Communicating Research to the General Public

The WISL Award for Communicating PhD Research to the Public launched in 2010, and since then over 100 Ph.D. degree recipients have successfully included a chapter in their Ph.D. thesis communicating their research to non-specialists. The goal is to explain the candidate's scholarly research and its significance—as well as their excitement for and journey through their area of study—to a wider audience that includes family members, friends, civic groups, newspaper reporters, program officers at appropriate funding agencies, state legislators, and members of the U.S. Congress.

WISL encourages the inclusion of such chapters in all Ph.D. theses everywhere, through the cooperation of PhD candidates, their mentors, and departments. WISL offers awards of $250 for UW-Madison Ph.D. candidates in science and engineering. Candidates from other institutions may participate, but are not eligible for the cash award. WISL strongly encourages other institutions to launch similar programs.

Wisconsin Initiative for Science Literacy

The dual mission of the Wisconsin Initiative for Science Literacy is to promote literacy in science, mathematics and technology among the general public and to attract future generations to careers in research, teaching and public service.

Contact: Prof. Bassam Z. Shakhashiri

UW-Madison Department of Chemistry

bassam@chem.wisc.edu

www.scifun.org

MASS SPECTROMETRY APPROACHES TO TECHNICAL AND BIOLOGICAL STUDIES WITH HUMAN URINE

by

Corinne Elise Moss

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Biochemistry)

at the

UNIVERSITY OF WISCONSIN-MADISON

2026

Date of final oral examination: 05/06/2026

The dissertation is approved by the following members of the Final Oral Committee:

Joshua J. Coon, Professor, Biomolecular Chemistry
Judith Simcox, Associate Professor, Biochemistry
Andrea Galmozzi, Assistant Professor, Department of Medicine
Lloyd M. Smith, Professor, Chemistry

CHAPTER 4: Mass Spectrometry and Urine: Modern Tools for Decoding the Molecular Language of Living Things

4.1: The Winding Path

One of the biggest challenges I have had in my education is a lack of decisiveness. My interest in the world around me, particularly the natural world, runs both deep and wide, and it has been that way my entire life. I struggle to pick which section of a museum to visit first (regardless of the type of museum), few books or magazines are safe from my wandering eyes, and even after my undergraduate university, Montana State, started sending letters to my house and threatened to prevent me from registering for classes, I still couldn't pick a major and ended up with two degrees. So, deciding to narrow in on a field of study or a career beyond undergraduate and consider one system for years on end was daunting, to say the least.

The throughline in my education has been an interest in living things. Living organisms and the ecosystems they inhabit are substantially more complex than we often give them credit for. As I found myself exploring ecosystems in my classes and outdoors, I became curious about the ways you can think about living things from both a very zoomed-out perspective and a super zoomed-in perspective. Are you (and every bear, bee, and flower) a speck in the universe? Are you your own complex machine full of molecular machines whirring along to keep you alive? Or are you intricately connected to individuals around you in more ways than you'll ever realize? From my perspective, the answer is 'yes' to all of the above. The remarkable thing about biological systems is that no matter how far you zoom in or out on the system, there are still interactions to observe, many of which can be described or understood using principles of chemistry.

As I stared down my decision for my studies and future career, indecision quickly became its own form of a decision. My interest in understanding the world around me at multiple scales drove my decision to study both biology and chemistry, including classes ranging from physical chemistry and metabolism to the ecology of fungi and soil biogeochemistry. Along the way, I

began to do extracurricular research in biology and chemistry, and also found myself more and more interested in the outdoors and the outdoor community around me.

One of the things that most deeply frustrated me about trying to understand both biology and chemistry was a sense of ambiguity about *how* scientists know things to be true. Sure, we have large textbooks full of information, but how, exactly, did we learn all of the information that we now accept to be true? My frustration at this disconnect began leaking into my personal life by sheer coincidence. After re-reading *Into the Wild* by Jon Krakauer, I noticed an update to the post-script in a later-published version regarding the detection of a plant toxin in the plants ingested by the main character in the book. The post-script included details about a technique known as "mass spectrometry" (MS) that compared masses of plant compounds and determined that the cause of death as different than Krakauer originally published. *Finally!* An observation of a biological result (death) with a chemical explanation (plant toxins) that included the steps scientists took to get there (MS analysis).

Further reading on this topic told me all about how plants contain many chemicals responsible for flavor, scent, toxicity, distress signals, nutrition, and many other things. While some of these are easily detectable to humans, such as the distinct spice of a pepper or scent of a flower, many are not. Organisms communicate through these chemical signals as if they are speaking a hidden language. Even more shocking to me was the idea that these chemical signals are changing around us all the time in the environment, completely undetectable to us without the use of complicated machines, like mass spectrometers, to measure and analyze them.

Reading about the use of MS to measure and understand chemicals in living things felt like the missing link explaining how scientists know the things they claim to know, and it turned my indecisive nature into enthusiastic focus. In my experience, I am not the only one missing this link,

and I believe that lack of fully understanding the mechanisms and processes behind scientific research impairs our ability to understand, believe in, and support scientific work. For this reason, I chose to write this chapter as a small step forward in communicating science to the public so that we all can understand scientific work and the world around us a bit more. This missing link is also where our journey to understand this analytical process really begins today. I am grateful to the mentors who helped me along the way as I honed this focus, and for the editing help from Elizabeth Reynolds at the Wisconsin Initiative for Scientific Literacy (WISL) as I have worked to put the journey on paper.

