

**DEVELOPMENT OF CHEMICAL PROTEOMIC APPROACHES TO  
STUDY VIRAL ENDOCYTOSIS AND PHOSPHOPROTEOMICS**

by

**Mayank Srivastava**

**A Dissertation**

*Submitted to the Faculty of Purdue University*

*In Partial Fulfillment of the Requirements for the degree of*

**Doctor of Philosophy**



Department of Chemistry

West Lafayette, Indiana

August 2019

**THE PURDUE UNIVERSITY GRADUATE SCHOOL**  
**STATEMENT OF COMMITTEE APPROVAL**

Dr. W. Andy Tao, Chair

Department of Chemistry

Dr. David Thompson

Department of Chemistry

Dr. Hilikka Kenttamaa

Department of Chemistry

Dr. Richard Kuhn

Department of Biological Sciences

**Approved by:**

Dr. Christine Hrycyna

Head of the Graduate Program

*This thesis is dedicated to my parents, who have made much bigger sacrifices for my PhD, and  
have been my role models since my childhood*

## ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. W. Andy Tao, for all his support throughout my graduate studies. I have seen him as a motivation for the great ideas he has as a scientist. I would take this moment to thank him for giving me a project that has inspired me every day and allowed me to explore my capabilities. I have had the opportunities to improve my skills in different research areas ranging from synthesis, mass spectrometry, to virology. I would also like to express my deepest sense of gratitude to Dr. Richard Kuhn, for the resources and his valuable feedbacks. I express my sincere thanks to my committee members Dr. David Thompson and Dr. Hilikka Kenttamaa, for their support and guidance. It was only because of the above advisors that my graduate study was possible.

I also need to acknowledge Dr. Devika Sirohi, Andrew Miller, Dr. Ying Zhang, and Dr. Linna Wang for their guidance and help in my Zika endocytosis and DiGEP projects. Special thanks to Andrew Miller for providing me the precious Zika virus samples, and Dr. Devika Sirohi for providing me the primers and the standard curve for qRT-PCR. Off course, I need to thank my lab mates who have treated me more like a family and a friend. Lastly, my parents Mr. Anil Kumar Srivastava and Mrs. Sarla Srivastava, who have inspired me to achieve more and taught me to live the life with integrity. My parents have let me pursue my goals and have made endless sacrifices for my sake. I dedicate my Ph.D. to my parents.

## TABLE OF CONTENTS

LIST OF FIGURES .....	8
LIST OF ABBREVIATIONS.....	11
ABSTRACT.....	12
CHAPTER 1. TRACING ENDOCYTOSIS BY MASS SPECTROMETRY .....	13
1.1 Introduction.....	13
1.2 Clathrin-mediated Endocytosis.....	13
1.2.1 Formation of Clathrin-Coated Vesicles .....	14
1.2.2 Vesicle Uncoating and Fusion with Endosomal Compartments .....	18
1.3 Mass Spectrometry as a Tool to Study Endocytosis.....	19
1.3.1 By Isolating Clathrin-coated Vesicles .....	19
1.3.2 By Using Chemical Proteomic Approaches .....	22
1.3.2.1 Receptor Identification by Ligand-Receptor Capture (LRC) Technology .....	22
1.3.2.2 Employing Functionalized Dendrimer .....	22
1.4 Conclusion and Future Directions .....	23
1.5 References.....	24
CHAPTER 2. MAPPING THE EARLY STAGE ENTRY OF ZIKA VIRUS INTO HOST CELLS .....	32
2.1 Summary.....	32
2.2 Introduction.....	32
2.3 Experimental Procedures .....	34
2.3.1 Materials .....	34
2.3.2 Mature Zika Preparation .....	35
2.3.3 Plaque Assay.....	35
2.3.4 Synthesis and Purification of Labeling Reagent.....	35
2.3.5 Virus Labeling .....	36
2.3.6 Virus Infection and Crosslinking of Host-Proteins .....	37
2.3.7 Sample Preparation for LC-MS Analysis .....	37
2.3.8 LC-MS/MS Analysis .....	38
2.3.9 Statistical Analysis.....	38

2.3.10	Antibody Inhibition of Virus Infection .....	39
2.3.11	RNA Extraction and RT-qPCR .....	39
2.3.12	Overexpression of NCAM in HEK 293T cells .....	39
2.3.13	Immunofluorescence .....	39
2.3.14	Western Blot.....	40
2.4	Results and Discussion .....	40
2.4.1	Design and Synthesis of Labeling Reagent .....	40
2.4.2	Labeling of a standard protein .....	41
2.4.3	Labeling of Zika E proteins .....	41
2.4.4	Infectivity testing of labeled Zika Virus .....	41
2.4.5	Identification of Zika receptors and other host factors/ direct interactors with Zika virus .....	42
2.4.6	NCAM is a receptor for Zika virus.....	44
2.5	References.....	45
CHAPTER 3. DEVELOPMENT OF DIGE (DIFFERENCE GEL ELECTROPHORESIS) FOR PHOSPHOPROTEOMICS .....		64
3.1	Introduction.....	64
3.1.1	Steps for 2-Dimensional Gel Electrophoresis (2-DE): .....	64
3.1.1.1	Sample Preparation.....	64
3.1.1.2	Isoelectric Focusing.....	65
3.1.1.3	Strip Equilibration .....	66
3.1.1.4	SDS-PAGE .....	66
3.1.1.5	Protein Detection and Identification.....	66
3.1.2	DiGE (Difference Gel Electrophoresis).....	67
3.1.2.1	Advantages of DiGE.....	68
3.1.2.2	DiGE Dyes:.....	68
3.1.3	DiGE in Phosphoproteomics and Goal of the study .....	69
3.2	Methods and Results .....	70
3.2.1	Design of Novel DiGE-P Reagent.....	70
3.2.2	Synthesis of Reagents .....	71
3.2.3	Phosphoprotein Labeling.....	72

3.2.3.1	Detection of Phosphoproteins in a Standard Protein Mixture .....	72
3.2.3.2	Labeling of Phosphoproteins in Complex Mixtures .....	72
3.2.3.3	Quantitative ability of DiGEP .....	73
3.3	Conclusion and Future Work .....	73
3.4	References .....	73

## LIST OF FIGURES

- Figure 1.1: Diagram showing the phases of endocytosis: Ligand binding, Coat proteins recruitment, Coat assembly, Scission leading to release of clathrin coated vesicle, Uncoating, and Fusion with early endosome. .... 29
- Figure 1.2: (A) Schematic representation of the functionalized dendrimer. (B) Experimental workflow for TITAN analysis. .... 30
- Figure 1.3: Spatiotemporal information for dendrimer-interacting proteins involved in the clathrin- and caveolar/raft-mediated endocytic pathways revealed by TITAN analysis..... 31
- Figure 2.1: Trifunctionalized reagent designed to label Zika ‘surface’ E proteins and capture virus-interacting proteins. (A) Design of Virus labeling reagent. Maleimide reacts with available cysteines on the virus surface under mild conditions, diazirine enables crosslinking proteins at fixed time points allowing tracing virus movement in ‘real-time’, and biotin acts as a handle for protein enrichment and downstream analysis. The three functionalities are separated by membrane-impermeable polyethylene glycol (PEG)- like linker offering the flexibility required for capturing interacting proteins. (B) Labeling of Zika E proteins. Purified zika virus was diluted in PBS pH 7 and rotated overnight in 4 °C with the labeling reagent. Reaction was quenched with three-fold excess cysteine for 1 hour. (C) Design and workflow for capturing virus receptor and mapping its entry. Labeled zika was diluted in DMEM and incubated with confluent cells for 1 hour at 4 °C. Additionally, cells were incubated in 37 °C for fixed time points to allow virus entry. Unbound virus was removed and cells were directly exposed to UV light, following which cells were harvested by scraping. Cells were lysed, and crosslinked proteins were captured on the neutravidin beads. Proteins were digested on-beads, using sequential Lys-C and Trypsin digestion, and analyzed by LC-MS/MS on an Orbitrap Velos Pro. Quantitation was performed using MaxQuant, and proteins identified uniquely in the UV compared to control, with FDR of 1%, were considered as crosslinked proteins..... 49
- Figure 2.2: Proteins crosslinked at different time points of infection. (A) Heat map for proteins crosslinked at 0, 4, and 8 minutes of infection (FDR 0.05) normalized to the control. Proteins in the significant region with ratios of 2.5 and higher for the UV at different time points compared to the control, are considered as crosslinked proteins and used for further analysis. (B) Principal Component Analysis, Cellular Component Distribution, and number of proteins implicated in viral infection (by Ingenuity Pathway Analysis) at different time point of infection..... 51
- Figure 2.3: NCAM is a receptor for Zika Virus (A) Anti-NCAM inhibits Zika infection of Vero cells. Immunofluorescence showing the binding of anti-NCAM antibody to Vero cells. After antibody inhibition of NCAM, about 50% reduction in viral infectivity

was observed. (B) Overexpression of NCAM1 in HEK293T increases Zika infection. Expression plasmid with NCAM1 was transfected into HEK293T cells. 48 hours after transfection, immunofluorescence was used to confirm surface expression of NCAM. qRT-PCR revealed an increase in Zika RNAs after NCAM overexpression. .... 53

Figure 2.4: Proposed Endocytosis pathway of Zika entry to host cells, highlighting proteins crosslinked at different time point. .... 55

Figure 2.5: Protein labeling by reagent. (A) Standard protein Bovine Serum Albumin (BSA) was labeled with the reagent in phosphate buffer pH 7 overnight, and the reaction was quenched with excess cysteine for 1 hour. The labeled proteins were enriched on streptavidin beads, followed by three 1% SDS washes and elution by boiling. The reagent labeling efficiency was quantified by comparing the protein band intensity with the input. A no-reagent control was employed to account for non-specific binders. In parallel, the BSA was reduced and alkylated prior to labeling, to confirm the labeling site on the protein by the maleimide-diazirine-biotin reagent. (B) Purified zika virus was labeled by the reagent using the same protocol as above. (C) The infectivity of labeled virus was tested by plaque assay. The number of plaque forming units (pfu) after virus labeling were compared with the unlabeled virus... 56

Figure 2.6: Synthesis, HPLC purification, and Characterization of the labeling reagent. (A) Synthesis was performed on the solid support, Rink-amide-AM resin, and cleaved from the resin using 95% TFA. (B and C)  $^1\text{H}$  NMR,  $^{13}\text{C}$  NMR and MALDI characterizations of Maleimide-Biotin and Maleimide-Biotin-Diazirine respectively. .... 58

Figure 2.7: Mass spectrometric identification of peptides from E protein of Zika..... 63

Figure 3.1: Principle and Steps of 2-DE.<sup>1</sup>..... 76

Figure 3.2: General Representation of DiGE.<sup>21</sup> ..... 76

Figure 3.3: Principle of 2D DiGE.<sup>22</sup> ..... 77

Figure 3.4: Structures of cyanine dyes used in DiGE.<sup>19,20</sup> ..... 77

Figure 3.5: DiGE-P Reagent (A= schematic representation, B= An example of actual DiGE-P reagent). .... 78

Figure 3.6: Synthetic Route for DiGE-P..... 78

Figure 3.7: MALDI for compound 2 (A) and 3 (B)..... 79

Figure 3.8: Another DiGE-P reagent and its MALDI.  $(M - \text{N}_2 + \text{H})^+$  peak observed at m/z at 1018.40..... 80

- Figure 3.9: Optimization of Phosphoprotein Labeling. Fluorescence image and Coomassie Blue stain of DiGE-P labeled 2  $\mu\text{g}$  four-protein mixture of standard phospho and non-phospho proteins (A), and 1  $\mu\text{g}$  phosphoproteins in complex mixture with human plasma (B)..... 81
- Figure 3.10: Selectivity (A) and Sensitivity (B) testing of DiGE-P using standard four-protein mixture. .... 82
- Figure 3.11: Selectivity testing of DiGE-P using 1  $\mu\text{g}$   $\alpha$ -casein spiked in 20  $\mu\text{g}$  human plasma. 83
- Figure 3.12: Fluorescence image and Coomassie Blue of DiGE-P (Cy3 and Cy5) labeled model phosphoprotein  $\alpha$ -casein in human plasma. .... 83

## LIST OF ABBREVIATIONS

ASKA	Analog-Sensitive Kinase Allele
DENV	Dengue Virus
EGFR	Epidermal Growth Factor Receptor
E-Syts	Extended-synaptotagmins
FDR	False discovery rate
FGFR	Fibroblast growth factors receptors
HCMV	Human cytomegalovirus
HHV-1	Herpes simplex virus 1
HHV-8	Herpes virus 8
JEV	Japanese encephalitis virus
GBS	Guillain-Barré syndrome
IPA	Ingenuity Pathway Analysis
ITGA3	Integrin alpha-3
HEK 293T	Human Embryonic Kidney Cells 293, transformed with large T antigen
NCAM1	Neural Cell Adhesion Molecule
vRNP	Viral ribonucleoprotein
WNV	West Nile virus
YFV	Yellow fever virus

## ABSTRACT

Author: Srivastava, Mayank,. PhD

Institution: Purdue University

Degree Received: August 2019

Title: Development of Chemical Proteomic Approaches to Study Viral Endocytosis and Phosphoproteomics.

Committee Chair: Weiguo Andy Tao

A significant development in mass spectrometry instrumentation and software in the past decade has led to its application in solving complex biological problems. One of the emerging areas is Chemical Proteomics that involves design and use of chemical reagents to probe protein functions in ‘a live cell’ environment. Another aspect of Chemical Proteomics is the identification of target proteins of a drug or small molecule. This is assisted by photoreactive groups, which on exposure to UV light, covalently link the target proteins that can be purified by affinity-based enrichment followed by mass-spectrometric identification. This phenomenon of Photoaffinity labeling (PAL) has been widely used in a broad range of applications. Herein, we have designed chemical tools to study Zika endocytosis and phosphoproteomics.

Zika virus has attracted the interest of researchers globally, following its outbreak in 2016. While a significant development has been made in understanding the structure and pathogenesis, the actual mechanism of Zika entry into host cells is largely unknown. We designed a chemical probe to tag the live virus, leading to the identification of the virus receptors and other host factors involved in viral entry. We further validated neural cell adhesion molecule (NCAM1) as a host protein involved in early phase entry of Zika virus into Vero cells.

The second aspect is the development of the DIGE (Difference Gel Electrophoresis) technology for phosphoproteomics. Phosphoproteins are known to be involved in various signaling pathways and implicated in multiple diseased states. We designed chemical reagents composed of titanium (IV) ion, diazirine and a fluorophore, to covalently label the phosphoproteins. Cyanine3 and cyanine5 fluorophores were employed to reveal the difference in phosphorylation between samples for the comparative proteomics. Thus far, we have successfully demonstrated the labeling of standard phosphoproteins in both simple and complex protein mixtures, and the future efforts are towards applying the technology to identify phosphoproteins in a cell lysate.

## CHAPTER 1. TRACING ENDOCYTOSIS BY MASS SPECTROMETRY

### 1.1 Introduction

Endocytosis can be described as a process by which a cell internalizes plasma membrane lipids, proteins, and extracellular materials via membrane-coated vesicles. The different endocytic pathways can be broadly classified into phagocytosis and pinocytosis. Phagocytosis is the engulfment of pathogens and other foreign agents by specialized certain types of cells like macrophages.<sup>1</sup> Pinocytosis is more general and can be observed in most of the mammalian cells, and it can be further divided into most widely studied clathrin-mediated endocytosis (CME), caveolae-mediated endocytosis, clathrin and caveolae-independent endocytosis, and micropinocytosis.<sup>2</sup> These pathways differ in the size of the internalized moiety, and composition of the invagination, as well as the environment inside the vesicles, and sequence of structures and proteins involved in the pathway. As CME is recognized as the most commonly used internalization mechanism by various ligands<sup>3</sup>, this chapter will focus on the details of CME, and the methods to map the pathway using mass spectrometry.

### 1.2 Clathrin-mediated Endocytosis

CME is the most extensively characterized cellular uptake mechanism and much progress has been made in deciphering the pathway for CME at the molecular level. Basically, CME requires the binding of ligands to specific receptors on the plasma membrane, to generate the signals required to initiate cellular uptake process. Some of the early examples of CME include clearance of lipoproteins from blood by induction of its internalization<sup>4</sup>, and antigen presentation on the cell surface by endocytosis of antigens for the activation of the immune system. The process is followed by recruitment of coat proteins to the membrane, and formation of specific coat at the cytoplasmic end of the plasma membrane. The coat is primarily composed of a cytoplasmic protein clathrin, which polymerizes at the plasma membrane only at the site of endocytosis. Clathrin assembly induces the membrane to bud inwards to form invaginations, which further grows to form clathrin-coated vesicles (CCVs) containing the ligand bound to the receptor. Eventually, the vesicle is pinched off from the membrane by another set of proteins led by a GTPase, dynamin, resulting in the internalization of the ligand-protein complex. The vesicle gets uncoated before

merging with different cellular compartments of the pathway, which hence decides the fate of the cargo (Figure 1).

### 1.2.1 Formation of Clathrin-Coated Vesicles

The forming of vesicle was observed using electron microscopy for the first time by Roth and Porter in 1964.<sup>5</sup> Using electron microscopy, researchers studied the uptake of yolk proteins by the oocyte of mosquito *Aedes aegypti* L. and observed the presence of pits on the surface of the oocytes, with a layer of 200 Å bristle-coat on the cytoplasmic face of the membrane. Structurally, these pits were found to be similar to the intracellular vesicles, suggesting that the vesicles were derived from the pits and represented a later stage of internalization. The vesicles closer to the membrane were about 140 - 150 μm in diameter, and the ones deeper into the cytoplasm were much larger in size and without the bristle-coat. This hints at the loss of the bristle-coat from the vesicles, which eventually fuse to deliver the yolk protein inside the oocyte.

Five years later, the coated vesicles were isolated from the nerve endings of guinea pig brain<sup>6</sup>, and Pearse further isolated them from pig brain in 1975<sup>7</sup>. Both of these studies were aimed at understanding the coat structure using electron microscopy. In the former, Kanaseki and Kadota described the vesicle as a basket like structure, composed of repeats of hexagons and pentagons, which was reconfirmed by Pearse. Besides the structure, Pearse also found that the coat is essentially made up of only one major protein with a very high molecular weight of about 180 kDa using gel electrophoresis. Also, the digestion of coated vesicles led to the loss of this protein band on the SDS gel, suggesting that the 180 kDa protein is specifically present on the outside of the vesicles (cytoplasmic face of membrane). Pearse named it as Clathrin. In later studies, purified coat proteins were further explored for their molecular arrangements. A purified clathrin exists as a trimer of heavy chains, each linked to a light chain.<sup>8,9</sup> The three heavy chains appear as three bent flexible legs diverging from a common point, hence called a triskelion. The heavy chains trimerize at a domain close to the C-terminus, while the light chains do not interact with each other.<sup>10</sup> Understanding the arrangement of clathrin units was challenging due to the heterogeneity of light chains. There are two types of light chains of clathrin (LC) in mammalian cells, LCa and LCb, which are randomly associated with the heavy chains in the triskelions, giving rise to four types of clathrin units- 3LCa, 2LCa-1LCb, 1LCa-2LCb, 3LCb. Moreover, the relative expressions of these light chains vary in different tissues. These light chains are considered to have regulatory

functions in vesicle coating and uncoating, as they possess the domains for phosphorylation, hsc70 (an uncoating protein) binding and calcium binding.<sup>11,12</sup> In 1992, through affinity purification of the fragments using the anti-clathrin antibody, Nathke *et al.* proved that the domains of clathrin heavy chains are involved in triskelion formation.<sup>10</sup> This was followed by another study on clathrin assembly in 1995 from the same group.<sup>13</sup> These studies indicated that the interacting domain (residues 1488-1587) of each heavy chain belongs to the carboxy-terminal third (residue 1074-1675), which also forms the 'hub' of the triskelion. The 'hub' is comprised of the vertex, the proximal portion of the legs extending to the bend region, and mediates all the important functions of the triskelion, including trimer formation, interaction with the light chains, and self-polymerization to form the cage.

Along with clathrin, there have been growing evidences of co-isolation of some other polypeptides or coat proteins. Another major protein associated with CCVs was identified as 'adaptor' or 'accessory' protein complex (AP complex) whose high molecular weight subunits of 100 kDa and 50 kDa were frequently identified in gel electrophoresis.<sup>8</sup> There are four different types of AP complexes: AP-1, AP-2<sup>14,15</sup>, AP-3<sup>16,17</sup>, and AP-4<sup>18,19</sup>, each containing four subunits. Three subunits are conserved and common to all, namely  $\sigma$  (20 kDa),  $\mu$  (50 kDa), and  $\beta$  (100 kDa). The fourth subunit of 100 kDa is characteristic to each AP complex, and is named after the Greek letters,  $\gamma$  for AP-1,  $\alpha$  for AP-2,  $\delta$  and  $\epsilon$  for AP-3 and AP-4, respectively. Even though the AP subunits have similar arrangement in the complexes, their localization inside the cell differs significantly. For example, out of the four, AP-1 and AP-2 are the only major AP complexes associated with CCVs. AP-1 plays an important role in *trans*-Golgi network (TGN), while AP-2 is critical in the cellular uptake process.<sup>14,15</sup> Both AP-1 and AP-2 were shown to initiate clathrin assembly and coat formation.<sup>20</sup> AP-2 is the primary focus here, as it is the most important adaptor protein complex involved in receptor-mediated endocytosis at plasma membrane.<sup>15</sup> AP-2 interacts with cytoplasmic domains of transmembrane receptor proteins and helps in the recruitment of cytosolic clathrin to the membrane. In fact, it acts as a linkage between the membrane and clathrin. This was confirmed by electron microscopy in which adaptor protein was found to be in between the vesicle membrane and the clathrin coat surrounding it.<sup>21</sup> The  $\alpha$  domain of AP-2 complexes with the membrane proteins and the  $\beta$  is on the side of clathrin. Even *in vitro*, AP-2 and clathrin were found to polymerize together to form a lattice similar to a vesicle coat, under physiological conditions.<sup>20,22</sup>

Besides the above two coat proteins, the involvement of a few more proteins has been identified in clathrin-coated vesicles such as dynamin, amphiphysin, synaptotagmin, AP180 or CALM5, synaptojanin, and receptor-specific arrestins, and Eps15. The next section will include the specific roles of some of these proteins.

Molecular mechanism for CCV formation: CME is highly complex and dynamic, but is highly regulated. A lot of understanding of the process resulted from the studies on synaptic-vesicle recycling in presynaptic neurons, and the proteins involved have isoforms commonly found in non-neuronal cells.<sup>23</sup> The formation of clathrin-coated vesicles can be broadly divided into three steps: Binding of the ligand to the receptor which initiates recruitment of coat proteins to the membrane, assembly of coat proteins, and budding of vesicle that result into its release into the cytoplasm. The vesicle then gets uncoated in an ATP-dependent process by a heat shock protein cognate, hsc70, in combination with auxilin, which is also reported to have a role in clathrin assembly.<sup>24,25</sup> The complexity of recruitment and assembly processes at the membrane require involvement of several other proteins which are discussed below.

A number of molecules have been identified in the initial recruitment of coat proteins, primarily AP-2 complex, to the membrane. Phosphoinositides are membrane phospholipids that interact with N-terminal region of the  $\alpha$  subunit of AP-2 with high affinity, and thus mediate its trafficking to the membrane.<sup>26,27</sup> In a similar context, protein phosphatases like phospholipase D (PLD) and synaptojanin 1 were also shown to play a role in AP-2 recruitment. West *et al.* studied the effect of PLD on the adaptor recruitment, and found that the addition of exogenous PLD stimulates the AP-2 recruitment, while Neomycin, an inhibitor of endogenous PLD activity, inhibits both endosomal and plasma membrane recruitment of AP-2.<sup>28</sup> Similarly, synaptojanin was also proposed to mediate the interaction between coat proteins and plasma membrane. Another protein, Synaptotagmin 1, was shown to be a key protein required for the 'high affinity' binding of AP-2 to plasma membrane.<sup>29</sup> In fact, it was considered as a docking site for AP-2. Synaptotagmin is a synaptic vesicle protein, the binding of AP-2 with synaptotagmin is stimulated by the tyrosine-rich endocytic motifs of cargo proteins.<sup>30</sup>

Besides coat protein recruitment to the membrane, a critical piece of information lies in the cytoplasmic domain of the receptor proteins. The sorting signals present on the cytoplasmic domain determine the selectivity of cargo. Some of the most widely studied sorting signals are rich in tyrosine<sup>31</sup> and C-terminal dileucine<sup>32</sup>. The sequence YXXO, where X is any amino acid and O

is an amino acid with bulky hydrophobic side chain, is recognized by the AP-2 via its  $\mu$  subunits, while different adaptor complexes bind to dileucines by diverse subunits.<sup>33</sup> These adaptor proteins then initiate clathrin assembly.

There are also other adaptor-like proteins that play the accessory role in the endocytosis process, by connecting the receptors with the clathrin coated vesicles, or directly interacting with clathrin or AP-2. This includes receptor-specific arrestins like  $\beta$ -arrestin and arrestin-3, which associate  $\beta$ -adrenergic receptors and other G protein-coupled receptors (GPCRs) to endosomal vesicles, and assist in their endocytosis.  $\beta$ -arrestin was shown to regulate desensitization and endocytosis of  $\beta$ 2-adrenergic receptor by binding with its phosphorylated form.<sup>34,35</sup> Another synaptic vesicle protein AP180 complexes with the AP-2 to enhance the clathrin coat assembly, and is also crucial for the endocytic mechanism, as its mutation in *Drosophila* led to impairment of synaptic vesicle endocytosis.<sup>36,37</sup> An analogue of AP180 in non-neuronal cells, clathrin assembly lymphoid myeloid leukemia (CALM) was shown to interact with clathrin heavy chain.<sup>38</sup> In a study by Benmerah *et al.* the epidermal growth factor receptor substrate 15 (Eps15), a tyrosine kinase substrate involved in EGFR signaling pathway, was found to be associated with AP-2 and hence involved in endocytosis.<sup>39</sup> The role of Eps15 in the cellular uptake process was further confirmed by its presence in the CCVs, and co-localization with both clathrin and AP-2 *in vivo*.<sup>40</sup> The N-terminal of Eps15 is composed of repeats of 70 amino acids that form EH (Eps15 homology) domain, while the C-terminus is rich in DPF sequence and possess the AP-2 binding site. EH is a highly conserved domain and is involved in multiple protein-protein interactions.<sup>41,42</sup> One of these proteins, Epsin, was identified as a binding partner for Eps15 and an important player in endocytic process. Epsins are widely distributed in tissues, and also interact with AP-2 by their central region, as evident by their co-precipitation along with both Eps15 and AP-2 from brain extracts.<sup>43</sup> Both Eps15 and epsin do not form major coat proteins, yet their interactions at the membrane promote clathrin lattice formation leading to invagination, and budding.

