APPLICATIONS OF CAPILLARY ELECTROPHORESIS AND MASS SPECTROMETRY FOR BIOANALYTICAL CHEMISTRY

by

Emil A. Zaripov

A thesis submitted in partial
fulfillment of the requirements for
Doctorate in Philosophy degree in Biology

Department of Biology
Faculty of Science
University of Ottawa

Declaration of Authorship

I solemnly affirm that this thesis is a unique creation of my own, with the exception of any duly credited references. I am cognizant of the university's policies regarding plagiarism, including the potential disciplinary measures. Any utilization of external sources has been properly acknowledged within the body of my work.

Abstract

This thesis explores the capabilities of capillary electrophoresis (CE) in analyzing various biomolecules for analytical applications. The first research chapter demonstrates the detection, separation and quantification of cannabinoids (small molecules) using Capillary Zone Electrophoresis coupled with UV detection. 14 cannabinoid standards were separated and their migration times and absorption spectra were determined. This was then applied to analyze cannabinoid content in herbal cannabis extracts, detecting 14 cannabinoids in one short run.

The second research chapter presents the hyphenation of CE to mass spectrometry (CE-MS) for cationic metabolome analysis of exosomes derived from breast cancer cells. Exosomes are extracellular vesicles released by cells that contain proteins, lipids, nucleic acids and metabolites indicative of their cell of origin. CE-MS enabled the separation and identification of over 200 cationic metabolites in exosome samples. Variables were observed in carnitine synthesis and purine metabolism pathways between cancerous and non-cancerous exosomes. This demonstrated the potential of CE-MS for biomarker discovery directly from exosomes.

The third research chapter highlights CE applications in analyzing large macromolecules. Specifically, CE was used for the selection of DNA aptamers against the nucleoprotein of SARS CoV 2, the virus causing the COVID-19 pandemic. Two SARS-CoV-2 nucleoprotein binding aptamers were identified and used to develop a Proximity Ligation Assay for sensitive viral protein detection. This assay was further coupled with quantitative PCR to detect both the viral nucleoprotein and viral RNA simultaneously through two fluorescent channels. This dual detection method shows promise for sensitive testing of other viruses.

The last chapter focuses on improvements of instrumentation and experimental methodologies. Since CE-MS has its own limitations and drawbacks, improvements had to be made to achieve experimental goals. This chapter describes the development and manufacturing a novel nano sheath liquid interface, allowing for sensitive detection of various compounds. In addition, the chapter describes sample normalization strategies and metabolite identification pipeline. Flaws of CE-MS such as run-to-run variation, as well as susceptibility to matrix effects, needed to be addressed to ensure reproducibility for complex experiments such as metabolomics. Normalization strategy to total UV signal as well as metabolomics pipeline are described in this chapter.

In summary, this thesis explores diverse applications of capillary electrophoresis in bioanalysis, from small molecule pharmaceutical analysis to cancer biomarker discovery. It demonstrates the potential of CE in addressing real-world bioanalytical problems and solidifies its versatility.

Acknowledgements

I would like to acknowledge Dr. Maxim Berezovski and his lab members, past and present, as well as collaborators, for their efforts in making this research possible. Author contributions and funding acknowledgements for each project are listed in each relevant chapter.

Dedication

For my family.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

The rapid advancement of analytical techniques in the field of bioanalytical chemistry has enabled researchers to tackle increasingly complex problems in areas such as drug discovery, disease diagnostics, and personalized medicine. These analytical methods must be able to accurately identify, quantify, and characterize a wide range of biomolecules - from small molecules to large proteins and nucleic acids - within complex biological matrices. Many traditional analytical techniques, while powerful, can suffer from limitations in terms of sensitivity, selectivity, speed, and sample preparation requirements when applied to these challenging bioanalytical applications.

One analytical technique that has emerged as a versatile tool for bioanalysis is capillary electrophoresis (CE). CE is a high-resolution separation method that can be coupled with a variety of detection systems to enable the analysis of diverse biomolecules. The small sample volumes, high separation efficiencies, and ability to manipulate the separation environment make CE an attractive option for applications where sample availability is limited or where the target analytes require specialized separation conditions.

1.1 Capillary electrophoresis for separation of molecules

Analysis of biomolecules can often be challenging due to the complexity of sample matrices. One way to avoid this is to separate molecules by exploiting one of their properties. For example, liquid chromatography modes such as reverse phase or hydrophilic interaction chromatography (HILIC) exploit the polarity of the molecule [1] while size exclusion chromatography is selective to the molecule's size [2, 3]. This thesis places emphasis on

capillary electrophoresis, a separation technique that enables the molecules to separate based on their charge and size in the presence of an electric field [4].

A CE system is composed of several components: high voltage power supply, autosampler, detector, pressure injection system and optionally cooling loop [Figure 1.1]. As described by its name, it is characterized by the use of a capillary: a narrow-fused silica tube with an inner diameter ranging from 10 to 100 micrometers and a length ranging from a few centimeters to several meters. Fused silica capillaries are coated with light impermeable polyimide on the outside, providing rigidity to the capillary. The chemical characteristics of the capillary are discussed further in section 1.1.1. The capillary is filled with a BGE buffer solution, and the sample is injected into one end of the capillary using an air pump, which applies pressure inside the vial, forcing the liquid inside the capillary due to pressure differences between the inlet and outlet. The ends of the capillary are usually immersed in buffer reservoirs, except for CE-MS hyphenation, where the outlet end is pointed into the MS orifice [5, 6]. An electric field is applied by a high voltage (30 kV) power supply through platinum or iridium electrodes. The application of the electric field initiates separation of charged molecules in the solution, where analytes travel to the detector with different velocities according to their electrophoretic mobility. Small ions or molecules with high charge to size ratios migrate faster than larger or less charged molecules. The migration of molecules is then detected, usually through a narrow window 10 centimeters before the outlet end of the capillary, where polyimide is burned off, allowing UV or visible light to pass through and reach the detector. In standalone CE systems, the usual modes of detection are either UV-VIS absorbance or laser-induced fluorescence (LIF). While this is often enough for simple matrices, it may be insufficient for complex systems; therefore, a detector such as MS is often employed where the mass-to-charge of the analyte is used for orthogonal separation.

1.1.1 Capillary wall chemistry

Capillary wall chemistry is an important aspect of CE that affects the separation performance and selectivity of the system. The capillary wall is usually made of fused silica that has a negatively charged surface due to the presence of silanol groups (-Si-OH) on the surface [7, 8]. As the pKa of silanol is 5.6, the silanol groups can dissociate to form negatively charged silanol ions (Si O-) under basic pH conditions [Figure 1.2 A]. Once the capillary is filled with a buffer solution, a high voltage (typically 10-30 kV) is applied across the capillary to create an electric field that drives the separation of compounds in the sample. Electrophoretic mobility of the analyte, however, is not the only factor affecting migration due to the presence of electroosmotic flow.

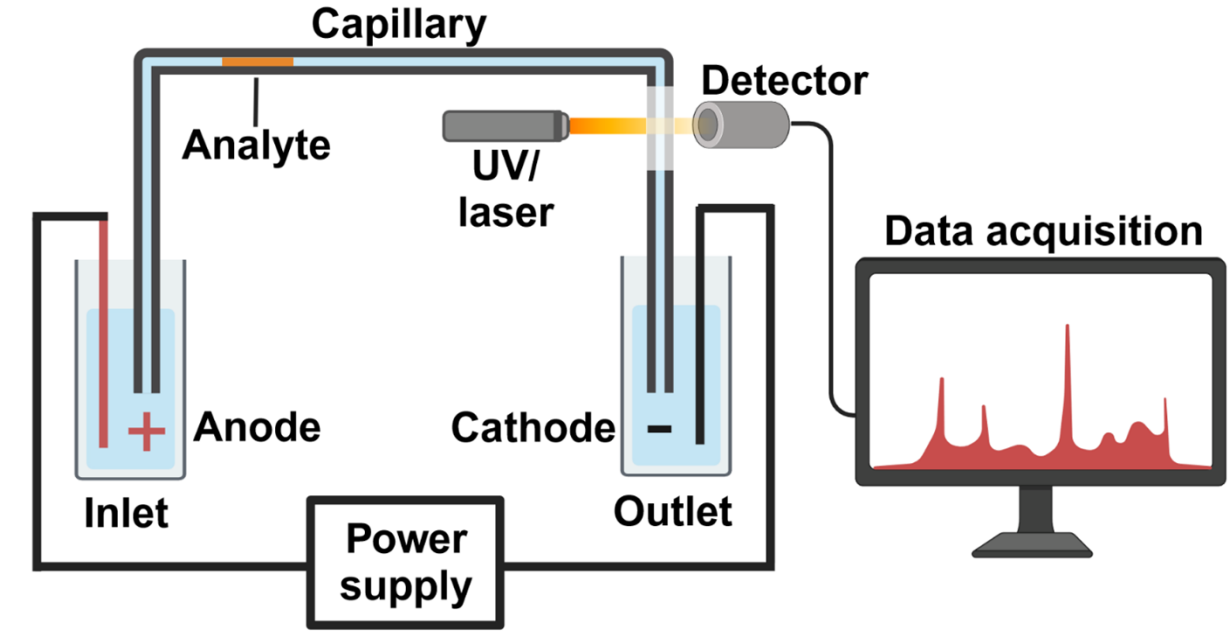

Figure 1.1: Schematic of a capillary electrophoresis (CE) system.

1.1.2 Electroosmotic flow

Electroosmotic flow (EOF) is a phenomenon that occurs in CE when the application of an electric field across the capillary results in the movement of the buffer solution as a whole towards the cathode [9]. This movement is due to the movement of the cations in the buffer along the negatively charged surface of the capillary wall, creating a moving layer of positively charged ions near the capillary wall, which drags the bulk of the solution. The movement of the buffer solution also drags the sample molecules along, resulting in a net movement of the sample towards the detector. This movement allows for the separation of molecules based on size and charge [Figure 1.2 B]. It is important to keep in mind that even cationic analytes can migrate toward a negative cathode if the magnitude of EOF is larger than the electrophoretic velocity of the analyte. In order to maintain EOF, it is vital to keep silanol groups deprotonated to maintain the wall's negative charge. The direction and magnitude of the EOF are influenced by several factors, including the pH and ionic strength of the buffer solution, the surface charge of the capillary wall, and applied electric field strength [9, 10]. The EOF can be controlled and manipulated by adjusting these parameters to optimize the separation conditions for a particular sample or application.

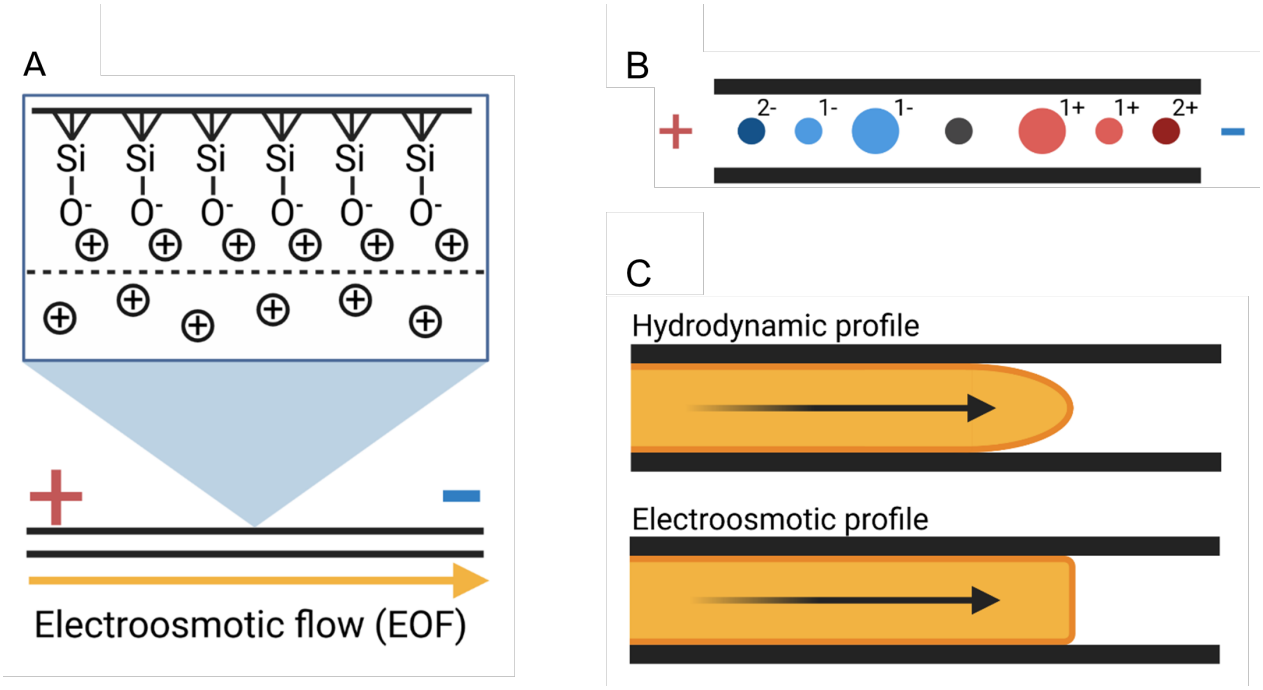

Figure 1.2: Principle of capillary electrophoresis. A. The chemical composition of a capillary wall under basic conditions (pH $>$ 7). B, Migration order taking into account size and charge of the analytes. C. Flat migration profile under electroosmotic flow compared to hyperbolic profile in hydrodynamic flow.

For example, pH and ionic strength of the buffer solution can affect the surface charge of the capillary wall and the mobility of the ions in the buffer solution, which can influence the magnitude and direction of the EOF. By adjusting the pH and ionic strength of the buffer solution, the EOF can be manipulated to optimize the separation efficiency and selectivity. Generally, EOF is a desirable feature of CE because it helps improve the separation efficiency and reproducibility of the system by reducing the diffusion broadening of the analyte bands and minimizing the band broadening due to thermal gradients and turbulence in the buffer solution. The EOF in CE forms a flat-front electroosmotic band profile compared to the parabolic profile of liquid chromatography [Figure 1.2 C] [11], thus reducing band broadening during long separations.

In certain scenarios, EOF is undesirable and must be suppressed. One method is to perform analysis in acidic conditions (pH $<$ 5), where silanol groups are fully protonated and stops the EOF completely, allowing for analytes to separate solely based on their in-

dividual electrophoretic mobilities; however, it is not possible to suppress EOF at high pH in a bare fused silica capillary since the silanol groups are always negatively charged. One method to eliminate or even reverse EOF for wide ranges of pH is to modify the capillary wall by coating it with a polymer or surfactant layer that can neutralize or reverse the surface charge, reducing the magnitude or direction of the EOF [12, 13]. The choice of coating material and the thickness of the layer can be optimized to control the EOF. Several different covalent and non-covalent coatings exist for suppression or reversal of EOF. The most prominent non-covalent coating is the successive multiple ionic-polymer layer, where cationic polybrene and anionic dextran sulfates are applied sequentially, resulting in multiple adsorbed layers on top of the capillary wall [14]. However, these layers do not stop the EOF completely, only changing the direction and increasing the stability over pH ranges of the EOF. To completely stop the EOF, a neutral coating is required. Two common neutral non-covalent coatings include polyethylene glycol (PEG) and polyvinyl alcohol (PVA). These coatings adsorb to the capillary wall and eliminate surface charge, suppressing the EOF [15]. However, with time, they wash out, and the EOF becomes prominent again. To avoid that, covalent coatings are required, which include crosslinked PVA or linear polyacrylamide (LPA) [16–18]. These coatings are covalently attached to silanol groups and are more stable than statically adsorbed coatings. Overall, the control of the EOF in CE is critical to achieving optimal separation efficiency and selectivity, and the parameters mentioned above can be manipulated to control the EOF and improve the separation performance.

1.1.3 Advantages and disadvantages of CE

Capillary electrophoresis (CE) is a powerful analytical technique with several advantages and disadvantages. The most prominent advantage is high separation efficiency. CE can achieve a high theoretical plate number due to the narrow diameter of the capillary, absence of stationary phase interactions, and flat electroosmotic band profile. This enables the separation of complex mixtures with high resolution and sensitivity and allows for the analysis of a wide range of analytes, including small molecules, peptides, proteins, nucleic acids, and even nanoparticles [19–21]. Furthermore, due to the absence of a stationary phase, the analytes mostly interact with the buffer around them, allowing for analyses in almost native conditions. For instance, it is possible to monitor interactions between targets and ligands during CE separation and quantify their kinetics [22]. Additionally, CE separations happen in a short span of time, with certain runs being as short as 5 minutes, which is an advantage for large throughput experiments. CE also offers low sample consumption compared to liquid chromatography (LC). Given the lack of stationary

phase which can trap analytes for elution later, sample injection volume in CE is very low, typically in the range of nanoliters. While this is a definite advantage for precious samples available in tiny volumes, it can also become problematic when large injections of low-concentration analytes are required for detection. Lastly, CE can arguably be cheaper than LC due to the smaller number of consumables required. Capillary itself is an open tube that is cheaper and easier to wash than LC columns with fittings and packing material. The solvent consumption of CE is also smaller since only inlet and outlet vials are required, which can be as low as 0.5 mL, as opposed to large bottles in LC.

Larger injections broaden the sample band, generating broad zones instead of peaks; however, this issue can be tackled with sample stacking, where the analytes are forced to accumulate into the narrow band at the boundary between the injection zone and buffer [23–26]. The small capillary diameter also imposes limitations on detection with UV or LIF since the light path is extremely short as opposed to conventional 1 cm cuvettes, requiring additional means of detection such as mass spectrometry. Stemming from low volumes is the inability of CE to be employed as a preparative technique, with the only exception of DNA that can be later amplified with PCR.

Perhaps the most important disadvantage is the presence of matrix effects and sensitivity to ionic strength. CE is highly sensitive to the ionic strength of the buffer and sample, which introduce variations to the current and as a result, reproducibility. Crude samples often contain buffer components such as salts, which affect the ionic strength of the sample. Higher ionic strength can lead to poor sample stacking and Joule heating, leading to broad peaks, poor separation, and irreproducible migration from run to run [27–29]. Therefore, it is vital to clean up the samples as much as possible using filters, solid phase extractions (SPE) or other means of enrichment.

Overall, CE is a powerful analytical technique but has limitations that must be considered when selecting an appropriate analytical method for the analysis of biological samples. In combination with other analytical tools, such as mass spectrometry, it is highly effective at detection and quantification of molecules even in very complex biological matrices.

1.2 Mass spectrometry

Mass spectrometry is an analytical technique that is used to measure the mass-to-charge (m/z) ratios of ions. In a mass spectrometer, molecules from a sample are ionized and converted to gas-phase ions which are then separated and detected by the mass analyzer based on their mass to-charge ratios. This provides molecular mass information that can

be used to identify unknown compounds or detect specific m/z. Main components of a mass spectrometer include an ionization source (i.e., electrospray ionization or matrix-assisted laser desorption/ionization), a mass analyzer (i.e., quadrupole, ion trap, orbitrap or time-of-flight) that separates ions, and a detector that registers the abundance of each ion. Mass spectrometry finds wide applications in fields like proteomics, metabolomics and drugs development due to its high sensitivity, accuracy and ability to provide molecular weight information. It allows quantitative or qualitative analysis of samples across diverse biological matrices.

1.2.1 Ionization

Ionization is the first step in mass spectrometry during which the analyte molecules are converted into gas phase ions that can be separated by mass analyzers. Different ionization techniques greatly vary in the types of ions that they produce [30]. Conventionally, there is a split between soft and hard ionization techniques. Soft ionization mostly yields intact molecular ions rather than an array of lower mass molecular fragments characteristic for hard ionization methods. As a result, soft ionization methods find wider use in analysis of biomolecules. Commonly used soft ionization methods like electrospray ionization (ESI) gently produce intact gaseous ions for large biomolecules from solution. In ESI, a charged spray is applied to induce droplet formation from which the ions evaporate prior to entering the MS orifice [31]. ESI has become one of the standards for biomolecule analysis in the fields of bottom-up proteomics and metabolomics where intact ions are vital for identification of the compound [32, 33]. Matrix-assisted laser desorption/ionization (MALDI), another soft ionization method, relies on ionization of analytes from a solid organic matrix, resulting in desorption and ionization events upon laser ablation [34].

There are two main modes of ionization: positive ion mode and negative ion mode. In positive ionization mode, the ionization source generates positively charged ions, such as protonated molecules or cations. These ions are attracted to the negatively charged electrode of the mass spectrometer, and they are separated based on their mass-to-charge ratio by the mass analyzer. In negative ionization mode, the ionization source generates negatively charged ions, such as deprotonated molecules or anions. These ions are attracted to the positively charged electrode of the mass spectrometer, and they are separated based on their mass-to-charge ratio by the mass analyzer.

The choice of ionization mode depends on the nature of the analyte and the desired information. Positive ionization is often used for molecules able to accept a proton to become a positively charged ion, such as an amine group containing metabolites, proteins,

or peptides. Negative ionization is often employed for the analysis of organic acids or, in the case of macromolecules, DNA due to the presence of a highly negative backbone. The choice of ionization mode can also affect the sensitivity and selectivity of the analysis, as some analytes may ionize more efficiently in one mode compared to the other.

1.2.2 Electrospray ionization (ESI)

Electrospray ionization (ESI) is a soft ionization technique used in mass spectrometry to produce gas phase ions from analytes in liquid solutions [Figure 1.3]. In ESI, the liquid sample is pumped through a narrow capillary tube with a high voltage applied to the tip. The strong electric field at the capillary tip charges the surface of the liquid emerging from the tip, dispersing it into an aerosol of highly charged droplets [35]. The droplets rapidly shrink in size as the solvent evaporates, increasing the charge density on their surface until they reach the Rayleigh limit. At this point, Coulombic repulsion causes the droplets to explode into smaller droplets. This process repeats until very small, charged droplets are formed, producing desolvated analyte ions in the gas phase which can then be mass analyzed and detected by the mass spectrometer. ESI allows the analysis of thermally labile macromolecules like proteins due to the soft ionization, making it ideal for biological applications [36, 37].

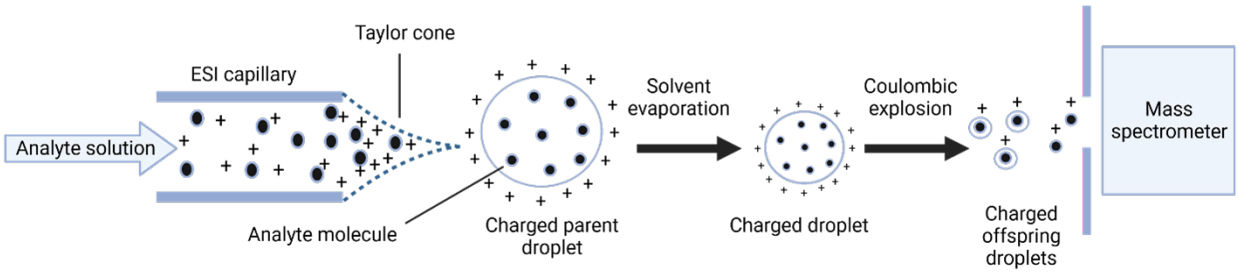

Figure 1.3: Schematic representation of the electrospray ionization process.

ESI can be used as an ionization source for mass spectrometry (MS) when coupled with various separation techniques, such as LC and CE. In LC-MS or CE-MS, the analyte is separated by liquid chromatography and capillary electrophoresis, respectively, and then introduced into the mass spectrometer via ESI. In MS/MS, ESI is used to generate precursor ions that are then fragmented and analyzed for structural elucidation. One advantage of ESI is its ability to analyze large, complex molecules that are difficult to analyze

using other ionization techniques [38]. ESI is also a highly sensitive technique, with limits of detection in the femtomolar range; however, ESI can be sensitive to contamination and requires careful sample preparation [39–41]. In addition, ESI can also be affected by the presence of salts and other non-volatile species in the sample, which can cause signal suppression and interference.

There is another disadvantage of ESI. For example, multi charged ions are produced during ionization which increases the complexity of mass spectra. This leads to a loss of sensitivity due to a spread of signal between differently charged species, and requires deconvolution methods [42, 43]. In addition, due to its nature of donating or removing protons, analytes require functional groups to be ionized, thus restricting the range of compounds. For example, squalene does not contain any basic or acidic functional groups; therefore, it is problematic to analyze it using ESI. Stemming from the functional group requirement, certain compounds may be fragmented during ionization due to a weak chemical bond at the ionization spot. Although ESI is indeed a soft ionization technique, it does not guarantee that all compounds will be ionized in their full molecular form. Lastly, ESI is prone to the formation of adduct ions with solvent molecules, which can complicate spectra instead of just producing $[M+H]^+$ or $[M-H]^-$ ions in case of intact ionization. The signal can be split among $[M+Na]^+$, $[M+K]^+$, $[M+NH4]^+$, etc., reducing sensitivity and ease of identification. In addition, certain molecules with labile bonds such as peptides can undergo neutral losses during ionization, leading to an even more complicated spectrum.

The choice of ionization technique depends on the properties of the analyte molecules, the desired sensitivity and resolution, and the specific application of the mass spectrometry analysis as each technique has its own advantages and disadvantages.

1.2.3 Mass analyzers

Mass analyzers are the component in a mass spectrometer that separate ions based on their mass to charge ratio (m/z). There are several types of mass analyzers that can be used in a mass spectrometer, each with its own advantages and disadvantages. Common mass analyzers used for biomolecule analysis include time-of-flight (TOF), quadrupole, orbitrap and ion trap analyzers.

Time-of-flight (TOF) mass analyzers work on the principle that the time it takes for an ion to travel a certain distance in a uniform electric field is inversely proportional to the square root of its m/z [44, 45]. The acceleration of ions is typically achieved using a high-voltage pulse, and the ions are then directed into a long flight tube. The flight tube is typically under vacuum, which provides a collision-free environment for the ions to

travel through. As the ions travel through the flight tube, they are detected at the end of the tube by a detector system. The detector system is typically a microchannel plate (MCP) detector, which amplifies the ion signal. The time that each ion takes to travel from the source to the detector is measured, and this information is used to calculate the mass-to-charge ratio of the ion. One of the advantages of TOF analyzers is their high mass range, which can extend up to several million Daltons. This makes them well-suited for the analysis of large biomolecules, such as proteins and nucleic acids. In addition, TOF analyzers have fast acquisition times, which can be important for high-throughput applications; however, TOF analyzers can have lower resolving power compared to other mass analyzers, which can limit their ability to distinguish between closely spaced ions. In addition, TOF analyzers can be affected by detector dead time, which is the time required for the detector to recover from the previous ion detection event. This can limit their ability to measure high ion currents and can result in a loss of sensitivity.

A quadrupole mass analyzer uses oscillating electric fields to selectively stabilize or destabilize the trajectories of ions passing through the analyzer [46]. It consists of four parallel metal rods with a direct current voltage and an alternating radio frequency voltage applied between opposite pairs of rods. The applied voltages affect the trajectory of ions traveling down the flight path between the four rods. Only ions of a certain mass-to-charge ratio have stable trajectories and reach the detector at the end of the quadrupole. All other ions have unstable trajectories and collide with the rods, never reaching the detector. The voltage can be varied to allow different mass-to-charge ratios to pass through the quadrupole at different times. This allows the quadrupole to act as a mass filter, scanning across a range of masses to produce a mass spectrum. Quadrupoles provide a fast and sensitive analysis of ions, and the combination of relatively low cost with capability for tandem mass spectrometry makes quadrupole analyzers popular in commercial instruments. It is very common to include quadrupole as a mass analyzer preceding higher resolution analyzer combined in one instrument to make ion isolations possible.

1.2.4 MS parameters

Mass spectrometry is characterized by a variety of parameters that determine its analytical performance: mass accuracy, resolution, sensitivity, dynamic range and mass range. Mass accuracy is the extent to which the measured mass of an ion corresponds to its true mass, thus playing an essential role in identifying unknown compounds [44, 47, 48]. Resolution, which is the ability to differentiate between two adjacent ions, is crucial for accurate mass determination particularly in complex mixtures. The higher the resolution, the greater

the ability to distinguish between ions of similar mass-to-charge ratios. The ability to detect low-abundance ions, or sensitivity, determines the lowest concentration of a substance that can be reliably detected. This parameter is critical in trace analysis and biomarker discovery. Dynamic range is the range within which the instrument can accurately measure ions of different abundances. This parameter allows for the simultaneous detection of both high and low-abundance ions. The wider the dynamic range, the better the performance in quantification tasks. Lastly, mass range dictates the types of molecules that can be analyzed. This parameter is the range of mass-to-charge ratios that the instrument can measure, and a wider mass range allows the analysis of heavier, more complex molecules [49]. Together, these parameters shape the efficacy of mass spectrometry in various applications, from proteomics to metabolomics.

1.2.5 Waters Synapt G2

The Waters Synapt G2 [Figure 1.4] is a hybrid quadrupole time-of-flight instrument suitable for analysis and characterization of biomolecules. It features a nanoelectrospray ionization source capable of soft ionization of small volumes of solutes. The interface contains two separate ionization sources, one emerging from either direct infusion or orthogonal separation technique, and the other, Lock Mass, is used to deliver internal calibrant to adjust calibration mid run. The two channels are separated by a switching baffle, preventing cross detection of ions.

The first mass analyzer of the instrument is a quadrupole, which is used as a mass filter to allow for the isolation of an individual ion. Following the quadrupole, there is an ion mobility (IMS) chamber consisting of three cells: trap, IMS, and transfer cells. IMS cell is filled with nitrogen gas and allows for gas phase separation of ions based on their collision cross section (CCS). Trap or transfer cells are fragmentation cells, where the user can choose whether to fragment ions before or after ion mobility separation, or both. IMS provides additional separation, often increasing sensitivity for co-eluting compounds, ensuring they reach the detector at different times.

In terms of mass analysis, the Synapt G2 couples a quadrupole filter with orthogonal time of flight (TOF) mass spectrometer, offering high transmission efficiency and resolving powers up to 40,000 FWHM (Full width at half maximum) for smaller compounds. TOF detector consists of a pusher, reflectron and detector plate. Pusher changes the direction of ions towards the reflectron, which then slows down ions and reverses their direction towards the detector plate. The instrument has two modes, “V” and “W”, which refer to number of bounces the ion undergoes. In “W” mode, the ion is reflected off the reflectron twice, increasing travel time and yielding higher resolution at the cost of lower sensitivity.

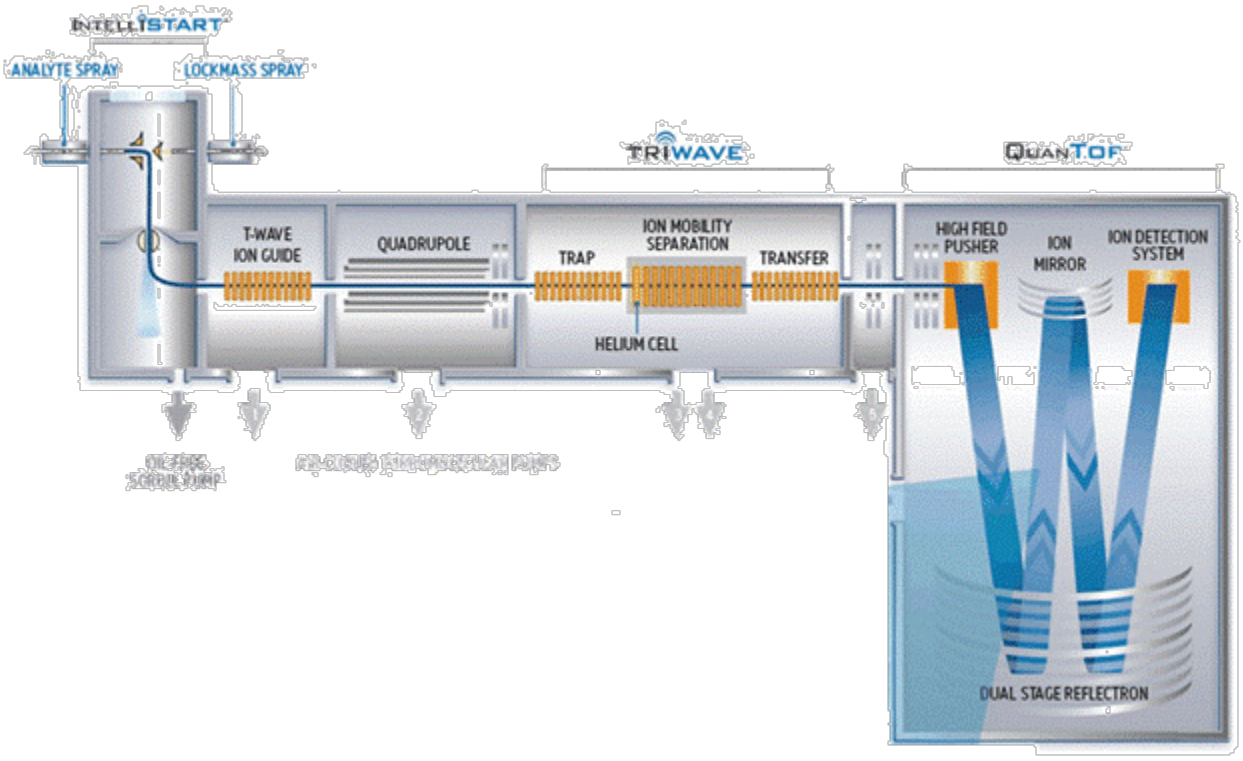

Figure 1.4: Schematic of a Waters Synapt-G2 instrument. Credit: Waters Corporation.

1.2.6 Coupling of Waters Synapt G2 to Beckman PA800 plus

This thesis employs Synapt G2 mass spectrometer coupled to Beckman PA800 plus for separation and detection of biomolecules across a few projects. Coupling of MS to CE [Figure 1.5] was performed using a custom sheath liquid interface, designed and manufactured in-house (Chapter 5). The sheath fluid interface between CE and MS utilizes a liquid junction between the capillary outlet and MS inlet ionization source. A coaxial sheath liquid flow allows for charged liquid to mix with the capillary buffer at the tip of the capillary, right before droplet formation, closing the circuit and allowing for electrophoresis to occur. This liquid sheath interface design effectively decouples the capillary EOF and analyte velocity from the ESI process, preventing analyte diffusion and band broadening at the MS interface compared to sheathless designs. The interface used for this particular instrument allows for nano volumes of sheath liquid (300-500 nL/min), reducing the volumes of solutes being sprayed inside the orifice. This system enables complex tasks for applications such as metabolomics, bottom-up proteomics or intact protein analyses.

