Enhancement Of Mass Spectrometry And Affinity Purification Methodologies Used To Acquire Novel Protein Identifications

by
Morgan Mitchell

A dissertation submitted to the Department of Biology and Biochemistry,

College of Natural Sciences and Mathematics

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in Biochemistry

Chair of Committee: Steven Bark

Committee Member: Chengzi Cai

Committee Member: Robert Schwartz

Committee Member: William Widger

University of Houston
August 2020

ABSTRACT

Mass spectrometry (MS)-based proteomics has become a primary tool to study proteins underlying infection disorders, antibiotic resistance mechanisms, and to identify significant new targets for drug discovery. Mass spectrometers must be capable of discriminating molecules of interest while maintaining a high level of robustness, sensitivity, and measurement accuracy. These specifications require a multi-faceted approach combining mass spectrometry and high performance analytical and biochemical separations.

The first study enhanced performance for a common mass spectrometer design. Significant variability in this instrument impaired consistent accurate mass measurements and severely reduced molecular identification. Spectral averaging provided a simple and highly effective strategy to mitigate these instrument limitations.

The second study improved consistency and reproducibility in Affinity Purification Mass Spectrometry (AP-MS), a powerful method to study proteins and protein interaction complexes. An optimized AP-MS method was developed through systematic evaluation of the most common resins and protein tag strategies, demonstrating that AP-MS experiments must consider the effect of solid-phase resin on the entire experimental design.

These improved methods were applied in two biological experiments. The first experiment applied optimized AP-MS to isolate phosphoinositide 3-kinase (PI3K) within pancreatic cancer BxPC-3 cells treated with different tyrosine kinase inhibitors targeting the epidermal growth factor receptor (EGFR). This experiment revealed a signaling processes within the EGFR/PI3K/Akt and ERK pathways as a mechanism of EGFR inhibitor response and provide new insight into the molecular basis tyrosine kinase inhibition and potential resistance mechanisms.

The second experiment utilized liquid chromatography-mass spectrometry (LC-MS) to study dormancy in *Micrococcus luteus (M. luteus),* a close relative of *Mycobacterium tuberculosis (M. tuberculosis).* Eighteen proteins upregulated in dormancy were identified, four of which linked to *M. tuberculosis* latency, and the remaining 14 are novel. These proteins are future targets for molecular and therapeutic studies.

These studies demonstrate improved methods for mass spectrometry-based proteomics and apply these methods to important molecular and drug-discovery systems, thereby progressing the field of systems biology.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1:

Introduction to Mass Spectrometry

1.1 Mass spectrometry

1.1.1 Ion generation: electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI)

Electropray ionization is a “soft” ionization process of a liquid being dispersed into sprayed droplets under the influence of an applied electric field. Soft ionization techniques have the advantage that they do not transfer large amounts of energy to the molecules to be analyzed. Molecular fragmentation through the ionization process is much more controlled and limited compared to more traditional electron impact and chemical ionization methods. The liquid sample is introduced into the high voltage region of the source through a capillary held at high voltage potential compared to a counter electrode (2-10kV). This induces high levels of charge within the liquid. The introduction of the liquid into this high potential field generates an equilibrium between the surface tension (holding the droplets together) and the electrostatic Coulombic attraction for the charged liquid to the counter electrode. This equilibrium is broken at a certain voltage (Taylor Cone Voltage) and, at that voltage and above the liquid it emitted from the capillary in a fine dispersion of micro-droplets (hundreds of nm or smaller). The point at which this applied electrical field counterbalances the surface tension is defined as the Rayleigh stability limit. This not only happens at the ionization source, but at the droplet level. Once these droplets are formed, solvent evaporation becomes a critical factor. Solvent molecules tend to be uncharged, so ionized analyte molecules separate to the surface of the droplets. As evaporation continues, the field density at the surface of the droplet increases to the Rayleigh limit again, and thus the droplets can break apart into smaller droplets and ultimately eject charged analyte ions into the gas phase, a process

often referred to as Coulombic fission. Most of the larger analyte molecules (above 500-1000 Da) appear to be generated by this evaporation-type mechanism, but multiple mechanisms have been proposed to be important (Ion Evaporation and Charge Residue Models). Ionization of analyte molecules in either positive or negative charge is accomplished using high positive or negative electric field potential relative to the counter-electrode. In proteomics experiments, peptides are usually observed more effectively in positive electrospray with some notable exceptions (for example, highly acidic phosphorylated and sulfated peptides). For this reason, the majority of proteomics experiments my mass spectrometry use highly acidic (pH 2.5) liquid phase for introduction of analytes into the ESI source. This is also extremely convenient for high pressure liquid chromatography (HPLC) and ultra-high pressure liquid chromatography (UPLC) methods, both of which use acidic pH liquid phases for high resolution separation of peptides. Typical ESI sources have liquid flow rates in the range of 100s to 1000 microliters/minute or higher, requiring physical assistance for generating solvent droplets suitable for ionization of analytes. This is often accomplished using nitrogen or argon nebulizer gases heated to 200 ºC, much like a paint sprayer. This process generates only a small percentage of the total solvent mass into droplets suitable for ionization (Yamashita and Fenn, 1984).

Nano-Electrospray Ionization sources use a much smaller diameter capillary for introduction of sample in much lower quantity than ESI sources (typically in 50 nanoliter-50 microliter/minute flow rates). The advantages of a Nano-ESI source are many including (1) much higher efficiency of ionization because most of the emitted liquid is dispersed into microdroplets suitable for ionization, (2) low flow rates consume much less material, and (3) the small scale of sample introduction increases relative concentration of analyte molecules in the gas phase. The combination of these advantages in real-world terms enhance sensitivity of Nano-ESI sources by 3-5 orders of magnitude compared to standard ESI sources.

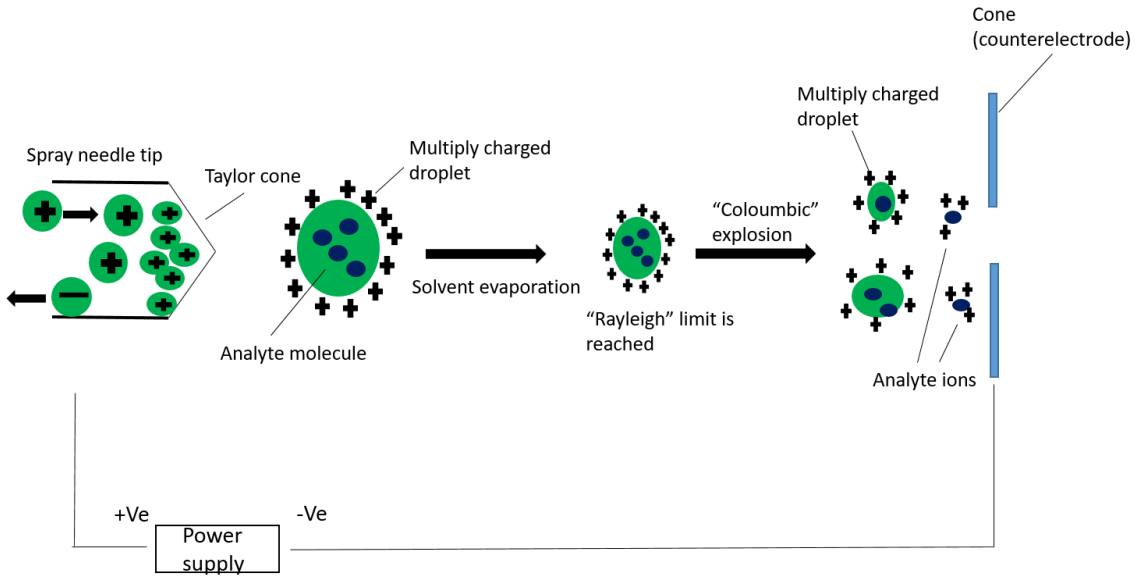


Figure 1.1 Schematic of electrospray ionization. A continual stream of the sample is passed through a capillary tube. Then a fine mist of the highly charged droplets is created, which has the same polarity as the capillary voltage. A nebulizing gas, such as nitrogen, surrounds the sample solution, creating a high flow rate. The charged droplets are generated at the capillary tip and pass down pressure and a potential gradient towards the mass analyzer. Finally, the droplet and its electric field strength reach a critical point, known as the Rayleigh limit, The Rayleigh limit is the maximum number of charges that a droplet can have before exploding into smaller droplets. This concept is known as "Coloumb fission" and occurs when the surface tension of the droplet can no longer sustain the Coloumb force of repulsion. Adapted from "Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte," by Banerjee, S and Mazumdar, S, 2012, *International journal of analytical chemistry,* 2012, 1-40. Copyright [2012] by Banerjee, Shibdas & Mazumdar, Shyamalava.

Matrix-Assisted Laser Desorption Ionization (MALDI) is another "soft" ionization technique wherein a MALDI source, the analyte must be mixed with a matrix material (Lewis *et al.,* 2000). A suitable matrix has a crystalline solid form, a moderate molecular weight to balance ease of vaporization without evaporating during sample preparation or within the high vacuum region of the MALDI source. Additionally, the matrix is usually acidic and acts as a proton source for the

ionization of the analyte. The matrix is usually made in a mixture of highly purified water and a solvent such as acetonitrile, ethanol, or trifluoroacetic acid. This allows solvation of hydrophilic and hydrophobic molecules. Many matrices are aromatic acids, with strong optical absorption in the UV or IR range to rapidly and efficiently absorb laser radiation. The acidic functional group also enhances transfer of charge to the analyte. The most common MALDI matrices that satisfy these properties are 3,5-dimethoxy-4-hydroxycinnamic acid, α-cyano-4-hydroxycinnamic acid, or 2,5-dihydroxybenzoic acid.

The analyte and matrix are mixed together on a metal plate and allowed to crystallize (Paulo *et al.,* 2012). Desorption occurs by absorption of ultraviolet (UV) or infrared (IR) laser light focused on the sample/crystalline matrix mixture on the plate, leading to the ablation of the upper layer (~1 μm) of the matrix mix (Zenobi and Knochenmuss, 1998). Although not as common, IR lasers have a softer ionization and allow for greater removal, less low mass contaminants, and compatibility with other matrix-free desorption methods. The ablation event produces many species including as neutral and ionized matrix molecules, matrix clusters, nanodroplets and analyte ions. MALDI ion sources produce predominantly singly charged ions, but multiply charged ions can be observed depending on specific matrix, the intensity of the laser, and the applied accelerating voltage. The efficient and direct energy transferred from the laser allows for a low concentration sample, usually from femtomoles to picomoles. The exact mechanism of MALDI ionization is not known but thought to be caused by laser excitation and ablation of the matrix. The advantages of MALDI include analysis of proteins up to 300,000 Daltons, sensitivity on the order of femtomole to picomole, soft ionization with little to no fragmentation, tolerance of salt contaminants within the millimolar concentration, and suitability for the analysis of complex mixtures (Geer *et al.,* 2004).

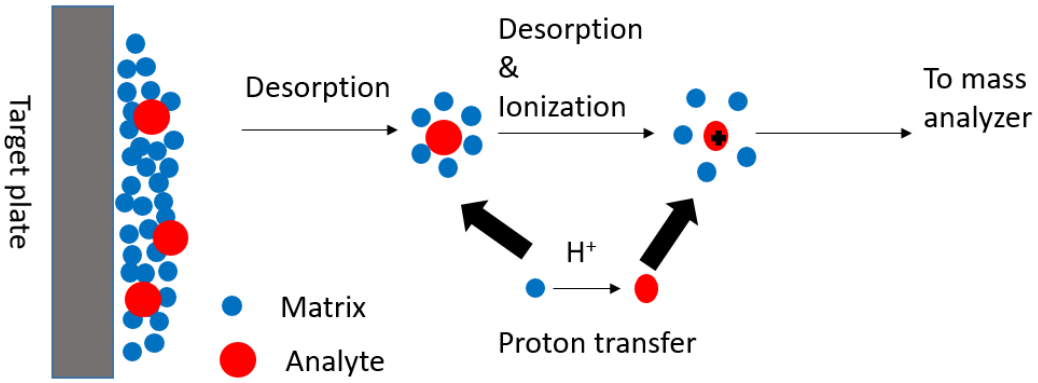


Figure 1.2 Schematic of MALDI ionization. The analyte is ionized and co-crystallized with the matrix. Ions observed are from an added proton or a removed proton. The exact mechanism of MALDI ionization is not known but thought to be caused by laser excitation and ablation of the matrix. Adapted from *AACLifeSci Course Companion Manual Advanced Analytical Chemistry for Life Sciences* (p.40) by P. Domingues, A. Garcia, E. Skrzydlewska, W. Łuczaj, & A. Gęgotek, & K. Bielawska, C. Barbas, D. Dudzik, F. Rey-Stolle, F. Rupérez, E. Maciel, E. Alves, R.M. Domingues, T. Melo, R. Ferreira, 2018, Copyright [2008-2020] by ResearchGate GmbH.

1.1.2 Ion separation: quadrupole, time-of-flight, and ion trap mass filters

As ions travel through a magnetic or electrical field, their movement is perturbed in direct relation to their mass-to-charge ratio, which underlines the main concept of ion separation in mass spectrometry. Therefore, a mass spectrometer can be considered a controlled electric or magnetic field generator that separates ions of different masses by altering the flight of ions in time or position. Because the mass spectrometer modification of electric or magnetic fields is carefully controlled, measurement of the flight characteristics of ions can be used to accurately calculate their mass-to-charge ratio. Many different mass spectrometer designs are available, with the designs relevant to these studies considered here: Quadrupole, Time-of-Flight, and Ion Trap instruments.

Quadrupole mass filter: A quadrupole consists of four parallel metal rods that are equidistant from each other. Often considered rods, the ideal cross-sectional area (from a plane cut through the quadrupole in the x-y plane) is actually a hyperbola. Each pair of opposite rods are electrically linked. An equal but opposite direct current (DC) voltage superimposed with a radio frequency (RF) alternating current (AC) is applied to the diagonal rods. The resulting electrical field causes the ions to move forward in the z-direction (Figure 1.2) with oscillatory motion in the x-y plane. The amplitude of oscillation is proportional to the m/z ratio and can be manipulated by the DC and RF voltages to select ions for analysis. Appropriate application of DC and RF voltages force undesirable ions into large and unstable trajectories that hit the rods, become neutralized, and fail to reach the detector (Fjeldsted, 2003).

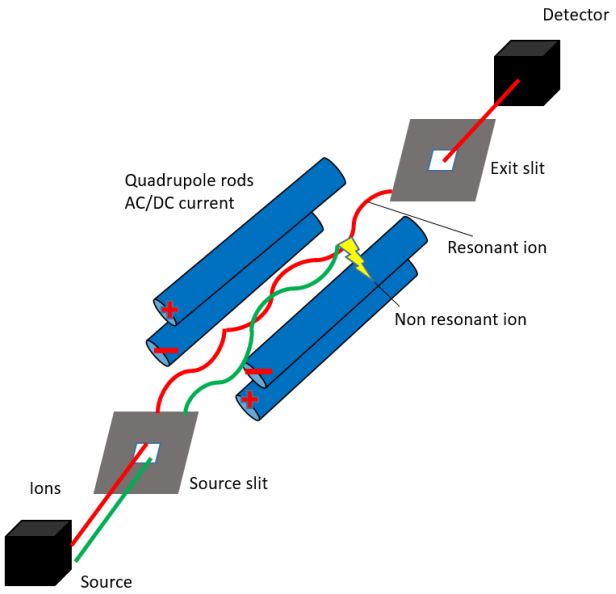


Figure 1.3 Schematic of quadrupole ion separator. A quadrupole consists of four parallel metal rods that are kept equidistant from each other. Each pair of opposite rods are connected electrically. The undesirable ions are large and unstable, thus they hit the rods, become neutralized and fail to reach the detector. Adapted from "Recent methodologies for studying the soil organic matter," by G. Santoiemma, 2017, *Applied Soil Ecology*,*123,*546-550. Copyright [2017] by Elsevier B.V.

Time-of-flight mass filter: The Time-of-Flight (TOF) is another ion separation mass filter. The TOF is perhaps the simplest to understand of mass filters. A TOF analyzer is under high vacuum ($<10^{-7}$ torr), but with no electrical field in the "field-free" region of the flight tube. The ions are accelerated in the ion source region of the mass spectrometer, then drift through this field-free flight tube with the kinetic energy derived from the potential energy of the accelerating electric field. While not completely accurate, the standard kinetic energy equation (Equation 1.4) provides an idea of how this type of mass filter functions.

$$eV=KE=mv^2/2$$

Equation 1.4 Mathematical calculation of kinetic energy. Used in time-of-flight ion separation, where eV is the potential energy of the electric field, KE is the kinetic energy, m is the m/z ratio, and v is the velocity.

If all ions obtain the same kinetic energy from the accelerating electric field, the ions of lesser m/z value will have a greater velocity than the ions of greater mass. Therefore, the ions separate as they travel through the field-free region of the analyzer. A detector at the end of the analyzer measures the arrival times of the ions. This is the basis for a simple linear Time-of-Flight mass filter. In reality, the acceleration voltage does not impart exactly the same kinetic energy to all ions, even ions of the same mass. For this reason, not all ions of the same mass reach their ideal velocities and hence do not arrive at the detector at the same time. To correct for this inhomogeneity in flight times, an ingenious method of "reflecting" ions back towards the ion source was developed (Schermann, J-P. 2008). This reflective TOF analyzer or "reflectron" consists of ring electrodes with high voltage that repulse the ions back through the flight chamber at a slightly

displaced angle. Ions of different kinetic energy penetrate the reflectron rings at different depths before they get repelled from the reflectron in the opposite direction. Faster ions carrying more kinetic energy travel a longer flight path than the slower ions, interacting for longer times with the reflectron so the detector will now receive ions of the same masses at essentially the same time. This modification has significantly increased both the sensitivity and resolution of the TOF mass filter (Lewis *et al.,* 2000).

Ion trap mass filter: The Ion Trap mass filter is yet another method of ion separation. The typical 3-Dimensional (3-D) ion trap consists of entrance end cap and an exit end cap electrode, and a central ring electrode. The cross-sectional profile of the ion trop is hyperbolic and corresponds precisely to the ideal cross sectional profile of a quadrupole. These electrodes form a cavity in which the ions are "trapped" and stored. Both end cap electrodes have small holes through which ions can travel. The ring electrode is located halfway between the two endcap electrodes. Various voltages are applied to the electrodes. Inside the trap, ions oscillate in a highly complex figure-8-shaped trajectory. The ring electrode RF potential, an AC potential of constant frequency, and varied amplitude are applied to the ring electrode to produce a three-dimensional quadrupolar potential field of the trap cavity. Ions are ejected in order of increasing m/z values based on the applied radiofrequency potentials (Scigelova and Makarov, 2006). Importantly, the essentially identical cross-sectional profile for the quadrupole and 3-Dimensional ion trap mass filters has provided new generations of ion trap mass filters based on quadrupole systems. These "linear" ion trap mass filters have higher sensitivity, faster scan rates, enhanced mass ranges, and other improved performance characteristics compared to the standard 3-D ion trap designs.

Q-TOF: The micrOTOF-Q features a quadrupole mass filter and a quadrupole collision cell for an accumulation of parent and fragment ions prior to mass analysis. The entire range of fragment ions is available at increased sensitivity for the time-of-flight mass analysis. The micrOTOF-Q reaches a mass accuracy of better than 3 ppm by time-of-flight systems with

innovative detection technologies. The instrument has extreme stability over a wide dynamic range without tedious recalibration routines. The mass position is kept stable for hours during temperature changes and over a wide dynamic range with significant variations in sample concentration. Its analytical performance consists of a mass range 20 – 40,000 m/z, mass accuracy in MS and MS/MS 3 ppm RMS error (internal), 5 ppm RMS error (external), mass resolution in MS and MS/MS 15,000 (FWHM), temperature compensated, and an acquisition rate (2 GHz sampling rate) 20 Hz (profile and peak detected spectra to disk) (Bristow *et al.,* 2008).

MALDI-TOF-TOF: Matrix-Assisted Laser Desorption/Ionization-Time-of-Flight-Time-of-Flight (MALDI-TOF-TOF) is commonly used for quality control of oligonucleotides. This spectrometer uses pulses of laser light to vaporize the sample and matrix to perform desorption. Some molecules become ionized through protonation. Ions are accelerated by an electrostatic field in the mass analyzer to a common kinetic energy. With the same kinetic energy, lighter ions travel faster and heavier ones travel more slowly. The mass spectrum is recorded in ion flux versus time (Lay, 2001).

1.1.3 Ion detection

After ions are generated by an ion source and separated in a mass filter, there is still the requirement of detection. The most common detector for mass spectrometry instruments is the electron multiplier. Two basic forms of electron multipliers are commonly used in mass spectrometry: the discrete-dynode electron multiplier (ETP) and the continuous-dynode electron multiplier (often referred to as a channel electron multiplier or CEM). Most standard mass spectrometer detectors utilize a discrete-dynode electron multiplier that consists of between 12 and 24 dynodes with progressive voltages applied to each dynode. When ions impact the target dynode of the electron multiplier, secondary electrons are released from the surface in a physical

process called secondary electron emission. The number of secondary electrons released depends on the type of incident primary particle, its energy, and characteristics and voltages on the dynode. The electrons are then transferred and impact the next and subsequent dynodes, generating a cascade of increasing numbers of electrons and amplifying the electrical signal. Amplification of detected ions, defined as the gain of the amplifier, range from 10^6 to 10^8 depending on specifics of the electron multiplier. In GC/MS applications, for example, the electron multiplier is typically operated in analog mode with a gain of around 10^5. For a new electron multiplier used in standard ion trap, TOF, or equivalent instruments, this gain is typically achieved with an applied high voltage of ~1400 volts (Yates, 2000). Importantly, the efficiency of the electron multiplier represents a potentially limiting factor on the overall system sensitivity and influence on the overall performance of the mass spectrometer.

1.1.4 Liquid chromatography-mass spectrometry

Liquid chromatography (LC) is a widely used method to separate components of a complex sample prior to analysis by other technologies including mass spectrometry. LC depends on the fundamental principle that compounds will partition between mobile phase (usually a liquid) and a stationary phase (usually a solid-phase material) based on the specific physical and chemical properties of the compound. This partitioning enables highly effective separation of diverse compounds based on choice of mobile and stationary phase solvents/materials. (Bruins *et al.,* 1987). LC has extraordinary advantages when linked to mass spectrometry (LC-MS). LC-MS offers broad separation capabilities for larger and non-volatile molecules such as proteins and complex peptides. Diverse column chemistries, including reverse-phase, normal-phase, ion-exchange, chiral, and size-exclusion, can be applied across wide ranges of biological samples. With higher resolution LC approaches (described below), separation of structural and chiral

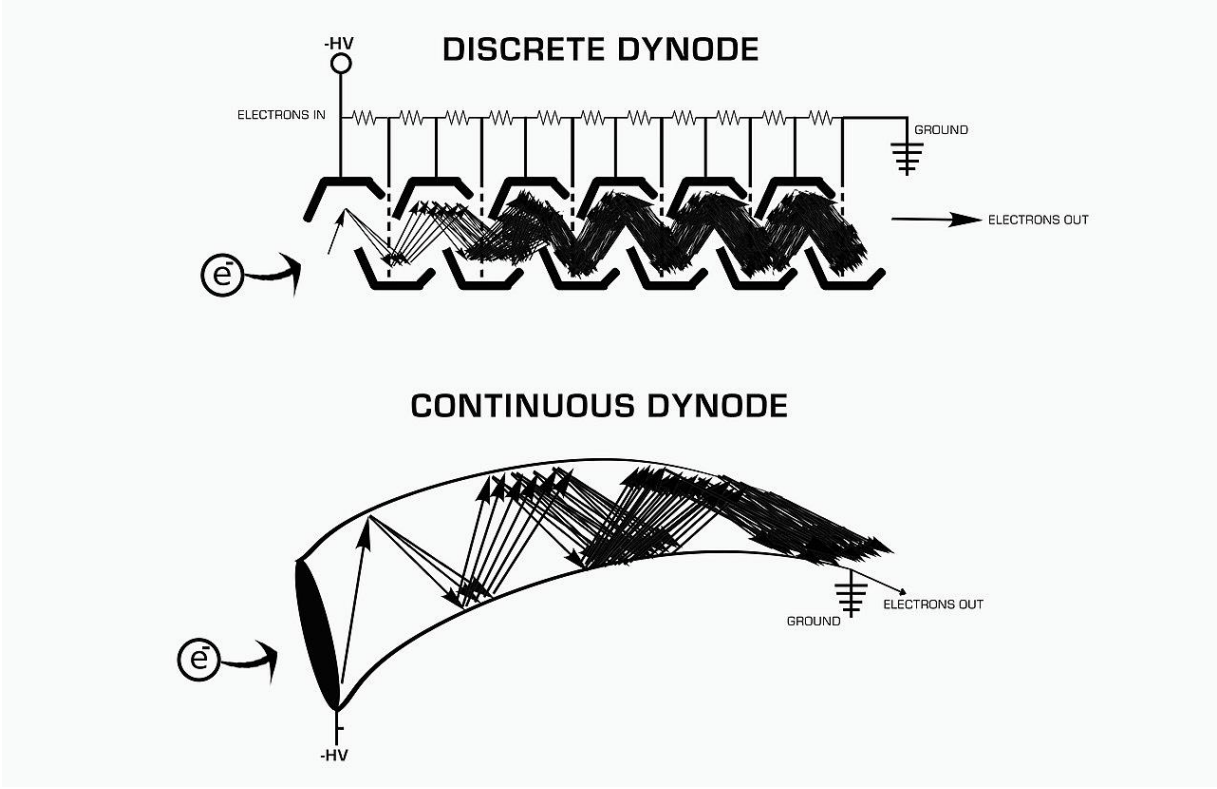


Figure 1.5 Electron multipliers. The discrete-dynode electron multiplier (DEM) and the continuous-dynode electron multiplier or a channel electron multiplier (CEM). Reprinted from "Photon and electron action spectroscopy of trapped biomolecular ions - From isolated to nanosolvated species," by M. Ranković, 2016, https://www.researchgate.net/figure/Schematic-representation-of-the-electron-multiplier-detector-with-conversion-dynodes-in_fig11_307606559. Copyright [2008-2020] by ResearchGate GmbH.

isomers can be achieved. This resolves a fundamental limitation for mass spectrometry, these isomers have the same mass and will not be differentiated (i.e., resolved) by a mass spectrometer. In fact, due to its superior resolving power and broad mass range, LC has largely replaced gel electrophoresis for molecular separation. Finally, the separation of compounds by LC prior to

mass spectrometry helps reduce ion suppression, which occurs when molecules interact with one another and impede the process of complete ionization (Roux, 2011).

High-performance liquid chromatography (HPLC) has improved upon and largely replaced LC for interfacing with mass spectrometry. HPLC operates at higher pressures ranging from 50-350 bar. In contrast, LC relies on gravity or lower-pressure pump systems for the passage of the mobile phase through the column (Xiao and Oefner, 2001). While standard ESI sources can generally handle flow rates up to 1 mL/min, lower flow rates in the range of 0.05 to 0.2 mL/min result in improved sensitivity. Therefore, lower flowrate HPLC systems using columns with 2.1- or 1.0-mm inner diameter (ID) are well suited to direct coupling with ESI sources found on many commercial mass spectrometers. Conversion of larger scale HPLC separations (4.6mm ID columns) to smaller 2.1-1.0 mm ID columns is predictable, and the resulting decrease in mobile phase consumption is a useful side benefit. Miniaturization to even smaller capillary columns (250-micron ID or smaller) can greatly enhance sensitivity and resolution on HPLC systems capable of sub µL/min flow rates. These separations are generally more demanding and require different ion sources and specialized HPLC pumps and column connections. However, these demanding separations are exceptionally valuable in the proteomics area where high sensitivity and resolution are required to identify as many components as possible. However, they are probably less suited for routine clinical laboratory work (Vestal, 1984).

Typical solvents used in reverse and normal phase HPLC (e.g., water, acetonitrile, methanol, ethanol, chloroform) are compatible with ESI. It should be noted that a grade of solvent (including water) that is suitable for a conventional LC separation may not always be suitable for an LC-MS based separation. The higher organic content of the mobile phase can result in improvements in ionization efficiency in ESI, and this may affect the choice of separation mode. For example, polar molecules are poorly retained on reverse phase columns and elute at the beginning of the chromatogram with low organic content. This will result in relatively poor ionization efficiency, and

polar interferences are more likely to co-elute in this region causing ion suppression. One way this can be overcome is to use hydrophilic interaction chromatography (HILIC) in which polar analytes are retained and elute with higher organic content of the mobile phase (Vestal, 1984).

1.1.5 Tandem mass spectrometry

Tandem mass spectrometry, or mass spectrometry/mass spectrometry (MS/MS), experiments are conducted by combining two mass analyzers in a single instrument. For example, a common MS/MS-capable instrument couples three quadrupole mass analyzers (i.e. “triple quadrupoles”) combined in series. However, only two of the quadrupoles are involved in mass analysis: the third (middle) quadrupole transmits all ions and is used to “activate” or transfer energy to fragment ions of interest selected in the first quadrupole. MS/MS has been described as ‘taking the mass spectrum of an ion in a mass spectrum.” The principle of MS/MS is that a sample is ionized and mass-analyzed in the first mass analyzer. A particular ion with specific indicative m/z value of interest is selected from the mass spectrum and is directed into a collision cell that contains a neutral gas. The ion is vibrationally excited by radio-frequency resonance or amplitude energy and collided with the target gas, a process known as collision-induced dissociation. This generates fragment ions that are separated and recorded by the second mass analyzer (Angel *et al.,* 2012).

Tandem mass spectrometry (MS/MS or MS^n) can take several forms. The first analyzer can be set to allow ions of only a specific *m/z* value to pass into the collision cell. The second analyzer is scanned to obtain a product-ion mass spectrum exhibiting ions of all *m/z* values resulting from CAD. This is called a product-ion analysis, and the results are a product-ion mass spectrum. The second *m/z* analyzer can be set to allow ions of only a single *m/z* value to pass to the detector. The first analyzer is scanned. This results in a signal for any ion of any *m/z* value that produces

product ions of a specific *m/z* value. This is called a precursor-ion analysis, and the results are a precursor-ion mass spectrum for a specific product ion. The process of fragmenting precursor ions is called collision-induced dissociation (CID). The two *m/z* analyzers can be scanned simultaneously with the second *m/z* analyzer having a lower starting *m/z* than the first. This will result in a signal for ions of any *m/z* value passing the first analyzer that undergo a loss of a specific mass (a neutral loss) in the collision cell. This is called a common-neutral-loss analysis, and the results are called a neutral-loss spectrum of a specified offset. The most widely used form of MS/MS is in a process called selected reaction monitoring (SRM). SRM is a process where an ion of a lesser *m/z* value is allowed to pass the second *m/z* analyzer when ions of a specified higher *m/z* value pass the first analyzer. This process is analogous to selected ion monitoring (SIM) however, because it involves the transition from an ion formed by an analyte to a fragment of that ion, it provides for a higher degree of selectivity. SIM is the practice of monitoring and recording ion currents at one or more selected ion m/z values with time, rather than recording full mass spectra, as sample is introduced into the ion source. Because the detector is integrating signal for a longer time at the relevant ion, limits of detection can be lowered, albeit at a cost of susceptibility of the experiment to unexpected interferences (Nesvizhskii *et al.,* 2007).

1.2 Proteomics

1.2.1 Mass spectra

A mass spectrum will usually be presented as a vertical bar graph, in which each bar represents an ion having a specific mass-to-charge ratio (m/z) and the length of the bar indicates the relative abundance of the ion. The most intense ion is assigned an abundance of 100, and it is referred to as the base peak. Most of the ions formed in a mass spectrometer have a single charge, so the m/z value is equivalent to mass itself. Modern mass spectrometers easily distinguish (resolve)

ions differing by only a single atomic mass unit (amu), and thus provide completely accurate values for the molecular mass of a compound. The highest-mass ion in a spectrum is normally considered to be the molecular ion, and lower-mass ions are fragments from the molecular ion, assuming the sample is a single pure compound. The following diagram displays the mass spectra of three simple gaseous compounds, carbon dioxide (CO_2), propane (C_3H_8) and cyclopropane (C_3H_6). The molecules of these compounds are similar in size. Carbon dioxide and propane both have a nominal mass of 44 amu, while cyclopropane has a mass of 42 amu. The molecular ion is the strongest ion in the spectra of carbon dioxide and cyclopropane, and it is moderately strong in propane. The unit mass resolution is readily apparent in these spectra (note the separation of ions having m/z=39, 40, 41 and 42 in the cyclopropane spectrum). Even though these compounds are very similar in size, it is a simple matter to identify them from their individual mass spectra. Even with simple compounds like these, it should be noted that it is rarely possible to explain the origin of all the fragment ions in a spectrum. Also, the structure of most fragment ions is seldom known with certainty. A similar bond cleavage in cyclopropane does not give two fragments, so the molecular ion is stronger than in propane, and is in fact responsible for the the base peak. Loss of a hydrogen atom, either before or after ring opening, produces the stable allyl cation (m/z=41). The third strongest ion in the spectrum has m/z=39 (C_3H_3). Its structure is uncertain, but two possibilities are shown in the diagram. The small m/z=39 ion in propane and the absence of an m/z=29 ion in cyclopropane are particularly significant in distinguishing these hydrocarbons (Kapp *et al.,* 2003).

Most stable organic compounds have an even number of total electrons, reflecting the fact that electrons occupy atomic and molecular orbitals in pairs. When a single electron is removed from a molecule to give an ion, the total electron count becomes an odd number, and we refer to such ions as radical cations. The molecular ion in a mass spectrum is always a radical cation, but the

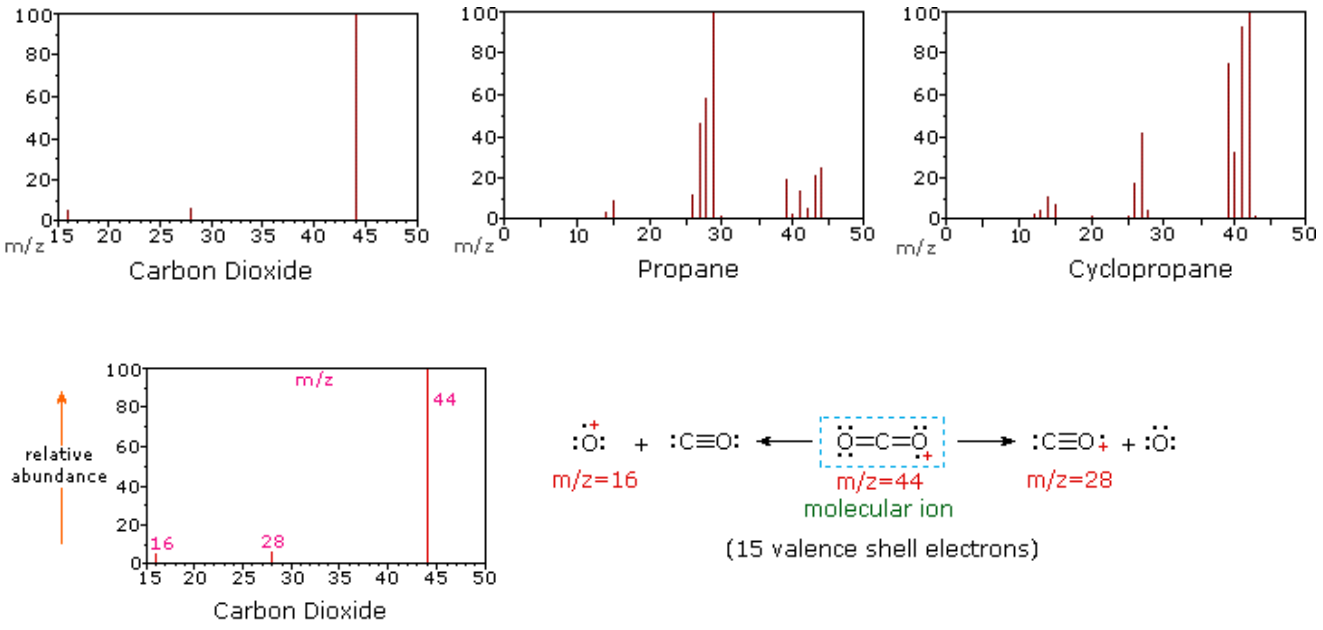


Figure 1.6. Mass spectra of the gaseous compounds, carbon dioxide, propane and cyclopropane. The molecules of these compounds are similar in size. Carbon dioxide and propane both have a nominal mass of 44 amu. Cyclopropane has a mass of 42 amu. Reprinted from "Mass Spectrometry" by W. Reusch, https://www2.chemistry.msu.edu/faculty/reusch/VirTxtJml/Spectrpy/MassSpec/masspec1.htm. Copyright [2013] by Michigan State University.

fragment ions may either be even-electron cations or odd-electron radical cations, depending on the neutral fragment lost. The simplest and most common fragmentations are bond cleavages producing a neutral radical (odd number of electrons) and a cation having an even number of electrons. A less common fragmentation, in which an even-electron neutral fragment is lost, produces an odd-electron radical cation fragment ion. Fragment ions themselves may fragment further. As a rule, odd-electron ions may fragment either to odd or even-electron ions, but even-electron ions fragment only to other

even-electron ions. The masses of molecular and fragment ions also reflect the electron count, depending on the number of nitrogen atoms in the species.

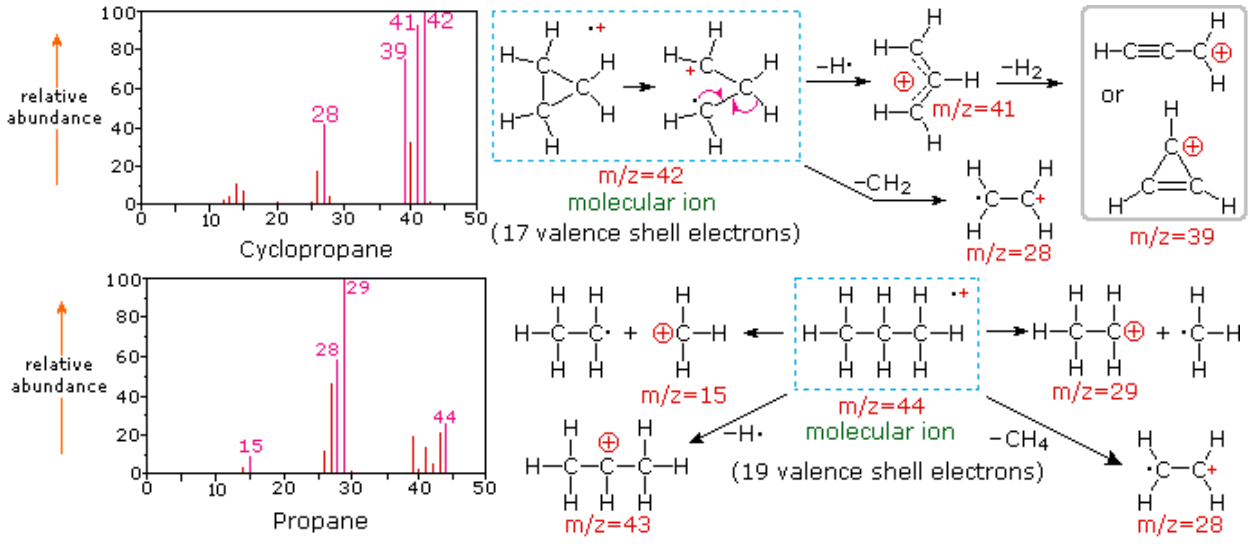
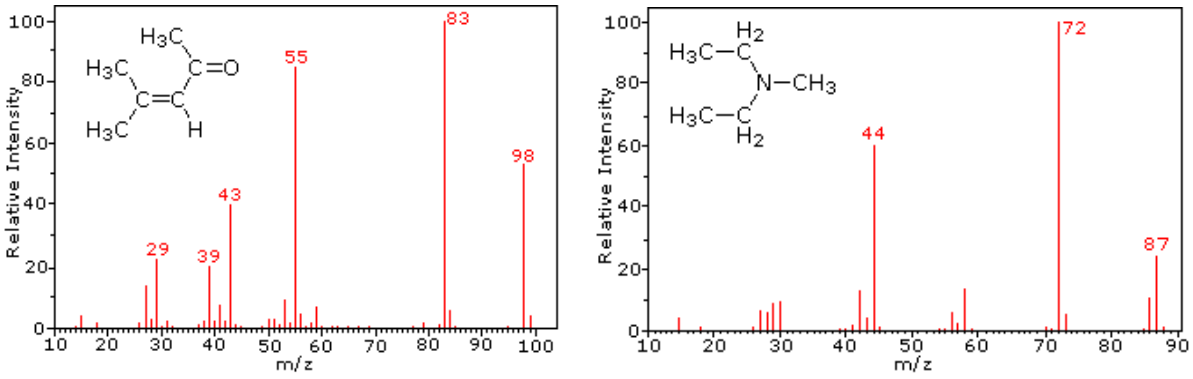


Figure 1.7 Mass spectra and fragmentation of propane and cyclopropane. The molecular ion of propane also has m/z=44, but it is not the most abundant ion in the spectrum. Cleavage of a carbon-carbon bond gives methyl and ethyl fragments, one of which is a carbocation and the other a radical. Both distributions are observed, but the larger ethyl cation (m/z=29) is the most abundant, possibly because its size affords greater charge dispersal. A similar bond cleavage in cyclopropane does not give two fragments, so the molecular ion is stronger than in propane, and is in fact responsible for the base peak. Reprinted from “Mass Spectrometry” by W. Reusch, https://www2.chemistry.msu.edu/faculty/reusch/VirtTxtJml/Spectrpy/MassSpec/massspec1. htm. Copyright [2013] by Michigan State University.

This distinction is illustrated nicely by the following two examples. The unsaturated ketone, 4-methyl-3-pentene-2-one, on the left has no nitrogen so the mass of the molecular ion (m/z = 98) is an even number. Most of the fragment ions have odd-numbered masses, and therefore are even-electron cations. N,N-diethylmethylamine, on the other hand, has one nitrogen and its molecular mass (m/z = 87) is an odd number. A majority of the fragment ions have even-

numbered masses (ions at m/z = 30, 42, 56 & 58 are not labeled), and are even-electron nitrogen cations. The weak even -electron ions at m/z=15 and 29 are due to methyl and ethyl cations (no nitrogen atoms) (Hsu and Turk, 2009).

Fragmentation: The fragmentation of molecular ions into an assortment of fragment ions is a mixed blessing. The nature of the fragments often provides a clue to the molecular structure, but if the molecular ion has a lifetime of less than a few microseconds it will not survive long enough to be observed. Without a molecular ion peak as a reference, the difficulty of interpreting a mass spectrum increases markedly. Fortunately, most organic compounds give mass spectra that include a molecular ion, and those that do not often respond successfully to the use of milder ionization conditions. (Hiroshi and Michihiko, 1873). Among simple organic compounds, the most stable molecular ions are those from aromatic rings, other conjugated pi-electron systems and cycloalkanes.



4-methyl-3-pentene-2-one **N,N-diethylmethylamine**

Figure 1.8 Mass spectra of 4-methyl-3-pentene-2-one and N,N-diethylmethylamine. The unsaturated ketone, 4-methyl-3-pentene-2-one, on the left has no nitrogen so the mass of the molecular ion (m/z = 98) is an even number. N,N-Diethylmethylamine has one nitrogen and its molecular mass (m/z = 87) is an odd number. Reprinted from "Mass Spectrometry" by W. Reusch, https://www2.chemistry.msu.edu/faculty/reusch/VirtTxtJml/Spectrpy/MassSpec/massspec1.htm.

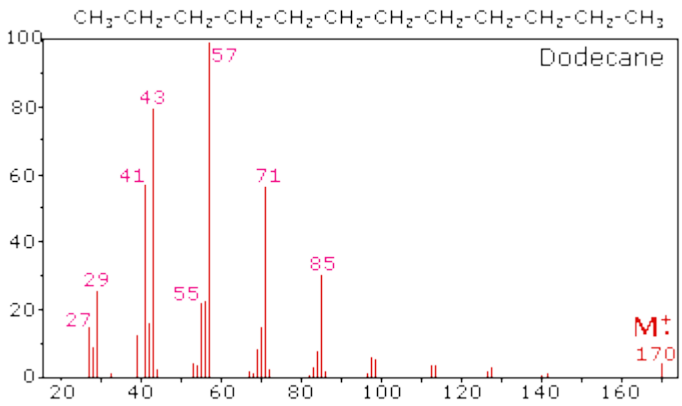


Figure 1.9 The mass spectrum of dodecane. This diagram illustrates the behavior of an unbranched alkane. Since there are no heteroatoms in this molecule, there are no non-bonding valence shell electrons. Consequently, the radical cation character of the molecular ion (m/z = 170) is delocalized over all the covalent bonds. Reprinted from "Mass Spectrometry" by W. Reusch, https://www2.chemistry.msu.edu/faculty/reusch/VirtTxtJml/Spectrpy/MassSpec/mass pec/massspec1.htm. Copyright [2013] by Michigan State University.

Alcohols, ethers and highly branched alkanes generally show the greatest tendency toward fragmentation. The mass spectrum of dodecane illustrates the behavior of an unbranched alkane. Since there are no heteroatoms in this molecule, there are no non-bonding valence shell electrons. Consequently, the radical cation character of the molecular ion (m/z = 170) is delocalized over all the covalent bonds. Fragmentation of C-C bonds occurs because they are usually weaker than C-H bonds, and this produces a mixture of alkyl radicals and alkyl carbocations. The positive charge commonly resides on the smaller fragment, so we see a homologous series of hexyl (m/z = 85), pentyl (m/z = 71), butyl (m/z = 57), propyl (m/z = 43), ethyl (m/z = 29) and methyl (m/z = 15) cations. These are accompanied by a set of corresponding alkenyl carbocations (e.g. m/z = 55, 41 &27) formed by loss of 2 H. All of the significant fragment ions in this spectrum are even-electron ions. In most alkane spectra the propyl and butyl ions are the most abundant. The presence of a functional group, particularly one having a heteroatom Y

with non-bonding valence electrons (Y = N, O, S, X etc.), can dramatically alter the fragmentation pattern of a compound. This influence is thought to occur because of a "localization" of the radical cation component of the molecular ion on the heteroatom. After all, it is easier to remove (ionize) a non-bonding electron than one that is part of a covalent bond. By localizing the reactive moiety, certain fragmentation processes will be favored. These are summarized in the following diagram, where the green shaded box at the top displays examples of such "localized" molecular ions. The first two fragmentation paths lead to even-electron ions, and the elimination (path #3) gives an odd-electron ion. Note the use of different curved arrows to show single electron shifts compared with electron pair shifts (Vaniya and Fiehn, 2015). The charge distributions shown above are common, but for each cleavage process the charge may sometimes be carried by the other (neutral) species, and both fragment ions are observed. Of the three cleavage reactions described here, the alpha-cleavage is generally favored for nitrogen, oxygen and sulfur compounds. Indeed, in the previously displayed spectra of 4-methyl-3-pentene-2-one and N,N-diethylmethylamine the major fragment ions come from alpha-cleavages (Chait and Kitson, 1970). The complexity of fragmentation patterns has led to mass spectra being used as "fingerprints" for identifying compounds. Environmental pollutants, pesticide residues on food, and controlled substance identification are but a few examples of this application. Extremely small samples of an unknown substance (a microgram or less) are sufficient for such analysis. The following mass spectrum of cocaine demonstrates how a forensic laboratory might determine the nature of an unknown street drug. Even though extensive fragmentation has occurred, many of the more abundant ions (identified by pink numbers) can be rationalized by the three mechanisms shown above. The m/z = 42 ion might be any or all of the following: C_3H_6, C_2H_2O or C_2H_4N. A precise assignment could be made from a high-resolution m/z value (Hunt *et al.,* 1986).

