Novel Mass Spectrometry-based Method Development and Applications to Signaling Peptides and Proteins

By

Christopher B. Lietz

A dissertation in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Chemistry)

at the

University of Wisconsin-Madison

2016

Date of final oral examination: 01/15/16

This dissertation was approved by the following members of the Final Oral Committee:

Lingjun Li, Professor, Pharmacy, Chemistry

Joshua Coon, Professor, Chemistry, Biomolecular Chemistry

Ying Ge, Associate Professor, Cell and Regenerative Biology, Chemistry

Joel Pedersen, Professor, Soil Science, Chemistry

Michael Sussman, Professor, Biochemistry

Acknowledgements

The research presented in my dissertation would not have been possible without the assistance of others. First and foremost, I would like to thank Dr. Lingjun Li for not only allowing me to join her group, but also for encouraging me to attend the University of Wisconsin. Her guidance, enthusiastic support, and scientific fearlessness has shaped me and taught me what it means to be a researcher. I cannot fully express how thankful I am.

Two fellowships were directly responsible for my funding. Not only did they make all my research possible, they also saved me from starvation. I would like to thank the Chemistry/Biology Interface Training fellowship (NIH grant number T32-GM008505) and the NSF Graduate Student Research Fellowship ((DGE-1256259).

I would also like to thank my committee members: Dr. Ying Ge, Dr. Michael Sussman, Dr. Joshua Coon, and Dr. Joel Pedersen. You are all wonderful role-models, and your time and feedback is instrumental to my development as a scientist.

Furthermore, I would like to acknowledge Dr. Sarah Trimpin for accepting me into her lab during my undergraduate studies at Wayne State University in Detroit, Michigan. She introduced me to mass spectrometry and taught me about hard work—not only how to do it, but how well it can pay off, too.

So many collaborators and labmates contributed to this dissertation in some way. None of the proteomics experiments would have been possible without the direction, motivation, and front-end biological preparation/experimentation by collaborators. I would like to thank the team at the Millennium Pain Center for all their exceptional work,

especially Dr. Dana M. Tilley, Dr. David L. Cedeño, Dr. Ricardo Vallejo, Courtney Kelley, and Dr. Ramsin Benyamin. I would also like the thank Dr. Xudong Shi for all his help with the restenosis project, as well as general assistance with cell culture. Dr. Robert Sturm was my first student mentor, and he helped me make a big transition from laser-based mass spectrometry to liquid chromatography-mass spectrometry. Dr. Tyler Greer provided both great advice and great friendship many times throughout my graduate studies. Erin Gemperline was extremely helpful and generous with her skills at making attractive scientific figures and schemes. I would also like to thank all other students in the Li Lab for support and helpful discussions.

I would like to give extra thanks to the Li Lab students who will take over my unfinished projects: Zhengwei Chen, Xiaofang Zhong, Fengfei Ma, and Qing Yu. I would like to give extra thanks to Xiaofang for being incredibly instrumental to the success of all my phosphoproteomic projects.

I would also like to thank Dr. Dustin Frost and Amanda Buchberger. They synthesized DiLeu tags for my experiments. Without the DiLeu, there wouldn't have been experiments.

I would like to thank Dr. Cameron Scarlett and Molly Pellitteri Hahn for running the School of Pharmacy's Analytical Instrumentation Center.

I was very lucky to get an opportunity to intern at Promega for three months. None of the work I did there is included in this thesis, but I learned so much while I was there and was able to gain some "molecular biology chops." I would like to give thanks to Dr. Danette Daniels, Dr. Marjeta Urh, Dr. Jacqui Mendez, and an especially big thanks to Nancy Murphy, for all their help and knowledge.

The impact of my family's love and emotional assistance cannot be understated. My sisters, Katie and Nicole, were always there to lift my spirits. My mother, Hendrika, has always been in awe of science. Her support of my graduate career was unwavering. Although it was a slow process, her passion for neuroscience has been passed on to me and helped direct my post-graduate path. My father, William, is an elementary school science teacher. His enthusiasm for science was always just a normal part of my life. He often tells a story about how he took my out to see a lunar eclipse when I was an infant. I obviously don't remember it, but such experiences throughout my life probably affected me more than I'll ever know.

My fiancé, Alicia Richards, was surely the most important influence on my graduate career. Without her helpful advice, encouragement, motivation, and patience, a Ph.D. would have not been possible. I aspire to be more like her in every way possible.

Finally, I would just like to thank nature. Thanks for being so neat.

Table of Contents

Page

Novel Mass Spectrometry-based Method Development and Applications to Signaling Peptides and Proteins

Christopher B. Lietz

Under the supervision of Professor Lingjun Li

At the University of Wisconsin-Madison

Abstract

This dissertation presents research centered on mass spectrometry (**MS**)-based methods for peptide and protein analysis. All investigations belong to one of two groups: innovative application of well-established MS approaches to signal transduction systems, or development of novel peptide and protein analytical methods that could be applied to signaling peptides and proteins in the future. In Chapter 1, the overall goals and key findings of each project are summarized. In Chapter 2, the use of ion mobility (**IM**)-MS to characterize cis/trans isomers of proline-containing peptides was given a critical examination. As a result, the investigation yielded the first concrete example of experimental IM-MS evidence suggesting cis/trans isomers while theoretical evidence suggested exclusively trans-proline conformations. In Chapter 3, the analytical utility of IM-MS was used to mitigate isobaric interference that arises during the use of quantitative tandem mass tags. When co-eluting, co-isolated labeled peptides are fragmented, they yield chimeric reporter ion spectra. If the precursors could be separated by IM prior to fragmentation, accurate reporter intensities were obtained. Chapter 4 demonstrates the synergistic combination of high-throughput bottom-up proteomics and ion collision cross-section (**CCS**) measurement via IM-MS. After developing and validating a new method,

it was used to create a peptide CCS database containing several thousand entries. Chapter 5 departs from IM-MS and method development to discuss the current state of proteomics in the field of chronic pain research. Chapter 6 then describes an application of multiplexed quantitative proteomics to the spared-nerve injury model of neuropathic pain. Among many proteins that were identified as having significant differential regulation, Bag3 showed the highest up-regulation in the pain model specimens and was orthogonally validated by Western blot. Chapter 7 reports on the application of multiplex quantitative proteomics to deciphering the complex signaling pathways of TGF-β and its role in restenosis. Representing the first biological application of dimethylated leucine (**DiLeu**) tags to phosphoprotein analysis, several hundred phosphoproteoforms were identified as exhibiting significant differential regulation between normal smooth muscle cells and cells stimulated with TGF-β. Finally, Chapter 8 concludes the dissertation with an evaluation of the contributions made to the broader field of analytical chemistry.

Chapter 1

Introduction and Summary of Research Projects

Introduction

Although the individual projects of this doctoral dissertation vary in scope, they are all earnest contributions to the field of mass spectrometry (**MS**)-based analysis of peptides and proteins involved in signal transduction—the process of cells receiving information via external stimuli to elicit a specific response.[1] Peptides and proteins are known to participate in virtually every step of signal transduction, from the extracellular ligands that carry the information,[2] to the membrane-bound receptors that receive the information,[3] and protein channels that allow signals to propagate across cell networks.[4] Although great strides have been made in elucidating the functional roles of numerous signaling peptides and proteins, the primary challenges in their analyses have remained largely unchanged. Even with detailed knowledge of the genes which encode them, the exact form and abundance of their bioactive forms, especially on the whole proteome and peptidome scale, are difficult to ascertain.[5-6]

MS has the potential to make tremendous contributions to the current knowledgebase. Therefore, the development and implementation of new MS methodology is the foundational motivation for the research described in this dissertation. A significant amount of effort has been directed to investigations that utilize ion mobility (**IM**)-MS, a MS method that integrates size/shape separation of gas-phase ions, to solve challenging biological problems. Projects can be further segregated into two groups: those that develop novel analytical tools and demonstrate proof-of-principle applications, and those which aim to employ cutting-edge proteomics technology to examine important and challenging biological systems.

This chapter will begin with an overview of important fundamentals, including the general foundations of MS and how MS technology is used to analyze peptides and proteins. This knowledge is crucial to all experiments described in this dissertation. Next, the chapter will move on to a qualitative and quantitative overview of IM-MS, important for **Chapters 2**, **3**, and **4**. Finally, the overall goals and MS approach of each project will be summarized.

Fundamentals of peptide and protein analysis via mass spectrometry

Over the past twenty years, MS has completely revolutionized the study of peptides and proteins. Prior to MS, analyses at the peptide and protein levels were limited by the discovery throughput of Edman Degradation[7] and the quantitative throughput of Western blots[8] or enzyme-linked immunosorbent assays (**ELISA**).[9] It wasn't until the development of "soft ionization" sources, such as electrospray ionization (**ESI**)[10] and matrix-assisted laser desorption ionization (**MALDI**),[11-12] that proteome- and peptidome-wide investigations were even considered a possibility. As methodology matured and MS-based proteomics and peptidomics expanded its influence, the field has now reached an exciting time where entire proteomes of simple organisms can be analyzed in just over one hour[13] and empirical draft maps of the human proteome have been proposed.[14]

All mass spectrometers primarily consist of the three components shown in **Scheme 1**. The first step of any MS analysis is to change neutral analyte molecules into ions. **Scheme 1A** shows the two most utilized ion sources for biomolecules: ESI and MALDI. In ESI, molecules in solution are ionized by a high voltage emitter tip and are sprayed into the gas-phase by repulsive coulombic forces.[15] In contrast, MALDI creates gas-phase ions from solid samples. Analyte molecules are co-crystalized with a small

organic molecule matrix and ablated by an ultra-violet or infrared laser. The matrix absorbs the laser energy and uses it to ionize the analyte via collisions in gas-phase matrix/analyte clusters.[16] Following ionization, ions enter the mass spectrometer and are transferred to the mass analyzer for measurement of their mass-to-charge ratios (***m*/*z***). Two common ways of achieving this step are shown in **Scheme 1B**. A time-of-flight (**TOF**) analyzer uses the time it takes an ion to complete a pre-defined flight path.[17] An orbital trap—known commercially as an Orbitrap—uses the *m*/*z*-dependent axial frequency of ions orbiting a hyperbolic central electrode.[18] Finally, these flight times and frequencies must be detected by generating an ion current or colliding ions into electron multipliers, represented in **Scheme 1C**.

All of the proteomic experiments in this dissertation utilize the shotgun bottom-up approach.[19] The general bottom-up workflow is depicted in **Scheme 2**. Proteomic samples can be prepared from virtually any material that contain proteins, such as cell culture, fluids/blood, or a piece of tissue. First, the sample is homogenized and the cells are lysed to enable the extraction of proteins from any cellular location. When working with tissues with high lipid content, such as brain tissue, additional protein extractions may be performed, such as a methanol/acetone precipitation.[20] Next, proteins are enzymatically digested into smaller peptide fragments. This step is the main difference between bottom-up proteomics and another approach called top-down proteomics.[21] Bottom-up proteomics indirectly identifies and quantifies proteins from enzymatic fragments, while top-down proteomics directly analyzes intact proteins.

Following digestion, peptides are purified via solid-phase extraction and then submitted to liquid chromatography-mass spectrometry (**LC-MS**) analysis. The

proteomics experiments in this dissertation use what is known as data-dependent acquisition (**DDA**). The one exception is **Chapter 4**, which used MS^E, a type of data-independent acquisition (**DIA**).[22] In DDA, peptide mixtures are commonly separated on a reversed-phase LC column with an integrated nano-ESI emitter tip. As eluting peptides are ionized and sprayed into the mass spectrometer, the instrument will go through cycles of choosing ions to isolate and fragment in tandem mass spectrometry (**MS/MS**). The resulting MS/MS spectrum of the fragment *m*/*z* can be used to determine the peptide's sequence. **Figure 1** shows an example of a single DDA cycle during analysis of total protein extract from HEK293 cells. The spectrum on the lower left is a single scan showing all the ionized peptides that eluted 36.9 minutes into a 90 minute LC gradient. In this particular DDA cycle, the instrument chooses the top 12 (number chosen by investigator) most abundant peptides and sequentially performs MS/MS on them. The fragmentation spectra of these peptides are shown in the upper right of **Figure 1**, nine of which yielded a peptide identification. The cycle would continue by taking a new precursor MS scan and choosing the next 12 most abundant precursors until the end of the analysis.

While it is possible to just use the *m*/*z* of fragment ions and sequence peptides *de novo*,[23] the use of a genome-predicted protein database drastically increases the efficiency of identification.[24] The proteome of the species being studied is first digested *in silico* to produce a list of all peptide sequences that could be created from intact proteins. Next, depending on the method of MS/MS fragmentation, a list of theoretical fragments and their *m*/*z* values are assigned to each predicted peptide. For example, MS/MS using higher-energy collision-induced dissociation (**HCD**) would assume b-type

and y-type fragments from cleavage of backbone amide bonds,[25] whereas electron transfer dissociation (**ETD**) would predict c-type and z•-type fragments from radical cleavage of N-C_α bonds.[26] Finally, the MS/MS spectra are matched to the theoretical spectra and given a score to assess the confidence of the match. If an identified peptide can only belong to one protein or one group of highly homologous proteins, that protein or protein group is also considered to have been identified. The experiments in this dissertation have primarily used the Open Mass Spectrometry Search Algorithm (**OMSSA**)[27] and the COMPASS software suite developed by Wegner et al.[28] The confidence of an entire dataset is often evaluated by calculating a false discovery rate (**FDR**) at the peptide and protein identification-level.[29]

Contrary to the wishes of many mass spectrometrists, bioinformatics does not end with protein identification. That is merely the beginning. After generating large datasets, it is important to have automated methods to perform pathway analysis and mine the data for biological meaning.[30] The first-generation of pathway analysis, called Over-Representation Analysis (**ORA**), is very common in MS-based proteomics.[31] The investigator chooses which proteins have significant differential regulation and catalogues them by broad properties such as gene ontology (**GO**) terms. ORA would then determine if any GO terms in the differentially regulated protein list are statistically over-represented compared to a “background” list used for control. The main weakness of ORA is that it does not consider if proteins within an over-represented group actually have any functional connection. Two different proteins may both have the GO term for being located in the cell membrane, but they may not have any consequential interactions. The second-generation of pathway analysis, called Functional Class Scoring (**FCS**),

addresses this problem.[32] Rather than independently choose which proteins have significant differential regulation, the investigator uses all quantitative data from the experiment. Proteins are placed in group where they are known to have some functional connectivity in a biological pathway or phenotype, and the quantitative data of each protein contributes to new quantitative data of a whole group. Thus, specific pathways or phenotypes can be considered differentially regulated and an empirical FDR-Q value can be calculated for each one. Proteomic experiments in this dissertation utilize Protein Set Enrichment Analysis (**PSEA**)-Quant,[33] a version of Gene Set Enrichment Analysis (**GSEA**)[34] modified for MS-based proteomic data.

As a final note on the fundamentals of MS-based proteomics, quantitative MS will be briefly addressed. Obtaining quantitative information about peptides and proteins with MS is not trivial. Differential ionization efficiencies and suppression is a challenge that is unique to MS. While there are effective ways to normalize ion intensities as a means to deal with these challenges,[35] many investigators choose to use chemical tags to assist quantification. Tag-based strategies are used for both precursor ion intensities[36] and MS/MS fragment intensities.[33] Further discussion of quantitative MS strategies can be found in upcoming chapters.

Fundamentals of ion mobility-mass spectrometry

IM is an electrophoretic separation method for gas-phase ions.[37] A weak electrical field pulls ions through a drift area filled with an inert buffer gas, and the time it takes to traverse this region ($\mathbf{t_D}$) is measured. Often, investigators may report Arrival Time, which is the sum of t_D and the mobility-independent travel time from the end of the IM cell to the MS detector. Due to the collisions with buffer gas molecules, ions will travel at different

speeds dictated by mass, charge, and collision cross section (**CCS**). **Scheme 3** illustrates such an IM separation. In **Scheme 3A**, two peptide ions enter an IM drift cell (a tunnel made of ring electrodes) filled with helium at 10 mTorr. Although the peptides have the exact same mass and charge, they have adopted different conformations. **Scheme 3B** shows that peptide 2, the globular conformation, has exited the drift region after 15.0 ms. Peptide 1 has not yet exited the drift cell. It has a more extended helical conformation and a larger CCS than peptide 2, thus it undergoes more collisions with He and travels at a slower speed.

The hypothetical experiment in **Scheme 3** demonstrates how IM-MS affords an additional dimension of separation over standard MS: ion size and shape. The quantitative relationship between t_D and CCS (denoted as Ω) is found in the Mason-Schamp equation:[37]

$$\Omega = \frac{ze}{16N}\frac{760}{P}\frac{T}{273.2}\frac{t_D E}{L}\left[\frac{18\pi}{k_b T}\left(\frac{1}{M}+\frac{1}{m}\right)\right]^{1/2} \qquad (1)$$

N is the number density of the buffer gas, *ze* is the charge of the ion, *P* is the pressure in the drift cell, *T* is the temperature of the drift cell, *L* is the length of the drift cell, *E* is the electric field, *M* is the mass of the analyte, *m* is the mass of a buffer gas molecule, and k_b is Boltzmann's constant. It becomes clear that by measuring an ion's t_D in an IM cell followed by *m*/*z* analysis in a subsequent mass analyzer, one can perform an absolute measurement (i.e. in concrete units such as $Å^2$ or nm^2) of ion shape and size via calculating the CCS. However, it is important to note that **Equation 1** only applies to IM in low-magnitude, static electric fields. The IM-MS experiments in this dissertation utilize a non-linear electric field IM method called traveling wave ion mobility (**TWIM**), wherein ions are pushed down the IM cell by periodic DC pulses that travel down the cell

electrodes like a wave.[38] For TWIM, **Equation 1** has been empirically modified to account for the non-linear electric field:[39]

$$\Omega' = A{t_D}^B \frac{1}{16N}\frac{760}{P}\frac{T}{273.2}\left[\frac{18\pi}{k_b T}\right]^{1/2} \qquad (2)$$

This equation is for Ω', or reduced CCS. Variable E and L have been collapsed into the constant A, and t_D has been raised to the exponent B, which is dependent on the shape of the TWIM waves. Both A and B must be found empirically through non-linear calibration with ions whose CCS are known from static-field IM measurements.[40] Often, all terms on the right side of **Equation 2** besides t_D^B are collapsed to a single constant (A') to simplify calibration. Reduced CCS, Ω', can be converted to Ω as follows:

$$\Omega = \Omega' ze\left(\frac{1}{M}+\frac{1}{m}\right)^{1/2} \qquad (3)$$

Calculation of a CCS can only tell one if an ion's conformation is relatively extended or compact. In-depth structural analysis via IM-MS requires molecular dynamics (**MD**) simulations.[41-42] A parallel experimental and theoretical workflow is shown in **Scheme 4**. The experimental portion is rather straightforward: samples are prepared for IM-MS, drift times are recorded under the desired conditions, and ion CCS are calculated. The first step of the theoretical portion is to create initial starting structures for the simulations. Next, the MD simulations are actually run. If the starting structure came from an empirical source, such as nuclear magnetic resonance or X-ray crystallography, gas-phase simulations or mixed solution and gas-phase simulations at experimental temperatures can be performed.[43] However, in the absence of an empirical starting structure, enhanced sampling methods must be performed to search as much of the potential energy landscape as possible for energetically reasonable conformations.[44-45] After MD, the

theoretical CCS of output structures are calculated.[46] Finally, putative structures with theoretical CCS can be assigned to ions with matching experimental CCS.