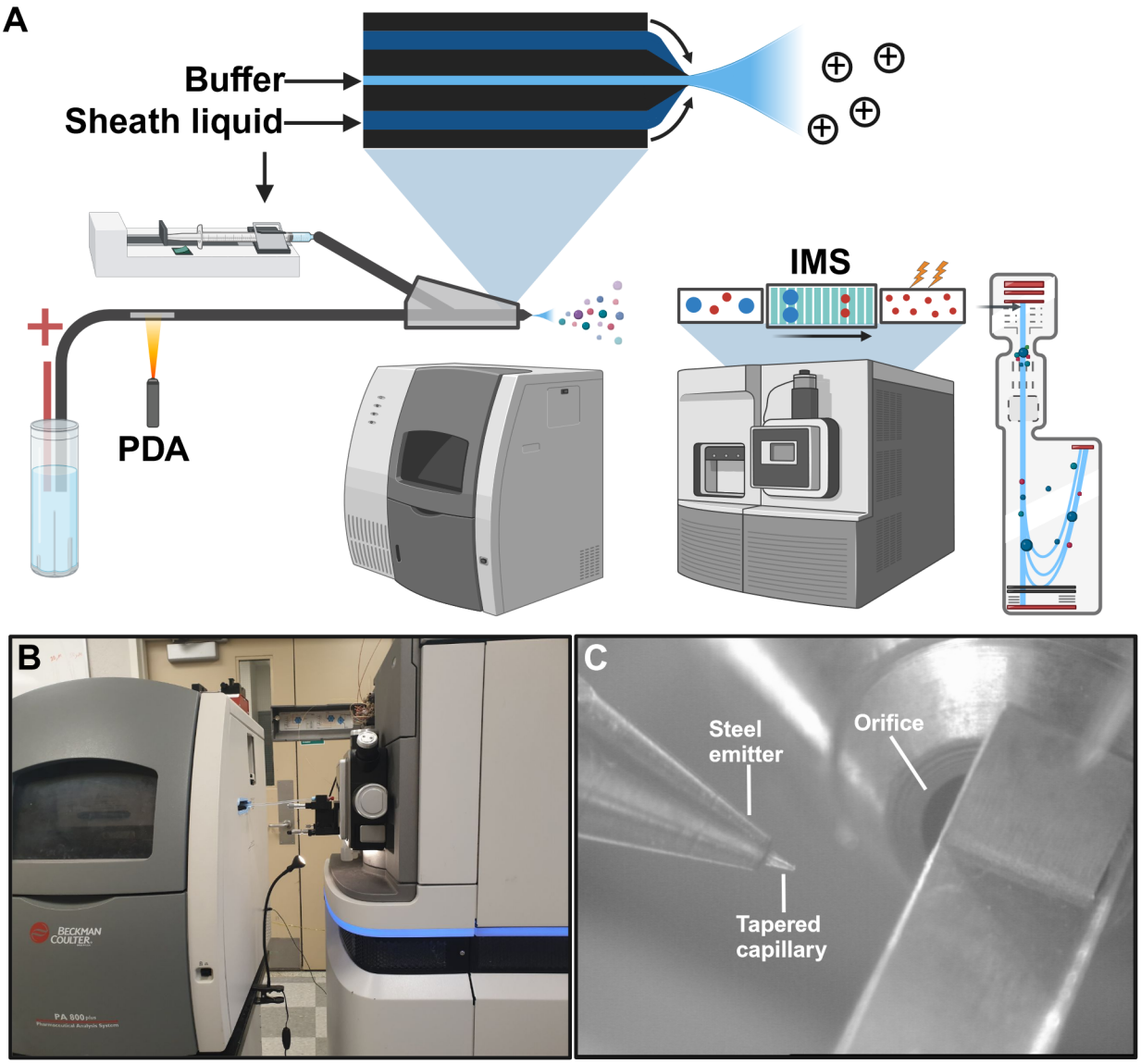

Figure 1.5: Overview of CE-MS system used in this thesis. A. Overall schematic of the CE-MS instrument. B. Photo of hyphenation of CE to MS. C. Photo of a spraying emitter with a tapered capillary end.

1.3 Application of CE-MS

CE-MS is a versatile tool for a variety of analyses, including proteomics and metabolomics. In this work, the human proteome and metabolome were explored, and the analytes of interest were successfully identified and quantified.

1.3.1 Proteomics

Proteomics is the large-scale study of proteomes, which are the entire collection of proteins expressed within a biological matrix. The central goals of proteomics are to identify, characterize, and quantify the proteins present within biological samples in order deduce their functions and interactions [50]. Proteomics studies rely on either top-down or bottom-up approaches [Figure 1.6]. Top-down proteomics involves the analysis of intact proteins and protein isoforms via high-resolution mass spectrometry; however, top-down approach has not reached maturity yet due to issues with sensitivity and signal deconvolution of multi-charged proteins. In contrast, bottom-up proteomics utilizes enzymatic protein digestion, commonly with trypsin which cleaves C-terminal to lysine and arginine residues [51, 52]. This produces a complex mixture of peptides for identification. in this work, CE is used to separate peptides before mass analysis. Tandem mass spectrometers such as quadrupole time-of-flight and Orbitrap instruments provide accurate mass measurements and fragmentation patterns of peptide precursor and product ions via collision induced dissociation (CID) or higher-energy collisional dissociation. Database search algorithms like Mascot, Sequest, and X!Tandem analyze the fragmentation spectra against a protein database to identify peptide and protein matches, while MaxQuant/Perseus and Fragpipe software packages employ quantitative algorithms, normalization methods and statistical analyses to identify differentially expressed proteins between experimental groups or conditions [53].

1.3.2 Metabolomics

Metabolomics encompasses the systematic study of small molecule metabolites using techniques such as nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry, but here we will focus specifically on mass spectrometry-based metabolomics approaches [54,55]. Unlike proteomics and peptides, metabolomics includes a very wide range of compounds with diametrically different properties, making analysis of total metabolome in one run difficult if not impossible with current technology [56]. A key first step in the

metabolomics workflow is sample preparation. Common extraction protocols involve protein precipitation, which utilizes organic solvents like methanol to denature proteins while retaining polar metabolites in the supernatant [57]. Alternatively, liquid-liquid extraction employs solvents of differing polarity to partition metabolites between aqueous and organic phases. Solid phase extraction cartridges containing sorbent resins also selectively retain molecules based on properties like polarity, molecular weight cut-offs, or functional groups. The choice of extraction impacts the subset of metabolites captured. Therefore, multi-step or sequential extraction strategies may be applied to maximize metabolic coverage. Extracts are analyzed by techniques such as LC-MS, GC-MS or CE-MS, providing information on masses of molecular ions and their fragments, enabling identifications using spectral databases and software [Figure 1.6].

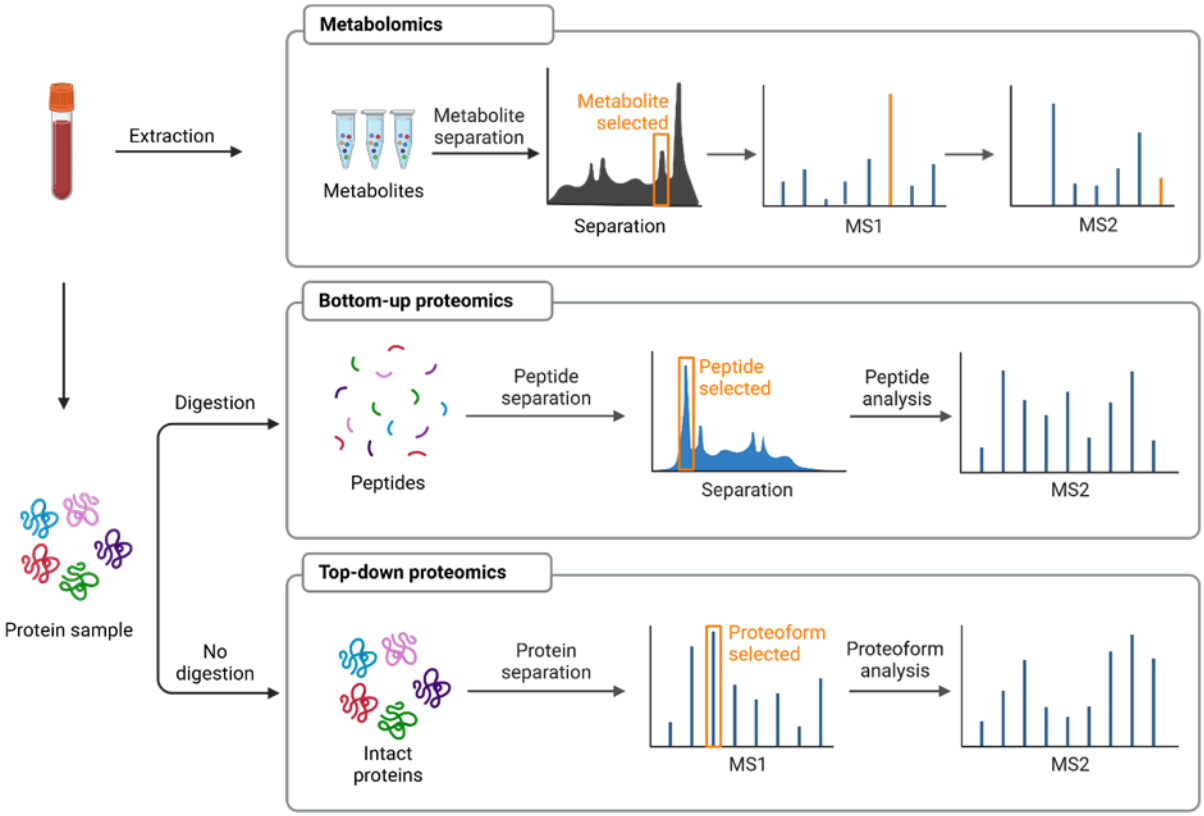

Figure 1.6: Schematics of proteomics and metabolomics workflows.

Software tools such as MSDial and mzMine aid raw data processing including feature detection and deconvolution, chromatogram/electropherogram alignment, normalization, and annotation against libraries like KEGG, HMDB or user-build library for a particular matrix [58]. Multivariate statistical analyses integrated with pathway databases characterize underlying biochemical networks. Method validations involve quality controls, standard compound spike-ins and matching fragmentation patterns to the databases.

1.4 Aptamers

Aptamers are short single-stranded synthetic oligonucleotides, usually composed of either DNA or RNA, that can bind to specific target molecules with high affinity and specificity [59]. Aptamers are explored in greater detail in chapter 4 of this thesis. DNA aptamers have emerged as promising recognition elements for a wide range of applications due to their unique characteristics. They are produced through an iterative in vitro selection process called SELEX (Systematic Evolution of Ligands by EXponential enrichment), which allows isolation of aptamers that bind to specified targets. The SELEX process starts with a large library of random sequence oligonucleotides [60, 61]. The library is incubated with the target molecule, then sequences that do not bind the target are removed while binders are recovered, amplified by polymerase chain reaction (PCR), and taken through additional iterative binding and amplification cycles. SELEX can be performed using a variety of techniques, including CE, where DNA is partitioned between free DNA and DNA-target complex inside the capillary, allowing for collection and amplification downstream.

The selected aptamers take on unique three-dimensional conformations that allow them to bind tightly and specifically to their targets [Figure 1.7]. Compared to antibodies and other protein-based recognition elements, DNA aptamers offer advantages such as smaller size, ease of modification, lack of immunogenicity, fast and inexpensive production, and excellent stability [62,63]. Their applications cover diverse fields including biosensors, drug delivery, imaging agents, and affinity purification. The high specificity and tight binding of DNA aptamers for targets including small molecules, proteins, cells, and tissues make them extremely useful as recognition elements for analytical, diagnostic, and therapeutic applications.

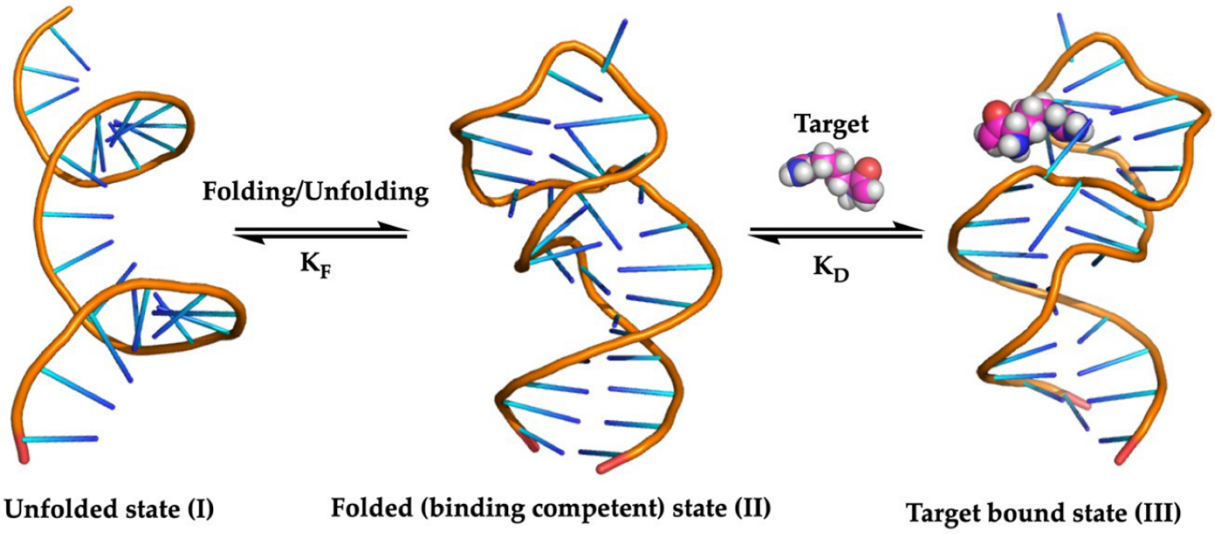

Figure 1.7: Visual representation of aptamer's 3D structure and its target binding. Reprinted from Elskens et Al. 2020 [64].

1.5 Thesis research objectives

In this thesis, small molecules and macromolecules in biological matrices are analyzed using analytical chemistry techniques. The introduction, Chapter 1, highlights the concepts of CE, MS, their coupling (CE-MS), and applications. Chapters 2, 3, and 4 are dedicated to research projects that were defined and achieved using one or a combination of the techniques mentioned above. Chapter 5 focuses on improvements of methodologies and instruments that were employed for the research chapters (2, 3 and 4).

Chapter 2 describes the CE analysis of small molecules (cannabinoids) with UV/PDA detection. The objective of this chapter is the exploration and solution of the challenges associated with the analysis of small molecules in biological samples. Specifically, it focuses on the development of a separation method for 14 cannabinoids, which are structurally similar molecules, using CE and detecting them with a UV detector. Upon successful separation of the standards, the method is validated using cannabinoids isolated from Cannabis sativa, a complex plant material. The research objective focuses on extraction, separation, and quantification, as well as the construction of calibration curves and calculation of LOD, LOQ, and precision.

In extension to the small molecule research, Chapter 3 describes CE-MS analysis of the human metabolome. Here, there is an emphasis on mass spectrometry as a detection method. Specifically, it focuses on the cationic polar metabolome of breast cancer cells and extracellular vesicles secreted by the cells. Most challenges come from the complexity of the matrices that need to be extracted in an optimal and reproducible manner, as well as the compound discovery and identification, since the study is purely untargeted. Mass spectrometry is the key to the untargeted identifications of this objective and is thus the primary focus. During this chapter, extracellular vesicles are successfully isolated, and they are validated using a proteomics approach where commonly known EV markers are discovered in the sample. Following the isolation, the polar metabolome is extracted from the EV samples and analyzed using CE-MS. Lastly, bioinformatic analysis is performed, and differentially present compounds between the cell lines are identified, as well as the integration of metabolomics and proteomics data into KEGG pathways.

Chapter 4 focuses on interactions of large macromolecules using CE and CE-MS analysis. Specifically, it works on the DNA aptamer interaction with the Nucleoprotein (N) of SARS-CoV-2 through the prism of CE. Emphasis on CE is critical since aptamer generation, as well as the epitope identifications, are all performed using either CE or CE-MS. Aptamers against the N protein of SARS-CoV-2 are successfully selected using CE and validated with biolayer interferometry. Using two of these clones, a proximity ligation assay is developed using these aptamers, where two aptamers located on one molecule of N protein can be ligated together, and this ligation product is detected using probe qPCR. Both the N protein and N RNA are detected in 2 different fluorescent channels using qPCR in a single reaction tube. This method is then successfully applied to a spiked saliva sample as an example of a complex matrix. In conclusion of this chapter, the epitopes for these 2 aptamers are identified using CE-MS, and a molecular docking model is proposed.

In chapter 5, the drawbacks of the instrumentation and methodologies are explored, capitalizing on implementing improvements to their design. Specifically, a novel nano flow interface for the hyphenation of CE with MS is created, allowing for higher sensitivity and reproducibility. It also takes advantage of the sequential connection of a UV detector prior to the MS interface, allowing for the detection of the same molecules in 2 different ways during one run, as well as sample normalization. In addition, minor improvements to sample preparation and analysis are also discussed, where a pipeline for CE-MS based metabolomics is presented.

Chapter 2

Single-Run Separation and Quantification of 14 Cannabinoids Using Capillary Electrophoresis

The chapter was originally published in *Separations* [65] in 2021. Authors: Emil A. Zaripov, Tiah Lee, Yuchu Dou, Cory S. Harris, Artem Egorov and Maxim V. Berezovski.

I was selected to deliver a spotlight talk at "CE in the Biotechnology and Pharmaceutical Industries 2023: Symposium on the Practical Applications for the Analysis of Proteins, Nucleotides and Small Molecules", Philadelphia, PA., Oct 2023.

2.1 Abstract

Quantification of major cannabinoids in cannabis products is normally performed using high-pressure liquid chromatography (HPLC)-based methods. We propose a cost-effective alternative method that successfully separates and quantifies 14 cannabinoids in a single run using capillary electrophoresis (CE) coupled with a UV detector in 18 min. The separation is carried out in 60% acetonitrile in the presence of 6.5 mM sodium hydroxide and 25 μM β-cyclodextrin, resulting in good separation of cannabinoids. Our CE method demonstrated the limit of detection between 1.2–1.8 μg/mL , with the linear range reaching up to 50 μg/mL. We validated the method performance by testing a plant extract and quantifying cannabinoid content. This method is the first to separate 14 cannabinoids in one run using a CE system with UV detection.

2.2 Introduction

Cannabis sativa L. is a flowering plant in the family Cannabaceae [66] and its taxonomy has been widely debated, with some scientists considering the plant to be a polytypic genus with three subspecies – *sativa, indica, and ruderalis* – and others considering these taxa to be different species [67]. In addition, the continual crossbreeding of *C. sativa*- and *C. indica*-like plants has led to many perceiving the group as one species, *C. sativa*. Phytochemically, *C. sativa* is recognized for its production of cannabinoids, which are unique to cannabis and underlie its therapeutic potential, including management of anxiety and stress-related symptoms [68], stimulation of appetite [69], pain relief [70], and promotion of sleep [71].

The Cannabis Act (Bill C-45) legalized adult access to cannabis in Canada on 17 October 2018 [72]. The bill permits the selling of fresh and dried cannabis, cannabis plants and seeds, and cannabis oil by an authorized entity with subsequent regulation of commercial edible products and concentrates the year following. In terms of the cannabinoid content, the regulation requires quantification of only tetrahydrocannabinol (THC), tetrahydrocannabinolic acid (THCA), cannabidiol (CBD), and cannabidiolic acid (CBDA).

In the USA, cannabis legalization is more nuanced. At the federal level, cannabis is recognized under the Controlled Substances act as a Schedule 1 substance, where it is considered to have a high potential for dependency and no accepted medical use. However, the medical use of cannabis is legalized in 33 states, with 11 states also legalizing it for recreational use [73]. In Europe, recreational cannabis is illegal, although several countries have legalized it for medical use or decriminalized its use, for instance, in The Netherlands. In Italy, for example, certain cannabis products are legal, however, the ratio of THC to CBD must be under 1, and total THC content must be under 0.6% [74]. This highlights the importance of precise quantification.

Phytocannabinoids are products naturally found in C. sativa that share a typical C21 terpenophenolic skeleton [75]. The plant synthesizes cannabinoids as acids, such as THCA, where decarboxylation of the 2-COOH by heat or light leads to the neutral cannabinoids, such as THC [76,77]. This thermal conversion is a process common to all acid-form cannabinoids. To date, 120 cannabinoids have been isolated, which has led to classification by 11 types: (-) Δ9 trans tetrahydrocannabinol (Δ9-THC), (-)-Δ8-trans-tetrahydrocannabinol (Δ8-THC), cannabigerol (CBG), cannabichromene (CBC), cannabidiol (CBD), cannabinodiol (CBND), cannabielsoin (CBE), cannabicyclol (CBL), cannabinol (CBN), cannabitriol (CBT), and the remaining cannabinoid isoform as miscellaneous.

Current analytical methods for quantifying cannabis are mostly based on high-performance liquid chromatography (HPLC) [78]; however, they are costly, especially those that rely on

ultra-high-performance liquid chromatography UHPLC systems. Capillary electrophoresis (CE) is an alternative separation technique that exploits intrinsic charges present on compounds and uses their different electrophoretic mobilities to separate them inside a narrow (in our case, a 75 μm internal diameter) capillary.

Our method extends to the previous work done in this area. A study dating back to 2002 performed a similar CE separation of cannabinoids extracted from hair [79]. The method described separated four cannabinoids (THC, CBD, CBN, THCA) using a very similar running background electrolyte with sodium hydroxide as an electrolyte, however, in a fully non-aqueous environment and with electrochemical detection in contrast to UV absorption. Another study with a similar method separated only THC and CBD in oral fluids; however, they employed LED-induced fluorescence for detection, with 280 nm excitation and 307 nm emission [80]. A different approach was picked in a study from 1998, where the stationary phase was involved in capillary electrochromatography to facilitate the separation of structurally similar cannabinoids [81]. Seven cannabinoids (6 decarboxylated and one acidic) were separated on a capillary packed with 3 μm C18 beads. The stationary phase helps separation due to the same charge on all cannabinoids and their similar polarity. Another approach with the exploitation of the hydrophobic nature of cannabinoids was undertaken in a study from 2012, where micellar electrokinetic chromatography with SDS was used to separate 10 cannabinoids [82]. The study demonstrated the separation of 10 cannabinoids in one run, which was not achieved before that.

Crude flower extracts sometimes yield very low amounts of certain cannabinoids, in contrast to large amounts of THC and sometimes CBD, making low concentrated compounds difficult to analyze with a UV detector. A study demonstrated a potential solution to that by introducing stacking in SDS presence to suppress electroosmotic flow [83]. They reported over a 2000-fold increase in sensitivity, overcoming one of CE's greatest problems, which is low injection volumes. In this work, we employed capillary electrophoresis, which relies on the acetonitrile-based background electrolyte in the presence of β-cyclodextrin (βCD). Cannabinoids are poorly soluble in water, requiring acetonitrile presence in the background electrolyte (BGE) for optimal solubility and separation. Additionally, cannabinoids share very similar ring structures and similar charges, rendering electrokinetic separation difficult to achieve or making separation time excessively long. βCD was also demonstrated to be an enhancer of cannabinoids solubility in aqueous solutions [84].

Accordingly, we added βCD, which brings an orthogonal separation media by transiently interacting with compounds based on their geometry and polarity. Besides, the CE technique is associated with extremely small sample consumption, where one injection in a particular case of this study is only 4 nL. While separation of these 14 compounds is possible in the absence of βCD, its presence shortens the analysis time by $\sim$ 10 min leading

to better resolution due to a lesser extent of longitudinal diffusion. CBDVA and CBGA, in the absence of βCD, migrate much later, making the analysis time very long. A drawback of our method is the low buffering capacity of BGE. Acetonitrile in our electrolyte tends to evaporate fast once the vial has been opened, leading to fast pH changes. Using fresh BGE is mandatory for reproducible data; reusing the same one twice is not recommended, as the pH will be different and data.

We propose a cost-effective, high throughput method using CE to accurately quantify 14 cannabinoids (THC, CBD, THCV, CBDV, CBG, CBN, CBC, and their respective acidic forms, see their structures in Figure 2.1) with detection by a UV detector at 230 nm. We separated deprotonated forms of these cannabinoids in highly basic conditions (pH $\sim$ 12); furthermore, we reduced the analysis time to under 20 min with the aid of β-cyclodextrin.

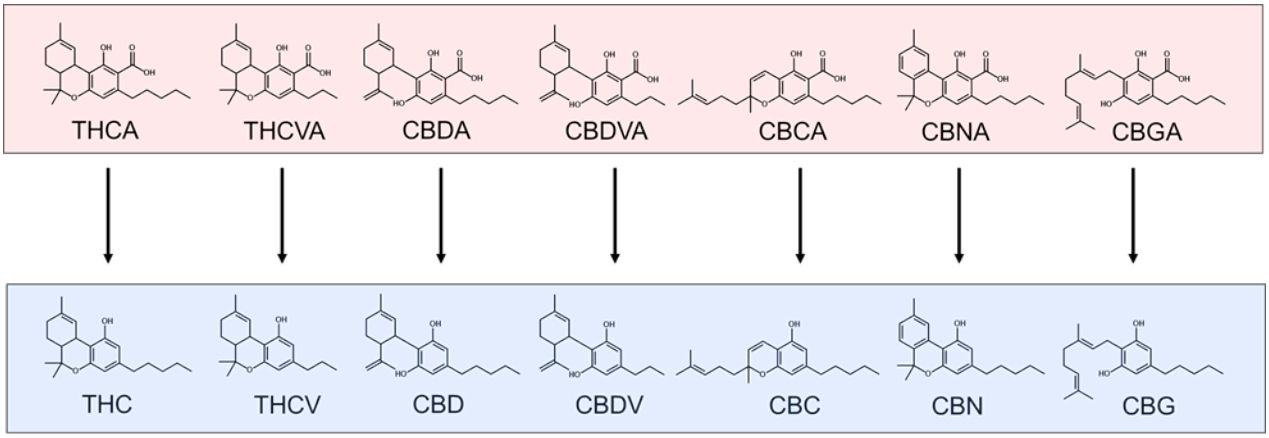

Figure 2.1: Structures of cannabinoids present in the mix. The cannabinoids in red are the acid forms and the cannabinoids in blue are the decarboxylated forms.

2.3 Materials and Methods

2.3.1 Plant Extraction

Lyophilized cannabis flowers were ground up with a hand grinder and extracted immediately. 5 mg of ground cannabis flower material was extracted in 1 mL of methanol (Optima LC/MS Grade methanol, A456212, Fisher Scientific, Ottawa, ON, Canada) for 30 min on a shaker at 200 rpm at room temperature. The extracted plant material was removed by

filtering through a 0.22 μm PTFE membrane syringe filter (Puradisc 25, WHA67842502, Sigma-Aldrich, Oakville, ON, Canada), and the filtrate was stored in an HPLC vial at – 20 °C until further analysis.

2.3.2 Capillary Electrophoresis

New bare silica capillaries (75 μm ID, Molex 1068150019/50M, Fisher Scientific, Ottawa, ON, Canada) were washed and conditioned at 2000 mbar pressure as following: 5 min wash with 1M HCl, followed by 2 min wash with ddH_2O, 10 min wash with 0.1M NaOH, 2 min wash with ddH_2O and 5 min with BGE (6 mM NaOH and 25 μM β-cyclodextrin in 60/40 acetonitrile/water). Samples were injected into the capillary electrophoresis system (Capel 205, Lumex Instruments Canada, Mission, BC, Canada) at 10 mbar for 3 s, forming a 2 mm plug of approximately 4 nL volume. Plug volume was calculated using CE Expert lite web app (Sciex, Concord, ON, Canada). 450 V/cm electric field was applied for 19 min to separate. Cannabinoid standards were diluted in methanol and injected the same way as the samples. Standards were purchased from Cayman Chemicals (Ann Arbor, Michigan, USA; THC, 12068; THCA, ISO60175, CBD, 90080; CBDA, 18090; CBG, 15293; CBGA, 20019; CBC, 26252; CBCA, 30879; CBN, 25495; THCV, 18091; CBDV, 29117; CBGVA, 29787).

2.3.3 HPLC

Extracts were analyzed using an HPLC system (1100 HPLC, Agilent, Santa Clara, CA, USA) coupled with a diode array detector (DAD, series G1315) and an autosampler (series G1313). Chromatography was performed on a Kinetex 2.6 μm C18 150 $\times$ 2.1 mm column (Phenomenex, Torrance, CA, USA) using water + 0.1% TFA (A) and methanol + 0.1% TFA (B) as the mobile phase with a linear gradient from 68% to 85% B in 13 min followed by seven minutes of isocratic conditions (85% B). The flow rate was set at 0.25 mL/min, and the column temperature was maintained at 60 °C with DAD monitoring at a wavelength of 230 nm.

2.3.4 Reproducibility

Precision was determined as a measure of how close the repeated injections are to each other in peak areas. The relative standard deviation was calculated to assess precision

according to the formula: RSD = (SD/x) × 100, where x is the arithmetic mean of all injections. Peak areas were normalized to the respective migration times [85].

2.3.5 Limit of Detection and Quantification

The limit of detection (LOD) was defined as the lowest concentration of an analyte that can still be detected by the instrument. We used a method based on standard deviation (SD) of multiple injections of a standard mix (in our case, seven injections of 5 μg/mL), where LOD = (3.3 × SD)/m, where m is the slope of the linear curve. The limit of quantification (LOQ) was calculated similarly; however, instead of a coefficient of 3.3, a coefficient of 10 was used.

2.4 Results and Discussion

2.4.1 Method

This study aimed to develop a reliable CE method for the separation and quantification of 14 cannabinoids from extracted cannabis samples. CE is a technique that separates compounds based on their charge; therefore, we used highly basic (apparent pH ∼ 12, measured with a pH meter) conditions to deprotonate all 14 standards. Cannabinoids share a highly similar ring structure and exhibit similar migration time in the capillary, causing overlapping peaks if the conditions are inadequate (incorrect pH, for instance).

A method published by Backofen [79] was used as a reference for method development described in this study. They have used a 1:1 mix of methanol and acetonitrile in the presence of 5 mM sodium hydroxide for the separation of 5 cannabinoids. In our case, 6.5 mM sodium hydroxide and high pH alone resulted in the adequate separation of all 14 peaks; however, CBDVA and CBGA were reaching the detector after 30 min of separation, making the analysis time excessive. We have also tested various concentrations of sodium dodecyl sulphate (SDS) in BGE ranging from fully aqueous to aqueous with organic modifiers, as well as fully non-aqueous BGEs; however, most additives resulted in a loss of separation of these 14 compounds, with some of them inevitably comigrating and not properly resolving. We have also tested various concentrations of sodium hydroxide; concentrations lower than 6.5 mM caused a lack of separation between CBG and THC, which started to comigrate. Concentrations of sodium hydroxide higher than 6.5 mM preserved that separation, but

the loss of separation between CBN and CBC was observed. Also, higher concentrations of sodium hydroxide extended the analysis time even further.

βCD, normally used as a chiral selector in CE applications, demonstrated very positive effects on reducing the analysis time without loss of separation. Since it is an inclusion complex with a hydrophobic core, the cannabinoids would transition between βCD and BGE throughout the whole run. As a result, cannabinoids that migrate late, around 30 min, would be "carried" by βCD and migrate slightly earlier. Indeed, with βCD added to the background electrolyte, all cannabinoids still showed great separation in a much shorter time (around 10 min less), except for THC and CBG, although partially resolved, had an overlap close to the baseline. CBDVA and CBGA, which migrated after 30 min of analysis time, in the presence of βCD, started to migrate around 14 and 17 min, respectively [Figure 2.2]. Fully non-aqueous electrolyte was not feasible since βCD was not soluble in the absence of water [86].

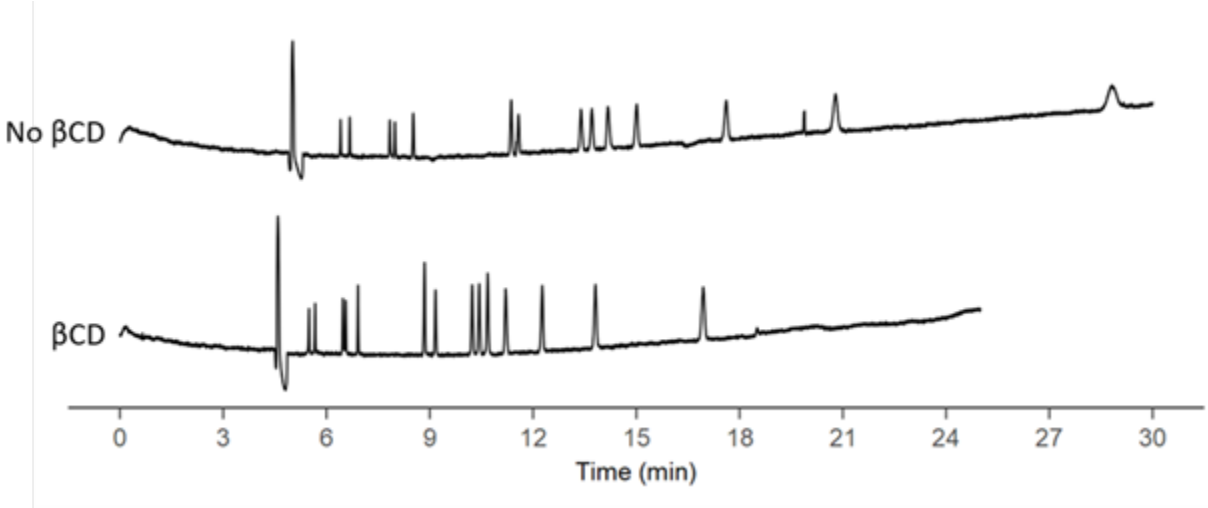

Figure 2.2: Electropherograms of 5 μg/mL standard mix separated in 6 mM NaOH in 60/40 acetonitrile/water with and without addition of 25 μM β-cyclodextrin.

We found that a combination of βCD and highly basic conditions resulted in a reproducible separation of all 14 compounds [Figure 2.3] in a 60 cm capillary, enabling further quantification. Only two compounds (CBG and THC) were not resolved to a baseline and had a resolution of Rs = 1; all other peaks were well resolved to a baseline. While feasible for separation of most compounds, shorter capillary length results in lower resolution of CBG and THC and loss of separation.

Individual standards demonstrated high migration time reproducibility compared to the more complex mixture [Figure 2.3], allowing for identifying each peak. All decarboxylated cannabinoids (CBD, CBDV, CBG, THC, THCV, CBN, CBC, in this order) tend to migrate before the acidic forms (CBCA, THCA, CBNA, THCVA, CBDA, CBDVA, CBGA, in this order), which can be attributed to the higher overall negative charge on acidic cannabinoids due to their carboxyl group, and longer migration time as a result.

