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

ADVANCING MASS SPECTROMETRY-BASED PROTEOMIC ANALYSIS STRATEGIES FOR THE INVESTIGATION OF HUMAN HEALTH AND DISEASE

by

Justin McKetney

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

Doctor of Philosophy

(Biochemistry)

at the

UNIVERSITY OF WISCONSIN–MADISON

2021

Date of final oral examination: 05/14/2021

The dissertation is approved by the following members of the Final Oral Committee:
Joshua J. Coon, Professor, Biomolecular Chemistry
Melissa M. Harrison, Associate Professor, Biomolecular Chemistry
Anjon Audhya, Professor, Biomolecular Chemistry
Mark E. Burkard, Professor, Medicine
John M. Denu, Professor, Biomolecular Chemistry

Title: Advancing Mass Spectrometry-based Proteomic Analysis Strategies for the Investigation of Human Health and Disease

Introduction

All human beings begin life as a single cell. The cell eventually divides into two cells, then four cells, all with nearly identical DNA. Over time these cells begin to diversify, some become skin cells while others become lung cells, organs with two very different appearances and functions. Yet these cells maintain their DNA similarity with only small differences that arise from errors in DNA replication as cells divide. Understandably some mistakes are made when copying 6 billion bases to replicate the DNA. If each of these bases were a word, it would be the equivalent of writing out Leo Tolstoy's War and Peace more than 10 times over. So how does the body develop the diversity to form these different cell types, when all cells read from the same DNA script? It would be like Mercedes Benz making every automobile they offer from a single blueprint. This differentiation occurs in part through the control of proteins, the molecular machinery of the cell, which are encoded in the genes of the DNA.

Where do proteins come from?

Proteins are made up of long chains of amino acids chemically bound together. Some people might be familiar with amino acids as a dietary supplement. Protein expression describes the generation of these proteins from the corresponding code in the DNA. Each cell must translate the nucleic acid sequence that make up the genes within DNA into the sequence of amino acids that make up proteins, similar to translating a book to another language. Although DNA is made up of sequences of four possible letters, or bases, Adenine (A), Guanine (G), Thiamine (T) and Cytosine (C), proteins are made up of a sequence of 20 possible amino acids. Three letter segments of nucleic acid encode a single amino acid letter to allow for this increase in vocabulary (**Figure 1**). Segments of DNA are used as a template to first generate an intermediate nucleic acid called RNA, before being translated into protein sequences (**Figure 1**).

