Université de Montréal

Optimization of Differential Ion Mobility and Segmented Ion Fractionation to Improve Proteome Coverage

By

Zhaoguan Wu

Département de chimie,

Faculté des arts et des sciences

Thèse présentée à en vue de l'obtention du grade de

Maître ès sciences (M.Sc.) en chimie

Avril 2022

Cette thèse intitulée

Optimization of Differential Ion Mobility and Segmented Ion Fractionation to Improve Proteome Coverage

Présenté par

Zhaoguan Wu

A été évaluée par un jury composé des personnes suivantes

Karen Waldron

Président-rapporteur

Pierre Thibault

Directeur de recherche

Pierre Chaurand

Membre du jury

Résumé

La sensibilité et la profondeur de l'analyse protéomique sont limitées par les ions isobares et les interférences qui entravent l'identification des peptides de faible abondance. Lorsque nous analysons des échantillons de grande complexité, une séparation extensive de l'échantillon est souvent nécessaire pour étendre la couverture protéomique. Ces dernières années, la spectrométrie de mobilité ionique à forme d'onde asymétrique à haut champ (FAIMS) a gagné en popularité dans le domaine de la protéomique pour sa capacité à séparer les ions isobares, à améliorer la capacité de pic et la sensibilité de la spectrométrie de masse (MS). Nous rapportons ici l'intégration d'un appareil FAIMS Pro™ à un Q-Exactive HF™ ainsi qu'un spectromètre de masse Orbitrap Exploris 480™. Des expériences protéomiques sur des digestions d'extraits protéiques issues de cellules Hela à l'aide d'un spectromètre de masse avec FAIMS ont amélioré le rapport signal sur bruit (S/N) et réduit les ions interférents, ce qui a entraîné une augmentation du taux d'identification des peptides de plus de 42 %. FAIMS est également combiné avec le fractionnement ionique segmenté (SIFT), qui utilise tour à tour une fenêtre de 100 ~ 300 m/z au lieu de la large plage traditionnelle (700 ~ 800 m/z), augmentant ainsi la profondeur de la couverture protéomique tout en réduisant la proportion de spectres MS/MS chimériques de 50% à 27%. Dans l'analyse quantitative, nous démontrons l'application de FAIMS pour améliorer les mesures quantitatives lorsque le marquage peptidique isobare est utilisé. Par rapport aux expériences LC-MS/MS conventionnelles, la combinaison des expériences FAIMS et SIFT réalisées sur un modèle à deux protéomes a montré une amélioration de 65 % de la précision des mesures quantitatives. Les digestions tryptiques d'extraits protéiques de différentes lignées cellulaires du cancer colorectal ont été utilisées pour l'évaluation de stratégie combinée FAIMS et SIFT sur un spectromètre de masse Orbitrap Exploris 480™ offre un gain d'identification de 70 % par rapport à l'approche conventionnelle et combinée aux données transcriptomiques elle facilite l'identification de variants protéiques.

Mots clés : Mobilité ionique à forme d'onde asymétrique à haut champ (FAIMS), spectrométrie de masse, fractionnement en phase gazeuse, protéomique quantitative, mutation protéique, polymorphisme d'un seul acide aminé, cancer du côlon, protéogénomique.

Summary

The sensitivity and depth of proteomic analysis in mass spectrometry (MS) is limited by isobaric ions and interferences that hinder the identification of low-abundance peptides. For high complexity samples, extensive separation is often required to expand proteomic coverage. In recent years, high-field asymmetric waveform ion mobility spectrometry (FAIMS) has gained popularity in the field of proteomics for its ability to resolve confounding ions, improve peak capacity, and sensitivity. This thesis presents the integration of a FAIMS Pro™ interface with electrical and gas embedded connections to a Q-Exactive HF™ as well as an Orbitrap Exploris 480™ mass spectrometer. Proteomic experiments on tryptic digests of HeLa cell line using a FAIMS integrated mass spectrometer improved signal-to-noise ratio (S/N) and reduced the occurrence of interfering ions. This enabled a 42% increase in peptide identification rate. Also, FAIMS was combined with segmented ion fractionation (SIFT), which in turn scans with windows of 100~300 m/z width instead of the traditional width (700~800 m/z), further increasing the depth of proteome coverage by a reducing from 50% to 27% in terms of MS/MS chimeric spectra numbers. The application of FAIMS gain improvement on quantitative measurements with TMT labeling method is presented. Compared to conventional LC-MS/MS tests, the combination of FAIMS and SIFT experiments showed a improvement by 65% in quantitative accuracy when performed on a human-yeast two-proteome model. As an application of the method, the tryptic digests from different colorectal cancer cell lines were used for the evaluation. FAIMS-SIFT-combined strategy on an Orbitrap Exploris 480™ mass spectrometer provides a 70% gain in identification compared to the conventional LC-MS/MS approach for the same sample amount and instrument time. This enhanced sensitivity facilitates single amino acid mutations confirmed by RNAseq analyses.

Key words: High field asymmetric waveform ion mobility (FAIMS), mass spectrometry, gas-phase fractionation, quantitative proteomics, protein mutation, single amino acid polymorphism, colon cancer, proteogenomics.

Contents

List of tables

List of figures

List of abbreviations

ACN	Acetonitrile
AGC	Automatic gain control
Asp-C	Endoproteinase at aspartic acid C terminal
CHAPS	3-[(3-cholamidopropyl) dimethylammonio] -1-propanesulfonate
CID	Collision induced dissociation
COSMIC	Catalogue of somatic mutations in cancer
CRC	Colorectal cancer
CV	Compensation voltage
DDA	Data dependent acquisition
DNA	Deoxyribonucleic Acid
DV	Dispersion voltage
EGFR	Epidermal growth factor
ERK	Extracellular signal regulated kinases
ESI	Electrospray ionization
ETD	Electron transfer dissociation
FAIMS	High field asymmetric waveform ion mobility spectrometry
FC	Fold change
FDR	False discovery rate
FT-ICR	Fourier transform–Ion cyclotron resonance
Glu-C	Endoproteinase at glutamic acid C terminal
GO	Gene ontology
GUI	Graphical user interface
HCD	High-energy collisional dissociation
HEK	Human embryonic kidney
HLB	Hydrophilic lipophilic balance
HRAM	High resolution accurate mass
HRMS	High resolution mass spectrometry
IFI	Interference free index
IM	Ion mobility
IMS	Ion mobility spectrometry
IT	Ion trap
iTRAQ	Isobaric tags for relative and absolute quantification
LC-MS	Liquid chromatography mass spectrometry
LC-MS/MS	Liquid chromatography tandem mass spectrometry
LFQ	Label free quantification
Lys-C	Endoproteinase at lysine C terminal

MAPK/MEK	Mitogen activated protein kinase
MHC	Major histocompatibility complex
MRM	Multiple reaction monitoring
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MSI	Microsatellite instability
MSS	Microsatellite stable
MU	Mutation
NCE	Normalized collision energy
OT	Orbi-trap
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PIF	Precursor/parent ion fraction
ppm	Part per million
PRM	Parallel reaction monitoring
PSM	Peptide spectrum match
Q	Quadrupole
qPCR	Quantitative polymerase chain reaction
RF	Radio frequency
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
RSD	Relative standard deviation
S/N	Signal-to-noise
SCX	Strong cation exchange
SD	Standard deviation
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SIFT	Segmented ion fractionation
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variants
SRM	Selective reaction monitoring
TMT	Tandem mass tag
TOF	Time of flight
WT	Wild type

Acknowledgements

First and foremost, I would like to thank Dr. ***Pierre Thibault***, my supervisor and mentor. We first met when he visited Huazhong University of Science and Technology in Wuhan in 2017. That was right after his splendid seminar, I rekindled my determination to go into academic research as a student even after stepping into the workplace. Thanks to him for giving me such an opportunity, I am honored to be his student. Even during the pandemic, and during that very difficult time, he has always been willing to continue to support my idea of completing the research projects. I thank him for his help in my project progressing, research articles publication, dissertation planning and writing. Without his patience and support, I could not have completed this dissertation. It worth my lifetime learning his rigorous approaches towards research works.

In particular, I would love to send my great thanks to Dr. ***Éric Bonneil***, who has been nothing but patient, supportive and understanding during my time at IRIC-UdeM. Éric is a good mentor. I learned a lot from working with him, not only in experimental technology and scientific thinking, but also in attitude towards family, work, and life. Together we face pressure from the tense research schedules or when projects hit the plateau. Together we optimize the performance of our instruments, and together we got solutions. I know it's not always easy working with me, but his patience always beats anxiety and gives me support as much as he can. At the same time, I thank him for revising the manuscript of my dissertation and articles.

My experience tells me that a good research group, where everyone gets along like a family, that is our Thibault Lab. I would present my thanks to everyone member as well my ex-collaborators here. ***Mirela***, ***Chantal***, ***Joel***, ***Mathieu***, ***Simon***, ***Chongyang***, ***Trent***, ***Jenna***, ***Charles***, ***Moise***, ***Robin, Cédric*** and ***Sibylle*** ! We shared so many great memories and the robust discussion with you always inspiring me. Mirela, thank her for always keeping everything organized and well prepared in the lab, our experiments are being guaranteed by her work perfectly. Although Chantal is always doing experiments in silence and seriousness, she is always willing to share her insights on projects and experiments with me at any time. Thank Joel for his help with my experiments by culturing the cell lines and sample preparation for the genomic experiments. Our beautiful afternoon at Joel's house was a wonderful leisure time, which is rare during a pandemic. I admire Mathieu's superb computer technology, which has provided great help for everyone's real data

processing. An enthusiastic friend from St. Catherine & Chambly, Simon, always brings us cheerful singing and any-time help! 感谢重阳(Dr. ***Chongyang Li***) 师兄一直以来归我的关心和指导，有幸共同在武汉和蒙特利尔两个城市学习实属缘分，祝他在科研路上乘风破浪，一路通达！Trent, we always inspire each other a lot, whether it's discussing experiments or beer time! Thanks Jenna a lot for reading my thesis, and my mom always ask me to send greeting to you, she really appreciates your good wishes during her difficult time. And now, we are, in the same "good trash", good luck to your future! Charles is a jelly good photographer, I should have learnt more from him on photography, I always cannot stop surfing his Instagram. And good luck to Robin and Moise on the proteomics associating projects! Ohooo, my "dangerous neighbor" Cédric, I sincerely thank him for the help on everything, he is a pretty much reliable friend, remember you said you will go to my hometown Wuhan one day. Sibylle, I admire her work and achievements on fancy FAIMS, it is right based on her previous work I could continue the research projects, we were neighbors here in Montreal, and I will never forget those good memories on food, chalet, and the lab time.

Thanks to Dr. ***Jianjun Li and his families***! Jianjun used to be my master supervisor who devoted everything to guiding me, and the professor who helped me a lot with my research while I was working. He and his family care about my life, health, and scientific research just like my families when it is so sad to be far away from home. Every visit to Ottawa gives me the illusion of being home. I hope they always be healthy, and everything goes well!

最后也是最必要的，我感谢我的**父母和家*(my parents and families)***对我的支持，他们在精神，物质各方面都倾囊支持，我一直很抱歉在武汉疫情最严重的时候我没能陪伴在他们身边。当我在蒙特利尔遭遇疫情或是科研进展不顺利而感到难过时，得到家人在远方的支持和关怀让我感觉时间和空间的阻隔都不复存在！没有他们如此地关怀，我甚至不能支撑自己的压抑情绪在这困难又难忘的留学时间！

Chapter 1

1 Introduction

1.1 Colorectal cancer

Colorectal cancer (CRC) ranks the third on most common cancers in the world. Epidemiological statistics show that the global incidence of colorectal cancer is fourth and third in men and women respectively. There are about 1 million new cases each year, and this number is increasing yearly posing a great threat to human health [1-3]. Studies have reported that the 5-year survival rate for early colorectal cancer is 80%-90%, and the 5-year survival rate for advanced colorectal cancer is less than 10%. However, only nearly 40% of patients can be diagnosed with colorectal cancer at an early stage[4, 5].

1.1.1 KRAS and BRAF mutations are significant indicators during CRC immunotherapy

The occurrence and development of colorectal cancer are the comprehensive results of genetics, environment, and lifestyle[6, 7]. In particular, KRAS and BRAF gene mutations play a key role in the colorectal cancer occurrence and development. These gene mutations are expressed in the early stage of adenoma, which may be an early event of colorectal cancer[8-10]. Both KRAS and BRAF genes can aquire oncogenic properties, and their mutations can cause colorectal adenoma and adenocarcinoma through self-activation and inhibition of tumor suppressor genes[11, 12]. In the study of the relationship between KRAS and BRAF gene mutations and colorectal cancer, it was found that there are multiple types of mutations in these two genes that cause tumors, and these mutations are not found in healthy tissue[13, 14]. KRAS and BRAF genes are major players in the epidermal growth factor receptor (EGFR) signaling pathway. EGFR is a receptor tyrosine kinase, which belongs to the ErbB receptor family. It phosphorylates downstream tyrosine residues by binding to specific ligands and provides binding sites for signal transducers of multiple signal transduction pathways (e.g. Ras/Raf/MAP/MEK/ERK pathways). This regulates the differentiation and proliferation of normal cells[15]. However, it has been found that abnormal EGFR signal pathway activation occurs in colorectal cancer cells, leading to cell proliferation, metastasis and escape from apoptosis or angiogenesis[15]. Therefore, EGFR has become the main target in the treatment of metastatic colorectal cancer. Anti-EGFR monoclonal antibody is key to metastatic colorectal cancer treatment and has achieved good clinical efficacy[16]. However, mutations in KRAS and BRAF genes can lead to the continuous activation of EGFR-mediated signal pathways

(***Figure 1-1***). Mutated KRAS gene is found in up to 50% of patients with colorectal cancer (CRC), while BRAF mutations are found in about 7-10% of patients with metastatic colorectal cancer, limiting anti-EGFR monoclonal antibody therapy to patient with KRAS and BRAS wild type. KRAS and BRAF mutations have become an indicator of poor prognosis in colorectal cancer[16-18].

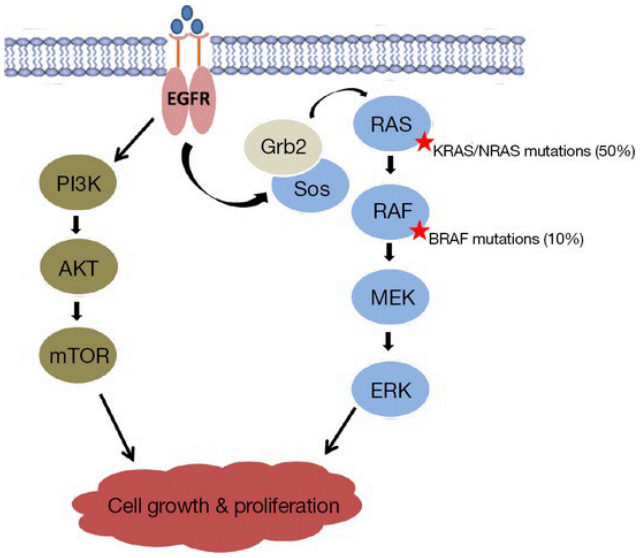


Figure 1-1: MAPK signaling pathway in CRC.[17]

Mutations in KRAS and BRAF lead to constitutive activation of signaling cascades, resulting in uninhibited cell proliferation and tumor growth. These effects occur downstream of the EGFR receptor, making this mutant tumor resistant to anti-EGFR therapies.[17]

The National Comprehensive Cancer Network (NCCN) clearly stated in the clinical practice guidelines of oncology, that KRAS and BRAF gene mutations should be detected when selecting anti-EGFR monoclonal antibody treatment making KRAS and BRAF gene mutations an important biological indicator for the treatment of colorectal cancer.

1.1.2 Research focusing on KRAS G12/13 and BRAF V600 mutations

The rate of KRAS/BRAF mutations varies among different countries and regions. KRAS gene mutation rate in colorectal cancer patients in the United Kingdom, Switzerland, and Spain is about 27.4%, 38%, and 41% respectively[19]. Ma's team reported that the BRAF gene mutation rate in colorectal cancer patients in China is a little bit lower than those in western countries (but still

higher than 5%)[20]. Multiple studies have also suggested that KRAS and BRAF gene mutations in colorectal cancer patients are likely to occur in women and elderly patients and are more common in poorly differentiated mucinous carcinoma and proximal colon cancer[19, 21]. They can be affected by factors such as age, gender, degree of differentiation, type of differentiation, and lesion location.

There are many types of mutations in KARS and BRAF genes, the most common ones are on KRAS codons 12, 13, 61 and 146, and BRAF codon 600[17, 19, 22]. In addition, KRAS gene also has multiple mutations (G12C, G12V, G13D, etc.) occurring simultaneously.

KRAS gene mutations in patients with sporadic colorectal cancer are more likely occurring in codons 12 and 13. BRAF gene mutations are mostly reported in the 600th codon, so these sites have become the focus of research[18, 20].

1.1.3 Molecular biological approach

For CRC gene mutation profiling, sample origin includes the paraffin-embedded pathological samples, surgical resection samples, endoscopic biopsy samples,[9, 23, 24]. Among them, endoscopic biopsy is the most convenient way to detect genetic mutations, but as the amount of diseased tissue in the biopsy sample is very low[25]. It is necessary to select high sensitivity and specific experimental methods in clinical trials.

Molecular approaches at present mainly include direct DNA sequencing and ordinary TaqMan PCR (among them, direct DNA sequencing is the gold standard for detecting gene mutations). However, their sensitivity is highly limited by the low occurrence of mutations in the samples[24]. The method TaqMan PCR mixes the fluorescein-labeled probe with the template DNA. As the DNA unwinding at high temperature (~95 °C), and re-combines (base complementary pairing rule) at low temperature (~60 °C), while the DNA is continuously multiplied, the TaqMan probe with fluorescein is released continuously from template DNA. The real-time intensity of the fluorescent signal is compared with the control sample. The copy number of the target gene in the sample could be measured. Even the ordinary TaqMan PCR method greatly improves the sensitivity and specificity of the experiment through fluorescence quantification and can detect gene mutations that cannot be detected by the traditional method like direct DNA sequencing [26, 27], this technology may also be limited by the sampling procedures.

1.1.4 High resolution mass spectrometry based proteogenomics

Besides DNA-sequencing-based approaches, high resolution mass spectrometry (HRMS) based high-throughput methods to investigate cancer proteome are also critical to identify changes on mutations for which therapeutic treatments are available.

Proteogenomics is a multi-omics analysis system that integrates genomic, transcriptomic, proteomic data, and post-translational modification identification (oncology research mainly focuses on phosphorylation). In the field of tumor research, proteogenomics redefines disease classification from molecular-level big data, mines potential therapeutic targets, and finally discovers and verifies tumor-related gene mutations, expression changes, and key molecular regulatory mechanisms at the protein level for further precise medication guidance and drug development. As reported, mRNA transcript abundance cannot reliably predict protein abundance differences between tumors while comprehensive proteomic analysis can provide functional context to explain genomic abnormalities[28]. Proteogenomics provides a new paradigm for cancer-related molecular biology research.

Most colon cancer cell lines have proved to be especially useful for preclinical research model systems[29]. In this study, we used representative COLO205, HCT116, SW620 and HIEC6 (control) cells lines. As Colo205 cells harbor a BRAF V600E mutation ($_{592}$IGDFGLATVK$_{601}$ to IGDFGLAT<u>E</u>K), HCT116 cells contain a KRAS G13D mutation ($_{6}$LVVVGAGGVGK$_{16}$ to LVVVGAG<u>D</u>VGK), SW620 cells contain a KRAS G12V ($_{6}$LVVVGAGGVGK$_{16}$ to LVVVGA<u>V</u>GVGK). These cell lines provide a window for generate and validate the HRMS based proteogenomics approaches to identify low abundant mutations in proteins/peptides.

1.2 Bottom-up Proteomics

Mass spectrometry (MS) is an analytical technique used to determine the elemental composition of test substances for both qualitative and quantitative applications. For example, MS can be used to identify unknown compounds and/or to determine the structure of a particular compound by observing its fragmentation pattern. In recent decades, MS has played an increasingly key role in proteomics due its speed, specificity, and sensitivity to characterize and identifying peptides and proteins[30-32].

One of those strategies is called "bottom-up" approach(***Figure 1-2***), in which proteins of interest are digested (e.g., via enzymes such as Trypsin, Lys-C, etc.) prior to peptide fragment analysis by MS analysis (MS^1) or tandem MS/MS analysis (MS^2 even MS^3).

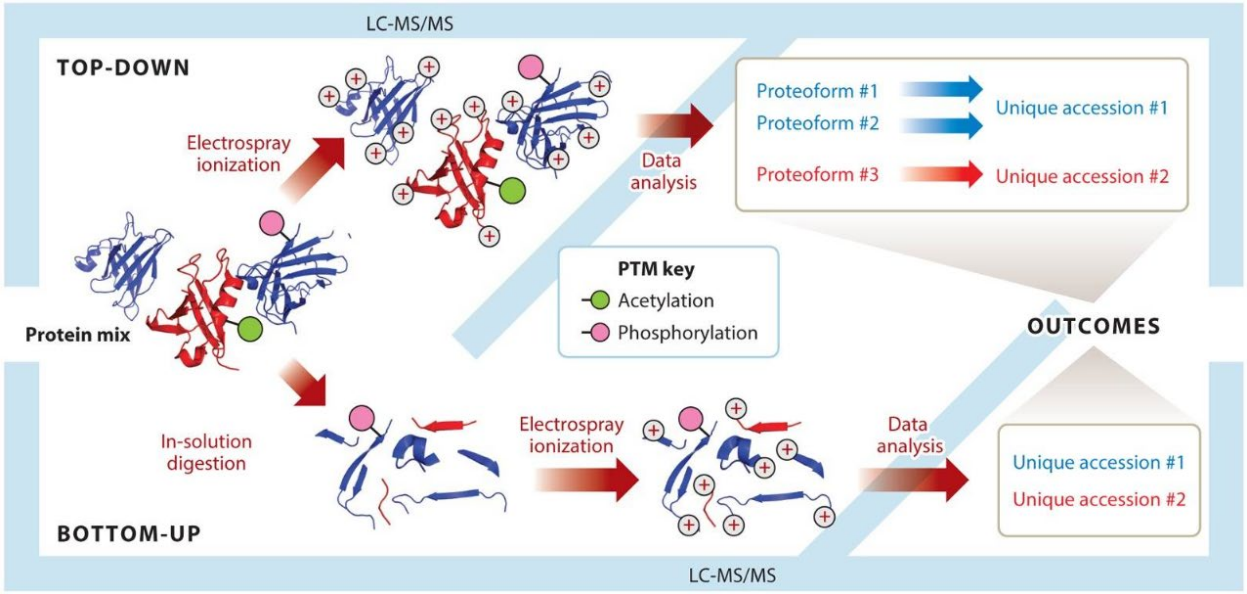


Figure 1-2: Bottom–up and top–down proteomics[33].

In the "bottom-up" MS^2 workflow, the dissociation techniques (e.g., collision-induced dissociation, CID, or higher-energy collisional dissociation, HCD, see ***Chapter 1.3.3. Tandem mass spectrometry***) are used to further fragment precursor peptides fragments selected in the primary MS stage into product ion fragments. During the dissociation, high-energy collisions between precursor ions and inert neutral gases and/or nitrogen molecules dissociate (cleave) the backbone amide bond, thereby producing b-type (n-terminal) and y- type (c-terminal) product ions.

The amino acid sequence of the precursor peptide ion can then be deduced from the masses of the product ion fragments. By identifying these product ion peptides and assembling a putative peptide sequence, the protein can be identified (e.g., by searching the putative sequence against known protein or genomic databases)[34, 35].

1.2.1 Sample preparation for bottom-up proteomics analysis

The solubility of proteins in the native state is generally low and increasing the solubility of the protein will improve digestion efficiency. Simultaneous solubilization of all proteins remains challenging due to the protein heterogeneity and the influence of non-proteinaceous impurities. At present, the principle of increasing protein solubility is to add reagents to destroy the internal protein bonds, including disulfide bonds, hydrogen bonds, van der Waals forces, ionic bonds, and hydrophobic bonds, etc., so that proteins are more easily cleaved into individual polypeptides[36] by enzymes like trypsin.

Urea is a neutral chaotropic agent. Urea at the concentration of 5-9 mol/L can effectively disrupt the secondary structure of protein, unfold the protein, and expose ionizable groups to enhance its solubility. Experiments have shown that adding thiourea to the urea solution will further increase the solubility of some water-soluble proteins and transmembrane proteins[37]. It is worth noting that urea and thiourea can be hydrolyzed at higher temperatures into cyanate and thiocyanate, respectively, resulting in artificial protein modification, so the sample temperature should not exceed 37 °C[36].

Improving the solubility of membrane-associated proteins that form complexes with membrane lipids is currently one of the biggest challenges in sample preparation for MS based proteomics. Detergents with the concentration generally at 1% to 4% are helpful for the extraction of membrane protein complexes. Ionic detergents (e.g., SDS) usually have favourable solubilization effects on hydrophobic proteins and membrane proteins but interfere with native electrophoresis and isoelectric focusing. To replace SDS-PAGE, zwitterionic (e.g. CHAPS) and non-ionic (e.g. TritonX-100) detergents are always the better choices[38].

In addition, specific protease inhibitors should be added after cell lysis or isolation. The human body contains more than 700 kinds of proteases, and the presence of proteases can cause the enzymatic hydrolysis of various proteins[39]. Protease inhibitors, such as methanesulfonyl

fluoride (MSF), 4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride (AEBSF), ethylenediaminetetraacetic acid (EDTA), benzamidine, etc., can effectively inhibit protease activity. Also, when extracting a small amount of protein samples, adding a tris buffer at relatively high pH(~8.2) after boiling the SDS buffer can also inhibit protease[40]. Study has shown that heat shock proteins can inhibit the action of proteases in vitro to protect proteins from enzymatic hydrolysis[41]. In the current thesis, urea was used as a protease inhibitor.

After the above treatment, the protein solubility increases, we use digestive enzymes to break down the protein into peptides. Trypsin is usually the first choice, which specifically cleaves the C-terminal side of lysine and arginine residues[42]. At the same time, other enzymes (e.g. Lys-C, Asp-N, Glu-C) can also be used in this process to increase the efficiency of digestion and thus improve the success rate of protein identification[43]. A two-enzyme digestion by Lys-C and trypsin was used in this thesis.

Before injecting the peptide mixture into the instrument, a clean-up step can be used to remove interfering substances, such as buffers, salts, lipids, carbohydrates, nucleic acids, and detergents introduced during the experimental processing, so that they do not hinder protein separation, ionization, and identification. This can improve detection sensitivity and success rate and can also reduce the complexity of data analysis. A “one-pot” sample preparation strategy is often employed because it reduces sample preparation steps, time, and losses[44, 45]. Stage tips[46, 47], functionalized beads[48], or functionalized nanoparticles[49] are often used for sample preparation to digest, label, and enrich low-volume samples, which can help when thousands proteins need to be identified[47].

1.3 Mass spectrometry instrumentation

Over two to three decades of development, mass spectrometry has gradually replaced Edman degradation sequencing for protein identification[50]. Peptide sequencing of tryptic peptides is performed based on the observation of fragment ions during tandem mass spectrometry[51]. The basic components of a mass spectrometer are the ion source, the mass analyzer, and the detector.

1.3.1 Electrospray ionization

Electrospray ionization (ESI) is currently the most commonly used ion source for liquid chromatography-mass spectrometry (LC-MS). As a soft ionization method, ESI can be used to study thermally unstable and polar compounds, such as proteins and other biological macromolecules. Fenn et al. first demonstrated the applicability of ESI for biomolecular ion analysis[52]. In the 1960s, Dole et al. produced the first Electrospray mass spectrometer[53]. In 1987, Bruins et al. introduced a pneumatically assisted nebulizer into the electrospray interface. The atomizing gas helps to stabilize the electrospray, keeping the flow rate approx. to 0.2 mL/min, which is suitable for liquid phase tandem ESI-MS. This design allows a larger distance between the nozzle and the counter electrode and reduces the occurrence of corona discharge[54].

In ESI-MS, the sample solution is passed through the capillary at a low flow rate (0.1~10 L/min). A high voltage (2~5 kV) is applied to the capillary, and the positive or negative voltage depends on the nature of the analyte. The voltage provides the electric field gradient required for the charge accumulation on the liquid surface. Under the action of the electric field, the liquid forms a "Taylor cone" at the capillary tip (***Figure 1-3***) [55]. The mechanism of Taylor cone formation is not yet clear, but it was shown that under certain conditions, the shape of the cone depends on the capillary voltage and is related to the pulsation of the fluid in the capillary[56]. When the solution at the Taylor cone tip reaches the Rayleigh limit, that is, the critical point where the coulomb repulsion of the surface charge is equivalent to the surface tension of the solution, the tip of the cone will produce droplets with a large number of charges. As the solvent evaporates, the droplet shrinks, and the repulsive force between charges in the droplet increases. When the Rayleigh limit is reached and exceeded, the droplet will undergo a Coulomb explosion, removing the excess charge on the droplet surface and generating smaller charged droplets[57]. Charged droplets undergo new rounds of explosion and finally gas-phase ions are obtained.

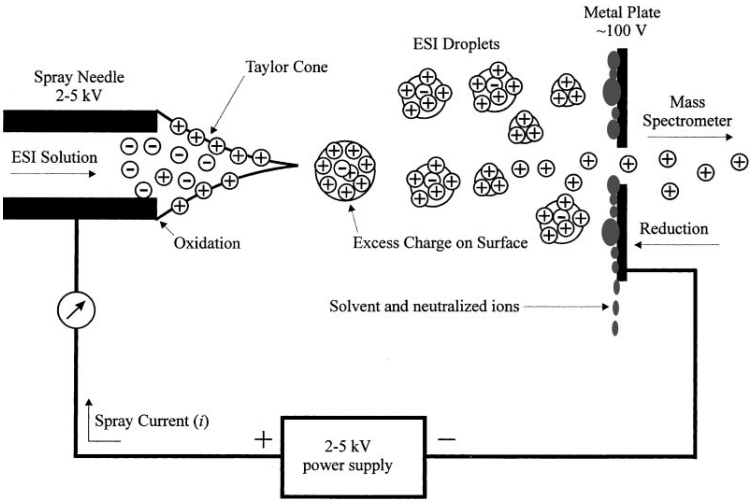


Figure 1-3: Schematic of the electrospray ionization process.[55]

For mass spectrometry-based proteomic analysis, flow rates are generally in the nanoflow range (nL/min), and such ion sources are also known as nano-electrospray (nano-ESI)[58]. When peptides are analyzed, a positive voltage of ~2-3 kV is typically applied between the capillary outlet and the heated MS entrance, which is generated by a high voltage applied at the emitter tip. For peptide analyses, positive mode is mostly used because peptides are usually diluted in low pH aqueous solutions, which make the peptide side chains and N-terminal amino groups protonated and the whole molecule is positively charged[59, 60].

1.3.2 Mass analyzer

Common mass analyzers used for proteomic analysis include quadrupole (Q), time of flight (TOF), Magnetic sector (B), ion trap (IT), orbital trap (OT) and Fourier transform ion cyclotron resonance (FT-ICR).[61] Each mass analyzer has its own physical design and has its own characteristics in terms of resolution, sensitivity, scan speed, etc[50, 62]. Based on the practical choices for this study, we focus on the Q and OT analyse.

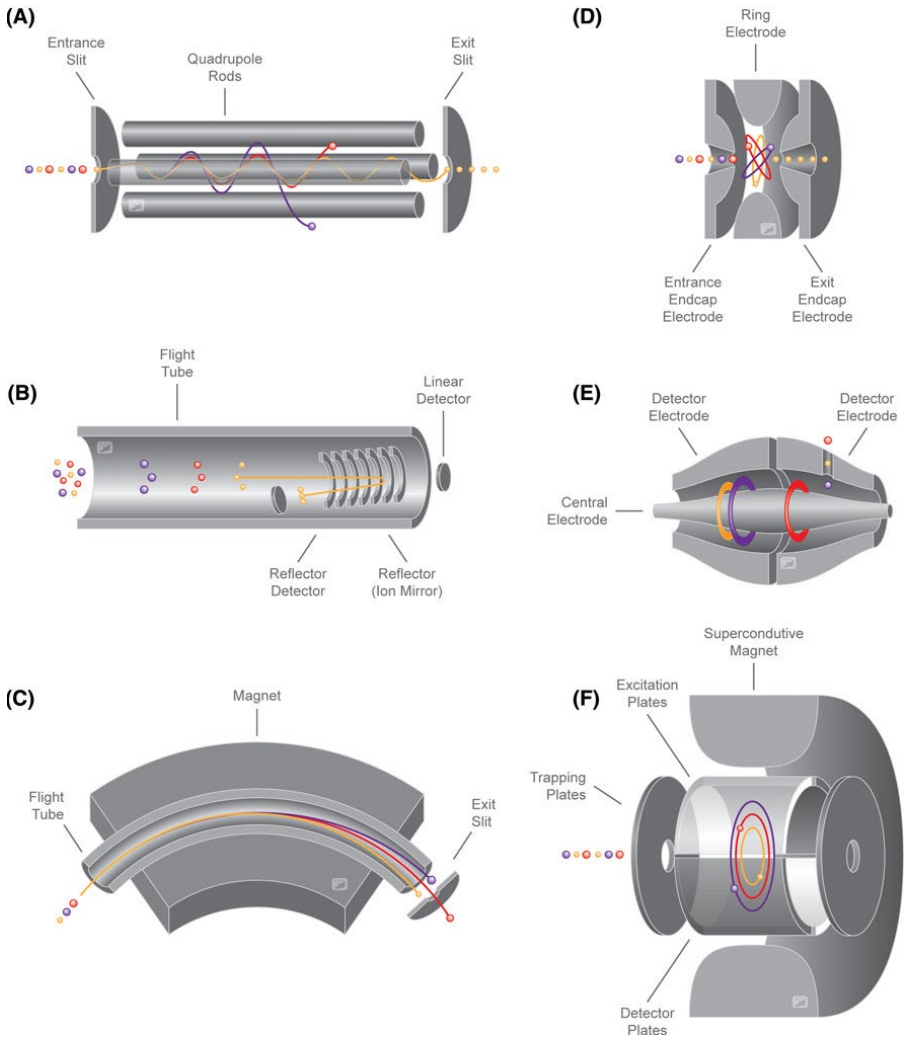


Figure 1-4: Schematic description of six mass analyzers used in mass spectrometers that are currently available on the market.[61]

a Quadrupole (Q).b Time-of-flight (TOF). c Magnetic sector (B). d Ion trap (IT). e Orbitrap (OT). f Ion cyclotron resonance (ICR).[61]

1.3.2.1 Linear quadrupole

The quadrupole is composed of two pairs of hyperbolic electrode rods with opposite polarities. An electric field is created by the combination of radio frequency (RF) and DC offset voltages applied between one rod pair and the other. (See ***Figure 1-4 a***) When a group of ions with different mass-to-charge ratios (m/z) enter the electric field, only the ions that meet certain conditions oscillate stably through the quadrupole rods and reach the detector[63].

