosylated proteins, targets the host nucleus during the invasion facilitated by rhoptry secretion. However, the invasion completion is not required for delivery of this effector protein [93] suggesting a possible epigenetic alteration of the host protein expression. Also, serine-threonine phosphatase 2C is found as part of the toxofilin–actin–PP2C complex where it selectively dephosphorylates toxofilin at serine 53 promoting toxofilin’s affinity to G-actin. [94]

One more enzyme found among glycosylated proteins is protein disulfide isomerase. Protein disulfide isomerase is localized on the surface of *Toxoplasma gondii* where it regulates interactions between parasite and the host cell also making it a good candidate for the vaccine antigen. [95]

Elongation factor 1-alpha (EF-1-ALPHA) is a putative protein in *Toxoplasma gondii* but it is characterized in other members of phylum where it is also involved in invasion pathway. [96]

Considering all these findings, it is interesting observation that initial adhesion of the parasite to the host cell, regulated by micronemes proteins, does not involve glycosylated proteins, while the subsequent steps such as formation of parasitophorous vacuole as well as tubule formation proteins and secretory proteins are glycosylated indicating those glycoproteins as a viable target for future studies in successful infection prevention. Also, glycosylation is found in proteins not associated with invasion but with maintenance of parasite within the host cell.

## Energy related glycosylated proteins

Several proteins associated with energy pathways were identified as glycosylated. One of these is adenosylhomocysteinase. This is a putative protein with catalytic activity and probably involved in energy transfer due to NAD-binding domains. Adenylate kinase, another putative protein, is a phosphotransferase. Dihydrolipoyllysine-residue succinyltransferase component of oxoglutarate dehydrogenase is acyltransferase that takes part in tricarboxylic acid cycle. Glyceraldehyde-3-phosphate dehydrogenase GAPDH1 is one of the key glycolytic enzymes. Another putative protein is GDP mannose 4,6-dehydratase that catalyzes hydrolysis of GDP-mannose to GDP-4-dehydro-6-deoxy-D-mannose. These findings indicate involvement of glycosylation in parasites metabolic pathways.

## DNA and translational machinery related glycosylated proteins

A number of DNA and translational machinery related proteins were identified as glycosylated.

One of DNA related proteins is high mobility group (HMG) box domain-containing protein class II, found in many eukaryotic chromosomal proteins and transcription factors, is a non-histone chromosomal protein that binds to the bent or distorted DNA.

SWI2/SNF2-containing PHD finger protein is a zinc-finger that exerts helicase activity, DNA binding and ATP binding activity. PHD finger protein from other organisms play a role in chromatin remodeling processes and histone acetylation. [97]

Histones are small, highly alkaline proteins responsible for packing DNA in nucleosomes. Epigenetic gene regulation heavily relies on posttranslational modifications of histones. Histones, H2A, H2B, and H4 are predicted for *Toxoplasma gondii*. The study described in this dissertation reveals glycosylation on H2A1, H2AZ, H4, and, surprisingly, H2Bb. H2B isoforms are expressed depending on the life cycle



stage. H2Ba is expressed in tachyzoites while H2Bb is expressed only during the sexual stage of the life cycle but not in tachyzoites nor bradyzoites. H4 is highly conserved while H2A and H2B are prone to changes due to their exposed N-terminal tail. [98]

Proteins involved in translation seem to have a glycosylated member at every stage of the process from initiation (eIF-5A, putative) through elongation (elongation factor 2 family protein, putative) to regulation. Prolyl-tRNA synthetase (ProRS) and tryptophanyl-tRNA synthetase (TrpRS2) are aminoacyl-tRNA synthetase located in cytoplasm involved in regulation of translational fidelity and tryptophanyl-tRNA aminoacylation respectively. Ribosomal protein RPL24 is located on the surface of the large subunit of a ribosome. RPL24 is involved in kinetics of peptide synthesis and probably in interactions between the large and small subunits. Ribosomal protein RPS13 is essential for binding of the small ribosomal subunit to intersubunit of a ribosome.

#### Other identified glycosylated proteins

GTP-binding protein putative regulates G proteins (guanosine nucleotide-binding proteins) that are known molecular switches.

Three chaperones were also identified as glycosylated proteins. Two of them are heat shock homologs HSP 70 and HSP 90 and the third is chaperonin protein BiP. HSP 70 is highly expressed by the host cells where inhibits mitochondrial apoptosis of the host. [99] HSP 90 chaperone machinery is involved in parasite development and can be found in a complex with HSP 70. [100] Since HSP 70 and HSP 90 can form a complex it is unclear if only one of the pair is glycosylated while the other is extracted only due to complex formation. Chaperonin protein BiP is found in endoplasmatic reticulum where it is involved in controlling protein folding and exit from the endoplasmatic reticulum. [101]

Perhaps the most important finding in this study is the identification of hypothetical proteins. In total 14 hypothetical proteins were identified not only in all three replicates but also majority of them with the high score.

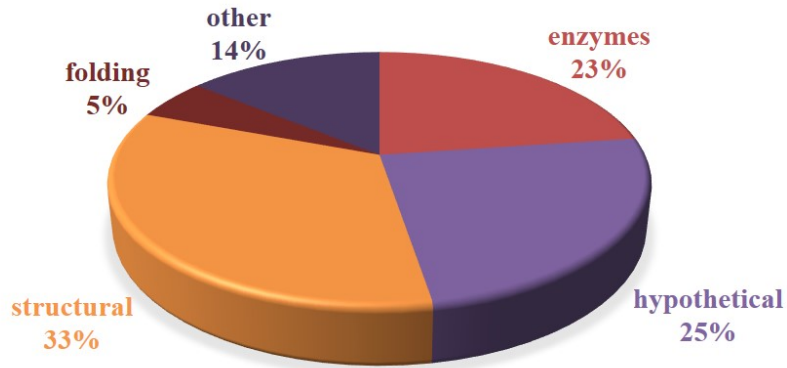


Figure 4.5. Distribution of identified glycoproteins according to function.

To easier comprehend the current body of knowledge of glycosylation the results are presented in a pie chart (Figure 4.5) having in mind the function of identified glycoproteins. One third of N-glycosylated proteins are structural proteins mainly involved in invasion and establishment of infection pointing to a group of proteins already under consideration as a vaccine antigen candidates. At the same time a prominent group of hypothetical proteins reveals that a field of glycoproteomics of *Toxoplasma gondii* is still in its infancy offering a chance for improvement, which was the goal of this dissertation.

Table 4.1. Proteins identified in three out of three replicates.

protein #	protein name	ID #	mass	Mowse score	queries matched
1	actin ACT1 (ACT1)	TGME49_209030	41907	532	52
2	actin depolymerizing factor ADF (ADF)	TGME49_220400	12922	46	17
3	adenosylhomocysteinase, putative	TGME49_225050	64105	39	53
4	adenylate kinase, putative	TGME49_224900	28560	100	25
5	alpha tubulin TUBA1 (TUBA1)	TGME49_316400	50113	223	35
6	ATPase family associated with various cellular activities (AAA) domain-containing protein	TGME49_230830	540153	35	195
7	chaperonin protein BiP	TGME49_311720	73252	286	52
8	dense granule protein GRA2 (GRA2)	TGME49_227620	19804	554	78
9	dense granule protein GRA7 (GRA7)	TGME49_203310	25919	160	29
10	dense granule protein GRA8 (GRA8)	TGME49_254720	28628	34	6
11	dihydrolipoyllysine-residue succinyltransferase component of oxoglutarate dehydrogenase	TGME49_219550	50125	150	44

Table 3.1. (continued)

12	elongation factor 1-alpha (EF-1-ALPHA), putative	TGME49_286420	49005	1111	128
13	GDP mannose 4,6- dehydratase, putative	TGME49_238940	41717	53	19
14	glyceraldehyde-3- phosphate dehydrogenase GAPDH1 (GAPDH1)	TGME49_289690	53339	60	50
15	GTP-binding protein, putative	TGME49_214350	44547	38	26
16	heat shock protein HSP70 (HSP70)	TGME49_273760	72880	438	51
17	heat shock protein HSP90 (HSP90)	TGME49_288380	81932	150	44
18	histone H2A1	TGME49_261250	19546	321	62
19	histone H2AZ	TGME49_300200	15920	373	98
20	histone H2Bb	TGME49_251870	12575	197	55
21	histone H4	TGME49_239260	11439	397	41
22	HMG (high mobility group) box domain- containing protein	TGME49_210408	10682	91	75
23	hypothetical protein	TGME49_294430	100125	53	58
24	hypothetical protein	TGME49_321680	40660	45	30
25	hypothetical protein	TGME49_297430	25336	70	20
26	hypothetical protein	TGME49_221470	61517	99	33
27	hypothetical protein	TGME49_275650	101158	92	50
28	hypothetical protein	TGME49_279420	144941	69	67
29	hypothetical protein	TGME49_212860	107698	53	62
30	hypothetical protein	TGME49_240080	122320	51	74

Table 3.1. (continued)