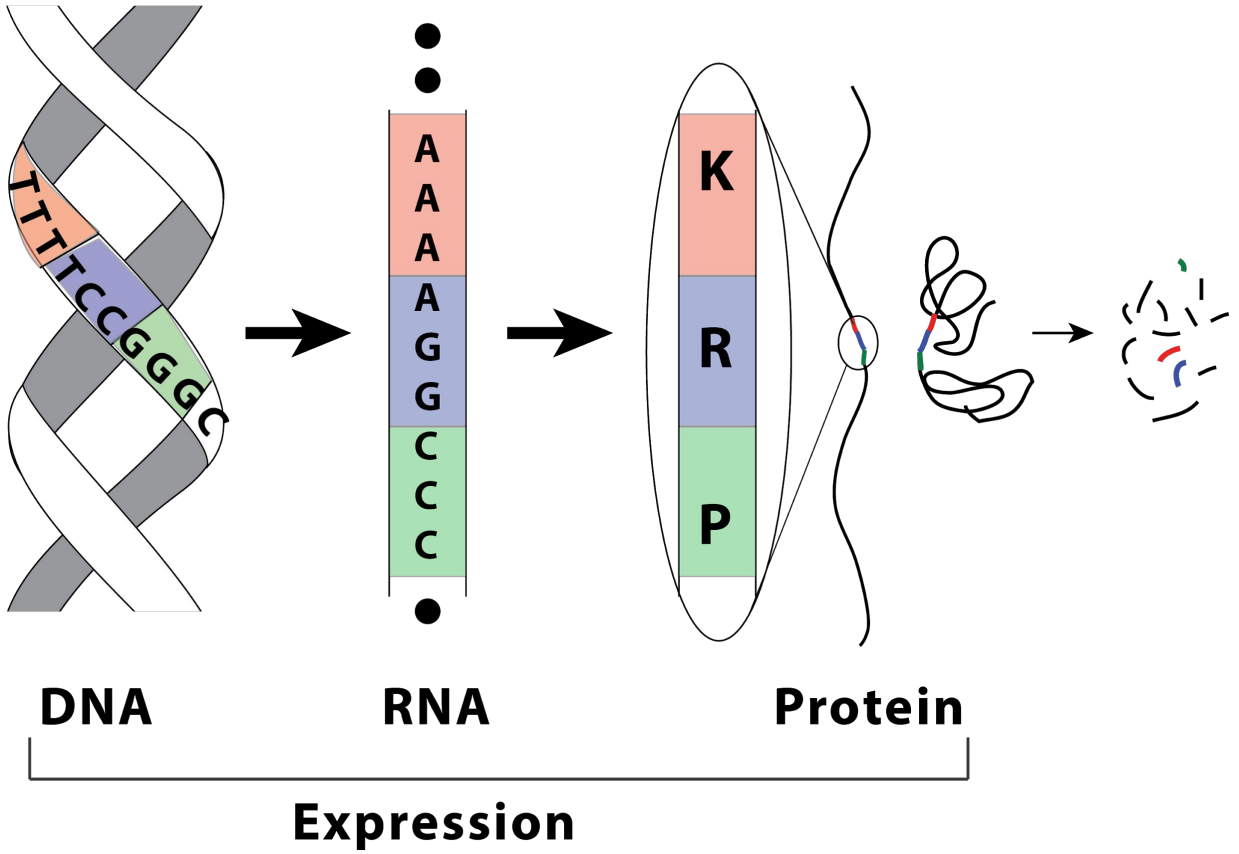

Figure 1: Translation, the process of generating proteins from DNA. Gene sequences within the DNA are first converted to sequences of RNA, another nucleic acid with a highly similar composition. This RNA is then used as a template for translation into the amino acid sequence of proteins. The protein strand folds into specific shapes that allow proteins to perform their functions. When proteins are no longer needed or start to malfunction, they are degraded back to their amino acid components by degradation machinery.

Many factors influence the quantity of proteins in a cell or tissue. Specific protein enzymes read the DNA, translate new proteins and degrade old proteins, controlling the protein life cycle. The accessibility of a DNA segment for reading, the speed of translation, and the speed of degradation all impact a protein's abundance (**Figure 1**). Some proteins travel to certain parts of the cell or exit the cell entirely. The different presence, levels and location of proteins in a cell differentiates the appearance and function of myocardial cells in the heart from neurons in the brain. By measuring the changing abundances of these proteins, researchers are able to better understand how cells function and whether they are sick or healthy. As a single cell divides, the resulting differential protein expression guides cells to form our lungs, hearts and brains. These proteins help form our "molecular self" and in organs like our brain they help form our memories, emotions and personalities. Similar to DNA, proteins make up who we are.

Proteins in the brain

In the brain, specialized proteins and protein structures allow the transmission of thoughts, storage of memories, control of movement, and the processing of sights, tastes and smells. Proteins extruded from the cell build scaffolding that guides the growth and spread of the tiny tendrils of neurons as the brain develops and matures[1]. This guide functions like a roadmap for growing cells, allowing the formation of the diverse structures of the brain. Protein receptors serve as docks for signaling molecules in the brain. These receptors allow brain cells to communicate feelings of happiness or anxiety, like telephone lines connecting buildings. Scientists have shown that proteins play a key role in storing our memories, with alterations in proteins leading to loss of memories in mice [4,5]. As you are reading this, protein shuttles transport the signal that tells your hands to move and allows your brain to process words. When researchers constructed a compendium of human proteins from 44 tissues, more than 30 proteins were found exclusively in the brain[6]. All of these molecules work cooperatively in your brain to construct your personality, and allow you to move, think and feel every day. Your brain is an incredibly precise instrument that is tuned and operated based on changes in protein activity, location, and abundance. Due to this precise operation, changes in protein characteristics can lead to disastrous and deadly diseases in the brain, such as Alzheimer's disease (AD).

