The Application of Liquid Chromatography-Tandem Mass Spectrometry to the Analysis of Biomedical Samples

Jonathan Oyler

A thesis presented in partial fulfilment of the
requirements of Nottingham Trent University for the
degree of Doctor of Philosophy

September 2023

The copyright in this work is held by the author. You may copy up to 5% of this work for private study, or personal, non-commercial research. Any re-use of the information contained within this document should be fully referenced listing the author, title, university, degree level, and pagination. Queries or requests for any other use, or if a more substantial copy is required, should be directed to the author.

Author Contributions

Chapters 1 and 5 were authored solely by Jonathan Oyler with review by David Kilgour, Benjamin Oyler, and Carole Perry.

Chapter 2 author contributions were as follows:

Jonathan Oyler – funding acquisition, study design, data acquisition, data analysis, data interpretation, manuscript authorship and submission.
Kathleen Maistros – sample preparation, data acquisition, data analysis, data interpretation, manuscript editing.
David P.A. Kilgour – manuscript editing.

Chapter 3 author contributions were as follows:

Jonathan Oyler – funding acquisition, study design, sample preparation, data acquisition, data analysis, data interpretation, manuscript authorship and submission.
Bao Tran – study design, sample preparation, data acquisition, data analysis, data interpretation.
David P.A Kilgour – data interpretation.

Chapter 4 author contributions were as follows:

Jonathan Oyler – Co-principle investigator, funding acquisition, study and methods design, sample preparation (orbitrap and QToF analyses), data acquisition (orbitrap analyses), data interpretation, manuscript authorship and submission.
Raymond Sullivan – Co-principle investigator, funding acquisition, study design, sample preparation (orbitrap analyses).
Bao Tran – LC/Orbitrap methods design, data acquisition (orbitrap analyses), data interpretation (orbitrap analyses), authorship of the orbitrap method section.
Dhanwin Baker – sample preparation (QToF analyses).
Clare Coveney – sample analyses (QToF analyses), authorship of QToF methods section in manuscript.
David Boocock – review data analysis results (QToF analyses).
Benjamin Oyler – data acquisition (repeat of small set of orbitrap analyses), data interpretation (small set of orbitrap data), manuscript editing.
Carole Perry – thesis project advisor, manuscript editing.
David Kilgour – thesis project advisor, data processing and interpretation, manuscript editing.

Outline

Figures and Tables

Chapter 1

Chapter 2

Chapter 3

3.3 **Figure 4** Number of RH peptides identified and sequence coverages for 1 and 4 h Flash, 24 h FASP, 1 and 4 h 100% can, and the new 15 min DOD digestion methods with disulfide reduction.

3.4 **Figure 5** Time course study comparing DOD digestion time to sequence coverage (A) and number of characteristic peptides identified (B) for BotNT-A, SEB, and ricin in a cocktail of the three proteins at a concentration of 22 µg/mL for BotNT-A and SEB, and 4.4 µg/mL for ricin.

Chapter 1

1.0 Introduction

Liquid chromatography (LC) and mass spectrometry (MS) have long been viable analytical tools alone or in combination with a variety of other techniques, but they could not be combined because there was no technique that efficiently and robustly interfaced LC, an ambient pressure method sample component separation method, and MS, a high vacuum (10^{-6} to 10^{-9} torr) detection method. The development of electrospray ionization (ESI), an atmospheric pressure ionization method (API), by Fenn, et.al.[1], made it possible to interface LC very efficiently and robustly with MS. ESI provided a method for transitioning ions generated in the liquid phase efficiently into the gas phase for analysis. This was a breakthrough in analytical chemistry which Dole first reported using in 1968[2,3] and for which Fenn received the 2002 Nobel Prize in chemistry because it became possible to detect, identify, and quantify many more molecular classes using the high specificity and sensitivity of MS. Samples previously not amenable to MS analysis, ranging from non-volatile, heat labile small molecules to large biopolymers like proteins, carbohydrates, large lipids, and intact RNA and DNA, to name a few, could be introduced in the liquid phase from a liquid chromatograph into a mass spectrometer for analysis in the gas phase[4,5,6,7,8,9,10,11,12,13,14]. MS analysis provides a very specific and characteristic spectrum for each molecular species, a "mass fingerprint", considered the "gold standard" for

molecular identification. Subsequently, methods were developed to use MS systems to quantify small molecule target analytes very accurately in samples using laboratory-formulated calibration curves constructed from pure standards of the analyte of interest. Employing different tandem mass spectrometric detectors provided a means for performing systematic experiments in the gas phase to determine more fully and with higher confidence fine molecular structure[15]. This manuscript describes three of the many diverse ways in which LC-MS can be used to answer analytical chemistry research questions very successfully. The first, the development of a small molecule quantification method, was part of an effort to determine the effect of organic solvents on the percutaneous absorption rate of carfentanil, a highly potent opioid, in humans. This method employed a triple quadrupole tandem instrument to very robustly identify carfentanil, caffeine and sufentanil over a relatively very wide concentration range. The initial impetus for the second, the development of a rapid protein identification method, was to develop a single, rapid, MS-based screening method for a wide range of protein toxins as a diagnostic tool to rapidly provide very accurate, specific, and sensitive analytical confirmatory results for use in clinical, forensic and attribution decision making where time to confident decision is critical to ensure patient safety and military personnel security. Subsequently, this method was used to rapidly prepare a variety of samples for LC-MSMS proteomic analysis to identify the proteins in complex biomedical matrices using two different tandem mass spectrometric, Orbitrap and hybrid quadrupole/time-of-flight (QTOF), methodologies.

The following is a brief functional description of LC, ESI and MS systems and how they have been effectively combined to produce a single, very sensitive and specific method that can be used to separate complex samples into their components and identify and accurately quantify with great sensitivity and with very high confidence a majority of the existing molecular classes. It has been one of the most important methods developed in biomedical research in the last two decades[16,17].

1.1 Liquid Chromatography (LC)

LC is an analytical chemistry chromatographic technique used routinely to separate complex mixtures of molecules in the liquid phase[18,19]. While still being very useful for the analysis of most small, semi-volatile analytes, it is the chromatographic method of choice and most common separation method used for polar and large molecules for which gas chromatography is not capable[20]. It is suitable for a much wider variety of analyte classes than GC and is especially useful for nonvolatile, thermally labile small molecules as well as large biopolymers and other large biomolecules which cannot be thermally induced to efficiently transition from the solid or liquid phase by GC into the gas phase as intact molecules (i.e., peptides, proteins, carbohydrates, DNA, RNA, polar lipids). With analytical LC, solvated complex samples are introduced under pressure into a cylindrical stainless steel or glass column packed with a solid, porous "stationary phase" of specific chemical composition. The initial solvent used to load the sample onto the column is one for which the target analytes have little affinity causing them to interact with or "partition" into the stationary phase through reversible chemical reactions[15,16]. This follows the long recognized "like into like" chemical principle[21]. Under initial conditions, as the target analyte-laden solvent passes through the column, target molecules characteristically interact through hydrophillic/hydrophobic, ionic, and/or other weaker chemical forces with the stationary phase and are retained and focused on the column. These chemical reactions are reversible with chemical equilibria which can be shifted depending on solvent conditions and temperature. As solvent and temperature conditions are manipulated during the chromatographic analysis, equilibria change and solute molecules partition from the stationary phase back into the solvent and elute into a detector. Each solute molecule interacts with the stationary phase in a characteristic way leading to a separation in time of different types of molecules as solvent composition and temperature are changed. Changing stationary phase particle size, column length and internal diameter, solvent, and temperature alter target analyte chromatographic characteristics. Historically, LC has been coupled with ultra-violet, infra-red,

electrochemical, and other "non-specific" detectors to identify the components of complex samples. The specificity of these methods relies almost solely on the LC separation or the chromatographic "retention time" and, consequently, they are considered by the analytical chemistry community to provide relatively low specificity. Consequently, efforts were made for years to develop robust interfaces called ion sources between LC and mass spectrometers so that the high specificity of MS could be applied to these analyses.

Typical LC analysis employs a minimum of two mobile phases; the aqueous mobile phase is referred to as "mobile phase A" (MP-A) and the organic mobile phase as "mobile phase B" (MP-B)[15,16]. Most LC separations are polarity-driven with analytes partitioning from the stationary phase into a "like" mobile phase to elute. LC separations can either be performed using "isocratic" solvent conditions where consistent mobile phase composition is maintained throughout chromatography or using a solvent "gradient" in which solvent composition changes during chromatography to more specifically and rapidly separate different solute molecules. Gradient conditions are commonly employed for most LC analyses because changing mobile phase composition generally increases elution strength which in turn speeds up the separation process while enhancing chromatographic peak resolution, and thereby separation efficiency (chromatographic specificity), and sensitivity. The initial solvent composition drives the partitioning of target analytes into the stationary phase where they are retained and focused on the head of the column in a narrow band. The solvent composition and conditions are maintained for a time to allow unwanted sample components that will not partition into the stationary phase under initial solvent conditions to elute prior to the chromatographic separation of target analytes. Solvent composition is then stepped or ramped, and target analytes partition back into the solvent as solvation strength is increased in order of their relative polarity or non-polarity, are swept thru the column, and eluted for detection.

LC can be conducted under two conditions. The most common, "reversed phase" chromatography, employs an initial solvent composed primarily of mobile phase A, the polar

solvent. The second is initiated under nonpolar conditions and is called “normal phase” LC. In reversed phase LC, relatively nonpolar analytes are introduced into the column in a relatively polar solvent where they partition into the nonpolar stationary phase and focus on the head of the column. At a point during chromatography, solvent is changed either gradually or in steps to a primarily organic, nonpolar composition inducing relatively nonpolar target analytes to partition back into the solvent according to solvation characteristics and elute. Normal phase chromatography employs the opposite solvent conditions. Most current analytical LC methods employ reversed phase gradient chromatography, where the solvent is changed gradually or in steps to a more nonpolar composition.

Current analytical LC/MS methods described in the literature almost exclusively employ high performance (formerly referred to as high pressure LC) liquid chromatography (HPLC) or ultra-performance liquid chromatography (UPLC). The primary difference between HPLC and UPLC is the maximum back-pressure that pumps can accommodate. UPLC pumps can reach a maximum pressure of ca. 16,000 psi while HPLC pumps operate at pressures up to 5000-6000 psi. UPLC allows for the use of smaller particle size stationary phases (reduction in particle size causes marked increases in the back-pressure required to push solvent through the column) for UPLC chromatography. The reduction in particle size provides a great increase in stationary phase surface area with a concomitant dramatic increase in chromatographic resolution (sharper peaks) and shorter analytical times. Willison, et.al., validated a method for the small molecule degradation products of chemical warfare agents for which they reported chromatography that was four times faster with much higher chromatographic resolution[22].

Both HPLC and UPLC consist of either binary or quaternary pumps which deliver mobile phase solvents through a degasser and solvent mixer into the analytical column. Employing valve switching, sample is loaded thru a sample loop into the LC flow path and onto an analytical column prior to chromatographic separation. For both LC techniques, under the correct initial solvent conditions, target analytes reversibly bind to the stationary phase at the head of the

column through polar, hydrophobic and/or ionic mechanisms prior to being chromatographically separated.

Analytical HPLC columns consist of a glass or stainless-steel tube with diameters ranging from 2-5 mm and lengths ranging from 30-150 mm. They are packed with silica particles ranging in diameter from 1.7-5 μm which are covalently coated with various chemical ligands. These ligand-coated particles form the stationary phase. Typical flow rates range from 0.1-1 mL/min, and pump back pressures are dependent on flow rate, particle size, column dimensions, solvent, and temperature. Use of an in-line guard column with similar/same stationary phase is usually used to prevent analytical column contamination.

UPLC columns are generally packed with sub-2 μm particles which increases pump pressures to ca. 16,000 psi with most commercially available systems. Because of the great increase in chromatographic resolution gained with UPLC as compared to HPLC, column length can often be reduced shortening run times while maintaining or even substantially improving peak resolution.

Stationary phases range from hydrophobic to hydrophilic for polarity-driven LC separations. Hydrophobic phases have alkyl groups like C30, C18, C8, cyano and phenyl groups covalently bound to the silica backbone. For polar analyses, hydrophilic interaction liquid chromatography (HILIC) employs bare silica or amide-bound silica particles as the stationary phase. A stationary phase is selected based on the relative polarity of the target analytes. Reversed phase chromatography uses a non-polar stationary phase with an initial polar mobile phase during sample loading to ensure partitioning or loading of target analytes onto the stationary phase. To induce elution, solvation strength is increased by decreasing the polarity of the mobile phase. Normal phase chromatography employs the opposite strategy using polar columns and an initial non-polar mobile phase with increased mobile phase polarity to induce elution.

1.2 Atmospheric Pressure Ionization (API), an LC-MS Interface

Multiple theories have been postulated to account for the high efficiency and very rapid conversion of liquid samples into the gas phase by ESI. The ionization and transition from liquid to gas phase that occurs in an ESI source has been observed in detail and theories that account for the observations have been postulated. The two most prominent theories are the ion evaporation model (IEM)[23,24] and the charged residue model (CRM)[25]. IEM theorizes that as solvent evaporation occurs and a droplet reaches a certain radius, called the Rayleigh limit[26,27], the field strength at the surface of the droplet becomes great enough to cause the desorption of single solvated ions. CRM theory states that electrospray droplets undergo evaporation and fission cycles, eventually leading to progeny droplets that contain an average of one analyte ion[28,29]. Gas-phase ions form after the remaining solvent molecules evaporate, leaving the analyte with the charges that the droplet carried. There is considerable evidence that demonstrates either directly or indirectly that small molecular ions transition from the liquid into the gas phase through IEM mechanisms, while larger molecular ions (i.e., folded proteins, DNA, and RNA for instance) form by CEM mechanisms. A third model proposing combined charged residue-field emission has been proposed[30], and another called the chain ejection model (CEM) theorizes the ionization and transition to the gas phase of disordered polymers (unfolded proteins)[31]. IEM is generally used to describe ESI of small molecules, and CRM is used to describe ESI of molecules like proteins. Suffice it to say, the application of a high voltage, low amperage direct current field to the tip of an LC column within the ion source of a mass spectrometer induces a process in which, first, solvated molecules capable of accepting or losing a proton or an electron become charged or ionized, and second, the disordered flow of the liquid as it exits the tubular column is changed into a relatively ordered process in which solvent is rapidly evaporated to release ionized molecules in the gas phase. Briefly, solvated ions of opposite charge to the ion source electromagnetic field accumulate at the solvent surface causing the liquid to assume a characteristic shape called the “Taylor cone” upon exiting the end of the column. Subsequently, sinusoidal waves called “waveform varicosities” form at the tip of the

Taylor cone which, as they increase in magnitude, eventually cause charged solvent droplets to "pinch off" and be ejected into the ion source. This occurs when ions of opposite charge to the ion source field accumulate just under the surface of the solvent. As the small droplets subsequently fly through the ion source electromagnetic field towards the MS inlet, the solvent evaporates while ions of like charge accumulate at the surface of each droplet. Eventually, as the ions of like charge become more and more crowded at the surface of the droplets, the repulsive coulombic forces between the ions overcome the liquid surface tension, the Rayleigh limit, causing the droplet to explode, a process called "coulombic explosion", to form smaller droplets[32]. This process is repeated extremely rapidly until single ions in the gas phase exist within the ion source (Figure 1). Because of the electromagnetic field that exists because of the DC voltage offset between the ion source and the mass spectrometer, the gas phase ions then migrate into the high vacuum of the mass spectrometer where they are sorted according to their mass-to-charge ratio using both direct and alternating current electromagnetic fields. During ESI, the kinetic energy involved in the process also characteristically fragments solvated molecules creating a set of characteristic fragment ions which constitute a highly characteristic mass spectrum for the original parent or precursor molecule; the spectrum is also very specific to the conditions under which it was produced. This spectrum is a highly specific "fingerprint" for the original molecule that can be used to not only identify it but to quantify it[32].

Positive ions observed from ESI by MS analysis include what were historically called quasi- or pseudo-molecular ions but which are now called precursor ions[33]. Under the influence of a negative electromagnetic field, they are formed by the addition of a hydrogen cation or of another cation such as a sodium ion symbolized as $[M+H]^+$ or $[M+Na]^+$, respectively. Negative precursor ions, formed when the ion source is under the influence of a positive electromagnetic field, are generated by the removal of a hydrogen nucleus, $[M-H]^-$. Multiply charged ions, denoted $[M+nH]n^+$, are often observed for larger molecules capable of accepting or losing multiple cations or hydrogen nuclei, respectively. For large macromolecules, such as proteins,

DNA, and RNA, there can be many charge states, resulting in a characteristic charge state envelope. All these are even-electron ion species where only electrons are added or removed, unlike in some other ionization sources.

The advent of electrospray ionization provided a robust interface between LC and MS systems. However, with many biological applications where target analyte concentrations in samples were extremely low or the sample size was extremely limited (low μL) making the absolute amount of target analyte available for analysis very low, LC-ESI/MSMS methods were not sensitive enough. In 1993, Gale and Smith reported significant sensitivity increases could be achieved using flow rates below conventional LC flow rates down to 200 nL/min[34]. The development of nano-electrospray or nanospray (nanoESI) tips was first described almost simultaneously by Wilm and Mann[35], and Emmett and Caprioli[36] in 1994. Wilm and Mann demonstrated that stable ESI could be sustained down to 25 nL/min using a self-fed system, and Emmett and Caprioli demonstrated that LC flows down to 300-800 nL/min markedly improved sensitivity for peptide and protein analyses. To successfully use nanoESI, very low diameter nanoESI emitter tips that provide consistent spray are required. Since ion desolvation efficiency during ESI is inversely proportional to the diameter of the initial solvent droplets produced[37], and since initial droplet size is directly proportional to the diameter of the Taylor cone formed which is, in turn, directly proportional to the diameter of the solvent column, the development of very low diameter emitter tips produced substantial sensitivity increases for mass spectrometric analysis. The manufacture of tips that spray consistently was the rate limiting step for the commercial success of nanoESI. Eventually, the availability of reproducible, robust nanoESI emitter tips contributed substantially to increases in LC-nanoESI/MS method sensitivity. Femtomolar sensitivity has been regularly reported[38,39,40,41], and, in my laboratory, we demonstrated sensitivity in the low zeptomole range for clean cerebrospinal fluid samples containing neurotransmitters and peptides at flow rates of 25 nL/min with self-feeding nanospray tips (data not published). In addition, since the flow volume was reduced substantially,

nanospray tips could be positioned closer to the MS inlet. With HPLC or UPLC, to take advantage of the sensitivity increases provided by nanoESI, post-column flow rate had to be reduced to flows which the tips could accommodate. This required post-column splitting with normal HPLC or UPLC in which a large proportion of the eluent was discarded leading to reduced sensitivity.

A new, miniaturized LC system called nanoLC was developed in 1988 by Karlsson and Novotny to provide greatly reduced flow rates[42]. Its development was primarily driven by biological applications requirements, particularly proteomics research. Due to lack of sensitivity, HPLC or UPLC techniques could not be successfully used to analyze biomedical samples containing low concentrations of proteins and peptides or to analyze the very small sample sizes often available in these applications. To interface nanoLC, which operates at ambient pressure, with a mass spectrometer, a miniature ion source had to be developed to reliably nebulize samples at the very low flow rates employed. The result was the nano-Electrospray Ionization (nanoESI) source that uses emitter tips with very small diameters. The marked sensitivity increases provided by nanoESI made nanoLC/MS ideal for these applications[35,41]. The development of nanoLC also necessitated the continued development of solid phases to maximize sample component recovery efficiency while maintaining sufficient column capacity[43]. This has resulted in the use of nonporous stationary phases coated with longer C-chain backbones including C30 stationary phases. Specifically, nanoLC has been very useful for the analysis of relatively large biopolymers including proteins, polynucleotides and lipids. This manuscript describes its use in the analysis of enzymatic sample digests to characterize proteins in clinical samples.

Following the development of ESI, another API method, atmospheric pressure chemical ionization (APCI), was developed in which ionization of solute molecules takes place following the initial nebulization and evaporation of the LC solvent by a relatively hot, inert sheath gas (Figure 1)[44]. Unlike ESI in which ions are produced in the solvent through the direct protonation

or deprotonation of target molecules, APCI ionization of target molecules takes place through an intermediate ion once the target molecules are aerosolized, hence the name chemical ionization. In the source, a corona discharge needle emits high energy electrons producing a glow discharge (corona) of high energy electrons which cause electron impact ionization of gas phase solvent molecules which, in turn, transfer protons to the molecules of interest in the gas phase. APCI, a 'softer' ionization method than ESI, generates primarily $[M+H]^+$ and $([M-H]^-$ ions in positive and negative modes, respectively. Consequently, APCI does not commonly produce multiply charged ions. In general, ESI is better suited for more polar, non-volatile molecules while APCI is better suited for ionization of non-polar, semi-volatile low molecular weight molecules.

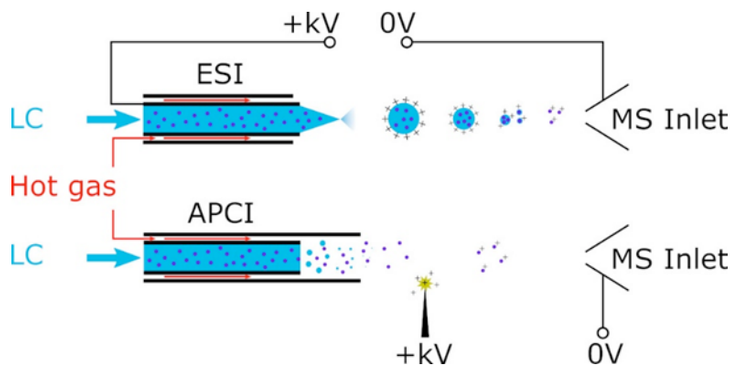


Figure 1. Cartoon of the atmospheric pressure ion sources, ESI and APCI, not drawn to scale.

To induce efficient ionization and avoid contamination of the high vacuum front end of the mass spectrometer, LC solvents and additives must be clean and volatile. Mobile phases commonly used for ESI-MS include water, acetonitrile (AcN), methanol (MeOH), ethanol (EtOH), isopropyl alcohol (IPA) and dichloromethane (DCM). Hydrophilic and hydrophobic analyte responses generally are optimal in polar and nonpolar solvents, respectively[45,46,47,48]. The pKa of target analytes should be considered and pH buffering used to optimize analyte solvation, ionization, and chromatography (peak shape). Buffers must also be volatile, and concentrations should not be above 25 mM[49]. Use of nonvolatile buffers or higher concentrations can lead to

substantial loss in ionization efficiency due to source contamination with the accumulation of buffer salts on source and MS front-end surfaces. Commonly used MS-friendly volatile buffers used include formic acid (FA), ammonium acetate, and ammonium formate. Trifluoroacetic acid, a weak acid commonly used with a corresponding base in biological applications as a buffer, suppresses ionization in ESI precluding its use in ESI. Phosphate buffers, again commonly used in biological experimental protocols, should also be avoided because phosphate salts accumulate on ion source and MS front-end surfaces quickly suppressing ionization.

1.3 Mass Spectrometry (MS)

Mass spectrometry is a highly specific and sensitive identification method in which ionized or charged molecules are manipulated in space and sorted according to mass-to-charge at high vacuum (i.e., 10^{-6}-10^{-9} Torr) using a combination of direct and alternating current electromagnetic or magnetic fields to indirectly determine molecular mass. Generally, DC fields are used to propel ions and RF fields are used to focus and direct ions. Molecular mass provides information about the chemical formula of the molecule, thereby, aiding identification. Because mass spectrometers measure mass-to-charge, the mass of an ion is inferred by the characteristic motion of the ions under an applied electromagnetic or magnetic field in a vacuum. A mass spectrometer is a system comprised of several subsystems, shown schematically in Figure 1. For most biomedical applications, the most common MS systems employ atmospheric pressure ionization (API), hence, Figure 2 illustrates that the ionization subsystem is outside the vacuum region; in other MS configurations, the ionization system can be inside the vacuum region.

Mass spectra contain very specific information based on the peak m/z assignments and intensities (areas or heights). Peak assignment is a measure of the m/z ratio of an ion, while intensity gives a measure of the relative abundance of the ion.

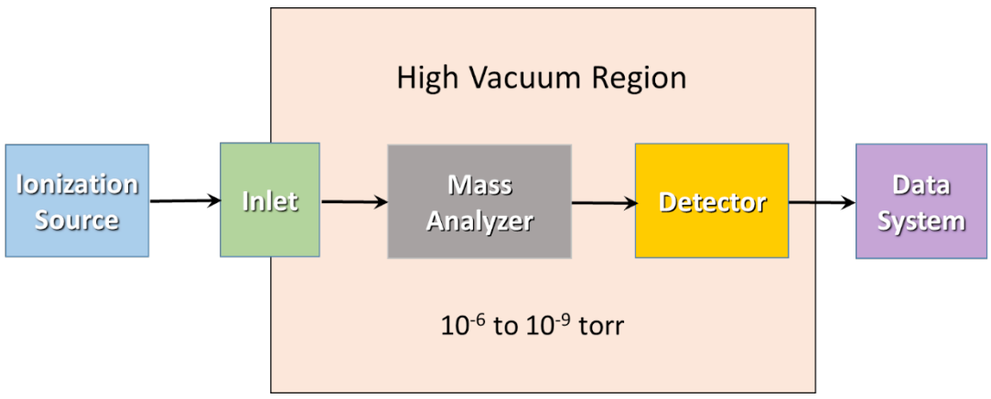


Figure 2. Schematic of a mass spectrometer

The intensity of a mass spectrum peak is not an absolute measure of the quantity of a specific compound in the sample. The quantity of ions detected for a specific compound from a given sample can vary as a result of many parameters that may differ across instruments, across different instrument setups, on the same instrument, or even with the presence of other compounds in the sample. Therefore, the measured ion abundance is only a relative measure of sample analyte concentration and requires calibration using standard materials before absolute quantification of CWAs, related chemicals and toxins can be attempted.

Resolution is important as it allows the user to determine if two m/z peaks can be successfully distinguished. Mass resolution or resolving power is usually measured by determining the ratio of the measured m/z at the MS peak apex to the peak width at half maximum height (full width at half maximum, FWHM) according to the following equation:

$$RFWHM = m/\Delta m$$

where RFWHM is resolution, m is the measured peak apical m/z and Δm is the FWHM. Resolution is important as it allows a user to determine if the mass analyzer will enable them to successfully distinguish between two peaks of similar mass. For molecules like many

biopolymers that can exist in multiple charge states, the charge states can only be resolved using high resolution instruments.

Mass spectrometers do not measure mass absolutely and must be mass calibrated regularly due to mass drift. Even for a well mass-calibrated instrument, a slight difference is expected between the measured ion mass and the exact mass of the ion calculated from the ionic formula accounting for the elemental known masses, their isotope masses, and what the charge state of the ion was. The difference is known as the mass error. It is expressed as a simple mass difference:

$$\Delta m = m_{\text{measured}} - m_{\text{calculated}}$$

where Δm is the mass error, $m_{measured}$ is the measured spectral mass and $m_{calculated}$ is the exact mass of the ion calculated from the known ionic formula. For singly charged species, the m/z value is equal to $m_{measured}$; for multiply charged ions, the measured m/z is multiplied by the z, or charge state to determine "$m_{measured}$" for the singly charged target molecule. Mass error may also be reported as the relative mass error (in this case expressed in parts per million (ppm) according to:

$$\Delta m = 10^6 \text{ x } \frac{m_{\text{measured}} - m_{\text{calculated}}}{m_{\text{calculated}}}$$

The confidence with which a specific ionic formula is assigned to a spectral peak is largely related to the mass error; the smaller the mass error, the more likely assignment is correct. If the mass error is small enough, then only one ionic formula may be possible. However, this level of mass accuracy is only achievable using high resolution, high mass accuracy time-of-flight (TOF) instruments and Fourier transform mass analyzers like Orbitrap and ion cyclotron instruments, and then only for molecules with molecular weights less than 1000 amu or Dalton (Da).

Most laboratories do not have access to high resolution/high mass accuracy instruments. To ensure that analytical specificity meets the requirements for confirmatory analysis with the more common nominal mass instruments available, the spectra from "unknowns" must be

compared to spectra from known authentic analytical standards. In addition, the use of analytical standards also allows for the construction of calibration curves from different known concentrations of the standards to use for quantitative analyses. A calibration curve is a plot of the instrument response as it changes with target analyte concentration. The responses are submitted to linear regression analysis[50,51,52]. The recognized requirements for valid quantification include that a curve be constructed of a minimum of five points to fit linear regression equations and six points to fit nonlinear regression equations. In addition, calibration samples must be made with target analyte-free sample matrix. When complicated sample preparation schemes are employed, an additional measure to ensure accuracy is the use of an internal standard, a chemical homolog isotopically labeled with deuterium or an analog for which the physical/chemical characteristics are very similar to the target analyte. Internal standard is added to all samples and calibrators prior to preparing them for analysis to account for any sample loss occurring during sample preparation.

1.3.0 Quadrupole Mass Spectrometry

Quadrupole mass spectrometers are the most common MS systems interfaced with LC. A quadrupole is comprised of four electrodes spatially arranged as shown in Figure 3[53]. The electrodes are hyperbolic as illustrated in Figure 3a. Other electrode or rod geometries are also possible but are rarely used. The rods are connected in opposite X and Y pairs and have RF and DC fields applied to them. Generally, RF is used to isolate or 'filter' ions, and DC fields are used to propel ions linearly toward a detector. The X and Y pairs oscillate through the RF field at opposite RF frequencies[54]. Ions are propelled down the axis of the instrument in the center of the quadrupole as illustrated in Figure 3. Mathematically, this can be modeled using the Mathieu differential equation[55]. Voltages applied to the two pairs of rods are selected so that ions of only one mass-to-charge (*m/z*) ratio can pass through the quadrupoles, from the ion source end of the instrument to a detector at the other end. Figure 4 is an actual picture of a set of quadrupole rods residing in an instrument. To produce a mass spectrum, RF voltages applied to the rods are

systematically stepped so that resonant ions of specific m/z ratios pass through the quadrupole at any given point in time. Each pass through the designated m/z range of this type is known as a scan. Multiple scans are averaged in order to produce a single mass spectrum.

A.) B.)

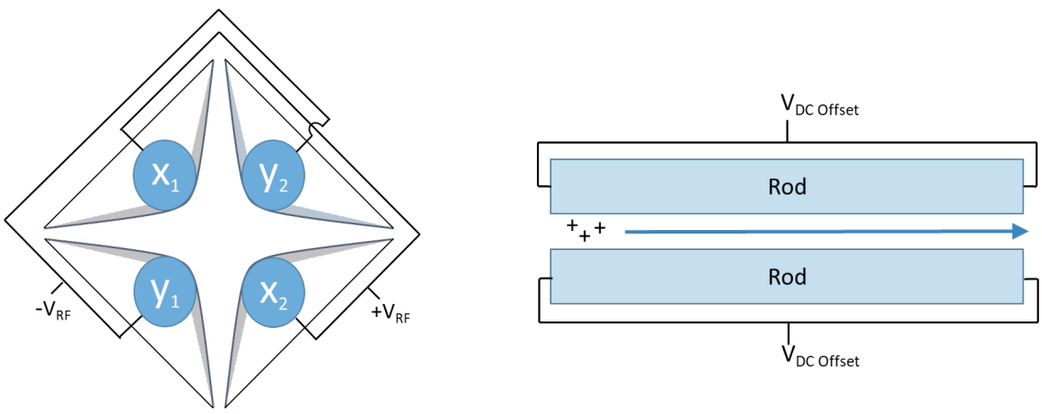


Figure 3. Cross (A) and sagittal (B) sections of a typical quadrupole analyzer where "X" and "Y" rod pairs and the applied RF and DC currents are identified. As illustrated, the "X" and "Y" rod pairs have out of phase RF current applied at any point in time.

The major limitations for 'single quad' instruments are that they are only capable of limited mass resolution and accuracy (hence they are often referred to as nominal mass resolution/accuracy instruments), and they are not capable of performing MSMS analyses to confirm molecular formula and fine structure. Multiple quads can, however, be linked together to perform MSMS analyses; these instruments are called triple quadrupole (QqQ) mass spectrometers.

