Mass Spectrometry and Fourier Transform Infrared Spectroscopy for Analysis of Biological Materials

By

Timothy J. Anderson

A dissertation submitted to the graduate faculty

In partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Analytical Chemistry

Program of Study Committee:

R. Sam Houk, Major Professor
Young-Jin Lee
Emily Smith
Joseph Burnett
Jesudoss Kingston

Iowa State University

Ames, Iowa

2014

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

GENERAL INTRODUCTION

Mass Spectrometry

Chemistry at its most basic foundation is the study of elements and compounds. Analytical chemistry strives to produce and utilize instruments or tools that assist scientists in qualitatively and quantitatively understanding matter. Few analytical tools have had as large an impact on chemistry as the mass spectrometer (MS). A MS analyzes materials based on their mass-to-charge ratio (m/z). As long as an element or a material has a positive or negative charge in gas phase, the MS has the potential to observe it.

The foundation of mass spectrometry began in the mid-19th century, first with the observation of canal rays by Goldstein in 1886 (1). Canal rays are beams of positive ions produced from gas discharge tubes. It was shown by Wien in 1898 that the trajectory of canal rays could be altered using electric and magnetic fields (1). This observation proved that canal rays were charged particles. In that same year Thomson was able to deduce the mass-to-charge ratio of the electron (1). In 1905 Thompson began to study canal rays by following Wien's experiments. The instrument Thomson developed used magnetic and electric fields to deflect gas ions from their original path. Improvements to the instrumentation allowed him to observe deflection parabolas of various ions from hydrogen, oxygen, chlorine, and phosgene. He replaced the photographic plate with a metal plate that had a slit, and then he placed a second plate attached to an electroscope to determine ion abundance vs mass (2).

In 1913 Thomson analyzed neon and observed parabolas at m/z 20 and 22. The m/z 22 isotope parabola in the spectrum of neon disproved the proposed view that all elements consisted

of atoms with a single mass number. Thomson's assistant, Aston, continued detailed studies of isotopes. In 1919 Aston built a new mass spectrograph which could collimate the parabolas into discrete lines, and had a resolution of 130. The instrument first used an electric field between parallel plates to separate the ions, second an electromagnet with a gap between the poles focused the ions to a spectrographic photoplate. The new instrument had the capability to not only accurately separate isotopes of neon, it could record the isotope abundances. By 1924, through the use of the mass spectrograph, fifty three of the known eighty elements had been measured for mass and abundance (1, 2). Based on the pivotal work of these early pioneers, MS has become critical for analytical chemistry and the broader scientific community from World War II into present times. A comprehensive history of critical MS discoveries can be found elsewhere (2).

Modern Mass Spectrometers and Detectors

In the past twenty years the use of MS for biological materials has matured. A variety of commercial instruments have emerged to meet these needs. MS instruments produced today can have high resolution, high mass ranges, rapid scan rates, and are relatively user friendly. However, modern instruments still produce results similar to Thomson's first mass spectrometer. MS instruments still measure intensity and m/z of the ionized elements or chemical compounds of interest. The following descriptions are brief functional overviews of the most common modern MS instruments used for biological analysis.

Quadrupole Mass Spectrometer

In the early 1950's a German physicist named Paul developed the quadrupole mass analyzer (3-5). Figure 1-1 shows that the quadrupole mass selectors consist of four hyperbolic or cylindrical metal rods equivalently spaced from each other (6). The rods positioned directly opposite each other carry the same positive or negative voltage Φ_0. The adjacent rods share the same magnitude of applied voltage; however they carry the inverse charge, considering there is no bias voltage. The constant voltage Φ_0 on the quadrupole rod is the summation of a direct current (DC) voltage U and radio frequency (RF) voltage V with a driving frequency ω, as shown below:

$$\Phi_0 = U + V\cos\omega t. \quad (7)$$

The applied voltages on the rods produce a field that deflects the ions as they pass between the rods. The field that is produced from a respective applied voltage to the rods permits stable oscillations of only a narrow range of m/z ions. All other ions that traverse the quadrupole field oscillate with increasingly unstable amplitudes until they hit the rods (7). The quadrupole can thus be used for MS to select masses and measure their intensity with the appropriate detector.

A popular version of the quadrupole mass spectrometer is the triple quadrupole mass spectrometer used for tandem mass spectrometry (8). The first quadrupole in the instrument is generally a mass selector for a m/z species of interest. The selected ions pass into a second multipole with only a RF voltage that is set to excite the ions for fragmentation (9). A buffer gas of either N_2, He, or Ar is introduced into the second quadrupole. The transfer of kinetic energy

from collisions with the buffer gas causes ion fragmentation. After fragmentation, the ions are transferred to a third quadrupole that measures the ion fragments of the parent ion.

Time-of-Flight Mass Spectrometer

Time-of-flight (TOF) MS was first proposed in 1946 by William Stephens (10). By 1948 a functional TOF instrument called the Velocitron was built by Cameron and Eggers (11). TOF instruments function as indicated by their name; they measure the flight time of a packet of ions from the injection source to the detector. An extraction pulse accelerates ions, then a m/z dependent separation occurs as they then traverse a field free zone to reach the detector (12). Due to the function of the TOF instrument, it has many proposed benefits, such as microsecond scan times for entire spectra, nearly unlimited mass range, and high resolution (12). The TOF instrument can achieve fast scan times because it can measure the entire m/z range in a single scan; however, there is a limit to the scan speed of the instruments. To prohibit spectral overlap with the previous scan, the instrument cannot begin the next scan pulse until the slowest ion from the previous scan has reached the detector (13). The high mass range for TOF is achieved by altering the extraction pulse and other various instrument parameters to allow massive ions to be separated. Ions as large as intact virus capsids have been measured using TOF MS (14).

TOF instruments began as low resolution instruments. It was common for basic linear TOF instruments to have a resolution of approximately 100 (15). Effective flight length, the thickness of the injected ion packet, and the extraction pulse shape are factors that affect TOF instrument resolution. The relation of effective flight length (L_{eff}) and thickness of ion packet (Δz) to TOF resolution (R_{FWHM}) is shown below (13):

$R_{FWHM} \approx L_{eff} / 2\Delta z$.

To improve TOF resolution, innovations to instrumentation have been developed to increase L_{eff} and decrease Δz (12, 15). To improve resolution of TOF instrumentation, the reflectron was devised to compensate for kinetic energy spread. Figure 1-2 details a reflectron instrument, which allows the bending and altering of the ion path so as to make as efficient use of the chamber space as possible and optimize L_{eff}. For early instruments Δz was a major hindrance because it was extremely difficult to focus all of the ions on the same "starting line". Orthogonal time of flight helped alleviate this issue by using lenses and slits to produce a narrow axial ion packet prior to the entrance of the TOF. The ions are injected by an ion modulator which orthogonally accelerates the ions (16). Most modern TOF instruments combine orthogonal ion injection and reflectron configurations to maximize the L_{eff} and Δz relationship to produce TOF resolutions greater than 10,000 (13).

Fourier Transform Ion Cyclotron Resonance and Orbitrap Mass Spectrometers

The highest resolution MS are Fourier transform ion cyclotron resonance (FT-ICR) instruments reaching resolutions of 1,000,000 or greater (17). The instrument was invented in 1974 by Marshall and Comisarow (18). FT-ICR MS functions as a trapping and analyzing device for ions in three-dimensional space, as shown in Figure 1-3. Ions are injected into a cell with a strong magnetic field applied in the same direction as injected ions. The magnetic field exerts a rotational force on the ions perpendicular to the magnetic field. As the ion packet is injected into the cell, the fore and aft plates are charged to trap the ions in the cell (19).

Many aspects of FT-ICR MS make it a powerful technique for biological analysis. The ultra-high resolution is obtained because the ions trapped within the ICR cell move extremely fast and they have small orbital radii, also, the ion cyclotron frequencies are almost completely independent of the spread of velocity in the axial direction. An ion can oscillate between the detector plates ~500,000 times per second (19). Such a large number of data points in such a short time can produce incredible signal-to-noise ratios. The FT-ICR instrument is also a popular choice for tandem mass spectrometry. The excitation plates can increase the kinetic energy of ions at selected m/z values for CID and subsequent MS^n experiments can be carried out (19).

The Orbitrap instrument is the newest high resolution MS. The first practical orbital trapping instrument, used purely for ion trapping was first devised by Kingdon in 1923 (20). Later, in 1981 Knight produced a modified trap device that improved upon the instrument proposed by Kindgon (21). It wasn't until 2000 that Makarov demonstrated with an early orbital trapping instrument that the orbitrap could be used as a MS (22). Figure 1-4 shows that the Orbitrap consists of a Saturn shaped central electrode that is slim at both ends, but converges into a slightly larger diameter at its center. The central electrode is surrounded by two symmetrical electrodes (22-24). Ions enter the Orbitrap off axis from the center rod's equatorial position and are allowed to settle into ring-like orbits around the rod. The rings oscillate back and forth over the electrode at a frequency dependent upon their m/z. The current image of the charged rings is recorded by the split outer electrodes and converted into signal amplitude vs time. The spectrum is converted via Fourier transform into intensity vs frequency, and then the plot is inverted into intensity vs m/z (22-24).

Detectors

One of the most important, but least appreciated components of a MS is the detector. The electron multiplier (EM) is the most widely used detector for biological TOF and quadrupole MS instruments. A conventional EM detector for many scanning quadrupole instruments is comprised of a dynode that is positioned to collect the ions of interest from the MS. The EM is built with many dynodes, each subsequent dynode has a lower voltage until terminating with an anode (25). The ions that hit the dynode release secondary electrons from the initial collision producing a cascade of electrons. The cascade then reaches the detector anode which converts the electron current into signal to be analyzed. EM detectors are useful for biological instruments because scanning across a spectrum may provide few ions for detection. The EM amplifies the signal and allows low ion counts to be recorded. A negative aspect of EM detectors is the ability for loss of signal linearity when the detector is saturated with too many ions and the secondary electron cascade overwhelms the anode (25).

For TOF instruments a different EM detector called the microchannel plate (MCP) is a popular choice. The MCP is a thin disk that consists of numerous parallel semiconducting channels that traverse the plate. As an ion hits the entrance of the cavity, secondary electrons are released and start a cascade down the detector (25, 26). It also has provides an extremely fast response time. However, there are many downsides to MCP detectors. If an ion hits a cavity, it drains the electron density of adjacent cells for several microseconds (26). If another ion hits at the same point within that period it is not counted, and causes decreased detection linearity. Lastly, MCP plates are relatively fragile and must be handled carefully and maintained in low pressure conditions.

Cryogenic detectors are another detector variant useful for biological TOF MS instruments. The cryogenic detector measures the thermal energy produced by ions impinging upon a superconducting film (26). They have been used successfully in TOF devices to measure extremely large molecules greater than a MDa (27). Cryogenic detectors also have the capability of discerning the charge of large molecular species (25, 26). The cryogenic detector has a major downside; the detector must be maintained at very low temperature (~2 K). Despite the difficulties with combining an MS instrument to a cryogenic detector, a commercial matrix assisted laser desorption ionization (MALDI) TOF instrument called the "macromizer" has been produced (25).

MS instruments that utilize image-current detectors have begun to be used heavily for biological MS. As described above, the image-current devices detect ion frequencies between two electrodes as they oscillate in an ion trap. The major benefits to image-current devices are due to their ruggedness, high resolution, and non-destructive ion analysis. The downsides are that FT-ICR instruments must be cryogenically cooled and have space charge constraints. The major downsides to Orbitrap instrument are a) space charge concerns, and b) acquisition times can be greater than a second to obtain high accuracy and resolution.

Separation Techniques Coupled to Mass Spectrometry

From the 1920's until the 1950's mass spectrometry was pivotal for many important breakthroughs to better understanding elemental ions and simple molecular ions. It wasn't until 1955, at the Dow Chemical Company, that MS got a new wind when gas chromatography (GC) device and MS were coupled together (28). The pairing of these two devices showed that by applying a separation prior to analysis that compound identities and quantitative measurements

could be made. Since then a multitude of separation techniques have been coupled to MS. The following are brief discussions of the major separation techniques paired with MS for biological analysis.

Gas Chromatography

The GC excels at separating a wide variety of volatile compounds (29). Samples are primarily dissolved and prepared in volatile solvents. The solvent is then injected by syringe into a heated port that vaporizes the solvent along with the dissolved analytes. The sample then flows through an oven-heated fused silica capillary tube that is commonly coated or filled with a stationary phase (30). The stationary phase is a material that has the potential to interact with the various analytes based on its unique side-group specificity (30). If an analyte interacts with the stationary phase it is apt to flow more slowly than other analytes, thus allowing a separation to occur. As the analytes leave the column they are directed to the MS. If the separation has been performed adequately, the individual compounds can be identified in order of retention without overlap using electron impact ionization (EI) described below.

Liquid Chromatography

For separating compounds which are high molecular weight, non-volatile, or unstable at high temperatures, liquid chromatography (LC) is a favorable technique (31). Sample preparation is generally simple. A sample can either be prepared by extraction from a biological material, or by dissolving desired compounds into the appropriate solvent. The sample “plug” is then injected into the LC solvent flow that runs through the system. The solvent flow is set to a certain pressure, temperature, and solvent gradient prior to reaching the column (31). With basic

503 reverse-phase LC setups, the column is generally packed with porous silica particles that are
504 typically 2 to 50 μm and bonded with C_{18} stationary phase. The sample analytes interact with the
505 stationary phase and the polar compounds pass through the column more quickly than the non-
506 polar compounds which interact strongly with the stationary phase (31). Commonly, a solvent
507 gradient that gradually increases the concentration of less-polar solvent is utilized. As the LC
508 solvent becomes more concentrated with a less-polar solvent, the interaction of the analyte with
509 the mobile phase increases, while the interaction of the analyte with the stationary phase is
510 lessoned. This allows for faster run times and better resolution of eluent peaks (31). After
511 separation, the analytes must be converted to the gas phase for MS detection. Electrospray
512 ionization (ESI) is a common method of interfacing LC and MS and is discussed below.
513 LC excels in the breadth of the analytes it can separate. It can separate the smallest of
514 dissolved chemical species to extremely high molecular weight proteins. To improve and speed
515 up the separation of analytes there are mainly three improvements applied to LC. The first is to
516 alter the stationary phase bead. By decreasing the size of the bead the surface area is increased
517 to allow analyte interaction (31). However, the beads are also porous and dwell time is an issue.
518 The interaction can be improved by using a solid core and porous coating to decrease the dwell
519 time (32). The second is to utilize high pressure instrumentation that reduces dead volume and
520 allows the sample to move quickly through a column and decrease longitudinal band broadening
521 (31). Also, high pressure instrumentation is important with smaller stationary phase beads, as
522 new volume flow constraints are produced. The third is the use of specialized stationary phases
523 for enhanced separation for specific analytes. In the case of chiral separation, a special chiral
524 stationary phase must be used to provide an interaction that separates the nearly identical
525 analytes (33).

For basic protein analysis by gel filtration/size exclusion chromatography, a porous gel stationary phase is utilized to separate the proteins based on shape and size (34). Larger proteins elute first, while smaller proteins dwell in the pores and thus elute more slowly. Along with the pore-size, varying types of stationary phase can be entrapped within the gel to interact with the protein functional groups to further enhance the separation.

Ion Mobility Spectrometry

For complex biological mixtures, ion mobility spectrometry (IMS) has the capability to separate a variety of compounds, for example, isomeric reverse peptides, proteins by charge state, and large protein complexes (35-37). IMS in conjunction with MS can provide important structural information about gas-phase biomolecular ions (38, 39). A variety of many conformation-based biomolecular studies have utilized IMS MS to probe protein folding, oligosaccharide characterization, and even determine whether proteins retain solution phase characteristics in the gas phase (36, 40-47).

Commonly, IMS utilizes a drift tube. The drift tube is filled with a buffer gas, typically He, and has a uniform electric field applied axially down the drift tube. As an ion moves through the drift tube its drift time through the IMS device is related to the number of collisions with buffer gas, the electric field applied, its initial velocity, and the length of the drift tube (48). This relation allows ions to be separated based upon their charge state and size. Ions which are highly charged traverse the drift tube faster than low charge state ions. Ions which are compact have fewer collisions with the buffer gas and traverse the drift tube more quickly than those with elongated conformations.

Recently, another form of IMS, called traveling wave IMS has been utilized. A photograph of a traveling wave device is shown in Figure 1-5. Figure 1-6 shows a diagram of travelling wave IMS separation. Traveling wave IMS follows the same principles of basic IMS by separating by charge and size, however it utilizes time varying AC symmetrical potential waves for separation (49). The wave speed and wave amplitude can both be modulated to produce a desired separation of ions. The separation can be visualized by thinking of ions as buoys on the ocean. A small buoy will not have the ability to crest a wave that washes over it and will be taken along with the wave. Larger buoys will crest the wave and propagate more slowly. Based upon this separation method smaller, and highly charged ions are carried further per wave event than larger less charged ions and thus produce a sizable separation after many wave events.

Ionization Techniques

The field of mass spectrometry relies heavily on the production of ions. Because of the various properties inherent to each analyte compound, the production of ions can be one of the most difficult and frustrating aspects of mass spectrometry. Compounds which do not have such satisfactory side-groups have to be ionized using various high energy ionization techniques, such as, electron ionization (EI), which is discussed further below. A great number of ionization techniques exist for mass spectrometry, some extremely specialized for the production of a specific ion species.

Electron Ionization

The ionization method EI was the first major biological ionization method, and is still one of the most popular (50). It was first used by Dempster (51) and was further developed by Nier (52). Currently, EI is the most commonly used ion source utilized with GC-MS instruments. GC excels at volatilizing complex solvent mixtures prior to the column for selective separation, and then injection to the EI source for ionization (53). The EI source requires that an analyte vapor is passed through a reduced pressured chamber that crosses an electron beam produced by a tungsten or rhenium filament (50). The electron beam is most commonly tuned to approximately 70 eV for optimal analyte ionization (54). The ionization process causes the neutral analyte molecule to lose an electron through a molecule/electron collision interaction (50). The ionization process is very energetic and it is common that the ion produced is unstable and fragments into smaller ionized species. Due to the large number of highly energetic ions produced from an analyte, fragmentation is extensive and a nearly complete distribution of the analyte's unique structural characteristics can be elucidated from the mass spectrum (50). The mechanism of analyte ionization by EI is deceptively complex and has been thoroughly described elsewhere (55).

Electrospray Ionization

Electrospray ionization (ESI), unlike EI, is a "soft" ionization technique which rarely causes large scale fragmentation of analyte molecules. ESI is also a much more selective ionization method than EI. ESI negative and positive ion production is most commonly centered around the abstraction or addition of protons to the analyte, respectively. Because of this reliance on protons for ionization, ESI is typically useful for protein and peptide analytes due to

the abundance of acidic and basic side groups. Regular ESI consists of a solvent system that includes the analyte of interest. The solvent is pumped through a capillary at flow rates of 1 to 100 µL/min, along with an applied voltage to the solution of approximately 2 to 5 kV. As the solution is electrically charged, a gradient is produced within the capillary and a Taylor cone is produced as the solution is expelled (56). The excess charge in the solution migrates to the exterior of the solvent. As the charge builds upon the surface of the solvent, droplets are expelled from the Taylor cone in a stream. The rate of the droplet formation increases until the Rayleigh limit, where the positive and negative charge hit equilibrium in conjunction with the surface tension of the solution. The highly charged droplets burst repeatedly as the repulsive charge splits the droplets apart until only gas phase ions of the analyte remain. Due to the excess charge placed within each droplet, it is common for peptides or protein analyte ions of multiple charge states to be produced.

In many situations only very small samples of protein or peptides are available for analysis. A popular form of ESI that accommodates this scenario is nano-ESI. The flow rates utilized for nano-ESI are substantially lower, approximately 10 to 40 nL/min. The inner diameter of the capillary used for nano-ESI is also significantly decreased to 200 nm, in comparison to 1 to 2 µm for ESI (57). The combination of the low flow rate and the decreased capillary size typically allows nano-ESI to produce finer droplets than ESI, which accounts for the generally more stable spray observed with nano-ESI.

Laser Desorption/Ionization

MALDI is a popular technique to ionize biological analytes. Typically proteins or peptides are ionized using MALDI, but many other biological compounds can also be ionized.

The matrix is the key player in the ionization of the analyte. MALDI matrices are usually organic acids, such as 2,5-dihydroxybenzoic acid (DHB) or α-cyano-4-hydroxycinnamic acid (CHCA) that absorb the wavelength of light emitted by a UV laser (58, 59). The sample preparation entails co-crystallization of the analyte with the matrix at a ratio that can be as high as 1000:1 matrix to analyte. During analysis, the laser irradiates the sample, the matrix absorbs the light and causes a small desorption event that releases matrix and analyte. The energetically excited matrix donates or abstracts a proton from the analyte causing it to become ionized. The analyte is usually singly charged, but in some instances, MALDI ions are doubly charged (60). The basic understanding of the matrix and analyte ionization is relatively well understood; however, the exact mechanism of the charge transfer from the matrix to analyte is still uncertain (61-63).

MALDI is a powerful ionization technique for biological sample analysis; but, it has a few major disadvantages. Some of the most glaring issues lie with the matrix. The matrix, as described previously, must be used in amounts that are much higher than the analyte. Since the matrix is ionized along with the analyte, it is common for large background signals from matrix and its fragment ions to drown out low molecular weight (M_r) analytes. Careless MALDI sample preparation commonly produces homogenous crystals that are not co-crystallized with the analyte, thus producing poor analyte ionization, reproducibility, and signal stability. Some analyte compounds, such as small nonpolar lipids, will not co-crystallize at all with traditional MALDI matrix and produce poor signal quality. Last, MALDI most often performs best at modest vacuum pressures (~1 torr). At these pressures it is common to lose many volatile analytes of interest prior to sample analysis.

Because of complications with MALDI discerning volatile analytes, atmospheric pressure laser desorption (AP LDI) has gained popularity (64, 65). Imaging of biological tissues has been popular for MALDI, but a great deal of focus has been placed on AP LDI because of the potential to observe both volatile and nonvolatile analytes (66-70). However, AP LDI has a variety of negative aspects. Because AP LDI is performed at atmospheric conditions, ionization and sensitivity are significantly reduced in comparison to MALDI. At atmospheric pressure, an ion encounters countless collisions with the gases in the atmosphere. Many of these atmospheric components will buffer the ions from reaching the MS source, or scavenge the charge from the ion itself. It is a highly unlikely event that ions from AP LDI will reach the MS. To remedy these issues, a variety of post ionization techniques have been developed. Laser ablation electrospray ionization (LAESI), ablates large portions of sample, then utilizes ESI to ionize neutral analyte from the ablation plume (71, 72). Electrospray assisted desorption/ionization (ELDI) uses low laser power to desorb analyte from a sample and then uses ESI to post ionize, similarly to LAESI (73). Laser ablation atmospheric pressure photoionization (LA APPI) ablates the sample, and as it is liberated a vacuum UV lamp is used to assist in ionizing UV absorbing analyte molecules (74). APPI post ionization can be utilized as a secondary ionization method when a UV absorbing dopant, such as toluene is used. This dopant can absorb the UV light, ionize, and then collide and transfer charge to the analyte before it enters the MS.

Mass Spectrometry for Proteomics

MS is a powerful tool that has been used to assist in the identification of proteins. The first step to comprehensively identify a protein is to use the M_r. High resolution MS instrumentation and the use of deconvolution can assist in the determination of a protein M_r;

662 however, due to the complexity of proteins, it is not possible to accurately identify the protein's
663 primary sequence. A common and more involved method of using MS to identify the primary
664 sequence of a protein is bottom-up proteomics. Generally, a purified unknown protein is
665 subjected to a proteolytic digestion that lyses the protein into peptide fragments. The digest is
666 then analyzed using LC-MS, or LC-MS/MS for a more complete view of the protein sequence
667 (75). The LC separation allows for separation of isobaric peptides prior to MS/MS analysis.
668 MS/MS is performed using collision-induced dissociation (CID). CID is performed in a
669 multipole and contains a collision gas of N_2 or Ar. The peptide ions are accelerated within the
670 multipole and will fragment as they strike the collision gas, resulting in b- and y-ion fragments
671 (76). Through the use of various software and databases, the fragment ion information can be
672 used with M_r to determine the de novo sequence of the protein (77-79).
673 A more recent method is top-down proteomics developed by McLafferty et al. (80) The
674 whole protein is introduced into the instrument and each of the charge states is isolated and
675 dissociated. The accurate mass of the parent protein and the fragments are then scrutinized and
676 the protein is identified. The main benefit to top-down proteomics is that there is little sample
677 preparation; however, a specialized high-resolution instrument that can isolate and dissociate
678 proteins is necessary. Top-down proteomic techniques have shifted, and the most common are
679 electron transfer dissociation (ETD) (81) and electron capture dissociation (ECD) (82-84). In
680 ETD an ion/ion reaction occurs as the protein is trapped in the instrument with a radical anion,
681 such as azobenzene ($[(C_6H_5)_2N_2]^{-\bullet}$). The radical anion trapped with the protein donates the spare
682 electron to the protein cation, which destabilizes the protein backbone and causes it to fragment.
683 ECD accomplishes the same fragmentation mechanism as ETD, yet uses free electrons captured

and trapped with the proteins. Since ETD and ECD fragment through a different mechanism than CID, the fragments produced are c- and y-ions (64).

Mass Spectrometry for Metabolomics

Metabolomics is currently considered the final major "omics". Metabolomics does not neatly fit within the same categorical structure as the other "omics", such as genomics, transcriptomics, and proteomics which study DNA, all forms of RNA, and proteins, respectively. The first three "omics" encompass only characterizing the nucleic acids and the combination of the 22 amino acids. However, metabolomics encompasses all compounds which are produced by living cells. Because of this broad classification, and the wide variety of compounds that can be observed as metabolites, metabolomics is a still growing and very complex field of biology.

There are two forms of metabolomics that are generally utilized. The first is targeted analysis, where the identification and quantification of metabolites have been pre-selected prior to the analysis (85). Targeted analysis is used often for drug or chemical studies where a pathway is understood and the metabolites of that drug have been identified. The second is metabolite profiling, which is the indiscriminate analysis of all metabolite compounds observable using a particular analytical technique (85).

MS is useful for both targeted analysis and metabolite profiling. Targeted analysis of a specific metabolite can be performed using the appropriate form of chromatography with the proper MS instrumentation. However, metabolite profiling requires complex chromatography and MS instrumentation that is suitable to observe a wide range of metabolites. FT-ICR MS is popular for metabolite profiling because of its high resolution (86). Ultra-high resolution FT-ICR MS can observe all ionized metabolites using direct infusion. Broad profiling by FT-ICR

MS can roughly identify metabolites using accurate mass measurements, but it does not have the ability to discern isomers or isobaric ions. LC-TOF MS has recently grown into a technique which can perform metabolite profiling (87). The draw to LC-TOF MS recently is due to the increasingly high resolutions (~30,000) available and the fast scan speeds. By separating metabolites and observing metabolites over time, the TOF MS is not hampered by its lower resolution and can profile metabolites by accurate mass similar to FT-ICR MS. LC separation can also be utilized to assist in accurate mass identification if retention times of known metabolite standards are aligned with the retention times of the unknown sample metabolites. GC-TOF MS is extremely useful for profiling derivatized or volatile metabolites (85). A major benefit to GC is that it allows for the complete identification of the profiled metabolites through the use of GC libraries.

Fourier Transform Infrared Spectroscopy

For comprehensive characterization of biological materials, such as proteins, lipids, and carbohydrates it is amenable to utilize other analytical techniques that complement MS. MS excels at inferring information from individual ions, techniques such as Fourier Transform Infrared Spectroscopy (FTIR) analyze the entire sample simultaneously and infer information from all of the analyte species.

The layout of the common double-beam interferometer, which most FTIR instruments use today, was developed in 1891 by Michelson (88). For a detailed explanation of how the Michelson interferometer is configured and produces an interferogram, please see Griffiths and de Haseth (88). The interferogram produced by the two-beam instrument is transformed using Fourier transform computation to produce an FTIR spectrum (88).

FTIR instrumentation has evolved greatly since the introduction of the Michelson interferometer. After Michelson designed the interferometer and introduced the interferogram, he showed using FT spectroscopy that they could discern the structures of atomic lines. However, due to lack of computational resources, interferometry was not widely used for nearly 50 years. In the 1950's Fellgett showed that all spectral information was simultaneously recorded by the interferogram. This was a huge informational advantage to the interferogram, especially in IR (88). Soon after, Jacquinot determined that the maximum solid angle of collimated light passed through an interferometer was greater than the same beam used with a prism or grating in a monochromator (88). In the 1960's FTIR gained traction with the introduction of compact computers and reliable lasers. The small lab-sized computers allowed for computational processing of the interferogram just after analysis. The laser allowed for precise measurements of the mirror position at equal time intervals, which produces an internal standard for the wavenumber for all measurements (88).