Summary of Research Projects

The projects detailed in this dissertation were selected for their overall quality and connection to structural and quantitative analysis of peptides and proteins. A complete list of my published and/or presented work can be found in **Appendix I**. My publication with the most citations, a review about MS imaging,[47] can be found in **Appendix II**. **Chapter 2** through **Chapter 7** contain results from original research. **Chapter 8** concludes the dissertation with a perspective on the impact of this research and suggested routes for follow-up experiments.

Chapter 2 is an analysis of peptide structures by IM-MS. It focuses on a widely used IM-MS "signature" of cis/trans proline isomerization. This signature was observed for a proline-containing mutant of Pre-pro-Neuropeptide Y's signal peptide region. Through combined theoretical and experimental methods, it was suggested that the observed "signature" may have actually resulted from conformations stabilized by trans-proline but not by trans-leucine. This investigation was submitted to *Analyst* for publication.[48]

In **Chapter 3**, the focus of IM-MS is on its analytical utility. IM separation was used to help mitigate inaccuracies caused by isobaric interference in tandem mass tag-based quantitative proteomics and peptidomics. Co-isolated isobaric peptides were able to be separated by IM if they had a difference in charge or size/shape. This investigation resulted in a co-first author publication in *Rapid Communications in Mass Spectrometry*.[49]

The final investigation to use IM-MS in this dissertation is **Chapter 4**. For the first time, the high-throughput peptide identification power of bottom-up proteomics was combined with accurate CCS measurement to create the largest peptide CCS database to-date. Using these databases, we demonstrated how one could find patterns in large datasets to make hypotheses on the structural preferences of peptide groups. This investigation resulted in a first author publication in *The Journal of the American Society for Mass Spectrometry*.[50]

The second half of this dissertation begins with **Chapter 5**: a review on MS-based proteomic applications to the study of chronic pain. In collaboration with the Millennium Pain Center of Bloomington, IL, several proteomic investigations of neuropathic pain were performed. Using the spared-nerve injury (**SNI**) model of neuropathic pain, the initial experiments documenting changes in global protein expression as the SNI phenotype sets can be found in **Chapter 6**. The results were submitted to the *Journal of Proteome Researc*h for publication.[48] Preliminary data from subsequent investigations of molecular mechanisms in neuropathic pain mitigation, via electrical stimulation, can be found in **Appendix IV** and **Appendix V**.

Chapter 7 reports on the final investigation of this dissertation work. Here, novel quantitative 12-plex *N,N*-dimethyl leucine (**DiLeu**) tags, developed by our lab,[51] were applied to global PTM analysis for the first time. To achieve this, custom software called *The DiLeu Tool* was created in C#, the source code of which can be found in **Appendix III**. Using this software, we were able to quantify several thousand phosphorylated proteoforms. The data is currently part of a co-first author manuscript being prepared for *Molecular Cellular Proteomics*.[52]

The work in this dissertation offers contributions both in terms of analytical method development and biological application. The novel use of IM-MS produced new avenues for greater quantitative accuracy from isobaric tags on quadrupole-TOF instruments and performing large-scale peptide structure analysis. Additionally, critical examination of IM-MS cis/trans proline analysis provided a strong argument for always following up with theoretical simulations. Through quantitative proteomics, this dissertation demonstrated that MS-based strategies have a lot to offer the field of pain research, and that use of higher multiplexing affords multi-group comparisons for complex signal pathways.

References

(1) Liddle, R. A. (2015) Peptide hormone signal transduction and regulation, In *UpToDate* (Raby, B. A., Ed.), Waltham, MA, USA.

(2) Krumm, B. E., and Grisshammer, R. *Front. Pharmacol.* **2015** *6*, 48.

(3) Katritch, V., Cherezov, V., and Stevens, R. C. *Annu. Rev. Pharmacol. Toxicol.* **2013** *53*, 531-556.

(4) Kurachi, Y., and North, A. *J. Physiol.* **2004** *554*, 245-247.

(5) Li, L., and Sweedler, J. V. *Annu. Rev. Anal. Chem.* **2008** *1*, 451-483.

(6) Romanova, E. V., and Sweedler, J. V. *Trends Pharmacol. Sci.* **2015** *36*, 579-586.

(7) Edman, P. *Arch. Biochem.* **1949** *22*, 475.

(8) Towbin, H., Staehelin, T., and Gordon, J. *Proc. Natl. Acad. Sci. U.S.A.* **1979** *76*, 4350-4354.

(9) Yalow, R. S., and Berson, S. A. *J. Clin. Invest.* **1960** *39*, 1157-1175.

(10) Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., and Whitehouse, C. M. *Science* **1989** *246*, 64-71.

(11) Tanaka, K., Waki, H., Ido, Y., Akita, S., Yoshida, Y., Yoshida, T. *et al. Rapid Commun. Mass Spectrom.* **1988** *2*, 151-153.

(12) Karas, M., and Hillenkamp, F. *Anal. Chem.* **1988** *60*, 2299-2301.

(13) Hebert, A. S., Richards, A. L., Bailey, D. J., Ulbrich, A., Coughlin, E. E., Westphall, M. S. *et al. Mol. Cell. Proteomics.* **2014** *13*, 339-347.

(14) Wilhelm, M., Schlegl, J., Hahne, H., Gholami, A. M., Lieberenz, M., Savitski, M. M. *et al. Nature* **2014** *509*, 582-587.

(15) Dole, M., Mack, L. L., and Hines, R. L. *J. Chem. Phys.* **1968** *49*, 2240-&.

(16) Knochenmuss, R. *Analyst* **2006** *131*, 966-986.

(17) Wolff, M. M., and Stephens, W. E. *Rev. Sci. Instrum.* **1953** *24*, 616-617.

(18) Zubarev, R. A., and Makarov, A. *Anal. Chem.* **2013** *85*, 5288-5296.

(19) Zhang, Y., Fonslow, B. R., Shan, B., Baek, M.-C., and Yates, J. R. *Chem. Rev.* **2013** *113*, 2343-2394.

(20) Shevchenko, G., Musunuri, S., Wetterhall, M., and Bergquist, J. *J. Proteome Res.* **2012** *11*, 2441-2451.

(21) Catherman, A. D., Skinner, O. S., and Kelleher, N. L. *Biochem. Biophys. Res. Commun.* **2014** *445*, 683-693.

(22) Silva, J. C., Denny, R., Dorschel, C. A., Gorenstein, M., Kass, I. J., Li, G.-Z. *et al. Anal. Chem.* **2005** *77*, 2187-2200.

(23) Medzihradszky, K. F., and Chalkley, R. J. *Mass Spectrom. Rev.* **2015** *34*, 43-63.

(24) Yates, J., III. *J. Am. Soc. Mass Spectrom.* **2015** *26*, 1804-1813.

(25) Olsen, J. V., Macek, B., Lange, O., Makarov, A., Horning, S., and Mann, M. *Nat. Methods* **2007** *4*, 709-712.

(26) Syka, J. E. P., Coon, J. J., Schroeder, M. J., Shabanowitz, J., and Hunt, D. F. *Proc. Natl. Acad. Sci. U.S.A.* **2004** *101*, 9528-9533.

(27) Geer, L. Y., Markey, S. P., Kowalak, J. A., Wagner, L., Xu, M., Maynard, D. M. *et al. J. Proteome Res.* **2004** *3*, 958-964.

(28) Wenger, C. D., Phanstiel, D. H., Lee, M. V., Bailey, D. J., and Coon, J. J. *PROTEOMICS* **2011** *11*, 1064-1074.

(29) Elias, J. E., and Gygi, S. P. *Nat. Methods* **2007** *4*, 207-214.

(30) Khatri, P., Sirota, M., and Butte, A. J. *Plos Comput. Biol.* **2012** *8*.

(31) Khatri, P., and Draghici, S. *Bioinformatics* **2005** *21*, 3587-3595.

(32) Ackermann, M., and Strimmer, K. *BMC Bioinformatics* **2009** *10*, 47.

(33) Lavallée-Adam, M., Rauniyar, N., McClatchy, D. B., and Yates, J. R. *J. Proteome Res.* **2014** *13*, 5496-5509.

(34) Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A. *et al. Proc. Natl. Acad. Sci. U.S.A.* **2005** *102*, 15545-15550.

(35) Cox, J., Hein, M. Y., Luber, C. A., Paron, I., Nagaraj, N., and Mann, M. *Mol. Cell. Proteomics* **2014** *13*, 2513-2526.

(36) Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A. *et al. Mol. Cell. Proteomics* **2002** *1*, 376-386.

(37) Revercomb, H. E., and Mason, E. A. *Anal. Chem.* **1975** *47*, 970-983.

(38) Shvartsburg, A. A., and Smith, R. D. *Anal. Chem.* **2008** *80*, 9689-9699.

(39) Smith, D. P., Knapman, T. W., Campuzano, I., Malham, R. W., Berryman, J. T., Radford, S. E. *et al. Eur. J. Mass Spectrom.* **2009** *15*, 113-130.

(40) Bush, M. F., Campuzano, I. D. G., and Robinson, C. V. *Anal. Chem.* **2012** *84*, 7124-7130.

(41) Tao, L., Dahl, D. B., Pérez, L. M., and Russell, D. H. *J. Am. Soc. Mass Spectrom.* **2009** *20*, 1593-1602.

(42) Wyttenbach, T., Pierson, N. A., Clemmer, D. E., and Bowers, M. T. *Annu. Rev. Phys. Chem.* **2014** *65*, 175-196.

(43) Chen, S.-H., Chen, L., and Russell, D. H. *J. Am. Chem. Soc.* **2014** *136*, 9499-9508.

(44) Chen, L., Shao, Q., Gao, Y.-Q., and Russell, D. H. *J. Phys. Chem. A* **2011** *115*, 4427-4435.

(45) Baumketner, A., Bernstein, S. L., Wyttenbach, T., Bitan, G., Teplow, D. B., Bowers, M. T. *et al. Protein Sci.* **2006** *15*, 420-428.

(46) Mesleh, M. F., Hunter, J. M., Shvartsburg, A. A., Schatz, G. C., and Jarrold, M. F. *J. Phys. Chem.* **1996** *100*, 16082-16086.

(47) Lietz, C. B., Gemperline, E., and Li, L. *Adv. Drug Delivery Rev.* **2013** *65*, 1074-1085.

(48) Lietz, C. B., Tilley, D. M., Kelley, C. A., Cedeno, D. L., Williams, J., Benyamin, R. *et al. J. Proteome Res.* **2015** *Submitted*.

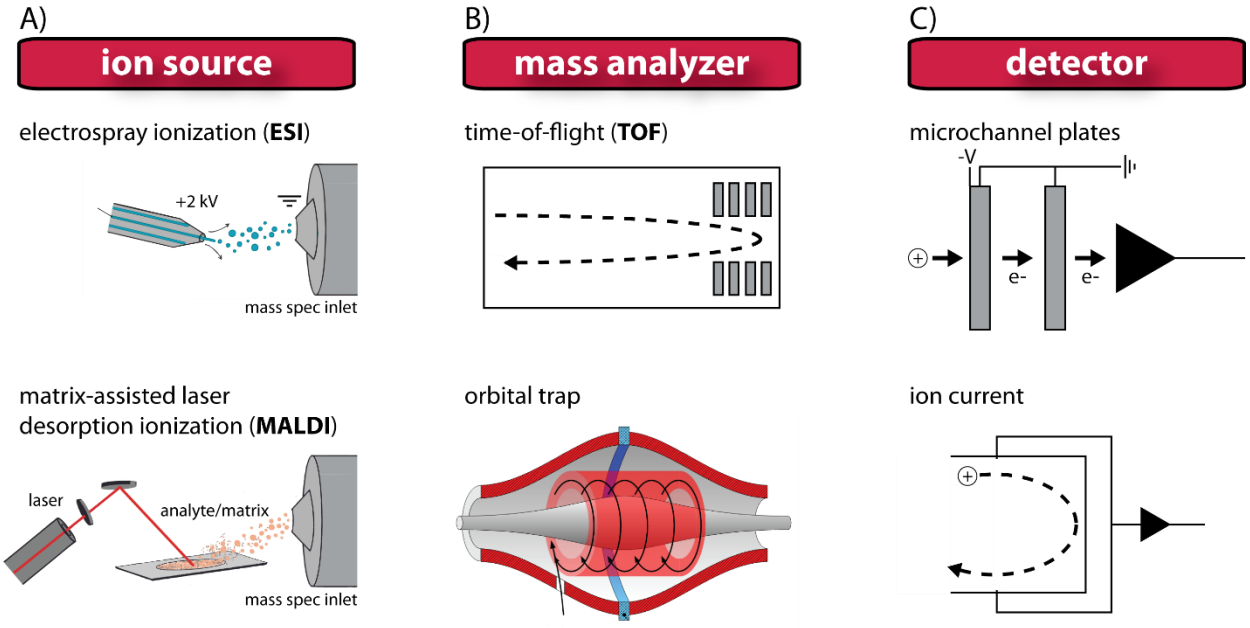
(49) Sturm, R. M., Lietz, C. B., and Li, L. *Rapid Commun. Mass Spectrom.* **2014** *28*, 1051-1060.

(50) Lietz, C. B., Yu, Q., and Li, L. *J. Am. Soc. Mass Spectrom.* **2014** *25*, 2009-2019.

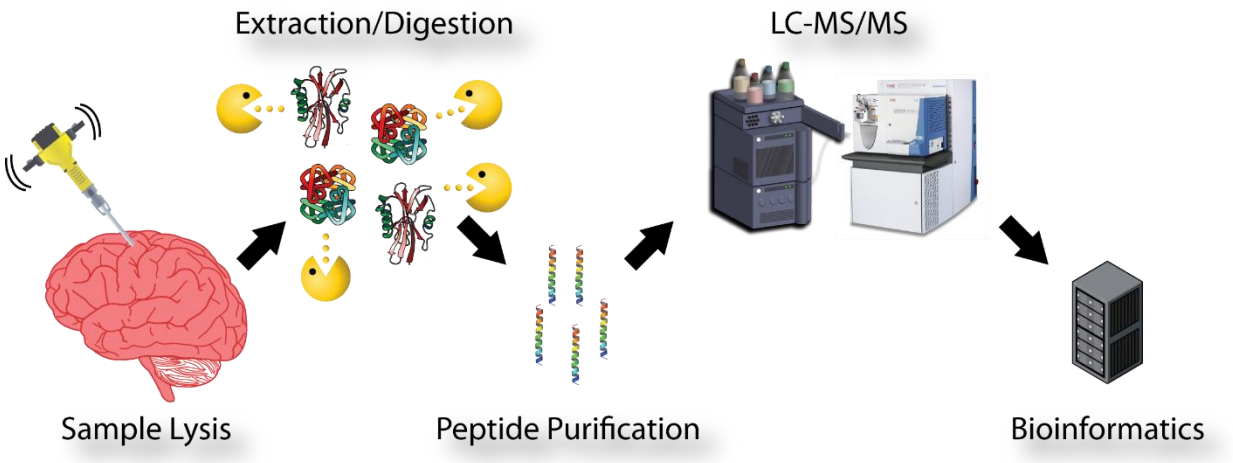
(51) Frost, D. C., Greer, T., and Li, L. *Anal. Chem.* **2015** *87*, 1646-1654.

(52) Zhong, X., Lietz, C. B., Shi, X., Buchberger, A., Frost, D. C., Kent, K. C. *et al. In preparation* **2015**.