Proteins in Alzheimer's

The two primary proteins in AD, tau protein and amyloid beta protein, damage the brain by clumping together in a process called aggregation (**Figure 2**). As AD progresses, sticky, thread-like tendrils of tau protein spread within neurons, eventually leading to cell death, through a largely unknown mechanism. Outside of the cell, amyloid protein aggregates to form large sticky clumps to which other proteins and amyloid molecules adhere. Many times, these clumps grow large enough to impede communication between neurons, like sticking a piece of gum on a circuit board. Interneuron communication plays an important role in transmission of thoughts and emotions as well as promoting growth of the neurons. Besieged by growing clusters from both inside and outside, many neurons will cease to function and die as AD progresses, inhibiting cognitive abilities. The death of these neurons has disastrous consequences as mature neurons no longer divide, meaning this cerebral deforestation is largely irreversible. The brain attempts to fight back against this aggregate onslaught, marshalling immune cells to remove the amyloid aggregates and damaged or dying neurons, like bulldozers removing debris of a demolished building. These cells are often overwhelmed by the sheer scope of growing protein clusters[7,8]. The challenging nature of this battlefield is reflected in the fact that no drug currently exists to substantially perturb or reverse the death of these brain cells, or neurodegeneration, associated with Alzheimer's. Understanding how different cell populations change and work to remove these disease aggregates plays a crucial role in advancing treatments for Alzheimer's disease. By measuring the protein levels in the nervous system, we can identify AD-related changes in these cells.

