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

STRUCTURAL ANALYSIS OF A UBIQUITOUS PROTON PUMP BY MASS SPECTROMETRY AND CRYO-ELECTRON MICROSCOPY

By

Matthew R.F. Blackburn

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Biochemistry)

at the

UNIVERSITY OF WISCONSIN-MADISON

2024

Date of final oral examination: 2/19/2024

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

Sebastian Y. Bednarek, Professor, Biochemistry
Katherine A. Henzler-Wildman, Professor, Biochemistry
Lingjun Li, Professor, Pharmaceutical Sciences and Chemistry
Alessandro Senes, Professor, Biochemistry
Michael R. Sussman, Professor, Biochemistry

5 ADDRESS & SUMMARY FOR THE GENERAL PUBLIC

5.0 An open letter to the public. My two siblings and I were raised by a single mother who was in constant custody battles with our (now long estranged) alcoholic father. While things were hard, science provided an escape for me. I've also always been a curious person, and the evidence-driven tendency of science's methods towards truth was reassuring to me. *Bill Nye the Science Guy* and PBS' *Nova* were the meat and potatoes of my television diet, and I read every children's science book I could get my hands on. I shared nearly everything I learned to willing (and unwilling) audiences alike. If I read about how a cone snail eats clams by drilling into them, you were going to hear about it. Did you know Jupiter's moon Io is covered in enormous lava lakes and its eruptions can be seen from space? You do now! (You're welcome). I was a manic trivial fact-finder and probably annoyed a lot of other kids with that but oh well, I couldn't help myself; this was my outlet. In 7th grade, I finally knew what I wanted to do with my science obsession. We were learning *cytology*, (*cyto*- "cell", -*ology*, "study of"), and while I knew by then that living material was built from cells, I had no notion of what that meant. Realizing that cells are such dense hives of activity blew my mind. Picture the busiest metropolis you can think of – perhaps New York. Contemplate its electrical grid, shipyards, railways, and airports. Now consider that inside of a living cell, you have the same type of specialization in both space and structure – but that it's molecules, and that it's also far, far more dynamic. This is what made me realize that I wanted to be a scientist, and over the years that interest cemented into biochemistry as a specific career choice.

Like any highly technical discipline, biochemistry is not something that can be thoroughly explained in an elevator pitch. I have long believed the dictum “if you can’t explain it well, then you don’t completely understand it”, but at the beginning of my PhD, I naïvely expected that it would become easier to explain my work as I became more versed in it. Last year, I confessed to my younger brother that it was getting harder for me to efficiently summarize my research to people outside of my field. The context needed context, and even that needed context! He replied, “It seems like for any very technical field, the further in you go, the more [conceptual] distance you have between you and people not in your field. It’s like you’re speaking a different language.”

Well, it turns out he’s right. You, the public, are entitled to the knowledge and theory produced by scientific effort and also to understand it, because you are in equal parts both benefactor and beneficiary of basic scientific research. Your dollars funded everything we now know about insulin (which only costs a few dollars to make *per vial*) and the sequencing of the human genome, which birthed many of the biotechnologies of the 21st century. Your investments are too numerous to list because basic scientific research is foundational to nearly every significant improvement to the human condition around the world across history, and you are entitled to this knowledge and its products.

Sadly, there are often very wide gaps in the general public’s understanding of the value of basic research. The press is ill-equipped to accurately report scientific findings in the sober, neutral tones that are demanded by the scientific process when results are appraised. Sensationalism simply sells better, and chasing a hot scoop leaves little time to get into the deep technical detail that is demanded by professional scientific practice. Scientists, for their part, are often unable to make the time to engage in public outreach,

and developing the skills to engage with the public and bridge that gap in communication takes time and energy that they often don't have. Everybody is just too busy to talk about it!

Scientists and the public need each other for progress to continue. In the absence of communication, conspiracies and disinformation spring up and cause real damage, where great benefits could be realized instead. History is rife with examples where science has benefitted humanity greatly, but there are also many stark examples where intentionally ignoring scientifically vetted research (or outright distorting it for political advantage) led to dire consequences. We don't need to look very far back to see how bad actors often take advantage of the gap between scientific work and how it is communicated to the public (if at all). In the last few decades, the United States negligently botched policy responses to both the AIDS epidemic and the COVID pandemic. Ignorance and malignant refusal of scientific expertise by the leadership of this nation prevailed in each case, leading to unconscionable and completely avoidable losses of American life. And just last year, the ocean waters off the coast of Florida *boiled* and wildfires became so common around the world that our own skies were darkened with smoke from another country. And now, a significant population of this country contends that truth is not truth, evidence is in the eye of the beholder, and insurrection isn't insurrection. Americans value freedom, but it cannot come at the cost of intelligence, competence, safety, basic decency, and human rights.

Science is the purest form of empirical practice. A society which values science benefits from building itself with policy and technology that are well-vetted by evidence. The project of scientific progress must continue, and for it to succeed it must be boosted

in the public consciousness. As the great scientific orator Carl Sagan noted, “it is better to light one candle than to curse the darkness.” In that spirit, I am extremely grateful to WISL for this opportunity. Without further ado, I will highlight some core concepts that will be useful for the reader to keep in their back pocket, and then I will dive into a high-level summary of my thesis work. Through this chapter, I hope I can provide a small glimpse into the world of basic scientific research and the value it provides.

5.1 Biochemistry: the “sparknotes” version. Biochemistry occupies a space between the life sciences and analytical chemistry. Another way to think about biochemistry is that it is the application of methods and concepts in chemistry to study and describe biological phenomena at the molecular level – that is, what are the molecules of biology? How do they react with each other? How does that chemistry contribute to what we understand as “Life”? Living systems are both highly diverse and very complex. As we become more adept at exploring and studying our planet, we continue to discover new environments with exotic organisms that sometimes defy prediction. All of this means that biochemistry’s scope is massive and will only continue to grow, and its adoption of emerging technologies from other fields (like chemistry and physics) have led to biochemistry becoming highly interdisciplinary today.

Biochemistry’s principles are applicable to all forms of life as we know it. The stuff of life – the molecules that are the bricks, mortar, timber, plumbing and engines of living cells – is built from large, carbon-based, complex molecules consisting principally of hydrogen (H), carbon (C), nitrogen (N), oxygen (O), phosphorus (P), and sulfur (S). These elements are frequently built into four types of large molecules, called

macromolecules[1] (listed in order of increasing size and complexity): lipids, carbohydrates, nucleic acids, and proteins. **Lipids** include steroids, cholesterol, and fatty acids, and can be used as hormones, energy storage, and for building the membranes of cells and cellular organelles. **Carbohydrates** are essentially sugars and their polymers, which include cellulose (one of the main components of wood), sucrose (table sugar), starch, glucose (the preferred energy source of our cells), and hyaluronic acid (the gel-like material in our eyes and joints). **Nucleic acids** are more complex than carbohydrates and lipids, and may be far larger, most often functioning as information storage (DNA, genes, and RNA), but can also act as scaffolds or signaling molecules (noncoding RNAs), or may catalyze reactions (ribozymes and ribosomes). Finally, **proteins** – which are the focus of my research – are the cell's jack-of-all-trades molecules. Proteins can be small hormones (like insulin), act as receptors for hormones (like the insulin receptor, which detects the signal from insulin by physically binding the insulin molecule), structural cabling inside of cells (actin and tubulin), storage and packaging (like gluten and histone proteins), carriers (hemoglobin, albumin), channels, transporters, antibodies, and enzymes[2].