31	hypothetical protein	TGME49_268760	31988	36	31
32	hypothetical protein	TGME49_268260	86736	51	60
33	hypothetical protein	TGME49_223640	180412	35	89
34	hypothetical protein	TGME49_310970	370390	40	234
35	hypothetical protein	TGME49_315570	9273	140	26
36	hypothetical protein	TGME49_248160	122417	52	71
37	inositol polyphosphate kinase	TGME49_268920	50014	36	33
38	NAC domain-containing protein	TGME49_205558	20507	66	21
39	prolyl-tRNA synthetase (ProRS)	TGME49_219850	93478	74	73
40	protein disulfide isomerase	TGME49_211680	52801	247	32
41	roptry kinase family protein ROP11 (incomplete catalytic triad) (ROP11)	TGME49_227810	58034	97	27
42	roptry kinase family protein ROP40 (incomplete catalytic triad) (ROP40)	TGME49_291960	57947	160	35
43	roptry neck protein RON4 (RON4)	TGME49_229010	107174	65	36
44	roptry protein ROP1 (ROP1)	TGME49_309590	47994	259	56
45	roptry protein ROP15 (ROP15)	TGME49_211290	36588	146	35

Table 3.1. (continued)

46	rhopty protein ROP2A (ROP2A)	TGME49_215785	49154	184	41
47	rhopty protein ROP7 (ROP7)	TGME49_295110	63394	52	18
48	rhopty protein ROP8 (ROP8)	TGME49_215775	65912	126	41
49	ribosomal protein RPL24 (RPL24)	TGME49_244320	32693	35	21
50	ribosomal protein RPS13 (RPS13)	TGME49_270380	17176	84	22
51	serine-threonine phosphatase 2C (PP2C)	TGME49_231850	36790	195	22
52	SWI2/SNF2-containing PHD finger protein	TGME49_236970	277892	35	160
53	translation elongation factor 2 family protein, putative	TGME49_205470	106571	104	62
54	translation initiation factor eIF-5A, putative	TGME49_251810	17468	323	25
55	tryptophanyl-tRNA synthetase (TrpRS2)	TGME49_288360	77057	67	46
56	dense granule protein GRA1 (GRA1)	TGME49_270250	20225	52	10
57	UBA/TS-N domain- containing protein	TGME49_270595	849189	46	320

## Bradyzoites related glycosylated proteins

The analysis of protein glycosylation of bradyzoites (table 3.2) mostly resulted in identification of 28 proteins out which 9 are hypothetical and 6 are putative. Identified proteins are mainly enzymes associated with translation and transcription and protein domains. However, very little is known about the role of these proteins in *Toxoplasma gondii* and their function, structure or location is mostly inferred by homology.

For example, DnaJ domain-containing protein is identified but its function is not known. However, since DnaJ proteins are highly conserved in eukaryotes their function can be inferred for *Toxoplasma gondii*. DnaJ in complex with HSP 40 play crucial roles in protein translation, folding, unfolding, translocation, and degradation.

Transcription is a complex process that involves many enzymes as herein identified UvrD/REP helicase domain-containing protein involved in DNA damage and repair. Transcription is initiated by RNA polymerase II apparatus but the role of identified mediator complex subunit MED14 is not explained for *Toxoplasma gondii*. In humans this complex is a component of the mediator complex responsible for transcription regulation of almost all RNA polymerase II-dependent genes. [102] Epsilon subunit of eukaryotic initiation factor-2B is identified as putative but it has an important role in regulating mRNA translation in eukaryotes. TrmH family of proteins is involved in tRNA recognition and one of the members, RNA methyltransferase, is also identified. [103]

Mitochondrial proteins responsible for protein import are found in the earliest eukaryotes. For that reason, even though mitochondrial inner membrane translocase subunit TIM17 is identified in this study as putative, its function can be narrowed down to either facilitator of the translocation of proteins across the inner membrane or facilitator of the insertion of proteins into the inner mitochondrial membrane. [104]

Reversible protein phosphorylation is used for cellular information processing and has to be tightly controlled. That is easier achieved by formation of complexes of kinases and other proteins. One of such complexes is CMGC kinase group, identified here as putative. [105]

As a part of S8 peptidase family, subtilisin SUB2 is a protease that contains a catalytic triad of serine, aspartate and histidine. [106]

Sec1 family protein is a part of Sec1/Munc18 (SM) family proteins that are essential for every vesicle fusion pathway. [107]

GYF domains are present in most eukaryotic species and recognize proline-rich sequences. Even though not fully elucidated for *Toxoplasma gondii*, GYF domain-containing protein is identified in this study. [108]

WD domain is found as a repeat in a large number of diverse eukaryotic proteins but what they have in common is folding into beta propeller that serves as a part of the platform used for reversible assembly of many protein complexes in almost all the major pathways. G-beta repeat-containing protein is identified.

Myosin K, identified in this study, is a part of motility system of *Toxoplasma gondii*, utilized for invasion as well as the escape from the host cells but also for general motility. [109]

Structural proteins such as clathrin and trichohyalin are also identified in this study but as putative. In humans clathrin is predominant component of polyhedral coat of coated pits and vesicles. Components of clathrin are 3 heavy chains and 3 light chains. Trichohyalin is responsible for organizing the intermediate filaments in human hair follicle inner root giving the shape to the hair fiber. [110]

Even more importantly, out of 28 identified proteins 9 are hypothetical leaving a large part of glycosylated proteins under mysterious veil impossible even to infer. Interestingly one of the hypothetical proteins was identified with the highest score comparing to all other identified proteins.

However, one should keep in mind that these results are obtained from only one sample and not confirmed with replicates due to the lack of material. This sample was obtained from brains of 26 mice with the total number of  $7 \times 10^4$  cysts. Total amount of protein was 1.6 mg and that amount was too small to divide in replicates.

Also, results in the table 4.2 show low scores for the most proteins identified by the Mascot search engine. Even though all proteins above the threshold are identified with 95% confidence, proteins close to the threshold still have a high probability of being a random sequence.



Putative proteins recognized in other organisms need to be further confirmed and characterized by studying mutant parasites lacking the relevant gene and identifying the proteins involved in their pathways.

## Conclusion

Even though ME49 strain is most prevalent in clinically diagnosed cases of toxoplasmosis, most proteomic studies are performed on RH strain because large quantities are easily obtained in tissue culture. Moreover, bradyzoite stage of parasites life cycle, which develop in mice brains after approximately two to four weeks, is significantly less studied comparing to rapidly growing tachyzoite stage. Considering all this, it is not surprising that the glycosylation machinery and its effects in *Toxoplasma gondii* are still largely unknown. To that end, the aim of this study is to contribute to current knowledge database by identifying the possible candidates for the future studies.

In total 394 glycosylated tachyzoite proteins and 28 glycosylated bradyzoite proteins were identified utilizing the method developed in the chapter 2 of this dissertation. Due to sample availability tachyzoite proteins were analyzed in triplicate and 57 out of 394 were identified in all replicates suggesting glycosylation as their post-translational modification. However, bradyzoites were studied utilizing only one sample due to the time required for their generation and the cost. Nevertheless, 28 proteins were identified. The lower number does not necessarily indicate low glycosylation but could also be due to low sample amount where majority of proteins are below the identification threshold of mass spectrometer. For that reason, more experimental proof is required.

Table 4.2. Proteins identified in tissue cysts.

protein #	protein name	ID #	mass	Mowse score	queries matched
1	hypothetical protein	TGME49_248160	122417	126	110
2	mediator complex subunit MED14 (MED14)	TGME49_229310	464457	41	165
3	myosin K	TGME49_206415	262859	38	169
4	Sec1 family protein	TGME49_271060	72640	37	47
5	trichohyalin, putative	TGME49_242790	200679	37	102
6	DnaJ domain-containing protein	TGME49_313310	65713	35	60
7	histone H4	TGME49_239260	11439	35	20
8	hypothetical protein	TGME49_219218	175383	35	74
9	alpha-tubulin N-acetyltransferase, putative	TGME49_319600	98566	35	68
10	hypothetical protein	TGME49_253020	51621	34	26
11	UvrD/REP helicase domain-containing protein	TGME49_277550	342488	34	140
12	clathrin heavy chain, putative	TGME49_290950	194501	34	66
13	WD domain, G-beta repeat-containing protein	TGME49_213060	328236	34	127
14	eukaryotic initiation factor-2B, epsilon subunit, putative	TGME49_272640	92848	34	75
15	mitochondrial inner membrane translocase subunit TIM17, putative	TGME49_312220	22612	33	21

Table 3.2. (continued)

16	DUF367 domain-containing protein	TGME49_204540	63641	33	50
17	CMGC kinase, putative	TGME49_250850	286625	33	123
18	subtilisin SUB2 (SUB2)	TGME49_314500	141581	33	77
19	hypothetical protein	TGME49_224830	121076	33	68
20	WD domain, G-beta repeat-containing protein	TGME49_272040	358195	33	91
21	KRUF family protein	TGME49_251180	29178	33	20
22	GYF domain-containing protein	TGME49_298610	237893	33	132
23	hypothetical protein	TGME49_306440	96284	33	37
24	hypothetical protein	TGME49_231815	431546	33	240
25	hypothetical protein	TGME49_210700	926533	32	420
26	hypothetical protein	TGME49_259870	230595	32	128
27	hypothetical protein	TGME49_233430	410963	32	207
28	RNA methyltransferase, TrmH family protein	TGME49_216390	89523	32	46

## CHAPTER 5

### CONCLUSION

*Toxoplasma gondii* is intracellular parasite that has several life stages. Humans are impacted by the two stages that allow asexual expansion of the parasite: tachyzoite and bradyzoite stages. Tachyzoite is a rapidly growing stage responsible for dissemination during acute infection while bradyzoite is a slow growing stage when tissue cysts are formed.