4.2: The Secret Language of Living Things

On a casual walk through a forest, it would be easy to say that the complexity of one plant is less than the complexity of the entire forest. Looking more closely, however, if we consider the cells and molecules that make up that plant, it is actually just as complicated as the whole group of different plants and animals that make up the forest. The best part about studying and observing living systems is that we can adjust our scale to be as zoomed-in or -out as we would like, and still have something new to discover.

To fully understand the *eureka!* moment that occurred as I read *Into the Wild*, we can start by zooming in on Alaska. Alaskan wilderness encompasses many aquatic and terrestrial environments, one of which is the Boreal Forest, an ecosystem that covers about 70% of the total land area in Alaska.[1] Boreal forests include a huge range of species, including animals like moose, grizzly bears, and great horned owls, and plants like spruce, willow, and alder trees, amongst many smaller species. Zooming in even further, there is a species of wild potato commonly found in this ecosystem. Known to scientists as *Hedysarum alpinum* and to casual naturalists as alpine sweetvetch or bear root, this small plant is partially edible. Like any other plant, it contains roots, leaves, and flowering parts, all of which are built up of many, many plant cells (**Picture 4.1**). Inside

these cells is a machine-like network of large molecules whirring along to keep the plant alive and growing. These large molecules interact with each other and interact with smaller molecules, known as metabolites. Zooming in as far as we can go, we find a small molecule in sweetvetch with a very unique role, *L-canavanine*.

Picture 4.1: Alpine Sweetvetch in bloom in Denali National Park. *Image credit: Denali National Park and Preserve, Jacob W. Frank*

L-canavanine is classified as a secondary metabolite in plants. Secondary metabolites, unlike primary metabolites, have functions that are considered non-essential to a plant's growth and reproduction. These functions can include warding off predators, responding to environmental changes, signaling changes or distress to surrounding organisms, attracting pollinators, and other helpful but non-essential functions. In the case of *L-canavanine*, it functions as a "defensive" molecule. It is found primarily in the seeds of alpine sweetvetch (and related legume species) and

can be broken down by the cells within the seed pod to be used as a source of nitrogen for growing a new plant.[2] When the seeds are eaten, it is believed that L-canavanine (likely in tandem with other molecules) gives the seeds a bitter taste to most organisms. The bitter taste deters organisms from eating sweetvetch seeds, which helps the seeds to survive so the plant can reproduce.[3]

More dangerously, however, *L-canavanine* mimics an essential metabolite and takes its place in the cell, but it does not function the same as the essential metabolite. The imposter metabolite integrates itself into larger molecules, disrupting their activity and degrading cellular functions in the process.[4,5] For one molecule in one cell, the consequences are high. Consider, however, that one seed may contain more than a quadrillion of these molecules, and once they are consumed by another living thing, they can enter cells all over that creature's body to limit or eventually completely end cell activity. When we zoom back out from our individual molecule and re-consider the whole complex organism, the consequences become disastrous.

When consumed by humans, the incorporation of *L-canavanine* into large molecules as an imposter metabolite causes other cells in the body to view the altered molecules as invaders, and begin attacking them. Cells get destroyed, resulting in autoimmune symptoms similar to those found in diseases like Lupus.[6] The incorporation of *L-canavanine* is increased when the human body is stressed or lacking in protein, heightening these symptoms and altering cell function so severely that it can lead to death. In humans and other animals, these consequences are also irreversible. Mammals do not have any natural way to break down or transform *L-canavanine*, so the only way to avoid the toxic effects is to stop eating the plant.

Even looking at just this one exemplary molecule, we can find multiple functions depending on the bigger context the molecule is in. These complex functions serve as an outstanding example of how living things function in ways well beyond what humans can naturally

perceive. Plants and animals regularly transmit chemical signals in a sort of secret language. *Into the Wild* tells the tale of an outdoorsman, Chris McCandless, who mistakenly ingested large amounts of sweetvetch seeds and died as a result. Even after skilled forensic analysis of McCandless' death, it took almost two decades for the cause of death to be determined. In a wild goose chase-style scientific quest to answer an unanswerable question, author Jon Krakauer and several scientists went back and forth, comparing results with each other, consulting Indigenous knowledge sources to understand probable causes of death, and re-testing and re-validating samples found with McCandless. The technology used to analyze evidence down to a molecular level was still expanding and growing during the initial investigation in the early 90's, and was not nearly as accessible, widely used, or powerful as it is today. Advancements in technology over the following decades enabled much more sensitive analysis and the detection of compounds not initially known to be present in plants that had been eaten by McCandless. In the most recent analysis, performed some 20 years after McCandless' death, high-performance MS detected *L-canavanine* in potentially lethal amounts.[7] Without the ability to zoom all the way into the molecular level, his death would remain a mystery, and the secret language of these plants would have remained a secret.

4.3: The Ion's Journey

As it turns out, measuring chemical signals using MS is much more common than most people realize (more on that later), but we often think of it based on the results or outcome, rather than thinking about the instrument itself. Likewise, though the most easily digestible portion of my work lies in the biological outcomes, much of my time as a graduate researcher has been spent working to thoroughly understand how a mass spectrometer works and how to best use it for a

specific biological system. To get meaningful results and understand how these results come to be, we can turn to the samples and the instrument that makes remarkable results possible.

Many of the samples that can be analyzed using MS, such as plant tissue or human biofluids, are very "complex". In the context of MS, complexity refers to the number of different types of molecules present within them, and higher complexity samples can be harder to analyze. If you wanted to read a book, it might be challenging to understand if every sentence of the book was written on the same page to be read at the same time. In MS, it can be difficult to measure all of the molecules in a complex sample if they are all input into the mass spectrometer all at once. To help separate out molecules like an author might separate the chapters of a book, we use additional tools known as separation techniques.