After coat proteins recruitment and assembly that result in the formation of clathrin-coated pit, the next step is the scission to dissociate clathrin-coated vesicles from the membrane. A handful of proteins are known to be involved in this fission process to yield the clathrin vesicles. As mentioned previously, a key mediator is the protein dynamin, a 100kDa GTPase, with several isoforms present in mammalian cells.<sup>44</sup> Initially identified as a microtubule-binding protein, dynamin was later shown to assemble at the plasma membrane to form helical tubes or ‘collars’

around the neck of invagination, which then constrict and hence pinch the vesicle from the membrane.<sup>45,46</sup> The role of dynamin was understood from studies on a homologous gene product from *Drosophila shibire*. The adult *Drosophila* with the *shibire* mutation suffers from temperature sensitive paralysis, which is the result of impairment in endocytosis.<sup>47</sup> Further studies by Damke *et al.* on HeLa cells overexpressing mutant dynamin which lacks the GTP binding activity revealed a defect in the pit constriction and vesicle budding in mutant cells compared to the wild type. No effect on coat protein recruitment or assembly was observed, indicating dynamin's importance in CCV formation.<sup>48</sup> Amphiphysin is another protein known to involve in the fission process. Indeed, amphiphysin links CCV budding with dynamin activity and mediates clathrin interaction with dynamin. In vitro, amphiphysin alone was shown to form narrow tubules from spherical liposomes (a property of dynamin), and in combination with dynamin, assembles around the tubule in the form of ring to promote the fragmentation.<sup>49</sup>

### 1.2.2 Vesicle Uncoating and Fusion with Endosomal Compartments

Clathrin uncoating is an essential step as it allows the fusion of vesicles to the endosomal compartments. Hsc70 is recognized as the major protein responsible for the uncoating, with the help of auxilin and ATP.<sup>24,25,50</sup> During this process, initially, an AP-clathrin-Hsc70-ADP complex formation takes place using the uncoated clathrin triskelions, which is followed by the exchange of ADP by ATP. The new complex AP-clathrin-Hsc70-ATP is a stable steady state complex and inhibits further clathrin uncoating activity of Hsc70. In the absence of AP, Hsc70 dissociates from the uncoated clathrin and continues to uncoat more clathrin from vesicles.<sup>51</sup>

After uncoating, the vesicle fuses with the early endosomes to deliver the cargo, from where it is either recycled to the surface (recycling endosomes) or sorted for degradation (late endosomes or lysosomes). It might also include some specialized compartments like MHC class II required for antigen presentation in immune cells, or secretory lysosomes.<sup>52</sup> It is relatively easier to investigate these late endocytic compartments by subcellular fractionation and morphological studies. However, since most of the molecules are degraded quickly after reaching lysosomes, the in vitro interaction studies are quite challenging. Lysosome contains an acidic medium and has multiple hydrolases to perform ultimate degradation of any endocytosed cargo. Usually, these late compartments receive the macromolecules from early lysosome and *trans*-golgi network (TGN), and there is an exchange of contents among late endosome and lysosomes.<sup>53</sup>

The whole intracellular trafficking of vesicles is tightly regulated by Rab GTPases. They are the master regulators that not only control the movement of vesicles between organelles, but also ensure their directionality. Different types of Rabs are located in different cellular compartments and have varying functions. For example, Rab5 is involved in early endocytic steps and modulates the merging of vesicles to form early endosomes, while Rab7 and Rab11 are linked with late and recycling endosomes.<sup>54,55</sup>

### 1.3 Mass Spectrometry as a Tool to Study Endocytosis

Mass spectrometry has been previously used as a tool to highlight perturbations in endocytic machinery in various diseased states. For example, changes in expression of CME proteins like dynamin-1<sup>56-58</sup>, AP-2<sup>59</sup>, and amphiphysin<sup>60</sup> were observed in schizophrenia and bipolar disorders, by mass-spectrometric methods. Additionally, there have also been advancements in understanding endocytosis using proteomic approaches, which can be broadly classified into the following categories:

#### 1.3.1 By Isolating Clathrin-coated Vesicles

One of the early mass spectrometric methods to reveal the players in clathrin-mediated endocytosis was by enrichment of vesicles into cellular fractions followed by their mass spectrometric analyses. This approach was applied to CCVs isolated from rat brain.<sup>61,62</sup> Using quantitative proteomics, by either labeling with isotopic (SILAC) or isobaric (iTRAQ/TMT) tags, it provides information about the organellar localization of proteins.<sup>63-66</sup> From an adult rat brain, CCVs were purified by differential centrifugation combined with a series of density-gradient centrifugations. The samples were further separated by gel electrophoresis, and proteins were identified by excising bands, in-gel digestion, and downstream mass spectrometric analysis. In total, 209 proteins were identified of which 92 were previously known CCV-associated proteins, including both heavy and light chains of clathrin, and all four subunits of AP-2 and AP-1 complex. Besides, 25 proteins were contaminants, leaving rest 92 as potential CCV-linked proteins.<sup>62</sup> A similar proteomic approach was applied by Girard *et al.* to CCVs isolated from adult rat liver.<sup>67</sup> They identified 346 proteins with high reproducibility, encompassing all the known CCV coat proteins and 21 novel proteins. Through label-free quantitative proteomics, they revealed a relatively lower expression of clathrin light chains (CHLs) compared to heavy chain (CHC) in

CCVs enriched from liver, in contrast to previously established 1:1 ratio between CHC:CHLs in brain CCVs. Furthermore, the higher (3:1) ratio of AP2:AP1 in CCVs from brain was found to be reversed for CCVs purified from liver, suggesting that most of the CCVs from liver originate from TGN pathway. In a separate study, Enthoprotin was discovered as a novel clathrin vesicle-associated protein from rat brain and liver, by mass spectrometry.<sup>68</sup> Enthoprotin was found to colocalize with clathrin and AP-1, and further stimulate clathrin assembly by interacting with the clathrin heavy chain. In general, mass spectrometry can be used as a tool to discover novel CCV-associated proteins.

However, the mixing of organelles while tissue homogenization or cell lysate preparation, imperfect purification and sample contamination by membrane and abundant proteins are common problems encountered in subcellular proteomics. This might lead to assignment of false positive identifications to organelles. The issue was partially resolved by comparing the proteins identified in CCVs with vesicles obtained from clathrin-depleted cells using quantitative proteomics to avoid contaminants.<sup>69</sup> The comparative proteomic strategy involved enrichment of CCVs from HeLa cells and mock CCVs from clathrin-knockdown cells, and employed two dimensional difference gel electrophoresis (DIGE) and iTRAQ for quantitative measurement of protein ratios. In total, 522 proteins were quantified by iTRAQ labeling, out of which 53 had ratios of 2 or higher for CCVs/Mock CCVs. Half of these proteins were known CCVs related proteins, four were contaminants, and remaining were novel candidates. However, the purity of CCVs obtained was much lower than the previous studies on rat brain and liver. Borner *et al.* utilized an improved CCV purification method, and performed an extensive quantitative proteomic study on CCVs from HeLa cells.<sup>70</sup> Besides siRNA knockdown of clathrin, the researchers also used auxilin depleted cells. As previously discussed, auxilin is known to have a role in vesicle uncoating. Hence, auxilin knockdown cells will have an accumulation of clathrin coat rich in clathrin-interacting proteins, but devoid of the cargo. Moreover, these clathrin cages will represent the proteins from endocytic vesicles rather than intracellular vesicles, hence assist in distinguishing between the two CCVs. Protein ratios were measured for CCVs relative to the ones obtained from clathrin depleted cells (CCVs/clathrin-knockdown) and auxilin depleted cells (CCVs/auxilin-knockdown) by SILAC-mass spectrometry. As expected, majority of known coat components were found to have a high ratio for CCVs/clathrin-knockdown and low ratio for CCVs/auxilin-knockdown cells, while contaminants did not change much between samples. Specifically, AP-1 and AP-2 complex

proteins had inverse ratio profiles, as they belong to intracellular and endocytic CCVs respectively. The method further utilizes the protein profiling through cluster analysis. The quantified proteins meeting the selection criteria were grouped and visualized using Principal Component Analysis, and the proteins in the close proximity were found to exhibit similar functions. Clathrin, AP-1, and AP-2 were clustered separately and away from other proteins. These complexes were surrounded by some known interacting proteins and some novel candidate CCV proteins. Based on cluster profiling, 136 proteins were considered as CCV-associated proteins, which included >93% of known components and 36 new candidates. Hence by using comparative proteomic approach, the proteins specific to CCVs could be identified.

As a further improvement in quantitative proteomic method to identify the true components of CCVs, Borner *et al.* developed fractionation profiling<sup>71</sup>, a universal approach based on the principle of quantifying the fractionation behavior of CCVs, instead of obtaining an absolutely pure CCV fraction. The gene silencing studies do provide reliable information about the coat components of CCVs, yet they cannot be universally applied to all cell types. The method is highly specific for cell lines amenable to perturbations. In contrast, fractionation profiling is applicable to a wide range of cell lines. The method utilizes differential centrifugation along with SILAC for quantitation. HeLa cells were metabolically labeled with heavy amino acids, and fractions were obtained from both light and heavy labeled HeLa cells. The 'light' fraction was further sub-fractionated into three fractions of similar total protein amount, using a series of increasing speed centrifugations, and the ratio of proteins in the three subfractions were quantitated against the reference 'heavy' labeled fraction. As expected, the abundance of known CCV proteins, like CHC, AP-1 and AP-2, reduced proportionally with increasing centrifugation speed, hence distinguishing them from non-CCV proteins with different distribution profiles. The diverse applicability of the approach was demonstrated on *Drosophila* S2 cells, being the first method to reveal CCV proteome of an insect cell. The investigators profiled more than 3500 and 2400 proteins from CCVs obtained from HeLa and S2 cells respectively. Therefore, the measurement of relative abundance distribution across fractions enabled the identification of true CCV components with high sensitivity and selectivity.

### 1.3.2 By Using Chemical Proteomic Approaches

#### 1.3.2.1 Receptor Identification by Ligand-Receptor Capture (LRC) Technology

As discussed previously, the process of endocytosis begins by the activation of receptors to initiate sorting signals. Receptors are particularly interesting as they are the first point of contact between the cell and various ligands including proteins, drugs and pathogens. This binding triggers the endocytosis of interacting cargoes; hence receptors are of great importance in the process. A landmark study to reveal the membrane proteins or receptors for an array of ligands was published by Wollscheid and colleagues. The Chemical Proteomic technology, called TRICEPS, utilizes a trifunctionalized reagent to covalently label the ligand of interest. The reagent comprises of a NHS-based reactive ester to label the free amines on proteins (or other ligands), a protected hydrazine to capture the oxidized receptor glycoproteins, and a biotin tag for affinity purification and downstream mass spectrometric identification.<sup>72</sup> Another similar study was recently published called HATRIC, which makes use of click chemistry for enrichment, hence reducing the false-positive identifications due to non-specific protein binding of streptavidin.<sup>73</sup>

#### 1.3.2.2 Employing Functionalized Dendrimer

All previous developments based on vesicle isolation provide comprehensive analysis of CCV proteome with high confidence, however, lack the ability to identify transiently associated proteins to CCVs. For example, the cargo cell surface proteins like receptors are known to reside in clathrin coated vesicles for a shorter time, and eventually move to early, late, or recycling endosomes. On the same lines, the above methods are unable to offer any information about the changes in CCV proteome with time, which is often more important in the study of cellular uptake process. We developed a novel strategy to decipher endocytosis by chemical crosslinking based mass spectrometry, enabling identification of protein changes at multiple time points.<sup>74</sup> The method called TITAN (Tracing Internalization and TrAfficking of nanomaterials), relies on the functionalization of a polyamidoamine generation 3 (PAMAM G3) dendrimer with a UV-reactive photocrosslinker, a fluorophore, and an affinity handle for both covalently-linking the proteins and mapping the endocytic pathway of nanoparticle internalization (Figure 2A). The PAMAM dendrimers are potential carrier molecules for therapeutic agents. G3-G1 PAMAM dendrimer conjugate, with G1 dendrimer linked to a fluorescence-labeled new generation taxoid, was shown to internalize efficiently in ovarian and breast cancer cells at a 20  $\mu\text{M}$  concentration.<sup>75</sup> The

modified dendrimer was allowed to enter HeLa cells at 37 °C for 0.5-2 hour, and UV light was irradiated to crosslink dendrimer-interacting proteins at fixed time points, followed by affinity purification and quantitative mass spectrometric analyses (Figure 2B). The aldehyde-hydrazone bioconjugation was employed for affinity purification, which enabled stringent washing conditions of 2% SDS, 8 M urea, and 3 M NaCl to remove non-specific interactors. The proteins were identified by LC-MS/MS. Using label-free Quantitation (LFQ), any protein which increased by two-fold in UV compared to No UV control, was considered as a crosslinked protein. In total, 809 proteins were crosslinked, including TFRC, CLTC (clathrin), and LAMP1 as marker proteins for different stages of endocytosis. Besides, a vast majority of other players in endocytosis were identified with high confidence (Figure 3). Hence, TITAN enabled an unbiased and high-throughput identifications of endocytic proteins. The method offers an additional advantage of providing information at spatial resolution, hence revealing direct interactors during the internalization of nanoparticles.

#### 1.4 Conclusion and Future Directions

Endocytosis is a complex cellular uptake process that comprises recruitment of several cytoplasmic proteins to the cell surface and induction of different signaling processes. The internalization begins by the attachment or binding of the cargo to the membrane ‘receptor’ protein, and formation of a pit enclosing the cargo, followed by the release of the invaginations to yield coated vesicles. The major proteins that form the vesicle coat are clathrin and adaptor proteins primarily AP-2. After release, eventually, the vesicle gets uncoated and fused to the early endosome. Subsequently, the cargo might follow the lysosomal degradation path via late endosome or the recycling pathway, assisted by the Rab7 or Rab11 respectively.

Mass Spectrometry offers a powerful and sensitive way to investigate the endocytosis process, besides the traditional methods. While there are a handful of mass spectrometric studies on endocytosis, based on nanoparticle internalization, it is only recently that a time-resolved proteomic strategy was developed, called TITAN, to trace nanoparticle internalization. This high throughput method provides unbiased identification of proteins involved in nanoparticle entry, besides offering the advantage the visualization to monitor the process. The method relies on the functionalization of a PAMAM dendrimer with a UV-photocrosslinker and a fluorophore for mass spectrometric protein identification and visualization respectively. It is a first proteomic study of

its kind to provide protein identifications at spatial resolution, as it allows the capture of direct interactors of nanoparticle during its entry into the live cell.

With the recent advancements in mass spectrometric instrumentation, and software for chemical proteomic and crosslinking use, the understanding of the complex endocytosis mechanism has truly enhanced. The idea of functionalization of nanoparticles could be further expanded to infectious viruses or pathogenic bacteria. This will serve as a useful tool to decipher the pathways utilized by various infectious agents to cause pathogenesis.

## 1.5 References

- (1) Aderem, A.; Underhill, D. M. Mechanisms of Phagocytosis in Macrophages. *Annual Review of Immunology* **1999**, *17* (1), 593–623.
- (2) Conner, S. D.; Schmid, S. L. Regulated Portals of Entry into the Cell. *Nature* **2003**, *422* (6927), 37–44.
- (3) Bitsikas, V.; Jr, I. R. C.; Nichols, B. J. Clathrin-Independent Pathways Do Not Contribute Significantly to Endocytic Flux. *eLife Sciences* **2014**, *3*, e03970.
- (4) Anderson, R. G. W.; Brown, M. S.; Goldstein, J. L. Role of the Coated Endocytic Vesicle in the Uptake of Receptor-Bound Low Density Lipoprotein in Human Fibroblasts. *Cell* **1977**, *10* (3), 351–364.
- (5) Roth, T. F.; Porter, K. R. Yolk Protein Uptake in the Oocyte of the Mosquito *Aedes Aegypti*. *The Journal of Cell Biology* **1964**, *20* (2), 313–332.
- (6) Kanaseki, T.; Kadota, K. THE “VESICLE IN A BASKET”: A Morphological Study of the Coated Vesicle Isolated from the Nerve Endings of the Guinea Pig Brain, with Special Reference to the Mechanism of Membrane Movements. *The Journal of Cell Biology* **1969**, *42* (1), 202–220.
- (7) Pearse, B. M. F. Coated Vesicles from Pig Brain: Purification and Biochemical Characterization. *Journal of Molecular Biology* **1975**, *97* (1), 93–98.
- (8) Ungewickell, E.; Branton, D. Assembly Units of Clathrin Coats. *Nature* **1981**, *289* (5796), 420–422.
- (9) Kirchhausen, T.; Harrison, S. C. Protein Organization in Clathrin Trimers. *Cell* **1981**, *23* (3), 755–761.
- (10) Näthke, I. S.; Heuser, J.; Lupas, A.; Stock, J.; Turck, C. W.; Brodsky, F. M. Folding and Trimerization of Clathrin Subunits at the Triskelion Hub. *Cell* **1992**, *68* (5), 899–910.
- (11) Acton, S. L.; Brodsky, F. M. Predominance of Clathrin Light Chain LCB Correlates with the Presence of a Regulated Secretory Pathway. *The Journal of Cell Biology* **1990**, *111* (4), 1419–1426.
- (12) Brodsky, F. M.; Hill, B. L.; Acton, S. L.; Näthke, I.; Wong, D. H.; Ponnambalam, S.; Parham, P. Clathrin Light Chains: Arrays of Protein Motifs That Regulate Coated-Vesicle Dynamics. *Trends in Biochemical Sciences* **1991**, *16*, 208–213.
- (13) Liu, S.-H.; Wong, M. L.; Craik, C. S.; Brodsky, F. M. Regulation of Clathrin Assembly and Trimerization Defined Using Recombinant Triskelion Hubs. *Cell* **1995**, *83* (2), 257–267.

- (14) Robinson, M. S. 100-KD Coated Vesicle Proteins: Molecular Heterogeneity and Intracellular Distribution Studied with Monoclonal Antibodies. *The Journal of Cell Biology* **1987**, *104* (4), 887–895.
- (15) Ahle, S.; Mann, A.; Eichelsbacher, U.; Ungewickell, E. Structural Relationships between Clathrin Assembly Proteins from the Golgi and the Plasma Membrane. *EMBO J* **1988**, *7* (4), 919–929.
- (16) Simpson, F.; Peden, A. A.; Christopoulou, L.; Robinson, M. S. Characterization of the Adaptor-Related Protein Complex, AP-3. *J Cell Biol* **1997**, *137* (4), 835–845.
- (17) Dell’Angelica, E. C.; Ohno, H.; Ooi, C. E.; Rabinovich, E.; Roche, K. W.; Bonifacino, J. S. AP-3: An Adaptor-like Protein Complex with Ubiquitous Expression. *EMBO J* **1997**, *16* (5), 917–928.
- (18) Dell’Angelica, E. C.; Mullins, C.; Bonifacino, J. S. AP-4, a Novel Protein Complex Related to Clathrin Adaptors. *J. Biol. Chem.* **1999**, *274* (11), 7278–7285.
- (19) Hirst, J.; Bright, N. A.; Rous, B.; Robinson, M. S. Characterization of a Fourth Adaptor-Related Protein Complex. *Mol Biol Cell* **1999**, *10* (8), 2787–2802.
- (20) Shih, W.; Gallusser, A.; Kirchhausen, T. A Clathrin-Binding Site in the Hinge of the 2 Chain of Mammalian AP-2 Complexes. *J. Biol. Chem.* **1995**, *270* (52), 31083–31090.
- (21) Vigers, G. P.; Crowther, R. A.; Pearse, B. M. Location of the 100 Kd-50 Kd Accessory Proteins in Clathrin Coats. *EMBO J* **1986**, *5* (9), 2079–2085.
- (22) Chang, M. P.; Mallet’, W. G.; Mostov, K. E.; Brodsky, F. M. Adaptor Self-Aggregation, Adaptor -Receptor Recognition and Binding of  $\alpha$ -Adaptin Subunits to the Plasma Membrane Contribute to Recruitment of Adaptor (AP2) Components of Clathrin-Coated Pits. 12.
- (23) Saheki, Y.; De Camilli, P. Synaptic Vesicle Endocytosis. *Cold Spring Harb Perspect Biol* **2012**, *4* (9).
- (24) Greene, L. E.; Eisenberg, E. Dissociation of Clathrin from Coated Vesicles by the Uncoating ATPase. *J. Biol. Chem.* **1990**, *265* (12), 6682–6687.
- (25) Ungewickell, E.; Ungewickell, H.; Holstein, S. E. H.; Lindner, R.; Prasad, K.; Barouch, W.; Martini, B.; Greene, L. E.; Eisenberg, E. Role of Auxilin in Uncoating Clathrin-Coated Vesicles. *Nature* **1995**, *378* (6557), 632–635.
- (26) Gaidarov, I.; Chen, Q.; Falck, J. R.; Reddy, K. K.; Keen, J. H. A Functional Phosphatidylinositol 3,4,5-Trisphosphate/Phosphoinositide Binding Domain in the Clathrin Adaptor AP-2  $\alpha$  Subunit. IMPLICATIONS FOR THE ENDOCYTTIC PATHWAY. *J. Biol. Chem.* **1996**, *271* (34), 20922–20929.
- (27) Gaidarov, I.; Keen, J. H. Phosphoinositide–Ap-2 Interactions Required for Targeting to Plasma Membrane Clathrin-Coated Pits. *J Cell Biol* **1999**, *146* (4), 755–764.
- (28) West, M. A.; Bright, N. A.; Robinson, M. S. The Role of ADP-Ribosylation Factor and Phospholipase D in Adaptor Recruitment. *J Cell Biol* **1997**, *138* (6), 1239–1254.
- (29) Zhang, J. Z.; Davletov, B. A.; Südhof, T. C.; Anderson, R. G. W. Synaptotagmin I Is a High Affinity Receptor for Clathrin AP-2: Implications for Membrane Recycling. *Cell* **1994**, *78* (5), 751–760.
- (30) Haucke, V.; Camilli, P. D. AP-2 Recruitment to Synaptotagmin Stimulated by Tyrosine-Based Endocytic Motifs. *Science* **1999**, *285* (5431), 1268–1271.
- (31) Ohno, H.; Stewart, J.; Fournier, M. C.; Bosshart, H.; Rhee, I.; Miyatake, S.; Saito, T.; Gallusser, A.; Kirchhausen, T.; Bonifacino, J. S. Interaction of Tyrosine-Based Sorting Signals with Clathrin-Associated Proteins. *Science* **1995**, *269* (5232), 1872–1875.

- (32) Heilker, R.; Manning-Krieg, U.; Zuber, J. F.; Spiess, M. In Vitro Binding of Clathrin Adaptors to Sorting Signals Correlates with Endocytosis and Basolateral Sorting. *The EMBO Journal* **1996**, *15* (11), 2893–2899.
- (33) Marks, M. S.; Ohno, H.; Kirchhausen, T.; Bonracino, J. S. Protein Sorting by Tyrosine-Based Signals: Adapting to the Ys and Wherefores. *Trends in Cell Biology* **1997**, *7* (3), 124–128.
- (34) Jr, O. B. G.; Krupnick, J. G.; Santini, F.; Gurevich, V. V.; Penn, R. B.; Gagnon, A. W.; Keen, J. H.; Benovic, J. L.  $\beta$ -Arrestin Acts as a Clathrin Adaptor in Endocytosis of the B2-Adrenergic Receptor. *Nature* **1996**, *383* (6599), 447–450.
- (35) Lin, F.-T.; Krueger, K. M.; Kendall, H. E.; Daaka, Y.; Fredericks, Z. L.; Pitcher, J. A.; Lefkowitz, R. J. Clathrin-Mediated Endocytosis of the  $\beta$ -Adrenergic Receptor Is Regulated by Phosphorylation/Dephosphorylation of  $\beta$ -Arrestin1. *J. Biol. Chem.* **1997**, *272* (49), 31051–31057.
- (36) Zhang, B.; Koh, Y. H.; Beckstead, R. B.; Budnik, V.; Ganetzky, B.; Bellen, H. J. Synaptic Vesicle Size and Number Are Regulated by a Clathrin Adaptor Protein Required for Endocytosis. *Neuron* **1998**, *21* (6), 1465–1475.
- (37) Hao, W.; Luo, Z.; Zheng, L.; Prasad, K.; Lafer, E. M. AP180 and AP-2 Interact Directly in a Complex That Cooperatively Assembles Clathrin. *J. Biol. Chem.* **1999**, *274* (32), 22785–22794.
- (38) Tebar, F.; Bohlander, S. K.; Sorkin, A. Clathrin Assembly Lymphoid Myeloid Leukemia (CALM) Protein: Localization in Endocytic-Coated Pits, Interactions with Clathrin, and the Impact of Overexpression on Clathrin-Mediated Traffic. *Mol Biol Cell* **1999**, *10* (8), 2687–2702.
- (39) Benmerah, A.; Gagnon, J.; Bègue, B.; Mégarbané, B.; Dautry-Varsat, A.; Cerf-Bensussan, N. The Tyrosine Kinase Substrate Eps15 Is Constitutively Associated with the Plasma Membrane Adaptor AP-2. *The Journal of Cell Biology* **1995**, *131* (6), 1831–1838.
- (40) van Delft, S.; Schumacher, C.; Hage, W.; Verkleij, A. J.; Henegouwen, P. M. P. van B. en. Association and Colocalization of Eps15 with Adaptor Protein-2 and Clathrin. *J Cell Biol* **1997**, *136* (4), 811–821.
- (41) Benmerah, A.; Bègue, B.; Dautry-Varsat, A.; Cerf-Bensussan, N. The Ear of  $\beta$ -Adaptin Interacts with the COOH-Terminal Domain of the Eps15 Protein. *J. Biol. Chem.* **1996**, *271* (20), 12111–12116.
- (42) Wong, W. T.; Schumacher, C.; Salcini, A. E.; Romano, A.; Castagnino, P.; Pelicci, P. G.; Fiore, P. D. A Protein-Binding Domain, EH, Identified in the Receptor Tyrosine Kinase Substrate Eps15 and Conserved in Evolution. *PNAS* **1995**, *92* (21), 9530–9534.
- (43) Chen, H.; Fre, S.; Slepnev, V. I.; Capua, M. R.; Takei, K.; Butler, M. H.; Fiore, P. P. D.; Camilli, P. D. Epsin Is an EH-Domain-Binding Protein Implicated in Clathrin-Mediated Endocytosis. *Nature* **1998**, *394* (6695), 793–797.
- (44) Cao, H.; Garcia, F.; McNiven, M. A. Differential Distribution of Dynamin Isoforms in Mammalian Cells. *Mol Biol Cell* **1998**, *9* (9), 2595–2609.
- (45) Kosaka, T.; Ikeda, K. Possible Temperature-Dependent Blockage of Synaptic Vesicle Recycling Induced by a Single Gene Mutation in *Drosophila*. *Journal of Neurobiology* **1983**, *14* (3), 207–225.
- (46) Sweitzer, S. M.; Hinshaw, J. E. Dynamin Undergoes a GTP-Dependent Conformational Change Causing Vesiculation. *Cell* **1998**, *93* (6), 1021–1029.

