# 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 Strategies for Identifying Endogenous Peptides and Determining Protein Turnover Rates

By

Zach Rolfs

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: November 16th, 2020

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

Ying Ge, Professor, Cell and Regenerative Biology and Chemistry

Lingjun Li, Professor, Pharmaceutical Sciences and Chemistry

Lloyd M. Smith, Professor, Chemistry

Nathan V. Welham, Associate Professor, Surgery

# 6 CHAPTER FOR THE NONSCIENTIFIC PUBLIC

## 6.1 Introduction

The purpose of this chapter is to explain my research to a broader, nonscientific audience. I chose to write this chapter because I feel that science has become increasingly unapproachable for the general public. Technology is advancing so fast and has become so complex that it's easy for people to be intimidated and distrustful of science. I'm incredibly saddened for the people who hold these views because I've always viewed science as a means for understanding our world. It's a way to ask the question "why" without getting the response "because I said so." Although misinformation is on the rise, I largely blame the public distrust of science on our education system. In my high school science classes, I was bombarded by facts which I needed to memorize. Once these facts were regurgitated to pass my exams, they could be immediately ejected from my brain. I strongly feel that a science curriculum centered around critical thinking would be much more beneficial to the general public for understanding the nature of science and scientific thinking.

In my high school physics class, I learned that the speed of light is approximately 300 million meters per second, but I was never taught how we figured out that number. Similarly, it's great that I learned that the world is approximately 4.5 billion years old, but how do we know that? If we take these facts by faith instead of by evidence, then it becomes "just as valid" to say that the world is 2020 years old as it is 4.5 billion years old. Public distrust of science stems from a failure of transparency and communication in the scientific community. Logistically, there are not enough hours in a K-12 education to teach people how humanity arrived at all of our scientific breakthroughs, but it is my hope that this chapter for nonscientific audiences provides a transparent and accessible means to demystify the research that I have accomplished during my time in graduate school.

I would like to thank my nonscientific family members for their continuous bravery over the past five years. They have been steadfast in asking me to explain my research at every Christmas and family gathering and they have been instrumental in helping me learn how to talk about science to nonscientific audiences. I would like to thank the Wisconsin Initiative for Science Literacy at the University of Wisconsin-Madison for providing this important platform and for their sponsorship and support throughout the writing of this chapter.

## 6.2 DNA Encodes for Proteins

DNA has been referred to as the "blueprints of life" but they're actually the blueprints for creating proteins, which are the molecules that provide structure, function, and regulation for all cells. Proteins are much more difficult to study than DNA, but it's important to study proteins because the proteins don't always look like the blueprints intended. You can think of DNA as being directions for building houses and proteins as those houses. Sometimes a homeowner will tear down a wall or add on a new addition. These changes to the house aren't reflected in the blueprints, but they can dramatically change the structure and the function of the house. Similarly, proteins can undergo changes within the cell that aren't specified by the DNA. Additionally, just because you have blueprints doesn't mean that the house was actually constructed. All of your cells share the same DNA, but there's an incredible amount of biological diversity between your tissues because of the differences in which blueprints your cells choose to use. Your genome is the collection of all your DNA and goes largely unchanged throughout your life. However, your proteome (which is the collection of all of your proteins) can be highly variable in response to stress, disease, and environmental factors. In terms of healthcare,

genomics is useful for telling you what could happen, while proteomics is useful for telling you what's happening right now.

### 6.3 Cancer Immunotherapy

Your immune system uses proteins to monitor all of your cells and determine which cells are healthy and unhealthy. Each cell is responsible for identifying itself to the immune system by cutting up some of their own proteins into small molecules called peptides. These peptides are then transported to the cell surface and displayed so that the immune system can recognize if there's a problem with a cell. If the immune system finds a peptide that it doesn't expect, then it kills the cell that's displaying the unusual peptide. Your immune system is very effective at doing this and is able to kill most cells that develop dangerous mutations, but it's not a perfect system. Cancer occurs when a cell mutates to become immortal and is able to hide this mutation from the immune system. There's at least one peptide produced by the cancer cell that isn't present in normal healthy cells, but the cancer cell is preventing the immune system from recognizing the unusual peptide responsible for the cancer. Scientists are still trying to figure out how cancers hide from the immune system, but there's evidence that cancers can display fewer unusual peptides (like a disguise) or by disabling the immune system.[1]

