Improving Traveling Wave Ion Mobility Mass Spectrometry for Proteomics

by

Sarah E. Haynes

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemistry)
in the University of Michigan
2019

Doctoral Committee:

Assistant Professor Brent R. Martin, Chair
Professor Philip C. Andrews
Professor Alexey I. Nesvizhskii
Professor Brandon T. Ruotolo

Sarah E. Haynes

hayse@umich.edu

ORCID iD: 0000-0003-3225-1691

© Sarah E. Haynes 2019

Dedication

To my wonderful teachers who have made these 20+ years of school fly by.

Acknowledgements

I would like to thank my dissertation committee, Brent, Phil, Brandon, and Alexey, for their invaluable help and guidance. Brent provided me with the opportunities to work on the interesting projects described here, and forged additional collaborations that I was fortunate to be a part of. I am grateful to Bob Kennedy and Kicki Håkansson, who provided assistance on a number of occasions. I was fortunate to receive funding from the NIH Chemistry-Biology Interface Training Program. I would also like to thank Liz Oxford, Katie Foster, and Heather Hanosh, some of the best support staff we could ask for.

I am grateful to Jaimeen Majmudar, who was my mentor when I first joined the Martin Lab, for his great skill and kindness. Dmitriy Avtonomov developed the software described in Chapter 4 in its entirety. Dan Polasky and Sugyan Dixit were instrumental in the work described in Chapter 2, but in truth they helped throughout.

I was lucky to find a wonderful friend and supporter in Melanie Cheungseekit, my labmate and deskmate for the past few years. I would like to thank my lovely mother in law, Mary, for helping me get started with Python. My grandmother, Sally, paved the way for my undergraduate and graduate studies, and I am eternally grateful for her help and encouragement. My parents, Mark and Caroline, and my sister Ginny have been incredibly supportive and I can't believe my luck at being part of this family.

My husband Paul embarked on this PhD journey with me, and has been a true partner throughout. I am so fortunate to have him in my life.

Table of Contents

List of Figures

List of Tables

Abstract

Large-scale analysis of proteins is a critical tool in the life sciences, guiding drug development and elucidating important cellular processes. These measurements are accomplished with liquid chromatography-mass spectrometry (LC-MS), where proteins are enzymatically digested into peptides, separated via liquid chromatography, and analyzed by mass spectrometry.

Typically, the most abundant peptide ion at a given time is selected for sequence identification, but limited instrument scan speed often results in under-sampling, compromising data completeness and reproducibility. In contrast, all-ion acquisition methods bypass peptide ion selection, measuring peptides ions across broad mass ranges. Despite capturing data for all events, peptide annotation is limited by inadequate separation prior to fragmentation, which results in interfering peaks in fragment spectra. Ion mobility spectrometry (IMS), where peptide ions are separated based on cross-sectional size and charge in the gas phase, adds an orthogonal analytical dimension to LC-MS proteomics. In-line ion mobility spectrometry provides additional separation without increasing analysis times, reducing spectral interference to improve reproducibility, peak capacity, and peptide identifications.

However, these ion mobility separations were not optimized for complex peptide mixtures, and the true peak capacity of the LC-IMS-MS platform was unknown. Additionally, fragment ion information acquired during protein quantification experiments using stable isotope labeling by amino acids in cell culture (SILAC) was underutilized. Rigidity and opacity of

proprietary data analysis software also presents a barrier to improving LC-IMS-MS proteomics measurements.

Chapter 1 presents the first quantitative characterization of traveling wave ion mobility separation in the context of real, complex proteomics samples. Taking into account the orthogonality of the LC, IMS, and MS separations, we found that IMS doubles the peak capacity of LC-MS alone under standard traveling wave settings. Seeking to improve the IMS separation, we discovered IMS settings that reproducibly increased peptide and protein identifications by over 40%. Chapter 2 describes a protein-centric statistical filtering method to leverage fragment ion quantification information. This filtering method reduces coefficients of variation by 4-fold, increasing confidence in differential protein measurements. Chapter 3 explores a new LC-IMS-MS software tool, focusing on 3D peak detection parameters, and reports the first database searches of LC-IMS-MS data performed entirely with free, open-source tools.

Chapter 1 Introduction

Genome sequencing can provide important clues about the inner workings of the cell, but DNA is merely the template for the diverse set of proteins that actually perform biological functions. These proteins, sequences of amino acids of that fold to assume distinct three-dimensional structures, function to catalyze chemical reactions, transfer signals by associating with other proteins, and provide structural support within the cell. Almost all drugs target proteins due to their active role in controlling biological function. To understand disease or answer other types of important biological questions, it is crucial to study proteins directly.

While there are about 20,000 protein-coding human genes,[1, 2] alternative splicing of RNA causes these genes to be translated into a total of about 70,000 different amino acid sequences.[2] Additionally, these proteins are frequently post-translationally modified by the addition of various molecules, such as a phosphoryl group or fatty acid, which can strongly influence their activity. Changes in the abundance of a protein or the way it is modified can have significant impacts on cellular function, and it is important for researchers to be able to detect these variations.

To measure proteins between experimental conditions, such as different disease states, mutations, or drug treatments, proteins must be identified and quantified. In the 1970s, two-dimensional polyacrylamide gel electrophoresis enabled researchers to separate and visualize about 1000 distinct proteins from cells.[3-5] In these experiments, protein extracts were separated in one dimension by isoelectric point and in the other by mass, such that each protein type

occupied one spot in a two-dimensional gel. To visualize the spots, the gels were stained with dyes that specifically bind to proteins or the gels were imaged with autoradiography (**Figure 1**).

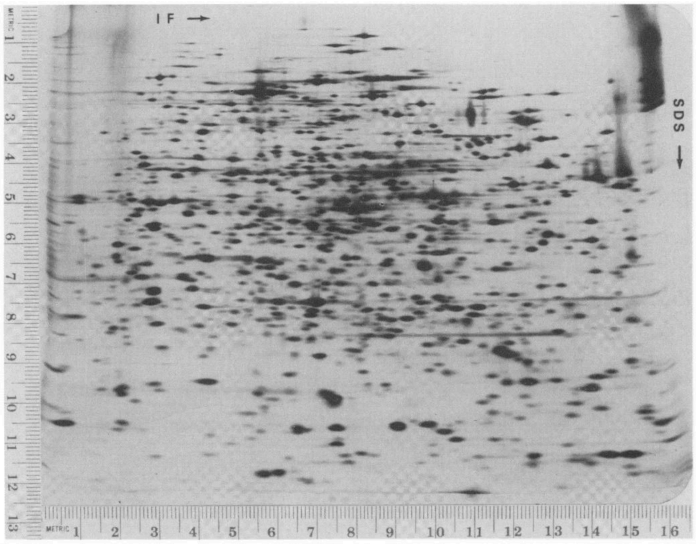

Figure 1. The first proteome-wide measurements. E. coli proteins are separated by isoelectric point (left to right) and molecular weight (decreasing from top to bottom). *This research was originally published in the Journal of Biological Chemistry. P.H. O'Farrell. High resolution two-dimensional electrophoresis of proteins. J. Biol. Chem. 1975; 250(10):4007-4021. © The American Society for Biochemistry and Molecular Biology.*

These were the first 'proteomics' studies because they demonstrated the ability to observe the abundance of many proteins at once. However, only the most abundant proteins could be seen, and determining the identity of an individual protein spot was challenging. This often required detection with a specific antibody, the addition of a known internal protein standard for comparison, or identification by additional chromatography experiments to map radiolabeled pieces of the proteins to images from previous experiments.[4-9] The amino acid sequence of a single protein spot could also be determined in a semi-automated manner using a combination of Edman degradation (sequential release of amino acids) and chromatography.[10, 11]

In the late 1980s and early 1990s, the use of mass spectrometry to analyze protein extracts significantly accelerated the process of sequencing and identifying proteins.[12, 13] The first experiments during this time period used ionization techniques such as secondary ion mass

spectrometry and fast atom bombardment to energize proteins immobilized on a prepared surface, charging the protein molecules so that they could be analyzed by mass spectrometry.

There are a number of different types of mass analyzers that can be used, but the basis of all types of mass spectrometry is the separation of ionized atoms or molecules by their mass-to-charge ratio. This is most often accomplished by controlled electric fields within the instrument, and enables rapid, high-resolution measurements of charged molecules. Once they have been separated by their mass-to-charge (m/z) ratios, the charged molecules are detected by the instrument. The strength of the signal detected for each m/z ratio can be used to quantify the amount of a molecule of interest, a crucial feature for proteomics studies.

The application of electrospray ionization to mass spectrometry for protein analysis promised even higher throughput.[14-17] In electrospray ionization (**Figure 2**), a solution containing the protein (or other analyte) of interest flows out of a high voltage (a few kV) needle or capillary.

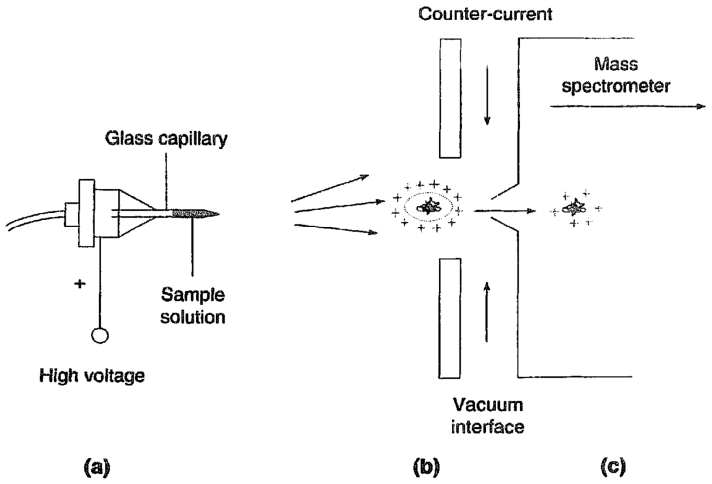

Figure 2. Electrospray ionization of proteins coupled to mass spectrometry. (a) Charged droplets emanate from the capillary in a strong electric field. (b) Liquid evaporates from ionized protein droplets (assisted in this image by a flow of dry N_2 gas). (c) Protein ions enter the mass spectrometer. *Reprinted from Trends in Biochemical Sciences, 20(6), Mann, M. & Wilm, M., Electrospray mass spectrometry for protein characterization, 219-224, Copyright (1995), with permission from Elsevier.*

The electric field at the needle tip charges the emerging liquid, dispersing it into a fine spray of charged droplets via Coulomb forces. Through a combination of splitting and desolvation (drying), the droplet spray results in charged protein molecules with most or all of the initial solvent removed. This process is gentle, introducing charged molecules into the gas phase without breaking or fragmenting them.[14, 18, 19] Electrospray also introduced the possibility of imparting multiple charges to intact protein molecules, which enabled the analysis of larger biomolecules and more advanced manipulation within the mass spectrometer.

Electrospray ionization paired naturally with small-scale capillary separations of liquid protein samples, further increasing the specificity and throughput of protein analyses via mass spectrometry. The miniaturization of the electrospray apparatus, with flow rates of tens of nL per minute and capillary tip diameters of about 1 μm further improved sensitivity, allowing the detection of very small (femtomolar) amounts of proteins.[20, 21]

These studies worked well for analyzing purified proteins, but the detected signals of thousands of proteins from complex cell extracts would overlap, making it difficult to measure each protein. Separation prior to mass analysis is crucial to identifying and quantifying individual proteins. Liquid chromatography at flow rates amenable to miniaturized electrospray ionization[22-24] is a powerful separation technique that can be added in-line prior to mass spectrometry, but mixtures of intact proteins are difficult to separate using this technique due to their widely varying sizes (ranging from about 1 to 1000 kDa, or about 10 to 1000 amino acid residues). To address this, proteins can be enzymatically cut into similarly sized pieces (usually from about 500 to 2500 Da, or about 5 to 25 amino acid residues) that can be efficiently separated by in-line liquid chromatography.

By the late 1990s and early 2000s, proteomic measurements were made by enzymatic cleavage, or digestion, of protein samples to create peptides, which were then separated with liquid chromatography and analyzed with electrospray ionization-mass spectrometry.[25-27] In addition to using mass spectrometry to determine the mass-to-charge ratios of the ionized peptides, the peptide ions can be energized to break them apart, generating fragment ions that can be used to find the amino acid sequence of the originating peptide.[28-30] This process of protein digestion, separation via liquid chromatography, and mass analysis with fragmentation is called 'bottom-up' or 'shotgun' proteomics (**Figure 3**), and is widely used today.

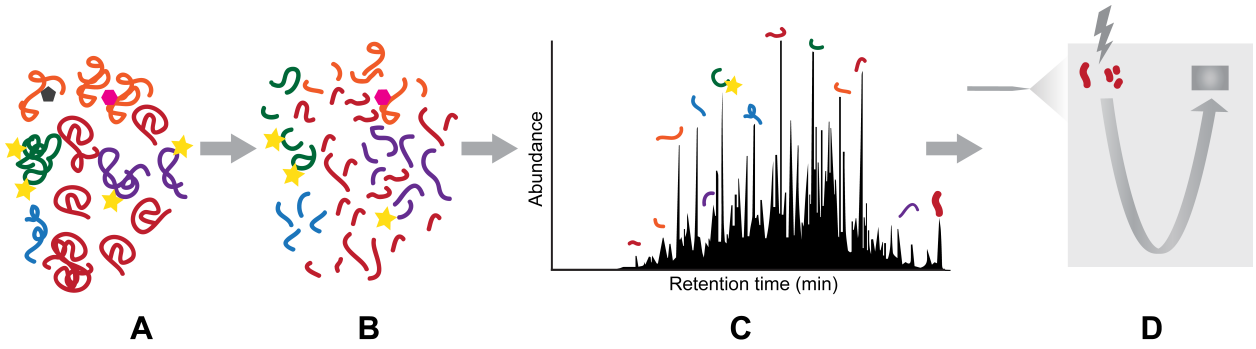

Figure 3. Workflow of a typical proteomics experiment. Protein extracts from cells or tissue samples (**a**) are enzymatically digested into peptides (**b**). These peptides are separated via liquid chromatography (**c**). As peptides elute from the chromatography column, they are electrosprayed into a mass spectrometer (**d**). Fragment energy is also applied to measure the m/z values of peptide ion fragments for sequence determination (also **d**).

To identify and quantify proteins from laboratory cultures or tissue samples, cells are lysed (broken open) to release their aqueous contents, which include fatty acids, salts, and metabolites in addition to proteins. A proteolytic enzyme, usually trypsin, which specifically cleaves the proteins at the amino acids arginine and lysine, is added to digest the proteins in the lysate solution (**Figure 3a,b**), usually overnight. Fatty acids, salts, metabolites and other non-protein components of the digested lysate are then removed from the sample.

The cleaned-up peptide mixture is then placed into an autosampler, which injects the sample onto a narrow-bore capillary column (50-100 μm inner diameter) that is packed with small particles (1-5 μm diameter) of reversed-phase[31] resin. The most popular resin for proteomics is C18 resin, which is made up of silica spheres covered with hydrocarbon chains, each of which contains 18 carbons. Because peptides are slightly hydrophobic, they mostly partition out of the surrounding water into interactions with the hydrocarbon chains of the resin. Over the course of the separation, an automatic pump gradually adds organic solvent to the column. As the percentage of the organic solvent surrounding the resin increases, the peptides partition out of the resin to rejoin the flow of liquid through the column.[32, 33]

Separation occurs as different peptides rejoin the flow at rates depending on their hydrophobicity (determined by the amino acid composition) and size.[34, 35] As a result, smaller, more hydrophilic peptides will elute, or emerge, from the column sooner than larger, more hydrophobic peptides (**Figure 3c**). Later eluting peptides have larger 'retention times', since they are retained by the column longer. The length of this separation of is generally between one and three hours for most liquid chromatography-mass spectrometry (LC-MS) proteomics experiments.

As peptides elute from the column over the course of the LC-MS analysis, they are ionized and introduced into the mass spectrometer via electrospray (**Figure 3c**). The mass spectrometer scans the entire peptide mass range (usually between 100 to 2000 m/z) and determines the most abundant peptide ions in the scan. A multipole mass analyzer (such as a quadrupole[36, 37]) is then used to select the most abundant peptide ions, which are fragmented for sequence determination. The most commonly used fragmentation technique for peptides is collision-induced dissociation, which uses electric fields to accelerate peptide ions into a heavy

neutral gas (such as argon), transferring kinetic energy to raise the internal energy of the peptide ion until the bonds between amino acid residues are broken.[13, 38, 39] This process of gaining further information about an ion via fragmentation is commonly called 'tandem mass spectrometry' or 'MS2', referring to the second mass spectrum measurement.

The mass spectrometer records the resulting m/z spectra from all intact peptide ions and all MS2 spectra. Once the LC-MS acquisition is finished, the data can be processed and searched against a sequence database. Because peptides contain nitrogen and carbon, a population of peptides has a statistical distribution of masses due to the natural abundance of ^{13}C and ^{15}N.[40] An example of this distribution can be seen in **Figure 4**, which shows the detected signal from a single tryptic peptide.

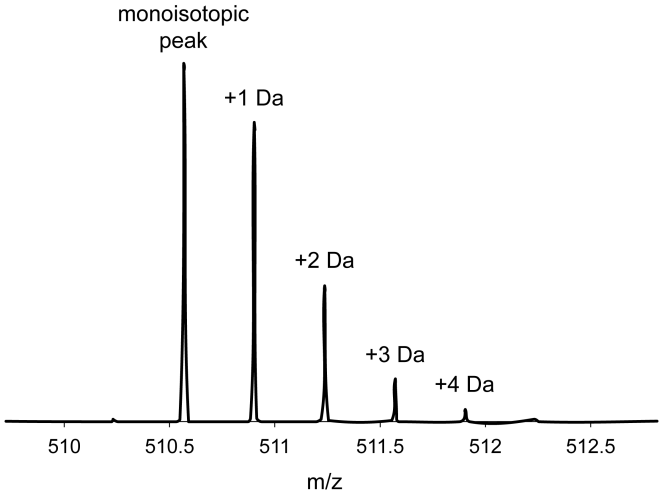

Figure 4. Mass-to-charge pattern of a single tryptic peptide. A high resolution mass spectrum of a tryptic peptide. The far left monoisotopic peak represents the population of the peptide that contains no heavy isotopes (^{13}C or ^{15}N). The next peak represents the population of the peptides that contain one ^{13}C or ^{15}N atom (adding 1 Dalton to the mass of the peptide). Data is from a HEK293 cell digest measured with an Orbitrap Fusion Lumos mass spectrometer.

These m/z distributions can be used to determine the charge and molecular mass of every peptide and fragment feature recorded by the instrument. Fortunately, peptide ions fragmented by collision-induced dissociation break in a predictable and reproducible manner between amino

acid residues. If fragmentation occurs between all or most of the amino acids in the peptide, the sequence can be confidently determined by matching against a database of known protein sequences.

Databases of all known protein sequences from an organism can be downloaded from a public repository such as Uniprot.[1] To match protein sequences in the database to the peptide and fragment masses catalogued in the LC-MS experiment, each protein sequence in the database is subjected to a trypsin digest *in silico*, and each *in silico* peptide is then fragmented to generate a theoretical fragment spectrum.

The experimental spectra are then compared to the theoretical spectra, and similarity scores are assigned for each comparison.[28-30] The peptide sequence corresponding to the highest scoring theoretical spectrum is stored. From the list of peptide identifications, the proteins in the original sample are inferred.[41] To validate whether the matched peptides and inferred proteins are correct identifications, each match is evaluated against all other matches in the experiment, as well as some matches that are necessarily false (e.g. to theoretical spectra from reversed protein sequences). Statistical modeling of the distribution of scores from all of these matches is used to determine a minimum score cutoff for true identifications, helping ensure that the final list of peptides and proteins reported by the software can be trusted.[42-44]

Identification of peptides and proteins from LC-MS was originally done by hand, but is now almost always performed computationally. Depending on the software used, the process of obtaining validated lists of peptide and protein identifications from a single raw LC-MS data file can be completed in just a few minutes.[45]

During the computational analysis of the raw LC-MS data, the measured abundance of each peptide ion can be stored. The amount of each identified protein in the originating cell or

tissue sample can be quantified from these ion abundances. This empowers researchers with the ability to assess large-scale differential protein abundances between experimental conditions, helping answer important questions about biological function or the mechanisms of disease.

Despite the sensitivity and selectivity of modern LC-MS proteomics technologies, the sheer complexity of the proteome presents a formidable barrier to identifying and quantifying the entire set of proteins in a whole cell lysate. Tryptic digestion increases the complexity even more. For example, from the 20,422 reviewed human proteins in the Uniprot database, a tryptic digest yields 604,246 peptides between 5 and 50 amino acids in length (the range generally detectable by LC-MS methods), about a 30-fold increase in complexity.

While many of these proteins (and peptides) are present in such low copy numbers in the cell that they are effectively unobservable by LC-MS, many of these tryptic peptides are abundant enough to be detected. Consequently, there are numerous detectable peptides similar in hydrophobicity and mass, and they elute from the chromatography column at the same time. When one peptide ion is mass-selected for fragmentation, there are often other co-eluting peptide ions with similar m/z values that are selected and fragmented as well. This co-fragmentation phenomenon is exacerbated by poor chromatographic separation and imprecise quadrupole isolation.

Insufficient separation prior to mass analysis leads to complex fragment ion spectra that contain products of multiple peptide ions. One in-depth study of human cell lysate found that signal from the target peptide ion was usually less than 50% of the ion intensity within the selection window.[46] Contaminating peaks in the spectra lead to lower database search scores, increasing the rate of false negatives and reducing the number of peptides that can be identified from the sample.[47] Additionally, co-fragmentation can skew protein quantification.

Ion mobility spectrometry has been added to the LC-MS workflow, introducing an additional dimension of separation to improve proteomics data acquisition. A post-ionization separation in the gas phase, ion mobility spectrometry disperses ions by their collisional cross sections (**Figure 5**).

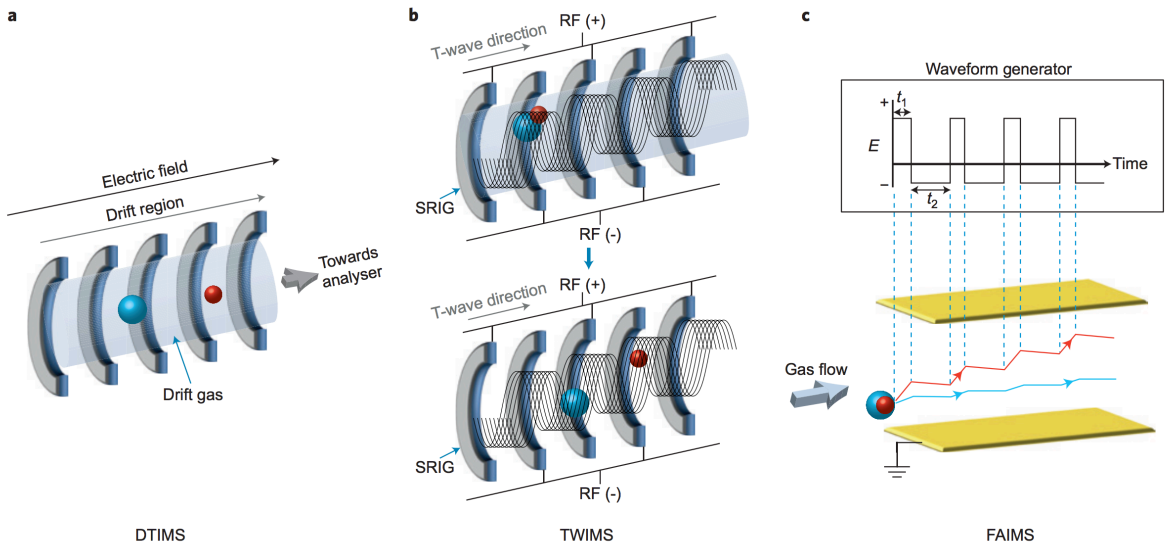

Figure 5. Three popular modes of ion mobility spectrometry. Ions with small collisional cross sections are in red, large ions are in blue. (a) Drift-tube ion mobility passes ions through a drift gas along the axis of an applied electric field. The red ion undergoes fewer interactions with the drift gas, and traverses the device faster than the blue ion. (b) Traveling wave ion mobility separation occurs as a travelling voltage wave is applied to a series stacked ring ion guides (SRIGs). Small ions are pushed forward by the wave, while large ions roll over the wave, taking longer to exit the device. (c) Field-asymmetric ion mobility employs an alternating, asymmetric electric field, causing ions to drift towards two electrodes (yellow) at different rates, separating ions spatially. *Reprinted by permission from: Springer Nature, Nature Chemistry, "The power if ion mobility mass spectrometry for structural characterization and the study of conformational dynamics", F. Lanucara, S.W. Holman, C.J. Gray, C.E. Eyers. Copyright 2014.*

Because the gas-phase cross-section of a peptide ion does not perfectly correlate with retention time and mass-to-charge, ion mobility brings additional resolving power to LC and MS analysis. There are a number of different modes of ion mobility spectrometry, all of which operate on the principle that the size-to-charge ratio of an ion influences the time it takes to traverse through a buffer gas in the presence of an electric field.

