CHAPTER 1: INTRODUCTION

"The protein is becoming the centre of attention again. After all, the word is derived from the Greek, 'Proteios' meaning 'of the first rank' – a position it fully deserves."

Peter James 1997 [1]

This thesis is a culmination of works in three distinct and yet similar areas focussed on method development for proteomics with respect to sample preparation and analysis. Chapter 2 explores the advantages of using surface chemistries attached to MALDI analysis plates; Chapter 3 is focused on the advantages of protein fractionation compared to peptide fractionation being top-down versus bottom-up proteomics with ESI as applied to human blood plasma; and Chapter 4 examines the advantages of spectral counting quantification compared to isotope affinity tag quantification with ESI as applied to temperature stress in rice leaves.

Partial content of this chapter was published in;

Neilson, K. A., Ali, N. A., Muralidharan, S., Mirzaei, M., Mariani, M., Assadourian, G., Lee, A., van Sluyter, S. C., Haynes, P. A., Less label, more free: Approaches in label-free quantitative mass spectrometry. *Proteomics* **2011**, 11, (4), 535-553.

1.1 Proteomic Research

Long before the term 'Biology' was popularised by Jean-Baptiste Lamarck over two hundred years ago – (from Greek, *bios*, "life"; *logia*, "study of") – humankind had eternally strived to understand the inner workings of the natural world in which they live [2, 3]. This journey of inquiry, to seek understanding of nature and life, to observe and to communicate these observations, is difficult to allocate a definable beginning to, as it is to assign a beginning to 'Science' – (from Latin, *sciential*, "knowledge") [4, 5]. Though, if one believes it is an inalienable tenet in all human beings to have an inquisitive characteristic, no matter at whatever level it is expressed, then the study of Science and Biology has been a journey as long as the existence of human awareness and thought.

It can be argued that the majority of significant advancements in biological understanding have come about within the past 300 years. The major reason for this, in my opinion, is the joint progress of complementary and burgeoning scientific disciplines of the likes of Physics, Mathematics, Engineering, Chemistry and Medicine. One of the first and arguably the most significant example of complementary advancement with respect to biology was the invention of the optical (light) microscope. The invention of the optical (light) microscope during the sixteenth century by Zaccharias and Hans Janssen, further enhanced by Robert Hooke, lead to the establishment of cell theory and the discovery of bacteria by Antony van Leeuwenhoek in the seventeenth century [6]. More modern examples of this complementary partnership are shown with the development of the non-optical (electron) microscopes [6, 7], carbon dating [8, 9], nuclear magnetic resonance (NMR/MRI) spectroscopy [10, 11], liquid chromatography [12-15], mass spectrometry [16, 17], polymerase chain reaction (PCR) [18, 19] and X-ray crystallography [20, 21]. It is the advancement of these and many other complementary scientific disciplines that have worked side by side enabling current biological knowledge and the lifestyle our society enjoys today. No greater metaphor for such is the discovery of the double helix structure of deoxyribonucleic acid (DNA) by X-ray crystallography [22] for the fields of biology, biotechnology, pharmacy, medicine, physics, philosophy and many others.

It can be said of the field of Proteomics, that it too has reached its current level of advancement, even its very manifestation, due to the advancement of complementary scientific disciplines. This relatively new and flourishing field of science, proteomics, has emerged over the past ~40 years, ever since the attempt to identify a complete complement of proteins arose with the development of two-dimensional gel electrophoresis (2D-GE) [23-25]. Proteins at an individual level have been studied in some form over the past 200 years [26].

The term Proteome was first introduced by Marc Wilkins and Keith Williams at the Siena 2D Electrophoresis meeting in 1994, to describe the entirety of all proteins encoded in a single genome expressed under distinct conditions [27]. The field of Proteomics is the study of the proteome, which is analogous to the field of Genomics, which is the study of an organism’s genome. Proteomics is a dynamic study that is

both spatially and temporally dependent on both the nature and nurture of the biological entity being analysed.

Proteomics distinguishes itself from other forms of traditional biochemistry and protein analysis primarily due to the desire to identify and quantify the entirety of proteins in a sample at a particular point in time, less so than the isolation of a singular protein from a biological sample. This ability to analyse several hundred to several thousand proteins simultaneously has been made possible by advancements in the analytical disciplines of gene sequencing, mass spectrometry, chromatography and computing [28]. Arguably, the most important innovation was the introduction in the 1980s of soft ionisation mass spectrometry of biomolecules such as proteins, peptides, DNA, RNA and viruses [29, 30].

1.2 Soft Ionisation Mass Spectrometry and Proteomics

Mass spectrometry (MS) began to develop in the early twentieth century and Francis William Aston won the Nobel prize for Chemistry in 1922 for his work on mass spectrometry [31, 32]. Mass spectrometers utilise magnetic and electric fields to exert forces on ions travelling in a vacuum-like atmosphere. By measuring the arc in which they travel under these conditions, a mass over charge ratio (m/z) is calculated for the ion [16]. Even though mass spectrometry is the most accurate analytical mass measurement instrument developed, the use of such an instrument traditionally meant having to work out complex fragmentation patterns that were generated by the destructive ionisation process of the earliest techniques [33]. This meant most analysis was limited to robust chemical compounds or gaseous and heat-volatile samples of simple compounds. However, many thermally labile analytes decompose upon heating, especially biomolecules, such as DNA, RNA, proteins and peptides. Thus, these samples required a softer ionisation technique, where the energy transfer methods (desolvation or desorption) are non-destructive or non-degrading if they are to be analysed intact by a mass spectrometer.

It was not until the introduction of two soft ionisation techniques in the 1980s, Electrospray Ionisation (ESI) and Matrix Assisted Laser Desorption Ionisation (MALDI), that the analysis of biomolecules became possible [29, 30]. Despite

ionisation and desolvation/desorption being a separate process, the term “ionisation method” is commonly used to refer to both, especially for proteomics. The mass spectrometer consists of three main sections, the Ion Source (front-end), the Mass Analyser and the Detector [33]. The different types of ion sources are central to the origins of proteomics and are fundamental to the work presented within Chapter 2.

The term MALDI was first coined by Prof. Franz Hillenkamp and Prof. Michael Karas in 1985 [34]. They were the first to show that small biomolecules could be ionised more effectively when mixed with tryptophan and irradiated with a pulsed laser at 226 nm [29]. However, it was Koichi Tanaka who was able to ionise biomolecules as large as 34 kDa in 1987 by mixing 30 nm cobalt particles in glycerol and irradiating with a nitrogen laser at 337 nm, a technique that he called Soft Laser Desorption (SLD) [35], giving him a share of the 2002 Nobel Prize in chemistry, which overlooked Hillenkamp and Karas [36].

Electrospray ionisation (ESI) dates back to the middle of the twentieth century, introduced by Dole and co-workers in 1968 [37], after the discovery of the Taylor cone by Sir Geoffery Taylor [38], though it was not till 1984 that Fenn and Yamashita adapted it for MS and to its current state for large biomolecules [30, 39]. Prof. John B. Fenn was also awarded the joint Nobel Prize for Chemistry in 2002 [40]. ESI is also a soft ionisation technique that differs from MALDI because the sample does not need to have a matrix added to the sample and the sample does not need to be dried or crystallised. It remains in a liquid aerosol form till it transfers into the secondary and tertiary compartments of the instrument [16]. The MALDI and ESI techniques were revolutionary analytical advancements that can be considered as the ‘quantum leap’ that initiated the field of proteomic research.

The MALDI technique first needs a sample (the analyte) that is mixed with its respective matrix in an acidic volatile solvent and dried to form a co-crystallised substance on the sample presentation surface (sample plate). The sample plate is then placed inside the instrument and the atmospheric pressure is reduced to a vacuum of less than 10^{-5} torr. A laser of particular chemical composition (N_2, Nd, YAG) and wavelength (337, 335, 349 and 226 nm) is fired at the sample, a small explosion occurs and matrix molecules co-mingled with analyte molecules are blown off the surface in a plasma plume [41, 42]. The instrument can then use a high

potential difference (delayed extraction), electric or magnetic fields to direct the moving ions into the secondary part of the instrument away from the primary explosion site [16, 33].

The current technique of MALDI includes a multitude of potential matrix molecules and mixtures that are typically conjugated aromatic ring structures with optimal laser wavelengths for irradiation [16, 43]. This is helpful when developing a commercial analytical method that is looking for only one or a few specific molecules, though in exploratory proteomics research the investigator is mainly looking for unknowns or a way to retrieve as much information from the sample as possible. Hence, there now is a short list of standard matrix compounds initially used by the typical proteomics investigator, with 2,5-dihydroxy benzoic acid (DHB) and α-cyano-4-hydroxy-cinnamic acid (CHCA) for peptides being the first choices [43-45]. The chemical structure of several of the most popular matrices are depicted in Figure 1.1.

The specific physics and chemistry behind how the technique of MALDI works is still not completely understood, simply because it is difficult to run experiments on these molecules in a such harsh environment of high vacuum, molecular speed, laser energy, micro-explosions, magnetic and voltage fields [46-48]. It is generally acknowledged that when the laser is fired at the sample plate, where the matrix and analyte are crystallised together, that a thermodynamic cascade reaction occurs [41]. Firstly, the laser beam must hit a matrix crystal. An explosion occurs to blow the analyte and matrix molecules off the surface of the sample plate, creating a plasma plume, either simultaneously with or preceding the next two steps. The matrix molecule absorbs energy from the laser or plasma plume in the form of ionisation, establishing a charge on the matrix molecule [41]. Then a transfer of energy from the matrix molecule to the analyte molecule occurs, known as desorption, taking place through the transfer of charge, thus ionising the analyte molecule and protecting if from the potentially destructive energy of the laser [42]. Figure 1.2 shows a schematic representation of the MALDI process.

Generally, ions observed are $[M+H]^+$ which has an added proton, $[M+Na]^+$ which has an added sodium ion and $[M-H]^-$ with a proton removed. MALDI primarily creates singly charged ions $[M+H]^+$, but multiply charged ions $[M+nH]^{n+}$, can also be created to a lesser extent. This is another distinguishing feature between MALDI and ESI, for

ESI mainly produces multiply charged analyte ions $[M+nH]^{n+}$, where MALDI does not [16]. What is found to be fundamental in creating these differences in charge are the different ionisation techniques in conjunction with the respective experimental settings for laser intensity, solvent, voltage and the type of matrix employed [33].

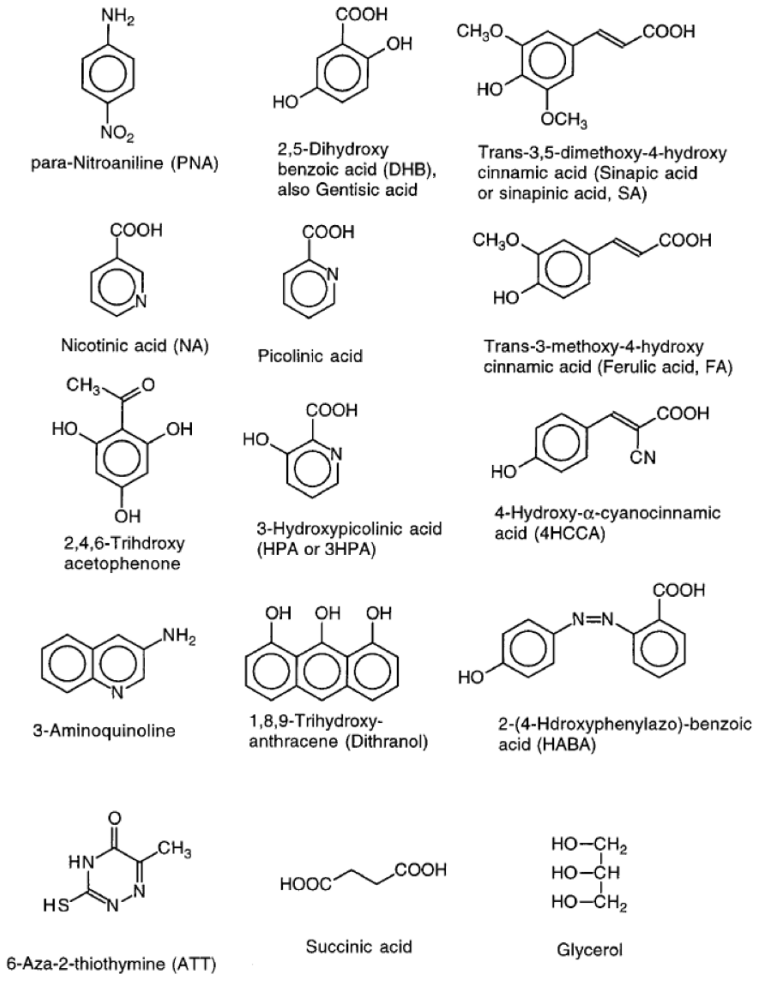


Figure 1.1 Structures, chemical names, trivial names and abbreviations of frequently used MALDI matrices. Figure adapted from R. Zenobi *et al.* 1998, *Mass Spectrometry Reviews* Vol. 17, Page 337-66 [43].

MALDI has an intriguing difference to ESI in that the sample in MALDI is effectively frozen, suspended in time, when it is crystallised on the sample plate. This enables the investigator to go back to a specific sample spot after the primary analysis and conduct secondary analysis that can be more directed towards the unique nature of the analyte detected in the primary analysis, whether this is to increase the sequence coverage of the primary identification, enhance quantification or elucidate post translational modifications (PTM) [49]. Alternatively, with ESI you only get one

chance to collect the information you are looking for and most if not all of the sample is put through the mass spectrometer in that one run. Continuous advancements in instrument duty cycle coupled with improvements in micro- and nano-liquid chromatography have resulted in less of the sample information being missed [50, 51].

In ESI, the analyte is dissolved in a volatile acidic solvent that creates anions and cations and is pushed through a small charged capillary [30, 39, 52]. Desolvation occurs as the liquid exits the capillary. These ions repel and form a mist of small droplets called an aerosol. This aerosol is also partially produced by the ‘Taylor Cone’ [38] and the ‘Jet’ from the tip of the cone that form the final plume with the mist of small droplets of charged analyte within the instrument [33]. Figure 1.3 shows a generalised schematic of the ESI process. Figure 1.4 shows the two current models for how ESI actually occurs, the ion evaporation model and the charge residue model [53].

The electrospray process is the preferred method of choice for the transfer of non-covalent complexes into the gas phase of the instrument without disrupting the inter-molecular interactions of biomolecules [54, 55]. MALDI is able to ionise non-covalent complexes, for example, the calcium-induced tetramers of MRP8 and MRP 14 are also confirmed by ESI MS [56, 57]. However, not all weak interactions can be preserved from the cell to the mass spectrometer with either MS approach [55].

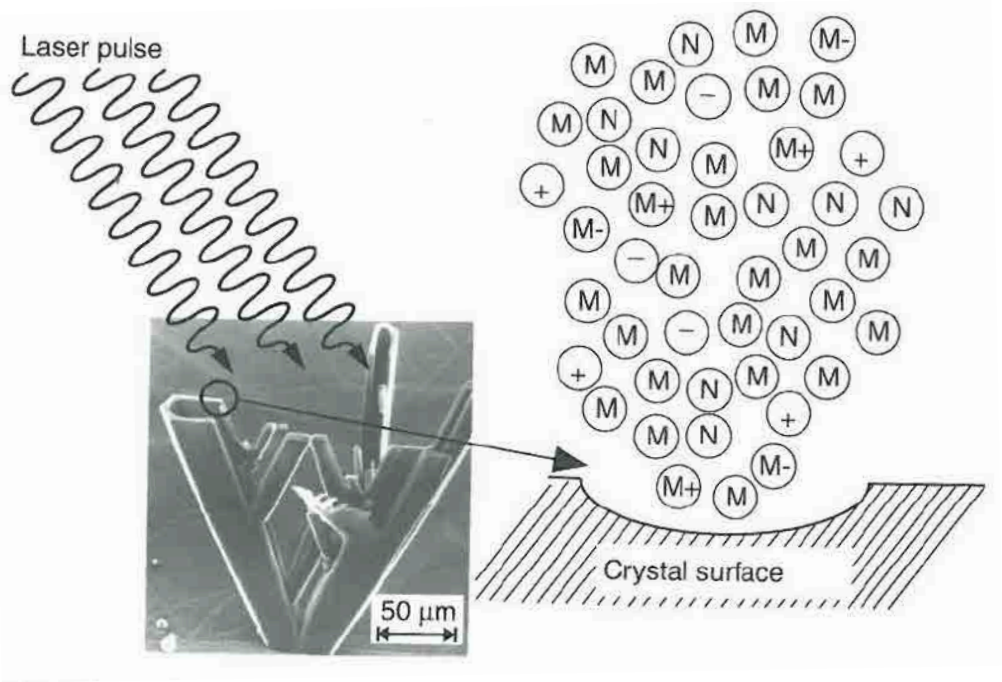


Figure 1.2 Schematic of a matrix-assisted laser desorption/ionization (MALDI) event. The SEM micrograph depicts sinapinic acid-equine myoglobin crystal from a sample prepared according to the dried drop sample preparation method. In the desorption event neutral matrix molecules (M), positive matrix ions (M+), negative matrix ions (M-), neutral analyte molecules (N), positive analyte ion (+), and negative analyte ions (-) are created and/or transferred to the gas phase. Figure adapted from A. Westerman-Brinkmalm and G. Brinkmalm 2002. In Mass Spectrometry and Hyphenated Techniques in Neuropeptide Research, publishers John Wiley and Sons, New York [58].

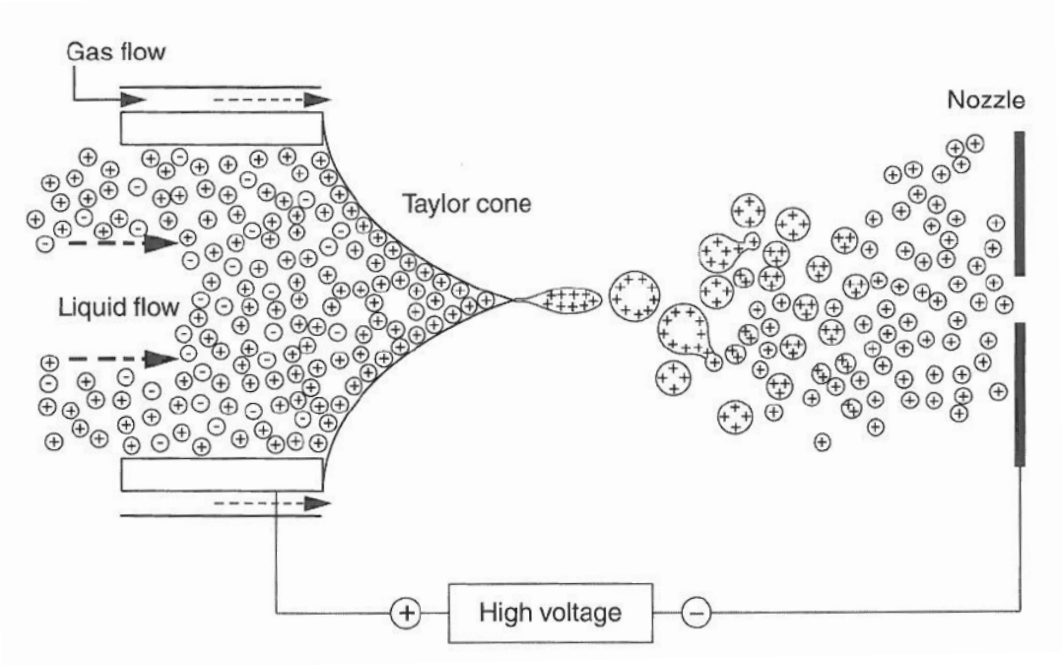


Figure 1.3 Schematic of an electrospray ionization (ESI) source. Figure adapted from A. Westerman-Brinkmalm and G. Brinkmalm 2002. In Mass Spectrometry and Hyphenated Techniques in Neuropeptide Research, publishers John Wiley and Sons, New York [58].

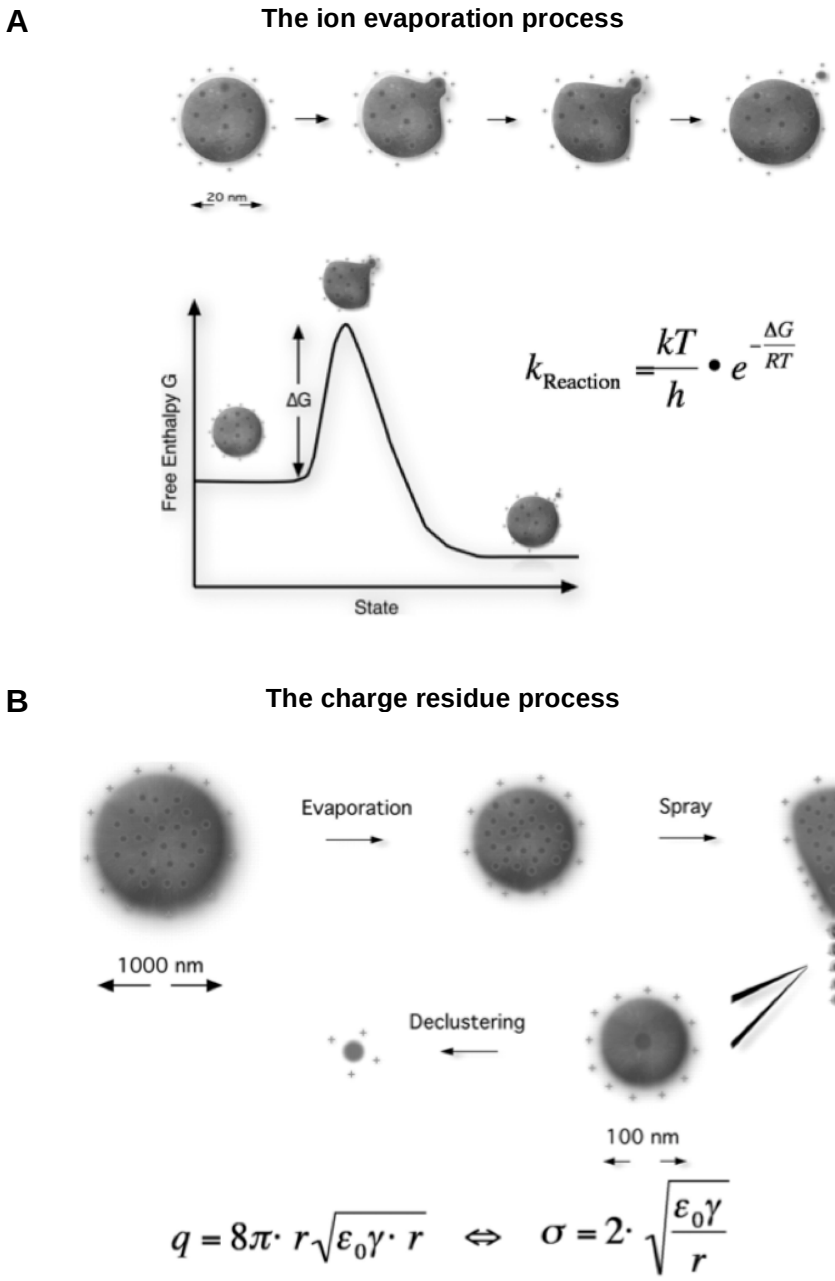


Figure 1.4 (A) The ion evaporation process. An individual ion leaves the charged droplet in a solvated state. The electric field strength at the surface of a droplet is so high that the energy required to increase the droplet surface is rapidly compensated by the gain because of Coulombic repulsion. $k_{Reaction}$, reaction rate constant; k, Boltzmann constant; T, temperature; h, Planck's constant; R, ideal gas constant. (B) The charge residue process. A highly charged droplet shrinks by solvent evaporation until the field strength at the location with the highest surface curvature is so large that a Taylor Cone forms. From the tip of the Taylor Cone, other highly charged smaller droplets are emitted. This process can repeat itself until droplets are formed that contain only one analyte molecule. This molecule is released as an ion by solvent evaporation and declustering. The equation describes the maximum charge a droplet can carry before the Coulomb repulsion overcomes the surface tension. Locally, it is the condition for the formation of a Taylor Cone. q, droplet charge at the Rayleigh instability limit; r, droplet radius; ε_0, electric permittivity of the surrounding medium; γ, surface tension; σ, surface charge density. Figure adapted from M. Wilm 2011, *Molecular and Cellular Proteomics* Vol.10, Issue 7 [53].

1.2.2 Surface Chemistry (SC) MALDI

One of the aims of this thesis was to assess new types of MALDI as compared to more traditional or established MALDI methods. One of these new forms of MALDI is what I term Surface Chemistry MALDI (SC-MALDI), where the surface of the sample plate is chemically altered. More specifically, the technique used within Chapter 2 is termed Spherically Concentric Surface Chemistry MALDI (SCSC-MALDI). SC-MALDI has recently been termed “Lab-On-A-Plate” and a detailed overview of the entire field has been reviewed by Urban *et al.*; the current status of the particular SC-MALDI technique used within has been published by Navare *et al.* [59, 60].

Traditionally, MALDI experiments are conducted on sample plates that are smooth flat metal surfaces without any functional chemistry attached or exposed to the analyte on the surface. This enables the sample plates to be used multiple times and play no role in the purification or concentration of the sample. In contrast, SC-MALDI has unique functional chemistry attached to the surface of the sample plate, which comes into contact with either the analyte alone or the analyte mixed with matrix. Either way, the embodiment of SC-MALDI is to devise specific surface chemistries, that when placed on the sample plate, will interact with the sample so as to achieve one of, or any combination of the following: affinity capture, movement, concentration, purification, reduced sample loss, increased sensitivity of detection or enhanced ionisation efficiency.

The AnchorChip^{TM} was one of first and perhaps the most well known patented system of SC-MALDI, with multiple patents for the company Bruker Daltonik [61]. Wells are composed primarily of two functional group chemistries, with a hydrophilic central concentric region and more hydrophobic outer region, so as to contain larger volumes of analyte than normal flat smooth metal surfaces [59]. It also allowed the concentration of the analyte by standard evaporative processes on the MALDI sample plate into the central concentric region defined by the hydrophilic chemistry.
Use of on-plate surface affinity chromatographic approaches is another SC-MALDI approach. The best known of these is termed surface enhanced laser desorption ionisation (SELDI^{TM}) [62]. In this method, the sample presenting surface plays an active role in the purification and extraction of the analyte whilst allowing for desorption and ionisation from the surface; it does not concentrate the sample as the

AnchorChip™ does [63]. In theory, the surface can be chemically manipulated to allow for designed surface chemistries for the selective retention of proteins/peptides with either broad (RP and polyclonal antibodies) or highly specific interaction chemistries (IMAC and monoclonal antibodies) [64]. Alternatively, it can be utilised through multiple additions, binding and washing steps to concentrate a high volume sample through the affinity capture process. However, this is time consuming and not the standard practice. It has been demonstrated that there is inefficient ionisation of strongly surface-bound proteins in MALDI experiments [65, 66]. It follows, therefore, that efficient decoupling mechanisms are critical prior to ionisation so as to overcome this inadequacy in the SELDI™ approach.

Chips or biochips are terms used in this field to describe biomolecular interactions on surfaces or MALDI plates with surface chemistries on them. Pre-spotted matrix chips offer uniformity of the matrix crystals and produce finer crystals than are generally produced in the lab by the other methods mentioned above [67]. It is inferred that the uniformity and size of the crystals in these pre-spotted chips are the largest factor contributing towards the increased ionisation efficiency and increased limit of detection (LOD) rather than the increased sample confinement abilities of these pre-spotted surfaces compared to standard MALDI plates [67]. Additionally, in traditional LC-MALDI the solvent concentration and crystal formation will change across the chromatographic gradient as the solvent and analyte concentrations change resulting in differential crystallisation and ionisation in the MS. Implementation of these pre-spotted matrix chips for LC-MALDI should eliminate this variability.

1.2.3 Spherical Concentric Surface Chemistry (SCSC) MALDI

The technology used in Chapter 2 of this thesis is referred to as Spherical Concentric Surface Chemistry (SCSC-MALDI), or more loosely as the Biochip, based on the patterning of three or more unique self-assembled monolayers (SAMs) on flat metal surfaces. The monolayer patterning exploits differential surface tensions created by the exposed functional groups on the flat two-dimensional gold covered surface. This method provides a greater working volume through enhanced sample containment and concentration of the analyte with matrix into a central analysis zone that is analyte binding resistant. This type of concentration event is a co-elution of both analyte and matrix under mild acid conditions, combined with a surface that is analyte binding resistant, and is unique to an SC-MALDI plate. The concentration event moves at a rate that is more analogous to the washing, or tidal flowing of solvent, analyte and matrix to the centre, rather than the normal evaporative crawl found in other concentration chip methods.

The Mass•Spec•Focus Chip^{TM} (X3) is the simplest embodiment of the technique described here and depicted in Figure 1.5. Each zone is created by the juxtaposition of distinct surface chemistries on a flat surface in a discrete concentric region. The central zone is the ‘analysis zone’; it is the most hydrophilic and is resistant to analyte binding. A less hydrophilic ‘liquid retention zone’ surrounds the analysis zone. The liquid retention zone is composed of differing chemistries that can be either resistant to analyte binding or provide a functionalised surface for affinity capture. The ‘boundary zone’ is hydrophobic and analyte binding resistant encircling the liquid retention zone. This unique architecture, which is distinct from SELDI^{TM} or the AnchorChip^{TM}, allows for selective screening of peptides in the liquid retention zone, as well as the ability for deposition of larger volumes of sample. In addition, the different zones allow for the subsequent concentration of the analyte with little or no sample loss during the process.

The SAM can be assembled on either silicon or metallic surfaces. Typically, metal plates coated with thin gold surface are used [68-75]. The SAM is bound to the gold surface through the thiol functional group as depicted in Figure 1.6. The gold thiolate bond and self-assembly process orientates the alkylthiol such that the distal end is away from the surface forming a monolayer [72, 76]. The SAMs used in this work

were patterned into concentric circles by photo-oxidation of the SAM by exposure to ultraviolet (UV) irradiation through a stencil mask to obtain the desired geometry. A replacement of the SAM in the ablated area was performed followed by a repeat of the photo-oxidation of the SAM with a different shaped stencil mask [77]. The functional groups of the three SAM regions result in the liquid droplet being exposed to the differing surface energies. Similar SAMs are of interest in the materials sciences due to their wettability [72, 78], frictional [79], adhesional [80] and corrosion resistant properties [81, 82]. Implementation of methods that modify the SAMs after their formation will be important for biological and biochemical studies in the future with the development of surfaces that allow for specific selectivity of large ligands and biomolecules found in complex mixtures such as phosphopeptides or His-Tagged proteins to mention a few.

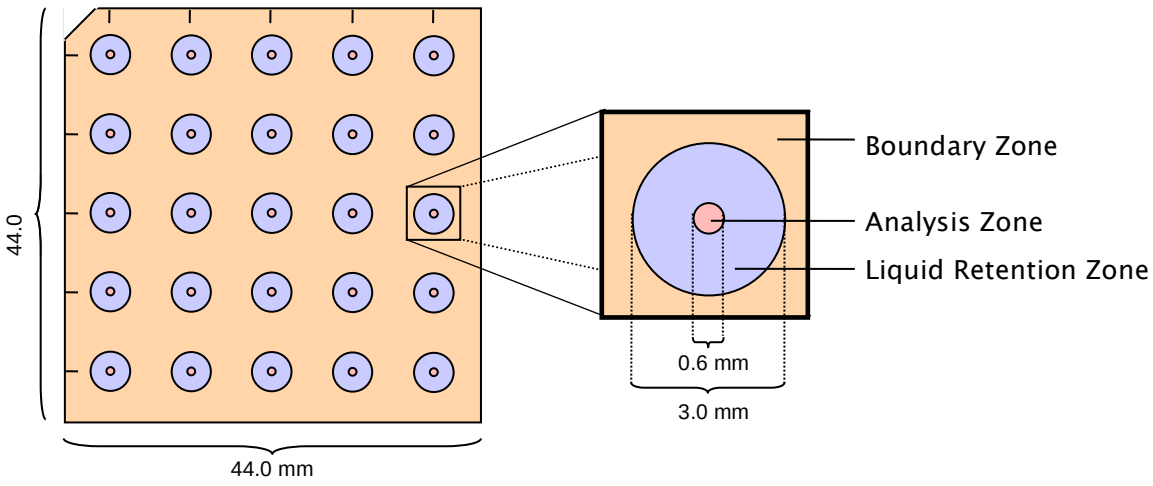


Figure 1.5 Schematic representation of the general outlay and orientation of the concentrating SAM chips (SC-MALDI) used in these experiments. The SAM was patterned onto a gold-coated metal plate to form sample application sites consisting of three distinct regions: Boundary, Liquid Retention and Analysis zones. The boundary zone is a hydrophobic non-binding area, the liquid retention zone is where the liquid sample is applied and contained, and is more hydrophilic then the boundary zone but more hydrophobic than the analysis zone. The analysis zone has the greatest hydrophilic area where the sample and matrix are concentrated and crystallised for MALDI-MS analysis. Figure adapted from Qiagen (Hilden Germany).

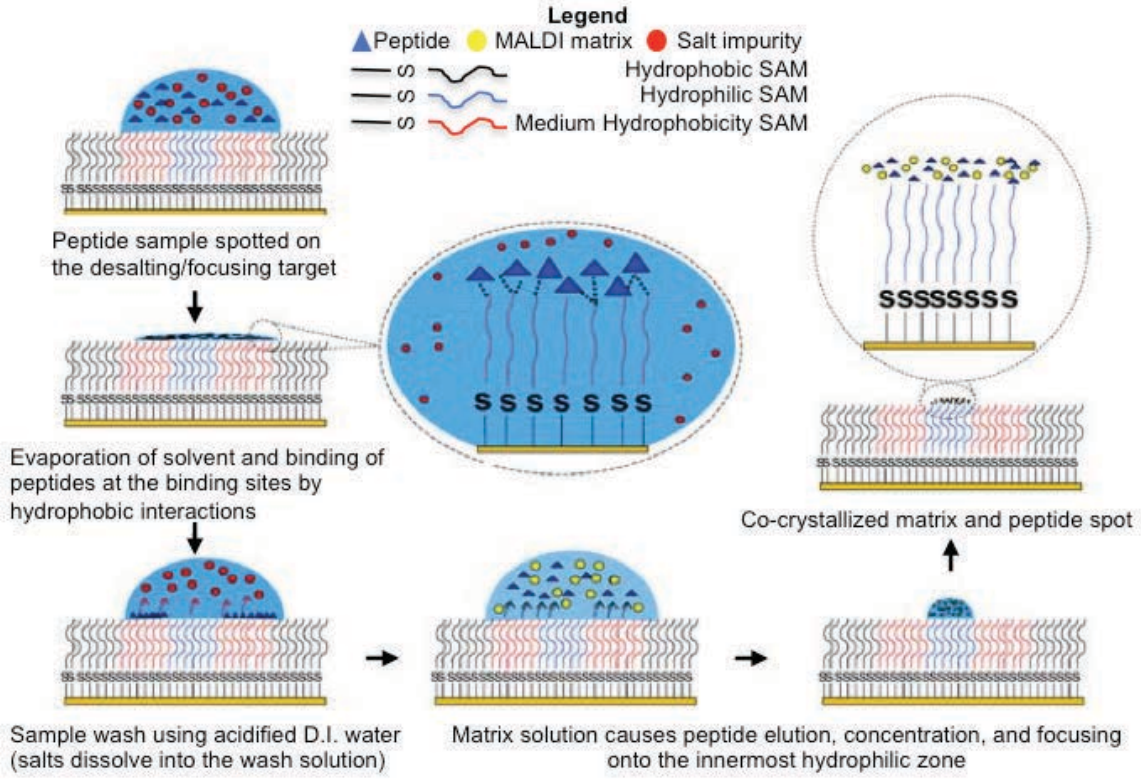


Figure 1.6 Illustration of the SAM on a gold-coated surface. The addition of a liquid sample, the binding of the analyte, the decoupling of the analyte from the liquid retention zone, followed by concentration and crystallisation into the analysis zone. Figure adapted from A. Navare *et al.* 2009, *Rapid Communications in Mass Spectrometry* Vol. 23, Page 477-86 [60].

The washing event is believed to be caused by differential localisation of hydrophilic and hydrophobic surfaces. The liquid retention zone (mixture of exposed carboxyl and methyl groups) has less hydrophilic chemistry compared to the central analysis zone (primarily hydroxyl groups), which is the most hydrophilic. The liquid retention zone effectively traps the liquid into not moving, or crawling into the centre, as the evaporation event begins to occur, until the volume of the liquid is such that the natural force and energy state of the liquid assumes a spherical shape. This is more thermodynamically advantageous than the energy to contain the liquid droplet spread out across the liquid retention zone, in what now looks like a wet surface with a flat droplet that is approximately 70% of the liquid volume relative to what was applied to the to surface. Once this point is reached the liquid floods the analysis zone at a rate that can be difficult to see to the untrained human eye. This motion of the liquid at such a rapid pace is what is referred to as the washing event and distinguishes this technique to any other. This results in a smaller droplet that has now returned to a spherical shape and covers only the analysis zone and proceeds to evaporate under normal conditions.

In essence the ability of this SC-MALDI plate to capture, concentrate and present the analyte in a small analysis zone (500 µm diameter) in the mass spectrometer without the losses associated with conventional methods should afford heightened sensitivity and improved data for biological analysis. This should provide investigators with greater levels of sensitivity of detection while also allowing for the development of numerous analyte capture chemistries. These highly functionalised surfaces within the liquid retention zone could be developed without the reduced ionisation efficiencies observed in other systems that do not separate the affinity capture region from the MS analysis region and keep the analyte bound to the affinity surface, e.g. SELDI^{TM} [65, 66]. There is still debate as to the “killer application” for SC-MALDI [83]. However, the introduction of disposable SC-MALDI that are inexpensive could spark a revival of the once revolutionary concept of disposable cartridges for clinical assays and lab-on-plate MALDI MS as envisaged in the beginning of the development of SC-MALDI [59].

1.2.4 Protein Identification with Mass Spectrometry for Proteomics

In the early implementations of proteomics, an investigator would use two-dimensional gel electrophoresis (2D-GE) to separate the sample into protein spots within the gel matrix, then spots were cut out and digested to give a number of peptides that relate uniquely to the respective protein [84]. MALDI-TOF MS analysis was used to generate a Peptide Mass Fingerprint (PMF), which gave the m/z of each of the intact peptides present, then the mass of each peptide identified from the gel spot were searched as a group against bioinformatics databases to get an identity match against theoretical protein digests [85]. Mass accuracy of earlier instruments and the potential for gel spots to contain more than one protein species limited PMF [86]. A second stage of MS analysis was to utilise the inherent post source decay (PSD) of the instrument and to mass focus on a particular peptide (select or isolate), which prevented other peptides from entering the tertiary section of the instrument [33, 87]. No matter which soft ionisation method is used and no matter what conditions are employed, some fragmentation of the analyte occurs, which is known as PSD [88, 89]. PSD is the inherent and erratic fragmentation (cleavage) of biomolecules. Due to the intrinsic strength of the biomolecules covalent and non-covalent structure (inter- and intra-molecular forces), when placed in high energy

environments (temperature, pressure, acceleration, speed, voltage, electric and magnetic fields) they will degrade and fragment over time [90]. PSD was used more in the formative years of proteomics as a preliminary and then complementary technique to Collision Induced Dissociation (CID) in TOF instruments to help elucidate the sequence structure of the peptides analysed [91]. It is less commonly used now. What was seen was a complex set of peaks representing the PSD fragmentation of the particular peptide into smaller peptides and amino acids. These PSD fragments could be used to piece together, like a jig saw puzzle, the entire peptide sequence from the fragments, adding weight (higher confidence) to the argument of the investigator that the peptide identified was actually the same as the PMF data inferred [88, 92].

Collision induced dissociation (CID), also called collision-activated dissociation, is undertaken when specific molecules, mainly inert gases like He, Ne, or Ar, are brought into contact with the travelling analyte ions producing inelastic collisions; energy is transferred into the analyte ions from these collisions leading to subsequent chemical-bond breaking and reduction of the molecular ion into smaller fragments [93, 94]. These physical contacts that occur result in a more detailed and reproducible fragmentation of a peptide backbone than in PSD. CID is the standard method for inducing fragmentation of a peptide backbone into primarily b- and y-ions for identification [87]. Figure 1.7 and 1.8 depicts the nomenclature for the fragmentation of peptides in a mass spectrometer under CID conditions [95].

Controllable and reproducible fragmentation of peptides using CID was another important leap for the field of proteomics. Measuring ion mass and then mass of fragments resulting from CID of that ion is referred to as MS/MS, to denote the running of two MS experiments on one molecule in one ionisation event through the mass spectrometer. MS/MS is the preferred technique for the identification of peptides and their modifications in proteomics, because the information generated by MS/MS is much richer and reproducible, therefore, more useful for identification when compared to PSD and PMF [96]. There are several other types of MS/MS fragmentation that have been finding an important place within proteomics: electron capture dissociation (ECD) [97], electron transfer dissociation (ETD) [98], infrared multiphoton dissociation (IRMPD) [99], and blackbody infrared radiative dissociation (BIRD) [100].

Finally, there is the concept of MS^3 or MS^n, conducted by ion trapping instruments, which denotes the ability to run greater than one MS/MS on a single biomolecule [101]. For example, phosphopeptides often fragment in CID to produce only one major ion resulting from loss of the phosphate group with little or no fragment ion information generated regarding the peptide backbone sequence. By performing MS^3 on the major ion produced by MS/MS, it is possible to fragment the peptide backbone and hence identify the peptide [102].