1.3.2.2 Orbitrap

Orbitrap is an innovative mass analyzer invented by Makarov[64] based on the early Kingdon trap[65]. Orbitrap can be regarded as a variant of the quadrupole ion trap, the difference is that the orbitrap uses a logarithmic electrostatic field, while the quadrupole ion trap uses a radio frequency electric field. After tangential injection in the Orbitrap, ions are moving around and along the inner electrode and are contained in the space defined by the outer electrode.

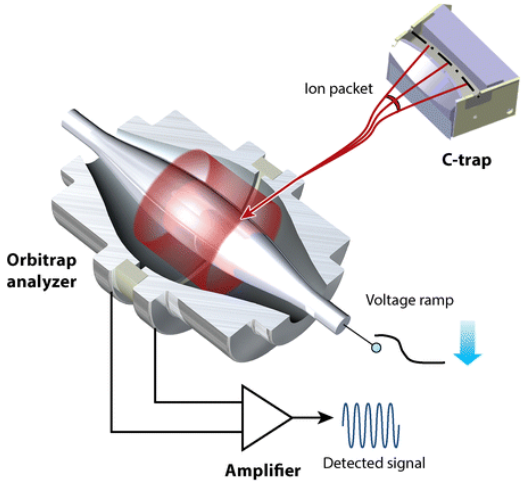


Figure 1-5: The schematic of Orbitrap. [66]

The mass-to-charge ratio of an ion is only related to the motion frequency of the ion in the z direction[66], as shown in equation (1-1), ω is the frequency of axial oscillations(in rad/s), k is the field curvature, z for ionic charge, m for ionic mass:

$$\omega = \sqrt{(z/m)k} \tag{1-1}$$

Similar to FT-ICR MS, the Orbitrap uses the image current to detect the ions in the trap, and the obtained time domain transient is converted into a frequency signal by Fourier transform. Although the radial motion frequency and angular frequency of ions are also related to the mass, the axial frequency is completely independent of the ion energy and the spatial position of the ion, so the axial frequency is used for detection. The frequency of ion motion is independent of the ion energy, so the Orbitrap has the advantages of high resolution and mass accuracy. Since the trap potential is independent of the mass-to-charge ratio and has a larger trap volume than the FT-ICR

and Paul ion traps, the Orbitrap has a high space charge storage capacity at the high mass end. The Orbitrap has a higher mass range than other trap mass spectrometers[65, 66].

At present, the Orbitrap mass analyzer has been successfully commercialized by Thermo Fisher, and a number of mass spectrometry instruments based on Orbitrap technology have been launched. For example, the Q-Exactive mass spectrometer combines the target ion-selective performance of a quadrupole with high-resolution, high-accuracy Orbitrap technology; The Orbitrap Lumos and Eclipse combine linear ion trap and Orbitrap to provide the best structural characterization of the analyte, which is achieved using different kinds of ion fragmentations when performing tandem mass spectrometry analysis. [*http://www. thermoscientific. com*.].

1.3.3. Tandem mass spectrometry and ion dissociation

In tandem mass spectrometry, multiple mass analyzers are connected in series. Such a design allows precursor ions (or: parent ions) to be isolated by the first mass analyzer, fragmented into product ions in a collision cell, and fragment ion m/z are then measured in another mass analyzer, which can provide more structural information.

Fragmentation of ions in tandem mass spectrometry is usually initiated by gas collisions, electron interactions or photon absorption[67, 68].

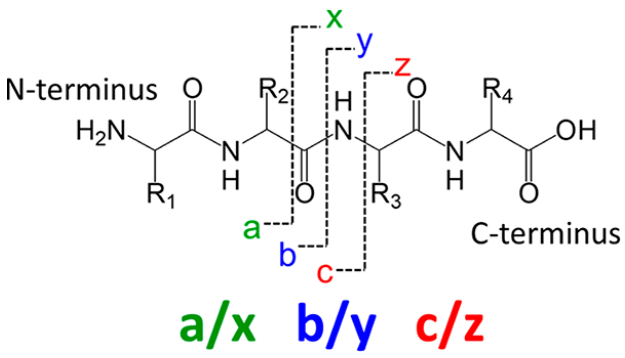


Figure 1-6: Nomenclature for peptide and protein fragmentation.[68]

Normally, fragmentation achieved by collision-induced dissociation (CID)， higher energy collision-induced dissociation (HCD), and electron transfer dissociation (ETD) could be used for proteomics analysis. As shown in ***Figure 1-6***, there are generally three fragmentation patterns for peptides, which generate a/x, b/y, and c/z ions[68]. Usually during HCD, collisions between ionized

precursor ions and inert neutral gases convert kinetic energy into internal energy which ultimately dissociate (cleaves) the backbone amide bond, thereby producing b-type (n-terminal) and y- type (c-terminal) product ion. Sometimes HCD also produces x-type ions due to CO loss from b-type ions. The a- and x- ions result from the cleavage of the Cα-C bond in close proximity to the b-/y- and c-/z- ions.

In CID, ions are excited slowly; as energy builds up during resonance excitation, the ions collide with the entire target gas and have time to redistribute the energy, which means that only the weakest bonds are broken[69-71]. However, in HCD mode, higher energy brings the precursor ion to a higher excited state. High energy state leads to multiple bond breaking during dissociation step, providing more fragment ions[72].

1.3.4. Hybrid mass spectrometer

As a practical example of a tandem mass spectrometer, the hybrid mass spectrometry system commercialized by Thermo scientific with the Q-Exactive series[73, 74] and the newer generation of mass spectrometers like the Exploris series[75] combines a quadrupole and an orbitrap. The quadrupole filters the ion beam before its accumulation in the C-trap and analysis in the orbitrap. A schematic diagram of a typical hybrid mass spectrometer (Exploris 480™) is shown in ***Figure 1-7***.

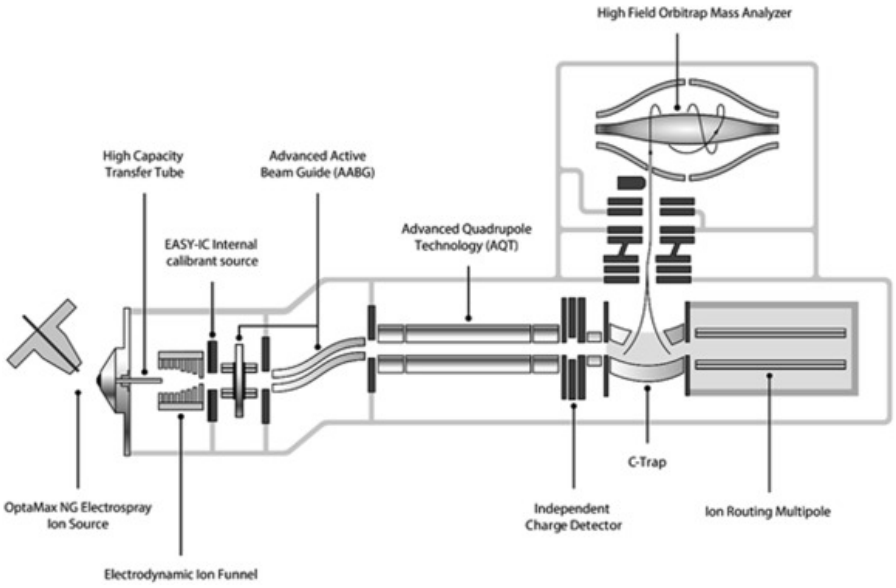


Figure 1-7: Scheme of the hybrid Orbitrap mass spectrometer using in this thesis. [75]

The Exploris 480 is an upgraded version of the Q-Exactive HF. Both instruments are equipped with a high-performance quadrupole and a high-resolution accurate mass (HRAM) Orbitrap detector. Such a hybrid mass spectrometer design mainly includes ion source, stacked ring ion guide (S-lens), quadrupole mass filter, curved linear trap (C-trap), a high energy collision dissociation (HCD) chamber and an Orbitrap mass analyzer. The sample can be introduced into the ion source after ionization. A beam of vaporized charged particles flies into the mass spectrometer under the action of the electric field between the capillary outlet and the MS ion transfer tube concurrently under the action of vacuum. S-Lens in the ion source is focusing the ion cloud into an ion beam. The bent flat-pole transports ions and removes neutral gas and solvent droplets. The quadrupole acts as an ion transport device that filters the transmitted ions based on their mass-to-charge ratio. The ions are transferred into the C-Trap and then injected into the Orbitrap mass analyzer to obtain a mass spectrum. In addition, sample ions can enter the HCD collision cell via the C-trap, which consists of a straight multipole rod mounted inside a metal tube. The front of the tube houses a lens that regulates the delivery and jetting of the C-Trap. A potential gradient is applied to the collision cell to rapidly extract ions. Fragmentation spectra generated in the HCD chamber and detected in the Orbitrap showed fragmentation patterns comparable to typical triple quadrupole mass spectra. The Q-Exactive HF system can perform 18 complete MS/MS scans in 1 second with a maximum resolution of 240,000. The Exploris 480 can complete data acquisition at faster scan rates of up to 40 MS/MS scans per second, with a higher resolution of 480,000. [73-75]

1.3.5. Analysis mode

A mass spectrometer usually provides different analysis modes[76]. In discovery mode, the duty cycle consists of the acquisition of a MS scan followed by the MS/MS acquisition of the most abundant precursor ions detected in the MS scan. In targeted analysis, the ions of interests with known m/z can be selectively targeted by using the analysis mode such as selected reaction monitoring (SRM), multiple reaction monitoring (MRM), or parallel reaction monitoring (PRM)[77].

The SRM/MRM method separates and resolves predefined precursor ions and monitors predefined product ions derived from the precursor ions. PRM separates and resolves predefined

precursor ions in the same manner as SRM/MRM but records all product ions of the predefined precursor ions. ***Figure 1-8***

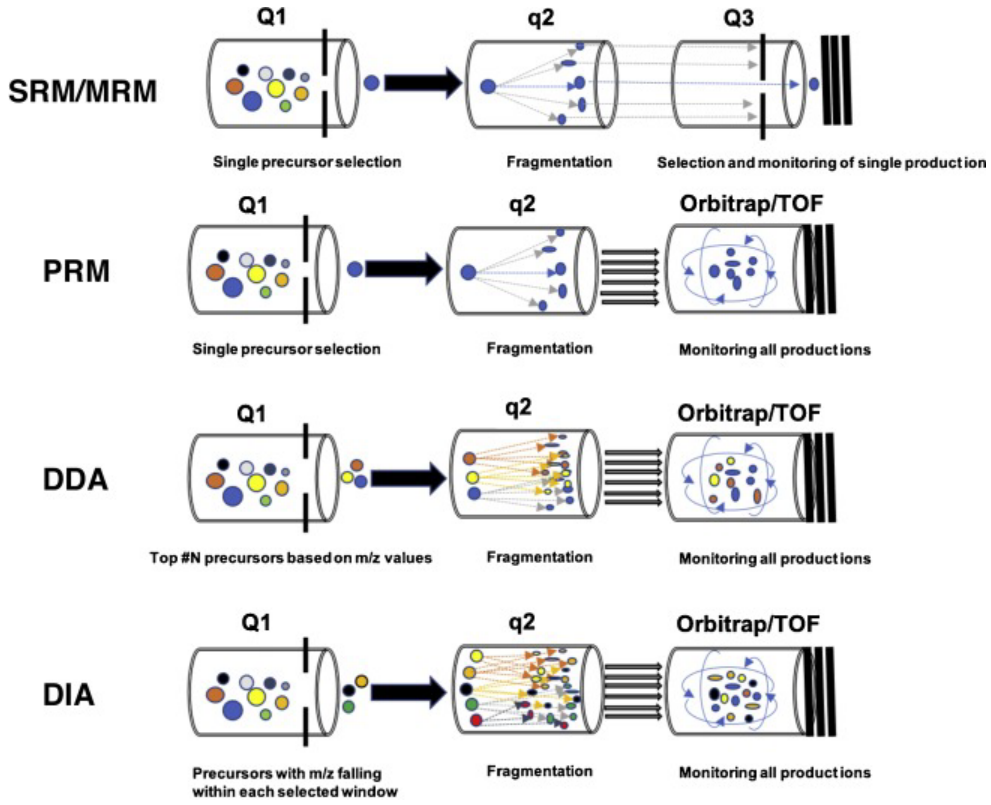


Figure 1-8: Schematic of SRM/MRM, PRM, DDA, and DIA proteomics.[77]

1.3.5.1. Data Dependent Acquisition (DDA)

Two types of data acquisition are encountered in proteomics: DDA (Data Dependent Acquisition) or DIA (Data Independent Acquisition). The choice is made when designing experiments. At present, there are many studies based on these two acquisition modes. In DIA all ions in a selected m/z range are fragmented simultaneously and their fragments are detected by tandem mass spectrometry[78]. Fragmentation information of all ions can theoretically be comprehensively obtained. But the data analysis usually requires professional analysis software and a library[79, 80] of fragmentation patterns.

DDA can simultaneously obtain primary mass spectrum and fragmentation information of the detected peptides[81]. The screening of precursor ions mainly depends on parameters pre-set by the user, such as resolution, isotope distribution, ion intensity threshold, selection of top-N, etc.

This method can provide researchers with higher-quality fragmentation information because a narrow m/z (usually ~1 Da) window is used to isolate targeted ions and reduce the presence of interfering ions[82]. Meanwhile, the main challenge of DDA is that the screening of targeted ions is stochastic, and ions with higher intensities are more likely to be selected for tandem mass spectrometry information acquisition and sometimes the quality of MS/MS scans are not sufficient to get confident identification[82]. Besides, if the number of precursor ions detected is too large, it will cause insufficient sampling of available precursors limiting the depth of the analysis[83]. Therefore, the users must decide how to optimize the acquisition method to obtain the most sample information.

1.3.5.2. Full and segmented mass acquisition modes

The high complexity of the sample usually results in a highly populated MS scan. Even the Orbitrap mass spectrometer has made great strides in sensitivity and acquisition speed, enabling deeper proteome coverage. These advances were largely made at the MS/MS level, and ion acquisition for primary MS scans is still very inefficient. In 2018, Mann Lab introduced a data acquisition method called BoxCar ***(Figure1-9)***, which is a segmented MS acquisition method, increasing the average ion injection time by more than 10 times compared with the standard full scan. In a 1-h analysis of a human cancer cell line, the method identified more than 90% of the proteins previously identified in 24 fractions by DDA and quantified more than 6200 proteins. In mouse brain tissue, more than 10,000 proteins were identified in just 100 min, extending sensitivity down to atto-molar levels[84]. This BoxCar method significantly improves the depth of protein identification in samples with a large range of concentration, such as plasma or tissue samples[85, 86].

It seems that BoxCar does provide a good opportunity to collect low-abundance peptides by accumulating a small scan mass range. However, the interpretation of data requires a spectral library, which makes the entire data analysis complex and time-consuming. In addition, the instrument produces many gaps in the whole scanning cycle due to frequent switching of scanning windows during acquisition. Potentially, we lose some acquisition time. Maybe we need larger m/z ranges, based on the complexity of the sample, which is the experimental technique which is called segmented ion fractionation (SIFT) pioneered by our group that I am pursuing in the experiments presented in chapters 2 and 3 based on gas phase fractionation technology[87].

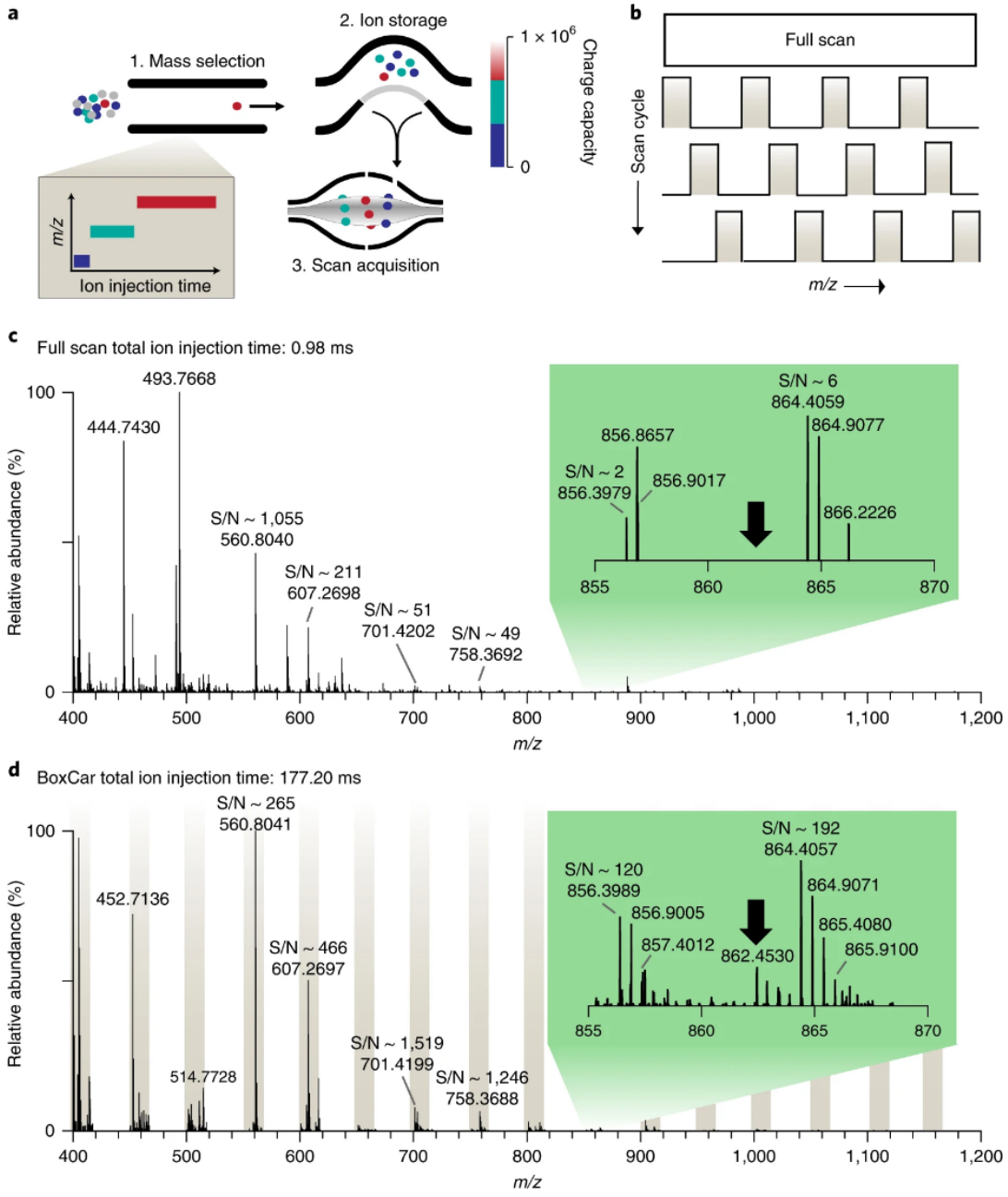


Figure 1-9: The BoxCar acquisition method. [84]

1.4 Quantitative proteomics

Qualitative analysis of protein types and modifications is insufficient for many applications in scientific research. The quantitative aspect of proteomics is to monitor protein and peptide abundance changes from peak intensities in either the MS scan or the MS/MS scan. Proteomic

quantitation mainly includes non-targeted relative quantitative technology and targeted absolute quantitative technology[88, 89].

Quantitative proteomics has high requirements on the reproducibility, sensitivity, and detection efficiency of mass spectrometry data. Two kinds of targeted mass spectrometry quantification are often adopted: multiple reaction monitoring (MRM) and parallel reaction monitoring (PRM). Technology[90, 91].

1.4.1. Label free

The label-free quantification technology (LFQ) of proteome is a relative quantification technology that directly relies on the ion signal intensity. This technology does not require expensive peptide labeling reagents and is suitable for the relative quantification of any type of protein, with high quantification accuracy.

The signal intensity method was first proposed and verified by Chelius et al[92]. It is based on the intensity of precursor ions in the full-scan spectrum, and quantitatively monitor protein changes by comparing the area or signal intensity of precursor ions across samples. Based on the principle of bottom-up method, DDA collects precursor ions from high to low on intensity for secondary fragmentation, but the result is random to a certain extent, technical/biological replicates are always needed. Besides, we may rely on a calibration curve to provide an absolute quantitation.

1.4.2. Tandem Mass Tags

To obtain more accurate quantitative results, the quantification could be carried out by analyzing the product ions of chemically labeled peptides using the reporter ions generated in the tandem mass spectrum. Typical isobaric tag for relative and absolute quantification includes isobaric tag (iTRAQ™) [93], and tandem mass tags (TMT™) [94], they are a batch of chemical labeling reagents called tag with the same masses but with isotopes located on different atoms along the tag. This allows the same peptides from different components of the proteins to be selected by primary mass analyzer at the same time after being tagged with isobaric labels, and the relative quantification can be accomplished by comparing reporter ion intensities. ***(Figure1-10)***At present, the commercialized iTRAQ tags are generally 4 markers and 8 markers, and TMT (pro) commonly used 4/6/10/16 tags[94].

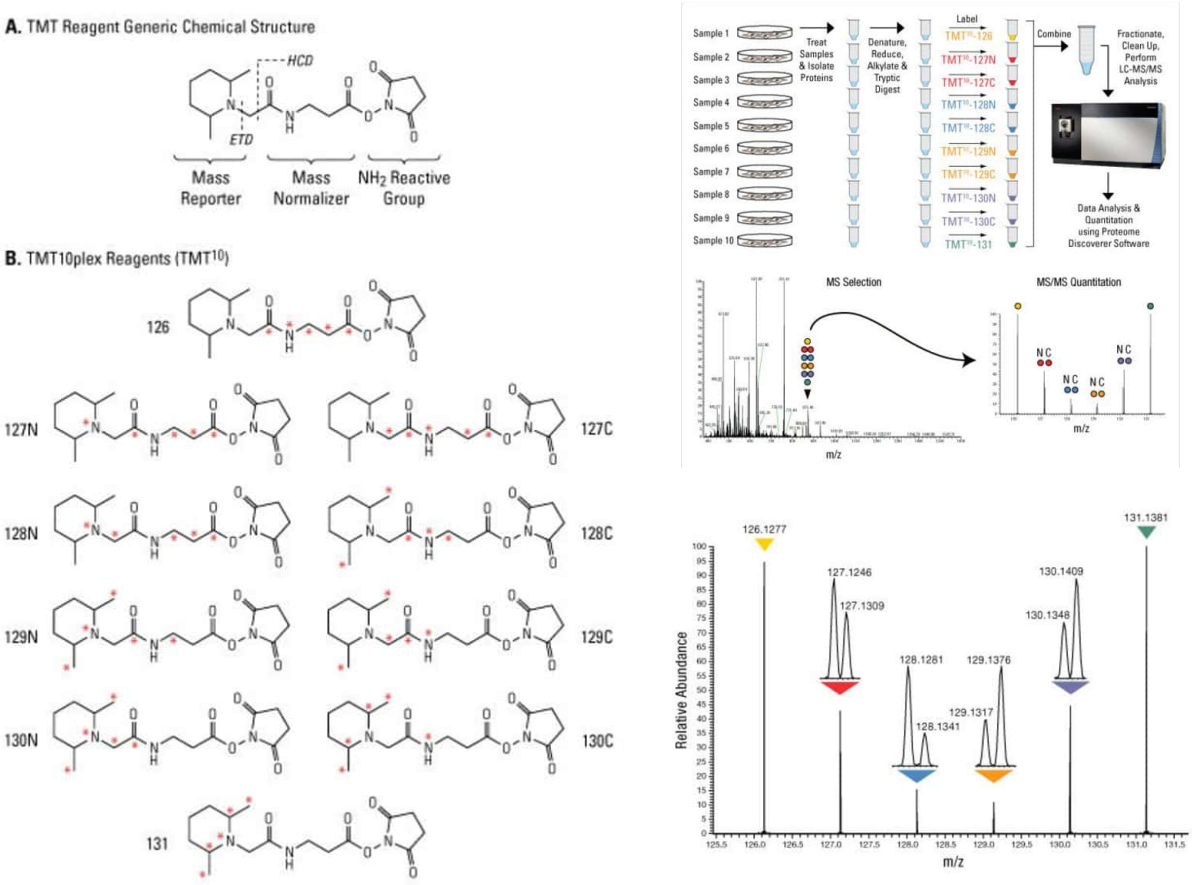


Figure 1-10: TMT labeling reagent. [94]

The benefits of TMT are the multiplexing capability and decreased number of missing values, and the disadvantages are the price, the high resolution needed in MS which slows down the duty cycle and ratio compression.

1.5 Data analysis

Proteomics data processing relies heavily on computer-based software, and the methods mainly include in silico database search and *de novo* sequencing.

The basic principle of in silico database search is to match the acquired MS/MS spectra with computationally generated spectra deduced from the proteome database. At present, the existing database search scoring algorithms mainly include four types of models: descriptive models, explanatory models, stochastic process models and statistical probability models[95]. Based on these scoring algorithms, various database search software has been developed. Among them, the most widely used are SEQUEST[96], MASCOT[97] and X!Tandem[98]. SEQUEST and MASCOT are commercial software, while X!Tandem is an open-source software.

The basic idea of *de novo* sequencing is to find a primary amino acid sequence that is derived from the MS/MS spectrum by taking into account the precursor m/z, charge and m/z differences between y and b ions. A variety of de novo sequencing software have been developed, such as MSNovo[99], PepNovo[100], RAId[101] and PEAKS[102, 103]. The de novo sequencing method does not rely on the existing database. According to the characteristics of the regular fragmentation of the peptide, the sequence of the peptide can be directly deduced from the spectrum. It is an important method to identify new proteins or potential mutations that have not been included in the sequence canonical database (e.g. Uniprot/Swissport). In this thesis, PEAKS™ and Proteome Discoverer™ have been used for MS data processing.

1.6 High field asymmetric-waveform ion mobility spectrometry (FAIMS)

Ion mobility (IM) refers to the moving speed of positive ions or negative ions when the electric field strength is 1 V/m or the electric field force is 1 N, and the unit is m/V. In an ion mobility spectrometry (IMS), the ions are accelerated by the electric field and let into a drift tube. Here, they collide with the buffer gas molecules in the drift zone to generate resistance and reduce the speed. The kinetic energy lost by the ions during the collision can be converted into internal energy to increase the temperature of the ions, and the second collision can transfer the increased internal energy to the gas molecules. Therefore, ion temperature and velocity do not remain constant during traveling into the drift tube. There may also be electrostatic attraction and Coulomb repulsion between ions and buffer gas. Ion movement in the drift region is extremely complex and can only be determined by ion average velocity (ion mobility, IM) or ion drift time. For comparison purposes between experiments, the ion mobility is usually converted to the reduced ion mobility, which is the IM at a temperature of 273 K and a pressure of 760 Torr. Ion size and shape can be measured by the average available cross section (collision Cross section, n) when the ions collide with the buffer gas. [104-107]

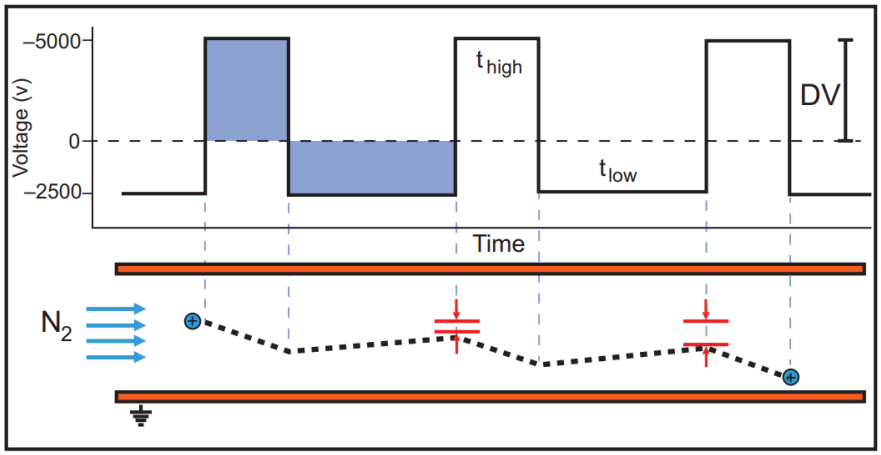


Figure 1-11: FAIMS Asymmetric Waveform.

Ion displacement using a dispersion voltage (DV) [108]

The main difference between an ion mobility mass spectrometer and a conventional mass spectrometer is that the former adds an ion mobility component between the ion source and the mass analyzer. As described above, ion mobility separation is mainly based on the shape and size of the ions. Therefore, this separation method has unique advantages for the analysis of isomers or complexes that cannot be distinguished by conventional mass spectrometry methods.

Field Asymmetric Waveform Ion Mobility Spectrometry (FAIMS), FAIMS is a design based on differential ion mobility, where the analyte is carried by a gas into an electric field alternating between high and low regime between two electrodes. This causes the charged particles to undergo a small deflection displacement during each period of the asymmetric field, as shown in ***Figure 1-11*** and ***Figure 1-12***.[108]

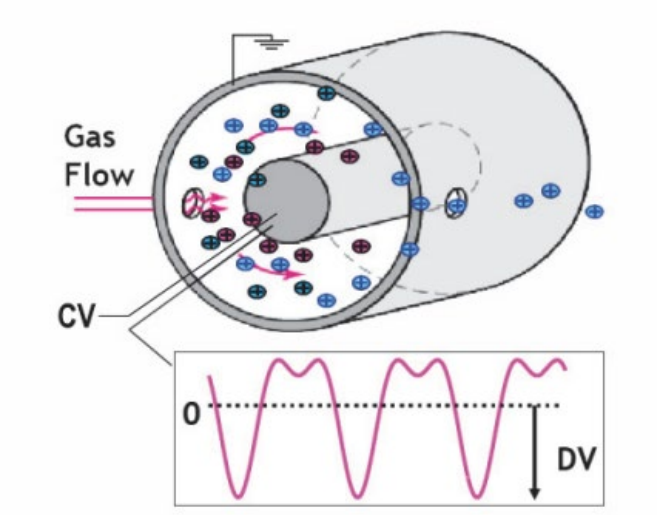


Figure 1-12: FAIMS electrodes and the gas flow in between. [108]

Each ion can only pass through the field smoothly under the effect of a specific compensation voltage. This allows FAIMS to be viewed as an ion filter, transmitting only a fraction of the ion beam. Ion transmission is not based on m/z like mass analyzers, but is based on shape, mass, aggregation and dipole moment between ions and neutral gas molecules.[108, 109]

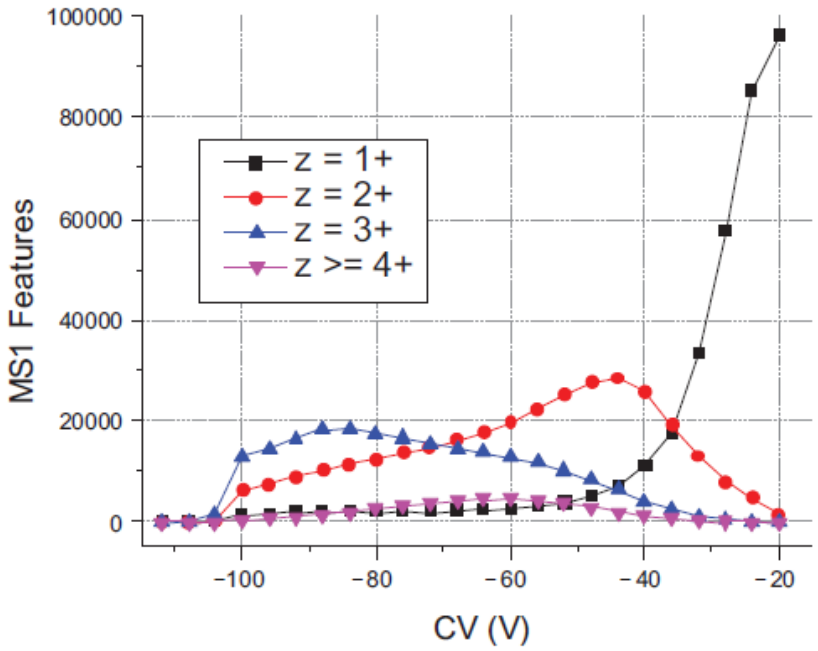


Figure 1-13: An untargeted FAIMS operation as CV fractionation.[108]

FAIMS can separate charge states before entering the mass spectrometer and can be used to selectively transmit multiply charged tryptic peptides. As shown in ***Figure 1-13***, it is encouraging that most of the singly charged ions that are interference can be blocked from the mass spectrometer at CV voltages lower than -30V, which makes FAIMS a background ion filter, this means that it filters out interfering ions before entering the MS[110-112]. We reasoned that it could be an asset for our project.