- (47) Blik, A. M. van der; Meyerowitz, E. M. Dynamin-like Protein Encoded by the *Drosophila* Shibire Gene Associated with Vesicular Traffic. *Nature* **1991**, *351* (6325), 411–414.
- (48) Damke, H.; Baba, T.; Warnock, D. E.; Schmid, S. L. Induction of Mutant Dynamin Specifically Blocks Endocytic Coated Vesicle Formation. *The Journal of Cell Biology* **1994**, *127* (4), 915–934.
- (49) Takei, K.; Slepnev, V. I.; Haucke, V.; Camilli, P. D. Functional Partnership between Amphiphysin and Dynamin in Clathrin-Mediated Endocytosis. *Nature Cell Biology* **1999**, *1* (1), 33–39.
- (50) Rothman, J. E.; Schmid, S. L. Enzymatic Recycling of Clathrin from Coated Vesicles. *Cell* **1986**, *46* (1), 5–9.
- (51) Jiang, R.; Gao, B.; Prasad, K.; Greene, L. E.; Eisenberg, E. Hsc70 Chaperones Clathrin and Primes It to Interact with Vesicle Membranes. *J. Biol. Chem.* **2000**, *275* (12), 8439–8447.
- (52) Kleijmeer, M. J.; Morkowski, S.; Griffith, J. M.; Rudensky, A. Y.; Geuze, H. J. Major Histocompatibility Complex Class II Compartments in Human and Mouse B Lymphoblasts Represent Conventional Endocytic Compartments. *The Journal of Cell Biology* **1997**, *139* (3), 639–649.
- (53) Huotari, J.; Helenius, A. Endosome Maturation. *EMBO J* **2011**, *30* (17), 3481–3500.
- (54) Rodman, J. S.; Wandinger-Ness, A. Rab GTPases Coordinate Endocytosis. *J Cell Sci* **2000**, *113* (2), 183–192.
- (55) HUTAGALUNG, A. H.; NOVICK, P. J. Role of Rab GTPases in Membrane Traffic and Cell Physiology. *Physiol Rev* **2011**, *91* (1), 119–149.
- (56) Pennington, K.; Beasley, C. L.; Dicker, P.; Fagan, A.; English, J.; Pariante, C. M.; Wait, R.; Dunn, M. J.; Cotter, D. R. Prominent Synaptic and Metabolic Abnormalities Revealed by Proteomic Analysis of the Dorsolateral Prefrontal Cortex in Schizophrenia and Bipolar Disorder. *Molecular Psychiatry* **2008**, *13* (12), 1102–1117.
- (57) Föcking, M.; Dicker, P.; English, J. A.; Schubert, K. O.; Dunn, M. J.; Cotter, D. R. Common Proteomic Changes in the Hippocampus in Schizophrenia and Bipolar Disorder and Particular Evidence for Involvement of Cornu Ammonis Regions 2 and 3. *Arch Gen Psychiatry* **2011**, *68* (5), 477–488.
- (58) Schubert, K. O.; Föcking, M.; Prehn, J. H. M.; Cotter, D. R. Hypothesis Review: Are Clathrin-Mediated Endocytosis and Clathrin-Dependent Membrane and Protein Trafficking Core Pathophysiological Processes in Schizophrenia and Bipolar Disorder? *Molecular Psychiatry* **2012**, *17* (7), 669–681.
- (59) Föcking, M.; Lopez, L. M.; English, J. A.; Dicker, P.; Wolff, A.; Brindley, E.; Wynne, K.; Cagney, G.; Cotter, D. R. Proteomic and Genomic Evidence Implicates the Postsynaptic Density in Schizophrenia. *Mol. Psychiatry* **2015**, *20* (4), 424–432.
- (60) English, J. A.; Dicker, P.; Föcking, M.; Dunn, M. J.; Cotter, D. R. 2-D DIGE Analysis Implicates Cytoskeletal Abnormalities in Psychiatric Disease. *PROTEOMICS* **9** (12), 3368–3382.
- (61) Ritter, B.; Blondeau, F.; Denisov, A. Y.; Gehring, K.; McPherson, P. S. Molecular Mechanisms in Clathrin-Mediated Membrane Budding Revealed through Subcellular Proteomics. *Biochemical Society Transactions* **2004**, *32* (5), 769–773.
- (62) Blondeau, F.; Ritter, B.; Allaire, P. D.; Wasiak, S.; Girard, M.; Hussain, N. K.; Angers, A.; Legendre-Guillemain, V.; Roy, L.; Boismenu, D.; et al. Tandem MS Analysis of Brain

- Clathrin-Coated Vesicles Reveals Their Critical Involvement in Synaptic Vesicle Recycling. *Proceedings of the National Academy of Sciences* **2004**, *101* (11), 3833–3838.
- (63) Christoforou, A.; Mulvey, C. M.; Breckels, L. M.; Geladaki, A.; Hurrell, T.; Hayward, P. C.; Naake, T.; Gatto, L.; Viner, R.; Arias, A. M.; et al. A Draft Map of the Mouse Pluripotent Stem Cell Spatial Proteome. *Nature Communications* **2016**, *7*, 9992.
- (64) Itzhak, D. N.; Tyanova, S.; Cox, J.; Borner, G. H. Global, Quantitative and Dynamic Mapping of Protein Subcellular Localization. *eLife Sciences* **2016**, *5*, e16950.
- (65) Jadot, M.; Boonen, M.; Thirion, J.; Wang, N.; Xing, J.; Zhao, C.; Tannous, A.; Qian, M.; Zheng, H.; Everett, J. K.; et al. Accounting for Protein Subcellular Localization: A Compartmental Map of the Rat Liver Proteome. *Mol Cell Proteomics* **2017**, *16* (2), 194–212.
- (66) Thul, P. J.; Åkesson, L.; Wiking, M.; Mahdessian, D.; Geladaki, A.; Blal, H. A.; Alm, T.; Asplund, A.; Björk, L.; Breckels, L. M.; et al. A Subcellular Map of the Human Proteome. *Science* **2017**, eaal3321.
- (67) Girard, M.; Allaire, P. D.; McPherson, P. S.; Blondeau, F. Non-Stoichiometric Relationship between Clathrin Heavy and Light Chains Revealed by Quantitative Comparative Proteomics of Clathrin-Coated Vesicles from Brain and Liver. *Mol Cell Proteomics* **2005**, *4* (8), 1145–1154.
- (68) Wasiak, S.; Legendre-Guillemain, V.; Puertollano, R.; Blondeau, F.; Girard, M.; de Heuvel, E.; Boismenu, D.; Bell, A. W.; Bonifacino, J. S.; McPherson, P. S. Enthoprotin. *J Cell Biol* **2002**, *158* (5), 855–862.
- (69) Borner, G. H. H.; Harbour, M.; Hester, S.; Lilley, K. S.; Robinson, M. S. Comparative Proteomics of Clathrin-Coated Vesicles. *J Cell Biol* **2006**, *175* (4), 571–578.
- (70) Borner, G. H. H.; Antrobus, R.; Hirst, J.; Bhumbra, G. S.; Kozik, P.; Jackson, L. P.; Sahlender, D. A.; Robinson, M. S. Multivariate Proteomic Profiling Identifies Novel Accessory Proteins of Coated Vesicles. *J Cell Biol* **2012**, *197* (1), 141–160.
- (71) Borner, G. H. H.; Hein, M. Y.; Hirst, J.; Edgar, J. R.; Mann, M.; Robinson, M. S. Fractionation Profiling: A Fast and Versatile Approach for Mapping Vesicle Proteomes and Protein–Protein Interactions. *Mol Biol Cell* **2014**, *25* (20), 3178–3194.
- (72) Frei, A. P.; Jeon, O.-Y.; Kilcher, S.; Moest, H.; Henning, L. M.; Jost, C.; Plückthun, A.; Mercer, J.; Aebersold, R.; Carreira, E. M.; et al. Direct Identification of Ligand-Receptor Interactions on Living Cells and Tissues. *Nature Biotechnology* **2012**, *30* (10), 997–1001.
- (73) Sobotzki, N.; Schafroth, M. A.; Rudnicka, A.; Koetemann, A.; Marty, F.; Goetze, S.; Yamauchi, Y.; Carreira, E. M.; Wollscheid, B. HATRIC-Based Identification of Receptors for Orphan Ligands. *Nature Communications* **2018**, *9* (1), 1519.
- (74) Wang, L.; Yang, L.; Pan, L.; Kadasala, N. R.; Xue, L.; Schuster, R. J.; Parker, L. L.; Wei, A.; Tao, W. A. Time-Resolved Proteomic Visualization of Dendrimer Cellular Entry and Trafficking. *J. Am. Chem. Soc.* **2015**, *137* (40), 12772–12775.
- (75) Wang, T.; Zhang, Y.; Wei, L.; Teng, Y. G.; Honda, T.; Ojima, I. Design, Synthesis, and Biological Evaluations of Asymmetric Bow-Tie PAMAM Dendrimer-Based Conjugates for Tumor-Targeted Drug Delivery. *ACS Omega* **2018**, *3* (4), 3717–3736.

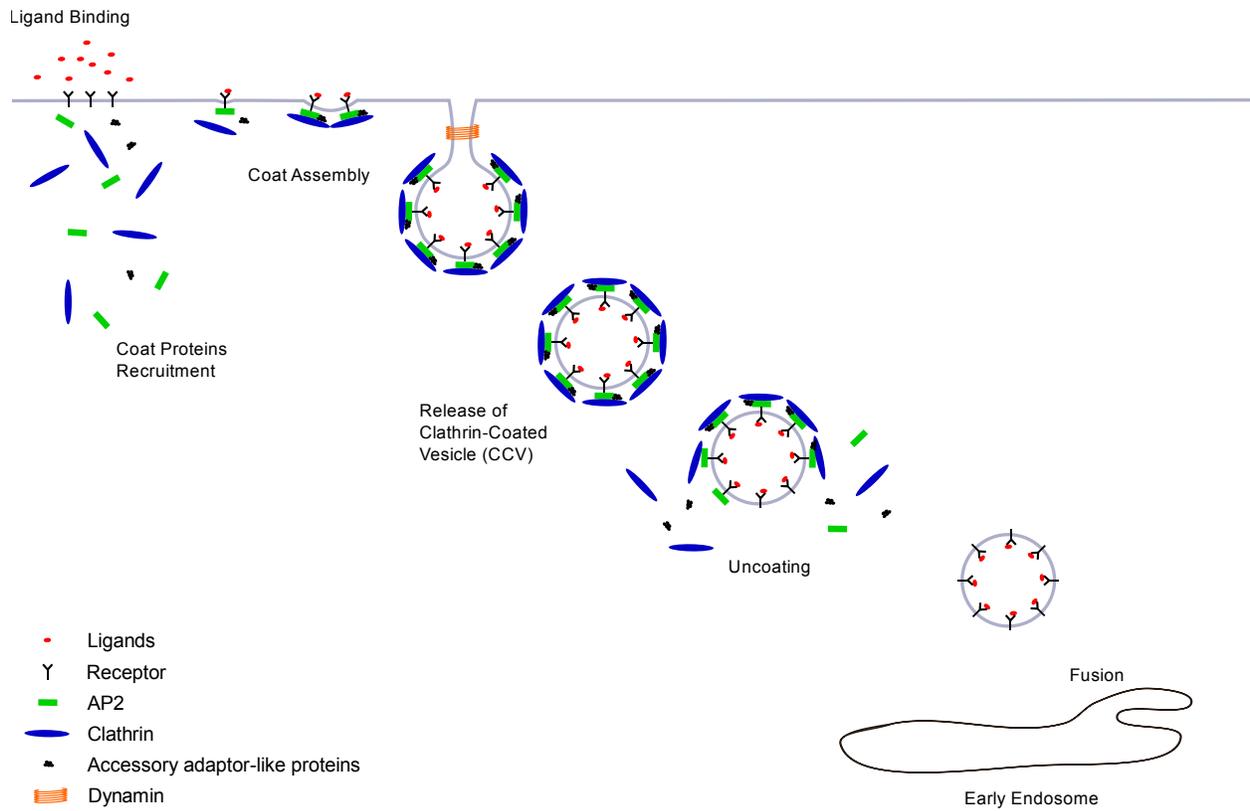


Figure 1.1: Diagram showing the phases of endocytosis: Ligand binding, Coat proteins recruitment, Coat assembly, Scission leading to release of clathrin coated vesicle, Uncoating, and Fusion with early endosome.

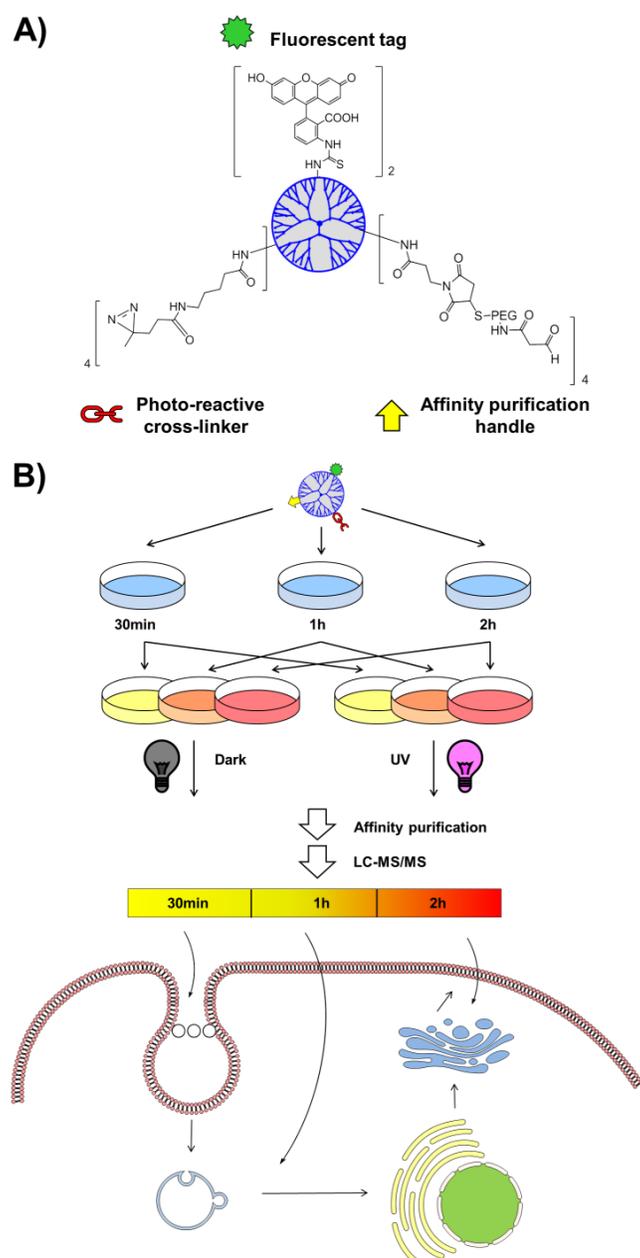


Figure 1.2: (A) Schematic representation of the functionalized dendrimer. (B) Experimental workflow for TITAN analysis.

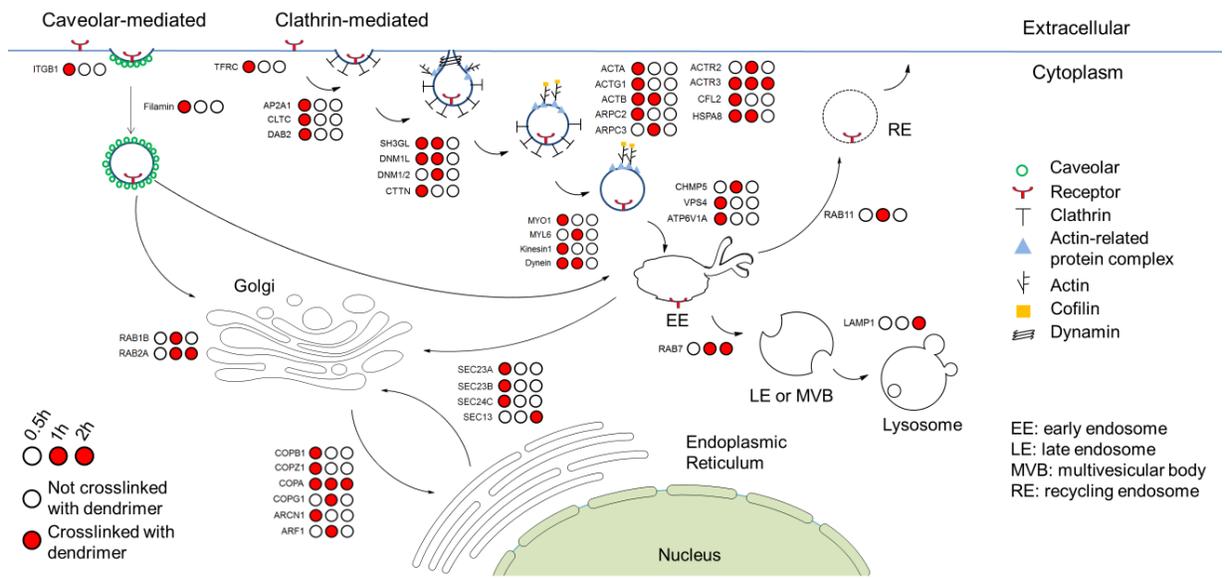


Figure 1.3: Spatiotemporal information for dendrimer-interacting proteins involved in the clathrin- and caveolar/raft-mediated endocytic pathways revealed by TITAN analysis.

## CHAPTER 2. MAPPING THE EARLY STAGE ENTRY OF ZIKA VIRUS INTO HOST CELLS

### 2.1 Summary

The outbreak of Zika in early 2016 created worldwide urgency and demanded rapid development in research pertaining to Zika virus. Till date, this has led to the elucidation of structure of Zika virus, development of neutralizing antibodies, and exploration of the host factors involved in Zika entry into various cell lines. The later, though, is not completely unraveled, primarily due to the lack of a robust and high-throughput method to study the enveloped virus internalization. Here, we present a novel technology to trace the early stage entry of Zika virus into host cells. Zika virus was labeled on its surface with a chemical probe, which carries a UV-photocrosslinker to covalently link any virus-interacting proteins on UV exposure, and a biotin tag for subsequent enrichment and mass spectrometric identification. The 'surface modified' virus was allowed to either attach or enter Vero cells, followed by UV-photocrosslinking at certain time points, to reveal the receptor or other host proteins critical for virus internalization. We identified Neural Cell Adhesion Molecule (NCAM1) as the potential attachment factor/ receptor for Zika virus, which we validated using antibody inhibition in Vero cells and overexpression into HEK 293T cells. Furthermore, using the technology we isolated other direct interactors of Zika virus, which might be critical for its pathogenesis. The method could serve as a universal tool to map the entry pathway of other enveloped virus, including Dengue and West Nile from the family *Flaviviridae*.

### 2.2 Introduction

Zika was first discovered in the Zika valley of Uganda in 1947 (1), however, only after the recent outbreak in South Americas it was declared a health emergency (2). Early cases were observed with skin rash, fever and other mild symptoms, but later a close relation between Zika infection and microcephaly in newborns and Guillain-Barré syndrome (GBS) in adults were reported (3,4). The primary mode of transmission to humans is through an infected mosquito, *Aedes aegypti*.

Zika belongs to *Flaviviridae*, the same family as dengue virus (DENV), West Nile virus (WNV), Japanese encephalitis virus (JEV), and yellow fever virus (YFV). It is a positive-strand RNA virus with a genome of about 11,000 nucleotides, which are translated into a single polyprotein, further processed by viral and host proteases to form three structural and seven non-structural proteins. Like other flavivirus, a mature Zika virus is composed of a nucleocapsid enclosed in an icosahedral shell containing 180 copies (90 dimers) each of the envelope E and membrane M proteins (5). E protein is the key surface glycoprotein involved in receptor-binding and membrane fusion with the host cell (6).

Zika virus has been the focus of immense investigation since the recent epidemic. Initial studies on Zika structure revealed its structural similarity with other virus from *Flaviviridae*, besides the peculiar Asn154 glycosylation site present on each of the E proteins (7), the thermal stability, and the compact surface of the virus (8). Progress has also been made towards developing therapeutics, by isolating neutralizing monoclonal antibodies from infected human subjects (9). However, a much deeper understanding of the immune response elicited and the various virus-host interactions are required for the rapid development of antiviral agents.

Identification of direct interactors of Zika virus during its entry into the host cells not only suggests different molecular pathways manipulated by the virus, but also provides immense opportunity to develop antivirals by offering new potential drug targets. However, owing to the transient nature of these interactions and the extreme rapidness in the flavivirus entry, in general, tracing the direct interactors of virus is a formidable task. One of the key proteins that assist in virus binding and entry is a receptor protein. Furthermore, mapping the host factors critical for viral infection will highlight the molecular pathways manipulated by the virus to maneuver to its benefits.

Here, we present a novel chemical proteomic technology to trace the virus entry and identify virus-interacting proteins. The virus ‘surface proteins’ were chemically labeled using a probe that bears a UV-photocrosslinker and a biotin handle (Figure 1B). The trifunctionalized probe contains a cysteine-reactive maleimide group to label the virus surface proteins, a diazirine for covalently crosslinking the virus-interacting proteins on exposure to UV light, and a biotin tag for enrichment and downstream mass spectrometric analysis (Figure 1A). The three functionalities are separated by a polyethylene glycol (PEG)-like linker to confer membrane impermeability so as to allow endocytic viral entry, while offering the flexibility for efficient

crosslinking and enrichment. The labeled Zika was used to infect Vero cells and proteins were crosslinked at fixed time points to identify the virus receptors and elucidate the virus entry mechanism (Figure 1C). Vero cells were chosen due to the well-established high infectivity of Zika, and the diazirine group was selected due to its high selectivity in protein crosslinking (10). The diazirine group allows tracing the virus movement in ‘real-time’, which is challenging due to the highly dynamic nature of the process and the transient virus-host protein interactions (11). Moreover, the crosslinking chemistry permits the identification of receptors, which otherwise being a hydrophobic membrane protein, presents its own challenges (12, 13). Compared to the previously reported mass spectrometric method (14) for identification of receptors, specific to glycoproteins, our method offers the additional advantage of covalently linking proteins at different time points, thus serving the dual purpose of identification of receptors and other host factors involved in different stages of virus entry. Lastly, this novel technology can be applied to relatively unstable enveloped viruses, owing to the minimal labeling by cysteine-reactive maleimide group.

## 2.3 Experimental Procedures

### 2.3.1 Materials

**Cells and Reagents:** Vero (African green monkey kidney) and HEK 293T (human embryonic kidney) cells were maintained in DMEM media supplemented with heat-inactivated 10% FBS, at 37 °C and under 5% CO<sub>2</sub>. Low passage cells were used for the virus propagation and all other infection experiments. The anti-NCAM monoclonal antibody for blocking cell surface protein and immunofluorescence assay was purchased from BD Biosciences (Cat. No. 559043), while the antibody for Western blot was obtained from Cell Signaling (Cat. No. 3576). Streptavidin-HRP antibody was bought from R&D Systems (Cat. No. DY998), and 4G2 antibody was generously provided by Richard Kuhn, Purdue University. All reagents for synthesis were obtained from Sigma-Aldrich, ChemPep Inc, Peptides International Inc, Novabiochem (EMD Millipore), and Alfa Aesar.

### 2.3.2 Mature Zika Preparation

Approximately  $1 \times 10^9$  Vero-Furin cells (15) were infected at a multiplicity of infection (MOI) of 0.1 with Zika virus (strain H/PF/2013) at 37 °C (16). Virus particles were purified from media collected at 60 and 72 hours post infection (hpi) according to (17). Briefly, virus particles were precipitated from the media with 8% polyethylene glycol (PEG) 8000 overnight at 4 °C, pelleted at 8891xg for 50 minutes at 4 °C. Re-suspended particles were pelleted through a 24% sucrose cushion, re-suspended in 0.5 mL NTE buffer (20mM Tris pH 8.0, 120mM NaCl, 1mM EDTA) and purified with a discontinuous gradient in 5% intervals from 35% to 10% K-tartrate, 20mM Tris pH 8.0, 1mM EDTA. Mature virus was extracted from the gradient, concentrated and buffer exchanged into NTE buffer.

### 2.3.3 Plaque Assay

The plaque assay was performed as described previously with some modification (18). Purified Zika virus was diluted serially in the order of ten folds, and incubated with monolayers of Vero cells for 1 hour at room temperature. Cells were layered with agarose, and incubated at 37 °C for 3 days. Plaques were counted following cell staining using Neutral red.

### 2.3.4 Synthesis and Purification of Labeling Reagent

The virus-labeling reagent was synthesized on the Rink-Amide-AM-Resin (200-400 mesh) 1% DVB, manually, using standard solid phase peptide synthesis approach. A 20% piperidine solution in DMF (N,N-Dimethylformamide) was used to deprotect the fmoc (9-Fluorenylmethoxycarbonyl) groups, while 95% TFA (Trifluoroacetic acid) was used for boc (*tert*-Butoxycarbonyl) group deprotection. HCTU (O-(1H-6-Chlorobenzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate) was utilized as an activating agent for the carboxyl group on the incoming reactant, in presence of the base NMM (4-methylmorpholine). The synthesis was performed on the 30  $\mu$ mol scale, and using 2.5 times excess of the reagents compared to the resin. Each step involved the deprotection of amine group, activation of carboxyl group followed by coupling reaction. The excess reagents were removed by thorough washing of beads by DMF. Ninhydrin test was performed after each deprotection and coupling reaction.

The synthesis was performed using the strategy, as previously described (19). 80 mg (30  $\mu\text{mol}$ ) of Rink-Amide-AM-Resin was added to the fritted reaction vessel. Beads were conditioned with DMF for 15 minutes. The solution was removed by filtration, and 20% piperidine was added to the beads for fmoc deprotection. The mixture was rotated for 30 minutes, and solution was removed followed by beads washing with DMF. A reaction mixture of Fmoc-Lys-Biotin-OH (44.60 mg, 75  $\mu\text{mol}$ ), HCTU (31.03 mg, 75  $\mu\text{mol}$ ), and NMM (16.49  $\mu\text{l}$ , 150  $\mu\text{mol}$ ) in DMF was added to the resin, and rotated for 4 hours at room temperature. The resin was washed with DMF, and fmoc deprotection steps were repeated. A solution of N-Fmoc-N''-succinyl-4,7,10-trioxa-1,13-tridecanediamine (40.70 mg, 75  $\mu\text{mol}$ ), HCTU (31.03 mg, 75  $\mu\text{mol}$ ), and NMM (16.49  $\mu\text{l}$ , 150  $\mu\text{mol}$ ) in DMF was rotated with the resin for 4 hours. Excess reagents were removed and the resin was washed with DMF. Similarly, Fmoc-Lys(Boc)-OH (35.14 mg, 75  $\mu\text{mol}$ ), N-Fmoc-N''-succinyl-4,7,10-trioxa-1,13-tridecanediamine (40.70 mg, 75  $\mu\text{mol}$ ), and 6-Maleimidohexanoic acid (15.84 mg, 75  $\mu\text{mol}$ ) were coupled after fmoc deprotection. The resin was washed with DMF and dichloromethane, and the molecule was cleaved from the resin using 95:5 mixture of TFA:TIS (Triisopropylsilane) for 1.5 hours. The cleavage step also deprotects the boc group, making available a free amine in the product. The crude product was concentrated and HPLC (Agilent 1100) purified using a gradient of 5-85% B (A: 0.1% TFA/H<sub>2</sub>O, B: 0.1% TFA/CH<sub>3</sub>OH) for 30 minutes on Waters XBridge Prep BEH130 C18 column 5  $\mu\text{m}$ , 10 X 250 mm. MALDI-MS was performed to characterize the product.

In order to add the diazirine functionality, the pure maleimide-biotin product was dissolved in DMF and reacted with excess NHS-LC-Diazirine (succinimidyl 6-(4,4'-azipentanamido)hexanoate) in phosphate buffer pH 8, for 2 hours at room temperature. The product was purified by directly loading the mixture on the column and using similar HPLC conditions. The final product was characterized by MALDI-MS, dissolved in DMF and used for virus labeling.

### 2.3.5 Virus Labeling

Purified Zika virus was diluted to 500  $\mu\text{l}$  with PBS pH 7, and mixed with the synthesized labeling reagent in the concentration of 0.1  $\mu\text{M}$ . The virus was labeled by end to end rotation in 4 °C overnight. The labeling was initiated a day before the cells reached confluency for

infection. The reaction was quenched by adding three times higher concentration of cysteine, 0.3  $\mu\text{M}$ .

### 2.3.6 Virus Infection and Crosslinking of Host-Proteins

Vero cells were first grown in T-150 flasks in DMEM supplemented with 10% FBS, then passaged to the 15 cm plates and grown to confluency. Cells were washed with cold PBS twice, and cooled down in 4 °C for 5 minutes. The labeled virus was diluted in DMEM and added to the cooled cells at an MOI of 5. Cells were rocked for 1 hour in 4 °C, to allow for virus attachment. For the receptor crosslinking, the unbound virus was removed, cells were washed once with cold PBS, and directly exposed to the UV light for 15 minutes on ice. All the above operations were performed on ice and using cold PBS to minimize any virus entry. To understand the virus internalization mechanism, virus was allowed to enter cells by incubation in 37 °C for 4 and 8 minutes, after pre-attachment for an hour at 4 °C. Subsequent to UV photocrosslinking, cells were collected by scraping in PBS, and stored in -80 °C until further processing. As a control, cells treated with the labeling reagent and exposed to UV were included.

### 2.3.7 Sample Preparation for LC-MS Analysis

Frozen cells were lysed in 1% SDS supplemented with protease inhibitor on ice, using sonication (10 cycles for 10 seconds each, with an interval of 10 seconds). Cell lysates were cleared by centrifugation at 14000 rpm to pellet down cell debris, and supernatant were used for the biotin-Neutravidin affinity purification. Bicinchoninic acid (BCA) assay was performed for protein quantitation, and the lysates equivalent to 1 mg protein for each sample were reduced and alkylated by boiling at 95 °C, in 10mM TCEP (Tris(2-carboxyethyl)phosphine) and 40mM CAA (chloroacetamide) respectively. The lysates were then diluted to 0.1% SDS and rotated with 50  $\mu\text{l}$  preconditioned Neutravidin beads in 4 °C overnight. The beads were washed three times with 0.1% SDS in Tris pH 8.5, and then transferred to the low protein binding eppendorf tubes, where they were further washed three times with 25mM ammonium bicarbonate (ABC) buffer pH 8. 200  $\mu\text{l}$  ABC buffer was added to the beads, and proteins were digested on-bead at 37 °C using 2  $\mu\text{l}$  Lys-C for 3 hours and 200 ng trypsin for 12 hours. The supernatant containing peptides was collected and beads were washed twice with 50  $\mu\text{l}$  ABC buffer, further pooled with the

supernatant. Peptides were acidified and desalted using in-house StageTips with SDB-XC (3M). The peptides were dried, resuspended in 0.1% formic acid, and analyzed by LC-MS/MS.

### 2.3.8 LC-MS/MS Analysis

The peptides were loaded on a 45 cm column, in-house packed with C18 resin (2.2  $\mu\text{m}$ , 100Å, Michrom Bioresources), and separated using the gradient system 10-30% B (solvent A: Water/ 0.1% formic acid, solvent B: 80% Acetonitrile/ 0.1% formic acid) over 60 minutes at a flow rate of 250 nl/min. Liquid Chromatography was performed on an EASY-nLC 1000 (Thermo) connected to the Orbitrap Velos Pro mass spectrometer (Thermo) by a nanospray source. Data acquisition was performed in data-dependent mode, in which a full scan (range from m/z 350-1800 at a resolution of 30000) was followed by MS/MS scans of top 10 abundant peptides with a dynamic exclusion for 60s and dynamic list of 500.