---

[1] **Macromolecule:** a category of molecules (often polymers) which are so "large" that they gain properties that are not found in "small" molecules with similar chemical composition. For example, water (3 atoms), carbon dioxide (3 atoms), ammonia (the original active ingredient of Windex, 4 atoms), hydrogen sulfide (2 atoms), and acetic acid (found in vinegar, 8 atoms) are all considered small molecules, and each has different properties that influence their behavior. The peptide hormone insulin is built from *all* the same elements as each of those molecules and has nearly 800 atoms in its mature form (and among all known proteins, that's on the smaller side! It's like the size difference between a brick and an entire house).

[2] **Enzyme:** a biological catalyst (usually a protein) that forces a specific chemical reaction to occur at a highly accelerated rate in a process known as **catalysis**. Catalysts participate in chemical reactions *but are not consumed by the reaction*, meaning they can accelerate a reaction to completion multiple times. Life cannot function without enzymatic catalysis because the reactions that enable life must happen at a controlled rate and be tightly coordinated. An everyday example of a *non-biological* catalyst is the catalytic converter of gasoline-consuming vehicles, which uses the catalyst palladium and/or platinum to degrade chemicals in fuel exhaust, limiting air pollution.

This list of roles that proteins fulfill only scratches the surface, and often multiple functions are bundled into the same protein. For example, the job performed by the type of protein that is the subject of my thesis work is pushing **protons**[3] (hydrogen **ions**[4]) out of the cell. That process – moving cargo from one compartment into another – might sound simplistic on paper. What the pump does to achieve this outcome is more complicated: catalyzing two different reactions, binding ions, exchanging molecules, and regulating its own activity while also acting like a gated conduit for protons to exit the cell. As protons accumulate outside the cell, they form a powerful reservoir that presses against the cell's outer membrane, very similar to how the water in a reservoir presses against the back of a dam. Just as a dam can harness the controlled release of water from its reservoir through hydroelectric turbines, the cell's membrane taps its proton reservoir to supply power elsewhere in the cell at a moment's notice. Life would have a very hard time if it could not convert and store energy in this way, and indeed, our very survival depends on this pump because it is required for plants to live. The oxygen in the air we breathe and the carbon of our food, drugs, clothes, fuel – even the wood and polymers of our homes and devices – have their origin in the photosynthetic activity of plants. And photosynthesis is made possible by the activity of this pump! Placed in this context, it's easy to see why basic research on a single protein matters so much.

---

[3] **Proton:** a hydrogen ion (H^+). Protons are also used by ion pumps in our stomachs to produce stomach acid (HCl) and in our muscles to store reservoirs of calcium ions (Ca^{2+}), which are tapped whenever we use our muscles.

[4] **Ion:** a particle – an atom or molecule – that carries an electric charge. For example, sodium chloride (a.k.a. table salt), is made up of tightly interacting sodium and chloride ions (Na^+ and Cl^- respectively).

5.2 A picture is worth a thousand words: how to visualize complex biological molecules. Have you ever wondered how we know what DNA looks like? The double helix of DNA has been established for so long that it is now fair to suggest it may be the most famous molecule in the world; it certainly is the most recognizable. The image of the double helix appears throughout pop culture with numerous references in film, marketing campaigns, television crime dramas, and even music. Its image is so prevalent that we don't ever question it, yet the groundbreaking X-ray crystallography experiment by Rosalind Franklin – which gave James Watson and Francis Crick the final piece of data needed to propose the 3D structure of DNA – happened only 70 years ago. In fact, DNA's entire history as an object of scientific investigation spans nearly two centuries, and it goes on because the combined efforts of multiple scientists and ongoing research efforts continue to show that DNA still has surprises in store for us. Regardless, the centuries of evidence contributing to our understanding of DNA and incessant re-validation of the double helix model around the world allow us to take DNA's elegant structure for granted[5]. This includes post-Millenium technological developments in **structural biology**, a field at the intersection of biology, chemistry, and physics. Structural biologists' goal is to identify the structure of a macromolecule, connect how its structural features and their movements explain known properties about the macromolecule, and finally to make educated predictions (hypotheses) about what different structural features might do. For example, the pharmaceuticals industry is very

[5] DNA is not a "naked" molecule inside of the cell, where it contacts numerous proteins, enzymes, and other nucleic acids. Whenever DNA-interacting proteins have been structurally characterized while bound to DNA, the DNA's structure is observed along with that of the protein and so far has always reinforced the 1953 Watson-Crick model even when new biochemical properties of DNA are observed. More recent examples of how DNA is continually revalidated through independent research in the modern era are the discovery and application of CRISPR and the development of mRNA-based vaccines, both of which depend on the nearly 200 years of experimental analysis of DNA.

interested in the interactions between drugs and the receptor proteins that bind them.

Among macromolecules, complex 3D structures lead to emergent properties. And why not? The shape of a spoon gives it new properties that the material it is made from does not have without shape – it can be gripped easily, it can temporarily store and carry liquids or food, and it can be packaged symmetrically with other spoons in a drawer. The same holds for a baseball glove, and indeed any structured object people can dream up. That link between structure and function applies here as it does to proteins, which have the highest structural diversity among the biological macromolecules, and the larger the protein the more complex its structure can become. At minimum, proteins comprise a single, long polymer called a **polypeptide**[6] (a continuous chain of **amino acids**) that has folded on itself to adopt a 3-dimensional structure, not unlike when a piece of string gets bunched up or knotted on itself **(Figure 1)**. A ball of yarn is a simple way to visualize this. Balls of yarn are spherical (3-dimensional) because the string (1-dimensional) has been wound on itself ("folded") many times. That 3D shape now gives the yarn ball new properties that it lacked when it was just a string – it can roll, bounce, and it drives cats mad! Just so, a polypeptide is also string-like and has no special properties beyond that, but once it is folded properly into its biologically active structure, it becomes a protein and gains new properties **(Figure 1)**.

[6] **Polypeptide:** a polymer (chain) of amino acids of varying length and sequence, which has not yet begun to fold into a **protein**, the 3-dimensional equivalent of the peptide (in addition to other modifications). The polypeptide is distinguished from other polypeptides only by its length and sequence, which is also known as the **primary structure** of the protein. Polypeptides are the molecular product when protein-encoding genes (DNA sequences) are expressed.