One of the most widespread and well-known techniques is called liquid chromatography (LC). In a liquid chromatography instrument, sample molecules are added to a column-shaped compartment that contains a substance that can bind to many types of molecules, and then slowly washed out of the compartment with chemicals chosen by the researcher, depending on what molecules the researcher is interested in measuring. Molecules will wash out and separate slower or faster depending on how well they are bound to the substance inside the compartment. These types of separations can be as short as a minute or two, or up to two or more hours long.

Another less common, but very interesting separation technique is called capillary electrophoresis (CE). In CE, sample molecules are added to a liquid and then put into a very, very, *very* narrow glass tube. Electricity and pressure are applied to the tube, and the molecules get pushed through and out of the tube at different speeds, depending on their size and several other properties that impact how they act when electrified. Early versions of this system were very challenging to use, but recent changes have made it as easy to use this technique as it is to use LC.

With either of these separation techniques, the molecules in a complex sample are introduced into the mass spectrometer over a period of time, rather than all at once. These techniques are the first step in transforming the unknown contents of a living thing into pieces of information that we know how to read and understand. After separation, molecules are input into the mass spectrometer, where the real magic happens (**Picture 4.2**).

Picture 4.2: A liquid chromatography instrument (left) attached to a mass spectrometer (right). This MS instrument stands approximately 55 inches tall and weighs 1,600 lbs. *Image credit: Corinne Moss*

Most of the information about how a mass spectrometer works is contained in the name. Mass refers to how much of something there is, or more specifically, the strength of gravitational

attraction something has to other bodies. Spectrometer comes from the Latin word *spectrum*, to see, and the Latin root *-metria*, to measure. So that's all there is- seeing and measuring the mass! Almost. Mass, as we think of it on a zoomed-out level, is different depending on the item being measured. We intuitively know that a sequoia tree has more mass than a rabbit, and given the names of the two organisms and their masses, it would not be a challenge to correctly assign the masses to them. Even for items of a more similar size, such as the marshmallow, chocolate, and graham cracker that make up a s'more, it would be possible to distinguish the three elements based on just their mass (assuming you only have the ingredients for a single s'more).

When we zoom in to a molecular level, this idea of unique masses still holds true. Molecules are composed of individual atoms, all with known masses. When we have the mass of a molecule and a set of potential atoms that may be in the molecule, it becomes possible to piece together what the molecule is, like a sort of chemical-mathematical puzzle. In living organisms, these atoms are carbon, oxygen, hydrogen, nitrogen, and occasionally sulfur or other nonmetals. These atoms can be combined in tens of thousands of ways to create molecules with a huge range of different masses. Essentially, knowing the mass of a molecule is the biggest first step in knowing that molecule's identity.

Mass is also interesting to consider in the context of molecular identifications because it is not the only unique thing about many molecules, but it is one of the most straightforward to make use of. Many properties of molecules that differ between them, like the temperature at which they boil or freeze, how they react with other substances, or how well they dissolve, are changed if the molecule or sample has contamination in it. Some other properties of molecules, like the movement of their component atoms, are less susceptible to contamination, but just as challenging to measure as mass. Other techniques that measure the mass are often imprecise or don't work

with extremely complicated mixtures or with molecules that are gases. The mass, however, only changes if a molecule reacts with another molecule, but this reaction changes the identity of the original molecule as well, making it still possible to use mass as a distinguishing feature. With known mass information, the identification of a molecule can be confirmed, even when in a mixture of multiple things or when its other properties are unknown.

Having established that mass is a useful property to glean molecular identifications, it then becomes a bit more complicated to understand how that is achieved. Within one molecule, there may be a few or many atoms that bond together. Within each of these atoms, there are smaller particles, known as sub-atomic particles, that hold together to form the atom (**Figure 4.1**). Each type of atom, known as an element, has a different number of sub-atomic particles, which give that element its defining properties, including the mass. These subatomic particles also change how the atom and the molecule it is in behave and respond to changes in their environment.

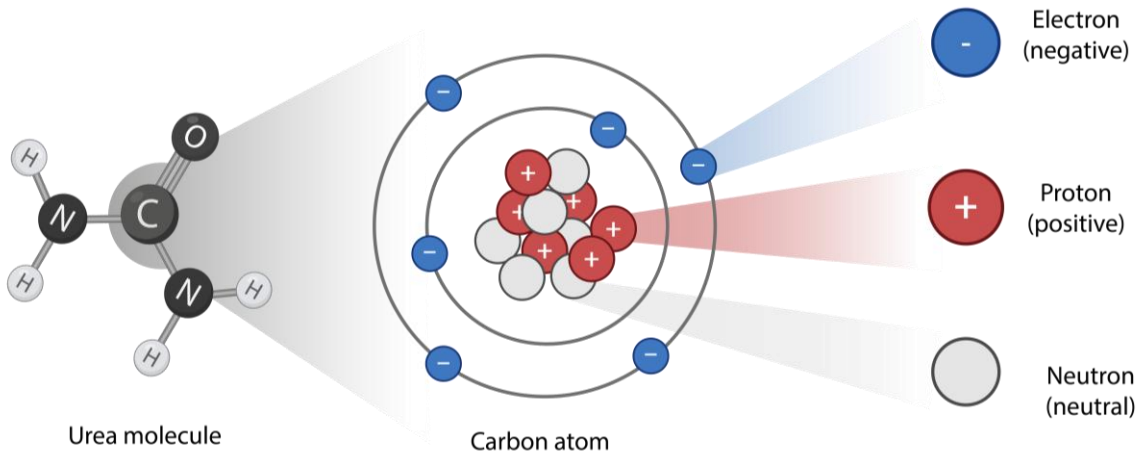

Figure 4.1: An example metabolite, urea (far left) is shown to illustrate the composition of molecules. Each molecule contains atoms (center) which each contain sub-atomic particles (right) that are responsible for the behaviors and charge of every molecule.