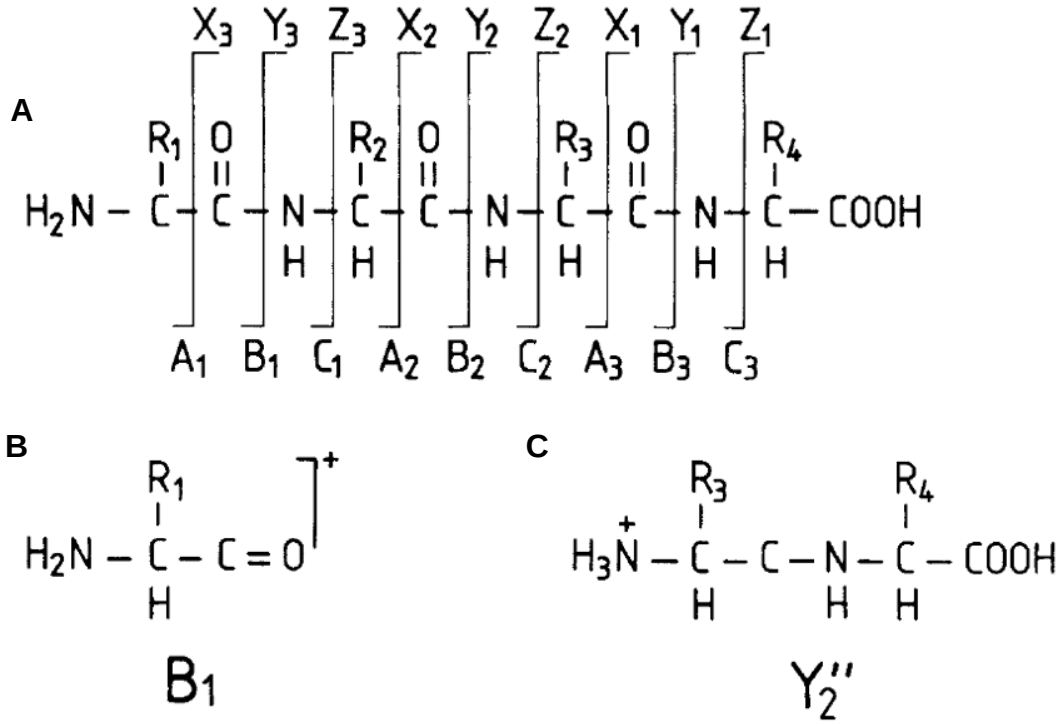


Figure 1.7 (A) The three possible cleavage points of the peptide backbone are called A, B and C when the charge is retained at the N-terminal fragment of the peptide and X, *Y* and Z when the charge is retained by the C-terminal fragment. The numbering indicates which peptide bond is cleaved counting from the N- and C-terminus respectively, and thus also the number of amino acid residues in the fragment ion. The number of hydrogens transferred to or lost from the fragment is indicated with apostrophes to the right and the left of the letter respectively. Thus the acylium ion is named B_n (Scheme B) and the most common C-terminal fragment ion in both fast atom bombardment (FAB) and CI Y_n" (Scheme C). A C-terminal N-C cleaved residue would be 'Z_1 etc. Figure adapted from P. Roepstorff 1984, *Biological Mass Spectrometry* Vol. 11, Page 601 [95].

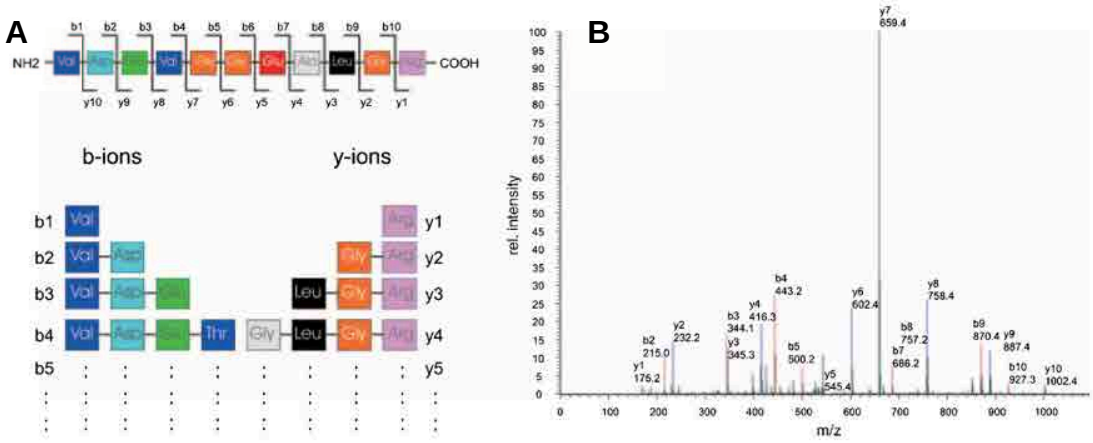


Figure 1.8 (A) Peptide fragmentation: the peptide ions are most likely to break at the peptide bond and produce two major series of fragment ions, called b-ion and y-ions. (B) Example of an interpreted MS/MS spectrum; by comparing real spectra with theoretical spectra. The figure shows a recorded MS/MS spectrum of an 11 amino acid long tryptic peptide, and the annotating of y- and b-ions performed by an MS/MS database search tool (SEQUEST). Figure adapted from T. Fröhlich *et al.* 2006, *Journal of Neural Transmission* Vol. 113, Page 973-974 [103].

1.2.5 Mass Spectrometry Analysers for Proteomics

The secondary part of the instrument, the mass analyser, varies greatly and can be broken up into various sub-sections, magnetic or electric, scanning or non-scanning (pulse based), and trapping or non-trapping. These explain how the instrument purifies the analyte (parent ion selection), while the sample is moving at high speeds, in the vicinity of 300 to 1700 m/s [33]. The acceleration, energy and momentum induced by the electric and magnetic fields are dependent on the charge on the ion and all molecules must be ionised (charged) before they are amenable to MS analysis [16].

The mass analysers can vary greatly with respect to how they work; these variations and advancements in recent years have lead to the development of specific instrumentation for advancement of proteomic analysis [104]. I will divide up the mass analysers into seven major classes for proteomics:

1) time-of-flight (TOF) mass analysers, which separate ions according to the time difference of ions to traverse a defined distance and when the ions hit the detector [105];
2) ion mobility (IM) mass analysers, which separate ions based on their size/charge ratio and interactions with a buffer gas studying ions mobilities in a drift cell [106];

3) magnetic/electric sector mass analysers, which accelerate ions similar to a TOF but passes through a magnetic sector, where the field is applied perpendicular to the direction of the ion beam, inducing a circular motion with radius dependant on the magnetic field strength, mass-to-charge ratio and velocity of the ions [17];
4) quadrupole mass filters, which employ a combination of direct current (DC) and radio frequency (RF) potentials across four parallel rods. It is the relationship between the DC and RF potential and the RF frequency that allow ions of a certain m/z range to pass through the quadrupole [107, 108];
5) quadrupole ion traps (QIT), which traps and stores ions in a potential well; ejection of ions from the potential well is conducted in order of ascending m/z for detection. The cylindrical quadrupole ion trap is the same principle as the quadrupole mass filter with different geometry. A recent variant, the linear quadrupole ion trap is similar to a quadrupole mass filter with extra trapping end electrodes for the axial direction, ideally trapping moving ions indefinitely under stable conditions [109];
6) Orbitrap, which traps ions in a potential well measuring the frequency of the trapped oscillating ions rather than ejecting them like the QIT. It is a pure electrostatic device inspired by the Kingdon trap [110], using a logarithmic electrostatic field between its inner and outer electrodes, as well as a quadrupolar field between the end caps [111, 112]; and
7) Fourier Transform Ion Cyclotron Resonance (FTICR), ion cyclotron resonance (ICR) analysers, which trap ions by strong magnetic fields [113, 114], causing the ions to move in circular motion with a frequency dependent on their m/z, using opposing plates for trapping, excitation, and detection. The analyser takes advantage of the same physical laws as the magnetic sector, except with a 360^{o} arc [115, 116].

The FTICR and Orbitrap analysers outperform any other commonly used mass spectrometer with respect to the maximum mass resolution and accuracy routinely achievable [117]. Prof. Han Georg Dehmelt and Prof. Wolfgang Paul were awarded the 1989 Nobel Prize in Physics for their discovery of the ion trap technique [118, 119].

The third element to the instrument is the detector, which converts the energy of the incoming particles into a signal current that is registered on electronic devices for use by the computer of the mass spectrometers acquisition system. A thorough review on detectors has been published by Koppenaal *et al* [120].

Either of these soft ionisation techniques (MALDI and ESI) are not entirely complementary or ‘stand alone’, as some biomolecules and ion types are instrument and settings limited, exhibiting variable ionisation efficiencies or ion suppression events. This means that any single analysis on a particular mass spectrometer platform will inherently miss something from a complex proteomic sample when compared against another platform [121].

1.3 Top-Down or Bottom-Up Proteomics

The diversity of a proteome is determined by three main processes: firstly, at the deoxyribonucleic acid (DNA) level (i.e. gene polymorphisms), secondly, at precursor messenger ribonucleic acid (pre-mRNA) or messenger ribonucleic acid (mRNA) level (i.e. alternative splicing) and, thirdly, at the protein level with post translational modifications (PTM). Currently proteomics technologies allow identification, characterization and quantitation in the order of thousands of proteins [122, 123].

Today, the field of proteomics by mass spectrometry can be divided up into two main groups, those that do top-down proteomics and those that do bottom-up proteomics [124]. Top-down proteomics works at the protein level for both the fractionation (decomplexation) and mass spectrometric analysis; the term is attributed to Prof. Fred W. McLafferty [125]. Alternatively, bottom-up proteomics employs the digestion of the proteome into peptides with endoproteases for the mass spectrometric analysis of the peptides alone [126]. Recently, middle-down proteomics has been coined for analysis on large peptides or sub-groups of a protein for PTM analysis without the limitations of handling intact mass proteins [127].

It is viewed by many that 1D- and 2D-GE belong to the top-down experimental group, mainly because fractionation is conducted on the proteins and information pertaining to isoforms and species can be derived [126]. However, since the proteins in most

1D- and 2D-GE experiments are digested when extracted from the gel and the mass spectrometric analysis is conducted on these peptides, it could be argued that they belong to the bottom-up experiment group, from a purist mass spectrometry point of view. Alternatively, 1D- and 2D-GE experiments could be considered as a bottom-up sub-group, with the distinguishing factor being the fractionation step in a gel matrix of proteins opposed to today's traditional bottom-up experiments with liquid chromatography conducted on peptides being the other sub-group.

Based on the International Union of Pure and Applied Chemistry (IUPAC) rules, the term "isoform" is used for different proteins derived from genetic variation. The term protein "species" is used more generally to define proteins bearing any chemical modification, including those derived from alternative splicing and post translational modifications (PTMs) [128]. Based on this definition, protein isoforms may be considered as a subset of protein species because they are also chemically different. Two or more proteins derived after alternative splicing and/or bearing one or more PTM represent different protein species; the term protein isoform should be restricted to those proteins where the source of the chemical difference is genetic [129, 130]. Hence, if the source of the chemical variation is unknown, the term protein species should be used.

PTMs are a major mechanism of speciation and protein function [131]. This process is more general than alternative splicing because it occurs not only in eukaryotes, but also in prokaryotes and archaea [132]. More than 1,000 PTMs (December 2010), have been referenced and assembled in UNIMOD (www.unimod.org). PTMs are a dynamic process and many proteins can undergo multiple PTMs at the same time. Thus, PTMs as a source of protein speciation should not be ignored and are a driving force behind the development and advancement of proteomic techniques for clinical applications [126].

1.3.2 Fractionation – Decomplexation

The isolation of a singular protein, free of all other biomolecules, is one of the primary objectives in traditional biochemistry and it has taken over 200 years to move from the view that there was “a single protein” to over a million potential protein species in the human body alone [26]. Proteomics strives towards this utopian goal by employing various separation procedures that are designed to exploit any distinguishing features of the target protein, such as its size, its physio-chemical properties and its binding affinity. However, the enormous size of a proteome precludes the expectation of achieving this singular isolation of proteins. Removing sample complexity through fractionation and enhancing duty cycles so that the mass spectrometer can analyse the proteins as if they were supplied in a singular fashion is the pragmatic goal of proteomics practitioners of today [133, 134].

This difficulty in handling and purifying proteins guided most research in the early stages of the 20th century towards proteins that could be found in large quantities from blood, egg white, various toxins, and enzymes obtained from slaughter houses [26]. These traditional biochemists were isolating and trying to understand the basic structure and function of single proteins or small protein complexes with purification methods such as centrifugation, precipitation, UV spectral measurements, dye-binding, electrophoresis and numerous forms of liquid chromatography. More modern, though traditional, protein analysis methods in biochemistry include Circular Dichroism (CD) [135], Nuclear Magnetic Resonance Spectroscopy (NMR) [10, 11], X-ray Crystallography [20, 21] and Immunohistochemistry [136]. These types of experiments require a sample to be limited in the complexity of the molecular entities present and in the majority of instances to be found in the most singular or purist form, whereas mass spectrometry based proteomics approaches do not demand such stringent restrictions.

The ability to separate proteins based on electrophoretic mobility triggered the characterisation of isoenzymes found in different organisms [137]. Several isoenzymes of human creatine phosphokinase were described, differing in their amino acid composition [138], catalytic constants [139] and electrophoretic mobility [140]. These findings and the introduction of SDS-PAGE for molecular weight determination [141, 142], made electrophoretic mobility the preferred physical

property to study the occurrence of multiple protein forms [26]. The subsequent introduction of two-dimensional electrophoresis in a gel matrix (2D-GE) enabled the separation of multiple proteins based on isoelectric focusing in the first dimension, taking advantage of the proteins isoelectric point (pI), followed by relative molecular weight (M_r) in the second dimension. This allowed the identification and characterisation of different protein species and isoforms [24]. 2D-GE represented a monumental leap for the sample handling and removal of complexity for proteomic samples enabling the study of the proteome to truly begin [23, 143].

Since peptides are more easily solubilised than proteins and have less of a tendency to aggregate out of solution, their handling is preferred, especially with regards to the complex nature and sheer number of proteins a proteomics researcher aims to deal with [144], hence the development and proliferation of the generic Shotgun proteomics experiments of today developed from ‘MudPIT’ (Multidimensional Protein Identification Technology) [145-147]. There is, however, a disconnect in the information generated in a mass spectrometer when working with peptides towards characterising the intact mass of the original protein and the protein species or isoforms present, enabling 1D- and more so 2D-GE to maintain a necessary place within the proteomics laboratories of today [148]. Both techniques have their limitations, with 2D-GE having limited solubility of hydrophobic proteins and low throughput [149, 150], while LC-MS/MS of peptides does not achieve full sequence coverage plus limited information for protein species, isoforms and the intact mass [151, 152]. Alternatively, the use of 1D SDS-PAGE on the proteins as a preliminary fractionation technique and coupled with LC-MS/MS of the peptides has shown great success, though it too suffers from the limitations mentioned above [153, 154]. Currently, top-down suffers from limited sensitivity, inability to work with detergents (SDS), throughput and limited bioinformatics [155]. However, the Kelleher research group has used gel-based and gel-free techniques for protein separation and top-down MS analysis [134, 156]. It is believed that the top-down LC-MS/MS approach can remove most of the shortcomings of both bottom-up and gel electrophoresis analysis, and at the same time enhance the quantity and type of information generated [134].

As an alternative to gel-based experiments, the analysis of protein mixtures utilising bottom-up LC-MS/MS experiments has become popular during the past decade

because it overcame some of the limitations of the 2D-GE technique [157]. LC-MS/MS experiments typically conduct multidimensional liquid chromatography on peptides, employing configurations of strong cation exchange (SCX) followed by C_{18} reverse phase (RP) in-line.

Commonly, the field is termed ‘shotgun proteomics’. Shotgun proteomics can be considered as any multidimensional fractionation of the proteins or peptides in-line or off-line for a proteome sample followed by MS/MS identification based on the peptides, not limited to ESI, also encompassing LC-MALDI. Generally, the fractionation is off-line, having many configurations, the high resolution RP/RP-MS/MS, affinity-capture (AC)/RP-MS/MS or the lower resolution ion exchange (IEX)/RP-MS/MS. In Chapter 3, I employed the digestion of the proteome into peptides and conducted a SCX/RP-LC-MS/MS methodology as the standard comparison experiment [158], comparing it to the system I developed, being SCX/SAX on the whole protein (intact-mass and native form), followed by digestion of each protein LC fraction and RP-LC-MS/MS on the peptides.

To provide a more biocompatible, high-resolution separation of biopolymers, including proteins, Pharmacia LKB (Uppsala, Sweden) developed fast protein liquid chromatography (FPLC) in 1982 [159]. The chromatographic modes can cover ion exchange [160], chromatofocusing [161], gel filtration [162], hydrophobic interaction [163], and reverse phase [164]. Systems have evolved from FPLC to high-pressure liquid chromatography (HPLC) through the introduction of smaller diameter resins (5 - 40-µm) in the columns and higher pressure (up to 400 bar), though they are limited in sample loading and difficult to scale up to industrial scale compared to FPLC [165]. The cost per test can be nearly 30 times cheaper for FPLC compared to HPLC and the costs of the columns for FPLC are 10 times less than HPLC [166]. Reversed-phase chromatography of proteins has been historically difficult because carryover, peak splitting, and broad or misshapen peaks are common [167]. The latest evolution to solve this is with Ultra-performance liquid chromatography (UPLC), using even smaller diameter resins (< 2-µm) at increased pressures and flow rates for enhanced resolution and decreased elution gradient time, with less mobile phase volume than HPLC [168-170].

Ion-exchange chromatography (IEX) is the method of choice I used to fractionate the intact proteins of depleted human plasma in two dimensions. IEX allows for the separation of ionisable molecules on the basis of difference in charge properties [148]. It has large sample handling capacity, broad applicability, moderate cost, powerful resolving ability and ease of scale-up and automation enabling it to become one of the most versatile and widely used of all LC approaches [171]. With origins dating back to the early half of the 20th century, IEX was designed to separate differentially charged or ionisable molecules [172, 173]. IEX has been employed for the purification of proteins [174, 175], enzymes [176, 177], antibodies [178], peptides [179] amino acids, nucleic acids [180, 181], simpler carbohydrates and organic compounds [182, 183]. Similar to most other forms of LC, IEX has both mobile and stationary phases, the former typically an aqueous buffer which the mixture to be resolved is introduced, and the latter usually an inert organic matrix chemically derivatised with ionisable functional groups that carry a displaceable oppositely charged counter ion. These counter ions exist in a state of equilibrium between the mobile and stationary phases, resulting in the two IEX formats of anion- and cation-exchange.

1.3.3 The Protein Inference Problem

Since the affiliation of the peptides to their original protein is lost after endoprotease digestion in most bottom-up experiments, the assignment of protein species and isoforms after the analysis becomes a challenging task [184]. Therein lies one of the insidious problems in proteomics today, the potential disconnect between information generated from analysing peptides so as to understand the proteins in a proteome, better known as the protein inference problem [185].

Simply put, how many peptides and of what type, unique or non-unique, justifies a confident identification or quantification? A myriad of bioinformatic tools have been developed that are able to reconstruct the puzzle of peptide sequences identified and build up a list of candidate proteins present in the sample [186]. This list of proteins is inferred after the comparison of the data acquired in a mass spectrometer against a protein sequence database using one or a combination of search engines, such as Mascot [187], Sequest [188], VEMS [189], X!Tandem [190], OMSSA [191] and many

more. Simplistically, if one or more peptide spectra match a peptide sequence within a protein sequence in the database, then it is inferred that the protein bearing that peptide in the database is present in the sample. The unambiguous identification of a single protein relies on the identification of at least one peptide uniquely found in that protein species. The peptides found only in a certain protein species are termed as “proteotypic”, “unique” or “discriminant” peptides [192]. Conversely, those peptides present in all protein species predicted for a given gene are referred to as “constitutive”. If a peptide is present in more than one but not every species, it is termed as a “semi-constitutive” peptide [192]. The term ‘non-unique’ has been employed when a distinction between constitutive or semi-constitutive is not known or necessary for the purpose of the discussion. Thus, in order to identify one protein species, detection of at least one unique peptide is a prerequisite. Detection of unique peptides in complex mixtures may turn into a difficult task since the peptide(s) may be present at low stoichiometric concentrations, especially since current data dependent acquisition (DDA) with LC-MS/MS is biased towards abundant proteins [93, 193].

Recent advances in off-gel isoelectric focusing (IEF) have enhanced the identification of low concentration species, with the identification of 5,111 proteins from mouse embryonic stem cells [123]. Additionally, combinatorial style methods called filter-aided sample preparation (FASP), combine in-gel and in-solution ideologies, for sample preparation for bottom-up LC-MS/MS. The mammalian cells are solubilised in SDS, retained and concentrated with ultrafiltration. The filter unit conducts detergent removal, buffer exchange, chemical modification and protein digestion. The technique found 7,093 proteins from HeLa cells compared to 3,979 with IEF, representing a crucial enabling sample preparation technology for organelle studies [122]. Despite the impressive number of proteins identified, discrimination between protein species is not included in the two studies above. Graumann *et al.* acknowledged the need for further technological development for a more comprehensive coverage of the proteome and especially for the identification and quantitation of specific protein species [123].

1.3.4 Comparative Solutions

The advancement of mass spectrometry in recent years, with respect to duty cycle, resolution, limit of detection (LOD) and mass accuracy, have arguably been the areas of most advancement for MS proteomics and in particular the field of biomarker discovery. These advances in instrumentation have enabled mass spectrometry proteomics to study complex samples with a resolution of more than 10,000, mass accuracy of less than 3 ppm, femtomole limit of detection and with a dynamic range of 3-4 orders of magnitude [93].

The emerging proteomic strategy, termed “top-down” proteomic analysis as mentioned earlier, applies to the characterisation of intact proteins and protein complexes [124, 194, 195]. The major advantage compared to the bottom-up approach is the removal of the requirement of endoproteases [196]. Thus now, in top-down analysis the affiliation of the peptides to their original proteins is known, which allows the unambiguous assignment of peptides to protein isoforms, including PTMs. The molecular weight ranges of current top down experiments are broadening and the traditional limit of 50 kDa is being pushed upwards, with the characterisation of proteins greater than 200 kDa being completed [197]. The expected standard is still approximately half this and is greatly dependent on sample, instrumentation and practitioner [198]. The ability to conduct top-down experiments becomes more difficult as the complexity of the sample increases and the need to extend top-down approaches to address the proteome of a whole cell is still a challenge [184]. It is believed that the handling of complex mixtures may need one or two orders of magnitude more material than traditional bottom-up experiments and the separation methods of intact proteins compatible with MS still need to be improved [198, 199].

In top-down proteomics, IEX or RP column chromatographic separation of the proteins is achieved in-line to a high-resolution mass spectrometer for LC-MS [200, 201]. After ionisation in the positive mode, multiple charge states for the proteins in the mixture are detected and are acknowledged by the characteristic “crown-shaped” spectra showing the several m/z ion charge states of the protein(s), and the monoisotopic mass is deciphered after deconvolution of the data [202]. However, accurate mass determination of a protein is not sufficient for confident protein identification, though it can be valuable to refer back to this information as it relates to

protein species differentiation. The loss of N-terminal methionine at the protein level is a source of speciation frequently found in eukaryotic organisms and is an example of how the use of top-down MS can help distinguish between these protein species [203]. Additionally, MS/MS fragmentation of at least one charge species and the detection of the fragment ions can be carried out automatically inside the mass spectrometer [194]. Combining the neutral deconvoluted intact monoisotopic mass and the subsequent fragment ions the identity of the protein(s) and quantification in a mixture can be obtained [204]. This process should lead towards the identification of protein species and the specific amino acid residues bearing PTMs [198, 205, 206]. Together with CID MS/MS, novel ways to activate fragment peptides, such as electron capture dissociation (ECD) and electron transfer dissociation (ETD) have emerged as valuable alternatives for analysing proteins rather than using MS alone [207, 208].

It is now recognised that protein fractionation steps prior to top-down MS analysis is difficult and a prerequisite [134], and is a basis for the development of my work. Recently, solution-phase isoelectric focusing (sIEF) has shown promise for top-down proteomics, which is the separation of proteins based on their pI while in solution [209]. A related approach called Gel-eluted liquid fraction entrapment electrophoresis (GELFrEE), which consists of the separation of proteins based on their molecular weight [210]. Both techniques have been used in combination or alone for the analysis of cell lysates [134, 198, 211]. With respect to the more traditional chromatography in top-down experiments, the introduction of polymeric reversed-phase (PLRP) columns outperformed the typical C4 reversed-phase columns used in most analysis [198].

Casado-Vela *et al.* in a recent review this year, stated that the characterisation of whole proteomes using the top-down method is still in the early stages, and as both bottom-up and top-down have proven valuable tools for protein species characterisation, it is likely they will co-evolve in the future [184]. Alternatively, several researchers foresee the two approaches meeting halfway as a hybrid approach in which large fragments or whole domains of proteins are analysed [124, 194, 212]. The approach has been termed “middle-down” [213, 214]. Alternatively, the hybridisation of both techniques to be run in parallel has also been proposed [215].

As the field of proteomics matures, it is being shown that it is not just about generating more protein identities from a sample, it is also about generating information that is more detailed (speciation), of a higher confidence and having a greater relevance to the function and endogenous form types of the proteins within a proteome for greater direct biological or bio-medical relevance [216-218]. My work in Chapter 3 moves towards such goals on human plasma.

1.4 Quantitative Proteomics – To Label or Not To Label

Ever since it has been known that mRNA levels do not relate directly to the level of proteins translated in a cell, implying that there must be additional, unknown post-transcriptional and post-translation mechanisms at play to determine protein amounts that regulate biochemical pathways, the field of proteomics and quantitative proteomics has been required [219, 220]. Proteomics has moved beyond simply protein identification to be now driven to identify accurately and reliably the quantitative differences in protein abundance for biological systems simultaneously [221-223].

It is important to note that the majority of quantitative proteomics is relative quantification and not absolute. The former will be a focus of this thesis. To observers outside the field of proteomic quantification it may seem that absolute quantification is the more important or of higher priority compared to relative quantification. However, to researchers working within this field, relative quantification does supply similar information about the stoichiometric differences between the protein species present within a sample, be that between a control and test subject, or across a time course experiment or other multi-variant experiments, and has shown to correlate well to absolute quantification [224, 225].

Quantitative proteomics can be divided into two major approaches; stable isotope labelling and label-free techniques. Common labelling techniques involve modifying proteins or peptides with isotope-coded affinity tags (ICAT) [226], isobaric tags for relative and absolute quantification (iTRAQ) [227], isotopomer labels, referred to as tandem mass tags (TMTs) [228], or metabolically labelling proteins through N^{15}, C^{13},

or stable isotope labels with amino acids in cell culture (SILAC) [229-232]. Label-free quantification can be divided into two groups: (1) peak area or area under the curve (AUC), is the integration of their signal intensity measurements based on precursor ion spectra [233, 234] and (2) spectral counting (SC), which is based on counting the number of peptides assigned to a protein in an MS/MS experiment [235].

Traditionally, proteomic quantitation utilising dyes, fluorophores, or radioactivity provided good sensitivity, linearity and dynamic range, although they have two shortcomings [236]. The first being they require high-resolution protein separation normally from 2D-GE, which is limited to abundant and soluble proteins, while being labour intensive and time consuming. Secondly, they do not reveal the identity of the underlying protein without MS analysis. With the development of gel-free techniques such as shotgun proteomics, there has been a rapid growth in quantitating proteins at the MS level rather than the laborious visual comparison of gel based methods [237]. In this thesis spectral counting and the labelling technique of iTRAQ are employed in the experiments in Chapter 4 on temperature stressed rice leaf tissue samples.

1.4.2 Label-free - Spectral Counting

Spectral counting relies on the observation that greater abundant peptides will be selected for fragmentation and produce an elevated abundance of MS/MS spectra, and is proportional to protein amount in DDA [235]. The spectral counting method has developed from summing spectra to modifying counts with a normalisation factor and combining strategies for increased accuracy. Lundgren *et al.* in a recent review outlined the many spectral counting strategies and statistical tools for analysing spectral count data [238]. Spectral counting has been modified to take into consideration the length of a protein will affect the number of spectral counts (SpC) such that a longer protein will generate more MS/MS spectra. A normalised spectral abundance factor (NSAF) provides an improved measure for relative abundance by taking into account the length of the protein, which is calculated by dividing the SpC for a protein by its length (L) [239-241]. This value is then normalised by dividing by the sum of all SpC/L for all proteins in the experiment. The dynamic range for NSAF is approximately four orders of magnitude and abundance changes as low as 1.4-fold can be detected [240]. NSAF values have been used in a study that showed spectral count data share similar statistical properties with transcript abundance value [242].

NSAF values have been used extensively in a broad range of applications for the representation of protein abundance. Such projects include peptide immobilised pH gradient isoelectric focusing (IPG-IEF) analysis of rat liver membrane proteins [243], the subcellular analysis of nuclear proteins in *Saccharomyces cerevisiae* [244], temperature stress response in rice [245], comparison of evolutionary adaptation of Pachyclaydon species through differential protein expression [246], construction of a probabilistic human protein interaction network [247], and the analysis of a wide range of mouse renal cortex proteins [248].

1.4.3 Labelling - iTRAQ

Labelling approaches are generally considered to be similar in accuracy for quantifying protein abundances, however these strategies require more time and more complex sample preparation, and the use of expensive isotope labels in conjunction with specific software for the analysis of the data [249]. Additionally, the number of samples that can be analysed is limited and some of the labelling approaches cannot be applied to all types of sample. Metabolic labelling, for example, is difficult to perform on whole organisms and especially for mammals and plants, where as label-free approaches are more versatile and reliable for measuring global response in protein expression, while being relatively inexpensive and applicable to any sample type [237].

The chemical labelling reagents used in iTRAQ conjugate isotope tags that target the N-terminus and the ε-amino group of lysine residues of peptides [227]. Isobaric mass tags have a mass reporter region (M), a cleavable linker region (F), a mass normalization region (N) and a protein reactive group (R). Figure 1.9 shows a schematic of the labeling technique. Each label subtype should have an identical chemical structure and contain isotopes substituted at various positions on the tag, so that the mass reporter and mass normalization have different molecular masses, resulting in a total molecular mass for the label that is equal and only distinguishable at the isotopic level between each label subtype when cleaved. Isobaric mass tagging produces labelled peptides that precisely co-migrate in LC separations and were first introduced by Thompson and co-workers [228]. It is only upon peptide

fragmentation that the different tags are distinguished by the mass spectrometer. This allows the simultaneous determination of both identity and relative abundance within tandem mass spectra of peptide pairs. The iTRAQ technique has a further refinement of allowing multiplexed quantitation of up to eight samples and the labelling of whole proteins, too [250]. The ability to multiplex has turned out to be useful for following biological systems over multiple time points [236].

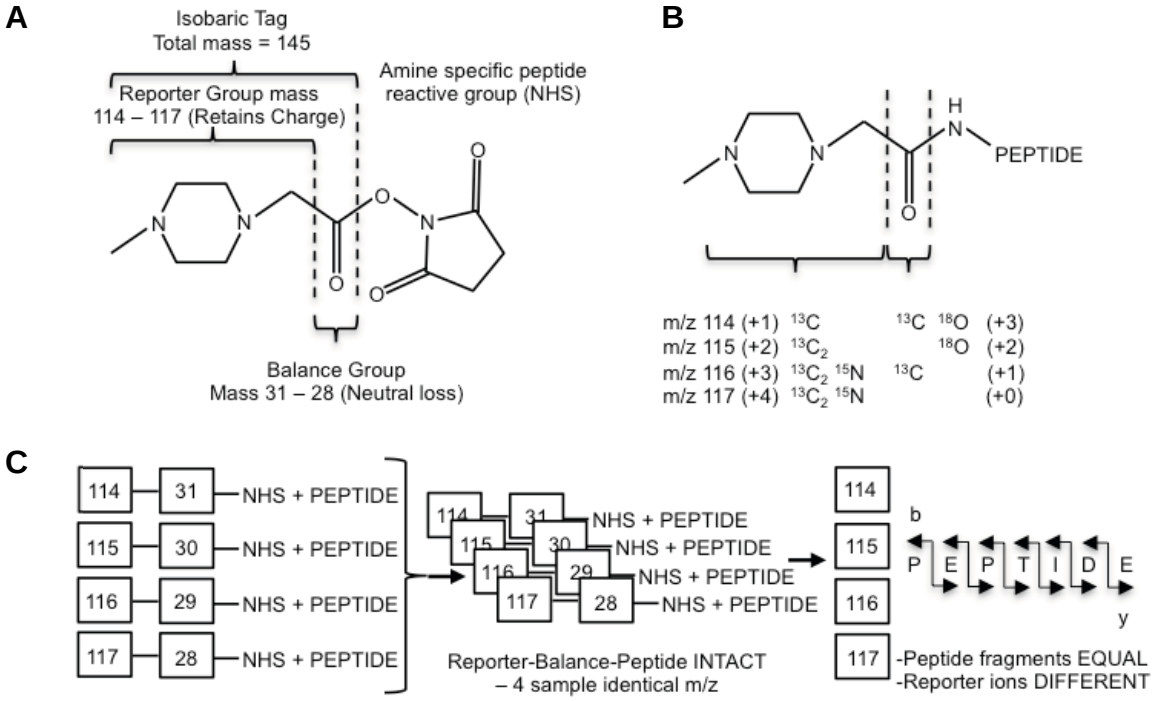


Figure 1.9 (A) Diagram showing the components of the multiplexed isobaric tagging chemistry for iTRAQ labeling. The complete molecule consists of a reporter group (based on *N*-methylpiperazine), a mass balance group (carbonyl), and a peptide-reactive group (NHS ester). The overall mass of reporter and balance components of the molecule are kept constant using differential isotopic enrichment with 13C, 15N, and 18O atoms. The reporter group ranges in mass from *m/z* 114.1 to 117.1, while the balance group ranges in mass from 28 to 31 Da, such that the combined mass remains constant (145.1 Da) for each of the four reagents. (B) When reacted with a peptide, the tag forms an amide linkage to any peptide amine (N-terminal or _ amino group of lysine). These amide linkages fragment in a similar fashion to backbone peptide bonds when subjected to CID. Following fragmentation of the tag amide bond, however, the balance (carbonyl) moiety is lost (neutral loss), while charge is retained by the reporter group fragment. The numbers in parentheses indicate the number of enriched centers in each section of the molecule. (C) Illustration of the isotopic tagging used to arrive at four isobaric combinations with four different reporter group masses. A mixture of four identical peptides each labeled with one member of the multiplex set appears as a single, unresolved precursor ion in MS (identical *m/z*). Following CID, the four reporter group ions appear as distinct masses (114–117 Da). All other sequence-informative fragment ions (b-, y-, etc.) remain isobaric, and their individual ion current signals (signal intensities) are additive. The relative concentration of the peptides is thus deduced from the relative intensities of the corresponding reporter ions. In contrast to ICAT and similar mass-difference labeling strategies, quantitation is thus performed at the MS/MS stage rather than in MS. Figure adapted from P.L. Ross *et al.* 2004 *Molecular and Cellular Proteomics* Vol. 3, Page 1154-69 [227].

The reporter ions used for iTRAQ quantification of MS/MS spectra should not interfere with ordinary peptide fragments and this is why the m/z region of 114-117 was chosen. There are still some interferences identified, notably the 116.1 Da y(1) fragment ion of peptides containing a C-terminal proline residue [95]. The use of a 1 Da mass shifts between the tags means that the instrument used in the analysis must be of a high mass accuracy.

The labelling of intact proteins can be advantageous since it accommodates further protein separation steps on the combined samples, potentially leading to the characterisation of protein isoforms/species [251]. There are two important considerations to iTRAQ protein labelling: one is that trypsin will not cleave at modified lysines which will result in longer peptides that are more difficult to detect by MS; second is that high labelling efficiencies are required in the case of further protein separation pre-MS, since incomplete labelling impairs resolving power [250].

1.4.4 Obstacles Intrinsic to Mass Spectrometry

There are two major obstacles or concerns with respect to quantitative proteomics. The first applies to both labelling and label-free approaches when working in a bottom-up style of analysis, this being the problem of shared peptides, or non-unique peptides, as discussed earlier regarding the protein inference problem [185, 192]. A single gene can result in hundreds of different proteins derived from a multitude of mechanisms that are either pre- or post translational, making the distinction of protein species and protein isoforms indistinguishable when dealing with incomplete sequence coverage [184, 252]. Thus, there are several options to deal with the conundrum of shared peptides for proteomic quantification; (1) ignore them and count them several times for each protein identification, (2) discard them and only focus on unique peptides, or (3) distribute them across the homologous proteins identified. Either way chosen it will affect quantification, as was highlighted in a recent review on the subject by Podwojski *et al.* [253].

Early attempts to handle the protein inference problem as it pertained to quantitation tended to disregard the shared peptides and only work with unique peptides, which is not ideal as this under-represented the true amounts of proteins present [254]. Other

methods presented for spectral counting were to weight the abundances of shared peptides based on the total SpC per protein [255], or alternatively distribute shared peptides based on the abundance of unique peptides identified for these proteins with shared peptide sequences identified [256, 257]. Others have tried to take advantage of biologically related families of proteins that have similar abundance ratios and to incorporate the shared peptides into these peptide-sharing closure groups [258]. A program has been developed called PANORAMICS, that calculates probability scores to determine if a peptide is correctly assigned to a protein [259]. This probability algorithm improved accuracy of identifications and diminished ambiguity regarding shared peptides from homologous groups of proteins. Recently, a study compared these different methods of dealing with shared peptides and surmised that distributing shared SpC based on the presence of unique peptides generated superior results when using spectral counting [260].

The second obstacle is intrinsic to mass spectrometry analysis. Most MS analysis is conducted in a DDA fashion, where a parent ion scan is run in MS mode and selects the most abundant ions on which to conduct subsequent MS/MS fragmentation scans; typically 5-9 ions are scanned in MS/MS before the instrument returns to another parent ion scan. This type of acquisition may introduce a bias in the data for co-eluting peptides towards omitting the lower abundant peptides from ever having MS/MS conducted on them [261]. This bias creates effectively an invisible subset of proteins due to the resultant level of detection limit of the mass spectrometer to not analyse these peptides [262]. Saturation of signal at elevated protein abundance can be due to limitations in the ion-trapping capacity, duty cycle, or ionisation efficiency of the particular mass spectrometer. Repeated sequencing and sequential extraction in various buffer conditions can lessen this problem [235, 263, 264]. It has been shown that sample loading optimisation could increase the upper limit of detection [265]. Optimised dynamic exclusion duration depends largely on sample complexity and is proportional to the average chromatographic peak width at base. Thus, optimised dynamic exclusion duration will significantly increase the detection number of peptides and spectra for low abundance proteins [266]. However, these limitations compress the dynamic range of quantifiable proteins; two to four orders of magnitude have been reported [235, 267].

1.4.5 Relative Solutions

A comprehensive comparative study by the Association of Biomolecular Resources Facilities to assess available methodologies for quantitating eight different proteins in sample mixtures across 52 proteomics groups was conducted in 2006-2007 [268]. The groups were able to choose their form of analysis, with 37% employing gel electrophoresis based methods, while 42% employed MS-based labelling approaches, and 21% used MS-based label-free methods. The study revealed considerable variability for the quantification results. Gel-based approaches exhibited larger inconsistency of percentage error compared to the MS-based techniques. The ratios obtained by both label-free MS AUC and SC were closer to the expected values compared to the ratio obtained by stable isotope labelling approaches. Stable isotope labelling exhibited uniform distribution of percentage error among all proteins except for the lowest concentration protein, with AUC and SC displayed an equal distribution of percentage error for all proteins. Undertaking triplicate analysis by MS-based methods improved results; however there was no further enhancement in accuracy after triplicate or quadruplicate analysis. Triplicate analysis with stable isotope labelling presented greater consistency compared to label-free replicates.

Another study compared ICAT with AUC and SC [269]. The results presented that both label-free methods were as accurate as ICAT for the detection of the standard proteins. However, ICAT failed to accurately detect proteins with four-fold differences in concentration. The performance difference of AUC and SC were examined in more detail to reveal that AUC was slightly closer to the actual known concentrations than SC. Other reports have too shown supporting results that AUC is slightly more accurate than SC [270, 271].

In another study, three MS-based quantification techniques were compared, AUC, $LCMS^E$ and iTRAQ, to both identify and relatively quantitate the proteins levels within a methanotrophic bacterium grown on varying substrates [272]. High-performance liquid chromatography was coupled to a Q-TOF tandem MS for all three quantitative methods. Total numbers of proteins identified by each method were similar and 49% of the proteins were common to all three techniques. $LCMS^E$ provided greater sequence coverage and a greater number of average peptides per protein identification compared to AUC and iTRAQ. Data-independent analysis, particularly

the LCMS^{E} approach introduced by Waters has been suggested to be of great promise for the future of label-free quantification [261].

Venable *et al.* suggested data-independent acquisition (DIA) as an alternative to DDA [261]. DIA does not undertake parent ion scans, the instrument operates constantly in MS/MS mode. DIA can increase the signal to noise by 3-5 fold and identify peptides undetected in a normal parent ion scan of a DDA experiment [273]. An enhanced DIA method, PAcIFIC (precursor acquisition independent from ion count) has been introduced, involving the acquisition of mass spectra at every m/z value [274]. In PAcIFIC, a sample is analysed in 10 continuous 1.5 m/z intervals in a 15 m/z window. This process is repeated for the subsequent 15 m/z window and so on until the m/z range of a typical precursor ion scan is covered, this acquisition process is time consuming and often requires 2-3 days of continuous data acquisition. This method achieved a dynamic range of eight orders of magnitude in a human cell lysate system, which is known to contain several high abundant proteins that would normally restrict comprehensive proteome analysis. This work demonstrate that DIA using PAcIFIC has surpassed DDA methods in terms of proteome coverage, dynamic range and the number of proteins identified, while enabling quantification in a label free form and represents one of the most promising development for label-free quantification for the future of the field.

1.5 Plasma Proteomics and Biomarker Discovery

The human plasma proteome is a major focus of proteomic studies towards disease biomarker discovery, since blood is easily and regularly collected and likely to contain multiple biomarkers [275, 276]. When applied to human plasma proteomics, a biomarker can be considered as anything that can be used as an indicator of a particular disease state or physiological state of an organism, arising from at least three sources: (1) the primary cells for the disease, such as neoplastic cells or endothelial cells; (2) the microenvironment of the primary cells; and (3) systemic responses to the altered protein expression, namely, acute-phase reactant proteins or immunoglobulin. In particular for proteomics, ‘anything’ is denoted as identification and/or quantification of a protein or peptide [277]. There is some dispute as to the potential of plasma or serum to generate clinical biomarkers as opposed to other

bodily fluids, for example urine and cerebrospinal fluid (CSF), although, as a bio-fluid for developmental research it is well suited [278].

It was once thought that the human genome contained between 50,000 – 100,000 genes [279]. The human genome is currently believed to consists of 20,000 – 30,000 genes, while coding for over 500,000 different proteins [280]. It has been demonstrated that alternative splicing of pre-mRNA and mRNA is a widespread process in most eukaryotes and is a major source for protein species generation [281]. Alternative splicing and processing of pre-mRNAs explains the discrepancy between the number of genes and the proteome complexity in multicellular eukaryotic organisms [282]. In humans, recent high-throughput sequencing studies indicate that alternative splicing occurs in 40-60% of all genes [283-286] and >95% of all human genes containing ≥ 2 exons, yielding multiple mRNAs, and therefore multiple protein products [287].