The structure that a protein folds into is governed by the number and sequence of amino acids making up its chain or chains. The unique properties of each of the twenty common amino acids is a story for another time, but I will say this – all amino acids have the exact same core structure, enabling them to be seamlessly linked together like pearls on a string. The defining feature of each of the natural amino acids that are used to build proteins is a **sidechain** – a branch on the amino acid that can be nearly *anything* and therefore defines the identity of the amino acid. For example, if the sidechain is simply another hydrogen atom, that amino acid is the simplest of them all: glycine. By contrast, the defining sidechain of tryptophan is a pair of joined ring structures. Thus, a polypeptide can have any sequence and proportion of the available natural amino acids (20 in total). The *length* (total number of amino acids) of the polypeptide is also variable – proteins may consist of a folded polypeptide with as few as 50 amino acids or as many as >1000 amino acids. These features all come into play during protein folding, resulting in a vast number of structural possibilities. This might seem abstract but think about it – this is just like the wide variety of buildings that humans construct. Just as skyscrapers are very different from a house or a shed, while all are each a type of building with different purposes, proteins can be small hormones like insulin or huge molecular engines that are many orders of magnitude larger and more complex than molecules like water, and everything in between (**Figure 2**).

As you can see in **Figure 2**, the structures of large biological molecules are elaborate and unique, so visualizing them can be a complicated task. In other words, it's relatively intuitive to visualize the structure of a molecule with the formula "H_2O" – a central sphere with a small sphere stuck to it on each side at an angle, looking like a

wide ‘V’ (**Figure 2A**). Things become more challenging for proteins like insulin or AHA2, where describing the molecular structure in a single image or viewing angle is not so simple (**Figure 2B, C**). Indeed, different areas of a protein often have a specific job, and their structure reflects the task, much like a baseball glove is optimized for catching and gripping a baseball.

How do biochemists visualize the structure of a macromolecule like a protein? How did I create the images in **Figure 2**? In *Ye OIden Tymes* (before the widespread adoption of molecular visualization software), physical models of the molecule’s structure were produced for handling and visualizing. For example, the original structural model of the DNA molecule was conceptualized using pieces of cardboard. Nowadays, this is accomplished using molecular visualization software (like PyMOL) which produces an interactive, 3-dimensional representation of the molecule using actual data collected from independent structural biology experiments. Molecular visualization is done using a variety of representation styles. Several such styles are depicted in **Figure 3**, which highlights a few common ways a researcher can visualize the same part of a molecule and how different styles reveal completely different types of information. In the following section, where I discuss my thesis work, you will see examples of how displaying different levels of information about a structural model can be used to convey data.

5.3 My thesis: Structural analysis of a ubiquitous proton pump by mass spectrometry and cryo-electron microscopy. The title of my thesis describes exactly what I have done: I used two methods of structural biology (mass spectrometry and cryo-electron microscopy) to analyze the structural details of a specific type of protein –

the plasma membrane proton pump of plants. As you read this, keep in mind that this work (regardless of the method) is essentially an effort to record the pump's movements and to photograph it in groups to better understand it. Let's dive in.

5.3.1 Background and relevance. Cells require lots of energy and they need to tightly manage their energy budget to ensure all necessary biochemical activity in the cell can happen smoothly and in balance. This means that cells must also be able to store and convert their available energy into other forms that can be independently used. Consider a cell that many of us may recall from high school biology: the neuron (nerve cell). Neurons are a good example of a cell that must convert a very large portion of its available energy into another form and store it for later use. Nerve signal propagation – a.k.a. the action potential – is an electrical pulse that zips from neuron to neuron along a nerve or neural circuit. That pulse is not something that happens on its own, and for it to be passed from neuron to neuron those cells each must be "charged up" (very similar to a battery) first. That charge is built up on the neuron's membrane, much like a battery is charged (filled with stored energy). Discharging that stored energy – like how a battery releases its stored energy when it powers a device – is what allows nerve impulses to occur. Nerves fire very rapidly, so a neuron's membrane needs to be "recharged" quickly and often. How does this happen? The answer is that a protein called the sodium potassium pump (Na^+/K^+ pump), which is anchored in the cell's membrane, spends energy derived from food (ATP) transporting ions (sodium, Na^+, and potassium, K^+) across the membrane, building electrical reservoirs that can be rapidly tapped later, much like a hydroelectric dam opens its sluiceways and harnesses the energy in the controlled flow of the water from the reservoir behind it. In this analogy, you could think

of these pumps as the headwaters or source of the river flowing into the reservoir behind the dam (**Figure 4**).

The neuronal Na^+/K^+ pump is an archetypal example of how cells energize their membranes, but there are many others. In plants and fungi, that job is performed by the **plasma membrane proton ATPase (PM H^+-ATPase)**[7] – the subject of my thesis. Any **ion** can be used to build a reservoir like this, and in the case of plants the solution that evolved for this job was the proton pump[8]. Plants rely so heavily on PM H^+-ATPases that they have multiple genes encoding a different version of it. For example, the plant *Arabidopsis thaliana* expresses 11 versions of this pump! PM H^+-ATPases are responsible for energizing many important processes in plants, making them relevant in both medicine and agriculture due to the myriad ways disease and environmental conditions affect crop plants through the proton pump. My thesis research focused on analyzing the various structures that AHA2 can form, both while it does the work of proton transport, and when it forms large groups known as macromolecular complexes. Plant proton pumps have been studied over several decades, but our understanding of how their structures contribute to their function is sparse, and so important questions remain.

---

[7] **Plasma membrane H^+-ATPase:** proteins are often named for the action they perform and/or characteristics observed about them. In the case of this protein, **plasma membrane** refers to the outermost membrane of the cell, where this protein is localized. **H^+** refers to the ion cargo that the pump transports across the membrane. Finally, **ATPase** uses the commonplace enzyme naming convention of adding the suffix **-ase** to the name of the compound or process the enzyme catalyzes. “ATPase” indicates that this protein breaks apart the energy source ATP to access the energy needed to push protons out of the cell.

[8] For example, fermentative yeasts produce organic acids as a waste product, and the yeast plasma membrane H^+ pump eliminates extra acidity by pushing protons out of the cell. Simultaneously, the buildup of protons outside of the cell results in a potent membrane potential that can be tripped when the cell needs to work to acquire other materials like the preferred food molecule, glucose.

First, two important regions (or “domains”, in biochemical parlance) of AHA2 are required for controlling the pumping activity, but their structure and mode of regulating the pump are still not fully understood. These regions are rope-like extensions at each end of the pump’s polypeptide: the very first ~10 amino acids and the final ~100 amino acids each constitute a flexible extension from the main structure of the pump, and those extensions have been confirmed to regulate how fast the pump does its work by tucking themselves into the moving parts of the pump, impeding its movements and shifting it into a lower gear – kind of like how the gearbox of a car controls the link between the fuel consumption by the car’s engine and the car’s actual speed. So collectively, you can think of these regulatory components of the proton pump as its “gearbox”. In AHA2, we call the shorter ~10-amino acid extension the “N-terminal domain” or NTD, because this extension hangs from the so-called “amino-, or N-terminal” part of the pump (the “beginning” of the protein, if you read its amino acid sequence of the pump from left to right like an English sentence). Similarly, the longer ~100-amino acid extension is called the “C-terminal domain” or CTD, because it extends from the “carboxy-, or C-terminal" end of the pump (the opposite end of the protein’s amino acid sequence). Second, the pump is a machine with moving parts, which means that it can adopt multiple structures, or catalytic intermediates. These intermediates have not been directly observed (we can make informed guesses, but that’s it). You may ask why we should care about this – isn’t knowing most of the protein’s structure enough?

