Protein Analysis by Liquid Chromatography-Tandem Mass Spectrometry:

Detecting and Profiling Protein S-Palmitoylation

and Host Cell Protein Characterization

by

John E. Crellin

A dissertation in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemical Biology)
in the University of Michigan
2020

Doctoral Committee:

      Professor Kristina I. Håkansson, Co-Chair
      Assistant Professor Brent R. Martin, Co-Chair
      Professor Robert T. Kennedy
      Professor Neil G. Marsh
      Professor Brandon T. Ruotolo

John E. Crellin

jcrellin@umich.edu

ORCID iD: 0000-0002-0919-5479

To Sarah, with cherished love.

You championed me through my deepest failures

and celebrated my smallest achievements.

I wish only to return the favor.

Acknowledgements

I would like to express my gratitude to the many people who supported and helped me through my graduate studies. The wealth of knowledge, emotional support, and guidance you have provided me with has been essential to my success working towards my PhD.

I appreciate the opportunity that University of Michigan Rackham Graduate School and the Program in Chemical Biology have given me. The faculty and staff of the department who provided administrative guidance were outstanding and reliable. I want to thank my PI, Dr. Brent R. Martin, for his guidance and mentorship. He was readily available and energetic about our research, and he has provided me with confidence and direction through the difficulties in my projects. I'd like to give a special thanks to my previous and current lab mates: Dr. Hashim Motiwala, Dr. Nicholas Borotto, Dr. Bo Peng, Dr Yu-Hsuan Kuo, Dr. Christina TMB Tom, Dr. Jaimeen Majmudar, Dr. Sarah Haynes, Dr. Melanie Cheungseekit, William Walker, and Dongyu Zhao. All of whom contributed in some way to this thesis work. It has been a pleasure working with them, and I wish the very best to their future careers. I would also like to thank my Co-Chair, Dr. Kristina Håkansson, for guiding me through the final steps of my PhD work, and providing me the opportunity to expand my focus by inviting me to weekly group meetings.

Thank you to my Dissertation Committee members: Professor Robert T. Kennedy, Professor Neil G. Marsh, and Professor Brandon T. Ruotolo for all of your time and

Jack Crellin

February 5, 2020

Ann, Arbor Michigan

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Abs   Antibodies

BMS   Bristol-Myers Squibb

BMS mAb 1-2   BMS proprietary recombinant IgG

BMS Fp 1-3   BMS proprietary fusion protein

BSA   Bovine serum albumin

CHO   Chinese hamster ovary

CID   Collison induced dissociation

CZE   Capillary zone electrophoresis

DIA   Data-independent acquisition

DDA   Data-dependent acquisition

DMSO   Dimethyl sulfoxide

DS   Drug substance

DTT   Dithiothreitol

ECD   Electron-capture dissociation

EIC   Extracted ion chromatogram

ELISA   Enzyme-linked immunosorbent assay

ETD   Electron-transfer dissociation

FA	Formic acid
FDR	False discovery rate
HCD	High energy collision dissociation
HCP	Host cell protein
HPLC	High-performance liquid chromatography
HILIC	Hydrophilic interaction chromatography
ICAT	Isotope-coded affinity tag
IEX	Ion exchange chromatography
IT	Ion trap mass analyzer
iTRAC	Isobaric tag for relative and absolute quantitation
LC	Liquid-chromatography
LMW	Low molecular weight
LOD	Limit of detection
LOQ	Limit of quantitation
mAb	Monoclonal antibody
MALDI-TOF	Matrix-assisted laser desorption/ionization-time-of-flight
MeCN	Acetonitrile
MRM	Multiple reaction monitoring
MS	Mass spectrometry
MS^n	Tandem mass spectrometry
MS/MS	Tandem mass spectrometry
MudPIT	Multidimensional Protein Identification Technology
NIST Ab	Humanized IgG1κ monoclonal antibody standard RM 8671

N$_3$-	Azido-
OT	Orbitrap mass analyzer
pAb	polyclonal antibody
PAGE	Polyacrylamide gel electrophoresis
PDA	Photodiode Array Detector
pI	Isoelectric point
PTM	Post-translational modification
QEP	Q Exactive Plus mass spectrometer
RAC	Resin-assisted capture
RP	Reverse phase
SDS	Sodium dodecyl sulfate
SEC	Size-exclusion chromatography
SILAC	Stable isotope labeling with amino acids in culture
SPE	Solid-phase extraction
SRM	Selected reaction monitoring
S-palm	Cysteine palmitoylation
TFA	Trifluoroacetic acid
TMT	Tandem mass tag
WB	Western blot
17-ODYA	17-octadecynoic acid
2D-DIGE	2D-differential in-gel electrophoresis

Abstract

Proteins perform many biological functions and characterizing their role in the cellular environment is critical in understanding a protein's role in disease. 'Bottom-up' liquid chromatography-tandem mass spectrometry (LC-MS/MS) offers the most effective strategy for large-scale global analysis of highly complex protein mixtures, i.e., proteomics. However, LC-MS/MS workflow strategies must be carefully adapted to maintain experimental reproducibility and overcome inherent limitations in detecting low abundance proteins in concentrated mixtures and labile posttranslational modifications (PTMs). This dissertation presents LC-MS/MS based analytical strategies for the direct detection and site-specific profiling of S-palmitoylated peptides, and for characterizing low abundance host cell proteins (HCPs) in purified biopharmaceutical drug substances (DS).

Chapter 1 presents an introduction to quantitative LC-MS/MS-based protein analysis, including instrumentation and techniques relevant to the work presented in Chapters 2 and 3.

Chapter 2 discusses S-palmitoylation, the reversible and dynamic modification of a cysteine sulfhydryl with a 16-carbon fatty acid. The dynamic cycling between a protein's palmitoylated and depalmitoylated states regulates several intracellular events such as protein activity, localization, and protein-protein interactions. N-Ras, a well-known driver of many cancers, is S-palmitoylated near its C-terminus, and despite the therapeutic potential of targeting this modification, direct detection of this modified peptide by LC-MS/MS has not previously been

successful. In fact, robust methods for direct LC-MS/MS-based analysis of S-palmitoylated proteins have been lacking. We addressed this issue by developing an LC-MS/MS "bottom-up" workflow that mitigates sample processing issues associated with S-palmitoylation, including peptide solubility, stability, LC column retention, and MS/MS-based sequence analysis. We successfully applied this workflow to directly detect and annotate endogenous acylation sites on recombinant N-Ras.

Chapter 3 discusses the adaptation of an LC-MS/MS workflow for HCP characterization in biopharmaceutical drug substances. HCPs are native proteins derived from a host organism used to express biotherapeutic proteins and may be co-purified with a drug substance as process related impurities. The presence of these impurities in drug formulations may present problems in product performance and affect the health of patients. Therefore, effective monitoring of HCP levels in drug substances is essential. Common procedures for quantifying HCP content are limited to immunocapture detection methods, but serious efforts are underway in developing LC-MS/MS strategies for characterizing HCPs. Problematically, HCPs can be 10^6-fold less concentrated than the biotherapeutic protein in downstream products presenting a dynamic range challenge for electrospray ionization-MS. We developed a generalized, semi-automated, and plate-based hydrophilic interaction liquid chromatography (HILIC) fractionation workflow to deplete biotherapeutic protein from DS with limited sample-specific optimization. Our method is practical, quantitative, and demonstrates industry leading HCP detection sensitivity by LC-MS/MS.

Chapter 4 provides a concluding summary of the novel contributions of our work and the broader impact on the field of LC-MS/MS based protein analysis. The developed method for direct detection of palmitoylated peptides promises to expand analytical profiling capabilities of

an under-represented PTM. Furthermore, the demonstrated robustness and dynamic range

improvements for HCP analysis show promise for assisting downstream process development of

biotherapeutic drugs as part of an MS-based analytical platform.

Chapter 1

Introduction

1.1 Proteome Analysis

A proteome corresponds to the complex milieu of proteins in a cell or tissue under a given set of conditions. These proteins function to catalyze chemical reactions, transfer signals, associate with other proteins, and provide structural support within the cell. As an example, the human genome encodes approximately 20,000 genes which can be transcribed into approximately 100,000 transcripts after post-transcriptional processing, such as alternative splicing and mRNA editing (Figure 1.1) [1,2]. These transcripts are translated into amino acid sequences which are folded into distinct three-dimensional (3D) protein structures. Amino acids comprising a protein can be modified with distinct chemical moieties, or post-translational modifications (PTM), further increasing proteome complexity to more than 1 million potential variants in human cells. These PTMs alter the chemical and physical properties of proteins, having a significant impact on their activity and function. Similarly, changes in each protein's activity or expression level can have a significant impact on cellular function. Therefore, directly studying the proteome is crucial in understanding cellular function and disease pathology.

Major challenges in proteomics, the study of the global expression of proteins in a cell or tissue, are the complexity of the mixture and the several order of magnitude abundance range

over which each protein is present. Strategies for detecting and quantifying proteins between experimental conditions minimally require a separation component and a detection component. An early example of a proteomics strategy involved two-dimensional polyacrylamide gel electrophoresis (2D-PAGE), developed in the 1970's. This approach enabled researchers to separate and detect up to 1,000 distinct proteins from cells [3-5]. In these experiments, protein extracts are separated in two dimensions by isoelectric point and molecular weight, respectively. Each protein ideally occupies one spot in a two-dimensional gel and is visualized with protein binding dyes or radiography. This strategy demonstrated the ability to detect and measure the abundance of many proteins at once. However, there are some limitations to the method in that only highly abundant proteins were detectable, and protein identification remained challenging.
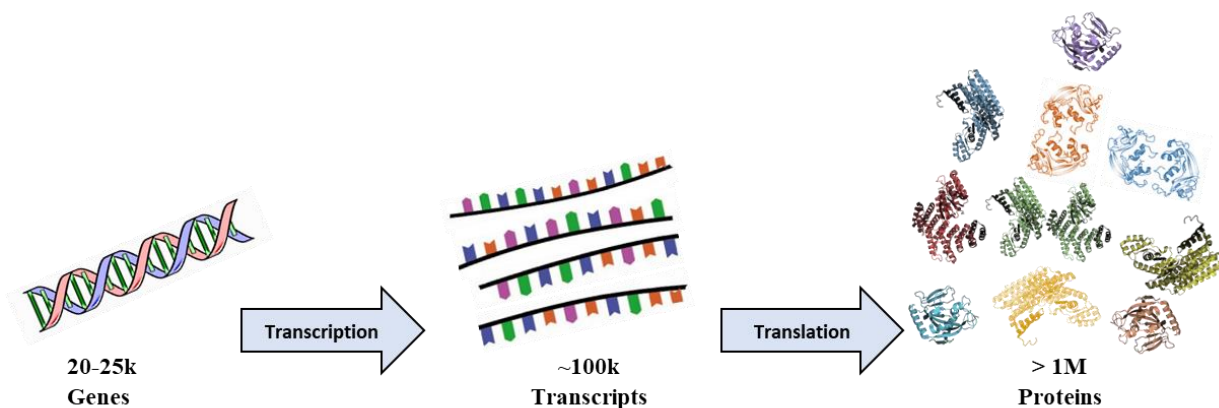
Figure 1.1 Proteomes are complex. The human genome encodes 20-25,000 genes which, after transcriptional processing and post-translational modification, can produce a proteome of over 1 million proteins.

Protein identification could be achieved by using internal protein standards, mapping the radiolabeled proteins with images from previous experiments, using fluorescently labelled chemical tags and antibodies specific to proteins of interest [4-9]. After transferring the proteins to a polyvinylidene difluoride (PVDF) blotting membrane, Edman degradation could also be used to determine the N-terminal amino acid sequence of a protein. However, chemically

modified N-termini will not react with phenyl isothiocyanate, the chemical reactant responsible

for the required degradation reaction, and sequence coverage is limited to ~30 amino acids due to

steric hindrance inhibiting the kinetics of cyclic derivatization. [10-11]. Digestion of the proteins

into their constituent peptides on the other hand, offers greater information for sequence

identification, however the analytical complexity greatly increases and requires more powerful

tools such as mass spectrometry.

1.2 Mass Spectrometry-Based Proteomics

The underlying principle of mass spectrometry (MS) is the separation of gaseous ions in

electric and/or magnetic fields based on their mass-to-charge (m/z) ratios. The introduction of

electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) in the late

1980s for efficient ionization and volatilization of proteins opened up MS to high throughput,

high sensitivity, label free protein analysis [18-23]. The parallel developments in high throughput

genome sequencing allowed generation of databases containing all possible protein sequences

for matching with MS data [50-53]. Because modern MS-based proteomic analysis typically

involves liquid chromatography separation (see section 1.3), ESI is most suitable as ionization

occurs directly from the liquid phase. In ESI, a protein solution is pumped through a needle with

a few kV potential applied with respect to the entrance of the mass spectrometer. Due to the

generated high electric field, solution-phase charged species accumulate at the tip of the needle,

resulting in charged droplet ejection due to Coulomb destabilization. Solvent evaporation

generates smaller droplets that split and further evaporate until gaseous protein ions are formed

and enter the mass spectrometer. This technique is considered "soft" as it ionizes biomolecules

without fragmenting them [21,22]. ESI typically imparts multiple charges on large biomolecules

such as peptides and proteins in a statistical manner with higher charge states observed for larger

molecules. Thus, protein signals are spread over multiple charges states, a characteristic that can be useful but also limits ESI-MS signal-to-noise ratios. Therefore, proteins are typically digested into proteolytic peptides for improved ionization.

The multiple charging inherent to ESI allows for efficient tandem MS (MS/MS), in which sequencing of peptide ions is achieved through gas-phase fragmentation. In an MS/MS experiment, a peptide precursor ion is selected based on its m/z ratio (e.g., with a quadruple mass filter) in MS^1 and activated with some form of energy introduction, resulting in cleavages of backbone chemical bonds. The resulting MS^2 spectrum is comprised of these peptide ion fragments, which can be cross-referenced with theoretical m/z values of possible fragment ions from genome-derived databases to identify the peptide. Ideally, backbone cleavage should occur at each possible inter-residue position; however, bioinformatics approaches allow high confidence peptide identification from partial sequence information [51-53].

The most common type of MS/MS activation is collision-induced dissociation (CID), which accelerates the peptide ions with a static or radiofrequency electric field to induce energetic collisions with a neutral gas, such as helium, nitrogen, or argon. The kinetic energy transferred from colliding with the gas raises the internal energy of the peptide ion and leads to fragmentation of the peptide backbone [24,25]. The "mobile proton model" (Figure 1.2) is the generally accepted mechanism for dissociation of peptide cations under collisional activation. This model postulates that, upon vibrational excitation, a proton of a cationic peptide will migrate to various, less basic, protonation sites, including the backbone amide bond. The carbonyl oxygen of the N-terminal, proximal peptide bond nucleophilically attacks the carbon center of the protonated amide bond forming a protonated oxazolone. This cyclic complex dissociates into a linear C-terminal fragment, and a cyclic N-terminal fragment (Figure 1.2).

4

Either or both fragments may carry charge (and thus be observable) depending on the charge state of the precursor peptide and the proton affinity of each fragment [25-30]. N- and C-terminal fragment ions from the associated amide bond cleavage are termed *b*- and *y*-type ions, respectively (Figure 1.3).
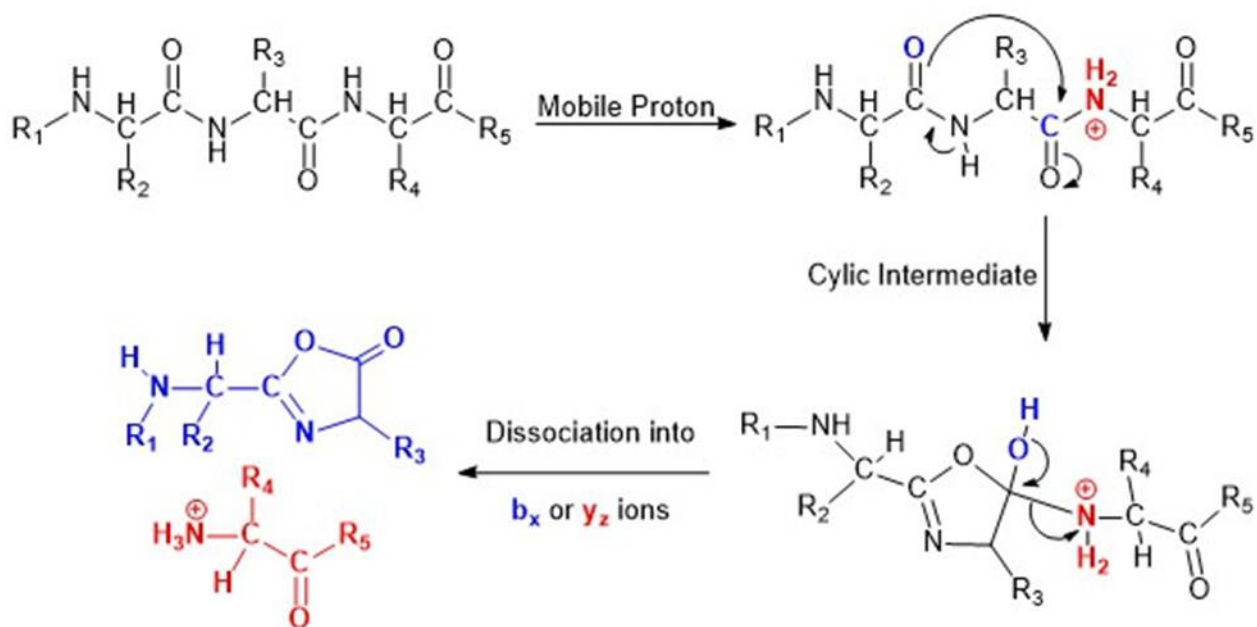
Figure 1.2 Fragmentation of a peptide bond according to the mobile proton model. Subsequent dissociation of the precursor ion forms either a *b*- or *y*- type fragment ion.

Higher-energy collision induced dissociation (HCD), a CID variant, uses static field acceleration into a separate collision cell to fragment precursor ions, in contrast to ion trap-type CID which uses an increased RF amplitude within the trap. The use of a separate collision cell avoids loss of lower molecular mass fragments due to changes in the trap stability region upon the RF amplitude change. Retention of low m/z fragments is useful for detecting protein modifications and isobaric labels [31]. Additionally, HCD (also referred to as beam-type CID) shows complementary fragmentation compared with ion trap-type CID, e.g., a greater abundance of *a*-type ions, resulting from further fragmentation of *b* ions via CO loss. However, one

5

drawback to CID/HCD-induced fragmentation methods is that they often induce cleavage of

peptide and protein PTMs, thus precluding direct determination of PTM sites.
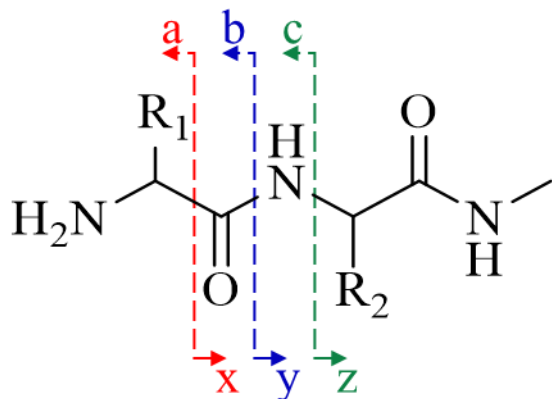
Figure 1.3 Peptide fragmentation patterns produced under different fragmentation conditions. CID and HCD produce mainly b- and y-type ions, while ETD produces mainly c- and z-type ions.

In contrast, electron transfer dissociation (ETD) induces fragmentation by transferring

electrons from a radical electron carrier, typically polycyclic aromatic compounds such as

fluoranthene, to higher charge state cationic molecules, including peptides and whole proteins.

ETD involves fragmentation of backbone N-Cα bonds in peptides/proteins, generating

complementary c- and z-type fragment ions (Figure 1.3), rather than the b- and y-type fragment

ions formed in CID/HCD methods. ETD is advantageous because it can retain labile PTMs

[33,34].

1.3 Liquid Chromatography- Mass Spectrometry for Proteomics

In a sample containing thousands of peptides, separation prior to MS is required to

provide resolution between isobaric and isomeric species. In addition, ionization suppression, an

inherent property of ESI, limits global peptide detection in complex samples. Analytes are

competing for charge and surface space in rapidly vaporizing ESI droplets, with analyte

saturation at the droplet surface preventing ejection of gas phase ions from analytes further inside the droplets [24]. Therefore, it is important to limit the abundance and co-ionization of competing analytes in a sample. Coupling liquid chromatography (LC) in-line with mass spectrometry (LC-MS) offers a powerful solution.

LC-MS/MS-based Proteomic Analysis

Cell Lysis   Protein Extraction,   Nano-UPLC   Annotation of MS² scans
                    Enrichment,   Orbitrap MS Analysis
                    and Digestion
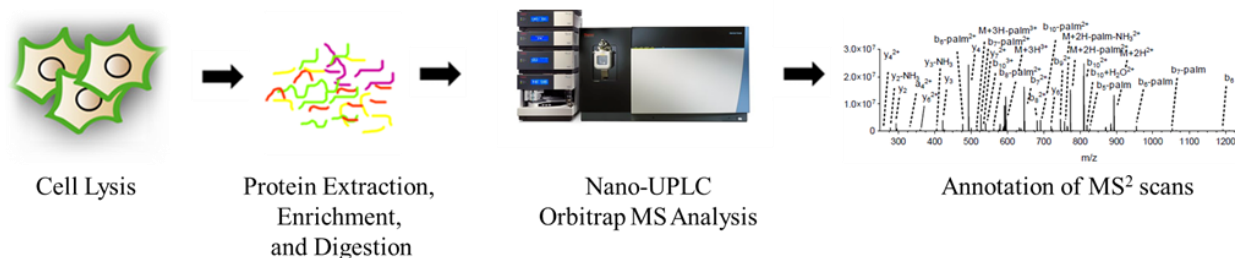
Figure 1.4 Schematic overview of the key steps in LC-MS/MS based shotgun/bottom-up proteomics.

The process of digesting protein samples into peptides, separating the peptides with LC, and analyzing the sample with inline ESI-MS/MS [35-37] is called "bottom-up" or "shotgun" proteomics. Generally, an LC-MS/MS "bottom-up" workflow (Figure 1.4) begins with the lysis of cells or tissues of interest followed by extraction of the host proteome and enzymatic cleavage, or digestion, of the proteins into their constituent peptides, typically less than 30 amino acids long. Protein digestion into similar sized peptides is also advantageous from an LC standpoint as intact proteins vary widely in size and physical properties and may therefore not all separate optimally under the same LC conditions.

Peptide mixtures are typically separated over a narrow-bore capillary column (75-100 µm inner diameter) packed with a particulate resin (1-5 µm diameter) at flow rates ~200 nL/min, referred to as nanoflow (nano)-LC. Reverse phase (RP) resins are silica particles modified with hydrocarbon chains, with 18-carbon chains (C-18) being the most common. Peptides interact

7

with the hydrocarbon chains and partition out of the liquid mobile phase. Mobile phase composition is changed over the course of the separation in a gradient manner, with a pump gradually adding organic solvent (typically acetonitrile or methanol) to the column. As the percentage of the organic solvent surrounding the resin increases, peptides partition back into the mobile phase and elute off the column. The principle behind separation is that strongly interacting solutes (i.e., more hydrophobic ones) retain on the column longer, thus effectively separating solutes in the order of their hydrophobicity [38-40].

While high-performance liquid chromatography (HPLC) affords excellent resolution, the high complexity of proteomic samples still presents a significant challenge. In complex digests, numerous peptides may have similar hydrophobicities and co-elute during the experiment. In addition, the abundance of each protein in a sample varies widely and low abundance peptides are typically not selected for MS/MS as instrument duty cycle determines how many MS/MS spectra can be generated per time unit. To combat these issues, there are many enrichment strategies intended to reduce sample complexity. Most widely used is affinity capture which immobilizes proteins of interest on a resin before washing away uncaptured background proteins. Proteins can be targeted for enrichment through strategies such as immunoprecipitation, chemical modification, or genetic engineering to contain enrichment tags [40-43]. Captured proteins are released from the resin through chemical or enzymatic degradation of the linker, proteolytic digestion, or eluting through other means. However, these enrichment strategies require proteins to be modified or to contain a common epitope. In untargeted studies, or studies of native proteins, such enrichment strategies may not be suitable. Therefore, adding additional dimensions of LC separation can also be exploited to improve bottom-up proteomics. Alternative

separation strategies such as capillary zone electrophoresis (CZE) have also recently been shown to be effective [108-110].

Common strategies for multidimensional separation include prefractionation on a resin complementary to the one used for LC/MS, or inline two-dimensional liquid chromatography (MudPIT) [125]. Chromatographic techniques complementing RP LC include size exclusion chromatography (SEC, which separates by molecular weight), ion-exchange chromatography (IEX, which separates by ionic strength), and hydrophilic interaction liquid chromatography (HILIC, which separates by hydrophilic interaction with a polar stationary phase). Pre-fractionation has the effect of greatly reducing chromatographic complexity in LC/MS but also decreases experimental throughput substantially depending on the number of fractions collected. 2D-LC-MS is an alternative to off-line fractionation and involves the sometimes-complicated addition of a second column to a traditional LC-MS setup. This strategy is limited by the compatibility of both columns with the mobile phase composition, and often requires plumbing of additional diverter valves to ensure that peptides are retained and that the columns are equilibrated [45-46]. The main advantage of 2D-LC-MS is improved detection limits and decreased sample loss. However, this strategy requires longer uninterrupted mass spectrometer operation.
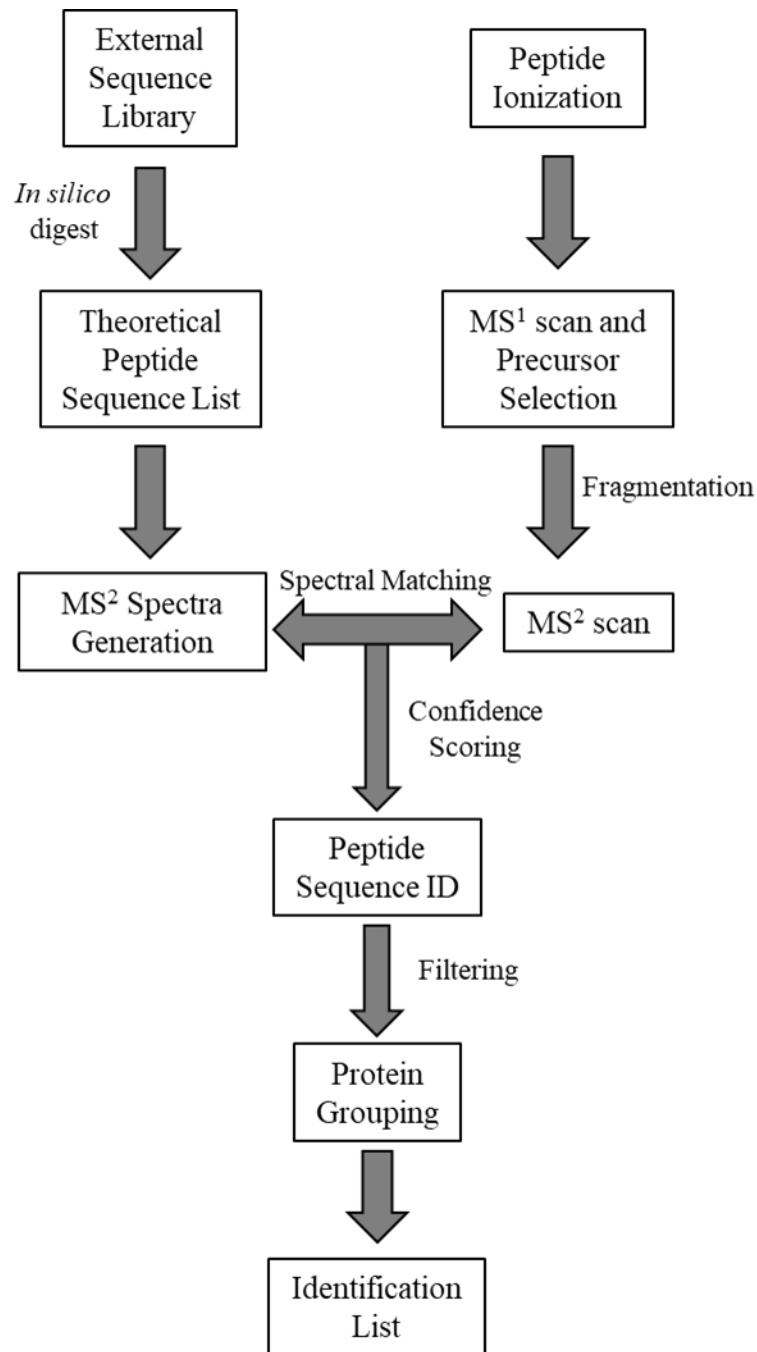
Figure 1.5 Computational analysis of LC-MS/MS data matches peaks from tandem MS scans against an *in silico* generated MS^2 spectral library. Spectral matches are scored for confidence and confident hits are filtered by search parameters. Peptide sequences are matched with protein groups, and a protein list is inferred.