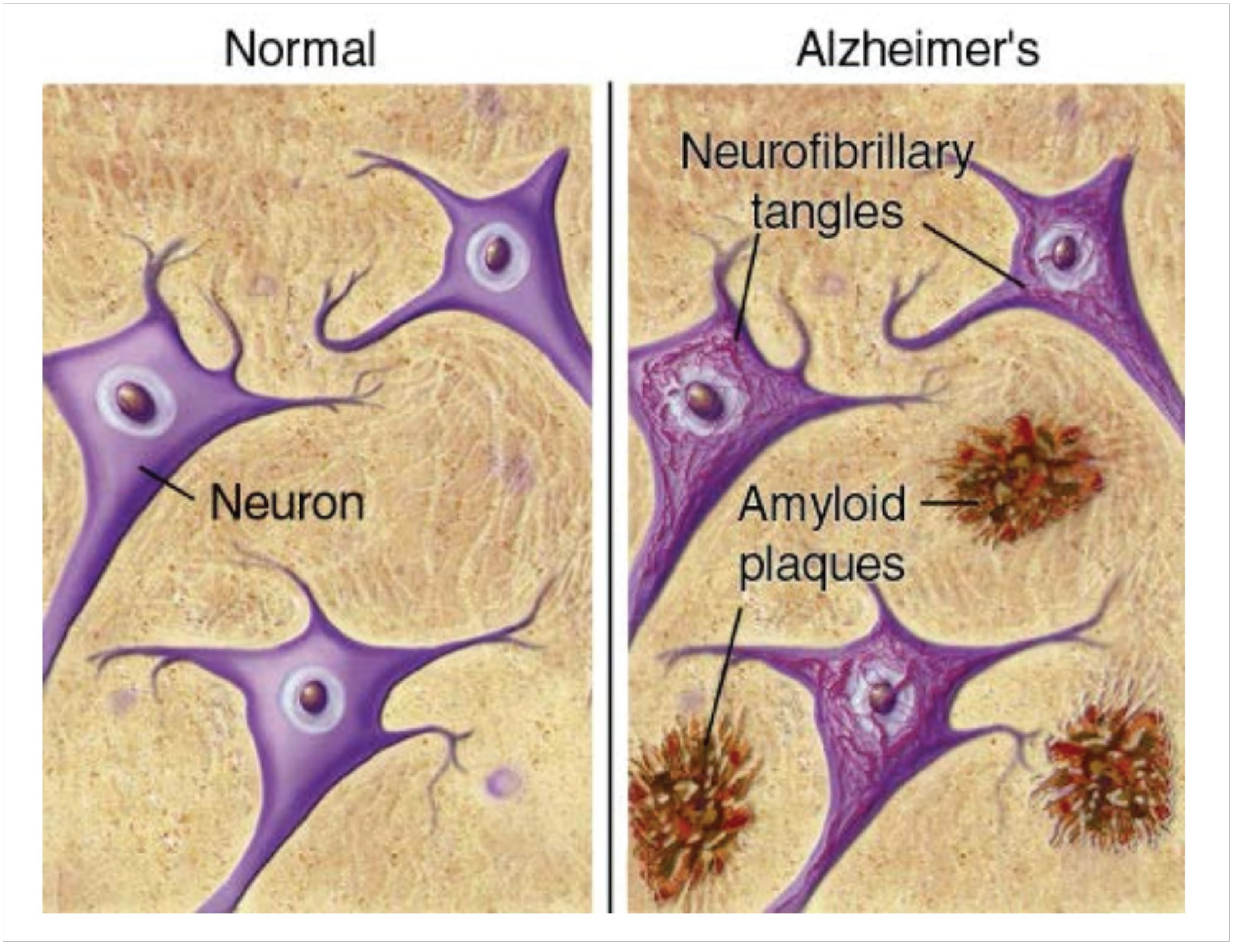

Figure 2: Aggregate formation in Alzheimer's disease. Tau protein clusters together inside of cells to form strand-like neurofibrillary tangles, while amyloid protein forms amorphous clumps outside the neurons, the primary cell of the brain.

Although our analysis occurs at a molecular level, that should not undercut the devastating cost of Alzheimer's on afflicted individuals and their families. In the US alone, an individual is diagnosed with Alzheimer's almost every 65 seconds[9]. As patients develop these protein clusters, they begin to struggle to identify faces and remember people. They find it more difficult to think of words and to speak. Many times, afflicted individuals will eventually struggle to dress themselves and perform basic functions, like climbing stairs [10]. Yet other patients, with similar genetic backgrounds or family history, show less severe symptoms or symptoms that advance more slowly[11–13]. The diverse effects and progression of AD makes developing treatment challenging[14], but also provides hope that AD can be managed by drugs or other therapies. If the field can identify the protein expression differences between rapid and slow declining groups, we can target the affected systems with drugs to slow the disease. We investigate the protein differences and similarities caused by disease using an analytical technique called mass spectrometry.

Quantifying protein levels with Mass Spectrometry

Mass spectrometry identifies and quantifies proteins by relying on the predictable behavior of charged molecules, or ions, in an electric field. If we think back to our introductory physics course, or our younger siblings' first relationship, opposites attract. This means that a positively charged molecule in an electric field will move towards the negative pole (**Figure 3a**). The speed of the molecule will depend on two factors: the mass of the molecule and its charge. If we imagine the molecule as a car (**Figure 3b**), then the charge would be the number of wheels. If the charge is higher, then the force pushing the car increases, but if the car is heavier (greater mass) it takes more energy to speed it up and slow it down (**Figure 3 c,d**). Due to this relationship, the ratio of mass to charge can be determined by applying an alternating electric field of known strength and measuring how ions accelerate and decelerate. Given that charge is a whole number, ions can often be identified using deductive reasoning for very simple mixtures with compounds of known mass.