Cancer immunotherapy is a promising treatment option for many patients and several drugs are already available. The idea behind cancer immunotherapy is to figure out which unusual peptide is being produced by the cancer cells and then train the patient's immune system to specifically attack the cells that are displaying this peptide, sort of like a vaccine. There are huge benefits for this therapy in comparison to traditional methods like chemotherapy. Chemotherapy indiscriminately kills both normal and cancer cells, but cancer immunotherapy only kills cancer cells. The biggest challenge for

developing new cancer immunotherapy treatments is to identify the unusual peptide that's being presented by the cancer cell. Cancer is actually many different diseases. There are countless mutations which allow a cell to become cancerous and each mutation requires a unique solution. Unfortunately, there are many types of cancer for which researchers have been unable to find a tumor-specific peptide. As a workaround, researchers will sometimes train the patient's immune system to recognize peptides which are displayed in large amounts on cancer cells and low amounts on healthy cells. This workaround has been effective for treating some cancers, such as HER2+ breast cancer, but it also kills healthy cells in the process. HER2+ breast cancer is a type of breast cancer that displays large amounts of the protein "**h**uman **e**pidermal growth factor **r**eceptor **2"** (HER2). Normal cells also display HER2, but at much lower levels. Training the patient's immune system to focus on HER2 allows the immune system to kill HER2+ breast cancer, but it also makes the immune system kill healthy cells. The destruction of healthy cells can have serious side effects and even result in death.[2]

### 6.4 Identifying Peptides using Mass Spectrometry

My research has been developing methods to identify unusual tumor-specific peptides. Just like how a house is made out of multiple rooms, peptides are made out of multiple pieces called amino acids. Each amino acid is linked to another amino acid to create a train. We need to identify all amino acids and the correct order of these amino acids to be able to identify a peptide. We can do this using an instrument called a mass spectrometer. Mass spectrometers are basically molecular scales which can determine the mass of a molecule. We can measure the mass for our peptide to help us determine which amino acids are in our peptide, but we need more information to determine the order of these amino acids. Inside the mass spectrometer, we blast the peptides with energy to

break a single chemical bond on each peptide and create fragments (Figure 6.1). We can then determine the order of the amino acids based on the masses of the fragments.