In the case of MS, the key quality to understand about subatomic particles and atoms is "charge". Charge refers to the way a particle responds in an electric field and how it responds to other particles. Most people associate charge with magnets, which repel or attract each other, and scientists think of charge in terms of "positive" or "negative". It is a fundamental property of matter, and can be characterized for everything in the known universe (although sometimes the charge is neutral!). In some cases, a neutral thing can be given a charge, or charged things can be made neutral. In these cases a charged molecule is referred to as an ion, whether it is positively or negatively charged. Taking out our magnifying glass again to zoom in as far as we can, we can see that the number of subatomic particles present in an atom and the charge of those particles determine the charge of an atom, and the bonding and arrangement of atoms determines the charge of molecules. The cumulation of all of these charges then changes how molecules interact with each other and their surroundings.

The reason that understanding charge is so important, and the big catch with MS, is that MS doesn't actually measure mass; it measures a ratio of *mass to charge* on a molecule. Scientists can manipulate electric fields and measure the actions of the molecules within them because charges cause molecules to react to electric fields. As long as the molecules are charged, then we can manipulate and measure them.

The first step in an MS analysis is to impart a charge onto the molecules you'd like to measure, known as ionization. Ionization can be done in a variety of ways, most commonly by applying electricity to the sample directly before it is input into the instrument. One of the most prolific methods to do this is called electrospray ionization (ESI). In ESI, the electricity is applied just before a liquid sample is sprayed through a very narrow pointed opening and into the MS. As the sample is pushed through this narrow point, the electric charge and motion of the ions forms a

cone of liquid, known as a Taylor cone, which reaches a very narrow point where the liquid separates into aerosolized droplets, similar to a perfume atomizer. The exact physics behind Taylor cone formation is still being discovered by scientists (**Figure 4.2**). As the aerosolized droplets disperse, the liquid evaporates, leaving behind only the charged molecules from the sample, which enter the MS, now known as ions.

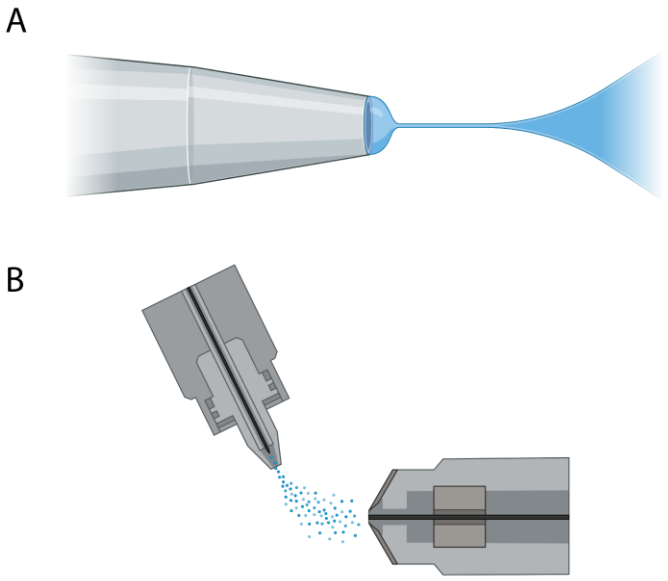

Figure 4.2: a) Taylor Cone formation from an electrospray source, where the tip of the cone begins at the spray source. b) An electrospray source is positioned near the inlet of the mass spectrometer to input molecules to the instrument.

Once ions are inside the MS, a long series of electrical systems helps move them around, sort them, break them into pieces, and measure their mass and charge. Each instrument contains a different configuration of systems depending on the intended system use, though most mass spectrometers have a similar general configuration.

The first stages of the electric systems that most molecules encounter inside the MS are the lenses. If you think of a traditional glass lens, you may think of focusing sunlight into one beam,

like using a magnifying glass to light an ant on fire. When you use these lenses, you are actually employing similar physics principles that get employed within an MS. But instead of using light, MS uses electricity to take a scattered group of ions and focus them into one very concentrated beam of ions that gets sent deeper into the instrument.

This beam is guided into mass analyzers, which can take many forms. These analyzers do more than just analyze, though. They can filter the ions and select ions with specific mass-to-charge ratios or a range of mass-to-charge values. They can also serve as chambers to hold ions and facilitate other chemical reactions, if needed for analysis.

Before ions are ready for further movement or analysis, they need to be at the correct energy levels. Several chambers can add energy to the ions and then direct them to the mass analyzer. In some cases, this may be the same mass analyzer as the first, and in other cases, a mass spectrometer will have two or even three mass analyzers. In this final mass analyzer stage, the electrical signal given off by the ions can be read and mathematically transformed to yield a mass-to-charge value that we can read and interpret to identify the molecule it corresponds to. These signals are known as mass spectra, and contain mass-to-charge values, time stamps, and the number of particles that were measured (intensity).