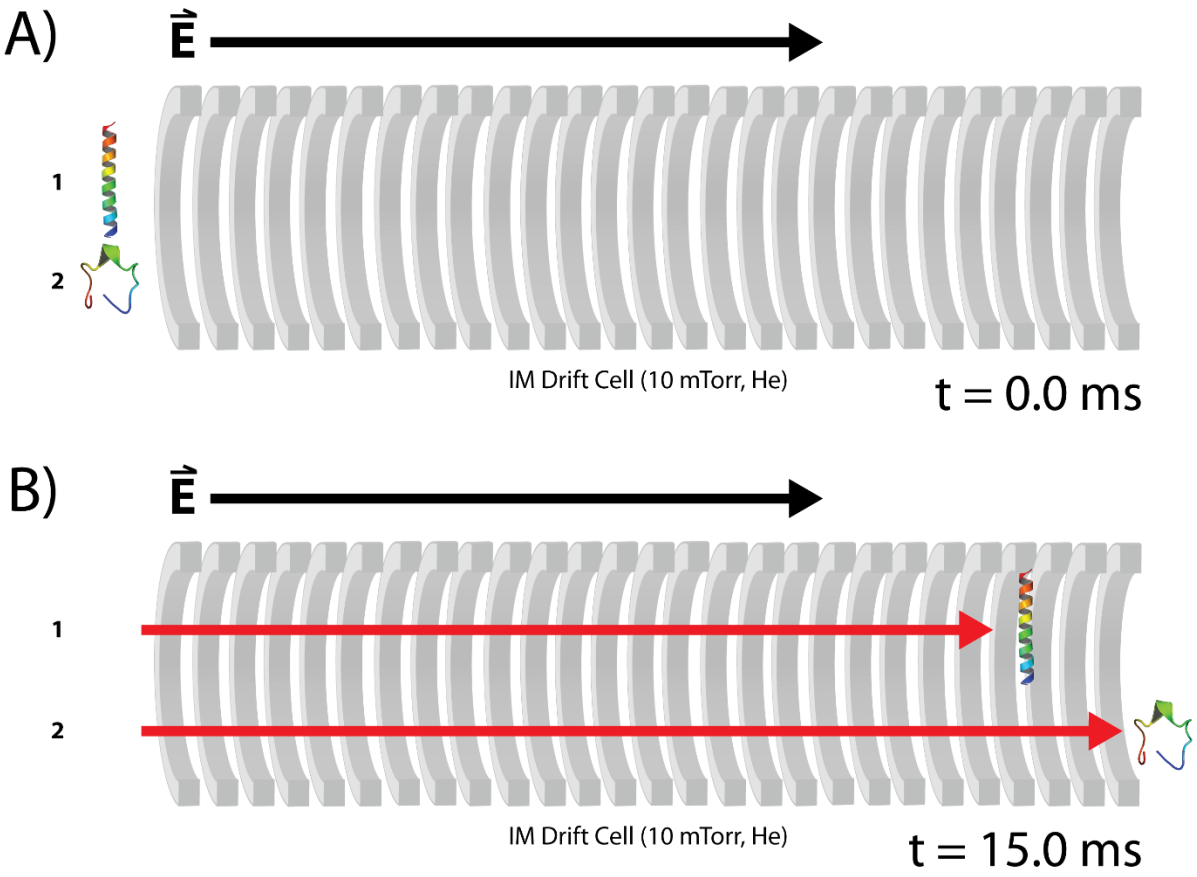
Scheme 1. The three primary components of a mass spectrometer: **A)** ion source, **B)** mass analyzer, and **C)** detector.



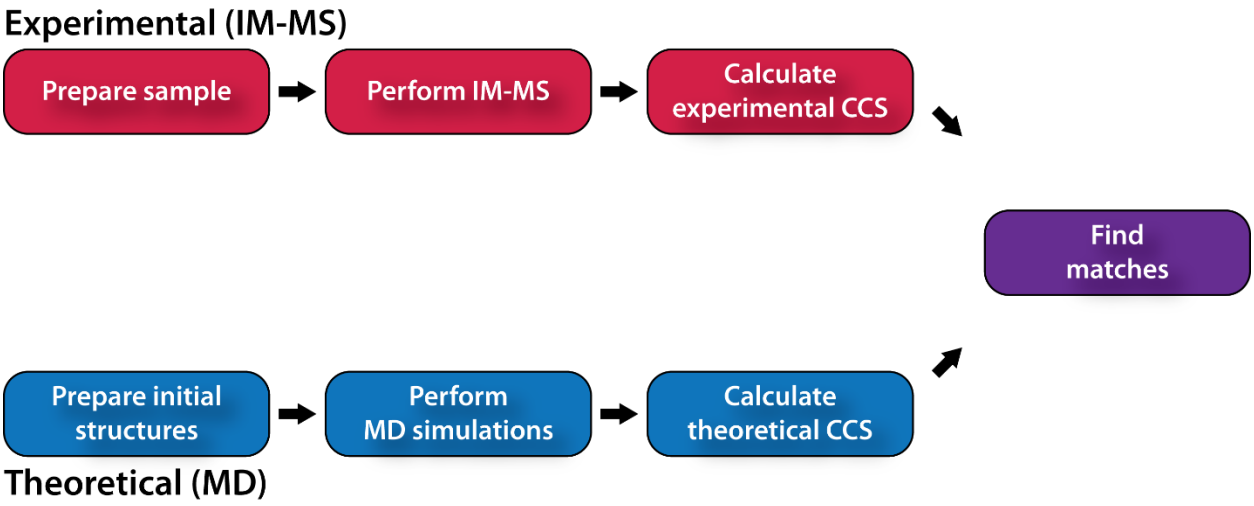
Scheme 2. The general workflow for bottom-up shotgun proteomics.



Scheme 3. IM-MS affords separation of gas-phase molecules by size and shape. **A)** Two peptides with identical mass and charge enter a low-pressure IM drift cell at the same time. **B)** Because ion 2 has a more compact conformation than ion 1, it collides with fewer helium molecules and exit the drift tube first.



Scheme 4. The parallel theoretical and experimental workflows for conformational analysis via IM-MS and MD simulations.



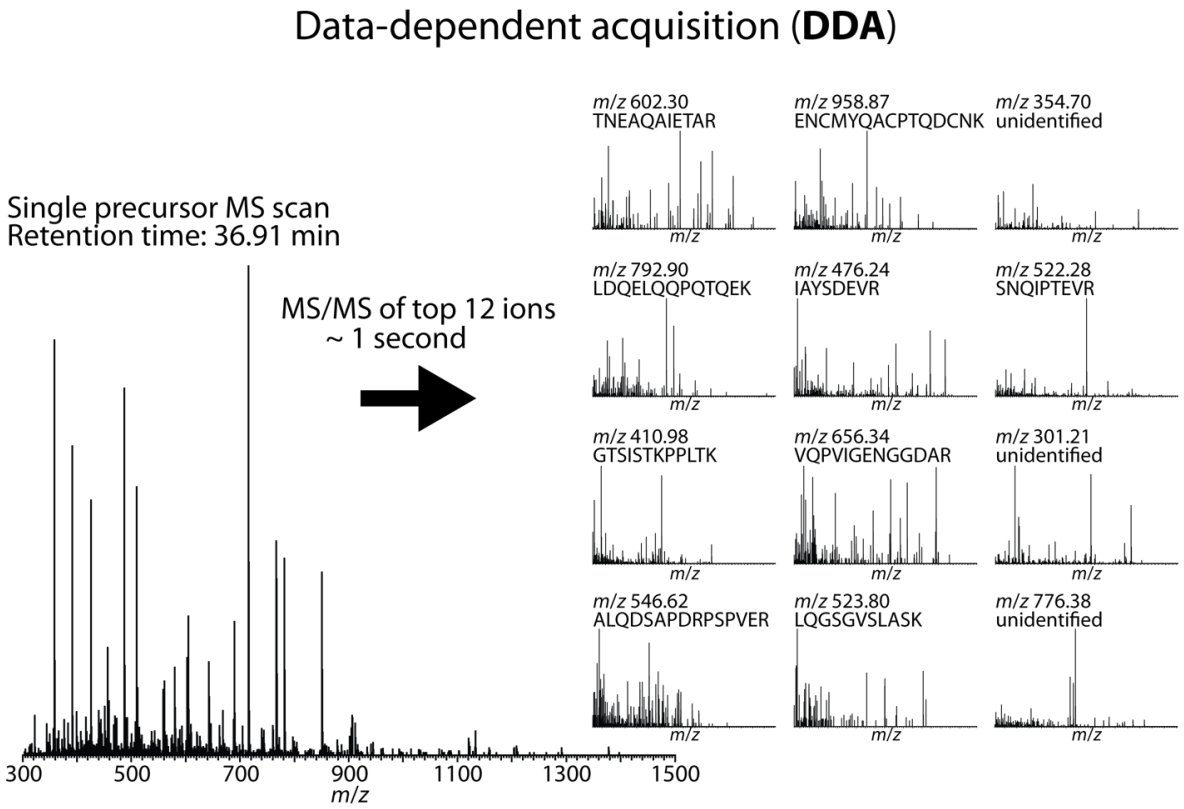


Figure 1. Data-dependent acquisition cycle. **A)** A single precursor scan from the bottom-up proteomic analysis of HEK293 total protein. **B)** Resulting MS/MS spectra of the top 12 most abundant ions chosen from the precursor scan.

Chapter 2

Multiple gas-phase conformations of proline-containing peptides: Is it always cis/trans isomerization?

Adapted from: Christopher B. Lietz, Zhengwei Chen, ChangYun Son, Xueqin Pang, Qiang Cui, and Lingjun Li. Multiple gas-phase conformations of proline-containing peptides: Is it always cis/trans isomerization? *Analyst*. **2015**, submitted.

Abstract

Ion mobility-mass spectrometry (**IM-MS**) is often employed to look at the secondary, tertiary, and quaternary structure of naked peptides and proteins in the gas-phase. Recently, it has offered a unique glimpse into proline-containing peptides and their cis/trans Xxx-Pro isomers. An experimental "signature" has been identified wherein a proline-containing peptide has its Pro residues substituted with another amino acid and the presence or absence of conformations in the IM-MS spectra are observed. If the proline-containing peptide displays all the same IM features that the substituted peptides display, but also displays additional features, those additional peaks are hypothesized to originate from cis-Pro conformers.

Investigators are often careful to state that, despite their confidence, it is possible that cis/trans isomers are not the cause of the additional conformations in proline-containing peptides. However, no one has ever explicitly shown evidence for such a system. Herein, we present the IM-MS analysis of Neuropeptide Y's wild-type (**WT**) signal sequence and Leu7Pro (**L7P**) mutant. Although comparison of arrival times and collision cross sections of $[M+4H]^{4+}$ ions yield the cis/trans "signature", molecular dynamics indicates that a cis-Pro7 is not very stable and that trans-Pro7 conformations of the same cross section arise with equal frequency. We believe this work further stresses the importance of theoretical calculations in IM-MS structural assignments.

Introduction

Ion mobility (**IM**)-mass spectrometry (**MS**) is a gas-phase electrophoretic separation method. Ions are pulled through a buffer gas by a weak electric field and separated by charge and collision cross section (**CCS**).[1] As its implementation expands, IM-MS continues to establish itself as an important tool for the structural characterization of biomolecules.[2-7] One area to which it has provided particular advantage is the analysis of small, flexible, highly disordered peptides. When careful measures are taken to ensure gentle ionization and ion transport, IM-MS can help identify the distributions of preferred peptide conformations and intermediates that may not be resolved in more traditional techniques like NMR.[8]

Proline-containing peptides,[9-14] small proteins,[15-16] and derivatized amino acids[17] have been a frequent target of IM-MS studies. Through the creation and analysis of a large database, Counterman and Clemmer were the first to reveal that proline-containing tryptic peptides were more likely to display multiple IM peaks than those without proline.[18] The investigation ultimately suggested that these multiple peaks arose from populations of cis-proline and trans-proline conformers. Therefore, peptide ions with both cis- and trans-proline may be relatively common in the gas-phase. In a later study, Pierson et al. developed a method to more accurately catalogue which IM peaks originated from cis/trans isomers.[19] The IM-MS spectrum of a proline-containing peptide is compared to the IM-MS spectrum of the same peptide but with its proline residues substituted with a different amino acid (for example, APAAA and AxAAA, where $x \neq P$). Since proline is unique in its ability to have appreciable populations of cis peptide bonds,[20] IM peaks shared with the substituted peptide are thought to emerge from trans-proline

conformations, and peaks unique to the proline-containing peptide are thought to come from cis-proline conformations. This approached helped to solidify the experimental IM-MS "signature" of cis/trans prolines and led to novel insights on the energetics of isomerization[21] and the effect of proline sequence position,[22] a phenomenon which may have consequences on the conformations of endogenous neuropeptides.[23]

Proline isomerization plays a crucial role in the rate of protein folding.[20] Although IM-MS may be poised to play a complementary role for proline analysis in peptide and protein systems, investigators often show prudence by mentioning the possibility that proline-induced multiple conformations may arise from events other than cis/trans isomerization. Therefore, cis/trans proline experiments often incorporate molecular dynamics (**MD**) simulations[24-26] to garner further support for specific designations of cis or trans conformations. Briefly, MD simulations are run in parallel with IM-MS data collection. The theoretical CCS of structures obtained from MD are then matched to the calculated CCS of ions observed in experiments. To our knowledge, no study has presented evidence for phenomena other than cis/trans-isomerization in the appearance of multiple peptide conformations after proline-substitution.

Here, we present the investigation of a proline-containing peptide system with the experimental hallmarks of cis/trans-proline conformers. We hypothesized isomerization as the explanation, but theoretical simulations yielded evidence for an alternative. Given the rise in use and maturation of commercial instrumentation, we believe such a communication is important to provide a concrete example of what previous reports have cautioned.

Experimental

Peptide Synthesis

Peptides were synthesized by the University of Wisconsin Biotechnology Peptide Synthesis Lab (Madison, WI, USA). We chose the 28-residue wild-type signal sequence of Neuropeptide Y (**NPY**) (MLGNKR**L**GLSGLTLALSLLVCLGALAEA), hereafter referred to as WT, and its naturally-occurring Leu7Pro mutant (MLGNKR**P**GLSGLTLALSLLVCLGALAEA), hereafter referred to as L7P, for use in this study. The peptides only differ in the identity of the seventh residue: Leu7 for WT and Pro7 for L7P. Because of the proline substitution, this system has the potential for displaying the IM-MS signatures of cis/trans isomerization.

It is interesting to note that the L7P mutation has strong biological consequences. NPY is initially translated as an inactive pro-hormone, called pro-NPY, and is guided by its N-terminal signal sequence to the endoplasmic reticulum for processing into bioactive NPY.[27] Upon entry into the endoplasmic reticulum, the signal sequence is cleaved from pro-NPY by signal peptidases. As a result of unknown molecular mechanisms, individuals with the L7P mutation have higher levels of bioactive NPY in their blood serum and are more likely to suffer from medical conditions such as diabetes than individuals with the WT sequence.[28-31]

Ion mobility-mass spectrometry

All IM-MS was conducted on the Synapt G2 HMDS (Waters Corporation, Milford, MA, USA) in positive mode with nitrogen as the IM buffer gas. Analyte molecules were dissolved in 50:50 water: methanol to a concentration of 30 µM and injected into mass spectrometer by direct infusion at the flow rate of 0.4 µL/min. DL-Polyalanine (Sigma-Aldrich, St. Louis, MO, USA) was suspended in 50% ACN, 1% HAc at 0.1 mg mL^{-1} and

subsequently used for CCS calibration. Calibration was performed with DriftScope and custom *pepCCScal* software that has been characterized previously.[13] Methods are based off the calibrations described by Bush et al. and used the helium CCS values for calibrant ions.[32] The IM-MS data of both the analyte and calibrant ions were acquired under five different wave velocity (**WV**, m s^{-1})/wave height (**WH**, V) ratios: 500/30, 500/35, 600/35, 700/40 and 800/40. CCS values listed in figure and the main text are the mean values calculated from these settings. Ion transmission settings for the spray voltage, cone voltage, extraction cone, and helium cell DC were kept at 1.5 kV – 2.0 kV, 20 V, 4 V, and 15 V, respectively. Trap bias voltage was usually kept at 35 V – 38 V, except when increased to 60V to induce collisional activation.

As explained by Bush et al.,[32] measuring ATDs in N_2 and using the He-CCS values of polyalanine attempts to account for systematic differences between N_2- and He-CCS and allow the calculation of calibrated He-CCS for unknowns. We acknowledge that there is no fundamental theoretical basis for this, and that recent studies have experimentally identified differences in analyte/gas collisions in N_2 and He.[33] However, we believe the empirical success by Bush et al. and in our own previous studies[13] justifies its use in the current investigation.

Replica-exchange molecular dynamics

Initial structures without hydrogen atoms were created using Avogadro and consisted of the WT and L7P sequences in an ideal linear configuration. Trans-Pro7 peptide ω- and φ-dihedrals were set to 180º and 149º, respectively. Cis-Pro7 peptide ω- and φ-dihedrals were set to 0º and 158º, respectively. MD was then carried out using GROMACS (version 4.6.5) on the University of Wisconsin-Madison Department of

Chemistry Phoenix cluster (NSF Grant CHE-0840494). Hydrogens were added to the initial structures and charging protons were placed on the N-term, Lys5, and Arg6. For simulations of N4H structures, a charging proton was placed on the amide of Asn4. For simulations of L25H structures, a charging proton was placed on the backbone amide of Leu25.

Linear structures were energy-minimized to a tolerance of 10.0 kJ mol^{-1} nm^{-1} (this tolerance was used for all energy minimizations) and then subjected to simulated annealing for 1.55 ns. The cycle began at 300 K and increased by 50 K increments to 1000 K and then back down to 300 K. Increments happened over 9 ps and the system equilibrated for 2 ps after each increment. The final 300 K structure was energy-minimized and used as the "random" starting structure for replica-exchange molecular dynamics (**REMD**).

Each random starting structure was assigned to 34 parallel simulations with the following reference temperatures: 80.00 K, 90.65 K, 101.35 K, 113.45 K, 126.51 K, 140.24 K, 154.64 K, 169.74 K, 185.59 K, 202.26 K, 219.79 K, 238.22 K, 257.62 K, 278.05 K, 299.57 K, 322.25 K, 346.17 K, 371.39 K, 397.97 K, 426.01 K, 455.60 K, 586.80 K, 519.75 K, 554.51 K, 591.23 K, 629.85 K, 670.74 K, 713.9 K, 759.51 K, 807.65 K, 858.48 K, 912.20 K, 968.91 K, and 1028.80 K. The reference temperatures were chosen with the help of free online tools (http://folding.bmc.uu.se/remd/). After allowing the starting structures to equilibrate to their reference temperatures for 200 ps, REMD simulations were carried out for 100 ns. An exchange was attempted every 100 fs. Every 22.5 ps, the current configuration in the 299.57 K simulation (referred to as the "300 K window" in the main text) was output and energy-minimized. When the simulations finished, each

WT ensemble consisted of 4444 outputs and each L7P ensemble consisted of 8888 outputs (from combining the outputs of cis-P7 and trans-P7 initial structures with a given charge configuration).