As FTIR was able to take hold as a suitable lab technique, many other techniques that utilize FTIR have been introduced to study biological samples. A common FTIR technique employed is transmission FTIR. The technique measures the amount of IR radiation that is absorbed by a sample as IR radiation is modulated over the spectral range of interest. The FT transformed output describes the bulk properties of the sample material (89). Other sample specific FTIR methods are specular reflection spectroscopy, attenuated total reflectance spectroscopy, diffuse reflectance spectroscopy, and photoacoustic spectroscopy.

Due to its unique ability to be paired with nearly any sample type, photoacoustic spectroscopy was utilized in Chapter 4 of this dissertation, and will be described more fully in the next section.

Fourier Transform Infrared-Photoacoustic Spectroscopy

Fourier transform infrared-photoacoustic spectroscopy (FTIR-PAS) is an absorption method. PAS is based on the photoacoustic effect, which was first described by Alexander Graham Bell (90). He was able to successfully demonstrate the effect with his "spectrophone" (91). It would take nearly a century until PAS would gain popularity again. In the 1980's as FTIR use increased, PAS proved to be a natural pairing with the innate IR modulation and multiplex advantage.

FTIR-PAS can be applied to nearly any solid or semi-solid material (92). This gives FTIR-PAS a large advantage over transmission spectroscopy and reflective spectroscopy methods, which must be used with select sample types. FTIR-PAS also needs very little sample preparation and is non-destructive, the sample must merely fit within the PA sample cell. It performs well with opaque, dark, and highly scattering materials. This is possible because FTIR-PAS does not detect IR radiation; rather it detects the heat produced by the absorption of the IR radiation by the sample. PAS is performed generally with a PAS accessory, shown in Figure 1-7. The PAS cell is placed within a suitable FTIR instrument. The accessory consists of a sealed cell that has been purged with helium gas. The sample material is placed within the cell under a salt window which passes IR radiation. The modulated IR radiation produced by the FTIR passes through the window and impinges upon the sample material. As a discrete wavelength of IR radiation illuminates the sample, the sample will heat if the respective IR radiation is absorbed. Rapidly, the heat will evolve from the sample and produce a pressure wave within the cell. The resultant sound waves are detected by a sensitive microphone and the FTIR and microphone will produce an absorption spectrum usually in the IR range (93, 94). For a thorough explanation of the PAS theory please see the article of Rosencwaig and Gersho (92).

FTIR-PAS Utilized for Biological Materials

Most FTIR techniques work well characterizing individual analyte compounds; however, many fail when confronted with complex biological mixtures. FTIR-PAS works well with these tough samples because tissues can be dried, ground, and then placed directly within the accessory with little prep and can be reused for another methodology. For the analysis of these complex biological mixtures, the pairing of FTIR-PAS and chemometrics is ideal to assist with qualitative and quantitative analysis. Currently, FTIR-PAS is attracting the most attention as a complement to magnetic resonance imaging (MRI) for microscopic imaging of tissues (95-97). Mid-IR FTIR-PAS has become popular in the food science field. It has been employed to analyze intact tissue samples of beef and pork (98). FTIR-PAS has been used to analyze cheddar cheese and the packaging of the cheese (99). The analysis of the cheese using the depth profiling properties of FTIR-PAS showed that moisture was lost near the surface of the cheese, also lipids and proteins were found to migrate into the packaging itself. FTIR-PAS has been able to differentiate between organic and conventional coffee varieties (100). The method has been shown to quantitatively measure the amount of lipid in chocolate (101). FTIR-PAS can glean viable information using an extremely small amount of sample. One study using peas was able to determine total starch, total lipid, and total protein, while using such small amount of sample that the peas were still viable (102). Chapter 4 of this dissertation expanded on the utilization of FTIR-PAS for biological materials. It was used to quantitatively and qualitatively measure resistant starch in unfractionated rat cecal contents for the first time.

Dissertation Overview

This dissertation focuses on the analysis of plant and animal extracts utilizing MS and FTIR-PAS. Chapter 2 details the use of high resolution TOF MS to analyze metabolites extracted from rat cecal and distal colon contents. The goal of the study is to determine whether rats fed four different starch diets have metabolic profiles that could be considered significantly different using partial least squares-discriminate analysis (PLS-DA) software. Chapter 3 delves into a comprehensive identification of alpha-zein proteins using LC separation coupled to a high resolution TOF MS. This is the first study to have comprehensively identified alpha-zein proteins with accurate mass capabilities. The study aims at determining whether alpha-zein protein extract profiles are different when extracted from corn gluten mean and dried distillers grains with solubles in combination with varying extraction conditions. Chapter 4 revisits the analysis of rat cecal and distal colon contents; however, the analysis focuses on starch determination utilizing FTIR-PAS. Many starch analysis techniques are time consuming and expensive, FTIR-PAS is explored to speed up analysis time and decrease cost on large sample runs. FTIR-PAS spectra of the cecal and distal colon contents are processed using PLS to determine quantitative starch values and principal component analysis (PCA) is applied to determine qualitative starch differences.

References

1. Griffiths, I. W. J. J. Thomson-the centenary of his discovery of the electron and of his invention of mass spectrometry. *Rapid Commun. Mass Sp.* **1997**, *11*, 2-16.
2. *Measuring mass: from positive rays to proteins*, Grayson, M. A., Ed.; Chemical Heritage Press: Philidelphia, PA, **2002**.
3. Friedburg, H.; Paul, W. Optische abbildung mit neutralen atomen. *Naturwissenschaften* **1951**, *38*, 159-160.
4. Bennewitz, H. G.; Paul, W. Eine method zur bestimmung von kernmomenten mit fokussiertem atomstrahl. *Z. Phys.* **1954**, *139*, 489.
5. Bennewitz, H. G.; Paul, W. Fokussierung polarer molecule. *Z. Phys.* **1955**, *141*, 6.
6. Giese, C. F. Strong focusing ion source for mass spectrometry. *Rev. Sci. Instrum.* **1959**, *30*, 260-261.
7. Paul, W. Electromagnetic traps for charged and neutral particles. *Rev. Mod. Phys.* **1990**, *62*, 531-540.
8. Yost, R. A.; Enke, C. G. Triple quadrupole mass spectrometry for direct mixture analysis and structure elucidation. *Anal. Chem.* **1979**, *51*, 1251A-1264A.
9. Douglas, D. J.; Aaron, J. F.; Dunmin, M. Linear ion trap in mass spectrometry. *Mass Spectrom. Rev.* **2005**, *24*, 1-29.
10. Stephens, W. E. A pulsed mass spectrometer with time dispersion. *Phys. Rev.* **1946**, *69*, 691.
11. Cameron, A. E.; Eggers, D. F. An ion "velocitron". *Rev. Sci. Instrum.* **1948**, *19*, 605-607.
12. Mamyrin, B. A.; Karataev, V. I.; Shmikk, D. V.; Zagulin V. A. The mass-reflectron, a new nonmagnetic time-of-flight mass spectrometer with high resolution. *Sov. Phys. - JETP*. **1973**, *37*, 45-48.
13. Chernushevich, I. V.; Loboda, A. V.; Thomson, B. A. An introduction to quadrupole-time-of-flight mass spectrometry. *J. Mass Spectrom.* **2001**, *36*, 849-865.
14. Tito, M. A.; Tars, K.; Valegard, K.; Hajdu, J.; Robinson, C. V. Electrospray time-of-flight mass spectrometry of the intact ms2 virus capsid. *J. Am. Chem. Soc.* **2000**, *122*, 3550-3551.
15. Wang, T.-I.; Chu, C.-W.; Hung, H.-M; Kuo, G.-S.; Han, C.-C. Design parameters of dual-stage ion reflectrons. *Rev. Sci. Instrum.* **1994**, *65*, 1585-1589.
16. Dodonov, A. F.; Kozlovski, V. I.; Soulimenkov, I. V.; Raznikov, V. V.; Loboda, A. V.; Zhen, Z.; Horwath, T.; Wollnik, H. High-resolution electrospray ionization orthogonal-injection time-of-flight mass spectrometry. *Eur. J. Mass Spectrom.* **2000**, *6*, 481-490.
17. Marshall, A. G. Fourier transform ion cyclotron resonance mass spectrometry. *Acc. Chem. Res.* **1985**, *18*, 316-322.
18. Comisarow, M. B.; Marshall, A. G. Fourier transform ion cyclotron resonance spectroscopy. *Chem. Phys. Lett.* **1974**, *25*, 282-283.
19. Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom. Rev.* **1998**, *17*, 1-35.
20. Kingdon, K. H. A method for the neutralization of electron space charge by positive ionization at very low gas pressures. *Phys. Rev.* **1923**, *21*, 408-418.
21. Knight, R. D. Storage of ions from laser-produced plasmas. *Appl. Phys. Lett.* **1981**, *38*, 221-223.
22. Makarov, A. Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Anal. Chem.* **2000**, *72*, 1156-1162.

23. Hardman, M.; Makarov, A. A. Interfacing the orbitrap mass analyzer to an electrospray ion source. *Anal. Chem.* **2003**, *75*, 1699-1705.
24. Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Cooks, R. G. The orbitrap: a new mass spectrometer. *J. Mass. Spectrom.* **2005**, *40*, 430-443.
25. Koppenaal, D. W.; Barinaga, C. J.; Denton, M. B.; Sperline, R. P.; Hieftje, G. M.; Schilling, G. D.; Andrade, F. J.; Barnes, J. H. MS detectors. *Anal. Chem.* **2005**, *77*, 419A-427A.
26. Westman-Brinkmalm, A.; Brinkmalm, G. A mass spectrometer's building blocks. In *Mass spectrometry: instrumentation, interpretation, and applications*, Elkman, R.; Silberring, J.; Westman-Brinkmalm, A.; Kraj, A, Eds.; John Wiley & Sons: Hoboken, NJ, **2009**; pp 15-71.
27. Wenzel, R. J.; Matter, U.; Schultheis, L.; Zenobi, R. Analysis of megadalton ions using cryodetection MALDI time-of-flight mass spectrometry. *Anal. Chem.* **2005**, *77*, 4329-4337.
28. Gohlke, R. S.; McLafferty, F. W. Early gas chromatography/mass spectrometry. *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 367-371.
29. *Handbook of instrumental techniques for analytical chemistry*, Settle, F. A. Ed.; Prentice Hall: Upper Saddle River, NJ, **1997**.
30. McNair, H. M.; Miller, J. M. *Basic gas chromatography*, 2nd ed.; Wiley-Interscience: Hoboken, NJ, **2009**.
31. Snyder, L. R.; Kirkland, J. J.; Dolan, J. W. *Introduction to modern liquid chromatography*, 3rd ed.; Wiley: Hoboken, NJ, **2010**.
32. Coutinho, F. M. B.; Carvalho, D. L.; La Torre Aponte, M. L.; Barbosa, C. C. R. Pellicular ion exchange resins based on divinylbenzene and 2-vinylpyridine. *Polymer* **2001**, *42*, 43-48.
33. Pirkle, W. H.; Pochapsky, T. C. Consideration of chiral recognition relevant to the liquid chromatographic separations of enantiomers. *Chem. Rev.* **1989**, *89*, 347-362.
34. Janson, J.-C.; Jönsson, J. Å. Introduction to chromatography. In *Protein purification: principles, high resolution methods, and applications*, 3rd ed.; Janson, J.-C., Ed.; John Wiley & Sons: Hoboken, NJ, **2011**; pp 25-50.
35. Wu, C.; Siems, W. F.; Klasmeier, J.; Hill, H. H. Separation of isomeric peptides using electrospray ionization/high-resolution ion mobility spectrometry. *Anal. Chem.* **2000**, *72*, 391-395.
36. Shelimov, K. B.; Clemmer, D. E.; Hudgins, R. R.; Jarrold M. F. Protein structure *in vacuo*: gas-phase conformation of BPTI and cytochrome *c*. *J. Am. Chem. Soc.* **1997**, *119*, 2240-2248.
37. Ruotolo, B. T.; Benesch, J. L. P.; Sandercock, A. M.; Hyung, S.-J.; Robinson, C. V. Ion mobility-mass spectrometry analysis of large protein complexes. *Nat. Protoc.* **2008**, *3*, 1139-1152.
38. Wu, C.; Siems, W. F.; Asbury, G. R.; Hill, H. H. Electrospray ionization high-resolution ion mobility spectrometry-mass spectrometry. *Anal. Chem.* **1998**, *70*, 4929-4938.
39. Henderson, S. C.; Valentine, S. J.; Counterman, A. E.; Clemmer, D. E. ESI/ion trap/ion mobility/time-of-flight mass spectrometry for rapid and sensitive analysis of biomolecular mixtures. *Anal. Chem.* **1999**, *71*, 291-301.

40. Mesleh, M. F.; Hunter, J. M.; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F. Structural information from ion mobility measurements: effects of the long-range potential. *J. Phys. Chem.* **1996**, *100*, 16082-16086.
41. Chen, Y.-L.; Collings, B. A.; Douglas, D. J. Collision cross sections of myoglobin and cytochrome c ions with Ne, Ar, and Kr. *J. Am. Soc. Mass Spectrom.* **1997**, *8*, 681-687.
42. Clemmer, D. E.; Jarrold, M. F. Ion mobility measurements and their applications to clusters and biomolecules. *J. Mass. Spectrom.* **1997**, *32*, 577-592.
43. Hudgins, R. R.; Woenckhaus, J.; Jarrold, M. F. High resolution ion mobility measurements for gas phase proteins: correlation between solution phase and gas phase conformations. *Int. J. Mass Spectrom.* **1997**, *165*, 497-507.
44. Clemmer, D. E.; Liu, Y. Characterizing oligosaccharides using injected-ion mobility/mass spectrometry. *Anal. Chem.* **1997**, *69*, 2504-2509.
45. Liu, Y.; Valentine, S. J.; Counterman, A. E.; Hoaglund, C. S.; Clemmer, D. E. Injected-ion mobility analysis of biomolecules. *Anal. Chem.* **1997**, *69*, 728A-735A.
46. Shelimov, K. B.; Jarrold, M. F. Conformations, unfolding, and refolding of apomyoglobin in vacuum: an activation barrier for gas-phase protein folding. *J. Am. Chem. Soc.* **1997**, *119*, 2987-2994.
47. Hoaglund, C. S.; Valentine, S. J.; Sporleder, C. R.; Reilly, J. P.; Clemmer, D. E. Three-dimensional ion mobility/TOFMS analysis of electrosprayed biomolecules. *Anal. Chem.* **1998**, *70*, 2236-2242.
48. *Transport properties of ions in gases*, Mason, E. A.; McDaniel, E. W., Eds.; Wiley: New York, NY, **1988**.
49. Shvartsburg, A. A.; Smith, R. D. Fundamentals of travelling wave ion mobility spectrometry. *Anal. Chem.* **2008**, *80*, 9689-9699.
50. *Practical organic mass spectrometry: a guide for chemical and biochemical analysis*, 2nd ed.; Chapman, J. R., Ed.; Wiley: New York, NY, **1993**.
51. Dempster, A. J. Positive ray analysis of lithium and magnesium. *Phys. Rev.* **1921**, *18*, 415.
52. Nier, A. O. A mass spectrometer for isotope and gas analysis. *Rev. Sci. Instrum.* **1947**, *18*, 398.
53. Gohlke, R. S. Time-of-flight mass spectrometry and gas-liquid partition chromatography. *Anal. Chem.* **1959**, *31*, 525-541.
54. *Interpretation of mass spectra*, 2nd ed.; McLafferty, F. W., Ed.; University Science Books: Mill Valley, CA, **1980**.
55. Peterkop, R. K. *Theory of ionization of atoms by electron impact*, Translation by Aronson, E.; Hummer, D. G., Ed.; Colorado Associated University Press: Boulder, CO, **1977**.
56. Taylor, G. Disintegration of water drops in an electric field. *P. R. Soc. London* **1964**, *280*, 383-397.
57. Wilm, M.; Mann, M.; Analytical properties of the nanoelectrospray ion source. *Anal. Chem.* **1996**, *68*, 1-8.
58. Beavis, R. C.; Chaudhary, T.; Chait, B. T. α-Cyano-4-hydroxycinnamic acid as a matrix for matrix assisted laser desorption mass spectrometry. *Org. Mass Spectrom.* **1992**, *27*, 156-158.

59. Kampmeier, J.; Dreisewerd, K.; Schürenberg, M.; Strupat, K. Investigations of 2,5-DHB and succinic acid as matrices for IR and UV MALDI. Part: I UV and IR laser ablation in the MALDI process. *Int. J. Mass Spectrom.* **1997**, *169*, 31-41.
60. Karas, M.; Glückmann, M.; Schäfer, J. Ionization in matrix-assisted laser desorption/ionization: singly charged molecular ions are the lucky survivors. *J. Mass Spectrom.* **2000**, *35*, 1-12.
61. Karas, M.; Krüger, R. Ion formation in MALDI: the cluster ionization mechanism. *Chem. Rev.* **2003**, *103*, 427-439.
62. Glückmann, M.; Pfenninger, A.; Krüger, R.; Thierolf, M.; Karas, M.; Horneffer, V.; Hillenkamp, F.; Strupat, K. Mechanisms in MALDI analysis: surface interaction or incorporation of analytes? *Int. J. Mass Spectrom.* **2001**, *210*, 121-132.
63. Knochenmuss, R. Ion formation mechanisms in UV-MALDI. *Analyst*, **2006**, *131*, 966-986.
64. Weston, D. J. Ambient ionization mass spectrometry: current understanding of mechanistic theory; analytical performance and application areas. *Analyst*, **2010**, *135*, 661-668.
65. Cooks, R. G.; Ouyang, Z.; Takats, Z.; Wiseman, J. M. Ambient mass spectrometry. *Science*, **2006**, *311*, 1566-1570.
66. Wiseman, J. M.; Ifa, D. R.; Zhu, Y.; Kissinger, C. B.; Manicke, N. E.; Kissinger, P. T.; Cooks, R. G. Desorption electrospray ionization mass spectrometry: imaging drugs and metabolites in tissues. *Proc. Natl. Acad. Sci.* **2008**, *105*, 18120-18125.
67. Esquenazi, E.; Yang, Y.-L.; Watrous, J.; Gerwick, W. H.; Dorrestein, P. C. Imaging mass spectrometry of natural products. *Nat. Prod. Rep.* **2009**, *26*, 1521-1534.
68. Zimmerman, T. A.; Monroe, E. B.; Tucker, K. R.; Rubakhin, S. S.; Sweedler, J. V. Imaging of cells and tissues with mass spectrometry: adding chemical information to imaging. *Method Cell Biol.* **2008**, *89*, 361-390.
69. Vickerman, J. C. Molecular imaging and depth profiling by mass spectrometry-SIMS, MALDI or DESI? *Analyst*, **2011**, *136*, 2199-2217.
70. Wiseman, J. M.; Ifa, D. R.; Song, Q.; Cooks, R. G. Tissue imaging at atmospheric pressure using desorption electrospray ionization (DESI) mass spectrometry. *Angew. Chem. Int. Ed.* **2006**, *45*, 7188-7192.
71. Shrestha, B.; Nemes, P.; Nazarian, J.; Hathout, Y.; Hoffman, E. P.; Vertes, A. Direct analysis of lipids and small metabolites in mouse brain tissue by AP IR-MALDI and reactive LAESI mass spectrometry. *Analyst.* **2010**, *135*, 751-758.
72. Nemes, P.; Vertes, A. Laser ablation electrospray ionization for atmospheric pressure, in vivo, and imaging mass spectrometry. *Anal. Chem.* **2007**, *79*, 8098-8106.
73. Shiea, J.; Huang, M.-Z.; Hsu, H.-J.; Lee, C.-Y.; Yuan, C.-H.; Beech, I.; Sunner, J. Electrospray-assisted laser desorption/ionization mass spectrometry for direct ambient analysis of solids. *Rapid Commun. Mass Spectrom.* **2005**, *19*, 3701-3704.
74. Vaikkinen, A.; Shrestha, B.; Kauppila, T. J.; Vertes, A.; Kostiainen, R. Infrared laser ablation atmospheric pressure photoionization mass spectrometry. *Anal. Chem.* **2012**, 84, 1630-1636.
75. Claassen, M. Inference and validation of protein identifications. *Mol. Cell Proteomics* **2012**, *11*, 1097-1104.
76. Roepstorff, P. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass Spectrom.* **1984**, *11*, 601.

77. Dančík, V.; Addona, T. A.; Clauser, K. R.; Vath, J. E.; Pevzner, P. A. *De novo* peptide sequencing via tandem mass spectrometry. *J. Comput. Biol.* **1999**, *6*, 327-342.
78. Dahiyat, B. I.; Mayo, S. L. De novo protein design: fully automated sequence selection. *Science* **1997**, *278*, 82-87.
79. Shevchenko, A.; Chernushevich, I.; Ens, W.; Standing, K. G.; Thomson, B.; Wilm, M.; Mann, M. Rapid '*de novo*' peptide sequencing by a combination of nanoelectrospary, isotopic labelling and a quadrupole/time-of-flight mass spectrometer. *Rapid Commun. Mass Sp.* **1997**, *11*, 1015-1024.
80. Kelleher, N. L.; Lin, H. Y.; Valaskovic, G. A.; Aaserud, D. J.; Fridriksson, E. K.; McLafferty, F. W. Top down versus bottom up protein characterization by tandem high-resolution mass spectrometery. *J. Am. Chem. Soc.* **1999**, *121*, 806-812.
81. Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci.* **2004**, *101*, 9528-9533.
82. Zubarev, R. A.; Horn, D. M.; Fridriksson, E. K.; Kelleher, N. L.; Kruger, N. A.; Lewis, M. A.; Carpenter, B. K.; McLafferty, F. W. Electron capture dissociation for structural characterization of multiply charged protein cations. *Anal. Chem.* **2000**, *72*, 563-573.
83. Zubarev, R. A.; Kruger, N. A.; Fridriksson, E. K.; Lewis, M. A.; Horn, D. M.; Carpenter, B. K.; McLafferty, F. W. Electron capture dissociation and gaseious multiply-charged proteins is favored at disulfide bonds and other sites of high hydrogen atom affinity. *J. Am. Chem. Soc.* **1999**, *121*, 2857-2862.
84. Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. Electron capture dissociation of multiply charged protein cations. A nonergodic process. *J. Am. Chem. Soc.* **1998**, *120*, 3265-3266.
85. Villas-Bôas, S. G.; Mas, S.; Åkesson, M.; Smedsgaard, J.; Nielsen, J. Mass spectrometry in metabolome analysis. *Mass Spectrom. Rev.* **2005**, *24*, 613-646.
86. Asharoni, A.; Ric de Vos, C. H.; Verhoeven, H. A.; Maliepaard, C. A.; Kruppa, G.; Bino, R. Goodenowe, D. B. Nontargeted metabolome analysis by use of Fourier transform ion cyclotron mass spectrometry. *Omics* **2002**, *6*, 217-234.
87. Ric de Vos C. H.; Moco, S.; Lommen, A.; Keurentjes, J. J. B.; Bino, R. J.; Hall R. D. Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nat. Protoc.* **2007**, *2*, 778-791.
88. Griffiths, P. R.; de Haseth, J. A. *Fourier transform infrared spectroscopy*, Elving, P. J.; Winefordner, J. D.; Kolthoff, I. M., Eds.; John Wiley & Sons: New York, NY, **1986**.
89. Mirabella, F. M. Introduction. *In Modern techniques in applied molecular spectroscopy*, Mirabella, F. M., Ed.; John Wiley & Sons: New York, **1998**, pp 1-10.
90. Bell, A. G. On the production and reproduction of sound by light. *Am. J. Sci.* (Series 3) **1880**, *20*, 305-324.
91. Bell, A. G. Upon the production of sound by radiant energy. *Phil. Mag.* (Series 5) **1881**, *11*, 510–528.
92. Rosencwaig, A.; Gersho, A. Theory of the photoacoustic effect with solids. *J. Appl. Phys.* **1976,** *47*, 64-69.
93. McClelland, J. F.; Jones, R. W.; Luo, S.; Seaverson, L. M. A Practical Guide to FT-IR Photoacoustic Spectroscopy. In *Practical Sampling Techniques for Infrared Analysis*, Coleman, P. B., Ed.; CRC Press: Boca Raton, FL, **1993**; pp 107-144.

94. McClelland, J. F.; Bajic, S. J.; Jones, R. W.; Seaverson, L. M. Photoacoustic spectroscopy. In *Modern Techniques in Applied Molecular Spectroscopy,* Mirabella, F. M., Ed.; Wiley-Interscience: New York, NY, 1**998**; pp 221-265.
95. *Photoacoustic imaging and spectroscopy,* Wang, L. V., Ed.; CRC Press: Boca Raton, FL, **2009**.
96. Cox, B.; Laufer, J. G.; Arridge, S. R.; Beard, P. C. Quantitative spectroscopic photoacoustic imaging: A Review. *J. Biomed. Opt.* **2012**, *17*, No. 061202.
97. Hu, S.; Wang, L. V. Optical-resolution photoacoustic microscopy: auscultation of biological systems at the cellular level. *Biophys. J.* **2013**, *105*, 841-847.
98. Yang, H.; Irudayaraj, J. Characterization of beef and pork using Fourier-transform infrared photoacoustic spectroscopy. *Lebensm.-Wiss. Technol.* **2001**, *34*, 402-409.
99. Irudayaraj, J.; Yang, H. Analysis of cheese using step-scan Fourier transform infrared photoacoustic spectroscopy. *Appl. Spectrosc.* **2000**, *54*, 595-600.
100. Gordillo-Delgado, F.; Marín, E.; Cortés-Hernández, D. M.; Mejía-Morales, C.; García-Salcedo, A. J. Discrimination of organic coffee via Fourier transform infrared-photoacoustic spectroscopy. *J. Sci. Food Agric.* **2012**, *92*, 2316-2319.
101. Belton, P. S.; Saffa, A. M.; Wilson, R. H. The potential of Fourier transform infrared spectroscopy for the analysis of confectionery products. *Food Chem.* **1988**, *28*, 53-61.
102. Letzelter, N. S.; Wilson, R. H.; Jones, A. D.; Sinnaeve, G. Quantitative determination of the composition of individual pea seeds by Fourier transform infrared photoacoustic spectroscopy. *J. Sci. Food Agric.* **1995**, *67*, 239-245.
103. Soyk, M. W. Instrumentation development for coupling ion/ion reactions and ion mobility in biological mass spectrometry. PhD Dissertation. Iowa State University: Ames, IA, **2008**.
104. Olsen, J. V.; de Godoy, L. M. F.; Li G.; Macek, B.; Mortensen, P.; Pesch, R.; Makarov, A.; Lange, O.; Horning, S.; Mann, M. Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell Proteomics* **2005**, *4*, 2010-2021.
105. Giles, K.; Pringle, S. D.; Worthington, K. R.; Little, D.; Wildgoose, J. L.; Bateman, R. H. Applications of a travelling wave-based radio-frequency-only stacked ring ion guide. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 2401-2414.

