Improved sample preparation method for LC-MS/MS based cellular proteomics and co-immunoprecipitation cross-linking proteomics study of Toll-like receptor 4.

by

ADA Shahinuzzaman

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON
August 2019

Acknowledgments

I want to grab this moment to leave a permanent thanksgiving note, commemorating everyone who has made me succeed up to this point defending my thesis at UT Arlington. First, I am thankful to almighty Allah, for my life and every moment and every achievement in this life. Then I want to thank Dr. Saiful Chowdhury, my advisor, for his patience and support towards me in this program. I probably would not have made it till this far, without his conscious monitoring over me. I am sincerely grateful that he has entrusted his support on me despite the hardships in the doctoral program at UT Arlington. I want to thank all the professors in my doctoral degree committee for their moral support and concern towards me and for the valuable advice they have given me on different occasions during this journey. I am thanking Dr. Purnendu (Sandy) Dasgupta, Dr. Kevin A. Schug, Dr. Subhrangsu S. Mandal, Dr. Robin Macaluso, Dr. Peter Kroll, Dr. Frank foss, for their valuable advising on different occasions. I want to especially thank Dr. Dasgupta and Dr. Carl J. Lovely for their classes. The analytical chemistry class with Dr. Dasgupta helped me learn a different dimension of teaching and exam system. I loved every moment doing all his take-home exams. Dr lovely allowed me to seat and audit his organic chemistry classes. I am sincerely thankful to him for this gesture. I am grateful to Dr. Bill Cleaver, Dr. Cynthia Griffith, and Dr. Heidi Conrad, for their support in my chemistry prep room works.

I want to thank our colleague Dr. Abu Hena Kamal for his encouragements and help understanding science in a better way. I want to leave a thankful note for my friend and lab mate Jayanta Chakrabarty, who has been a big help with managing grad life. I want to thank Dr. Maheshika Wanigasekara for her positive motivation and smile every moment

we were around. I want to thank Dr. Ruchika Bhawal for supporting me in her busy schedule in the final semester, while we just came in UTA. I thank Zixiang Fang for his help in lab life and office life. I want to thank Mohammad Kabir Hossain for teaching me cooking and thus survive in the USA unless otherwise, food is costly around here. I remember him nearly getting killed in the process of teaching me driving. I want to thank Mohammad Showkat Hossain (Sohail) for his support with the initial help in relocation and settling down in Texas. My friend Raisa Abrar from UTAH state helped in the beginning. I remember her thankfully. Dr. Noor e Alam Siddique and Dr. Sajeeb Raihan also helped me initially settle down and finding an apartment. I remember them thankfully. I also want to thank Deborah, Jill, Beth, and Stephanie, for their constant support. I am thankful to the fantastic people of the United States for having me here and helping explore their culture and life. It was a pleasure living here. I am thankful for the emigrant south Asian and Arab community for the opportunity to explore the excellent food they brought to DFW with them from their country. I am grateful to my mother and sister for their tolerance back home while I am doing my Ph.D. in the US. I am grateful to Dr. Md Hossain Sohrab for his constant encouragements to go back and work for my country once I would be done here. Finally, I want to leave a note of gratitude for my late father. He dreamt of seeing me a graduate someday, but that day never ended up coming in his lifetime.

Abstract

Mass spectrometry-based methods are now widely practiced in large-scale proteomics studies. Unfortunately, sample loss in proteomics hinders complex analysis in proteomics, and immune precipitation (IP) mass spectrometry-based proteomics is getting popular in studying comprehensive signaling pathways. To contribute to these demanding research areas, I have completed two experimental projects in this thesis.

In our first project, we developed a sample preparation method to identify proteins from detergent-containing cell lysate samples efficiently. Methanol-chloroform purification is an essential step in LC-MS/MS based cellular proteomics experiments since it removes salts and detergents from the sample. The cell lysis buffer contains a high concentration of detergents. Hence, proteins with large hydrophobic domains can remain dissolved in the lysis buffer after cell lysis. However, after methanol-chloroform purification, solubilization of the precipitated proteins back to the solution in the absence of detergents is difficult. Through a BCA protein quantification assay, we have seen that around two-thirds of the weight of protein become insoluble after methanol-chloroform purification. Thus, we decided to use a phase transfer surfactant, deoxycholic acid (DCA) to improve re-solubilization and protein identification after tryptic digestion. We took mouse macrophage cell lysates and performed methanol-chloroform purification followed by reconstitution in the presence of deoxycholic acid (DCA). Upon mass spectrometric analysis, for this modified method, we observed a higher number of identified proteins compared to the unmodified method. This trend was observed further when equal amounts of precipitated and resolubilized proteins from modified and unmodified methods were tryptic digested and analyzed in high-resolution mass spectrometry with an increase of 14% in protein

identification with this new method. Additionally, we compared this modified sample preparation method with other popular methods in bottom-up proteomics analysis, and we conclusively confirmed that this proposed method identifies the highest number of proteins for proteomics sample preparation. Our method solves the bottleneck of re-solubilizing methanol-chloroform precipitated proteomics for any large-scale proteomics analysis where methanol-chloroform purification and in-solution digestion are required.

Additionally, in my second project, I worked to understand the immune signaling through Toll-like receptor 4 (TLR-4). Toll-like receptors are essential immune sensors of the innate immune mechanism. They are transmembrane receptors on immune cells and organelle surfaces. Toll-like receptor 4 (TLR4) is a highly expressed receptor on immune cells and can get activated through both exogenous bacterial infections and endogenous damage signaling molecules. Here, we used a co-immunoprecipitation (IP) based cross-linking proteomics and label-free quantitation approach to reveal the TLR-4 interactome in the presence of Statin and LPS, alone or in combination on HA-TLR4 transfected HEK293 cells. A total of 712 differentially expressed proteins were identified and quantified in this study. I selected two candidate proteins from cross-linking proteomics data. Using biochemical validation through immunoblotting, we confirmed macrophage myristoylated alanine-rich C kinase substrate (MacMARCKS) as a member in TLR4 specific immune signaling network.

I believe these two studies contribute significantly in sample preparations and signaling-pathway analysis by mass spectrometry.

Table of Contents

List of Illustrations

List of Tables

List of Abbreviations

Abbreviations	Name
CCL2	C-C Motif Chemokine Ligand 2
CD14	Cluster of differentiation 14
CID	Collision-induced dissociation
CXCL8	C-X-C Motif Chemokine Ligand 8
ESI	Electrospray ionization
ETD	Electron-transfer dissociation
FASP	Filter Assisted Sample Preparation
IFN-γ	Interferon-gamma
IL-12	Interleukin 12
IL-6	Interleukin 6
iNOS	Inducible nitric oxide synthase
LBP	LPS binding protein
LIT	Linear ion trap
LPS	Lipopolysaccharides
MALDI	matrix-assisted laser desorption/ionization
MARCKSL1 / MacMARCKS	MARCKS-related protein / Macrophage myristoylated alanine-rich C-kinase substrate
MD-2	Myeloid differentiation 2
MHC	Major histocompatibility complex Protein
MyD88	Myeloid differentiation primary response 88
NK Cells	Natural Killer cells
QIT	Quadrupole ion trap
SDS	Sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate–polyacrylamide gel electrophoresis
TIRAP	Toll-interleukin 1 receptor (TIR) domain-containing adaptor protein
TNF-a	Tumor necrosis factor-alpha
TRAM	Translocating chain-associated membrane protein
TRIF	TIR-domain-containing adapter-inducing interferon-β

Chapter 1

Introduction

1.1. Mass Spectrometry

Mass spectrometry is an analytical technique that uses gas-phase ions and sorts them according to their mass to charge (m/z) ratio, using the electric field and magnetic field and provides signal (peak) output from which mass to charge ratio and abundance of ionic species are determined. [1]

1.2. Organization of mass spectrometer

A mass spectrometer has six major components. 1) System for sample introduction, 2) ion source where analytes vaporized to gas phase to produce ions, 3) mass analyzer where ions are separated according to their mass to charge ration, 4) ion detector, where we detect signal intensities of the separated m/z values, 5) vacuum system prevents the loss of ions through collision with neutral gas, 6) computers to control operation of instrument, record and process data. [2]

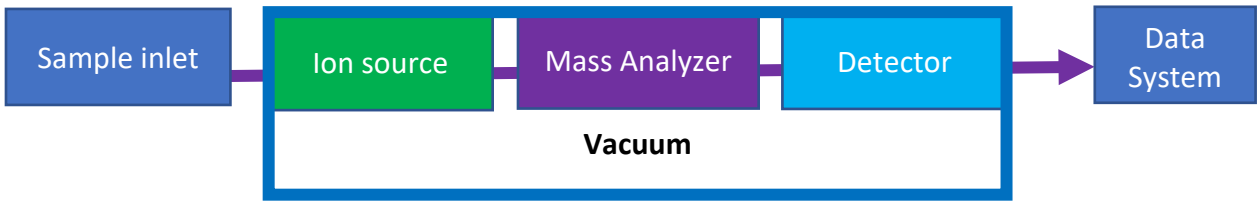


Figure 1.1. Components of a mass spectrometer.

Briefly, samples are introduced through different methods of choices based on the instrument platform, sample complexity, volatility, polarity, etc. Volatile compounds are preferred to add through Gas chromatography, non-polar compounds through robust probes, polar non-volatile biomolecules are preferred delivered through Liquid

Chromatography (LC). During sample introduction, samples are ionized in the solid plate (e.g., MALDI) or in an ion source inlet (ESI). Ions are separated according to their m/z ration in the mass analyzer and transferred to the detector for the detection or in collision cell for further fragmentation into fragments and then transferred to a same or separate analyzer for sorting according to their m/z and detected into the detector. [2]

1.3. Sample ionization

There are different ionizing methods available for different types of samples. For proteomics sample due to their non-volatile nature and thermostability, MALDI and ESI are used. In MALDI samples are co-crystallized with a matrix molecule, e.g., Sinapinic acid or dihydroxybenzoic Acid and irradiated with a pulse of laser to evaporate the matrix along with some sample molecule. The reagent ions formed from the matrix protonates the sample.[3] In ESI, analyte peptides are separated through a reverse phase column and separated based on their retention time. A high potential difference of 2.5-3KV is maintained between a hollow needle to an inlet orifice. Samples are passed through the hollow needle along with heat and nebulizing gas N_2 to nebulize and desolvate the charged droplets, causing them to shrink and subdivide and finally emit charged analyte molecules to enter the mass spectrometer. [4] All samples prepared for the research project presented in this thesis were collected using electrospray ionizations liquid chromatography-tandem mass spectrometry.

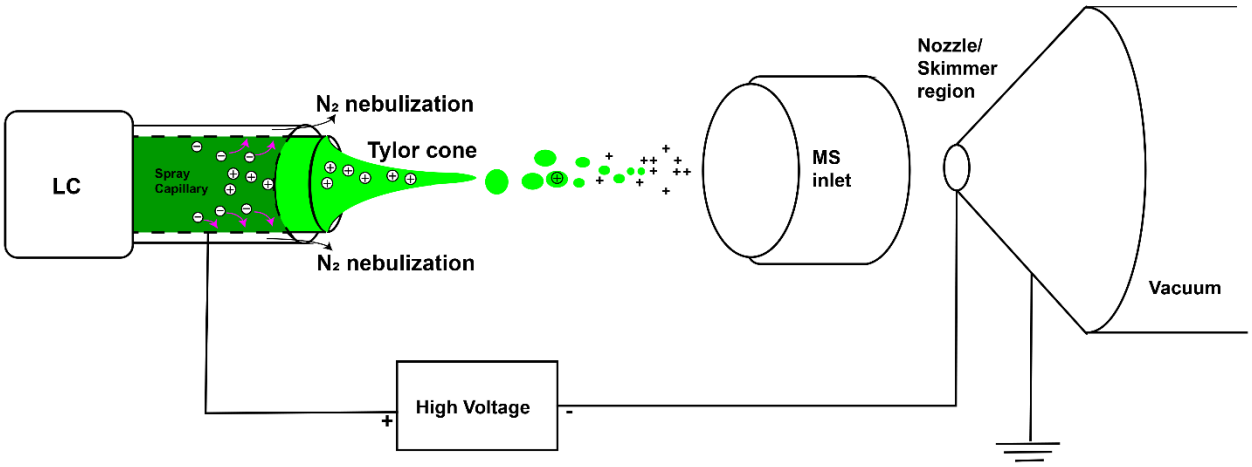


Figure 1.2. Electrospray ionization interface.

1.4. Tandem Mass Spectrometry

Tandem mass spectrometry (MS/MS or MS^n) enables two mass separations consecutively within the same instrument. Here, at least two stages of mass analysis occur with at least one step in conjugation with a fragmentation process. We have two types of instruments, e.g., tandem in space and tandem in time for tandem mass spectrometry. For tandem in space instruments, from the full scan ion spectrum, a choice of specific ions is selected in the first analyzer followed by selected ions undergoing collision-induced dissociation (CID) in the pressurized collision cell, followed by an analysis of productions in the second analyzer. Examples of such instruments will be Triple Quadrupole (QqQ) system. In tandem in time instruments, the same analyzer is used for both scans but at different interval of time. This selection of ions and fragmentation cycle can go onto for n times hence we get (MS^n) Spectra. Example of tandem in time will be an ion-trap instrument. These (MS/MS or MS^n) spectra gives essential information about the identification of small molecules (drug leads) or large biomolecules (protein or peptides). [2, 5]

1.5. Mass spectrometry (MS) based bottom-up proteomics

Proteomics is a large-scale systematic study of proteins where protein structure, function, expression, modification, interaction are studied. [6] In mass spectrometry-based proteomics, we use mass spectrometric analysis often combined with liquid chromatography to explore such features of proteins. The most commonly used method of proteomics sample analysis is bottom-up proteomics or shotgun proteomics. In all projects in this thesis, we used bottom-up proteomics approach. In bottom-up proteomics, we take the proteins (pure or cell lysate), tryptic digest them into small peptides using proteolytic enzymes, e.g. Trypsin. Different enzymes have different preferential sites for proteolytic cleavage. For trypsin, cleavage happens in carboxyl group of lysine (Lys) and arginine (Arg), resulting in a positively charged amino acid at the c-terminus. Peptides are desalted in a solid phase C18 bead tip and loaded on UHPLC system coupled to a mass spectrometer. Peptide separation takes place in a hydrophobic C18 stationary phase column using the reverse-phase gradient of 5%-70% acetonitrile before release and electrospray ionization and tandem mass spectrometry fragmentation in a Mass spectrometer coupled to the LC system. [7]

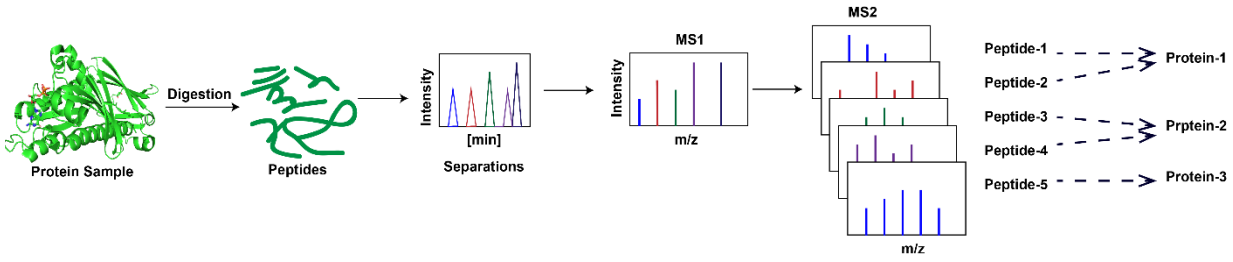


Figure 1.3. Bottom-up proteomics.

1.6. Protein Purification for mass spectrometric analysis

Protein purification is an essential step in a mass spectrometry-based bottom-up cellular proteomics experiment. The proteins in the cell are localized largely in subcellular locations, e.g., cytosol or organelles. Some are transmembrane proteins embedded in membranes. To analyze the proteins in the sample, the membrane needs to be disrupted through lysis buffer that contains a high concentration of detergents. Hence this makes the cell lysate unsuitable for tryptic digestion and subsequent mass spectrometric analysis unless the purified proteins from the lysate are collected and then tryptic digested for mass spectrometric analysis. There are different methods available that take advantage of liquid-liquid partitioning or Microcon-filters' absorption properties to separate the proteins from the detergents in cell lysate samples and prepares them for analysis. Popular methods of choices for this type of sample purification step are methanol-chloroform purification, acetone purification, detergent removal spin column filtration, filter assisted sample preparation, etc. [8-10]

1.7. Protein identification by peptide mass spectrometry

There are two ways proteins are identified from peptide mass spectrometry. These are by determination of molecular masses (peptide mass fingerprinting) or by generating sequence-specific peptide fragments. The unknown sample of proteins once tryptic digested into small peptides their absolute masses can be measured with a mass spectrometer. Their masses are searched against the theoretical masses of the proteins stored in the database. This method works well for isolated pure proteins and protein

sequences that are already stored in the database. For relatively complex sample mixture, peptides need to be fragmented in tandem mass spectrometry.[11]

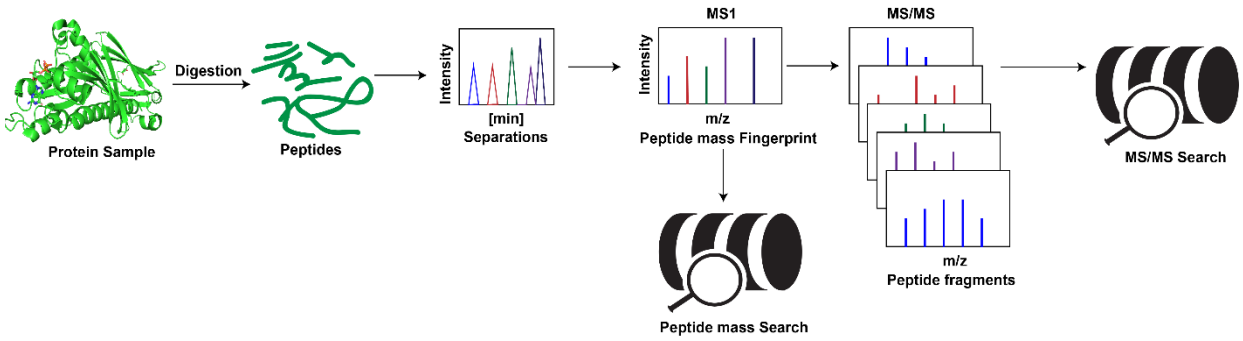


Figure 1.4. Workflow for protein identification through peptide sequencing.

Fragmentation of peptides in tandem mass spectrometry can take place through different ion activation methods. The most popular methods for ion activation are collision-induced dissociation (CID) and Electron transfer dissociation (ETD). In CID, the analyte molecules collide with neutral atoms with high ionization potential in the collision cell. The kinetic energy of the ions is then converted to vibrational energy, leading to collisional activation of the molecule. Once the excess energy acquired is enough to cleave chemical bonds the ions undergo collision-induced dissociation forming fragment ions through cleavage of C-N bonds producing “b” and “y” ions. In ETD reagent molecules such as fluoranthene are converted to their radical anions through corona discharge. Once the radicals are mixed with the analyte molecules that were positively ionized through ESI, electron passes from the radicals to the analyte molecules leading to fragments of the peptides in C-C bonds forming “c” and “z” ions.[2] There are other high energy fragmentation methods such as UVPD that creates b/y, c/z and through homolytic cleavage of the C_α – C bonds, a/x ions.[12] The MS/MS spectra can be searched against a database and matched with the peptide sequences in the database, and proteins are identified.[11] When protein

sequence cannot be found in the database, de novo sequencing of the entire sequence must be derived from mass spectra. In de novo sequencing, different digestion enzymes are used, and cross-referencing of sequence chains obtained from MS/MS fragmentation are compared to get the full sequence. [13]

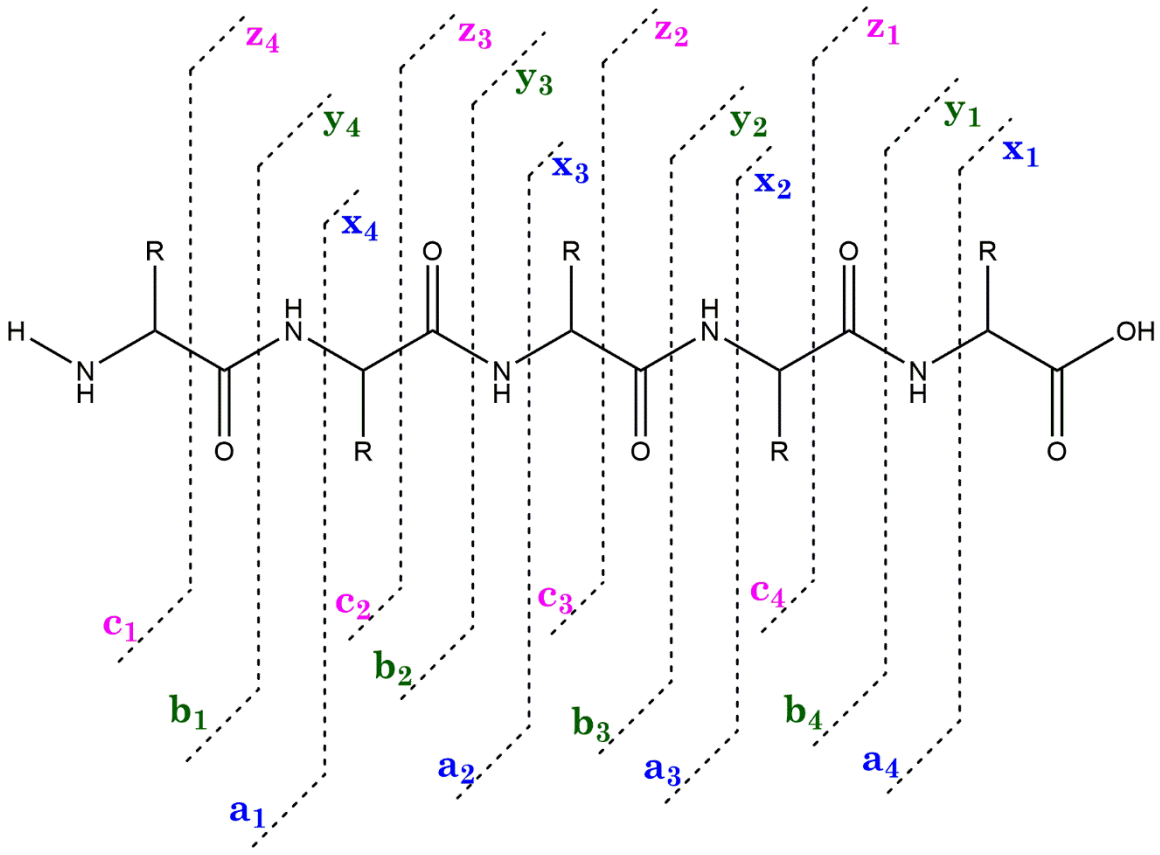


Figure 1.5. Interpretation of peptide MS/MS fragments through different ion activation.

1.8. Label-Free Quantitation

Label-free quantification is a simple, cost-efficient method for determining the relative quantification of proteins present in a sample compared to that of control set in the experiment. There are two reviewed methods available for doing label-free quantification. For label-free quantification, equal amounts of proteins are tryptic digested and loaded into LC-MS/MS for analysis. Peptide abundances can be attributed from the intensity of

MS peaks or from the count of the number identified MS/MS spectra from each peptide. These are known as quantification by peak intensity of the peptide and quantification by a spectral count of MS/MS spectra of peptides. Both correspond to the abundance of the corresponding proteins present in the sample. [14-15]

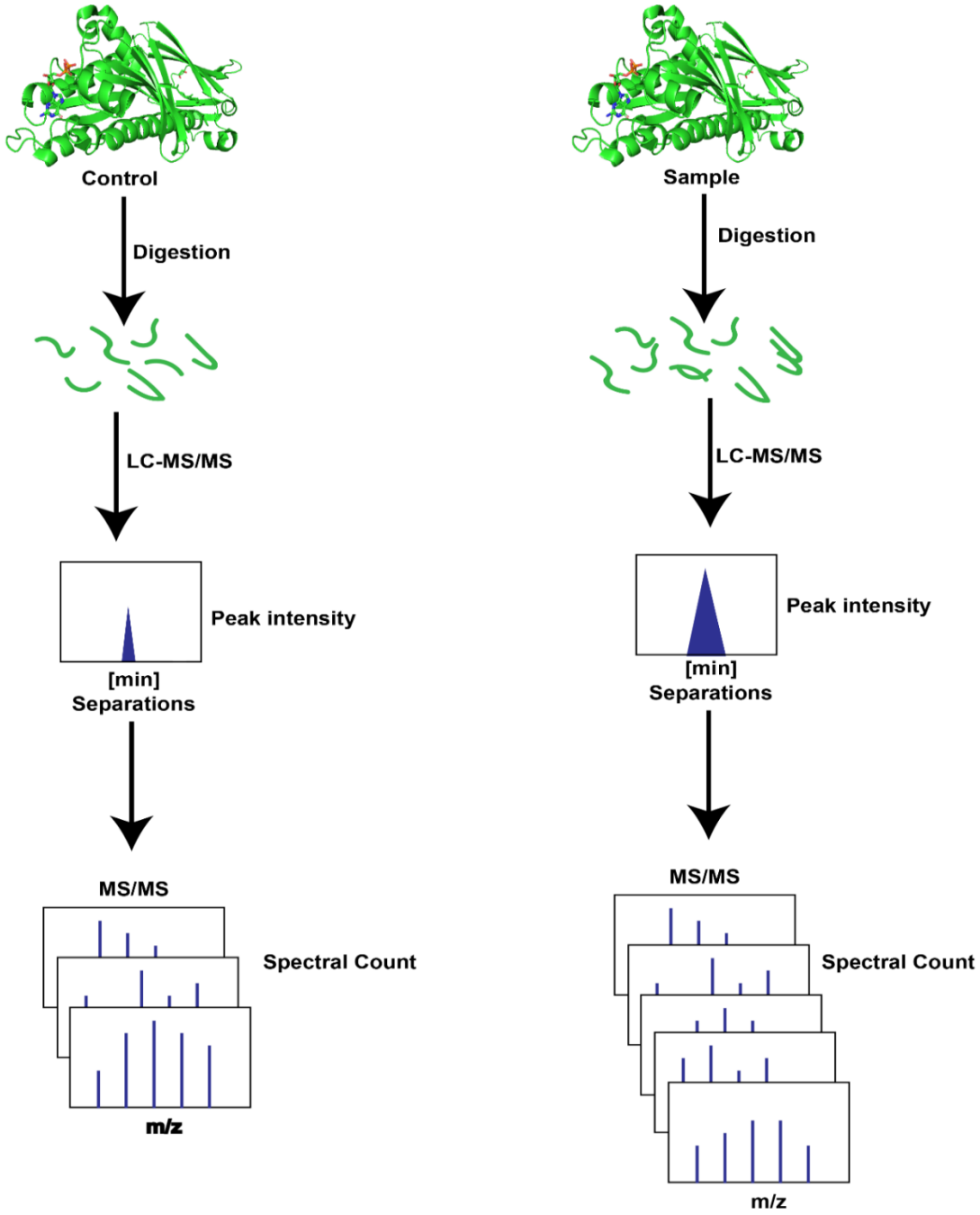


Figure 1.6. Label-free Quantification based on peak intensity and spectral count.