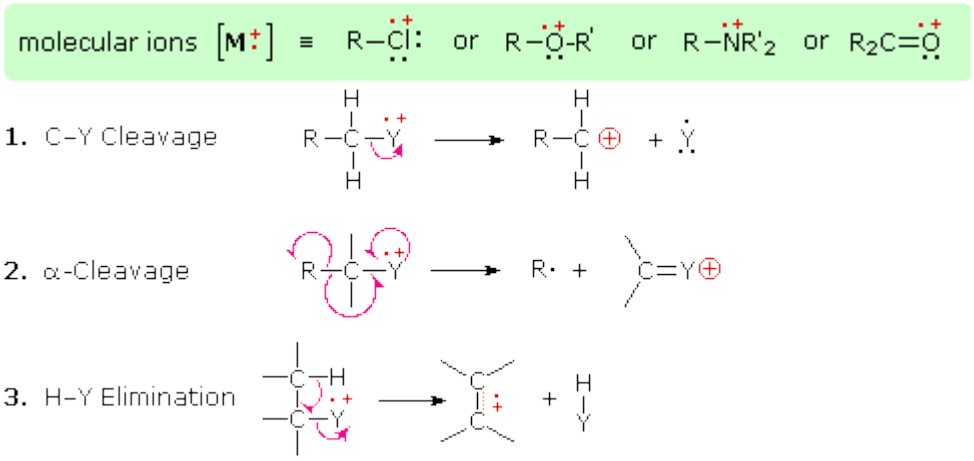


Figure 1.10 Common fragmentation paths. The green shaded box displays examples of "localized" molecular ions. The first two fragmentation paths lead to even-electron ions, and the elimination path (path #3) gives an odd-electron ion. Note the use of different curved arrows to here, the alpha-cleavage is generally favored for nitrogen, oxygen and sulfur compounds. Reprinted from “Mass Spectrometry” by W. Reusch, https://www2.chemistry.msu.edu/faculty/reusch/VirtTxtJml/Spectrpy/MassSpec/masspec1.htm. Copyright [2013] by Michigan State University.

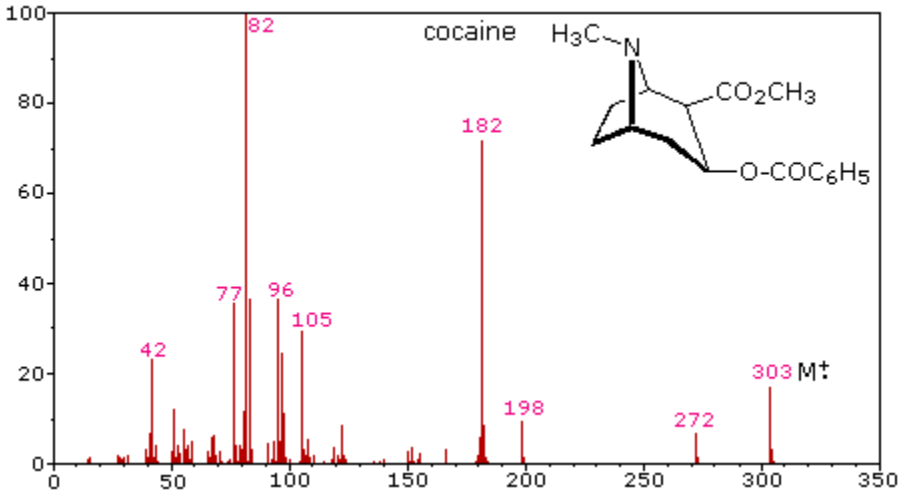


Figure 1.11 Mass spectrum of cocaine. Even though extensive fragmentation has occurred, many of the more abundant ions (pink numbers) can be rationalized by the three mechanisms discussed previously. The m/z = 42 ion might be any or all of the following: C_3H_6, C_2H_2O or C_2H_4N. Reprinted from “Mass Spectrometry” by W. Reusch, https://www2.chemistry.msu.edu/faculty/reusch/VirtTxtJml/Spectrpy/MassSpec/masspec1.htm. Copyright [2013] by Michigan State University.

1.2.2 Computational databases

Because there are large numbers of spectra generated in high throughput proteomics, manual interpretation is impractical, so it is essential that the matches from these algorithms are scored with a threshold or thresholds that allow few false positives. The first computer program to use a database search procedure to identify fragmentation spectra was SEQUEST. It connects each tandem mass spectrum not only to a protein but also to an organism that synthesized that protein. It uses two scoring functions. The first one is used to rapidly determine a few hundred peptide candidates for each spectrum (preliminary score, Sp) while the second uses cross-correlation of the experimental and theoretical spectra (Xcorr). The preliminary score takes into account the sum of matched fragment ion intensities, the number of total and matched fragment ions, and the factor that rewards continuity of matching for each ion series (b and y ions). The final score, Xcorr, is achieved by converting the expected masses of fragment ions predicted for any database peptide into a theoretical spectrum and by computing a cross-correlation between the theoretical spectrum and the experimental spectrum. In addition, SEQUEST exports the normalized difference between the best and the second- best scores (delta Cn), which is useful to determine match uniqueness relative to near misses for a given database (Noble *et al.,* 2012).

The Open Mass Spectrometry Search Algorithm (OMSSA) is an algorithm based on an explicit model of matching statistics similar to the type of statistical model used in BLAST. Probability-based matching allows such thresholds to be set in terms of the number of false positives allowed, leading to the development of several statistical scoring algorithms–in addition to existing probability-based search algorithms. OMSSA is an attempt to create a fast search algorithm whose results are scored using a classical statistical model using assumptions taken directly from the experimental setup and allowing for experimental noise. OMSSA takes experimental ms/ms spectra, filters noise peaks, extracts *m*/*z* values, and then compares these *m*/*z* values to calculated *m*/*z* values derived from peptides produced by an in silico digestion of a protein

sequence library. The theoretical peptides must have a mass within a user specified tolerance of the precursor mass. The resulting search hits are then statistically scored (Geer *et al.,* 2004).

X!Tandem is an open source software that can match tandem mass spectra (usually the experiment) with peptide sequences from a database. This process allows identification of proteins in one or more samples. The X!Tandem search engine calculates a statistical confidence (expectation value) for all of the individual spectrum-to-sequence assignments. Some spectra might map to more than one protein, in this case X!Tandem will pick the best match with the highest confidence score. This database also reassembles all of the peptide assignments in a data set onto the known protein sequences and assign the statistical confidence that this assembly and alignment is non-random. Therefore, separate assembly and statistical analysis software do not need to be used (Geer *et al.,* 2004).

Using a 1% MS^2 FDR are very consistent for each of the algorithms. X!Tandem and SEQUEST FDR yields roughly the same amount of distinct peptides and proteins. However, OMSSA identifies about 5% more distinct peptides and proteins. The protein FDR for each algorithm is consistent as well. In all cases, combining a 1% MS^2 FDR with the requirement for two distinct peptides per protein lowered the protein FDR to an effective rate of 0%. This method enables production of datasets with very high confidence identifications. However, the method also greatly increases the number of false negatives. Other measures of data quality are the average number of MS^2 hits and distinct peptide identifications per protein. Increases in either of these measures is interpreted as an increase in the confidence of a protein identification. In particular, distinct peptide identifications are valued as they lead to increased confidence in and sequence coverage of the predicted protein. By this measure OMSSA and X!Tandem outperform SEQUEST with respect to MS^2 identifications per protein. The number of MS^2 identifications becomes very important when implementing spectral counting-based quantification because the expression levels of each protein are determined by the number of MS^2 identifications. All three algorithms

are much more comparable with regard to the number of distinct peptides identified per protein. Through the use of the target-decoy search approach the relative specificity exhibited by all algorithms is excellent at roughly 99%. The relative sensitivity of SEQUEST is at about 71%, whereas X!Tandem yields 74% and OMSSA 98%. OMSSA delivers the best performance in terms of MS^2 discoveries regardless of MS^2 FDR. Even at very low false discovery rates (0.1%) OMSSA performed with significantly higher sensitivity than SEQUEST, or X!Tandem. Furthermore, when evaluating proteins by the number of distinct peptides mapping to each, OMSSA similarly returned the best performance.

1.3 Quantitative proteomics

1.3.1 Normalized spectral abundance factor

Quantitative information is valuable and important, but MS is not inherently quantitative at measuring abundances from ion intensities. Resolution of this problem include, absolute quantitation approaches based on isotopic labeling, multiple peptide counting, and spectral counting. Absolute quantification is one such strategy to determine numbers of proteins and their modification states. Peptides are synthesized with incorporated stable isotopes as ideal internal standards to mimic native peptides formed by proteolysis. These synthetic peptides can also be prepared with covalent modifications (e.g., phosphorylation, methylation, acetylation, etc.) that are chemically identical to naturally occurring posttranslational modifications. Such internal standard peptides are then used to precisely and quantitatively measure the absolute levels of proteins and post-translationally modified proteins after proteolysis by using a selected reaction monitoring analysis in a tandem mass spectrometer (Gerber et al., 2003).

Multiple peptide counting theorizes that the more abundant a peptide ion is in a mixture, the higher the rate of sampling during the course of a shotgun proteomics experiment. This random sampling

suggests the probability of a protein being identified is directly related to a protein's abundance. More abundant proteins have a higher probability of identification and thus are observed more reproducibly with the acquisition of spectra for more peptides. However, the redundancy of peptide ion acquisition for abundant peptides decreases the efficiency of data acquisition for peptide ions present at low abundance in a predictable manner. This model suggests that the number of spectra acquired from a complex peptide mixture is the most invariant property of the analysis. Spectral counting (i.e. number of tandem mass spectra collected from peptides) displayed near perfect linearity with respect to concentration and therefore is an improvement from peptide counting (Liu *et al.,* 2004).

Normalized spectral abundance factor (NSAF) is a label free spectral counting approach for protein quantification. In this method, the total number of MS/MS spectra matching the peptides from a protein is used to compare the relative abundance of the protein among samples. During protein identification, a larger number of peptides will be observed for large proteins compared to smaller proteins. To account for this, the number of spectral counts is normalized to the protein size or length (Powell *et al*., 2004). The NSAF value of a protein stands for the ratio of total number of spectra observed for that protein compared to the total number of spectra observed from the sample (Equation 1.6). Essentially, NSAF value reflects the percentage of a protein relative to the total amount of all proteins in a sample. With this spectral counting, data from both labeled or label free approaches can be directly compared for quantitative analysis. Along with good statistical analysis, label free spectral counting is a valuable technique for quantitative proteomics analysis allowing one to draw biologically significant inferences (Zybailov *et al*., 2006).

$$(\mathrm{NSAF})_J = \frac{(\mathrm{Sc}/L)_J}{\sum_{i=1}^{N} (\mathrm{Sc}/L)_i}$$

Equation 1.6 Normalized spectral abundance factor. The equation defines the NSAF, where Sc is the spectral count of protein *J, L* is the length of protein *J,* and *N* is the total number of proteins. Reprinted from "Statistical Analysis of Membrane Proteome Expression Changes in Saccharomyces cerevisiae," by B. Zybailov, A. Mosley, M. Sardiu, M. Coleman L. Florens, M. Washburn, 2006, *Journal of Proteome Research*, 5(9), 2339-2347. Copyright [2006] by American Chemical Society.

Chapter 2:

Matrix-Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry

Preface

The work outlined in this chapter has been published in PLOS One (**Mitchell M,** Mali S, King CC, Bark SJ. 2015. Enhancing maldi time-of-flight mass spectrometer performance through spectrum averaging. *PloS One* 10(3)).

2.1 Introduction

Matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometers are highly robust and capable instruments for biomolecular analysis. While the first practical TOF mass spectrometer was developed in the 1950s (Wiley and McLaren, 1955), it was the advent of new ionization techniques compatible with large biological molecules like MALDI (Karas and Hillenkamp, 1988) and electrospray ionization (Fenn *et al.*, 1989) that revolutionized the biological applications of mass spectrometers including TOF instruments. The rapid advancement of biological mass spectrometry is largely attributable to these ionization techniques and their impact on vastly improved mass spectrometer instrumentation performance. The importance of MALDI-TOF mass spectrometry can be appreciated through a brief literature review for diverse fields including identification of bacterial and viral pathogens, clinical pathology, imaging mass spectrometry, biochemistry, and natural products (Yang *et al.*, 2009). Modern MALDI-TOF mass spectrometers use delayed extraction and ion reflector systems to enhance instrument resolution and accuracy, enabling accurate mass measurements of peptides and molecules (Takach *et al.,* 1997). Routine performance specifications for reflector MALDI-TOF instruments often exceed 15,000 for resolution measured by full-width at half-maximum (FWHM) and <5 parts-per-million

(ppm) for accuracy with internal calibration. This performance is sufficient for most biological applications including protein identification by peptide mass fingerprinting, a technique that is highly dependent on high accuracy mass measurements of the component peptides (Zhang and Chait, 2000). The MALDI-TOF/TOF mass spectrometer used for these studies is a high-performance reflectron instrument with specifications at the level described in the previous paragraph in TOF mode. However, we have observed significant variability in replicate mass measurements from under 1 to 20 ppm or greater on this instrument, even in internally calibrated spectra (Tables 2.1-2.2). For example, considering replicate measurements for multiple different peptides, 30–50% of individual measurements exhibited errors above 5 ppm. We have made similar observations for multiple MALDI-TOF-type instruments from different manufacturers, which suggests these factors are intrinsic to this mass spectrometer design. Additionally, mass measurements for multiple different peptides within a single mass spectrum often exhibit uncorrelated errors. Increasing the number of laser shots for each spectrum did not resolve this variability. Unfortunately, there is no a priori method to define the accuracy of an unknown peptide mass measurement and, therefore, these observed mass deviations cannot be compensated or mitigated. We hypothesized that understanding the basis for the observed variability in replicate mass measurements could suggest a method to mitigate these errors and improve the consistency of MALDI-TOF measurements. To this end, trypsin digests of both a standard protein mixture and proteins derived from a biological immunoprecipitation experiment were analyzed using a high-resolution MALDI-TOF/TOF mass spectrometer in TOF mode with 5-point internal calibration. The same sample was also analyzed using a quadrupole-time-of-flight (Q-TOF) mass spectrometer coupled to an HPLC system. Direct comparison of these different mass spectrometry platforms enhanced the identification of peptides and provided high confidence for evaluating the accuracy and performance of the MALDI-TOF mass spectrometer (Mitchell *et al.,* 2015).

These data demonstrate significant variability in observed peptide masses and the discontinuous nature of the analog-to-digital (AD) detector system in the MALDI-TOF mass spectrometer. When restarting acquisition, the AD detector system resets the position of the bins within the electronic error of the system, thus shifting the data by a small amount (usually less than the width of a single bin). This error impacts both flight time measurement and calibration function, both of which require interpolation from the discontinuous data observed in the mass spectrum. The data suggest this small error is still significant and contributes to the observed variability in the MALDI-TOF data. While the mechanisms underlying the variability observed in the MALDI-TOF data appear complex, the data indicate the method to resolve this variability is simple. The bin repositioning for each independent spectrum and calibration follows a normal Gaussian distribution. Therefore, mass spectral measurements can be analyzed by averaging populations of individual spectra and using simple descriptive statistics appropriate for normally distributed data. This simplified statistical approach provides enhanced instrument performance more consistent with the accuracy specifications of a high-resolution MALDI-TOF mass spectrometer. An unexpected advantage to this approach was that statistics identified mass measurements exhibiting high standard deviations, suggesting peptide measurements with high potential for error. Finally, we demonstrate this method enhances the identification of proteins in trypsin digestion of standard protein samples and a biological sample from an immunoprecipitation experiment. We anticipate that simplified averaging and calibration algorithms including spectrum averaging and descriptive statistical measures can be readily incorporated into automated acquisition software, providing enhanced performance for any MALDI-TOF mass spectrometer (Mitchell *et al.*, 2015).

2.2 Materials and methods

2.2.1 Isolation and culture of human embryological cells

HEK293 cells were purchased from ATCC and grown at 37 °C in DMEM/high glucose media with 10% fetal bovine serum and 1% penicillin-streptomycin. At 90% confluence, cells were trypsinized, washed three times with warm deionized phosphate-buffered saline, then lysed in buffer containing 1% sodium deoxycholate, 1X protease inhibitor cocktail, and 1% nuclease. Total protein concentration was estimated using 280 nm absorbance. All reagents were purchased from Thermo Scientific except for Penicillin-Streptomycin, which was purchased from Gemini BioProducts.

2.2.2 Cell lysate pre-clearing

Cell lysate (1.275 mg) was centrifuged at 16,000 x g for 40 min at 4 °C and the supernatant was added to phosphate-buffered saline (PBS) containing 0.1% Tween-20 (0.1% PBST) and allowed to bind to 50 µl of Protein A/G agarose beads (Santa Cruz) for 2 h at 4 °C with mixing. After incubation with resin, the supernatant was isolated by centrifugation and used for subsequent experiments immediately or stored at −80 °C.

2.2.3 Preparation of antibody-linked Dynabeads

3 mg of Protein G Dynabeads (Life Technologies) were washed three times with 200 µl of PBST. After washing, 24 µg anti-β-tubulin rabbit polyclonal antibody (Abcam) was diluted in 0.1% PBST (200 µl) and incubated with the washed Dynabeads at 37 °C for 1 h with shaking at 1400 rpm. Beads were washed three times with 200 µl of PBST, then conjugation buffer (20 mM sodium phosphate, 0.15 M NaCl, pH = 7–9) for crosslinking. The antibody was crosslinked to the

Dynabeads using 250 µl of a freshly prepared solution of 5 mM Bis (Sulfosuccinimidyl) substrate (BS3, Thermo Scientific) and incubated at room temperature for 30 min with shaking at 1400 rpm. The reaction was quenched by adding 1 M Tris HCl, pH = 7.5 (12.5 µl) for 15 min at room temperature. Beads were subsequently washed three times with 200 µl of 0.1% PBST.

2.2.4 Immunoprecipitation

Anti-tubulin antibody-coupled Dynabeads were resuspended in precleared cell lysate and incubated with rotation at 4 °C overnight. After 18 h, the supernatant was decanted and saved for analysis. Beads were washed with 0.1% PBST (200 µl), then 500 mM NaCl (3 x 200 µl), PBS (200 µl) and H_2O (200 µl). Proteins were eluted using two washes in 20 µl of 5% NH_4OH for 5 min each wash. Elution fractions were combined and neutralized to pH 7 by the addition of 40 µl of 1 M NH_4HCO_3 and concentrated in vacuo.

2.2.5 Proteolytic digests

Immunoprecipitated proteins: Immunoprecipitated tubulin was dissolved in 10 µl of 50% acetonitrile in 100 mM ammonium bicarbonate, pH 7.2 and reduced and alkylated with TCEP and iodoacetamide. Proteins were precipitated using chloroform/methanol to eliminate residual reagents, detergents, lipids, and salts. The protein was then dissolved in 100 mM NH_4HCO_3 (60 µl) and 10 µl of a 25 ng/µl trypsin stock solution in 100 mM NH_4HCO_3 was added. After 18 h incubation at 37 °C, 10 µl of 10% formic acid was added to quench the reaction.

Purified proteins: A mixture of bovine serum albumin, myoglobin and β-Casein at 1 mg/ml for each protein was dissolved in 100 mM ammonium bicarbonate with 1% sodium deoxycholate. Fifty microliters of this standard protein solution were reduced and alkylated with TCEP and iodoacetamide. The deoxycholate concentration was reduced to 0.5% by adding 50 µl of 100 mM

ammonium bicarbonate and trypsin was added to the concentration of 1/50-1/100 by weight of the protein sample. The digestion was incubated at 37 °C for 18 h. After incubation, the reaction was quenched with 10 µl of 10% formic acid.

2.2.6 Western blot analysis

Immunoprecipitated tubulin proteins were separated by standard gel electrophoresis on Novex 4–12% precast gels and transferred to PVDF membrane using the Novex MiniCell blot module (Life Technologies). The blocking agent was 5% dry nonfat milk in tris-buffered saline (TBS) with 0.1% Tween 20 (TBST). The primary anti-β-tubulin antibody (Abcam) was used at 1:3000 dilution in a 5% solution of bovine serum albumin dissolved in 0.05% TBST. The secondary antibody was horse anti-mouse IgG horseradish peroxidase linked antibody (Cell Signaling) at 1:1500 dilution in a 5% solution of non-fat milk powder dissolved in 0.05% TBST for one hour with gentle shaking. The membrane was developed using the SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific) according to the manufacturer's instructions.

2.2.7 Liquid chromatography-tandem mass spectrometry

Liquid chromatography-tandem mass spectrometry was performed on a high-resolution Bruker MicrOTOF-Q (Q-TOF) mass spectrometer equipped with an Agilent 1200 Nano HPLC system using Solvent A, water with 0.25% formic acid and Solvent B, acetonitrile with 0.25% formic acid. The gradient was 5% Solvent B to 30% Solvent B over 85 min at a column flow rate of 1.4 µl/min on a home-packed 75 µm x 100 mm nanobore C18 column, Bruker captive spray source voltage at 1400 Volts, dry gas at 3.0 L/min, and capillary temperature at 150 °C. Data was collected in data-dependent mode fragmentation (MS/MS) on the 3 most intense spectra identified in MS mode. Spectra from the HPLC-MS/MS run were output into an MGF format for analysis in X!

Tandem (Current GPM2013.09.07). Settings for database search were high-resolution Q-TOF (MS tolerance = 100 ppm), fragment mass tolerance = 100 ppm, and trypsin digestion with semi-cleavage allowed, complete acetamidomethylation of cysteine, and potential modifications for methionine and tryptophan (oxidation) and asparagine and glutamine (deamination). Injection volume for the BSA sample was 0.05 µl (50 fmoles) and 8 µl of tubulin protein sample.

2.2.8 MALDI time-of-flight mass spectrometry

MALDI-TOF mass spectrometry was performed on an AB Sciex 4800 MALDI-TOF/TOF mass spectrometer using AB Sciex 4000 Series Data Explorer control and processing software (V3.7.1 Build 1, AB Sciex). Protein samples were diluted to approximately 1 pmol/µl, then 0.5 µl of the sample was combined with 0.1 µl of a 5-peptide calibration mixture (described below) and 0.5 µl of a saturated solution of α-cyano-4-hydroxycinnamic acid (Sigma Aldrich) dissolved in 50% acetonitrile in water with 1% formic acid and spotted onto a 384-well AB Sciex MALDI plate. Appropriate levels for accurate calibration were determined empirically. Each replicate spectrum acquired was a composite of 500 laser shots followed by 5-point non-linear calibration using the Data Explorer internal calibration procedure. Calibration employed a mixture of 5 standard peptides at approximately 100 ng peptides per 10 µl of α-cyano-4-hydroxycinnamic acid solution: Des-Arg(9) Bradykinin (MH+ = 904.4676), Human Angiotensin 1 (MH+ = 1296.6848), Glu-1-Fibrinopeptide B (MH+ = 1570.6768), ACTH 1–17 (MH+ = 2093.0862), ACTH 18–39 (MH+ = 2465.1983). All peptides were purchased from Bachem. These peptides were used for plate calibration of the AB Sciex 4800 mass spectrometer prior to analysis followed by close external and internal calibration of all spectra. Close external calibration was acquired by spotting the calibration peptide mixture in close proximity to the standard protein digest sample on the plate. Internal calibration incorporated the calibration peptide mixture directly into the protein digest

sample on the plate. Internal calibration provided more accurate data, but close external calibration provided increased mass spectral signal, presumably from less charge competition between standard and sample peptides. Calibration required 3 bins across peak-width-at-half-height for bin sizes of 0.5 and 1.0 nanosecond (ns) and 1 bin for 2.0 ns bin size. No calibration was achieved for 4.0 ns bin size. Bin size modification to 0.5, 1.0, 2.0 and 4.0 ns scale utilized the Digitizer Bin Menu in the Digitizer section of the AB Sciex 4000 Series Data Explorer mass spectrometer operating software. All spectra were required to pass internal calibration to be considered in subsequent analysis. For all experiments, 23 individual spectra were used for analysis to provide enough individual measurements for reliable statistics. However, not all peptides were observed in each spectrum (not uncommon in MALDI-TOF mass spectrometers). Peptide signals exhibiting signal-to-noise lower than approximately 5/1 were not included in reported measurements because of the difficulty in reliable peak assignment. Therefore, the number of individual spectral measurements for each peptide was between 10 and 23 different measurements. To evaluate the contribution of the calibration procedure on the observed variability, the initial calibration function derived from Data Explorer was deleted for multiple spectra and the spectra were subjected to the same calibration procedure as though the spectrum was acquired de novo. Multiple recalibrations of the same data provided the same calibration functions demonstrating that the calibration algorithm itself is not contributing to the observed data variability.

2.2.9 Calibration

Routine calibration of mass spectrometers is required for optimal performance. Numerous variables can be calibrated such as mass assignment, peak heights, quantitation, peak shape, and ion source. All mass spectrometers need to be calibrated to specify how masses are assigned

to peaks. Calibration compounds are used by a mass spectrometrist to adjust the mass calibration scale, as well as the relative intensities of the ions, to match that of known entities. This operation is performed on all mass spectrometers because subtle changes in electronics, cleanliness of surfaces, and a lab's ambient conditions can affect the instrument's ability to reproduce a meaningful measurement. For the least demanding analyses on nominal-mass instruments, the need for calibration can be infrequent and a check of its response more frequently. Nevertheless, high mass accuracy requires constant surveillance for minute changes. This is most commonly done by constructing a calibration curve based upon the mass spectrum of a known reference standard. Peak heights can also be calibrated. Quadrupole mass spectrometers require calibration to achieve a standardized abundance-versus-mass response. This is done by tuning peak heights for a standard sample. Quantitative methods such as require the creation of a calibration curve based on the response of varying quantities of a standard sample. These methods usually also require some calibration of the mass spectrometer so that the masses or reactions will be accurately monitored. The mass spectrometer parameters (ion source voltages, lens potentials, and also excitation and detection conditions in trapped-ion methods) all have to be adjusted to give good sensitivity and good peak shapes for a standard reference compound at some specified resolution value. Internal calibration is common throughout mass spectrometry and consists of mixing several compounds of known molecular weight with the analyte prior to mass analysis and then using the known masses to calculate the unknown ones. The improvement in mass accuracy is significant (usually ~10-fold) and very reliable (Connor and Costello, 2000). For the MALDI-TOF-TOF an internal calibration of a solution containing Des-Arg(9) bradykinin, human angiotensin 1, Glu-1-fibrinopeptide B, Adrenocorticotropic hormone (ACTH) including amino acids 1–17, and ACTH including amino acids 18–39. The calibration solution used for the LTQ consists of caffeine, L-methionyl-arginyl-phenylalanyl-alanine acetate-water (MRFA), and Ultramark 1621. Ultramark 1621 is a

commercially available mixture of fluorinated phosphazines and is a useful calibration compound for negative and positive ion fast-atom bombardment (FAB) high-resolution mass spectrometry. For the MicroTOFQ a commercial solution containing nine phosphazines and purine purchased from Agilent Technologies was used (Bark and Hook, 2007).

2.2.10 Descriptive statistical analysis

Descriptive statistics were calculated using the statistical functions in Excel (primarily AVERAGE and STDEV) for averaging, standard deviation and descriptive statistical calculation data (Excel 2010, Microsoft Corporation). More advanced statistical calculations such as Shapiro-Wilk normality, p-Value, and Grubbs Outlier tests used the SciStat Calc online statistics resource and online documentation of the statistical model algorithms (scistatcalc.blogspot.co.uk). The specific Shapiro-Wilk normality calculation algorithm has been previously adapted from Royston *et al.,* 1982 and distributed as a MatLab m-file in Matlab File Exchange.

2.3 Results and discussion

2.3.1 Replicate mass measurements and statistical analysis of MALDI-TOF data

Understanding the underlying mechanism influencing the observed variability in the MALDI-TOF instrument was of significant aim. Standard proteins were digested. A compilation of mass measurements for 11 different identified peptides from 23 individual spectra followed. Not all peptides were observed in each spectrum and the number of spectral measurements used for the analysis of each peptide is indicated in Table 2.1, Column N. These measurements underwent descriptive statistics calculations including mean, standard deviation, Shapiro-Wilk's normality test, and Grubbs outlier tests (Table 2.1). In addition to MALDI-TOF analysis, we also subjected these same digests to analysis on a high-resolution Q-TOF LC-MS/MS system for comparison

and confirmation of mass measurements. For most peptides, the combined replicate MALDI-TOF mass measurements exhibited a normal distribution as evidenced by Shapiro-Wilk's normality (Table 2.1). This suggests that several component error factors occur in combination and are essentially random and independent of each other. An immediate observation from these calculations was the improved consistency of the mean peptide mass compared to many of the measured masses from individual spectra. This consistency was observed even in the cases where Shapiro-Wilk's normality was not confirmed. Absolute values of measured errors evaluate the dispersion of the data and indicate higher variability in the more inaccurate data (Figure 2.2) with the standard deviations for the data consistently in the 0.010–0.014 amu range. The peptide mass measurement with the highest standard deviation (0.022 amu) was also the measurement most in error (8.996 ppm, Table 2.1) and had the greatest dispersion in absolute values of the errors (Figure 2.2). It is noted that a high standard deviation and high dispersion of data may still provide an accurate average but was not observed in these data. Therefore, these data suggest statistical analysis on a population of individual mass spectral measurements provides more consistently accurate MALDI-TOF data and that data with higher standard deviations may identify mass measurements with a higher probability of error.

Table 2.1 Data and statistical calculations for protein standard and tubulin IP

Protein Standard Peptides										
Peptide sequence	Protein	Protein Prospector	Q-TOF	MALDI MEAN	MALDI SD (amu)	Error (PPM)	N	Ave. S/N	Shapiro-Wilk (PASS)	Grubbs Outlier
1. RHPYFYAPELLYYANK	BSA	2045.0280	2045.028	2045.019	0.013	-4.358	16	822.188	0.9277(YES)	NO
2. LGEYGFQNALIVR	BSA	1479.7954	1479.795	1479.794	0.013	-3.624	16	454.813	0.9480(YES)	NO
3. DAFLGSFLYEYSR	BSA	1567.7427	1567.743	1567.739	0.014	-2.292	16	2039.375	0.8519(NO)	NO
4. KVPQVSTPTLVEVSR	BSA	1639.9377	1639.938	1639.933	0.013	2.293	15	381.133	0.9279(YES)	NO
5. VPQVSTPTLVEVSR	BSA	1511.8428	1511.8428	1511.829	0.022	8.996	16	23.125	0.9612(YES)	NO
6. GLSDGEWQQVLNVWGK	Myoglobin	1815.9024	1815.902	1815.896	0.012	-3.304	16	391.125	0.8488(NO)	NO
7. VEADIAGHGQEVLIR	Myoglobin	1606.8547	1606.855	1606.851	0.011	2.246	23	1991.957	0.9738(YES)	NO
8. KGHHEAELKPLAQSHATK	Myoglobin	1982.0566	1982.057	1982.053	0.014	0.891	19	1870.947	0.9102(YES)	NO
9. GHHEAELKPLAQSHATK	Myoglobin	1853.9617	1853.926	1853.957	0.012	2.584	20	154.850	0.9379(YES)	NO
10. HKIPIKYLEFISDAIIHVLHSK	Myoglobin	2601.4915	2601.492	2601.483	0.013	-3.200	12	229.167	0.9707(YES)	NO
11. YLEFISDAIIHVLHSK	Myoglobin	1885.0218	1885.022	1885.014	0.013	-4.042	15	403.200	0.9476(YES)	NO
Tubulin IP Peptides										
12. LHFFMPGFAPLTSR	Tubulin	1620.8355	1620.836	1620.832	0.008	-2.369	20	271.500	0.9284(YES)	NO
13. NSSYFVEWIPNNVK	Tubulin	1697.8170	1697.817	1697.813	0.014	-2.560	18	87.000	0.9723(YES)	NO
14. FPGQLNADLR	Tubulin	1130.5953	1130.595	1130.589	0.008	-5.917	20	43.350	0.9679(YES)	NO
15. YMACCLLYR	Tubulin	1249.5526	1249.553	1249.585	0.017	25.574	18	136.667	0.9179(YES)	NO
16. LAVNMVPFPR	Tubulin	1143.6343	1143.634	1143.632	0.013	-1.867	20	28.350	0.7533(NO)	NO
17. EVDEQMLNVQNK	Tubulin	1446.6893	1446.689	1446.743	0.016	37.229	17	45.294	0.8645(NO)	NO
18. SYELPDGQVITIGNER	Actin	1790.8919	1790.892	1790.894	0.014	1.251	20	94.250	0.9478(YES)	NO
19. IWHHTFYNELR	Actin	1515.7492	1515.749	1515.747	0.006	-1.528	18	95.278	0.8983(YES)	NO
20. AMGIMNSFVNDIFER	Hist2H2BF	1743.8193	1743.819	1743.824	0.009	2.667	20	151.850	0.9841(YES)	NO
21, LDIDSPPITAR	PKM2	1197.6474	1197.647	1197.655	0.018	5.904	17	19.177	0.9180(YES)	NO
22. LISWYDNEFGYSNR	GAPDH	1763.8024	1763.802	1763.822	0.016	11.041	20	52.000	0.9447(YES)	NO
23. THNLEPYFESFINNLR	Keratin	1993.9767	1993.977	1993.972	0.015	-2.649	17	109.176	0.9548 (YES)	NO
24. THNLEPYFESFINNLR	Keratin	1993.9767	1993.977	1993.873	0.298	-52.268	19	98.947	0.3951(NO)	YES

Peptides identified in both MALDI-TOF and LC-MS/MS experiments were subjected to multiple replicate measurements on the MALDI-TOF instrument and descriptive statistics were calculated on the population of data for each peptide (12–23 individual measurements from 23 complete spectra). The exact mass for each peptide was calculated in Protein Prospector (UCSF and compared to the masses observed for the Q-TOF LC-MS/MS system (Q-TOF) and the mean of replicate measurements for the MALDI-TOF instrument (MALDI MEAN). The standard deviation (MALDI SD) is reported in amu and the error (Error) was calculated in PPM. The number of replicate measurements (N), the raw Shapiro-Wilk threshold criteria (Shapiro-Wilk (PASS)) and Grubbs Outlier results are reported in their respective columns. Note that Peptide THNLEPYFESFINNLR is listed in line 23 and 24 with all mass spectral measurements (N = 19) for this peptide used for the statistical data in line 24 (denoted with a *). However, Grubbs Outlier Test analysis indicated positive outliers for two spectral measurements in these data, which were removed (N = 17) to provide the statistical information presented in line 23. Reprinted from "Enhancing MALDI Time-Of-Flight Mass Spectrometer Performance through Spectrum Averaging," by M. Mitchell, S. Mali, C.C. King, S.J. Bark, 2015, *PLoS One*, 10(3):1-15. Copyright [2015] by Mitchell et al.

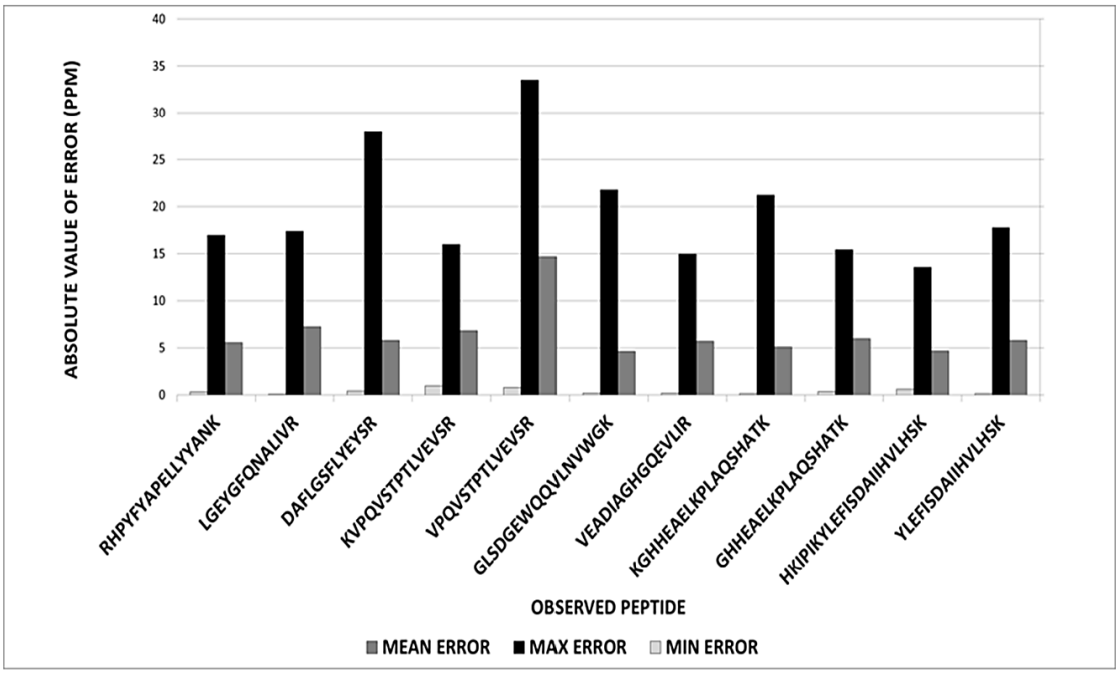


Figure 2.2 Absolute values for mean, maximum and minimum errors observed for peptides from a standard protein trypsin digestion. Data demonstrate higher variability in more inaccurate data. Measured errors were converted to absolute values to evaluate the dispersion of data measurements and plotted according to the peptide. The uneven distribution of maximumand minimum errors is expected because of zero as a lower bound for minimum error. Note, absolute value transformation of the data eliminates negative values and the mean errors reported here are higher than the mean reported for the observed peptide masses, which contain both positive and negative errors. Reprinted from “Enhancing MALDI Time-Of-Flight Mass Spectrometer Performance through Spectrum Averaging,” by M. Mitchell, S. Mali, C.C. King, S.J. Bark, 2015, *PLoS One*, 10(3):1-15. Copyright [2015] by Mitchell et al.

2.3.2 MALDI-TOF data are comprised of discontinuous bins

Specific evaluation of peak shape in all observed spectra demonstrated discontinuous intensity measurements or sampling points across every peak, suggesting the data was parsed into bins. In the particular mass spectrometer employed for these studies, we observed the discontinuous spacing or bins at 0.0178 amu for Des-Arg(9) Bradykinin (MH+ = 904.4676), 0.0214 amu for Human Angiotensin 1 (MH+ = 1296.6848), and 0.0272 amu for ACTH 1–17 (MH+ = 2093.0862)

(Figure 2.3). The differences in bin spacing are correlated with the difference in flight times for ions of different masses by a simple ratio equation that accurately relates the bin spacing to 5 decimal places, suggesting strongly that the differences in the ion flight times in the instrument and the binning of data are related (Figure 2.3). The actual signal from flight times for ions in the mass spectrometer should produce a continuum peak shape (Figure 2.3). However, the AD detector system can only measure the intensity of ions in discrete packets of signals across the time spectrum defined by the high-speed internal clock (500–1000 MHz or higher) and internal electronics (Figure 2.3). The data bins widths observed in Figure 2.3 correspond to 0.0178 amu for a mass of 904.4729 in the mass spectrometer and increasing for higher masses (0.0214 amu at 1296.6829 and 0.0272 at 2093.0735). These bin widths correlate to 19.7 ppm, 16.5 ppm, and 13.0 ppm, respectively. The resulting bin spacing is reasonably consistent with the standard deviation for the peptide measurements (Table 2.1). These observations are also consistent with our observation that the number of laser shots improves signal-to-noise of the data with little impact on accuracy.

2.3.3 Multiple mechanisms may impact variability in MALDI-TOF data consistency

TOF mass spectrometers measure ions by detecting the flight time of the ions from the time of laser shot through delayed extraction to impact at the detector. The data demonstrate a shifting of bin positions between replicate measurements, which we initially interpreted as variation in the flight time measurements. With each laser shot, the start of acquisition defines the time range and sets the position of the bins. Each subsequent laser shot is an independent measurement and sets new bin positions within the inherent error of the electronics. However, this variability is quite small and usually limited to less than the width of a single bin. For this reason, most MALDI-TOF software appears to ignore these variations, but it cannot be rigorously eliminated as a possible

source of the variation we observed in our data. What is unequivocal is that the final data exhibits binning and, considering the bin width observed between 13 and 20 ppm, variation errors less than one bin width could have a significant impact. While our initial data suggested to us variation in the flight time measurement is important in our observed data variability, the further evaluation suggests calibration may be more critical. After collection of all laser shots, the data are combined and calibrated through a calibration function based on known internal masses (5-point nonlinear internal calibration in our experiments). The discontinuous binning of the flight time measurements would impact the peak interpolation algorithms for peptide masses and calibration simultaneously. In these experiments, we are averaging data from multiple spectra of 500 laser shots/spectrum with 5-point internal calibration and a unique calibration for each spectrum (averaging 10 unique spectra with 10 unique calibration functions). This is fundamentally different than collecting individual spectra averaging 10,000 or more laser shots prior to calibration, which would provide the same number of combined laser shots, but with a single calibration function.

To evaluate these factors, we performed two experiments. First, we collected data at bin widths of 0.5, 1.0, 2.0 and 4.0 ns for 4 different peptides with 10 replicate measurements (Figure 2.4). We anticipated that the standard deviation as a measure of variability would increase with increased bin width if binning were a factor contributing to variability.

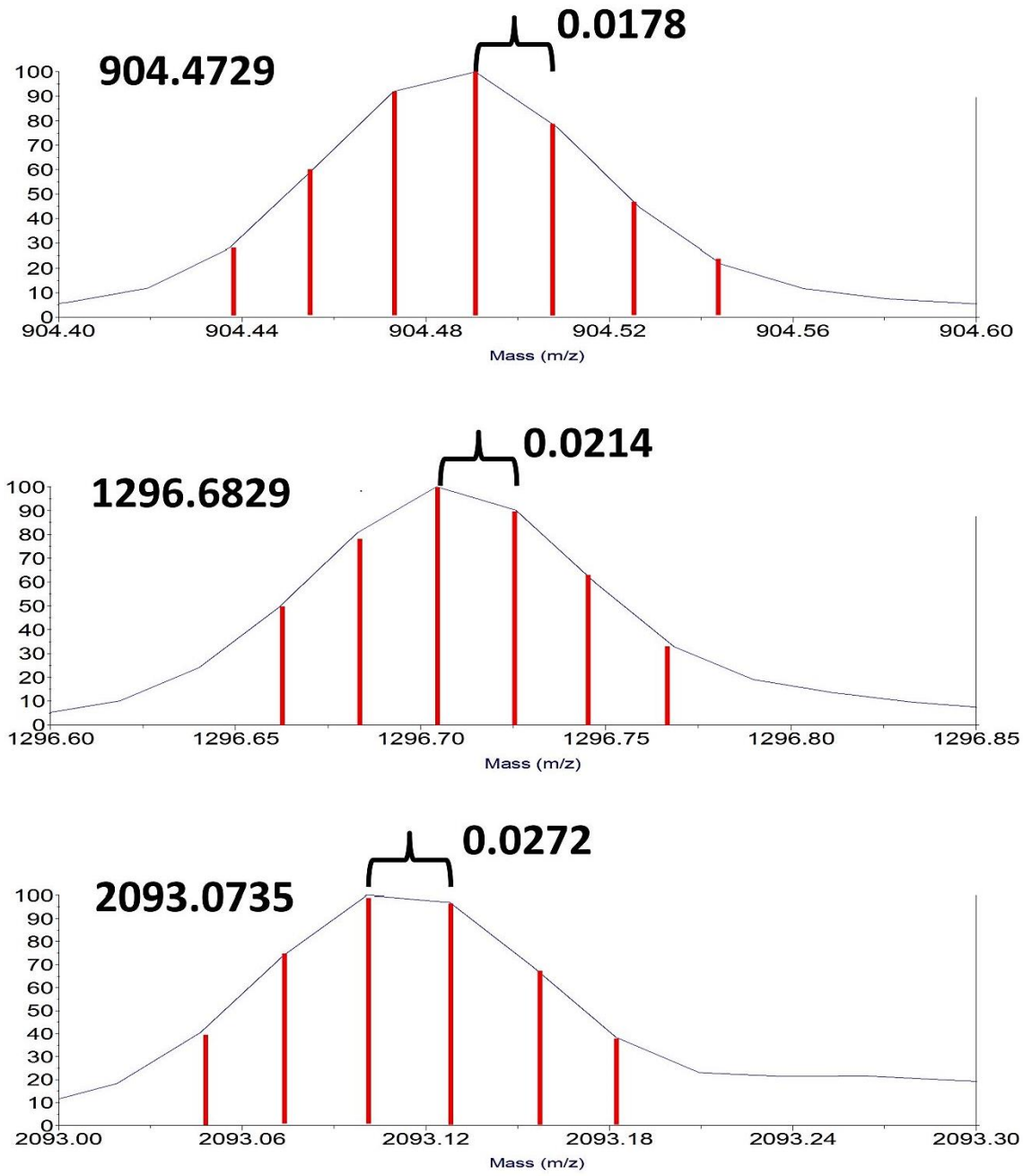


Figure 2.3 Evidence for discontinuous binning of MALDI-TOF mass spectrometry data for Des-Arg(9) Bradykinin, Angiotensin 1, and ACTH 1–17. Representative monoisotopic peaks for Des-Arg(9) Bradykinin (MH+ = 904.4676), Angiotensin 1 (MH+ = 1296.6848) and ACTH 1–17 (MH+ = 2093.0862). The red vertical lines are fitted to the bins evident in the observed peak shapes. The spacing of bins in mass units is larger for higher mass ions and can be accurately calculated by the relationship between flight times and the ratio of masses of the molecular ions according to the equation $\Delta t2/\Delta t1 = (M2/M1)1/2$. Reprinted from "Enhancing MALDI Time-Of-Flight Mass Spectrometer Performance through Spectrum Averaging," by M. Mitchell, S. Mali, C.C. King, S.J. Bark, 2015, *PLoS One*, 10(3):1-15. Copyright [2015] by Mitchell et al.