Acute infection invokes the response of the host's immune system and can be treated with combinations of antibiotics and antimalarial medications. Even though a large majority of tachyzoites are eliminated this way, a very few are required for the establishment of bradyzoites enclosed within the tissue cysts and responsible for the chronic phase of infection when the parasite becomes invisible for the host's immune system. This, so called, "dormant phase" has only recently been discovered to interfere with the host's behavioral changes, most probably by altering neurotransmitter signal transduction and increasing the production of dopamine. [83]

Although there are three types of lineages of *Toxoplasma gondii* (type I (RH and GT-1 strains), type II (ME49 strain) and type III (VEG strain)), the clinical cases of toxoplasmosis are most commonly due to type II (ME49) isolates. ME49 strain, while significantly less virulent in mice, readily forms the cysts. [111] For that reason ME49 is chosen for the study described in this dissertation.

Inferring from the genome of *Toxoplasma gondii*, it is believed that the parasite can express about 6500 proteins. But 72 already identified proteins have no identified transcripts. This can be attributed to posttranslational modifications that often have a functional significance and can help to understand the biology of the parasite. [112] However, glycosylation of the proteome of *Toxoplasma gondii* is still largely unknown with even more scarce evidence of N-glycosylation. Studies performed on RH strain of the parasite identified 26 N-glycosylated proteins indicating involvement of N-glycosylation in gliding motility and host cell invasion. [85] Three of those 26 glycoproteins were identified utilizing method developed in this dissertation. Two rhoptry proteins of tachyzoites are associated with the invasion of the host cell, but the

third protein DnaJ was identified only in bradyzoites that were not analyzed in above mentioned study.

However, one striking structural difference between tachyzoites and bradyzoites raises the question if the glycosylation plays a significant role in bradyzoite stage of parasite's life. That difference is the location of Golgi complex.

In bradyzoites both endoplasmatic reticulum and Golgi complex are located near the basal end of the parasite. However, in tachyzoites Golgi complex is located near to the apical end with numerous organelles separating it from endoplasmatic reticulum. [83] Considering that N- and O-glycosylation take place in both endoplasmatic reticulum and Golgi complex, it would be hard to comprehend the significance of organelle separation if glycosylation played an important role in tachyzoites. However, bradyzoites, developed from tachyzoites in the later stage of parasites life, have Golgi complex and endoplasmatic reticulum in close proximity. One could speculate the importance of this movement and wonder if N- and/or O-glycosylation play a significant role in cyst formation or maintenance.

In general, the array of glycans displayed on the surface of various viruses (HIV) or even cell types (blood group) can silence or induce the immune response of the host. Tachyzoite stage of *Toxoplasma gondii* is recognized by the host and induces the immune response that can be treated but bradyzoites are not recognized. Current treatment of tachyzoites is not efficient enough and some of the surviving tachyzoites evolve into bradyzoite stage. It is speculated that the parasite recruits some of the host proteins and uses them for its own growth and maintenance but that could also be the case with the glycans. If the host's glycans are displayed at the surface of the cyst the immune system would not recognize it as a foreign object.

Figure 4.2, in chapter 4, represents the bradyzoites containing tissue cyst stained with two lectins, DBA and Con A. DBA binds to N-acetyl-D-galactosamine (GalNAc) residues typical for O-linked glycoproteins but not found in N-linked glycoproteins, while Con A specifically recognizes the trimannoside core of N-linked glycans. It is obvious from the figure 4.2 that O-linked glycans are not observed anywhere within the cyst but are abundantly present on the cyst wall. If the O-linked glycans were produced or transported in any vesicle within the cyst the green stain would be obvious, but it is

not. This raises the question if O-linked glycans are recruited from the host thus making a parasite invisible to the host's immune system.

Glycosylation plays a crucial role in cell signaling. Utilizing the method described in this dissertation, N-glycosylated proteins are identified in both tachyzoites and bradyzoites but only one glycoprotein is found in both groups and it is not yet annotated. Considering parasite's different morphology and the growth rate in these two life stages, different signaling pathways are easily envisioned. However, the glycoprotein present in the both groups could be the key to understanding of the transition from tachyzoites to bradyzoites and provide the good target to prevent cyst formation and eliminate the parasite before it becomes invisible to the host's immune system.

In my opinion the key to successful eradication of infection caused by *Toxoplasma gondii* lays not so much in the recognition of the differences between tachyzoites and bradyzoites but their similarities. If the tissue cysts recruit and display on their surface components derived from the host's organism, damaging the tissue cyst would inevitably damage the host as well. However, if one would identify components within the tachyzoites, while the parasite is still visible to the immune system and thus different then the host, then those components could possibly become the target for the treatment that would be least damaging for the host. It would be fortunate to find such components among the proteins that are unique to the parasite but not present in the host. It was such notion that drove the development of this dissertation.

Due to the growing body of evidence pointing out the importance of glycosylation of tachyzoites, the methods described in this dissertation were developed in an effort to provide an improvement of purification and enrichment of glycoproteins that subsequently could point out the candidate glycoproteins for future studies that will shed the light on the biology of *Toxoplasma gondii* and thus help develop means for toxoplasmosis eradication.

APPENDIX

Table A.1. Proteins identified in two out of three replicates.

protein #	protein name	ID #	mass	Mowse score	queries matched
1	AMP-binding enzyme domain-containing protein	TGME49_247760	86986	40	37
2	calmodulin, putative	TGME49_269442	9780	75	10
3	cyclin protein	TGME49_293280	273014	56	150
4	DDHD domain-containing protein	TGME49_313600	161510	32	108
5	DEAD (Asp-Glu-Ala-Asp) box polypeptide DDX39 (DDX39)	TGME49_216860	49113	51	24
6	dense granule protein GRA5 (GRA5)	TGME49_286450	12838	92	11
7	dihydropteroate synthase	TGME49_259550	82684	60	28
8	elongation factor 1-gamma, putative	TGME49_300140	44039	36	18
9	eukaryotic initiation factor-4A, putative	TGME49_250770	46673	58	47
10	GTP-binding nuclear protein ran/tc4	TGME49_248340	25997	41	22
11	heat shock protein 90, putative	TGME49_244560	96822	79	62
12	histone H2Ba	TGME49_305160	12602	197	55
13	histone H2Bv	TGME49_209910	13732	197	44
14	histone H3	TGME49_261240	15402	91	52
15	hypothetical protein	TGME49_232120	222553	34	73
16	hypothetical protein	TGME49_312420	117806	44	50
17	hypothetical protein	TGME49_278960	58350	34	30

Table A.1. (continued)

18	nucleosome assembly protein (nap) protein	TGME49_244110	48586	38	25
19	PEP-carboxykinase I	TGME49_289650	75342	84	38
20	rhoptry protein ROP5 (ROP5)	TGME49_308090	61101	71	57
21	ribosomal protein RPS14 (RPS14)	TGME49_263700	16301	53	33
22	ribosomal protein RPS30 (RPS30)	TGME49_291850	6600	66	13
23	SF-assemblin/beta giardin protein	TGME49_307840	23344	37	21
24	thioredoxin, putative	TGME49_209950	47699	43	47
25	miconeme protein MIC1 (MIC1)	TGME49_291890	48629	37	23
26	MRP family domain-containing protein	TGME49_318590	70666	57	27
27	heat shock protein HSP60 (HSP60)	TGME49_247550	60913	107	49
28	HECT-domain (ubiquitin- transferase) domain-containing protein	TGME49_275630	185550	39	64
29	histone H3 centromeric CENH3	TGME49_225410	23282	47	39
30	histone H3.3	TGME49_218260	15413	91	51
31	hypothetical protein	TGME49_321410	212035	38	124
32	hypothetical protein	TGME49_221220	213039	35	100
33	hypothetical protein	TGME49_297210	444121	32	216
34	hypothetical protein	TGME49_215980	24466	34	16
35	PIK3R4 kinase-related protein (incomplete catalytic triad)	TGME49_310190	317227	32	86
36	protein kinase	TGME49_226540	54688	93	59



Table A.1. (continued)

37	regulator of chromosome condensation (RCC1) repeat-containing protein	TGME49_310290	80702	40	74
38	rhoptry protein ROP4 (ROP4)	TGME49_295125	42389	66	20
39	ribosomal protein RPL10A (RPL10A)	TGME49_215470	24603	33	20
40	SAG-related sequence SRS34A (SRS34A)	TGME49_271050	19106	59	6
41	anonymous antigen-1, putative	TGME49_312630	286215	69	185
42	AP2 domain transcription factor AP2VIII-4 (AP2VIII4)	TGME49_272710	356669	43	242
43	ATPase family associated with various cellular activities (AAA) subfamily protein	TGME49_242625	976079	36	505
44	DEAD/DEAH box helicase domain-containing protein	TGME49_203840	196955	34	149
45	dynein light chain, putative	TGME49_285350	25051	37	27
46	dynein, axonemal, heavy chain 2 family protein	TGME49_235920	514463	49	335
47	EF-1 guanine nucleotide exchange domain-containing protein	TGME49_226410	36049	56	23
48	histone acetyltransferase TAF1/250	TGME49_276180	296268	33	151
49	histone family DNA-binding protein	TGME49_227970	25188	66	14
50	histone H2AX	TGME49_261580	14510	35	28
51	hypothetical protein	TGME49_221280	439285	33	284
52	hypothetical protein	TGME49_240060	88532	69	48

Table A.1. (continued)