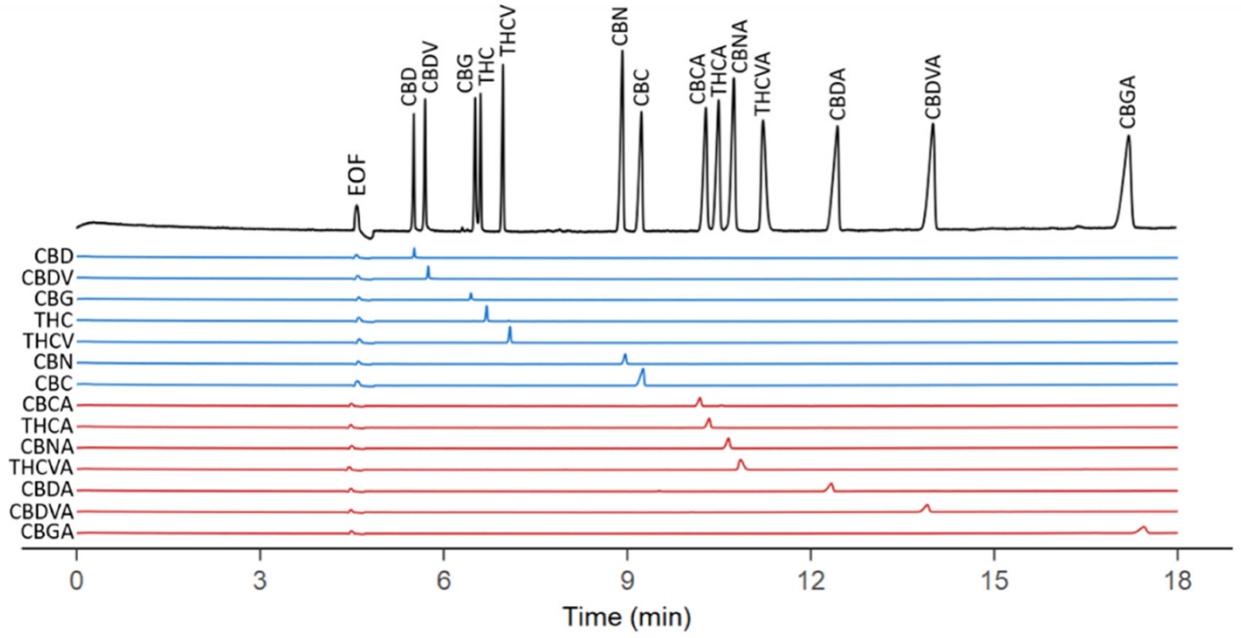

Figure 2.3: CE electropherograms of 14 cannabinoids in a standard mix and individually. 50 μg/mL of compounds either individually or in a mix were injected and detected with a UV detector at 230 nm. Background electrolyte (BGE): 6mM NaOH and 25 μM β cyclodextrin in 60/40 acetonitrile/water. 450 V/cm field was applied to the capillary for 19 min. The first small peak present in all injections represents the injection plug reaching the detector via electroosmotic flow. Blue electropherograms represent decarboxylated forms of cannabinoids, while red electropherograms represent carbonylated forms. EOF represents and injection plug reaching the detector.

We performed HPLC analysis to compare with our CE method [Figure 2.4], which provided comparable separation over a similar run time. However, some compounds were not fully resolved with this method (THCV, CBD, CBG; CBGA and CBN; THC and THCVA). Cannabinoids share highly similar structure with similar polarity, making them

difficult to separate on C18 stationary phase. Reducing the column's particle size and increasing pressure may result in better separation and a more efficient method than the one we present in the paper; besides, more expensive UHPLC systems can be used to further improve the separation and analysis time [78].

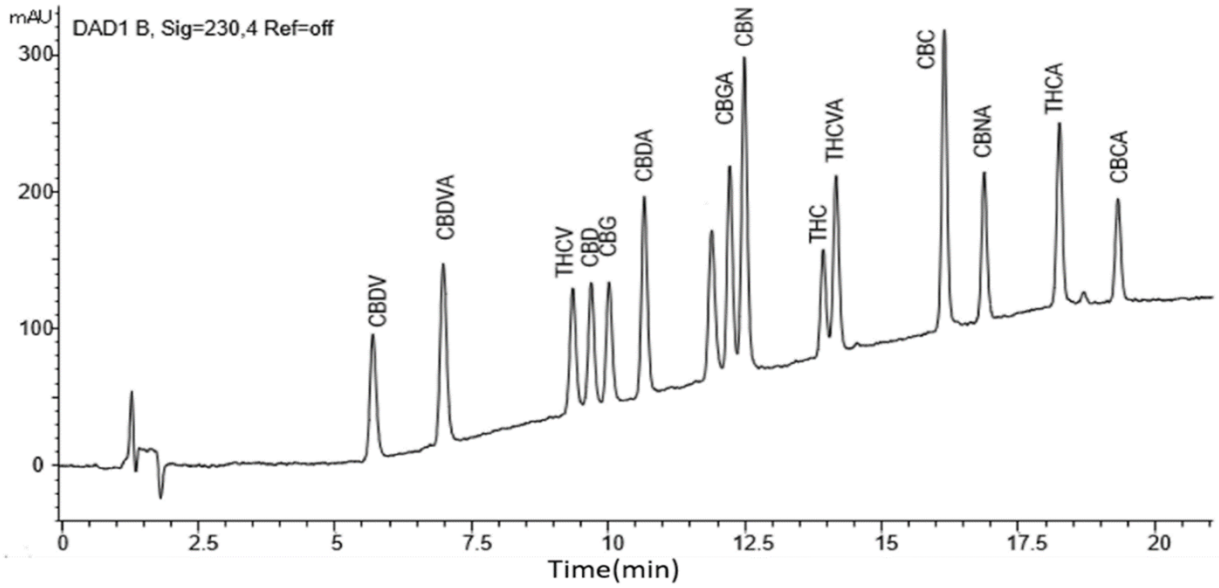

Figure 2.4: A representative HPLC chromatogram of a 14-cannabinoid standard mix. Fifty μg/mL mix was injected and detected with a DAD detector at 230 nm. Chromatography was performed using water + 0.1% TFA (A) and methanol + 0.1% TFA (B) as the mobile phase with a linear gradient from 68% to 85% B in 13 min followed by seven minutes of isocratic conditions (85% B). The flow rate was set at 0.25 mL/min, and the column temperature was maintained at 60 °C.

2.4.2 Quantitation of cannabinoids in samples

We demonstrated the applicability of the CE technique for cannabinoids separation in a mixture of synthetic standards. Extracts of a real cannabis flower are rich in THCA and CBDA compared to other classes of cannabinoids. This situation sometimes leads to results where the peaks representing THC, THCA, CBD, and CBDA are magnitudes larger than anything else present in the sample [Figure 2.5]. In some cases, these peaks may overlap adjacent peaks representing low abundant cannabinoids, especially in the case of

THCA, which can overlap CBCA and CBNA. However, dilution of the sample may result in loss of detection of low abundant cannabinoids. To solve the problem, two injections may be necessary for the detection and quantification of all 14 compounds: injection of concentrated sample and injection of its 10-fold dilution. In the case of our sample, this was not necessary. Additionally, concentrating the sample using solid phase extraction or simple evaporation under vacuum may assist in detection of low abundant cannabinoids, however, one must be aware of abundant peaks overlapping adjacent smaller peaks.

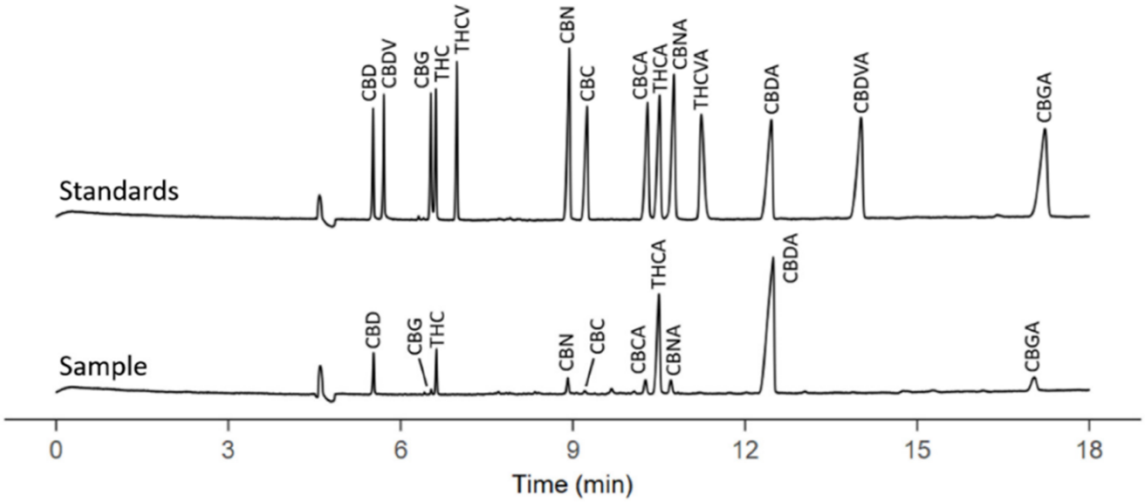

Figure 2.5: Electropherograms of a cannabis extract and standard mix detected at 230 nm. Both cannabis extract and standard mix were injected from methanol; 50 μg/mL mix was injected for migration reference standard and detected at 230 nm. BGE: 6 mM NaOH and 25 μM β-cyclodextrin in 60/40 acetonitrile/water. 450 V/cm field was applied to the capillary for 19 min.

For all our calibration curves, we injected each concentration 3 times, except for 5 μg/mL, which we injected 7 times for accurate limits estimations [Figure 2.6]. The CE method demonstrated the limit of detection between 1.2–1.8 μg/mL, the limit of quantitation between 3.7 5.4 μg/mL, and a linear range reaching up and exceeding 50 μg/mL [Table 2.1]. Our standard mix showed decent precision at all concentrations except 10 μg/mL, where it had around 30% variation [Figure 2.7].

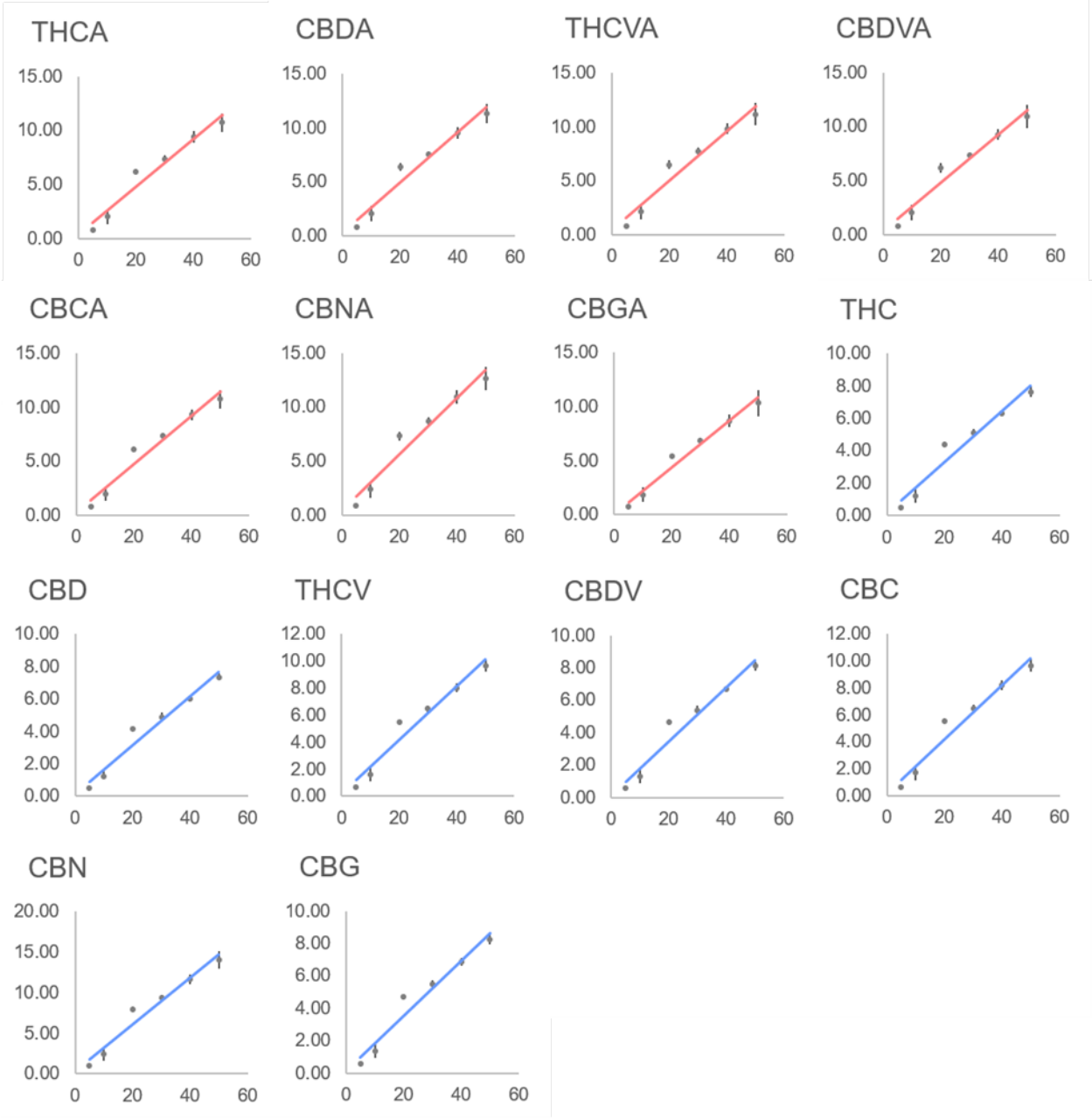

Figure 2.6: Calibration curves of all 14 cannabinoids. Each concentration was injected as a mixture three times (n = 3), with the exception of 5 μg/mL, which was injected seven times (n = 7). Peak areas were normalized to migration times.

Table 2.1: LOD and LOQ values for cannabinoid compounds (n = 7).

Compound Name	LOD (μg/mL)	LOQ (μg/mL)
CBD	1.3	3.8
CBDV	1.2	3.7
CBG	1.5	4.6
THC	1.2	3.7
THCV	1.3	4.0
CBN	1.4	4.1
CBC	1.2	3.7
CBCA	1.4	4.3
THCA	1.3	3.8
CBNA	1.4	4.3
THCVA	1.6	4.8
CBDA	1.4	4.1
CBDVA	1.8	5.4
CBGA	1.7	5.0

All 14 compounds showed highly similar RSD values, with more variation between concentrations rather than between compounds. We observed some loss of precision in the upper ranges, likely due to the higher ionic strength of the mixture injected, greatly reducing sample plug stacking. However, in complex sample mixtures, other compounds may be migrating simultaneously with the compounds of interest, interfering with quantification. We demonstrated that our CE method also provided reliable means of detecting and quantitating cannabinoids in real samples instead of a standard mix. Our extracted sample was analyzed using CE. We quantified all detected cannabinoids in the sample and reported our findings in Figure 2.8 in quantities of each compound per gram of dry flower.

While the standard mix shows all 14 compounds, in our samples, we found no traces of THCV, CBDV, THCVA, and CBDVA. We hypothesize that there may be two underlying reasons for this. First, our samples may be completely devoid of these four compounds. In this case, different strains with the presence of these compounds should be analyzed for their detection. The second reason, which we think to be the case, is the very low abundance of these compounds in the samples. In this case, more sensitive detection methods should be employed. While it is possible to use a wider capillary to increase the UV detection path's length, more sensitive detectors such as mass spectrometers should be employed to detect and quantify these four compounds in the samples.

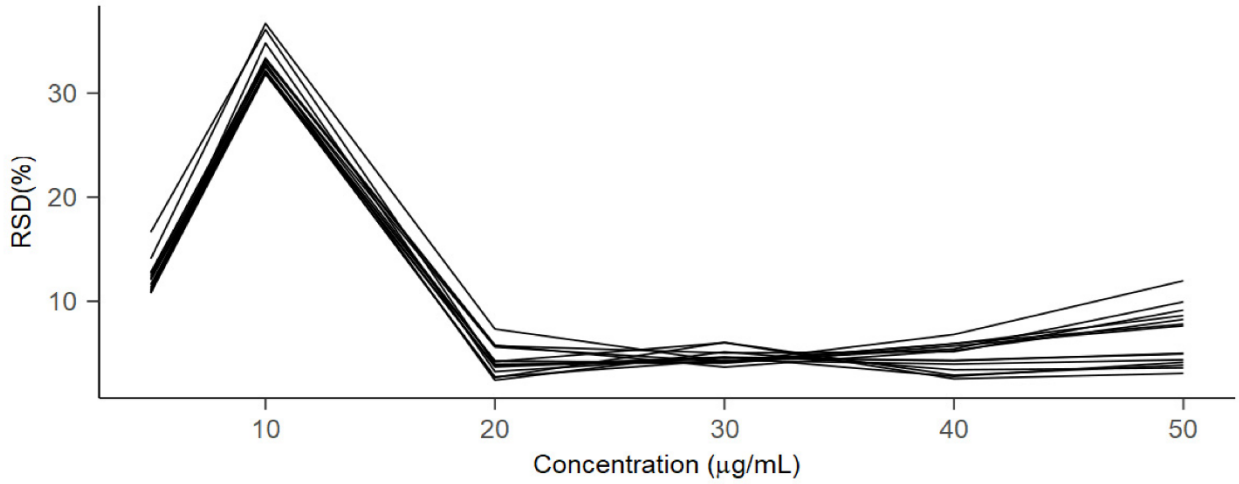

Figure 2.7: Relative standard deviations (RSD) for the peak areas (precision) for each cannabinoid at different concentrations (n = 3). Peak areas were normalized to migration times.

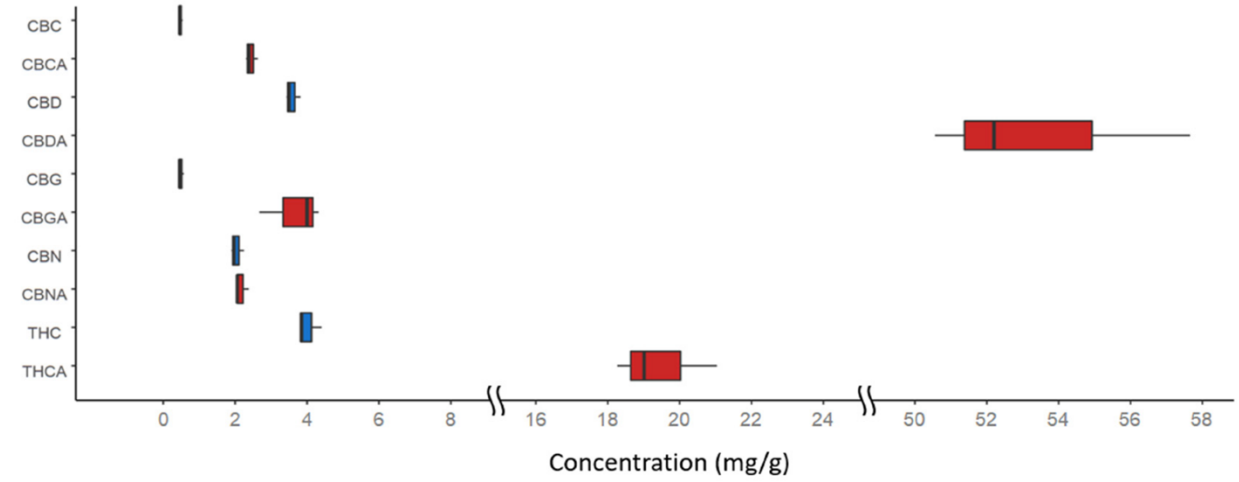

Figure 2.8: Boxplot representations of cannabinoid quantities per one gram of dry cannabis flower (n = 3).

2.5 Conclusions

We propose a fast and cost-efficient method of separation and quantitation of 14 cannabinoids in refined products and crude extracts. CE is a powerful tool in analytical chemistry; however, only a handful of studies employing this technique exist, with fewer cannabinoids detected and separated than the current study. This study covers seven major classes of cannabinoids and their acidic forms, which has only been previously performed on a costly UHPLC system. Thus, we propose CE as an alternative and demonstrate its feasibility. We also perform the first separation of 14 cannabinoids using CE, with prior methods separating 10 cannabinoids only.

2.6 Author Contributions

Conceptualization, M.V.B., E.A.Z., and C.S.H.; Methodology, E.A.Z.; Software, A.E.; Validation, E.A.Z., T.L., and Y.D.; Formal Analysis, E.A.Z., and Y.D.; Investigation, E.A.Z.; Resources, M.V.B., A.E. and C.S.H.; Data Curation, E.A.Z.; Writing—Original Draft Preparation, E.A.Z., and T.L.; Writing—Review & Editing, E.A.Z., and Y.D.; Visualization, E.A.Z.; Supervision, M.V.B.; Project Administration, E.A.Z. and M.V.B.; Funding Acquisition, M.V.B. All authors have read and agreed to the published version of the manuscript.

2.7 Funding

This work was funded by an NSERC Collaborative Research and Development grant (grant number CRDPJ/491434-2016). E.A.Z. was supported by Queen Elizabeth II Graduate Scholarship in Science & Technology and funded by the Government of Ontario (Canada) and Lumex Instruments Canada Inc.

Chapter 3

CE-MS-Based Cationic Metabolomics Reveals Alterations in Purine Metabolism and Carnitine Shuttle Pathway Signatures in MDA MB 231 Breast Cancer Exosomes

3.1 Abstract

A nano sheath-flow capillary electrophoresis mass spectrometry (CE-MS) system with electrospray ionization was used to profile cationic metabolite cargo in exosomes secreted by non-tumorigenic MCF-10A and tumorigenic MDA-MB-231 breast epithelial cells [Figure 3.1]. CE-MS-based metabolomics revealed increased purine synthesis intermediates and increased carnitine synthesis metabolites in MDA-MB-231 exosomes, with pathway enrichment indicating activation of de novo purine pathways and upregulating of carnitine metabolism. In addition, differential expression of ecto-5'-nucleotidase (NT5E) and mitochondrial aldehyde dehydrogenase (ALDH2) demonstrated metabolic alterations in related enzymatic steps. This study demonstrates the application of nano sheath-flow CE-MS for comprehensive and quantitative exosome metabolomics, uncovering metabolic reprogramming in purine and carnitine pathways between normal and cancerous breast cell lines and providing insight into exosome-mediated signaling of breast cancer metabolism.

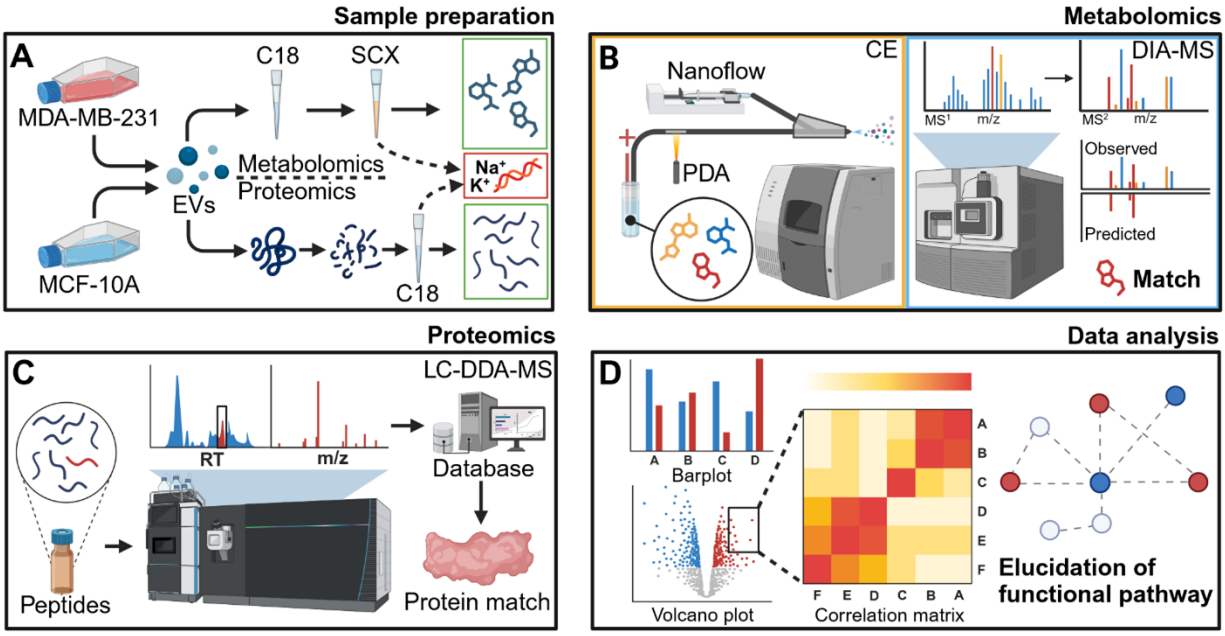

Figure 3.1: Overview of the experimental pipeline for proteomic and metabolomic analysis of EVs. A. Sample preparation procedure using in-house made SCX (strong cation exchange) SPE tip. B. Schematic of instrument setup. PDA data is read online prior to the nano sheath flow interface with MS. MS employs the DIA approach with the IMS function. C. Fragmentation post drift tube assigns equivalent drift time to both parent and product ions, allowing accurate association of both. D. Data analysis workflow of metabolomics data with pathways elucidation.

3.2 Introduction

Extracellular vesicles (EVs) are nanosized lipid vesicles secreted by cells that have emerged as important mediators of intercellular communication. By transferring molecular cargo between cells, EVs can modulate recipient cell function and play central roles in both physiological and pathological processes. EVs have been demonstrated to carry various cargo including proteins, lipids, RNA, and metabolites [87, 88]. As EVs reflect the functional state of their cells of origin, they hold great promise as circulating biomarkers for non-invasive diagnostics and as therapeutic targets [89–92]; however, realizing this potential will require in-depth characterization of EV composition through high-resolution omics approaches.

Metabolomics is the analysis of small molecules or metabolites within biological systems [93] using various analytical chemistry techniques such as mass spectrometry (MS) and nuclear magnetic resonance (NMR) [94]. Metabolites play essential roles in the cellular and biochemical processes that underlie the function of living organisms. The metabolome, which represents the complete set of metabolites in a cell, is the final step in the DNA-transcript-protein pathway, making metabolomics a powerful approach for deciphering complex biological systems [95]. By capturing metabolic phenotypes and flux, metabolomics can reveal insights inaccessible by other omics approaches such as transcriptomics or proteomics due to its focus on small molecules – final member of central dogma. As EVs are demonstrated to be involved in cellular communication, performing untargeted metabolomic profiling on them in an unbiased manner may uncover disease-associated metabolic signatures and novel biomarker candidates. However, the use of metabolomics in the study of EVs faces unique challenges due to the nanoscale size, heterogeneity, and low abundance of EVs. Sensitive analytical techniques such as nano-LC-MS or CE-MS have been successfully employed for low abundance sample matrices before, making them viable for untargeted metabolomic profiling of EVs.

CE-MS has become an invaluable platform for metabolomics research [96–100] due to its high separation efficiency, resolution, mass accuracy and ability to detect charged metabolites at trace levels. In CE, analytes are separated based on charge and size within a narrow bore capillary filled with electrolyte solution. This enables separation of complex biological mixtures into discrete zones detected by MS with high selectivity and sensitivity. The coupling of CE to electrospray ionization MS provides data on mass-to-charge and electrophoretic mobility to enhance metabolite identification and enable comprehensive profiling and analysis of metabolites. A key advantage of CE-MS for EV analysis is extremely low sample consumption, as well as high efficiency of separation of polar compounds, making CE-MS a preferable choice over LC-MS for this study.

In this study, we performed untargeted cationic metabolome profiling of EVs derived from MDA MB 231 breast cancer and MCF-10A normal-like breast epithelial cell lines using CE-MS. Given scarcity of EVs in biological matrices, breast cell lines were selected as a well characterized model for this study. EVs were first isolated from cell culture media by size exclusion chromatography, a gentle separation technique that prevents EV rupture or loss [Figure 3.1 A]. Isolated EVs were validated by proteomics, fluorescent and electron microscopy to confirm enrichment of exosomal markers and morphology. Untargeted metabolomics was performed using CE-MS [Figure 3.1 B], and bottom-up proteomics was performed using conventional LC-MS approach [Figure 3.1 C], allowing for the integration with the obtained metabolomics data, as well as detecting common exosomal markers. Integration of proteomics and metabolomics data of EVs uncovered affected pathways in

cancerous cell lines as opposed to non-cancerous [Figure 3.1 D].

Metabolomic profiling revealed L-carnitine synthesis pathway enriched significantly between the cancer and noncancerous EVs. Pathway analysis was performed by integrating proteomic data to gain a systems view of altered pathways and processes. L-carnitine synthesis pathway was found to be positively correlating with lysine degradation pathway and GABA synthesis, as well as purine metabolism and upregulated adenine in cancer cells. Candidate biomarkers discovered require further validation in larger cohorts, while metabolite trends may uncover new insights into cancer pathogenesis and EV signaling mechanisms. This study highlights the feasibility of our CE-MS workflow for in-depth characterization of the cationic EV metabolome through an untargeted approach. The methodology and cationic profile generated can help elucidate EV biology and metabolic crosstalk in cancer.

3.3 Materials and Methods

3.3.1 Cell culture and conditioned media preparation

The MDA-MB-231 cell line was obtained from ATCC (Cat. # HTB-26). 4 million cells were seeded in a T175 flask into 30 mL of DMEM/F12 media (Thermo Fisher, Cat. # 11330032) supplemented with 10% EV depleted fetal bovine serum (Sigma Aldrich, Cat. # F1051), 100 μg/mL streptomycin, 100 μg/mL penicillin, and 25 ng/mL of Amphotericin B (Thermo Fisher, Cat. #15240096). In order to deplete commercial FBS of bovine EVs, it was diluted 1:1 with DMEM/F12 media, and then ultracentrifugated at 100,000g for 20 hours.

At around 90% cell confluency, the media was aspirated, followed by cell washing with PBS (Thermo Fisher, Cat. #10010023) and replenishment with 30 mL of fresh media. Conditioned media was then collected 72h hours later, spun down at 2000g for 10 minutes and left at 4°C overnight in the presence of 12% PEG-8000. The mixture was then centrifugated at 4000g for 30 minutes, and the pellet was saved for further use.

3.3.2 EV purification and extraction

The pellets were resuspended in 1 mL of 150 mM ammonium acetate, pH $\sim$ 7 and injected into a column packed with Sephacryl S 500 HR (40 cm x 2 cm, Sigma Aldrich, Cat. #S500HR). Separation was carried out in 150 mM ammonium acetate, pH $\sim$ 7 at 1 mL/min

flow rate. Fractions were monitored with UV absorbance at 280 nm, and collected in increments of 2 mL. Fractions corresponding to EV's UV signal were concentrated on 30 kDa MWCO concentrators (Millipore, Cat. # UFC803024) at 4000 g until the total volume was 500 μL. EVs were extracted with 50% acetonitrile at 4 °C for 10 minutes. The extraction mixture was then collected and centrifugated at 21,000g for 20 minutes at – 5 °C, and the supernatant was collected. The supernatant was dried in a vacuum centrifuge overnight to produce dry pellets. The pellets were resuspended in 50 μL of 500 mM formic acid and passed through C18 micro SPE tip. C18 beads were first conditioned by washing them 3 times with 50 μL acetonitrile, followed by 3 washed with 50 μL of 500 mM formic acid. The sample was then passed through the tip and collected. The beads were then flushed 2 more times with 50 μL of 500 mM formic acid, and the flowthrough was combined with the sample.

3.3.3 Cation exchange micro SPE

SPE microextraction tips were manufactured in-house. 2 mm polypropylene frit was installed into a 200 μL pipette tip. Polysulfoethyl A beads (20 μm diameter, 100 Å pore size, (PolyLC, BMSE20-03) were suspended in water at 500 mg/mL. 100 μL of the suspension was added to a fritted tip prepared earlier, and the solution was pushed with a syringe until all the liquid was removed. Traces of water were airdried for 48 h.

Prior to metabolite extraction, the packed bed was conditioned 3 times with 50 μL of 500 mM formic acid, followed by loading of sample. The flowthrough from the C18 cleanup step was loaded onto the column in 3 increments of 50 μL. The packed bed was then washed 3 times with 50 μL of 500 mM formic acid, followed by collected elution with 2 M ammonia in 50% acetonitrile. The collected samples were then dried overnight under vacuum and stored at -80 °C until the injection.

3.3.4 Capillary electrophoresis

Internal electrodes were switched, where inlet was connected to the outlet, and outlet was connected to inlet. All separations were then performed using the outlet tray. New capillary (40 μm I.D., 360 μm O.D., 70 cm length) was cut, and the window was burned 10 cm away from the capillary end. Capillary was then fed through the outlet of the cartridge to the MS source, aligning the window with the PDA probe (Figure 3 1-A).

The new capillary was conditioned with 60 min wash with isopropanol, 60 min wash with 5M ammonium hydroxide and 60 min wash with 1 M formic acid. All separations

were carried out in 0.5 M formic acid at $\sim$ 357 V/cm electric field.

Dry sample was resuspended in 5 μL of 500 mM formic acid. Prior to sample injection, 500 mM ammonium hydroxide was injected at 2 psi for 20 seconds. Sample injection was carried out at 2 psi for 40 seconds. PDA was monitored at 214 nm for metabolites and 195 nm for ammonium injection plug.

3.3.5 Electrospray mass spectrometry

Sheath liquid (30% isopropanol with 10 mM ammonium acetate) was delivered at a rate of 600 nL/min. Source temperature was maintained at 130 °C. The capillary voltage was set to 2.5 kV. The sampling cone voltage was set to 30 V. Extraction cone voltage was set to 5 V. Drying gas, nebulizing gas, and purge gas were all 0 L/h. Waters MSe method was used with 1.5 Hz scan rate and a range of 50-800 *m/z*. Fragmentation was carried out in a transfer cell with a voltage ramp between 5-25 V. Lockmass solution consisting of 200 pg/mL Leucine Enkephalin was delivered perpendicularly at 100 nL/min, collecting 1 scan (1 s/scan) every 5 minutes for *m/z* correction.

3.3.6 Preprocessing

Raw Waters files were converted to mzML for downstream normalization using ROMANCE software. Time conversion to mobility was performed using arginine as a standard with known mobility and that was present in all the samples. mzML files with converted mobilities were then imported into MSDIAL using the all-ion fragmentation (AIF) protocol. Mass tolerance was set to 20 ppm and mobility tolerance was 3 $mm^2kV^{-1}s^{-1}$ (MSDIAL parameter was still RT due to absence of mobility parameter). $[M+H]^{1+}$,$[M+H-H_2O]^{1+}$, $[2M+H]^{1+}$ and $[M+2H]^{2+}$ adducts were included in the analysis. The alignment result was then exported with MS^2 as MSP files and imported into MSFINDER for identification. Identification was restricted to human databases, notably, HMDB, with MS^2 abundance cutoff of 0.1%. Identified compounds were used to build a library for MSDIAL peak matching. An aligned data set with identifications was exported, and the values were then normalized to the area of the UV peak at 280 nm wavelength. Normalized values were then used for all statistical analyses.

3.3.7 Validation with standards

Purified metabolite standards were prepared at 1 μM in 0.5 M formic acid spiked with 1 μM L-arginine prior to injecting. Injection and processing were performed in the same manner as previously described for all samples. Spectral matching was performed against a library samples. Two criteria were used for validation: match of the MS^2 fragments between purified standard and the sample, as well as match of the electrophoretic mobilities. Electrophoretic mobility values were obtained by converting time scale to effective mobilities with ROMANCE algorithm using L-arginine as the internal standard. L-arginine was ubiquitously present in all EV samples as well, allowing for migration time correction even in absence of spiked internal standard.

3.3.8 Sample Preparation for Proteomics Analysis

The sample preparation used the filter-aided sample processing (FASP) protocol. Frozen extracellular vesicle (EV) pellets were resuspended in a 100 μL solubilization buffer (5% SDS, 50 mM TRIS HCl, pH 8) and subsequently transferred to a 30 kDa molecular weight cut-off (MWCO) filter (Microcon® 30K device, Cat No. MRCF0R030, Millipore). Following a 10 minute incubation, 400 μL denaturation buffer (8 M urea, 50 mM TRIS HCl, pH 8) was added, and the mixture was centrifuged for 15 minutes at 14,000 g. Removal of residual SDS was achieved through iterative centrifugation with 400 μL denaturation buffer. Protein reduction was initiated by combining 8 μL of 100 mM TCEP with 200 μL denaturation buffer, followed by a 30 minute incubation at RT and a subsequent centrifugation step of 15 minutes at 14,000 g. Alkylation of reduced proteins involved mixing 8 μL of 500 mM IAA with denaturation buffer, incubating for 45 minutes in the dark at RT, and centrifugation for 15 minutes at 14,000 g. The addition of 200 μL digestion buffer (50 mM TRIS HCl, 0.6% glycerol, pH 8) to the samples was followed by centrifugation for 15 minutes at 14,000 g. Residual urea was eliminated through a repetitive step using 200 μL solubilization buffer. Subsequently, 200 μL digestion buffer containing 600 ng trypsin/Lys-C was added to the samples, and digestion ensued at 37 °C for 12 hours. Peptide collection required centrifugation for 20 minutes at 14,000 g, followed by the addition of 40 μL digestion buffer and a subsequent centrifugation step of 20 minutes at 14,000 g. Acidification of the peptides was achieved using 2 μL of 100% formic acid, followed by desalting via in-house C18 TopTip microcolumns. The C18 microcolumns were conditioned through triple washing with 50 μL 70% acetonitrile containing 0.1% formic acid and subsequently thrice with 50 μL 0.1% formic acid. The samples were then passed through the C18 microcolumns, followed by a triple wash using 0.1% formic acid. Finally, peptide

elution was performed using a 70% acetonitrile solution containing 0.1% formic acid. The purified peptides underwent vacuum drying and were stored at -80 °C.