1.9. Co-Immunoprecipitation shotgun proteomics

Proteins are the structural and functional units of the living system. They perform varieties of roles in the living system, e.g., metabolism regulation, immune signaling, maintenance of cellular homeostasis, etc. The complete set of such interactions that takes place in a living system for a protein are its interactome. So, identifying a protein's interacting partners or interactome is crucial for understanding more about a proteins function, mechanism, activation, and inhibition, etc. Co-immunoprecipitation coupled to shotgun proteomics has become a gold standard to characterize the protein interactome using mass spectrometry. Typically, in this method, a cell lysate is pre-cleared with non-immune antibody bound fusion protein A/G coated beads to get rid of native immunoglobulins' bound proteins and then incubated with A/G coated bead-bound antibody against the target protein. The immune complex is then washed followed by 1D / 2D SDS-PAGE and then shotgun proteomics and computational analysis, network visualization and biochemical validation through western blots. [16]

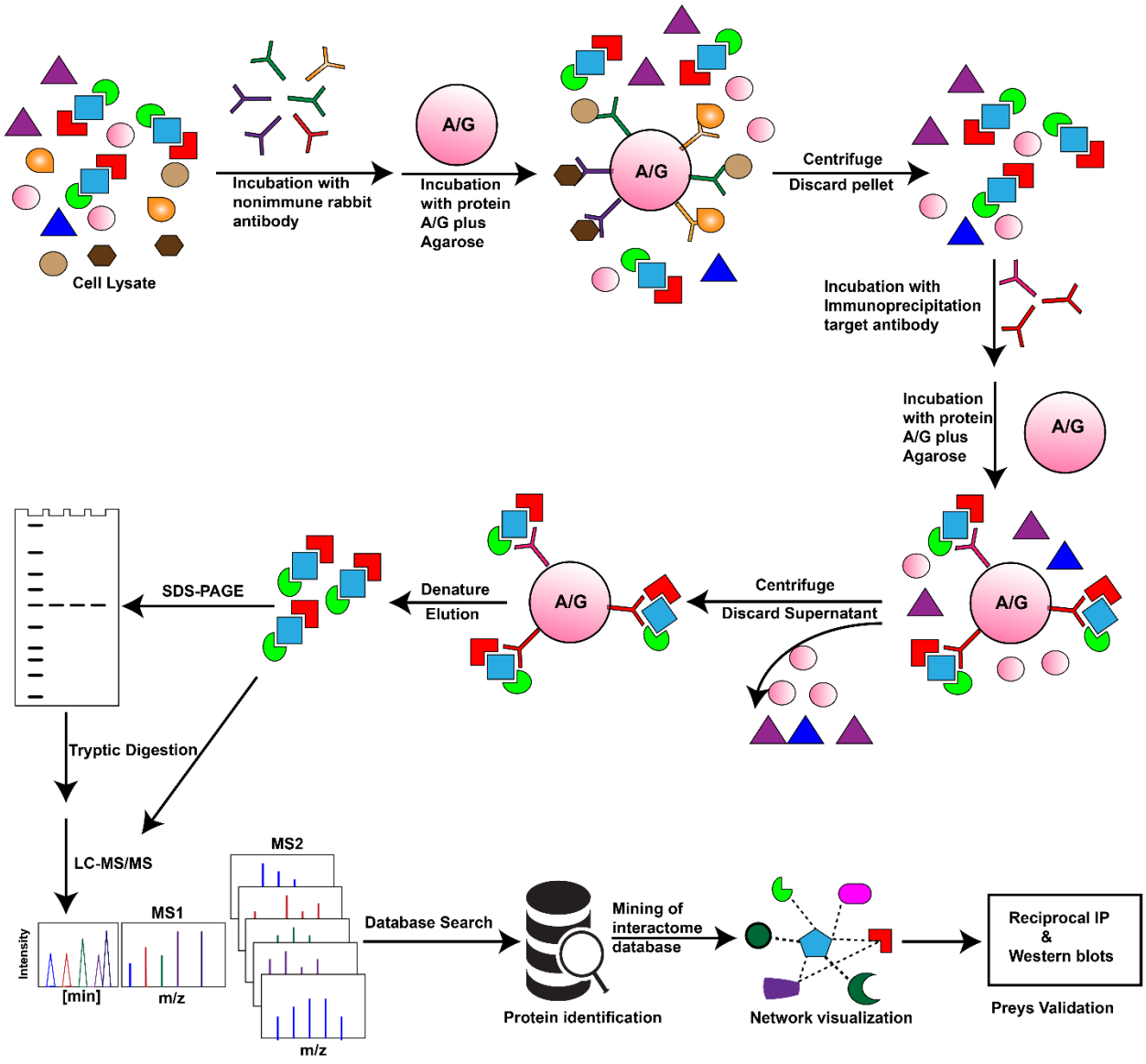


Figure 1.7. Co-immunoprecipitation shotgun proteomics.

1.10. Adaptive and Innate immune response

There are two types of immune response present in the living system, adaptive and innate immune response. Adaptive immunity develops as an adaptation to infection from pathogen leading to lifelong memory against the same type of pathogen. The adaptive immune response includes activation of T and B lymphocytes, production and release of antibody, cytotoxic T lymphocyte-mediated destruction of infected cells. On the contrary,

the innate immune response is inherently present against a wide range of pathogens. It's more of a general response and does not lead to the production of memory in the immune response. Skin functions as the first line of defense in innate immunity. Inside the skin, natural killer cells, macrophages encounter and recognizes structures specific to the microbial origin. This, in turn, activates them for phagocytosis or secretion of cytokines and interferons for inflammation or antiviral response. [17-18]

1.11. Toll-Like Receptor (TLR) Signaling

A major part of the innate immune signaling is imparted through the Toll-like receptors on the surfaces of macrophages and antigen-presenting cells and in endosomes inside these cells. There are 10 different TLRs identified in human. TLR 3, 7, 8, and 9 resides in the endosome since they recognize viral nucleic acids. Others are on the surface of macrophage and other antigen-presenting cells, and they can recognize different microbial origin molecules and lead to activation of macrophage and immune response. For example, TLR1, 2 and 6 are activated by bacterial lipopeptides, TLR2 also can be activated by bacterial peptidoglycan, lipopolysaccharides of gram-negative bacteria, and TLR5 can activate TLR4 can be activated by bacterial flagellin. Once activated through the ligands, TLRs, through their intracellular toll interacting domain, can activate downstream adapter proteins to produce cytokines, e.g. TNF-α, IL-6, IL-12, etc. These produce inflammation, enhance adaptive immunity, and complement activation through the alternate pathway can take place followed by phagocytosis. On the other hand, IL-12 activates natural killer cells that produce IFN-γ. This leads to the production of reactive oxygen species in macrophages and can help the killing of invading microbes. [17-18]

The innate immune response through Toll-like receptor 4 is particularly notable, as TLR-4 can be activated by wide varieties of ligands including bacterial lipopolysaccharides (LPS) that are commonly found in the outer membrane of all gram-negative bacteria. TLR-4 signaling is particularly most complex as well as it requires four adapter proteins for signaling after activation and can function in two distinct pathway through either TRIF or the MyD88 pathway. [19]

Briefly, LPS are components on the cell membranes of gram-negative bacteria. Upon infection with bacteria, LPS comes in contact with cells. With assistance from LPS binding protein (LBP), LPS is transferred to CD14 on the macrophage cell membrane. From CD14, LPS is then transferred to myeloid differentiation (MD2) protein. LPS-MD2 complex then interacts with TLR4 to cause them to dimerize and form the TLR4-MD2-LPS complex. The complex can then initiate downstream signaling through the recruitment of adapter proteins.[20] TLR4 mediated signaling can work through two different pathways, e.g, MYD88 and TRIF dependent pathway. Upon activation, through TLR4, both pathways lead to the activation of nuclear transcription factor NF-$\kappa\beta$. This leads to the production of proinflammatory cytokines, e.g. (IL-1, IL-6, IL-12, etc.), chemokines (CCL2, CXCL8, etc.) and other pro-inflammatory molecules to cause inflammation or activation of adaptive immunity through differentiation of inflammatory T cells and activation of memory T cells.

On the contrary, through endosome formation, in TRIF dependent pathway, activated TLR4 can lead to the production of Type 1 interferons, e.g. IFN-α and IFN-β. Production of Type 1 interferons play different roles in different viral or bacterial infections. With viral

infection, the immune responses include, inhibition of viral gene expression and protein synthesis, degradation of mRNA, activation of NK cells, MHC-I expression and CD8 cytotoxic T-cell mediated killing of infected cells, etc. Type 1 interferon production has been reported with both protective and detrimental effects in case of bacterial infection based on the type of bacteria and infection. Protective effects include maturation and activation of dendritic cells and thus enhanced antigen presentation, T-cell activation, and adaptive immunity. Type-1 interferon also helps IL-12 and IFN-γ production. IFN-γ can also exert immune response through cytokine receptor and activation of antimicrobial genes such as inducible nitric oxide synthase (iNOS), phagocyte oxidase followed by subsequent production of Nitric Oxide and reactive oxygen species to kill bacteria phagocyted through the complement-dependent pathway. Type 1 IFNs also inhibit bacterial invasion through the protection of epithelial barrier, reducing intracellular spore germination. Additionally, Type 1 IFNs lower the survival of infected cells and can thus help limit disease progression. [21-27]

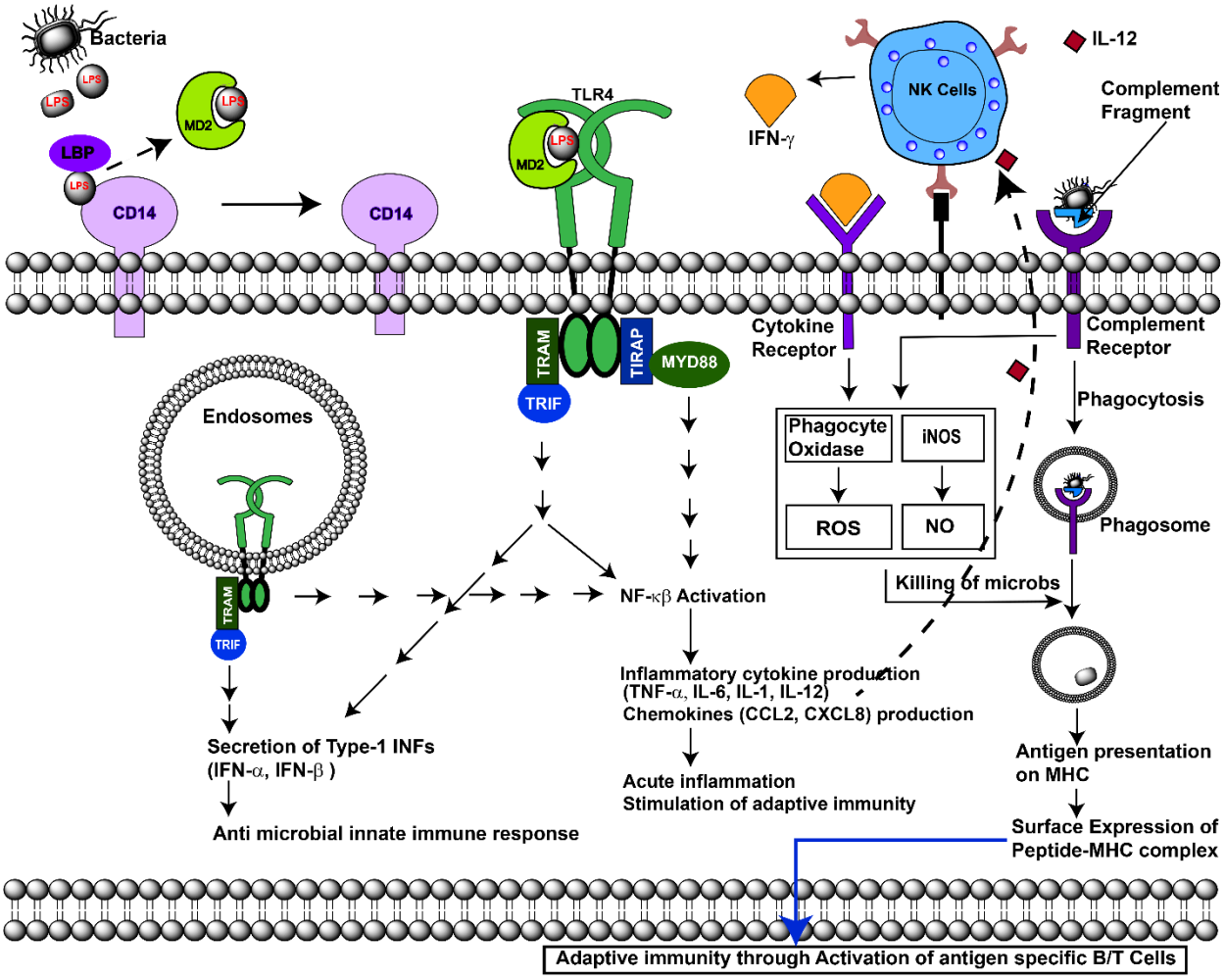


Figure 1.8. Immune signaling through Toll-like receptor 4.

Hence, TLR4 signaling has roles in both innate and adaptive immunity and not every detail about TLR4 mechanism is known; hence new interacting partners are sought which can lead to the discovery of a new drug target. Different TLR agonists are already sought as cancer drugs.[28] Infectious diseases that target the immune signaling network as such of Dengue and Ebola can be better tackled with a better understanding of immune signaling network. Hence, TLR4 was a good pick for immune signaling network analysis study.

1.12. Thesis organization

This dissertation is focused on two projects. Chapter 2 focuses on the improvement of sample preparation method for mass spectrometry-based cellular proteomics. After methanol-chloroform based protein purification, the detergents are removed from the cell lysate. But for analyzing the proteins through mass spectrometry, they need to be re-solubilized before tryptic digestion. But not all the cell lysate proteins would go back to the solution phase again in the absence of detergents. Here we show how the introduction of a mass spectrometry compatible surfactant can improve the re-solubilization of proteins followed by higher identification of proteins.

Chapter 3 discusses, the immune signaling network study of toll-like receptor 4. Here we show targeted immune precipitation of toll-like receptor 4 (TLR4) along with chemical crosslinker captures the stable and transiently interacting protein partners in TLR4 mediated immune signaling network. From label-free quantitation of bottom-up proteomics data, we identify some candidate partners as putative partners in TLR4 interacting immune signaling network. In the final step, one of the targets is biochemically validated through western blot, and we confirm the target as a partner in TLR4 mediated signaling network. Chapter 4 summarizes both the projects with a concluding note.

Chapter-2

Improved in-solution trypsin digestion method for Methanol-chloroform precipitated cellular proteomics sample

2.1 Abstract

Methanol-chloroform based protein precipitation is an essential step in many LC-MS/MS based cellular proteomics application. This precipitation helps remove detergents and salts present in the cell lysis buffer and hence has been a method of choice for sample preparation before mass spectrometric analysis. Many membrane proteins and proteins with large hydrophobic domains may behave as poorly soluble if not insoluble in water after having been precipitated from detergent rich solution. Therefore, re-solubilization of the total precipitate is not possible using regular in-solution digestion protocol. We used a phase transfer surfactant Deoxycholic acid (DCA) to improve re-solubilization of Methanol-chloroform precipitated proteome, which increased more protein identification in bottom-up proteomics. We demonstrated our modified method using an equal amount of macrophage cell lysate. Detailed in-solution trypsin digestion studies were presented on un-normalized/normalized methanol-chloroform precipitated samples with or without DCA treatments. The mass spectrometric analysis confirmed a higher number of protein identification in modified samples compared to controls. This improved digestion method leads to increased protein identification and can be used in any type of large-scale bottom-up proteomics study where Methanol-chloroform purification and in-solution digestion will be utilized.

2.2 Introduction

Solution-based bottom-up proteomics approach is a popular method of choice for global identification and quantification of proteins in complex biological samples. Highly complex protein samples are digested into peptides and then analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS). Based on the MS/MS spectra and corresponding database search, the peptides are sequenced, and proteins present in the sample are identified. [29-33] Efficiency of bottom-up proteomics application depends on many aspects, e.g., cell type, cell lysis, protein extraction/solubilization, proteolysis conditions, salt, and detergent removal, fractionation methods, etc. [34-36] Mammalian proteome is diverse, complex and dynamic like expression. Membranes and organelles have distinct roles in the cell. Membrane proteins are essential components of the membrane and play different roles as immune receptors, ion channels, transporter and adapters of the cellular signaling network, etc. [37-41] Proteins found in organelle membranes also perform similar functions as ion channels, transporter and, receptors adaptors of signaling, etc. [42-46]

In most global proteomics study, we want to study the overall cellular function that includes the membrane and organelle associated proteins and cytosolic soluble proteins. Sometimes biochemical methods are used to purify target organelle or to isolate the membranes alone to work on the specific component. [47] However, many membrane and organelle associated proteins are hydrophobic. To get them into solution, protein extraction buffer for cell lysis includes detergents and protein denaturing reagents. Sodium dodecyl sulfate (SDS) is a strong detergent used in the lysis buffer for protein

extraction and solubilization. SDS in average binds after every two amino acid in the polypeptide chain. Thus, proteins attain approximately constant charge per unit of mass with an overall negative charge, and this helps denature and dissolve proteins in solution. [48-49]

However, there are some problems associated with downstream mass spectrometry analysis of samples that are prepared with SDS. It has been shown that the presence of SDS can introduce up to 15 dodecyl sulfates adducts along with Na^+ ion adducts into the molecular ion of protein myoglobin. [50] The presence of highly surface-active anionic surfactants can also interfere with electrospray ionization in positive ion mode. This can lead to reduced spray efficiency. [51] Moreover, it has been observed that the presence of SDS in samples reduces the signal to noise ratio. [52-53] In addition to this, for bottom-up proteomics, presence of SDS reduces trypsin activity. [54] Thus, SDS needs to be removed before the bottom-up proteomics experiment.

There are many ways that have been reported as a method for removing SDS before mass spectrometric analysis. These include electrophoretic separation, [55] hydrophobic interaction chromatography, [56] organic solvent precipitation, [57-58] ion-pairing reagents, [9] potassium dodecyl sulfate precipitation, [59] oligosaccharides based spin column, [60] filter assisted sample preparation (FASP), [61] and enhanced filter assisted sample preparation (eFASP)[62], etc.

These methods have different advantages and disadvantages. Different organic solvents, in combination with salts or acids, are reported to be used in organic solvent-based precipitation of proteins and thus remove SDS. Two most popular among these methods are acetone purification and Methanol-chloroform precipitation. These methods are a good choice for protein precipitation for bottom-up proteomics because of low cost, low sample preparation time required, minimizes protein degradation, and relatively simple preparation procedure. [9-10, 63-67]

However, bottom-up proteomics experiment following organic-solvent precipitation of proteins suffers from under-representation of precipitated proteins in MS detection platform. The MS-compatible solvent system is not able to solubilize all the hydrophobic component of the precipitate. [68] So, here, our objective is to improve the solubilization and identification of proteins that have been purified through Methanol-Chloroform purification.

It has been shown in previous work by Takeshi Masuda et al. that a phase transfer surfactant, sodium deoxycholate can improve protein identification in isolated membrane enrich fractions of E.coli and HeLa cells. [69] Short gun proteomics on standard filtration device has been developed with improved identification when the exchange buffer for filtration device had 8M Urea or Deoxycholic acid. [62, 70-73]

Here we demonstrate an improved Methanol-chloroform precipitation method using mouse RAW 264.7 macrophage proteome and solubilize the proteome in Deoxycholic

acid or Sodium Deoxycholate solution to further tryptic digest and analyze the efficiency of this method compared to the conventional methods. Detailed experiments were conducted using Methanol-chloroform precipitation on equal amount macrophage cell lysate, or equal amount of proteins normalized after methanol chloroform purification and compared with protein identification in DCA treated and untreated solution. We also evaluated the effectiveness of Deoxycholic acid-mediated digestion on the leftover digested pellet of regular Methanol-chloroform precipitated samples. Our study helps establish this modified method of Methanol-chloroform precipitation followed by resuspension and tryptic digestion in Deoxycholic acid as an efficient method for in solution digestion-based bottom-up proteomics application in global proteomics study.

2.3 Materials and Methods

2.3.1 Chemicals:

Sodium Deoxycholate (SDC), Deoxycholic acid (DCA), Iodoacetamide (IAA), Formic acid (FA), Ammonium Bicarbonate (NH_4HCO_3), were purchased from Sigma-Aldrich (St. Louis, MO, USA). Dulbecco's Modified Essential Medium (DMEM) was purchased from Corning Cell gro (Tewksbury, MA, USA) Methanol; Acetonitrile was purchased from Alfa Aesar by thermo scientific (Tewksbury, MA, USA). Proteomics sequencing grade modified trypsin was purchased from Promega Corporation (Madison, WI, USA). Pierce BCA Protein Assay kit was purchased from Thermo Fisher Scientific (Waltham, MA, USA).

2.4 Experimental Procedures

2.4.1 Cell Culture:

Raw 264.7 macrophage cells were grown in Dulbecco's Modified Essential Medium (DMEM) in the presence of 10% fetal bovine serum and 1% penicillin/streptomycin. Cells were grown in an incubator with 37^{o} C in humidified condition with a continuous supply of 5% CO_2. After a passage of culture, once cells grown to near confluence, they are rinsed three times with 1X Phosphate buffer saline (PBS) and then scrapped and hence collected for lysis and proteomics analysis.

2.4.2 Protein Extraction, quantification

Ripa lysis buffer, in combination with protease inhibitor, was used to lyse the cells at 4^{o} C for 15 minutes. The cells suspension was then sonicated for 15 minutes. This is followed by a final incubation of 30 minutes at 4^{o} C. The supernatant was collected into a new tube, and cell debris was saved separately. The protein concentration in the supernatant was measured with a BCA Assay kit, using bovine serum albumin as standard according to the manufacturer’s protocol.

2.4.3 Protein purification, digestion:

2.4.3.1 Methanol-chloroform purification and digestion in ammonium bicarbonate (ABC)

In the first phase, 150 µg of Raw 264.7 macrophage cells, in three replicates (Tag: MeOH-Child) were purified using Methanol-chloroform method as described. [74] Briefly, Proteins are diluted to 1µg/ µL. One volume of protein is mixed with 4 volume of methanol and

briefly vortexed for 30 seconds. Then, combined with one volume of chloroform and vortexed for 30 seconds. Afterward, three-volume of water is added and vortexed for 30 seconds and centrifuged at 20,000 × g for 10 minutes at room temperature. The upper phase is discarded, keeping the white precipitate. Then, three-volume of methanol is added, gently mixed, without breaking the pellet. Finally, centrifugation at 20,000 × g precipitates the proteins as a final pellet. The supernatant is discarded, and proteins in the pellet are taken for bottom-up proteomics. The proteins were then reduced and alkylated, then digested with trypsin (MS Grade) at a 1:100 enzyme/protein concentration for 16 h at 37 °C. Formic acid was added afterward to drop the pH and to stop trypsin activity. The samples were then desalted using a C18 desalting column (Thermo Scientific, IL, USA). After completely drying by speed vacuum, peptides were dissolved in 0.1% formic acid, and stored at -20 °C before LC-MS/MS analysis.

In the second phase, 250 µg of Raw 264.7 macrophage cells, in nine aliquots were taken and then purified using Methanol-chloroform method as described. [74] They were then suspended in 300 µL, 50 mM NH_4HCO_3. From these nine replicates, we prepared three replicates by combining three replicates in one microcentrifuge tube. Proteins were normalized in each replicates, then reduced, alkylated, digested, desalted, centrifuged and the supernatants were stored at -20°C before LC-MS/MS analysis.

2.4.3.2 Deoxycholic acid (DCA) Assisted Tryptic digestion of methanol-chloroform purified proteins:

The extracted proteins (150 µg and 250 µg), in three biological replicates, (Tag: MeOH-Chl-DCA) were purified using Methanol-chloroform method as described. [74] This is

followed by our modified method, where the sample was air-dried for 5 minutes. All Solutions: 10 mM DTT, 10 mM Iodoacetamide, Trypsin stock solution (20 µg /200 µL) were prepared in 50 mM NH_4HCO_3, pH 7.5. Two % Deoxycholic Acid (DCA) were prepared in 7N NH_4OH in water, and then NH_4HCO_3 was added so that the final solution will have 2% Deoxycholic acid in 50 mM NH_4HCO_3. Then pH adjusted to ~ 7.0 -8.0. (With pH paper) (conc. NH_4OH was used to adjust pH). 100 µL of 2% Deoxycholic Acid (DCA) in 50 mM NH_4HCO_3 was added to the sample, and then the sample was rested at 25º C for 30 minutes. 300 µL of 50 mM NH_4HCO_3 was added to the sample and rinsing with a 200 µL pipette tip; the samples were dissolved. For 250 µg samples, protein amount in three biological replicates was normalized, so the same amount of proteins are used for digestion in regular and DCA treated solution. After reduction and alkylation, the final volume is increased to 1 mL with 50 mM NH_4HCO_3. The proteins were then digested with trypsin (MS Grade) at a 1:100 enzyme/protein concentration for 16 h at 37 °C. Formic acid was added afterward to drop the pH, to stop trypsin activity, and to precipitate the DCA. Next, centrifugation at 20000 rpm for 30 minutes at 4 ^oC removes the surfactant DCA. The samples were then desalted using a C18 desalting column (Thermo Scientific, IL, USA). After completely drying by speed vacuum, peptides were dissolved in 0.1% formic acid, and stored at -20 °C before LC-MS/MS. A detailed protocol of each step is included in the supporting information page 1-2.

2.4.3.3 Sodium deoxycholate (NaDCO) Assisted Tryptic digestion of methanol-chloroform purified proteins

The extracted proteins (150 µg and 250 µg), in three biological replicates, (Tag: MeOH-Chl-NaDCO) were purified using Methanol-Chloroform method as described. [74] This is followed by our modified method. Where the sample was air-dried for 5 minutes. All Solutions: 10 mM DTT, 10 mM Iodoacetamide, Trypsin stock solution (20 µg /200 µL) were prepared in 50 mM NH_4HCO_3, pH 7.5. 2% Sodium Deoxycholate were prepared in 7N NH_4OH in water, and then NH_4HCO_3 was added so that the final solution will have 2% Sodium Deoxycholate in 50 mM NH_4HCO_3. Then pH adjusted to ~ 7.0 -8.0. (With pH paper) (conc. NH_4OH was used to adjust pH). 100 µL of 2% Sodium Deoxycholate in 50 mM NH_4HCO_3 was added to the sample, and then the sample was rested at 25º C for 30 minutes. 300 µL of 50 mM NH_4HCO_3 was added to the sample and rinsing with a 200 µL pipette tip; the samples were dissolved. After reduction and alkylation, the final volume is increased to 1 mL with 50 mM NH_4HCO_3. The proteins were then digested with trypsin (MS Grade) at a 1:100 enzyme/protein concentration for 16 h at 37 °C digestion. Formic acid was added afterward to drop the pH, to stop trypsin activity, and to precipitate the NaDCO. Next, centrifugation at 20000 rpm for 30 minutes at 4 ^oC removes the surfactant NaDCO. The samples were then desalted using a C18 desalting column (Thermo Scientific, IL, USA). After completely drying by speed vacuum, peptides were dissolved in 0.1% formic acid, and stored at -20 °C before LC-MS/MS. A detailed protocol of each step is the same as that of DCA, provided in the supporting information (section 2.7)

DCA Assisted digestion of discarded protein pellets after Methanol-chloroform purification [74] (Tag: Digested-Pellet-DCA), detailed methods for Acetone Purification and digestion

and Detergent removal Spin Column Filtration (DRSC) and digestion in Ammonium Bicarbonate solutions were provided in the supporting information. [75]

2.4.4 LC-MS/MS experiments

Detailed experimental methods were provided in the supporting information for data acquisition by Thermo Velos Pro and Orbitrap Lumos.

2.4.5 Data Analysis

Proteins were identified through Proteome Discoverer software (ver. 2.1, Thermo Fisher Scientific) and a mouse (Mus musculus) UniProt protein sequence database (16973 sequences, and 9602132 residues). The reviewed protein sequences of the mouse were downloaded from UniProt protein database (www.uniprot.org) November 30, 2018. The considerations in SEQUEST searches for normal peptides were used carbamidomethylation of cysteine and oxidation of methionine as dynamic modifications. Trypsin was indicated as the proteolytic enzyme with two missed cleavages. Peptide and fragment mass tolerance were set at ± 1.6 and 0.6 Da, and precursor mass range of 350-5000 Da and peptide charges were set excluding +1. SEQUEST HT results were filtered with the Percolator-based scoring to improve the sensitivity and accuracy of the peptide identification. Using a decoy search strategy, target false discovery rates for peptide identification of all searches were utilized at less than 1% with at least two peptides per protein, and the results were strictly filtered by Δ Cn (< 0.01), Xcorr (≥ 1.5) for peptides, and peptide spectral matches (PSMs ≥ 5) with high confidence with q value (< 0.05). Protein quantification was conducted using the total spectrum count of identified proteins.

Additional criteria were applied to increase confidence that PSMs must be present in at least two biological replicate samples.

2.4.6 Bioinformatics Analysis

Proteins were functionally categorized using gene ontology system by PANTHER classification system based biological processes, molecular activity, cellular components, and protein classes. [76]

2.4.7 Statistical Analysis

The quantitative analysis of proteins as PSMs was performed using built-in-statistical packages in Proteome Discoverer (Ver. 2.1). Results were considered statistically significant if $q \leq 0.05$. Scatter plots, and pairwise correlation matrices were generated data corresponding to the number of peptides identified using the R package, where results were considered if correlation coefficient (R) was > 0.80 and are shown in supporting Figures S2.3A and S2.3B.

2.5 Results and Discussion

In our regular experiments with cell lysates, we always observe the Methanol-chloroform precipitated sample behave recalcitrant in dissolving into the tryptic digestion buffer of 50 mM NH_4HCO_3. We know from the literature review, Sodium Deoxycholate is used to dissolve isolated membrane and Deoxycholic acid is used to equilibrate micron filters of a detergent removal method, enhanced Filter Assisted Sample Preparation (eFASP). We came up with a modification in the method of tryptic digestion where precipitated pellet

samples were dissolved in 0.5% Deoxycholic acid in 50 mM NH_4HCO_3 and digested in a total concentration of 0.2% Deoxycholic Acid in 50 mM NH_4HCO_3. Similarly, pellets were dissolved with 0.5% Sodium Deoxycholate in 50 mM NH_4HCO_3 and digested in 0.2% Sodium Deoxycholate in 50 mM NH_4HCO_3. Here, we expected, the use of surfactants, i.e., Deoxycholic acid or Sodium Deoxycholate can take the precipitated proteins of methanol chloroform precipitate fully into the solution and hence this would lead to improved protein solubilization, digestion, and identification. Another advantage is DCA can be precipitated from the solution after digestions.

To test this, we first took 150 µg of RAW 264.7 cell lysates in three replicates for control and our modified method. We first did Methanol-chloroform precipitation of proteins and dissolved and tryptic digested the pellets in regular tryptic lysis buffer, 50 mM NH_4HCO_3 (Figure 2.1A). Similarly, we dissolved the Methanol-chloroform precipitated proteins in 0.5% Deoxycholic Acid (DCA) in 50 mM NH_4HCO_3 or 0.5% Sodium Deoxycholate (Na-DCO) in 50 mM NH_4HCO_3 (Figure 2.1B). Finally, the dissolved proteins were diluted and tryptic digested in 0.2% respective surfactant solution in 50 mM NH_4HCO_3. Acidification followed by centrifugation removed the surfactants, and after desalting, samples were analyzed in LC-ESI-Thermo Velos Pro mass spectrometer. For the purpose of comparison, we also used other purification methods for purification of 150 µg of cell lysate and did the sample preparation according to Acetone purification, Detergent removal spin column purification followed by tryptic digestion in Ammonium Bicarbonate (see the supporting information page S4-S6).