53	hypothetical protein	TGME49_239410	256130	39	198
54	hypothetical protein	TGME49_205130	518657	37	277
55	hypothetical protein	TGME49_203980	201277	44	78
56	hypothetical protein	TGME49_230350	138895	35	79
57	hypothetical protein	TGME49_228070	441563	37	240
58	hypothetical protein	TGME49_315610	15971	132	48
59	hypothetical protein	TGME49_316470	363875	36	148
60	hypothetical protein	TGME49_276860	55516	34	37
61	NAC domain-containing protein	TGME49_257090	38782	56	26
62	poly(ADP-ribose) glycohydrolase	TGME49_280380	61836	47	42
63	Pumilio-family RNA binding repeat-containing protein	TGME49_260600	172516	37	81
64	ribosomal protein RPL12 (RPL12)	TGME49_254440	17821	38	14
65	ribosomal protein RPP2 (RPP2)	TGME49_309810	11768	43	21
66	ribosomal protein RPS12 (RPS12)	TGME49_205340	15399	67	9
67	RNA methyltransferase	TGME49_218580	310819	73	111
68	RNA recognition motif- containing protein	TGME49_262620	31760	58	63
69	SAG-related sequence SRS22F (SRS22F)	TGME49_238500	19504	39	9
70	translation initiation factor IF-2, putative	TGME49_214270	126067	35	91
71	XRN 5'-3' exonuclease N- terminus protein	TGME49_242830	223762	35	107

Table A.2. Proteins identified in one out of three replicates.

protein #	protein name	ID #	mass	Mowse score	queries matched
1	2-oxo acid dehydrogenases acyltransferase (catalytic domain) domain-containing protein	TGME49_31 9920	70289	34	35
2	3-hydroxyisobutyrate dehydrogenase	TGME49_26 3430	34159	35	26
3	5'-3' exonuclease, N-terminal resolvase family domain-containing protein	TGME49_28 4010	124798	34	75
4	ABC transporter, putative	TGME49_20 8050	381288	39	224
5	ABC1 family protein	TGME49_21 3620	211215	41	135
6	acetyltransferase, GNAT family protein	TGME49_30 5450	56862	40	36
7	adaptin c-terminal domain-containing protein	TGME49_27 2600	182951	32	93
8	AMP-binding enzyme domain-containing protein	TGME49_31 0080	87847	42	46
9	AMP-binding enzyme domain-containing protein	TGME49_21 9230	87016	36	34
10	AP2 domain transcription factor AP2III-4 (AP2III4)	TGME49_29 9020	171537	38	112
11	AP2 domain transcription factor AP2V-2 (AP2V2)	TGME49_28 5895	365853	38	167
12	arginyl-tRNA synthetase	TGME49_27 0690	130254	34	99

Table A.2. (continued)

13	argonaute AGO (AGO)	TGME49_31 0160	105190	34	75
14	beta tubulin	TGME49_26 6960	50073	61	44
15	beta-tubulin, putative	TGME49_22 1620	50037	61	45
16	bifunctional GMP synthase/glutamine amidotransferase protein	TGME49_23 0450	62294	34	41
17	Brf1p family coiled coil protein	TGME49_23 2440	54044	34	76
18	calcium dependent protein kinase CDPK7 (CDPK7)	TGME49_22 8750	219083	34	95
19	calcium dependent protein kinase CDPK8 (CDPK8)	TGME49_29 2055	161419	36	77
20	cAMP-dependent protein kinase regulatory subunit	TGME49_24 2070	42796	35	42
21	cell-cycle-associated protein kinase, putative	TGME49_28 1450	50168	36	44
22	ClpB, putative	TGME49_27 5690	114549	36	80
23	cold-shock DNA-binding domain- containing protein	TGME49_32 0600	23060	87	21
24	CRIP1, putative	TGME49_25 0350	10076	32	7
25	cyclin protein	TGME49_29 3280	273014	68	181
26	cyclophilin	TGME49_22 1210	19604	48	24

Table A.2. (continued)

27	cyclophilin precursor	TGME49_20 5700	38196	49	31
28	D-3-phosphoglycerate dehydrogenase	TGME49_23 9820	65193	58	3
29	DEAD/DEAH box helicase domain- containing protein	TGME49_26 4160	120180	33	62
30	DEAD-family helicase	TGME49_29 8020	130966	33	80
31	dense granule protein GRA3 (GRA3)	TGME49_22 7280	23875	58	16
32	DNA-dependent RNA polymerase (POLRMT)	TGME49_24 6060	232348	37	123
33	DnaJ family chaperone, putative	TGME49_31 1240	47448	33	39
34	DnaK family protein	TGME49_22 6830	103191	41	49
35	elongation factor TS protein	TGME49_20 9010	52348	37	26
36	endonuclease/exonuclease/phosphat ase domain-containing protein	TGME49_23 8400	290205	40	141
37	endonuclease/exonuclease/phosphat ase family protein	TGME49_25 9560	164511	42	54
38	enolase 1	TGME49_26 8860	48341	36	25
39	enolase 2	TGME49_26 8850	52113	36	32
40	ERCC4 domain-containing protein	TGME49_30 5310	223459	40	136

Table A.2. (continued)

41	ethylene-responsive RNA helicase, putative	TGME49_31 3240	85533	34	43
42	eukaryotic aspartyl protease superfamily protein	TGME49_20 9620	43488	33	16
43	eukaryotic initiation factor-2 gamma, putative	TGME49_23 5970	50508	39	26
44	eukaryotic translation initiation factor 3 subunit 7, putative	TGME49_31 7720	68766	39	62
45	Fe-S metabolism associated domain-containing protein	TGME49_27 7010	36177	32	31
46	FHA domain-containing protein	TGME49_26 7600	84416	33	47
47	FUSE-binding protein 2 / KH-type splicing regulatory protein (FUBP2)	TGME49_21 6670	100093	43	99
48	glucose-6-phosphate 1- dehydrogenase	TGME49_29 4200	62736	39	47
49	glutamine synthetase, type I, putative	TGME49_27 3490	65146	35	49
50	glutaminyl-tRNA synthetase (GlnRS)	TGME49_21 7460	95526	34	57
51	glycine cleavage T-protein (aminomethyl transferase) domain- containing protein	TGME49_28 3820	195707	36	96
52	GTPase RAB7 (RAB7)	TGME49_24 8880	47712	35	33
53	guanylyl cyclase	TGME49_25 4370	477032	39	161
54	HEAT repeat-containing protein	TGME49_24 4040	132694	32	39

Table A.2. (continued)

55	HEAT repeat-containing protein	TGME49_22 9180	136163	39	595
56	heat shock protein	TGME49_25 1780	78250	38	85
57	HECT-domain (ubiquitin-transferase) domain-containing protein	TGME49_29 5710	851243	46	475
58	helicase, putative	TGME49_20 9770	150821	35	68
59	helix-hairpin-helix motif domain-containing protein	TGME49_28 5490	117256	37	64
60	histone lysine methyltransferase, SET, putative	TGME49_20 1250	177893	39	114
61	HMG (high mobility group) box domain-containing protein	TGME49_21 9828	11381	50	34
62	hypothetical protein	TGME49_32 6600	17775	32	46
63	hypothetical protein	TGME49_29 4860	254850	38	109
64	hypothetical protein	TGME49_29 5370	29729	32	21
65	hypothetical protein	TGME49_29 3252	166690	41	64
66	hypothetical protein	TGME49_29 3560	229122	37	66
67	hypothetical protein	TGME49_20 9470	57616	39	18
68	hypothetical protein	TGME49_22 0890	51195	36	28

Table A.2. (continued)

69	hypothetical protein	TGME49_22 1560	58420	50	54
70	hypothetical protein	TGME49_22 2060	103670	34	70
71	hypothetical protein	TGME49_22 2250	167959	36	109
72	hypothetical protein	TGME49_25 3990	231484	41	101
73	hypothetical protein	TGME49_29 9270	119375	34	65
74	hypothetical protein	TGME49_29 9130	346626	38	210
75	hypothetical protein	TGME49_27 6200	448604	36	317
76	hypothetical protein	TGME49_21 1700	273962	35	124
77	hypothetical protein	TGME49_26 5470	43627	36	39
78	hypothetical protein	TGME49_26 5280	133185	37	66
79	hypothetical protein	TGME49_26 4890	199927	36	106
80	hypothetical protein	TGME49_27 9380	34769	73	24
81	hypothetical protein	TGME49_28 9520	355979	39	137
82	hypothetical protein	TGME49_28 9970	24448	53	24



Table A.2. (continued)

83	hypothetical protein	TGME49_29 0310	200813	34	138
84	hypothetical protein	TGME49_29 0620	296416	37	187
85	hypothetical protein	TGME49_29 2360	117449	35	47
86	hypothetical protein	TGME49_26 7670	90813	35	51
87	hypothetical protein	TGME49_21 3635	325486	33	145
88	hypothetical protein	TGME49_28 6510	53152	35	20
89	hypothetical protein	TGME49_28 5820	19055	35	21
90	hypothetical protein	TGME49_24 0980	151943	33	83
91	hypothetical protein	TGME49_24 3310	132517	32	82
92	hypothetical protein	TGME49_24 3780	91568	42	47
93	hypothetical protein	TGME49_24 4120	248641	34	194
94	hypothetical protein	TGME49_23 9340	211263	36	145
95	hypothetical protein	TGME49_20 5320	70849	33	36
96	hypothetical protein	TGME49_20 3362	24667	36	16

Table A.2. (continued)

97	hypothetical protein	TGME49_20 3090	60624	34	63
98	hypothetical protein	TGME49_20 2170	94060	33	32
99	hypothetical protein	TGME49_26 1740	14158	40	10
100	hypothetical protein	TGME49_26 1390	28567	34	14
101	hypothetical protein	TGME49_25 8670	187479	41	94
102	hypothetical protein	TGME49_25 8420	259194	35	101
103	hypothetical protein	TGME49_25 8170	43915	35	27
104	hypothetical protein	TGME49_26 3200	118612	35	61
105	hypothetical protein	TGME49_26 3080	14407	91	10
106	hypothetical protein	TGME49_23 0950	240359	37	91
107	hypothetical protein	TGME49_23 1200	298665	37	141
108	hypothetical protein	TGME49_22 9390	103131	39	40
109	hypothetical protein	TGME49_22 9750	146074	33	83
110	hypothetical protein	TGME49_26 9413	11746	41	6