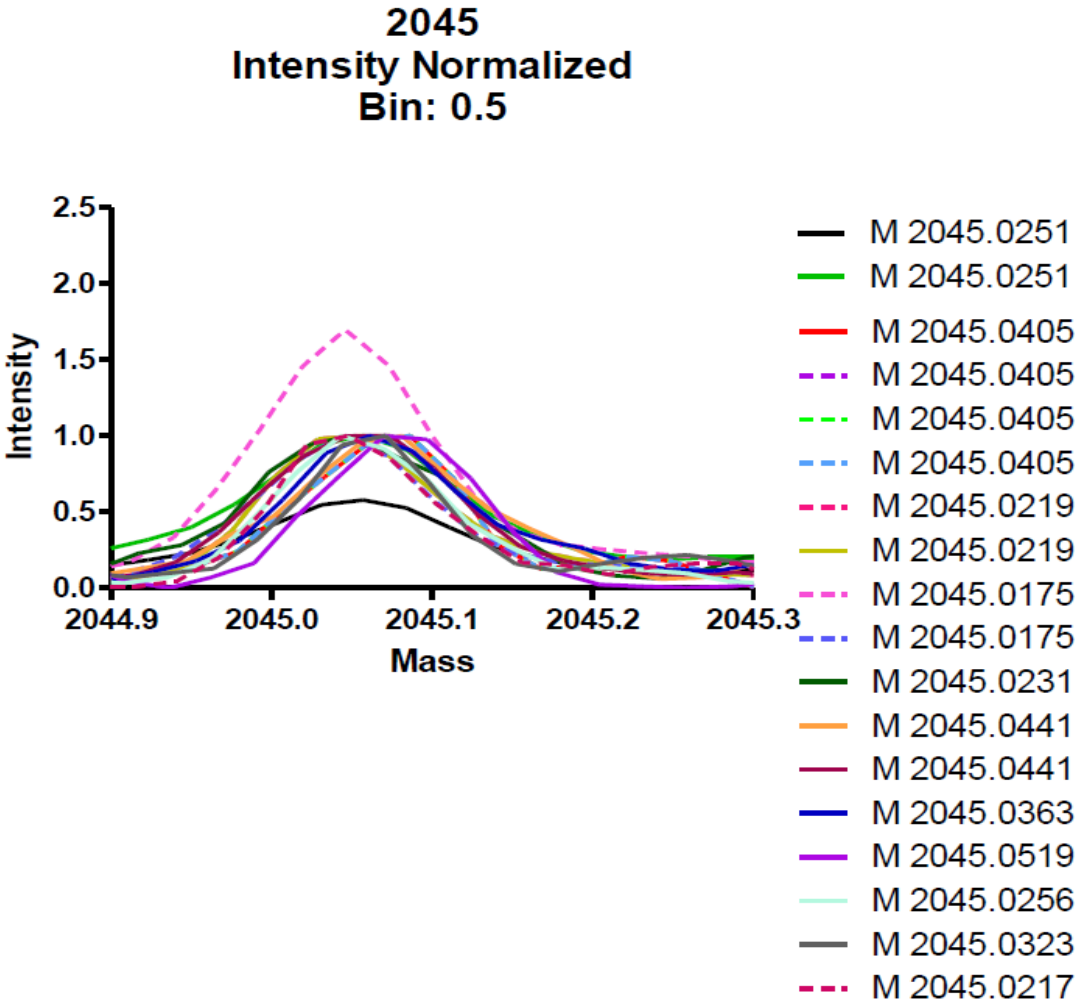
Our data shows that standard deviation is little impacted by increased bin width from 0.5 to 1.0 ns,but increased from 1.0 to 2.0 ns (Figure 2.4). This data suggests a significant tolerance by the software interpolation algorithms for the number of data points across peaks when there are enough data points to define the peak. However, by 2.0 ns bin width, the limits of the interpolation software are reached. For this reason, bin widths for data above 2.0 ns were unable to be calibrated. In the second experiment, we attempted to perform multiple calibrations on a single spectrum to determine if the calibration algorithm itself was contributing to the variances observed in the spectral data. However, recalibrating a single spectrum multiple times derived the identical calibration function and mass measurement, indicating that the calibration algorithm itself is not contributing to the observed variability. Initially, we thought this indicated that calibration was not important. However, calibration may be important when considering subtle differences present in multiple independent spectra. Therefore, we anticipate the observed variability in data is attributable to both the unique spectrum and the unique calibration function based on that spectrum. However, we also note that definitively determining the sources of the variation we observe in our MALDI–TOF data are beyond the capabilities of our instrumentation and resources.

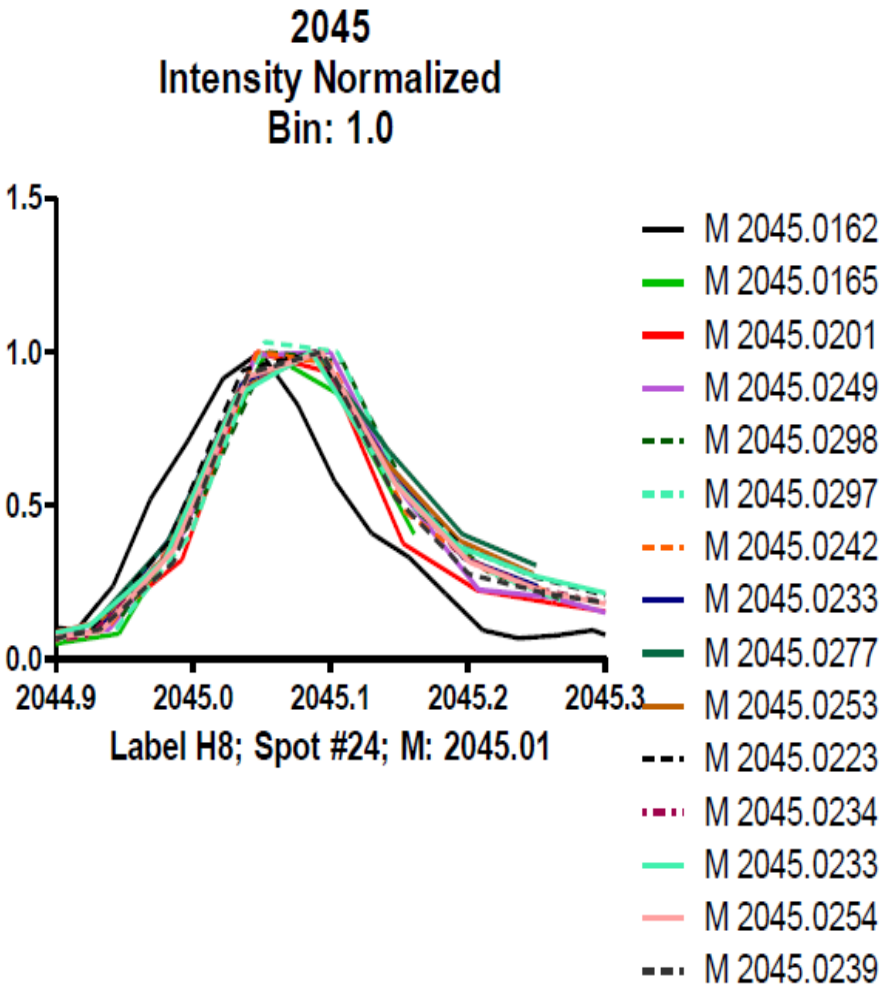
2.3.4 Spectral averaging improves the accuracy and performance of MALDI-TOF mass spectrometry in protein identification

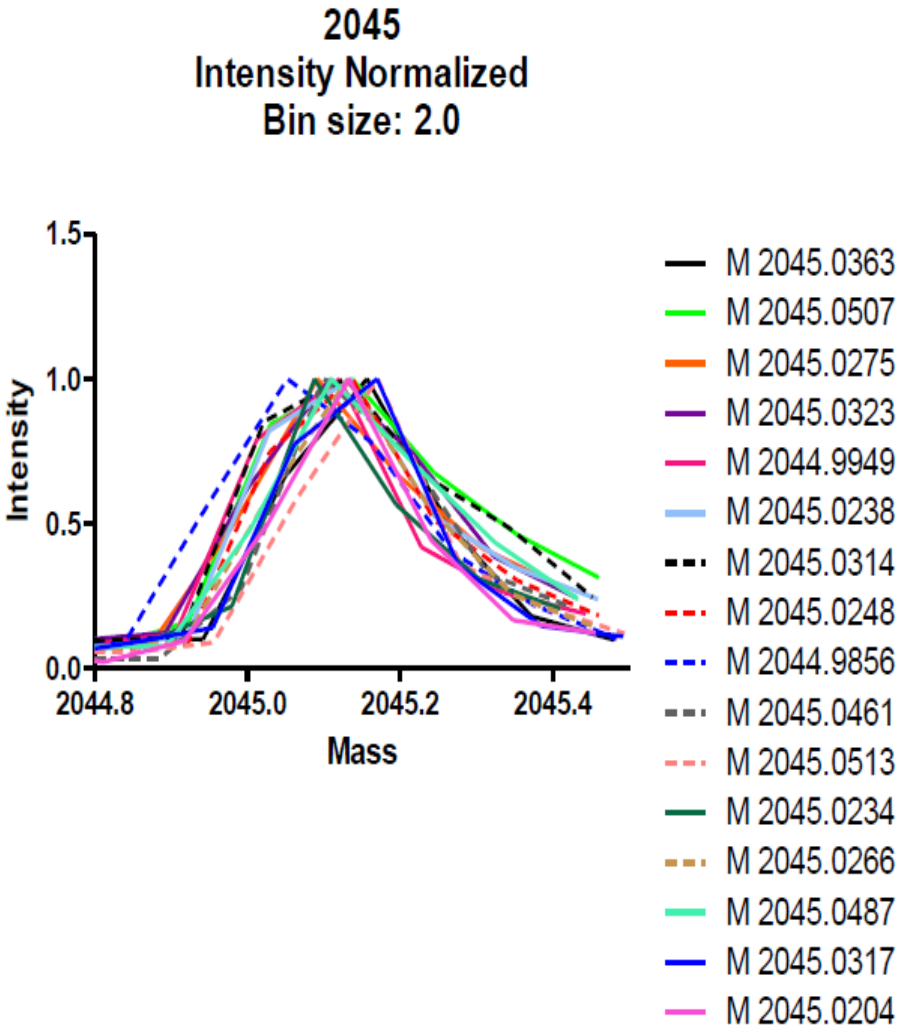
The Gaussian distribution of the observed data variability suggested that averaging multiple individual spectra would result in enhanced MALDI-TOF performance. The assignment of the MALDI-TOF data into bins places a significant limitation on the accuracy of the instrument. However, the electronic variability in the AD detector and calibration system can be exploited to

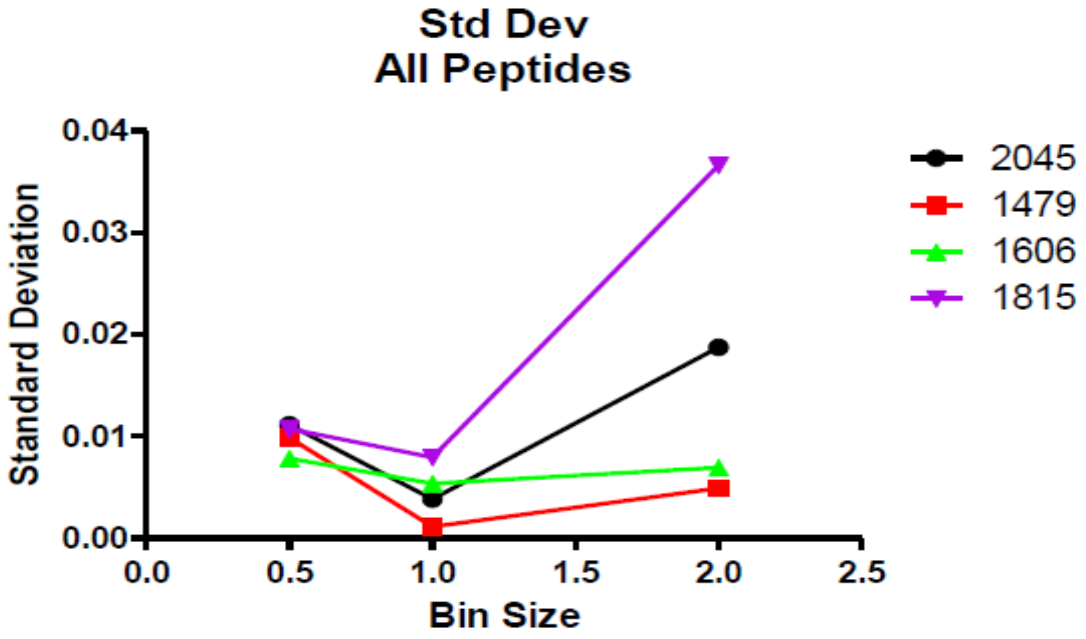
provide random repositioning of the data bin positions by combining multiple independently calibrated acquisitions. This explains why our statistical mapping of the population of data for

Figure 2.4 Overlay of 10 individual spectra for the monoisotopic peak at 2045 at bin size 0.5 nanoseconds, 1.0 nanoseconds, 2.0 nanoseconds and bin width and standard deviation correlation. The intensity was normalized to 1.0 for each spectral peak. Increasing bin width to 1.0 nanoseconds initially improves standard deviation but extending to wider bin widths to 2.0 nanoseconds provides fewer sampling points across the peak leading to increased data variability. This variability is quantitated by the standard deviation for the data. Reprinted from "Enhancing MALDI Time-Of-Flight Mass Spectrometer Performance through Spectrum Averaging," by M. Mitchell, S. Mali, C.C. King, S.J. Bark, 2015, *PLoS One*, 10(3):1-15. Copyright [2015] by Mitchell et al.




2045
Intensity Normalized
Bin: 1.0
1.5
1.0
0.5
0.0
2044.9
2045.0
2045.1
2045.2
2045.3
Label H8; Spot #24; M: 2045.01
M 2045.0162
M 2045.0165
M 2045.0201
M 2045.0249
M 2045.0298
M 2045.0297
M 2045.0242
M 2045.0233
M 2045.0277
M 2045.0253
M 2045.0223
M 2045.0234
M 2045.0233
M 2045.0254
M 2045.0239


2045
Intensity Normalized
Bin size: 2.0
Intensity
1.5
1.0
0.5
0.0
2044.8
2045.0
2045.2
2045.4
Mass
M 2045.0363
M 2045.0507
M 2045.0275
M 2045.0323
M 2044.9949
M 2045.0238
M 2045.0314
M 2045.0248
M 2044.9856
M 2045.0461
M 2045.0513
M 2045.0234
M 2045.0266
M 2045.0487
M 2045.0317
M 2045.0204


Std Dev
All Peptides
Standard Deviation
0.04
0.03
0.02
0.01
0.00
0.0
0.5
1.0
1.5
2.0
2.5
Bin Size
2045
1479
1606
1815

standard protein digest peptides provided more consistent mass measurements than individual spectra. We tested this approach by identifying a protein by peptide mass fingerprinting from an immunoprecipitation experiment for β-tubulin. β-tubulin is an abundant cytoskeletal protein important in multiple cellular functions including cell structure, morphology, and intracellular trafficking. Using a rabbit polyclonal antibody to this protein immobilized on protein G Dynabeads, we immunoisolated β-tubulin from HEK293 cell lysate and confirmed using Western blot (Figure 2.5). The remaining immunoisolated β-tubulin was reduced and alkylated prior to digestion with trypsin and analysis by MALDI-TOF mass spectrometry. As described for the standard protein digest, 20 mass spectra were obtained for this digest, between 17 and 20 individual mass measurements were observed for each peptide, and the data were subjected to descriptive statistical analysis as described for data from the standard protein digestion (Table 2.1). The β-tubulin digest was also analyzed using high-resolution LC- For the majority of peptides analyzed, these data demonstrate the same advantages from the spectrum averaging approach as described above for the standard protein digests. These peptide mass data were employed for peptide mass fingerprinting using ProFound and successfully identified β-tubulin from the immunoprecipitation sample using mass data from all observed peptides, even those peptides identified from LC-MS/MS as not derived from tubulin (Table 2.1). We note that the identification of a protein from accurate mass measurements where peptides from multiple proteins are represented is inherently problematic for peptide mass fingerprinting.

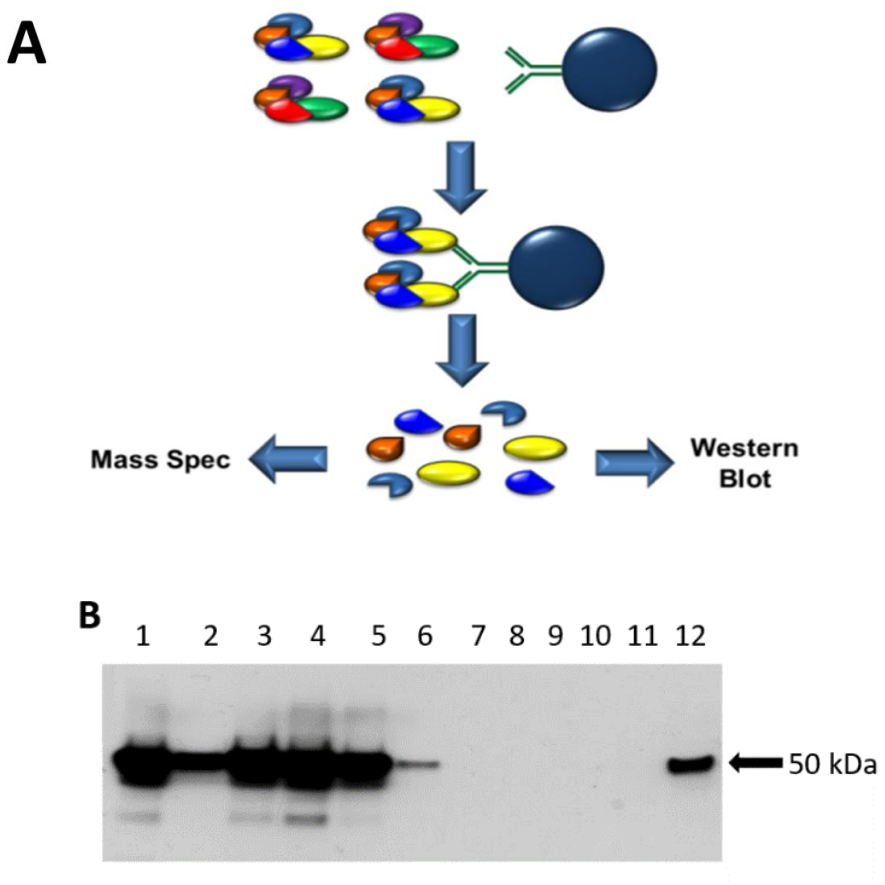


Figure 2.5 Immunoaffinity isolation of β-tubulin. (A) Schematic for immunoaffinity isolation. The antibody is immobilized on solid resin support and the specific protein interacting with this antibody is isolated. Under appropriate conditions, proteins interacting with the specific target protein can also be isolated (co-immunoisolation or coimmunoprecipitation). (B) Western blot for immunoaffinity isolation of β-tubulin. Lane 1: cell lysate, Lane 2: unbound supernatant, Lane 3-11: wash steps, Lane 11: acid elution, Lane 12: base elution.MS/MS for confirmation of mass measurement accuracy obtained on the MALDI-TOF instrument. Reprinted from "Enhancing MALDI Time-Of-Flight Mass Spectrometer Performance through Spectrum Averaging," by M. Mitchell, S. Mali, C.C. King, S.J. Bark, 2015, *PLoS One*, 10(3):1-15. Copyright [2015] by Mitchell et al.

As expected, the identification of tubulin became more tenuous with increased stringency on accuracy. However, employing some of the tools suggested by the spectrum averaging experiments for the standard digests provided enhanced identification of β-tubulin. The standard peptide experiments demonstrated a standard deviation between 0.010–0.014 amu. In the β-

tubulin IP experiment, 5 of the 12 peptides identified, exhibit deviations above this range. Two of these peptides were identified by LC-MS/MS as derived from tubulin. Four deviations are minor, but with restricting the peptides for the second round of analysis to those identified to tubulin in the initial fingerprinting analysis and removing the two peptides with high standard deviations, the identification of tubulin was enhanced. Such an approach improves the confidence of correct identification, but further experiments using alternative techniques are necessary for confirmation. While protein identification using ProFound was effective for the β-tubulin IP, there was an unexpected error (Table 2.6). The peptide identified as ALTVSELTQQMFDSK (MH+ = 1697.8415) in ProFound was alternatively identified using LC-MS/MS sequence fragmentation data as NSSYFVEWIPNNVK (MH+ = 1696.8329) with an N-terminal asparagine deamidation to produce the sequence DSSYVEWIPNNVK (MH+ = 1697.817) (Table 2.6). The observed ion mass in the MALDI-TOF instrument was 1697.813, which is an error of −2.4 ppm with the peptide ion sequence assigned by LC-MS/MS analysis, but 17 ppm error for the assigned sequence from ProFound ([ion mass observed—ion mass theoretical] /ion). In addition to the issues noted for peptide mass fingerprinting with peptides derived from multiple proteins, these data are important reminders of the limitations for protein identification based on accurate mass measurements alone. We note that several immunoprecipitation peptide data standard deviations were not grossly out of proportion compared to test peptides, yet three peptides exhibited errors > 10 ppm (YMACCLLYR line 15, EVDEQMLNVQNK line 17, and LISWYDNEFGYSNR line 22, Table 2.1). Analysis of the raw data for these measurements demonstrated a systematic error resulting in the shifting of all measurements in the same direction. Unfortunately, systematic errors in a mass spectrometer are difficult to identify without external mass information and statistical analysis cannot resolve this problem. However, we were able to identify these errors effectively using the LC-MS/MS identifications.

Table 2.6 Identification of β-tubulin peptides by ProFound peptide mass fingerprinting using immunoprecipitation experiments and averaged MALDI-TOF/TOF data

Mass Tolerance	Top Protein ID	Expectation Value	Peptide Set	Search Mass Range/pI Range
50ppm	TUBB	2.2x10-3	All	0-3000kDa/0-14
25ppm	TUBB2A	0.011	All	0-3000kDa/0-14
10ppm	Tubulin Beta Proteins	0.074	All	0-3000kDa/0-14
50ppm	TUBB2A	2.7x10-4	Tubulin	0-3000kDa/0-14
25ppm	TUBB2A	1.4x10-3	Tubulin	0-3000kDa/0-14
10ppm	Tubulin Beta Proteins	7.8x10-3	Tubulin	0-3000kDa/0-14
5ppm	Tubulin Beta Proteins	0.011-0.021	Tubulin	0-3000kDa/0-14

β-tubulin was immunoprecipitated from HEK293 cell lysate using a rabbit polyclonal antibody and protein A/G Dynabeads. After isolation, all eluted proteins were reduced and alkylated, digested with trypsin, and analyzed on an ABI 4800 MALDI-TOF/TOF mass spectrometer. Multiple individual spectra were acquired with internal calibration and between 10 and 23 individual measurements for each peptide were used for calculating the average observed masses for each peptide. Mass tolerance is in parts-per-million (ppm), Peptide Set defines the peptides included for search and search mass range/pI range are input parameters for ProFound. Top protein ID and expectation value were calculated within ProFound from the mass spectrometry data using the IPI human database. Reprinted from “Enhancing MALDI Time-Of-Flight Mass Spectrometer Performance through Spectrum Averaging,” by M. Mitchell, S. Mali, C.C. King, S.J. Bark, 2015, *PLoS One*, 10(3):1-15. Copyright [2015] by Mitchell et al.

For one specific peptide, THNLEPYFESFINNLR (Line 23 and 24, Table 2.1), the magnitude of standard deviation was an effective indicator of measurement errors. Using all mass measurement data for this peptide (N = 19), we observed a standard deviation of 0.298 ppm and a calculated error of −52.268 ppm (Line 24, Table 2.1). These values are vastly greater than expected from standard peptide data. The Grubbs outlier test indicated two individual measurements were outside of the Grubbs test limit. We note that the Grubbs test is only valid for a single outlier measurement and we did not apply the test recursively.

Table 2.7 β-tubulin identification and limitations of peptide mass fingerprinting

Mass Observed	Mass Expected	Error (PPM)	Residue Number	Peptide Sequence
1129.581	1129.588	-6	242-251	FPGQLNADLR
1142.624	1142.627	-2	253-262	LAVNMVPFPR
1445.735	1445.681	37	325-336	EVDEQMLNVQNK
1619.824	1619.828	-2	263-276	LHFFMPGFAPLTSR
1696.805*	1696.833*	-17*	283-297*	ALTVSELTQQMFDSK*

The identification of tubulin from peptide mass fingerprinting matches with LC-MS/MS data for the majority of peptides. However, the assignment for the 1249.585 and 1696.805 peptides was inaccurate. The 1249.585 peaks were not assigned in the MALDI-TOF data. The observed mass of 1696.805 and data in Table 1 are from ProFound and assigned to peptide sequence ALTVSELTQQMFDSK (data denoted with an asterisk *). Peptide fragmentation data using LC-MS/MS suggests this assignment is incorrect and that the correct sequence is NSSYFVEWIPNNVK with deamidation at the amino-terminus yielding the sequence DSSYFVEWIPNNVK (MH+ = 1697.817 expected, 1697.813 observed, −2.4 ppm error). Reprinted from “Enhancing MALDI Time-Of-Flight Mass Spectrometer Performance through Spectrum Averaging,” by M. Mitchell, S. Mali, C.C. King, S.J. Bark, 2015, *PLoS One*, 10(3):1-15. Copyright [2015] by Mitchell et al.

Removal of these outlier measurements (N = 17) vastly improved the quality of the data for this peptide. (standard deviation = 0.015 amu, error = −2.649 ppm, Line 23, Table 2.1). The averaging approach used in this study was tested up to 2465.1983 amu (ACTH 18–39 Human) because of our confirmation of the correct identification of peptides by MS/MS. The tuning and calibration for our 4800 MALDI-TOF/TOF system are optimized for this lower mass range because of high performance in the lower masses and problematic and unreliable peptide fragmentation above approximately 3500–4000 amu. However, the reflectron system on this instrument is capable of higher mass analyses for mass measurements of parent masses, but not for fragment masses. We tested the higher mass reflectron system on our instrument using insulin (monoisotopic mass = 5803.6376, average mass = 5807.57) with the acknowledged limitations that the reflectron system was not tuned for this higher mass. As a consequence, the observed data has lower

sensitivity, the resolution is unable to define monoisotopic peaks, and observed masses are consistent with average mass rather than monoisotopic mass. Despite these issues, we observed the same type of variability as observed in the lower mass range with a majority of single mass observations inferior to the average of the population of mass observations (Table 2.1). We fully expect that proper tuning and calibration of the reflectron system of the 4800 MALDI-TOF/TOF for higher masses such as insulin would provide similar advantages as observed in the properly tuned and calibrated lower mass region. However, the capabilities of this instrument to provide high accuracy reflectron data across a wide mass range from <1000 to >5000 amu will require further experiments. These data demonstrate that, given that similar analog-to-digital detection systems are utilized in all current MALDI-TOF instruments and that the same detection systems are used in both reflectron and linear modes, the averaging methods described in this article should be applicable across all mass ranges and modes. Reflectron data will be limited to lower mass range than linear data but exhibit much higher accuracy as long as the reflectron system is tuned and calibrated for the appropriate mass ranges under study. Analog-to-digital detection systems are not limited to MALDI-TOF instruments but are also incorporated into LC-TOF and LC-Q-TOF instruments as well. We expect that the same averaging approach used to enhance the mass accuracy and consistency in the AB Sciex 4800 MALDI-TOF/TOF mass spectrometer can be applied to mass spectrometers of LC-TOF and LC-Q-TOF design. These studies did not utilize peak intensity as a parameter for averaging. Peak intensity can provide valuable data for quantitation and more consistent and reproducible quantitation would be highly advantageous.

Table 2.8 Raw data for higher mass range insulin data

RAW DATA FOR HIGHER MASS RANGE INSULIN DATA

Obs. Mass	S/N	Δ (amu)	Δ(ppm)				
5808.31	49	0.73	125.697795	5807.74	19	0.16	27.55020163
5809.20	38	1.62	278.9457915	5809.86	17	2.28	392.5903733
5809.70	31	2.12	365.0401716	5807.26	13	-0.32	-55.10040327
5809.73	23	2.15	370.2058344	5811.25	24	3.67	631.93275
5807.85	37	0.27	46.49096526	5808.18	24	0.6	103.3132561
5809.90	13	2.32	399.4779237	5808.00	55	0.42	72.31927929
5808.14	23	0.56	96.42570572	5807.98	10	0.4	68.87550408
5807.44	14	-0.14	-24.10642643	5809.13	13	1.55	266.8925783
5806.59	29	-0.99	-170.4668726	5810.21	15	2.63	452.8564393
5809.73	13	2.15	370.2058344	5808.42	21	0.84	144.6385586
5806.92	36	-0.66	-113.6445817	5806.70	15	-0.88	-151.526109
5807.48	19	-0.1	-17.21887602	5810.10	17	2.52	433.9156757
5805.71	11	-1.87	-321.9929816	5808.55	22	0.97	167.0230974
5807.73	22	0.15	25.82831403	5809.54	10	1.96	337.48997
5804.94	13	-2.64	-454.5783269	5808.68	24	1.1	189.4076362

Insulin Monoisotopic Mass = 5803.638 amu
Insulin Average Mass = 5807.58 amu

Data Average Mass = 5808.366 amu
Data Average Delta (ppm) = 135.283

To confirm the observed variability at the lower mass range also occurs at higher mass range using the reflectron system in the 4800 MALDI-TOF/TOF mass spectrometer, we added insulin (monoisotopic mass = 5803.6376, average mass = 5807.57) to our standard peptide mixture for internal calibration and the data for 30 replicate measurements for insulin were acquired and internally calibrated as described above. All data reported and calculated in this table use the same methods as described for Table 2.1. Reprinted from “Enhancing MALDI Time-Of-Flight Mass Spectrometer Performance through Spectrum Averaging,” by M. Mitchell, S. Mali, C.C. King, S.J. Bark, 2015, *PLoS One*, 10(3):1-15. Copyright [2015] by Mitchell et al.

Additionally, peak intensity averaging could be readily incorporated into automated software platforms. Careful analysis of the raw peptide intensity data demonstrates that the variability in intensity measurements are much more pronounced than variability in mass measurements in the MALDI-TOF instrument (Appendix, Table 1). This does not imply that averaging would not improve the consistency of intensity measurements and quantitation capabilities of this instrument, but that the larger standard deviations observed would be potentially problematic for comparing multiple data sets. We note that the peak intensities observed in the LC-Q-TOF

instrument were far more consistent and reproducible. However, given the high performance of current LC-TOF type instruments such as the Bruker MicroTOF-Q instrument used in these studies, the gain in data quality by averaging would be expected to be much smaller than observed for MALDI-TOF instruments.

2.4 Conclusions

We demonstrate a simple technique to improve MALDI-TOF instrument performance using spectral averaging. While LC-MS approaches have become predominant in recent biological research, enhanced performance MALDI-TOF instruments have unique and important capabilities for multiple biological fields including biochemistry, cell and molecular biology, chemistry, protein sciences, natural products, and microbial research. The theoretical limits for MALDI-TOF mass spectrometers far exceed any currently available instrument and we anticipate the future development of instruments with enhanced performance characteristics. The spectrum averaging and descriptive statistics analysis described here can be easily incorporated into automated data collection algorithms and enhance the consistency and data quality for both current and future generation MALDI-TOF instruments.

Chapter 3:

Affinity Purification Mass Spectrometry

Preface

Work outlined in this chapter has been published in Analytical Biochemistry (Mali S, Moree WJ, **Mitchell M**, Widger W, Bark SJ. 2016. Observations on different resin strategies for affinity purification mass spectrometry of a tagged protein. Analytical Biochemistry. 15(515):26-32.)

3.1 Introduction

Co-affinity purification mass spectrometry (CoAP-MS) is a highly effective method for isolating and identifying protein interactions from a complex biological sample. CoAP-MS uses a solid-support affinity system to selectively bind the protein of interest. Successful CoAP-MS experiments require affinity purification (AP) of a bait protein, which is often achieved using well documented high-affinity antibodies to unique peptide sequence tags such as FLAG (DYKDDDDK), c-Myc (EQKLISEEDL), HA (YPYDVPDYA), or V5 (GKPIPNPLLGLDST). Protein expression with these tags enables selective isolation without requiring an antibody specific for the bait protein (Chang, 2006). Evaluation of the AP of the bait protein by mass spectrometry is useful to identify factors influencing the AP procedure, which is the focus of this article. For clarity, we denote AP-MS to be consistent with the alternative IP and CoIP terminology.

Immobilization of the antibody to a support resin enables selective enrichment of the tagged protein through noncovalent or covalent methods. Noncovalent immobilization usually uses Protein A and G covalently attached to a resin matrix (agarose, sepharose, or paramagnetic bead) (Nathan *et al.,* 1992). Unfortunately, elution conditions that destabilize the antibody-antigen

interaction often destabilize the protein A/G-antibody interaction, resulting in co-elution of the antibody. Diverse covalent antibody-immobilization strategies include prelinked resin-antibody or resin-protein conjugated beads (for example, EZ View and streptavidin-sepharose), use of a bifunctional crosslinking agent such as dimethyl pimelimidate (DMP) or bis(sulfosuccinimidyl) suberate (BS^3) to couple the antibody to protein A and G, or beads with functionally derivatized surfaces (acids, amines, maleimides, succinimidyl esters, epoxy, or aldehydes) (Maertens *et al.,* 2015). Each strategy has unique advantages and challenges that change antibody immobilization and subsequent affinity purification of the bait protein. The impact of the solid-support resin is often overlooked. In this article, we evaluated protein A/G agarose, protein G Dynabeads and AminoLink aldehyde-functionalized resins under different crosslinking conditions (Figure 3) using two protein tag strategies: Anti-FLAG M2 Clone antibody/N-terminal tagged FLAG-p53 (N-FLAGp53) and Anti-HA antibody/N-terminal tagged HA-PTEN (N-HA-PTEN), both expressed in HEK293 cells. PTEN and p53 are tumor suppressor proteins mutated in many cancer types and have several known protein interactions. Confining our experiments to these specific antibody/labeled protein pairs, we were able to compare widely used commercial solid-supports under similar conditions. Our initial AP-MS experiments used widely available epoxy functionalized Dynabeads. However, subsequent experiments over 6-8 weeks exhibited progressively reduced affinity purification, likely because of epoxide hydrolysis. This prompted us to evaluate protein A/G Agarose, protein G Dynabeads and AminoLink aldehyde-functionalized resins to optimize resin immobilization, protein binding, and elution conditions using N-FLAG-p53 as a model. We then used N-HA-PTEN for more detailed comparisons. We note that all tested resins can perform adequately for AP-MS under compatible conditions, but require optimization of tag, crosslinking strategies, lysis buffers, and binding and elution conditions (LaCava *et al.,* 2015). However, under the conditions of our experiments, AminoLink resin provided a more robust platform for affinity purification of tagged proteins. The objective of this article is to provide insight

into our experimental observations and discuss the impact of solid-support resins on AP-MS experiments.

3.2 Materials and methods

3.2.1 Antibodies and reagents

PTEN and p53 antibodies were purchased from Cell Signaling Technology (p53 #9282 Rabbit Ab, PTEN (D4.3) XP I #9188 Rabbit mAb). Anti-FLAG antibodies were purchased from Cell Signaling Technology (DYKDDDDK Tag 9A3 #8146 Mouse Ab) or from Sigma Aldrich (FLAG Ab #F1804 Mouse M2 Clone). HA and V5 antibodies were purchased from Sigma Aldrich (HA #H3663 Mouse HA-7 Clone, V5 #V8012 Mouse V5-10 Clone). Mouse monoclonal His-H8 antibody was purchased from Abcam. Secondary HRP conjugated anti-rabbit and anti-mouse antibodies were purchased from Cell Signaling Technology (Anti-Mouse IgG HRP-linked Antibody #7076S, Anti-Rabbit IgG HRP-linked Antibody #7074S). Buffers, reagents, chemicals were purchased from Sigma Aldrich, Fisher, or VWR and were ACS certified reagent grade or better. Solvents were purchased from Fisher and were HPLC grade except for solvents used in mass spectrometry, which were LC/MS Grade. The M-270 Epoxy Dynabeads Co-IP Kit 14321D, reagents and protein G Dynabeads were purchased from Life Technologies. Protein A/G PLUS-agarose resin (SC-2003) was purchased from Santa Cruz Biotechnologies. AminoLink was purchased from ThermoScientific. Ni-NTA spin columns (300 mg protein binding capacity) were purchased from Qiagen. Spin columns were purchased from Pierce.

3.2.2 Cell culture and transfection

HEK293 cells were purchased from ATCC and maintained in DMEM/high glucose media with 10% fetal bovine serum and 1% L-glutamine- penicillin-streptomycin at 37 ºC in 5% CO_2. Cells

were seeded in a T75 mm^2 flask such that it will be 70%-90% confluent after 24 h of incubation. 15-20 mg of plasmid DNA and Lipofectamine ® 2000 (Invitrogen) at a ratio of 1:3 was used for transfection following the manufacturer's instruction. 500 mL of Opti-MEM medium was used to re-suspend both plasmid and Lipofectamine® 2000. Before transfection, the medium was changed to serum-free and antibiotic-free DMEM, DNA/lipid complex was added and further incubated for 24 h.

3.2.3 Cell lysate preparation

After transfection, cells were trypsinized, washed two times with Phosphate Buffered Saline (PBS), then lysed in 1 mL of 1X IP Buffer (M-270 Epoxy Dynabeads CoIP Kit 14321D, Life Technologies) to which we added 1X Halt Protease inhibitor cocktail (ThermoScientific) on ice for 30 min with occasional vortexing. The lysate was centrifuged at 13,000 rpm for 10 min at 4 ºC and transferred to the new LoBind Eppendorf tube. Total protein concentration was estimated using absorbance at A280. The cell lysate was precleared by incubation at 4 ºC for 2 h with 100 mL of protein A/G agarose beads.

3.2.4 Resin loading and binding capacity standardization

Resin loading capacity for resins was first determined based on product literature for each resin: AminoLink resin at 10 mg antibody/mg of resin, protein G Dynabeads at 8 mg antibody/mg resin, protein A/G resin at 10 mg antibody/mg of resin, M-270 epoxy Dynabeads at 5-7 mg antibody/mg of resin. The estimated resin loading was lowest for M-270 epoxy Dynabeads and protein G Dynabeads, we initially coupled 8 mg of antibody to each resin and the extent of coupling was determined by UV absorbance at 280 nm. Coupling efficiencies were measured at >60% for multiple reactions on both M-270 epoxy and protein G Dynabeads. This level of coupling was

consistent with product literature for M-270 epoxy Dynabeads, which had the lowest reported loading capacity for all tested resins. For all subsequent experiments, antibodies were coupled at 8 mg antibody/mg of beads. AP experiments were standardized using 3 mg of resin beads and 24 mg of antibody (crosslinked and non-crosslinked experiments). Subsequent absorbance measurements at A280 were consistent with preliminary experiments and we noted no significant deviations except in the complete failure of the protein A/G agarose crosslinking AP experiment. In several cases, we also checked antibody binding by Western Blot, but the high sensitivity of this technique made quantitation difficult for even low levels of unbound antibody where several micrograms antibody could remain in solution. To facilitate comparisons, protein concentrations for cell lysates were determined by UV absorbance at A280 and similar levels of protein were used in each experiment.

3.2.5 M-270 epoxy Dynabeads affinity purification procedure

The initial experiments using M-270 epoxy Dynabeads followed the standard procedure and utilized the reagents included in the M-270 epoxy Dynabeads Co-IP Kit 14321D. No modifications were made to this procedure. We noted decreased effectiveness for AP of the bait proteins after storage for 6-8 weeks. M-270 epoxy Dynabeads aliquoted into Eppendorf tubes and sealed under nitrogen exhibited minimal degradation over 8 weeks but have not been tested after extended storage.

3.2.6 Preparation of antibody-coupled solid-support resins

3.2.7 AminoLink resin (condition A)

AminoLink aldehyde-functionalized resin (100 mL resin) was washed 3 times with 300 mL PBS for each wash to remove storage buffer. The appropriate antibody (24 mg) was dissolved in extra

PBS to generate a 300 mL total volume. This antibody solution was added to the washed AminoLink resin and 3 mL of sodium cyanoborohydride ($NaCNBH_3$) solution (5 M $NaCNBH_3$ in 1 M NaOH) was added to generate a final $NaCNBH_3$ concentration of 50 mM. The reaction was mixed by end-over-end rocking for 1-2 h at room temperature. Vigorous shaking such as used for the mechanically stronger Dynabeads can result in a significant breakdown of the AminoLink resin into microparticles. Therefore, end-over-end rocking or gentle agitation is preferred. The supernatant was saved for analysis of antibody coupling efficiency and the antibody coupled resin was washed two times with 300 mL PBS. Remaining aldehyde functional groups were blocked by incubation with 300 mL of 1M TRIS with 3 mL of $NaCNBH_3$ solution. The reaction was mixed gently for 30 min by end-over-end rocking at room temperature before washing four times with 300 mL of 1 M NaCl.

3.2.8 Protein G Dynabeads and protein A/G agarose without BS^3 crosslinking (condition B)

Protein G Dynabeads slurry or protein A/G agarose beads (100 mL, 3.0 mg) were transferred into a LoBind Eppendorf tube. An isolation magnet was used for isolating Dynabeads and centrifugation at 1000 x g was used for isolation of protein A/G agarose beads. The beads were washed three times with PBST, then 24 mg of antibody diluted in 200 mL PBS was added at 37 ºC for 1 h with mild agitation. After incubation, the antibody-coupled resin was directly used for affinity purification.

3.2.9 Protein G Dynabeads and protein A/G agarose with BS^3 crosslinking (condition C)

Protein G Dynabeads slurry or protein A/G agarose beads were washed and antibody was coupled as described above without crosslinking. After antibody incubation, the antibody-coupled resin was washed 3 times with PBST and once with BS^3 conjugation buffer (20 mM sodium

phosphate, 0.15 M NaCl, pH 7-9). Crosslinking used 250 mL of BS^3 working solution (5 mM BS^3 in BS^3 conjugation buffer) and incubated at room temperature for 30 min with shaking at 1400 rpm. The crosslinking reaction was quenched by adding 12.5 mL BS^3 quenching buffer (1 M Tris HCl, pH 7.5) for 15 minutes. After crosslinking, the resin was washed three times with 200 mL of PBST.

3.2.10 Affinity purification

All of the resin preparation conditions (A, B and C above) utilized this same affinity purification procedure. The antibody-coupled resin was equilibrated by 2 PBS washes followed by two washes with 1:1 PBS/1x IP buffer and once with undiluted 1x IP buffer (300 mL/wash). The precleared cell lysate (200-500 mg total protein) dissolved in 1x IP buffer was incubated with antibody-coupled resin at 4 ºC with gentle end-over-end rocking. The resin was washed with 300 mL 1:1 PBS/1xIP buffer, 300 mL PBS, 300 mL 1M NaCl, and 300 mL HPLC-grade water. The resin was transferred to a new LoBind Eppendorf tube and bound proteins were eluted by two 5 min washes of 50 mL of 5% NH_4OH in HPLC water. A wash with 300 mL water followed by elution in 100 mM glycine buffer, pH 3, was attempted to determine if all protein was eluted under basic conditions. A sample of resin was treated with LDS at 95 ºC to assess if any residual protein remains bound to the resin after elution. Control AP followed the same procedure without the use of an antibody.

3.2.11 Western blot

Western blots were performed using standard procedures for the Invitrogen Xcell SureLock Electrophoresis System. LDS sample buffer, running buffers, and transfer buffers were purchased from Invitrogen. Electrophoresis was run on Novex 12% or 4-20% Precast Bis-Tris gels in either

1x MOPS or MES running buffers at voltages between 125 and 175 V and run times of between 30 and 90 min. Proteins were transferred from gels to PVDF membranes using the standard Xcell Blot Module at 30 V for 1 h. All western blots used 5% dry non-fat milk for blocking and secondary antibody incubation, 5% BSA for primary antibody incubation, and ECL-pico chemiluminescent reagent for detection.

3.2.12 Sample preparation for mass spectrometry

Protein samples after lyophilization were reconstituted in 10-20 mL of a mixture of 25-50% acetonitrile with 100 mM ammonium bicarbonate, pH 7.2 based on protein solubility. The reconstituted AP protein sample was reduced for 15 min at 37 ºC using 50 mM TCEP in ammonium bicarbonate, pH 7.2. Alkylation was performed for 30 min at room temperature without light exposure using 50 mM iodoacetamide in ammonium bicarbonate, pH 7.2. The protein sample was diluted to 300 mL with HPLC-grade water and the proteins were precipitated by chloroform-methanol. Precipitated proteins were reconstituted in 80 mL of 100 mM ammonium bicarbonate and digested using 10 mL of a 20 ng/mL stock solution of sequencing-grade trypsin (200 ng, Promega) for 16-24 h at 37 ºC. The digest reaction was quenched by addition of 10 mL of 10% formic acid and the sample was lyophilized to remove ammonium bicarbonate prior to reconstitution in 20 mL of 1% formic acid for analysis by mass spectrometry with analysis in triplicate.