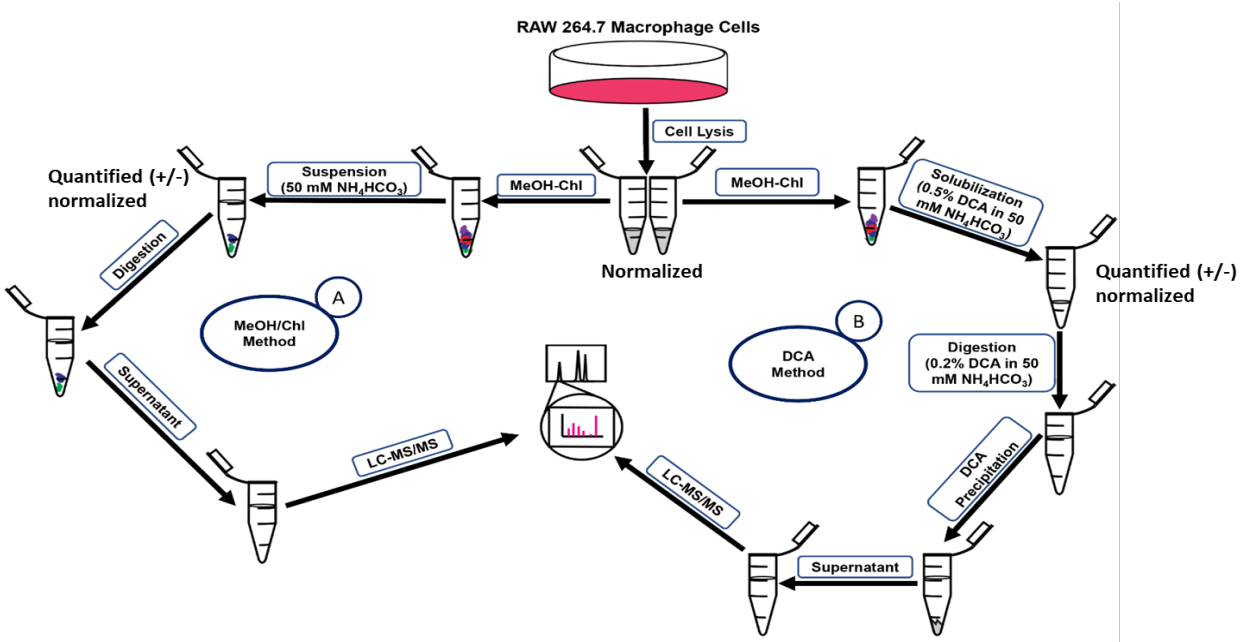


Figure 2.1. Two experimental schemes with steps of sample preparations are shown A) Shows Methanol-chloroform purification followed by digestion in 50 mM Ammonium Bicarbonate. B) Shows Methanol-chloroform-DCA method where Methanol-chloroform precipitated pellet was dissolved in Deoxycholic acid and tryptic digested. Please see supplementary data S1 for more information.

In another experiment, three biological replicates samples were prepared using 250 µg of cell lysate. As mentioned, significant loss occurs after methanol-chloroform purification. Thus 9 samples were prepared and after Methanol-chloroform purification, were combined to make three replicates for regular NH_4HCO_3 method whereas three 250 µg of cell lysate replicates were used for DCA treated solution. This way, it is easier for us to normalize the number of proteins in each replicate. The proteins amounts were normalized to an equal amount before digestion. LC-ESI-Thermo Orbitrap Fusion Lumos Tribrid Mass Spectrometer was used for high-resolution data. We also prepared a third

set of samples in this phase of the experiment. We took the left-over pellet after regular tryptic digestion and used a surfactant to dissolve this leftover pellet and re-digested this one in the surfactant solution (supporting Figure S2.1C).

2.5.1 Visual Assessment and Biochemical Quantification

In the first stage of our experiment, we took 150 µg of RAW 264.7 macrophage cell lysate, performed Methanol-chloroform purification and then reconstituted the pellets in regular Tryptic digest buffer 50 mM Ammonium Bicarbonate and our modified buffer 0.5% Deoxycholic Acid or 0.5 % Sodium Deoxycholate in 50 mM Ammonium Bicarbonate buffer, pH ~ 7.5. The visual assessment showed (Figure 2.2) both Deoxycholic Acid and Sodium Deoxycholate took all the pellets into the solution. This was further validated through BCA assay of the dissolved proteins to quantify roughly how much proteins have been solubilized. Similar Analysis was performed using 250 µg of proteins. From both the analysis, we confirmed ~ three times more proteins are dissolved when assisted with either of the surfactant addition (Supplementary table S2.1).

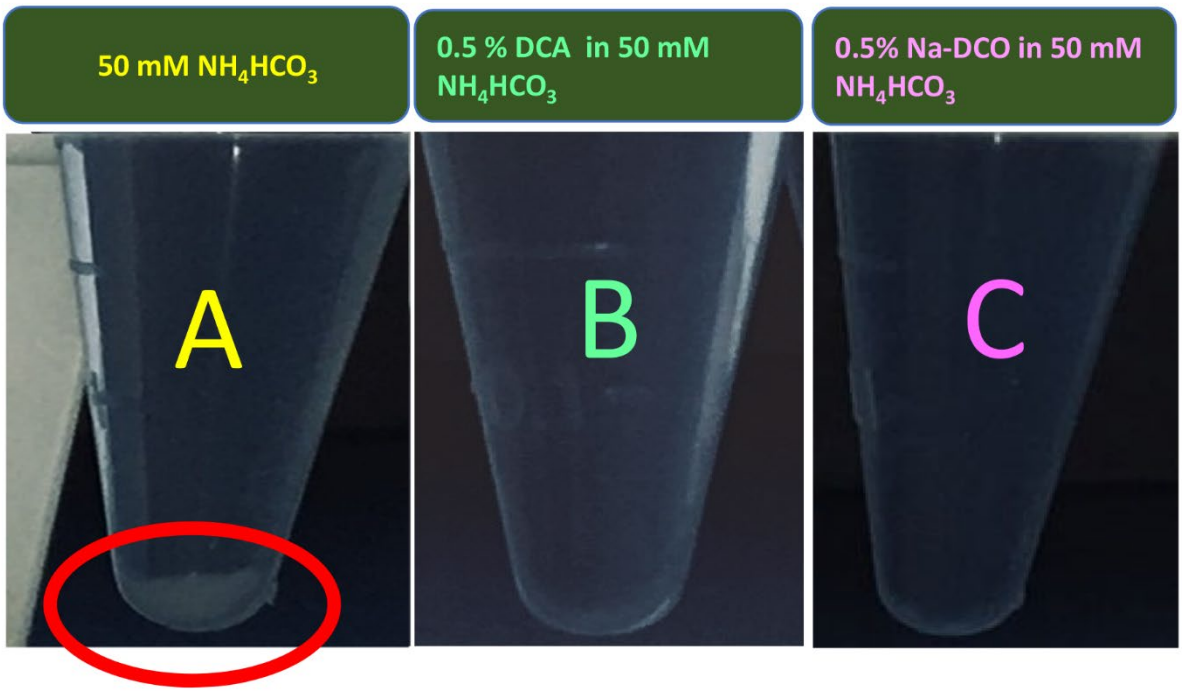


Figure 2.2. Visual assessment of Methanol-chloroform precipitated proteins dissolved in (A) 50 mM NH_4HCO_3, (B) Deoxycholic acid (DCA) in 50 mM NH_4HCO_3, or (C) Sodium Deoxycholate (Na-DCO) in 50 mM NH_4HCO_3. Visual Assessment confirmed the use of (B) Deoxycholic acid (DCA) or (C) Sodium Deoxycholate (Na-DCO) took all the protein pellets form Methanol-chloroform precipitation, back into the solution compared to (A) regular tryptic lysis buffer 50 mM NH_4HCO_3.

2.5.2 Protein and peptide identification

We compared two conventional sample preparation methods (e.g., Acetone purification, DRSC filtration) in combination with proper tryptic digestion in 50 mM Ammonium Bicarbonate versus our two modified sample preparation methods using Deoxycholic acid and Sodium Deoxycholate. (Figure 2.3) 150 µg of RAW 264.7 Macrophage cell lysates were taken, and different sample purification methods, e.g., Methanol-chloroform

purification, Acetone purification, Detergent removal Spin column filtration (DRSC) were used to remove detergents from samples followed by in-solution digestion in 50 mM Ammonium Bicarbonate (ABC). At the same time, we took proteins for Methanol-chloroform precipitation and after precipitation, used 0.5% Deoxycholic Acid or 0.5 % Sodium Deoxycholate in 50 mM Ammonium Bicarbonate (ABC) buffer, pH ~ 7.5 for dissolving the precipitate and diluted to 0.2% respective surfactant concentration with 50 mM Ammonium Bicarbonate for tryptic digestion.

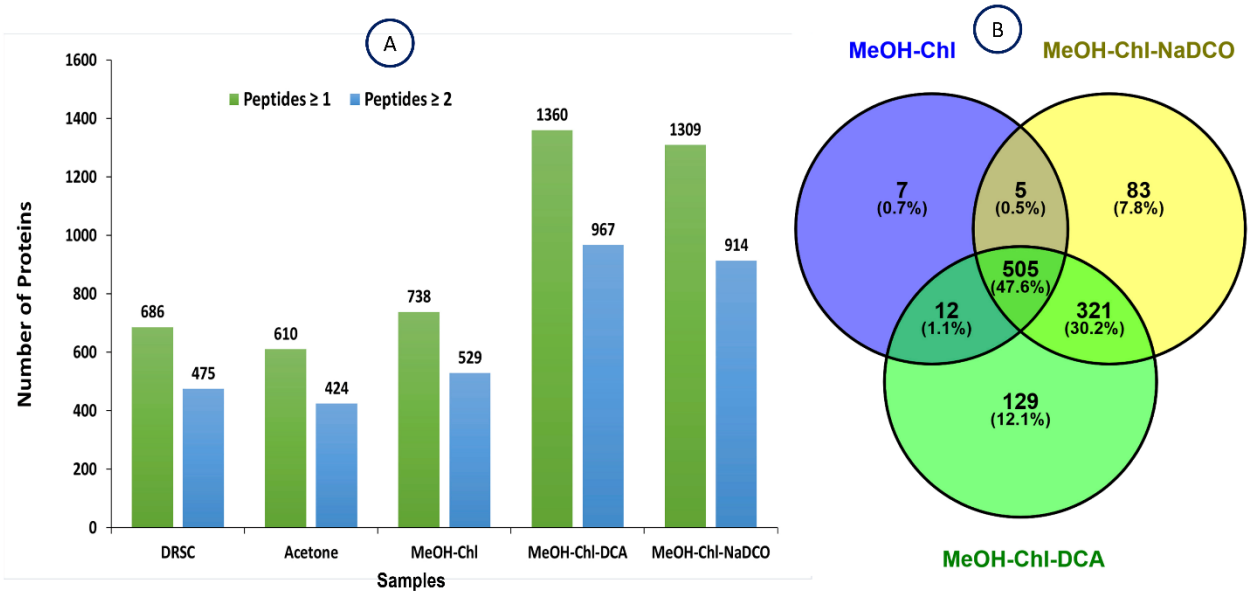


Figure 2.3. Commonly identified proteins and Venn diagram of protein IDs from RAW 264.7 Macrophage cell lysates prepared through different protein precipitation methods (un-normalized) followed by in-solution digestion. A) Shows commonly identified proteins among all biological replicates after different precipitation and improved methods using LC-ESI-Thermo Velos Pro Mass Spectrometer (un-normalized, after precipitation). B) Venn diagram shows proteins upon Methanol-chloroform purification, and tryptic digestion in 0.2% deoxycholic acid in 50 mM Ammonium Bicarbonate identified the

highest number of unique proteins compared to tryptic digestion in 50 mM Ammonium Bicarbonate or 0.2% Sodium Deoxycholate in 50 mM Ammonium Bicarbonate.

Finally, the results of three biological replicates were merged, and we show the data sets where at least one peptide or two peptides are identified for each identified protein from the three replicates of each sample set and controls. Protein ID's that were not represented in all the three replicates were excluded from analysis for these datasets. (Figure 2.3A and data not shown) In this set of data, we observed, regular in-solution digestion in 50 mM NH_4HCO_3, samples prepared through Methanol-chloroform purification gave higher protein identification (738 and 529, at least one or two peptides respectively from each protein identified) compared to Acetone or DRSC method. The protein identification from Methanol-chloroform based precipitation pellet increased once DCA or NaDCO in 50 mM NH_4HCO_3 was used for digestion. We observed the highest number of identification (1360 and 967, for at least one and two peptides respectively per identified proteins) from DCA mediated tryptic digest sample and (1309 and 914, for at least one and two peptides respectively per identified proteins) for NaDCO mediated tryptic digest sample. This is an 84% (number of identified peptides per protein $\geq$ 1) or 82% (number of identified peptides per protein $\geq$ 2) increase in protein identification for DCA compared to regular MeOH-Chl with NH_4HCO_3 buffer method. For NaDCO the identification is 77% (number of identified peptides per protein $\geq$ 1) or 72% (number of identified peptides per protein $\geq$ 2) higher compared to regular MeOH-Chl with NH_4HCO_3 buffer method.

We also generated a Venn diagram with identified proteins from Methanol-Chl method with Methanol-Chl-DCA and Methanol-Chl-NaDCO method where the number of peptides identified per protein is ≥ 2. (Figure 2.3B) From the Venn diagram 450 proteins (321 +129) are uniquely identified with Methanol-Chl-DCA method compared to the control (Methanol-Chl). On the contrary, only 12 proteins (7+5) are identified uniquely to Methanol-Chl method that is not identified in Methanol-Chl-DCA method. Thus, this is 37.5 (450/12) times more identification of unique protein in our modified method with DCA, compared to the control with ABC. (Supplementary Table S2.3)

For Methanol-Chl-NaDCO method, 404 proteins (321 + 83) were uniquely identified with Methanol-Chl-NaDCO method compared to the control (Methanol-Chl). On the contrary, only 19 proteins (12 +7) are uniquely identified in Methanol-Chl method that is not identified in Methanol-Chl-NaDCO method. Therefore, this is 21.2 (404/19) times (number of peptides identified per protein ≥ 2) more identification of unique protein in our modified method with NaDCO, compared to that of control with ABC. This clearly suggests that the modified method is highly efficient in identifying more proteins. Methanol-chloroform purification technique is a widely used protein precipitation technique, but we clearly see a significant amount of protein loss during precipitation.

Since Sodium Deoxycholate is the sodium salt of Deoxycholic acid, and we got similar results from both, we decided to proceed further with Deoxycholic acid for the next stage of the experiment.

To further evaluate proteins identified by the modified method, we quantified and normalized the same amount of proteins in Methanol-chloroform, and Methanol-chloroform DCA treated samples. Samples preparations details were described before. Three replicates samples were dissolved in 0.5% Deoxycholic acid in 50 mM NH_4HCO_3. This was further diluted to 0.2% Deoxycholic Acid solution with a 50 mM NH_4HCO_3 and tryptic digested in this condition. We have found previously that resuspension in DCA yields thrice as much proteins dissolved into the solution phase compared to Ammonium Bicarbonate buffer alone. Thus, in this stage, we ensured we had an equal amount of cell lysate proteins, and after Methanol-chloroform precipitation, we normalized protein amount in control, and DCA treated samples. This way, we can assess the categories of proteins which are coming into the solution due to the DCA treatment. After digestion, these three replicates were sent for analysis with high-resolution mass spectrometry. One replicate was an outlier, so we excluded that from the analysis. The left-over pellets of these three biological replicates, after overnight digestion were washed with methanol and solubilized with 0.5% Deoxycholic acid in 50 mM NH_4HCO_3 and diluted with 50 mM NH_4HCO_3 to 0.2% Deoxycholic acid in 50 mM NH_4HCO_3 solution for tryptic digestion. All samples from this phase were sent to Purdue Proteomics Facility for analysis in LC-ESI-Thermo Orbitrap Fusion Lumos Tribrid Mass Spectrometer.

The results of all biological replicates were combined, and the protein IDs that were commonly identified in all replicates are shown in Figure 2.4A. Methanol-chloroform purification in combination with digestion in DCA in 50 mM ammonium bicarbonate (MeOH-Chl-DCA) method, ends up identifying 3217 proteins (at least one peptide

identified per protein) or 2560 proteins (at least two peptides identified per protein). On the contrary, Methanol-chloroform purification in combination with ABC (MeOH-Chl) identified 2872 proteins (at least one peptide identified per protein) or 2242 proteins (at least two peptides identified per protein). This is 12% to 14% less than our modified method. The left-over pellet after digestion of Methanol-Chl method identified 2439 proteins (at least one peptide identified per protein) or 1782 proteins (at least two peptides identified per protein). This confirmed that a large number of proteins are not taken for in-solution digestion unless DCA is used for dissolving the precipitation.

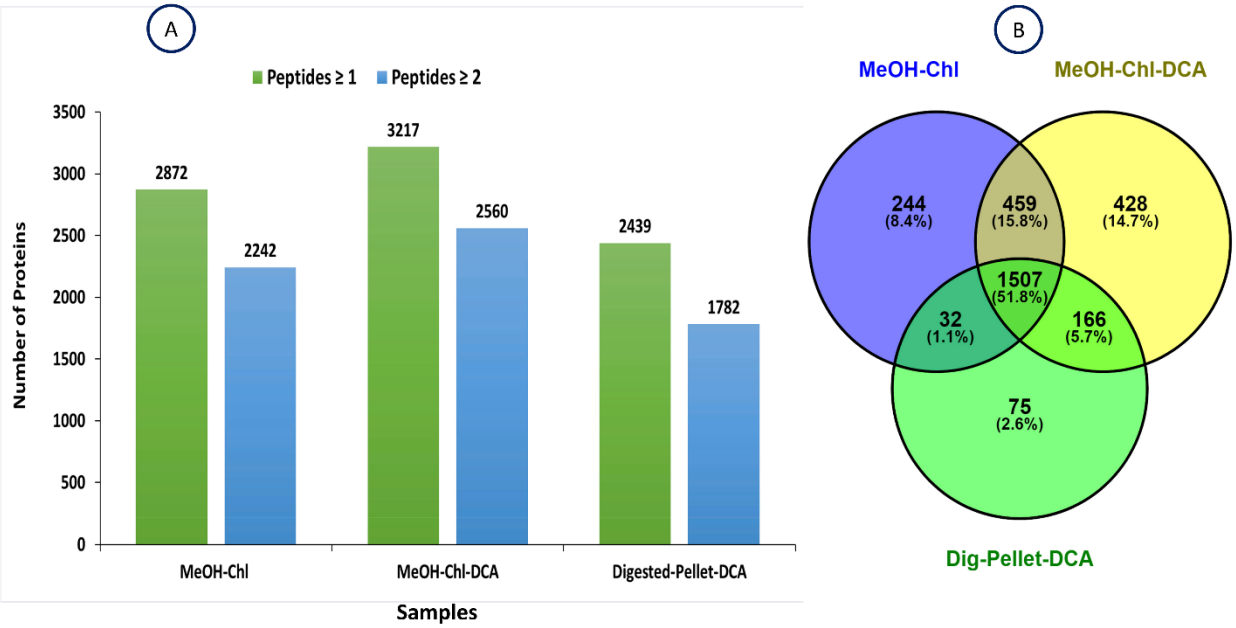


Figure 2.4. Commonly identified proteins and Venn diagram of protein ID from Raw 264.7 Macrophage cell lysate after Methanol-chloroform or improved Methanol-chloroform precipitation methods (normalized) followed by in-solution digestion. A) Shows commonly identified proteins among all biological replicates using LC-ESI-Thermo Lumos Mass Spectrometer after Methanol-chloroform purification and normalization. B) Shows upon

Methanol-chloroform purification, normalization of protein amount, and tryptic digestion in 0.2% deoxycholic acid in 50 mM Ammonium Bicarbonate identified more unique proteins compared to proteins digested in Ammonium bicarbonate alone or left-over pellet digested in the presence of 0.2% deoxycholic acid in 50 mM Ammonium Bicarbonate.

To compare these data sets, we generated a Venn diagram. (Figure 2.4B). From the Venn diagram, with at least two peptides identified per proteins, we can project, at least 594 proteins (428+166) are identified uniquely with the DCA method compared to the control, Methanol-Chl method. These are the proteins we did not identify in Methanol-Chl method. On the contrary, only 276 proteins are identified uniquely in Methanol-Chl method, which is not identified in the DCA method. This is 2.15 times (594/276) more identification of unique proteins in our modified method with DCA compared to the control with ABC. Also, in the left-over pellet digested with DCA method, and 241 (166+75) proteins were uniquely identified. From this data, we confirmed twice as much as higher unique protein identification in our modified method with MeOH-Chl-DCA. (Supplementary Table S2.3)

2.5.3 Comparative Analysis

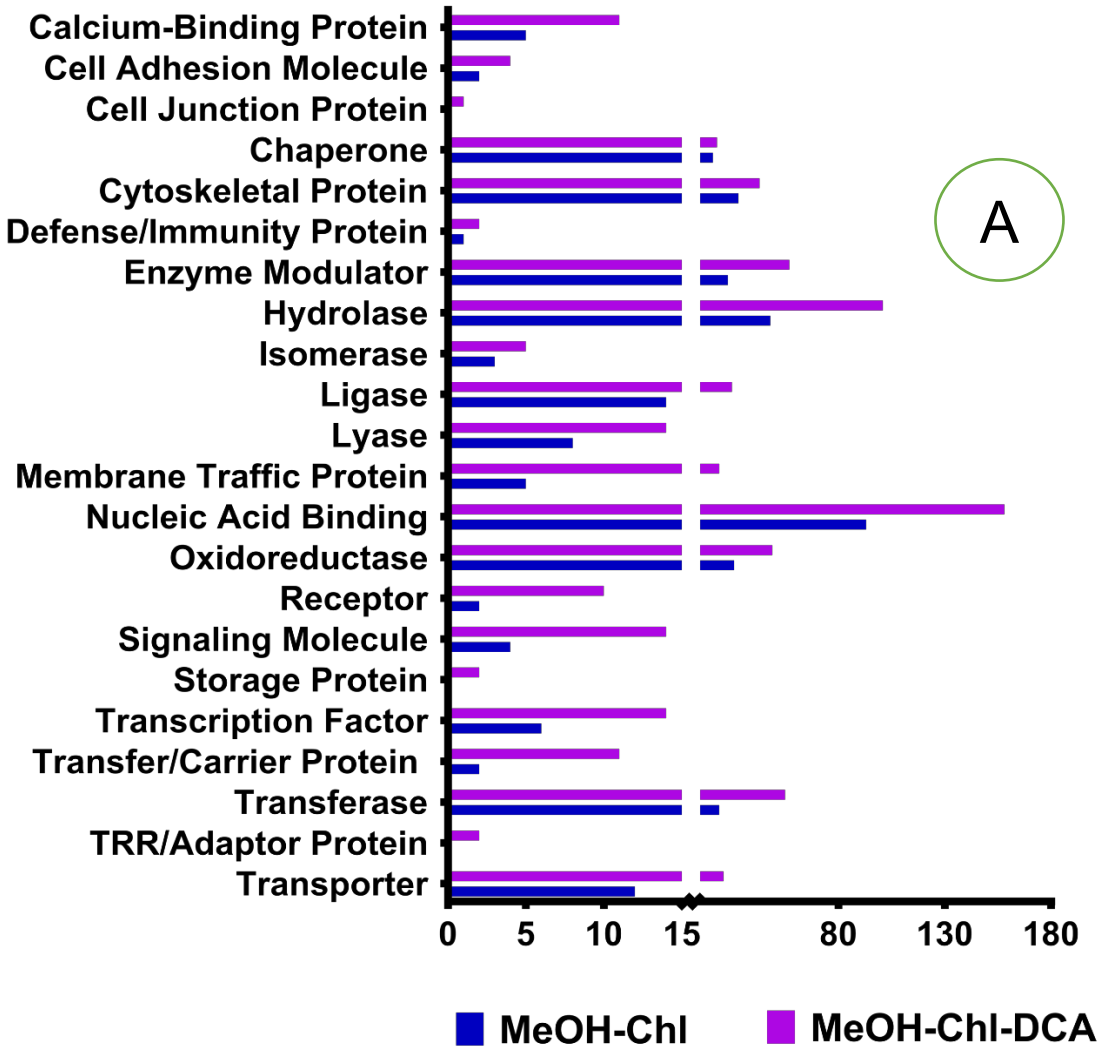
We summarize the identification of proteins among Methanol-Chl method and our modified Methanol-chloroform-DCA method. (Table S2.2) When we start with an equal amount of starting material (150 µg), we observed our modified method, yields 82 % (number of peptides identified per protein ≥ 2) increase in protein identification compared to the unmodified method. Since we confirmed through BCA Assay that incorporation of surfactants takes in three times more proteins into the solution and mass spectrometry

analysis shown 82% increase in protein ID; we decided to normalize protein amount in both samples after Methanol-chloroform precipitation. This way, we can see different functional categories of proteins uniquely identified in the samples due to our modified method. We found, there is an increase of 14% (number of peptides identified per protein ≥ 2) in protein identification in our modified method, and among these proteins, 27% are unique protein ID. The proteins are then classified according to their gene ontology functional classification to classify the overall category of protein identified due to the surfactant treatments.

2.5.4 Functional Classification

Proteins present in all biological replicates with at least 2 peptides identified per proteins were filtered. This filtered data set from control (Methanol-Chl) and modified method with DCA (Methanol-Chl-DCA) of both the experiment stages were used for generating gene ontology classification based on panther classification system into the molecular function, biological process, cellular component, and protein class. Gene ontology classification based on protein classes is shown in Figure 2.5A and 2.5B. In Figure 2.5A we show, in the protein class functional classification, with DCA we have higher identification in protein type in all subcategories, whereas in Figure 2.5B in the normalized sample, we see higher identification in most of the subcategories. Proteins that belong to nucleic acid binding, transfer/carrier protein, membrane and organelle proteins, membrane traffic protein, cytoskeletal protein, transporter, receptor protein class showed higher identification in both experiments. Other functional classifications' comparative diagrams are shown in

supporting Figures S2.2A and S2.2B. Protein Search Data file lists for all functional classes and their subcategories are not shown.



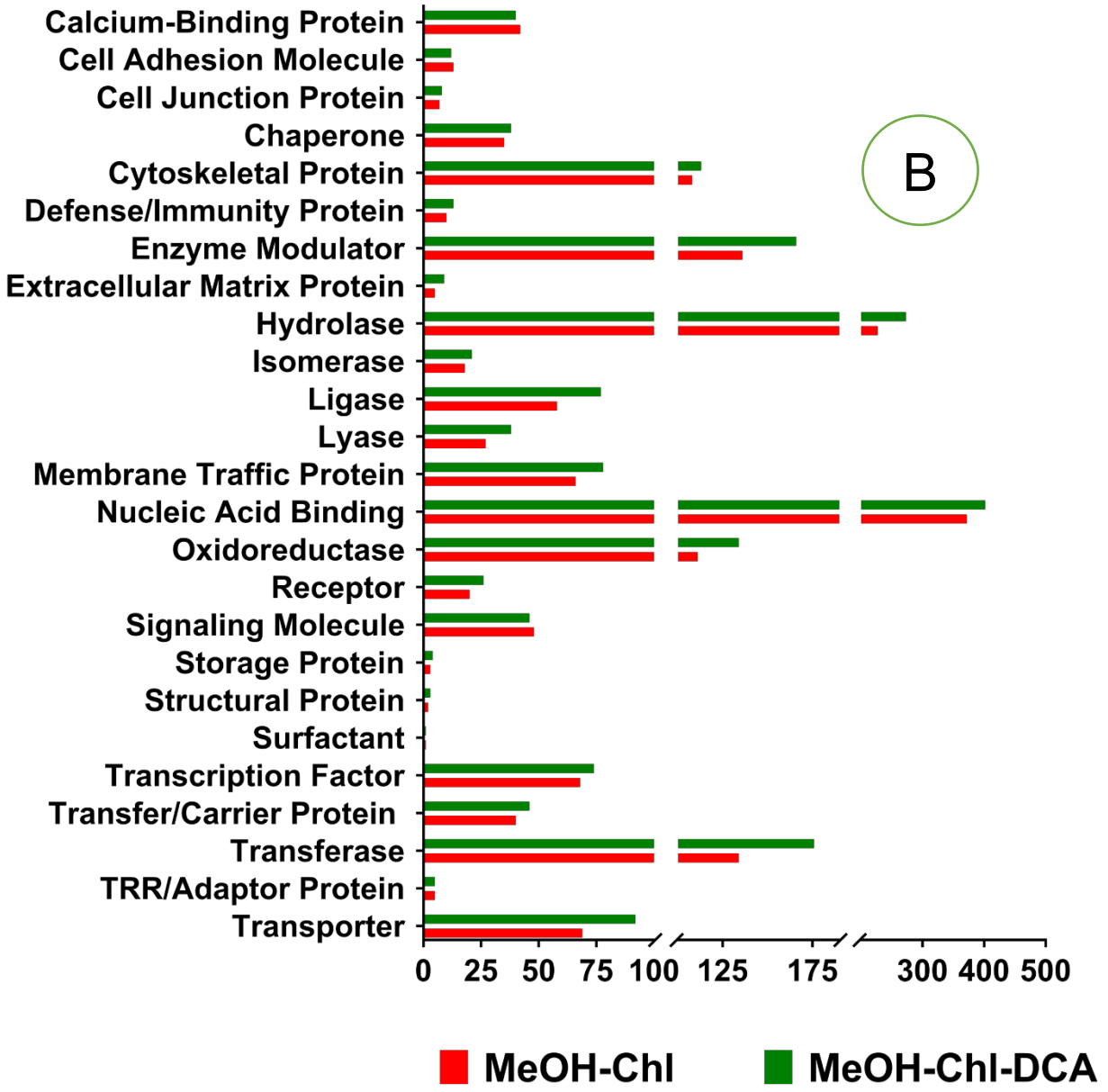


Figure 2.5. Gene ontology functional classification (Protein Class) by Panther Gene classification system, across control (MeOH-Chl) and improved method (MeOH-Chl-DCA). A) Shows MeOH-Chl-DCA method (un-normalized) identified a higher number of proteins compared to MeOH-Chl method in Gene ontology protein-class functional classification for each subcategory. B) Shows MeOH-Chl-DCA method (normalized)

identified a higher number of proteins compared to MeOH-ChI method in Gene ontology protein-class functional classification for most of the sub-classification.