1.7 Research objectives

In this thesis, we have evaluated the combination of FAIMS and SIFT for proteomic analyses, in order to improve the coverage of the proteome. We need to optimize the experimental parameters on existing Thermo Scientific hybrid mass spectrometers (including Q-Exactive HF™ and Orbitrap Exploris480™), as few proteomic studies have been made using FAIMS. By comparing the experimental results, we could have an in-depth understanding of the gain with FAIMS and/or SIFT with optimized experimental conditions. These assessments are based on qualitative and quantitative analysis techniques on experimental samples. The combination of FAIMS and SIFT will be used to analyze the proteome of colorectal cancer cell lines. Briefly, this thesis has the following objectives:

i. Integration of the FAIMS interface to the Q-Exactive HF and Orbitrap Exploris 480 mass spectrometer.

ii. Optimization of the LC-FAIMS-MS/MS system parameters, associating with the SIFT acquisition, to improve the proteome coverage.

iii. Evaluate the advantages of the FAIMS and/or SIFT strategies in qualitative and quantitative analysis compared to conventional approaches.

iv. Develop a proteomic assay combining FAIMS and SIFT to increase proteome coverage when analyzing the protein digests of CRC cell lines.

v. Combine proteomic analysis and molecular biology experiments (such as RNA sequencing) to complete multi-omics combined analysis of biological samples, and explore the mutational landscape in colon cancer cells

1.8 Overview of thesis

This dissertation will be presented in four chapters. ***Chapter One*** is an introduction to briefly state the experimental methods involved in this study and the relevant research background. This includes the selection of experimental samples, cell lines, sample preparation, basic configuration of mass spectrometers, SIFT acquisition technology and the use of FAIMS interface. In ***Chapter two***, we integrated the FAIMS interface to the Q-Exactive HF mass spectrometer, combing with the SIFT to get a better proteome coverage as well the higher sensitivity and accuracy for quantitative analyses. In ***Chapter Three***, we applied the FAIMS and SIFT to the Orbitrap Exploris 480 mass spectrometer, to evaluate the performance when analyzing the CRC cell line proteome and explore the mutational landscape in CRC cells by using a proteogenomic approach. ***Chapter four*** provides conclusions and perspectives.

1.9 References

1. Voutsina, A.; Tzardi, M.; Kalikaki, A.; Zafeiriou, Z.; Papadimitraki, E.; Papadakis, M.; Mavroudis, D.; Georgoulias, V. Combined analysis of KRAS and PIK3CA mutations, MET and PTEN expression in primary tumors and corresponding metastases in colorectal cancer. *Mod Pathol* **2013**, *26* (2), 302-313.
2. Jobin, C. Colorectal cancer: looking for answers in the microbiota. *Cancer Discov* **2013**, *3* (4), 384-387.
3. Center, M. M.; Jemal, A.; Ward, E. International trends in colorectal cancer incidence rates. *Cancer Epidemiol Biomarkers Prev* **2009**, *18* (6), 1688-1694.
4. Tjalsma, H.; Scholler-Guinard, M.; Lasonder, E.; Ruers, T. J.; Willems, H. L.; Swinkels, D. W. Profiling the humoral immune response in colon cancer patients: diagnostic antigens from Streptococcus bovis. *Int J Cancer* **2006**, *119* (9), 2127-2135.
5. Abdulamir, A. S.; Hafidh, R. R.; Abu Bakar, F. The association of Streptococcus bovis/gallolyticus with colorectal tumors: the nature and the underlying mechanisms of its etiological role. *J Exp Clin Cancer Res* **2011**, *30* (1), 11.
6. Bauer, V. P.; Papaconstantinou, H. T. Management of serrated adenomas and hyperplastic polyps. *Clin Colon Rectal Surg* **2008**, *21* (4), 273-279.
7. Armaghany, T.; Wilson, J. D.; Chu, Q.; Mills, G. Genetic alterations in colorectal cancer. *Gastrointest Cancer Res* **2012**, *5* (1), 19-27.
8. Arrington, A. K.; Heinrich, E. L.; Lee, W.; Duldulao, M.; Patel, S.; Sanchez, J.; Garcia-Aguilar, J.; Kim, J. Prognostic and predictive roles of KRAS mutation in colorectal cancer. *Int J Mol Sci* **2012**, *13* (10), 12153-12168.
9. Velho, S.; Moutinho, C.; Cirnes, L.; Albuquerque, C.; Hamelin, R.; Schmitt, F.; Carneiro, F.; Oliveira, C.; Seruca, R. BRAF, KRAS and PIK3CA mutations in colorectal serrated polyps and cancer: primary or secondary genetic events in colorectal carcinogenesis? *BMC Cancer* **2008**, *8*, 255.
10. Jin, Y. M.; Li, B. J.; Qu, B.; Du, Y. J. BRAF, K-ras and BAT26 mutations in colorectal polyps and stool. *World J Gastroenterol* **2006**, *12* (32), 5148-5152.
11. Thiel, A.; Ristimaki, A. Toward a Molecular Classification of Colorectal Cancer: The Role of BRAF. *Front Oncol* **2013**, *3*, 281.
12. Petanidis, S.; Anestakis, D.; Argyraki, M.; Hadzopoulou-Cladaras, M.; Salifoglou, A. Differential expression of IL-17, 22 and 23 in the progression of colorectal cancer in patients with K-ras mutation: Ras signal inhibition and crosstalk with GM-CSF and IFN-γ. *PLoS One* **2013**, *8* (9), e73616.
13. Phipps, A. I.; Buchanan, D. D.; Makar, K. W.; Burnett-Hartman, A. N.; Coghill, A. E.; Passarelli, M. N.; Baron, J. A.; Ahnen, D. J.; Win, A. K.; Potter, J. D.; et al. BRAF mutation status and survival after colorectal cancer diagnosis according to patient and tumor characteristics. *Cancer Epidemiol Biomarkers Prev* **2012**, *21* (10), 1792-1798.

14. Peeters, M.; Douillard, J. Y.; Van Cutsem, E.; Siena, S.; Zhang, K.; Williams, R.; Wiezorek, J. Mutant KRAS codon 12 and 13 alleles in patients with metastatic colorectal cancer: assessment as prognostic and predictive biomarkers of response to panitumumab. *J Clin Oncol* **2013**, *31* (6), 759-765.

15. Yokota, T. Are KRAS/BRAF mutations potent prognostic and/or predictive biomarkers in colorectal cancers? *Anticancer Agents Med Chem* **2012**, *12* (2), 163-171.

16. Martinetti, D.; Costanzo, R.; Kadare, S.; Alimehmeti, M.; Colarossi, C.; Canzonieri, V.; Berretta, M.; Memeo, L. KRAS and BRAF mutational status in colon cancer from Albanian patients. *Diagn Pathol* **2014**, *9*, 187.

17. Clarke, C. N.; Kopetz, E. S. BRAF mutant colorectal cancer as a distinct subset of colorectal cancer: clinical characteristics, clinical behavior, and response to targeted therapies. *J Gastrointest Oncol* **2015**, *6* (6), 660-667.

18. Sotelo, M. J.; Garcia-Paredes, B.; Aguado, C.; Sastre, J.; Diaz-Rubio, E. Role of cetuximab in first-line treatment of metastatic colorectal cancer. *World J Gastroenterol* **2014**, *20* (15), 4208-4219.

19. Tong, J. H.; Lung, R. W.; Sin, F. M.; Law, P. P.; Kang, W.; Chan, A. W.; Ma, B. B.; Mak, T. W.; Ng, S. S.; To, K. F. Characterization of rare transforming KRAS mutations in sporadic colorectal cancer. *Cancer Biol Ther* **2014**, *15* (6), 768-776.

20. Li, L.; Ma, B. B. Colorectal cancer in Chinese patients: current and emerging treatment options. *Onco Targets Ther* **2014**, *7*, 1817-1828.

21. Kumar, S. S.; Price, T. J.; Mohyieldin, O.; Borg, M.; Townsend, A.; Hardingham, J. E. KRAS G13D Mutation and Sensitivity to Cetuximab or Panitumumab in a Colorectal Cancer Cell Line Model. *Gastrointest Cancer Res* **2014**, *7* (1), 23-26.

22. Saridaki, Z.; Tzardi, M.; Sfakianaki, M.; Papadaki, C.; Voutsina, A.; Kalykaki, A.; Messaritakis, I.; Mpananis, K.; Mavroudis, D.; Stathopoulos, E.; et al. BRAFV600E mutation analysis in patients with metastatic colorectal cancer (mCRC) in daily clinical practice: correlations with clinical characteristics, and its impact on patients' outcome. *PLoS One* **2013**, *8* (12), e84604.

23. Yang, S.; Farraye, F. A.; Mack, C.; Posnik, O.; O'Brien, M. J. BRAF and KRAS Mutations in hyperplastic polyps and serrated adenomas of the colorectum: relationship to histology and CpG island methylation status. *Am J Surg Pathol* **2004**, *28* (11), 1452-1459.

24. Chan, T. L.; Zhao, W.; Leung, S. Y.; Yuen, S. T. BRAF and KRAS mutations in colorectal hyperplastic polyps and serrated adenomas. *Cancer Res* **2003**, *63* (16), 4878-4881.

25. Roma, C.; Esposito, C.; Rachiglio, A. M.; Pasquale, R.; Iannaccone, A.; Chicchinelli, N.; Franco, R.; Mancini, R.; Pisconti, S.; De Luca, A.; et al. Detection of EGFR mutations by TaqMan mutation detection assays powered by competitive allele-specific TaqMan PCR technology. *Biomed Res Int* **2013**, *2013*, 385087.

26. Negru, S.; Papadopoulou, E.; Apessos, A.; Stanculeanu, D. L.; Ciuleanu, E.; Volovat, C.; Croitoru, A.; Kakolyris, S.; Aravantinos, G.; Ziras, N.; et al. KRAS, NRAS and BRAF mutations in Greek and Romanian patients with colorectal cancer: a cohort study. *BMJ Open* **2014**, *4* (5), e004652.

27. Chretien, A. S.; Harle, A.; Meyer-Lefebvre, M.; Rouyer, M.; Husson, M.; Ramacci, C.; Harter, V.; Genin, P.; Leroux, A.; Merlin, J. L. Optimization of routine KRAS mutation PCR-based testing procedure for rational individualized first-line-targeted therapy selection in metastatic colorectal cancer. *Cancer Med* **2013**, *2* (1), 11-20.

28. Zhang, B.; Wang, J.; Wang, X.; Zhu, J.; Liu, Q.; Shi, Z.; Chambers, M. C.; Zimmerman, L. J.; Shaddox, K. F.; Kim, S.; et al. Proteogenomic characterization of human colon and rectal cancer. *Nature* **2014**, *513* (7518), 382-387.

29. Mouradov, D.; Sloggett, C.; Jorissen, R. N.; Love, C. G.; Li, S.; Burgess, A. W.; Arango, D.; Strausberg, R. L.; Buchanan, D.; Wormald, S.; et al. Colorectal cancer cell lines are representative models of the main molecular subtypes of primary cancer. *Cancer Res* **2014**, *74* (12), 3238-3247.

30. Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R., 3rd. Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol* **1999**, *17* (7), 676-682.

31. Aebersold, R.; Mann, M. Mass-spectrometric exploration of proteome structure and function. *Nature* **2016**, *537* (7620), 347-355.

32. Catherman, A. D.; Skinner, O. S.; Kelleher, N. L. Top Down proteomics: facts and perspectives. *Biochem Biophys Res Commun* **2014**, *445* (4), 683-693.

33. Toby, T. K.; Fornelli, L.; Kelleher, N. L. Progress in Top-Down Proteomics and the Analysis of Proteoforms. *Annu Rev Anal Chem (Palo Alto Calif)* **2016**, *9* (1), 499-519.

34. Roepstorff, P.; Fohlman, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom* **1984**, *11* (11), 601.

35. Biemann, K. Contributions of mass spectrometry to peptide and protein structure. *Biomed Environ Mass Spectrom* **1988**, *16* (1-12), 99-111.

36. Matt, P.; Fu, Z.; Fu, Q.; Van Eyk, J. E. Biomarker discovery: proteome fractionation and separation in biological samples. *Physiol Genomics* **2008**, *33* (1), 12-17.

37. Joo, J. I.; Kim, D. H.; Choi, J. W.; Yun, J. W. Proteomic analysis for antiobesity potential of capsaicin on white adipose tissue in rats fed with a high fat diet. *J Proteome Res* **2010**, *9* (6), 2977-2987.

38. Luche, S.; Santoni, V.; Rabilloud, T. Evaluation of nonionic and zwitterionic detergents as membrane protein solubilizers in two-dimensional electrophoresis. *Proteomics* **2003**, *3* (3), 249-253.

39. Wu, C. H.; Apweiler, R.; Bairoch, A.; Natale, D. A.; Barker, W. C.; Boeckmann, B.; Ferro, S.; Gasteiger, E.; Huang, H.; Lopez, R.; et al. The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Res* **2006**, *34* (Database issue), D187-191.

40. Ahn, S.-M.; Simpson, R. J. Proteomic Strategies for Analyzing Body Fluids. In *Proteomics of Human Body Fluids: Principles, Methods, and Applications*, Thongboonkerd, V. Ed.; Humana Press, 2007; pp 3-30.

41. Han, M. J.; Lee, J. W.; Lee, S. Y. Enhanced proteome profiling by inhibiting proteolysis with small heat shock proteins. *J Proteome Res* **2005**, *4* (6), 2429-2434.

42. Tsiatsiani, L.; Heck, A. J. Proteomics beyond trypsin. *FEBS J* **2015**, *282* (14), 2612-2626.

43. Trevisiol, S.; Ayoub, D.; Lesur, A.; Ancheva, L.; Gallien, S.; Domon, B. The use of proteases complementary to trypsin to probe isoforms and modifications. *Proteomics* **2016**, *16* (5), 715-728.

44. Mayne, J.; Ning, Z.; Zhang, X.; Starr, A. E.; Chen, R.; Deeke, S.; Chiang, C. K.; Xu, B.; Wen, M.; Cheng, K.; et al. Bottom-Up Proteomics (2013-2015): Keeping up in the Era of Systems Biology. *Anal Chem* **2016**, *88* (1), 95-121.

45. Kanshin, E.; Tyers, M.; Thibault, P. Sample Collection Method Bias Effects in Quantitative Phosphoproteomics. *J Proteome Res* **2015**, *14* (7), 2998-3004.

46. Rappsilber, J.; Mann, M.; Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* **2007**, *2* (8), 1896-1906.

47. Kulak, N. A.; Pichler, G.; Paron, I.; Nagaraj, N.; Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat Methods* **2014**, *11* (3), 319-324.

48. Hughes, C. S.; Foehr, S.; Garfield, D. A.; Furlong, E. E.; Steinmetz, L. M.; Krijgsveld, J. Ultrasensitive proteome analysis using paramagnetic bead technology. *Mol Syst Biol* **2014**, *10*, 757.

49. Geyer, P. E.; Kulak, N. A.; Pichler, G.; Holdt, L. M.; Teupser, D.; Mann, M. Plasma Proteome Profiling to Assess Human Health and Disease. *Cell Syst* **2016**, *2* (3), 185-195.

50. Domon, B.; Aebersold, R. Mass spectrometry and protein analysis. *Science* **2006**, *312* (5771), 212-217.

51. Haag, A. M. Mass Analyzers and Mass Spectrometers. *Adv Exp Med Biol* **2016**, *919*, 157-169.

52. 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.

53. Dole, M.; Mack, L. L.; Hines, R. L.; Mobley, R. C.; Ferguson, L. D.; Alice, M. B. Molecular Beams of Macroions. *J Chem Phy* **1968**, *49* (5), 2240-2249.

54. Bruins, A. P.; Covey, T. R.; Henion, J. D. Ion spray interface for combined liquid chromatography/atmospheric pressure ionization mass spectrometry. *Anal Chem* **1987**, *59* (22), 2642-2646.

55. Cech, N. B.; Enke, C. G. Practical implications of some recent studies in electrospray ionization fundamentals. *Mass Spectrom Rev* **2001**, *20* (6), 362-387.

56. Juraschek, R.; Röllgen, F. W. Pulsation phenomena during electrospray ionization. *Int J Mass Spec* **1998**, *177*, 1-15.

57. Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S. Unraveling the mechanism of electrospray ionization. *Anal Chem* **2013**, *85* (1), 2-9.

58. 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), 466-469.

59. Mano, N.; Goto, J. Biomedical and biological mass spectrometry. *Anal Sci* **2003**, *19* (1), 3-14.

60. Wilm, M. Principles of electrospray ionization. *Mol Cell Proteomics* **2011**, *10* (7), M111.009407.

61. Pól, J.; Strohalm, M.; Havlíček, V.; Volný, M. Molecular mass spectrometry imaging in biomedical and life science research. *Histochem Cell Biol* **2010**, *134* (5), 423-443.

62. Mallick, P.; Kuster, B. Proteomics: a pragmatic perspective. *Nat Biotechnol* **2010**, *28* (7), 695-709.

63. March, R. E.; Todd, J. F. *Quadrupole Ion Trap Mass Spectrometry*; John Wiley & Sons, 2005.

64. Hardman, M.; Makarov, A. A. Interfacing the orbitrap mass analyzer to an electrospray ion source. *Anal Chem* **2003**, *75* (7), 1699-1705.

65. Kingdon, K. H. A Method for the Neutralization of Electron Space Charge by Positive Ionization at Very Low Gas Pressures. *Physical Rev* **1923**, *21* (4), 408-418.

66. Eliuk, S.; Makarov, A. Evolution of Orbitrap Mass Spectrometry Instrumentation. *Annu Rev Anal Chem (Palo Alto Calif)* **2015**, *8*, 61-80.

67. Yates, J. R., 3rd. Recent technical advances in proteomics. *F1000Res* **2019**, *8*.

68. Brodbelt, J. S. Ion Activation Methods for Peptides and Proteins. *Anal Chem* **2016**, *88* (1), 30-51.

69. Batista da Cunha, D.; Pupo Silvestrini, A. V.; Gomes da Silva, A. C.; Maria de Paula Estevam, D.; Pollettini, F. L.; de Oliveira Navarro, J.; Alves, A. A.; Remedio Zeni Beretta, A. L.; Annichino Bizzacchi, J. M.; Pereira, L. C.; et al. Mechanistic insights into functional characteristics of native crotamine. *Toxicon* **2018**, *146*, 1-12.

70. Teixeira, T. L.; Oliveira Silva, V. A.; da Cunha, D. B.; Polettini, F. L.; Thomaz, C. D.; Pianca, A. A.; Zambom, F. L.; da Silva Leitao Mazzi, D. P.; Reis, R. M.; Mazzi, M. V. Isolation, characterization and screening of the in vitro cytotoxic activity of a novel L-amino acid oxidase (LAAOcdt) from Crotalus durissus terrificus venom on human cancer cell lines. *Toxicon* **2016**, *119*, 203-217.

71. Muller, S. P.; Silva, V. A. O.; Silvestrini, A. V. P.; de Macedo, L. H.; Caetano, G. F.; Reis, R. M.; Mazzi, M. V. Crotoxin from Crotalus durissus terrificus venom: In vitro cytotoxic activity of a heterodimeric phospholipase A(2) on human cancer-derived cell lines. *Toxicon* **2018**, *156*, 13-22.

72. Jedrychowski, M. P.; Huttlin, E. L.; Haas, W.; Sowa, M. E.; Rad, R.; Gygi, S. P. Evaluation of HCD- and CID-type fragmentation within their respective detection platforms for murine phosphoproteomics. *Mol Cell Proteomics* **2011**, *10* (12), M111 009910.

73. Rose, R. J.; Damoc, E.; Denisov, E.; Makarov, A.; Heck, A. J. High-sensitivity Orbitrap mass analysis of intact macromolecular assemblies. *Nat Methods* **2012**, *9* (11), 1084-1086.

74. Scheltema, R. A.; Hauschild, J. P.; Lange, O.; Hornburg, D.; Denisov, E.; Damoc, E.; Kuehn, A.; Makarov, A.; Mann, M. The Q Exactive HF, a Benchtop mass spectrometer with a pre-filter, high-performance quadrupole and an ultra-high-field Orbitrap analyzer. *Mol Cell Proteomics* **2014**, *13* (12), 3698-3708.

75. Hauschild, J.-P.; Peterson, A. C.; Couzijn, E.; Denisov, E.; Chernyshev, D.; Hock, C.; Stewart, H.; Hartmer, R.; Grinfeld, D.; Thoeing, C.; et al. *Preprints* **2020**, 2020060111.

76. Kockmann, T.; Trachsel, C.; Panse, C.; Wahlander, A.; Selevsek, N.; Grossmann, J.; Wolski, W. E.; Schlapbach, R. Targeted proteomics coming of age - SRM, PRM and DIA performance evaluated from a core facility perspective. *Proteomics* **2016**, *16* (15-16), 2183-2192.

77. Li, J.; Smith, L. S.; Zhu, H. J. Data-independent acquisition (DIA): An emerging proteomics technology for analysis of drug-metabolizing enzymes and transporters. *Drug Discov Today Technol* **2021**, *39*, 49-56.

78. Masselon, C.; Anderson, G. A.; Harkewicz, R.; Bruce, J. E.; Pasa-Tolic, L.; Smith, R. D. Accurate mass multiplexed tandem mass spectrometry for high-throughput polypeptide identification from mixtures. *Anal Chem* **2000**, *72* (8), 1918-1924.

79. 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.

80. Bilbao, A.; Varesio, E.; Luban, J.; Strambio-De-Castillia, C.; Hopfgartner, G.; Muller, M.; Lisacek, F. Processing strategies and software solutions for data-independent acquisition in mass spectrometry. *Proteomics* **2015**, *15* (5-6), 964-980.

81. 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), 2343-2394.

82. Hebert, A. S.; Thoing, C.; Riley, N. M.; Kwiecien, N. W.; Shiskova, E.; Huguet, R.; Cardasis, H. L.; Kuehn, A.; Eliuk, S.; Zabrouskov, V.; et al. Improved Precursor Characterization for Data-Dependent Mass Spectrometry. *Anal Chem* **2018**, *90* (3), 2333-2340.

83. 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. *J Proteome Res* **2011**, *10* (4), 1785-1793.

84. Meier, F.; Geyer, P. E.; Virreira Winter, S.; Cox, J.; Mann, M. BoxCar acquisition method enables single-shot proteomics at a depth of 10,000 proteins in 100 minutes. *Nat Methods* **2018**, *15* (6), 440-448.

85. Wewer Albrechtsen, N. J.; Geyer, P. E.; Doll, S.; Treit, P. V.; Bojsen-Moller, K. N.; Martinussen, C.; Jorgensen, N. B.; Torekov, S. S.; Meier, F.; Niu, L.; et al. Plasma Proteome Profiling Reveals Dynamics of Inflammatory and Lipid Homeostasis Markers after Roux-En-Y Gastric Bypass Surgery. *Cell Syst* **2018**, *7* (6), 601-612 e603.

86. Doll, S.; Dressen, M.; Geyer, P. E.; Itzhak, D. N.; Braun, C.; Doppler, S. A.; Meier, F.; Deutsch, M. A.; Lahm, H.; Lange, R.; et al. Region and cell-type resolved quantitative proteomic map of the human heart. *Nat Commun* **2017**, *8* (1), 1469.

87. Kennedy, J.; Yi, E. C. Use of gas-phase fractionation to increase protein identifications : application to the peroxisome. *Methods Mol Biol* **2008**, *432*, 217-228.

88. Gerber, S. A.; Rush, J.; Stemman, O.; Kirschner, M. W.; Gygi, S. P. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc Natl Acad Sci U S A* **2003**, *100* (12), 6940-6945.

89. Piersma, S. R.; Fiedler, U.; Span, S.; Lingnau, A.; Pham, T. V.; Hoffmann, S.; Kubbutat, M. H.; Jiménez, C. R. Workflow comparison for label-free, quantitative secretome proteomics for cancer biomarker discovery: method evaluation, differential analysis, and verification in serum. *J Proteome Res* **2010**, *9* (4), 1913-1922.

90. Griffin, N. M.; Yu, J.; Long, F.; Oh, P.; Shore, S.; Li, Y.; Koziol, J. A.; Schnitzer, J. E. Label-free, normalized quantification of complex mass spectrometry data for proteomic analysis. *Nat Biotechnol* **2010**, *28* (1), 83-89.

91. Asara, J. M.; Christofk, H. R.; Freimark, L. M.; Cantley, L. C. A label-free quantification method by MS/MS TIC compared to SILAC and spectral counting in a proteomics screen. *Proteomics* **2008**, *8* (5), 994-999.

92. Chelius, D.; Bondarenko, P. V. Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *J Proteome Res* **2002**, *1* (4), 317-323.

93. Wiese, S.; Reidegeld, K. A.; Meyer, H. E.; Warscheid, B. Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. *Proteomics* **2007**, *7* (3), 340-350.

94. Thermo Fisher Scientific Inc. *Label Reagents—higher multiplex quantitation for up to 16 samples. Next generation of TMT labeling reagents for higher throughput and greater quantitative accuracy for multiplexed proteome analysis [Product Brochure]*. 2019. https://assets.thermofisher.com/TFS-Assets/BID/brochures/tmtpro-label-reagents-higher-multiplex-quantitation-brochure.pdf. (last consulted: Aug 30, 2022)

95. Sadygov, R. G.; Cociorva, D.; Yates, J. R., 3rd. Large-scale database searching using tandem mass spectra: looking up the answer in the back of the book. *Nat Methods* **2004**, *1* (3), 195-202.

96. 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. *J Am Soc Mass Spectrom* **1994**, *5* (11), 976-989.

97. Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999**, *20* (18), 3551-3567.

98. Craig, R.; Beavis, R. C. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **2004**, *20* (9), 1466-1467.

99. Mo, L.; Dutta, D.; Wan, Y.; Chen, T. MSNovo: a dynamic programming algorithm for de novo peptide sequencing via tandem mass spectrometry. *Anal Chem* **2007**, *79* (13), 4870-4878.

100. Frank, A.; Pevzner, P. PepNovo: de novo peptide sequencing via probabilistic network modeling. *Anal Chem* **2005**, *77* (4), 964-973.

101. Alves, G.; Yu, Y. K. Robust accurate identification of peptides (RAId): deciphering MS2 data using a structured library search with de novo based statistics. *Bioinformatics* **2005**, *21* (19), 3726-3732.

102. Ma, B.; Zhang, K.; Hendrie, C.; Liang, C.; Li, M.; Doherty-Kirby, A.; Lajoie, G. PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun Mass Spectrom* **2003**, *17* (20), 2337-2342.

103. Ma, B.; Zhang, K.; Liang, C. An effective algorithm for peptide de novo sequencing from MS/MS spectra. *J Comput Syst Sci* **2005**, *70* (3), 418–430.

104. Collins, D. C.; Lee, M. L. Developments in ion mobility spectrometry-mass spectrometry. *Anal Bioanal Chem* **2002**, *372* (1), 66-73.

105. Hill, H. H., Jr.; Siems, W. F.; St Louis, R. H.; McMinn, D. G. Ion mobility spectrometry. *Anal Chem* **1990**, *62* (23), 1201a-1209a.

106. Baumbach, J. I.; Eiceman, G. A. Ion mobility spectrometry: arriving on site and moving beyond a low profile. *Appl Spectrosc* **1999**, *53* (9), 338a-355a.

107. Ewing, R. G.; Atkinson, D. A.; Eiceman, G. A.; Ewing, G. J. A critical review of ion mobility spectrometry for the detection of explosives and explosive related compounds. *Talanta* **2001**, *54* (3), 515-529.

108. Thermo Fisher Scientific Inc. *FAIMS Pro and FAIMS Pro Duo User Guide, Operator's Manual, 98100-98031, Revision A*. 2021. https://assets.thermofisher.com/TFS-Assets/CMD/manuals/man-98100-98031-faims-pro-faims-pro-duo-user-guide-man98100-98031-en.pdf. (last consulted: Aug 30, 2022)

109. Kolakowski, B. M.; Mester, Z. Review of applications of high-field asymmetric waveform ion mobility spectrometry (FAIMS) and differential mobility spectrometry (DMS). *Analyst* **2007**, *132* (9), 842-864.

110. Bonneil, E.; Pfammatter, S.; Thibault, P. Enhancement of mass spectrometry performance for proteomic analyses using high-field asymmetric waveform ion mobility spectrometry (FAIMS). *J Mass Spectrom* **2015**, *50* (11), 1181-1195.

111. Pfammatter, S.; Bonneil, E.; McManus, F. P.; Prasad, S.; Bailey, D. J.; Belford, M.; Dunyach, J. J.; Thibault, P. A Novel Differential Ion Mobility Device Expands the Depth of Proteome Coverage and the Sensitivity of Multiplex Proteomic Measurements. *Mol Cell Proteomics* **2018**, *17* (10), 2051-2067.

112. Pfammatter, S.; Bonneil, E.; McManus, F. P.; Thibault, P. Accurate Quantitative Proteomic Analyses Using Metabolic Labeling and High Field Asymmetric Waveform Ion Mobility Spectrometry (FAIMS). *J Proteome Res* **2019**, *18* (5), 2129-2138.

Chapter 2

2 Integration of Segmented Ion Fractionation and Differential Ion Mobility on a Q-Exactive Hybrid Quadrupole Orbitrap Mass Spectrometer

Sibylle Pfammatter[1,2]; **Zhaoguan Wu[1,2](Co-first)**; Eric Bonneil[1]; Derek J. Bailey[3]; Satendra Prasad[3]; Michael Belford[3]; Jonathan Rochon[4]; Pierre Picard[4]; Jean Lacoursière[4]; Jean-Jacques Dunyach[3]; Pierre Thibault[1,2*]

[1]Institute for Research in Immunology and Cancer (IRIC); [2]Department of Chemistry, Université de Montréal, Montréal, QC, [3]ThermoFisher Scientific, San Jose, CA, [4]Phytronix Technologies, Québec, QC

Publication Date:July 2, 2021

Anal. Chem. 2021, 93, 28, 9817–9825

https://doi.org/10.1021/acs.analchem.1c01376

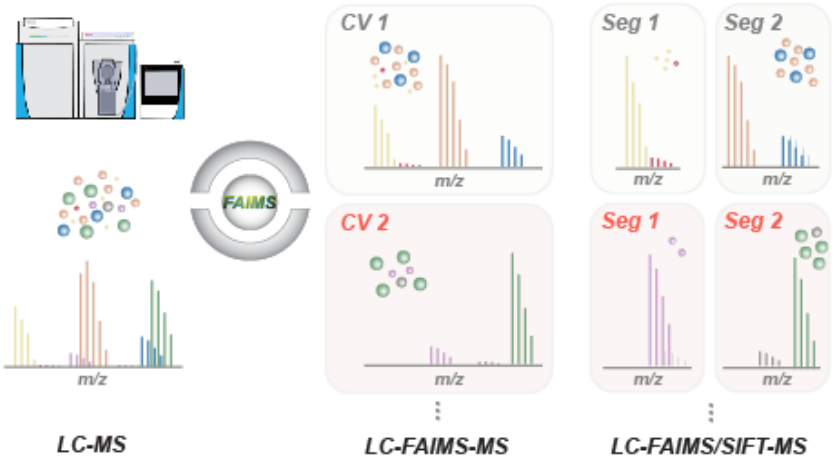
Author contributions:

Zhaoguan Wu and Sibylle Pfammatter (former PhD student) prepared and analyzed samples, interpreted results, prepared all figures, and participated in the editing of the manuscript. Eric Bonneil (MS platform manager) supervised the MS analysis, Pierre Thibault (supervisor) and Jean-Jacques Dunyach (Industrial partner) developed the concept and managed the project. Satendra Prasad, Derek J. Bailey, Michael Belford, and Jean-Jacques Dunyach build the new FAIMS device. Jonathan Rochon, Pierre Picard and Jean Lacoursiere designed a FAIMS support flange.

2.1 Abstract

High-field asymmetric waveform ion mobility spectrometry (FAIMS) has gained popularity in the proteomics field for its capability to improve mass spectrometry sensitivity and to decrease peptide co-fragmentation. The recent implementation of FAIMS on Tribrid Orbitrap instruments enhanced proteome coverage and increased the precision of quantitative measurements. However, the FAIMS interface has not been available on older generation Orbitrap mass spectrometers such as the Q-Exactive. Here, we report the integration of the FAIMS Pro™ device with embedded electrical and gas connections to a Q-Exactive HF mass spectrometer. Proteomic experiments performed on HeLa tryptic digests with the modified mass spectrometer improved signal to noise and reduced interfering ions resulting in an increase of 42% in peptide identification. FAIMS was also combined with segmented ion fractionation where 100 m/z windows were acquired in turn to further increase the depth of proteome analysis by reducing the proportion of chimeric MS/MS spectra from 50% to 27%. We also demonstrate the application of FAIMS to improve quantitative measurements when using isobaric peptide labeling. FAIMS experiments performed on a two-proteome model revealed that FAIMS Pro™ provided a 65 % improvement in quantification accuracy compared to conventional LC-MS/MS experiments.

Key words: High field asymmetric waveform ion mobility (FAIMS), mass spectrometry, gas-phase fractionation, tandem mass tags (TMT), quantitative proteomics.