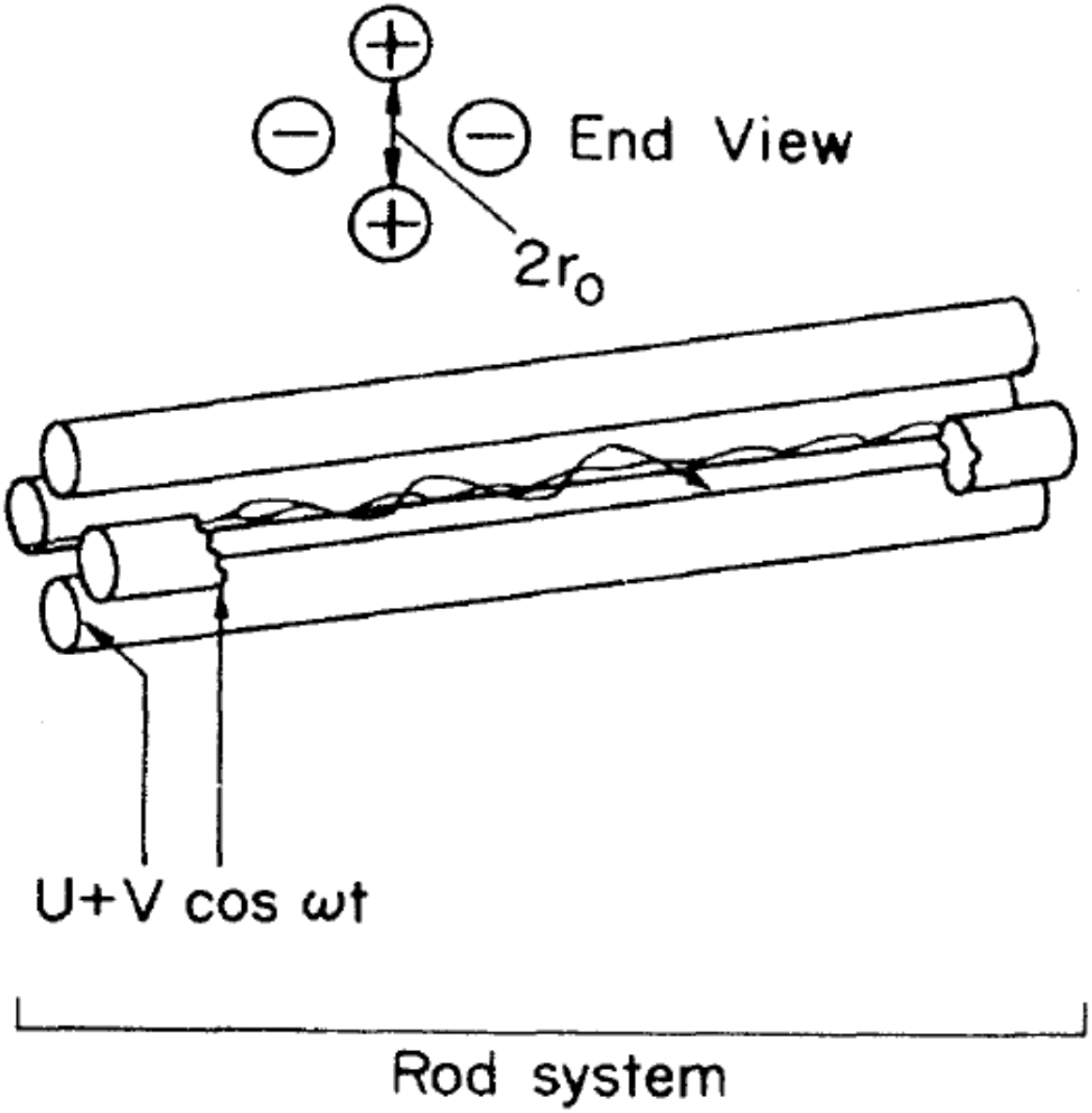


Figure 1-1 Side and end view of quadrupole rod system (7).

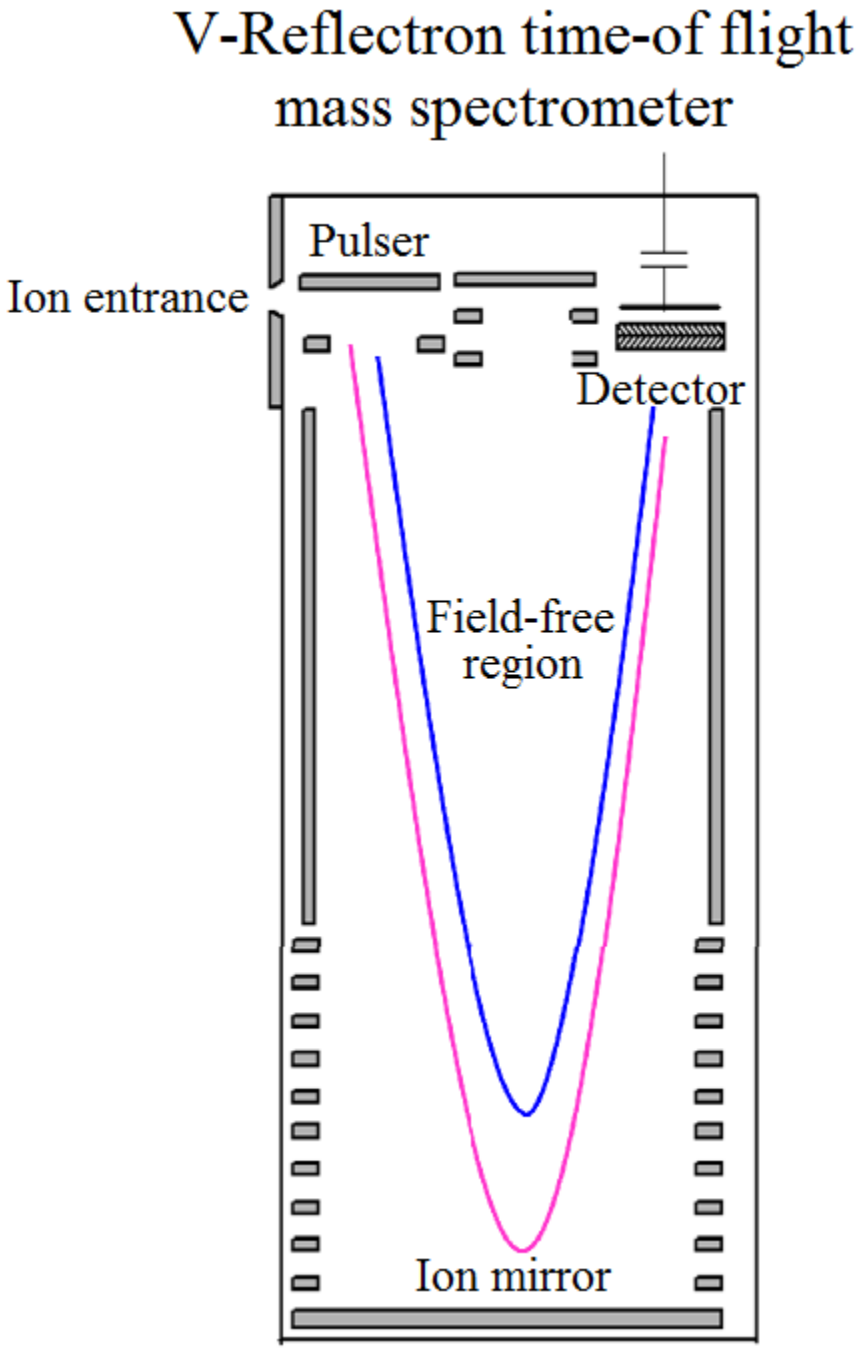


Figure 1-2 Time of flight reflectron, the blue and pink lines detail the flight path of two ions with different kinetic energy profiles traveling from the pulser to the detector (103).

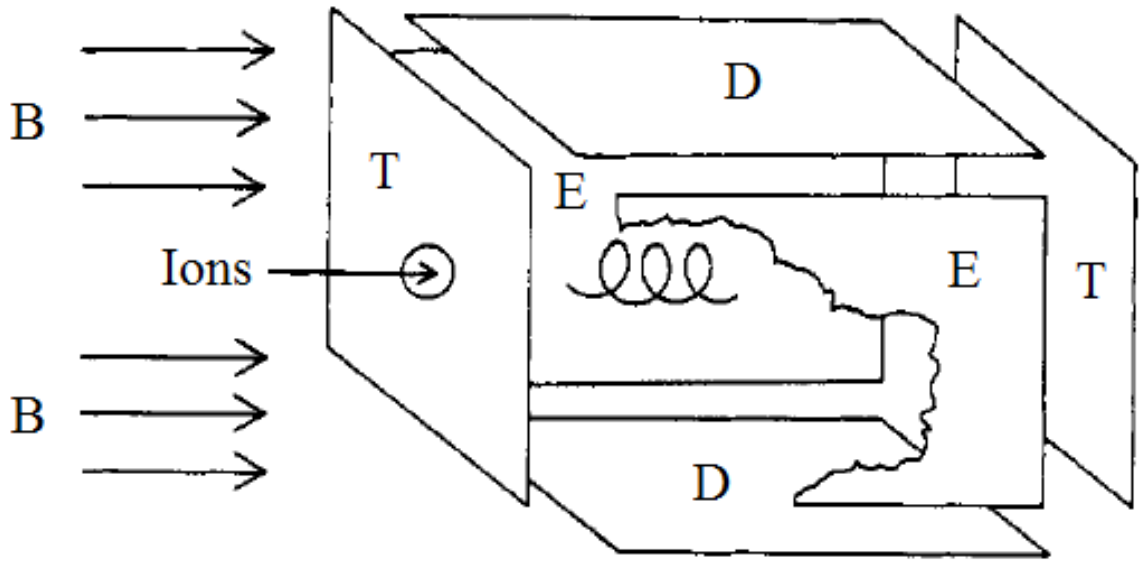


Figure 1-3 Diagram of FT-ICR cell. The first trapping plate (T) contains a hole for the ions to enter the cell. The excitation plates (E) modulate the radii of the trapped ions. The detection plates (D) measure the image-current of the ions as they oscillate in within the trap. The whole time a magnetic field (B) is applied axially to the trapped ion flight path (17).

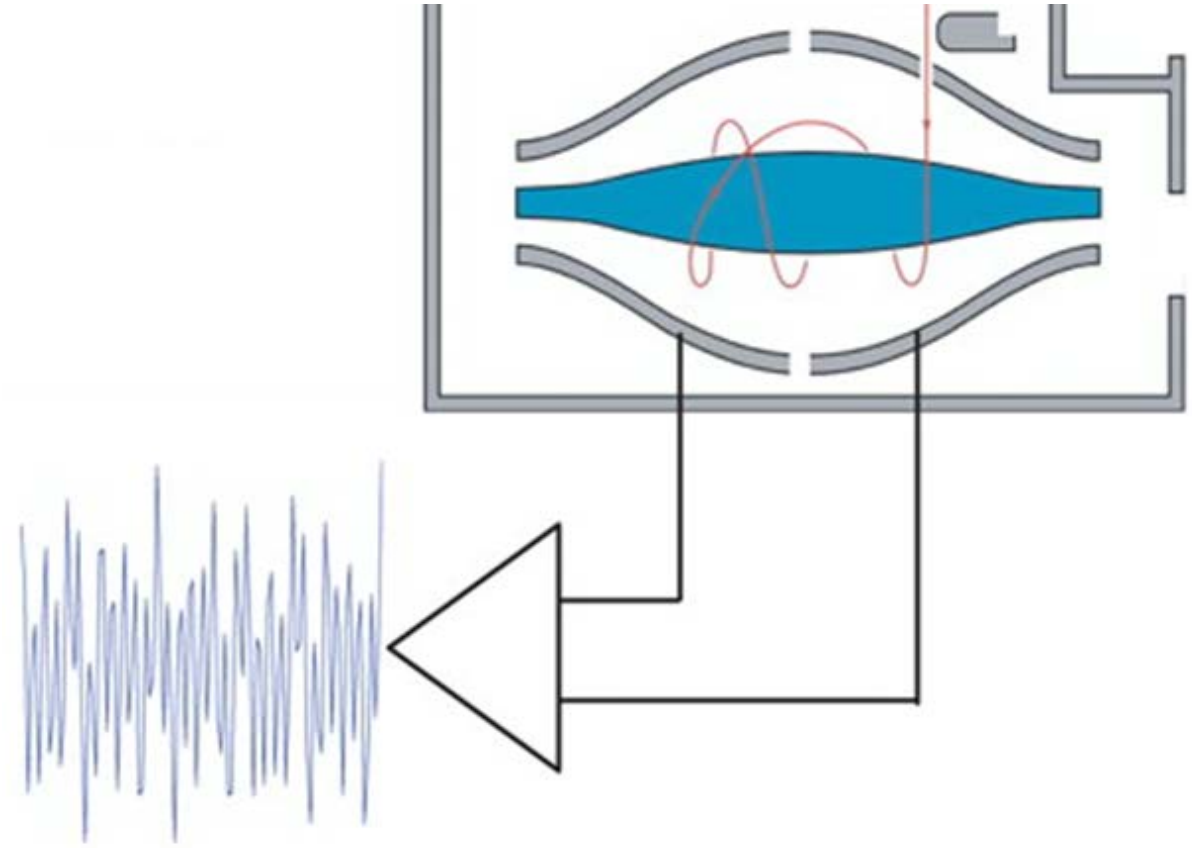


Figure 1-4 Schematic of Orbitrap detector detailing the entrance of the ions into the detector, orbit about the center electrode, and subsequent image-current frequency detected by the symmetric exterior electrodes (104).

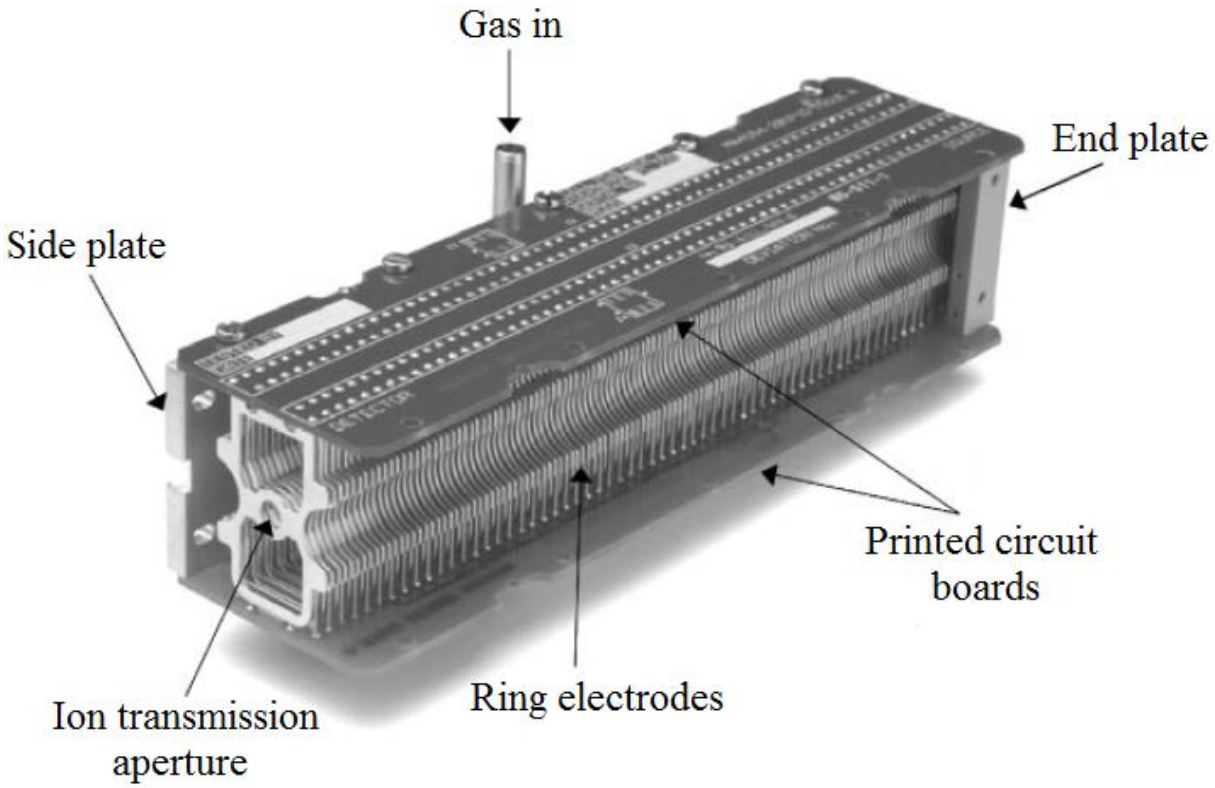


Figure 1-5 Photograph of the travelling wave ion mobility device (105).

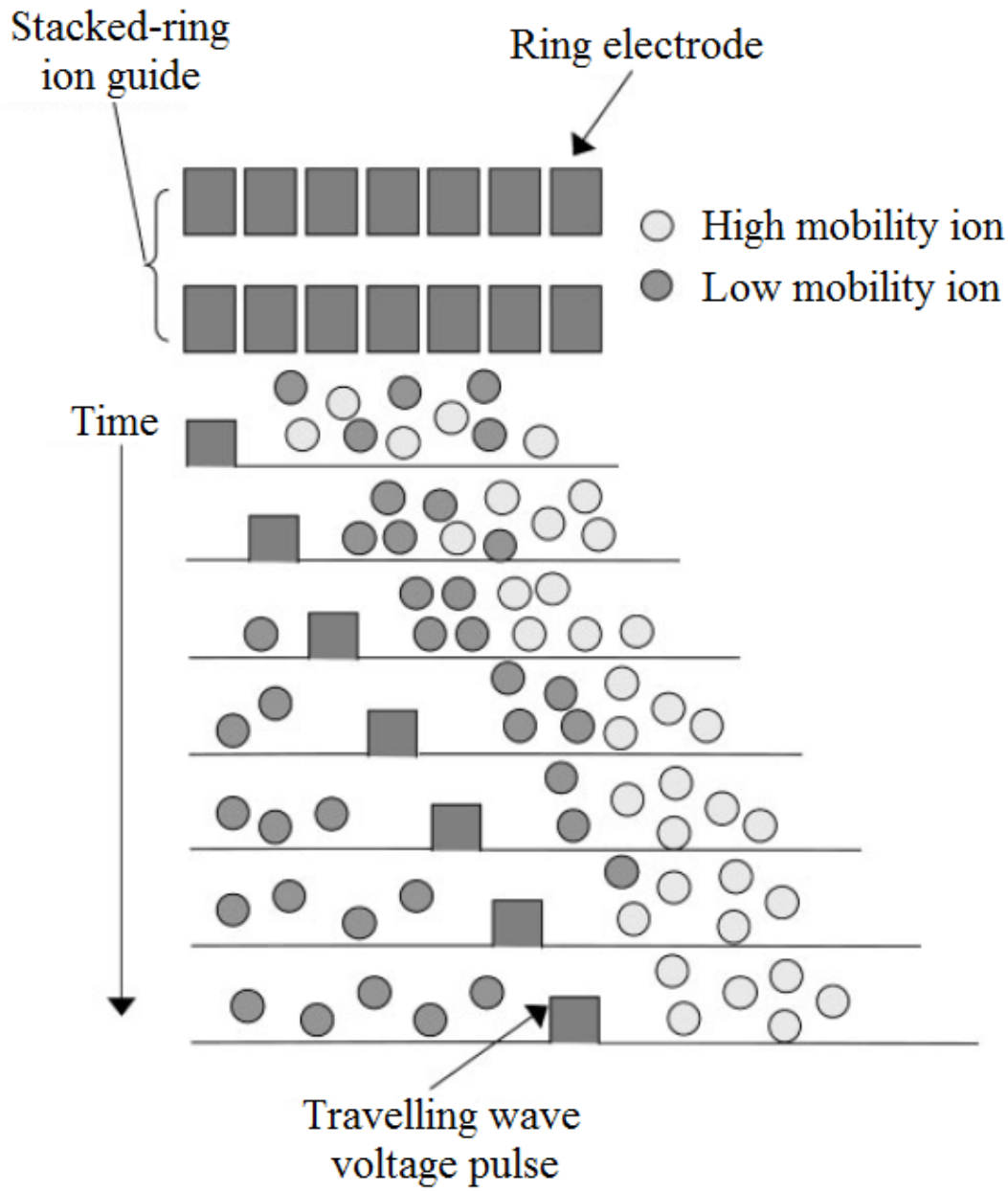
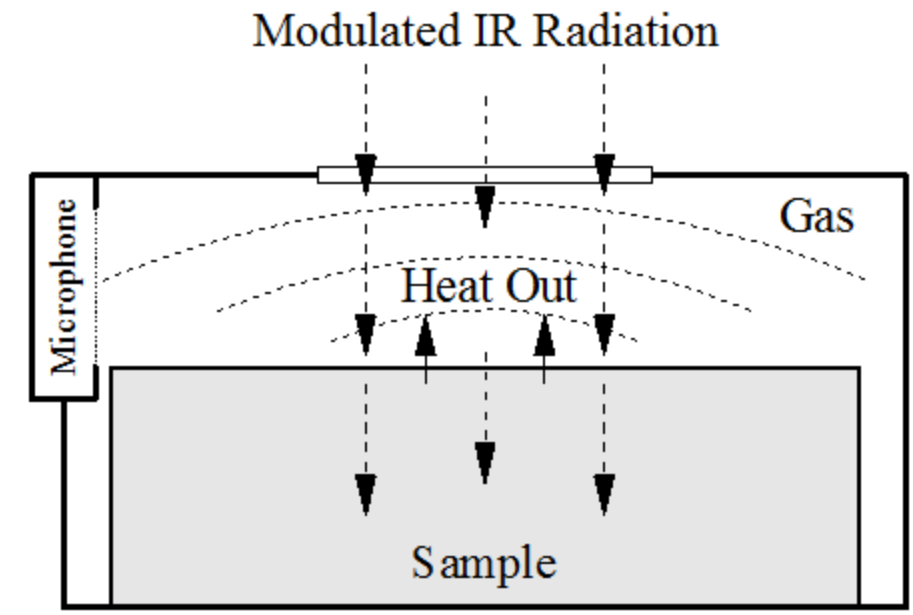


Figure 1-6 Diagram of the separation performed within a travelling wave ion mobility cell (105).



A

B

Figure 1-7 Diagram A shows how the PAS cell functions in the presence of modulated IR radiation (not to scale). Picture B is an exterior view of a functioning PAS cell nestled within a FTIR instrument. Diagram and photograph courtesy of the McClelland group.

CHAPTER 2

HIGH RESOLUTION TIME-OF-FLIGHT MASS SPECTROMETRY FINGERPRINTING OF METABOLITES FROM CECUM AND DISTAL COLON CONTENTS OF RATS FED RESISTANT STARCH

Modified from a paper published in *Analytical and Bioanalytical Chemistry*[1]

By

Timothy J. Anderson,[2,3*] Roger W. Jones,[2] Yongfeng Ai,[4] Robert S. Houk,[2,3] Jay-lin Jane,[4]

Yinsheng Zhao,[4] Diane F. Birt,[4] John F. McClelland[2,5]

[1]Reprinted with permission of *Analytical and Bioanalytical Chemistry 406(3):745-756*

[2]Ames Laboratory-USDOE, Iowa State University, Ames, Iowa 50011

[3]Department of Chemistry, Iowa State University, Ames, Iowa 50011

[4]Department of Food Science and Human Nutrition, Iowa State University, Ames, Iowa 50011

[5]Department of Mechanical Engineering, Iowa State University, Ames, Iowa 50011

*Corresponding Author, Email timma@iastate.edu, Phone 515-294-4285

Abstract

Time-of-flight mass spectrometry along with statistical analysis was utilized to study metabolic profiles among rats fed resistant starch (RS) diets. Fischer 344 rats were fed four starch diets consisting of 55% (w/w, dbs) starch. A control starch diet consisting of corn starch was compared against three RS diets. The RS diets were high-amylose corn starch (HA7), HA7 chemically modified with octenyl succinic anhydride, and stearic-acid-complexed HA7 starch. A subgroup received antibiotic treatment to determine if perturbations in the gut microbiome were long lasting. A second subgroup was treated with azoxymethane (AOM), a carcinogen. At the end of the eight week study, cecal and distal-colon contents samples were collected from the sacrificed rats. Metabolites were extracted from cecal and distal colon samples into acetonitrile. The extracts were then analyzed on an accurate-mass time-of-flight mass spectrometer to obtain their metabolic profile. The data were analyzed using partial least-squares discriminant analysis (PLS-DA). The PLS-DA analysis utilized a training set and verification set to classify samples within diet and treatment groups. PLS-DA could reliably differentiate the diet treatments for both cecal and distal colon samples. The PLS-DA analyses of the antibiotic and no antibiotic treated subgroups were well classified for cecal samples and modestly separated for distal-colon samples. PLS-DA analysis had limited success separating distal colon samples for rats given AOM from those not treated; the cecal samples from AOM had very poor classification. Mass spectrometry profiling coupled with PLS-DA can readily classify metabolite differences among rats given RS diets.

Keywords Resistant Starch, Mass Spectrometry, Metabolites, PLS-DA

Introduction

Food is composed of three main components, which are protein, lipid and starch. Cereal grains are one of the world's most important food sources; of these, its most abundant component is typically starch. Amylopectin and amylose make up the bulk of starch granules; their relative concentration can vary depending on source material. Amylopectin comprises typically about 70% of the starch granule but can reach 100% for waxy corn varieties. Amylose levels are typically about 15 to 30% of the starch granule (1-3).

Both amylopectin and amylose are comprised of polymeric glucose. Recent studies have shown that amylose forms double helical structures, and the hydroxyl groups of the individual glucose units may be esterified with a variety of lipid-like compounds (3). Amylose can also form single helical complexes with free fatty acids and iodine. These modified amylose structures along with food materials that contain up to 85% amylose are examples of resistant starch (RS) (3), which effectively resist digestion in the stomach and small intestine, and thus are processed farther along the gut (4).

RS has been studied for a variety of health benefits. RS may aid diabetics, decrease energy density, and also act as a prebiotic (5-10). Prebiotics are materials which resist digestion by the host and are metabolized in the gut by microorganisms (10). The microorganisms release compounds, such as short chain fatty acids, which are believed to be beneficial to the host organism (11, 12). The interplay between microorganisms and host is complex; the microorganisms may even interact with host diet to increase deleterious conditions such as intestinal inflammation (13). Some materials, such as oligofructose, lactulose, and transgalacto-oligosaccharides are well characterized and known to be prebiotics, but only recently has RS been classified as a prebiotic (14-20).

The catabolism of prebiotics by microorganisms produces a vast array of chemical components, which are classified as metabolites (21). Metabolites from gut microorganisms can be obtained from blood, urine or feces (22-24). Blood and urine reflect metabolic changes that occur in the gut from direct absorption through the intestinal lining (25, 26). Feces are complex media from which a direct interplay between microorganism and host can be observed (27).

There are five RS categories. Whole grains and ground legumes contain type 1 RS. Type 1 RS is amylose protected by cell wall material and other plant materials that are difficult to digest. Type 2 RS retains a crystalline structure, which resists enzyme hydrolysis, and is found primarily in banana and potato starch. Type 3 RS is retrograded starch; this occurs when the starch granule has gelatinized and reformed a crystalline structure (28). Type 4 RS is a chemically modified starch molecule. To produce the type 4 RS, a lipid compound such as octenyl succinic anhydride is generally bound to the various hydroxyl groups of glucose within the starch amylose molecule (14, 29). Type 5 RS is formed from a physical interaction of debranched helical starch amylose and a long chain fatty acid, such as palmitic or stearic acid (30).

Mass spectrometry has emerged as one of the most important platforms for metabolite analysis due to improvements in ionization methodology and high resolution instruments. A variety of metabolite separation methods can be coupled with mass spectrometers. The coupling allows comprehensive analysis through fingerprinting or profiling of metabolites (31, 32). Gas chromatography-mass spectrometry (GC-MS) is useful for the analysis of thermally stable volatile compounds. GC-MS often utilizes electron impact (EI) ionization. The GC separation and subsequent EI fragmentation often allow metabolite identification through database searching (27, 33-35). Liquid chromatography-mass spectrometry (LC-MS) is very useful for

metabolite analysis because specialized columns can be used to separate either polar or non-polar compounds (25, 26, 36). LC-MS can also be coupled with collision induced dissociation (CID) that provides fragment ions for potential metabolite identification (28, 37). Capillary electrophoresis (CE) is another separation technique that can be used prior to mass spectrometry analysis (38, 39). Work performed with CE and LC-MS has attempted to identify metabolites using standard retention time libraries and elemental formulas from accurate mass measurements (40). Ultra-high resolution mass spectrometry using Fourier transform ion cyclotron resonance (FT-ICR) instruments has been used for metabolite analysis (41, 42). Even matrix assisted laser desorption ionization (MALDI) imaging has recently been coupled with mass spectrometry to obtain metabolite insight (43).

Statistical tools have become important in the post examination of metabolites analyzed using mass spectrometry. Tools such as principal component analysis (PCA), soft independent modeling of class analogy (SIMCA), partial least-squares discriminant analysis (PLS-DA), linear discriminant analysis (LDA), and artificial neural networks all allow differentiation and group prediction (26, 27, 34, 35, 39, 44-46). Using these tools, difficult and complex analyses of large amounts of data can be more easily graphed and visualized.

The current study describes such methods for profiling the metabolites from rat cecal and distal colon contents to deduce differences among RS diets. Three different RS and a control corn starch were fed to rats. The first RS was a type 2 RS that consisted of high amylose corn starch (HA7). The second RS was a type 4 RS that was HA7 corn starch modified with octenyl succinic anhydride (OS-HA7). The last RS was a type 5 RS that was produced by complexing HA7 with stearic acid (StA-HA7). The main goal of this study was to determine whether statistical tools, such as PLS-DA, could observe differences among RS starch diets from rat cecal

and distal colon contents using mass spectrometry. Another goal was to understand whether statistical tools could accurately group diets and target their individual biomarkers that vary the most among the diets.