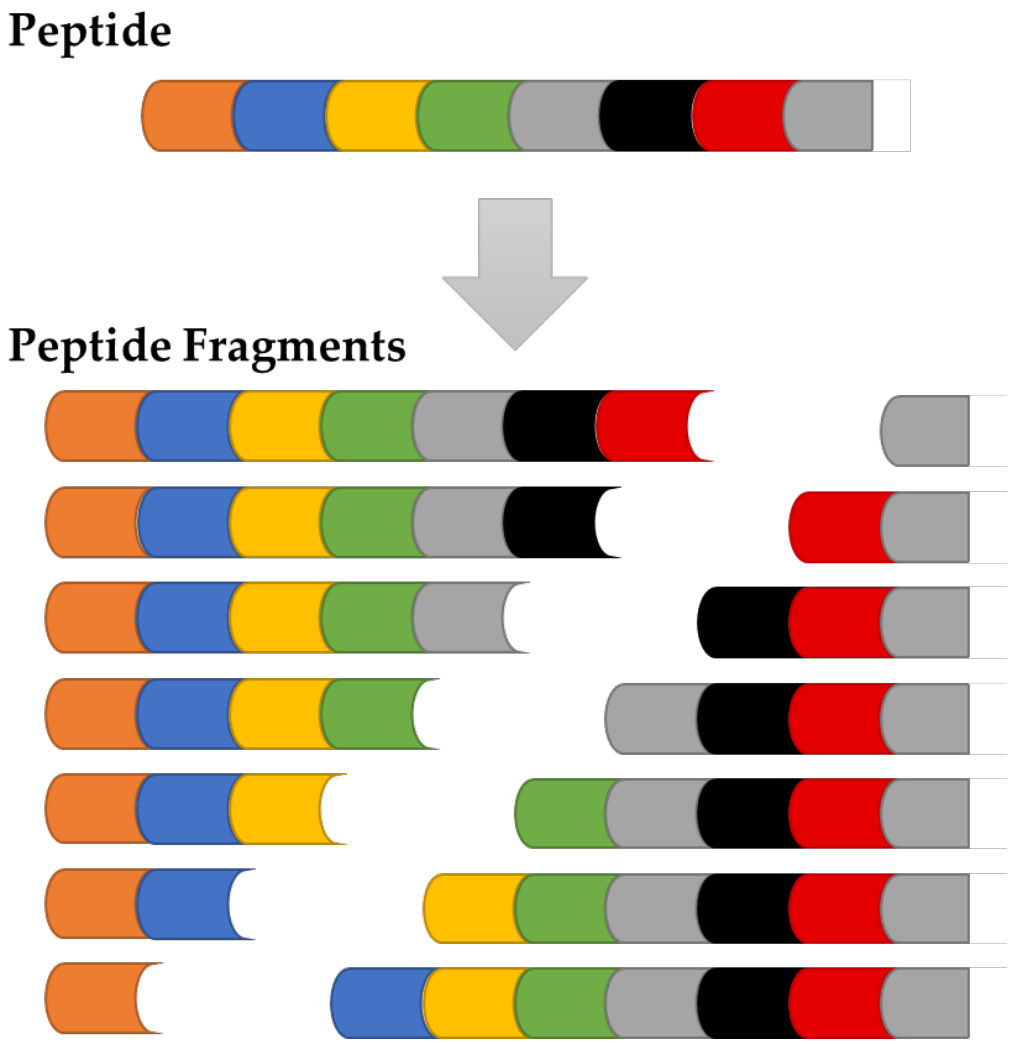


**Figure 6.1** Cartoon representation of how a peptide fragments. Each colored cylinder represents a different amino acid. A mass is measured for the intact peptide as well as for each of its 14 fragments.

Mass spectrometers are incredibly fast and can identify tens of thousands of peptides per hour. They generate too much data for researchers to analyze by hand, so software programs have been developed to quickly analyze the data and identify each peptide. These programs work by taking advantage of the fact that DNA encodes the instructions for proteins. First, they translate the DNA to get a database of all proteins in the sample.

Next, they virtually cut up these proteins into peptides and predict what the data should look like for each peptide. They then use a pattern matching algorithm to compare the observed and expected list of masses for each peptide and determine the best identification for each peptide. This approach works well for identifying peptides that look like their blueprints, but it is unable to identify peptides that don't have blueprints (recall that peptides can be changed after they're made).

### 6.5 Identifying Post-Translationally Spliced Peptides

In 2016, a scientific paper was published claiming that a large fraction (30%) of the peptides displayed on the cell surface were different from their instructions.[3] These modified peptides are called post-translationally spliced peptides (PSPs) because they come from two different peptides being spliced (cut and pasted) together after the DNA has been translated by the cell to make proteins (Figure 6.2). There weren't any methods available for researchers to identify PSPs because they're different from their instructions and the instructions are needed by existing methods to identify peptides. I developed a software program that can identify these PSPs and allow cancer researchers to determine if any of them are useful for cancer immunotherapy.

Although PSPs are not in the blueprint database, the two halves of each PSP do have blueprints. If you were trying to identify a mermaid, you might incorrectly say that it's a fish or a human. I created a program that could identify PSPs by identifying the two halves (fish and human) and then putting the pieces back together. First, I made two