Unfortunately, the answer is no. These terminal regulatory domains are necessary for the pump to function properly, and its catalytic intermediates (the different

shapes it adopts while it is working) are directly affected by the terminal domains. When plants fail to regulate the activity of their proton pumps, they succumb to drought, salt, and disease far more easily. For example, a fungus known as *Phomopsis amygdali* attacks peach and almond trees with devastating effect by secreting a potent chemical known as fusicoccin (FC). FC hijacks the natural function of the plant proton pump of those trees by permanently disabling the pump's regulatory C-terminal domain. The immediate result is the pump enters an irreversible, high activity state, and this ultimately leads to catastrophic wilting and tissue death in the trees. Similarly, if the reservoir behind a hydroelectric dam fills too fast, the dam's turbines could cause a power surge and damage the local electrical grid, or the dam itself could breach. This underscores the relevance of the plant proton pump's C-terminal domain in agriculture. We can't fully understand how the C-terminal domain exerts these important regulatory effects without observing its structure and precise interactions. My work is an attempt to fill these knowledge gaps.

5.3.2 Identification of a new regulatory interaction for the CTD.

Here, I used an application of mass spectrometry known as **protein footprinting** as my method for detecting structural variations in the pump because it is sensitive to small changes in protein structure, it is easy to use, and it requires far less protein compared to other structural biology methods. Here, I used a chemical treatment to add molecular extensions to accessible surfaces on the protein. Since these extensions add a specific amount of additional mass to these sites, those changes can be detected by mass spectrometry (MS) because that method excels in finding and identifying altered masses in a sample.

That's very abstract, so another way to think about it is to imagine you have a 3-year-old in the family with a devious sense of humor, and they throw paint at you, and it gets all over you. Which parts of your body get paint on them? Anywhere you have exposed skin. Everywhere else is shielded by fabric. Conceptually, footprinting is just that – "splashing paint" (footprinting agent) on a protein and "seeing" (via MS) where the paint sticks. Anywhere the footprinting agent sticks must have been an exposed surface at the time; anywhere that was shielded, or buried, at the time will not have anything stuck to its surface. A protein alternating between structural intermediates, or participating in an interaction with another molecule, will also alter its topology (exposed surfaces) in the process, and so those alterations can be detected with footprinting. This allows us to get a series of snapshots about the protein's shape at a moment in time and begin to assemble an informed description – a model – of how it operates.

What do I mean by "comparing structural intermediates" or "detecting structural shifts"? Imagine you and your family are on vacation and are posing for a group photo. First, you want everybody to pose for the "normal" photo with everyone smiling, and hopefully nobody is blinking. Next, you want someone to pinch their fingers in the air lined up with a building in the distance, to look like they're picking up the building with their fingers. And maybe, for the last shot, you want everybody to hold their favorite souvenir aloft, or hold hands. You take the shots, and now you have a collection of images of your family in various poses. If you think about your family members as 3D entities with visible surfaces, then you can also think of each group photo as an image of your family members participating in a single valid structural intermediate.

In a similar sense, I used protein footprinting combined with MS to take snapshots of AHA2 in various poses. AHA2 belongs to a huge family of transport proteins that is found throughout life – in plants, fungi, and animals, including you and me. These diverse proteins are grouped as a family because they have similar structures and perform the same biochemical reactions while they pump their ion cargo. In short, they share the same core domains. These "conserved" core domains are shown on AHA2 in **Figure 5** – red, blue, and yellow.

AHA2's similarity to many other pumps in the P-type ATPase family – like the neuronal Na^+/K^+ pump I mentioned earlier – means I could use published research on those pumps for guidance on trapping my proton pump into a specific pose. The strategy I used was to treat AHA2 with special chemicals known as **inhibitors** to bind AHA2 and to get it stuck in a specific configuration – long enough for a photo. Different inhibitors create different poses, so I used three inhibitors in my work with AHA2 and "photographed" its structural details with footprinting. I followed this up with an MS analysis to identify what changed due to the inhibitor. Two of those inhibitors are versions of the fuel that the pump specifically relies upon for it to do its work, and the last inhibitor effectively stops the pump from resetting itself after it has performed transport, jamming it. **Figure 6** highlights what I found, and I will summarize it here.

The most interesting footprints I observed suggested that, in each AHA2 intermediate, a portion of the regulatory C-terminal domain was consistently buried – in other words, I was seeing the pump's "gearbox" at work. Part of the pump's "Nucleotide-binding domain" (or "fuel tank") – named for its job of grabbing and storing the fuel ATP – shared a highly similar footprint with the C-terminal domain. This suggests a physical

interaction between the two domains at this site (zoomed-in portion of AHA2 in **Figure 3**; red portion of AHA2 in **Figure 5**; dark blue patch on AHA2 in **Figure 6**). In other words, the footprinting pattern (or “paint splash”) showed that a piece of AHA2’s “gearbox” and part of its “fuel tank” were covered up at the same time. Other regions of AHA2 showed different amounts of paint-splash, and for some poses of AHA2 these regions showed less, while others showed more. However, this coinciding pattern between the CTD and the patch on the Nucleotide-binding domain (dark blue patch in **Figure 6**) was the most consistent footprint.

This is exciting for several reasons. First, while there is currently no experimentally determined structure of the C-terminal domain (“gearbox”) for any plant proton pump, this finding illustrates a potential site – part of the “fuel tank”, or Nucleotide-binding domain – where it can jam the pump to slow it down. Interestingly, the specific piece of the C-terminal domain which had this altered footprint is already known to be important for regulation, supporting the possibility that this shared footprint between “fuel tank” and “gearbox” is due to a physical interaction. Second, we know that the C-terminal domain is long, highly flexible, and capable of wrapping itself around the main body of the pump at multiple sites, like a how belt can be wrapped and tightened around one’s waist or hang unbuckled. Those previously known points of contact include areas very close to this Nucleotide-binding domain (“fuel tank”) site that had the shared footprint with the C-terminal domain (“gearbox”). A final detail which is relevant about the C-terminal domain is that this “gearbox” domain is not directly affected by these inhibitors, meaning any changes it experienced – more buried versus less buried – must be linked to other parts of the pump away from the C-terminal

domain. This leads to a possible scenario where the only way a portion of the C-terminal domain could become buried at the same time as another part of the pump is if it is packed tightly against that part of the pump when the footprinting experiment is occurring, causing the interacting pieces to partially block each other from the footprinting agent (“paint”). The region where this altered footprint in the Nucleotide-binding domain (“fuel tank”) matched the C-terminal domain (“gearbox”) is indicated in **Figure 6**. Follow-up experiments are needed to fully map the response of the C-terminal domain to structural shifts in the other domains of AHA2, but this is an important step forward because it adds new information to what previous studies have identified.