1.4 Bioinformatics for LC/MS/MS-Based Proteomics

Modern academic and commercial software algorithms can process raw files of LC-MS/MS data and produce statistically valid lists of protein and peptide identifications in just minutes (Figure 1.5). Online repositories, such as Uniprot [1], provide databases of all protein sequences in an organism, and *in silico* digests of a downloaded proteome are stored as theoretical peptide lists. From these peptide lists, MS^2 spectra are generated *in silico* with predictable fragmentation patterns and are matched to raw MS^2 scans to identify the precursor peptides [48,50]. Scores are assigned to spectral matches and the highest scored sequence is stored in a list of peptides. Protein identifications are inferred by matching sequenced peptides in the theoretical peptide list grouped into proteins [50]. Peptide sequences and protein groupings are validated by evaluating each match against all other matches in the experiment. For statistical modeling, false spectra are generated from a database of reversed protein sequences and added to the theoretical peak list. Spectral matches to these reverse sequences are false discoveries. Statistical modeling of distribution scores from all matches is used to determine the minimum confidence score for valid hits and further filtering to remove identifications scored lower than the top percentile of false discovery scores, the false discovery rate (FDR), is performed to ensure the validity of identification lists [51-53].

1.5 Quantitative Proteomics

Mass spectrometry is not inherently quantitative and, thus, measuring protein abundance in LC-MS/MS experiments is a unique challenge. The detector measures the ion current throughout an analysis with the detector response being linearly proportionate to the number of ions within its dynamic range. This property is useful in quantitation where the relative intensity of a peak can be used to determine a change in abundance (i.e., relative quantitation). However,

the signal intensity produced per peptide depends on individual ionization efficiency, which varies based on the chemical environment and is subject to systematic error between experiments [54,57]. Thus, in the absence of a constant detector response, it is challenging to accurately measure a peptide's abundance using only signal intensity. Two types of strategies commonly used to circumvent this issue are 'labeling' and 'label-free' methods. Labeling works by altering the stable heavy isotope content/distribution in peptides, which changes their mass/mass distribution but not their ionization efficiency. Such labeling can be performed either via cell culture in standard/isotopically enriched media (stable isotope labeling of amino acids in cell culture (SILAC)) [126], by introducing "light" and "heavy" chemical tags, or tags with different isotope distribution designed to fragment easily during MS/MS. Labeled proteins from multiple experiments or conditions are mixed and enriched together for decreased experimental error. An early example of this strategy, developed in the late 1990's, is isotope coded affinity tags (ICAT) that utilizes isotopically labeled/unlabeled tags to derivatize and enrich free cysteines [58]. However, because cysteine is a relatively rare amino acid, this strategy is not universally applicable. By contrast, SILAC can label virtually all proteins but is only compatible with cell culture experiments. Also, both ICAT and SILAC double the complexity of LC-MS spectra as each peptide will appear as two distinct m/z signal clusters with the added signals potentially overlapping with other peptides. Isotopic labeling methods are useful but require additional experimental steps with expensive reagents. Additionally, stable isotope labels need to be designed carefully to prevent systematic errors caused by dissimilar behavior. For example, hydrogen/deuterium substitution is known to affect the retention time of the labeled peptides, while $^{12}C/^{13}C$ substitution does not show such an effect [60]. Newer technologies are commercially available that circumvent these issues by utilizing MS cleavable isobaric reagents

to label the primary amines of peptides and proteins. Isobaric tags for relative and absolute quantification (iTRAQ) [128] and tandem mass tags (TMT) [127] facilitate quantitative analysis with reporter groups that are generated upon fragmentation in the mass spectrometer. The fragmentation of the attached tag generates a low molecular mass reporter ion that can be used to relatively quantify the peptides and the proteins from which they originated. These tags can be designed for analysis of up to 16-multiplexed samples in a single experiment.

Label-free methods for quantitation are often used when peptide modification is undesirable or the cost of labeling is prohibitive [61-63]. In label-free workflows, samples are analyzed separately, and every workflow needs to be optimized for reproducibility to reduce systematic error. In contrast to labelling methods, each peptide ion MS peak is integrated and used as a measure of quantity [64,65]. The "Hi-3", or "Top-3," method [66] can be used to approximate the concentration of proteins in a sample. This method integrates the MS peaks of the top three ionizing peptides from each positive protein identification, averages the three signal intensities, and calculates the stoichiometry relative to the top three ionizing peptides from an internal standard. The internal standard, usually a tryptic digest of a protein standard from a different host organism, is spiked into the sample at a known molar concentration prior to LC-MS/MS analysis. Since the detected abundance of each peptide from a protein digest is influenced by workflow factors, such as digestion and ionization efficiency, measuring only the top three peptides reduces the risk for significant under-estimation that results from considering low intensity peptides in abundance calculations. Similarly, by averaging at least three peptides from each protein, the variability in detector response to peptide ions can be mitigated and a close approximation to absolute abundances can be more confidently made.

1.6 Approaches for Annotating Protein S-Palmitoylation

S-palmitoylation, here referred to as palmitoylation, is a reversible and dynamic post-translational modification (PTM) that results from the thioesterification of palmitate, a saturated sixteen-carbon fatty acid, to a cysteine residue of a protein. Protein palmitoylation was first reported in 1979, the same year as the discovery of tyrosine phosphorylation [67-69]. Similar to protein phosphorylation, it was quickly realized that palmitoylation is essential for intracellular signaling, and it is also critical in regulating the function of the Ras family of GTPases, Src-family kinases, G-proteins, and G-protein coupled receptors [70-72].
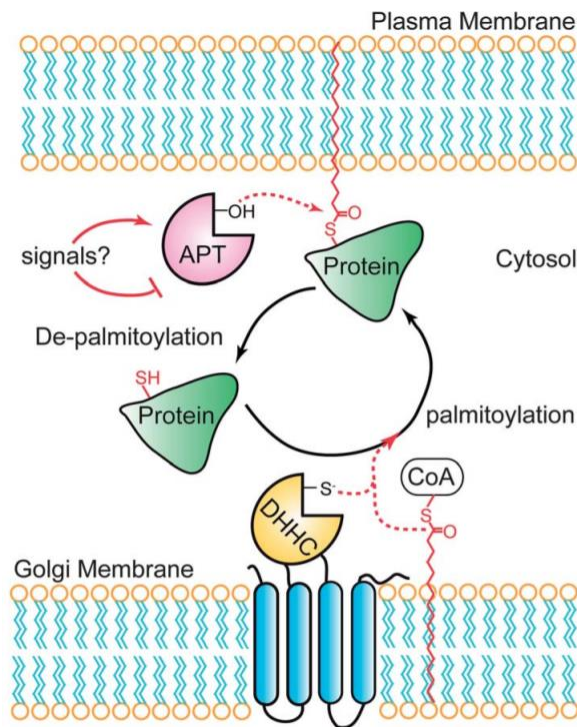
Figure 1.6 Dynamic cycling of protein palmitoylation is regulated by protein acyltransferases (PATs) and acyl protein thioesterases (APTs). Permission for reuse requested Graphic credit: Martin *et al.* [78].

The dynamic cycling between a protein's palmitoylated and depalmitoylated states (Figure 1.6) regulates several intracellular events such as protein trafficking, spatial localization,

membrane tethering, and the resultant cellular signaling and protein-protein interactions [70-76].

The potential for reversibility of the thioester bond, a labile and high-energy bond [76,77], between a palmitate group and a free cysteine residue is unique among the characterized lipid modifications. Importantly, it is this reversibility of palmitoylation that enables dynamic control of both spatial and temporal protein function [78-81]; however, the importance of palmitoylation has only recently gained significant attention.

This historical lack of investigational attention may be partly attributed to a shortage of robust strategies for directly profiling this non-polar, non-antigenic modification [81,82]. Palmitoylated peptides can be challenging to detect directly because the acyl-chain is highly non-polar and lacks sufficient antigenicity, rendering traditional PTM-specific enrichment strategies such as immunocapture techniques impractical. Additionally, thioester bonds are labile and prone to hydrolysis under conditions present in many protein analysis workflows, further complicating quantitative analyses and the detection of low abundant species.
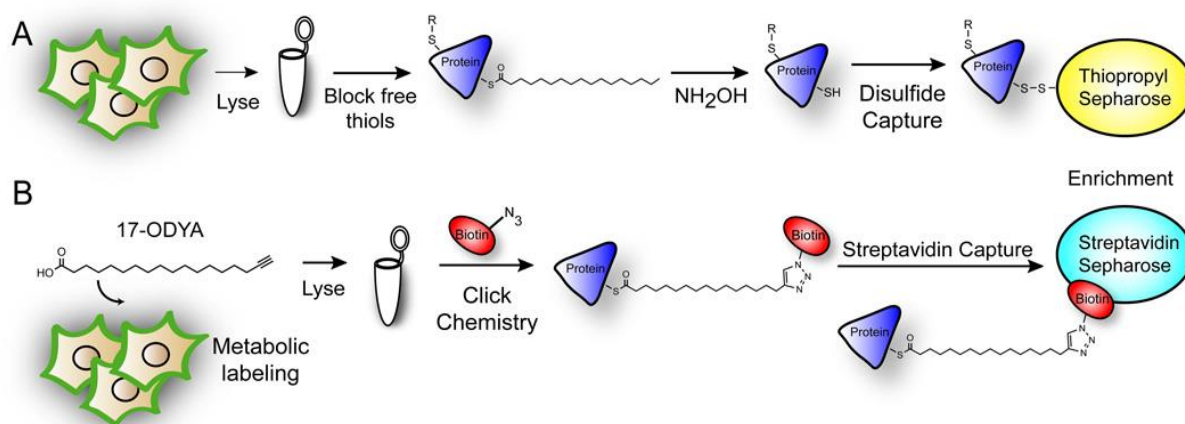
Figure 1.7 Current S-palmitoylation detection methods A) The acyl-exchange method blocks free thiols, hydrolyzes the thioesters, and then labels or captures the resulting free thiols for visualization and/or enrichment. B) The "click chemistry" method metabolically labels palmitoylation sites with 17-ODYA and uses copper catalysis to bioorthoganally ligate labeled sites to N_3-modified tags for visualization and/or enrichment. Permission for reuse requested. Graphic credit: Hernandez *et al.* [81].

Originally, the most direct way to monitor protein palmitoylation was through metabolic labelling with [³H]- palmitate, followed by lengthy exposure times ranging from days to weeks [67-69, 83]. Recent advances in chemical tagging and MS-based proteome-profiling techniques have greatly expanded databases, such as Swiss Palm [82], which catalogs reported palmitoylation sites, to include hundreds of putative targets for palmitoylation [84-92]. Common approaches to enriching for and profiling the palmitoylome involve affinity capture techniques such as acyl-biotin exchange (ABE) followed by fluorescent detection or resin assisted capture of S-acylation sites (acyl-RAC) [92,93]. ABE traditionally involves a three-step procedure (Figure 1.7(a), [81]) that includes: complete blocking of free thiols with N-ethylmaleimide (NEM), hydrolyzing all thioesters with hydroxylamine [76], and labeling free sulfhydryl groups with a fluorescent or biotinylated marker [86].

Another approach utilizes bioorthogonal chemical ligation, or "click chemistry" (Figure 1.7B,) [94-97]. This strategy involves metabolically labeling cells with alkynyl-analogues of palmitic acid, such as 17-octadecynoic acid (17-ODYA) [78,91]. The labeled proteins are then subjected to copper catalysis with azide (N_3-) containing detection markers, such as fluorescent tags or biotin, for enrichment and/or visualization. This approach has shown to be highly successful and proteomics analysis using metabolic labeling and enrichment of 17-ODYA labeled proteins was able to annotate more than 300 palmitoylated proteins in neuronal stem cells [78]. In contrast to ABE methods, the metabolic incorporation of 17-ODYA into cell cultures minimizes false positives generated by ABE protocols due to incomplete alkylation of free cysteines or capture of endogenous hydroxylamine-sensitive thioesters. There are limitations to this approach, however, and poor metabolic incorporation rates of the alkynyl-analogues, metabolic degradation of alkynyl-fatty acids, and aberrant labeling of the N-terminus by 17-

ODYA have been reported. Considering its potential for off-site labeling, sites reported as labeled with this analogue need further validation experiments to confirm the site is natively palmitoylated.

As novel approaches are developed the list of palmitoylated proteins continues to expand and the critical role of protein palmitoylation in the function of many membrane localized proteins is difficult to ignore. Therefore, in order to better understand the nature and extent of palmitoylation in biological systems, it is necessary to develop analytical strategies that can globally profile acylation, differentiate between types of acylation, localize palmitoylation sites, and quantitatively measure changes in protein palmitoylation under various conditions.

1.7 Host-Cell Protein Characterization in Biopharmaceutical Drug Substances

Host cell proteins (HCP) are native proteins derived from a host organism used to express biotherapeutic proteins for drug substances (DS), such as therapeutic antibodies and fusion proteins [98,99]. HCPs may be unintentionally co-purified with a DS as process-related impurities and the presence of these impurities may present problems in drug performance through eliciting immune responses in patients or contributing to the instability of drug formulations [99-102]. Due to their potential to affect product safety and efficacy, the concentration of residual HCPs in a drug product is generally considered a critical quality attribute (CQA) [103-105]. The ICH, a council of pharmaceutical regulatory bodies, defines a CQA as "a physical, chemical, biological or microbiological property or characteristic that should be within an appropriate limit, range, or distribution to ensure the desired product quality" [105]. Additionally, it is a US federal regulatory requirement to monitor the removal of HCPs in drug product during bioprocess development. Therefore, measuring HCP levels is essential

throughout manufacturing process development and at release testing to ensure adequate removal

of these impurities [106,107].
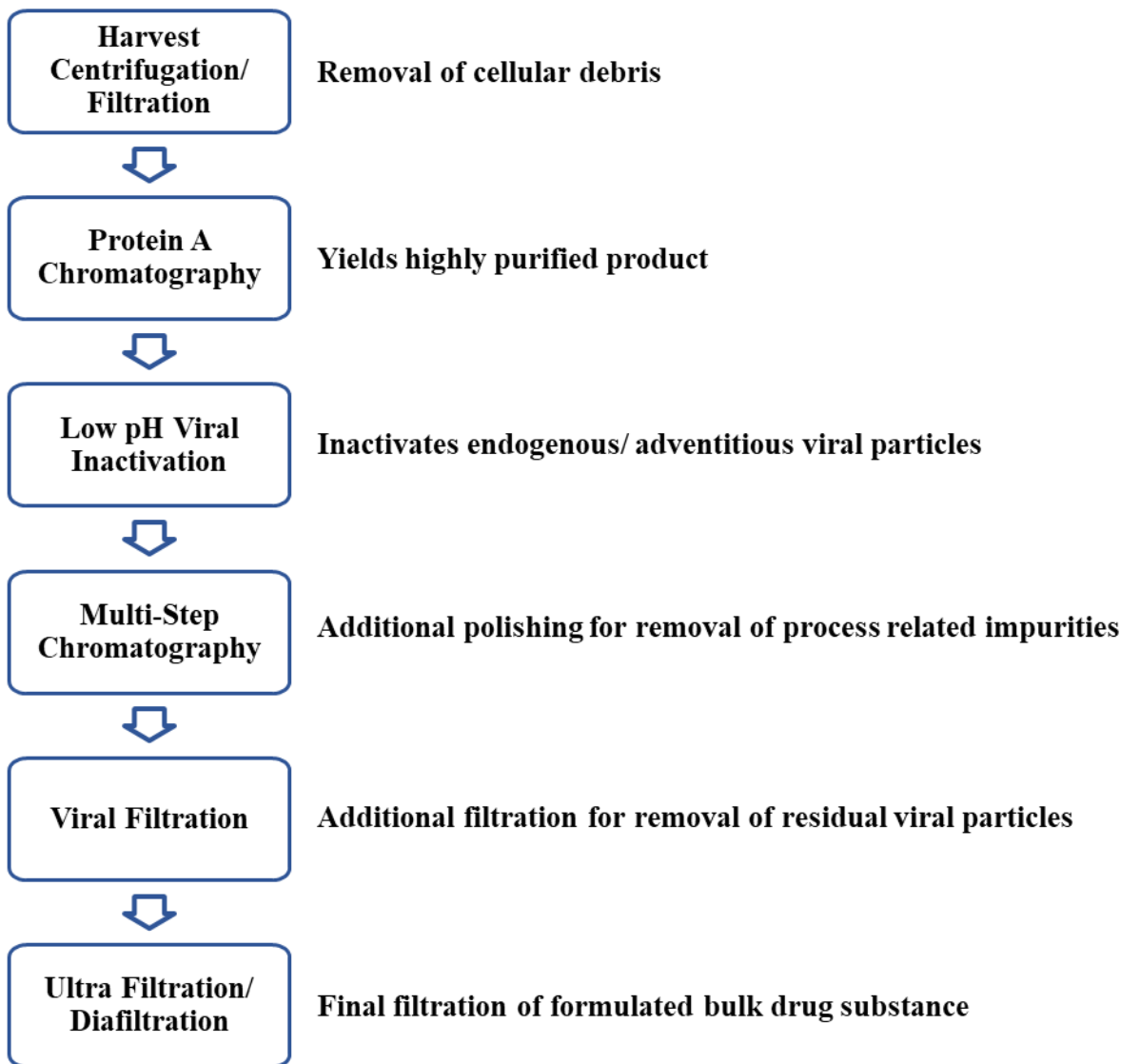
Figure 1.8 Downstream purification process for industrial-scale manufacturing of biotherapeutic antibodies. [123]

Multiple purification steps (Figure 1.8) ensure that residual HCPs are typically at very

low levels relative to the therapeutic protein, often 100 parts per million (ppm) or less

[106,122,123]. While these low levels are desirable, they present an analytical challenge for HCP

identification and quantitation. Enzyme-linked immunosorbent assay (ELISA) is currently the most common HCP analysis technique in industry as such assays offer high sensitivity, parallelizability, and quantitation of HCP content making them useful for support of routine analyses and process development [98,99,108]. Generic ELISAs for HCPs are typically created using polyclonal antibodies raised against antigen HCPs produced in culture from a process-representative null host cell line. However, it remains challenging to achieve complete antibody coverage of all possible HCP species present in a cell line, and due to the fact that not every HCP is immunogenic, even the best process-specific HCP ELISA cannot detect 100% of possible HCPs [107]. Furthermore, detection sensitivity may vary for specific HCPs depending on their respective immunogenicity, resulting in dilutional nonlinearity and potentially an underestimate of residual HCPs [108].

Orthogonal strategies have emerged with protein separation methods such as 1D and 2D sodium dodecyl sulfate- polyacrylamide gel electrophoresis (SDS-PAGE), 2D-differential in-gel electrophoresis (2D-DIGE), western blotting (WB), and capillary zone electrophoresis [108-110]. However, each of these methods presents practical limitations, such as poor sensitivity and/or the inability to distinguish drug related fragments from HCPs.

LC-MS/MS-based "bottom-up" proteomics offers a highly sensitive, semi-quantitative method for HCP detection and identification which can be effectively used to support downstream purification process development [111,112]. Using a 'bottom-up' approach, HCPs can be identified by separating proteins or peptides with 1D or 2D-LC [113], analyzing peptides by MS/MS [114], and conducting database searches against the entire proteome of the host organism. Measurements of absolute HCP concentrations can be approximated with the 'Top-3' quantitation method [66]. This technique is less accurate than other approaches such as surrogate

19

peptide methods, but it is advantageous for proteomic-type HCP analyses since specific standards are not required for each analyte [113,114].

Biopharmaceutical product typically contains HCP concentrations less than 100 ppm (ng of HCP per mg biotherapeutic protein), which presents a dynamic range problem for ESI LC-MS/MS analysis. Depletion of the biotherapeutic protein is necessary even when using state-of-the-art mass spectrometers in order to lower dynamic range to detect HCPs at concentrations less than 10 ppm [114-117]. Adapting chromatography with long LC gradients and multiple separation dimensions such as 2D-LC [113] and/or ion mobility separation [114] to resolve the low abundance peptides from the digested DS has been an effective approach. However, these strategies involve complex instrument modifications, and decrease processing throughput which limits general practicality. Other approaches include depleting monoclonal antibody (Ab) with affinity purification [113-116]; however, HCPs that strongly interact or aggregate with the biotherapeutic protein may not be enriched without strong denaturing conditions, limiting potential enrichment options. Furthermore, affinity purification strategies require careful sample-specific optimizations which are difficult to adapt into standardized protocols for downstream product analysis.

Another LC-MS/MS approach to enrich for residual HCPs, is to precipitate biotherapeutic Ab from a sample after overnight native trypsin digestion prior to analysis [117]. Developed at Eli Lilly, the principle hypothesis of this method is that the Ab in its rigid, folded structure will have a slower kinetic rate of digestion by trypsin than natively folded HCPs, and as a result, HCPs will be preferentially digested under native conditions. After digestion, the undigested proteins are heat precipitated and pelleted with centrifugation leaving the HCP-enriched supernatant for recovery and LC-MS/MS analysis. Benchmarking of this method with a

high-purity monoclonal antibody, from the National institute of Standards and Technology (NIST), resulted in detection of 60 HCPs which is twice as many as previously reported with 2D-LC/MS/MS approaches. In contrast to traditional 'bottom-up' approaches, this depletion strategy can lower the dynamic range up to two orders of magnitude for HCP detection. This approach does have a few drawbacks however, including the potential to lose heat-labile HCPs, co-precipitation of the HCPs with the Ab, and an inherent lack of general applicability for characterizing HCP content in non-antibody samples, such as fusion proteins, which may not have the same resistance to proteolysis [116,118]. Given the limitations of existing methods, novel strategies for sensitive and comprehensive characterization of HCP content in biotherapeutic drug substances are needed.

1.8 Dissertation Outline

The aim of the work described in this thesis is to develop novel LC-MS/MS workflow strategies, including sample preparation, chromatographic separation, and MS/MS activation, for characterizing and profiling challenging proteins and labile PTMs. Specifically, Chapter 2 describes work to adapt LC-MS/MS 'bottom-up' workflows for directly detecting native protein palmitoylation in proteomic analyses and Chapter 3 discusses our efforts developing a one plate LC-MS/MS workflow utilizing chromatographic fractionation of undigested drug substance for characterizing host-cell protein content in biotherapeutic drug substances. These findings can be applied broadly to inform workflow optimization for studying labile protein modifications and detecting low abundant proteins in complex mixtures. Chapter 4 provides a concluding summary of the novel contributions of our work and the broader impact on the field of LC-MS/MS based protein analysis.

1.9 References

1. UniProt: the universal protein knowledgebase. Nucleic acids research 2018, 46 (5), 2699.

2. Gaudet, P.; Michel, P.-A.; Zahn-Zabal, M.; Britan, A.; Cusin, I.; Domagalski, M.; Duek, P. D.; Gateau, A.; Gleizes, A.; Hinard, V., The neXtProt knowledgebase on human proteins: 2017 update. Nucleic acids research 2016, 45 (D1), D177-D182.

3. O'Farrell, P. H., High resolution two-dimensional electrophoresis of proteins. Journal of biological chemistry 1975, 250 (10), 4007-4021.

4. Milman, G.; Lee, E.; Ghangas, G. S.; McLaughlin, J. R.; George, M., Analysis of HeLa cell hypoxanthine phosphoribosyltransferase mutants and revertants by two-dimensional polyacrylamide gel electrophoresis: evidence for silent gene activation. Proceedings of the National Academy of Sciences 1976, 73 (12), 4589-4593.

5. Reeh, S.; Pedersen, S.; Friesen, J. D., Biosynthetic regulation of individual proteins in relA+ and relA strains of Escherichia coli during amino acid starvation. Molecular and General Genetics MGG 1976, 149 (3), 279-289.

6. Gibson, W., Polyoma virus proteins: a description of the structural proteins of the virion based on polyacrylamide gel electrophoresis and peptide analysis. Virology 1974, 62 (2), 319-336.

7. Garrels, J. I.; Gibson, W., Identification and characterization of multiple forms of actin. Cell 1976, 9 (4), 793-805.

8. Pedersen, S.; Reeh, S. V.; Parker, J.; Watson, R. J.; Friesen, J. D.; Fiil, N. P., Analysis of the proteins synthesized in ultraviolet light-irradiated Escherichia coli following infection with the bacteriophages λdrif d 18 and λdfus-3. Molecular and General Genetics MGG 1976, 144 (3), 339-343.

9. Gershoni, J. M.; Palade, G. E., Protein blotting: principles and applications. Analytical biochemistry 1983, 131 (1), 1-15.

10. Hewick, R. M.; Hunkapiller, M. W.; Hood, L. E.; Dreyer, W. J., A gas-liquid solid phase peptide and protein sequenator. Journal of Biological Chemistry 1981, 256 (15), 7990-7997.

11. Edman, P.; Begg, G., A protein sequenator. European Journal of Biochemistry 1967, 1 (1), 80-91.

12. Hunt, D. F.; Yates, J. R.; Shabanowitz, J.; Winston, S.; Hauer, C. R., Protein sequencing by tandem mass spectrometry. Proceedings of the National Academy of Sciences 1986, 83 (17), 6233-6237.

13. Biemann, K., Contributions of mass spectrometry to peptide and protein structure. Biological Mass Spectrometry 1988, 16 (1-12), 99-111.

14. Barber, M.; Bordoli, R. S.; Sedgwick, R. D.; Tyler, A. N. (September 1981). "Fast atom bombardment of solids as an ion source in mass spectrometry". Nature. 293 (5830): 270–275.

15. Stoll, R.G.; Harvan, D.J.; Hass, J.R.. Liquid secondary ion mass spectrometry with a focussed primary ion source. International Journal of Mass Spectrometry and Ion Processes.1984 61 (1): 71–79.

16. Dominic M. Desiderio . Mass Spectrometry of Peptides. CRC Press. 1990, 174.

17. De Pauw, E.; Agnello, A.; Derwa, F.Liquid matrices for liquid secondary ion mass spectrometry-fast atom bombardment: An update. Mass Spectrometry Reviews. 1991, 10 (4): 283–301.

18. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M., Electrospray ionization for mass spectrometry of large biomolecules. Science 1989, 246 (4926), 64-71. 85

19. Loo, J. A.; Edmonds, C. G.; Smith, R. D., Primary sequence information from intact proteins by electrospray ionization tandem mass spectrometry. Science 1990, 248 (4952), 201- 204.

20. Smith, R. D.; Loo, J. A.; Loo, R. R. O.; Busman, M.; Udseth, H. R., Principles and practice of electrospray ionization—mass spectrometry for large polypeptides and proteins. Mass Spectrometry Reviews 1991, 10 (5), 359-452.

21. Smith, R. D.; Loo, J. A.; Edmonds, C. G.; Barinaga, C. J.; Udseth, H. R., New developments in biochemical mass spectrometry: electrospray ionization. Analytical Chemistry 1990, 62 (9), 882-899.

22. Wilm, M., Principles of electrospray ionization. Molecular & Cellular Proteomics 2011, mcp. R111. 009407.

23. Karas M. and Hillenkamp, F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. Anal. Chem. 1988, 60, 20, 2299–2301

24. Annesley, Thomas M. Ion Suppression in Mass Spectrometry. Clinical Chemistry. 2003, 49 (7): 1041–1044

25. Johnson, R.S.; Martin, S.A.; Biemann, K. Collision-Induced Fragmentation of (M+H)+ Ions of Peptides. Side Chain Specific Sequence Ions. Int. J. Mass Spectrom. Ion Processes. 1988, 86: 137-54.

26. Wysocki, V.H.; Tsaprailis, G.; Smith, L.L.; et al. Mobile and localized protons: a framework for understanding peptide dissociation. J Mass Spectrom. 2000 35(12): 1399-406.

27. Harrison, A.G.; Yalcin, T. Proton mobility in protonated amino acids and peptides. Int J Mass Spectr Ion Process. 1997, 165: 339-47.

28. Johnson, R.S.; Krylov, D.; Walsh, K.A. Proton mobility within electrosprayed ions. J Mass Spectrom. 1995, 30: 386-7.

29. Jorgensen, T.J.; Bache, N.; Roepstorff, P.; et al.. Collisional activation by MALDI tandem time-offlight mass spectrometry induces intramolecular migration of amide hydrogens in protonated peptides. Mol Cell Proteomics. 2005, 4(12): 1910-9.

30. Jorgensen, T.J.; Gardsvoll, H.; Ploug, M.; et al. Intramolecular migration of amide hydrogens in protonated peptides upon collisional activation. J Am Chem Soc. 2005, 127(8): 2785-93.

31. Olsen JV, Macek B, Lange O, Mathias, M. Higher-energy C-trap dissociation for peptide modification analysis. Nat Methods. 2007, (4): 709-712

32. Gross, M.J. Charge-remote fragmentation: an account of research on mechanisms and applications. Int. J. Mass Spec. 2000, (200) 611–624

33. Han H, Xia Y, McLuckey SA. Ion Trap Collisional Activation of c and z Ions Formed via Gas-Phase Ion/Ion Electron Transfer Dissociation J Proteome Res. 2007, 6(8): 3062-3069.