The first ion mobility measurements were made as early as the late 1800s, but the technique was first coupled to mass spectrometry in the 1960s (**Figure 6**).[48, 49]

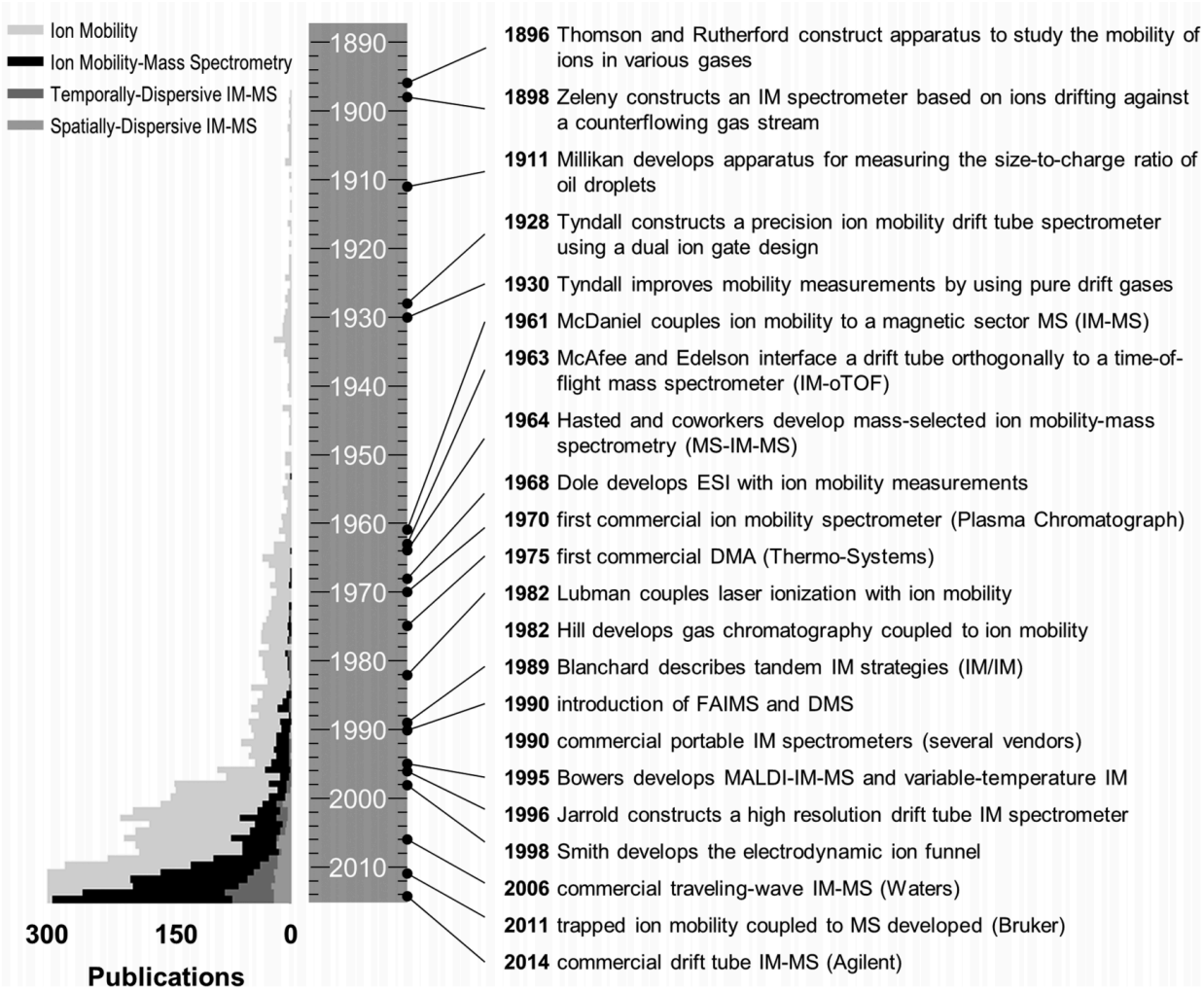

Figure 6. Historical timeline of ion mobility-mass spectrometry. The number of peer-reviewed publications in ion mobility and ion mobility-mass spectrometry is plotted against a timeline of important developments in ion mobility and ion mobility-mass spectrometry instrumentation. *Reproduced with permission from May, J. C., & McLean, J. A. (2015). Ion mobility-mass spectrometry: time-dispersive instrumentation. Analytical chemistry, 87(3), 1422-1436. Direct link: https://pubs.acs.org/doi/full/10.1021/ac504720m. Further permissions related to the material excerpted should be directed to the ACS*

Ion mobility-mass spectrometry (IMS-MS) technologies have enabled high-throughput, sensitive and selective measurements of biomolecules.[50-55] Many of these measurements were made in IMS-MS prototypes constructed in research laboratories, but there are now three major

commercially available IMS-MS instruments. These are the Waters traveling wave ion mobility-time of flight mass spectrometer,[56] Bruker trapped ion mobility-time of flight mass spectrometer,[57] and Agilent drift-tube ion mobility-time of flight mass spectrometer.[58] The first of these, introduced in 2006, employs traveling wave ion mobility spectrometry (**Figure 5b**), which disperses ions based on their ability to 'surf' on a traveling electric field.

The traveling wave device is a series of ring ion guides filled with a neutral gas, usually a mixture of He and N_2. Ions are confined along the axis of separation by a radiofrequency field, while a repeating pattern of DC voltages is sequentially applied to the stacked ring ion guides, creating a wave that travels at speeds of hundreds of meters per second.[56, 59, 60] As they experience fewer collisions with the buffer gas, ions with smaller collisional cross sections (higher mobilities) are pushed further by the voltage wave before rolling over behind it. Consequently, lower-mobility ions roll over behind the wave more frequently and take longer to transit the device.[61]

In the commercial traveling wave-ion mobility mass spectrometer, the Waters Synapt G2-S, ions are introduced by electrospray, and neutral species are filtered out using a stepped ion guide (**Figure 7**). A quadrupole acts to guide ions to a trapping region, where they are accumulated then released into the traveling wave ion mobility device. As ions exit the device, they are transferred into an orthogonal acceleration[62] time-of-flight mass analyzer.[63, 64]

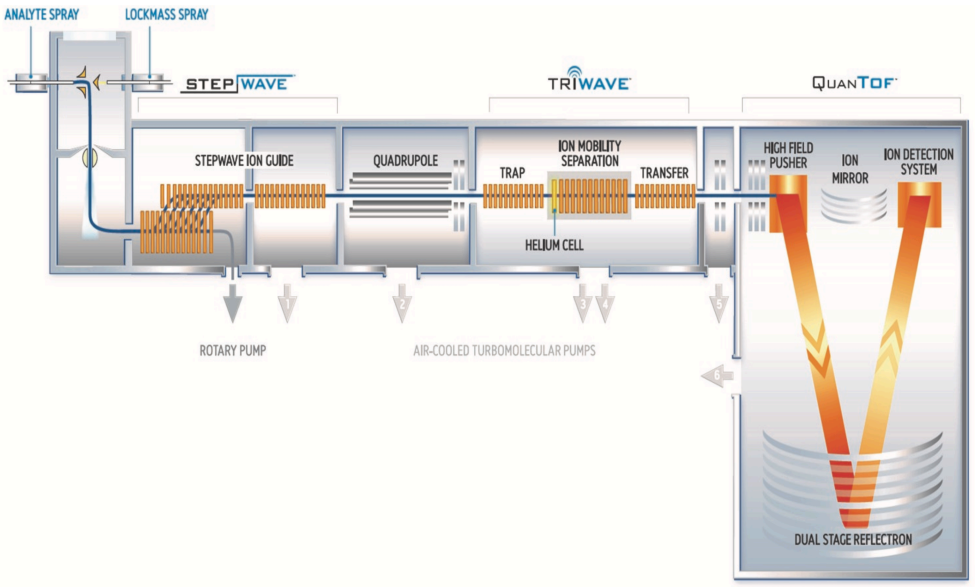

Figure 7. Schematic of the traveling wave ion mobility-time of flight instrument (Synapt G2-S). After being separated by the traveling wave device ("Triwave"), m/z values are measured with the time-of-flight mass analyzer ("QuanTOF"). Copyright 2011, Waters Corporation.

In the Synapt, intact peptide ions are fragmented after ion mobility separation in the transfer region. Because of this, collision voltages can be ramped up throughout the ion mobility separation, such that higher energies can be applied to larger peptide ions with later drift times. Drift time-specific collision energy can significantly improve the efficiency of peptide fragmentation, and thus greater signal-to-noise ratios of peaks in the resulting fragment spectra.[65] The application of drift time-specific collision energies alone has been shown to improve protein identifications by over 40%. The combination of ion mobility separation and drift time-specific collision energy improves proteome coverage by over 120% compared to the same acquisition strategy with no ion mobility separation.[65]

In addition to increased fragmentation efficiency, such improvements in peptide and protein identifications can largely be attributed to the selectivity afforded by the ion mobility separation. Indeed, the additional dimension of separation has been shown to decrease ambiguity

in peptide-fragment matching and significantly increase the number of fragment peaks that can be associated with their originating peptides, improving database search scores.[66]

The Synapt has been used to make a number of important proteome measurements, including the correlation of lipid post-translational modification with the expression of an oncogenic transcription factor,[67] protein expression profiling of different strains of the human cytomegalovirus,[68] and the identification of biomarkers to distinguish between pancreatic cancer and pancreatitis.[69]

The motivation of the work described in this thesis document is to benchmark, study, and improve the platform for proteomics experiments, with a particular focus on whole proteome measurements. Specifically, this thesis explores the true resolving power of the three-dimensional LC-IMS-MS system for highly complex mixtures of tryptic peptides and introduces a statistical filtering method to enhance quantitative proteome measurements. This work also reports attempts to develop flexible, open-source tools to help interrogate LC-IMS-MS data for analytical method development and routine proteomics experiments.

These findings can be applied to ion mobility-enhanced proteomics broadly, as they provide a framework for optimization of ion mobility parameters for peptide identification, a detailed method for measuring and tracking peak capacity during method development (work previously published as **Haynes, S.E.; Polasky, D.A.; Dixit, S.M.; Majmudar, J.D.; Neeson, K.; Ruotolo, B.T.; Martin, B.R. Anal. Chem. 2017, 89(11), 5669-5672**), a means to leverage all identified peaks to improve data-independent stable isotope quantification (work previously published as **Haynes, S.E.; Majmudar, J.D.; Martin, B.R. Anal. Chem. 2018, 90(15), 8722-8726**), and an approach for tuning software to improve performance of feature detection from multidimensional LC-IMS-MS data.

Chapter 2 Quantification of LC-IMS-MS resolving power for complex proteomics

Introduction

One of the primary challenges to acquiring complete and quantitative proteome measurements is the enormous complexity of the protein digests. Efficient separation of tryptic peptides via liquid chromatography prior to electrospray ionization is the first key step to tackling this complexity. Once peptide ions have entered the mass spectrometer, they must be separated from one another and fragmented individually to provide sequence information for each. As discussed in Chapter 1, the original approach to separating peptides within the instrument is to sequentially isolate and fragment only those peptides within a narrow mass-to-charge window.[70, 71] While this selection process allows the instrument to distinguish between co-eluting peptides of different masses, the instrument only has time to isolate and fragment peptide ions with the highest abundance. As a result, the instrument is unable to acquire high-quality spectra for many of the peptide ions, leading to a stochastic selection process that compromises both reproducibility and completeness of the protein measurements.

An alternative approach to performing proteome-wide measurements is all-ion fragmentation, where no mass-to-charge discrimination is used prior to fragmentation. This approach is termed MSE, and is used to acquire fragmentation spectra for all peptides eluting at a given time.[72] The mass spectrometer cycles between acquiring spectra of intact peptide ions (low collision energy) and fragment ions (high collision energy). In this way, mass spectra for all peptide and fragment ions are recorded, but the resulting fragment ion spectra are highly

complex. Deconvolution of the spectra requires precise liquid chromatography separation so that chromatographic peak shapes can be used to match fragments to their respective intact peptide.[73] Given the high complexity of most proteomic samples, multiple peptides often elute simultaneously and are co-fragmented, which can make the fragment ion spectra difficult to interpret. This phenomenon of ion interference complicates identification and quantification of individual features, reducing depth and accuracy of proteome coverage.[47, 74-78] Without extensive off-line fractionation prior to analysis, liquid chromatography coupled to MSE mass spectrometry is unable to resolve all features in a complex proteomics sample.

The combination of liquid chromatography and MSE alone does not have sufficient peak capacity to separate all tryptic peptides from a whole cell digest. Peak capacity is the number of components that can theoretically be resolved by a specific separation technique.[74] For example, in liquid chromatography, the peak capacity can be defined as the total time of the separation divided by the width (in time) of the average chromatographic peak. Co-elution and co-fragmentation of peptides is a symptom of insufficient peak capacity in proteomics analyses.

One way to increase peak capacity is to introduce in-line ion mobility separation. This orthogonal separation dimension prior to fragmentation helps deconvolute complex mass spectra, increasing the number of individually resolved peptides and precursor ions.[79] In traveling-wave ion mobility spectrometry (TWIMS), ions are confined on-axis by a radiofrequency field, and direct current voltages are sequentially applied to stacked-ring ion guides. This creates a potential wave that pushes ions along through neutral gas. Ions with larger collisional cross sections fall behind the traveling wave more frequently and arrive later to the detector. This arrival time, or drift time, is related to the size, shape, and charge of each ion. Importantly, TWIMS separation can resolve isobaric features with distinct cross sectional areas.[56, 60] In

addition, since larger peptides have later arrival times and also require more energy for fragmentation, adjusting collision energy as a function of drift time improves fragmentation efficiency and significantly increases the number of peptide and protein identifications when employed with TWIMS.[65]

Previous reports have sought to quantify the benefit of the addition of ion mobility to liquid chromatography-mass spectrometry for proteomic analyses in terms of peak capacity, often relying on extrapolation from individual studies of simple sample mixtures in order to predict outcomes for whole-cell tryptic digests.[80-85] Such results are highly dependent on ion mobility resolution, defined as the centroid arrival time divided by the width of arrival time distribution at half height. For TWIMS, a drift-time ion mobility resolution of 45 has been reported for peptides on current generation equipment, with larger values possible when resolution is instead measured in cross-sectional space.[60, 86] Despite significant commercial and user interest in ion mobility separation, there is little information on how much selectivity TWIMS adds to liquid chromatography-mass spectrometry analyses of highly complex proteomics samples. Additionally, it is unclear whether one fixed set of TWIMS parameters can provide adequate separation across the entire diversity of peptides from a whole cell tryptic digest, and there are few reported attempts to optimize TWIMS parameters for proteomics. [87]

To address these issues, we benchmarked TWIMS separation using a commercial HeLa cell tryptic digest. Two of the primary traveling wave ion mobility parameters that can be tuned to affect separation are the wave height (in volts) and wave velocity (in m/s) of the voltages that travel along the stacked-ring ion guides of the device.[55, 86] Published TWIMS settings were surveyed, and we found that many reports use fixed wave height and velocity settings of 40 V and 600 m/s respectively, although there is no clear consensus (**Table 1**).

Table 1. Profile of reported Waters Synapt TWIMS settings. Wave velocity and heights reported for various proteomic, metabolomic and lipidomic studies performed with the Synapt G2 and later instrument models, may not be an exhaustive list. Many proteomics studies employed a constant wave velocity of ~600 m/s and 40 V wave height. A number of publications state that 'optimized mobility conditions' were used, but limited information about the optimization process is given.

Synapt model	Wave velocity (m/s)	Wave height (V)	citation
modified G2	**300**	1.7	Lermyte, Frederik, et al. "ETD allows for native surface mapping of a 150 kDa noncovalent complex on a commercial Q-TWIMS-TOF instrument." *J. Am. Soc. Mass Spectrom.* 25.3 (2014): 343-350.
modified G2-S	**300**	20	Song, Yang, et al. "Refining the Structural Model of a Heterohexameric Protein Complex: Surface Induced Dissociation and Ion Mobility Provide Key Connectivity and Topology Information." *ACS Cent. Sci.* 1.9 (2015): 477-487.
G2-S	**500**	40	Brantley, Matthew, et al. "Automated deconvolution of overlapped ion mobility profiles." *J. Am. Soc. Mass Spectrom.* 25.10 (2014): 1810-1819.
G2-S	**500**	28	Thomas, Andreas, et al. "Simplifying and expanding the screening for peptides< 2 kDa by direct urine injection, liquid chromatography, and ion mobility mass spectrometry." J. Sep. Sci. 39.2 (2016): 333-341.
G2-S	**600**	n/a	Cao, Li, et al. "Alterations in molecular pathways in the retina of early experimental glaucoma eyes." *Int. J. Physiol., Pathophysiol. Pharmacol.* 7.1 (2015): 44.
G2	**600**	40	Cortes, Diego F., Miranda K. Landis, and Andrew K. Ottens. "High-capacity peptide-centric platform to decode the proteomic response to brain injury." *Electrophoresis* 33.24 (2012): 3712-3719.
G2-Si	**600**	n/a	Reis, Ricardo Souza, et al. "Putrescine induces somatic embryo development and proteomic changes in embryogenic callus of sugarcane." *J. proteomics* 130 (2016): 170-179.
G2-S	**600**	n/a	Wang, Hong, and Sam Hanash. "Mass spectrometry based proteomics for absolute quantification of proteins from tumor cells." *Methods* 81 (2015): 34-40.
G2-S	**650**	n/a	Craxton, A., et al. "XLS (c9orf142) is a new component of mammalian DNA double-stranded break repair." *Cell Death Differ.* 22.6 (2015): 890-897.
G2-S	**650**	40	Gelb, Abby S., et al. "A study of calibrant selection in measurement of carbohydrate and peptide ion-neutral collision cross sections by traveling wave ion mobility spectrometry." *Anal. Chem.* 86.22 (2014): 11396-11402.
G2-S	**650**	40	Soderquist, Ryan G., et al. "Development of advanced host cell protein enrichment and detection strategies to enable process relevant spike challenge studies." *Biotechnol. Prog.* 31.4 (2015): 983-989.
G2-S	**650**	40	Thomas, Andreas, Wilhelm Schänzer, and Mario Thevis. "Determination of human insulin and its analogues in human blood using liquid chromatography coupled to ion mobility mass spectrometry (LC-IM-MS)." *Drug Test. Anal.* 6.11-12 (2014): 1125-1132.
G2-S	**650**	40	Wortham, Noel C., et al. "Stoichiometry of the eIF2B complex is maintained by mutual stabilization of subunits." *Biochem. J.* 473.5

			(2016): 571-580.
G2-S	**650**	40	Zhang, Linwen, et al. "In Situ metabolic analysis of single plant cells by capillary microsampling and electrospray ionization mass spectrometry with ion mobility separation." *Analyst* 139.20 (2014): 5079-5085.
G2-S	**650**	40	Zhang, Qingchun, et al. "Characterization of the co-elution of host cell proteins with monoclonal antibodies during protein A purification." *Biotechnol. Prog.* (2016).
G2-S	**652**	40	Jin, Ya, et al. "Analysis of low-density lipoprotein-associated proteins using the method of digitized native protein mapping." *Electrophoresis* (2016).
G2-S	**652**	40	Jin, Ya, et al. "Proteomic analysis of cellular soluble proteins from human bronchial smooth muscle cells by combining nondenaturing micro 2DE and quantitative LC-MS/MS. 1. Preparation of more than 4000 native protein maps." *Electrophoresis* 36.15 (2015): 1711-1723.
G2	**800**	32	Jia, Chenxi, et al. "Site-specific Localization of D-Amino Acids in Bioactive Peptides by Ion Mobility Spectrometry." (2016).
G2-S	**900**	n/a	Parviainen, Ville I., et al. "Label-free mass spectrometry proteome quantification of human embryonic kidney cells following 24 hours of sialic acid overproduction." *Proteome sci.* 11.1 (2013): 1.
G2-S	**1000**	40	Sturm, Robert M., Christopher B. Lietz, and Lingjun Li. "Improved isobaric tandem mass tag quantification by ion mobility mass spectrometry." *Rapid Commun. Mass Spectrom.* 28.9 (2014): 1051-1060.
G2-S	**1200**	n/a	Aird, Steven D., et al. "Snake venoms are integrated systems, but abundant venom proteins evolve more rapidly." *BMC genomics* 16.1 (2015): 1.
G2-S	**1300**	30.5	Geromanos, Scott J., et al. "Using ion purity scores for enhancing quantitative accuracy and precision in complex proteomics samples." *Anal. Bioanal. Chem.* 404.4 (2012): 1127-1139.
G2-S	**1200-300**	40	Milic, Ivana, et al. "Separation and characterization of oxidized isomeric lipid–peptide adducts by ion mobility mass spectrometry." *J. Mass Spectrom.* 50.12 (2015): 1386-1392.
G2-S	**1200-400**	38	Helm, Stefan, et al. "Protein identification and quantification by data-independent acquisition and multi-parallel collision-induced dissociation mass spectrometry (MS E) in the chloroplast stroma proteome." *J. proteomics* 98 (2014): 79-89.
G2-Si	**1200-400**	40	Scheerlinck, Ellen, et al. "Minimizing technical variation during sample preparation prior to label-free quantitative mass spectrometry." *Anal. Biochem.* 490 (2015): 14-19.
G2-S	**1250, 1000**	40	Harper, Brett, Mahsan Miladi, and Touradj Solouki. "Loss of internal backbone carbonyls: additional evidence for sequence-scrambling in collision-induced dissociation of y-type ions." *J. Am. Soc. Mass Spectrom.* 25.10 (2014): 1716-1729.
G2-S	**250, 540, 900**	20, 30, 40 respectively	Kune, Christopher, Johann Far, and Edwin De Pauw. "Accurate drift time determination by traveling wave ion mobility spectrometry: The concept of the diffusion calibration." *Anal. Chem.* (2016).
G2	**500, 500, 600, 700, 800**	30, 35, 35, 35, 40 respectively	Lietz, Christopher B., Qing Yu, and Lingjun Li. "Large-scale collision cross-section profiling on a traveling wave ion mobility mass spectrometer." *J. Am. Soc. Mass Spectrom.* 25.12 (2014): 2009-2019.
G2-S	**800-200**	40	Gavriilidou, Agni FM, Basri Gulbakan, and Renato Zenobi. "Influence of Ammonium Acetate Concentration on Receptor–Ligand Binding Affinities Measured by Native Nano ESI-MS: A Systematic Study." *Anal. Chem.* 87.20 (2015): 10378-10384.

G2-S	**800-500**	40	Distler, Ute, et al. "Drift time-specific collision energies enable deep-coverage data-independent acquisition proteomics." *Nat. Methods* 2 (2014): 167-170.
G2-S	**800-500**	40	Docter, Dominic, et al. "Quantitative profiling of the protein coronas that form around nanoparticles." *Nat. Protoc.* 9.9 (2014): 2030-2044.
G2-S	**800-500**	40	Schwarz, Alexandra, et al. "A systems level analysis reveals transcriptomic and proteomic complexity in Ixodes ricinus midgut and salivary glands during early attachment and feeding." *Mol. Cell. Proteomics* 13.10 (2014): 2725-2735.
G2-S	**800-500**	40	Tenzer, Stefan, et al. "Proteome-wide characterization of the RNA-binding protein RALY-interactome using the in vivo-biotinylation-pulldown-quant (iBioPQ) approach." *J. Proteome Res.* 12.6 (2013): 2869-2884.
G2-S	**between 1500 and 1000**	between 25 and 40	Rathore, Deepali, Forouzan Aboufazeli, and Eric D. Dodds. "Obtaining complementary polypeptide sequence information from a single precursor ion packet via sequential ion mobility-resolved electron transfer and vibrational activation." *Analyst* 140.21 (2015): 7175-7183.

As the mechanisms of separation in reversed-phase liquid chromatography (LC), traveling wave ion mobility spectrometry (TWIMS), and mass spectrometry (MS) each leverage mass-to-charge ratio, cross-sectional area, and hydrophobicity to different extents. The difference, or orthogonality, between these separation processes determines the extent to which resolving power will increase when the techniques are used in tandem.[88]

Taking into account the degree of orthogonality observed between the LC, TWIMS and MS dimensions, we calculated a three-dimensional separation space to provide a quantifiable peak capacity. Importantly, the majority of ions fell within a limited area of the TWIMS separation space. Systematic optimization of TWIMS wave height and wave velocity parameters identified ramped wave velocity settings that increased the drift space occupancy, yielding a 40% increase in peptide annotations and a 50% increase in the theoretical peak capacity. Under these conditions, the velocity of the DC pulses within the TWIMS separator are ramped in a time domain that allows high mobility species to experience primarily higher wave velocities and lower mobility peptides to experience progressively lower velocities, each of which can be optimized for the mobility range of the peptides separated. Overall, this analysis establishes a

quantitative benchmark for TWIMS separation, highlighting customizable settings to optimally resolve peptide features in complex proteomics samples.

Ion mobility spectrometry separation dramatically improves the number of peptide and protein identifications in all-ion fragmentation (MSE) shotgun proteomics. Unfortunately, IMS-resolved MSE data files acquired on the Waters Synapt platform are stored in a proprietary file format, and existing tools for data conversion perform poorly with ion mobility-mass spectrometry data. For example, msconvert[89], a popular tool for converting between LC-MS file formats, collapses TWIMS separation information entirely, so that peptide ion drift times cannot be found and used.

While individual peptide features can be manually interrogated using the Waters proprietary DriftScope software, it is not technically feasible to manually extract comprehensive information for each of the hundreds of thousands of features across an IMS-MS proteomics data set.

To address this issue, we designed a user-friendly interface (TWIMExtract) around a minimal raw data extraction tool provided by Waters (see Data Analysis section, below). TWIMExtract is a Java-based graphical user interface (GUI) for extracting data from Waters' proprietary .raw format. It uses an executable to query the proprietary format and return information requested by the user. The Java GUI allows users to quickly process large numbers of data files and extract whole or targeted data sets in an automated fashion. The user selects the data file(s) to query, which are displayed in a table with accompanying information, and range file(s) specifying the regions of the three-dimensional data to extract. Any of the three dimensions of data (retention time, m/z, and ion mobility arrival time) can be extracted individually. This enables users interested in obtaining retention time chromatograms, arrival

time distributions, or mass spectra of any or all features in their raw data set to rapidly and automatically extract those features for further analysis. TWIMExtract provides raw data without further processing and is intended to act as a flexible piece of larger downstream data analysis pipelines. Importantly, TWIMExtract can extract hundreds of thousands of user-defined slices of liquid chromatography-ion mobility-mass spectrometry (LC-IMS-MS) data in just a few hours. Users can specify any combination of ranges of chromatographic retention time, IM drift time, and m/z and process any number of raw data files with any number of slices. Key parameters can be exported with the data, including collision energies and TWIMS settings, and multiple extractions from the same raw file can be automatically stitched together into a single .csv output file. TWIMExtract also has batch processing capabilities to enable large-scale automated data extraction.

Experimental

Reagents and solvents were purchased from Fisher unless otherwise indicated.

Mass spectrometry

HeLa protein cell digest standard (20 µg, Pierce, 88329) was resuspended in LC-MS grade water supplemented with 10 fmol / µL alcohol dehydrogenase, 3% acetonitrile, and 0.1% formic acid. For each experiment, 2 µL (200 ng) of the reconstituted HeLa standard was injected and separated using a one-dimensional Waters nanoAcquity UPLC system fitted with a 5 µm C18 (180 µm x 20 mm) trap column and a 1.8 µm HSS T3 analytical column (75 µm x 250 mm). Tryptic HeLa peptides were loaded onto the trap column over 3 minutes, followed by analytical separation over a 110 minute gradient (3% acetonitrile to 40% acetonitrile over 92 minutes) at a flow rate of 0.5 µL / min. Eluting peptides were introduced to the Synapt-G2S HDMS traveling

wave ion mobility quadrupole time-of-flight mass spectrometer (Waters) by electrospray ionization using a pre-cut uncoated 10 μm ID SilicaTip nanospray emitter (New Objective). The Synapt-G2S auxiliary pump was used to deliver the lock mass compound [Glu1]-Fibrinopeptide B (100 fmol / μL; Waters) at 400 nL / min through the reference sprayer of the nanoLockSpray source. The nanoESI source was set to 70°C with a 3 kV applied potential. Data-independent acquisition (MSE) was performed in positive mode. The quadrupole mass analyzer was set to transmit all precursor ions (50-2000 m/z) through to the TOF mass analyzer. Precursor ions traveled through the quadrupole mass analyzer, and were then separated in the traveling wave device using the manufacturer recommended constant wave height of 40 V and wave velocity of 600 m/s, which was later optimized. Low and elevated-energy scans were alternated every 0.5 seconds (0.05 s interscan delay) to acquire sequential precursor and product ion data. Collision energy was applied after ion mobility separation in the transfer cell to fragment peptide ions via collision-induced dissociation. In the low-energy mode, data were collected at constant collision energy (CE) of 4 eV across all 200 drift bins (TOF scans). In high-energy mode, no CE was applied in bins 0 – 19. Collision energy ramping was applied for the 1000-600 m / s wave velocity condition as follows (collision energies used with the 600 m / s condition in parentheses): bins 20 – 119, 16.4 – 58.5 eV (11.0 – 44.1 eV); bins 120 – 200, 58.9 – 63.9 eV (44.4 – 49 eV). Wave velocities in the trap and transfer cell were set to 311 m / s and 380 m / s, respectively, with wave heights of 4 V. In the trap cell, the N2 gas flow rate was 2.0 mL / min. In the ion mobility cell, the He and N2 flow rates were set to 180 and 90 mL / min, respectively. The mass spectrometer was operated in Resolution mode (V-mode), with a typical resolving power of at least 20,000 FWHM. Data were mass corrected post-acquisition by comparison to

the doubly-charged monoisotopic ion of [Glu1]-Fibrinopeptide B (m/z = 785.843), which was sampled from the nanoLockSpray source every 30 seconds throughout data acquisition.