2.6 Conclusion

Methanol-chloroform based protein precipitation is a widely used protein sample preparation methods. A significant amount of sample loss occurs during this purification processes. Here in this report, we showed the addition of surfactant such as Deoxycholic acid increased protein solubility and more protein ID in LC-MS/MS based protein identification methods. We designed two main experiments; in one, cell lysates with an equal amount of proteins after Methanol-chloroform precipitations, treated with or without DCA. In another experiment, an equal amount of cell lysates were taken, and after methanol chloroform precipitations, proteins were normalized and treated with or without DCA. In both cases, we observed a significant increase in protein ID as well as unique protein IDs in the modified method. Our comparison of different methods with cell lysates showed 82% more protein identification in the presence of Deoxycholic acid during digestion compared to Methanol-chloroform precipitate digested in Ammonium bicarbonate buffer. Normalization of protein amount after precipitations gave us a 14% increase in protein ID as well as 27% unique protein ID. The left-over pellet after digestion in control, upon dissolved in DCA; when running in mass spectrometry; showed a good number of proteins identified from there; further validating the fact that there is sample loss involved in the traditional method. We believe this method will help in cellular sample preparations, where the detergent is quite common for protein solubility. Another major advantage is that Deoxycholic acid can be precipitated from the sample before mass

spectrometric analysis. The objective of this experiment was to improve the Methanol-chloroform based sample preparation and protein identification in cellular proteomics sample. From all results being summarized, we conclude that the modified method will increase protein identification from commonly used Methanol-chloroform purification based cellular proteomics.

2.7 Supplementary Information

2.7.1 Methods

2.7.1.1 Protocol for Deoxycholic acid-mediated cellular proteome digestion method.

1. Culture RAW 264.7 cells.
2. Lysis cells using regular RIPA buffer.
3. BCA assay for protein quantification.
4. All Solutions: 10 mM DTT, 10 mM Iodoacetamide, Trypsin stock solution (20 mg/200 mL) were prepared in 50 mM NH_4HCO_3 pH 7.5.
5. 150 µg / 250 µg protein was taken for further work in a microcentrifuge tube.
6. Methanol chloroform purification and the precipitation were performed for further work.
7. The sample was air-dried for 5 minutes.
8. 2% Deoxycholic Acid (DCA) was prepared in 7N NH_4OH in water, and then NH_4HCO_3 was added so that the final solution will have 2% Deoxycholic acid and 50 mM NH_4HCO_3. Then pH adjusted to ~ 7.0 -8.0. (With pH paper) (conc. NH_4OH was used to adjust pH)
9. Solution vortexed for properly dissolving the solid DCA.
10. Separately 50 mM NH_4HCO_3 was prepared.

11. 100 µL of 2% Deoxycholic Acid (DCA) in 50 mM NH_4HCO_3 was added to the sample pellet, and then the sample was rested at room temperature / 25º C for 30 minutes.
12. The pellet was rinsed for 5 minutes with the flow of the 100 µL tip to dissolve it, rinse volume was set to 50 µL.
13. 300 µL NH_4HCO_3 added and rinsed similarly for 10 minutes with the flow of the 100 µL tip to dissolve it, rinse volume was set to 100 µL.
14. 5 µL, 10 mM DTT (added to the solution).
15. Thermomix (56 ^{o} C, 45 minutes)
16. 25 µL 10 mM Iodoacetamide (added) / 1:5 ratio with DTT.
17. 30 minutes (incubated in dark) 37 C at 600 rpm.
18. Trypsin added so that Trypsin: Protein = 1:100 (20 mg/200 mL stock solution =15 mL transferred.)
19. 50 mM NH_4HCO_3 (pH 7.5) was added to dilute the solution up to 1mL.
20. 37° C, 500 rpm, 16-hour incubation.
21. 100 µL 20 % Formic acid (added).
22. Centrifuge 20,000 rpm, 4 ^{0}C, 30 minutes.
23. Take supernatant.
24. 100 µL 20 % Formic acid (added). A cloudy white solid DCA precipitate will be formed.

25. Centrifuge 20,000 rpm, 4 ^{0}C, 30 minutes.
26. Take supernatant.
27. Speed Vac to dry the sample.
28. 300 µL 0.1% Formic acid (added) → Mix properly → Proceed to desalt with C18 Column.
29. Zip tip needs to be done twice, with two C18 Column per sample and in last step addition of 80%, Acetonitrile to release the bound peptide was done thrice for each C18 column used.
30. Reconstitute in 20 µL 0.1% formic acid.
31. Tap the microcentrifuge tube multiple time to ensure peptides attached to the wall are brought to the solution.
32. Spin down the sample for 10 seconds.
33. Use 2 µL sample to measure the amount of peptide present in the sample using UV-Vis in nanodrop.
34. Load remaining 18 µL for LC-MS/MS such that loading volume has 1ug peptide delivered to the mass spectrometer.

2.7.1.2 Acetone Purification and digestion in Ammonium Bicarbonate

The extracted proteins (150 µg), three replicates (Tag: Acetone) were purified using Acetone purification method as described. Briefly, Proteins are diluted to 1µg/µL. Briefly, Acetone and Acetone/ water (4:1) solution were prepared and stored in -20^{o} C. Four

volume of -20^{o} C Acetone was added to the sample. Samples were centrifuged for 15 minutes at 13000 rpm; then the top layer was decanted. 400 µL of cold acetone was used for additional washing, and without any mixing, immediate centrifugation at 13000 rpm is performed, and the top acetone layer is removed. The proteins in the pellets are taken for bottom-up proteomics. The proteins were then reduced and alkylated, then digested with trypsin (MS Grade) at a 1:100 enzyme/protein concentration for 16 h at 37 °C. Formic acid was added afterward to drop the pH and to stop trypsin activity. The samples were then desalted using a C18 desalting column (Thermo Scientific, IL, USA). After completely drying by speed vacuum, peptides were dissolved in 0.1% formic acid, and stored at -20°C before LC-MS/MS experiments.

2.7.1.3 Detergent removal Spin Column Filtration (DRSC) and digestion in Ammonium bicarbonate

The extracted proteins (150 µg), in three replicates (Tag: DRSC), were filtered through Thermo Scientific Pierce detergent removal spin column according to the manufacturer's protocol. Briefly, the spin columns were equilibrated with equilibration buffer. Then, the sample is equilibrated in the filter for 2 minutes and centrifuged at 1500g to get a detergent free sample. The proteins were then reduced and alkylated, then digested with trypsin (MS Grade) at a 1:100 enzyme/protein concentration for 16h at 37 °C. Formic acid was added afterward to drop the pH and to stop trypsin activity. The samples were then desalted using a C18 desalting column (Thermo Scientific, IL, USA). After completely drying by speed vacuum, peptides were dissolved in 0.1% formic acid, and stored at -20 °C before LC-MS/MS.

2.7.1.4 DCA Assisted digestion of discarded protein pellets after Methanol-chloroform purification and digestion in Ammonium Bicarbonate

250 µg of Raw 264.7 macrophage cells, in nine aliquots, were taken and then purified using Methanol-Chloroform method. They were then suspended in 300 µL, 50 mM NH_4HCO_3. From these nine, we prepared three replicates by combining three aliquots in one microcentrifuge tube. Proteins were normalized, then reduced, alkylated, digested, centrifuged, and the pellets are isolated, and 100 µL of 2% Deoxycholic Acid (DCA) in 50 mM NH_4HCO_3 was added to the sample. The sample was rested at 25º C for 30 minutes. (Tag: Digested-Pellet-DCA) The samples were then dissolved in 300 µL of 50 mM NH_4HCO_3. After reduction and alkylation, the final volume is increased to 1 mL with 50 mM NH_4HCO_3. The proteins were then digested with trypsin (MS Grade) at a 1:100 enzyme/protein concentration for 16 h at 37 °C digestion. Formic acid was added afterward to drop the pH, to stop trypsin activity, and to precipitate the DCA. Next, centrifugation at 20000 rpm for 30 minutes at 4ºC removes the surfactant. The supernatant samples were then collected, desalted using a C18 desalting column (Thermo Scientific, IL, USA). After completely drying by speed vacuum, peptides were dissolved in 0.1% formic acid, and stored at -20 °C before LC-MS/MS.

2.7.1.5 Mass analysis

2.7.1.5.1 Nano-LC-MS/MS using Thermos Velos Pro

Digested peptides were analyzed by nano-LC-MS/MS using a Velos Pro Dual-Pressure Linear Ion Trap Mass Spectrometer (Thermo Fisher Scientific, MA) coupled to a UHPLC (UltiMate 3000, Thermo Fisher Scientific, MA). Peptides were loaded onto the analytical

column and separated by reversed-phase chromatography using a 15-cm column (Acclaim PepMap RSLC) with an inner diameter of 75 µm, packed with 2 µm C18 particles (Thermo Fisher Scientific, MA). Nano column was eluted with multi-step gradient of 4-90% solvent B (A: 0.1% formic acid in 18 Mohm Milli-Q water; B: 95% acetonitrile and 0.1% formic acid in 18 Mohm Milli-Q water) over 70 min with a flow rate of 300 nL/min with a total run time of 90 min. The mass spectrometer was operated in positive ionization mode with the nano-spray voltage set at 2.50 kV and source temperature at 275 ºC. The instrument was operated in a data-dependent mode in which the three precursor ions with the most intense signal in a full MS scan were consecutively isolated and fragmented to acquire their corresponding MS2 scans. Full MS scans with 1 micro scan (µs) were at a resolution of 3,000, and a mass range of m/z 350-1500. Normalized collision energy (NCE) was used at 35%. Fragment ion spectra produced via high-energy collision-induced dissociation (CID) was acquired in the Linear Ion Trap mass analyzer with the resolution of 0.05 FWHM (full-width half-maximum) with Ultra ZoomScan between m/z 50-2000. A maximum injection volume of 5 µL was used during data acquisition with partial injection mode. The mass spectrometer was controlled in a data-dependent mode that toggled automatically between MS and MS/MS acquisition. MS/MS data acquisition and processing were performed by Xcalibur™ software (Thermo Fisher Scientific, MA).

2.7.1.5.2 LC-MS/MS data collection by Orbitrap Lumos

Samples were analyzed by reverse-phase LC-ESI-MS/MS system using the Dionex UltiMate 3000 RSLC nano System coupled to the Orbitrap Fusion Lumos Mass Spectrometer (Thermo Fisher Scientific). Peptides were loaded onto a trap column (300 µm ID x 5 mm) packed with 5 µm 100 Å PepMap C18 medium, and then separated on a

reverse-phase column (50-cm long × 75 μm ID) packed with 2 μm 100 Å PepMap C18 silica (Thermo Fisher Scientific) using a 120 min gradient. The column temperature was maintained at 50 °C. Mobile phase solvent A was 0.1% formic acid (FA) in water, and solvent B was 0.1% FA in 80% acetonitrile. Peptides were loaded to the trap column in 100% buffer A for 5 min at 5 μl/min flow rate, and eluted from the analytical column at a flow rate of 300 nl/min with a linear 80 min gradient of 5-30% of buffer B, then changing to 45% of B at 91 min, 100% of B at 93 min at which point the gradient was held for 7 min before reverting to 95% of A at 100 min. The mass spectrometer was operated using standard data-dependent mode. The full scan MS spectra were collected at a resolution of 120,000 at 200 m/z with AGC target of $4.0e^5$, a mass tolerance of 10 ppm, and Advanced Peak Detection (APD) activated. Fragmentation of precursor ions was performed by high-energy C-trap dissociation (HCD) in Orbitrap with the normalized collision energy of 30% and a resolution of 15,000 at m/z 200. The dynamic exclusion was set at the 30s to avoid repeated scanning of identical peptides. All the MS measurements were performed in the positive ion mode. The instrument was calibrated at the start of the experiment, and the performance of the instrument and data repeatability was evaluated by using Hela cell digest before running the samples.

2.7.2 Supplementary Figures

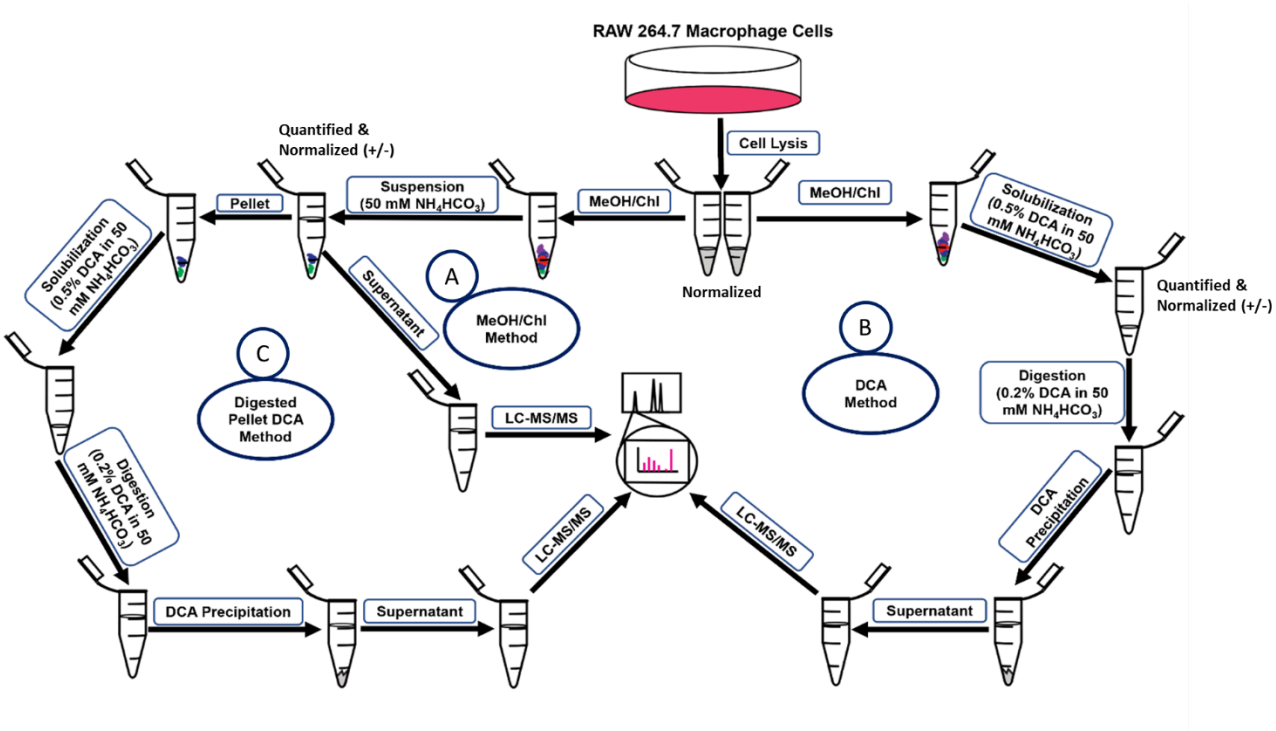


Figure S2.1. Steps of sample preparations are shown in the schemes. A) Shows Methanol-chloroform purification followed by digestion in 50 mM Ammonium Bicarbonate. B) Shows Methanol-Chloroform-DCA method where Methanol/Chloroform precipitated pellet was dissolved in Deoxycholic acid and tryptic digested. C) Shows the left-over pellet after tryptic digestion from A were used for dissolving in Deoxycholic acid and digestion.

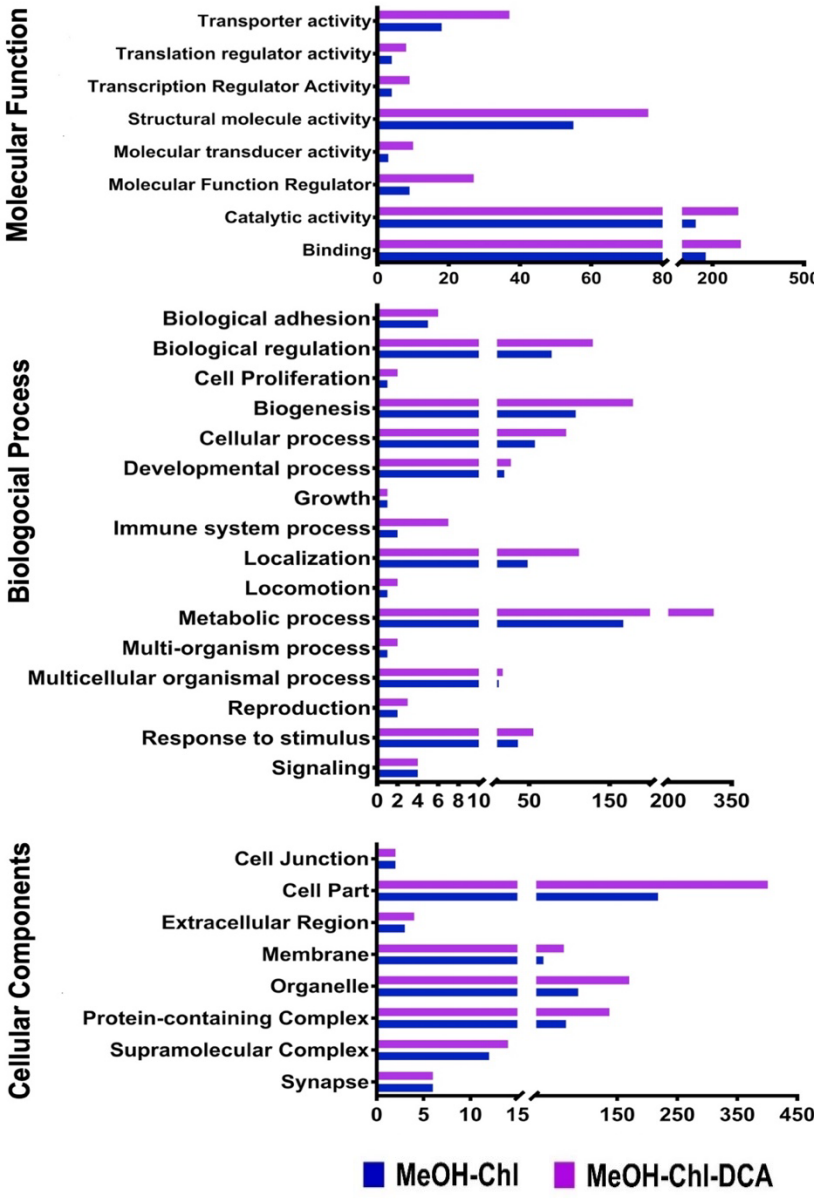


Figure S2.2A. Panther gene classification system based functional classification of identified proteins across control (MeOH-Chl) and optimized method (MeOH-Chl-DCA).

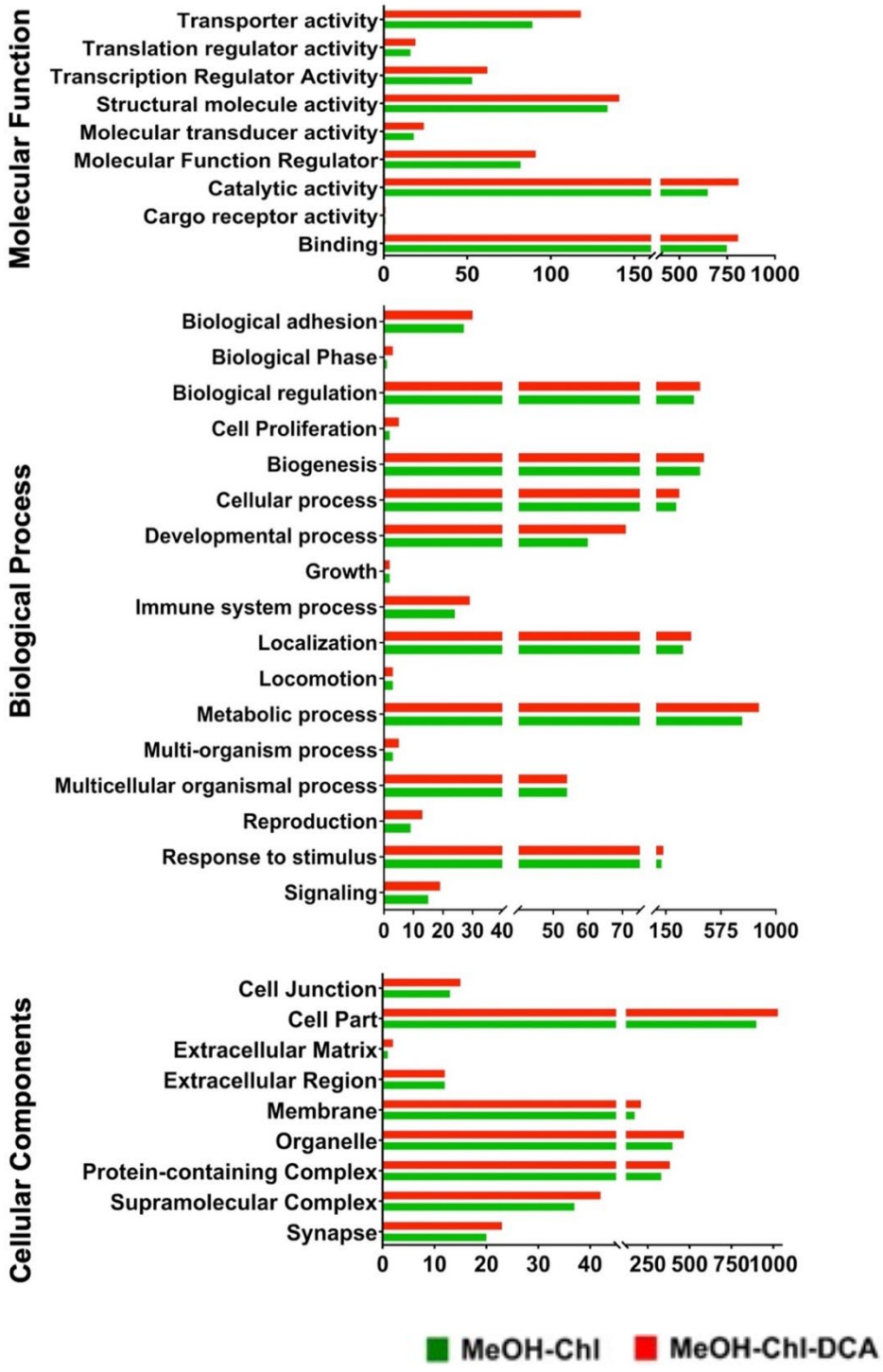


Figure S2.2B. Panther gene classification system based functional classification of identified proteins across control (MeOH-Chl) and optimized method (MeOH-Chl-DCA).

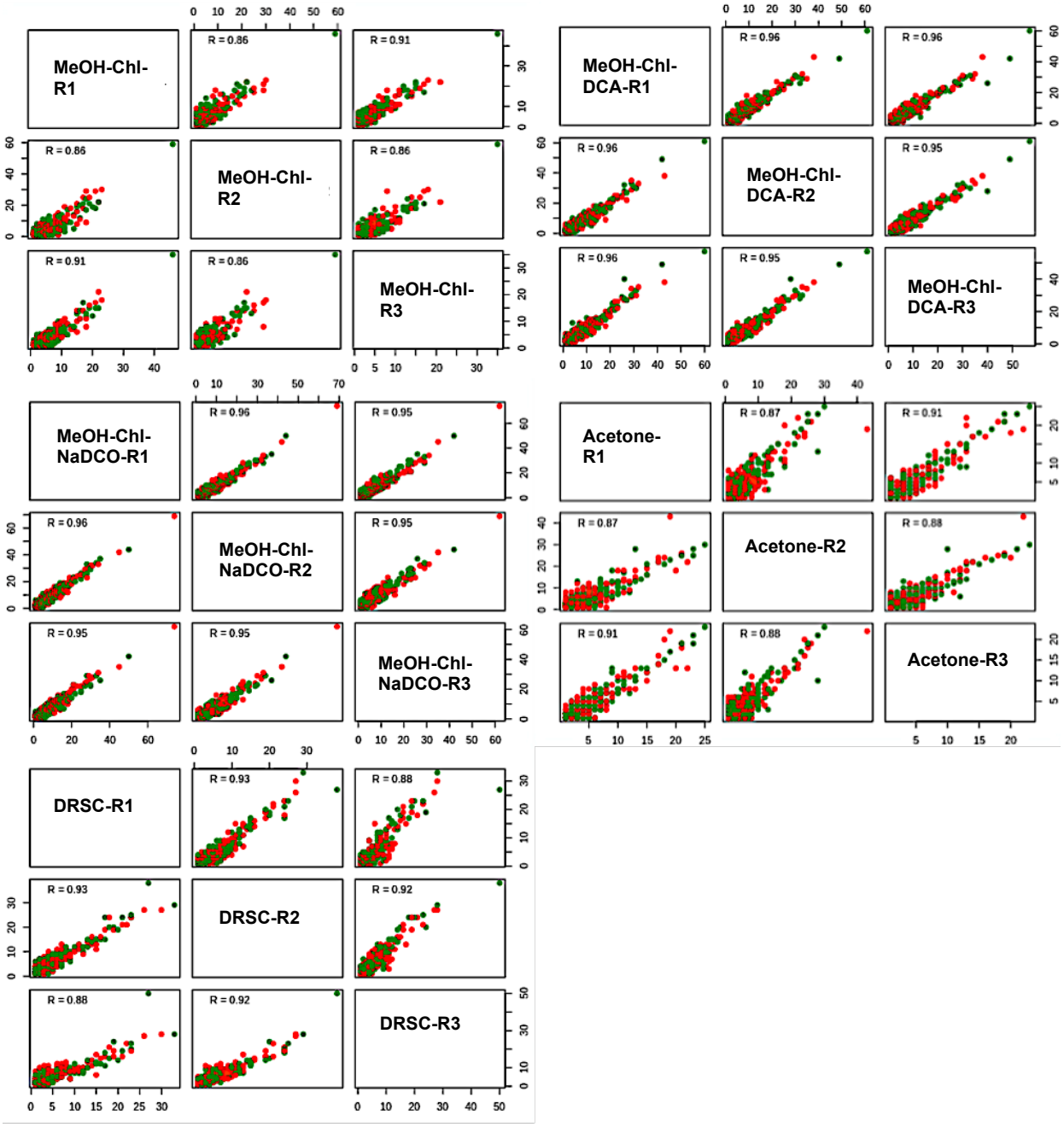


Figure S2.3A. Scatter plot and pairwise correlations showed a significant correlation among all the biological replicates in specific sample preparation procedures. All samples were prepared through an equal amount of RAW 264.7 cell lysate proteins purified through different methods and ran in LC-ESI-Thermo Velos pro mass spectrometer after in-solution digestion. The data corresponding to the number of peptides identified per

identified proteins was plotted against each other among the biological replicates, on x-axis and y-axis. Every spot symbolizes the abundance of protein and corresponds to Pearson's correlation coefficient (R) to 1. The scatter plot and pairwise correlation were generated by R package ver 3.5.2.