3.2.13 Liquid chromatography-tandem mass spectrometry

Liquid chromatography-tandem mass spectrometry was performed on either a high-resolution Bruker MicrOTOF-Q (Q-TOF) or a Thermo Finnigan LTQ Linear Ion Trap mass spectrometer. MicrOTOF-Q: The Bruker Q-TOF was equipped with an Agilent 1200 Nano HPLC system using

Solvent A, water with 0.25% formic acid and Solvent B, acetonitrile with 0.25% formic acid. Gradient: 5% Solvent B to 30% Solvent B over 85 min at a column flow rate of 1.4 mL/min on a home-packed 75 µm x 100 mm nanobore C18 column, Bruker captive spray source voltage at 1400 V, dry gas at 3.0 L/min, and capillary temperature at 150 ºC. Data was collected in data-dependent mode fragmentation (MS/MS) on the three most intense spectra identified in MS mode. Data was output to MGF file format using the Compass Data Analysis software. LTQ Ion Trap: The Thermo Finnigan LTQ Ion Trap mass spectrometer is equipped with an Agilent 1290 Infinity UPLC system using Solvent A, water and 0.1% formic acid and Solvent B, methanol and 0.1% formic acid. The gradient was 20% Solvent B to 100% Solvent B over 60 min at a flow rate of 40 mL/min on a home-packed 500 µm x 6 mm C18 reversed-phase column. Electrospray voltage was set to 3.78 kV with sweep, auxiliary, and sheath gas set to 0 on a standard IonMax ESI Source (Thermo Finnigan). The capillary temperature was set to 250 ºC and the mass spectrometer was set for Dynamic Exclusion Data Dependent MS/MS with the three highest intensity masses observed in the MS scan targeted for MS/MS fragmentation. The RAW data files were transformed to either mzXML or MGF format using the MSConvert utility program from the ProteoWizard program suite (http://proteowizard.sourceforge.net/tools.shtml). Spectra from the HPLC-MS/MS run were output into MGF format for analysis in X!Tandem (Current GPM 2013.09.07). Each database search used trypsin digestion and default Q-TOF and Ion Trap methods.

3.3. Results and discussion

3.3.1 AP-MS for N-FLAG-p53 on protein G Dynabeads and protein A/G agarose without crosslinking

The tumor suppressor protein p53 is known to have regulatory roles in the suppression of cancer and to have diverse protein interactions (Joerger and Fersht, 2016). Therefore, we used an N-terminal FLAG-tagged p53 (N-FLAG-p53) for initial AP studies. AP for N-FLAG-p53 was successful using both protein G Dynabeads and protein A/G agarose without covalent cross-linking (Figure 3.4). Protein G Dynabeads and protein A/G agarose both depleted essentially all N-FLAG-p53 in (Figure 3.4, Lanes 3 and 7).

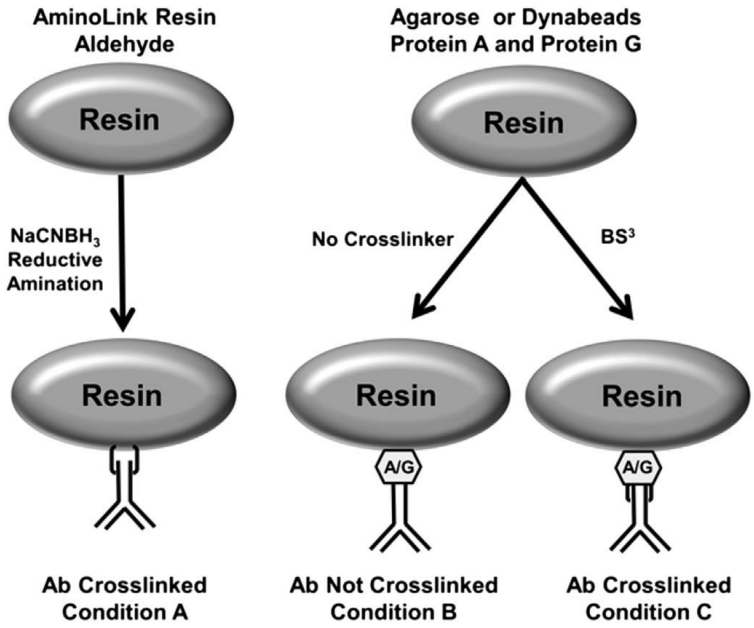


Figure 3.1 Schematic representation of AminoLink (left), agarose, and Dynabead (right) resins and crosslinking conditions. AminoLink resin uses reductive amination between resin-linked aldehyde and antibody amine functional groups to immobilize antibodies (Condition A). Agarose and Dynabeads were covalently linked to protein A and protein G, which enables non-covalent attachment to antibodies (Condition B) or the addition of crosslinkers including BS^3 for covalent attachment through protein A and G (Condition C). The antibody-coupled resins were used for affinity purification of either FLAG-tagged p53 or HA-tagged PTEN proteins under identical conditions. Reprinted from "Observations on different resin strategies for affinity purification mass spectrometry of a tagged protein," by S. Mali, M. Mitchell, W. Widger, S.J. Bark, 2016, *Analytical Biochemistry*, *515,* 26-32. Copyright [2016] by Mali et al.

However, significant differences in resin behavior were observed. Elution from Dynabeads gave a single band and elution from protein A/G agarose two bands, the higher molecular weight band being approximately twice the expected molecular weight for monomeric N-FLAG-p53. After loading LDS treated resin samples from both the Dynabeads and protein A/G agarose experiments, an obvious p53-intense band was observed only in the well region of the gel (Figure 3.4, Lane 6), suggesting an irreversible binding of the N-FLAG-p53 to Dynabeads. This irreversible binding was observed for crosslinked and non-crosslinked Dynabeads on several occasions and protein A/G agarose on one occasion after using dilution with PBST to reduce detergent concentration. IP Buffer for dilution reduced but did not eliminate, this irreversible binding. As expected, the non-crosslinked experiments showed antibody elution with N-FLAG-p53 (Figure 3.4).

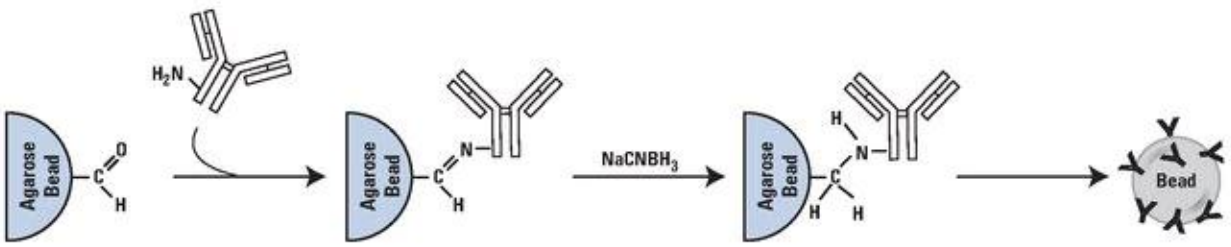


Figure 3.2 Schematic of reductive amination on AminoLink resin. Protein primary amines on antibody react spontaneously with aldehyde groups on resin resulting in Schiff-base bonds. Sodium cyanoborohydride reduces Schiff base to stable secondary amine bond. Many aldehyde groups per bead and several amines per antibody result in many antibody molecules per bead. Adapted from "AminoLink™ Coupling Resin" by ThermoFisher Scientific, 2020, https://www.thermofisher.com/order/catalog/product/20381#/20381. Copyright [2020] by ThermoFisher Scientific.

3.3.2 AP-MS for N-FLAG-p53 on protein G Dynabeads and protein A/G agarose with crosslinking

Crosslinking strategies have been used extensively for protein structural studies. For AP-MS, crosslinking the antibody to the resin beads through protein A or protein G prevents antibody contamination of the eluted protein. However, AP of N-FLAG-p53 using BS^3 crosslinking reversed the effectiveness of AP in a resin-dependent manner (Figure 3.4). High levels of N-FLAG-p53 were observed in the unbound supernatant, which contrasts with the N-FLAG-p53 AP without crosslinking. Protein G Dynabeads isolated the N-FLAG-p53 protein, but with an increased quantity of the higher molecular weight band compared to non-crosslinked protein G Dynabeads. Protein A/G agarose failed, yielding only N-FLAG-p53 in the unbound lysate fraction. Reprobing with anti-mouse antibody indicated that protein G Dynabeads retained high levels of antibody while protein A/G agarose did not (Figure 3.4). We also attempted a more specific two-step crosslinking strategy by first attaching BS^3 to the resin followed by several washes to remove excess cross-linking agent before antibody attachment. These approaches were not successful. Homobifunctional crosslinkers such as BS^3 crosslink proteins based on the accessibility of lysine sidechains and have been observed to negatively impact antibody activity (Trahan and Oeffinger, 2016). It was unexpected that AP-MS performance upon crosslinking would vary by resin type. No crosslinking suggested that equal or better performance for protein A/G agarose beads compared to protein G Dynabeads (Figure 3.4). However, after crosslinking, protein G Dynabeads exhibited better performance without a significant impact on the activity of the antibody, unlike the negative effect we observed with protein A/G agarose (Figure 3.4). This differential effect on AP-MS between protein A/G agarose and protein G Dynabeads cannot be explained by crosslinking because we used the same antibody, buffers, and conditions for all experiments and the reactivity of BS^3 should negatively impact the binding site of the antibody consistently across all

experiments. These data suggest that different resins may impact the success of the AP-MS experiment.

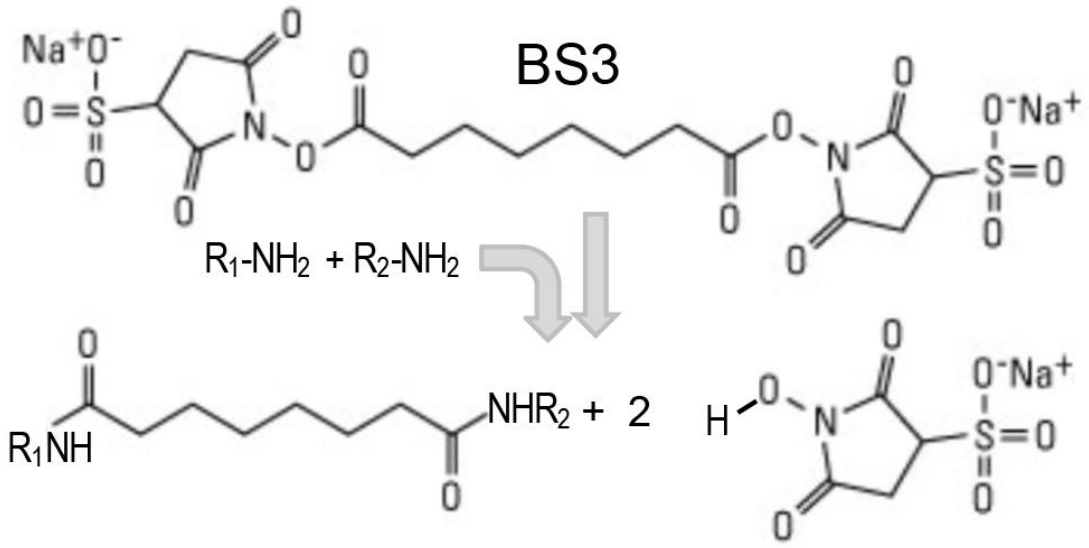


Figure 3.3 Schematic of crosslinking reaction on AminoLink resin. Homobifunctional crosslinkers such as BS^3 crosslink proteins based on the accessibility of lysine sidechains and have been observed to negatively impact antibody activity. Adapted from "BS3 (bis(sulfosuccinimidyl)suberate)," by ThermoFisher Scientific, 2020, https://www.thermofisher.com/order/catalog/product/21580#/21580. Copyright [2020] by ThermoFisher Scientific.

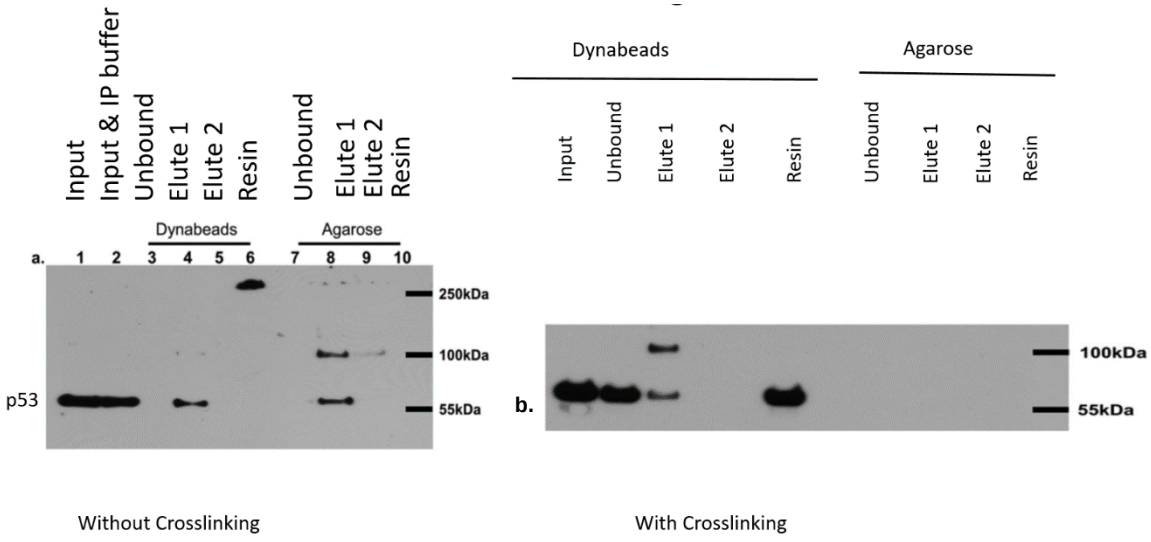


Figure 3.4 Affinity purification western blot for overexpressed FLAG-p53 in HEK293 cells. Without crosslinking (a) and with crosslinking (b) of M2 clone anti-FLAG antibody to protein G Dynabeads or protein A/G agarose beads. Reprinted from "Observations on different resin strategies for affinity purification mass spectrometry of a tagged protein," by S. Mali, M. Mitchell, W. Widger, S.J. Bark, 2016, *Analytical Biochemistry*, *515,* 26-32. Copyright [2016] by Elsevier Inc.

3.3.3 Improved AP-MS for N-FLAG-p53 on AminoLink resin

Epoxy-functionalized Dynabeads were initially used for covalent linking of antibody to the resin, but this resin degraded over time, likely because of epoxide hydrolysis. Fresh aliquots of resin samples were sealed in Eppendorf microcentrifuge tubes to limit contact with air and reduce this degradation. However, we desired a more stable resin system to immobilize antibodies that could be tailored to different antibodies and experimental designs without hydrolysis. AminoLink resin satisfied these requirements because of the Schiff's base formation between aldehyde antibody amine functional groups (Ren *et al.,* 2003). This reversible reaction allows multiple different linking conformations for the antibody and reduction with sodium cyanoborohydride is tolerant of water and results in irreversible linking of the antibody to the resin.

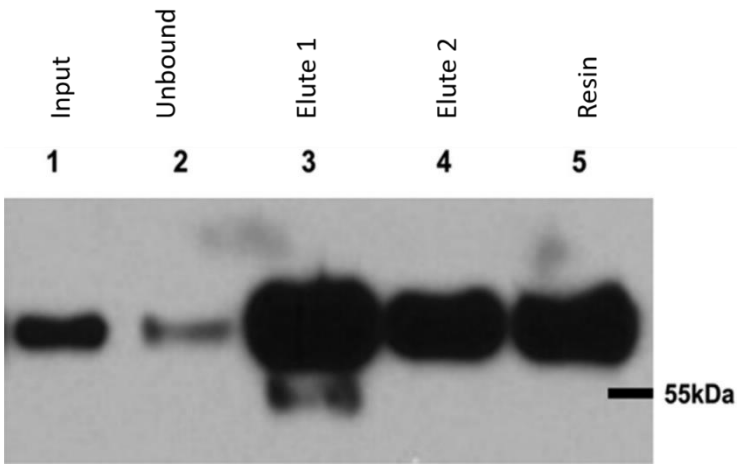


Figure 3.5 Affinity purification western blot for overexpressed N-FLAG-p53 in HEK293 cells using AminoLink resin. M2 Clone Anti-FLAG antibody was immobilized on AminoLink resin and used for affinity purification of FLAG-P53 protein expressed in HEK293 cells while western blot used anti-P53 antibody for detection. The western blot demonstrated high levels of P53 in the cell lysate (Lane 1) that was depleted upon incubation with anti-FLAG AminoLink resin (Lane 2). After washes, elution under 5% NH_4OH efficiently eluted the bound FLAG-P53 (Lane 3). Because further elution using acidic 100 mM glycine, pH 3.0 exhibited no further elution of FLAG-P53 protein, a second base elution was performed and eluted further FLAG-P53 immobilized on the resin (Lane 4). FLAG-P53 protein remaining on AminoLink resin was eluted at 95 ºC in LDS buffer and demonstrated non-specific resin binding (Lane 5). Reprinted from "Observations on different resin strategies for affinity purification mass spectrometry of a tagged protein," by S. Mali, M. Mitchell, W. Widger, S.J. Bark, 2016, *Analytical Biochemistry*, *515,* 26-32. Copyright [2016] by Elsevier Inc.

Some of these linking conformations will not be effective for binding the antigen, but the majority of linking conformations should orient the antibodies reasonably for antigen binding. AP of pre-cleared HEK293 cell lysate overexpressing N-FLAG-p53 with M2 Antibody-linked AminoLink resin exhibited a decrease of N-FLAG-p53 in the supernatant after 16 h binding at 4 ºC (Figure 3.5). Elution under basic conditions showed a robust signal for specifically bound N-FLAG-p53 protein both in the first and second elution steps. Elution of the remaining protein using LDS buffer at 95 ºC showed that some N-FLAG-p53 remained on the resin after base elution. This resin-bound N-

FLAG-p53 migrated at the expected molecular weight and was solubilized by LDS. In contrast, AP of this protein using Dynabeads or protein A/G agarose without optimized binding conditions resulted in variable and sometimes irreversible binding to the resin. We also observed multiple bands with one approximately twice the expected molecular weight of NFLAG-p53, suggesting modification or dimerization after AP (Figure 3.4, Lane 8 and Figure 3.5, Lane 3). Minimal "bleed off" of the anti-FLAG antibody from the resin was observed only after LDS treatment. Mass spectrometry analysis successfully identified p53 (TP53) in these experiments in X!Tandem based on the highest log of expectation I value (Table 3.6).

3.3.4 AP-MS for N-HA-PTEN on protein G Dynabeads, protein A/G agarose, and AminoLink solid-support resins

The experiments for N-FLAG-p53 suggested a question of the generality of our results. Another widely mutated tumor suppressor protein is phosphatase and tensin homolog (PTEN). We constructed the coding sequence of PTEN with an N-terminal HA-tag sequence, an N-terminal V5 tag sequence, and the N-terminal and C-terminal FLAG tag sequence in pGCN, pcDNA3, or pRK5 vectors. Transient transfection of the C-FLAG-PTEN and N-HA-PTEN protein in HEK293 cells was highly effective as evidenced by western blot, but N-V5-PTEN did not express under these conditions and N-FLAG-PTEN was expressed without the tag. Comparisons for affinity purification using tagged PTEN constructs used the optimized procedure described for N-FLAG-p53 on protein A/G agarose, protein G Dynabeads, or AminoLink resin. The unbound fractions contained PTEN in both C-FLAG-PTEN and N-HA-PTEN experiments, but not for N-V5-PTEN (Figure 3.4, Lanes 1-3). Base elution readily liberated the N-HA-PTEN and mass spectrometry identified PTEN as a major component of the protein eluent (Table 3.6). The mass spectrometry data supplements the western blot data.

Table 3.6 Affinity purification-mass spectrometry data summary for N-HA-PTEN experiments using different resins

N-HA-PTEN										
Rank	Protein	Unique	Total	log(e)	%/%	FPR%	MS	Resin	Crosslink	Condition
2	PTEN	14	18	−133.6	42/70	0.85	QTOF	AminoLink	+	A
2	PTEN	15	21	−142.3	34/57	0.85	QTOF	AminoLink	+	A
2	PTEN	17	22	−161.5	43/72	0.76	QTOF	AminoLink	+	A
2	PTEN	6	7	−58.2	15/25	1.69	LTQ	Dynabeads	−	B
2	PTEN	5	6	−52.3	12/20	2.22	LTQ	Dynabeads	−	B
2	PTEN	5	6	−48.5	12/20	2.15	LTQ	Dynabeads	−	B
2	PTEN	6	7	−63.5	14/23	1.29	LTQ	Agarose	−	B
2	PTEN	6½	7	−63.8	14/23	1.21	LTQ	Agarose	−	B
2	PTEN	4	4	−37.7	11/19	1.15	LTQ	Agarose	−	B
5	PTEN	3	4	−24.4	7.4/12	1.52	LTQ	Dynabeads	+	C
6	PTEN	3	4	−28.4	8.4/14	1.34	LTQ	Dynabeads	+	C
4	PTEN	3	3	−18.4	6.5/11	1.78	LTQ	Dynabeads	+	C
9	PTEN	4	5	−40.7	8.4/14	0.93	LTQ	Agarose	+	C
20	PTEN	3	4	−20.2	5.7/10	1.4	LTQ	Agarose	+	C
11	PTEN	4	5	−32.4	10/17	1.75	LTQ	Agarose	+	C

PTEN protein was affinity purified using AminoLink resin (Condition A), protein G Dynabeads or protein A/G agarose without crosslinking (Condition B), or protein G Dynabeads or protein A/G agarose with crosslinking (Condition C). The PTEN protein (Protein) identification was statistically ranked with the top hit ranked 1 (Rank). Unique and total are the number of unique spectra and total redundant spectra annotated to PTEN. Log(e) is the log expectation value and %/% refers total percent sequence identified and percent sequence identified corrected for peptides unlikely to be identified by tryptic digest. FPR% = false positive rate, MS= instrument used for the experiment, QTOF= MicroTOF-Q and LTQ = Ion trap. Resin = resin used, Crosslink = + for yes, - for no, Condition = resin conditions used for AP-MS experiment, A = AminoLink with crosslinking, B = protein A/G agarose or protein G Dynabeads without BS^3 crosslinking, C = protein A/G agarose or protein G Dynabeads with BS^3 crosslinking. Reprinted from "Observations on different resin strategies for affinity purification mass spectrometry of a tagged protein," by S. Mali, M. Mitchell, W. Widger, S.J. Bark, 2016, *Analytical Biochemistry*, *515,* 26-32. Copyright [2016] by Elsevier Inc.

Protein identification was equally achieved in both non-crosslinked and crosslinked protein A/G agarose and protein G Dynabeads, which contrasts with NFLAG-p53. The significantly improved data from the AminoLink PTEN AP-MS experiment was unexpected because N-FLAG-p53 experiments showed a smaller advantage for this resin. These data suggest that the specific tag

and protein combination impacts the quality of the AP experimental data as previous reports suggest.

3.4 Limitations

These studies evaluated several solid-support matrices in AP-MS of tagged proteins. We used low quantities of resin and antibody sufficient for the identification of the target proteins, but the data were insufficient for the rigorous identification of protein interactions. We observed high levels of BSA in the N-HA-PTEN experiments, which we attribute to the stabilization of the anti-HA antibody with BSA. This problem was not observed in the FLAG-tagged AP-MS experiments. We also observed that the level of tagged-protein expression correlated directly to the quality of AP-MS as has been noted previously. Compared to N-HA-PTEN and C-FLAG-PTEN, the N-FLAG-p53 protein was expressed at a lower level after transfection and exhibited weaker AP. We also observed the N-terminal HA worked better than C-terminal FLAG, despite similar levels of expression. Some mass spectrometry data was acquired on a MicroTOF-Q instrument rather than the LTQ ion trap instrument. We evaluated potential data quality differences indicating that the LTQ is more sensitive under our experimental conditions. Therefore, the higher unique and total peptides annotated to PTEN in Table 3.6 (Rows 1-3) cannot be attributed to the Q-TOF instrument alone, but the lower FPR% is directly correlated to the higher mass accuracy of the Q-TOF instrument. Another limitation of this study is the qualitative nature of the results presented. While these results are consistent and reproducible, we caution that quantitative comparisons using these types of experiments are difficult. These limitations cannot account for nor counter our observation that AminoLink resin performed better than other tested resins for recovery of the tagged proteins under our tested conditions.

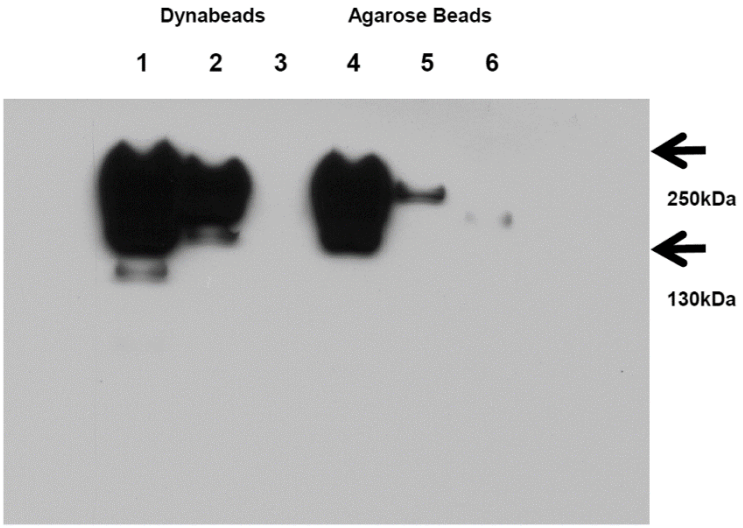


Figure 3.7 Analysis of antibody contamination from protein G Dynabeads (Lanes 2-5) and protein A/G agarose (Lanes 6-9) without crosslinking. Antibody levels in unbound, elution and resin fractions were determined using an anti-mouse HRP-linked secondary antibody for western blot. Input cell lysate has no added antibody (Lane 1). Unbound supernatant after incubation with antibody linked protein G Dynabeads indicates some "bleed off" of antibody (Lane 2). Base elution releases significant quantities of antibody (Lane 3). A second base elution step (Lane 4) and resin analysis (Lane 5) indicate the majority of antibody was eluted in the first base elution step. The equivalent experiment for protein A/G agarose beads demonstrates the same basic profile. Unbound supernatant after incubation with antibody linked protein A/G agarose beads indicates lower "bleed off" of antibody from resin than observed for Dynabeads (Lane 6). Base elution steps release copious quantities of antibody in both the first elution (Lane 7) and second elution steps (Lane 8). Significant antibody remains bound to the protein A/G agarose resin even after base elution (Lane 9). Reprinted from "Observations on different resin strategies for affinity purification mass spectrometry of a tagged protein," by S. Mali, M. Mitchell, W. Widger, S.J. Bark, 2016, *Analytical Biochemistry*, *515,* 26-32. Copyright [2016] by Elsevier Inc.

3.5 Conclusions

Our studies suggest that the matrix used to immobilize the antibody is an important component for the success of AP-MS. We evaluated multiple resin types including epoxy Dynabeads, protein G Dynabeads, protein A/G agarose, and AminoLink resin. While all resins performed effectively under appropriate conditions, non-optimal conditions produced high variability and non-specific protein binding in the AP-MS experiments. Variability was also observed arising from the tag and

protein expression, buffers, reagents, and conditions for binding, washes, and elution as previously observed in other laboratories. Our data suggest that, for affinity purification, AminoLink resin required less optimization for effective recovery of tagged proteins. Therefore, this resin may be useful for initial AP-MS trials when no other information is available. Future experiments will evaluate the performance of these resins for CoAP-MS starting from the procedures developed in this study.

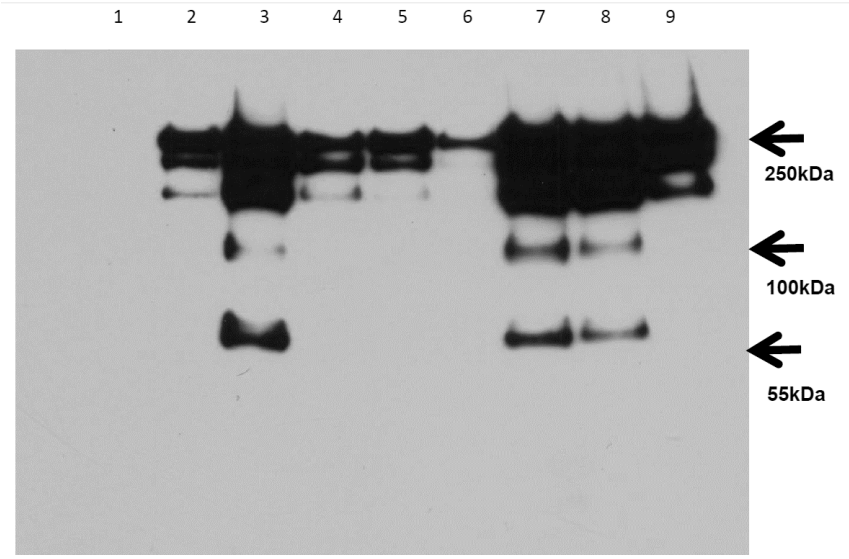


Figure 3.8 Comparison of M2 anti-FLAG antibody binding to protein G Dynabeads and protein A/G agarose beads during BS^3 crosslinking experiment. Failure of AP for N-FLAG-p53 using protein A/G agarose beads suggested problems with antibody binding or crosslinking. This was confirmed using an anti-mouse HRP-linked antibody for western blot against the M2 clone anti-FLAG antibody used for AP. These data show essentially equivalent levels of anti-FLAG antibody before crosslinking to protein G Dynabeads (Lane 1) and protein A/G agarose beads (Lane 4). However, after washes to remove unbound antibody, elution analysis suggests Higher levels of antibody bound to the protein G Dynabeads (Lane 2) versus protein A/G agarose beads (Lane 5). In each case, essentially no antibody elution occurred after crosslinking with BS3 (Lanes 3 and 6). Reprinted from "Observations on different resin strategies for affinity purification mass spectrometry of a tagged protein," by S. Mali, M. Mitchell, W. Widger, S.J. Bark, 2016, *Analytical Biochemistry*, *515,* 26-32. Copyright [2016] by Elsevier Inc.

Table 3.9 Mass spectrometry data for identification of N-FLAG-p53 and C-FLAG-PTEN proteins by AP-MS

N-FLAG-p53										
Rank	**Protein**	**Unique**	**Total**	**log(e)**	**%/%**	**FPR%**	**MS**	**Resin**	**Crosslink**	**Condition**
1	TP53	8	9	-64.3	22/36	0.83	QTOF	AminoLink	+	A
1	TP53	7	10	-59.3	20/33	0.62	QTOF	AminoLink	+	A
1	TP53	6	8	-55.2	20/32	1.01	LTQ	AminoLink	+	A
C-FLAG-PTEN										
6	PTEN	5	6	-51.2	13/21	0.55	QTOF	AminoLink	+	A
9	PTEN	5½	5½	-47	14/23	0.067	QTOF	AminoLink	+	A

Tagged proteins were affinity purified using an optimized procedure with crosslinking for AminoLink resin (condition A as described in materials and methods). Rank is the numerical ranking of the protein identification (i.e. rank 2 is the second-highest ranked identification in the data set), log(e) is the log of the expectation value for the protein identification, **%/%** is the percentage sequence coverage for peptides identified before and after correction for sequences unlikely to be observed, unique and total denote the number of unique peptides identified for the protein identification and the total number of spectral counts for those peptide identifications, FPR% is the false positive rate calculated from X!Tandem on the data set, MS is the mass spectrometry platform used for the analysis of the sample (note that there are no appreciable differences between these mass spectrometers for protein identification purposes), resin defines which resin was used for the AP-MS experiment, crosslink indicates if crosslinking was used for antibody immobilization. Condition = resin conditions used for AP-MS experiment, A = AminoLink with crosslinking, which was used for all experiments in this table and are defined in materials and methods. Reprinted from “Observations on different resin strategies for affinity purification mass spectrometry of a tagged protein,” by S. Mali, M. Mitchell, W. Widger, S.J. Bark, 2016, *Analytical Biochemistry*, *515,* 26-32. Copyright [2016] by Elsevier Inc.

Table 3.10 Primer, sequence, and vector information for tagged protein constructs

Primer	Sequence	Vector
PTEN_N-HA_F	GCCA *TCTAGA* **atg**acagccatcatcaaagag	pCGN
PTEN_N-HA_R	GTGT *GGATCC* **tca**gacttttgtaatttgtgtatgc	pCGN
PTEN_C-FLAG_F	GAAT *GAATTC* ccacc**atg**acagccatcatcaaagag	pRK5
PTEN_C-FLAG_R	GTAT *TCTAGA* gacttttgtaatttgtgtatgctgatcttc	pRK5
PTEN_N-FLAG_F	GAAT *GGATCC* acagccatcatcaaagag	pcDNA3
PTEN_N-FLAG_R	GGAG *TCTAGA* ccagtgtgatggatatctgc	pcDNA3
PTEN_N-V5_F	TTAT *AAGCTT* GGTACCGAC**atg**ggtaagcctatccctaaccc tctcctcggtctcgattctacg *GGATCC* **atg**acagccatcatc	pcDNA3
PTEN_N-V5_R	GGAG *TCTAGA* ccagtgtgatggatatctgc	pcDNA3

The tagged proteins were generated using forward and reverse primers for PCR amplification and cloning into the appropriate vectors. Restriction enzyme sites are denoted with underline and italics. Priming sites are indicated in lower case letters. Reprinted from “Observations on different resin strategies for affinity purification mass spectrometry of a tagged protein,” by S. Mali, M. Mitchell, W. Widger, S.J. Bark, 2016, *Analytical Biochemistry*, *515,* 26-32. Copyright [2016] by Elsevier Inc.sol

Chapter 4:

Identification of Bacterial Dormancy Signature Proteins in *Micrococcus luteus* Using Mass Spectrometry

Preface

Work outlined in this chapter has been published in the Journal of Bacteriology (Mali S, **Mitchell M**, Havis S, Bodunrin A, Rangel J, Olson G, Widger WR, Bark SJ. 2017. A proteomic signature of dormancy in the Actinobacterium *Micrococcus luteus*. *J Bacteriol* 199(14):206-217.)

4.1 Introduction

Dormancy is an important stress response in which bacteria curtail metabolic activity to survive potentially fatal external stresses. Dormancy has broad implications for many pathogenic bacteria, including *Mycobacterium tuberculosis*, *Staphylococcus aureus*, *Treponema pallidum* (syphilis), and *Borrelia burgdorferi* (Lyme disease) (Lewis, 2007). Additionally, stress responses appear to have importance in bacterial life processes like quorum sensing and biofilm formation (Harms *et al.,* 2016). The recognized importance of dormancy does not provide a simple definition for this stress state. Recent evidence suggests that dormancy consists of a continuum that spans bacterial stress responses, including the recognized viable but nonculturable (VBNC) and persistence states (Ayrapetyan *et al.,* 2015). Persister cells have been identified as a “small subpopulation of cells that spontaneously enter a dormant, nondividing state”, a definition based on the survival of bacterial cell populations after antibiotic treatment. A complementary definition

for VBNC is living cells that have lost the ability to grow on routine media (Li *et al.*, 2014). This VBNC definition is based on evidence that these cells are metabolically active, respiration competent, and capable of incorporating amino acids into proteins (Maisonneuve *et al.,* 2014). However, these definitions have caveats for defining actual dormancy states. Differentiating and isolating VBNC and persister cells is difficult at best because both states exhibit enhanced survival under antibiotic treatment. It is also likely that these states coexist and stochastically arise in any dormant cell population (Keren *et al.,* 2004). There is general agreement that VBNC and persistence states are interrelated, have reduced metabolic activity, and are important for antibiotic tolerance, latent infection, and the reemergence of active infections after the resolution of stress (Gengenbacher and Kaufmann, 2012). From these considerations and for this article, we propose to describe persister cells as proliferation-competent cells that can spontaneously resuscitate after the resolution of stress. VBNC cells are induced into a nonproliferative state by external stress and will not resuscitate spontaneously until an appropriate external signal is provided. These definitions were chosen to specifically address the experimental system described in this article, but they are consistent with observations in the field and provide a viable starting point for experiments and discussion. Dormancy has been implicated in persistence and VBNC stress responses in chronic infection, latency, and resuscitation of active infections (Barry *et al.,* 2009). Targeting resuscitation has been a primary focus by the World Health Organization (WHO) as an avenue for the treatment of widespread tuberculosis infection (Lipworth *et al.,* 2016). A link between latency and dormancy was suggested by Wayne from results showing initiation of

dormancy in oxygen-poor (hypoxic) tissue environments, an observation leading to an *in vitro* hypoxia *M. tuberculosis* model of dormancy. The actual state and underlying mechanisms of tuberculosis latency are not well understood, and diverse definitions exist for latent and dormant phenotypic states (Besnard *et al.,* 2002). However, other bacteria, including *Mycobacterium bovis*, *Mycobacterium smegmatis*, *Escherichia coli*, *Listeria monocytogenes*, and *Micrococcus luteus*, can exhibit VBNC, persistence, and other dormancy stress responses under nutrient deprivation and long-term antibiotic treatment (Zhang, 2004). These observations are not unique, and several recent reviews compiling data from many laboratories using different bacterial strains and conditions suggest that different stresses may not generate equivalent dormancy states (Leung and Lévesque, 2012). There is even a diversity of specific genetically encoded pathways that are critical for the regulation of persistence, including quorum sensing, (p)ppGpp, toxin-antitoxin systems, and even the age of the inoculation (Gaca *et al.,* 2005). These data suggest diverse stress response states that are dependent on diverse sets of proteins. We note that drawing general conclusions from combining unrelated studies using different bacteria and experimental conditions should be approached with caution. However, this level of complexity presents experimental challenges, and our definitions may require amendments in different bacterial systems. Rather than focusing on diversity, we seek to determine if there is a common core of stress response proteins across different dormancy states and conditions. A conserved dormancy-related set of proteins has remarkable implications. These protein data would be complementary to and enhance genetic and transcriptomic profiles, because the relationship

between transcripts and expressed proteins may not be directly comparable in a metabolically reduced state. These proteins would be a signature for the dormant state and facilitate diagnostics for latent bacterial infections. The biological functions of these proteins could define baseline mechanisms that drive the dormancy stress response. Extension to structural studies and protein-protein interactions can provide viable targets for drug development against bacteria exhibiting reduced metabolic activity, which is a limitation of current antibiotics. There is even the potential for the determination of the origin of persister cells, i.e., a protein that is essential for maintaining the dormant state (Page and Peti, 2016). These considerations suggested the importance of identifying conserved dormancy-related proteins. Here, we report using *Micrococcus luteus* NCTC 2665 (MI-2665) as a nonpathogenic actinobacterial model to identify VBNC stress response proteins. MI-2665 has unique advantages for these studies, including rapid growth (doubling time of ~40 min compared to over 24 h for *M. tuberculosis*), a small, 2.5-Mb genome comprising a “minimal set” of gene products and proteins necessary for dormancy, and homology with pathogenic *M. tuberculosis*, *M. bovis*, and *Mycobacterium leprae*. Importantly, MI-2665 exhibits a well-defined and reproducible VBNC state under nutrient deprivation (Wayne and Lin, 1982). Finally, the resuscitation-promoting factor (Rpf) protein from MI-2665 can induce resuscitation from the VBNC state in *M. tuberculosis* (Zhu *et al.,* 2003). These reports suggest that mechanisms driving VBNC and, potentially, other dormancy states may be conserved among actinobacteria and that MI-2665 has a uniquely tractable VBNC state for study. An additional consideration was that MI-2665 VBNC bacteria do not resuscitate spontaneously, which

circumvents the potential complication presented by persister cells combined with VBNC cells in the biological sample, based on our definition. We note that many other bacterial systems can resuscitate spontaneously after the removal of external stress, but it is unclear to us whether the restored viability is from VBNC or persistence cell states. We used liquid chromatography-tandem mass spectrometry (LC-MS/MS) to quantitatively identify proteins in MI-2665 grown in logarithmic-growth and dormant phases, using nutrient deprivation to generate a physiologically relevant and long-term dormancy state (Kaprelyants and Kell, 1993). We have identified 18 proteins that are differentially increased in dormant MI-2665. Given the caveats for generalization of our data identified above, we noted that four tuberculosis latency-related proteins, including universal stress protein A (UspA) and proteins involved in the anaplerotic glyoxylate pathway, were quantitatively increased in the MI-2665 VBNC state (Wayne and Lin, 1982). The other 14 proteins we observed to be quantitatively increased in the MI-2665 VBNC state have not been previously reported but are conserved across many, if not a majority, of bacterial species and correlate to known important amino acid metabolism and ribosomal regulation pathways. These proteins present a proteomic signature for the VBNC state in MI-2665 that complements other transcriptomic studies (Young *et al.,* 2010). Our data support the idea that conserved proteins and pathways could regulate dormancy across actinobacteria and, potentially, across diverse bacterial species. While it is not possible at this time to conclusively generalize our results, the conservation of the proteins identified in these data warrants further exploration. The unique properties of MI-2665 present a relevant and tractable bacterial system for these future studies.

4.2 Materials and methods

Buffers, reagents, chemicals were purchased from Sigma Aldrich, Fisher, or VWR and were ACS certified reagent grade or better. Solvents were purchased from Fisher and were HPLC grade except for solvents used in mass spectrometry, which were LC/MS or Fisher Optima Grade. Halt protease inhibitors were purchased from Thermo Scientific. Trypsin was purchased from Promega or Worthington Biochemicals. *Micrococcus luteus* NCTC 2665 was purchased from ATCC (www.atcc.org) and stored as glycerol stocks at 80 ºC until plating and culture experiments.

4.2.1 Exponential phase and dormant cultures of *M. luteus* MI-2665 on acetate minimal medium

Micrococcus luteus NCTC 2665 was plated from a glycerol stock onto a rich medium plate. A single colony was used to inoculate 5 ml of rich medium broth starter culture. After incubating the culture at 30 °C for 5 h (optical density at 660 nm [OD660] of between 0.3 and 0.6) with shaking (265 rpm), 2 ml of starter culture was inoculated into 500 ml of acetate minimal medium (AMM) and incubated at 30 °C with shaking. The bacterial growth was monitored by OD660 from time point zero until the growth curve entered late stationary phase. The samples were collected every 2 h for OD measurements. When the bacterial growth was in logarithmic (log) phase (i.e., OD660 of 0.5 to 0.9), 40 ml of culture was harvested and collected by centrifugation at 10,000 x *g* for 10 min at 4 °C. The cell pellet was resuspended in 1 ml of Tris-EDTA (TE) buffer (100 mM Tris-HCl and 10 mM EDTA, pH 8) and used immediately or stored at -80 °C. To collect dormant *M. luteus* samples, the culture was incubated at 30 °C with agitation for 2 months and then kept at room temperature without agitation for up to 1 month. The viable but nonculturable (VBNC) state of dormant bacteria was confirmed by plating experiments on rich medium agar plates. The dormant culture was harvested similarly to the logarithmic-phase sample.

4.2.2 Rich medium

In a total volume of 500 ml, 8.5 g tryptone, 1.5 g yeast extract, 1.25 g dextrose, 2.5 g NaCl, and 1.25 g potassium phosphate dibasic (K_2HPO_4) were added, adjusted to pH 7.4, and autoclaved. Rich medium agar plates contained 1.5% agar.

4.2.3 Acetate minimal medium

A 500-ml volume of acetate minimal medium (AMM) containing 4 g/liter NH_4Cl, 1.4 g/liter K_2HPO_4 at pH 7.4, 0.1 M sodium acetate trihydrate, 0.5 g inosine, 0.5 g yeast extract (0.1%), and 5 ml 100x trace metal stock solution (14.3 g/liter MgSO4 7H2O, 0.00375 g/liter $CuSO_4$ $5H_2O$, 0.079 g/liter $MnCl_2$ $4H_2O$, 0.183 g/liter $FeSO_4$ $7H_2O$, 0.025 g/liter Na_2MoO_4, and 0.005 g/liter $ZnSO_4$ $7H_2O$ was autoclaved, and 0.5 ml filter-sterilized 1,000x vitamin supplement consisting of 0.2 g/10 ml methionine, 0.4 g/10 ml thiamine, and 0.05 g/10 ml biotin dissolved in 0.1 M $NaPO_4$ at pH 7.4 was added.

4.2.4 Cellular lysate preparation

Sample preparation was modified from the procedures in a proteomic study in *Streptococcus pyogenes* (Malström *et al.,* 2012). *M. luteus* MI-2665 cells were harvested from exponentially growing (50 mg) and dormant (200 mg) cultures by centrifugation at 10,000 x *g* for 5 min, followed by 3 washes with 50 mM Tris HCl, pH 7.2. The cells were lysed in 1 ml of lysis buffer (6 M urea, 50 mM Tris HCl with 1x Halt protease inhibitors [Thermo Scientific]) using a Mini-BeadBeater (BioSpec Products) at 4,200 rpm, with five cycles of 3 min of bead beating and 3 min of incubation on ice. Cellular debris was pelleted by centrifugation (16,000 x *g* for 10 min), and the protein supernatant isolated and reduced using 5 mM 2-triscarboxyethyl phosphine hydrochloride (TCEP)

at 37 °C for 30 min. The sample was alkylated by adding iodoacetamide (IAA) to 5 mM and incubated at room temperature for 30 min in the dark. The reduced/alkylated sample was concentrated to 25 µl using a 10-kDa molecular-mass-cutoff filter (GE Healthcare), and protein concentrations were determined by absorbance at 280 nm. The concentrated protein sample (700 µg) was mixed with Laemmli buffer, boiled at 95 °C for 10 min, and separated on a 10-well, 4 to 12% Bis-Tris NuPAGE gel (Invitrogen) at 200 V for 40 min using 1x MOPS (morpholinepropanesulfonic acid) running buffer. The gel was stained with Coomassie brilliant blue for an hour and destained overnight.

4.2.5 In-gel digestion

Gel lanes were cut into 16 molecular-weight-range fractions and placed in separate LoBind Eppendorf tubes. The gel piece was minced into small fragments and washed in 100 mM ammonium bicarbonate-acetonitrile in a 3:1 (vol/vol) ratio at 37 °C for 30 min. The cycle was repeated 3 times or until the gel became colorless. The gel was dehydrated with acetonitrile, rehydrated with 100 µl of trypsin (200 ng) in 100 mM ammonium bicarbonate buffer at pH 7.4, and incubated overnight at 37 °C. The supernatant was transferred to a new LoBind Eppendorf tube and extracted with 100 µl of 1% formic acid-acetonitrile in a 3:1 (vol/vol) ratio. The extraction was repeated two more times, and the pooled supernatant was lyophilized and stored at -80 °C or directly resuspended in 28 µl of 1% formic acid and used for mass spectrometry analysis. For each biological replicate, 48 total liquid chromatography- tandem mass spectrometry (LC-MS/MS) runs were performed for three independent technical replicates.