This process all occurs within fractions of a second and repeats until the entire sample has been analyzed. In short, to read the secret language of living things, we take advantage of the electrical properties of molecules to transport and manipulate them until this biological language is written in a language we can read. This process usually works great, but it doesn't address one major issue that occurs within mass spectrometry- what if two *different* molecules have the same mass?

4.4: Distinguishing Ions- Demolition Derby Style

Before considering how an ion travels through an MS, we established that biological molecules are composed of a handful of different atoms that get combined in thousands of different ways. In each molecule, there can be multiples of each type of atom. In some cases, these molecules bind together to form large molecules with hundreds to thousands of atomic components! The arrangement of each component atom in a molecule is also crucially important. This arrangement is known as molecular structure, and is dictated by fundamental properties of each element and its subatomic particles. Getting into these properties would really require its own chapter, but it is helpful to remember that the arrangement of molecules and the behavior of their atoms is not all random. The structure and behavior of molecules follow predictable patterns. A different molecular structure, even if it is just two of the same atoms in different positions, can result in two very similar molecules having separate identities and unique properties. If we were to only measure the mass-to-charge ratio of these two similar molecules, it would be the same value, making them indistinguishable.

To tackle this hurdle, one of the primary strategies used in MS work is known as fragmentation. Fragmentation is a process of breaking each molecule into pieces, and then measuring those pieces and using them, with the help of a computer, to put the molecule back together, like assembling a puzzle or a Lego kit. Similarly to how atoms follow specific patterns when forming molecular structures, they also follow specific patterns when molecular structure is being destroyed. Understanding these patterns and rules, along with the known masses of every

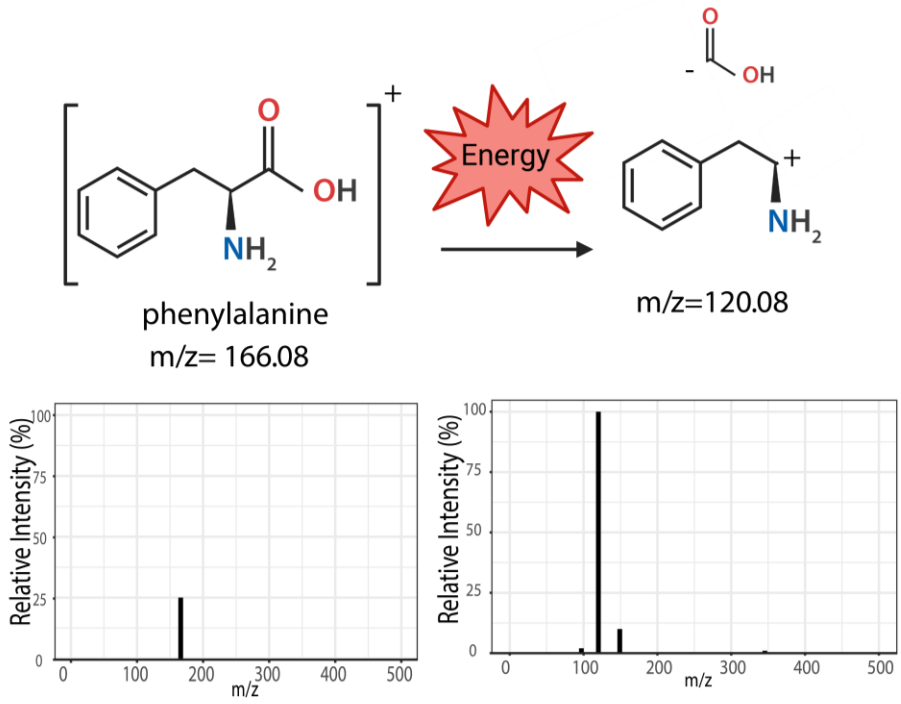

Figure 4.3: An intact molecule, such as phenylalanine, will give a signal at one m/z value (left). When it is fragmented, it will give signal at its fragmented mass values (right).

atom, helps us make educated guesses about what a fragment might contain, and combined with knowledge of the molecules' unbroken structure, we can determine the identity and structure of thousands of molecules.

Fragmentation can be done in several ways, and in several locations in the mass spectrometer. In a mass spectrometer, fragmentation happens after the first mass-to-charge ratio

measurement of the ions. One of the most common ways of fragmenting molecules is with a collision gas in a collision chamber. As the ions move around the chamber, they bump into and collide with the gas molecules. Just like a demolition derby where cars collide and break apart, these collisions break the ions into pieces. After enough collisions have happened, the fragments are moved back into the mass analyzer a second time so that the mass-to-charge ratio of these fragment ions can be measured. In an experiment that measures fragments in addition to the intact ions, this repeats for every group of ions in the entire sample.

Fragmentation can also be done using lasers, which add large pulses of energized particles to the ions. As the ions absorb this energy, their own energy increases. If the ion gets enough energy, it will fragment into pieces. Fragmentation can also be achieved by adding specific ions to the sample ions that can react with and break apart the sample ions. Each of these, including adding a gas to molecules, works best with specific types of molecules and specific mass spectrometer instrument set-ups.

Using fragmentation and intact ion data gives researchers all of the puzzle pieces needed to identify molecules in biological systems and begin to read their secret language. Though challenging, improvements in these methods over the past few decades have greatly increased our ability to get more information from any given sample. Once we understand these tools, we can apply them to just about any living thing!

4.5: MS as a Tool for Human Health

My interest and initial understanding of MS was framed in the context of plant ecology and toxins that impact human health, however, plant metabolism and toxicology are not the only fields of research that benefit from using MS to decode the molecules at work. MS can be applied to

environmental questions, fundamental science and physics, pharmaceuticals, and is very commonly applied to human biology and challenges in human medicine.