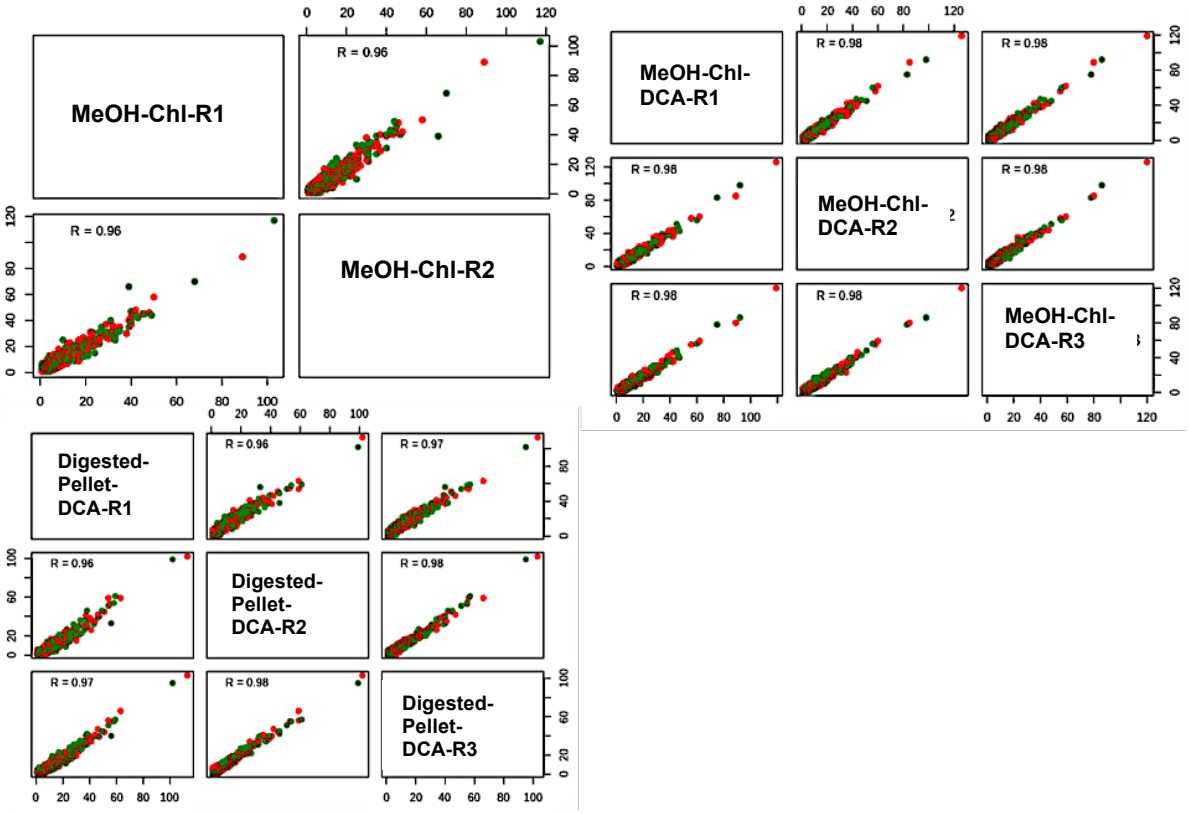


Figure S2.3B. Scatter plot and pairwise correlations show a significant correlation among all the biological replicates in specific sample preparation procedures. Samples were prepared through 1) RAW 264.7 cell lysate proteins purified through Methanol-chloroform purification followed by dissolved in DCA (normalized) and bottom-up proteomics and 2) RAW 264.7 cell lysate proteins purified through Methanol-chloroform purification followed by suspended in ammonium bicarbonate (normalized) and bottom-up proteomics and 3) Leftover pellet after tryptic digestion from 2 were collected and dissolved in DCA followed by bottom-up proteomics. All samples were analyzed in LC-ESI-Thermo Orbitrap Fusion

Lumos Tribrid Mass Spectrometer after in-solution digestion. The data corresponding to the number of peptides identified per identified proteins was plotted against each other among the biological replicates, on x-axis and y-axis. Every spot symbolizes the abundance of protein and corresponds to Pearson's correlation coefficient (R) to 1. The scatter plot and pairwise correlation were generated by R package ver 3.5.2.

2.7.3 Supplementary Tables

Dissolving Solution	Initial amount 150 µg	Initial amount 250 µg
A) 400 µL 50 mM NH_4HCO_3	A) 35 µg quantified	A) 68 µg quantified
B) 400 µL 0.5 % DCA in 50 mM NH_4HCO_3	B) 91.7 quantified	B) 162.5 µg quantified
C) 400 µL 0.5% Na-DCO in 50 mM NH_4HCO_3	C) 96.7 µg quantified	C) 168.0 µg quantified

Table S2.1: Amount of proteins quantified in different amount cell lysate using BCA method using with or without deoxycholic treatment.

			LTQ-Velos Pro	
Samples	**Peptides ≥ 1**	**% increase**	**Peptides ≥ 2**	**% increase**
MeOH-Chl-DCA	1360	84%	967	82%
MeOH-Chl	738		529	

			Thermo Lumos	
Samples	**Peptides ≥ 1**	**% increase**	**Peptides ≥ 2**	**% increase**
MeOH-Chl-DCA	3217	12%	2560	14%
MeOH-Chl	2872		2242	

Table S2.2: Comparative identification of proteins with or without deoxycholic Acid. Among two experiment conditions in Methanol-chloroform purification followed by tryptic digestion in 50 mM Ammonium Bicarbonate and tryptic digestion in 0.2% Deoxycholic Acid in 50 mM Ammonium Bicarbonate shows higher identification of total proteins in both cases with deoxycholic Acid.

Increase in total protein identification with number of Identified peptides ≥ 2 per protein id				
Samples	**Unnormalized Set**	**% increase**	**Normalized Set**	**% increase**
MeOH-Chl-DCA	**967**	**82%**	**2560**	**14%**
MeOH-Chl	**529**		**2242**	

Increase in unique protein identification with number of identified peptides ≥ 2 per protein id				
Samples	**Unnormalized Set**	**Times increase**	**Normalized Set**	**Times increase**
MeOH-Chl-DCA	**450**	**37.5**	**594**	**2.15**
MeOH-Chl	**12**		**276**	

Table S2.3: Comparative portray of total identified proteins and uniquely identified proteins. Incorporation of deoxycholic acid brings higher identification of both total and unique proteins in both normalized and unnormalized experiment set.

Acknowledgment: The study was partially funded from the grant from UA5GM113216-01, NIGMS, NIH. We acknowledge start-up funding from the UT-Arlington to Dr. Chowdhury. We recognize the Shimadzu Institute for Advanced Analytical Chemistry at the University of Texas at Arlington for mass spectrometry support. The authors acknowledge support from the UT systems proteomics networks for an LTQ Velos Pro mass spectrometer. Purdue Proteomics Facility, Bindley Bioscience Center, Purdue University was utilized for collection of high-resolution LC-MS/MS data.

Chapter-3

Identification of Toll-like receptor 4 interacting partners through immunoprecipitation-based chemical cross-linking proteomics

3.1 Abstract

Toll-like receptor 4 (TLR-4) are receptors on immune cell that can recognize invasion of bacteria through their attachment with bacterial lipopolysaccharides. Statin, on the other hand, are lipid-lowering drugs and can also dampen immune cell response. Here we wanted to study how the effect of pro and anti-inflammatory stimulus function in immune signaling. We selected human embryonic kidney (HEK 293) cells that are engineered to express HA-tagged TLR-4. Upon treatment with Statin or LPS or Statin+LPS, cell lysates were further used for co-immunoprecipitation using monoclonal antibody directed against HA-tagged TLR-4. We performed this study, both in the presence of a chemical crosslinker (ETD cleavable) or in the absence of the crosslinker. The goal is to understand the TLR-4 mediated signaling and trap the transiently interacting partners using chemical crosslinkers. All samples were tryptic digested with trypsin under regular in solution digestion conditions, and the peptides were analyzed through liquid chromatography-tandem mass spectrometry (LC-MS/MS). We used label-free quantitation to understand comparative expression among treatments and controls and investigated stable and transient interactions. Using the same proteomics data analyzed through IPA, A total of 712 differentially expressed proteins were identified and quantified in this study. From these, we selected macrophage myristoylated alanine-rich C kinase substrate and

creatine kinase for biochemical validation. Western blot analysis confirmed, myristoylated Alanine Rich Protein Kinase C Substrate (MARCKS) is a candidate in Statin-induced anti-inflammatory response mechanism with possible links to TLR4 networks.

3.2 Introduction

Our immune system works in a complex manner where various immune and effector cells work in coherence to deter invading pathogens. Macrophages are immune effector cells that perform varieties functions, e.g., innate immune response, inflammatory response and wound healing, etc. [77] Activation of macrophages is a complicated process and can happen in different ways. In multiple ways, different endogenous and exogenous stimulus and immune effector cells work to activate macrophages for their function. Activation through Toll-like receptor/(s) (TLRs) is one of them. [77-78] Toll-like receptors are type I transmembrane glycoproteins. The extracellular domain, rich with leucine repeats can recognize, foreign materials known as pathogen-associated molecular patterns (PAMPs) or endogenous ligands, i.e., damage-associated molecular patterns (DAMPs) produced through tissue damage, necrosis, and infection. [28, 79] Engagement of ligands to TLR or TLR complex with accessory proteins generate signals, which are then transferred, through transmembrane domain and cytoplasmic toll-interleukin 1 Receptor (TIR) domain to downstream adaptor proteins. Through this signaling mechanism, TLRs leads to activation of transcription factors, e.g. NF$-\kappa$B, IRFs (interferon regulatory factors), etc. These induce the expression of type 1 interferon (IFN α, β), different proinflammatory cytokines, e.g. TNF$-\alpha$, IL-1, IL-6, IL-12, nitric oxide, reactive oxygen species, etc. [80] Secretion of IFN γ, IL-4, IL-10, etc. from other effector cells, e.g., NK cells, Th cells,

regulatory T cells, etc. can also lead to activation of tissue macrophages for different immune response. These all responses include, microbicidal activity and tissue repair and wound healing, inflammation, immune suppression, etc. based on the types and sites of the infection and involvements of immune cells and their effectors. [19, 77-78, 81-82] Apart from innate immune response, Toll-like receptor response in adaptive immunity also plays a role through Type 1 interferon production and proliferation of memory T cells. [83] In addition to the innate and adaptive immune response through Toll-like receptors (TLRs) signaling; inappropriate activation of TLRs can lead to autoimmune disease. [84] Studying the toll-like receptor activation signaling network is hence essential to find new therapeutic approach against viral or bacterial infection or autoimmune disorder.

In human, so far, 10 functional Toll-like receptors have been identified. Apart from endosomal compartment localized TLRs 3, 7, 8, 9; rest of the six TLRs are expressed as membrane-spanning proteins on different types of immune cell surfaces. Each TLR recognizes specific types of ligands from invading microorganisms or specific types secretion after tissue damage. [79, 85-87] Among all these TLRs, TLR2 and TLR4 has gained most of the attention due to their ability to recognize a diverse array of pathogenic ligands. Toll-like receptor 4, in association with host accessory proteins MD-2 and CD-13, can recognize the lipopolysaccharide (LPS) from the outer membrane of gram-negative bacteria.[88-89] TLR4 can also recognize various endogenous ligands, e.g., fibronectin, fibrinogen, heparan sulfate, hyaluronic acid, HSP60, HSP70, and other ligands from the secretion of tissue damage. Upon activation, TLR4 can exert functions through two distinctive pathways, e.g., MyD88 dependent pathway and TRIF-dependent pathway.

Hence, immune signaling through TLR4 is diverse and complicated. [87, 90] Thus, we were interested in studying TLR4 interacting immune signaling networks comprehensively.

LPSs are an essential component found on the outer membranes of gram-negative bacteria. The composition of LPS varies among different bacteria but essentially consists of a polysaccharide core chain attached to amphipathic lipid, and a variable fatty acid chain. During infection, it can get detached from the membrane and transferred to the TLR4 complex with MD-2 via LPS binding protein and CD14. Once bound to TLR4-MD-2 complex, TLR-4 mediated signaling is activated with the overwhelming secretion of cytokines followed by immune-inflammatory or host defense response from the immune cells. [88, 91] A few quantitative proteomics studies have been reported where LPS treatment on the immune cell was analyzed to study differentiation and activation of monocytes in pro-inflammatory states. A label-free proteomics study on LPS treated human monocyte suggested human monocyte (THP-1 cell line) suggested a potential therapeutic application of monocyte in tumor treatment. [92] Raw macrophage cells treated with LPS were analyzed in 2D gel electrophoresis combined with mass spectrometry. A total of 11 differentially expressed proteins were successfully identified in gel spots and mass spectrometry. [93] Stable isotope labeling with amino acid (SILAC) based quantitative proteomics studies were performed on nuclear and cytosolic fractions of LPS stimulated macrophages. A 10-minute LPS exposure had led to modulation of several mitogen-activated protein kinases (MAPK) and NF-kB signaling pathways on LPS stimulated macrophage. This study provided the first system-wide insight of cross-talk between signaling pathways, transcription factors, and activated pro-inflammatory genes. [94] There

are reports of quantitative proteomics experiments on isolated lipid rafts from LPS treated RAW 264.7 macrophage cell line. [95-97] Proteomics study on the isolated raft of ABCA1-deficient primary mouse macrophage cells upon LPS-stimulation identified 383 unique proteins. [97] In LPS and interferon-γ (IFNγ) activated macrophage, 409 associated microtubule proteins were identified. [98] Proteomics analysis of Salmonella-infected macrophage cells identified 244 significantly altered proteins in a time-dependent manner. [99] Isotope Coded Affinity Tagging (ICAT) profiling of Raw 264.7 macrophage cells identified 36 differentially expressed proteins upon LPS stimulation. [100]

Statins are inhibitors of 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) reductase. Hyperlipidemic patients are treated with statins to reduce serum cholesterol. Apart from this, statins have been reported with additional immunomodulatory activities. For example, human monocyte-derived dendritic cells' maturation is suppressed with statins. [101] In different immunological disorders, such as autoimmune encephalomyelitis, arterial graft disease (GAD) and sepsis, statins have been shown to ameliorate inflammation.[102-105] Simvastatin, a variant of Statin was reported to suppress LPS induced phosphorylation of Akt protein in monocytes. [106] Statins also downregulate Toll-like receptor 4 and NF-κB phosphorylation leading to control of DNA transcription and reduction in lymphocyte cytokines' production. Thus, statins play an effector role in the amelioration of inflammation. [107-111] Hence statins are a choice of a modulator to study anti-inflammatory response.

There have not been however, many studies on proteomics analysis in statin stimulated model system. Study on low-density lipoprotein receptor-negative mouse, receiving cholesterol-rich diet and subsequent treatment with statins identified proteomics network in macrophage as a model for atherosclerosis development. [112] Healthy individual treated with rosuvastatin followed by label-free spectral counting approach, identified significant changes in lipoprotein pools along with a marked expression of alpha 1-antitrypsin in HDL-L fractions. [112] Lipoprotein proteomics performed on a human blood sample to understand the effects of statin (rosuvastatin) in the immune system identified 154 proteins. [113]

As summarized here, there have been several proteomics studies on a pro-inflammatory model system with LPS being used a pro-inflammatory stimulant in Raw 264.7 macrophage cell line and very few studies on an anti-inflammatory model system with a statin as a choice of the effector. But, no other group have performed a combined experiment with an anti-inflammatory stimulant, statin, and pro-inflammatory stimulant, LPS been used in tandem to see the transition in the proteomic network and better understand the model system. However, our group did perform a combined experiment before, which is the only reported one of its kind, where we had studied anti-inflammatory stimulant, statin, and pro-inflammatory stimulant, LPS used alone and in tandem to generate proteomics network of an immune infection model in macrophages. However, this study did not use any targeted network analysis; rather, the whole work was limited to global profiling approach. Additionally, our group's previous work did not take into

account of capturing the protein-protein interaction networks that were transiently interacting with one another. [114]

Here we wanted to make more of a targeted approach to study toll-like receptor 4 interactome network. Identifying protein interacting partners in complex biological samples is a challenging task. There are several methods available to undertake such challenges. [115-117] In a native natural system, co-immunoprecipitation (IP)-based identification of protein interaction networks has become a gold standard for studying protein interactome. [115, 118] In (IP)-based proteomics, a target protein and it's interacting partners are purified as a complex through affinity or antibody-based purification. However, optimization of wash conditions through both removal of nonspecific interactions and preserving the transient and weak interactions at the same time, in the co-IP method makes it amenable to study only stable interactions. Combination of cross-linking proteomics to attach proximal protein binding partners in co-IP covalently have recently been employed to improve the efficiency of the methods. [119-120] Since, cross-linking theoretically captures transient and weekly interacting proteins, incorporation of cross-linking allows use of strong detergent washing step to remove non-specific interactions. Use of cross-linkers also gives a two-factor advantage of data analysis, both through the identification of individual peptides from cross-linked proteins and through the fragmentations of cross-linked peptides. [119, 121] However, due data complexity and lack of efficient software platform, in most cases for complex cellular sample, the analysis may only be limited to identification of individual peptides through bottom-up proteomics. Here we used this approach to design and analyze our experiment and datasets.

We chose a model system, HEK293 stable cell line incorporating atypically expressing HA-tagged toll-like receptor 4 (TLR-4). Activation of immune signaling through TLR-4 and antibody-based purification of TLR-4 complex along with stable and weak interacting partners will allow us to do a targeted study on TLR-4 mediated signaling mechanism in the immune response.

In our experiment, cells were treated with statin or LPS in alone or in combination in tandem, with or without cross-linkers. This would allow us to study expression in all possible conditions: e.g. pro-inflammatory status in presence of LPS alone, anti-inflammatory response in presence of statin alone, transition from anti-inflammatory towards pro-inflammatory condition in stain-LPS treatment made in tandem, stable interaction of proteins in all above conditions followed by purification through co-IP and finally weakly interacting partners along with stable interacting partners in presence of our in-house chemical cross-linker treatment followed by purification through co-IP. The antibody is specific against the HA-tagged extension of toll-like receptor 4. So, through this experiment design, antibody-based purification will allow us to study protein interactomics in the transition of the immune signaling response mechanism from anti-inflammatory towards pro-inflammatory status.

From preliminary proteomics data, two proteins, e.g., macrophage myristoylated alanine-rich C kinase substrate/ MARCKS related protein (MacMARCKS/MRP/MARCKSL1), Creatine kinase were selected as tentative interactors in the immune signaling pathway. However, when biochemical validation was combined with discovery-based proteomics, we were able to confirm, MacMARCKS alone is a statin-responsive protein that serves to

modulate TLR-4 mediated signaling. We hope our study will help bring a translational implication for the overwhelming number of statin users across the globe someday.

3.3 Experimental procedures

3.3.1 HA-TLR4-HEK-293 cell line:

HA-tagged human TLR4 gene transfected stable HEK-293 cell line was purchased from Invivogen (Catalog # 293-htlr4ha). We grew and maintained the cell line in DMEM containing 10% FBS, 100 U/mL penicillin with the addition of 100µg/mL streptomycin, supplemented with Blasticidin 10 µg/mL and Hygromycin 50 µg/mL to maintain selective pressure on transfected DNA.

3.3.2 Cell culture and protein preparation

We maintained the Hemagglutinin (HA) tagged-TLR4-human embryonic kidney (HEK)293 cells in DMEM medium supplemented with 10% fetal bovine serum, 1% penicillin/streptomycin and antibiotics (50 µg/ml hygromycin and 10 µg/ml blasticidin) in a humidified atmosphere of 5% CO_2 at 37^{o} C. Eight experiment sets were prepared with three replicates in each of them.

We treated the cells with 10 µM simvastatin (Sigma) for 24 hours, then stimulated with 1 µg/ml Lipopolysaccharides (LPS-EB, InvivoGen) for 1 hour in the freshly supplied medium. Next, we treated the cells with our in-house ETD crosslinker (XL) added at a final concentration of 1µmol/ml for 30 min, followed by quenching the reaction with 50 mM Tris-HCl, pH 8.0. Similarly, we treated cells with simvastatin for 24 hour or

Lipopolysaccharides for 1 hour, followed by treatment with ETD crosslinker. Also, we prepared control cell lines with or without treatment of ETD crosslinker.

In the next step, we did IP-pull down for proteomics. We lysed the cells with immunoprecipitation (IP)- lysis buffer supplemented with protease inhibitors at 4°C for 15 mins followed by sonication for another 15 mins and incubation at 4°C for another 30 mins. Finally, we centrifuged the cells at 20,000×g, 4°C for 30 mins and collected the supernatant for measuring the protein concentration with a BCA protein assay kit, using bovine serum albumin as a standard.

3.3.3 Separating the TLR4-interacting partners using immunoprecipitation

We washed anti-HA magnetic beads (Thermo Scientific, MA) with a gentle vortex in 0.05% TBS-T buffer. We collected the suspended magnetic beads using a magnetic stand for 5 minutes at room temperature (RT). We mixed the pre-washed magnetic beads with HA-tagged cell lysate protein sample and kept for gentle rotation at 4°C overnight in a rotating shaker. The next day, we collected the beads with a magnetic stand and washed with TBS-T buffer and ultrapure water twice, followed by elution in a Laemelli buffer (95^{o} C for 5 minutes). After centrifugation, we purified the samples through Methanol-chloroform purification followed by in-solution tryptic digestion and mass spectrometric analysis and label-free spectral counting based on quantification of matched peptides. [114]

3.3.4 In-solution digestion and mass analysis (nano-LC-MS/MS)

We took the immunoprecipitated proteins and reduced and alkylated them followed by trypsin (porcine) (MS Grade) at 37°C for overnight. [114] Next, we added formic acid to the

peptides to bring pH < 3 and stop trypsin activity. Finally, we dried the samples by speed vacuum and then dissolved in 0.1% formic acid solution followed by centrifugation at 20,000 × g for 30 min at 4°C.

We analyzed the digested peptides by nano-LC-MS/MS using a Velos Pro Dual-Pressure Linear Ion Trap Mass Spectrometer (ThermoFisher Scientific, MA) coupled to an Ultimate 3000 UHPLC (ThermoFisher Scientific, MA). We loaded the peptides into an analytical column and separated them by reverse phase chromatography using a 15 cm column (Acclaim PepMap RSLC) with an inner diameter of 75 µm and packed with 2 µm C_{18} particles (Thermo Fisher Scientific, MA). We used multi-step gradients of 4-90% solvent B (A: 0.1% formic acid in water; B: 95% acetonitrile and 0.1% formic acid in water) over 70 min with a flow rate of 300 nL/min with a total run time of 90 min and thus eluted the peptides from the nanocolumn. We kept the mass spectrometer operating in a positive ionization mode with nanospray voltage set at 2.50-3.00 kV and source temperature at 275^{o}C. We set the method such that, in a full MS scan, the three precursor ions with the most intense signal were consecutively isolated and fragmented to acquire their corresponding MS^2 scans. Full MS scans were performed with 1 micro scan at a resolution of 3000, and a mass range of m/z 350-1500. Normalized collision energy (NCE) was set at 35%. Fragment ion spectra produced via high-energy collision-induced dissociation (CID) was acquired in the Linear Ion Trap with a resolution of 0.05 FWHM (full-width half-maximum) with an Ultra Zoom-Scan between *m/z* 50-2000. A maximum injection volume of 5 µl was used during data acquisition with partial injection mode. The mass spectrometer was controlled in a data-dependent mode that toggled automatically

between MS and MS/MS acquisition. MS/MS data acquisition and processing were performed by Xcalibur^{TM} software, ver. 2.2 (ThermoFisher Scientific, MA).

3.3.5 Database search

We used Proteome Discoverer software (ver. 2.1, Thermo Fisher Scientific) and UniProt human (*Homo sapiens*) protein sequence database (120672 sequences, and 44548111 residues) to search and match our raw files to the database and identify the proteins present in our samples. The reviewed protein sequences of human were downloaded from UniProt protein database (www.uniprot.org) on August 12, 2016. The considerations in SEQUEST searches for normal peptides were used with carbamidomethylation of cysteine as the static modification and oxidation of methionine as the dynamic modification. Trypsin was indicated as the proteolytic enzyme with two missed cleavages. Peptide and fragment mass tolerance were set at ± 1.6 and 0.6 Da, and precursor mass range of 350-3500 Da and peptide charges were set excluding +1 charge state. SEQUEST results were filtered with the target PSM validator to improve the sensitivity and accuracy of the peptide identification. Using a decoy search strategy, target false discovery rates for peptide identification of all searches were < 1% with at least two peptides per protein, a maximum of two missed cleavage, and the results were strictly filtered by ΔCn (< 0.01), Xcorr (≥ 1.5) for peptides, and peptide spectral matches (PSMs) with high confidence, that is, with *q*-value of ≤ 0.05. Proteins quantifications were conducted using the total spectrum count of identified proteins. Additional criteria were applied to increase confidence that PSMs must be present in all three biological replicates samples. Normalization of identified PSMs among LC-MS/MS runs was done by dividing individual PSMs of proteins with total PSMs and average of % PSM count was utilized for

calculating fold changes for different treatment conditions [107, 114]. For contrasting relative intensities of proteins between control, LPS, Statin, and Statin-LPS groups, all with or without crosslinkers were evaluated using cumulative confident normalized PSMs value.

3.3.6 Gene ontology and protein interaction analysis

We functionally categorized the protein-encoding genes using gene ontology systems by PANTHER classification system based biological process, molecular function, cellular components.[122] We used the heatmap to visualize protein abundances, and the cluster was generated by MeV software (ver. 4.9; http://www.tm4.org/).[123] We used the proteomic data set with UniProt identifiers and fold changes of total protein to generate core analysis through Ingenuity Pathway Analysis (IPA) (Ingenuity Systems, Redwood City, CA). TLR4 protein interaction networks, according to biological as well as molecular functions, were generated from the matched proteins in the Ingenuity Knowledge BASE. The settings of indirect and direct relationships between molecules based on experimentally observed data and other data sources were considered in the human database in the Ingenuity Knowledge Base to generate the core analysis. [124]

3.3.7 Immunoblotting

We washed the cells twice with 1X PBS buffer and then lysed with RIPA lysis buffer (same as protein preparation). We prepared the protein samples in the 2X Laemelli buffer and heated for 5 min at 95^{o}C. We separated the proteins on a 12% polyacrylamide gel. We transferred the proteins to a 0.45 µm nitrocellulose membrane for 1.5 hr at 100 V. Next, we blocked the nitrocellulose membrane with skim milk (5%) in TBST buffer for 2 hr at

room temperature (RT) and incubated with primary antibodies against Creatine Kinase (ab108388; Abcam), MARCKSL1 (ab184546; Abcam) in bovine serum albumin (5%) at 4°C for overnight. Goat anti-rabbit IgG secondary antibody conjugated to HRP (Abcam) was then used for 2 hr at RT. β-actin (ab8227; Abcam), the antibody was used as a loading control. The targeted protein bands were visualized using clarity western enhanced chemiluminescent substrate (BioRad).

3.3.8 Fluorescence staining

We grew the cells on 1M HCl treated glass slides and then used chilled 4% paraformaldehyde for 10 min at RT to fix the cells in the slides. Subsequently, we permeabilized the cells with 0.1% Triton X-100 in 1x PBS for 10 min. Next, we washed the cells with PBS and stained with Alexa Fluor 488@ phalloidin 2 μL / 2 mL in each well for 25 minutes at RT. Subsequently, we washed the cells with PBS and stained with Propidium iodide for 5 minutes and washed in PBS finally before fixed into coverslips. The cells were then imaged with a **Leica DMi8** confocal microscope (**Leica, IL, USA**). The images were analyzed using Lax X (Leica, IL, USA).

3.3.9 Statistical analysis

We used a built-in statistical package in Proteome Discoverer (Ver. 2.1) for the quantitative analysis of proteins as PSMs. Statistically significant results with $q \leq 0.05$ considered for analysis. We used the R package (ver. 3.5.3) to generate scatter plots and pairwise correlation matrices. The results here were only considered if the correlation coefficient (R^2) was > 0.80 (n=3).

3.4 Results

3.4.1 Identification of TLR-4 interacting proteins.

We wanted to understand the immune responsive interactome in TLR4 mediating immune signaling. We started with co-IP proteomics on HA-TLR4-MD2-CD14-HEK293 cells from four exposure conditions (control; LPS; statin; statin-LPS) with or without post-exposure treatment of ETD crosslinker (ETD-XL) (Figure 3.1).

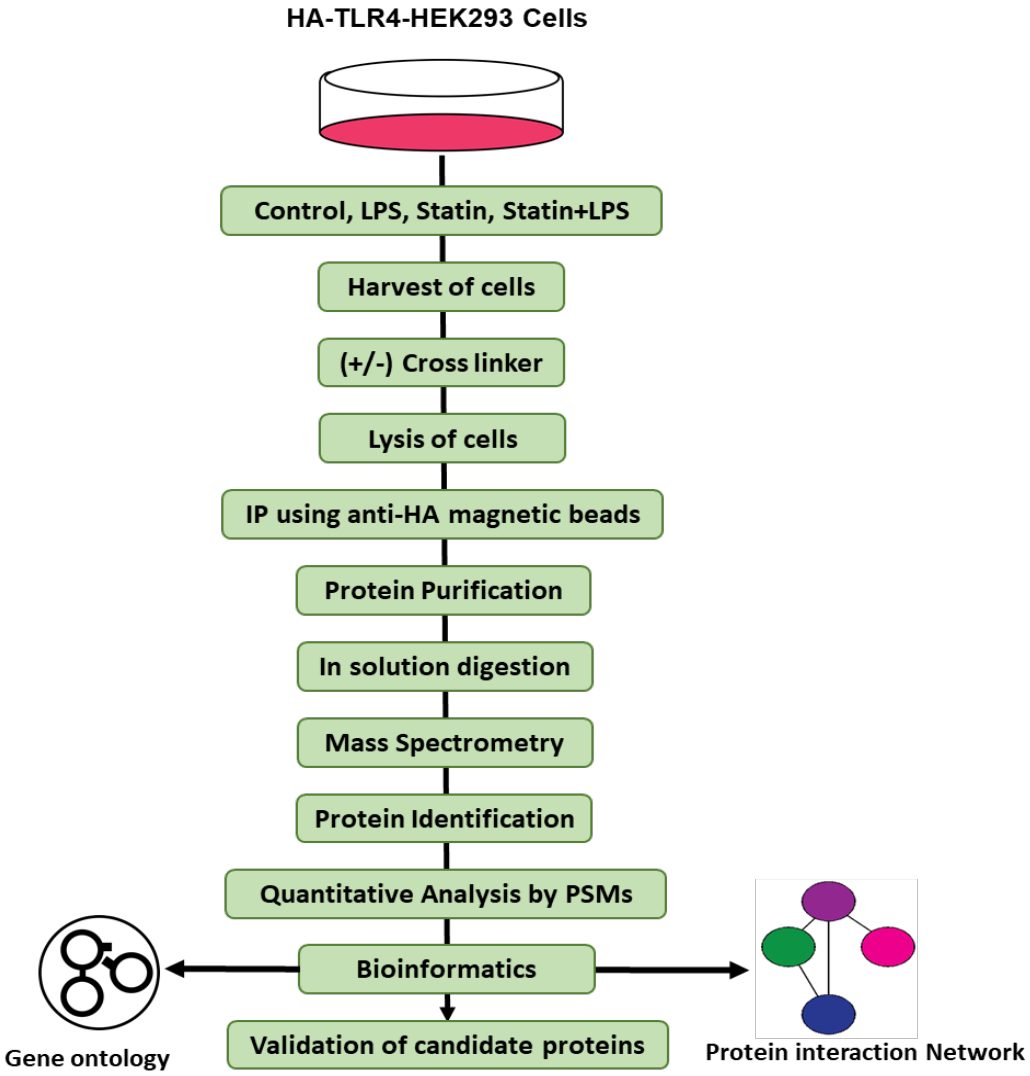


Figure 3.1. Experimental Procedure of the IP-cross linked MS-based proteomics Analysis. HEK293 cells treated with Statin, LPS, alone or in tandem and with or without cross-

linkers as depicted. Pull down samples were in solution tryptic digested and analyzed in nano-LC-MS/MS, followed by PSMs based quantitative analysis. Biochemical validation was performed with Western blot.