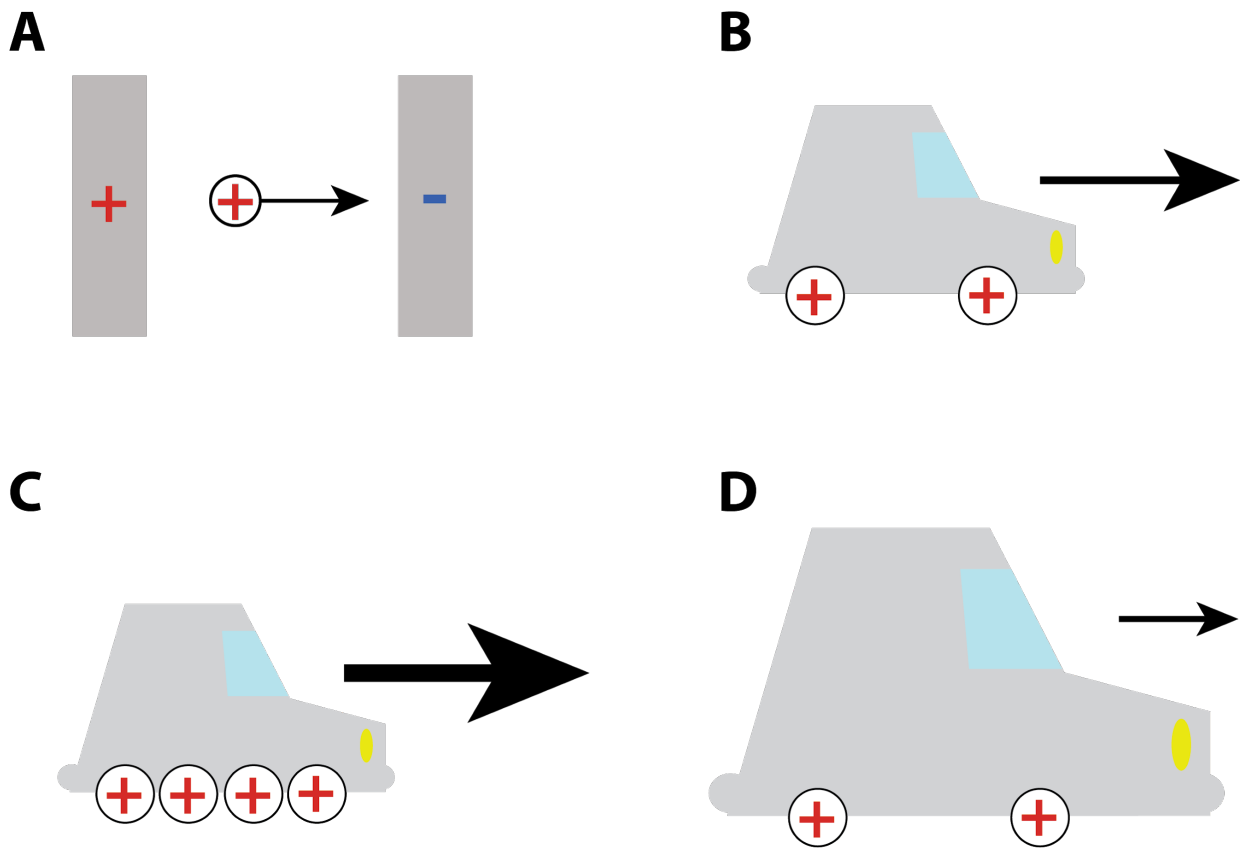

Figure 3 Predictable movement of charged molecules. (A) Positively charged ion moves towards negative pole in an electric field. (B) If we imagine our charged molecule or ion as a car, wheels(charge) and weight(mass) of the car determine how fast it can accelerate (C) If there is more charge but same weight, car can accelerate much more quickly (D) but a much larger car with the same number of wheels accelerates much more slowly.