One important application of MS that many have heard of, but may not realize uses MS, is drug testing. When individuals intake drugs or pharmaceutical substances, these molecules get processed within the body and excreted through the urine. MS can analyze many types of biological samples and identify thousands of small-molecule types, making it an outstanding tool to use for drug detection from urine samples. Labs across the world regularly test samples from professional athletes, as well as from workplaces that require drug testing. Every Olympics, the World Anti-Doping Agency (WADA) collects thousands of urine samples from athletes and processes them in local WADA-accredited laboratories using MS techniques, and has been doing so since the mid 1970's.[8] These labs contain many types of different mass spectrometers and other tools to help prepare and analyze the samples, and some WADA labs even use the same Exploris™ MS instrument I use in my own work. Any major doping scandal you've read about in the past 50 years has been uncovered with the help of MS.

Urine samples contain much more information than just the presence or absence of drugs, however, and have taught me a lot about using the language of molecules to understand biology. Following my initial interest in plant ecology studies, my fascination with MS took me down a very unexpected path to understand how urine can be used to its maximum potential to understand human biology.

Drug testing sites, clinical uses of urine, and scientific studies have many similarities, despite using urine for different purposes. In the case of most drug testing uses, unlike for elite athletes, urine samples are taken at random intervals, not on a regular basis. In medical clinics, urine samples are usually taken as a way to achieve a specific diagnosis at a specific time in a

patient's life, and multiple samples per person are rarely taken in a short period of time. In many scientific studies that use urine samples, this is also the case. There are exceptionally few studies that use urine to look at changes on a day-to-day basis, even though the molecules present in urine are directly related to things we consume daily (food and drugs). Many urine-based studies are also experimental rather than observational, meaning that the researchers set specific conditions that they think will cause a response, whereas an observational study does not set any conditions and simply measures outcomes. The lack of regular sampling at short intervals as a way to observe changes without deliberately causing them appeared to be a large gap in current research, even though there is enough established knowledge about urine to think that daily samples could show daily changes.

Urine is such a notable type of sample to analyze because it contains molecules that come from other areas of the body, not just from the urinary tract. As blood flows through every part of the body, it picks up waste molecules that the body is no longer using. These waste molecules get filtered out of the blood in the kidneys and form the molecular component of urine. Waste molecules can also come from changes that happen as a result of something else a person did, like hard exercise or a night of poor sleep. In some cases, the waste molecules in urine are caused by problems in a body system far from the urinary tract. For example, when a person develops diabetes, the way their cells process glucose, a sugar molecule, changes. This change causes an increase in specific sugar and sugar-adjacent molecules in the urine. In the case of diabetes, the increase in sugar molecules as measured in urine is so large that it doesn't need to be compared to urine samples from that individual that were taken on other days. Many other molecules found in urine do not change quite as dramatically, however, and smaller changes may indicate a change in health or lifestyle, or they may be the result of natural biological differences between people.

Measuring urine samples on a much more regular basis, such as daily or multiple times a day, would make smaller health changes and the relation of urine molecules to lifestyle choices clearer, potentially giving substantial insight into human health.

To investigate this idea further, I needed to gather both urine samples and data to coincide with the samples. Humans have a huge amount of biological and lifestyle differences between them, so observing just a few people would not yield much information, however, observing many people becomes incredibly logistically challenging. To balance these challenges, I had 24 people help me in the study. Each person wore a FitBit fitness tracker device and used a smartphone app to record everything they ate or drank for a month. Within that month, they also had a two-week period where they collected three to five urine samples per day and stored them in the freezer until they could be transported to the lab.

Once in the lab, I, our study director, and a post-baccalaureate research assistant, catalogued, sorted, organized, and labeled all 1,054 samples. We also downloaded and organized gigabytes of FitBit and smartphone app data, with hundreds of data points about things like grams of protein eaten in a day or minutes of vigorous exercise. For every single urine sample, I removed a small amount of the sample and set it up in a container to be sampled by the mass spectrometer. I then spent a few months analyzing every single sample two times- one using LC-MS and a second time using CE-MS. Because the two methods work slightly differently, they also end up separating and identifying different sets of sample molecules, helping me identify a broad range of molecules overall.

After MS data generation, I needed to sort through the data and make sense of it. Though MS translates the hidden language in living things, reading it can still take a lot of time. I started by going through all of the data produced from the mass spectrometer and the identifications that

had been hypothesized by our computer program which we use for analyzing MS data. After going through hundreds of pieces of data over the course of many weeks, then going back and forth with several collaborators who went through the same process for the data produced from their method, we had our two urine data sets from our two analytical methods. I then took our molecule data and our FitBit and smartphone data, and began searching for interesting patterns, stand-out trends, or anything else that could help us determine if and how daily urine tracking could be used to improve human health monitoring.

I first started by comparing all of the lifestyle data values to all of the metabolite values, and filtering them out to see which ones were most closely correlated. From these identified pairs, I was able to look at how they changed over time for each person. Over the two weeks that our study participants collected samples, I can graph each variable. This shows how their activity, diet, and urine molecules changes from day to day, and shows how some of these changes are highly related to each other (**Figure 4.4A,B**). I worked with a data scientist to analyze this for everyone in our 24-person cohort, and determined whether each trend was random or an interesting and potentially useful finding. For the useful findings, I then searched through other published papers

to see if other scientists understand what the metabolite tells us about the body, and how that relates to the changes in lifestyle.