4.2.6 Liquid chromatography-tandem mass spectrometry

LC-MS/MS was performed on a ThermoFinnigan LTQ (linear trap quadrupole) linear ion trap mass spectrometer coupled to an Agilent 1290 Infinity ultraperformance liquid chromatography (UPLC) system using solvent A (water plus 0.1% formic acid) and solvent B (methanol plus 0.1% formic acid). The gradient was 20% solvent B to 90% solvent B over 60 min at a flow rate of 40 µl/min on a 500-µm by 6-cm C_{18} reversed-phase column packed in-house. The electrospray voltage was set to 3.78 kV, with sweep, auxiliary, and sheath gas set to 0 on a standard IonMax electrospray ionization (ESI) source. The capillary temperature was 250 °C, and data were acquired using data-dependent dynamic exclusion MS/MS on the three highest intensity masses observed in the MS scan. The RAW data files were transformed to MGF format using the MSConvert utility program from ProteoWizard.

4.2.7 Data analysis

MGF data files were searched using OMSSA against a database of 2,196 *M. luteus* proteins (NCBI RefSeq GCF_000023205.1 ASM2320v1) appended with a randomized decoy database (Geer *et al.,* 2004 and Elias *et al.,* 2005). MGF files for all 16 fractions for each technical replicate were combined into a single merged file prior to search. Default settings were employed for the ion trap, as follows: trypsin digestion with a maximum of 1 missed cleavage, precursor *m/z* tolerance of 2 Da, product ion tolerance of 0.8 Da, and carbamidomethyl C as a variable modification. The E value threshold was adjusted to a false discovery rate of $\leq$1%. Quantitative-analysis-normalized spectral abundance factor (NSAF) values were natural log transformed, and the Shapiro-Wilk test performed to confirm normal distribution of data. Proteins with no spectral counts in either the logarithmic or dormant states were replaced by 0.16 spectral counts as previously described to enable natural log transformation (Zybailov *et al.,* 2006). Statistical

analysis was performed using Microsoft Excel to identify proteins that were significantly upregulated (*P* value <0.05) in all three biological replicates in either dormant or log phase. For the data shown in Fig. 2, negative natural log transformation was used to establish positive *x* and *y* axes to simplify visualization of the data.

4.2.8 Visualization and comparison of protein structures

Protein structures were visualized and comparisons performed in Chimera. Structure files from the Protein Data Bank (PDB) were imported into Chimera based on their PDB identification numbers and manipulated with the core toolset available in the program. Overlays were used for comparison of protein structures from different organisms and were generated using MatchMaker with the following parameters: best-aligning pair of chains between reference and match structure, Needleman-Wunsch alignment algorithm, BLOSUM-62 matrix, and iterate by pruning long atom pairs until no pair exceeds 3.0 Å. All other parameters were the defaults. The root mean square deviations (RMSD) for overlays were calculated at 1.139 Å for 101 atom pairs (UspA structures) (Fig. 4), 1.029 Å for 356 atom pairs (Icl1 structures), 1.311 Å for 235 atom pairs (CysK1 structures), and 1.205 Å for 216 atom pairs (homoserine sulfhydrylase structures).

4.3 Results and discussion

4.3.1 Nutrient starvation induces VBNC dormant state in *M. luteus*

M. luteus cells were grown in minimal acetate medium to induce a transition to VBNC dormancy by nutrient starvation (Kaprelyants and Kell, 1993). After inoculation of *M. luteus* from rich to minimal acetate medium, the growth curve was monitored for more than 15 days by measuring OD_{600}, colony-forming units (CFU) and respiratory activity using TMRM dye at different time points

(Figure 4.1). The optical density of the culture increased up to four days, then decreased slightly and stayed constant during the rest of the experiment. In contrast, both colony-forming units and respiratory activity increased exponentially as expected in the logarithmic growth phase and decreased as the cells entered the stationary phase (Day 6). Significant loss of cell viability after Day 10 suggests that the majority of the remaining cells have entered the viable but non-culturable (VBNC) dormant state. The ability of the VBNC dormant cells to resuscitate back to active cells was confirmed by the addition of purified Rpf (Zhu *et al.,* 2003).

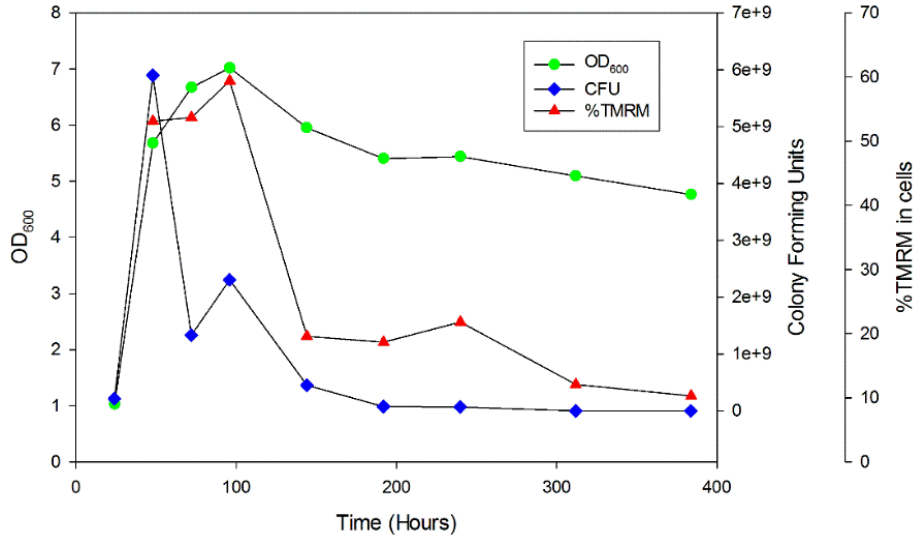


Figure 4.1 *M. luteus* growth curve in minimal acetate media. *M. luteus* was cultured in minimal acetate medium to induce nutrient-starved VBNC dormant cells. Optical density (OD_{600}, circle), colony-forming units (CFU, diamond) and respiratory activity (TMRM, triangle) were monitored at the indicated time points. (Figure Courtesy of Abiodun Bodunrin and William Widger, University of Houston).

The Wayne Model (oxygen limitation) and Loebel Model (nutrient deprivation) are two common *in vitro* methods to induce dormancy in many bacteria including *M. tuberculosis (*Wayne and Hayes,

1996). In the Loebel model, cells from nutrient-rich media are washed and directly transferred to phosphate-buffered saline (Loebel *et al.,* 1933). Instead of abrupt nutrient starvation, we utilized a long-term starvation model following Kaprelyants and Kell, 1993 to reflect *in vivo* conditions. This proteomics study used *M. luteus* cells cultured in minimal acetate medium for at least a month before harvesting dormant samples.

4.3.2 Reduced global protein diversity in the VBNC state

M. luteus cells grown in minimal acetate medium were harvested in logarithmic (log), stationary and VBNC dormant phases. The cells were lysed by bead beating and the total protein was separated by gel electrophoresis. After in-gel digestion and tandem mass spectrometry, the number of proteins in each growth phase was identified by performing database searches against the total *M. luteus* proteome using OMSSA. Global protein diversity was greatly decreased in the dormant phase compared to log and stationary phases based on Coomassie gel staining (Figure 4.2). This result was consistent with mass spectrometry identification of 200-300 proteins in the dormant phase compared to around 700 proteins in log and stationary growth phases. We also observed a lower total number of peptide spectra annotated to proteins in the dormant phase that larger quantities of the dormant sample (4 times the weight of log cell pellet) could not resolve. To evaluate potential bias in our protein-sampling procedures, we performed Gene Ontology analysis to check if any functional protein classes are overrepresented in the data. We did not find any statistically significant overrepresentation of any functional classifications indicating that our experimental approach did not enrich any protein class above the background levels.

The gel electrophoresis and mass spectrometry data strongly indicated the drastic loss in protein diversity as well as protein quantities in *M. luteus* VBNC dormant state. These results suggest a global shutdown of metabolic and protein synthesis machinery in VBNC state which correlates

with VBNC states in other bacterial species, such as *Vibrio vulnificus* and *M. smegmatis* (Mulcahy *et al.,* 2010). The most abundant protein band in the dormant gel lane was identified as a bifunctional catalase (KatG). Catalase was also highly abundant in log and stationary phase (Figure 4.2). KatG has peroxidase activity and is an important defense against reactive oxygen species (ROS). Retaining KatG protein at high levels in VBNC state might be advantageous to combat ROS during oxidative stress in dormancy and could be related to stress survival as observed in *V. vulnificus* (Mulcahy *et al.,* 2010).

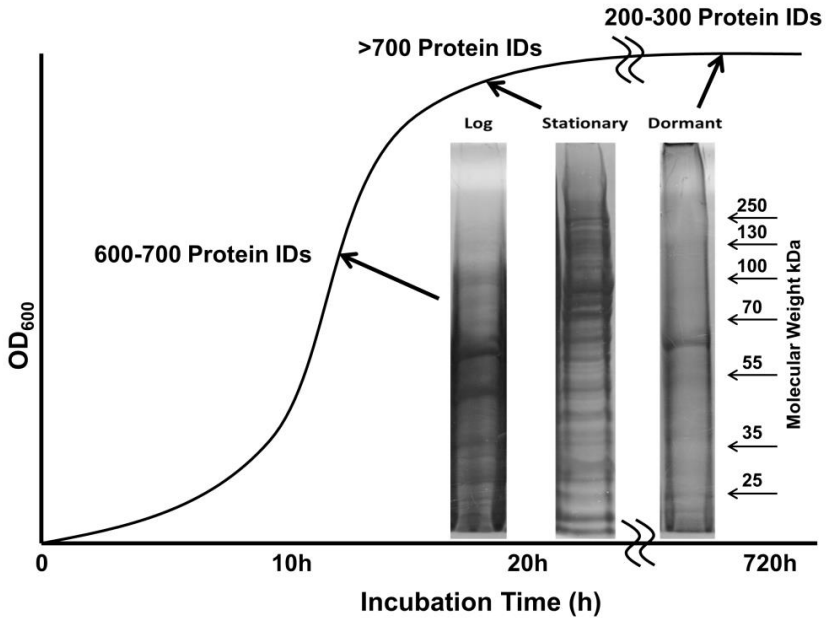


Figure 4.2 Total protein comparison of *M. luteus* in log, stationary and dormant growth phase by gel electrophoresis and mass spectrometry. Total protein is significantly reduced in dormant (200 to 300 proteins) compared to log (~700 proteins) and stationary phase (~700 proteins). Note: 60 kDa intense protein band in the dormant sample was identified as catalase by mass spectrometry. Reprinted from "A proteomic signature of dormancyin the actinobacterium of micrococcus luteus," by S. Mali, M. Mitchell, S. Havis, A. Bodunrin, J. Rangel, G. Olson, W. Widger, S.J. Bark, 2017, *Journal of Bacteriology*, *199(*14*),* 206-217. Copyright [2016] by American Society for Microbiology.

4.3.3 VBNC protein identification overlap is similar to logarithmic (log) and stationary phase

Protein identification (ID) overlaps between dormant, log and stationary phase were compared using a Venn diagram (Figure 4.3). Surprisingly, protein ID overlap is similar between VBNC and stationary phase (314 shared protein IDs) and VBNC and log phase (315 shared protein IDs). Among all three states, 300 proteins were shared (88% of VBNC proteins, 44% of stationary, 39% of log). These data suggest that the VBNC state is similarly related to log and stationary growth phases, but 461 protein identifications were lost in VBNC compared to log phase data. We observed a very high overlap of protein IDs between log and stationary phase (603 proteins) that may be related to the time proximity of the two growth phases. However, we have noted that *M. luteus* in the stationary phase has limited growth potential when transferred to a nutrient-rich medium.

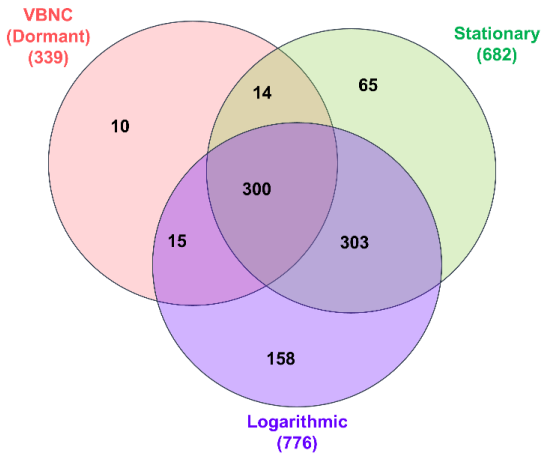


Figure 4.3 Venn diagram representing the number of proteins overlap between log, stationary and dormant growth phases. Proteins identified by at least two independent spectra were only considered to generate this diagram. The protein identification overlap of the dormant phase is similar to both log and stationary phase. The unique proteins observed in either growth phase may not be statistically significant. Reprinted from “A proteomic signature of dormancyin the actinobacterium of micrococcus luteus,” by S. Mali, M. Mitchell, S. Havis, A. Bodunrin, J. Rangel, G. Olson, W. Widger, S.J. Bark, 2017, *Journal of Bacteriology*, *199(*14), 206-217. Copyright [2016] by American Society for Microbiology.

This suggests that *M. luteus* cells in the stationary phase may already undergo a metabolic shift towards the VBNC state, but the mechanisms constituting this shift are not understood.

The Venn diagram indicates 10 unique protein IDs for the dormant state, but these unique IDs are not the whole story. Proteins may demonstrate statistically significant upregulation in a dormant state, but the Venn data does not reflect the number of protein levels. Quantitative proteomics analysis can more effectively determine statistically significant proteins that were upregulated in the dormant state compared to log growth phases in all three biological replicates.

4.3.4 Multiple proteins are differentially expressed in VBNC and log phases

Quantitative proteomics analysis was performed by label-free normalized spectral abundance factors (NSAF) (Zybailov *et al.,* 2016). The NSAF value is a relative measurement of protein abundance represented by the ratio of the number of spectra contributed by a protein compared to the total number of spectra from all proteins in the sample. NSAF normalizes for protein molecular weight and the total amount of protein in the sample, which enables more effective comparisons across experiments. The natural log-transformed NSAF value was used to calculate P-value and determine the proteins that were significantly upregulated in the dormant phase compared to the log phase in all biological and technical replicates. The NSAF values for all proteins are percentages and add to 1 (or 100%), so the natural log transformation generates a normalized distribution with all abundances represented by a negative numerical value. We graphed these values as the negative of the ln(NSAF) to generate a more effective graphical representation of the data (Figure 4.4). The scatter plot analysis comparing negative ln(NSAF) values showed a wide distribution of protein quantities in dormant and log phases (Figure 4.4) as expected from the large differences in protein quantities present in a biological sample (Figure 4.2). In this scatter plot, increasing distance from the diagonal correlation line (i.e. ≥ 2-fold change)

represents the upregulation of protein in either of the growth phase with larger ln(NSAF) indicating lower protein abundance. Figure 4.4 shows the comparison of log and dormant protein levels in one biological replicate with many proteins upregulated in dormant and log growth phase (large distance from diagonal). However, the rigorous statistical analysis only confirmed proteins upregulated in the dormant and log phase as indicated by colored circles in Figure 4.3. We found 18 proteins that were statistically more abundant in the dormant phase compared to the log phase (blue) and 46 proteins more abundant in log compared to the dormant growth phase (yellow).

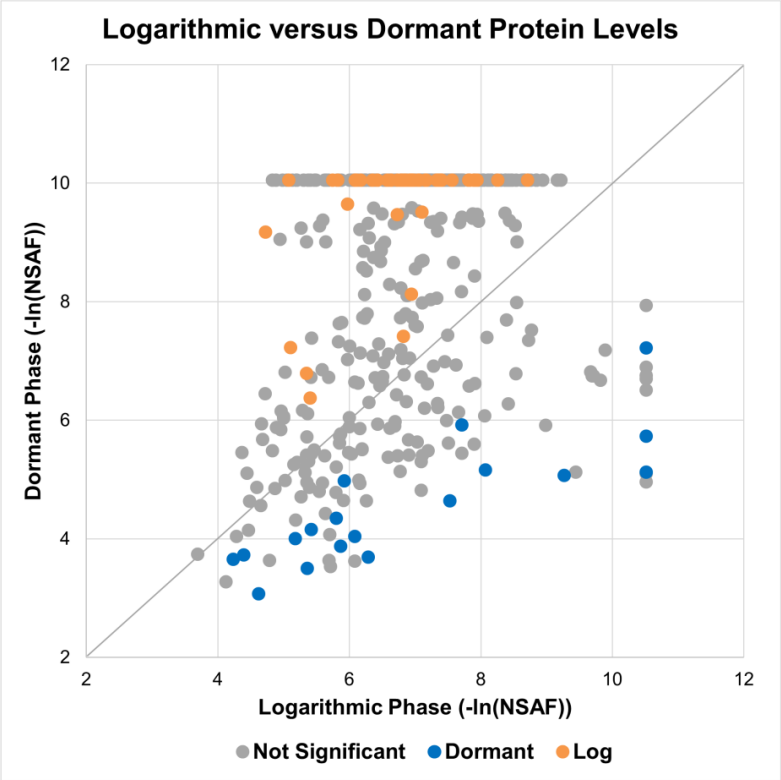


Figure 4.4 Scatter plot of –ln(NSAF) values to compare proteins upregulated in log and dormant phase. Proteins that are significantly upregulated (P values <0.05) in dormant and log phase are colored blue and yellow respectively. Note that negative values of ln(NSAF) were used to generate positive x and y axes for simplicity but larger numbers represent smaller protein abundance. Reprinted from "A proteomic signature of dormancyin the actinobacterium of micrococcus luteus," by S. Mali, M. Mitchell, S. Havis, A. Bodunrin, J. Rangel, G. Olson, W. Widger, S.J. Bark, 2017, *Journal of Bacteriology*, *199(*14), 206-217. Copyright [2016] by American Society for Microbiology.

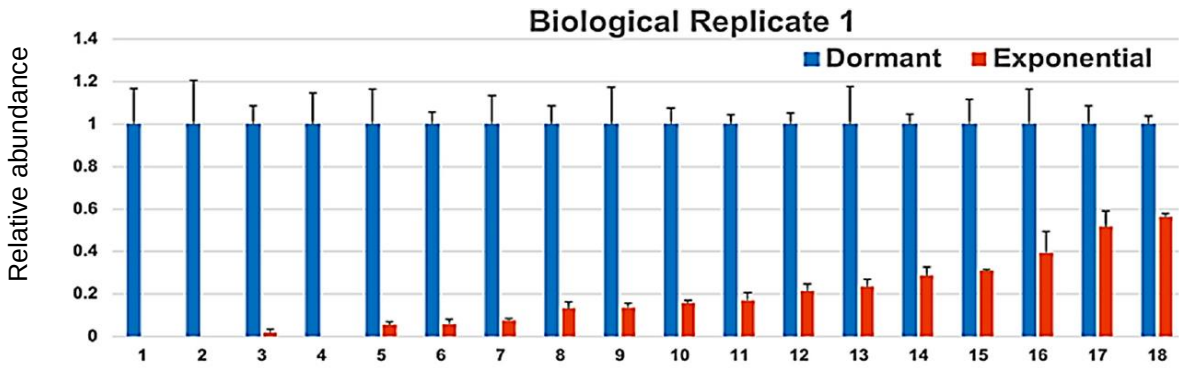
The small number of differentially regulated proteins in the dormant phase precluded functional organization by gene ontology. However, the quantitative differences in protein levels suggest many biological processes including metabolic and ribosomal regulation are important for the VBNC dormant state.

4.3.5 Eighteen proteins are upregulated in VBNC compared to log phase

The eighteen proteins upregulated in the dormant phase represent a molecular signature for the VBNC dormant stress response in *M. luteus*. The relative abundance of these proteins is presented in the bar diagram and heat map in Figure 4.5. Four of these proteins are already known to be associated with hypoxia and latency stress responses in *M. tuberculosis*. These eighteen dormancy signature proteins are homologous to proteins from other bacteria, especially within actinobacteria phylum including pathogenic *M. tuberculosis*. Interestingly, Universal Stress Protein A, isocitrate lyase, cysteine synthase, and O-Acetylhomoserine sulfhydrylase are highly homologous among mycobacteria and reported to be important in *M. tuberculosis* latency. While the other 14 dormancy signature proteins are novel, some of those proteins are known to be involved in different stress responses but a direct link with dormancy has not been studied.

4.3.5.1 Universal stress protein A (UspA)

M. luteus genome encodes three different universal stress protein A (UspA) proteins. In our global quantitative proteomics study, only one UspA (WP_010079616.1) was upregulated in the VBNC dormant state (Figure 4.5). Multiple sequence alignment showed that *M. luteus* UspA sequence is homologous with other bacteria including *M. tuberculosis* Rv2623 with 35% sequence identity and 50% similarity (Figure 4.6). Rv2623 is an ATP-binding protein and facilitates chronic and persistent tuberculosis infection in an animal model study (Drumm *et al*., 2009).



	Protein identification	Replicate 1		Replicate 2		Replicate 3	
		Log	Dormant	Log	Dormant	Log	Dormant
1	Isocitrate lyase	-10.51	-5.12	-8.33	-3.77	-9.73	-4.84
2	Hypothetical protein Mlut_03760	-10.51	-5.72	-8.01	-6.14	-7.56	-5.05
3	O-acetylhomoserine sulfhydrolase	-9.264	-5.07	-10.80	-6.31	-9.50	-5.15
4	ABC transporter substrate-binding protein	-10.51	-7.21	-6.31	-5.76	-9.26	-6.55
5	Acetyl-CoA synthetase	-8.06	-5.16	-6.59	-5.32	-8.64	-6.50
6	Ribosomal biogenesis protein	-7.52	-4.64	-7.45	-5.61	-5.87	-4.92
7	Universal stress protein UspA	-6.28	-3.69	-4.84	-3.89	-4.86	-3.54
8	Aldolase	-6.08	-4.03	-8.86	-4.50	-9.61	-5.04
9	Integration host factor	-5.86	-3.87	-4.87	-3.23	-4.96	-4.39
10	Aldehyde dehydrogenase	-5.35	-3.50	-4.71	-3.43	-6.36	-3.36
11	Amidohydrolase	-7.70	-5.92	-6.52	-5.77	-7.45	-6.40
12	Hypothetical protein Mlut_10450	-4.62	-3.07	-3.85	-2.90	-4.39	-3.47
13	Cysteine synthase	-5.79	-4.34	-6.47	-5.09	-6.40	-4.78
14	Alcohol dehydrogenase	-5.41	-4.16	-4.93	-4.12	-6.54	-4.38
15	Ketol-acid reductoisomerase	-5.17	-4.00	-5.51	-4.46	-5.72	-4.28
16	Superoxide dismutase	-5.92	-4.97	-5.12	-4.41	-5.20	-3.90
17	ATP synthase subunit beta	-4.39	-3.72	-4.37	-3.79	-4.68	-4.04
18	Molecular chaperone GroEL	-4.23	-3.65	-3.93	-3.34	-3.84	-3.59

Figure 4.5 Eighteen proteins are significantly upregulated in the dormant versus log phase. (Top) Bar graph comparison of protein level in dormant versus log phase after normalizing each dormant protein level as 1. Error bars represent standard deviations within technical replicates. (Bottom) Heat map comparison of the natural log-transformed data between log and dormant growth phase. All eighteen proteins were significantly upregulated in dormant versus log phase with P-value <0.05 in three biological replicates. Numbering in the bar graph corresponds to the same number in the heat map. Reprinted from "A proteomic signature of dormancyin the actinobacterium of micrococcus luteus," by S. Mali, M. Mitchell, S. Havis, A. Bodunrin, J. Rangel, G. Olson, W. Widger, S.J. Bark, 2017, *Journal of Bacteriology*, *199(*14), 206-217. Copyright [2016] by American Society for Microbiology.

In general, Usp-type proteins are known to be differentially regulated under various stress conditions including hypoxia, nutrient starvation, heat, and antibiotics (Kvint *et al*., 2003). To further investigate the structural homology across species, UspA crystal structures from *M. tuberculosis*, *M. smegmatis,* and *L. plantarum* were overlaid using Chimera (Figure 4.7). Interestingly, the structures were highly similar and the Rossman fold was structurally conserved, which is important for nucleotide-binding proteins (Drumm *et al*., 2009). Usp proteins can exist as single or double domain structures. Although many actinobacteria contain double domain Usp-protein like Rv2623, *M. smegmatis* and *L. plantarum* have single domain Usp variants (Drumm *et al*., 2009). Both sequence alignment and structural overlay suggest that *M. luteus* UspA protein is similar to *M. tuberculosis* Rv2623 and contains two Rossman fold ATP-binding domains. The biochemical function of Usp proteins is not known. However, based on the sequence homology with Rv2623, we propose that *M. luteus* UspA may regulate the VBNC transition through signaling interactions with different stress response proteins. In support of this hypothesis, we report that *E. coli* UspA-type proteins can be phosphorylated and in *E. coli* UspA knockout, the levels of acetyl-CoA synthase and the glyoxylate shunt protein isocitrate lyase are modulated (Nyström and Neidhardt, 1993; Nachin *et al*., 2005). Coincidentally, both of these proteins are upregulated along with the upregulation of UspA in our global proteomics study (Figure 4.5).

UspA has been associated with oxidative stress response in *E. coli*. UspA also appears to be regulated by proteins regulating the bacterial cold shock stress response (Phadtare, 2004). These reports indicate that UspA may be an important component mechanistically linked to many different stress responses. In *M. tuberculosis*, 10 different Usp-type proteins are present with possible functional redundancy. However, genetic studies suggest that individual Usp proteins may have incomplete redundancy and specific Usp proteins might be functional under different stress conditions (O'Toole and Williams, 2003). The data linking Rv2623 to chronic infection indicate this protein may have specific importance in the infection process that may not be

reflected by other Usp variants. Therefore, the combination of our data and the supporting literature indicate that UspA has a role in modulating the VBNC state in response to nutrient deprivation and maybe other stress responses as well. The UspA homolog Rv2623 may be associated with the same type of response in *Mycobacterium tuberculosis*.

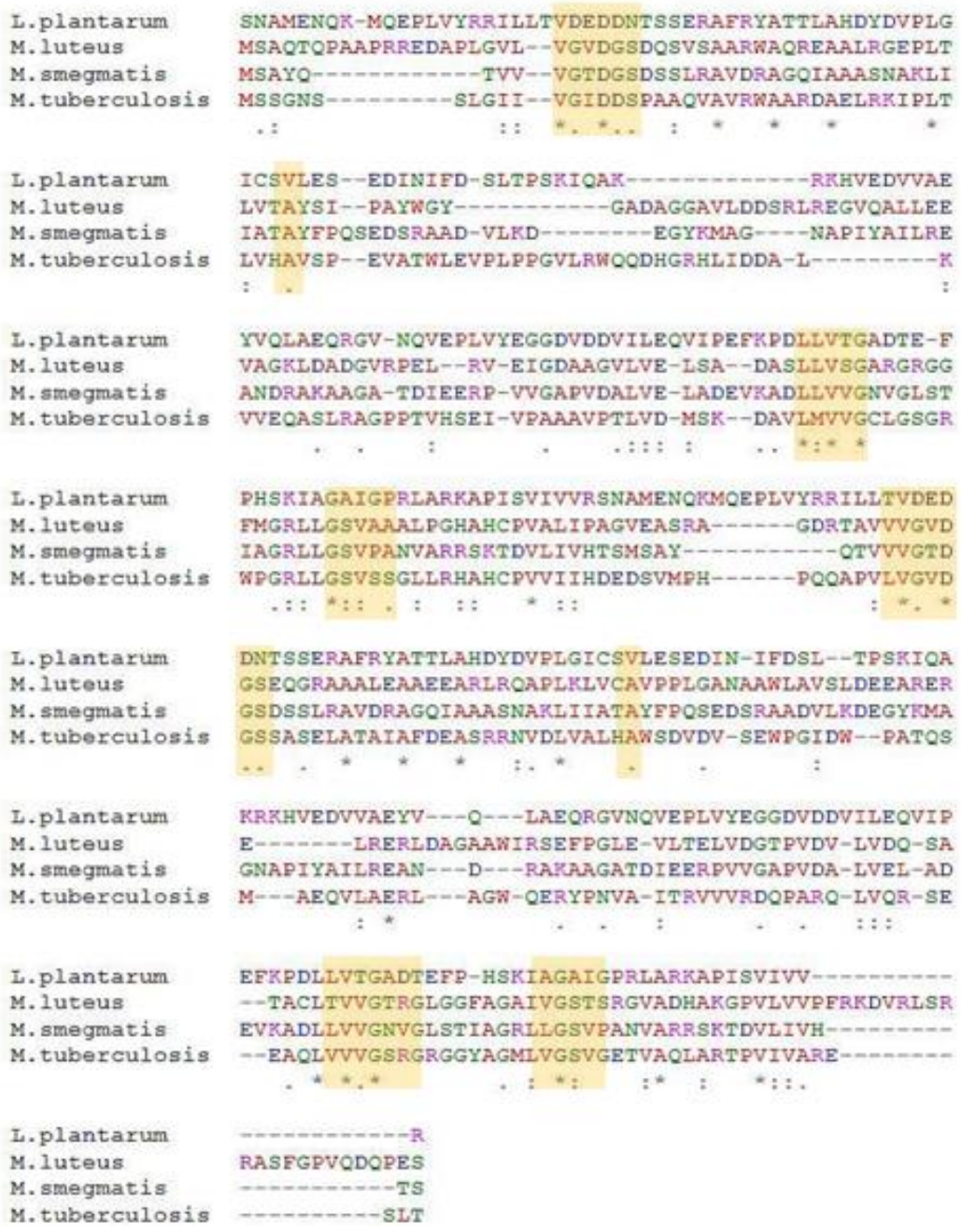


Figure 4.6 Alignment of UspA proteins. Multiple sequence alignment of UspA proteins from *M. luteus*, *M. tuberculosis*, *M. smegmatis*, and *L. plantarum.* Highlighted regions are conserved ATP and Mg^{2+} binding site residues. Color Code: R, K = Pink; D, E = Blue; Q, N, H, S, T, Y, C, G = Green; M, W, A, I, L, F, V, P = Red. Reprinted from "A proteomic signature of dormancyin the actinobacterium of micrococcus luteus," by S. Mali, M. Mitchell, S. Havis, A. Bodunrin, J. Rangel, G. Olson, W. Widger, S.J. Bark, 2017, *Journal of Bacteriology*, *199(*14*),* 206-217. Copyright [2016] by American Society for Microbiology.

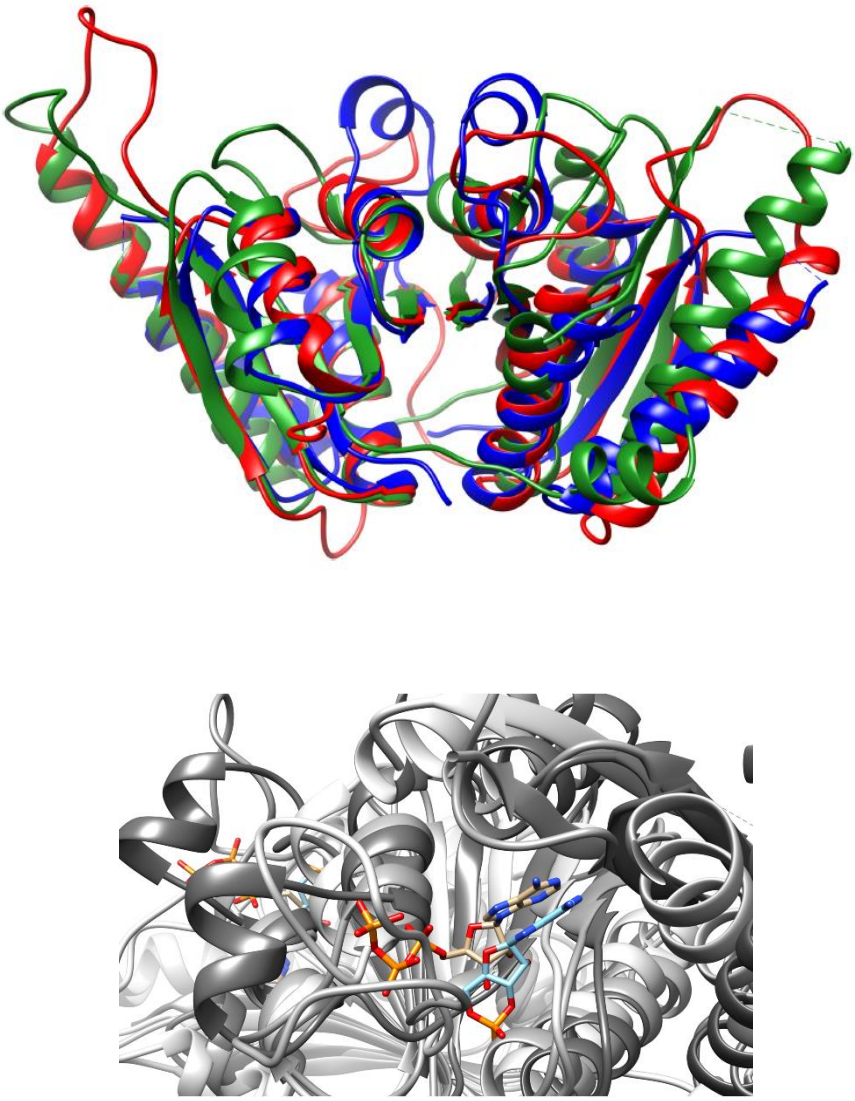


Figure 4.7 Overlay of UspA protein structures. (Top) Usp protein structures from *M. tuberculosis* (PDB: 3cis; red), *M. smegmatis* (PDB: 5ahw; blue), and *L. plantarum* (PDB: 3fg9; green) were overlaid in Chimera using a 3 Å cutoff. (Bottom) Enlarged image of the same alignment to show the overlay of the nucleotide substrate binding region. The cyclic AMP substrate from *M. smegmatis* is shown in light blue color. Reprinted from "A proteomic signature of dormancyin the actinobacterium of micrococcus luteus," by S. Mali, M. Mitchell, S. Havis, A. Bodunrin, J. Rangel, G. Olson, W. Widger, S.J. Bark, 2017, *Journal of Bacteriology*, *199(*14*),* 206-217. Copyright [2016] by American Society for Microbiology.

4.3.5.2 Isocitrate lyase

We observed a significant upregulation of isocitrate lyase in dormancy. Isocitrate lyase is a primary enzyme in the glyoxylate shunt, an anaplerotic metabolic pathway. This enzyme is also induced under hypoxic growth conditions in *M. tuberculosis* (Wayne and Lin, 1982). During hypoxia, the citric acid cycle would be downregulated since oxidative phosphorylation requires oxygen. Isocitrate lyase, along with malate synthase, catalyzes the conversion of isocitrate to malate through glyoxylate and succinate intermediates. This anaplerotic pathway bypasses two oxidative decarboxylation steps in the citric acid cycle, preserving a limited carbon pool during starvation and reducing NADH formation. We noted that malate synthase was significantly upregulated in the dormant phase in two of the three biological replicates in our study (Figure 4.8). The *M. luteus* isocitrate lyase (Icl) structure is not known. However, the structural comparison of Icl from distant species (*M. tuberculosis* and *Aspergillus nidulans*) shows remarkable structural conservation of the enzymatic core (Figure 4.9). Multiple sequence alignment of Icl from *M. luteus* with *M. tuberculosis* and *A. nidulans* showed high sequence identity including the conserved active site residues and metal-binding sites (Figure 4.10). *M. luteus* Icl sequence had 79% identity and 87% similarity with *M. tuberculosis* and 39% identity and 54% similarity with *A. nidulans* Icl proteins. Based on the protein homology, we hypothesize that these proteins perform the same function across actinobacteria. Moreover, a significant mutation of the Icl core structure would result in a compromised enzyme that would not support survival under stress conditions including nutrient starvation. This hypothesis is consistent with the knockout of the *icl1* gene in *M. tuberculosis,* which severely impaired survival in both acute and persistent infection models (Muñoz-Elías and McKinney, 2005).

Protein Identifications	Biological Replicates		
	Rep#1	Rep#2	Rep#3
WP_002857590.1 MULTISPECIES: molecular chaperone GroES	0.0298	0.0033	0.4212
WP_010078529.1 dihydrolipoamide dehydrogenase	0.0014	0.0072	0.0556
WP_002854699.1 MULTISPECIES: 30S ribosomal protein S20	0.0301	0.0491	0.1055
WP_012750946.1 ornithine carbamoyltransferase	0.0001	0.0001	0.3449
WP_010079812.1 glycerol kinase	0.0231	0.0871	0.0161
WP_029248252.1 flavin-nucleotide-binding protein	0.0045	0.2412	0.0204
WP_010080229.1 DNA-binding ferritin-like protein	0.0269	0.0396	0.0981
WP_010079157.1 MULTISPECIES: elongation factor Ts	0.1095	0.0027	0.0238
WP_010080546.1 ABC transporter substrate-binding protein	0.0331	0.0239	0.0580
WP_036335912.1 malate synthase	0.0012	0.0975	0.0039
WP_010079369.1 ABC transporter	0.0201	0.0683	0.0210
WP_010078687.1 glucose-1-phosphate adenylyltransferase	0.0042	0.0008	0.5663
WP_010080230.1 molecular chaperone DnaK	0.0698	0.0076	0.0350
WP_010079704.1 hypothetical protein	0.0329	0.1178	0.0022
WP_010078640.1 menaquinol-cytochrome C reductase cytochrome b subunit	0.2581	0.0005	0.0011

Figure 4.8 List of proteins overexpressed in the dormant phase in two biological replicates. Green cells indicate P-value < 0.05 and red cells indicate > 0.05 P-value in three biological replicates (Rep). Reprinted from “A proteomic signature of dormancyin the actinobacterium of micrococcus luteus,” by S. Mali, M. Mitchell, S. Havis, A. Bodunrin, J. Rangel, G. Olson, W. Widger, S.J. Bark, 2017, *Journal of Bacteriology*, *199(*14), 206-217. Copyright [2016] by American Society for Microbiology.

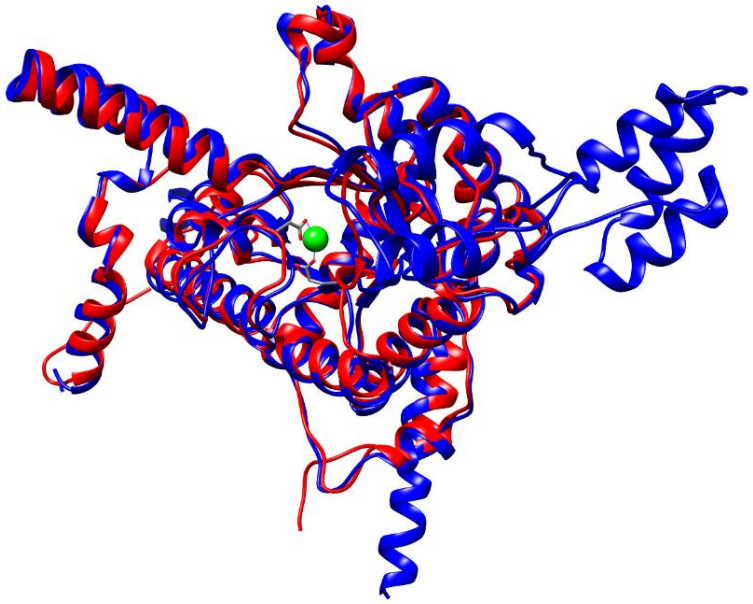


Figure 4.9 Overlay of isocitrate lyase protein structures. Isocitrate lyase 1 (Icl1) protein structures from *M. tuberculosis* (1f61, red) and *A. nidulans* (1dqu, blue) were overlaid in Chimera using a 3 Å cutoff. The green sphere is a magnesium ion coordinated by Glu155 and Asp108 in the *M. tuberculosis* structure and is enlarged to simplify visualization. Reprinted from "A proteomic signature of dormancyin the actinobacterium of micrococcus luteus," by S. Mali, M. Mitchell, S. Havis, A. Bodunrin, J. Rangel, G. Olson, W. Widger, S.J. Bark, 2017, *Journal of Bacteriology*, *199(*14*),* 206-217. Copyright [2016] by American Society for Microbiology.

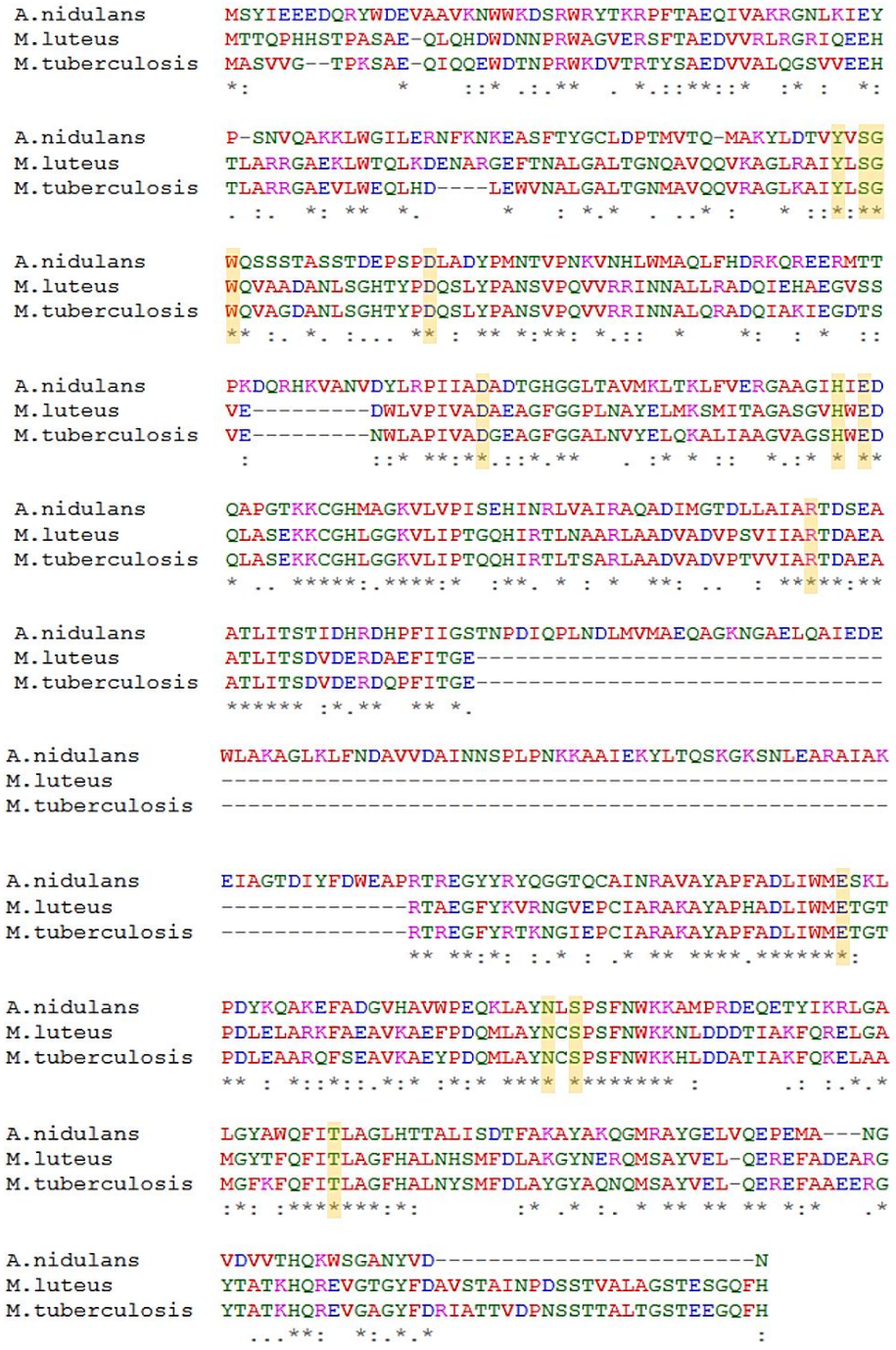


Figure 4.10 Alignment of isocitrate lyase proteins. Multiple sequence alignment of isocitrate lyase proteins from *M. luteus*, *M. tuberculosis*, and *Aspergillus nidulans.* Highlighted regions are conserved active site residues and Mg^{2+}/Mn^{2+} binding site residues. Color code is the same as Figure 4.6. Reprinted from "A proteomic signature of dormancyin the actinobacterium of micrococcus luteus," by S. Mali, M. Mitchell, S. Havis, A. Bodunrin, J. Rangel, G. Olson, W. Widger, S.J. Bark, 2017, *Journal of Bacteriology*, *199(*14*),* 206-217. Copyright [2016] by American Society for Microbiology.

4.3.5.3 Cysteine and methionine synthase

Two related proteins involved in amino acid metabolism were upregulated in *M. luteus* VBNC state. Cysteine synthase (CysK1, or *O*-acetylserine sulfhydrylase) is a critical enzyme in cysteine metabolism. CysK1 catalyzes the synthesis of cysteine from *O*-acetylserine and hydrogen sulfide. An alignment between CysK1 homologous proteins from *M. luteus* and *M. tuberculosis* showed a 70% identity and 85% sequence similarity (Figure 4.11). Similar levels of sequence homology are observed among mycobacterial species. The overlay of several crystal structures demonstrates structural conservation for CysK1 across different mycobacteria (Figure 4.12). A related enzyme involved in methionine biosynthesis, *O*-acetylhomoserine sulfhydrylase, was also observed to be greatly increased in the VBNC state. This enzyme catalyzes a reaction similar to CysK1 but uses *O*-acetylhomoserine condensation with methylthiol to generate methionine. An alignment between *M. luteus* and *M. tuberculosis O*-acetylhomoserine sulfhydrylase proteins demonstrates 65% identity and 77% sequence similarity (Figure 4.13). Because the structure of *M. tuberculosis O*-acetylhomoserine sulfhydrylase was not known, we used the homologous protein *O*-succinylhomoserine sulfhydrylase for structural comparison (Figure 4.14). We hypothesize that these two-important amino acid metabolism enzymes would signal a pivot away from the complete catabolic breakdown of carbon sources to the conservation of carbon species under nutrient deprivation. Moreover, these enzymes synthesize cysteine and methionine, which would also be important to stabilize the oxidation-reduction environment within dormant bacterial cells under starvation or oxidative stress condition. The upregulation of superoxide dismutase in dormancy and retention of high levels of KatG supports this hypothesis.