Data analysis

Raw data files were analyzed with ProteinLynx Global SERVER version 3.0.2 (PLGS, Waters Corporation), searching against the Uniprot database of the human proteome. The following criteria were applied to perform the search: (i) trypsin as digestion enzyme, (ii) one missed cleavage allowed, (iii) carbamidomethyl cysteine as a fixed modification and methionine oxidation as a variable modification, (iv) a minimum of two identified fragment ions per peptide and a minimum of five fragments per protein, and (v) at least two identified peptides per protein. The global protein false discovery rate (FDR) was set at 1% using a reversed sequence database. Any identified peptides with a calculated mass error greater than 10 ppm were not considered. Accurate mass retention time (AMRT) and peptide information was extracted using TWIMExtract. The TWIMExtract tool, along with instructions for installation and operation, can be downloaded from https://sites.lsa.umich.edu/ruotolo/software.

Also available on this site is the Python script used to measure drift arrival time distributions and ion mobility resolution following the use of TWIMExtract. Briefly, ion mobility arrival time distributions for each extracted feature are analyzed to find the highest intensity peak, from which the arrival time and peak width at half the maximum intensity are found. From the arrival time and width, ion mobility resolution of the feature is calculated.

Source code for TWIMExtract can be found at www.github.com/dpolasky/TWIMExtract.

Results and Discussion

We applied TWIMExtract to evaluate the arrival time distributions of peptide ion features across ±0.005 m/z and ±0.2 min from their respective centroid values. We then implemented a custom peak-processing algorithm in Python (see Experimental section) to extract the centroid drift time, peak width, and resolution of the dominant peak within the arrival time distribution. Due to the complexity of the HeLa digest, the m/z and retention time windows occasionally extracted multiple peptide features in a single ion mobility arrive time distribution. In the rare event (<0.1%) when an ion clearly fell outside the linear m/z−drift time trendline of the assigned charge state, the event was flagged as an incorrect assignment and excluded from further analysis.

We started by comparing the ion mobility drift times with mass to charge (m/z) and LC retention time of a commercial HeLa cell proteome tryptic digest, plotting each separation in two dimensions. Drift time information was extracted for all annotated peptides and compared to the accurate mass retention time (AMRT) features. The lists of AMRTs, which can be found in one of the .csv files resulting from PLGS analysis, include all deisotoped features detected in both low (peptide) and high (fragment) ion spectra. The AMRT datasets provide another means of benchmarking the LC-TWIMS-MS system by reporting all peptide-like features, whether or not those features resulted in database search hits. When ion mobility drift times are compared to m/z, tryptic peptides are primarily in the 2+ charge state and share a common trendline (**Figure 8a**).

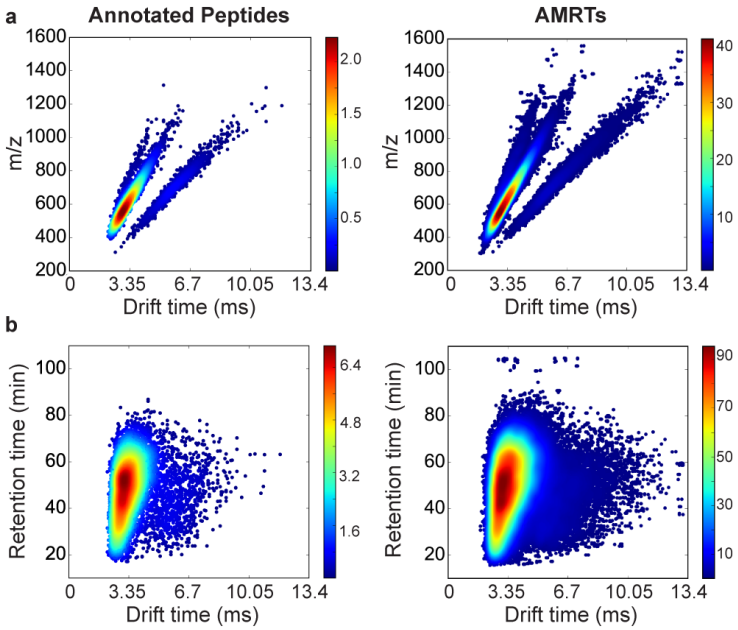

Figure 8. Distribution of annotated peptides and AMRTs in three dimensions. (a) Two dimensional plot of m/z and drift separates ions by charge state. 1+ peptides have longer drift times and are entirely separated from the 2+ and higher trendlines. (b) Two dimensional plot of liquid chromatography versus drift separation reveals greater orthogonality than mass-to-charge separation. Representative data from three independent replicates.

Importantly, both AMRTs and annotated peptides show nearly identical distributions in the drift and mass dimensions, and in both instances the majority of features occupy a small, highly dense region of separation space. Liquid chromatography separation is significantly more orthogonal to ion mobility separation and distributes features more evenly across the separation space, a trend that can be observed with both annotated peptides and AMRTs (**Figure 8b**). Indeed, ~90% of the annotated peptides or AMRTs occupy less than 25% of the overall drift space (**Figure 9a**). Under these traveling wave settings, tryptic peptides do not access the majority of TWIMS separation space, highlighting poor separation efficiency for complex mixtures of peptides with similar collisional cross sections.

Next, we evaluated TWIMS resolving power for both AMRTs and annotated peptides in the HeLa cell tryptic digest. After optimization of wave height and wave velocity, synthetic

peptide standards can reportedly achieve IMS resolution between 25 and 45.[60, 86] In a complex tryptic digest, our analysis returned significantly lower values, with a mean resolution of 18.6 and a large standard deviation of 5.5 (**Figure 9b**).

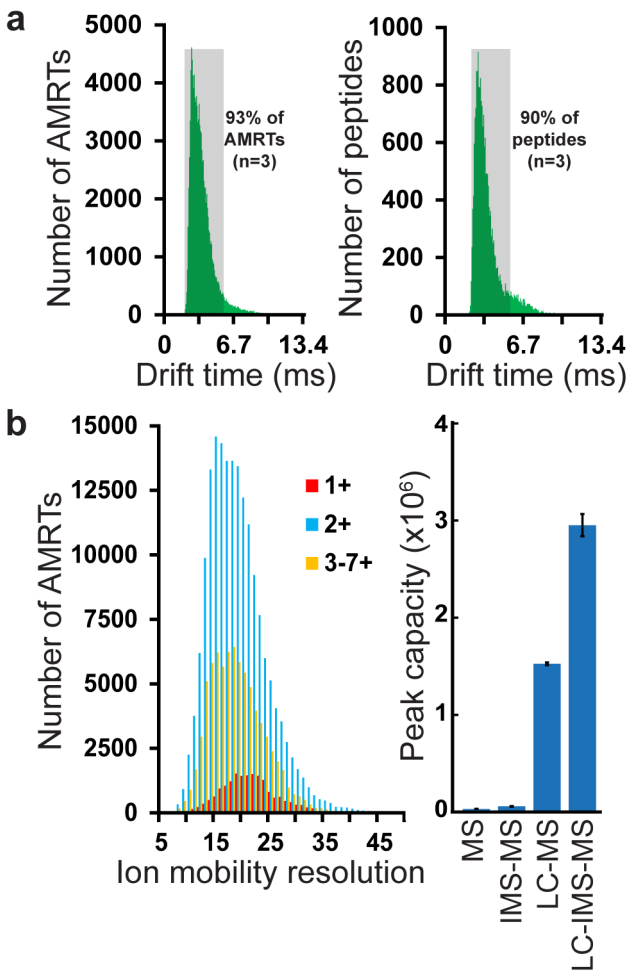

Figure 9. Tryptic peptides occupy a limited area of TWIMS separation space. (a) Distribution of AMRTs and peptides in drift time. Shaded gray areas define 25% of the total TWIMS separation space, corresponding to approximately 90% of the total AMRT and peptide events. (b) Ion mobility resolution of >106 AMRTs, displayed by charge state. (c) TWIMS doubles the peak capacity of LC-MS analysis. Data are representative of 3 independent replicates.

Peptides with different charge states had significantly different apparent TWIMS resolutions. For example, 1+ peptides exhibited higher mean resolution (21.1 ± 0.4) than 2+

(18.6 ± 0.3) and higher charge state (18.4 ± 0.2) ions ($p < 0.001$, N = 3). When the same peptide was compared across different charge states, the species with higher charge generally exhibited lower mobility resolution (**Table 2**). This is most likely due to the broadening of the arrival time distributions of higher charge features due to the presence of unresolved conformers. Highly charged ions may adopt extended conformations in order to reduce charge repulsion, increasing the measured width of the arrival time distribution.[90]

Table 2. Ion mobility resolution of identical peptides in different charge states.

peptide sequence	1+ drift time (ms)	1+ FWHM (ms)	1+ resolution	2+ drift time (ms)	2+ FWHM (ms)	2+ resolution	*resolution difference*
ELNYFAK	7.4	0.402	**19.2**	2.4	0.201	**12.0**	***7.2***
EIAQDFK	7.2	0.402	**17.8**	2.3	0.201	**11.3**	***6.5***

peptide sequence	2+ drift time (ms)	2+ FWHM (ms)	2+ resolution	3+ drift time (ms)	3+ FWHM (ms)	3+ resolution	*resolution difference*
GHLENNPALEK	3.2	0.134	**24.0**	2.3	0.201	**11.7**	***12.3***
AQIHDLVLVGGSTR	4.1	0.201	**20.3**	2.5	0.268	**9.5**	***10.8***
GQPIYIQFSNHK	3.7	0.134	**27.5**	2.3	0.134	**17.0**	***10.5***
EGMAALQSDPWQQELYR	5.4	0.201	**27.0**	3.4	0.201	**16.7**	***10.3***
EILGTAQSVGCNVDGR	4.2	0.201	**21.0**	2.9	0.268	**11.0**	***10.0***
HEQNIDCGGGYVK	3.8	0.134	**28.5**	2.5	0.134	**18.5**	***10.0***
GITINAAHVEYSTAAR	4.9	0.201	**24.3**	2.9	0.201	**14.7**	***9.7***
ADDGRPFPQVIK	3.6	0.134	**27.0**	2.5	0.134	**18.5**	***8.5***
SPFVVQVGEACNPNACR	4.9	0.201	**24.3**	3.2	0.201	**16.0**	***8.3***
GSYVSIHSSGFR	3.6	0.134	**26.5**	2.6	0.134	**19.5**	***7.0***
GHLENNPALEK	3.2	0.134	**24.0**	2.3	0.134	**17.5**	***6.5***
VPADTEVVCAPPTAYIDFAR	5.9	0.268	**22.0**	3.3	0.201	**16.3**	***5.7***
EILGTAQSVGCNVDGR	4.3	0.268	**16.0**	2.9	0.268	**10.8**	***5.3***
TIGGGDDSFNTFFSETGAGK	5.0	0.268	**18.8**	2.8	0.201	**14.0**	***4.8***
LAGTQPLEVLEAVQR	4.4	0.268	**16.5**	2.7	0.201	**13.3**	***3.2***
ALTVPELTQQMFDAK	4.6	0.201	**23.0**	2.7	0.134	**20.5**	***2.5***
ALTVPELTQQMFDAK	4.6	0.201	**22.7**	2.7	0.134	**20.5**	***2.2***
EHALLAYTLGVK	3.6	0.268	**13.5**	2.3	0.201	**11.3**	***2.2***
ETAENYLGHTAK	3.8	0.201	**18.7**	2.2	0.134	**16.5**	***2.2***

ITGEAFVQFASQELAEK	4.9	0.268	**18.3**	3.3	0.201	**16.3**	***1.9***
ALQEGEGDLSISADR	4.0	0.201	**20.0**	2.5	0.134	**18.5**	***1.5***
RAAEDDEDDDVDTK	4.1	0.201	**20.3**	2.6	0.134	**19.5**	***0.8***
ALTVPELTQQMFDSK	4.5	0.335	**13.4**	2.7	0.201	**13.3**	***0.1***
KYDAFLASESLIK	4.1	0.201	**20.3**	2.8	0.134	**21.0**	***-0.7***
QAGEVTYADAHK	3.2	0.201	**16.0**	2.3	0.134	**17.0**	***-1.0***
AQTAHIVLEDGTK	3.8	0.201	**19.0**	2.8	0.134	**21.0**	***-2.0***
TIGGGDDSFNTFFSETGAGK	5.0	0.268	**18.5**	2.8	0.134	**21.0**	***-2.5***
LAGTQPLEVLEAVQR	4.4	0.268	**16.3**	2.7	0.134	**20.0**	***-3.8***
YPLFEGQETGKK	3.4	0.268	**12.8**	2.3	0.134	**17.5**	***-4.8***
FDHIFYTGNTAVGK	4.1	0.268	**15.3**	2.9	0.134	**21.5**	***-6.3***
peptide sequence	**3+ drift time (ms)**	**3+ FWHM (ms)**	**3+ resolution**	**4+ drift time (ms)**	**4+ FWHM (ms)**	**4+ resolution**	***resolution difference***
VHIPNDDAQFDASHCDSDK	3.5	0.134	**26.0**	2.7	0.201	**13.3**	***12.7***
VIHLSNLPHSGYSDSAVLK	3.4	0.134	**25.0**	2.9	0.201	**14.3**	***10.7***
KPTDGASSSNCVTDISHLVR	3.3	0.201	**16.7**	2.8	0.201	**14.0**	***2.7***
LMLSTSEYSQSPKMESLSSHR	3.4	0.201	**17.0**	3.2	0.201	**16.0**	***1.0***
FVPQEMGPHTVAVKYR	3.1	0.201	**15.7**	2.5	0.134	**18.5**	***-2.8***
EATTDFTVDSRPLTQVGGDHIK	4.0	0.268	**15.0**	2.9	0.134	**22.0**	***-7.0***

The ‘resolution difference’ column contains the resolution of the higher charge state subtracted from that of the lower charge state (e.g. [1+ resolution] – [2+ resolution]). These resolution differences between charge states of the same peptide are shown in **Figure 10**. For the same peptide identified in different charge states (n=41 comparisons), the lower charge state generally exhibited higher ion mobility resolution, likely due to fewer stable conformational states.

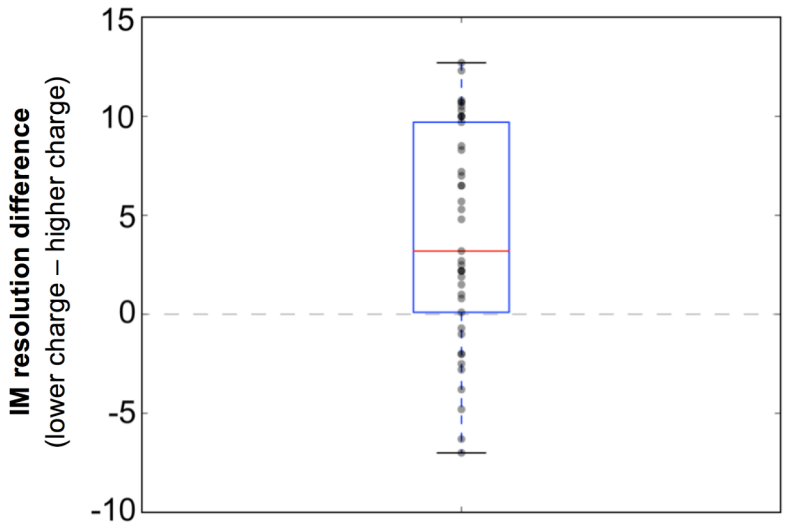

Figure 10. Ion mobility resolution of identical peptides in different charge states. IM resolution difference between higher and lower charge states (comparing 1+ to 2+, 2+ to 3+ , and 3+ to 4+ , as shown in the table above (e.g. [1+ resolution] – [2+ resolution]). Median is displayed as the red line, upper and lower quartiles are bounded by the blue box.

Using the extracted TWIMS resolution data, we next sought to estimate the total three-dimensional peak capacity of the LC-TWIMS-MS analysis. Peak capacity of a single dimension of separation represents the total separation length divided by the average peak width. If LC, IMS, and MS were fully orthogonal to one another, the total peak capacity of the LC-IMS-MS system would be the product of each of the peak capacities from each dimension.[88] Because LC, IMS, and MS separate peptides on the basis of related physicochemical properties, we accounted for their empirical correlation in our estimates of three-dimensional system peak capacity (**Figure 9**).

First, we calculated linear fits of m/z and drift time separately for 1+, 2+, and combined 3−7+ charge states (white dashed line, step **Figure 11.1**). From these linear fits, a mean m/z value was calculated for each drift time "bin", and a conservative separation length was defined as the average deviation from the mean m/z.[82] From these values, two-dimensional drift time –

m/z peak capacity was calculated for each charge trendline by multiplying the correlation-adjusted m/z peak capacity by the drift time peak capacity (orange shaded regions in **Figure 11.1**) demarcating the actual area of two-dimensional IMS-MS separation space occupied (orange shaded regions). Average peak widths across both the m/z and drift time dimensions were fit into the separation area to give the two-dimensional peak capacity for each charge trendline (**Figure 11.2**).

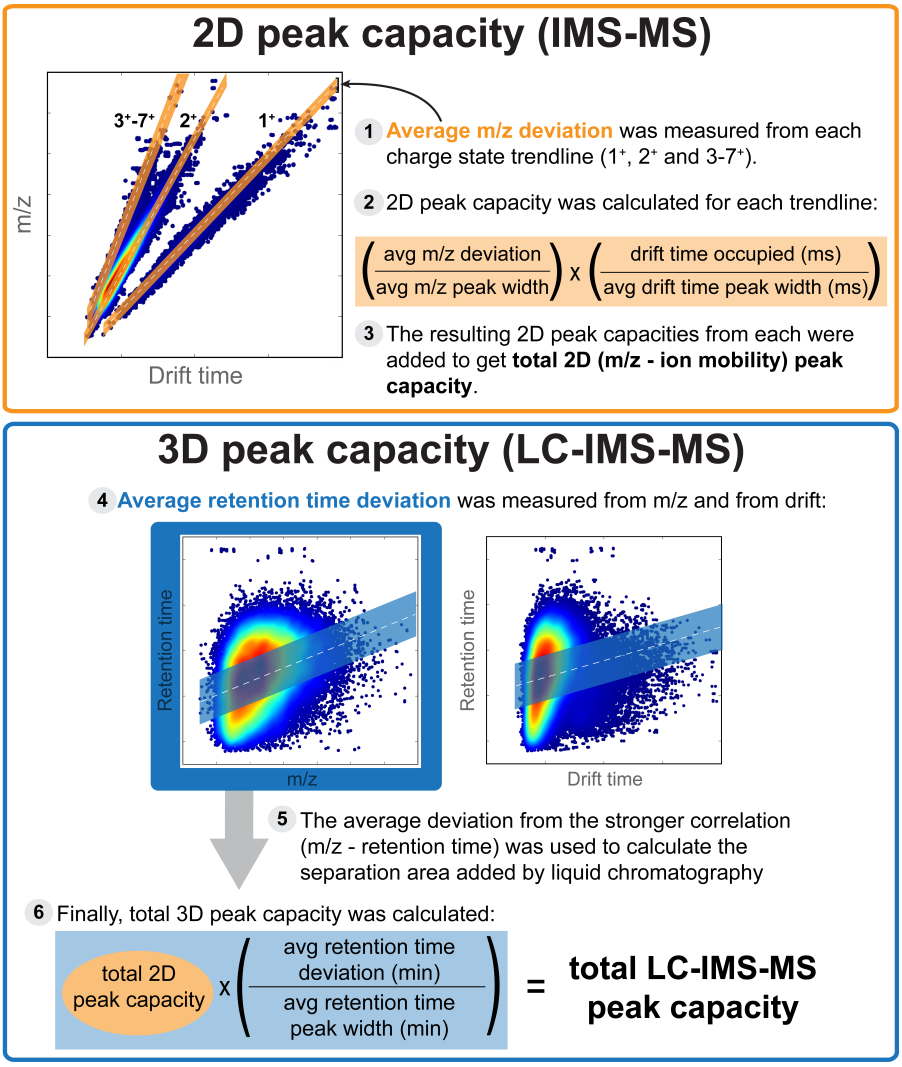

Figure 11. Calculation of peak capacity in three dimensions.

Summing the three two-dimensional peak capacities yielded a two-dimensional peak capacity of $5.67 \pm 0.27 \times 10^4$ peaks across the m/z and ion mobility dimensions (**Figure 11.3**),

nearly twice the $2.93 \pm 0.01 \times 10^4$ peaks defined by mass analysis alone (**Figure 9**). Next, the degree of correlation between LC, IM, and MS separation was defined (**Figure 11.4**). Both m/z and drift time are almost completely uncorrelated with retention time. To provide the most conservative estimate of three-dimensional peak capacity, we estimated the contribution of the LC dimension from the slightly stronger retention time−m/z correlation (**Figure 11.6**). This added separation "length" was then divided by average chromatographic peak width (12 s), resulting in 53-fold more peak capacity than the two-dimensional IM-MS separation. The 2-dimensional peak capacity multiplied 53-fold produced a total three-dimensional peak capacity estimate of $2.95 \pm 0.11 \times 10^6$ using the standard TWIMS settings or twice the $1.53 \pm 0.01 \times 10^6$ afforded by LC-MS alone (**Figure 9**).

From our analysis, the starting TWIMS settings (wave height of 40 V, wave velocity of 600 m/s) do not efficiently occupy the available IM separation space. Therefore, we sought to optimize TWIMS separation using the HeLa cell protein digest standard.

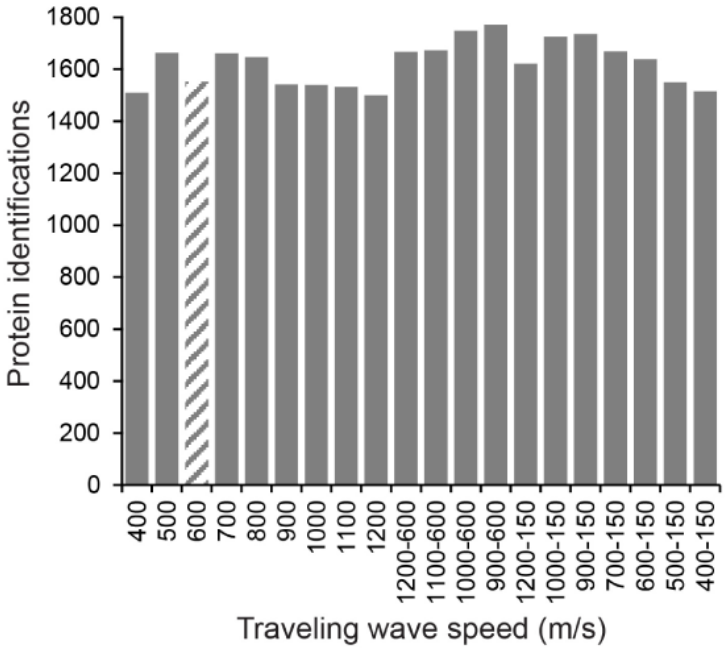

Figure 12. Protein identifications using differential wave velocities.
Constant wave height (40 V) was used for all experiments.

Our initial survey tested 20 different combinations of traveling-wave velocities, which overall achieved largely the same number of protein annotations when the same collision energy profile was used (**Figure 12**).

The total number of annotated proteins from different TWIMS wave velocities demonstrates a greater number of protein identifications with variable velocities versus a constant collision energy gradient. These data represent the first iteration of experiments with modified traveling wave speed. From here, we chose the three traveling wave conditions that yielded the most protein identifications (900 – 600, 900 – 150 and 1000 – 600 m / s) for further optimization. For each of these three conditions (in addition to 600 m / s), drift-time specific collision energy profiles were generated following the protocol outlined by Distler et al.[65]

Interestingly, linearly decreasing the traveling-wave velocity during the 13.4 ms separation increased the number of annotated proteins by approximately 10%. Therefore, we selected the 1000−600 m/s wave velocity (and 40 V wave height) was chosen for further analysis, optimizing the drift time−specific collision energy profile to increase fragmentation efficiency. Following this optimization, the 1000−600 m/s wave velocity reproducibly increased the mean number of annotated peptides from $1.25 \pm 0.05 \times 10^4$ to $2.00 \pm 0.14 \times 10^4$ peptides, or more than 60% (**Figure 13a**).

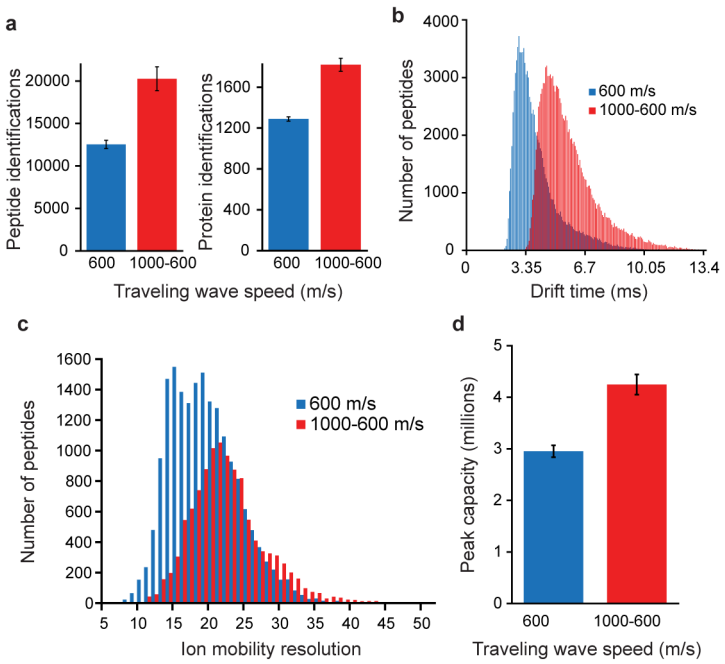

Figure 13. Variable-velocity TWIMS separation improves peptide analysis. (a) Peptide and protein identifications are increased with variable-velocity TWIMS. (b) Optimized wave velocity settings increase drift space occupancy. (c) Enhanced TWIMS resolution from variable−velocity separation. (d) LC-IMS-MS peak capacity is increased with optimized TWIMS settings. Data are representative of three independent replicates.