5.3.3 AHA2 pumps form trimers by interacting through adjacent domains.

The second major thrust of my research is to capture a group of multiple AHA2 proteins interacting with each other. In contrast to the footprinting and mass spectrometry experiments I described above – which are good at detecting structural changes in response to different conditions – this approach uses a special type of microscopy called cryo-electron microscopy, which is very good at completely imaging really large molecules (or large clumps of molecules) all at once.

In many respects, this is also a simpler experimental process to describe. Briefly, it entails isolating the target protein (in this case, groupings of AHA2), freezing it at extremely low temperatures, and then scanning it with a microscope that also records the images it scans. Those images are then computationally sorted according to their appearance into categories, and then stacked to build 2D image averages (**Figure 7**, *top*). You could think of these images as the average of many thousands of individual photographs of the molecule/molecular complex. If there are enough of those images in

a specific category – for example, 6 molecules of AHA2 arranged in a dumbbell shape – they can be assembled into a 3D model of the object that was scanned (**Figure 7**, *middle* and *bottom*). The most common “particle class” that we observed in AHA2 was an unusual arrangement of two valid, 3-pump groups of AHA2 proteins (aka., two trimers of AHA2), with each group interlocking (red, yellow, and blue domains, as shown in Figures 5 and 7).

This result was surprising, but also validating in several respects. The reason it is surprising is that the dumbbell shape formed by two interacting trimers is unlikely to reflect what AHA2 really does in living cells. However, we found two interesting details in the trimer that validate our expectations based on what is already known about AHA2. First, there are clear interactions between adjacent AHA2 proteins through their other domains (blue and yellow domains in **Figure 5**). Second, we did not observe the C-terminal domain directly, but the site where it would extend from the main body of AHA2 is oriented towards the adjacent pump in agreement with earlier work from our lab. That study showed that the C-terminal domain can daisy-chain adjacent AHA2 molecules in living cells, forming trimers like the ones we observed in this cryo-EM work. To imagine how C-terminal domain can bridge adjacent pumps, imagine how a sports team forms a huddle to hype each other up, with each player’s arms (C-terminal domains) across the shoulders of the adjacent player.

Future work with the plant proton pumps will focus on stabilizing the trimer unit enough to completely resolve the interactions between adjacent pumps. We also hope to uncover the exact structure of the C-terminal domain while it bridges adjacent molecules, which reflects the C-terminal domain’s function as it behaves like a

“gearbox” for the pump next to it. We are also engaged in a collaboration to analyze the opposite form of this structure – a fully activated group of plant proton pumps where their C-terminal domains are pulled away and tethered by other regulating proteins. There is more work ahead, but these results show we are on the right path toward finally resolving the complete structure of the plant proton pump, including the functional groups it can form.

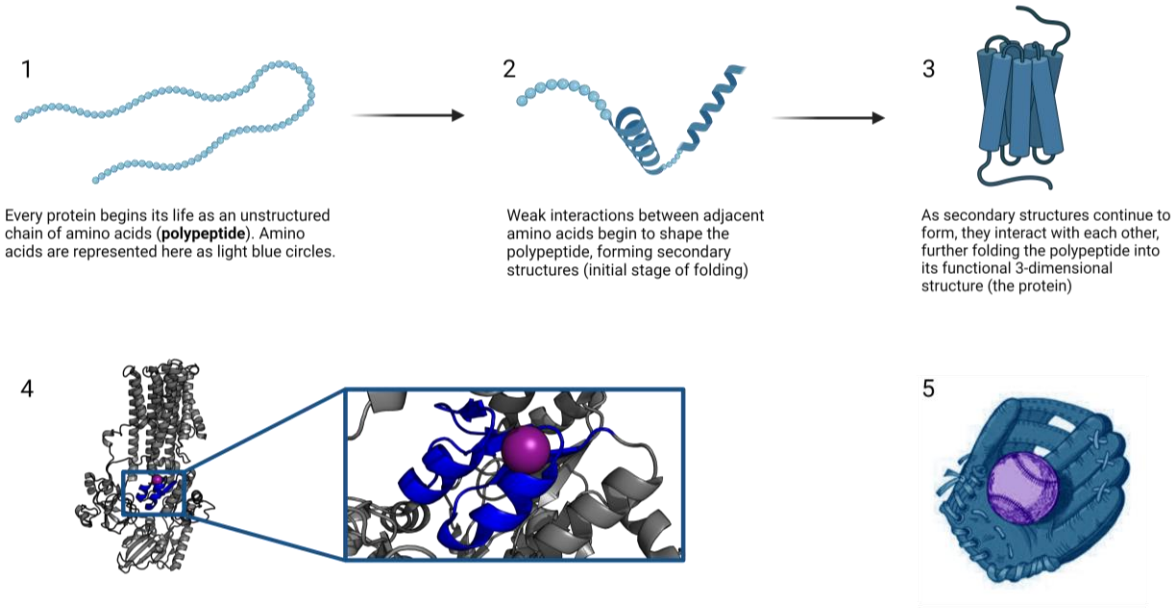

Figure 1. Protein folding results in a 3D structure with new properties that did not exist in the original 1D polypeptide. (1-3): Diagram of the stages of protein folding, culminating in a final example structure shown in (3). **(4):** Structure of the proton pump AHA2, with a zoom-in to reveal how individual regions of the protein are granted function through their structure. Displayed here is a potassium ion (K^+, purple sphere) gripped by a conserved fold of the protein. 5: This is analogous to how a baseball glove's shape is optimized for gripping a baseball.