### 2.3.9 Statistical Analysis

Raw files were processed with MaxQuant v1.5.5.1 (20), and the Label-free Quantitation (LFQ) (21) was performed. The raw data was searched against UniProtKB green monkey (*Chlorocebus sabaesus*) FASTA database with Andromeda search engine (22), using standard parameters. Carbamidomethylation was set as a fixed modification for cysteines, while oxidation of methionine and acetylation at N-terminus were selected as variable modifications. Enzyme specificity was set to trypsin with maximum two missed cleavages. The search was performed with 1% false discovery rate (FDR) at both peptide and protein levels. The identifications were transferred from the sequenced peaks to the unidentified peaks of the same m/z within a time window of 0.7 minutes (match between runs) across samples. All potential contaminants were removed, along with proteins only identified by site. Proteins with valid values in minimum two out of three replicates in at least one group were only considered, and values were imputed for all missing values based on normal distribution. Data was analyzed by Perseus (23), and t-test was performed to identify proteins with significant changes (permutation based FDR 5%). Proteins in the significant region with at least 2.5-fold change for UV at different time points vs control were considered as crosslinked proteins, which were further analyzed by Ingenuity Pathway Analysis, DAVID (Database for Annotation, Visualization, and Integrated Discovery), and

STRING. Proteins were matched to their homologs from UniProt human database for the analysis.

#### 2.3.10 Antibody Inhibition of Virus Infection

Antibody inhibition assay was performed according to the protocol described before (24). Vero cells in six-well plate were preincubated with 30  $\mu$ g/ml of anti-NCAM or control IgG in DMEM for 45 min at room temperature. Cells were then infected with purified zika at an MOI of 0.1 in the presence of antibody. Cellular RNA was purified after 24 h, and RT-qPCR was performed to measure viral RNA.

#### 2.3.11 RNA Extraction and RT-qPCR

RNA was extracted using RNeasy mini kit (Qiagen, Valencia, CA) as per manufacturer's protocol, and RT-qPCR was performed using SuperScript III Platinum SYBR Green One-Step qRT-PCR kit (Invitrogen, Grand Island, NY). The purified total RNA in water, were normalized and used to generate cDNA. Gene expression was measured using Applied Biosystems 7300 real-time PCR system. The conditions used were 4 minutes at 50 °C, 5 minutes at 95°C, and 40 cycles of 15 seconds at 95 °C and 1 minute at 60 °C. The number of viral RNA copies was determined using a standard curve.

#### 2.3.12 Overexpression of NCAM in HEK 293T cells

HEK 293T cells at the confluency of 60% were transfected with an expression plasmid with NCAM1 (GenScript, OHu00262D), using lipofectamine 2000 reagent. 48 hours post transfection, cells were challenged with Zika virus at MOI of 0.1. The virus was incubated with the cells at 4 °C for 1 hour, and increase in viral attachment was measured using RT-qPCR.

#### 2.3.13 Immunofluorescence

Vero cells or HEK 293T cells (for transfection with NCAM1) were seeded on cover slips in 24-well plate. Cells were washed with PBS and fixed with 3.7% paraformaldehyde for 10 minutes at room temperature. Cells were again washed with PBS three times and blocked with 2% BSA in PBS for 1 hour. Anti-NCAM antibody in blocking solution was incubated with the cells for 1 hour at room temperature. Cells were washed three times and incubated with anti-mouse FITC or anti-mouse Alexa Fluor 488 for 1 hour at room temperature. DAPI staining was

performed for 10 minutes, followed by final three PBS washes. Cover slips were mounted on glass slide and images were captured using Olympus IX81 fluorescence microscope with a 60X oil immersion objective.

#### 2.3.14 Western Blot

Following overexpression of NCAM1 in HEK 293T, cells were lysed at 48 hours post transfection. The samples were boiled at 95 °C in gel loading buffer and 1,4-Dithiothreitol (DTT) for 5 minutes. The cell lysates were separated on the precast NuPAGE 4-12% Bis-Tris polyacrylamide gels (Invitrogen) for 90 minutes at constant voltage of 150V. A MOPS solution (50 mM MOPS, 50 mM Tris-base, 1 mM EDTA, 0.1% SDS) was used as a running buffer. The proteins were transferred onto polyvinylidene fluoride membranes in Bicine-Bis-Tris transfer buffer containing 12% methanol, for 75 minutes at a constant current of 275 mA. The membrane was blocked with 2% BSA in TBST, and probed with anti-human NCAM (Cell Signaling) for 1 hour at room temperature. Following washings, anti-mouse IgG HRP-conjugated secondary antibody (Cell Signaling) was utilized for visualization.

## 2.4 Results and Discussion

### 2.4.1 Design and Synthesis of Labeling Reagent

The labeling reagent is designed so as to be biocompatible with desired aqueous solubility. Maleimide was preferred as the virus 'surface' labeling functionality, due to the minimal labeling of cysteine residues. Zika, like other flavi viruses and enveloped viruses in general, are quite unstable and prone to undergo structural changes under external influence. Considering the virus stability, and in order to ensure virus infectivity, we preferred minimal labeling of virus under mild conditions at neutral pH. The other important functionality is a UV-photocrosslinker, diazirine which was employed due to its high selectivity attribute. The carbene formed after exposing to UV light crosslinks any protein in close-proximity to Zika virus, which is likely receptors or other host proteins from pathways manipulated by the virus in the infection process. The short half-life and reactive nature of carbene ensures a fewer non-selective crosslinking of proteins besides virus-interacting proteins. The two functional moieties are separated by a biocompatible and polar linker to impart flexibility for efficient crosslinking. The third moiety is biotin, required for affinity purification of viral E protein and the crosslinked host proteins,

which were subsequently identified by quantitative proteomics. The synthesis of reagent was performed on the solid phase resin. The identity of intermediate and the final product were confirmed by NMR and Maldi-MS.

#### 2.4.2 Labeling of a standard protein

Bovine serum albumin (BSA) was used as a model protein to test the labeling capability of the reagent. BSA was incubated with 1 mM of reagent in phosphate buffer pH 7, overnight in 4 °C. Reaction was quenched, and biotinylated proteins were enriched on streptavidin beads. The captured proteins were eluted by boiling, and separated on SDS-PAGE. According to quantitation based on band intensities, the labeling efficiency was estimated to be 25-30% (supplementary fig). Separately, prior to labeling, BSA protein sample was pretreated with TCEP and CAA to reduce and alkylate the cysteine residues. As shown in supplementary fig., the BSA labeling by the reagent was blocked by the alkylation of cysteine residues, suggesting the reagent labeling site as the available cysteines.

#### 2.4.3 Labeling of Zika E proteins

Pure Zika virus was labeled using the maleimide-based labeling reagent by gentle end-to-end rotation overnight. After quenching the reaction with excess cysteine for 1 hour, the particles were denatured with 0.2% SDS, and the labeled surface proteins were captured on high capacity streptavidin agarose (Thermo Scientific) for 1 hour at room temperature. The beads were pelleted by gentle centrifugation. Flow through was removed and beads were washed three times with harsh washing condition of 1% SDS for five minutes each. The beads were boiled in gel loading buffer to elute the labeled E proteins, which were separated on gel and visualized by silver stain. We also confirmed the identity of E proteins by immunoblotting using 4G2 antibody and a HRP-conjugated secondary antibody (Supplementary Figure 1C). Note that we used a higher capacity streptavidin beads here to maximize enrichment, with the additional advantage of stringent washing condition to avoid any non-specific binding. We detected no loss of labeled proteins with this washing condition, during our optimizations, which is similar to the reported data (25).

#### 2.4.4 Infectivity testing of labeled Zika Virus

One of the challenges in a chemical proteomic based study of virus-host interactome for flavivirus, an enveloped virus, is their inherent instability to mechanical or chemical agents.

Hence, after chemically labeling the surface proteins of Zika, we confirmed the infectivity of virus particles and compared it with the ‘unlabeled’ virus using plaque assay. No loss of infectivity was observed under labeling conditions for 1 mM reagent concentration.

Furthermore, several unique peptides from E protein of Zika were identified by mass spectrometry with high confidence. The labeled virus diluted in DMEM, when added to Vero cells, successfully entered the cells confirming the infectious nature of labeled Zika. About 11 unique peptides from E protein were identified and quantified across the three time points of infection. Moreover, we did not identify any peptides from membrane (M) protein, capsid, or any of the non-structural proteins of Zika, further confirming the exclusive tagging of virus surface with the reagent. It is primarily owing to the membrane impermeable attribute of the reagent imparted by the PEGylated linkers. Hence, we concluded that the ‘minimal’ labeling of virus achieved by cysteine-reactive maleimide group does not perturb the infectivity of the virus.

#### 2.4.5 Identification of Zika receptors and other host factors/ direct interactors with Zika virus

Flavi virus are quite promiscuous in their selection of receptors for entry to different cells (26). The complex entry mechanism might involve multiple receptor interactions to help virus internalize. Though some previous studies have identified AXL, TIM, and TAM to be putative receptors for Zika, some conflicting evidences suggest that the virus might also employ some different class of receptors. Furthermore, while most of the virus are believed to enter cells by Clathrin-mediated endocytosis, there is no evidence suggesting the absence of any parallel mode of virus entry. The complexity of flavi virus entry mechanism hints at the presence of varied virus-protein interactions after recruitment of host cellular proteins following viral infection. In order to identify receptors, we allowed the virus to attach to the cells in 4 °C for 1 hour, immediately followed by UV photocrosslinking. For the virus entry, we chose 4 and 8 minutes as previous study indicates the localization of a similar flavi virus to late endosomes in 5-6 minutes post incubation at 37 °C (11). After crosslinking proteins at certain time points of attachment or entry, cells were harvested and proteins were extracted followed by enrichment using neutravidin beads. Proteins were identified by shotgun proteomic strategy, and quantitated using Label-free quantitation. In total we identified about 300 crosslinked proteins, out of which more than 70 proteins are implicated in virus infection across different time points (Figure 2). Notable we identified a putative receptor at 0 minutes, Neural Cell Adhesion Molecule (NCAM1), which is

also a receptor for rabies virus (27). Besides, we identified integrin alpha-3 (ITGA3), at 0 and 4 minutes of infection. Integrins are highly conserved transmembrane glycoproteins which serve as receptors for herpes virus 8 (HHV-8), herpes simplex virus 1 (HHV-1), Human cytomegalovirus (HCMV), adenoviruses, coxsackievirus A9 and B1, Foot-and-mouth disease virus, Hantavirus, Rotavirus, and human echovirus 1 (28). A relation between West Nile, a flavi virus, infectivity and integrins has also been established (29). Another protein CD81 was identified with high confidence only at 0 minutes, suggesting its interaction with the Zika E protein. CD81 is a well-known co-receptor for HCV. We posit that NCAM1, integrins and CD81 are putative attachment factors or receptors for Zika virus (Figure 3).

Furthermore, we identified RPSA at all the three time points with high confidence, indicating its possible role in the early stage infection of Zika virus. Ribosomal protein SA (RPSA) is a protein found in different cellular compartments, and serves a wide range of different functions. One of them is its receptor activity for laminin, hence also termed as a laminin receptor. Besides laminin, the protein has also been shown to act as a receptor for a wide range of viruses including Sindbis virus (30), Dengue virus (31, 32), Venezuelan equine encephalitis virus (33), West Nile virus (34), Japanese encephalitis virus (35) and various serotypes of Adeno-associated virus (36).

A few other proteins crosslinked at 0 minutes are AP2M1 and CALM1. The  $\mu$  domain of adaptor protein-2 complex interacts with the tyrosine-rich or dileucine motifs on cytoplasmic tail of receptor proteins. AP2M1 has further been shown to be modulate infectious entry of many more viruses including . CALM1 was also reported by a recently published study on Zika host factors.

Synaptotagmins are another class of membrane trafficking proteins which are characterized by N-terminal transmembrane domain and two C2 domains at C-terminus, responsible for calcium-dependent phospholipid binding. A related class of proteins recently discovered is called extended-synaptotagmins (E-Syts) marked by three or five C2 domains (37). Out of the three extended-synaptotagmins (E-Syt1, 2, and 3), E-Syt2 is essentially required for clathrin-mediated endocytosis of fibroblast growth factors receptors (FGFR) (38). In the same study, the E-Syt2 was found to interact selectively with activated FGFR, EGFR (epidermal growth factor), and adaptor protein AP-2. We identified E-Syt2 exclusively at 4 and 8 minutes of Zika infection, suggesting the possible role of this transmembrane protein in Zika endocytosis.

At 8 minutes of infection, we identified an isoform of a key component of recycling endosomes, RAB11A. The role of RAB11A in transport of viral ribonucleoprotein (vRNP) or core proteins to plasma membrane for the generation of new virus particles of influenza and HCV respectively, is well documented (39). However, some virus may also utilize recycling endosomal pathway to evade lysosomal degradation. Our study suggests the use of recycling pathway by Zika virus after entry. Besides Rab11, we also observed RAB5A at 8 minutes. Several studies have established the importance of Rab5 in the entry of Zika virus, Dengue virus, and West Nile virus (40, 41). Our observation for Rab5 and Rab11 at late time point of entry is consistent with the published data on Japanese Encephalitis virus (JEV), another neurovirulent pathogen from flavi virus family structurally similar to the Zika (42).

HSPA8, also known as Hsc70 (heat shock cognate 70), is known for its role in the later stage of endocytosis, more precisely in vesicle uncoating. In our study, It was crosslinked only at 8 minutes of Zika entry, demonstrating clathrin-mediated pathway employed by the virus to infect Vero cells.

We also identified ARPC3, a component of ARP 2/3 complex known to mediate actin polymerization, at a late time point of infection (43). ARP 2/3 complex is critical for the human immunodeficiency virus (HIV-1) infection, as its inhibition limits the viral infection. A similar reduction in infection of DENV was reported after siRNA knockdown of ARPC1B, another component of the ARP 2/3 complex emphasizing its critical role in flavi virus infection (44-47). Studies hint at the role of this complex at a late stage in Clathrin-mediated endocytosis, which is in accordance with our observation.

#### 2.4.6 NCAM is a receptor for Zika virus

Since NCAM1 is abundantly expressed in brain and Zika causes neurological disorders, we tested NCAM as a potential candidate for Zika receptor. As NCAM is expressed on the plasma membrane, and hence was crosslinked at 0 minutes, we investigated if antibody blocking of the NCAM causes any reduction in the Zika infectivity. For this purpose, we employed CD56 antibody which binds to the extracellular immunoglobulin-like domains present in all three isoforms of the NCAM, and is recommended for the functional assays. We first tested the ability of the antibody to bind to the Vero cell membrane using immunofluorescence without membrane permeabilization, and observed strong fluorescence signal (Figure 4A). Then, we incubated Vero

cells with anti-NCAM or control IgG for 45 minutes at room temperature, prior to infection by Zika at MOI of 0.1. 24 hours post-infection, cellular RNAs were extracted and Zika RNAs were quantitated using qRT-PCR. We noticed a significant reduction in internalized Zika RNAs, across three biological replicates, which indicate NCAM as a host factor involved in Zika infection (Figure 4B).

To further confirm the receptor activity of NCAM, we overexpressed NCAM1 in HEK 293T cells and performed attachment assays with Zika virus. HEK 293T cells lack NCAM protein and have minimal infectivity by Zika (22), which makes them an ideal system to test the receptor activity for Zika virus. HEK 293T cells were, initially, transfected with expression plasmid encoding NCAM1, and the protein overexpression was confirmed by Western blotting 48 hours post-transfection (Figure 5A). Furthermore, we tested the localization of overexpressed NCAM using immunofluorescence and found it on the membrane of transfected cells (Figure 5B). Two days post-transfection, HEK 293T cells were incubated with Zika virus for 1 hour at 4 °C. Cells were washed and harvested, and qRT-PCR was used to quantitate Zika RNAs. We found a two to three fold increase, confirming NCAM as a Zika receptor (Figure 5C).

## 2.5 References

- (1) Dick, G.; Kitchen, S.; Haddock, A. Zika Virus (I). Isolations and serological specificity. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 1952, *46*, 509-520.
- (2) World Health Organization. Zika strategic response framework and joint operations plan, January-June 2016.
- (3) Talan, J. Epidemiologists Are Tracking Possible Links Between Zika Virus, Microcephaly, and Guillain–Barré Syndrome. *Neurology Today* 2016, *16*, 1.
- (4) Cauchemez, S.; Besnard, M.; Bompard, P.; Dub, T.; Guillemette-Artur, P.; Eyrolle-Guignot, D.; Salje, H.; Van Kerkhove, M.; Abadie, V.; Garel, C. et al. Association Between Zika Virus and Microcephaly in French Polynesia, 2013–2015. *Obstetrical & Gynecological Survey* 2016, *71*, 512-514.
- (5) Flaviviridae: The Viruses and Their Replication. Brett D. Lindenbach Heinz Jurgen Thiel Charles M. Rice.
- (6) Modis, Y.; Ogata, S.; Clements, D.; Harrison, S. Structure of the dengue virus envelope protein after membrane fusion. *Nature* 2004, *427*, 313-319.
- (7) Sirohi, D.; Chen, Z.; Sun, L.; Klose, T.; Pierson, T.; Rossmann, M.; Kuhn, R. The 3.8 Å resolution cryo-EM structure of Zika virus. *Science* 2016, *352*, 467-470.
- (8) Kostyuchenko, V.; Lim, E.; Zhang, S.; Fibriansah, G.; Ng, T.; Ooi, J.; Shi, J.; Lok, S. Structure of the thermally stable Zika virus. *Nature* 2016, *533*, 425-428.

- (9) Sapparapu, G.; Fernandez, E.; Kose, N.; Bin Cao; Fox, J.; Bombardi, R.; Zhao, H.; Nelson, C.; Bryan, A.; Barnes, T. et al. Neutralizing human antibodies prevent Zika virus replication and fetal disease in mice. *Nature* 2016, *540*, 443-447.
- (10) Sakurai, K.; Ozawa, S.; Yamada, R.; Yasui, T.; Mizuno, S. Comparison Of The Reactivity Of Carbohydrate Photoaffinity Probes With Different Photoreactive Groups. *ChemBioChem* 2014, *15*, 1399-1403.
- (11) van der Schaar, H.; Rust, M.; Chen, C.; van der Ende-Metselaar, H.; Wilschut, J.; Zhuang, X.; Smit, J. Dissecting The Cell Entry Pathway Of Dengue Virus By Single-Particle Tracking In Living Cells. *PLoS Pathogens* 2008, *4*, e1000244.
- (12) Elschenbroich, S.; Kim, Y.; Medin, J.; Kislinger, T. Isolation Of Cell Surface Proteins For Mass Spectrometry-Based Proteomics. *Expert Review of Proteomics* 2010, *7*, 141-154.
- (13) Helbig, A.; Heck, A.; Slijper, M. Exploring The Membrane Proteome—Challenges And Analytical Strategies. *Journal of Proteomics* 2010, *73*, 868-878.
- (14) Frei, A.; Jeon, O.; Kilcher, S.; Moest, H.; Henning, L.; Jost, C.; Plückthun, A.; Mercer, J.; Aebersold, R.; Carreira, E. et al. Direct Identification Of Ligand-Receptor Interactions On Living Cells And Tissues. *Nature Biotechnology* 2012, *30*, 997-1001.
- (15) Mukherjee, S.; Sirohi, D.; Dowd, K.; Chen, Z.; Diamond, M.; Kuhn, R.; Pierson, T. Enhancing Dengue Virus Maturation Using A Stable Furin Over-Expressing Cell Line. *Virology* 2016, *497*, 33-40.
- (16) Sirohi, D.; Chen, Z.; Sun, L.; Klose, T.; Pierson, T.; Rossmann, M.; Kuhn, R. The 3.8 Å resolution cryo-EM structure of Zika virus. *Science* 2016, *352*, 467-470.
- (17) Kuhn, R.J. et. al. *Cell*. 2002 Mar 8;108(5):717-25.
- (18) Attenuation markers of a candidate dengue type 2 vaccine virus, strain 16681 (PDK-53), are defined by mutations in the 5' noncoding region and nonstructural proteins 1 and 3. *J Virol*. 2000 Apr;74(7):3011-9.
- (19) Wang, L.; Pan, L.; Tao, W. Specific Visualization And Identification Of Phosphoproteome In Gels. *Analytical Chemistry* 2014, *86*, 6741-6747.
- (20) Cox, J.; Mann, M. Maxquant Enables High Peptide Identification Rates, Individualized P.P.B.-Range Mass Accuracies And Proteome-Wide Protein Quantification. *Nature Biotechnology* 2008, *26*, 1367-1372.
- (21) Cox, J.; Hein, M.; Lubner, C.; Paron, I.; Nagaraj, N.; Mann, M. Accurate Proteome-Wide Label-Free Quantification By Delayed Normalization And Maximal Peptide Ratio Extraction, Termed Maxlfr. *Molecular & Cellular Proteomics* 2014, *13*, 2513-2526.
- (22) Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R.; Olsen, J.; Mann, M. Andromeda: A Peptide Search Engine Integrated Into The Maxquant Environment. *Journal of Proteome Research* 2011, *10*, 1794-1805.
- (23) Tyanova, S.; Temu, T.; Sinitcyn, P.; Carlson, A.; Hein, M.; Geiger, T.; Mann, M.; Cox, J. The Perseus Computational Platform For Comprehensive Analysis Of (Prote)Omics Data. *Nature Methods* 2016, *13*, 731-740.
- (24) Richard, A.; Shim, B.; Kwon, Y.; Zhang, R.; Otsuka, Y.; Schmitt, K.; Berri, F.; Diamond, M.; Choe, H. AXL-Dependent Infection Of Human Fetal Endothelial Cells Distinguishes Zika Virus From Other Pathogenic Flaviviruses. *Proceedings of the National Academy of Sciences* 2017, *114*, 2024-2029.

- (25) Hulce, J.; Cognetta, A.; Niphakis, M.; Tully, S.; Cravatt, B. Proteome-Wide Mapping Of Cholesterol-Interacting Proteins In Mammalian Cells. *Nature Methods* 2013, *10*, 259-264.
- (26) Biology of Zika Virus Infection in Human Skin Cells.
- (27) The neural cell adhesion molecule is a receptor for rabies virus.
- (28) Stewart, P.; Nemerow, G. Cell integrins: commonly used receptors for diverse viral pathogens. *Trends in Microbiology* 2007, *15*, 500-507.
- (29) Schmidt, K.; Keller, M.; Bader, B.; Korytar, T.; Finke, S.; Ziegler, U.; Groschup, M. Integrins modulate the infection efficiency of West Nile virus into cells. *Journal of General Virology* 2013, *94*, 1723-1733.
- (30) Wang, K. S., Kuhn, R. J., Strauss, E. G., Ou, S., and Strauss, J. H. (1992) High-affinity laminin receptor is a receptor for Sindbis virus in mammalian cells. *J. Virol.* 66, 4992–5001.
- (31) Thepparit, C., and Smith, D. R. (2004) Serotype-specific entry of dengue virus into liver cells: Identification of the 37-kilodalton/67-kilodalton high-affinity laminin receptor as a dengue virus serotype 1 receptor. *J. Virol.* 78, 12647–12656.
- (32) Tio, P. H., Jong, W. W., and Cardosa, M. J. (2005) Two dimensional VOPBA reveals laminin receptor (LAMR1) interaction with dengue virus serotypes 1, 2 and 3. *Virol. J.* 2, 25.
- (33) Ludwig, G. V., Kondig, J. P. & Smith, J. F. (1996). A putative receptor for Venezuelan equine encephalitis virus from mosquito cells. *Journal of Virology* 70, 5592–5599.
- (34) Bogachek, M. V., Protopopova, E. V., Loktev, V. B., Zaitsev, B. N., Favre, M., Sekatskii, S. K. & Dietler, G. (2008). Immunochemical and single molecule force spectroscopy studies of specific interaction between the laminin binding protein and the West Nile virus surface glycoprotein E domain II. *Journal of Molecular Recognition* 21, 55–62.
- (35) Thongtan, T., Wikan, N., Wintachai, P., Rattanarungsan, C., Srisomsap, C., Cheepsunthorn, P. & Smith, D. R. (2012). Characterization of putative Japanese encephalitis virus receptor molecules on microglial cells. *Journal of Medical Virology* 84, 615–623.
- (36) Akache, B., Grimm, D., Pandey, K., Yant, S. R., Xu, H., and Kay, M. A. (2006) The 37/67-kilodalton laminin receptor is a receptor for adeno-associated virus serotypes 8, 2, 3, and 9. *J. Virol.* 80, 9831–9836.
- (37) Herdman, C.; Moss, T. Extended-Synaptotagmins (E-Syts); the extended story. *Pharmacological Research* 2016, *107*, 48-56.
- (38) Jean, S.; Mikryukov, A.; Tremblay, M.; Baril, J.; Guillou, F.; Bellenfant, S.; Moss, T. Extended-Synaptotagmin-2 Mediates FGF Receptor Endocytosis and ERK Activation In Vivo. *Developmental Cell* 2010, *19*, 426-439.
- (39) Einfeld, A.; Kawakami, E.; Watanabe, T.; Neumann, G.; Kawaoka, Y. RAB11A Is Essential for Transport of the Influenza Virus Genome to the Plasma Membrane. *Journal of Virology* 2011, *85*, 6117-6126.
- (40) Meertens, L.; Labeau, A.; Dejarnac, O.; Cipriani, S.; Sinigaglia, L.; Bonnet-Madin, L.; Le Charpentier, T.; Hafirassou, M.; Zamborlini, A.; Cao-Lormeau, V. et al. Axl Mediates ZIKA Virus Entry In Human Glial Cells And Modulates Innate Immune Responses. *Cell Reports* 2017, *18*, 324-333.

- (41) Krishnan, M.; Sukumaran, B.; Pal, U.; Agaisse, H.; Murray, J.; Hodge, T.; Fikrig, E. Rab 5 Is Required For The Cellular Entry Of Dengue And West Nile Viruses. *Journal of Virology* 2007, 81, 4881-4885.
- (42) Liu, C.; Zhang, Y.; Li, Z.; Hou, J.; Zhou, J.; Kan, L.; Zhou, B.; Chen, P. Rab5 And Rab11 Are Required For Clathrin-Dependent Endocytosis Of Japanese Encephalitis Virus In BHK-21 Cells. *Journal of Virology* 2017, 91, e01113-17.
- (43) Collier, K.; Heaton, N.; Berger, K.; Cooper, J.; Saunders, J.; Randall, G. Molecular Determinants and Dynamics of Hepatitis C Virus Secretion. *PLoS Pathogens* 2012, 8, e1002466.
- (44) Kadiu, I.; Gendelman, H. Human Immunodeficiency Virus type 1 Endocytic Trafficking Through Macrophage Bridging Conduits Facilitates Spread of Infection. *Journal of Neuroimmune Pharmacology* 2011, 6, 658-675.
- (45) Komano, J. Inhibiting the Arp2/3 Complex Limits Infection of Both Intracellular Mature Vaccinia Virus and Primate Lentiviruses. *Molecular Biology of the Cell* 2004, 15, 5197-5207.
- (46) Ang, F.; Wong, A.; Ng, M.; Chu, J. Small interference RNA profiling reveals the essential role of human membrane trafficking genes in mediating the infectious entry of dengue virus. *Virology Journal* 2010, 7, 24.
- (47) Merrifield, C. Neural Wiskott Aldrich Syndrome Protein (N-WASP) and the Arp2/3 complex are recruited to sites of clathrin-mediated endocytosis in cultured fibroblasts. *European Journal of Cell Biology* 2004, 83, 13-18.