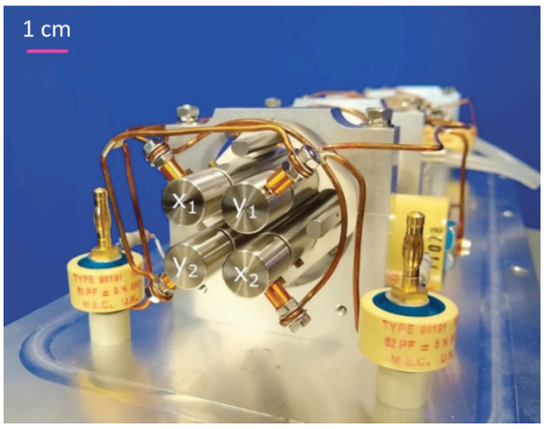


Figure 4. The end view of quadrupoles in a commercial quadrupole instrument.

1.3.1 Triple Quadrupole (QqQ) Mass Spectrometry

The development of triple quadrupole (QqQ) mass spectrometers[56], the first widely used tandem (MSMS) instruments, allowed scientists to conduct gas phase experiments in which precursor ions could be fragmented in a much more controlled and systematic way. A QqQ system is comprised of several differentially pumped mass filters and focusing units along with a collision cell[57]. A QqQ mass spectrometer conceptually consists of three sequential units instead of the single mass analyzer illustrated in Figure 2. In the first, a mass analyzer, specific precursor ions (previously called parent ions) or precursor ions of a specific mass range are given kinetic energy through the application of a direct current field (DC) offset across the analyzer causing them to fly from one end of the mass analyzer to the other. They are then filtered and isolated. A radio frequency alternating current (RF) field is applied to the mass analyzer to focus the ions into a beam, and resonant ions remain in the analyzer until being injected into the next MS unit by changing the field frequency, while non-resonant ions fly out of the analyzer to vacuum where they are discarded. In the second, called a "collision cell", precursor ions are systematically fragmented under somewhat controlled conditions into product ions (previously referred to as fragment or daughter ions). In the third, product ions are isolated by a mass

analyzer to identify either a single, diagnostic ion fragment or to produce a characteristic spectrum composed of all the product ions generated and identified from the fragmentation of the precursor. In some classes of tandem mass spectrometers, these three stages are performed in three separate, but sequential, mass analyzer elements. For example, in triple quadrupole (QqQ) or quadrupole/time-of-flight (QToF) instruments in which the ion flight path is linear, precursor ions are selected in an initial mass filter and injected into the second unit, a mass filter/collision cell. (In other mass spectrometer classes where the ion flight path is nonlinear, i.e., ion trap (IT), orbitrap (OT) and ion cyclotron resonance (FTICR) instruments, the three stages of analysis are conducted in the same mass analyzer). A highly pure, nonreactive gas such as Argon or Nitrogen is pumped into the collision cell to be used as a collision gas. As more or less kinetic energy is imparted to the precursor ions by increasing or decreasing the DC offset, respectively, across the cell, they collide with collision gas atoms or molecules with greater or lesser force causing somewhat controlled precursor ion fragmentation. The characteristic product ions are then filtered and scanned out of the cell for MS analysis in the third element, another mass analyzer. In this third stage, product ions are filtered to isolate and identify a diagnostic fragment(s) or to collect and compile all the product ions generated to form a characteristic precursor spectrum. The product ions are scanned out of the third mass analyzer into the detector where they are collected and converted into an electronic signature called a mass spectrum. The ESI interface along with MSMS provided the capability to conduct gas phase experiments in-order-to determine molecular fine structure. The result was that proteins, the most diverse bio-macromolecule, and the second most abundant macromolecule of mammalian life-forms, could be sequenced and fine structure, like post-translational amino acid residue modifications, verified.

Because the ability to separate ion species, resolution, is directly dependent on the length of the ion flight path, nonlinear mass spectrometers achieve their high mass resolution and mass accuracy through the theoretical ability to maintain ion flight paths of infinite length. These

instruments include three-dimensional IT, OT, and FTICR instruments. In these instruments, rather than flying down a linear space at high vacuum to a detector, the ions orbit within a space defined by an electromagnetic or magnetic field at high vacuum. Ions resonant with the field remain in the space and non-resonant ions are ejected either to vacuum or to a linear mass analyzer or series of mass analyzers/filters for fragmentation/focusing and then to a detector. These instruments can, therefore, resolve the charge envelope of very highly charged biomacromolecules like proteins, polar lipids, carbohydrates, etc.

The QqQ mass spectrometer was developed from the single quadrupole mass analyzer[56]. A QqQ MSMS is comprised of three quadrupoles in series (Q1, q2, and Q3). Usually, Q1 acts as a mass filter set to specific RF voltages through which only resonant ions of one m/z ratio or a range of ratios effectively pass and with a DC offset to propel ions towards the detector. As selected precursor ions are propelled into q2, a collision cell, they collide with inert gas molecules with specific kinetic energy imparted by adjusting the DC voltage offset applied to the cell[57]. Inert or relatively inert gases most commonly employed in a collision cell include argon, nitrogen, and less commonly helium[58]. Collisions impart internal energy to the precursor ions which results in their fragmentation into product ions; this is known as collisionally induced dissociation (CID)[59] or collisionally activated dissociation (CAD)[60]. Q3 functions as a second mass filter from which product ions are propelled for mass analysis in a dectector[56]. Product ion masses provide information about the formula and structure of the original precursor ion. Often, q2 is replaced with a hexapole or octapole ion guide which are like quadrupoles in design, except they consist of 6 or 8 electrodes, respectively, in a hexagonal or octagonal arrangement, and cannot be used as mass analyzers.

Normal QqQ MSMS processes conducted include product ion scans, precursor scans, neutral loss scans, and selected/multiple reaction monitoring[59,60,61,62]. During product ion scanning (Figure 5), the RF field in Q1 is set to select a single resonant ion at a known m/z, which is subsequently fragmented in q2[63]. All the product ions are then injected into Q3 in

which the RF is stepped through the frequency of the entire potential m/z range of all product ions[64,65]. This produces a mass spectrum of the precursor ion. Because the RF frequency must be stepped through the entire frequency range (generally commercial instruments step through the frequency range in 0.1 m/z increments per scan from low m/z to high), QqQ scan rates are relatively slow. It can take between 0.02 and 0.1 seconds to perform each quadrupole scan. Since an LC peak may only be a few seconds wide, and each peak must be sampled at least 10 times across the product ion m/z range to accurately measure a peak area for quantification purposes, this severely restricts the number of complete product ion scans that can be collected for each peak. For example, it would take a QqQ scanning across a 450 m/z range (4500 samplings/scan cycle) 9 seconds to produce a product ion spectrum from precursor ion isolation. Therefore, the other QqQ functions have been developed to restrict the number of scanning cycles required to produce the required spectrometric information. Some classes of compounds produce common MSMS fragment ions. Therefore, the precursor scan was developed to identify all precursors that produce a common fragment ion[65]. To accomplish this, the RF field in Q3 is set to filter only that common m/z while Q1 scans across a mass range thought to encompass the mass range of any targeted precursor. By restricting m/z scanning to only that required to filter specific ions through Q3, the time is greatly reduced as compared to full product ion scanning. This can only be used for targeted analysis. A third mode is called neutral loss scanning. During a neutral loss scan, both Q1 and Q3 are scanned simultaneously but using RF fields that select for a constant m/z difference (Q3 is always selecting ions at an m/z that is a specific m/z less than that selected by Q1). This method selectively identifies all ions which lose a common neutral fragment during q2 fragmentation. As compared to the precursor scan, the coordination of Q1 and Q3 scanning in a neutral loss scan reduces the time required to make one full mass scan, resulting in a significant increase in data acquisition rate. As with the precursor scan, the limitation for neutral loss scanning is that only molecules capable of losing the specific neutral fragment are filtered and identified. A fourth option is selected/multiple reaction monitoring

(SRM/MRM). In SRM/MRM, the RF in Q1 is set to select for a single precursor m/z, and the Q3 RF is set to filter only one, single reaction monitoring (SRM), or multiple reaction monitoring (MRM), precursor ion m/z, greatly reducing the time required to complete a scan cycle. In SRM/MRM, the QqQ will detect only targeted precursor ions in a sample which can be fragmented into preselected product ion(s). The use of MRM instead of SRM increases the specificity of the method because valid MRM identification requires not only the identification of characteristic precursor and product ions, but product ion relative abundances must be characterized and match to those from a reference standard to increase identification confidence.

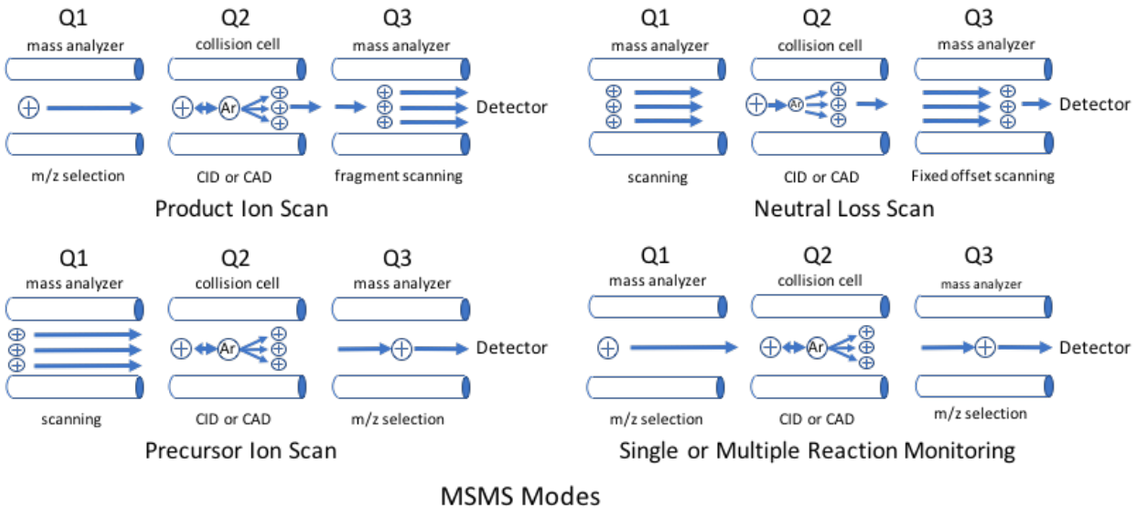


Figure 5. Operational modes in a triple quadrupole MSMS instrument. During product ion scanning, Q1 operates under a fixed RF frequency to select target ions of a specific m/z, Q2 operates as a collision cell to fragment all transitioning ions, and Q3 operates in scan mode over the expected fragment ion m/z range to obtain a full scan of the precursor ion. During precursor ion scanning, Q1 operates in full scan mode over the desired m/z range, Q2 operates as a collision cell to fragment all transitioning ions, and Q3 operates in single reaction or multiple reaction monitoring mode to isolate specific fragment ions. During neutral loss scanning, Q1

operates in full scan mode, Q2 operates as a collision cell, and Q3 operates scanning with a fixed RF offset. Finally, during SRM or MRM mode, Q1 operates as a mass filter selecting specific precursor ions which are fragmented in Q2. Q3 also operates as a mass filter monitoring specific single of multiple product ions. SRM or MRM mode is the most sensitive mode but least specific regarding the identification of precursor ions and product ion scanning mode is the least sensitive but most specific mode for identifying precursor ions.

1.3.2 **Time of Flight (ToF) Mass Spectrometry**

QqQ mass analyzers are limited in two ways. One of these weaknesses mentioned previously is that in a QqQ, two quadrupoles must be scanned in time, and as quadrupole scanning is relatively time consuming as compared to the width of LC peaks, this restricts the amount of m/z data that can be collected for each LC peak. The second weakness is that QqQ MS instruments are nominal mass instruments with limited mass resolution and mass accuracy capable of fully resolving only peaks of single m/z difference. Time of Flight (ToF) instruments are capable of relatively high mass resolution which is dependent almost exclusively, as with all mass spectrometers, on the ion flight path length. In a simple ToF instrument post-ionization, a packet of ions is accelerated by a repeller set to a high voltage (DC) relative to the draw out grid using the same charge state as the target ions. The ions initially experience a potential of 2.5 kV propelling them down the DC field gradient so that all ions accumulate 2.5 keV in kinetic energy by the time they reach and pass through the grid. Figure 6 is a cartoon illustrating a very simple ToF mass analyzer. The ions are propelled down the ToF through an RF field-free region to a detector. Normally, velocity (v) is directly related to kinetic energy (Ek), where $Ek=1/2mv^2$, but not all ions of the same m/z have exactly the same Ek so are propelled with a distribution of velocities. As ions travel through the field-free region, they are separated according to their velocities arriving at the detector at different times. Arrival time distributions can be mass calibrated providing all ions start traveling at the same time. Therefore, voltages in the ToF

pusher or repeller region are pulsed – the starting time of each ion packet through the ToF corresponding to a specific pulse.

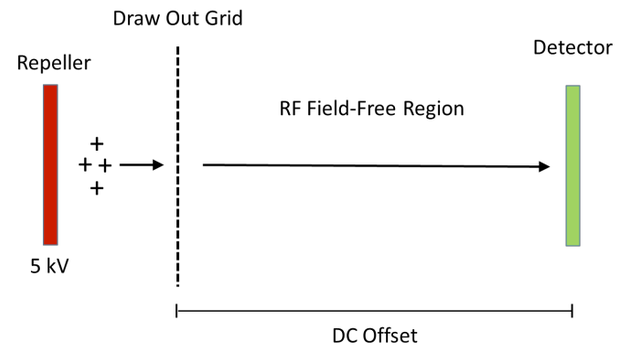


Figure 6. The schematic of a simple ToF instrument.

The two QqQ weaknesses can be overcome by adding a ToF mass analyzer to a QqQ[65,66]. This hybrid instrument is called a quadrupole-time of flight (QToF) instrument. In a QToF, a packet of ions is ejected from the terminal Q region and propelled orthogonally down the ToF, employing pulsed DC current from a "pusher plate". The mass resolution of a ToF primarily depends on the length of the ion flight path, but in part, also depends on the Ek distribution (and directly related velocity distribution) of the ions. Because the ions are ejected from the QqQ into the ToF orthogonally, the Ek distribution of the ions as they originate from the QqQ is minimized. One way of increasing resolution in a ToF is to position a reflectron or ion mirror at the end of the flight tube[67]. It consists of a stack of ring electrodes held at constant voltage in line with the ion flight path. The ions enter the stack with a range of kinetic energies, and as they penetrate the stack they slow down and are repelled in the opposite direction. The ions with the highest kinetic energy penetrate the furthest into the stack. As they reverse flight direction and exit the ring stack, the relative linear spatial position of the ions is minimized increasing ToF resolution. A reflectron or ion mirror provides two purposes, it enables the ion flight path to be essentially doubled without physically doubling the instrument length, and it reduces the spatial

distribution of the ions resulting from the Ek distribution which improves resolution. There are instruments that also allow for multiple passes through the ToF to improve mass resolution even further.

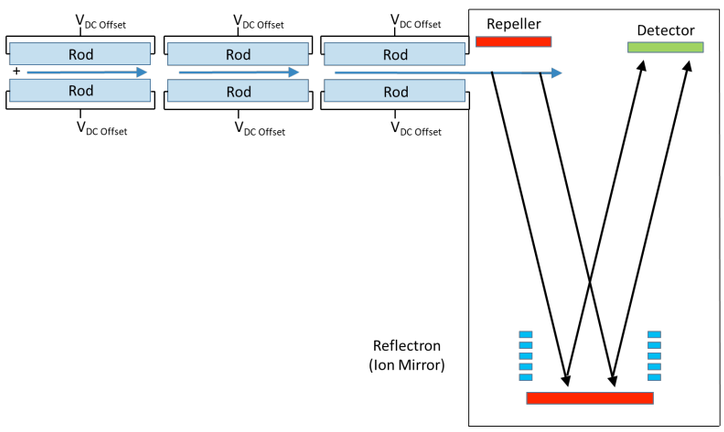


Figure 7. Schematic of a simple QqQ-TOF instrument. This is a hybrid instrument using quadrupole technology to perform low resolution ion separation and CID to perform fragmentation experiments prior to orthogonal introduction of ions into the TOF for high m/z ion sorting and analysis.

1.3.3 Orbitrap Mass Spectrometry

The development of the Orbitrap, a nonlinear, Fourier transform mass analyzer, has revolutionized the application of mass spectrometry to biomedical analysis[68]. Current instruments routinely capable of $\geq 10^6$ resolution and mass accuracy to parts per billion are being employed today. This provides the ability to analyze both small molecules and large biomolecules with very high specificity. In current Orbitrap instruments, ions are formed in an ion source, packets of ions are selected in a quadrupole or linear ion trap, and collected in a C-Trap, a curved field linear ion trap. The ions are then injected into the Orbitrap where they are trapped in an electrostatic field formed by specially shaped electrodes[69,70]. Figure 8 is a schematic of an Orbitrap instrument. The ions separate axially along the axis of the inner

electrode according to m/z, and the frequency of the ion oscillations around the spindle-shaped electrode inside the trap is mass calibrated to their m/z. Like FTICR instruments, the ion motion generates a current which is captured and measured in an ion transient which must be Fourier transformed to produce a mass spectrum. Unlike quadrupole, three-dimensional trap, linear ion trap, and ToF instruments, ions are not detected and measured by directly striking an electron multiplier or multichannel plate detector in an Orbitrap. Though Orbitrap instruments provide slightly less performance than FTICR MS instruments, they are cheaper, require less resources and space to maintain, and are still able to provide sufficient resolution to determine isotopic fine structure.

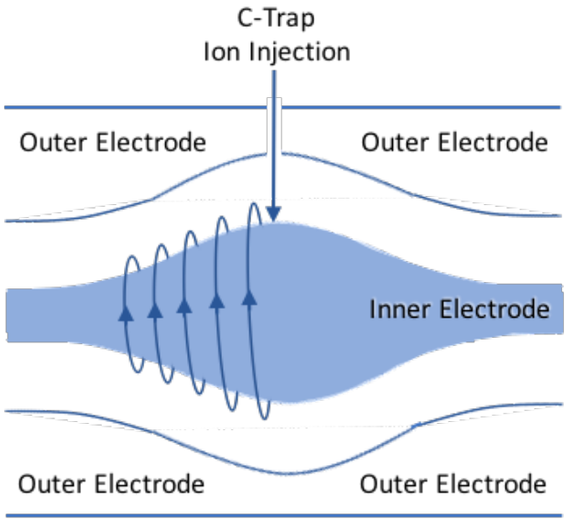


Figure 8. Cartoon of an Orbitrap instrument with the inner and outer electrodes showing the injection of ions from a C-Trap into the mass spectrometer. Ions are injected orthogonal to the axis of the electrodes and revolve around the inner electrode dispersing axially according to their m/z. As in FTICR instruments, ion movement generates a current called a transient which can be measured and m/z calibrated.

1.3.4 Thesis Aims

This thesis briefly describes the development and validation of two LC-MSMS regimens used for very different analytical purposes. It demonstrates the versatility and utility of interfacing LC with quadrupole, quadrupole-TOF and hybrid Orbitrap MSMS instruments to provide highly specific and sensitive analysis of biomedical samples for both targeted small and large molecular species and for the proteomic characterization of a variety of complex biomedical samples.

The goals of the first application of LC-MSMS to the analysis of biomedical samples were:

1. the development of a unique LC-QqQ nominal mass resolution MSMS method with associated sample preparation strategies to absolutely quantify, with high sensitivity, carfentanil, an extremely potent and clinically toxic semi-volatile small molecule opioid.
2. Determine and describe the percutaneous absorption kinetics of carfentanil in a live human epidermal model.

The goals of the second application of LC/MSMS to the analysis of biomedical samples were:

1. The development and validation of a unique sample preparation method that enabled samples to be rapidly prepared for LC-QTOF MSMS or LC-Orbitrap MSMS proteomic analysis.
2. The identification and characterization of targeted toxin proteins in complex biomedical samples.
3. To describe the proteomes of a variety of highly complex biomedical matrices.

Chapter 2 describes the validation of a quantitative carfentanil method that was published in 2022 by editor invitation in a special issue of the peer reviewed journal, *Rapid Communications in Mass Spectrometry,* celebrating the advent of quadrupole mass spectrometry[72]. A triple quadrupole MSMS method was developed and used to measure and

describe percutaneous carfentanil absorption kinetics in a live human epidermal model. Results from these studies have been published in US Army Toxicity Report No. S.0055513-18, July - September 2018 and in a journal article in 2019 entitled "In vitro dermal absorption of carfentanil" in the peer reviewed journal, *Toxicology In Vitro*[73]. The impetus for the study was to determine whether alcohol-containing skin decontaminating solutions would endanger law enforcement and medical first responders through enhancing percutaneous absorption of carfentanil should they be exposed. The results from these experiments demonstrated the capability of LC coupled with nominal resolution QqQ to absolutely quantify small semi-volatile drugs with relatively high sensitivity in biomedical samples.

Chapters 3 and 4 describe a tryptic digestion method developed for use in bottom-up LC-MSMS protein identification. Chapter 3 describes its use in the identification, with high confidence, specific target proteins in biomedical samples[74], and chapter 4 describes results from its use to perform proteomic characterization of several complex biomedical matrices. It is a very rapid and simple method requiring no specialized equipment or expensive reagents. A manuscript describing the targeted protein data was published in the peer-reviewed journal, *Analytical Chemistry*, in 2021. The method is currently being used very successfully by Merck and Company for targeted protein analysis in one of their high-volume protein discovery pipelines. The method has proven to provide substantial time and monetary cost savings according to a personal communication from the Laboratory Director to a colleague. A manuscript describing the application of the digest method to the proteomic characterization of complex biomedical samples was published in the peer reviewed journal, *Analytical Chemistry,* in 2023[75].

Bibliography

1. J.B. Fenn, M. Mann, C.K. Meng, S.F. Wong, C.M. Whitehouse. Electrospray Ionization for Mass Spectrometry of Large Biomolecules, *Science,* 6 Oct 1989, Vol. 246, Issue 4926, pp.

64-71
DOI: 10.1126/science.2675315.

2. M. Dole, L.L. Mack, R.L. Hines, R.C. Mobley, L.D. Ferguson, M.B. Alice. Molecular Beams of Macroions. *J. Chemical Physics,* 1968, 49 (5): 2240–2249. Bibcode:1968JChPh49.2240D, doi:10.1063/1.1670391.
3. Birendra N. Pramanik, A.K. Ganguly, Michael L. Gross. Applied Electrospray Mass Spectrometry: Practical Spectroscopy Series. CRC Press, 28 February 2002, p. 4, ISBN 978-0-8247-4419-9.
4. Andrej Frolov, Anja Henning, Christoph Böttcher, Alain Tissier, Dieter Strack. An UPLC-MS/MS Method for the Simultaneous Identification and Quantitation of Cell Wall Phenolics in Brassica napus Seeds, *Journal of Agricultural and Food Chemistry*, Volume 61 (6), pp.1219-1227, 2013, 10.1021/jf3042648.
5. Hui Wei, Guodong Chen, Adrienne A. Tymiak, Characterization of Impurities and Degradation Products in Small Molecule Pharmaceuticals and Biologics, Mass Spectrometry for Drug Discovery and Drug Development, pp. 191-220, 2013, 10.1002/9781118516157.ch7.
6. Gerd Ulrich Balcke, Vinzenz Handrick, Nick Bergau, Mandy Fichtner, Anja Henning, Hagen Stellmach, Alain Tissier, Bettina Hause, Andrej Frolov. An UPLC-MS/MS method for highly sensitive high-throughput analysis of phytohormones in plant tissues, Plant Methods, Volume 8 (1), pp. 8-47, 2012, 10.1186/1746-4811-8-47.
7. Eric C. Huang, Jack D. Henion. LC/MS and LC/MS/MS Determination of protein tryptic digests, *J. American Soc. Mass Spec.*, Volume 1, Issue 2, March–April 1990, Pages 158-165.
8. Charles J. Barinaga, Charles G. Edmonds, Harold R. Udseth, Richard D. Smith. Sequence determination of multiply charged peptide molecular ions by electrospray–ionization

tandem mass spectrometry, *Rapid Communications in Mass Spectrometry*, Vol 3 (5), May 1989, pp. 160-164.

9. Matthias Mann, Chin Kai. Meng, John B. Fenn. Interpreting Mass Spectra of Multiply Charged Ions, *Anal. Chem*. 61, 15, 1989, pp. 1702–1708.
10. Leena Valmu, Henrik Alfthan, Kristina Hotakainen, Steven Birken, Ulf-Håkan Stenman, Site-specific Glycan Analysis of Human Chorionic Gonadotropin β-Subunit from Malignancies and Pregnancy by Liquid Chromatography—Electrospray Mass Spectrometry, *Glycobiology*, Volume 16, Issue 12, 1 December 2006, pp. 1207–1218, https://doi.org/10.1093/glycob/cwl034.
11. Han X, Gross RW. Electrospray Ionization Mass Spectroscopic Analysis of Human Erythrocyte Plasma Membrane Phospholipids, *Proc Natl Acad Sci USA,* 1994; **91**: pp. 10635 – 10639.
12. Hsu FF, Turk J. Characterization of Phosphatidylinositol, Phosphatidylinositol-4-phosphate, and Phosphatidylinositol-4,5-bisphosphate by Electrospray Ionization Tandem Mass Spectrometry: A Mechanistic Study, *J Am Soc Mass Spectrom.* 2000; 11: pp. 986 – 999.
13. E.L Esmans, D Broes, I Hoes, F Lemière, and K Vanhoutte. Liquid Chromatography–Mass Spectrometry in Nucleoside, Nucleotide and Modified Nucleotide Characterization. *J. Chromy A*, Vol 794, Issues 1–2, 23 Jan 1998, pp. 109-127.
14. Anders Nordström, Petr Tarkowski, Danuse Tarkowska, Karel Dolezal, Crister Åstot, Göran Sandberg, Thomas Moritz. Derivatization for LC-Electrospray Ionization-MS: A Tool for Improving Reversed-Phase Separation and ESI Responses of Bases, Ribosides, and Intact Nucleotides., *Analytical Chemistry,* Vol 76, 10, April 20, 2004, pp. 2869-2877.
15. Marvin C. McMaster, LC/MS A Practical User's Guide. Wiley Interscience, A John Wiley and Sons, Inc., publication, 2005, Ch. 1, pp. 1-7.

16. Kumar P.R., Dinesh S.R. and Rini R. LCMS – A Review and Recent Update, *World J. of Pharmacy and Pharmaceutical Sciences,* Volume 5, Issue 5, pp. 377-391.
17. Hans H. Maurer, Thomas Kraemer, Oliver Ledvinka, Christian J Schmitt, Armin A. Weber. Gas Chromatography-Mass Spectrometry (GC-MS) and Liquid Chromatography-Mass Spectrometry (LC-MS) in Toxicological Analysis Studies on the Detection of Clobenzorex and Its Metabolites Within a Systematic Toxicological Analysis Procedure by GC-MS and by Immunoassay and Studies on the Detection of α- and β-Amanitin in Urine by Atmospheric Pressure Ionization Electrospray LC-MS, *J. Chromatography B: Biomedical Sciences and Applications,* Volume 689, Issue 1, 7 February 1997, pp. 81-89.
18. Shibdas Banerjee and Shyamalava Mazumdar. Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte. *Int. J. Anal. Chem.*, Vol 2012, ID 282574, https://doi.org/10.1155/2012/282574.
19. Shibdas Banerjee. Empowering Clinical Diagnostics with Mass Spectrometry. *ACS Omega*, 5, pp. 2041-48.
20. Bilin Zhuang, Gabriele Ramanauskaite, Zhao Yuan Koa, Zhen-Gang Wang. Like Dissolves Like: A First-Principles Theory for Predicting Liquid Miscibility and Mixture Dielectric Constant, *Sci. Adv*., 2021; 7: eabe7275.
21. S.A. Willison, Wipe Selection for the Analysis of Surface Materials Containing Chemical Warfare Agent Nitrogen Mustard Degradation Products by Ultra-High Pressure Liquid Chromatography-Tandem Mass Spectrometry. *J. Chromatogr. A*, 1270, pp. 72–79 (2012). DOI: 10.1016/j.chroma.2012.11.013.
22. Matthias Wilm. Principles of Electrospray Ionization. *Mol. Cell Proteomics*, 2011 Jul;10(7):M111.009407. doi: 10.1074/mcp.M111.009407.
23. Elnaz Aliyari and Lars Konermann. Formation of Gaseous Proteins via the Ion Evaporation Model (IEM) in Electrospray Mass Spectrometry. *Anal. Chem.*, 2020 Aug 4, 92(15), p. 10807-10814. doi: 10.1021/acs.analchem.0c02290, Epub 2020 Jul 15.

24. Fernandez de la Mora. Electrospray Ionization of Large Multiply Charged Species Proceeds Via Dole's Charged Residue Mechanism. *Analytica Chimica Acta*., 2000, 406:93104, doi: 10.1016/S0003-2670(99)00601-7.
25. L. Rayleigh. On the Equilibrium of Liquid Conducting Masses charged with Electricity. *Philosophical Magazine*, 1882, Vol 14, Issue 1, pp. 184–186.
26. A. Gomez and K. Tang, Charge and Fission of Droplets in Electrostatic Sprays. *Physics of Fluids*, 1994, Vol 6, Issue 1, pp. 404–414.
27. S. Nguyen, J.B. Fenn. Gas-Phase Ions of Solute Species from Charged Droplets of Solutions. *Proc. Natl. Acad. Sci., USA*, 2007, Vol 104, Issue 4, pp. 1111–7, Bibcode: 2007PNAS.104.1111N, doi: 10.1073/pnas.0609969104, PMC 1783130, PMID 17213314.
28. M. Gamero-Castaño, Direct Measurement of Ion Evaporation Kinetics from Electrified Liquid Surfaces. *J. Chem. Phys*., 2000, Vol 113, Issue 2, p. 815, Bibcode: 2000JChPh.113.815G, doi:10.1063/1.481857. S2CID 36112510.
29. C.J. Hogan, J.A. Carroll, H.W. Rohrs, P. Biswas, M.L. Gross. Combined Charged Residue-Field Emission Model of Macromolecular Electrospray Ionization. *Anal. Chem.,* Jan 2009, Vol 81, Issue 1, pp. 369–77. doi:10.1021/ac8016532. PMC 2613577. PMID 19117463.
30. Lars Konermann. "Unraveling the Mechanism of Electrospray Ionization". *Anal. Chem*., 2013, 85 (1): pp. 2–9. doi:10.1021/ac302789c. PMID 23134552.
31. Takashi Nohmi and John B. Fenn. Electrospray Mass Spectrometry of Poly(ethylene glycols) with Molecular Weights Up to Five Million. *J. Amer. Chem. Soc.*, 1992, 114, 9, pp. 3241-3246, publication date (Print): April 1, 1992, doi: 10.1021/ja00035a012.
32. Neil Loftus. Gold Standard: Mass Spectrometry and Chromatography. Biochem (Lond) 2002, Vol 24, Issue 1, pp. 25–27.
33. Arnold Steckel, Gitta Schlosser. An Organic Chemist's Guide to Electrospray Mass Spectrometric Structure Elucidation, *Molecules*, 2019 Feb 10;24(3), p. 611. doi: 10.3390/molecules24030611.