3.3.9 LC-MS/MS bottom-up proteomics

Protein samples, with a maximum content of 50 μg as determined by the BCA assay, underwent analysis utilizing an Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific) coupled to an UltiMate 3000 nanoRSLC system (Dionex, Thermo Fisher Scientific). Peptides were chromatographically separated on an in-house packed column (Polymicro Technology) with dimensions of 15 cm × 70 μm ID, Luna C18(2), 3 μm, 100 Å (Phenomenex), employing a gradient of water/acetonitrile/0.1% formic acid. The sample injection onto the column occurred over a duration of 105 min at a flow rate of 0.30 μL/min. The peptide separation involved an initial 2% acetonitrile stage for the first 7 min, followed by a linear gradient from 2% to 38% acetonitrile over 70 min. Subsequently, a gradient from 38% to 98% acetonitrile was applied for 9 min, followed by a 10 min stage at 98% acetonitrile. Finally, a gradient from 98% to 2% acetonitrile over 3 min was executed, and a 10 min wash at 2% acetonitrile concluded the chromatographic run. Eluted peptides were introduced into the mass spectrometer using nanoelectrospray ionization (nESI) in positive mode at an ion source temperature of 250 °C and an ion spray voltage of 2.1 kV. The Orbitrap Fusion mass spectrometer operated in the top speed mode, acquiring full-scan MS spectra in the m/z range of 350–2000 at a resolution of 60,000. Precursor ions were selectively filtered based on monoisotopic precursor selection, charge state (+2 to +7), and dynamic exclusion (30 s with a 10 ppm window). Automatic gain control settings were set at 5×10^5 for full FTMS scans and 5×10^3 for MS/MS scans. Peptides were fragmented with CID in the linear ion trap, with precursors isolated using a 2 m/z window and fragmented with a normalized collision energy of 35%.

3.3.10 MS data processing for proteomics

Mass spectrometry (MS) raw files were analyzed using MaxQuant software (version 2.3.0.0) and the Andromeda search engine. Peptide sequences were queried against a human UniProt FASTA file comprising 20,424 entries as of August 30, 2023, along with a default contaminants database. Default parameters were employed unless otherwise specified. Variable modifications included N-terminal acetylation and methionine oxidation, while cysteine carbamidomethylation was set as a fixed modification. Peptides were required to have a minimum length of 7 amino acids, and a false discovery rate (FDR) of 0.01 was

applied at both the protein and peptide levels. FDR was determined by searching against a decoy database of reverse sequences. Enzyme specificity was defined as C-terminal to arginine and lysine for conventional proteomic samples. Up to two missed cleavages were allowed. Peptide identification employed an initial precursor mass deviation of up to 10 ppm and a fragment mass deviation of 0.5 Da. The 'Match between runs' algorithm in MaxQuant was executed across all samples to enhance peptide identification rates. Proteins and peptides corresponding to the reverse database were excluded from further analysis. For label-free quantification, a minimum ratio count of 2 was deemed necessary.

The MaxQuant output tables, "proteinGroups.txt", was imported into the R. Subsequently, proteins identified as potential contaminants and reverse proteins by MaxQuant were deliberately excluded from the ensuing analysis.

3.4 Results

3.4.1 EV isolation and characterization

EVs precipitated overnight in the presence of PEG-8000 were resuspended in ammonium acetate buffer and separated on SEC column with fraction collection [Figure 3.2 A]. First peak corresponding to EVs was collected for all the further analysis [Figure 3.2 B]. To confirm correct size distribution, dynamic light scattering was performed, showing size distribution with the median of 80-90 nm, in accordance with exosomal size reported in literature [Figure 3.2 C]. Both MDA-MB-231 and MCF-10A derived EVs showed nearly identical size distribution.

Isolated fractions were stained with green RNA dye (SYTO RNASelect, S32703) and red membrane dye (BODIPY TR Ceramide, D7540) and visualized on fluorescence microscope [Figure 3.2 D]. Although the resolution was insufficient to visualize the vesicles, fluorescent signals were observed both for membrane dye and RNA dye, indicating presence of both, as is expected for EVs. Lastly, TEM image of vesicles stained with uranyl acetate was obtained to visualize the vesicle and observe the membrane structure.

Furthermore, we aimed to identify exosomal markers in our isolations to confirm exosomal presence. We performed bottom-up proteomic profiling of the collected fractions, and identified 86 out of 100 commonly found exosomal markers as listed on ExoCarta database [Figure 3.2 E]. Among these are CD9 and CD81, commonly used biomarkers for the confirmation of exosomal enrichment. Therefore, we were able to confidently claim exosomal presence in our samples and proceed with further experiments.

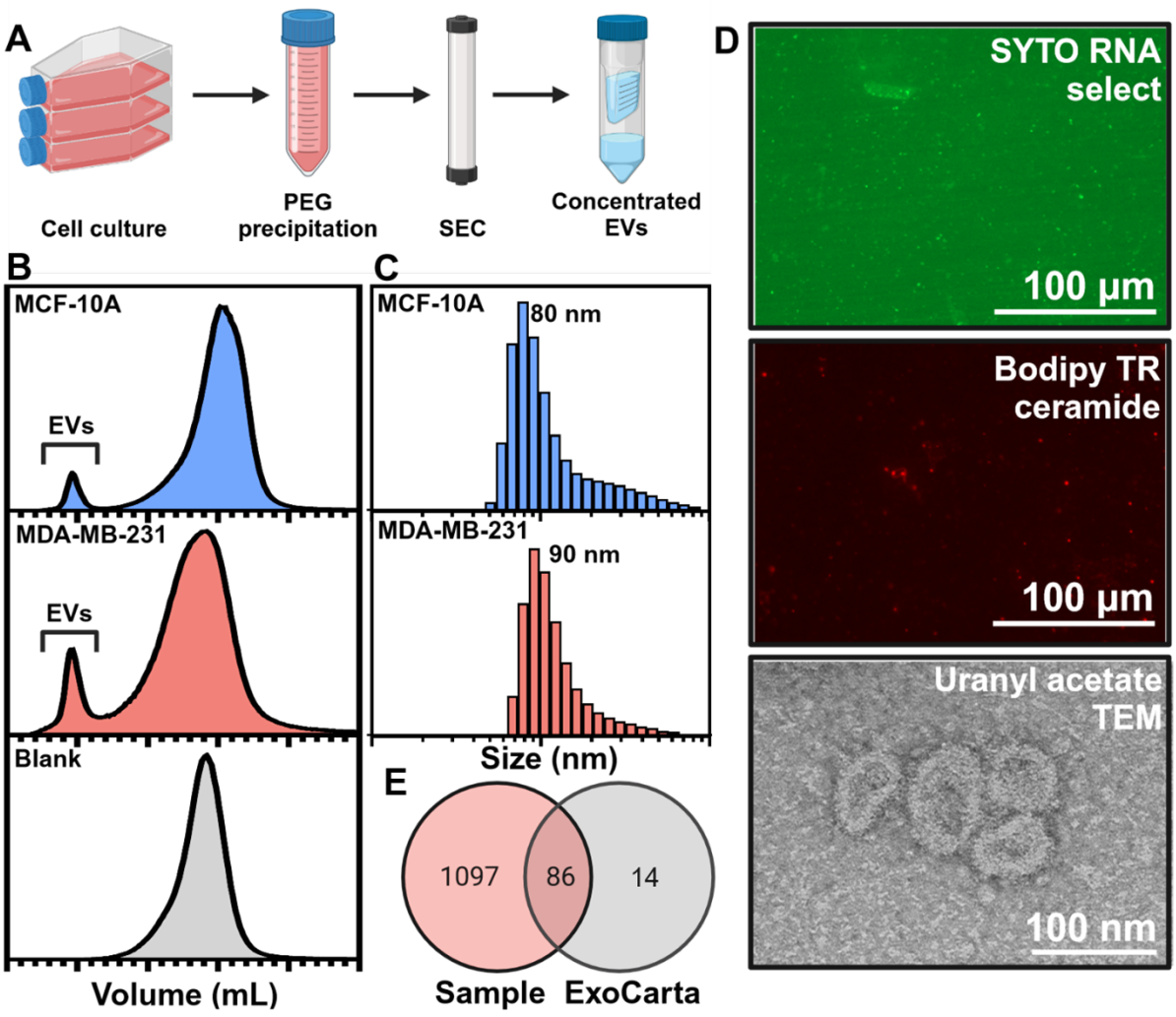

Figure 3.2: Characterization of EVs isolated through size exclusion chromatography. A. Schematic of EV preparation. B. Small peak was detected at A280 and collected in ammonium acetate buffer. C. Dynamic light scattering confirmed exosomal size distribution with median size of 80-90 nm. D. Fluorescent microscopy confirmed presence of RNA and lipid membranes. TEM confirmed the expected exosomal size as well as visualizing EV's membrane. E. Proteomic analysis of small peak revealed 86 exosomal markers in ExoCarta database.

3.4.2 Untargeted identification of enriched candidate molecules

Following the identification of metabolites and normalization of intensities, class enrichment was performed. Three most enriched metabolic classes included amino acids, indoles and purines [Figure 3.3 A]. A Whitney-Mann test was performed to determine the significance threshold for each metabolite. The volcano plot revealed L-carnitine, butyro-betaine, delta-trimethyllysine and propionylcarnitine as the most significant compounds with the highest fold change that are part of the carnitine synthesis pathway in MDA-MB-231 EVs [Figure 3.3 B]. In addition to statistical significance, all these compounds were at least 4 times more present in MDA-MB-231 EVs compared to MCF-10A EVs. Another compound above the threshold was adenine, a part of the purine metabolism pathway. Notably, NT5E protein from the same pathway was also enriched in MDA MB 231 EVs with the highest fold change in the matrix [Figure 3.3 C]. We selected 2 pathways to focus on, carnitine synthesis and purine metabolism pathways due to their high enrichment in MDA-MB-231 EVs.

We then performed a correlation analysis of metabolomics data with hierarchical clustering. Metabolites were first ranked according to the Pearson correlation method, followed by the complete linkage method for hierarchical clustering. Purine metabolism and carnitine synthesis pathways clustered next to each other on the correlation matrix [Figure 3.3 D]. All of the metabolites in these 2 groups showed a high degree of enrichment in MDA-MB-231 as opposed to MCF-10A. We chose to focus on 2 discovered pathways: carnitine synthesis and purine metabolism due to their high degree of correlation, as well as presence of the most significant metabolites within the sample matrix [Figure 3.3 E].

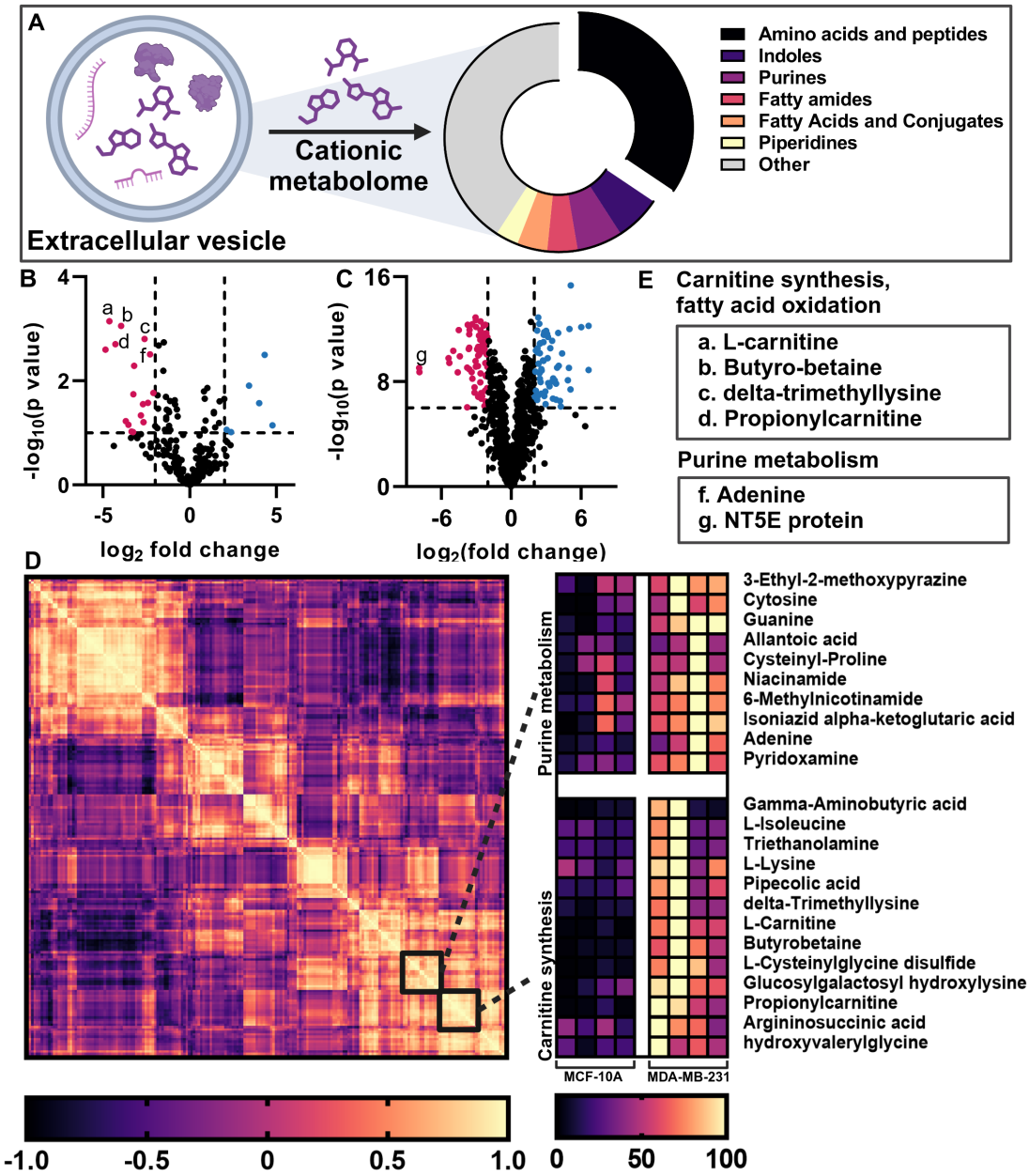

Figure 3.3: Untargeted metabolomics revealed metabolic pathways enriched in MDA-MB-231 breast cancer derived EVs. A. Class distribution of total EV metabolome from both cell lines. B. Volcano plot of metabolites outlining significant compounds with larger presence in MDA-MB-231 EVs. C. Volcano plot of proteins with high presence of NT5E protein in MDA-MB-231. D. Correlation matrix showing hierarchical clustering of metabolites involved in two most enriched KEGG pathways. Heatmaps for each of the clusters depict relative intensities of each of the member's metabolites between cell lines E. Significant metabolites and proteins that show high hold change in MDA-MB-231 vesicles and are part of the enriched pathways.

3.4.3 Metabolite Validation

We validated 3 discovered metabolites using molecular standards. A high degree of certainty was observed for L-carnitine since all MS^2 fragments of the standard were observed for spectrum obtained from the library sample [Figure 3.4 A]. Additionally, mobility of the standard and sample L-carnitine was matched, confirming the identity of the compound.

Adenine was the 2nd validated compound, however, its low ionization efficiency in our system provided noisier spectra compared to L-carnitine. Most MS^2 peaks from the standard were matched to putative adenine in the sample. Standard spectrum demonstrated additional peak at 87.044 *m/z* [Figure 3.4 B]. This peak was attributed to the buffer interference as it was ubiquitously present throughout the entire run. Validation and samples were analyzed on different days, resulting in different background signals. Mobility matching and construction of extracted ion electropherograms revealed presence of 2 peaks with the same *m/z* but drastically different mobilities. Smaller peak with higher mobility corresponded to adenine signal in the standard run, while the larger peak was found to be adenosine. 136.061 *m/z* was observed for adenosine due to its spontaneous fragmentation during ESI process, resulting in fragment that is identical to molecular adenine, hence, 2 peaks were observed. Adenosine was also validated using MS^2 of the standard spectrum. In order to avoid confusion with adenine, mobilities were matched. The low mobility peak for adenine was indeed adenosine, as its mobility matches adenosine standard [Figure 3.4 C].

3.4.4 Two pathways enriched in EVs derived from MDA-MB-231

We focused our analysis on two pathways that showed strong correlations between enriched metabolite levels in exosomes derived from MDA-MB-231 and MCF-10A cells. The carnitine biosynthesis pathway, which involves the degradation of lysine, displayed significant enrichments of three compounds in exosomes from MDA-MB-231 cells [Figure 3.5 A]. While quantitative analysis detected enrichment of the first (delta-trimethyllysine), last (L-carnitine) and second last (Butyro-betaine) metabolites of this pathway, levels of intermediates were below detection limits. However, the enzyme gamma-butyrobetaine dioxygenase (BBOX1) that catalyzes the conversion of butyro-betaine to L-carnitine was not detected in exosomes from either cell line. The sole protein detected from this pathway was aldehyde dehydrogenase 2 (ALDH2), which was downregulated at the protein level in MDA-MB-231 exosomes compared to MCF-10A, in contrast to the upregulated metabolite levels. This indicates divergent regulation of metabolites and proteins in this pathway between normal and cancerous cell lines.

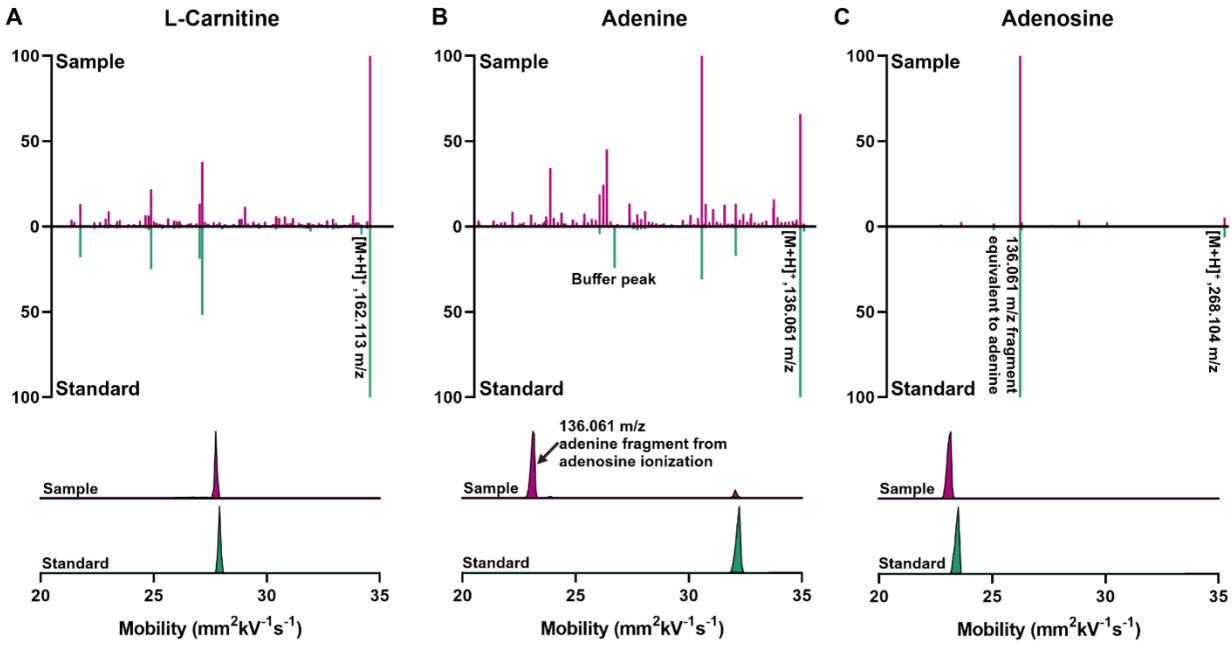

Figure 3.4: Validation of three candidate metabolites. A. MS^2 spectrum and mobility matching between L-carnitine and library sample. B. MS^2 spectrum and mobility matching between adenine and library sample. C. MS^2 spectrum and mobility matching between adenosine and library sample.

The exosomes secreted by MDA-MB-231 also cells showed enrichment of adenine and NT5E, a critical enzyme in the purine synthesis pathway [Figure 3.5 B]. Measurement of adenosine levels within the exosomes found no significant difference between the two cell lines. However, analysis of proteins involved in purine metabolism revealed downregulation of the enzymes PNP and APRT in MDA-MB-231 exosomes. The purine synthesis pathway also showed strong correlation with carnitine synthesis pathway and clustered together on the correlation matrix.

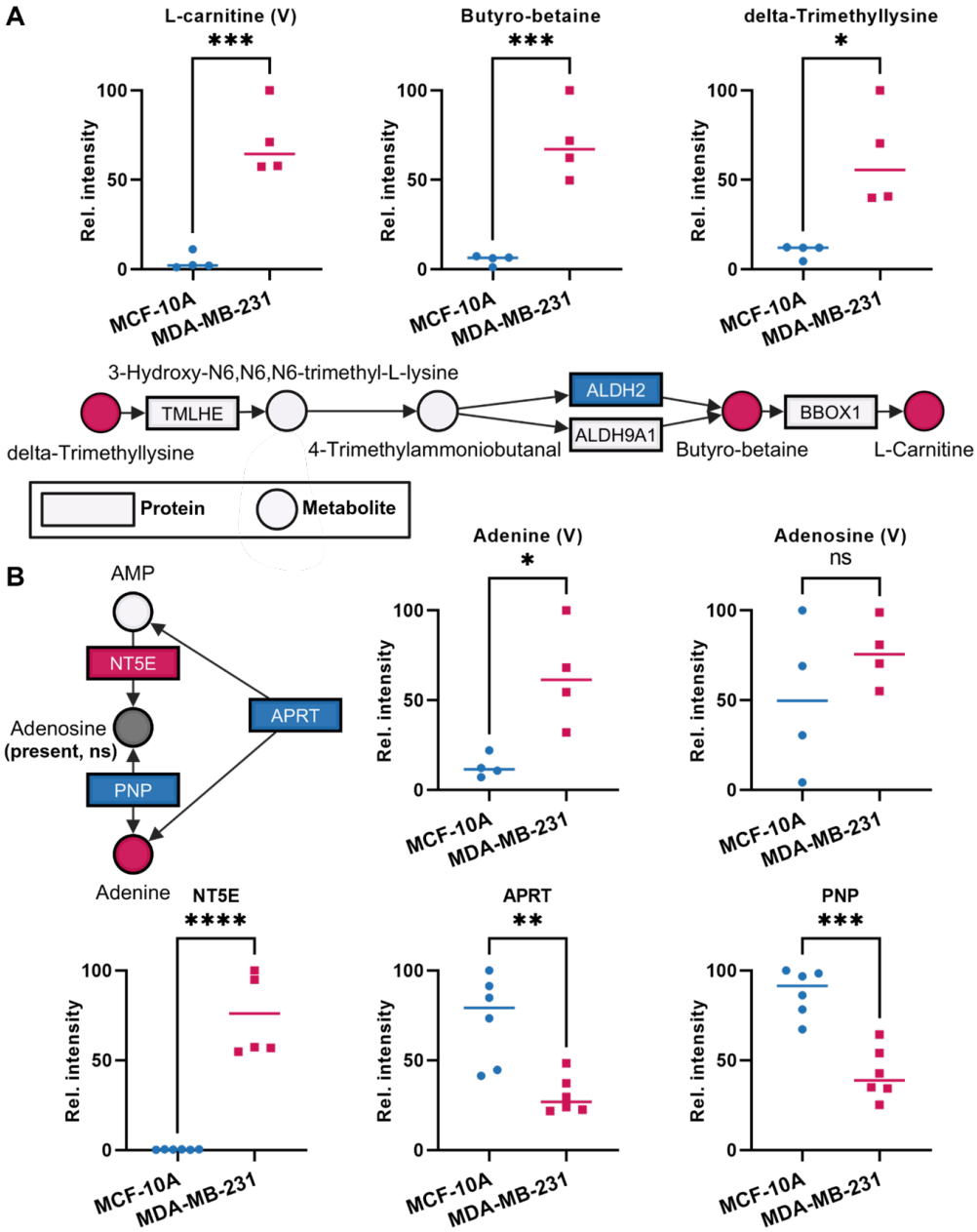

Figure 3.5: Pathways with enriched metabolites in MDA-MB-231 EVs compared to MCF-10A Blue color indicates enrichment in MCF-10A, while red color indicated enrichment in MDA-MB-231. A. Carnitine synthesis pathways revealed 3 upregulated compounds, and downregulated ALDH2 protein. B. Purine metabolism pathway revealed increased levels of adenine and NT5E protein in MDA-MB-231, as well as downregulated ecto-5'-nucleotidase (NT5E) and mitochondrial aldehyde dehydrogenase (ALDH2) proteins. Adenosine was detected in both groups without fold change difference. (V) indicates metabolites validated with standards. N = 4 for metabolomics. N = 6 for proteomics.

3.5 Discussion

The difference between cancer and non-cancer cell lines were explored through their EV composition. Carnitines play an important role in cancer metabolism and progression [101, 102]. Carnitines are small molecules that transport fatty acids into the mitochondria where they can be broken down to produce energy. Carnitines play a key role in the β -oxidation of fatty acids, which generates acetyl CoA that feeds into the citric acid cycle and oxidative phosphorylation to produce large amounts of ATP needed for rapid cancer cell growth and division [103–105]. Tumor cells often undergo metabolic reprogramming to increase fatty acid oxidation as a source of energy and biomass production to fuel rapid growth and proliferation [106–108]. Cancer cells are known to secrete large quantities of EVs to interact with and alter the tumor microenvironment [109, 110]. We have found that EVs contain various metabolites, including carnitines. The elevated levels of carnitines within tumor cell derived EVs suggests they may facilitate fatty acid fueling of recipient cells. By transferring carnitines systemically via EVs, tumor cells can remodel the metabolism of distant tissues to support metastasis and progression. We have not discovered any proteins related to this pathway aside from downregulated ALHD2 in EVs despite the presence of the metabolites. We hypothesize that EVs are packed with members of carnitine pathway at the stage of release by the tumor cells, rather than them acting as individual metabolic units and producing these metabolites on their own with pre-packed enzymes. It is important to mention that carnitine pathway members, especially PNP and ALDH2, are known to be localized to mitochondria. An argument can be made that these discoveries originate from mitochondria rather than EVs, however, after rigorous cleanup, our sample was mostly enriched for particles in the size range of 100 nm, corresponding to expected exosome size. Mitochondrial size ranges between 0.5-1 μm, which is outside of the range observed with DLS, therefore, we remain confident that our sample is indeed an EV fraction.

Increased levels of adenine have also been detected within EVs secreted from various tumor types. Adenine is a critical precursor for numerous metabolic processes and its efficient recycling is crucial to sustaining rapid proliferation [111,112]. The NT5E protein, also known as CD73, plays a key role in adenine salvage by catalyzing the conversion of AMP to adenosine. NT5E also has been demonstrated to be associated with tumor proliferation in cancers, including breast cancer [113–115]. Intriguingly, proteomic analyses have revealed elevated levels of NT5E within tumor cell derived EVs compared to non-malignant cell EVs. While we detected elevated levels of adenine and the NT5E protein within tumor cell EVs, we did not find phosphorylated AMP present based on our mass spectrometry results, possibly due to omission during sample preparation with cation exchange SPE. This is notable because NT5E catalyzes the dephosphorylation of AMP to produce adenosine. The

selective enrichment of adenine without its phosphorylated product AMP suggests active transport and packaging of specific metabolites into EVs occurs. This suggests cancer cells may exploit EV-mediated intercellular communication to optimize adenine metabolism in the tumor microenvironment. Specifically, the transfer of adenine, along with carnitines and increased NT5E via EVs is hypothesized to synchronously reprogram recipient cells towards a catabolic, autophagic state that maximizes adenine recycling and energy production from diverse substrates like fatty acids and amino acids. Taken together, these metabolic adaptations mediated by EV cargo could cooperatively enhance nutrient availability and support the malignant growth and progression. The combined results indicate MDA-MB-231 cells have increased de novo purine synthesis through the NT5E pathway while decreasing purine salvage through lower PNP and APRT expression.

3.6 Conclusions

In this study, we applied a nano sheath-flow CE-MS platform to conduct metabolomic profiling of exosomes derived from non-tumorigenic MCF-10A and tumorigenic MDA-MB-231 breast epithelial cells. The results revealed distinct metabolic signatures between the two cell types, highlighting altered purine and carnitine metabolism in the cancer exosomes. Specifically, MDA-MB-231 exosomes displayed elevated levels of purine synthesis intermediates and enrichment of the NT5E pathway involved in de novo purine biosynthesis. In contrast, carnitine shuttle proteins like ALDH2 were downregulated. Together with the differential expression of related enzymes NT5E and ALDH2, these findings demonstrate metabolic reprogramming in purine and carnitine pathways associated with the breast cancer phenotype. This work provides novel insights into how breast cancer cells rewire cellular metabolism to impact exosome content as evident from 2 cell line system.

3.7 Acknowledgements

Conceptualization, Maxim Berezovski and Emil Zaripov; Methodology, Emil Zaripov; Experimental, Emil Zaripov, Abdullah Khraibah and Petr Kasyanchik; Data Analysis, Emil Zaripov, Resources, Maxim Berezovski; Data Curation, Emil Zaripov; Writing—Original Draft Preparation, Emil Zaripov; Writing—Review & Editing, Maxim Berezovski and Emil Zaripov; Visualization, Emil Zaripov, Abdullah Khraibah and Petr Kasyanchik; Supervision, Maxim Berezovski; Project Administration, Maxim Berezovski and Emil Zaripov; Funding Acquisition, Maxim Berezovski.

Chapter 4

Chapter 4: Dual Detection of SARS-CoV-2 Nucleoprotein and RNA Using Proximity Ligation of Aptamers Developed with CE-SELEX

4.1 Abstract

The global COVID-19 pandemic stressed the need for accurate and sensitive diagnostic tools to detect SARS-CoV-2 infection. While a real-time reverse-transcription polymerase chain reaction (RT-qPCR) laboratory test remains the gold standard, antigen testing provides a faster and simpler alternative despite reduced sensitivity. Beyond these methods, aptamers have emerged as promising substitutes for antibodies in viral detection. This study aimed to develop aptamers targeting the nucleocapsid (N) protein of SARS-CoV-2 and apply them in a novel test called proximity ligation aptamer-mediated RT-qPCR (PLA/RT-qPCR) for simultaneous detection of N protein and viral RNA. Capillary electrophoresis was used to select six high-affinity aptamers to the N protein. Two aptamers, ECK1 and ECK4, were selected based on their superior performance in a proximity ligation assay (PLA). These were adapted into a PLA assay using one aptamer with a forward primer and the other with a reverse primer, enabling ligation and subsequent detection by real-time PCR [Figure 4.1]. Concurrently, viral RNA was detected via RT-qPCR with

a Taqman probe in the same tube in a different color channel. The PLA/RT-qPCR performance was validated in spiked saliva, demonstrating feasibility in complex matrices. Finally, aptamer epitopes were mapped to the N protein using pulldown proteomics after trypsin digestion, providing insights into their binding sites. This work presents a novel aptamer-based PLA/RT-qPCR diagnostic strategy for sensitive dual detection of SARS-CoV-2 protein and RNA in a single test.

4.2 Introduction

The global outbreak of the Coronavirus disease (COVID-19) has stressed the urgent need for accurate and sensitive diagnostic tools for detecting the presence of the SARS CoV 2 virus. Effective detection methods are crucial for the timely identification of infected individuals to enable appropriate containment measures and treatment interventions. The detection of SARS CoV 2, the causative agent of COVID-19, is critical for diagnosis and disease surveillance. Real time reverse transcription polymerase chain reaction (RT-qPCR) detection of viral RNA remains the gold standard for SARS-CoV-2 diagnosis as it is highly sensitive and can detect low viral loads, making it well-suited for diagnosis in the early stages of infection. Rapid antigen tests offer simpler, faster, and cheaper options for SARS-CoV-2 detection, though they are generally less sensitive than RT-qPCR. Antigen tests detect SARS-CoV-2 nucleocapsid protein and can provide results in under an hour; however, they have reduced sensitivity compared to RT-qPCR, especially when viral loads are lower. Beyond RT-PCR and antigen testing, other molecular methods such as CRISPR and next-generation sequencing show promise for SARS-CoV-2 detection and variant characterization. Serological antibody tests are also important for understanding population exposure and immune response.

Traditional protein detection approaches primarily rely on antibodies, which have certain limitations such as batch-to-batch variability and the need for specific handling and storage conditions. In recent years, aptamers have emerged as a promising alternative to antibodies for the detection and treatment of viral infections [116, 117]. Aptamers are single-stranded oligonucleotides with the unique ability to bind specifically to target molecules, including viral proteins. Their high affinity and specificity, combined with their ease of synthesis and modification, make aptamers an attractive option for diagnostic applications [118–121]. Aptamers are generated through a systematic evolution of ligands by exponential enrichment (SELEX), an iterative process of generation of high affinity aptamers from a large pool of randomized DNA.

