Communicating Research to the General Public

At the March 5, 2010 UW-Madison Chemistry Department Colloquium, Prof. Bassam Z. Shakhashiri, the director of the Wisconsin Initiative for Science Literacy (WISL), encouraged all UW-Madison chemistry Ph.D. candidates to include 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 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.

Over 50 Ph.D. degree recipients have successfully completed their theses and included such a chapter.

WISL encourages the inclusion of such chapters in all Ph.D. theses everywhere through the cooperation of Ph.D. candidates and their mentors. WISL is now offering additional awards of $250 for UW-Madison chemistry Ph.D. candidates.

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.

UW-Madison Department of Chemistry
1101 University Avenue
Madison, WI 53706-1396
Contact: Prof. Bassam Z. Shakhashiri
bassam@chem.wisc.edu
www.scifun.org

Novel Proteomic Approaches to Characterize Endogenous Membrane Proteins

By

Kyle Brown

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

(Chemistry)

At the

UNIVERSITY OF WISCONSIN-MADISON

2020

Date of final oral examination: 08/12/2020

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

Ying Ge, Professor, Cell and Regenerative Biology and Chemistry

Sam Gellman, Professor, Chemistry

Lingjun Li, Professor, Pharmacy and Chemistry

Song Jin, Professor, Chemistry

Developing New Soaps to Access Hard-to-Study Proteins

Kyle Brown, Ge Group

In many cases, it can be challenging to provide enough detail and context in a scientific publication to explain the work to non-experts while maintaining conciseness (otherwise every publication would be a book). Moreover, small advancements in a particular area of science, while important, may only be of interest to those in that field. Therefore, it is practical for us to tailor our publications, to a certain degree, to our expertise. On the other hand, we must engage the broader scientific and non-scientific communities to explain the importance of our work. Fostering widespread interest and faith in science is vital for its future success. To this end, I have written a chapter of my thesis explaining my research to an audience of non-experts to make it more accessible than a traditional scientific publication. I hope that my family, friends, and anyone curious about understanding more about science will be able to learn something about what I have accomplished over the course of my PhD. I would like to thank the Wisconsin Initiative for Science Literacy at the University of Wisconsin-Madison for supporting this chapter and for their important work of promoting the communication of science outside the academic setting.

I have always been interested in understanding how the world around me works. In particular, I am fascinated by the complexity we observe in biology. There is so much going on in our bodies that we cannot begin to explain or understand. It can be daunting to know where to

begin when trying to explain why people have certain traits or get sick or any number of interesting questions regarding our nature. For this reason, I decided to get a PhD focusing on life science.

The simplest explanation of our approach to better understanding human biology is we try to study the most fundamental components and work our way backward. We can remember from our early biology classes that humans are composed of various organs, which are made up of cells (the basic unit of all life). An important part of our cells is DNA, the genetic blueprint of an individual. The role of DNA is to code for RNA that makes protein out of chains of amino acids called peptides. Each protein has a specific function, or set of functions, within cells that keeps us alive and allows us to respond to our environment. Thus, proteins represent an important starting point towards understanding our cells at a molecular level.

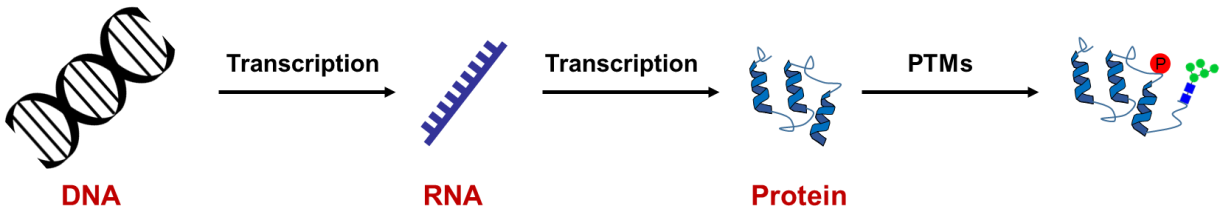

Figure 1. Scheme illustrating the process of DNA transcription into RNA, which is then translated to make proteins. Proteins perform many vital functions in cells and post-translational modification (PTMs) of proteins can change their functions giving rise to much of the biological complexity we observe in cells.