34. D.C. Gale, R.D. Smith. Small Volume and Low Flow Rate Electrospray Ionization Mass Spectrometry for Aqueous Samples. *Rapid Commun. Mass Spectrom.*, 1993, 7 (11): pp. 1017–1021. Bibcode:1993RCMS7.1017G, doi:10.1002/rcm.1290071111.
35. M.S. Wilm, M. Mann. Electrospray and Taylor-Cone Theory, Dole's Beam of Macromolecules at Last?. *Int. J. Mass Spectrom. Ion Process.*, 1994, 136 (2–3): pp. 167–180. Bibcode:1994IJMSI.136167W, doi: 10.1016/0168-1176(94)04024-9.
36. M.R. Emmett, R.M. Caprioli. Micro-Electrospray Mass Spectrometry: Ultra-High-Sensitivity Analysis of Peptides and Proteins. *J. Am. Soc. Mass Spectrom.*, 1994, 5 (7): 605–613, doi: 10.1016/1044-0305(94)85001-1, PMID 24221962.
37. Denis Duft, Tobias Achtzehn, Rene Müller, Bernd A. Huber, Thomas Leisner. Coulomb Fission: Rayleigh Jets from Levitated Microdroplets. *Nature*, Vol. 421, 9 Jan 2003, p. 128, www.nature.com/nature.
38. M.T. Smoluch, M. Sällberg, J. Silberring. Nanospray Mass Spectrometry for Identification of Peptides. Application of a Novel Interface. *Acta Biochim. Pol.*, 1999, 46(2), p. 475-81.
39. C.L. Gatlin, G.R. Kleemann, L.G. Hays, A.J. Link, J.R. Yates 3rd. Protein Identification at the Low Femtomole Level from Silver-Stained Gels Using a New Fritless Electrospray Interface for Liquid Chromatography-Microspray and Nanospray Mass Spectrometry. *Anal. Biochem.,* 1 Oct 1998, 263(1), pp. 93-101, doi: 10.1006/abio.1998.2809.
40. J. Raymackers, A. Daniels, V. De Brabandere, C. Missiaen, M. Dauwe, P. Verhaert, E. Vanmechelen, L. Meheus. Identification of Two-Dimensionally Separated Human Cerebrospinal Fluid Proteins by N-terminal Sequencing, Matrix-Assisted Laser Desorption/Ionization-Mass Spectrometry, Nanoliquid Chromatography-Electrospray Ionization-Time of Flight-Mass Spectrometry, and Tandem Mass Spectrometry. *Electrophoresis*. 2000 Jun, 21(11), pp. 2266-83.

41. M.R. Emmett, R.M. Caprioli. Micro-Electrospray Mass Spectrometry: Ultra-High-Sensitivity Analysis of Peptides and Proteins, *J. Am. Soc. Mass Spectrom.*, 1994. DOI: 10.1016/1044-0305(94)85001-1.
42. K.E. Karlsson, M. Novotny. Separation Efficiency of Slurry-Packed Liquid Chromatography Microcolumns with Very Small Inner Diameters, *Anal. Chem.,* 1 Sept 1988, 60(17), pp. 1662-5, doi: 10.1021/ac00168a006.
43. Yusuke Kawashima and Osamu Ohara. Development of a NanoLC–MS/MS System Using a Nonporous Reverse Phase Column for Ultrasensitive Proteome Analysis, *Analytical Chemistry*, 2018, 90, 21, pp. 12334–12338.
44. E.M. Thurman, I. Ferrer, D. Barceló, Choosing Between Atmospheric Pressure Chemical Ionization and Electrospray Ionization Interfaces for the HPLC/MS Analysis of Pesticides, *Anal. Chem.*, 2001, 73(22), pp. 5441–5449, DOI: 10.1021/ac010506f.
45. Isaac Omari, Parmissa Randhawa, Jaiya Randhawa, Jenny Yu, and J. Scott McIndoe. Structure, Anion, and Solvent Effects on Cation Response in ESI-MS, *J. Amer. Soc. Mass Spectrometry,* 2019, 30:1750Y1757.
46. Risto Kostiainen and Tiina J. Kauppila. Effect of Eluent on the Ionization Process in Liquid Chromatography–Mass Spectrometry, *Journal of Chromatography A*, Vol 1216, Issue 4, 23 January 2009, pp. 685-699.
47. Anneli Kruve. Influence of Mobile Phase, Source Parameters and Source Type on Electrospray Ionization Efficiency in Negative Ion Mode, *Journal of Mass Spectrometry*, Volume 51, Issue 8, August 2016, pp. 596-601.
48. Ami Kageyama (Kaneshima), Akira Motoyama, and Mitsuo Takayama. Influence of Solvent Composition and Surface Tension on the Signal Intensity of Amino Acids in Electrospray Ionization Mass Spectrometry, Mass Spectrom (Tokyo). 2019 Nov 30;8(1): A0077. doi: 10.5702/massspectrometry.A0077.

49. Allison Brookhart, Mahika Arora, Michael McCullagh, Ian D. Wilson, Robert S. Plumb, Johannes PC Vissers, and Nikung Tanna. Understanding Mobile Phase Buffer Composition and Chemical Structure Effects on Electrospray Ionization Mass Spectrometry Response, Journal of Chromatography A, Vol 1696, 10 May 2023, 463966.
50. S. Berthier. Drug Monitoring and Clinical Chemistry. *Handbook of Analytical Separations,* 2004, pp. 113-128.
51. Oskar González, Rosa María Alonso. Methods of Therapeutic Drug Monitoring Including Pharmacogenetics. *Handbook of Analytical Separations*, 2020, pp. 115-134.
52. JoEllyn McMillan. Principles of Analytical Validation. *Proteomic Profiling and Analytical Chemistry*, 2013, pp. 205-215.
53. Paul, Wolfgang and Steinwedel, Helmut. Apparatus for Separating Charged Particles of Different Specific Charges, Patent US 2939952, published 1960-06-07, assigned to Wolfgang Paul.
54. de Hoffmann, Edmond, Vincent Stroobant. *Mass Spectrometry: Principles and Applications* (Second ed.). Toronto: John Wiley & Sons, Ltd., 2003, p. 65. ISBN 978-0-471-48566-7.
55. Gerald Teschl. Ordinary Differential Equations and Dynamical Systems. *Providence: American Mathematical Society*. 2012, ISBN 978-0-8218-8328-0.
56. R.A. Yost, C.G. Enke. Triple Quadrupole Mass Spectrometry for Direct Mixture Analysis and Structure Elucidation. *Anal. Chem.*, 1979, 51(12), pp. 1251–1264 DOI:10.1021/ac50048a002.
57. Jonathan Oyler, Kathleen Maistros, Benjamin Oyler, David Kilgour. Liquid Chromatography–Mass Spectrometry in Analysis of Chemicals Related to the Chemical Weapons Convention. *Wiley Encyclopedia of Analytical Chemistry: Applications, Theory, and Instrumentation*, 13 March 2020, https://doi.org/10.1002/9780470027318.a0408.pub3.

58. Fred W. MClafferty. Tandem Mass Spectrometry, *Science,* 16 Oct 1981, Vol 214, Issue 4518, pp. 280-287.

59. Wells JM, McLuckey SA (2005). Collision-Induced Dissociation (CID) of Peptides and Proteins, *Biological Mass Spectrometry. Methods in Enzymology*. Vol. 402. pp. 148–85. doi:10.1016/S0076-6879(05)02005-7. ISBN 9780121828073. PMID 16401509.

60. Sleno L, Volmer DA. Ion Activation Methods for Tandem Mass Spectrometry, *Journal of Mass Spectrometry*. Vol 39, Issue 10, 2004, pp. 1091–112.

61. Kermit K. Murray, Robert K. Boyd , Marcos N. Eberlin , G. John Langley , Liang Li, and Yasuhide Naito. Definitions of Terms Relating to Mass Spectrometry (IUPAC Recommendations 2013), *Pure and Applied Chemistry*, June 6 2013, Vol. 85, No. 7, pp. 1515–1609.

62. G. Kaklamanos, E. Aprea, and G. Theodoridis. Tandem-in-Space and Tandem-in-Time Mass Spectrometry: Triple Quadrupoles and Quadrupole Ion Traps, *Analytical Chemistry.* Vol 62, Issue 20, 1990, pp. 2162–2172.

63. G. Kaklamanos, E. Aprea, and G. Theodoridis. Mass Spectrometry: Principles and Instrumentation, *Encyclopedia of Food and Health*, 2016, pp. 661-668.

64. P.H. Dawson and D.J. Douglas. Quadrupoles, Use of In Mass Spectrometry, *Encyclopedia of Spectroscopy and Spectrometry (Third Edition)*, 2017, pp. 802-810.

65. H. Steen, B. Küster, M. Mann. Quadrupole Time-of-Flight Versus Triple-Quadrupole Mass Spectrometry for the Determination of Phosphopeptides by Precursor Ion Scanning, *Journal of Mass Spectrometry*, July 2001, Vol. 36, Issue 7, pp. 782-90, doi: 10.1002/jms.174.

66. Darren R Allen and Brett C McWhinney. Quadrupole Time-of-Flight Mass Spectrometry: A Paradigm Shift in Toxicology Screening Applications, *Clinical Biochemistry Review*, 2019 Aug, Vol 40, Issue 3, pp. 135–146. doi: 10.33176/AACB-19-00023.

67. Boris Mamyrin. Time-of-Flight Mass Spectrometry (Concepts, Achievements, and Prospects), *International Journal of Mass Spectrometry*, 22 Mar 2001, Vol 206, Issue 3, pp. 251–266.

68. S. Eliuk, A. Makarov. Evolution of Orbitrap Mass Spectrometry Instrumentation, *Annu. Rev. Anal. Chem.*, 2015, 8(1), pp. 61–80, DOI: 10.1146/annurev-anchem-071114040325.

69. Q. Hu, R.J. Noll, H. Li, A. Makarov, M. Hardman, R. Graham Cooks. The Orbitrap: A New Mass Spectrometer. *Journal of Mass Spectrometry*, 2005, Vol 4, Issue 4, pp. 430–443, DOI:10.1002/jms.856.

70. Roman A. Zubarev and Alexander Makarov. Orbitrap Mass Spectrometry, *Analytical Chemistry.,* 2013, Vol 85, Issue 11, pp. 5288–5296.

71. Michael W. Senko, Philip M. Remes, Jesse D. Canterbury, Raman Mathur, Qingyu Song, Shannon M. Eliuk, Chris Mullen, Lee Earley, Mark Hardman, Justin D. Blethrow, Huy Bui, August Specht, Oliver Lange, Eduard Denisov, Alexander Makarov, Stevan Horning, Vlad Zabrouskov. Novel Parallelized Quadrupole/Linear Ion Trap/Orbitrap Tribrid Mass Spectrometer Improving Proteome Coverage and Peptide Identification Rates, *Analytical Chemistry,* 2013, Vol 85, Issue 24, pp. 11710–11714.

72. Oyler JM, Maistros KJ, Kilgour D. Validation of a Quantitative Liquid Chromatography/ Hybrid Quadrupole-Linear Ion Trap MS/MS Method for Carfentanil and Caffeine in a Live Human Epidermis Model., *Rapid Communications in Mass Spectrometry*, 2022; e9420, doi:10.1002/rcm.9420.

73. Emily M. Lent, Kathleen J. Maistros, Jonathan M. Oyler. In Vitro Dermal Absorption of Carfentanil. *Toxicology in Vitro*, 2020 Feb 1, Vol 62, p. 104696.

74. Jonathan M. Oyler, Bao Q. Tran, and David P. A. Kilgour. Rapid Denaturing Organic Digestion Method for Targeted Protein Identification and Characterization, *Analytical Chemistry,* 2021, Vol 93, Issue 12, pp. 5046–5053.

75. Jonathan Oyler, Raymond F. Sullivan, Bao Q. Tran, Dhanwin Baker, Clare Coveney, David Boocock, Benjamin Oyler, Carole C. Perry, David P.A. Kilgour. Proteomic Interrogation of Complex Biomedical Samples Using the Rapid Denaturing Organic Digestion (DOD) Method, *Journal of Proteomics,* 2025, Vol 312, p. 105359.

Chapter 2

Validation of a Quantitative LC/Hybrid Quadrupole-Linear Ion Trap MSMS Method for Carfentanil and Caffeine in a Live Human Epidermis Model

Jonathan M. Oyler[1], Kathleen J. Maistros[2], David Kilgour[1]

[1]Nottingham Trent University, Clifton Campus, Nottingham, UK

[2]Laboratory Services Department, U.S. Army Medical Research Institute of Chemical Defense, Aberdeen Proving Ground, MD, USA

2.0 Abstract

In this study, a quantitative liquid chromatography/hybrid triple quadrupole-linear ion trap method was developed for carfentanil, an extremely toxic opioid. The method was also validated for caffeine, a molecule routinely used to monitor epidermal cell-culture viability. The method employed reversed-phase chromatography coupled with positive electrospray ionization and MRM monitoring to quantify carfentanil and caffeine against calibration curves formulated from authentic standards. Limits of detection (LOD) for the two compounds were determined employing 10:1 signal-to-noise requirements for all product ions with relative peak areas equal to ±20% of those observed for a mid-level calibrator. Calibration curves for seven formulated replicates of the two compounds met linearity requirements over at least four orders of magnitude concentration range. The accuracy was within ±20% of the actual, precision across results (%CV) was ≤15%, curve coefficients of determination (r^2) were ≥ 0.980 (correlation coefficient r > 0.990), and relative ion ratios of all qualifier ions were within ±20% of those for a mid-level standard. Blanks contained no target analyte or internal standard signal above baseline. Limits of quantification (LOQs) for carfentanil and caffeine were 230 pg/mL and 12 ng/mL, respectively. Precision and accuracy were determined by analyzing positive controls formulated in quintuplicate, by a different analyst, at three concentrations bracketing the method dynamic range. Intra-day accuracies (mean concentrations) for carfentanil and caffeine ranged

from 90.1-100.8% and 87.1-108.9%, respectively; inter-day accuracies ranged from 98.7-100.4% and 97.5-101.7%. Intra-day precision (%CV) over the dynamic range ranged from 1.31-8.88 and 1.49-6.72 for carfentanil and caffeine, respectively. Inter-day precision was evaluated using data collected from three separate days of analyses and ranged from 4.7-9.9 and 7.6-12.1 %CV for carfentanil and caffeine, respectively. The method was subsequently used to evaluate the percutaneous absorption kinetics of carfentanil in solution as a function of `solvent` composition using an *in vitro*, live human epidermis model. Counterintuitively, as previously reported, the addition of organic solvents to the formulations decreased rather than increased the percutaneous absorption rate of the ultra-potent opioid, carfentanil.

2.1 INTRODUCTION

One of the major strengths of quadrupole mass spectrometry is the ability to rapidly and reproducibly scan through relatively wide mass-to-charge ranges with great sensitivity and speed, especially when targeting limited characteristic ions using a selected ion monitoring mode. The subsequent development of triple quadrupole instruments resulted in the ability to reproducibly fragment a precursor molecule under controlled conditions and quantitatively measure the resultant product ions to characterize and quantify the parent molecule with even higher specificity and sensitivity[1]. Consequently, triple quadrupole instruments have become the "gold standard" for drug and small molecule quantification in toxicological, forensics, drug monitoring, and pharmaceutical analyses. In this study, a highly sensitive quantitative liquid chromatography/hybrid triple quadrupole-linear ion trap MS/MS method was validated for carfentanil and caffeine. Subsequently, the method was used in an *in vitro* protocol designed to evaluate percutaneous absorption kinetics of carfentanil in a live human epidermal model, as a function of solvent composition.

Fentanyl analogs, like other opiates and opioids, act directly on central nervous system μ-opioid receptors to induce bradycardia, respiratory depression, hypotension, depressed cough reflex, miosis, and sedation[2,3,4]. An extremely potent fentanyl analog, carfentanil (methyl 4-(1-

oxopropyl) phenylamino-1-(2-phenylethyl)-4-piperidine carboxylate-2-hydroxy-1,2,3-propane tricarboxylate), is a DEA Schedule II opioid (synthetic drug) that has been estimated to be 100 times more potent than fentanyl and 10,000 times more potent than morphine, a naturally occurring opiate[5]. The drug has until recently been considered a United States federally classified compound so there is little data available in the open literature on its physical/chemical or pharmacokinetic characteristics. The citrate salt was originally formulated (Wildnil) for use in veterinary medicine as a large animal tranquilizer but has not been commercially available since 2003. The intravenous LD_{50} of the citrate salt in mice and rats has been reported to be 3.39 and 18.75 mg/kg, respectively[6], and the subcutaneous median effective dose (ED_{50}) required to induce severe intoxication in African green monkeys was found to be 0.706 μg/kg[7]. Due to its ultra-potency and because it has no licit clinical uses, the clinical LD_{50} of carfentanil has not been determined, but the estimated human lethal dose is 20 μg (0.286 μg/kg), with significant clinical effects observed with exposure to as little as 1 μg[5]. This toxicity is within the range of classic organophosphate nerve agents, making it a potential agent of warfare and terrorism. There have been terrorist incidents in which humans have been exposed to carfentanil. In 2002, Chechen hostages in a theater in Moscow were presumably exposed to vaporized or aerosolized carfentanil and remifentanil, another highly potent opioid, resulting in 170 human deaths. Urine and clothing residue collected from three victims, analyzed by scientists at Porton Down, Wiltshire, England, contained the two fentanyl analogs[8,9].

Recently in the United States, the incidence of fentanyl-related overdose deaths has dramatically risen[10]. In the past, this resulted from the exposure of individuals to fentanyl-adulterated heroin, and cocaine and other psychostimulants (primarily methamphetamine)[11], but increasingly, especially during specific localized outbreaks, fentanyl analogs alone, including carfentanil, have been implicated[4]. With this dramatic increase in overdose cases, concern has grown for the safety of responding law enforcement and emergency medical personnel due to potential occupational exposure since these highly potent opioids are known to be readily

absorbed through skin and mucous membranes. Incidents of percutaneous and dermal fentanyl exposure have been reported. There have been multiple reports from emergency response personnel describing symptoms consistent with opioid exposure presumably from percutaneous, dermal and mucous membrane contact with on-scene material and/or potentially contaminated surfaces[12,13]. In a confirmed case resulting from skin and mucous membrane exposure, a veterinarian's facial skin and mucous membranes of the eye and mouth were exposed to a large animal veterinary formulation which splashed from a misfired dart. The individual exhibited classic opioid symptoms including nausea, respiratory depression, miosis, and hypotension[14].

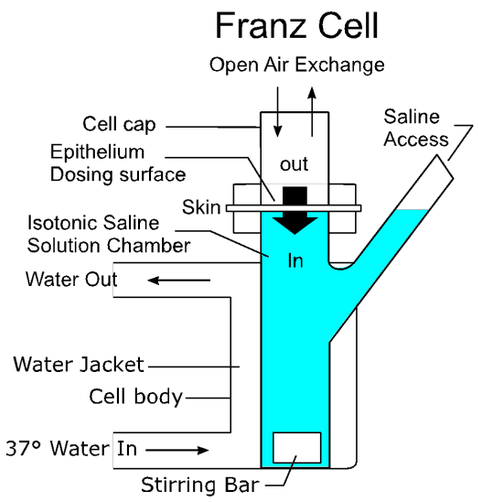
In addition, since fentanyl analogs are lipophilic and readily absorbed through the skin, concern has been generated in the medical community that the routine use of alcohol-containing hand sanitizers by law enforcement and emergency medical personnel could enhance dermal absorption rates. For this reason, the NIOSH, NIH, the US DEA, and the NIEH, NIH, have issued guidance documents strongly discouraging the use of alcohol-based hand sanitizers by first responders who are at risk of being exposed occupationally to fentanyl-related substances[15,16,17].

Due to its potency, until recently no human data from controlled studies existed describing percutaneous or dermal absorption of carfentanil. For this reason, an *in vitro* epidermal dosing study was designed and conducted using a live reconstructed human epidermal model (RhE) to measure percutaneous carfentanil permeability coefficients, absorption rates and permeation lag times as a function of the solvent in which it was formulated.

The *in vitro* Franz Cell system consisted of two chambers with an air-liquid interface between the upper and lower chambers (Figure 1). Live, cultured human epidermis with intact stratum corneum was suspended, stratum corneum side up (air side), at the air/liquid interface on Krebs-Ringer bicarbonate solution where cell integrity and viability could be maintained; caffeine was employed as a cell viability marker[18]. Dosing was performed in the upper chamber, onto the air-exposed stratum corneum, and samples were drawn from the lower chamber through

a port, prior to and at different time points after dosing. Epidermal cells were also dosed with caffeine, a drug routinely used to monitor epidermal viability in this model. This provided a live human model in which absorption kinetics could be studied as a function of dosing formulation solvent composition. In the *in vitro* study, carfentanil was formulated in four vehicles, water, ethanol, and two ethanol-containing hand sanitizers, for dosing. The pharmacokinetic results have been reported[19]. The goal of the study was to determine the effect of the use of hand sanitizers containing a lipophilic organic solvent as a "decontamination agent" on the percutaneous absorption of the highly potent opioid, carfentanil, in humans should an epidermal surface become contaminated with carfentanil in the course of entering and examining a forensic scene.

Figure 1. Franz Cell live human epidermal model.



Prior to conducting the pharmacokinetic study, a sensitive LC/hybrid triple quadrupole-linear ion trap (LC-MS/MS) method for direct quantification of carfentanil and caffeine in the RhE buffer, was designed and validated. The method incorporated another fentanyl analog, sufentanil, as an internal standard. (Theoretically, use of an isotopically labeled carfentanil internal standard should produce comparable or better method performance provided the label

remained attached to the fragments used.) This report describes the method validation process and method performance in detail.

2.2 MATERIALS AND METHODS

Carfentanil citrate (CAS# 61380-27-6; purity: 96.9%) was obtained from Edgewood Chemical and Biological Center. Caffeine (CAS# 58-08-2) was purchased from Sigma (St. Louis, MO). Sufentanil citrate (CAS# 60561-17-3) was purchased from Cerilliant Corporation (Round Rock, TX) for use as the analytical internal standard (IS). Water (18 MΩ) was produced by an in-house Millipore system acquired from MilliporeSigma, Burlington, MA. LCMS-grade methanol, acetonitrile, and formic acid were purchased from Fischer Scientific (Hampton, NH). All test systems, reagents, and chemicals were maintained according to manufacturer's instructions.

2.2.0 ANALYTICAL STANDARDS

Carfentanil and caffeine primary stock solutions were gravimetrically formulated from neat stock in ethanol and stored at -20 ºC in borosilicate vials until use. The IS, sufentanil, was commercially acquired in solution at 100 µg/mL in methanol (Lot FE013012-01), in flame-sealed borosilicate vials. Stock standards were formulated volumetrically using calibrated Eppendorf pipets. Calibration standards and controls were diluted volumetrically in water containing 0.1% formic acid fresh each day of analysis from primary standards. Calibration curve standards for carfentanil and caffeine ranged in concentration from 0.23-920 and 11.88-19,800 ng/mL, respectively.

2.2.1 ANALYTICAL METHODS:

2.2.1.0 LC Method:

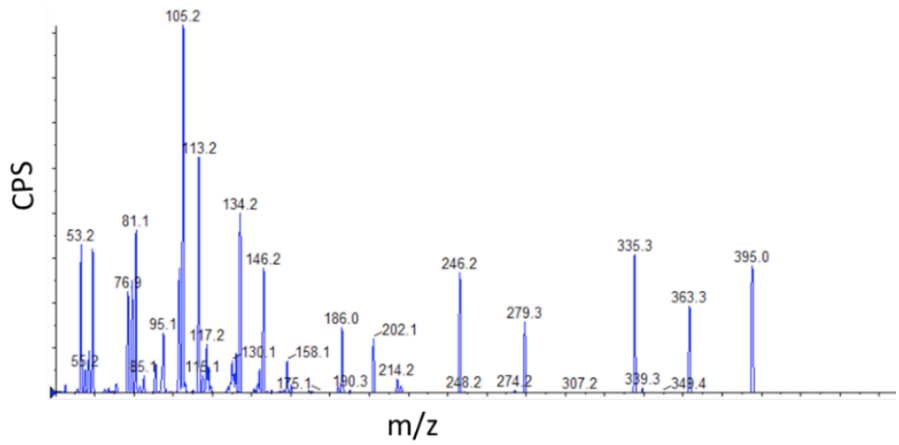
The LC system used was a Shimadzu Nexera UHPLC. A 5 µL sample aliquot was analyzed on a Phenomenex Kinetex column (2.1 x 100 mm x 1.7 µm F5, PN: 00D-4722-AN, SN# H18-042762). Mobile phase A consisted of MS-grade water with 0.1% formic acid, and

mobile phase B was MS-grade acetonitrile containing 0.1% formic acid. The mobile phase flow rate was 0.4 mL/min using an initial 0.1-minute hold time at 2% mobile phase B, a gradient profile of 2-90% mobile phase B over 4 minutes, and a 30°C column temperature.

2.2.1.1 MS/MS Method Development:

The MS/MS system employed for carfentanil and caffeine quantification was a SCIEX 4000 QTRAP, a nominal mass accuracy/resolution hybrid quadrupole-linear ion trap, operated in positive ESI mode. To develop the method, standards at 2 µg/mL in methanol with 0.1%FA were directly infused at 15 µL/min and analyzed in Q1 full scan mode to acquire a background-subtracted spectrum for each compound. The $(M+H)^+$ species for each compound was identified to use as precursor ion for MS/MS analysis; standards were directly infused to optimize source and Q1 conditions for each respective precursor. Again, samples were directly infused in product ion scan mode (precursors selected in Q1 followed by collision assisted dissociation (CAD) in Q2 with the resulting product ions scanned out of Q3 to identify the most abundant characteristic product ions for use in multiple reaction monitoring (MRM) analysis. During product ion scanning for both analytes, the declustering potential (DP) was held at 50, and the collision energy (CE) was ramped in 5 V steps from 0-120V for carfentanil and 0-60V for caffeine. CAD of carfentanil produced many highly characteristic product ions to use for MRM analysis; again, CAD conditions were optimized for the selected ions. A heat plot of caffeine CAD produced a more limited number of prominent characteristic fragment ions. Figure 2 illustrates cumulative product ion scan spectra over the respective CE ranges for carfentanil and caffeine, and Figure 3 illustrates product ion intensities for carfentanil as a function of CE.

A)



B)

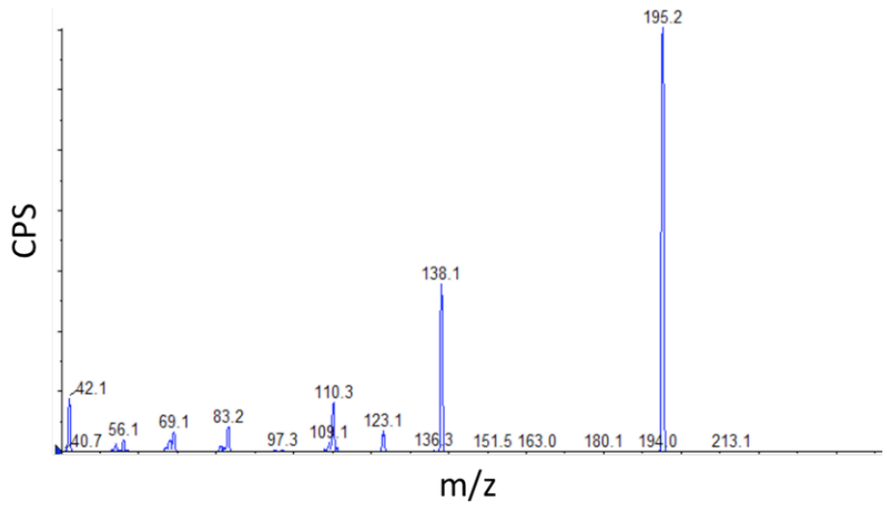


Figure 2: Summed product ion scan spectra for carfentanil and caffeine; A) Carfentanil at 2 µg/mL in methanol with 0.1% formic acid infused at 15 µL/min (MS2 Scan of *m/z* 395.0, declustering potential = 50, CE ramped from 0-120V, and CAD = MED); and b) Caffeine at 20

µg/mL in methanol with 0.1% formic acid infused at 15 µL/min (MS2 Scan of *m/z* 198.0, declustering potential = 50, CE ramped from 0-60V, and CAD = MED).

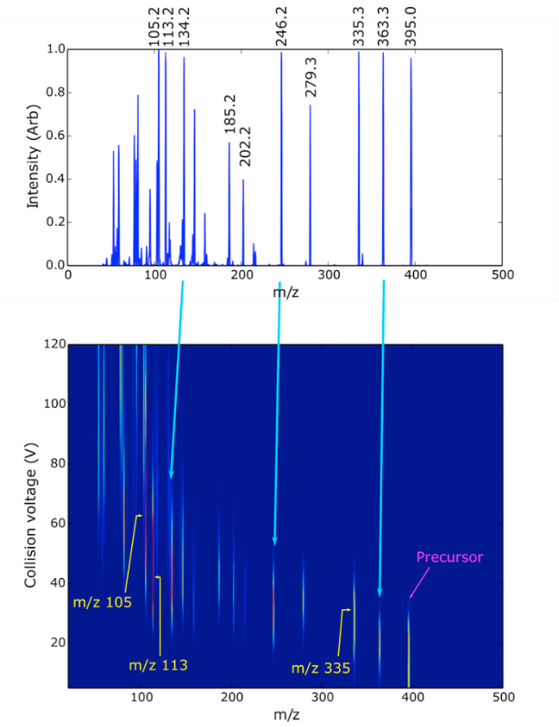


Figure 3. Carfentanil total fragment ion mass spectrum and heat map of product ion intensity as a function of CE.

The most abundant and characteristic ions were then selected for MRM method development. Table 1 lists the characteristic precursor and product ion m/z values for carfentanil, D_5-carfentanil, caffeine, and $^{13}C_3$-caffeine with select optimal acquisition conditions.

Q1 m/z	Q3 m/z	Time (msec)	ID	DP (volts)	CE (volts)
395.0	335.0	45	Carfentanil	60.0	30.0
395.0	113.0	45	Carfentanil	60.0	43.0
395.0	105.0	45	Carfentanil	60.0	63.0
400.0	340.0	45	D_5-Carfentanil	60.0	26.0
400.0	113.1	45	D_5-Carfentanil	60.0	46.0
400.0	105.0	45	D_5-Carfentanil	60.0	65.0
195.0	138.0	45	Caffeine	60.0	15.0
195.0	110.0	45	Caffeine	60.0	20.0
195.0	83.0	45	Caffeine	60.0	41.0
198.0	140.0	45	$^{13}C_3$-Caffeine	65.0	29.0
198.0	112.0	45	$^{13}C_3$-Caffeine	65.0	44.0
198.0	84.0	45	$^{13}C_3$-Caffeine	65.0	37.0

Table 1. Precursor and product ion m/z values along with important optimal acquisition conditions and parameters (DP = declustering potential, CE = collision energy).

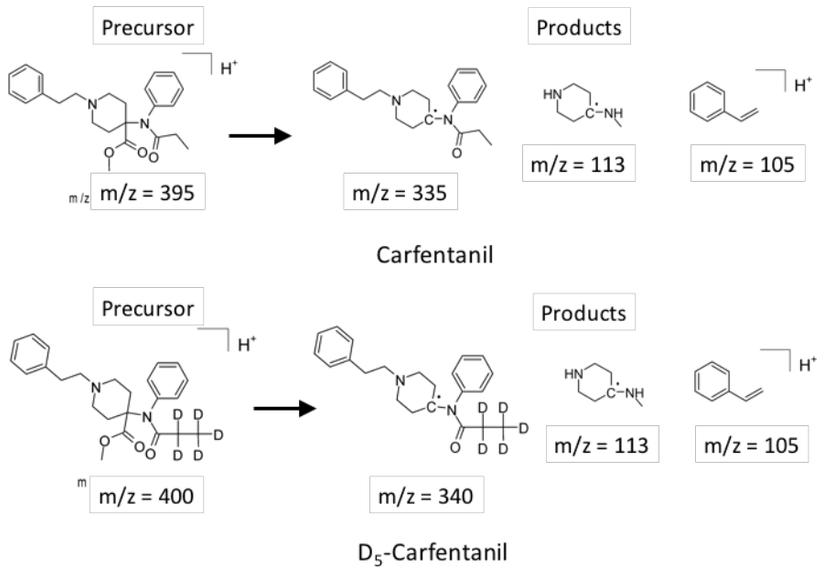


Figure 4. Precursor and product ion structures and m/z values for carfentanil and D_5-carfentanil.