One good example of these relationships is in a metabolite called xanthurenic acid. In our

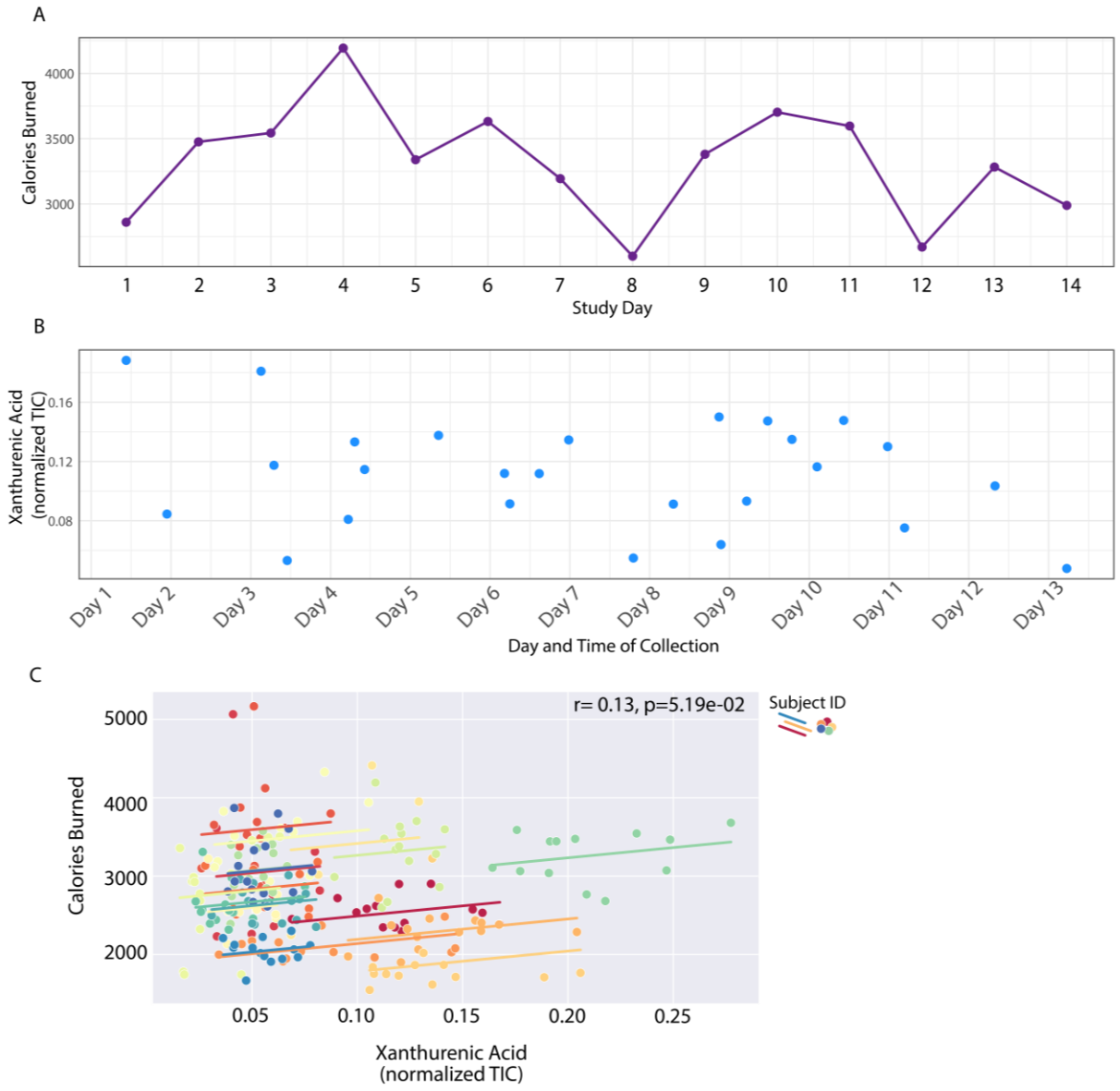

Figure 4.4: Longitudinal tracking shows daily changes and larger-scale patterns. a) Total calories burned over each day of the study for one participant. b) Measured levels of xanthurenic acid in the urine samples for every sample in the study for one participant. c) Correlation analysis of xanthurenic acid and calories burned shows that as more calories are burned, xanthurenic acid excretion increases.

study, we found that when participants exercised and burned more calories, there was more

measurable xanthurenic acid in their urine (**Figure 4.4C**). Other research has found that this molecule is involved in some activities in the cell that happen during and after a long period of exercise. The other molecules involved in these reactions in the cell are known to help cells produce energy, which is needed to keep exercising. If you were training regularly and hoping to see if your cells were adapting to the training and improving in their energy usage, you could look at xanthurenic acid levels. A steady increase with exercise tells us that cells are switching to energy production as a response to exercise, but an absence of this trend could tell you that something else is preventing your training from having the desired effect. Combining the molecular data with lifestyle data helps paint a clearer and more specific picture of what is happening inside the body, which can help us change our daily behaviors.

While initially looking through the data and talking about it with other researchers, we began to wonder if there were other, more subtle relationships that we could find from our data. A lot of recent research has been done on specific types of molecules that do not break down in the environment, known as polyfluoroalkyl substances (PFAS), or "forever" chemicals. The compounds have been widely reported on in the news because they are known to be harmful to the environment and human health, but they are very useful in lots of industrial and commercial settings. What news outlets often don't specify, though, is that many PFAS can dissolve into water, which to me means that if some of them were found in the human body, they could potentially be filtered into the urine (which is mostly water). After a few ominous discussions with other scientists, I began to search through recent scientific studies to see if there were other researchers looking into this. A lot of reading showed me that there are many people looking into this, but it is a challenging problem to tackle because it is so novel, and it is *highly* frowned upon to use toxic

chemicals in experiments with people. This makes observational studies, such as the one I did here, much more useful.