Theoretical CCS Calculations

The theoretical CCS were calculated via the trajectory method (**TM**) using MOBCAL.[34] The program was compiled to run 2.5 x 10^7 trajectories per output at 300 K. Running MOBCAL TM calculations for 26664 structures was accomplished through the use of the Center for High Throughput Computing (**CHTC**) in Madison, WI. We were allotted an average of 2000 CPUs for parallel calculations, and at times were able to obtain over 4000 CPUs.

Conformation Cluster Analysis

All structures with theoretical CCS that matched within ± 3% of experimental values were kept for conformation cluster analysis by the MaxCluster algorithm (http://www.sbg.bio.ic.ac.uk/~maxcluster/). Clustering was performed as an "All-vs-All" analysis within each CCS-filtered ensemble. Cis-L7P and trans-L7P structures were analyzed separately. Average-linkage clustering was performed by calculating the RMSD of the N, C_α, C, O, and C_β backbone atoms and adjusting the RMSD threshold until an initial cluster containing at least 1/3 of the CCS-filtered ensemble was obtained. The "representative" structure of a cluster is the centroid that has the lowest average RMSD when compared to every other structure in the cluster.

Results and Discussion

IM-MS Results

Figure 1 shows the sequences of the WT and L7P peptides and their nano electrospray ionization (**ESI**) mass spectra. As seen in **Figure 1B**, the most abundant ion species in both spectra are the triply protonated monomers. An inset of the yellow-highlighted region near *m*/*z* 700 is shown in **Figure 1C** and **Figure 1D** and contains 4+ L7P and WT monomers, respectively. Both peptides displayed peaks of slightly higher abundance corresponding to $[M+3H+K]^{4+}$ compared to the $[M+4H]^{4+}$ ions (labeled as $L7P^{4+}$ or WT^{4+}). **Figure 1E** and **Figure 1F** shows the final area of interest, the green-highlighted region near *m*/*z* 1400. These spectra contain the 2+ monomers and low abundance signals from 4+ dimers.

Since protonated monomers were observed for all charge states, we focused further analysis on the $[M+nH]^{n+}$ ions. Subsequently, IM arrival time distributions (**ATDs**) were acquired and CCS were calculated to search for evidence of cis/trans isomers. We hypothesized that peaks with nearly identical CCS common to both WT ad L7P may represent similar conformations with the seventh residue in a trans peptide bond, and additional peaks unique to L7P may indicate a cis Arg6-Pro7 bond. It is important to note that we cannot directly calculate CCS from ATDs acquired on our unmodified SYNAPT G2.[35] We utilized ions with known CCS and a common calibration strategy[32, 36-37] to calculate calibrated CCS for WT and L7P. Our group has previously used this method and obtained peptide CCS with well under 3% error.[13]

The ATDs and calibrated CCS of the 2+ and 3+ ions are shown in **Figure S1**. At these charge states, we did not see any convincing evidence of cis/trans isomers. The broad peaks in these ATDs suggest the presence of multiple unresolved features, thus

we cannot discount the possibility that high resolution instrumentation may reveal a cis/trans-like signature.

Figure 2 shows the ATDs and calibrated CCS of the 4+ WT and L7P ions. These spectra have a very different appearance than their 2+ and 3+ counterparts. Both peptides display a single narrow peak with identical mean CCS of 527 $Å^2$, as well as more extended conformations at 626 $Å^2$ (WT) and 627 $Å^2$ (L7P). Additionally, the $L7P^{4+}$ spectrum contains a third peak at 610 $Å^2$. This was precisely the type of IM-MS profile that would suggest the presence of cis-proline and trans-proline conformers in the $L7P^{4+}$ ions. Therefore, we made the following hypothesis: the $L7P^{4+}$ ions at 527 $Å^2$ and 627 $Å^2$ are trans-proline conformations because they are also found in the WT^{4+} spectrum; the unique $L7P^{4+}$ peak at 610 $Å^2$ is a cis-proline conformation. To test this hypothesis, we performed MD simulations of WT^{4+} and $L7P^{4+}$ ions.

Collisional Activation

As a final consideration before MD, we wanted to see if we could detect any kinetically trapped conformations in our gas-phase ions that may have originated from their conformations in the electrospray solvent.[8] The absence or presence of solution-state memory would determine whether we could carry out our simulations entirely in the gas-phase or if simulations in solution would also be necessary. We accomplished this by using collisional activation, similar to methods used by others.[38] The results can be seen in **Figure 3**. Two ATDs are shown for each peptide, one acquired with a 38 V Trap bias and one acquired with a 60 V Trap bias. The Trap bias is the ion optics voltage that injects ions from a low-pressure argon region to the high-pressure helium region near the entrance of the IM cell. Ions injected with a higher bias will heated to higher temperatures

before undergoing collisional cooling in the helium cell.[39] Activation of WT^{4+} and $L7P^{4+}$ ions both result in a decreased population of the compact conformations and an increased population of extended conformations. This indicates that the 527 $Å^2$ conformations may kinetically trapped intermediates between solution-state and gas phase conformations, but the more extended ions are likely the preferred gas-phase conformations. Therefore, we opted to perform all our MD in the gas-phase.

REMD Results

In building our initial MD structures, we first had to determine the protonation sites. The obvious choices in the WT and L7P sequences (**Figure 1A**) were the N-terminus, K5, and R6. Additionally, the carboxylic acids at E27 and the C-terminus were protonated to make them neutral. However, this only produces a 3+ charge, and so the additional protonation of a neutral residue was required. Using a simplified adaptation of Zhang's calculations for gas phase basicities of peptides,[40] we determined that N4 was a probable location for sidechain protonation (**N4H**), and L25 was probable for backbone amide protonation (**L25H**).

Our computational strategy utilized REMD to achieve broad sampling of the conformational space.[41-43] Simulations were performed separately on six initial structures: WT^{4+} (N4H), WT^{4+} (L25H), $L7P^{4+}$ (N4H) with trans-Pro7, $L7P^{4+}$ (N4H) with cis-Pro7, $L7P^{4+}$ (L25H) with trans-Pro7, and $L7P^{4+}$ (L25H) with cis-Pro7. Each simulation produce 4444 outputs with cis and trans ensembles of the same protonation configuration being combined. A summary of the outputs is shown in **Table 1**. No potential was put in place to confine the R6-P7 bond in cis or trans configuration, and we observed a lot of cis/trans isomerization that initiated in the high temperature windows and trickled down

to 300 K. Therefore, we combined all the L7P outputs of the same charge configuration, leading to the four ensembles listed in **Table 1**. With the exception of WT^{4+} (L25H), nearly 50% of the outputs remained after CCS-filtering. Ratios of cis-Pro7 and trans-Pro7 outputs were nearly the same with or without CCS-filtering, with trans-Pro7 structures representing 85% to 95% of the populations. Here, we see the first indications that a cis-proline may not be very energetically favorable for $L7P^{4+}$ ions.

Figure S2 contains plots of calculated potential energy versus theoretical CCS for all REMD outputs. The cis and trans L7P structures are shown in separate plots. From the distributions, we saw that both WT^{4+} and $L7P^{4+}$ ions cluster much closer to the 527 $Å^2$ experimental CCS with a protonated N4. In fact, the N4H ensembles do not contain any theoretical CCS within 3% of the experimental CCS of the more extended ions. Conversely, the distributions of L25H ions are densely distributed towards larger CCS with the exception of some cis-Pro7 structures that form small clusters near both sets of experimental values. We interpreted these data as evidence for differential charge location producing two sets of 4+ ion conformations present in both WT^{4+} and $L7P^{4+}$ IM-MS spectra. We hypothesize that during electrospray or along the gas-phase unfolding pathway, the N4 sidechain is the preferred protonation site due to accessibility. When collisionally activated, the proton migrates to the more basic L25 backbone amide and reduces coulombic repulsion in the charge-dense N-terminal region.

Conformation Cluster Results

Initial analysis of the REMD simulations leaves the question of cis/trans isomers in $L7P^{4+}$ unanswered. In fact, with the extended $L7P^{4+}$ conformers only having a 2.7% difference in CCS, they sit right on the edge of being theoretically indistinguishable. Our

hope was that MaxCluster analysis of the CCS-filtered ensembles would reveal whether or not cis and trans structures would form tight, discrete clusters around a particular experimental CCS.

Representative backbone structures of the most populated WT^{4+}, trans $L7P^{4+}$, and cis $L7P^{4+}$ clusters with L25H protonation are shown in **Figure 4**. Backbone ribbons are color-coded to denote the N-terminus (blue) and C-terminus (red). The structures are largely characterized by a N-terminal hairpin with β-strands from residues 1-5 and 12-15, followed by an elongated middle region and a sharp turn near the C-terminus. Intramolecular solvation of the termini are remarkably similar between Clusters 1, 3, and 5. The charged N-terminal amine is primarily stabilized by neutral carboxylic acids from E27 and the C-term, and some combination of backbones from residues 15-17 provide additional N-term solvation. In Cluster 1, the C-term acid appears to only form polar contacts with the N-term, whereas the C-terms of Cluster 3 and Cluster 5 are close enough to be solvated by the L16 backbone as well. Cluster 6, a cis-proline cluster, is the only L25H cluster that comes within 3% of 527 $Å^2$. While it bears some resemblance to the other double hairpin structures, its broad N-term turn and tangled mid-region cause it to adopt a more globular conformation with a lower CCS.

The WT^{4+} peak at 626 $Å^2$ in **Figure 2** is narrow enough to be explained by a single dominant species, thus we assign it the only dominant Cluster from the CCS-filtered WT^{4+} L25H ensemble, Cluster 1. For the $L7P^{4+}$ counterpart at 627 $Å^2$, we can assign Cluster 3. Although Cluster 1 and Cluster 3 have very strong visual similarities, we do note that the mean theoretical CCS are different. Our experiments indicate they should be nearly identical. One possibility could be error in our calibrated CCS of the extended 4+ ions.

Using the same calibration strategy, Bush et al. noted higher calibrated CCS error in some 4+ peptides.[32] The source of the error was not conclusively determined, but it was suggested that the lack of 4+ calibrants may have played a role. It's possible that our experimental calibrated CCS of the extended 4+ conformations contains similar errors and the actual values may be smaller. Additionally, it's important to remember that the CCS-filtered $L7P^{4+}$ ensembles contained structures that matched within ± 3% of 610 $Å^2$. If we include structures in that same range for the CCS-filtered WT4+ L25H ensemble as well, the MaxCluster results are shown in **Figure S3**. The total population of the CCS-filtered ensemble increases to 1368 with primary cluster's mean CCS at 612 $Å^2$.

The assignment of 610 $Å^2$ $L7P^{4+}$ ion in **Figure 2** is left to two remaining options: the trans-proline Cluster 4 and the cis-proline Cluster 5. Their slightly smaller theoretical CCS is owed to a shallow bend in the peptides' mid-regions. The N-term hairpin peaks at Gly8 in Cluster 3, but the cis-proline of Cluster 5 seems to force the turn one residue earlier at Pro7. In order for the favorable N-term/C-term interactions to still occur, the entire structure must bend in the middle to bring the termini close together. A similar phenomena could be caused in Cluster 4 by differences in the C-term turning point, also forcing a bend to maintain the N-term/C-term contacts. We note that similar backbone bends are seen in the alternative WT^{4+} L25H cluster (**Figure S3**), but C-terminal coil and perpendicular curvature of the N-terminal region may explain the larger theoretical CCS.

Both Clusters 4 and 5 have low relative populations, and both have the same relative difference in theoretical CCS compared to Cluster 3. Combined with the apparent general instability of cis-Pro7 in $L7P^{4+}$ ions, we are forced to admit that our theoretical evidence suggests the conformation at 610 $Å^2$ may not contain a cis-proline. The

hypothesis that the multiple peaks of $L7P^{4+}$ are due to cis/trans proline isomerization was not conclusively supported.

Summary of Structural Assignments

Our $L7P^{4+}$ assignments are summarized in **Figure 5**. Clusters 9 and 11 come from the N4H clusters in **Figure S4**. Their compact shapes likely arise from all charge being densely located near the N-terminus, allowing residues near the C-terminus to provide backbone solvation in globular conformations. Due to similar theoretical CCS and lower relative populations of the dominant clusters, we hypothesize that any of these clusters—or combinations of clusters—could be present in the peak. Additionally, we do not discount the possibility that small amounts of Cluster 6-like structures could be present in that same region. In it interesting to consider the implications of cis and trans structures existing in the narrow peak at 3.0 ms in **Figure 5**. It is possible the globular nature of the N4H conformations effectively masks what would show up as CCS differences in more elongated conformations.

The extended $L7P^{4+}$ conformation with the largest experimental CCS was assigned Cluster 3, and the assignment of the smaller extended peak remains ambiguous. From the universally low populations of cis structures for $L7P^{4+}$, our inclination is that Cluster 4 is more likely. One could certainly make the argument for a combination of Clusters 4 and 5, but the fact that evidence suggests cis/trans isomerization may not be the cause for multiple conformations in $L7P^{4+}$ remains unchanged.

Although we are confident in our conclusions, reached after careful examination of the experimental and theoretical evidence, alternatives could be offered. One possibility

for the multiple conformations in $L7P^{4+}$ could be additional protonation sites. Just as N4H and L25H configurations lead to different conformations, another backbone or sidechain protonation site, somehow stabilized by the proline-substitution, might do the same. While our data and considerations from Zhang[40] support N4H and L25H as very reasonable assignments, different charge sites would still support the conclusion that our cis/trans isomerization-like experimental data could be explained by alternative phenomena.

If the elongated conformations in the $L7P^{4+}$ ATDs only contain trans peptide bonds between residues 6 and 7, a new question arises: why are WT^{4+} ions unable to display two extended conformations? It would likely be the case that Pro7—be it cis or trans—is somehow essential to preventing full elongation of the backbone between the N-terminal and C-terminal turns. It's possible that the intrinsic rigidity of proline increases the propensity of the hairpin's apex to occur at the seventh residue compared with leucine. Thus, WT4+ might always produce a turn with the apex at Gly8 and never be forced to bend its elongated backbone for the N-term/C-term interactions. This, of course, is just speculation. The effect of proline-substitution on other peptides with N-terminal and C-terminal turns would provide an interesting system for future experiments.

Conclusion

All evidence considered, the answer to the question posed by our title, "is it always cis/trans isomerization?" is no. To be fair, however, previous investigations never said it was. Our aim was to explicitly demonstrate that systems with uncanny experimental resemblance to the cis/trans isomerization "signature" can reveal a different origin when pressed further. One cannot solely rely on experimental IM-MS data for structural

analysis. This study should not be seen as refutation of any previously published IM-MS cis/trans assignments, as many of those investigations do provide theoretical validation. Some experiments even utilized additional MS methods for validation, such as ultraviolet photodissociation[15-16] and ion spectroscopy,[12] which can take an investigation much further than MD alone. As more researchers utilize IM-MS for cis/trans analysis, we encourage that trend to continue.

Acknowledgements

This work is supported in part by the National Science Foundation grant (CHE-1413596 to LL, CHE-0840494 to the UW Phoenix cluster) and the National Institutes of Health grants (1R01DK071801 and 1R56DK071801). It also utilized the compute resources and assistance of the UW-Madison Center For High Throughput Computing (CHTC) in the Department of Computer Sciences. The CHTC is supported by UW-Madison, the Advanced Computing Initiative, the Wisconsin Alumni Research Foundation, The Wisconsin Institutes for Discovery, and the National Science Foundation, and is an active member of the Open Science Grid, which is supported by the National Science Foundation and the U.S. Department of Energy's Office of Science. C.L. acknowledges an NIH-supported Chemistry Biology Interface Training Program Predoctoral Fellowship (grant number T32-GM008505) and an NSF Graduate Research Fellowship (DGE-1256259). LL acknowledges an H. I. Romnes Faculty Research Fellowship.

References

(1) Revercomb, H. E., and Mason, E. A. *Anal. Chem.* **1975** *47*, 970-983.

(2) von Helden, G., Wyttenbach, T., and Bowers, M. T. *Science* **1995** *267*, 1483-1485.

(3) Clemmer, D. E., Hudgins, R. R., and Jarrold, M. F. *J. Am. Chem. Soc.* **1995** *117*, 10141-10142.

(4) McLean, J. A., Ruotolo, B. T., Gillig, K. J., and Russell, D. H. *Int. J. Mass Spectrom.* **2005** *240*, 301-315.

(5) Benesch, J. L. P., and Ruotolo, B. T. *Curr. Opin. Struct. Biol.* **2011** *21*, 641-649.

(6) Jurneczko, E., and Barran, P. E. *Analyst* **2011** *136*, 20-28.

(7) Lanucara, F., Holman, S. W., Gray, C. J., and Eyers, C. E. *Nat. Chem.* **2014** *6*, 281-294.

(8) Pierson, N. A., Chen, L., Valentine, S. J., Russell, D. H., and Clemmer, D. E. *J. Am. Chem. Soc.* **2011** *133*, 13810-13813.

(9) Counterman, A. E., and Clemmer, D. E. *J. Phys. Chem. B* **2004** *108*, 4885-4898.

(10) Kim, T. Y., Valentine, S. J., Clemmer, D. E., and Reilly, J. P. *J. Am. Soc. Mass Spectrom.* **2010** *21*, 1455-1465.

(11) Shi, L., Holliday, A. E., Glover, M. S., Ewing, M. A., Russell, D. H., and Clemmer, D. E. *J. Am. Soc. Mass Spectrom.* **2015**.

(12) Masson, A., Kamrath, M. Z., Perez, M. A. S., Glover, M. S., Rothlisberger, U., Clemmer, D. E. *et al. J. Am. Soc. Mass. Spectrom.* **2015** *26*, 1444-1454.

(13) Lietz, C. B., Yu, Q., and Li, L. J. *J. Am. Soc. Mass Spectrom.* **2014** *25*, 2009-2019.

(14) Counterman, A. E., and Clemmer, D. E. *J. Am. Chem. Soc.* **2001** *123*, 1490-1498.

(15) Warnke, S., Baldauf, C., Bowers, M. T., Pagel, K., and von Helden, G. *J. Am. Chem. Soc.* **2014** *136*, 10308-10314.