34. Zhurov KO, Fornelli L, Wodrich MD et al. Principles of electron capture and transfer dissociation mass spectrometry applied to peptide and protein structure analysis. Chem Soc Rev. 2013, 42: 5014-5030.

35. Gatlin, C. L.; Kleemann, G. R.; Hays, L. G.; Link, A. J.; Yates III, J. R., Protein identification at the low femtomole level from silver-stained gels using a new fritless electrospray interface for liquid chromatography–microspray and nanospray mass spectrometry. Analytical biochemistry. 1998, 263 (1), 93-101.

36. Martin, S. E.; Shabanowitz, J.; Hunt, D. F.; Marto, J. A., Subfemtomole. MS and MS/MS peptide sequence analysis using nano-HPLC micro-ESI fourier transform ion cyclotron resonance mass spectrometry. Analytical Chemistry 2000, 72 (18), 4266-4274.

37. Ducret, A.; Oostveen, I. V.; Eng, J. K.; Yates III, J. R.; Aebersold, R.. High throughput protein characterization by automated reverse-phase chromatography/electrospray tandem mass spectrometry. Protein Science. 1998, 7 (3), 706-719.

38. Karlsson, K. E.; Novotny, M., A miniature gradient elution system for liquid chromatography with packed capillary columns. Journal of High Resolution Chromatography 1984, 7 (7), 411-413.

39. Covey, T. R.; Lee, E. D.; Bruins, A. P.; Henion, J. D., Liquid chromatography/mass spectrometry. Analytical chemistry 1986, 58 (14), 1451A-1461A.

40. Zhao Y, Gu H, Zheng N, Zeng J. Critical considerations for immunocapture enrichment LC–MS bioanalysis of protein therapeutics and biomarkers. Bioanalysis 2018, 10(13), 987–995

41. Fung EN, Bryan P, Kozhich A. Techniques for quantitative LC–MS/MS analysis of protein therapeutics: advances in enzyme digestion and immunocapture. Bioanalysis 2016, 8(8), 847–856

42. Becker JO, Hoofnagle AN. Replacing immunoassays with tryptic digestion-peptide immunoaffinity enrichment and LC–MS/MS. Bioanalysis 2012 4(3), 281–290

43. Sucharski FK, Meier S, Miess C et al.Development of an automated, interference-free, 2D-LC–MS/MS assay for quantification of a therapeutic Ab in human sera. Bioanalysis 2018 10(13), 1023–1037

44.   Kemp, G. et al. Capillary electrophoresis: a versatile family of analytical techniques. Biotechnology and Applied Biochemistry. 1998, 27 (1): 9–17

45. Shi T, Qian W-J. Antibody-free PRISM-SRM for multiplexed protein quantification: is this the new competition for immunoassays in bioanalysis? Bioanalysis 2013 5(3), 267–269

46. Sandra K, Mortier K, Jorge L et al.LC–MS/MS quantification of next-generation biotherapeutics: a case study for an IgE binding Nanobody in cynomolgus monkey plasma. Bioanalysis 2014, 6(9), 1201–1213

47. McCormack, A. L.; Schieltz, D. M.; Goode, B.; Yang, S.; Barnes, G.; Drubin, D.; Yates, J. R., Direct analysis and identification of proteins in mixtures by LC/MS/MS and database searching at the low-femtomole level. Analytical chemistry 1997, 69 (4), 767-776.

48. Eng, J. K.; McCormack, A. L.; Yates, J. R., An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. Journal of the American Society for Mass Spectrometry 1994, 5 (11), 976-989.

49. Yates, J. R.; Eng, J. K.; McCormack, A. L.; Schieltz, D., Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. Analytical chemistry 1995, 67 (8), 1426-1436.

50. Nesvizhskii, A. I.; Aebersold, R., Interpretation of shotgun proteomic data the protein inference problem. Molecular & cellular proteomics 2005, 4 (10), 1419-1440.

51.   Keller, A.; Nesvizhskii, A. I.; Kolker, E.; Aebersold, R., Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Analytical chemistry 2002, 74 (20), 5383-5392.

52. Nesvizhskii, A. I.; Vitek, O.; Aebersold, R., Analysis and validation of proteomic data generated by tandem mass spectrometry. Nature methods 2007, 4 (10), 787.

53. Choi, H.; Ghosh, D.; Nesvizhskii, A. I., Statistical validation of peptide identifications in large-scale proteomics using the target-decoy database search strategy and flexible mixture modeling. Journal of proteome research 2007, 7 (01), 286-292.

54. Nikolov M, Schmidt C, Urlaub H. Quantitative mass spectrometry-based proteomics: an overview. Methods in Molecular Biology. 2012, 893. pp. 85–100

55. Addona, TA, et al. Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. NatBiotechnol. 2009, 27:633–41.

56. Liu H, Sadygov RG, Yates JR., 3rd A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal Chem. 2004;76:4193–201.

57. Oda Y, Huang K, Cross FR, Cowburn D, Chait BT. Accurate quantitation of protein expression and site-specific phosphorylation. Proc Natl Acad Sci USA. 1999;96:6591–6

58. Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat. Biotechnol. 1999, 17, 994– 999

59. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics. 2002 May; 1(5):376-86

60. Yi EC, Li XJ, Cooke K, Lee H, Raught B, Page A, Aneliunas V, Hieter P, Goodlett DR, Aebersold, R. Increased quantitative proteome coverage with (13)C/(12)C-based, acid-cleavable isotope-coded affinity tag reagent and modified data acquisition scheme. Proteomics. 2005; 5(2):380-7.

61. Schulz-Knappe P, Zucht HD, Heine G, Jurgens M, Hess R, Schrader M. Peptidomics: the comprehensive analysis of peptides in complex biological mixtures. Comb Chem High Throughput Screen. 2001;4:207–17.

62. Wang W, Zhou H, Lin H, Roy S, Shaler TA, Hill LR, Norton S, Kumar P, Anderle M, Becker CH. Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. Anal Chem. 2003;75:4818–26.

63. Wiener MC, Sachs JR, Deyanova EG, Yates NA. Differential mass spectrometry: a label-free LC-MS method for finding significant differences in complex peptide and protein mixtures. Anal Chem. 2004;76:6085–96.

64. Clough T., Key M., Ott I., Ragg S., Schadow G., Vitek O. Protein quantification in label-free LC-MS experiments. J. Proteome Res. 2009;8, 5275–5284

65. Griffin N. M., Yu J., Long F., Oh P., Shore S., Li Y., Koziol J. A., Schnitzer J. E. Label-free, normalized quantification of complex mass spectrometry data for proteomic analysis. Nat. Biotechnol. 2010; 28, 83–89

66. Fabre, B., Lambour, T., Bouyssié, D., Menneteau, T., Monsarrat, B., Burlet-Schiltz, O. & Bousquet-Dubouch, M.-P. Comparison of label-free quantification methods for the determination of protein complexes subunits stoichiometry. EuPA Open Proteomics 2014; 4, 82–86

67. Schmidt, M.F.G., Schlesinger, M.J., Fatty acid binding to vesicular stomatitis virus glycoprotein: a new type of post-translational modification of the viral glycoprotein. Cell, 1979. 17: p. 813-819.

68. Schmidt MF, Bracha M, Schlesinger MJ. Evidence for covalent attachment of fatty acids to Sindbis virus glycoproteins. Proc Natl Acad Sci U S A. 1979;76:1687–1691.

69. Eckhart, W., Hutchinson, M.A., Hunter, T., An activity phosphorylating tyrosine in polyoma T antigen immunoprecipitates. Cell, 1979. 18: p. 925-933.

70. Smotrys JE, Linder ME. Palmitoylation of intracellular signaling proteins: regulation and function. Annu Rev Biochem. 2004; 73():559-87.

71. Eisenberg, S. et al. The role of palmitoylation in regulating Ras localization and function. Biochem. Soc. Trans. 2013; 41, 79–83

72. Bizzozero, O.A., Malkoski, S. P., Mobarak, C., Bixler, H. A., Evans, J. E. , Mass-spectrometric analysis of myelin proteolipids reveals new features of this family of palmitoylated membrane proteins. J. Neurochem. , 2002. 81: p. 636−645.

73. Roos, M., Soskic, V., Poznanovic, S., Godovac-Zimmermann, J., Post-translational Modifications of Endothelin Receptor B from Bovine Lungs Analyzed by Mass Spectrometry. J. Biol. Chem. , 1998. 273: p. 924−931

74. Salaun, C.G., J.; Chamberlain, L. H., The intracellular dynamic of protein palmitoylation. J. Cell Biol., 2010. 191: p. 1229−1238.

75. Linder, M.E. & Deschenes, R.J. Palmitoylation: policing protein stability and traffic Nat. Rev. Mol. Cell Biol. 2007; 8, 74–84 .

76. Magee AI, Koyama AH, Malfer C, Wen D, Schlesinger MJ. Release of fatty acids from virus glycoproteins by hydroxylamine. Biochimica et Biophysica Acta (BBA) - General Subjects. 1984;798:156–166.

77. Rose JK, Adams GA, Gallione CJ. The presence of cysteine in the cytoplasmic domain of the vesicular stomatitis virus glycoprotein is required for palmitate addition. Proc Natl Acad Sci U S A. 1984;81:2050–2054.

78. Martin BR, Wang C, Adibekian A, Tully SE, Cravatt BF Global profiling of dynamic protein palmitoylation. Nat Methods. 2011; 9(1):84-9.

79. Topinka, J.R. & Bredt, D.S. N-Terminal Palmitoylation of PSD-95 Regulates Association with Cell Membranes and Interaction with K1 Channel Kv1.4. Neuron. 1998; 20, 125–134

80. Chan, P. et al. Autopalmitoylation of TEAD proteins regulates transcriptional output of the Hippo pathway. Nat. Chem. Biol. 2016; 12, 282–289.

81. Hernandez, J.L., Majmudar, J.D., Martin, B.R., Profiling and Inhibiting Reversible Palmitoylation. Curr Opin Chem Biol, 2013. 17(1): p. 20-26

82. Blanc, M., David, F., Abrami, L., Migliozzi, D., Armand, F., Bürgi, J., van der Goot, F. G. , SwissPalm: Protein Palmitoylation database. F1000Research, 2015. 4(261).

83. Sefton, B.M., Trowbridge, I.S., Cooper, J.A., The transforming proteins of Rous sarcoma virus, Harvey sarcoma virus and Abelson virus contain tightly bound lipid. Cell, 1982. 31: p. 465–474.

84. Kathayat, R., Elvira, P. & Dickinson, B. A fluorescent probe for cysteine depalmitoylation reveals dynamic APT signaling. Nat Chem Biol. 2017; 13, 150–152

85. Hang HC, Linder ME. Exploring protein lipidation with chemical biology. Chem Rev. 2011; 111(10):6341-58.

86. Drisdel, R.C., Green, W.N., Labeling and quantifying sites of protein palmitoylation. Biotechniques, 2004. 36: p. 276–285.

87. Charron, G., Zhang, M.M., Yount, J.S., et al., Robust fluorescent detection of protein fatty-acylation with chemical reporters. J Am Chem Soc, 2009. 131: p. 4967–4975.

88. Peng, T., Thinon, E., Hang, H.C. Proteomic analysis of fatty-acylated proteins. Curr. Opin. Chem. Biol. 2016 30, 77–86

89. Tate EW, Kalesh KA, Lanyon-Hogg T, Storck EM, Thinon E. Global profiling of protein lipidation using chemical proteomic technologies. Curr Opin Chem Biol. 2015; 24:48-57.

90. Kang, R., Wan, J., Arstikaitis, P., Neural palmitoyl-proteomics reveals dynamic synaptic palmitoylation. Nature, 2008. 456: p. 904-909

91. Roth, A.F., Wan, J., Bailey, A.O., et al, Global analysis of protein palmitoylation in yeast. Cell. 2006;125:1003–1013. Cell, 2006. 125: p. 1003-1013.

92. Roth, A.F., Wan, J., Bailey, A.O., et al, Global analysis of protein palmitoylation in yeast. Cell. 2006;125:1003–1013. Cell, 2006. 125: p. 1003-1013.

93. Forrester MT, Hess DT, Thompson JW, Hultman R, Moseley MA, Stamler JS, Casey PJ. Site-specific analysis of protein S-acylation by resin-assisted capture. J Lipid Res. 2011; 52(2):393-8.

94. E. Saxon, C.R. Bertozzi. Cell surface engineering by a modified Staudinger reaction. Science. 2000; 287. 2007-2010

95. Q. Wang, et al. Bioconjugation by copper(I)-catalyzed azide-alkyne [3 + 2] cycloaddition. J. Am. Chem. Soc. 2003;125, pp. 3192-3193

96. Raghavan AS, Hang HC. Seeing small molecules in action with bioorthogonal chemistry. Drug Discov Today. 2009;14(3-4):178-84.

97. Zhang MM, Tsou LK, Charron G, Raghavan AS, Hang HC. Tandem fluorescence imaging of dynamic S-acylation and protein turnover. Proc Natl Acad Sci U S A. 2010;107(19):8627-32.

98. Wang, F., Richardson, D., Shameem, M., Host-Cell Protein Measurement and Control. BioPharm Int., 2015. 28: p. 32-38.

99. Hogwood, C.E.M., Bracewell, D.G., Smales, C.M., Host cell protein dynamics inrecombinant CHO cells. Bioengineered, 2013. 4: p. 288-291.

100. Abiri, N., Pang, J., Ou, J., Shi, B., Wang, X., Zhang, S., Sun, Y., Yan, D. , Assessment of the immunogenicity of residual host cell protein impurities of OsrHSA. PLOS ONE, 2018. 13.

101. Moussa EM, Panchal JP, Moorthy BS, et al. Immunogenicity of therapeutic protein aggregates. J Pharm Sci 2016; 105: 417-430

102. Jefferis R. Posttranslational modifications and the immunogenicity of biotherapeutics. J Immunology Res 2016; article ID 5358272

103. Atl N, Zhang TY, Motchnik P, et al. Determination of critical quality attributes for monoclonal antibodies using quality by design principles. Biologicals 2016; 44: 291-305.

104. Vandekerckhove K, Seidl A, Gutka H, et al. Rational selection, criticality assessment, and tiering of quality attributes and test methods for analytical similarity evaluation of biosimilars. AAPS Journal 2018; 20: 68.

105. ICH Q8 (R2) Pharmaceutical development. 2009

106. Feist, P., Hummon, A.B., Proteomic Challenges: Sample Preparation Techniques for Microgram-Quantity Protein Analysis from Biological Samples Int. J. Mol. Sci., 2015. 16: p. 3537-3563.

107. Bracewell, D.G., Francis, R., Smales, C.M., The future of host cell protein (HCP) identification during process development and manufacturing linked to a risk-based management for their control. Biotechnol. Bioeng., 2015. 112: p. 1727-1737.

108. Krawitz, D.C., Forrest, W., Moreno, G.T., Kittleson, J., Champion, K.M., Proteomic studies support the use of multi-product immunoassays to monitor host cell protein impurities. Proteomics, 2006. 6: p. 94-110

109. Berkelman, T., Harbers, A., Bandhakavi S, Proteomic Profiling Methods in Molecular Biology. Humana Press, New York, NY, 2015.

110. Schenauer, M., Flynn, G., Goetze, A., Identification and quantification of host cell protein impurities in biotherapeutics using mass spectrometry. Anal. Biochem., 2012. 428: p. 150-157.

111. Zhang, Q., Goetz, A.M., Cui, H., Wylie, J., Trimble, S., Hewig, A., Flynn, G.C., Comprehensive tracking of host cell proteins during monoclonal antibody purifications using mass spectrometry. Abs, 2014. 6: p. 659-670.

112. Doneanu, C., Xenopoulos, A., Fadgen, K., Murphy, J., Skilton, S.J., Prentice, H., Stapels, M., Chem, W., Analysis of host-cell proteins in biotherapeutic proteins by comprehensive online two-dimensional liquid chromatography/mass spectrometry. Abs, 2012. 4: p. 24-44.

113. Madsen, J.A., Frauntin, V., Carbeau, T., Wudyka, S., Yin, Y. Smith, S., Anderson, J., Capila, I., Toward the complete characterization of host cell proteins in biotherapeutics via affinity depletions, LC-MS/MS, and multivariate analysis. Abs, 2015. 7: p. 1128-1137.

114. Doneanu, C.E., Anderson, M., Williams, B.J., Lauber, M.A., Chakraborty, A., Chen, W., Enhanced Detection of Low-Abundance Host Cell Protein Impurities in High-Purity Monoclonal Antibodies Down to 1 ppm Using Ion Mobility Mass Spectrometry Coupled with Multidimensional Liquid Chromatography. Anal. Chem., 2015. 87: p. 10283-10291.

115. Lagassé, H.A.D., Alexaki, A., Simhadri, V.L., Katagiri, N.H., Jankowski, W., Sauna, Z.E., Kimchi-Sarfaty, C., Recent advances in (therapeutic protein) drug development. F1000Research, 2017. 6: p. 113-114.

116. Huang, L., Wang, N., Mitchell, C.E., Brownlee, T., Maple, S.R., De Felippis, M.R., A Novel Sample Preparation for Shotgun Proteomics Characterization of HCPs in Antibodies. Anal. Chem., 2017. 89: p. 5436-5444.

117. Moreland, L.W., et al. Treatment of Rheumatoid Arthritis with a Recombinant Human Tumor Necrosis Factor Receptor (p75)–Fc Fusion Protein. N. Engl. J. Med., 1997. 337: p. 141-147.

118. Wang, S., Liu, A.P., Yan, Y., Daly, T.J., Li, N., Characterization of product-related low molecular weight impurities in therapeutic monoclonal antibodies using hydrophilic interaction chromatography coupled with mass spectrometry J. Pharm. Biomed. Anal., 2018. 154: p. 468-475.

119. Wang, Q., Protein-Protein Interaction Analysis: Expanded Hydrogen/Deuterium Exchange Tandem Mass Spectrometry and Host Cell Protein Characterization, in Chemistry. 2019, University of Michigan. p. 172.

120. lpert, A. J. Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. Journal of Chromatography. 1990; (499) 177–196

121. K. Gekko, E. Ohmae, K. Kameyama, T. Takagi. Acetonitrile-protein interactions: amino acid solubility and preferential solvation. Biochim Biophys Acta, 1998; 1387 pp. 195-205

122. Shukla AA, Hubbard B, Tressel T, Guhan S, Low D. Downstream processing of monoclonal antibodies--application of platform approaches. Journal of Chromatography. B, Analytical

Technologies in the Biomedical and Life Sciences. Polyclonal and Monoclonal Antibody

Production, Purification, Process and Product Analytics. 2007; 848 (1): 28–39

123. Liu HF, Ma J, Winter C, Bayer R. Recovery and purification process development for

monoclonal antibody production. mAbs. 2009; 2 (5): 480–99

124. McLuckey, S.A. and Goeringer D. E. Slow Heating Methods in Tandem Mass

Spectrometry. J. Mass Spectrom. 1997; 32, 461-474

125. Lohrig, K., Wolters, D. Multidimensional Protein Identification Technology. 2009; 564,

143-153

126. Ong S. and Mann, M. Stable Isotope Labeling by Amino Acids in Cell Culture for

Quantitative Proteomics. Methods Mol Biol. 2007; 359, 37-52.

127. Schäfer T.A, *et al.* Tandem mass tags: a novel quantification strategy for comparative

analysis of complex protein mixtures by MS/MS. Anal. Chem. 2003; 75 (8), 1895–904

128. Wiese, S., Reidegeld K.A., Meyer, H.E., Warscheid B. Protein Labeling by iTRAQ: A New

Tool for Quantitative Mass Spectrometry in Proteome Research Proteomics. 2007; 7 (3)

340-50

Chapter 2

Direct Identification and Site-Specific Analysis of Recombinant N-Ras Palmitoylation

by Liquid Chromatography-Tandem Mass Spectrometry

2.1 Introduction

S-palmitoylation, referred to as palmitoylation, is a reversible and dynamic post-translational modification (PTM) that results from the thioesterification of palmitate, a saturated sixteen-carbon fatty acid, to a cysteine residue of a protein. This dynamic modification (Figure 1.6) is essential for intracellular signaling, and is critical in regulating the function of membrane-associated signaling proteins such as the Ras family of signaling proteins (Figure 2.1) [1,2].

The Ras family of proteins consists of small GTPases (guanosine triphosphatases) that play a role in signaling, cell proliferation, differentiation, and survival [3]. They are well-known drivers of many cancers and thus represent attractive targets for the development of anticancer therapeutics. All four Ras proteins (Figure 2.1) are farnesylated on a C-terminal CaaX motif, which is important for membrane association. However, the farnesyl group alone confers weak membrane association, and palmitoylation is needed for stable membrane attachment. In fact, drugs inhibiting farnesylation of palmitoylated Ras have failed to promote significant membrane dissociation of targeted Ras proteins, and S-palmitoylation was sufficient to retain membrane

localization and signaling behavior [4]. However, relatively little is known about cellular regulation of palmitoylation, and strategies for inhibiting Ras palmitoylation remain elusive. Therefore, studying the dynamics of palmitoylation is vital for developing new therapies targeting protein palmitoylation.
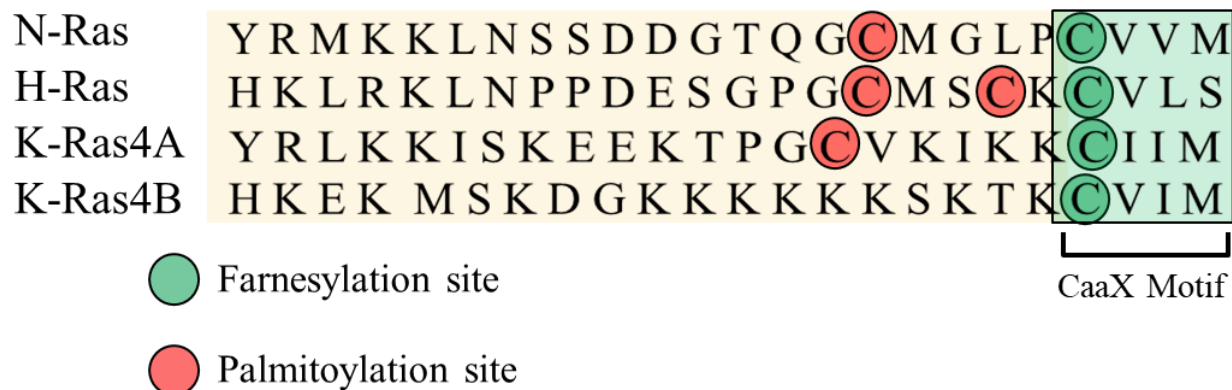
Figure 2.1 N-Ras is post-translationally farnesylated and then methylated at the clipped C-terminal CaaX motif (green) before being palmitoylated (red).

Common biochemical analysis strategies for S-palmitoylation analysis involve chemolytic removal of palmitate followed by labeling of the free sulfhydryl with either a fluorescent marker for detection, or resin assisted capture of S-acylation sites (acyl-RAC) with a sulfhydryl reactive reagent (Figure1.7(a)) [5-7,10]. S-palmitoylation sites have also been identified in large scale studies through metabolic labeling in culture with alkynyl analogues of palmitic acid (Figure 1.7(b))[8,9,13]. These alkynyl analogues can be coupled to an azide-linked fluorescent tag or an affinity reagent via copper-catalyzed 'click chemistry' [9-14] Site specific analysis typically has involved the introduction of point mutations, which can be tedious or impractical, and this approach is particularly challenging when a protein contains dozens of cysteine residues but may only undergo a single modification [10].

Modern high-resolution, liquid chromatography-tandem mass spectrometry (LC-MS/MS) techniques enable large scale protein sequencing with site-specific resolution for abundant and

stable PTMs) [15-17]. In "bottom-up" proteomics, proteins and their PTMs are profiled by separating proteolytic peptides over a LC column, analyzing them by online MS/MS, and compiling lists of proteins identified through database searches against the entire proteome of the host organism (Figure 1.4). However, robust and sensitive methods for the direct detection and profiling of palmitoylated proteins by LC-MS remain elusive. Palmitoylation sites are predicted for more than 1,000 proteins, yet only a small percentage have been experimentally validated [21-23].

LC-MS/MS 'bottom up' proteomics protocols calls for the initial enrichment of protein classes of interest from cell lysates, digestion of the proteins with trypsin into their constituent peptides, sample clean up steps to remove buffer contaminants and concentrate the sample, separation on reverse phase (RP) columns with high-performance LC, and subsequent analysis with high resolution MS/MS. While these methods are robust and highly generalizable for typical proteomics experiments, optimization of each step based on sample-specific considerations is necessary. Each of the aforementioned steps presents specific PTM-related challenges for analyzing S-palmitoylated proteins that are not easily enriched and can undergo palmitoyl loss during sample processing [20], hydrophobic peptides are difficult to maintain in solution, are strongly retained on RP columns, and labile PTMs are typically lost during conventional MS/MS analysis..

S-palmitoylated proteins are typically membrane bound (Figure 1.6) and can be insoluble in aqueous buffers potentially making them prone to aggregation and resistant to digestion by trypsin without the use of a surfactant. Acyl thioesters are labile functional groups that have a propensity to hydrolyze in the presence of bases and/or heat [24,25], and they are known to be labile in the presence of the reducing agent dithiothreitol (DTT) [20]. Also, lipidated peptides

may be retained on solid-phase extraction (SPE) columns during clean up steps. Furthermore, the long fatty acyl chain may contribute to poor ionization efficiency, a prerequisite for sensitive detection and efficient fragmentation in MS^2. Additionally, acyl thioester linkages have been reported to be labile under collision-induced dissociation (CID), leading to neutral loss of the fatty acyl group, further complicating identification of acylated peptides [20].

Important considerations in limiting palmitoyl loss are the desalting and buffer exchange steps during sample cleanup for LC-MS based workflows. After protein digestion and/or enrichment, the presence of reagents such as alkylators, reductants, denaturants, and buffer salts in digested samples can lead to column contamination, ion suppression, and/or fouling of the MS source. Typically, cleanup steps, such as SPE or affinity chromatography, are employed to remove these contaminants, however these methods must be compatible with palmitoylated peptides. Using SPE filters presents a similar problem to RP chromatography in that it can potentially retain hydrophobic peptides. In contrast to RP chromatography, SPE columns are not commonly conditioned to prevent strong adsorption of the peptides to the resin and requires additional sample transfers between fresh vials, risking further loss of low abundant species to vial adsorption. Additionally, any surfactant must be removed before SPE clean up risking precipitation of the more insoluble peptides in aqueous buffers. Furthermore, affinity enrichments targeting palmitoylated peptides are challenging because palmitate lacks functional groups, which typically promote antigenicity for immunocapture methods.

These challenges may contribute to the poor LC-MS detection sensitivity of low abundance peptides, such as palmitoylated peptides, or introduce bias in quantitative measurements towards more compatible peptides. In our studies, we opted to eschew these steps and instead employ an inline sample cleanup with a pre-column filter, or trap column, within our

LC setup. Using an isocratic wash step after LC injection, we can concentrate the digested peptides on the trap column and divert the contaminated sample buffer to waste, avoiding the analytical column and mass spectrometer entirely.

Another area of concern is the removal of surfactant, used to solubilize hydrophobic peptides in digestion buffer, prior to LC injection which may result in aggregation or precipitation of insoluble peptides. Such removal is a problem as precipitation may result in loss of acylated peptides before they can be analyzed, and it could also result in clogging of the LC plumbing, especially at higher concentrations. To circumvent this issue, samples were diluted into mixed organic-phase buffers containing 35% acetonitrile/ water prior to the surfactant degradation in 1% TFA. Furthermore, TFA is known to be difficult to remove from RP columns, requiring long wash cycles and contributing to ion suppression. Therefore, it is especially important that an inline desalting step is used before introduction of the sample to the analytical RP column.

In this chapter, we identify and address key steps in LC-MS 'bottom up' workflows that may contribute to loss of palmitoylated species and develop a novel workflow for directly detecting and annotating palmitoylated peptides in complex mixtures. Specifically, this novel workflow is applied to neuroblastoma Ras viral oncogene homolog (N-Ras), an important therapeutic target [3,4] that is palmitoylated proximal to the C-terminal cysteinyl site of farnesylation [4,26] (Figure 2.1). No direct detection of a palmitoylated N-Ras peptide has been previously reported, presumably due to the inherently challenging characteristics of this non-antigenic and greasy peptide [22,23].

2.2 Experimental

2.2.1 Preparation of Palmitoylated Peptides.

Three custom peptides were purchased from ABM (>90% purity) with the following

sequences: PDFRIAFQELLCLR, MGCVQCKDKEA, ARAWCQVAQKF, based on a previous

study [20]. Each peptide standard (200 µg) was reacted with 1 µL of palmitoyl chloride (Sigma

Aldrich) in 10 µL of 100% TFA (Sigma Aldrich) for 10 min at room temperature. The resulting

mixture was dried under a nitrogen flow. The sample was suspended in 400 µL of 30%

acetonitrile (singly palmitoylated peptides) or 40% acetonitrile (doubly palmitoylated peptides),

sonicated for 1 min, centrifuged for 10 min at 13000 RCF. Twice more the supernatant was

collected and centrifuged for 10 min and transferred to another vial. Palmitoylation was validated

by mass analysis with an Agilent 6500 series QTOF mass spectrometer. Aliquots were frozen at

−80 °C for later use.