Co-IP proteomics samples of same four exposure conditions untreated of crosslinkers were also prepared. After pull-down with anti-HA magnetic beads, precipitated proteins were washed and dissolved in Laemelli buffer followed by methanol-chloroform purification and reconstitution in trypsin lysis buffer (50 mM Ammonium bicarbonate) for in-solution digestion. After in-solution digestion with trypsin, the peptides were analyzed by nano-LC-MS/MS and database searching (UniProt). Peptide Spectrum Matches (PSMs) were used for quantification of peptides. Correlation matrix comparisons among three biological replicates are shown in Figure S3.1 and S3.2. Pairwise correlation coefficients among the biological replicates showed a high correlation with an R^2 value of >0.80.

Overall, 712 proteins were identified and quantified altogether across all conditions. The data set was filtered using at least one unique peptide per proteins and a false discovery rate of 1%. Details about the identified proteins and peptides are big database search files which are not shown here. First, we consider all four-exposure condition for HA-TLR4 pulldown samples, where no crosslinker treatment was involved. Altogether in total, 425 proteins were identified across four conditions. Out of 425 proteins, 158 were commonly identified in all four exposures without cross-linker involvement.

On the contrary, 175 proteins were exclusive to single exposure (156 in control, 10 in LPS, 6 in Statin, and 3 in Statin-LPS) (Figure 3.2A). Next, we added into consideration of

the cross-linker treated samples in all four conditions. When cross-linker treated and untreated LPS exposure and Control samples were compared together, we found altogether 460 proteins are identified across these mentioned conditions. Of these, 165 proteins are commonly identified among these conditions. 10 proteins were identified, exclusively in LPS treated samples in the presence of crosslinkers (Figure 3.2B).

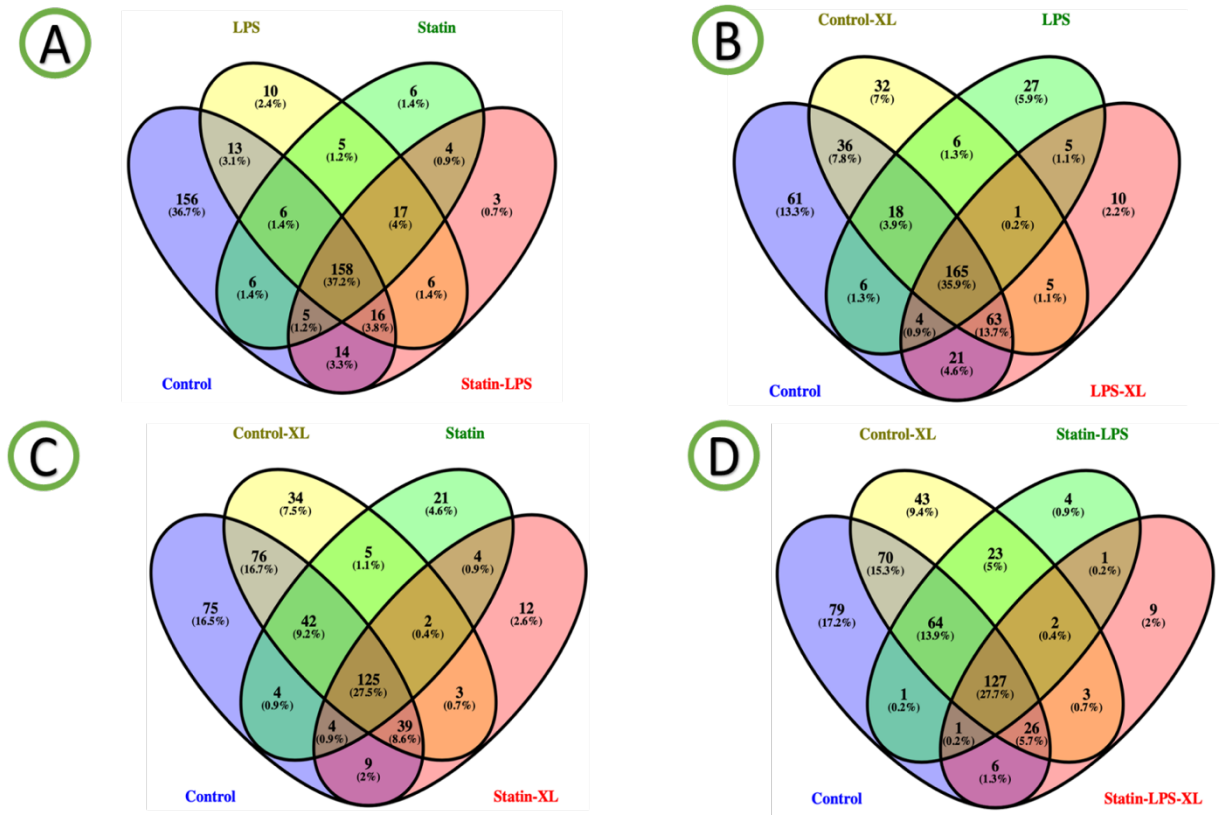


Figure 3.2. Venn diagrams showing distributions of identified proteins across all treatment conditions. The diagram shows the distribution of the total and uniquely identified proteins in HEK293 cells upon treatment with (A) Statin, LPS, Statin+LPS (B) Control cross-linker, and LPS cross-linker (C) Control cross-linker and Statin cross-linker (D) Control cross-linker and Statin+LPS cross-linker.

Similarly, when Statin or Statin-LPS were compared with cross-linker treated and untreated control, we identified, 12 or 9 proteins were exclusively identified in the presence of cross-linkers in Statin or Statin-LPS treatment, respectively (Figure 3.2C and 3.2D).

TLR4 interacting proteins' (N=712) relative expression (normalized PSMs) value (data not shown) was used to visualize a heatmap. The heatmap suggested that treatment with LPS, Statin, Statin-LPS introduced distinct biological states in the cell (Figure 3.3). The differential proteomics expression pattern was evident across different treatment condition.

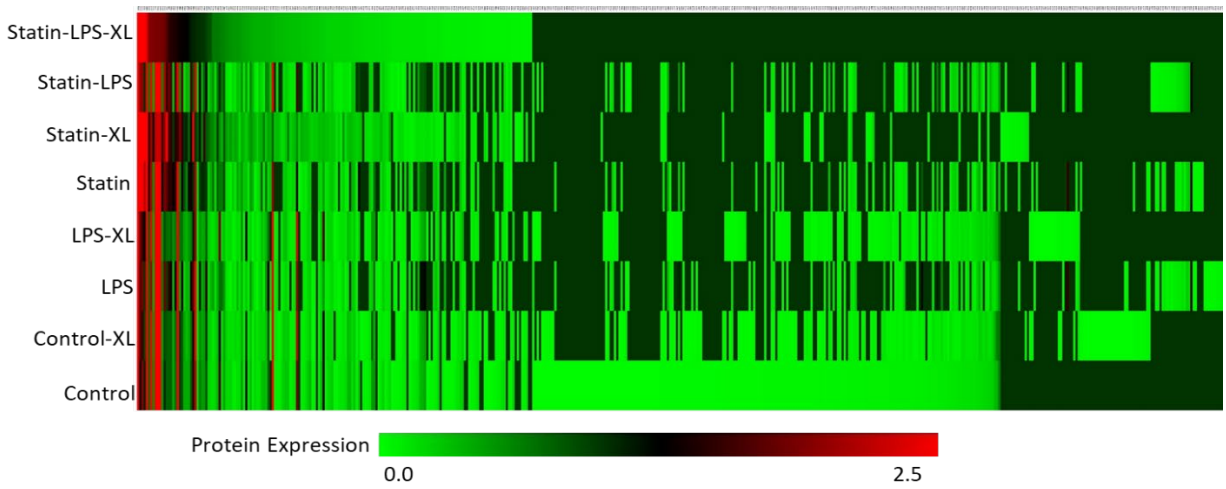


Figure 3.3. Heatmap showing relative expression of proteins across different treatment conditions. Proteins were differentially expressed with or without the presence of cross-linker across the different treatment conditions, e.g., Statin, LPS, or Statin+LPS.

All 425 proteins identified in the absence of cross-linker, across different treatment conditions, were categorized into gene ontology using panther classification system. [125]

The gene ontology categories: cellular components, biological process, molecular functions got represented in 6, 10, and 7 ontology pathways. All ontology pathways show differential representation across different treatment conditions (Figure 3.4).

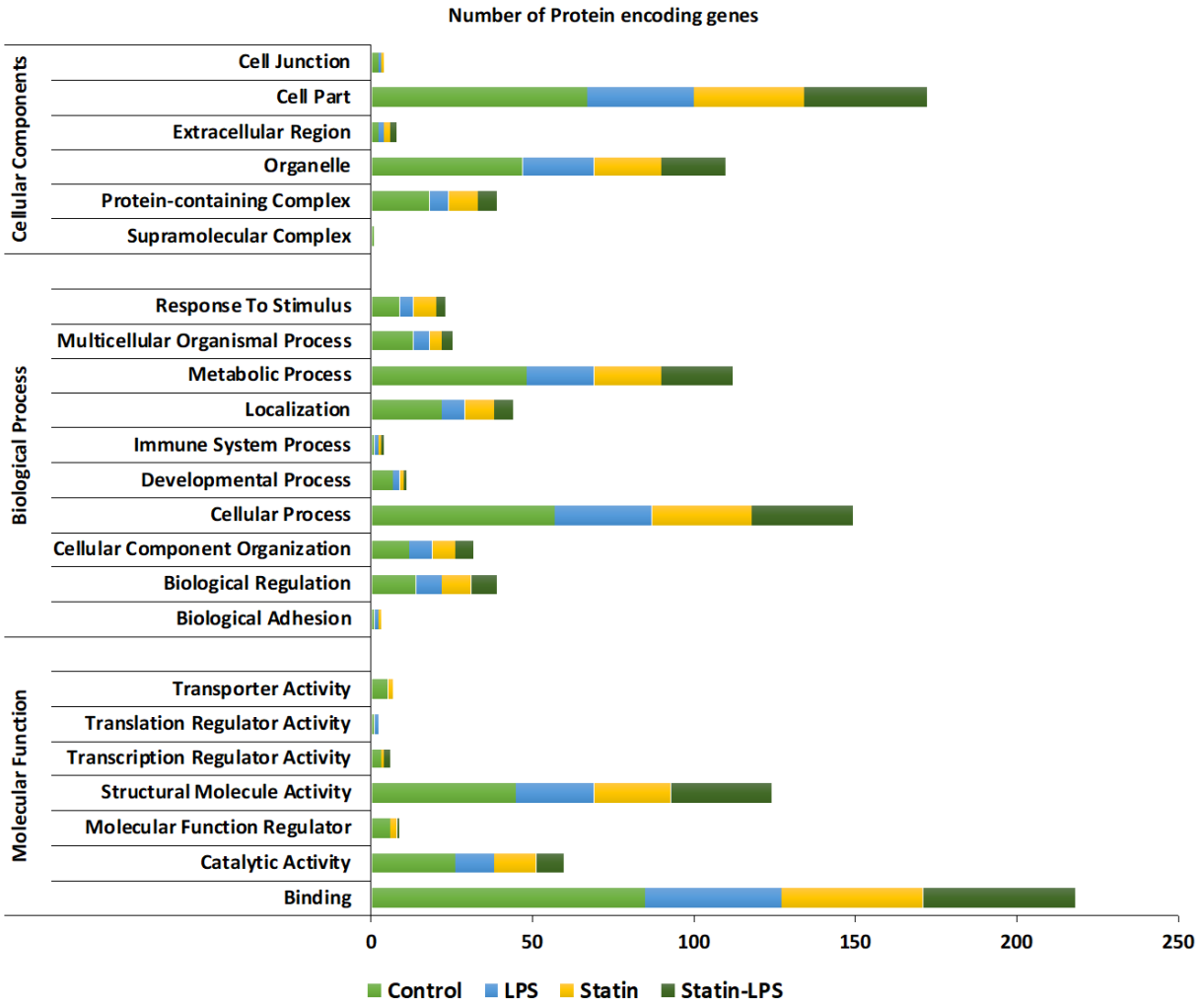


Figure 3.4. Gene ontology enrichment analysis. All proteins identified and quantified were subjected to gene ontology enrichment analysis based on molecular functions, biological process, and cellular components.

3.4.2 IPA-based TLR4-targeted protein interactions network

IPA (Ingenuity Systems) was used to perform core analysis, and hypothetical interaction networks, canonical pathways, functional and disease pathways were constructed, and putative upstream regulators among the TLR4 interactome proteins were identified. [126] In this analysis, we found 14 classified network types according to top disease and functions in each of LPS, Statin, or Statin-LPS treated data set. In all three treatment types, a TLR4 protein-interacting network was ranked number 4 (data not shown). The TLR4 network is centered on the cell to cell signaling interaction and signaling in all three-stimulus condition. With LPS and Statin+LPS stimulus, the TLR4 interacting network is centered additionally on cancer and organismal injury and abnormalities pathways on LPS and Statin+LPS responsive samples. With Statin stimulus, the TLR4 interacting network is centered additionally on DNA replication, recombination, and repair, RNA post-transcriptional modification (data not shown). Our work concern in this project was very specific to targeted pull-down of TLR4 and studied the TLR4 interactome. Thus, on the next step, we performed a TLR4 based network analysis (Figure 3.5, Supplementary S3.3, and S3.4). Here we observed 35 interacting protein partners. Expression of TLR4 itself was downregulated in the presence of statin (Figure 3.5) and highly upregulated with LPS, and Statin+LPS with expression been the highest in LPS stimulated cells. (Figure Supplementary S3.3 and S3.4).

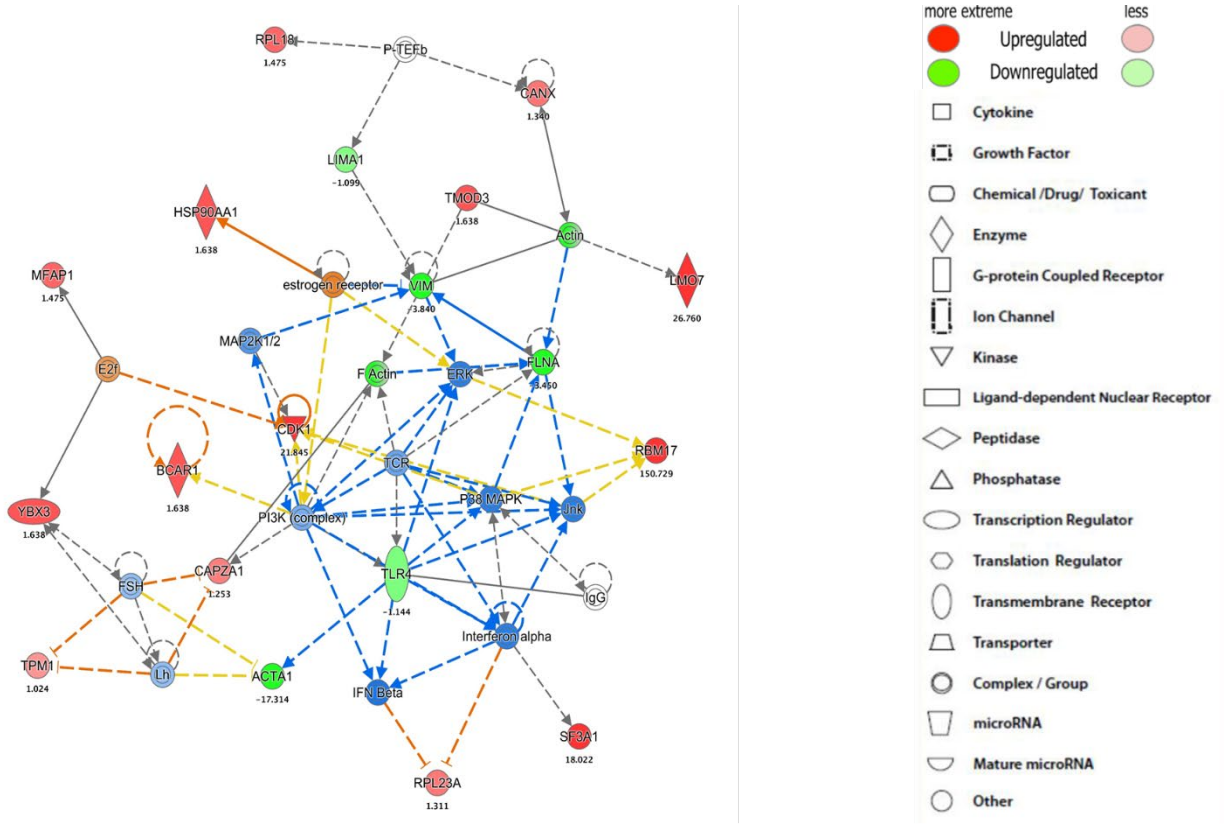


Figure 3.5. TLR4 targeted protein network with expression profile. The interaction network shown was generated using IPA bioinformatics software upon treatment of Statin in HA-TLR4-HEK293 cells.

In this putative network, TLR4 interacted directly with IgG and indirectly with eight different proteins. These include IFN Beta, Interferon-alpha, Actin Alpha 1 (ACTA1), Phosphoinositide 3-kinase (PI3K complex), T-cell receptor (TCR), p38 mitogen-activated protein kinases (P38MAPK), c-Jun N-terminal kinase (JNK) and extracellular signal-regulated kinase (ERK1). Among these, four are important classes of protein kinases. These kinases then intern interact with other cellular kinase-like Cyclin-dependent kinase 1 (CDK1) or transcription factors like E2F, receptor protein like Breast cancer anti-estrogen resistance protein 1 (BCAR1), filamentous proteins like ACTA1, Filamin-A

(FLNA), Vimentin (VIM), F-actin-capping protein subunit alpha-1 (CAPZA1) and so on. In our pulldown, we were able to experimentally observe kinase CDK1, filamentous proteins like ACTA1, Filamin-A, Vimentin, F-actin-capping protein subunit alpha-1 (CAPZA1), etc. Hence, proteins with different important cellular functions co-appears in our network. Additionally, we observed, in our study, the expression of filamentous proteins going up with LPS treatment, aka activation of TLR4 signaling, goes down with Statin and goes up again with the addition of LPS. This phenomenon of filamentous proteins was further confirmed through Fluorescence staining, where actin filaments and nucleus were stained.

3.4.3 Fluorescence staining

The phenomena mentioned above were further confirmed when Alexa Fluor 488® phalloidin, a high-affinity filamentous actin probe was used to stain the F-actins selectively and Propidium iodine was used to bind the DNA in the nucleus (Figure 3.6). From Fluorescence staining, it's evident that treatment of LPS increases filamentous protein production, increases in cell and nuclear size to activate the cells for immune response. So, the model expression system can work as an ideal model to study atypical expression, activation, and suppression of immune response through TLR4 signaling interactome.

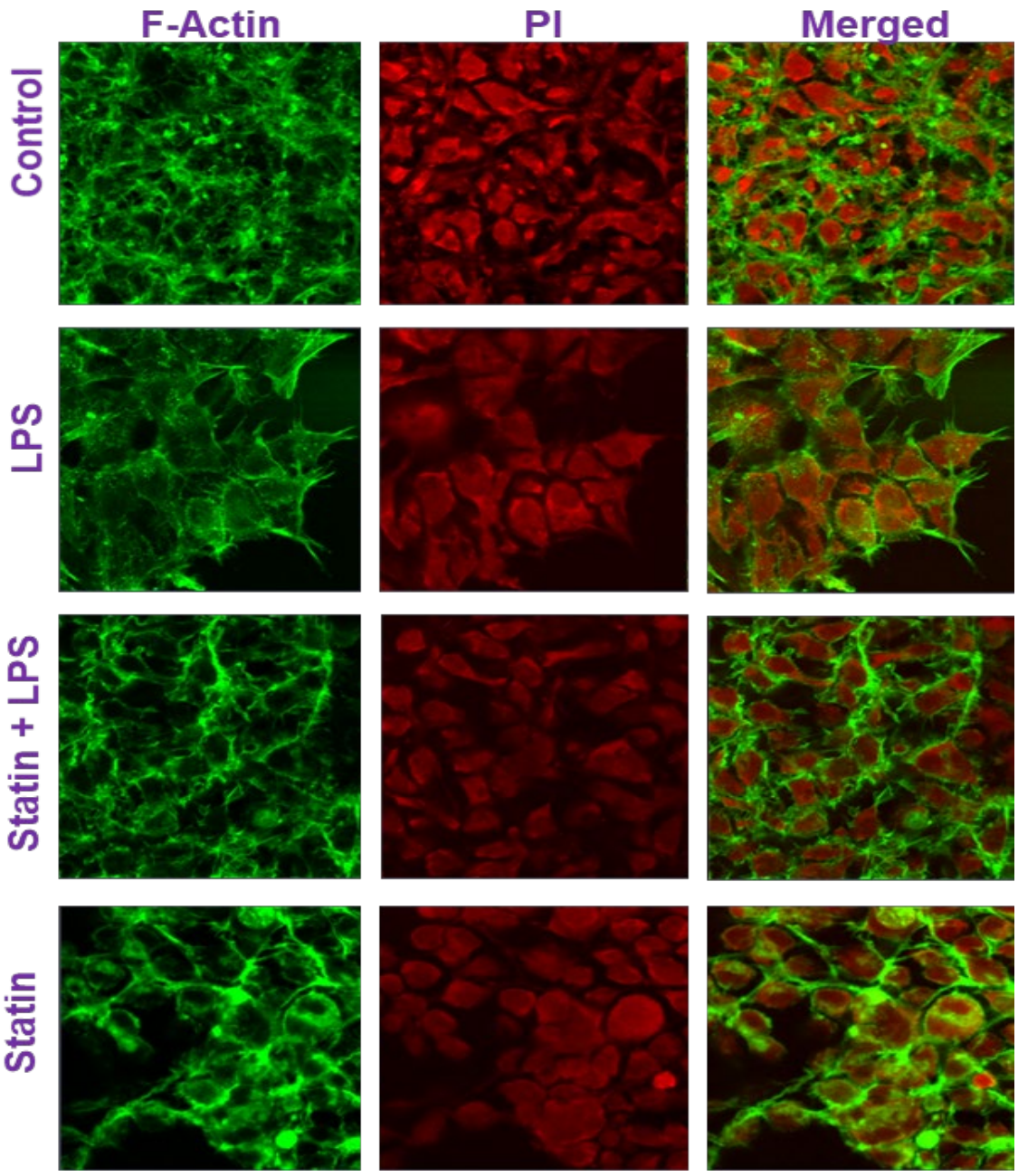


Figure 3.6. Fluorescence staining of F-actin and Nucleus. HA-TLR4-HEK293 cells were stained with Alexa 488 conjugated phalloidin probe (Green) and propidium iodide (Red) in all four treatment conditions. LPS induced immune stimulation shows even for HEK cells, filamentous protein expression is increased.

3.4.4 Protein identification and interactions after cross-linking study

In this study, we have used our lab developed in house, ETD cleavable cross-linker to covalently capture the low abundance, weakly interacting proteins in TLR4 interactome. After cross-linker treatment, we identified 244, 214, 120, and 116 proteins, respectively, after control (no stimulant), LPS, Statin, or Statin+LPS stimulation (Figure 3.7).

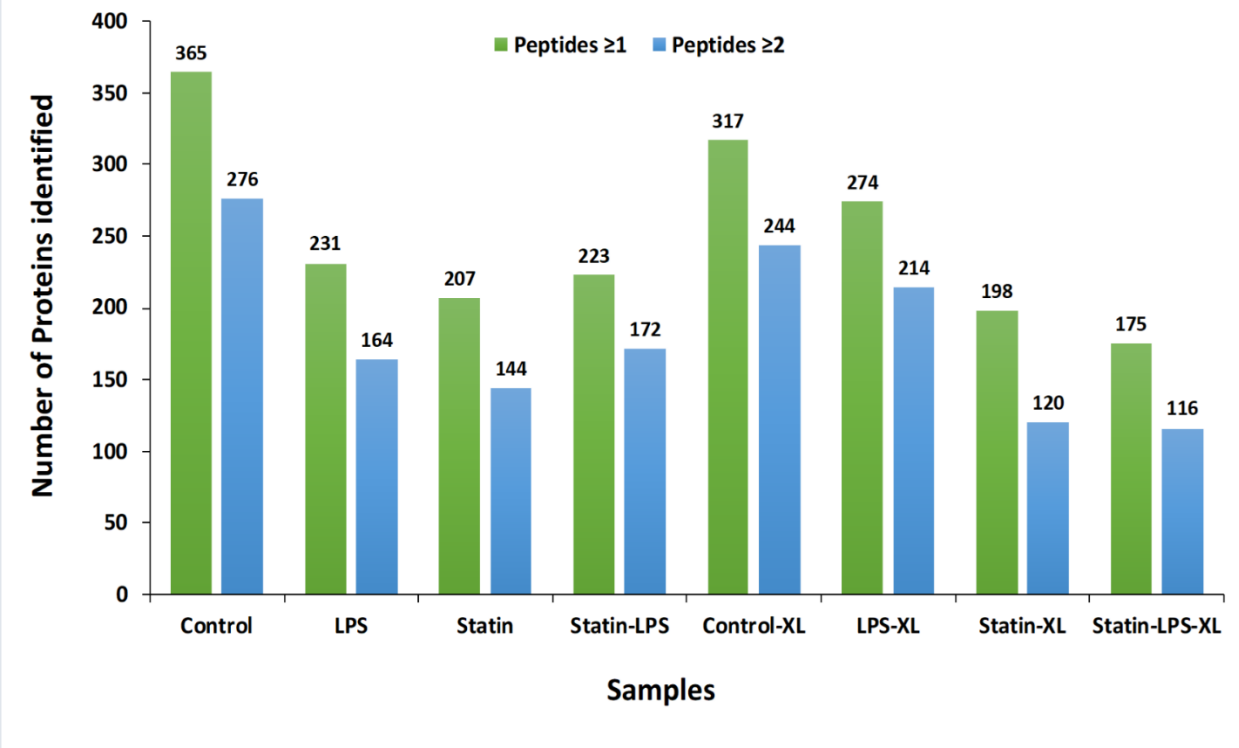


Figure 3.7. The number of proteins identified with or without cross-linkers across each treatment conditions where at least 1 or 2 peptides identified for each sample.

After stringent filtering among these and comparing with control and control cross-linked sample, we identified 10, 12 and 9 proteins are exclusively identified respectively for LPS, Statin and Statin+LPS treated samples (Figure 3.2B, 3.2C, 3.2D). From this list, we selected two proteins, Creatine kinase and Macrophage Myristoylated alanine-rich C-

kinase substrate (MacMARCKS) for further validation. Creatine kinase brain isoform was observed exclusively in the presence of Statin, and cross-linked treatment and Creatine kinase U-type mitochondrial was observed exclusively in the presence of LPS and cross-linked treatment. MacMARCKS was identified in cross-linker treated samples stimulated by both Statin and Statin+LPS. Exclusive identification in the presence of crosslinker for one stimulus suggested possible distinct patterns of responsiveness of these proteins.

3.4.5 Validation of selected proteins

We wanted to verify our mass spectrometry-based proteomics data. So, in the final step of our experiment, we performed immunoblot of candidate proteins that we have selected for validation in the previous step. We used whole-cell lysate after respective treatments of stimulant, e.g., LPS, Statin, or Statin+LPS for immunoblot analysis to establish the expression status of these three proteins. β-Actin was used as a loading control for immunoblot analysis, and all experiments were performed three times to confirm the expressions. After performing the immunoblots in three replicates, we concluded that macrophage myristoylated alanine-rich C-kinase substrate (MacMARCKS) protein, upon treatment with Statin, shows an irreversible decrease in expression and this trend continues down to even lower despite the reintroduction of LPS in tandem after removal of Statin (Figure 3.8).