What is not as commonly known is that proteins can be modified by specific chemical groups, which can dramatically change their function (**Figure 1**). The chemical groups that are added to proteins are called post-translational modification (PTMs). A single gene, therefore, can produce many different protein forms, called proteoforms, giving us a lot of our biological complexity. We often think of PTMs as the body's barcode system, where a protein is produced

and then labeled with different codes resulting in them going to different locations in the cell or performing different roles.

The study of proteins, referred to as proteomics, seeks to understand the function of proteins in our cells and determine how they are regulated by PTMs and other biological events. In particular, we tend to study proteins by comparing two conditions, like health versus disease. Identifying changes from the normal state, for instance, lower abundance of a protein or the presence/absence of modification enables us to identify markers of disease progression and potential targets for new therapeutics.

A common tool used to study proteins is mass spectrometry, which can be thought of as a "molecular-scale." Mass spectrometry allows us to identify proteins in a sample, determine their abundance, and evaluate their modifications all by simply measuring their mass (measured in Daltons [Da]). To study proteins using mass spectrometry, we first have to extract them from the sample of interest. Then, we separate the proteins so that we can analyze them individually. The proteins are then ionized by a process called electrospray ionization (ESI) and a mass spectrometer is used to determine their mass (**Figure 2**).

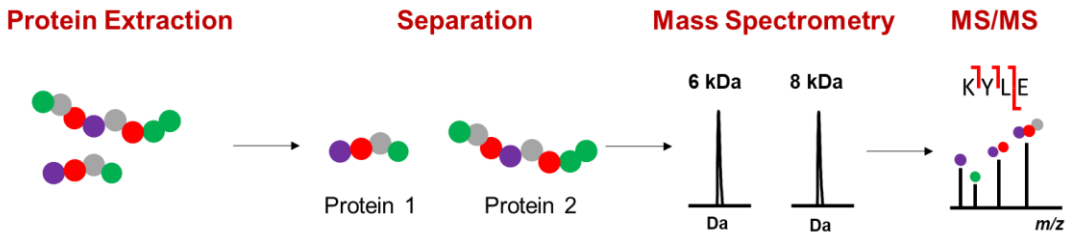

Figure 2. Process of proteins sequencing by mass spectrometry. Proteins are extracted from cells, separated, and weighed using mass spectrometry. Finally, proteins can be fragmented in the mass spectrometer and the pieces weighed to gain sequence information for more confident protein identification.

As we discussed earlier, proteins are made up of a unique sequence of amino acids and therefore have a unique mass. The predicted masses of all proteins are known since their amino acid sequences are determined by DNA, and we sequenced the entire human genome in 2002. Therefore, we can compare the masses we observed by mass spectrometry to the predicted masses to identify them. Unfortunately, modifications, mutations, and similarities in protein sequences can result in proteins having almost identical masses or at least result in their masses differing from the predicted values. Therefore, we sequence proteins, meaning we determine their amino acid composition, by fragmenting them in the mass spectrometer (**Figure 2**). Protein fragmentation (also referred to as tandem MS or MS/MS) is performed by colliding the protein ions with inert gas molecules (e.g., nitrogen) and measuring the mass of the resulting pieces. This helps us put together the puzzle, so to speak. Importantly, if the mass of the protein, or the mass of its fragments, is different from the expected value, we can match it to a modification. For example, if a protein mass is 80 Dalton more than the predicted value then we know it has been modified by a phosphate group. Examples of common protein modifications and their masses can be seen in **Figure 3**. In summary, mass spectrometry allows us to identify, sequence, and determine the modification of proteins.

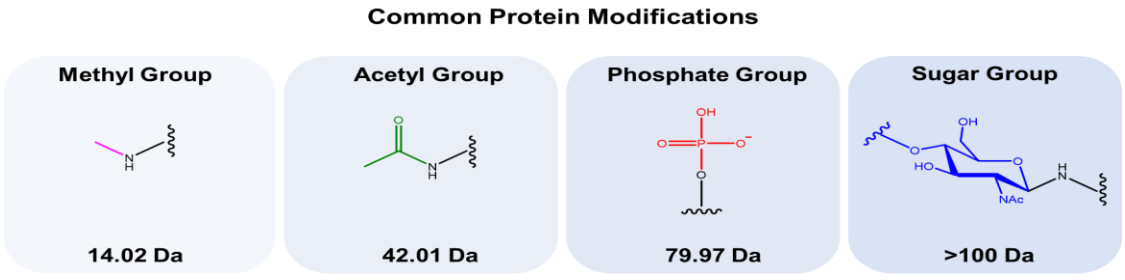

Figure 3. Examples of common protein modifications (methyl, acetyl, phosphate, and sugar groups) and their corresponding masses.