We prepare samples by first breaking protein sequences into smaller sequences called peptides at specific amino acids (**Figure 4**). We then spray the peptides as tiny, charged droplets into the mass spectrometer (**Figure 4**). As the droplets fly through the air, they lose liquid until only the charged peptide remains (**Figure 4**). Inside of the instrument we determine the different peptides' mass-to-charge ratio. Once we have determined the mass-to-charge value we match that against a database of all human proteins and their component peptides in order to identify the peptide. The mass spectrometer uses an electric field to guide the peptides to strike a detector, which generates a signal proportional to the number of peptides colliding with it. We compare the relative abundances using this signal level. We determine protein relative abundance from their component peptide abundances.

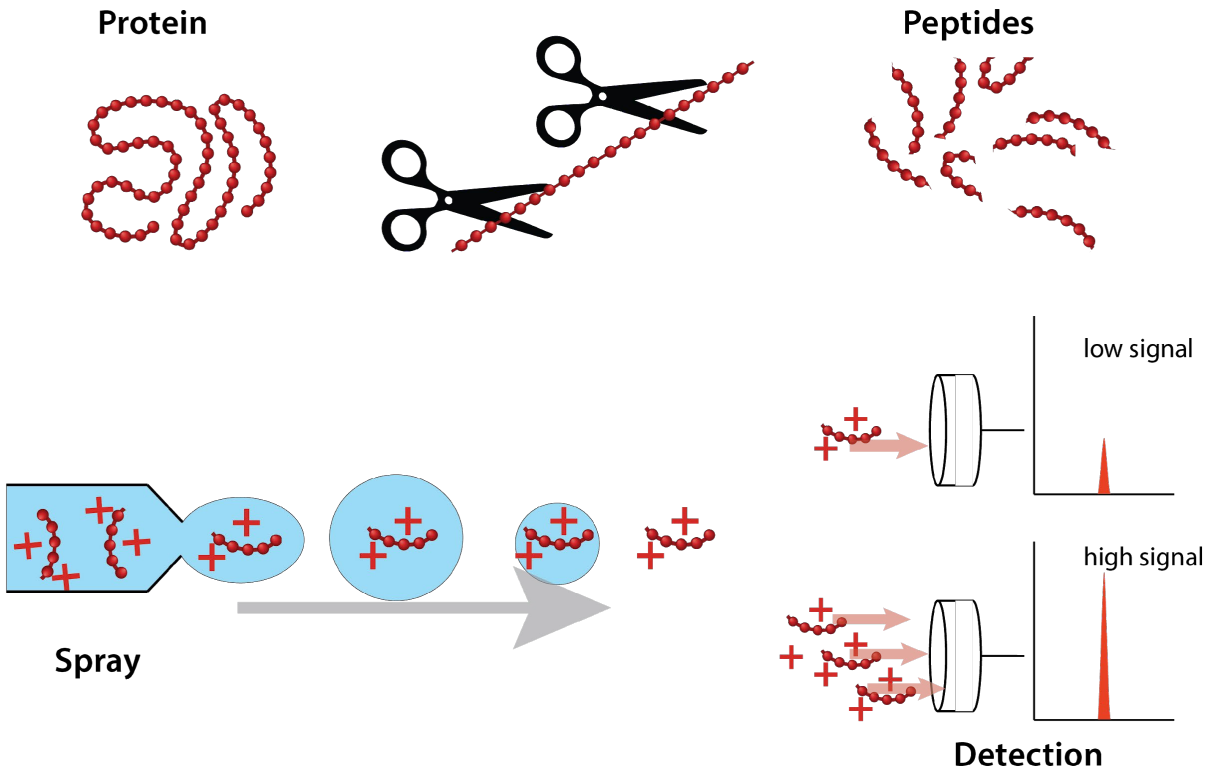

Figure 4. Detecting protein amounts. Proteins are first cut into shorter sequences called peptides. These peptides are then sprayed in charged droplets which dry as they fly through the air, eventually losing all of their liquid and leaving behind the charged peptide. Inside of the mass spectrometer, these peptides strike a detector and generate an electrical signal proportional to the number of peptide ions.

Application of Mass Spectrometry to Alzheimer's disease

We apply these protein measurements to the AD problem in two ways: by examining the different AD effects in different regions of the brain and by aligning the protein changes to clinical symptoms.