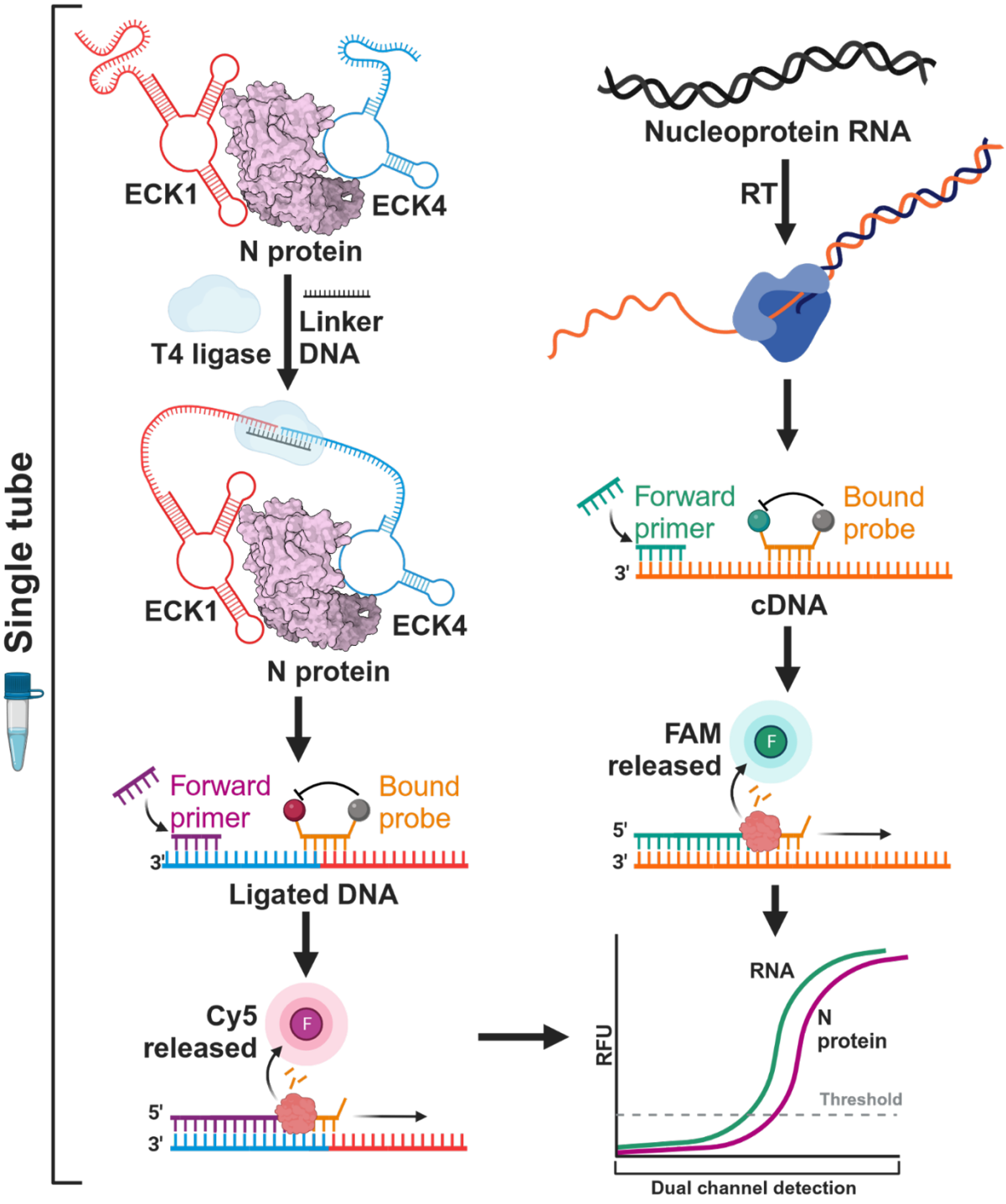

Figure 4.1: Single tube dual detection of SARS-CoV-2 RNA and N protein through PLA/RT-qPCR.

The ongoing research of SARS-CoV-2 aptamers has contributed significantly to the development of innovative diagnostic strategies. The exploration of aptamers targeting various SARS-CoV-2 proteins, mostly Spike (S) protein and Nucleocapsid (N) protein, and other viral components, has paved the way for the development of sensitive and reliable diagnostic tools. S protein carries therapeutic potential as it is the protein involved in human ACE-2 recognition and cell's first infection. Several papers, therefore, focus on aptamers against receptor binding domain (RBD) of S protein [122–125]. Nucleocapsid is another common target due to its high evolutionary conservation and abundancy, making it a great candidate for detection applications. Aptamers have been developed to N protein, with some involved in applications in biosensors [126–128]. Among the various viral proteins, the N protein plays a pivotal role in virus replication, assembly, packaging, and RNA transcription [129]. Notably, this protein exhibits a lower propensity for mutations compared to other essential SARS-CoV-2 proteins [130], making it a reliable target for detection.

In this study, we aimed to discover aptamers against N protein and use them for the detection of the protein through proximity ligation assay (PLA) followed by real-time PCR. PLA is a molecular technique employed to assess interactions between proteins or nucleic acids in proximity. PLA has been widely employed through antibody-DNA conjugates, where two different antibodies recognize the target and bring their conjugated oligonucleotides together, allowing for ligation to happen [131,132]. Aptamers are used a lot rarer as primary ligands; however, several studies demonstrated their feasibility [133–137]. Our novel approach to SARS-CoV-2 detection involves simultaneous detection of viral RNA through TaqMan probe mediated RT-qPCR, as well as detection of N protein by PLA involving two aptamers that were selected [Figure 4.1]. Detection of both protein and RNA in one PCR reaction can provide more confidence in virus detection than RNA alone. Capillary electrophoresis (CE) was employed as a platform for aptamer selection [Figure 4.2], and the estimation of their epitopes. Six aptamers that interact with N protein with a high affinity were discovered, and then two of them were used to develop our PLA approach.

ECK1 and ECK4 aptamers with a singular primer site each were used to achieve N protein detection through ligation of the two aptamers together using a linker sequence. Subsequently, Cy5 TaqMan probe RT-qPCR was used to detect the ligation site. At the same time, RNA was detected with RT-qPCR in the FAM channel, enabling simultaneous detection of both the viral protein and RNA in a single RT-qPCR assay. A completely novel approach to detection of both viral protein and RNA in a single vial simultaneously is suggested, opening new avenues for diagnostics and its applications.

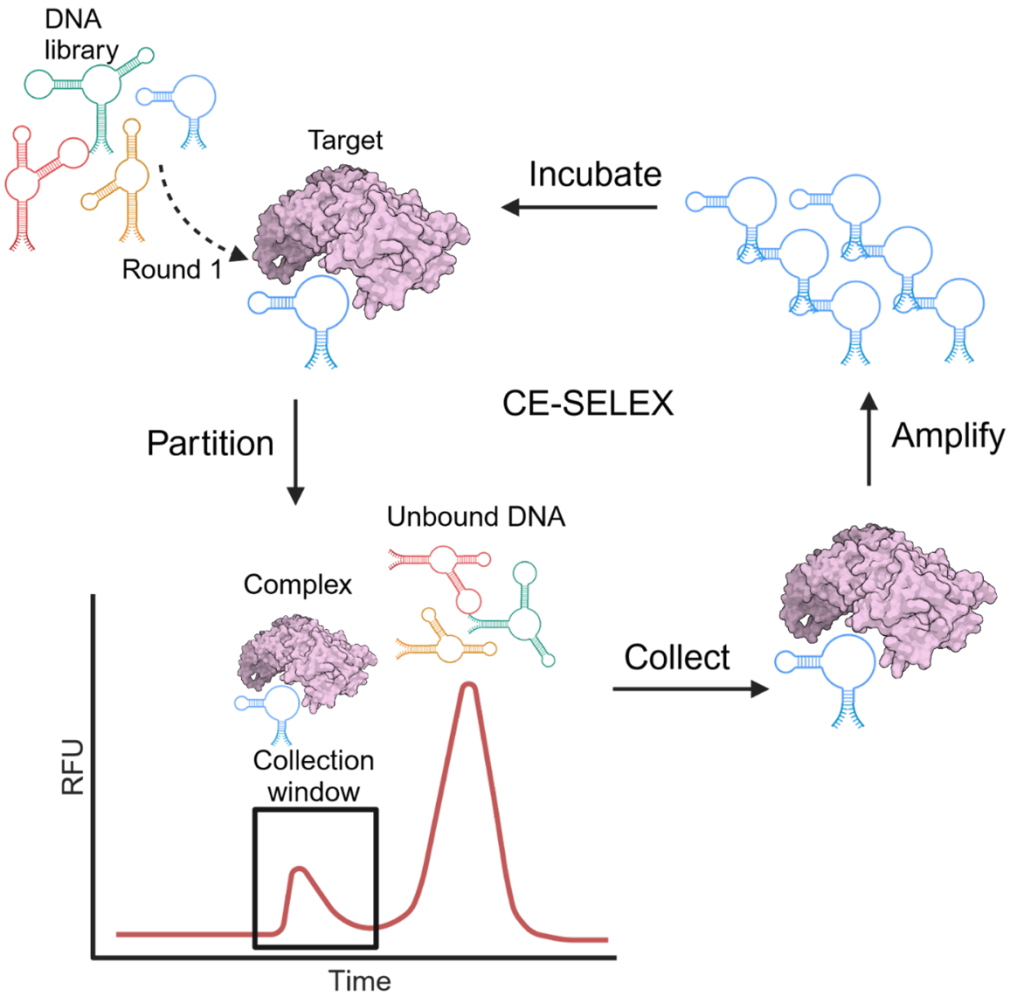

Figure 4.2: Schematic of CE-SELEX of DNA aptamers to a protein target.

PLA/RT-qPCR method was further validated using saliva spiked with N protein. Saliva was chosen as a complex matrix that is generally collected for SARS-CoV-2 detection during RT-PCR testing. Without any sample cleanup steps, it was possible to distinguish saliva negative for N protein from the positive one using our method, proving robustness outside of simple buffer systems.

Lastly, the binding epitope of the aptamers on the N protein was determined. This was achieved by performing trypsin digestion of the N protein in the presence of the aptamer followed by pulldown experiment. By utilizing bottom-up proteomics approach,

peptides that were enriched in the pulldown fraction were identified. This approach enabled determination of the epitope of the N protein recognized by the aptamer. This information was then used to construct a hypothetical molecular docking *in silico*.

4.3 Materials and Methods

4.3.1 Protein expression

DH5α strain with a plasmid encoding for N protein was received from Rick Tarleton (Addgene, Watertown, MA). An LB agar plate containing 100 μg/mL kanamycin (Sigma-Aldrich, St. Louis, MO) was inoculated with the culture and incubated at 37°C for 24 hours. A single colony was picked and inoculated into 5 mL LB with 100 μg/mL kanamycin, which was incubated at 300 rpm and 37°C overnight. The plasmid was extracted using a miniprep kit (Qiagen, Germantown, MD), and 10 ng of the plasmid was used to transform into competent DE3 cells (Thermo Fisher Scientific, Waltham, MA) according to the manufacturer's protocol. The transformed competent cells were cultured in the same conditions as DH5α and were frozen in 50% glycerol/LB.

The frozen cells were used to inoculate 500 mL LB with 100 μg/mL kanamycin. The culture was incubated at 37°C on a magnetic stir plate at 300 rpm until an OD600 reached 1. The culture was then supplied with 1mM IPTG (Sigma-Aldrich, St. Louis, MO) and further incubated for 3 hours. Cells were spun down at 4,000 g for 30 minutes and resuspended in T300 buffer (30mM triethanolamine, 300 mM NaCl, pH 7.8). Aliquots of 1mg/mL of lysozyme and 1mM PMSF (Sigma Aldrich, St. Louis, MO) were added before sonication for 30 minutes on ice. The lysate was centrifugated at 21,000 g for 40 minutes, and pellets were resuspended in T300 with 8 M urea (Sigma-Aldrich, St. Louis, MO). The solubilized pellets were then filtered through a 0.45μm nylon membrane.

The C 10/10 chromatography column (Cytiva, Marlborough, MA) was packed with 7.85 mL of Ni-NTA agarose beads (Thermo Fisher Scientific, Waltham, MA). Packed media was equilibrated with T300 buffer containing 8 M urea for 4 column volumes. 10 mL of solubilized sample was injected into the column, which was then washed 4 more column volumes of the same buffer containing urea. The buffer was slowly replaced with T300 by applying gradient for 8 column volumes to gradually remove urea and refold protein on the column. The column was then washed with T300 with 30 mM imidazole for 4 column volumes. The protein was then eluted with 4 column volumes of elution buffer (T300 and 300mM imidazole) and concentrated on a 10 kDa filter (Sigma-Aldrich, St. Louis, MO). All chromatography steps were performed using a flow rate of 3 mL/min.

In the next step, gel filtration chromatography was performed using a XK 16/70 column (Cytiva, Marlborough, MA) which was packed with 140 mL of Sephacryl S-100 HR resin (Cytiva, Marlborough, MA). The column was washed with 2 volumes of 150 mM ammonium acetate, pH 7.0. 1 mL of concentrated sample was injected and separated at a flow rate of 0.5 mL/min. A280 was monitored during the run to collect fractions of N protein. The collected protein was then concentrated on a 10 kDa filter to the final volume of 1 mL and stored at -20 °C until further use. The protein concentration was then determined using a nanodrop spectrophotometer. 2 μL of protein mixture was loaded onto a pedestal, and A280 reading was taken. With a calculated extinction coefficient (43.9 x 103 M-1cm-1), the concentration was calculated and the protein was aliquoted for further use.

Following protein expression, the protein's purity was assessed through a direct infusion into Synapt G2 mass spectrometer (MS). Protein was diluted to 0.5 μg/mL in 50% acetonitrile with 0.1% formic acid and was infused at 1 μL/min. Capillary voltage was set to 3.3 kV, and cone voltage was set to 60V. Drying nitrogen gas was delivered at 10 L/h. Range of 300-3000 m/z was recorded at 1 Hz scan rate for 30 seconds to obtain the spectrum. Deconvolution was performed with UniDec software. Mass range was set to 500-1600 m/z, and the charge range was set to 10-80. Potential size range was set to 5-200 kDa.

4.3.2 CE-SELEX

N40 DNA library (CTCCTCTGACTGTAACCACG-N40-GCATAGGTAGTCCAGAAGCC) with 5' Cy5 fluorescent tag was synthesized by IDT DNA Technology (Iowa, USA) by a custom order. It was diluted to a concentration of 5 μM in 20 mM ammonium acetate (pH = 7.0). The library was then heated to 95 °C for 2 minutes and slowly cooled down to 4 °C for 5 minutes. 1 μM of N protein was added and the mixture was incubated at RT for 30 minutes.

New CE capillary (60 cm, 360 μm O.D., 75 μm I.D.) was first washed with 1M NaOH for 30 minutes at 20 psi, and then prefilled with separation buffer (20 mM ammonium acetate, pH = 9.2) for 5 min at 20 psi. Prepared library-protein mixture was injected at 0.5 psi for 10 seconds, sandwiched between 2 plugs of water injected with the same parameters. Separation was carried out at 20 °C with 500 V/cm electric field for 15 minutes. Collection was performed in 50 μL of separation buffer placed in a sample vial. The collected vial was then dried under vacuum, producing dry DNA pellet.

4.3.3 Emulsion PCR and ssDNA generation

Emulsion PCR was used for all amplifications. Oil-surfactant mix was prepared by adding 4.5% (v/v) Span-80, 0.4% (v/v) Tween 80, 0.05% (v/v) Triton X-100, and topping the mixture to 50 mL with mineral oil. Q5 high fidelity PCR master mix was prepared according to the manufacturer's protocol. 200 μM of dNTPs were added, as well as 500 nM of forward and reverse primers each, 0.02 U/ μL of Q5 polymerase and 10 μg/ μL of BSA. 150 μL of master mix was added to previously dried pellet of collected DNA and mixed well. 300 μL of oil-surfactant mix was aliquoted into 1.5 mL glass autosampler vial with a magnetic stir bar, and placed on a stir plate at 1000 rpm. 150 μL of master mix was added dropwise to a rotating oil-surfactant mix over the course of 2 minutes, emulsifying the mix. Master mix emulsion was then added to 8 x PCR tube strips, 50 μL in each. PCR was performed in an Eppendorf thermocycler. PCR mix was first heated to 98 °C for 30 seconds for polymerase activation. 40 cycles of PCR were performed: 98 °C denaturation for 5 seconds, 56 °C annealing for 10 seconds and 72 °C elongation for 10 seconds. Final extension was done for 1 min at 72 °C prior to cooling samples down to 4 °C. Emulsion mix was then combined and spun down for 15 min at 20,000 g. Supernatant (mineral oil) was aspirated, and the pellet was extracted with 500 μL of diethyl ether twice to remove the surfactants.

The samples were subjected to electrophoresis on a 1.2% agarose gel for a duration of 30 minutes at 12 V/cm in TBE buffer (100 mM Tris, 100 mM boric acid and 2 mM EDTA). Prior to gel electrophoresis, the samples were stained with 1X SYBR gold (Invitrogen, Cat No. 41003). The resulting gel was viewed under UV light and the 100 bp band was excised and extracted using Monarch DNA Gel Extraction Kit (NEB #T1020). 3 mL of gel dissolving buffer was added to 1 mL of excised agarose and heated to 50 °C for 30 minutes. The mix was then passed through a silica column for 1 min at 10,000 g and washed with 100 μL of wash buffer twice. DNA was then eluted with 10 μL of deionized water and quantified on Nanodrop One.

New master mix for asymmetric non-emulsion PCR was then prepared to generate ssDNA pool for next round of SELEX. 200 μM of dNTPs, 400 nM of forward primer, 20 nM of reverse primer, and 0.02 U/ μL of Q5 polymerase were added along with 1-10 ng of template previously prepared. 25 cycles of PCR were performed with the same cycling parameters described in emulsion PCR section. The product was then subjected to gel electrophoresis and extraction from gel using the same procedure outlined above. Cleaned up product was stored at – 80 °C until further use for the next round of SELEX.

4.3.4 Preparation for Illumina sequencing

Three pools as well as the library itself were amplified using emulsion PCR protocol outlined above. The primers were modified to include adapters, where forward primer was 5'ACACTCTTTCCCTACACGACGCTCTTCCGATCTTTCTCCTCTGACTGTAACCACG3' and the reverse primer was 5'GACTGGAGTTCAGACGTGTGCTCTTCCGATCTTTTGGCTT CTGGACTACCTATGC3'. The amplification products were separated on a 1.2% agarose gel in TBE buffer and 150 bp band was excised and extracted using gel extraction kit listed in SELEX section. 10 ng of dry DNA for each of the pools was shipped for sequencing.

4.3.5 Selection of clones from sequencing results

FASTQ files were first converted to FASTA using *fastaptamer_count* function of FASTAptamer software package [138]. The sequences with lengths between 78 and 82 were selected, and the rest were omitted because 80 nts band was excised from during every stage of selection and amplification prior to sequencing [Figure 4.3]. Sequences not containing both primer binding sites were also removed from the analysis. Then, the sequences with primer repetitions in the variable region were removed, leaving only the sequences with true random variable region. The sequences were then enriched according to the pool progression using *fastaptamer_enrich* function of FASTAptamer software package. Sequences present in both pools were sorted by their rank in pool 2, and top sequences were selected and synthesized for further analysis. Among these sequences, we picked the 6 most abundant in pool corresponding to round 2 of SELEX. We decided to use round 3 as our primary pool for sequence selection because round 4 did not bring improvement to the binding of the pool. We anticipated that an extra round of PCR would introduce more by-products and decided on round 3 instead.

Figure 4.3: Workflow of sequencing analysis using FAST Aptamer.

4.3.6 Biolayer interferometry analysis

For evaluating the binding strength between the aptamer pools or full-length clones and N protein, Ni-NTA biosensors and the °Ctet N1 BLI instrument were employed. Initially, DNA-enriched pools or full-length aptamers were prepared in PBS, pH 7.4, 1 mM MgCl2, 1 mM CaCl2, 0.02% Tween 20 and 0.2% BSA. The Ni-NTA biosensor tips were loaded with his-tagged N protein, and the binding assay was initiated by establishing a baseline in the binding buffer. Subsequently, the association between the DNA and N protein on the biosensors was monitored. The association step involved four different samples: N40 library, round 1, round 2, and round 3. The assay was performed at room temperature, with a 30-second baseline, 120-second loading, 100-second association, and 150-second dissociation periods. For truncated clones' validation, 400 nM biotinylated clones were prepared in PBS, pH 7.4, 1 mM MgCl2, 1 mM CaCl2, 0.02% Tween 20 and 0.2% BSA. °Ctet Streptavidin sensor was first hydrated in the same buffer for 10 minutes, and then incubated with the clone for 60 seconds, after which it was immersed in the buffer again. For the interaction kinetics experiment, the baseline was first established in the buffer for 30 seconds, after which the association was monitored in varying concentrations of N protein (0-125 nM) in the same buffer for 180 seconds. Dissociation was performed in the buffer for 60 seconds, followed by 30 seconds of regeneration of the sensor surface in 10 mM Glycine-HCl (pH = 2.0). Four concentrations of protein as well as blank were monitored, each repeated 3 times. For the competitive displacement experiment, 125 nM N protein was loaded onto the sensor for 180 seconds, followed by baseline establishment in the buffer for 60 seconds. Competitive binding was then monitored by association or dissociation in presence of 100 nM of the aptamer clone.

4.3.7 Proximity ligation assay

Stock solutions of aptamer pairs were prepared at the concentrations of 10 nM in ddH_2O in presence of 0.001% PEG8000 to prevent tube wall interactions. Reaction solutions were prepared in 25 mM ammonium acetate (pH = 9.2) and 0.001% PEG8000 with 10 nM of N protein or BSA for control samples, and 1 nM of each of the aptamer. Incubation was performed at RT for 1 h, after which ligation master mix was added according to manufacturer's recommendations (NEB T4 ligase), as well as 3 nM linker. Ligation was performed for 5 minutes at 37 °C, followed by heat deactivation at 65 °C for 20 minutes. Reaction mixture was then diluted 5-fold with ddH_2O and used directly for rPCR experiment.

During concentration titrations, all of the above conditions were kept the same except the concentration of N protein, which ranged between 1 nM and 25 nM, and 0 nM for

reference sample. For ligation time optimization, all the above conditions were kept the same, and the concentration of N or BSA proteins were 10 nM.

4.3.8 Real-time RT-PCR

RT-qPCR was performed using Promega GoTaq 1-step RT-PCR master mix. Master mix was prepared according to the manufacturer (Promega) in presence of 500 nM concentration of each primer and 125 nM of probe. Two sets of primers/probe were added, one for ligation product detection, and another for RNA detection. 1 μL of ligation reaction product as well as 3000 cp/μL RNA template (EURM019, Sigma) were added per 10 μL of reaction mixture. Two channels, FAM and Cy5, were monitored during PCR amplification. Samples were first held at 40 °C for 10 min to reverse transcribe RNA into cDNA, followed by 50 cycles of 95 °C hold for 3 sec and 50 °C annealing/extension for 45 seconds. Readings were taken at the end of each cycle and used to construct amplification curves.

4.3.9 Epitope pulldown

Biotinylated aptamers were prepared at 2 μM concentration in 25 mM ammonium acetate (pH = 9.2) and heated to 95 °C, followed by gradual cooling to 4 °C. 1 μM of N protein was added and incubated with aptamer for 30 minutes at RT. 100 ng of trypsin/LysC was added along with 10 mM tris-HCl (pH = 8.1) to a total reaction volume of 50 μL. Samples were incubated at 37 °C for 30 minutes, followed by rapid mixing with 10 μL of streptavidin-agarose resin (Pierce, 20353) and incubation for 2 min. The slurry was then loaded onto a spin column and washed twice with 10 μL of 25 mM ammonium acetate buffer (pH = 9.2), followed by elution with 50 μL of 1 M formic acid. The sample was then immediately purified using an in-house prepared micro SPE tip with C18 silica. SPE tip was first activated with 3 x 50 μL of 70% acetonitrile with 0.1% formic acid, followed by washing with 3 x 50 μL of 0.1% formic acid. 50 μL sample was then loaded and passed through the silica. The silica was then washed again with 3 x 50 μL of 0.1% formic acid, followed with peptide elution with 3 x 50 μL of 70% acetonitrile with 0.1% formic acid. Samples were then dried under vacuum stored for further CE-MS analysis.

4.3.10 Capillary electrophoresis-mass spectrometry

The new capillary was conditioned with 60 min wash with 1M ammonium hydroxide, followed by 60 min wash with 1 M formic acid. All separations were carried out in 0.5

M formic acid at 250 V/cm electric field. 80 cm capillary with 360 μm O.D. and 40 μm I.D. was used. Dry sample was resuspended in 10 μL of water. Prior to sample injection, 50 mM ammonium hydroxide was injected at 1 psi for 20 seconds. Sample injection was carried out at 1 psi for 40 seconds.

Sheath liquid (30% isopropanol with 25 mM formic acid) was delivered at a rate of 500 nL/min. Source temperature was maintained at 110 °C. Capillary voltage was set to 3.3 kV. Sampling cone voltage was set to 30 V. Extraction cone volage was set to 5 V. Cone gas was set to 10 L/h. Waters MS method was used with 1 Hz scan rate and a range of 200 1200 *m/z*. Lockmass solution consisting of 200 pg/mL Leucine-Enkephalin was delivered perpendicularly at 100 nL/min, collecting 1 scan (0.5 seconds) every 3 minutes for *m/z* correction.

Peak picking was performed using MzMine 3 using exact mass algorithm [139]. ADAP chromatogram builder was used to construct extracted ion electropherograms for each detected peptide. Mass tolerance was set to 0.005 Da or 20 ppm, with RT window set to 1 min. Minimum peak height was set to e^3, with $5e^2$ set for the noise cutoff. The data was processed as is, without normalization or gap filling. Fold change analysis and t-test were performed to determine statistically significant peptides. Identification was performed using MS1 *m/z* values and pre-made list of expected *m/z* values based on Expasy PeptideMass web software.

4.3.11 Molecular docking

Solved N protein structure was retrieved from PDB (ID = 8fdt). Aptamer sequences were first folded with RNAfold web server using Matthews model for ssDNA at 37° and 0.025 mM ionic strength. RNAcomposer was then used to obtain three-dimensional structure in .pdb format. HADDOCK web server [140] was used to first simulate molecular docking between ECK1 and N protein. For N protein, AYNVTQAFGR peptide (267-276) was specified as the receptor peptide, while for the aptamer, the entire sequence (1-40) was used. Following protein-DNA complex obtainment, second docking was performed, where N-ECK1 complex was loaded as a receptor with protein-nucleic acid type of molecule. The 3D structure of ECK4 was obtained in the same way as ECK1 and was added as a ligand for the N-ECK1 complex. For ECK4, HIDAYKTFPPTEPKKDK peptide (356-372) was specified as the ligand peptide, while for the aptamer itself, the entire sequence was used (1-40).

4.4 Results

4.4.1 Aptamer Selection

Protein expression

Nucleoprotein was validated using top-down approach and direct infusion ESI-MS. Deconvoluted spectrum demonstrated presence of 51.4 kDa entity, which corresponds to the calculated size of N protein from the sequence in the plasmid ([Figure 4.4]. The sequence of the protein was also validated by bottom up approach, confirming the correct sequence of N protein. The presence of additional entities was negligible, confirming clean preparation of the N protein.

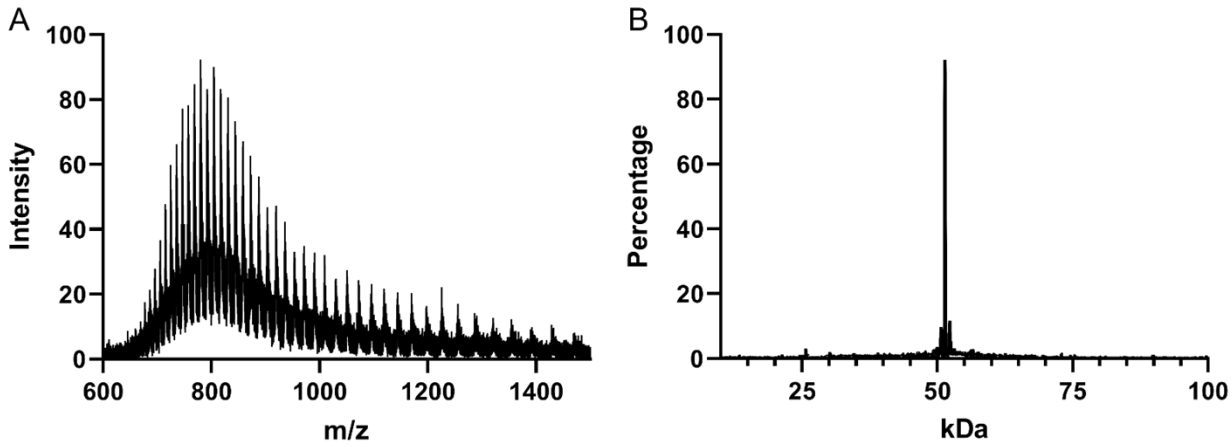

Figure 4.4: N protein validation using ESI-MS. A. Convoluted multicharged spectrum of denatured N protein. B. Deconvoluted spectrum showing abundant 51.4 kDa protein.

Systematic evolution of ligands by exponential enrichment

The first round of 5 μM N40 library with 1 μM of N protein showed a complex at 4 minutes as opposed to a library only run [Figure 4.5]. It is worth noting that 2 complex peaks were observed, presumable due to single or multiple DNA molecules co-migrating with a singular protein. Ammonium acetate also allowed for a fast electroosmotic flow due to its high pH, making the separations fast (under 8 minutes) and allowing to perform 10-20 collection runs to accumulate more product. It is noticeable that while round 2 showed better binding

compared to round 1, round 3 provided marginal improvement, where 160 nM of N protein in round 3 yielded equivalent peak area of protein-DNA complex compared to 180 nM of N protein in round 2.

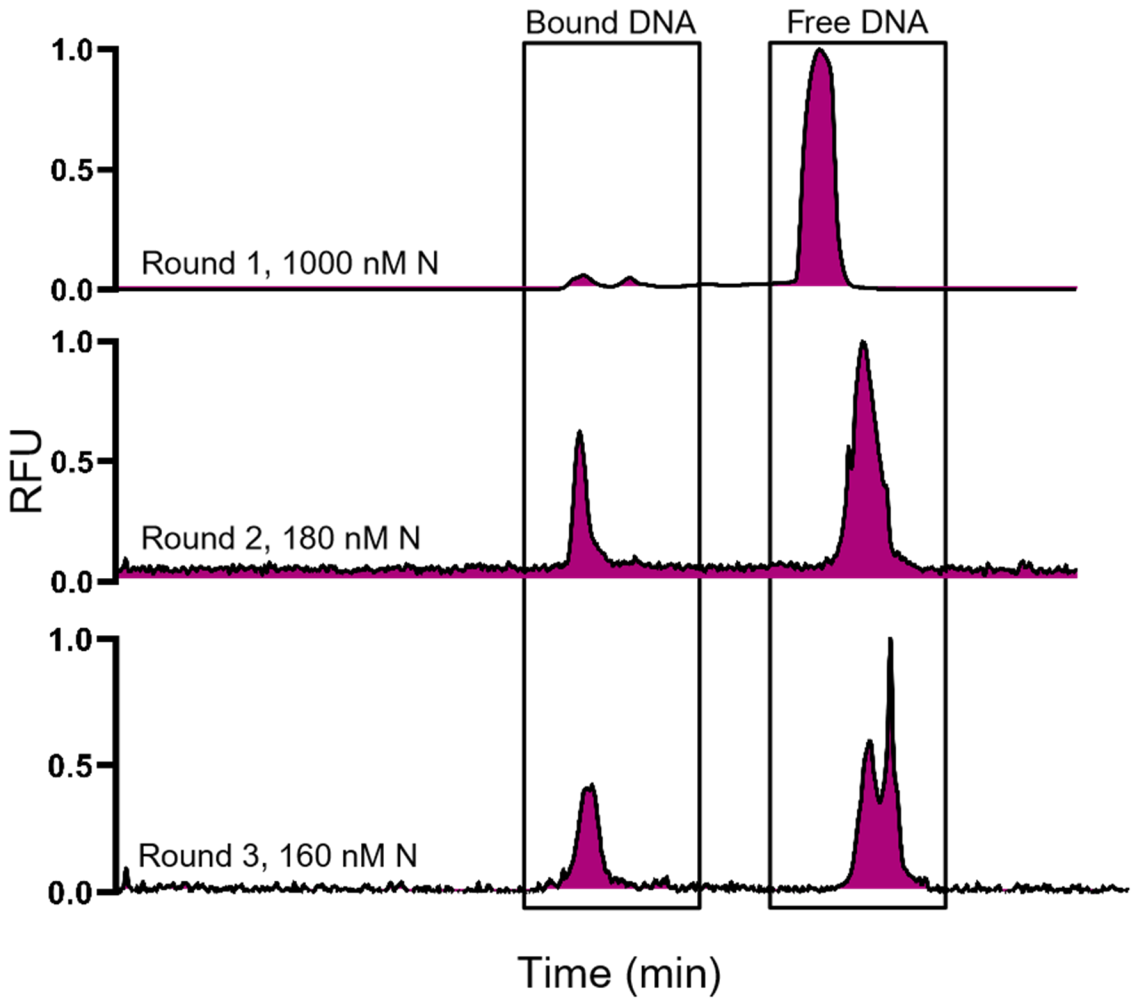

Figure 4.5: Representative electropherograms of three rounds of selection. Highlighted windows reflect collected fractions for bound DNA and discarded fractions of free DNA.

BLI validation of SELEX pools

In addition to pool affinity monitored by CE, we performed BLI as an orthogonal method to confirm improved affinity of the pools compared to initial library. We used Ni-NTA sensor with immobilized 6xHis tagged N protein to monitor binding affinity of the DNA pools. Our first pool showed lower affinity than the starting library, however, all subsequent pools exhibited improved binding, confirming CE data [Figure 4.6 A].

Biolayer interferometry on six full-length clones

Six clones that were selected from sequencing analysis were tested on BLI [Table 4.1]. We used random scrambled 80-nt sequences as a negative control, and previously published tNSP3 aptamer as a positive control. Scrambled DNA showed very low affinity compared to all 6 of the aptamers, as well as the tNSP3 positive control [Figure 4.6 A]. As with pools, we used Ni-NTA sensors with immobilized N protein. The affinity curves confirmed that our full-length aptamers bind better when compared to the scrambled DNA or the library, however, we aimed to produce shortened truncated version of our clones without the primer binding sites, thus reducing the length by 40 nts.

4.4.2 Aptamer Characterization

Truncated sequences validation

Aiming to reduce the length and cost of the aptamers, we performed truncation of the aptamers by removing forward and reverse primer binding sites, yielding 40-nt sequences. These sequences were ordered in biotinylated forms for validation of the binding. Streptavidin Octet BLI sensors were used to immobilize the DNA aptamers via the biotin label. Varying concentrations of the N protein were used to monitor binding kinetics with each of the aptamers. Figure 4.6 B demonstrates 4 different concentrations of N protein as well as a blank to each of the aptamers. All 6 aptamers have showed adequate binding, with KD ranging from 4.8 nM for ECK1, and 11.9 nM for ECK6. We also employed a previously selected positive control aptamer, tNSP3 [141], which demonstrates comparable binding affinity to our best clones. All the binding curves were performed in triplicates, with kinetics values present in Table 4.2.