Materials and Methods

Animal Study

Male Fischer 344 rats were housed and fed according to the procedure of Zhao et al. (47). The study initially used 90 rats (two died before sacrifice) that were obtained at five weeks old, and were placed on specific feeding regimens for eight weeks, as shown in the flow diagram in Figure 2-1. The animals were randomly separated into four starch diet groups: Control, HA7, OS-HA7, and StA-HA7. The control and StA-HA7 starch diet groups each had 29 rats, and both diets contained four further subgroup treatments. Rats fed HA7 and OS-HA7 starch diets had 15 rats each, and both diets contained only two further subgroup treatments. All four starch diet groups were divided into two treatment subgroups, which consisted of 9 to 10 rats given either two injections of the carcinogen azoxymethane (AOM, Midwest Research Institute, Kansas City, MO) at a dose of 20 mg AOM/kg rat body weight and 4 to 5 rats given two injections of saline administered eight weeks prior to sample collection, following the method of Zhao et al. (47). In the control and StA-HA7 diet groups, both of the injection subgroups were further split into two more subgroups consisting of rats either administered an antibiotic consisting of a combination of imipenem and vancomycin at a dose of 50 µg/mL each in the drinking water or not administered any antibiotic in week 2 (Figure 2-1).

Starch Diet Materials

Normal corn starch (NCS, Cargill Gel ™) and HA7 (AmyloGel ™) were purchased from Cargill Inc. (Minneapolis, MN), 2-Octen-1-ylsuccinic anhydride (OSA), pullulanase from *Bacillus acidopullulyticus* and stearic acid (SA) were purchased from Sigma-Aldrich Co. (St. Louis, MO). All starch was cooked prior to addition to the rat diets, in accordance with the procedure of Zhao et al. (47). The non-starch component of the rat diet was prepared in compliance with the standards of the American Society for Nutritional Sciences for mature rats (AIN-93M) (48). The diets were prepared every other day and served fresh to the rats.

Preparation of OS-HA7 Starch Diet

OS-HA7 was produced from the HA7 through modification with OSA (49). The HA7 was suspended in water at a concentration of 35% (w/w, dsb). The pH of the starch slurry was adjusted to 8.0 by adding a sodium hydroxide aqueous solution (3%, w/w), and the temperature was maintained at 35 °C. OSA (10%, w/w, dsb) was added to the slurry dropwise while maintaining the pH at 8.0 and 35 °C. After the reaction was completed, the pH of the starch slurry became stable and was then adjusted to 6.5 by adding 1.0 M hydrochloric acid. The starch was recovered using filtration, washed with distilled water and 100% ethanol, dried at 37 °C, and then ground to fine powder.

Preparation of StA-HA7 Starch Diet

StA-HA7 was prepared from the HA7 following the method of Hasjim et al. (30) with modifications. The HA7 was suspended in water at a concentration of 10% (w/w, dsb). The starch slurry was pre-heated to 80 °C, and the temperature maintained for 1 h with stirring.

Pullulanase (5 units/g starch, dsb) was added to the slurry to hydrolyze α 1-6 branch linkages of the starch at 60 °C for 24 h. Stearic acid (SA, 10%, w/w, dsb) was added, and the mixture was kept at 80 °C for 1 h with vigorous stirring for amylose-SA complex formation. After cooling to room temperature, the StA-HA7 was recovered using centrifugation, washed with 50% (v/v) aqueous ethanol solution, dried at 37 °C, and then ground.

Rat Cecal Samples

The rat cecum was removed just after sacrifice. The cecum contents were placed in 15-mL centrifuge tubes (Corning, Tewksbury, MA) and stored on dry ice until placed in long term storage at -80 °C. Of the 88 rats sampled, 81 cecal samples were analyzed. For control, HA7, OS-HA7, and StA-HA7 diets, 24, 15, 14, and 28 samples were analyzed, respectively. The lower number of samples analyzed for the control diet was due to reduced sample volume. All RS fed rats had ceca of substantial size; on average they were nearly four times the mass of the control.

Rat Distal Colon Samples

After sacrifice, the rat colon was opened. Fecal pellets were collected from the large intestine starting from the anus to 5 cm up the colon and were placed in 15-mL centrifuge tubes containing phosphate buffer. The samples were placed on dry ice, and then stored long term at -80 °C. Seventy-two distal colon samples were analyzed from the original 88 rats. From control, HA7, OS-HA7, and StA-HA7 diets, 22, 13, 11, and 26 samples were analyzed, respectively. The drop in samples analyzed in respect to cecum samples was due to random amounts of fecal

pellets in the colon. Some rats contained many pellets, while some had none at the time of sacrifice.

Metabolite Extraction for Cecal and Distal Colon Samples

The metabolite extraction was an adaption of the method of Antunes et al. (42). Frozen cecal and distal colon samples were thawed and approximately 150 mg of sample was placed in a 2.0-mL microcentrifuge tube (Eppendorf, Hamburg, Germany). The tube was then half-filled with 1 mm zirconia/silica beads (Biospec Products, Bartlesville, OK) and 1 mL of HPLC grade acetonitrile (Fischer Scientific, Fair Lawn, NJ) was introduced. The microcentrifuge tubes were placed on a vortex apparatus (Fisherbrand, Bohemia, NY) with an orbital bead homogenizing adaptor (Mo Bio Laboratories Inc., Carlsbad, CA). The samples were vortexed until homogenized. The liquid portion of each sample was then pipetted into a clean microcentrifuge tube. The samples were placed in a centrifuge (Model 5415C, Eppendorf, Westbury, NY) for 5 minutes at 12,000 rpm. From the centrifuge tube, 500 µL of supernatant was placed in a clean centrifuge tube, then vacuum centrifuged (Labconco Corporation, Kansas City, MO) at 50 °C and depressurized with a dry vacuum pump (Welch Rietchle Thomas, Skokie, IL). The dried metabolite extracts were stored at -20 °C.

The dried metabolite extracts were reconstituted for mass spectrometry analysis by adding 500 µL of a 60% (v/v) aqueous acetonitrile solution. The extract was vigorously vortexed and then centrifuged for 2 minutes at 12,000 rpm. 100 µL of supernatant was removed and placed in a clean microcentrifuge tube. An additional 500 µL of 60% (v/v) aqueous acetonitrile was added to dilute the sample. Formic acid was added to produce a 0.2% formic acid sample solution to aid in positive ion production by electrospray ionization (ESI). Lastly,

the solution was vortexed and transferred into 12x35 mm clear glass vials to be placed in an LC autosampler.

Mass Spectrometry

From the reconstituted cecal and distal colon metabolite extracts, 20 µL of sample was injected into an Agilent 1260 Infinity LC system (Agilent Technologies, Inc., Santa Clara, CA), which was used as an autosampler. The samples were direct-infused from the LC and ionized using ESI in positive mode on an Agilent 6224 time-of-flight mass spectrometer operated at an acquisition rate of 4 GHz. A sample run lasted 5 minutes with a mobile phase of 60% (v/v) aqueous acetonitrile. A second mobile phase of 60% (v/v) aqueous acetonitrile with 20 ppm of a 50/50 mixture of polyethylene glycol (PEG) 200 and 600 standards (Hampton Research, Aliso Viejo, CA) was used for calibration. At 48 seconds into a sample run, the PEG-containing mobile phase was used for 12 seconds. At 1 minute the primary mobile phase was switched inline and flushed the PEG calibrant from the system prior to the next sample.
The software used was Agilent MassHunter Qualitative Analysis. Each spectrum was background subtracted, and the m/z scale was recalibrated using 17 ammoniated PEG-adduct peaks ranging from m/z 168.1236 to 872.5430. The calibrated spectra were centroided and exported as text files. The text files for each diet and subgroup were averaged using custom software. The averaged text files were then uploaded to the MassTRIX website, and accurate mass data were compared to rat metabolites in the KEGG database (50).

Statistical Analysis

Separate PLS-DA classification models were developed using commercial software (Pirouette Version 4.5; Infometrix, Inc.; Bothell, WA) for diet, AOM treatment, and antibiotic treatment. The cecal and distal fecal datasets were modeled separately. The cecal and distal fecal datasets consisted of spectra from 81 and 72 rats, respectively. These sets included data from AOM-treated and antibiotic-treated rats, as well as untreated ones. The verification set consisted of 25% of the total number of samples and the remaining 75% of samples were placed in the training set. The verification set was comprised of as close to one quarter of diet, AOM, and antibiotic treated samples as possible to ensure equal weighting during classification. The appropriate number of rats from each treatment was randomly selected for the verification set without previous knowledge of the spectral results. Each spectrum was normalized so that the base peak had an intensity of 100. Small peaks with intensities below one were then eliminated. The data were aligned using a mass tolerance of 0.005 m/z so that peaks in different spectra within a range of m/z ~ 0.010 were assigned the same (average) mass. In PLS-DA, each training-set spectrum is assigned a class-fit value of one for any class to which its sample belongs and a class-fit value of zero for classes to which its sample does not belong. When the verification set is analyzed, the resulting model determines the class-fit values of each spectrum for every class. Each verification-set spectrum is then assigned to the class for which its class-fit value is closest to one, which usually means the one class for which its class-fit value is above 0.5.

Prior to modeling, the spectra were always mean centered and for some models one-component orthogonal signal correction was used (51). The resulting regression vectors for the classes (e.g., diets) were used to identify those peaks that most strongly differentiate the classes, and the corresponding species were examined as possible biomarkers.

Results and Discussion

Mass Spectrometry

Even without the mass spectrometry analysis, there were obvious differences in the digestive-tract contents of rats fed different diets. As described previously, the RS fed rats had ceca which were as much as four times greater in mass than the rats on the control diet. The increased mass may have been due to increased fermentation time in the cecum to obtain nutrients from resilient RS components. The color, consistency, and aroma of the cecal contents were much different for the control compared to the RS diets. These differences may be linked to the amount of the starch fermented, and thus tied to the gut microorganisms and metabolites. The primary difference among distal colon contents was that the various diets had distinct colors.

Mass spectra of metabolites from the cecal contents of rats that were fed with different diets and received no AOM or antibiotic were averaged and are compared in Figure 2-2. The main emphasis of the study was to determine differences among diets in respect to their metabolite profiles. A metabolite peak was considered specific to a diet group if the peak was approximately three times the intensity of the same peak in the other diets. The peaks are given letter labels in the figure, and Table 2-1 lists their m/z values and the diets in which they were found. First, the peaks that were specific to only one diet group were inspected. The control group had four specific peaks labeled *y*, *z*, *aa*, and *ai*. The HA7 group had no specific peaks. The OS-HA7 group contained six specific peaks marked *c*, *p*, *w*, *ad*, *ag*, and *aj*. The last group, StA-HA7, had four specific peaks labeled *t*, *u*, *ah*, and *ak*. Far fewer metabolite peaks were shared between any two of the groups. The *ab* peak was shared by the control and HA7 groups. Another peak, labeled *f*, was in both the control and StA-HA7 groups. The HA7 and StA-HA7 groups had five mutual peaks marked *e*, *i*, *j*, *n* and *r*. Fewer peaks were observed among three

common groups. The control, HA7, and StA-HA7 diets shared the *g* peak. The control, OS-HA7, and StA-HA7 diets contained a single peak in common, labeled *h*. The HA7, OS-HA7 and StA-HA7 diet groups shared the three peaks *b*, *k* and *s*. Only one prominent metabolite was observed in all groups; it was the *d* peak.

Mass spectra of the diet treatment metabolites in the distal colon contents are compared in Figure 2-3. Similarly to the cecal contents, the mass spectra for the distal colon contents were averaged from rats that received no AOM or antibiotic. The peaks were also classified as specific to diet groups if they were approximately three times the intensity of the other diet groups. Like the labeled cecal-contents peaks in Figure 2-2, further information about the lettered peaks in Fig 2-3 can be found in Table 2-1. No peaks were observed to be specific to the control group. The HA7 group had only one exclusive peak, labeled *af*. The OS-HA7 group had seven peaks—*m*, *p*, *q*, *w*, *ad*, *ae*, and *aj*. Lastly, StA-HA7 had the six peaks *k*, *l*, *o*, *t*, *ac*, and *ak*. Fewer metabolite peaks were shared between any two groups than the peaks specific to a single group. The control and StA-HA7 diet groups shared the four peaks *a*, *g*, *x* and *ah*. The HA7 and StA-HA7 diets contained five common peaks—*j*, *n*, *r*, *s*, and *v*. Only the metabolite *h* was shared among the control, OS-HA7, and StA-HA7 diet groups. The two peaks *b* and *d* were shared among all four diet groups.

Analysis of the mass spectra in Figure 2-1 and 2-2 highlighted modest differences between the cecal and distal colon contents. Spectra for the control and StA-HA7 diet group for both digestive locations have been placed in Figure 2-4 for comparison. Specific peaks across digestive locations for a diet group were chosen with the same rule as described above for Figure 2-2, and Figure 2-3, and the peaks are again identified in Table 2-1. The control group for the cecal contents contained peaks labeled *y*, *z*, *aa*, *ab*, and *ai*, which were more intense than the

same peaks in the distal colon contents. The distal colon contents contained the four peaks labeled *g*, *l*, *x*, and *ac*, which were more intense than those same peaks observed in the cecal contents. Both digestive locations shared the peaks *a*, *b*, *d*, *f*, *h*, and *ah* with similar intensities. The StA-HA7 group for both cecal and distal colon digestive locations was more complex than the control group. The cecal contents contained only one peak, labeled *ah*, that was more intense than in the distal colon contents. Peaks *g*, *l*, *o*, and *ac* had higher intensities in the distal colon contents than in the cecal contents. Peaks labeled *a*, *b*, *d*, *f*, *h*, *j*, *k*, *n*, *r*, *s*, *t*, *u* and *ak* were at similar intensities in the cecal and distal colon contents.

From this scrutiny of the mass spectra of the cecal contents and distal colon contents for the diets, it is apparent that RS diets altered the metabolism of the rats. In Table 2-1 the cecal contents from each diet had a distinct metabolite profile. Of the 27 labeled metabolites, 52% were specific to only one diet group. These differences among diets persisted into the distal colon contents, for which 54% of the 26 labeled metabolites in Table 2-1 were exclusive to individual diet groups. Figure 2-4 compared the cecal and the distal colon contents of the control diet, Table 2-1 described that 77% of the 13 labeled metabolites are specific to one or the other digestive location. Figure 2-4 also compared spectra of cecal and distal colon contents of the StA-HA7 diet, and 22 metabolites from Table 2-1 were specific to only 45% of either digestive location. For the control, there appears to be substantial change to the metabolite profile. However, for RS, such as StA-HA7, there was less variability in the metabolite profile from cecum to distal colon.

In a companion study to this one that used the same samples, Anderson et al. (52) determined through enzymatic assay that the average cecal starch content (dsb) for the control, HA7, OS-HA7, and StA-HA7 diets were 0.7%, 18.3%, 48.8%, and 21.9%, respectively. The

residual starch differences are most likely attributed to the ability of each starch to be fermented in the cecum. The control and OS-HA7, which were the most and least fermented diets, respectively, had more diet-specific metabolites for both cecal and distal contents. The HA7 and StA-HA7 RS diets had nearly the same degree of fermentation and also shared many metabolites in the cecal and distal colon contents. The HA7 and StA-HA7 similarities also could be due to the relationship between the diets; StA-HA7 was produced from a complex of HA7 and stearic acid, and StA-HA7 may share some of the fermentation products of its parent RS. Regardless, the degree of fermentation and the nature of the starch source are probably the largest factors regarding the divergence of the metabolite profiles.

Statistical Analysis of Cecal Contents

Classification analysis of the mass spectra for the RS diet groups was performed with PLS-DA for the cecal extracts. PLS-DA modeling was done for the cecal contents with a 61 sample training set, and a 20 sample verification set. The verification set contained 6, 4, 3, and 7 samples from the control, HA7, OS-HA7, and StA-HA7 diets, respectively. The PLS-DA analysis of cecal contents for diet correctly classified all training-set samples; all verification-set samples were classified properly. Figure 2-5 displays the PLS-DA results and the quality of separation for the cecal contents RS diet groups. The closed and open symbols correspond to verification-set samples that do and do not belong to the indicated diet, respectively. For the control and OS-HA7 diets, all samples were spaced well away from the 0.5 nominal threshold for class differentiation. StA-HA7 had nearly all closed symbols positioned well above the 0.5 line, but one sample appeared slightly above the nominal threshold. HA7 had three of its samples positioned above the 0.5 line, but one sample appeared below the threshold line.

Nevertheless, it is readily classified as HA7 because its class-fit values for all of the other diets were even lower.

PLS-DA fingerprint analysis for the cecal contents was also carried out for the antibiotic subgroup treatments. The cecal contents PLS-DA antibiotic analysis had no training-set misclassifications. In Figure 2-6 only two verification-set samples were misclassified. Thus, the PLS-DA statistical classification separated the cecal contents antibiotic and saline treatments well.

Classifying the cecal-content samples by AOM subgroup treatment was much less successful (data not shown). PLS-DA was able to produce a model for AOM treatment classification that properly predicted all of the training-set samples, but its analysis yielded 13 misclassifications among the 20 verification-set samples. Clearly, at the time the animals were sacrificed and the samples taken, eight weeks after the last AOM injection, the AOM treatment had too little effect on the cecal contents to be observed via mass spectrometry.

Statistical Analysis of Distal Colon Contents

Classification analysis was also performed for the spectra of the distal colon contents RS groups. PLS-DA was performed with a 54-sample training set and 18-member verification set. The verification set contained 6, 3, 3, and 6 samples from the control, HA7, OS-HA7, and StA-HA7 diets, respectively. The PLS-DA diet group analysis of the distal colon contents misclassified only one verification-set sample. Figure 2-7 shows the PLS-DA verification-set analysis of the distal colon contents. The distal-colon-contents categorization for the diet samples was not as clean as it was for the cecal contents. The HA7, OS-HA7, and StA-HA7 diets have all of their markers in the correct region above or below the 0.5 class-fit threshold, but

many markers occur very close to the threshold. The control diet classification clearly shows all of the errant samples. In the control diet classification, one control sample fell below the 0.5 class-fit threshold, but it still correctly classified due to its class-fit value being closest to one for the control group. The other errant sample in the control-diet classification is the open, inverted triangle (StA-HA7 sample) at a control-diet class-fit value of 0.76. This same sample corresponds to the highest filled triangle in the StA-HA7 classification (class-fit value 1.45). Because the class-fit value for the sample in the control diet class is closer to one, the sample was misclassified as belonging to the control diet.

PLS-DA analysis of the distal-colon-contents training set for antibiotic treatment had no misclassified samples. The PLS-DA verification set in Figure 2-6 contained four misclassified samples. The PLS-DA classification was able only to modestly separate distal colon antibiotic and saline subgroup treatments. It has been observed that antibiotics have a profound effect on the microbial gut community and metabolic processes in mice (42). A cultured gut microbial community can return to pre-antibiotic treatment levels approximately three weeks after treatment (53, 54). A metabolic study also determined that urinary and fecal metabolites of mice return to near control conditions three weeks after antibiotic treatment (55). However, a comprehensive pyrosequencing study of the gut microbiome found that antibiotic recovery did not return to the initial conditions, even after six weeks (56).

Our findings are consistent with long-term antibiotic perturbations to the gut microbiome. Eight weeks after the antibiotic treatment, the cecal contents antibiotic treatment was still well separated from the saline treatment, and distal colon antibiotic treatment was modestly different from the saline treatment (Figure 2-6).

The distal colon PLS-DA modeling of the training set for AOM treatment had no misclassifications. The verification set for the PLS-DA analysis had four misclassifications (data not shown). AOM is a carcinogen and promotes the growth of preneoplastic lesions in the colon. Previous work had suggested that a RS diet decreased the levels of preneoplastic lesions in rats given AOM (47) but preneoplastic lesions in the animals used for this study were not appreciably different among the four diet treatments (unpublished data). As described above, the cecum metabolite profile was not greatly affected by the AOM treatment. The PLS-DA distal colon analysis does appear to show that AOM may have had a minor effect on the metabolite profile in the distal colon. This could be attributed to the possibility that lesions would appear in the colon, rather than the cecum. These lesions thus would most likely not have a large effect upstream on the cecal metabolites.

Biomarkers from Resistant Starch Diets

An effort was made to locate a small group of distinctive metabolites from the cecal and distal colon contents diet treatments as indicators or biomarkers for the diets. The statistical analysis of the cecal and distal colon contents yielded a vector regression plot for each diet (data not shown). The regression vectors indicated how strongly each m/z peak contributed to the PLS-DA classification of the parent diet in relation to the three diets by assigning each m/z peak a relative intensity. Mass spectral peaks that contributed a substantial weight to the variation and classification of the diets gave a strong, positive vector intensity for that diet. The absolute normalized intensity of a m/z peak within a specific diet in relation to the other diets did not necessarily predict a large vector regression coefficient. In many cases, m/z peaks had analogous intensities among multiple groups, but was considered a biomarker for only a single group. That

biomarker had a high regression coefficient because the m/z peak was a key component in classifying that diet group from the others.

For both cecal and distal colon contents analysis, the five or six metabolites with the strongest positive regression-vector values were examined as potential biomarkers. The accurate m/z values of these potential biomarkers were compared against the KEGG database using MassTRIX (50). The accurate mass of the biomarkers matched within an accuracy of ±0.005 Da.

The potential biomarkers from each RS diet for cecal and distal colon contents are shown in Table 2-2. The biomarkers comprise a variety of compound classes ranging from amino acids and glucose to various steroids. Amino acids appear often among many of the diet categories; proline, leucine, isoleucine, valine, methionine, and phenylalanine are all present. In many cases the amino acid biomarkers that are prevalent in the cecal contents are not prevalent enough to be chosen as biomarkers for the distal colon contents. The control diet was the only diet to retain an amino acid biomarker, proline, between cecal and distal colon contents. By contrast, the control diet has leucine or isoleucine only in the cecum, while methionine appears only in the distal colon contents list. It is not apparent why the RS diets would have such variations in the prevalence of amino acid biomarkers, but the large differences between the cecum and the distal colon could be due to host absorption or differences in the bacterial community induced by the different diets. Table 2-2 also shows that many biomarkers are conserved in both the cecal and distal colon contents. Biomarkers with the m/z of 116.071, 148.134, 251.128, 321.240, 336.228, 357.239, and 595.352 were observed in the cecal and distal contents in their respective diet groups. The similarity in the biomarkers between the cecal and distal colon contents correspond well with Figure 2-4. The Figure 2-4 control and StA-HA7 diets contained many of the same

cecal and distal-colon contents m/z peaks. Thus, many of the most prominent biomarkers could transit the large intestine.

Conclusions

Metabolite extracts of rat cecal and distal colon contents from the starch diets could be accurately fingerprinted using PLS-DA. The PLS-DA verification-set plots showed that the diet groups for the cecal and distal colon contents could be distinguished from each other. Utilizing PLS-DA, the antibiotic subgroup treatments were divided well for the cecal contents and modestly separated for the distal colon contents. Thus, the digestive system of the rats had most likely regained normal function following antibiotic treatment, but had different microbial communities. The AOM and saline treated rats partially separated based on the distal colon contents, but failed to separate for the cecal contents. The AOM treatment may not have had an effect on the cecum due to AOM targeting the colon or the time frame of AOM induced carcinogenesis may not have been reached.

Future proposed work will involve studying changes in metabolites over time within the period immediately following administration of the diet. Compounds which appear to change over time will be compared to biomarkers in this study in an attempt to improve understanding of RS digestion. Also due to the variety of biomarker compounds, parallel companion studies will be run to observe polar or non-polar metabolites and to definitely identify observed compounds using MS/MS or GC/MS.

Acknowledgement

We would like to thank Dr. Gregory J. Phillips for the use of his lab to extract the fecal metabolites (Department of Veterinary Science and Medicine, Iowa State University). We would also like to give our appreciation to Herman S. Sahota for the use of custom software to average mass spectra (Department of Computer Science, Iowa State University). This project was supported by the Iowa State University Plant Sciences Institute and supported in part by the Department of Agriculture, CSREES award number 2009-65503-05798. This research was performed in part at the Ames Laboratory. Ames Laboratory is operated for the U.S. Department of Energy by Iowa State University under contract no. DE-AC02-07CH11358.

References

1. Sajilata, M. G.; Singhal, R. S.; Kulkarni, P. R. Resistant starch-a review. *Compr. Rev. Food Sci. Food Saf.* **2006**, *5*, 1-17.
2. Atkin, N. J.; Cheng, S. L.; Abeysekera, R. M.; Robards, A. W. Localisation of amylose and amylopectin in starch granules using enzyme-gold labelling. *Starch-Stärke* **1999**, *51*, 163-172.
3. Jane, J. Structural Features of Starch Granules II. In: *Starch*: *Chemistry and Technology*; 3rd Edn. BeMiller J. N.; Whistler, R. L., Eds. Academic Press Inc.: Orlando, **2009**; pp 193-227.
4. Champ, M.; Langkilde, A-M.; Brouns, F.; Kettlitz, B.; Le Bail-Collet, Y. Advances in dietary fibre characterisation. 2. Consumption, chemistry, physiology and measurement of resistant starch; implications for health and food labeling. *Nutr. Res. Rev.* **2003**, *16*, 143-161.
5. Robertson, M. D. Dietary-resistant starch and glucose metabolism. *Curr. Opin. Clin. Nutr. Metab. Care* **2012**, *15*, 362-367.
6. Saltiel, A. R.; Kahn, C. R. Insulin signalling and the regulation of glucose and lipid metabolism. *Nature* **2001**, *414*, 799-806.
7. Ranganathan, S.; Champ, M.; Pechard, C.; Blanchard, P.; Nguyen, M.; Colonna, P.; Krempf, M. Comparative study of the acute effects of resistant starch and dietary fibers on metabolic indexes in men. *Am. J. Clin. Nutr.* **1994**, *59*, 879-883.
8. Eckel, R. H.; Grundy, S. M.; Zimmet, P. Z. The metabolic syndrome. *Lancet* **2005**, *365*, 1415-1428.
9. Bodinham, C. L.; Frost, G. S.; Robertson, M. D. Acute ingestion of resistant starch reduces food intake in healthy adults. *Brit. J. Nutr.* **2010**, *103*, 917-922.
10. Gibson, G. R.; Probert, H. M.; Van Loo, J.; Rastall, R. A.; Roberfroid, M. B. Dietary modulation of the human colonic microbiota: updating the concept of prebiotics. *Nutr. Res. Rev.* **2004**, *17*, 259-275.
11. Jenkins, D. J. A.; Vuksan, V.; Kendall, C. W. C.; Würsch, P.; Jeffcoat, R.; Waring, S.; Mehling, C. C.; Vidgen, E.; Augustin, L. S. A.; Wong, E. Physiological effects of resistant starches on fecal bulk, short chain fatty acids, blood lipids and glycemic index. *J. Am. Coll. Nutr.* **1998**, *17*, 609-616.
12. Nilsson, A. C.; Östman, E. M.; Knudsen, K. E. B.; Holst, J. J.; Björck, I. M. E. A cereal-based evening meal rich in indigestible carbohydrates increases plasma butyrate the next morning. *J. Nutr.* **2010**, *140*, 1932-1936.
13. Ding, S.; Chi, M. M.; Scull, B. P.; Rigby, R.; Schwerbrock, N. M. J.; Magness, S.; Jobin, C.; Lund, P. K. High-fat diet: bacteria interactions promote intestinal inflammation which precedes and correlates with obesity and insulin resistance in mouse. *PLOS One* **2010**, *5*, e12191.
14. Brown, I. L.; Wang, X.; Topping, D. L.; Playne, M. J.; Conway, P. L. High amylose maize starch as a versatile prebiotic for use with probiotic bacteria. *Food Aust.* **1998**, *50*, 603-610.
15. Roberfroid, M. B. Functional foods: concepts and applications to inulin and oligofructose. *Brit. J. Nutr.* **2002**, *87*, S139-S143.
16. Tuohy, K. M.; Ziemer, C. J.; Klinder, A.; Knöbel, Y.; Pool-Zobel, B. L.; Gibson, G. R. A human volunteer study to determine the prebiotic effects of lactulose powder on human colonic microbiota. *Microb. Ecol. Health Dis.* **2002**, *14*, 165-173.