2.2 Introduction

Recent technological advances in mass spectrometry (MS) in terms of acquisition speed and sensitivity have extended the breadth of proteomic analyses [1-3]. The capability to perform multiplex quantitative analyses using isobaric peptide labeling facilitated the large-scale comparison of protein abundance to an unprecedented depth [4]. Tens of thousands of peptides can be identified and quantified at the MS2 stage with ion intensities spanning over more than three orders of magnitude. However, several shortcomings including low quality spectra, wrong charge assignment and co-fragmentation of isobaric precursors limit the precision and accuracy of quantitative measurements using isobaric peptide labeling [5, 6]. Furthermore, limited space charge capacity in the ion trap results in signal suppression of low abundant species when they are trapped with more abundant precursor ions. Sample pre-fractionation using high pH reversed phase or strong cation exchange (SCX) chromatography can alleviate some of these issues while reducing peptide co-fragmentation. Unfortunately, those strategies require higher amounts of material and dilute peptides by splitting them into consecutive fractions.

Another approach to reduce sample complexity and increase peak capacity in proteomic analyses is the combination of ion mobility separation with MS [3, 7-11]. Optimized gas flow and increased ion transmission using high field asymmetric waveform ion mobility spectrometry (FAIMS) [12] led to the development of an improved device with enhanced sensitivity and precision in quantitative proteomic analyses [13]. In contrast to other ion mobility devices, FAIMS can perform gas phase ion fractionation by transmitting in turn different ion populations based on their changes in mobility at high and low electric fields [14, 15]. In the context of proteomic analyses, FAIMS can be advantageously exploited to reduce the occurrence of interfering ions, improve signal-to-noise and enhance the detectability of low abundance peptide ions. Several reports have already demonstrated the advantages of this new interface, termed FAIMS-Pro™, for single shot proteomics [16], isobaric peptide labeling [17, 18], and short LC-MS/MS analyses [19].

The current version of the FAIMS-Pro™ device is only available on recent generations of mass spectrometers such as the Tribrid Orbitrap and the Exploris 480. However, many laboratories are still using older generation instruments such as the hybrid Q-Exactive mass spectrometers for their capabilities to acquire high-resolution tandem mass spectra, albeit at lower MS/MS repetition

rates [20, 21]. The application of isobaric labeling for quantitative proteomics on these older instruments has been reported previously, but the frequent occurrence of peptide ion co-fragmentation lead to ratio compression of reporter ions and affect the accuracy of quantitative measurements [22, 23]. Clearly, gas phase ion fractionation using FAIMS could be advantageous for older generation mass spectrometers to improve peptide identification and quantitative proteomic analyses.

In the present study, we developed an interface that enabled direct mounting of the FAIMS-Pro™ device on a Q-Exactive HF and compared its performance to that of the unmodified mass spectrometer. LC-MS/MS experiments performed using tryptic digests of HeLa cells indicated that the combination of FAIMS increased peptide identification by more than 40%. In an effort to further improve MS sensitivity and proteome coverage, we integrated a data acquisition method, termed segmented ion fractionation (SIFT), in which multiple narrow m/z scans were analyzed at different compensation voltages (CVs) with FAIMS. We found that SIFT increased the mean injection time and improved peptide detection consistent with that observed previously for the Boxcar method [24]. Importantly, the combination of FAIMS and SIFT increased peak capacity while reducing the occurrence of co-selected peptide ions during MS/MS analysis. The analytical benefits of this combined approach in terms of number of quantifiable peptides and accuracy of quantitative measurements is presented for 10-plex isobaric peptide labeling using a two-proteome model.

2.3 Materials and Methods

2.3.1 Protein extraction and enzymatic digestion.

Direct infusion experiments were performed on a tryptic digest of bovine serum albumin (BioShop, ALB00150) diluted to 0.4 µg/µL in 50% methanol. For LC-MS benchmark tests, 500ng of Thermo Scientific™Pierce™HeLa protein digest standard (Thermo Fisher Scientific, 88329) diluted in 0.2% aqueous FA was used for each injection. For tandem mass tag (TMT) experiments, *Saccharomyces cerevisiae* and human embryonic kidney 293 (HEK293) tryptic digests were prepared in 200mM HEPES, pH 8.2 prior to labeling.

For LC-MS benchmark tests, we have injected 500ng of Thermo Scientific™ Pierce™ HeLa protein digest standard (Thermo Fisher Scientific, 88329) diluted in 0.2% aqueous FA. For tandem mass tag (TMT) experiments, Saccharomyces cerevisiae and human embryonic kidney 293 (HEK293) cells were twice washed with cold Phosphate Buffered Saline (PBS) (Fisher Scientific, BP399-1) and pelleted by centrifugation (1000 rpm, 5min). Lysis buffer containing 8M Urea (Fisher Scientific, BP169), 50mM HEPES (Bio Basic Inc., HB0264) and 75mM sodium chloride (Fisher Scientific, BP358), pH 8.2 was added to the cell pellets. Yeast cells were mechanically lysed with 10 min bead beating, HEK293 cells were mechanically lysed with 2x 10s sonication bursts. Lysates were centrifuged at 14,000 g for 10min, and protein concentration of clear lysate was determined by Bradford assay. Proteins were precipitated using methanol/chloroform (lysate/methanol/chloroform/water 1:4:1:3 volume parts). Samples were vortexed and centrifuged for 5min at 14,000 g. The precipitated protein layer was washed twice with four volume parts of methanol and air-dried. For tryptic digestion, proteins were dissolved in 200mM HEPES, pH 8.2 and reduced, alkylated and trypsinized (enzyme/protein ratio 1:50) overnight at 37°C.

2.3.2 Tandem mass tag (TMT) labeling.

Equal amounts of peptide aliquots were labeled with the individual isobaric TMT10plex tags (Thermo Fisher Scientific, 90110). Briefly, 100 µL of peptides (100 µg) dissolved in 200mM HEPES at pH 8.2 were combined with 40µl of TMT label (200 µg) in anhydrous acetonitrile for 90 minutes at room temperature. Reaction was quenched with 1 µL of 50% hydroxylamine. A pool

of human peptides was created by mixing 10:10:10:10:10:10:10:10:10:10 (v:v) for TMT 126 :127N :127C :128N :128C :129N :129C :130N :130C :131. Yeast peptides were pooled at a 0:1:1:2:2:5:5:10:10:0 volume ratio. For the final two proteome sample the stock solutions were mixed at 0.265 yeast/human ratio (µg/µg).

2.3.3 Direct infusion.

Bovine Serum Albumin (BioShop, ALB00150) was resuspended in 50 mM ammonium bicarbonate (Sigma Aldrich, A6141), reduced with 5mM TCEP (Thermo Fisher Scientific, 77720) and alkylated with 10mM 2-chloroacetamide (Sigma Aldrich, C0267) prior to overnight incubation with trypsin (Promega, V511A, enzyme/protein ratio 1:50) at 37°C. The protein digest was desalted on an Oasis HLB extraction cartridge (Waters, WAT094225). For direct infusion, peptides were diluted to 0.4 µg/µL in 50% methanol (Fisher Scientific, A452-4)/0.2% formic acid (FA) (Fisher Scientific, AC147930010).

2.3.4 LC-MS.

All analyses were performed on a Q Exactive™ HF BioPharma platform (Thermo Fisher Scientific). HeLa peptide separation was performed on an EASY-Spray column PepMap™ RSLC C_{18} (2µm, 100Å, 75µm x 50cm) at 50˚C connected to a nano EASY-Spray™ source (Thermo Fisher Scientific). Peptides were loaded on-column and eluted at 300 nL/min with a linear gradient of 5-24 % acetonitrile, ACN (0.2% FA) in 80 min, followed by 24-95 % ACN (0.2% FA) in 10 min. MS survey scans were acquired from *m/z* 350-1150 at a resolution of 60,000, an automatic gain control (AGC) of $3x10^6$ and a maximum injection time (max IT) of 50ms. For SIFT analyses, the *m/z* range was divided in 7 segments (6 x *m/z* 100 wide and one *m/z* 200 wide) each with 2 *m/z* overlap. ***Supplementary table 2-1*** lists the different mass ranges used for all MS experiments. Top 15 precursors (charge 2-5) were selected for MS/MS acquisition at 30,000 resolution, $2x10^4$ AGC and a maximal injection time of 200 ms for full range acquisition and 800 ms for segmented acquisition. The isolation window was set to 1.3 Th with 0.3 Th offset with 28% HCD normalized collision energy (NCE). A dynamic exclusion time of 20s and 10s was used for full scan and segmented MS/MS acquisition, respectively.

TMT experiments used similar LC-MS/MS set-up except that LC separation used a C_4 guard precolumn (Optimize Technologies, Oregon City, OR) and a custom 150 µm ID x 20 cm

analytical column (Jupiter C18, 3 µm, 300 Å, Phenomenex, Torrance, CA). The C_{18} column was connected to a nano Flex source (Thermo Fisher Scientific). TMT labeled peptides (500 ng/injection) were loaded on the precolumn and eluted at 600 nL/min with a linear gradient of 5-38 % ACN (0.2% FA) over 75 min, followed by 38-95 % ACN (0.2% FA) ramp in 10 min.

2.3.5 FAIMS.

A customized mounting bracket enabled the coupling of the FAIMS Pro™ interface (Tribrid Orbitrap series) on the Q-Exactive MS instrument (***Figure 2-1 a***). This interface is an experimental research system exclusively developed in the context of the present project, and is not commercially available. Detailed information is provided in ***Supplementary Figure 2-1*** on the production of the mounting flange and necessary parts to build this customized adapter. The FAIMS electrodes were spaced by a 1.5 mm gap and maintained at 100°C. Dispersion voltage (DV) was maintained at -4800V and the flow of the nitrogen carrier gas was set at 1.6 L/min. FAIMS was controlled by a custom graphical user interface (GUI) enabling the change of the CV setting automatically in a pre-acquisition parameter file prior to each injection. For LC-FAIMS-MS/MS performed with SIFT, each of the 7 m/z ranges were acquired at 6 CV values from -23V to -58V with a 7V step (***Supplementary Table 2-1***). Alternatively, these analyses were performed at fixed CV where several m/z segments were scanned during a single run.

2.3.5 Data Analysis and Visualization.

All raw files were searched with PEAKS engine (Bioinformatics Solutions, Inc., Version 10) against the Uniprot human and/or yeast database (June 08, 2018). Maximal tolerances for precursor and fragments were 10 ppm and 0.01 Da, respectively. Search parameters included trypsin with a maximum of three missed cleavages per peptide. A maximum of 3 variable modifications was allowed per peptide, and included oxidation (M), deamidation (NQ), carbamidomethylation (C) and phosphorylation (STY). For quantitative analyses, TMT was selected as a fixed modification and raw files were searched against a concatenated Uniprot human and yeast database. False discovery rate (FDR) was set to 1% for peptide spectrum matches. Data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD019848.

2.4 Result and Discussion

2.4.1 Benchmark experiments with FAIMS interfaced to the Q-Exactive HF

A custom mounting flange with embedded electrical and gas connections, a plug adapter, a transfer tube and mounting features was built to interface the FAIMS Pro^{TM} device to the Ion Max NG source of the Q-Exactive mass spectrometer (***Figure 2-1a***). All the source parameters including gas flow rate and electrospray voltage can be controlled through the Tune software (version 2.8 SP1 build 2806). No ion source or optic modifications are required. The only source modification is an extended (10.7 cm) ion transfer tube to guide the ions from the flange to the S-Lens. FAIMS parameters were controlled through a GUI interface enabling the tuning of the entrance plate voltage, carrier gas flow rate, dispersion voltage and compensation voltage. A pre-acquisition script was used with the GUI to change the CV values before each injection.

We performed a series of benchmark experiments with the direct infusion of a BSA tryptic digest **(*Supplementary Figure 2-2a*)**. The distribution of peptide ions according to charge state is similar to that observed previously for the FAIMS Pro^{TM} coupled to a tribrid Orbitrap Fusion mass spectrometer [13]. The majority of singly charged ions are transmitted above -20V and are separated from multiply-charges ions that are distributed from-20V to -80V. The transmission of selected peptide ions of different charge states is shown as an inset to ***Supplementary Figure 2-2b*** and the average peak width was determined to be 11.1±1.6 V (5.5±1.5 V, half height).

Next, we investigated the LC-MS/MS analysis of 500 ng of a HeLa tryptic digest with and without FAIMS to benchmark the performance of each system (***Figure 2-1b***). We first compared the ion transmission of 21'893 peptide ions common to both FAIMS and non FAIMS LC-MS/MS experiments, and determined that FAIMS provided an average ion transmission of 58% across 4 orders of magnitude in ion intensity (***Figure 2-1c***). The lower ion transmission observed here compared to the Tribrid Fusion instrument [13] is partly explained by the longer transfer tube capillary and the lower DV value (-4800 V) required to maintain voltage stability.

In spite of the lower transmission, we observed that FAIMS enabled the identification of 44'975 peptides (3891 proteins) compared to 31'267 peptides (2924 proteins) without FAIMS (***Figure 2-1d***, ***Supplementary Table 2-2***). The reduction of the peptide co-fragmentation and

interfering ions obtainable with FAIMS provided an increase of 44 % in the number of peptide identified, consistent with that observed previously using the Tribrid Orbitrap mass spectrometer[13]. As expected, we noted that the distribution of peptides identified with FAIMS varied across CVs and charge over the range from -23 to -58 V (***Supplementary Figure 2-3a***). Out of all identifications (54'349 peptides) approximately 40% were common while 43% and 17 % were unique to FAIMS and non FAIMS experiments, respectively (***Figure 2-1c***). The large majority of peptides common to FAIMS and non FAIMS analyses have been identified in two or more LC-MS/MS replicates (84%) whereas peptides unique to LC-MS/MS were mostly identified in one replicate (40%). A closer examination of peptides unique to LC-MS/MS revealed that they were of low abundance and not reliably detected across all runs (***Supplementary Figure 2-3b***), We also plotted the intensity distribution of peptides obtained with both strategies, and found that FAIMS typically provided an additional order of magnitude in peptide identification compared to non-FAIMS (***Figure 2-1c***). The enhanced sensitivity of FAIMS enabled an improvement in signal to noise detection that facilitated the identification of low abundance peptides.

We also compared the precursor ion fraction (PIF) between experiments to determine the proportion of ion co-selection during MS/MS, where a PIF value of 1 indicates that no ion other than the selected target ion is selected within the isolation window (***Figure 2-1d***). We observed that FAIMS significantly reduced the occurrence of precursor co-selection (average PIF value: 0.92) compared to conventional LC-MS/MS experiments (average PIF value: 0.80). Altogether, these results indicated that the integration of the FAIMS Pro^{TM} device on the Q-Exactive HF enabled the isolation of precursor ions with lower occurrence of interfering ions thus improving the identification of low abundance peptide ions.

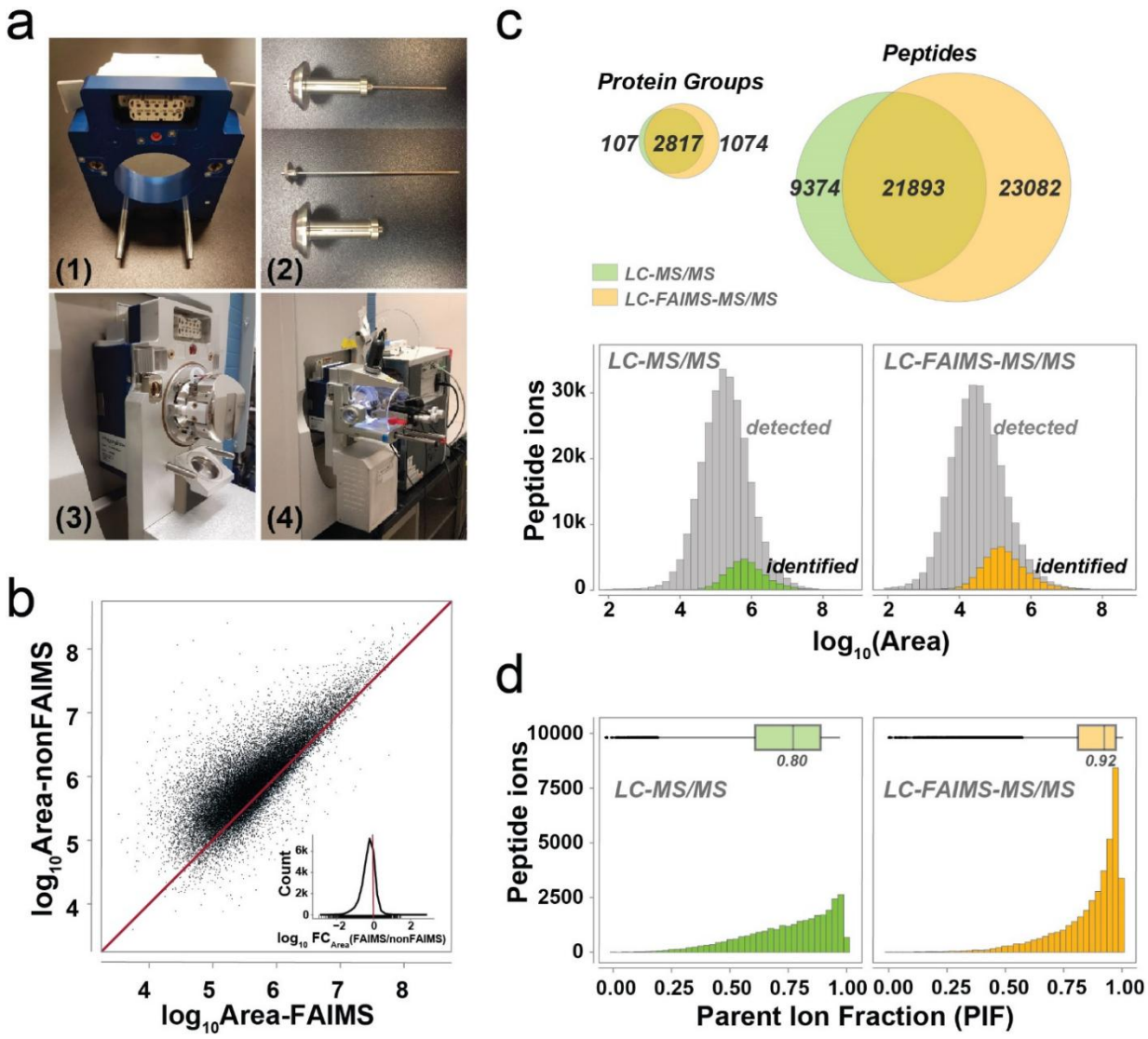


Figure 2-1: FAIMS Pro™ device interfaced to a Q-Exactive HF mass spectrometer.

(a) Photograph of the mounting flange (1), extended transfer tube (2), FAIMS Pro™ device installed on the mounting flange (3), and combined with nanoFlex electrospray source (4). (b) Distribution of log10 fold change (FC) and scatter plot of peptide intensities with and without FAIMS. (c) Overlap of identified peptides and proteins with and without FAIMS. (d) Distribution of precursor ion fraction (PIF) for identified peptides with and without FAIMS.

2.4.2 Separation through segmented ion fractionation (SIFT)

The depth of proteomic analyses is limited to the most abundant peptides leaving low abundance peptides largely unidentified. This limitation is associated with the ion capacity of the

C-trap ($1x10^6$ ions) and the duty cycle of the MS instrument. During the MS acquisition, the C-trap is filled with all eluting ions including peptide and interfering ions. Precursor ions are typically selected for MS/MS based on their charge state and intensity, thus leaving peptides ions of low abundance underrepresented. To enhance the detection and selection of low abundance peptide ions, we compared LC-MS/MS experiments performed under full- and segmented-scan acquisitions, the latter consisting of breaking down the scanning range of *m/z* 350 – 1150 into six windows of 100 Th except for the last segment of 200 Th. The narrow m/z scan enabled the longer accumulation of low peptide ion signals thereby providing better isotopic peak definition and charge state assignment. However, we noted that the extent of background ions contributing to the baseline is still significant and prevented the detection of low abundance peptide ions. To reduce the contribution of interfering ions, we combined FAIMS with segmented MS acquisition windows, a method that we referred to as segmented ion fractionation (SIFT). In SIFT, each narrow m/z scan is analyzed at different compensation voltages (CVs) with FAIMS, thus enabling peptide and interfering ions of different mobilities at high and low electric fields to be separated from each other. An experimental outline of the full scan LC-MS/MS analyses and segmented MS scans performed with or without FAIMS are depicted in ***Figure 2-2a***. We selected CV steps of 7V increments based on our benchmark experiments with the infusion of the BSA tryptic digest. Each m/z segment was analyzed at 6 different fixed CV values spanning across the CV transmission domain. This represents a total of 42 injections (7 m/z ranges and 6 CV values).

To evaluate the analytical benefits of a segmented mass range on peptide identification, we performed LC-MS/MS analyses with and without FAIMS and SIFT for replicate analyses of 500 ng of HeLa tryptic digest (***Supplementary Table 2-3***). The same number of injections were performed for FAIMS and non-FAIMS experiments (ie 6 fixed CV analyses for each 7 m/z segments vs 6 replicates for each 7 m/z segments for non-FAIMS). Significant gains in the number of detected multiply-charged features were noted when FAIMS, SIFT or the combination of the two is used (***Figure 2-2b***). Indeed, FAIMS or SIFT provides 2-2.5-fold increase in the number of peptide features, whereas the combination of FAIMS-SIFT enabled up to 3-fold increase in peptide features (e.g. 500828 and 145377 features for LC-FAIMS-SIFT-MS/MS and LC-MS/MS, respectively). Similarly, the use of SIFT in LC-MS/MS experiments provided an 87% increase in peptide identification (58557 unique peptides) compared to LC-MS/MS (31267 unique peptides) (***Figure2- 2b***). The combination of FAIMS and SIFT further enhanced the number of identification

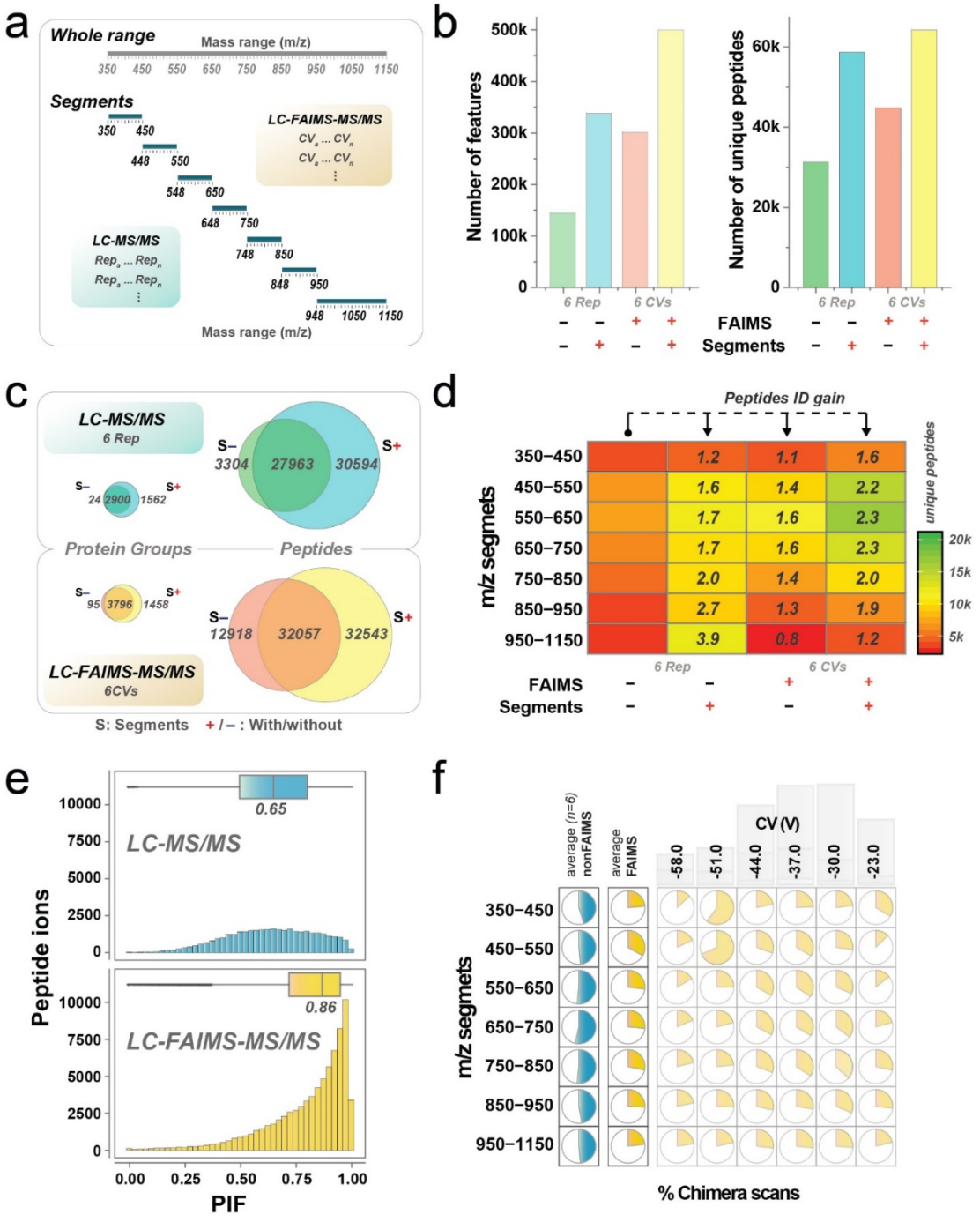


Figure 2-2: Improvement of proteome coverage using SIFT.

(a) Experimental scheme used for segmented ion fractionation (SIFT). (b) Number of unique features and peptides for segmented and non-segmented LC-MS/MS with and without FAIMS. (c) Venn diagram showing the peptide and protein overlap for segmented and non-segmented analyses with and without FAIMS. (d) Heatmap showing the number of identified peptides for the 4 experiments. Numbers represent peptide identification fold increase. (e) PIF distribution for the SIFT analysis with and without FAIMS (f) % of chimera scans for SIFT analyses with and without FAIMS.

to 64600 unique peptides. Next, we compared the overlap in peptide and protein identification for LC-MS/MS experiments performed with and without FAIMS and SIFT (***Figure 2-2c***). We obtained a good overlap of identified proteins groups and peptides where typically less than 10% identification were observed with SIFT for LC-MS/MS experiments performed with or without FAIMS. We also evaluated the gain in number of unique peptide identified for individual m/z segments, and noted that the most significant gains were observed for densely populated segments between *m/z* 450-750 (***Figure 2-2d***). In this m/z region, the combination of FAIMS-SIFT more than doubled the number of identified peptides compared to LC-MS/MS experiments. We noted again that FAIMS reduced the number of interfering ions in LC-MS/MS experiments performed using SIFT where we observed a PIF value of 0.85 compared to 0.65 for experiments with and without FAIMS, respectively (***Figure 2-2e***). Background and contaminating ions are more prevalent when sequencing lower abundance ions with SIFT. However, the median of the PIF values (0.86) observed with FAIMS-SIFT is comparable to that obtained with FAIMS only (median PIF 0.92). Without FAIMS, approximately half of the identified PSMs were chimeric consistent with previous reports[13] (***Figure 2-2f***). The lower occurrence of precursor co-isolation provided with FAIMS reduces by half the number of chimeric scans (27% of the PSMs were chimeric) for all CVs and m/z tiles.

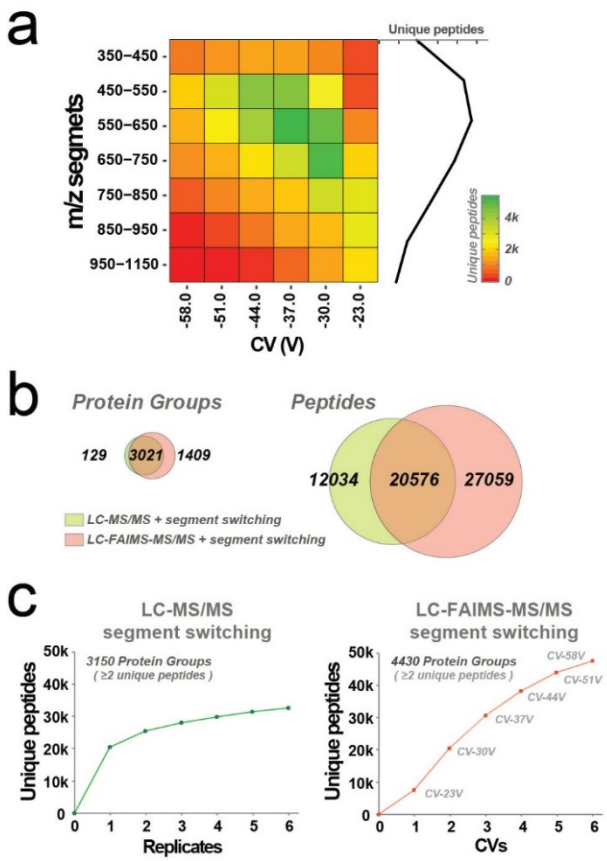


Figure 2-3: Impact of m/z segmentation on peptide identification.

(a) Heatmap displaying the variation of peptide density based on m/z segment and CV (left panel). Number of unique peptide across CVs for each m/z segment (right panel). (b) Comparison of identified peptides and proteins for LC-MS/MS SIFT experiments performed with and without FAIMS. (c) Cumulative increase of unique peptides across replicates or CV values.

The very large number of injections required to complete the LC-FAIMS-SIFT-MS/MS experiments is not time efficient for large-scale proteomic studies. While the current software does not enable efficient programming of CV stepping during the LC-MS/MS experiment, we stepped across *m/z* windows instead of CVs to decrease the number of injections. We used the previous dataset to display peptide density for each CV and m/z segment (***Figure 2-3a***). A heatmap distribution of detected features and peptide identification across m/z, CVs and retention time is presented in ***Supplementary Figures 2-4a and 2-4b***. No clear relationship between m/z, Rt and CV was found, and consequently we made our m/z segment combination based on peptide distribution per m/z segment and CVs (***Supplementary Figure 2-4c***). Based on peptide distribution for LC-MS/MS analyses shown in ***Figure 2-2d***, we selected to step across one populated segment

(m/z 448-450) and two less populated segments (m/z 350-450 and m/z 948-1150) within the same LC-MS/MS run. Similarly, the second analysis cycles through the populated m/z 548-650 and the less populated m/z 848-950 tiles. Finally, the third analysis step through the m/z 648-750 and m/z 748-850. We performed 6 replicates of this experiment giving a total of 18 LC-MS/MS runs. For FAIMS, we took advantage of the distribution of peptide population based on CV and m/z tile to group LC-MS/MS runs in a similar fashion to non-FAIMS experiments. For instance, two LC-FAIMS-SIFT-MS/MS were performed at CV -23V: one for tiles m/z 350-450, 548-650, 648-750, 950-1150 whereas the second injection groups m/z 448-550, 748-850 and 848-950 MS segments. This enabled us to cycle through m/z tiles and CVs in 13 injections. The comparison of m/z segment stepping with and without FAIMS is presented in ***Figure 2-3b*** and the corresponding identifications are listed in ***Supplementary Table 2-4***. In total, we identified 32610 peptides (3150 proteins) without FAIMS compared to 47635 peptides (4430 proteins) with FAIMS, and 20576 peptides (3021 proteins) were identified with both approaches. Once again, FAIMS enhanced the number of identified peptides and proteins with an additional 27059 (45%) peptides and 1409 (31%) proteins. However, we noted that the number of peptides identified without FAIMS decreased by 44 % from 58557 (4462 proteins) to 32610 peptides (3150 proteins) when comparing identification obtained with combined MS segments to those acquired separately (***Figure 2-2c***). While the identification rate remained the same with or without segments, the additional time required to acquire MS scan for each segment reduced the overall number of acquired MS/MS spectra. It is noteworthy that without FAIMS, the combination of MS segments within the same LC-MS/MS run reached a plateau after the third replicate with approximately 30000 unique peptides identified (***Figure 2-3c***). In contrast, we observed a regular increase in the number of identified peptides for each individual CV extending to more than 47000 peptides (***Figure 2-3c***). This highlights the different peptide populations transmitted at different CVs even when different m/z segments are selected. With FAIMS, the combination of MS segments within the same run also reduced the total number of identification albeit to a lower extent than that observed for non-FAIMS experiments. Indeed, we observed that the combination of MS segments resulted in a 26% reduction of identification from 64600 to 47635 peptides. Although a lower number of identification is observed when using combined MS segments, this approach enabled a comprehensive proteome coverage with 13 injections instead of 42.

2.4.3 TMT quantification

To determine if the combination of SIFT and FAIMS can reduce the extent of precursor co-isolation and improve quantitative measurements, we conducted LC-MS/MS experiments using isobaric peptide labeling with a 10-plex tandem mass tag (TMT). Previous studies have reported that FAIMS can reduce the occurrence of chimeric MS/MS spectra to alleviate TMT ratio compression in multiplex quantitative proteomic experiments [13, 18]. Accordingly, we devised a series of experiments based on a two proteome model [25]where the ratio of yeast to human tryptic digests varied from 0 to 10 (***Figure 2-4a***). Yeast tryptic peptides were not labeled with TMT-126 and TMT 131 to evaluate the extent of interferences from the human proteome. We then compared the number of quantifiable peptides from LC-MS/MS analyses performed with or without SIFT and FAIMS (***Supplementary Table 2-5***). In conventional LC-MS/MS experiments, SIFT extended the quantitative measurements to 4568 peptides (789 proteins) compared to 2452 peptides and 527 proteins without MS segments (***Figure 2-4b***). However, a more sizable gain in the number of quantifiable peptides is obtained using LC-FAIMS-MS/MS with 16489 peptides (2278 proteins) compared to 13710 peptides (1912 proteins) without SIFT. The combination of FAIMS and SIFT provided the maximum coverage of quantifiable peptides compared to other LC-MS/MS configurations. Furthermore, we noted that FAIMS and SIFT also improved the quality of MS/MS spectra as reflected by increased PEAKS identification scores across almost 4-orders of magnitude in precursor ion intensity (***Figure 2-4c***). The enhanced identification score observed with FAIMS and SIFT arise from reduced co-selection of precursor ions and higher PIF values as noted previously (***Figure 2-2e***).