What are the challenges?

Studying proteins gives us a window into what is happening to the cells at a very fundamental level; however, there are many challenges. For example, there are roughly 20,000 different proteins that can be mutated, shortened, and modified by many chemical groups resulting in a huge number of possible proteoforms (all the different forms of a protein). The result of this protein level diversity is a huge number of possibilities and differences even between individuals. With so many proteins, we can't study all of them at once; so we have to separate them away from each other to study a few at a time. However, there is a wide range of proteins with different properties (size, charge, water-solubility, etc.), so one method cannot effectively separate them all. This is further complicated by the fact that some proteins are produced in very high numbers and others in much lower numbers. However, lower abundance does not mean they are less important.

One common way to handle this complexity is to digest (chop up into smaller pieces) proteins into peptides (small chains of amino acids). Digesting proteins into peptides makes the mixture more alike as the resulting peptides will have similar sizes and charges. The smaller size of peptides compared to proteins makes it easier to separate and measure by mass spectrometry (kind of like trying to weigh a lion vs a cat). This approach, called bottom-up proteomics, allows us to study thousands of proteins from a sample and is particularly useful for identifying changes in protein abundance.

However, there are significant drawbacks to studying peptides rather than intact protein. When proteins are digested you can't always put the sequence back in the correct order to get the true state of the proteins. The confusion regarding the true protein state that results from digestion is best illustrated in **Figure 4**. As you can see, even if we identify all the peptides associated with the protein, we lose information about how the protein was modified. The exact state of a protein

can be critical for understanding its function and role in disease; therefore, we want to precisely know its sequence and modification state.

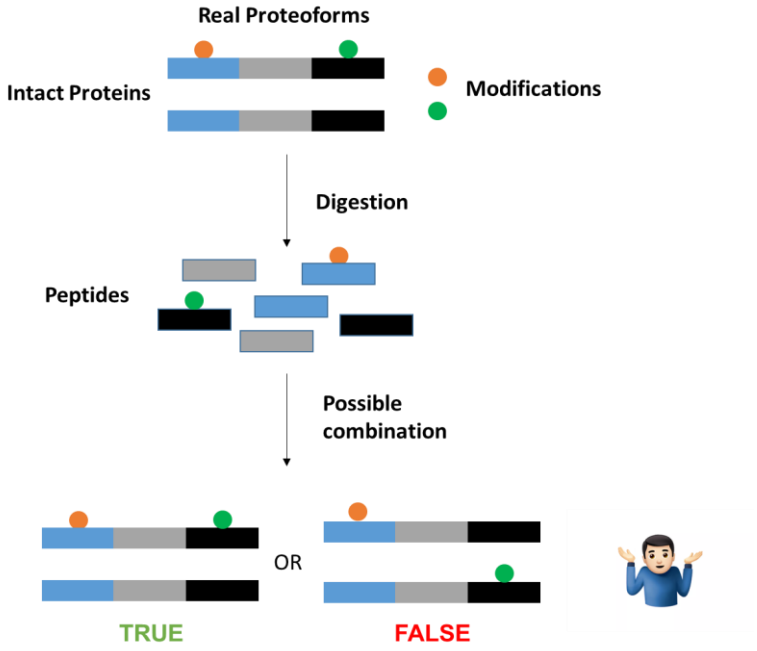

Figure 4. Diagram illustrates the challenge that arises from digesting proteins into peptides. Although we can identify all the peptides, we don't know whether the proteins exist as a doubly modified species, two single modified, or some combination. Therefore, it is best to analyze intact proteins.

What is my research about?

At this point, we've established that it is important to study proteins and that there is a great benefit to looking at them intact, but this makes them more difficult to separate and analyze by mass spectrometry. My research focuses on developing methods to better extract, separate, and sequence intact proteins to gain a better understanding of their role in diseases. More specifically, I have focused on an important class of protein located in the cell membrane. Membrane proteins account for about 50% of drug targets and one-third of our genes code for them. However, working with

membrane proteins presents unique challenges as they are insoluble in water (like oil) and there are generally fewer copies per cell relative to other types of proteins.