This translated to an increase from 1289 ± 21 to 1820 ± 64 proteins and increased occupancy throughout the ion mobility separation dimension with no loss in peak integrity (**Figure 13b** and **Figure 14**). An increased spread in drift time, an overall shift to later drift times, and an increased number of detected features (particularly in the higher mass range) can be observed using the 1000-600 m/s condition.

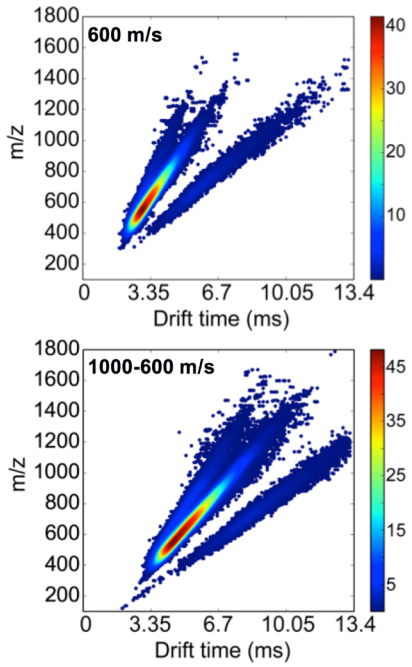

Figure 14. Variable-velocity TWIMS enhancement in occupied separation space. Comparison of accurate mass retention time pair (AMRT) drift time and m/z distributions between 600 m/s (top) and 1000-600 m/s (bottom) wave velocity.

In fact, we observe an overall increase in the mean resolution from 18.8 ± 0.6 to 21.9 ± 0.2 (**Figure 13c**), which significantly shifts the median drift time toward larger values while maintaining peak width, yielding improved overall resolution. Furthermore, the resulting mean peak capacity is $4.25 \pm 0.2 \times 10^6$ or 44% higher than the default traveling-wave velocity settings (**Figure 13d**). Importantly, the variable drift time settings (1000−600 m/s) not only improved ion dispersal but also improved transmission efficiency (**Figure 15**). We also examined the rate of peptide identifications from all peptide-like features (AMRT) across drift time and retention time (**Figure 16).** Relative to total ion current, the peptide annotation rate was higher at longer drift times. In addition, the 1000−600 m/s wave velocity settings increased the mean size of identified peptides from 649 ± 2 to 692 ± 5 m/z, which is likely related to more effective drift time-dependent collision energy assignment resulting from more efficient ion dispersal. Clearly, even

modest improvements in TWIMS separation can greatly enhance the resultant LC-IM-MS peak capacity, affording better precursor-fragment alignment, peak annotation, and ultimately a larger number of high-confidence protein identifications.

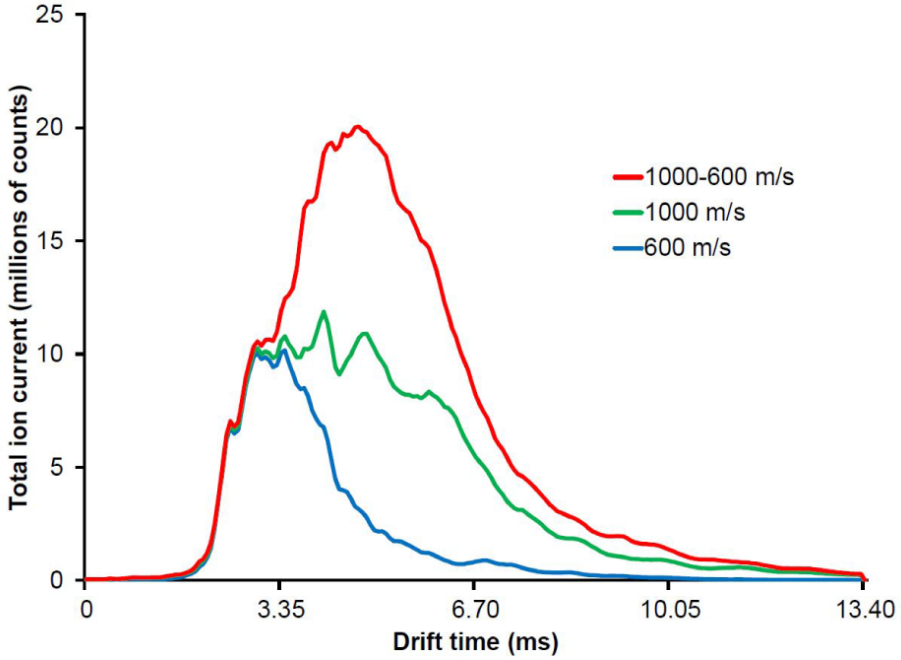

Figure 15. Total ion drift time distributions under different traveling wave velocities.

Overall, this chapter presents a detailed analysis of the resolving power and peak capacity of LC-TWIMS-MS for cellular mixtures of tryptic peptides. On the basis of this analysis, TWIMS is only marginally orthogonal to LC-MS analysis, yet as previously reported, TWIMS separation increases the number of protein identifications from all-ion fragmentation methods. Additionally, the optimized, wave velocity ramped TWIMS conditions increased the overall peak capacity by 2.8-fold over LC-MS analysis alone, which significantly enhances peptide and protein identifications. Improvements in ion mobility separation technologies promise to further improve the peak capacities of LC-IMS-MS systems for the analysis of tryptic peptides, which

will continue to reduce ion interferences that are currently commonplace in most complex data-independent acquisition proteomics analyses.

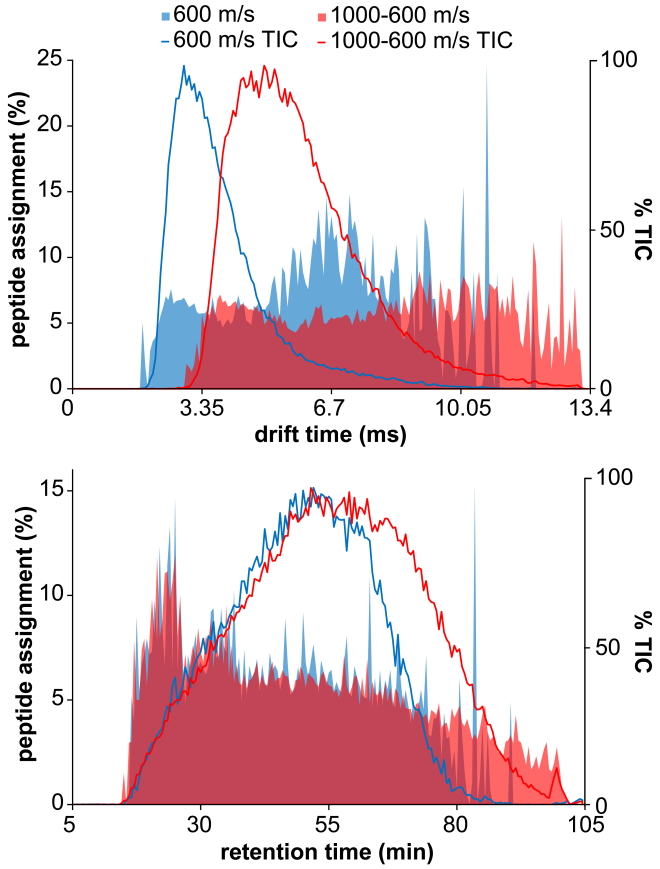

Figure 16. Peptide assignment efficiency as a function of drift time (top) and retention time (bottom). The percentage of annotated peptides out of all AMRTs per drift bin or for 0.5 minutes of retention time for both traveling wave settings is shown shaded and corresponds to the left axis. The percentage of the total ion current over drift or retention time for each condition is shown as line traces and corresponds to the right axis.

Chapter 3 Development of post-processing methods to improve DIA-SILAC quantification

Introduction

As discussed in Chapter 1, the standard method for acquiring LC-MS proteomics data relies on repeating acquisition cycles initialized by a scan of peptide ions followed by dynamic selection, isolation, and fragmentation of a fixed number of the most abundant ions for sequence identification.[91] In regions of high chromatographic complexity, the rate of the data-dependent acquisition (DDA) duty cycle limits isolation, fragmentation, and subsequent identification to only the most intense precursor ions. This stochastic sampling method ties the number of analyzed peptides to the instrument scan speed, and fundamentally limits proteome coverage.[92] Additionally, the low mass resolution of multipole ion selection often produces undesirable ion interference, yielding chimeric fragmentation spectra across >50% of MS2 spectra.[47] The resulting composite fragment spectra incur scoring penalties during database searching, which limits confident peptide annotations. These variables diminish quantitative accuracy and reproducibility between analyses, particularly replicate LC-MS experiments, prompting the need for alternative acquisition methods.

To address this gap, data-independent acquisition (DIA) methods have been developed that bypass abundance-based ion selection. Instead of iteratively isolating the most abundant ions for fragmentation, all-ion fragmentation methods, such as MSE, use alternating MS scans collected at low and high collision energies, generating precursor (MS1) and product (MS2) ion spectra across the entire mass range.[72, 93, 94] Post-acquisition alignment of MS1 and MS2 spectra

peptide elution profiles allows accurate assignment of precursors to their corresponding product ions.[73]

Recent commercial mass spectrometers incorporating ion mobility spectrometry (IMS) separation provide an orthogonal approach to reduce ion interference for DIA workflows. In general terms, IMS describes a gas-phase separation that resolves ions based on their size, shape, and charge on a millisecond time scale. These inherent physical properties impart distinct IMS drift times, providing an orthogonal analytical dimension that increases peak capacity (as quantified in Chapter 1), reduces ion interference, and delivers multiparameter peptide analytics (retention time, drift time, precursor and product m/z) that can be used for library building.[81, 95] Furthermore, drift-time dependent collision energy assignment produces more efficient fragmentation by delivering higher collision energies to larger peptides with longer drift times.[65] Altogether, IMS-DIA methods yield a 2.3-fold higher annotation rate than the theoretical maximum for a first generation quadrupole Orbitrap instrument using higher energy collisional dissociation (HCD).[65] This results in 1/3 more annotated proteins and twice as many annotated peptides. Based on these values, IMS separation improves both peak capacity and fragmentation efficiency to provide a unique addition to DIA analysis platforms.

Label-free quantification methods typically measure the sum of the abundances of three ions with the highest intensity per protein (top-3 analysis), which generally correlates with protein abundance.[93] This procedure immediately triages the bulk of the peptide features from the final measurement, diminishing the statistical power of the measurement needed for confident cross-experiment analyses. In SWATH acquisition experiments, where 10-25 m/z windows are used in place of abundance-based ion selection, proteins can be similarly quantified by summing the five most intense product ions from the three most intense peak groups. This

approach yields intraday coefficients of variation of less than 20% across replicate runs, but variance increases significantly between different laboratories operating the same instrument.[96] Many sources of the variance observed between facilities in this study could be corrected by incorporating internal standards.

We hypothesized that stable isotope labeling by amino acids in cell culture (SILAC) analysis might improve reproducible DIA quantitation across the proteome, especially since comparative ratios are internally quantified and eliminate many sources of experimental error. In this approach, control and experimental cells are grown separately in "Light" media or stable isotope-labeled "Heavy" media supplemented with L-arginine ($^{13}C_6$, $^{15}N_4$) and L-lysine ($^{13}C_6$, $^{15}N_2$). The cell lysates are then mixed and digested with trypsin for comparative LC-MS proteomic analysis. Since trypsin cuts proteins at arginine and lysine residues, the ion abundance of each control (Light) and experimental (Heavy) peptides can be quantified and compared across all tryptic peptides. In one study, when SILAC is combined with IMS-DIA methods, even the point in the LC gradients where the most peptides are eluting, IMS separation limits interference of SILAC ions to only 5.5%.[75] The addition of ion mobility spectrometry enhances peak capacity to resolve much of the complexity introduced through incorporation of 2-plex stable isotope labels. Nevertheless, it remained unclear if this purity is sufficient for accurate quantitative analysis. This chapter describes the benchmarking of stable isotope quantitation using IMS-DIA methods, particularly for accurate quantification across a range of fractional abundance changes.

Experimental

Reagents and solvents were purchased from Fisher unless otherwise indicated.

Cell culture and sample preparation

HEK 293T cells were grown for more than 7 passages in Light or Heavy SILAC DMEM (Thermo Scientific, #88364) supplemented with Penicillin-Streptomycin (Gibco), 10% v/v dialyzed fetal bovine serum (Atlanta Biologicals), and 1.1 mM arginine and lysine. Heavy arginine (+10 Da: $^{13}C_6$, $^{15}N_4$ labeled L-arginine hydrochloride) and Heavy lysine (+8 Da: $^{13}C_6$, $^{15}N_2$ labeled L-lysine hydrochloride) were purchased from Sigma. For protein turnover experiments, cells grown in Heavy SILAC media were gently rinsed with 1x PBS (Gibco), and light media was added for 2, 4 or 6 hours. Cell pellets were lysed by sonication in 6 M urea in 1x PBS and quantified by BCA (BioRad). Light and Heavy lysates were combined at 5 different Light : Heavy ratios (1/5, 1/3, 1/2, 1, 2, 3, 5). Each of these samples was reduced with 10 mM dithiothreitol (DTT) for 30 minutes at room temperature, then alkylated with 50 mM iodoacetamide for 30 minutes at room temperature in the dark. Samples were diluted with PBS so that the final concentration of urea was 1.2 M. Trypsin/LysC mix (Promega) was added to digest the samples overnight at 35°C with agitation. Digests were desalted using Waters Oasis HLB μElution plate (2 mg sorbent per well, 30 μm particle size). Sorbent was conditioned with UPLC-grade acetonitrile and water containing 0.1% trifluoroacetic acid (Acros Organics), digests were loaded onto the sorbent, washed first with UPLC-grade water containing 0.1% trifluoroacetic acid, then with UPLC-grade water, and peptides were finally eluted with 70% acetonitrile in water. Peptide eluate was dried using a vacuum centrifuge, then resuspended in UPLC-grade water containing 3% acetonitrile and 0.1% (v/v) formic acid.

LC-IMS-MS data acquisition

In triplicate, 2 µL (~350 ng) of each sample was injected onto Waters nanoAcquity UPLC system fitted with a 5 µm C18 (180 µm x 20 mm) trap column and a 1.8 µm HSS T3 analytical column (75 µm x 250 mm). Peptides were loaded onto the trap column over 3 minutes, followed by analytical separation over a 105 minute gradient (10 minute trapping, 3% to 35% acetonitrile over 95 minutes), at a flow rate of 0.5 µL / minute and an average pressure of ~5800 psi. The lock mass compounds [Glu1]-Fibrinopeptide B and Leu-enkephalin (each ~100 fmol / µL) were delivered by the auxiliary pump of the LC system at 0.4 µL / min to the nanoLockSpray reference source of the mass spectrometer. The nano electrospray ionization source was set to 70° C and a 3 kV potential was applied. Peptides were analyzed in positive mode on a Waters Synapt G2-S HDMS traveling wave ion mobility time-of-flight (IMS-TOF) mass spectrometer. Precursor ions entering the instrument were separated by traveling wave ion mobility, using 40 V wave height and 600 m/s wave velocity. Ions between 50-2000 m/z exiting the ion mobility region were transferred to the TOF mass analyzer with a wave height of 4 V and a wave velocity of 380 m/s. Fragmentation energy was alternately applied post-ion mobility separation in the transfer cell. Alternating low and elevated-energy scans were performed every 0.5 seconds (0.05 s interscan delay) to obtain precursor and fragment information, respectively. In low energy MS mode, data were collected at constant collision energy (CE) of 4 eV across all 200 drift bins (TOF scans). In high-energy mode, no collision energy was applied in bins 0-19; a CE ramp from 17 to 60 eV was applied from bins 20-119, and a CE ramp from 60 to 72 eV was applied from bins 120-200. For mass analysis of precursor and fragment ions, the TOF was operated in V-mode with a typical resolving power of at least 20,000 FWHM. Data were mass corrected post-acquisition by comparison to the doubly-charged monoisotopic ion of [Glu1]-Fibrinopeptide B (m/z = 785.843), which was sampled every 30 seconds during acquisition.

Database searching and quantification

Raw data were analyzed with ProteinLynx Global SERVER version 3.0.2 (PLGS, Waters Corporation), searching against a database of the human proteome (downloaded from UniProt on February 11, 2016). The following criteria were applied in the search: (i) trypsin as digestion enzyme, (ii) one missed cleavage allowed, (iii) carbamidomethylated cysteine, Heavy ($^{13}C_6$, $^{15}N_2$ +8.014 Da) lysine, and Heavy ($^{13}C_6$, $^{15}N_4$ +10.021 Da) arginine as fixed modifications and methionine oxidation as a variable modification, (iv) a minimum of two identified fragment ions per peptide and a minimum of five fragments per protein, and (v) at least two identified peptides per protein. The false discovery rate for protein identifications was set at 1% using a decoy database of reversed protein sequences. Any identified peptides with a calculated mass error greater than 10 ppm were not considered.

DIA-SIFT: Extraction of y*-ion ratios and quartile filtering*

Precursor SILAC ratios from the PLGS search contained Light / Heavy ratios, but there was no information on y-ion ratios. Using a custom Python script, y-ion SILAC ratio measurements were extracted from the .csv files containing fragment results. The script searched for all heavy-labeled y-ion fragments, then compared the ion abundance of each heavy-labeled product ion to the intensity of the matching light-labeled product ion. The resulting y-ion Light / Heavy ratios were then added to a database already containing precursor Light / Heavy ratios from the peptide PLGS result files. The pooled precursor and fragment ratios from three technical replicates (replicate injections) were concatenated. A second script pooled all precursor and fragment Light / Heavy ratios for each identified protein. The distribution of Light / Heavy ratios for each protein was fit with quartiles, and outliers were filtered out according to an

optimized quartile width. Because the observed distribution of Light / Heavy ratios was unique for each protein, the absolute width of the quartile filter varied between proteins. The lower and upper bounds of Light / Heavy ratio measurements that passed the quartile filter for each protein were determined as follows:

$$lower = q1 - iqr\_factor(q3 - q1)$$

$$lower = q3 + iqr\_factor(q3 - q1)$$

where q1 is the first quartile (25th percentile), q3 is the third quartile (75th percentile), and iqr_factor was 0.2. This value for the iqr_factor was selected after testing values between 0 and 1. Values less than 0 were not tested in order to maintain a substantial portion (at least 50 %) of the total measurements for each protein. The Python scripts used to perform fragment ratio extraction and quartile filtering, along with instructions for use, can be found at github.com/martin-lab/DIA-SIFT.

Results and Discussion

Since SILAC light and heavy peptide pairs share common retention and ion mobility drift times, the paired peptides co-fragment, and the resulting *y*-ion fragment spectra include light and heavy peak pairs derived from the C-terminal arginine or lysine labels. In current IMS-DIA proteomic analysis workflows, these product ion pairs are automatically binned together by drift and retention time during peptide annotation (**Figure 17a,b**).

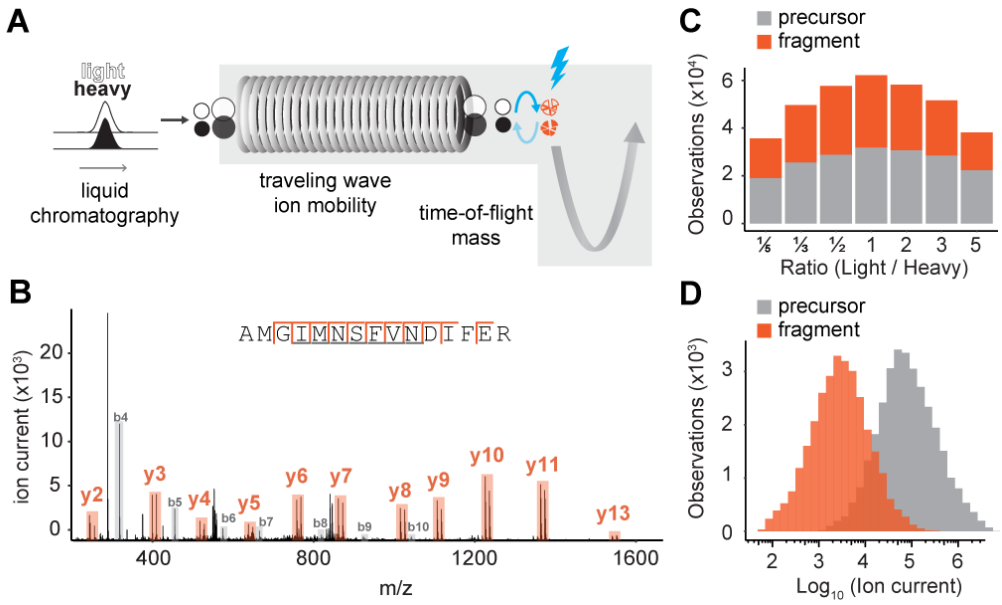

Figure 17. *y*-Ion quantitation increases measurable observations in DIA-SILAC analysis. (a) Schematic of SILAC IMS-DIA analysis. SILAC pairs share a common drift time and retention time for both precursor and product ions. (b) MS2 spectra of SILAC analysis (1:1) of histone 2B tryptic peptide demonstrating "Light" and "Heavy" (13C6, 15N4) *y*-ion pairs (shaded orange). (c) SILAC quantitation of *y*-ion ratios increases the number of quantifiable observations. Data represent aggregate observations across three replicates. (d) Fragment and precursor ion intensity profiles derived from SILAC IMS-DIA analysis.

To access these product ion ratios for quantitation, we developed a SILAC analysis pipeline to process output files generated by the APEX algorithm provided in the PLGS software (Waters).[73, 97] This post-processing algorithm retroactively extracts *y*-ion peak pair assignments, and since the DIA workflow samples peptide fragments across their chromatographic elution profile, the peak area can be quantified for all assigned *y*-ion pairs.

A series of human SILAC-labeled 293T cell tryptic digests of known Light / Heavy ratios were analyzed using a quadrupole IMS time-of-flight mass spectrometer (Synapt G2-S HDMS; Waters). Across different SILAC mixtures, MS2 *y*-ion ratio measurements nearly doubled the number of quantifiable observations, producing up to 60 000 quantifiable ratios summed across 3 replicates using a 105 min gradient (**Figure 17c**). Across this triplicate dilution series, 1254 ±

153 (SD) proteins were identified with an average of 19 precursor and 21 *y*-ion SILAC ratios per protein. Fragment ion intensities average 1−2 orders of magnitude lower than precursor ions, yet remain well within the dynamic range of the detector, allowing accurate quantitation (**Figure 17d**). By including these intrinsic *y*-ion SILAC ratios, on average, each peptide is measured ~5 times, providing inherent validation of the quantified precursor ratio.

Label-free MS2 quantitation of IMS-DIA data reportedly increases the accuracy of protein quantification, even when protein abundance spans a wide dynamic range.[98] In order to explore this further, we quantified the distribution of SILAC ratios across peptides from the different SILAC mixtures. In analyses of predefined mixtures spanning a 25-fold range, MS1 quantitation yielded a broad distribution of ratios with a 95% confidence interval for the coefficient of variation between 28 and 30% (**Figure 18a**), signifying extensive ion interference amplified by the increased sample complexity.

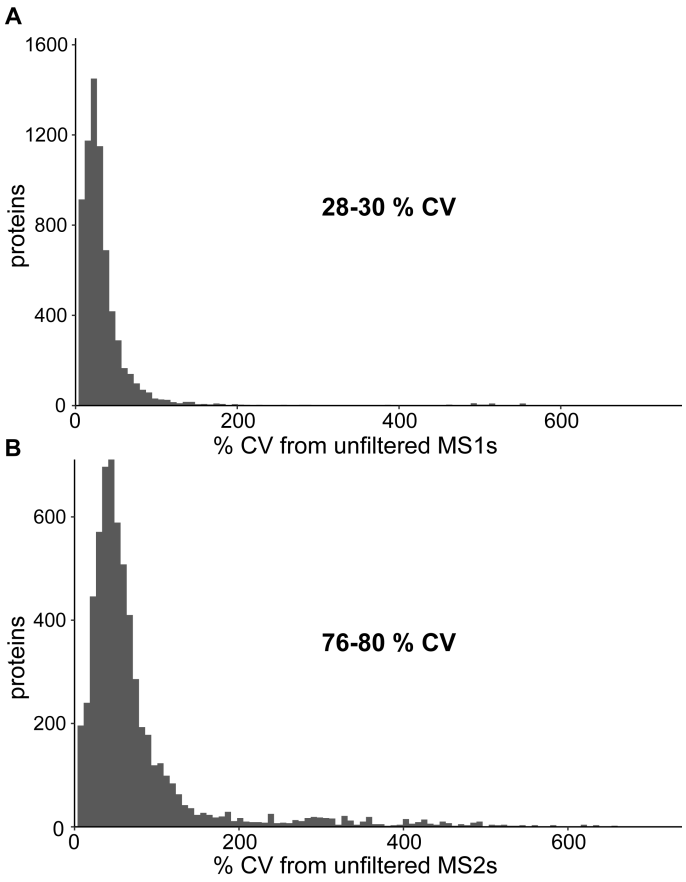

Figure 18. Distribution of precursor and *y*-ion SILAC ratios. Protein level CVs from (A) unfiltered precursors or (B) unfiltered fragments pooled across samples (293T cell lysates with Light / Heavy ratios 1/5, 1/3, 1/2, 1, 2, 3, 5). The 95% confidence interval is shown for each analysis.

In comparison, MS2 quantitation of *y*-ion ratios acquired with IMS-DIA methods yielded a broad distribution of measurements with a 95% confidence interval for the coefficient of variation between 76 and 80% (**Figure 18b**). However, the quantified ratios across individual proteins were largely Gaussian, with outliers amplifying the overall variance. These protein-level CVs are similar to retrospective analysis across several archived DDA experiments acquired on various Orbitrap instruments (**Figure 19**).[99-101] Raw data from 3 different SILAC experiments were downloaded from the ProteomeXchange[102] repository (http://www.proteomexchange.org/). Replicates were searched independently in Proteome Discoverer 2.2 using Sequest HT and

Percolator with the following search parameters: trypsin as digestion reagent, 1 missed cleavage allowed, 10 ppm precursor tolerance and 0.05 Da fragment tolerance, fixed carbamidomethyl cysteine, variable methionine oxidation, and appropriate isotopic labels. Percolator was used to filter identifications at 1% FDR. Precursor area (MS1) quantification was used for each analysis. Protein level % CV from pooled replicates for each experiment is shown, along with the respective 95% confidence interval (bold text).