Table A.2. (continued)

111	hypothetical protein	TGME49_26 8950	60196	35	25
112	hypothetical protein	TGME49_23 0180	57330	34	26
113	hypothetical protein	TGME49_22 6690	113185	37	65
114	hypothetical protein	TGME49_22 6320	73663	36	22
115	hypothetical protein	TGME49_22 5380	51498	36	56
116	hypothetical protein	TGME49_22 5130	148061	40	74
117	hypothetical protein	TGME49_22 8630	34812	115	23
118	hypothetical protein	TGME49_22 4320	262381	39	156
119	hypothetical protein	TGME49_22 3760	168702	32	87
120	hypothetical protein	TGME49_22 3725	449056	32	188
121	hypothetical protein	TGME49_22 3500	120960	34	66
122	hypothetical protein	TGME49_23 4250	248625	35	182
123	hypothetical protein	TGME49_23 4590	30345	38	28
124	hypothetical protein	TGME49_23 5580	79102	33	62

Table A.2. (continued)

125	hypothetical protein	TGME49_23 6800	52349	36	22
126	hypothetical protein	TGME49_23 7195	99915	44	45
127	hypothetical protein	TGME49_22 8065	10166	34	6
128	hypothetical protein	TGME49_21 4950	21501	36	2
129	hypothetical protein	TGME49_21 5030	199439	33	62
130	hypothetical protein	TGME49_31 0790	24708	51	28
131	hypothetical protein	TGME49_31 1270	203341	34	108
132	hypothetical protein	TGME49_31 2500	169890	39	113
133	hypothetical protein	TGME49_31 2580	96685	34	48
134	hypothetical protein	TGME49_31 2905	261708	41	202
135	hypothetical protein	TGME49_31 3340	161141	36	187
136	hypothetical protein	TGME49_31 3430	593550	43	219
137	hypothetical protein	TGME49_24 6190	530021	37	228
138	hypothetical protein	TGME49_24 6580	42820	35	27

Table A.2. (continued)

139	hypothetical protein	TGME49_24 7520	26535	34	9
140	hypothetical protein	TGME49_30 7860	434056	42	235
141	hypothetical protein	TGME49_25 0090	88129	35	52
142	hypothetical protein	TGME49_21 9640	197263	34	153
143	Kazal-type serine protease inhibitor domain-containing protein	TGME49_22 4080	214943	33	83
144	kelch repeat-containing protein	TGME49_26 4940	214243	33	81
145	kelch repeat-containing protein	TGME49_22 9290	104328	37	39
146	kinesin, putative	TGME49_28 6660	76150	36	30
147	lactate dehydrogenase LDH1 (LDH1)	TGME49_23 2350	35548	32	56
148	lactate dehydrogenase LDH2 (LDH2)	TGME49_29 1040	35334	43	44
149	leucine rich repeat-containing protein	TGME49_26 2530	108056	34	29
150	leucyl aminopeptidase LAP (LAP)	TGME49_29 0670	83236	34	68
151	lipase	TGME49_26 9300	126439	35	82
152	long-chain fatty acid CoA ligase, putative	TGME49_24 3800	83507	49	53

Table A.2. (continued)

153	mannosyl-oligosaccharide glucosidase	TGME49_24 2020	151503	33	47
154	MC family transporter, putative	TGME49_20 8790	33687	53	26
155	Met-10+ like-protein	TGME49_24 3280	98907	35	74
156	methyltransferase MTA70, putative	TGME49_21 7350	87997	39	49
157	microneme protein MIC2 (MIC2)	TGME49_20 1780	82633	34	43
158	MoaC family protein	TGME49_24 0930	41013	35	38
159	Mpv17 / PMP22 family protein	TGME49_27 3290	93582	33	54
160	Myb family DNA-binding domain- containing protein	TGME49_32 1450	437016	33	237
161	myosin A	TGME49_23 5470	93319	36	52
162	myosin D	TGME49_26 3180	91051	36	82
163	Nin one binding (NOB1) Zn-ribbon family protein	TGME49_21 8570	61202	43	37
164	non-specific serine/threonine protein kinase	TGME49_26 8370	904528	36	419
165	Not1 N-terminal domain, CCR4-Not complex component protein	TGME49_23 3020	83853	34	34
166	notchless, putative	TGME49_21 5740	57710	59	23

Table A.2. (continued)

167	oxidoreductase, short chain dehydrogenase/reductase family protein	TGME49_31 3050	40367	36	30
168	PCI domain-containing protein	TGME49_29 2220	61439	45	45
169	PCI domain-containing protein	TGME49_21 7820	51722	55	23
170	PDI family protein	TGME49_23 2410	24909	35	18
171	peptidase M16 family protein, putative	TGME49_23 6210	56915	36	38
172	peptidase M16 inactive domain-containing protein	TGME49_22 7948	142762	34	74
173	peroxiredoxin PRX3 (PRX3)	TGME49_23 0410	30534	59	11
174	phosphatidylinositol 3- and 4-kinase	TGME49_26 6010	963722	35	519
175	phosphatidylinositol 3- and 4-kinase	TGME49_21 5700	316905	36	133
176	phosphatidylserine decarboxylase	TGME49_26 9920	107987	34	79
177	phospholipase, patatin family protein	TGME49_21 2130	109733	35	60
178	phospholipase/carboxylesterase	TGME49_25 4690	54235	33	29
179	proliferation-associated protein 2G4, putative	TGME49_27 9390	50341	33	20
180	proteasome 26S regulatory subunit	TGME49_31 3410	117708	42	82

Table A.2. (continued)

181	protein kinase	TGME49_23 9420	183676	37	73
182	protein phosphatase 2C domain-containing protein	TGME49_20 1630	41457	32	31
183	protein phosphatase 2C domain-containing protein	TGME49_20 1520	42094	32	35
184	protein phosphatase 2C domain-containing protein	TGME49_23 2340	59867	83	31
185	protein phosphatase 2C domain-containing protein	TGME49_27 0320	58674	61	49
186	Rad9 protein	TGME49_25 5430	88589	35	53
187	RAP domain-containing protein	TGME49_23 9350	297817	37	112
188	RAP domain-containing protein	TGME49_23 7100	181991	38	99
189	RecF/RecN/SMC N terminal domain-containing protein	TGME49_25 7180	181309	35	99
190	rhoptry neck protein RON5 (RON5)	TGME49_31 1470	187408	33	103
191	rhoptry protein ROP16 (ROP16)	TGME49_26 2730	76215	36	37
192	ribonuclease type III Dicer	TGME49_26 7030	491971	39	171
193	ribosomal L37ae family protein	TGME49_32 5400	13049	44	11
194	ribosomal protein RPL11 (RPL11)	TGME49_30 9820	20199	52	10



Table A.2. (continued)

195	ribosomal protein RPL14 (RPL14)	TGME49_26 7060	17878	40	13
196	ribosomal protein RPL17 (RPL17)	TGME49_29 9050	22269	69	25
197	ribosomal protein RPL23A (RPL23A)	TGME49_23 8010	18561	37	24
198	ribosomal protein RPL3 (RPL3)	TGME49_22 7360	44056	55	50
199	ribosomal protein RPL37A (RPL37A)	TGME49_30 0190	10404	44	20
200	ribosomal protein RPL4 (RPL4)	TGME49_30 9120	49258	54	55
201	ribosomal protein RPL6 (RPL6)	TGME49_31 3390	31222	111	23
202	ribosomal protein RPL7 (RPL7)	TGME49_31 4810	30121	43	26
203	ribosomal protein RPP0 (RPP0)	TGME49_21 8410	34157	75	27
204	ribosomal protein RPS10 (RPS10)	TGME49_27 5810	17398	41	19
205	ribosomal protein RPS11 (RPS11)	TGME49_22 6970	18677	41	20
206	ribosomal protein RPS17 (RPS17)	TGME49_20 7840	15228	99	21
207	ribosomal protein RPS18 (RPS18)	TGME49_22 5080	17722	46	13
208	ribosomal protein RPS2 (RPS2)	TGME49_30 5520	29336	46	21

Table A.2. (continued)

209	ribosomal protein RPS20 (RPS20)	TGME49_22 3050	25899	83	27
210	ribosomal protein RPS25 (RPS25)	TGME49_23 1140	18682	46	9
211	ribosomal protein RPS26 (RPS26)	TGME49_24 3570	12725	42	7
212	ribosomal protein RPS28 (RPS28)	TGME49_20 9290	13214	61	11
213	ribosomal protein RPS4 (RPS4)	TGME49_20 7440	35507	74	55
214	ribosomal protein RPS7 (RPS7)	TGME49_23 9100	22587	47	19
215	ribosomal protein RPSA (RPSA)	TGME49_26 6060	31512	45	19
216	ribosomal-ubiquitin protein RPS27A (RPS27A)	TGME49_24 5620	17321	53	15
217	RNA polymerase-associated protein RTF1 (RTF1)	TGME49_24 4210	84698	35	58
218	RNA recognition motif-containing protein	TGME49_32 1500	50938	40	57
219	RNA recognition motif-containing protein	TGME49_21 1420	20398	64	14
220	RNA recognition motif-containing protein	TGME49_32 0100	128921	32	105
221	RNA recognition motif-containing protein	TGME49_26 5250	160178	34	149
222	RNA recognition motif-containing protein	TGME49_26 4610	44974	37	26