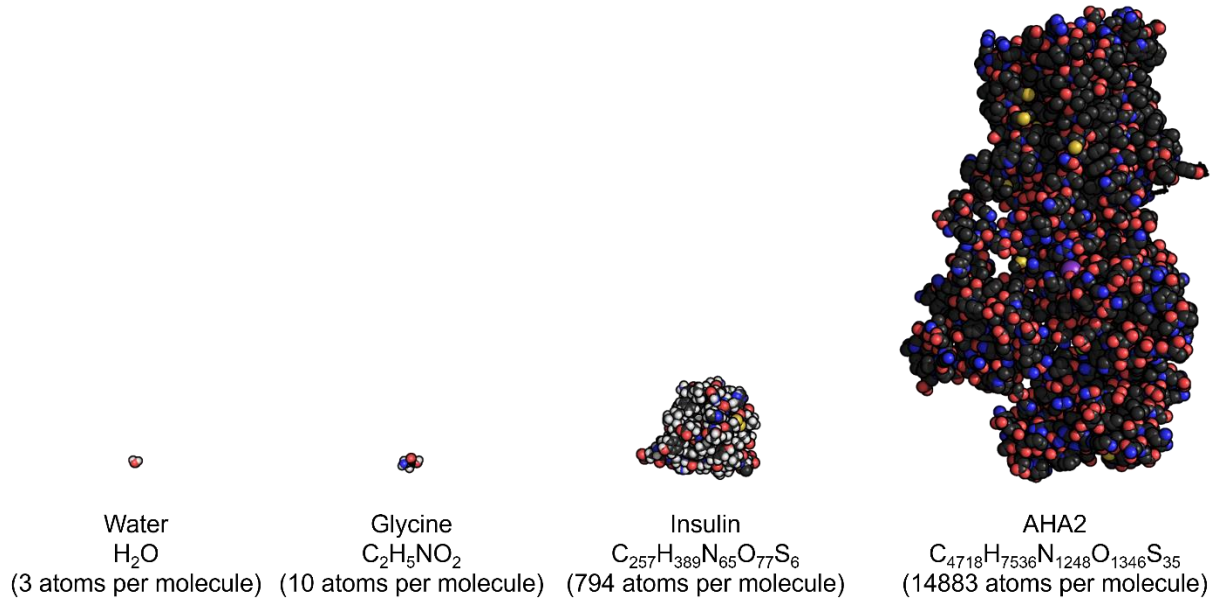

Figure 2. Biological molecules vary significantly in size and structure. From left to right, a size comparison between water, the amino acid glycine, the peptide hormone insulin, and the proton pump AHA2. These four molecules are shown with their atoms color-coded by element: grey for hydrogen (H), black for carbon (C), red for oxygen (O), blue for nitrogen (N), and yellow for sulfur (S). Hydrogens are not modeled in AHA2 not because they are absent, but because there was insufficient resolution in the experiment that produced the model.

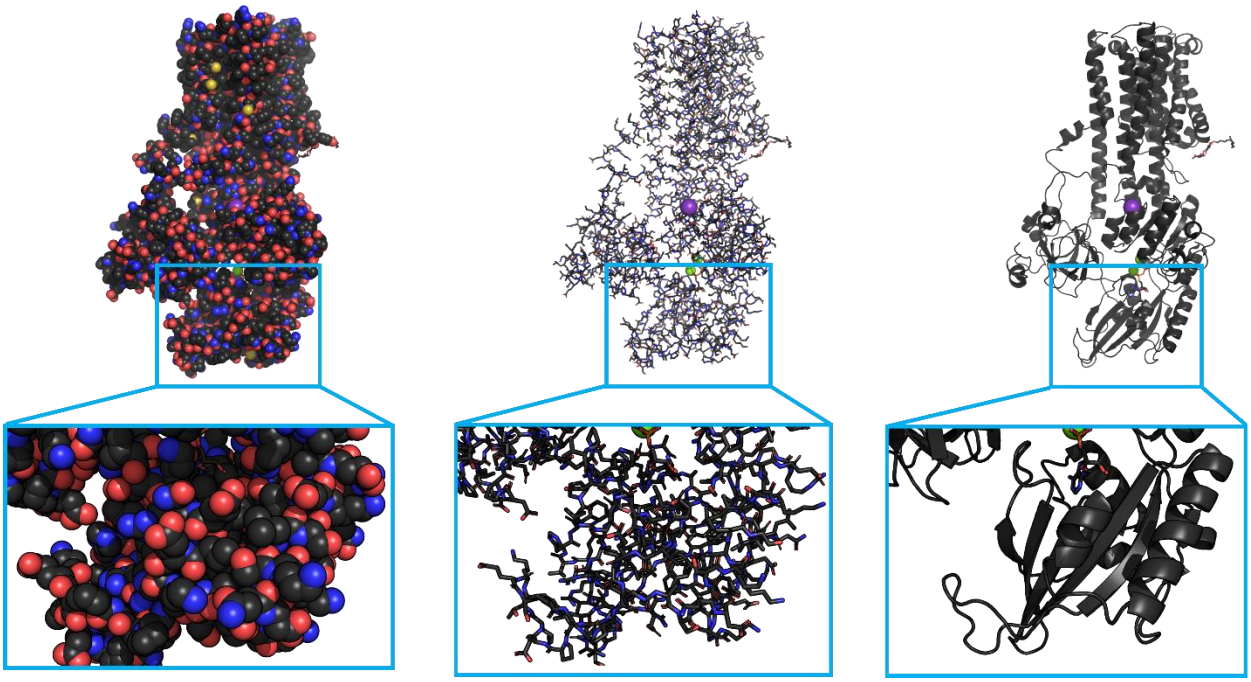

Figure 3. Molecular features can be highlighted at different levels of detail. *Left*: AHA2's atoms are modeled as spheres, representing a rough estimate of the actual bulk of the protein. *Middle*: AHA2's atoms are modeled as sticks, highlighting the geometry of the chemical bonds between the atoms. *Right*: the protein is displayed in a "cartoon" style where the overall contours of the protein's folds are emphasized: **helical** secondary structures that look like rotini, **β-sheets** that look like fettuccine, and **loops** that look like spaghetti (you're welcome). This highlights the shape of the **polypeptide backbone** (like a "skeletal" view of the protein). Protein structures are often depicted this way.