17. Ito, M.; Kimura, M.; Deguchi, Y.; Miyamori-Watabe, A.; Yajima, T.; Kan, T. Effects of transgalactosylated disaccharides on the human intestinal microflora and their metabolism. *J. Nutr. Sci. Vitaminol.* **1993**, *39*, 279-288.
18. Marotti, I.; Bregola, V.; Aloisio, I.; Di Gioia, D.; Bosi, S.; Di Silvestro, R.; Quinn, R.; Dinelli, G. Prebiotic effect of soluble fibres from modern and old durum-type wheat varieties on *Lactobacillus* and *Bifidobacterium* strains. *J. Sci. Food Agric.* **2012**, *92*, 2133-2140.
19. Kritchevsky, D. (1995) Epidemiology of fibre, resistant starch and colorectal cancer. *Eur. J. Cancer Prev.* **1995**, *4*, 345-352.
20. van Munster, I. P.; Tangerman, A.; Nagengast, F. M. Effect of resistant starch on colonic fermentation, bile acid metabolism, and mucosal proliferation. *Digest Dis. Sci.* **1994**, *39*, 834-842.
21. Hamer, H. M.; De Preter, V.; Windey, K.; Verbeke, K. Functional analysis of colonic bacterial metabolism: relevant to health? *Am. J. Physiol. Gastrointest. Liver Physiol.* **2012**, *302*, G1-G9.
22. Yin, P.; Zhao, X.; Li, Q.; Wang, J.; Li, J.; Xu, G. Metabonomics study of intestinal fistulas based on ultraperformance liquid chromatography coupled with Q-TOF mass spectrometry (UPLC/Q-TOF MS). *J. Proteome Res.* **2006**, *5*, 2135-2143.
23. Williams, R. E.; Lenz, E. M.; Evans, J. A.; Wilson, I. D.; Granger, J. H.; Plumb, R. S.; Stumpf, C. L. A combined ^{1}H NMR and HPLC-MS-based metabonomic study of urine from obese (fa/fa) Zucker and normal Wistar-derived rats. *J. Pharm. Biomed. Anal.* **2005**, *38*, 465-471.
24. Poroyko, V.; Morowitz, M.; Bell, T.; Ulanov, A.; Wang, M.; Donovan, S.; Bao, N.; Gu, S.; Hong, L.; Alverdy, J. C.; Bergelson, J.; Liu, D. C. Diet creates metabolic niches in the "inmature gut" that shape microbial communities. *Nutr. Hosp.* **2011**, *26*, 1283-1295.
25. Mellert, W.; Kapp, M.; Strauss, V.; Wiemer, J.; Kamp, H.; Walk, T.; Looser, R.; Prokoudine, A.; Fabian, E.; Krennrich, G.; Herold, M.; van Ravenzwaay, B. Nutritional impact on the plasma metabolome of rats. *Toxicol. Lett.* **2011**, *207*, 173-181.
26. Legido-Quigley, C.; Stella, C.; Perez-Jimenez, F.; Lopez-Miranda, J.; Ordovas, J.; Powell, J.; van-der-Ouderaa, F.; Ware, L.; Lindon, J. C.; Nicholson, J. K.; Holmes, E. Liquid chromatography-mass spectrometry methods for urinary biomarker detection in metabonomic studies with application to nutritional studies. *Biomed. Chromatogr.* **2010**, *24*, 737-743.
27. Ventura, M.; Turroni, F.; Canchaya, C.; Vaughan, E. E.; O'Toole, P. W.; van Sinderen, D. Microbial diversity in the human intestine and novel insights from metagenomics. *Front Biosci.* **2009**, *14*, 3214-3221.
28. Englyst, H. N.; Kingman, S. M.; Cumming, J. H. Classification and measurement of nutritionally important starch fractions. *Eur. J. Clin. Nutr.* **1992**, *46*, S33-S50.
29. Baghurst, P. A.; Baghurst, K. I.; Record, S. J. Dietary fibre, non-starch polysaccharides and resistant starch-a review. *Food Aust.* **1996**, *48*, S3-S35.
30. Hasjim, J.; Lee, S-O.; Hendrich, S.; Setiawan, S.; Ai, Y.; Jane, J. Characterization of novel resistant-starch and its effects on postprandial plasma-glucose and insulin responses. *Cereal Chem.* **2010**, *87*, 257-262.
31. Dettmer, K.; Aronov, P. A.; Hammock, B. D. Mass spectrometry-based metabolomics. *Mass Spectrom. Rev.* **2007**, *26*, 51-78.

32. Scalbert, A.; Brennan, L.; Fiehn, O.; Hankemeier, T.; Kristal, B. S.; van Ommen, B.; Pujos-Guillot, E.; Verheij, E.; Wishart, D.; Wopereis, S. Mass-spectrometry-based metabolomics: limitations and recommendations for future progress with particular focus on nutrition research. *Metabolomics* **2009**, *5*, 435-458.
33. Vitali, B.; Ndagijimana, M.; Maccaferri, S.; Biagi, E.; Guerzoni, M. E.; Brigidi, P. An *in vitro* evaluation of the effect of probiotics and prebiotics on the metabolic profile of human microbiota. *Anaerobe* **2012**, *18*, 386-391.
34. Blokland, M. H.; Van Tricht, E. F.; Van Rossum, H. J.; Sterk, S. S.; Nielen, M. W. F. Endogenous steroid profiling by gas chromatography-tandem mass spectrometry and multivariate statistics for the detection of natural hormone abuse in cattle. *Food Addit. Contam. A.* **2012**, *29*, 1030-1045.
35. Tsugawa, H.; Tsujimoto, Y.; Arita, M.; Bamba, T.; Fukusaki, E.; GC/MS based metabolomics: development of a data mining system for metabolite identification by using soft independent modeling of class analogy (SIMCA). *BMC Bioinformatics* **2011**, *12*, 131.
36. Plumb, R. S.; Granger, J. H.; Stumpf, C. L.; Johnson, K. A.; Smith, B. W.; Gaulitz, S.; Wilson, I. D., Castro-Perez, J. A rapid screening approach to metabonomics using UPLC and oa-TOF mass spectrometry: application to age, gender and diurnal variation in normal/Zucker obese rats and black, white and nude mice. *Analyst* **2005**, *130*, 844-849.
37. Mateos-Martín, M. L.; Pérez-Jiménez, J.; Fuguet, E.; Torres, J. L. Non-extractable proanthocyanidins from grapes are a source of bioavailable (epi)catechin and derived metabolites in rats. *Br. J. Nutr.* **2012**, *108*, 290-297.
38. Ohashi, Y.; Hirayama, A.; Ishikawa, T.; Nakamura, S.; Shimizu, K.; Ueno, Y.; Tomita, M.; Soga, T. Depiction of metabolome changes in histidine-starved *Escherichia coli* by CE-TOFMS. *Mol. Biosyst.* **2008**, *4*, 135-147.
39. Matsumoto, M.; Kibe, R.; Ooga, T.; Aiba, Y.; Kurihara, S.; Sawaki, E.; Koga, Y.; Benno, Y.; Impact of intestinal microbiota on intestinal luminal metabolome. *Scientific Reports* **2012**, *2*, 233.
40. Ooga, T.; Sato, H.; Nagashima, A.; Sasaki, K.; Tomita, M.; Soga, T.; Ohashi, Y. Metabolomic anatomy of an animal model revealing homeostatic imbalances in dyslipidaemia. *Mol. Biosyst.* **2011**, *7*, 1217-1223.
41. Hasegawa, M.; Ide, M.; Kuwamura, M.; Yamate, J.; Takenaka, S. Metabolic fingerprinting in toxicological assessment using FT-ICR MS. *J. Toxicol. Pathol.* **2010**, *23*, 67-74.
42. Antunes, L. C. M.; Han, J.; Ferreira, R. B. R.; Lolić, P.; Borchers, C. H.; Finlay, B. B. Effect of antibiotic treatment on the intestinal metabolome. *Antimicrob. Agents Chemother.* **2011**, *55*, 1494-1503.
43. Rath, C. M.; Alexandrov, T.; Higginbottom, S. K.; Song, J.; Milla, M. E.; Fischbach, M. A.; Sonnenburg, J. L.; Dorrestein, P. C. Molecular analysis of model gut microbiotas by imaging mass spectrometry and nanodesorption electrospray ionization reveals dietary metabolite transformations. *Anal. Chem.* **2012**, *84*, 9259-9267.
44. Chen, Y.; Zhu, S. B.; Xie, M. Y.; Nie, S. P.; Liu, W.; Li, C.; Gong, X. F.; Wang, Y. X. Quality control and original discrimination of *Ganoderma lucidum* based on high-performance liquid chromatographic fingerprints and combined chemometrics methods. *Anal. Chim. Acta.* **2008**, *623*,146-156.

45. Fremout, W.; Kuckova, S.; Crhova, M.; Sanyova, J.; Saverwyns, S.; Hynek, R.; Kodicek, M.; Vandenabeele, P.; Moens, L. Classification of protein binders in artist's paints by matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry: an evaluation of principal component analysis (PCA) and soft independent modeling of class analogy (SIMCA). *Rapid Commun. Mass Spectrom.* **2011**, *25*, 1631-1640.
46. Benabdelkamel, H.; Di Donna, L.; Mazzotti, F.; Naccarato, A.; Sindona, G.; Tagarelli, A.; Taverna, D. (2012) Authenticity of PGI "Clementine of Calabria" by multielement fingerprint. *J. Agric. Food Chem.* **2012**, *60*, 3717-3726.
47. Zhao, Y.; Hasjim, J.; Li, L.; Jane, J.; Hendrich, S.; Birt, D. F. Inhibition of azoxymethane-induced preneoplastic lesions in the rat colon by a cooked stearic acid complexed high-amylose cornstarch. *J. Agric. Food Chem.* **2011**, *59*, 9700-9708.
48. Reeves, P. G. Components of the AIN-93 diets as improvements in the AIN-76A diet. *J. Nutr.* **1997**, *127*, 838S-841S.
49. Zhang, B.; Huang, Q.; Luo, F.; Fu, X.; Jiang, H.; Jane, J. Effects of octenylsuccinylation on the structure and properties of high-amylose maize starch. *Carbohyd. Polym.* **2011**, *84*, 1276-1281.
50. Suhre, K.; Schmitt-Kopplin, P. MassTRIX: mass translator into pathways. *Nucleic Acids Res.* **2008**, *36*, W481-W484 (Web Server Issue).
51. Westerhuis, J. A.; de Jong, S.; Smilde, A. K. Direct orthogonal signal correction. *Chemometr. Intell. Lab Sys.* **2001**, *56*, 13-25.
52. Anderson, T. J.; Ai, Y.; Jones, R. W.; Houk, R. S.; Jane, J.; Zhao, Y.; Birt, D. F.; McClelland, J. F. Analysis of resistant starches in rat cecal contents using Fourier transform infrared photoacoustic spectroscopy. *J. Agric. Food Chem.* **2013**, *61*, 1818-1822.
53. Croswell, A.; Amir, E.; Teggatz, P.; Barman, M.; Salzman, N. H. Prolonged impact of antibiotics on intestinal microbial ecology and susceptibility to enteric *Salmonella* infection. *Infect. Immun.* **2009**, *77*, 2741-2753.
54. Manichanh, C.; Reeder, J.; Gibert, P.; Varela, E.; Llopis, M.; Antolin, M.; Guigo, R.; Knight, R.; Guarner, F. Reshaping the gut microbiome with bacterial transplantation and antibiotic intake. *Genome Res.* **2010**, *20*, 1411-1419.
55. Zhung, X.; Xie, G.; Zhao, A.; Zhao, L.; Yao, C.; Chiu, N. H. L.; Zhou, Z.; Bao, Y.; Jia, W.; Nicholson, J. K.; Jia, W. The footprints of gut microbial-mammalian co-metabolism. *J. Proteome Res.* **2011**, *10*, 5512-5522.
56. Antonopoulos, D. A.; Huse, S. M.; Morrison, H. G.; Schmidt, T. M.; Sogin, M. L.; Young, V. B. Reproducible community dynamics of the gastrointestinal microbiota following antibiotic perturbation. *Infect. Immun.* **2009**, *77*, 2367-2375.

Table 2-1

Metabolite Peaks Found in (C) Cecal and (D) Distal Colon Samples as Shown in Figures 2-1 through 2-3

Peak	m/z	Control	HA7	OS-HA7	StA-HA7
a	116.071	D			D
b	118.087	D	C D	C D	C D
c	121.065			C	
d	132.102	C D	C D	C D	C D
e	141.066		C		C
f	141.112	C			C
g	148.134	C D	C		C D
h	166.087	C D		C D	C D
i	173.129		C		C
j	195.113		C D		C D
k	212.14		C	C	C D
l	217.104				D
m	229.151			D	
n	230.187		C D		C D
o	242.187				D
p	251.128			C D	
q	273.108			D	
r	276.134		C D		C D
s	277.129		C D	C	C D
t	321.240				C D
u	335.219				C
v	336.228		D		D
w	357.239			C D	
x	373.274	D			D
y	377.265	C			
z	378.269	C			
aa	379.280	C			
ab	393.242	C	C		
ac	409.164				D
ad	479.263			C D	
ae	501.246			D	
af	563.268		D		
ag	591.322			C	
ah	595.352	D			C D
ai	733.559	C			
aj	737.286			C D	
ak	839.565				C D

Table 2-2

Biomarkers with Substantial Contribution to Class Differentiation Matched to KEGG Database for Cecal Contents

Cecal Contents	m/z	Compound[a]	Molecular Formula
Control	116.071	Proline $[M+H]^+$	$C_5H_9NO_2$
	132.102	Leucine $[M+H]^+$	$C_6H_{13}NO_2$
	132.102	Isoleucine $[M+H]^+$	$C_6H_{13}NO_2$
	148.134	No Database Match	Unknown
	377.265	3-Acetyl-5alpha-androstane-3beta,17beta-diol 3-acetate $[M+H]^+$	$C_{23}H_{36}O_4$
	595.350	L-Stercobilin $[M+H]^+$	$C_{33}H_{46}N_4O_6$
HA7 Diet	118.087	Valine $[M+H]^+$	$C_5H_{11}NO_2$
	141.113	No Database Match	Unknown
	161.093	Alanyl-alanine $[M+H]^+$	$C_6H_{12}N_2O_3$
	161.093	Nonadienal $[M+Na]^+$	$C_9H_{14}O$
	203.053	Glucose $[M+Na]^+$	$C_6H_{12}O_6$
	336.228	12-hydroxyicosa-trienoic acid $[M+H]^+$	$C_{20}H_{31}O_4$
OS-HA7	251.129	Methylripariochromene A $[M+H]^+$	$C_{14}H_{18}O_4$
	294.157	No Database Match	Unknown
	357.238	Allotetrahydrodeoxycorticosterone $[M+Na]^+$	$C_{21}H_{34}O_3$
	479.263	No Database Match	Unknown
	591.322	Urobilin $[M+H]^+$	$C_{33}H_{42}N_4O_6$
	591.322	Deoxycholic acid 3-glucuronide $[M+Na]^+$	$C_{30}H_{48}O_{10}$
StA-HA7	242.187	No Database Match	Unknown
	276.134	Alanyltryptophan $[M+H]^+$	$C_{14}H_{17}N_3O_3$
	277.129	Homoanserine $[M+Na]^+$	$C_{11}H_{18}N_4O_3$
	321.240	Oxooctadecanoic acid $[M+Na]^+$	$C_{18}H_{34}O_3$
	321.240	hydroxy-Eicosatetraenoic acid $[M+H]^+$	$C_{20}H_{32}O_3$
	335.219	hydroperoxy, octadecadienoic acid $[M+Na]^+$	$C_{18}H_{32}O_4$

[a]Notation refers to the stated molecule (M) and the observed adduct bracketed with the positive charge

Table 2-3

Biomarkers with Substantial Contribution to Class Differentiation Matched to KEGG Database for Distal Colon Contents

Distal Contents	m/z	Compound[a]	Molecular Formula
Control	116.071	Proline $[M+H]^+$	$C_5H_9NO_2$
	141.111	No Database Match	Unknown
	148.134	No Database Match	Unknown
	150.059	Methionine $[M+H]^+$	$C_5H_{11}NO_2S$
	595.353	L-Stercobilin $[M+H]^+$	$C_{33}H_{46}N_4O_6$
HA7 Diet	230.187	No Database Match	Unknown
	254.155	Gamma-Aminobutyryl-lysine $[M+Na]^+$	$C_{10}H_{21}N_3O_3$
	255.147	Homoanserine $[M+H]^+$	$C_{11}H_{18}N_4O_3$
	336.228	12-hydroxyicosa-trienoic acid $[M+H]^+$	$C_{20}H_{31}O_4$
	563.268	Protoporphyrin $[M+Na]^+$	$C_{34}H_{34}N_4O_4$
OS-HA7	118.087	Valine $[M+H]^+$	$C_5H_{11}NO_2$
	212.140	No Database Match	Unknown
	251.127	Methylripariochromene A $[M+H]^+$	$C_{14}H_{18}O_4$
	294.162	No Database Match	Unknown
	357.240	Allotetrahydrodeoxycorticosterone $[M+Na]^+$	$C_{21}H_{34}O_3$
	737.286	No Database Match	Unknown
StA-HA7	116.071	Proline $[M+H]^+$	$C_5H_9NO_2$
	124.039	Picolinic acid $[M+H]^+$	$C_6H_5NO_2$
	166.087	Phenylalanine $[M+H]^+$	$C_9H_{11}NO_2$
	166.087	Stachydrine $[M+Na]^+$	$C_7H_{13}NO_2$
	212.140	No Database Match	Unknown
	321.240	Oxooctadecanoic acid $[M+Na]^+$	$C_{18}H_{34}O_3$
	321.240	hydroxy-Eicosatetraenoic acid $[M+H]^+$	$C_{20}H_{32}O_3$
	409.164	Burseran $[M+Na]^+$	$C_{22}H_{26}O_6$
	409.164	1,2-Dihydro-5-hydroxy-2-(1-hydroxy-1-methylethyl)-4-(isobutyryl)-6-phenylfurano [2,3-h][1]benzopyran-8-one $[M+H]^+$	$C_{24}H_{24}O_6$

[a]Notation refers to the stated molecule (M) and the observed adduct bracketed with the positive charge

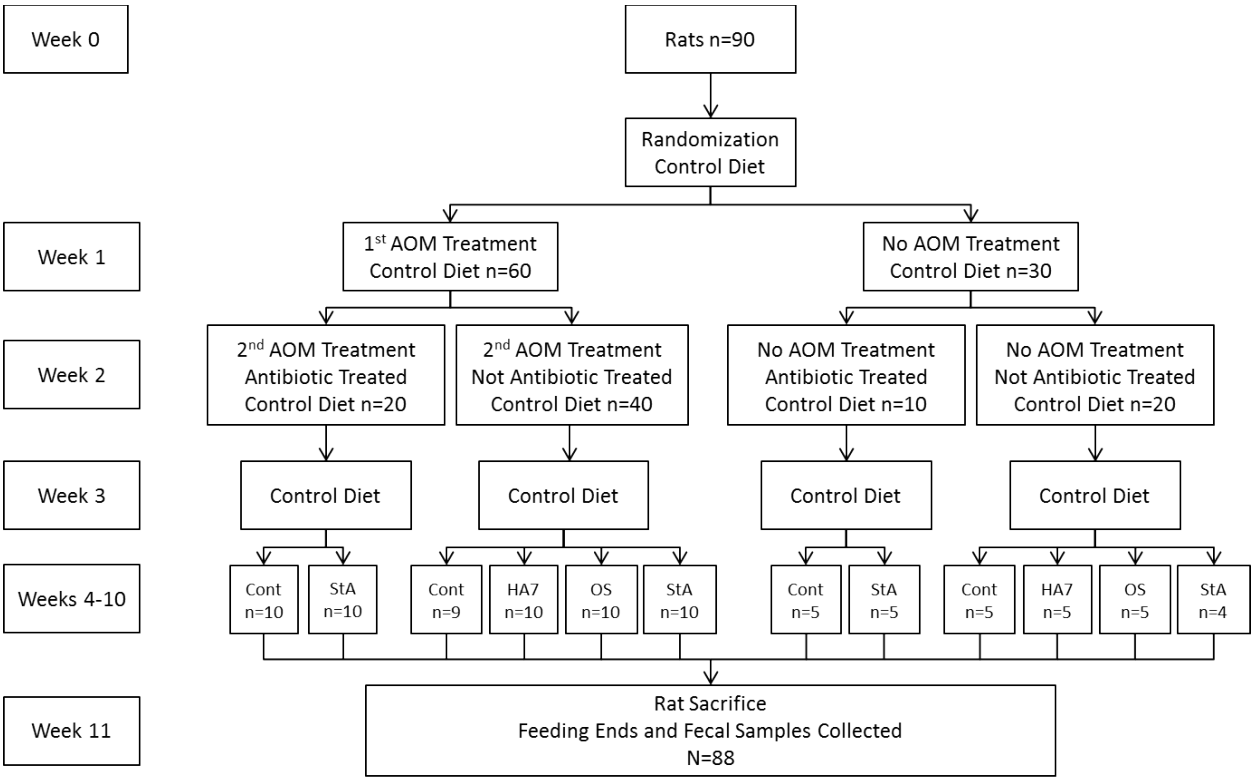


Figure 2-1 Flow diagram of the animal study detailing the treatment schedule for AOM, antibiotic, and diet up until sacrifice. The diets that were given for weeks 4-10 are represented as the abbreviations cont (Control), HA7 (HA7), OS (OS-HA7), and StA (StA-HA7). The amount of rats in each group of the flow chart are represented by the letter "n"

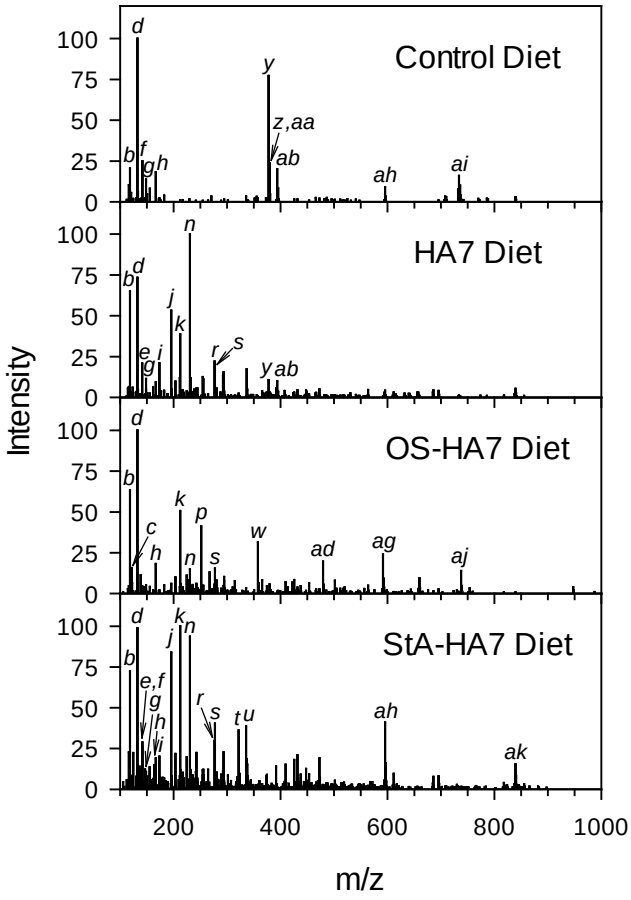


Figure 2-2 Averaged mass spectra of cecal content samples from each diet group with no AOM or antibiotic treatments. The peak intensities have been normalized against the strongest peak in the spectrum for each diet. The most prominent metabolite peaks in each diet have been labeled and can be matched to their respective m/z in Table 2-1

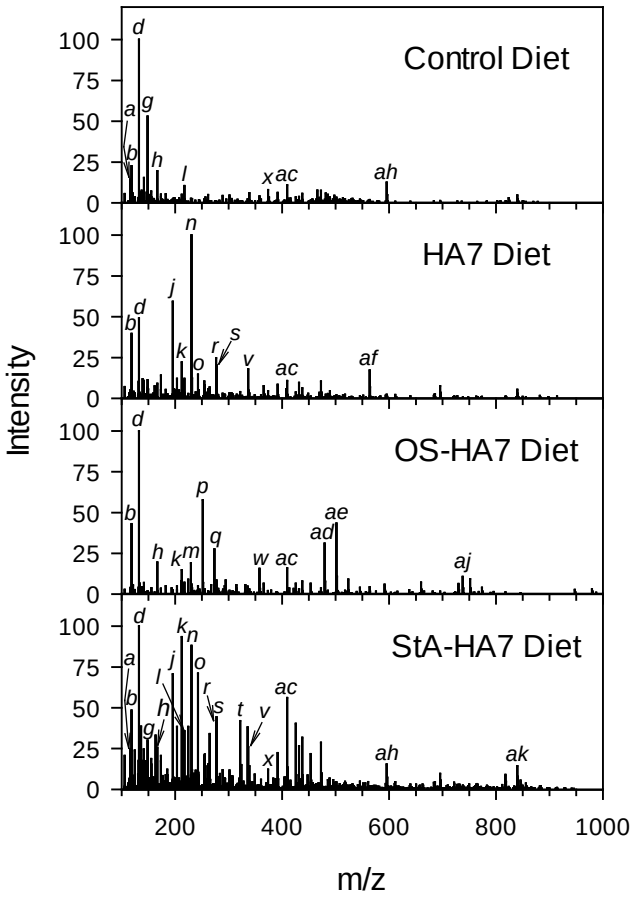


Figure 2-3 Averaged mass spectra of distal colon content samples from each diet group with no AOM or antibiotic treatments. The peak intensities have been normalized against the strongest peak in the spectrum for each diet. The most prominent metabolite peaks in each diet have been labeled and can be matched to their respective m/z in Table 2-1

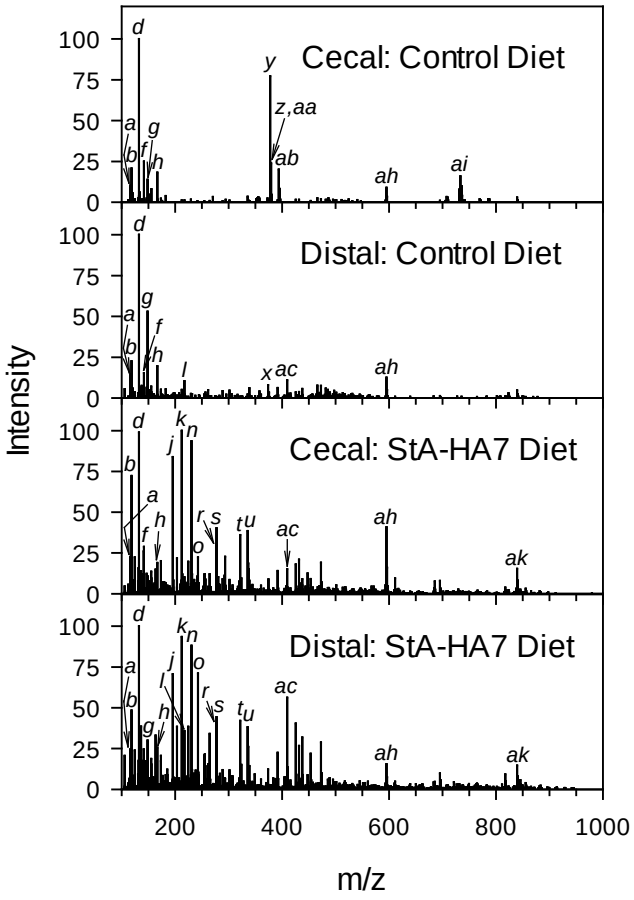


Figure 2-4 Averaged mass spectra comparison of cecal and distal colon content samples from two example diets with no AOM or antibiotic treatments. The peak intensities have been normalized against the strongest peak in the spectrum for each diet. The most prominent metabolite peaks in the spectra have been labeled and can be matched to their respective m/z in Table 2-1

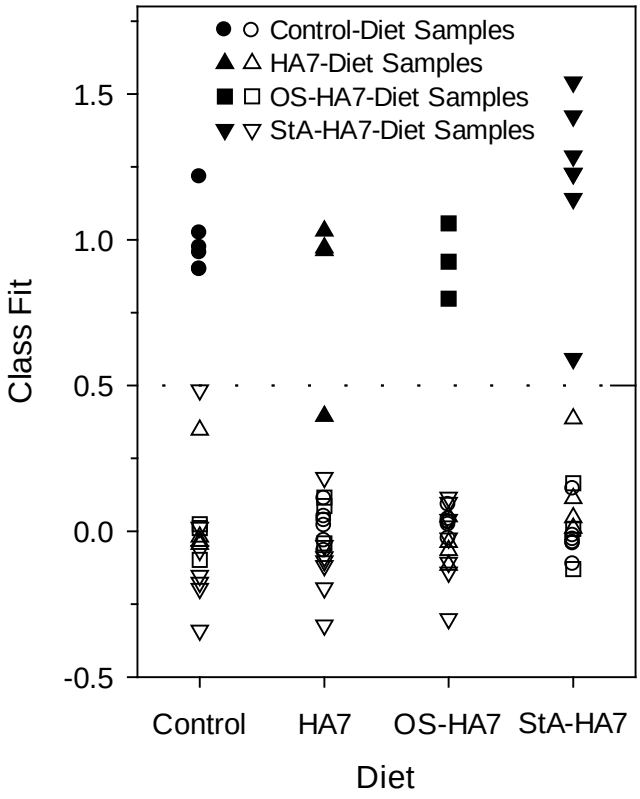


Figure 2-5 PLS-DA analysis of cecal-content verification-set samples for diet treatments, including AOM and antibiotic treated samples. The solid markers represent the samples that belong to the tested class and should ideally have a class fit value of 1.0. The open markers correspond to samples which do not belong to the tested class and ideally have a class fit of 0