What is our approach?

To study membrane proteins with mass spectrometry, we must first extract them into a water solution from cell or tissue samples. Just like in everyday life, we use soaps (also called detergents or surfactants) to remove difficult or insoluble material. Soaps have unique structures and properties because they contain both a water-soluble group (head) and a water-insoluble group (tail) (**Figure 5a**). When added to water, individual soap molecules come together to form structures called micelles, whereby the insoluble tail groups are in the middle and the soluble head groups are on the outside (**Figure 5b**). A micelle is a natural capsule that can protect insoluble material on the inside and shield it from water. Using soap, insoluble proteins, like those found in membranes, can readily be dissolved just like an annoying stain. As seen in **Figure 5b**, the soap naturally forms a protective sphere (micelles) that protect the insoluble portions of proteins from water thereby allowing them to be soluble in water. Think of how oil on a pot or pan does not come off until you add soap. Not only do soaps help solubilize membrane proteins, but they also break down cells, helping release all the proteins within, making them a valuable tool for studying proteins.

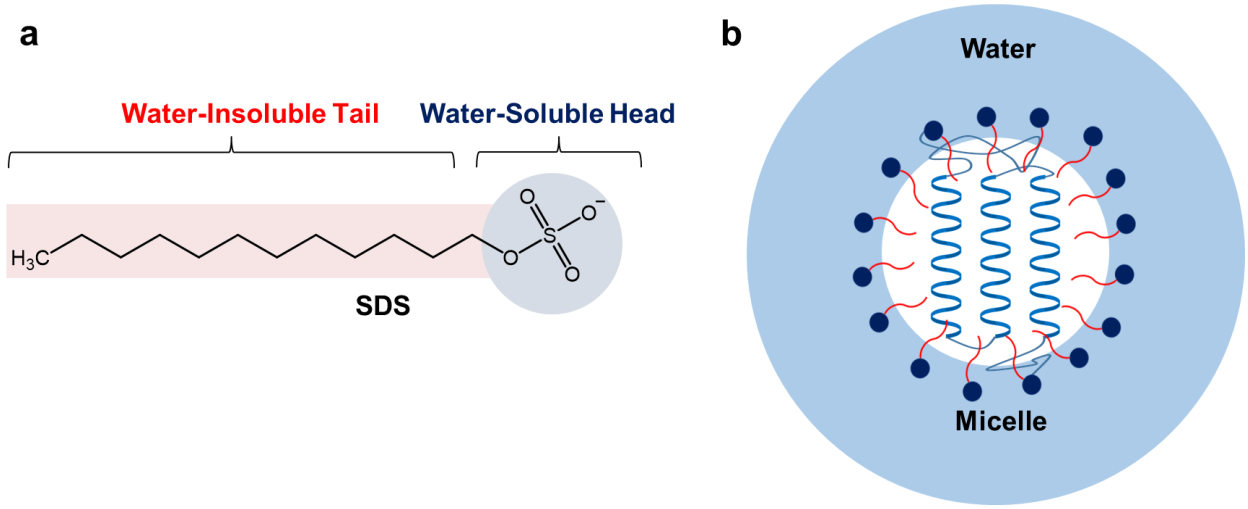

Figure 5. Structure of a common soap, sodium dodecyl sulfate (SDS). Illustration of how soap surrounds proteins by forming a micelle, thereby protecting the hydrophobic portions from water and keeping it in solution.

Unfortunately, the presence of soap inhibits our ability to measure the mass of proteins, because they are surrounded by soap molecules and "trapped" in the micelles. Soap is also challenging to remove since it surrounds and interacts very strongly with proteins. Therefore, I engineered a soap with a photocleavable bond (that is, it breaks apart/degrades with light) between the head group and the tail group. The photocleavable group enables us to extract proteins and then quickly degrade it using UV light thereby "freeing" the proteins (**Figure 6**). With the soap degraded it can be easily removed allowing us to sequence proteins without interference. Overall, the photocleavable soap technology allows us to analyze many hard-to study proteins, like membrane protein, that were previously insoluble in water, bringing us closer to the goal of being able to characterize the entire human proteome.