Though I did not specifically design our study or conduct our data collection with PFAS in mind, the data that I had was compatible with software that could search for PFAS molecules in every sample, kind of like using Control-F to find a certain word in a large document. Though part of me hoped that I would find nothing at all, I was not surprised that our urine samples had hundreds of potential PFAS in them. Some of the identified molecules have known environmental presence with established effects in humans, while many others are poorly understood or are not thought to be harmful.

Looking through all of my data, it also began to become clear that the identified molecules were not found in every sample, and not every PFAS was found in every person's samples. There was a lot of variation in how many molecules were found, and how many times over the course of the two-week study they were found. This variation made me wonder if there is day-to-day variation in PFAS excretion, similar to the day-to-day variation I saw with metabolites. When I chose a PFAS molecule of interest, and then graphed the levels I measured over each day of the study, it appears very plausible that there are environmental exposures (e.g., food packaging, air pollution, etc) that result in higher PFAS excretion. Other urine-based PFAS studies have found PFAS excretion, but few/none have shown that PFAS excretion can change on a day-to-day basis, and further validation and combination with more detailed daily activity tracking could help people understand how to lessen their exposure to potentially harmful "forever" chemicals. More broadly speaking, this type of information could also give valuable information about which communities are more affected by these chemical exposures, and encourage public health action to mitigate everyone's exposure to them.

Looking at the range of identified relationships in the urine study made it clear that there is both a large amount of variation in human metabolism and that these differences in metabolites can be used to understand why biological changes or responses to lifestyle changes are happening. In the example relationships of xantheurenic acid to exercise and PFAS changes over time, it becomes clear that the hidden language of biological systems is at play not just around us, but within us every day. Good identification and tracking of these molecules would help to decode and read this language, and help us make informed health and lifestyle decisions. All of this, however, relies on the ability to identify the molecules in urine, which is done very effectively with a solid understanding of MS.

4.6: From the Perspective of a Lifelong Learner

Understanding the instruments that help us answer biological questions can be as complicated as the answers themselves, but when we understand how to apply MS in one context, it becomes easier to see how it might apply in another context. I was initially blown away by the ways that MS could be used to detect molecules, like the toxins in *h. alpinum*, and then that knowledge could be used to answer questions and solve problems. As my understanding of MS grew, I saw the potential to answer a huge swath of questions in biology. Beyond understanding how plants communicate with their surroundings and can defend themselves, MS can be used to analyze the molecules that make any living creature function and tell us all about how those functions are changing internally and in ways we otherwise can't observe. My research in tracking urine metabolite changes over time is just one example of the power of molecular instrumentation.

The instrument described here represents one of dozens of possible combinations of molecule separation and MS techniques. We selected these to use based on their compatibility with urine and the time it would take to analyze the large number of samples I was working with. In

other situations, it is possible to use much larger and more complex instruments, more or less precise separation techniques, no separation technique, and many, many different configurations of mass spectrometers. Changing each aspect has the potential to answer different types of questions from all sorts of different samples. As researchers, we can tailor our MS experiments and begin to see what is hidden in the molecules there.

As I move beyond my formal education, I look forward to learning about new ways researchers are using MS systems to answer their questions, and want to demystify these complex techniques that give insight into processes happening in and around us every day. My indecisiveness will be much less of an issue moving forward, since I will get to move from conducting extremely specific research into educating undergraduates on the essentials of chemistry as an assistant professor. Knowing how many biological questions can be answered with MS, I could never leave the instrument behind, but I hope to expand my own scope of understanding and continue exploring all sorts of questions and ideas that help us understand and use these systems better, and give undergraduates the knowledge they need to do the same.

4.7: References

1. Bonanza Creek LTER. https://www.lter.uaf.edu/boreal-forest/about.
2. Rosenthal, G. A., Berge, M. A., Ozinskas, A. J. & Hughes, C. G. Ability of L-canavanine to support nitrogen metabolism in the jack bean, Canavalia ensiformis (L.) DC. *J. Agric. Food Chem.* **36**, 1159–1163 (1988).
3. Meadows, R. Bad Taste Protects Fruit Flies from Eating a Toxic Amino Acid in Plants. *PLoS Biol* **7**, e1000128 (2009).
4. Rosenthal, G. A. l-Canavanine: a higher plant insecticidal allelochemical. *Amino Acids* **21**, 319–330 (2001).
5. Bleiler, J. A., Rosenthal, G. A. & Janzen, D. H. Biochemical Ecology of Canavanine-Eating Seed Predators. *Ecology* **69**, 427–433 (1988).
6. Akaogi, J. *et al.* Role of non-protein amino acid l-canavanine in autoimmunity. *Autoimmunity Reviews* **5**, 429–435 (2006).
7. Krakauer, J., Long, Y., Kolbert, A., Thanedar, S. & Southard, J. Presence of L-Canavanine in Hedysarum alpinum Seeds and Its Potential Role in the Death of Chris McCandless. *Wilderness & Environmental Medicine* **26**, 36–42 (2015).
8. Ljungqvist, A. Brief History of Anti-Doping. *Med Sport Sci* **62**, 1–10 (2017).