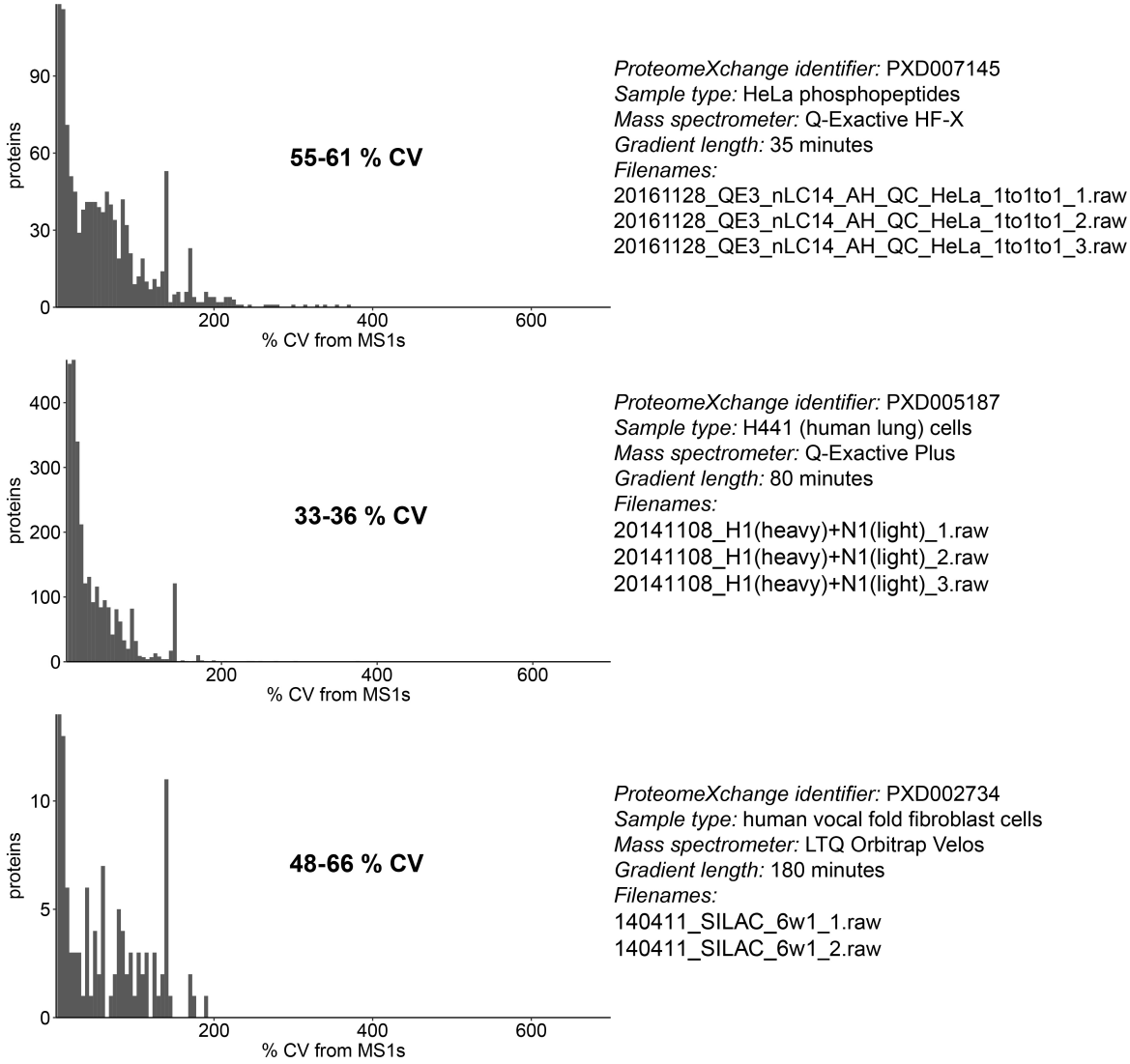

Figure 19. Distribution of precursor SILAC ratios obtained from three publicly available data-dependent (DDA) studies.

In both our DIA data and data from three publicly available DDA experiments, protein-level SILAC measurements exhibit significant coefficients of variation. To address this issue within our DIA SILAC data, we introduced a simple statistical filter, termed DIA-SIFT, to eliminate outliers from pooled MS1 and MS2 ratios (**Figure 20a**).

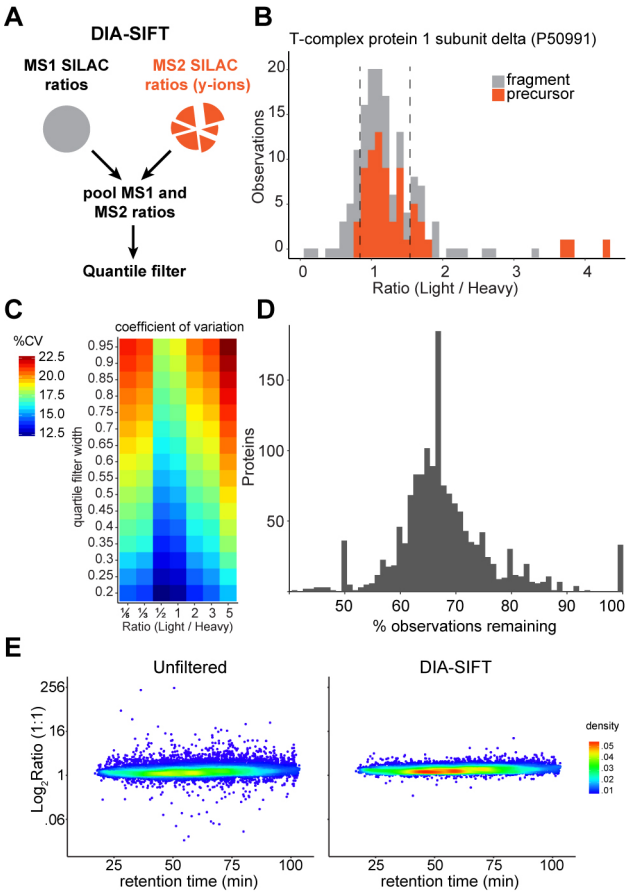

Figure 20. Quantile filtering of pooled MS1 and y-ion SILAC ratios improves accuracy. (A) Measured ratios are pooled together for quantile filtering of each protein. (B) Elimination of outlier ratios (those outside the dashed lines) by quantile filtering of pooled SILAC ratios for accurate protein quantitation. (C) Systematic analysis of quantile filter window relationship to the percent CV across the analysis at different defined SILAC mixtures. (D) Profile of observations remaining after quantile filtering MS1 and y-ion ratios by protein using a stringent 0.2 i.q.r. filter. (E) Quantile filtering reduces outlying peptide ratios to narrow the overall distribution of SILAC ratios toward the true value across the chromatographic gradient. Data compiled from three independent replicates. The color heatmap corresponds to a two-dimensional kernel density estimate.

Since MS1 and MS2 ratios are both intrinsic to the actual ratio, precursor and *y*-ion measurements are weighted equally. The pooled measurements for each protein are then quantile filtered, removing any statistical outliers from the overall protein measurement. For example, the chaperone protein CCT4 reports a largely Gaussian distribution of both precursor and *y*-ion ratios in a 1:1 (Light:Heavy) analysis. DIA-SIFT filtering removes interfered outliers while retaining sufficient measurements for accurate quantitation (**Figure 20b**). Pooled protein measurements are only filtered when there are at least three observations, and no quantified proteins are outright eliminated from the analysis.

MS1-based SILAC measurements are typically reported as median values, which can complicate later statistical analysis. Here, a sufficient number of measurements are retained to average any inconsistencies and more accurately reflect the true SILAC ratio. Applying a 0.2 inner quartile range (IQR) filter (see Experimental Methods in the Supporting Information) retains 67 ± 10% (mean ± standard deviation) of the observations in the final protein quantitation (**Figure 20c,d**), while reducing the coefficient of variation >4-fold across the SILAC dilution series. The selected filter thresholds balance stringency with retention of quantified observations. Outlier ratios are largely eliminated after quantile filtering and are equally distributed across the chromatographic gradient (**Figure 20e**). Overall, DIA-SIFT restores accurate protein quantitation in SILAC IMS-DIA analysis without removing any protein identifications.

Across the tested dilution series, DIA-SIFT dramatically improves accurate quantitation of fractional changes across the proteome (**Figure 21a**, **Figure 22**, and **Figure 23**).

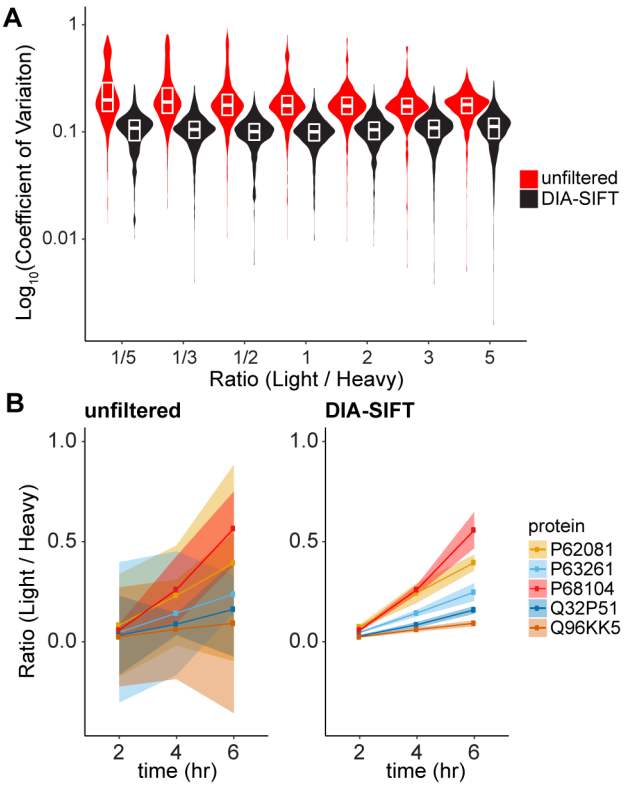

Figure 21. DIA-SIFT reduces variance to more accurately measure changes in protein levels. (A) Protein quantitative accuracy is improved by DIA-SIFT. Observed coefficients of variation across a defined SILAC dilution series are broad and harbor a number of outliers. DIA-SIFT reduces variation equivalently across the series. Violin plots illustrate the distribtion of ratios. The median, 25th and 75th percentiles are shown in the white box plot insets. (B) PulseSILAC labeling measurements are resolved by DIA-SIFT analysis. Five proteins are shown across three time points after cell growth media was switched from heavy to light. Shaded error represents the standard deviation for each protein.

For example, in a 1:1 mixture the mean protein-level coefficient of variation is reduced from 0.55 to 0.12, and the mean peptide-level variation is reduced from 0.24 to 0.06. The overall result is a reduction in outliers, which is essential for accurately measuring biologically significant changes.

Due to the larger observed variance of protein SILAC ratios when quantified from MS2-observations alone compared to MS1 measurements alone, we tested whether weighing MS1

observations more heavily would further improve quantification. In **Figure 23**, distributions of mean Light / Heavy protein ratios are shown across samples of different defined ratios when single, double, and triple MS1 weight was used with DIA-SIFT. Weighting MS1 ratios more heavily than *y*-ion ratios had no effect on the overall accuracy, highlighting the broad dynamic range of TOF-based SILAC quantitation. Overall, quantile filtering reduces the mean coefficient of variation to less than 15% across all samples across the dilution series.

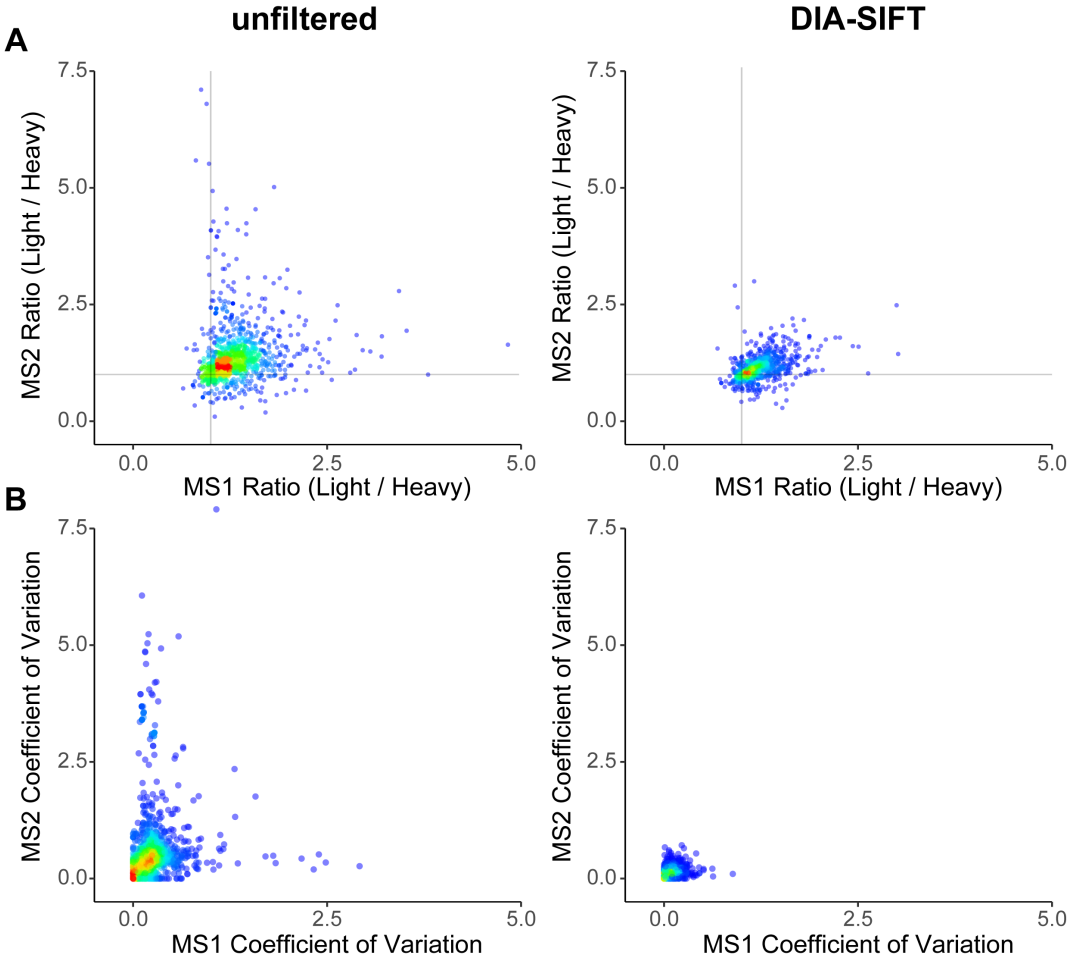

Figure 22. DIA-SIFT improves precursor and y-ion SILAC ratio correlation. (A) Correlation of precursor and y-ion ratios improve with DIA-SIFT processing. (B) The correlated coefficients of variation are reduced by DIA-SIFT processing. (1 point = 1 protein, color corresponds to two-dimensional kernel density estimate, which is calculated separately for each plot). Data shown are from a 1:1 (Light / Heavy) 293T lysate.

As further validation, human 293T cells initially grown in heavy SILAC media (with arginine +10 and lysine +8) were switched to light media and collected after different time intervals. The rate of heavy to light amino acid incorporation reports the ratio of old and newly synthesized proteins (**Figure 21b**). Unprocessed quantification data is highly variable and cannot be reasonably interpreted. After processing the data with DIA-SIFT, individual proteins are more accurately quantified to determine their rate of synthesis and turnover.

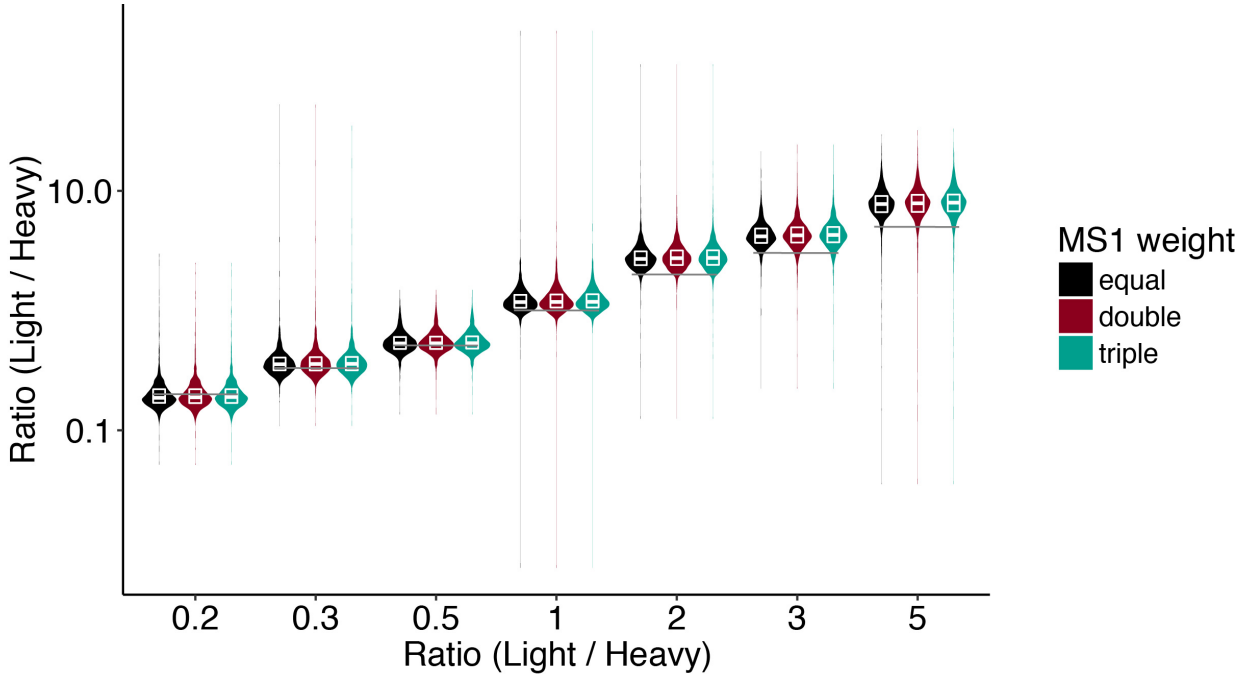

Figure 23. DIA-SIFT with increased weighting of MS1 observations. Observed ratio distributions (y-axis) are plotted against the true ratio (x-axis).

Overall, DIA-SILAC methods offer several unique advantages for proteomics data analysis. Recent NeuCoDIA methods also quantify *y*-ion ratios, leveraging neutron-encoded SILAC labels to avoid increased sample complexity.[103] Despite the promise of this approach, its practical application is limited by the long acquisition times required for high-resolution analysis (120 000 resolution at 200 m/z) of the neutron-encoded labels. In order to keep pace with the liquid chromatography gradient, only a narrow mass range (100 m/z) with four 27 m/ z-wide

SWATH windows is acquired per experiment. In comparison, IMS-DIA SILAC methods profile the entire mass range, collect more frequent MS1 and MS2 scans for better chromatographic peak definition, and yield significantly more quantifiable observations than NeuCoDIA analysis. Nonetheless, both DDA and DIA SILAC analysis reduce the total number of annotated proteins by ~20%.[75, 104] When accuracy is critical, this fractional loss in proteome coverage is less of a concern. In addition, the largest ratio changes are typically the most biologically interesting. Thus, implementing DIA-SIFT will reduce false positives and direct further biological validation to bona fide targets.

Here we demonstrate that DIA analysis yields *y*-ion ratio measurements that enhance SILAC quantitation. *y*-Ions can be easily identified as fragments pairs with matched drift and retention times. Since the *y*-ion pairs are matched using drift and retention time peaks, they represent multiple measurements and not single scans typical of DDA spectra. Beyond protein quantification, these *y*-ion pairs could also be leveraged to enhance confidence in peptide search algorithms, since they provide an orthogonal validation of accurate peptide fragmentation.[105] Implementing *y*-ion pair statistics in label-assisted de novo sequencing (LADS) algorithms could further enhance peptide annotation, quantitation, and reproducibility. Overall, DIA-SIFT algorithms can be adapted for SWATH analysis of SILAC or NeuCode-labeled peptides, providing a simple filter to leverage *y*-ion ratios for more accurate quantitation.

Chapter 4 Optimization of open-source LC-IMS-MS proteomics data analysis software

Introduction

Data-dependent acquisition is the current standard method in LC-MS protein annotation and quantification. In high complexity samples, limited duty cycle and co-isolation during abundance-based ion selection often results in under-sampling. In contrast, all-ion fragmentation (data-independent acquisition) methods bypass peptide ion selection, alternating low and high collision energies to analyze precursor and fragment ions across wide mass ranges. Traveling wave ion mobility spectrometry (IMS) adds an additional dimension of separation, reducing interference of precursor ions to improve reproducibility, peak capacity, and peptide identifications. Despite the advantages of ion mobility separation, these multidimensional proteomics data are not compatible with the majority of database search engines.

While the software tool TWIMExtract[95] (Chapter 2) can extract LC-IMS-MS features in any of the three dimensions from a predefined list, it cannot discern these features directly from the raw data file. In order to perform a database search on LC-IMS-MS data, each peptide and fragment feature must be detected across all three dimensions of separation to retain the full information content of the acquired data. Outside of the proprietary software packages offered by the Waters Corporation, there are no tools to analyze bottom-up LC-IMS-MS proteomics data.

This chapter presents the initial testing of an informatics pipeline for peak extraction and correlated peptide and fragment analysis from data-independent liquid chromatography-ion mobility-mass spectrometry (LC-IMS-MS) data. Peak detection from raw LC-IMS-MS data occurs across the three dimensions.[106] Features are first de-isotoped (monoisotopic mass is

determined from the isotopic distribution of each feature, see **Figure 4**), followed by grouping of fragment ions (from high energy scans) to their respective precursors (intact peptide ions from low energy scans). The data are then exported as standardized file format (.mgf), which is compatible with most proteomics search engines.[107]

This chapter details parameter optimization studies of an open-source data analysis tool previously published for IMS-MS studies of intact proteins. Raw LC-IMS-MS proteomics data quality can be highly variable, fluctuating based on sample complexity, instrument tuning, calibration quality, and liquid chromatography performance. For this reason, there is a need to empirically test the use of different parameters within the software to identify those will perform well for LC-IMS-MS proteomics data.

Data generated in a number of different laboratories was used to test the software parameters for feature detection and precursor-fragment grouping. This chapter reports a high-throughput approach to test many unique combinations of parameters that control the peak detection and precursor-fragment grouping steps of LC-IMS-MS proteomics data analysis. The findings from this chapter can be used to guide users of the open-source tools in parameter selection to improve their data analysis.

Using this pipeline, which is shown in **Figure 24**, we can successfully generate a common proteomics file format from liquid chromatography-ion mobility-mass spectrometry data. These files can be fed into almost any database search engine, and for these studies we selected MSFragger[45] to perform ultrafast database searching, and Philosopher, which integrates PeptideProphet and ProteinProphet for statistical validation with convenient reporting tools. This pipeline provides the first open-source approach towards transparent and flexible data analysis from the Waters LC-IMS-MS platform.

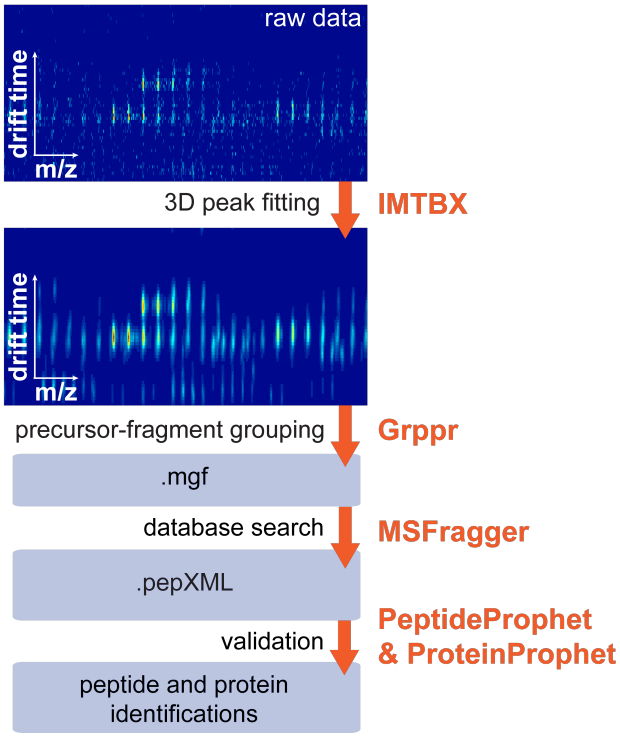

Figure 24. Overview of LC-IMS-MS data analysis pipeline testing. IMTBX is used to extract features from three dimensions, and Grppr correlates precursors and fragments. MSFragger is used to perform database searches, and Philosopher (which integrates PeptideProphet and ProteinProphet) is used to validate database search hits.

Experimental

To test the general utility and robustness of the pipeline for analyzing data across different laboratories and sample types, six '.raw' data files were tested (**Table 3**). Three of these files were downloaded from the publicly accessible proteomics data repository, proteomeXchange.[102] Two of the remaining three files were generated and analyzed in Chapter 2,[95] and the third was provided by Waters via personal communication. The raw files tested each consisted of a human protein sample analyzed by the MSE acquisition method[108] over reversed-phase liquid chromatography gradients ranging in length from 50 to 120 minutes. Only one of

these raw files (sample label CRC) was subjected to pre-fractionation prior to LC-MS analysis, the remaining five are whole, unfractionated human cell digests.

Table 3. List of raw data files used to test the data analysis pipeline. The sample label is used to refer to each file throughout this chapter.

proteomeXchange identifier	citation	sample type	instrument model	sample label	gradient length (min)
PXD005735	Quesada-Calvo et al. Clin. Proteom. 2017.[109]	Human serum samples, 1/5 high-pH reversed phase fractions	Synapt G2	**CRC**	80
PXD005551	Piltti et al. Sci. Rep. 2017.[110]	Human chondrosarcoma 2/8 cell culture	Synapt G2-Si	**HCS28**	120
(N/A)	Haynes et al. Anal. Chem. 2017.[95]	Human HeLa standard digest	Synapt G2-S	**H1000**	80
(N/A)	Haynes et al. Anal. Chem. 2017.[95]	Human HeLa standard digest	Synapt G2-S	**H600**	80
(N/A)	(N/A)	Human HeLa digest from Waters for ABRF study	Synapt G2-Si	**Waters**	95
PXD004696	Cassoli et al. Proteomics 2017.[111]	Human MO3.13 oligodendrocyte cell culture	Synapt G2-Si	**MO3**	50

The testing pipeline was composed of three major parts, (1) raw feature extraction and grouping, (2) database searching with MSFragger,[45] and (3) validation of search results using PeptideProphet[42] and ProteinProphet[112] via Philosopher. The first step in the pipeline, feature extraction and precursor-fragment grouping, was performed by the combination of IMTBX / Grppr, which have been applied previously for intact protein analyses.[106] More detailed descriptions of the IMTBX / Grppr tools can be found in the original publication and at https://dmtavt.com/IMTBX.