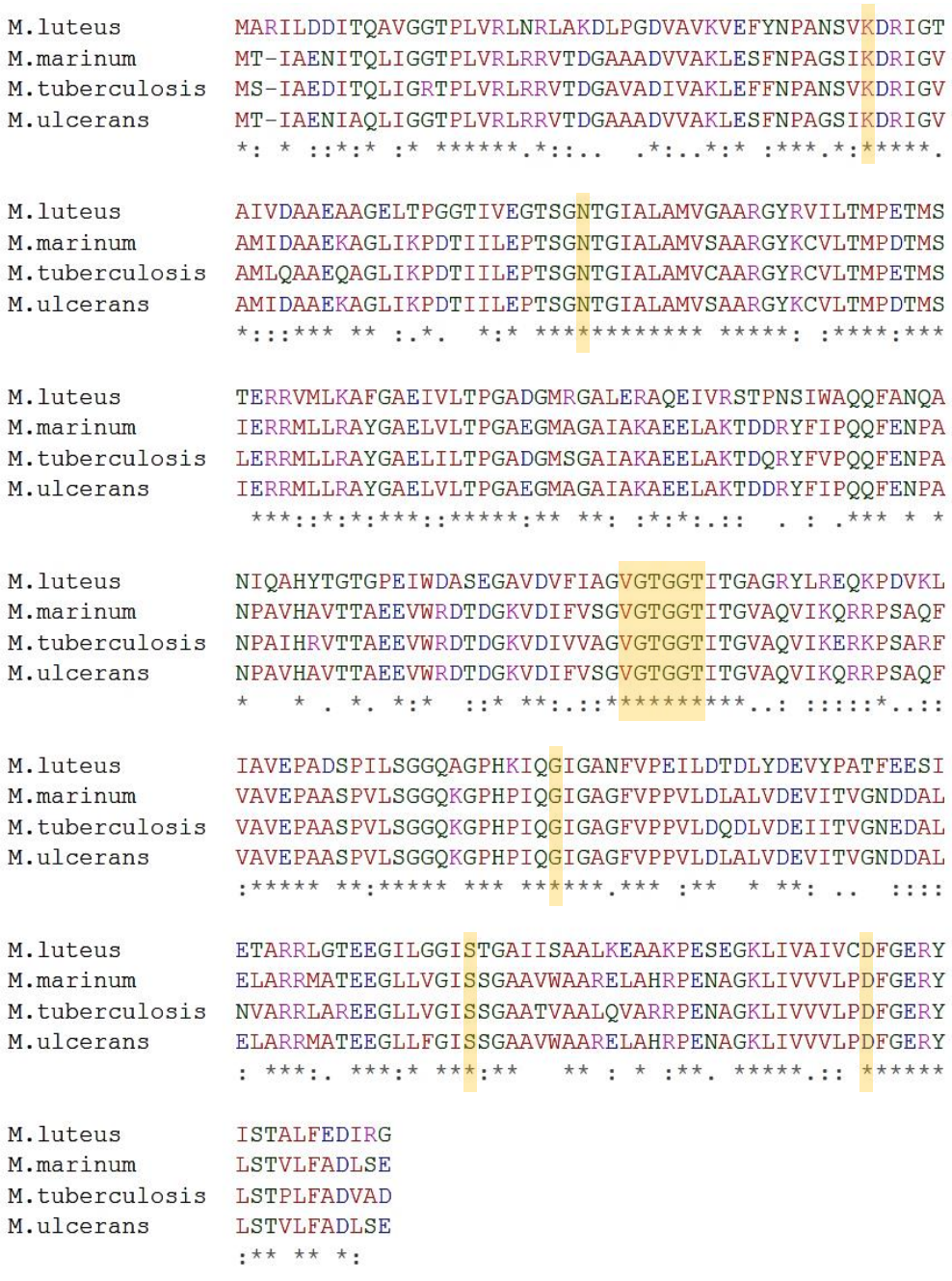


Figure 4.11 Alignment of cysteine synthase. Multiple sequence alignment of cysteine synthase proteins from *M. luteus*, *M. tuberculosis*, *M. ulcerans*, and *M. marinum.* Highlighted regions are conserved active sites and pyridoxal 5'-phosphate-binding site residues. The color code is the same as Figure 4.6. Reprinted from "A proteomic signature of dormancyin the actinobacterium of Reprinted from "A proteomic signature of dormancyin the actinobacterium of micrococcus luteus," by S. Mali, M. Mitchell, S. Havis, A. Bodunrin, J. Rangel, G. Olson, W. Widger, S.J. Bark, 2017, *Journal of Bacteriology*, *199(*14), 206-217. Copyright [2016] by American Society for Microbiology.

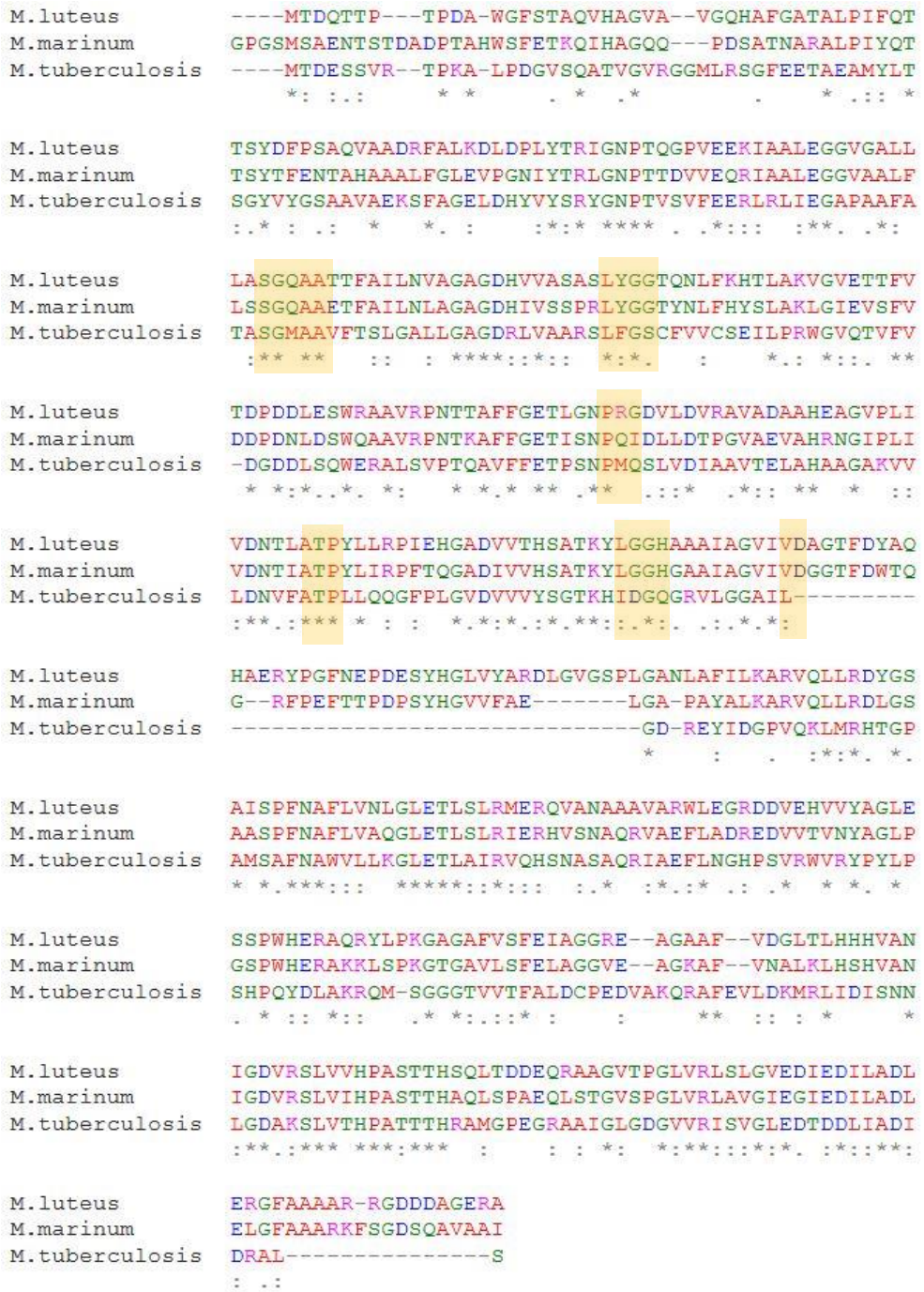


Figure 4.12 Alignment of O-succinylhomoserine sulfhyrylase. Multiple sequence alignment of O-succinylhomoserine sulfhydrylase proteins from *M. luteus*, *M. tuberculosis*, and *M. marinum.* Highlighted regions are conserved substrate-cofactor active sites. The color code is the same as Figure 4.6. Reprinted from “A proteomic signature of dormancyin the actinobacterium of micrococcus luteus,” by S. Mali, M. Mitchell, S. Havis, A. Bodunrin, J. Rangel, G. Olson, W. Widger, S.J. Bark, 2017, *Journal of Bacteriology*, *199(*14*),* 206-217. Copyright [2016] by American Society for Microbiology.

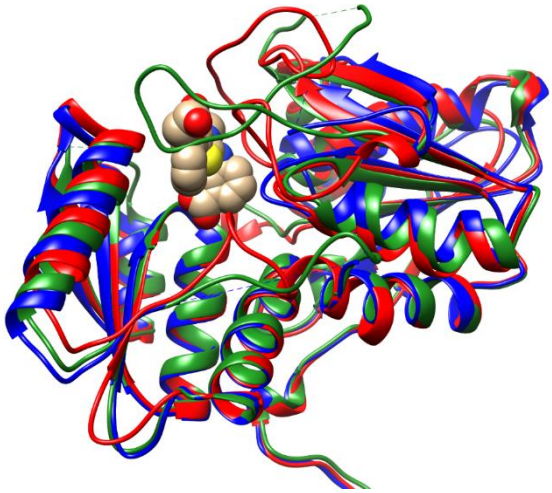


Figure 4.13 Overlay of cysteine synthase structures. Cysteine synthase (CysK1) protein structures from *M. tuberculosis* (3zei, red), *M. ulcerans* (4i1y, blue), and *M. marinum* (3rr2, green) were overlaid in Chimera using a 3 Å cutoff. The molecule presented in the sphere is a small-molecule thiazolidine inhibitor (CPK) in the crystal structure of CysK1 from *M. tuberculosis.* Reprinted from "A proteomic signature of dormancyin the actinobacterium of micrococcus luteus," by S. Mali, M. Mitchell, S. Havis, A. Bodunrin, J. Rangel, G. Olson, W. Widger, S.J. Bark, 2017, *Journal of Bacteriology*, *199(*14)*,* 206-217. Copyright [2016] by American Society for Microbiology.

Figure 4.14 Overlay of homoserine sulfhydrylase structures. O-succinylhomoserine sulfhydrylase from *M. tuberculosis* (3ndn, red) and O-acetylhomoserine sulfhydrylase from *M. marinum* (4kam, blue) were overlaid in Chimera using a 3 Å cutoff. Amino acid side chains lining the active site residues are presented in ball-and-stick form for *M. marinum* O-acetylhomoserine sulfhydrylase. Reprinted from "A proteomic signature of dormancyin the actinobacterium of micrococcus luteus," by S. Mali, M. Mitchell, S. Havis, A. Bodunrin, J. Rangel, G. Olson, W. Widger, S.J. Bark, 2017, *Journal of Bacteriology*, *199(*14)*,* 206-217. Copyright [2016] by American Society for Microbiology.

4.3.5.4 O-acetylhomoserine and O-succinylhomoserine sulfhyrylase

Several proteins directly or indirectly related to energy metabolism were upregulated in VBNC *M. luteus*. In general, the enzymes involved in alternate and anaplerotic pathways were increased, signaling a pivot away from catabolic metabolism towards anabolic construction of metabolites for stress survival. In addition to the upregulation of isocitrate lyase and malate synthase (discussed above), aldolase, another enzyme related to this pathway, was also increased. *M. luteus* aldolase has sequence homology with citrate lyase which generates oxaloacetate and acetyl-CoA from citrate. This metabolic process would complement the glyoxylate shunt by further metabolizing citrate to acetyl-CoA. Levels of aldehyde and alcohol dehydrogenases were also significantly increased. These enzymes affect glyceraldehyde-3-phosphate oxidation and lactate fermentation. Another enzyme amidohydrolase has homology with acyl-amino acid degradation enzyme which links this enzyme with ketol-acid reductoisomerase and the *O*-acetylserine and *O*-acetylhomoserine sulfhydrylases (described above) for amino acid metabolism. Overall, these data suggest that *M. luteus* under nutrient deprivation is shifting metabolism away from the complete combustion of sugars towards the utilization of carbon sources for anabolic metabolism.

4.4 Conclusions

Our optimized liquid chromatography-MS strategies allowed us to visualize proteins in all three growth phases and important proteins involved in *M. luteus* and *M. tuberculosis* dormancy. Our data at the protein level have importance when correlated to several published studies on gene regulation in *M. tuberculosis* under different stress conditions, including hypoxia and macrophage engulfment. The initial comparisons with these genetic studies were responsible for identifying UspA, isocitrate lyase, and both sulfhydrylase enzymes as potential targets for protein regulation in the *M. luteus* VBNC dormancy state. Many of the genes differentially regulated in these genetic

studies were not observed in our protein-based experiments. Among the reasons for this may be that stress conditions for these genetic studies were not equivalent to our conditions or that the dynamic range available for genomic studies is higher than that of the protein analysis technologies we employed in this study. More comprehensive mass spectrometry technologies are available once a complete proteomic analysis of a bacterial sample is completed, and these may be useful for future experiments in *M. luteus*, especially with the incorporation of multiple different sample preparations. Despite these limitations, our data provide many insights for future experiments based on the 18 proteins we rigorously identified as quantitatively increased in the VBNC state. We propose the following experiments: (i) knockout of proteins to identify causative agents and regulators for the VBNC state, (ii) proteomic analysis of alternative stress systems like hypoxia, acidity, and antibiotics to identify core proteins involved in stress responses, (iii) identification of protein interactions and signaling processes that initiate and maintain the VBNC state, (iv) proteomic and functional analysis of the proteins lost in VBNC transition and how loss of these functions affects dormancy, (v) identification of the protein signature and functional mechanisms of resuscitation, and (vi) development of novel strategies to therapeutically target proteins that drive the VBNC and persistent bacterial states and facilitate resuscitation from dormancy. These experiments will answer critically important questions about the proteins enforcing dormancy and resuscitation states and how these proteins affect bacterial survival and antibiotic tolerance. The studies provide the foundation for building a molecular picture of dormancy; they establish a baseline protein signature for a uniquely well-defined and reproducible VBNC state and mark MI-2665 *M. luteus* as a relevant reference organism for understanding the molecular basis of this critical bacterial survival mechanism.

Chapter 5:

Targeted Proteomics of Universal Stress Proteins in Micrococcus Luteus using Mass Spectrometry

5.1 Introduction

5.2 Universal stress proteins

Universal stress proteins (USPs) also known as starvation-inducible proteins are induced or upregulated in bacterial cells by different stress conditions, such as nutrient starvation, exposure to heat, oxidants, metals, uncouplers, polymyxin, cycloserine, ethanol, antibiotics, and other stimulants (Nystrom and Neidhardt, 1992). These stress conditions have led to bacterial growth arrest, a physiological state which aids the organism in surviving extreme environmental conditions. The Universal stress proteins can be found in a diverse set of organisms from archaea to bacteria, fungi, and plants. This diversity suggests their critical role in the three domains of life. Although found to be upregulated in different stress conditions, the *uspA* gene was found to be repressed under extreme temperatures and tetracycline exposure in *Salmonella typhimurium* (Liu *et al.,* 2007). USPs found in different organisms vary considerably. USP containing organisms are usually equipped with several USP genes despite exhibiting extensive similarity in the function that borders on redundancy. The *E. coli* (strain K-12) genome encodes six USP genes—UspA,-C,-D,-E,-F, and –G—their expression is triggered by many environmental stressors such as oxidative stress, heat, nutrient deprivation. These USPs have been shown to have a differential function in oxidative stress resistance, adhesion, and motility (O'Toole and Williams, 2003). *Mycobacterium tuberculosis, M. smegmatis, M. bovis, M. leprae* encodes nine, eight, five, and one USP, respectively while *Streptomyces coelicolor* is predicted to encode as many as 12 USPs (Nachin *et al.,* 2005).

The first *E. coli* USP discovered was a cytoplasmic protein with 144 amino acids. Its expression was induced in a nutrient-deprived culture media (Kvint *et al.,* 2003). Inactivation of *E. coli* UspA was found to affect bacterial survival under different growth impeding conditions such as carbon source starvation, oxidizing agents such as hydrogen peroxide, osmotic shock, low pH, heat and high salt (Tkaczuk *et al.,* 2013). UspA is thought to have DNA binding properties (Xu *et al.,* 2016). Although phenotypical changes were not observed in the global proteomic profile of inactivated UspA in *E. coli* during steady-state growth, however this mutant, exhibited a distinct alteration in global protein expression suggesting protein-protein binding interactions. The Universal Stress Protein, MJ0577 in *Methanococcus jannaschii*, forms homodimers while UspG of *E. coli* has been found in complexes with the GroEL isolated from stationary phase cultures (Bochkareva *et al.,* 2002). USPs can also be divided into ATP binding (UspFG-type in *E. coli*, RV2623 in TB) and Non-ATP binding (UspA's and UspA-like group in *E. coli* and *H. influenza*) (Nyström and Neidhardt, 1996). ATP-binding has been known to regulate bacillary growth in *Mycobacterium tuberculosis* and phosphorylated in response to stasis in *E. coli* thus reinforcing the idea of phosphorylation of serine and threonine residues in *E. coli* (Drumm *et al.,* 2009).

M. luteus is characterized by a high genomic GC content (above 70%). It has a large habitat preference including soil, hydrocarbon-impacted environments, and human skin-associated sources (Kaprylants *et al.,* 1993). Infections have been reported in patients with a lowered immunity. Genomic analysis of strains belonging to the genus *Micrococcus* can contribute to our understanding of the molecular mechanisms of opportunistic pathogenesis and infection and subsequently reduce the occurrence and/or mitigate the severity of such infections (Young *et al.,* 2010).

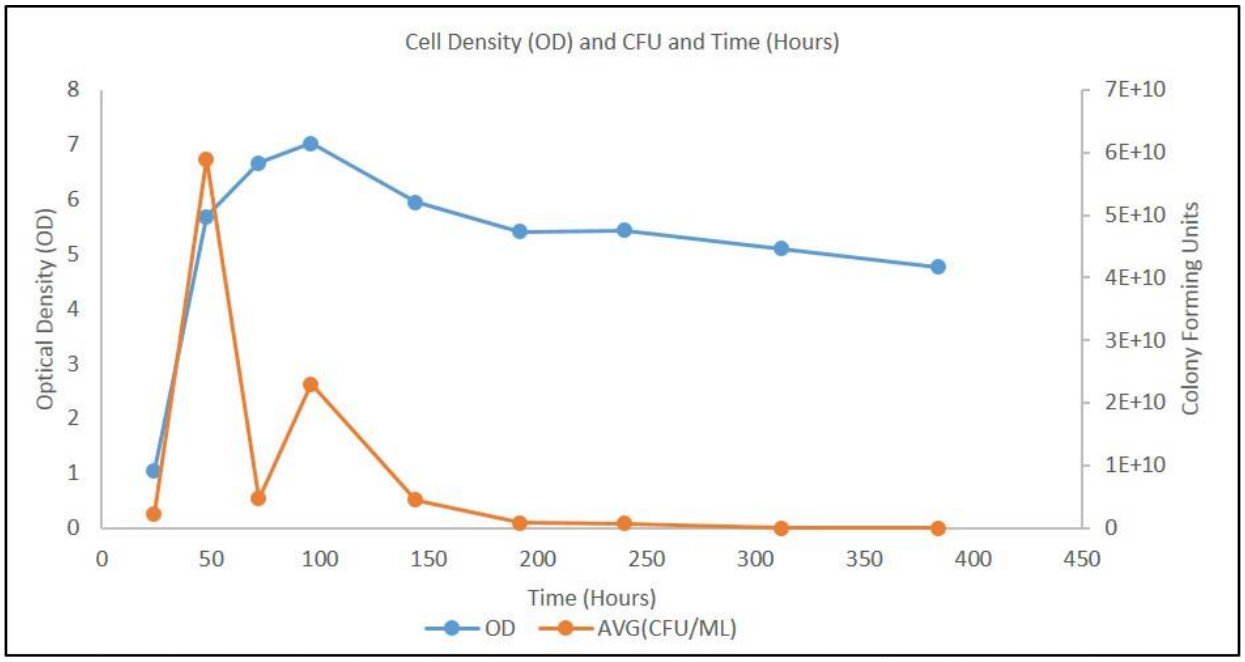


Figure 5.1 Optical density and average CFUs vs time measurement of *M. luteus*. OD is in blue and CFU is in red. Shows dormant yet viable state of MI-2665. (Figure courtesy of Jonathan Rangel and William Widger, University of Houston).

Unlike *Bacillus subtilis*, *M. luteus* does not form spores to escape stresses however they can enter a dormant state (Hanafy *et al.,* 2016). The *M. luteus* genome encodes three USPs named Usp616, Usp184 and Usp712 based on their WP numbers. USPs encode either a small Usp protein (around 14 to 15 kDa) containing only one Usp domain (Usp712 and Usp184), a larger version (around 30 kDa) consisting of two Usp domains in tandem (Usp616), or large proteins in which the Usp domain is present together with other functional domains (Kvint *et al.,* 2003). Dormancy is a protective mechanism in which bacteria, including *M. luteus* and *M. tuberculosis*, reduce metabolic activity upon encountering stress like antibiotics. This has led to latent bacterial infections, unculturable micro-organisms, multi-drug tolerance, and persisters (Wayne *et al.,* 1996). The ability of *M. tuberculosis* to enter into latency (same as dormancy but occurring in vivo)

has made TB resistant to antibiotics. Approximately one-third of the world's population is infected with latent *M. tuberculosis* while possible TB reactivation occurring in 2-23% of cases (Alnimr, 2015). With regards to *M. tuberculosis*, the terms "dormant", "latent", or "persistent" have been used interchangeably. Latency is a clinical condition in which an organism (in this case the TB bacteria) persists for many years within the host without causing the disease or showing reactivity in laboratory testing. The dormant bacilli are characterized by slow in-vitro growth, a downshift of metabolic pathways, altered staining features, an inability to be cultivated on solid media, and resistance to antimycobacterial agents (Lewis *et al.,* 2007). Persisters, on the other hand, are a minute number of slow dividing cells pre-formed in the early exponential growth stage rather than being produced in response to antibiotics (Laal, 2012). Latency and resuscitation play important roles in the pathogenicity of tuberculosis (Flynn and Chan, 2001). The long term goal of our research is to prevent bacteria from entering dormancy or prevent resuscitation from dormancy hence, providing a unique treatment of pathogenic infections.

While there is general agreement on their importance, the biological functions of universal stress proteins (USPs) have yet to be clearly defined. One reason for this situation is that most bacteria have multiple related USPs with diverse and redundant functionality (Nachin *et al.,* 2005). A consistent feature is that they facilitate resistance to external stress and long-term growth arrest. USPs are universally found in prokaryotes and there is ample evidence for phosphorylation, protein interactions, and binding of different signaling molecules including ATP, cAMP, and lipids (Kvint *et al.,* 2003). Usp signaling is also supported by *in vivo* Ser/Thr phosphorylation of the homologous UspE protein in *E. coli* and alteration of the isoelectric point of many proteins after overexpressing *E. coli* USPs (Zarembinski *et al.,* 1998). Therefore, we expect that UspA-type proteins require ATP-binding, phosphorylation, protein interactions, or a combination of these, which can be studied using biochemical and biophysical approaches.

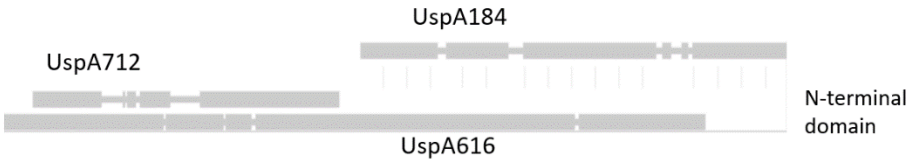


Figure 5.2 Structural alignment of Usp712, Usp616, and Usp184. The structural alignment of the domains of USPs shows a significant overlap. (Figure courtesy of Abiodun Bodunrin and William Widger, University of Houston).

The USPs of *Mycobacterium tuberculosis* and the Usp616 of *M. luteus* share a 33% sequence identity with the RV2623 of TB while Usp184 shares a 15% homology with RV2319c and RV1996.

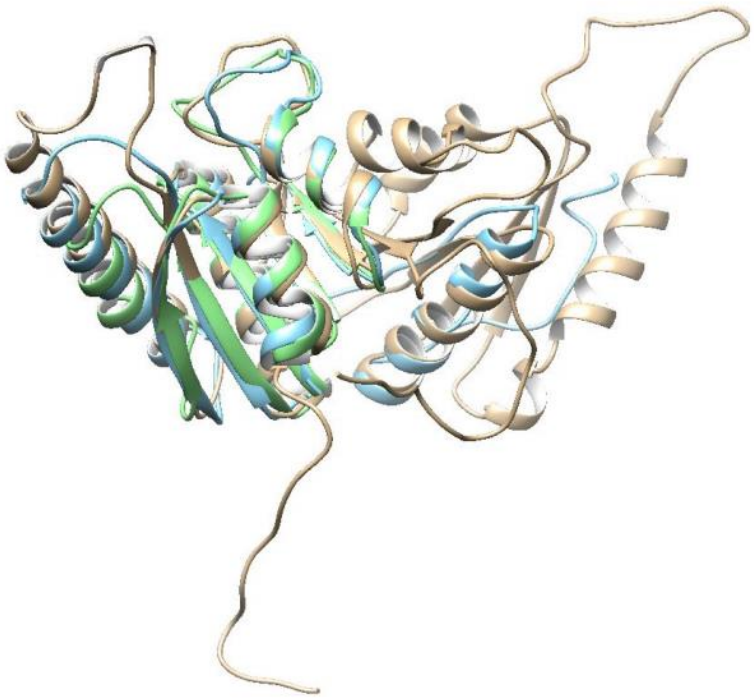


Figure 5.3 Structure alignment of the three USPs in *M. luteus*. UspA616 (grey), UspA184 (light blue) & UspA712 (light green). UspA616 and UspA184 have the walker A & B sequences that confer nucleotide-binding properties. UspA712 does not have the same sequence similarity and hence does not bind ATP. The figure was produced using Chimera.

A comparison of the conserved sequences among Usp184, UspE of *E. coli* and RV2623 in TB showed that Usp184 is closely related to the RV2623 in both domain and binding motifs.

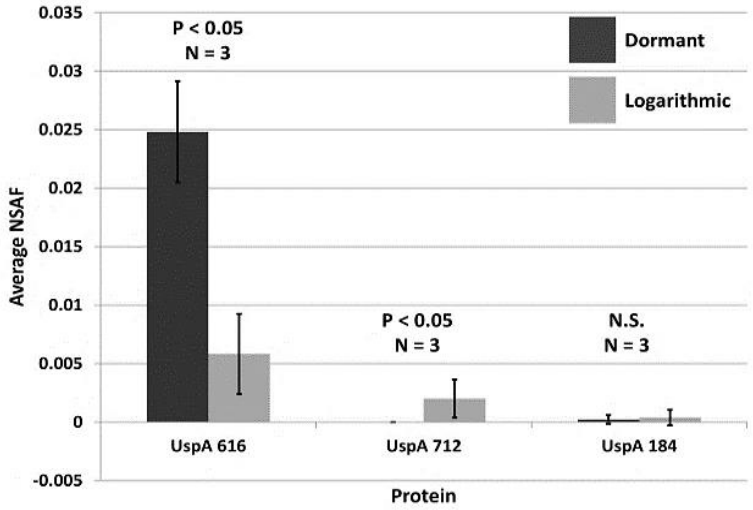


Figure 5.4 Mass spectrometry quantitation of *M. luteus* universal stress proteins (USPs) in dormant and logarithmic growth phases. Data shows the upregulation of UspA616 and the downregulation of UspA712 in dormancy. Relative abundances for UspA616, UspA712, and UspA184 were measured by Normalized Spectral Abundance Factor (NSAF) levels. Each protein was measured using three biological replicates, with each biological replicate analyzed in triplicate. P-values were <0.05 for UspA616 and UspA712 across biological replicates, but not statistically significant (N.S.) for UspA184. Adapted from "A proteomic signature of dormancyin the actinobacterium of micrococcus luteus," by S. Mali, M. Mitchell, S. Havis, A. Bodunrin, J. Rangel, G. Olson, W. Widger, S.J. Bark, 2017, *Journal of Bacteriology*, *199(*14*),* 206-217. Copyright [2016] by American Society for Microbiology.

5.2.1 Universal stress protein A in *M. luteus*

We can only infer the physiological role of the Universal Stress Proteins in *M. luteus*. A more detailed quantitative analysis of our proteomics data revealed that UspA616 is markedly increased in dormancy**,** UspA712 is more prevalent in the logarithmic state, and UspA184 was unchanged (Figure 5.5). These are the only three Usp proteins expressed by *M. luteus* compared to 9 in *M. tuberculosis* and 14 in *M. smegmatis*. Therefore, we predicted that knockout of UspA

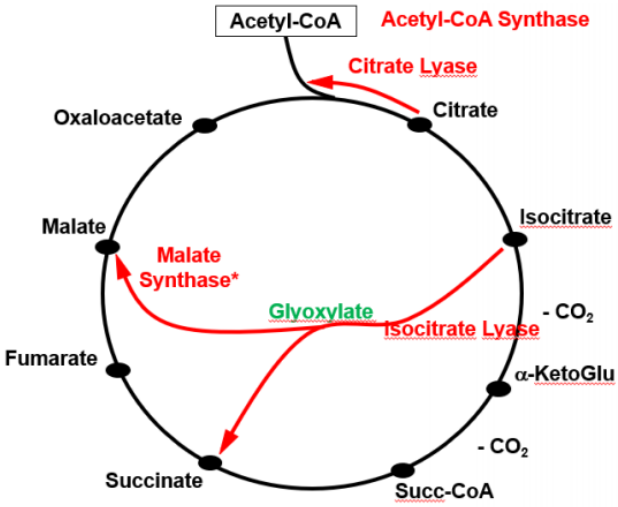


Figure 5.5 Glyoxylate shunt. Illustrates glyoxylate shunt and TCA proteins that are upregulated (red) in the VBNC state of MI-2665. Adapted from " A universal stress protein that controls bacteria stress survival in micrococcus luteus," by S. Havis, A. Bodunrin, J. Rangel, R. Zimmerer, J. Murphy, J. Storey, T. Duong, B. Mistretta, P. Gunaratne, W. Widger, S.J. Bark, 2019, *Journal of Bacteriology,* 201(24), 497-519. Copyright [2019] by American Society for Microbiology.

5.3 Role of glyoxylate shunt

Proteins in *M. luteus* would generate a more definitive phenotype than found in bacteria with many more potentially redundant Universal Stress Proteins. In *E. coli*, there are six Usp proteins with overlapping and distinct functions (Nyström and Neidhardt, 1994). Both UspA and UspD are required in the defense against superoxide-generating agents, and UspD appears also important in controlling intracellular levels of iron. In contrast, UspC is not involved in stress resistance or iron metabolism but is essential for cellular motility, like UspE (Xu *et al.,* 2016). Elucidating the role of the USPs of *M. luteus* should not be markedly dissimilar from the above functions.

We observed glyoxylate shunt enzyme isocitrate lyase upregulated in dormant *M. luteus*. This shunt is an anaplerotic pathway that bypasses the oxidative metabolism of isocitrate and ketoglutarate in the TCA cycle (Figure 5.3) (Kornberg, 1996). We proposed that this bypass of CO_2-generating steps would conserve critical carbon resources and reroute metabolism towards a synthesis of molecular species necessary for survival (Mali *et al.,* 2017). We observed a significant upregulation of isocitrate lyase in dormancy. This enzyme is also induced under hypoxic growth conditions in *M. tuberculosis* (Wayne and Lin, 1982). During hypoxia, the citric acid cycle would be downregulated since oxidative phosphorylation requires oxygen. Isocitrate lyase, along with malate synthase, catalyzes the conversion of isocitrate to malate through glyoxylate and succinate intermediates. This anaplerotic pathway bypasses two oxidative decarboxylation steps in the citric acid cycle, preserving a limited carbon pool during starvation and reducing NADH formation. We noted that malate synthase was significantly upregulated in the dormant phase in two of the three biological replicates in our study (Figure 4.5).

The *M. luteus* isocitrate lyase (Icl) structure is not known. However, the structural comparison of Icl from distant species (*M. tuberculosis* and *Aspergillus nidulans*) shows remarkable structural conservation of the enzymatic core (Figure 4.9). Multiple sequence alignment of Icl from *M. luteus* with *M. tuberculosis* and *A. nidulans* showed high sequence identity including the conserved active site residues and metal-binding sites (Figure 4.10). *M. luteus* Icl sequence had 79% identity and 87% similarity with *M. tuberculosis* and 39% identity and 54% similarity with *A. nidulans* Icl proteins. Based on the protein homology, we hypothesize that these proteins perform the same function across actinobacteria. Moreover, a significant mutation of the Icl core structure would result in a compromised enzyme that would not support survival under stress conditions including nutrient starvation. This hypothesis is consistent with the knockout of the *icl1* gene in *M. tuberculosis,* which severely impaired survival in both acute and persistent infection models.

Isocitrate lyase is critical for *M. tuberculosis* survival in hypoxia and persistence in macrophages and mice. Isocitrate lyase has been proposed to be upregulated in response to acetate metabolism, but we observed a lower quantity of this protein in minimal acetate medium until starvation (Kornberg, 1966). We can predict that the ablation of isocitrate lyase would block glyoxylate metabolism, reducing the fitness of the *M. luteus* under hypoxia and nutrient starvation, but probably not under normal growth conditions. For this reason, we would predict that knockout of isocitrate lyase would have normal logarithmic growth properties but could not survive dormancy stress. Loss of fitness would be evident by increased cell death and loss of colony-forming units when exposed to stress for extended periods. However, the mechanisms by which isocitrate lyase enables survival are unknown, so experiments relating this protein to growth, dormancy, and resuscitation phases would provide insight into the survival function of this anaplerotic pathway.

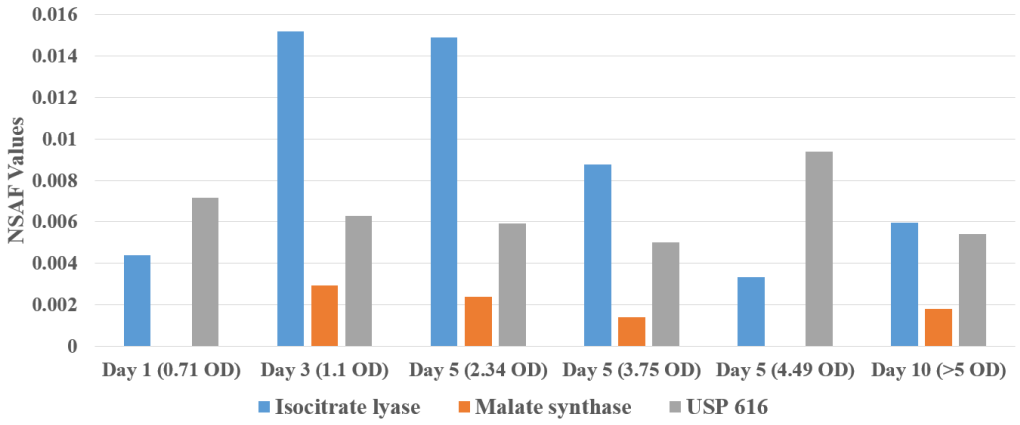


Figure 5.6 NSAF quantitation values of isocitrate lyase, malate synthase, and UspA616 in *M. luteus* with UspA712 knock-out. UspA712 KO shows a compensatory mechanism for UspA616 and upregulation of key glyoxylate shunt proteins. Three biological replicates and 3 technical replicates at each OD were recorded.

5.4 Conclusions

Previous data confirms that Usp-type proteins can interact with each other in both homodimeric and heterodimeric forms (Nachin *et al.,* 2008). The preliminary *in vivo* data presented here from UspA712 knock-out affinity purification reveals UspA712 may interact directly with UspA616, consistent with observations by Nachin, et al. The importance of this potential interaction cannot be overstated. First, UspA616 is upregulated in dormancy (Mali *et al.,* 2017). However, UspA712 is upregulated in logarithmic phase (Figure 5.4). Thus, it appears that these proteins act in opposition to each other. UspA712 knockout causes an extended lag phase growth phenotype (Abiodun Bodunrin and William Widger, personal communication). In contrast, UspA616 knockout demonstrated no significant lag phase and inability to survive external stress. Second, UspA616 is a critical factor in the VBNC state of *M. luteus* and has been identified as a metabolic switch that controls survival by regulating the glyoxylate shunt (Havis *et al.,* 2019). Under external stress, this pathway bypasses the oxidative degradation and decarboxylation steps and conserves carbon nutrients in preparation into the VBNC state. The recent paper by Havis, et al. demonstrates that isocitrate lyase and malate synthase, the two critical enzymes in the glyoxylate shunt, are regulated by UspA616. However, there is no previous data in the literature suggesting that UspA712 interacts with isocitrate lyase or malate synthase. Therfore, our UspA712 affinity purification data suggests that the regulation of isocitrate lyase and malate synthase by UspA616 may be a direct protein-protein interaction. Indeed, the data we have to date presented here and in our recent publications are consistent with this hypothesis.

Table 5.7 Protein summary of APMS of UspA712 knock-out Table displays the proteins that were co-precipitated with UspA712 knock-out. The yellow color indicates proteins that are non-background proteins. The green color indicates background proteins. The red box shows that UspA616 is present and indicates that UspA712 interacts with UspA616.

Summary View for OMSSA file: 20190122_712_1ml.omx

	Sum log E-Value	Protein Score	%Coverage	#Pep.	#Unique Pep.	Protein Description
1	-67	-258.93	34.58	12	10	WP_012751032.1 molecular chaperone GroEL [Micrococcus luteus]
2	-43.99	-169.15	37.88	9	8	WP_010080381.1 elongation factor Tu [Micrococcus luteus]
3	-23.6	-89.65	24.29	7	6	WP_010080430.1 iron ABC transporter substrate-binding protein [Micrococcus lute
4	-15.69	-60.66	14.02	3	2	WP_012750705.1 MULTISPECIES: isocitrate lyase [Bacteria]
5	-19.27	-59.63	17.05	6	6	WP_010080380.1 elongation factor G [Micrococcus luteus]
6	-12.74	-54.82	40.86	3	2	WP_002858339.1 MULTISPECIES: integration host factor [Bacteria]
7	-10.57	-45.73	8.26	1	1	WP_010080400.1 50S ribosomal protein L17 [Micrococcus luteus]
8	-12.21	-44.65	12.96	3	3	WP_010079477.1 urease subunit alpha [Micrococcus luteus]
9	-10.5	-42.18	10.44	2	2	WP_010078780.1 thiosulfate sulfurtransferase [Micrococcus luteus]
10	-9.39	-39.47	3.59	1	1	WP_010079313.1 aspartate ammonia-lyase [Micrococcus luteus]
11	-8.97	-37.76	5.42	1	1	WP_002857515.1 DNA-directed RNA polymerase subunit alpha [Micrococcus luteus
12	-9.62	-36.8	12.17	2	2	WP_010079726.1 hypothetical protein [Micrococcus luteus]
13	-9.18	-34.74	14.8	3	3	WP_010079910.1 ABC transporter substrate-binding protein [Micrococcus luteus]
14	-8.66	-33.94	12.42	2	2	WP_012750749.1 succinyl-CoA synthetase subunit alpha [Micrococcus luteus]
15	-7.84	-33.34	7.11	1	1	WP_010080279.1 Crp/Fnr family transcriptional regulator [Micrococcus luteus]
16	-10.7	-31.72	11.11	4	4	WP_010079363.1 catalase [Micrococcus luteus]
17	-6.39	-26.15	6.38	1	1	WP_010080375.1 50S ribosomal protein L1 [Micrococcus luteus]
18	-7.43	-25.07	5.58	2	2	WP_012751153.1 FAD-dependent oxidoreductase [Micrococcus luteus]
19	-7.73	-24.84	11.57	3	3	WP_010079404.1 succinyl-CoA ligase subunit beta [Micrococcus luteus]
20	-5.9	-23.78	6.51	1	1	WP_010078717.1 phosphoglycerate kinase [Micrococcus luteus]
21	-5.45	-23.04	22.73	1	1	WP_002857494.1 MULTISPECIES: 30S ribosomal protein S8 [Bacteria]
22	-5.55	-22.64	5.56	1	1	WP_010079453.1 phosphoglyceromutase [Micrococcus luteus]
23	-5.99	-22.58	1.76	1	1	WP_010079378.1 isocitrate dehydrogenase [Micrococcus luteus]
24	-5.38	-22.43	16.28	1	1	WP_010080162.1 hypothetical protein [Micrococcus luteus]
25	-5.16	-21.45	9.05	1	1	WP_002855988.1 MULTISPECIES: ATP-dependent Clp protease proteolytic subunit [
26	-5.47	-20.72	4.95	1	1	WP_010079421.1 molecular chaperone GroEL [Micrococcus luteus]
27	-4.99	-20.39	15	2	1	WP_010080408.1 large conductance mechanosensitive channel protein MscL [Micr
28	-4.88	-20.32	10	1	1	WP_012750855.1 hypothetical protein [Micrococcus luteus]
29	-4.7	-20.05	11.05	1	1	WP_012751017.1 ferritin [Micrococcus luteus]
30	-5.49	-20.04	19.23	2	2	WP_010080379.1 MULTISPECIES: 30S ribosomal protein S7 [Bacteria]
31	-4.81	-19.64	9.5	1	1	WP_010079904.1 NAD-dependent dehydratase [Micrococcus luteus]
32	-4.59	-19.56	11.63	1	1	WP_010080376.1 50S ribosomal protein L7/L12 [Micrococcus luteus]
33	-6.04	-19.46	6.8	2	2	WP_012750867.1 30S ribosomal protein S1 [Micrococcus luteus]
34	-5.08	-19.08	2.47	1	1	WP_041779816.1 peptide ABC transporter ATPase [Micrococcus luteus]
35	-5.26	-18.98	23.7	2	2	WP_002854782.1 MULTISPECIES: 50S ribosomal protein L10 [Bacteria]
36	-5.09	-18.9	2.76	1	1	WP_036335912.1 malate synthase [Micrococcus luteus]
37	-4.76	-18.76	7.69	1	1	WP_010078585.1 30S ribosomal protein S4 [Micrococcus luteus]
38	-7.42	-18.54	5.11	3	3	WP_010078557.1 aconitate hydratase [Micrococcus luteus]
39	-6.9	-18.51	8.82	3	3	WP_010079017.1 F0F1 ATP synthase subunit alpha [Micrococcus luteus]
40	-4.36	-17.26	8.42	1	1	WP_010080124.1 alanine dehydrogenase [Micrococcus luteus]
41	-4.46	-16.7	7.35	1	1	WP_002857479.1 MULTISPECIES: 30S ribosomal protein S3 [Bacteria]
42	-3.81	-14.08	4.85	1	1	WP_010080575.1 hypothetical protein [Micrococcus luteus]
43	-3.28	-13.3	6.73	1	1	WP_010078719.1 superoxide dismutase [Micrococcus luteus]
44	-3.26	-13.2	15.87	1	1	WP_010079555.1 hypothetical protein [Micrococcus luteus]
45	-3.56	-13.15	5.24	1	1	WP_012750737.1 hypothetical protein [Micrococcus luteus]
46	-2.95	-12.65	15.38	1	1	WP_002858335.1 MULTISPECIES: 50S ribosomal protein L28 [Bacteria]
47	-3.46	-11.99	3.94	1	1	WP_012751114.1 aldehyde dehydrogenase [Micrococcus luteus]
48	-2.86	-11.99	7.39	1	1	WP_010079725.1 MULTISPECIES: peptidyl-prolyl cis-trans isomerase [Bacteria]
49	-2.8	-10.76	16.22	1	1	WP_010079129.1 ribosome-binding factor A [Micrococcus luteus]
50	-2.68	-10.37	16.34	1	1	WP_010079796.1 50S ribosomal protein L9 [Micrococcus luteus]
51	-3.67	-9.58	5.88	2	1	WP_010079332.1 enolase [Micrococcus luteus]
52	-2.53	-9.15	12.2	1	1	WP_010078853.1 DSBA oxidoreductase [Micrococcus luteus]
53	-2.95	-9.14	2.5	1	1	WP_010078865.1 ABC transporter ATP-binding protein [Micrococcus luteus]
54	-2.14	-7.66	14.81	1	1	WP_010080403.1 30S ribosomal protein S9 [Micrococcus luteus]
55	-2.29	-7.48	11.64	1	1	WP_010079653.1 hypothetical protein [Micrococcus luteus]
56	-2.12	-7.13	5.37	1	1	WP_010080395.1 adenylate kinase [Micrococcus luteus]
57	-2.31	-6.06	1.63	1	1	WP_010079483.1 phosphoenolpyruvate carboxykinase [Micrococcus luteus]
58	-1.42	-5.49	21.82	1	1	WP_002858336.1 MULTISPECIES: 50S ribosomal protein L33 [Bacteria]
59	-1.74	-4.88	6.78	1	1	WP_012750914.1 cell division protein FtsZ [Micrococcus luteus]
60	-1.88	-4.38	2.76	1	1	WP_010079328.1 NADH dehydrogenase [Micrococcus luteus]
61	-1.17	-2.89	4.72	1	1	WP_010078702.1 MULTISPECIES: Fe-S cluster assembly ATPase SufC [Bacteria]
62	-1.18	-2.3	3.4	1	1	WP_010079616.1 universal stress protein UspA [Micrococcus luteus]
63	-1.41	-2.04	2.39	1	1	WP_010079359.1 hypothetical protein [Micrococcus luteus]
64	-1.15	-1.45	2.06	1	1	WP_010079015.1 ATP synthase subunit beta [Micrococcus luteus]
65	-1.22	-0.65	3.59	1	1	WP_010079121.1 polyribonucleotide nucleotidyltransferase [Micrococcus luteus]
66	-1.01	-0.03	2.34	1	1	WP_010078578.1 ATPase [Micrococcus luteus]

For example, UspA616 is upregulated in dormancy, UspA712 is upregulated in logarithmic phase, and UspA184 levels remain unchanged throughout logarithmic, stationary, and dormant phases. Because *M. luteus* only has three universal stress proteins, this model organism represents an ideal system to study and regulation of growth and dormancy by universal stress proteins. As indicated above, other well-studied bacteria have many USPs with overlapping functions (*E. coli* has 6, *M. tuberculosis*, *M. smegmatis* has 14), making study of these proteins difficult in these bacterial species. Future studies include affinity purification of UspA712 wild type to decipher normal protein interactions for comparative studies to the UspA712 knock-out affinity purification. Additionally, UspA184 knock out and wild type studies need to be completed.