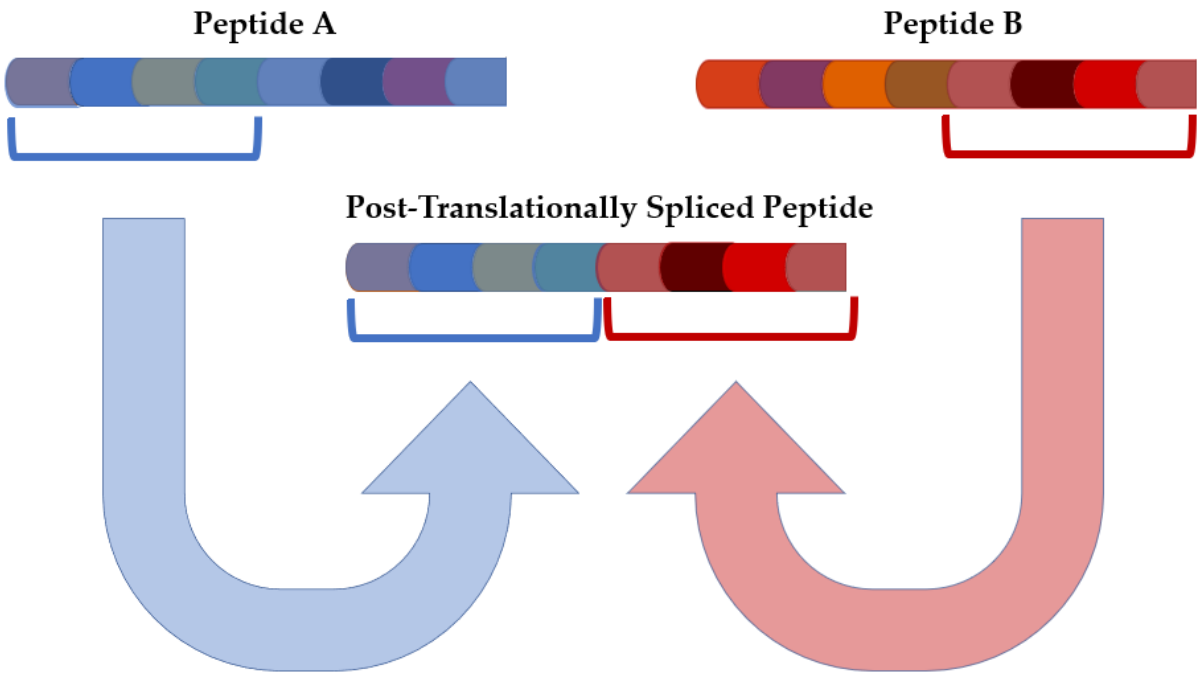


**Figure 6.2** Cartoon representation of a post-translationally spliced peptide. Two existing peptides both contribute sequences which are then pasted together. The new post-translationally spliced peptide still looks like a normal peptide, but there are no instructions for this peptide in the DNA.

separate lists of fragment masses for each peptide in the database. All of the fragments that contain the left end were put in one list and all of the fragments that contain the right end were put in a separate list (Figure 6.1). For the mermaid analogy, you can think of this as putting all the features above the waist in one search and all the features below the waist in another search (Figure 6.3). The idea was to use the list of masses from the left side to identify the original left peptide and use the list of masses from the right side to identify the original right peptide. The observed mass list contains both sets of fragments, so neither the left nor the right pattern is a perfect match for PSPs. For example, comparing a fish tail with the mermaid shows similarities, but the mermaid still has a top half that's different from the fish tail. The closest matches from both sides are then combined in all possible combinations until the mass of the combination matches the observed peptide mass.