2.2.2 Synthesis of Acid Labile Surfactant
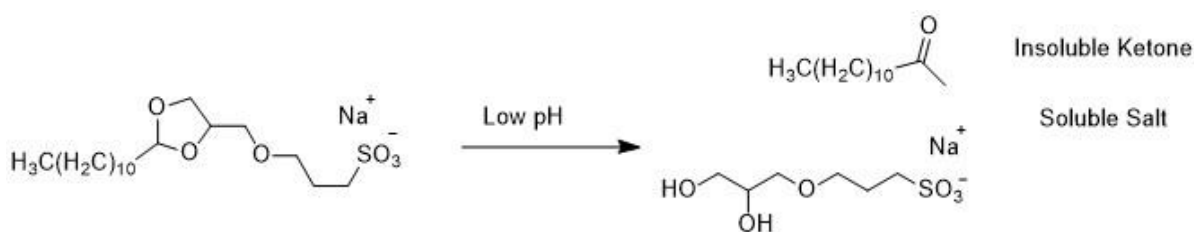
Figure 2.2 RapiGest SF (Waters) is an acid-labile surfactant that hydrolyzes into 2-tridecanone and an LC-MS compatible salt.

Synthesis of (2-methyl-2-undecyl-1,3-dioxolan-4-yl)methanol

To a dried 50 mL round bottom flask containing 35 mL anhydrous toluene and a stir bar

was added 2-tridecanone (1.0 eq, 5.0 g, 25.2 mmol), glycerol (2 eq, 3.7 mL, 50.4 mmol), and *p*-

toluenesulfonic acid (0.02 eq., 96 mg, 0.50 mmol). The flask containing the reaction mixture was

attached to a dean stark apparatus and was heated to reflux at 120 °C for 4.5 h. Every hour, 10 mL of water/toluene mixture was removed from the collection tube of the dean-stark apparatus and then 10 mL of anhydrous toluene was added back to the reaction flask. The reaction mixture was washed with (2 x 15 mL) of 5%(aq) sodium bicarbonate and (3 x 15 mL) of distilled water. The organic layer was then dried over $MgSO_4$ and concentrated under vacuum affording a clear yellow-tinged oil. The oil was purified by flash chromatography on a Telodyne CombiFlash over a silica column with a gradient of ethyl acetate and hexanes. Solvent was removed under reduced pressure affording a clear white oil. Identity of the product was confirmed by HNMR and CNMR in $CDCl_3$.

Synthesis of 2-((2-methyl-2-undecyl-1,3-dioxolan-4-yl)methoxy)ethane-1-sulfonate

To an oven dried flask equipped with a stir bar was added sodium hydride (88.1 mg, 2.20 mmol, 1.2 eq) in anhydrous THF (2.0 mL). Then (2-methyl-2-undecyl-1,3-dioxolan-4-yl) methanol (500 mg, 1.84 mmol, 1.0 eq) was added to the solution and stirred under an argon atmosphere for 15 min at room temperature. To the stirring solution, 1,3 propane sultone (161 μL, 1.84 mmol 1.0 eq) was added slowly. The solution was reacted at room temperature for 20 h. The reaction progress was monitored by TLC. The reaction was quenched by the slow addition of methanol to the stirring mixture. The resulting suspension was then evaporated under reduced pressure. The resulting solid residue was then resuspended in ethanol, heated to reflux and cooled to room temperature and a white precipitate formed. The suspension was centrifuged at 3000 rpm in a conical vial for 5 min and then the supernatant was removed, and the wash was repeated once more. The pelleted solid was dissolved in dichloromethane and then dried under vacuum. The dried residue was then dissolved in distilled water and again centrifuged at 5000 rpm for 5 min. The aqueous layer was removed, and the solution was lyophilized. Identity of the product was confirmed by HNMR and CNMR in D_2O.

2.2.3 pH Stability Test of Palmitoylated Peptides

Into separate Eppendorf tubes containing 400 µL of 50% acetonitrile (aq.) was added 400 µg of each palmitoylated peptide, referenced and prepared as described, and vortexed until dissolved. Then 5 µL of each palmitoylated peptide solution was transferred from these stocks to each of 15 vials (3x5 vials) each containing 15 µL acetonitrile and 75 µL of 100 mM PBS where the pH was adjusted to pH 5.0, 6.0, 7.0, 8.0, 9.0 respectively. The samples were vortexed and then allowed to incubate for 60 min before 5 µL of formic acid was added to each vial. Then 5uL of each stock solution was added to a vial containing 15% acetonitrile in 100 mM PBS (5% formic acid) as controls. Control groups were prepared for each peptide by adding 5 µL of each palmitoylated stock solution to vials containing 15 µL of acetonitrile and 75 µL of 100 mM PBS and acidified with 1 µL formic acid. Then 10 µL from each vial was then injected onto a Waters 5x100 mm CSH C-18 3.5 µm UPLC column and analyzed by an Agilent 6500 series QTOF mass spectrometer. Extracted ion chromatograms were analyzed for changes in peak area corresponding to the palmitoylated peptides compared to control groups.

2.2.4 Chromatographic separations and MS/MS analyses

All chromatographic separations were conducted with a Thermo UltiMate3000 Nano LC pump with house-packed 75 µm x 10 cm 3.5 µm C-8 or C-18 nano capillary columns and an IDEX C-8 or C-18 2x20 mm 7 µm trap column. Conditions were flow rate: 0.300 µL/min, temperature: 60 ℃, Solvent A: water (0.1% formic acid), Solvent B: acetonitrile (aq,0.1% formic acid).

Mass analyses were conducted with a Thermo Orbitrap Fusion Lumos mass spectrometer. MS[1] conditions were ESI voltage: 1780 V, vaporizer temperature: 150 ℃, ion transfer tube temperature: 300 ℃, sheath gas: 10, Aux gas: 1, sweep gas: 0, Orbitrap precursor selection: 150

k resolution, scan range: 200-2000, max inject time: 50 ms, AGC target: 4×10^5, RF lens: 30%, dynamic exclusion: 60 s, mass tolerance: 10 ppm. MS^2 conditions were: detector: Orbitrap 60k resolution, activation: HCD (17, 24, 30), scan rate: rapid, max injection time: 35 ms, isolation window: 1.4 Da. Database searches were conducted with Thermo Proteome Discoverer against a customized protein database including the synthetic peptide sequences or downloaded from UniProt.org. Search parameters were a mass tolerance of 10 ppm for precursors and fragments and fully specific peptide termini with ≤ 3 missed cleavages. Variable modifications included methionine oxidation and palmitoylation and/or carbamidomethylation depending on the experimental conditions.

2.2.5 Evaluation of Chromatographic Conditions for Palmitoylated Synthetic Peptides

For evaluation purposes, the three synthetic peptides mentioned above were chemically palmitoylated as previously described, combined in 1:1:1 ratio, and diluted to 1 µg/µL in 35% acetonitrile/water (0.1% formic acid). The samples were cooled to 5 °C before 1 µg was injected onto the column. The gradient for testing column stationary phase was 5-90% B over 60 min and the gradient for testing separation temperatures was 25-90% B over 60 min. Extracted ion chromatograms were obtained for each peptide with Thermo Xcalibur.

2.2.6 Evaluation of MS^2 Fragmentation Conditions for Synthetic Palmitoylated Peptides

For evaluation purposes, the three synthetic peptides mentioned above were chemically palmitoylated as previously described and diluted to 1 µg/µL in 35% acetonitrile/water (0.1% formic acid). Samples were directly injected for MS analysis and $[M+3H]^{3+}$ precursor ions were selected for MS^2 fragmentation under CID, HCD, and ETD conditions.

2.2.7 Evaluation of Chromatographic Conditions for Palmitoylated Bovine Serum Albumin Digest

Bovine serum albumin (40 µg, Fisher Scientific) was dissolved into 4 µL of 50 mM Tris Buffer (pH 7.6) containing 2 mM TCEP (Sigma Aldrich). The sample was incubated at 60 °C for 1 h before 8 µL of 50 mM Tris Buffer (pH 7.6) containing 6 M Urea (Fisher Scientific) was added. The sample was heated to 90 °C for 10 min, removed from the heat, and 28 µL of 50 mM Tris Buffer (pH 7.6) was added. The sample was vortexed, and 2 µg of trypsin (Promega) was added before incubation on an orbital shaker at 37 °C for 18 h and desalting on an Oasis HLB Elution solid phase extraction plate according to manufacturer's protocol. The digest mixture was then aliquoted into two equal portions and dried on a Speed-Vac evaporator at 40 °C. One aliquot was reconstituted in 10 µL trifluoroacetic acid (Sigma Aldrich) and then palmitoylated following the previously described protocol. The other aliquot was reconstituted in 10 µL of 50 mM Tris buffer, pH 7.4 and treated with 20 mM iodoacetamide for 1h at 37 °C. The sample was then acidified with 1 µL of 1% formic acid. 10 µg of protein from each sample was combined and diluted to a final concentration of 40 ng/µL in 35% acetonitrile/ water (0.1% formic acid). 100 ng of protein mixture was injected for each replicate.

2.2.8 Site-Specific Annotation of Palmitoylated Recombinant N-Ras by LC-MS/MS

Overexpression and enrichment of recombinant N-Ras

Plasmids expressing N-terminally His-tagged N-Ras with a TEV-linker were generated via restriction digest and ligation into p6xHis, pMBP and pGB1 (Protein Core in the Center for Structural Biology, Life Sciences Institute, University of Michigan). The human N-Ras gene was ligated to the plasmids using SspI and BamHI restriction sites and amplified before transformation into BL21 (DE3) *E.coli* electrocompetent cells. Colonies were selected on (-)His

agar plates and BL21 (DE3) *E. coli* cultures were grown at 37 ˚C to OD600 = 0.8 and the temperature was reduced to 18 ˚C for 1 h before induction with 0.4 mM IPTG for 16 hours at 18 ˚C. Cell pellets were resuspended in 50 mM HEPES buffer containing 300 mM NaCl, 10% glycerol, 2 mM 2-mercaptoethanol, 0.1% RapiGest, and 0.005% IGEPAL. The cells were lysed using a microfluidizer and centrifuged at 16000 rpm for 30 min. Protein concentration was measured using DC protein assay (Biorad). The cleared supernatant was incubated with Talon cobalt affinity beads (Clontech) for 1 h, washed with 50 mM HEPES, 150 mM NaCl , and then incubated for 16 hours with TEV-protease (1:100 w/w, Sigma Aldrich) at 4 °C. The proteins were eluted using 50 mM Tris pH 7.0 containing 0.2% Rapigest, and protein concentrations were measured using DC protein assay (Biorad).

Digestion and LC-MS Analysis

Thawed samples were incubated with 2 mM TCEP at 37 °C for 30 min and diluted 2-fold in 50 mM Tris pH 7.0. The protein mixture was digested for 4 h with trypsin (1:20; Promega) at 37 °C. Samples were concentrated with a Thermo Savant SpeedVac and then reconstituted in 35% acetonitrile/water containing 1% trifluoracetic acid (v/v) and incubated for another 30 min at 37 °C. Precipitated surfactant was pelleted by centrifugation at 13000 rcf for 10 min and the supernatants were transferred to an LC-MS vial for analysis.

Protein samples were injected (10 µL) with a Thermo UltiMate 3000 NanoLC pump and separated over a house-packed nanocapillary column (C-8 or C-18, 75 µm x 15 cm; 3 µm particles, 100 Å pore density) and a matching stationary phase 2 cm trap column (IDEX) with a 500 nL/min flow rate at 55 °C. The mobile phase gradient consisting of Buffer A: water (0.1% formic acid) and Buffer B: acetonitrile (0.1% formic acid) was initially held at 1% Buffer B for 5 min with flow diverted from the trap column to waste. At 5 min, flow was diverted to the

analytical column and the gradient was brought to 35% Buffer B (5-10min), then gradually raised to 75% Buffer B (10-65 min), and then raised to 95% and held there (65-75 min). MS analysis and a database search was performed as stated above with the exception of using HCD step-fragmentation (17, 24, 30%) for MS^2 analysis.

2.3 Results and Discussion

Previous work by Ji *et al.* [18] identified three key steps that contributed to the failure of direct palmitoylation detection for three synthetic peptides. These include sample processing steps that contribute to S-palmitoyl hydrolysis, LC conditions that lead to strong retention and associated poor resolution of the acylated peptides, and MS^2 fragmentation conditions that lead to neutral loss of the palmitoyl group. We sought to build upon this work by addressing each sample processing condition that may lead to loss of this labile group, optimizing those conditions to increase the probability of detection by LC-MS, and apply this newly developed method for the novel direct detection and site-specific annotation of palmitoylated N-Ras.

For evaluating each sample processing step, we used the same three peptide standards (Peptides 1-3, Figure 2.3) developed by Ji *et al.* based on non-tryptic sequences of known palmitoylation sites in native proteins. These non-tryptic peptides were chosen because they afford good signal during positive ion mode ESI-MS analysis due to having multiple proton carrying amino acids. These peptides also model different types of peptide acylation with one or two sites available for *in situ* chemical palmitoylation. Peptide 3 contains a tryptophan residue, which is electrophilically labeled by palmitoyl chloride, forming an irreversible and stable acyl ketone on the indole ring, thus modeling doubly acylated peptides with a single reversible palmitoylation and a second more stable fatty acid, i.e., prenylation or myristylation.

Peptide 1- PDFRIAFQELL**C**LR

Peptide 2- MG**C**VQ**C**KDKEA

Peptide 3- ARA**WC**QVAQKF

Figure 2.3 Three synthetic peptides standards each contain one or two sites (highlighted in yellow) for chemical palmitoylation and can be used in evaluating S-palmitoylation stability throughout the LC-MS workflow.

2.3.1 Cell Lysis and Protein Extraction

We began by evaluating each step in an LC-MS/MS workflow for conditions that could lead to loss of palmitoylated peptides. Generally, the first step of a typical LC-MS/MS-based proteomic workflow (Figure 1.4) is the lysis of cells of interest followed by extraction of the host proteome. Since S-palmitoylated proteins are hydrophobic and most likely membrane-bound (Figure 1.6), they are largely contained within the insoluble lipid layers of a cell lysate and may exhibit poor solubility in aqueous lysis buffers. In targeted studies of membrane proteins, hydrophobic proteins are solubilized into extraction buffers with the use of surfactants, such as sodium dodecyl sulfate (SDS). These surfactants are important for not only the initial extraction but also the continued solubilization of these hydrophobic proteins and peptides throughout the workflow. However, there are caveats to using surfactants in LC-MS based analysis, such as interference with trypsin digestion, contributing to ion suppression, and adversely impacting chromatographic retention, resolution, and peak shape. For example, surfactants can disrupt the interactions between a peptide and a column's stationary phase leading to poor column retention and peak broadening, thereby decreasing chromatographic resolution, and increasing the likelihood of peptides co-eluting. Peak broadening can also lower the maximum analyte signal intensity following ionization, which can reduce the probability of detecting low abundance

peptides. Additionally, ionizable surfactants, such as SDS, can be incompatible with MS analysis due to ion suppression and adduct formation. These factors necessitate steps to remove surfactant from the sample prior to LC-MS analysis resulting in lower throughput and unpredictable protein losses, with implications for downstream detection and quantitation.

LC-MS compatible surfactants offer a viable alternative and have been shown to effectively improve protein digestion efficiency in proteomic workflows [19-21] .The analogue, 2-((2-methyl-2-undecyl-1,3-dioxolan-4-yl)methoxy)ethane-1-sulfonate, or RapiGest SF (Waters) has surfactant, denaturant, and electrophoretic properties comparable to SDS. This detergent is acid-labile, and upon acidification hydrolyzes into 2-tridecanone, an easily removed precipitate, and an LC-MS compatible salt (Figure 2.2). This degradation permits direct MS analysis without further sample processing resulting in minimal sample loss or introduced bias for quantitative measurements.

Peptide—SH $\xrightarrow[\substack{\text{TFA, 10min} \\ >99\% \text{ yield}}]{\text{Palmitoyl chloride}}$ Peptide—S $\overset{\displaystyle O}{\underset{\displaystyle }{\|}}$ (CH$_2$)$_{14}$CH$_3$
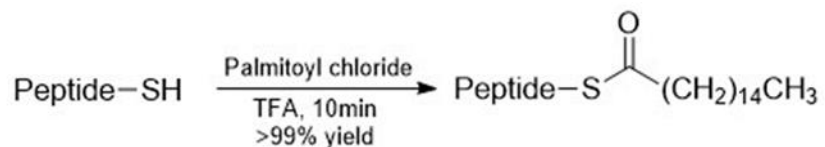
Figure 2.4 Chemical palmitoylation scheme for (S-) and (W-) palmitoylation.

RapiGest SF provides a facile solution for extracting and solubilizing S-palmitoylated proteins and peptides throughout the workflow. It is a proprietary compound that can be purchased through vendors, and it is optimally used at concentrations similar to SDS. However, at approximately $50 per milligram, the cost of the detergent limits its practicality in laboratories where overhead costs are a concern. To lower the per unit cost, we synthesized our own in-house 2-((2-methyl-2-undecyl-1,3-dioxolan-4-yl)methoxy)ethane-1-sulfonate). The resulting product

can be made in a high-yield, two-step synthesis (Fig. 2.2) at a cost of approximately $50 per gram, about three orders of magnitude more cost efficient.

2.3.2 Evaluation of S-palmitoylation pH Stability

Under basic pH conditions, thioester bonds are susceptible to nucleophilic hydrolysis from hydroxide ions present in aqueous buffers [29]. Most 'bottom up' workflows use trypsin to digest proteins into their constituent peptides. Trypsin digestions are optimally carried out between pH 7.6-8.0 for at least 4 h, and it is highly likely that hydrolysis of palmitoylated residues in these digests is substantial. Ji *et al* previously reported this behavior for the three peptide standards, and we sought to validate these results by testing the pH stability of our chemically modified peptides over a similar range of buffers (Figure 2.5) and quantify the loss of palmitate over the course of 4 hours.
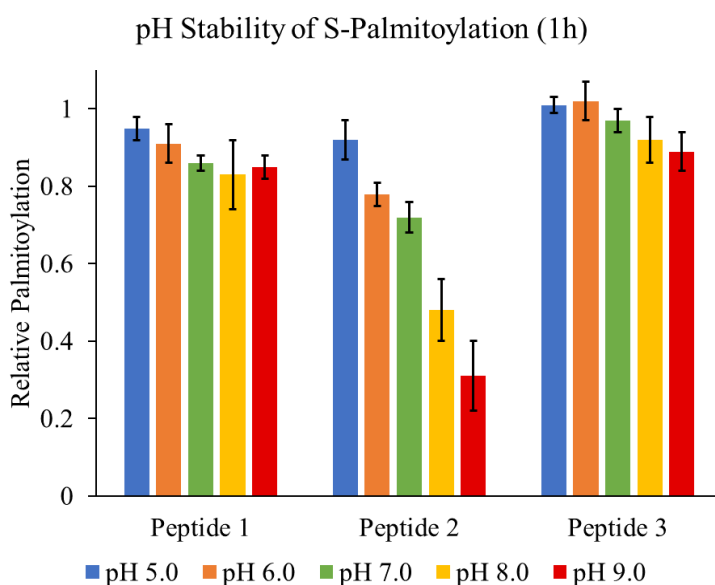
Figure 2.5 Relative retention of S-palmitoylation on three synthetic peptides PDFRIAFQELLCLR, MGCVQCKDKEA, ARAWCQVAQKF (Peptides 1-3 respectively) in 100 mM PBS buffer over 1h. *n=3*

Palmitoylation stability was measured by incubating palmitoylated peptides 1-3 (Figure 2.5), each in 100 mM PBS buffer at pH 5.0-9.0, for one hour before quenching with formic acid. Quantitative analysis was performed by LC-MS on an Agilent QTOF mass spectrometer. Peak area ratios integrated from extracted ion chromatograms (EICs) were normalized to the respective control groups that were incubated in formic acid buffer. Our results match those by Ji *et al* and indicate that thioesterified palmitate is sensitive to basic buffer conditions, and the trends in all three cases indicate that higher pH values tend to increase the rate of palmitate hydrolysis. Interestingly, Peptides 1 and 3 were least affected by the higher pH conditions retaining at least 85% of their palmitate labels over the course of four hours at room temperature and pH 9.0. Interestingly, the behavior of each peptide in these buffers varied. Peptide 1 lost approximately 15% of the palmitate label at pH 7.0 while Peptide 3 retained greater than 90% of palmitoylation. The data also indicate that Peptide 2, a doubly S-palmitoylated peptide, was more sensitive to high pH conditions and only retained 30% of total palmitoylation at pH 9.0. It is important to note that the steepest change in palmitate retention occurred between pH 7.0-8.0 where the total label retention dropped from approximately 78 – 45% respectively.

We hypothesize that the difference in the observed pH sensitivities may be due to the neighboring amino acids present in each peptide. Peptide 2, the most labile peptide, contains two lysines which have approximate pKa values of 10. Therefore, at pH 8-9 greater than 90% of the lysines will remain protonated in solution. This protonation potentially introduces hydrogen bonding interactions between the alkyl oxygen of the thioester and the terminal ammonium of the lysine, thus increasing the electrophilicity of the alkyl carbon center and the kinetic rate of palmitate hydrolysis. Another possibility is that electrostatic interactions between positively charged ammonium of the lysine and free hydroxide may also increase the rate of hydrolysis by concentrating the hydroxide anions near the thioester. Peptide 1, however, does not contain

lysine which may explain its relative stability compared to Peptide 2. Peptide 3 does contain a single lysine, but the steric hinderance of the adjacent acylated tryptophan may serve to protect the thioester and slow the rate of hydrolysis.

While Ji et al. also observed a similar trend in pH sensitivity between the peptides, they did not quantify their results, nor did they report such significant loss of peptide 2 palmitoylation until after six hours. These data are helpful in evaluating buffer pH conditions most suitable for palmitoylation analysis, especially during trypsin digestion. While these higher pH buffers are ideal, trypsin digests can still be carried out in neutral buffers with minimal impact on overall digestion efficiency. The manufacturer's protocol for Promega Sequencing Grade Modified Trypsin states that trypsin is maximally active between a pH range of 7- 9. Therefore, in order to retain palmitoylation throughout our workflow, going forward we carried out all sample preparation steps in pH ≤ 7 buffers, including all trypsin digests of palmitoylated proteins in neutral pH buffer.

2.3.3 Chromatographic effects of stationary phase and temperature on palmitoylated peptides

Having adapted our sample processing workflow to limit palmitoyl-loss from a whole cell lysate up to the LC separation step, we next explored LC conditions for compatibility with palmitoylated peptides. We initially hypothesized that the underrepresentation of palmitoylated peptide identifications in LC-MS/MS based proteome analyses may be primarily due to strong peptide retention on C-18 resins. Currently, reverse phase HPLC is the most common LC platform for the separation of water-soluble proteins and peptides. Typically, mobile phase gradients comprised of formic acid buffered water and a miscible organic eluent, such as acetonitrile, are used to elute solutes with increasing strength. This combination of stationary and

mobile phases can be adjusted and optimized to fit the specific needs of a study. However, C-18 columns are readily available and have proven to be reliable and robust for most proteomic analyses studying soluble peptides. In cases where membrane proteins are of interest and high sequence coverage of hydrophobic regions is desired, gradients need to be strengthened to elute the hydrophobic or lipidated peptides. Mobile phase composition can also be strengthened with the addition of stronger organic solvents, such as isopropanol, if acetonitrile alone is insufficient. However, there are drawbacks to changing mobile phase composition, and the process of changing and equilibrating mobile phases, with capillary LC especially, can be time intensive or impractical if using a shared instrument with limited availability. Additionally, strengthening the organic phase can result in poor chromatographic resolution of more hydrophilic peptides resulting in poor sequence coverage during analysis. Often, a more practical solution is to use a different column with a less retentive stationary phase, such as C-8 or C-4, or to use high separation temperatures to increase solubility.

It is reasonable to expect that multiply lipidated peptides with a high degree of nonpolar amino acid composition, such as membrane associated palmitoylated peptides, may be difficult to elute from C-18 columns. We hypothesized that the combination of a C-8 nano-LC column with a formic acid buffered water and acetonitrile mobile phase would be capable of eluting multiply lipidated peptides and still provide good resolution for both hydrophobic and hydrophilic peptides. To test this hypothesis, we compared the LC retention profiles of our three standard palmitoylated peptides on both C-8 and C-18 columns (Figure 2.6).
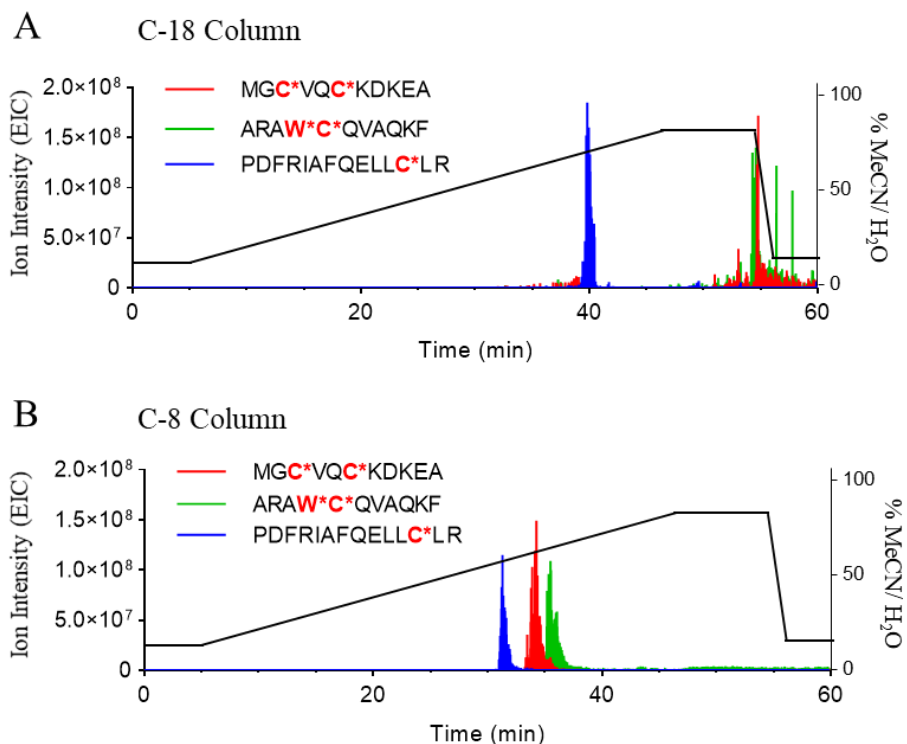
Figure 2.6 Peak retention profiles from extracted ion chromatograms (EICs) of three palmitoylated peptides. Peptides were separated at 35°C with a Thermo UltiMate 3000 Nano-LC fitted with a **A)** house-packed C-18, 75 μm x 10 cm nanocapillary column **B)** house-packed C-8, 75 μm x 10 cm nanocapillary column. Flow rate 300 nL/min.

Figure 2.6 shows the separation of a sample containing an equiproportional mixture of our three palmitoylated peptides at 35 °C with water/acetonitrile on a 10 cm nanocapillary column packed with either C-18 or C-8 resin. The graphs illustrate the EICs of the $[M+2H]^{2+}$ ions of the palmiotoylated peptides after separation. Figure 2.6(a) indicates that the peptide containing a single palmitoyl group eluted in approximately 80% acetonitrile. However, both of the doubly palmitoylated peptides failed to elute on the C-18 resin until the gradient was held at 95% acetonitrile for 10 min at which point they co-eluted with poor signal-to-noise ratios. A subsequent injection of a blank containing 50% aqueous methanol reulted in carry-over of these peptides into the next LC run, indicating that some fraction of these peptides were retained on the column. This result helped confirm our intial hypothesis that C-18 columns may not be an

optimal choice for separation of lipidated peptides, especially peptides with more than one lipid group.

We then evaluated the effects of separating our peptides on a C-8 column under the same conditions. Figure 2.6(b) illustrates a significantly improved retention profile with all three peptides eluting between 50-65% acetonitrile with good resolution and peak shapes. A subsequent injection of a blank containing 50% aqueous methanol showed no carry-over of these peptides, indicating that all three peptides were quantitatively eluted from the column. The latter behavior is important when quantitative measurements of palmitoylated peptides are desired. Overall, we determined that use of C-18 chromatography with an acetonitrile mobile phase potentially leads to loss of palmitoylated peptides in LC-MS analyses. Additionally, we determined that a C-8 resin is a better choice for separating lipidated peptides under these conditions.