Chapter 6:

Proteomic Profiling of SH2 Domain-Containing Proteins in Human Pancreatic Adenocarcinoma Cells Using Affinity Purification

Preface

The work outlined in this chapter has been submitted and accepted to the Journal of Proteome Research (Vo H, Robertson C, **Mitchell M**, Modi P, White A, Hong A, Zhu R, Zhou X, Gao X. 2020. Proteomic profiling of SH2 domain-containing proteins and cellular signaling pathways using microfluidic peptide microarray (μPepArray™). Journal of Proteome Research (in revision).

6.1 Introduction

The efficacy of targeted cancer therapy is limited due to intrinsic or acquired resistance, cross-talk, and persistence of cellular signaling pathways. Cancer genomics has provided useful information to understand the roles of kinases, their mutations in cancer, and assistance in drug design (Lal *et al.,* 2013). However, the proteomic complexity and dysregulated signaling networks derived from cancer genome alterations, reflecting therapeutic response and resistance in different cancer types, have not been fully understood. Targeting proteins individually may lead to the reactivation of compensatory pathways, thus hindering drug efficacy and conferring resistance Thus, there is a need to understand the signaling pathways and cross-talk in cancer in response to specific therapeutic agents. Such a comprehensive posttranslational picture can be used to assist in therapeutic strategy, reduce adverse drug effects and decrease anti-cancer drug resistance, thereby enhancing treatment efficacy.

Well-established technologies available for proteomic studies include mass spectrometry (MS) and reverse phase protein microarray (RPPA). MS provides a powerful tool to identify and quantify proteins based on their mass-to-charge ratio (m/z) (Aebersold *et al.,* 2003). This technology offers a sensitive, quantitative method to study cancer proteomics. However, there are drawbacks to this approach including the financial cost, and complexity in sample preparation and operation. RPPA is a protein array method that also allows multiplex detection of several hundred cellular proteins as well as their post-translational modification states. This approach is also limited by cost, time, and requirement of diverse antibodies (Krishnamoorthy *et al.,* 2013).

Cellular signaling pathways mediated by receptor tyrosine kinases (RTK) and protein kinases play essential roles in cellular physiology including growth, proliferation, migration, and apoptosis. Src Homology 2 (SH2) domains are small domains that bind to specific phosphotyrosine protein motifs and link-activated RTK to several downstream pathways. Studying of SH2 domain-containing proteins implicated in cancer signaling networks under given therapeutic conditions such as tyrosine kinase inhibition can yield significant insight into the molecular basis of drug mechanism and treatment response, thus having valuable clinical applications towards personalized anti-cancer treatment. Here, we demonstrate a microfluidic, peptide microarray biochip technology (μPepArray™) for the label-free and multiplex detection of the endogenous SH2 domain-containing proteins and signaling pathways from total protein lysate. The biochip contains 3968 isolated reaction chambers in which the phosphopeptides derived from the phosphotyrosine motifs of protein kinases and phosphoproteins are used as the molecular probes to capture the present SH2 domain proteins. Our detection method utilizes GRB2 as a molecular reporter, which ubiquitously interacts with the SH2 domain protein and phosphopeptide probe signaling complex(es). Using pancreatic cancer BxPC-3 cells treated with different tyrosine kinase inhibitors (TKI) targeting the epidermal growth factor receptor (EGFR), we identified the differentially expressed SH2 domain proteins such as SHP2, PLCG1, PI3K, and therefore the

downstream phosphotyrosine-mediated signaling events occurred in response to TKI treatments. Our findings revealed a signaling cross-talk and compensation between the EGFR/PI3K/Akt and ERK pathways as a mechanism of EGFR inhibitor response and potential resistance in the pancreatic cancer cells. Our findings also highlighted potential roles for PKCα and mTORC2 in ERK signaling reactivation to confer drug resistance. We propose that the proteomic and signaling pathway information provided by μPepArray™ proteomic technology could hold clinical significance by assisting in the cancer biomarker identification and guiding future treatment decisions. Our method uses total cell lysate samples without purification and the detection does not require a panel of antibodies, thus alleviating time and financial restraints. In the present work, we aimed to elucidate the mechanisms underlying the drug response in pancreatic cancer cells that undergo molecularly targeted therapy by using μPepArray™ for SH2 domain protein and signaling pathway profiling.

6.2 Materials and methods

6.2.1 Cell culture, inhibitors, chemicals and antibodies

Pancreatic cancer BxPC-3 cells were purchased from the American Type Culture Collection. The cells were grown in DMEM F12 media (10% FBS) at 37 ºC and supplied with 5% CO_2. Tyrosine kinase inhibitors (Erlotinib, Afatinib, AZD9291) were purchased from Selleckchem. Thiazolyl Blue Tetrazolium Bromide (MTT agent), β-glycerophosphate disodium salt hydrate (BPG), 4-(2-Hydroxyethyl) piperazine-1-ethane sulfonic acid (HEPES), sodium pyrophosphate tetrabasic decahydrate, sodium fluoride (NaF), sodium orthovanadate, phenylmethanesulfonyl fluoride (PMSF), tris (2-carboxyethyl) phosphine hydrochloride (TCEP), 0.4% Trypan Blue solution, and phosphatase inhibitor cocktail tablets (PhosSTOP) were purchased from Sigma-Aldrich. EGFR, p-EGFR, HER2, p-HER2, GRB2, PI3K, Akt, p-Akt (Ser473), mTOR, p-mTOR (Ser2448), S6, p-

S6, SHP2, PLCG1, PKCα, p- PKCα (Thr638/641), MEK1/2, p-MEK1/2(Ser217/221), ERK1/2, p-ERK1/2 (Thr202/Tyr204), β-Actin, horseradish peroxidase-conjugated secondary antibodies were purchased from Cell Signaling Technology. Anti-Rabbit IgG (H+L) secondary antibody- Alexa Fluor 594, SuperSignal West Pico Chemiluminescent Substrate were purchased from Thermo Scientific.

6.2.2 Immunoblotting

The cells were washed with cold PBS three times and lysed using 2X Laemmli buffer (4% SDS, 10% β-mercaptoethanol, 20% glycerol, 0.004% bromophenol blue, 0.125 M Tris HCl, pH 6.8). The lysate was centrifuged at room temperature, 13500 rpm for 20 min. Supernatants were filtered through a 0.22 µM membrane. Total protein concentration was determined by a Nanodrop spectrophotometer using an absorbance of 280 nm. SDS-PAGE gels were loaded with 35 µg of total protein per lane. After separation, proteins were transferred onto PVDF membranes, blocked in 5% BSA for 1 h at room temperature, washed three times with TBST (0.01 M Tris HCl, 0.15 M NaCl, 0.05% Tween-20) and incubated with primary antibodies (1:2000 dilution) overnight at 4 ^{o}C. After washing three times with TBST, the membranes were incubated with horseradish peroxidase-conjugated secondary antibodies (1:5000 dilution) for 1 h at room temperature. Membranes were washed three times with TBST and incubated with an enhanced chemiluminescent substrate for 3 min at room temperature and developed on X-ray films.

6.3 Results

Our first goal was to identify phosphotyrosine peptide-binding proteins containing SH2 domains. A silicone surfaced immobilized phosphotyrosine containing peptides was used to capture the

SH2 domain-containing proteins in the microfluidic apparatus setting. We delegated PI3K as one of the candidate proteins with the desirable SH2 domain. Then, BxPC-3 (human pancreatic adenocarcinoma) cell lysate is used to pull down the PI3K using a biotinylated phosphor-tyrosine peptide and resin conjugated to an anti-biotin antibody. We initially attempted to use streptavidin resin but that only resulted in failed elution, as the incredibly strong streptavidin-avidin bond could not be hindered, even when boiled to 90 ºC (Figure 6.2).

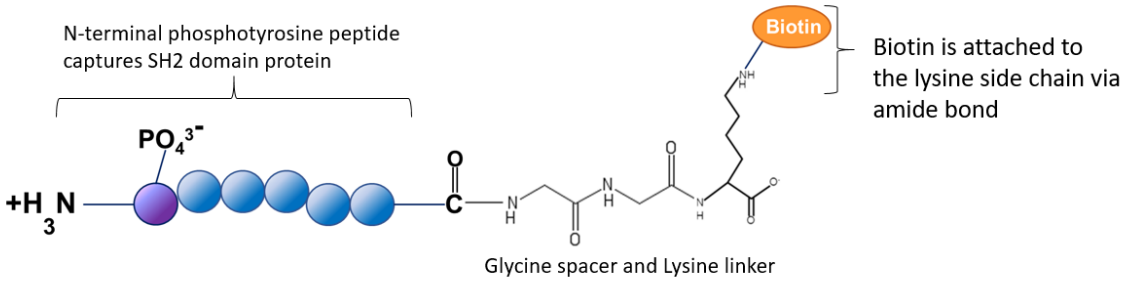


Figure 6.1 Construction of C-terminal biotinylated peptide. The peptide is pYMPMNQGG-K(biotin)-amide. The main peptide is six amino acids long which contains an N'-phosphotyrosine in a similar fashion as the peptides on the chip. The sequence of the peptide is chosen based on the peptide that shows significant differences between treated and untreated cells (previously identified by peptide array). Two additional glycine residues are added toward the C-terminal as a spacer. A C-terminal lysine is used as a biotin conjugate. The biotin is attached to the lysine side chain via an amide bond. Lysine's positive charge is removed. pY = phosphotyrosine, N-term: free amine, C-term: K(biotin)-amide, purity: peak area by HPLC ≥ 95%, quantity (gross peptide): 5 mg.

6.4 Conclusions

Signaling proteins containing the SH2 domains are heavily involved in tyrosine kinase signaling pathways that play critical roles in cellular physiology. Their interaction with GRB2 adaptor protein has been reported as a molecular probe for the detection of signaling pathway activation (Saito *et al.,* 2010). In this study, we have described a proteomic, biochip-technology based on a

microfluidic peptide microarray for the label-free, high-throughput profiling of SH2 proteins in the cellular lysate. The detection method utilizes an antibody that recognizes GRB2 bound to the signaling complex of SH2 proteins, which have been captured by the peptide probes. The identity of the protein is determined based on a specific peptide sequence that it interacts with.

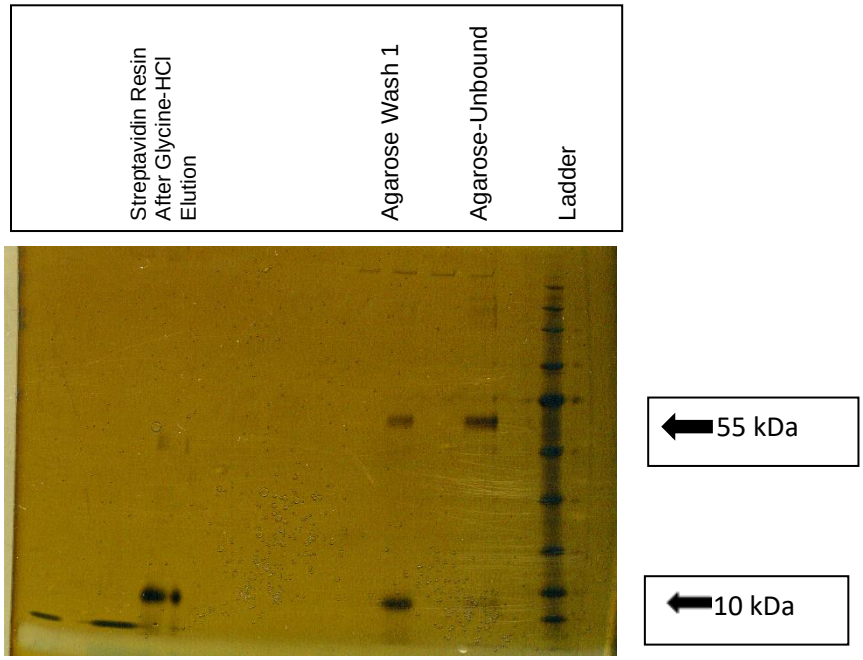


Figure 6.2 Affinity purification of PI3K on agarose and streptavidin resins. Lane 1: empty Lane 2: ladder Lane 3: empty Lane 4: agarose unbound (proteins that did not bind to the resin) Lane 5: streptavidin unbound. Lane 6: agarose wash 1 (PBS, 0.1 M phosphate, 0.15 M NaCl, pH 7.2) Lane 7: streptavidin wash 1 Lane 8: agarose glycine-HCl elute (0.1 M glycine-HCl, pH 2.8) Lane 9: agarose guanidine-HCl elute (8 M guanidine-HCl, pH 1.5) Lane 10: streptavidin glycine-HCl elute Lane 11: streptavidin guanidine-HCl elute Lane 12: agarose glycine-HCl resin after elution Lane 13: agarose guanidine-HCl resin after elution Lane14: streptavidin glycine-HCl resin after elution Lane 15: streptavidin guanidine-HCl resin after elution. Protein is not eluting under the two conditions that we used, which were specified in the protocol.

GRB2 has been reported to interact with a wide range of signaling proteins implicated in major cell signaling pathways (Bisson *et al.,* 2011). Consistently many of these pathways were identified

in our novel method of detection. Unlike other proteomic methods that require a panel of antibodies, our technology allows the detection of multiple PPBD proteins at the same time using only one anti-GRB2 primary antibody and fluorescent-conjugated secondary antibody. Also, signaling protein detection was performed in microfluidic, distinct reaction chambers with consistent, reprogrammed flow rate and controlled temperature, which minimize the risk of sample denaturation, oxidation, and degradation.

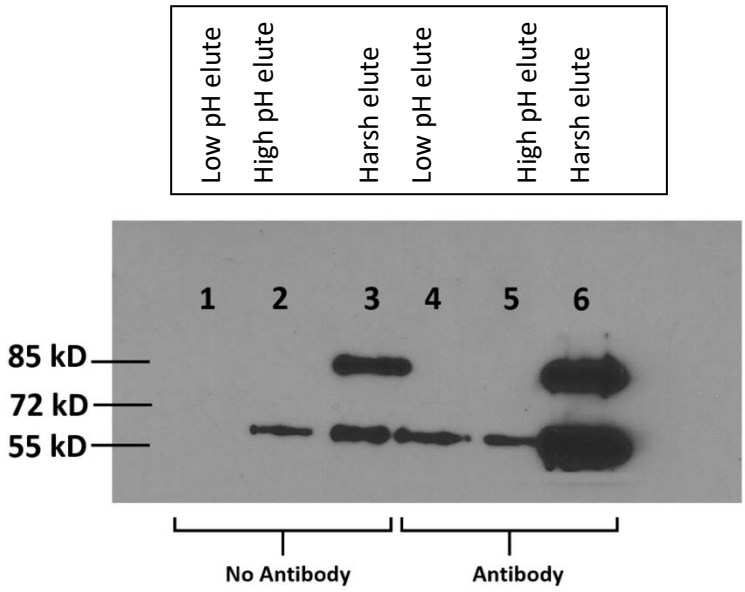


Figure 6.3 Western blot of PI3K with and without anti-biotin antibody. Lane 1: without antibody, low pH elution. Lane 2: without antibody, high pH elution. Lane 3: without antibody, harsh elution. Lane 4: with antibody, low pH elution. Lane 5: with antibody, high pH elution. Lane 6: with antibody, low pH elution. Low pH: 0.1 M glycine-HCl, pH 2.8. High pH: 5% Ammonium hydroxide, pH 12. Harsh elution conditions: 2% SDS, 30 mM biotin, 50 mM phosphate, 100 mM NaCl, 6 M Urea, 2 M thiourea, pH 12. Phosphatidylinositol 3-kinase is composed of an 85 kDa regulatory subunit and a 110 kDa catalytic subunit. 72 kDa is biotin.

Our future studies aim to overcome some technical limitations to improve protein detection and analysis. One limitation of the technology is that the peptide library was designed based on known

annotations from various publications and publicly available databases of existing protein-protein interaction.

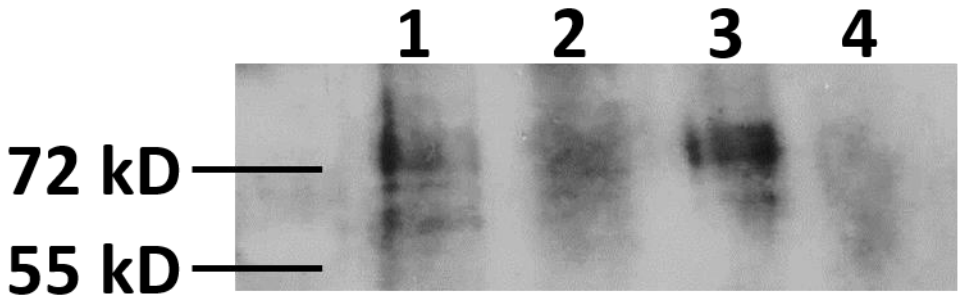


Figure 6.4 Affinity purification of PI3K on AminoLink resin. Lane 1**:** elute, no peptide, AminoLink resin, anti-biotin antibody, harsh elution. Lane 2: empty, no protein Lane 3: elute, with peptide, AminoLink resin, anti-biotin antibody, harsh elution Lane 4: empty, no protein. The lyophilized protein sample was reconstituted by adding 80 µL of Tris 30 mM, pH 8. Before loading on the gel, 30 µL of the sample was mixed with 30 µL of Laemalie loading buffer and heated at 90 ^{o}C for 7 min. 38 µL of each sample was loaded in each lane. 72 kDa is biotin.

Changes in the signal from the chip represent a combination of the change in protein expression and the change in peptide-protein binding. Additionally, there is the possibility that an unknown protein binding to the pY peptide has not been identified, precluding detection. It is therefore critical that the analysis proceeds with stringent validation. Our previous studies using affinity-tagged recombinant SH2 proteins including GRB2 have shown that PPBD protein binding to their relevant pY peptides on the chip has high specificity and affinity. The results were consistent with the reported literature (Krishnamoorthy *et al.,* 2013). In this current study, we used conventional western blot to validate the endogenous protein expression in different EGFRi treatments. Such validation methods are specific and reliable but suffer from being low-throughput and time-consuming. Our future studies, therefore, aim to improve our peptide library design by utilizing

other high-throughput proteomic technology such as mass spectrometry to experimentally validate all those known peptide-protein interactions of interest. Since the technology uses total cellular lysate, the peptide and PPBD protein interaction may also involve multiple protein binding. This is similar to the recruitment and formation of a signaling complex in the cell at the phosphotyrosine motifs. Identifying these proteins and their interactome under specific conditions or in different cancer subtypes would yield valuable insights into the underlining mechanism that requires signaling multicomplex assembly. A well-validated peptide library will, therefore, allow better and faster high-throughput readout of protein expression changes and hold valuable implications in proteomic-based diagnosis and treatment monitoring. In the present study, this technology allows us to monitor the differential expression of SH2 domain proteins and phosphoproteins in the pancreatic cancer cells after treatment with EGFR inhibitors. The phosphopeptide probe and functional cellular protein information provide a molecular signature, which reflects the cancer status and allows assessment of molecularly targeted treatment efficacy including kinase inhibitor response. µPepArray™ proteomic and signaling pathway results thus hold implications for the biomarker identification in pancreatic cancer therapy, as well as facilitating molecularly targeted cancer treatment. The peptide array results, therefore, provide biomarkers for the detection of a signaling cross-talk between PI3K/Akt/mTORC1 pathway and ERK pathway as a response to EGFR inhibition in pancreatic cancer cells.

We also detected an increased SHP2 expression and suggested its role in drug response. SHP2 is a tyrosine phosphatase encoded by the gene PTPN11, a proto-oncogene reported in different types of human cancer including leukemia, lung, laryngeal, cervical, and breast (Chan and Feng, 2007). SHP2 involves the cellular signaling downstream of growth factor receptor and promotes MAPK signaling through various mechanisms to support tumor cell growth, proliferation, invasion, metastasis, survival, and suppression of apoptosis. Future studies to elucidate the functional

significance of SHP2 and targeting this protein in pancreatic cancer treatment would be of our interest.

The cross-talking between PI3K/mTOR and Ras-ERK pathways influences each other's signaling activity and plays important roles in cancer cell survival, proliferation, metabolism, and motility. This cross-talking varies under specific conditions in each cancer subtype. Different mechanisms include cross-inhibition, cross-activation, and common substrates between the two pathways (Mendoza *et al.,* 2011). Detection of signaling pathway activities in cancer under specific stimulation or drug treatment would offer a promising approach to assist therapeutic strategy, intervention, and improve treatment efficacy. Our current study used BxPC-3 pancreatic cancer cells that lack KRAS mutations leading to constitutively active MAPK signaling. Interestingly, MAPK signaling pathway reactivation was observed as a response mechanism to EGFRi, as mediated by PKCα, SHP2, PLCG1, and mTORC2. Our studies using a combination of EGFR and mTORC2 inhibitors suggest a potential therapeutic approach for the treatment of pancreatic cancer. In conclusion, our peptide microarray technology uncovers the post-translational signaling pathway signatures in the pancreatic cancer cells as functions of drug response, thus providing biomarkers for the determination of signaling cross-talk, prediction of therapeutic responsiveness, and possible therapeutic combinations. This will benefit the treatment of different cancer types with specific inhibitors, thus facilitating personalized anti-cancer therapy.

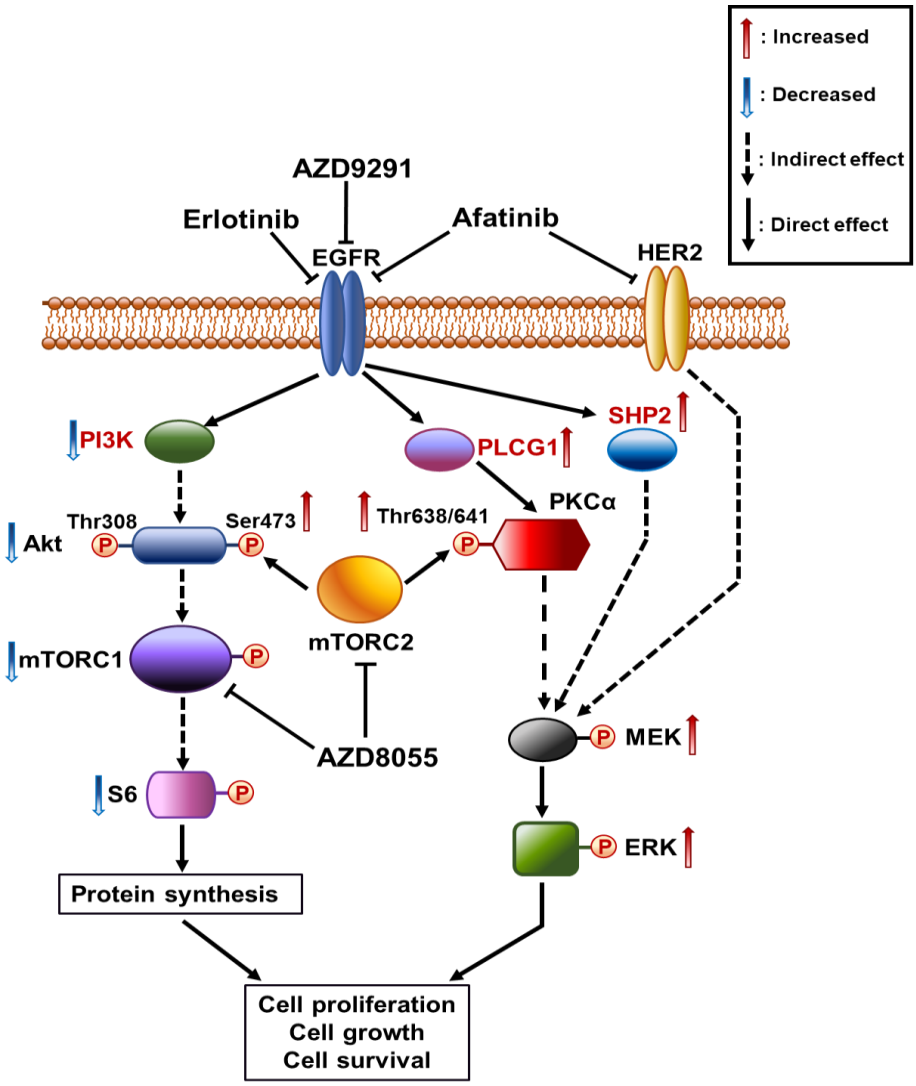


Figure 6.5 Proposed mechanism of kinase inhibitor response and signaling pathways in pancreatic cancer cells. The EGFR inhibitors inhibit EGFR phosphorylation, leading to the reduced downstream PI3K/Akt/mTORC1 signaling while inducing ERK signaling activation through the upregulation of SHP2, PLCG1, and mTORC2 activity. Afatinib is a dual inhibitor of EGFR and HER2 while Erlotinib and AZD9291 only inhibit EGFR. AZD8055 is a dual mTORC1/mTORC2 inhibitor that synergistically inhibits the pancreatic cancer cells in combination with the EGFR inhibitor.

Chapter 7:

Summary and Future Directions

7.1 Mass spectrometry instrument optimization

Matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometers are used for the analysis of biologically relevant molecules in diverse fields including pathogen identification, imaging mass spectrometry, and natural products chemistry. The identification of these molecules relies on instruments that can achieve remarkably high resolution and accuracy. Despite high nominal resolution and accuracy, we have observed significant variability where 30–50% of individual replicate measurements have errors above 5 parts-per-million, even when using 5-point internal calibration. Our data suggest a major component of variability is the binning of the raw flight time data by the electronics and clock speed of the analog-to-digital (AD) detection system, which requires interpolation by automated peak fitting algorithms and impacts both calibration and the observed mass spectrum. Importantly, the variation observed is predominantly normal in distribution, which implies multiple components contribute to the observed variation and suggests a method to mitigate this variability through spectrum averaging. Restarting the acquisition impacts each spectrum within the electronic error of the AD detector system and defines a new calibration function. Therefore, averaging multiple independent spectra and not a larger number of laser shots leverages this inherent binning error to mitigate variability in accurate MALDI-TOF mass measurements.

7.2 Affinity purification optimization

Co-affinity purification mass spectrometry is a highly effective method for identifying protein complexes from a biological sample and inferring important interactions, but the impact of the solid-support is usually not considered in the design of such experiments. Affinity purification experiments typically utilize a bait protein expressing a peptide tag such as FLAG, c-Myc, HA or V5 and high-affinity antibodies to these peptide sequences to facilitate isolation of a bait protein to co-purify interacting proteins. We observed significant variability for isolation of tagged bait proteins between Protein A/G Agarose, Protein G Dynabeads, and AminoLink resins. While previous research identified the importance of tag sequence and their location, crosslinking procedures, reagents, dilution, and detergent concentrations, the effect of the resin itself has not been considered. Our data suggest the type of solid-support is important and, under the conditions of our experiments, AminoLink resin provided a more robust solid-support platform for affinity purification mass spectrometry. Our studies suggest that the matrix used to immobilize the antibody is an important component for the success of AP-MS. We evaluated multiple resin types including epoxy Dynabeads, protein G Dynabeads, protein A/G agarose, and AminoLink resin. While all resins performed effectively under appropriate conditions, non-optimal conditions produced high variability and non-specific protein binding in the AP-MS experiments. Variability was also observed arising from the tag and protein expression, buffers, reagents, and conditions for binding, washes, and elution as previously observed in other laboratories. Our data suggest that, for affinity purification, AminoLink resin required less optimization for effective recovery of tagged proteins. Therefore, this resin may be useful for initial experimental trials when no other information is available.

7.3 Identification of bacterial dormancy signature proteins in *M. luteus* using AP-MS

Major pathogenic bacteria can survive severe external stresses including host immunity, starvation, and antibiotics by entering a reduced metabolic state called dormancy. Many non-pathogenic bacteria also use dormancy and dormancy mechanisms for survival. Because actively growing bacteria are most susceptible to external stresses, the reduced metabolic activity associated with the dormant (latent) state is a highly effective survival strategy. Recent evidence suggests that dormant bacteria contribute to the evolution of authentic antibiotic resistance by allowing bacteria to survive long enough to acquire antibiotic resistance genes, which implies that dormancy is “hardwired” into diverse bacterial species. Dormant bacteria can remain in this protective state until suitable growth conditions signal resuscitation from dormancy, thus reactivating active transmissible infection years or even decades later. Despite the obvious importance of dormant bacteria, we understand very little about the proteins and molecular mechanisms that enable bacteria to become dormant or resuscitate from dormancy. If we knew the proteins and molecular mechanisms that drive dormancy, we would understand this important mechanism of bacterial survival; knowledge that would completely transform our vision of how bacteria adapt to external stresses, how bacteria may evolve better survival mechanisms in the future, and how to combat dormant bacteria that plague over a quarter of the world’s population. To begin addressing this lack of knowledge, we anticipated that a bacterium with a verifiable dormancy state would have specific proteins upregulated in response to external stress that would be important for initiating dormancy or resuscitation survival mechanisms. *M. luteus* is one of the few bacterial species with a well-defined and reproducible dormancy state that is amenable to molecular studies. Additionally, *M. luteus* also has Resuscitation Promoting Factor (Rpf), a secreted protein reported to reactivate dormant bacteria including *M. luteus* and various Mycobacteria. These unique properties allow us to know we are studying dormancy and enable linking proteins and molecular studies to biological functions and dormancy. Our proteomic

studies in *M. luteus* successfully identified 18 proteins consistently upregulated in dormancy. These proteins have direct homologs with highly similar sequences and structures in other gram-positive and even gram-negative bacterial species. Universal Stress Protein A (UspA), isocitrate lyase (Icl), and cysteine synthase (CysM) have been correlated to *M. tuberculosis* latency, which is assumed to indicate dormancy. Ten of the remaining proteins are homologs to stress response proteins not previously correlated with dormancy. We hypothesize that the proteins upregulated in dormant *M. luteus* are required for dormancy and survival under external stress. We predict that the ablation of proteins that cause dormancy should block bacterial survival.

7.4 Targeted proteomics of universal stress protein A 712 *M. luteus* using AP-MS

UspA712 knock-out affinity purification reveals UspA712 may interact directly with UspA616. However, UspA712 is upregulated in logarithmic phase. UspA712 knockout causes an extended lag phase growth phenotype (Abiodun Bodunrin and William Widger, personal communication). In contrast, UspA616 knockout demonstrated no significant lag phase and inability to survive external stress. Also, UspA616 is a critical factor in the VBNC state of *M. luteus* and has been identified as a metabolic switch that controls survival by regulating the glyoxylate shunt (Havis *et al.,* 2019). Under external stress, this pathway bypasses the oxidative degradation and decarboxylation steps and conserves carbon nutrients in preparation into the VBNC state. The recent paper by Havis, et al. demonstrates that isocitrate lyase and malate synthase, the two critical enzymes in the glyoxylate shunt, are regulated by UspA616. However, there is no previous data in the literature suggesting that UspA712 interacts with isocitrate lyase or malate synthase. Therfore, our UspA712 affinity purification data suggests that the regulation of isocitrate lyase and malate synthase by UspA616 may be a direct protein-protein interaction. Finally, there are only 3 different known classes of universal stress proteins and the

three *M. luteus* USPs (UspA616, UspA712, and UspA184) are unique members of each class. Based on this observation, we hypothesized in earlier manuscripts (Mali *et al.,* 2017, Havis *et al.,* 2019) that *M. luteus* would not exhibit the redundancy observed in Usp studies in other bacterial models. Future studies include affinity purification of UspA712 wild type to decipher normal protein interactions for comparative studies to the UspA712 knock-out affinity purification. Additionally, UspA184 knock out and wild type studies need to be completed.

LIST OF PUBLICATIONS

Mitchell M, Mali S, King CC, Bark SJ. 2015. Enhancing MALDI time-of-flight mass spectrometer performance through spectrum averaging. **PLoS One**. 23;10(3):e0120932.

Mali S, **Mitchell M**, Havis S, Bodunrin A, Rangel J, Olson G, Widger WR, Bark SJ. 2017. A Proteomic Signature of Dormancy in the Actinobacterium Micrococcus luteus. **Journal of Bacteriology**. 199(14). pii: e00206-17.

Mali S, Moree WJ, **Mitchell M**, Widger W, Bark SJ. 2016. Observations on different resin strategies for affinity purification mass spectrometry of a tagged protein. **Analytical Biochemistry**. 15;515:26-32.

Manuscripts

Vo H, Robertson C, **Mitchell M**, Modi P, White A, Hong A, Zhu R, Zhou X, Gao X. 2020. Proteomic profiling of SH2 domain-containing proteins and cellular signaling pathways using microfluidic peptide microarray (μPepArray™). **Journal of Proteome Research** (in revision).

APPENDICES

Table 1 Raw data for standard peptide measurements

Raw data for standard peptide measurements				
RHPYFYAPELLYYANK				
Obs. Mass	Intensity	S/N	Δ(amu)	Δ(ppm)
2045.0316	633.7260	561.0000	0.0036	1.7604
2045.0383	1118.2000	732.0000	0.0103	5.0366
2045.0219	390.9020	453.0000	-0.0061	-2.9828
2045.0319	639.4900	672.0000	0.0039	1.9071
2044.9935	2036.0800	1102.0000	-0.0345	-16.8702
2045.0161	1871.6900	1061.0000	-0.0119	-5.8190
2045.0286	960.6280	682.0000	0.0006	0.2934
2045.0248	281.7250	236.0000	-0.0032	-1.5648
2045.0099	811.2940	522.0000	-0.0181	-8.8507
2045.0137	553.4120	462.0000	-0.0143	-6.9926
2045.0248	1295.0600	615.0000	-0.0032	-1.5648
2045.0089	2138.3500	1084.0000	-0.0191	-9.3397
2045.0294	1817.1000	1073.0000	0.0014	0.6846
2044.9933	2977.8800	1373.0000	-0.0347	-16.9680
2045.0239	1626.3500	1180.0000	-0.0041	-2.0049
2045.0148	2976.0000	1347.0000	-0.0132	-6.4547
LGEYGFQNALIVR				
Obs. Mass	Intensity	S/N	Δ(amu)	Δ(ppm)
1479.7771	616.7840	322.0000	-0.0183	-12.3666
1479.8064	1926.9000	585.0000	0.0110	7.4335
1479.7968	70.2745	33.0000	0.0014	0.9461
1479.7969	217.0980	145.0000	0.0015	1.0137
1479.7762	2511.0600	773.0000	-0.0192	-12.9748
1479.7689	3049.4100	1064.0000	-0.0265	-17.9079
1479.7799	532.0780	323.0000	-0.0155	-10.4744
1479.7697	764.8630	533.0000	-0.0257	-17.3673
1479.7953	1262.4300	486.0000	-0.0001	-0.0676
1479.7944	1064.1600	457.0000	-0.0010	-0.6758
1479.7815	1686.5900	461.0000	-0.0139	-9.3932
1479.7906	555.2940	157.0000	-0.0048	-3.2437
1479.7919	688.3140	328.0000	-0.0035	-2.3652
1479.8104	507.6080	125.0000	0.0150	10.1365
1479.8003	2870.5900	749.0000	0.0049	3.3113
1479.8043	2751.3700	736.0000	0.0089	6.0143

Table 1 Continued

DAFLGSFLYEYSR				
Obs. Mass	Intensity	S/N	Δ(amu)	Δ(ppm)
1567.7467	3195.6100	1698.0000	0.0040	2.5514
1567.7439	3013.0200	2029.0000	0.0012	0.7654
1567.7460	6049.8800	2539.0000	0.0033	2.1049
1567.7473	4340.0800	1865.0000	0.0046	2.9342
1567.7498	12876.5000	4012.0000	0.0071	4.5288
1567.7433	6143.3700	2600.0000	0.0006	0.3827
1567.7288	14824.8000	3303.0000	-0.0139	-8.8663
1567.7440	13323.9000	4164.0000	0.0013	0.8292
1567.7418	11107.1000	2983.0000	-0.0009	-0.5741
1567.7306	9724.2400	3854.0000	-0.0121	-7.7181
1567.7367	614.2750	456.0000	-0.0060	-3.8272
1567.7588	1517.1800	742.0000	0.0161	10.2695
1567.7484	564.7060	408.0000	0.0057	3.6358
1567.6989	203.2940	156.0000	-0.0438	-27.9383
1567.7388	3209.4100	1556.0000	-0.0039	-2.4877
1567.7219	428.5490	265.0000	-0.0208	13.2675
KVPQVSTPTLVEVSR				
Obs. Mass	Intensity	S/N	Δ(amu)	Δ(ppm)
1639.9293	1487.0600	542.0000	-0.0084	-5.1221
1639.9491	1701.0200	455.0000	0.0114	6.9515
1639.9393	198.9020	87.0000	0.0016	0.9756
1639.9106	1242.3500	372.0000	-0.0271	-16.5250
1639.9260	371.4510	203.0000	-0.0117	-7.1344
1639.9492	181.9610	92.0000	0.0115	7.0125
1639.9453	404.0780	212.0000	0.0076	4.6343
1639.9283	1813.3300	410.0000	-0.0094	-5.7319
1639.9274	997.6470	344.0000	-0.0103	-6.2807
1639.9115	1619.4500	577.0000	-0.0262	-15.9762
1639.9308	1723.6100	664.0000	-0.0069	-4.2075
1639.9518	1286.9000	441.0000	0.0141	8.5979
1639.9446	1273.7300	418.0000	0.0069	4.2075
1639.9255	2705.5700	273.0000	-0.0122	-7.4393
1639.9404	2628.3900	627.0000	0.0027	1.6464
VPQVSTPTLVEVSR				
Obs. Mass	Intensity	S/N	Δ(amu)	Δ(ppm)
1511.8123	93.4090	13.0000	-0.0305	-20.1741
1511.8214	114.1960	19.0000	-0.0214	-14.1549
1511.8217	75.9216	16.0000	-0.0211	-13.9565
1511.8177	213.9610	39.0000	-0.0251	-16.6023
1511.8218	72.7843	23.0000	-0.0210	-13.8903
1511.8285	225.8820	40.0000	-0.0143	-9.4587

Table 1 Continued				
1511.8440	133.0200	26.0000	0.0012	0.7937
1511.8293	80.3137	25.0000	-0.0135	-8.9295
1511.8754	65.8824	22.0000	0.0326	21.5631
1511.8131	62.7451	17.0000	-0.0297	-19.6449
1511.7979	77.1765	24.0000	-0.0449	-29.6989
1511.8529	300.5490	40.0000	0.0101	6.6806
1511.8536	175.6860	16.0000	0.0108	7.1436
1511.8566	253.4900	21.0000	0.0138	9.1279
1511.7921	115.4510	13.0000	-0.0507	-33.5352
1511.8289	128.6270	16.0000	-0.0139	-9.1941
GLSDGEWQQVLNVWGK				
Obs. Mass	Intensity	S/N	Δ(amu)	Δ(ppm)
1815.9091	1092.3900	463.0000	0.0067	3.6896
1815.9031	399.0590	208.0000	0.0007	0.3855
1815.9071	1638.2700	659.0000	0.0047	2.5882
1815.9006	2598.2700	728.0000	-0.0018	-0.9912
1815.8884	2239.3700	590.0000	-0.0140	-7.7097
1815.8798	862.7450	245.0000	-0.0226	-12.4456
1815.9034	1567.3700	734.0000	0.0010	0.5507
1815.8903	3737.1000	881.0000	-0.0121	-6.6634
1815.8629	249.0980	119.0000	-0.0395	-21.7523
1815.8971	154.9800	71.0000	-0.0053	-2.9187
1815.9027	140.5490	64.0000	0.0003	0.1652
1815.8977	1240.4700	305.0000	-0.0047	-2.5882
1815.9015	1169.5700	294.0000	-0.0009	-0.4956
1815.8918	2152.1600	584.0000	-0.0106	-5.8373
1815.8987	373.3330	137.0000	-0.0037	-2.0376
1815.9082	410.9800	176.0000	0.0058	3.1940
VEADIAGHGQEVLIR				
Obs. Mass	Intensity	S/N	Δ(amu)	Δ(ppm)
1606.8483	7198.7500	2599.0000	-0.0064	-3.9829
1606.8442	8803.7600	2697.0000	-0.0105	-6.5345
1606.8591	1118.1200	633.0000	0.0044	2.7383
1606.8544	4875.2900	1163.0000	-0.0003	-0.1867
1606.8459	2202.9800	1212.0000	-0.0088	-5.4765
1606.8688	1528.4700	687.0000	0.0141	8.7749
1606.8646	2230.5900	1118.0000	0.0099	6.1611
1606.8480	7589.0200	1589.0000	-0.0067	-4.1696
1606.8469	4859.6100	1627.0000	-0.0078	-4.8542
1606.8550	8126.7500	2234.0000	0.0003	0.1867
1606.8508	5423.0600	1808.0000	-0.0039	-2.4271
1606.8712	6115.1400	2035.0000	0.0165	10.2685

1606.8643	5677.1800	1889.0000	0.0096	5.9744
Table 1 Continued				
1606.8452	9697.2500	2165.0000	-0.0095	-5.9122
1606.8599	11731.5000	2746.0000	0.0052	3.2361
1606.8307	14322.8000	2693.0000	-0.0240	-14.9360
1606.8580	5672.1600	1515.0000	0.0033	2.0537
1606.8309	10701.8000	2007.0000	-0.0238	-14.8115
1606.8342	14566.3000	2479.0000	-0.0205	-12.7578
1606.8390	10336.6000	2264.0000	-0.0157	-9.7706
1606.8527	15350.6000	2693.0000	-0.0020	-1.2447
1606.8511	17577.4000	3169.0000	-0.0036	-2.2404
1606.8519	18396.2000	2793.0000	-0.0028	-1.7425
KGHHEAELKPLAQSHATK				
Obs. Mass	Intensity	S/N	Δ(amu)	Δ(ppm)
1982.0525	6893.1800	2254.0000	-0.0041	-2.0686
1982.0385	8739.1400	2719.0000	-0.0181	-9.1319
1982.0703	7203.7700	1959.0000	0.0137	6.9120
1982.0646	4702.7500	1784.0000	0.0080	4.0362
1982.0452	9069.1800	3053.0000	-0.0114	-5.7516
1982.0422	2302.1200	1173.0000	-0.0144	-7.2652
1982.0432	1356.5500	903.0000	-0.0134	-6.7607
1982.0653	2581.3300	972.0000	0.0087	4.3894
1982.0607	10199.8000	2609.0000	0.0041	2.0686
1982.0608	11536.9000	2395.0000	0.0042	2.1190
1982.0583	7072.6300	2287.0000	0.0017	0.8577
1982.0564	5056.6300	1773.0000	-0.0002	-0.1009
1982.0587	3272.7800	1676.0000	0.0021	1.0595
1982.0338	7292.2400	2261.0000	-0.0228	-11.5032
1982.0533	3527.5300	1628.0000	-0.0033	-1.6649
1982.0146	1180.8600	571.0000	-0.0420	-21.1901
1982.0697	1595.6100	1257.0000	0.0131	6.6093
1982.0602	1921.8800	2311.0000	0.0036	1.8163
1982.0535	5150.1200	1963.0000	-0.0031	-1.5640
GHHEAELKPLAQSHATK				
Obs. Mass	Intensity	S/N	Δ(amu)	Δ(ppm)
1853.9457	307.4510	83.0000	-0.0160	-8.6302
1853.9475	542.7450	129.0000	-0.0142	-7.6593
1853.9702	490.6670	110.0000	0.0085	4.5848
1853.9718	353.2550	84.0000	0.0101	5.4478
1853.9487	868.6470	220.0000	-0.0130	-7.0120
1853.9603	1500.2400	311.0000	-0.0014	-0.7551
1853.9509	681.4120	156.0000	-0.0108	-5.8254
1853.9520	841.4120	198.0000	-0.0097	-5.2320

1853.9475	739.1370	194.0000	-0.0142	-7.6593
1853.9502	860.2350	255.0000	-0.0115	-6.2029
Table 1 Continued				
1853.9714	581.6470	144.0000	0.0097	5.2320
1853.9841	666.3530	181.0000	0.0224	12.0822
1853.9543	385.8820	202.0000	-0.0074	-3.9915
1853.9581	143.0590	71.0000	-0.0036	-1.9418
1853.9528	573.4900	224.0000	-0.0089	-4.8005
1853.9331	65.2549	29.0000	-0.0286	-15.4264
1853.9512	304.9410	153.0000	-0.0105	-5.6635
1853.9529	305.5690	143.0000	-0.0088	-4.7466
1853.9731	98.5098	74.0000	0.0114	6.1490
1853.9624	304.3140	136.0000	0.0007	0.3776
HKIPIKYLEFISDAIIHVLHSK				
Obs. Mass	Intensity	S/N	Δ(amu)	Δ(ppm)
2601.4846	559.6860	369.0000	-0.0069	-2.6523
2601.4795	731.6080	523.0000	-0.0120	-4.6127
2601.4983	355.7650	259.0000	0.0068	2.6139
2601.4563	183.8430	182.0000	-0.0352	-13.5307
2601.4900	317.4900	256.0000	-0.0015	-0.5766
2601.4800	607.3730	404.0000	-0.0115	-4.4205
2601.4700	220.8630	183.0000	-0.0215	-8.2645
2601.5078	267.9220	215.0000	0.0163	6.2656
2601.4880	101.0200	107.0000	-0.0035	-1.3454
2601.4756	129.8820	102.0000	-0.0159	-6.1119
2601.4844	149.3330	75.0000	-0.0071	-2.7292
2601.4836	116.7060	75.0000	-0.0079	-3.0367
YLEFISDAIIHVLHSK				
Obs. Mass	Intensity	S/N	Δ(amu)	Δ(ppm)
1884.9935	868.1370	331.0000	-0.0283	-15.0131
1885.0079	619.9220	338.0000	-0.0139	-7.3739
1884.9883	1184.6300	558.0000	-0.0335	-17.7717
1885.0328	1035.9200	342.0000	0.0110	5.8355
1885.0220	1812.7100	605.0000	0.0002	0.1061
1885.0074	767.3730	356.0000	-0.0144	-7.6392
1885.0027	766.7450	300.0000	-0.0191	-10.1325
1885.0236	915.4510	276.0000	0.0018	0.9549
1885.0151	759.2160	413.0000	-0.0067	-3.5543
1885.0131	1985.2600	612.0000	-0.0087	-4.6153
1885.0260	1153.2500	389.0000	0.0042	2.2281
1885.0145	904.7840	346.0000	-0.0073	-3.8726
1885.0149	415.3730	254.0000	-0.0069	-3.6604
1885.0264	276.0780	197.0000	0.0046	2.4403

1885.0245	2362.3500	731.0000	0.0027	1.4323

Raw mass spectrometry measurements for observed mass (Obs. Mass), intensity, signal-to-noise (S/N), the difference between calculated and measured masses (Δ or Delta) in amu and ppm are reported. Observed mass, signal-to-noise, and intensity measurements are derived from the AB Sciex 4000 Series Data Explorer Software used for data acquisition and processing. The Δ mass differences were calculated using the standard equation: Δ = (mass observed – mass calculated)/ mass calculated. Where mass calculated is the accurate monoisotopic mass calculated from the primary sequence of the observed peptide. This Δ was multiplied by 10^6 to convert to parts-per-million (ppm).