(16) Warnke, S., von Helden, G., and Pagel, K. *PROTEOMICS* **2015** *15*, 2804-2812.

(17) Flick, T. G., Campuzano, I. D. G., and Bartberger, M. D. *Anal. Chem.* **2015** *87*, 3300-3307.

(18) Counterman, A. E., and Clemmer, D. E. *Anal. Chem.* **2002** *74*, 1946-1951.

(19) Pierson, N. A., Chen, L., Russell, D. H., and Clemmer, D. E. *J. Am. Chem. Soc.* **2013** *135*, 3186-3192.

(20) Wedemeyer, W. J., Welker, E., and Scheraga, H. A. *Biochemistry* **2002** *41*, 14637-14644.

(21) Pierson, N. A., and Clemmer, D. E. *Int. J. Mass Spectrom.* **2015** *377*, 646-654.

(22) Glover, M. S., Shi, L., Fuller, D. R., Arnold, R. J., Radivojac, P., and Clemmer, D. E. *J. Am. Soc. Mass Spectrom.* **2015** *26*, 444-452.

(23) Glover, M. S., Bellinger, E. P., Radivojac, P., and Clemmer, D. E. *Anal. Chem.* **2015** *87*, 8466-8472.

(24) Tao, L., Dahl, D., Pérez, L., and Russell, D. *J. Am. Soc. Mass Spectrom.* **2009** *20*, 1593-1602.

(25) Fernandez-Lima, F. A., Wei, H., Gao, Y. Q., and Russell, D. H. *J. Phys. Chem. A* **2009** *113*, 8221-8234.

(26) Wyttenbach, T., Pierson, N. A., Clemmer, D. E., and Bowers, M. T. *Annu. Rev. Phys. Chem.* **2014** *65*, 175-196.

(27) Pedrazzini, T., Pralong, F., and Grouzmann, E. *Cell. Mol. Life Sci.* **2003** *60*, 350-377.

(28) Karvonen, M. K., Pesonen, U., Koulu, M., Niskanen, L., Laakso, M., Rissanen, A. *et al. Nat. Med.* **1998** *4*, 1434-1437.

(29) Mitchell, G. C., Wang, Q., Ramamoorthy, P., and Whim, M. D. *J. Neurosci.* **2008** *28*, 14428-14434.

(30) Ukkola, O., and Kesaniemi, Y. A. *Eur. J. Clin. Nutr.* **2007** *61*, 1102-1105.

(31) Ding, B., Kull, B., Liu, Z., Mottagui-Tabar, S., Thonberg, H., Gu, H. F. *et al. Regul. Pept.* **2005** *127*, 45-53.

(32) Bush, M. F., Campuzano, I. D. G., and Robinson, C. V. *Anal. Chem.* **2012** *84*, 7124-7130.

(33) Larriba-Andaluz, C., Fernandez-Garcia, J., Ewing, M. A., Hogan, C. J., and Clemmer, D. E. *Phys. Chem. Chem. Phys.* **2015** *17*, 15019-15029.

(34) Mesleh, M. F., Hunter, J. M., Shvartsburg, A. A., Schatz, G. C., and Jarrold, M. F. *J. Phys. Chem.* **1996** *100*, 16082-16086.

(35) Shvartsburg, A. A., and Smith, R. D. *Anal. Chem.* **2008** *80*, 9689-9699.

(36) Ruotolo, B. T., Benesch, J. L. P., Sandercock, A. M., Hyung, S.-J., and Robinson, C. V. *Nat. Protoc.* **2008** *3*, 1139-1152.

(37) Smith, D., Knapman, T., Campuzano, I., Malham, R., Berryman, J., Radford, S. *et al. Eur. J. Mass Spectrom.* **2009** *15*, 113-130.

(38) Pierson, N. A., Valentine, S. J., and Clemmer, D. E. *J. Phys. Chem. B* **2010** *114*, 7777-7783.

(39) Merenbloom, S., Flick, T., and Williams, E. *J. Am. Soc. Mass Spectrom.* **2012** *23*, 553-562.

(40) Zhang, Z. *Anal. Chem.* **2004** *76*, 3908-3922.

(41) Baumketner, A., Bernstein, S. L., Wyttenbach, T., Bitan, G., Teplow, D. B., Bowers, M. T. *et al. Protein Sci.* **2006** *15*, 420-428.

(42) Albrieux, F., Calvo, F., Chirot, F., Vorobyev, A., Tsybin, Y. O., Lepère, V. *et al. J. Phys. Chem. A* **2010** *114*, 6888-6896.

(43) Chirot, F., Calvo, F., Albrieux, F., Lemoine, J., Tsybin, Y., and Dugourd, P. *J. Am. Soc. Mass Spectrom.* **2012** *23*, 386-396.

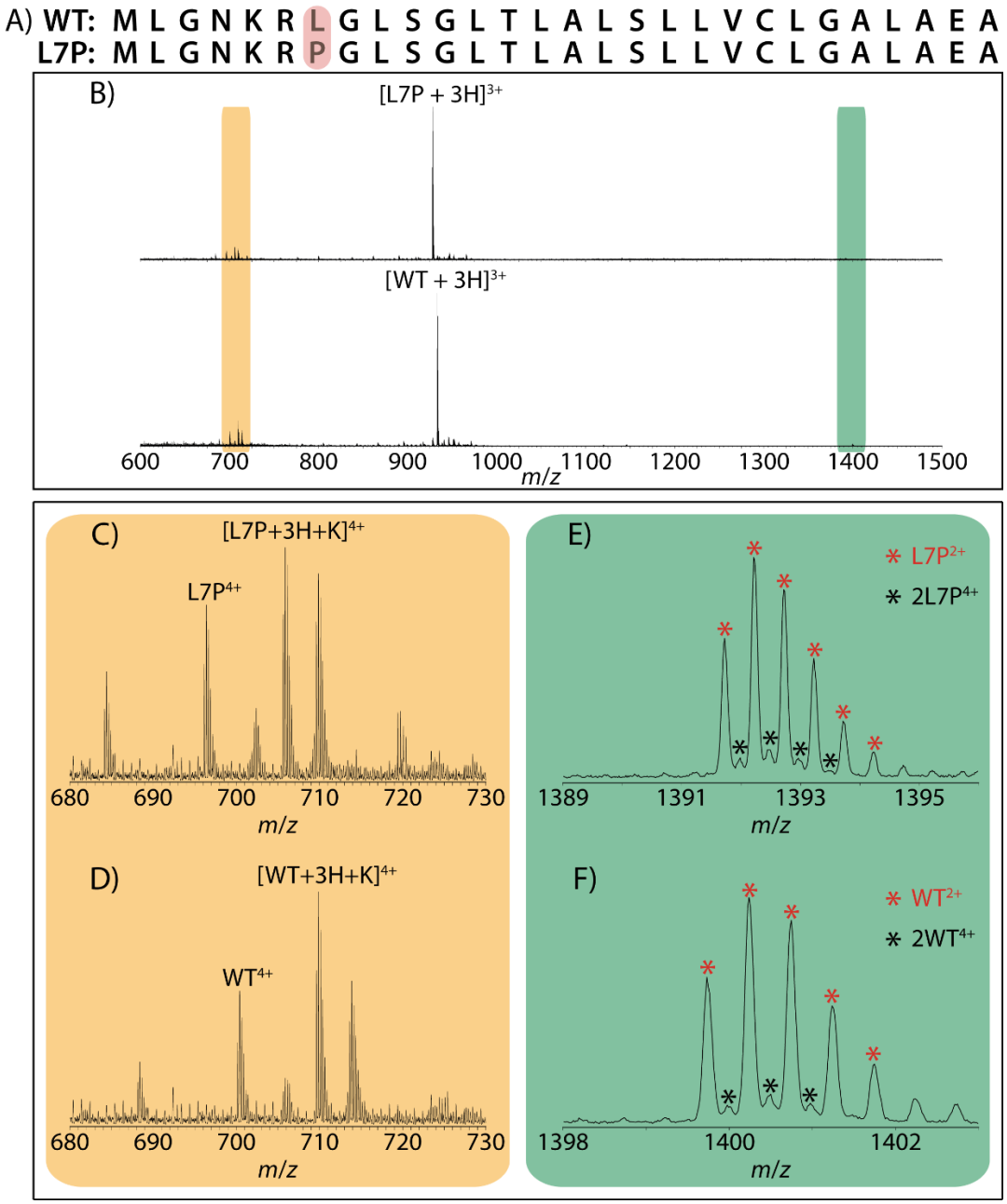


Figure 1. Mass spectrometry of WT and L7P peptides. A) The sequences of WT and L7P, which only differ in the identity of the seventh residue (highlighted in red). **B)** The direct-infusion nano electrospray ionization mass spectra of WT and L7P. Insets of spectral regions containing 4+ ions (highlighted in yellow) and 2+ ions (highlighted in green) are shown below in **C) – D)** and **E) – F)**, respectively.

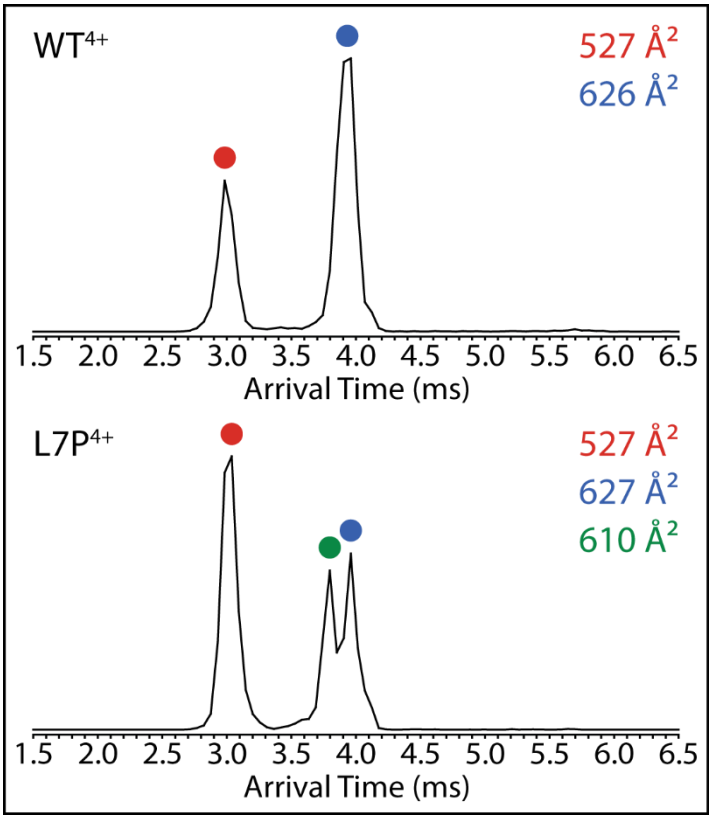


Figure 2. ATDs and calibrated CCS for 4+ peptide ions. ATDs were acquired at IM cell wave height and wave velocity settings of 40 V and 600 m s^{-1}, respectively. Reported CCS are mean values from calibrations and measurements taken at several different wave heights and wave velocities.

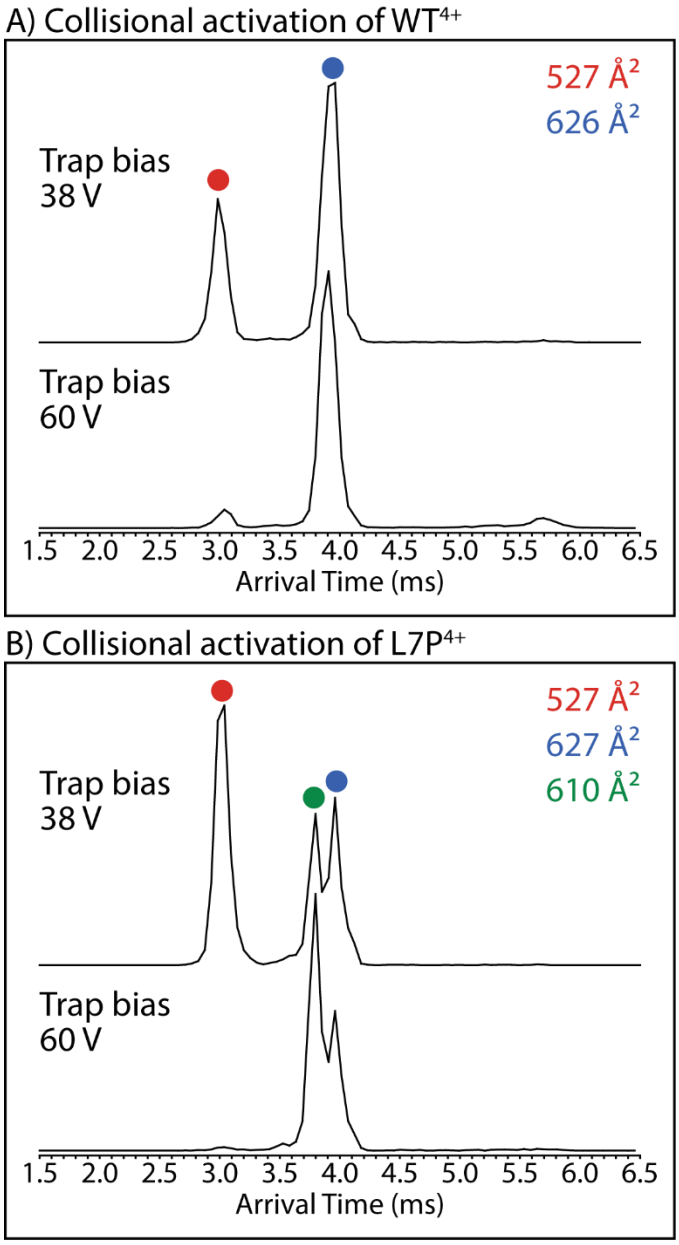


Figure 3. Collisional activation of 4+ peptide ions. ATDs of **A)** WT^{4+} and **B)** $L7P^{4+}$ ions resulting from measurements taken with Trap Bias settings of 38 V and 60 V. ATDs were acquired at IM cell wave height and wave velocity settings of 40 V and 600 m s^{-1}, respectively. Reported CCS are mean values from calibrations and measurements taken at several different wave heights and wave velocities.

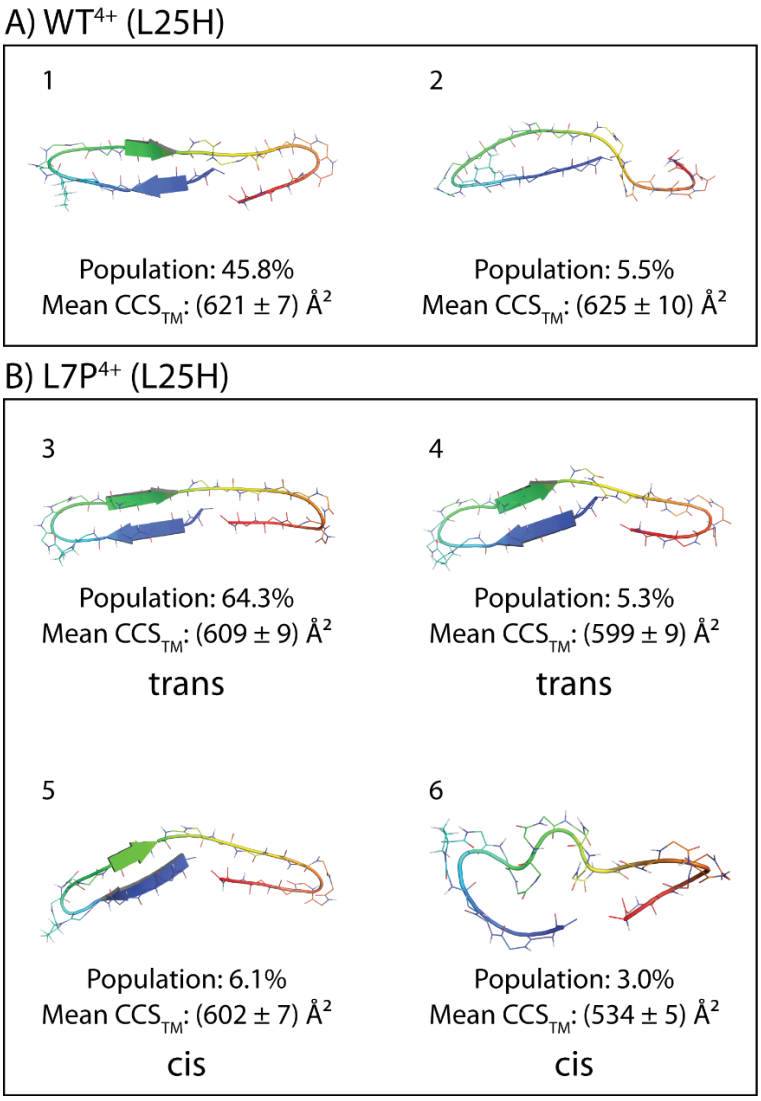


Figure 4. Candidate structures for L25H 4+ ions. Representative centroid structures from the highest population conformation clusters of **A)** WT^{4+} and **B)** $L7P^{4+}$ L25H ions. Atoms from the entire backbone and seventh residue's sidechain are displayed. Structures are color-coded to denote N-terminal (blue) and C-terminal (red) regions. Relative cluster population and mean theoretical CCS of the cluster are listed below the centroid structure. The total populations of the WT^{4+} and $L7P^{4+}$ L25H CCS-filtered ensembles were 758 and 4240, respectively.