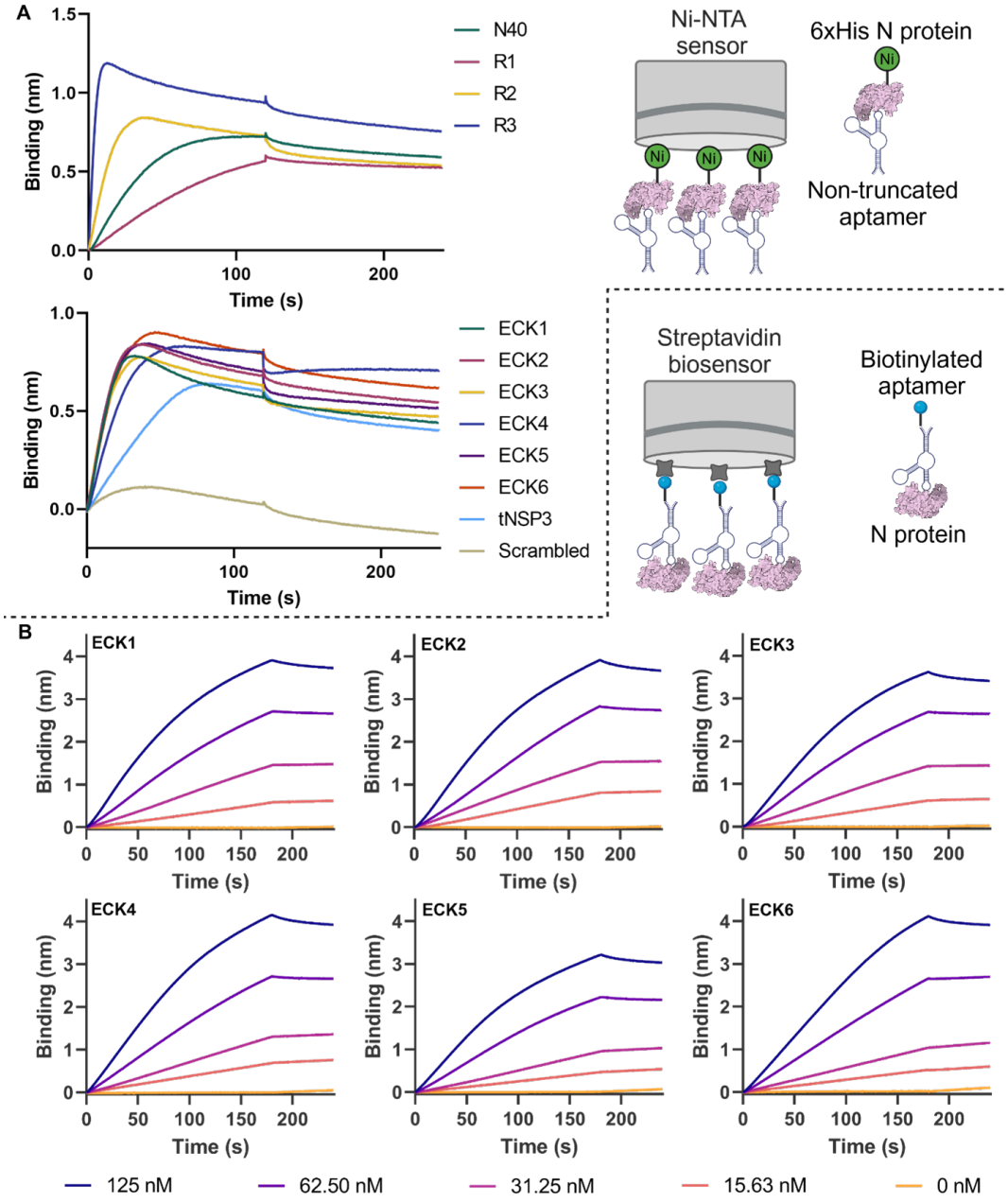

Figure 4.6: Biolayer interferometry panel. A. BLI experiments performed with a Ni-NTA biosensor immobilizing N protein through a histidine tag. Representative binding curves of 3 aptamer pools compared to the library demonstrate improving affinity through the course of SELEX. Full-length aptamer clones are presented in comparison against positive tNSP3 control as well as scrambled DNA negative control. B. BLI experiments performed on streptavidin sensor and immobilized biotinylated truncated clones. Titration of 4 different concentrations as well as negative control are presented for each truncated clone. All reactions were carried out in DPBS with 0.05% Tween 20 and 0.2% BSA. N = 3.

Table 4.1: Six full-length sequences selected based on data analysis. Primer binding sites surrounding variable N40 region are highlighted in bold.

Clone	Sequence (5' - 3')
fECK1	**CTCCTCTGACTGTAACCACG**TCATCGCTTTGCATATTCTATGTCCTGTCCATACTCGACA**GCATAGGTAGTCCAGAAGC**C
fECK2	**CTCCTCTGACTGTAACCACG**TTCACCGCTGCCCATCGGCATGGTTACCCGCGGTCCTTTT**GCATAGGTAGTCCAGAAGC**C
fECK3	**CTCCTCTGACTGTAACCACG**CCTCCTATTGCGCCCCAACGCATAGGCAATGGCCTACAGG**GCATAGGTAGTCCAGAAGCC**
fECK1	**CTCCTCTGACTGTAACCAC**GTTCGAACGAGGTTGCTCTACCGGCCAGGCCCTAGGTGCGT**GCATAGGTAGTCCAGAAGCC**
fECK2	**CTCCTCTGACTGTAACCACG**GTTATACAGTGTTGTGACCGAGGCACATCCGCTATCACAG**GCATAGGTAGTCCAGAAGCC**
fECK3	**CTCCTCTGACTGTAACCACG**AGTCAGGCGCTCATTCAAAATGAGGATCGTCATCATCCAA**GCATAGGTAGTCCAGAAGCC**

Table 4.2: Binding kinetics of 6 truncated aptamer clones to N protein. Error is represented with standard deviation. N = 3.

Clone	**Truncated sequence**	**KD (nM)**	**Ka (1/Ms)**	**Kd (1/s)**
ECK1	TCATCGCTTTGCATATTCTATGTCCTGTCCATACTCGACA	4.8 ± 0.47	$5.7 \times 10^{4} \pm 7.0 \times 10^{2}$	$2.8 \times 10^{-4} \pm 3.6 \times 10^{-5}$
ECK2	TTCACCGCTGCCCATCGGCATGGTTACCCGCGGTCCTTTT	9.2 ± 0.75	$4.8 \times 10^{4} \pm 7.9 \times 10^{2}$	$4.4 \times 10^{-4} \pm 4.1 \times 10^{-5}$
ECK3	CCTCCTATTGCGCCCCAACGCATAGGCAATGGCCTACAGG	10 ± 0.55	$5.2 \times 10^{4} \pm 6.8 \times 10^{2}$	$5.2 \times 10^{-4} \pm 3.5 \times 10^{-5}$
ECK1	TTCGAACGAGGTTGCTCTACCGGCCAGGCCCTAGGTGCGT	10.5 ± 0.75	$4.3 \times 10^{4} \pm 7.5 \times 10^{2}$	$4.5 \times 10^{-4} \pm 3.8 \times 10^{-5}$
ECK2	GTTATACAGTGTTGTGACCGAGGCACATCCGCTATCACAG	6.9 ± 0.58	$5.5 \times 10^{4} \pm 7.3 \times 10^{2}$	$3.8 \times 10^{-4} \pm 3.7 \times 10^{-5}$
ECK3	AGTCAGGCGCTCATTCAAAATGAGGATCGTCATCATCCAA	11.9 ± 1.15	$2.4 \times 10^{4} \pm 9.4 \times 10^{2}$	$2.9 \times 10^{-4} \pm 4.2 \times 10^{-5}$

Epitope estimation and molecular docking

Epitope estimation was performed by incubating the N protein with either ECK1 or ECK4 with a biotin tag, followed by direct addition of trypsin. The presence of aptamer during digestion event was thought to shield the digestion sites in case the epitope contained K or R residues. In this case, trypsin cleaves the protein at that site, rendering epitope incomplete and eliminating the interaction. Following quick 30 min digestion, pulldown on streptavidin agarose beads was performed, and the peptides were directly analyzed with CE-MS. We identified 1 peptide for each of the aptamers: AYNVTQAFGR (267-276) for ECK1 and HIDAYKTFPPTEPKKDK (356-372) for ECK4. We want to point out that ECK4 epitope contains 3 potential cleavage sites in its sequence (K360, K368 and K369). In absence of shielding aptamer, these sequences would become a target for trypsin and this peptide would likely be excluded from analysis, or at the very least, present in very low concentrations. Following identification of the peptides, we performed molecular docking using HADDOCK web server. We folded the aptamers using RNAfold webserver and Matthews DNA model to generate 2D structures. Three dimensional structures were generated by RNAcomposer, provided with the 2D structure from RNAfold. Two step docking was performed, where ECK1 was first docked with N protein, and obtained structure was then subsequently docked with ECK4. Figure 4.7 suggests a molecular model of ECK1 and ECK4 simultaneous binding event to N protein of SARS-CoV-2.

4.4.3 PLA/RT-qPCR

Selection of aptamer pair for PLA

We performed tests with multiple aptamer pairs between ECK1-6 [Figure 4.8 A] and ended up using ECK1 and ECK4 due to the greatest Ct difference in presence of N protein versus in its absence. We hypothesize that other aptamer pairs experience steric hindrance and are unable to interact with 1 molecule of N protein at the same time, possibly due to them sharing same or similar epitope on the protein. It is worth noting that ECK2 and ECK4 pair exhibited positive result as well, however, the Ct difference between sample and control is lower than for ECK1 and ECK4 pair. Other pairs either showed no Ct difference or higher Ct for samples with present N protein. It is possible that for these aptamers, the ligation site is not exposed very well, and in presence of protein, the ligation efficiency is decreased as opposed to expected increase.

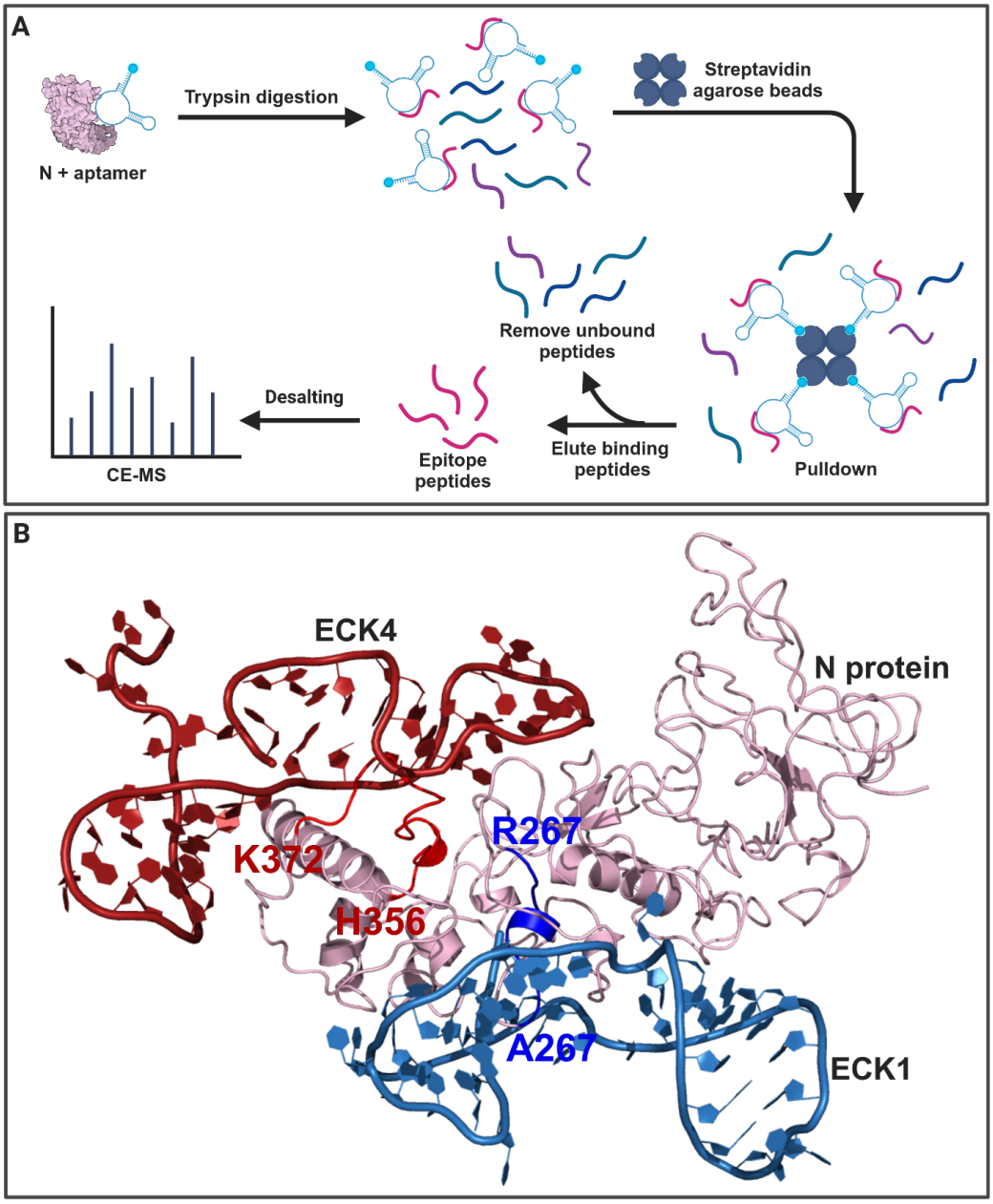

Figure 4.7: Epitope estimation of ECK1 and ECK4 on N protein. A. Schematic of a pulldown experiment to isolate epitope peptides. B. Proposed molecular docking model with both aptamers bound to the protein.

Limit of detection for dual channel detection

To determine the lowest concentration of N protein that is possible to detect with our method, we conducted a series of ligations with varying concentrations of N protein [Figure 4.8 B]. With increasing concentrations of N, the Ct number is going down, as expected with an increasing amount of starting template due to higher efficiency ligation. Concentrations above 20nM resulted in the plateau, making quantification in that range not possible; however, a yes-or-no answer would still be possible as the Ct would always be lower than the blank. For the lower range of concentrations, we performed PLA in presence of as low as 1 nM of N protein, which was still significant compared to the blank.

Ligation time

We performed multiple ligation reactions with identical conditions, varying the reaction time. We determined that prolonged ligation time exhibited almost no effect on the sample with N protein present [Figure 4.8 C]. The Ct between 5 min ligation and 3h ligation is in the range of 24-26 cycles; however, ligation time mattered for the BSA controls. The shortest ligation time for BSA resulted in the highest difference of Ct between N and BSA, while prolonged ligation rendered both BSA and N to have similar Ct values.

Simultaneous detection of N protein and its RNA

We used synthetic RNA standard with N1 sequence for our RT-PCR. Two sets of primers and probes were used, one for the N1 region of the RNA with FAM fluorophore, and another for the ligated region between two aptamers, tagged with Cy5 fluorophore. Amplification revealed that both RNA and N protein detection are possible in 1 vial [Figure 4.9 A]. We aimed to imitate the natural ratio of N protein to the RNA observed in real virus particles by adding 1000 copies of N protein per 1 copy of RNA, which was considered to be 1 single "viral particle" [Figure 4.9 B]. Regardless of RNA presence, both positive and negative ligation products were able to amplify with different Ct values. In positive ligation samples, the Ct values were in the range of 24-26, while in negative ligation samples, the Ct was 30-32. We were able to achieve linearity of response in the range between 6-120 million particles, with LOD calculated to be around 3 million particles [Figure 4.9 C]. It is worth noting that LOD was calculated solely based on the PLA response since negative control also resulted in amplification curve with higher Ct value. We hypothesize LOD of RNA alone to be much lower, however, we do not report it because outside of PLA range, duality of detection is lost.

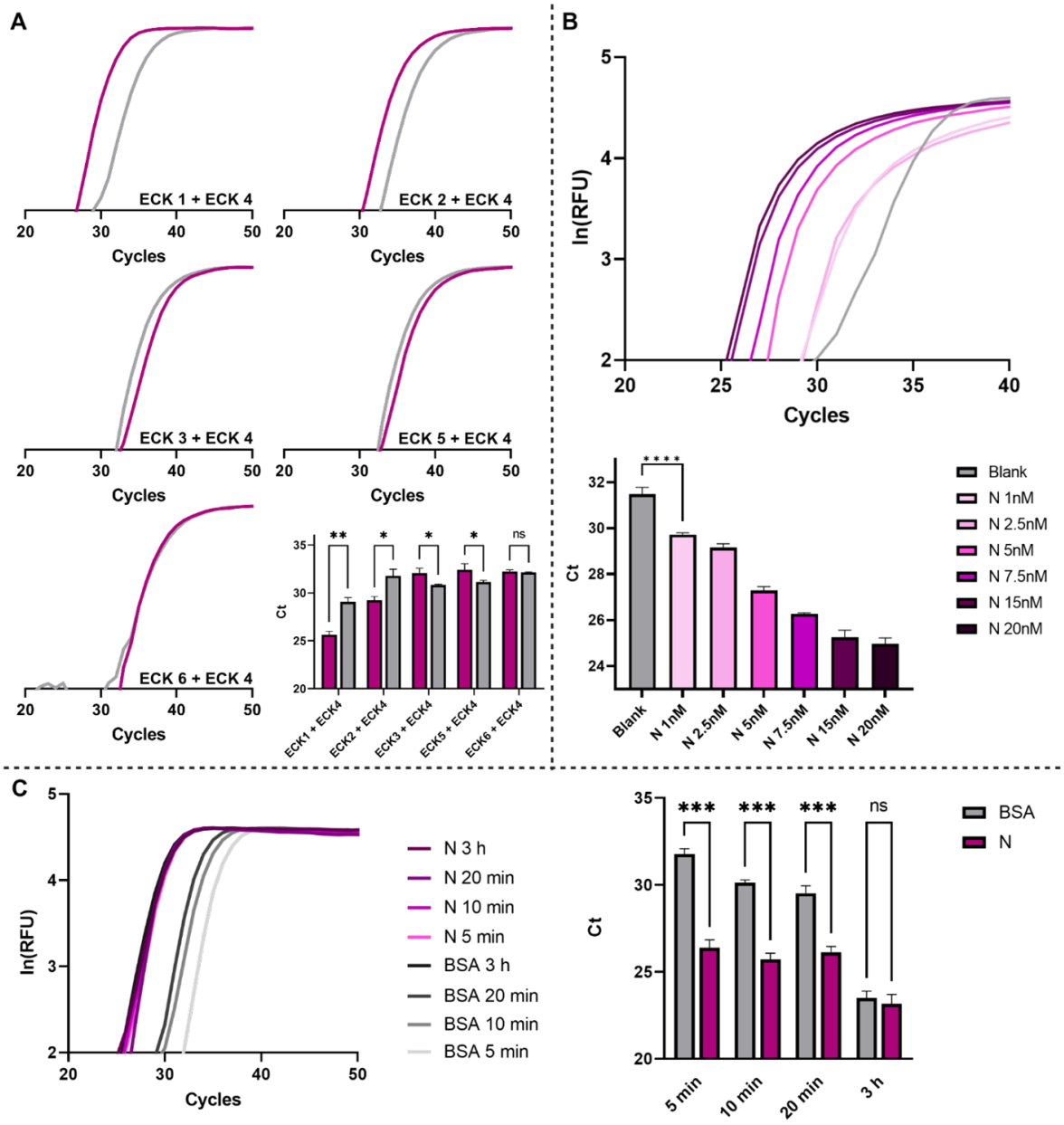

Figure 4.8: Proximity ligation assay panel. A. PLA assay with 5 different aptamer combinations and their respective Ct thresholds with N protein or BSA control. B. qPCR on PLA samples with different concentrations of N protein. Concentrations above 2.5 nM yielded significant difference of Ct threshold in presence of N protein as opposed to BSA. C. Influence of ligation reaction time on the Ct difference between samples and controls. The lowest ligation time of 5 min was the most effective at yielding the highest Ct difference between N protein and BSA ligations. N = 3.

Detection in saliva

While detection of N protein with PLA methodology worked in our buffer system, functionality could be diminished in complex biological matrices, from which the detection is generally desirable. Since SARS-CoV-2 detection is generally performed from saliva of the patient, we picked saliva as our biological matrix suitable for spiking with N protein. We were able to successfully detect 60 million particle equivalents [Figure 4.9 D] in saliva with only 1 centrifugation cleaning step and addition of RNAse inhibitors, suggesting that our method is robust in human saliva. The Ct difference between positive and negative saliva sample was still around 5 cycles, suggesting that saliva matrix does not cause interference with our method.

4.5 Discussion

The library tagged with Cy5 was employed for selection since Cy5 was detectable with both 635 nM laser in our CE system, as well as allowing simultaneous visualization with SYBR gold for the purposes of PCR during the selection process. CE-SELEX was performed in 25 mM ammonium acetate (pH = 9.2) for a few reasons; first, this buffer system is volatile which allows for easy cleanup of the DNA by vacuum evaporation; second, it does not leave behind any salts; and lastly, there is downstream compatibility with CE MS, since analysis by CE-MS for epitope analysis was originally planned at the start of the selection. One of the concerns of using ammonium acetate was the absence of sodium and potassium ions, which are commonly used for most applications of SELEX [142, 143].

Emulsion PCR was employed in the entire SELEX process because high molecular weight by-products were accumulating during regular PCR. CE collections yielded extremely low amounts of DNA due to small injection volumes of around 10 nL. This scarcity of DNA caused accumulation of PCR byproducts in early cycles, partially suppressing amplification of 80 nts desired product. Emulsion PCR has proven effective in amplifying small amounts of DNA to yield a singular 80 nts DNA population [144, 145]. We split the PCR into 2 stages; in stage one, we accumulated dsDNA using emulsion PCR, and in stage two, we performed asymmetric non-emulsion PCR with large amount of dsDNA template yielded in stage one. This approach has allowed us to avoid generation of high molecular weight by-products and reduce the number of cycles to only three. Three rounds of SELEX were performed because the binding was becoming prominent in sub 100 nM concentrations of the protein, and starting with round 4, loss of binding was observed. These three rounds were then sequenced using Illumina next gen sequencing.

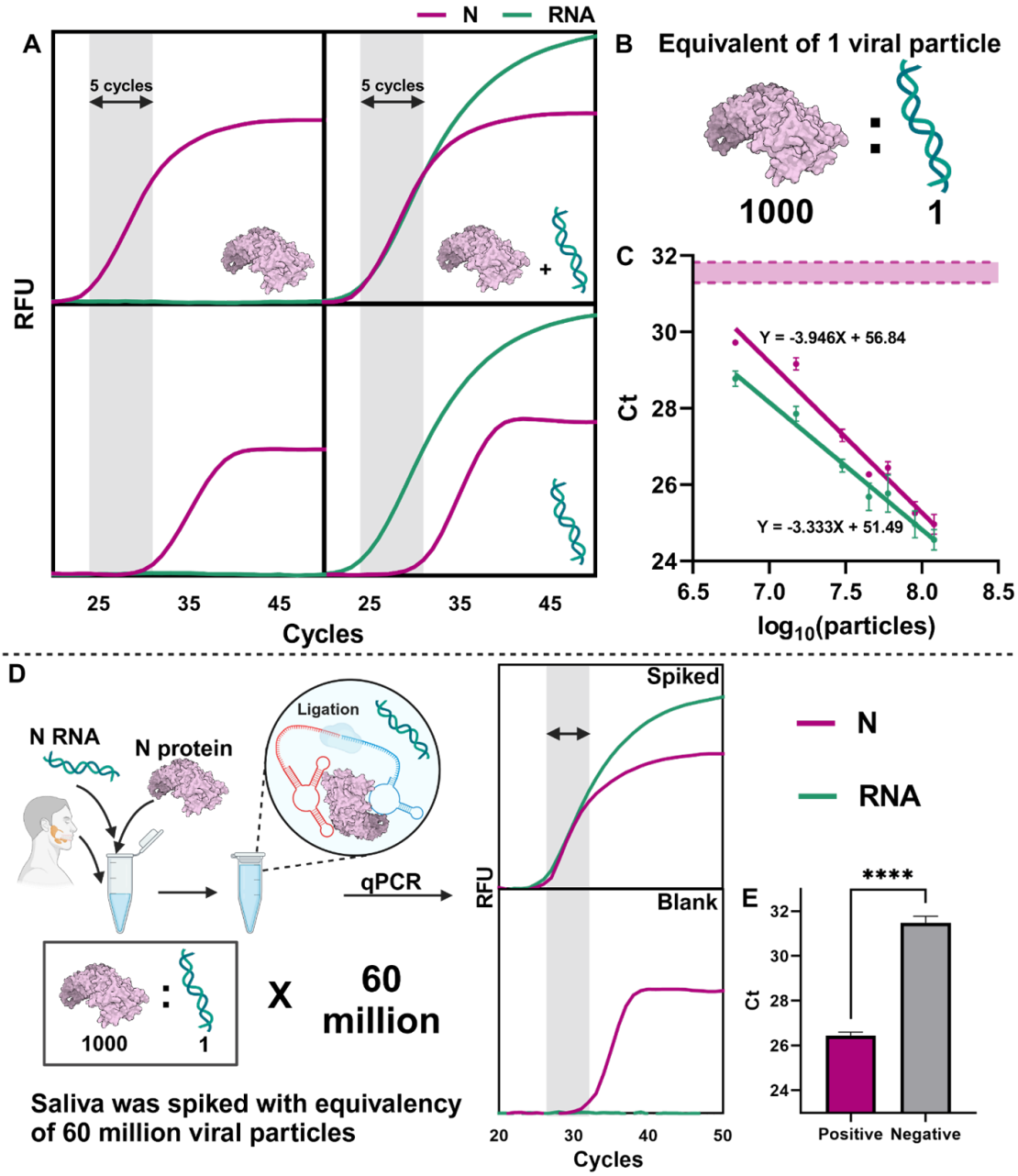

Figure 4.9: Dual channel detection panel. A. Amplification curves of PLA products and RNA. 4 sample groups were amplified with all possible combinations of RNA and protein present. Representative curves were obtained with 10 nM N and 10 pM RNA. B. Ratio of N protein to the RNA was 1000:1 to imitate the natural ratio present in the viral particle. C. Calibration curve for both RNA and protein. Region corresponding to the blank signal highlighted in red is the apparent LOD. D. Dual detection in saliva spiked with RNA and protein following natural viral ratio. An equivalent of 60 million particles was added. E. Ct difference of PLA amplification signal between spiked and non-spiked saliva.

N protein of SARS-CoV-2 was selected in this study due to its high abundance and sequence conservation. N is an RNA binding protein of SARS-CoV-2; thus, it is conserved and makes a good target as it is not prone to mutations. Several articles were published featuring aptamers against N protein [146–149]. One novelty of this work comes from six newly discovered aptamers that interact with N protein with high affinity in a volatile buffer system, offering multiple applications for any future work involving mass spectrometry or applications requiring salt elimination. Another novelty that we claim is the new method of detection of N protein through ligation of 2 aptamers bound to the same protein. While a few studies involving PLA and aptamers already exist [150–152], this study is the first to perform detection of both viral RNA and protein in 1 vial through probe qPCR, where 2 sets of primers, one for RNA and one for protein, are employed. Each set of primers and probe is monitored in separate channels, green and red, allowing for independent monitoring of either PCR reaction.

We aimed to use our aptamers for the detection of viral protein. Aptamer facilitated protein detection has been routinely performed in many applications such as ELASA, flow cytometry or biosensor-based detection [153–155]. Our goal, however, was to use our aptamers for the detection through probe qPCR simultaneously with the RNA detection. To achieve that, we opted for ligation of 2 aptamers that bind simultaneously to one protein molecule, bringing the aptamers into proximity with one another. Two aptamers, one with reverse primer binding site (60 nts) and the other with 5' phosphorylated forward primer (60 nts), were synthesized. We operated under the assumption that primer binding sites would not interfere with the binding, since BLI data suggests that truncated aptamers bind in absence of the primers. 20 nts linker with half of its sequence (10 nts) complementary to the reverse primer binding site, and with the other half complementary to the forward primer, was also synthesized. In absence of N protein, the chance of linker bringing 2 aptamers together was assumed to be lower than in presence of the protein. Compared to the full 20 nts sequence of the linker, its 10 nts half would have a much weaker interaction with its complementary sequence, making this interaction more transient. As such, without of N protein, both aptamers would be free in solution, with occasional linker interaction. However, in presence of the N protein, the aptamers interacting with N protein will spend large amounts of time near one another, allowing their 5' and 3' ends to meet linker a lot more frequently. Ligation time also proved to be important for the experiment. We hypothesize that for prolonged reactions, even withoute of N protein, high degree of ligation is still achieved due transient hybridization of linker to both aptamer's tails. For short reaction times, however, the likelihood of aptamers being in proximity is low, producing lower degrees of ligation. We therefore conclude that it is beneficial to keep ligation time short to achieve the highest difference of Ct values between positive and negative samples.

In presence of T4 ligase, linker and N protein phosphorylated 5' of ECK4 and 3' end of ECK1 would come in contact and ligate more often than in absence of N protein, producing 120 nts ligated product. This product can then be amplified and detected using probe qPCR, where the probe is complementary to the junction site between 2 aptamers and would not hybridize if ligation event did not happen. Moreover, reverse primer of this assay can only hybridize to reverse complementary sequence of the ligation product, meaning that in absence of ligation, only half of the product will be synthesized during PCR with linear amplification as opposed to exponential.

Our PLA method can detect as low as 500 pg of N protein when using ECK1 and ECK4 aptamers pair, including detection in complex matrices such as saliva. A study reported that per one virion particle, 1000-2000 copies of N protein are present [156]. Theoretically, our method, therefore, can detect around 3 million virion particles with high confidence. The PLA approach is an excellent orthogonal method of detection; however, it has its own disadvantages. First, quantification above 25 nM, or 150 million virions becomes problematic. While the method indicates whether the protein is present or absent, the quantification of the viral load remains problematic. Above 25 nM of N, Ct values will reach diminishing returns and will not reduce further, thus making 25 nM and, for instance, 250 nM virtually indistinguishable. For clinical application, however, a yes or no answer would provide valuable information.

Second, for the lower end of the concentration range, detection is poor below 500 pM or 3 million virion particles. While qPCR is one of the most sensitive analytical techniques, this method does not start with a cDNA template present at low concentration. Rather, protein-DNA interaction is necessary for the ligation reaction to occur, which is not as sensitive as amplification of a few copies of cDNA. Nonetheless, detection of both N protein and RNA encoding for N would provide invaluable information, especially in the early stages of infection, when large number of virion particles are still present and carry N protein in them.

Quantification in saliva was the final goal since saliva is the matrix routinely collected in the clinical setting for SARS-CoV-2 detection. Since saliva is a complicated matrix, introduction of RNAse inhibitors is mandatory due to abundance of RNAses. In absence of inhibitors, the RNA signal is lost completely, however, the PLA signal is unaffected. We suggest that detection of both protein and RNA in one reaction provides higher confidence of diagnosis as RNA detection alone is prone to false positive results [157,158]. The amplification of both products, however, should increase detection confidence. A single greatest drawback of our method, however, is the narrow linear range, making quantification problematic should the viral load be outside of this range.

For future developments of the method, it might be beneficial to design primers not requiring ligation step, therefore, allowing for polymerase to amplify from DNA directly and from lower quantities. While PLA has been demonstrated to work, the limits of detection still depend largely on the ligation reaction. Additionally, introduction of surface for ligation reaction might improve LOD due to ability to wash the unbound aptamers, reducing the overall false positive signal.

4.6 Conclusion

Our goal was to develop a novel method for SARS-CoV-2 detection using aptamers. To achieve that, we aimed to discover at least two DNA aptamers binding simultaneously to different peptide epitopes of N protein. This would bring these aptamers in proximity in the solution and enable a ligation event between these two aptamers, which will produce a single long DNA molecule. The following PCR amplification and real-time detection of this ligated DNA molecule would indicate the presence of N protein in solution as opposed to the absence of protein, where the ligation would occur at lower efficiency. By employing PLA/RT-qPCR, we can indirectly detect viral protein by detecting ligated DNA. This approach offers a novel method of detection of SARS-CoV-2, where the PLA product is amplified alongside of viral genome in 2 separate channels, allowing detection of both nucleoprotein and RNA simultaneously.

4.7 Acknowledgements

Conceptualization, Maxim Berezovski, Emil Zaripov and Carlos Gu; Methodology, Emil Zaripov, Yuchu Dou, Carlos Gu; Experimental, Emil Zaripov, Carlos Gu, Elnaz Yaghoobi, Petr Kasyanchyk and Abdullah Khraibah; Data Analysis, Emil Zaripov and Carlos Gu, Resources, Maxim Berezovski; Data Curation, Emil Zaripov; Writing—Original Draft Preparation, Emil Zaripov; Writing—Review & Editing, Maxim Berezovski, Emil Zaripov and Carlos Gu; Visualization, Emil Zaripov and Carlos Gu; Supervision, Maxim Berezovski; Project Administration, Maxim Berezovski and Emil Zaripov; Funding Acquisition, Maxim Berezovski.

Chapter 5

Chapter 5: Instrumentation and methodology improvement

5.1 Introduction

In this chapter, we present several key advancements in the instrumentation and methodology of capillary electrophoresis (CE) that have significantly enhanced its capabilities and applications. Specifically, we describe the development of a novel electrospray ionization (ESI) interface for mass spectrometry (MS) detection, a novel normalization strategy to the total injected UV signal using an online UV detector, and the establishment of a comprehensive pipeline for metabolomic sample preparation and identification.

The integration of CE with MS is a powerful analytical technique that provides exceptional separation efficiency and direct structural information. To further improve the coupling of CE and MS, we have designed and implemented a new nano sheath liquid ESI interface, leading to enhanced sensitivity and robustness.

Additionally, we introduce a novel normalization strategy that utilizes an online UV detector to adjust the CE-MS signal to the total injected sample. This approach addresses the inherent challenges of sample-to-sample variation in CE, enabling more accurate quantitative analysis and improved data quality.

Finally, we present a comprehensive pipeline for preparing and identifying metabolomic samples in the context of CE-MS analysis. This includes optimized sample extraction, desalting, and separation protocols, as well as the development of robust data processing and compound identification workflows.

The advancements described in this chapter contribute to the ongoing evolution of capillary electrophoresis, expanding its analytical capabilities and opening new avenues for its application in various fields, including metabolomics, proteomics, and beyond.

5.2 Modification of ESI interface

The interface between capillary electrophoresis (CE) and mass spectrometry (MS) plays a critical role in achieving high-performance analysis and detection of analytes. The conventional CE-MS interface utilizes a sheath fluid as the interface medium, which requires a high flow rate (3 mL/min) of sheath liquid; however, this conventional approach presents several limitations and challenges that compromises the overall performance of the CE-MS system. We developed a custom CE-MS interface using stainless steel and PETG polymer to address many of these limitations [Figure 5.1].

5.2.1 Limitations of the Conventional CE-MS Interface

Conventional Sheath Fluid

The use of a conventional sheath fluid in the CE-MS interface necessitated a high flow rate, which posed practical challenges and limited the efficiency of the system. The requirement for a large volume of sheath liquid at high flow rates increased operational costs and introduced additional complexity to the system, while reducing sensitivity and reproducibility.

Nebulizing Gas Requirement

To maintain electrospray ionization (ESI) in the conventional interface, a constant nebulizing gas stream was required. This gas stream was necessary to create a fine aerosol of the sample solution and facilitate ionization. However, the continuous supply of nebulizing gas resulted in an unstable CE current, leading to poor migration time reproducibility between runs. The intermittent interruptions caused by the nebulizing gas flow impacted the overall accuracy and reliability of the CE-MS analysis.