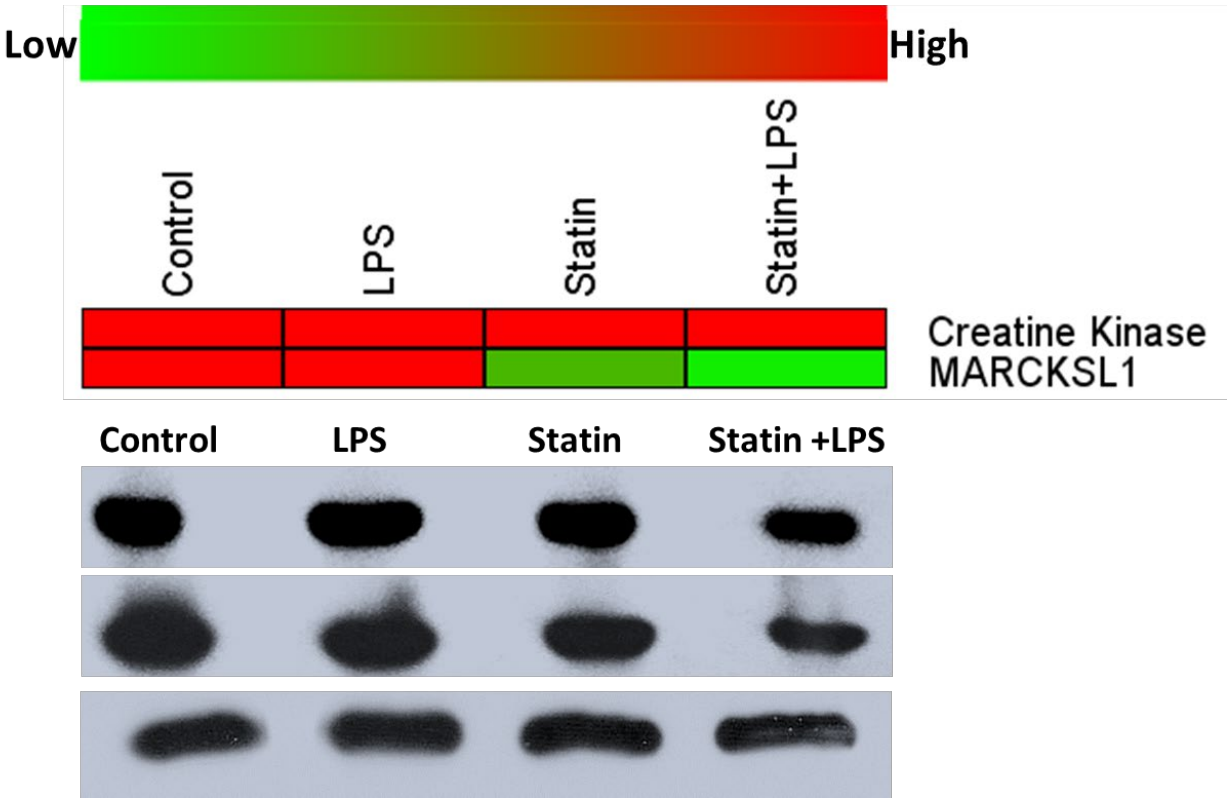


Figure 8. Western blot analysis of Creatine Kinase and MacMARCKS (MARCKSL1) and β-Actin used as a loading control.

3.5 Discussion

In this study, we used, HA-TLR4 transfected HEK293 cell as an immune signaling model system to study the TLR4 mediated immune interactome through the atypically expressed HA-TLR4 on transfected HEK293 cells. HEK cells are easy to maintain and easy to transfection manipulation through foreign DNA. Hence, HEK cells are a popular choice for heterologous proteins' expression and their functional characterization. [127-129] Also, due to the suboptimal level performance of IP antibodies against TLRs, we had to consider a stable epitope (HA) tagged TLR4 expressing cell line for our TLR4 immune interactome study design. Also, the HEK 293 cells do not have any native TLRs

expressing on its membrane, yet they produce downstream functional signaling molecules, pro-inflammatory cytokines after TLR4 ligand stimulation. [130-131] This makes them an ideal system to control the expression of a selected TLR in the absence of endogenous background receptors. Hence, we selected HA-TLR4 transfected HEK293 cells for targeted immune precipitation against HA-tagged TLR4 and cross-linking proteomics experiment to study TLR4 mediated immune signaling.

We have used two antagonistic stimulants, e.g., LPS a pro-inflammatory stimulant [94] and Statin, an anti-inflammatory stimulant [107-108] and in tandem, the first Statin then LPS to incorporate a bigger picture, where we can see immune signaling in pro-inflammatory condition and anti-inflammatory condition and also in transition from anti-inflammatory to pro-inflammatory condition. Introduction of cross-linker into the study design enabled us to capture and enrich low abundance, transiently interacting protein partners during these different three above mentioned immune signaling condition.

After immune precipitation, tryptic digestion, LC-MS/MS analysis, and database search; we were able to get lists of proteins expressed in mentioned conditions. We use the PSMs value to calculate normalized PSM percentage and fold changes among these conditions compared to the control sample. From this data set, we used the gene accession number lists (data not shown) of specific treatment conditions to generate Venn diagram and identified the proteins that are exclusively expressed in each specific condition. (Figure 3.2) (data not shown). In this step, we identified 10, 12, and 9 proteins exclusively identified with cross-linker in the presence of LPS, Statin, and Statin+LPS. The normalized PSM percentage values from different conditions were used to create a heat map (Figure 3.3) (data not shown) that shows differential expression comparison of each

protein across all treatment conditions. The gene accession number lists were also used to generate gene ontology information of identified proteins across all treatment conditions and classified the proteins according to cellular components, biological process, and molecular functions. All ontology conditions showed differential expression across the three ontology categories and each of their subcategories. (Figure 3.4) (data not shown).

The fold change data (data not shown) we calculated from PSMs were used to generate core analysis through Ingenuity Pathway Analysis, and hypothetical interaction networks, canonical pathways, functional and disease pathways were constructed, and putative upstream regulators of TLR4 interactome were identified (Figure 3.5 and Supplementary Figures S3.3, S3.4). In this step, TLR4 showed LPS dependent higher expression. The network showed direct and indirectly interacting proteins to TLR4, and at the same time, their comparative expressions. Notable interacting partners included different cytokines (e.g., IFN–α, IFN–β), different kinases (e.g., p38 mitogen-activated protein kinases, c-Jun N-terminal kinase, extracellular signal-regulated kinase, etc.), different filamentous proteins (e.g., Actin Alpha 1, Filamin-A, Vimentin, F-actin-capping protein subunit alpha-1) and so on.

The appearances of different types of proteins starting from receptors to transcription factors to kinases and filamentous proteins in the network suggests the potentials for a complex interaction between proteins in the proposed TLR4 interactome (Figure 3.5, Supplementary S3 and S4). In this study, we also observed that the filamentous proteins'

(e.g., Vimentin, F-Actin, Filamin) expression goes up with LPS stimulation, goes down with Statin stimulation and goes up again compared to Statin once LPS is reintroduced in tandem after removal of Statin. (Figure 3.5, Supplementary S3.3 and S3.4). This corresponds to similar published literature, though on macrophage cellular system where Actin, Filamin A or Vimentin expression were reported to be directly related to macrophage activation and function. [132-134] This trend was further confirmed through Fluorescence staining experiment where Alexa Flour 488@ phalloidin selectively probes F-Actins, and Propidium Iodine bound the DNA in the HEK293 nucleus. (Figure 3.6) We observed an increase in filamentous protein production, cell, and nuclear size upon LPS stimulation. So, our model system with atypically expressed HA-TLR4 in HEK293 cell shows similar patterns of immune responsive gene expression as regular macrophage and thus is a good model system to further the analysis to find a novel response protein in the TLR4 interactome.

As mentioned earlier, incorporation of the cross-linker in the experiment was to enrich low abundance transiently interacting proteins in the TLR4 interactome. So, from the protein lists of exclusively identified proteins with cross-linker (Figure 3.2) (data not shown) in LPS or Statin or Statin+LPS treated cells, we wanted to choose candidate proteins for biochemical validation for the next step.

From the list (data are not shown), we selected Creatine kinase and macrophage myristoylated alanine-rich C-kinase substrate (MacMARCKS) protein for further biochemical validation. Creatine kinase is a marker of kidney function. [135] It has been reported that due to trauma and muscle injury, Creatine kinase level can be elevated. [136]

This enzyme catalyzes the reversible transfer of γ-phosphate group of ATP to guanidino group of Creatine to yield phosphocreatine (PCR). In skeletal muscle, a large pool of phosphocreatine is used for ATP regeneration. [137] As we have observed the presence of different kinases in our TLR4 network generated through IPA analysis, and Creatine kinase appeared in our lists of exclusively identified transiently interacting proteins in the presence of crosslinker, we thought Creatine Kinase may have some role in TLR4 interactome as well, which may or may not be for ATP regeneration. Therefore, we chose Creatine Kinase as a potential candidate for validation through western blots.

The second candidate that we selected for biochemical validation is macrophage-enriched myristoylated alanine-rich C kinase substrate (MARCKS-related protein / MRP/ MacMARCKS/MARCKSL1). Both myristoylated alanine-rich C kinase substrate (MARCKS) and MacMARCKS proteins are protein kinase C (PKC) substrates that take part in a myriad of functions in the living system. Both share identical effector domains, binding to calmodulin in a phosphorylation-dependent manner. [138] MacMARCKS have been implicated with membrane-cytoskeletal signaling, integrin activation, cell spreading, cell-cell adhesion, migration, phagocytosis. [139-141]

The cellular responsiveness to stimulus varies upon many factors, e.g., expression of its cellular receptors, localization, degradation, etc. Here, in our biochemical validation through western blot (Figure 3.8), for creatine kinase, we observed no variation of expression among Statin, LPS, or Statin+LPS treatments compared to that of the untreated control sample. This means, from the transition of anti-inflammatory status to pro-inflammatory status, Creatine kinase does not undergo gene expression regulation and hence, are not a partner in TLR4 interacting immune signaling mechanism. However,

on the contrary, for MacMARCKS, in our study with HA-TLR4-HEK293 Cells, we observed the inclusion of Statin, lowers MacMARCKS expression compared to Control or LPS. However, LPS treatment did not lead to any change of expression of MacMARCKS protein in our sample. Therefore, when we compared to control with LPS treated sample, the expression of MacMARCKS are same in both. Treatment of Statin, an anti-inflammatory effector lowered the expression of MARCKS, and since LPS does not have any effect on MacMARCKS expression, hence, re-introduction of LPS in tandem after Statin treatment could not help bring up the MacMARCKS expression, rather it followed the same trend to go down even lower. Although MARCKS had been reported to be expressed in macrophage through LPS stimulation, in Madin–Darby canine kidney (MDCK) epithelial cells and renal tubule cells, they are endogenously expressed. [142-143] Thus, is it in our case with HA-TLR4 transfected Human embryonic kidney (HEK293) cells: LPS stimulation did not bring in any change in expression of MacMARCKS. In 2018, Christina M. Van Itallie et al. showed treatment with IFN-γ and TNF-α in epithelial cells, increase MacMARCKS (MRP/MARCKSL1) expression.

On the contrary, when they deleted MacMARCKS expression, this did not abrogate cytokine production. Their suggestion was MacMARCKS was not required for occludin dependent IFN/TNF response or cytokine production, unlike their similar family protein MARCKS. Treatment of Statin ensures direct suppression of cytokines (e.g., IFN-gamma, tumor necrosis factor (TNF)-alpha, interleukin (IL)-2, and IL-4. [109] In our case, with Statin treatment, expression of MacMARCKS is reduced and that too in an LPS independent manner. So, we can say, inductions of cytokines are required for MacMARCKS expression. In Macrophage cell line, LPS stimulation increases the production of

cytokines that in turn, is responsible for increased expression of MacMARCKS. In cells with endogenous expression of MacMARCKS and Toll-like receptor 4, Statin induction coupled by suppression of cytokine production is pivotal to repression of MacMARCKS. All of these establish MacMARCKS as a member in the Toll-like receptor 4 mediated signaling pathway.

3.6 Supplementary Information

3.6.1 Supplementary Figures

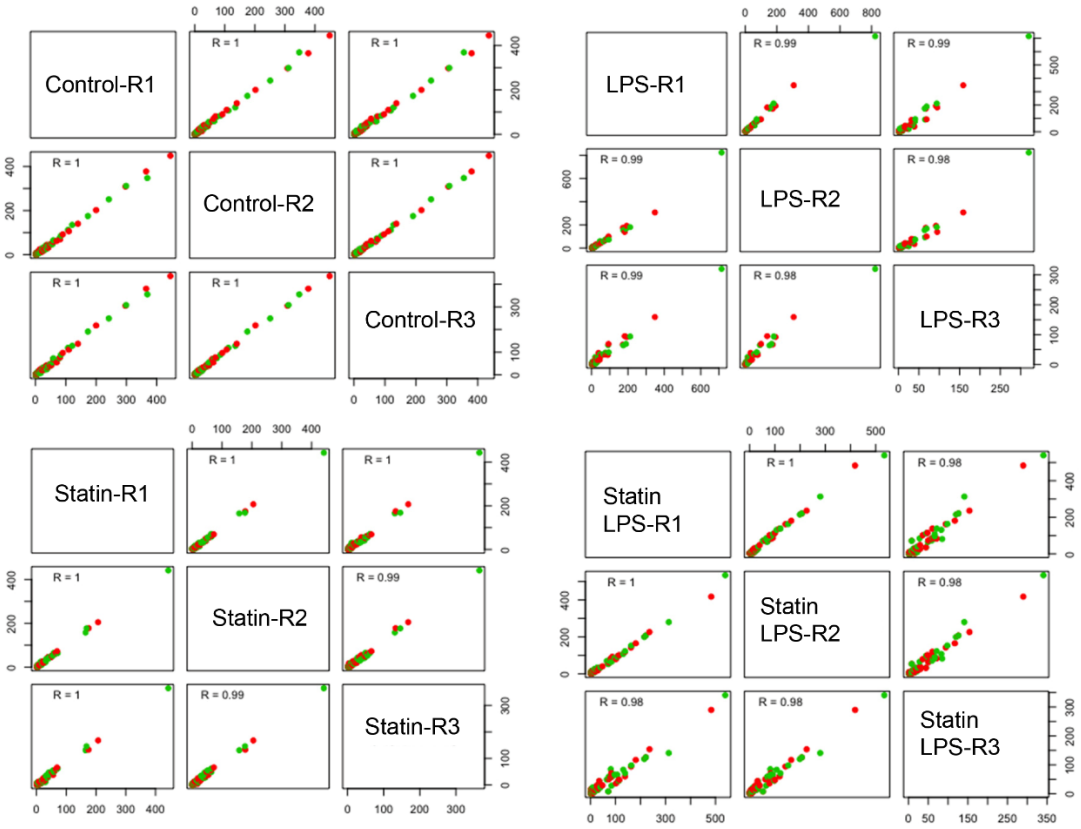


Figure S3.1. Scatter plots and pairwise correlations among the biological replicates showing significant correlation patterns among replicates of Control sample, LPS, Statin and Statin+LPS treated samples without cross-linker. The PSMs (Spectral counts) of replicates are plotted against each other across x and y-axis, correspondingly. Every spot

symbolizes the abundances of proteins and corresponds to Pearson's correlation coefficient (R^2) of 1. The scatter plot and pairwise correlation analysis were performed by R package ver. 3.5.3.

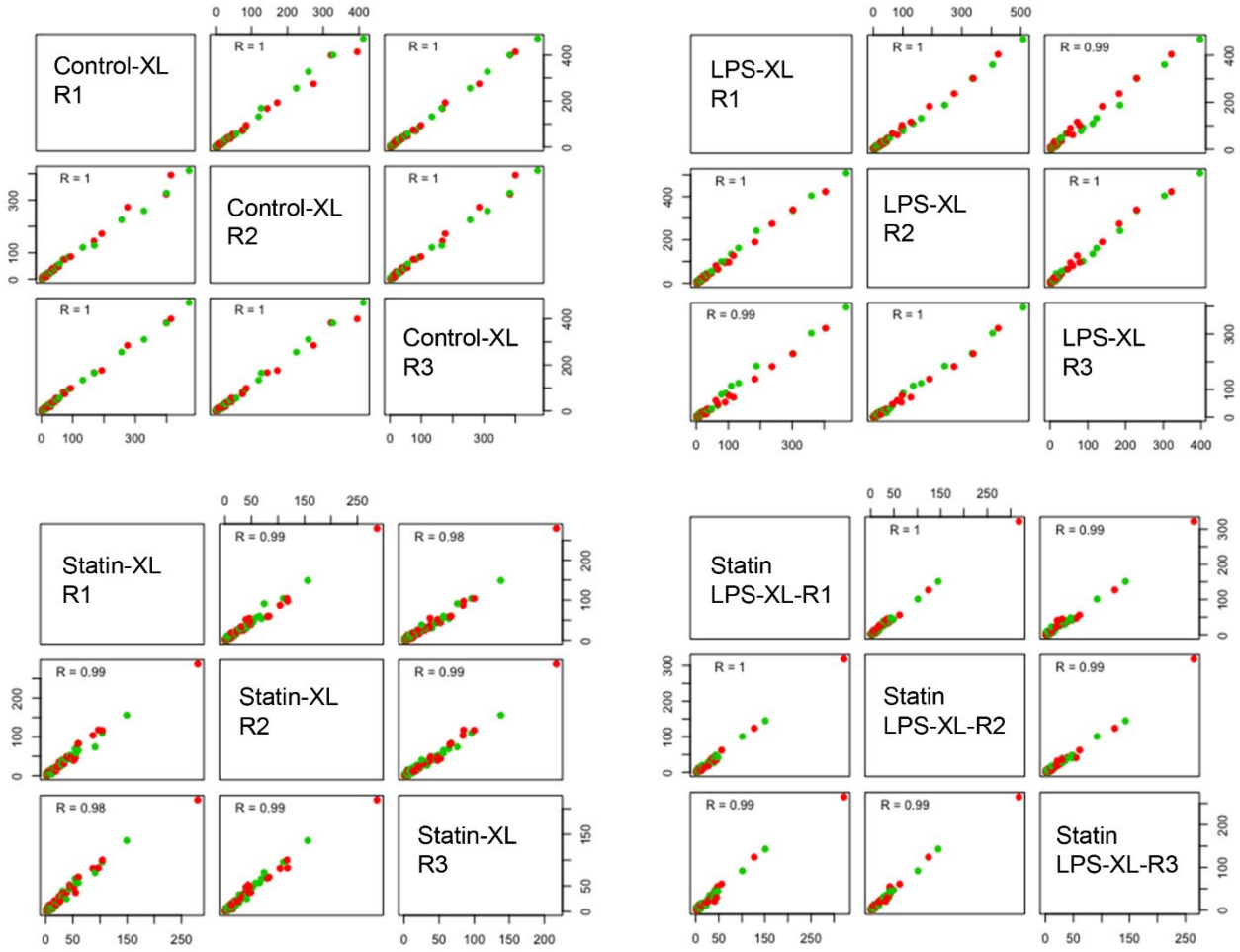


Figure S3.2. Scatter plots and pairwise correlations among the biological replicates showing significant correlation patterns among replicates of Control sample, LPS, Statin, and Statin+LPS treated samples with cross-linker treatment. The PSMs (Spectral counts) of replicates are plotted against each other across x and y-axis, correspondingly. Every spot symbolizes the abundances of proteins and corresponds to Pearson's correlation coefficient (R^2) of 1. The scatter plot and pairwise correlation analysis were performed by R package ver. 3.5.3.

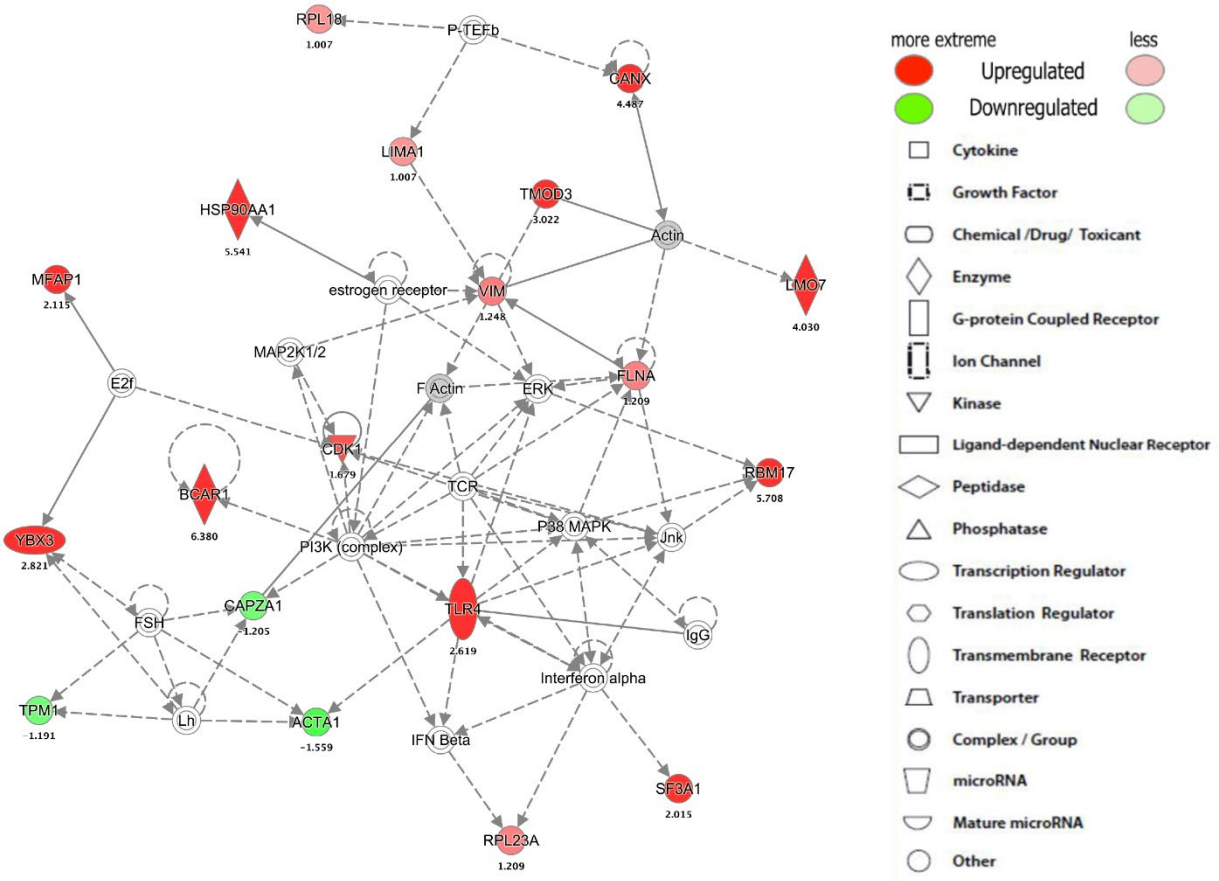


Figure S3.3. TLR4 targeted protein network with expression profile. The interaction network shown was generated using IPA bioinformatics software upon treatment of LPS in HA-TLR4-HEK293 cells.

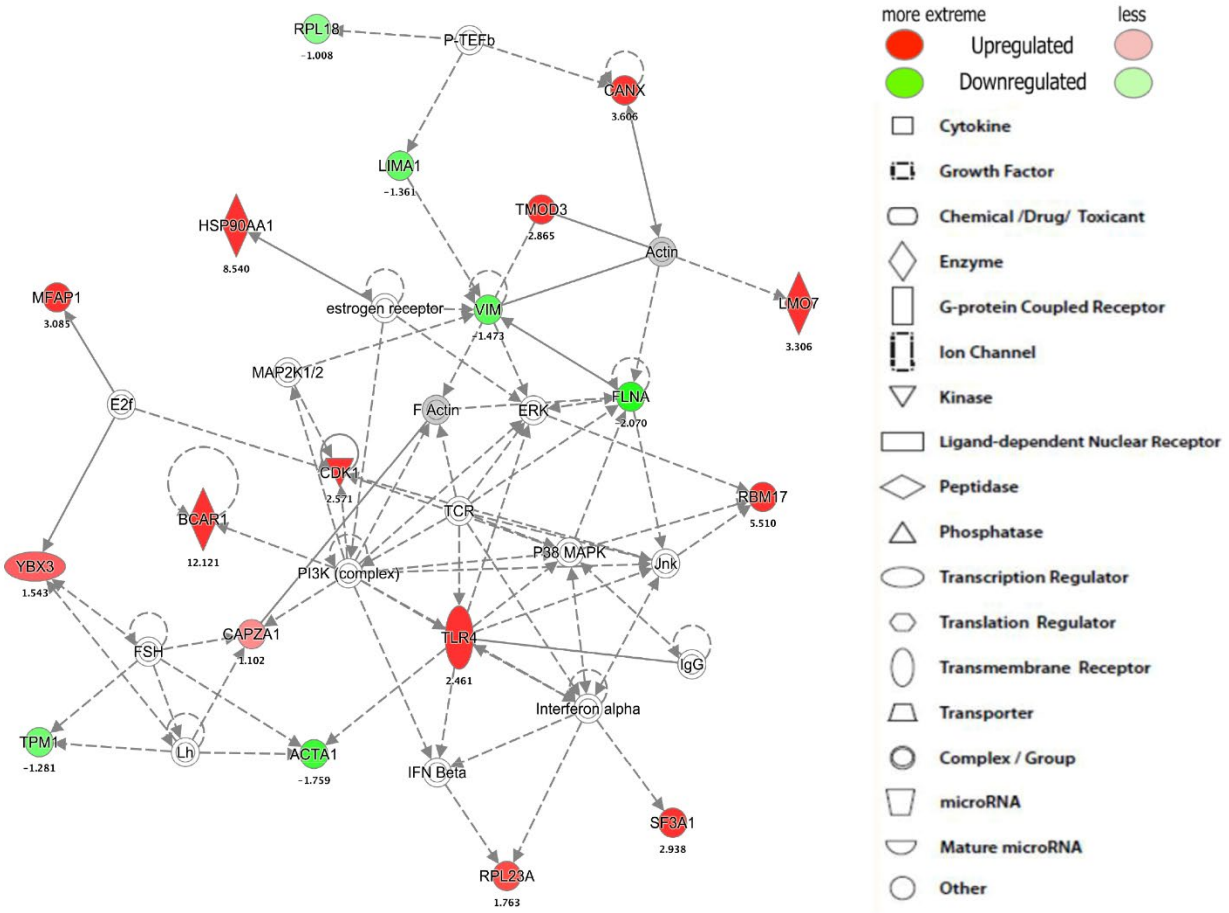


Figure S3.4. TLR4 targeted protein network with expression profile. The interaction network shown was generated using IPA bioinformatics software upon treatment of Statin+LPS in HA-TLR4-HEK293 cells.

Acknowledgments: We acknowledge support from NIGMS, NIH grant number 1UA5GM113216-01, for funding of this research. Authors also recognize start-up funds supports to PI Dr. Chowdhury from UT Arlington. UT Systems Shared Proteomics Networks are also acknowledged for a Thermo LTQ-velos pro mass spectrometer.

Chapter 4

Concluding Summary

The thesis focuses on two projects. In the first project described in chapter 2, we wanted to address and solve a bottleneck in mass spectrometry-based bottom-up cellular proteomics sample preparation. For any experiment design with the complex cellular proteome, we must disrupt the cell membranes to get the proteins out of the cells. The cell lysis buffer for this purpose needs to have high SDS concentration (10 % in Ripa lysis buffer). Incorporation of the SDS in the lysis buffer assists dissolving both membrane-bound and cytosolic proteins into the solution. However, presence of detergent in the solvent can interfere with both tryptic digestion and downstream mass spectrometric analysis through ion suppression in positive ion mode and through the formation of adduct ion and through a low signal to noise ratio. Hence, different sample purification approaches are in practice that purifies the cell lysate proteins as soluble precipitates out from the lysis buffer. Popular methods of purification are methanol-chloroform purification, acetone purification, detergent removal spin column purification, etc. The problem with precipitate proteins is, once they are precipitated out, it is no longer possible to take back all the proteins into the solution phase in the absence of detergent owing to the complex hydrophobic nature of some proteins. Therefore, a good amount of proteins ends up being discarded from the bottom-up proteomics samples as they could not be dissolved back to the digestion solution. Hence, we came up with a simple solution to this issue, where we used mass spectrometry compatible surfactant deoxycholic acid (0.2% in 50 mM Ammonium bicarbonate) to dissolve the methanol-chloroform purification protein precipitatebefore tryptic digestion. This is followed by a simple acidification step after the

digestion to remove the deoxycholic acid as a precipitate. The peptides in the solution phase are then desalted in C18 column and run in LC-MS/MS for identification of the proteins. We compared our modified method with other routine methods like methanol-chloroform purification, acetone purification, detergent removal spin column purification, etc. and we conclude that the modified method improved significantly improved protein identification compared to control. This work, we believe, will bring a change in the bottom-up proteomics sample preparation strategy in the scientific community.

In chapter 3, a second project is described where, we studied Toll-like receptor 4 immune signaling through chemical cross-linking, co-immunoprecipitation proteomics. TLR-4 is an immune signaling receptor on macrophage and other antigen-presenting cells. TLR-4 is activated through lipopolysaccharides (LPS) found on the outer membranes of a wide range of gram-negative bacteria. TLR-4 mediated signaling is mediated through two different pathways with assistance from 4 adapter's proteins. This makes TLR-4, the most important and complex immune signaling receptor of the innate immune response mechanism. However, the detailed TLR-4 interactome is yet to decipher. Hence, we wanted to study the TLR-4 interactome and find a possible interacting partner in the TLR-4 network. We selected an HA-TLR-4 transfected Human embryonic kidney cell line for our experiment. This transfected cell does not express any other kind of toll-like receptors but TLR-4. Hence cross-reactivity with other TLRs can be avoided.