The dynamic range of blood proteins, known to span 9-12 orders of magnitude, the sheer magnitude of proteins, potentially >500,000 candidates, and the extensive physiological variation among patients has complicated the systematic discovery of potential biomarker candidates [277, 288, 289]. Known disease biomarkers, such as prostate-specific antigen (PSA), carcinoembryonic antigen (CEA), and alpha-fetoprotein (AFP), are low abundance proteins in the ng/ml to pg/ml concentration range and it is believed that other disease specific biomarkers are likely to be at similar concentrations [275, 290, 291]. Figure 1.10 outlines the dynamic range of protein present within human blood and highlights the abundance of some known clinical biomarkers. Most of these biomarkers have been elucidated by technologies other than mass spectrometry or proteomics, and the *in vitro* diagnostics (IVDs) tests are in the form of immunologic assays.

Clinical chemistry, as opposed to proteomics, has achieved fundamental importance in medicine, with a spectrum of protein tests, for example C-reactive protein (CRP) for the prediction of disease risk after myocardial infarction and coronary disease, and the levels of the protein thyroglobulin for disease recurrence in metastatic thyroid cancer after the thyroid has been removed [292]. Currently there have been more than 3,000 different proteins successfully detected in the plasma proteome, through an accumulation of various methods [293]. The identification of peptides is only the

first and easiest step for the characterisation of the plasma proteome. Distinguishing the molecular forms (protein species) and the concentrations of the proteins are also necessary so as to understand the diversity and diagnostic potential of the individual protein components of the plasma proteome [294].

Better performing markers have been difficult to obtain judging from the decline in biomarker clearance by government agencies in the past decade [295]. Contrary to this trend, the FDA recently approved a proteomics diagnostic test, OVA1, based on a panel of proteins that aid in the identification of preoperative malignant ovarian tumours [296]. This test was not intended to be a stand-alone OVA1 test and it is not used as such [297]. The introduction of new protein analytes has averaged 1.5 per year and this rate has remained flat over the past 15 years. This rate of discovery is not able to meet the projected medical requirements and highlights deficiencies in the protein biomarker pipeline from which no proteomics-discovered analytes have yet emerged [298]. Current proteomics still has difficulties detecting the low abundant species that can be currently measured by specific immunoassay or functional assays [294].

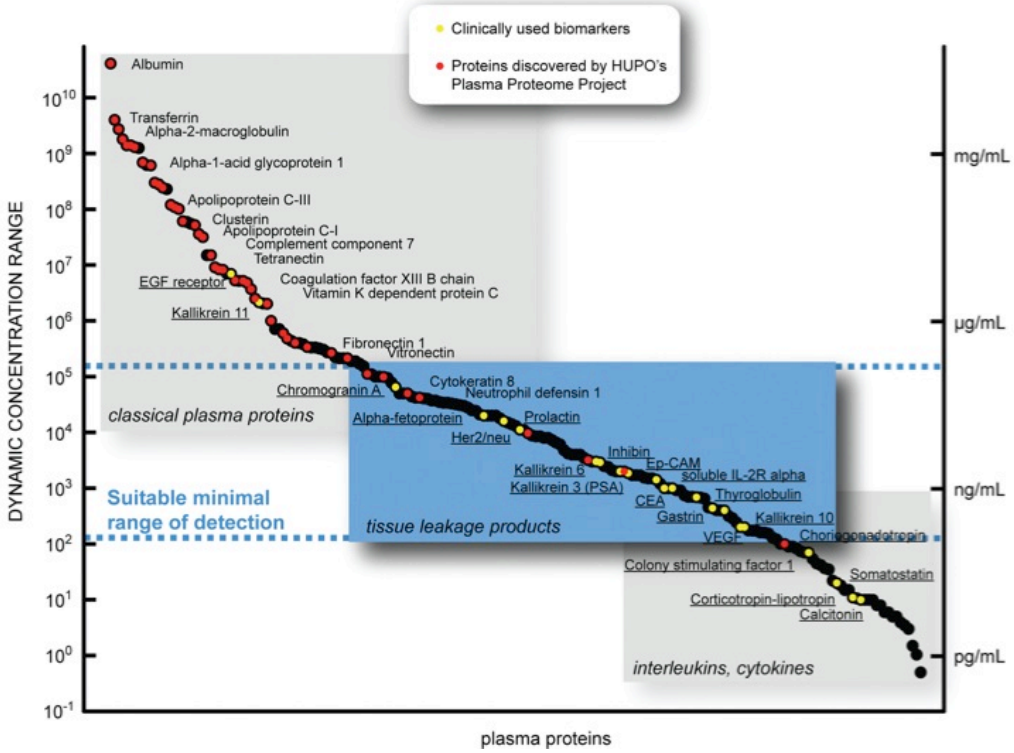


Figure 1.10 Dynamic plasma protein concentration range and the three main plasma protein categories are shown as reported by Anderson *et al.* [275]. Red dots indicate proteins identified by the HUPO plasma proteome project (PPP) [293] and yellow dots represent currently used biomarkers in the clinic. [218] Suitable minimal range of detection for biomarker targeting in plasma is shown with dotted lines. Figure adapted from S. Surinova *et al.* 2010, *Journal of Proteome Research* Vol. 10, Page 5-16 [295].

Leigh Anderson has stated recently that there are several factors that have been identified as limiting progress in identifying new clinical biomarkers using proteomics-based mass spectrometry methods [298]: the lack of an effective technology platform to verify candidate markers in large datasets [299], difficulty surrounding access to clinical samples without significant bias [300], the absence of an organised biomarker pipeline [291, 301], and the absence of anything approaching a useful theory of biomarkers [298].

The requirement for new and enhanced clinical biomarkers and assays has not gone unnoticed by the proteomics community, in particular the Human Proteome Organisation (HUPO, http://www.hupo.org/), which has begun to establish coordinated efforts to address these needs [302]. HUPO is a young organisation of approximately 10 years and has established numerous goals with regards to bioinformatics and the use of plasma and serum [277, 303]. Their most ambitious endeavour to date was the announcement that they will now actively start to map the entire human proteome [304]. This announcement is not only a sign of the importance and need for the information stored within the human proteome, but that the technological platforms have possibly reached a maturity to enable such an endeavour to generate tangible data [305].

Strategies to overcome sample collection, preparation, analysis and interpretation are the main areas for HUPO to address, as they are for any individual lab or group wanting to map a proteome, in particular human plasma. I have focused on sample preparation and data interpretation through liquid chromatography (LC) and a unique graphical presentation of the bioinformatics for the analysis of the whole plasma proteome in Chapter 3.

Sample collection will always be hampered by the genetic and environmental difference between people of variable nationality and health at the time of sample collection. The largest area that can be tackled in a tangible way is the uniformity of collection and storage and collating of information about the donor for selection by the researcher, based on whether they want to conduct a broad study or a more detailed disease state for example [291].

Sample preparation has seen some of the largest developments within the past 10 years, particularly with respect to the various forms of chromatography, be it gel or liquid based [306]. There are a multitude of issues to overcome when dealing with human plasma. As mentioned earlier the dynamic range of plasma is greater than 10 orders of magnitude and this is a major sample preparation and handling obstacle for analysis. One strategy to tackle the dynamic range problem is to use multi-dimensional separation strategies involving orthogonal protein or peptide fractionation steps [307-311].

One particular strategy is the removal of high abundant proteins, referred to as a depletion strategy. This reduces the dynamic range of the proteins in the sample, resulting in the identification of low concentration proteins that would otherwise not be detected without the implementation of a depletion strategy. While removing high abundant proteins to identify the lower abundant proteins, there is a trade off, for there is also some co-depletion of lower abundance proteins in the process. Until approximately 2003, Cibacron® blue dye was the most effective way of removing albumin and Protein A or G, which constitutes ~65% and ~15% of the plasma proteome respectively. Unfortunately, this did not remove the entire albumin component, while at the same time removing lower abundance proteins [312]. Agilent was the first company to bring out a commercial product that simultaneously depleted six abundant proteins, known as the Multiple Affinity Removal System (MARS), reducing the protein content of human serum or plasma by ~85% [313]. The ideal goal is to target both high abundant and medium abundant proteins so as to remove 98-99% of the proteins [307, 312]. Other manufactures have brought out antibody columns that target between 6 - 20 abundant proteins, [Aurum™ Affi-Gel® Blue mini kit (Bio-Rad), Vivapure® anti-HSA/IgG kit (Sartorius Stedim Biotech), Qproteome albumin/IgG depletion kit (Qiagen), MARC (human 6) (Agilent Technologies), Seppro® MIXED12-LC20 column (GenWay Biotech) and ProteoPrep® 20 plasma immunodepletion kit (Sigma-Aldrich)] [314].

The abundance removal field is moving towards medium range proteins too, for example the recently introduced Seppro® Supermix column (Sigma-Aldrich) is an immunoaffinity column designed to be used in tandem with IgY columns, targeting medium-abundant proteins. It has been reported that IgY columns are reproducible

and greatly increase detection of lower-abundant proteins by removing approximately 60 medium- to high-abundant proteins [315].

Following the depletion of the high- to medium-abundant proteins, a number of strategies for further fractionation that exploit the orthogonal physiochemical properties of the molecules are the basis for the design and implementation. The most common shotgun proteomics approach involves separation of tryptic peptides by SCX followed by RP-LC [146, 316, 317]. This approach was implemented as the standard method for the comparative work against the method I developed which employed SCX and SAX on the plasma proteins after depletion followed by tryptic digestion of the protein fractions and RP-LC-MS/MS on the peptides.

Since tryptic peptides have limited orthogonal physiochemical properties, separation methods that employ both proteins and peptides can maximise the effectiveness of fractionation strategies [318]. Shotgun proteomics is suited for discovery-driven profiling achieving thousands of identifications. It relies on DDA, which selects ions for fragmentation based on abundance dependent heuristics [319]. The technique is ideal for discovering novel proteins, but limited to detection of high abundant species within the sample with reduced reproducibility of peptide identifications [320].

Multidimensional fractionation, in either a gel matrix (1D- or 2D-GE) or in a liquid chromatographic system, is the main method for the fractionation and decomplexation of plasma samples before MS analysis. The development of microfluidics has a myriad of applications that have grown exponentially over the past 15 years and separation chemistry has had some of the greatest benefits from the fabrication of micro-electro-mechanical system (MEMS) that have the potential to revolutionise protein chromatography [321].

With respect to the identification of biomarkers and mass spectrometry, there are three main advancements currently underway. The first being the introduction of top-down proteomics [134]. Secondly, the introduction of tissue imaging MALDI [322, 323], with recent advancements in 3D imaging holds great promise for the future [324]. Thirdly, a recent alternative based on a more targeted approach called selected reaction monitoring (SRM/MRM) has been employed to address some of these holes within the field by targeting *a priori* selected protein sets repeatedly and

generating more consistent data [325, 326]. To date the most sensitive method of SRM is with a triple quadrupole (QQQ) MS platform [327]. Protein quantification in plasma by multiplexed SRM has covered a dynamic range of 4-5 orders of magnitude with a limit of quantification (LOQ) of 1 µg/ml [328, 329]. Implementation of the isolation of a sub-proteome based on phosphorylation and in particular glycosylation achieved the detection of relevant proteins at the required sensitivity to a LOD of 5 ng/ml [330-332]. This highlights how affinity capture liquid chromatography coupled to SRM mass spectrometry can begin to work at the necessary LOD and LOQ for biomarker studies, thus becoming an important and burgeoning methodology for proteomics and biomarker research in the near future.

1.6 Agricultural Proteomics – Temperature Stress in Rice

In a recent editorial presentation on plant proteomics, plants were referred to as “our bread and butter” and being the most important species on our planet, that control our food production and sustenance [333]. Of these plants, rice is one of the most important food crops in the world and is a staple for approximately half the world’s population [158]. The particular strain of rice used in our experiments in Chapter 4 was *Oryza sativa Nipponbare* cultivar, a Japonica varietal. Rice is divided into two main varieties, Japonica and Indica, with Japonica grown in temperate regions like Australia, while Indica is grown in more tropical climates.

It has been estimated that global mean temperatures will increase between 2-4.5 ^{o}C during this century, with associated extreme temperature and environmental events occurring during this period, bringing about climate change [334]. With the current population of the world estimated to be ~6.8 billion and its projected growth to 9 billion by 2040, combined with climate change, there is a definite need for adaptation and to increase the production of food if we are to move forward as a sustainable society. Despite these predictions, the current importance of plant biology research seems to be underestimated, even in light of the fact that there are more deaths from hunger than disease in the world today [335].

Climate change will bring about increases in both the frequency and amplitude of severe temperature events, and these fluctuations will have profound effects on the

natural ecosystem and agriculture [334]. Nutrient availability, water, salinity and temperature changes influence agricultural production across the globe [336]. Plants are sedentary and have a finite capacity to acclimatise to rapid (shock) changes in the environment, in particular temperature stresses, even though long-term adaptation has been established in systems with extreme biomes. Cold snaps, where the temperature is suboptimal for several days, inhibit rice development, particularly during the reproductive stage where low temperatures can render the plant sterile reducing rice yields by 30-40% [337, 338]. Damage to the early stages of male gametophyte development are, caused by abiotic stress and occurs in many important crops such as rice, wheat, maize, barley, sorghum and chickpea [339]. Therefore, it is the rate of change coupled to the frequency and amplitude of these temperature fluctuations from the mean values, which is critical for productivity of all plant species, including rice.

Traditionally, efforts to increase the yield of rice have ranged from cross mating and hybridisation of two or more different varietals based on phenotype however, advance in genomics allow studies for the relationship of genotype with phenotype representing a paradigm shift in plant breeding [340]. The advances of old, for example; cutting off the tip left behind after harvest, undertaking multiple harvests in a calendar year, moving cultivation to regions of higher rain fall or optimal temperatures, can not keep up with the increased demands of the future unless a major change or shift in the strain (type) of rice occurs, similar to the increased production due to the introduction of ‘dwarf wheat’ in the middle of the 20th century [341, 342]. Minimal advances have been made to enhance the genome of rice so as to increase its ability to resist infection, handle unique climate changes and raise the nutritional content [343]. These advances and the techniques used to create these modifications are the foundations on which the future of food crops and their accelerated adaptation and evolution will need to be built upon to meet the demands of the 21st century [343].

There are many arguments for and against the intervention into the genome of food crops and it is not the focus of this thesis to analyse, nor is it the premise under which these experiments were devised [344, 345]. I believe empowerment and emancipation of the individual, or a society, through knowledge and awareness, with

scientific research being the search for truth and knowledge, will allow more informed decisions that produce better projected and resultant outcomes for our community.

In addition to the social, nutritional and economic importance of rice, it has become a favoured and attractive model system for cereal genomic research, due primarily to its relatively small genome (approx. 32 000 genes), high genomic synteny with other cereal crops, compatibility with genetic transformation and available sequenced genome [346]. When the rice genome was sequenced it was, and still is, considered the “Rosetta Stone” of cereal biotechnology [347].

Transcriptomics and proteomics have been applied to identify the stress-responsive genes and proteins that are effected by elevated temperatures, salinity, cold and water deficit in both rice and wheat [348-351]. Proteomics approaches have also been applied to ascertain the stress related proteins [352-358]. Stress-related genes and proteins have been identified, involved in abscissic acid and jasmonic acid biosynthesis and signalling, redox homeostatis, energy metabolism, polysaccharide and cell wall metabolism and defence [336].

To date, only a small number of genes have been isolated as being stress responsive in both the vegetative and reproductive stages. For example, OsSALT (encoding a 15-kDa mannose-binding lectin protein) and OsNac6 (encoding an apical meristem transcription factor) are induced by abiotic stress in rice [359]. However, the impact of the stress responses at the reproductive stage is different to the vegetative stage, for it is clear that abiotic stresses affect grain-yield more than vegetative growth [360]. Despite the rice genome being successfully sequenced, there are still approximately one third of the proteins uncharacterised, despite the recent efforts of rice proteomics [346, 361]. This lack of characterised proteins from the rice genome was evident in the analysis conducted within Chapter 4.

Despite the lack of characterised proteins for the genome of rice, there is evidence that the mechanisms at play with respect to abiotic stress or any biochemical event are multifaceted and may require both extensive interrogation and interpolation. For example, in programmed cell death, several proteins identified were cold-responsive proteins, including *Oryza sativa* cold-induced anther protein (OsCIA) [362]. This protein is present in panicles, leaves, and seedlings under normal conditions and

was induced in anthers after cold stress. There was no observed change in the mRNA levels for OsCIA in the anthers, panicles, leaves and seedlings after cold treatment, exemplifying that OsCIA has differential tissue-specific expression that is regulated by some unknown post-transcriptional mechanism. A vital movement to advance from genomics through functional genomics towards systems biology is the definition of protein interactions in living cells [363], which is one of the main challenges in plant research for the next few years and proteomic research as a whole [364].

Methodologically, 1D- and 2D-Gel electrophoresis coupled to MS is the staple for plant proteome analysis [361]. Applications of gel-free protein separation and the so called “second generation” proteomics techniques of the likes of MudPIT (shotgun) with labelled or label-free quantification were stated as being still anecdotal by Jorrín-Novo *et al* in 2009 [335]. Both comments I believe have shown to be out-dated and are not representative of the current status or the future direction of plant proteomics [336, 365, 366]. Yes, these “second generation” techniques are in their infancy for plant proteomics though they are no longer anecdotal as I hope our research shows [158, 237]. This is a great example of how fast the field of plant proteomics and associated methodologies are evolving and how that, with some adaptations, the approaches used in analysis of yeast and mammalian proteomes can be incorporated into the study of plants.

1.7 Concluding Remarks

The understanding of the structure of DNA by X-ray crystallography in the middle half of the 20th century brought about a golden age for the field of genetics of approximately 50 years, culminating in the mapping of the human genome at the beginning of the 21st century. It could be said that since this time the field of proteomics has been and is undergoing a similar golden age, with the setting of the task in 2010 to map the human proteome. If history is any guide, the project should be complete by approximately 2050, or at the least, similarly to the human genome, a first draft should be complete and heralding novel techniques not foreseen at the beginning of the journey.

Proteomics has ventured a long way in a short time, from the primitive days of eluding to the primary structure of a protein to detailed tertiary and quaternary information along with post translational modifications. Temporal, spatial and quantitative proteomics leads to the next significant leap, systems biology. Proteomics is currently driven by mass spectrometry, with the associated areas of sample preparation and bioinformatic interpretation of the data produced being the main areas of technical development for the betterment of the field.

In the future it is not necessarily mass spectrometry that will be the driving force of proteomics as it has been since the 1980s. It is envisioned that further advancements in computational capabilities and bioinformatics platforms will better disseminate the information proteomics can and is producing today. These will allow researchers to conduct whole scale protein-protein, protein-nucleic acid or protein-small molecule interaction studies, coupled with transcriptomics and metabolomics for a complete systems biology approach. Proteomics and systems biology will become increasingly important in a range of biological, bio-medical or agricultural applications.

CHAPTER 2: DEVELOPMENT OF NOVEL SURFACE CHEMISTRIES FOR ENHANCED MATRIX ASSISTED LASER DESORPTION IONISATION MASS SPECTROMETRY

2.1 ABSTRACT

The field of proteomics escalated in the latter part of the 20th century due to the implementation and development of soft ionization mass spectrometry of which Matrix Assisted Laser Desorption and Ionization (MALDI) is one of the two most prominent techniques used today in the field of proteomics for mass spectrometric analysis of proteins and peptides. Early success and development of enhanced forms of MALDI with surface chemistries applied to the plate surface achieving concentration, purification or affinity capture of a sample sparked a multitude of alternative methods trying to capitalize on the potential of MALDI to become in the 21st century the analysis method enabling the discovery of new biomarkers, bio-chemical warfare agents and even surpassing the ELISA technique. This study investigated the potential viability of a unique form of MALDI with surface chemistries applied that we call spherically concentric surface chemistry MALDI (SCSC-MALDI). Three particular archetypes were studied: the first was the X3 concentrating Biochip, the second was the RP3 desalting Biochip, and the third was the HA3 affinity capture Biochip.

The study showed that the X3 concentrating Biochip had a limit of detection (LOD) of 1 x 10^{-18} mol/µl for single peptide samples, which represents a 10-fold (for ACTH) and 100-fold (for bradykinin, angiotensin I and neurotensin) increase in sensitivity of detection compared to standard MALDI. With a four peptide mixture the LOD was 10 x 10^{-18} mol/µl for Bradykinin, Angiotensin I and Neurotensin while for ACTH it was 30 x 10^{-18} mol/µl. The relative differences in the LOD between the X3 Biochip and Std-MALDI similarly displayed a 100-fold increase for the peptides Bradykinin, Angiotensin I and Neurotensin while ACTH 18-39 displayed a 10 fold increase.

A comparison between the X3 Biochip, AnchorChip and Std-MALDI for the identification of peptide digests, Human Serum Albumin and Enolase found that both the X3 Biochip and AnchorChip out performed the Std-MALDI method. While both the X3 Biochip and AnchorChip perform similarly across the middle and top end of the concentration gradient for the identification of the digests used, the X3 Biochip displayed a potential benefit by outperforming the AnchorChip at the lower end of the concentration gradient analysed on both digests.

The HA3 Biochip was able to selectively capture phosphorylated peptides with a LOD of 0.75 femtomole applied. The HA3 Biochip was also able to selectively capture phosphorylated peptides in an environment of competing peptides from a Bovine Serum Albumin digest (10 molar excess) at the low end of the femtomole range.

The RP3 desalting Biochip was problematic when working on protein digests from SDS-PAGE gel fractionated samples, with unexplained contamination and interference of the concentration event, which could not be isolated beyond potential manufacturing, sample or handling variables.

It was observed that SCSC-MALDI is highly susceptible to contamination that adversely affects the ability of the Biochip to concentrate or purify a sample on the surface leading to ambiguous and non-reproducible results limiting the application of the technique. In my opinion, at the time of these experiments the SCSC-MALDI technique needed further development before implementation as a robust analytical technique, particularly the RP3 and HA3 archetypes. However the X3 format did display superior results and promise for the analysis of purified peptide samples compared to Std-MALDI.

2.2 PREAMBLE

The technology used within Chapter 2 is referred to as Spherical Concentric Surface Chemistry MALDI (SCSC-MALDI), or more loosely as the Biochip; of which there has so far been only one peer-reviewed publication of this technology [60]. I will provide a detailed explanation of the workings of such technology, while highlighting both the positives and negative of such a technique. In this thesis I describe the first independently conducted experiments on this sample preparation and presentation Biochip for MS in proteomics, based on the patterning of three or more unique self-assembled monolayers (SAMs) on flat metal surfaces, the SCSC-MALDI technology, marketed now as the Mass•Spec•Focus Chip™ (LCI/Qiagen).

At the beginning of these experiments in July of 2004, there was burgeoning interest in enhancing standard MALDI (Std-MALDI), conducted mostly on flat metal surfaces, by way of altering the surface chemistry of the MALDI plate (SC-MALDI). The AnchorChip™ (Bruker) and SELDI™ (Ciphergen) techniques were leaders in the field [61, 64]. Simultaneously, high sensitivity MALDI TOF and high throughput MALDI TOF/TOF analysers became available. The desire to take advantage of this high throughput MS/MS to conduct larger scale screening type experiments coupled with superior limits of detection (LOD) created a potential need for an enhanced surface chemistry MALDI [367, 368]. The aim of this was to reduce sample loss and handling and increase the LOD and dynamic range of the mass analysers, whilst being reproducible, quick and cost effective compared to nano-Electrospray Ionisation (nano-ESI) and Std-MALDI mass spectrometry [59]. I utilised a self assembled monolayer (SAM) chemistry patterned on a gold covered metal plate that is identified within as a SAM Biochip as the novel approach. Alternatively, the technique is also referred to as a spherically concentric surface chemistry MALDI (SCSC-MALDI) due to orientation of the surface chemistries on the surface of the MALDI plate in these experiments.

The major abilities required of any surface chemistry MALDI are liquid containment, manipulation and presentation within the one surface (location) through either concentration of the sample (similar to AnchorChip™), purification through cleaning/desalting of the sample (similar to RP C_{18} chromatography) or affinity

capture (similar to SELDI™), and presentation of analyte and matrix in a co-crystallised form conducive to ionisation and desorption for MALDI mass spectrometry [59]. Figure 2.0.1 illustrates the comparison between Std-MALDI, AnchorChip™, SELDI™ and SAM Biochip used in these experiments. These capabilities should reduce sample handling and the sample losses associated with standard sample preparation. This has the potential to result in increased ion intensity (signal strength) of the analytes at lower concentrations, which should increase the dynamic range of the mass spectrometer or at the very least increase the LOD for the mass spectrometer, providing more information for investigators than was previously available from similar sample sets.

The desire to conduct one or more of the standard sample preparation steps for a proteomics sample within the same location on a MALDI plate, also known as the target, well or spot, is paramount to making any of the advancements mentioned earlier an attractive endeavour worth investigating. The distinguishing factors of this particular type of SC-MALDI used are: the type, orientation and number of surface chemistries employed.

In this study I investigated alkylthiols patterned onto a gold covered metal surface as a self assembled monolayer (SAM). The chemistries are orientated into concentric circles on the surface with varying degrees of hydrophilicity. There are three (or potentially more) distinct chemistry regions for each well (spot), which are referred to as a chemically defined virtual well (CDVW).

I tested the viability of three particular platform types of SCSC-MALDI supplied by our collaborators at Lumicyte Inc. (LCI), which became a Qiagen/Lumicyte venture in 2006 after Qiagen acquired Lumicyte (Qiagen, Germany and Lumicyte, USA). Figure 2.0.2 outlines the general architecture of the technique. The first type tested was the earliest incarnation of the technology, the concentration chip or X3 Biochip (Mass•Spec•Focus Chip™). The second type tested was a sample clean-up chip, analogous to a reverse phase purification step, identified as the RP3 Biochip. The third type tested was an affinity capture chip, specifically for phosphorylated peptides or proteins utilising immobilised metal affinity chromatography (IMAC) principles, specifically by employing nitrilotriacetic acid (NTA) and identified as the NTA3 Biochip.

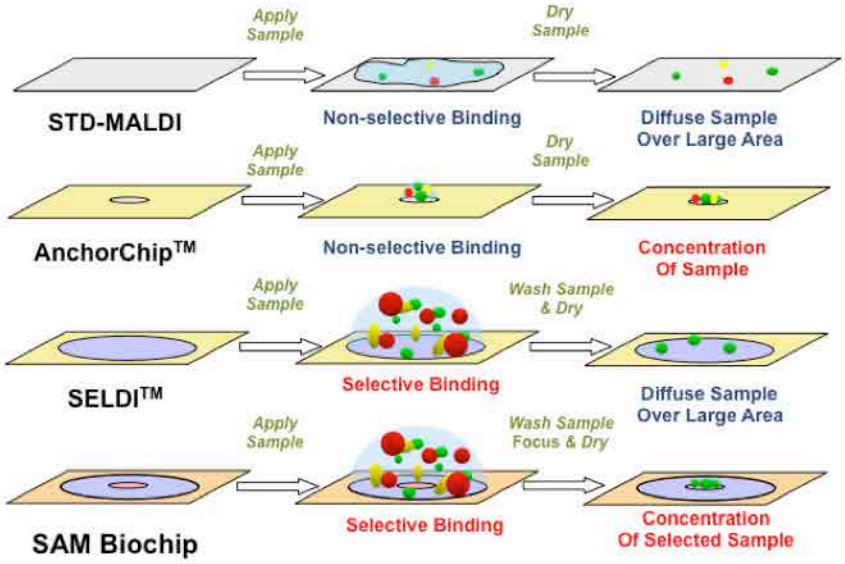


Figure 2.0.1 Illustration highlighting the theoretical differences between the main types of SC-MALDI. The first is Std-MALDI, which has non-selective binding and the sample is diffused over a large area dependant on the spotting technique employed. The second is the AnchorChip™, which has non-selective binding in its simplest form and can concentrate the sample due to large differences in hydrophilicity between the two chemistries, with the external chemistry being more hydrophobic compared to the centre. The AnchorChip™ has been shown to conduct the desalting of samples is some incarnations [369]. The third is SELDI™, which does achieve affinity capture and the removal of contaminants, though does not concentrate. The fourth is the SAM Biochip in concentric circles, which has been shown to concentrate samples due to its three unique chemistries, both in this thesis and in the literature [60]. I have shown that it can also achieve affinity capture and limited concentration on the one surface.

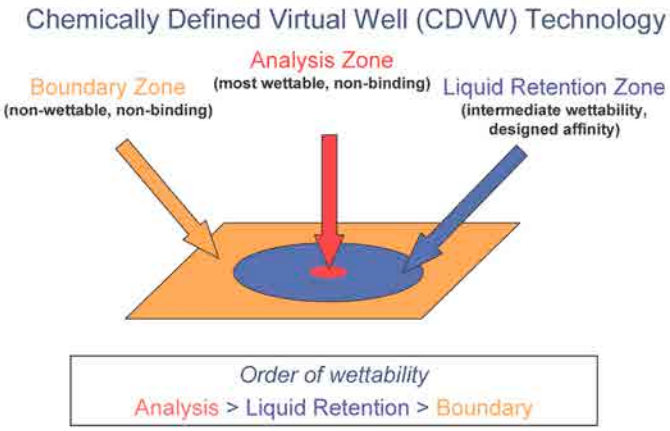


Figure 2.0.2 Illustration of the three regions (zones) that make up the CDVW for the SAM Biochip in concentric circles used in these experiments. There is a hydrophobic, non-wettable and non-binding "Boundary Zone". Followed by an intermediate hydrophilic/hydrophobic "Liquid Retention/Capture Zone". Lastly, the central "Analysis Zone" which is of the highest hydrophilicity (wettability) and is non-binding. These different zones are invisible to the naked eye until liquid is applied, then the Boundary Zone and Liquid Retention Zone are visible due to the containment of the liquid in the droplet that is formed. Only once the matrix solution is added the unique 'snap-lock' concentration event occurs and the Analysis Zone becomes visible, due to the residual precipitation and crystallisation of the analyte and matrix in the area.

All the work presented here was conducted using an Applied Biosystems (ABI) 4700 MALDI TOF/TOF Analyser at either the Australian Proteomics Analysis Facility (APAF) within Macquarie University (MQ), Australia, or at the demonstration labs of ABI in Fremont, California, USA. An additional comparison of the X3 Biochips was conducted against the Bruker AnchorChip^{TM} and Std-MALDI, undertaken within an ABI 4800 MALDI TOF/TOF mass analyser at the John Curtin Medical School at the Australian National University (ANU), Australia, requiring me to design the first hybridised Bruker/ABI (AnchorChip/4800) MALDI plate system for the 4800 mass analyser.

This chapter is written in a more narrative style compared to the other chapters in this thesis with the primary goal of trying to communicate the benefits and liabilities of this particular type of platform. I have tried to highlight the main methodologies for this type of SC-MALDI in a more traditional “Material and Methods” section within the chapter. I have also tried to supply mainly experiments that showed positive results, thus omitting the plethora of unsuccessful experiments that were performed. However, I will make reference to some of these where necessary, elaborating on why I believe they were unsuccessful.

It is important to acknowledge that much of the methodology for the initial handling of samples and manipulation of the Biochips came from LCI, since they created the technology. All of the methods presented required some form of adjustment and adaptation by myself, some minor for the X3 Biochip and some major for the NTA3 and RP3 Biochips, which were done as part of a research collaboration between myself and the scientists of LCI including, Dr. Christopher M. Belisle, Dr. John A. Walker II, Dr. Mark J. Levy, Dr. Douglas P. Greiner and others.

2.3 MATERIALS AND METHODS

2.3.1 Materials and Reagents

All chemicals and reagents were purchased from Sigma-Aldrich (St. Louis, Missouri, USA) unless otherwise stated.

All solvents used were HPLC grade or higher. Acetonitrile (ACN), ethanol (EtOH) and ammonium hydrogen citrate (AHC) were 98% ACS reagent, 1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC), N-hydroxysuccinimide (NHS), octyl-β-glucopyranoside (OBG), β-Casein digest and bovine serum albumin (BSA) protein were all from Sigma-Aldrich (St. Louis, Missouri, USA). Trifluoroacetic acid (TFA) was from Pierce (USA). Reverse phase C_{18} ZipTips were from Millipore (USA). Concentrating Self Assembled Monolayer (SAM) Biochips and α-cyano-4-hydroxy-cinnamic acid (CHCA) were from Lumicyte (USA) and Qiagen Biosciences (Germany). The peptides, Bradykinin (1061), Angiotensin I (1297), Neurotensin (1673), ACTH 18-39 (2465), T6 β-Casein (2061), T1-2 β-Casein (3122) and organic molecule N-(5-Amino-1-carboxypentyl)iminodiacetic acid (AB-NTA) were from Auspep (AUS).

2.3.2 Limit of Detection Comparison for Peptides Utilising the Concentrating (X3) Biochip Compared to Standard MALDI within an Applied Biosystems 4700 Analyser

2.3.2.a Preparation of Peptides

Peptides were solubilised in 25% ACN : 0.1% TFA (1:1) at concentrations ranging from 0.1-1 mg/ml and stored at -20^{o}C. A concentration and purification step of the sample using μC_{18} ZipTips was conducted to remove low-level impurities. The concentrated and desalted solution was then quantified by amino acid analysis.

2.3.2.b Preparation and Deposition of Single Peptide Samples

All working solutions were made up fresh on the day of the experiment. Six dilutions of 1 in 10 of the peptides were solubilised in 25% ACN / 0.1% TFA, starting at $1x10^{-14}$ mol/µl to 1 $x10^{-19}$ mol/µl (10 nano molar to 0.1 pico molar). The dilutions used were applied to the X3 Biochip in triplicate at a volume of 10 µl per well. The chip was then placed in a desiccator (without desiccant) and placed under a mild vacuum with a diaphragm vacuum pump to speed up the evaporation process. Once dry, the Biochip was placed inside a humid chamber. Aliquots (2 µl) of the matrix solution (0.063 mg/ml) solubilised in (84:13:3 v/v ACN:EtOH:5 mM ammonium hydrogen citrate in 0.1% TFA) were added to each site and the spots were allowed to concentrate. Once dried the chip was used for analysis by MALDI-MS.

The Std-MALDI samples were spotted down in triplicate from the same dilutions used for the X3 Biochip depositions. Each dilution was pre-mixed with matrix in a ratio of 1:1 (v/v) using CHCA at a concentration of 4 mg/ml and 1 µl of the mixture was spotted down.

2.3.2.c Preparation of Four Peptide Mixture

The methods used were the same as above for the single peptide samples except that the four peptides, bradykinin (1060.5692^{+} m/z), angiotensin I (1269.6853^{+} m/z), neurotensin (1672.9175^{+} m/z) and ACTH 18-39 (2465.2004^{+} m/z), were mixed together in a ratio of 1:1:1:4 (mole) respectively. Six dilutions of 1 in 10 of the mixture were diluted, starting from $1x10^{-13}$ mol/µl to $1x10^{-18}$ mol/µl (0.1 micro molar to 1 pico molar).

2.3.2.d Mass Spectrometric Parameters (4700 TOF/TOF)

Sample were analysed using a 4700 Proteomics Analyser MALDI-TOF/TOF (Applied Biosystems Inc., USA). New plate templates had to be generated within the 4700 software due to the orientation and size difference of the X3 (5 x 5) Biochip wells compared to the Std-MALDI plates used. All spectra were recorded in positive

reflectron mode. MS data acquisition was set up with 120 laser shots per 40 locations of laser positioning on the sites. An unbiased random firing (raster) pattern was used, focusing in on the smaller analysis zone of 500-600 µm in diameter compared to the larger standard diameter of 1000-1500 µm for the Std-MALDI plate, and no sub-spectra were excluded from the final accumulated spectrum. In order to keep continuity in the comparison all instrument parameters were kept the same with the exception of the position of the first mirrors, X_1 and Y_1 and laser intensity when it was deemed necessary. These three parameters were optimised each time an experiment was run. The same Opti-Tof holder was used for both the X3 Biochip and Std-MALDI inserts. A conservative limit of ≥10:1 signal-to-noise ratio was used as a cut-off point to reconcile if a peak was considered to be valid and acceptable; anything below this was excluded for the analysis when drawing comparisons of the data collected. Signal intensity in the form of the signal to noise ratio (S/N) were recorded and graphed relative to the particular concentrations applied as shown in Figure 2.1.2 -2.1.4, 2.1.7 and 2.1.8.

2.3.3 Limit of Detection and Identification Comparison of Peptide Standards and Protein Digests utilising a Novel Hybridised AnchorChip/ABI 4800 MALDI Plate versus the Concentrating (X3) Biochip within an Applied Biosystems 4800 Analyser

2.3.3.a Incorporation of the AnchorChip and X3 Biochip into an Applied Biosystems 4800 Analyser

The AnchorChip in the micro-titre plate (MTP) format had the magnets removed and guide holes were drilled for the alignment poles to sit in as orientation points for mounting, and secondary holes for the attachment screws were counter sunk so that the screws would sit flush, to below, the top surface, were also machined. A standard 4800 plate holder in the MTP format was modified by firstly removing the magnets, followed by machining down the top surface so that the combined height of the hybrid system when joined with the modified AnchorChip on top and the modified 4800 plate on the bottom, would be the same height as an original 4800 plate. The alignment (guide) poles were attached to the modified holder and threaded holes for the

attachment screws were created. Refer to Figures 2.2.1 and 2.2.2 for a pictorial representation of this novel hybrid MALDI platform. All of the external dimensions of the final hybrid plate/holder were the same as the original 4800, allowing it to fit inside the machine. The weight of a standard 4800 plate holder in the microtitre plate format with an LC MALDI plate attached is ~280 grams, while the novel hybrid system is 380 grams. Plate templates were made within the 4800 software for the new orientation of the wells in this novel hybrid MALDI platform.

The 4800 came with a holder that would take the standard 4700 Opti-Tof insert plates, so the X3 Biochips and Std-MALDI plates could be run in the 4800 without any need for physical modification.

2.3.3.b Preparation and Deposition of Protein Digests

MassPREP^{TM} Enolase digest from Yeast and Human Serum Albumin (HSA) digestion standards (Waters, USA) were used as the protein digests in the comparison. Initial solubilisation of the lyophilized digested proteins was conducted as per the manufactures specifications to 1 x 10^{-12} mol/µl. Eight serial dilutions of 1 in 10 were made using 25% ACN / 0.1% TFA ranging from 1 x 10^{-15} mol/ µl to 1 x 10^{-22} mol/ µl (1 nanomolar to 0.1 femtomolar).

Std-MALDI with Opti-Tof insert in the 24 x 16 configuration were dried droplet applied with 1 µl of sample followed by 0.5 µl of matrix 4 mg/ml CHCA in a ratio 7:3 - ACN : 0.1% TFA (5 mM dibasic ammonium citrate).

Bruker AnchorChip^{TM} in 24 x 16 configuration were dried droplet applied as per the manufactures specifications with 5 µl of sample followed by 1 µl of matrix 0.3 mg/ml CHCA in a ratio 7:3 - ACN : 0.1% TFA (5 mM dibasic ammonium citrate).

X3 Biochips in the 8 x 8 and 9 x 9 configuration were used as per the earlier experiments on single peptides standards, with 10 µl of sample applied followed by 2 µl of matrix 0.063 mg/ml CHCA in a ratio 84:13:3 - ACN : EtOH : 0.1% TFA (5 mM dibasic ammonium citrate).

2.3.3.c Mass Spectrometric Parameters (4800 TOF/TOF)

Samples were analysed using a 4800 Proteomics Analyser MALDI-TOF/TOF (Applied Biosystems Inc., USA). New plate templates had to be generated within the 4800 software for the novel hybrid AnchorChip/4800 plate holder in MTP format, and X3 Biochips in 8 x 8 and 9 x 9 format. Single peptides (bradykinin, angiotensin I, neurotensin and ACTH 18-39) and mixtures of these peptides in a ratio of 1:1:1:3 respectively, were made similar to earlier experiments and used to optimise the deflectors, signal resolution, laser intensity and shot numbers for MS and MS/MS, facilitating one optimal acquisition method for all three platforms. All spectra were recorded in positive reflectron mode. MS and MS/MS data acquisition was set up with 25 laser shots per 20 locations and 40 laser shots per 50 locations of laser positioning on the sites respectively. An unbiased random firing (raster) pattern was utilised and no sub-spectra were excluded from the final accumulated spectrum. The same Opti-Tof holder was used for both the Std-MALDI and X3 Biochip inserts.

2.3.3.d Database Searching and Presentation of the PMF Identification Data

Processing of the data for Enolase and HSA (SwissProt P00924 and P02769) was conducted using the Mascot search engine against the SwissProt database with 0.2 Da mass error for MS with potential modification of Methionine oxidation and alkylation of cysteine residues with iodoacetamide. Spectra were manually interrogated for presentation as shown in Figures 2.2.3 – 2.2.8.

2.3.4 Limit of Detection Comparison of Phosphorylated Peptides Utilising the Concentrating and Affinity Capture (NTA3) Biochip with Immobilised Metal Affinity Chromatography within an Applied Biosystems 4700 Analyser

NTA3 Biochips composed of 10%, 15% and 20% exposed carbonyl groups (COOH) in the capture (liquid retention) zone, were activated to covalently link the nitrilotriacetic acid (NTA) group to these COOH groups via an 1-Ethyl-3-(3-

dimethylaminopropyl) carbodiimide hydrochloride (EDC) and N-hydroxysuccinimide (NHS) mediated reaction before every experiment. The sample with phosphorylated peptides and non-phosphorylated peptides were added to the wells and allowed to incubate for approximately 20 min. All other contaminants were then washed away and the remaining bound species (phosphorylated peptides) were decoupled from the NTA-Fe^{3+} activated capture zone by changing the pH to an acidic environment and allowing evaporation and concentration of the sample with matrix into the central analysis zone for crystallisation and mass spectral analysis. Below in Table 2.1 is the final step-by-step protocol for conducting such an experiment that produced the optimal results.

Two particular NTA3 Biochips utilised are highlighted in Table 2.1.2 and 2.1.3, representing four experiments across the surface analysed to establish the LOD of the Biochips for affinity capture of phosphorylated peptides in the ABI 4700. These experiments were prepared at the Lumicyte laboratories in San Jose, California, USA, and the sample were analysed at the ABI demonstration labs in Fremont, California, USA. Two tables showing the layout of the four experiments are below.