BIBLIOGRAPHY

Aebersold R, Mann M. 2003. Mass spectrometry-based proteomics. *Nature* 422(6928):198-207.

Alfassi, Z. 2004. On the normalization of a mass spectrum for comparison of two spectra. *Jour of Amer Soc for Mass Spec* 15:385-387.

Alnimr, AM. 2015. Dormancy models for Mycobacterium tuberculosis: A minireview. *Brazil Jour of Microb* 46(3)641-647.

Angel T, Aryal U, Hengel S, Baker E, Kelly R, Robinson E, Smith R. 2012. Mass spectrometry based proteomics: existing capabilities and future directions. *Chem Soc Rev* 41(10):3912-3928.

Ayrapetyan M, Williams TC, Baxter R, Oliver JD. 2015. Viable but nonculturable and persister cells coexist stochastically and are induced by human serum. *Infect Immun* 83:4194–4203.

Ayrapetyan M, Williams TC, Oliver JD. 2014. Interspecific quorum sensing mediates the resuscitation of viable but nonculturable vibrios. *Appl Environ Microbiol* 80:2478 –2483.

Ayrapetyan M, Williams TC, Oliver JD. 2015. Bridging the gap between viable but non-culturable and antibiotic persistent bacteria. *Trends Microbiol* 23:7–13.

Azam TA, Ishihama A. 1999. Twelve species of the nucleoid-associated protein from Escherichia coli. Sequence recognition specificity and DNA binding affinity. *J Biol Chem* 274:33105-33113.

Azam TA, Iwata A, Nishimura A, Ueda S, Ishihama A. 1999. Growth phase-dependent variation in protein composition of the Escherichia coli nucleoid. *J Bacteriol* 181:6361-6370.

Bacon J, James BW, Wernisch L, Williams A, Morley KA, Hatch GJ, Mangan JA, Hinds J, Stoker NG, Butcher PD, Marsh PD. 2004. The influence of reduced oxygen availability on pathogenicity and gene expression in Mycobacterium tuberculosis. *Tuberculosis (Edinb)* 84:205–217.

Balgey BM, Laudeman T, Yang L, Song T, Lee CS. 2007. Comparative evaluation of tandem MS search algorithms using a target-decoy search strategy. Molec & Cell Prot 6:1599-1608.

Bark SJ, Hook V. 2007. Differential recovery of peptides from sample tubes and the reproducibility of quantitative proteomic data. *Journ Pro Res* 6(11):4511-4516.

Barry CE, Boshoff HI, Dartois V, Dick T, Ehrt S, Flynn J, Schnappinger D, Wilkinson RJ, Young D. 2009. The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. *Nat Rev Microbiol* 7:845– 855.

Bättig P, Hathaway LJ, Hofer S, Mühlemann K. 2006. Serotype-specific invasiveness and colonization prevalence in *Streptococcus pneumoniae* correlate with the lag phase during *in vitro* growth. *Microbes Infect* 8:2612–2617.

Bernier SP, Lebeaux D, DeFrancesco AS, Valomon A, Soubigou G, Coppée J-Y, Ghigo J-M, Beloin C. 2013. Starvation, together with the SOS response, mediates high biofilm-specific tolerance to the fluoroquinolone ofloxacin. *PLoS Genet* 9:1003-1144.

Besnard V, Federighi M, Declerq E, Jugiau F, Cappelier J-M. 2002. Environmental and physico-chemical factors induce VBNC state in Listeria monocytogenes. *Vet Res* 33:359 –370.

Betts JC, Lukey PT, Robb LC, McAdam RA, Duncan K. 2002. Evaluation of a nutrient starvation model of Mycobacterium tuberculosis persistence by gene and protein expression profiling. *Mol Microbiol* 43:717–731.

Bhagwat AA, Tan J, Sharma M, Kothary M, Low S, Tall BD, Bhagwat M. 2006. Functional heterogeneity of RpoS in stress tolerance of enterohemorrhagic Escherichia coli strains. *Appl Environ Microbiol* 72:4978–4986.

Bhowmick T, Ghosh S, Dixit K, Ganesan V, Ramagopal UA, Dey D, Sarma SP, Ramakumar S, Nagaraja V. 2014. Targeting Mycobacterium tuberculosis nucleoid-associated protein HU with structure-based inhibitors. *Nat Commun* 5:4124.

Bisson N, James DA, Ivosev G. 2011. Selected reaction monitoring mass spectrometry reveals the dynamics of signaling through the GRB2 adaptor. *Nat Biotechnol* 29(7):653-658.

Blumenthal A, Trujillo C, Ehrt S, Schnappinger D. 2010. Simultaneous analysis of multiple Mycobacterium tuberculosis knockdown mutants in vitro and in vivo. *PLoS One* 5:156-167.

Boaretti M, Del Mar Lleò M, Bonato B, Signoretto C, Canepari P. 2003. Involvement of rpoS in the survival of Escherichia coli in the viable but nonculturable state. *Environ Microbiol* 5:986–996.

Bochkareva ES, Girshovich AS, and Bibi E. 2002. Identification and characterization of theEscherichia colistress protein UP12, a putative in vivo substrate of GroEL. *Europ Jour of Biochem* 269(12):3032-3040.

Boesl U, Weinkauf R, Schlag EW. 1992. Reflectron time-of-flight mass spectrometry and laser excitation for the analysis of neutrals, ionized molecules and secondary fragments. *Int Jour of Mass Spec and Ion Proc*. 112(2-3):121-166.

Bonnefoy E, Rouviere-Yaniv J. 1992. HU, the major histone-like protein of E. coli, modulates the binding of IHF to oriC. EMBO J 11:4489 – 4496.

Boon C, Dick T. 2002. Mycobacterium bovis BCG response regulator essential for hypoxic dormancy. *J Bacteriol* 184:6760–6767.

Boon C, Dick T. 2012. How Mycobacterium tuberculosis goes to sleep: the dormancy survival regulator DosR a decade later. *Future Microbiol* 7:513–518.

Bristow T, Constantine J, Harrison M, Cavoit F. 2008. Performance optimization of a new-generation orthogonal-acceleration quadrupole-time-of-flight mass spectrometer. *Rap Comm in Mass Spec* 22(8): 1-21.

Browning DF, Grainger DC, Busby SJ. 2010. Effects of nucleoid-associated proteins on bacterial chromosome structure and gene expression. *Curr Opin Microbiol* 13:773–780.

Bruins A, Covey T, Henion J. 1987. Ion spray interface for combined liquid chromatography/atmospheric pressure ionization mass spectrometry. *Anal Chem* 59:2642-2646.

Carvalhais V, Cerca N, Vilanova M, Vitorino R. 2015. Proteomic profile of dormancy within Staphylococcus epidermidis biofilms using iTRAQ and label-free strategies. *Appl Microbiol Biotechnol* 99:2751–2762.

Chait EM, Kitson FG. 1970. Field-ionization mass spectrometry-fragmentation and metastables. *Org Mass Spec* 3(5):533-547.

Chan RJ, Feng G. 2007. Review article PTPN11 is the first identified proto-oncogene that encodes a tyrosine phosphatase. *Blood* 109(3):862-868.

Chang IF. 2006. Mass spectrometry-based proteomic analysis of the epitope-tag affinity purified protein complexes in eukaryotes. *Proteo* 6:6158-6166.

Chao MC, Rubin EJ. 2010. Letting sleeping dos lie: does dormancy play a role in tuberculosis? *Annu Rev Microbiol* 64:293–311.

Chaparian RR, Olney SG, Hustmyer CM, Rowe-Magnus DA, van Kessel JC. 2016. Integration host factor and LuxR synergistically bind DNA to coactivate quorum-sensing genes in Vibrio harveyi. *Mol Microbiol* 101: 823–840.

Chelius D, Zhang T, Wang G, Shen RF. 2003. Global protein identification and quantification technology using two-dimensional liquid chromatography nanospray mass spectrometry. *Anal Chem 75*:6658-6665.

Chen CY, Schwartz RJ. 1995. Identification of novel DNA binding targets and regulatory domains of a murine tinman homeodomain factor, nkx-2.5. *J Biol Chem* 270:15628-33.

Chen JM, Ren H, Shaw JE, Wang YJ, Li M, Leung AS, Tran V, Berbenetz NM, Kocincova D, Yip CM, Reyrat JM, Liu J. 2008. Lsr2 of Mycobacterium tuberculosis is a DNA-bridging protein. *Nucleic Acids Res* 36:2123–2135.

Colangeli R, Haq A, Arcus VL, Summers E, Magliozzo RS, McBride A, Mitra AK, Radjainia M, Khajo A, Jacobs WR, Salgame P, Alland D. 2009. The multifunctional histone-like protein Lsr2 protects mycobacteria against reactive oxygen intermediates. *Proc Natl Acad Sci U S A* 106:4414–4418.

Connor PB, Costello CE. 2000. Internal calibration on adjacent samples (Incas) with fourier transform mass spectrometry. *Anal Chem* 72(24):5881-5885.

Correia FF, D'Onofrio A, Rejtar T, Li L, Karger BL, Makarova K, Koonin EV, Lewis K. 2006. Kinase activity of overexpressed HipA is required for growth arrest and multidrug tolerance in Escherichia coli. *J Bacteriol* 188:8360–8367.

Corrigan RM, Bellows LE, Wood A, Gründling A. 2016. ppGpp negatively impacts ribosome assembly affecting growth and antimicrobial tolerance in Gram-positive bacteria. *Proc Natl Acad Sci U S A* 113: E1710–E1719.

Craig NL, Nash HA. 1984. E. coli Integration Host Factor Binds to Specific Sites in DNA. *Cell* 39:707-716.

Cunningham E, O'Byrne C, Oliver JD. 2009. Effect of weak acids on Listeria monocytogenes survival: evidence for a viable but nonculturable state in response to low pH. *Food Control* 20:1141–1144.

Dean ACR, Hinshelwood CN. 1957. The adaptation of bacterial cultures during the lag phase in media containing new substrates or antibacterial agents. *Proc R Soc Lond B Biol Sci* 147:247–257.

Dhar N, McKinney JD. 2007. Microbial phenotypic heterogeneity and antibiotic tolerance. *Curr Opin Microbiol* 10:30–38.

Dietz L. 1965. Basic properties of electron multiplier ion detection and pulse counting methods in mass spectrometry. *Rev of Sci Instr* 36: 1763.

Dillon SC, Dorman CJ. 2010. Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat Rev Microbiol* 8:185-195.

Dörr T, Vulic´ M, Lewis K. 2010. Ciprofloxacin causes persister formation by inducing the TisB toxin in Escherichia coli. *PLoS Biol* 8:1000-10317.

Drumm JE, Mi K, Bilder P, Sun M, Lim J, Bielefeldt-Ohmann H, Basaraba R, So M, Zhu G, Tufariello JM, Izzo AA, Orme IM, Almo SC, Leyh TS, Chan J. 2009. Mycobacterium tuberculosis universal stress protein Rv2623 regulates bacillary growth by ATP-binding: requirement for establishing chronic persistent infection. *PLoS Pathog* 5:1000-1460.

Durfee T, Hansen A-M, Zhi H, Blattner FR, Jin DJ. 2008. Transcription profiling of the stringent response in Escherichia coli. *J Bacteriol* 190: 1084–1096.

Dutta NK, Karakousis PC. 2014. Latent tuberculosis infection: myths, models, and molecular mechanisms. *Microbiol Mol Biol Rev* 78:343–371.

Elias JE, Haas W, Faherty BK, Gygi SP. 2005. Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nat Methods* 2:667– 675.

Fahrner RL, Cascio D, Lake JA, Slesarev A. 2001. An ancestral nuclear protein assembly: Crystal structure of the Methanopyrus kandleri histone. *Protein Sci* 10:2002–2007.

Fakruddin M, Mannan KS, Andrews S. 2013. Viable but nonculturable bacteria: food safety and public health perspective. *ISRN Microbiol* 2013:703813.

Farewell A, Diez AA, DiRusso CC, Nyström T. 1996. Role of the Escherichia coli FadR regulator in stasis survival and growth phase-dependent expression of the uspA, fad, and fab genes. *J Bacteriol* 178:6443–6450.

Farmer TB, Caprioli RM. 1998. Determination of protein–protein interactions by matrix-assisted laser desorption/ionization mass spectrometry. *Jour Mass Spec* 33(8): 697-704.

Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. 1989. Electrospray ionization for mass spectrometry of large biomolecules. *Science* 246(4926):64–71.

Fjeldstead J. 2003. Time-of-flight mass spectrometry. *Agilent Technologies* 1-6.

Flynn J and Chan J. 2001. Tuberculosis: Latency and Reactivation. Infect Immun 69(7):4195–4201.

Frenkiel-Krispin D, Ben-Avraham I, Englander J, Shimoni E, Wolf SG, Minsky A. 2004. Nucleoid restructuring in stationary-state bacteria. *Mol Microbiol* 51:395-405.

Fridman O, Goldberg A, Ronin I, Shoresh N, Balaban NQ. 2014. Optimization of lag time underlies antibiotic tolerance in evolved bacterial populations. *Nature* 513:418–421.

Frimodt-Møller N, Sebbesen O, Thomsen VF. 1983. The *Pneumococcus* and the mouse protection test: importance of the lag phase *in vivo*. *Chemotherapy* 29:128–134.

Gaca AO, Kudrin P, Colomer-Winter C, Beljantseva J, Liu K, Anderson B, Wang JD, Rejman D, Potrykus K, Cashel M, Hauryliuk V, Lemos JA. 2015. From (p)ppGpp to (pp)pGpp: characterization of regulatory effects of pGpp synthesized by the small alarmone synthetase of Enterococcus faecalis. *J Bacteriol* 197:2908 –2919.

Gao B, Gupta RS. 2012. Phylogenetic framework and molecular signatures for the main clades of the phylum Actinobacteria. *Microbiol Mol Biol Rev* 76:66-112.

Geer LY, Markey SP, Kowalak JA, Wagner L, Xu M, Maynard DM, Yang X, Shi W, Bryant SH. 2004. Open mass spectrometry search algorithm. *J Proteome Res* 3:958 –964.

Gengenbacher M, Kaufmann SHE. 2012. Mycobacterium tuberculosis: success through dormancy. *FEMS Microbiol Rev* 36:514 –532.

Gerber SA, Rush J, Stemman O, Kirshner MW, Gygi SP. 2003. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *PNAS* 100(12):6940-6945.

Gerdes K, Maisonneuve E. 2015. Remarkable functional convergence: alarmone pp;gpp mediates persistence by activating type I and II toxinantitoxins. *Mol Cell* 59:1–3.

Greenblatt CL, Baum J, Klein BY, Nachshon S, Koltunov V, Cano RJ. 2004. Micrococcus luteus—survival in amber. *Microb Ecol* 48:120 –127.

Grove A. 2011. Functional evolution of bacterial histone-like HU proteins. *Curr Issues Mol Biol* 13:1-12.

Gustavsson N, Diez A, Nyström T. 2002. The universal stress protein paralogues of Escherichia coli are co-ordinately regulated and cooperate in the defence against DNA damage. *Mol Microbiol* 43:107–117.

Hanafy RA, Couger MB, Baker K, Murphy C, O'Kane, SD, Budd C, Youssef N. 2016. Draft genome sequence of Micrococcus luteus strain O'Kane implicates metabolic versatility and the potential to degrade polyhydroxybutyrates. *Geno Dat 9*:148-153.

Harms A, Maisonneuve E, Gerdes K. 2016. Mechanisms of bacterial persistence during stress and antibiotic exposure. *Science* 354:42-68.

Hingley-Wilson SM, Lougheed KEA, Ferguson K, Leiva S, Williams HD. 2010. Individual Mycobacterium tuberculosis universal stress protein homologues are dispensable in vitro. *Tuberculosis (Edinb)* 90:236 –244.

Hiroshi I, Michihiko O. 1973. A theoretical treatment of the fragmentation rules in the mass spectrometry of organic compounds. *CSJ* 46(6):1873-1874.

Hu X, Li X, Huang L, Chan J, Chen Y, Deng H, Mi K. 2015. Quantitative proteomics reveals novel insights into isoniazid susceptibility in mycobacteria mediated by a universal stress protein. *J Proteome Res* 14: 1445–1454.

Hunt DF, Yates JR, Shabanowitz J, Winston S, Haur CR. 1986. Protein sequencing by tandem mass spectrometry. *Proc Natl Acad Sci* 83:6233-6237.

Jakob U, Gaestel M, Engel K, Buchner J. 1993. Small heat shock proteins are molecular chaperones. *J Bio Chem* 268(3):1517-20.

Joerger AC, Fersht AR. 2016. The p53 pathway: origins, inactivation in cancer, and emerging therapeutic approaches. *Annu Rev Biochem* 815-847.

Källberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, Xu J. 2012. Template-based protein structure modeling using the RaptorX web server. *Nat Protoco* 7:1511–1522.

Kamashev D, Balandina A, Rouviere-Yaniv J. 1999. The binding motif recognized by HU on both nicked and cruciform DNA. *EMBO J* 18:5434–5444.

Kapp EA, Schütz F, Reid GE, Eddes JS, Moritz RL, O'Hair RA, Speed TP, Simpson RJ. 2003. Mining a tandem mass spectrometry database to determine the trends and global factors influencing peptide fragmentation. *Anal Chem* 75(22):6251-6264.

Kaprelyants AS, Gottschal JC, Kell DB. 1993. Dormancy in non sporulating bacteria. *FEMS Microbiol Rev* 104:271-286.

Kaprelyants AS, Kell DB. 1993. Dormancy in stationary-phase cultures of Micrococcus luteus: flow cytometric analysis of starvation and resuscitation. *Appl Environ Microbiol* 59:3187–3196.

Kaprelyants AS, Mukamolova GV, Davey HM, Kell DB. 1996. Quantitative analysis of the physiological heterogeneity within starved cultures of Micrococcus luteus by flow cytometry and cell sorting. *Appl Environ Microbiol* 62:1311–1316.

Karas M, Hillenkamp F. 1988. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* 60(20):2299–301.

Keren I, Shah D, Spoering A, Kaldalu N, Lewis K. 2004. Specialized persister cells and the mechanism of multidrug tolerance in Escherichia coli. *J Bacteriol* 186:8172– 8180.

Kim JS, Chowdhury N, Yamasaki R, Wood TK. 2018. Viable But Non-Culturable and Persistence Describe the Same Bacterial Stress State. *Environ Microbiol* 1111:1462-2920.

Kong I-S, Bates TC, Hülsmann A, Hassan H, Smith BE, Oliver JD. 2004. Role of catalase and oxyR in the viable but nonculturable state of Vibrio vulnificus. *FEMS Microbiol Ecol* 50:133–142.

Kornberg, HL. 1966. The role and control of the glyoxylate cycle in *Escherichia coli*. *Biochem J* 99(1):1–11.

Krishnamoorthy S, Liu Z, Hong A. 2013. A novel phosphopeptide microarray based interactome map in breast cancer cells reveals phosphoprotein-GRB2 cell signaling networks. *PLoS One* 8(6):67634.

Kumar S, Sardesai AA, Basu D, Muniyappa K, Hasnain SE. 2010. DNA clasping by mycobacterial HU: the C-terminal region of HupB mediates increased specificity of DNA binding. *PLoS One* 5:12551.

Kvint K, Nachin L, Diez A, Nyström T. 2003. The bacterial universal stress protein: function and regulation. *Curr Opin Microbiol* 6:140 –145.

Laal, S. 2012. How Does *Mycobacterium tuberculosis* establish infection? *The Jour of Infect Dis* 206(8):1157–1159.

Lal H, Kolaja KL, Force T. 2013. Cancer genetics and the cardiotoxicity of the therapeutics. *J Am Coll Cardiol*. 61(3):267-274.

LaCava J, Molloy KR, Taylor MS, Domanski M, Chait BT, Rout MP. 2015. Affinity proteomics to study endogenous protein complexes: pointers, pitfalls, preferences and perspectives. *BioTechni* 58:103-119.

Lafleur MD, Qi Q, Lewis K. 2010. Patients with long-term oral carriage harbor high-persister mutants of Candida albicans. *Antimicrob Agents Chemother* 54:39–44.

Lang B, Blot N, Bouffartigues E, Buckle M, Geertz M, Gualerzi CO, Mavathur R, Muskhelishvili G, Pon CL, Rimsky S, Stella S, Babu MM, Travers A. 2007. High-affinity DNA binding sites for H-NS provide a molecular basis for selective silencing within proteobacterial genomes. *Nucleic Acids Res* 35:6330–6337.

Lay JO. 2001. Maldi-tof mass spectrometry of bacteria. *Mass Spec Rev* 20(4) 1-44.

Lee BH, Murugasu-Oei B, Dick T. 1998. Upregulation of a histone-like protein in dormant Mycobacterium smegmatis. *Mol Gen Genet* 260:475-479.

Lee SY, Lim CJ, Droge P, Yan J. 2015. Regulation of Bacterial DNA Packaging in Early Stationary Phase by Competitive DNA Binding of Dps and IHF. *Sci Rep* 5:18146.

Leung V, Lévesque CM. 2012. A stress-inducible quorum-sensing peptide mediates the formation of persister cells with noninherited multidrug tolerance. *J Bacteriol* 194:2265–2274.

Lewis J, Wei J, Siuzdak G. 2000. Matrix-assisted laser desorption/ionization mass spectrometry in peptide and protein analysis. *Ency of Anal Chem* 5880-5894.

Lewis K. 2007. Persister cells, dormancy and infectious disease. *Nat Rev Microbiol* 5:48 –56.

Lewis K. 2010. Persister cells*. Annu Rev Microbiol* 64:357–372.

Li L, Mendis N, Trigui H, Oliver JD, Faucher SP. 2014. The importance of the viable but non-culturable state in human bacterial pathogens. *Front Microbiol* 5:258.

Lin J, Chen H, Droge P, Yan J. 2012. Physical organization of DNA by multiple non-specific DNA-binding modes of integration host factor (IHF). PLoS One 7:498-585.

Lipworth S, Hammond RJH, Baron VO, Hu Y, Coates A, Gillespie SH. 2016. Defining dormancy in mycobacterial disease. *Tuberculosis (Edinb)* 99: 131–142.

Liu H, Sadygov RG, Yates JR. 2004. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* 76(14):4193-4201.

Liu W-T, Michail H, Karavolos, Bulmer DM, Allaoui A, Carlos, RD, Hormaeche E, Jeong JL, Khan CM. 2007. Role of the universal stress protein UspA of *Salmonella* in growth arrest, stress and virulence. *Micro Pathos* 42(1):2-10.

Liu Y, Wang H, Cui T, Zhou X, Jia Y, Zhang H, He ZG. 2016. NapM, a new nucleoid-associated protein, broadly regulates gene expression and affects mycobacterial resistance to anti-tuberculosis drugs. *Mol Microbiol* 101:167–181.

Liu Y, Zhou S, Deng Q, Li X, Meng J, Guan Y, Li C, Xiao C. 2016. Identification of a novel inhibitor of isocitrate lyase as a potent antitubercular agent against both active and non-replicating Mycobacterium tuberculosis. *Tuberculosis (Edinb)* 97:38–46.

Lleò MM, Pierobon S, Tafi MC, Signoretto C, Canepari P. 2000. mRNA detection by reverse transcription-PCR for monitoring viability over time in an Enterococcus faecalis viable but nonculturable population maintained in a laboratory microcosm. *Appl Environ Microbiol* 66:4564–4567.

Lleó MM, Tafi MC, Canepari P. 1998. Nonculturable Enterococcus faecalis cells are metabolically active and capable of resuming active growth. *Syst Appl Microbiol* 21:333–339.

Loebel RO, Shorr E, Richardson HB. 1933. The influence of adverse conditions upon the respiratory metabolism and growth of human tubercle bacilli. *J Bacteriol* 26:167–200.

Luidalepp H, Jõers A, Kaldalu N, Tenson T. 2011. Age of inoculum strongly influences persister frequency and can mask effects of mutations implicated in altered persistence. *J Bacteriol* 193:3598 –3605.

Luijsterburg MS, Noom MC, Wuite GJ, Dame RT. 2006. The architectural role of nucleoid-associated proteins in the organization of bacterial chromatin: a molecular perspective. *J Struct Biol* 156:262-272.

Maertens B, Spriestersbach A, Kubicek J, Schafer F. 2015. Strep-tagged protein purification. *Meth Enzymol* 559(2015):53-69.

Magnusson LU, Farewell A, Nyström T. 2005. ppGpp: a global regulator in Escherichia coli. *Trends Microbiol* 13:236–242.

Maisonneuve E, Castro-Camargo M, Gerdes K. 2013. (p)ppGpp controls bacterial persistence by stochastic induction of toxin-antitoxin activity. *Cell* 154:1140 –1150.

Maisonneuve E, Gerdes K. 2014. Molecular mechanisms underlying bacterial persisters. *Cell* 157:539 –548.

Mali S, Mitchell M, Havis S, Bodunrin A, Rangel J, Olson G, Widger WR, Bark SJ. 2017. A proteomic signature of dormancy in the Actinobacterium *Micrococcus luteus*. *J Bacteriol* 199(14):206-217.

Malmström J, Karlsson C, Nordenfelt P, Ossola R, Weisser H, Quandt A, Hansson K, Aebersold R, Malmström L, Björck L. 2012. Streptococcus pyogenes in human plasma: adaptive mechanisms analyzed by mass spectrometry-based proteomics. *J Biol Chem* 287:1415–1425.

Mangan MW, Lucchini S, Danino V, Cróinín TO, Hinton JCD, Dorman CJ. 2006. The integration host factor (IHF) integrates stationary-phase and virulence gene expression in Salmonella enterica serovar Typhimurium. *Mol Microbiol* 59:1831–1847.

Mann M, Meng CK, Fenn JB. 1989. Interpreting mass spectra of multiply charged ions. Anal Chem 61(15):1702-1708.

Marchler-Bauer A, Bryant SH. 2004. CD-Search: protein domain annotations on the fly. *Nucleic Acids Res* 32:W327–W331.

Matsumoto S, Furugen M, Yukitake H, Yamada T. 2000. The gene encoding mycobacterial DNA-binding protein I (MDPI) transformed rapidly growing bacteria to slowly growing bacteria. *FEMS Microbiol Lett* 182:297-301.

Mazumder M, Gourinath S. 2016. Structure-based design of inhibitors of the crucial cysteine biosynthetic pathway enzyme O-acetyl serine sulfhydrylase. *Curr Top Med Chem* 16:948 –959.

Mendoza MC, Er EE, Blenis J. 2011. The Ras-ERK and PI3K-mTOR pathways: Cross-talk and compensation. *Trends Biochem Sci* 36(6):320-328.

Meyer JG, Schilling B. 2017. Clinical applications of quantitative proteomics using targeted and untargeted data-independent acquisition techniques. *Expert Rev Proteomics* 14:419-429.

Micka B, Marahiel MA. 1992. The DNA-binding protein HBsu is essential for normal growth and development in Bacillus subtilis. *Biochimie* 74:641-650.

Mishra A, Vij M, Kumar D, Taneja V, Mondal AK, Bothra A, Rao V, Ganguli M, Taneja B. 2013. Integration host factor of Mycobacterium tuberculosis, mIHF, compacts DNA by a bending mechanism. *PLoS One* 8:699-785.

Mitchell M, Mali S, King CC, Bark SJ. 2015. Enhancing maldi time-of-flight mass spectrometer performance through spectrum averaging. *PLoS One* 10(3).

Möker N, Dean CR, Tao J. 2010. Pseudomonas aeruginosa increases formation of multidrug-tolerant persister cells in response to quorumsensing signaling molecules. *J Bacteriol* 192:1946 –1955.

Mukamolova GV, Kormer SS, Yanopolskaya ND, Kaprelyants AS. 1995. Properties of dormant cells in stationary-phase cultures of *Micrococcus luteus* during prolonged incubation. *Mikrobiologia* 64:284-288.

Mukamolova GV, Yanopolskaya ND, Kell DB, Kaprelyants AS. 1998. On resuscitation from the dormant state of Micrococcus luteus. *Antonie Van Leeuwenhoek* 73:237–243.

Mulcahy LR, Burns JL, Lory S, Lewis K. 2010. Emergence of Pseudomonas aeruginosa strains producing high levels of persister cells in patients with cystic fibrosis. *J Bacteriol* 192:6191–6199

Muñoz-Elías EJ, McKinney JD. 2005. Mycobacterium tuberculosis isocitrate lyases 1 and 2 are jointly required for in vivo growth and virulence. *Nat Med* 11:638–644.

Muttucumaru DGN, Roberts G, Hinds J, Stabler RA, Parish T. 2004. Gene expression profile of Mycobacterium tuberculosis in a non-replicating state. *Tuberculosis (Edinb)* 84:239 –246.

Nachin L, Nannmark U, Nyström T. 2005. Differential roles of the universal stress proteins of Escherichia coli in oxidative stress resistance, adhesion, and motility. *J Bacteriol* 187:6265– 6272.

Nandakumar M, Nathan C, Rhee KY. 2014. Isocitrate lyase mediates broad antibiotic tolerance in Mycobacterium tuberculosis. *Nat Commun* 5:4306.

Nathan I, Aharon M, Reisenfeld G, Dvilansky A. 1992. A novel agarose acrobeads protein A column for selective immunoadsorbance of whole blood: performance, specificity and safety. *Biomater. Artif. Cells Immobil Biotechnol Off J Int Soc Artif Cells Immobil Biotechnol* 23-30.

Nesvizhskii A, Vitek O, Aebersold R. 2007. Analysis of proteomic data generated by tandem mass spectrometry. *Nat Meth* 4: 787-797.

Nguyen HH, de la Tour CB, Toueille M, Vannier F, Sommer S, Servant P. 2009. The essential histone-like protein HU plays a major role in Deinococcus radiodurans nucleoid compaction. *Mol Microbiol* 73:240-252.

Nikitushkin V, Demina G, Shleeva M, Kaprelyants A. 2013. Peptidoglycan fragments stimulate resuscitation of “non-culturable” mycobacteria. *Antonie Van Leeuwenhoek* 103:37–46.

Nilsson L, Oliver JD, Kjelleberg S. 1991. Resuscitation of Vibrio vulnificus from the viable but non culturable state. *J Bacteriol* 173:5054–5059.

Noble WS, MacCoss MJ. 2012. Computational and statistical analysis of protein mass spectrometry data. *PLoS Comput Biol* 8(1): e1002296.

Nyka W. 1974. Studies on the effect of starvation on mycobacteria. *Infect Immun* 9:843 – 850.

Nyström, T, Neidhardt FC. 1992. Cloning, mapping, and nucleotide sequencing of a gene encoding a universal stress protein in Escherichia coli. *Mol Microbiol* 6:3187-3198.

Nyström T, Neidhardt FC. 1993. Isolation and properties of a mutant of Escherichia coli with an insertional inactivation of the uspA gene, which encodes a universal stress protein. *J Bacteriol* 175:3949 – 3956.

Nyström T, Neidhardt FC. 1994. Expression and role of the universal stress protein, UspA, of Escherichia coli during growth arrest. *Mol Microbiol* 11:537–544.

O'Toole R, Williams HD. 2003. Universal stress proteins and Mycobacterium tuberculosis. *Res Microbiol* 154:387–392.

Oberto J, Nabti S, Jooste V, Mignot H Rouviere-Yaniv J. 2009. The HU regulon is composed of genes responding to anaerobiosis, acid stress, high osmolarity and SOS induction. *PLoS ONE* 4:4367.

Ohniwa RL, Morikawa K, Kim J, Ohta T, Ishihama A, Wada C, Takeyasu K. 2006. Dynamic state of DNA topology is essential for genome condensation in bacteria. *EMBO J* 25:5591-602.

Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M. 2002. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* 1:376−386.

Ong SE, Mann M. 2005. Mass spectrometry-based proteomics turns quantitative. *Nat Chem Biol* 1:252–262.

Page R, Peti W. 2016. Toxin-antitoxin systems in bacterial growth arrest and persistence. *Nat Chem Biol* 12:208 –214.

Papayannopoulos I. 1995. The interpretation of collision-induced dissociation tandem mass spectra of peptides. *Mas Spec Rev* 14(1):1-3.

Paulo J, Kadiyala V, Banks P, Steen H, Conwell D. 2012. Mass spectrometry-based proteomics for translational research: a technical overview. *Yale Journ of Bio and Med* 85:59-73.

Pedulla ML, Hatfull GF. 1998. Characterization of the mIHF gene of Mycobacterium smegmatis. *J Bacteriol* 180:5473–5477.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera - A Visualization System for Exploratory Research and Analysis. *J Comput Chem* 25:1605-1612.

Phadtare S. 2004. Recent developments in bacterial cold-shock response. *Curr Issues Mol Biol* 6:125–136.

Pinto D, Almeida V, Almeida Santos M, Chambel L. 2011. Resuscitation of Escherichia coli VBNC cells depends on a variety of environmental or chemical stimuli. *J Appl Microbiol* 110:1601–1611.

Polikanov YS, Blaha GM, Steitz TA. 2012. How hibernation factors RMF, HPF, and YfiA turn off protein synthesis. *Science* 336:915–918.

Powell DW, Weaver CM, Jennings JL, McAfee KJ, He Y, Weil PA, Link, A. 2004. Cluster analysis of mass spectrometry data reveals a novel component of SAGA. J *Mol Cell Biol 24*:7249-7259.

Prakash A, Tomazela DM, Frewen B, Maclean B, Merrihew G, Peterman S, Maccoss MJ. 2009. Expediting the development of targeted SRM assays: Using data from shotgun proteomics to automate method development. *J Proteome Res* 8:2733–2739.

Qiu Y, Tereshko V, Kim Y, Zhang R, Collart F, Yousef M, Kossiakoff A, Joachimiak A. 2006. The crystal structure of Aq_328 from the hyperthermophilic bacteria Aquifex aeolicus shows an ancestral histone fold. *Proteins* 62:8 –16.

Ramage HR, Connolly LE, Cox JS. 2009. Comprehensive functional analysis of Mycobacterium tuberculosis toxin-antitoxin systems: implications for pathogenesis, stress responses, and evolution. *PLoS Genet* 5:1000-1067.

Raman N, Black PN, DiRusso CC. 1997. Characterization of the fatty acid-responsive transcription factor FadR. Biochemical and genetic analyses of the native conformation and functional domains. *J Biol Chem* 272:30645–30650.

Rao NV, Shashidhar R, Bandekar JR. 2014. Induction, resuscitation and quantitative real-time polymerase chain reaction analyses of viable but nonculturable Vibrio vulnificus in artificial sea water. *World J Microbiol Biotechnol* 30:2205–2212.

Rauniyar N. 2015. Parallel Reaction Monitoring: A Targeted Experiment Performed Using High Resolution and High Mass Accuracy Mass Spectrometry. *Int J Mol Sci* 16:28566-28581.

Ren L, Emery D, Kaboord B, Chang E, Qoronfleh MW. 2003. Improved immunomatrix methods to detect protein: protein interactions. *J Biochem Biophys Methods* 57:143-157.

Rice PA, Yang S, Mizuuchi K, Nash HA. 1996. Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell* 87:1295–1306.

Rosenkrands I, Slayden RA, Crawford J, Aagaard C, Barry CE, Andersen P. 2002. Hypoxic response of Mycobacterium tuberculosis studied by metabolic labeling and proteome analysis of cellular and extracellular proteins. *J Bacteriol* 184:3485–3491.

Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, Purkayastha S, Juhasz P, Martin S, Bartlet-Jones M, He F, Jacobson A, Pappin DJ. 2004. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* 3:1154−1169.

Roux A, Lison D, Junot C, Heiler J. 2011. Applications of liquid chromatography coupled to mass spectrometry-based metabolomics in clinical chemistry and toxicology: a review. *Clin Biochem* 44(1):119-135.

Ruggiero A, Marchant J, Squeglia F, Makarov V, De Simone A, Berisio R. 2013. Molecular determinants of inactivation of the resuscitation promoting factor B from Mycobacterium tuberculosis. *J Biomol Struct Dyn* 31:195–205.

Saito Y, Furukawa T, Arano Y. 2010. Basic study on SH2 domain of Grb2 as a molecular probe for detection of RTK activation. *Int J Oncol* 37(2):281-287.

Sassetti, CM, Boyd DH, Rubin EJ. 2003. Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol* 48:77–84.

Sato YT, Watanabe S, Kenmotsu T, Ichikawa Y, Teramoto J, Imanaka T, Ishihama A, Yoshikawa K. 2013. Structural change of DNA induced by nucleoid proteins: growth phase-specific Fis and stationary phase-specific Dps. *Biophysical journal* 105:1037-1044.

Schnappinger D, Ehrt S, Voskuil MI, Liu Y, Mangan JA, Monahan IM, Dolganov G, Efron B, Butcher PD, Nathan C, Schoolnik GK. 2003. Transcriptional adaptation of Mycobacterium tuberculosis within macrophages: insights into the phagosomal environment. *J Exp Med* 198:693–704.

Schermann, J-P. 2008. Experimental Methods. *Spec and Model of Biomolec Buil Bloc* 129-207.

Schubert OT, Ludwig C, Kogadeeva M, Zimmermann M, Rosenberger G, Gengenbacher M, Gillet LC, Collins BC, Röst HL, Kaufmann SHE, Sauer U, Aebersold R. 2015. Absolute proteome composition and dynamics during dormancy and resuscitation of Mycobacterium tuberculosis. *Cell Host Microbe* 18:96 –108.

Schubert OT, Mouritsen J, Ludwig C, Röst HL, Rosenberger G, Arthur PK, Claassen M, Campbell DS, Sun Z, Farrah T, Gengenbacher M, Maiolica A, Kaufmann SHE, Moritz RL, Aebersold R. 2013. The Mtb proteome library: a resource of assays to quantify the complete proteome of Mycobacterium tuberculosis. *Cell Host Microbe* 13:602– 612.

Scigelova M, and Makarov, A. 2006. Orbitrap mass analyzer-overview and applications in proteomics. *Pract Proteo* 1-6.

Senoh M, Ghosh-Banerjee J, Ramamurthy T, Hamabata T, Kurakawa T, Takeda M, Colwell RR, Nair GB, Takeda Y. 2010. Conversion of viable but non culturable Vibrio cholerae to the culturable state by co-culture with eukaryotic cells. *Microbiol Immunol* 54:502–507.

Shleeva M, Mukamolova GV, Young M, Williams HD, Kaprelyants AS. 2004. Formation of “non-culturable” cells of Mycobacterium smegmatis in stationary phase in response to growth under suboptimal conditions and their Rpf-mediated resuscitation. *Microbiology* 150:1687–1697.

Smaoui S, Siala M, Hadj Fredj S, Kammoun S, Marouane C, Hachicha S, Ghorbel A, Gdoura R, Slim L, Ben Messaoud T, Messadi F. 2016. Molecular characterization of Mycobacterium tuberculosis strains resistant to isoniazid. *Int J Mycobacteriol* 5(Suppl 1):S151.

Smeulders MJ, Keer J, Speight RA, Williams HD. 1999. Adaptation of Mycobacterium smegmatis to stationary phase. *J Bacteriol* 181:270–283.

Swiercz JP, Nanji T, Gloyd M, Guarné A, Elliot MA. 2013. A novel nucleoid-associated protein specific to the actinobacteria. *Nucleic Acids Res* 41:4171-84.

Takach EJ, Hines WM, Patterson DH, Juhasz P, Falick AM, Vestal ML. 1997. Accurate mass measurements using MALDI-TOF with delayed extraction. *J Protein Chem* 16(5):363–9.

Tan X, Varughese M, Widger WR. 1994. A light-repressed transcript found in Synechococcus PCC 7002 is similar to a chloroplast-specific small subunit ribosomal protein and to a transcription modulator protein associated with sigma 54. *J Biol Chem* 269:20905–20912.

Tang XX, Fok KL, Chen H, Chan KS, Tsang LL, Rowlands DK, Zhang XH, Dong JD, Ruan YC, Jiang X, Yu SS, Chung YW, Chan HC. 2012. Lymphocyte CFTR promotes epithelial bicarbonate secretion for bacterial killing. J Cell Physiol 227:3887–3894.

Telkov MV, Demina GR, Voloshin SA, Salina EG, Dudik TV, Stekhanova TN, Mukamolova GV, Kazaryan KA, Goncharenko AV, Young M, Kaprelyants AS. 2006. Proteins of the Rpf (resuscitation promoting factor) family are peptidoglycan hydrolases. *Biochemistry* 71:414–422.

Teramoto J, Yoshimura SH, Takeyasu K, Ishihama A. 2010. A novel nucleoid protein of Escherichia coli induced under anaerobiotic growth conditions. *Nucleic Acids Res* 38:3605–3618.

Theophel K, Schacht VJ, Schlüter M, Schnell S, Stingu CS, Schaumann R, Bunge M. 2014. The importance of growth kinetic analysis in determining bacterial susceptibility against antibiotics and silver nanoparticles. *Front Microbiol* 5:544.

Trahan C, Oeffinger M. 2016. Targeted cross-linking-mass spectrometry determines vicinal interactomes within heterogeneous RNP complexes. *Nucleic Acids Res* 44:1354-1369.

Ueta M, Ohniwa RL, Yoshida H, Maki Y, Wada C, Wada A. 2008. Role of HPF (hibernation promoting factor) in translational activity in Escherichia coli. *J Biochem* 143:425– 433.

Ueta M, Yoshida H, Wada C, Baba T, Mori H, Wada A. 2005. Ribosome binding proteins YhbH and YfiA have opposite functions during 100S formation in the stationary phase of Escherichia coli. *Genes Cells* 10: 1103–1112.

Vaniya A, Fiehn O. 2015. Using fragmentation trees and mass spectral trees for identifying unknown compounds in metabolomics*. Anal Chem* 69:52-61.

Vestal ML. 1984. High-performance liquid chromatography-mass spectrometry. *Science* 226(4672):275-281.

Voskuil MI, Visconti KC, Schoolnik GK. 2004. Mycobacterium tuberculosis gene expression during adaptation to stationary phase and low-oxygen dormancy. *Tuberculosis (Edinb)* 84:218 –227.

Vulin C, Leimer N, Huemer M, Ackermann M, Zinkernagel AS. 2018. Prolonged bacterial lag time results in small colony variants that represent a sub-population of persisters. *Nat Commun* 9:4074.

Wada M, Kano Y, Ogawa T, Okazaki T, Imamoto F. 1988. Construction and characterization of the deletion mutant of hupA and hupB genes in Escherichia coli. *J Mol Biol* 204:581-591.

Wayne LG, Hayes LG. 1996. An in vitro model for sequential study of shiftdown of Mycobacterium tuberculosis through two stages of nonreplicating persistence. *Infect Immun* 64:2062–2069.

Wayne LG, Lin KY. 1982. Glyoxylate metabolism and adaptation of Mycobacterium tuberculosis to survival under anaerobic conditions. *Infect Immun* 37:1042–1049.

Wayne LG. 1994. Dormancy of Mycobacterium tuberculosis and latency of disease. *Eur J Clin Microbiol Infect Dis* 13:908 –914.

Wiley WC, McLaren IH. 1995. Time-of-flight mass spectrometer with improved resolution. *Rev Sci Instrum* 26:1150.

Wong HC, Wang P, Chen SY, Chiu SW. 2004. Resuscitation of viable but non-culturable Vibrio parahaemolyticus in aminimum salt medium. *FEMS Microbiol Lett* 233:269–275.

Xiao W, Oefner P. 2001. Denaturing high performance liquid chromatography: a review. *Hum Mut* 17(6):439-474.

Xu Y, Guo J, Jin X, Kim, JS, Ji Y, Fan S, Quan CS. 2016. Crystal structure and functional implications of the tandem-type universal stress protein UspE from Escherichia coli. *BMC Struct Biol 16*(3)53-59.

Yang Y-L, Xu Y, Straight P, Dorrestein PC. 2009. Translating metabolic exchange with imaging mass spectrometry. *Nat Chem Biol* 5(12):885–7.

Yates J. 2000. Mass spectrometry from genomics to proteomics. *Trends in Gen* 16(1):5-8.

Young M, Artsatbanov V, Beller HR, Chandra G, Chater KF, Dover LG, Goh EB, Kahan T, Kaprelyants AS, Kyrpides N, Lapidus A, Lowry SR, Lykidis A, Mahillon J, Markowitz V, Mavromatis K, Mukamolova GV, Oren A, Rokem JS, Smith MC, Young DI, Greenblatt CL. 2010. Genome sequence of the Fleming strain of Micrococcus luteus, a simple free-living actinobacterium. *J Bacteriol* 192:841-860.

Yuan Y, Crane DD, Barry CE III. 1996. Stationary phase-associated protein expression in Mycobacterium tuberculosis: function of the mycobacterial alpha-crystallin homolog. *J Bacteriol* 178:4484–4492.

Zarembinski TI, Hung LW, Mueller-Dieckmann HJ, Kim KK, Yokota H, Kim R, Kim, SH. 1998. Structure-based assignment of the biochemical function of a hypothetical protein: a test case of structural genomics. *Proc of the Nat Acad of Scie* 95(26):15189-15193.

Zenobi R, Knochenmuss R. 1998. Ion formation in MALDI mass spectrometry. *Mass Spec Rev* 17(5):337-366.

Zhang W, Chait BT. 2000. ProFound: an expert system for protein identification using mass spectrometric peptide mapping information. *Anal Chem* 72(11):2482–9.

Zhang Y. 2004. Persistent and dormant tubercle bacilli and latent tuberculosis. *Front Biosci* 9:1136 – 1156.

Zhu W, Plikaytis BB, Shinnick TM. 2003. Resuscitation factors from mycobacteria: homologs of Micrococcus luteus proteins. *Tuberculosis (Edinb)* 83:261–269.

Zubarev R and Makarov, A. 2013. Orbitrap mass spectrometry. *Anal Chem* 85(11):5288-5296.

Zybailov B, Mosley AL, Sardiu ME, Coleman MK, Florens L, Washburn MP. 2006. Statistical analysis of membrane proteome expression changes in Saccharomyces cerevisiae. *J Proteome Res* 5:2339 –2347.