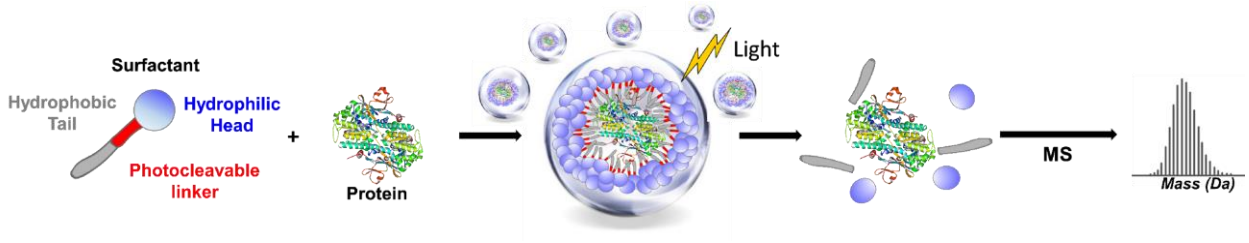

Figure 6. Photocleavable soap aids in breaking down the cell to extract and solubilize proteins. Afterward, the soap can be easily degraded with UV light allowing for protein characterization by mass spectrometry.

With an improved method for extracting proteins, I next developed different ways to separate proteins. Protein separation is generally accomplished using chromatography. A protein solution is introduced to solid particles that can be made of a variety of materials, but most commonly consists of silica particles (similar to sand) coated with carbon (**Figure 7**). The proteins interact with the solid particles with varying degrees of strength. As the liquid is added, it washes the proteins off the particles. Thus, proteins that have weaker interactions with the particles come off first, followed by the proteins with stronger interaction that "hold on" the longest and come off last. An example we commonly use to demonstrate chromatography is putting dye on a coffee filter then introducing it to water. You can then see the dye start to separate into its different components based on differing interaction strengths they have with the paper (**Figure 7**).