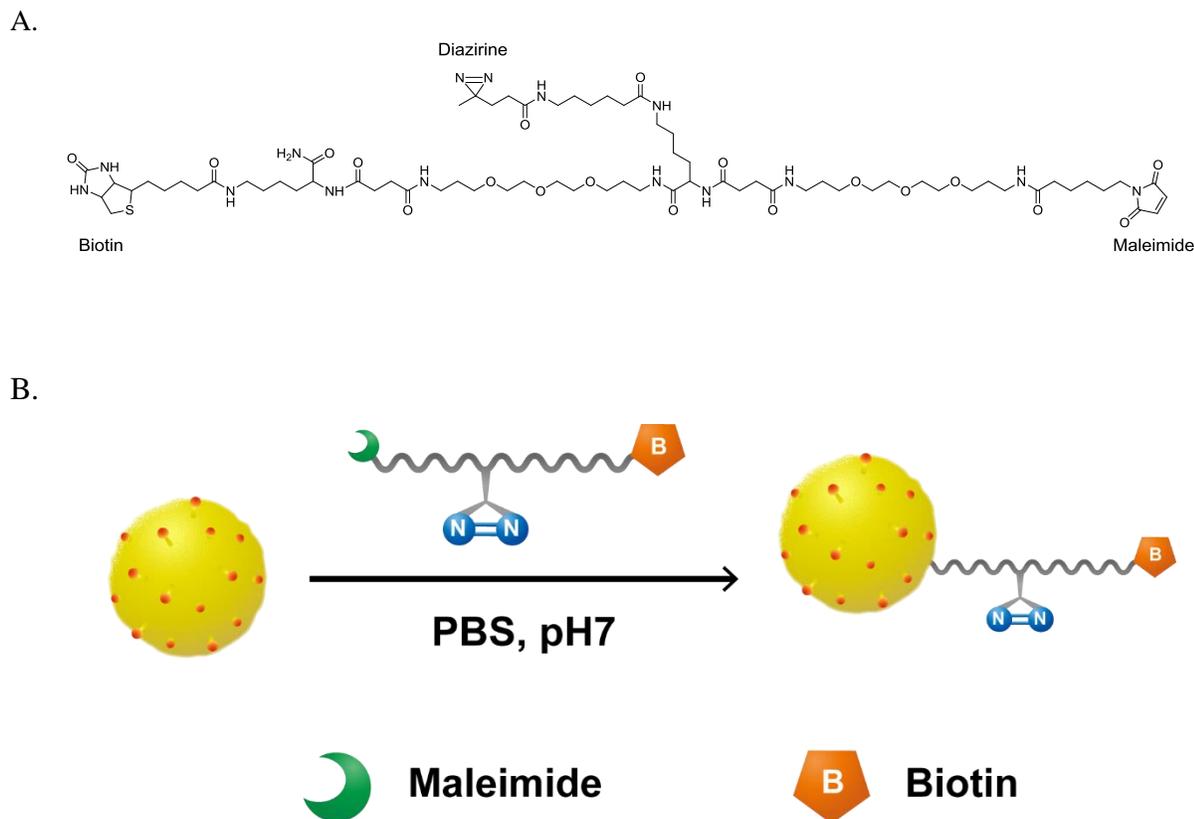
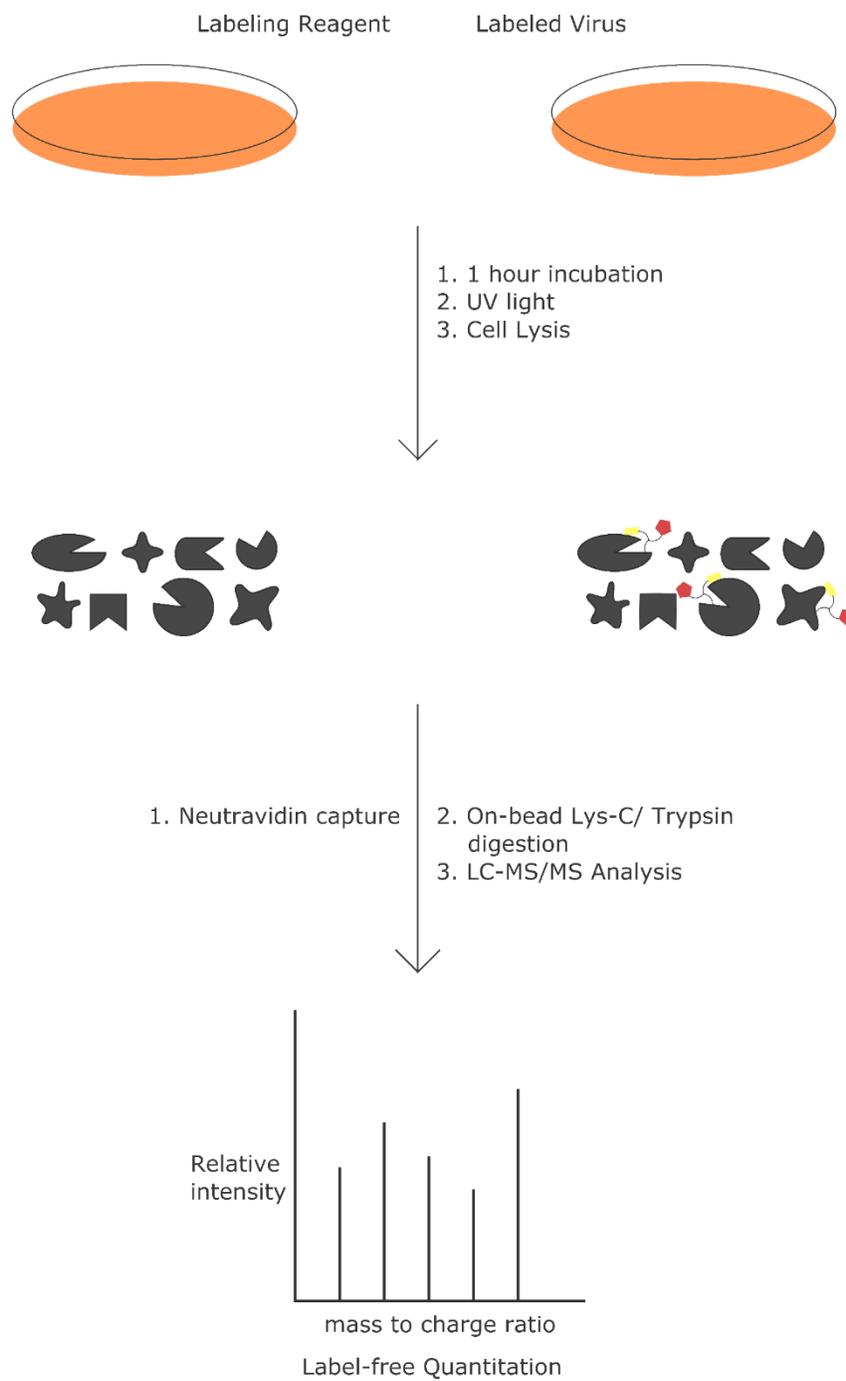


Figure 2.1: Trifunctionalized reagent designed to label Zika ‘surface’ E proteins and capture virus-interacting proteins. (A) Design of Virus labeling reagent. Maleimide reacts with available cysteines on the virus surface under mild conditions, diazirine enables crosslinking proteins at fixed time points allowing tracing virus movement in ‘real-time’, and biotin acts as a handle for protein enrichment and downstream analysis. The three functionalities are separated by membrane-impermeable polyethylene glycol (PEG)- like linker offering the flexibility required for capturing interacting proteins. (B) Labeling of Zika E proteins. Purified zika virus was diluted in PBS pH 7 and rotated overnight in 4 °C with the labeling reagent. Reaction was quenched with three-fold excess cysteine for 1 hour. (C) Design and workflow for capturing virus receptor and mapping its entry. Labeled zika was diluted in DMEM and incubated with confluent cells for 1 hour at 4 °C. Additionally, cells were incubated in 37 °C for fixed time points to allow virus entry. Unbound virus was removed and cells were directly exposed to UV light, following which cells were harvested by scraping. Cells were lysed, and crosslinked proteins were captured on the neutravidin beads. Proteins were digested on-beads, using sequential Lys-C and Trypsin digestion, and analyzed by LC-MS/MS on an Orbitrap Velos Pro. Quantitation was performed using MaxQuant, and proteins identified uniquely in the UV compared to control, with FDR of 1%, were considered as crosslinked proteins.

Figure 2.1 Continued

C.



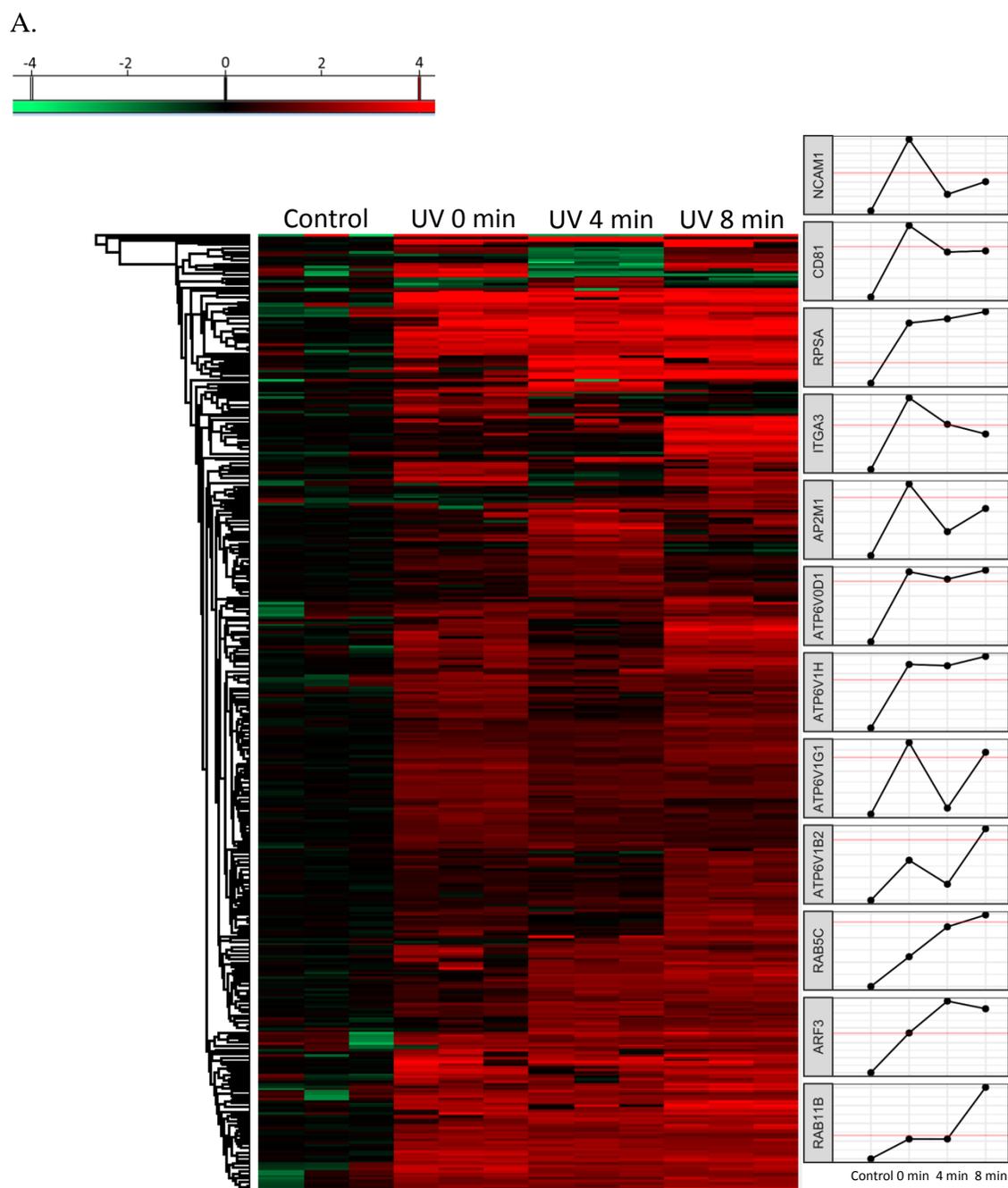
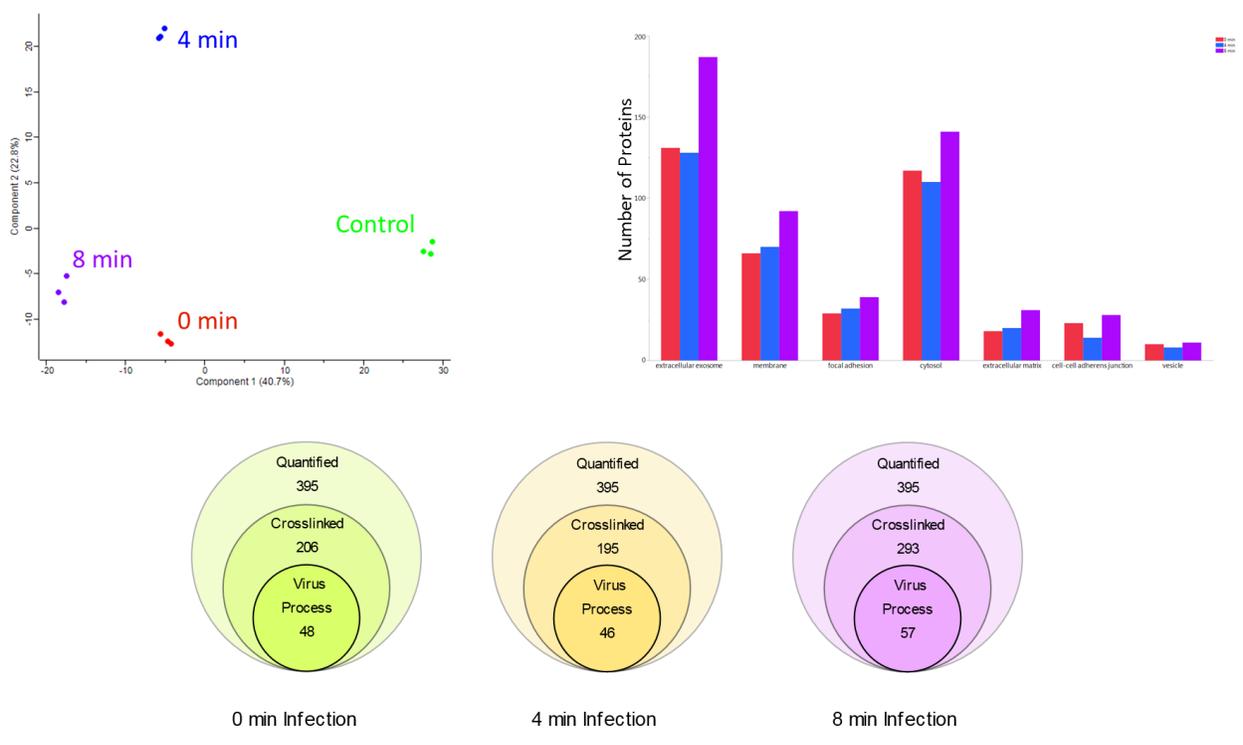


Figure 2.2: Proteins crosslinked at different time points of infection. (A) Heat map for proteins crosslinked at 0, 4, and 8 minutes of infection (FDR 0.05) normalized to the control. Proteins in the significant region with ratios of 2.5 and higher for the UV at different time points compared to the control, are considered as crosslinked proteins and used for further analysis. (B) Principal Component Analysis, Cellular Component Distribution, and number of proteins implicated in viral infection (by Ingenuity Pathway Analysis) at different time point of infection.

Figure 2.2 Continued

B.



A.

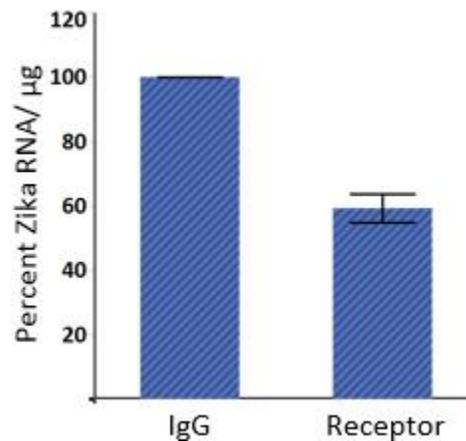
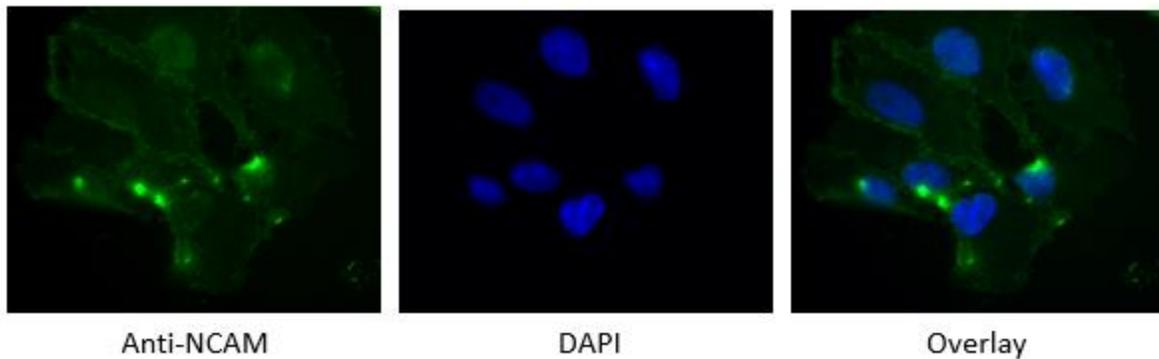
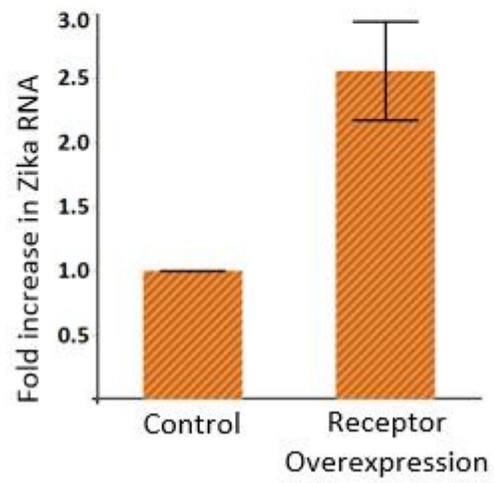
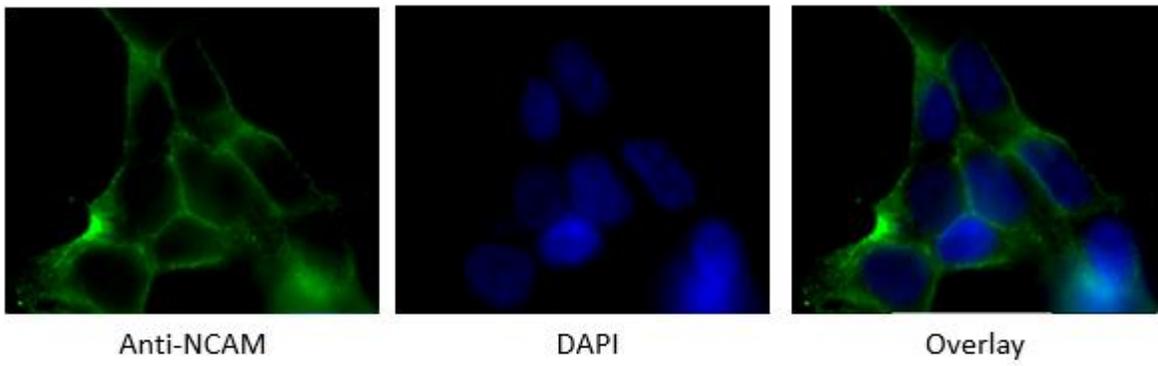


Figure 2.3: NCAM is a receptor for Zika Virus (A) Anti-NCAM inhibits Zika infection of Vero cells. Immunofluorescence showing the binding of anti-NCAM antibody to Vero cells. After antibody inhibition of NCAM, about 50% reduction in viral infectivity was observed. (B) Overexpression of NCAM1 in HEK293T increases Zika infection. Expression plasmid with NCAM1 was transfected into HEK293T cells. 48 hours after transfection, immunofluorescence was used to confirm surface expression of NCAM. qRT-PCR revealed an increase in Zika RNAs after NCAM overexpression.

Figure 2.3 Continued

B.



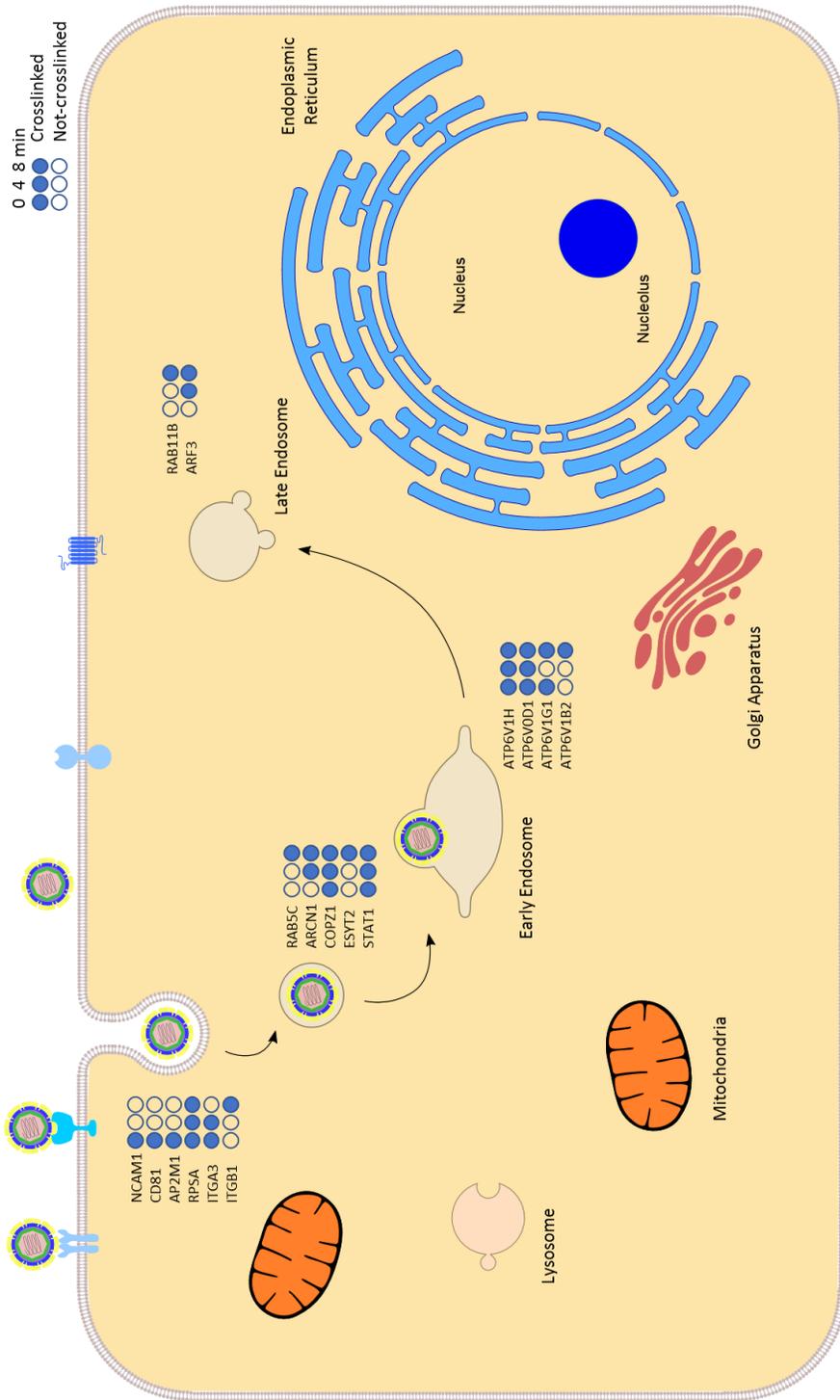
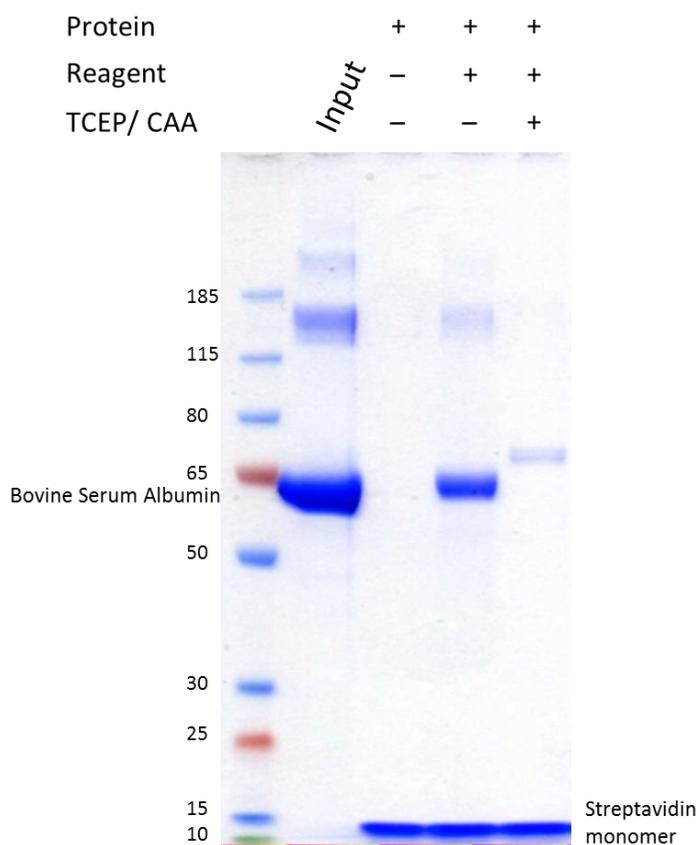


Figure 2.4: Proposed Endocytosis pathway of Zika entry to host cells, highlighting proteins crosslinked at different time point.

A.



B.

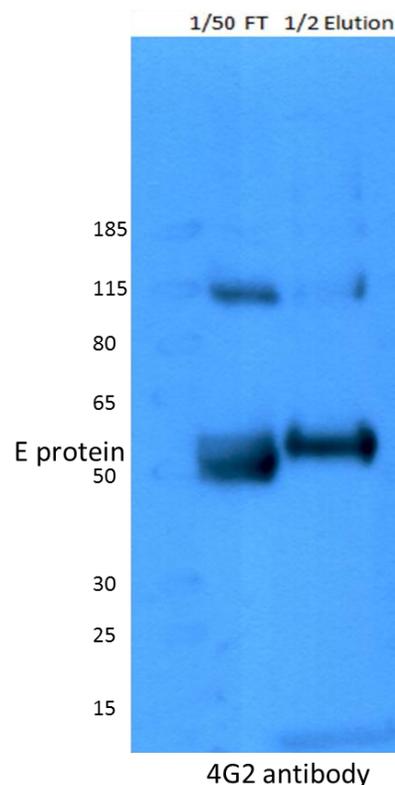
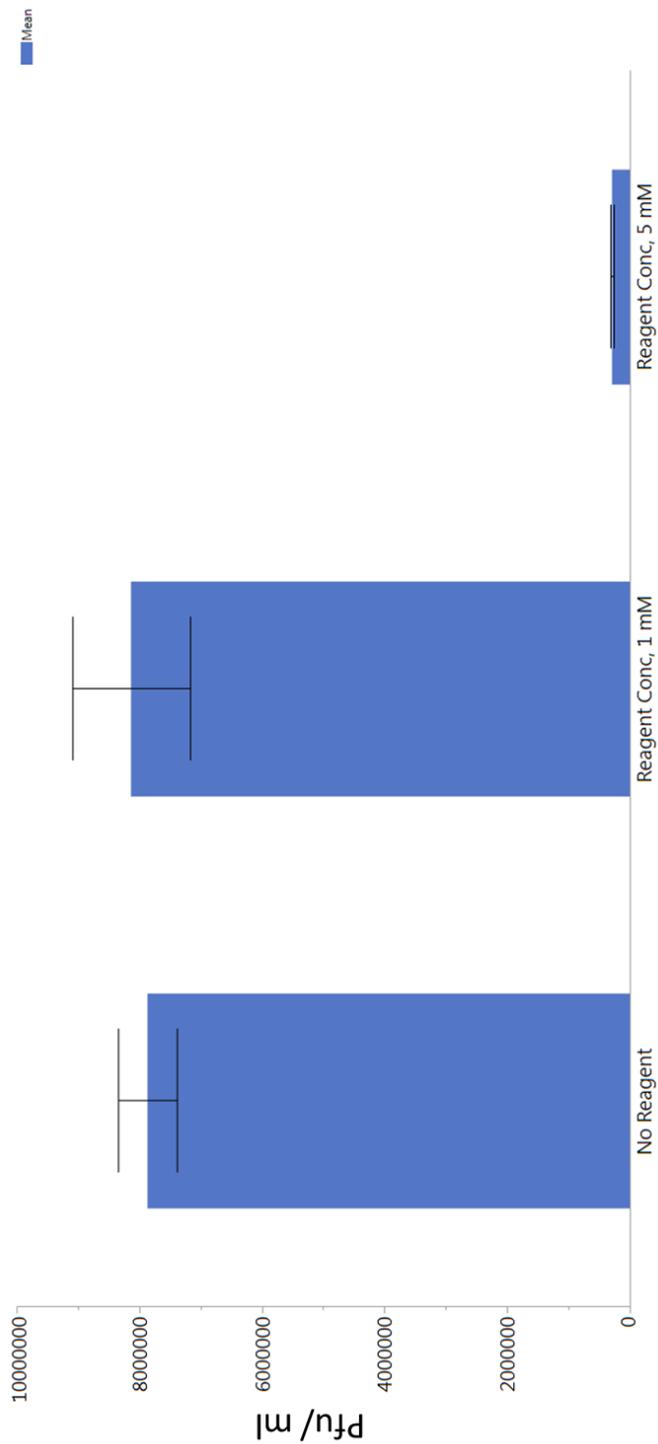


Figure 2.5: Protein labeling by reagent. (A) Standard protein Bovine Serum Albumin (BSA) was labeled with the reagent in phosphate buffer pH 7 overnight, and the reaction was quenched with excess cysteine for 1 hour. The labeled proteins were enriched on streptavidin beads, followed by three 1% SDS washes and elution by boiling. The reagent labeling efficiency was quantified by comparing the protein band intensity with the input. A no-reagent control was employed to account for non-specific binders. In parallel, the BSA was reduced and alkylated prior to labeling, to confirm the labeling site on the protein by the maleimide-diazirine-biotin reagent. (B) Purified zika virus was labeled by the reagent using the same protocol as above. (C) The infectivity of labeled virus was tested by plaque assay. The number of plaque forming units (pfu) after virus labeling were compared with the unlabeled virus.