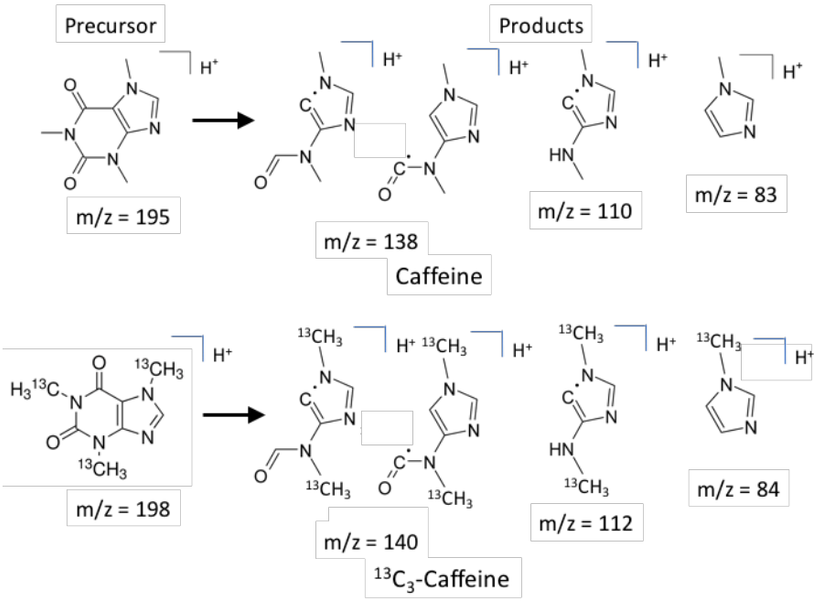


Figure 5. Precursor and product ion structures and m/z values for caffeine and $^{13}C_3$-caffeine.

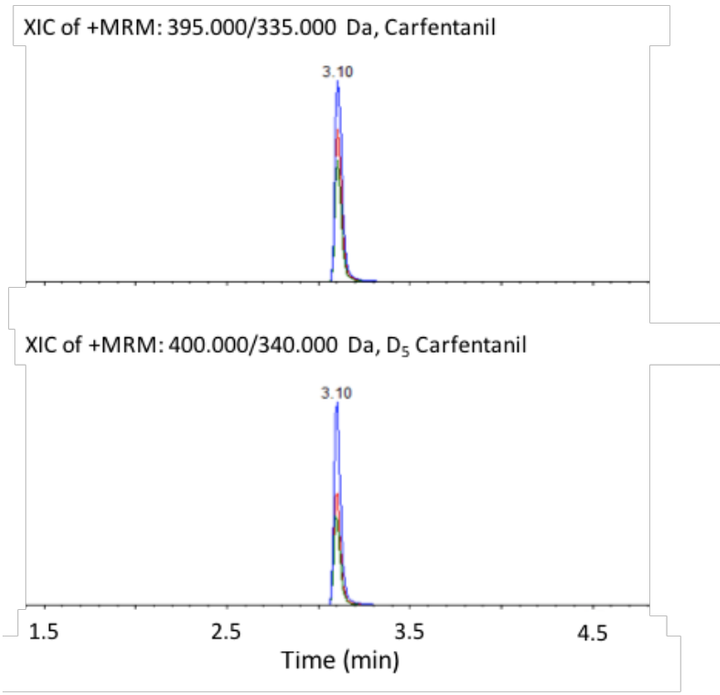


Figure 6. MRM chromatogram of carfentanil and D_5-carfentanil.

2.2.1.2 MRM Method Validation

MRM method quantification was performed with newly formulated calibration standards, quality control samples (QCs), and blank samples on a minimum of three separate days to measure inter-day method performance; calibration standard and QC samples were formulated from stock solutions and analyzed each day. Calibration and QC stock solutions were formulated by different analysts. QCs consisted of low, medium and high levels, bracketing the calibration curves, and were formulated by a different analyst. The internal standard, sufentanil, was spiked into the samples at a final concentration of 100 ng/mL prior to sample analysis.

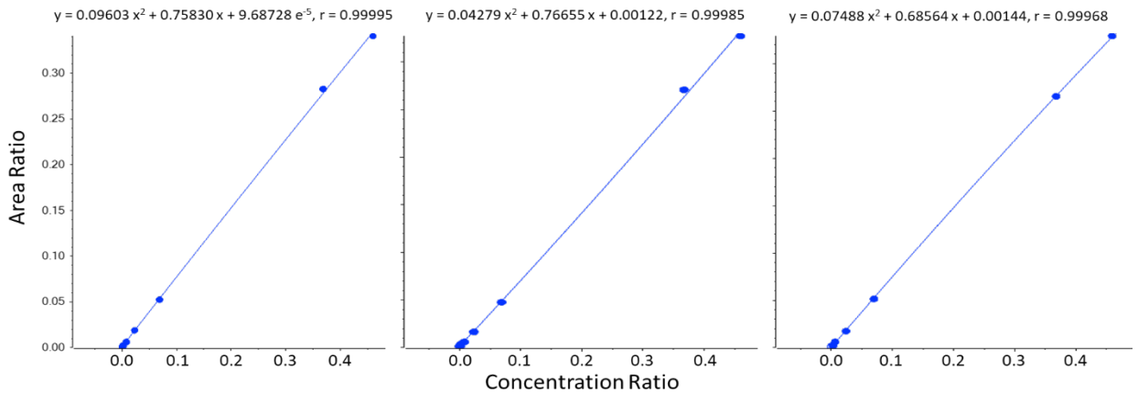
A minimum of three characteristic product ions for each compound were selected for MRM monitoring. From the infusion experiments, characteristic product ions for both compounds were identified along with their corresponding optimal CEs. Subsequently, a high standard for

each compound was analyzed using the LC gradient described above employing the identified optimal CEs. The most abundant characteristic fragment ion was selected as the quantitative transition; the second and third most abundant characteristic fragment ions were selected as qualifier transitions for MRM spectral matching to ensure method specificity.

To determine the limit of detection (LOD) for carfentanil and caffeine, a minimum of seven daily replicates were analyzed. All blanks, controls, and samples had to contain no peaks interfering with the target analyte or internal standard. At the method LOD concentration, a signal-to-noise criterion of ≥10:1 and a minimum of 10 points had to be identified across each m/z peak for quantifier and qualifier ions (sensitivity requirement) had to be met, and qualifier ion relative peak area ratios had to be within ±20% of those for a mid-level calibration curve standard (spectral matching specificity requirement).

To identify the method limit of quantification (LOQ) and dynamic range, quantification criteria included the general requirements used routinely by the U.S. FDA and the Federal Drug Testing Program that all measured replicate results were accurate within ±20% of the actual, precision across results (%CV) was ≤15%, the curve coefficient of determination or r^2 (measure of curve linearity) was ≥ 0.980 (correlation coefficient or r > 0.990), ion ratios of all qualifier ions were within ±20% of that for a mid-level standard, and blanks contained no target analyte or internal standard signal above baseline. Two carfentanil calibration curves meeting all analytical criteria were employed in order to accurately quantify over the broad concentration range encountered with the study samples. Carfentanil low and high calibration curve concentrations ranged from 230-4600 pg/mL and 46-920 ng/mL, respectively, each comprised of a minimum of six calibration points. A single calibration curve meeting analytical criteria was used to quantify caffeine over a concentration ranging from 12-19,800 ng/mL. Figure 4 illustrates carfentanil calibration curves for the two dynamic concentration ranges employed in the three-day method validation. The curves were best described mathematically with quadratic formulae, using 1/x weighting for the area ratio term.

A) Low curves



B) High curves

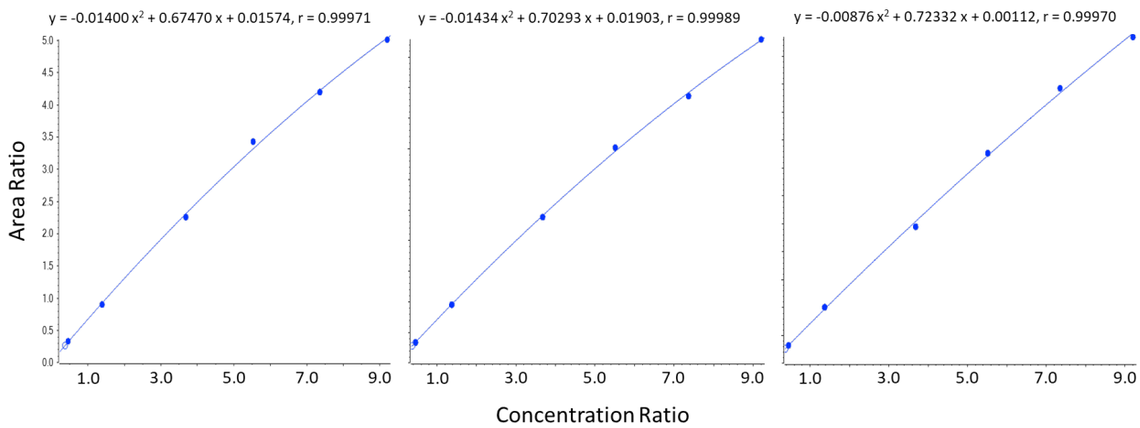


Figure 7. Daily carfentanil calibration curves; A) 230-4600 pg/mL (low curves) and B) 46-920 ng/mL (high curves).

To determine method intra- and inter-day precision and accuracy, a minimum of five replicate quality control samples at three concentrations bracketing the method dynamic concentration range were analyzed on three separate days. Intra- and inter-day results were

evaluated using LOD and LOQ requirements. Quality controls were formulated by a different analyst and evaluated against LOD and LOQ requirements. All calibrator and quality control results were required to meet the following accuracy and precision criteria: 1) all individual measured quantitative results had to quantify within ±20% of the expected concentration, 2) the mean measured concentration for each level had to quantify within ±15% of the expected concentration, and 3) the % coefficient of variation (CV) across replicate results had to be <15%.

2.3 RESULTS:

Optimal source parameters for the target analytes included the following; curtain gas (CUR) - 55; ionspray voltage (IS) - 5500; turbo gas temperature (TEM) - 700; gas one (GS1) - 60; gas 2 (GS2) - 60; interface heater (ihe) - ON; collision-assisted-dissociation (CAD) voltage - MED; entrance potential (EP) - 10; and collision cell exit potential (CXP) - 9. Table 1 lists the optimal MSMS voltages for precursor and product ions.

Table 2: Optimal MSMS Parameters

Transitions	Precursor $(M+H)^+$	Product Ion	Dwell (msec)	Declustering Potential (DP)	Collision Energy (CE)
Carfentanil	395	335	62	60	30
		113			43
		105			63
Caffeine	195	138	62	60	15
		110			20
		83			41
Sufentanil	387	238	62	30	23
		355			23
		111			52

Applying both intra- and inter-day linearity, accuracy, and precision acceptance criteria, carfentanil low and high calibration curve concentrations ranged from 230-4600 pg/mL and 46-920 ng/mL, respectively, each comprised of a minimum of six calibration points. A single

calibration curve ranging from 11.88-19,800 ng/mL met analytical criteria and was used to quantify caffeine. All curves were evaluated as quadratic expressions with 1/x weighting, and correlation coefficients for all calibration curves exceeded 0.995. Intra-day quantitative data listed in Tables 2-4 were evaluated using the curve regression equations; all calibration curve data met intra- and inter-day precision and accuracy requirements.

Table 3: Carfentanil Low Calibration Curve Accuracy

Expected Conc. (ng/mL)	Day 1		Day 2		Day 3	
	Mean Conc. (ng/mL)	Accuracy (%)	Mean Conc. (ng/mL)	Accuracy (%)	Mean Conc. (ng/mL)	Accuracy (%)
0.046	**0.09**	**193.0**	**0.00**	**0.00**	0.042	91.25
0.069	**0.20**	**292.4**	0.068	98.53	0.065	94.62
0.23	0.21	91.8	0.24	104.25	0.27	116.43
0.69	0.74	107.5	0.68	98.76	0.72	104.80
2.30	2.34	101.9	2.31	100.25	2.17	94.35
6.90	6.83	98.9	6.72	97.44	6.73	97.51
36.8	36.73	99.8	37.55	102.04	37.77	102.63
46.0	46.06	100.1	45.42	98.74	45.27	98.41

*A minimum of 7 replicates/calibration level/day were analyzed to determine the LOD and LOQ. Data in **red** did not meet accuracy requirements but did meet detection requirements. The valid carfentanil curve dynamic concentration range was 0.23-46 ng/mL.*

Table 4: Carfentanil High Calibration Curve Accuracy

Expected Conc. (ng/mL)	Day 1		Day 2		Day 3	
	Mean Conc. (ng/mL)	Accuracy (%)	Mean Conc. (ng/mL)	Accuracy (%)	Mean Conc. (ng/mL)	Accuracy (%)
46	46.9	101.93	46.2	100.39	45.6	99.21
138	135.1	97.87	137.5	99.60	141.3	102.39
368	359.5	97.69	364.5	99.04	354.3	96.29

522	573.0	103.8	563.8	102.1	559.7	101.40
736	731.9	99.45	724.9	98.5	750.9	102.02
920	913.3	99.28	923.3	100.4	908.1	98.71

A minimum of 7 replicates/calibration level/day were analyzed to determine the LOD and LOQ. All data met accuracy requirements.

Table 5: Caffeine Calibration Curve Accuracy

Expected Conc. (ng/mL)	Day 1		Day 2		Day 3	
	Mean Conc. (ng/mL)	Accuracy (%)	Mean Conc. (ng/mL)	Accuracy (%)	Mean Conc. (ng/mL)	Accuracy (%)
7.9	7.3	92.0	**3.3**	**41.9**	**11.4**	**143.5**
11.9	10.9	92.0	11.9	86.1	13.1	110.4
39.6	42.3	106.8	39.6	110.1	42.0	106.0
59.4	62.3	104.8	59.4	92.8	59.3	99.8
198	205.6	103.8	198.0	100.9	178.8	90.3
597	601	100.7	597.0	104.2	565.3	94.7
1584	1579	99.7	1584	105.7	1571	99.2
1980	2003	101.1	1980	102.5	1914	96.7
5940	5886	99.1	5940	97.5	6367	107.2
7920	7802	98.5	7920	101.8	7453	94.1
11880	12340	103.9	11880	100.9	11970	100.7
15840	15310	96.6	15840	94.5	16240	102.5
19800	20020	101.1	19800	103.0	19480	98.4

*A minimum of 7 replicates/calibration level/day were analyzed to determine the LOD and LOQ. Data in **red** did not meet accuracy requirements.*

In addition, all daily QC sample results met precision and accuracy criteria during the validation. Intra-day precision and accuracy data for carfentanil QC samples from the three days of analysis are listed in Table 5, and inter-day precision and accuracy data for carfentanil and caffeine QC samples over the three days are listed in Table 6.

Table 6: Intra-day Carfentanil QC Accuracy and Precision

Expected Conc. (ng/mL)	Day 1		Day 2		Day 3	
	Measured Conc. (ng/mL)	**Accuracy (%)**	**Measured Conc. (ng/mL)**	**Accuracy (%)**	**Measured Conc. (ng/mL)**	**Accuracy (%)**
1.54	1.65	107.1	1.52	98.77	1.59	103.07
	1.51	98.3	1.51	98.20	1.66	108.00
	1.45	94.3	1.77	115.17	1.57	101.88
	1.48	96.1	1.44	93.44	1.42	92.24
	1.39	90.2	1.45	93.94	1.38	89.90
	Mean Accuracy = 97.20		**Mean Accuracy = 99.90**		**Mean Accuracy = 99.02**	
	%CV = 6.46		**%CV = 8.88**		**%CV = 7.73**	
40.48	36.15	89.31	39.50	97.57	38.48	95.06
	36.27	89.60	39.68	98.03	39.92	98.63
	36.79	90.87	38.34	94.72	41.62	102.81
	36.04	89.03	40.41	99.82	38.59	95.32
	37.17	91.83	40.03	98.89	40.07	98.98
	Mean Accuracy = 90.13		**Mean Accuracy = 97.81**		**Mean Accuracy = 98.16**	
	%CV = 1.31		**%CV = 1.97**		**%CV = 3.23**	
625.6	601.8	96.20	639.54	102.23	611.90	97.81
	597.1	95.45	609.06	97.36	595.30	95.15
	601.6	96.16	654.88	104.68	NA	NA
	622.6	99.53	615.65	98.41	597.80	95.55
	606.4	96.94	634.43	101.41	591.40	94.53
	Mean Accuracy = 96.86		**Mean Accuracy = 100.82**		**Mean Accuracy = 95.76**	
	%CV = 1.64		**%CV = 6.46**		**%CV = 1.49**	

Table 7: Intra-day Caffeine QC Accuracy and Precision

Expected Conc. (ng/mL)	Day 1		Day 2		Day 3	
	Measured Conc. (ng/mL)	**Accuracy (%)**	**Measured Conc. (ng/mL)**	**Accuracy (%)**	**Measured Conc. (ng/mL)**	**Accuracy (%)**
132.78	117.4	88.38	152.4	114.79	126.6	95.31
	118.1	88.97	147.9	107.6	131.5	99.03
	116.7	87.92	156.7	118.03	129.3	97.39
	113.0	85.13	135.3	101.87	131.0	98.63
	113.2	87.23	135.7	102.2	123.0	92.62
	Mean Accuracy = 87.13		**Mean Accuracy = 108.9**		**Mean Accuracy = 96.6**	
	%CV = 2.08		**%CV = 6.72**		**%CV = 2.75**	
1742	1522	87.37	1895	108.77	1615	92.71
	1563	89.72	1867	107.12	1655	94.99
	1536	88.16	1806	103.62	1751	100.49
	1542	88.51	1891	108.51	1638	94.00
	1623	93.17	1850	106.18	1766	101.36
	Mean Accuracy = 89.39		**Mean Accuracy = 106.84**		**Mean Accuracy = 96.71**	
	%CV = 2.55		**%CV = 1.95**		**%CV = 4.08**	
13464	12660	94.02	13711	101.84	14950	111.02
	12670	94.10	12867	95.57	13590	100.94
	12720	94.47	13891	103.17	NA	NA
	12670	94.13	13475	100.08	13200	98.06
	12260	91.08	13375	99.34	13780	102.34
	Mean Accuracy = 93.56		**Mean Accuracy = 100.00**		**Mean Accuracy = 103.09**	
	%CV = 1.49		**%CV = 2.89**		**%CV = 5.41**	

Table 8: Inter-day Carfentanil and Caffeine QC Accuracy and Precision

Analyte	Expected Conc. (ng/mL)	Mean Conc. (ng/mL)	Accuracy (%)	%CV
Carfentanil	1.54	1.52 ± 0.11	98.7	7.25
	40.48	38.74 ± 1.82	95.7	4.71
	625.60	628.10 ± 61.86	100.4	9.85
Caffeine	132.78	129.50 ± 13.50	97.5	10.40
	1742.0	1707.0 ± 130.0	98.0	7.63
	13464.0	13663.0 ± 12.1	101.5	12.12

2.4 Discussion

A relatively sensitive and rapid quantitative LC/hybrid triple quadrupole-linear ion trap MS/MS method was developed and validated for the extremely potent opioid, carfentanil, in a live human epidermal model previously designed and employed to perform pharmacokinetic measures following cutaneous drug exposure. The method was also validated for caffeine which was, in turn, used to validate the epidermal model for EtOH-containing solvents by verifying epidermal viability throughout the experiments. Another fentanyl analog, sufentanil, was employed as the internal standard for both target analytes. The method met rigorous sensitivity, specificity, accuracy and reproducibility criteria required for quantitative methods used in toxicology, drug monitoring, pharmacokinetics, pharmacodynamics, and forensics. These data demonstrated the relative high sensitivity, accuracy and reproducibility capabilities of triple quadrupole instruments for quantitative interrogation of samples for drugs and small molecules and illustrate why triple quadrupole instruments are considered the "gold standard" for small molecule quantification in these disciplines.

The mass spectrometer used was a nominal mass accuracy/resolution hybrid quadrupole-linear ion trap instrument. Not surprisingly, a single calibration curve proved to be too

inaccurate at the lower and upper concentration ranges over the extended concentration range needed for carfentanil quantification in the experimental samples. Therefore, "low" and "high" concentration curves were validated and employed. The curves, however, did each bracket at least four orders of magnitude concentration providing a very wide method dynamic range for the study. The caffeine calibration curve proved to be accurate and precise over the more than four orders of magnitude concentration range needed to verify epidermal cell viability during the exposures. Method accuracy and precision for carfentanil and caffeine met all validation acceptance criteria.

Counterintuitively, as reported previously[19], alcohol modifiers appeared to reduce permeation rates and peak concentrations in the RhE culture medium prolonging permeation lag times. The authors postulated that this was due to slowed partitioning from the hand sanitizers, due to reduced solvent polarity as compared to water alone. The results appeared to indicate that the concern over increased risk for law enforcement and emergency personnel using these hand sanitizers in environments where they may be at risk from percutaneous carfentanil exposure may be unwarranted. These conclusions appeared to support previous results[20].

Bibliography

1. Jodie V. Johnson and Richard A. Yost. Tandem Mass Spectrometry for Trace Analysis. *Anal. Chem.,* 1985, 57, 7, pp. 758A-768A, https://doi.org/10.1021/ac00284a718.
2. D.A. Volpe, G.A.M. Tobin, R.D. Mellon. Uniform Assessment and Ranking of Opioid Mu Receptor Binding Constants for Selected Opioid Drugs. *Regul. Toxicol. Pharmacol.,* 2011, 59(3), pp. 385-390, doi:10.1016/j.yrtph.2010.12.007.
3. Y. Higashikawa, S. Suzuki. Studies On 1-(2-phenethyl)-4-(N-propionylanilino)piperidine (fentanyl) and Its Related Compounds. VI. Structure-Analgesic Activity Relationship for Fentanyl, Methyl-Substituted Fentanyls and Other Analogues. *Forensic Toxicol.,* 2008, 26(1), pp. 1-5, doi:10.1007/s11419-007-0039-1.

4. Jessica L. S. Leen, David N. Juurlink. Carfentanil: A Narrative Review of Its Pharmacology and Public Health Concerns. *Can. J. Anesth.,* 2019, 66, pp. 414–421.
5. John F. Casale, Jennifer R. Mallette, Elizabeth M. Guest. Analysis of Illicit Carfentanil: Emergence of the Death Dragon. *Forensic Chemistry*, March 2017, Volume 3, pp. 74-80.
6. J.H. Manthei, D.H. Heitkamp, K.P. Cameron, D.M. Bona, J.L. Hill. In: Center, E.C.B. (Ed.), Intravenous Toxicity of Selected EA Compounds in Male Institute of Cancer Research (ICR) Mice Tested During FY90-91. Technical Report ECBC-TR-242, 2002, Edgewood Research, Development and Engineering Center, Aberdeen Proving Ground, MD.
7. Jeffrey L. Langston, Mark C. Moffett, Jennifer R. Makar, Bradley M. Burgan, and Todd M. Myers. Carfentanil Toxicity in the African Green Monkey: Therapeutic Efficacy of Naloxone. *Toxicol. Lett.*, 1 June 2020, 325, pp. 34-42.
8. James R. Riches, Robert W. Read, Robin M. Black, Nicholas J. Cooper and Christopher M. Timperley. Analysis of Clothing and Urine from Moscow Theatre Siege Casualties Reveals Carfentanil and Remifentanil Use. *Journal of Analytical Toxicology*, Nov 2012, 36(9), pp. 647–656, doi:10.1093/jat/bks078. PMID 23002178.
9. P.M. Wax, C.E. Becker, S.C. Curry. Unexpected "Gas" Casualties in Moscow: a Medical Toxicology Perspective. *Ann. Emerg. Med.,* 2003, 41, pp. 700–705.
10. CDC, 2018. Rising Numbers of Deaths Involving Fentanyl and Fentanyl Analogs, Including Carfentanil, and Increased Usage and Mixing with Non-Opioids. Centers for Disease Control. https://emergency.cdc.gov/han/han00413.asp.
11. Chris Delcher, Yanning Wey, Russell S. Vega, John Halpen, R. Matthew Gladden, Julie K. O'Donnell, Jessica A. Hvozdovich, Bruce A. Goldberger. Carfentanil Outbreak – Florida 2016-2017, *Morb. and Mort. Wkly. Report*, Feb 7, 2020, Vol 69, No. 5, pp. 125-129.
12. A. McDaniels, E. Deputy Butler. Two EMS Providers Treated for Overdose Symptoms Responding to Call. *Baltimore Sun*, May 23, 2017, https://www.baltimoresun.com/health/bs-md-haRFord-opioid-exposure-20170523-story.html, Accessed January 21, 2020.

13. J. Queally. LAPD Officers Hospitalized After Exposure to Fentanyl in North Hollywood. *Los Angeles Times*, 6 September 2019. https://www.latimes.com/california/story/2019-09-06/lapd-officers-hospitalized-after-exposure-to-fentanyl-in-north-hollywood.
14. A.V. George, J.J. Lu, M.V. Pisano, J. Metz, T.B. Erickson. Carfentanil-An Ultra-Potent Opioid. *Am. J. Emerg. Med.,* 2010, 28, pp. 530–532.
15. Fentanyl: Emergency Responders at Risk, NIOSH, CDC, 11 Feb 2020, https://www.cdc.gov/niosh/topics/fentanyl/risk.html.
16. DEA, 2017. Drug Enforcement Agency. (Fentanyl. A Briefing Guide for First Responders).
17. NIEHS, 2018. Prevention of Occupational Exposure to Fentanyl and Other Opioids. National Institute of Environmental Health.
18. Emily May Lent, Kathleen J. Maistros, Jonathan M. Oyler. In Vitro Dermal Absorption of Carfentanil. *Toxicology in Vitro*, 2020, 62, pp. 1-6.
19. J.B. Cole, L.S. Nelson. Controversies and Carfentanil: We have Much to Learn About the Present State of Opioid Poisoning. *Am. J. Emerg. Med.,* 2017, 35, pp. 1743–1745.

Chapter 3

The Rapid Denaturing Organic Digestion Method for Targeted Protein Identification and Characterization

Jonathan M. Oyler[1,2,3*], Bao Q. Tran[4], and David P.A. Kilgour[3]

[1]US Army Medical Research Institute of Chemical Defense, [2]The US Army Inspector General Agency, Technical Inspections Division (SAIG-TI), [3]Nottingham Trent University, [4]US Army 20th CBRNE Analytical and Remediation Activity

[1] 8350 Ricketts Point Road, Aberdeen Proving Ground, MD 21010-5400

[2] 2530 Crystal Drive, Suite 12039, Arlington, VA 22202-3912

[3] Clifton Campus, College Drive, Nottingham, NG11 8NS, UK

[4] Building E 1947, 5183 Blackhawk Road, Aberdeen Proving Ground, MD 21010-5424

[*]Corresponding Author: jonathan.m.oyler.civ@mail.mil

DOI 10.1021/acs.analchem.0c04143

3.0 Acknowledgements

The authors acknowledge Raymond Sullivan, Ph.D., Biologist at the US Army Combat Capabilities Development Command Chemical Biological Center, for conducting bench work associated with this manuscript and for contributions to writing of the manuscript.

This research was made possible by funding provided by the Defense Threat Reduction Agency-Joint Science and Technology Office for Chemical and Biological Defense. Conclusions and opinions presented here are those of the authors and are not the official policy of the U.S. Army Medical Research Institute for Chemical Defense, US Army 20th CBRNE

Analytical and Remediation Activity, Joint Program Executive Office for Chemical Biological Radiological Nuclear Defense or the U.S. Department of Defense. Information in this report is cleared for public release and distribution is unlimited.

3.1 Abstract

Bottom-up mass spectrometry-based protein analysis methods employing protease digestion are routinely used to identify and characterize proteins with high specificity and sensitivity. Method performance is generally measured by sequence coverage capability and the total number of characteristic peptides identified, when compared to predicted databases. Limitations to commonly used solvent-based digestion methods currently employed include long digest times (18-24 h or more), leading to protease autolysis which also precludes automation, decreases sensitivity, and increases both intra- and inter-day performance variability. This report describes the development and validation of a simple, 5 min tryptic Denaturing Organic Digestion (DOD) method for use with tandem mass spectrometry in bottom-up protein identification and characterization. It has been evaluated across select protein toxins and diagnostic clinical protein targets substantially improving digestion performance when compared to other solution- based and enzyme immobilized methods. The method was compared to two currently used bottom-up methods, the 24 h Filter Aided Sample Prep (FASP) and Flash Digest (1 h and 4h) methods. Single proteins used to compare the methods included ricin light chain, ricin heavy chain, ricin holotoxin, serotype A Clostridium botulinum toxin, Staphylococcus enterotoxin B, ribonuclease A, and thyroglobulin. In tests across the target proteins investigated, the 5 min DOD digestion method resulted in sequence coverages ranging from 55-100%, with relative high reproducibility and precision; results were better than or equal to FASP method results and were greatly enhanced when compared to Flash method results. Importantly, DOD method intra- and inter-day precision was much improved as compared to results for both FASP and Flash digestions.

These data indicated that the DOD method, when compared to the FASP and Flash Digest methods, dramatically reduced digestion time while maintaining or improving the ability to detect and characterize targeted proteins, and reduced analytical variability for tryptic digestion resulting in markedly faster and more precise analyses.

3.2 Introduction

Protein identification and quantification is a challenging task. The most commonly used mass spectrometry-based protein analysis approach, bottom-up LC-tandem MS protein analysis, does not have a requirement for high resolution mass spectrometry, usually requires a much shorter analysis time, and has historically been much more sensitive than top-down approaches. Because the mass range of many of the mass spectrometers used is limited, bottom-up methods initially employ proteolytic digestion to produce characteristic peptides, unique to the protein targeted, and amenable to analysis by these instruments. To identify the peptides, their masses and their tandem mass spectra are then compared to sequence database or annotated peptide spectral library predictions. The peptide data is then compiled to predict the sequence of the parent protein. These methods, however, have generally been limited by an inability to provide sufficient sequence coverage to fully characterize targeted proteins and are limited in their ability to fully characterize PTM maps and point mutations located in digestion fragments that are not easily ionized. In addition, bottom-up methods have had relatively poor reproducibility, often attributed to the tendency for protease autolysis during lengthy digestion times, especially at high relative protease concentrations, and inconsistent target analyte recoveries inherent with lengthy multi-step processes. To improve sequence coverage, multi-protease strategies have been devised which have proven to dramatically improve results[1,2,3], but the additional processes can reduce sensitivity, make results more variable due to compounded inherent recovery losses, and dramatically increase analytical time by increasing sample preparation time from hours to days.

The most common protease used in targeted protein analysis is trypsin which cleaves the amide bond on the C-terminal side of lysine and arginine residues, except when they are

immediately followed by a proline. Many previously published benchtop tryptic digestion methods require 18-24 h digestion times[4]. Trypsin autolysis, directly correlated with trypsin concentration and digestion time, is the primary factor contributing to inter- and intra-day variability[5,6].

Recent descriptions of methods employing on-column and bead surface immobilization of protease[7] have dramatically reduced digestion times as well, improved method reproducibility and almost eliminated autolysis. These methods, however, often provide limited sequence coverage as compared to in-solution methods, likely due to steric hindrance issues. Another limitation to column-immobilized protease digestion is the inability to multiplex sample preparation.

The effects of organic modifiers in solvent systems or elevated temperatures have been previously used to speed protein digestion, improve digestion reproducibility, provide increased sequence coverage for protein characterization, and to identify conditions under which the protein folding could be characterized. Data demonstrated that trypsin proteolytic activity was maintained across a relatively wide temperature range in aqueous environments and in a variety of organically modified solvent systems, and that organic solvents denatured proteins exposing otherwise protected domains. In addition, work had been conducted using either a combination of low organic solvent concentration and temperatures up to 40°C or higher organic solvent concentrations alone. In 1987, Fink et al.[8] investigated the denaturation or unfolding of carboxymethylated, disulfide- reduced ribonuclease A in methanol-water cryosolvent systems. The goal was to identify cryosolvent conditions under which the protein was stabilized, in multiple partially folded intermediate forms in order to examine the thermodynamics, kinetics, and structural aspects of protein folding mechanisms. They monitored the degree of exposure or burial of tyrosine residues to measure the degree of protein unfolding. The results demonstrated that there was a positive correlation between increased unfolding and both methanol (MeOH) concentration and temperature and that unfolding increased even more with increases in both

MeOH content and temperature. In the course of their studies, they discovered that with MeOH as co-solvent, protein unfolding was 100% reversible up to a 60% MeOH concentration (v:v) over a temperature range of -40 to 70 °C. Ultimately, the authors selected a 50% (v:v) concentration with which to conduct further studies.