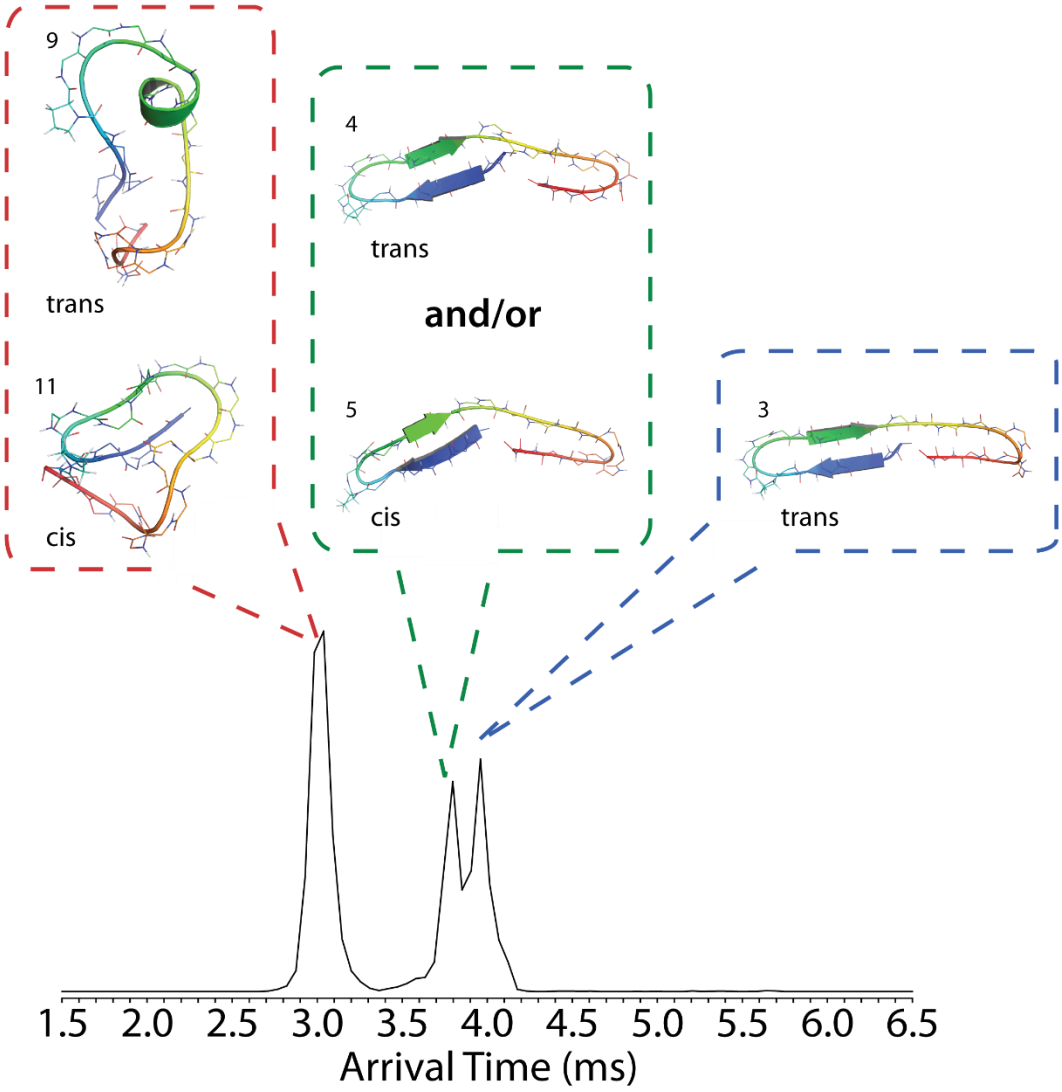


Figure 5. Putative structural assignments for L7P^{4+} ions. Representative cluster centroids are shown by the L7P^{4+} ATD to which they were putatively assigned. Identity of the proline-type in for the peak near 3.8 ms remains ambiguous. ATDs were acquired at IM cell wave height and wave velocity settings of 40 V and 600 m s^{-1}, respectively.

Table 1. Summary of REMD Ensembles

Ensemble	Total Population	Total TRANS/CIS	CCS-filtered Population	CCS-filtered TRANS/CIS
WT^{4+} (N4H)	4444	-	2513 (56.5%)	-
$L7P^{4+}$ (N4H)	8888	7931/957 (89.2%,10.8%)	4565 (51.3%)	3870/695 (84.8%/15.2%)
WT^{4+} (L25H)	4444	-	758 (17%)	-
$L7P^{4+}$ (L25H)	8888	8040/848 (90.4%/9.6%)	4240 (47.7%)	3760/480 (88.7%/11.3%)

Supplemental Information

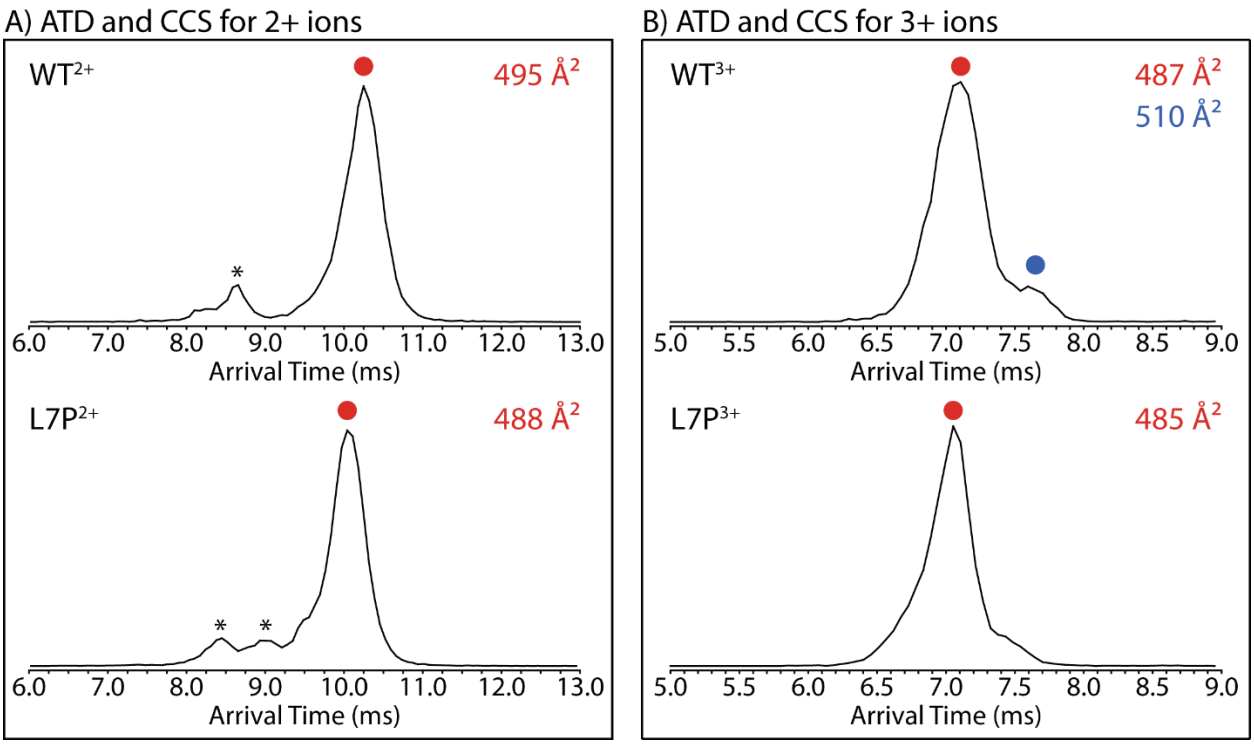


Figure S1. IM-MS ATDs and CCS for 2+ and 3+ ions. A) The ATDs and experimental calibrated CCS for WT^{2+} and $L7P^{2+}$ ions. Black asterisks denote peaks corresponding to 4+ dimers. The 2+ ATDs were acquired at wave height and wave velocity settings of 35 V and 500 m s^{-1}, respectively. **B)** The ATDs and experimental calibrated CCS for WT^{3+} and $L7P^{3+}$ ions. The 3+ ATDs were acquired at wave height and wave velocity settings of 30 V and 500 m s^{-1}, respectively. All reported CCS are mean values from acquisitions at wave height and wave velocity settings specified in **Experimental** section.

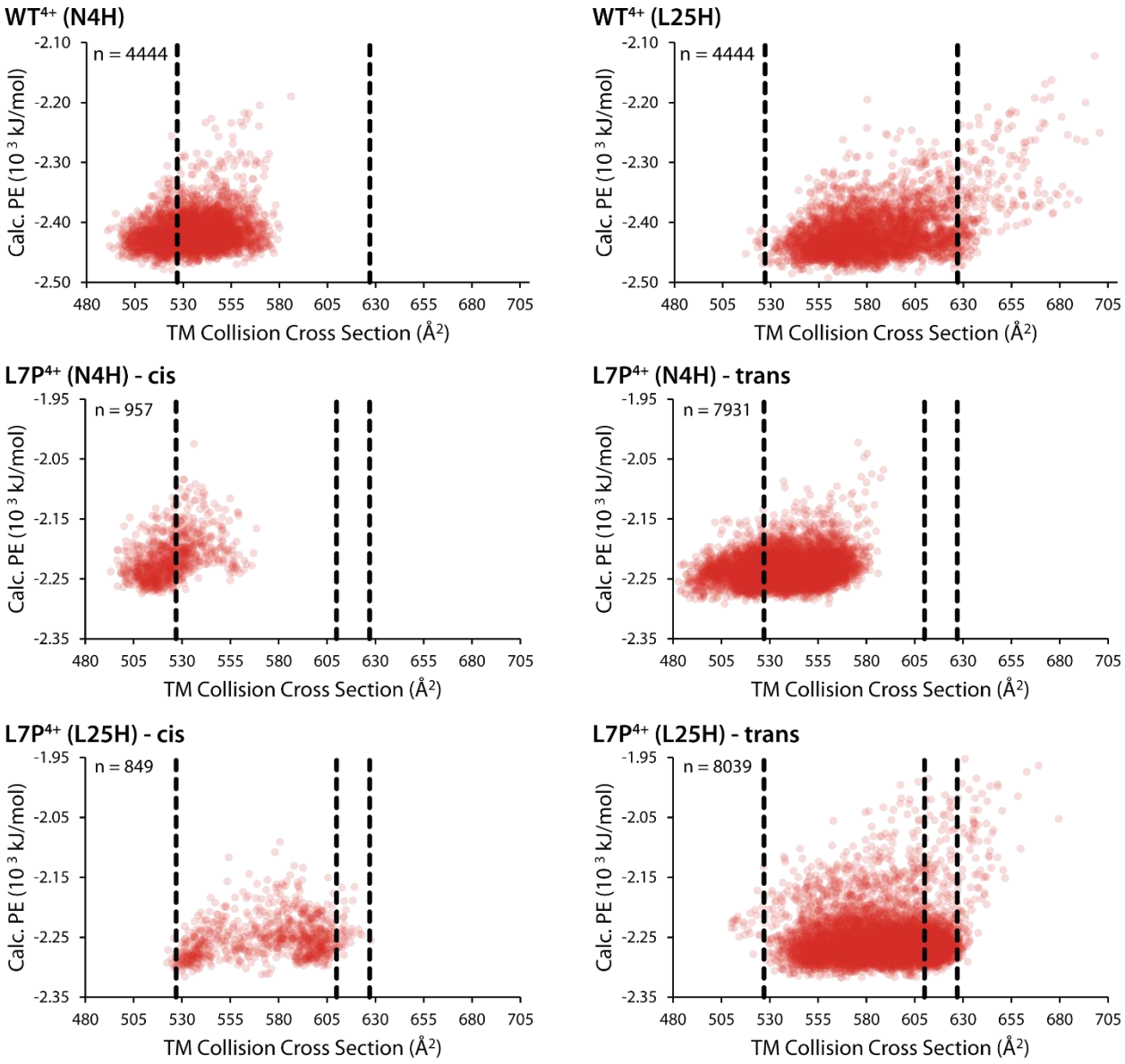


Figure S2. PE versus CCS. Plots of calculated minimized potential energy versus theoretical CCS for all REMD outputs. Dotted lines denote the experimental calibrated CCS.

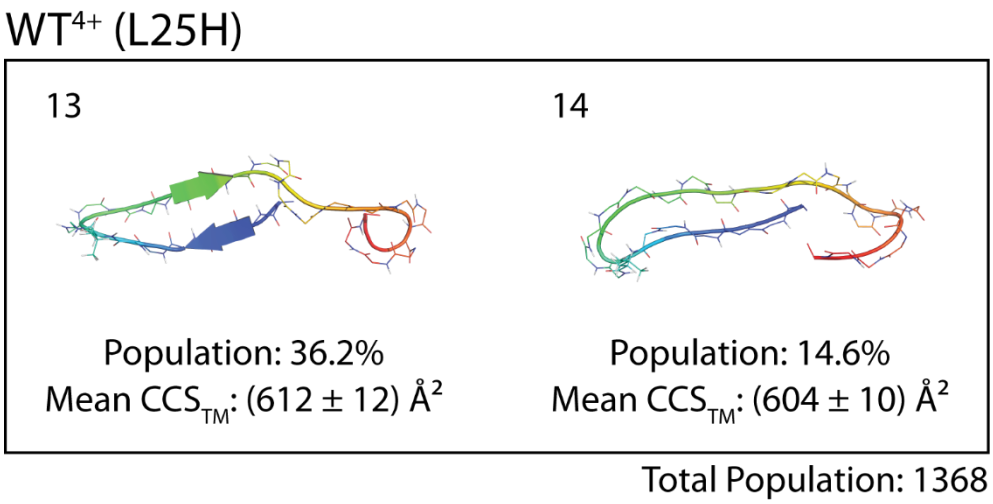


Figure S3. Alternative WT^{4+} clusters for L25H structures. Representative centroid structures for WT^{4+} L25H clusters when outputs matching within ± 3% of 610 $Å^2$ in addition to those that match within ± 3% of 626 $Å^2$. Atoms from the sidechain of the seventh residue and the entire backbone are displayed. The backbone is color-coded to denote the N-terminus (blue) and C-terminus (red).

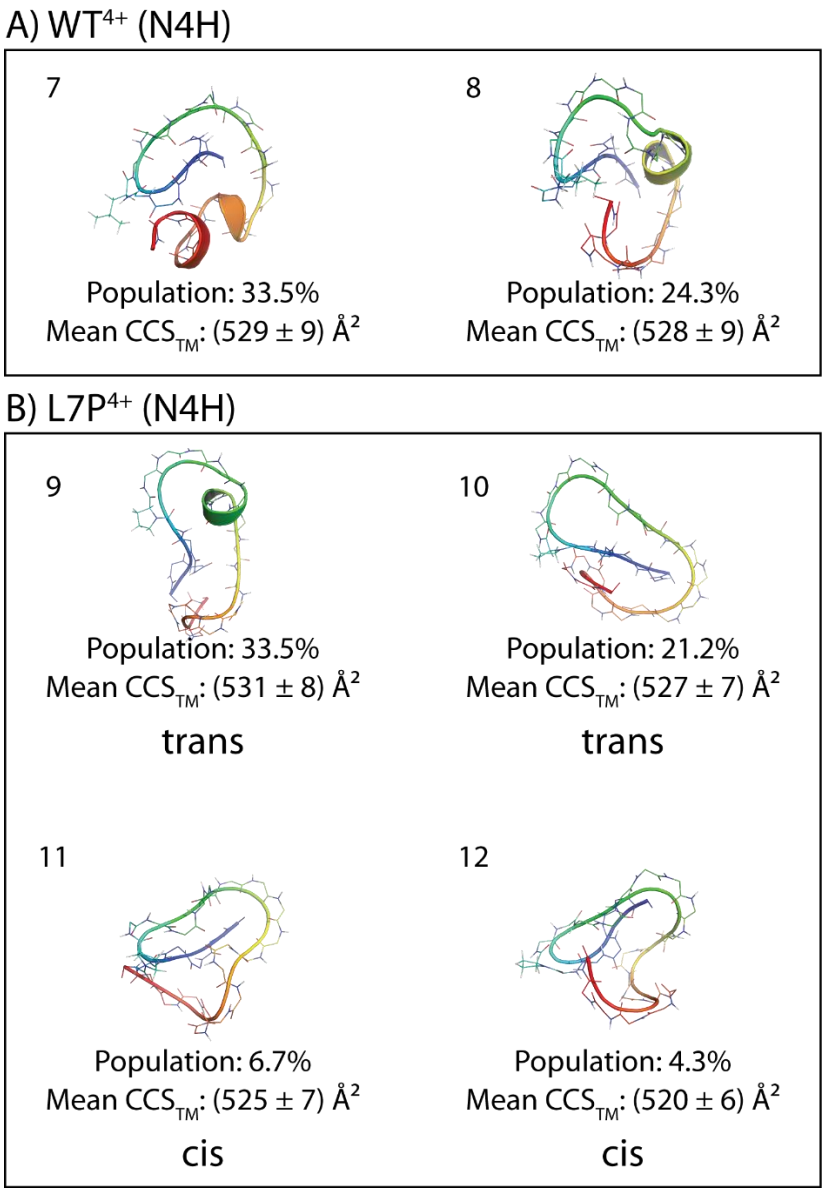


Figure S4. Candidate structures for N4H 4+ ions. Representative centroid structures from the highest population conformation clusters of **A)** WT^{4+} and **B)** $L7P^{4+}$ N4H ions. Atoms from the entire backbone and seventh residue's sidechain are displayed. Structures are color-coded to denote N-terminal (blue) and C-terminal (red) regions. Relative cluster population and mean theoretical CCS of the cluster are listed below the centroid structure. The total populations of the WT^{4+} and $L7P^{4+}$ N4H CCS-filtered ensembles were 2513 and 4565, respectively.

Chapter 3

Improved Isobaric Tandem Mass Tag Quantification by Ion Mobility Mass Spectrometry

Adapted from: Robert M. Sturm,* Christopher B. Lietz,* Lingjun Li. Improved isobaric tandem mass tag quantification by Ion Mobility Mass Spectrometry. *Rapid Comm. Mass Spectrom.* **2014** *28*, 1051-1060.

* Indicates equal contribution

Abstract

Isobaric tandem mass tags are an attractive alternative to mass difference tags and label-free approaches for quantitative proteomics due to the high degree of multiplexing that can be performed with their implementation. A drawback of tandem mass tags are that the co-isolation and co-fragmentation of labeled peptide precursors can result in chimeric tandem mass (MS/MS) spectra that can underestimate the fold-change expression of each peptide. Ion mobility (IM) separations coupled to quadrupole time-of-flight (Q-TOF) instruments have the potential to mitigate MS/MS spectra chimeracy since IM-MS has the ability to separate ions based on charge, m/z, and collision cross section (CCS).