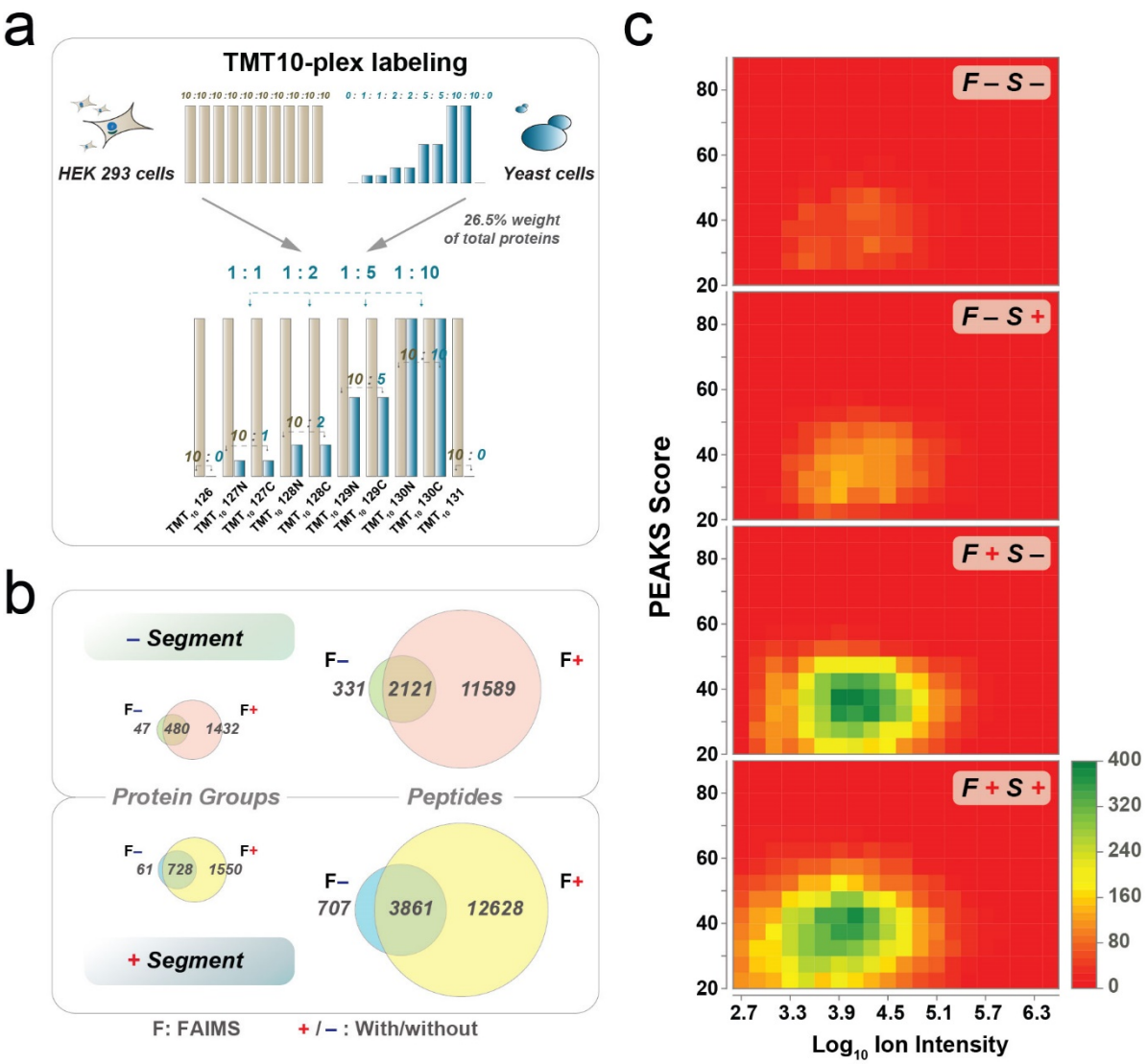


Figure 2-4: Quantification of human and yeast extracts using FAIMS and SIFT.

(a) Schematic of the TMT labeling scheme for human and yeast proteome mix. (b) Distribution of identified peptides for human and yeast proteins for the four different LC-MS/MS configurations. (c) Heatmaps displaying peptide intensity vs PEAKS identification score.

Next, we evaluated the accuracy of quantitative measurements by comparing the theoretical and observed fold change (FC) values of TMT reporter ions for experiments performed with and without SIFT and FAIMS (***Supplementary Figure 2-5a***). As expected, a progressive TMT ratio compression was observed with increasing dilution of the yeast sample as reflected by a larger deviation between the ratios of theoretical and observed FC values. While SIFT increased the number of quantifiable peptides, we observed that selecting narrow MS segments did not improve

the accuracy of quantitative measurements, and comparable log_2FC ratios were observed for both FAIMS and non-FAIMS experiments. Closer examination of these results revealed that MS segments increased the quantification of lower intensity peptides ions though the decreased occurrence of peptide ion co-isolation. However, FAIMS significantly improved the accuracy of quantitative measurements with a lower extent of TMT ratio compression. For instance, we measured a FC value of 9.1 ($log_2FC_{obs/theo}$= - 0.14) for a theoretical ratio of 10 when using FAIMS, compared to a FC value of 5.5 ($log_2FC_{obs/theo}$= - 0.85) without FAIMS (***Supplementary Figure 2-5a***), thus representing a 65% improvement in accuracy of quantitative measurement. We calculated the interference free index (IFI) for yeast peptides identified with and without FAIMS and SIFT [26]. An IFI of 1 indicates that no reporter ion intensity is detected in the yeast peptide empty channels (ie TMT-126 and TMT 131). The scatterplots of IFI values for all experiments performed with and without FAIMS are shown in ***Supplementary Figure 2-5b***, and indicated that a higher IFI value is obtained with FAIMS, further confirming the reduced contribution of precursor ion co-selection. Scatter plots of IFI and fold changes obtained with and without SIFT and FAIMS are displayed in **Supplementary *Figure 2-6***. These plots clearly indicate that SIFT does not improve IFI (***Supplementary Figure 2-6a***) or accuracy of quantitative measurements (***Supplementary Figure 2-6b***), whereas observed fold changes are very close to the expected values when using FAIMS.

To improve the accuracy of quantitative measurements, we correlated IFI values with FC observed on TMT ratios for all four instrument configurations. We plotted FC distribution within IFI bins and noted that larger deviation of FC ratios was observed with decreasing IFI values (***Supplementary Figure 2-7***). This is especially true for higher fold changes where the intensity of the low abundance TMT channel approached the background level resulting in higher contribution of interfering ions. The application of a specific IFI to filter out distorted FC ratios could thus improve the accuracy of quantitative measurements although this would reduce the quantification coverage. Accordingly, we evaluated the impact of IFI values on the number of quantifiable peptides and proteins (***Figure 2-5a***). When the IFI is progressively increased from 0 to 0.95, we observed that the number of quantifiable peptides decreased from 5666 to 3884 (750 to 674 proteins) and from 1821 to 1178 (321 to 248 proteins) for SIFT experiments performed with and without FAIMS, respectively. Based on these results, we selected all quantifiable peptides with IFI values ≥ 0.95 and plotted the distribution of FC ratios (***Figure 2-5b***). By filtering FC ratios

based on a IFI value of 0.95, we significantly improved the accuracy of quantitative measurements for all four configurations. For example, the application of an IFI of 0.95, provided an average FC ratios of 0.95, 1.45, 3.77, and 7.65 (theoretical ratios of 1, 2, 5, and 10) for experiments performed without SIFT and FAIMS compared to 1.0, 1.33, 2.75 and 5.31 without IFI filter. Further improvement of accuracy was achieved using SIFT and FAIMS where we obtained FC ratios of 0.96, 1.56, 3.98, and 8.17 for an IFI value of 0.95. Additionally, interquartile ranges and empty channel contamination is smaller for peptides and proteins with IFI>0.95 for all configurations. Altogether, these analyses indicate that improvement in the comprehensiveness and accuracy of quantitative proteomics can be achieved on the Q-Exactive HF using SIFT and FAIMS. While SIFT provides marginal gains in the number of quantifiable peptides, we anticipate a faster duty cycle and the use of phase constrained spectrum deconvolution [5] could further enhance the depth of quantitative proteomics.

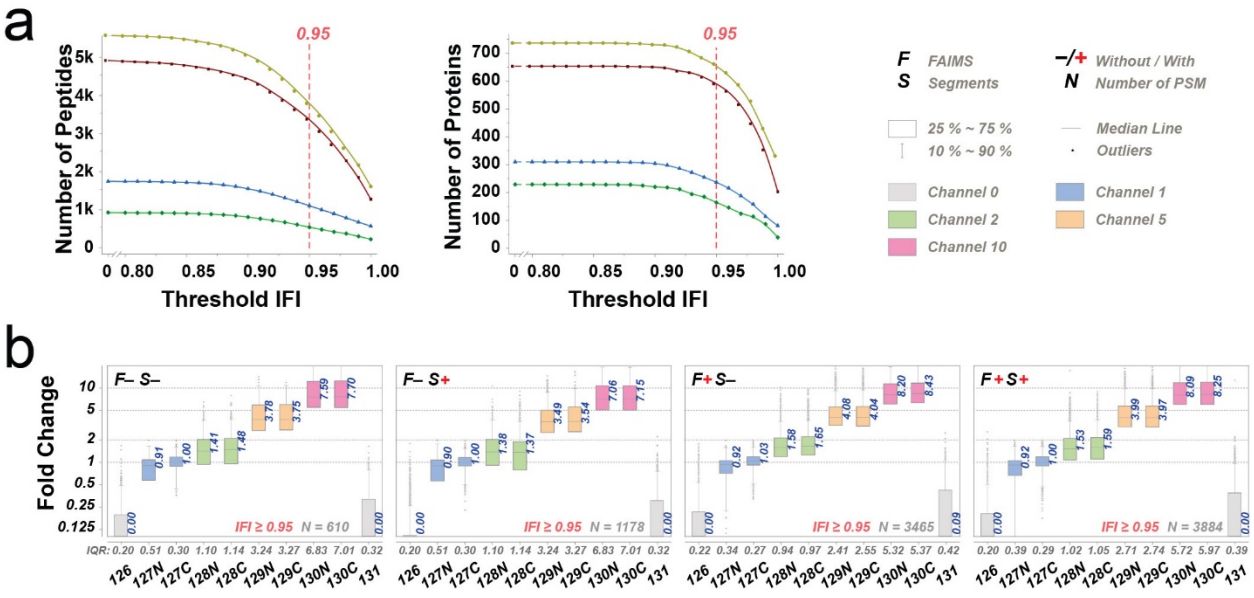


Figure 2-5: Impact of IFI values on the number of quantifiable peptides and accuracy of quantitative measurements.

(a) Number of peptide and proteins quantified for different IFI filters. (b) Boxplot distribution of fold changes for all quantified yeast peptides with an IFI>0.95 against the average of TMT reporter intensities for channel 127C and 127N.

2.5 Conclusion

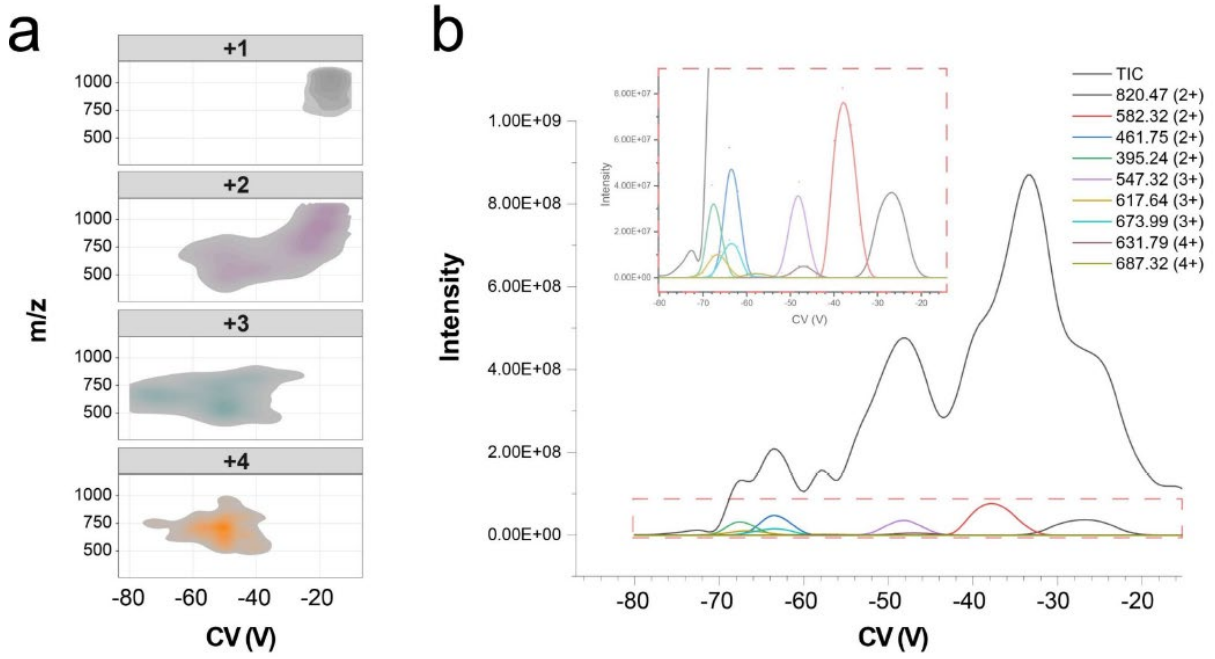
We report on a customized (non-commercial) interface enabling the coupling of the FAIMS Pro™ Interface to older generation Q-Exactive Orbitrap series with a minor modification on the mass spectrometer ion source (i.e a longer transfer tube). FAIMS parameters are controlled by a simple user interface that allow the selection of CV values during LC-MS/MS experiments and can be programmed to change CV values between runs. FAIMS on the Q-Exactive HF provides charge state separation comparable to that observed on the latest generation of Orbitrap instruments (e.g Eclipse and Exploris instruments). Reduction of background ions enhances peptide identification by 42% with close to 60% ion transmission. FAIMS is especially beneficial when combined with SIFT where a higher proportion of low abundance ions are sequenced by MS/MS. FAIMS decreases the extent of chimeric tandem mass spectra by two-fold providing a 50% increase in the number of identified proteins. FAIMS is also beneficial for TMT quantification, and alleviates distortion of TMT reporter ion ratios commonly found with quantitative proteomics analyses performed using isobaric peptide labeling. While SIFT typically extend the number of quantifiable peptides, the reduction of TMT ratio compression available with FAIMS enable a 65% improvement in accuracy of quantitative measurements. Further improvement in accuracy can be achieved by filtering FC measurements using IFI values ≥ 0.95.

While FAIMS enhanced both the depth and accuracy of quantitative measurements on the Q-Exactive HF, we expect that the ability to use DV voltage higher than -4800V would improve ion transmission and sensitivity of this system. Furthermore, additional flexibility in the software to facilitate the selection of different CVs within the same LC-MS/MS run would expand the breath of ion population for optimal sample use. Integration of the FAIMS voltage control to the Tune file would facilitate the design of acquisition methods where CV switching can be integrated in the method duty cycle.

2.6 Acknowledgements

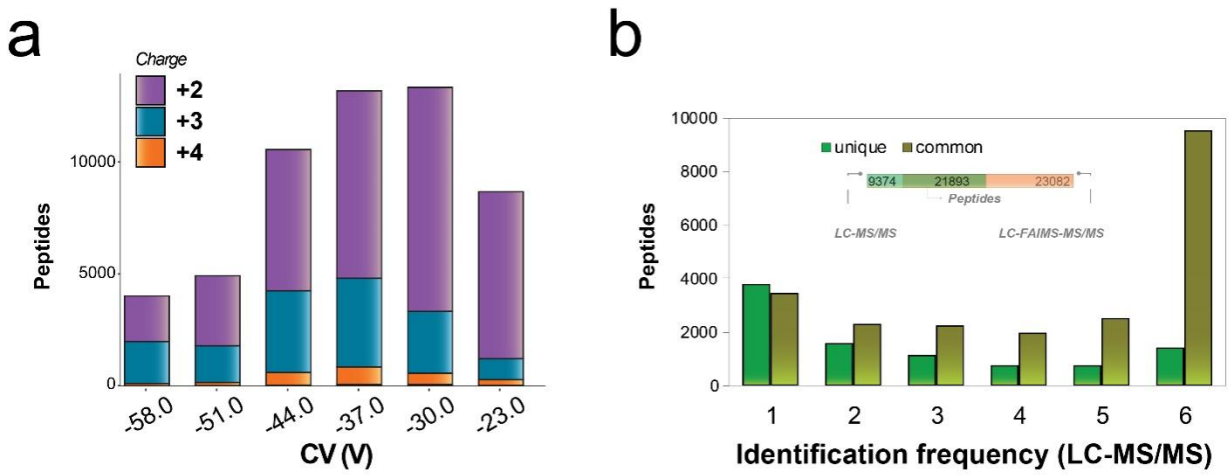
This work was carried out with financial support from the Natural Sciences and Engineering Research Council (NSERC 311598) and the Genomic Applications Partnership Program (GAPP) of Genome Canada. The Institute for Research in Immunology and Cancer (IRIC) receives infrastructure support from IRICoR, the Canadian Foundation for Innovation, and the Fonds de Recherche du Québec - Santé (FRQS). IRIC proteomics facility is a Genomics Technology platform funded in part by Genome Canada and Genome Québec.

2.7 Supplementary Infomation



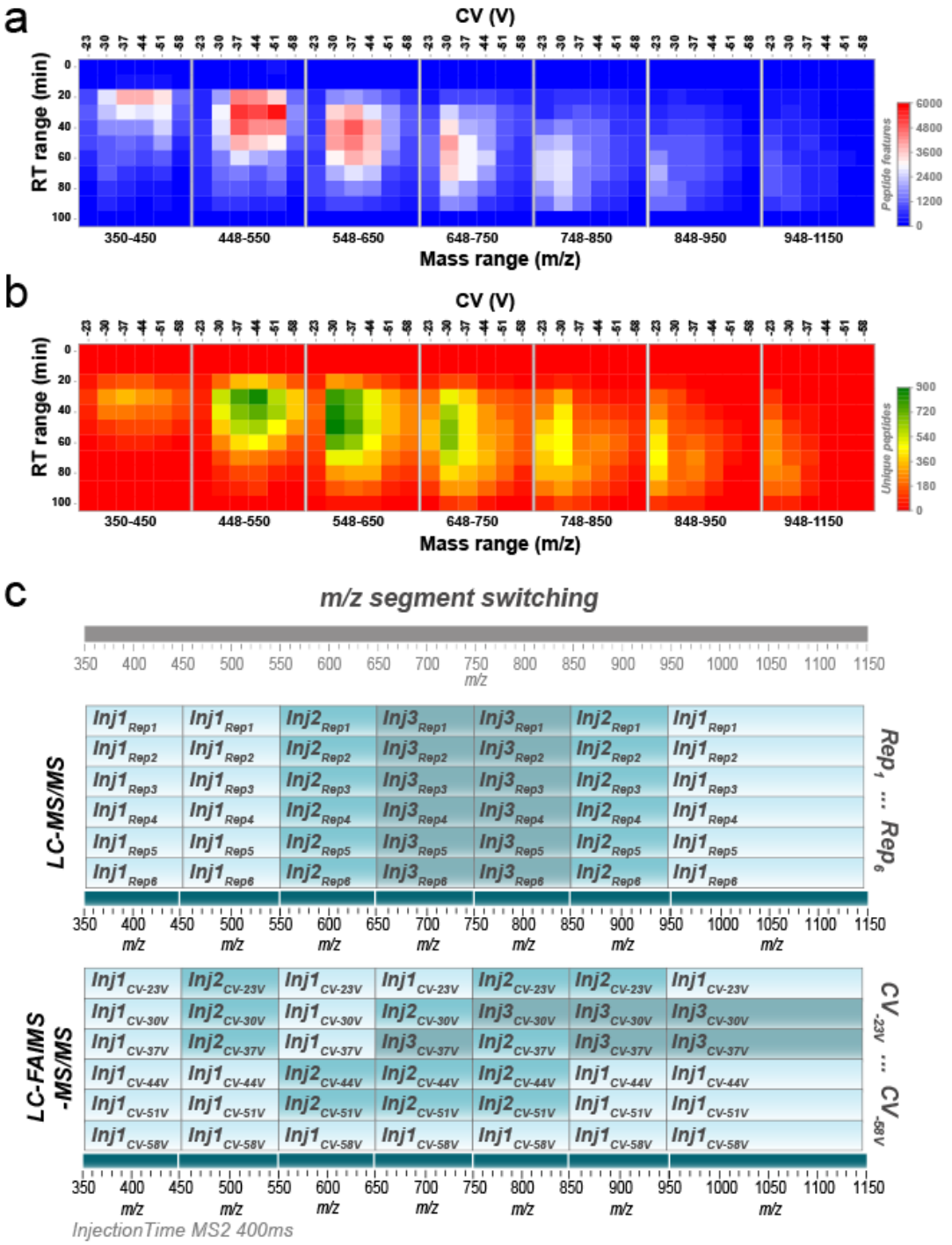
Supplementary Figure 2-1: Distribution of BDA tryptic peptides according the CV.

Direct infusion of BSA tryptic digest by ESI-FAIMS-MS by switching the CV in 2-V steps from -10 V up to -80 V at DV -4800. (a) Heat map showing transmission domain for each charge state. (b) Total ion chromatogram (black) with representative extracted ion chromatograms of selected BSA tryptic peptides.



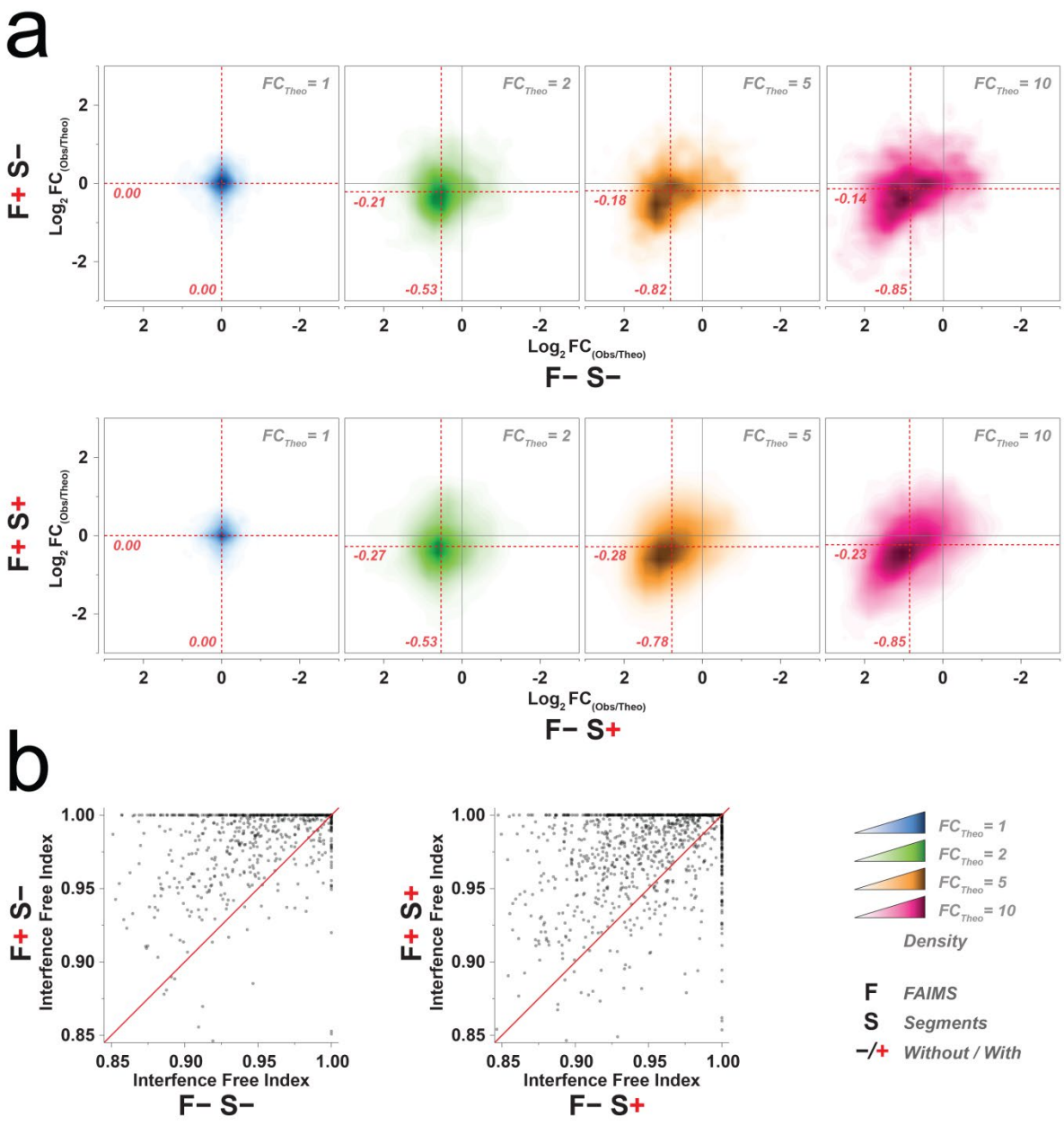
Supplementary Figure 2-2: Distribution and frequency of tryptic peptides from HeLa cells identified in LC-MS/MS experiments with and without FAIMS.

(a) Charge state distribution of HeLa peptides identified with FAIMS. (b) Identification frequency of peptides identified by LC-MS/MS only or identified by LC-MS/MS and LC-FAIMS-MS/MS,



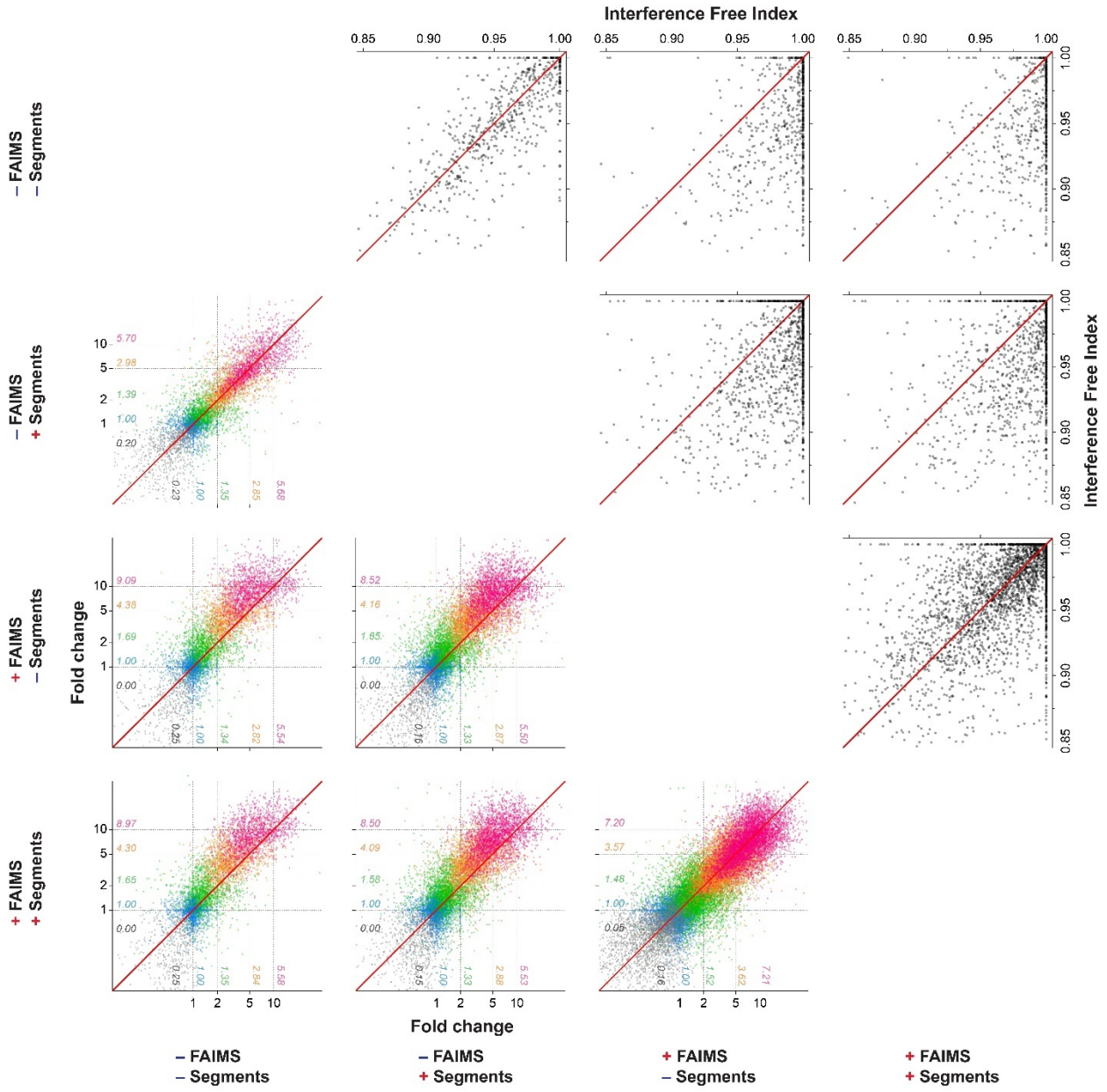
Supplementary Figure 2-3: Identification of HeLa tryptic peptides by LC-MS-MS using SIFT and FAIMS.

Heatmap of detected features (a) and unique peptides (b) according to CV, m/z tiles and retention time bins. (b) Schematic representation of the m/z cycling per replicate injections with and without FAIMS.

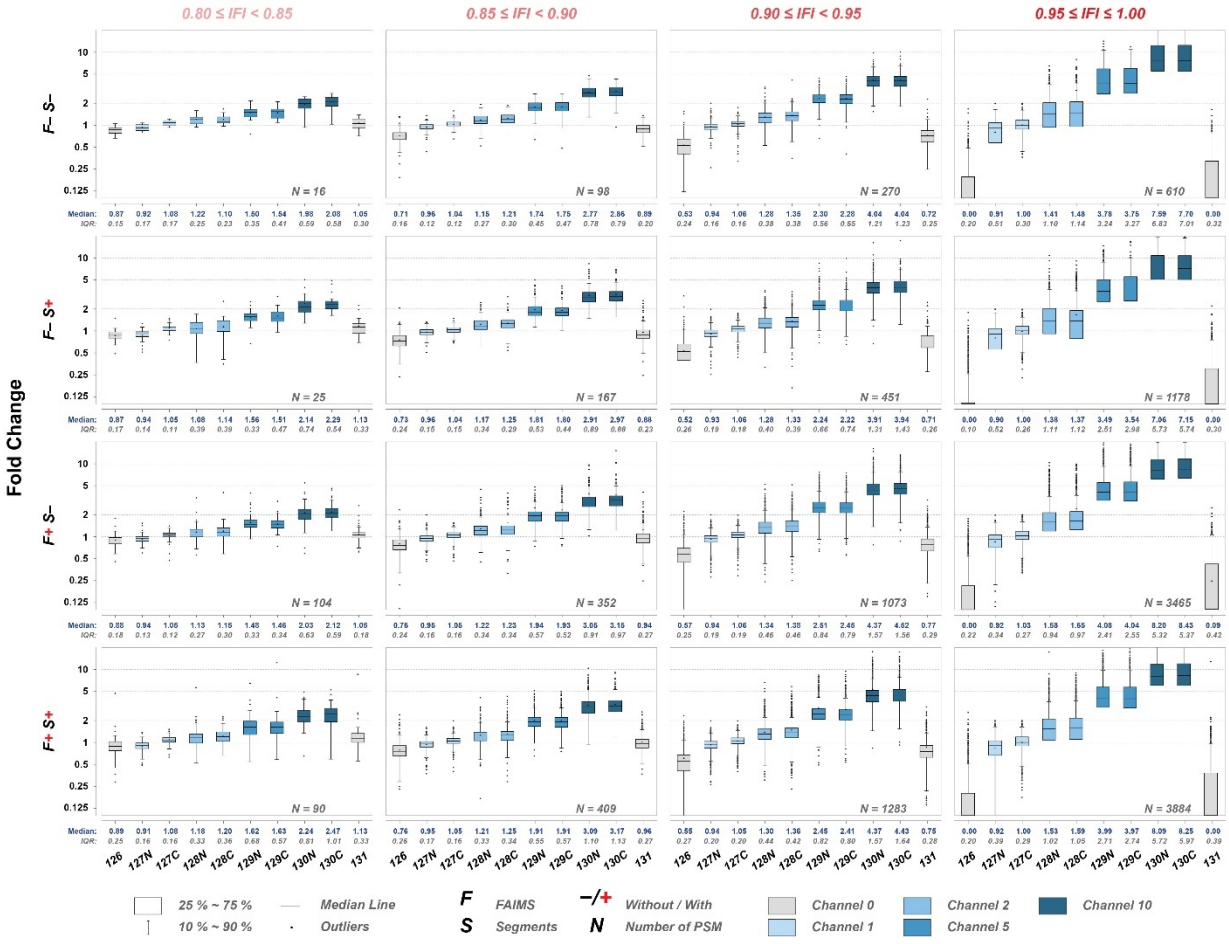


Supplementary Figure 2-4: Impact of interference on quantitative measurements.

(a) Deviation of observed from expected TMT ratios with and without FAIMS. (b) Scatter plot of interference free index for common peptides identified with and without SIFT and FAIMS.



Supplementary Figure 2-5: Scatter plots of interference free index and fold changes obtained with and without FAIMS and SIFT.



Supplementary Figure 2-6: Boxplot distribution of fold changes for peptides grouped by IFI.

Numbers (N) of peptides per groups are indicated.