Table A.2. (continued)

223	RNA recognition motif-containing protein	TGME49_29 1930	73117	35	93
224	rRNA pseudouridine synthase	TGME49_21 4210	56744	43	45
225	SAG-related sequence SRS14A (SRS14A)	TGME49_25 4060	40805	34	19
226	SAG-related sequence SRS29C (SRS29C)	TGME49_23 3480	39119	83	17
227	Sec1 family protein	TGME49_24 0740	71938	33	34
228	site-specific recombinase, phage integrase family protein	TGME49_25 9230	112628	34	88
229	SNARE associated Golgi protein	TGME49_27 9370	46581	37	28
230	STE kinase	TGME49_22 5960	412543	36	171
231	subtilisin SUB2 (SUB2)	TGME49_31 4500	141581	34	67
232	surp module domain-containing protein	TGME49_24 6500	71701	34	36
233	sushi domain (scr repeat) domain-containing protein	TGME49_22 3480	507026	38	274
234	syntaxin protein	TGME49_20 9820	73624	38	39
235	TBC domain-containing protein	TGME49_20 3910	228838	34	115
236	T-complex protein 10 C-terminus protein	TGME49_25 8710	190704	35	108

Table A.2. (continued)

237	tetratricopeptide repeat-containing protein	TGME49_30 5150	153101	32	85
238	tetratricopeptide repeat-containing protein	TGME49_24 2360	80215	33	46
239	tetratricopeptide repeat-containing protein	TGME49_24 9480	126391	32	109
240	<i>Toxoplasma gondii</i> family E protein	TGME49_24 0360	51331	39	49
241	transmembrane protein	TGME49_31 3930	94912	65	51
242	tRNA ligases class II (D, K and N) domain-containing protein	TGME49_22 0350	127446	36	63
243	Tubulin-tyrosine ligase family protein	TGME49_22 8410	125827	32	70
244	UvrD/REP helicase domain-containing protein	TGME49_27 7550	342488	33	124
245	Vps52 / Sac2 family protein	TGME49_25 8832	86819	33	59
246	V-type H(+)-translocating pyrophosphatase VP1	TGME49_24 8670	85295	35	25
247	WD domain, G-beta repeat-containing protein	TGME49_31 5140	155379	37	72
248	zinc finger (CCCH type) motif-containing protein	TGME49_25 0690	368046	34	161
249	zinc finger, C3HC4 type (RING finger) domain-containing protein	TGME49_21 5640	160053	37	53

## REFERENCES

1. Thelen, J.J. and J.A. Miernyk, The proteomic future: where mass spectrometry should be taking us. *The Biochemical journal*, 2012. **444**(2): p. 169-81.
2. Angel, T.E., et al., Mass spectrometry-based proteomics: existing capabilities and future directions. *Chemical Society reviews*, 2012. **41**(10): p. 3912-28.
3. Brewis, I.A. and P. Brennan, Proteomics technologies for the global identification and quantification of proteins. *Advances in protein chemistry and structural biology*, 2010. **80**: p. 1-44.
4. Sharma, P. and C.E. Chitnis, Key molecular events during host cell invasion by Apicomplexan pathogens. *Curr Opin Microbiol*, 2013. **16**(4): p. 432-7.
5. Han, X., A. Aslanian, and J.R. Yates, 3rd, Mass spectrometry for proteomics. *Curr Opin Chem Biol*, 2008. **12**(5): p. 483-90.
6. Hecker, M., A proteomic view of cell physiology of *Bacillus subtilis*--bringing the genome sequence to life. *Advances in biochemical engineering/biotechnology*, 2003. **83**: p. 57-92.
7. Yates, J.R., C.I. Ruse, and A. Nakorchevsky, Proteomics by mass spectrometry: approaches, advances, and applications. *Annu Rev Biomed Eng*, 2009. **11**: p. 49-79.
8. Sterling, H.J., et al., Supercharging protein complexes from aqueous solution disrupts their native conformations. *Journal of the American Society for Mass Spectrometry*, 2012. **23**(2): p. 191-200.
9. Aebersold, R. and M. Mann, Mass spectrometry-based proteomics. *Nature*, 2003. **422**(6928): p. 198-207.
10. Sterling, H.J., et al., Protein conformation and supercharging with DMSO from aqueous solution. *Journal of the American Society for Mass Spectrometry*, 2011. **22**(7): p. 1178-86.
11. Wisniewski, J.R., Proteomic sample preparation from formalin fixed and paraffin embedded tissue. *Journal of visualized experiments : JoVE*, 2013(79).

12. Goldstein, I.J. and R.N. Iyer, Interaction of concanavalin A, a phytohemagglutinin, with model substrates. *Biochimica et biophysica acta*, 1966. **121**(1): p. 197-200.
13. Xiao, Y., L. Guo, and Y. Wang, A Targeted Quantitative Proteomics Strategy for Global Kinome Profiling of Cancer Cells and Tissues. *Molecular & cellular proteomics : MCP*, 2014.
14. Wang, J.M., et al., Conformation of concanavalin A and its fragments in aqueous solution and organic solvent-water mixtures. *Journal of protein chemistry*, 1992. **11**(2): p. 157-64.
15. Pallarola, D., et al., Redox-active concanavalin A: synthesis, characterization, and recognition-driven assembly of interfacial architectures for bioelectronic applications. *Langmuir : the ACS journal of surfaces and colloids*, 2010. **26**(16): p. 13684-96.
16. Navarro, P., et al., General Statistical Framework for Quantitative Proteomics by Stable Isotope Labeling. *Journal of proteome research*, 2014.
17. Annesley, T.M., Ion suppression in mass spectrometry. *Clinical Chemistry*, 2003. **49**(7): p. 1041-1044.
18. Blundell, T., New dimensions of structural proteomics: exploring chemical and biological space. *Structure*, 2007. **15**(11): p. 1342-3.
19. Delahunty, C.M. and J.R. Yates, 3rd, MudPIT: multidimensional protein identification technology. *Biotechniques*, 2007. **43**(5): p. 563, 565, 567 passim.
20. Lomeli, S.H., et al., New reagents for increasing ESI multiple charging of proteins and protein complexes. *Journal of the American Society for Mass Spectrometry*, 2010. **21**(1): p. 127-31.
21. Drissi, R., M.L. Dubois, and F.M. Boisvert, Proteomics Methods for Subcellular Proteome Analysis. *FEBS J*, 2013.
22. Chen, E.I., et al., Large scale protein profiling by combination of protein fractionation and multidimensional protein identification technology (MudPIT). *Mol Cell Proteomics*, 2006. **5**(1): p. 53-6.
23. Link, A.J., et al., Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol*, 1999. **17**(7): p. 676-82.

24. Wisniewski, J.R., K. Dus, and M. Mann, Proteomic workflow for analysis of archival formalin-fixed and paraffin-embedded clinical samples to a depth of 10 000 proteins. *Proteomics Clin Appl*, 2013. **7**(3-4): p. 225-33.
25. Zhou, H., et al., The GlycoFilter: A Simple and Comprehensive Sample Preparation Platform for Proteomics, N-Glycomics and Glycosylation Site Assignment. *Molecular & cellular proteomics : MCP*, 2013.
26. Ho, C.S., et al., Electrospray ionisation mass spectrometry: principles and clinical applications. *The Clinical biochemist. Reviews / Australian Association of Clinical Biochemists*, 2003. **24**(1): p. 3-12.
27. Leon, I.R., et al., Quantitative assessment of in-solution digestion efficiency identifies optimal protocols for unbiased protein analysis. *Molecular & cellular proteomics : MCP*, 2013.
28. Yan, A. and W.J. Lennarz, Unraveling the mechanism of protein N-glycosylation. *The Journal of biological chemistry*, 2005. **280**(5): p. 3121-4.
29. Lotan, R., et al., Activities of lectins and their immobilized derivatives in detergent solutions. Implications on the use of lectin affinity chromatography for the purification of membrane glycoproteins. *Biochemistry*, 1977. **16**(9): p. 1787-94.
30. Goldstein, I.J., C.E. Hollerman, and E.E. Smith, Protein-Carbohydrate Interaction. Ii. Inhibition Studies on the Interaction of Concanavalin a with Polysaccharides. *Biochemistry*, 1965. **4**: p. 876-83.
31. Portman, N., et al., Evidence for Intraflagellar Transport and Apical Complex formation in a Free Living Relative of the Apicomplexa. *Eukaryotic cell*, 2013.
32. King, R., et al., Mechanistic investigation of ionization suppression in electrospray ionization. *Journal of the American Society for Mass Spectrometry*, 2000. **11**(11): p. 942-950.
33. Thornalley, P.J., et al., Quantitative screening of advanced glycation endproducts in cellular and extracellular proteins by tandem mass spectrometry. *The Biochemical journal*, 2003. **375**(Pt 3): p. 581-92.
34. Graves, P.R. and T.A. Haystead, Molecular biologist's guide to proteomics. *Microbiol Mol Biol Rev*, 2002. **66**(1): p. 39-63; table of contents.