1. NTA3 Biochip # 1, experiment # 1, is a dilution series of two phosphorylated peptides T6 and T2.1.
2. NTA3 Biochip # 1, experiment # 2, is a dilution series of the two phosphorylated peptides T6 and T2.1 in the presence of BSA in 10 fold molar excess.
3. NTA3 Biochip # 2, experiment # 3, is a dilution series of the digested phospho-protein, β-casein.
4. NTA3 Biochip # 2, experiment # 3, is a dilution series of the digested phospho-protein β-casein in the presence of BSA in 10 fold molar excess.

1	Pre-wash	10 µl per well of 3% NH_4OH for 5 min
2	Blow Dry	N_2 gas
N.B.		Once liquid is added to the CDVW's, do not allow them to dry out till the end of the activation and binding – just before the decoupling and concentration with the addition of matrix procedure
4	Equilibrate	20 µl per well of 25 mM NaPi, pH 8 in 0.1 % OBG for 10 min
5	Wash 1	10 µl per well of 0.1% OBG for 2 min, mixing up and down 3 times when removing
6	Activation	10 µl per well of 50 mM NHS and 200 mM EDC in 0.1 % OBG for 20 min
7	Wash 2	10 µl per well of 0.1 % OBG for 2 min, mixing up and down 3 times when removing
8	Immobilise	10 µl per well of 20 mM AB-NTA in 25 mM NaPi at pH 8 in 0.1 % OBG for 30 min
9	Wash 3	10 µl per well of 0.1 % OBG for 2 min, mixing up and down 3 times when removing
10	Wash 4	10 µl per well of 100 mM AcOH in 0.1% OBG for 2 min, mixing up and down 3 times when removing
11	Charge 1a	9 µl per well of 1 mM AcOH at pH 3 in 0.1% OBG
12	Charge 1b	1 µl per well of 0.1 mM $FeCl_3$
13	Wash 5	10 µl per well of 100 mM AcOH in 0.1 % OBG for 2 min, mixing up and down 3 times when removing
14	Wash 6	10 µl 100 mM AcOH, in 1 M Urea and 0.1 % OBG for 2min, mixing up and down 3 times when removing
15	Bind	5 µl per well of sample in 100 mM AcOH, in 1 M Urea and 0.1 % OBG for 20 min
16	Wash 7	10 µl per well of 100 mM AcOH, in 1 M Urea in 0.1 % and OBG for 2 min, mixing up and down 3 times when removing
17	Wash 8	10 µl per well of 100 mM AcOH for 2 min, mixing up and down 3 times when removing
18	Dry	Ambient conditions or under vacuum
19	Pre-elute	2 µl 9:1 ACN:0.1% PA and dry on bench top
20	Co-concentrate Matrix	2 µl of matrix 0.126 mg/ml CHCA in a ratio 84:13:3 - ACN : EtOH : 0.1% TFA (5 mM dibasic ammonium citrate)
21	Dry	Ambient conditions or under vacuum

Table 2.1.1 Step-by-step methodology that was co-developed by LCI and myself over a period of 6 months for the NTA3 Biochip. This method presented the optimal results as shown in the results section that follows.

	NTA3 #1	1	2	3	4	5
	NTA	No	Yes	Yes	Yes	Yes
	Fe	No	Yes	Yes	Yes	Yes
	Cal Mix	Yes	No	No	No	No
A	T6	No	100 fmol	100 fmol	50 fmol	50 fmol
	T1-2	No	100 fmol	100 fmol	50 fmol	50 fmol
	BSA	No	No	No	No	No
	Focus	84:13:3	84:13:3	84:13:3	84:13:3	84:13:3
	NTA	Yes	Yes	Yes	Yes	Yes
	Fe	Yes	Yes	Yes	Yes	Yes
	Cal Mix	No	No	No	No	No
B	T6	10 fmol	10 fmol	1 fmol	1 fmol	0.75 fmol
	T1-2	10 fmol	10 fmol	1 fmol	1 fmol	0.75 fmol
	BSA	No	No	No	No	No
	Focus	84:13:3	84:13:3	84:13:3	84:13:3	84:13:3
	NTA	Yes	Yes	Yes	Yes	Yes
	Fe	Yes	Yes	Yes	Yes	Yes
	Cal Mix	No	No	No	No	No
C	T6	0.75 fmol	0.5 fmol	0.5 fmol	100 fmol	100 fmol
	T1-2	0.75 fmol	0.5 fmol	0.5 fmol	100 fmol	100 fmol
	BSA	No	No	No	1000 fmol	1000 fmol
	Focus	84:13:3	84:13:3	84:13:3	84:13:3	84:13:3
	NTA	Yes	Yes	Yes	Yes	Yes
	Fe	Yes	Yes	Yes	Yes	Yes
	Cal Mix	No	No	No	No	No
D	T6	50 fmol	50 fmol	10 fmol	10 fmol	1 fmol
	T1-2	50 fmol	50 fmol	10 fmol	10 fmol	1 fmol
	BSA	500 fmol	500 fmol	100 fmol	100 fmol	10 fmol
	Focus	84:13:3	84:13:3	84:13:3	84:13:3	84:13:3
	NTA	Yes	Yes	Yes	Yes	Yes
	Fe	Yes	Yes	Yes	Yes	Yes
	Cal Mix	No	No	No	No	No
E	T6	1 fmol	0.75 fmol	0.75 fmol	0.5 fmol	0.5 fmol
	T1-2	1 fmol	0.75 fmol	0.75 fmol	0.5 fmol	0.5 fmol
	BSA	10 fmol	7.5 fmol	7.5 fmol	5 fmol	5 fmol
	Focus	84:13:3	84:13:3	84:13:3	84:13:3	84:13:3

Table 2.1.2 Chip 1, NTA3 Biochip with two dilution series across the surface to highlight the affinity capture and LOD of the Biochip for the detection of phosphorylated peptides T6 and T2.1, both pure (blue) and in the presence of a contaminant, BSA in 10 molar excess (red). The recorded amount of sample and contaminant applied is the total number of moles and the Yes/No recording indicate if the particular species is applied or not to the surface. Focus means the type of solution the matrix was applied in. Cal Mix is analogous to Pep Mix, being the peptide standard for calibration.

	NTA3 #2	1	2	3	4	5
	NTA	No	Yes	Yes	Yes	Yes
	Fe	No	Yes	Yes	Yes	Yes
	Cal Mix	Yes	No	No	No	No
A	Casein	No	100 fmol	100 fmol	50 fmol	50 fmol
	BSA	No	No	No	No	No
	Focus	84:13:3	84:13:3	84:13:3	84:13:3	84:13:3
	NTA	Yes	Yes	Yes	Yes	Yes
	Fe	Yes	Yes	Yes	Yes	Yes
	Cal Mix	No	No	No	No	No
B	Casein	10 fmol	10 fmol	1 fmol	1 fmol	0.75 fmol
	BSA	No	No	No	No	No
	Focus	84:13:3	84:13:3	84:13:3	84:13:3	84:13:3
	NTA	Yes	Yes	Yes	Yes	Yes
	Fe	Yes	Yes	Yes	Yes	Yes
	Cal Mix	No	No	No	No	No
C	Casein	0.75 fmol	0.5 fmol	0.5 fmol	100 fmol	100 fmol
	BSA	No	No	No	1000 fmol	1000 fmol
	Focus	84:13:3	84:13:3	84:13:3	84:13:3	84:13:3
	NTA	Yes	Yes	Yes	Yes	Yes
	Fe	Yes	Yes	Yes	Yes	Yes
	Cal Mix	No	No	No	No	No
D	Casein	50 fmol	50 fmol	10 fmol	10 fmol	1 fmol
	BSA	500 fmol	500 fmol	100 fmol	100 fmol	10 fmol
	Focus	84:13:3	84:13:3	84:13:3	84:13:3	84:13:3
	NTA	Yes	Yes	Yes	Yes	Yes
	Fe	Yes	Yes	Yes	Yes	Yes
	Cal Mix	No	No	No	No	No
E	Casein	1 fmol	0.75 fmol	0.75 fmol	0.5 fmol	0.5 fmol
	BSA	10 fmol	7.5 fmol	7.5 fmol	5 fmol	5 fmol
	Focus	84:13:3	84:13:3	84:13:3	84:13:3	84:13:3

Table 2.1.3 Chip 2, NTA3 Biochip with two dilution series across the surface to highlight the affinity capture and LOD of the Biochip for the detection of phosphorylated peptides within Casein (digested), both pure (green) and in the presence of a contaminant, BSA in 10 molar excess (purple). The recorded amount of sample and contaminant applied is the total number of moles and the Yes/No recording indicate if the particular species is applied or not to the surface. Focus means the type of solution the matrix was applied in. Cal Mix is analogous to Pep Mix, being the peptide standard for calibration.

2.3.4.a Preparation of Phosphorylated Peptides and Peptides Mixtures

All peptides and proteins were diluted and kept at -20^{o}C as per the manufactures specification and were made fresh on the day from these frozen standards by thawing to RT, solubilised and diluted in 100 mM AcOH, 0.1% OBG and 1 M Urea to the working concentrations.

2.3.4.b Mass Spectrometric Parameters (4700 TOF/TOF)

The same methods outlined earlier for single peptide analysis of non-phosphorylated peptides used on the 4700 in Sydney were applied to the phosphorylated peptide analysis used on the 4700 in California.

2.3.5 Limit of Detection and Capabilities of the Concentrating Desalting (RP3) Biochip on Contaminated Peptide Samples within an Applied Biosystems 4700 Analyser

2.3.5.a 1D Gel Purification of a Standard Protein Mixture

A low molecular weight marker (LMWM) calibration kit (Amersham Bioscience, GE Biosciences, SWE) composed of 6 proteins (phosphorylase b, albumin, ovalbumin, carbonic anhydrase, trypsin inhibitor and α-lactalbumin) was applied to One Dimensional SDS Gel Electrophoresis (1D GE), NuPAGE® Novex 12% Bis-Tris (Invitrogen, USA). The initial preparation and solubilisation was conducted as per manufactures specification, a dilution series of the LMWM kit was prepared to give eight differing loading amounts. The first dilution was created by taking 10 µl of the LMWM solution and mixing it with 90 µl of boiling SDS-PAGE sample buffer, followed by another 1 in 10 dilution and then 3 subsequent 1 in 4 dilutions, representing a total amount of protein loaded per lane of 57.6, 5.76, 1.44, 0.36 and 0.09 µg of the LMWM kit. The running of the gel was performed as per manufacturers specifications. The gel was stained with Coomassie Brilliant Blue.

2.3.5.b Digestion of Proteins and Extraction of Peptides from Gel

Row 4 of the LMWM kit was manually cut from the lanes of the gel, representing the protein carbonic anhydrase at differing loading amounts, ranging from 8.3, 0.83, 0.2075, 0.051875, 0.01296875 µg, for lanes 2 to 6 respectively. Each well plug was placed in 500 µl of ACN : 25 mM NH_4HCO_3 (1:1) and shaken for 30 minutes. The supernatant was discarded and the process repeated three times. The gel plugs were dried out in a speed vac. The gel plugs had 10 µl of trypsin solution (15 ng/µl in 25 mM NH_4HCO_3) applied and then placed at 4^oC for 60 minutes. The gel plugs had an additional 20 µl of 25 mM NH_4HCO_3 applied and then incubated at 37^oC overnight.

2.3.5.c Utilisation of RP3 Biochip for Purification and Analysis

Each well was pre-activated with the application of 5 µl of ACN : 0.1% TFA (1:1). A portion of the digest supernatant (10 µl) was applied to one well and then mixed on surface with a pipette, and 5 µl of this mixture was transfer to the proceeding well (second well), which already had 5 µl of the pre-activation solution on the surface. The second well was mixed on the surface identical to the preceding well and 5 µl was transferred to the third well, which had 5 µl of the pre-activation solution. This on-chip manipulation utilizing three wells per supernatant was repeated for each of the gel plug dilutions. The layout of the RP3 Biochip is depicted in Table 2.1.4. The solution was allowed to incubate for 20 min on the Biochip at RT under a dust cover. Any remaining solution was removed from the wells after the 20 min incubation and the wash solution was applied, 10 µl of 0.1% TFA. The wash solution was removed after 10 – 30 seconds and a second fresh wash solution was applied and removed after 10 - 30 sec. The CDVWs were allowed to air dry under a dust cover. The matrix solution was applied to each CDVW, 2 µl of 0.063 mg/ml CHCA in a ratio 84:13:3 - ACN : EtOH : 0.1% TFA (5 mM dibasic ammonium citrate).

RP3	1	2	3	4	5
A	Not used	Lane 2 Dil 1	Lane 2 Dil 2	Lane 2 Dil 3	Not used
B	Lane 3 Dill 1	Lane 3 Dill 2	Lane 3 Dill 3	Lane 4 Dil 1	Lane 4 Dil 2
C	Not used	Lane 4 Dil 3	Lane 5 Dil 1	Lane 5 Dil 2	Lane 5 Dil 3
D	Lane 6 Dil 1	Lane 6 Dil 2	Lane 7 Dil 3	Pep-Mix	Pep-Mix
E	Not used	Pep-Mix	Not used	Pep-Mix	Pep-Mix

Table 2.1.4 RP3 Biochip layout for the LOD study on Carbonic Anhydrase extracted from a 1D Gel of a LMWM mixture. Lane 2 - 6 are the five dilutions of the sample applied to the respective lanes from most concentrated to least. Each lane had the protein digest applied to three wells marked as Dil 1 – 3. The wells marked Pep-Mix were the four peptide mixture with salt contaminants added. Well marked as not used were not used due to pre-scribing of the chips at manufacture and removing the concentration ability of these wells.

2.3.5.d Mass Spectrometric Parameters (4700 TOF/TOF)

Acquisition of data was conducted utilising the same methodology as outlined above for the detection of peptides employing the X3 Biochip.

2.3.5.e Database Searching and Presentation of the PMF Identification Data

All spectra were internally calibrated with a minimum of two peptides from a theoretical digest of carbonic anhydrase (P00921), though the majority of calibration performed utilised four peptides from the ones used (831.4835, 1012.5421, 1141.5286, 1581.8172, 1838.0071, 2198.2179, 2253.1549 m/z) for the PMF calibration. Processing of the data was conducted utilising the Mascot Daemon software and search engine against the SwissProt database for mammals with a mass error of 0.2 Da, with potential modifications of methionine oxidation. Spectra were manually interrogated for presentation as shown in Figures 2.4.2 – 2.4.4.

2.4 RESULTS AND DISCUSSION

2.4.1 Limit of Detection Comparison for Peptides Utilising the Concentrating (X3) Biochip Compared to Standard MALDI within an Applied Biosystems 4700 Analyser

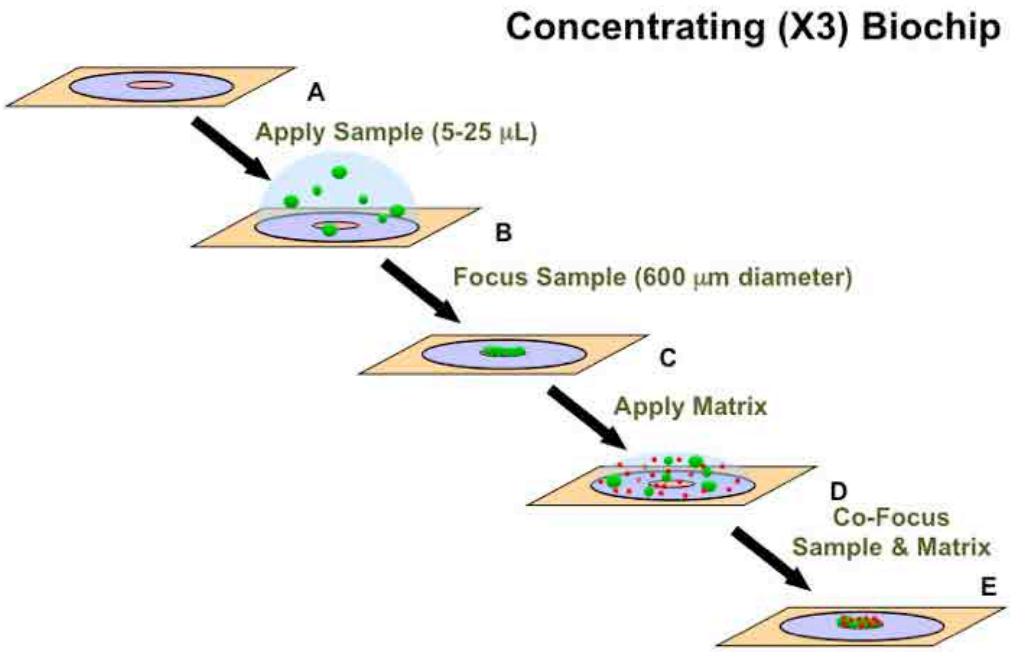


Figure 2.1.1 Schematic representation of the general methodology for the utilisation of the concentrating X3 Biochip on pure (clean) peptide samples. (A) Highlighting the three differing chemistries, analysis zone, liquid retention zone and boundary zone from centre projecting to the outside respectively. (B) The sample is applied to surface with a pipette and contained on the flat surface by the liquid retention zone (3 mm diameter), produced by the differing surface chemistry of the liquid retention zone, being intermediately hydrophilic compared to the external plate surface, being hydrophobic in nature. The peptides are shown as green balls within a liquid droplet shown as an opaque blue semi-sphere. Depending on the size of the CDVW and the composition of the solvents used, the recommended volume that can be contained and used for individual sample application is between 5 – 25 µl in this embodiment. (C) The sample is allowed to pre-concentrate under ambient conditions into the small analysis zone that is typically 600 µm in diameter and of the highest hydrophilic nature compared to the other surfaces. (D) The matrix solution is applied to the surface, denoted by the small red balls. It is believed that the original sample is re-solubilised and a co-migratory event and washing event, due to variable surface tension created by the various surface chemistries, to carry the peptides and matrix from the liquid retention zone into the analysis zone. This mechanism of action should also undertake a washing event that is different to the standard evaporative crawl seen on non-segmented surfaces. Refer to the power point movie supplied in the appendix that highlights this unique washing event. (E) The analyte and matrix are now concentrated in the analysis zone where the final stages of evaporation and crystallisation of the sample occurs, making it ready for MALDI MS analysis.

2.4.1.a Limit of Detection Comparison for Single Peptide Samples

The total signal strength generated from the mass spectrometer for each of the dilutions were plotted as shown in Figure 2.1.2 and 2.1.3. The LOD when using the X3 Biochip was 1 x 10^{-18} mol/µl, which represents a 10-fold (for ACTH) and 100-fold (for bradykinin, angiotensin I and neurotensin) increase in the sensitivity of detection ability of the mass spectrometer compared to the Std-MALDI approach. One of the main reasons for this observed increase in sensitivity could be due to a loading effect resulting from the increased working volume and thus the increased number of moles that can be applied to the Biochip compared to Std-MALDI. In these experiments the Biochip has 10 µl added to the surface, where the Std-MALDI uses 1 µl of a 1:1 mixture of analyte to matrix. This means that the Biochip allows for approximately 20 times more analyte of an equal concentration to be applied to the surface compared to this Std-MALDI method. Provided that there is minimal to no sample loss in the concentration event, the X3 Biochip should present the increased analyte in the analysis zone for detection by the mass spectrometer, resulting in a proportionately higher ion count. As there is theoretically 20 times more analyte on the surface of the X3 Biochip one would expect approximately a 20 fold increase in sensitivity of detection limit, which does not entirely explain the 100-fold increase observed for three of the four peptides used.

This additional increase in sensitivity may also be attributed to the final size of the analysis zone, which is between 0.5 – 0.6 mm in diameter compared to the Std-MALDI of 1.4 – 2 mm in diameter, the difference in the areas between the two are in the ratio range of 1 : 7.8 and 1 : 11.1, respectively. This results in a higher concentration (density) of analyte per unit of surface area for the Biochip, supplying one reason for the increased signal strength observed that is non-proportional to the sample application amount differences. The working example (1) in Table 2.1.5 calculates the total theoretical number of analyte molecules per cubic µm for the X3 Biochip and Std-MALDI, highlighting that there is at least a 400-fold increase in the density of the analyte in the final analysis zone of the X3 Biochip compared to Std-MALDI. Similarly, the working example (2) in Table 2.1.6 calculates the total theoretical number of matrix molecules in volume, which is at least 80 times higher in the X3 Biochip compared to the Std-MALDI employed within. It is known that there is an optimal range for the matrix-to-analyte ratio (M/A), being from 1,000 – 100,000 : 1,

which would mean lowering the amount of matrix applied as you reduce the amount of analyte to be analysed [370]. The ratio of matrix-to-analyte in the two examples is approximately 47,000,000 : 1 for the Biochip and 200,000,000 : 1 for Std-MALDI. Both of these are greater than 100,000 with the Std-MALDI being four times more than the Biochip, though it does not account for the large difference in sensitivity. It would seem that either the ratio of matrix-to-analyte does not apply when you concentrate the sample and matrix as is done with the X3 Biochip, or there is a new ratio needed when the sample is concentrated. Either way, the contributing effect must be indistinguishable based on these experiments when compared to the greater density of the analyte alone, which is at least 400 times greater for the Biochip compared to Std-MALDI.

Lastly, the final crystal structure of the matrix and analyte may be altered in the X3 Biochip, whether it is due to the novel combination of solvents employed, the rate of crystallisation, the ratio of matrix to analyte used or the unique migration of analyte and matrix in the concentration process [42, 46]. Thus, it is to be expected that the shape and type of crystals generated by the X3 Biochip will be different to the Std-MALDI. Irrespective of the shape of the crystals generated, it is still a potential factor working towards the increase in sensitivity or even working in conjunction with the other variables mentioned above.

X3 Biochip	Std-MALDI
molecules = mol x Avogadro = (1 x 10^{-15} mol) x (6.02214179 x 10^{23} mol^{-1}) = 6.022214179 x 10^{8} molecules	molecules = mol x Avogadro = (0.05 x 10^{-15} mol) x (6.02214179 x 10^{23} mol^{-1}) = 3.01110709 x 10^{7} molecules
volume = π x r^2 x height = 22/7 x 0.3^2 x 0.05 = 0.014137 mm^3	volume = π x r^2 x height = 22/7 x 1.4^2 x 0.05 = 0.307876 mm^3
density = molecules x volume = 6.022214179 x 10^{8} molecules / 14.137 μm^3 = 42,598,954 molecules/μm^3	density = molecules x volume = 3.01110709 x 10^{7} molecules / 307.876 μm^3 = 97,803 molecules/μm^3

Table 2.1.5 Example 1: if a 10 µl sample of peptides at a concentration of 100 attomol/µl (1 femtomole in total) was applied to an X3 Biochip and allowed to concentrate, that would equate to approximately 42,598,954 molecules/μm^3 while for Std-MALDI at 0.5 µl would be 97,803 molecules/μm^3. This highlights that there is theoretically greater than 400 times the density of analyte molecules for the X3 Biochip after concentration compared to Std-MALDI.

X3 Biochip	Std-MALDI
mass = density x volume = (0.126 x 10^{-3} g) x (2 x 10^{-6} l) = 2.52 x 10^{-10} g	mass = density x volume = (0.126 x 10^{-3} g) x (0.5 x 10^{-6} l) = 6.3 x 10^{-11} g
n = mass x F.W. = (2.52 x 10^{-10} g) x (189.17 g mol^{-1}) = 4.767084 x 10^{-8} mol	n = mass x F.W. = (6.3 x 10^{-11} g) x (189.17 g mol^{-1}) = 1.191771 x 10^{-8} mol
molecules = mol x Avogadro = (4.767084 x 10^{-8} mol) x (6.02214179 x 10^{23} mol^{-1}) = 2.870840 x 10^{16} molecules	molecules = mol x Avogadro = (1.191771 x 10^{-8} mol) x (6.02214179 x 10^{23} mol^{-1}) = 7.1771 x 10^{15} molecules
volume = π x r^2 x height = 22/7 x 0.3^2 x 0.05 = 0.014137 mm^3	volume = π x r^2 x height = 22/7 x 1.4^2 x 0.05 = 0.307876 mm^3
density = molecules x volume = 2.870840 x 10^{16} molecules / 14.137 μm^3 = 2.0 x 10^{15} molecules/μm^3	density = molecules x volume = 7.1771 x 10^{15} molecules / 307.876 μm^3 = 2.3 x 10^{13} molecules/μm^3

Table 2.1.6 Example 2: if 2 µl of the matrix solution for the X3 Biochip (0.126 mg/ml) was applied and allowed to concentrate there would be 2 x 10^{15} molecules/μm^3 of matrix, while for Std-MALDI at 0.5 ml (4 mg/ml) would be 2.3 x 10^{13} molecules/μm^3 of matrix. This highlights that there is theoretically greater than 80 times the density of matrix molecules for the X3 Biochip after concentration compared to Std-MALDI.

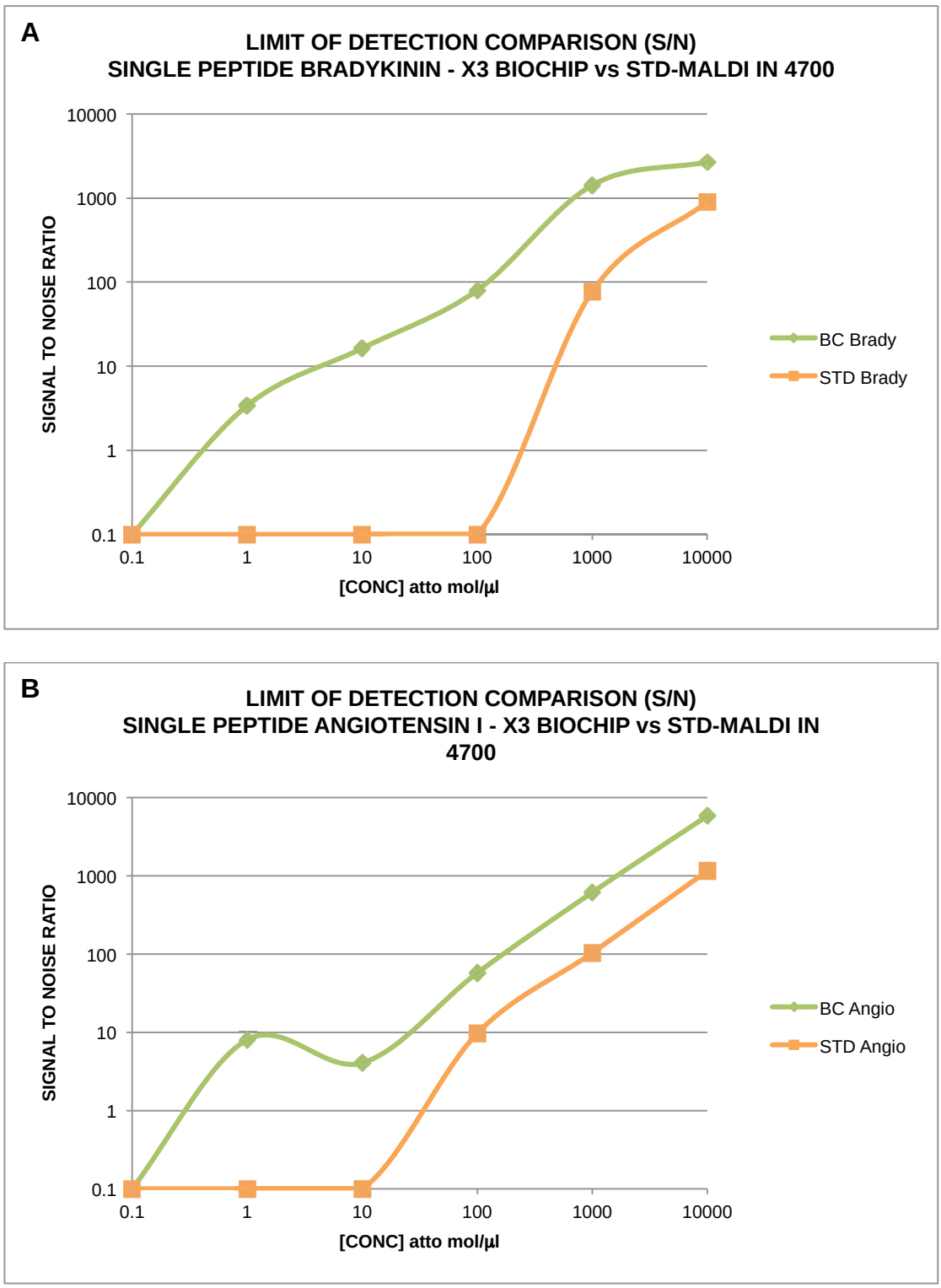


Figure 2.1.2 Comparison of the total signal strength generated when using an X3 Biochip compared to Std-MALDI over a dilution series for single peptide samples. The y-axis denotes the mean signal to noise ratio generated for each peptide (3 replicates), whilst the x-axis for individual concentrations of the peptides used in the study in atto (10^{-12}) moles per micro litre. (A) The top graph shows the response relationship when using the peptide Bradykinin (m/z=1061). (B) The bottom graph shows the response relationship when using the peptide Angiotensin I (m/z=1297). Both peptides show the X3 Biochip producing a 10 fold increase in sensitivity of detection compared to Std-MALDI which drops off between 100 – 10 attomol/µl while the X3 Biochip signal drops off between 1 – 0.1 attomol/µl.

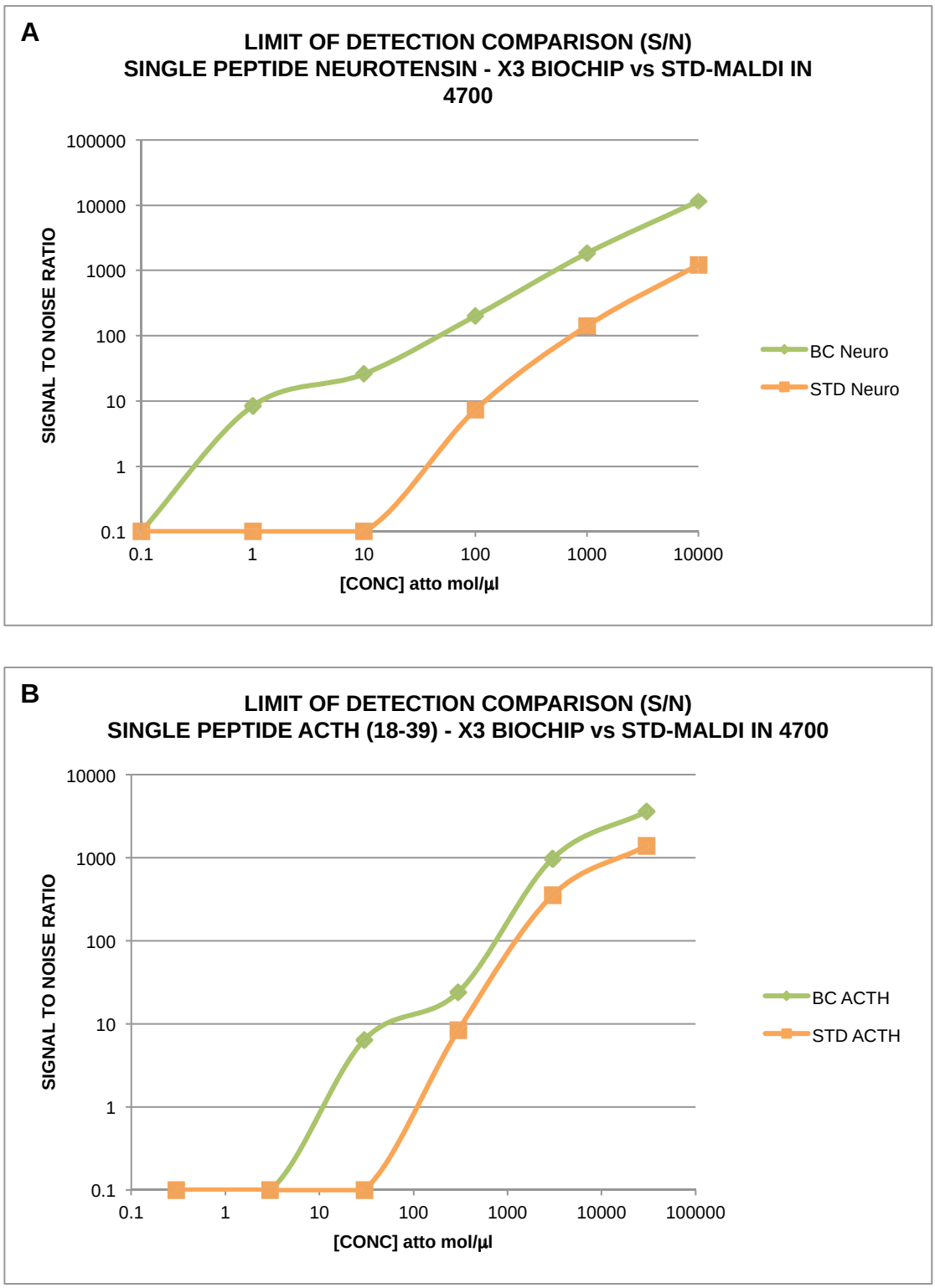


Figure 2.1.3 Comparison of the total signal strength generated when using an X3 Biochip compared to Std-MALDI over a dilution series for single peptide samples. The y-axis denotes the mean signal to noise ratio for each peptide (3 replicates), whilst the x-axis represents the individual concentrations of the peptides used in the study in atto (10^{-12}) moles per micro litre. (A) The top graph shows the response relationship when using the peptide Neurotensin (m/z=1673). The X3 Biochip produces a drop off in signal between 10 – 0.1 attomol/µl and 100 – 10 attomol/µl for Std-MLADI. (B) The bottom graph shows the response relationship when using the peptide ACTH 18-39 (m/z=2465). The X3 Biochip produces a drop off in signal between 30 – 3 attomol/µl and 300 – 30 attomol/µl for Std-MALDI.

2.4.1.b Sensitivity of Detection Comparison for a Four Peptide Mixture

The total signal strength generated from the mass spectrometer for each of the peptides in the four peptide mixture across the dilutions were plotted as shown in Figure 2.1.4. The LOD for the X3 Biochip was 10 x 10^{-18} mol/µl for Bradykinin, Angiotensin I and Neurotensin while for ACTH it was 30 x 10^{-18} mol/µl, which is 10 fold lower than that for the single peptide experiments conducted earlier. The relative differences in the LOD between the X3 Biochip and Std-MALDI were still 100-fold for the peptides Bradykinin, Angiotensin I and Neurotensin while for ACTH 18-39 it was 10 fold.

The difference in absolute LOD observed here for the four peptide mixture compared to the single peptide analysis may be attributed to ion suppression events brought on by the mixture, similar to the ion suppression event created by the addition of an internal standard, hence reducing the ionisation efficiencies of the individual peptides when in a mixture [371]. The similar relative enhanced sensitivity differences of 100 and 10 fold observed for the four peptide mixture, as was seen with the single peptide analysis when comparing the X3 Biochip to Std-MALDI, are explained by the same factors discussed earlier for the single peptide comparison experiment.

Upon deeper manual inspection of the spectra generated from the four peptide mixtures ionised from either of the two surfaces (X3 Biochip or Std-MALDI), comparable isotopic resolution of signal with negligible discernable differences at both high and low concentrations was observed, shown in Figure 2.1.5. The same was observed for the single peptide samples (results not shown). The observed generation of isotopic resolution suggest that the surface of the SAM is of comparable flatness to the Std-MALDI metal surface and that the SAM is not adversely affecting the ability of the mass spectrometer to resolve the peaks at the isotopic level.

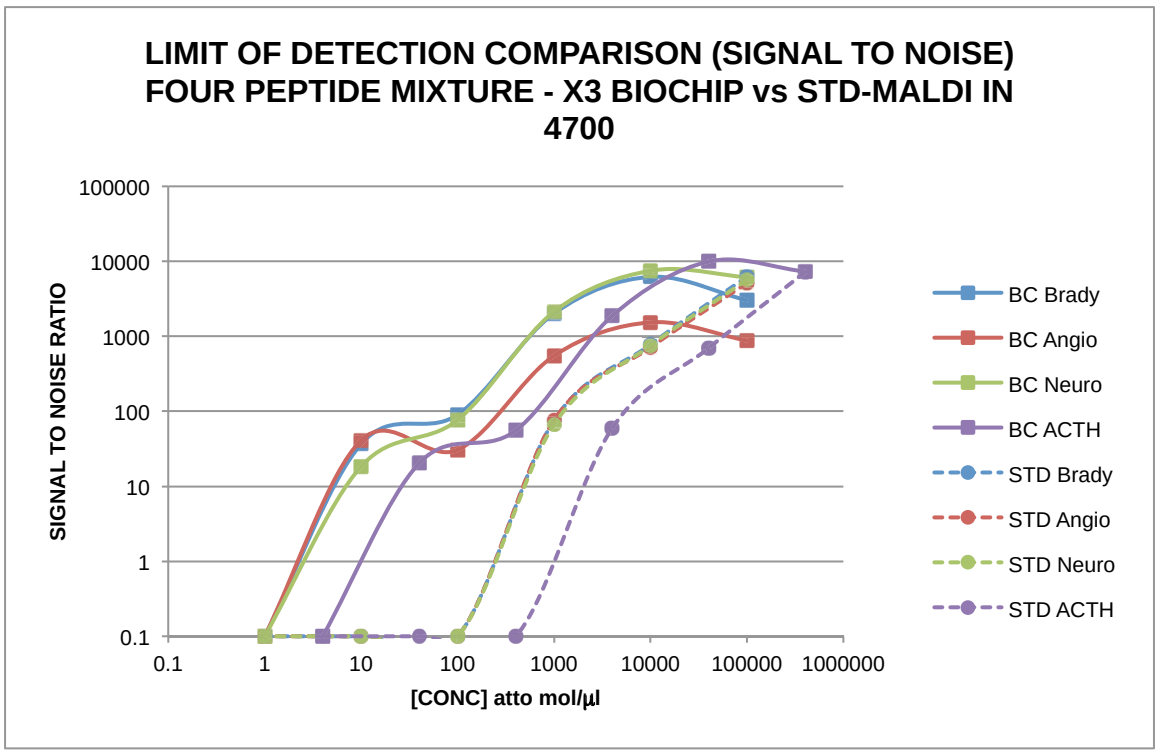


Figure 2.1.4 Comparison of the total signal strength generated when using an X3 Biochip compared to Std-MALDI over a dilution series on a four peptide mixture of Bradykinin (m/z=1061), Angiotensin I (m/z=1297), Neurotensin (m/z=1673) and ACTH 18-39 (m/z=2465) in the ratio, 1:1:1:4 respectively. The y-axis denotes the mean signal to noise ratio for each peptide (3 replicates), whilst the x-axis represents the individual concentrations of the peptides used in the mixture across the dilution series in atto (10^{-12}) moles per micro litre. The absolute LOD for each peptide is 10 fold less than the corresponding signal generated for the peptide when analysed alone. The X3 Biochip produces 100 fold increase in the LOD for Bradykinin, Angiotensin I and Neurotensin with the signal dropping off between 10 - 1 attomol/µl compared to Std-MALDI which was between 1000 – 100 attomol/µl. While ACTH 18-39 produced a 10 fold increase in the LOD with the signal dropping off between 40 – 4 attomol/µl compared to Std-MALDI, which was between 400 – 40 attomol/µl.

When comparing the spectra at very low concentrations, typically at or below 100 x 10^{-18} mol/µl (10 attomole in total), the sporadic appearance of polymer peaks when using the X3 Biochips was observed. These polymer peaks are evenly spaced by 44 Da and were observed in the range of 1300-1800 m/z, thereby disrupting the visualisation of peaks of low signal intensity in spectra obtained from samples of 10^{-18} to 10^{-21} mol/µl of peptides, as shown in Figure 2.1.6. It was observed that whenever the polymer appeared on the Biochips, it was less prevalent on the Std-MALDI surface. The most likely reason for this lies in the concentration event of the X3 Biochip, which, while increasing the signal intensity of the peptide due to concentration of the analyte on the surface, will also concentrate any contaminants similarly. This may lead to the greater observed prevalence and deleterious effect (ion suppression) of polymers when using the Biochip rather than Std-MALDI. The

size of the polymer units and the known molecular weight of the SAM indicate that the polymer is not from the SAM dislodging off the surface due to the laser or other factors in the ionisation/desorption process [372]. Polymers are ubiquitous contaminants and even though their appearance seemed to be a random event, it reinforces the need for the highest quality of all reagents and handling equipment when conducting these types of experiments. Pipette tips that are made of virgin polypropylene and void of any coatings are imperative to the successful use of the Biochips. The occurrence of these polymers in conjunction with the use of these Biochips made many of these experiments unreadable. Despite our best efforts to identify a single source of contaminant we were not able to do so conclusively, which strongly suggests that they were an artefact of the manufacturing process, and hence unavoidable.

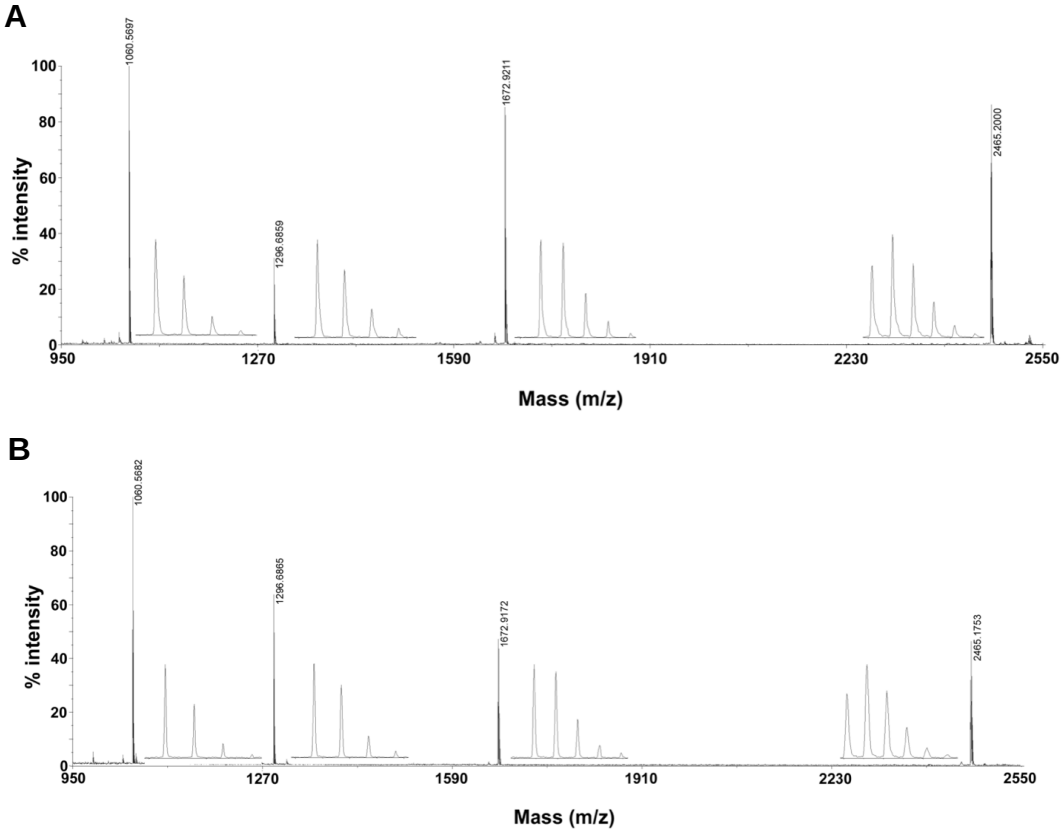


Figure 2.1.5 Example of the quality of mass spectrum for the four peptides Bradykinin (m/z=1061), Angiotensin I (m/z=1297), Neurotensin (m/z=1673) and ACTH 18-39 (m/z=2465), which were used in both the individual LOD analysis and the four peptide mixture LOD analysis. Both spectra have superimposed next to each peak a zoomed-in view of the quality of the spectra, which on visual inspection highlight no discernable difference between each surface and present sharp and definable isotopic resolution. The y-axis is the signal intensity expressed in percentage of strength and the x-axis is the m/z ratio. (A) X3 Biochip with 10 µl of peptides at 1 femtomol/µl (10 femtomoles added). (B) Std-MALDI plate with 1 µl of peptides at 5 femtomol/µl (5 femtomoles added).