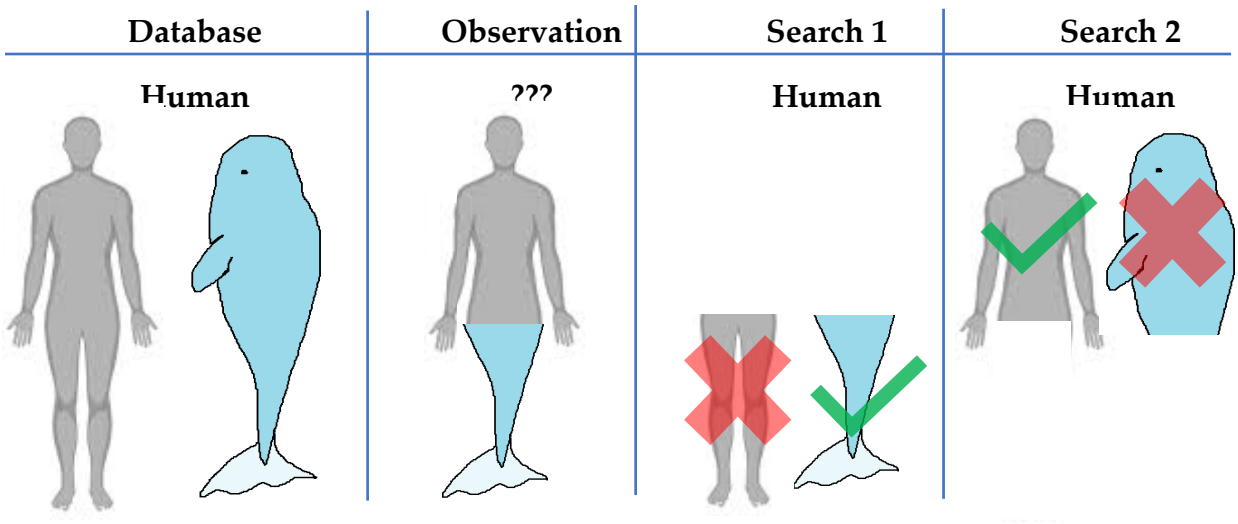


**Figure 6.3** Finding a PSP is like finding a mermaid. The two halves come from known peptides (animals), but the PSP (mermaid) doesn't look exactly like either of them. Using two searches for each half allows us to identify the two parts and then put them together to solve the puzzle.

After I made this tool, I started looking for PSPs. Lots of mass spectrometry data has been made publicly available and I took advantage of this free data. I downloaded data from both normal cells and tumor cells and analyzed it with my tool to find PSPs. I was expecting about 30% of the identified peptides in each dataset to be PSPs because of the previous paper, but less than 1% of all the peptides I identified were PSPs. This was concerning because it meant that either the previous paper was incorrect or my tool was not very good at finding PSPs. I simulated a large amount of PSP data and showed that my program was effective at finding PSPs, so I needed to figure out what was wrong with the previous paper, which claimed that 30% of peptides were PSPs. That paper made it through the peer review process, but sometimes reviewers can miss things. Mistakes happen and it's an important part of the scientific process to constantly be self-correcting and searching for the truth. I discovered that the original researchers had a statistical issue in their analysis that caused them to greatly overestimate the number of PSPs in their sample. An easy way to see this statistical issue is to look at the order in which the peptides were identified (Figure 6.4). Peptides are sorted based on how well they dissolve

in water before they enter the mass spectrometer. Peptides that dissolve well in water are identified first and peptides that don't are identified last. We have equations than can predict when a peptide should be identified and we can use this information to determine if our identification is correct or not.[4] I teamed up with another lab that was investigating this same statistical issue and we worked together to publish rebuttals showing what went wrong.[5-7] These rebuttal papers are important to the field so that cancer researchers don't waste time and resources looking for PSPs that don't exist.

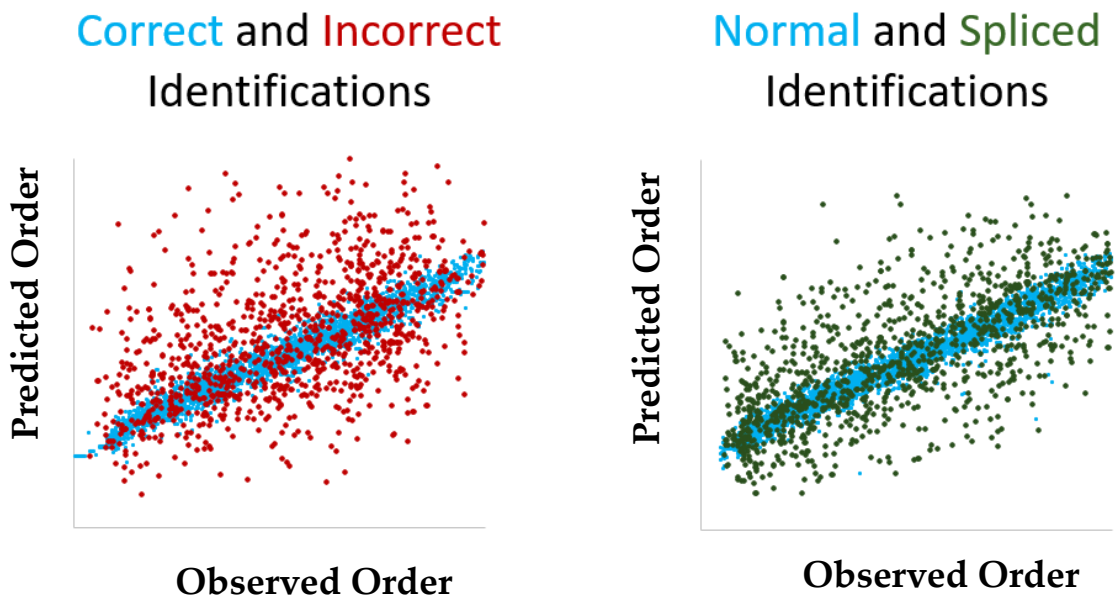


**Figure 6.4** Peptides are separated before analysis. The order in which they separate is related to the chemical properties of each peptide. We predicted the separation order and compared it with the observed separation order. Correct identifications produce a nice straight line and incorrect identifications are a random scatter. Similarly, the PSPs reported in the previous paper resemble incorrect identifications.