Gradient optimization is an imporant consideration in any LC separation, and due to the high concentration of organic phase required to elute our lipidated peptides coupled with the close retention times, we increased the organic phase concentration throughout the gradient to hopefully achieve better resolution. Therefore, the gradient was adjusted to a 15% acetonitrile intial isocratic hold and quickly increased to 35% in the first few minutes followed by more gradual increase to 70% over 45 min. We expected these changes to improve the performance of the separation and increase resolution between the lipidated peptides. Additionally, temperature is a another conditon that affects chromatography. Higher column temperatures can be utilized in LC separations to decrease peptide retention times, potentially increase peptide resolution, and contribute to sharper peak profiles. Using our adjusted gradient, we explored the effects of increasing column temperature from 35 $^\circ$C to 55 $^\circ$C for our palmitoylated peptides (Figure 2.7).
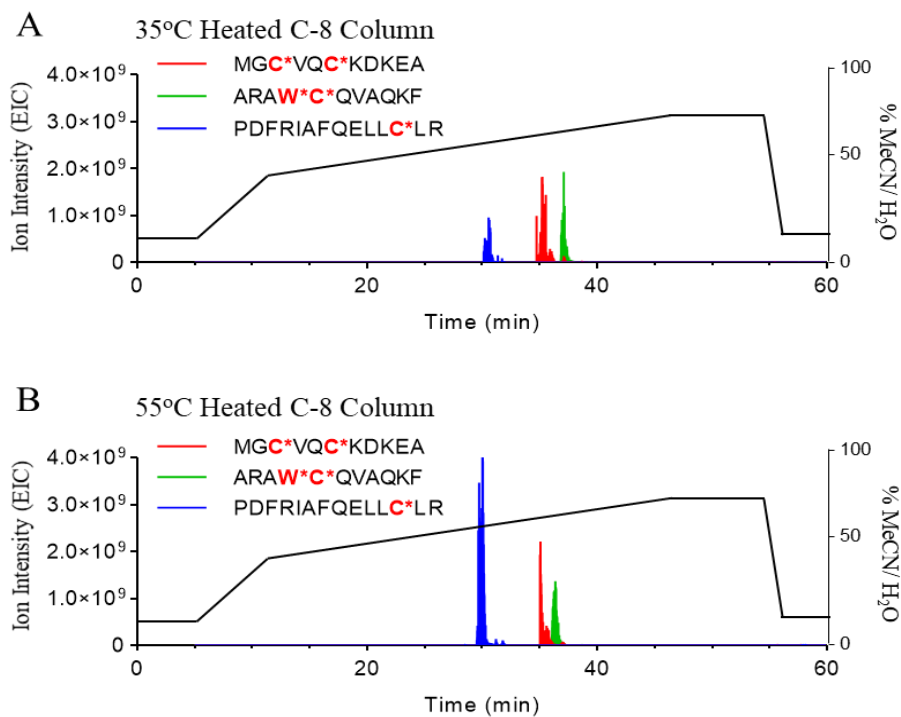
Figure 2.7 Peak retention profiles from extracted ion chromatograms (EICs) of three palmitoylated peptides. Peptides were separated on house-packed C-8, 75 μm x 10 cm nanocapillary column. with a Thermo UltiMate 3000 Nano-LC heated to **A)** 35°C, or **B)** 55°C. Flow rate 300 nL/min.

Comparing results between experiments (Figures 2.6(b) and 2.7(a)), we evaluted the effects of our gradient optimization on the retention times of the three lipidated peptides. As expected, the optimized gradient provided improved separation with sharper chromatographic peaks. Importantly, the increased resolution minimized any co-elution between the peptides. Comparing Figures 2.7 (a-b), the temperature increase from 35 ^{o}C to 55 ^{o}C improved the signal of the singly palmitoylated peptide but the other two peaks reamined almost unchanged. Interestingly, there was an approximate 2 min shift in the retention time of the singly lipidated peptide, in blue, but the doubly lipidated peptides seemed unaffected. The increased signal for the singly palmitoylated peptide may be a result of a temperature dependent unfolding of the peptide structure leading to improved ionization. Further work needs to be performed to explain this phenonmenon. Overall, we determined that our optimized gradient offers significant

improvement to our peak resolution while heated chromatography offers, albeit minor, improvements in resolution profiles.

2.3.4 Evaluating and optimizing MS/MS fragmentation for S-palmitoylated peptides

Sequencing a peptide with site-specific resolution of PTMs is achieved through fragmenting the peptide with MS/MS. A peptide precursor ion is selected in MS^1 and then fragmented with some form of dissociative energy introduction resulting in cleavages of the chemical bonds making up the peptide, and in ideal circumstances, these cleavages occur between the amino acids in the sequence. The subsequent MS^2 spectrum is then comprised of these ion fragments which can be annotated to sequence the peptide. Ideally, a cleavage is introduced between each amino acid in the sequence allowing for complete annotation of the peptide. However, the energy levels used in fragmentation need to be tuned to prevent cleaving bonds with low activation energy barriers, such as labile PTMs. If the PTM is cleaved without retaining a charge, then the MS^2 scan will not detect the PTM and is considered a neutral loss event. This is potentially a major hurdle in the MS analysis of palmitoylation, and Ji *et al.* [18] reported that palmitoylated peptides were susceptible to neutral loss of the palmitate under multiple types of fragmentation methods, including collision induced dissociation (CID), higher-energy collision induced dissociation (HCD), and electron-capture dissociation (ECD) but showed less neutral loss with electron transfer dissociation (ETD). While this phenomenon was well-characterized in the study, we sought to evaluate CID, HCD, and ETD on a different instrument platform, the Orbitrap Fusion Lumos to select the MS/MS method that provides the best combination of good sequence coverage and the least neutral palmitate loss.
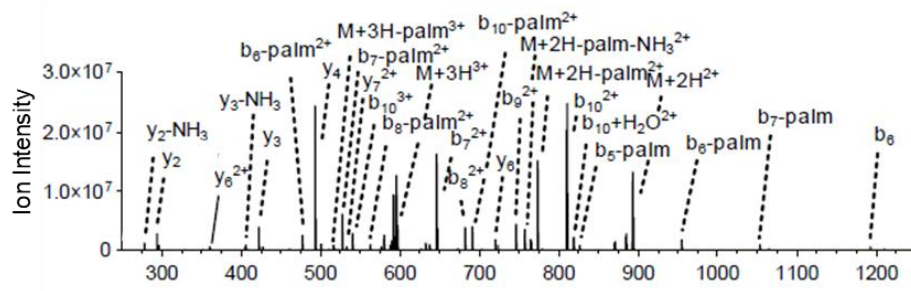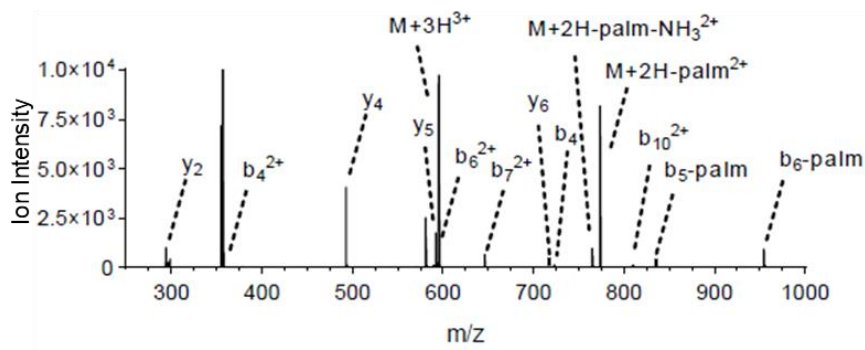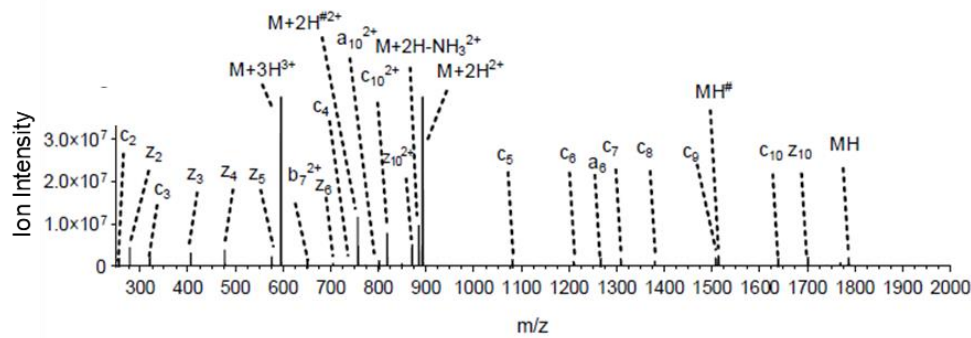
Figure 2.8 MS/MS spectra and annotations of MS^2 fragmentations of [M+3H]$^{3+}$ precursor ion ARAW$^{(+238.22)}$ C$^{(+238.22)}$QVAQKF. **A)** CID (30%) **B)** HCD (20%) **C)** ETD (15 ms). [x -palm] denotes neutral loss of palmitate [m -238.22].

* = palmitoylated amino acid
# = neutral loss of palmitate [m -238.22] in CID and HCD or [m -271.21] in ETD

Figure 2.8 compares the MS/MS spectra of the peptide ARAW$^{(palm)}$C$^{(palm)}$QVAQKF from CID, HCD and ETD. CID fragmentation (Fig. 2.8(a)) showed significant neutral palmitate (palm) loss, however HCD (Fig. 2.8(b)) resulted in moderate neutral loss. The fragmentation profiles observed from both methods provided sufficient sequence coverage for peptide identification, however, only HCD provided site-specific annotation of the palmitoylation sites. Similar differences between CID and HCD have been observed for other labile PTMs, e.g., fucosylation [31]. However, complete coverage of all amino acids in the peptide was not observed. Figure 2.8(c) shows the ETD MS/MS spectrum for the same peptide. From ETD, complete coverage of the peptide was achieved with site-specific resolution allowing for the identification of the palmitoylated amino acids. Similar to the results by Ji et. al., almost no neutral loss was observed with only two peaks corresponding to the loss of palmitate from the precursor ions. Unlike CID or HCD, which tend to fragment the thioester bond of the palmitate, ETD tends to fragment the bond between the β-carbon and the sulfur atom of a palmitoylated cysteine producing a slightly different neutral loss signature. Overall, we observed ETD outperforming both CID and HCD in retaining the palmitate and in providing complete peptide coverage. However, there are some important drawbacks with ETD, which involves introduction of a radical anion and transfer of an electron from the radical carrier to a multiply positively charged peptide inducing fragmentation of the charge reduced peptide ion. It has been shown that peptides best suited for this type of fragmentation are highly charged [30] and/or have relatively low m/z ratio (<1,000). While our chemically palmitoylated peptides fall within this range, larger scale studies of lipidated peptides using tryptic digests may not. S-palmitoylated peptides from trypsin digests tend to have high m/z ratios due to containing one or more large acyl chains, and due to originating from membrane associating regions lacking polar amino acids.
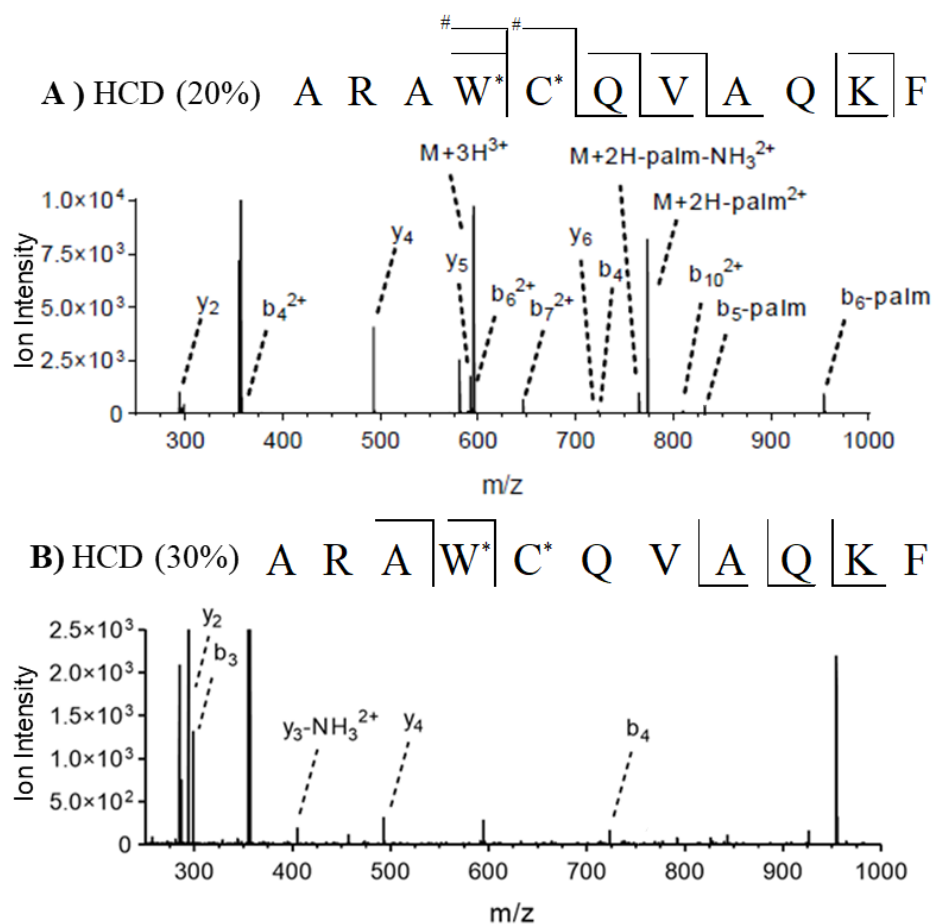
Figure 2.9 MS/MS spectra and annotations of MS^2 fragmentations of [M+3H]$^{3+}$ precursor ion ARAW$^{(+238.22)}$C$^{(+238.22)}$QVAQKF with HCD collisional activation. The effect of collisional energies were compared **A)** 20% **B)** 30%. [x -palm] denotes neutral loss of palmitate [m -238.22].

* = palmitoylated amino acid
# = neutral loss of palmitate [m -238.22]

A) HCD (20%)   P|D|F| R I A|F| Q|E|L|L|C*|L R
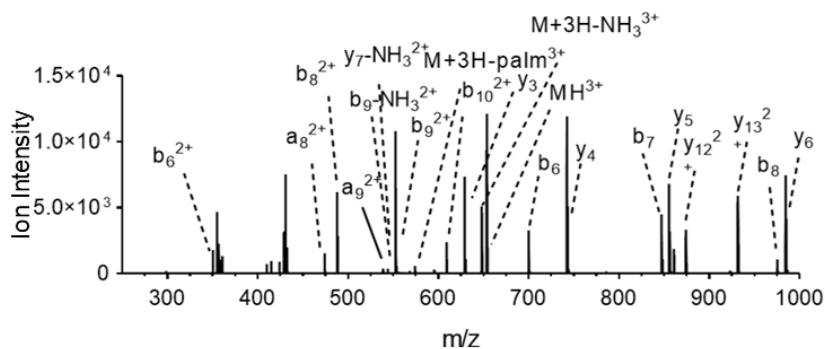
B) HCD (30%)   P|D|F|R| I | A| F Q|E|L|L|C*|L R
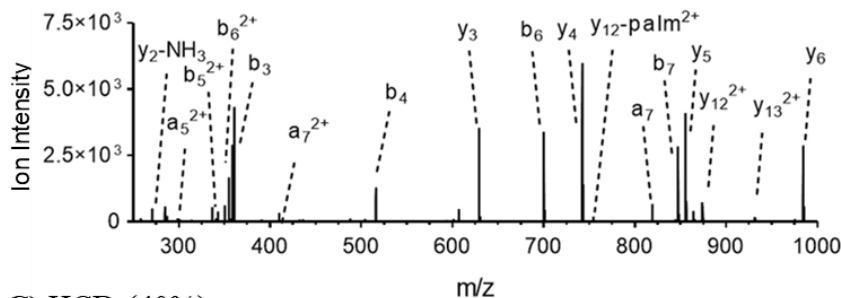
C) HCD (40%)
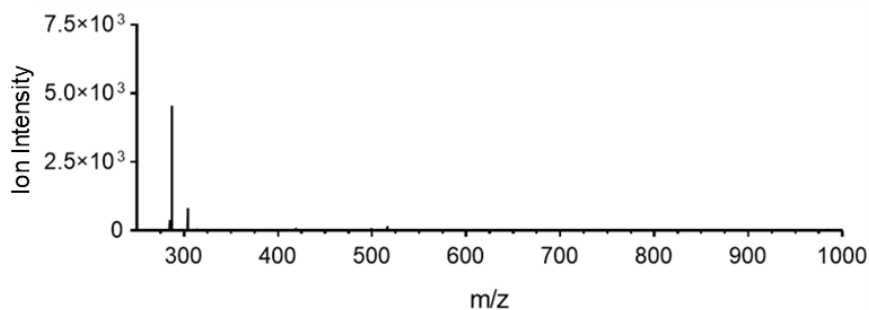
Figure 2.10 MS/MS spectra and annotations of MS^2 fragmentations of [M+3H]$^{3+}$ precursor ion PDFRIAFQELLC$^{(+238.22)}$LR with HCD collisional activation. The effect of collisional energies were compared **A)** 20% **B)** 30% **C)** 40%. [x -palm] denotes neutral loss of palmitate [m -238.22].

* = palmitoylated amino acid
\# = neutral loss of palmitate [m -238.22]
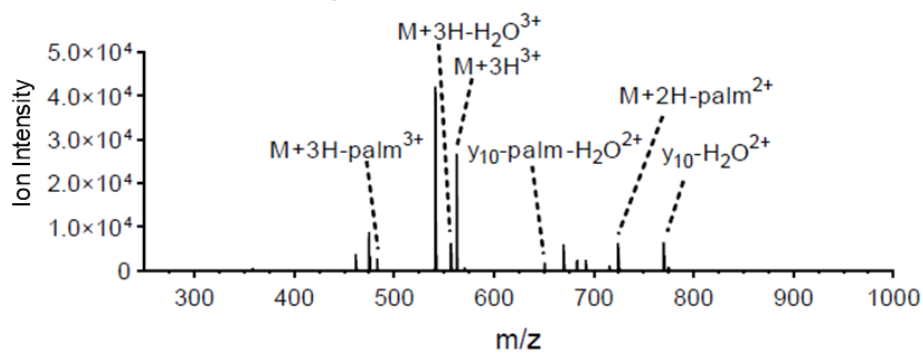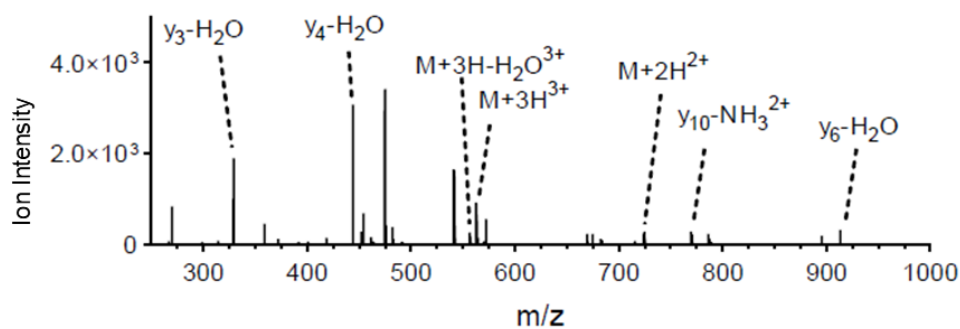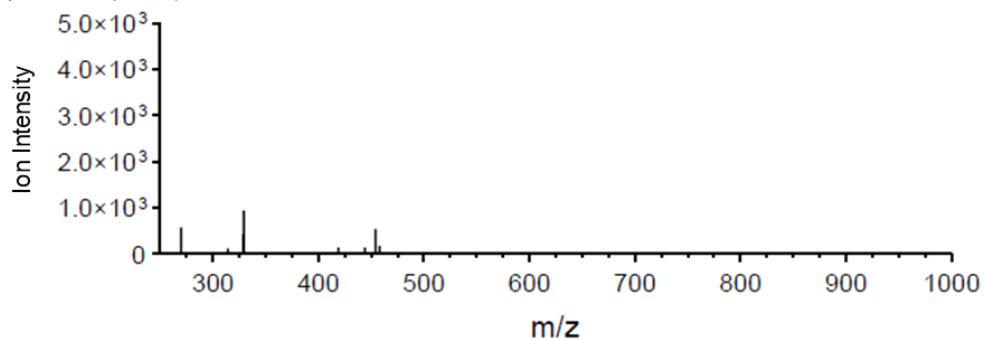
Figure 2.11 MS/MS spectra and annotations of MS^2 fragmentations of [M+3H]$^{3+}$ precursor ion MGC$^{(+238.22)}$VQC$^{(+238.22)}$KDKEA with HCD collisional activation. The effect of collisional energies were compared **A)** 20% **B)** 30% **C)** 40%. [x -palm] denotes neutral loss of palmitate [m -238.22].

* = palmitoylated amino acid
\# = neutral loss of palmitate [m -238.22]

Thus ETD may not be the best overall choice for palmitoylation analysis, and therefore, we chose to further explore HCD.

Figure 2.9 shows the MS^2 spectra from HCD fragmentation of the peptide ARAW(palm)C(palm)QVAQKF with 20 and 30% collisional activation energies. While both energy levels resulted in cleavages between the modified tryptophan and cysteine residues, 30% collision energy resulted in lower peptide coverage and did not allow unambiguous identification of the cysteine palmitoylation site. By contrast, 20% energy did allow site-specific resolution of the modified cysteine despite also yielding palmitate neutral loss. These results suggested that optimization of collision energy is important and we proceeded to evaluate the other two peptides PDFRIAFQELLC(palm)LR (Figure 2.10) and MGC(palm)VQC(palm)KDKEA (Figure 2.11) at different levels of collisional energy. Interestingly, the latter two peptides showed improved fragmentation behavior at 30% collisional energy than at 20%, unlike ARAW(palm)C(palm)QVAQKF. At 40% collisional activation, both MS^2 scans resulted in no identifiable fragment ions. Overall, we concluded that no single energy level would produce optimal fragmentation behavior for every peptide, and therefore, in more complex mixtures, HCD step-fragmentation may be ideal. In such approaches, precursor ions are fragmented sequentially at multiple energy levels. Although this instrument operation increases duty cycle, the probability of acquiring optimum MS/MS data also increases For optimum performance, this strategy may need to be coupled with sample enrichment or longer LC gradients to reduce the likelihood of co-eluting peptides.

2.3.5 Evaluating LC-MS workflow with S-Palmitoylated Bovine Serum Albumin

Having systematically evaluated and optimized our LC-MS workflow for the direct detection of three non-tryptic palmitoylated peptides, we sought to test these conditions on a

more complex mixture of chemically palmitoylated tryptic peptides. Bovine serum albumin

(BSA) contains 35 cysteines, and after digestion with trypsin affords 24 peptides, each with 1-3

cysteine sulfhydryls available for chemical palmitoylation. This model system increases sample

complexity and provides an opportunity to evaluate our conditions on a mixture of tryptic

peptides of varying palmitoylation status. BSA was alkylated with either palmitate or

iodoacetamide. The presence of carbamidomethylated peptides provides a positive control for the

analysis and allows us to evaluate the behavior of non-lipidated peptides with our

chromatography method.
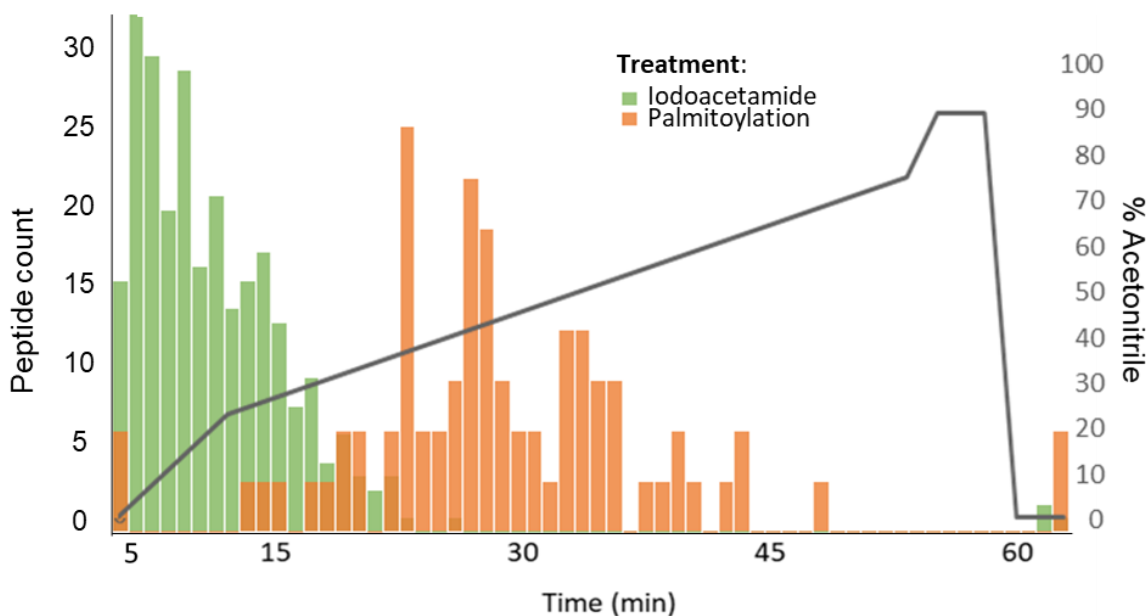
Figure 2.12 Chart shows the total count of peptide identifications in 90 s bins from a mixture of carbamidomethylated (green) and palmitoylated (orange) tryptic digest of bovine serum albumin. Peptides were separated over the course 65 min on a house-packed 15 cm C-8 nanocapillary column at 55 °C. Peptides were identified by tandem MS with HCD step-fragmentation (17, 24, 30 %).

Palmitoylated and carbamidomethylated BSA tryptic peptides were combined and

separated over a C-8 nano-LC column at 55 °C. Peptides were analyzed by tandem MS with

HCD step-fragmentation at 17, 24, and 30% collisional activation energy. The ".raw" files were subjected to a database search with Thermo Proteome Discoverer. Retention times of the identified peptides were then compared between the palmitoylated and carbamidomethylated peptides. Figure 2.12 shows the total identification count for all peptide ions in 90 second bins. As expected, the average retention time of palmitoylated peptides was longer than the carbamidomethylated peptides, and a majority of the non-lipidated peptides eluted within the first 10 min of the gradient. Interestingly, our LC conditions resulted in an effective enrichment of lipidated peptides during the separation. Such enrichment is important for complex mixtures where lipidated peptides are less abundant than non-lipidated peptides, and co-elution of abundant unmodified peptides with lipidated peptides could lead to ion suppression or lack of palmitoylated peptide precursor selection for MS/MS. Unfortunately, we only detected 14 out of 24 of the expected unique palmitoylated peptides. Further investigations failed to identify the cause of these missed identifications. It is possible that some peptides were lost during the chemical palmitoylation step due to being too insoluble for reconstitution in 35% acetonitrile. Other possibilities include the loss of these peptides during the LC separation, or their thioesters were not stable and were hydrolyzed prior to analysis. However, complete profiling of artificially palmitoylated BSA peptides was only peripheral to the focus of our study, and we did illustrate promise of our method for site-specific analysis of a mixture containing tryptic palmitoylated peptides. Interestingly, two of the cysteine-containing peptides that were missed (MPCTEDYLSLILNR and PCFSALTPDETYVPK) both contain a proline residue immediately preceding the palmitoylated cysteine and the truncated non-tryptic peptides TEDYLSLILNR and FSALTPDETYVPK were both detected. Further work is needed to determine the root cause of the existence of these non-tryptic peptides as new information may be gained about the stability of palmitoylated peptides near proline residues.

2.3.7 Direct Detection and Site-Specific Analysis of Palmitoylated N-Ras

The successful direct, site-specific annotation of 17 palmitoylated peptides, 14 from BSA and 3 from synthetic peptides spiked into the sample, encouraged us to apply our method to achieve direct detection of natively palmitoylated N-Ras. To obtain this sample, we over-expressed recombinant N-Ras containing a His-tag, extracted the membrane fraction with RapiGest, enriched with a nickel affinity column, then performed trypsin digestion at pH 7.0. The surfactant was hydrolyzed in TFA prior to LC-HCD MS/MS analysis with either a C-8 or C-18 column. The LC-MS '.raw' files were searched with Thermo Proteome Discoverer, and search results were filtered for our peptide of interest (Table 2.1). Unfortunately, Proteome Discoverer failed to annotate a fully modified N-Ras C-terminal peptide. However, two peptides of interest were identified, including a prenylated and methylated C-terminal peptide and a palmitoylated and methylated peptide (highlighted in yellow in Table 2.1).

Interestingly, a peptide was identified containing both the C-terminal methylation and S-palmitoylation but lacked prenylation. This result was unexpected because farnesylation occurs through the acylation of a cysteine sulfhydryl, not through formation of a thioester bond. Thioether bonds are significantly more stable then thioester bonds and are therefore not prone to hydrolytic cleavage, heat labile, or sensitive to reducing agents. Furthermore, the palmitoylated peptide had a retention time of 63 min, near the end of the gradient where, in previous experiments, doubly or triply lipidated species tended to elute. This behavior suggested that farnesylation may have been lost post column prior to MS[1] precursor selection, however, further work is needed to validate the existence of this peptide.