C.

Figure 2.5 Continued



A.

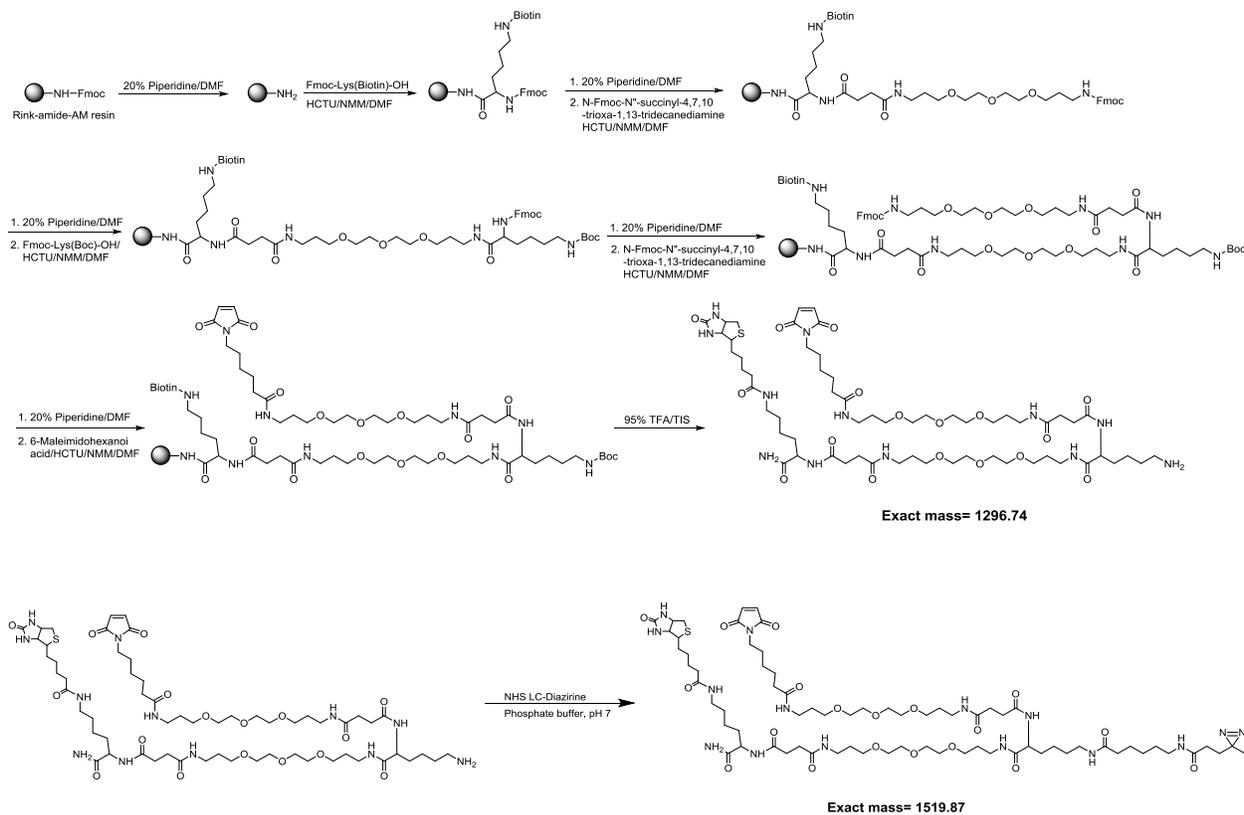
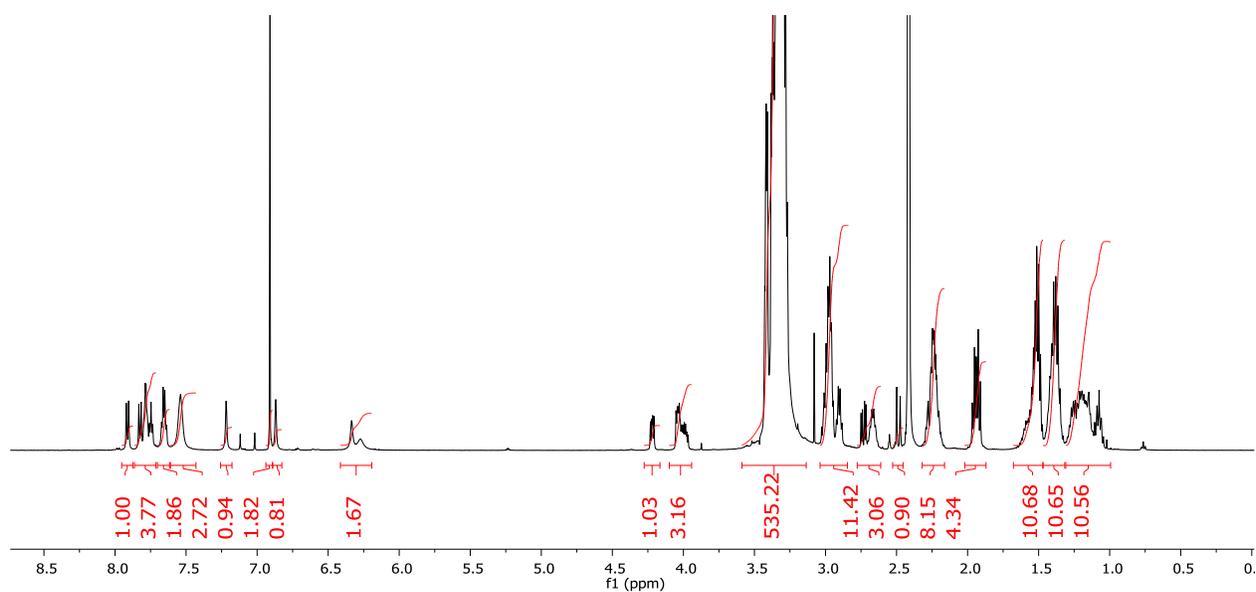


Figure 2.6: Synthesis, HPLC purification, and Characterization of the labeling reagent. (A) Synthesis was performed on the solid support, Rink-amide-AM resin, and cleaved from the resin using 95% TFA. (B and C)  $^1\text{H}$  NMR,  $^{13}\text{C}$  NMR and MALDI characterizations of Maleimide-Biotin and Maleimide-Biotin-Diazirine respectively.

Figure 2.6 Continued

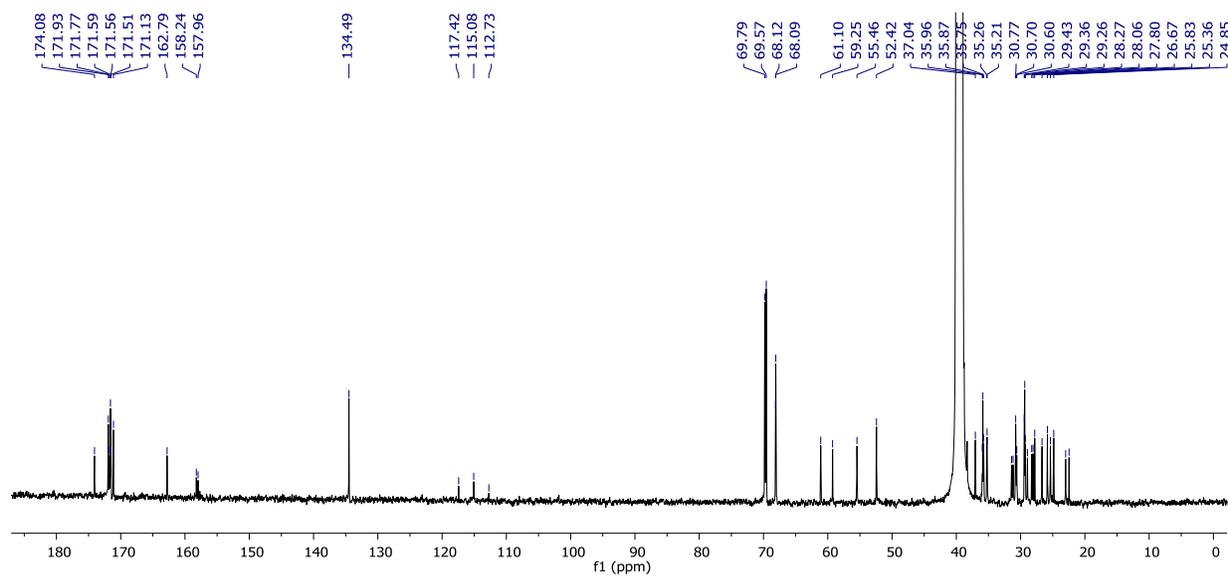
B.

Maleimide-Biotin



<sup>1</sup>H NMR (500 MHz, DMSO-*d*<sub>6</sub>) δ 7.91 (d, *J* = 8.1 Hz, 1H), 7.86 – 7.71 (m, 4H), 7.66 (q, *J* = 5.9 Hz, 2H), 7.54 (bs, 3H), 7.26 – 7.18 (m, 1H), 6.91 (s, 2H), 6.89 – 6.83 (m, 1H), 6.42 – 6.20 (m, 2H), 4.28 – 4.16 (m, 1H), 4.10 – 3.94 (m, 3H), 3.49 – 3.18 (m, 24H), 3.04 – 2.85 (m, 11H), 2.70 (dd, *J* = 15, 5.6 Hz, 1H), 2.71 – 2.61 (m, 2H), 2.49 (d, *J* = 12.4 Hz, 1H), 2.33 – 2.16 (m, 8H), 1.94 (dt, *J* = 13.4, 7.4 Hz, 4H), 1.68 – 1.47 (m, 11H), 1.46 – 1.32 (m, 11H), 1.31 – 0.99 (m, 11H).

Figure 2.6 Continued



<sup>13</sup>C NMR (126 MHz, DMSO)  $\delta$  174.1, 172.0, 171.8, 171.6, 171.5, 171.1, 162.8, 158.2, 158.0, 134.5, 117.4, 115.1, 112.7, 69.8, 69.6, 68.1, 68.1, 61.1, 59.3, 55.5, 52.4, 37.0, 36.0, 35.9, 35.8, 35.3, 35.2, 31.4, 31.2, 30.8, 30.7, 30.6, 29.4, 29.4, 29.3, 29.0, 28.3, 28.1, 27.8, 26.7, 25.8, 25.4, 24.9, 23.0, 22.5.

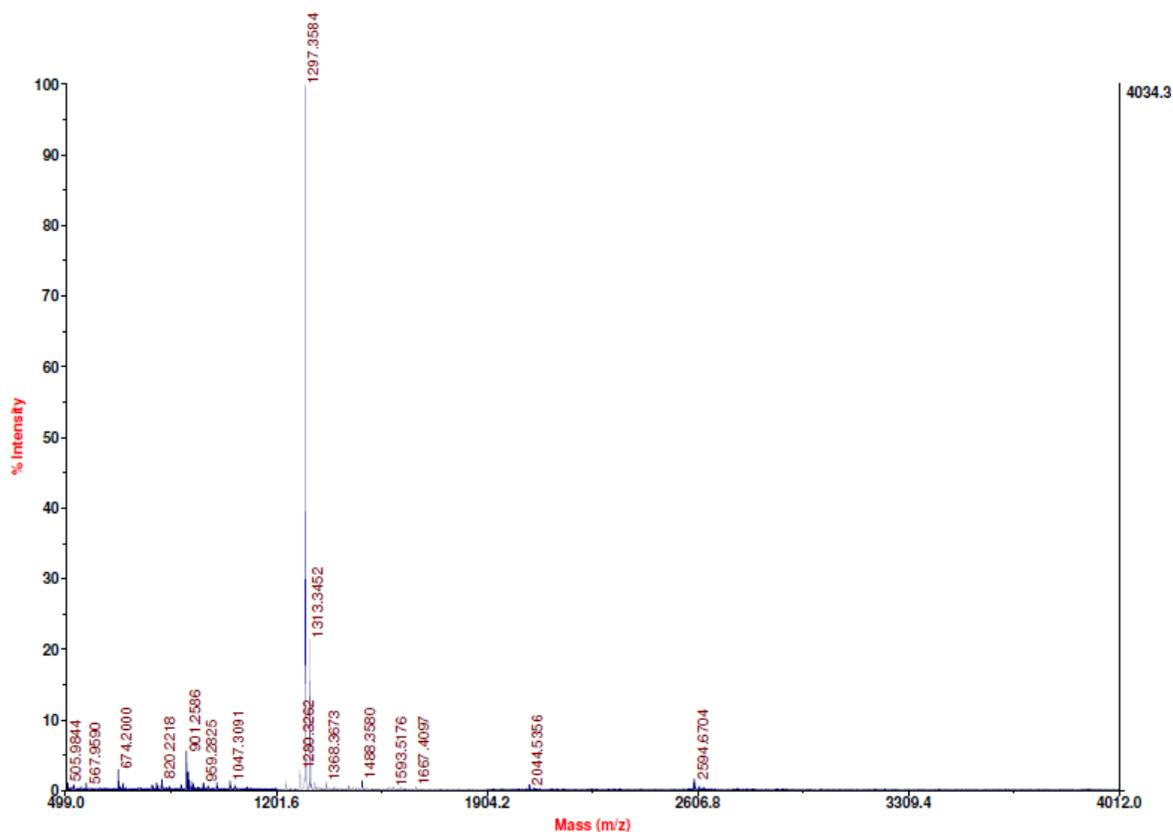
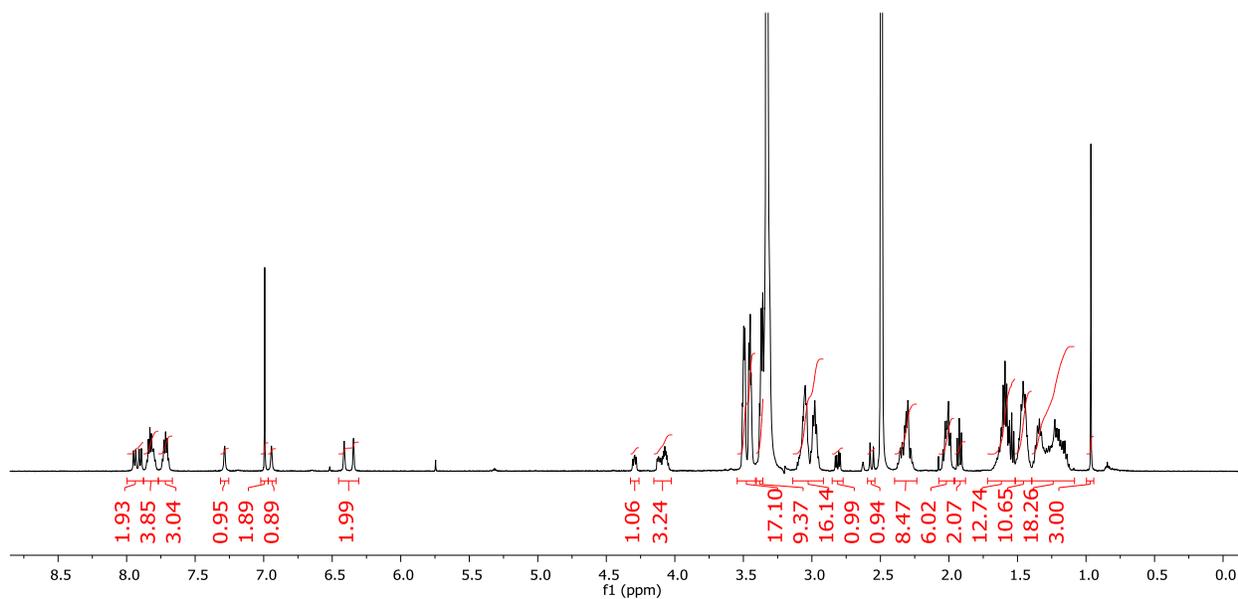


Figure 2.6 Continued

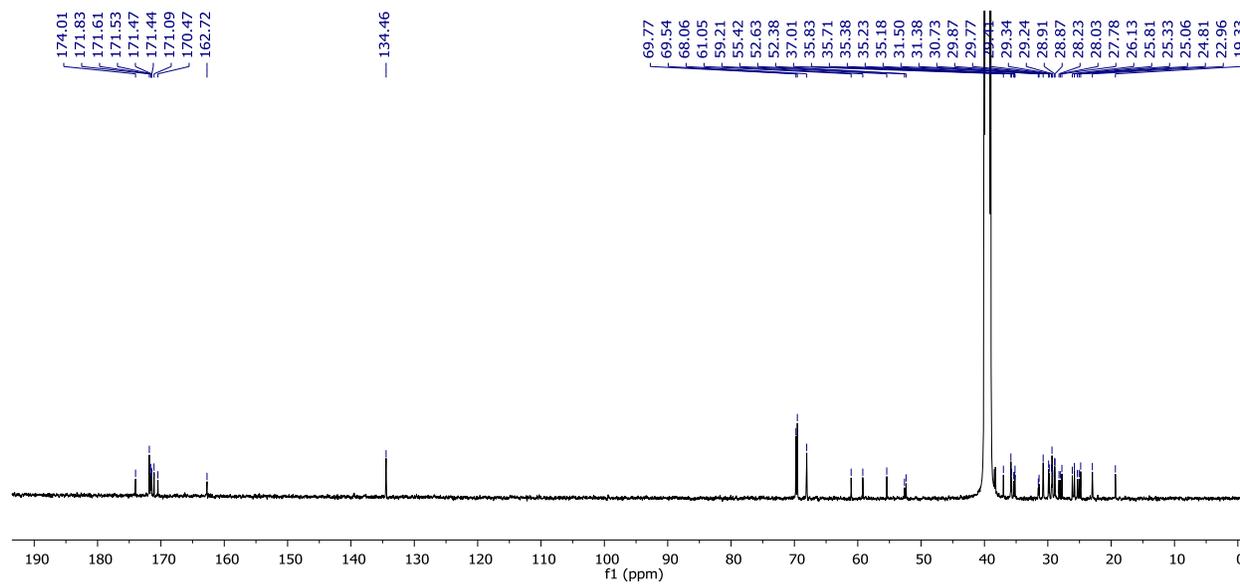
C.

Maleimide-Biotin-Diazirine



<sup>1</sup>H NMR (500 MHz, DMSO-*d*<sub>6</sub>) δ 8.00 – 7.88 (m, 2H), 7.88 – 7.77 (m, 4H), 7.72 (p, *J* = 5.8 Hz, 3H), 7.28 (p, *J* = 2.7, 2.2 Hz, 1H), 6.99 (s, 2H), 6.94 (s, 1H), 6.45 – 6.31 (m, 2H), 4.32 – 4.26 (m, 1H), 4.15 – 4.03 (m, 3H), 3.54 – 3.41 (m, 17H), 3.40 – 3.24 (m, 7H), 3.14 – 2.92 (m, 16H), 2.81 (dd, *J* = 12.4, 5.1 Hz, 1H), 2.56 (d, *J* = 12.4 Hz, 1H), 2.40 – 2.23 (m, 8H), 2.06 – 1.96 (m, 6H), 1.96 – 1.88 (m, 2H), 1.72 – 1.52 (m, 13H), 1.52 – 1.40 (m, 10H), 1.40 – 1.08 (m, 18H), 0.96 (s, 3H).

Figure 2.6 Continued



$^{13}\text{C}$  NMR (126 MHz, DMSO)  $\delta$  174.0, 171.8, 171.6, 171.5, 171.5, 171.4, 171.1, 170.5, 162.7, 134.5, 69.8, 69.5, 68.1, 61.1, 59.2, 55.4, 52.6, 52.4, 37.0, 35.8, 35.7, 35.4, 35.2, 35.2, 31.5, 31.4, 30.7, 29.8, 29.4, 29.3, 29.2, 28.9, 28.9, 28.2, 28.0, 27.8, 26.1, 25.8, 25.3, 25.1, 24.8, 23.0, 19.3.

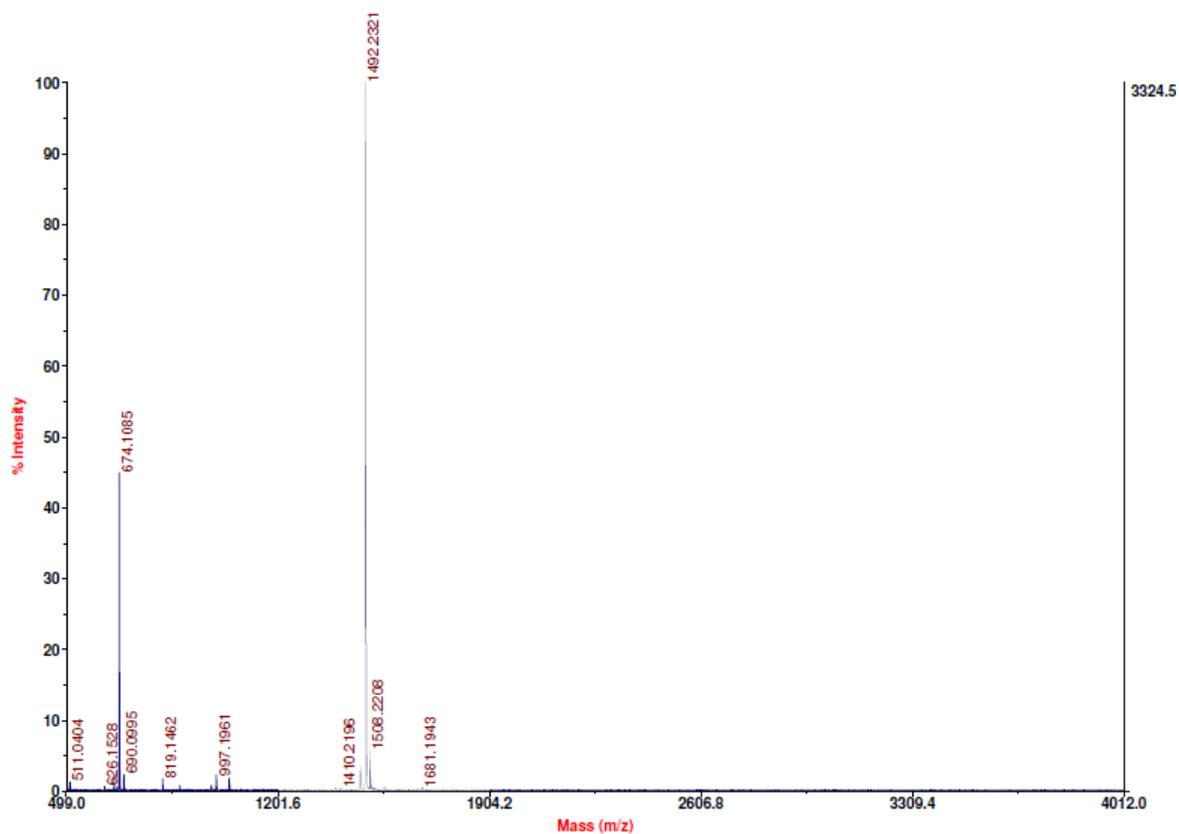




Figure 2.7: Mass spectrometric identification of peptides from E protein of Zika.

## CHAPTER 3. DEVELOPMENT OF DIGE (DIFFERENCE GEL ELECTROPHORESIS) FOR PHOSPHOPROTEOMICS

### 3.1 Introduction

Since, DiGE is a type of two-dimensional gel electrophoresis, it is important to briefly review this classical and versatile method of protein separation.

The evolution of 2-Dimensional gel electrophoresis took place alongside proteomics. Though proteomics, with the advancement of the field, utilizes much broader techniques, 2-D gel electrophoresis still finds its use as a tool for simplifying complex protein mixtures before subjecting them to mass spectrometry for protein identification. When it comes to biological samples, the complexity demands a high resolution protein separation method and 2-D gel electrophoresis indeed solves this purpose.<sup>1,2</sup> However, this method is a little complex compared to the normal gel electrophoresis (1-D) and does involve multiple steps, which are discussed below.

#### 3.1.1 Steps for 2-Dimensional Gel Electrophoresis (2-DE):

In general, 2-D gel electrophoresis (2-DE) involves the separation of proteins, initially, based on their isoelectric point (pH at which the protein remains in its electrically neutral form) followed by the separation according to their molecular weight (Figure 3.1).<sup>1</sup>

##### 3.1.1.1 Sample Preparation

A perfect sample preparation method would be the one that involves the dissolution of all the proteins quantitatively without any modification to them, leaving behind any other interfering substances. Also, what is more important is the sample compatibility with the following step (in this case, isoelectric focusing). Since, in the next step, proteins are separated based on their isoelectric point (pI), the primary requirement is not to alter their charges in any way, during sample preparation. This excludes the use of most commonly used detergent, SDS (Sodium Dodecyl Sulfate) or any other ionic detergent for dissolving proteins. Hence, the best solution here is the use of chaotropes (urea and thiourea), which are very powerful protein denaturing agents due to their ability to increase disorder in water molecules and cause unfolding of proteins.<sup>3,4</sup> However, chaotropes alone are not capable of maintaining the dissolved state of proteins under

conditions of isoelectric focusing. Hence, non-ionic detergents (primarily CHAPS), are used along with urea.<sup>1</sup>

Above mixture of urea and CHAPS works well in most of the cases. However, there is an inherent problem associated with the use of urea. The probability of carbamylation of proteins increases, especially during storage, with the use of urea.<sup>5</sup> So, the problem of protein modification comes into picture.

When the pH of the system shifts to the basic side, there is another problem of thiol and phenol ionization. Tyrosine phenol ionization is less common compared to that of cysteine thiol. These ionizations can again alter the overall charge on the protein, leading to erroneous results. In order to prevent this, it is not a bad idea to convert all the cysteine thiols to disulfide bonds at this stage.<sup>6,7</sup>

### 3.1.1.2 Isoelectric Focusing

Usually, the sample protein solution is mixed with the strip rehydration solution, and the immobilized pH gradient (IPG) strip is incubated with it under electric field for several hours. The process is convenient, but it involves some significant protein loss.<sup>8</sup>

Another important consideration in 2-DE is the application of a very high electric field. About 170 V/cm field strengths are common in case of 2-DE, compared to just 15 V/cm for general electrophoresis (1-D). On the IPG strip, a protein moves as long as it remains in its ionized form (at any certain pH) due to electric field. However, this movement stops when protein reaches its isoelectric point (neutral form). As the protein gets closer to its pI, the speed of migration goes down and that's why, there is a requirement of much higher field strength for isoelectric focusing.<sup>1,2</sup> After multiple hours of operation, all the proteins get stabilized at their pI after which there is no further movement and proteins get separated on 1<sup>st</sup> dimension (termed as isoelectric focusing).