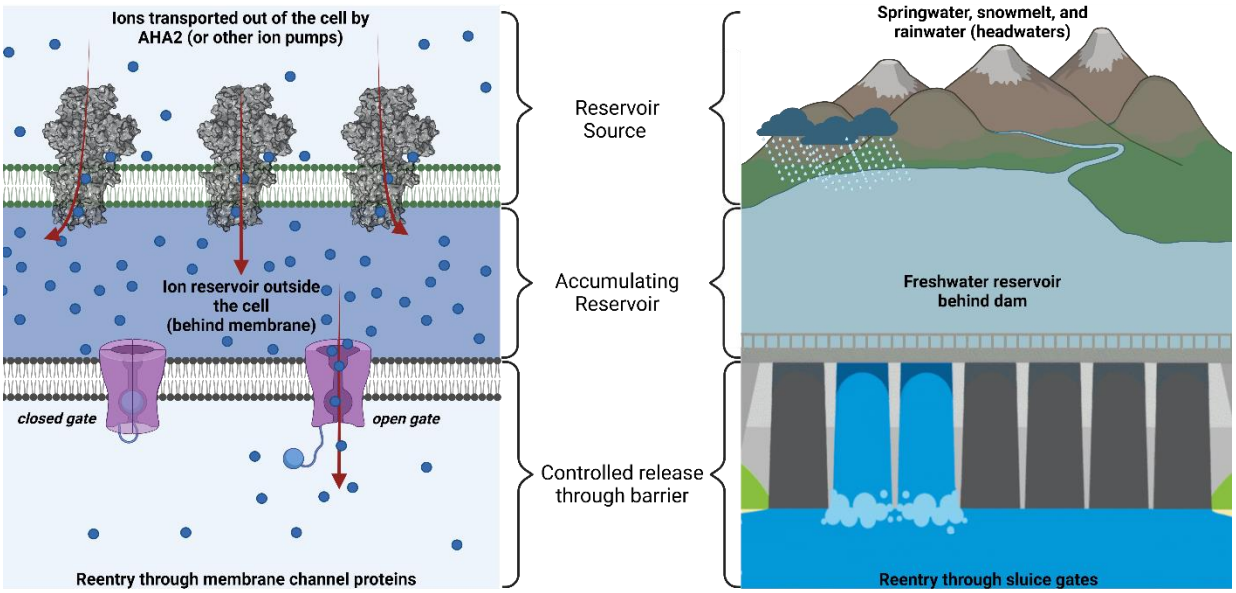

Figure 4. Reservoirs are stores of potential energy that can be released and harnessed at a controlled rate. *Left:* An ion reservoir generated by multiple ion transporters, like the proton pump AHA2, can be released across a membrane through specialized conduits, like gated ion channels. *Right:* A water reservoir generated by multiple natural freshwater sources can be released across a dam through sluice gates. In both cases, the downward flow from the reservoir through a point of egress can be used to power various kinds of work.

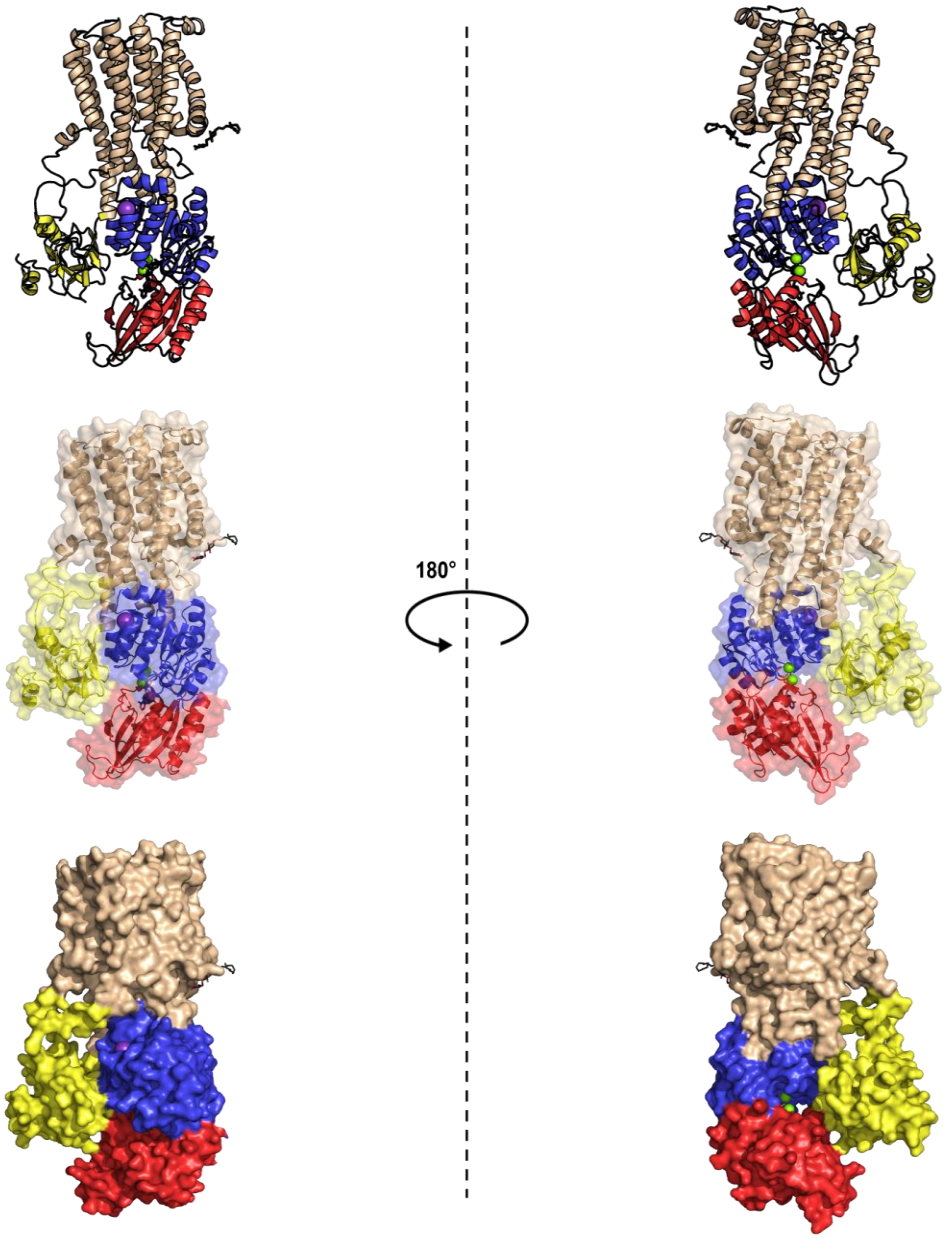

Figure 5. Visualizing the major functional regions (domains) of AHA2. The current structural model AHA2 is shown with cartoon (polypeptide backbone geometry) and surface topology overlaid. Red: Nucleotide-binding domain, Blue: Phosphorylation domain, and Yellow: Actuator domain. These three domains hang down into the cell like an iceberg's bulk hangs beneath the surface of the water. Wheat/Beige: Transmembrane (TM) domain, which is what embeds the pump in the outer membrane of living cells. The true structure of the C-terminal domain, which is missing here, has never been solved.