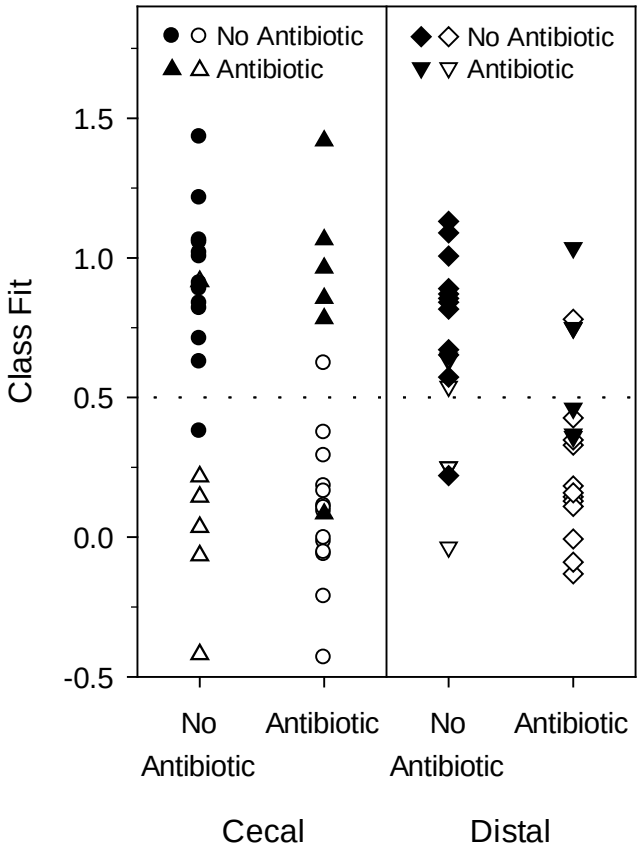


Figure 2-6 PLS-DA analysis of cecal-content and distal-colon-content verification-set samples for antibiotic treatment, including diets and AOM treated samples. The solid markers represent the samples that belong to the tested class and should ideally have a class fit value of 1.0. The open markers correspond to sample which do not belong to the tested class and ideally have a class fit of 0

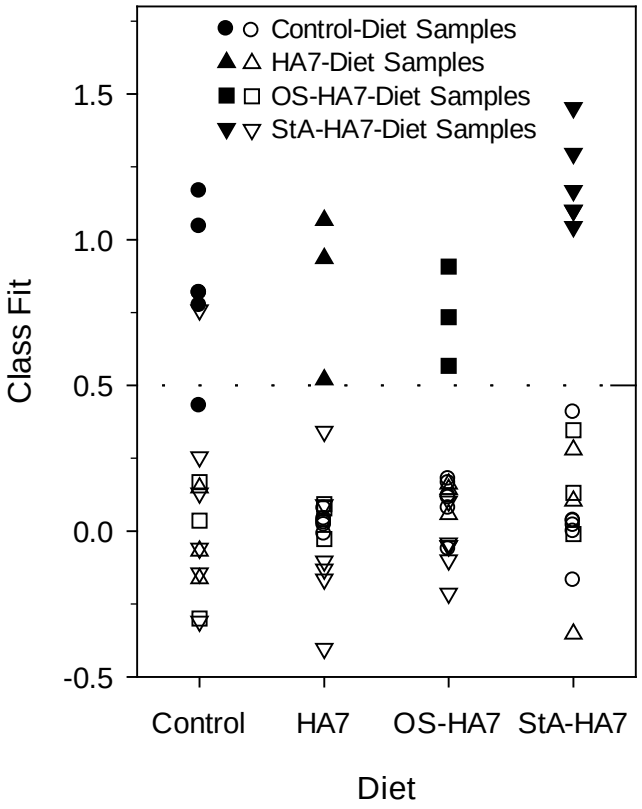


Figure 2-7 PLS-DA analysis of distal-colon-content verification-set samples for diet treatments, including AOM and antibiotic treated samples. The solid markers represent the samples that belong to the tested class and should ideally have a class fit value of 1.0. The open markers correspond to samples which do not belong to the tested class and ideally have a class fit of 0

CHAPTER 3

COMPREHENSIVE IDENTIFICATION OF ALPHA-ZEIN PROTEINS BY HIGH - PERFORMANCE LIQUID CHROMATOGRAPHY ACCURATE MASS TIME-OF-FLIGHT MASS SPECTROMETRY

Paper submitted to *The Journal of Agricultural and Food Chemistry* .

By

Derrick L. Morast,[1,2,3,4] Timothy J. Anderson,[1,2,4] R. S. Houk[1,2*]

[1]Ames Laboratory-USDOE, Iowa State University, Ames, Iowa 50011

[2]Department of Chemistry, Iowa State University, Ames, Iowa 50011

[3]Present address: Exxon Mobil Refining & Supply, Baytown TX 77520 USA

[4]These authors contributed equally to this work

* Corresponding Author. Tel: 1-515-294-9462 Fax -0050 rshouk@iastate.edu

Abstract

High-resolution mass spectrometry (HRMS) was used to detect many proteins in α -zein samples extracted from corn gluten meal (CGM) and distillers' dried grains with solubles (DDGS). High-performance liquid chromatography (LC) was utilized for the separation of the complex protein samples. Proteins were introduced into the mass spectrometer via electrospray ionization (ESI). Measured molecular weights (M_r) were compared with protein database values and previous mass spectrometric α-zein analyses. Overall, 95 α-zein proteins were identified; 49 of these had M_r values similar to those of previously reported proteins. The remaining 46 proteins are identified for the first time. Interestingly, the α-zein extracted from DDGS contained many of the same intact proteins observed in the CGM extract, despite the fermentation and thermal processes used to convert dry milled corn to DDGS.

Keywords Zein, corn gluten meal, CGM, distillers dried grains with solubles, DDGS, high-resolution mass spectrometry, high-performance liquid chromatography

Introduction

Zeins, also known as corn prolamines, are alcohol-soluble storage proteins. Their main function is to provide nitrogen during germination. Zein proteins have a high content of glutamine and proline and are insoluble in water. There are currently many commercial applications for zeins, especially in adhesives and plastics. There is considerable interest in utilizing zein as a polymeric material for films, coatings, and plastics since it is both biodegradable and renewable (1-4).

Several methods have been reported for the extraction of zeins from maize endosperm (5-8), along with corresponding nomenclature systems. The most widely used nomenclature, described by Esen and used herein, classifies zein fractions as α-, β-, γ-, and δ-zeins based on differences in solubility and amino acid sequence (9, 10). Of the four fractions, α-zein is the most abundant, containing 75 to 85% of total zein proteins. They range in size from 210 to 245 amino acids and are commonly divided into two groups, M_r ~ 22 kDa and M_r ~ 19 kDa, based on relative SDS-PAGE migration. The protein sequences are ~ 90% homologous within each group and ~ 60% homologous between the two groups (11). The α-zein proteins contain high levels of glutamine (~25%), leucine (~20%), alanine (~15%), and proline (~11%) (12).

Zein is produced from corn using four different methods, two of which will be discussed further. Corn wet-milling and the dry-grind ethanol process both generate coproducts which contain α-zein proteins. CGM, a coproduct of the corn wet-milling process, can contain between 62 to 74% protein (13). Many commercially useful applications have been found for CGM and most commercial zein extraction methods remove mainly α-zein proteins from CGM (14, 15). For this reason, CGM is fairly valuable and currently sells for around $550/ton.

The dry-grind ethanol process used for the production of ethanol from corn generates DDGS as a coproduct. Protein content in DDGS is about 28 to 30% (16). During the ethanol

process, a high-temperature cooking step (90 to 105 °C for 1 to 3 hrs) is used, which may alter the α-zein proteins. Only small scale extractions of α-zein proteins from DDGS have been reported (17). DDGS has found little commercial application outside of use as livestock feed and currently sells for only $100/ton or less.

These α-zein extraction procedures yield complex mixtures of many proteins. Mass spectrometry (MS) is a powerful tool for the analysis of such protein mixtures. When ionized by electrospray ionization (ESI), protein cations are highly charged and are observed in a wide variety of charge states. This charge state distribution allows for the molecular weight of the protein to be calculated from many measured values of mass-to-charge ratio (m/z). A high-resolution mass spectrometer (HRMS) measures m/z values more accurately than lower resolution instruments. In favorable cases, individual proteins can often be identified in complex mixtures by HRMS alone (18, 19), especially if isotope peaks within individual charge states can be resolved, and protein extracts can be sufficiently purified for a predicted protein class.

High-performance liquid chromatography (LC) is very useful for separation of protein mixtures. However, the complexity of the α-zein sample and the sequence homology of the α-zein proteins make it difficult to get complete separation with LC alone. A previous 2-dimensional electrophoresis study identified 41 α-zein proteins and it was noted that this number was well below the estimated number of α-zein genes in the genome (20). One benefit of using HRMS is that baseline LC separation of the proteins is often not necessary. Also, deconvolution software can classify most co-eluting proteins.

To our knowledge, all previous maize protein separations by LC used UV absorbance detection instead of MS. Although LC/UV has many useful applications, the lack of mass information limits its ability to identify proteins in such complex mixtures. In particular, it is

difficult to identify proteins from retention time alone unless the analytes are almost baseline resolved and retention time is compared against purified reference standards of each protein of interest. Given the inability of LC to baseline separate each α-zein protein and the lack of standards, identification of coeluting compounds in very complex mixtures by UV absorbance detection alone is difficult.

Capillary electrophoresis (CE) has been coupled to a mass spectrometer via ESI for the analysis of maize proteins (21, 22). CE generally provides a faster separation with higher resolution than LC, however coupling CE to a mass spectrometer is not as straightforward. For that reason, LC is a more widely used technique with MS detection.

The present work compares LC-MS results for zeins to those reported by others using CE-MS and MALDI-MS. Many new zein proteins are reported here.

Materials and Methods

Alpha-Zein Protein Extraction

CGM was obtained from Cargill, Inc. (Eddyville, IA) and DDGS was obtained from Lincolnway Energy (Nevada, IA). The CGM and DDGS extraction methods were reported previously (23, 24). A variety of solvent systems were employed. For CGM, a total of ten samples were extracted using five different solvent systems: 88% aqueous 2-propanol (IPA), 70% aqueous IPA, 55% aqueous IPA, 70% aqueous IPA with 22.5% glycerol, and 70% aqueous ethanol (EtOH). Each extraction was done with or without sodium hydroxide (NaOH) and a reductant (sodium bisulfite). For DDGS, a total of 6 samples were extracted using three different solvent systems: 88% aqueous IPA, 70% aqueous IPA, and 70% aqueous EtOH. These solvents were chosen based on performance results from previous CGM extractions (23). Half of the 6

samples were pretreated with a combination of the enzyme mixtures Multifect GX GC (0.4%) and Multifect pectinase FE (0.1%) (Genencor International, New York, NY), while the other half had no enzyme pretreatment. These enzyme mixtures were employed to remove any crystalline cellulose matrix which may have impeded extraction of zeins.

After extraction with the solvent of interest, the solution was centrifuged for 15 min at room temperature and 8000 x *g* (Beckman, Palo Alto, CA). The supernatant was cold precipitated at -20 °C overnight. The suspended precipitate was centrifuged at -20 °C and 8000 x *g* to remove excess solvent and impurities, leaving a purified zein protein pellet. This pellet was dissolved in 88% IPA in water and dried in a vacuum oven at 50 °C at 0.6 bar. Dried samples were kept at 4 °C until use.

Figure 3-1 shows two representative SDS-PAGE gels of CGM and DDGS extraction products. The two bands observed for each sample correspond to the 19 and 22 kDa α-zein protein fractions. All five CGM samples were extracted without sodium hydroxide or a reducing agent. A similar band pattern was observed for the five corresponding extractions with sodium hydroxide and the reducing agent. For DDGS, samples 1-3 had no enzyme pretreatment prior to extraction while samples 4-6 were pretreated.

HPLC-TOF MS

Fifty milligrams of α-zein protein extract were dissolved in 55% aqueous IPA with 5% (v/v) 2-mercaptoethanol. A 5 µL sample was injected onto a reversed phase LC column (3.0 x 150 mm, 3.5 µm Agilent Zorbax 300 Å Stable Bond-C3). Proteins were eluted using a 60 min linear gradient from 50% to 60% acetonitrile, with 0.1% trifluoroacetic acid (TFA), on an Agilent 1260 Infinity LC system (Agilent Technologies). The accurate-mass Agilent 6224 time-

of-flight mass spectrometer was operated at a data acquisition rate of 4 GHz. Mass spectra were deconvoluted using Agilent MassHunter BioConfirm software. Protein molecular weights were measured using a minimum of ten consecutive charge states, a signal-to-noise ratio (S/N) of thirty or more, and a protein fit score of at least 9 (out of 10).

Results and Discussion

Previous MS Studies of Zein Proteins

Previous zein protein studies utilizing mass spectrometry involved low resolution instruments and relied on either matrix-assisted laser desorption/ionization (MALDI) or CE coupled to ESI-MS (21, 22, 25, 26). There are several potential issues with each of these techniques, especially when coupled to a low resolution mass spectrometer. First, some α-zein proteins differ in molecular weight by only 2 Da. With MALDI, each laser shot has the potential to ionize all proteins present in the sample. Without the separating power of a high resolution mass spectrometer, the direct analysis of every protein in the complex zein sample would be very difficult. This is evident when referring to mass spectra generated via MALDI-MS by Wang et al. and Adams et al (25, 26). They saw broad mass spectral peaks spanning ~100 Da that are probably the sum of peaks from 10 or more α-zein proteins instead of individual proteins. A second concern when using MALDI directly on complex samples is the ability to detect low concentration proteins within the mixture of α-zein proteins. Since no analytical separation is performed prior to ionization, most of the signal generated will be from proteins at higher concentrations.

Third, there could be issues with the proteins co-crystallizing with the matrix when using MALDI. Alpha-zein proteins have peculiar properties, especially their solubility characteristics.

As solvent is removed, they tend to aggregate into spherical structures or arrange into ordered films (27). It is possible that during solvent evaporation, some α-zein proteins aggregate rather than co-crystallize with the matrix. These proteins would not be ionized or detected efficiently by MALDI.

CE-ESI is a better alternative than MALDI when using a low resolution mass spectrometer. One advantage of CE over MALDI is the ability to separate proteins in the complex mixture prior to analysis. However, CE-ESI-MS has lower sensitivity than LC-ESI-MS. The number of charge states generated for each α-zein protein via CE-ESI appears to be lower as well, perhaps because of the buffers and background electrolyte used during CE. Erny et al. had to use as few as three charge states for protein identification, whereas with LC-ESI we require at least ten observable peaks from consecutive charge states in the present work (21). For many proteins, over 20 charge states were observed. Signal-to-noise ratio appears to be much lower for the previous CE studies than our current LC study as well.

Measured Mass Spectra and M_r Values

As an example of our results, Figure 3-2a shows a typical mass spectrum obtained during an eluting fraction that lasted ~ 60 s. The resulting deconvoluted spectrum for protein M_r 24,137.61 Da is also shown. At least three other proteins co-eluted during this timeframe. However, the deconvolution software recognizes the peaks corresponding to each protein of interest (bottom panel). The inset displays a zoomed in view of the area around the 17+ peak showing that the peaks from other proteins are almost baseline resolved by the MS.

Using M_r data from the LC-ESI-MS measurements, GenBank and UniProt were searched for amino acid sequences corresponding to zein proteins expected from *zea mays*. Molecular

weights of each protein, without the signal peptide, were calculated. The signal peptide for α-zein proteins is typically 21 amino acids in length. This signal peptide is cleaved off naturally before the zeins are extracted and is therefore absent from the mature protein (28).

Comparison of Observed Alpha-Zein Proteins with Database Entries

Table 3-1 lists the observed molecular weights of those α-zein proteins, measured via high-resolution mass spectrometry, which closely match database search results. In several cases, the amino acid sequence of different database entries vary only due to the loss of the signal peptide. This resulted in multiple "matches" for the mature protein. For that reason, the literature column lists the reference cited for each entry (21, 22, 25-44). The CGM and DDGS columns list the number of samples (out of 10 and 6, respectively) in which each protein was identified.

For the proteins listed in Table 3-1, 71% (24 of 34) were in at least one CGM and DDGS sample. However, of those 24 proteins, 18 of them were in at least half of all CGM and DDGS samples. Four of the proteins were found only in CGM while six were only found in DDGS. Most of the proteins (85%) have observed molecular weights which are within 2 Da of the calculated values. Fourteen of the proteins listed in Table 3-1 were identified in a previous mass spectrometry analysis of zein (21, 22, 25, 26). The remaining 18 proteins are identified and reported here for the first time and are italicized in Table 3-1.

The results in Table 3-1 lead to the following key points. First, this and subsequent tables report only M_r values, not actual sequences. The absolute sequence of a protein cannot be established via mass spectrometry without either digesting the protein and sequencing the resulting peptides (bottom-up proteomics) or fragmenting the intact protein using dissociation

techniques, usually electron-capture/electron-transfer (top-down proteomics). Neither of these techniques was utilized for the present study. Digestion and CID of LC α-zein protein fractions would potentially provide more definitive assignments. However typical enzyme digestions occur in aqueous solutions and α-zein proteins are not soluble in water.

The mass spectrometry studies cited above definitively match many specific database entries, namely those from Woo et al (29). In some cases, our experimental M_r values vary by 10s to 100s of Da from supposedly matched or identified proteins. Note that each protein listed in the present work was observed in at least ten MS peaks from consecutive charge states with a minimum protein fit score of nine and a signal-to-noise ratio of at least thirty.

Thus, we are confident that each protein identified represents an actual protein in a CGM sample, a DDGS sample, or both. However, we cannot definitively say each protein listed is the exact protein cited from the database.

One special case is the two possible assignments for the 26,752.24 Da protein, noted with asterisks in Table 3-1. This predicted protein was near two observed ones, 1.26 Da from 26,750.98 and 0.65 Da from 26,752.89. Although these two α-zein proteins share ~85% of their sequence, they also have 37 different amino acids. Of the many proteins found, this is the only case in which an observed protein closely matches two database proteins. Without at least partial digestion and sequence analysis, we cannot match the sequence in the observed peaks to those in the database, so both are listed.

Second, there are a few cases in which a database protein was found in only one sample. These proteins tended to be lower abundance components of the total sample. As such, they only met the rigorous deconvolution criteria in one (or a few) sample(s). When the criteria were

lowered to a S/N of 10 and only 5 consecutive charge states, many of the low-abundance proteins were observed in multiple samples (data not shown).

For those proteins which were found in only some of the CGM or DDGS samples, the sodium bisulfite and NaOH treatments did affect whether a protein was observed or not. In other words, one protein might not be observed in CGM extracted with 70% aqueous IPA treated with NaOH and sodium bisulfite, while that same protein would be present in CGM treated the same way. This observation was found for the DDGS samples with or without enzyme treatment (data not shown). A more detailed analysis of each sample extraction procedure, including protein purity, yield, and recovery assessments, can be found in previous publications (23, 24).

There were 15 proteins from Table 3-1 which were also identified in one of the previous mass spectrometry publications. These proteins are listed in Table 3-2. All but two of the proteins listed in Table 3-2 were identified in over half of our CGM and DDGS samples. In addition, 12 of the 15 proteins were identified in 8 or more CGM samples while 13 of the 15 proteins were found in 4 or more of the DDGS samples. We consider this reasonable agreement with previous work.

Some proteins have been reported in multiple publications. For those proteins, both the observed molecular weights are listed. In some cases the original publication did not directly cite the specific database entry shown in Table 3-1. In other cases, the original publication does cite the database entry listed in Table 3-1, even though their observed molecular weights sometimes deviate from the calculated value by a large margin. For these entries, we simply report the previous data and do not imply that both groups actually identify the same protein.

As an extreme example, Adams et al. report GenBank accession number AF371268, with a calculated M_r of 24,705.60 Da (25). Their observed M_r was 24,515 Da. Our data, taken on an

instrument with much higher resolution and mass accuracy, measured an α-zein protein with M_r 24,706.84 Da. We did not observe a protein fitting our acceptance criteria with M_r 24,515 Da, as discussed below.

For the data listed in Table 3-2, previous analysis via CE-ESI-MS generated data with observed molecular weights within ~1.4 Da of the calculated values, much closer than previous MALDI-MS experiments (~43.6 Da) (21, 22). Of course the deviation from the calculated value for MALDI analysis is skewed by the small sample size and the 191 Da deviation listed above. If that single aberrant value is removed, the deviation drops to 14.2 Da. Although this value is better, it is still much larger than that for the data generated via HRMS (1.2 Da, present work) and CE-MS (21-22). We believe this point illustrates the previous stated concerns regarding the analyses of a complex mixture of very similar proteins, such as α-zein proteins, via MALDI.

Alpha-Zein Proteins without Database Entries

Table 3-3 compares molecular weights of α-zein proteins measured via HRMS with M_r values similar to those found in previous MS studies, but which were not in the database. Nine of the 15 proteins listed in Table 3-3 were found in over half of the CGM and DDGS samples. Moreover, 10 of the 15 were found in 9 or more of the CGM samples and 11 of 15 were found in four or more of the DDGS samples. One protein, M_r 23,217.76, was found in all 10 of the CGM samples but none of the DDGS samples. Of the α-zein proteins which were compared to database entries or previous mass spectrometry studies (Tables 3-1 and 3-3), 27 of the 49 (55%) proteins were found in over half of the CGM and DDGS samples. Twenty-nine proteins were identified in 8 or more of the CGM samples and thirty-four were found in four or more of the DDGS samples.

It is interesting to compare our list of proteins in Tables 3-1 to 3-3 to previous data. In some previous studies a protein is assigned to a specific GenBank accession number, implying that the protein observed has the same sequence as that given in the database. For example, Adams et al. observed (25), via MALDI, an α-zein protein with M_r 24,069 Da. They assigned it to GenBank accession number AF371271 with a calculated M_r 24,087 Da. However, our study shows two proteins with slightly different M_r values (24,071.90 Da and 24,087.90 Da), both of which were present in every CGM and DDGS sample. Wang et al (26). report an α-zein protein with M_r 24,097 Da, also via MALDI, and assign it to the same protein (AF371271, M_r 24,087 Da). We find an α-zein protein having a mass of 24,096.55 Da. Erny et al (21). also reported an α-zein protein of mass 24,695 Da, observed via CE-ESI-MS. This protein is compared to GenBank accession number AF371268 from Woo et al (29). having a calculated M_r 24,706 Da. It is also compared to a protein found by Adams et al (25). of mass 24,644 Da. Adams, however, actually compared their observed 24,644 Da protein to GenBank accession number AF371267, with a calculated mass of 24,818 Da. Our study found α-zein proteins having a mass of 24,694.62 Da and 24,096.55 Da. We did not observe an α-zein protein having a mass of 24,818 Da.

Table 3-4 contains a list of 47 observed proteins observed in the present work that fit the deconvolution criteria. These proteins did not match proteins from either the database or previous mass spectrometry studies. It is possible that some of these "new" proteins are Na^+ or K^+ salt adducts of expected proteins (45). Overall, only 10 of these 47 proteins were found in over half of the CGM and DDGS samples. Fifteen of the 47 "new" proteins were found in 7 or more of the CGM samples and 18 of the 47 were found in 4 or more of the DDGS samples. Twelve of the 47 proteins were found only in a single CGM or DDGS sample; two of these, Mr

26,375 and 26,419.76, were identified in one CGM and one DDGS sample. Three of the proteins were identified in every DDGS sample but none of the CGM samples. Only one protein was found in every CGM sample but none of the DDGS samples.

Comparison of Proteins from CGM and DDGS

One goal of the present work was to compare the individual α-zein proteins extracted from both CGM and DDGS. Most commercial zein is produced from CGM and as such, CGM commands a higher market price. Heating during the fermentation process is thought to adulterate α-zein proteins in DDGS. However, previous studies have shown the ability to extract a high purity zein protein fraction from the much cheaper DDGS, although no molecular weight information has been described other than that provided by SDS-PAGE (24, 46, 47).

Of the α-zein proteins previously reported (Tables 1-3), about 71% (34 of 48) were found in both CGM and DDGS. Overall, 56 of the 95 (59%) α-zein proteins that fit our deconvolution criteria were found in both CGM and DDGS. Also, 22 proteins were found only in DDGS samples and 17 proteins were found only in CGM samples.

Based solely on these results, there is no simple answer to the question of whether α-zein proteins are adulterated during conversion of dry milled corn to DDGS. However, we believe our findings, in support with the information from previous publications, suggest conventional DDGS contains many viable α-zein proteins.

Acknowledgements

The authors would like to thank Professor Buddhi Lamsal (Iowa State University) for his guidance in the alpha zein protein extractions. We also thank Kamel Harrata, Instrument Services group, Department of Chemistry, Iowa State University for providing access to the instrumentation. Derrick Morast and Timothy Anderson would like to acknowledge the GAANN Fellowship (Iowa State University, 2009) for financial support.

References

1. Lawton, J. W. Zein: A history of processing and use. *Cereal Chem.* **2002**, *79*, 1-18.
2. Wu, Q. X.; Sakabe, H.; Isobe, S. Studies on the toughness and water resistance of zein-based polymers by modification. *Polymer* **2003**, *44*, 3901-3908.
3. Wang, Y.; Padua, G. W. Tensile properties of extruded Zein sheets and extrusion blown films. *Macromol. Mater. Eng.* **2003**, *288*, 886-893.
4. Lawton, J. W. Plasticizers for zein: Their effect on tensile properties and water absorption of zein films. *Cereal Chem.* **2004**, *81*, 1-5.
5. Wilson, C. M. A nomenclature for zein polypeptides based on isoelectric focusing and sodium dodecyl sulfate polyacrylamide gel electrophoresis. *Cereal Chem.* **1985**, *62*, 361-365.
6. Wilson, C. M. Serial analysis of zein by isoelectric focusing and sodium dodecyl sulfate gel electrophoresis. *Plant Physiol.* **1986**, *82*, 196-202.
7. Hastings, H.; Bonanomi, S.; Soave, C.; Di Fonzo, N.; Salamini, F. Mapping of genes for minor zein sodium dodecylsulfate subunits and revision of zein gene nomenclature. *Genet. Agrar.* **1984**, *38*, 447-464.
8. Esen, A. Separation of alcohol-soluble proteins (zeins) from maize into three fractions by differential solubility. *Plant Physiol.* **1986**, *80*, 623-627.
9. Thompson, G. A.; Larkins, B. A. Structural elements regulating zein gene expression. *Bioessays* **1989**, *10*, 108-113.
10. Esen, A. A proposed nomenclature for the alcohol-soluble proteins (zeins) of maize (Zea mays L.). *J. Cereal Sci.* **1987**, *5*, 117-128.
11. Coleman, C. E.; Larkins, B. A. The prolamins of maize. In *Seed Proteins*, Shewry, P. R.; Casey, R., Eds. Kluwer: Dordrecht, Netherlands, **1999**; pp 109-139.
12. Thompson, G. A.; Larkins, B. A. Characterization of zein genes and their regulation in maize endosperm. In *The Maize Handbook*, Freeling, M.; Walbot, V., Eds. Springer-Verlag: New York, **1994**; pp 639-647.
13. Wu, S. W.; Myers, D. J.; Johnson, L. A. Factors affecting yield and composition of zein extracted from commercial corn gluten meal. *Cereal Chem.* **1997**, *74*, 258-263.
14. Carter, R.; Reck, D. R. Low temperature solvent extraction process for producing high purity zein. U.S. Patent 3,535,305, Oct. 20, 1970.
15. Takahashi, H.; Yanai,, N. Process for refining zein. U.S. Patent 5,342,923, Aug. 30, 1994.
16. Singh, V.; Moreau, R. A.; Hicks, K. B.; Belyea, R. L.; Staff, C. H. Removal of fiber from distillers dried grains with solubles (DDGS) to increase value. *Trans. ASAE* **2002**, *45*, 389-392.
17. Anderson, T. J.; Lamsal, B. P. Zein extraction from corn, corn products, and coproducts and modifications for various applications: A review. *Cereal Chem.* **2011**, *88*, 159-173.
18. Jensen, P. K.; Paša-Tolić, L.; Anderson, G. A.; Horner, J. A.; Lipton, M. S.; Bruce, J. E.; Smith, R. D. Probing proteomes using capillary isoelectric focusing-electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* **1999**, *71*, 2076-2084.
19. Mann, M.; Hendrickson, R. C.; Pandey, A. Analysis of proteins and proteomes by mass spectrometry. *Annu. Rev. Biochem.* **2001**, *70*, 437-473.