In 1988, Welinder, et al.[9], studied the activity and specificity of eight proteases when digesting reduced and carboxymethylated ribonuclease using organic solvent-modified systems. The authors evaluated MeOH, ethanol, isopropanol, and acetonitrile (AcN) as organic modifiers, and digests were analyzed by amino acid sequencing using commercially available sequencers; results were compared to aqueous digestion. Experiments were conducted to determine if digestion could be performed directly in the chromatographic mobile phases in which samples were collected following protein isolation. The authors stated that at least partial trypsin activity was maintained in solvents containing up to 40% organic (v:v), but digestion efficiency was only evaluated at 22 and 37 ℃ using 2 and 18 h digestion times. The authors concluded that trypsin activity was reduced with organic solvent modification; activity was most reduced using AcN, and at 37 ℃, proteolysis was even more reduced.

In 2000, Park et al.[10] showed that thermal protein denaturation prior to digestion generally increased the digestion efficiency of soluble proteins otherwise resistant to proteolytic digestion in aqueous solvents, and similar results were demonstrated by the same group with mixtures of proteins[11]. Though it improved the capability to digest resistant soluble proteins, thermal denaturation alone did not result in an improved ability to digest membrane proteins.

Subsequently, Russell, et al.[12], investigated the use of organic-modified solvent systems with proteolytic digestion for use in high through-put bottom-up peptide mapping. Their goal was to reduce digestion time in order to more rapidly characterize proteins. Investigated proteins included rabbit phosphorylase, bovine serum albumin, bovine hemoglobin, rabbit aldolase, chicken ovalbumin, rabbit carbonic anhydrase, horse myoglobin, horse cytochrome C, chicken lysozyme and bovine ubiquitin. With MeOH, acetone, 2-propanol, or AcN addition, digestion

rate dramatically increased, the sequence coverage for the proteins also increased, and proteins resistant to proteolysis under aqueous conditions could be digested to yield sequence data. Though digestion times were dramatically improved with no need for detergents or other chemical denaturants, the increases in digestion efficiency observed across the eleven proteins investigated were generally relatively modest.

Proc, et al.[13], evaluated the effects of chaotropic agents, surfactants and two organic solvents on trypsin digestion efficiency. They used a method employing isotopically labelled characteristic peptides as internal standards in order to quantitatively measure the "absolute amounts" of peptides produced. Forty-five clinically important plasma proteins were digested. Results were compared to those from a common denaturing digestion method using sodium deoxycholate. Solvents evaluated were MeOH and AcN. Digestion was conducted at 37°C using 40% and 20% AcN and MeOH (v:v), respectively, with digestion times ranging from 0.5 to 23 h. The organic solvent was added to the digestion solution following reduction and alkylation and just prior to the addition of trypsin, to "prevent protein precipitation". The proteins were divided into three groups based on the results: "rapidly digested", "moderately digested" and "resistant to digestion". For most proteins, digestion "signal" peaked at 4 h and plateaued, though for a few proteins, digestion continued throughout the 23 h incubation without plateau. They also observed that after 4 h, the digestion signal for some proteins decreased. With AcN as solvent, though digestion did occur, it was significantly reduced compared to the sodium deoxycholate denaturation method; the authors said of the AcN results, "in our experiments, a significant reduction in digestion efficiency was observed for all 45 analytes in the presence of 40% v/v acetonitrile, even though no protein precipitation was observed. Moderate digestion of some proteins, such as haptoglobin however, was observed when using AcN denaturation."; MeOH results were similar.

More recently, methods have been designed and tested to shorten optimal digestion times. In 2017, Zhang, et.al., validated a tryptic digestion method for five target proteins employing a

20-min digestion preceded by heat denaturation, without the use of chemical denaturants or reduction/alkylation steps[14]. It was validated for the detection of five plasma proteins in diluted 1 µl plasma samples. It employed trypsin pretreated with tosyl phenylalanyl chloromethyl ketone (TPCK) and required several additional pre-digestion steps. In 2023, Shuford, et.al., also explored the use of TPCK-treated trypsin, TPCK-treated trypsin/Lys C, and dimethylated trypsin with several strategies in a 96-well plate format to shorten and simplify tryptic digestion for a single target protein, thyroglobulin, in clinical serum samples[15]. Importantly, digestions were performed only at 35 or 37°C and the formation of characteristic thyroglobulin peptides was monitored to measure performance. Four different digestion time course experiments were conducted. The four digestion variables assessed in each experiment were digestion additive, trypsin type, incubator/reactor, and trypsin concentration. All digestions were performed using TPCK-treated bovine trypsin at an enzyme to protein ratio of 1:200 (w/w)." Additives tested included, "urea, guanidine, thiourea, deoxycholate, CHAPS, trifluoroethanol, ACN, MeOH, and 1-propanol. They observed that maximum recoveries for four of the eight characteristic thyroglobulin peptides occurred with digestions conducted for at least 4 h using any of the additives while the other additive conditions required much longer digestions. Interestingly, the most efficient digestions were generally achieved with the lower of two additive concentrations. The effect of trypsin type was also investigated. Samples were digested with five different trypsins employing five different digestion times ranging from 0.5 to 18 h on a Thermomixer C and employing an enzyme-to-protein ratio of 1:200 (w/w). Several 'sequencing' grade trypsins as well as a trypsin/LysC mixture were tested to determine their effect on digestion performance. Trypsin grade did not appear to improve digestion performance and required up to 18 h for optimal performance. The proteomics-grade dimethylated porcine trypsin tested did decrease the digestion time slightly, but a trypsin/Lys C enzyme combination achieved the most rapid digestion time course with optimal digestion occurring at 2 h for some peptides. Different enzyme reactor systems were also evaluated including the Thermomixer C heat block, the two

microwave systems tested, the Rapid Enzyme Digestion System (REDS) and microwave acid digestion system (MARS). The pressurized system shortened optimal digestion time to 15-75 min across the peptides assayed. The authors stated, however, "further optimization of the enzyme reactors was not attempted as similar, rapid digestion speeds were achievable with conventional incubation by simply increasing the amount of trypsin. In fact, increasing the concentration of trypsin 10-fold with conventional incubation was feasible due to the relatively low cost of TPCK-treated bovine trypsin and was more favorable for a high-throughput clinical lab environment compared to using specialized reactors and associated consumables. An important finding was also noted; use of "sequencing" or "MS" grade was not required for optimal digestion results.

We investigated the hitherto unexplored combined use of high solvent concentrations (consisting of 40-100% AcN (v:v)), along with "high" temperatures (ranging from 40-80°C) to digest select single protein toxins and diagnostically important proteins and cocktails of the same for use in bottom up protein analysis. Digest products were analyzed by liquid chromatography/tandem mass spectrometry as part of an effort to develop a rapid diagnostic method for targeted proteins. Proteins investigated included the toxins ricin, botulinum neurotoxin serotype A (BotNT-A) and staphylococcus enterotoxin B (SEB).

Ricin holotoxin (RH), a lectin found in castor beans, was initially investigated as a "proof of concept" experiment. RH is a protein heterodimer composed of an A and B chain (RAC and RBC, respectively) linked through a disulfide bond. It is a ribosome-inactivating protein that disrupts cellular protein synthesis[14,15]. It has significant clinical toxicity, is highly stable, and there are no effective therapies. BotNT-A is one in a family of toxins produced by gram positive anaerobic bacteria in the genus Clostridium. BotNTs are the most toxic substances known to man and, like ricin, potential weapons of terror and mass destruction[16,17,18]; BotNT-A is one of four clinically significant serotypes. Like RH, BotNTs are heterodimers consisting of heavy (HC) and light chains (LC) linked by disulfide bridges. SEB is an enterotoxin produced by the

gram-positive bacterium Staphylococcus aureus. It induces severe diarrhea and nausea, is a superantigen that can lead to severe gastroenteritis[19], and is a cause of toxic shock syndrome. It is also quite stable, even capable of withstanding boiling in aqueous solution at 100 ºC for several minutes. Due to the ease with which it can be produced, its stability, and its capability for causing human morbidity and mortality, it is also considered a potential weapon of terror and mass destruction[20]. For these reasons, it was important to develop rapid, sensitive assays for these protein toxins.

After the initial successful tests with toxins, the high solvent/high temperature digestion method was used to digest two clinically important diagnostic proteins, thyroglobulin (TG) and ribonuclease A (RNAse A) for which routinely used clinical digestion methods have lacked sensitivity, specificity, and reproducibility largely due to digestion method inefficiency and lack of reproducibility. TG, a ~660 kDa glycoprotein homodimer, is produced mainly by the thyroid, is involved in the synthesis of thyroxine and triiodothyronine, and acts as a storage site for thyroid hormone and iodine[21]. Detection and quantification of TG in the clinical laboratory can be problematic especially for patients that express anti-TG antibodies (ATAs) which can inteRFere with the immunoaffinity methods commonly used[22]. Since successful long- term patient monitoring using these methods can be difficult, the American Thyroid Association has emphasized the need for development of sensitive TG methods, not affected by ATAs, in clinical samples. RNAse A, a relatively small protein (124 residues, ~13.7 kDa), is a very stable endoribonuclease typically prepared in the lab for analysis by boiling for 30 min to remove deoxyribonuclease[23]. Blood levels have been used in combination with other biomarkers to diagnose a variety of cancers and infectious diseases. Current, largely activity-based indirect methods, as with TG, are relatively long and have lacked the specificity, sensitivity and reproducibility needed.

We describe the development and validation of the simple, 5-minute rapid proteolytic Denaturing Organic Digestion method (DOD), using trypsin, that was compared briefly to a

100% AcN solvent method and more rigorously to the Filter Aided Sample Prep (FASP) and Flash methods, two commonly used tryptic digest methods used in MS-based protein analysis. (The known sequences and tryptic cleavage sites for the proteins studied are illustrated in the supplemental information published with the original manuscript.) FASP is a tryptic digest method developed to generate peptides from crude cell lysates for subsequent MS analyses. It employs a detergent to disrupt cells exposing cellular proteins to proteolytic digestion. Critical method steps occur in a filter using 8 M urea, hence the name. Method advantages include that genomic material is removed from samples prior to digestion, samples are not subjected to precipitation, the method can be used successfully with samples containing strong detergents at relatively high concentrations, samples containing a relatively wide range of protein concentrations can be prepared using a single filter device, and targeted protein concentrations are kept relatively high[24,25,26]. The Flash Digest method employs immunoaffinity-immobilized protease. Advantages include very rapid digestion times resulting from a very high enzyme-to-target protein concentration ratio, steric hindrance of protease autolysis resulting in good reproducibility, and the ability to reuse enzyme[27,28]. The optimized DOD method employed a digestion solvent composed of a high relative amount of AcN (60% v:v) and a high relative (60 °C) incubation temperature. In this study, method digestion efficacies were compared for select toxins and two clinically previously difficult to digest diagnostic proteins by measuring and comparing the sequence coverage and number of characteristic peptides identified using each method. Samples containing single proteins or protein cocktails were digested and analyzed and results compared.

3.3 Methods

3.3.0 FASP 24 h Digestion Method

Samples were processed by the FASP digestion method according to the protocol described by Wizniewski et al.[29]. Briefly, proteins were introduced onto a Microcon YM-10 (Millipore, Cat. No. number 42407) filter and reduced with 20 mM DTT (prepared fresh daily)

for 1 h at 60 ºC followed by alkylation with 55 mM iodoacetamide for 45 min at room temperature. Samples were digested overnight at 37 ºC using 0.02 µg/µL trypsin prepared in 50 mM ammonium bicarbonate. Post-digestion, the samples were vacuum centrifuged to dryness for 40 min and stored at -20 ºC until analysis. Prior to nanoLC/MSMS analysis, samples were suspended in 100 µL of buffer consisting of 95% LC/MS-grade water/5% acetonitrile with 0.5% formic acid.

3.3.1 Flash Digestion Method

Samples were digested according to the method described by Griffiths, et al.[30]. To each Flash Digest tube was added a 150 µL aliquot of Flash Digest buffer and 50 µL of 50 mM triethyl ammonium bicarbonate containing 20 µg of each target protein. Samples were placed in an Eppendorf ThermoMixer C equipped with a 96-well PCR plate shaker with a heated lid (pre-heated to 70^{o}C) and digested for 1 h constantly stirred at 1400 RPM. Digestion was terminated by adding 400 µL of 0.1 % TFA with agitation. Samples were allowed to stand for 40 minutes at 37 ^{o}C followed by filtration through 3kD and 10kD membranes to remove particulates. Each supernatant was transferred to an autosampler vial and diluted to the desired target concentration with LC/MSMS buffer prior to nanoLC/MSMS analysis.

3.3.2 Denaturing Organic Digestion (DOD) and 100% AcN Methods

Sample digestion using the optimized DOD and 100% (v:v) AcN methods was performed as follows. A master mix was formulated consisting of 10-70 µg of targeted protein, 22 µL of 1.0 M DTT (final concentration 20 mM), 660 µL of AcN and 418 µL LC/MS grade water (60% AcN ((v:v)) for the DOD method or 1078 µL of 100% (100% AcN method) for a total sample volume of 1100 µL. A 1 µg trypsin/µL LC/MS-grade water solution was prepared fresh daily. Aliquots (50 µL or 450-3180 ng of protein) of the master mix were transferred to separate 1.5 mL Eppendorf snap-cap tubes. Digestion was initiated by adding trypsin at a 1:3 molar ratio to target protein(s). Vial caps were affixed, and samples were incubated at 60 ºC at 700 rpm in an Eppendorf ThermoMixer C equipped with a heated lid for five minutes (During method

optimization, digestion times and temperatures were varied from 5-60 min and 40-80 °C, respectively. Our first digestion attempts were conducted using an Eppendorf ThermoMixer C equipped *without* a heated lid and failed; results improved markedly after procuring a heated lid.). Digestion was terminated by adding 1 µL of 50% TFA to each tube. Samples were vacuum-dried for 30 minutes in a Speedvac under medium heat and stored at -20 ºC until analysis. Prior to nanoLC/MSMS analysis, samples were suspended in 50 µL LC/MSMS buffer consisting of 95% LC/MS-grade water/5% AcN with 0.5% formic acid.

3.4 Results

3.4.0 Ricin

Since previous studies had shown that protease digestion efficiency for peptides and proteins with known potential cleavage sites appears to be limited by steric hindrance resulting from protein tertiary structure[31], we reasoned that a simple treatment of RH with AcN, an organic solvent routinely employed to crudely "crash" proteins out of aqueous solutions through tertiary structure disruption, might expose potential cleavage sites providing trypsin more access thereby improving bottom-up sequence coverage. Following DTT reduction, RH was digested using the 24 h FASP, 1 and 4 h Flash, and 100% AcN method, digesting for 1 and 4 h. Samples were digested in quintuplicate. The mean number of characteristic peptides identified for the FASP, 1 and 4 h Flash and 1 and 4 h 100% AcN methods were 29, 7, 6, 12 and 9, respectively, and mean sequence coverages for the methods were 72, 11, 9, 14 and 13%, respectively.

Subsequently, RAC and RBC were digested using solvents containing 50-80% AcN (v:v) incubated for 15-60 min at 40-70 ºC; quadruplicate biological replicates were digested under each condition. Both sequence coverage and number of unique peptides identified significantly improved when compared to results from previous 24 h FASP and 1 and 4 h Flash experiments. Optimal results were observed when employing 60-70% AcN (v:v) at 60-70ºC (Figure 1 illustrates 15 min digestion results). The data also suggested that AcN content was the principal

variable affecting digestion across this temperature range and that digestion could be conducted in as little as 15 min without affecting performance.

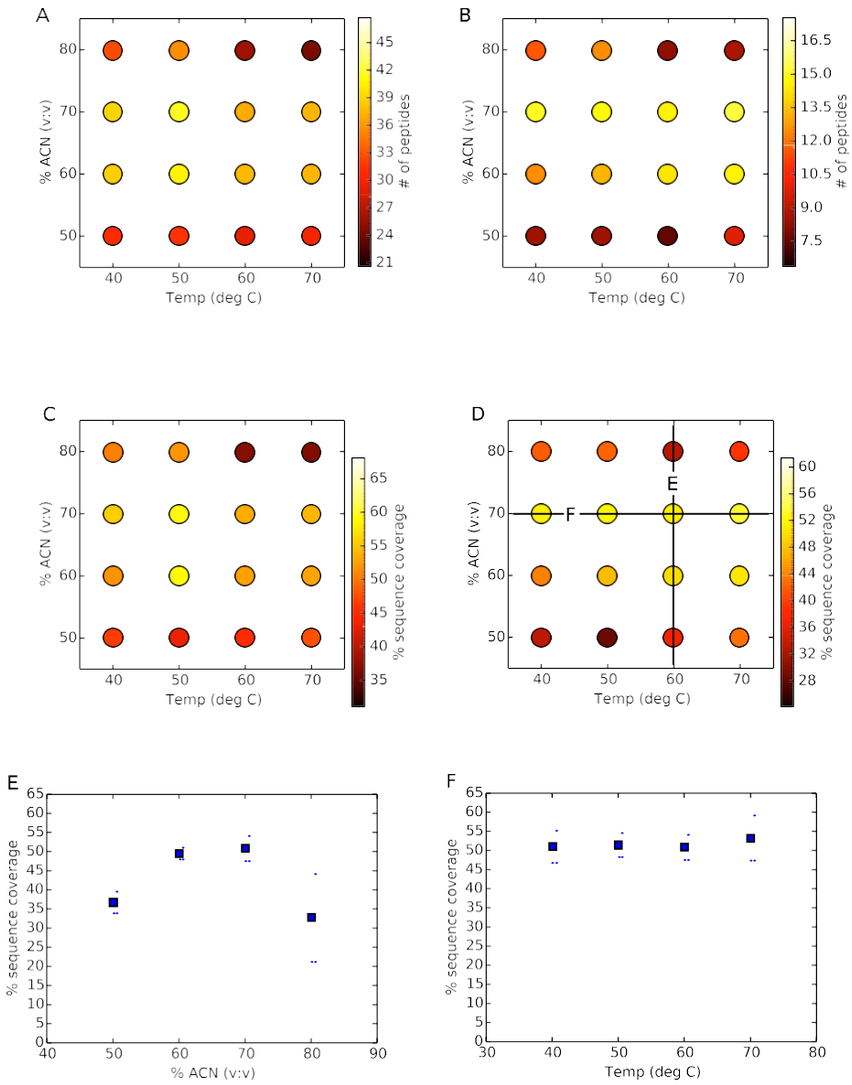


Fig. 1: Number of characteristic peptides identified and sequence coverages from 15 min digestion of RAC and RBC varying temperature and AcN concentration; proteins were not reduced or alkylated in this experiment. Panels A and B illustrate AcN-assisted 15 min digestion of RAC and RBC, with respect to number of unique peptides identified. Panels C and D illustrate % sequence coverage results for RAC and RBC, respectively. Panels E and F show transects through (D) to illustrate the shape of the surface as a function of AcN concentration (E) and temperature (F). Error bars illustrate the standard deviation from 4 biological replicates.

To confirm that DTT reduction could be performed in 60% AcN, we alkylated, and digested RH, RAC and RBC with and without DTT reduction in the solvent. Results from four replicates per condition are illustrated in Figure 2. Across proteins, substantial improvements in sequence coverage and the number of characteristic peptides identified were observed with disulfide reduction. Not surprisingly, the number of characteristic peptides improved mostly with RBC, in which four of the five holotoxin disulfide bonds lie. Additionally, across the proteins, precision, as determined by the calculating the RSD across replicates, improved with disulfide reduction, especially for sequence coverage.

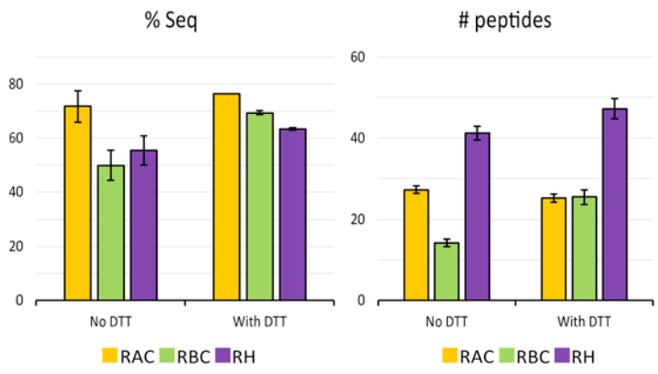


Fig. 2. DOD method sequence coverages and number of characteristic peptides identified for four replicate digestions each of Ricin-A chain, Ricin-B chain, and ricin holotoxin with and without 0.5 M DTT. Error bars represent the calculated RSD for the replicate data.

Subsequently, separate RAC, RBC, and RH samples were reduced and digested in quintuplicate using the DOD method employing 15-, 30-, and 60-min digestions. Data were compared to that from 24 h FASP digests collected previously, as illustrated in Figure 3 - results for RAC and RBC from 60 min DOD digestion were comparable and are not shown. The DOD method, for all digestion times, showed consistently increased numbers of characteristic peptides detected and much better precision across replicates, as compared to those attained with the FASP method.

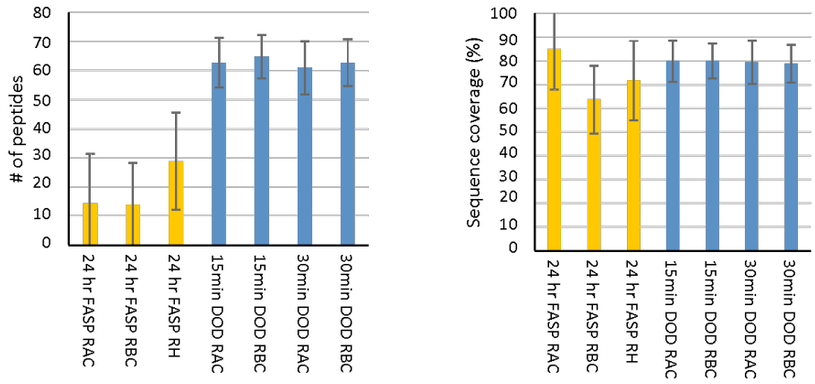


Fig. 3: Mean number of peptides identified (left) and sequence coverage (right) for RAC, RBC and RH, respectively, using the 24 h FASP and 15- and 30-min DOD digestion methods. Results are mean data across five biological replicates with variability expressed as %CV (error bars).

The average number of characteristic peptides identified and average % sequence coverages for the 1 and 4 h Flash, 24 h FASP, 1 and 4 h Flash, and 1 and 4 h 100% AcN methods collected from the initial experiment and the results for the 15 min DOD method are illustrated in Figure 4. Specific sequence coverages for RH using the 24 h FASP and the 15 min DOD

methods are illustrated in Figure 1 of the supplemental information. Compared to the 24 h FASP method, the DOD method provided equivalent sequence coverages for RAC, RBC and better results for RH, with a substantial increase in the number of characteristic peptides identified. Compared to the Flash method, the % sequence coverage and number of characteristic peptides identified improved dramatically. Additionally, intra-day data variability across replicates improved as compared to that for the FASP or Flash methods.

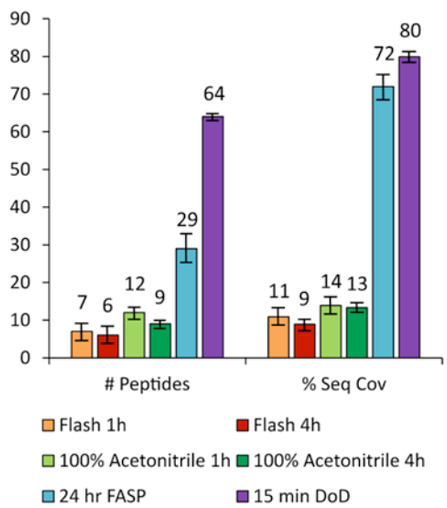


Fig. 4: Number of RH peptides identified and sequence coverages for 1 and 4 h Flash, 24 h FASP, 1 and 4 h 100% can, and the new 15 min DOD digestion methods with disulfide reduction.

3.4.1 Toxin Cocktail

DOD method performance for a protein cocktail containing RH, BotNT-A, and SEB was then evaluated. A 1 mL sample containing 22 µg/mL each of RH and BotNT-A, and 4 µg/mL of SEB was reduced and digested in quintuplicate using the FASP, Flash and DOD methods. Results confirmed our preliminary results using RAC, RBC and RH. Using the DOD method, sequence coverages for all three toxins were comparable to those attained with the 24 h FASP

method, but there was a >20% increase in detected peptides for BotNT-A, and a >25% increase in detected peptides for Ricin. In addition, across all targeted proteins, the DOD method reproducibility was substantially better than that resulting from the FASP or Flash methods. The DOD method produced only slight increases for SEB, but SEB was well digested with both the FASP and Flash methods. The number of characteristic peptides identified and percent sequence coverages for RH and BotNT-A using the 1 h Flash method were substantially lower (data not illustrated).

3.4.2 Digestion Time Optimization

Remarkably, a digestion time course study of RH, BotNT-A and SEB in a cocktail comparing the FASP method to the DOD method using digestion times ranging from 5-60 minutes, demonstrated that 5-minute digestion produced equivalent results to 60 min digestion. With the DOD method, across digestion times, the number of peptides identified and percent sequence coverages obtained were consistent down to as little as 5 minutes with comparable precision (Figure 5).

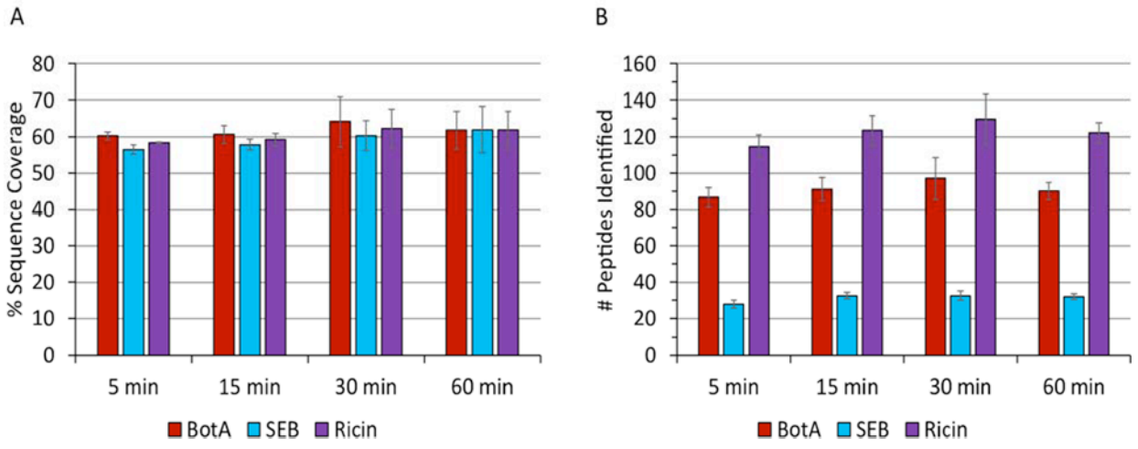


Fig 5: Time course study comparing DOD digestion time to sequence coverage (A) and number of characteristic peptides identified (B) for BotNT-A, SEB, and ricin in a cocktail of the three proteins at a concentration of 22 µg/mL for BotNT-A and SEB, and 4.4 µg/mL for ricin.

3.4.3 Thyroglobulin and Ribonuclease A

Using 5-, 15-, 30-, and 60-min digestions, samples in quintuplicate containing 1 µg each of TG and RNAse A were digested and analyzed as technical duplicates (total of 40 samples). Method precision was very high, both within and across digestion times. Between 260-280 characteristic peptides were identified for TG, and sequence coverages ranged from 50-55% with no significant variability in the data across digestion times. For RNAse A, the number of characteristic peptides identified per sample ranged from 35-45 with no significant difference within and across digestion times, and sequence coverages for all 40 samples were 100%. Sequence coverages and number of peptides identified under each condition for TG and RNAse A are illustrated in Figure 6 and Table 1. Sequence coverage results were so consistent across digestion times that we believe digestion time could likely be reduced even further without affecting method capability for these proteins. In fact, the data from RNAse A digestion indicated that at least for this protein, the number of characteristic peptides produced was highest with only a 5-minute digestion.

The effect of pH and temperature on DOD method digestion efficacy was further investigated by digesting quadruplicate biological replicates containing 1 µg each of TG and RNAse A in a cocktail under neutral pH conditions and under acidic conditions at 60 and 25ºC. Digestion was also compared employing trypsin-to-protein ratios of 1:5 and 1:1. Differences observed across different trypsin-to-protein ratios were small (data not shown) indicating that a stoichiometric excess of trypsin was efficacious even at a trypsin-to-protein ratio of 1:5. Method performance was optimal at 60ºC and neutral pH. The greatest numbers of characteristic peptides identified, and greatest sequence coverage were produced at 60ºC under neutral pH conditions. Though the samples consisted of only spiked proteins containing no clinical matrix, sequence coverage was extremely good for these two proteins using the DOD method. Additionally, the data from RNAse A digestion indicated that at least for this protein, the number of characteristic peptides produced was highest with only a 5-min digestion. Experiments were

conducted separately digesting at room temperature and under acidic conditions. The method worked at room temperature with somewhat less efficiency, but no detectable digestion occurred under acidic conditions.

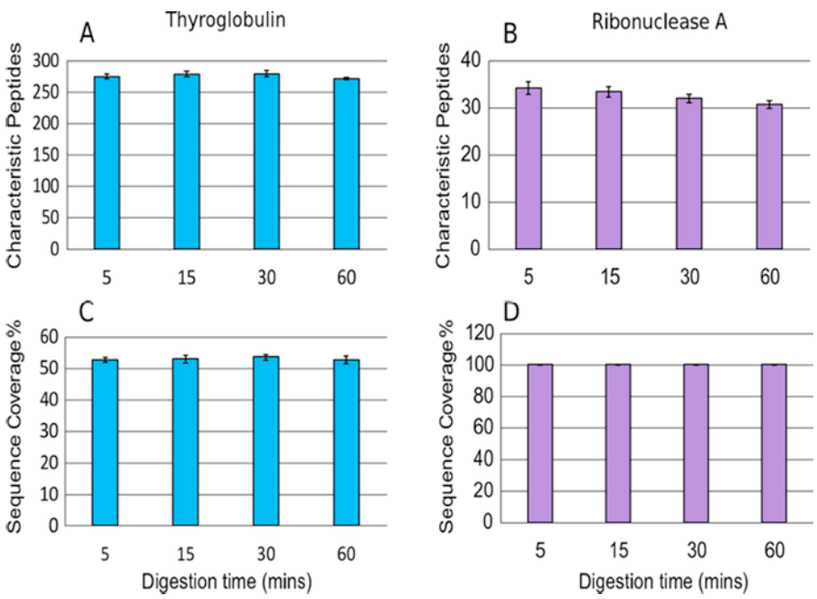


Fig. 6. Number of characteristic peptides and sequence coverages for TG and RNAse A over different digestion times using the DOD method. Mean data with error bars representing %CV for quintuplicate biological replicates analyzed as technical duplicates is illustrated.

Digestion time	5 min	15 min	30 min	60 min
TG peptides	273	279	280	271
TG sequence coverage (%)	53	54	54	53
RNAse A peptides	35	35.0	34	33
RNAse A sequence coverage (%)	100	100	100	100

Table 1. The number of peptides identified and sequence coverage for TG and RNAse A using the optimized DOD method. The table lists mean data for number of characteristic tryptic

peptides identified and the sequence coverage for each protein from technical duplicates digesting quintuplicate biological replicates and analyzing 1 µg of each protein.