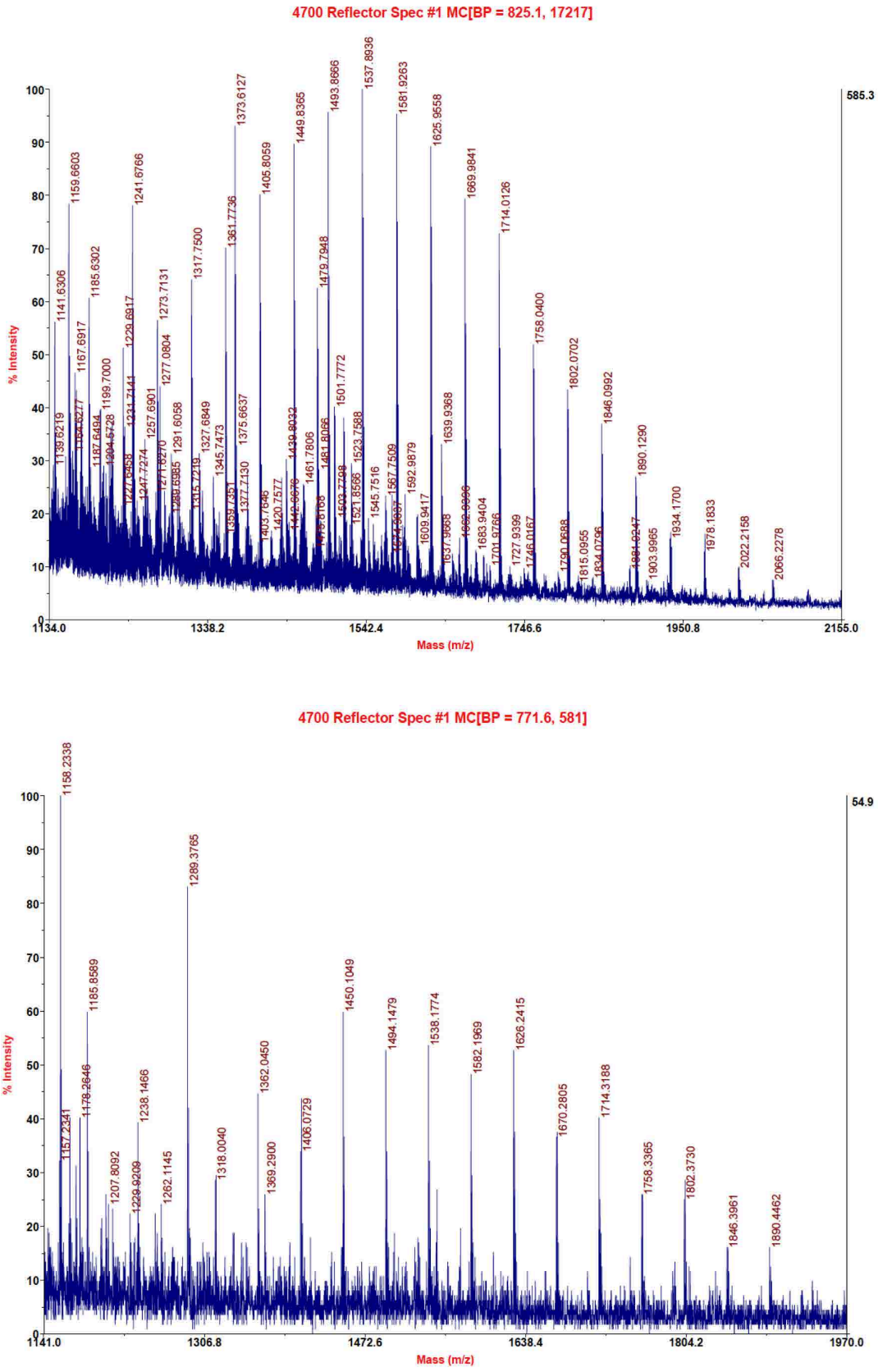


Figure 2.1.6 Example of the quality of spectra seen at various times on numerous occasions throughout the project. The repetitive peaks across the m/z range 1229-2066 (top) and 1229-1890 (bottom) with a mass difference of 44 Da is indicative of polymer contamination. This contamination was seen mostly when dealing close to the LOD at and around concentrations of 100 x 10^{-18} mol/µl and lower. It was also seen when dealing with peptides extracted from gel plugs or when incorrect plastics for sample handling were used, causing ion suppression and the omission of the data collected for any experiment containing such contamination.

When a comparison is made between the signal strength and the total number of moles added to the different surfaces, as opposed to comparing signal strength versus concentration of the samples applied, the differences in the observed limits of detection are now less than or equal to 5 times as indicated in Figure 2.1.7. The linear regression for the data is greater than 0.98 for both forms of analysis across the scale presented. The high linearity in conjunction with the similar ion count reinforces the observation that the SAM has no discernable deleterious affect on the ionisation and desorption of the analyte at these amounts of analyte. This data also reinforces the point that the SAM in the analysis zone is resistant to analyte binding. While this is not a detailed study on the mechanism of how the ionisation and desorption process works with SAMs, the results suggest that using a SAM surface has no discernable negative or differential effect on the ionisation and desorption abilities of the peptides we examined.

Thus, if SAMs are employed in concentric circles with variable exposed surface chemistries of variable hydrophilicity, there is a discernable and differential effect, which will result in increased LOD, though at a disproportionately lower effect than the loading effect of comparing absolute analyte amounts. Since there are 20 times more volume (moles) of analyte added to the surface of the X3 Biochip compared to Std-MALDI, there is less than 20 times, actually only a 5-fold increase in the sensitivity of detection.

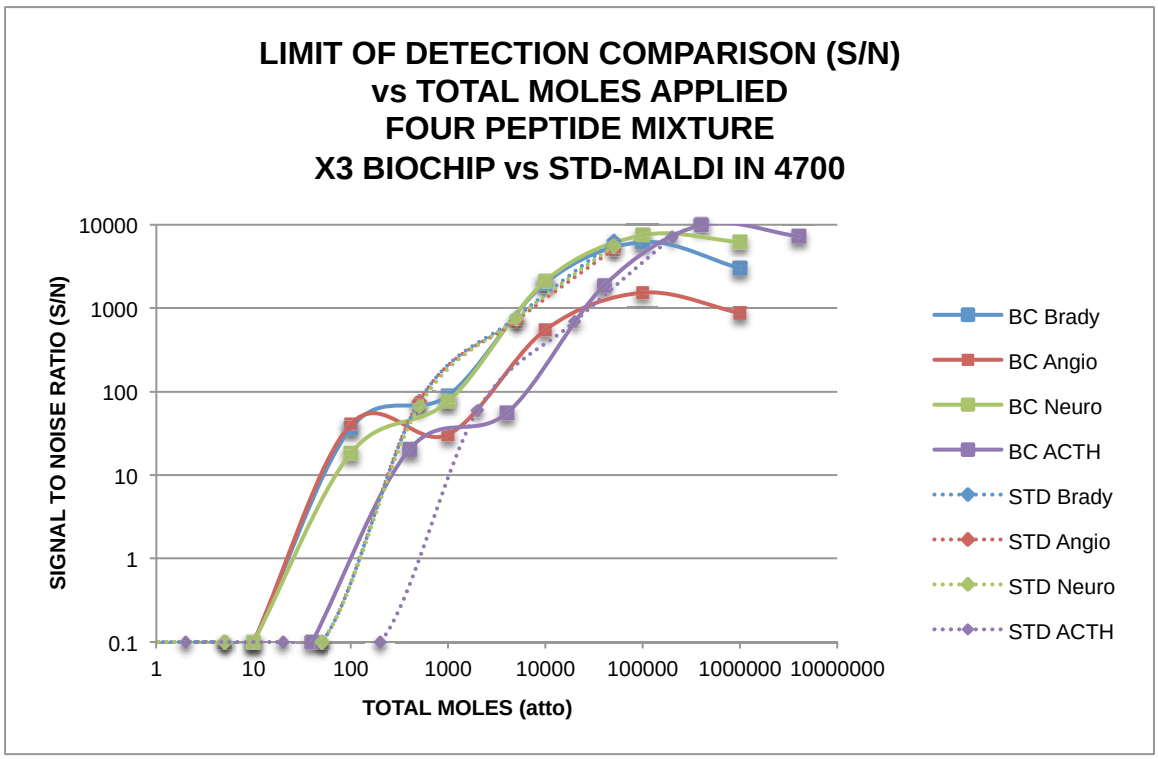
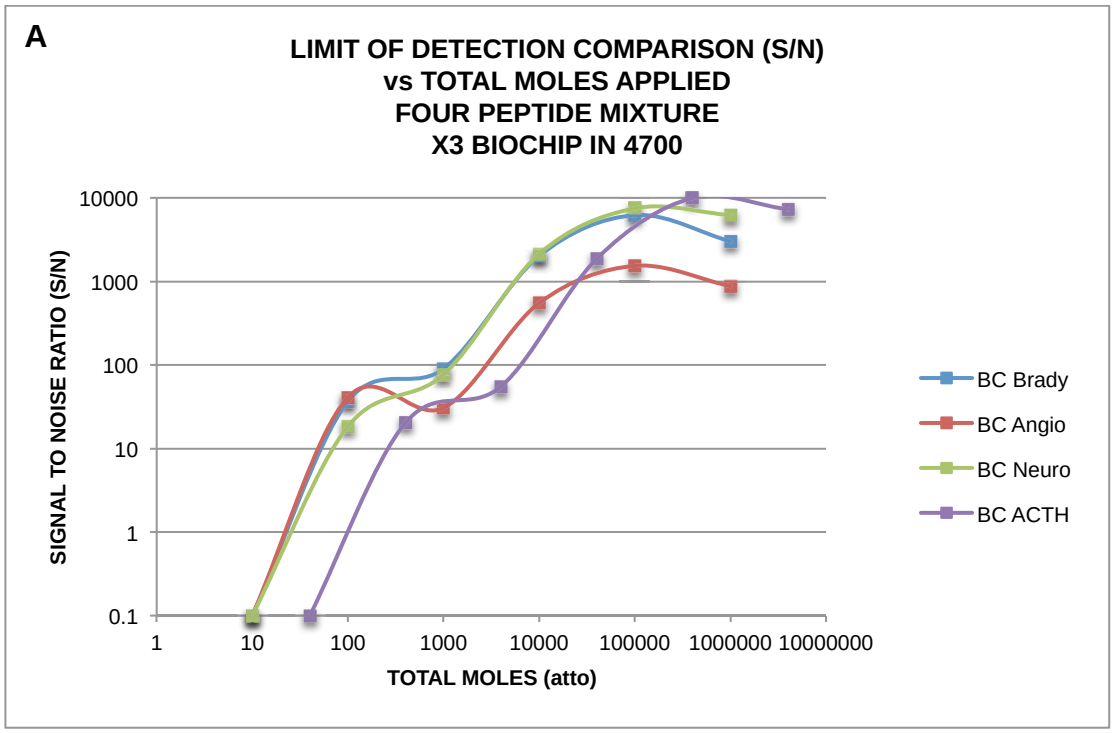


Figure 2.1.7 Comparison of total signal strength generated when using an X3 Biochip compared to Std-MALDI with a four-peptide mixture, Bradykinin (m/z=1061), Angiotensin I (m/z=1297), Neurotensin (m/z=1673) and ACTH 18-39 (m/z=2465) in the ratio 1:1:1:4 respectively. The y-axis denotes the mean signal to noise ratio (3 replicates), whilst the x-axis represents the total number of moles added, as opposed to expressing it in concentration as in the earlier graphs. This different way of presenting the data, so as the loading affect of the X3 Biochip is removed, provides a clearer interpretation of the analysis. The resultant LOD is less than or equal to a 5 fold increase for the X3 Biochip compared to Std-MALDI



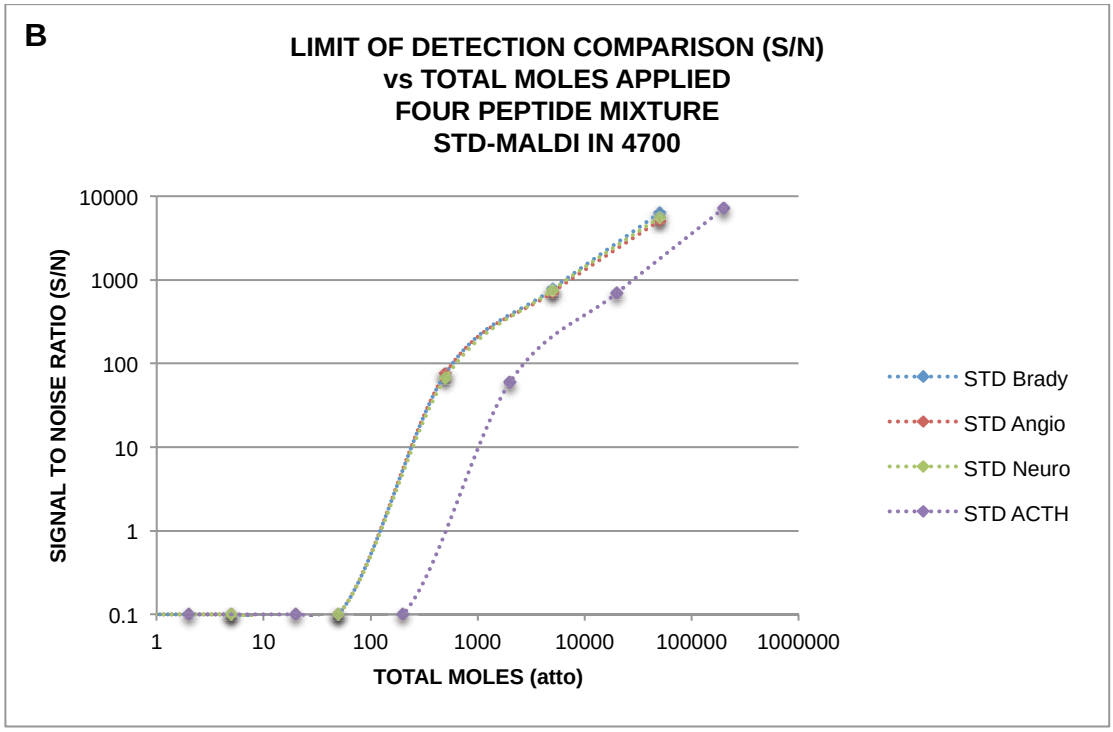


Figure 2.1.8 Due to the overlap of the lines on the preceding graph, (Figure 2.1.7) they have been presented here in two parts. The y-axis denotes the mean signal to noise ratio (3 replicates), the x-axis represents the total number of moles added. (A) The signal response shown by the X3 Biochip alone with the Std-MALDI data omitted and shown below. (B) The signal response shown by Std-MALDI alone with the X3 Biochip data omitted and shown above.

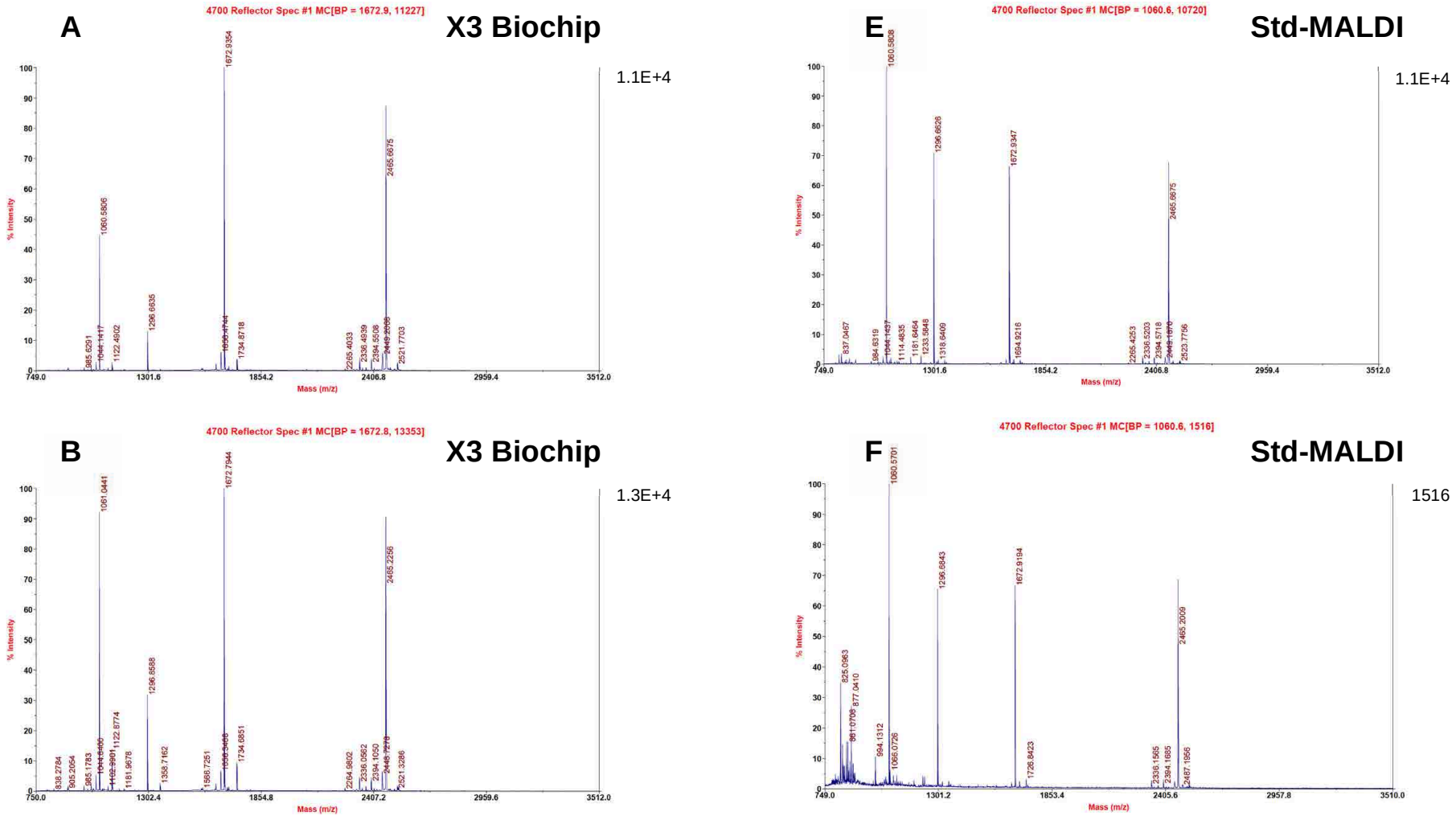


Figure 2.1.9 Spectra from the four peptide mixture at concentration 1 x 10^{-13} mol/µl for (A) and (E) being X3 Biochip and Std-MALDI respectively, both showing a similar ion count for the largest peak, being 1.1 x 10^{4}. Concentration 1 x 10^{-14} mol/µl for (B) and (F) being X3 Biochip and Std-MALDI respectively, the difference in ion count is being to become evident with the highest peak being 1.3 x 10^{4} for X3 Biochip and 1516 for Std-MALDI. Note the ionization ratio differences, the X3 Biochip shows suppression of m/z 1060 and 1269 compared to m/z 1672 and 2465. Alternatively, Std-MALDI shows suppression of m/z 1672 and 2465 compared to m/z 1060 and 1269.

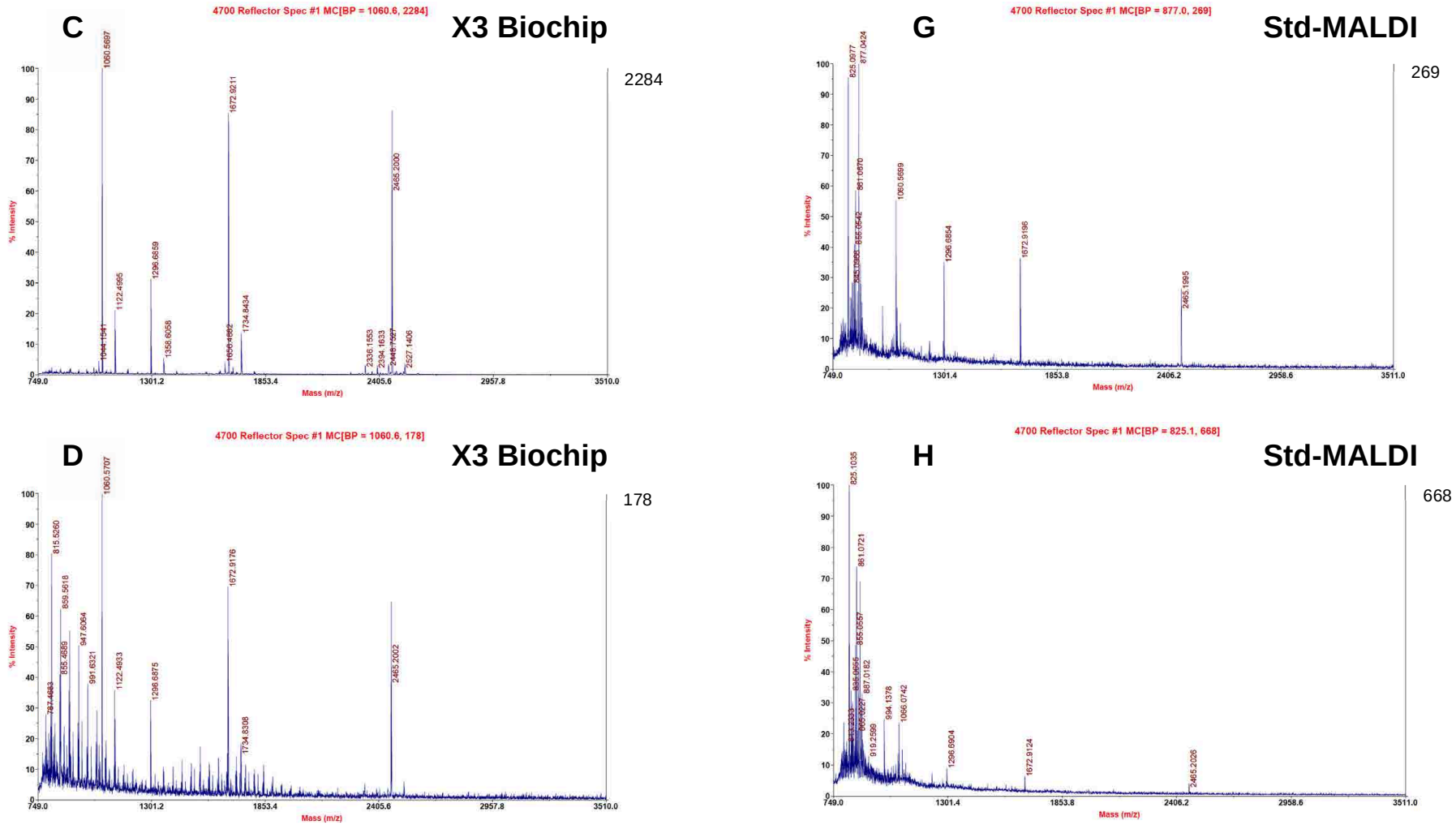


Figure 2.1.10 Spectra from the four peptide mixture at concentration 1 x 10^{-15} mol/µl for (C) and (G) being X3 Biochip and Std-MALDI respectively, difference in ion count is shown by the X3 Biochip displaying ~2000 ion counts for the highest ion compared to ~150 for Std-MALDI. Concentration 1 x 10^{-16} mol/µl for (D) and (H) being X3 Biochip and Std-MALDI respectively, difference in ion count is shown by the X3 Biochip displaying ~170 ion counts for the highest ion compared to ~50 for Std-MALDI. Note that the matrix adducts are masking out the signal for Std-MALDI compared to the matrix adducts for X3 Biochip.

2.4.2 Limit of Detection and Identification Comparison of Peptide Standards and Protein Digests utilising a Novel Hybridised AnchorChip/ABI 4800 MALDI Plate versus the Concentrating (X3) Biochip within an Applied Biosystems 4800 Analyser

2.4.2.a Incorporation of the Bruker AnchorChip and X3 Biochip into an Applied Biosystems 4800 Analyser

In order to extend the results generated thus far by the X3 Biochip against Std-MALDI and show that they were actually unique and differentiated to other concentration chips, I decided to compare the industry standard Bruker AnchorChip^{TM}, against the X3 Biochip platform.

The Bruker AnchorChip^{TM} was developed only to fit into Bruker mass spectrometers. While LCI wanted to develop their Biochips for all mass spectrometers, at this stage of development they only had designs for Shimadzu and ABI instruments. Rather than getting LCI to manufacture new chips to accommodate a Bruker instrument, I decided to work with an ABI instrument and try to modify a Bruker AnchorChip^{TM} to work within this system. The latest ABI 4800 had just been released with LC-MALDI applications considered a priority in its design. It had the ability to take sample plates that were of MTP format and shape. This meant that the Bruker AnchorChip^{TM} (which did come in MTP format) could possibly fit and work inside an ABI 4800. I designed, manufactured and tested a novel hybrid AnchorChip^{TM} / ABI 4800 plate holder, to generate the results presented within, with the assistance of the Macquarie University Engineering and Technical Services (METS) personnel and Dr Peter Milburn from the John Curtin Medical School at the Australian National University (ANU). Figure 2.2.1 shows the final hybrid-plate system assembled.

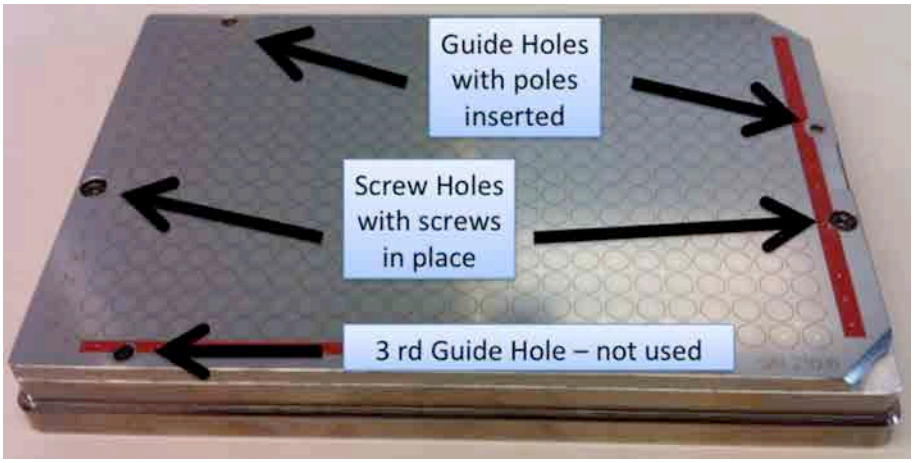


Figure 2.2.1 Picture of the final and assembled novel hybrid Bruker AnchorChip™ and ABI 4800 Plate Holder used in these experiments. Note that both the screws and the guide poles are below (flush with) the surface of the AnchorChip. The length and width of the AnchorChip are shown to be flush with the sides of the ABI 4800 Plate Holder and the final height of the hybrid system is the same as the original ABI 4800 holder with LC-MALDI plate attached (not shown). The final weight of the hybrid holder was 380 g, which is approximately 40% heavier than the original holder with an LC-MALDI plate attached.

A risk assessment was conducted with respect to the envisaged additional weight of the hybrid-plate system before the final design, manufacture and use within the 4800 was conducted. Primarily, the risk assessment was conducted with respect to not damaging either the hybrid-plate system, or more importantly the mass spectrometer. There was concern with the lifting mechanism or other internals of the instrument, specifically the chance of the plate being dropped due to weight issues, or the plate coming into contact with the internals of the instrument due to insufficient clearances. Since the external dimensions of the hybrid-plate system were designed to be identical to the original ABI 4800 plate, the risk due to clearance was mitigated. The shape of the bottom section of the hybrid-plate system had not been changed, and thus there was minimal risk associated with the grabbing mechanism being able to attach to and pick up the plate. Figure 2.2.2 shows the hybrid-plate system in the two pieces that were made and brought together to make a working system.

The weight difference, ~280 g for the standard plate, while the hybrid-plate system was ~380 g, was a potential risk. However, it was hypothesized that the design of the ABI 4800 instrument should be most likely able to accommodate the additional weight as the design principles for incorporation of safety factors in lab instrumentation typically allow for between 1.5 to at least 3 times the standard operational strain or weight of daily operation. Since the hybrid system was only 1.4

times heavier I went forward with the manufacture and my risk assessment argument was agreed to by Dr Peter Milburn who allowed me to use the 4800 at ANU based on these justifications.

There were no observable adverse issues for the mass spectrometer with respect to the generation of signal and resolution when tested on standards loaded on the hybrid plate holder. There did however seem to be a larger than normal clunking noise that occurred when the plate was taken from the loading bay to the grab arm, which I attribute to the extra weight and different surface properties of the hybrid-plate system.

Since the implementation of this design in 2006, the Australian Proteomics Analysis Facility (APAF) housed at Macquarie University (MQU), have implemented the manufacture and use of the hybrid-plate system as their standard operational procedure for MALDI mass spectrometry conducted within their lab on an ABI 4800. They have reported ease of use, increased LOD and signal strength for comparable samples on standard MALDI plates and no observable adverse affects on the instrument operation throughout this period.

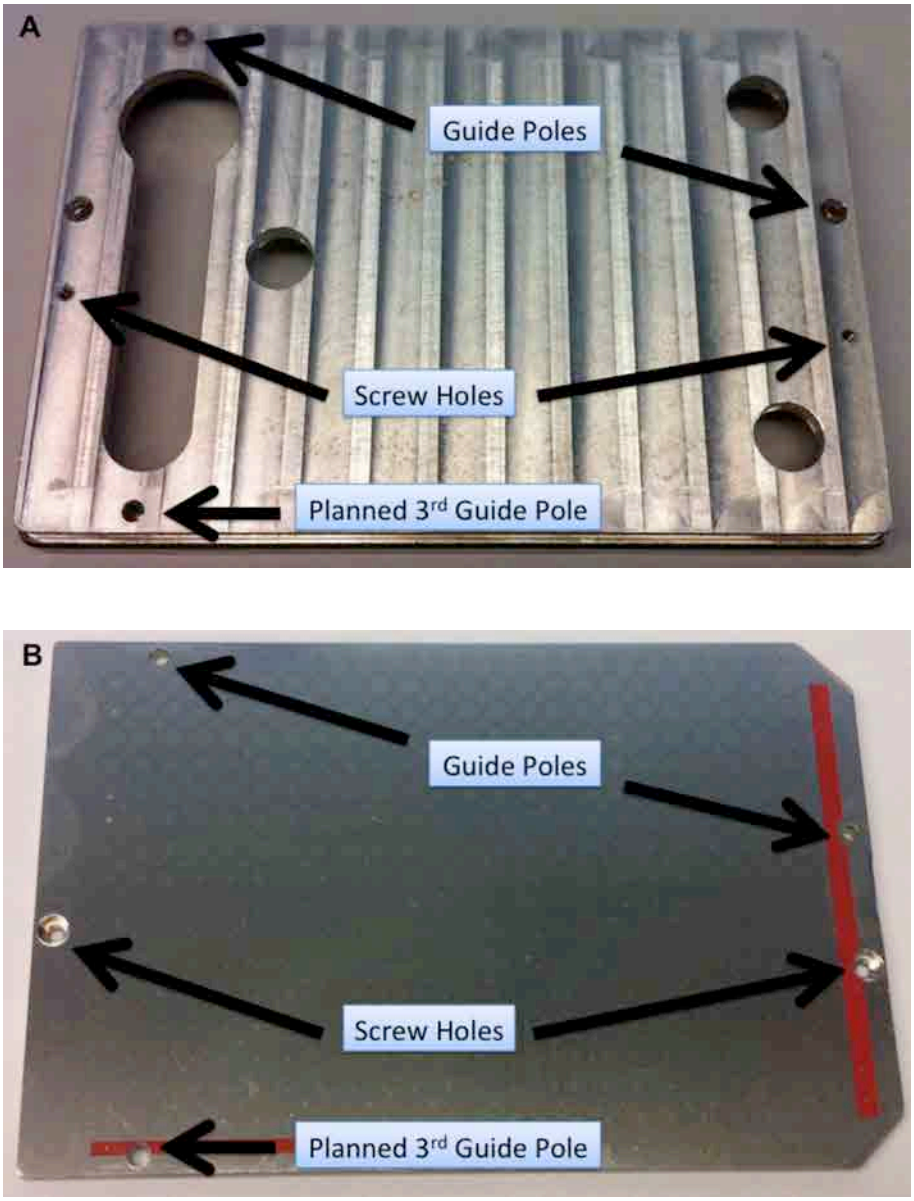


Figure 2.2.2 Picture of the modified 4800 plate holder that was developed to house the AnchorChip™ for use in the 4800 mass spectrometer. (A) There are three guide poles shown above, only two out of the planned three were used. The two that worked, shown on the top left and right hand side of the picture above were slightly different in shape and method of attachment to each other, though they were sufficient for this application. There are two screw holes shown above, it was deemed not necessary to create any more than two since there were guide holes/poles designed to work in conjunction with the screw holes for alignment and fixing to the surface. The hole directly above the left screw hole was a machining mistake that did not interfere with the use of the plate holder. Similarly the other larger holes were made from the removal of the magnets that were in the original holder and the need to create hold points for the fabrication process. The machine lines that run vertically down the plate holder are from the removal of the top surface to reduce the height of the holder so that when the AnchorChip is attached to this surface the final height of the combined product is the same as the original holder. (B) The AnchorChip has had the magnets removed from its under side. Three holes were drilled out for the guide poles for the modified holder to go into, but the third guide hole was not used. Two screw holes were drilled and counter sunk so that the screws would sit flush with surface of the chip once affixed. Despite the heavy nature of the AnchorChip is was decided not to machine the underside down or take out material from it due to possible heat transfer and metal distortion of the shape and wells on the chip from the process.

2.4.2.b MS/MS Identification Comparison of the Hybrid AnchorChip, X3 Biochip and Std-MALDI in an Applied Biosystems 4800 Analyser

With respect to the analysis of the Enolase digest shown in Figure 2.2.3, the X3 Biochip and AnchorChip outperform the Std-MALDI across the entire concentration gradient applied, as expected. The AnchorChip and Std-MALDI did generate a similar result at the 0.1 attomol/µl level, though on closer inspection of the slopes of the curves, it would seem that the AnchorChip generated an anomalous reading here, lower than trend scores at this concentration for unknown reasons. The X3 Biochip and AnchorChip generated a similar response at the high end of the concentration gradient (10000 – 1 attomol/µl). While at the lower end of the gradient (0.1 – 0.01 attomol/µl) the X3 Biochip outperforms the AnchorChip. The standard deviations in the Mascot scores generated are displayed in Figure 2.2.5 and are consistent for the X3 Biochip, generating reproducibility better than the AnchorChip^{TM}, though not better than Std-MALDI. Std-MALDI has the lowest standard deviations for the scores generated, while the AnchorChip has the largest deviations.

Figure 2.2.4 is similar to Figure 2.2.3 in that only the x-axis has been changed from concentration to total amount added to the surface, so as to try and remove any loading effects that may be skewing the interpretation of the data. Since 10 µl of analyte was applied to the X3 Biochip, 5 µl to the AnchorChip and 1 µl for Std-MALDI, there is potentially a loading effect in the order of 2:1 for X3 Biochip to AnchorChip and 10:1 for X3 Biochip to Std-MALDI. The AnchorChip to Std-MALDI has a potential loading effect of 5:1. I decided to use 1 µl of sample for the Std-MALDI method, rather than the 0.5 µl (1:1 mixture) used in the earlier experiments, to try and reduce the loading effect, if any, between the methods. I also planned to use 10 µl of sample on the AnchorChip, though the SOP suggest 5 µl as being the optimal amount and I did not want to deviate from this by placing too much on the surface and causing the droplets to break out of the wells. Alternatively, I considered loading 5 µl, waiting for it to concentration then applying another 5 µl followed by the matrix. Both alternative methods would have been non-standard and not a true representation of the how the techniques compare to the SOP and any results generated could be questioned as skewed.

Viewing the data with total moles on the x-axis rather than concentration, as in Figure 2.2.4 leads to a subtle difference in the interpretation. Firstly, the performance of the X3 Biochip compared to the AnchorChip is less obviously different. Also, the performance of the AnchorChip and Std-MALDI seems to come closer together, over a wider range.

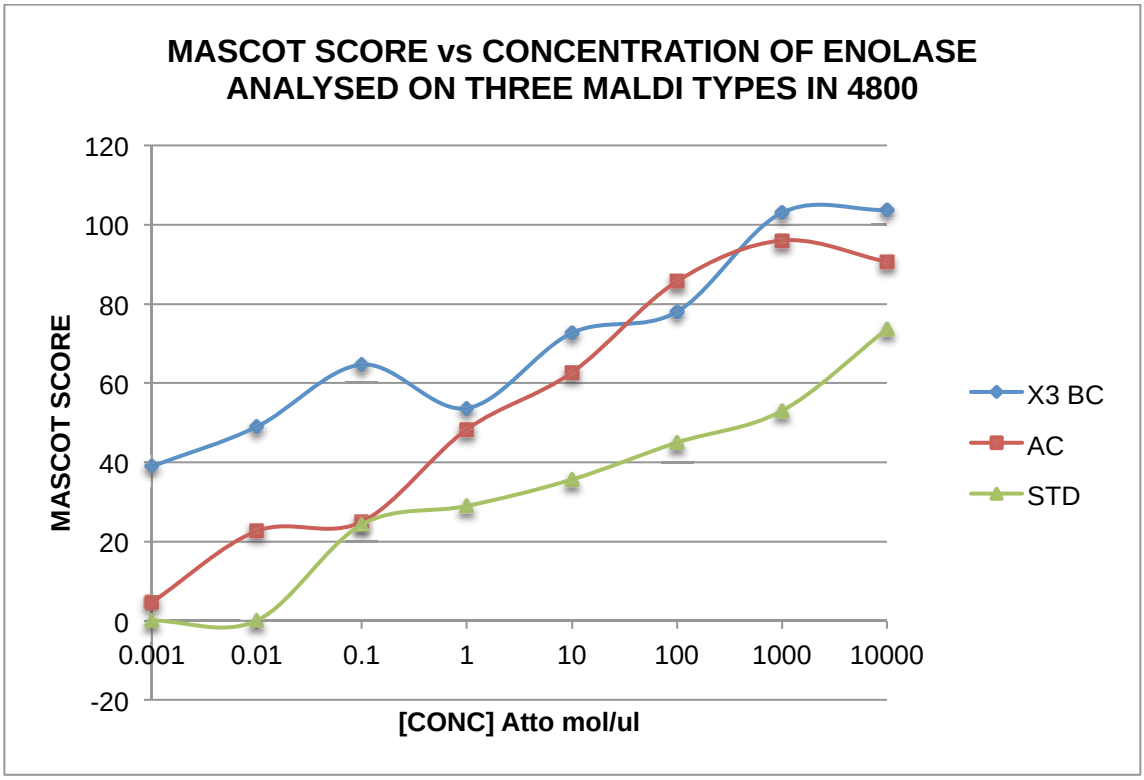


Figure 2.2.3 Graph depicting the response of the mass spectrometer in identification of a protein digest (Enolase) across a concentration gradient with respect to the mascot score generated from the MS/MS analysis. The x-axis represents the concentration of the protein digest applied and the y-axis represents the mean identification score (3 replicates), delivered from the Mascot search engine from the MS/MS data generated. The blue line represents the X3 Biochip, the red line the AnchorChip and green line the Std-MALDI. Both the X3 Biochip and AnchorChip show superior results across the entire concentration gradient compared to the Std-MALDI technique. The X3 Biochip performed similarly to the AnchorChip across the higher concentrations applied, though it seems to out perform the AnchorChip at the LOD range of the concentration gradient tested here.