## 6.6 Conclusions and Future Directions

I developed a tool to identify post-translationally spliced peptides (PSPs), but I found that there weren't many PSPs displayed on the cell surface. This means that PSPs are unlikely to be useful for cancer immunotherapy and the tool probably won't be used by researchers to find tumor-specific peptides. However, it was an important discovery to show that very few PSPs are displayed and that has been my biggest contribution to the field.

Although PSPs might not be useful for cancer immunotherapy, researchers found that PSPs might be responsible for type 1 diabetes.[8] Type 1 diabetes is an auto-immune disease, where the cells that produce insulin are destroyed by the patient's immune system. A few papers have been published reporting that one or more PSPs are being recognized by the immune system. Our hope is that we could help cure type 1 diabetes if we can figure out what these PSPs are and how they're being made. The work that I've done to identify PSPs is still being used now to find PSPs for type 1 diabetes and we're excited to see what we find!

## 6.7 References

(1) Waldman, A.D.; Fritz, J.M.; Lenardo, M.J. A guide to cancer immunotherapy: from T cell basic science to clinical practice. *Nat. Rev. Immunol.* **2020,** *20*, 651–668.

(2) Morgan, R. A.; Yang, J. C.; Kitano, M.; Dudley, M. E.; Laurencot, C. M.; Rosenberg, S. A. Case Report of a Serious Adverse Event Following the Administration of T Cells Transduced With a Chimeric Antigen Receptor Recognizing ERBB2. *Mol. Ther.* **2010**, *18* (4), 843–851.

(3) Liepe, J.; Marino, F.; Sidney, J.; Jeko, A.; Bunting, D. E.; Sette, A.; Kloetzel, P. M.; Stumpf, M. P. H.; Heck, A. J. R.; Mishto, M. A Large Fraction of HLA Class I Ligands Are Proteasome-Generated Spliced Peptides. *Science,* **2016**, 354 (6310).

(4) Krokhin, O. v.; Ying, S.; Cortens, J. P.; Ghosh, D.; Spicer, V.; Ens, W.; Standing, K. G.; Beavis, R. C.; Wilkins, J. A. Use of Peptide Retention Time Prediction for Protein Identification by Off-Line Reversed-Phase HPLC-MALDI MS/MS. *Analytical Chemistry* **2006**, 78 (17), 6265–6269.

(5) Mylonas, R.; Beer, I.; Iseli, C.; Chong, C.; Pak, H.-S.; Gfeller, D.; Coukos, G.; Xenarios, I.; Müller, M.; Bassani-Sternberg, M. Estimating the Contribution of Proteasomal Spliced Peptides to the HLA-I Ligandome. *Molecular & Cellular Proteomics* **2018**, mcp.RA118.000877.

(6) Rolfs, Z.; Solntsev, S. K.; Shortreed, M. R.; Frey, B. L.; Smith, L. M. Global Identification of Post-Translationally Spliced Peptides with Neo-Fusion. *Journal of Proteome Research* **2019**, *18* (1), 349–358.

(7) Rolfs, Z.; Müller, M.; Shortreed, M.R.; Smith, L.M.; Bassani-Sternberg, M. Comment on "A subset of HLA-I peptides are not genomically templated: Evidence for cis- and trans-spliced peptide ligands". *Sci. Immunol*. **2019** *4* (38), eaaw1622.

(8) Delong, T.; Wiles, T. A.; Baker, R. L.; Bradley, B.; Barbour, G.; Reisdorph, R.; Armstrong, M.; Powell, R. L.; Reisdorph, N.; Kumar, N.; et al. Pathogenic CD4 T Cells in Type 1 Diabetes Recognize Epitopes Formed by Peptide Fusion. *Science*. **2016**, 351 (6274), 711–714.