Peak detection from raw LC-IMS-MS data

IMTBX is written in C# (.NET 4.5.1), and was operated in the 3D peak detection mode. Version 4.4.1 was used for the studies reported in this chapter. A full IMS scan is represented as a sparse matrix of N rows and M columns, where N is the number of IM drift bins (typically 200), and M is the number of bins for the m/z axis (typically 100000−300000 bins). Each of these IMS scans take approximately 1 second to acquire during the LC-MS analysis. Low and high-energy IMS scans are alternated during acquisition, and are separated from each other (stored as function 1 and function 2, respectively) in the raw file.

To process the LC-IMS-MS data, a 3-D Gaussian filter of user-specified size and shape is applied to smooth the data. The size and shape of the Gaussian filter corresponds to the number of neighboring scans taken into account and their weighting, respectively. A lone data point removal step, controlled by user-specified parameters, is then used to filter out nonzero intensity data points that only have few neighboring nonzero points. Signal-to-noise ratio thresholds and absolute intensity cutoffs can also be specified by the user for each function. Local maxima in each scan are found and used as seed points for fitting the peaks with 3D Gaussians. All detected features from function 1 and function 2 are written to .txt files for precursor-fragment grouping.

Peaks in the m/z spectra throughout the LC-MS file can be calibrated post-acquisition by comparison to the lockmass spray, which is sampled every 30 seconds throughout the acquisition process. All six raw files included in this study were calibrated using the doubly-charged monoisotopic peak of [Glu1]-fibrinopeptide-B (785.8426 m/z).

Precursor-fragment grouping

Fragment ions from function 2 were correlated to precursor ions from function 1 using Grppr version 0.3.17. Grppr takes as input the set of 2D function 1 and function 2 peaks containing peak intensities as well as locations and widths in both dimensions (m/z and drift time). The 'Averagine' algorithm was used for isotopic cluster detection, which uses the Averagine[113] model amino acid to calculate theoretical isotopic distributions for each peak to compare to the data. This algorithm determines the charge and monoisotopic mass of each feature, which is then recorded to a .mgf file for database searching.

Ions in function 2 can often be found with shifted drift times relative to function 1. Grppr provides adjustable tolerances for the maximum allowable width of this drift time shift. For these studies, the maximum drift shift allowed was 3.5 bins. The minimum number of peaks required to write a cluster to the output file can also be specified to reduce output file size. A minimum of three fragment peaks was used.

Database searching

Prior to database searching, .mgf files output from IMTBX / Grppr were converted to the .mzXML format. The msconvert tool from ProteoWizard package[89] was used with default parameters. ProteoWizard can be downloaded from http://proteowizard.sourceforge.net/download.html.

Database searches were performed using MSFragger.[45] This search algorithm features an optimized fragment ion indexing method that makes it roughly two orders of magnitude faster than other tools. While it is well-suited for the challenging task of 'open' searching, where

allowed mass tolerances can be of hundreds of Daltons, it performs well for closed searches, and its speed was crucial to the feasibility of this study.

A database of reviewed human protein sequences (20410 sequences downloaded from Uniprot[1] on October 29, 2018) was used. Trypsin was specified as the digesting enzyme, with 2 missed cleavages allowed. Fixed carbamidomethyl cysteine was specified as a fixed modification, and methionine oxidation and acetyl N-termini were specified as variable modifications, with a maximum combination of variable modifications for each peptide set at 5000. Closed searches were performed, with precursor and fragment mass tolerance of 50 and 40 ppm, respectively. Standard ^{13}C isotope error was allowed. The minimum and maximum lengths of peptides searched were 7 and 50 residues, and the mass range was 500 to 5000 Da. A minimum of 3 peaks per spectrum was required to be included in the search, and only the top 100 peaks from each spectrum were used.

The March 16, 2018 release of MSFragger was used. For the work described in this chapter, a standalone version of the tool was run as part of the testing pipeline, but a user-friendly interface has since been developed for more routine use. This tool, called FragPipe, can be found at https://nesvilab.github.io/FragPipe/ along with instructions for use.

Validation of peptide and protein identifications

Database search hits, in the form of '.pepXML' files resulting from MSFragger analyses, were validated using PeptideProphet[42] and ProteinProphet[112]. Validation is crucial to deriving meaningful results from resulting lists of peptides and proteins, particularly in this work, where changes to feature detection parameters will significantly influence the quality of the resulting spectra and thus the distributions of search scores. PeptideProphet and ProteinProphet validation

helps ensure that the peptide identifications can be compared as raw data processing parameters are changed.

PeptideProphet models the distributions of search scores, mass error, and other attributes of every peptide-spectrum match (PSM) returned by MSFragger. This includes those matches to the reversed protein sequences, which are necessarily incorrect. Using machine learning techniques to find a model that best distinguishes between correct and incorrect peptide assignments, PeptideProphet assigns a probability of being correct to each PSM. Using similar machine learning techniques, ProteinProphet is then used to assign probabilities that each identified protein is truly present in the sample. From these probabilities, lists of PSMs and identified peptides and proteins can be reported.

A tool called Philosopher was used to run PeptideProphet and ProteinProphet, facilitating automation of peptide and protein validation. Philosopher was also used to perform the reporting step, taking the validated peptide and protein outputs and generating .tsv files, which were used for all downstream data analysis. While Philosopher was only used for validation and reporting in these studies, it can also be used as an all-in-one data analysis toolkit for both open and closed proteomics searches. Philosopher is currently unpublished, but it can be downloaded at https://nesvilab.github.io/philosopher/.

High-throughput testing approach

To find optimal parameters for IMTBX across the six different raw data files, a range of values for each of 12 different parameters were tested. These 12 parameters and the ranges of values tested for both IMTBX can be seen in **Table 4**.

Table 4. IMTBX parameters tested. Minimum and maximum values, increment size, and the resulting number of possible values for each parameter are shown.

parameter	min	max	increment	# possible
mz	1	3	1	3
im	1	3	1	3
rt	1	3	1	3
hyper_mz	0.2	0.8	0.1	7
hyper_im	0.2	0.8	0.1	7
hyper_rt	0.2	0.8	0.1	7
ln	4	14	1	11
lr	1	8	1	8
npts	7	12	1	6
noiseWnd1	0.5	3	0.5	6
noiseWnd2	1	6	1	6
snr	1	3	0.2	11

The first three parameters (mz, im, and rt) in **Table 4** control the width of the Gaussian filter in terms of points in each dimension (m/z, ion mobility, and retention time, respectively). The next three parameters (hyper_mz, hyper_im, and hyper_rt) control the shape of the Gaussian filter in each dimension. The values tested, between 0.2 and 0.8, correspond to weight of the farthest filter point in each direction. The next three parameters (ln, lr, and npts) are used to distinguish true peaks from noise. The minimum number of non-zero neighbors is controlled by ln, and the taxi-cab radius in which this criterion applies is lr. An overall threshold is provided by npts, which is the minimum number of points needed for a peak to be reported. noiseWnd1 and noiseWnd2 are the widths of the noise estimation window in the m/z dimension. Finally, a signal-to-noise threshold is applied using snr.

From these ranges, there are approximately 1.9 billion possible combinations of parameters. Testing all combinations is not feasible, so a scheme was devised to sample this 12-

dimensional parameter space. A Halton sequence[114] was used to generate quasi-random combinations of values of these 12 parameters that were evenly sampled relative to a more random approach.

A wrapper script written in Python 3 was used to run IMTBX, Grppr, msconvert, MSFragger, and Philosopher in a semi-automated manner on a desktop computer (2-core, 96 GB RAM).

Results and Discussion

A single test involving peak detection, precursor-fragment grouping, file conversion, database searching, and validation from a full LC-IMS-MS raw file took between 2 and 5 hours to complete, and required 5 to 15 GB of storage space. A method of sampling only a section of scans from the center of the raw file was used to facilitate higher throughput sampling of the parameter space.

To validate this approach, 100 IMTBX parameter combinations were used to extract sections of variable sizes from the center of the LC-IMS-MS run. The center of the gradient was chosen because the highest density of peptides elute at this time, and the peptide features in the middle of the gradient are representative of average peptides. **Figure 25** demonstrates that the number of peptide-spectrum matches (PSMs) obtained from the small section (250 scans, or approximately 5 minutes) scaled linearly to those from a much larger section (2000 scans, or approximately 40 minutes) from the center of the file. Peptide and protein identifications also scaled linearly, with correlation coefficients of 0.84 and 0.76, respectively.

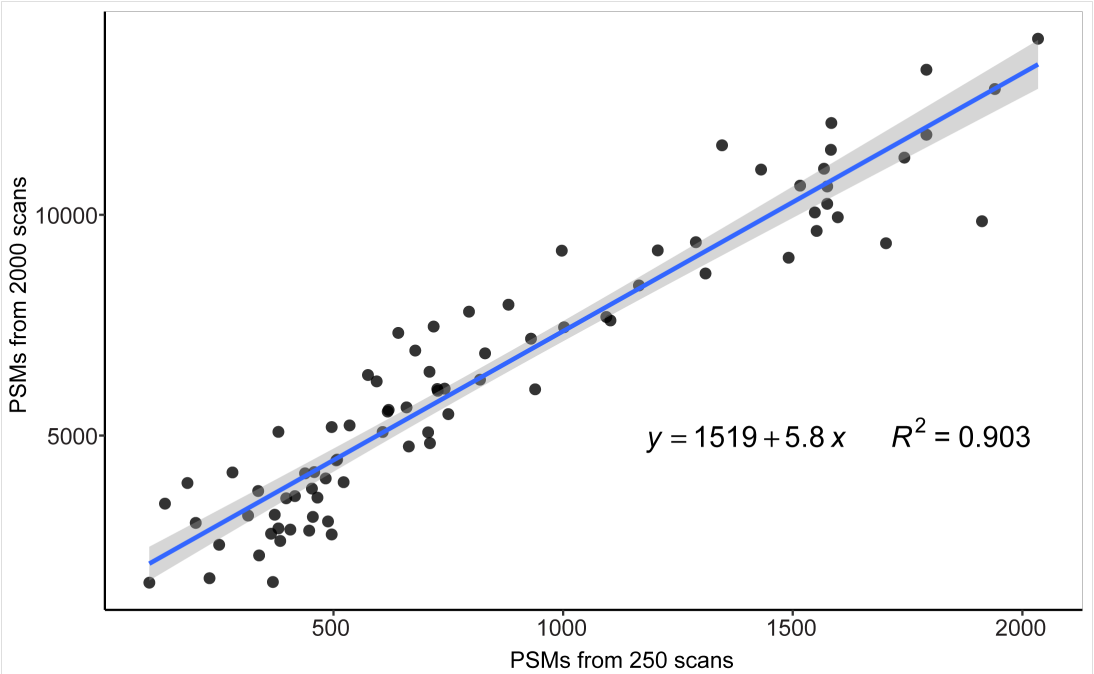

Figure 25. PSMs obtained between 5 minute (250 scan) and 40 minute (2000 scan) sections of an LC-IMS-MS raw file. Points represent the number of peptide-spectrum matches from 100 unique IMTBX parameter combinations tested on raw file H1000 (80 minute total gradient length). OLS line of best fit is shown in blue with the 95% confidence interval shaded gray.

Because the number of identified features scaled well with increasing amount of the raw file used, the 5 minute (250 scan) sections were used to significantly increase throughput of parameter testing. Prior to large-scale testing of IMTBX parameters, the criteria for accepting isotopic clusters was first tested on a smaller scale. The other Grppr options were used as specified in the Experimental section. In Grppr, the minimum number of peaks required to identify an isotopic cluster (see **Figure 4**) can be specified separately for function 1 (peptide) and function 2 (fragment) spectra. Using 100 unique sets of IMTBX parameters generated according to **Table 4**, three different options were tested (**Figure 26**).

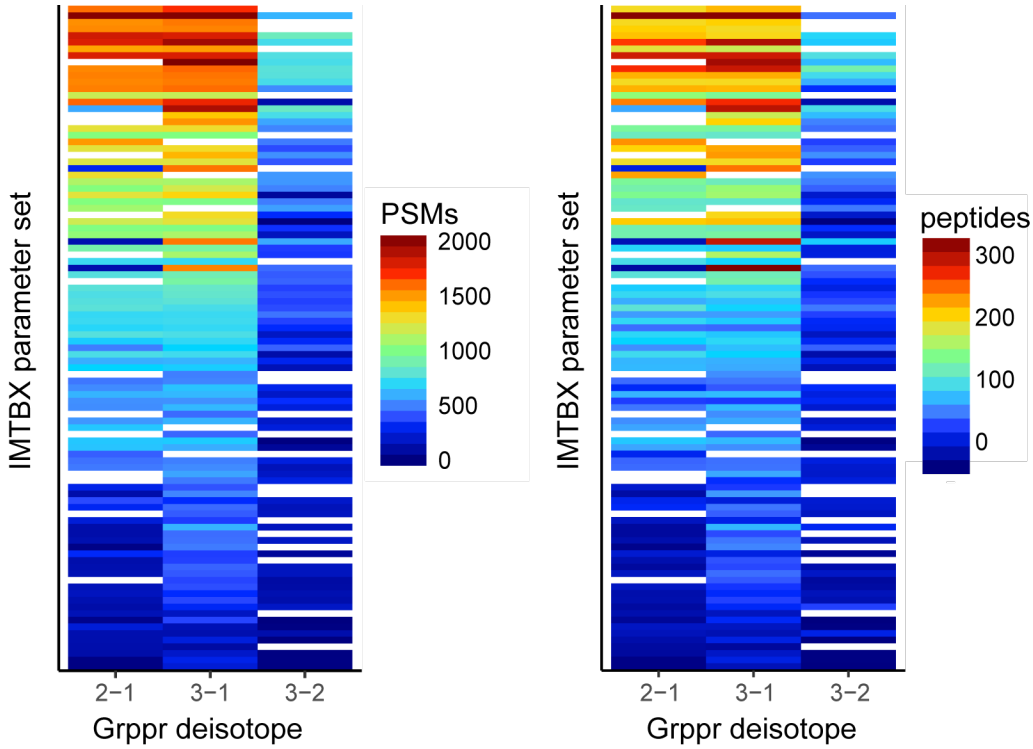

Figure 26. Peptide-spectrum matches (left) and unique peptide sequences (right) observed with three Grppr deisotope options for 100 parameter sets. Each colored bar indicates the number of features identified for one IMTBX parameter set according to each color scale (PSMs and peptides, respectively).

For each option tested (2-1, 3-1, 3-2), the first number in the pair (2, 3, 3) indicates the number of isotopic peaks required to accept a cluster from function 1 spectra, and the second number in the pair (1, 1, 2) indicates the number of isotopic peaks required to accept a cluster in function 2. The numbers of parameter sets that failed to return any identified features were 14, 5, and 21, from deisotope options 2-1, 3-1, and 3-2, respectively. The numbers of peptide-spectrum matches were 722, 867, and 331; and the numbers of unique peptide sequences identified were 106, 130, and 50 for deisotope options 2-1, 3-1, and 3-2, respectively. Because option 3-1 had the fewest number of failed parameter sets and the highest average number of PSMs and peptide identifications, this option was used for the subsequent large-scale test of IMTBX parameters.

For the large-scale test, a total of 5000 unique IMTBX parameter combinations generated using the Halton method (see Experimental section) were tested using five minute sections from the center of each of the six LC-IMS-MS raw files. The numbers of peptide-spectrum matches and unique sequences identified for each of these is shown in **Figure 27**.

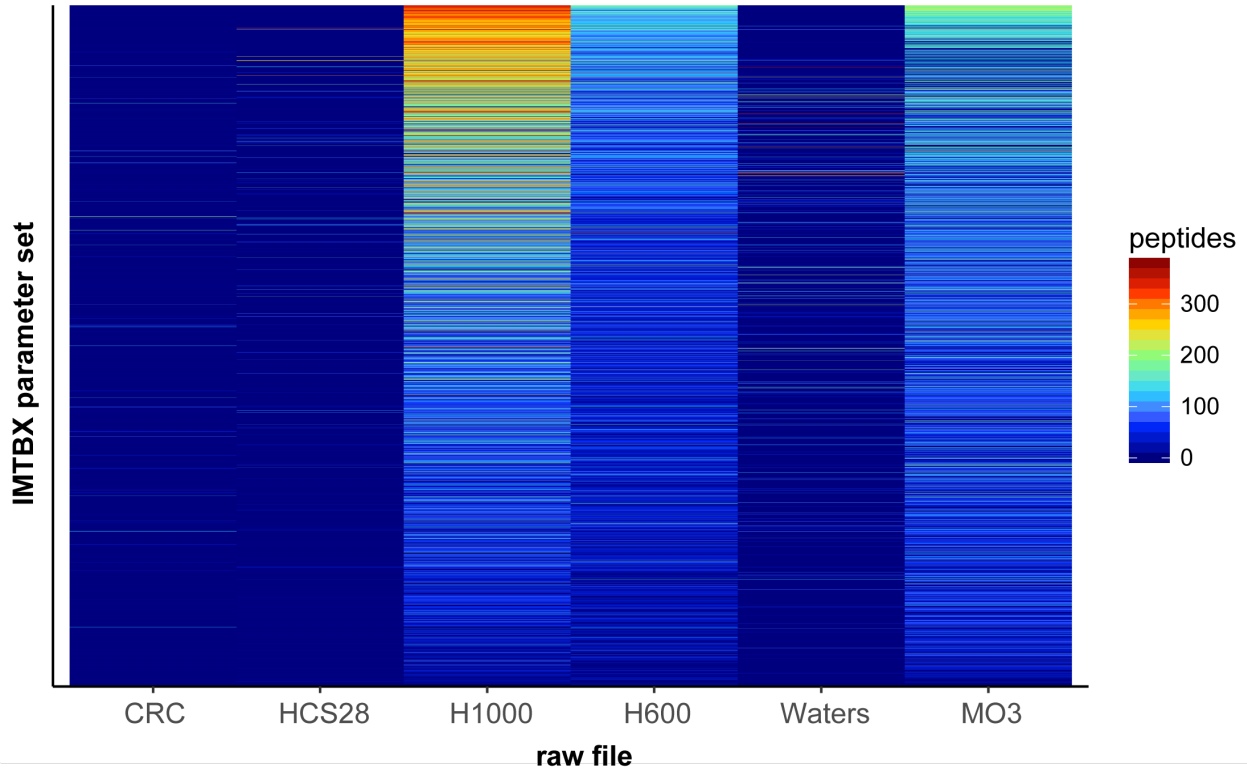

Figure 27. Unique peptide identifications obtained from each raw LC-IMS-MS file using 5000 different IMTBX parameter sets. Color indicates the number of sequences identified for each test according to the color scale at right.

The raw files returned very different numbers of identified features, which is to be expected given the different sample types and three different versions of the Synapt (G2, G2-S, and G2-Si). The H1000 file returned the most peptide identifications (87 sequences on average), and the CRC file returned the lowest (an average of 1.8 sequences). Overall, the top-performing parameter sets for the H1000 file tended to also work well for the H600 and MO3 files (**Figure**

27). It is likely that the CRC file returned so few peptides because the originating sample was pre-fractionated serum, which would be expected to contain fewer unique peptides overall. The CRC file was also the only one acquired on the oldest instrument model of the three, the Synapt G2. Behind the H1000 file, MO3 returned the second most peptide identifications (49 sequences on average), and this data was acquired on the newest instrument version, the Synapt G2-Si.

The highest performing parameter set from the 5000 that were tested was 1-1-1-0.7-0.2-0.5-5-7-12-1.5-2-2.2 (in the order mz-im-rt-hyper_mz-hyper_im-hyper_rt-ln-lr-npts-noiseWnd1-noiseWnd2-snr). The full raw file H1000 (80 minute gradient of a standard HeLa digest acquired on a Synapt G2-S) was extracted using these IMTBX settings and compared to the Waters proprietary software, ProteinLynx Global Server (PLGS) version 3.0.2.

The PLGS software was used to both for feature detection and database searching. PLGS can generate the .mgf file format to use with other database searches, with options for intensity thresholds in function 1 and function 2. In addition to searching the raw file H1000 to obtain peptide and protein reports, PLGS was also used to generate an .mgf file (with default intensity thresholds) to run through the database searching and validation portion of the analysis pipeline for side-by-side comparison.

For these comparisons, the database search was performed on the full H1000 file according to the methods specified in the Experimental section. The PLGS search was performed using the same mass tolerances as the MSFragger searches (50 ppm and 40 ppm for peptide and fragment spectra, respectively), with three minimum fragment ions to identify a peptide and at least two peptides required to identify a protein. A global false discovery rate was set to 1% in PLGS using the reverse sequences appended to the protein database.

The peptide and protein results from these comparisons are shown in area-proportional Venn diagrams (**Figure 28**).

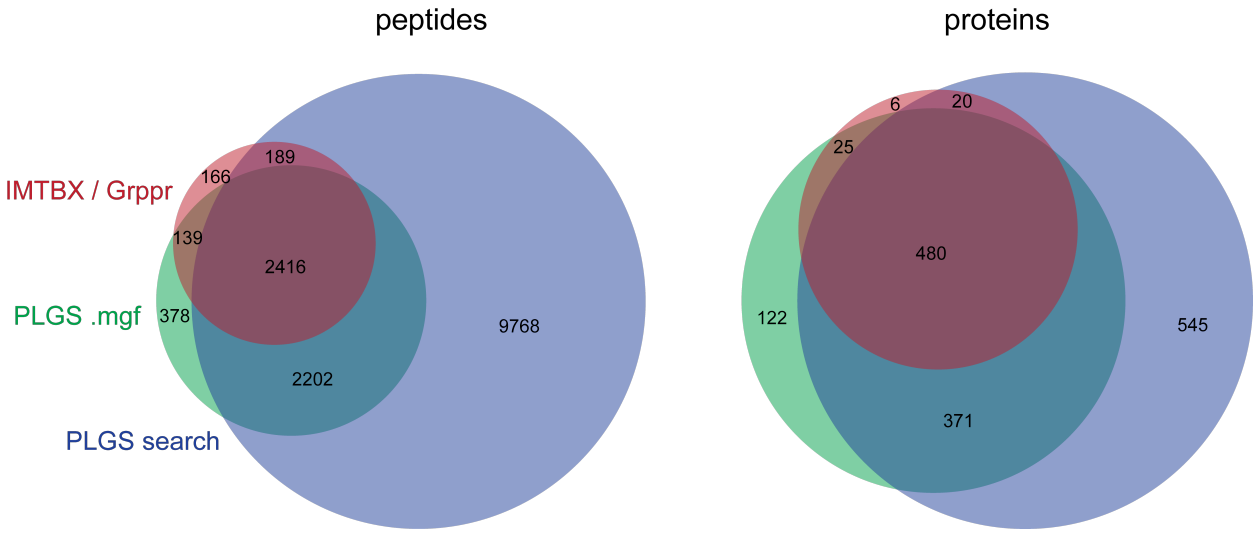

Figure 28. Peptide and protein identifications from the best-performing IMTBX / Grppr parameters (red), the .mgf file generated using PLGS (green), and a PLGS search (blue). The numbers of identifications are represented proportionally by the area of each region of the Venn diagram.

The IMTBX / Grppr analysis pipeline returned only 57% and 20% of the total unique peptide sequences and 53% and 38% of the total proteins identified from the PLGS .mgf and the PLGS search, respectively. While the PLGS search significantly outperformed the MSFragger and Philosopher search and validation pipeline in terms of the numbers of identifications, the proprietary software may not include rigorous statistical validation of the database search hits. PLGS does not report the same search scores as most other search engines, so these scores cannot be directly compared to those from MSFragger. Distributions of selected search scores (hyperscore, nextscore, and expectation value) from validated peptide identifications from IMTBX / Grppr and PLGS .mgf are shown in **Figure 29**.

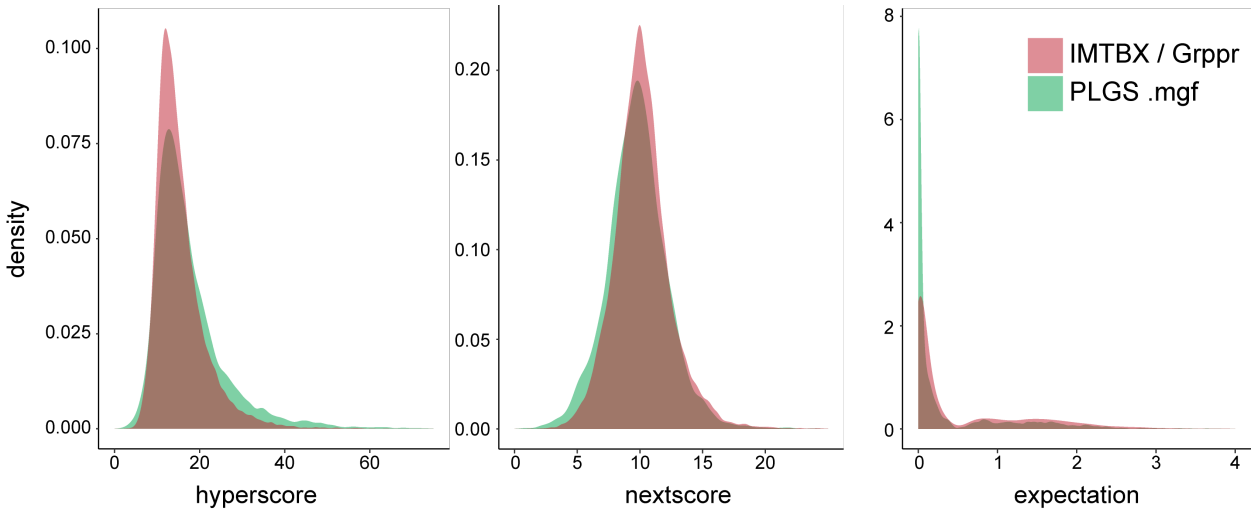

Figure 29. Comparison of MSFragger search score distributions of validated peptide-spectrum matches from IMTBX and PLGS. Density plots are shown, where the shaded areas are normalized (equal) between the IMTBX and PLGS data.

Briefly, hyperscore takes into account the number and intensities of matched fragment ions. Nextscore is the score of the second-highest scoring peptide match for a given spectrum. The expectation value is derived from these other search scores, and indicates the number of peptides with scores equal to or better than the observed score that would be expected if peptides match the experimental spectrum purely by random chance.

The hyperscore and nextscore distributions between PLGS and IMTBX feature detection (.mgf files) are similar, but peptides identified with PLGS have significantly lower expectation values (0.34 vs 0.51 on average). The average charge of a validated peptide-spectrum match was slightly higher when PLGS feature detection was used compared to IMTBX (2.23 and 2.17 average charge, respectively).

Compared to the PLGS-generated .mgf file, IMTBX / Grppr was only able to extract about half of the peptide and fragment features overall. However, given that only 5000 out of the over 9 billion possible parameter combinations were sampled, it is likely that the performance of the IMTBX / Grppr pipeline could still be improved.