Supplementary Table 2-1: Scanning ranges used for full MS scan and SIFT with and without FAIMS

FAIMS − SIFT −

Experiment	Injection	CV (V)	m/z range
Rep 1	1	—	350-1150
Rep 2	2	—	350-1150
Rep 3	3	—	350-1150
Rep 4	4	—	350-1150
Rep 5	5	—	350-1150
Rep 6	6	—	350-1150

FAIMS + SIFT −

Experiment	Injection	CV (V)	m/z range
CV1	1	-23	350-1150
CV2	2	-30	350-1150
CV3	3	-37	350-1150
CV4	4	-44	350-1150
CV5	5	-51	350-1150
CV6	6	-58	350-1150

FAIMS − SIFT +

Experiment	Injection	CV (V)	m/z range
Rep 1	1	—	350-450, 448-550, 948-1150
	2	—	548-650, 848-950
	3	—	648-750, 748-850
Rep 2	4	—	350-450, 448-550, 948-1150
	5	—	548-650, 848-950
	6	—	648-750, 748-850
Rep 3	7	—	350-450, 448-550, 948-1150
	8	—	548-650, 848-950
	9	—	648-750, 748-850
Rep 4	10	—	350-450, 448-550, 948-1150
	11	—	548-650, 848-950
	12	—	648-750, 748-850
Rep 5	13	—	350-450, 448-550, 948-1150
	14	—	548-650, 848-950
	15	—	648-750, 748-850
Rep 6	16	—	350-450, 448-550, 948-1150
	17	—	548-650, 848-950
	18	—	648-750, 748-850

FAIMS + SIFT +

Experiment	Injection	CV (V)	m/z range
CV1 Seg1	1	-23	350-450, 548-650, 648-750, 948-1150
CV1 Seg2	2	-23	448-550, 748-850, 848-950
CV2 Seg1	3	-30	350-450, 548-650, 748-850
CV2 Seg2	4	-30	448-550, 648-750
CV2 Seg3	5	-30	848-950, 948-1150
CV3 Seg1	6	-37	350-450, 548-650
CV3 Seg2	7	-37	448-550, 748-850
CV3 Seg3	8	-37	648-750, 848-950, 948-1150
CV4 Seg1	9	-44	350-450, 448-550, 848-950, 948-1150
CV4 Seg2	10	-44	548-650, 648-750, 748-850
CV5 Seg1	11	-51	350-450, 448-550, 548-650
CV5 Seg2	12	-51	648-750, 748-850, 848-950, 948-1150
CV6 Seg1	13	-58	350-450, 448-550, 548-650
CV6 Seg2	14	-58	648-750, 748-850, 848-950, 948-1150

The following tables are too large to be contained in the thesis, they are avalibale online as

Supporting Information at:

https://pubs.acs.org/doi/10.1021/acs.analchem.1c01376.

Supplementary Table 2-2 (.xls): LC-MS/MS analyses of HeLa digest with and without FAIMS (separate file)

Supplementary Table 2-3 (.xls): LC-MS/MS analyses of HeLa digest with and without FAIMS using mz segments (SIFT) (separate file)

Supplementary Table 2-4 (.xls): LC-MS/MS analyses of HeLa digest with and without FAIMS using mz segment stepping. (separate file)

Supplementary Table 2-5 (.xls): LC-MS/MS analyses of two proteome model with and without FAIMS using m/z segment stepping. (separate file)

2.8 References

1. Cong, Y.; Liang, Y.; Motamedchaboki, K.; Huguet, R.; Truong, T.; Zhao, R.; Shen, Y.; Lopez-Ferrer, D.; Zhu, Y.; Kelly, R. T. Improved Single-Cell Proteome Coverage Using Narrow-Bore Packed NanoLC Columns and Ultrasensitive Mass Spectrometry. *Anal Chem* **2020**, *92* (3), 2665-2671. DOI: 10.1021/acs.analchem.9b04631.

2. Levy, M. J.; Washburn, M. P.; Florens, L. Probing the Sensitivity of the Orbitrap Lumos Mass Spectrometer Using a Standard Reference Protein in a Complex Background. *J Proteome Res* **2018**, *17* (10), 3586-3592. DOI: 10.1021/acs.jproteome.8b00269.

3. Meier, F.; Brunner, A. D.; Koch, S.; Koch, H.; Lubeck, M.; Krause, M.; Goedecke, N.; Decker, J.; Kosinski, T.; Park, M. A.; et al. Online Parallel Accumulation-Serial Fragmentation (PASEF) with a Novel Trapped Ion Mobility Mass Spectrometer. *Mol Cell Proteomics* **2018**, *17* (12), 2534-2545. DOI: 10.1074/mcp.TIR118.000900.

4. Nusinow, D. P.; Szpyt, J.; Ghandi, M.; Rose, C. M.; McDonald, E. R., 3rd; Kalocsay, M.; Jane-Valbuena, J.; Gelfand, E.; Schweppe, D. K.; Jedrychowski, M.; et al. Quantitative Proteomics of the Cancer Cell Line Encyclopedia. *Cell* **2020**, *180* (2), 387-402 e316. DOI: 10.1016/j.cell.2019.12.023.

5. Kelstrup, C. D.; Aizikov, K.; Batth, T. S.; Kreutzman, A.; Grinfeld, D.; Lange, O.; Mourad, D.; Makarov, A. A.; Olsen, J. V. Limits for Resolving Isobaric Tandem Mass Tag Reporter Ions Using Phase-Constrained Spectrum Deconvolution. *J Proteome Res* **2018**, *17* (11), 4008-4016. DOI: 10.1021/acs.jproteome.8b00381.

6. Sonnett, M.; Yeung, E.; Wuhr, M. Accurate, Sensitive, and Precise Multiplexed Proteomics Using the Complement Reporter Ion Cluster. *Anal Chem* **2018**, *90* (8), 5032-5039. DOI: 10.1021/acs.analchem.7b04713.

7. Baker, E. S.; Livesay, E. A.; Orton, D. J.; Moore, R. J.; Danielson, W. F., 3rd; Prior, D. C.; Ibrahim, Y. M.; LaMarche, B. L.; Mayampurath, A. M.; Schepmoes, A. A.; et al. An LC-IMS-MS platform providing increased dynamic range for high-throughput proteomic studies. *J Proteome Res* **2010**, *9* (2), 997-1006. DOI: 10.1021/pr900888b.

8. Guevremont, R.; Barnett, D. A.; Purves, R. W.; Vandermey, J. Analysis of a tryptic digest of pig hemoglobin using ESI-FAIMS-MS. *Anal Chem* **2000**, *72* (19), 4577-4584. DOI: 10.1021/ac0000271.

9. Helm, D.; Vissers, J. P.; Hughes, C. J.; Hahne, H.; Ruprecht, B.; Pachl, F.; Grzyb, A.; Richardson, K.; Wildgoose, J.; Maier, S. K.; et al. Ion mobility tandem mass spectrometry enhances performance of bottom-up proteomics. *Mol Cell Proteomics* **2014**, *13* (12), 3709-3715. DOI: 10.1074/mcp.M114.041038.

10. Valentine, S. J.; Plasencia, M. D.; Liu, X.; Krishnan, M.; Naylor, S.; Udseth, H. R.; Smith, R. D.; Clemmer, D. E. Toward plasma proteome profiling with ion mobility-mass spectrometry. *J Proteome Res* **2006**, *5* (11), 2977-2984. DOI: 10.1021/pr060232i.

11. Venne, K.; Bonneil, E.; Eng, K.; Thibault, P. Improvement in peptide detection for proteomics analyses using NanoLC-MS and high-field asymmetry waveform ion mobility mass spectrometry. *Anal Chem* **2005**, *77* (7), 2176-2186. DOI: 10.1021/ac048410j.

12. Prasad, S.; Belford, M. W.; Dunyach, J. J.; Purves, R. W. On an aerodynamic mechanism to enhance ion transmission and sensitivity of FAIMS for nano-electrospray ionization-mass spectrometry. *J Am Soc Mass Spectrom* **2014**, *25* (12), 2143-2153. DOI: 10.1007/s13361-014-0995-8.

13. Pfammatter, S.; Bonneil, E.; McManus, F. P.; Prasad, S.; Bailey, D. J.; Belford, M.; Dunyach, J. J.; Thibault, P. A Novel Differential Ion Mobility Device Expands the Depth of Proteome Coverage and the Sensitivity of Multiplex Proteomic Measurements. *Mol Cell Proteomics* **2018**, *17* (10), 2051-2067. DOI: 10.1074/mcp.TIR118.000862.

14. Dodds, J. N.; Baker, E. S. Ion Mobility Spectrometry: Fundamental Concepts, Instrumentation, Applications, and the Road Ahead. *J Am Soc Mass Spectrom* **2019**, *30* (11), 2185-2195. DOI: 10.1007/s13361-019-02288-2.

15. Guevremont, R. High-field asymmetric waveform ion mobility spectrometry: a new tool for mass spectrometry. *J Chromatogr A* **2004**, *1058* (1-2), 3-19.

16. Hebert, A. S.; Prasad, S.; Belford, M. W.; Bailey, D. J.; McAlister, G. C.; Abbatiello, S. E.; Huguet, R.; Wouters, E. R.; Dunyach, J. J.; Brademan, D. R.; et al. Comprehensive Single-Shot Proteomics with FAIMS on a Hybrid Orbitrap Mass Spectrometer. *Anal Chem* **2018**, *90* (15), 9529-9537. DOI: 10.1021/acs.analchem.8b02233.

17. Schweppe, D. K.; Rusin, S. F.; Gygi, S. P.; Paulo, J. A. Optimized Workflow for Multiplexed Phosphorylation Analysis of TMT-Labeled Peptides Using High-Field Asymmetric Waveform Ion Mobility Spectrometry. *J Proteome Res* **2020**, *19* (1), 554-560. DOI: 10.1021/acs.jproteome.9b00759.

18. Schweppe, D. K.; Prasad, S.; Belford, M. W.; Navarrete-Perea, J.; Bailey, D. J.; Huguet, R.; Jedrychowski, M. P.; Rad, R.; McAlister, G.; Abbatiello, S. E.; et al. Characterization and Optimization of Multiplexed Quantitative Analyses Using High-Field Asymmetric-Waveform Ion Mobility Mass Spectrometry. *Anal Chem* **2019**, *91* (6), 4010-4016. DOI: 10.1021/acs.analchem.8b05399.

19. Bekker-Jensen, D. B.; Martinez-Val, A.; Steigerwald, S.; Ruther, P.; Fort, K. L.; Arrey, T. N.; Harder, A.; Makarov, A.; Olsen, J. V. A Compact Quadrupole-Orbitrap Mass Spectrometer with FAIMS Interface Improves Proteome Coverage in Short LC Gradients. *Mol Cell Proteomics* **2020**, *19* (4), 716-729. DOI: 10.1074/mcp.TIR119.001906.

20. Eliuk, S.; Makarov, A. Evolution of Orbitrap Mass Spectrometry Instrumentation. *Annu Rev Anal Chem (Palo Alto Calif)* **2015**, *8*, 61-80. DOI: 10.1146/annurev-anchem-071114-040325.

21. Michalski, A.; Damoc, E.; Hauschild, J. P.; Lange, O.; Wieghaus, A.; Makarov, A.; Nagaraj, N.; Cox, J.; Mann, M.; Horning, S. Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Mol Cell Proteomics* **2011**, *10* (9), M111 011015. DOI: 10.1074/mcp.M111.011015.

22. Martinez-Val, A.; Garcia, F.; Ximenez-Embun, P.; Ibarz, N.; Zarzuela, E.; Ruppen, I.; Mohammed, S.; Munoz, J. On the Statistical Significance of Compressed Ratios in Isobaric Labeling: A Cross-Platform Comparison. *J Proteome Res* **2016**, *15* (9), 3029-3038. DOI: 10.1021/acs.jproteome.6b00151.

23. Williamson, J. C.; Edwards, A. V.; Verano-Braga, T.; Schwammle, V.; Kjeldsen, F.; Jensen, O. N.; Larsen, M. R. High-performance hybrid Orbitrap mass spectrometers for quantitative proteome analysis: Observations and implications. *Proteomics* **2016**, *16* (6), 907-914. DOI: 10.1002/pmic.201400545.

24. Meier, F.; Geyer, P. E.; Virreira Winter, S.; Cox, J.; Mann, M. BoxCar acquisition method enables single-shot proteomics at a depth of 10,000 proteins in 100 minutes. *Nat Methods* **2018**, *15* (6), 440-448. DOI: 10.1038/s41592-018-0003-5.

25. Ting, L.; Rad, R.; Gygi, S. P.; Haas, W. MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nat Methods* **2011**, *8* (11), 937-940. DOI: 10.1038/nmeth.1714.

26. Paulo, J. A.; O'Connell, J. D.; Gygi, S. P. A Triple Knockout (TKO) Proteomics Standard for Diagnosing Ion Interference in Isobaric Labeling Experiments. *J Am Soc Mass Spectrom* **2016**, *27* (10), 1620-1625. DOI: 10.1007/s13361-016-1434-9.

Chapter 3

3 Proteogenomics and differential ion mobility enables the exploration of the mutational landscape in colon cancer cells

Zhaoguan Wu[1,2]; Eric Bonneil[1]; Michael Belford[3]; Cornelia Boeser[3]; Maria Virginia Ruiz Cuevas[1,4], Sebastien Lemieux[1,4], Jean-Jacques Dunyach[3]; Pierre Thibault[1,2*]

[1]Institute for Research in Immunology and Cancer (IRIC); [2]Department of Chemistry, Université de Montréal, Montréal, QC, [3]ThermoFisher Scientific, San Jose, CA, [4]Department of biochemistry, Université de Montréal, Montréal, QC

Publication Date:August 22, 2022

Anal. Chem. 2022, 94, 35, 12086–12094

https://doi.org/10.1021/acs.analchem.2c02056

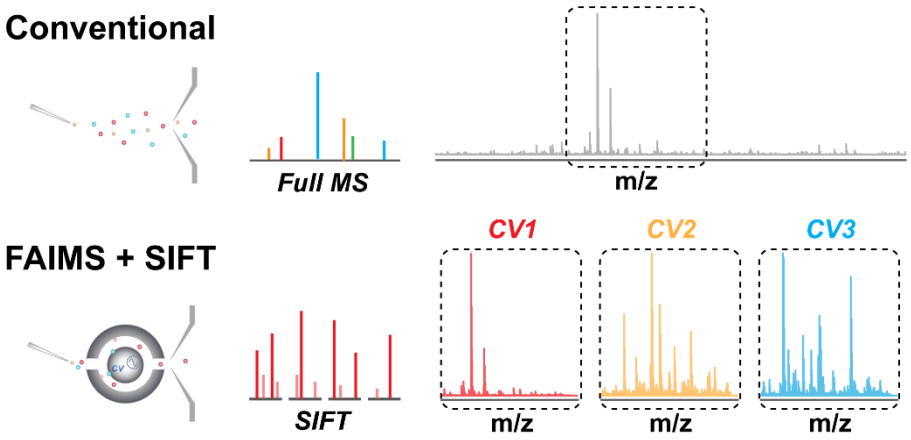
Author contributions:

Zhaoguan Wu prepared and analyzed samples, interpreted results, prepared all figures, and participated in the editing of the manuscript. Eric Bonneil (MS platform manger) supervised the MS analysis, Thibault (supervisor) and Jean-Jacques Dunyach (Industrial partner) developed the concept and managed the project. Michael Belford, Cornelia Boeser and Jean-Jacques Dunyach provided new FAIMS device and participated in the analysis of the data. Maria Virginia Ruiz Cuevas and Sébastien Lemieux provided bioinformatic analyses and correlated mutations with RNASeq data.

3.1 Abstract

The sensitivity and depth of proteomic analyses is limited by isobaric ions and interferences that preclude the identification of low abundance peptides. Extensive sample fractionation is often required to extend proteome coverage when sample amount is not a limitation. Ion mobility devices provide a viable alternate approach to resolve confounding ions, improve peak capacity and mass spectrometry (MS) sensitivity. Here, we report the integration of differential ion mobility with segmented ion fractionation (SIFT) to enhance the comprehensiveness of proteomic analyses. The combination of differential ion mobility and SIFT, where narrow windows of ~ m/z 100 are acquired in turn, is found particularly advantageous in the analysis of protein digests, and typically provided 70% gain in identification compared to conventional single-shot LC-MS/MS. The application of this approach is further demonstrated for the analysis of tryptic digests from different colorectal cancer cell lines where the enhanced sensitivity enabled the identification of single amino acid variants that were correlated with the corresponding transcriptomic datasets.

Key words: High field asymmetric waveform ion mobility (FAIMS), mass spectrometry, gas-phase fractionation, proteogenomics, protein mutation, single amino acid polymorphism, single amino acid variant, colon cancer



3.2 Introduction

Technological advances in mass spectrometry (MS) enables the acquisition of mass spectra with great speed and sensitivity to enhance the depth of proteomic analyses. Single-shot LC-MS/MS using data-dependent acquisition (DDA) still remains one of the preferred analytical strategy in proteomic research, though only about half of the expressed proteome is typically accessible with this approach [1, 2]. Although this limitation is partly accounted for by the MS/MS acquisition rate and the sensitivity of the mass spectrometer, the overwhelming complexity of protein digests also represents a sizeable difficulty towards complete proteome coverage. Several prior reports indicated the challenges in sequencing peptide ions of low abundance that often overlap with isobaric multiply-charged ions, and remain inaccessible or unidentified by MS/MS [3-6]. The proportion of conflicting ions increases with decreasing abundance, a situation that can also affect quantitative measurements when interfering ions are wrongly assigned to target ions. This problem is particularly important in multiplex quantitative proteomics using isobaric peptide labeling due to the frequency of precursor ion co-selection that can lead to distorted reporter ion ratios and ratio compression effects [7, 8].

Different analytical approaches can be used to improve peak capacity and resolve isobaric peptide ions. For example, liquid chromatography prefractionation using ion exchange, hydrophilic-interaction or high pH reverse phase were reported to extend the depth of proteomic analyses[9-11]. Although effective, these approaches require significant sample amounts and can be laborious to integrate in a seamless fashion with LC-MS/MS systems. Alternatively, the combination of ion mobility with MS offers a viable option to perform fractionation of peptide ions in the gas phase, and can be integrated with different types of mass spectrometers to enhance proteome coverage [12-14]. Among the different types of ion mobility devices, field asymmetric waveform ion mobility spectrometry (FAIMS) enables gas phase separation by transmitting ions based on their changes in mobility at high and low electric fields [15, 16]. This analytical characteristic is leveraged in proteomic research to extend the dynamic range of detection, resolve isobaric ions and facilitate the identification of low abundance peptides [17-19].

While ion mobility devices can improve sensitivity and tease out peptide ions from interfering signal, the ability to detect multi-isotope patterns for subsequent identification is

challenging in view of the variability of weak ion signal. This difficulty is further compounded with the limited ion capacity of trapping device where low abundance peptide ions are underrepresented when co-isolated with highly abundant multiply-charged species. To alleviate this problem, a data acquisition method, termed segmented ion fractionation (SIFT), was recently introduced to separate the MS acquisition range into multiple narrow m/z scans [20]. SIFT can be combined with FAIMS to enhance the detection of low abundance ion signal and resolve overlapping peptide features in order to improve precursor ion purity for MS/MS analysis. Although this approach extended the dynamic range of peptide identification compared to conventional LC-MS/MS [20], these preliminary experiments were performed on a Q-Exactive HF with limited FAIMS functionality.

In the present study, we provide a comprehensive evaluation of the SIFT and FAIMS method on an Exploris 480 mass spectrometer where different compensation voltages (CVs) are fully integrated and optimized for separate m/z segments. When compared to the conventional DDA approach, the combination of SIFT and FAIMS enhanced peptide and protein identification by more than 40% and 70%, respectively. The application of this method is further demonstrated for the identification of polymorphic variants in colorectal cancer cell lines where point mutations were correlated with transcriptomic analyses.

3.3 Materials and Methods

3.3.1 Cell cultures.

Three human colorectal cancer cell lines (COLO 205, HCT 116 and SW620) and one normal human fetal small intestine cell line (HIEC6) were obtained from the American Type Culture Collection (ATCC). COLO205, HCT116, and SW620 were grown in RPMI-1640 (Gibco) supplemented with 10% Fetal bovine serum (FBS), while HIEC-6 was grown in OptiMEM 1 Reduced Serum Medium (Gibco) supplemented with 20 mM HEPES (Gibco), 10 mM GlutaMAX (Gibco), 10ng/mL epidermal growth factor (EGF) (Gibco), and FBS to a final concentration of 4%. All cells were maintained at 37°C with 5% CO_2.

Cells were rinsed with warm phosphate-buffered saline (PBS) before being trypsinized with TrypLE™ Express Enzyme (1×) (Gibco) for 5-15 minutes at 37°C with 5% CO_2. Harvested material was then spun at 1000 rpm for 5 minutes, rinsed once with warm PBS, then resuspended in ice-cold PBS. After cell count, replicates of 2×10^8 cells were pelleted and frozen at -80°C until further use.

3.3.2 RNA extraction.

Cells were collected (1-2 million each) and washed once with ice-cold PBS. Cells were then resuspended in Trizol (Invitrogen). Total RNA was isolated using the RNeasy Mini kit (Qiagen) as recommended by the manufacturer.

3.3.3 RNA sequencing.

500 ng of total RNA was used for library preparation. RNA quality control was assessed with the Bioanalyzer RNA 6000 Nano assay on the 2100 Bioanalyzer system (Agilent Technologies) and all samples had an RNA integrity number (RIN) of at least 9.6 for all cell lines. Libraries were prepared with the KAPA mRNAseq Hyperprep kit (Roche). Ligation was made with Illumina dual-index UMI (IDT). After being validated on a BioAnalyzer DNA1000 chip and quantified by QuBit and qPCR, libraries were pooled to equimolar concentration and sequenced with the Illumina Nextseq500 using the Nextseq High Output 150 (2×75bp) cycles kit. A mean of

129 million paired-end PF reads were generated for the cell lines. Library preparation and sequencing were performed at the Genomic Platform of the Institute for Research in Immunology and Cancer (IRIC).

3.3.4 Protein extraction and enzymatic digestion.

Cells were washed twice with cold Phosphate Buffered Saline (PBS) (Fisher Scientific, BP399-1) and pelleted by centrifugation (1000 rpm, 5min). Lysis buffer containing 8M Urea (Fisher Scientific, BP169), 50mM Tris (Thermofisher Scientific) pH 8.2 was added to the cell pellets. Cells were mechanically lysed with 2 × 10s sonication bursts. Lysates were centrifuged at 14,000 g for 10min, and protein concentration of clear lysate was determined by Bradford assay. Samples were diluted to 1M urea 50mM Tris and Lys-C (Wako Pure Chemical Corp.) (enzyme/protein ratio 1:100) digestion was performed for 3h at 37°C. Trypsin (Promega) digestion (enzyme/protein ratio 1:50) was performed overnight at 37°C. Samples were desalted with OASIS HLB columns (Waters) and dried in a speed-vac. Samples were reconstituted in Formic acid 4% (EMD Millipore Corporation).

3.3.5 LC-MS/MS.

All analyses were performed on an Orbitrap Exploris 480™ platform (Thermo Fisher Scientific). Peptide separation was performed on a home-made C_{18} column (Jupiter C18, Phenomenex, 3 µm, 300 Å, 150 µm × 25 cm) connected to a nanoflex electrospray source (Thermo Fisher Scientific). Peptides were loaded on a C_4 precolumn (Optimize Technologies) and eluted at 600 nL/min with a linear gradient of 5–38% solvent B (80% acetonitrile, 0.2% FA) in 56 min, followed by 38%–100% B in 2 min and held at 100% B for 12 min. For conventional full scan analyses, MS survey scans were obtained from m/z 350 to 890 at a resolution of 480000, a standard automatic gain control (AGC) target, and the Auto mode for the maximum injection time. MS/MS scans were acquired for a maximum of 3 seconds with an intensity threshold at 8E4, a MS/MS resolution at 30000, a normalized AGC target at 50% and a maximum injection time at 50ms. For SIFT analyses, MS parameters were the same except that the m/z range was divided in four m/z segments (350-453, 451-542, 540-661, 659-890) in 4 injections. MS/MS spectra of precursor ions with charge state 2-4 were acquired for a maximum of 3 s at a resolution of 30000, intensity

threshold 4E4, normalized AGC target at 100%, and a maximum injection time at 100 ms. FAIMS acquisitions without SIFT were conducted with the same MS parameters except that the resolution was set to 60000. The full mass range and segmented acquisitions used an intensity threshold at 1E4. MS/MS spectra were acquired with a normalized AGC target of 50% and maximum injection time of 75 ms. For FAIMS acquisition with SIFT, MS/MS spectra were acquired with a normalized AGC target 100% and maximum injection time at 100 ms. The isolation window was set to 1.0 Th with 34% HCD normalized collision energy (NCE). A dynamic exclusion time of 30 s was used for all MS/MS acquisitions.

3.3.6 FAIMS.

The FAIMS Pro™ (Thermo Fisher Scientific) inner and outer electrodes were separated by a 1.5 mm gap. The inner and outer electrodes were maintained at 100 °C to maximize ion transmission. Nitrogen (N_2) was used as a carrier gas with a flow rate of 1.6 L/min for temperature controlling and a flow rate of 4 L/min for user carrier. The dispersion voltage (DV) was set to −5000 V with a 3 MHz frequency for the high electric field. The FAIMS transit time was 40 ms. The compensation voltage (CV) for optimal transmission of target peptide ions in LC-FAIMS-MS/MS experiments was determined for full range and SIFT acquisition by running LC-FAIMS-MS/MS of Colo205 tryptic digest at a single CV per run every 5V from -25V to -95V (***Supplementary Figure 3-1***).

3.3.7 Data Analysis and Visualization.

All raw files were searched with PEAKS engine (Bioinformatics Solutions, Inc., Version 10) against the Uniprot human database (Jan 2021). Additional searches were performed using PEAKS and SPIDER[21] to identify potential amino acid variants. Maximal tolerances for precursor and fragments were 10 ppm and 0.01 Da, respectively. Search parameters included trypsin with a maximum of three missed cleavages per peptide. A maximum of 3 variable modifications was allowed per peptide, and included oxidation (M), deamidation (NQ), carbamidomethylation (C) and phosphorylation (STY). For SPIDER search, de novo score (%) threshold was set to 15, L equals I and Q equals K was allowed. False discovery rate (FDR) was set to 1% for peptide

spectrum matches. Data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD033710.

3.3.8 Bioinformatic search of amino acid variants.

Amino acid variants were validated by evaluating the RNA expression of the peptide bearing them in the corresponding colorectal RNASeq sample [22]. For this purpose, each peptide was reverse translated to obtain all possible nucleotide sequences, taking special care to trim peptide sequences larger than 16 amino acids. We then mapped these sequences to the GRCh38.p10 genome version (GENCODE v.26) with STAR v. 2.7.1a, to locate all genomic regions capable of encoding a given peptide without mismatches. The alignment was performed taking into consideration variants from dbSNP (release 155). From these genomic regions, we counted the number of reads present in each RNA-seq sample. Finally, peptides that had a total read count from all different coding regions and sequences ≥ 5 and that were not contained in any proteins in the human UniProtKB 2021 or in translated proteins from the Ensembl 88 (GENCODE v.26) annotations were validated as peptides bearing amino acid variants.

3.4 Result and Discussion

3.4.1 Optimization of segmented ion fractionation and FAIMS

To evaluate the analytical benefits of segmented ion fractionation and FAIMS on the Exploris 480 mass spectrometer, we first conducted a series of conventional LC-MS/MS experiments with a tryptic digest of the colorectal cancer cell line Colo205. Replicate injections of 2 μg were performed using a 70 min gradient elution and a full scan acquisition (m/z 350-890) to determine the distribution of peptide ions across the m/z range (***Supplementary Figure 3-1a***). Replicate LC-MS/MS injections were performed to further improve the number of identified peptides, but provided only 20 % gain in new identifications. Closer examination of replicate analyses revealed the reacquisition of abundant ions with few additional identifications from low abundance peptide ions (***Figure 3-1a***). The additional increase in new identifications followed the normal distribution of detected features, and displayed the expected decrease in matching rate with lower peptide ion intensity (***Figure 3-1b***). The modest increase in identification observed here is partly attributed to the limited space charge capacity of the Orbitrap, which results in the under sampling of low abundance species. To improve the dynamic range of peptide identification, we performed experiments using narrow m/z scans (SIFT) with and without FAIMS. The use of narrow MS scan alone enabled the transmission of a wider distribution of peptide ions for each MS segment, and facilitated the detection of low abundance ions (***Figure 3-1c***). Generally, this translated into an increase of 7 % in peptide identification. However, we noted that the extent of spectral complexity increased in each MS segment giving rise to a higher proportion of co-eluting isobaric peptides as intensity decreased. In contrast, the combination of SIFT and FAIMS significantly reduced the fraction of isobaric peptides while providing new identifications for each CV (***Figure 3-1d***). The improved signal to noise observed with FAIMS facilitated the transmission of peptides not typically detected under conventional LC-MS/MS experiments.

Following these preliminary experiments, we optimized the number of m/z segments and CV steps to properly integrate SIFT and FAIMS into a reduced number of LC-MS/MS experiments. More than 95% of identified peptides (all precursor ion charge stated considered) were detected in the range of m/z 350-890. Accordingly, we first determined the distribution of identified peptides across the m/z range, and selected up to four segments of approximately equal

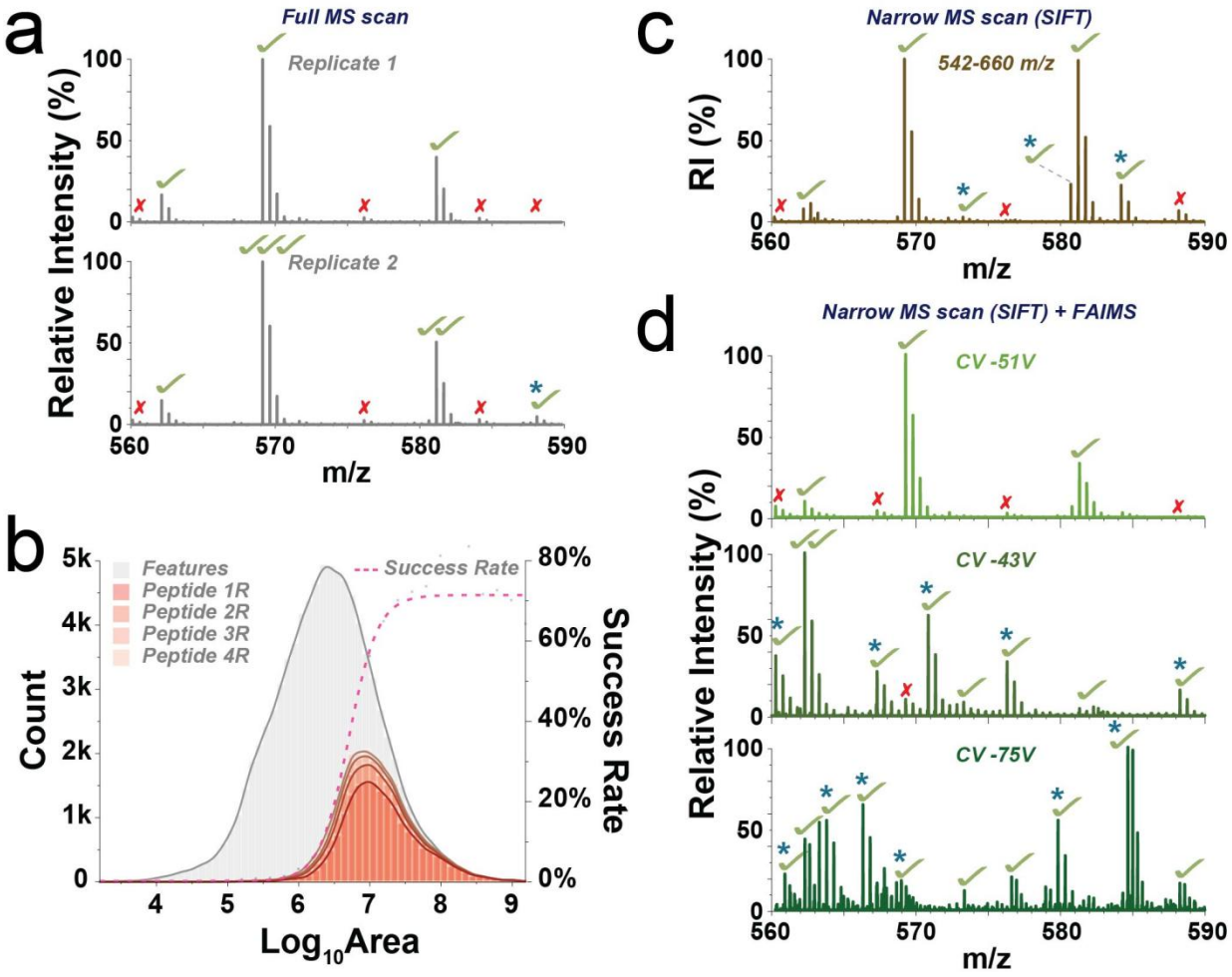


Figure 3-1: Improvement in peptide identification using FAIMS and SIFT in LC-MS/MS analyses of Colo205 tryptic digests.

a) Replicate LC-MS/MS analyses showing a narrow segment of the full mass spectrum. b) Distribution of features, identified peptides and MS/MS success rate vs. intensity for replicate LC-MS/MS injections. c) LC-MS/MS analysis using SIFT. d) LC-MS/MS analysis using FAIMS and SIFT. Check marks and cross signs indicate identified and unassigned precursor ions, while asterisks show new identifications from consecutive analyses.

number of identification (***Supplementary Figure 3-1a***). We also ensured that each segment overlapped by 2 m/z unit to maximize peptide identification. We evaluated the impact of MS resolution on the number of features detected for one to four segments (***Supplementary Figure 3-1b***). Both MS resolution and the number of segments improved feature detection, although this resulted in longer acquisition time. For replicate analyses performed at a single m/z segment we observed a progressive increase in the number of identified features, varying from 55796 features (16338 peptides) to 114778 features (15280 peptides) for resolution settings of 60000 to 480000,

respectively. MS analyses conducted with up to four m/z segments at resolution of 480000 further increased the number of features to 151511 features (26764 peptides).