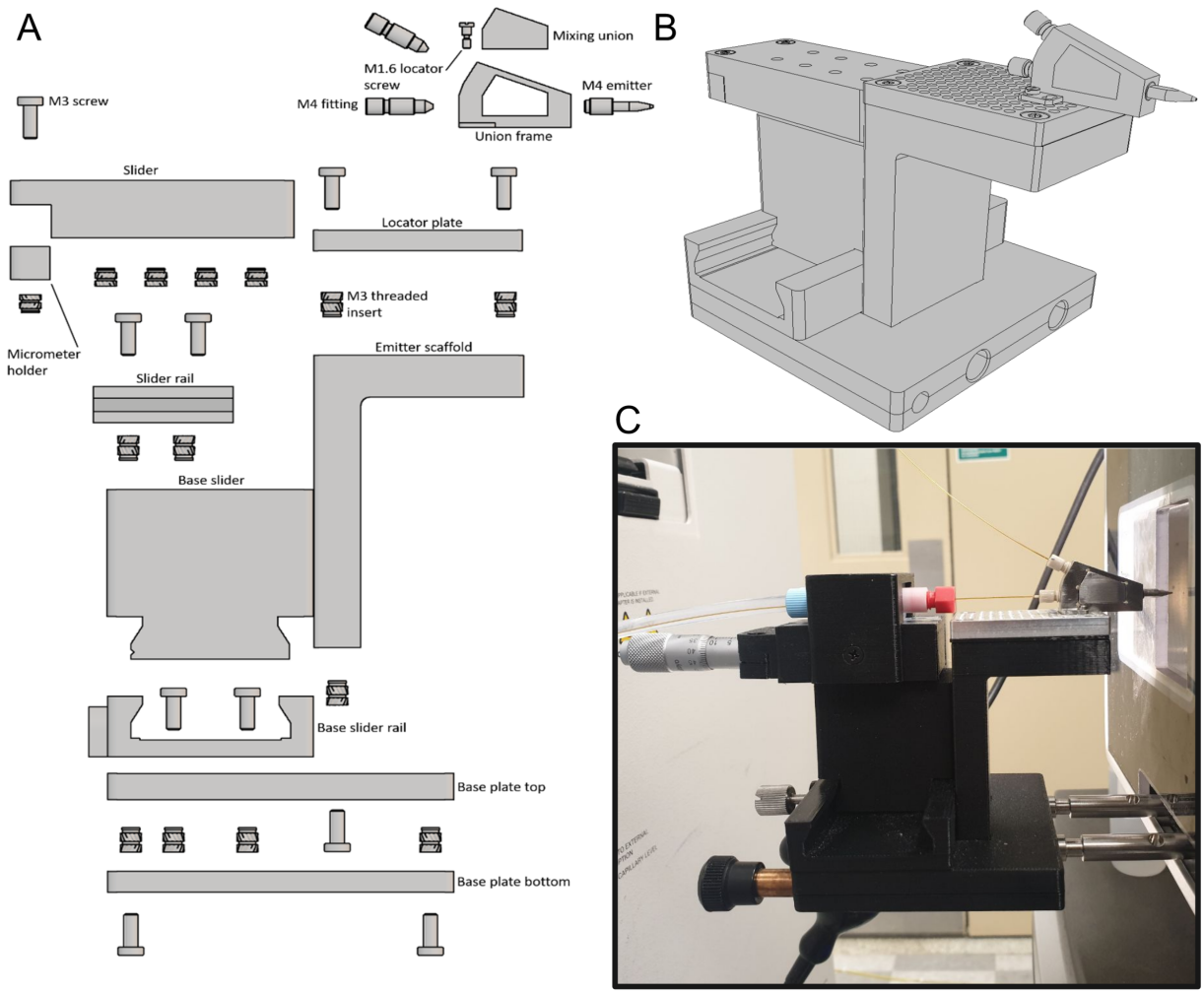

Figure 5.1: Structural diagrams of the modified nano sheath flow interface. A. Exploded view of the interface outlining all components used in the fabrication process. B. Assembled isometric overview of the interface. C. Photo of the interface from the side.

5.2.2 Limitations of the Positioning Table

The positioning table in the conventional CE-MS interface was equipped with a crude micrometer, which caused issues during operation. The vacuum pump vibrations induced slow movement of the capillary spraying needle mid-run, resulting in even poorer repro-

ducibility and ionization efficiency. This lack of stability and precision in the positioning table hindered the accurate alignment and positioning of the capillary, adversely affecting the overall performance of the CE-MS system.

5.2.3 Proposed Modifications to the CE-MS Interface

Considering the limitations associated with the conventional CE-MS interface, a new design was undertaken with the aim of improving system performance. Several key modifications were implemented, taking into consideration the need for rigid stability, reduced sheath liquid consumption, and enhanced ionization efficiency.

Rigid Stability of the Spraying Unit

One of the primary objectives of the interface redesign was to achieve rigid stability of the spraying unit, minimizing any movement even in the presence of vibrations. By employing fixed positioning surface with a grid of threaded positions, the spraying unit is no longer connected to any micrometer, and instead is anchored to a selected place on the locator plate. The plate itself as well as the mixing unit are manufactured out of stainless steel, while everything else including the sliding adjustors were 3D printed from PETG polymer [Figure 5.1]. The new interface ensures stable alignment and positioning of the capillary, thereby enhancing reproducibility and minimizing ionization variability.

Tapered Steel Emitter

To overcome the issue of sheath liquid buildup on the capillary spraying needle, a tapered end was incorporated into the design. The tapered steel emitter facilitates a smoother flow of the sheath liquid, preventing the accumulation of liquid and resulting in improved performance and reduced signal fluctuations. The needle itself also contains internal taper reducing from 0.79 mm (conventional 1/32-inch PEEK sleeves) to 0.36 mm (standard fused silica tubing outer diameter), allowing to protect the capillary with a sleeve until the sheath liquid makes contact about 2 mm before the capillary tip.

Elimination of Nebulizing Gas and Reduced Flow Rate

The conventional reliance on nebulizing gas was eliminated in the modified interface design. Instead, the flow rate of the sheath liquid was significantly reduced to the nano

range, approximately 400 nL/min. This reduction in flow rate not only stabilized the current, minimizing signal fluctuations, but also enhanced the ESI process due to the reduced volume of ionized liquid. The lower flow rate also led to reduced operational costs and simplified the experimental setup. Lower sheath liquid requirement also allowed for miniaturization of the capillary internal diameter, making 25 μm I.D. capillaries possible as opposed to the conventional I.D. of 75 or 50 μm.

5.2.4 Experimental Validation and Results

The modified CE-MS interface was subjected to rigorous experimental validation to evaluate its performance and compare it with the conventional interface. Extensive analyses were performed to assess reproducibility, ionization efficiency, and overall system stability. As a result, we can confidently state that the system is suitable for both proteomic and metabolomics analysis of complex mixtures, providing excellent detection sensitivities. The modified interface also demonstrated a significant improvement in migration time reproducibility between runs compared to the conventional interface.

The analysis of a standard mix comprising 20 amino acids and 8 nucleic acids using capillary electrophoresis coupled to mass spectrometry (CE-MS) demonstrated promising results [Figure 5.2]. Without any normalization applied, the base peak intensity electropherograms showed excellent peak intensity reproducibility across multiple injections. The migration time difference observed between four consecutive runs was approximately 2 minutes out of a 15-minute total analysis time, indicating good temporal stability of the system. Furthermore, the peak areas exhibited exceptional reproducibility, with a CV of $\sim 5\%$. These findings suggest that the CE-MS platform provides highly consistent and reliable performance in the separation and detection of the targeted amino acids and nucleic acids, laying a robust foundation for further applications and method development.

The redesign of the CE-MS interface addressed the limitations of the conventional system and resulted in improvements in performance. The rigid stability of the spraying unit, the incorporation of a tapered steel emitter, and the elimination of nebulizing gas led to reproducibility under nano flow of sheath liquid. The modifications presented in this study provide a promising solution for overcoming the challenges associated with the conventional CE-MS interface.

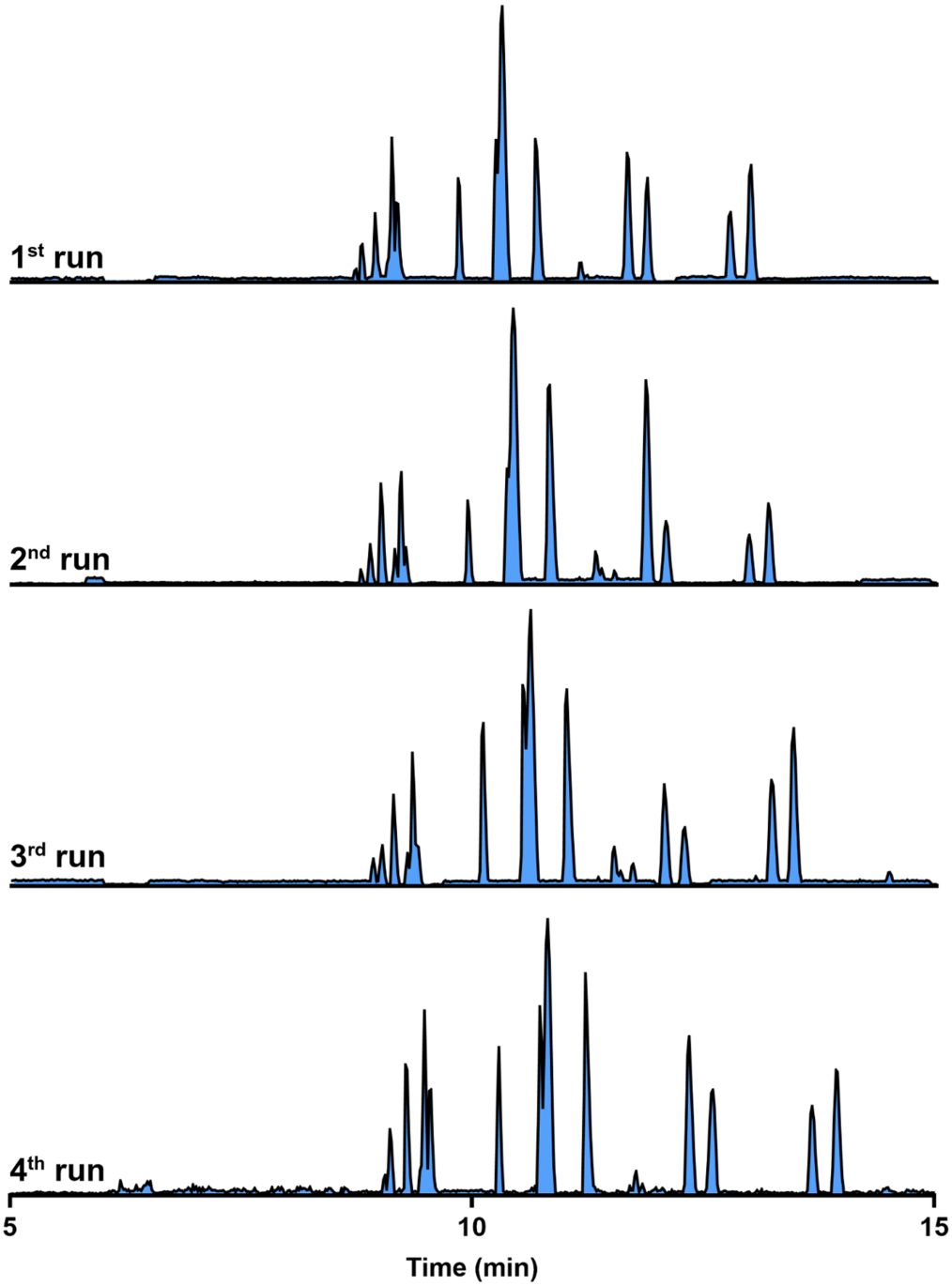

Figure 5.2: Repeated injections of the cationic standard mix.

5.3 Quantitative normalization to UV absorbance

In CE analysis, injection plug volumes can vary between injections due to different viscosities of samples and evaporation during long sequences. Eventual evaporation leads to poor reproducibility of peak areas, making analysis more difficult. Evaporation could be prevented using mineral oil on the top of the samples at the cost of potential metabolite loss, introduction of contaminants or inability to salvage the sample for further use. To maintain the reproducibility of peak areas, we introduced a new online normalization strategy to a total injected UV signal [Figure 5.3 A]. This method was the basis of all EV sample normalization in Chapter 3. Therefore, the metabolome of EV data was employed. The UV normalization method showed improvement in peak area discrepancy between injections. Since the UV signal of total injected sample is directly proportional to the amount of sample injected, as well as the total MS response, we deemed this a feasible normalization method. An online PDA detector was introduced 10 cm away from the inlet, allowing for the detection of the sample plug early on during the start of the run [Figure 5.3 B]. Since the detector is placed so early on, separation does not happen, and all the compounds are detected as one peak. For injection normalization, all the MS1 peak areas were divided by the total peak area of the UV. This strategy allowed us to account for sample evaporation during long runs and improve precision. We selected 280 nm wavelength due to high baseline fluctuations at the lower wavelength, making quantification problematic. While 280 nm does omit certain metabolites that lack aromatic structures, total metabolome is still expected to have a variety of compounds with the structures capable of absorbing 280 nm light. The entire matrix of molecular features was selected as opposed to a narrow list of putatively identified compounds for normalization and estimation of CV. The rationale for this stemmed from the fact that injection does not include EV metabolites only, but also compounds collected during the process of sample preparation. These compounds contribute to the absorbance signal as well, and if omitted, they would introduce additional error. Since the real proportion of EV compounds and irrelevant compounds is unknown, it would be problematic to calculate the CV based on just the EV compounds alone. In addition, MCF-10A and MDA-MB-231 produce different quantities of EVs which reflect the number of SEC fractions collected, as well as the SEC injection volume. This difference could be the reason for different degrees of efficiency of this normalization method when applied to MCF-10A sample matrix as opposed to MDA-MB-231.

The performance of this normalization method was investigated by estimating the total coefficient of variance (CV) for each cell line before and after normalization to a total 280 nm signal [Figure 5.3 C]. The entire matrix of molecular features prior to identification was employed, which included features that were not part of EV metabolome. For EVs from both cell lines, median CV lowered overall, leading to more credible peak areas used in further analyses. MCF-10A demonstrated greater difference in CV before and after normalization, reducing discrepancy of the features that had CV greater than 150%. The median was shifted from 115% to 78%, compared to non-normalized data. For MDA-MB-231, however, the change was smaller, with the median shifting from 55% to 48%. While there is a quantitative difference, the overall change was lower compared to MCF-10A. In addition, high discrepancy molecular features (CV greater than 150%) exhibited almost no difference between normalized and non-normalized data. Implementation of this normalization technique improved run-to-run variation issues related to peak quantification.

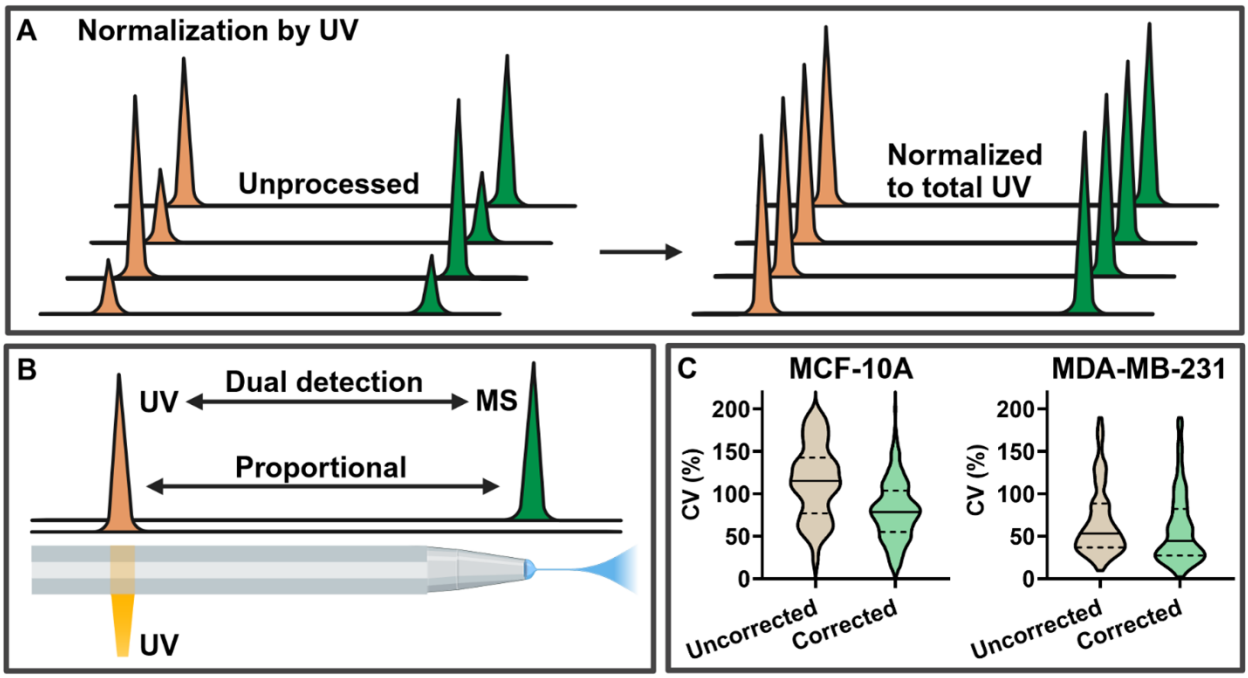

Figure 5.3: Online normalization to a total UV signal. A. Schematic of normalization concept. B. Instrumental setup permitting online UV detection. C. Violin plots depicting lowered mean coefficient of variance following the normalization.

5.4 Identification pipeline

To address the challenges associated with matrix effects in CE-MS analysis, it is crucial to ensure clean and reproducible sample preparation procedures. We implemented a highly efficient and standardized extraction method using a polysulfoethyl aspartamide resin-based cation exchange [159] micro SPE (solid-phase extraction) tip. This approach allowed for the reproducible enrichment of cationic molecules, which are particularly well-suited for subsequent capillary electrophoresis (CE) separation. By employing the cation exchange micro SPE tip extraction, we achieved enhanced purification and concentration of the target metabolites, thereby improving the sensitivity and accuracy of their subsequent analysis by CE-MS.

To perform compound identifications and further standard validations, we performed an experiment to build a spectral library for identifications. All the samples were mixed into one quality control (QC) sample with a high concentration of metabolites [Figure 5.4 A]. The QC sample was injected, and migration time was then converted to mobility scale with ROMANCE algorithm using known metabolites in the sample or internal standards. The library sample underwent AIF (All Ion Fragmentation) MS2 pipeline in MSDIAL followed by MSFINDER for peak picking and identification, respectively [Figure 5.4 B-C]. The resulting array of peak identities with top ranking candidates as putative identifications was then referred to as library, and all the subsequent samples from the same matrix were queried against it to perform identification. To perform validation, we injected molecular standards into CE-MS, and queried these runs against the matrix library [Figure 5.4 D]. In case of a match, the standard was recognized as a metabolite from the spectral library and was considered validated [Figure 5.4 E].

We chose AIF approach due to generally low quantities of metabolites in biological matrices, to not lose any ions during selective quadrupole transmittance. As CE is a non gradient technique, the solution reaching ESI emitter would be the same regardless of the stage of the run, leading to a very uniform background that can be easily subtracted. Unlike LC, where the columns can leach multiple compounds and make AIF problematic, in CE the AIF spectra are easier to process, while still allowing for detection of all ions that get ionized.

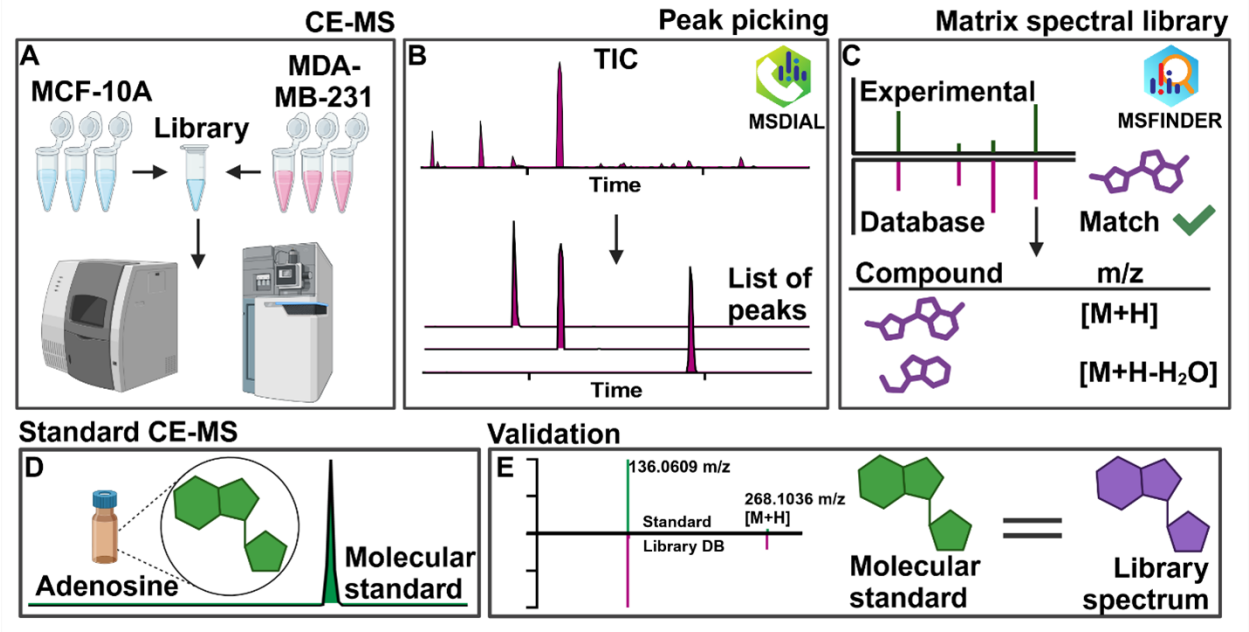

Figure 5.4: Sample processing and identification pipeline. A. Matrix library is generated through QC sample. B. Peak list generation using MSDIAL. C. Peak identifications using DIA MS2 in MSFINDER. D. Molecular standard injection for validation. E. Query of molecular standard against the spectral database.

5.5 Conclusion

CE-MS encounters certain challenges that can impede accurate identification and quantification of metabolites [160, 161]. Among these are observed discrepancies in migration time between different injections, which can hinder the reproducibility and reliability of metabolomic studies [162–164]. Another issue is the lower sensitivity of detection of the sheath fluid interfaces due to dilution of the analyte, which restricts analysis of low abundant metabolites [165–167]. Furthermore, electrophoretic separation itself is subject to matrix effects of the injection plug, requiring precise injection amounts to negate the discrepancies [168]. All these issues have been addressed in this chapter, where improvements were made in multiple areas. As such, a new ESI interface has been made, enabling nanoflow and high reproducibility run-to-run. A UV detector has been connected online, allowing for normalization of the injected sample to a total UV signal, improving quantification. Lastly, an optimized pipeline for small molecule sample preparation and analysis

has been made, allowing for robust metabolite identification. All these improvements have been used throughout the chapters of these thesis, enabling high quality data to be collected.

Chapter 6

Conclusion

This thesis sought to advance multiple areas of bioanalytical chemistry through innovative applications of capillary electrophoresis coupled with various detectors. Capillary electrophoresis is a versatile separation technique with continuous improvements being made to its analytical capabilities. The work presented in this thesis explored new fronts for capillary electrophoresis including metabolomics, phytochemical analysis, and antiviral development. Each study built upon the latter to demonstrate deeper exploration of important challenges in these domains. The first research chapter focused on establishing a foundational CE-UV method for cannabinoid analysis in cannabis. Subsequent chapters then progressively increased the scope and sophistication of capillary electrophoresis coupling it with mass spectrometry for metabolite profiling and employing it for aptamer selection against SARS-CoV-2.

The first research chapter presented a CE with UV detection method, which was successfully developed and optimized to achieve baseline separation of 14 major cannabinoids within a run time of less than 20 minutes. This provided a simple yet powerful method for quality control and quantitative analysis in the growing cannabis industry. The high-resolution separation of this challenging mixture of structurally similar compounds highlighted the effectiveness of CE for analyzing small molecules in complex matrices.

The second research chapter delved into metabolomics by utilizing CE-MS for the untargeted analysis of extracellular vesicles. Through coupling the excellent separation power of CE with the identification capabilities of high-resolution MS, dozens of metabolites were detected and relatively quantified between cancerous and non-cancerous EV samples. Two pathways were hypothesized to be altered between cancerous EVs and non-cancerous. Several compounds showed potential as novel cancer biomarkers, demonstrating how the

combined techniques of CE-MS can uncover disease relevant molecular information from basic research discoveries.

The third research chapter proceeded to apply CE and CE-MS in developing antiviral strategies by isolating DNA aptamers targeting the SARS-CoV-2 nucleocapsid protein. Aptamers with high affinity and specificity were selected from a library using CE alone. Then, to attain structural insights, CE-MS was employed to pinpoint the precise binding epitope on the protein surface. This validated the functionality and clinical utility of the identified aptamers. These aptamers were then developed into rapid diagnostic assay using RT-PCR and proximity ligation phenomena, allowing for dual detection of viral RNA and nucleocapsid simultaneously in 2 different channels. Overall, this chapter demonstrated how aptamer selection through CE can enable directed diagnostic development, with CE-MS providing invaluable molecular details.

The research chapters were followed by a descriptive section on improvement of CE and CE-MS instrumentation, as well as the sample processing pipeline. First, it presented a manufactured nano flow interface for CE-MS, which allows coupling of CE to MS through ESI. This interface was the basis of all work related to metabolomics and aptamer epitope identification in chapters 3 and 4, respectively. Second, the presented the heavily modified CE instrument that allowed for online detection of analytes through UV absorbance during CE-MS run. This modification allowed for a normalization strategy to UV signal, which has been experimentally demonstrated to reduce the variability between injections. Lastly, it demonstrated the development of the metabolomics pipeline from sample preparation to metabolite identification, underlining CE-MS as self-sufficient platform for complex techniques like metabolomics.

In conclusion, this body of work exemplified how capillary electrophoresis coupled with suitable detection techniques, whether UV, MS, or other options, enables high resolution analysis across diverse applications in biomedicine, phytochemistry, and more. The techniques proved adaptable from basic research uses like metabolite profiling to larger pursuits including biomarker discovery, quality control assays, and targeted antiviral development. This thesis thereby highlighted the remarkable versatility and growing importance of capillary electrophoresis and affiliated technologies.

Chapter 7

List of publications

Hüttmann, N., Li, Y., Poolsup, S., Zaripov, E., D'Mello, R., Susevski, V., Minic, Z., and Berezovski, M. Surface Proteome of Extracellular Vesicles and Correlation Analysis Reveal Breast Cancer Biomarkers. Cancers (Basel) 16, 520 (2024).

Uppal, G. K., Poolsup, S., Zaripov, E., Gu, Y. & Berezovski, M. Comparative analysis of aptamers binding to SARS-CoV-2 N protein using capillary electrophoresis and bio-layer interferometry. *Anal Bioanal Chem* 416, 1697–1705 (2024).

Muratov, K., Zaripov, E., Berezovski, M. & Gagosz, F. DFT-Enabled Development of Hemilabile (P ∧ N) Ligands for Gold(I/III) RedOx Catalysis: Application to the Thiotosylation of Aryl Iodides. *J Am Chem Soc* 146, 3660–3674 (2024).

Poolsup, S., Zaripov, E., Hüttmann, N., Minic, Z., Artyushenko, P. V., Shchugoreva, I.A., Tomilin, F.N., Kichkailo, A.S., and Berezovski, M. Discovery of DNA aptamers targeting SARS-CoV-2 nucleocapsid protein and protein-binding epitopes for label-free COVID-19 diagnostics. *Mol Ther Nucleic Acids* 31, 731–743 (2023).

Minic, Z., Hüttmann, N., Poolsup, S., Li, Y., Susevski, V., Zaripov, E., and Berezovski, M. Phosphoproteomic Analysis of Breast Cancer-Derived Small Extracellular Vesicles Reveals Disease-Specific Phosphorylated Enzymes. *Biomedicines* 10, 408 (2022).

Zaripov, E., Lee, T., Dou, Y., Harris, C.S., Egorov, A., and Berezovski, M. Single-Run Separation and Quantification of 14 Cannabinoids Using Capillary Electrophoresis. *Separations* 8, 30 (2021).

Dou, Y., Ren, L., Kulabhusan, P., Zaripov, E. & Berezovski, M. Quantitative Capillary Electrophoresis for Analysis of Extracellular Vesicles (EVqCE). *Separations* 8, 110 (2021).

Grechkin, Y.A., Grechkina, S.L., Zaripov, E.A., Fedorenko, S. V., Mustafina, A.R., and Berezovski, M. Aptamer-Conjugated Tb(III)-Doped Silica Nanoparticles for Luminescent Detection of Leukemia Cells. *Biomedicines* 8, 14 (2020).

References

[1] D.-Q. Tang, L. Zou, X.-X. Yin, and C.N. Ong. Hilic-ms for metabolomics: An attractive and complementary approach to rplc-ms. *Mass spectrometry reviews*, 35(5):574–600, 2016.

[2] D. Berek. Critical assessment of size exclusion chromatography. *Polimery/Polymers*, 49(5):313–320, 2004.

[3] D. Berek. Size exclusion chromatography - a blessing and a curse of science and technology of synthetic polymers. *Journal of Separation Science*, 33(3):315–335, 2010.

[4] H. Wätzig and S. Günter. Capillary electrophoresis - a high performance analytical separation technique. *Clinical Chemistry and Laboratory Medicine*, 41(6):724–738, 2003.

[5] A. Stolz, K. Jooß, O. Höcker, J. Römer, J. Schlecht, and C. Neusüß. Recent advances in capillary electrophoresis-mass spectrometry: Instrumentation, methodology and applications. *Electrophoresis*, 40(1):79–112, 2019.

[6] W. Zhang, T. Hankemeier, and R. Ramautar. Next-generation capillary electrophoresis–mass spectrometry approaches in metabolomics. *Current Opinion in Biotechnology*, 43:1–7, 2017.

[7] M. Buszewska-Forajta, M.J. Markuszewski, and R. Kaliszan. Free silanols and ionic liquids as their suppressors in liquid chromatography. *Journal of Chromatography A*, 1559:17–43, 2018.

[8] Jorge E. Gómez and Junior E. Sandoval. The effect of conditioning of fused-silica capillaries on their electrophoretic performance. *Electrophoresis*, 29(2):381–392, January 2008.

[9] Sandip Ghosal. Fluid mechanics of electroosmotic flow and its effect on band broadening in capillary electrophoresis. *Electrophoresis*, 25(2):214–228, January 2004.

[10] Yuejun Kang, Chun Yang, and Xiaoyang Huang. Dynamic aspects of electroosmotic flow in a cylindrical microcapillary. *International Journal of Engineering Science*, 40(20):2203–2221, December 2002.

[11] Xiangchun Xuan, Bo Xu, David Sinton, and Dongqing Li. Electroosmotic flow with joule heating effects. *Lab on a Chip*, 4(3):230–236, May 2004.

[12] Charles A. Lucy, Amy M. MacDonald, and Makedonka D. Gulcev. Non-covalent capillary coatings for protein separations in capillary electrophoresis. *Journal of Chromatography A*, 1184(1–2):81–105, March 2008.

[13] Jeremy E. Melanson, Nicole E. Baryla, and Charles A. Lucy. Dynamic capillary coatings for electroosmotic flow control in capillary electrophoresis. *TrAC Trends in Analytical Chemistry*, 20(6–7):365–374, June 2001.

[14] M. Weinbauer and H. Stutz. Successive multiple ionic polymer layer coated capillaries in the separation of proteins - recombinant allergen variants as a case study. *Electrophoresis*, 31(11):1805–1812, 2010.

[15] C.A. Lucy, A.M. MacDonald, and M.D. Gulcev. Non-covalent capillary coatings for protein separations in capillary electrophoresis. *Journal of Chromatography A*, 1184(1–2):81–105, 2008.

[16] D. Belder, A. Deege, H. Husmann, F. Kohler, and M. Ludwig. Cross-linked poly(vinyl alcohol) as permanent hydrophilic column coating for capillary electrophoresis. *Electrophoresis*, 22(17):3813–3818, 2001.

[17] M. Beneito-Cambra, P. Anres, J. Vial, P. Gareil, and N. Delaunay. Stability and effectiveness of linear polyacrylamide capillary coating to suppress eof in acidic media in the presence of surfactants, ionic liquids and organic modifiers. *Talanta*, 150:546–552, 2016.

[18] F. Steiner and M. Hassel. Control of electroosmotic flow in nonaqueous capillary electrophoresis by polymer capillary coatings. *Electrophoresis*, 24(3):399–407, 2003.

[19] K. Adachi, N. Noda, M. Nakashige, S. Tsuneda, and T. Kanagawa. Affinity capillary electrophoresis with a dna-nanoparticle conjugate as a new tool for genotyping. *Journal of Chromatography A*, 1109(2):127–131, 2006.

[20] N. Li, S. Zeng, L. He, and W. Zhong. Probing nanoparticle-protein interaction by capillary electrophoresis. *Analytical Chemistry*, 82(17):7460–7466, 2010.

[21] N. Na, Y. Hu, J. Ouyang, W.R.G. Baeyens, J.R. Delanghe, Y.E.C. Taes, M. Xie, H. Chen, and Y. Yang. On the use of dispersed nanoparticles modified with single layer -cyclodextrin as chiral selecor to enhance enantioseparation of clenbuterol with capillary electrophoresis. *Talanta*, 69(4):866–872, 2006.

[22] S.N. Krylov. Nonequilibrium capillary electrophoresis of equilibrium mixtures (neceem): A novel method for biomolecular screening. *Journal of Biomolecular Screening*, 11(2):115–122, 2006.

[23] Y. Chen, W. Lü, X. Chen, and M. Teng. Review of recent developments of on-line sample stacking techniques and their application in capillary electrophoresis. *Central European Journal of Chemistry*, 10(3):611–638, 2012.

[24] D.-S. Lian, S.-J. Zhao, J. Li, and B.-L. Li. Progress in stacking techniques based on field amplification of capillary electrophoresis. *Analytical and Bioanalytical Chemistry*, 406(25):6129–6150, 2014.

[25] X.-M. Yang, S.-H. Zhang, C. Wang, and Z. Wang. Development of micellar to solvent stacking on-line preconcentration technique in capillary electrophoresis. *Fenxi Huaxue/ Chinese Journal of Analytical Chemistry*, 41(12):1939–1946, 2013.

[26] W. Zhang and C. Cao. A review on stacking of analytes in high salt sample in capillary electrophoresis. *Fenxi Huaxue*, 33(2):267–271, 2005.

[27] B. Cetin and D. Li. Effect of joule heating on electrokinetic transport. *Electrophoresis*, 29(5):994–1005, 2008.

[28] C.J. Evenhuis and P.R. Haddad. Joule heating effects and the experimental determination of temperature during ce. *Electrophoresis*, 30(5):897–909, 2009.

[29] A.S. Rathore. Joule heating and determination of temperature in capillary electrophoresis and capillary electrochromatography columns. *Journal of Chromatography A*, 1037(1–2):431–443, 2004.

[30] M. Donegan and M. Browning. A review recent developments in sample ionization interfaces used in mass spectrometry. *Journal of Liquid Chromatography and Related Technologies*, 35(16):2345–2363, 2012.

[31] M. Wilm. Principles of electrospray ionization. *Molecular and Cellular Proteomics*, 10(7), 2011.

[32] B. Challen and R. Cramer. Advances in ionisation techniques for mass spectrometry-based omics research. *Proteomics*, 22(15–16), 2022.

[33] J.B. Fenn, M. Mann, C.K. Meng, S.F. Wong, and C.M. Whitehouse. Electrospray ionization for mass spectrometry of large biomolecules. *Science*, 246(4926):64–71, 1989.

[34] Q.A. Farooq, N. ul Haq, A. Aziz, S. Aimen, and M.I. ul Haq. Mass spectrometry for proteomics and recent developments in esi, maldi and other ionization methodologies. *Current Proteomics*, 16(4):272–281, 2019.

[35] A. El-Aneed, A. Cohen, and J. Banoub. Mass spectrometry, review of the basics: Electrospray, maldi, and commonly used mass analyzers. *Applied Spectroscopy Reviews*, 44(3):210–230, 2009.