Two complex protein mixtures, labeled with DiLeu isobaric tandem mass tags in opposite ratios, were mixed together and analyzed by data-dependent LC/IM-MS/MS. The accuracy of reporters from interfering pairs was compared with and without IM separation. IM separation was able to mitigate isobaric interference from differentially charged interfering ion pairs, as well as pairs of the same charge. Of the eight example precursors shown, only one had reporters that remained compressed below the significance threshold after IM separation.

The results of this investigation demonstrate proof-of-principle that IM separation of tagged precursors prior to MS/MS fragmentation can help mitigate quantitative inaccuracies caused by isobaric interference. Future improvements of the method would include software for automated correction and use of higher resolution IM instrumentations.

Introduction

Mass spectrometry (MS)-based proteomics has matured from a technique that could only identify single purified proteins to a technique that can now identify and quantify thousands of proteins at once.[1-4] Many different quantitation strategies are now commonly used in proteomics experiments. There are label-free techniques that measure protein abundance based on peptide ion intensities or spectral counts.[5-7] Mass-difference labeling approaches introduce light and heavy isotopic forms of an isotopic label into peptides to allow binary or tertiary comparisons to be made within the same experiment during MS analysis.[8-12] Isobaric labeling methods allow for the greatest amount of experiment multiplexing enabling quantification measurements of four samples,[13-14] six samples,[15-17] eight samples,[18] sixteen samples,[19-20] and even 54 samples[21] at the same time.

Each of these quantitative approaches has advantages and disadvantages. Label-free methods have the advantage of measuring peptides in their native state, but lack the high throughput of mass-difference and isobaric tagging methods. Mass-difference tags allow for binary or tertiary comparisons, but introduce increased complexity in MS acquisitions that can decrease the confidence and accuracy of quantitation and also limit the sampling depth of proteome. Isobaric tags have the advantage of multiplexing multiple samples together in one run without dramatically affecting MS complexity since quantitation is performed at the tandem mass spectrometry (MS/MS) level. The drawback of this method is that concurrent isolation of multiple precursor ions in the MS scan, termed isobaric interference, can create a chimeric MS/MS spectrum leading to an

underestimation of protein/peptide fold-changes and compression of the quantitative data towards unity.[22-25]

Isobaric interference is one of the most difficult problems facing isobaric tandem mass tag quantification.[24] Simple and effective methods to deal with it have been proposed, such as offline pre-fractionation.[26] However, in exceedingly complex samples, such as those composed of multiple proteomes, pre-fractionation has been shown to make no significant improvement to average ratio accuracy.[27] Savitski et al. have recently characterized two solutions, one involving smaller precursor isolation windows and optimal fragmentation times.[28] Smaller isolation windows can provide a modest improvement to reporter ratios, but may also cause an identification reduction as large as 48.5%.[27] The second method estimates the amount of protein ratio compression using the amount of interference in the precursor scan.[29] While this is a sophisticated approach to the problem, its inherent limitation is that it does not actually remove interference from a sample. Interference is accounted for and its impact is estimated.

As a comparison to our approach, we will focus on two acquisition-method-based strategies that have been proposed in the literature. The first solution, named 'QuantMode' by the authors, involves using gas-phase proton-transfer reactions to reduce the charge state of the target species, separating it from the interfering ion, and performing higher-energy collision-induced dissociation (HCD) on the charge-reduced target ion to improve quantification accuracy.[27] The second solution, commonly called the MS^3 method, improves quantification accuracy by performing an additional round of tandem MS on a fragment ion in the *m/z* range 110–160% relative to the precursor ion *m/z* and then measuring the isobaric tag intensity.[30] Both methods have documented

success in the literature but like all methods possess minor shortcomings. QuantMode works well for different charge precursor ions but can only resolve interference of precursors with the same charge if the charge-reduced *m*/*z* falls outside the precursor isolation window, whereas the triple-stage MS performed by MS^3 may produce reporters with greatly reduced intensity leading to decreased quantitative sensitivity.

QuantMode and MS^3 rely heavily on ion trap and orbitrap mass spectrometers. This leaves an unfilled niche for quadrupole-time-of-flight (Q-TOF) users wanting to improve isobaric quantitation accuracy. This work investigates the use of traveling wave ion mobility (TWIM)-MS on a Q-TOF mass spectrometer (SYNAPT G2) to improve peptide/protein quantification. Using ion mobility mass spectrometry (IM-MS), molecular ions can be separated by *m/z*, charge state, and collisional cross-section (shape and size).[31-34] Scheme 1a illustrates the consequences analyzing two hypothetical co-eluting peptides with similar *m*/*z*. Both Peptide A ($z = +2$) and Peptide B ($z = +3$) were labeled with *N,N*-dimethyl leucine (DiLeu) isobaric tags[14] and yield four MS/MS reporters ranging from 115 to 118 Da. Peptide A should display an expression ratio of 5:1:5:1, and Peptide B should display a ratio of 1:5:1:5. However, because their *m/z* values are within a 3 Th isolation window, they are co-isolated and co-fragmented. The resulting chimeric reporter spectrum is the sum of reporter intensities from both precursors. Scheme 1b illustrates how IM separation prior to MS/MS can mitigate the interference. Peptides A and B can be separated in the IM drift cell and therefore enter the fragmentation cell separately. Each precursor will be individually fragmented and mass analyzed upon exiting the drift cell, thus mitigating the chimeracy observed in Scheme 1a.

The parallel fragmentation of mobility-separated precursors, referred to as time-aligned parallel (TAP) fragmentation, is a vital component to our method.[35-37] It has proven to be a powerful method in proteomics,[38] and has been applied to several large-scale investigations.[39-41] IM-MS has also been used for quantitative analysis involving chemical tags, including isotopic labels[42-43] and multiplex tags meant to induce mobility differences.[44-46] In 2011, further application to label-based quantification was shown when Waters published a technology briefing detailing the ability of a LC/IM-MS method to separate tandem mass tag (TMT)-labeled bovine serum albumin (BSA) peptides from non-tagged *Escherichia coli* peptides.[47] The brief report showed evidence that acquisitions not utilizing IM-MS resulted in chimeric MS/MS spectra whereas utilizing IM-MS cleaned up the MS/MS spectra to show purer MS/MS sequence fragments. An interesting point not made in this report was whether IM-MS has the ability to clean up chimeric MS/MS spectra of two peptides with different isobaric tag ratios. Here, data-dependent analysis (DDA) with and without the use of precursor IM separation is investigated for its ability to correct quantitative inaccuracies caused by isobaric interference.

Experimental

Chemicals and Materials

Anhydrous acetonitrile, triethylammonium bicarbonate (TEAB), 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride (DMTMM), Trizma hydrochloride (Tris HCl, ≥99.0%), iodoacetamide (IAA), acetone (≥99.5%), anhydrous *N,N*-dimethylformamide (DMF, ≥99.8%), Triton X-100 (laboratory grade), bovine serum albumin (BSA, ≥96%), bovine apo-transferrin (≥97%), bovine beta-lactoglobulin (≥90%),

horse myoglobin (≥90%), and bovine cytochrome C (≥95%) were purchased from Sigma-Aldrich (St. Louis, MO, USA). Urea, formic acid (Optima LC/MS grade), acetonitrile, and water for LC/MS solvents (Optima LC/MS grade) were purchased from Fisher Scientific (Fair Lawn, NJ, USA). Deionized water (18.2 MΩ·cm) was prepared with a Milli-Q system (Millipore, Billerica, MA, USA). DL-Dithiothreitol (DTT), trypsin gold (mass spectrometry grade), and rLys-C (mass spectrometry grade) were purchased from Promega (Madison, WI, USA). N-Methylmorpholine was purchased from TCI America (Portland, OR, USA). Sodium dodecyl sulfate (SDS, ≥99.8%) was purchased from US Biological (Marblehead, MA, USA). The BCA protein assay kit and 1X protease and phosphatase inhibitor cocktail were purchased from Thermo Pierce (Rockford, IL, USA). Yeast peptone dextrose (YPD) media was prepared in 1 L deionized water using 10 g yeast extract, 20 g peptone from Becton, Dickinson and Company (Sparks, MD, USA), and 20 g D-(+)-glucose (≥99.5%, Sigma-Aldrich). For solid-phase extraction (SPE), Oasis HLB 1 cc (10 mg) extraction cartridge were purchased from Waters (Milford, MA, USA) and strong-cation-exchange (SCX) spintips were purchased from Protea Biosciences (Morgantown, WV, USA).

Yeast Lysate Protein Preparation

The yeast samples were prepared using the protocol of Miller and Cross with modification.[48] Yeast strain s288c was inoculated with yeast peptone dextrose (YPD) media and shaken for 72 h. After shaking, an aliquot of 100 optical density units (ODU) of the yeast culture was transferred to a test tube containing 2% glucose. The yeast sample was concentrated using a Beckman-Coulter J6B centrifuge (Indianapolis, IN, USA) at 2500 *g* for 3 min. The media was decanted and the yeast pellet was flash frozen in liquid nitrogen and stored at −80 °C until further use. The yeast pellets were allowed to

thaw on ice for 10 min before being resuspended in 1 mL of lysis buffer (50 mM Tris HCl, 0.1% Triton X-100, 0.5% SDS, pH 8.0) and 1X protease and phosphatase inhibitor cocktail. The resuspended yeast was transferred to a new microcentrifuge tube and 200 µL of glass beads were added and vortexed for 2.5 min at 4 °C. The combined sample was inverted, and the bottom of the microcentrifuge tube was pierced with a hot 23 gauge needle. The tube was placed atop a 5 mL culture tube and the sample was centrifuged at 2500 *g* for 3 min at 4 °C. The collected protein lysate from the pierced microcentrifuge tube was centrifuged at 3000 *g* for an additional 5 min and the supernatant was stored at −80 °C until further use.

Because detergents can be detrimental to MS analysis, yeast proteins (~2 mg per sample) were purified from cell lysis buffer by acetone precipitation. Four parts −20 °C acetone were added to one part yeast protein, the microcentrifuge tube was inverted three times, and protein precipitation was allowed to proceed for 3 h at −20 °C. The samples were then centrifuged at 13000 *g* for 5 min and the supernatant was removed leaving a precipitated protein pellet. The pellet was reconstituted in 100 µL 8 M urea, 50 mM Tris HCl (pH 8) and a BCA protein assay was run. The reconstituted yeast protein was divided into 400 µg aliquots for digestion.

Mammalian Protein Mixture Preparation

The five-protein equimolar mix was prepared by weighing out 1 mg of BSA, apo-transferrin, beta-lactoglobulin and 0.5 mg of myoglobin, and cytochrome C and resuspending each in 84.5 µL, 75.5 µL, 294 µL, 171.25 µL, and 250 µL 8 M urea, 50 mM Tris HCL (pH 8), respectively. Then 8 µL of each reconstituted protein were combined to give approximately 267 µg that was used for digestion.

Protein Digestion

The yeast proteins and the five mammalian protein mix were digested separately using identical protocols. Cysteine residues were reduced by addition of 4 µL 50 mM DTT and incubated for 1 h at room temperature, then alkylated by addition of 5 µL 85 mM IAA for 20 min in the dark at room temperature. An additional 2 µL aliquot of 50 mM DTT was added to each sample to quench the alkylation reaction. Proteins were digested with rLys-C (1:100 enzyme/protein) for 2 h at 37 °C. Prior to trypsin digestion, each digest was diluted with 50 mM Tris HCl (pH 8) to reduce the urea concentration to ≤1 M. Trypsin digestion (1:50 enzyme/protein) was carried out in a 37 °C water bath for 16 h. Each digest was acidified to pH 2 by addition of 10% formic $acid_{(aq)}$. Each sample was desalted using Oasis HLB, 1 cc (10 mg) SPE cartridges following the manufacturer's protocol. The eluted peptides were dried down by vacuum centrifugation using a speedvac (Thermo Scientific, Waltham, MA, USA). Peptides were resuspended in 0.5 M TEAB to a concentration of 5 $\mu g \cdot \mu L^{-1}$.

DiLeu Labeling

DiLeu isobaric tags are isotopically encoded *N,N*-dimethyl leucines. They were activated by triazine esterification and then attached to primary amines on peptide N-termini and lysine residues. For more details regarding the DiLeu reagent, see Xiang *et al*.[14] Four 50 µg aliquots of yeast protein digest and four 50 µg aliquots of the five-protein mix digest were prepared. Each aliquot was labeled with either the 115, 116, 117, or 118 DiLeu label. For each labeling reaction, 1 mg of dried DiLeu label was activated with 50 µL of activation solution (1.86 mg DMTMM, 0.74 µL NMM, 51.5 µL dried DMF) for 1 h at room temperature, with shaking. After 1 h, 20 µL anhydrous acetonitrile and 25 µL of the

appropriate activated DiLeu label were added to each 50 µg protein digest aliquot. Labeling was carried out for 2 h at room temperature, with shaking. The labeling reaction was quenched by addition of 100 µL deionized water, and the samples were shaken at room temperature for an additional 30 min. The labeled peptide solutions were concentrated to dryness using a speedvac.

Each dried, labeled peptide sample was resuspended in 100 µL of SCX resuspension buffer (Protea Biosciences). Alternatively, labeled (115, 116, 117, or 118) yeast protein digest samples were aliquoted and combined to give mass ratios of 1:5:1:5 for each respective reporter ion channel. Alternatively, labeled five-protein mix digest samples were aliquoted and combined to give mass ratios of 5:1:5:1. Residual labeling chemicals were then removed from each combined sample using SCX spintips following the manufacturer's protocol. The eluate was dried down by speedvac, resuspended in 500 µL in 0.1% formic $acid_{(aq)}$ (v/v), and desalted using Oasis HLB, 1 cc (10 mg) SPE cartridges following the manufacturer's protocol. The eluate was dried down by speedvac and resuspended in 3% acetonitrile/0.1% formic $acid_{(aq)}$ (v/v) so that the concentration of each sample was approximately 0.5 $\mu g \cdot \mu L^{-1}$ assuming minimal loss from sample preparation steps. Finally, a mixture of the labeled yeast protein digest and five-protein mix digest was prepared by adding one part five-protein mix digest to five parts yeast protein digest.

Mass Spectrometry

A Waters SYNAPT G2 Q-TOF mass spectrometer coupled to a Waters nanoAcquity ultra-performance liquid chromatography (UPLC) system was used for nano-liquid chromatography electrospray ionization tandem mass spectrometry

(nanoLC/ESI-MS/MS) analysis. Mobile phase A was water in 0.1% $FA_{(aq)}$ (v/v), and mobile phase B was ACN in 0.1% $FA_{(aq)}$ (v/v). For each analysis, approximately 800 ng of DiLeu-labeled protein digest was loaded onto a Symmetry C18 nanoAcquity trap column (180 µm × 20 mm, 5 µm) at a flow rate of 5 $\mu L \cdot min^{-1}$ of 97% mobile phase A for 3 min. Peptides were separated using a 1.7 µm BEH C18 column (75 µm × 100 mm) with a 60 min gradient. Mobile phase B was linearly ramped from 9 to 35% B over 60 min. The flow rate was 300 $nL \cdot min^{-1}$ and the column temperature was 35 °C. Electrospray emitter tips were prepared in house from 75 µm i.d., 360 µm o.d. capillary tubing (Polymicro Technologies, Phoenix, AZ, USA) using a Sutter P-2000 laser capillary puller (Novator, CA, USA).

Data were acquired in resolution mode using DDA and mobility DDA with a precursor isolation window of ~3 Th. The capillary voltage was set to 2.80 kV, sampling cone voltage to 30 V, extraction cone voltage to 4.0 V, and source temperature to 70 °C. A 1 Hz survey scan was followed by three MS/MS scans of the top three ions with charge states +2, +3, or +4. Dynamic exclusion was set to 60 s. Mobility DDA utilized high-purity N_2 as the drift gas in the TWIM cell. Pressures in the helium cell and TWIM cell (also called the IM-MS cell) were 1.46×10^3 mbar and 3.61 mbar, respectively. Trap DC bias was 48.0 V, IM-MS wave velocity was 1000 ms^{-1}, IMS wave height was 40 V, and IMS wave delay was 450 µs. A look-up table was created and optimized to use specified MS/MS collision energies at specific drift times. The look-up table applies a specific collisional energy to the transfer cell based on the IM-MS bin. Assuming a rough proportionality between *m*/*z* and drift time, collision energies can be tailored to specific *m*/*z*. Further details can be found in Supplementary Figure 1.

Data Analysis

Acquired data were analyzed using MassLynx 4.1 software and DriftScope (version 2.2) software. The LC/MS peak and MS/MS scans of select peptides were manually extracted as two-dimensional (2-D) *m*/*z* versus drift time data in DriftScope. Reporter ion ratios were then calculated manually in MassLynx. Reporter ion spectra were smoothed once across ± 2 channels. Data were then median-centered using a minimum peak width at half height of four channels. The centered spectra were based on peak areas. Centroid intensities of the reporter ions were placed into an Excel spreadsheet that calculated the reagent purity-corrected isobaric tag ratios using correction factors and solving four equations simultaneously as outlined previously.[49] A detailed description of the purity-correction calculations can be found in Supplementary Figure 2.

Mammalian peptides exhibiting isobaric interference were identified by precursor accurate mass matching (<20 ppm) to tryptic peptides predicted by Protein Prospector and manual MS/MS annotation of centroid fragment ions (<0.05 Th accuracy). The annotated spectra are displayed in Supplementary Figs. S3–S5 (see Supporting Information). No yeast peptide sequences were identified.

Experimental Plan

The experimental workflow and design is shown in Supplementary Fig. S6 (see Supporting Information). The concept is to create easily identifiable isobaric interferences and to maximize its occurrence. The approach is similar to that used by Wegner *et al*.[27] and Ting *et al*.[30] All labeled mammalian peptides should theoretically yield a 5:1:5:1 reporter ratio. By mixing the mammalian peptides with an entire proteome yielding the

opposite 1:5:1:5 ratio, the sum of reporter intensities from mutually interfering peptides will trend towards 1:1:1:1. If the reporter ratios extracted at the drift times of interfering parents are closer to 5:1:5:1 and 1:5:1:5 than reporters without IM extraction, it is evidence that IM-MS can partially mitigate isobaric interference. Yeast was selected as the interfering proteome because of its low protein sequence homology to mammals, thus removing some ambiguity when identifying mammalian peptides.