35. Gupta, R. and S. Brunak, Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac Symp Biocomput*, 2002: p. 310-22.
36. Kononkov, N.V., M. Sudakov, and D.J. Douglas, Matrix methods for the calculation of stability diagrams in quadrupole mass spectrometry. *Journal of the American Society for Mass Spectrometry*, 2002. **13**(6): p. 597-613.
37. Eng, J.K., A.L. McCormack, and J.R. Yates, An Approach to Correlate Tandem Mass-Spectral Data of Peptides with Amino-Acid-Sequences in a Protein Database. *Journal of the American Society for Mass Spectrometry*, 1994. **5**(11): p. 976-989.
38. Nesvizhskii, A.I., O. Vitek, and R. Aebersold, Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nature methods*, 2007. **4**(10): p. 787-797.
39. Cottrell, J.S., Protein identification using MS/MS data. *Journal of proteomics*, 2011. **74**(10): p. 1842-51.
40. Matrix Science. Available from: <http://www.matrixscience.com/>.
41. Pappin, D.J., P. Hojrup, and A.J. Bleasby, Rapid identification of proteins by peptide-mass fingerprinting. *Curr Biol*, 1993. **3**(6): p. 327-32.
42. Pappin, D.J.C., et al., Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, 1999. **20**(18): p. 3551-3567.
43. Shental-Bechor, D. and Y. Levy, Folding of glycoproteins: toward understanding the biophysics of the glycosylation code. *Curr Opin Struct Biol*, 2009. **19**(5): p. 524-33.
44. Wang, Y.C., S.E. Peterson, and J.F. Loring, Protein post-translational modifications and regulation of pluripotency in human stem cells. *Cell Res*, 2014. **24**(2): p. 143-60.
45. Lu, D., C. Yang, and Z. Liu, How hydrophobicity and the glycosylation site of glycans affect protein folding and stability: a molecular dynamics simulation. *J Phys Chem B*, 2012. **116**(1): p. 390-400.
46. Moremen, K.W., M. Tiemeyer, and A.V. Nairn, Vertebrate protein glycosylation: diversity, synthesis and function. *Nat Rev Mol Cell Biol*, 2012. **13**(7): p. 448-62.



47. Spiro, R.G., Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology*, 2002. **12**(4): p. 43R-56R.
48. Warren, G., Size and position matter. *Nat Rev Mol Cell Biol*, 2013. **14**(12): p. 755-7.
49. Brandizzi, F. and C. Barlowe, Organization of the ER-Golgi interface for membrane traffic control. *Nat Rev Mol Cell Biol*, 2013. **14**(6): p. 382-92.
50. Shental-Bechor, D. and Y. Levy, Effect of glycosylation on protein folding: a close look at thermodynamic stabilization. *Proc Natl Acad Sci U S A*, 2008. **105**(24): p. 8256-61.
51. Berg JM, T.J., Stryer L., *Biochemistry* 5th edition. Lectins Are Specific Carbohydrate-Binding Proteins. 2002, New York: W H Freeman.
52. Dalziel, M., et al., Emerging principles for the therapeutic exploitation of glycosylation. *Science*, 2014. **343**(6166): p. 1235681.
53. McNaught, A.D., A. Wilkinson, and International Union of Pure and Applied Chemistry., *Compendium of chemical terminology : IUPAC recommendations*. 2nd ed. 1997, Oxford England ; Malden, MA, USA: Blackwell Science. vii, 450 p.
54. Bertozzi, C.R. and D. Rabuka, Structural Basis of Glycan Diversity, in *Essentials of Glycobiology*, A. Varki, et al., Editors. 2009: Cold Spring Harbor (NY).
55. Stanley, P., H. Schachter, and N. Taniguchi, N-Glycans, in *Essentials of Glycobiology*, A. Varki, et al., Editors. 2009: Cold Spring Harbor (NY).
56. Brockhausen, I., H. Schachter, and P. Stanley, O-GalNAc Glycans, in *Essentials of Glycobiology*, A. Varki, et al., Editors. 2009: Cold Spring Harbor (NY).
57. Brewer, C.F., M.C. Miceli, and L.G. Baum, Clusters, bundles, arrays and lattices: novel mechanisms for lectin-saccharide-mediated cellular interactions. *Curr Opin Struct Biol*, 2002. **12**(5): p. 616-23.
58. Abbott, K.L. and J.M. Pierce, Lectin-based glycoproteomic techniques for the enrichment and identification of potential biomarkers. *Methods in enzymology*, 2010. **480**: p. 461-76.

59. Agrawal, B.B. and I.J. Goldstein, Protein-carbohydrate interaction. VII. Physical and chemical studies on concanavalin A, the hemagglutinin of the jack bean. Archives of biochemistry and biophysics, 1968. **124**(1): p. 218-29.
60. Bryce, R.A., I.H. Hillier, and J.H. Naismith, Carbohydrate-protein recognition: Molecular dynamics simulations and free energy analysis of oligosaccharide binding to Concanavalin A. Biophysical Journal, 2001. **81**(3): p. 1373-1388.
61. Battistel, E., et al., Physico-chemical and thermodynamic properties of monomeric concanavalin A. European biophysics journal : EBJ, 1985. **13**(1): p. 1-9.
62. Derewenda, Z., et al., The structure of the saccharide-binding site of concanavalin A. EMBO J, 1989. **8**(8): p. 2189-93.
63. Kaushik, S., D. Mohanty, and A. Surolia, The role of metal ions in substrate recognition and stability of concanavalin A: a molecular dynamics study. Biophys J, 2009. **96**(1): p. 21-34.
64. Cummings, R.D. and M.E. Etzler, Antibodies and Lectins in Glycan Analysis, in Essentials of Glycobiology, A. Varki, et al., Editors. 2009: Cold Spring Harbor (NY).
65. Labrou, N.E., Design and selection of ligands for affinity chromatography. Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences, 2003. **790**(1-2): p. 67-78.
66. Luo, Q., et al., Analysis of the glycoproteome of *Toxoplasma gondii* using lectin affinity chromatography and tandem mass spectrometry. Microbes Infect, 2011. **13**(14-15): p. 1199-210.
67. Ngoka, L.C., Sample prep for proteomics of breast cancer: proteomics and genontology reveal dramatic differences in protein solubilization preferences of radioimmunoprecipitation assay and urea lysis buffers. Proteome science, 2008. **6**: p. 30.
68. Peach, M., N. Marsh, and D.J. Macphee, Protein solubilization: attend to the choice of lysis buffer. Methods in molecular biology, 2012. **869**: p. 37-47.
69. Abcam: Western blotting – a beginner's guide. 9-24-12]; Available from: <http://www.abcam.com/ps/pdf/protocols/WB-beginner.pdf>.

70. Winter, D. and H. Steen, Optimization of cell lysis and protein digestion protocols for the analysis of HeLa S3 cells by LC-MS/MS. *Proteomics*, 2011. **11**(24): p. 4726-30.
71. Hames, B.D. and D. Rickwood, *Gel electrophoresis of proteins : a practical approach*. 2nd ed. The Practical approach series. 1990, Oxford ; New York: IRL Press at Oxford University Press. xviii, 383 p.
72. Weber, K. and M. Osborn, The reliability of molecular weight determinations by dodecyl sulfate-polyacrylamide gel electrophoresis. *J Biol Chem*, 1969. **244**(16): p. 4406-12.
73. Liu, T., et al., Three-layer sandwich gel electrophoresis: a method of salt removal and protein concentration in proteome analysis. *Journal of proteome research*, 2008. **7**(10): p. 4256-65.
74. Luo, J.Y., et al., Diffusion dialysis-concept, principle and applications. *Journal of Membrane Science*, 2011. **366**(1-2): p. 1-16.
75. Evans, D.R.H., J.K. Romero, and M. Westoby, Chapter 9 Concentration of Proteins and Removal of Solutes, in *Methods in Enzymology*, R.B. Richard and P.D. Murray, Editors. 2009, Academic Press. p. 97-120.
76. Laemmli, U.K., Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 1970. **227**(5259): p. 680-5.
77. Luk, F.C., T.M. Johnson, and C.J. Beckers, N-linked glycosylation of proteins in the protozoan parasite *Toxoplasma gondii*. *Mol Biochem Parasitol*, 2008. **157**(2): p. 169-78.
78. Chen, R., et al., Glycoproteomics analysis of human liver tissue by combination of multiple enzyme digestion and hydrazide chemistry. *J Proteome Res*, 2009. **8**(2): p. 651-61.
79. Christianson, J.C., et al., OS-9 and GRP94 deliver mutant alpha1-antitrypsin to the Hrd1-SEL1L ubiquitin ligase complex for ERAD. *Nat Cell Biol*, 2008. **10**(3): p. 272-82.
80. Zhang, H., et al., Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotechnol*, 2003. **21**(6): p. 660-6.

81. Prandovszky, E., et al., The neurotropic parasite *Toxoplasma gondii* increases dopamine metabolism. Plos One, 2011. **6**(9): p. e23866.
82. Blader, I.J. and J.P. Saeij, Communication between *Toxoplasma gondii* and its host: impact on parasite growth, development, immune evasion, and virulence. APMIS : acta pathologica, microbiologica, et immunologica Scandinavica, 2009. **117**(5-6): p. 458-76.
83. Dubey, J.P., D.S. Lindsay, and C.A. Speer, Structures of *Toxoplasma gondii* tachyzoites, bradyzoites, and sporozoites and biology and development of tissue cysts. Clinical microbiology reviews, 1998. **11**(2): p. 267-99.
84. Zhang, Y.W., et al., Initial characterization of CST1, a *Toxoplasma gondii* cyst wall glycoprotein. Infect Immun, 2001. **69**(1): p. 501-7.
85. Fauquenoy, S., et al., Proteomics and glycomics analyses of N-glycosylated structures involved in *Toxoplasma gondii*--host cell interactions. Molecular & cellular proteomics : MCP, 2008. **7**(5): p. 891-910.
86. Garenaux, E., et al., The dual origin of *Toxoplasma gondii* N-glycans. Biochemistry, 2008. **47**(47): p. 12270-6.
87. Takemae, H., et al., Characterization of the interaction between *Toxoplasma gondii* rhoptry neck protein 4 and host cellular beta-tubulin. Sci Rep, 2013. **3**: p. 3199.
88. Che, F.Y., et al., Comprehensive proteomic analysis of membrane proteins in *Toxoplasma gondii*. Molecular & cellular proteomics : MCP, 2011. **10**(1): p. M110 000745.
89. Quan, J.H., et al., Induction of protective immune responses by a multiantigenic DNA vaccine encoding GRA7 and ROP1 of *Toxoplasma gondii*. Clin Vaccine Immunol, 2012. **19**(5): p. 666-74.
90. Zhang, J., et al., Evaluation of the immune response induced by multiantigenic DNA vaccine encoding SAG1 and ROP2 of *Toxoplasma gondii* and the adjuvant properties of murine interleukin-12 plasmid in BALB/c mice. Parasitol Res, 2007. **101**(2): p. 331-8.