20. Consoli, L.; Damerval, C. Quantification of individual zein isoforms resolved by two-dimensional electrophoresis: Genetic variability in 45 maize inbred lines. *Electrophoresis* **2001**, 2983-2989.(21) Erny, G. L.; Marina, M. L.; Cifuentes, A. Capillary-electrophoresis mass spectrometry of zein proteins from conventional and transgenic maize. *Electrophoresis* **2007**, *28*, 4192-4201.
22. Erny, G. L.; Leon, C.; Marina, M. L.; Cifuentes, A. Time of flight versus ion trap mass spectrometry coupled to capillary electrophoresis to analyse intact proteins. *J. Sep. Sci.* **2008**, *31*, 1810-1818.
23. Anderson, T. J.; Lamsal, B. P. Development of new method for extraction of α-zein from corn gluten meal using different solvents. *Cereal Chem.* **2011**, *88*, 356-362.
24. Anderson, T. J.; Ilankovan, P.; Lamsal, B. P. Two fraction extraction of alpha-zein from DDGS and its characterization. *Ind. Crops Prod.* **2012**, *37*, 466-472.
25. Adams, W. R.; Huang, S. S.; Kriz, A. L.; Luethy, M. H. Matrix-assisted laser desorption ionization time-of-flight mass spectrometry analysis of zeins in mature maize kernels. *J. Agric. Food Chem.* **2004**, *52*, 1842-1849.
26. Wang, J. F.; Geil, P. H.; Kolling, D. R. J.; Padua, G. W. Analysis of zein by matrix-assisted laser desorption/ionization mass spectrometry. *J. Agric. Food Chem.* **2003**, *51*, 5849-5854.
27. Guo, Y.; Liu, Z.; An, H.; Li, M.; Hu, J. Nano-structure and properties of maize zein studied by atomic force microscopy. *J. Cereal Sci.* **2005**, *41*, 277-281.
28. Pedersen, K.; Devereux, J.; Wilson, D. R.; Sheldon, E.; Larkins, B. A. Cloning and sequence analysis reveal structural variation among related zein genes in maize. *Cell* **1982**, *29*, 1015-1026.
29. Woo, Y. M.; Hu, D. W. N.; Larkins, B. A.; Jung, R. Genomics analysis of genes expressed in maize endosperm identifies novel seed proteins and clarifies patterns of zein gene expression. *Plant Cell* **2001**, *13*, 2297-2317.
30. Alexandrov, N. N.; Brover, V. V.; Freidin, S.; Troukhan, M. E.; Tatarinova, T. V.; Zhang, H.; Swaller, T. J.; Lu, Y.-P.; Bouck, J.; Flavell, R. B.; Feldmann, K. A. Insights into corn genes derived from large-scale cDNA sequencing. *Plant Mol. Biol.* **2009**, *69*, 179-194.
31. Soderlund, C.; Descour, A.; Kudrna, D.; Bomhoff, M.; Boyd, L.; Currie, J.; Angelova, A.; Collura, K.; Wissotski, M.; Ashley, E.; Morrow, D.; Fernandes, J.; Walbot, V.; Yu, Y. Sequencing, mapping, and analysis of 27,455 maize full-length cDNAs. *Plos Genetics* **2009**, *5*.
32. Geraghty, D.; Peifer, M. A.; Rubenstein, I.; Messing, J. The primary structure of a plant storage protein: Zein. *Nucleic Acids Res.* **1981**, *9*, 5163-5174.
33. Marks, M. D.; Lindell, J. S.; Larkins, B. A. Nucleotide sequence analysis of zein messenger RNAs from maize endosperm. *J. Biol. Chem.* **1985**, *260*, 6451-6459.
34. Kriz, A. L.; Boston, R. S.; Larkins, B. A. Structural and transcriptional analysis of DNA sequences flanking genes that encode 19 kDa zeins. *Mol. Gen. Genet.* **1987**, *207*, 90-98.
35. Marks, M. D.; Larkins, B. A. Analysis of sequence microheterogeneity among zein messenger RNAs. *Journal of Biological Chemistry* **1982**, *257*, 9976-9983.
36. Kim, C. S.; Hunter, B. G.; Kraft, J.; Boston, R. S.; Yans, S.; Jung, R.; Larkins, B. A. A defective signal peptide in a 19-kD alpha-zein protein causes the unfolded protein response and an opaque endosperm phenotype in the maize De*-B30 mutant. *Plant Physiol.* **2004**, *134*, 380-387.

37. Viotti, A.; Cairo, G.; Vitale, A.; Sala, E. Each zein gene class can produce polypeptides of different sizes. *EMBO J.* **1985**, *4*, 1103-1110. (38) Geraghty, D. E.; Messing, J.; Rubenstein, I., Sequence analysis and comparison of CDNAs of the zein multigene family. *EMBO J.* **1982**, *1*, 1329-1335.
39. Song, R. T.; Llaca, V.; Linton, E.; Messing, J. Sequence, regulation, and evolution of the maize 22-kD alpha zein in gene family. *Genome Res.* **2001**, *11*, 1817-1825.
40. Thompson, G. A.; Siemieniak, D. R.; Sieu, L. C.; Slightom, J. L.; Larkins, B. A. Sequence analysis of linked maize 22 kDa alpha-zein genes. *Plant Mol. Biol.* **1992**, *18*, 827-833.
41. Liu, C. N.; Rubenstein, I. Molecular characterization of two types of 22-kDa alpha-zein genes in a gene cluster in maize. *Mol. Gen. Genet.* **1992**, *234*, 244-253.
42. Spena, A.; Viotti, A.; Pirrotta, V. A homologous repetitive block structure underlies the heterogeneity of heavy and light chain zein genes. *EMBO J.* **1982**, *1*, 1589-1594.
43. Hu, N. T.; Peifer, M. A.; Heidecker, G.; Messing, J.; Rubenstein, I. Primary structure of a genomic zein sequence of maize. *EMBO J.* **1982**, *1*, 1337-1342.
44. Coleman, C. E.; Lopes, M. A.; Gillikin, J. W.; Boston, R. S.; Larkins, B. A. A defective signal peptide in the maize high lysine mutant Floury-2. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 6828-6831.
45. Tong, H.; Bell, D.; Tabei, K.; Siegel, M. M. Automated data massaging, interpretation, and E-mailing modules for high throughput open access mass spectrometry. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 1174-1187.
46. Xu, W.; Reddy, N.; Yang, Y. An acidic method of zein extraction from DDGS. *J. Agric. Food Chem.* **2007**, *55*, 6279-6284.
47. Lawton, J. W. Isolation of zein using 100% ethanol. *Cereal Chem.* **2006**, *83*, 565-568.

Table 3-1

A Comparison of Molecular Weights of α-Zein Proteins Measured via High-Resolution Mass Spectrometry to Protein Database Search Results

Molecular weight (obs., HRMS)	Molecular weight (GenBank)	Difference	CGM samples	DDGS samples	Literature
23231.84	23230.87	0.97	8/10	6/6	(20, 26, 27)
23327.99	23324.99	3.00	0	6	(28)
23344.93	23346.09	-1.17	9	6	(19, 26, 29)
23358.33	23358.96	-0.63	9	6	(20, 23, 25, 26)
23361.24	23360.04	1.20	9	6	(19, 20, 24, 26, 30)
23366.54	23367.03	-0.49	3	2	(20, 26, 30)
23420.76	23419.06	1.70	9	6	(40)
23493.78	23496.30	-2.52	0	6	(31, 32)
23527.20	23527.33	-0.12	0	6	(36)
23568.96	23567.33	1.64	0	1	direct submission
24020.97	24019.79	1.18	10	6	(33)
24032.98	24036.82	-3.84	10	5	(34)
24077.73	24076.84	0.89	0	4	(35)
24087.90	24086.86	1.04	10	6	(19, 20, 23, 25, 26)
24137.61	24135.91	1.70	10	6	(20, 26)
24326.40	24325.13	1.27	1	0	direct submission
24375.20	24377.15	-1.95	1	0	direct submission
24424.43	24423.12	1.31	10	6	(19, 20, 26)
24706.84	24705.60	1.24	1	2	(23, 25)
26334.03	26336.59	-2.55	7	1	(41)
26359.84	26358.52	1.32	6	4	(20, 23, 25, 26)
26590.24	26586.87	3.37	10	1	direct submission
26708.58	26709.86	-1.28	8	0	(25)
26752.24*	26750.98	1.26	10	6	(23, 25, 36)
26752.24*	26752.89	-0.65	10	6	(37)
26761.12	26760.01	1.11	10	6	(19, 20, 36)
26779.45	26779.04	0.42	2	5	(26, 38)
26820.73	26819.12	1.61	10	6	(20, 26, 32, 36)
26837.16	26838.17	-1.01	10	6	(24, 27)
26859.75	26860.00	-0.25	4	0	(34, 39)
26891.66	26891.06	0.60	6	2	(36)
26906.45	26905.08	1.37	0	2	(36, 38)
26924.46	26923.17	1.30	10	5	(19, 20, 25)
27129.02	27127.53	1.49	9	5	(19, 20, 40)

*Observed protein can be classified to two GenBank accessions

Italicized proteins are reported for the first time in this study by mass spectrometry

Table 3-2

A Comparison of Molecular Weights of α-Zein Proteins Measured via High-Resolution Mass Spectrometry to Database Proteins Previously Identified via Mass Spectrometry

Molecular weight (obs., HRMS)	Molecular weight (GenBank)	Molecular weight (reported)	Analysis type	CGM samples	DDGS samples	Literature
23231.84	23230.87	23232	CE-MS	8/10	6/6	(20)
23358.33	23358.96	23318/23358.5	MALDI-MS/CE-MS	9	6	(20, 23)
23361.24	23360.04	23362/23361	MADLI-MS/ CE-MS	9	6	(20, 24)
23366.54	23367.03	23365.3	CE-MS	3	2	(20)
24087.90	24086.86	24069/24085.7	MALDI-MS/CE-MS	10	6	(20, 23)
24137.61	24135.91	24137.3	CE-MS	10	6	(20)
24424.43	24423.12	24425/24426 (±4)	CE-MS/ CE-MS	10	6	(19, 20)
24706.84	24705.60	24515	MALDI-MS	1	2	(23)
26359.84	26358.52	26308	MALDI-MS	6	4	(23)
26752.24	26750.98	26741	MALDI-MS	10	6	(23)
26761.12	26760.01	26758.8	CE-MS	10	6	(20)
26820.73	26819.12	26817	CE-MS	10	6	(20)
26837.16	26838.17	26838	MALDI-MS	10	6	(24)
26924.46	26923.17	26922.8/26925 (±1)	CE-MS/CE-MS	10	5	(19, 20)
27129.02	27127.53	27128.6	CE-MS	9	5	(19, 20)

Table 3-3

A Comparison of Molecular Weights of α-Zein Proteins Measure via High-Resolution Mass Spectrometry to Non-Database Proteins Previously Identified via Mass Spectrometry

Molecular weight (obs., HRMS)	Molecular weight (reported)	Analysis type	CGM samples	DDGS samples	Literature
23217.76	23216.3	CE-MS	10/10	0/6	(20)
23141.98	23140.2	CE-MS	10	6	(20)
23379.00	23377	MALDI-MS	10	6	(24)
23404.27	23401 (±1)	CE-MS	2	5	(19)
23437.09	23436.4	CE-MS	9	6	(20)
23995.94	23993.6	CE-MS	10	6	(20)
24071.90	24069	MALDI-MS	10	6	(23)
24096.55	24097	MALDI-MS	1	4	(24)
24559.48	24558.8	CE-MS	9	6	(20)
24694.62	24695 (±1)	CE-MS	0	4	(19)
26386.02	26383.2	CE-MS	0	2	(20)
26632.19	26630.6	CE-MS	9	4	(20)
26811.46	26813 (±4)	CE-MS	9	5	(19)
26828.15	26831.1	CE-MS	9	3	(20)
27186.14	27185.5	CE-MS	1	0	(20)

Table 3-4

Molecular Weights of α-Zein Proteins Which Met Deconvolution Criteria, Not Previously Reported Via Mass Spectrometry Analysis and No Database Entry

Molecular weight (obs., HRMS)	CGM samples	DDGS samples	Molecular weight (obs., HRMS)	CGM samples	DDGS samples
23241.83	1/10	0/6	24447.34	10/10	6/6
23308.89	9	4	24455.85	10	2
23331.97	0	2	24500.48	8	2
23394.45	0	2	24523.26	0	2
23407.36	0	5	24531.5	7	0
23442.62	6	6	24635.5	1	0
23455.25	2	6	26257.69	8	5
23458.20	10	6	26375.94	1	1
23470.26	1	0	26419.76	1	1
23476.11	0	6	26435.98	7	2
23510.95	0	6	26514.26	10	3
23519.50	10	0	26529.8	3	0
23523.17	0	6	26655.09	4	4
23554.11	0	3	26734.28	0	3
23599.38	2	4	26776.27	2	4
24010.05	10	4	26796.51	0	1
24108.57	8	6	26855.12	0	1
24148.26	2	0	26878.48	0	1
24155.42	5	6	26915.7	3	0
24164.22	10	6	26942.24	1	2
24195.96	0	2	26973.8	1	0
24213.00	10	6	27016.8	1	0
24240.38	1	0	27205.19	8	1
24441.77	0	1			

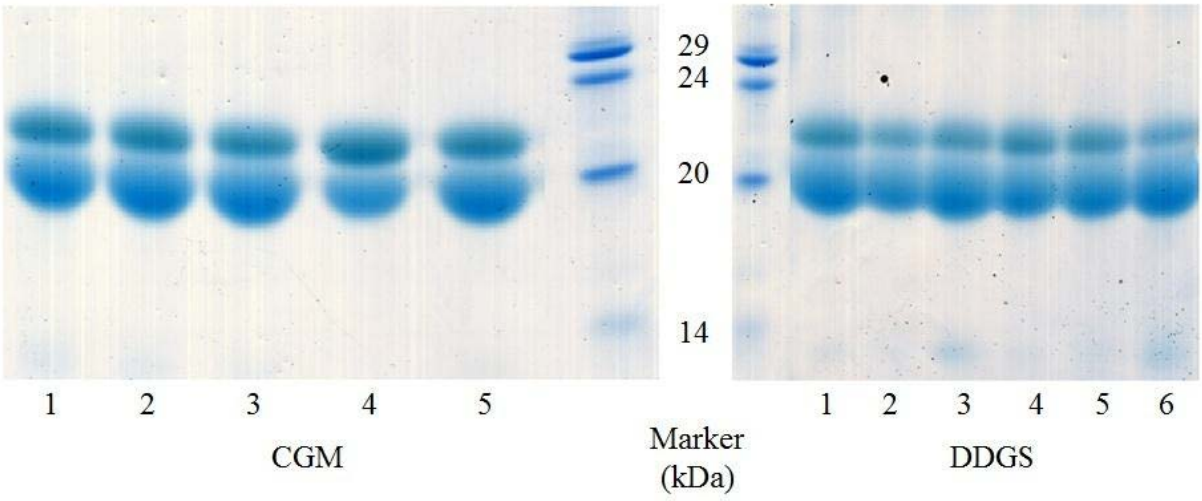


Figure 3-1. SDS-PAGE of α-zein protein extracts from CGM and DDGS. For the five CGM samples, all were extracted without sodium hydroxide and sodium bisulfite. A similar gel was observed for the five identical extractions which used sodium hydroxide and sodium bisulfite. CGM lane 1 was extracted with 88% aqueous 2-propanol, lane 2 with 70% aqueous 2-propanol, lane 3 with 55% aqueous 2-propanol, lane 4 with 70% 2-propanol,22.5% glycerol, and 7.5% water, and lane 5 with 70% aqueous ethanol. For the six DDGS samples, lanes 1-3 were not pretreated with enzyme while lanes 4-6 were pretreated. Lanes 1 and 4 were extracted with 88% aqueous 2-propanol, lanes 2 and 5 were extracted with 70% aqueous 2-propanol, and lanes 3 and 6 were extracted with 70% aqueous ethanol. CGM gel adapted from Anderson and Lamsal (21) and DDGS gel adapted from Anderson et al. (22)

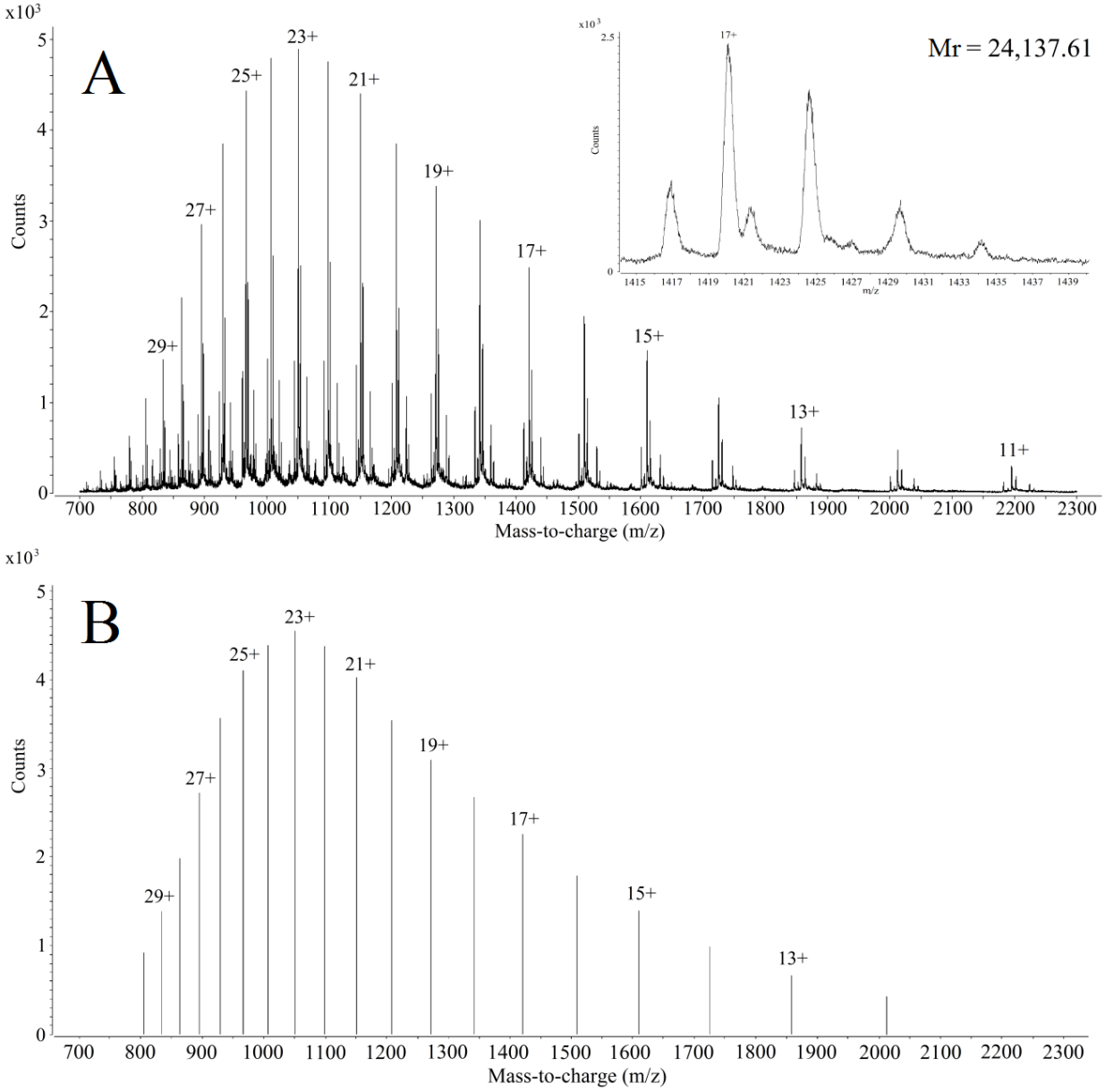


Figure 3-2. Mass spectrum of a mixture of α-zein proteins. Peaks for the protein with Mr = 24,137.61 Da are labeled. Spectrum A is raw data integration of total ion chromatograph from 41.6 to 42.7 minutes. A zoomed in view of the 17+ peak is displayed in the inset. The peak labeled 17+ belongs to the charge state progression for the protein shown in the deconvoluted spectrum B. The other peaks in the inset belong to charge state progressions from other proteins.

CHAPTER 4

ANALYSIS OF RESISTANT STARCHES IN RAT CECAL CONTENTS USING FOURIER TRANSFORM INFRARED PHOTOACOUSTIC SPECTROSCOPY

Modified from a paper published in *Journal of Agricultural and Food Chemistry*[1]

By

Timothy J. Anderson,[2,3] Yongfeng Ai,[4] Roger W. Jones,[2] Robert S. Houk,[2,3] Jay-lin Jane,[4] Yinsheng Zhao,[4] Diane F. Birt,[4] and John F. McClelland[2,5*]

[1]Reprinted with permission of *Journal of Agricultural and Food Chemistry 61(8):1818–1822*

[2]Ames Laboratory-USDOE, Iowa State University, Ames, Iowa 50011

[3]Department of Chemistry, Iowa State University, Ames, Iowa 50011

[4]Department of Food Science and Human Nutrition, Iowa State University, Ames, Iowa 50011

[5]Department of Mechanical Engineering, Iowa State University, Ames, Iowa 50011

*Corresponding Author, Email johnfm@iastate.edu, Phone 515-294-7948, Fax 515-294-4748

Abstract

Fourier transform infrared photoacoustic spectroscopy (FTIR-PAS) qualitatively and quantitatively measured resistant starch (RS) in rat cecal contents. Fisher 344 rats were fed diets of 55 % (w/w, dry basis) starch for eight weeks. Cecal contents were collected from sacrificed rats. A corn starch control was compared against three RS diets. The RS diets were high-amylose corn starch (HA7), HA7 chemically modified with octenyl succinic anhydride, and stearic-acid-complexed HA7 starch. To calibrate the FTIR-PAS analysis, samples from each diet were analyzed using an enzymatic assay. A partial least squares cross-validation plot generated from the enzymatic assay and FTIR-PAS spectral results for starch fit the ideal curve with an R^2 of 0.997. A principal component analysis plot of components 1 and 2 showed that spectra from diets clustered significantly from each other. This study clearly showed that FTIR-PAS can accurately quantify starch content and identify the form or starch in complex matrices.

Keywords Resistant starch, fecal analysis, Fourier transform infrared photoacoustic spectroscopy, partial least squares, principal component analysis

Introduction

Foodstuffs contain a varied mixture of complex compounds and materials. One of these compounds, starch, has been characterized and studied for decades. Starch is commonly found in many foods as starch granules, which are a combination of amylose and amylopectin. The ratio of these two compounds varies with source material, but the amount of amylose in the normal starch granule is typically 15-30%, while amylopectin can reach 70% or 100% for waxy starch (1-3). Amylose can form single helical complexes with many chemicals, such as free fatty acids and iodine, and it can also form double helices (3). Recently, resistant starches containing high concentrations of amylose (up to 85%, high-amylose corn starch) and those with chemical modifications have increasingly been investigated (3). These starches have been dubbed resistant starch (RS), due to the fact that they resist degradation and absorption in the small intestine (4).

The decreased digestibility of RS has garnered attention from researchers who study diabetes (5). With easily digestible starch, diabetics have difficulty controlling their blood glucose levels, but RS may impart many beneficial effects for diabetics through reduction in blood-glucose spikes (5-8). Another benefit of RS is the potential to control energy intake. Many researchers are attempting to find foods that digest slowly and decrease energy intake, which could help with weight maintenance (9). RS can also play a role as a prebiotic. Prebiotics encompass many of the dietary fibers, including RS, which are not readily digestible by humans. The undigested RS can be utilized by microbes within the gut and may release beneficial compounds for the host organism (10-13).

There are five varieties of RS. Type 1 RS can be found in coarsely ground legumes or whole grain. The cell wall surrounding the Type 1 RS makes the starch physically inaccessible

to digestion. Type 2 RS is individual C-type or B-type crystalline starch granules. Type 2 RS typically is raw banana and potato starch, and high amylose corn starch that retains the crystalline structure. Type 3 RS refers to retrograded amylose (14). Type 4 RS is chemically modified starch (10, 15). The latest RS is Type 5 RS, which is an amylose-lipid complex (16).

Animal studies are often performed to evaluate RS digestibility from analysis of fecal samples, which are complex materials containing protein, carbohydrate, and lipid. Common quantitative methods for analyzing starch content are starch-hydrolysis enzyme assays. The enzyme assays are useful for starch quantification, but have many negative aspects for fecal studies. The enzyme assays cost approximately $3 per sample, take about 20 minutes per sample, and consume at least 0.2 g dry sample for analyses in duplicate. Studies with mice or rats tend to produce large numbers of small samples, which may become time consuming and costly when hundreds of samples need to be analyzed.

Our alternative proposed method of analysis is Fourier transform infrared photoacoustic spectroscopy (FTIR-PAS). Conventional FTIR relies on transmission of IR light through the sample to measure the absorption bands of the compounds of interest. Conventional FTIR does not work well with many food products due to their opaque nature, light scattering properties, and difficulties with sample preparation (17-19). Alternatively, FTIR-PAS directly measures the IR absorbance spectrum of opaque samples, needs minimal sample preparation, and is fast and nondestructive (20). FTIR-PAS uses a PAS accessory, which has a sample cell with a window to allow a modulated IR beam from the spectrometer to enter and illuminate the sample (21). The IR light absorbed by the sample heats it. The heat migrates to the gas/sample interface and produces a pressure wave in proportion to the absorbance by the sample. The resultant pressure signal is then picked up by a sensitive microphone, and the signal is converted into a

wavenumber versus absorbance intensity spectrum (21). For further information pertaining to FTIR-PAS theory or explanation of various experimental methods, please see References 22 and 23.

A handful of studies have successfully analyzed starch and other food-based components using FTIR-PAS. One of the first food analyses utilized IR-PAS with a near infrared monochromator to determine the moisture content of protein powders (24). Later researchers applied FTIR-PAS to analyze protein and carbohydrate, but lacked statistical power to quantify the data (25). It was not until the mid-1990's that FTIR techniques with food materials began to couple spectral results with statistical techniques such as partial least squares (PLS) (26). PLS uses a small training set of samples analyzed via a non-FTIR standard method to calibrate the FTIR analysis. A multivariate model of the spectral data with the quantitative values can be produced to create a calibration to predict the composition of unknown samples from their spectra. This approach has been confirmed for determining lipid, protein, and carbohydrate concentrations in pea seeds (27).

The present study went beyond food and single identity starch analysis by quantifying modified starch in rat cecal contents. Three types of RS were studied along with a control corn starch. The first RS studied was high amylose corn starch (HA7), a Type 2 RS. The second RS was octenyl succinic high-amylose corn starch (OS-HA7) which is a Type 4 RS. OS-HA7 is obtained from modifying starch with octenyl succinic anhydride, which forms ester bonds with hydroxyl groups of starch molecules. The third RS was high amylose corn starch complexed with stearic acid (RS5-HA7), a Type 5 RS. RS5-HA7 is based on a physical complex between amylose and stearic acid rather than chemical bonds.

The main goal of this study was to determine if FTIR-PAS and PLS was a practical alternative to the enzymatic assay for starch content. To achieve this, we needed to determine whether FTIR-PAS analysis could produce a linear correlation while accounting for potential interferences from the complex sample matrix and RS modification. Beyond a quantitative fit of the starch, principal component analysis (PCA) was tested to determine if the different diets could be differentiated qualitatively.

Materials and Methods

Rat Animal Study

Fischer 344 rats were housed following the procedure of Zhao et al (28). The animals were on the feeding regimen for eight weeks before the animals were sacrificed. The trial contained 90 rats total (2 rats died before sacrifice), which were randomly assigned to four diet groups. The four diets consisted of the control (corn starch), HA7, OS-HA7, and RS5-HA7 diets described below. For purposes important to other companion studies based on this same diet trial, the control and RS5-HA7 diet groups each contained 29 rats and were broken down further into four subgroups per diet. The rats were given two injections of either saline or the carcinogen azoxymethane (AOM, Midwest Research Institute, Kansas City, MO) administered following the method of Zhao et al (28)., and some were fed an antibiotic treatment mixture of vancomycin and imipenem. The treatments resulted in the four subgroups within the control and RS5-HA7 diets that consisted of rats given both AOM and antibiotic, AOM and no antibiotic, saline and antibiotic, and only saline. Both HA7 and OS-HA7 diets contained 15 rats per diet group and were divided into only two subgroups. They were given either AOM and no antibiotic or neither. For purposes of the tests reported here, we have grouped samples only according to

diet and not according to AOM or antibiotic treatment. The animal studies were performed in compliance with the guidelines of The Institutional Animal Care and Use Committee of Iowa State University.