While there were too few individual raw files tested to discern any real trends in the performance of IMTBX for data acquired using different instrument models or sample types, the results of sampling 5000 different parameter combinations were used to attempt to predict more optimal combinations. To fit the 12-dimensional test data, three different modeling approaches were used: multiple linear regression, multivariate adaptive regression splines[115, 116], and gradient boosted regression[117-120]. To test each of the three modeling approaches, 80% of the 5000 parameter sets was used to train, and 20% of the data was used for testing. Models were evaluated based on the coefficient of determination between the true number of peptides identified and those predicted by the model.

Multiple linear regression was performed with the LinearRegression function from the scikit-learn package in Python with default settings. Comprehensive parameter optimization of the multivariate adaptive regression splines (MARS) and gradient boosted regression models was performed to provide the best fit for each of these models. The Python implementation of MARS (found in the py-earth package https://contrib.scikit-learn.org/py-earth/) was used with the following parameters: fast_K=3, endspan_alpha=0.001, minspan_alpha=0.001, allow_missing=True, max_terms=50, max_degree=50, and penalty=3.0. Gradient boosted regression was performed with the GradientBoostingRegressor function from scikit-learn with the following parameters: learning_rate=0.01, max_depth=3, min_samples_leaf=3, max_features=1.0, subsample=1.0, loss='ls', min_weight_fraction_leaf=0.1. The coefficients of determination (R^2 values) for the linear regression, MARS, and gradient boosted regression predictions from five independent train/test split trials were 0.21 ± 0.01, 0.23 ± 0.01, and 0.24 ± 0.01 (mean ± standard deviation), respectively. Both the MARS and gradient boosted regression models performed better than the simple linear regression model. Though it gave a marginally

better fit than MARS, gradient boosted regression yielded the highest R^2 value and thus provided the most accurate model of IMTBX parameter performance.

From the gradient boosted regression model, the number of peptide identifications was predicted for each of the IMTBX parameter combinations in the testing portion of the dataset. The results of these predictions are plotted against the actual number of peptides identified from the IMTBX parameter set for all six raw LC-IMS-MS raw files together (**Figure 30a**). The predictive power of the model, which is quite low when all raw files are examined together, is significantly increased when only one raw file is used (**Figure 30b**).

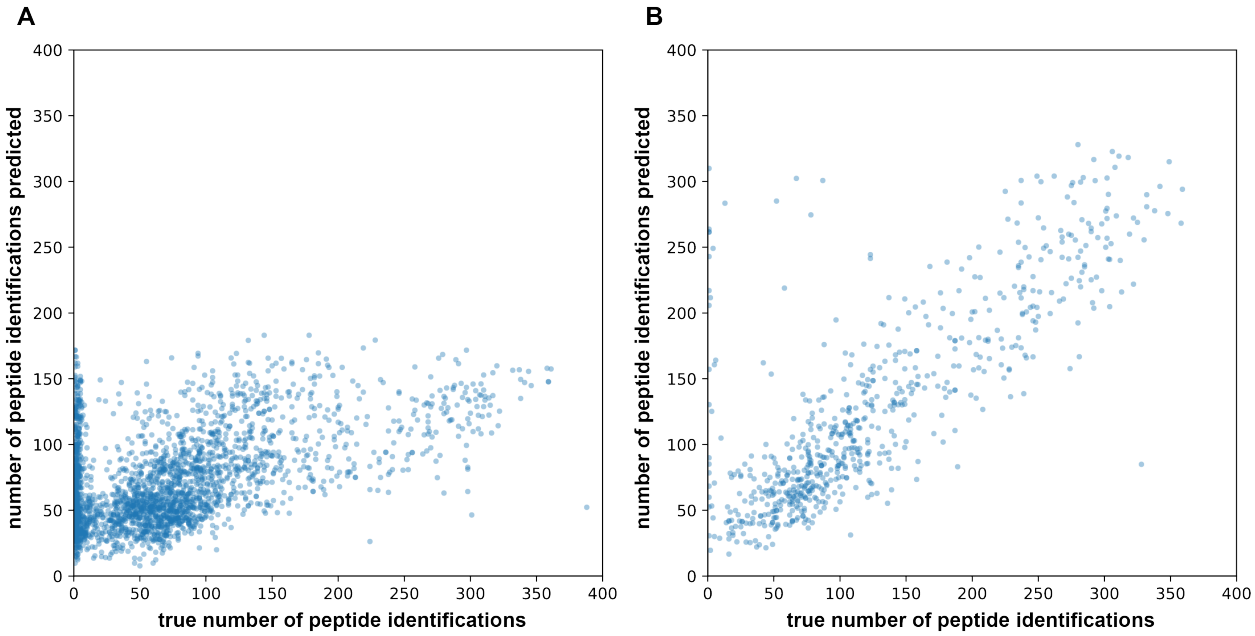

Figure 30. Predicted number of peptide identifications as a function of the true number of peptide identifications for each IMTBX parameter combination. Predictions based on results of the 5000 different IMTBX parameter sets (a) from all six raw files, and (b) from file H1000 alone.

The average cross-predicted score (coefficient of determination of the predicted vs. actual results) was 0.24 ± 0.01 (n=5 individual train/test splits) when all files were used, compared to 0.70 ± 0.03 when only H1000 was used to train and test the model.

Using the fit from the gradient boosted regression model of all six raw files, response in terms of peptide identifications across the range of values tested for each of the 12 IMTBX parameters is shown in **Figure 31**.

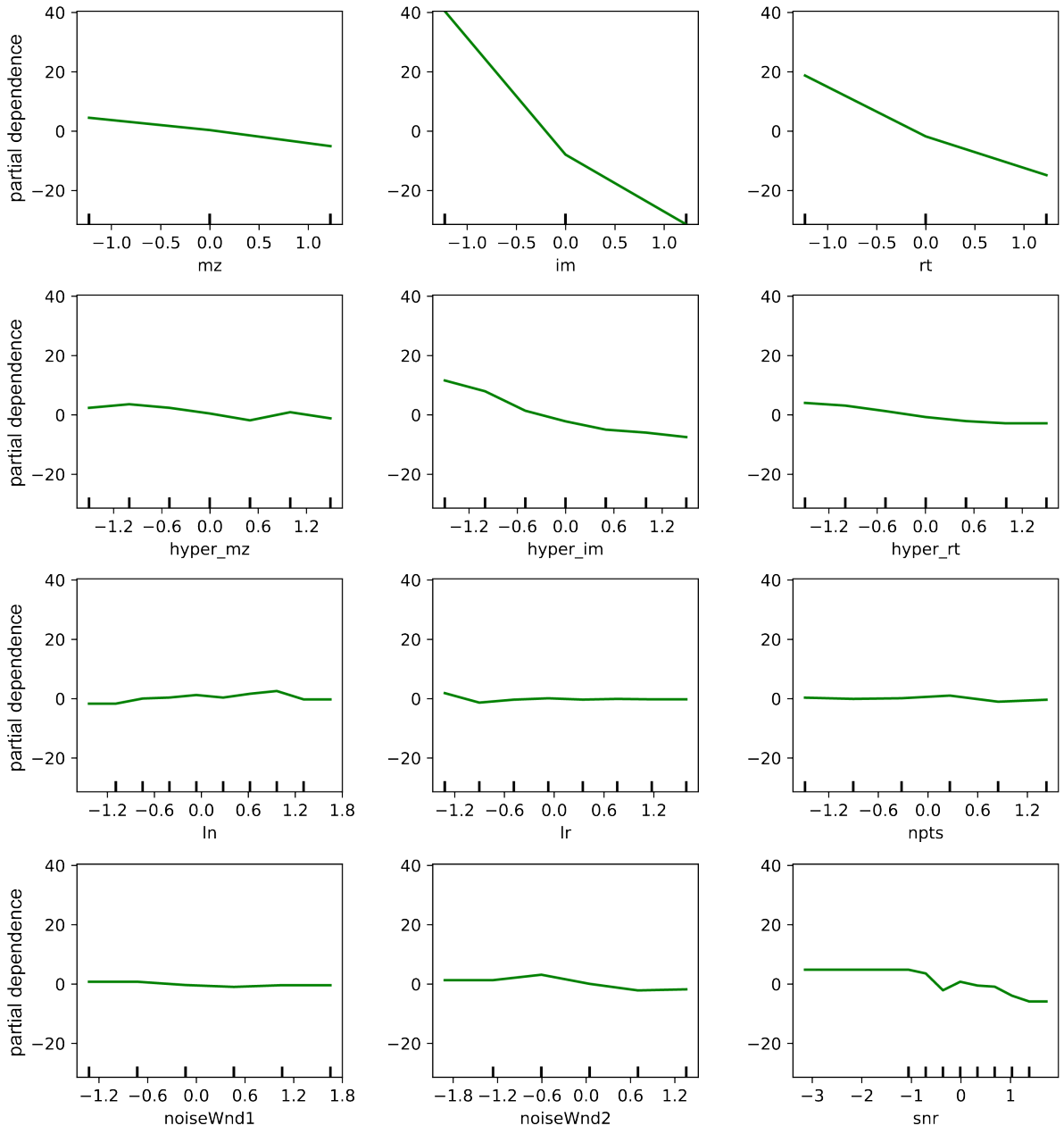

Figure 31. Peptide identification response across the values tested for each IMTBX parameter. Partial dependence values (y-axis) indicate the change in peptide identifications relative to the performance of the average parameter set.

Some significant relationships can be observed, such as a decrease in peptide identifications as the values of im, rt, and hyper_im increase. Other effects on feature identification, such as snr and hyper_mz, appear to be complex, likely representing complex relationships with other variables. Other IMTBX parameters such as ln, npts, and noiseWnd1, do not appear to have a significant effect, but it is possible that interactions between parameters obscure individual effects when visualized in a single dimension. Though the complexities of the interactions between IMTBX parameters cannot be visualized due to the high dimensionality of the parameter space, the gradient boosted regression model can be used to determine the parameters that are most influential in accurately detecting peptide features from the raw LC-IMS-MS data.

The relative importance of each IMTBX parameter is shown in **Figure 32**.

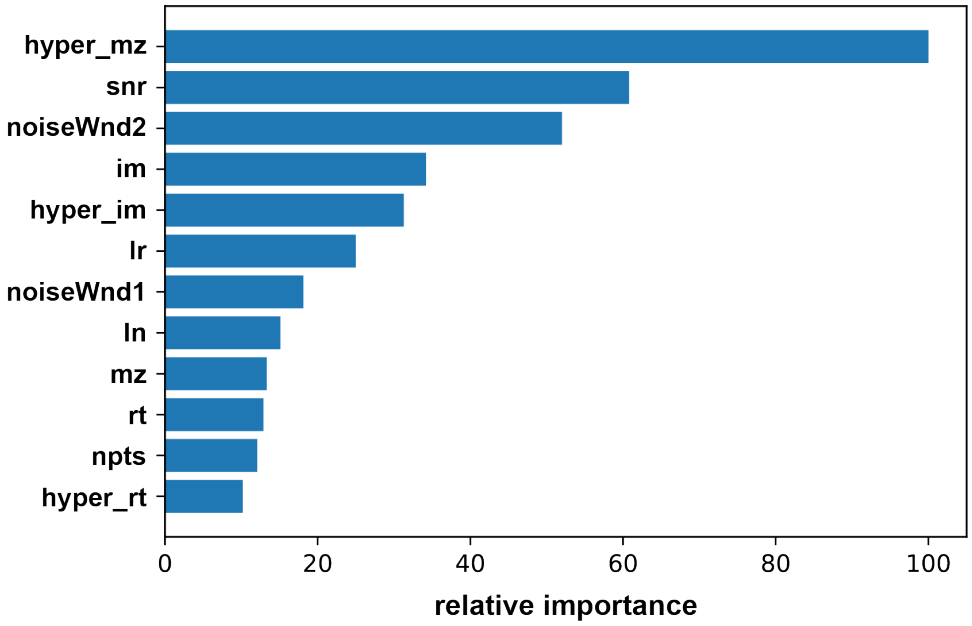

Figure 32. Importance of each IMTBX parameter as determined by the gradient boosted regression model fit to all six raw files. Relative importance (x-axis) is plotted as the percentage relative to the maximum for all parameters.

According to the model, hyper_mz, snr, and noiseWnd2 are the most influential parameters. Interestingly, these three most important parameters do not have pronounced effects when visualized alone in the partial dependence plots (**Figure 31**). This indicates that complex interactions between parameters occur, which makes it challenging to predict optimal parameter combinations. Additionally, the overall predictive power of the model is quite low ($R^2 < 0.25$, **Figure 30a**) when used for all raw files, indicative of a poor fit to the data. Given that the fit and predictive power is greatly improved when only a single raw file is used ($R^2 = 0.7$, **Figure 30b**), it is likely that the peptide and fragment ion features are too different between raw files, and IMTBX parameters suitable for one file are unsuitable for another.

Overall, the partial dependence plots of mz, im, and rt indicate that the values of these parameters (as well as their respective 'hyper' parameters) should generally be kept low to improve feature detection. Fully automated tuning of IMTBX feature detection parameters, where grid searching could be applied on-line to each raw file individually (or batch of raw files in an experiment), is theoretically feasible, but would require significant time, memory, and computational resources.

This chapter demonstrates initial use and testing of the first open-source data analysis software for LC-IMS-MS proteomics. A framework for exploring complex parameter spaces of new tools is presented. In future work, the process of parameter optimization could be fully automated, with the results of improved multiple regression modeling used to guide new sampling of the parameter space. Results of initial fitting of one 5000-sample test can be used to test the feasibility of an automated process. Even without further optimization, the IMTBX / Grppr tools are well suited to analyses of intact proteins on the IMS-MS platform. For analyses

of simple peptide mixtures, this pipeline offers free, open-source proteomics data analysis and feature detection capabilities for analytical method development.

Chapter 5 Conclusion

Mass spectrometry-based proteomics is a crucial tool for research in biomedicine and biology. The work described in this dissertation centers around one of the commercially available mass spectrometry platforms currently being used to conduct proteomics experiments. This instrument platform, the Waters Synapt G2-S, integrates traveling-wave ion mobility spectrometry with liquid chromatography-mass spectrometry, leveraging the orthogonality of the ion mobility separation to improve selectivity and proteome coverage.

The sheer complexity of the trypsin-digested proteome, which likely contains millions of different uniquely modified peptides, will remain a barrier to identifying and quantifying the entire set of proteins in a whole cell lysate. Without additional in-line separation, co-elution of peptides of similar mass from the liquid chromatography column leads to co-fragmentation of peptide ions, generating complex fragment ion spectra that contain products of multiple peptide ions. These contaminated spectra result in lost identifications and skewed protein quantification.

Complex proteomics samples can be fractionated (using techniques such as gel electrophoresis, high-pH reversed phase liquid chromatography, or isoelectric focusing) prior to LC-MS analysis to reduce complexity and achieve deeper proteome coverage. However, these methods require multiplying the LC-MS analysis time by the number of fractions, imposing a steep time cost on increased selectivity.

In-line separations, such as ion mobility spectrometry, do not require additional separation time, as these separations occur over milliseconds, and fit neatly between the LC time scale (seconds) and the MS timescale (microseconds). As mass spectrometry-based proteomics continues towards becoming a routine tool in drug discovery, biological investigations, and personalized medicine, instrument platforms that employ such time-saving separations are likely to increase in popularity. But to maximize their performance, the true selectivity of such systems will need to be studied analytically with measures such as total peak capacity for tryptic peptides.

The work discussed in Chapter 2 presents a framework for the analytical study of a three-dimensional LC-IMS-MS system, and demonstrates that even modest improvements in selectivity can increase peak capacity to enhance proteome coverage by significant margins. Following on this framework for monitoring system peak capacity, continued work to optimize ion mobility separations for complex proteomics promises further improvements in selectivity. Additionally, because of the relationship between ion mobility drift times and mass-to-charge, peptide ions separated by ion mobility prior to fragmentation can be dissociated with collision energies tuned to provide high-quality fragment ion spectra.

Given the enhanced selectivity and fragmentation efficiency afforded by ion mobility spectrometry, instruments incorporating IMS have the potential to become the top-performing platforms for high-throughput proteomics studies. Indeed, a trapped ion mobility spectrometry-time of flight (TIMS-TOF) instrument recently introduced by Bruker has demonstrated promising results. The TIMS-TOF platform uses ion mobility and subsequent quadrupole m/z filtering to isolate peptide ions prior to fragmentation, decreasing co-isolation to improve fragment ion spectral quality (**Figure 31**).[121-123]

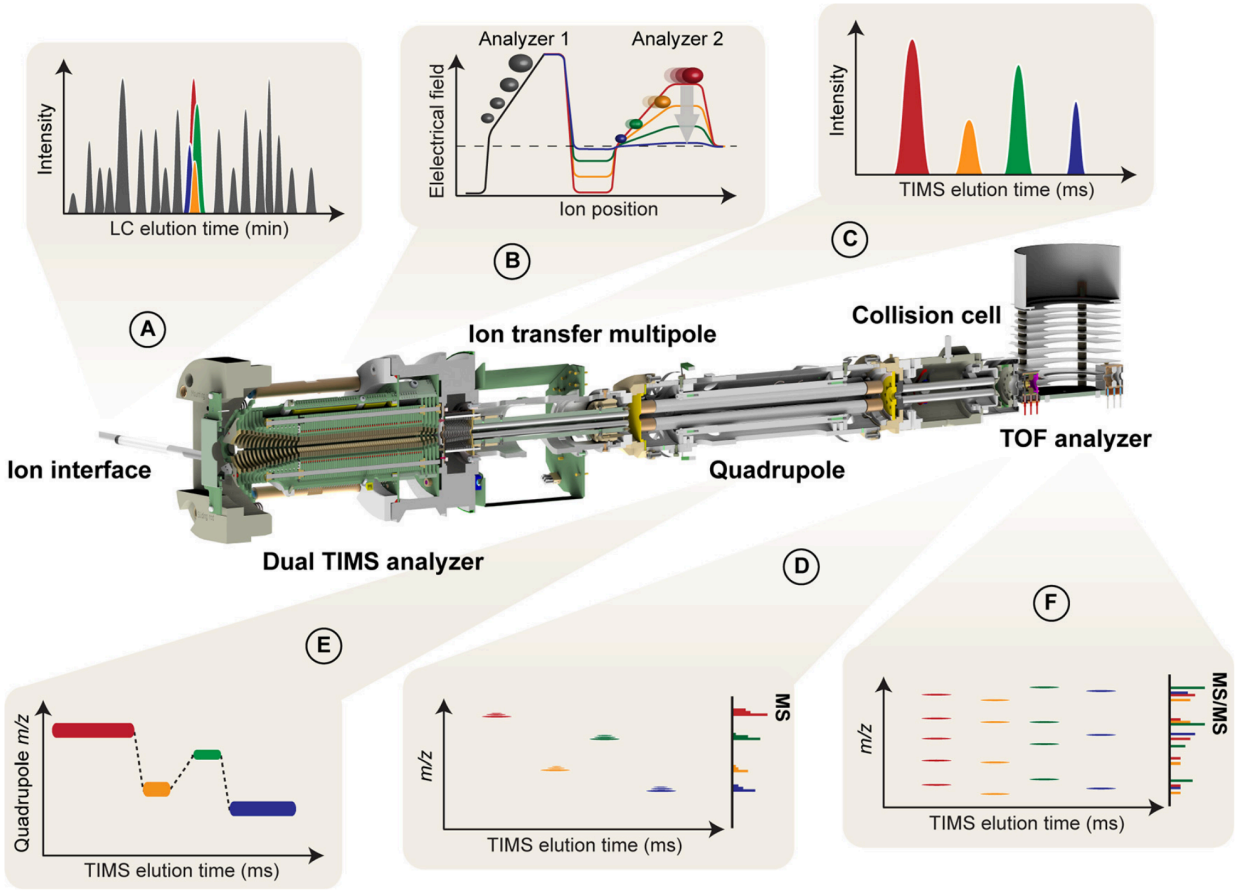

Figure 31. Schematic of the trapped ion mobility-time of flight instrument and proteomics acquisition method. (a) Eluting peptides are ionized and enter the mass spectrometer. (b) Trapping mechanism of the dual ion mobility analyzer is shown. The first regions traps peptide ions, and the second performs the mobility separation. (c) Ions exit the ion mobility region and (d) are mass analyzed. (e) To acquire fragment spectra, the quadrupole is synchronized to the ion mobility device, isolating peptide ions by mobility and m/z prior to fragmentation. (f) This yields mobility-resolved fragment spectra. *This research was originally published in Molecular & Cellular Proteomics. Meier, F., Brunner, A.D., Koch, S., Koch, H., Lubeck, M., Krause, M., Goedecke, N., Decker, J., Kosinski, T., Park, M.A. and Bache, N. "Online parallel accumulation–serial fragmentation (PASEF) with a novel trapped ion mobility mass spectrometer". Mol Cell Proteomics. 2018; 17.12:2534-2545. Copyright the American Society for Biochemistry and Molecular Biology.*

Currently, the instrument can obtain roughly two times greater proteome coverage than the previous state-of-the-art platform, the Orbitrap Fusion Lumos.[124] The authors reported identification of 2500 proteins from complex HeLa cell digests over short, 30-minute LC gradients, and over 6000 proteins when 120-minute gradients were used.

Other manufacturers in addition to Bruker are also pursuing high-resolution ion mobility separations coupled to mass spectrometry. A commercial drift tube ion mobility-time of flight instrument has recently been introduced[125], as well as a cyclic traveling wave ion mobility-time of flight platform[126]. The cyclic ion mobility instrument is shown in **Figure 32**. An extension of the linear traveling wave device described in Chapter 1, the cyclic traveling wave device consists of stacked ring ion guides over which a voltage wave is passed within a confining radiofrequency field. In contrast to the linear ion mobility setup employed by the Synapt, where the path length of the separation is limited to the physical length of the ion mobility region, ions can be sent for multiple passes through the cyclic ion mobility device. This allows for the possibility of extremely long separation path lengths and increases the flexibility of the traveling wave device to improve separation efficiency.

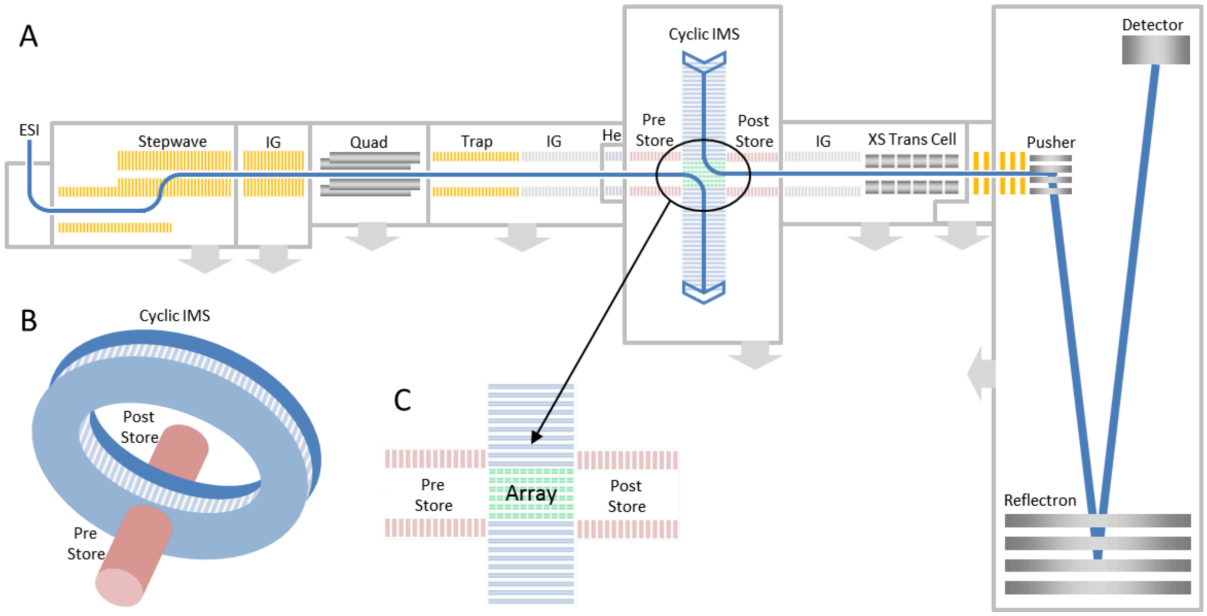

Figure 34. Cyclic traveling wave ion mobility device coupled to a time-of-flight mass analyzer. (a) Schematic of the the quadruopole-cyclic ion mobility-mass spectrometry geometry. (b) Side-on schematic of the cyclic ion mobility device with pre- and post-separation trapping regions show. (c) Location of electrode array for switching traveling wave direction. *Reprinted from ChemRxiv via Creative Commons 4.0. Eldrid, C., Ujma, J., Kalfas, S., Giles, K., & Thalassinos, K. (2018). Gas Phase Stability of Protein Ions in a Cyclic Ion Mobility Spectrometry Travelling Wave Device.*

The resolution of the ion mobility separations (measured by the drift time over the drift time peak width, or $R = t_d/t_{FWHM}$) achieved by drift tube and cyclic traveling wave devices (75[125] and 350[126], respectively) may significantly exceed that of the current generation Synapt instruments, which we measured to be less than 20 for tryptic peptides. If separation times remain in the millisecond timescale, such massive improvements in resolving power will surely lead to significant increases in the number of features that can be identified from complex proteomics samples.

Ideally, some of these new ion-mobility enabled instrument platforms would retain the capability to perform data-independent acquisition (DIA), where no abundance-based m/z isolation is performed prior to peptide fragmentation. Data-independent acquisition generates highly complex fragment spectra that can be challenging to deconvolve, but data-dependent

acquisition sacrifices cross-replicate reproducibility and may also suffer from interference from co-isolated peptides ions. To devise the best possible workflows for proteome analyses, overall quantification accuracy and the balance between spectral quality and reproducibility should be carefully monitored as different acquisition schemes are explored on these new ion mobility-enabled platforms.

Data-independent acquisition stores all peptide and fragment information, increasing the total number of signals available for protein quantification. The work described in Chapter 3 leveraged the y-ion peaks measured by DIA analysis to improve protein quantification by stable isotope labeling in cell culture (SILAC). The statistical filtering method demonstrated in Chapter 3 could also be applied to improve SILAC or dimethyl protein quantification when other data-independent acquisition methods, such as SWATH[70, 127], are used.

A major bottleneck in advancing ion mobility-mass spectrometry proteomics is data analysis. In contrast to most LC-MS data formats, few free tools are available to use with ion mobility-mass spectrometry data for feature extraction and conversion to widely-accepted file formats. This limitation hinders analytical method development, such as optimization of ion mobility separations and assessing fragmentation efficiency. Proprietary analysis software also imposes hard limitations on the types of search algorithms and validation tools that can be used to analyze LC-IMS-MS proteomics data.