91. Beckers, C.J., et al., The *Toxoplasma gondii* rhostry protein ROP 2 is inserted into the parasitophorous vacuole membrane, surrounding the intracellular parasite, and is exposed to the host cell cytoplasm. *J Cell Biol*, 1994. **127**(4): p. 947-61.
92. Ahn, H.J., et al., Interactions between secreted GRA proteins and host cell proteins across the parasitophorous vacuolar membrane in the parasitism of *Toxoplasma gondii*. *Korean J Parasitol*, 2006. **44**(4): p. 303-12.
93. Gilbert, L.A., et al., *Toxoplasma gondii* targets a protein phosphatase 2C to the nuclei of infected host cells. *Eukaryot Cell*, 2007. **6**(1): p. 73-83.
94. Jan, G., et al., The toxofilin-actin-PP2C complex of *Toxoplasma*: identification of interacting domains. *Biochem J*, 2007. **401**(3): p. 711-9.
95. Wang, H.L., et al., *Toxoplasma gondii* protein disulfide isomerase (TgPDI) is a novel vaccine candidate against toxoplasmosis. *PLoS One*, 2013. **8**(8): p. e70884.
96. Matsubayashi, M., et al., Elongation factor-1alpha is a novel protein associated with host cell invasion and a potential protective antigen of *Cryptosporidium parvum*. *J Biol Chem*, 2013. **288**(47): p. 34111-20.
97. Halbach, T., N. Scheer, and W. Werr, Transcriptional activation by the PHD finger is inhibited through an adjacent leucine zipper that binds 14-3-3 proteins. *Nucleic Acids Research*, 2000. **28**(18): p. 3542-3550.
98. Nardelli, S.C., et al., The histone code of *Toxoplasma gondii* comprises conserved and unique posttranslational modifications. *MBio*, 2013. **4**(6): p. e00922-13.
99. Hwang, I.Y., et al., *Toxoplasma gondii* infection inhibits the mitochondrial apoptosis through induction of Bcl-2 and HSP70. *Parasitol Res*, 2010. **107**(6): p. 1313-21.
100. Echeverria, P.C., et al., The Hsp90 co-chaperone p23 of *Toxoplasma gondii*: Identification, functional analysis and dynamic interactome determination. *Mol Biochem Parasitol*, 2010. **172**(2): p. 129-40.
101. Hager, K.M., et al., The nuclear envelope serves as an intermediary between the ER and Golgi complex in the intracellular parasite *Toxoplasma gondii*. *J Cell Sci*, 1999. **112** ( Pt 16): p. 2631-8.
102. Toth, J.I., et al., Selective coactivator interactions in gene activation by SREBP-1a and -1c. *Mol Cell Biol*, 2004. **24**(18): p. 8288-300.

103. Ochi, A., et al., The catalytic domain of topological knot tRNA methyltransferase (TrmH) discriminates between substrate tRNA and nonsubstrate tRNA via an induced-fit process. *J Biol Chem*, 2013. **288**(35): p. 25562-74.
104. Duncan, O., M.W. Murcha, and J. Whelan, Unique components of the plant mitochondrial protein import apparatus. *Biochim Biophys Acta*, 2013. **1833**(2): p. 304-13.
105. Varjosalo, M., et al., The protein interaction landscape of the human CMGC kinase group. *Cell Rep*, 2013. **3**(4): p. 1306-20.
106. Polgar, L., The catalytic triad of serine peptidases. *Cellular and molecular life sciences : CMLS*, 2005. **62**(19-20): p. 2161-72.
107. Yu, H., et al., Comparative studies of Munc18c and Munc18-1 reveal conserved and divergent mechanisms of Sec1/Munc18 proteins. *Proc Natl Acad Sci U S A*, 2013. **110**(35): p. E3271-80.
108. Kofler, M.M. and C. Freund, The GYF domain. *FEBS J*, 2006. **273**(2): p. 245-56.
109. Pomel, S., F.C. Luk, and C.J. Beckers, Host cell egress and invasion induce marked relocations of glycolytic enzymes in *Toxoplasma gondii* tachyzoites. *PLoS Pathog*, 2008. **4**(10): p. e1000188.
110. Westgate, G.E., N.V. Botchkareva, and D.J. Tobin, The biology of hair diversity. *Int J Cosmet Sci*, 2013. **35**(4): p. 329-36.
111. Weiss, L.M., et al., *Toxoplasma gondii* proteomics. *Expert Rev Proteomics*, 2009. **6**(3): p. 303-13.
112. Xia, D., et al., The proteome of *Toxoplasma gondii*: integration with the genome provides novel insights into gene expression and annotation. *Genome Biol*, 2008. **9**(7): p. R116.

## VITA

### Educational institutions attended and degrees already awarded:

1. M.S. Physical chemistry, Faculty of Physical Chemistry, Dept. of Biophysical chemistry, University of Belgrade, 2006.
2. Specialist of toxicological chemistry, National Center for Poisoning Control, Military Medical Academy, Belgrade 2005.
3. B.S. Physical chemistry, Faculty of Physical Chemistry, University of Belgrade, 2000.

### Professional positions:

1. Specialist in Toxicological laboratory, Military Medical Academy, National Center for Poisoning Control, Belgrade, Serbia, 2002-2005.
2. Trainee at the Institute of Radiology, Dept. of Radiotherapy, Military Medical Academy, Belgrade, Serbia, 2001-2002.
3. Trainee at the Institute of Nuclear Medicine, Military Medical Academy, Belgrade, Serbia, 2001.

### Scholastic and professional honors:

1. Research Challenge Trust Fund (RCTF) Fellowship, University of Kentucky, 2012

2. Research Challenge Trust Fund (RCTF) Fellowship, University of Kentucky, 2011

Publications:

1. S. Trajkovic, X. Zhang, S. Daunert, Y. Cai, Atomic force microscopy study of the conformational change in immobilized calmodulin, Langmuir 2011 Sep 6;27(17):10793-9
2. A. Djordjevic, B. Ajdinovic, M. Dopudja, S. Trajkovic, Z. Milovanovic, T. Maksin, O. Neskovic, V. Bogdanovic, Dj. Tripkovic, J. Cveticanin, Scintigraphy of the Domestic Dog Using [ $^{99m}\text{Tc}(\text{CO})_3(\text{H}_2\text{O})_3$ ]- $\text{C}_{60}(\text{OH})_{22-24}$ , Digest Journal of Nanomaterials and Biostructures 2010, 6(1), 99-106
3. J. Orbulescu, M. Micic, M. Ensor, S. Trajkovic, S. Daunert, R. M. Leblanc, Human Cardiac Troponin I: A Langmuir Monolayer Study, Langmuir 2010, 26(5), 3268–3274
4. S. Trajkovic, S. Dobric, V. Jacevic, V. Dragojevic-Simic, Z. Milovanovic, A. Djordjevic, Tissue-protective effects of fulleranol  $\text{C}_{60}(\text{OH})_{24}$  and amifostine in irradiated rats, Colloids and Surfaces B: Biointerfaces 58 (2007) 39–43
5. S. Trajkovic, S. Dobric, A. Djordjevic, G. Nisevic, Z. Milovanovic, V. Dragojevic-Simic, Comparative efficiency of fulleranol and amifostine in protection of hematopoiesis against harmful effects of ionizing radiation in rats, Materials Science Forum Vol. 518 (2006)
6. S. Trajkovic, S. Dobric, A. Djordjevic, V. Dragojevic-Simic, Z. Milovanovic, Radioprotective efficiency of fulleranol in irradiated mice, Materials Science Forum Vol. 494 (2005) pp. 549-554.
7. S. Trajkovic, S. Dobric, A. Djordjevic, V. Dragojevic-Simic, Z. Milovanovic, Radioprotective efficiency of fulleranol  $\text{C}_{60}(\text{OH})_{24}$  during in vivo conditions, Novi Materijali (yu issn 0040-2176) 14 (2005) 2.



8. B Ajdinovic, T Dragovic, S Trajkovic, GFR Normalization by Valsartan in Patients with Incipient Diabetic Nephropathy, European Journal of Nuclear Medicine and Molecular Imaging, Vol 30, supplement 2, august 2003.
9. B Ajdinovic, T Dragovic, S Trajkovic, <sup>51</sup>Cr-EDTA effect of angiotensin II receptor blockers on glomerular filtration rate in patients with incipient diabetic nephropathy, RAS vol. 11, No 2, 89-94 (2002).

Sanja Trajkovic, Ph.D.