Regional Brain Protein Analysis

A diverse set of structures and cell populations make up the different regions of the brain similar to neighborhoods in a city. The cerebellum, which controls aspects of movement and language, contains a high density of cells, like the busy downtown. While the caudate nucleus, which controls aspects of learning, consists of a long, stranded structure like a coastal peninsula (**Figure 5a**). Both of these structures contain differing neuron populations, which facilitate their different functions. The caudate nucleus houses a high proportion of spiny projection neurons, while the cerebellum contains many more Purkinje cells, a highly branched type of neuron. Differences like this exist for many of the structures and areas of the brain. Given these differences, we hypothesized that the damage caused by the protein aggregates described above may lead to unique responses for the different areas of the brain. We also theorized that differential regional effects in brain proteins could contribute to the wide diversity of clinical symptoms in AD. We identified overlap between proteins specific to certain brain regions and those affected by disease, suggesting that regional disease effects do exist. We also identified region-specific proteins that were altered by the normal aging process, a process closely intertwined with the development of Alzheimer's.

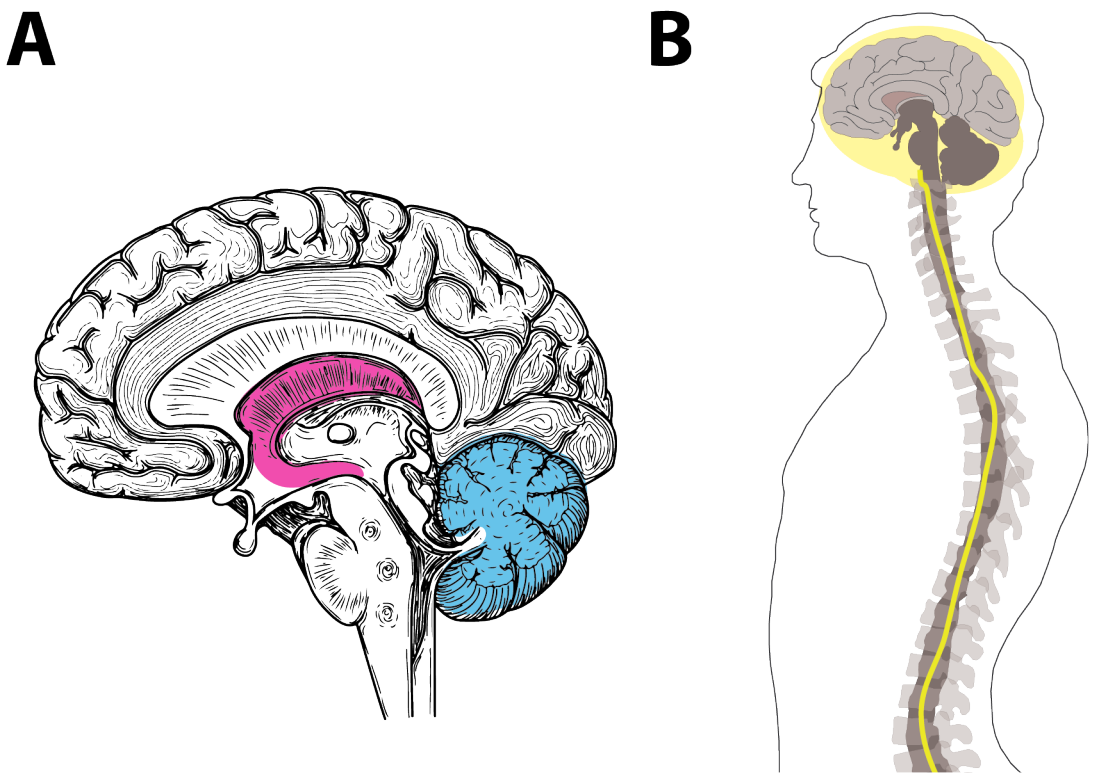

Figure 5: Diverse structures of the human nervous system. (A) Different regions of the brain take different shapes, with the cerebellum, pictured in blue, roughly circular and highly dense, and the caudate nucleus, pictured in pink, more strand-like. (B) Cerebrospinal fluid, indicated in yellow, surrounds the brain and flows up and down the spinal cord, delivering nutrients and removing waste.