Table 2.1 Filtered search results of identified lipidated N-Ras peptides by LC HCD MS/MS

Annotated Sequence	Modifications	# Missed Cleavages	Charge (+)	m/z	RT [min]
KLNSSDDGTQGCmGLPc	M13(Oxidation); C17(Farnesyl)	1	2	973.45	24.4
LNSSDDGTQGCmGLPc	M12(Oxidation); C16(Farnesyl)	0	2	909.41	28.0
KLNSSDDGTQGCMGLPc	C17(Farnesyl)	1	2	965.45	25.2
KLNSSDDGTQGCMGLPc	C17(Farnesyl)	1	2	965.46	54.6
LNSSDDGTQGCMGLPc	C16(Farnesyl)	0	1	1801.8	31.1
LNSSDDGTQGCMGLPc	C16(Farnesyl)	0	2	901.41	30.9
KLNSSDDGTQGcMGLPc	C12(Palmitoyl); C-Term(Methyl)	1	2	989.49	63.4
KLNSSDDGTQGcMGLPc	C12(Farnesyl); C-Term(Methyl)	1	2	972.47	26.8
KLNSSDDGTQGcMGLPc	C12(Farnesyl); C-Term(Methyl)	1	2	972.47	26.8

Additionally, it should be noted that most of the identified peptides contained a missed cleavage site, resulting in an N-terminal lysine being present. This missed cleavage is important because lysines are basic and sequester positive charges under ESI (+) ionization, thus effectively increasing the total charge of the peptide ion. Additional basic residues typically results in higher signal intensity and improved HCD fragmentation behavior, which both help improve peptide annotations. This consideration is important when building inclusion lists for targeted proteomics analyses, and proteases such as Lys-N, which cleaves N-terminal to lysine, can potentially be leveraged for more sensitive detection.

In an effort to determine the root cause of Proteome Discoverer's failure to identify our targeted peptide, we manually searched the MS[1] scans for a precursor ion matching the m/z value of fully modified N-Ras C-terminal peptide. We were successful in identifying a precursor matching the theoretical mass of our targeted peptide, however, the only matching precursor ion contained the missed lysine cleavage in its sequence, similar to the peptides identified in Table

2.1. In fact, not only was the peptide ion signal abundant, but it had also been selected for fragmentation. The manually annotated resulting MS^2 scan is shown in Figure 2.13 for . matching the [M+2H]$^{2+}$ precursor ion of KLNSSDDGTQGC$^{(palm)}$MGLPC$^{(farnesyl)}$-OMe, the fully modified N-Ras C-terminal peptide.
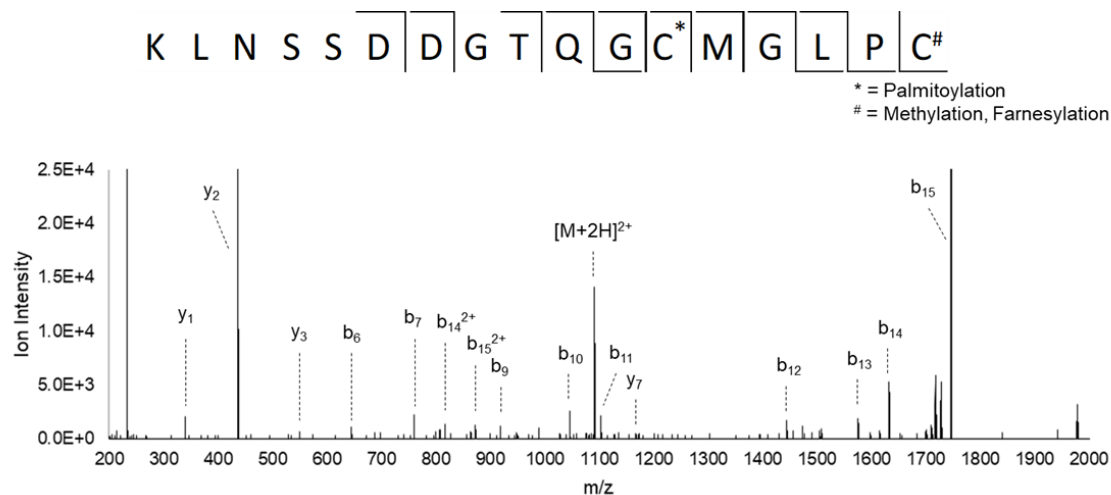
Figure 2.13 Fragmentation diagram and annotation of the MS^2 scan of a [M+2H]$^{2+}$ precursor ion of the fully modified N-Ras C-terminal peptide KLNSSDDGTQGCMGLPC with HCD collisional activation. Annotation confirms direct detection of S-palmitoylation with site-specific resolution of each modification.

2.4 Conclusion

Based on efforts to develop a method for directly detecting palmitoylation events in protein digests, we report the first direct identification of N-Ras palmitoylation by LC-HCD MS/MS with sufficient sequence coverage for site-specific resolution of each modified amino acid. While the targeted peptide ion was present at relatively high signal abundance and was selected for MS^2 fragmentation producing a high quality spectrum, searches with Proteome Discoverer using a variety of search parameters, such as filtering criteria, signal to noise thresholds, removal of false discovery rate filtering, adjusting ppm tolerances, and even the use of targeted inclusion lists did not yield an identification. The reasons underlying the failure of the software remain elusive, but manual searches of the MS^1 scans by mass range did require

tightening the ppm tolerances to 0.10 ppm from 0.50 ppm to return the target spectrum which may hold clues to the underlying cause. Future work should entail the comparison and evaluation of different search algorithms for this type of work, as it is possible that other software may not have the same challenges.

Additionally, to further test our initial hypothesis that poor chromatography underlies much of the failure to identify S-palmitoylated peptides in proteomics experiments, we injected the same N-Ras digest onto a C-18 column. Both Proteome Discoverer and manual search of the corresponding data failed to yield the palmitoylated N-Ras peptide. This result suggests that C-18 chromatography with an acetonitrile/water mobile phase gradient is not suitable for LC-MS analysis of S-palmitoylated peptides. Overall, we successfully achieved the novel direct detection and site-specific analysis of N-Ras palmitoylation by LC-MS and have built a foundation for future LC-MS strategies targeting analysis of global S-palmitoylation and other lipid PTMs.

2.4 References

1. Smotrys JE, Linder ME. Palmitoylation of intracellular signaling proteins: regulation and function. Annu Rev Biochem. 2004; 73():559-87.

2. Bizzozero, O.A., Malkoski, S. P., Mobarak, C., Bixler, H. A., Evans, J. E. , Mass-spectrometric analysis of myelin proteolipids reveals new features of this family of palmitoylated membrane proteins. J. Neurochem. , 2002. 81: p. 636−645.

3. Eisenberg, S. et al. The role of palmitoylation in regulating Ras localization and function. Biochem. Soc. Trans. 2013; 41, 79–83

4. Lin, D. T. S., Davis, N. G., Conibear, E. (2017). Targeting the Ras palmitoylation/ depalmitoylation cycle in cancer. Biochem Soc Trans, **45**(4): p. 913-921

5. Roth, A.F., Wan, J., Bailey, A.O., et al, Global analysis of protein palmitoylation in yeast. Cell. 2006;125:1003–1013. Cell, 2006. 125: p. 1003-1013.

6. Forrester MT, Hess DT, Thompson JW, Hultman R, Moseley MA, Stamler JS, Casey PJ. Site-specific analysis of protein S-acylation by resin-assisted capture. J Lipid Res. 2011; 52(2):393-8.

7. Drisdel, R.C., Green, W.N., Labeling and quantifying sites of protein palmitoylation. Biotechniques, 2004. 36: p. 276–285.

8. Martin BR, Wang C, Adibekian A, Tully SE, Cravatt BF Global profiling of dynamic protein palmitoylation. Nat Methods. 2011; 9(1):84-9.

9. Zhang MM, Tsou LK, Charron G, Raghavan AS, Hang HC. Tandem fluorescence imaging of dynamic S-acylation and protein turnover. Proc Natl Acad Sci U S A. 2010;107(19):8627-32.

10. Hang HC, Linder ME. Exploring protein lipidation with chemical biology. Chem Rev. 2011; 111(10):6341-58.

11. Kathayat, R., Elvira, P. & Dickinson, B. A fluorescent probe for cysteine depalmitoylation reveals dynamic APT signaling. Nat Chem Biol. 2017; 13, 150–152

12. E. Saxon, C.R. Bertozzi. Cell surface engineering by a modified Staudinger reaction. Science. 2000; 287. 2007-2010

13. Q. Wang, et al. Bioconjugation by copper(I)-catalyzed azide-alkyne [3 + 2] cycloaddition. J. Am. Chem. Soc. 2003;125, pp. 3192-3193

14. Raghavan AS, Hang HC. Seeing small molecules in action with bioorthogonal chemistry. Drug Discov Today. 2009;14(3-4):178-84.

15. Martin, S. E.; Shabanowitz, J.; Hunt, D. F.; Marto, J. A., Subfemtomole. MS and MS/MS peptide sequence analysis using nano-HPLC micro-ESI fourier transform ion cyclotron resonance mass spectrometry. Analytical Chemistry 2000, 72 (18), 4266-4274.

16. Ducret, A.; Oostveen, I. V.; Eng, J. K.; Yates III, J. R.; Aebersold, R.. High throughput protein characterization by automated reverse-phase chromatography/electrospray tandem mass spectrometry. Protein Science. 1998, 7 (3), 706-719.

17. Covey, T. R.; Lee, E. D.; Bruins, A. P.; Henion, J. D., Liquid chromatography/mass spectrometry. Analytical chemistry 1986, 58 (14), 1451A-1461A.

18. Fung EN, Bryan P, Kozhich A. Techniques for quantitative LC–MS/MS analysis of protein therapeutics: advances in enzyme digestion and immunocapture. Bioanalysis 2016, 8(8), 847–856

19. Becker JO, Hoofnagle AN. Replacing immunoassays with tryptic digestion-peptide immunoaffinity enrichment and LC–MS/MS. Bioanalysis 2012 4(3), 281–290

20. Ji, Y., Leymarie, N., Haeussler, D.J., Bachschmid, M.M.,Costello, C.E., Lin, C., Direct Detection of S-Palmitoylation by Mass Spectrometry. Anal. Chem., 2013. 85: p. 11952−11959.

21. Salaun, C.G., J.; Chamberlain, L. H., The intracellular dynamic of protein palmitoylation. J. Cell Biol., 2010. 191: p. 1229−1238.

22. Hernandez, J.L., Majmudar, J.D., Martin, B.R., Profiling and Inhibiting Reversible Palmitoylation. Curr Opin Chem Biol, 2013. 17(1): p. 20-26.

23. Blanc, M., David, F., Abrami, L., Migliozzi, D., Armand, F., Bürgi, J., van der Goot, F. G. , SwissPalm: Protein Palmitoylation database. F1000Research, 2015. 4(261).

24. Magee AI, Koyama AH, Malfer C, Wen D, Schlesinger MJ. Release of fatty acids from virus glycoproteins by hydroxylamine. Biochimica et Biophysica Acta (BBA) - General Subjects. 1984;798:156–166.

25. Rose JK, Adams GA, Gallione CJ. The presence of cysteine in the cytoplasmic domain of the vesicular stomatitis virus glycoprotein is required for palmitate addition. Proc Natl Acad Sci U S A. 1984;81:2050–2054.

26. Liang, X., Lu, Y., Neubert, T. A., Resh, M. D., Mass Spectrometric Analysis of GAP-43/Neuromodulin Reveals the Presence of a Variety of Fatty Acylated Species. J. Biol. Chem., 2002. 277: p. 33032−33040.

27. Cadaweller, K .A. et al. N-Terminally Myristoylated Ras Proteins Require Palmitoylation or a Polybasic Domain for Plasma Membrane Localization. Am. Soc. Micro. 1994 (14) 7, 4722-4730

28. F1000Res. 2015 Jul 16;4:261

29. Jencks, W.P., Carriuolo, J. General Base Catalysis of Ester Hydrolysis. J Am Chem Soc. 1961; 83 (7), 1743-1750

30. Good, DM, Wirtala, M, McAlister, GC, Coon, JJ. Performance characteristics of electron transfer dissociation mass spectrometry. Mol Cell Proteomics. 2007; 6 (11), 1942-1951

31. Macek, B, Hofsteenge, J, Peter-Katalinic, J. Direct determination of glycosylation sites in O-fucosylated glycopeptides using nano-electrospray quadrupole time-of-flight mass spectrometry. Rapid Comm in Mass Spec. 2001; 15(10), 771-777

Chapter 3

Improving Detection Sensitivity for Host-Cell Protein Characterization with Liquid Chromatography-Tandem Mass Spectrometry

3.1 Introduction

Host cell proteins (HCPs) are native proteins derived from a host organism used to express biotherapeutic proteins for drug substances (DS), such as therapeutic antibodies and fusion proteins [1,2]. HCPs may be unintentionally co-purified with a DS as process-related impurities, and the presence of these impurities in a DS may present problems in drug performance through eliciting immune responses in patients or contributing to the instability of drug formulations [2-5]. Therefore, measuring HCP levels to ensure depletion throughout manufacturing process development and at release testing is critical [6,7]. Purification steps during the manufacturing of a DS nearly completely deplete HCPs, and they are, therefore, at very low levels, often 100 parts per million (ppm) or less relative to the therapeutic protein [6]. However, these low levels present an analytical challenge for HCP identification and quantitation, as the high levels of therapeutic proteins can interfere with HCP detection.

Liquid chromatography- tandem mass spectrometry (LC-MS/MS) "bottom-up" proteomics offers a highly sensitive, semi-quantitative method for HCP characterization to

support downstream purification process development [8,9]. Using an LC-MS/MS 'bottom-up'

approach, HCPs can be identified by separating proteolytic peptides with one dimensional (1D)

or two dimensional (2D) LC [10], analyzing them withMS/MS [11], and conducting database

searches against the entire proteome of the host organism.

One approach to enrich for residual HCPs prior to LC MS/MS analysis, is to precipitate

biotherapeutic antibodies (Abs) from a sample after overnight native trypsin digestion (Figure

3.1) [12]. Developed at Eli Lilly, the principle hypothesis of this method is that the Ab in its

rigid, folded structure will have a slower kinetic rate of digestion by trypsin than natively folded

HCPs, and as a result, HCPs will be preferentially digested under native conditions. After

digestion, the undigested proteins are heat precipitated and pelleted with centrifugation leaving

the HCP-enriched supernatant for recovery and LC-MS/MS analysis.
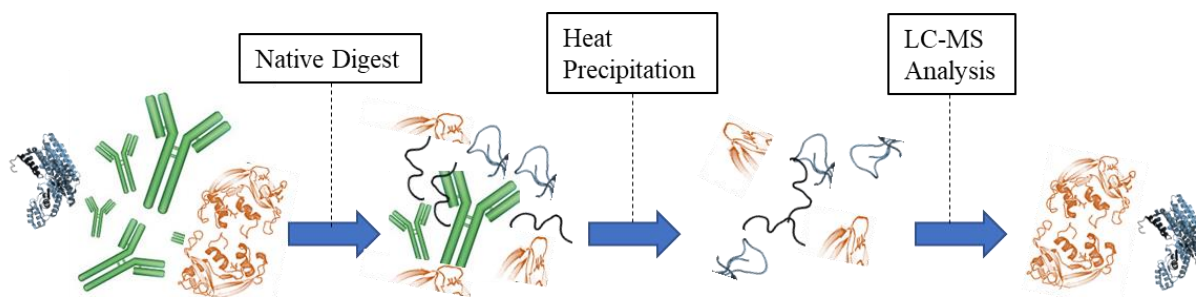
Figure 3.1 Native trypsin digestion of biotherapeutic proteins and their host-cell protein (HCP) content. Tightly folded antibodies have slower digestion kinetics than the HCPs, enabling enrichment of HCP peptides after heat precipitation of undigested proteins. Image Credit: Huang *et al.*

In contrast to traditional 'bottom-up' approaches, this depletion strategy can increase

dynamic range up to two orders of magnitude for HCP detection. This approach does have a few

drawbacks however, including the potential to lose heat-labile HCPs, co-precipitation of the

HCPs with the Ab, and an inherent lack of general applicability for characterizing HCP content

in non-antibody samples, such as fusion proteins, which may not have the same resistance to

proteolysis [13,14].

Recently, a novel hydrophilic interaction chromatography (HILIC) fractionation approach has shown potential for characterization of impurities in Ab products [14,15]. In HILIC separations, the strength of a protein's interaction with the polar stationary phase of the HILIC column is largely dependent on protein size and the number of charged residues [16]. In practice, the elution order has been observed to be mostly determined by molecular weight, with low molecular weight Ab related species eluting first [14,15].

Wang *et al.* recently implemented off-line HILIC fractionation to separate HCPs from the therapeutic protein [15]. This fractionation approach further improves the dynamic range with enhanced depletion of therapeutic proteins from the sample. In this method fractions were collected over the course of a HILIC separation with fractions containing therapeutic protein pooled and analyzed separately from HCP fractions. The low protein concentration of the HCP fractions enables higher volume sample loading during subsequent LC-MS/MS analysis and effectively increases the detection sensitivity for low level impurities. Additionally, the denaturing effects of high-acetonitrile concentrations [17] in HILIC separations ensures dissociation of strongly interacting HCPs that may co-precipitate. However, this approach has some practical drawbacks. The first concern is related to the multiple transfers of fractions between vials. Sample manipulation, such as these transfers, carry the risk of HCP loss due to adsorption to the vials they are transferred between and stability loss after the pH changes and concentration steps [6]. Second, the throughput is lower than the Huang *et al.* [13] native digestion method, and the manual workload is higher. Third, the method needs to be optimized for each therapeutic protein because the initial high organic phase in HILIC gradients can cause many therapeutic proteins to precipitate causing potential loss of HCPs and clogging of LC columns. Also, the LC gradients and fraction collection windows need to be optimized for proper separation of each sample. Ideally, a general protocol could be developed which limits sample

specific considerations and optimizations.   Here, we developed an automated HILIC

enrichment LC-MS/MS workflow that builds upon the previous offline HILIC fractionation

method. To improve the throughput and practicality of this method, we aimed to streamline the

workflow and develop a generalized platform for LC-MS/MS based HCP characterization and

quantitation that limits protein specific optimization requirements, increases throughput, and

decreases overall handling of the samples between steps.

3.2 Experimental

3.2.1 Materials

Recombinant IgG1 Abs (BMS mAb 1 purified drug substance and BMS mAb 2 purified by

protein A column), fusion proteins (BMS Fp 1, BMS Fp 2, BMS Fp 3), and Chinese hamster

ovary (CHO) null strain material (containing HCPs without drug substance) produced by Bristol-

Myers Squibb (Hopewell, NJ) were used in this work. Humanized IgG1κ monoclonal antibody

standard RM 8671 was purchased from the National Institute of Standards and Technology

(NIST, Gaithersburg, MD, NIST Ab). All chemicals were reagent grade or better. Ammonium

formate, ammonium hydroxide solution (~1 M in water), DL-dithiothreitol (DTT), dimethyl

sulfoxide (DMSO), and Trizma pre-set crystals (pH 7.6) were purchased from Sigma-Aldrich

Co. (St. Louis, MO). Water and acetonitrile (LC-MS grade), formic acid (FA), and

trifluoroacetic acid (TFA) were obtained from Thermo Fisher Scientific (Rockford, IL). Urea

(crystalline) was purchased from Alfa Aesar (Tewksbury, MA). Sequencing grade modified

trypsin was obtained from Promega (Madison, WI). Pre-mixed chromatographic solvents (water

containing 0.1% TFA, and acetonitrile containing 0.1% TFA), were from J. T. Baker Chemical

Co. (Phillipsburg, NJ).

3.2.2 Offline HILIC Sample Preparation

For assessment of biotherapeutic protein solubility in mixed solvent solutions, BMS mAb 1 (88 mg/mL) was diluted to 4 mg/mL into sample buffer containing aqueous acetonitrile with 0.1% trifluoroacetic acid (TFA) to achieve final compositions of 10, 15, 20, and 25% (vol./vol.) water. Protein solubility was assessed visually at room temperature and at 4 $^\circ$C over 24 h. NIST Ab, BMS mAb 2 and BMS Fp 1-3 were diluted from their respective stocks to 5 mg/mL into 25% water/75% acetonitrile (0.1% TFA) and the solubility of each sample was assessed observationally at room temperature over 24 h.

For assessment of gradient and chromatographic resolution of HCP proteins from biotherapeutic protein, 20 µg of BMS mAb 1 (88 mg/mL) was diluted into 200 µL of 25% water/ 75% acetonitrile (0.1% TFA) with or without 20 µg of spiked HCP from CHO standard null strain material. A 200 µL injection volume was utilized for each sample.

3.2.3 HILIC Separation Parameters and Fraction Collection

A Waters ACQUITY UPLC I-Class system was used for all HILIC separations with an ACQUITY UPLC BEH glycoprotein amide column (300 Å, 1.7 µm, 2.1 x 150 mm, Waters, Milford, MA). The column was set at 60 °C and the mobile phase solvents were HPLC grade water (0.1% TFA) as Buffer A and acetonitrile (0.1% TFA) as Buffer B. The samples were separated over 24 minutes at a flow rate of 0.200 mL/min. Gradient: Initial 28% A (0-2 min) and then ramped 28-42% A (2-24 min). To rinse the column of any residual contaminants the gradient was ramped 42-99% A (2 min) held at 99% A (3 min), and then re-equilibrated at 28% A (11 min). UV chromatograms were acquired using a PDA detector at 280 nm and 215 nm wavelengths using Waters Empower Software. Fractions were collected by an Acquity Fraction Manager in 5 mL wells of a 48- deep well plate. Fraction 1 was collected from 2.00-11.50

minutes, Fraction 2 was collected from 11.50-18.50 minutes, and Fraction 3 was collected from 18.50-24.00 minutes. For HCP characterization, all samples were prepared as a final concentration of 5 mg/mL in 25% water and 75% acetonitrile and 0.1% TFA. NIST Ab (10 mg/mL stock) was prepared at 2.5 mg/mL. For each sample, 200 μL injections were performed until 1.0 mg of protein was loaded onto the column, and the fractions were collected from one HILIC run.

3.2.4 Trypsin Digestion

Ammonium hydroxide solution (10% in acetonitrile) was added to each fraction to adjust pH to 6-8 (~50 μL per mL) before each fraction was concentrated to less than 50 μL under nitrogen flow using a Glas-Col gas inlet, heated Basic Heated ZipVap Evaporator for 96-well plates set to 40°C. Concentrated or dried samples were reconstituted with 200 μL of digestion buffer containing 2.0 M urea and 50 mM tris-HCl (pH 7.6). Fractions were mixed to dissolve all protein, then combined with 2 μL of 50 mg/mL DTT solution and mixed again. Fractions were then incubated for 30 min at 60 °C. After cooling to room temperature, each fraction was treated with 2.5 µg of trypsin, mixed, and incubated at 37 °C for 18 h.

3.2.5 LC-MS/MS Analysis of Digested HILIC Fractions

The total volume of the digest for each HILIC fraction (~200 µL) was injected for UPLC-MS/MS analysis unless otherwise specified. Tryptic digests were separated using a Waters ACQUITY UPLC CSH C18 column (1.7 μm, 2.1 x 100 mm) on a Waters ACQUITY UPLC system (Milford, MA). The mobile-phase Buffer A was water (0.1% FA) and Buffer B was acetonitrile (0.1% FA). Separation was performed at a column temperature of 60 °C with a mobile-phase gradient as reported by Lagassé *et al* [115]. On-line mass spectrometric analysis was conducted using an Orbitrap Fusion Lumos Mass Spectrometer (Thermo Scientific,

Waltham, MA) in positive ion mode utilizing an orbitrap for MS1 scans and an ion trap for precursor selection and MS^2 scans. MS1 settings were as follows: Electrospray voltage : +3500 V, sheath gas: 30, Aux gas: 7, sweep gas: 1, ion transfer tube: 300 °C, Vaporizer: 150 °C, Orbitrap precursor selection: 120k resolution, scan range 230-1500 m/z, max inject time: 50 ms, AGC target: 4x10^5, RF lens: 30%, dynamic exclusion: 60 s, mass tolerance: 10 ppm, MS^2 settings were: activation: HCD, scan rate: rapid, max injection time: 35 ms, Isolation window: 1.4 Da

3.2.6 Peptide Identification and Protein Quantitation

MS/MS data were searched for HCP identification using Byonic 3.1.0 (Protein Metrics, San Carlos, CA) against a customized protein database including the therapeutic protein sequence and the CHO or mouse proteome, as appropriate, downloaded from UniProt.org. Search parameters were set as follows: mass tolerance as 10 ppm for precursors and 0.4 Da for fragments, fully specific peptide termini with ≤ 3 missed cleavages, and protein FDR cutoff at 1%. N-terminal glutamine conversion to pyroglutamate was considered a common variable modification. Protein hits with |Log Prob| score lower than 3.0, decoys, and common protein contaminants were excluded. For quantitation, only proteins with at least 3 unique peptides of sequence length 6 or greater were included. All digested samples were spiked with 5000 fmol of Waters MassPrep enolase digestion standard (dissolved in 1 mL water) to be used as internal calibrant. Absolute quantitation was completed with Progenesis QI.

3.2.7 Digestion of Proteins in Reducing and Denaturing Conditions Following the "Native Digestion" Protocol

All samples were diluted to 5.0 g/L in 50 mM tris-HCl (pH 7.6) to a final volume of 200 μL. For samples in reducing conditions, 2 μL of 50 mg/mL DTT solution was added to the buffer. For samples in denaturing conditions, 2.0 M urea was added to the 50 mM tris-HCl (pH

7.6) buffer. Each sample was digested, reduced, and precipitated according to the protocol described in Huang *et al* [13].

3.3 Results and Discussion

The aim of this work was to develop an LC-MS/MS based, semi-automated platform for HCP analysis in biotherapeutic drug substances. We built upon a previously described HILIC fractionation method [15] to develop robust sample preparation and HILIC separation conditions which can be broadly applied to characterize HCP content in most biotherapeutics with minimum sample specific optimizations. Additionally, we evaluate both the qualitative and quantitative sensitivity of our advanced HILIC method compared to native digestion combined with LC-MS/MS analysis. A flow chart of our developed HILIC fractionation method and LC-MS/MS analysis is shown in Figure 3.2. Our focus was primarily on optimizing the HILIC fractionation procedure to increase analytical throughput, with a plate-based approach, and optimizing separation parameters for greater generalizability across different DS and protein classes.
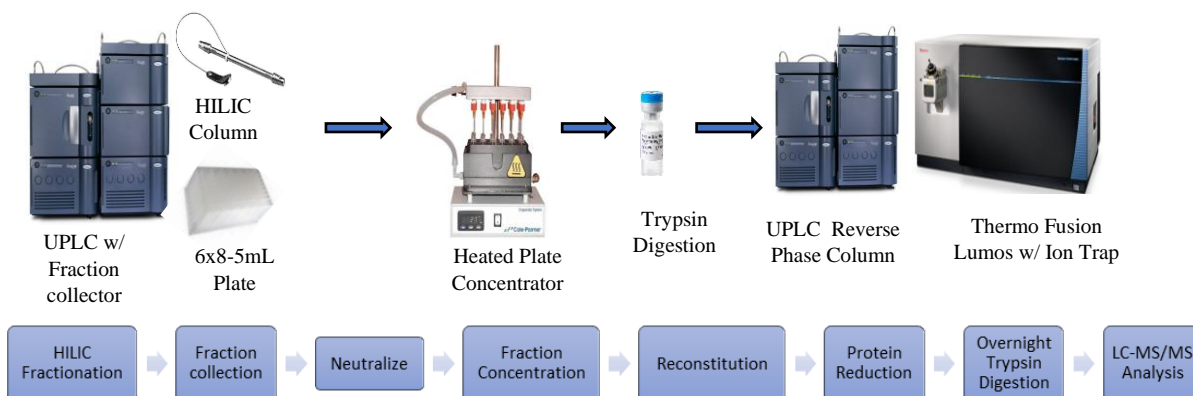
Figure 3.2. Advanced HILIC fractionation method workflow for HCP characterization in biotherapeutic proteins.