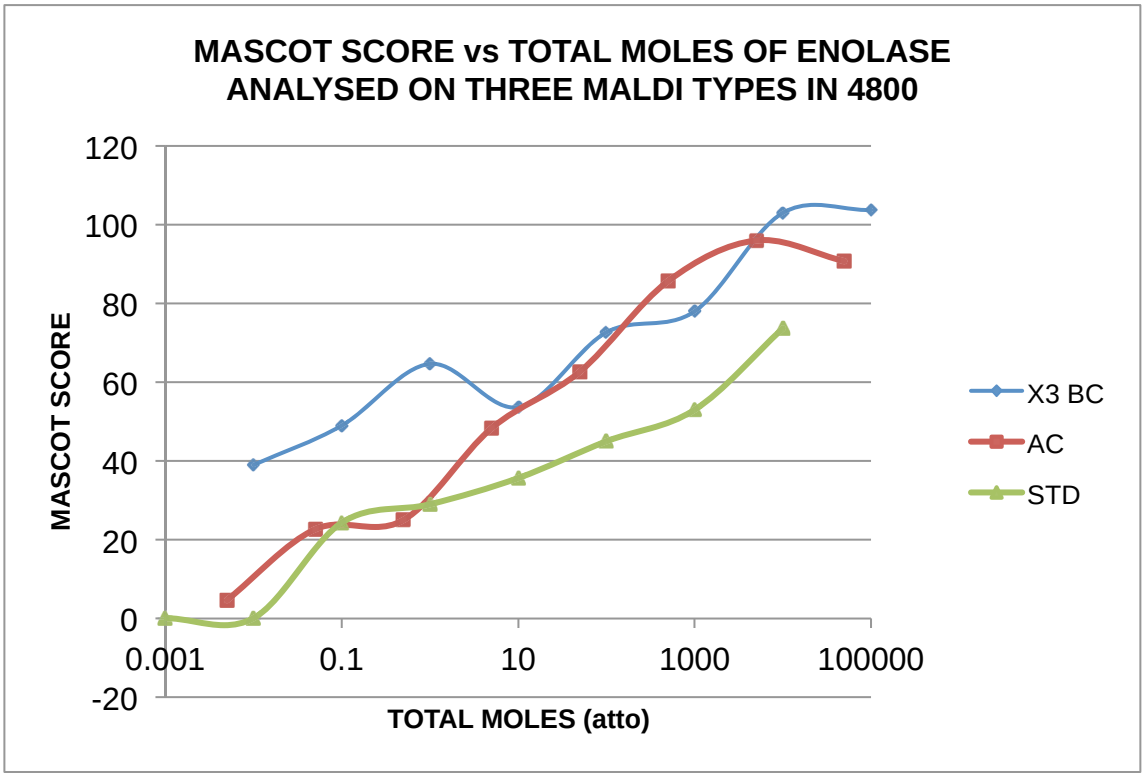


Figure 2.2.4 Graph depicting the response of the mass spectrometer identification of a protein digest (Enolase) across a concentration gradient with respect to the mascot score generated from the MS/MS analysis and the total number of moles applied to the surface. The x-axis represents the total moles of the protein digest applied and the y-axis represents the mean identification score (3 replicates), delivered from three the Mascot search engine from the MS/MS data generated. The blue line represents the X3 Biochip, the red line the AnchorChip and green line the Std-MALDI. Both the X3 Biochip and AnchorChip show superior results across the higher end of concentration gradient compared to the Std-MALDI technique. The X3 Biochip performed similarly to the AnchorChip across the higher concentrations applied, though it seems to outperform the AnchorChip at the LOD range of the concentration gradient tested here.

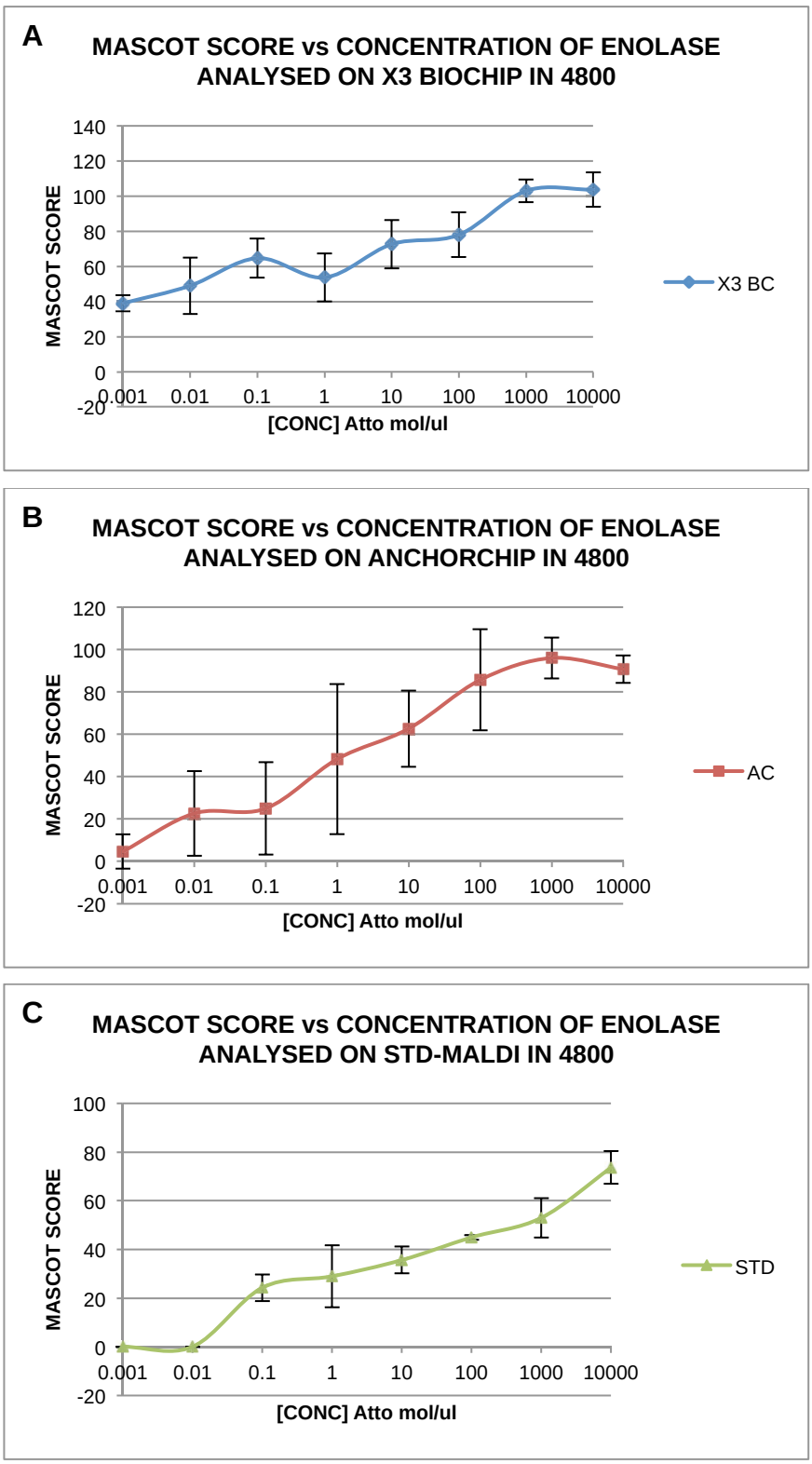


Figure 2.2.5 The three graphs displayed here are the same information presented in Figure 2.2.3 except in three separate graphs, one for each technique, to highlight the standard deviation for the Mascot scores generated across the concentration gradient (3 replicates). (A) The blue line representing the X3 Biochip shows similar standard deviation across the entire concentration gradient and is less than the AnchorChip (red), though more than Std-MALDI (green). (B) The AnchorChip has the highest standard deviation across the majority of the concentration gradient applied. (C) The Std-MALDI shows the lowest standard deviation across the concentration gradient applied.

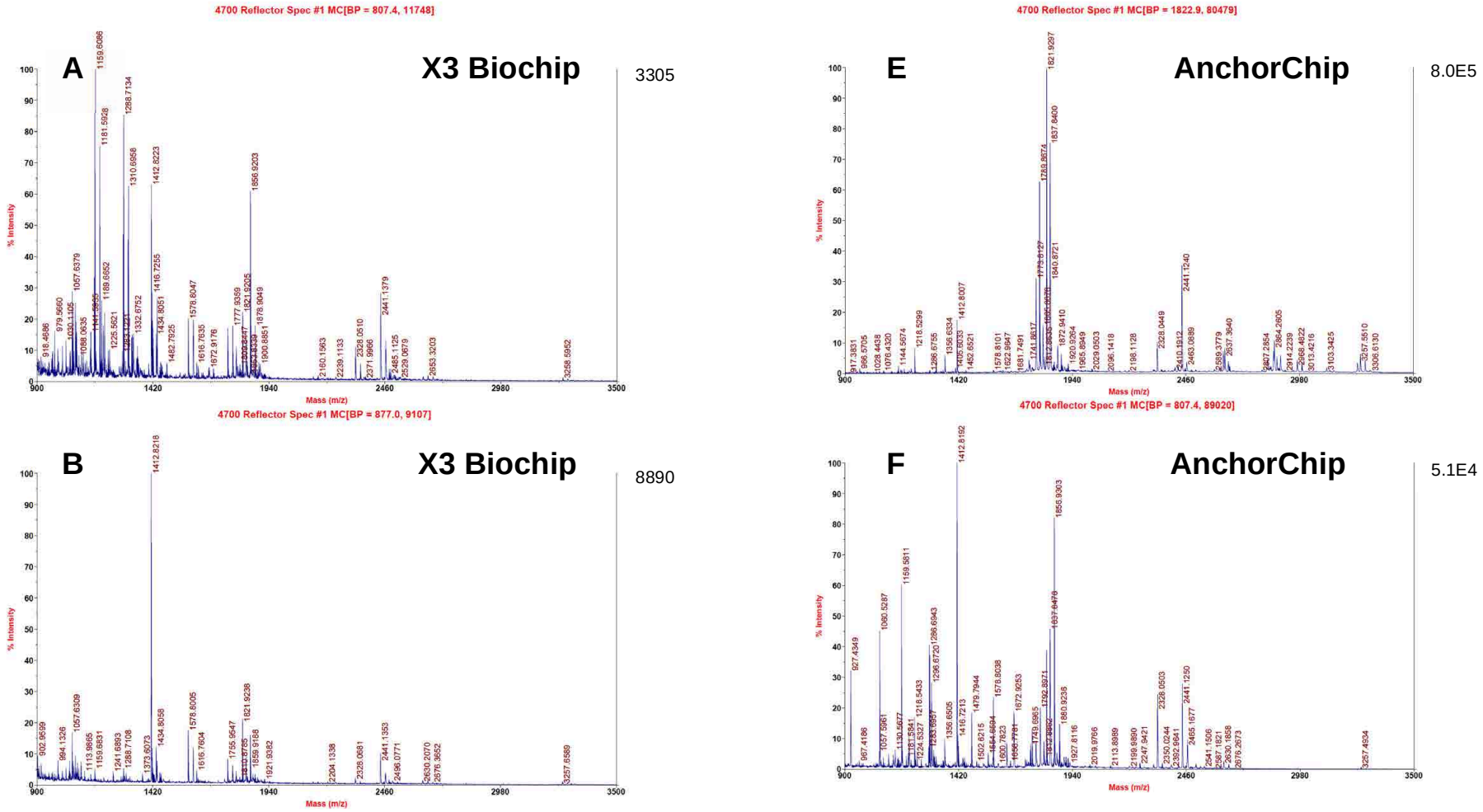


Figure 2.2.6 Spectra from the Enolase digest. Spectra (A) and (B) are the X3 Biochip at the first and third dilutions respectively. Spectra (E) and (F) are the AnchorChip at the first and third dilutions respectively. The ion count for the AnchorChip is significantly greater than for the X3 Biochip by a factor of 10. Note that the AnchorChip is showing residual 'ghosting' of the peptide mixture (m/z 1060, 1296, 1672 and 2465) that was used in earlier testing which could be causing ion suppression and skewing the data.

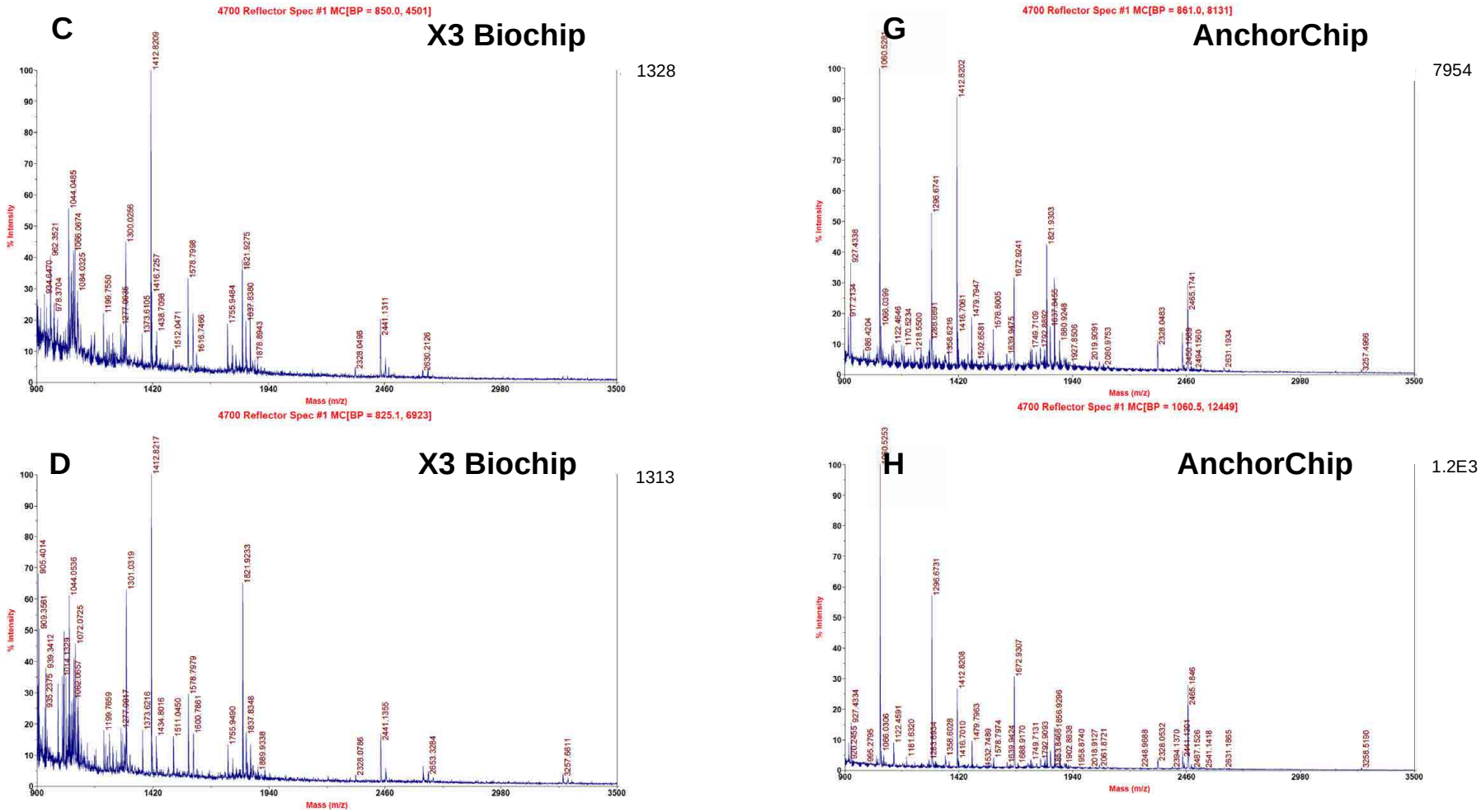


Figure 2.2.6 (continued) Spectra from the Enolase digest. Spectra (C) and (D) are the X3 Biochip at the fifth and seventh dilutions respectively. Spectra (G) and (H) are the AnchorChip at the fifth and seventh dilutions respectively. The ion count for the AnchorChip is significantly greater than for the X3 Biochip by a factor of 10. The AnchorChip is presenting superior ion count despite generating comparable mascot scores for identification. Note that the AnchorChip is showing residual 'ghosting' of the peptide mixture (m/z 1060, 1296, 1672 and 2465) that was used in earlier testing which could be causing ion suppression and skewing the data.

With respect to the analysis of the HSA digest shown in Figure 2.2.7, the X3 Biochip outperforms the AnchorChip across the last 4 dilutions, where in the Enolase it was the last three. The Std-MALDI outperforms the AnchorChip at the 1 attomol/µl level and also the X3 Biochip at the 10 attomol/µl. The AnchorChip outperforms both the X3 Biochip and Std-MALDI at the higher end of the concentration gradient (100 – 10000 attomol/µl). There is an anomalous reading for Std-MALDI at the 1000 attomol/µl mark that skews the data and trend line for the Std-MALDI results.

The AnchorChip and Std-MALDI generate similar scores across the mid range of the concentration gradient (0.1 – 10 attomol/µl), though the X3 Biochip generates superior results for the last four concentrations, compared to the Enolase results. The X3 Biochip outperforms the other two methods at the LOD level similar to the Enolase results, and outperforms the AnchorChip across the mid-range for HSA, where for Enolase it was only in the last three concentrations, rather than being the last two (0.01 – 0.001 attomol/µl) for HSA.

The differences in the standard deviation of mascot scores generated for HSA identification are minimal, as shown in Figure 2.2.9. The general deviations are lower across all three methods compared to the Enolase results with only a few large deviations by each method across the scale at differing points. The standard deviation for the 1000 attomol/µl is low because all three were showing the effect of the anomalous reading.

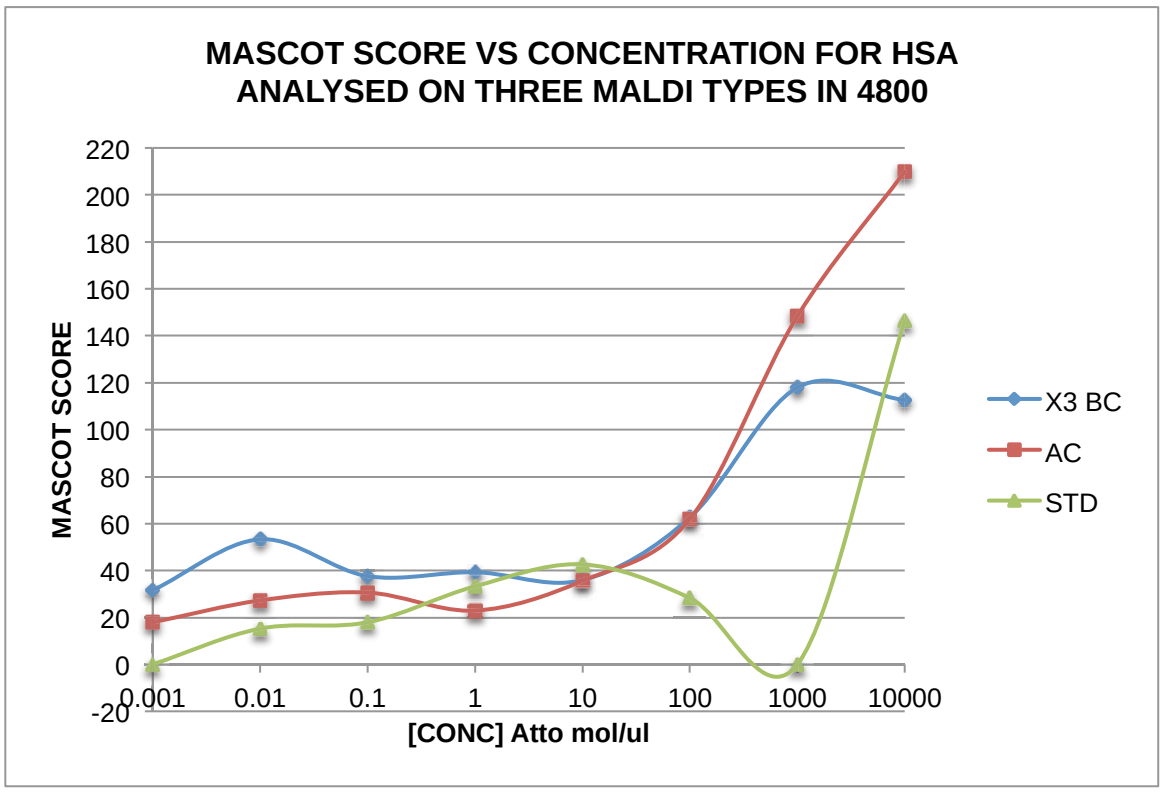


Figure 2.2.7 Graph depicting the response of the mass spectrometer to identify a protein digest (HSA) across a concentration gradient with respect to the mascot score generated from the MS/MS analysis. The x-axis represents the concentration of the protein digest applied and the y-axis represents the identification score delivered from the mascot search engine from the MS/MS data generated (3 replicates). The blue line represents the X3 Biochip, the red line the AnchorChip and green line the Std-MALDI. Both the X3 Biochip and AnchorChip generated higher mascot scores across the lower end of the concentration gradient applied, with the Std-MALDI showing similar performance across the mid range of the concentration gradient. The mascot score for Std-MALDI at the 1000 attomol/µl is an anomalous reading that skews the data for Std-MALDI on HSA. The AnchorChip out performs both the X3 Biochip and Std-MALDI at the higher concentrations tested. The X3 Biochip generates the best performance at the LOD end of the concentration gradient, similar to the Enolase digest in Figure 2.2.3.

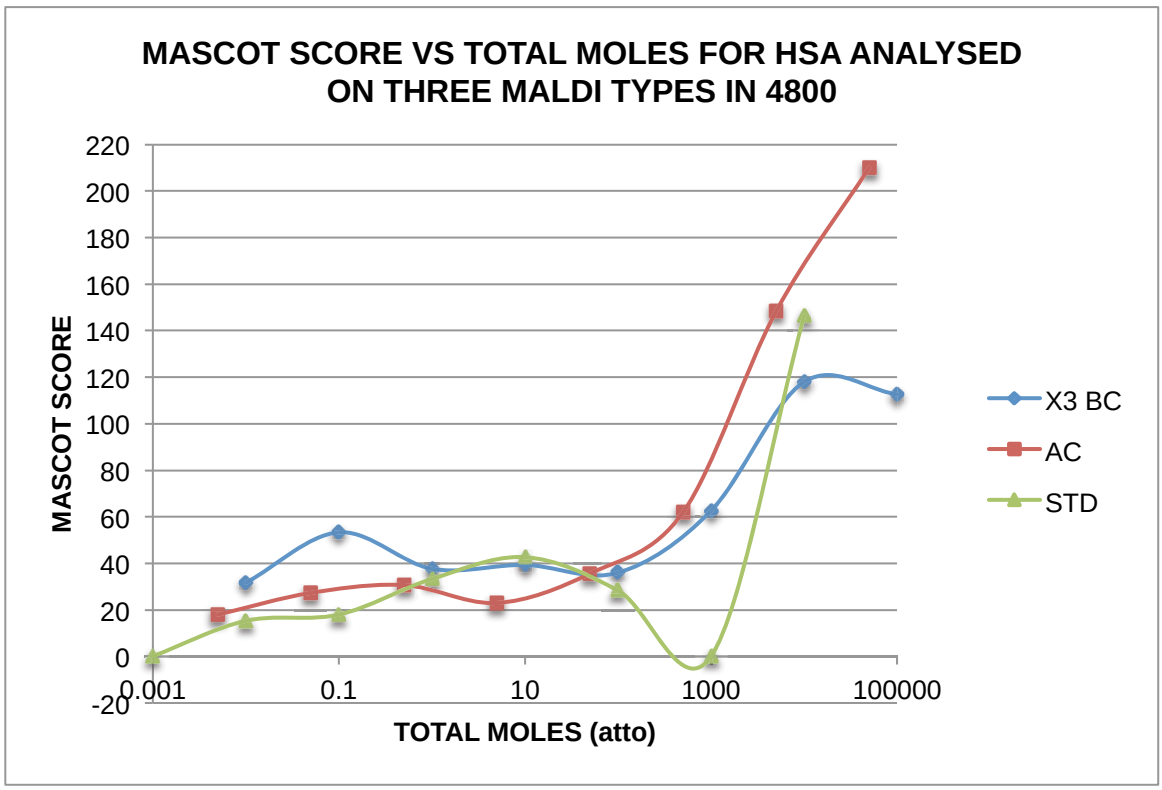


Figure 2.2.8 Graph depicting the response of the mass spectrometer to identify a protein digest (HSA) across a concentration gradient with respect to the mascot score generated from the MS/MS analysis and the total number of moles applied to the surface. The x-axis represents the total moles of the protein digest applied and the y-axis represents the identification score delivered from the mascot search engine from the MS/MS data generated (3 replicates). The blue line represents the X3 Biochip, the red line the AnchorChip and green line the Std-MALDI. Both the X3 Biochip and AnchorChip generated higher mascot scores across the lower end of the concentration gradient applied, with the Std-MALDI showing similar to performance across the mid range of the concentration gradient. The mascot score for Std-MALDI at the 1000 attomoles is an anomalous reading that skews the data for Std-MALDI for HSA. The AnchorChip out performs both the X3 Biochip and Std-MALDI at the higher concentrations tested. The X3 Biochip generates the best performance at the LOD end of the concentration gradient, similar to the Enolase digest in Figure 2.2.4.

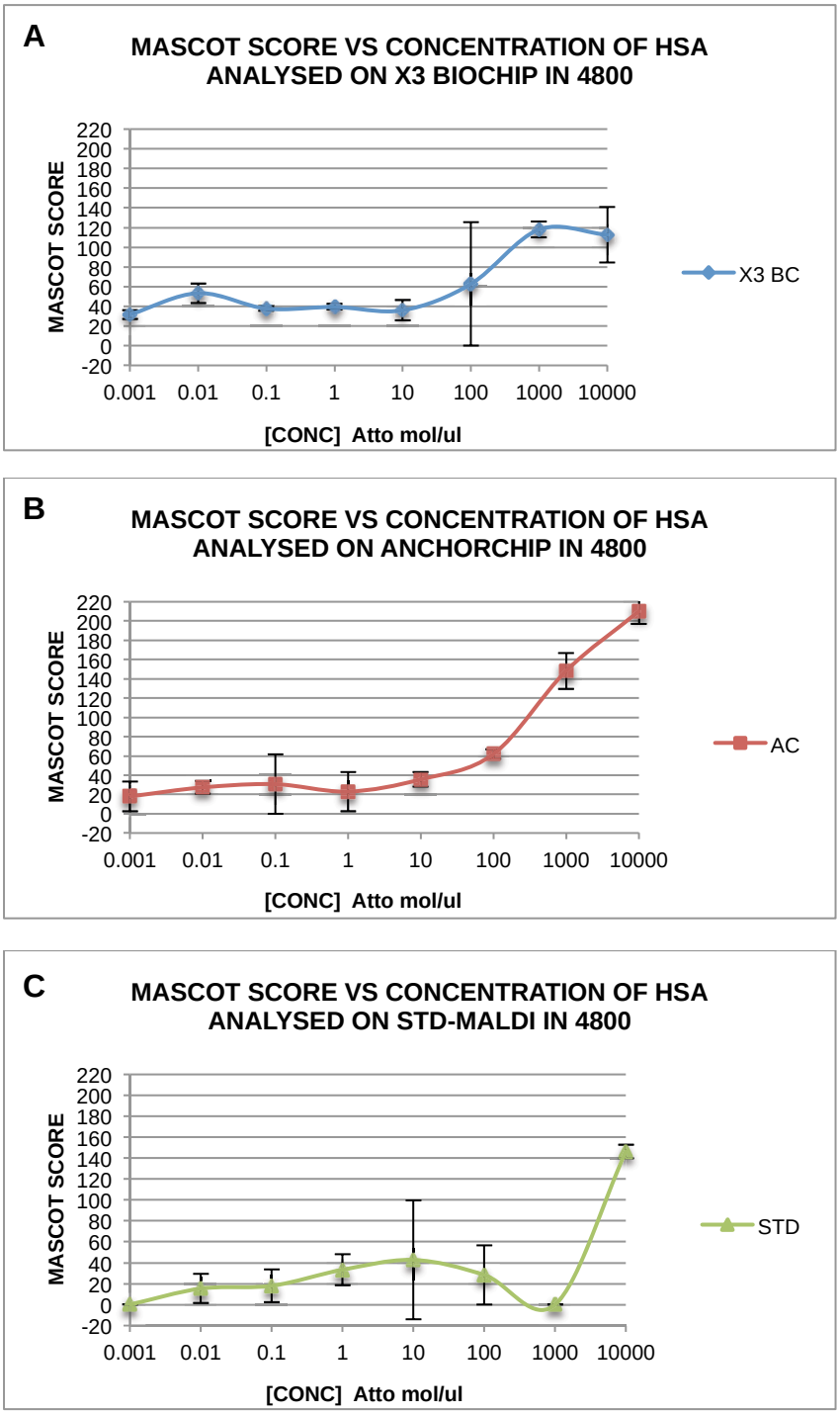


Figure 2.2.9 The three graphs displayed here are the same information presented in Figure 2.2.7 though presented here in three separate graphs to highlight the standard deviation for the mascot scores generated across the concentration gradient (3 replicates). (A) The blue line representing the X3 Biochip shows low and similar standard deviation across the entire concentration gradient except for a large standard deviation at the 100 attomol/µl reading and medium reading at 10000 attomol/µl. (B) The AnchorChip has a average standard deviation response across the concentration gradient, from highly reproducible with very low standard deviation at 0.01, 10 and 100 attomol/µl to intermediate standard deviation compared to the other techniques. (C) The Std-MALDI shows high to medium standard deviations, with the anomalous reading at 1000 attomol/µl skewing the response curve shape.

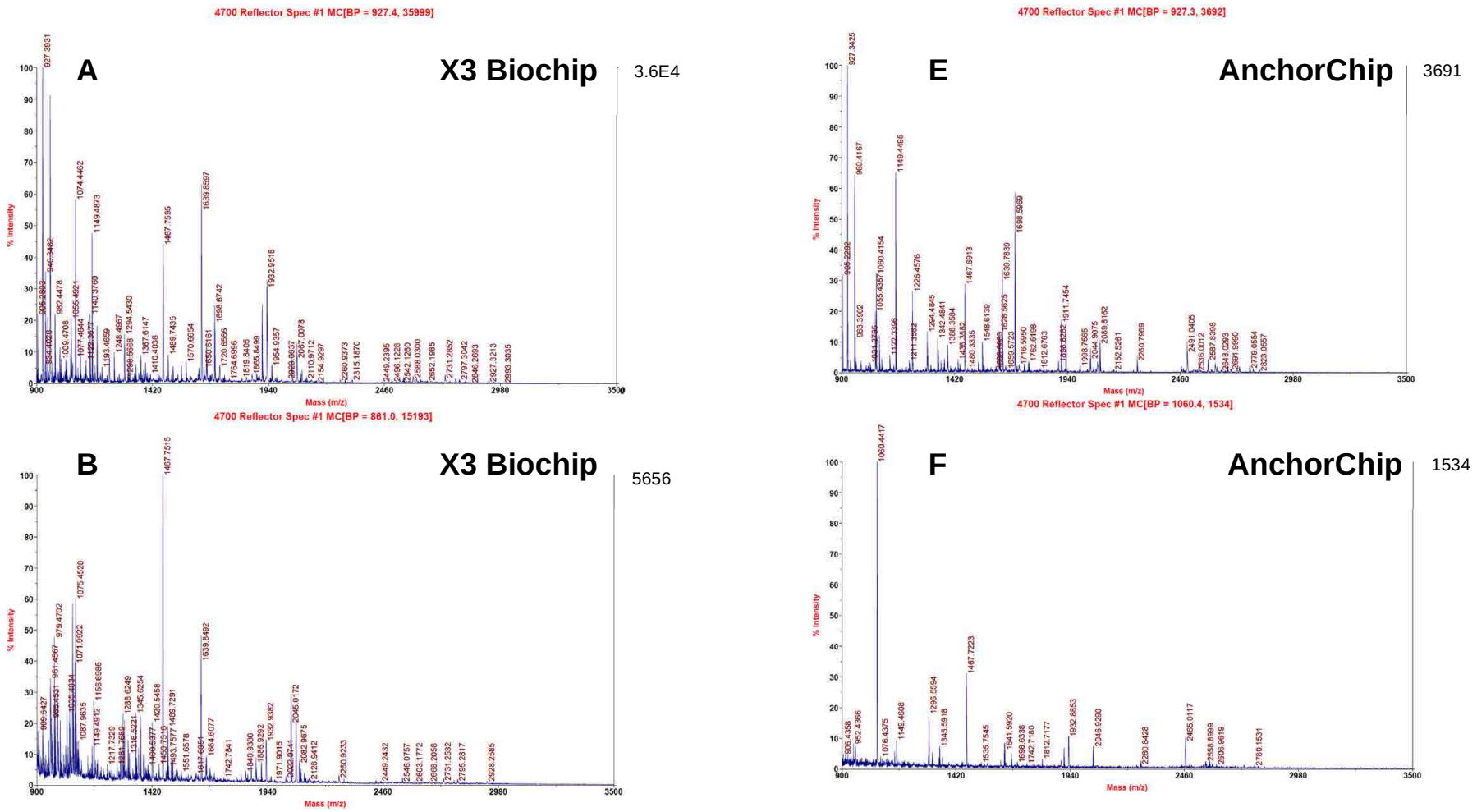


Figure 2.2.10 Spectra from the HSA digest. Spectra (A) and (B) are the X3 Biochip at the first and third dilutions respectively. Spectra (E) and (F) are the AnchorChip at the first and third dilutions respectively. The ion count for the X3 Biochip is significantly greater than for the AnchorChip by a factor of 5-10. Note that the AnchorChip is showing residual 'ghosting' of the peptide mixture (m/z 1060, 1296, 1672 and 2465) that was used in earlier testing, which could be causing ion suppression and skewing the data.

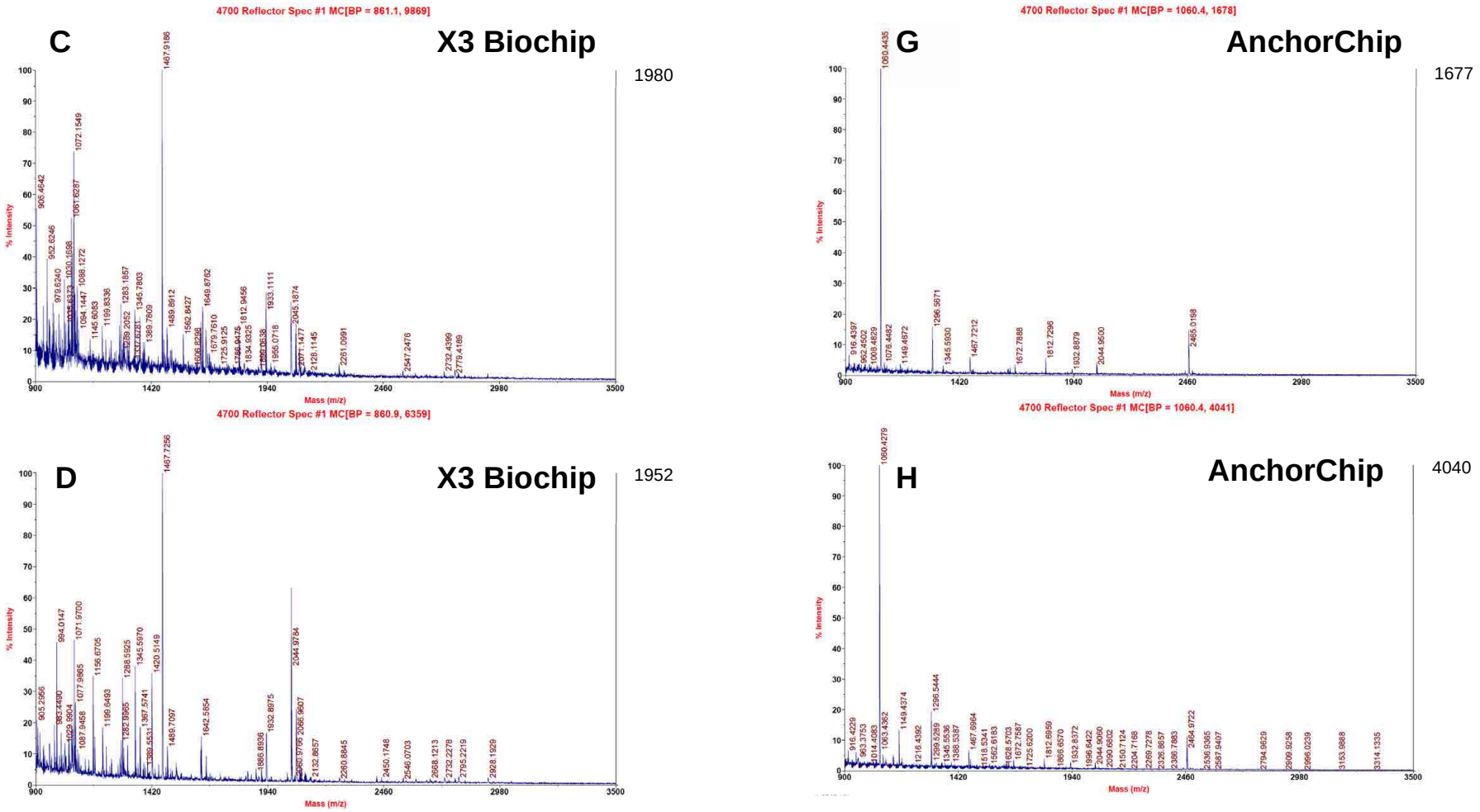


Figure 2.2.10 (continued) Spectra from the HSA digest. Spectra (C) and (D) are the X3 Biochip at the fifth and seventh dilutions respectively. Spectra (G) and (H) are the AnchorChip at the fifth and seventh dilutions respectively. The ion count is similar for both methods. Note that the AnchorChip is showing residual 'ghosting' of the peptide mixture (m/z 1060, 1296, 1672 and 2465) that was used in earlier testing, which could be causing ion suppression and skewing the data.

Due to reproducibility issues with the X3 Biochips I have not presented the BSA and α-lactalbumin experiments, which were undertaken at the same time as these experiments with Enolase and HSA.

These three types of MALDI hold advantages in three different types of environments. The Std-MALDI gives the most reproducible generation of signal across a concentration range, then dropping off sharply once the LOD has been reached, but it is the easiest and quickest method to implement. Across the middle to lower concentration range, the AnchorChip will generate superior results to Std-MALDI, while generating comparable results to the X3 Biochip.

The observation of the peptide mixture still attached to the AnchorChip after cleaning could be causing ion suppression of the digest and skewing the data in favour of the X3 Biochip. The presence of the peptide mixture signal in these samples is an example of the necessity to undertake stringent cleaning beyond the manufacturers specification, since I did follow the standard protocol for cleaning (sonication). These peptide mixture signals are also an example of the need for single use surfaces when working at the lower end of the LOD with current instrumentation that can detect into and beyond the attomole level.

Daily operation in a normal laboratory environment is typically not about LOD, but more about generating a stronger and more reproducible signal within the standard concentration ranges. Since the AnchorChip meets this need and is also re-useable, it can be considered superior for this type of application. Finally, any investigator trying to identify unknown proteins at the lowest of concentrations has the best chance of generating a higher signal with the X3 Biochip compared to the AnchorChip and Std-MALDI.

The final point to note regarding these LOD experiments is the improved LOD achieved here compared to the earlier experiments conducted on the 4700, even though these samples are more complex in nature. This is primarily due to the 4800 having an approximate 10-fold increase in sensitivity of detection compared to the 4700. According to the manufacturer this is brought about by improved optics leading to enhanced ion transmission. I have not tried to highlight the absolute LOD for the methods displayed here because the absolute figures could be in error by a

significant margin due to the variability seen when using commercial protein digests that are not made primarily for quantitative purposes, as was done here. This was a study of the relative differences between the methods and all care to produce quantitative technical replicates for the most meaningful relative comparisons were undertaken, which is why I have discussed the relative differences between these techniques rather than the absolute LOD.

2.4.3 Limit of Detection Comparison of Phosphorylated Peptides Utilising the Concentrating and Affinity Capture (NTA3) Biochip with Immobilised Metal Affinity Chromatography within an Applied Biosystems 4700 Analyser

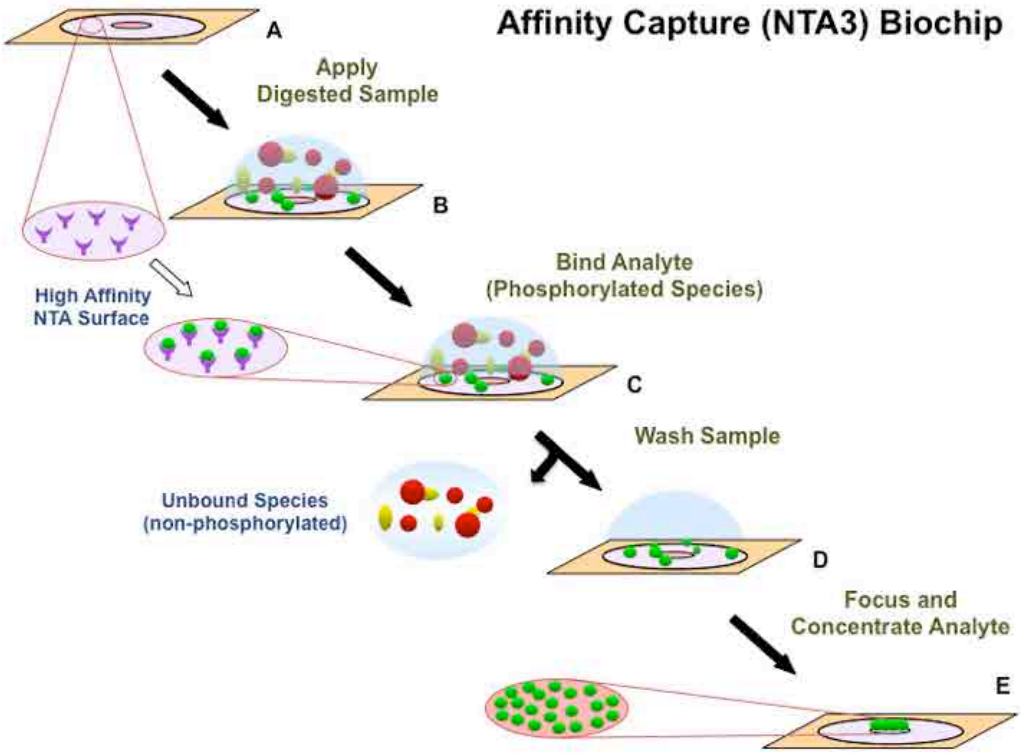


Figure 2.3.1 Schematic representation of the general methodology for the utilisation of the affinity capture concentrating Biochip with NTA conjugated to the surface for the application of IMAC on phosphorylated peptide samples (NTA3 Biochip). (A) The liquid retention zone is populated with the affinity capture ligand, referred to here as the capture zone, either at the point of manufacture or in the lab by the investigator, as I did by conjugating the NTA molecule (purple fork shape) to the COOH groups exposed on the surface of the SAM in the capture zone. (B) The sample is applied to the surface once the surface chemistry has been prepared and activated, with other contaminants present in the sample, being generally other non-phosphorylated peptides or salts, shown as red and yellow balls and phosphorylated peptides as green balls. (C) The mixture is allowed to incubate and equilibrate, allowing the phosphorylated peptides to bind to the NTA ligands. (D) A washing step is applied to remove any unbound species, leaving behind primarily phospho-peptides. (E) A solution containing the matrix at an acidic pH is applied, facilitating the uncoupling of the phosphorylated peptides from the NTA ligands and the concentration of these peptides and the matrix in a co-migratory event towards the centre of the CDVW. This results in the concentration, crystallisation and presentation of the phosphorylated peptides in the central analysis zone for MALDI MS analysis.