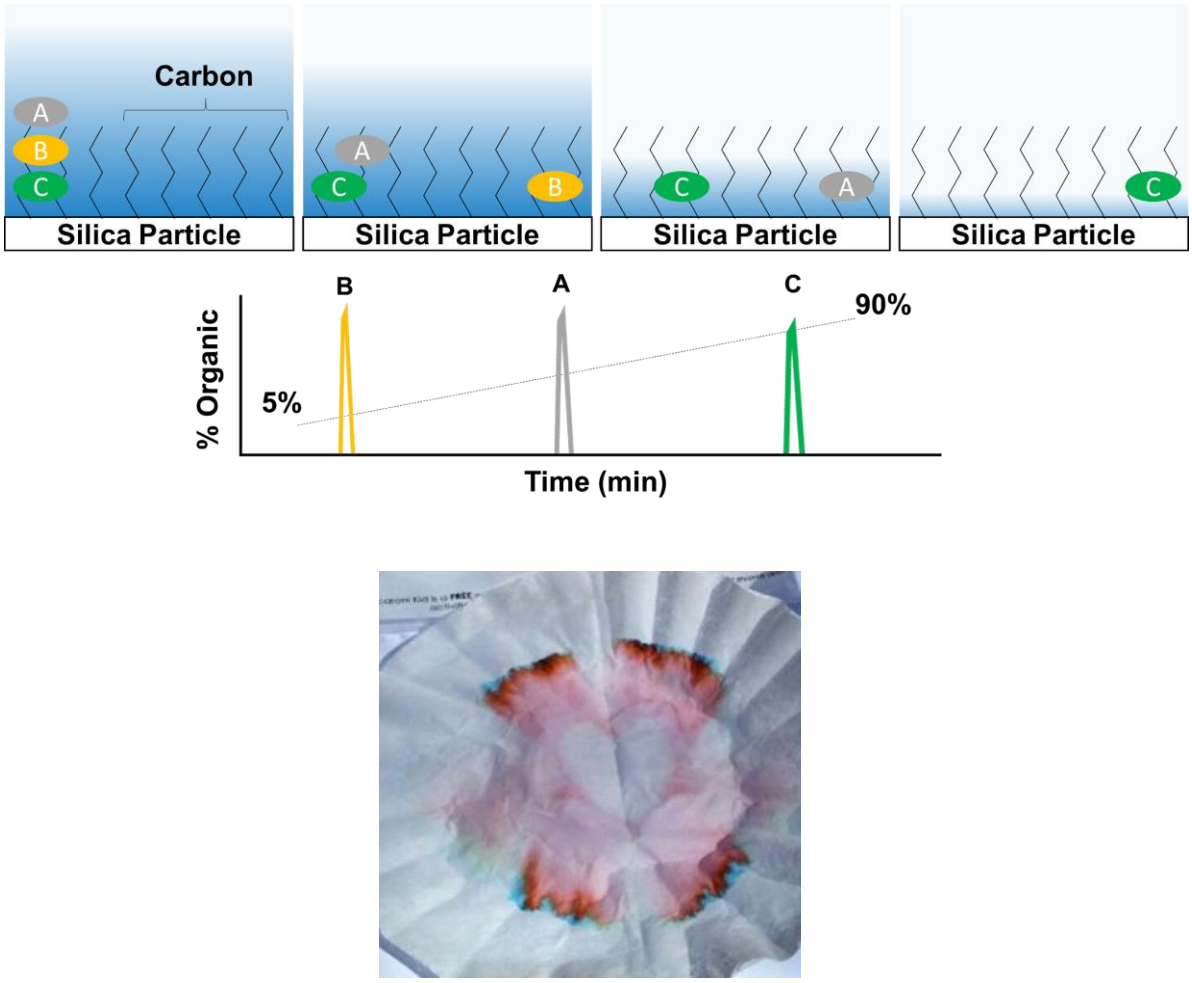

Figure 7. Illustration of separating proteins by chromatography (top). Proteins bind to the carbon with different strengths. As liquid is added and it washes the proteins off the solid. The proteins that interact the strongest stay on longer while the proteins that interact less come off first. A common demonstration of this is separating dyes using a coffee filter (bottom). As you can see, different dyes migrate more or less once the liquid is added. The different colors separating from each other based on the strength of their interaction with the filter.

Another common chromatography technique to separate proteins is called size-exclusion chromatography. In this technique, proteins are introduced to porous, non-reactive solid particles. Bigger proteins will not be able to enter the pores (they will be excluded) and therefore come out earlier while smaller proteins will enter the pores and come out later. Size-exclusion

chromatography allows us to sort proteins based on their size, making it easier to study more proteins from a sample. My research combined multiple different chromatography techniques to increase the number of proteins we could study. Additionally, I had to optimize them for membrane proteins, so that I can study hundreds of proteins from a sample after using the photocleavable soap to extract them.

How does this help disease research?

Our lab focuses on studying proteins involved in cardiac diseases. We want to understand how proteins change as cardiac diseases progress or after a major cardiac event (e.g., a heart attack). A better understanding of a protein's roles in cardiac disease will allow us to identify the ones that cause dysfunction or that are most affected by the disease, resulting in long-term health consequences.

The technology I developed helps extract and separate membrane proteins from heart tissues to discover new protein changes that result from or contribute to the disease. We commonly extract proteins from animals where we have simulated a specific heart disease or even from donated patient samples. We then look at how proteins are changing before and after the disease (**Figure 8**). This knowledge allows us to determine what role certain proteins play in causing diseases. Once we identify a specific protein (or proteins) involved, we can target them to better detect the disease in its early stages or to develop better treatments for the disease. Because there is protein variability from person to person, being able to sequence an individual's proteome could also allow us in the future to design person-specific treatment plans that enable precision medicine.

Figure 8. Illustration of a protein modification change that can be a factor-contributing factor to disease.

What is the next step?

Moving forward, the technology I developed in my PhD research for extracting and separating membrane proteins can be applied to almost any disease to help discover how certain proteins are being altered. Since membrane proteins are common drug targets, it is critical to be able to determine their role in diseases.

Although I was able to improve the separation of intact proteins, enabling the identification of hundreds from a given sample, there is still significant work towards the ultimate goal of being able to sequence the entire human proteome. I anticipate further development in protein separation will significantly improve our ability to characterize protein changes.

During my time in graduate school, I have had a great experience learning about proteins, diseases, and much more. I am very excited to move on to a postdoc position where I will continue to study proteins and their role in diseases. I am very thankful for the mentorship I have received and the friends I have made!

Images credit

Figure 6: Michelle Melnik. Paper Chromatography Coffee Filters. https://deerfield.macaronikid.com/articles/5826504224d8f4da3aafb36e/paper-chromatography-coffee-filters. (Accessed May 28, 2020).