3.3.1. Developing a Universal, Automated HILIC Fractionation Method for HCP Characterization

The first step in our workflow is offline- HILIC fractionation of diluted DS. In contrast to reverse phase (RP) chromatography, HILIC separation uses a polar stationary phase, and therefore, the gradients are initiated with a high concentration of organic mobile phase and increase in polar mobile phase to elute retained polar analytes. Drug substances can be carefully formulated to achieve protein concentrations up to 100 g/L or more [18]. This high concentration is problematic for separations on typical analytical columns, which have limited loading capacity, and thus dilution prior to injection is necessary. Additionally, HILIC columns require long equilibration periods to re-establish the desired mobile phase concentration and water layers on the polar resin bed. Disruption of this layer can lead to retention time shifts between injections and poor reproducibility [19]. Injections of samples composed of high aqueous concentrations can disrupt the equilibrated column leading to partial breakthrough of unretained proteins [20,21]. This problem can be avoided simply by preparing protein samples in a buffer closely matching the initial mobile phase composition. This solvent composition matching is especially important with high injection volumes that exceed the total column volume. However, biotherapeutic proteins, such as antibodies which exhibit high water solubility due to their relatively low isoelectric point (pI) values [18], lack the solubility necessary to dissolve in high acetonitrile concentrations. Wang *et al.* addressed this issue by carefully formulating buffers composed of solubilizing reagents, such as trifluoroacetic acid and DMSO, or by using multiple small injections of highly concentrated aqueous DS to avoid disrupting the equilibrated column. However, these conditions are sample specific, thereby increasing the method development time and limiting the overall feasibility of this method as a universal platform across different protein classes. We sought to improve upon this strategy by evaluating the solubility of multiple DS in

various buffer compositions in order to identify a single sample buffer composition that is compatible with the HILIC mobile phase buffers and maintains solubility across different types of biotherapeutic proteins.
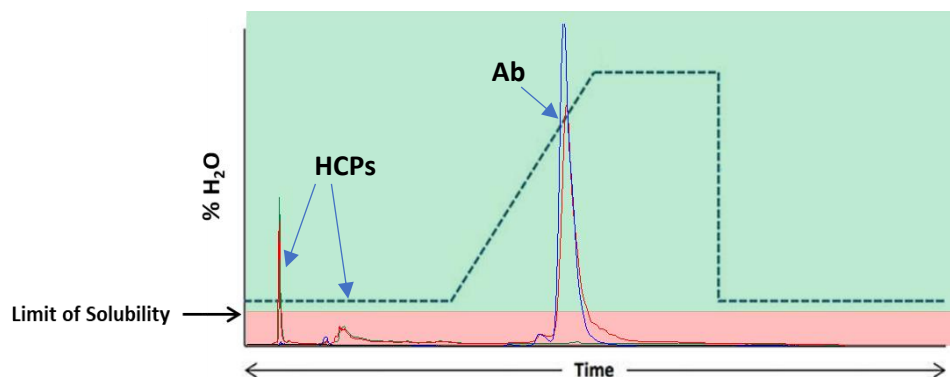
Figure 3.3. Sample solubility is dependent on buffer composition which needs to be balanced with an initial high organic mobile phase for HILIC separation of biotherapeutic proteins from residual HCPs.

Additionally, the sample must remain soluble on the column throughout the gradient to prevent potential column clogging from precipitated material and to avoid loss of the sample to maintain accuracy of absolute peptide quantitation in the later steps. Each protein sample has a 'limit of solubility' in high organic concentration mobile phase which needs to be considered when developing the HILIC gradient (Figure 3.3). Therefore, an optimized gradient needs to be designed with a high enough aqueous phase throughout the separation to maintain solubility of the samples and provide sufficient resolution for fractionation across multiple biotherapeutic proteins.

To address these concerns, we first aimed to determine the minimum concentration of water needed to completely solubilize most biotherapeutic proteins in our mobile phase buffer (H_2O/acetonitrile, 0.1% TFA) and then match the initial gradient conditions on the HILIC column with the same buffer composition. It is important to note that, when preparing the samples, order of addition is important to ensure the protein remains soluble and to prevent

precipitation of the proteins during the dissolution of organic solvents or highly concentrated acids. Pre-mixing the buffer components prior to the addition of the DS to the sample showed to be the best practice to avoid precipitation of the protein samples during addition of organic solvent or concentrated acid.

We evaluated the solubility of the Bristol-Myers Squibb (BMS) proprietary therapeutic monoclonal antibody (mAb) BMS mAb 1 (88 g/L stock, 5 g/L final) over a range of organic phase buffer concentrations (Figure 3.4, Table 3.1). We monitored the solubility of the mAb over 24 h by visual inspection of the mixed sample at 5 and 25 $^{\circ}$C. The presence of white, cloudy precipitate indicates a high degree of precipitation, and a completely homogenous and clear solution was considered indicative of good solubility (Figure 3.4). Our observations indicated that BMS mAb 1 was insoluble in less than 20% aqueous buffers and remained partially soluble until a composition of 25% aqueous phase or greater was reached. Additionally, it was observed that this solubility was temperature dependent, and when these samples were placed in a refrigerated autosampler (5 $^{\circ}$C), the samples would become insoluble and precipitate (Table 3.1) at these organic phase concentrations. This temperature sensitivity explains the necessity for buffer additives in the method published by Wang et al. [15] which reported solubilizing some samples in up to 28% aqueous buffer. Therefore, by preventing refrigeration of the mixtures, we can potentially improve sample solubility and eliminate the need for disrupting equilibrated column conditions with compounds such as DMSO.
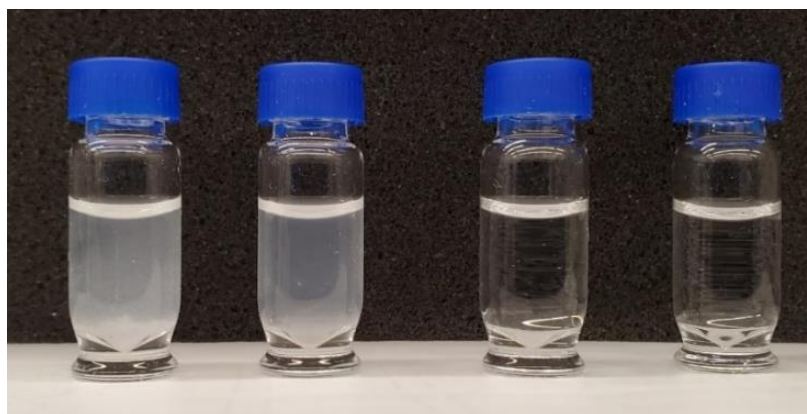
Aqueous composition: 10%   15%   20%   25%

Figure 3.4. BMS-mAb 1 solubility in mixed phase buffers (5 g/L) in H_2O/MeCN (0.1% TFA).

Table 3.1 Solubility of BMS-mAb 1 (5 mg/mL) in 0.1% trifluoracetic acid in water over 24 hours.

H_2O (%)	25°C	5°C
10	Cloudy	Cloudy
15	Cloudy	Cloudy
20	Translucent Particles	Cloudy
25	Cloudy	Cloudy
28	Cloudy	Cloudy

BMS mAb 1 was determined to be soluble at room temperature over 24 h in 25% aqueous buffer which fits well within the limitations imposed in a HILIC gradient separation. Although these solvent conditions are compatible with this particular DS, to evaluate their wider applicability we compared the same conditions for six different proteins representing two distinct biotherapeutic protein classes, mAbs and fusion proteins (Fp). All proteins were added to room temperature buffers containing 25% water in 0.1% TFA and assessed for solubility (Table 3.2). Five proteins were observed to be soluble in these conditions with only one protein (BMS Fp 3) lacking sufficient solubility. This higher insolubility is possibly attributable to a high degree of sialylation of the O-glycans of this particular protein, further lowering its pI, and decreasing the

solubility in acidified acetonitrile or non-polar mobile phases. Additional considerations would

be required for analyzing this specific sample, such as deglycosylation with PNGase F or

marginally increasing the water content of the buffer. Overall, these buffer conditions appeared

mostly universal across two different protein classes reducing the need for sample specific

optimization except in specific circumstances like our highly glycosylated protein.

Table 3.2 Therapeutic protein solubility in 25% H_2O/acetonitrile (0.1%TFA).

Protein Sample	Concentration (g/L)	Temperature (^{o}C)	Soluble
NIST mAb	2.5	25	Yes
BMS-mAb 1	5.0	25	Yes
BMS-mAb 2	5.0	25	Yes
BMS-Fp 1	5.0	25	Yes
BMS-Fp 2	5.0	25	Yes
BMS-Fp 3	1.0	25	No

The next areas we addressed were the selection of a universally applicable HILIC

gradient and automation of the fraction collection. In the Wang *et al.* method, fractions were

manually collected with microcentrifuge tubes and then pooled into three distinct fractions: a

pre-mAb fraction, the mAb fraction, and a post-mAb fraction (Figure 3.5). The gradient was

previously optimized for different therapeutic proteins based on their buffer compositions. For

example, BMS mAb 1 was injected in pure water at 100 mg/mL whereas NIST Ab was injected

in 28% water with 5% DMSO. Both of these samples required different HILIC gradients to

prevent chromatographic breakthrough with the some of the mAb eluting with the solvent front

and to retain resolving capacity of the column. In order to reduce workload requirements and

improve the throughput of the method, we added a Waters Acquity fraction manager, which

enables automated fraction collection based on time dependent collection parameters. Due to the

high degree of resolution achieved with the HILIC gradient used in Figure 3.5, we devised a set

of fraction collection parameters that work across a wide variety of similarly sized proteins using

the same generalized gradient.
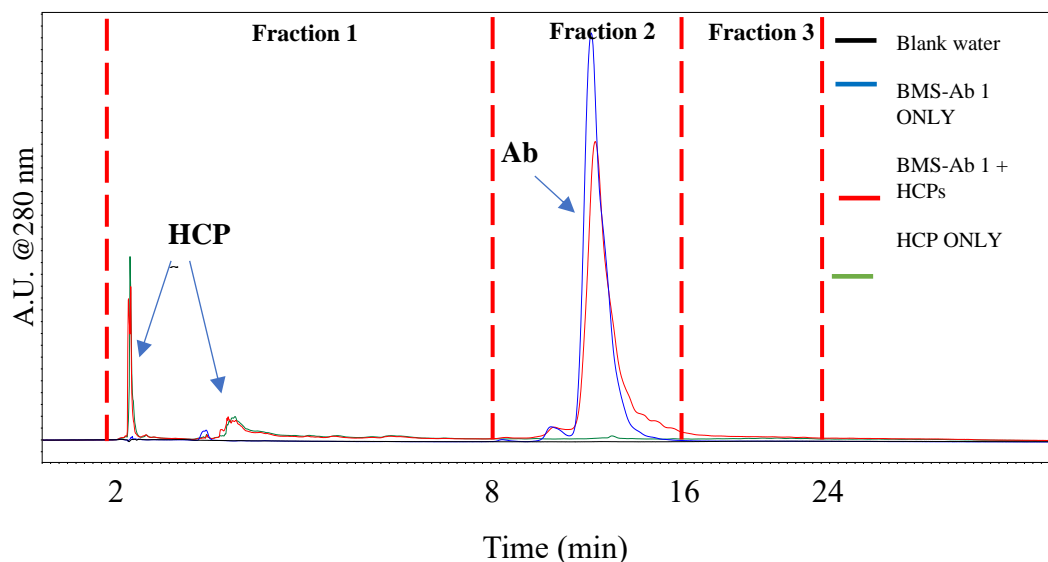
Figure 3.5. HILIC separation of biotherapeutic proteins and HCPs offers sufficient resolution to clearly separate fractions into three collection windows. Graphic credit: Wang *et al.*

An advantage of offline fractionation is that larger volumes of DS can be fractionated, and thus higher volumes of low abundant HCPs will be available for digestion and analysis. Additionally, if a high degree of separation is achieved and DS is effectively depleted from HCP fractions, higher volume injection of the corresponding proteolytic peptides is feasible in the LC-MS/MS analysis, increasing the abundance of HCPs. For comparison, the native digestion protocol by Huang *et al.* [13] digests 1.0 mg of Ab and injects 1/4 of the total volume for LC-MS/MS analysis because samples with higher concentrations would require further depletion of the mAb from the sample in order to prevent overloading of the analytical RP-LC column and increasing background peptides from the digested Ab. We expected that injecting a greater amount of HCP enriched fractions for LC-MS/MS analysis should further increase our sensitivity for HCPs in the sample, however, it would also require high loading conditions, more than 250 µg protein, during the HILIC step which could potentially affect the chromatography and the quantitative capability of the method.
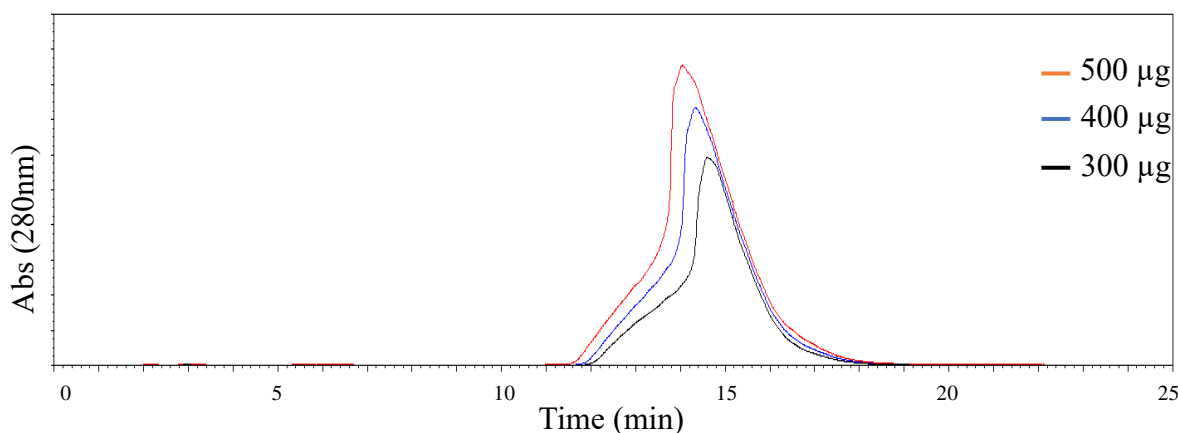
Figure 3.6. UV chromatograms (280 nm) of BMS-Ab 1 on a 2.1x100 mm BEH glycoprotein amide column.

Table 3.3 Peak area aatios of BMS-mAb 1 (280 nm) on a 2.1 x 100 mm BEH glycoprotein amide column

Mass (µg)	Peak Area (kAU*s)	RSD (%)
1000	430	1.2
500	210	3.2
400	170	1.5
300	130	2.4
R^2		1
		n = 3

To evaluate our analytical HILIC column's capacity to both resolve and quantitatively elute our samples, increasing volumes of BMS mAb 1 (5 mg/mL) were injected and separated with an initial mobile phase of 28% water. Figure 3.6 shows example UV-chromatograms from 300, 400, and 500 µg injections, acquired at 280 nm. The peak areas were integrated (Table 3.3) to validate that no loss of protein to precipitation or column retention occurred. These results indicate that quantitative linearity was maintained up to 1000 µg of BMS mAb, however larger injection volumes might require a larger diameter column to prevent band broadening.

Having developed working HILIC separation conditions, we next sought to automate the fraction collection parameters which may be sample specific. How strongly a column retains

material with a given gradient is dependent on the number and strength of hydrophilic interactions between the protein and the stationary phase [14,20]. HILIC column loading conditions involve high organic phase solvent which may denature proteins. Acetonitrile disrupts the hydrogen bonding interactions between amino acids and limits higher order structure formation; therefore, column retention is typically a function of the number of polar groups available for interaction with the stationary phase, such as polar side chains and the amides in the protein backbone. The number of interacting groups is therefore protein size dependent, and smaller proteins tend to elute first. However, there are other variables, such as the abundance of highly polar amino acids or PTMs which increase the overall polarity and retention profile of a protein, which should be taken into consideration when predicting retention characteristics [20,21].

Timing the fraction collection windows such that one set of parameters will apply across different samples necessitates that the main therapeutic protein peaks, like those shown in Figure 3.6, exhibit similar retention times between samples. We designed our collection parameters for a 1.0 mg injection of BMS mAb 1 to collect three fractions, where the first and third fractions contain HCPs, while the second fraction contains the therapeutic protein. We set our fraction collection windows based on the retention profile of BMS mAb 1, where the main therapeutic peak eluted from 11.5-17 min. We then compared the retention time profiles between three mAbs (Figure 3.7). UV-chromatograms of each sample shows similar profiles and retention characteristics between mAbs, which was expected as they are all similarly sized proteins.
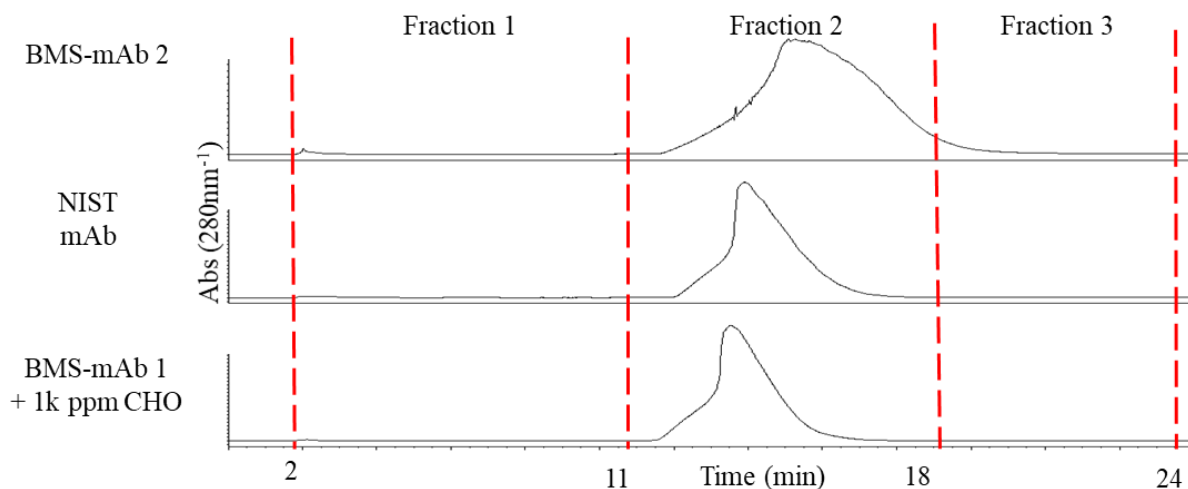
Figure 3.7. UV chromatograms (280 nm) of sample Abs on 2.1x100 mm BEH glycoprotein amide column show similar retention profiles between antibodies. Fraction windows can thus be generalized between Abs.

Fraction windows for these samples indicated by the red dotted- lines in Figure 3.7 showed effective fractionation for all three proteins. NIST mAb, a well-characterized and purchasable monoclonal IgG standard purified from a Chinese hamster ovary (CHO) culture, exhibited significant retention profile overlap with BMS mAb 1 and eluted from 12-17.5 min. However, BMS mAb 2, another BMS proprietary therapeutic antibody, exhibited a broader peak profile and eluted from 11.5 to 19 min indicating the need to widen the fraction 2 collection window another 30 sec. According to two previous studies [14,15], HCPs almost entirely elute in fraction 1 (Figure 3.6), and subsequent studies by Wang *et al.*, characterizing HCP content in NIST mAb and BMS mAb 1 using HILIC fractionation, reported less than 2% of total HCP content detected in fraction 3. Therefore, it is crucial to accurately define the collection endpoint for fraction 1 to separate the therapeutic protein from HCP content. However, since fraction 3 offers little informative value, optimizing the collection endpoint for fraction 2 after the main peak is of lesser importance.

We concluded, from these results, that these parameters required no further optimization for IgG-based mAbs. We next examined how general these parameters are across different types of proteins by characterizing two BMS proprietary fusion proteins, BMS Fp 1-2 (Figure 3.8). Interestingly, both fusion proteins eluted later than NIST mAb. BMS-Fp 1 eluted from 17-25 min, and BMS Fp 2 eluted from 14.5-19.5 min, indicating that some optimization between different protein types is necessary. However, separation of the fusion proteins from the early eluting HCPs was observed. Notably, both fusion proteins showed significant peaks in the fraction 1 collection window. These peaks were later determined to be low-molecular weight (LMW) fragments of the fusion proteins. This observation is a concern when considering this fractionation method, as LMW are common in biotherapeutic DS. Therefore, sample-specific optimizations may be necessary when contaminating LMWs are present. However, widely available automated fraction collection systems that can monitor peak formation and separate fractions accordingly may be better suited as part of a HCP characterization platform.

Furthermore, both fusion proteins are smaller than the mAbs we tested, and the longer retention times were therefore unexpected. This behavior is possibly due to the slightly higher degree of glycosylation of the two fusion proteins. Additionally, the lower abundance of disulfide bonds in the fusion proteins compared to the mAbs enables more complete denaturation of the proteins in organic buffers and increases the protein-column interactions. Overall, our chosen parameters worked across each of the Abs we characterized and provided a basis for our HILIC platform.
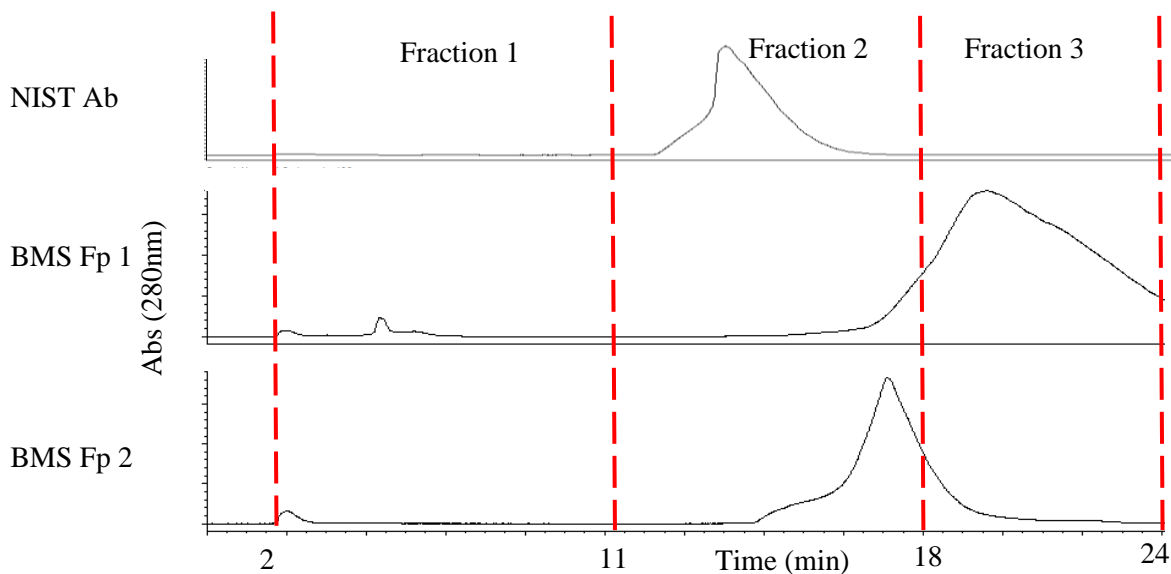
Figure 3.8. UV chromatograms (280 nm) of NIST Ab and BMS Fp 1-2 on a 2.1x100 mm BEH glycoprotein amide column show similar retention profiles between the fusion proteins but not for the Ab. Fraction windows need to be optimized between proteins classes.

Proteins can adsorb to the surfaces of plastic and glass containers [22-27] used in sample processing which contributes to quantitation errors and poor detection sensitivity of low abundance species. This problem is compounded with multiple transfers between containers and, similarly, this phenomenon is likely to occur when passing a protein or peptide solution through a filter or membrane. Therefore, limiting the number of sample transfer steps, including removal of filtration and buffer exchange steps, is important in limiting the potential loss of HCPs. In our efforts to improve upon the HILIC fractionation method reported by Wang *et al*. [15], we sought to eliminate as many sample transfers as possible by using a plate-based workflow that collects each fraction into a single well (Figure 3.2). We implemented a 48- deep well (5 mL) plate that can pool each fraction into a single well, resulting in three distinct fractions at the end of our gradient and eliminating the pooling step. Furthermore, the following steps from sample concentration through trypsin digestion could be completed without transferring the samples

between vials or plates, reducing any sample handling that can result in protein loss. This strategy also increases the throughput of the method where up to sixteen samples per plate can be fractionated in tandem, concentrated and digested simultaneously, and then sequentially analyzed with LC-MS/MS.

3.3.2. Identifying HCPs in Abs by HILIC Fractionation LC-MS-MS

Ultimately, our aim with the developed HILIC fractionation method is to achieve industry leading sensitivity for detecting and identifying low abundance HCPs in purified downstream biotherapeutic protein samples. However, most DS have been subjected to multiple purification steps (Figure 1.8) [6,28,29] and only contain a few detectable HCPs, if any at all. This low HCP concentration presents a problem for evaluating the sensitivity of our method with a purified DS which may contain only one or two detectable HCPs, not ideal when building a statistically relevant comparison of the methods. Therefore, for method evaluation purposes, we analyzed the HCP content in an upstream product, BMS mAb 2 that contained approximately 3000-5000 ppm HCP after Protein A purification [28,29]. Similarly, NIST mAb is a well-characterized mAb standard with approximately 2000 ppm HCP [30], and it can be used as a benchmark for evaluating our method's performance [30].

Our strategy to achieve even greater sensitivity in analysis, was to perform our experiments on a Thermo Oribtrap Fusion Lumos mass spectrometer, a sensitive instrument with a near-state-of-the art 40 Hz acquisition rate and capable of both ion trap collision induced dissociation (CID) and higher energy CID (HCD). Resulting '.raw' files were searched against a compiled '.fasta' database using the Byonic (Protein Metrics, Cupertino, CA) search algorithm outputting a list of protein identifications. Comparing our HILIC fractionation method result list to the native digestion method [13] provides a benchmark from which we can evaluate the

sensitivity of our method. The native digestion method offers the highest sensitivity LC-MS/MS analysis to date, with exception of the Wang *et al.* method [15] that serves as the foundation of our workflow.

We began assessing our automated HILIC fractionation method by analyzing the HCP content of BMS mAb 2, the upstream purification product, and NIST mAb with our defined workflow (Figure 3.2), filtering the resulting lists to remove false positive identifications. The resulting lists were then compared to a parallel experiment that followed the native digestion protocol. Both experiments were repeated in triplicate for BMS mAb2 and NIST mAb, and the total number of positive hits identified by each method was averaged and compared (Table 3.4, yellow columns).

Our fractionation method on average identified 477 HCPs in BMS mAb 2 and 45 HCPs in a NIST mAb sample over the three runs. Conversely, the native digestion method averaged only 281 HCP identification in BMS mAb 2 and 41 HCPs in the NIST mAb sample. Our results indicate that our HILIC fractionation method resulted in more total positive identifications than the native digestion method for both proteins, and BMS mAb 2 showed an almost 2-fold increase in total positive identifications.

Table 3.4. Comparison of average HCP identifications between our HILIC method and the native digestion method. *n=3*

Samples (n=3)	Native Digestion		HILIC Method		HILIC Fractions		
	Total	Unique	Total	Unique	1	2	3
BMS-mAb 2	281	81	477	276	358	85	194
NISTmAb	41	24	45	28	30	7	17

After filtering out the proteins that were identified by both methods, the resulting lists were comprised of protein identifications that were unique to their respective methods. Again, the total unique positive identifications were totaled and averaged (Table 3.4, green columns). In

both samples, the HILIC method outperformed the native digestion method and totaled more unique identifications with more than 50% of the NIST mAb total identifications being unique to their respective methods. Similarly, for BMS mAb 2, the HILIC method resulted in 276 unique identifications that were not detected with the native digestion method. However, it is important to note that the native digestion method did identify proteins in both samples that the HILIC method did not detect.

Overall, the HILIC method resulted in more sensitive detection and more positive identifications, but it is important to consider lack of overlap in identifications. The heat precipitation step in the native digestion method is a potential cause of HCP loss as is the concentration step in the HILIC fractionation method due to changing buffer conditions and potential HCP insolubility at low volumes. These factors may lead to effective 'blind spots' in HCP characterization and limit the sensitivity of their respective methods. This result suggests that the two methods should be used together to obtain a more complete overall profile of HCP content in more complex Ab samples. However, our HILIC fractionation method is a practical option for non-Ab type species.

3.3.3. Absolute Quantitation of HCPs in Therapeutic Abs with "Hi-3" method

While residual HCP identifications are useful, regulatory standards require reporting the total concentration of HCPs in a DS [31-33]. With sensitive assays, such as ELISA, offering high parallelizability and quantitative analysis, a competitive LC-MS method needs to provide not only identification lists of detectable HCPs but quantitative measurements of each identified protein as well. The "Hi-3" method [34] can be used to approximate the concentration of proteins in a sample taking the top three most ionizable peptides from each positive HCP identification and calculating their molar abundance based on the top three ionizing peptides from a spiked

internal standard at a known molar concentration. The concentrations can then be converted to ppm values (ng HCP/ mg therapeutic protein) for reporting. Quantifiable proteins were selected by including more stringent filtering conditions than the previous identification lists including a minimum of three unique peptides per protein in order to confidently use the "Hi-3" quantitation method. Also, a minimum peptide length of six amino acids was required to avoid grouping short, conserved sequences to multiple proteins or measuring the additive signal of a conserved sequence from multiple proteins, both of which contribute to artificially inflated abundances.

We quantified the HCP content in both BMS mAb 2 and NIST mAb with both methods and compared the HILIC fractionation method to the native digestion method. Unsurprisingly, proteins that were identified and quantified in both BMS mAb 2 and NIST mAb were consistently measured in higher concentrations by the HILIC fractionation method which approximated 1828 ± 1003 ppm HCP concentration in NIST mAb compared to 575 ± 49 ppm HCP approximated by the native digestion method. Not only was the overall total concentration measured to be higher with HILIC fractionation, but almost every identified protein was individually measured at higher concentrations as well.