The experiments in the following section were performed at Lumicyte in San Jose, California, and the ABI demonstration laboratory in Fremont, California. The data were presented as a poster at The American Society for Mass Spectrometry (ASMS) conference in San Antonio, Texas, June 2005.

The initial plan was to have the High Affinity (HA3) Immobilised Metal Affinity Chromatography (IMAC) chips supplied pre-fabricated at the point of manufacture with the nitrilotriacetic acid (NTA) molecules covalently bound to the surface, so I did not have to do this conjugation step in the lab prior to each experiment. This was not available, so I had to conjugate the NTA molecule to the SAM surfaces exposed carbonyl (COOH) groups via a 1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC) and N-hydroxysuccinimide (NHS) mediated reaction. I had no way of testing the yield of such a reaction, which introduced a variable to the experiments that has not been accounted for and may explain some of the variability I experienced in the early stages of the method development. Another variable that made it difficult to method develop was the amount of COOH groups on the surface in the capture zone. This varied from 10% in some to 15% and 20% COOH groups in others, changing the spatial distribution of the NTA molecule across the surface and changing the binding capacity of the chips for the NTA, $FeCl_3$ and phosphorylated species. The introduction of differing amounts of COOH groups was done at the point of manufacture with the goal to increase the number of NTA molecules bound and increase the binding capacity of the Biochip surface for phosphopeptides. The change in number of available COOH groups and bound NTA moieties altered the concentration effect of the normal hydrophilic gradient when decoupling the analyte from the NTA ligand. I observed that as the amount of COOH groups in the capture zone was increased, the concentration event resembled more of a normal evaporative process and that the snap-lock/washing effect unique to these Biochips was lost. Additionally, when the Fe-NTA was bound to the high COOH chips (20%) there was almost no concentration seen at all.

Since this was a newly manufactured product at the time, the HA3 IMAC chip protocol we began with was not optimised for Fe-NTA, but for a Streptavidin chip. We had to test the variables that made these Fe-NTA chips different and workable, particularly the activation of the surface with the $FeCl_3$ and the wash buffers so as to

make sure that we were activating the surface correctly and not de-activating the chip and/or removing the Fe^{3+} species from the NTA-Fe complex before the desired time.

Initially, I started with a 10 mM $FeCl_3$ solution that was added to the surface, as was in the Ni-NTA protocol for the addition of the 10 mM $NiSO_4$, but I found that this was too much and the surface of the capture zone seemed to react (oxidize) prematurely before the addition of the sample. It also seemed to precipitate out of solution in the concentration step, inhibiting concentration and leaving a yellow-brown precipitate over the capture zones with matrix and sample spread throughout. These coagulations did give a signal on some wells when I tried to analyse them in the mass spectrometer, but they showed no significant advantage in detection of peptides.

I eventually came up with a method that seemed to supply some positive results that were reproducible enough to give several replicates, but this was at the expense of many Biochips and samples. According to popular wisdom, a researcher only needs to show something work three times reproducibly, while a product developer needs to show it works at least 9 times out of 10. On the other hand, a proof of concept only needs to show it works once or twice.

In my opinion, this technique was still in the proof of concept stage at the time of these experiments in 2005. I may have achieved an early proof of concept, but the technology and technique is far from ready for use in the general scientific community and not even yet at a beta-test stage.

Figure 2.3.2 outlines the results generated in California, USA. These four experiments across two NTA3 Biochips show that the NTA3 Biochip was able to selectively capture the mono-phosphorylated peptide T6 (m/z=2061) of the protein β-casein at a sensitivity of identification with MS/MS of 50 femtomoles total applied to the surface for all four experiments. The addition of BSA in a 10 fold molar excess to the CDVWs did not have any deleterious effects on these results, for the same sensitivity of identification was observed for the spiked samples as was seen without any competing background. The quadruply phosphorylated peptide T1-2 (m/z=3122) was not detected in any of these experiments. It is known in the literature that Fe^{3+}-IMAC (NTA) has a greater co-ordination chemistry affinity for singularly

phosphorylated species compared to multiply phosphorylated species, and hence we were not expecting to see comparable signals, though I was expecting to see something at the higher concentrations applied, which I did not. The affinity of Fe^{3+}-NTA to phosphorylated peptides could be altered by these flat surfaces compared to standard substrates on spherical surfaces, in conjunction with the absence of microfluidics generated in most liquid chromatography, having advantageous or deleterious compounding effects for this technique [373, 374].

Figure 2.3.2 B) shows the difference in the amount of ions generated in MS/MS mode for the four experiments. Experiments 1 and 3 show similar amounts of ions generated at the higher end of the concentration range, 50-100 femtomoles total applied. The fact that these two experiments did not have any BSA added to the background may explain the similar results, as with experiments 2 and 4, which also show similar results between the 50-100 femtomoles total applied range. Figure 2.3.2 B) also highlights that two of the experiments generated MS/MS ions at the 10 femtomole total applied level, indicating that the LOD for this technique for MS/MS selection was 10 femtomoles, although the MS/MS spectra generated were not of a high enough quality to generate a matching score in the Mascot search engine.

Since our first aim was to detect known phosphopeptides, for the purposes of this experiment we decided that the LOD did not require MS/MS identification, if upon manual inspection a m/z peak of a known m/z was observed with a signal to noise ratio of 10 or greater, then this is stated as the level at which the technique has achieved LOD in MS mode. Figure 2.3.3 shows that experiment 3 generated a LOD of 0.75 femtomole total applied (750 attomole) for the selection of the phospho-peptide T6 (m/z=2061). Experiments 1, 2 and 4 generated LODs of 10, 10 and 50 femtomole respectively.

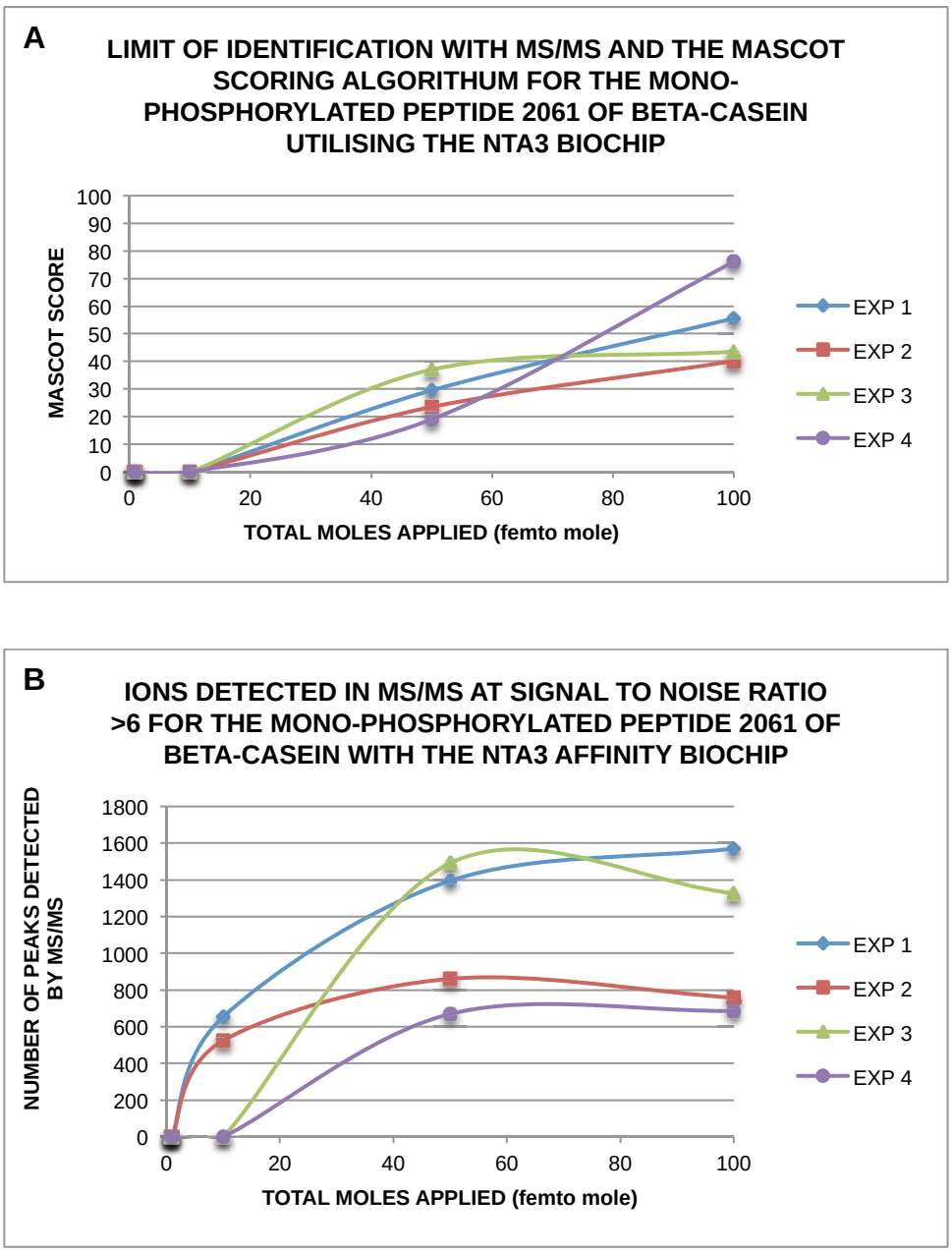


Figure 2.3.2 Graphical representation of the limit of identification for the mono-phosphorylated peptide (m/z=2061) of β-casein to be 50 femtomoles across all four experiments, irrespective of the addition or not of a BSA digest in 10 molar excess to act as background contamination. The quadruply phosphorylated peptide (m/z=3122) of β-casein was not detected in any of the experiments. (A) The x-axis represents the total moles of analyte applied in femtomoles, the y-axis is the number score generated by the mascot search engine. Experiment 1 had T6 (1P) (M/Z=2061) and T1-2 (4P) (m/z=3122) spiked onto the surface of the NTA3 Biochip without any background contamination. Experiment 2 had T6 (1P) (m/z=2061) and T1-2 (4P) (m/z=3122) spiked onto the surface of the NTA3 Biochip with the additional of a BSA digest in a 10 molar excess. Experiment 3 had a digest of β-casein applied to the NTA3 Biochip without background contamination. Experiment 4 had a digest of β-casein applied to the NTA3 Biochip with the addition of a BSA digest in a 10 molar excess. All four experiments generated the same LOD for identification with MS/MS, though the reproducibility of the signals was highly variable. (B) Graphical representation of the total number of MS/MS peaks generated at the various amounts of analyte applied. Both experiments 1 and 2 show peaks at the 10 femtomole amount, indicating that ionisation and selection for the phospho-peptide was achieved at this lower level, but not enough for the generation of identification by Mascot.

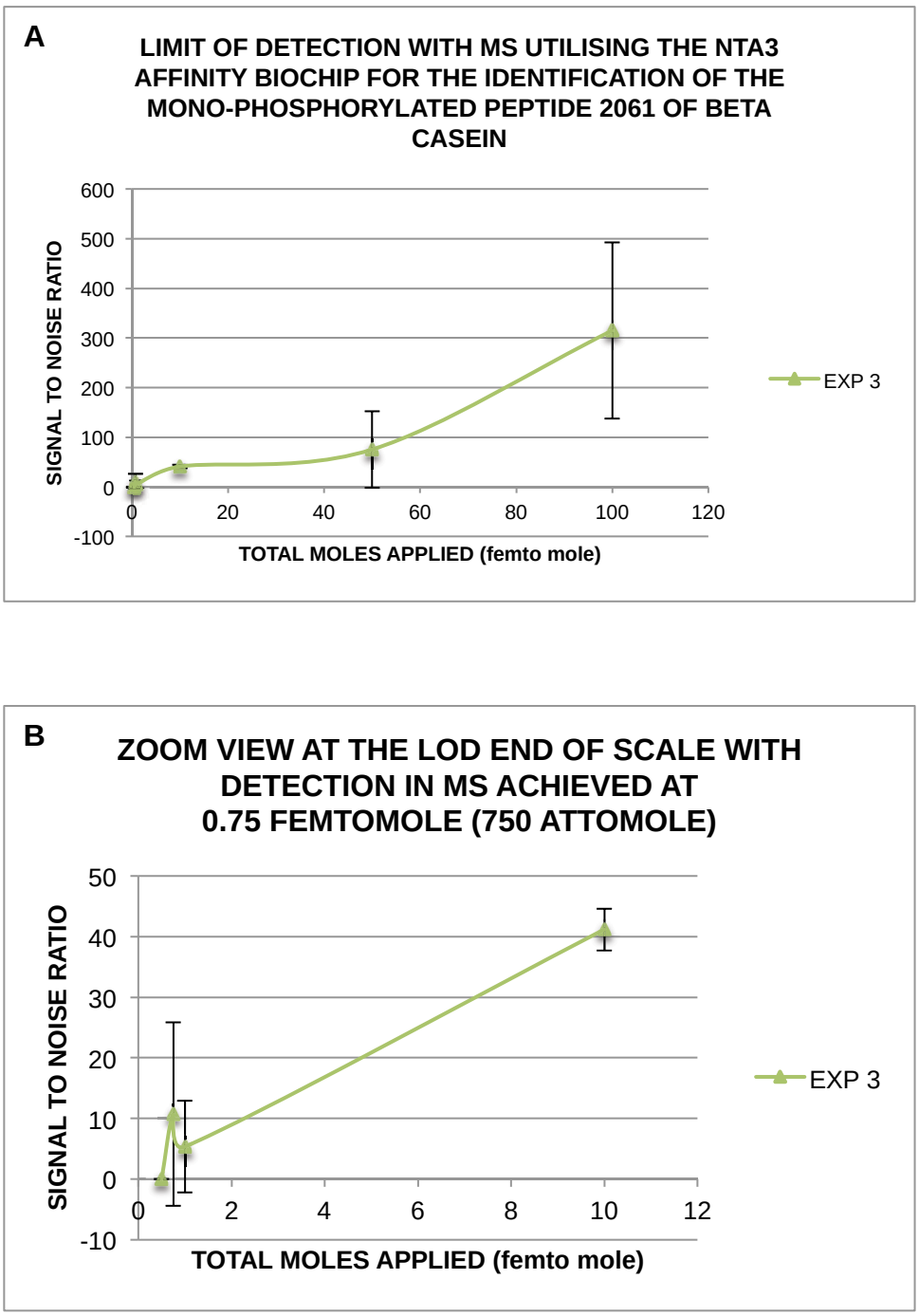


Figure 2.3.3 The LOD for the affinity capture and ionisation in MALDI MS of the singly phosphorylated peptide (m/z=2061) on a 4700 mass spectrometer was 0.75 femtomole (750 attomole) of β-casein digest applied, without any background contamination apart from the non-phosphorylated peptides in the β-casein digest. The standard deviation at both the 1 and 0.75 femtomole level are high because the peptide was only identified once, rather than twice, as it was for all other signals presented in these experiments, and I have calculated the standard deviation over two potential samples with one generating zero to highlight that these results were singular events. The actual signal to noise reading was 10.7 and 21.45 at 1 and 0.75 femtomole respectively. Experiments 1, 2 and 4 did not identify 2061 at this LOD, their LOD were 10, 10 and 50 femtomole respectively. (A) The MS signal to noise response across the full scale of the experiment 3. (B) A zoomed in view to highlight the information at the LOD range that is not easily visualised in graph A above for experiment 3.

Figure 2.3.4 highlights a selection of the spectra generated for experiment 3 and the generation of the LOD at 750 attomole for the mono-phosphorylated peptide 2061 of β-casein. The metastable ion fragments are shown to justify that the peak is phosphorylated because I was not able to generate MS/MS information at this level. Alternatively, Figure 2.3.5 shows two MS/MS spectra generated in experiment 4 at the 50 and 100 femtomole total applied level.

As stated earlier, the NTA3 Biochip did not show any generation of signal or selective binding of the quadruply phosphorylated peptide T1-2 (m/z=3122) of β-casein. However, I was able to generate a signal from a doubly phosphorylated peptide on several occasions as part of the method development for the NTA3 Biochip. The doubly phosphorylated (pS, pY) peptide of Map Kinase Substrate was shown to selectively bind to the NTA3 Biochip after washing the surface several times, with the successful generation of signal in MS. Figure 2.3.6 shows one such spectrum generated and highlights the metastable fragment ions, which I used to prove that the peptide was phosphorylated at the point of ionisation without MS/MS data.

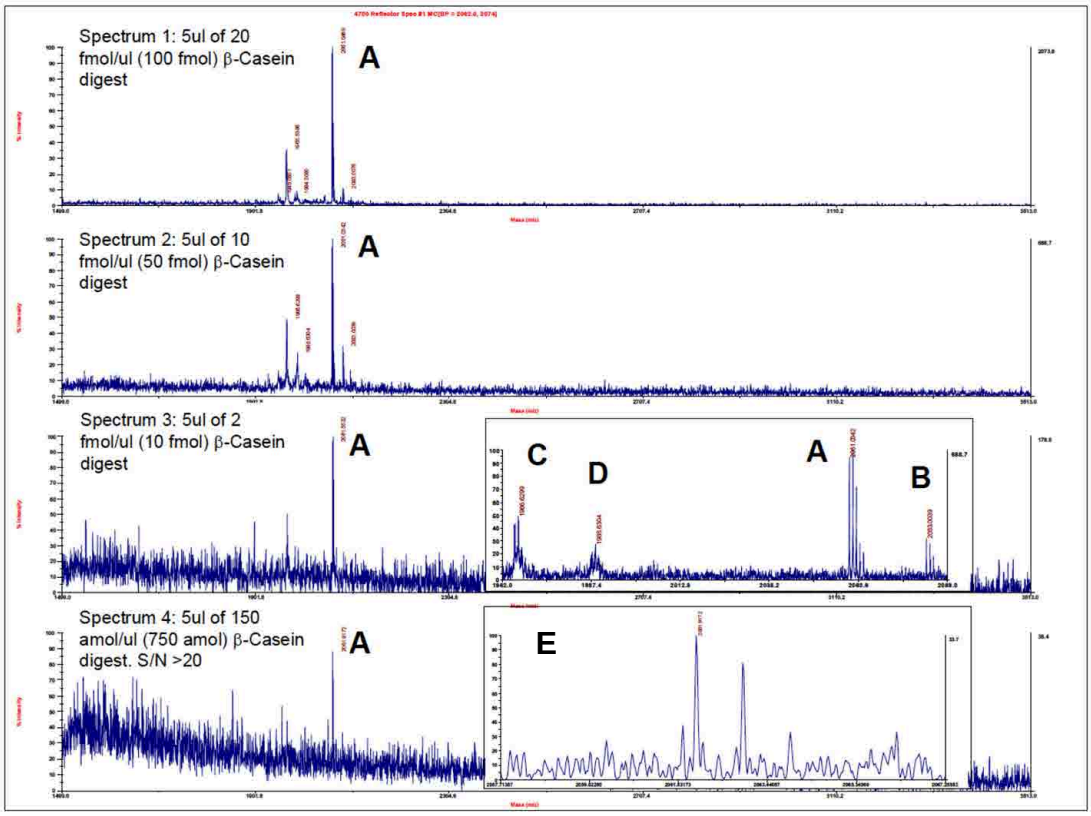


Figure 2.3.4 Spectra from experiment 3 highlighting the LOD for phospho-peptide 2061 from a β-casein digest utilising the NTA3 Biochip. Peak (A) represents the parent ion of the singly phosphorylated peptide at m/z = 2061 found across the concentration gradient. The bottom spectrum shows a signal to noise ratio of greater than 20 was achieved when 5 µl of 150 attomol/µl (750 attomole) was applied to the surface. The spectrum second from the bottom highlights the fragment ions (metastables) generated by the loss of the phosphate group in MS mode, labelled as (C) and (D). Label (B) represents a water adduct on the phospho-peptide. Label (E) is a zoomed-in view of the 2061 peak at 750 attomoles.

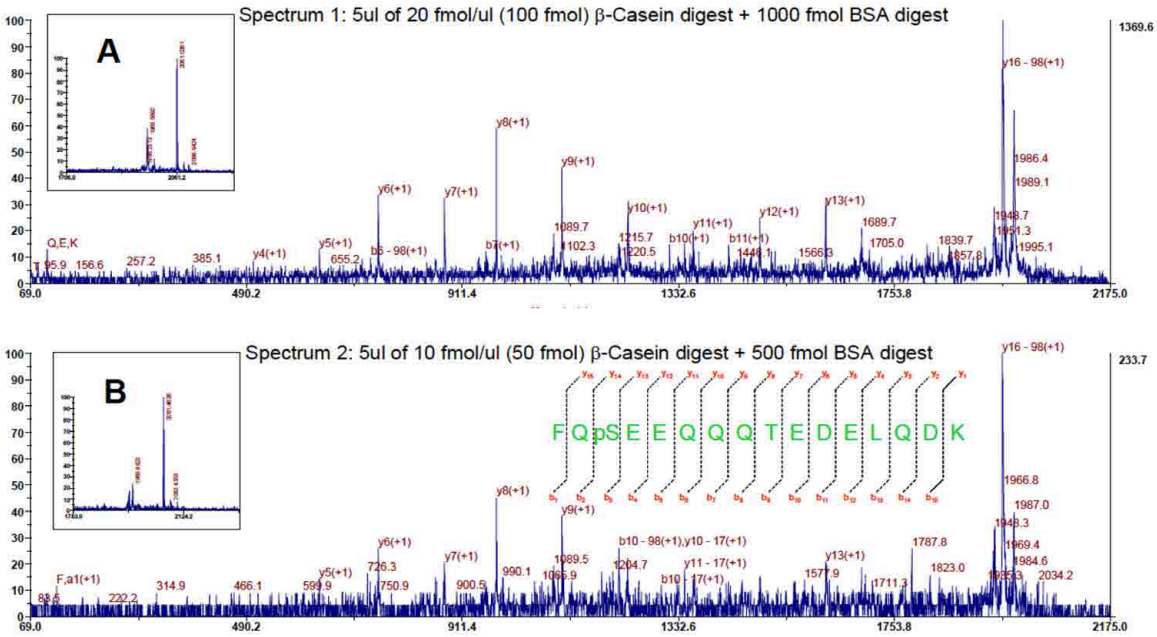


Figure 2.3.5 Two spectra from experiment 4 highlighting the MS/MS of the 2061 ion from low femtomole amounts of β-Casein digest applied to the NTA3 Biochip in the presence of a 10 fold molar excess of BSA digest. Spectrum 1 is 5 µl of 20 femto mol/µl β-Casein digest (100 femtomol total) + 1000 femtomol BSA digest applied. Spectrum 2 is 5 µl of 10 femtomol/µl β-Casein digest (50 femtomol total) + 500 femtomol BSA digest applied. Inserts (A) and (B) show the parent ion MS for the respective MS/MS spectra. The sequence of the peptide is superimposed in spectrum 2 in green depicting the theoretical y- and b-ion series, which relate to the shown assignment of y- and b-ions in both spectrum 1 and 2 above in red writing.

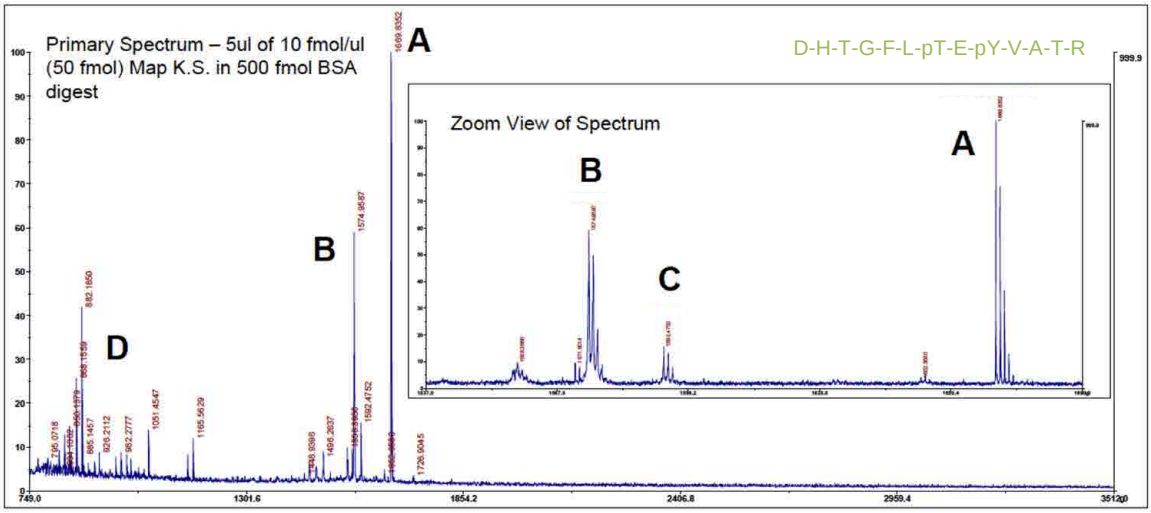


Figure 2.3.6 This spectrum was generated as part of the method development for the NTA3 Biochip, which was conducted on a synthetically generated doubly phosphorylated (pT, pY) peptide Map Kinase Substrate, sequence in green. Peak (A) m/z = 1669.83 represents 5 µl of 10 femtomole/µl (50 femtomole) peptide applied to the NTA3 Biochip in the presence of a 10-fold molar excess of BSA. Peaks (B) m/z = 1574.96 and (C) m/z = 1592.47 represent metastable fragment ions generated by the loss of the phosphate in MS mode. Peaks in the region marked (D) are primarily matrix adducts. This example shows that the affinity capture abilities of the NTA3 Biochip are not limited to singly phosphorylated peptides.

As mentioned earlier the NTA3 Biochip was at an early proof of concept stage when I finished working with it. With the benefit of hindsight I believe that the variability in concentration I observed could be fixed by returning the surface of the capture zone back to a similar composition to the X3 Biochip in order to create a similar hydrophilic gradient and facilitate a concentration event that is not based on a normal evaporative crawl, but similar to the unique snap-lock/washing event of the X3 Biochip.

I propose that whatever ligand is conjugated to the capture zone would be more efficient if a photo-cleavable linker group was used between the SAM and the ligand. This linker group would enable the ligand to be removed after the analyte has conducted the affinity capture and the contaminants were removed so as to facilitate co-migration of the analyte, ligand and matrix into the central analysis zone. I believe that this would remove the problem of non-concentration seen by the 20% COOH chips with NTA attached.

2.4.4 Limit of Detection and Capabilities of the Concentrating Desalting (RP3) Biochip on Peptide Samples from SDS-PAGE gel separated proteins using Applied Biosystems 4700 Analyser

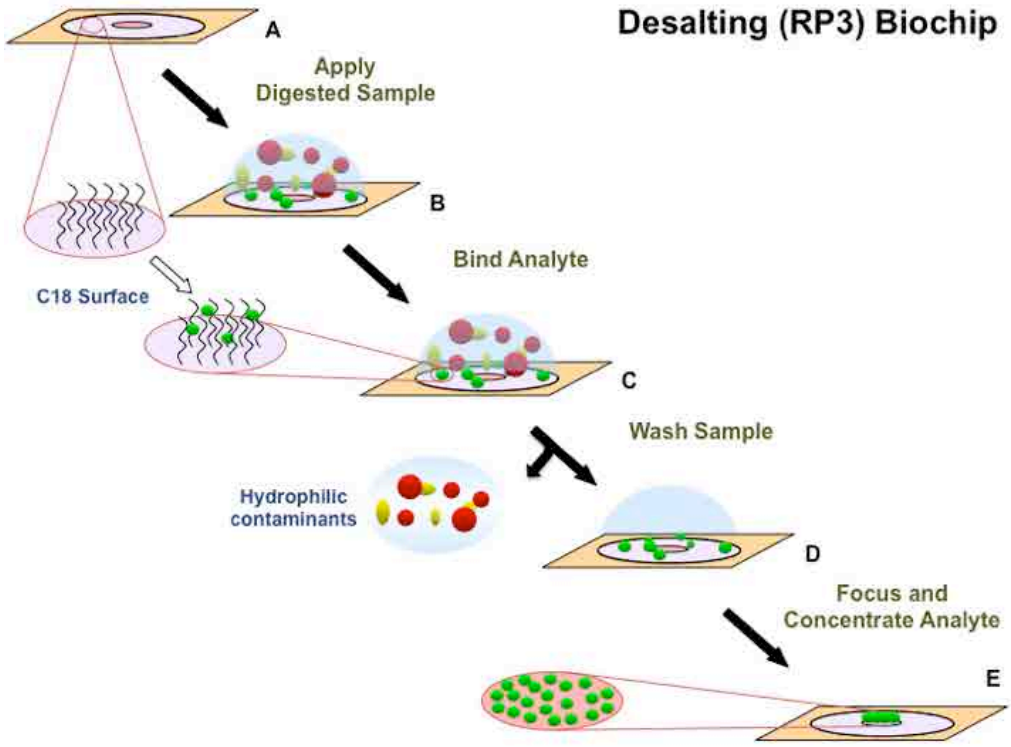


Figure 2.4.1 Schematic representation of the general methodology for the utilisation of the reverse phase (hydrophobic) capture and concentrating RP3 Biochip. (A) The liquid retention zone is populated with the C18 surface, shown as black wavy lines. (B) The sample is applied to the surface once the surface chemistry has been prepared and activated, with other contaminants present in the sample, being generally salts, shown as red and yellow balls and peptides as the green balls. (C) The mixture is allowed to incubate and equilibrate, allowing the peptides to associate (bind) through the hydrophobic interaction chemistry of the C18 surface. (D) A washing step is applied to remove any unbound species, which are hydrophilic in nature, leaving behind primarily peptides. (E) A solution containing the matrix facilitates the concentration of these peptides and the matrix in a co-migratory event towards the centre of the CDVW. This results in the concentration, crystallisation and presentation of the peptides and matrix in the central analysis zone for MALDI MS analysis.

The RP3 Biochip presented a multitude of problems from the earliest days of use, which I attribute in some degree to an ideology within the research group that the platform “already worked” and there was no real method development need. This resulted in pushing the experiments towards more complex and interesting samples, rather than performing more replicates of the normal standard experiments on peptide standards with and without different salts at various concentrations. A good example of this is that I observed from my initial experiment on standard proteins fractionated on BioRad Criterion gels that the resulting in-gel digest samples caused a film across the CDVW and no concentration and no acquisition of signal. I then switched to Invitrogen gels and this film was not evident. What caused this difference between these two types of gels and what was causing this film that interfered with the concentration event was not further investigated since the problem was alleviated by the use of Invitrogen gels. It would have made more sense, in hindsight, to try and determine the cause of this issue and see if it could be avoided, as it is possible it was occurring to some degree in our ongoing experiments and negatively impacting on our subsequent results. However, the lack of reproducibility of the Biochips, whether due to sensitivity to contamination or unspecified manufacturing irregularities, made it extremely difficult to validate this platform for use in this particular application.

Early experiments conducted at Lumicyte in Feb 2005 were done on gel plugs of BSA, and various contaminants were seen in the spectra. We then analysed a series of five 1 in 3 dilutions across the surface of the chip and noticed that the contamination seemed to disappear on the 3rd and 4th dilutions. The number of dilutions may have increased the signal of the peptides of interest and reduced the contaminants, but it equalled out with normal levels of sensitivity of detection. This led me to postulate that the contamination showed a concentration dependent effect and it would be difficult to show any advantage of utilising this particular aspect of the SAM Biochip technology.

Figures 2.4.2 - 2.4.4 are three peptide mass finger print (PMF) spectra from a 1D Gel band of Carbonic Anhydrase that was applied across the gel in a dilution series to ascertain the LOD of peptides using the RP3 Biochip. I was only able to generate a mascot score for the highest concentration sample, which were replicates applied to wells A2, A3 and A4. The other dilutions did not generate any viable results. The

identification of keratin in the spectra accounts for some of the observed contaminants but there were also a large number of unassigned peaks that could not be attributed to a particular contaminant, despite our best efforts at spectral interpretation. In addition, I have very rarely had problems with keratin contamination when preparing digests of gel separated proteins in the past, yet significant keratin contamination occurred consistently when I used the RP3 Biochips for analysis of gel separated proteins. These three spectra are proof that the RP3 Biochip could capture peptides on the surface from a gel digest, remove any salts through the washing steps and concentrate the sample for MALDI MS. However, the RP3 Biochip seemed to also concentrate whatever contaminants that were present too, reducing the ability of the technique to work at the desired low levels of detection beyond gel staining limits of detection.

The four peptides used to identify Carbonic Anhydrase were the same for all three replicates, but the signal strength and identity of the strongest Carbonic Anhydrase peaks are not the same for all three replicates; the most intense peaks are, m/z=1581.81 (A2), m/z=2198.23 (A3) and m/z=2198.23 (A4). The fact that there are two different peptides with significant ionisation differences from one well to the other leads to three immediate possibilities. The first is that the level of keratin and unknown contamination is different in each due to the on chip dilution series, and this has had a variable ion suppression effect on the samples within the mass spectrometer. Secondly, the Keratin and unknown contaminants may have a higher affinity to the SAM chemistry than the peptides of interest, hence the contamination is stronger in well A2 than A3 and A4, in that order. Thirdly, this variable contamination may have affected the RP3 Biochips ability to reproducibly retain the peptides on the surface of the CDVW throughout the washing procedure, resulting in a different number or ratio of peptides available for ionisation.

During the course of the experiment presented here for Carbonic Anhydrase I also tested a four peptide mixture similar to the X3 Biochip experiments, except that 25 mM of NH_4HCO_3 and NaCl were added to simulate a salt contaminated environment. Figure 2.4.5 shows the spectra generated on the four peptide sample with salts added. It is evident that there are little to no salt adducts and that the RP3 Biochip has removed these from the sample while retaining the peptides for analysis. Why the ACTH 18-39 peptide showed such a low ionisation compared to the other three

peptides is unknown. Ion suppression is unlikely because the known concentration and ratios used in earlier experiments generated equal ionisation and not the disproportionate ionisation observed here. One possibility is that the majority of the ACTH 18-39 peptide was removed in the washing steps. This could be attributed to the difference in the charge and the hydrophobicity of the peptides. The pI for Bradykinin, Angiotensin I and Neurotensin are approximately 12, 7 and 9, while ACTH 18-39 is 4. Additionally, ACTH 18-39 has a considerable number of acidic amino acids and this charge density may also contribute to a differing binding capacity on these Biochips.

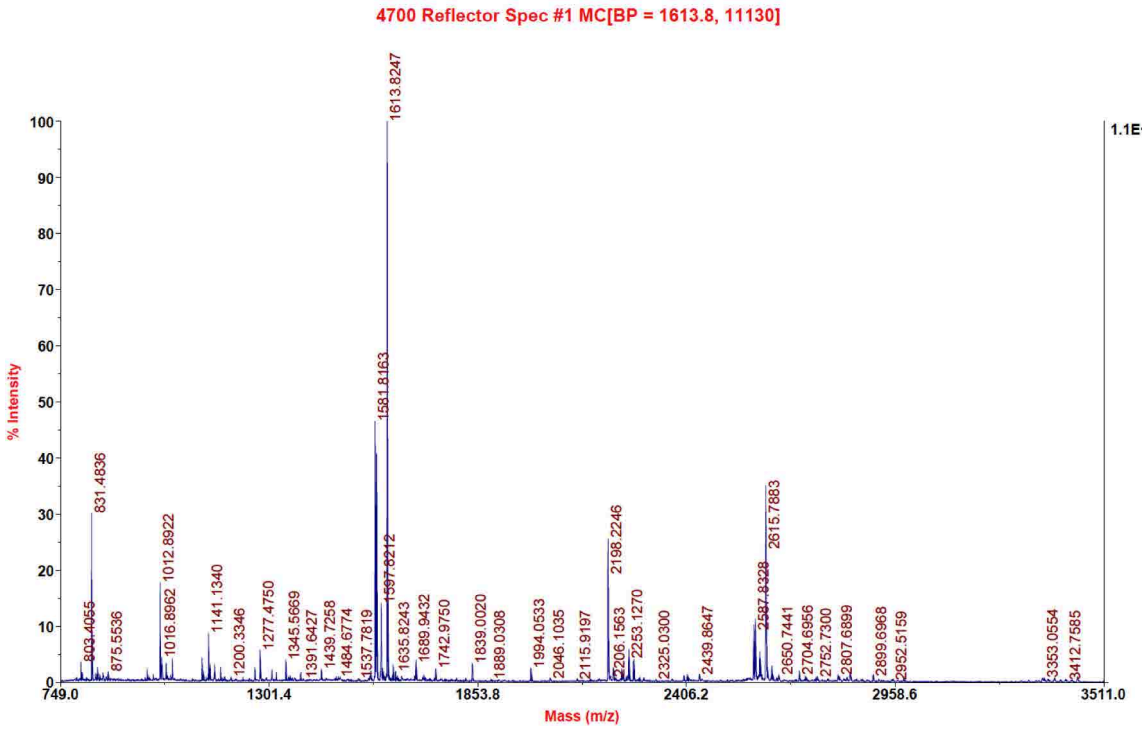


Figure 2.4.2 MS spectra from a 1D gel digest of Carbonic Anhydrase (bovine) applied to and purified by a RP3 Biochip and analysed in an Applied Biosystems 4700 mass analyser. This is the first of the three technical replicates (A2) of the digested supernatant that was analysed in adjacent wells on the same RP3 Biochip. The Mascot score for the PMF search identified Carbonic Anhydrase with a score of 44. Despite the large number of peaks present only 4 peptides were used in the identification. Upon closer inspection there is considerable Keratin contamination as well as other unknown contaminants. Re-searching in Mascot after removing the known keratin peaks increased the score to 47. The source of the additional peaks that seemed to suppress the ionisation and identification of the expected peptides from Carbonic Anhydrase could not be identified. Note the signal intensity of the highest peak in the spectrum was 11000 for ion m/z=1613.82, which is a Keratin peak and the second largest peak was at approximately 45% of this for ion m/z=1581.82, which is a Carbonic Anhydrase peak. There were no noticeable salt adducts upon visual inspection of the spectrum, showing that the RP3 Biochip can remove salt contaminants from a complex peptide sample.

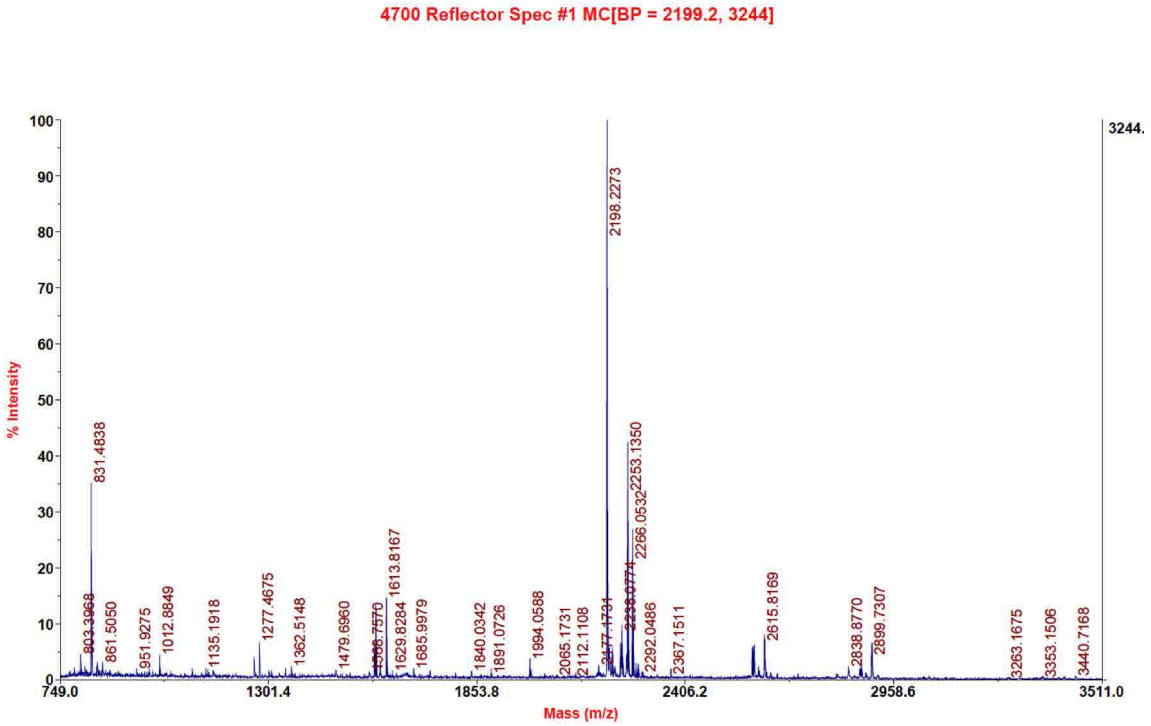


Figure 2.4.3 MS spectra from a 1D gel digest of Carbonic Anhydrase (bovine) applied to and purified by a RP3 Biochip and analysed in an Applied Biosystems 4700 mass analyser. This is the second of the three technical replicates (A3) at dilution 1:2 of (A2) of the digested supernatant that was applied and analysed by an adjacent well on the same RP3 Biochip. The Mascot score for the PMF search identified Carbonic Anhydrase with a score of 43. Despite the large number of peaks present only 4 peptides were used in the identification. Upon closer inspection there is considerable Keratin contamination as well as other unknown contaminants, similar to the first sample as shown above, although of a less pronounced nature. Re-searching in Mascot after removing the known keratin peaks increased the score to 47. The source of the additional peaks that seemed to suppress the ionisation and identification of the expected peptides from Carbonic Anhydrase could not be identified. Note the signal intensity of the highest peak in the spectrum was 3244 for ion m/z=2198.23, which is a Carbonic Anhydrase peak and the Keratin peak m/z=1613.82 is 15% of this, though is pronounced compared to the other peaks in this spectrum. There were no noticeable salt adducts upon visual inspection of the spectrum, showing that the RP3 Biochip can remove salt contaminants from a complex peptide sample.