Chapter 4 of this thesis examines a new, open-source tool[106] for analyzing traveling wave ion mobility-mass spectrometry data. While the tool, IMTBX / Grppr, has been successfully used to automatically analyze IMS-MS experiments of intact proteins, it has not been examined in the context of LC-IMS-MS proteomics data. The appropriate peak detection parameters for extracting peptide and fragment ion features from the multidimensional data were unknown, and

a sampling approach was used to empirically test values for these parameters. From tests of thousands of combinations of values, the best-performing parameters of IMTBX / Grppr recovered only half the number of protein identifications as the proprietary peak detection software.

While the open-source tools did not perform as well as the proprietary tools, it is likely that the performance of IMTBX / Gppr could be improved with further sampling following the process outlined in Chapter 4. These initial results also provide users of the LC-IMS-MS platform with the tools to analyze simple peptide mixtures without having to purchase proprietary software.

Flexible, open-source tools will be necessary to carry out analytical studies of current and next-generation ion mobility mass spectrometers in the context of complex proteome measurements. Optimizing and testing such peak detection tools will present a challenge, and the empirical sampling approach discussed in Chapter 4 could be applied to improve software performance in a variety of other situations where values of multiple parameters are unknown.

Large-scale measurements of protein expression, interactions, and modifications will remain a crucial tool in the life sciences. In-line ion mobility separations are a promising direction for bottom-up proteomics. As LC-IMS-MS instrumentation advances, careful analytical studies of peak capacity and optimization of selectivity for tryptic peptides will be required, along with statistical methods for improving confidence in quantitative proteomics results and flexible, open-source tools to analyze the data. By applying these concepts to one LC-IMS-MS instrument platform, the work described in this thesis demonstrates the promise of such careful analytical studies in pursuit of proteome-wide measurements.

Bibliography

1. Consortium, U., 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. 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.

15. 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.
16. 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.
17. 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.
18. Wilm, M., Principles of electrospray ionization. *Molecular & Cellular Proteomics* **2011**, mcp. R111. 009407.
19. Mann, M.; Wilm, M., Electrospray mass spectrometry for protein characterization. *Trends in biochemical sciences* **1995,** *20* (6), 219-224.
20. Wilm, M.; Shevchenko, A.; Houthaeve, T.; Breit, S.; Schweigerer, L.; Fotsis, T.; Mann, M., Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* **1996,** *379* (6564), 466.
21. Wilm, M.; Mann, M., Analytical properties of the nanoelectrospray ion source. *Analytical chemistry* **1996,** *68* (1), 1-8.
22. Kennedy, R. T.; Jorgenson, J. W., Preparation and evaluation of packed capillary liquid chromatography columns with inner diameters from 20 to 50 micrometers. *Analytical chemistry* **1989,** *61* (10), 1128-1135.
23. 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.
24. Covey, T. R.; Lee, E. D.; Bruins, A. P.; Henion, J. D., Liquid chromatography/mass spectrometry. *Analytical chemistry* **1986,** *58* (14), 1451A-1461A.
25. 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.
26. 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.
27. 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.
28. 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.
29. 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.
30. 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.

31. Dorsey, J. G.; Dill, K. A., The molecular mechanism of retention in reversed-phase liquid chromatography. *Chemical Reviews* **1989,** *89* (2), 331-346.
32. Snyder, L.; Dolan, J.; Gant, J. R., Gradient elution in high-performance liquid chromatography: I. Theoretical basis for reversed-phase systems. *Journal of Chromatography A* **1979,** *165* (1), 3-30.
33. Karger, B. L.; Gant, J. R.; Martkopf, A.; Weiner, P. H., Hydrophobic effects in reversed-phase liquid chromatography. *Journal of Chromatography A* **1976,** *128* (1), 65-78.
34. Guo, D.; Mant, C. T.; Taneja, A. K.; Parker, J. R.; Rodges, R. S., Prediction of peptide retention times in reversed-phase high-performance liquid chromatography I. Determination of retention coefficients of amino acid residues of model synthetic peptides. *Journal of Chromatography A* **1986,** *359*, 499-518.
35. Sasagawa, T.; Okuyama, T.; Teller, D. C., Prediction of peptide retention times in reversed-phases high-performance liquid chromatography during linear gradient elution. *Journal of Chromatography A* **1982,** *240* (2), 329-340.
36. March, R. E., An introduction to quadrupole ion trap mass spectrometry. *Journal of mass spectrometry* **1997,** *32* (4), 351-369.
37. Chernushevich, I. V.; Loboda, A. V.; Thomson, B. A., An introduction to quadrupole–time-of-flight mass spectrometry. *Journal of mass spectrometry* **2001,** *36* (8), 849-865.
38. Papayannopoulos, I. A., The interpretation of collision-induced dissociation tandem mass spectra of peptides. *Mass spectrometry reviews* **1995,** *14* (1), 49-73.
39. Sleno, L.; Volmer, D. A., Ion activation methods for tandem mass spectrometry. *Journal of mass spectrometry* **2004,** *39* (10), 1091-1112.
40. Covey, T. R.; Bonner, R. F.; Shushan, B. I.; Henion, J.; Boyd, R., The determination of protein, oligonucleotide and peptide molecular weights by ion-spray mass spectrometry. *Rapid communications in mass spectrometry* **1988,** *2* (11), 249-256.
41. Nesvizhskii, A. I.; Aebersold, R., Interpretation of shotgun proteomic data the protein inference problem. *Molecular & cellular proteomics* **2005,** *4* (10), 1419-1440.
42. 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.
43. Nesvizhskii, A. I.; Vitek, O.; Aebersold, R., Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nature methods* **2007,** *4* (10), 787.
44. 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.
45. Kong, A. T.; Leprevost, F. V.; Avtonomov, D. M.; Mellacheruvu, D.; Nesvizhskii, A. I., MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry–based proteomics. *Nature methods* **2017,** *14* (5), 513.
46. Michalski, A.; Cox, J.; Mann, M., More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC− MS/MS. *Journal of proteome research* **2011,** *10* (4), 1785-1793.
47. Houel, S.; Abernathy, R.; Renganathan, K.; Meyer-Arendt, K.; Ahn, N. G.; Old, W. M., Quantifying the impact of chimera MS/MS spectra on peptide identification in large-scale proteomics studies. *J Proteome Res* **2010,** *9* (8), 4152-60.
48. May, J. C.; McLean, J. A., Ion mobility-mass spectrometry: time-dispersive instrumentation. *Anal Chem* **2015,** *87* (3), 1422-36.

49. McDaniel, E.; Martin, D.; Barnes, W., Drift tube-mass spectrometer for studies of low-energy ion-molecule reactions. *Review of Scientific Instruments* **1962,** *33* (1), 2-7.
50. Wu, C.; Siems, W. F.; Klasmeier, J.; Hill, H. H., Separation of isomeric peptides using electrospray ionization/high-resolution ion mobility spectrometry. *Analytical chemistry* **2000,** *72* (2), 391-395.
51. Koomen, J. M.; Ruotolo, B. T.; Gillig, K. J.; McLean, J. A.; Russell, D. H.; Kang, M.; Dunbar, K. R.; Fuhrer, K.; Gonin, M.; Schultz, J. A., Oligonucleotide analysis with MALDI-ion-mobility-TOFMS. *Anal Bioanal Chem* **2002,** *373* (7), 612-7.
52. Verbeck, G.; Ruotolo, B.; Sawyer, H.; Gillig, K.; Russell, D., A fundamental introduction to ion mobility mass spectrometry applied to the analysis of biomolecules. *Journal of biomolecular techniques: JBT* **2002,** *13* (2), 56.
53. Woods, A. S.; Ugarov, M.; Egan, T.; Koomen, J.; Gillig, K. J.; Fuhrer, K.; Gonin, M.; Schultz, J. A., Lipid/peptide/nucleotide separation with MALDI-ion mobility-TOF MS. *Anal Chem* **2004,** *76* (8), 2187-95.
54. Fenn, L. S.; McLean, J. A., Biomolecular structural separations by ion mobility-mass spectrometry. *Anal Bioanal Chem* **2008,** *391* (3), 905-9.
55. Thalassinos, K.; Grabenauer, M.; Slade, S. E.; Hilton, G. R.; Bowers, M. T.; Scrivens, J. H., Characterization of phosphorylated peptides using traveling wave-based and drift cell ion mobility mass spectrometry. *Analytical chemistry* **2008,** *81* (1), 248-254.
56. Pringle, S. D.; Giles, K.; Wildgoose, J. L.; Williams, J. P.; Slade, S. E.; Thalassinos, K.; Bateman, R. H.; Bowers, M. T.; Scrivens, J. H., An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. International Journal of Mass Spectrometry 2007; Vol. 261, pp 1-12.
57. Fernandez-Lima, F.; Kaplan, D. A.; Suetering, J.; Park, M. A., Gas-phase separation using a trapped ion mobility spectrometer. *International Journal for Ion Mobility Spectrometry* **2011,** *14* (2-3), 93-98.
58. May, J. C.; Goodwin, C. R.; Lareau, N. M.; Leaptrot, K. L.; Morris, C. B.; Kurulugama, R. T.; Mordehai, A.; Klein, C.; Barry, W.; Darland, E., Conformational ordering of biomolecules in the gas phase: nitrogen collision cross sections measured on a prototype high resolution drift tube ion mobility-mass spectrometer. *Analytical chemistry* **2014,** *86* (4), 2107-2116.
59. Giles, K.; Pringle, S. D.; Worthington, K. R.; Little, D.; Wildgoose, J. L.; Bateman, R. H., Applications of a travelling wave-based radio-frequency-only stacked ring ion guide. *Rapid Commun Mass Spectrom* **2004,** *18* (20), 2401-14.
60. Giles, K.; Williams, J. P.; Campuzano, I., Enhancements in travelling wave ion mobility resolution. *Rapid Commun Mass Spectrom* **2011,** *25* (11), 1559-66.
61. Shvartsburg, A. A.; Smith, R. D., Fundamentals of traveling wave ion mobility spectrometry. *Anal Chem* **2008,** *80* (24), 9689-99.
62. Dawson, J.; Guilhaus, M., Orthogonal-acceleration time-of-flight mass spectrometer. *Rapid Communications in Mass Spectrometry* **1989,** *3* (5), 155-159.
63. Boyle, J. G.; Whitehouse, C. M., Time-of-flight mass spectrometry with an electrospray ion beam. *Analytical chemistry* **1992,** *64* (18), 2084-2089.
64. Guilhaus, M.; Selby, D.; Mlynski, V., Orthogonal acceleration time-of-flight mass spectrometry. *Mass spectrometry reviews* **2000,** *19* (2), 65-107.

65. Distler, U.; Kuharev, J.; Navarro, P.; Levin, Y.; Schild, H.; Tenzer, S., Drift time-specific collision energies enable deep-coverage data-independent acquisition proteomics. *Nat Methods* **2014,** *11* (2), 167-70.
66. Shliaha, P. V.; Bond, N. J.; Gatto, L.; Lilley, K. S., Effects of traveling wave ion mobility separation on data independent acquisition in proteomics studies. *J Proteome Res* **2013,** *12* (6), 2323-39.
67. Hernandez, J. L.; Davda, D.; Majmudar, J. D.; Won, S. J.; Prakash, A.; Choi, A. I.; Martin, B. R., Correlated S-palmitoylation profiling of Snail-induced epithelial to mesenchymal transition. *Molecular BioSystems* **2016,** *12* (6), 1799-1808.
68. Büscher, N.; Paulus, C.; Nevels, M.; Tenzer, S.; Plachter, B., The proteome of human cytomegalovirus virions and dense bodies is conserved across different strains. *Medical microbiology and immunology* **2015,** *204* (3), 285-293.
69. Saraswat, M.; Joenväärä, S.; Seppänen, H.; Mustonen, H.; Haglund, C.; Renkonen, R., Comparative proteomic profiling of the serum differentiates pancreatic cancer from chronic pancreatitis. *Cancer medicine* **2017,** *6* (7), 1738-1751.
70. Venable, J. D.; Dong, M. Q.; Wohlschlegel, J.; Dillin, A.; Yates, J. R., Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat Methods* **2004,** *1* (1), 39-45.
71. Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M. C.; Yates, J. R., 3rd, Protein analysis by shotgun/bottom-up proteomics. *Chem Rev* **2013,** *113* (4), 2343-94.
72. Bateman, R.; Carruthers, R.; Hoyes, J.; Jones, C.; Langridge, J.; Millar, A.; Vissers, J., A novel precursor ion discovery method on a hybrid quadrupole orthogonal acceleration time-of-flight (Q-TOF) mass spectrometer for studying protein phosphorylation. *Journal of the American Society for Mass Spectrometry* **2002,** *13* (7), 792-803.
73. Geromanos, S. J.; Vissers, J. P.; Silva, J. C.; Dorschel, C. A.; Li, G. Z.; Gorenstein, M. V.; Bateman, R. H.; Langridge, J. I., The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS. *Proteomics* **2009,** *9* (6), 1683-95.
74. Horvath, C. G.; Lipsky, S. R., Peak capacity in chromatography. *Analytical chemistry* **1967,** *39* (14), 1893-1893.
75. Geromanos, S. J.; Hughes, C.; Ciavarini, S.; Vissers, J. P.; Langridge, J. I., Using ion purity scores for enhancing quantitative accuracy and precision in complex proteomics samples. *Anal Bioanal Chem* **2012,** *404* (4), 1127-39.
76. Rardin, M. J.; Schilling, B.; Cheng, L.-Y.; MacLean, B. X.; Sorenson, D. J.; Sahu, A. K.; MacCoss, M. J.; Vitek, O.; Gibson, B. W., MS1 Peptide Ion Intensity Chromatograms in MS2 (SWATH) Data Independent Acquisitions. Improving Post Acquisition Analysis of Proteomic Experiments. *Molecular & Cellular Proteomics* **2015**, mcp. O115. 048181.
77. Gallien, S.; Duriez, E.; Demeure, K.; Domon, B., Selectivity of LC-MS/MS analysis: implication for proteomics experiments. *Journal of proteomics* **2013,** *81*, 148-158.
78. Paulo, J. A.; O'Connell, J. D.; Gygi, S. P., A triple knockout (TKO) proteomics standard for diagnosing ion interference in isobaric labeling experiments. *Journal of the American Society for Mass Spectrometry* **2016,** *27* (10), 1620-1625.
79. Baker, E. S.; Livesay, E. A.; Orton, D. J.; Moore, R. J.; Danielson III, W. F.; Prior, D. C.; Ibrahim, Y. M.; LaMarche, B. L.; Mayampurath, A. M.; Schepmoes, A. A., An LC-IMS-MS platform providing increased dynamic range for high-throughput proteomic studies. *Journal of proteome research* **2010,** *9* (2), 997-1006.

80. Liu, X.; Valentine, S. J.; Plasencia, M. D.; Trimpin, S.; Naylor, S.; Clemmer, D. E., Mapping the human plasma proteome by SCX-LC-IMS-MS. *Journal of the American Society for Mass Spectrometry* **2007,** *18* (7), 1249-1264.
81. Valentine, S. J.; Kulchania, M.; Barnes, C. A. S.; Clemmer, D. E., Multidimensional separations of complex peptide mixtures: a combined high-performance liquid chromatography/ion mobility/time-of-flight mass spectrometry approach. *International Journal of Mass Spectrometry* **2001,** *212* (1), 97-109.
82. Ruotolo, B. T.; Gillig, K. J.; Stone, E. G.; Russell, D. H., Peak capacity of ion mobility mass spectrometry: separation of peptides in helium buffer gas. *J Chromatogr B Analyt Technol Biomed Life Sci* **2002,** *782* (1-2), 385-92.
83. Ruotolo, B. T.; McLean, J. A.; Gillig, K. J.; Russell, D. H., The influence and utility of varying field strength for the separation of tryptic peptides by ion mobility-mass spectrometry. *J Am Soc Mass Spectrom* **2005,** *16* (2), 158-65.
84. Ruotolo, B. T.; McLean, J. A.; Gillig, K. J.; Russell, D. H., Peak capacity of ion mobility mass spectrometry: the utility of varying drift gas polarizability for the separation of tryptic peptides. *J Mass Spectrom* **2004,** *39* (4), 361-7.
85. Fenn, L. S.; Kliman, M.; Mahsut, A.; Zhao, S. R.; McLean, J. A., Characterizing ion mobility-mass spectrometry conformation space for the analysis of complex biological samples. *Anal Bioanal Chem* **2009,** *394* (1), 235-44.
86. Zhong, Y.; Hyung, S. J.; Ruotolo, B. T., Characterizing the resolution and accuracy of a second-generation traveling-wave ion mobility separator for biomolecular ions. *Analyst* **2011,** *136* (17), 3534-41.
87. Distler, U.; Kuharev, J.; Navarro, P.; Tenzer, S., Label-free quantification in ion mobility–enhanced data-independent acquisition proteomics. *nature protocols* **2016,** *11* (4), 795.
88. Giddings, J. C., Two-dimensional separations: concept and promise. *Anal Chem* **1984,** *56* (12), 1258A-1260A, 1262A, 1264A passim.
89. Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J., A cross-platform toolkit for mass spectrometry and proteomics. *Nature biotechnology* **2012,** *30* (10), 918.
90. Fernandez-Lima, F. A.; Blase, R. C.; Russell, D. H., A study of ion-neutral collision cross-section values for low charge states of peptides, proteins, and peptide/protein complexes. *International journal of mass spectrometry* **2010,** *298* (1-3), 111-118.
91. Pandey, A.; Mann, M., Proteomics to study genes and genomes. *Nature* **2000,** *405* (6788), 837.
92. Shishkova, E.; Hebert, A. S.; Coon, J. J., Now, more than ever, proteomics needs better chromatography. *Cell systems* **2016,** *3* (4), 321-324.
93. Silva, J. C.; Gorenstein, M. V.; Li, G. Z.; Vissers, J. P.; Geromanos, S. J., Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Mol Cell Proteomics* **2006,** *5* (1), 144-56.
94. Levin, Y.; Hradetzky, E.; Bahn, S., Quantification of proteins using data-independent analysis (MSE) in simple andcomplex samples: A systematic evaluation. *Proteomics* **2011,** *11* (16), 3273-3287.
95. Haynes, S. E.; Polasky, D. A.; Dixit, S. M.; Majmudar, J. D.; Neeson, K.; Ruotolo, B. T.; Martin, B. R., Variable-velocity traveling-wave ion mobility separation enhances peak capacity for data-independent acquisition proteomics. *Analytical Chemistry* **2017**.

96. Collins, B. C.; Hunter, C. L.; Liu, Y.; Schilling, B.; Rosenberger, G.; Bader, S. L.; Chan, D. W.; Gibson, B. W.; Gingras, A.-C.; Held, J. M., Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH-mass spectrometry. *Nature communications* **2017,** *8* (1), 291.
97. Huang, X.; Liu, M.; Nold, M. J.; Tian, C.; Fu, K.; Zheng, J.; Geromanos, S. J.; Ding, S.-J., Software for quantitative proteomic analysis using stable isotope labeling and data independent acquisition. *Analytical chemistry* **2011,** *83* (18), 6971-6979.
98. Daly, C. E.; Ng, L. L.; Hakimi, A.; Willingale, R.; Jones, D. J., Qualitative and Quantitative Characterization of Plasma Proteins When Incorporating Traveling Wave Ion Mobility into a Liquid Chromatography–Mass Spectrometry Workflow for Biomarker Discovery: Use of Product Ion Quantitation As an Alternative Data Analysis Tool for Label Free Quantitation. *Analytical chemistry* **2014,** *86* (4), 1972-1979.
99. Li, Q.; Chang, Z.; Oliveira, G.; Xiong, M.; Smith, L. M.; Frey, B. L.; Welham, N. V., Protein turnover during in vitro tissue engineering. *Biomaterials* **2016,** *81*, 104-113.
100. Kathiriya, J. J.; Nakra, N.; Nixon, J.; Patel, P. S.; Vaghasiya, V.; Alhassani, A.; Tian, Z.; Allen-Gipson, D.; Davé, V., Galectin-1 inhibition attenuates profibrotic signaling in hypoxia-induced pulmonary fibrosis. *Cell death discovery* **2017,** *3*, 17010.
101. Hogrebe, A.; von Stechow, L.; Bekker-Jensen, D. B.; Weinert, B. T.; Kelstrup, C. D.; Olsen, J. V., Benchmarking common quantification strategies for large-scale phosphoproteomics. *Nature communications* **2018,** *9* (1), 1045.
102. Vizcaíno, J. A.; Deutsch, E. W.; Wang, R.; Csordas, A.; Reisinger, F.; Rios, D.; Dianes, J. A.; Sun, Z.; Farrah, T.; Bandeira, N., ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nature biotechnology* **2014,** *32* (3), 223.
103. Minogue, C. E.; Hebert, A. S.; Rensvold, J. W.; Westphall, M. S.; Pagliarini, D. J.; Coon, J. J., Multiplexed quantification for data-independent acquisition. *Analytical chemistry* **2015,** *87* (5), 2570-2575.
104. Li, Z.; Adams, R. M.; Chourey, K.; Hurst, G. B.; Hettich, R. L.; Pan, C., Systematic comparison of label-free, metabolic labeling, and isobaric chemical labeling for quantitative proteomics on LTQ Orbitrap Velos. *Journal of proteome research* **2012,** *11* (3), 1582-1590.
105. Devabhaktuni, A.; Elias, J. E., Application of de novo sequencing to large-scale complex proteomics data sets. *Journal of proteome research* **2016,** *15* (3), 732-742.
106. Avtonomov, D. M.; Polasky, D. A.; Ruotolo, B. T.; Nesvizhskii, A. I., IMTBX and Grppr: Software for Top-Down Proteomics Utilizing Ion Mobility-Mass Spectrometry. *Analytical chemistry* **2018,** *90* (3), 2369-2375.
107. Vizcaíno, J. A.; Perkins, S.; Jones, A. R.; Deutsch, E. W., Data Formats of the Proteomics Standards Initiative. In *Proteome Informatics*, 2016; pp 229-258.
108. Rodriguez-Suarez, E.; Hughes, C.; Gethings, L.; Giles, K.; Wildgoose, J.; Stapels, M.; E Fadgen, K.; J Geromanos, S.; PC Vissers, J.; Elortza, F., An ion mobility assisted data independent LC-MS strategy for the analysis of complex biological samples. *Current Analytical Chemistry* **2013,** *9* (2), 199-211.
109. Quesada-Calvo, F.; Massot, C.; Bertrand, V.; Longuespée, R.; Blétard, N.; Somja, J.; Mazzucchelli, G.; Smargiasso, N.; Baiwir, D.; De Pauw-Gillet, M.-C., OLFM4, KNG1 and Sec24C identified by proteomics and immunohistochemistry as potential markers of early colorectal cancer stages. *Clinical proteomics* **2017,** *14* (1), 9.

110. Piltti, J.; Bygdell, J.; Fernández-Echevarría, C.; Marcellino, D.; Lammi, M. J., Rho-kinase inhibitor Y-27632 and hypoxia synergistically enhance chondrocytic phenotype and modify S100 protein profiles in human chondrosarcoma cells. *Scientific reports* **2017,** *7* (1), 3708.
111. Cassoli, J. S.; Brandão-Teles, C.; Santana, A. G.; Souza, G. H.; Martins-de-Souza, D., Ion Mobility-Enhanced Data-Independent Acquisitions Enable a Deep Proteomic Landscape of Oligodendrocytes. *Proteomics* **2017,** *17* (21), 1700209.
112. Nesvizhskii, A. I.; Keller, A.; Kolker, E.; Aebersold, R., A statistical model for identifying proteins by tandem mass spectrometry. *Analytical chemistry* **2003,** *75* (17), 4646-4658.
113. Senko, M. W.; Beu, S. C.; McLaffertycor, F. W., Determination of monoisotopic masses and ion populations for large biomolecules from resolved isotopic distributions. *Journal of the American Society for Mass Spectrometry* **1995,** *6* (4), 229-233.
114. Halton, J. H., On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik* **1960,** *2* (1), 84-90.
115. Friedman, J. H., Multivariate adaptive regression splines. *The annals of statistics* **1991,** *19* (1), 1-67.
116. Bjorck, A., *Numerical methods for least squares problems*. Siam: 1996; Vol. 51.
117. Friedman, J. H., Greedy function approximation: a gradient boosting machine. *Annals of statistics* **2001**, 1189-1232.
118. Friedman, J. H. *Stochastic gradient boosting. Department of Statistics*; Stanford University, Technical Report, San Francisco, CA: 1999.
119. Hastie, T.; Tibshirani, R.; Jerome, H., Friedman (2002) Elements of Statistical Learning. Springer, NY.
120. Ridgeway, G., Generalized Boosted Models: A guide to the gbm package. *Update* **2007,** *1* (1), 2007.
121. Meier, F.; Brunner, A. D.; Koch, S.; Koch, H.; Lubeck, M.; Krause, M.; Goedecke, N.; Decker, J.; Kosinski, T.; Park, M. A.; Bache, N.; Hoerning, O.; Cox, J.; Rather, O.; Mann, M., Online Parallel Accumulation-Serial Fragmentation (PASEF) with a Novel Trapped Ion Mobility Mass Spectrometer. *Mol Cell Proteomics* **2018,** *17* (12), 2534-2545.
122. Meier, F.; Beck, S.; Grassl, N.; Lubeck, M.; Park, M. A.; Raether, O.; Mann, M., Parallel accumulation–serial fragmentation (PASEF): multiplying sequencing speed and sensitivity by synchronized scans in a trapped ion mobility device. *Journal of proteome research* **2015,** *14* (12), 5378-5387.
123. Silveira, J. A.; Ridgeway, M. E.; Laukien, F. H.; Mann, M.; Park, M. A., Parallel accumulation for 100% duty cycle trapped ion mobility-mass spectrometry. *International Journal of Mass Spectrometry* **2017,** *413*, 168-175.
124. Senko, M. W.; Remes, P. M.; Canterbury, J. D.; Mathur, R.; Song, Q.; Eliuk, S. M.; Mullen, C.; Earley, L.; Hardman, M.; Blethrow, J. D., Novel parallelized quadrupole/linear ion trap/Orbitrap tribrid mass spectrometer improving proteome coverage and peptide identification rates. *Analytical chemistry* **2013,** *85* (24), 11710-11714.
125. Ibrahim, Y. M.; Baker, E. S.; Danielson III, W. F.; Norheim, R. V.; Prior, D. C.; Anderson, G. A.; Belov, M. E.; Smith, R. D., Development of a new ion mobility time-of-flight mass spectrometer. *International journal of mass spectrometry* **2015,** *377*, 655-662.
126. Eldrid, C.; Ujma, J.; Kalfas, S.; Giles, K.; Thalassinos, K., Gas Phase Stability of Protein Ions in a Cyclic Ion Mobility Spectrometry Travelling Wave Device. **2018**.

127. Gillet, L. C.; Navarro, P.; Tate, S.; Rost, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R., Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol Cell Proteomics* **2012,** *11* (6), O111.016717.