

*Application for Admission to the University of Washington Genome Sciences Department*

As structural prediction tools in protein science improve, there is now no shortage of interesting questions to ask. Yet in my work, I have repeatedly seen that it is the analytics side -- the way we answer these questions -- which needs improvement. It is my goal to combine the tools of analytical chemistry to answer questions in structural biology, and I believe the proteomics work in development in the Genome Sciences department at UW is a compelling place to accomplish this.

During my freshman year of college, my research advisor sat me down in front of the modeling software "Coot" and taught me the basics of solving the crystal structures of proteins. As I studied the electron densities and ensured that the ball-and-stick model aligned to the x-ray data, it was as if all of biochemistry suddenly clicked for me. Visualizing the structure of an enzyme, observing the residues in their three-dimensional form, I fell in love and knew that the questions I wanted to ask were in the structure and function of proteins.

In my first post-baccalaureate research posting, I worked in an institute that used high-resolution mass spectrometry to study the metabolomics of rare populations of cells. It was there that I learned of the power of mass spectrometry to give clear, quantitative answers to some of the most foundational of questions under the most challenging of circumstances.

Currently, I work as a research scientist at Mopac Biologics, a spin-out from the Institute for Protein Design. It is this work that has allowed me to not only apply my analytical skills to structural biology questions but to also see ahead to the work in analytics still yet to be done and so desperately needed.

Mopac Biologics seeks to develop a protein-based, computationally-designed oral IL-23R inhibitor for the treatment of inflammatory bowel disease. Developing analytics for this designed protein has been one of my main roles at Mopac, and it has been fraught with delightful challenges. Our lead protein design is unique: at only a few kDa, it's too large for typical peptide analyses to work but too small for typical protein analyses. With a high pI, it is remarkably positive and problematically sticky. Finally, with two disulfides, there is the opportunity for heterogeneity in production that must be monitored.

My first task with Mopac was to develop a size-exclusion chromatography (SEC) analysis method that could provide us with a purity measurement. Being so positively charged, our protein had a habit of sticking to the stationary phases of our columns, causing our resulting peaks to smear. After fixing the pumps on the HPLC and testing many different columns and mobile phases, I found that a high ionic strength running buffer kept non-specific interactions to a minimum and resulted in beautiful peaks. However, one contaminant persisted, and finding its identity became my next task.

Early work showed that treatment of our protein with a reducing agent caused our protein to behave in the SEC method similarly to this contaminant. This was my first hypothesis, that the contaminant was a disulfide isoform of our protein. Leveraging my experience with small molecule mass-spectrometry, I planned to use MS to analyze our protein sample and determine the mass, and therefore identity, of the contaminant.

Small molecules, however, behave very differently in MS than whole proteins like ours. This was a fact I didn't truly appreciate until I saw the tangled mess of spectra coming from my sample. The reduction of a disulfide bond results in a 2 Dalton mass shift which is very difficult to observe within the spectra of a several-thousand Dalton protein. It's possible with high-resolution MS, but not with the instruments I had access to.

I designed a workaround to this limitation. If the contaminant could be separated from the protein, even low-resolution MS could identify it. SEC couldn't effectively accomplish this, so reversed-phase chromatography was investigated. With again much persistence, I found a reversed-phase method that separated the contaminant from our protein.

I still remember vividly that Sunday evening in the lab when I used the reversed-phase method with the MS. For the first time, I saw the structural heterogeneity of this protein represented as clean peaks on the chromatogram, each with a mass difference of 2 Daltons corresponding to disulfide cleavage as hypothesized. I had successfully used my analytical chemistry skills to probe the structural features of this protein. Further testing with this method confirmed that the contaminant seen in our SEC assay was a disulfide isoform of our protein, as expected.

As I began to look around, I saw every new problem as a proverbial nail for my newfound hammer to solve. My new RP-HPLC-MS method could characterize our protein; the next step was to see if it could quantitate it from complex matrices. Yet, characterization from pure samples and quantitation from dirty samples are completely different beasts, and I wish I could tell you I was successful. It was not without trying; I spent many long days and nights ruining many columns before I came to the conclusion that this top-down, whole-protein approach to quantitation was unviable in this case, and that work would need to be restarted with a bottom-up, fragmented peptide approach. Unfortunately, my efforts were needed on other projects and this work was put on pause.

It is perhaps this unfinished project that motivates me the most to seek a PhD in Genome Sciences. We are but one of many protein-based startups coming out of the Institute for Protein Design asking all of the same analytical questions as our peers: how is our designed protein behaving, how much of it is in this sample, and how is it affecting other signaling molecules in the cell.

It's my belief that quantitative proteomics via mass spectrometry is the tool to answer these questions. LC-MS has the capability to separate, identify, and quantify protein-based analytes with ever increasing efficiency and accuracy, but work needs to be done to develop and apply these tools.

Here, in the Genome Sciences department at UW, are motivated researchers developing these tools, and I am seeking to be among them. Dr. Michael MacCoss and his team have been pioneers in advancing and democratizing proteomics research by producing Skyline and are currently working with cutting-edge instruments and techniques to improve sensitivity, accuracy, and throughput of proteomics analyses. Along this line, Drs. Judit Villen and Devin Schweppe are also working to develop new quantitation methods in this field. A PhD from any one of them would allow me to continue in my career.

After my PhD, I have two paths forward. One path is to take the analytical skills I will learn into the translational research space and continue my work with Mopac Biologics, the Institute for Protein Design, or another protein and analytics-focused startup such as Talus Bioscience. Another path after a PhD is to take an academic post-doctoral position to continue using the tools of proteomics to answer basic science questions in structural biology. This path would allow me to engage in my love of teaching, mentorship, and science communication by ultimately allowing me to pursue a faculty position at a university. In either case, a PhD is needed to further my training and career as a scientist, and the department of Genome Sciences at the UW has the expertise to train me to use the tools of analytical chemistry to answer questions in structural biology.