3.5 Discussion and Conclusion

As described, a new rapid, in-solution trypsin digestion method was successfully developed, optimized and validated for a select list of toxins and standard clinical diagnostic proteins with a proteolytic digestion time of as little as five minutes. Digesting spiked protein samples containing no environmental or clinical matrix, the DOD method produced sequence coverage results and sensitivities that were equivalent to or often greatly exceeded those provided by the 24 h FASP method across all samples, and dramatically exceeded those provided by the Flash 1 and 4 h digestion methods. In addition, compared to the FASP method, intra-day method reproducibility was in most cases substantially improved. The DOD method substantially reduced analysis time compared to traditional methods, which can require hours to more than a day to sufficiently digest targeted proteins. In addition, in most cases, the DOD method provided equal or better sequence coverages than the methods to which it was compared. For targeted analysis of single or multiple proteins, the DOD method, while greatly reducing digestion times, appeared to provide an increase in the number of characteristic peptides identified, increased sequence coverages for many, and there was a marked increase in intra- and inter-day precision. Prior to method use for specific applications however, digestion capabilities need to be verified with spiked samples containing the sample-specific environmental or biomedical matrix.

With its comparatively short digestion time, consistent capability for identifying relative high numbers of characteristic tryptic peptides and proteins, and its consistently high sequence coverage capabilities when compared to other bottom-up methods, the DOD method should dramatically affect the ability to rapidly identify target toxins and other peptides/proteins of interest without the requirement for sophisticated equipment or costly materials. Importantly, if

validated further it could also be used to great advantage in proteomics to rapidly identify biomarkers of disease, identify microorganisms and viruses of interest, and changes in protein sequence leading to disease. Along with increases in analytical performance, time and cost savings in diagnostic and research laboratories could be substantial and current processes simplified. In addition, the demonstrated improvements in precision inherent with this method could enhance the absolute quantification of characteristic peptides and their parent proteins.

To date, we have also performed bottom-up protein analysis with the DOD tryptic digestion method employing it to characterize other target proteins for which characteristic peptide identification and resulting sequence coverage from currently used longer digestion methods is relatively poor and to characterize the proteomes of two complex systems. Using a study design similar to that described in this report, initial results with select complex samples were compared to results from other digestion methods currently in use. These results are described in the following chapter.

3.6 References

1. Shevchenko, A., Wilm, M., Vorm, O., and Mann, M., "Mass Spectrometric Sequencing of Proteins from Silver-Stained Polyacrylamide Gels". *Anal. Chem.*, 1996, Vol 68, pp. 850–858.

2. Wisniewski, J.R., Rakus, D., "Multi-enzyme digestion FASP and the 'Total Protein Approach'-based Absolute Quantification of the Escherichia coli Proteome", *Proteomics,* 2014, Vol 109, pp. 322-331.

3. Vazquez-Cintron, E.J., Beske, B.H., Tran, B.Q., Oyler, J.M., Angeles, C.A., Tenezaca, L., Syngkon, A., Mukherjee, J., Kalb, S.R., Band, P.A., McNutt, P.M., Shoemaker, C.B., Ichtchenko, K.. "Atoxic Derivative of Botulinum Neurotoxin C1 as a Molecular Vehicle for Targeted Delivery to the Neuronal Cytoplasm". *Scientific Reports*, 21 Mar, 2017, pp. 1-17.

4. Nagaraj, N., Lu, A., Mann, M., and Wisniewski, J.R.. "Detergent-Based but Gel-Free Method Allows Identification of Several Hundred Membrane Proteins in Single LC-MS Runs", *J. Proteo. Res.*, 2008, Vol 7, pp. 5028–5032.

5. Addona, T.A., Abbatiello S.E., Schilling, B, et al. “Multi-Site Assessment of the Precision and Reproducibility of Multiple Reaction Monitoring Based Measurements of Proteins in Plasma”. *Nature Biotech.*, 2009, Vol 27, pp. 633–641.

6. Hoofnagle, A.N. “Quantitative Clinical Proteomics by Liquid Chromatography-Tandem Mass Spectrometry: Assessing the Platform”. *Clin. Chem.*, 2010, Vol 56, pp. 161–164.

7. Slysz, G.W., and Schriemer, D.C., On-Column Digestion of Proteins in Aqueous-Organic Solvents, *Rapid Commun. Mass Spectrom.*, 2003, Vol 17, pp. 1044-1050.

8. Fink, A.L. and Painter, B., “Thermodynamics of Protein Denaturation. III. Denaturation of Ribonuclease in Water and in Aqueous Urea and Aqueous Ethanol Mixtures”. *Biochem.*, 1987, Vol 26, pp. 1665-1671.

9. Welinder, K.G., “Generation of Peptides Suitable for Sequence Analysis by Proteolytic Cleavage in Reversed-phase High-performance Liquid Chromatography Solvents.” *Anal.Biochem.*, 1988, Vol 174, pp. 54-64.

10. Park, Z.Y., Russell, D.H. “Thermal Denaturation: a Useful Technique in Peptide Mass Mapping”. *Anal. Chem.*, Vol 72, pp. 2667-2670.

11. Park, Z-Y., and Russell, D.Y., “Identification of Individual Proteins in Complex Protein Mixtures by High-Resolution, High-Mass-Accuracy MALDI TOF-Mass Spectrometry Analysis of In-Solution Thermal Denaturation/Enzymatic Digestion”. *Anal. Chem.,* 2001, Vol 73, Issue 11, pp. 2558–2564.

12. Russell, W.K., Park, Z-Y., and Russell, D.H., “Proteolysis in Mixed Organic-Aqueous Solvent Systems: Applications for Peptide Mass Mapping Using Mass Spectrometry”. *Anal. Chem.*, 2001, Vol 73, pp. 2682-2685.

13. Proc, J.L., Kuzyk, M.A., Hardie, D.B., Yang, J., Smith, D.S., Jackson, A.M., Parker, C.E., and Borchers, C.H., “A Quantitative Study of the Effects of Chaotropic Agents, SuRFactants, and Solvents on the Digestion Efficiency of Human Plasma Proteins by Trypsin”. *J. Proteome. Res.*, 2010, Vol 9, Issue 10, pp. 5422-5437.

14. Yu Zi Zhenga, Mari L. DeMarco. Manipulating Trypsin Digestion Conditions to Accelerate Proteolysis and Simplify Digestion Workflows in Development of Protein Mass Spectrometric Assays for the Clinical Laboratory. *Clinical Mass Spectrometry,* December 2017 Vol 6, pp. 1-12.

15. Christopher M. Shuford and Russell P. Grant. Cheaper, Faster, Simpler Trypsin Digestion for High-Throughput Targeted Protein Quantification. *J Mass Spectrometry and Advances in the Clinical Lab*, Nov 2023, Vol 30, pp. 74-82.

16. Montanaro, L., Sperti, S., and Stirpe, F., "Inhibition by Ricin of Protein Synthesis *In Vitro*. Ribosome as the Target of Toxin". *Biochem. J*., November 1973, Vol 136, Issue 3, pp. 677-683.

17. Walsh, M.J., Dodd, J.E., Hautbergue, G.M., "Ribosome-Inactivating Proteins Potent Poisons and Molecular Tools". *Virulence,* Nov 15 2013, Vol 4, Issue 8, pp. 774-784.

18. Arnon S.S., Schechter R, Inglesby, T.V., Henderson D.A., Bartlett J.G., Ascher M.S., Eitzen E., Fine A.D., Hauer J., Layton M., Lillibridge S., Osterholm M.T., O'Toole T., Parker G., Perl T.M., Russell P.K., Swerdlow D.L., Tonat K. "Botulinum Toxin as a Biological Weapon: Medical and Public Health Management". *J. Amer. Med. Ass.*, Feb 2001, Vol 285, Issue 8, pp. 1059–70.

19. Koirala J., Basnet S., "Botulism, Botulinum Toxin, and Bioterrorism: Review and Update". *Infect. Med.,* Vol 2, Issue 6, pp. 284-290.

20. Beske, P.H., Bradford, A.B., Grynovicki, J.O., Glotfelty, E.J., Hoffman, K.M., Hubbard, K.S., Tuznik, K.M., and McNutt, P.M., "Botulinum and Tetanus Neurotoxin-Induced Blockade of Synaptic Transmission in Networked Cultures of Human and Rodent Neurons". *Toxicol. Sci.*, February 2016, Vol 149, Issue 2, pp. 503-15.

21. Acharya K.R., Papageorgiou A.C., Tranter H.S., "Crystal Structure of Microbial Superantigen Staphylococcal Enterotoxin B at 1.5 A Resolution: Implications for Superantigen Recognition by MHC Class II Molecules and T-cell Receptors". *J. Mol. Bio.*, 1998, Vol 277, Issue 1, pp. 61–79.

22. Nema, V., Agrawal, R., Kamboj, D.V., Goel, A.K., and Singh, L., "Isolation and Characterization of Heat Resistant Enterotoxigenic Staphylococcus Aureus from a Food Poisoning Outbreak in Indian Subcontinent". *Int. J. Food Microbio.*, June 2007, Vol 117, Issue 1, pp. 29–35.

23. Ain, K.B., "*Clinical Management of Thyroid Disease*". 2009, Chapter 27, p. 383-397. Hoofnagle, A.N. and Roth, M.Y., "Improving the Measurement of Serum Thyroglobulin with Mass Spectrometry". *J. Clin. Endocrin. Metab.*, April 2013, Vol 98, Issue 4, pp. 1343–1352.

24. Schein, C.H., "From Housekeeper to Microsurgeon: the Diagnostic and Therapeutic Potential of Ribonucleases", *Nat. Biotech.*, June 1997, Vol 15, Issue 6, pp. 529-36.

25. Aebersold, R. and Goodlett, D.R., "Mass Spectrometry in Proteomics". *Chem. Rev.*, 2001, Vol 101, pp. 269- 295.

26. Aebersold, R. and Mann, R., "Mass Spectrometry-Based Proteomics". *Nat.*, 2003, Vol 422, pp. 198-207.

27. Gundry, R.L., White, M.Y., Murray, C.I., Kane, L.A., Fu, Q., Stanley, B.A. and Van Eyk, J.E., "Preparation of Proteins and Peptides for Mass Spectrometry Analysis in a Bottom-Up Proteomics Workflow". *Curr. Prot. Mol. Bio*., 2009, Chapter 10, Unit 10.

28. Massolini, G. and Calleri, E., "Immobilized Trypsin Systems Coupled On-Line to Separation Methods: Recent Developments and Analytical Applications". *J. Sep. Sci.*, January 2005, Vol 28, Issue 1, pp. 7-21.

29. Shah, V., Lassman, M.E., Chen, Y., Zhou, H. and Laterza, O.F., "Achieving Efficient Digestion Faster with Flash Digest: Potential Alternative to Multi-Step Detergent Assisted In-Solution Digestion in Quantitative Proteomics Experiments". *Rapid Commun. Mass Spectrom.*, 30 Jan 2017, Vol 31, Issue 2, pp. 193- 199.

30. Wisniewski, J. and Mann, M. "Universal Sample Preparation Method for Proteome Analysis". *Nature Meth*, April 2009, 6, 5, p. 359-62. Griffiths, J.R., Connolly, Y., Cook, K.,

Meyer, K. and Smith D.L., “A New Enhanced, Rapid and Precise Sample Preparation Protocol for Label-free Protein Quantification”. *J. Anal. Bioanal. Tech*, 2014, Vol 5, Issue 5.

31. Hustoft, H.K., Wilson, S.R., Reubsaet, L., Lundanes, E., and Greibrokk, T.. “A Critical Review of Trypsin Digestion for LC-MS Based Proteomics”. *Integr. Proteo.*, 2012, Chapter 4, pp. 73-92, Editor Hon-Chiu Leung.

Chapter 4

Proteomic Interrogation of Complex Biomedical Samples Using the Rapid Denaturing Organic Digestion (DOD) Method

Jonathan Oyler[a], Raymond F. Sullivan[b], Bao Q. Tran[c], Dhanwin Baker[a], Clare Coveney[a], David Boocock[a], Benjamin Oyler[d], Carole C. Perry[a], David P.A. Kilgour[a,e]

[a]Nottingham Trent University, Clifton Lane, Nottingham NG11 8NS, United Kingdom, [b]U.S. Army DEVCOM-CCDC Chemical Biological Center, 5100 Blackhawk Rd., Building E3330, Aberdeen Proving Ground, Gunpowder, MD 21010, USA, [c]US Army 20th CBRNE Analytical and Remediation Activity, 5183 Blackhawk Rd., Building E3150, Aberdeen Proving Ground, Gunpowder, MD 21010, USA, [d]National Institute of Infectious Disease, NIH, 9 W. Watkins Mill Rd., Rm. 212, Gaithersburg, MD 20878, USA, [e]Waltham Pet Care Science Institute, Waltham on the Wolds, Melton Mowbray, Leics LE14 4RT, United Kingdom of Great Britain and Northern Ireland

4.0 Acknowledgements

This research was made possible by funding provided by the Defense Threat Reduction Agency − Joint Science and Technology Office for Chemical and Biological Defense. Conclusions and opinions presented here are those of the authors and are not the official policy of the U.S. Department of Defense. Information in this report is cleared for public release and distribution is unlimited.

4.1 Abstract

Limitations to many current aqueous-based tryptic digestion methods include lengthy digestion times and both relatively high inter- and intra-day variability for both characteristic peptides identified and sequence coverages. This report describes results from digestion of some complex biomedical samples using the rapid Denaturing Organic Digestion method (DOD), an organic solvent-modified digestion method previously optimized for targeted protein digestion. Advantages of the DOD method included a very rapid digestion only requiring inexpensive

solvents and reagents generally available in the laboratory, with no requirement for specialized equipment or expensive, specialized consumables. For this study, samples of *E. coli* and murine ileum protein extracts, and K562, a mass spectrometry-compatible human protein extract and reference standard routinely used to evaluate methods, were digested. Sequence coverage and characteristic peptide identification results were compared to those from 18 and 24h conventional aqueous-based digestion methods. Across the samples tested, though the number of characteristic peptides and sequence coverages produced by the 5 min DOD method were very similar to those produced by the aqueous-based digestion methods, the specific characteristic proteins and their corresponding tryptic peptides identified following DOD method digestion included more hydrophilic and less hydrophobic species. In addition, we explored the effect of increasing digestion times with complex samples from 5 to 30 and 90 minutes for the DOD method. Increasing the digestion time to ≥30 minutes resulted in improved intra-day precision and the identification of many more peptide products than the currently used aqueous methods to which it was compared. These results suggest that the DOD organic-modified digestion method could, while markedly reducing protein digestion time, also provide more precise analysis and access to a somewhat different area of the proteome than that provided by current aqueous-based digestion methods.

4.2 Introduction

Characterizing the proteome of a biological system is essential to describing and understanding it since proteins, as a class, are the second most abundant mammalian molecule after water[1,2,3], are the most diverse of all the bio-macromolecules[4], and function as enzymes, basic structural building blocks, hormones and cytokines, transport/trafficking vehicles, and primary immune agents, et cetera. Additionally, the identification and characterization of proteins can be very important for clinical diagnostics, medical research, clinical and forensic attribution, and remediation.

Protein identification and quantification can be difficult. In Edmand degradation, a classic chemical method generally used for sequencing peptides up to 30-40 amino acids (AA) long used prior to the advent of MS-based methods, phenylisothiocyanate was reacted with the N-terminal AA under basic conditions forming a phenylthiohydantoin (PTH) derivative[5]. The pH was then adjusted to acidic conditions with TFA resulting in the cleavage of the N-terminal AA PTH derivative which was subsequently characterized by liquid chromatography or electrophoresis. This very lengthy process was repeated to sequence the peptide. The principal methods employing protease digestion used to characterize proteins prior to the advent of MS-based proteomics employed protease digestion followed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) separation or an immunoaffinity separation/detection method. These methods have been time consuming, required relatively expensive reagents that lacked the specificity needed to identify isoforms and post-translationally modified species (PTMs), and were relatively insensitive and irreproducible.

The development of electrospray ionization (ESI)[6] and matrix-assisted laser desorption/ionization (MALDI)[7,8] provided reliable, robust interfaces with mass spectrometers that could be used to measure non-volatile, high molecular weight compounds such as proteins. Traditional mass spectrometric approaches have included primarily bottom-up, middle-up/down and top-down proteomic analyses. The strength of top-down and middle-up/down proteomics as compared to bottom-up approaches has been the high specificity and sequence coverage capabilities of the methods, resulting from the high mass resolution and mass accuracy of the instruments used. In top-down approaches, intact protein ions are isolated and subsequent fragmentation occurs in a high-resolution mass spectrometer. By analyzing intact proteins, complete PTM maps and single amino acid mutations or substitutions can be identified with great accuracy[9,10]. The requirement for very high-resolution instruments like Fourier transform-ion cyclotron resonance (FT-ICR) instruments, due to the initial cost and the resources required to maintain them, is one limitation for top-down and middle-up/down analyses. Recently, higher

resolution Orbitraps and time-of-flight (TOF) instruments have been developed capable of performing the analyses. High-resolution instruments are required because of the need to resolve the high charge states of these large biopolymers when ionized by ESI. Another liability associated with these approaches is the inherent relatively long instrumental analysis times, with concomitant relatively poor sensitivities. This makes the identification of intact, low abundance proteins in complex mixtures difficult with these methods.

Bottom-up analysis is the most common approach employed for protein characterization and proteomics[11]. It does not have a requirement for high resolution mass spectrometry, usually requires a much shorter analysis time, and is, in practice, more sensitive. One major liability for bottom-up proteomic analysis is that sequence coverage is often not sufficient to fully characterize identified proteins and elucidate needed information. Another major liability is that lists of peptides from digested proteins cannot be used to define translated polypeptide sequences identified in the sample, *de novo*, since they are usually compared to *in silico* translated sequences. These sequence databases often do not contain sequence modifications or anomalies commonly observed in protein biosynthesis. Additionally, the inherent irreproducibility associated with bottom-up methods, which has been attributed to the tendency for protease autolysis during lengthy digestion times, especially at high relative protease concentrations, makes these methods less than optimal for many applications.

Trypsin, the most common protease used in MS-based protein analysis, is a digestive pancreatic serine endopeptidase that hydrolyses peptide bonds on the carboxyl side of lysine or arginine residues, except when either is followed by a proline residue. Previously published aqueous solvent-based methods have required 18-24 h digestion times[12]. Under the conditions used previously, unmodified trypsin has been prone to autolysis which is a primary factor contributing to inter- and intra-day variability[13]. Autolysis under these conditions has been directly correlated to trypsin concentration and digestion time. Unfortunately, reducing the trypsin concentration to reduce autolysis is counterproductive in that it also reduces proteolytic

efficiency. Recently, tryptic isoforms have been generated that are resistant to autolysis[14]. They have been generated through two mechanisms; acetylation or methylation of trypsin lysine residues or site-directed mutagenesis of trypsin. Though effective, the cost of these resistant trypsin isoforms adds significantly to the per sample cost of analysis.

With the advent of bottom-up proteomics has been the co-development of peptide databases that can be searched against an unknown peptide sequence in order to identify it and software applications which can perform denovo sequencing from raw MSMS data. Public proteomic databases include the Global Proteome Machine Database (GPMDB)[15], PeptideAtlas[16], the PRoteomics IDEntifications database (PRIDE)[17], Tranche (http://www.tranche.proteomecommons.org), and NCBI Peptidome[18]. There are many software applications using algorithms designed to perform de novo sequencing directly from the raw MSMS data including Antilope, MSNovo, NovoHMM, PepNovo, PEAKS, pNovo+, UniNovo, MaxQuant, and ProteinPilot.

The effect of organic-modified solvents and temperature on proteolytic digestion has been previously investigated. Fink et al. investigated the denaturation or unfolding of a single protein using methanol-water cryo-solvent systems[19]. They demonstrated that both the addition of methanol (MeOH) to the solvent and elevated temperature increased protein unfolding, ultimately choosing a solvent composed of 50% MeOH as the preferred solvent system for their studies. Welinder et al. studied the activity of eight proteases when digesting reduced and carboxymethylated ribonuclease, using organic solvent-modified systems[20]. The authors investigated four organic modifiers using two digestion temperatures, 22 and 37 ºC, and 2 and 18 h incubations. Amino acid analyses were carried out using N-terminal amino acid sequencing followed by reverse phase HPLC analyses. The analyses were performed using Waters ion-exchange chromatography for separation, followed by post-column derivation with o-phthaldialdehyde in non-halide buffers. The authors concluded that modification of the digestion

solution with organic solvent reduced trypsin activity and that activity was further reduced at a higher temperature.

In 2000 and 2001, Park et al.[21,22] investigated the effect of thermal denaturation of soluble proteins prior to digestion and found that, generally, digestion efficiency increased for single proteins and mixtures of proteins that were otherwise resistant to proteolytic digestion in aqueous solvents. Subsequently, Russell et al. investigated the use of organic-modified solvent systems for proteolytic digestion of 10 target proteins to increase method efficiency in high throughput peptide mapping[23]. Though digestion efficiency improvements were generally very modest, importantly, they discovered that along with digestion times as short as 5 minutes and solvent modification up to 80% organic, there was no need for detergents or other denaturants. Finally, in 2010, Proc et al.[24] evaluated the effects of two different organic solvents on the digestion efficiency of trypsin. Using isotopically labelled characteristic peptides specific to each protein as internal standards, solvent modifications of 40% and 20% AcN and MeOH (v:v), respectively, and a 37°C digestion temperature, they quantitatively measured the "absolute amounts" of peptides produced from the tryptic digestion of 45 clinically important plasma proteins over digestion times ranging from 0.5-23 h. They compared the results to those from an aqueous solvent digestion method employing sodium deoxycholate denaturation. To "prevent protein precipitation", the organic solvent was added to each sample following reduction and alkylation and just preceding the addition of trypsin. The proteins were categorized as "rapidly digested", "moderately digested" or "resistant to digestion" based on the results. Generally, digestion for most proteins was maximized at 4 h followed by a plateau. For a few proteins including fibrinogen γ-chain, however, digestion never plateaued throughout the 23 h and for some proteins the digestion signal decreased after 4 h. Importantly, regarding the use of AcN as a modifier, the authors stated, "a significant reduction in digestion efficiency was observed for all 45 analytes in the presence of 40% v/v acetonitrile, even though no protein precipitation was observed.".

Microwave heating of samples during chemical or enzymatic digestion has been employed reducing digestion time from hours to minutes. In 2006, Hua, et.al., concluded that microwave heating could greatly accelerate the in-gel digestion of myoglobin and BSA by mild acid, as well as proteins isolated from Escherichia coli K12 cells[25]. Subsequently, in 2019, Kim, et.al., investigated the effects of temperature and AcN modification of digestion solvent on the microwave-assisted digestion of horse skeletal muscle myoglobin and bovine serum albumin[26]. They employed temperatures of 25, 37, and 55 ^{o}C, using 0, 10 and 20% AcN over digestion times ranging from 10-50 minutes. They concluded that similar sequence coverages were observed across all conditions except with 20% AcN and 50 ^{o}C which "provided the lowest sequence coverages for both MYG and BSA for all investigated digestion times".

Lastly, the use of trypsin in combination with Lys C, a serine endoprotease that specifically cleaves proteins on the carboxyl side of lysine residues optimally at pH of 7-9, has historically been used to digest "difficult to digest proteins". In 2013, Saveliev, et.al., investigated the use of trypsin/Lys C for general proteomics[27]. They found that use of the combination of proteases digestion of produced increased number of identified peptides and proteins, higher analytical reproducibility and more accurate protein quantitation. They wrote, "Effectively, by supplementing trypsin with Lys-C, we create improved trypsin."

We previously reported on the development of the Denaturing Organic Digestion (DOD) method[28], a 5-min organic solvent-modified tryptic digestion method for targeted protein analysis. The DOD method was compared to two commonly used tryptic digestion methods, the Filter Aided Sample Prep (FASP)[29] and the Flash methods[30]. The FASP method employs a detergent to disrupt cells exposing cellular proteins to proteolytic digestion. Critical method steps occur in a 30-k filter, using 8 M urea, and a digestion time of 24 h. The Flash method, a rapid digestion method, is an on-column based method that employs immuno-affinity immobilized trypsin chemically bonded to the column stationary phase. The DOD method employs a relatively high temperature, 60 ^{o}C, and an organic solvent-modified system containing a

relatively high amount of AcN (60% v:v). Importantly, the use of the Eppendorf Thermomixer C with a heated lid was determined to be essential for consistent method performance. Initially, it was evaluated across protein toxins including ricin, ricin A chain, ricin B chain, *Clostridium botulinum* neurotoxin A, and *Staphylococcus* Enterotoxin B as part of an effort to develop a single highly specific and sensitive rapid screening assay for proteins having a potential for weaponization. Performance measures included protein sequence coverage and the number of characteristic peptides identified. Subsequently, it was employed to digest thyroglobulin and ribonuclease A, two clinically relevant proteins for which routinely used clinical digestion methods have lacked sensitivity, specificity, and reproducibility largely due to digestion method inefficiency and lack of repeatability[31,32]. DOD method performance equaled or bettered that of the aqueous solution-based and surface-immobilized trypsin digestion methods to which it was compared, and, for most of the targeted proteins, performance was substantially improved. The authors postulated that these improvements were in large part due to precise temperature control provided by the Thermomixer C employed.

To further evaluate the utility of the DOD method tryptic digestion performance capabilities for proteomic analysis of complex samples, we digested biological replicates of three complex samples, including *E. coli* and murine ileum protein extracts, and K562, a mass spectrometry-compatible human protein extract and reference standard routinely used to evaluate digestion methods. Results were compared to those from two routinely used proteolytic digestion methods developed to generate peptides from crude cell lysates which are currently used in mass spectrometry-based proteomics. The methods were the FASP method, described previously, and the "NTU Long Digestion Method", an overnight aqueous-based tryptic digestion method employed at the time this work was conducted in the Nottingham Trent University – Clifton Campus John van Geest Cancer Research Centre proteomics laboratory for routine proteomic analysis. Proteomic analyses of murine ileum determined that the proteome consists of >3500 individual proteinsFinally, samples of fibrinogen γ-chain, a plasma protein

categorized by Proc et. al. as “resistant to digestion” by two organic solvent-modified tryptic digestion methods[24], were digested using the DOD method and the NTU digestion method and peptide results compared.

4.3 Materials and Sample Preparation Equipment

The MS-compatible yeast and human protein extract protein standard reference, K562, was acquired from Promega (catalog #V6951).

4.3.0 FASP Method

The following materials and equipment were used: urea, iodoacetamide (Sigma Aldrich catalog #114959/0681253022), LC/MS grade water, lyophilized Promega Gold mass spectrometry-grade trypsin, NaCl, ammonium bicarbonate, sodium disulfide (SDS), Tris/HCl, dithiothreitol (DTT) (Sigma catalog #D9779), trifluoroacetic acid (TFA), Eppendorf Thermomixer C with heated lid, 1.5 mL Eppendorf snap-cap conical tubes, Speedvac, 30 k filter, and formic acid (FA).

4.3.1 DOD Method

The following materials and equipment were used: AcN, iodoacetamide (Sigma Aldrich catalog #114959/0681253022), LC/MS-grade water, lyophilized reagent-grade trypsin (Sigma Aldrich, catalog #T4799), DTT (Sigma Aldrich, catalog #D0632-25G), Eppendorf Thermomixer C with heated lid, 1.5 mL Eppendorf snap-cap conical tubes, Speedvac, FA, and TFA.

4.3.3 NTU Digestion Method

The following materials and equipment were used: AcN, iodoacetamide (Sigma Aldrich catalog #114959/0681253022), LC/MS-grade water, lyophilized Promega Gold mass spectrometry-grade trypsin, DTT (Sigma Aldrich, catalog #D0632-25G), 1.5 mL Eppendorf snap-cap conical tubes, BIOER MB-102 mixing block, ultracentrifuge, Speedvac, FA, and TFA.

4.4 Methods

4.4.0 FASP 24 h Digestion Method

Samples were processed by the filter-aided sample preparation (FASP) proteolytic digestion method according to the protocol described by Wizniewski, *et al*.[22]. Briefly, proteins were reduced with 20 mM DTT (prepared fresh daily) for 1 h at 60 ºC followed by alkylation with 55 mM iodoacetamide for 45 min at room temperature prior to proteolytic digestion. Additional method details are described in chapter 3 and in the manuscript describing the original work. Samples were digested overnight at 37 ºC using 0.02 µg trypsin/µL, prepared in 50 mM aqueous ammonium bicarbonate, at approximately 1:100 trypsin to protein ratio. Following digestion, samples were centrifuged under vacuum to dryness for 40 minutes and stored at -20 ºC until analysis. Prior to analysis, samples were suspended in 100 µL of LC mobile phase.

4.4.1 NTU Long Digestion Method

The NTU method employs an overnight aqueous-based tryptic digestion. Septuplicate biological replicates were digested according to the following regimen. A 30-µg sample of target protein was diluted in 100 µL of 50 mM triethyl ammonium bicarbonate. Sample proteins were reduced by adding 1 µL of 0.5 M DTT (Sigma Aldrich D9779 SLBH59545) with a 20-minute incubation at 56 °C shaking at 600 rpm on a BIOER MB-102 mixing block followed by a pulse centrifugation to collect lid condensate. To alkylate sample proteins, a 2.7 µL aliquot of 0.55M iodoacetamide (Sigma Aldrich 114959/0681253022) was added, and the sample was vortexed and incubated for 15 min at room temperature in the dark. A 1 µL aliquot of 1 µg trypsin /µL was added, and the sample was thoroughly vortexed and again incubated at 37 °C overnight (16.75 hrs). Pulse centrifugation was conducted to collect lid condensate and digestion terminated by adding 0.5 % TFA to a final concentration of 0.5 % (v:v). Using the Speedvac V-AQ setting, samples were vacuum-centrifuged to dryness for 45 minutes and stored at -20°C. Prior to micro LC/MSMS analysis, samples were thawed at room temperature, suspended in 25 µL of 10% AcN containing 0.1% FA, and shaken for 15 minutes at 5 °C. Sediment particles

were removed employing 13 G centrifugation prior to transferring the supernatant to a sample vial for micro LC/MSMS analysis.

4.4.2 DOD Method

Sample digestion was performed as follows (a fresh solution of 1.0 M dithiothreitol (DTT) in LC/MS-grade water was prepared prior to each digestion). A master mix was formulated consisting of 10-70 µg of targeted protein, 22 µL of 1.0 M DTT (final concentration 20 mM), 660 µL of AcN (60% v:v), and 396 µL LC-MS grade water for a total sample volume of 1100 µL. A 1 µg trypsin/µL LC/MS-grade water solution was prepared fresh daily. Aliquots (50 µL or 450-3180 ng of protein) of the master mix were transferred to separate 1.5 mL Eppendorf snap-cap tubes. Digestion was initiated by addition of trypsin at a 1:3 molar ratio to target protein(s). Vial caps were affixed, and samples were incubated at 60 ºC at 700 rpm in an Eppendorf ThermoMixer C equipped with a heated lid for 5, 30 or 90 minutes. Digestion was terminated by the addition of 1 µL of 50% TFA to each tube. Samples were dried under vacuum for 30 minutes in a Rotovap under medium heat until dry and stored at -20 ºC until analysis. Samples were then suspended in LC/MSMS mobile phase prior to nanoLC/MSMS analysis. Note that the DOD digestions involve much higher trypsin to protein ratios: 1:3 as opposed to the 1:100 or 1:50 in conventional digestions. However, the DOD digestions can be undertaken using low cost, reagent grade trypsin instead of the much more expensive grades normally used in protein digestions for LCMSMS analysis. Therefore, the DOD digestions were more cost effective, in terms of the consumables costs, as well as taking less time.

4.4.3 LC/MSMS Analysis

LC/MSMS analyses of the *E. coli* and murine samples were performed on an Orbitrap LC-MSMS. Digest products were suspended in 50 µL of 95/5/0.1 % H_2O/AcN/Formic Acid (v:v:v) as described above, loaded on a 100 µm id x 20 mm x 5 µm 200 Å AQC18 trap column, and separated by reversed-phase nanoLC, on a 75 µm id x 18 cm x 5 µm 100 Å AQC18 analytical column, using 0.1% formic acid in water and 0.1% formic acid in AcN as mobile phases A and

B, respectively. The nano-LC reverse-phase separation method was initiated using 100% solvent A with an initial hold for 2 min followed by a gradient over 45 min to a final mobile phase composition of 90% solvent B/10% solvent A. The nanoLC was interfaced with a Thermo Fisher Scientific Fusion Orbitrap mass spectrometer or a Q Exactive Plus Hybrid Quadrupole-Orbitrap mass spectrometer equipped with a nano ESI source. Mass spectrometric data were collected using a 3 sec, top-speed data-dependent acquisition (DDA) method in which full MS scans were collected at a resolving power of 120,000, over a 350-1500 m/z range. Peptide precursors were selected and subjected to collision-induced dissociation (CID), and MSMS spectra were collected from the ion trap, employing a rapid scan rate. The data were processed using MaxQuant software (version 1.5.3.30) which generates a tryptic database using predicted trypsin amino acid sequence cleavage sites for each protein. Spectra were searched against the MaxQuant-generated database for each respective proteome. MaxQuant parameters used to process the data were set to the default with a missed cleavage setting of 2, variable modifications of oxidation on methionine, and fixed modification of carbamidomethylation on cysteine. A 0.5 Da window was used to isolate product ions, and full scan spectra were matched employing a mass tolerance of 10 ppm. Peptide identification was performed with a false discovery rate of 1%.