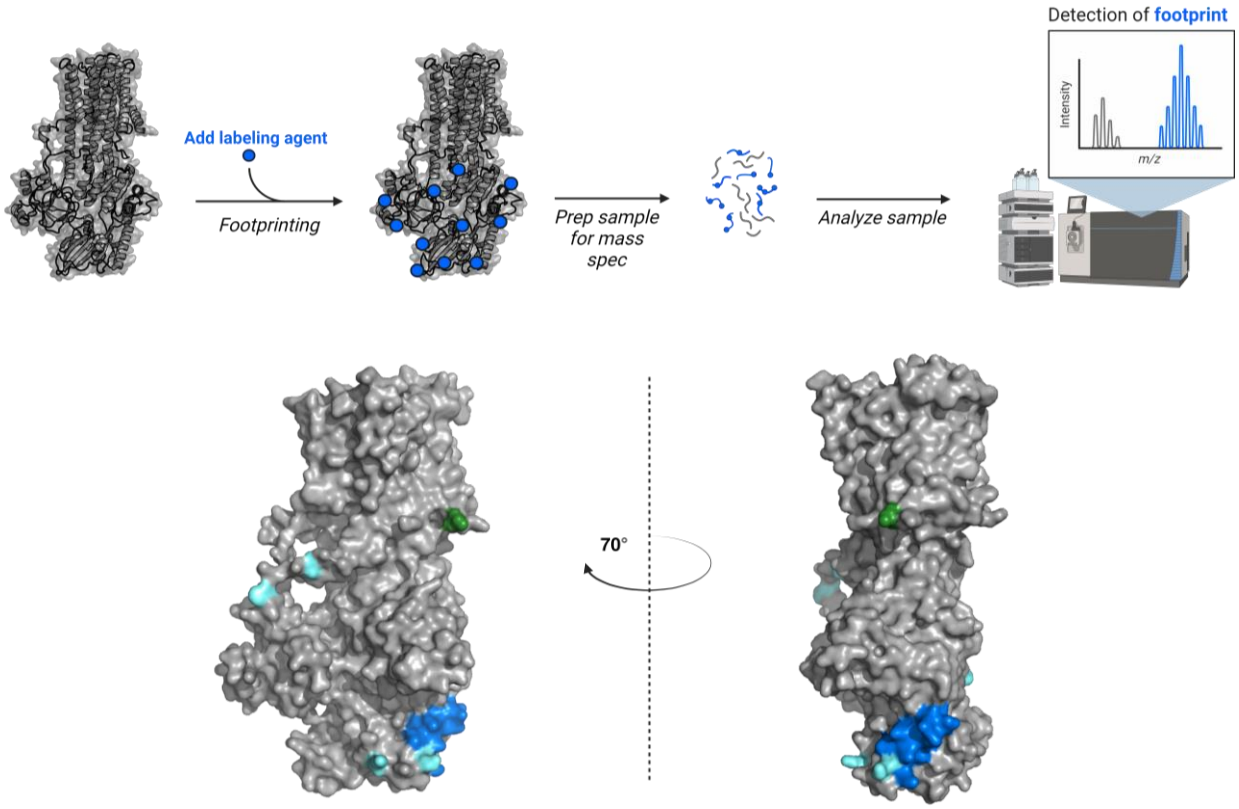

Figure 6. Footprinting reveals a possible binding cleft for the C-terminal domain. *Top:* a typical footprinting experiment involves mixing labeling agent ("paint") into a sample of protein and allowing the reaction to proceed for a determined length of time. The sample is then prepared and analyzed by mass spectrometry, which detects the footprint ("paint stains") on the protein. *Bottom:* model for where (dark blue) the C-terminal domain (which extends from the green patch) may be interacting with the pump, depicted with a solid surface. The specific structure of the C-terminal domain is still unknown, but our lab previously confirmed that specific contact points (light blue) between the pump and its C-terminal domain include the region I observed in my footprinting data (dark blue). For simplicity, the rest of the protein is shown in gray.

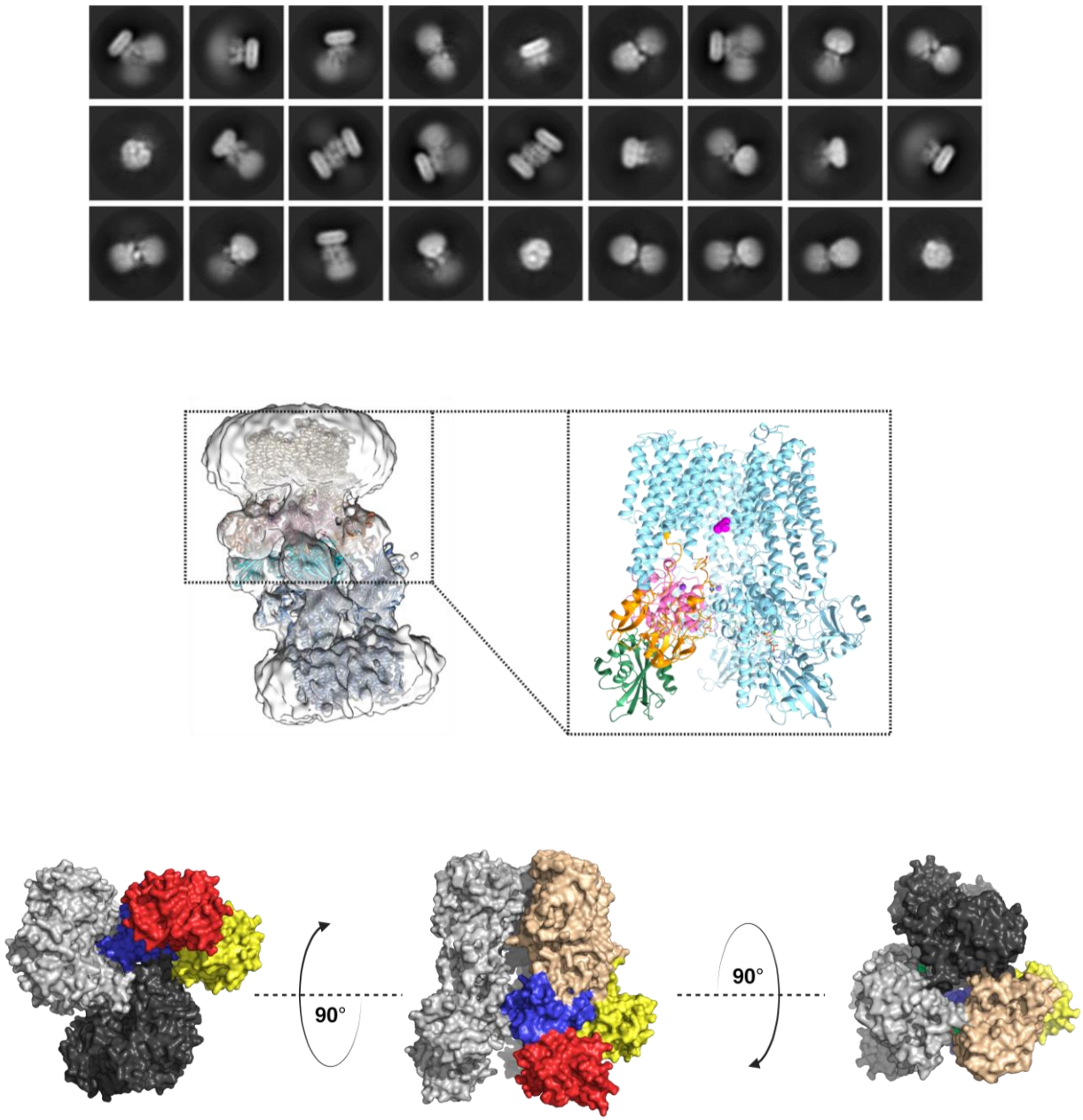

Figure 7. AHA2 may preferentially form trimers (groups of three proteins). *Top:* 2D class averages (microscope images) of grouped AHA2 molecules. *Middle:* the 3D model built from the most common images in the top panel reveals loose but stable packing between 3 molecules of the AHA2 protein. *Bottom:* the trimer (3-member group) of AHA2 is shown with a solid surface indicating the topology. Left to right: view from below, view from the side, and view from the top. For visual simplicity, only one AHA2 molecule in the group is color-coded by its domains like in Figure 4; the other two all gray or all black.

Figures in this chapter were prepared using BioRender and PyMOL.