Next, we optimized transmission conditions of peptide ions according to CV values for single and multiple m/z segments (***Supplementary Figures 3-1c and 3-1d***). LC-FAIMS-MS/MS experiments performed at individual CVs from -25 to -95V enabled the identification of a total of 39814 peptides, of which more than 98% are transmitted between -30 to -80V. An overlap of 3-36 % of identified peptides was observed between adjacent CVs, and this overlap did not exceed 19% for CV values spaced by 10V. Since no single CV can provide sufficient peptide coverage, we combined three non-overlapping CVs into four separate LC-MS/MS analyses for FAIMS experiments using a single m/z segment. Lastly, we selected up to six non-overlapping CV steps for optimal transmission of peptide ions in each of the four individual m/z segments with the number of CV steps scaled according to ion peptide density with m/z 451-542 and m/z 540-661 representing the most populated segments.

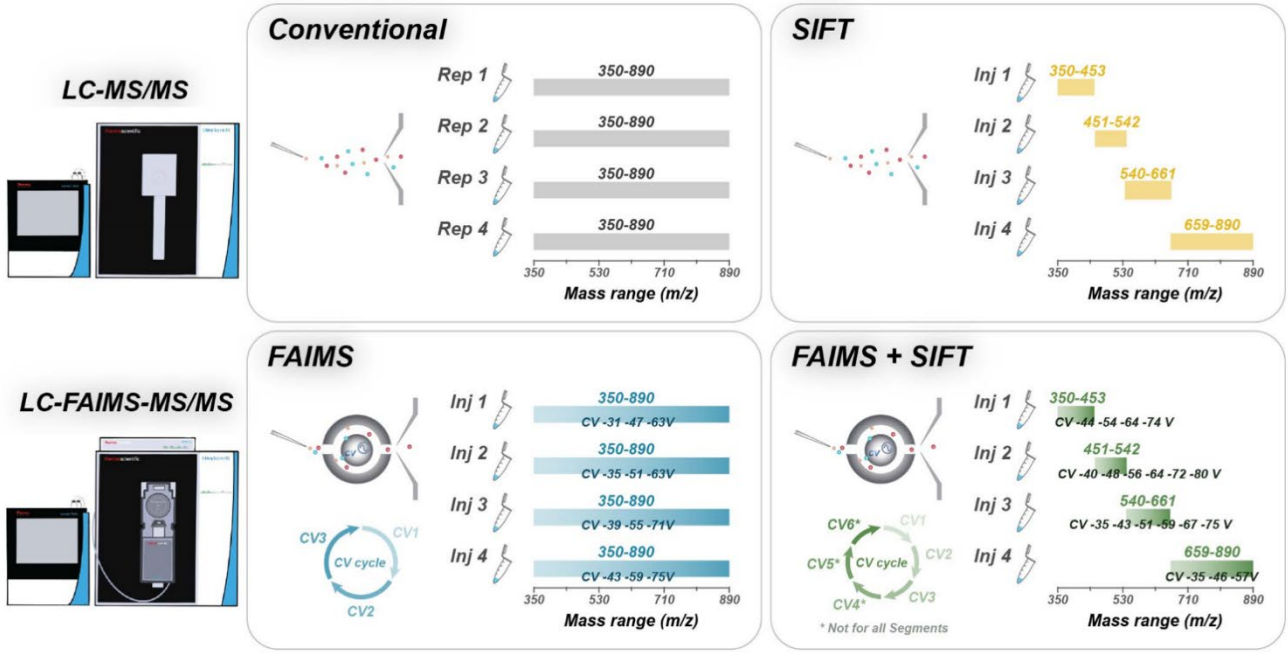


Figure 3-2: Schematic description of experimental conditions used with four acquisition methods.

See experimental section for more details.

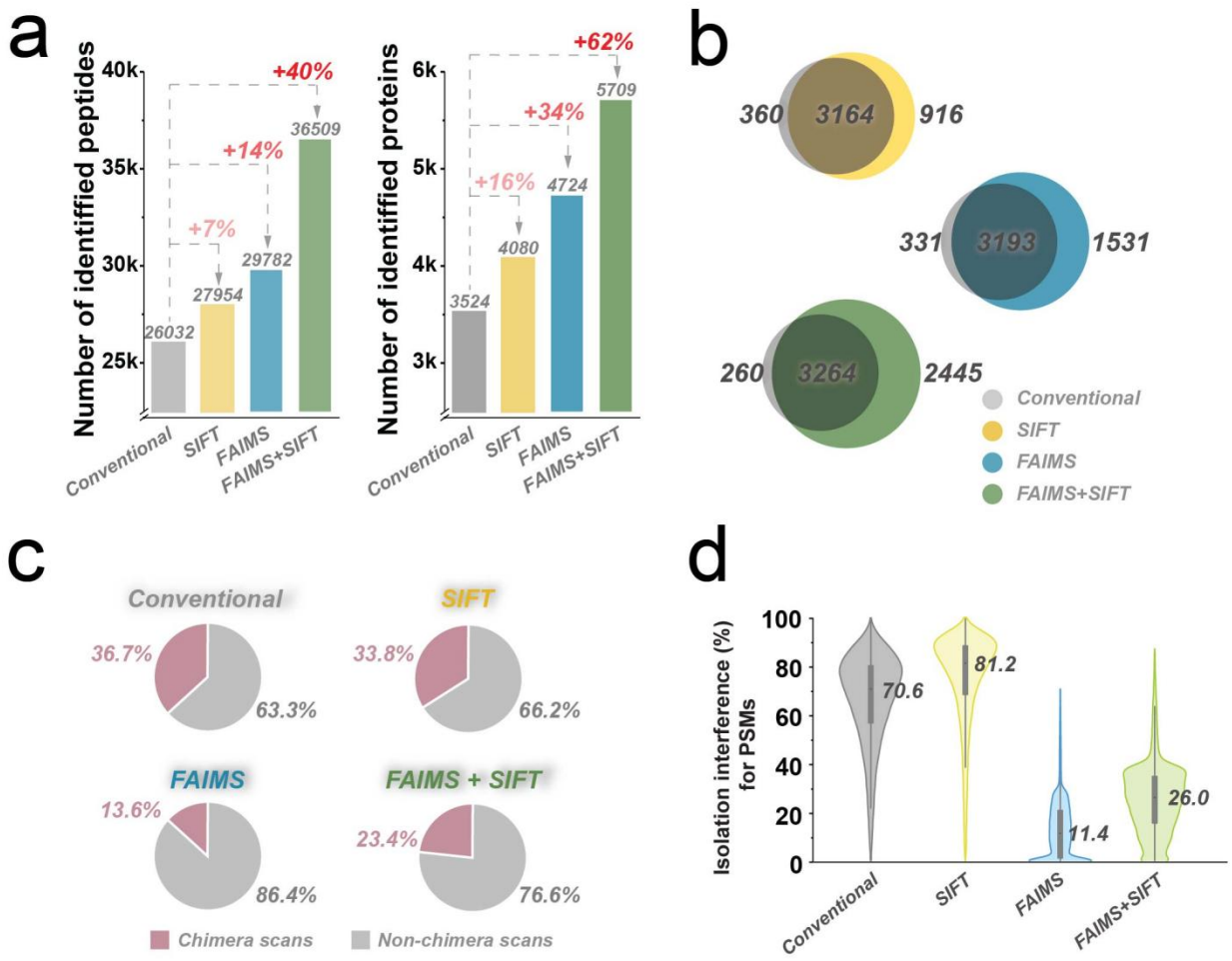


Figure 3-3: Improvement of proteome coverage in LC-MS/MS analyses of Colo205 tryptic digests using SIFT and FAIMS.

a) Distribution of identified peptides and proteins. b) Venn diagrams showing the protein overlap for segmented and non-segmented analyses with and without FAIMS. c) Pie chart representation of chimera and non-chimera scans for the four methods used. (d) Violin plots showing the proportion of interferences co-isolated with precursor ions for all methods.

A schematic representation used for the four different acquisition methods along with the corresponding parameter settings are shown in ***Figure 3-2***. We conducted additional experiments to evaluate the impact of injection time (IT) and the automatic gain control (AGC) on the number of identified peptides and matching rates for each acquisition method (***Supplementary Figure 3-2***). Optimal acquisition conditions were found using an AGC of 5×10^4 and an IT of 50-100 ms.

For each method, we performed LC-MS/MS experiments on a tryptic digest of Colo205 using a total of four injections. Conventional LC-MS/MS experiments enabled the identification of 26032 unique peptides corresponding to 3524 proteins (***Figure 3-3a***, ***Supplementary Table 3-1***). The number of identifications followed a normal distribution where a lower number of peptides were identified with decreasing precursor ion intensity (***Supplementary Figure 3-3***). SIFT using 4 MS segments improved the detection of low abundance ions and enabled a gain of 7 % and 16% in peptide and protein identification, respectively. Further gains in sensitivity were obtained using FAIMS either alone or in combination with SIFT. For instance, LC-FAIMS-MS/MS experiments identified a total of 29782 peptides (4724 proteins) corresponding to a 14 % increase in peptide identification (34% gain in identified proteins). However, the largest gains in sensitivity were obtained when FAIMS was combined with SIFT, which led to 40% and 62% increase in peptide and protein identification, respectively. This acquisition method increased proteome depth and extended detection by almost an order of magnitude in peptide intensity compared to conventional LC-MS/MS experiments (***Supplementary Figure 3-3***). Comparison of proteins identified by each method revealed that SIFT and FAIMS not only provided a 93% overlap, but also uncovered 2445 additional proteins that were not identified with the traditional DDA approach (***Figure 3-3b***). It is noteworthy that additional identifications observed with FAIMS and SIFT also improved sequence coverage by 26-82 %, and that on average peptide ions unique to this method were of lower abundance than peptides common with conventional LC-MS/MS (***Supplementary Figure 3-4***). The improved sensitivity achievable using SIFT and FAIMS enabled the identification of low abundance proteins including for C2H2 zinc finger, DNA binding, and general transcription factors (***Supplementary Figure 3-5***). Furthermore, we obtained comparable gains in peptide and protein identifications with FAIMS and SIFT when using different database search engines such as PEAKS, Sequest HT and COMET (***Supplementary Figure 3-6***).

The difficulty of sequencing lower intensity peptides is not only associated with sporadic fragment ion signals but also with conflicting isobaric species present in the MS spectrum. These low abundance interfering ions cannot be mitigated upon improvement in signal-to-noise ratios alone. Indeed, we noted that chimeric spectra represented 36.7 % of all acquired MS/MS spectra in conventional experiments, and this proportion remained relatively similar when using SIFT (***Figure 3-3c***). In contrast, FAIMS enabled the reduction of interferences by transmitting peptide ions at different CV values, thus reducing the number of chimeric MS/MS spectra to 13.6 %. The

increased ion signal available with SIFT was leveraged using FAIMS, though the proportion of chimeric spectra increased to 23.4 %. Similar observations were also noted when examining the relative amount of ion current associated with interfering ions within the isolation window where these proportions were 70.6, 11.4 and 26.0% for conventional, FAIMS and FAIMS with SIFT experiments, respectively (***Figure 3-3d***). Altogether these results indicated that the combination of FAIMS and SIFT enhanced the depth of proteome analysis by reducing isobaric interferences and increasing ion signals, which either method could not achieve separately.

3.4.2 Identification of amino acid mutations in colorectal cancer cell lines

The enhanced sensitivity observed by combining FAIMS and SIFT prompted us to evaluate its application to the analysis of polymorphic variants in different colorectal cancer cell lines. Non-synonymous mutations alter protein sequences and can lead to a gain or loss of function that play crucial roles in the development and progression of various types of cancer. Previous large-scale proteomic analyses have uncovered only a small proportion of non-synonymous single nucleotide variants (SNVs) detected by RNA sequencing [23]. The variable expression and low abundance of these mutations in biological samples often dictate the use of targeted analyses with either selected or parallel reaction monitoring for their successful detection. To explore the mutational landscape of colorectal cancer cells, we conducted LC-MS/MS analyses on the tryptic digests of colon epithelial cells from fetal small intestin (HIEC-6) and adenocarcinoma cells (HCT116, Colo205, SW620), and analyzed the corresponding MS/MS spectra with PEAKS *de novo* and SPIDER homology search to identify putative amino acid mutations.

The comparison of LC-MS/MS analyses is shown in ***Figure 3-4a***, and indicates that the combination of SIFT and FAIMS typically provides 39-54 % additional peptide identifications compared to the conventional DDA approach (***Supplementary Table 3-2***). For example, LC-MS/MS analyses performed on SW620 digest identified 25383 peptides (3052 proteins), whereas SIFT and FAIMS bolstered these numbers to 39154 peptides (5276 proteins). Proteins identified with both acquisition methods are presented in ***Supplementary Figure 3-4***. In comparing the number of peptides identified, we noted that 60-70% of identifications were common to both approaches (***Figure 3-4b***). While the selection of CVs partly explained this incomplete overlap, FAIMS clearly enabled the detection of low abundance peptides that were either undetected or

convoluted with interfering isobaric peptides. This is exemplified in ***Figure 3-4c*** for two common KRAS mutations G13D and G12V, which are present in 7.3 and 9.1% of all colorectal carcinoma patients [24]. The tryptic peptides comprising these mutations were not detected in conventional LC-MS/MS experiments, whereas the combination of SIFT and FAIMS enabled the identification of KRAS G13D and G12V mutations in cell lines HCT116 and SW620, respectively.

PEAKS *de novo* and SPIDER searches revealed more than 2200 putative amino acid variants in each cell lines with Colo205 containing 3791 variants. To confirm the occurrence of genuine amino acid mutations, we selected only those for which a RNA transcript was identified with more than 5 reads in the corresponding RNASeq database. Altogether, we correlated typically less than 5% of alleged variants resulting in all four cell lines (***Supplementary table 3-3***). The low correlation observed here is partly attributed to the stringent criteria used (e.g. ≥ 5 reads, no mutation contained in any proteins from UniProtKB 2021 or Ensembl 88 databases) and that variants from protein modifications (e.g. deamidation) or isomeric amino acid (e.g. Leu/Ile) were not considered. Next, we mapped mutations to the corresponding genes and used an upset plot to display the intersection of mutations across cell lines (***Figure 3-5a***). Additional information on the distribution of mutation across chromosomes is also available in Supplementary ***Figure 3-5***. Interestingly, the cell line HCT116 derived from primary tumors showed the highest number of mutated genes, of which 24 genes were shared with other cell lines. HCT116 also displays microsatellite instability (MSI), a predisposition to mutation which results from impaired DNA mismatch repair, while other cell lines are microsatellite stable (MSS).

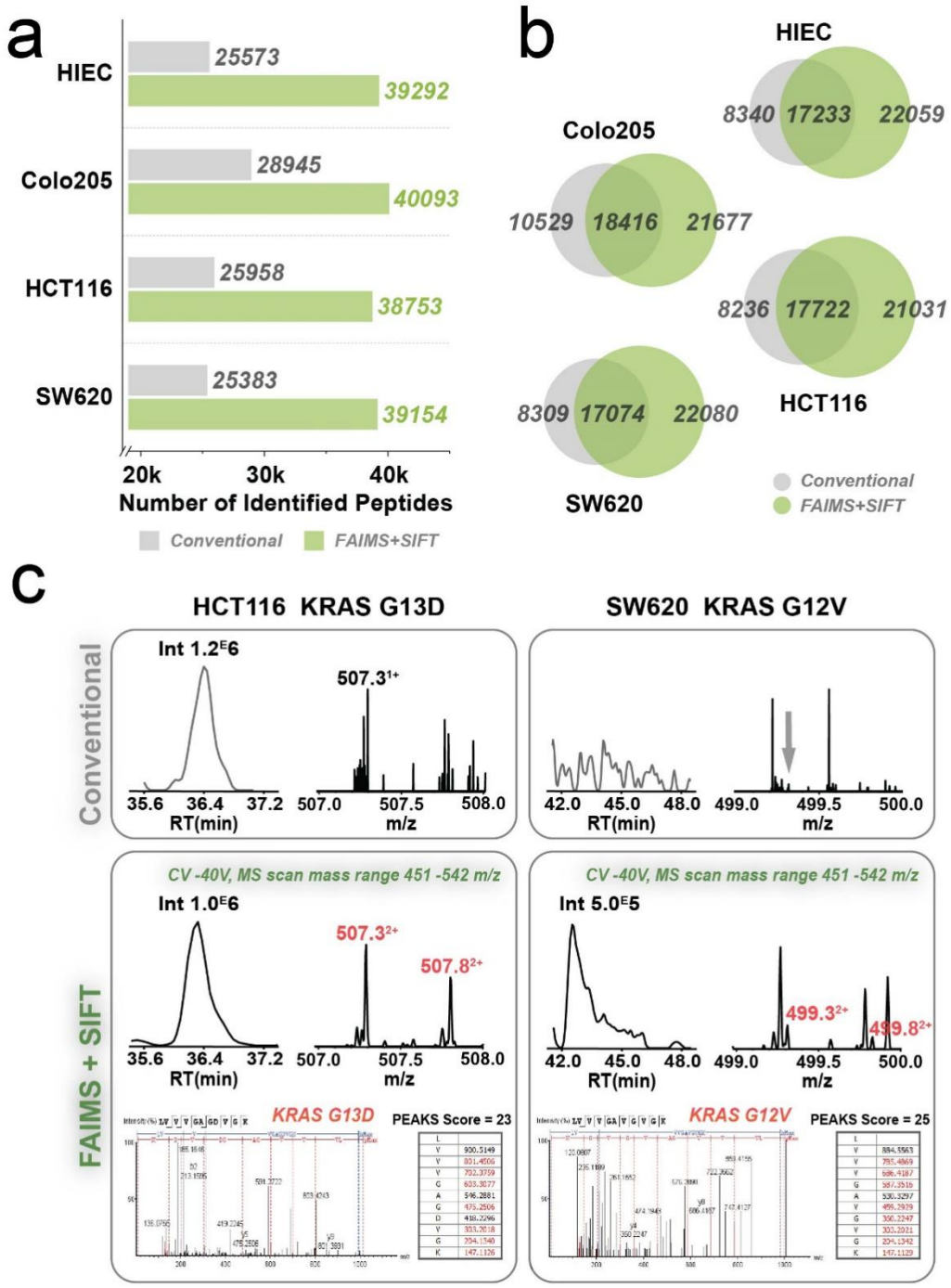


Figure 3-4: LC-MS/MS analyses of tryptic digests from 4 CRC cell lines with and without FAIMS and SIFT.

a) Distribution of identified peptides. b) Venn diagrams showing the peptide overlap for LC-MS/MS analyses performed with and without FAIMS and SIFT. c) Identification of KRAS G13D and G12V mutations in digests of HCT116 and SW620 with their corresponding MS/MS spectra. LC-MS/MS with (bottom) and without (top) FAIMS and SIFT.

Next, we performed gene ontology (GO) analyses to determine if mutated genes were enriched in specific molecular functions or biological processes (***Figure 3-5b***. ***Supplementary Table 3-4***). While each cell line had specific mutation signatures, we noted that several mutated genes found in CRC cell lines were associated with RNA binding (e.g. EEF1A1, EEF1A1P5, EIF4H), ribosome assembly (e.g. RPL5, RPL13, BMS1), cytoskeleton (e.g. ACTB, ACTG1, VIM), antigen processing and presentation (e.g. PSMD13, HLA-A, HLA-B, HLA-C). Also several mutated genes were common across cell lines (e.g. CAP1, CAPN2, LDHA, SEPT9, THRAP3, PDLIM5). To determine if SNVs identified were reported to have an impact on protein function, we checked individual variants against the catalog of somatic mutations in cancer COSMIC (https://cancer.sanger.ac.uk/cosmic) and used FATHMM-MKL [25] to predict the functional consequences of the corresponding mutations. The proportion of mutations with predicted pathogenic consequences (score ≥ 0.7) ranged between 5-21%, with most mutation being either neutral or having unpredicted impact on protein function (***Figure 3-5c***). The CRC cell line SW620 had the highest number of predicted pathogenic mutations. Aside from K-RAS G12V, a somatic mutation frequently observed in colon, lung and pancreas cancers, we also identified mutations where the substitution of amino acid could lead to changes in protein function. For example, the P132S mutation in the enhancer of mRNA-decapping protein 3 (EDC-3) and the S163P in the Ras association domain-containing protein 6 (RASSF6) could lead to changes in the phosphorylation and interactions with binding proteins. These results suggests that the increase sensitivity available using FAIMS and SIFT can facilitate the identification of low abundance protein variants associated with key molecular subtypes of CRC that remain undetectable with single-shot LC-MS/MS experiments.

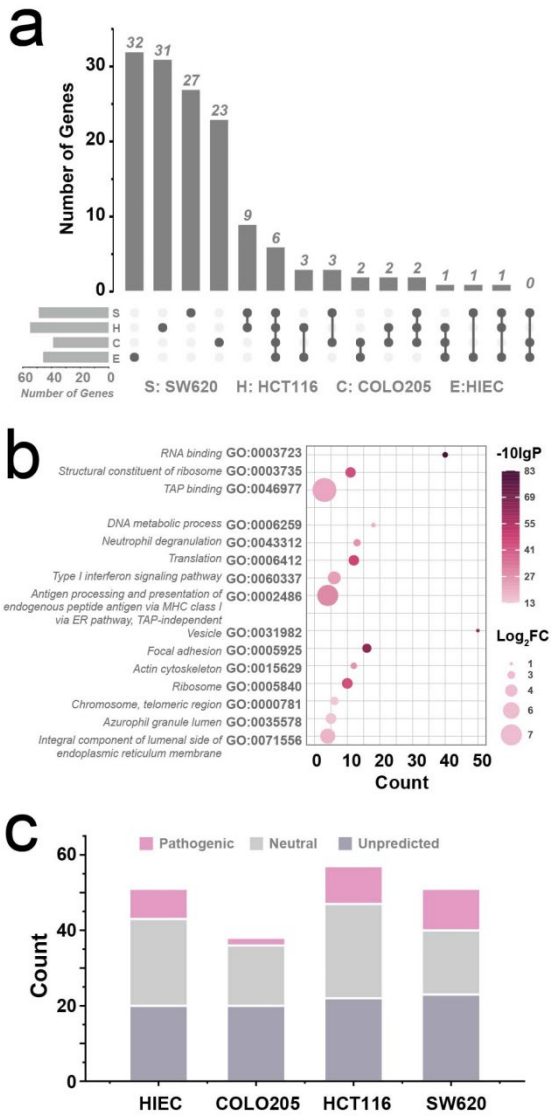


Figure 3-5: Identification of mutated proteins and correlation with RNASeq data.

a) Upset plot showing the distribution of identified mutations across four CRC cell lines. b) Distribution of enriched GO terms for molecular function, biological processes and cellular components for mutated genes. c) Stacked bar charts showing the mutational signatures across different CRC cell lines according to COSMIC and FATHMM predictions.

3.5 Conclusion

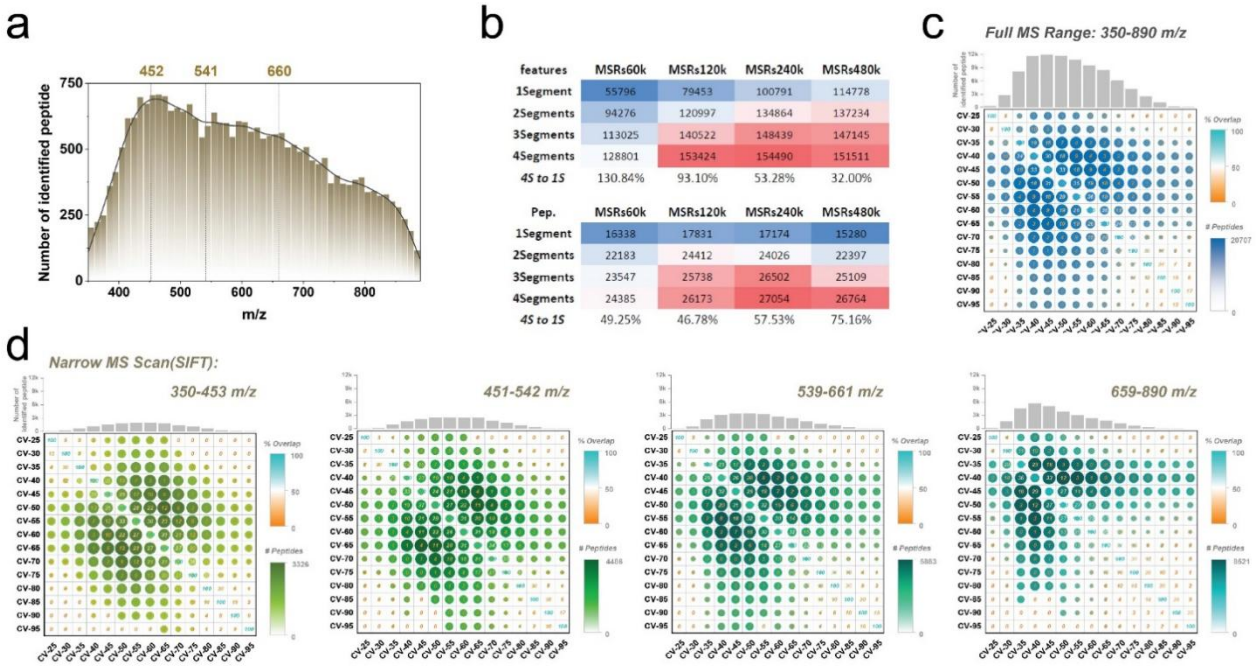
We describe the integration FAIMS and SIFT on an Exploris 480 mass spectrometer to enhance MS sensitivity and increase protein identification. SIFT enables the sequential transmission of ions from narrow mass ranges to favor ion accumulation of low abundance ions that could be under represented in wide MS scan acquisition. However, the corresponding gains in ion accumulation only provides marginal increases in new identifications when using conventional LC-MS/MS. This limitation is partly due to the overwhelming spectral complexity that gives rise to chimeric MS/MS spectra with decreasing precursor ion intensity. FAIMS can be used to reduce the occurrence of isobaric interferences by transmitting ions at different CV values. The combination of FAIMS and SIFT in LC-MS/MS experiments provides unparalleled sensitivity and proteome coverage that either method could not achieve separately. Under optimized conditions, FAIMS and SIFT extends the dynamic range of peptide detection by one order of magnitude, and provides more than 40% and 60% increase in peptide and protein identifications, respectively. While the present study did not examine the analytical benefits of FAIMS and SIFT for quantitative proteomic analyses, we anticipate that the decrease in chimeric tandem mass spectra achievable with this method will be beneficial for TMT quantification to minimize the distortion of TMT reporter ion ratios commonly found using isobaric peptide labeling [20]. This would not only extend the number of quantifiable peptides but also improve the accuracy of quantitative measurements.

The improvement of sensitivity available using FAIMS and SIFT was exploited to identify single nucleotide variants and confirmed the G12V and G13D KRAS mutations in SW620 and HCT116 CRC cell lines, which could not be detected using LC-MS/MS alone. The availability of RNASeq data also facilitated the correlation of other low abundance mutations including the EDC3 P132S and RASSF6 S163P predicted to be pathogenic. The combination FAIMS and SIFT with LC-MS/MS to identify mutations and uncover protein variants without immunoaffinity purification open up interesting perspectives for the exploration of mutational landscape in cancer research.

3.6 Acknowledgements

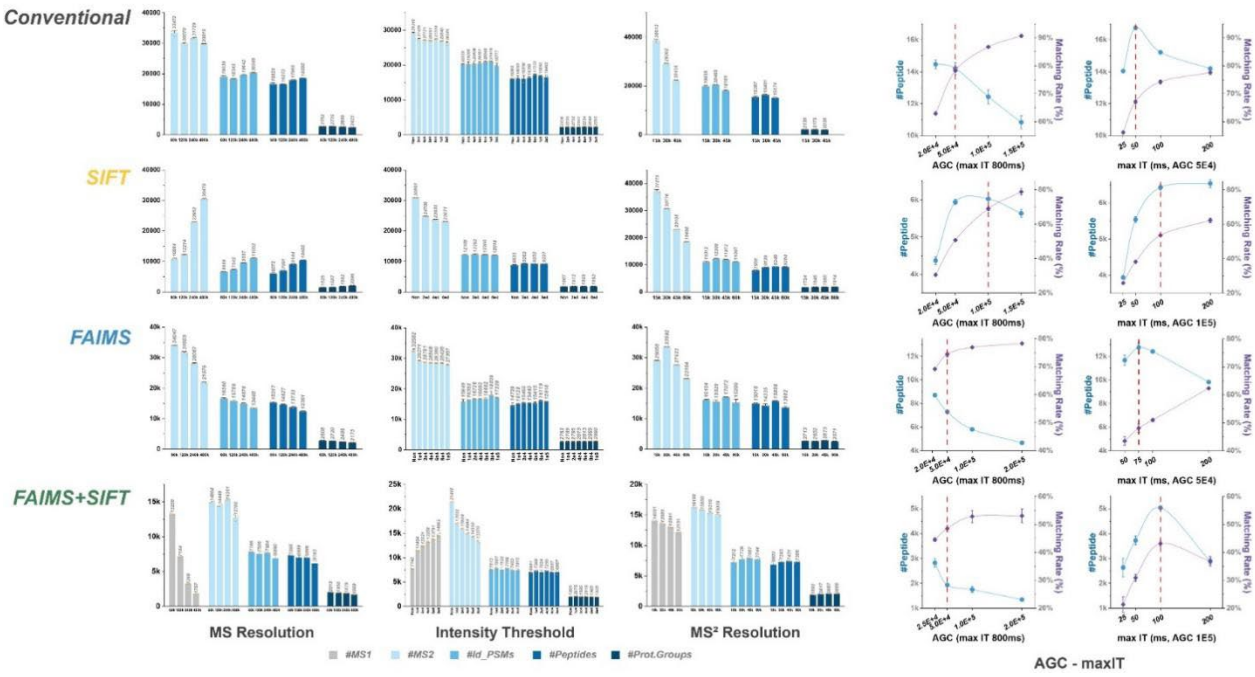
This work was carried out with financial support from the Natural Sciences and Engineering Research Council (NSERC 311598) and the Genomic Applications Partnership Program (GAPP) of Genome Canada. The authors thank Mirela Pascariu, Jenna Cleyle and Joel Lanoix for assistance with cell culturing and RNASeq analyses. The Institute for Research in Immunology and Cancer (IRIC) receives infrastructure support from IRICoR, the Canadian Foundation for Innovation, and the Fonds de Recherche du Québec - Santé (FRQS). IRIC proteomics facility is a Genomics Technology platform funded in part by Genome Canada and Genome Québec.

3.7 Supporting Information.



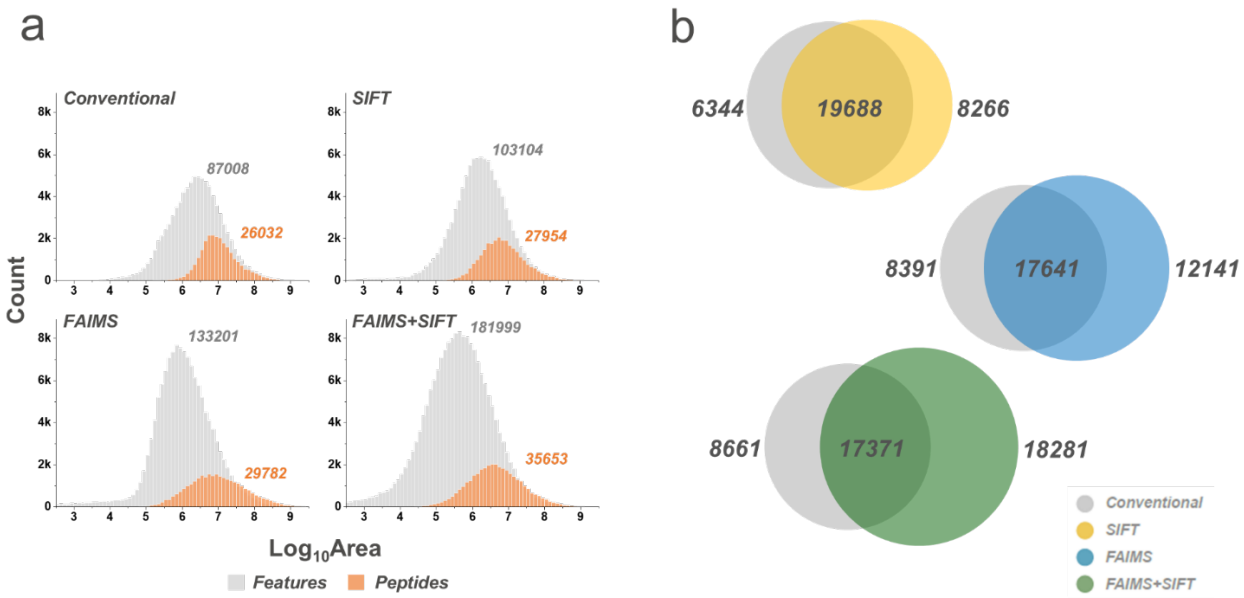
Supplementary Figure 3-1: Optimization of acquisition ranges and CV values for LC-MS/MS using for FAIMS and SIFT.

a) Distribution of identified peptides according to their m/z values. b) Number of detected features and identified peptides for different MS resolution settings and m/z segments. c) Bubble chart displaying the variation of identified peptides and overlap across different CV values for (c) a wide MS segment (m/z 350-890) and (d) several narrow segments (m/z 350-453, 451-542, 540-661, and 659-850). Bar chart shows the distribution of identified peptides for different CV values.