For the K562 NTU long digestion experiments, 30 µg samples of target protein were analyzed on a Sciex QTOF high resolution instrument. A sample was brought to 100 µL in 50 mM tetraethylammonium bromide. Each sample was subsequently reduced and alkylated by adding 1 µL of DTT followed by a 20-minute incubation at 56°C shaking at 600 rpm. Samples were briefly centrifuged to collect lid condensation, and 2.7 µL of 0.55 M iodoacetamide was added. Samples were then vortexed and incubated at room temperature for 15 minutes in darkness. A 1µL aliquot of 1µg/µL trypsin was added to each, and samples were thoroughly vortexed and incubated for 18 hours at 37°C overnight, shaking at 600 rpm. Condensate was collected with a pulse spin, and digestion was terminated by adding TFA to a final concentration

of 0.5%. Samples were vacuum-dried for 45 minutes in a Speedvac at 60°C using the "V-AQ" setting, visually checking at 20-minute intervals, then, frozen at -20°C. Prior to MS analyses, the samples were then resuspended in 25 µL of 0.1% FA in 10% AcN and shaken for 15 minutes at 5°C followed by the addition of 25 µL of 0.1% FA and a further 15-minute shaking at 5°C. Sediment particles were removed by a 13 G centrifugation prior to transferring the supernatant from each sample to LC vials for LC/MSMS analysis. Triplicate biological replicates were digested and analyzed in triplicate by micro LC/MSMS using a conventional ESI source on either a SCIEX 5600 or 6600 Triple TOF instrument. The data were processed using Protein Pilot v5.0.3 to identify proteins and evaluate the resulting sequence coverages and characteristic peptide fragments. Data were extracted from the "Protein Summary" and "Peptide Summary" pages in Protein Pilot. For identified proteins with a false discovery rate <1%, the % sequence coverage and number of characteristic peptides resulting from the DOD and NTU digestion methods were compared. For characteristic peptides, the peptide sequence coverages and peak intensities (areas) were also compared.

4.5 Results

4.5.0 E. coli Whole Cell Lysate Digestion

The DOD method was first applied to a protein extract from an *E. coli* whole cell lysate to evaluate optimum digestion time using a complex sample. Without alkylation, a total of 568 characteristic proteins were successfully identified using a 200-ng sample. Quadruplicate 200-ng samples were digested for 5, 15, 30 and 60 minutes, and the mean data are listed in Table 1. Across digestion times, the mean number of characteristic peptides and proteins identified, and the proteome sequence coverages, were very consistent at ca. 475 proteins and 12% sequence coverage, respectively. Protein identification was based on the identification of ≥2 characteristic peptides.

Digestion time	**5 min**	**15 min**	**30 min**	**60 min**

Peptides	3007	3127	3161	2783
Protein identifications	484	496	505	466
Proteome coverage (%)	12.1	12.4	12.6	11.7

Table 1. The effect of digestion time on the mean number of characteristic tryptic proteins identified with corresponding peptides, and the average proteomic sequence coverage obtained from *E. coli* tryptic digests prepared using the DOD method. Using the calculated RSD for each data set, differences observed across digestion times were insignificant.

4.5.1 Mouse Ileum Protein Extract Digestion

Subsequently, digestion efficiency of the DOD tryptic digestion method was compared to the commonly used 24 h FASP method using a second highly complex sample – whole mouse ileum protein extract. Initially, quadruplicate biological replicates were digested using the DOD method employing 5-, 15-, 30- and 60-min digestion times; analyses were performed on a Q Exactive Plus Hybrid Quadrupole-Orbitrap mass spectrometer. The total number of protein groups and proteins identified (≥2 characteristic peptides identified) was very consistent across digestion times. In addition, the number of proteins tentatively identified through the identification of only one characteristic peptide was also very consistent across digestion times as was the average number of proteins identified with ≥2 characteristic peptides. Interestingly, the average number of peptides identified per sample seemed to drop with a digestion time of 60 minutes. Quadruplicate biological replicates were then digested using the FASP 24 h method; analyses were also performed on a Q Exactive Plus Hybrid Quadrupole-Orbitrap mass spectrometer. Data from the experiments are listed in Table 2. When compared with the FASP digestions, the DOD digestion produced somewhat fewer protein and peptide identifications (Figure 1). The mean number of proteins (with ≥2 characteristic peptides identified) common to all DOD method samples across the four digestion times was 585 with 737 proteins identified across all samples as compared to 607 proteins common to all FASP samples and 801 proteins across quadruplicate biological replicates using the FASP 24 h digestion method. In one study, proteomic analyses of murine ileum determined that the proteome consists of >3500 individual

proteins[33]. The major reason for the observed relatively low identification rates with both methods is most likely because the copy number differences per sample across individual protein species in complex native tissue samples can range widely with as much as 8-fold differences occurring[34]. This has been well characterized and is widely known by those conducting tissue proteomics. Less likely possible reasons for the relatively low identification rates with both methods could include trypsin saturation during the digestion process, and/or ion suppression for species less easily ionized or in relative low abundance in the ion source of the mass spectrometer. Using additional methods to physically separate the proteins into further subsets prior to digestion and proteomic analysis could result in the identification of additional proteins.

Identification	5 min DOD	15 min DOD	30 min DOD	60 min DOD	24 h FASP
Total Proteins	660	665	717	663	1008
Total proteins ≥ 2 peptides	590	595	650	569	801
Total single hit proteins	70	70	67	74	207
Average # of proteins ≥ 2 peptides	546	537	552	497	731
Average No. of Peptides	1956	1809	1842	1595	3077
Average sequence coverage (%)	6.3	5.6	5.9	4.8	9.6

Table 2: Comparison of the total number of proteins and peptides identified and their respective sequence coverages from a tryptic digest of 400 ng of mouse ileum protein extract employing 5-, 15-, 30-, and 60-min digestion times using the DOD digestion method and with quadruplicate biological replicates employing the 24 h FASP tryptic digest with alkylation.

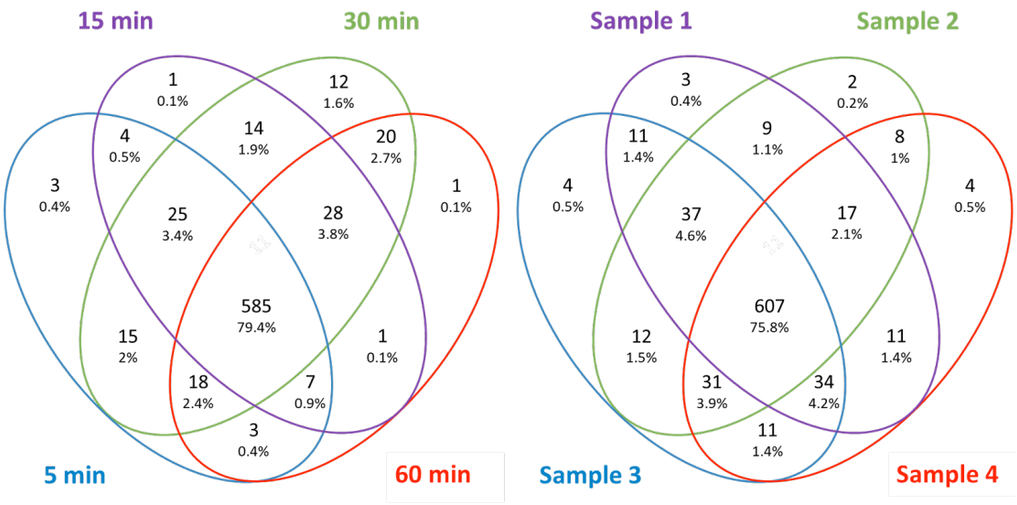


Figure 1: Venn diagrams of the number of mouse ileum proteins identified employing the 5-, 15-, 30-, and 60-min DOD tryptic digestion methods (left panel) and four biological replicates using the 24 h FASP method (right panel).

4.5.2 K562 Human Protein Digestion

The 5-minute DOD digestion method was then used to digest K562, a mass spectrometry-compatible human protein extract and reference standard routinely used to evaluate digestion method performance. Results were compared to those from an overnight aqueous-based tryptic digestion method routinely employed in the Nottingham Trent University (NTU) - John van Geest Cancer Research Centre proteomics laboratory for routine proteomic analysis. Sample replicates of 30 µg protein were digested with each method and results compared. Protein sequence coverages, number of characteristic peptides identified, and peptide LC-MSMS peak areas were compared. Proteins and peptides were considered "identified" only if they were found in all replicates. The resulting protein sequence coverages and numbers of characteristic peptides were similar (Figure 2) for the two methods with the longer method providing a somewhat greater number of characteristic peptide identifications but with less statistical certainty, as determined by the precision of peptide LC-MSMS peak area measures. There was

substantial overlap in common proteins identified (175), but also significant numbers of unique proteins; with the 5-minute DOD method identifying 115 unique proteins and the NTU method identifying 144 unique proteins. The number of proteins identified when results were combined (434) provided a substantial increase when compared to use of either single method alone (290 and 315) for the DOD and NTU methods, respectively).

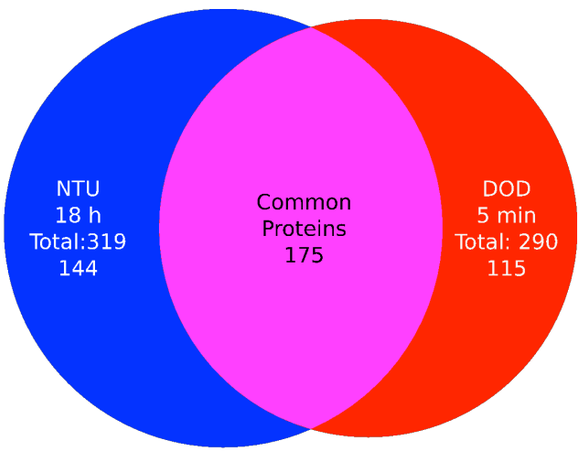


Figure 2. The DOD 5 min and NTU 18 h tryptic digestion method protein profiles for the human protein reference standard K562 (the top number in each area is the total number of proteins identified, and the lower number is the number of proteins identified unique to the digestion method).

For 3 biological replicates, each analyzed three times, the overall results are shown in Figure 3. In the data analysis, as described above, proteins and peptides were only classed as detected if they were detected in all replicates for a particular protocol; peptides that were only detected sporadically were ignored.

The results presented in Figure 3 are shown versus the protein or peptide rank because sorting the data in this way makes the trends easier to visualize. Rankings of said proteins or peptides are calculated with regards to the other metric being displayed in that graph (i.e. if the metric is peptide intensity then the most intense peptide gets ranked 1, next most intense as 2 etc.; if the metric is %RSD, then the best %RSD gets ranked as 1). A greater number of proteins

was identified using the 18-h NTU digestion method (315 vs 290), but for those proteins identified, the DOD 5 min digestion generally provided slightly more sequence coverage (Figure 3-A); similarly, for proteins identified in common across the two methods, sequence coverages for the DOD 5 min method are slightly superior for the majority of proteins, only losing out once the sequence coverage drops below 20% (Figure 3-B). The number of peptides per protein identified for the lower ranked proteins, however, was relatively low. Despite the 5-minute DOD digestion method brevity, there were very similar numbers of characteristic peptides (Figure 3-C). However, the replicate peptide LC-MSMS mean peak areas were lower for the 5-minute DOD method (Figure 3-D).

Though the total numbers of proteins identified, and the sequence coverages were similar for the two methods, LC/MSMS peak area precision (%RSD) was better using the NTU digestion method. (Figure 4). The distribution of peptide lengths was similar but the 5-minute DOD digestion method produced (~20%) longer tryptic peptides, on average. To evaluate and compare the raw data, David Kilgour developed an in-house algorithm.

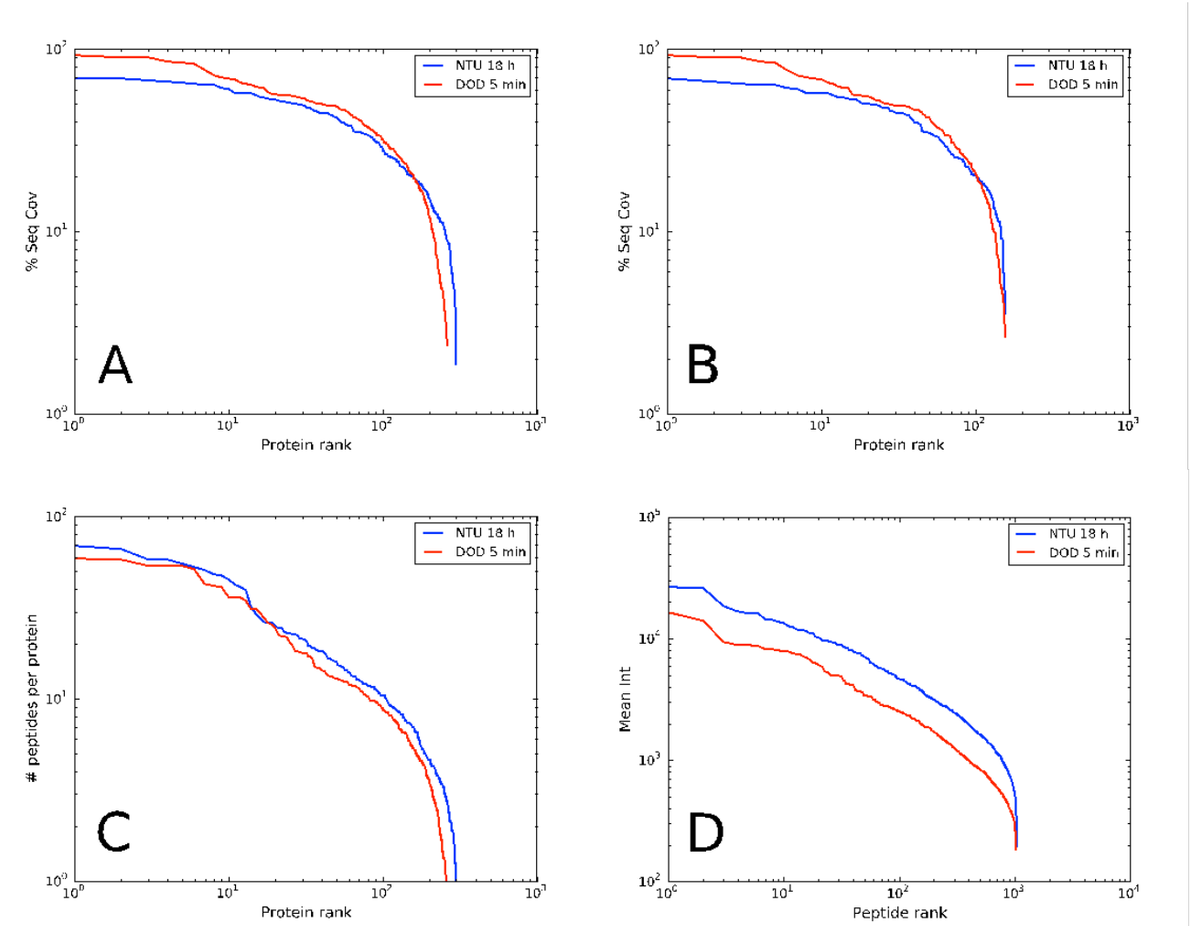


Figure 3. K562 DOD 5-min and NTU 18-h digestion results comparisons for (A) sequence coverages for all proteins identified by each method, (B) sequence coverages for proteins commonly identified across methods, (C) number of characteristic peptides identified per protein, and (D) mean peak intensities of unique characteristic peptides identified per protein.

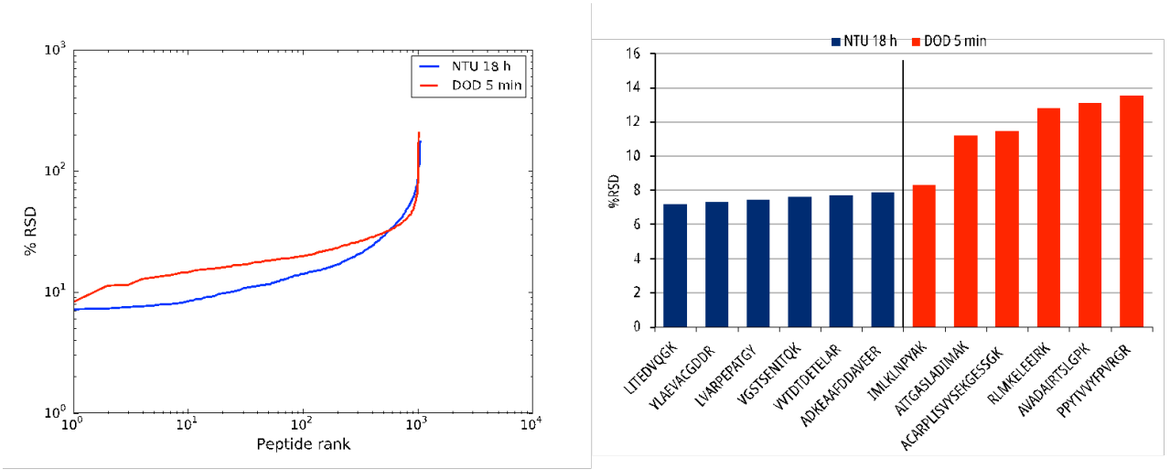


Figure 4. DOD 5-min and NTU 18-h LC/MSMS peak area precision (%RSD for peptide products) for the most precise data sets for each digestion method.

The proteins identified with the two methods were further characterized according to hydrophobicity based on their sequences and the method of Abraham and Leo, as is also used by https://web.expasy.org/protscale/. The hydrophobicity profile for proteins identified by each method (Figure 5) illustrated that, though there was considerable overlap in the specific proteins digested, the two methods did not digest the same set of proteins. The proteins unique to the DOD method tended to be slightly more hydrophilic proteins and those for the NTU method more hydrophobic, with the distribution of proteins that were seen by both methods sitting in the middle. In addition, tryptic peptide hydrophobicity profiles from each method were also determined. The distribution of the peptides generated by the DOD method tryptic peptides was also slightly more hydrophilic in character than peptides generated from the NTU method. We therefore conclude that combining the DOD method in parallel with more conventional aqueous digestions might provide even broader coverage of the proteins in complex samples than analysis by either individual method, and while the underlying mechanism is not yet fully understood, the extra information may still be valuable.

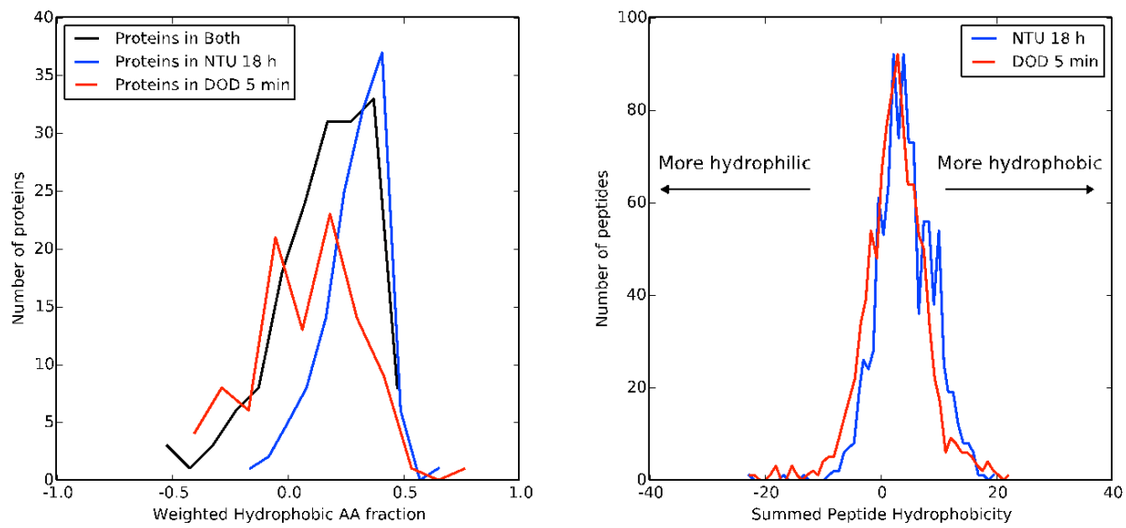


Figure 5. Relative hydrophobicity of proteins digested by the DOD and NTU methods.

The experiments were repeated with K562 increasing the 5-minute DOD method digestion time to 90 minutes - 18x longer than the original digestion time, but still ≥12x shorter than 18-h conventional digestion methods (Figure 6). The 90-minute digestion time shows similar performance to the DOD 5 min and conventional digestion in terms of the sequence coverage of the detected proteins (Figure 6-A). But, the DOD 90-min digestion shows the best overall performance when considering the combination of the number of proteins detected (NTU 18-h: 319 proteins, DOD 90-min: 330 proteins and DOD 5-min: 290 proteins), number of peptides per protein (Figure 6-B) and quantification precision (%RSD – Figure 6-C).

Figure 7 illustrates select peptide peak area precision differences when comparing the data for the most precisely measured peptides with the 5 and 90 min DOD, and NTU 18-h digestion methods.

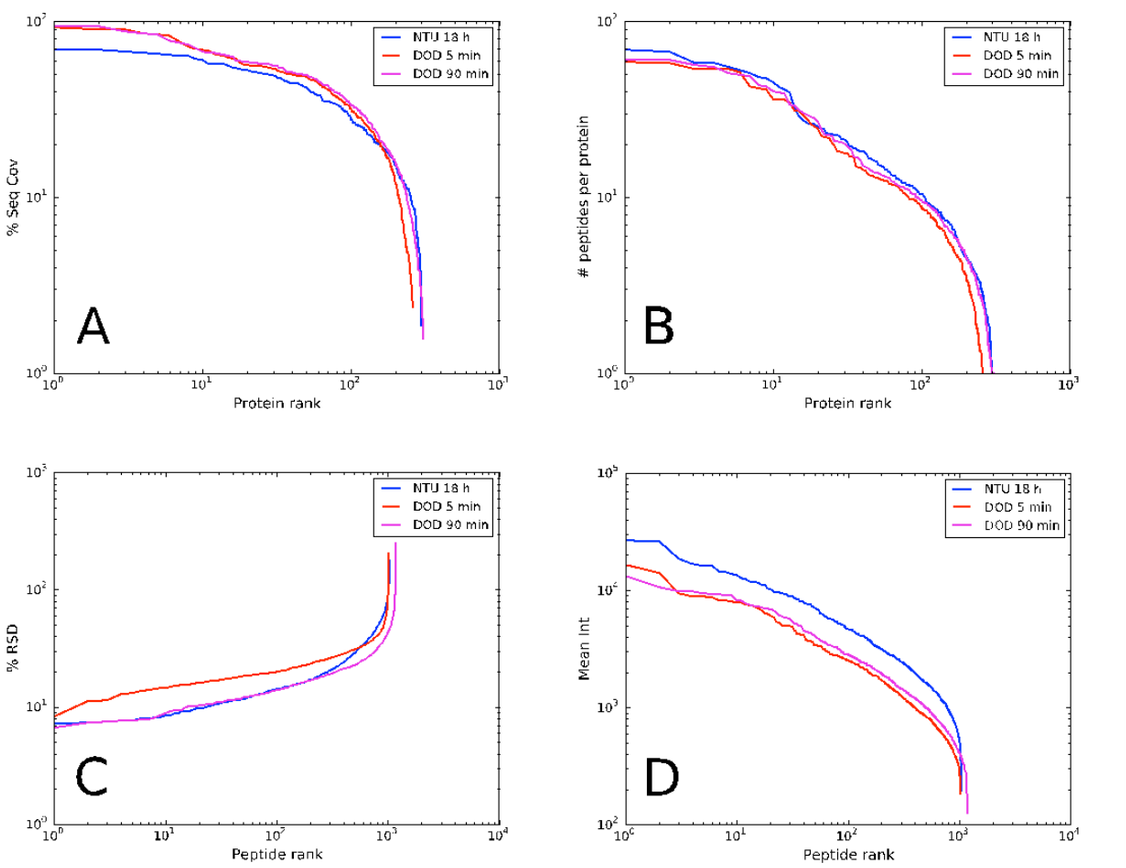


Figure 6. K562 DOD 90 min, 5 min; and 18-h NTU digestion results comparisons for (A) Sequence coverages for all proteins identified by the different methods, (B) number of peptides per protein, (C) the %RSD reproducibility of peptides detected by each method, (D) mean peak areas of unique characteristic peptides identified for each method.

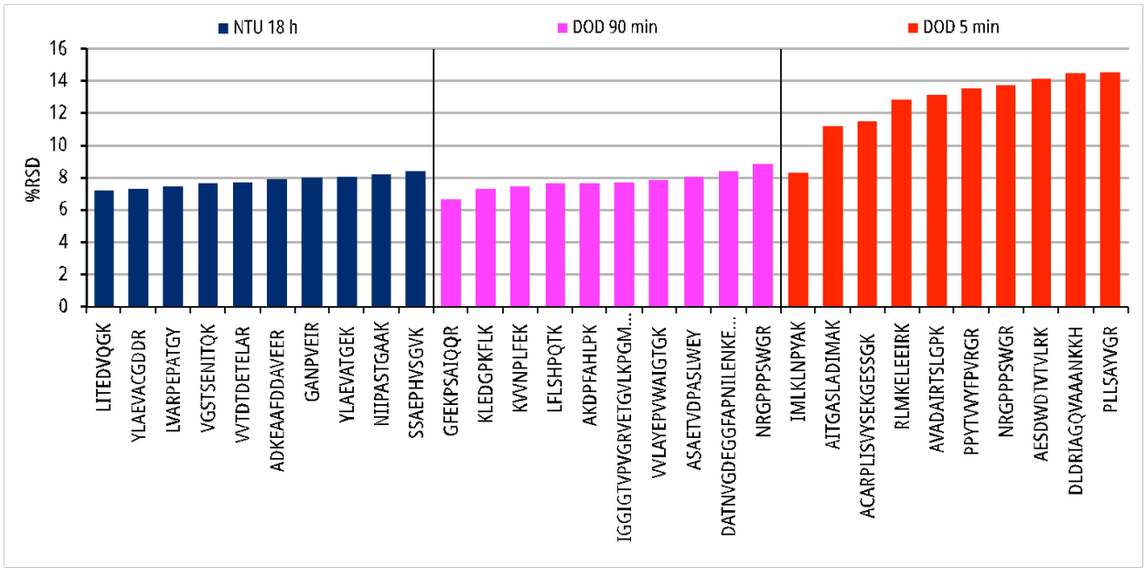


Figure 7. Comparison of the 10 most precise peptide LC-MSMS peak area data sets from digestion of K562 for the 18-h NTU, 90-min DOD and 5-min DOD digestions.

4.5.3 Fibrinogen Component Protein Digestions

Finally, 3 biological replicates of fibrinogen were digested in triplicate and analyzed. Fibrinogen is composed of 3 component proteins: the α, β and γ chains. This protein was selected because the fibrinogen α, β and γ chains were previously studied by Proc et. al.[24]. In part of that study, the authors were testing the utility of high solvent digestion methods for improving bottom-up digestions; however, the temperatures used were physiological (37 °C) and the trypsin to protein ratios were intermediate (1:20). Under those conditions, the authors concluded that high solvent digestion was unlikely to be generally applicable as they identified several proteins that were resistant to solvent-assisted digestion. They highlighted fibrinogen γ chain as a particularly difficult case. In the same study, they also tested the α and β chains and identified that these were proteins that did benefit from solvent assisted digestion. Therefore, fibrinogen was a useful test sample for characterizing such digestions as it contains a key target

of interest (the γ chain) and two controls (the α & β chains). We aimed to test whether the additional higher temperature and high trypsin concentrations of the DOD digestion protocol might improve the digestion results for fibrinogen γ chain.

The results of digesting fibrinogen are shown in Figure 8. As with the other analytes reported previously[28] or here, data precision for the longer DOD digestions (90-min in this case) exceed the analytical precision of the conventional digestion, even for fibrinogen γ chain. If the aim was number of peptides, then the 90-min DOD digestion provided 45% more fragments for fibrinogen α and similar values for β & γ compared to the conventional, 18-hour digestion. Sequence coverage using the DOD 90-min digestion was not quite as good as the sequence coverage achieved by the conventional 18-hour protocol, but it still provided useful coverage, coming at only 11%, 9% and 6% below the sequence coverage from 18-h digestion of fibrinogen α, β, and γ, respectively. For the shorter digestions (5 min and 30 min) the DOD digestions generally performed similar to the conventional 18-hour digestion in terms of the %RSD and hence, for targeted methods, it would be possible to achieve similar levels of precision from the 5-min DOD digestion as for the conventional 18-hour digestion for fibrinogen α & β.

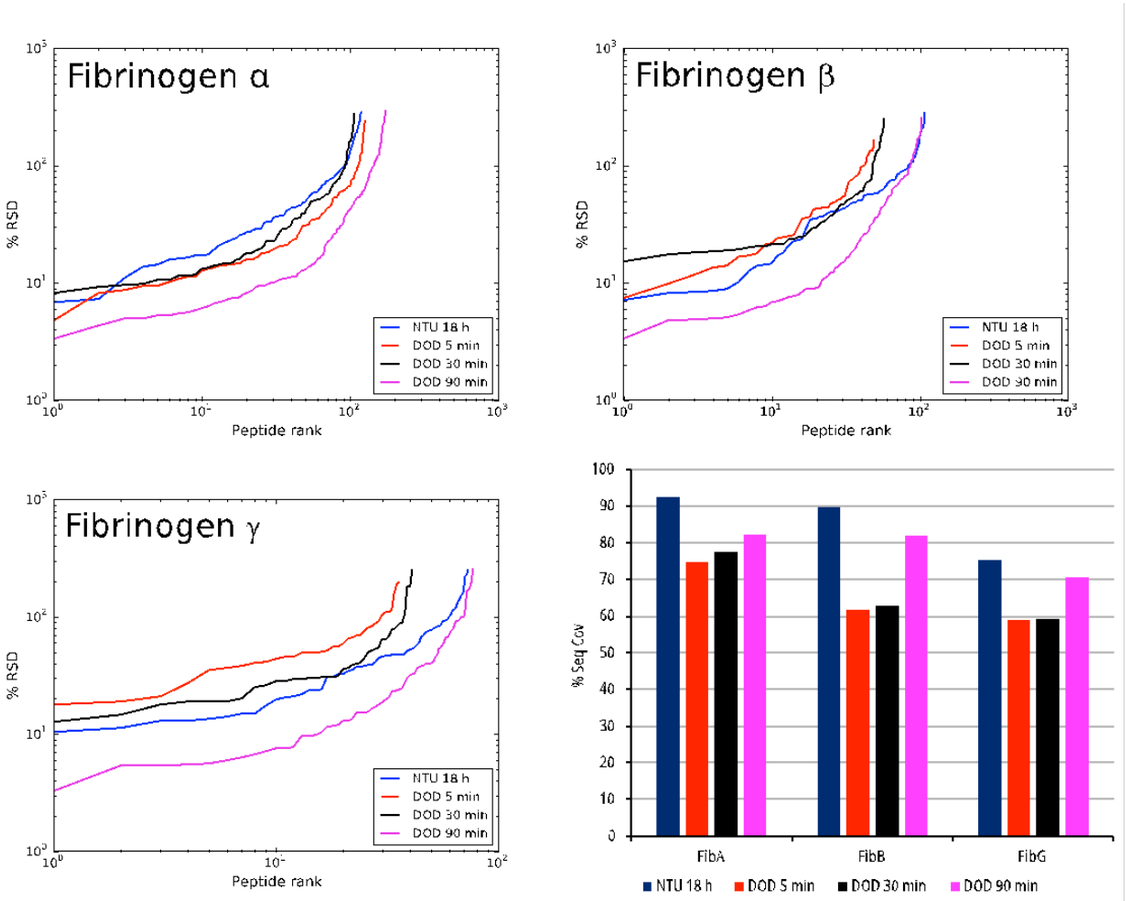


Figure 8. Comparison of the number of peptides seen in every biological replicate produced from digesting fibrinogen α, β and γ-chains, using the 18-h NTU, and the 5-, 30-, and 90-minute DOD tryptic digest methods for human fibrinogen component proteins.