Biological samples usually contain varying amount of salts. If these samples are not substantially desalted, there is a problem of excessive sample heating while applying strong electric field strengths. As a solution, usually a low to middle-range field strength is applied initially for some time (for instance 15 V/cm for 3 hours) to allow faster-moving salts to stabilize first, before increasing the voltage to very high.<sup>1</sup>

One of the biggest issues or inherent problems associated with 2-DE is the high probability of protein precipitation. The definition of pI itself highlights the least solubility of proteins at their isoelectric pH. At this pH, proteins have net neutral charge and hence, there is no (to least) repulsive forces between proteins, which is a potential reason behind their precipitation. This brings another limitation of 2-DE, which is its inapplicability to membrane proteins.<sup>9</sup>

#### 3.1.1.3 Strip Equilibration

It involves the treatment of IPG gel (strip) with SDS. During this process, the proteins separated on the strip will be covered with SDS. This is required for the second dimension molecular weight-based separation. The principle is same as normal gel electrophoresis, in which SDS imparts negative charge to all proteins.<sup>10</sup>

#### 3.1.1.4 SDS-PAGE

This is done according to the standard protocol. The only difference is that IPG strip is actually brought in contact with the top part of the gel, so that proteins earlier separated according to their pI, can now be separated further using normal SDS gel electrophoresis. Only precaution here is to start the gel at a fairly low voltage. This is because, as pointed out earlier, the solubility of proteins is quite low in IPG strip. Low voltage gives SDS sufficient time to re-dissolve these proteins.<sup>1</sup>

#### 3.1.1.5 Protein Detection and Identification

Multiple staining methods are available for detection of proteins. Nevertheless, none of them is perfect. Coomassie blue, the most standard protein staining reagent, suffers with its limitation of low sensitivity.<sup>11</sup> But, it is still the method of choice for proteomics, due to its compatibility with the mass spectrometry.<sup>12</sup> Other common methods are silver staining and fluorescence-based. Silver staining is predominantly used for its nanogram sensitivity.<sup>13</sup> However, it has poor linearity and mass spectrometry compatibility.<sup>14,12</sup> To overcome this, formaldehyde-free silver staining protocol was developed.<sup>15</sup> While, fluorescence method combines the advantages for both previous methods. It delivers an excellent sensitivity, linearity and compatibility with the downstream protein analysis methods.<sup>12</sup>

Though 2-DE is usually performed in the conventional order of isoelectric focusing separation first and SDS-PAGE next, the opposite order is also been described.<sup>16,17</sup> But, the

conventional order is preferred, as it is economical and also, due to the ease of gel staining and protein identification after SDS-PAGE.

It is noteworthy to mention here that in the past, the reproducibility of isoelectric focusing (first dimension separation) was very poor. This was due to the use of carrier ampholytes in isoelectric focusing, which was subsequently replaced by IPG strips with much better performance.<sup>18</sup> However, when it comes to detection of mutational level protein changes, this reproducibility is not good enough. Since, 2-D gel gives the global view of the entire proteome of the cell on just a rectangular slab, it is associated with a lot of inherent variability.

### 3.1.2 DiGE (Difference Gel Electrophoresis)

DiGE was developed to solve the variability or reproducibility problem of 2-DE. In order to detect the differential protein expression levels in two samples, the 2-D gel for both samples need to be compared. For this purpose, the overlay of the two gels is required. However, in a case of 2-D gel, which usually lacks good reproducibility in terms of spot position, combined with the complex nature of samples, it is extremely difficult to superimpose spots on two different gels. Sometimes, a number of replicate gels are run and an electronic average is taken. But, this method has a little success because the variations in 2-D gels are enormous. In order to avoid this gel-to-gel variability, it was proposed to run both samples (test and control) on the same gel. As an instance, both normal and cancerous cell extracts can be run on the same gel to detect the protein changes during cancer occurrence. This led to the invention of DiGE (Difference Gel Electrophoresis), which allows quantitative comparison of the protein changes, by running both samples on the same gel.<sup>19,20</sup>

The next challenge was to identify which spot belongs to which sample, for which it was essential to label both samples with different dyes (let's say one red and other blue) (Figure 3.2).<sup>21</sup>

As shown in above figure, in DiGE, two or three samples are labeled with different dyes and then they are mixed together before subjecting to isoelectric focusing. All remaining steps are same as 2-DE. A detailed representation of principle of DiGE is shown in Figure 3.3.<sup>22</sup>

In Figure 3.2, the two red spots indicate the proteins having higher expression level in test sample compared to the control. Similarly, light blue spots correspond to the proteins with reduced expression in the test sample (or control protein is higher). All other proteins with no difference in

expression levels between the test and control will appear as a violet spots (absolute overlap of red and blue spots).

### 3.1.2.1 Advantages of DiGE

Since the samples are mixed together before separation, if there is a sample loss, it will be same for both samples. This is just one of the many ways DiGE eliminates different variability. DiGE offers perfect spot to spot matching, to detect even the slight difference in protein expressions of the samples (<15%). It is a highly sensitive method and can detect 0.5 fmol of protein. The linearity of detection is also exceptionally high (>10,000 fold)<sup>19,23</sup>

### 3.1.2.2 DiGE Dyes:

Four parameters were chosen in the design of dyes for DiGE:

- 1) Both dyes should react with the same amino acid residue of proteins.
- 2) They should possess similar mass (molecular weight).
- 3) Dyes should not alter charges on proteins. This is paramount because of the reasons explained earlier. For 2-DE, maintaining the natural charge on proteins is pivotal, so as to allow them to separate based on their isoelectric point.
- 4) Lastly, they should have different fluorescent properties. Their excitation frequencies should be far enough to be able to excite only one dye at a time.

The first one is important, in order to have a similar extent of labeling on both samples. It is logical to state that protein labeling increases the mass of proteins (especially when fluorescent dyes are big molecules of considerable mass) and, hence, affects their migration during second dimension separation (SDS-PAGE) of 2-DE. Hence, it is important to have even labeling or same level of protein labeling in both samples. There are two types of labeling suggested for these dyes: Saturation labeling and Minimal labeling. In saturation labeling, all the reactive residues on protein are occupied by the dye, while in minimal labeling each protein is labeled by only one dye. If there is any level of labeling in between these two extremes, there tends to be mass differences among labeled proteins.<sup>19,20</sup> Two kinds of dyes were developed: Lysine reactive and Cysteine reactive.

Lysine-reactive dyes: Lysine residues are quite common in any protein. Initially, the saturation labeling was tried with lysine-reactive dyes. Unfortunately, it led to protein precipitation, as the protein molecular weight increased significantly due to presence of a number of lysine

residues per protein. Then, method was optimized for minimal labeling of proteins using same dyes. However, a minor issue was identified in this case. Since, it is a condition of minimal labeling, not all the proteins get labeled. As a result of this, the unlabeled proteins run with a slightly faster pace compared to labeled proteins. This observation is quite common, particularly in the case of low molecular weight proteins (below 30 kDa).

Cysteine-reactive dyes: There are relatively a much fewer number of cysteine in a protein, which makes the saturation labeling feasible for these dyes.<sup>19,20</sup> Due to this, cysteine-reactive dyes have shown a comparatively much higher sensitivity. This makes the labeling of samples, with extremely low protein concentrations, possible and it has been demonstrated in multiple studies.<sup>24-28</sup>

The dye masses need to be same, so that the labelled proteins have similar masses too (provided proteins have similar extent of labeling), which allows them to run at the same pace on SDS-PAGE gel. The most commonly used dyes for DiGE that possess above mentioned parameters are lysine-reactive Cy3, Cy5 and Cy2 (cyanine based dyes). As stated earlier, substoichiometric labeling or minimal labeling is preferred for these lysine-reactive dyes. These dyes are available in their NHS form (Figure 3.4), which undergo nucleophilic substitution with the amines (of lysine) to form amide bond. All of them are also single positively charged, which compensates for the loss of positive charge of lysine amines, keeping the overall charge of protein unchanged.<sup>19,20,22</sup>

### 3.1.3 DiGE in Phosphoproteomics and Goal of the study

DiGE has been applied to study the protein expression changes in the process of whole animal development (example *Drosophila melanogaster*<sup>23,29</sup> and zebrafish<sup>30</sup>), to identification of disease biomarkers in body fluids.<sup>31</sup> The application has been shown for the proteome change at tissue level, to subcellular level to body fluids. However, almost all of these studies detect the change at the protein expression level. It does not give any information about the post-translational modifications (PTMs) of these proteins, which in instances are more important in disease pathogenesis.

Protein phosphorylation has been identified as a very important PTM, involved in regulation of numerous pathways in a normal functioning of cell. Any change in the phosphorylation of proteins, mainly at serine, threonine and tyrosine, may lead to a perturbation in the signaling

pathway or even prognosis of disease (example cancer). A number of proteomic approaches have been used to identify this difference in phosphorylation levels between normal and cancerous cells. One of the commonly used approaches is gel-based proteomics, which involves the separation of complex protein mixtures on 1-D or 2-D gel, followed by mass spectrometric analysis (for protein identification). Nevertheless, these methods are not suitable for comparison of whole proteome of diseased and normal cells. This is due to the very fact that gels suffer from often poor reproducibility, which makes gel-to-gel comparison very difficult.

Here, for the first time, we are trying to apply DiGE principle to phosphoproteomics (DiGE-P) in order to solve this reproducibility problem. In other words, the aim is to label normal and cancer cell extracts differentially (using different dyes), with the amount of labeling proportional to their phosphorylation levels. Next steps would be to mix both labeled samples together and subjecting it to first and second dimensional separation. Again, the idea is to detect the ‘phosphorylation change’ instead of ‘protein change’ in diseased states.

## 3.2 Methods and Results

### 3.2.1 Design of Novel DiGE-P Reagent

The design of the DiGE-P reagent is similar to the VIPing reagent reported earlier.<sup>32</sup> However, instead of biotin, DiGE-P reagent has a photoactive linker. In general, the reagent has three functionalities as shown in figure 3.5A- Titanium (IV) for selective binding to phosphoproteins, a UV-photocrosslinker for covalent labeling of proteins, and a fluorophore for the visualization.

Initially, when the reagent is incubated with complex protein mixture, titanium (that acts a strong lewis acid) binds to phosphate groups by a weak coordinate bond (Figure 3.5A). Subsequently, after shining UV, the reagent is made covalently linked to the same phosphoprotein. Now, the covalently labeled phosphoproteins are separated on the gel which can be detected by the fluorescence. Ideally, only the phosphoproteins should be labeled by the reagent, out of the whole complex protein mixture. Using different fluorophores, different reagents can be synthesized to label the samples differentially (say normal cell with Cy 3, and cancer cell with Cy 5 reagent).

As shown in Figure 3.5B, the core of the molecule is peptidic in nature. It is chosen due to the well-established chemistry behind and the ease of solid phase peptide synthesis. This is again similar to VIPing reagent.<sup>32</sup> Also, for the photo-reactive UV linker, diazirine was used. On shining UV, it forms an extremely reactive carbene with the loss of N<sub>2</sub>, which quickly labels the protein. The selectivity of labeling using diazirine is reported to be the highest among all other photolinkers, owing to its very high reactivity. In other words, carbene is likely to form covalent bond only with the protein with which titanium is already bound. If there is no protein in the immediate vicinity, it reacts with water and gets quenched.

### 3.2.2 Synthesis of Reagents

As stated earlier, the core of the reagent is peptide in nature. Hence, its synthesis was performed by the well-established solid phase (peptide) chemistry. The molecule was built on the Rink amide resin, with each functionality added to the molecule core by the formation of a new peptide bond.<sup>32</sup>

The synthesis of the reagent was performed using the synthetic route shown in figure 3.6. Each amide bond formation step involved sequential deprotection of N-terminus, activation of C-terminus by HCTU, followed by coupling reaction. Ninhydrin test was performed to monitor different steps of the reaction. A positive test after the deprotection step (presence of free amine) and a negative test after the coupling step confirmed success of the respective reactions. The molecule was cleaved from the resin using 95% TFA to get compound 1 (Figure 3.6). The phosphonate in the molecule still had the protecting group, which was removed by treatment with TMSBr to yield compound 2. Lastly, the NHS-diazirine was reacted to give compound 3, which contains the fluorophore (TAMRA), UV reactive diazirine, and the phosphonate (for Titanium coordination).

As expected, MALDI showed (M + H)<sup>+</sup> peak at m/z 823.6 for compound 2 and (M - N<sub>2</sub> + H)<sup>+</sup> peak at m/z 906.4 for compound 3. It has been documented before that diazirine, under MALDI conditions, loses N<sub>2</sub> easily.<sup>33</sup> HPLC purification and MALDI analysis were performed for compound 2 and 3 (Figure 3.7).

Finally, the titanium coordination to phosphonate was performed by immobilizing the reagent on zip tip and incubating with an acidic solution of Titanium oxychloride. It is noteworthy

to mention here that, similar to above DiGE-P reagent, another reagent was synthesized with a longer chain between diazirine and amide (Figure 3.8).

### 3.2.3 Phosphoprotein Labeling

The labeling was performed using glycolic acid as the binding solution. In brief, protein sample was incubated with the reagent (0.4-0.5  $\mu\text{M}$ ) in glycolic acid for 15 minutes at room temperature. During this process titanium (of reagent) binds to phosphoproteins. Excess reagent was removed using zeba column and UV light was shone for 15 minutes for covalent labeling of proteins. SDS-PAGE was performed. Gel was scanned using Typhoon and stained with coomassie blue.

#### 3.2.3.1 Detection of Phosphoproteins in a Standard Protein Mixture

A four-protein mixture of two phosphoproteins (ovalbumin and  $\alpha$ -casein) and two non-phosphoproteins (BSA and  $\beta$ -lactoglobulin) was prepared and labeled with the DiGE-P reagent using above method. Each lane has 2  $\mu\text{g}$  of four-protein mixture (Figure 3.9A). First lane is ladder, and second lane is the loading control (no reagent). In results, the phosphoproteins were selectively labeled, with a high sensitivity. The  $\alpha$ -casein showed a much stronger signal compared to ovalbumin, due to higher number of phosphorylated residues in  $\alpha$ -casein. More importantly, it can be clearly observed that with increase in glycolic acid concentration, there is a better labeling of phosphoproteins (with optimum glycolic acid concentration of 3-5 M). The selectivity and sensitivity were further demonstrated in figure 3.10.

#### 3.2.3.2 Labeling of Phosphoproteins in Complex Mixtures

Next, the selectivity of DiGE-P labeling was investigated by spiking phosphoproteins in Human Plasma (1  $\mu\text{g}$  each of ovalbumin and  $\alpha$ -casein spiked in 33  $\mu\text{g}$  human plasma proteins) (Figure 3.9B and 3.11). Human plasma is a highly complex sample that contains huge amount of plasma proteins (primarily albumin), besides many other constituents. Also, the phosphorylation level in human plasma is quite low. In results, there was some non-specific albumin labeling observed, but phosphoproteins were detected with good signal strength.

### 3.2.3.3 Quantitative ability of DiGEP

Samples were labeled with Cy3 and Cy5 DiGEP reagents. Cy3- and Cy5-labeled samples were pooled for SDS-PAGE analysis. As shown in Figure 3.12,  $\alpha$ -casein was visible in fluorescence detection, again showing good selectivity. Quantitative analysis of images under two different fluorescence wavelengths revealed the ratio of two protein samples close to the theoretical value.

### 3.3 Conclusion and Future Work

DiGE, a modified form of 2-DE, overcomes the reproducibility problem of 2-D gels and allows the comparison of two or more samples by running them on the same gel. It offers great sensitivity and has an excellent linearity of detection.

This is for the first time, we are trying to apply DiGE technology to phosphoproteomics, so as to detect the phosphorylation differences between normal and diseased (cancerous) cells. Thus far, we have been successfully able to synthesize the reagent and use it to label the simple four-protein mixture. The method also works, to some extent, for the complex protein mixtures like human plasma that has a very low phosphorylation level. However, the method needs further optimization in case of complex protein mixtures.

In the future, we wish to apply this novel technology, DiGE-P, to label phosphoproteins in complex mixtures like *E.coli*, yeast, plant cell lysates and even extracts from human cancer cell lines. This will be followed by in-gel digestion and phosphoprotein identification using mass spectrometry.

### 3.4 References

- (1) Rabilloud, T.; Lelong, C. Two-dimensional gel electrophoresis in proteomics: A tutorial. *J. Proteomics* **2011**, *74* (10), 1829–1841.
- (2) López, J. L. Two-dimensional electrophoresis in proteome expression analysis. *J. Chromatogr. B* **2007**, *849* (1-2), 190–202.
- (3) Gordon, J. A.; Jencks, W. P. The Relationship of Structure to the Effectiveness of Denaturing Agents for Proteins. *Biochemistry* **1963**, *2* (1), 47–57.
- (4) Rabilloud, T.; Adessi, C.; Giraudel, A.; Lunardi, J. Improvement of the solubilization of proteins in two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* **1997**, *18* (3-4), 307–316.

- (5) Čejka, J.; Vodrážka, Z.; Salák, J. Carbamylation of globin in electrophoresis and chromatography in the presence of urea. *BBA - Protein Struct.* **1968**, *154* (3), 589–591.
- (6) Luche, S.; Diemer, H.; Tastet, C.; Chevallet, M.; Van Dorsselaer, A.; Leize-Wagner, E.; Rabilloud, T. About thiol derivatization and resolution of basic proteins in two-dimensional electrophoresis. *Proteomics* **2004**, *4* (3), 551–561.
- (7) Olsson, I.; Larsson, K.; Palmgren, R.; Bjellqvist, B. Organic disulfides as a means to generate streak-free two-dimensional maps with narrow range basic immobilized pH gradient strips as first dimension. *Proteomics* **2002**, *2* (11), 1630–1632.
- (8) Zhou, S.; Bailey, M. J.; Dunn, M. J.; Preedy, V. R.; Emery, P. W. A quantitative investigation into the losses of proteins at different stages of a two-dimensional gel electrophoresis procedure. *Proteomics* **2005**, *5* (11), 2739–2747.
- (9) Rabilloud, T. Membrane proteins and proteomics: Love is possible, but so difficult. *Electrophoresis* **2009**, *30* (S1), 174–180.
- (10) O'Farrell, P. H. High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* **1975**, *250*, 4007–4021.
- (11) Neuhoff, V.; Arold, N.; Taube, D.; Ehrhardt, W. Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis* **1988**, *9* (6), 255–262.
- (12) Chevalier, F.; Centeno, D.; Rofidal, V.; Tauzin, M.; Martin, O.; Sommerer, N.; Rossignol, M. Different Impact of Staining Procedures Using Visible Stains and Fluorescent Dyes for Large-Scale Investigation of Proteomes by MALDI-TOF Mass Spectrometry. *J. Proteome Res.* **2006**, *5* (3), 512–520.
- (13) Chevallet, M.; Lescuyer, P.; Diemer, H.; van Dorsselaer, A.; Leize-Wagner, E.; Rabilloud, T. Alterations of the mitochondrial proteome caused by the absence of mitochondrial DNA: A proteomic view. *Electrophoresis* **2006**, *27* (8), 1574–1583.
- (14) Rabilloud, T. Mechanisms of protein silver staining in polyacrylamide gels: A 10-year synthesis. *Electrophoresis* **1990**, *11* (10), 785–794.
- (15) Chevallet, M.; Luche, S.; Diemer, H.; Strub, J.-M.; Van Dorsselaer, A.; Rabilloud, T. Sweet silver: A formaldehyde-free silver staining using aldoses as developing agents, with enhanced compatibility with mass spectrometry. *Proteomics* **2008**, *8* (23-24), 4853–4861.
- (16) Tuszynski, G. P.; Buck, C. A.; Warren, L. A two-dimensional polyacrylamide gel electrophoresis (PAGE) system using sodium dodecyl sulfate—PAGE in the first dimension. *Anal. Biochem.* **1979**, *93*, 329–338.
- (17) Nakamura, K.; Okuya, Y.; Katahira, M.; Yoshida, S.; Wada, S.; Okuno, M. Analysis of tubulin isoforms by two-dimensional gel electrophoresis using SDS-polyacrylamide gel electrophoresis in the first dimension. *J. Biochem. Biophys. Methods* **1992**, *24* (3-4), 195–203.
- (18) Corbett, J. M.; Dunn, M. J.; Posch, A.; Görg, A. Positional reproducibility of protein spots in two-dimensional polyacrylamide gel electrophoresis using immobilised pH gradient isoelectric focusing in the first dimension: An interlaboratory comparison. *Electrophoresis* **1994**, *15* (1), 1205–1211.
- (19) Ünlü, M.; Morgan, M. E.; Minden, J. S. Difference gel electrophoresis. A single gel method for detecting changes in protein extracts. *Electrophoresis* **1997**, *18* (11), 2071–2077.

- (20) Minden, J. S.; Dowd, S. R.; Meyer, H. E.; Stühler, K. Difference gel electrophoresis. *Electrophoresis* **2009**, *30* (S1), 156–161.
- (21) Viswanathan, S.; Ünlü, M.; Minden, J. S. Two-dimensional difference gel electrophoresis. *Nat. Protoc.* **2006**, *1* (3), 1351–1358.
- (22) Tonge, R.; Shaw, J.; Middleton, B.; Rowlinson, R.; Rayner, S.; Young, J.; Pognan, F.; Hawkins, E.; Currie, I.; Davison, M. Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. *Proteomics* **2001**, *1* (3), 377–396.
- (23) Gong, L.; Puri, M.; Unlu, M.; Young, M.; Robertson, K.; Viswanathan, S.; Krishnaswamy, A.; Dowd, S. R.; Minden, J. S. Drosophila ventral furrow morphogenesis: a proteomic analysis. *Development* **2004**, *131* (3), 643–656.
- (24) Sitek, B.; Lüttges, J.; Marcus, K.; Klöppel, G.; Schmiegel, W.; Meyer, H. E.; Hahn, S. A.; Stühler, K. Application of fluorescence difference gel electrophoresis saturation labelling for the analysis of microdissected precursor lesions of pancreatic ductal adenocarcinoma. *Proteomics* **2005**, *5* (10), 2665–2679.
- (25) Sitek, B.; Potthoff, S.; Schulenburg, T.; Stegbauer, J.; Vinke, T.; Rump, L.-C.; Meyer, H. E.; Vonend, O.; Stühler, K. Novel approaches to analyse glomerular proteins from smallest scale murine and human samples using DIGE saturation labelling. *Proteomics* **2006**, *6* (15), 4337–4345.
- (26) Wilson, K. E.; Marouga, R.; Prime, J. E.; Pashby, P. D.; Orange, P. R.; Crosier, S.; Keith, A. B.; Lathe, R.; Mullins, J.; Estibeiro, P.; et al. Comparative proteomic analysis using samples obtained with laser microdissection and saturation dye labelling. *Proteomics* **2005**, *5* (15), 3851–3858.
- (27) Greengauz-Roberts, O.; Stöppler, H.; Nomura, S.; Yamaguchi, H.; Goldenring, J. R.; Podolsky, R. H.; Lee, J. R.; Dynan, W. S. Saturation labeling with cysteine-reactive cyanine fluorescent dyes provides increased sensitivity for protein expression profiling of laser-microdissected clinical specimens. *Proteomics* **2005**, *5* (7), 1746–1757.
- (28) Kondo, T.; Seike, M.; Mori, Y.; Fujii, K.; Yamada, T.; Hirohashi, S. Application of sensitive fluorescent dyes in linkage of laser microdissection and two-dimensional gel electrophoresis as a cancer proteomic study tool. *Proteomics* **2003**, *3* (9), 1758–1766.
- (29) Puri, M.; Goyal, A.; Senutovich, N.; Dowd, S. R.; Minden, J. S. Building proteomic pathways using Drosophila ventral furrow formation as a model. *Mol. BioSyst.* **2008**, *4* (11), 1126–1135.
- (30) Lucitt, M. B.; Price, T. S.; Pizarro, A.; Wu, W.; Yocum, A. K.; Seiler, C.; Pack, M. A.; Blair, I. A.; FitzGerald, G. A.; Grosser, T. Analysis of the Zebrafish Proteome during Embryonic Development. *Mol. Cell. Proteomics* **2008**, *7* (5), 981–994.
- (31) Pitteri, S. J.; Hanash, S. M. Proteomic approaches for cancer biomarker discovery in plasma. *Expert Rev. Proteomics* **2007**, *4* (5), 589–590.
- (32) Wang, L.; Pan, L.; Tao, A. W. Specific Visualization and Identification of Phosphoproteome in Gels. *Anal. Chem.* **2014**, *86* (14), 6741–6747.
- (33) Gomes, A. F.; Gozzo, F. C. Chemical cross-linking with a diazirine photoactivatable cross-linker investigated by MALDI- and ESI-MS/MS. *J. Mass Spectrom.* **2010**, *45* (8), 892–899s.

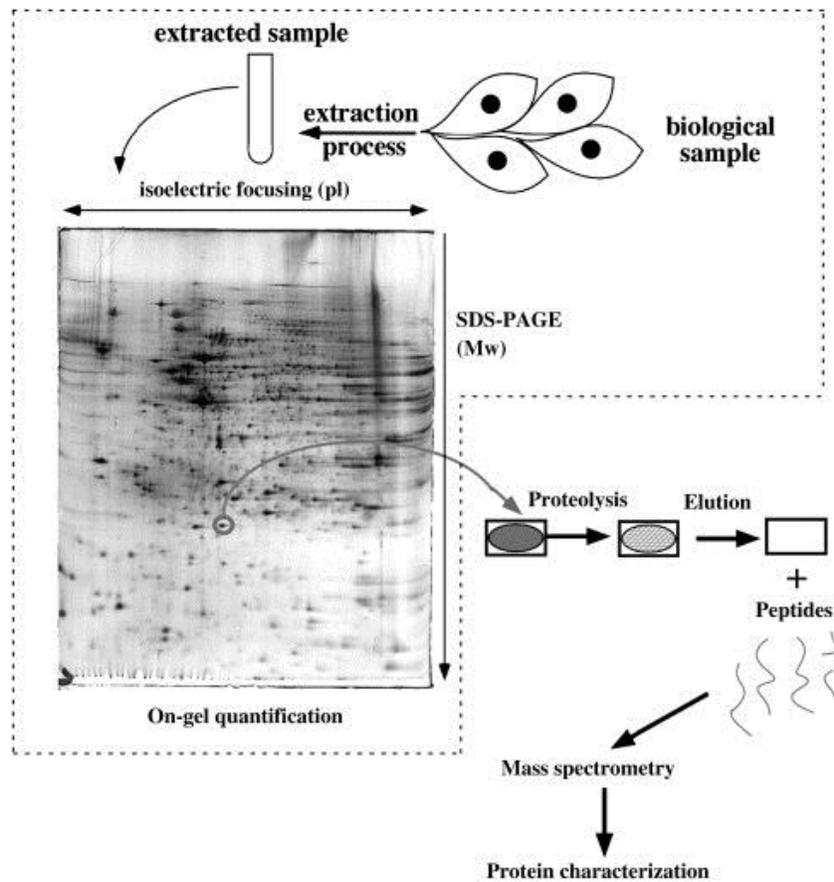


Figure 3.1: Principle and Steps of 2-DE.<sup>1</sup>

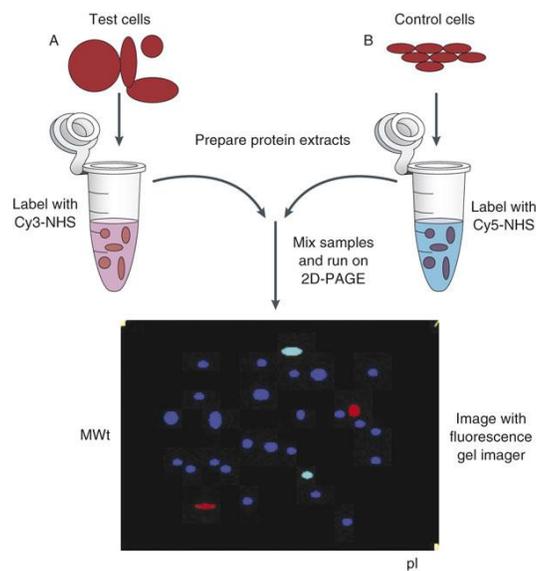


Figure 3.2: General Representation of DiGE.<sup>21</sup>

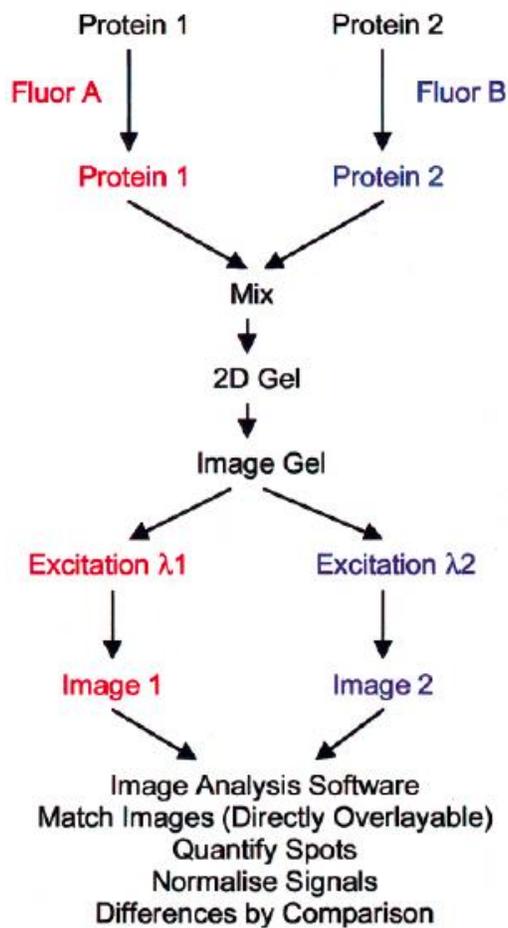


Figure 3.3: Principle of 2D DIGE.<sup>22</sup>

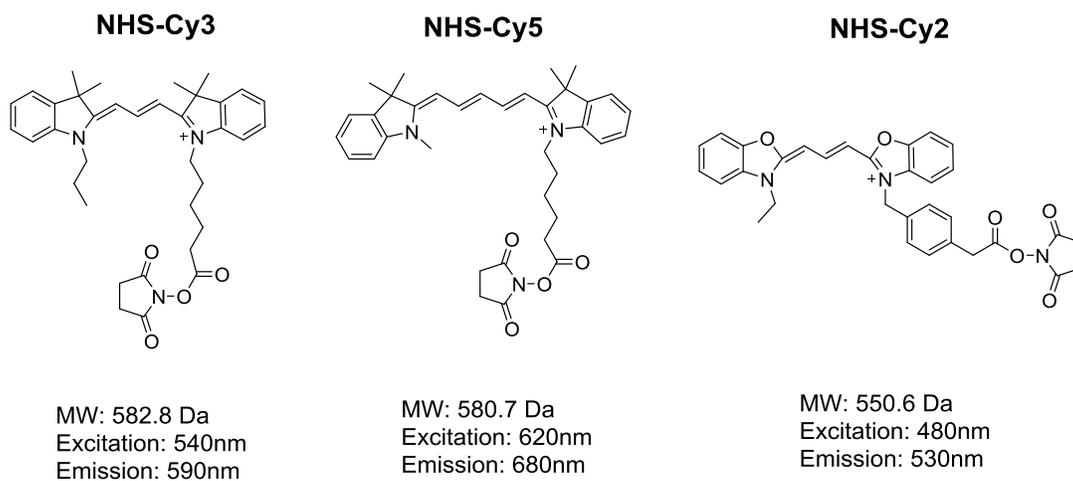
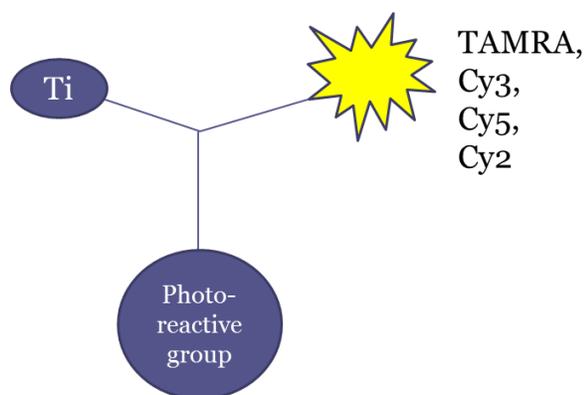


Figure 3.4: Structures of cyanine dyes used in DiGE.<sup>19,20</sup>

A.