[36] R. Grandori, C. Santambrogio, S. Brocca, G. Invernizzi, and M. Lotti. Electrospray-ionization mass spectrometry as a tool for fast screening of protein structural properties. *Biotechnology Journal*, 4(1):73–87, 2009.

[37] W.J. Griffiths, A.P. Jonsson, S. Liu, D.K. Rai, and Y. Wang. Electrospray and tandem mass spectrometry in biochemistry. *Biochemical Journal*, 355(3):545–561, 2001.

[38] S. Laughlin and W. David Wilson. May the best molecule win: Competition esi mass spectrometry. *International Journal of Molecular Sciences*, 16(10):24506–24531, 2015.

[39] S. Ahmad, H. Kalra, A. Gupta, B. Raut, A. Hussain, and M.A. Rahman. Hybridspe: A novel technique to reduce phospholipid-based matrix effect in lc-esi-ms bioanalysis. *Journal of Pharmacy and Bioallied Sciences*, 4(4):267–275, 2012.

[40] A. Cappiello, G. Famiglini, P. Palma, E. Pierini, V. Termopoli, and H. Trufelli. Overcoming matrix effects in liquid chromatography-mass spectrometry. *Analytical Chemistry*, 80(23):9343–9348, 2008.

[41] S. Özcan, S. Levent, and N.Ö. Can. Aspects of matrix effects in applications of liquid chromatography-mass spectrometry to catecholamine analysis-a review. *Current Analytical Chemistry*, 17(9):1305–1321, 2021.

[42] S.H. Lomeli, I.X. Peng, S. Yin, R.R. Ogorzalek Loo, and J.A. Loo. New reagents for increasing esi multiple charging of proteins and protein complexes. *Journal of the American Society for Mass Spectrometry*, 21(1):127–131, 2010.

[43] Z. Zhang and A.G. Marshall. A universal algorithm for fast and automated charge state deconvolution of electrospray mass-to-charge ratio spectra. *Journal of the American Society for Mass Spectrometry*, 9(3):225–233, 1998.

[44] U. Boesl. Time-of-flight mass spectrometry: Introduction to the basics. *Mass Spectrometry Reviews*, 36(1):86–109, 2017.

[45] C. Li, S. Chu, S. Tan, X. Yin, Y. Jiang, X. Dai, X. Gong, X. Fang, and D. Tian. Towards higher sensitivity of mass spectrometry: A perspective from the mass analyzers. *Frontiers in Chemistry*, 9, 2021.

[46] B.K. Choi, D.M. Hercules, T. Zhang, and A.I. Gusev. Comparison of quadrupole, time-of-flight, and fourier transform mass analyzers for lc-ms applications. *LCGC North America*, 19(5):514–524, 2001.

[47] G. Giorgi. *Mass spectrometry and high resolution mass spectrometry: An overview*. Number 9789402411. 2017.

[48] J.H. Gross. *Mass spectrometry: A textbook: Second edition*. 2011.

[49] W.M.A. Niessen. *MS-MS and MSn*. 2017.

[50] R. Aebersold and M. Mann. Mass-spectrometric exploration of proteome structure and function. *Nature*, 537(7620):347–355, 2016.

[51] J.M. Gilmore and M.P. Washburn. Advances in shotgun proteomics and the analysis of membrane proteomes. *Journal of Proteomics*, 73(11):2078–2091, 2010.

[52] Y. Zhang, B.R. Fonslow, B. Shan, M.-C. Baek, and J.R. Yates III. Protein analysis by shotgun/bottom-up proteomics. *Chemical Reviews*, 113(4):2343–2394, 2013.

[53] W. Zhu, J.W. Smith, and C.-M. Huang. Mass spectrometry-based label-free quantitative proteomics. *Journal of Biomedicine and Biotechnology*, 2010, 2010.

[54] C.H. Johnson, J. Ivanisevic, and G. Siuzdak. Metabolomics: Beyond biomarkers and towards mechanisms. *Nature Reviews Molecular Cell Biology*, 17(7):451–459, 2016.

[55] A. Shevchenko and K. Simons. Lipidomics: Coming to grips with lipid diversity. *Nature Reviews Molecular Cell Biology*, 11(8):593–598, 2010.

[56] G.J. Patti, O. Yanes, and G. Siuzdak. Innovation: Metabolomics: the apogee of the omics trilogy. *Nature Reviews Molecular Cell Biology*, 13(4):263–269, 2012.

[57] D. Vuckovic. Current trends and challenges in sample preparation for global metabolomics using liquid chromatography-mass spectrometry. *Analytical and Bioanalytical Chemistry*, 403(6):1523–1548, 2012.

[58] Y. Hasin, M. Seldin, and A. Lusis. Multi-omics approaches to disease. *Genome Biology*, 18(1), 2017.

[59] K.-M. Song, S. Lee, and C. Ban. Aptamers and their biological applications. *Sensors*, 12(1):612–631, 2012.

[60] M. Darmostuk, S. Rimpelova, H. Gbelcova, and T. Ruml. Current approaches in selex: An update to aptamer selection technology. *Biotechnology Advances*, 33(6):1141–1161, 2014.

[61] Z. Zhuo, Y. Yu, M. Wang, J. Li, Z. Zhang, J. Liu, X. Wu, A. Lu, G. Zhang, and B. Zhang. Recent advances in selex technology and aptamer applications in biomedicine. *International Journal of Molecular Sciences*, 18(10), 2017.

[62] A.D. Keefe, S. Pai, and A. Ellington. Aptamers as therapeutics. *Nature Reviews Drug Discovery*, 9(7):537–550, 2010.

[63] J. Zhou and J. Rossi. Aptamers as targeted therapeutics: Current potential and challenges. *Nature Reviews Drug Discovery*, 16(3):181–202, 2017.

[64] Jan P. Elskens, Joke M. Elskens, and Annemieke Madder. Chemical modification of aptamers for increased binding affinity in diagnostic applications: Current status and future prospects. *International Journal of Molecular Sciences*, 21(12):4522, June 2020.

[65] Emil A. Zaripov, Tiah Lee, Yuchu Dou, Cory S. Harris, Artem Egorov, and Maxim V. Berezovski. Single-run separation and quantification of 14 cannabinoids using capillary electrophoresis. *Separations*, 8(3), March 2021.

[66] Ernest Small and Arthur Cronquist. *A Practical and Natural Taxonomy for Cannabis*, volume 25. 1976.

[67] John M. McPartland. Cannabis systematics at the levels of family, genus, and species. *Cannabis and Cannabinoid Research*, 3(1):203–212, October 2018.

[68] José Alexandre Crippa, Antonio Waldo Zuardi, Rocio Martín-Santos, Sagnik Bhattacharyya, Zerrin Atakan, Philip McGuire, and Paolo Fusar-Poli. Cannabis and anxiety: A critical review of the evidence. *Human Psychopharmacology*, 24(7):515–523, 2009.

[69] Richard D Mattes, Karl Engelman, Leslie M SHAWt, Mahmoud A Elsohly, K Engelman, L M Shaw, and M A Elsohly. *Cannabinoids and appetite stimulation*, volume 49. 1994.

[70] Matthias Karst, Sonja Wippermann, and Jörg Ahrens. *Role of Cannabinoids in the Treatment of Pain and (Painful) Spasticity.*

[71] Kimberly A. Babson and Marcel O. Bonn-Miller. Sleep disturbances: Implications for cannabis use, cannabis use cessation, and cannabis use treatment. *Current Addiction Reports*, 1(2):109–114, June 2014.

[72] The cannabis act: The facts, June 2018.

[73] State medical marijuana laws, March 2021.

[74] Michele Dei Cas, Eleonora Casagni, Anna Saccardo, Sebastiano Arnoldi, Craig Young, Stefano Scotti, Edgardo Vieira de Manicor, Veniero Gambaro, and Gabriella Roda. The italian panorama of cannabis light preparation: Determination of cannabinoids by lc-uv. *Forensic Science International*, 307, February 2020.

[75] Mahmoud A. ElSohly, Mohamed M. Radwan, Waseem Gul, Suman Chandra, and Ahmed Galal. Phytochemistry of cannabis sativa l. *Progress in the chemistry of organic natural products*, 103:1–36, January 2017.

[76] Franz E. Dussy, Cornelia Hamberg, Marco Luginbühl, Thomas Schwerzmann, and Thomas A. Briellmann. Isolation of 9-thca-a from hemp and analytical aspects concerning the determination of 9-thc in cannabis products. *Forensic Science International*, 149(1):3–10, April 2005.

[77] Yoshisuke Tsuda and Masao Hattori. Tetrahydrocannabinolic acid, a genuine substance of tetrahydrocannabinol. *Chem. Pharm. Bull.*, 15(7):1075–1076, April 1967.

[78] Mara Mandrioli, Matilde Tura, Stefano Scotti, and Tullia Gallina Toschi. Fast detection of 10 cannabinoids by rp-hplc-uv method in cannabis sativa l. *Molecules*, 24(11), June 2019.

[79] Ulli Backofen, Frank-Michael Matysik, and Craig E Lunte. *Determination of cannabinoids in hair using high-pH* non-aqueous electrolytes and electrochemical detection Some aspects of sensitivity and selectivity.* 2002.

[80] Jekaterina Mazina, Anastassija Spiljova, Merike Vaher, Mihkel Kaljurand, and Maria Kulp. A rapid capillary electrophoresis method with led-induced native fluorescence detection for the analysis of cannabinoids in oral fluid. *Analytical Methods*, 7(18):7741–7747, September 2015.

[81] Ira S Lurie, Richard P Meyers, and Timothy S Conver. *Capillary Electrochromatography of Cannabinoids.*

[82] Rossella Gottardo, Anna Bertaso, Jennifer Pascali, Daniela Sorio, Giacomo Musile, Elisa Trapani, Catia Seri, Giovanni Serpelloni, and Franco Tagliaro. Micellar electrokinetic chromatography: A new simple tool for the analysis of synthetic cannabinoids in herbal blends and for the rapid estimation of their logp values. *Journal of Chromatography A*, 1267:198–205, December 2012.

[83] Hui Ling Cheng, Yi Hsuan Tsai, Wan Ling Hsu, and Yi Hui Lin. An on-line stacking capillary electrophoresis method for the analysis of 9-tetrahydrocannabinol and its metabolites. *Journal of Chromatography A*, 1426:226–232, December 2015.

[84] Ketan Hippalgaonkar, Waseem Gul, Mahmoud A Elsohly, Michael A Repka, and Soumyajit Majumdar. Enhanced solubility, stability, and transcorneal permeability of delta-8-tetrahydrocannabinol in the presence of cyclodextrins.

[85] Philippe Schmitt-Kopplin, Andrej V. Garmash, Alexey V. Kudryavtsev, Franz Menzinger, Irina V. Perminova, Norbert Hertkorn, Dieter Freitag, Valery S. Petrosyan, and Antonius Kettrup. Quantitative and qualitative precision improvements by effective mobility-scale data transformation in capillary electrophoresis analysis. *Electrophoresis*, 22(1):77–87, 2001.

[86] Walter F. da Silva Júnior, Jonas G. de Oliveira Pinheiro, Carlos D.L.F.A. Moreira, Fabia J.J. de Souza, and Ádley A.N. de Lima. Alternative technologies to improve solubility and stability of poorly water-soluble drugs. *Multifunctional Systems for Combined Delivery, Biosensing and Diagnostics*, page 281–305, January 2017.

[87] Guillaume van Niel, Gisela D'Angelo, and Graça Raposo. Shedding light on the cell biology of extracellular vesicles. *Nature Reviews Molecular Cell Biology*, 19(4):213–228, April 2018.

[88] Mercedes Tkach and Clotilde Théry. Communication by extracellular vesicles: Where we are and where we need to go. *Cell*, 164(6):1226–1232, March 2016.

[89] Takashi Kinoshita, Kenneth W Yip, Tara Spence, and Fei-Fei Liu. Micrornas in extracellular vesicles: potential cancer biomarkers. *Journal of Human Genetics*, 62(1):67–74, January 2017.

[90] R. E. Lane, D. Korbie, M. M. Hill, and M. Trau. Extracellular vesicles as circulating cancer biomarkers: opportunities and challenges. *Clinical and Translational Medicine*, 7(1), December 2018.

[91] Janusz Rak. Extracellular vesicles – biomarkers and effectors of the cellular interactome in cancer. *Frontiers in Pharmacology*, 4, 2013.

[92] Fumihiko Urabe, Nobuyoshi Kosaka, Kagenori Ito, Takahiro Kimura, Shin Egawa, and Takahiro Ochiya. Extracellular vesicles as biomarkers and therapeutic targets for cancer. *American Journal of Physiology-Cell Physiology*, 318(1):C29–C39, January 2020.

[93] G.J. Patti, O. Yanes, and G. Siuzdak. Innovation: Metabolomics: the apogee of the omics trilogy. *Nature Reviews Molecular Cell Biology*, 13(4):263–269, 2012.

[94] K. Dettmer, P.A. Aronov, and B.D. Hammock. Mass spectrometry-based metabolomics. *Mass Spectrometry Reviews*, 26(1):51–78, 2007.

[95] C.H. Johnson, J. Ivanisevic, and G. Siuzdak. Metabolomics: Beyond biomarkers and towards mechanisms. *Nature Reviews Molecular Cell Biology*, 17(7):451–459, 2016.

[96] A. Hirayama, K. Kami, M. Sugimoto, M. Sugawara, N. Toki, H. Onozuka, T. Kinoshita, N. Saito, A. Ochiai, M. Tomita, H. Esumi, and T. Soga. Quantitative metabolome profiling of colon and stomach cancer microenvironment by capillary electrophoresis time-of-flight mass spectrometry. *Cancer Research*, 69(11):4918–4925, 2009.

[97] T. Obata and A.R. Fernie. The use of metabolomics to dissect plant responses to abiotic stresses. *Cellular and Molecular Life Sciences*, 69(19):3225–3243, 2012.

[98] T. Soga, R. Baran, M. Suematsu, Y. Ueno, S. Ikeda, T. Sakurakawa, Y. Kakazu, T. Ishikawa, M. Robert, T. Nishioka, T. Nishioka, and M. Tomita. Differential metabolomics reveals ophthalmic acid as an oxidative stress biomarker indicating hepatic glutathione consumption. *Journal of Biological Chemistry*, 281(24):16768–16776, 2006.

[99] T. Soga, Y. Ohashi, Y. Ueno, H. Naraoka, M. Tomita, and T. Nishioka. Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. *Journal of Proteome Research*, 2(5):488–494, 2003.

[100] M. Sugimoto, D.T. Wong, A. Hirayama, T. Soga, and M. Tomita. Capillary electrophoresis mass spectrometry-based saliva metabolomics identified oral, breast and pancreatic cancer-specific profiles. *Metabolomics*, 6(1):78–95, 2010.

[101] Nahla E. El-Ashmawy and Rania M. Khalil. A review on the role of l-carnitine in the management of tamoxifen side effects in treated women with breast cancer. *Tumor Biology*, 35(4):2845–2855, April 2014.

[102] Raheleh Farahzadi, Zohreh Sanaat, Ali Akbar Movassaghpour-Akbari, Ezzatollah Fathi, and Soheila Montazersaheb. Investigation of l-carnitine effects on cd44+ cancer stem cells from mda-mb-231 breast cancer cell line as anti-cancer therapy. *Regenerative Therapy*, 24:219–226, December 2023.

[103] Barbara Juraszek, Joanna Czarnecka-Herok, and Katarzyna A. Nałecz. Glioma cells survival depends both on fatty acid oxidation and on functional carnitine transport by slc22a5. *Journal of Neurochemistry*, 156(5):642–657, March 2021.

[104] Marc R. McCann, Mery Vet George De la Rosa, Gus R. Rosania, and Kathleen A. Stringer. L-carnitine and acylcarnitines: Mitochondrial biomarkers for precision medicine. *Metabolites*, 11(1):51, January 2021.

[105] Dragana Savic, Leanne Hodson, Stefan Neubauer, and Michael Pavlides. The importance of the fatty acid transporter l-carnitine in non-alcoholic fatty liver disease (nafld). *Nutrients*, 12(8):2178, July 2020.

[106] Andrew J. Hoy, Shilpa R. Nagarajan, and Lisa M. Butler. Tumour fatty acid metabolism in the context of therapy resistance and obesity. *Nature Reviews Cancer*, 21(12):753–766, December 2021.

[107] Nikos Koundouros and George Poulogiannis. Reprogramming of fatty acid metabolism in cancer. *British Journal of Cancer*, 122(1):4–22, January 2020.

[108] Jieun Seo, Do-Won Jeong, Jong-Wan Park, Kwang-Woong Lee, Junji Fukuda, and Yang-Sook Chun. Fatty-acid-induced fabp5/hif-1 reprograms lipid metabolism and enhances the proliferation of liver cancer cells. *Communications Biology*, 3(1):638, October 2020.

[109] Laura E. Graves, Edgardo V. Ariztia, Jason R. Navari, Heather J. Matzel, M. Sharon Stack, and David A. Fishman. Proinvasive properties of ovarian cancer ascites-derived membrane vesicles. *Cancer Research*, 64(19):7045–7049, October 2004.

[110] M. Helena Vasconcelos, Hugo R. Caires, Artūrs Ābols, Cristina P.R. Xavier, and Aija Linē. Extracellular vesicles as a novel source of biomarkers in liquid biopsies for monitoring cancer progression and drug resistance. *Drug Resistance Updates*, 47:100647, December 2019.

[111] Lola E. Navas and Amancio Carnero. Nicotinamide adenine dinucleotide (nad) metabolism as a relevant target in cancer. *Cells*, 11(17):2627, August 2022.

[112] Cissy Zhang, Giang Hoang, Nabeel Attarwala, Arthur J. L. Cooper, Ryoichi Asaka, and Anne Le. Metabolic recycling enhances proliferation in myc-transformed lymphoma b cells. *Advanced Biology*, 7(2), February 2023.

[113] Ana Paula Santin Bertoni, Cleiton Felipe Valandro, Rafael Ávila Brasil, Felipe André Zeiser, Márcia Rosângela Wink, Tania Weber Furlanetto, and Cristiano André da Costa. Nt5e dna methylation in papillary thyroid cancer: Novel opportunities for precision oncology. *Molecular and Cellular Endocrinology*, 570:111915, June 2023.

[114] Xue-min Chen, Yu-yang Liu, Bing-yan Tao, Xin-miao Xue, Xin-xin Zhang, Lin-lin Wang, Hui Zhong, Jun Zhang, Shi-ming Yang, and Qing-qing Jiang. Nt5e upregulation in head and neck squamous cell carcinoma: A novel biomarker on cancer-associated fibroblasts for predicting immunosuppressive tumor microenvironment. *Frontiers in Immunology*, 13, August 2022.

[115] Young Ju Jeong, Hoon Kyu Oh, Hye Ryeon Choi, and Sung Hwan Park. Methylation of the nt5e gene is associated with poor prognostic factors in breast cancer. *Diagnostics*, 10(11):939, November 2020.

[116] S.Y. Toh, M. Citartan, S.C.B. Gopinath, and T.-H. Tang. Aptamers as a replacement for antibodies in enzyme-linked immunosorbent assay. *Biosensors and Bioelectronics*, 64:392–403, 2015.

[117] J. Zhou and J. Rossi. Aptamers as targeted therapeutics: Current potential and challenges. *Nature Reviews Drug Discovery*, 16(3):181–202, 2017.

[118] N. Cennamo, L. Pasquardini, F. Arcadio, L. Lunelli, L. Vanzetti, V. Carafa, L. Altucci, and L. Zeni. Sars-cov-2 spike protein detection through a plasmonic d-shaped plastic optical fiber aptasensor. *Talanta*, 233, 2021.

[119] N. Kacherovsky, L.F. Yang, H.V. Dang, E.L. Cheng, I.I. Cardle, A.C. Walls, M. McCallum, D.L. Sellers, F. DiMaio, S.J. Salipante, D. Veesler, and S.H. Pun. Discovery and characterization of spike n-terminal domain-binding aptamers for rapid sars-cov-2 detection. *Angewandte Chemie - International Edition*, 60(39):21211–21215, 2021.

[120] H. Kaur, J.G. Bruno, A. Kumar, and T.K. Sharma. Aptamers in the therapeutics and diagnostics pipelines. *Theranostics*, 8(15):4016–4032, 2018.

[121] Y. Zhang, B.S. Lai, and M. Juhas. Recent advances in aptamer discovery and applications. *Molecules*, 24(5), 2019.

[122] X. Liu, Y.-L. Wang, J. Wu, J. Qi, Z. Zeng, Q. Wan, Z. Chen, P. Manandhar, V.S. Cavener, N.R. Boyle, S.-H. Chen, and Y. Zu. Neutralizing aptamers block s/rbd-ace2 interactions and prevent host cell infection. *Angewandte Chemie - International Edition*, 60(18):10273–10278, 2021.

[123] A. Schmitz, A. Weber, M. Bayin, S. Breuers, V. Fieberg, M. Famulok, and G. Mayer. A sars-cov-2 spike binding dna aptamer that inhibits pseudovirus infection by an rbd-independent mechanism**. *Angewandte Chemie - International Edition*, 60(18):10279–10285, 2021.

[124] Y. Song, J. Song, X. Wei, M. Huang, M. Sun, L. Zhu, B. Lin, H. Shen, Z. Zhu, and C. Yang. Discovery of aptamers targeting the receptor-binding domain of the sars-cov-2 spike glycoprotein. *Analytical Chemistry*, 92(14):9895–9900, 2020.

[125] M. Sun, S. Liu, X. Wei, S. Wan, M. Huang, T. Song, Y. Lu, X. Weng, Z. Lin, H. Chen, Y. Song, and C. Yang. Aptamer blocking strategy inhibits sars-cov-2 virus infection. *Angewandte Chemie - International Edition*, 60(18):10266–10272, 2021.

[126] R. Chen, L. Kan, F. Duan, L. He, M. Wang, J. Cui, Z. Zhang, and Z. Zhang. Surface plasmon resonance aptasensor based on niobium carbide mxene quantum dots for nucleocapsid of sars-cov-2 detection. *Microchimica Acta*, 188(10), 2021.

[127] S. Ramanathan, S.C.B. Gopinath, Z.H. Ismail, M.K. Md Arshad, and P. Poopalan. Aptasensing nucleocapsid protein on nanodiamond assembled gold interdigitated electrodes for impedimetric sars-cov-2 infectious disease assessment. *Biosensors and Bioelectronics*, 197, 2022.

[128] J. Tian, Z. Liang, O. Hu, Q. He, D. Sun, and Z. Chen. An electrochemical dual-aptamer biosensor based on metal-organic frameworks mil-53 decorated with au@pt nanoparticles and enzymes for detection of covid-19 nucleocapsid protein. *Electrochimica Acta*, 387, 2021.

[129] J.-Y. Li, C.-H. Liao, Q. Wang, Y.-J. Tan, R. Luo, Y. Qiu, and X.-Y. Ge. The orf6, orf8 and nucleocapsid proteins of sars-cov-2 inhibit type i interferon signaling pathway. *Virus Research*, 286, 2020.

[130] N.K. Dutta, K. Mazumdar, and J.T. Gordy. The nucleocapsid protein of sars–cov-2: A target for vaccine development. *Journal of Virology*, 94(13), 2020.

[131] I. Dovgan, O. Koniev, S. Kolodych, and A. Wagner. Antibody-oligonucleotide conjugates as therapeutic, imaging, and detection agents. *Bioconjugate Chemistry*, 30(10):2483–2501, 2019.

[132] L. Xu, J. Duan, J. Chen, S. Ding, and W. Cheng. Recent advances in rolling circle amplification-based biosensing strategies-a review. *Analytica Chimica Acta*, 1148, 2021.

[133] Y. Fu, H. Qian, X. Zhou, Y. Wu, L. Song, K. Chen, D. Bai, Y. Yang, J. Li, and G. Xie. Proximity ligation assay mediated rolling circle amplification strategy for in situ amplified imaging of glycosylated pd-l1. *Analytical and Bioanalytical Chemistry*, 413(28):6929–6939, 2021.

[134] B. Lin, T. Tian, Y. Lu, D. Liu, M. Huang, L. Zhu, Z. Zhu, Y. Song, and C. Yang. Tracing tumor-derived exosomal pd-l1 by dual-aptamer activated proximity-induced droplet digital pcr. *Angewandte Chemie - International Edition*, 60(14):7582–7586, 2021.

[135] R. Liu, L. He, Y. Hu, Z. Luo, and J. Zhang. A serological aptamer-assisted proximity ligation assay for covid-19 diagnosis and seeking neutralizing aptamers. *Chemical Science*, 11(44):12157–12164, 2020.

[136] W. Yun, L. You, F. Li, H. Wu, L. Chen, and L. Yang. Proximity ligation assay induced and dnazyme powered dna motor for fluorescent detection of thrombin. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, 207:39–45, 2019.

[137] X. Zhao, C. Luo, Q. Mei, H. Zhang, W. Zhang, D. Su, W. Fu, and Y. Luo. Aptamer-cholesterol-mediated proximity ligation assay for accurate identification of exosomes. *Analytical Chemistry*, 92(7):5411–5418, 2020.

[138] Khalid K Alam, Jonathan L Chang, and Donald H Burke. Fastaptamer: A bioinformatic toolkit for high-throughput sequence analysis of combinatorial selections. *Molecular Therapy - Nucleic Acids*, 4:e230, 2015.

[139] Robin Schmid, Steffen Heuckeroth, Ansgar Korf, Aleksandr Smirnov, Owen Myers, Thomas S. Dyrlund, Roman Bushuiev, Kevin J. Murray, Nils Hoffmann, Miaoshan Lu, Abinesh Sarvepalli, Zheng Zhang, Markus Fleischauer, Kai Dührkop, Mark Wesner, Shawn J. Hoogstra, Edward Rudt, Olena Mokshyna, Corinna Brungs, Kirill Ponomarov, Lana Mutabdžija, Tito Damiani, Chris J. Pudney, Mark Earll, Patrick O. Helmer, Timothy R. Fallon, Tobias Schulze, Albert Rivas-Ubach, Aivett Bilbao, Henning Richter, Louis-Félix Nothias, Mingxun Wang, Matej Orešič, Jing-Ke Weng, Sebastian Böcker, Astrid Jeibmann, Heiko Hayen, Uwe Karst, Pieter C. Dorrestein, Daniel Petras, Xiuxia Du, and Tomáš Pluskal. Integrative analysis of multimodal mass spectrometry data in mzmine 3. *Nature Biotechnology*, 41(4):447–449, April 2023.

[140] G.C.P. van Zundert, J.P.G.L.M. Rodrigues, M. Trellet, C. Schmitz, P.L. Kastritis, E. Karaca, A.S.J. Melquiond, M. van Dijk, S.J. de Vries, and A.M.J.J. Bonvin. The haddock2.2 web server: User-friendly integrative modeling of biomolecular complexes. *Journal of Molecular Biology*, 428(4):720–725, February 2016.

[141] Suttinee Poolsup, Emil Zaripov, Nico Hüttmann, Zoran Minic, Polina V. Artyushenko, Irina A. Shchugoreva, Felix N. Tomilin, Anna S. Kichkailo, and Maxim V. Berezovski. Discovery of dna aptamers targeting sars-cov-2 nucleocapsid protein and protein-binding epitopes for label-free covid-19 diagnostics. *Molecular Therapy - Nucleic Acids*, 31:731–743, March 2023.

[142] Michael Kohlberger and Gabriele Gadermaier. Selex: Critical factors and optimization strategies for successful aptamer selection. *Biotechnology and Applied Biochemistry*, 69(5):1771–1792, October 2022.

[143] Cheng-Sheng Lin, Yi-Cheng Tsai, Keng-Fu Hsu, and Gwo-Bin Lee. Optimization of aptamer selection on an automated microfluidic system with cancer tissues. *Lab on a Chip*, 21(4):725–734, 2021.

[144] Grant C. O'Connell and Christine G. Smothers. Optimized methodology for product recovery following emulsion pcr: applications for amplification of aptamer libraries and other complex templates. *Journal of Biological Methods*, 7(1):e128, March 2020.

[145] Richard Williams, Sergio G Peisajovich, Oliver J Miller, Shlomo Magdassi, Dan S Tawfik, and Andrew D Griffiths. Amplification of complex gene libraries by emulsion pcr. *Nature Methods*, 3(7):545–550, July 2006.

[146] Zhiqiang Chen, Qihan Wu, Jing Chen, Xiaohua Ni, and Jianfeng Dai. A dna aptamer based method for detection of sars-cov-2 nucleocapsid protein. *Virologica Sinica*, 35(3):351–354, June 2020.

[147] Chenchen Ge, Juan Feng, Jiaming Zhang, Kai Hu, Dou Wang, Ling Zha, Xuejuan Hu, and Rongsong Li. Aptamer/antibody sandwich method for digital detection of sars-cov2 nucleocapsid protein. *Talanta*, 236:122847, January 2022.

[148] Nataly Kacherovsky, Lucy F. Yang, Ha V. Dang, Emmeline L. Cheng, Ian I. Cardle, Alexandra C. Walls, Matthew McCallum, Drew L. Sellers, Frank DiMaio, Stephen J. Salipante, Davide Corti, David Veesler, and Suzie H. Pun. Discovery and characterization of spike n-terminal domain-binding aptamers for rapid sars-cov-2 detection. *Angewandte Chemie International Edition*, 60(39):21211–21215, September 2021.

[149] Liyun Zhang, Xiaona Fang, Xingbo Liu, Huichao Ou, Haiyan Zhang, Jinjun Wang, Qian Li, Huanyi Cheng, Wenyi Zhang, and Zhaofeng Luo. Discovery of sandwich type covid-19 nucleocapsid protein dna aptamers. *Chemical Communications*, 56(70):10235–10238, 2020.

[150] Issam Hmila, Boutheina Marnissi, Masood Kamali-Moghaddam, and Abdeljelil Ghram. Aptamer-assisted proximity ligation assay for sensitive detection of infectious bronchitis coronavirus. *Microbiology Spectrum*, 11(1), February 2023.

[151] Ran Liu, Lei He, Yuansheng Hu, Zhaofeng Luo, and Jingjing Zhang. A serological aptamer-assisted proximity ligation assay for covid-19 diagnosis and seeking neutralizing aptamers. *Chemical Science*, 11(44):12157–12164, 2020.

[152] Boutheina Marnissi, Khouloud Khalfaoui, Tonge Ebai, Felipe Marques Souza de Oliveira, Abdeljelil Ghram, Masood Kamali-Moghaddam, and Issam Hmila. Accurate detection of newcastle disease virus using proximity-dependent dna aptamer ligation assays. *FEBS Open Bio*, 11(4):1122–1131, April 2021.

[153] Juan Carlos Abrego-Martinez, Maziar Jafari, Siham Chergui, Catalin Pavel, Diping Che, and Mohamed Siaj. Aptamer-based electrochemical biosensor for rapid detection of sars-cov-2: Nanoscale electrode-aptamer-sars-cov-2 imaging by photo-induced force microscopy. *Biosensors and Bioelectronics*, 195:113595, January 2022.

[154] Zahra Sadat Moshref, Tahmineh Jalali, Razieh Rezaei Adriani, Elahe Soltati, and Seyed Latif Mousavi Gargari. Aptamer-based diagnosis of various sars-cov2 strains isolated from clinical specimens. *Heliyon*, 9(6):e16458, June 2023.

[155] Lucy F. Yang, Nataly Kacherovsky, Nuttada Panpradist, Ruixuan Wan, Joey Liang, Bo Zhang, Stephen J. Salipante, Barry R. Lutz, and Suzie H. Pun. Aptamer sandwich lateral flow assay (aptaflow) for antibody-free sars-cov-2 detection. *Analytical Chemistry*, 94(20):7278–7285, May 2022.

[156] Yuka Nishibata, Shota Koshimoto, Kenta Ogaki, Erika Ishikawa, Kosuke Wada, Miku Yoshinari, Yuto Tamura, Ryo Uozumi, Sakiko Masuda, Utano Tomaru, and Akihiro Ishizu. Rnase in the saliva can affect the detection of severe acute respiratory syndrome coronavirus 2 by real-time one-step polymerase chain reaction using saliva samples. *Pathology - Research and Practice*, 220:153381, April 2021.

[157] Glenn D. Braunstein, Lori Schwartz, Pamela Hymel, and Jonathan Fielding. False positive results with sars-cov-2 rt-pcr tests and how to evaluate a rt-pcr-positive test for the possibility of a false positive result. *Journal of Occupational Environmental Medicine*, 63(3):e159–e162, March 2021.

[158] Francesca Falasca, Ilaria Sciandra, Daniele Di Carlo, Massimo Gentile, Alberto Deales, Guido Antonelli, and Ombretta Turriziani. Detection of sars-cov n2 gene: Very low amounts of viral rna or false positive? *Journal of Clinical Virology*, 133:104660, December 2020.

[159] Andrew J. Alpert. Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. *Journal of Chromatography A*, 499(C):177–196, January 1990.

[160] M.R.N. Monton and T. Soga. Metabolome analysis by capillary electrophoresis-mass spectrometry. *Journal of Chromatography A*, 1168(1–2):237–246, 2007.

[161] A. Stolz, K. Jooß, O. Höcker, J. Römer, J. Schlecht, and C. Neusüß. Recent advances in capillary electrophoresis-mass spectrometry: Instrumentation, methodology and applications. *Electrophoresis*, 40(1):79–112, 2019.

[162] N. Drouin, M. Van Mever, W. Zhang, E. Tobolkina, S. Ferre, A.-C. Servais, M.-J. Gou, L. Nyssen, M. Fillet, G.S.M. Lageveen-Kammeijer, S. Rudaz, and R. Ramautar. Capillary electrophoresis-mass spectrometry at trial by metabo-ring: Effective electrophoretic mobility for reproducible and robust compound annotation. *Analytical Chemistry*, 92(20):14103–14112, 2020.

[163] V. González-Ruiz, Y. Gagnebin, N. Drouin, S. Codesido, S. Rudaz, and J. Schappler. Romance: A new software tool to improve data robustness and feature identification in ce-ms metabolomics. *Electrophoresis*, 39(9–10):1222–1232, 2018.

[164] W. Zhang, K. Segers, D. Mangelings, A. Van Eeckhaut, T. Hankemeier, Y. Vander Heyden, and R. Ramautar. Assessing the suitability of capillary electrophoresis-mass spectrometry for biomarker discovery in plasma-based metabolomics. *Electrophoresis*, 40(18–19):2309–2320, 2019.

[165] P. Fang, J.-Z. Pan, and Q. Fang. A robust and extendable sheath flow interface with minimal dead volume for coupling ce with esi-ms. *Talanta*, 180:376–382, 2018.

[166] S. Ferré, N. Drouin, V. González-Ruiz, and S. Rudaz. Evaluation of a nanoflow interface based on the triple-tube coaxial sheath-flow sprayer for capillary electrophoresis-mass spectrometry coupling in metabolomics. *Journal of Chromatography A*, 1641, 2021.

[167] E. Sánchez-López, G.S.M. Kammeijer, A.L. Crego, M.L. Marina, R. Ramautar, D.J.M. Peters, and O.A. Mayboroda. Sheathless ce-ms based metabolic profiling of kidney tissue section samples from a mouse model of polycystic kidney disease. *Scientific Reports*, 9(1), 2019.

[168] K. Sasaki, H. Sagawa, M. Suzuki, H. Yamamoto, M. Tomita, T. Soga, and Y. Ohashi. Metabolomics platform with capillary electrophoresis coupled with high-resolution mass spectrometry for plasma analysis. *Analytical Chemistry*, 91(2):1295–1301, 2019.