Progressive Analysis of Cerebrospinal Fluid

Our investigation of brain region-specific effects in Alzheimer's relied entirely on post-mortem tissue. This type of analysis provides only a snapshot of the final stage of the disease, forcing investigators to piece together processes from the destruction left behind, like detectives at a crime scene. Developing effective treatments for AD requires the ability to detect, identify, and diagnose the disease in its earliest stages, based on both clinical symptoms and molecular indicators. This requires analysis tools and strategies with the capacity to measure protein changes from living participants as disease develops and progresses. We sought to meet this need with an analysis of cerebrospinal fluid, which can be collected from living participants, as they develop symptoms or age normally. Cerebrospinal fluid (CSF) flows along the spinal cord and surrounds the brain, physically protecting the brain, supplying nutrients and removing molecular waste (**Figure 5b**). By studying the CSF, we can better understand the function of the brain as a whole in both healthy and diseased states. We performed a pilot analysis of CSF from twenty individuals, half of whom were diagnosed with AD, in order to test our capacity to detect protein differences. We identified more than 80 proteins to be associated with disease. Starting from vials of CSF, we prepared the samples and collected and analyzed the data in only five days. Rapid analyses of this type allow researchers to expand the number of samples analyzed in a study or clinical trial, leading to more representative results. This analysis strategy provides excellent scalability, allowing the theoretical analysis of more than 100 samples in less than two weeks.

Conclusions and Future Directions

The brain serves as a vault in which we keep our most treasured pieces of ourselves, our memories, our experiences, our emotions. Although all human brains perform the same basic functions, each is incredibly unique and completely irreplaceable. The brain inspires wonder at the impressive feat of bioengineering that allows it to function, and fear surrounding our own molecular mortality. While it is one of the most heavily studied organs, there is still so much we don't know about the brain. This mystery becomes a challenge when studying diseases that affect the brain such as Alzheimer's. Our findings provide general information about the cellular effects of AD, as well as additional foundational tools that could help future researchers develop precise and personalized treatments.

The distinctiveness of each human brain leads to unique effects of neurodegenerative disease regarding both symptoms and progression. Currently, Alzheimer's works in a devastating and insidious manner, moving across the brain, leaving damaged and dead neurons in its wake. The unpredictability of its progression only adds to the emotional trauma of patients and their loved ones, making it difficult to know what symptoms will develop and how rapidly they will worsen. This uncertainty also presents a challenge to researchers and doctors, as they attempt to develop treatments that prevent and reverse the advancing neurodegeneration. As different brain regions are affected by disease, they cause different symptoms, while the widespread nature of aggregates drives the severity of these symptoms. Our streamlined analysis of cerebrospinal fluid in AD allows for the construction of a disease timeline, mapping specific protein changes to the different stages of disease and symptoms in the clinic. This analysis

allows for more than 100 samples to be analyzed in less than two weeks. When combining this timeline with a brain region-specific atlas, AD progression can be tracked by both severity and location. In the short-term, this information will help neurologists prescribe preventive measures using better predictions regarding disease developments. In the long-term, as researchers develop compounds and drugs to inhibit or alter specific processes in Alzheimer's, these drugs can be strategically targeted to patients that would benefit most. Protein resources such as the atlas and timeline can be combined with high-throughput genome sequencing to create detailed and personalized molecular snapshots for individual patients. Diagnoses of AD-associated dementia can then be paired with highly specific drug and lifestyle regiments to allow for improved outcomes. Information about affected regions, molecular timelines, genetics, and clinical symptoms can be used to build a map of the neurodegeneration landscape, allowing medical practitioners to equip brain cells to succeed against Alzheimer's with tactical precision.