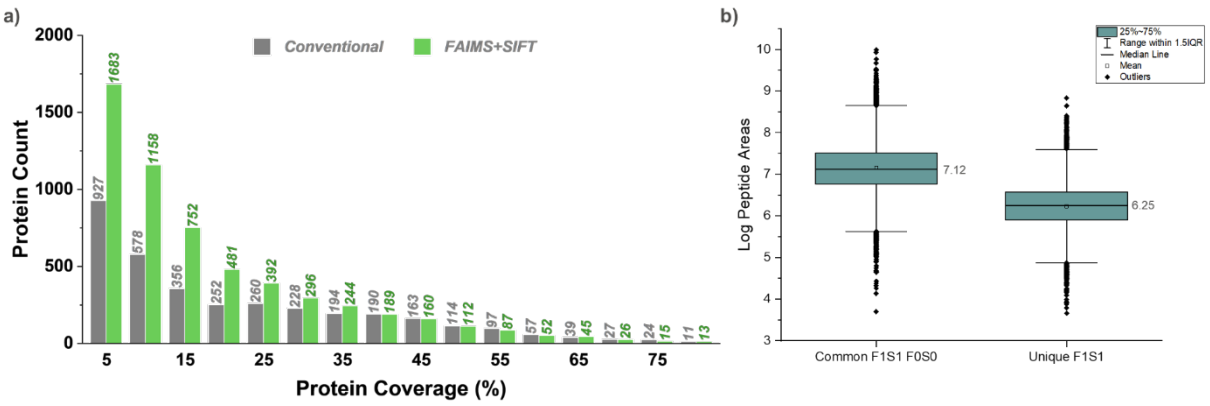
Supplementary Figure 3-2: Optimization of MS resolution and ion accumulation parameters for four different acquisition LC-MS/MS methods.

Each column represents a different optimization parameter (e.g. MS^1 resolution, intensity threshold for MS/MS triggering, MS^2 resolution) and its impact on the number of identified peptides for separate acquisition method. The last two columns represent the variation of automatic gain control (AGC) or maximum injection time (IT) and their impact on the number of identified peptides and MS/MS matching rate for each acquisition method.



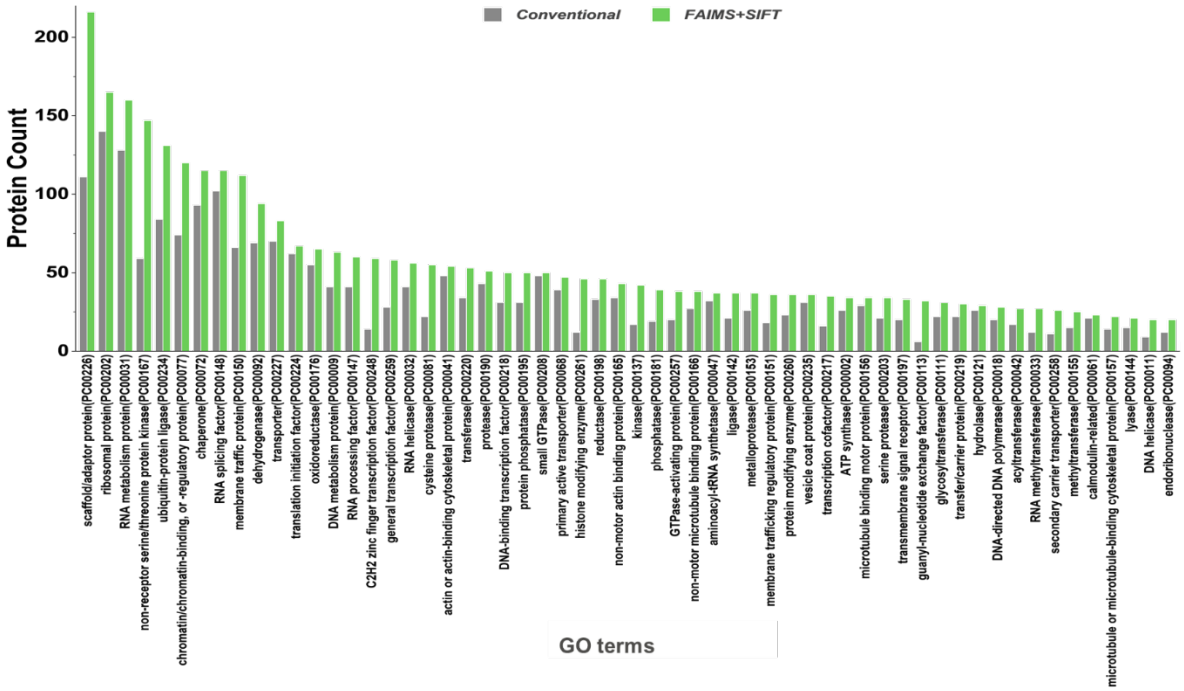
Supplementary Figure 3-3: Comparison of features and identified peptides of Colo205 tryptic digest for four different acquisition methods.

a) Distribution of features and identified peptides according to their intensities. b) Venn diagrams showing the intersections of peptides identified by conventional LC-MS/MS and each other method.

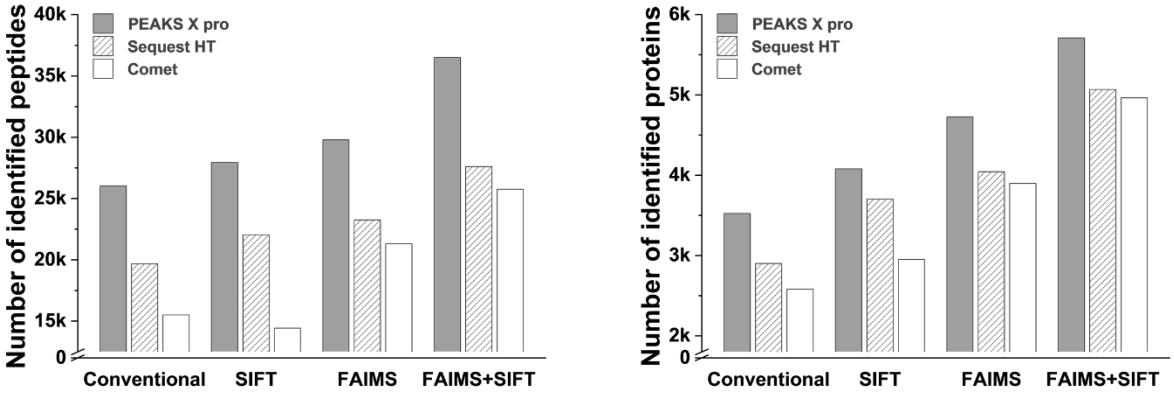


Supplementary Figure 3-4: Improved sequence coverage using FAIMS-SIFT LC-MS/MS.

a) Comparison of sequence coverage for proteins identified in Colo205 tryptic digest for conventional (F0S0) and FAIMS-SIFT (F1S1) LC-MS/MS analyses. b) Boxplots of peptide abundance (peak areas) for peptides common and unique peptides identified for both approaches.



Supplementary Figure 3-5: Distribution of Gene Ontology (GO) terms for groups containing at least 20 proteins identified in Colo205 tryptic digest for SIFT-FAIMS and conventional LC-MS/MS analyses.

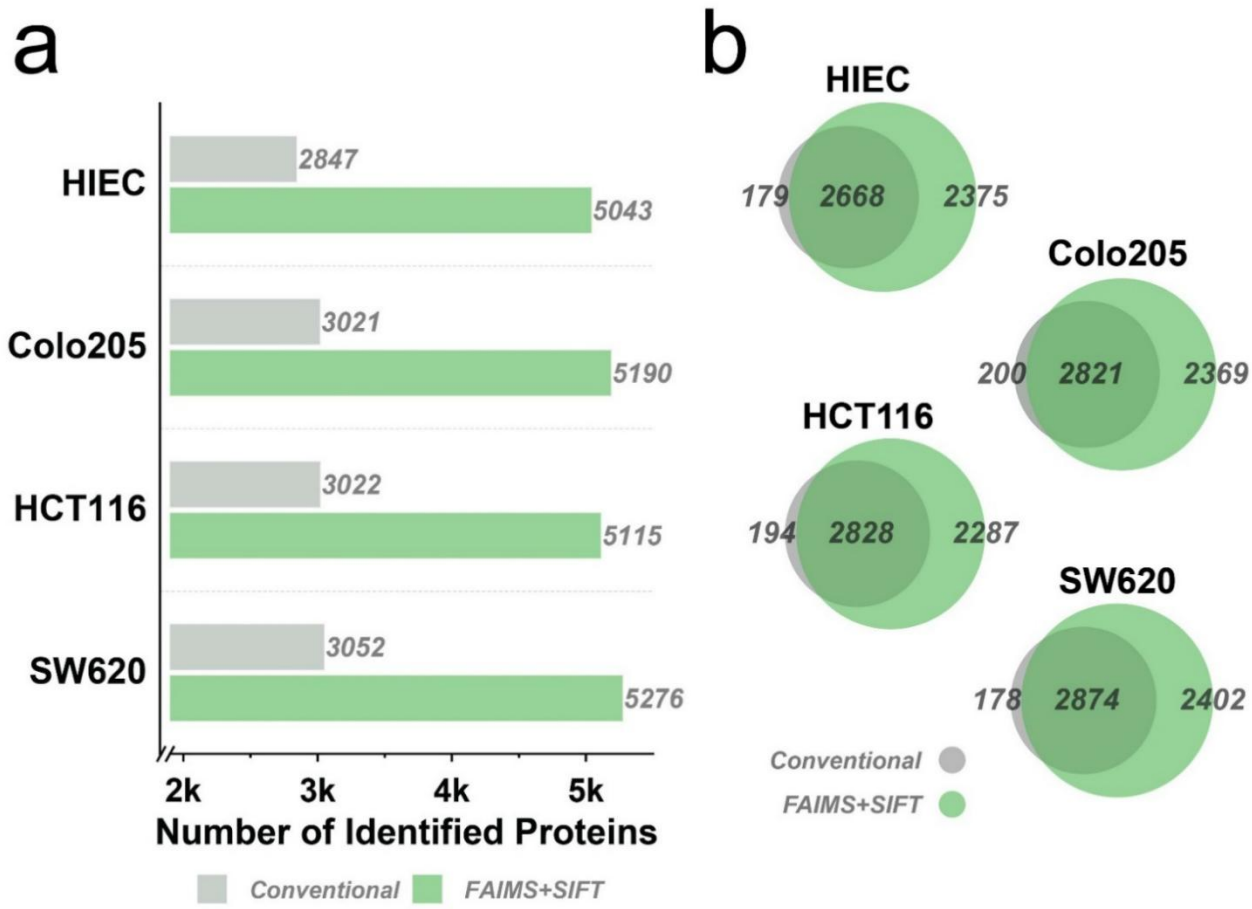


	Number of identified peptides						Number of identified proteins					
	PEAKS X pro	Δ*	Sequest HT	Δ	Comet	Δ	PEAKS X pro	Δ	Sequest HT	Δ	Comet	Δ
Conventional	26032	—	19685	—	15509	—	3524	—	2901	—	2582	—
SIFT	27954	7.38%	22049	12.01%	14432	- 6.94%	4080	15.78%	3702	27.61%	2952	14.33%
FAIMS	29782	14.41%	23252	18.12%	21324	37.49%	4724	34.05%	4044	39.40%	3899	51.01%
FAIMS+SIFT	36509	40.25%	27601	40.21%	25764	66.12%	5709	62.00%	5068	74.70%	4964	92.25%

Δ* compare to #id by conventional approach

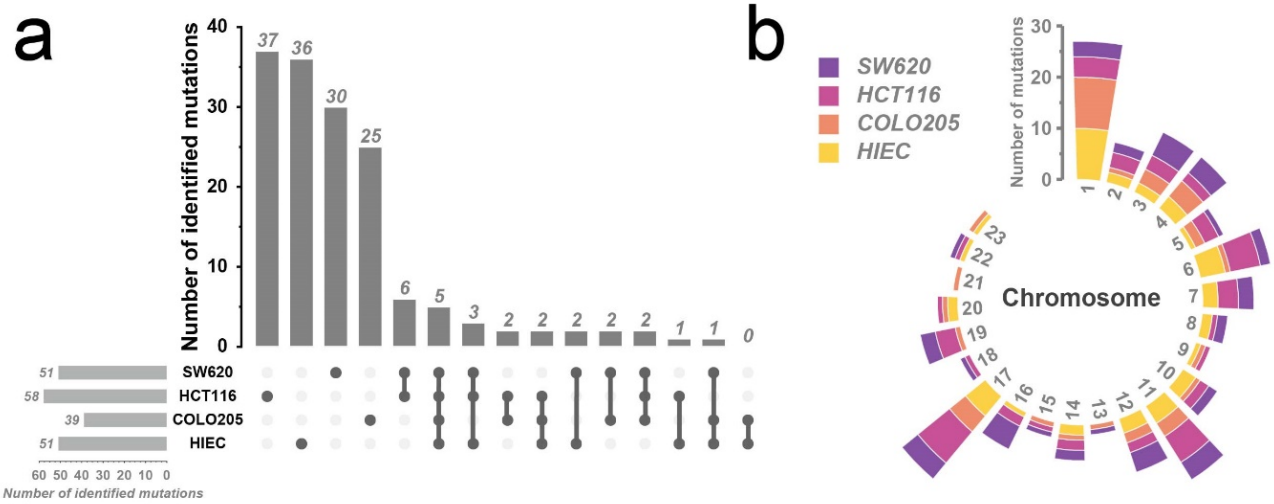
Supplementary Figure 3-6: FAIMS-SIFT enhances peptide and protein identifications irrespective of database search engines used.

Comparison of unique peptides (left) and proteins (right) identified using PEAKS, Sequest HT and COMET for all four acquisition approaches. Table at bottom summarizes changes in the number of identifications between search engines and acquisition approaches.



Supplementary Figure 3-7: Distribution of identified proteins for different CRC tryptic digests using different acquisition methods.

a) Comparison of identified proteins in LC-MS/MS experiments with and without FAIMS and SIFT. Venn diagrams showing the intersection of identified proteins for each CRC cell lines.



Supplementary Figure 3-8: Distribution of mutations identified across different CRC cell lines.

a) Upset plot showing the distribution of specific mutations found across cell lines. b) Distribution of mutations and their location on different chromosomes.

The following tables are too large to be contained in the thesis, they are avalibale online as

Supporting Information at:

https://pubs.acs.org/doi/10.1021/acs.analchem.2c02056.

Supplementary Table 3-1 (.xls): Detected features and identified peptides from LC-MS/MS analyses of Colo205 digest for four different acquisition methods (separate file)

Supplementary Table 3-2 (.xls): Identified peptides from LC-MS/MS analyses of 4 CRC cell lines digest with and without FAIMS and SIFT (separate file)

Supplementary Table 3-3 (.xls): Mutations identified by FAIMS-SIFT LC-MS/MS in 4 CRC cell lines and correlated with RNASeq. (separate file)

Supplementary Table 3-4 (.xls): GO term analyses of mutations identified in 4 CRC cell lines. (separate file)

3.8 References

1. Bian, Y.; The, M.; Giansanti, P.; Mergner, J.; Zheng, R.; Wilhelm, M.; Boychenko, A.; Kuster, B. Identification of 7000-9000 Proteins from Cell Lines and Tissues by Single-Shot Microflow LC-MS/MS. *Anal Chem* **2021**, *93* (25), 8687-8692.
2. Hebert, A. S.; Thoing, C.; Riley, N. M.; Kwiecien, N. W.; Shiskova, E.; Huguet, R.; Cardasis, H. L.; Kuehn, A.; Eliuk, S.; Zabrouskov, V.; et al. Improved Precursor Characterization for Data-Dependent Mass Spectrometry. *Anal Chem* **2018**, *90* (3), 2333-2340.
3. 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. *J Proteome Res* **2011**, *10* (4), 1785-1793.
4. Shishkova, E.; Hebert, A. S.; Westphall, M. S.; Coon, J. J. Ultra-High Pressure (>30,000 psi) Packing of Capillary Columns Enhancing Depth of Shotgun Proteomic Analyses. *Anal Chem* **2018**, *90* (19), 11503-11508.
5. Wang, X.; Shen, S.; Rasam, S. S.; Qu, J. MS1 ion current-based quantitative proteomics: A promising solution for reliable analysis of large biological cohorts. *Mass Spectrom Rev* **2019**, *38* (6), 461-482.
6. Wichmann, C.; Meier, F.; Virreira Winter, S.; Brunner, A. D.; Cox, J.; Mann, M. MaxQuant.Live Enables Global Targeting of More Than 25,000 Peptides. *Mol Cell Proteomics* **2019**, *18* (5), 982-994.
7. Wenger, C. D.; Lee, M. V.; Hebert, A. S.; McAlister, G. C.; Phanstiel, D. H.; Westphall, M. S.; Coon, J. J. Gas-phase purification enables accurate, multiplexed proteome quantification with isobaric tagging. *Nat Methods* **2011**, *8* (11), 933-935.
8. Wuhr, M.; Haas, W.; McAlister, G. C.; Peshkin, L.; Rad, R.; Kirschner, M. W.; Gygi, S. P. Accurate multiplexed proteomics at the MS2 level using the complement reporter ion cluster. *Anal Chem* **2012**, *84* (21), 9214-9221.
9. Bekker-Jensen, D. B.; Kelstrup, C. D.; Batth, T. S.; Larsen, S. C.; Haldrup, C.; Bramsen, J. B.; Sorensen, K. D.; Hoyer, S.; Orntoft, T. F.; Andersen, C. L.; et al. An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes. *Cell Syst* **2017**, *4* (6), 587-599 e584.
10. Di Palma, S.; Hennrich, M. L.; Heck, A. J.; Mohammed, S. Recent advances in peptide separation by multidimensional liquid chromatography for proteome analysis. *J Proteomics* **2012**, *75* (13), 3791-3813.
11. Roca, L. S.; Gargano, A. F. G.; Schoenmakers, P. J. Development of comprehensive two-dimensional low-flow liquid-chromatography setup coupled to high-resolution mass spectrometry for shotgun proteomics. *Anal Chim Acta* **2021**, *1156*, 338349.
12. Hebert, A. S.; Prasad, S.; Belford, M. W.; Bailey, D. J.; McAlister, G. C.; Abbatiello, S. E.; Huguet, R.; Wouters, E. R.; Dunyach, J. J.; Brademan, D. R.; et al. Comprehensive Single-Shot Proteomics with FAIMS on a Hybrid Orbitrap Mass Spectrometer. *Anal Chem* **2018**, *90* (15), 9529-9537.

13. Lennon, S.; Hughes, C. J.; Muazzam, A.; Townsend, P. A.; Gethings, L. A.; Wilson, I. D.; Plumb, R. S. High-Throughput Microbore Ultrahigh-Performance Liquid Chromatography-Ion Mobility-Enabled-Mass Spectrometry-Based Proteomics Methodology for the Exploratory Analysis of Serum Samples from Large Cohort Studies. *J Proteome Res* **2021**, *20* (3), 1705-1715.

14. Meier, F.; Brunner, A. D.; Koch, S.; Koch, H.; Lubeck, M.; Krause, M.; Goedecke, N.; Decker, J.; Kosinski, T.; Park, M. A.; et al. Online Parallel Accumulation-Serial Fragmentation (PASEF) with a Novel Trapped Ion Mobility Mass Spectrometer. *Mol Cell Proteomics* **2018**, *17* (12), 2534-2545.

15. Guevremont, R. High-field asymmetric waveform ion mobility spectrometry: a new tool for mass spectrometry. *J Chromatogr A* **2004**, *1058* (1-2), 3-19.

16. Shvartsburg, A. A.; Smith, R. D. Optimum waveforms for differential ion mobility spectrometry (FAIMS). *J Am Soc Mass Spectrom* **2008**, *19* (9), 1286-1295.

17. Bekker-Jensen, D. B.; Martinez-Val, A.; Steigerwald, S.; Ruther, P.; Fort, K. L.; Arrey, T. N.; Harder, A.; Makarov, A.; Olsen, J. V. A Compact Quadrupole-Orbitrap Mass Spectrometer with FAIMS Interface Improves Proteome Coverage in Short LC Gradients. *Mol Cell Proteomics* **2020**, *19* (4), 716-729.

18. Pfammatter, S.; Bonneil, E.; McManus, F. P.; Prasad, S.; Bailey, D. J.; Belford, M.; Dunyach, J. J.; Thibault, P. A Novel Differential Ion Mobility Device Expands the Depth of Proteome Coverage and the Sensitivity of Multiplex Proteomic Measurements. *Mol Cell Proteomics* **2018**, *17* (10), 2051-2067.

19. Schweppe, D. K.; Rusin, S. F.; Gygi, S. P.; Paulo, J. A. Optimized Workflow for Multiplexed Phosphorylation Analysis of TMT-Labeled Peptides Using High-Field Asymmetric Waveform Ion Mobility Spectrometry. *J Proteome Res* **2020**, *19* (1), 554-560.

20. Pfammatter, S.; Wu, Z.; Bonneil, E.; Bailey, D. J.; Prasad, S.; Belford, M.; Rochon, J.; Picard, P.; Lacoursiere, J.; Dunyach, J. J.; et al. Integration of Segmented Ion Fractionation and Differential Ion Mobility on a Q-Exactive Hybrid Quadrupole Orbitrap Mass Spectrometer. *Anal Chem* **2021**, *93* (28), 9817-9825.

21. Ma, B.; Johnson, R. De novo sequencing and homology searching. *Mol Cell Proteomics* **2012**, *11* (2), O111 014902.

22. Cleyle, J.; Hardy, M. P.; Minati, R.; Courcelles, M.; Durette, C.; Lanoix, J.; Laverdure, J. P.; Vincent, K.; Perreault, C.; Thibault, P. Immunopeptidomic analyses of colorectal cancers with and without microsatellite instability. *Mol Cell Proteomics* **2022**, 100228.

23. Lin, T. T.; Zhang, T.; Kitata, R. B.; Liu, T.; Smith, R. D.; Qian, W. J.; Shi, T. Mass spectrometry-based targeted proteomics for analysis of protein mutations. *Mass Spectrom Rev* **2021**, e21741.

24. Consortium, A. P. G. AACR Project GENIE: Powering Precision Medicine through an International Consortium. *Cancer Discov* **2017**, *7* (8), 818-831.

25. Shihab, H. A.; Rogers, M. F.; Gough, J.; Mort, M.; Cooper, D. N.; Day, I. N.; Gaunt, T. R.; Campbell, C. An integrative approach to predicting the functional effects of non-coding and coding sequence variation. *Bioinformatics* **2015**, *31* (10), 1536-1543.

Chapter 4

4 Conclusion and perspectives

4.1 Conclusion

This work presents the successful integration of a custom (non-commercial) FAIMS Pro™ interface to an older generation Q-Exactive Orbitrap series (Chapter 2). With the optimization of acquisition parameters (e.g. Resolution of MS and MS/MS scan, AGC and maximum injection time etc.), FAIMS increased the number of proteins identified by 50% and extended the dynamic range of peptide detection compared to conventional LC-MS/MS experiments. The integration of FAIMS also provided another dimension of ion separation whereby peptide precursors can be transmitted based on their changes in ion mobility at low and high electric fields. This capability facilitated the resolution of isobaric peptides that often limit the dynamic range of peptide detection leading to co-isolation of precursor ions and generation of chimeric MS/MS spectra. To extend the dynamic range of peptide detection we developed a new approach termed segmented ion fractionation (SIFT) where different m/z segments are analyzed in turn by the mass spectrometer. This technological advance circumvented the limited ion capacity of the Orbitrap mass spectrometer which currently lies in the range of 0.5−1 million elementary charges. The combination of FAIMS and SIFT can be advantageously exploited in quantitative proteomics to improve the accuracy of quantitative measurements as demonstrated for mixtures of human and yeast tryptic digests derivatized with isobaric labeling reagents. This approach led to a 65% improvement in the accuracy of fold-change abundance measurements and can be further improved by filtering data based on an interference free index (IFI).

While the integration of FAIMS and SIFT was accomplished on an old generation of Q-Exactive mass spectrometer, we faced important limitations in terms of integrating FAIMS scanning functions. In this context, Chapter 3 presented a comprehensive evaluation of different acquisition parameters on the Exploris 480 mass spectrometer, which represents the latest generation of Q-Exactive instrument. Under optimized conditions, proteomics analyses performed using FAIMS and SIFT enabled a 40 % and 60% increase in peptide and protein identification compared to conventional LC-MS/MS experiments. This gain in sensitivity also translated into an extension of at least an order of magnitude in peptide detection which also facilitated the identification of protein mutations in colorectal cancer cell lines. For example, our study showed that improvement of sensitivity conferred by FAIMS and SIFT enabled the identification of protein mutations that could not be detected by conventional LC-MS/MS (such as G13D and G12V

mutations in KRAS from HCT116 and SW620 cell lines, respectively). Similarly, this technique can also be used to identify low-abundance single-nucleotide variants without the need for complex, costly and time-consuming immunoaffinity purification. In Chapter 3, a multi-omics approach combining proteomic and genomic datasets facilitated the detection of pathogenic single-nucleotide variants such as EDC3 P132S and RASSF6 S163P.

Altogether, the work presented in this MSc thesis demonstrated the use and application of FAIMS and SIFT to expand the sensitivity and comprehensiveness of proteomic analyses on hybrid Orbitrap mass spectrometers (e.g. Q-Exactive HF and Orbitrap Exploris480). We anticipate that results obtained through this study will be of practical use to proteomic researchers and will expand the capability of current mass spectrometry instrumentation.

4.2 Perspectives

We anticipate that the implementation of FAIMS and SIFT will expand the breath of proteomic research and will pen up new perspectives in the identification of trace-level proteins. Examples of applications are highlighted below.

4.2.1 Immunopeptidomics

Tumor cells have many genetic mutations, resulting in mutant proteins that are presented on the cell surface by the major histocompatibility complex (MHC). These biomarkers are also called neoantigens[1]. The detection of neoantigens has greatly accelerated the development of individualized immunotherapy for cancer patients, for example, to predict the patient's response to immunotherapy[2, 3], or to design personalized tumor vaccines[4]. However, since these immune peptides are usually present in very low abundance and are non-tryptic peptides, their detection by mass spectrometry is limited. As in the study by Newey A, et al[5], 3 out of 612 non-silent mutations encoding neoantigens were detectable by MS.

In recent years, researchers have applied the LC-FAIMS-MS/MS system to the analysis of immunopeptidomics and have achieved improvements in both qualitative[6] and quantitative[7] analysis. But its coverage depth still needs to be improved. However, as concluded in this dissertation, the combination of FAIMS and SIFT can well increase the proteomic coverage, which benefits from a comprehensive and in-depth analysis of the sample. Such combinations have good potential in immunopeptidomics, providing a good platform for improving the coverage and quantitative analysis of immunopeptidome. Optimization of experimental parameters is essential because sample complexity and peptide population are different from typical proteomes.

4.2.2 Targeted analysis

Targeted analysis is commonly used in proteomics for quantitative analysis that requires high sensitivity and accuracy. Parallel reaction monitoring (PRM) was developed based on the high-resolution and accurate mass measurement of Orbitraps[8]. Typically, in PRM analysis, one or more unique peptides are selected in the inclusion list during the primary scan based on m/z and retention time for subsequent quantification based on ion intensity when comparing to an internal

standard or a calibration curve. FAIMS has potential to provide a better identification/quantitation for low abundance peptides/proteins[9]. A more recent example, optimized PRM analysis based on the LC-FAIMS-MS/MS system on the Orbitrap Exploris480 enabled the validation of 88 positive and 88 negative nasopharyngeal swab samples with sensitivities as high as 98% and 100%[10].

However, there are few basic studies based on FAIMS-PRM analysis strategies. The possibility of using FAIMS and SIFT would improve the sensitivity and accuracy of peptide quantification.

4.2.3 Clinical research

The use of proteome for clinical testing requires a complete set of simple and fast analytical workflows from sample preparation. A well-established strategy that takes only 3 hours to obtain quantitative results for more than 300 proteins, including nearly 50 FDA-approved disease markers was recently reported[11]. In the past decade the rapid development of mass spectrometry technology and its application to clinical cancer research (e.g. breast cancer[12], colorectal cancer[13], etc.) opens up new perspectives in early detection and disease prognosis. Recent studies comparing genomic and proteomic analyses of patient described that the proteome conceals additional information unavailable by genomics. The possibility of using proteogenomics to leverage both datasets provide valuable information that could lead to more effective treatments for patients[14].

Based on the proteogenomics exploration in Chapter 3, the combination of FAIMS and SIFT has been shown to have promising applications in the discovery of pathogenic mutations in colorectal cancer cell lines. The identification of patient somatic mutations could provide a new window for personalized immunotherapy to improve human health.

4.3 References

1. Schumacher, T. N.; Schreiber, R. D. Neoantigens in cancer immunotherapy. *Science* **2015**, *348* (6230), 69-74.
2. Schappert, A.; Schneck, J. P.; Suarez, L.; Oelke, M.; Schütz, C. Soluble MHC class I complexes for targeted immunotherapy. *Life Sci* **2018**, *209*, 255-258.
3. Sun, Z.; Chen, F.; Meng, F.; Wei, J.; Liu, B. MHC class II restricted neoantigen: A promising target in tumor immunotherapy. *Cancer Lett* **2017**, *392*, 17-25.
4. Chen, Z.; Zhang, S.; Han, N.; Jiang, J.; Xu, Y.; Ma, D.; Lu, L.; Guo, X.; Qiu, M.; Huang, Q.; et al. A Neoantigen-Based Peptide Vaccine for Patients With Advanced Pancreatic Cancer Refractory to Standard Treatment. *Front Immunol* **2021**, *12*, 691605.
5. Newey, A.; Griffiths, B.; Michaux, J.; Pak, H. S.; Stevenson, B. J.; Woolston, A.; Semiannikova, M.; Spain, G.; Barber, L. J.; Matthews, N.; et al. Immunopeptidomics of colorectal cancer organoids reveals a sparse HLA class I neoantigen landscape and no increase in neoantigens with interferon or MEK-inhibitor treatment. *J Immunother Cancer* **2019**, *7* (1), 309.
6. Klaeger, S.; Apffel, A.; Clauser, K. R.; Sarkizova, S.; Oliveira, G.; Rachimi, S.; Le, P. M.; Tarren, A.; Chea, V.; Abelin, J. G.; et al. Optimized Liquid and Gas Phase Fractionation Increases HLA-Peptidome Coverage for Primary Cell and Tissue Samples. *Mol Cell Proteomics* **2021**, *20*, 100133.
7. Pfammatter, S.; Bonneil, E.; Lanoix, J.; Vincent, K.; Hardy, M. P.; Courcelles, M.; Perreault, C.; Thibault, P. Extending the Comprehensiveness of Immunopeptidome Analyses Using Isobaric Peptide Labeling. *Anal Chem* **2020**, *92* (13), 9194-9204.
8. Barkovits, K.; Chen, W.; Kohl, M.; Bracht, T. Targeted Protein Quantification Using Parallel Reaction Monitoring (PRM). *Methods Mol Biol* **2021**, *2228*, 145-157.
9. Trujillo, E. A.; Hebert, A. S.; Rivera Vazquez, J. C.; Brademan, D. R.; Tatli, M.; Amador-Noguez, D.; Meyer, J. G.; Coon, J. J. Rapid Targeted Quantitation of Protein Overexpression with Direct Infusion Shotgun Proteome Analysis (DISPA-PRM). *Anal Chem* **2022**, *94* (4), 1965-1973.
10. Renuse, S.; Vanderboom, P. M.; Maus, A. D.; Kemp, J. V.; Gurtner, K. M.; Madugundu, A. K.; Chavan, S.; Peterson, J. A.; Madden, B. J.; Mangalaparthi, K. K.; et al. A mass spectrometry-based targeted assay for detection of SARS-CoV-2 antigen from clinical specimens. *EBioMedicine* **2021**, *69*, 103465.
11. Geyer, P. E.; Kulak, N. A.; Pichler, G.; Holdt, L. M.; Teupser, D.; Mann, M. Plasma Proteome Profiling to Assess Human Health and Disease. *Cell Syst* **2016**, *2* (3), 185-195.
12. Pozniak, Y.; Balint-Lahat, N.; Rudolph, J. D.; Lindskog, C.; Katzir, R.; Avivi, C.; Pontén, F.; Ruppin, E.; Barshack, I.; Geiger, T. System-wide Clinical Proteomics of Breast Cancer Reveals Global Remodeling of Tissue Homeostasis. *Cell Syst* **2016**, *2* (3), 172-184.

13. de Wit, M.; Fijneman, R. J.; Verheul, H. M.; Meijer, G. A.; Jimenez, C. R. Proteomics in colorectal cancer translational research: biomarker discovery for clinical applications. *Clin Biochem* **2013**, *46* (6), 466-479.

14. Zhang, B.; Whiteaker, J. R.; Hoofnagle, A. N.; Baird, G. S.; Rodland, K. D.; Paulovich, A. G. Clinical potential of mass spectrometry-based proteogenomics. *Nat Rev Clin Oncol* **2019**, *16* (4), 256-268.

观复 ***(Guan Fu****: observe the regularity of the cycle of things.)*

书法习练让我的内心感到平静。

《道德经》有云：万物并作，吾以观复。

我们在观测时间万物的同时，也何尝不是被万物所观测和度量呢？唯有致虚极，守静笃，方得从容面对一切！

Calligraphy builds my inner peace.

The ***Tao Te Ching*** says: "Usually everything in the world happens at the same time, and all we can do is observe the regularity of the cycle of things."

However, when we observe all things in time, we are also observed and measured by all things at the same time. Only by keeping the mind in a state of emptiness and tranquility, without being affected, can we face everything calmly!