One of the HCP identifications in our NIST mAb analysis led us to question the validity of our absolute concentrations because the presence of clusterin was detected by both methods in two out of the three replicates. Clusterin was measured to be near 100 ppm which made it one of the most abundant HCPs identified in the samples, which is unusual as clusterin has not been previously characterized in NIST mAb. At such a high abundance level, even less sensitive methods should have identified this protein if it was a true contaminant. Potential sample contamination issues may have led to this identification, and therefore, we chose not to publish our identification lists until further validation experiments could be carried out. Additionally, the

presence of exogenous contaminants precluded comparing our results to the published HCP concentrations [30] measured by ELISA, however, we can accurately compare our results between our methods as both analyses were carried out on the same stock samples.

The under approximation observed with the native digestion method may be explained by poor digestion efficiency of proteins in their native folded conditions which lowers the overall rate of digestion. This condition may have the effect of lowering the average ion signal of each peptide in the MS analysis and result in underestimation of protein concentrations. From these data, we can further surmise that the HILIC fractionation method is not only more sensitive, but potentially more accurate than the native digestion method. However, the precision and reproducibility of the measurements will need to be improved before such conclusions can be confidently asserted.

One of the current drawbacks to peptide quantitation with Progenesis QI, is the propensity for proteins to have conserved tryptic peptide sequences. The software assigns these peptides to each protein, causing high abundance peptides from one protein to also be attributed to other low abundance proteins. This artifact potentially results in overestimation of the individual concentrations of these low abundance peptides. Resolving these conflicts is carried out manually and for each conflicting peptide adds significant time to the analysis. The BMS mAb 2 sample had over 400 quantifiable proteins with hundreds of peptide conflicts that needed manual resolution. Our approach to improve the feasibility of this analysis was to filter for a list of known 'problem HCPs' that have been documented in drug development as commonly characterized impurities that are problematic in drug formulations. Included in this list are proteases that degrade protein stability, lipases that degrade the stabilizers in the drug formulations, and potential effectors of immunogenicity in drug administered patients.

Table 3.5 Quantitation of Problematic HCPs identified in BMS mAb2.

Accession	Protein	HILIC (ppm)	Native Digestion (ppm)
G3H6V7	Lipoprotein lipase	104.4 ± 15.7	10.8 ± 15.3
G3H1D5	Carboxypeptidase	94.7 ± 20.4	11.9 ± 16.8
G3HQY6	Lipase	4.4 ± 5.6	8.4 ± 11.9
G3H8V5	Carboxypeptidase	1.9 ± 0.1	
G3I6T1	Phospholipase B-like	1.8 ± 0.3	6.9 ± 9.8
A0A3L7HCS7	Carboxypeptidase	0.6 ± 0.5	

We compiled the results (Table 3.5) and compared both LC-MS methods. Six 'problem HCPs' were identified in BMS mAb 2 by the HILIC fractionation method while only four of those proteins were identified by the native digestion method. Again, it appears that our HILIC fractionation method outperforms the native digestion method with a more sensitive protein analysis. Similar to what we observed in our NIST mAb quantitative analysis, the HILIC method also measured higher concentrations of each peptide with lipoprotein lipase and carboxypeptidase (Table 3.5) both calculated to have about 10-fold higher ppm values than the native digestion method, and the lowest concentration measured by the method reached 0.6 ± 0.5ppm. These results indicate that our method lowers the detection limit approximately 10-fold compared to any other LC-MS method to date.

Overall, the results indicate that our HILIC fractionation method is a highly sensitive platform for HCP detection and absolute quantification in purified DS, and using this method we effectively lowered the limit of quantitation (LOQ) an order of magnitude. The native digestion method, on the other hand, offers a robust and technically simple protocol for HCP identification, but it lacks the ability to characterize non-Ab samples. When comprehensive HCP profiling in Ab samples is necessary, both methods can be used in conjunction to significantly

increase the number of positive identifications for Ab-type samples, however, when absolute quantitation is necessary, our method outperforms other LC-MS techniques.

3.3.4. Improving the Native Digestion of Therapeutic Antibodies for HCP Characterization

The native digestion protocol offers a practical and elegant solution to Ab depletion in HCP analysis. However, the previous evaluations indicate significant limitations with this method in detection sensitivity and accuracy of the absolute protein concentration measurements. We believe the observed underperformance may be a result of poor digestion efficiency, lowering the overall quantity of peptides available for ionization and we therefore sought to evaluate different digestion conditions to improve overall efficiency and potentially increase the detection sensitivity and quantitative accuracy of the method.

The Ab depletion mechanism underlying the native digestion method, presented by Huang *et al.* [13], is based on the slower kinetics of Ab digestion under native folded conditions. In their protocol, they forego the use of reducing and chaotropic reagents that disrupt native protein structures. However, this approach was observed to lower the total HCP peptide concentrations in our samples as HCPs may be similarly difficult for trypsin to digest under native conditions. Therefore, we explored the effects of adding reagents such as dithiothreitol (DTT) and urea to the samples prior to digestion. We hypothesized that by disrupting some of the native folding, but not all, we would see some improvement in digestion efficiency without sacrificing sensitivity by over-digesting the abundant therapeutic protein.

Replicate samples of NIST Ab and BMS mAb 2 were analyzed according to the Huang *et al.* protocol, however, in our experiments, they were digested overnight in buffers containing either 3.25 mM DTT, 2.0 M Urea, or a combination of both. These variables respectively represent reducing conditions, denaturing conditions, or both. Total HCP counts of all unique

proteins identified between replicates were compiled and compared across conditions (Figures 3.8 and 3.9). In contrast to the analysis in Table 3.4, compiling all unique proteins identified in our replicates effectively increased the total number of unique identifications made by the native digestion method compared to averaging the results.

Averaging total HCP counts between replicates helps indicate the robustness of our methods, however since the robustness of the native digestion protocol is well documented, adding all positive identifications in each replicate and removing duplicate identifications provides a more comprehensive profile of HCP content in our samples. Although, while this approach can provide an improved assessment of the sensitivity of the method, certain proteins that scored near our lower confidence threshold and that were identified in only one or two of the replicates, required further validation to ensure these hits are not false positives.

NIST mAb Native Digestion HCP count

DTT

	-	+
-	60	81
Urea		
+	46	47

$n = 3$

Figure 3.9. Average total HCPs identified in NIST Ab using the native digestion protocol with 2.0 M urea, 3.25 mM DTT, or a combination of both in the digest buffer. Green tiles indicate statistical significance ($p < 0.05$).

BMS-mAb 2 Native Digestion HCP count

DTT

	-	+
Urea -	362	350
Urea +	371	374

$n = 3$

Figure 3.10. Average total HCPs identified in BMS mAb 2 using the native digestion protocol with 2.0 M urea, 3.25 mM DTT, or a combination of both in the digest buffer. Green tiles indicate statistical significance ($p < 0.05$).

Between the two experiments, only the (+) DTT condition in our NIST mAb sample showed a statistically significant increase in the number of HCPs identified, indicated by the green tile (Figure 3.10). This result is most likely due to the reduction of disulfide bonds allowing for more efficient digestion of the HCPs, but validation experiments are needed to verify these results. The BMS mAb 2 sample, on the other hand, was unaffected by the additions. In all three replicates the urea-containing samples showed slightly more hits, but additional experiments will be necessary to reach statistically relevant conclusions. Since urea acts as a denaturant during digestion, the antibody should unfold and produce more proteolytic peptides which increases the overall background and reduces the detection sensitivity for low abundant peptides that may coelute.

The lack of variation between conditions indicated by our results was unexpected and challenges the underlying hypothesis proposed by Huang *et. al* which rationalizes native structural differences between Ab and HCPs underlying the slower Ab digestion kinetics compared to other proteins. If the tightly folded native structure is the driver of this phenomenon then the addition of urea, as a denaturing reagent, and DTT, as a reducing agent, should have

disrupted the native structure of the mAb species, thereby significantly increasing the number of observed mAb peptides. This increase should have the effect of increasing ion suppression of HCP peptides as well as background noise, and should have limited the detection of the HCPs in the samples. Instead, we saw little change in the detection limits in the denatured and reduced samples. In fact, the reduction of NIST mAb samples increased the number of HCPs identified, and the addition of both reagents to the BMS mAb 2 sample slightly increased the number of HCP identifications in our analysis. While further work needs to be performed to compare the overlap of identified peptides between conditions to validate our results, the lack of significant change between conditions was unexpected and potentially contradicts the proposed HCP enrichment mechanism. It is possible that the lack of heat and relatively low urea concentration (2.0 M) was insufficient to denature the tightly folded mAbs, and addition of DTT, which reduces disulfide bonds, may have improved HCP digestion in the NIST mAb sample while still leaving the heavy and light chains of the mAbs tightly wound and difficult to digest. Overall, denaturing and reducing conditions either significantly improved or had little effect on the detection sensitivity of the native digestion method. This approach may allow for more accurate estimates of protein concentrations. Therefore, the logical next step would be to quantify the identified peptides and evaluate the HCP concentration estimates. Our efforts have shown that native digestion conditions are not necessary for HCP peptide enrichment, and they have opened further questions about the underlying enrichment mechanism of the Huang *et al.* method.

3.4 References

1. Wang, F., Richardson, D., Shameem, M., Host-Cell Protein Measurement and Control. BioPharm Int., 2015. 28: p. 32-38.

2. Hogwood, C.E.M., Bracewell, D.G., Smales, C.M., Host cell protein dynamics in recombinant CHO cells. Bioengineered, 2013. 4: p. 288-291.

3. Abiri, N., Pang, J., Ou, J., Shi, B., Wang, X., Zhang, S., Sun, Y., Yan, D. , Assessment of the immunogenicity of residual host cell protein impurities of OsrHSA. PLOS ONE, 2018. 13.

4. Moussa EM, Panchal JP, Moorthy BS, et al. Immunogenicity of therapeutic protein aggregates. J Pharm Sci 2016; 105: 417-430

5. Jefferis R. Posttranslational modifications and the immunogenicity of biotherapeutics. J Immunology Res 2016; article ID 5358272

6. Feist, P., Hummon, A.B., Proteomic Challenges: Sample Preparation Techniques for Microgram-Quantity Protein Analysis from Biological Samples Int. J. Mol. Sci., 2015. 16: p. 3537-3563.

7. Bracewell, D.G., Francis, R., Smales, C.M., The future of host cell protein (HCP) identification during process development and manufacturing linked to a risk-based management for their control. Biotechnol. Bioeng., 2015. 112: p. 1727-1737.

8. Zhang, Q., Goetz, A.M., Cui, H., Wylie, J., Trimble, S., Hewig, A., Flynn, G.C., Comprehensive tracking of host cell proteins during monoclonal antibody purifications using mass spectrometry. mAbs, 2014. 6: p. 659-670.

9. Doneanu, C., Xenopoulos, A., Fadgen, K., Murphy, J., Skilton, S.J., Prentice, H., Stapels, M., Chem, W., Analysis of host-cell proteins in biotherapeutic proteins by

comprehensive online two-dimensional liquid chromatography/mass spectrometry. mAbs, 2012. 4: p. 24-44.

10. Madsen, J.A., Frauntin, V., Carbeau, T., Wudyka, S., Yin, Y. Smith, S., Anderson, J., Capila, I., Toward the complete characterization of host cell proteins in biotherapeutics via affinity depletions, LC-MS/MS, and multivariate analysis. mAbs, 2015. 7: p. 1128-1137.

11. Doneanu, C.E., Anderson, M., Williams, B.J., Lauber, M.A., Chakraborty, A., Chen, W., Enhanced Detection of Low-Abundance Host Cell Protein Impurities in High-Purity Monoclonal Antibodies Down to 1 ppm Using Ion Mobility Mass Spectrometry Coupled with Multidimensional Liquid Chromatography. Anal. Chem., 2015. 87: p. 10283-10291.

12. Moreland, L.W., et al.Treatment of Rheumatoid Arthritis with a Recombinant Human Tumor Necrosis Factor Receptor (p75)–Fc Fusion Protein. N. Engl. J. Med., 1997. 337: p. 141-147.

13. Huang, L., Wang, N., Mitchell, C.E., Brownlee, T., Maple, S.R., De Felippis, M.R., A Novel Sample Preparation for Shotgun Proteomics Characterization of HCPs in Antibodies. Anal. Chem., 2017. 89: p. 5436-5444.

14. Wang, S., Liu, A.P., Yan, Y., Daly, T.J., Li, N., Characterization of product-related low molecular weight impurities in therapeutic monoclonal antibodies using hydrophilic interaction chromatography coupled with mass spectrometry J. Pharm. Biomed. Anal., 2018. 154: p. 468-475.

15. Wang, Q., Protein-Protein Interaction Analysis: Expanded Hydrogen/Deuterium Exchange Tandem Mass Spectrometry and Host Cell Protein Characterization, in Chemistry. 2019, University of Michigan. p. 172.

16. lpert, A. J. Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. Journal of Chromatography. 1990; (499) 177–196

17. K. Gekko, E. Ohmae, K. Kameyama, T. Takagi. Acetonitrile-protein interactions: amino acid solubility and preferential solvation. Biochim Biophys Acta, 1998; 1387 pp. 195-205

18. Alves, N. J. Antibody conjugation and formulation. Antibody Therapeutics. 2019; 2 (1) p. 33–39

19. McCalley, D. A study of column equilibration time in hydrophilic interaction chromatography. Journal of Chromatography. 2018; 1554, p. 61-70

20. Buszewski, B., Noga, S. Hydrophilic interaction liquid chromatography (HILIC)—a powerful separation technique. Anal Bioanal Chem. 2012; 402, 231–247

21. Alpert, A.J. Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. J. Chrom. 1990; 499:177–196

22. Arbuzova A, Schwarz G. Pore-forming action of mastoparan peptides on liposomes: a quantitative analysis. Biochim Biophys Acta. 1999; 1420(1-2):139-52.

23. Grant E Jr, Beeler TJ, Taylor KM, Gable K, Roseman MA. Mechanism of magainin 2a induced permeabilization of phospholipid vesicles. Biochemistry. 1992; 31(41):9912-8

24. Persson D, Thorén PEG, Herner M, Lincoln P, Nordén B Application of a novel analysis to measure the binding of the membrane-translocating peptide penetratin to negatively charged liposomes. Biochemistry. 2003; 42:421–429

25. Chico DE, Given RL, Miller BT.Binding of cationic cell-permeable peptides to plastic and glass. Peptides. 2003; 24:3–9

26. Duncan MR, Lee JM, Warchol MP. Influence of surfactants upon protein/peptide adsorption to glass and polypropylene. Int J Pharm. 1998; 120:179–188.

27. Joosten HMLJ, Nuñez M. Adsorption of nisin and enterocin 4 to polypropylene and glass surfaces and its prevention by Tween 80. Lett Appl Microbiol 21:389–392.

28. Shukla AA, Hubbard B, Tressel T, Guhan S, Low D. Downstream processing of monoclonal antibodies--application of platform approaches. Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences. Polyclonal and Monoclonal Antibody Production, Purification, Process and Product Analytics. 2007; 848 (1): 28–39

29. Liu HF, Ma J, Winter C, Bayer R. Recovery and purification process development for monoclonal antibody production. mAbs. 2009; 2 (5): 480–99.

30. Schiel, J. E., Davis, D. L., Borisov, O. V., Eds.; State-of-the-Art and Emerging Technologies for Therapeutic Monoclonal Antibody Characterization. Volume 3. Defining the Next Generation of Analytical and Biophysical Techniques; ACS Symposium Series 1202; American Chemical Society, Washington, DC, 2015.

31. Atl N, Zhang TY, Motchnik P, et al. Determination of critical quality attributes for monoclonal antibodies using quality by design principles. Biologicals 2016; 44: 291-305.

32. Vandekerckhove K, Seidl A, Gutka H, et al. Rational selection, criticality assessment, and tiering of quality attributes and test methods for analytical similarity evaluation of biosimilars. AAPS Journal 2018; 20: 68.

33. ICH Q8 (R2) Pharmaceutical development. 2009

34. Fabre, B., Lambour, T., Bouyssié, D., Menneteau, T., Monsarrat, B., Burlet-Schiltz, O. & Bousquet-Dubouch, M.-P. Comparison of label-free quantification methods for the determination of protein complexes subunits stoichiometry. EuPA Open Proteomics 2014; 4, 82–86

Chapter 4

Conclusions and Future Directions

Liquid chromatography-mass spectrometry (LC-MS) remains an essential tool in biomedical research. The work described in this dissertation centers around leveraging this tool to develop 'bottom-up' LC-MS strategies for enhanced detection of labile, non-antigenic post-translational modifications (PTM) and trace protein impurities in biotherapeutic drug substances.

4.1 Direct Identification and Site-Specific Analysis of S- Palmitoylation

The strategies employed in the work described in Chapter 2 evaluated and optimized a liquid chromatography-mass spectrometry (LC-MS) based analytical workflow for the direct detection of S-palmitoylated peptides in complex mixtures. Our work defines sample preparation, chromatographic separation, and MS analysis considerations to limit the loss of labile palmitoylation from cell lyses through LC separation and MS/MS fragmentation (Figure 4.1).

Extraction
1. Lyse cells in pH neutral buffer
2. Extract proteins from membrane fraction with 0.1% Rapigest SF™

Enrichment
1. Maintain buffer pH $\leq$ 7
2. Elution buffers should contain $\geq$0.1% Rapigest SF™
3. Nucleophilic and thioester reactive reductants should be avoided entirely

Digestion
1. Maintain pH $\leq$ 7
2. TCEP is ideal for reduction of disulfides; avoid DTT
3. Digestion buffers should contain 0.5-0.1% Rapigest SF™
4. Limit incubation periods to 4 hours or less
5. Limit all incubation temperatures to $\leq$ 37°C

Liquid Chromatography
1. Before removing surfactant, adjust sample buffer composition to $\geq$ 35% acetonitrile to maintain solubility
2. Remove Rapigest SF™ with 0.1% trifluoroacetic acid at 37°C for 30 min
3. Avoid strong reverse phase columns, such as C18
4. Column temperatures $\geq$ 35°C improve palmitoyl peptide resolution
5. Avoid offline solid-phase filtration
6. For acetonitrile/water gradients, use formic acid as a buffer
7. Gradients should initiate at $\geq$15% acetonitrile if possible, in order to enrich for hydrophobic peptides and to avoid precipitation of any palmitoylated species

Mass Spectrometry
1. Fragmentation conditions should be optimized for peptide(s) of interest, when possible.
2. ETD fragmentation is best suited for palmitoylation analysis of highly enriched samples
3. HCD fragmentation is acceptable for large-scale analyses of complex samples
4. In collisionally induced fragmentation methods, fragmentation potentials should be stepped to achieve best results
5. Analysis software may require additional, vendor-specific parameter optimizations

Figure 4.1 Proposed workflow considerations for site-specific detection of S-palmitoylated proteins by LC-MS analysis

108

Drug Deformulation

- Dilute protein concentration to ≤ 10 g/L
- Adjust buffer composition to 25% acetonitrile in water (0.1%trifluoroacetic acid)
- Remove protein glycosylation if applicable
- Remove residual surfactants from buffer
- Avoid chilling deformulated sample

HILIC Fractionation

- Equilibrate 2.1 x 100 mm BEH glycoprotein amide column
- Initiate gradient at 25% water (0.1%trifluoroacetic acid) to prevent protein precipitation
- Inject 1 mg total protein
- Gradient increases to 45% water over 25 min at 0.5 mL/ min flow rate
- Heat column to ensure good resolution

Fraction Collection

- Use an automated fraction collector
- Collect fractions in a 6x8 plate with 5mL wells
- Collect 3 fractions:
    Fraction 1: 1.5- 11.5 min
    Fraction 2: 11.5-18 min
    Fraction 3: 18 - 25 min
- Avoid chilling collected fractions

Proteolytic Digestion

- Add 5 µL of MassPrep yeast enolase digest standard (1 pmol / µL)
- Add 2 µL of DTT (50 mg/mL) to each well
- Incubate plate at 60°C for 30 min
- Cool to 37°C before addition of 5 µg of Trypsin/Lys-C mix to each well
- Cover plate and incubate for 18 h at 37°C
- Transfer 200µL of each sample to HPLC vials containing 5 µL of 10% formic acid in water

Sample Reconstitution

- Reconstitute samples in 0.2 mL aqueous digestion buffer containing:
    Urea (2.0 M)
    Tris pH 8.0 (0.1 M)
- Ensure complete dissolution of proteins by mixing thoroughly and plate agitation for 5 min

Sample Concentration

- Neutralize fractions with a 10% ammonium hydroxide solution in water
- Concentrate fractions under nitrogen flow
- Avoid heating samples to ≥ 45°C
- Monitor pH during buffer concentration to avoid buffer acidification and protein precipitation
- Avoid completely drying fractions by ending concentration when fractions reach between 25-50 µL total volume

LC-MS Analysis

- Equilibrate 2.1 x 100 mm CSH C18 reverse-phase column at 60°C
- Inject 200 µL of sample
- Initiate gradient at 100% water (0.1%formic acid)
- Gradient increases to 45% acetonitrile over 60 min at 0.3 mL/ min flow rate
- MS/MS acquisition with ion trap at minimum 60K MS^1 resolution
- Set HCD fragmentation at 32% collision energy

Protein Identification

- Analyze with Byonic™(ProteinMetrics) software
- Search tolerances set to:
    10ppm for precursors
    0.4 Da for fragments
    1% FDR
- Filtering criteria:
    |LogProb| ≥ 3.00
    Unique peptides ≥ 3
- Manually validate any proteins scored lower than top decoy hits

HCP Quantitation

- Analyze Byonic™ search results with Progenesis QI™ (Waters)
- Absolute quantitation with Top-3 peptides analysis
- Enolase set as internal standard at 5000 fmol
- Peptide filtering conditions:
    Sequence length ≥ 6 amino acids
    Unique peptides ≥ 3
    |LogProb |≥ 3.00
- Resolve conflicting peptide grouping assignments by attributing all peptide identifications to highest scored protein group.
- Convert measured protein molarities results output to ppm values.

Figure 4.2 Proposed workflow for HCP detection and quantitation of drug substances using offline HILIC fractionation and LC-MS analysis

109

Each of these steps and conditions were evaluated and optimized to prevent hydrolysis of the S-palmitoyl thioester bond and to maintain the solubility of lipidated or otherwise hydrophobic peptides in sample solutions so that they can be analyzed by LC-MS. Figure 4.1 presents important conditions that should be considered by future researchers seeking to analyze S-palmitoylation with 'bottom-up' LC-MS/MS workflows. As proof of concept, we applied these method considerations and achieved the novel, direct detection of recombinant N-Ras palmitoylation in a complex mixture.

As the importance of the regulation of S-palmitoylation is continually realized, directly detecting these labile modifications in large scale studies of palmitoylated protein populations by tandem MS will be an available option. Future studies should expand on the foundation built in this work and seek to further optimize the workflow conditions and parameters to globally profile lipidation in a proteome. Several challenges can be expected when transitioning to proteome wide studies such as the low abundance of this PTM in most tissues. Generally, enrichment techniques are used to reduce background interference during analysis and to increase signal of low abundant species. New strategies will need to be employed to overcome this challenge while taking into consideration the labile and non-antigenic nature of this PTM. Our strategy was to leverage the hydrophobicity of the greasy acyl-chain and modify our chromatography to, in effect, wash the more hydrophilic background away. While this strategy works for single peptide analyses, analyses requiring high sequence coverage of proteins of interest and proteome-wide studies will need to develop additional strategies to improve signal/noise ratios for low abundance palmitoylated peptides in complex mixtures.

LC-MS/MS analyses of mouse brain tissues, which are known to contain higher levels of palmitoylation than other tissues, should be considered when studying dynamic palmitoylation

conditions or to directly detect novel sites of palmitoylation. Additionally, our workflow can serve as an outline for validating reported palmitoylation sites catalogued by the SwissPalm database with direct LC-MS analysis. Similarly, elements of the workflow strategies detailed in Chapter 2, such as using C-8 chromatography, may be used in future studies of the hydrophobic proteome which may advance our understanding of membrane protein dynamics, interactions, and structure.

The labile nature of palmitoylation will remain a challenge in direct detection of palmitoylation sites. Electron transfer dissociation (ETD) provided the highest sequence coverage for palmitoylated peptides with minimum neutral palmitate loss and, therefore, should be considered in future analyses. As instrumentation capable of ETD become more readily available, methodologies utilizing this technology may become more practical for general research purposes, however limitations do exist which must be overcome. Importantly, ETD adds a negative charge to precursor ions and therefore requires high charge states for peptides of interest. This charge state requirement can be a problem for complex samples where matrix analytes compete for charge during electrospray ionization (ESI) or for lipidated peptides which tend to exhibit poor ionization efficiency. One strategy that should be explored is post-column introduction of supercharging reagents to effectively increase the average charge of peptide ions and thus increase the effectiveness of ETD fragmentation. This strategy could provide an effective alternative to our HCD step-fragmentation method, detailed in Chapter 2, and increase global palmitoylated peptide fragmentation coverage. Overall, this work provides a foundation for directly detecting S-palmitoylated peptides and serves to explore orthogonal strategies to conventional indirect detection methods.

4.2 Improving Detection Sensitivity for Host-Cell Protein Characterization

The offline fractionation method and plate-based workflow, described in Chapter 3, serves as a sensitive LC-MS-based host cell protein (HCP) characterization platform. This platform can serve to support downstream analysis of HCP content in drug substances (DS) across multiple biotherapeutic protein classes. Furthermore, our proposed workflow limits most sample specific optimizations, achieves ppm level sensitivity for HCP detection and absolute measurements of residual HCP impurities. Our modifications of the Wang *et al.* method included implementation of automated fraction collection and a single-plate sample processing workflow that reduces HCP loss from multiple vial transfer steps and significantly reduces parameter optimization time required between samples, thereby improving the feasibility of this method for adoption in industrial process development. Additionally, our plate-based method improves processing throughput where up to 16 samples can be fractionated in sequentially for multiple sample analyses. Figure 4.2 presents a diagram outlining our proposed workflow for HCP characterization for use in downstream critical quality attribute analysis of biotherapeutic drug substances.

Although our method reliably measured higher concentrations of HCPs when compared to other LC-MS strategies, our method did not provide the robustness between runs that would be expected of a standardized workflow. Future experiments will need to be carried out to determine the root cause of variance between replicates in both the HILIC and native digestion methods. Future work should also examine the properties of specific HCPs detected by different methods/conditions to elucidate why some HCPs are poorly recovered by one method as opposed to the other, and compare HCP identifications between replicates to determine the reproducibility of each unique identification. Such experiments will serve to justify the use of one method over another comparable method and potentially explain causes underlying the

discrepancies between runs. Furthermore, it is important to note that our work failed to provide a benchmark for evaluating the absolute quantitative accuracy of each LC-MS method, and by comparing each method's HCP measurements with those of an ELISA will help support our method's quantitative accuracy and could potentially validate our results. Another complementary approach for validating our method's quantitative accuracy would be to measure the levels of multiple spiked proteins at known concentrations. Such experiments would help identify, with improved precision, the accuracy of our results.

Dynamic range is an important consideration when using ESI based ionization for LC-MS analyses and will continue to be problematic as higher purities of biotherapeutic proteins are achieved. Methods such as 2D-chromatogprahy and sample fractionation allow for separation of the trace impurities from the drug substance and improve the dynamic range of detection. However, certain limitations with these strategies exist, including the increased analysis time and, consequentially, lower throughput compared to other existing bioanalytical methods. Our workflow sought to mitigate these issues, specific to LC-MS analysis strategies but is still incapable of the level of parallelizability inherent with other bioanalytical techniques, such as the Enzyme-linked Immunosorbent assay (ELISA). Future work should seek to improve this workflow to further increase throughput and reduce analytical workload. A higher degree of automation, such as implementing software for fraction collection that automatically detects UV chromatographic peak formation during fraction collection instead of manually optimizing time-based fraction window parameters, could drastically improve throughput. Additionally, enhanced automation would eliminate sample specific optimization still needed in our HILIC fractionation step and serve to further generalize the application of this method as a platform for HCP characterization.

Future advancements will undoubtedly turn towards improving the limits of detection and quantitation to the ppb concentration range and will most likely require concentrating larger quantities of HCPs during the fractionation step. To achieve this goal, using a larger diameter HILIC column, such as a 4.6 mm inner diameter (ID) column compared to the 2.1 mm ID column employed in Chapter 3, will be necessary for larger volume injections.

While identifying and quantifying HCPs in biotherapeutics is demonstrably important, efforts are currently underway to improve the downstream purification process to further deplete these HCPs from the drug substance and thereby improving the overall purity of biologics products. LC-MS analysis can aid this endeavor, and again demonstrate the analytical capabilities exclusive to mass spectrometry, by identifying and characterizing specific proteins that evade purification. Identifying the HCPs that are consistently and reproducibly present between samples may help elucidate the underlying cause of HCP contamination and provide helpful information for process development in biotherapeutic manufacturing.

LC-MS based protein analysis will remain an essential tool in protein analysis for years to come. As computational algorithms and instrumentation evolve to match current analytical demands, novel strategies will need to be developed to fully capitalize on the detection capabilities and the analytical power inherent in this technique. Our methods add to the repertoire of analytical capabilities that LC-MS systems provide, and further illustrate the power of this technique for protein analysis. By applying the concepts and considerations outlined in the preceding chapters, new challenges can be addressed that will definitively advance the field of LC-MS based protein analysis.