B.

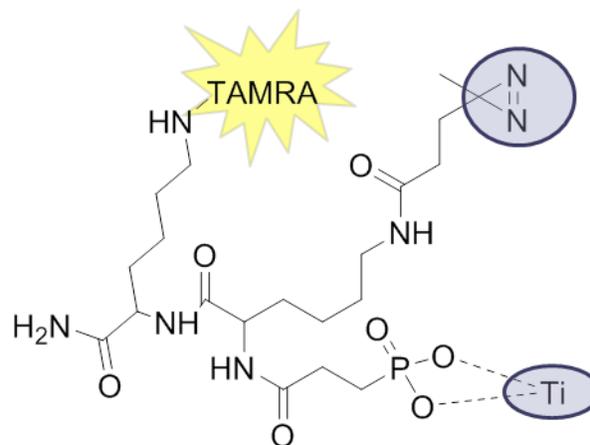


Figure 3.5: DiGE-P Reagent (A= schematic representation, B= An example of actual DiGE-P reagent).

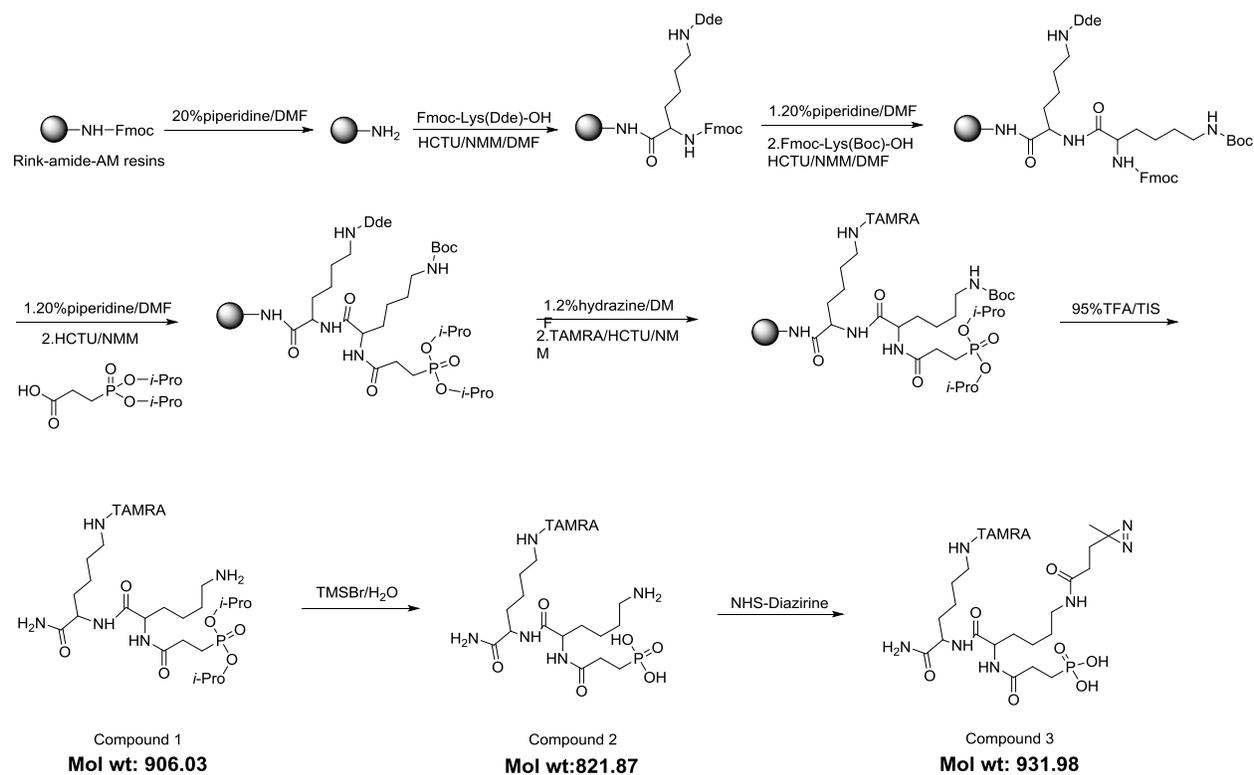
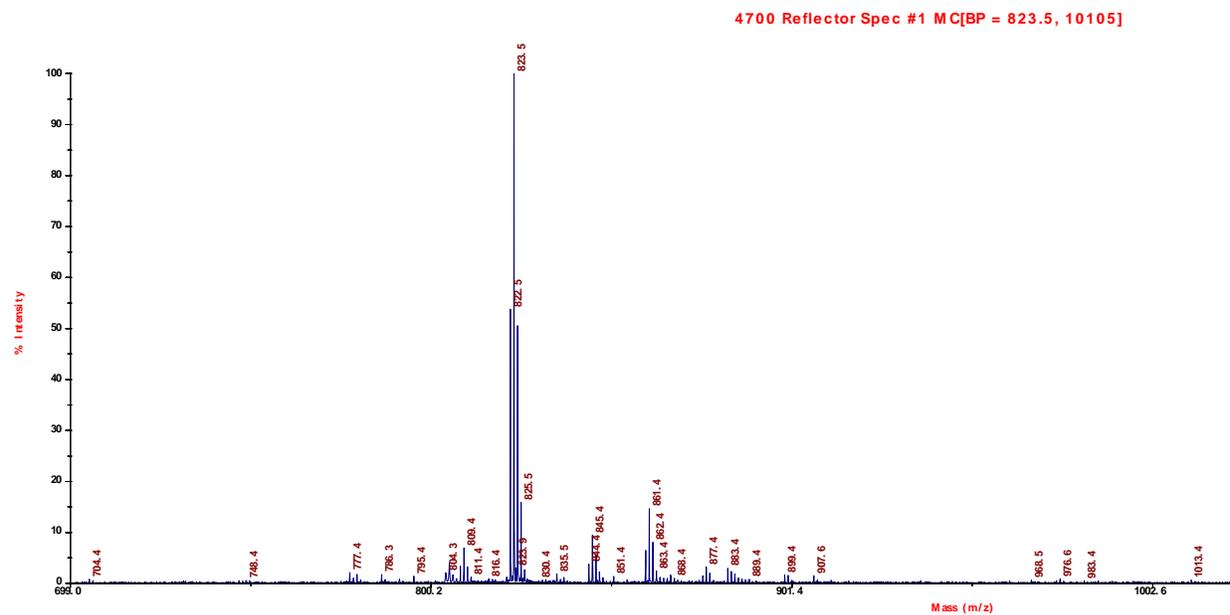


Figure 3.6: Synthetic Route for DiGE-P.

A.



B.

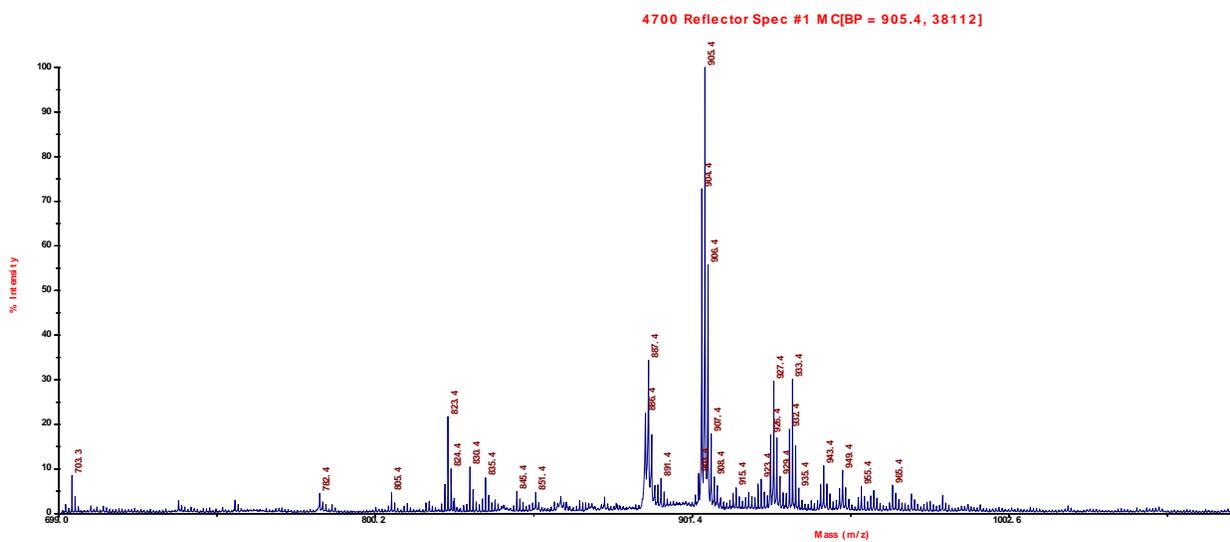


Figure 3.7: MALDI for compound 2 (A) and 3 (B).

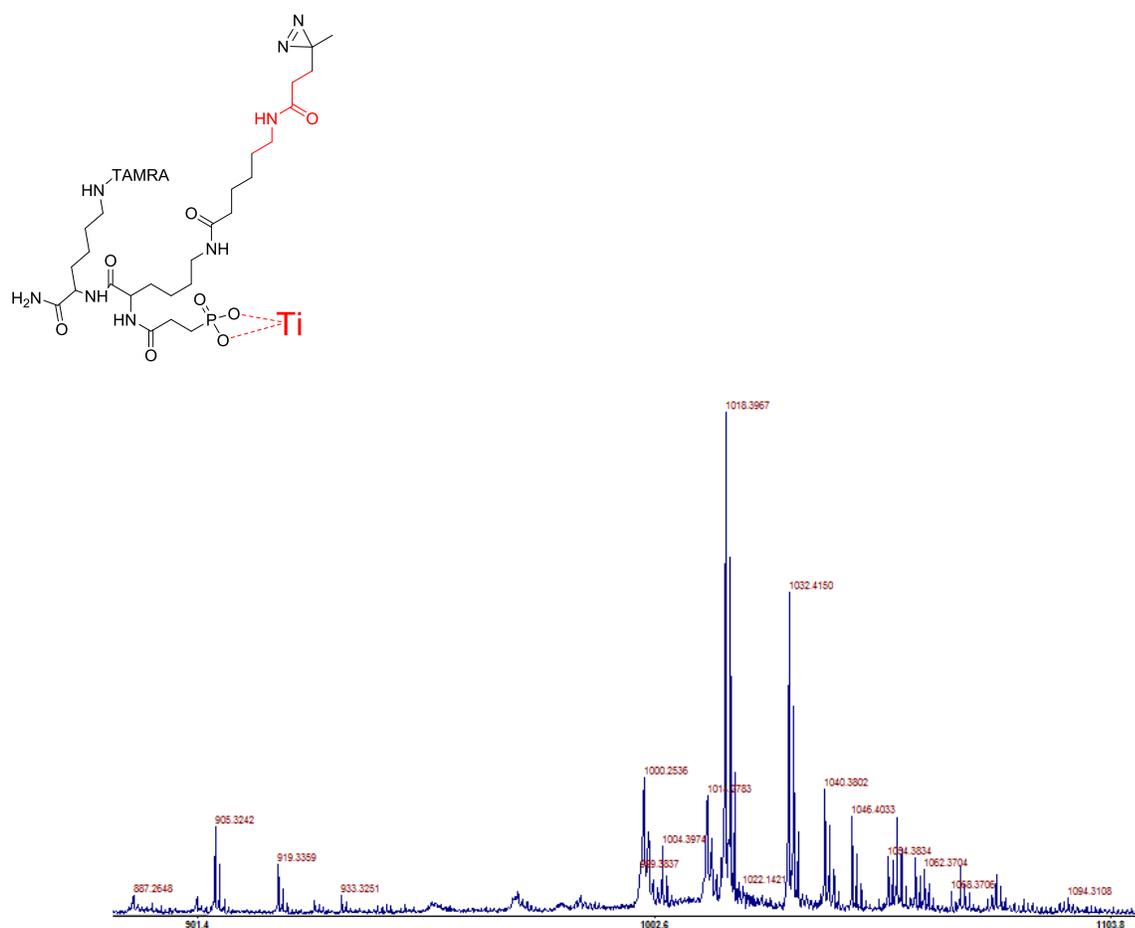
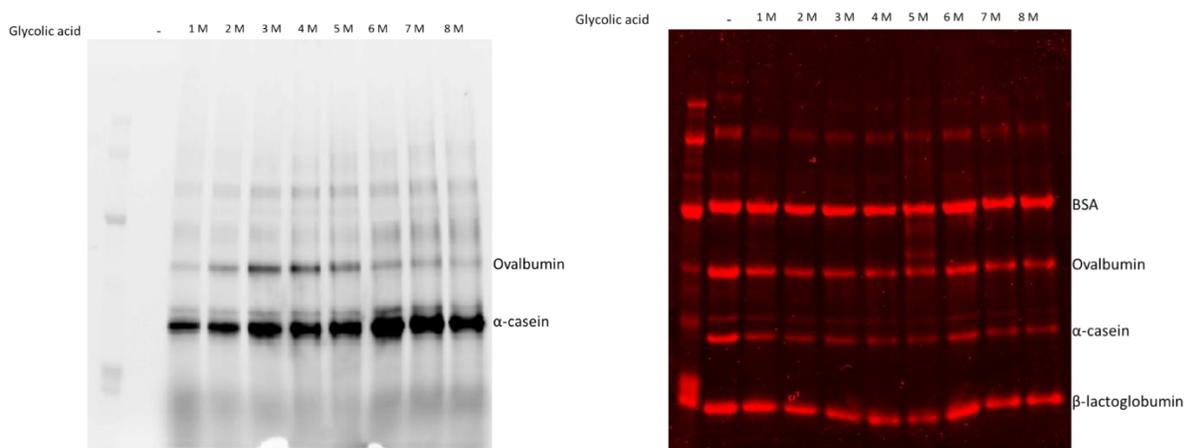


Figure 3.8: Another DiGE-P reagent and its MALDI. (M - N<sub>2</sub> + H)<sup>+</sup> peak observed at m/z at 1018.40.

A.



B.

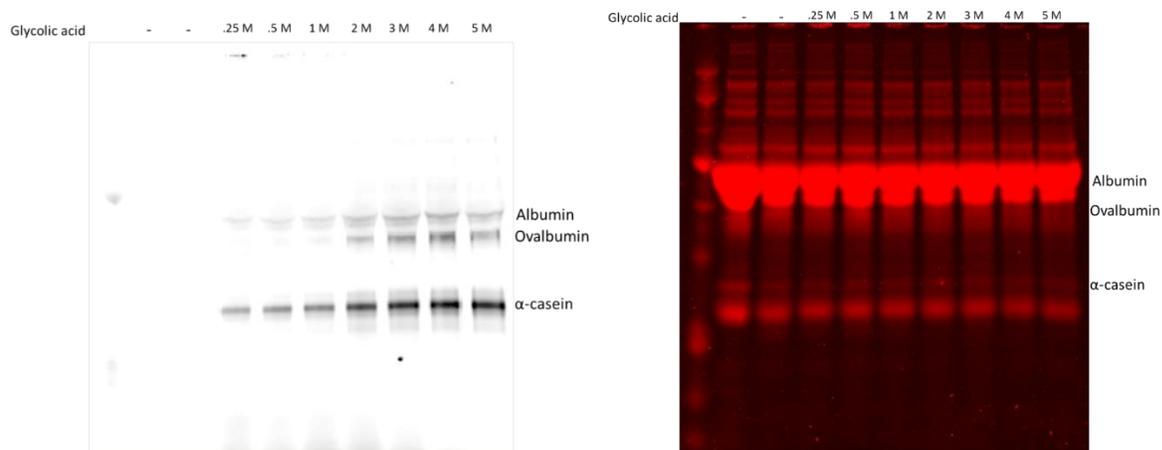
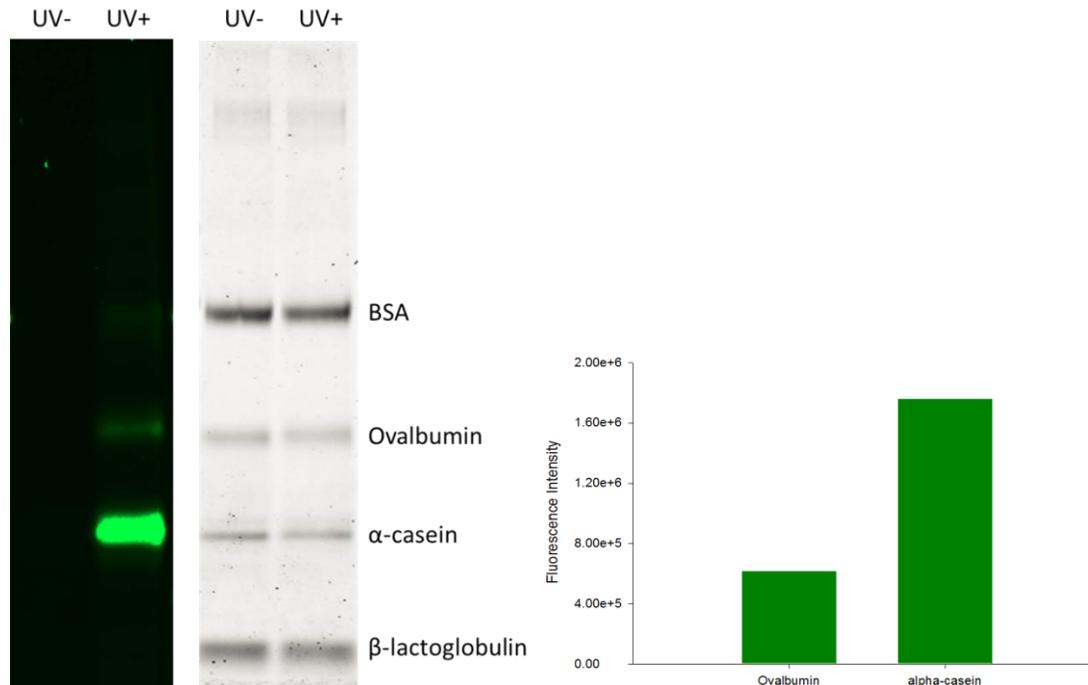


Figure 3.9: Optimization of Phosphoprotein Labeling. Fluorescence image and Coomassie Blue stain of DiGE-P labeled 2  $\mu$ g four-protein mixture of standard phospho and non-phospho proteins (A), and 1  $\mu$ g phosphoproteins in complex mixture with human plasma (B).

A.



B.

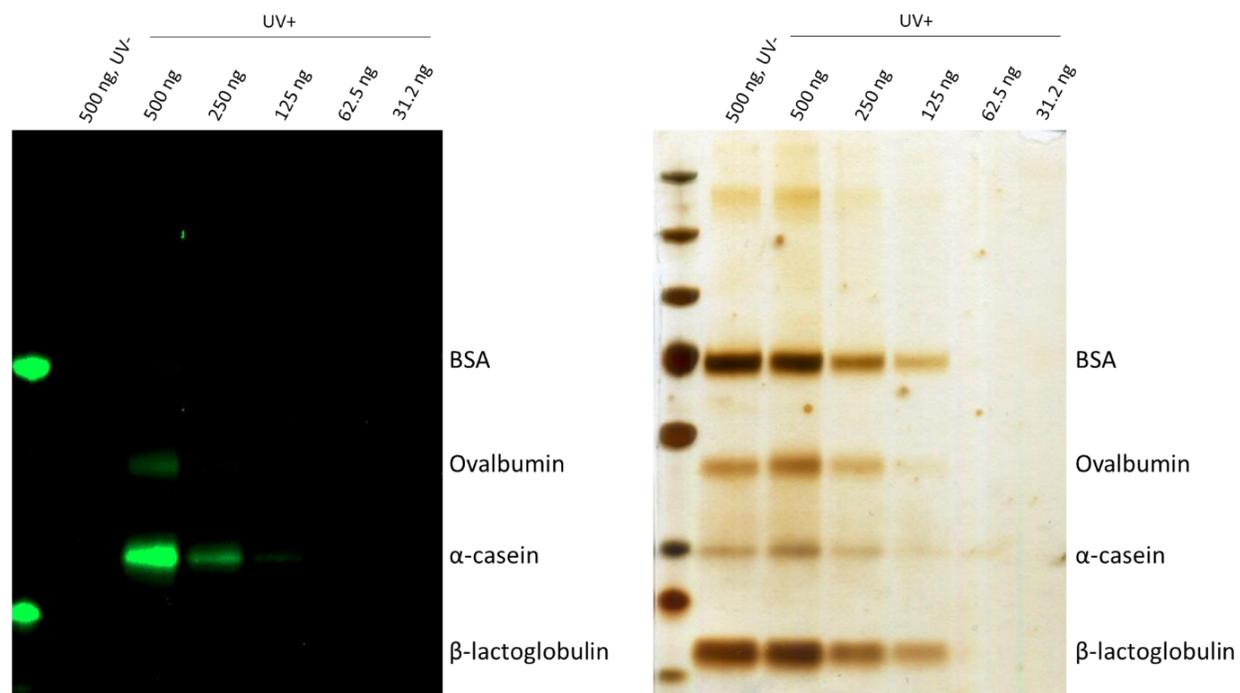


Figure 3.10: Selectivity (A) and Sensitivity (B) testing of DiGE-P using standard four-protein mixture.

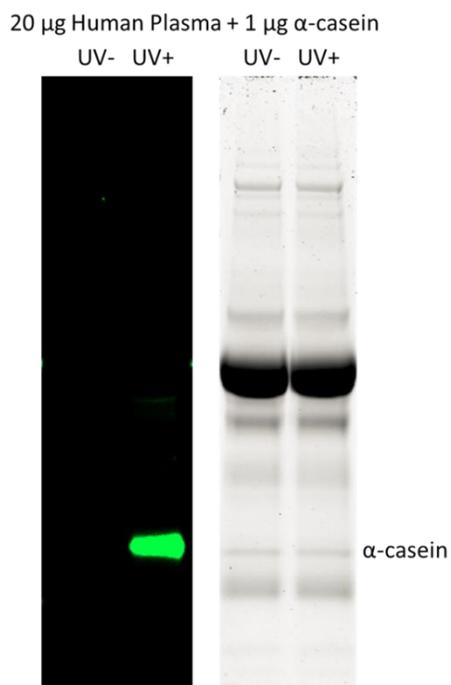


Figure 3.11: Selectivity testing of DiGE-P using 1  $\mu$ g  $\alpha$ -casein spiked in 20  $\mu$ g human plasma.

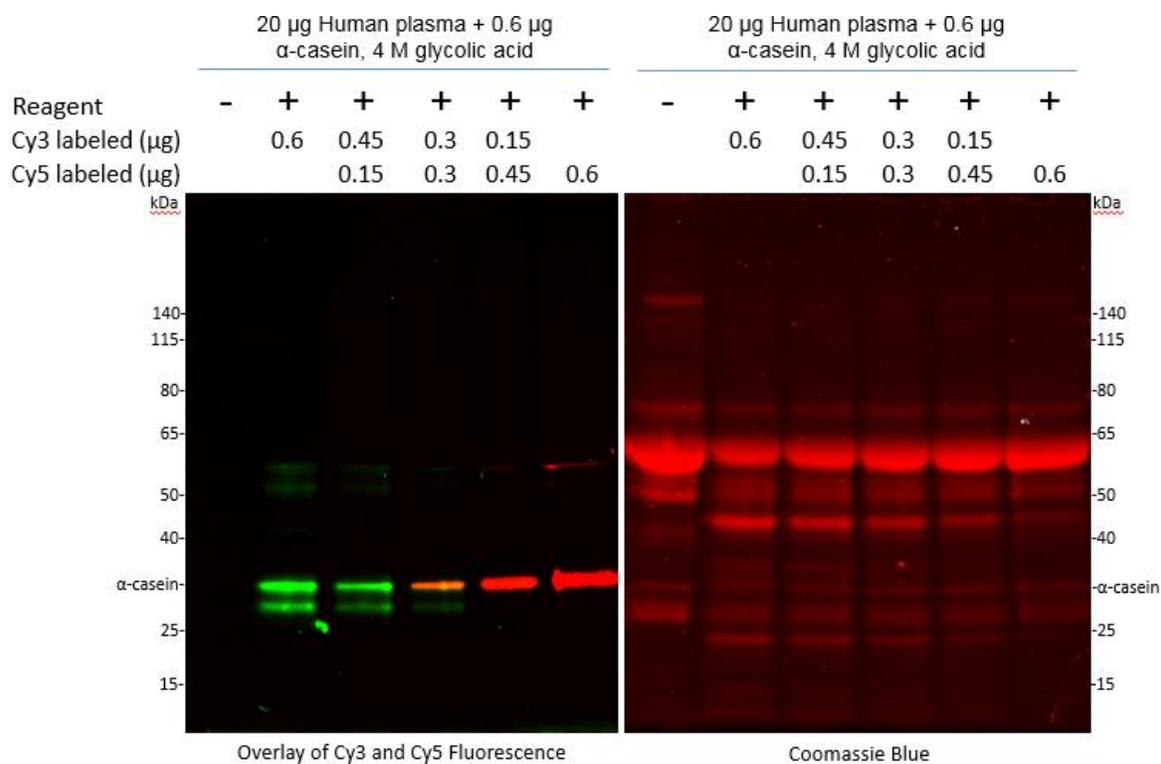


Figure 3.12: Fluorescence image and Coomassie Blue of DiGE-P (Cy3 and Cy5) labeled model phosphoprotein  $\alpha$ -casein in human plasma.