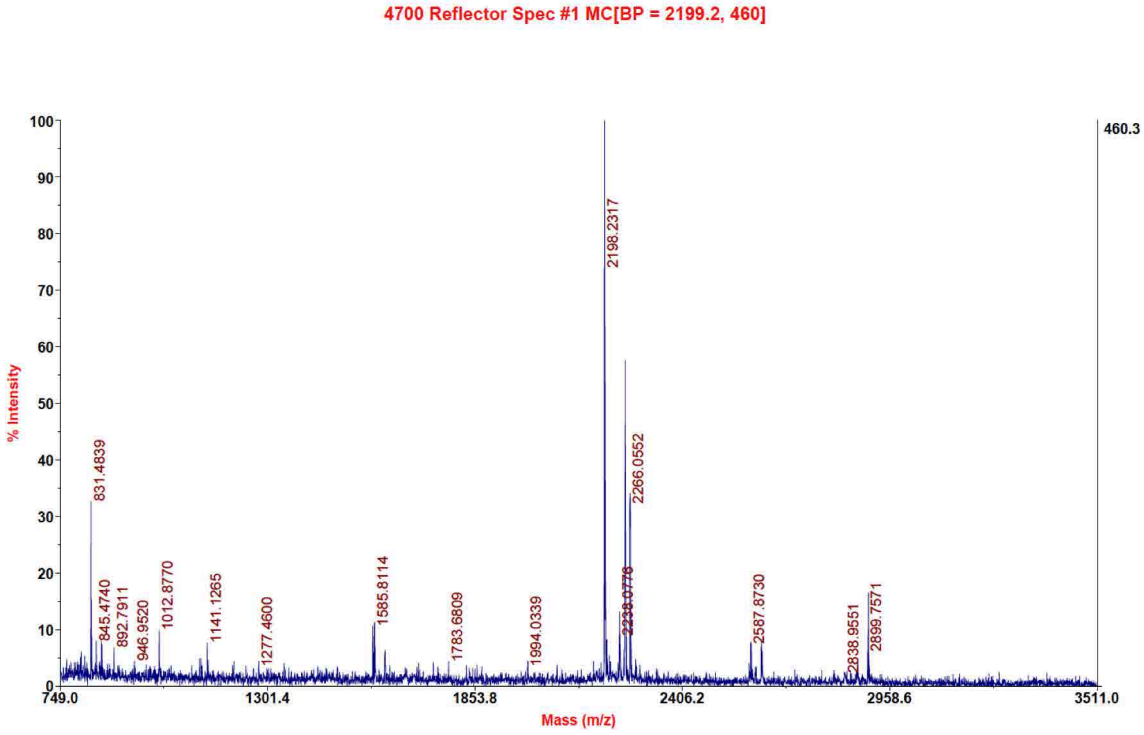


Figure 2.4.4 MS spectra from a 1D gel digest of Carbonic Anhydrase (bovine) applied to and purified by a RP3 Biochip and analysed in an Applied Biosystems 4700 mass analyser. This is the third of the three technical replicates (A4) at dilution 1:2 of (A3) of the digested supernatant that was applied and analysed by an adjacent well on the same RP3 Biochip. The Mascot score for the PMF search identified Carbonic Anhydrase with a score of 41. The same 4 peptides were used in this identification as for the other two replicates despite the reduced number of peptides present. The keratin contaminations as well as other unknown contaminants, similar to the first sample as shown above, were of a less pronounced nature. Re-searching in Mascot after removing the known keratin peaks increased the score to 44. The source of the additional peaks that seemed to suppress the ionisation and identification of the expected peptides from Carbonic Anhydrase could not be identified. Note the signal intensity of the highest peak in the spectrum was 460 for ion m/z=2198.23, which is a Carbonic Anhydrase peak and the Keratin peak m/z=1613.82 is not detected, though there were still several keratin peaks identified. There were no noticeable salt adducts upon visual inspection of the spectrum, showing that the RP3 Biochip can remove salt contaminants from a complex peptide sample.

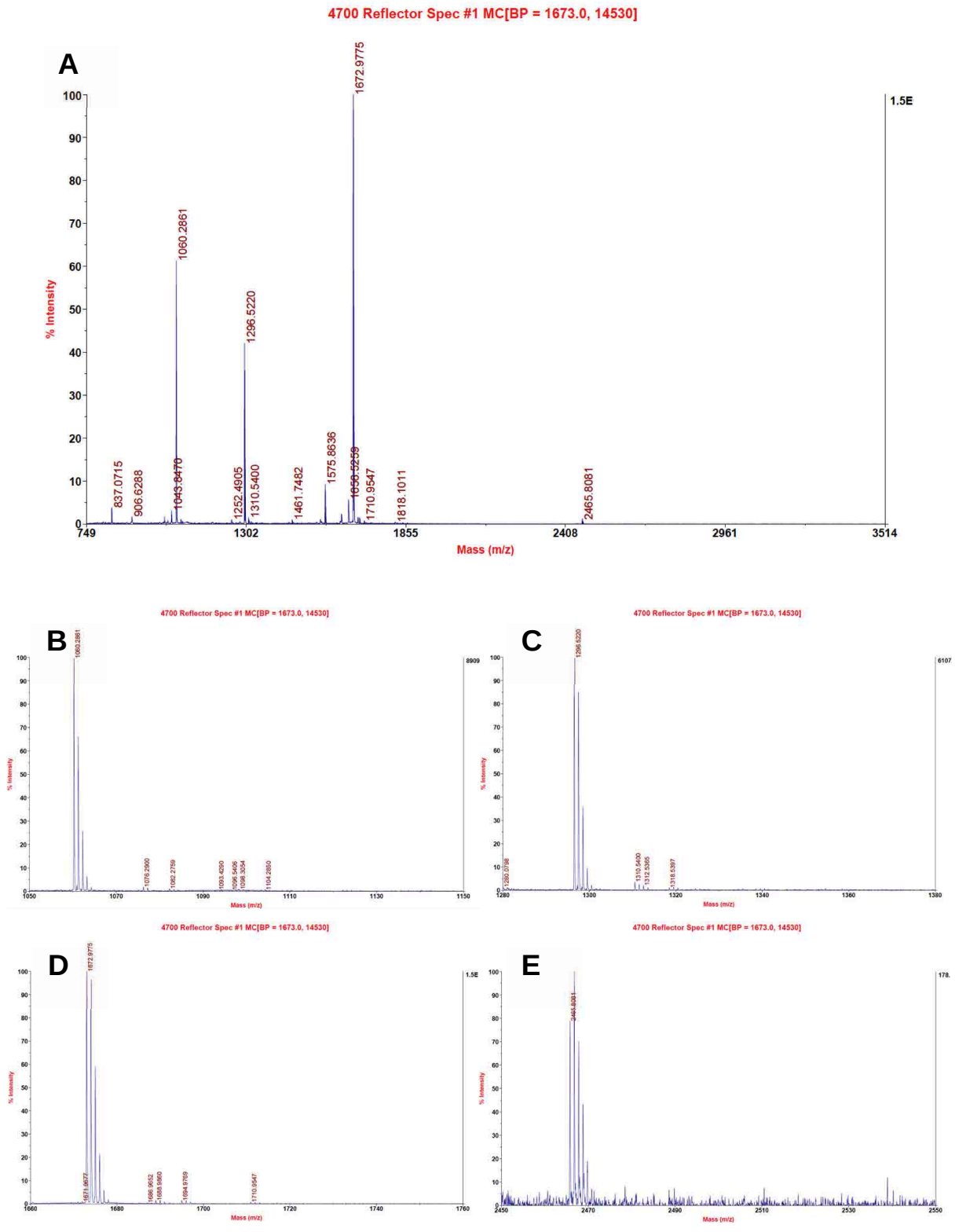


Figure 2.4.5 (A) MS spectra of a four peptide mixture Bradykinin (1060), Angiotensin I (1296), Neurotensin (1272) and ACTH 18-39 (2465) in a ratio of 1:1:1:3 solubilised in 25 mM NH_4HCO_3 and 25 mM NaCl, applied to and purified by an RP3 Biochip using the TFA wet method, analysed in an Applied Biosystems 4700 mass analyser. (B-E) There were no noticeable salt adducts upon visual inspection of the spectrum as shown in the superimposed zoomed view for each of the four peptides, showing that the RP3 Biochip can remove salt contaminants from a peptide sample. The low signal strength for ACTH relative to the other three peptides was unexpected when compared to the earlier X3 Biochips and Std-MALDI data for these three peptides mixed in the ratio 1:1:1:3.

2.4.5 Variability of the Concentration and Crystallisation Event

In the early stages of this project we were given detailed instructions on how to implement the Biochip technique, especially with respect to the addition of the matrix solution, which would induce the unique co-migration of the analyte and matrix, resulting in the concentration of the two followed by crystallisation of the matrix and analyte mixture. Due to the unique juxtaposition of surface chemistries in the orientation of concentric circles, this induced a unique concentration event termed a snap-lock/washing event. The surface tension differences of the exposed functional groups restrict flow of the droplet to the centre and holds it across the entire liquid retention zone till the volume is almost completely evaporated and the energy state for the droplet to be a micelle is more energy advantageous than to hold it across the surface. This creates a flow of liquid, analogous to a washing effect, across the surface due to the instantaneous manner in which it occurs, hence the term snap-lock is used. This snap-lock event is different to a normal evaporative crawl observed on functionalised surfaces like the AnchorChip^{TM} and on non-functionalised two-dimensional surfaces.

The primary matrix solution used for most experiments was 2 µl of 84:13:3 ACN:EtOH:0.1% TFA (aqueous component containing 5 mM Ammonium citrate) with 0.063 mg/ml of CHCA, as per the manufacturers specifications. A DHB matrix solution was also prepared, composed of 0.65 mg/ml DHB dissolved in 8:2 ACN:0.1% TFA (aqueous component containing 5 mM Ammonium citrate). Since the manufacturers of the 4700 recommended using CHCA and that it was superior to DHB for LOD studies, this study was directed towards using CHCA as the primary matrix solution, even though the majority of the method development prior had been with DHB. Upon later reflection the main reason for the recommendation from ABI to use CHCA over DHB was due to the CHCA forming more uniform and dispersed small crystal structures across a MALDI plate compared to the needle like crystals formed by DHB. Thus, using CHCA would remove the hot spots and ‘blank’ laser shots without ionisation in the areas where no crystals were when using DHB. Since we were concentrating the sample and matrix these hot spots and void areas are greatly reduced if not removed entirely and would not have posed a problem. Perhaps using DHB as opposed to CHCA may have reduced the variability of the

concentration event seen. Due to a limited supply of Biochips this type of detailed study could not be undertaken.
Initial testing of the X3 Biochip platform gave variable results with respect to signal strength for identically spotted samples in side by side wells with an LOD for Glu-1-Fibrinopeptide B at mid femtomole level. The expected formation of defined (tight) concentrated analysis zones of analyte and crystals of uniform shape was not observed, but protrusions permeating from the perimeter of the analysis zone were seen even though they could not be readily explained. Figures 2.5.1 and 2.5.2 below, illustrate some of this variability in the concentration event observed at the beginning of the project.

After satisfying myself that this was not simply due to human error or instrument tuning, I postulated that laboratory temperature and humidity might be playing a role. I purchased a temperature and humidity meter and made daily measurements to try and factor this into the method for matrix addition, concentration (focusing) and re-focusing. If there is a problem with the crystallisation of the CHCA, then refocus for low humidity ≤ 35% with 2 µl 9:1 (ACN:Water) and at high humidity ≥ 65% with 1 µl 98:2 (ACN:Water). This introduces another variable in the experimental process, as the size of the final concentration zone and the types of crystals are still inconsistent, which can cause ionisation variance.

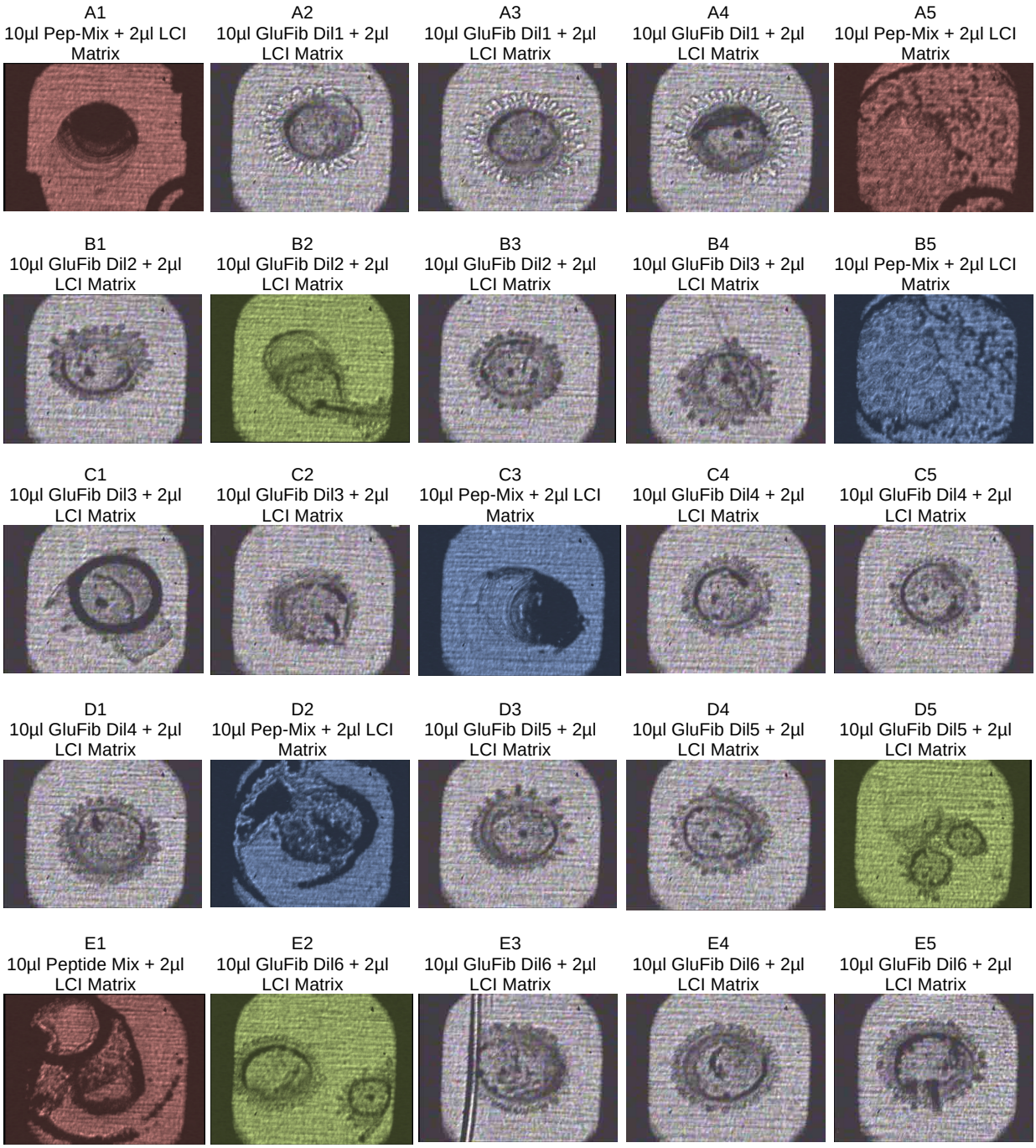


Figure 2.5.1 Images captured by the camera inside the 4700 instrument of the X3 Biochips before the identified oily residue and implementation of the DCM washes. Experiment was a simple LOD of Glu-1-Fib diluted 1 in 10, from Dil1 (3.2 picomol/µl) to Dil6 (32 attomol/µl). Wells A2-A4 exhibit opaque protrusions permeating outwards from the perimeter of the analysis zone. Well B2 shows a smearing of the matrix and analyte, while wells D5 and E2 show two separate concentration centres. All of the wells do not show clearly defined crystal spots but what appears to be an amorphous aggregation and precipitation. The red wells A1, A5 and E1 are pre-scribed at manufacture and used as calibration standard spots, because there are no uniform CDVW's here any abnormalities can be ignored. The blue wells B5, C3 and D2 are used as calibration standard wells and these too exhibit irregular concentration.

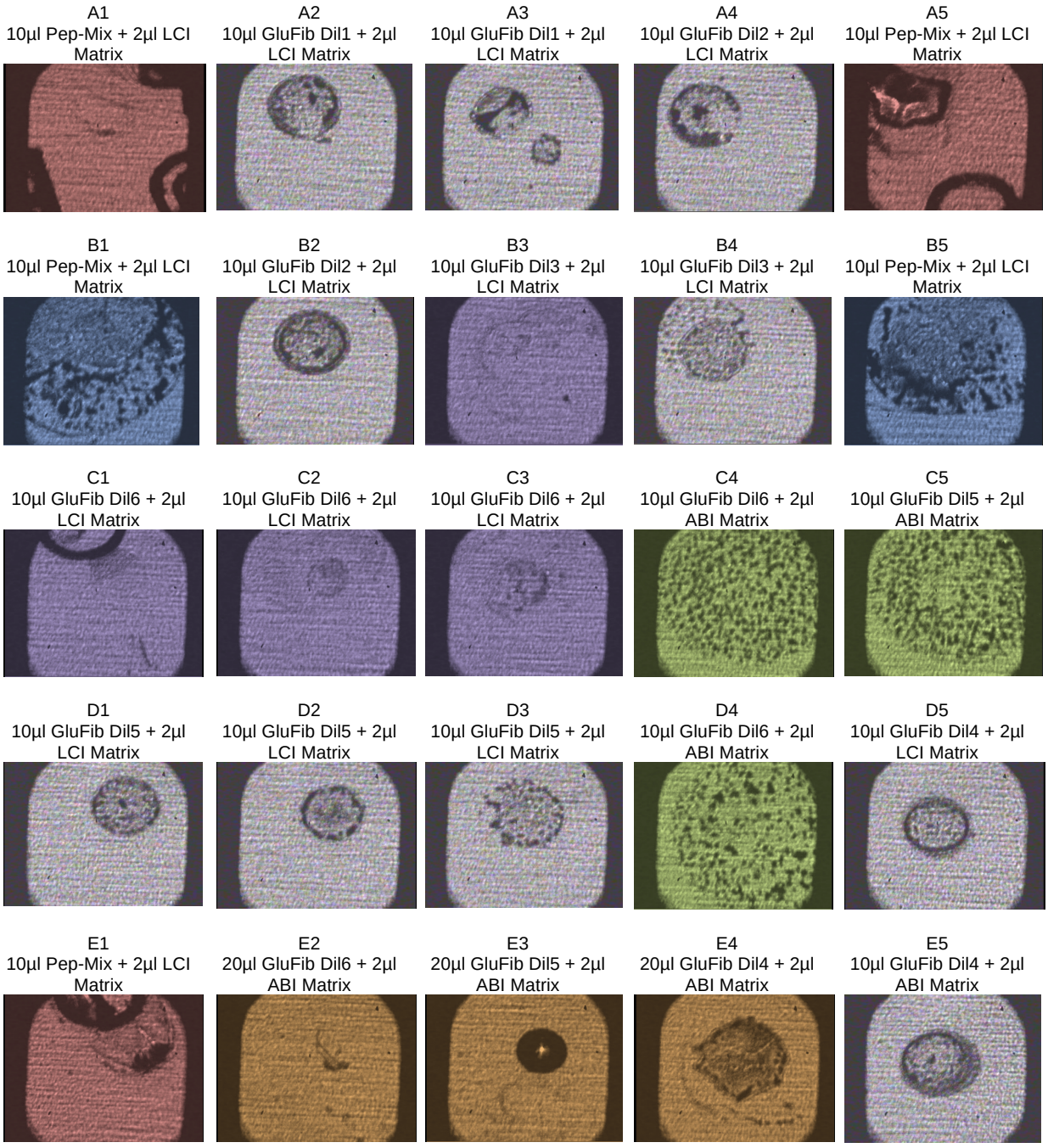


Figure 2.5.2 Images captured by the camera inside the 4700 instrument of the X3 Biochips before the identified oily residue and implementation of the DCM washes. Experiment was a simple LOD of Glu-1-Fib diluted 1 in 10, from Dil1 (3.2 picomol/µl) to Dil6 (32 attomol/µl), being a replicate experiment to the one shown above. Wells C4, C5 and D4 show more definable crystals though less of a concentration effect when the concentration of the matrix is increased to 4 mg/ml. Well E2-E4 had 20 µl analyte and the variability of crystal formation was not explainable at the time. Wells C1-C3 and B3 have reduced to no visible crystals or concentration. Well A3 has two separated aggregations of concentration. Wells B4 and D3 show undefined perimeters for the analysis zone. The red wells A1, A5 and E1 are pre-scribed at manufacture and used as calibration standard spots, because there are no uniform CDVW's here. The blue wells B5, C3 and D2 are used as calibration standard wells.

I have mentioned that the unique ‘snap-lock’ concentration event was not being observed and sometimes the migration was to the outside of the analysis zone and other times it was non-existent. I did notice that the packaging container for the Biochips had an oily residue inside. These containers were standard ABI plate packaging containers, since the Biochips were fashioned from blank ABI plates. Lumicyte used the ABI packaging and had not used them before, which lead me to the idea that the surface of the SAM had been contaminated with an oily residue of unknown nature from the packaging. I tested this by placing a single chip into a glass dish filled with dichloromethane (DCM) and within a minute I noticed the DCM solution turning milky, with the production of a cloudy/oily residue throughout the surface of the liquid the longer the Biochip remained within as I agitated the liquid across the surface of the Biochip.

I then implemented a pre-washing of all the chips that had been supplied in the ABI plate packaging with a DCM bath for 10 min followed by a second DCM bath for 10 min followed by an EtOH bath for 10 min, followed by soaking the chips in ultra pure water overnight. I then applied two more baths of 5 minutes each the day of the experiment, the first being in 3% NH_4OH and then ultra pure water. Argon or Nitrogen gas was blown over the surface to remove all liquid and dry the Biochip under a dust cover (desiccator without desiccant).

This washing regime seemed to recover the CDVWs, for they become visible as I moved the Biochips from wash to wash, and upon application of the primary matrix solution, with or without analyte. Based on visual inspection, the concentration event worked better than before, with clearer and more defined analysis zones, though there was still a high degree of variability, which was made clearer when viewed under the 4700 camera system as shown in Figures 2.5.3 and 2.5.4 below. The images depicted here outline a few troubleshooting and optimisation experiments conducted to test the optimal concentration or volume of the matrix applied to enhance the concentration (focusing) event and signal generation. These images are from RP3 Biochips.

Although the DCM washing procedure reduced the observed variability in surface properties, they were still relatively inconsistent. The most likely explanation was that the oily residue could not be completely removed during the wash steps and the

chips could not be returned to their original state in entirety. Over the course of time and many more experiments, I observed repeated high variability of the Biochips, with respect to their ability to focus (concentrate) on all three platforms.

New chips were made and encapsulated in different containers for all the future experiments. Figure 2.5.5 shows that the newly manufactured Biochips shipped in different transport containers performed as expected. There are tight and well defined concentrated spots of matrix and analyte with similar size, shape and crystal structure based on visual inspection. Additionally, the unique snap-lock concentration event was seen rather than the evaporative creep observed when the chips were not working optimally.

Despite the problem with the transportation of the Biochips in contaminated containers being solved, I observed on numerous occasions variability of concentration, which I have attributed to the inability of the surface chemistry to handle large deviations in contaminants and deviations in the surface chemistry. I conducted a simple experiment to test three different manufacturing dates of the chips to see if there was any significant variability, with respect to the manufacturing of one batch to another. Figure 2.5.6 is an example of three supposedly identical Biochips manufactured on three different dates all performing differently. The ability of the chips to concentrate across the one Biochip was reproducible but from chip to chip they were different, which strongly suggested that manufacturing variability was a significant factor in the inherent variability I had observed.

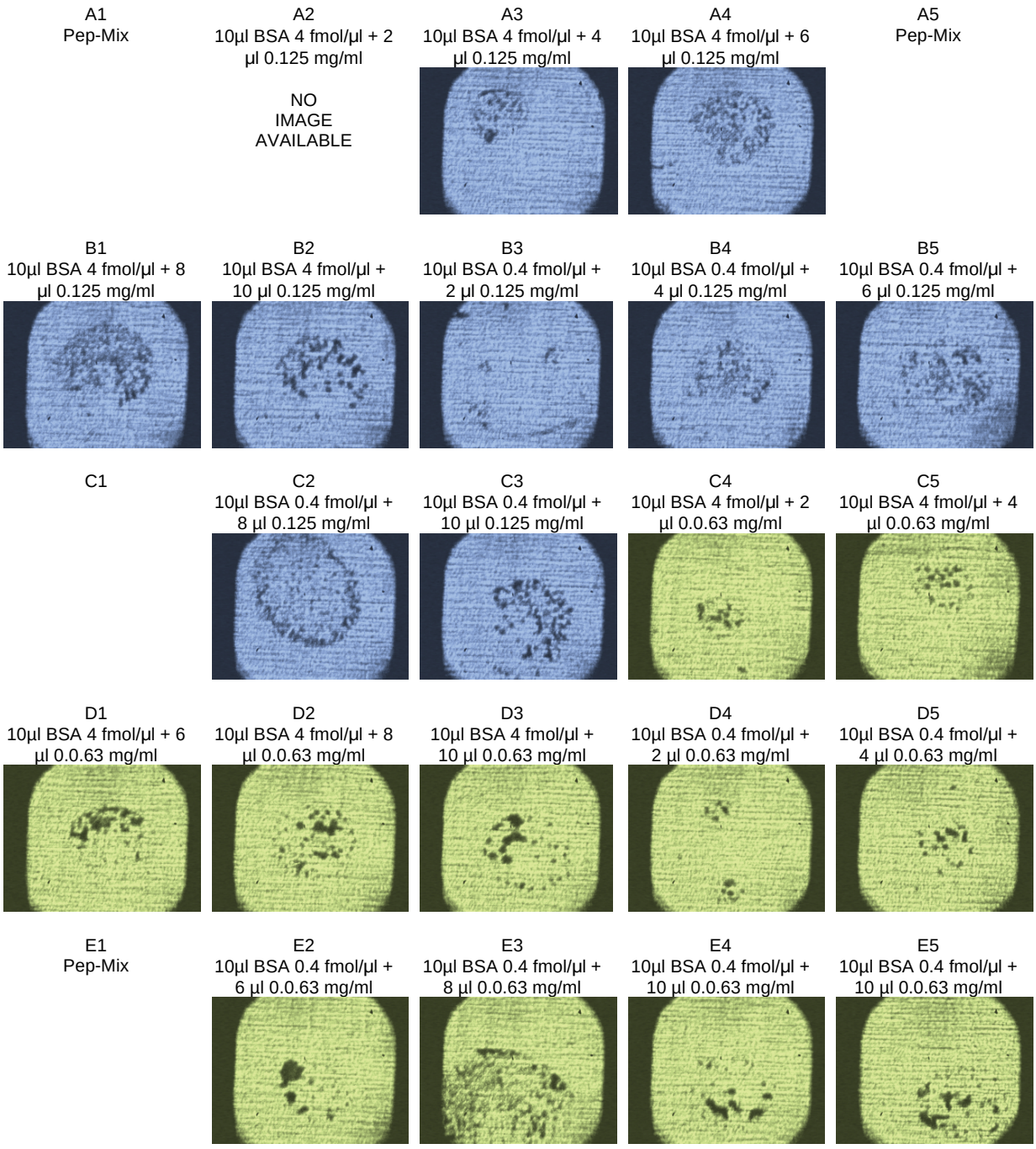


Figure 2.5.3 Images captured by the camera inside the 4700 instrument of the RP3 Biochips after the implementation of the DCM-EtOH-NH4OH-H2O washes. Experiment was a simple LOD of BSA in Urea and testing the effects of changes in matrix concentration or load amount added. The blue wells A3 - C3 are at a matrix concentration of 0.125 mg/ml. The green wells C4 – E5 are at the recommended matrix concentration of 0.063 mg/ml. The lack of reproducibility in the focusing event is shown in the images above. The wells A1, A5 and E1 are pre-scribed and used as calibration standard spots, because there are no uniform CDVWs here and are omitted from the images.

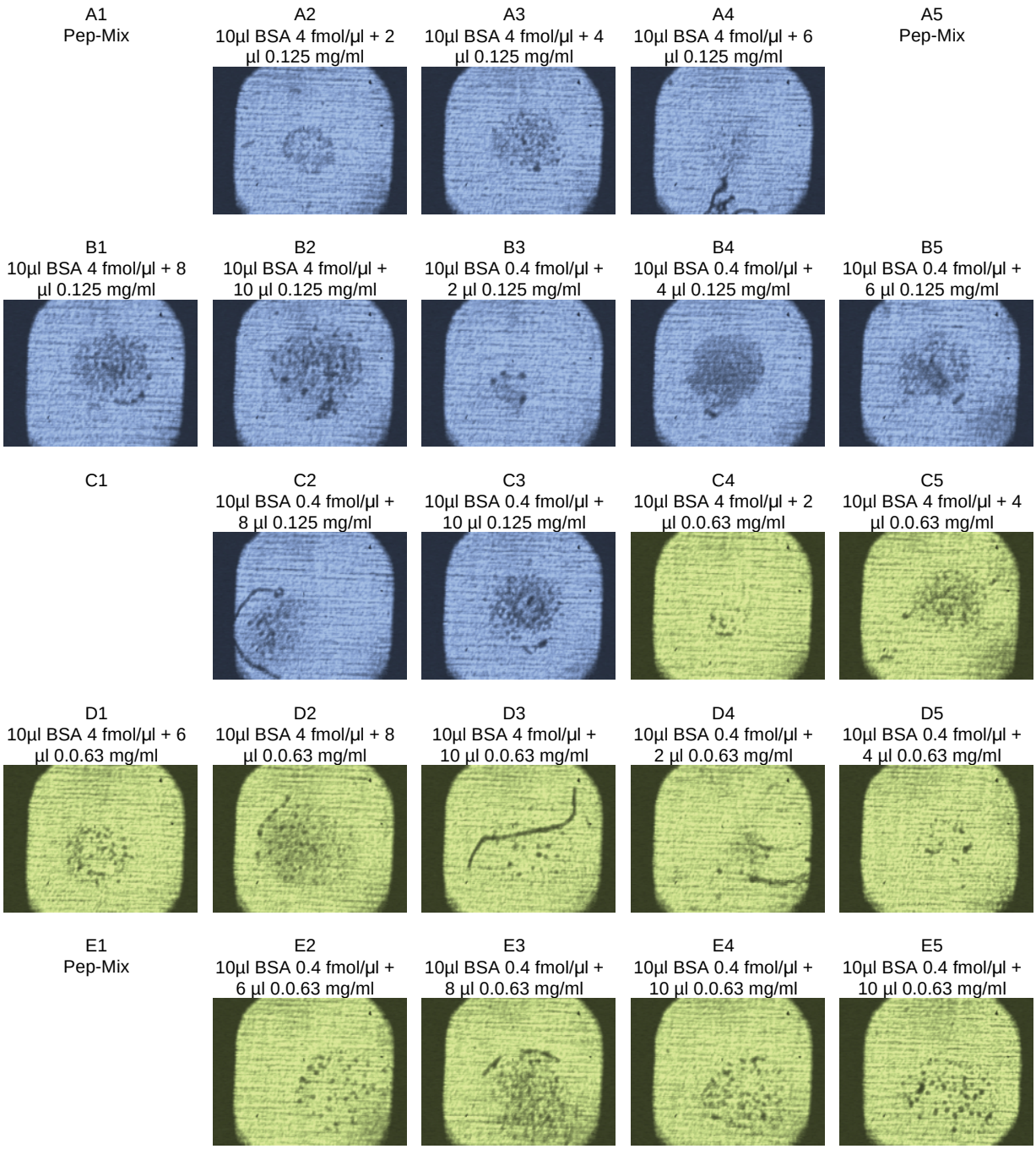


Figure 2.5.4 Images captured by the camera inside the 4700 instrument of the RP3 Biochips after the implementation of the DCM-EtOH-NH4OH-H2O washes. Experiment was a repeat of the earlier Figure 2.5.3, being a simple LOD of BSA in Urea and testing the effects of changes in matrix concentration or load amount added. The blue wells A3 - C3 are at a matrix concentration of 0.125 mg/ml. The green wells C4 – E5 are at the recommended matrix concentration of 0.063 mg/ml. The lack of reproducibility in the focusing event is shown in the images above. The wells A1, A5 and E1 are pre-scribed and used as calibration standard spots, because there are no uniform CDVWs here and are omitted from the images.

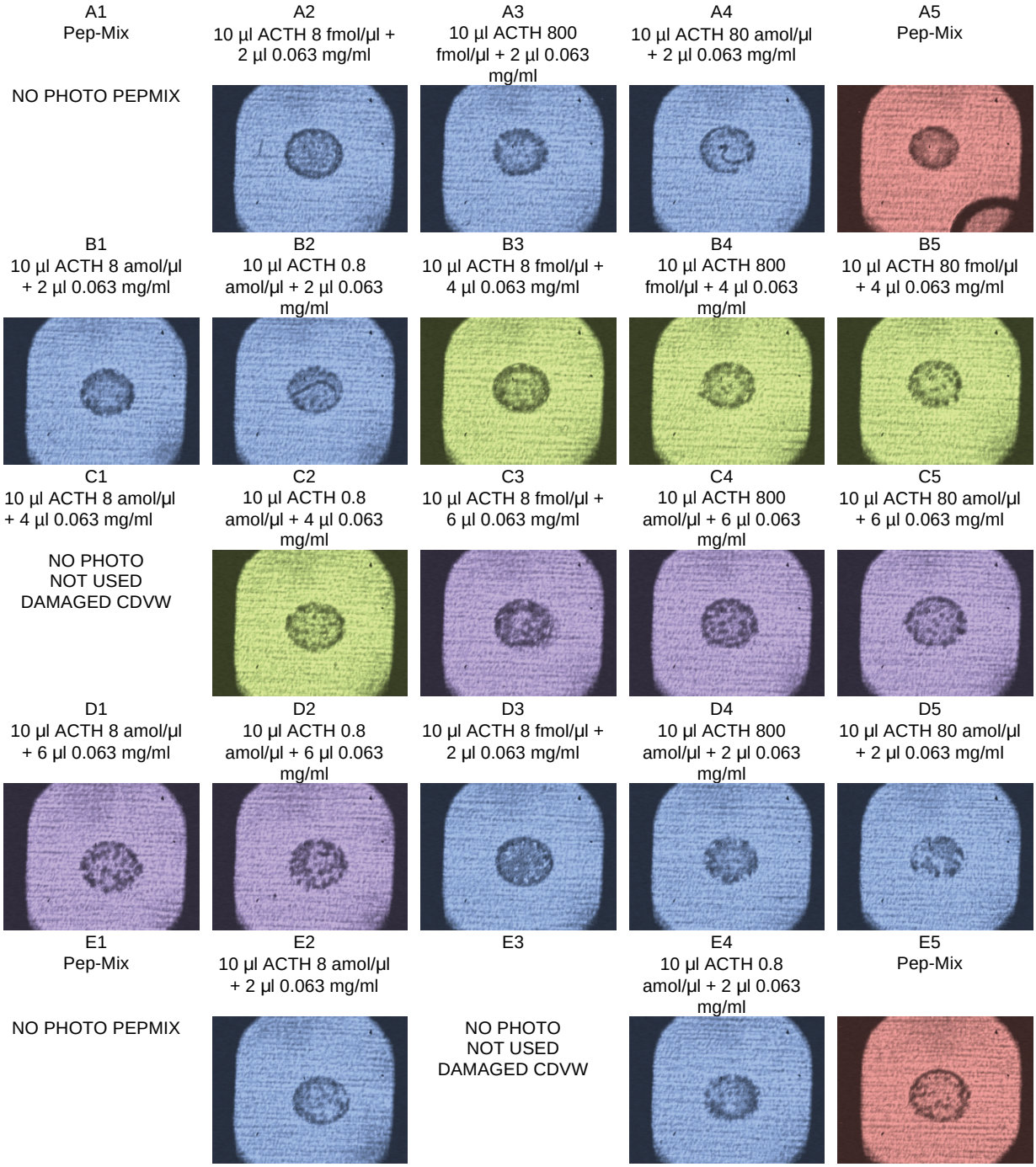


Figure 2.5.5 Images captured by the camera inside the 4700 instrument of the X3 Biochips after the manufacture of new chips and the transportation packaging being changed. Visual inspection shows tight and reproducible focusing of the matrix and analyte across all CDWV's and throughout the range of 2, 4 and 8 µl additions of the matrix at 0.063 mg/ml. LOD testing on ACTH (18-39) was 80 attomole here (10 µl at 8 amol/µl) across all three matrix amounts. Wells C1 and E3 were damaged and no photo was taken at the time of the experiment. Wells A1 and E1 were calibration standards of a standard four peptide mixture hence no photos were taken here either, although wells A4 and E5 showed equal focusing to the other wells, despite having a different sample to the ACTH alone.

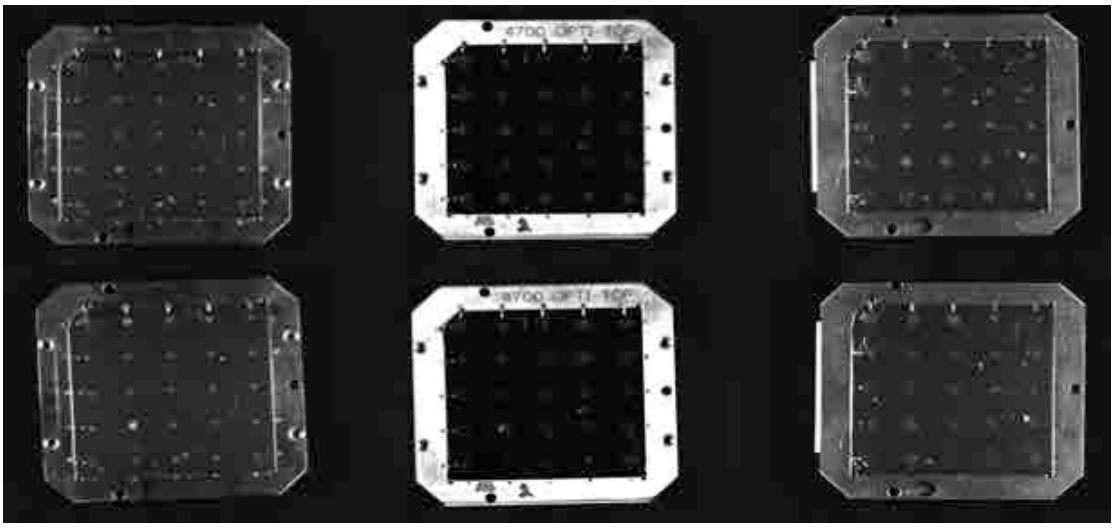


Figure 2.5.6 The top three images, from left to right, represent three X3 Biochips manufactured at three different times points across a four month period after a pure solution containing the peptide Angiotensin and matrix was applied for LOD detection studies and to assess the variability with respect to manufacture irregularities. The bottom three images, from left to right are the same Biochips as in the line above except they have undergone a re-focusing stage to try and get the sample and matrix to concentrate. Plate 1 showed acceptable concentration, though the signal generated was not discernibly different from plate 3. Plate 3 showed sporadic concentration and the re-focusing did not seem to fix the problem, though there was the generation of signal in the mass spectrometer at lower levels than would be expected for the amounts of peptide applied. Plate 2 did not concentrate and displayed a clearly visible film of unknown cause that could not be resolved by re-focusing and no signal could be generated. I handled all three Biochips in an identical manner, leaving me to deduce that some unknown variable in either the manufacture or handling of the Biochips can cause both variability of concentration and ionisation, with the complete failure of the sample to concentration and ionise in some cases.

2.5 FUTURE VIEWS AND DIRECTIONS

The theory of the CDVW as depicted in Figure 2.0.2, 2.1.1, 2.3.1 and 2.4.1 is intriguing, and during the visualisation of the snap-lock concentration event it is astonishing to see the liquid being manipulated in such a counterintuitive way. However, the sensitivity to disruption of the micro-fluidic flow that is created by the surface chemistries, solvent and analytes leaves the technique exposed and with a limited range of possible uses.

The concept of being able to incorporate both affinity capture and concentration onto the one surface as depicted in Figure 2.0.1 and 2.3.1 is an enchanting idea and perhaps is the ultimate goal of such a technology. Taking into consideration the amount of scientific information the AnchorChip^{TM} and SELDI^{TM} have been able to achieve separately, one would expect even more exciting outputs from a system that combined the best aspects of both these techniques. If these techniques could be combined the envisaged increase in scientific knowledge justifies further investigation of the technique in my opinion.

The changes brought about by the surface chemistry when affinity ligands are bound alters and in some cases removes the hydrophilic gradient across the surface and the concentration event is altered. I propose that if a reversible linker group, for example a photo-cleavable linker, could be conjugated to the surface of the capture zone, this photo-cleavable linker could enable affinity capture and the removal of contaminants. This could then be followed by photo-oxidation/reduction to remove the ligand and bring the surface chemistry back to a form that would facilitate the snap-lock concentration event so that the analyte of interest, affinity linker and matrix would co-migrate for presentation in the analysis zone for MALDI MS. The possibility of ion suppression due to the ligand being present in the analysis zone would have to be addressed and tested against various known compounds.

Additionally, the fact that the Biochip is single use in any or all of the chemistry types supplied needs to also be addressed, due to the cost per well for a user, making the technique hard to justify economically. Perhaps a washing protocol of numerous solvents, similar to the DCM washing protocol I developed coupled with sonication,

that brings the surface back to originality after use, would establish the re-useability of the technique and enhance the economic justification of further development.

In closing, this technique, at its core, aims to manipulate the flow of liquids and analytes across two-dimensional surfaces, so as to induce crystallisation at a higher concentration than the result of a normal evaporative effect. The external forces at play on solvents and analytes based on the temperature, humidity, pressure and solvent type, are all variables that need to be controlled if an investigator wishes to produce reproducible crystals and/or concentration. Despite the obvious and pragmatic technical need to control such factors as mentioned above, there are also the theoretical and unquantifiable aspects, based on the scientific knowledge at this point in time. Specifically, there remains significant gaps in the scientific knowledge base with respect to the complete mechanism that regulates the generation of a crystal, the nucleation event, and the facilitation of the growth of this singular crystal to grow and form the crystal mass [375]. One could postulate that the inability to understand the nucleation event is based on an inability of the instruments at this point in time to measure such a thing. Alternatively, it may be that the mechanism of transition from liquid to crystal is not a steady state flow which follows Newtonian physics and is measurable at all points, but perhaps involves a ‘packet’ like transition or quanta that is only explainable by quantum physics [376]. If there were a form of quantum mechanical mechanism at play, then one would not be able to observe the transition. It would simply occur, for it would be existing in both states simultaneously, till it is energetically favorable to be a crystal permanently, rather than a liquid or in this quantum mechanical state of flux. This renders the process not only difficult to understand at a mechanistic level but infinitely more difficult to control as this technique of a SAM in concentric circles sets out to achieve.