4.6 Discussion and Conclusion

To date, we have performed bottom-up proteomic analyses using the new, very rapid, simple, and inexpensive organic solvent-modified DOD tryptic digestion method, over a relatively wide range of targeted proteins and with complex samples. Results were compared to several aqueous-based solvent tryptic digestion methods currently in use. Experimental data indicated that the new organic solvent-modified DOD tryptic digestion method offers key

performance advantages over conventional aqueous-based protein digestion methods in terms of much shorter digestion times and improved reproducibility, and the new approach generally offers improved sequence coverages as well. As it only uses conventional consumables and widely available hardware, and, in our tests, performs well with cost effective, reagent grade trypsin, it could be easily and inexpensively adopted in laboratories performing targeted protein and proteomic analyses. With its short digest times, capability for identifying relatively high numbers of characteristic tryptic peptides and proteins, consistently relative high sequence coverage capabilities, and good precision relative to the methods to which it was compared, the DOD method could substantially affect not only the analysis of target proteins of interest but could also be used to great advantage in general proteomics and in the rapid identification of proteins in complex biomedical samples.

As described, we successfully optimized and validated the new rapid digestion method for complex samples (*E. coli* lysate, murine ileum protein extracts, and K562) with proteolytic digestion times as short as 5 minutes. The DOD method was much simpler and substantially reduced digestion time as compared to traditional methods that require hours to more than a day to sufficiently digest targeted proteins. Importantly, in most cases, the DOD method also provided equivalent or better sequence coverages as compared to other methods using a single protease, provided access to a somewhat different area of the proteome, and, importantly, provided substantially better precision across results.

The data indicated that for applications, such as clinical and forensic diagnostics, in which the goal is simply the rapid identification or quantification of specific targeted proteins, the DOD 5-minute digestion results would likely be sufficient, and the marked reduction in digestion time could lead to much earlier etiological agent identification. This would be very advantageous especially in cases where early diagnosis is essential for successful clinical outcomes or force/general population protection response.

For bottom-up proteomics, in which the goal is the full characterization of component proteins in complex biomedical samples, use of longer digestion times up to 90 minutes or more may be required to optimize method performance. Across the complex biomedical samples subjected to proteomic analysis, especially using digestion times of up to 90-minutes, DOD method performance was generally better in terms of number of proteins identified, the sequence coverages for said proteins, and the number of characteristic peptides produced. In addition, there was improved protein and peptide LC-MSMS peak area precision providing higher confidence in the results as compared to that of the aqueous-based methods we investigated.

DOD digestion, using a relatively high organic solvent content, also tended to digest slightly more hydrophilic proteins and produce slightly more hydrophilic peptides as compared to aqueous-based tryptic digestion methods. This resulted in the identification of a different subset of proteins than the aqueous-based digestion methods, when digesting complex samples, though the effect is relatively minor. The production and identification of more hydrophilic peptides could indicate that in an organic environment and/or a higher temperature environment, different areas of the protein had been accessed.

For complex biomedical samples, if the goal is to identify and fully characterize each individual protein and identify the maximum number of proteins within the proteome, use of the DOD tryptic digestion method in parallel with an aqueous-based method would likely result in the greatest number of proteins identified per sample along with better sequence coverages for those proteins than using any single proteolytic digestion method.

Since we limited our experiments to using the previously published protease-to-sample ratio, a future area to explore would be to increase the amount of trypsin used per sample with complex samples to determine if digestion efficiency was limited due to protease saturation. In addition, digestion pH could also be manipulated to determine whether resulting target protein tertiary structure changes would change the access of the protease to different protein domains.

This could result in changes to the protein profile and sequence coverage changes for the proteins identified.

4.7 References

1. Molecular Biology of the Cell. 4th edition. B. Alberts, A. Johnson, J. Lewis, New York: Garland Science; 2002.
2. J.R. Freitas, A. Robert. (1999). Nanomedicine. *Landes Bioscience*. Tables 3–1 & 3–2. ISBN 1-57059-680-8.
3. D. Voet, J.G. Voet. (2004). Biochemistry, Vol 1, 3rd ed. Wiley: Hoboken, NJ.
4. The Cell: A Molecular Approach. 2nd Ed, G.M. Cooper (editor), Sinauer Associates: 2000.
5. Harnimarta Deol, Ava Raeisbahrami, Phuoc H.T. Ngo, Jagannath Swaminathan, Ophelia Papoulas, Edward M. Marcotte, Eric V. Ansly. After 75 Years, an Alternative to Edman Degradation: A Mechanistic and Efficiency Study of a Base-Induced Method for N-Terminal Peptide Sequencing. *J. Am. Chem. Soc.*, 2025, Vol 147, Issue 16, pp. 13973–13982.
6. J.B. Fenn, M. Mann, C.K. Meng, S.F. Wong, and C.M. Whitehouse. Electrospray Ionization for Mass Spectrometry of Large Biomolecules, *Science,* 6 Oct 1989, Vol. 246, Issue 4926, pp. 64-71, DOI: 10.1126/science.2675315.
7. M. Karas, D. Bachmann, F. Hillenkamp. Influence of the Wavelength in High-Irradiance Ultraviolet Laser Desorption Mass Spectrometry of Organic Molecules. *Anal Chem,* 1985, Vol 57, Issue 14, pp. 2935-2939. *doi*:*10.1021/ac00291a042.*
8. K. Tanaka, H. Waki, U. Ido, S. Akita, Y. Yoshida, T. Yoshida, T. Matsuo. Protein and Polymer Analyses Up to *m/z* 100000 by Laser Ionization Time-of-Flight Mass Spectrometry. *Rapid Commun Mass Spectrom,* 1988, Vol 2, Issue 8, pp. 151-153.
9. A. Armirotti, G. Damonte. Achievements and Perspective of Top-Down Proteomics, *Proteomics*, 2010 Oct, Vol 10, Issue 20, pp. 3566-76.

10. B. Boqdanov, R.D. Smith. Proteomics by FTICR Mass Spectrometry: Top Down and Bottom Up. *Mass Spectrom Rev*, 2005 Mar-Apr; Vol 24, Issue 2, pp.168-200.
11. E. J. Dupree, M. Jayathirtha, H. Yorkey, M. Mihasan, B.A. Petre, and C.C. Darie. A Critical Review of Bottom-Up Proteomics: The Good, the Bad, and the Future of This Field. *Proteomes*. 2020 Sep; Vol 8, Issue 3, p. 14.
12. N. Nagaraj, J.R. Wisniewski, T. Geiger, J. Cox, M. Kircher, J. Kelso, S. Pääbo, M. Mann. Deep Proteome and Transcriptome Mapping of a Human Cancer Cell Line. *Mol Syst Biol.*, 8 Nov 2011, Vol 7, p. 548.
13. T.A. Addona, S.E. Abbatiello, B. Schilling. Multi-Site Assessment of the Precision and Reproducibility of Multiple Reaction Monitoring Based Measurements of Proteins in Plasma. *Nature Biotechnol*, 2009; Vol 27, pp. 633–641.
14. Beatrice Muriithi, Samantha Ippoliti, Abraham Finny, Balasubrahmanyam Addepalli, Matthew Lauber. Clean and Complete Protein Digestion with an Autolysis Resistant Trypsin for Peptide Mapping. *J. Proteome Res.*, 2024, Vol 23, Issue 11, pp. 5221–5228.
15. Craig R., Cortens J.P., Beavis R.C. Open Source System for Analyzing, Validating, and Storing Protein Identification Data. *J Proteome Res.*, 2004, Vol 3, pp. 1234–1242. doi: 10.1021/pr049882h.
16. Deutsch E.W., Lam H., Aebersold R. PeptideAtlas: A Resource for Target Selection for Emerging Targeted Proteomics Workflows. EMBO Rep. 2008, Vol 9, pp. 429–434. doi: 10.1038/embor.2008.56.
17. Martens L., Hermjakob H., Jones P., Adamski M., Taylor C., and States D. PRIDE: the Proteomics Identifications Database. *Proteomics*. 2005 Vol 5, pp. 3537–3545. doi: 10.1002/pmic.200401303.
18. Ji L., Barrett T., Ayanbule O., Troup D.B., Rudnev D., and Muertter R.N. NCBI Peptidome: a New Repository for Mass Spectrometry Proteomics Data. *Nucleic Acids Res.*, 2010, Vol 38, pp. D731–D735. doi: 10.1093/nar/gkp1047.

19. A.L Fink and B. Painter. Thermodynamics of Protein Denaturation. III. Denaturation of Ribonuclease in Water and in Aqueous Urea and Aqueous Ethanol Mixtures. *Biochem.*, 1987, Vol 26, pp. 1665-1671.
20. K.G. Welinder. Generation of Peptides Suitable for Sequence Analysis by Proteolytic Cleavage in Reversed-phase High-performance Liquid Chromatography Solvents. *Anal Biochem*, 1988, Vol 174, pp. 54-64.
21. Z.Y. Park, D.H Russell. Thermal Denaturation: a Useful Technique in Peptide Mass Mapping. *Anal Chem*, 2000, Vol 72, pp. 2667-2670.
22. Z.Y. Park, D.H Russell. Identification of Individual Proteins in Complex Protein Mixtures by High-Resolution, High-Mass-Accuracy MALDI TOF-Mass Spectrometry Analysis of In-Solution Thermal Denaturation/Enzymatic Digestion. *Anal Chem,* 2001, Vol 73, Issue 11, pp. 2558–2564.
23. W.K. Russell, Z.Y. Park, D.H Russell. Proteolysis in Mixed Organic-Aqueous Solvent Systems: Applications for Peptide Mass Mapping Using Mass Spectrometry. *Anal Chem*, 2001, Vol 73, pp. 2682-2685.
24. J.L. Proc, M.A. Kuzyk, D.B. Hardie, J. Yang, D.S. Smith, A.M. Jackson, C.E. Parker, and C.H. Borchers. A Quantitative Study of the Effects of Chaotropic Agents, Surfactants, and Solvents on the Digestion Efficiency of Human Plasma Proteins by Trypsin. *J Proteome Res*, 2010, Vol 9, Issue 10, pp. 5422-5437.
25. Lin Hua , Teck Yew Low, and Siu Kwan Sze. Microwave-Assisted Specific Chemical Digestion for Rapid Protein Identification. *Proteomics*, 2006 Jan, Vol 6, Issue 2, pp. 586-91, doi: 10.1002/pmic.200500304.
26. Yeoseon Kim, Dabin Lee , Jeongkwon Kim. Effects of Incubation Temperature and Acetonitrile Amount on Microwave-Assisted Tryptic Digestion of Proteins. *Anal Biochem*, 2019 Mar Vol 15, Issue 569, pp. 31-38. doi: 10.1016/j.ab.2019.01.009. Epub 2019 Jan 29.

27. Sergei Saveliev, Mark Bratz, Roman Zubarev, Matt Szapacs, Harshavardhan Budamgunta, and Marjeta Urh. Trypsin/Lys-C Protease Mix for Enhanced Protein Mass Spectrometry Analysis. *Nature Methods,* 2013, Vol 10, pp. i–ii.
28. J.M. Oyler, B.Q. Tran and D.P.A. Kilgour. The Rapid Denaturing Organic Digestion Method for Targeted Protein Identification and Characterization. *Anal Chem*, 17 Mar 2021, Vol 93, Issue 12, pp. 5046-5053.
29. J. Wisniewski and M. Mann. Universal Sample Preparation Method for Proteome Analysis. *Nature Meth*, April 2009, Vol 6, Issue 5, pp. 359-62.
30. V. Shah, M.E. Lassman, Y. Chen, H. Zhou, O.F. Laterza. Achieving Efficient Digestion Faster with Flash Digest: Potential Alternative to Multi-Step Detergent Assisted In-Solution Digestion in Quantitative Proteomics Experiments. *Rapid Commun Mass Spectrom*, 30 Jan 2017, Vol 31, Issue 2, pp. 193- 199.
31. A.N. Hoofnagle, and M.Y. Roth. Improving the Measurement of Serum Thyroglobulin with Mass Spectrometry. *J Clin Endocrin Metab*, April 2013, Vol 98, Issue 4, pp. 1343–1352.
32. C.H. Schein. From Housekeeper to Microsurgeon: the Diagnostic and Therapeutic Potential of Ribonucleases. *Nature Biotech*, June 1997, Vol 15, Issue 6, pp. 529-36.
33. Valentina Bianco, Monika Svecla, Giovanni Battista Vingiani, Dagmar Kolb, Birgit Schwarz, Martin Buerger, Giangiacomo Beretta, Giuseppe Danilo Norata**,** Dagmar Kratky. Regional Differences in the Small Intestinal Proteome of Control Mice and of Mice Lacking Lysosomal Acid Lipase. *J. Proteome Res.,* 2024, Vol 23, Issue 4, pp. 1506–1518.
34. Lloyd M. Smith, Neil L. Kelleher, Michal Linial, David Goodlett, Pat Langridge-Smith, Young Ah Goo, George Safford, Leo Bonilla*, George Kruppa, Roman Zubarev, Jon Rontree, Julia Chamot-Rooke, John Garavelli, Albert Heck, Joseph Loo, Deborah Penque, Martin Hornshaw, Christopher Hendrickson, Ljiljana Pasa-Tolic, Christoph Borchers, Dominic Chan, Nicholas Young, Jeffrey Agar, Christophe Masselon, Michael Gross, Fred McLafferty, Yury Tsybin, Ying Ge, Ian Sanders, James Langridge, Julian

Whitelegge, and Alan Marshall. Proteoform: A Single Term Describing Protein Complexity. *Nature Methods,* 27 Feb 2013, Vol 10, pp.186–187.

Chapter 5

5.0 Thesis Discussion and Conclusion

The work described in this thesis illustrates the diverse analytical capabilities of LC-MSMS for the analysis of biomedical samples across a variety of very diverse applications. In chapter 2, the first series of experiments illustrated the utility of LC interfaced with a nominal mass accuracy/resolution QqQ instrument for the very sensitive targeted quantitative analysis of a highly potent/toxic drug, carfentanil, and an experimental biomarker, caffeine, in a simple matrix collected from a live *in vitro* human epidermal model. The model was constructed as a clinically derived alternative to animal pharmacokinetics testing models to evaluate percutaneous absorption of drugs. The model was used in our studies to evaluate percutaneous absorption of carfentanil and the effect of organic solvents on its *in vitro* human pharmacokinetics. The second set of experiments was described in chapter 3 and described the successful validation of the DOD tryptic digestion method, a very rapid and simple tryptic digestion method, used to fragment proteins reproducibly for bottom-up nanoLC/high resolution (HR) MSMS analysis. A list of protein toxins was successfully identified and characterized very specifically using nanoLC-HRMSMS orbitrap analysis. The third results chapter, chapter 4 documented the successful use of the rapid DOD digestion method to perform proteomic analyses of several complex biomedical matrices to identify, with a high level of specificity and reproducibility, the proteins in each biomatrix. Two nanoLC-HRMSMS methods, orbitrap and HR QqQ/TOF methods were used to successfully perform bottom-up analysis with a high degree of specificity and sensitivity. The 5-minute DOD bottom-up experimental protocol was compared to 18-24 h protocols currently used in many laboratories.

Since the goal of the LC-QqQ experiments was the quantification of only two small molecule drug targets, the rapid scanning speed of QqQ instruments when scanning over such a limited m/z range provided a sensitivity advantage. Also, there was no need for the specificity of HRMSMS instrumentation since the use of authentic standard homologs to construct calibration

curves and the use of internal standards in the analyses provided additional specificity; furthermore, greater specificity was not required because the sample matrix contained few potential inteRFerents. The ability to compare the LC retention times and mass spectra of authentic standards to that of experimental sample peaks provided the specificity to confirm not only the presence or absence of carfentanil and caffeine in the experimental samples but also, to quantify with great accuracy and precision the amount of the target analytes in the samples. These data could then be used with high confidence to validate the live human epidermal model using the caffeine data while also describing the effect of different solvents on the percutaneous absorption kinetics of the highly potent opioid, carfentanil, in the model. Because of the great concern in the medical and law enforcement communities concerning the clinical toxicity of carfentanil and the risk of percutaneous exposure to these personnel, this information can be used along with other information to inform policy on the use of organic solvent-containing hand sanitizers by first responders in situations where there is high risk of percutaneous exposure to carfentanil but where the use of the sanitizers for other medical reasons is warranted.

The quantitative method validated for the *in vitro* study, met recognized forensic analytical requirements for specificity and sensitivity[1,2]. The appropriate negative and positive controls were employed to provide the required specificity and to provide the precision required to ensure confidence in the quantitative results. Calibration curves for both target analytes were constructed and met all forensic sensitivity, specificity, accuracy, and precision requirements. Specifically, lower limits of detection (LOD) and quantification (LOQ) for the method were established, the validity of the calibration curves over the concentration ranges required to perform the study were established and met mathematical requirements, and calibrators and quantitative controls over the calibration range met daily requirements for accuracy and precision. Ultimately, the method was used to demonstrate that, counterintuitively, the use of organic solvents did not accelerate percutaneous absorption of carfentanil through the epidermis but, instead, slowed it. The results have been published as a peer reviewed manuscript in the

journal, *Toxicology In Vitro*[3], and the model is currently being employed in additional percutaneous pharmacokinetic and toxicologic studies.

In chapter 3, as part of a protocol originally designed to rapidly and with high confidence screen clinical samples for a select list of protein toxins with a high potential for terrorist use, a very rapid 5-minute tryptic digest method was developed to replace the lengthy 18-24 h methods generally in use in many laboratories[4]. When used with MS-based protein identification methods, the commonly used 18-24 h methods were the rate limiting analytical step, expensive, required specialized equipment and supplies, and were also relatively irreproducible. The 5-minute DOD tryptic digestion method proved to not only be extremely rapid and simple, but coupled with nanoLC-orbitrap HRMSMS, proved to be very sensitive, specific, and reproducible. The combination of the nanoLC chromatographic resolution used to separate complex mixtures into their components with the mass resolution and mass accuracy of orbitrap HRMSMS provided the high specificity required for these analyses since the goal was to develop a screening method that could be used without the need for authentic protein toxin standards. The ability of HRMSMS to deconvolute the charge states of tryptic peptide fragments was required so that the characteristic peptide fragments and, in turn, the parent proteins could be identified with high confidence using easily accessed peptide libraries. The rapid, simple, high organic solvent/high temperature digestion DOD method was subsequently used to, very reproducibly, digest TG and RNAse A, two important clinical diagnostic proteins, for nanoLC-HRMSMS analysis and identification; the 5-minute DOD method was again compared to commonly used overnight, aqueous-based digestion methods which lacked sensitivity, specificity, and reproducibility. After the method was published, a personal communication was received from a colleague confirming that the 5-minute DOD method is currently being used successfully by Merck Pharmaceuticals in their high throughput protein discovery pipeline to identify and characterize target proteins[5].

The method proved to not only provide the capability to confidently meet the forensic

requirements for targeted protein identification using a limited number of characteristic peptide fragments but the capability to identify protein amino acid sequences more fully. The combined high resolution of nanoLC chromatography, and the high mass resolution and accuracy capability of the HR orbitrap MSMS method were required to resolve the charge envelopes of the proteins and peptide fragments. This was, in turn, essential for assigning their molecular masses. The DOD digestion method was again compared to aqueous-based tryptic digestion methods routinely used in proteomic analysis for its ability to determine protein sequence coverage. Across the target proteins analyzed which included ricin holotoxin, ricin A-chain, ricin B-chain, botulinum neurotoxin A, and staphylococcus enterotoxin B, and the clinical biomarker proteins, thyroglobulin and ribonuclease A, the 5-minute, 60 % organic/60 ºC DOD tryptic digestion method consistently equaled or outperformed the overnight, aqueous-based tryptic digestion methods with regards to number of characteristic peptides produced, sequence coverages and within-day reproducibility.

These results indicated the 5-minute DOD method could not only be used in forensics, medical diagnostics, and other applications where target protein identification is based on the identification of a limited number of characteristic fragments, but in applications where determining all or most of the protein sequence is important. Biomedically important proteins like exotoxins and other enzymes always exist as a family of proteins where more consistent amino acid sequence exists in the areas of the protein that confer activity, but where functionally silent areas of the protein are often highly substituted. For enzymes, a group of related catalytic enzymes that share a common function mediated by conserved elements dependent on amino acid sequence and related tertiary structure, but which catalyze different overall reactions is called a "mechanistically diverse supeRFamily"[6]. A supeRFamily consisting of enzymes with diverse functions can, in turn, be subdivided into families, where a "family" consists of a group of related enzymes in which its members catalyze the same overall reaction dependent on conserved catalytic elements[7]. For botulinum neurotoxins, exotoxins that are the most clinically

toxic substances known to man with an LD_{50} of 1 ng/kg[8] (assuming an average body weight of 75 kg, 1 kg can kill the entire human race) produced by anaerobic clostridia bacteria, there are eight confirmed serotypes, A-H[8]. Of the eight, five, A, B, E, F, and H, are clinically relevant while the others primarily affect animals[9]. The serotypes differ structurally in their amino acid sequences. They also differ in their endemic etiology and prevalence, so it can be important to quickly identify the specific serotype when BotNT intoxication is suspected. To have a fully validated, simple, and inexpensive digestion method coupled with the high specificity and sensitivity of nanoLC-HRMSMS analysis would provide the ability to rapidly identify the specific serotype without the need for the costly overnight mouse injection immunoassay currently used to confirm serotype. The 5-minute DOD digestion method coupled with nanoLC-HRMSMS has been used previously to characterize a triple amino acid point-mutated BotNT specie which was subsequently used as a transport vehicle for drugs across the blood brain barrier. Results were published in the journal *Scientific Reports*[10]. The method provided sequence coverages for the light and heavy chains of 96% and 98%, respectively which confirmed the point mutations. The DOD method could be used with nanoLC-HRMSMS analysis to characterize other clinically significant protein families. In addition, the method, due to its high specificity, could be used to identify changes to proteins which do not alter their amino acid sequences like post-translational amino acid modifications, but which often impact functional activity.

Finally, the DOD tryptic digestion method along with nanoLC-HRMSMS was used in the third project to perform targeted proteomic analyses on a variety of complex biomedical samples. Proteomic analyses are protein assays designed to rapidly identify the constituent proteins in an organism[11]. This is usually accomplished through the analysis of different biomedical samples originating from biological subsystems. Protein expression can vary greatly according to time and conditions[11]. The individual proteins and their relative and absolute amounts can vary based on many factors. Stage of development, cellular differentiation, cell cycle, carcinogenesis, and

environmental stressors are just a few of the factors that a cell or organism undergoes or is subjected to that can greatly influence the expression of different individual proteins and the relative amounts of those proteins. The goal of proteomic analysis can also be very specific as in the search for a biomarker of disease[12].

Following protein expression, post-translational modifications (PTMs) can occur which can markedly alter protein function. These include phosphorylation, ubiquitination (a type of conjugation), methylation, acetylation, glycosylation (addition of a carbohydrate moiety), oxidation, nitrosylation, and lipidation to name a few. Other PTMs include the cleaving of peptide bonds to remove amino acids from a propeptide resulting in the formation of the mature or functional form of the protein or the removal of an initiator methionine residue. Disulfide bond formation between cysteine residues may also be referred to as a post-translational modification[13]. For instance, after disulfide bonds are formed, the peptide hormone, insulin, is cut twice, and a peptide is removed from the middle of the chain. This produces the mature insulin protein consisting of two polypeptide chains connected by disulfide bonds. Phosphorylation is a common mechanism for controlling enzyme activity and is the most common PTM[14]. Entire laboratories can be devoted to determining the phosphorylation states of proteins in relation to a disease or condition. Some PTMs result from oxidative stress. Carbonylation is an example that labels the protein for degradation[15,16]. Specific amino acid modifications are also biomarkers indicating oxidative damage[17,18].

There are multiple ways to investigate the proteome. Generally, immunoassays using antibodies can be used to identify proteins, proteins can be separated by electrophoresis and detected, or mass spectrometry can be used. Of these, the most poweRFul current analytical tool is HRMSMS coupled with LC-API or matrix assisted laser desorption ionization. Because the resolution of HRMSMS provides the ability to resolve the charge state envelopes of proteins, and HRMSMS can assign molecular weight with such high mass accuracy, all the possible PTMs can be identified with great confidence. The use of LC and nanoLC using ESI provides another level

of specificity that contributes to the ability to separate very complex mixtures of biomolecules. The HRMSMS analysis of intact proteins, called top-down analysis, generally requires very expensive Fourier transform-ion cyclotron resonance (FT-ICR) instruments that are very expensive and costly to maintain. These instruments are generally not available to most laboratories. An alternative and the most commonly used strategy, called bottom-up protein analysis, is the analysis of proteolytic protein digests to identify characteristic peptide fragments. Some limitations for this method include lengthy digestion times, irreproducibility, and poor protein sequence coverage. Trypsin is the most commonly used proteolytic enzyme employed. Most current tryptic digestion methods are aqueous solvent-based methods that require overnight digestion times, specialized equipment, and costly reagents. The DOD tryptic digestion method, a high organic solvent/high temperature tryptic digestion method was developed as a very rapid alternative to current methods. In targeted protein analysis, it exceeded the performance of the overnight digestion methods to which it was compared. In the proteomic analysis of complex biomedical samples, it also provided equal or better performance than the overnight methods especially when the digestion time was increased from 5 to 90 minutes for specific biomatrices. Results showed that it did cleave some proteins at different sites likely due to the change in protein tertiary structure that occurred under the more hydrophobic organic solvent conditions. This resulted in a somewhat different set of characteristic peptide fragments. That said, for the majority of the biomedical samples investigated, the DOD method provided better sequence coverages and increased numbers of characteristic peptides identified. Since the DOD method is very rapid, and requires relatively inexpensive equipment and reagents, it should be a boon to protein discovery and proteomics laboratories as an added important tool in the analytical arsenal.

In the future, the addition of Lys C to the DOD method in a similar fashion to what is being routinely done with aqueous solvents. This should be investigated to see whether the addition of Lys C would improve digestion sequence coverages as it does in aqueous samples

while maintaining the method precisions that were observed with the DOD method.

5.0 Bibliography

1. Methods of Analysis/Drug Identification. U.S, Department of Justice, Office of Justice Programs, NCJ Number 218655, *Forensic Science Communications*, January 2005, Forensic Science Communications, N 1, V 7.
2. Lane Harper, Jeff Powell, and Em M. Pijl. An Overview of Forensic Drug Testing Methods and Their Suitability for Harm Reduction Point-of-Care Services. *Harm Reduct J.* 2017, Vol 14: p. 52, doi: 10.1186/s12954-017-0179-5.
3. Emily May Lent, Kathleen J. Maistros, Jonathan M. Oyler. In Vitro Dermal Absorption of Carfentanil. *Toxicology in Vitro*, 2020, Vol 62, pp. 1-6.
4. Jonathan M. Oyler, Bao Q. Tran, and David P. A. Kilgour. Rapid Denaturing Organic Digestion Method for Targeted Protein Identification and Characterization. *Anal. Chem.*, 17 Mar 2021, Vol 93, Issue 12, pp. 5046-5053.
5. Personal correspondence with David Colquhoun, Ph.D., Biologics/Advanced Workflows Specialist, Mid-Atlantic and Tristate Regions, Sciex, david.colquhoun@sciex.com, (443) 201-6166.
6. J.A. Gerlt, P.C. Babbitt: Divergent Evolution of Enzymatic Function: Mechanistically Diverse SupeRFamilies and Functionally Distinct Suprafamilies. *Annu Rev Biochem*. 2001, Vol 70: pp. 209-246. 10.1146/annurev.biochem.70.1.20.
7. Shoshana D Brown, John A Gerlt, Jennifer L Seffernick and Patricia C Babbitt. A Gold Standard Set of Mechanistically Diverse Enzyme SupeRFamilies. *Genome Biology*, Vol 7, Article number: R8 (2006).
8. D.O. Fleming. Biological Safety: principles and practices. Vol. 2000. ASM Press. p. 267.
9. Clostridium, Peptostreptococcus, Bacteroides, and Other Anaerobes. Ryan K.J., Ray C (Eds), *Sherris Medical Microbiology*, 5th ed., Chapter 19, 2010, ISBN 978-0-07-160402-4.

10. Edwin J. Vazquez-Cintron, Phillip H. Beske, Luis Tenezaca, Bao Q. Tran, Jonathan M. Oyler, Elliot J. Glotfelty, Christopher A. Angeles, Aurelia Syngkon, Jean Mukherjee, Suzanne R. Kalb, Philip A. Band, Patrick M. McNutt, Charles B. Shoemaker and Konstantin Ichtchenko. Engineering Botulinum Neurotoxin C1 as a Molecular Vehicle for Intra-Neuronal Drug Delivery. *Scientific Reports*, 21 February 2017, Vol 7:42923, DOI: 10.1038/srep42923.
11. N.L. Anderson, N.G. Anderson. Proteome and Proteomics: New Technologies, New Concepts, and New Words. *Electrophoresis*. August 1998, Vol 19, Issue 11, pp. 1853–1861. doi:10.1002/elps.1150191103. PMID 9740045. S2CID 28933890.
12. J.D. Anderson, H.J. Johansson, C.S. Graham, M. Vesterlund, M.T. Pham, C.S. Bramlett. Comprehensive Proteomic Analysis of Mesenchymal Stem Cell Exosomes Reveals Modulation of Angiogenesis via Nuclear Factor-KappaB Signaling. *Stem Cells*, March 2016, Vol 34, Issue 3, pp. 601–613. doi:10.1002/stem.2298. PMC 5785927. PMID 26782178.
13. P.R. Srinivas, M. Verma, Y. Zhao, S. Srivastava. Proteomics for Cancer Biomarker Discovery". *Clinical Chemistry*. August 2002, Vol 48, Issue 8, pp. 1160–1169. PMID 12142368.
14. G.A. Khoury, R.C. Baliban, C.A. Floudas. Proteome-Wide Post-Translational Modification Statistics: Frequency Analysis and Curation of the Swiss-Prot Database. *Scientific Reports*, September 2011, Vol 1, p. 90, Bibcode:2011NatSR...1E..90K. doi:10.1038/srep00090. PMC 3201773. PMID 22034591.
15. H. Lodish, A. Berk, S.L. Zipursky. Post-Translational Modifications and Quality Control in the Rough ER. *Molecular Cell Biology (4th ed.)*, 2000. New York: W. H. Freeman. ISBN 978-0-7167-3136-8.
16. I. Dalle-Donne, G. Aldini, M. Carini, R. Colombo, R. Rossi, A. Milzani. Protein Carbonylation, Cellular Dysfunction, and Disease Progression". *Journal of Cellular and*

Molecular Medicine. 2006, Vol 10, Issue 2, pp. 389–406, doi:10.1111/j.1582-4934.2006.tb00407.x. PMC 3933129. PMID 16796807.

17. P.A. Grimsrud, H. Xie, T.J. Griffin, D.A. Bernlohr. Oxidative Stress and Covalent Modification of Protein with Bioactive Aldehydes". *Journal of Biological Chemistry*, Aug 2008, Vol 283, Issue 32, pp. 21837-41, doi:10.1074/jbc.R700019200. PMC 2494933. PMID 18445586.
18. E. Gianazza, J. Crawford, I. Miller. "Detecting Oxidative Post-Translational Modifications in Proteins". *Amino Acids*, July 2007, Vol 33, Issue 1, pp. 51-6, doi:10.1007/s00726-006-0410-2. PMID 17021655. S2CID 23819101.