Starch Diets Fed to Rats

Four starch varieties were utilized for the feeding study: Control (corn starch, Cargill Gel 03420; Cargill Inc., Minneapolis, MN), HA7 (AmyloGel 03003; Cargill Inc.), OS-HA7 (processed HA7 bound to octenyl succinate in the Department of Food Science and Human Nutrition, Iowa State University), and RS5-HA7 (processed using HA7 and stearic acid in the Department of Food Science and Human Nutrition, Iowa State University) (16, 29). The starches were cooked before being added to the diets following the procedure of Zhao et al (28). The cooked starch was then added to a diet formulated on the basis of the standard diet recommended by the American Society for Nutritional Sciences for mature rats (AIN-93M) (30). Starch diets were prepared every other day and served fresh to the rats.

Rat Cecal Samples

This study collected only the rat ceca and placed the contents into Corning 15-mL centrifuge tubes (Tewksbury, Massachusetts) on dry ice before storage at -80 °C. Due to two other companion studies obtaining samples prior to this experiment, much of the cecal contents from the samples was exhausted. Adequate material from only twenty-eight samples, seven from each of the four feeding groups, could be randomly obtained. The wet cecal samples were placed in aluminum weighing pans and dried in an oven at 105 °C for three hours. After drying, the cecal material formed dry wafers, which were ground using mortar and pestle. The ground

cecal material was then placed in 1.7-mL microcentrifuge tubes purchased from Marsh Bio Products (Rochester, New York) and stored sealed at room temperature prior to analysis.

Enzymatic Assay for Starch Content

The starch content of the cecal materials was measured using Total Starch Assay Kit (Megazyme International Ireland Ltd., Co.,Wicklow, Ireland) following AACC Method 76-13 (31).

FTIR-PAS

The FTIR-PAS analysis was performed using an MTEC Photoacoustics PAC300 detector mounted in a Digilab FTS 7000 FTIR spectrometer. The sample detector has a 1-cm interior diameter and a window at the top for the infrared beam to enter the chamber and illuminate the sample. The dried and ground cecal material was placed in a disposable aluminum cup, which was fully illuminated by the infrared beam. Immediately before analysis the detector was purged with helium gas to remove atmospheric water vapor and carbon dioxide, which have strong mid-infrared absorptions. Also, a desiccant, magnesium perchlorate, was added beneath the sample to remove any moisture that might evolve from the sample during analysis. Spectra were taken at 8 cm^{-1} resolution and a 2.5 kHz scan speed, with the co-addition of 256 scans.

PLS and PCA

The spectra were correlated with starch levels determined by the enzymatic assay via PLS using commercial software (Thermo Galactic GRAMS/AI PLSplus IQ, Version 5.1) (32-34). PLS utilizes a training set of spectra from samples whose relevant properties are known and span

the range of interest. In the present case, the enzymatic assay provided the known property values. PLS modeling determines a small set of basis-vector spectra, called factors, by which it can describe all of the training set spectra. Each training set spectrum is then just a weighted sum of the factors. The factors with the smallest weightings consist mostly of noise and are dropped from the model. PLS then performs a multiple linear regression correlating the factor weightings with the known values of the property being predicted. Once the PLS model is built, the correlated property can be determined for unknown samples directly from the model, as long as the properties of the unknowns fall within the range of those covered by the original training set.

Because the starch level was determined for only 28 samples (seven per diet), the sample set was not split into separate training and validation sets. Instead, all of the samples were used in creating the PLS model, and a single-elimination cross validation was used to measure model quality. In such a cross validation, one member of the training set is removed, and a model is built from the remaining members. The removed spectrum is then analyzed as an unknown. The removed spectrum is returned to the training set, and then a different one is removed and the process is repeated. This is done until all training set members have been removed and analyzed as unknowns. Plots comparing the known values and the predicted values from the cross validations are included in Results and Discussion. The standard error of cross-validation (SECV) is a measure of model quality. It is the root-mean-square difference between the values of the predicted property determined during the cross-validation and their known values.

The model with the lowest prediction residual error sum of squares (PRESS) value was selected as the most accurate model. PRESS is given by

$$PRESS = \sum_{i=1}^{N} (k_i - p_i)^2$$

where k_i and p_i are the known and predicted values for the i^{th} sample, and there are N samples in the training set. In that most-accurate model, the 4000-397 cm^{-1} range of the spectra was used, the spectra were preprocessed using multiplicative scatter correction (MSC) (35) and by conversion to first derivatives (19-point Savitsky-Golay). The resulting model had ten factors.

Classification of the spectra according to diet was done using PCA (36-37). The same 4000-397 cm^{-1} range and the same first derivative and MSC preprocessing were applied to the data as in the PLS modeling. This was sufficient to cleanly separate the samples into clusters according to diet.

Results and Discussion

Enzymatic Assay for Starch Content

Starch contents of the cecal material from the rats fed different diets are shown in Table 5-1. The cecal content from the rats fed the OS-HA7 diet had the highest starch content, ranging from 47% to 50.1%, whereas that from the rats fed the control diet with normal corn starch had the lowest starch content, ranging from 0.3% to 1.1%. There was no significant difference among the food disappearance (used to estimate intake but includes loses) of the rats fed the different diets (data not shown). These results suggest that the OS-HA7 has the highest resistance to *in vivo* digestion, followed by RS5-HA7, HA7 and normal corn starch.

FTIR-PAS

The FTIR-PAS data were measured from 4000 to 397 cm^{-1}. Spectra from all four diets are shown in Figure 4-1. All samples show many bands in common, but in the fingerprint region (1800 to 397 cm^{-1}), there are visible differences among the cecal samples from different diets. A study by Irudayaraj and Yang using FTIR-PAS identified bands in pure starch and protein spectra (38). However, due to the complexity of the cecal samples, the present spectra have substantial peak overlap, so manual interpretation is not sufficient. The use of chemometric software can analyze the data and draw out the quantitative and qualitative data needed.

PLS was successfully used for starch to model the relation between the enzymatic assay results and the FTIR-PAS spectra of the rat cecal contents. Figure 4-2 shows the cross validation for the best fitting model. The plot correlates the known starch content (dry basis) with the starch content predicted by the PLS model. The diagonal line is the ideal (i.e., predicted = known). The SECV is 1.055 wt. % and R^2 is 0.997. The SECV is only 2% of the starch-content range in the sample set (0.3 to 50.1 wt. %), so the predictions are of good quality. The high quality of the predictions from the training set would allow unknown samples to be quantitatively analyzed for starch content using the chemometric model developed. Also since the model was able to accurately fit every modified starch, the model should be useful for any of the four starch diets utilized.

Besides the quantitative starch information, qualitative information to identify which starch was measured is very useful. The spectral data was analyzed by PCA to aid in sample identification. The first two principal components from the PCA of the spectra cleanly separated the samples according to diet, as shown in Figure 4-3. These two components account for 83.5%

of the variance in the data. The PCA analysis gives a simple and clearly visible means to match the cecal samples to the corresponding starch diets.

Despite the similarity of the measured spectra for the different cecal materials, chemometric analysis produced a successful model of the data. The FTIR-PAS data coupled with the enzyme starch assay results clearly were able to produce a cross-validation plot that gave high-quality quantitative results. The first two principal component scores also were able to show clustering that would allow qualitative identification of starch in future unknown cecal samples. No clustering among the antibiotic or AOM subgroup treatments was observed using the PCA components. This finding should give credence to the robustness of FTIR-PAS to see through minor effects even within complex matrix materials.

This study was proof of concept for FTIR-PAS analysis of starch to replace future high volume enzymatic assay analysis. Future work will incorporate timed fecal collections and FTIR-PAS starch analysis and metabolic analysis. This analysis would be used to track how the chemistry of the gut microbiome changes as the animal adapts over time to an RS diet.

Abbreviations

Resistant starch (RS), Fourier transform infrared photoacoustic spectroscopy (FTIR-PAS), partial least squares (PLS), high-amylose corn starch (HA7), octenyl succinic high-amylose corn starch (OS-HA7), high amylose corn starch complexed with stearic acid (RS5-HA7), principal component analysis (PCA), prediction residual error sum of squares (PRESS), multiplicative scatter correction (MSC), standard error of cross validation (SECV).

Funding Sources

This project was supported by the Iowa State University (ISU) Plant Sciences Institute and was supported in part by the Department of Agriculture, CSREES award number 2009-65503-05798.

Conflict of Interest Statement

John McClelland has a financial interest in MTEC Photoacoustics, Inc., the manufacturer of the photoacoustic detector used in this study.

Acknowledgement

This research was performed at the Ames Laboratory. Ames Laboratory is operated for the U.S. Department of Energy by Iowa State University under contract no. DE-AC02-07CH11358.

References

1. Sajilata, M. G.; Singhal, R. S.; Kulkarni, P. R. Resistant starch-A review. *Compr. Rev. Food Sci. Food Saf.* **2006**, *5*, 1-17.
2. Atkin, N. J.; Cheng, S. L.; Abeysekera, R. M.; Robards, A. W. Localisation of amylose and amylopectin in starch granules using enzyme-gold labelling. *Starch-Stärke* **1999**, *51*, 163-172.
3. Jane, J. Structural Features of Starch Granules II. In *Starch: Chemistry and Technology*; 3rd ed.; BeMiller, J. N.; Whistler, R. L., Eds.; Academic Press Inc.: Orlando, FL, **2009**; pp. 193-236.
4. Champ, M.; Langkilde, A.-M.; Brouns, F.; Kettlitz, B.; Le Bail-Collet, Y. Advances in dietary fibre characterisation. 2. Consumption, chemistry, physiology and measurement of resistant starch; implications for health and food labeling. *Nutr. Res. Rev.* **2003**, *16*, 143-161.
5. Robertson, M. D. Dietary-resistant starch and glucose metabolism. *Curr. Opin. Clin. Nutr. Metab. Care* **2012**, *15*, 362-367.
6. Saltiel, A. R.; Kahn, C. R. Insulin signalling and the regulation of glucose and lipid metabolism. *Nature* **2001**, *414*, 799-806.
7. Ranganathan, S.; Champ, M.; Pechard, C.; Blanchard, P.; Nguyen, M.; Colonna, P.; Krempf, M. Comparative study of the acute effects of resistant starch and dietary fibers on metabolic indexes in men. *Am. J. Clin. Nutr.* **1994**, *59*, 879-883.
8. Eckel, R. H.; Grundy, S. M.; Zimmet, P. Z. The metabolic syndrome. *Lancet* **2005**, *365*, 1415-1428.
9. Bodinham, C. L.; Frost, G. S.; Robertson, M. D. Acute ingestion of resistant starch reduces food intake in healthy adults. *Brit. J. Nutr.* **2010**, *103*, 917-922.
10. Brown, I. L.; Wang, X.; Topping, D. L.; Playne; M. J.; Conway, P. L. High amylose maize starch as a versatile prebiotic for use with probiotic bacteria. *Food Aust.* **1998**, *50*, 603-610.
11. Marotti, I.; Bregola, V.; Aloisio, I.; Di Gioia, D.; Bosi, S.; Di Silvestro, R.; Quinn, R.; Dinelli, G. Prebiotic effect of soluble fibres from modern and old durum-type wheat varieties on *Lactobacillus* and *Bifidobacterium* strains. *J. Sci. Food Agric.* **2012**, *92*, 2133-2140.
12. Kritchevsky, D. Epidemiology of fibre, resistant starch and colorectal cancer. *Eur. J. Cancer Prev.* **1995**, *4*, 345-352.
13. van Munster, I. P.; Tangerman, A.; Nagengast, F. M. Effect of resistant starch on colonic fermentation, bile acid metabolism, and mucosal proliferation. *Digest. Dis. Sci.* **1994**, *39*, 834-842.
14. Englyst, H. N.; Kingman, S. M.; Cummings, J. H. Classification and measurement of nutritionally important starch fractions. *Eur. J. Clin. Nutr.* **1992**, *46* (Suppl. 2), S33-S50.
15. Baghurst, P. A.; Baghurst, K. I.; Record, S. J. Dietary fibre, non-starch polysaccharides and resistant starch-a review. *Food Aust.* **1996**, *48* (Suppl.), S3-S35.
16. Hasjim, J.; Lee, S.-O.; Hendrich, S.; Setiawan, S.; Ai, Y.; Jane, J. Characterization of novel resistant-starch and its effects on postprandial plasma-glucose and insulin responses. *Cereal Chem.* **2010**, *87*, 257-262.
17. Colthup, N. B.; Daly, L. H.; Wiberley, S. E. *Introduction to Infrared and Raman Spectroscopy*, 3rd ed.; Academic Press: New York, US, **1990**.

18. *Laboratory Methods in Infrared Spectroscopy*, 2nd ed.; Miller, R. G. J., Stace, B. C., Eds.; Heyden and Sons: London, UK, **1979**.
19. *Practical FT-IR Spectroscopy: Industrial and Laboratory Chemical Analysis*, Ferraro, J. R.; Krishnan, K., Eds.; Academic Press: New York, US, **1990**.
20. Graham, J. A.; Grim, W. M., III; Fateley, W. G. Fourier transform infrared photoacoustic spectroscopy of condensed-phase samples. In *Fourier Transform Infrared Spectroscopy,* Ferraro, J. R., Basile, L. J., Eds.; Academic Press: Orlando, US, **1985**; Vol. 4; pp. 345-392.
21. McClelland, J. F.; Jones, R. W.; Luo, S.; Seaverson, L. M. A practical guide to FT-IR photoacoustic spectroscopy. In *Practical Sampling Techniques for Infrared Analysis*, Coleman, P. B., Ed.; CRC Press: Boca Raton, FL, **1993;** pp. 107-144.
22. McClelland, J. F.; Bajic, S. J.; Jones, R. W.; Seaverson, L. M. Photoacoustic spectroscopy. In *Modern Techniques in Applied Molecular Spectroscopy*, Mirabella, F. M., Ed.; John Wiley & Sons, Inc.: New York, US, **1998;** pp. 221-265.
23. McClelland, J. F.; Jones, R. W.; Bajic, S. J. Photoacoustic spectroscopy. In *Handbook of Vibrational Spectroscopy*, Chalmers, J. M.; Griffiths, P. R., Eds.; John Wiley & Sons Ltd.: Chichester, UK, **2002**; Vol. 2, pp. 1231-1251.
24. Castleden, S. L.; Kirkbright, G. F.; Menon, K. R. Determination of moisture in single-cell protein utilising photoacoustic spectroscopy in the near-infrared region. *Analyst* **1980**, *105*, 1076-1081.
25. Belton, P. S.; Saffa, A. M.; Wilson, R. H. Use of Fourier transform infrared spectroscopy for quantitative analysis: A comparative study of different detection methods. *Analyst* **1987**, *112*, 1117-1120.
26. McQueen, D. H.; Wilson, R.; Kinnunen, A. Near and mid-infrared photoacoustic analysis of principal components of foodstuffs. *TrAC-Trends Anal. Chem*. **1995**, *14*, 482-492.
27. Letzelter, N. S.; Wilson, R. H.; Jones, A. D.; Sinnaeve, G. Quantitative determination of the composition of individual pea seeds by Fourier transform infrared photoacoustic spectroscopy. *J. Sci. Food Agric*. **1995**, *67*, 239-245.
28. Zhao, Y.; Hasjim, J.; Li, L.; Jane, J.; Hendrich, S.; Birt, D. F. Inhibition of azoxymethane-induced preneoplastic lesions in the rat colon by a cooked stearic acid complexed high-amylose cornstarch. *J. Agric. Food Chem*. **2011**, *59*, 9700-9708.
29. Zhang, B.; Huang, Q.; Luo, F.; Fu, X.; Jiang, H.; Jane, J. Effects of octenylsuccinylation on the structure and properties of high-amylose maize starch. *Carbohyd. Polym*. **2011**, *84*, 1276-1281.
30. Reeves, P. G. Components of the AIN-93 diets as improvements in the AIN-76A diet. *J. Nutr*. **1997**, *127* (Suppl. 5), 838S-841S.
31. AACC International. *Approved Methods of the American Association of Cereal Chemists*, 10th Ed., Method 76-13. The Association: St. Paul, US. **2000**.
32. Haaland, D. M.; Thomas, E. V. Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. *Anal. Chem*. **1988**, *60*, 1193-1202.
33. Fredstrom, S. B.; Jung, H.-J. G.; Halgerson, J. L; Eyden, C. A.; Slavin, J. L. Trial of near-infrared reflectance spectroscopy in a human fiber digestibility study. *J. Agric. Food Chem*. **1994**, *42*, 735-738.

34. Franck, P. F.; Sallerin, J.-L.; Schroeder, H.; Gelot, M.-A.; Nabet, P. Rapid determination of fecal fat by Fourier transform infrared analysis (FTIR) with partial least-squares regression and an attenuated total reflectance accessory. *Clinical Chem.* **1996**, *42*, 2015-2020.
35. Geladi, P.; MacDougall, D.; Martens, H. Linearization and scatter-correction for near-infrared reflectance spectra of meat. *Appl. Spectrosc*. **1985**, *39*, 491-500.
36. Leardi, R. Chemometrics in data analysis. In *Food Authenticity and Traceability*, Lees, M., Ed.; Woodhead Publishing: Cambridge, UK, **2003;** pp. 299-320.
37. Roldán-Marín, E.; Jensen, R. I.; Krath, B. N.; Kristensen, M.; Poulsen, M.; Cano, M. P.; Sánchez-Moreno, C.; Dragsted, L. O. An onion byproduct affects plasma lipids in healthy rats. *J. Agric. Food Chem*. **2010**, *58*, 5308-5314.
38. Irudayaraj, J.; Yang, H. Depth profiling of a heterogeneous food-packaging model using step-scan Fourier transform infrared photoacoustic spectroscopy. *J. Food Eng*. **2002**, *55*, 25-33.

Table 4-1

Summary of enzymatic assay analysis of *in vivo* starch (dry basis) in cecal contents by rat

Diet	Rat#	Average Starch % (wt)	Standard Deviation
Control	87	0.3	0.2
	23	0.6	0.1
	27	0.8	0.0
	49	0.7	0.0
	84	0.8	0.0
	81	1.1	0.3
	70	1.0	0.1
	Average	0.7	
HA7	33	12.7	0.4
	40	19.4	0.5
	15	21.2	0.5
	9	21.3	0.5
	18	18.1	0.1
	22	20.5	0.1
	28	14.8	0.6
	Average	18.3	
OS-HA7	14	47.9	0.1
	16	47.9	0.1
	10	49.8	0.3
	44	47.1	0.4
	6	50.1	0.4
	50	50.0	0.1
	52	49.1	0.7
	Average	48.8	
RS5-HA7	12	24.0	0.1
	1	19.2	0.3
	20	17.1	0.6
	29	13.3	0.5
	43	28.1	0.2
	57	30.5	0.7
	64	21.5	0.4
	Average	21.9	

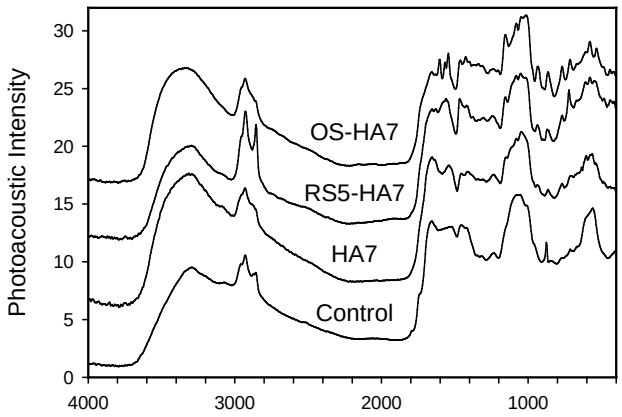


Figure 4-1. FTIR-PAS spectra collected from rat cecal contents. The spectra are from single representative rats from each of the diet groups. Spectra scaled and displaced vertically.

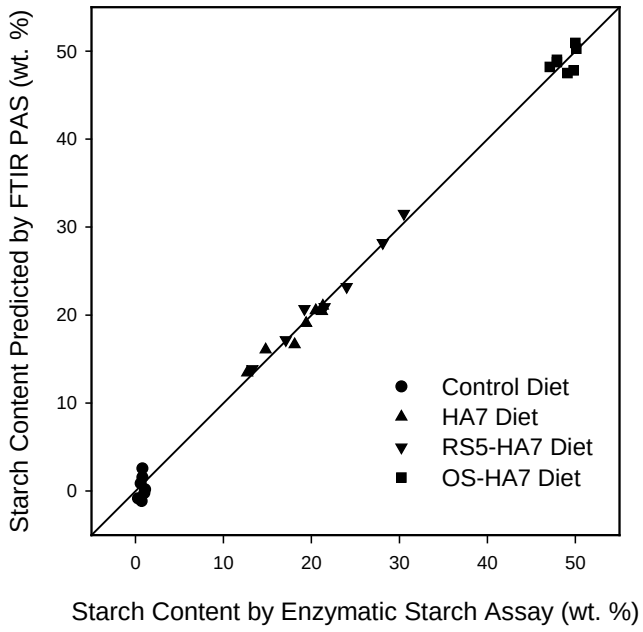


Figure 4-2. Plot of cross validation for starch content measured by enzymatic starch assay from each cecal sample (dry basis) from all four diets versus the predicted values by FTIR-PAS. The R^2 of the data to the ideal best fit line was 0.997.

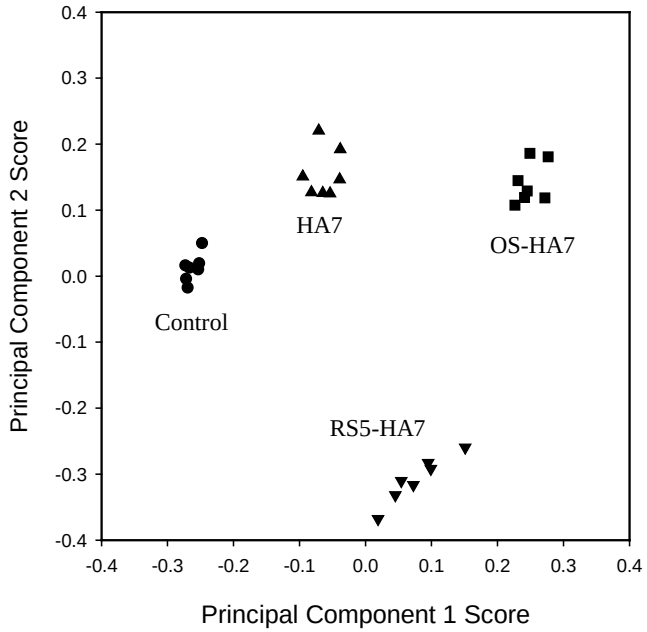


Figure 4-3. Scores for the first two principal components in the PCA modeling of the spectra of 28 dried cecal samples separate the samples according to the resistant-starch diets of the rats.

CHAPTER 5

General Conclusions

This dissertation concentrated on the use of mass spectrometry and FTIR-PAS for the analysis of metabolites, proteins, and starch from biological sources. Chapter 2 focused on the use of high resolution MS to determine if RS fed to rats had an effect on metabolites produced during digestion. It was startling to observe the sheer number of metabolites produced through direct infusion ESI MS. From the basic MS spectra that were obtained from the extracts from both cecal and distal colon contents, it was apparent the RS diets had an effect on the metabolites observed. The trend was that many of the low m/z ions for distal colon and cecal contents were similar despite diet, however many of the mid to high m/z ions were not shared. To determine if the apparent differences were statistically relevant PLS-DA was performed for cecal and distal colon content samples. The results showed that PLS-DA was able to accurately classify the cecal and distal colon content samples into distinct groups based upon the RS fed to the rats. The PLS-DA software was also able to determine biomarkers based upon accurate mass data. Future work would be to improve the fidelity of the study to include only RS parameters. The study included antibiotic and a carcinogen that potentially caused unnecessary metabolic variation among RS groups. Other interesting work would be to determine if metabolites shift over the time that an animal consumes the RS. By taking samples each week, rather than at the end of the study, these shifts would be apparent. Lastly, it would be paramount to attempt to accurately identify biomarker metabolites which consistently appear in multiple studies using LC-MS/MS in conjunction with standards.

Chapter 3 centered on the identification of zein proteins from extracts procured from CGM and DDGS using a variety of solvent systems and with or without the use of a reducing

agent. The zein extractions were tuned to extract primarily α-zein proteins. The use of high resolution MS was employed to find the accurate masses of the α-zein proteins in the extracts and to determine if the α-zein proteins varied based on extraction and source material. Previous MS studies had observed limited numbers of α-zein proteins, and generally had low resolution. The LC separation coupled to the high resolution MS discerned as many as 95 unique proteins, 49 of the proteins had similar M_r values to previously reported proteins. It was unexpected that 27 of the 49 proteins previously reported, were observed in both the CGM and DDGS extracts. The production of DDGS undergoes rigorous temperature and adulteration conditions, it was hypothesized that the zein extracted from DDGS many not contain fully intact zein proteins. Further work with α-zein extracts would include the full identification of the zein protein sequences using high resolution MS. The methods of choice would most likely include CID, ETD, or ECD. Such techniques would trap the charge state of choice for each zein protein of interest, fragment the protein, and then using software determine its protein sequence. These techniques would be amenable to protease digestion because most α-zein proteins would not be solubilized in aqueous solutions necessary for standard digestion conditions. Such digestions may be possible, but could be much more costly then seeking instrumental sequencing techniques.

Chapter 4 applied FTIR-PAS to the same cecal samples used in Chapter 2. Rather than study the metabolites in the rat cecal contents, the main focus was the starch content. Very few techniques have the ability to determine how much starch is in a sample, and to determine the starch type. FTIR-PAS was utilized successfully for qualitatively and quantitatively determining the amount and type of RS remaining in the cecum of the rat. FTIR-PAS was able to analyze the bulk sample with minimal sample prep, which included grinding and oven drying. The main

issue with FTIR-PAS was that it could not be utilized alone. A calibration had to be produced using approximately twenty cecal samples of known RS type, and measured for starch content using a bench-mark method. However, after calibrating for the twenty samples, it would be simple to measure the RS content and type for a near unlimited number of samples with minimal time and cost invested.

Mass spectrometry is a method which is nearly indispensable to biological sample analysis. With future advances in ionization, chromatography, sensitivity, and resolution it may be relatively simple in the future to understand the near complete profile of compounds within a complex biological analyte. However, parallel techniques such as FTIR can interrogate the bulk properties of these same complex biological analytes with minimal sample preparation. If future improvements to FTIR techniques can decrease the number of techniques and apply it to most sample types, it would be vastly more accessible. One technique can provide valuable information, however the combination of two techniques used appropriately, such as MS and FTIR in tandem can provide a vast wealth of information from biological samples.

ACKNOWLEDGEMENTS

A portion of this work was supported by the Iowa State University (ISU) Plant Sciences Institute and was supported in part by the Department of Agriculture, CSREES award number 2009-65503-05798. This research was performed at the Ames Laboratory. Ames Laboratory is operated for the U.S. Department of Energy by Iowa State University under contract no. DE-AC02-07CH11358. Other funding was provided by the GAANN Fellowship.

I would foremost like to thank Dr. Sam Houk for giving me the pleasure to join his lab while I overlapped the last year of my Masters in Food Science with my first year in the Chemistry PhD program. When things got rough during the time I had issues with my thesis he was always there with advice, assistance writing, and an untold number of Hershey chocolate kisses. During my PhD he allowed me a great deal of freedom to pursue my interests dealing with mass spectrometry and biological materials, while always being there with rock-solid advice. Second, I would like to thank Dr. John McClelland and Dr. Roger Jones for allowing me the ability to pursue and publish work within their lab. Dr. Jones' guidance and assistance writing two of my papers was exceptional. Third, I would like to thank Dr. Derrick Morast for teaching me everything he could about our instruments in Gilman and collaborating with me on our zein research paper. His mentoring and advice has been invaluable during my PhD.

I would like to thank the Houk group members for their friendship, including Jeneé Jacobs, Jonna Berry, Katherine-Jo Galayda, Patrick McVey, Dr. Chris Ebert, Dr. Travis Witte, and Dr. Megan Mekoli. They were always of great help whether it be brainstorming, technical advice, or a productive group lunch on Friday. I would also like to thank my committee members for their assistance, Dr. Young-Jin Lee, Dr. Joseph Burnett, Dr. Jesudoss Kingston, and Dr. Emily Smith. I would like to further show my appreciation to Dr. Lee for meeting with me

and discussing many ideas (some outlandish) pertaining to biological mass spectrometry and MALDI mass spectrometry. I would also like to thank Dr. Stan Bajic for his advice and assistance this year with our lab's ongoing the laser desorption project.

I would like to thank the many friends I have in the department and I have made while my second stint in Ames. They made this time in Ames extremely enriching and enjoyable. I would like to thank my parents Karl Anderson and Katherine Anderson for their love and support in the past, and as I came back to graduate school. I would like to thank my brother John Anderson for his persistent, near daily calls and friendship. Last, I would like to thank my lovely wife Yeran, you were the one who convinced me to scale the cliff I considered a PhD. Your love and affection saw me through my toughest and best days during my Masters and PhD, thank you.