At the same time, the HA tag part can be used as an affinity handle for immune precipitation experiment that native TLR-4 does not have. Thus, we used anti-HA antibody-bound magnetic beads to pull down TLR-4 interacting proteome. An in-house

chemical crosslinker was used as well to capture the transiently interacting proteins of this interactome network. All these were performed in four different conditions where A) no stimulus were present (Control), B) a pro-inflammatory TLR-4 immune signaling activating stimulus (LPS) was present, C) an anti-inflammatory cytokine lowering stimulus (Statin) was present and D) anti-inflammatory stimulus (Statin) followed by Pro-inflammatory stimulus (LPS) both were treated. We expected to understand the TLR-4 interactome better from this transition of signaling in TLR-4 specific network pull-down analysis. The immunoprecipitated samples were in-solution tryptic digested and analyzed in tandem mass spectrometry. We visualized the TLR-4 interacting network and from the cross-linked Statin and Statin-LPS sets, identified two putative candidates, e.g., Creatine kinase and MacMARCKS that may have a TLR-4 mediated signaling. Western blot analysis confirmed MacMARCKS expression lowers irreversibly with the treatment of Statin, independent of LPS stimulus. Thus, we confirm MacMARCKS as a partner in TLR-4 interactome that undergoes repression in anti-inflammatory condition.

Chapter 5

References

1. Finehout, E. J.; Lee, K. H., *Biochem Mol Biol Educ* **2004,** *32* (2), 93-100.

2. Greaves, J.; Roboz, J., *Mass spectrometry for the novice*. CRC Press: 2013.

3. Chang, W. C.; Huang, L. C.; Wang, Y. S.; Peng, W. P.; Chang, H. C.; Hsu, N. Y.; Yang, W. B.; Chen, C. H., *Anal Chim Acta* **2007,** *582* (1), 1-9.

4. Wilm, M., *Mol Cell Proteomics* **2011**.

5. Christian, G. D.; Dasgupta, P. K.; Schug, K. A., *Analytical Chemistry*. Wiley: 2013.

6. Pandey, A.; Mann, M., *Nature* **2000,** *405* (6788), 837-46.

7. Resing, K. A.; Ahn, N. G., *FEBS Lett* **2005,** *579* (4), 885-9.

8. Potriquet, J.; Laohaviroj, M.; Bethony, J. M.; Mulvenna, J., *PLoS One* **2017,** *12* (7), e0175967.

9. Crowell, A. M.; Wall, M. J.; Doucette, A. A., *Anal Chim Acta* **2013,** *796*, 48-54.

10. Vertommen, A.; Panis, B.; Swennen, R.; Carpentier, S. C., *Planta* **2010,** *231* (5), 1113-25.

11. Thiede, B.; Hohenwarter, W.; Krah, A.; Mattow, J.; Schmid, M.; Schmidt, F.; Jungblut, P. R., *Methods* **2005,** *35* (3), 237-47.

12. Brodbelt, J. S., *Anal Chem* **2016,** *88* (1), 30-51.

13. Guthals, A.; Clauser, K. R.; Frank, A. M.; Bandeira, N., *J Proteome Res* **2013,** *12* (6), 2846-57.

14. Deracinois, B.; Flahaut, C.; Duban-Deweer, S.; Karamanos, Y., *Proteomes* **2013,** *1* (3), 180-218.

15. Kall, L.; Vitek, O., *PLoS Comput Biol* **2011,** *7* (12), e1002277.

16. Maccarrone, G.; Bonfiglio, J. J.; Silberstein, S.; Turck, C. W.; Martins-de-Souza, D., *Methods Mol Biol* **2017,** *1546*, 223-234.

17. Murphy, K.; Weaver, C., Janeway's Immunobiology; Nineth Edition. Taylor & Francis: 2017.

18. Abbas, A. K.; Lichtman, A. H.; Pillai, S., *Basic Immunology E-Book: Functions and Disorders of the Immune System*. Elsevier Health Sciences: 2015.

19. O'Neill, L. A.; Bowie, A. G., *Nat Rev Immunol* **2007,** *7* (5), 353-64.

20. Awasthi, S., *Front Immunol* **2014,** *5*, 328.

21. Patel, P. S.; Buras, E. D.; Balasubramanyam, A., *J Obes* **2013,** *2013*, 616193.

22. Liu, T.; Zhang, L.; Joo, D.; Sun, S. C., *Signal Transduct Target Ther* **2017,** *2*.

23. Monroe, K. M.; McWhirter, S. M.; Vance, R. E., *Cell Microbiol* **2010,** *12* (7), 881-90.

24. Boxx, G. M.; Cheng, G., *Cell Host Microbe* **2016,** *19* (6), 760-9.

25. Kovarik, P.; Castiglia, V.; Ivin, M.; Ebner, F., *Front Immunol* **2016,** *7*, 652.

26. Trinchieri, G., *J Exp Med* **2010,** *207* (10), 2053-63.

27. Decker, T.; Muller, M.; Stockinger, S., *Nat Rev Immunol* **2005,** *5* (9), 675-87.

28. Shetab Boushehri, M. A.; Lamprecht, A., *Mol Pharm* **2018,** *15* (11), 4777-4800.

29. Wu, C. C.; MacCoss, M. J., *Curr Opin Mol Ther* **2002,** *4* (3), 242-50.

30. Ong, S. E.; Mann, M., *Nat Chem Biol* **2005,** *1* (5), 252-62.

31. Aebersold, R.; Mann, M., *Nature* **2003,** *422* (6928), 198-207.

32. Leon, I. R.; Schwammle, V.; Jensen, O. N.; Sprenger, R. R., *Mol Cell Proteomics* **2013,** *12* (10), 2992-3005.

33. Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M., *Anal Chem* **1996,** *68* (5), 850-8.

34. Weston, L. A.; Bauer, K. M.; Hummon, A. B., *Anal Methods* **2013,** *5* (18).

35. Glatter, T.; Ahrne, E.; Schmidt, A., *J Proteome Res* **2015,** *14* (11), 4472-85.

36. Mostovenko, E.; Hassan, C.; Rattke, J.; Deelder, A. M.; van Veelen, P. A.; Palmblad, M. J. E. O. P., **2013,** *1*, 30-37.

37. Yuseff, M. I.; Lankar, D.; Lennon-Dumenil, A. M., *Traffic* **2009,** *10* (6), 629-36.

38. Navratil, A. M.; Bliss, S. P.; Roberson, M. S., *Brain Res* **2010,** *1364*, 53-61.

39. Shi, H.; Hua, X.; Kong, D.; Stein, D.; Hua, F., *Neuropharmacology* **2019,** *145* (Pt B), 259-267.

40. Bjork, S.; Hurt, C. M.; Ho, V. K.; Angelotti, T., *PLoS One* **2013,** *8* (10), e76366.

41. Molina, J. R.; Morales, F. C.; Hayashi, Y.; Aldape, K. D.; Georgescu, M. M., *Cancer Res* **2010,** *70* (17), 6697-703.

42. Sloves, P. J.; Delhaye, S.; Mouveaux, T.; Werkmeister, E.; Slomianny, C.; Hovasse, A.; Dilezitoko Alayi, T.; Callebaut, I.; Gaji, R. Y.; Schaeffer-Reiss, C.; Van Dorsselear, A.; Carruthers, V. B.; Tomavo, S., *Cell Host Microbe* **2012,** *11* (5), 515-27.

43. Saito, S.; Matsui, H.; Kawano, M.; Kumagai, K.; Tomishige, N.; Hanada, K.; Echigo, S.; Tamura, S.; Kobayashi, T., *J Biol Chem* **2008,** *283* (10), 6584-93.

44. Kelly, R. B., *Nature* **1987,** *326* (6108), 14-5.

45. Chowdhury, C.; Chun, S.; Pang, A.; Sawaya, M. R.; Sinha, S.; Yeates, T. O.; Bobik, T. A., *Proc Natl Acad Sci U S A* **2015,** *112* (10), 2990-5.

46. Lim, K. H.; Staudt, L. M., *Cold Spring Harb Perspect Biol* **2013,** *5* (1), a011247.

47. Viswanathan, K.; Verweij, M. C.; John, N.; Malouli, D.; Fruh, K., *PLoS One* **2017,** *12* (11), e0187899.

48. Stephenson, F. H., *Calculations for molecular biology and biotechnology*. Third ed.; Academic press: 2016; p 50.

49. Zilberstein, G.; Korol, L.; Antonioli, P.; Righetti, P. G.; Bukshpan, S., *Anal Chem* **2007,** *79* (3), 821-7.

50. Fridriksson, E. K.; Baird, B.; McLafferty, F. W., *J Am Soc Mass Spectrom* **1999,** *10* (5), 453-5.

51. Rundlett, K. L.; Armstrong, D. W., *Anal Chem* **1996,** *68* (19), 3493-7.

52. Loo, R. R.; Dales, N.; Andrews, P. C., *Protein Sci* **1994,** *3* (11), 1975-83.

53. Clarke, N. J.; Li, F.; Tomlinson, A. J.; Naylor, S., *J Am Soc Mass Spectrom* **1998,** *9* (1), 88-91.

54. Sweadner, K. J., *Anal Biochem* **1991,** *194* (1), 130-5.

55. Bhattacharyya, D.; Basu, A.; Sen, P. C., *Indian J Biochem Biophys* **2007,** *44* (2), 122-5.

56. Jeno, P.; Scherer, P. E.; Manning-Krieg, U.; Horst, M., *Anal Biochem* **1993,** *215* (2), 292-8.

57. Li, F.; Dong, M.; Miller, L. J.; Naylor, S., *Rapid Commun Mass Spectrom* **1999,** *13* (5), 464-5.

58. Fic, E.; Kedracka-Krok, S.; Jankowska, U.; Pirog, A.; Dziedzicka-Wasylewska, M. J. E., **2010,** *31* (21), 3573-3579.

59. Zhou, J. Y.; Dann, G. P.; Shi, T.; Wang, L.; Gao, X.; Su, D.; Nicora, C. D.; Shukla, A. K.; Moore, R. J.; Liu, T.; Camp, D. G., 2nd; Smith, R. D.; Qian, W. J., *Anal Chem* **2012,** *84* (6), 2862-7.

60. Antharavally, B. S.; Mallia, K. A.; Rosenblatt, M. M.; Salunkhe, A. M.; Rogers, J. C.; Haney, P.; Haghdoost, N., *Anal Biochem* **2011,** *416* (1), 39-44.

61. Wisniewski, J. R.; Zougman, A.; Mann, M., *J Proteome Res* **2009,** *8* (12), 5674-8.

62. Erde, J.; Loo, R. R.; Loo, J. A., *J Proteome Res* **2014,** *13* (4), 1885-95.

63. Ciborowski, P.; Silberring, J., *Proteomic profiling and analytical chemistry: the crossroads*. Second Edition ed.; Elsevier: 2016; p 51-61.

64. Fic, E.; Kedracka-Krok, S.; Jankowska, U.; Pirog, A.; Dziedzicka-Wasylewska, M., *Electrophoresis* **2010,** *31* (21), 3573-9.

65. Feist, P.; Hummon, A. B., *Int J Mol Sci* **2015,** *16* (2), 3537-63.

66. Ferro, M.; Seigneurin-Berny, D.; Rolland, N.; Chapel, A.; Salvi, D.; Garin, J.; Joyard, J., *Electrophoresis* **2000,** *21* (16), 3517-26.

67. Jiang, L.; He, L.; Fountoulakis, M., *J Chromatogr A* **2004,** *1023* (2), 317-20.

68. Doucette, A. A.; Vieira, D. B.; Orton, D. J.; Wall, M. J., *J Proteome Res* **2014,** *13* (12), 6001-12.

69. Masuda, T.; Tomita, M.; Ishihama, Y., *J Proteome Res* **2008,** *7* (2), 731-40.

70. Coleman, O.; Henry, M.; Clynes, M.; Meleady, P., *Methods Mol Biol* **2017,** *1603*, 187-194.

71. Erde, J.; Loo, R. R.; Loo, J. A., *Methods Mol Biol* **2017,** *1550*, 11-18.

72. Wisniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M., *Nat Methods* **2009,** *6* (5), 359-62.

73. Liebler, D. C.; Ham, A. J., *Nat Methods* **2009,** *6* (11), 785; author reply 785-6.

74. Kamal, A. H.; Komatsu, S., *J Proteome Res* **2015,** *14* (5), 2219-36.

75. Botelho, D.; Wall, M. J.; Vieira, D. B.; Fitzsimmons, S.; Liu, F.; Doucette, A., *J Proteome Res* **2010,** *9* (6), 2863-70.

76. Mi, H.; Muruganujan, A.; Casagrande, J. T.; Thomas, P. D., *Nat Protoc* **2013,** *8* (8), 1551-66.

77. Mosser, D. M.; Edwards, J. P., *Nat Rev Immunol* **2008,** *8* (12), 958-69.

78. Gordon, S., *Nat Rev Immunol* **2003,** *3* (1), 23-35.

79. Najar, M.; Krayem, M.; Meuleman, N.; Bron, D.; Lagneaux, L., *Immune Netw* **2017,** *17* (2), 89-102.

80. Arango Duque, G.; Descoteaux, A., *Front Immunol* **2014,** *5*, 491.

81. Parker, L. C.; Prince, L. R.; Sabroe, I., *Clin Exp Immunol* **2007,** *147* (2), 199-207.

82. Martinez, F. O.; Helming, L.; Gordon, S., *Annu Rev Immunol* **2009,** *27*, 451-83.

83. Iwasaki, A.; Medzhitov, R., *Nat Immunol* **2004,** *5* (10), 987-95.

84. Krieg, A. M.; Vollmer, J., *Immunol Rev* **2007,** *220*, 251-69.

85. Patidar, A.; Selvaraj, S.; Sarode, A.; Chauhan, P.; Chattopadhyay, D.; Saha, B., *Cytokine* **2018,** *104*, 114-123.

86. Lauzon, N. M.; Mian, F.; Ashkar, A. A., *Adv Exp Med Biol* **2007,** *598*, 1-11.

87. Takeda, K.; Akira, S., *Int Immunol* **2005,** *17* (1), 1-14.

88. Park, B. S.; Song, D. H.; Kim, H. M.; Choi, B. S.; Lee, H.; Lee, J. O., *Nature* **2009,** *458* (7242), 1191-5.

89. Kim, H. M.; Park, B. S.; Kim, J. I.; Kim, S. E.; Lee, J.; Oh, S. C.; Enkhbayar, P.; Matsushima, N.; Lee, H.; Yoo, O. J.; Lee, J. O., *Cell* **2007,** *130* (5), 906-17.

90. Kawai, T.; Akira, S., *Immunity* **2011,** *34* (5), 637-50.

91. Beutler, B., *Curr Opin Microbiol* **2000,** *3* (1), 23-8.

92. Tarasova, N. K.; Ytterberg, A. J.; Lundberg, K.; Zhang, X. M.; Harris, R. A.; Zubarev, R. A., *J Proteome Res* **2015,** *14* (9), 3940-7.

93. Zhang, X.; Kuramitsu, Y.; Fujimoto, M.; Hayashi, E.; Yuan, X.; Nakamura, K., *Electrophoresis* **2006,** *27* (8), 1659-68.

94. Du, R.; Long, J.; Yao, J.; Dong, Y.; Yang, X.; Tang, S.; Zuo, S.; He, Y.; Chen, X., *J Proteome Res* **2010,** *9* (4), 1805-21.

95. Hartlova, A.; Link, M.; Balounova, J.; Benesova, M.; Resch, U.; Straskova, A.; Sobol, M.; Philimonenko, A.; Hozak, P.; Krocova, Z.; Gekara, N.; Filipp, D.; Stulik, J., *J Proteome Res* **2014,** *13* (2), 796-804.

96. Dhungana, S.; Merrick, B. A.; Tomer, K. B.; Fessler, M. B., *Mol Cell Proteomics* **2009,** *8* (1), 201-13.

97. Chowdhury, S. M.; Zhu, X.; Aloor, J. J.; Azzam, K. M.; Gabor, K. A.; Ge, W.; Addo, K. A.; Tomer, K. B.; Parks, J. S.; Fessler, M. B., *Mol Cell Proteomics* **2015,** *14* (7), 1859-70.

98. Patel, P. C.; Fisher, K. H.; Yang, E. C.; Deane, C. M.; Harrison, R. E., *Mol Cell Proteomics* **2009,** *8* (11), 2500-14.

99. Shi, L.; Chowdhury, S. M.; Smallwood, H. S.; Yoon, H.; Mottaz-Brewer, H. M.; Norbeck, A. D.; McDermott, J. E.; Clauss, T. R.; Heffron, F.; Smith, R. D.; Adkins, J. N., *Infect Immun* **2009,** *77* (8), 3227-33.

100. Swearingen, K. E.; Loomis, W. P.; Zheng, M.; Cookson, B. T.; Dovichi, N. J., *J Proteome Res* **2010,** *9* (5), 2412-21.

101. Yilmaz, A.; Reiss, C.; Tantawi, O.; Weng, A.; Stumpf, C.; Raaz, D.; Ludwig, J.; Berger, T.; Steinkasserer, A.; Daniel, W. G.; Garlichs, C. D., *Atherosclerosis* **2004,** *172* (1), 85-93.

102. Aktas, O.; Waiczies, S.; Smorodchenko, A.; Dorr, J.; Seeger, B.; Prozorovski, T.; Sallach, S.; Endres, M.; Brocke, S.; Nitsch, R.; Zipp, F., *J Exp Med* **2003,** *197* (6), 725-33.

103. Youssef, S.; Stuve, O.; Patarroyo, J. C.; Ruiz, P. J.; Radosevich, J. L.; Hur, E. M.; Bravo, M.; Mitchell, D. J.; Sobel, R. A.; Steinman, L.; Zamvil, S. S., *Nature* **2002,** *420* (6911), 78-84.

104. Jung, K. H.; Chu, K.; Jeong, S. W.; Han, S. Y.; Lee, S. T.; Kim, J. Y.; Kim, M.; Roh, J. K., *Stroke* **2004,** *35* (7), 1744-9.

105. Merx, M. W.; Liehn, E. A.; Janssens, U.; Lutticken, R.; Schrader, J.; Hanrath, P.; Weber, C., *Circulation* **2004,** *109* (21), 2560-5.

106. Patel, T. R.; Corbett, S. A., *J Surg Res* **2004,** *116* (1), 116-20.

107. Kamal, A. H. M.; Fessler, M. B.; Chowdhury, S. M., *PLoS One* **2018,** *13* (2), e0193104.

108. Kim, Y.; Lee, E. J.; Jang, H. K.; Kim, C. H.; Kim, D. G.; Han, J. H.; Park, S. M., *J Gastroenterol Hepatol* **2016,** *31* (6), 1220-8.

109. Zhang, J.; Osawa, S.; Takayanagi, Y.; Ikuma, M.; Yamada, T.; Sugimoto, M.; Furuta, T.; Miyajima, H.; Sugimoto, K., *Cytokine* **2013,** *61* (2), 540-5.

110. Chung, H. K.; Lee, I. K.; Kang, H.; Suh, J. M.; Kim, H.; Park, K. C.; Kim, D. W.; Kim, Y. K.; Ro, H. K.; Shong, M., *Exp Mol Med* **2002,** *34* (6), 451-61.

111. Yi, T.; Rao, D. A.; Tang, P. C.; Wang, Y.; Cuchara, L. A.; Bothwell, A. L.; Colangelo, C. M.; Tellides, G.; Pober, J. S.; Lorber, M. I., *Transplantation* **2008,** *86* (5), 719-27.

112. Becker, L.; Gharib, S. A.; Irwin, A. D.; Wijsman, E.; Vaisar, T.; Oram, J. F.; Heinecke, J. W., *Cell Metab* **2010,** *11* (2), 125-35.

113. Gordon, S. M.; McKenzie, B.; Kemeh, G.; Sampson, M.; Perl, S.; Young, N. S.; Fessler, M. B.; Remaley, A. T., *Mol Cell Proteomics* **2015,** *14* (12), 3247-57.

114. Kamal, A. H. M.; Chakrabarty, J. K.; Udden, S. M. N.; Zaki, M. H.; Chowdhury, S. M., *Sci Rep* **2018,** *8* (1), 164.

115. Ewing, R. M.; Chu, P.; Elisma, F.; Li, H.; Taylor, P.; Climie, S.; McBroom-Cerajewski, L.; Robinson, M. D.; O'Connor, L.; Li, M.; Taylor, R.; Dharsee, M.; Ho, Y.; Heilbut, A.; Moore, L.; Zhang, S.; Ornatsky, O.; Bukhman, Y. V.; Ethier, M.; Sheng, Y.; Vasilescu, J.; Abu-Farha, M.; Lambert, J. P.; Duewel, H. S.; Stewart, II; Kuehl, B.; Hogue, K.; Colwill, K.; Gladwish, K.; Muskat, B.; Kinach, R.; Adams, S. L.; Moran, M. F.; Morin, G. B.; Topaloglou, T.; Figeys, D., *Mol Syst Biol* **2007,** *3*, 89.

116. Stelzl, U.; Worm, U.; Lalowski, M.; Haenig, C.; Brembeck, F. H.; Goehler, H.; Stroedicke, M.; Zenkner, M.; Schoenherr, A.; Koeppen, S.; Timm, J.; Mintzlaff, S.; Abraham, C.; Bock, N.; Kietzmann, S.; Goedde, A.; Toksoz, E.; Droege, A.; Krobitsch, S.; Korn, B.; Birchmeier, W.; Lehrach, H.; Wanker, E. E., *Cell* **2005,** *122* (6), 957-68.

117. Rual, J. F.; Venkatesan, K.; Hao, T.; Hirozane-Kishikawa, T.; Dricot, A.; Li, N.; Berriz, G. F.; Gibbons, F. D.; Dreze, M.; Ayivi-Guedehoussou, N.; Klitgord, N.; Simon, C.; Boxem, M.; Milstein, S.; Rosenberg, J.; Goldberg, D. S.; Zhang, L. V.; Wong, S. L.; Franklin, G.; Li, S.; Albala, J. S.; Lim, J.; Fraughton, C.; Llamosas, E.; Cevik, S.; Bex,

C.; Lamesch, P.; Sikorski, R. S.; Vandenhaute, J.; Zoghbi, H. Y.; Smolyar, A.; Bosak, S.; Sequerra, R.; Doucette-Stamm, L.; Cusick, M. E.; Hill, D. E.; Roth, F. P.; Vidal, M., *Nature* **2005,** *437* (7062), 1173-8.

118. Lambert, J. P.; Picaud, S.; Fujisawa, T.; Hou, H.; Savitsky, P.; Uuskula-Reimand, L.; Gupta, G. D.; Abdouni, H.; Lin, Z. Y.; Tucholska, M.; Knight, J. D. R.; Gonzalez-Badillo, B.; St-Denis, N.; Newman, J. A.; Stucki, M.; Pelletier, L.; Bandeira, N.; Wilson, M. D.; Filippakopoulos, P.; Gingras, A. C., *Mol Cell* **2019,** *73* (3), 621-638 e17.

119. Chowdhury, S. M.; Shi, L.; Yoon, H.; Ansong, C.; Rommereim, L. M.; Norbeck, A. D.; Auberry, K. J.; Moore, R. J.; Adkins, J. N.; Heffron, F.; Smith, R. D., *J Proteome Res* **2009,** *8* (3), 1504-14.

120. Yu, C.; Huang, L., *Anal Chem* **2018,** *90* (1), 144-165.

121. Chakrabarty, J. K.; Naik, A. G.; Fessler, M. B.; Munske, G. R.; Chowdhury, S. M., *Anal Chem* **2016,** *88* (20), 10215-10222.

122. Schmidlin, T.; Garrigues, L.; Lane, C. S.; Mulder, T. C.; van Doorn, S.; Post, H.; de Graaf, E. L.; Lemeer, S.; Heck, A. J.; Altelaar, A. F., *Proteomics* **2016,** *16* (15-16), 2193-205.

123. Saeed, A. I.; Sharov, V.; White, J.; Li, J.; Liang, W.; Bhagabati, N.; Braisted, J.; Klapa, M.; Currier, T.; Thiagarajan, M.; Sturn, A.; Snuffin, M.; Rezantsev, A.; Popov, D.; Ryltsov, A.; Kostukovich, E.; Borisovsky, I.; Liu, Z.; Vinsavich, A.; Trush, V.; Quackenbush, J., *Biotechniques* **2003,** *34* (2), 374-8.

124. Ellis, N. A.; Glazer, A. M.; Donde, N. N.; Cleves, P. A.; Agoglia, R. M.; Miller, C. T., *Development* **2015,** *142* (14), 2442-51.

125. Mi, H.; Poudel, S.; Muruganujan, A.; Casagrande, J. T.; Thomas, P. D., *Nucleic Acids Res* **2016,** *44* (D1), D336-42.

126. Kramer, A.; Green, J.; Pollard, J., Jr.; Tugendreich, S., *Bioinformatics* **2014,** *30* (4), 523-30.

127. Ooi, A.; Wong, A.; Esau, L.; Lemtiri-Chlieh, F.; Gehring, C., *Front Physiol* **2016,** *7*, 300.

128. Dietmair, S.; Hodson, M. P.; Quek, L. E.; Timmins, N. E.; Gray, P.; Nielsen, L. K., *PLoS One* **2012,** *7* (8), e43394.

129. He, B.; Soderlund, D. M., *Neurosci Lett* **2010,** *469* (2), 268-72.

130. Kamal, A. H. M.; Aloor, J. J.; Fessler, M. B.; Chowdhury, S. M., *Mol Cell Proteomics* **2019**.

131. Aloor, J. J.; Azzam, K. M.; Guardiola, J. J.; Gowdy, K. M.; Madenspacher, J. H.; Gabor, K. A.; Mueller, G. A.; Lin, W. C.; Lowe, J. M.; Gruzdev, A.; Henderson, M. W.; Draper, D. W.; Merrick, B. A.; Fessler, M. B., *J Biol Chem* **2019,** *294* (6), 1997-2008.

132. Benes, P.; Maceckova, V.; Zdrahal, Z.; Konecna, H.; Zahradnickova, E.; Muzik, J.; Smarda, J., *Differentiation* **2006,** *74* (6), 265-76.

133. Bandaru, S.; Ala, C.; Salimi, R.; Akula, M. K.; Ekstrand, M.; Devarakonda, S.; Karlsson, J.; Van den Eynden, J.; Bergstrom, G.; Larsson, E.; Levin, M.; Boren, J.; Bergo, M. O.; Akyurek, L. M., *Circulation* **2019**.

134. Tu, Y.; Zhang, L.; Tong, L.; Wang, Y.; Zhang, S.; Wang, R.; Li, L.; Wang, Z., *Int Immunopharmacol* **2018,** *55*, 263-271.

135. Talving, P.; Karamanos, E.; Skiada, D.; Lam, L.; Teixeira, P. G.; Inaba, K.; Johnson, J.; Demetriades, D., *J Trauma Acute Care Surg* **2013,** *74* (3), 912-6.

136. Gao, Y. Y.; Zhang, D. F.; Li, H.; Liu, R.; Zhuang, Z. H.; Li, Q. F.; Wang, S. Y.; Peng, X. X., *J Proteome Res* **2007,** *6* (11), 4321-9.

137. Wyss, M.; Kaddurah-Daouk, R., *Physiol Rev* **2000,** *80* (3), 1107-213.

138. Li, J.; Aderem, A., *Cell* **1992,** *70* (5), 791-801.

139. Zhu, Z.; Bao, Z.; Li, J., *J Biol Chem* **1995,** *270* (30), 17652-5.

140. Li, J.; Zhu, Z.; Bao, Z., *J Biol Chem* **1996,** *271* (22), 12985-90.

141. Yue, L.; Lu, S.; Garces, J.; Jin, T.; Li, J., *J Biol Chem* **2000,** *275* (31), 23948-56.

142. Myat, M. M.; Chang, S.; Rodriguez-Boulan, E.; Aderem, A., *Curr Biol* **1998,** *8* (12), 677-83.

143. Huling, J. C.; Pisitkun, T.; Song, J. H.; Yu, M. J.; Hoffert, J. D.; Knepper, M. A., *Am J Physiol Renal Physiol* **2012,** *302* (4), F401-7.

Vita

ADA Shahinuzzaman

The author was born in Dhaka, Bangladesh and completed his early schooling at the University of Dhaka with a Bachelor and Master in Genetic Engineering and Biotechnology in Nov 2011. During his master program back in Bangladesh, he explored genomics science while working the mitochondrial DNA sequencing from the mastectomy samples of breast cancer patients. Soon after graduation, He joined as a research assistant at the Genetic Engineering and Biotechnology Research Laboratory in Center for Advanced Research and Sciences (CARS) in University of Dhaka. In 2013 February, he joined Bangladesh Council of Scientific and Industrial Research **(**BCSIR**)** as a scientific officer and got exposed to drug discovery research. In 2014 August, he took a sabbatical leave to come to UT Arlington to pursue his Ph.D. in Analytical chemistry with Dr. Saiful M Chowdhury. His research in the University of Texas At Arlington was on Improved sample preparation method for LC-MS/MS based cellular proteomics and co-immunoprecipitation cross-linking proteomics study of Toll-like receptor 4. He worked in Shimadzu Center for Advanced Analytical Chemistry (SCAAC) in Arlington, TX as a summer intern. After his Ph.D., he will do Pharmaceutical sciences research in his work station at BCSIR.