Results and Discussion

LC-IM-MS Results

Figure 1 shows the LC/IM-MS/MS analysis of two pairs of co-eluting peptides from mammalian and yeast proteins. Each pair contains peptide ions with different charge states (z = +2, +3). In Figure 1a, the triply charged tryptic BSA peptide HLVDEPQNLIK (*m/z* 532.673) and doubly charged yeast peptide (*m/z* 531.866) were co-isolated and co-fragmented to produce the 1.0:1.0:1.2:1.2 chimeric reporter ion spectrum (Figure 1b). Figure 1c shows the extracted ion drift time (t_D) distributions for each precursor. The BSA peptide and yeast peptide display apex drift times at 5.24 ms and 6.14 ms, respectively. Each trace was normalized to its own maximum. Figure 1d shows the reporter ratios extracted at the drift times of the BSA and yeast precursors. The ratio at 5.24 ms was 5.7:1.0:6.1:1.2, and the ratio at 6.14 ms was 1.1:3.4:1.0:4.1. Figures 1e – 1h show analogous spectra for the second peptide pair, triply charged transferrin tryptic peptide NYELLCGDNTRK (*m/z* 591.655) and a doubly charged yeast peptide (*m/z* 589.880). The initial reporter ratios were 1.0:1.7:1.2:1.2, indicating the possibility of interference. The apex drift time for the transferrin peptide was 6.07 ms and the apex drift time of the yeast

peptide was 6.49 ms. The extracted reporter ratios at 6.07 ms and 6.49 ms were 3.4:1.4:4.9:1.0 and 1.2:4.3:1.0:3.2, respectively.

Figure 2 shows results from analyzing interfering peptides of the same charge state (z = +4). Tryptic BSA peptide QEPERNECFLSHK (m/z 491.762) and a yeast peptide (m/z 492.257) yield a chimeric reporter ion spectrum with a 1.3:1.0:1.3:1.0 ratio. In Figure 2c, the apex drift time of the BSA peptide was 5.11 ms. The yeast peptide distribution also displays a peak at 5.11 ms, but the true apex was attributed to 5.38 ms. The reporter ratios extracted at 5.11 ms and 5.38 ms were 4.4:1.0:4.7:1.0 and 1.0:4.9:1.2:4.7, respectively.

The summaries of the isobaric interference examples, as well as additional considerations involving fold-change, are contained in Table 1. Reporter ion ratios 115/116 and 117/118 were considered independently. Each experimental ratio with IM (IM-MS/MS) and without IM (MS/MS) was compared to the theoretical ratio. The 'Log_2 Threshold' refers to whether or not an experimental ratio was greater than 1.5 on a Log_2 scale. A 'yes' or 'no' answer is proceeded by the actual ratio on a Log_2 scale. A 'yes' signifies that IM improved quantitative accuracy. An additional mammalian peptide in Table 1 at m/z 499.316 was clearly experiencing isobaric interference (Supplementary Figure 7). However, the peptide's mass did not match to any tryptic peptide, single, or double miscleavage for the five proteins within 20 ppm, and MS/MS fragments were not sufficient for full *de novo* sequencing. The extracted reporter ion ratios unambiguously identify its origin as the mammalian sample, and we hypothesize that the peptide was a fragment from a low-abundance protein in the BSA sample (~96% purity) or a peptide that was unintentionally modified during sample preparation.

Additional instances of isobaric interference were indicated by near-unity reporter ratios produced by other precursors. However, IM separation was unable to mitigate the interference (data not shown).

Co-isolation of interfering precursors with different charge states (z = +2, +3)

The spectra shown in Figure 1 exemplify the ability of IM to readily separate ions of different charge states. Both pairs of peptides display baseline separation of their extracted drift time distributions and were able to produce IM-MS/MS reporters that were more accurate than the chimeric MS/MS reporters. Even though the yeast peptide sequences were not identified, the pattern of their reporter ratios was strong evidence that they originated from the yeast sample. In Figure 1c, two additional small peaks in the BSA peptide's trace were present at the same drift times as the yeast peptide's peak. It is possible that they represent additional elongated structures of the BSA peptide. Reporters from the elongated BSA peptides could have contributed to the intensity from the yeast reporters and caused the underestimation present in the 6.14 ms reporters. However, similar reporter underestimation was present in both reporter spectra of Figure 1h, yet there was no evidence of precursor IM crossover in Figure 1g. Other possibilities for the reporter underestimation could be additional interfering precursors that cannot be resolved in IM, or incomplete labeling by the DiLeu tags.

Co-isolation of interfering precursors with the same charge states (z = +4)

The examples shown in Figure 1 demonstrate that IM can separate differently charged interfering precursors, but this is something that can also be readily accomplished by methods described previously by Wegner *et al*.[27] or Ting *et al*.[30] Even the mitigation of differently charged precursors with very similar *m/z*, such as the

unidentified peptide in Supplementary Figure 7 that differs from its interfering ion by only 40 mTh, would be separated by at least 3 Th after charge reduction. However, precursor IM separation can mitigate interferents of the same charge, which is still possible for the MS^3 method but only possible for QuantMode when the difference in *m/z* is sufficiently large.

The monoisotopic *m/z* of the quadruply charged BSA and yeast peptides shown in Figure 2 differ by approximately 0.5 Th. Even if these peptides were charge reduced to doubly charged precursors, the *m/z* difference would be well within a 3 Th isolation window. After IM separation of these precursors, the largest relative error for the BSA peptide (115:116) was reduced from 74% to 12%, and the largest yeast relative error (118:115) was reduced from 85% to 6%.

It is interesting to note that although Figure 2 arguably presents better reporter correction than Figure 1, the precursors in the former are less resolved by IM separation. The resolution is compounded by the two peaks displayed by the yeast peptide; however, we believe the yeast peak at 5.11 ms is a result of the BSA peptide's isotope at *m/z* 492.266. Decreased IM resolution would be expected for two interfering peaks of the same charge state compared to those with different charges. In a triangular waveform approximation of TW IM-MS, the transit time from the net drift velocity (*t*) scales proportionally with the cell length (*L*) and wave speed (*s*), and inversely with the square of mobility (*K*) and the net electric field (*E*):[50]

$$t = \frac{Ls}{(KE)^2} \quad (1)$$

The Mason-Schamp relationship says that *K* is directly proportional to ion charge (*z*) and inversely proportional to CCS (Ω):

$$K \propto \frac{z}{\Omega} \quad (2)$$

Therefore, we can say that t is directly proportional to Ω^2 and inversely proportional to z^2:

$$t \propto \left(\frac{\Omega}{z}\right)^2 \quad (3)$$

Suppose we call t_A the transit time of a doubly charged ion A and t_B the transit time of a triply charged ion B with the same CCS as ion A. Using Equations 1-3, and assuming all other pertinent variables to be held constant, it can be shown that t_B is approximately equal to 4/9 t_A. However, if the ions were the same charge and the CCS of ion A was 250 $Å^2$, ion B would need a CCS nearly 84 $Å^2$ smaller than ion A to exhibit the same degree of separation. Because of this, we expect that the IM-MS method would have a diminished performance for interferents of the same charge. Figure 2 is only one example of same charge interference and cannot be considered a statistically significant representation, but it does show that IM-MS can perform accurate corrections even when interfering peptides are of the same charge. It is important to note that TW have waveforms more similar to half-sinusoidal than triangular, thus the *t*/*K* relationship becomes more quadratic as *K* increases.[50]

Fold-change Masking

Quantitative proteomic experiments utilizing tandem mass tags are not often concerned with the absolute ratios displayed by reporters. An investigator might not be interested in whether a ratio is 1:5 or 1:8, but rather one might be interested in whether a protein is displaying overall differential regulation. In many instances, ratios are converted to a Log_2 scale to represent fold-change. In Table 1, the data was organized and framed relative to fold-change. Ratios were divided up into 115–116 and 117–118 pairs, and a

ratio of 1.5 on a Log_2 scale (~2.8) was set as a threshold for a peptide to be considered a candidate for up-regulation.

All labeled peptides were mixed to ratios of either 1:5:1:5 or 5:1:5:1 and therefore should surpass the Log_2 threshold. However, the fold-change for each peptide in the table was masked by interference and would have been incorrectly passed over for up-regulation candidacy. When the precursors are separated by IM, the ratios meet the Log_2 threshold in all examples except for the transferrin miscleavage NYELLCGDNTRK. This result demonstrates that IM-MS is able to correct isobaric interference to a degree that may have a significant impact on real proteomic investigations.

Conclusion

Isobaric interference is a very common occurrence in complex samples. In the extreme, Wenger *et al*. observed only 68% of intensity from the average reporter spectrum came from the intended precursor, and only 3% of MS/MS spectra were obtained from precursor isolations containing less than 1% interference.[27] The results of this investigation demonstrate proof-of-principle that IM separation of tagged precursors prior to MS/MS fragmentation can help mitigate quantitative inaccuracies caused by isobaric interference. IM-MS was able to resolve interference among differentially charged co-isolated peptides, and even among peptides of the same charge state. Contrary to the highlighted example, it is likely that the IM-MS method may have better overall performance with differentially charged interferants compared to pairs with the same charge. The instances of IM separation providing no reporter spectrum correction suggest that, just as is the case with QuantMode and MS^3, this method cannot universally solve all interferences.

We would like to emphasize that the IM-MS technique should not be seen as a 'competitor' to QuantMode or MS^3. Rather, IM-MS should be considered as a potential alternative for Q-TOF instrumentation. The limited examples we presented represent the types of interference that could be encountered, and we selected instances where interference was most obvious. To transform the IM-MS technique into a viable method, significant changes need to be made to the acquisition and analysis software. To the knowledge of the authors, it is not possible to automatically extract specific ion intensities from a single drift time during DDA in MassLynx, DriftScope, or the common search algorithm used for SYNAPT G2 mobility DDA, ProteinLynx Global Server. Consequently, all drift time extractions in this investigation had to be performed manually, making application to large datasets very difficult. In MassLynx, there is an acquisition option to only use MS/MS drift times that were present in the precursor scan. This option is not helpful in this application, because drift times from the interfering peptide would have also been present in the precursor scan. The authors propose that an option allowing only the use of the most abundant isolated peptide's drift time would remedy this situation. Lacking automated analysis to apply the method to large datasets, we cannot explicitly evaluate its utility for large-scale proteomics at this point.

Perhaps the most imminent need of the IM-MS method is the applicability to data-independent $HDMS^E$ analysis.[39, 51] The scan cycle of the mobility DDA commonly operates at 1 to 5 Hz, making it difficult to obtain a large number of peptide IDs and quantified proteins. Additionally, the millisecond time scale of IM intrinsically requires longer scan rates than analyses without IM. In its current state, precursor IM separation as part of mobility DDA may be appropriate for low-complexity mixtures or targeted

analyses. Since the performance of HDMS^E is largely independent of scan rate, the IM-MS method may be successful with large-scale investigations. The most significant challenge in coupling to HDMS^E would be that the lack of quadrupole isolation would mean all co-eluting labeled peptides would be interfering with each other. Future investigations will focus on the use of field-asymmetric ion mobility spectrometry (FAIMS) instruments. Although it is often difficult to predict how the separation of specific ions will compare between FAIMS and electrostatic or TW IM, FAIMS has demonstrated very high resolving powers.[52-55] The SYNAPT G2 is able to achieve drift time resolutions near 25, and based on isobaric interference predictions we made from HDMS^E analysis of yeast digest, this would be insufficient for 30.1% of interfering peptide pairs (Supplementary Table 1). Higher resolution than that of current TW IM instruments may be necessary to minimize apex drift time overlap of co-eluting peptides to make coupling a HDMS^E-like data-independent acquisition (DIA) method a possibility. Whether or not this can be achieved by FAIMS must be evaluated empirically.

Acknowledgements

The authors would like to thank Tyler Greer and Dustin Frost in the Li Research Group for synthesizing the DiLeu isobaric tags. This work is supported in part by a National Science Foundation Grant (CHE-0957784) and National Institutes of Health Grants (1R01DK071801 and 1R56DK071801). L. Li acknowledges an H.I. Romnes Faculty Research Fellowship. R.M.S. acknowledges the NIH-supported Clinical Neuroengineering Training Program Predoctoral Fellowship (T32 EB011434). C.B.L. acknowledges an NIH-supported Chemistry Biology Interface Training Program

Predoctoral Fellowship (Grant No. T32-GM008505) and an NSF Graduate Research Fellowship (DGE-1256259).

References

(1) Kline, K. G., and Sussman, M. R. *Annu. Rev. Biophys.* **2010** *39*, 291-308.

(2) Ong, S.-E., and Mann, M. *Nat. Chem. Biol.* **2005** *1*, 252-262.

(3) Schulze, W. X., and Usadel, B. *Annu. Rev. Plant Biol.* **2010** *61*, 491-516.

(4) de Godoy, L. M. F., Olsen, J. V., Cox, J., Nielsen, M. L., Hubner, N. C., Frohlich, F. *et al. Nature* **2008** *455*, 1251-1254.

(5) Lundgren, D. H., Hwang, S.-I., Wu, L., and Han, D. K. *Expert Rev. Proteomics* **2010** *7*, 39-53.

(6) Old, W. M., Meyer-Arendt, K., Aveline-Wolf, L., Pierce, K. G., Mendoza, A., Sevinsky, J. R. *et al. Mol. Cell. Proteomics* **2005** *4*, 1487-1502.

(7) Podwojski, K., Eisenacher, M., Kohl, M., Turewicz, M., Meyer, H. E., Rahnenführer, J. *et al. Expert Rev. Proteomics* **2010** *7*, 249-261.

(8) Boersema, P. J., Raijmakers, R., Lemeer, S., Mohammed, S., and Heck, A. J. R. *Nat. Protoc.* **2009** *4*, 484-494.

(9) Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. *Nat. Biotechnol.* **1999** *17*, 994-999.

(10) Hsu, J.-L., Huang, S.-Y., Chow, N.-H., and Chen, S.-H. *Anal. Chem.* **2003** *75*, 6843-6852.

(11) Li, J., Steen, H., and Gygi, S. P. *Mol. Cell. Proteomics* **2003** *2*, 1198-1204.

(12) Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A. *et al. Mol. Cell. Proteomics* **2002** *1*, 376-386.

(13) Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S. *et al. Mol. Cell. Proteomics* **2004** *3*, 1154-1169.

(14) Xiang, F., Ye, H., Chen, R., Fu, Q., and Li, L. *Anal. Chem.* **2010** *82*, 2817-2825.

(15) Dayon, L., Hainard, A., Licker, V., Turck, N., Kuhn, K., Hochstrasser, D. F. *et al. Anal. Chem.* **2008** *80*, 2921-2931.

(16) Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G. *et al. Anal. Chem.* **2003** *75*, 1895-1904.

(17) Zhang, J., Wang, Y., and Li, S. *Anal. Chem.* **2010** *82*, 7588-7595.

(18) Choe, L., D'Ascenzo, M., Relkin, N. R., Pappin, D., Ross, P., Williamson, B. *et al. PROTEOMICS* **2007** *7*, 3651-3660.

(19) Dephoure, N., and Gygi, S. P. *Sci. Signaling* **2012** *5*, rs2-rs2.

(20) Huttlin, E. J., Jedrychowski, M., Kuhn, K., Dai, C., Ting, L., McAlister, G. C. *et al.* Enhanced isobaric labeling enables 18-plexed quantitative exploration of the hsf1-dependent cellular response to multiple proteotoxic stresses at a proteomic scale. *Proc. 60th ASMS. Conf. Mass Specrometry and Allied Topics*. **2012** Vancouver, Canada.

(21) Everley, R. A., Kunz, R. C., Mcallister, F. E., Kuhn, K., Rogers, J. C., Pike, I. *et al.* Increasing throughput in kinase inhibition assays: 54-plex quantitation in a single MS run. *Proc. 60th ASMS. Conf. Mass Specrometry and Allied Topics*. **2012** Vancouver, Canada.ThP05 112

(22) Bantscheff, M., Boesche, M., Eberhard, D., Matthieson, T., Sweetman, G., and Kuster, B. *Mol. Cell. Proteomics* **2008** *7*, 1702-1713.

(23) Karp, N. A., Huber, W., Sadowski, P. G., Charles, P. D., Hester, S. V., and Lilley, K. S. *Mol. Cell. Proteomics* **2010** *9*, 1885-1897.

(24) Ow, S. Y., Salim, M., Noirel, J., Evans, C., Rehman, I., and Wright, P. C. *J. Proteome Res.* **2009** *8*, 5347-5355.

(25) Shirran, S. L., and Botting, C. H. *J. Proteomics* **2010** *73*, 1391-1403.

(26) Ow, S. Y., Salim, M., Noirel, J., Evans, C., and Wright, P. C. *PROTEOMICS* **2011** *11*, 2341-2346.

(27) Wenger, C. D., Lee, M. V., Hebert, A. S., McAlister, G. C., Phanstiel, D. H., Westphall, M. S. *et al. Nat. Methods* **2011** *8*, 933-935.

(28) Savitski, M. M., Sweetman, G., Askenazi, M., Marto, J. A., Lang, M., Zinn, N. *et al. Anal. Chem.* **2011** *83*, 8959-8967.

(29) Savitski, M. M., Mathieson, T., Zinn, N., Sweetman, G., Doce, C., Becher, I. *et al. J. Proteome Res.* **2013** *12*, 3586-3598.

(30) Ting, L., Rad, R., Gygi, S. P., and Haas, W. *Nat. Methods* **2011** *8*, 937-940.

(31) McLean, J. A., Ruotolo, B. T., Gillig, K. J., and Russell, D. H. *Int. J. Mass Spectrom.* **2005** *240*, 301-315.

(32) Pringle, S. D., Giles, K., Wildgoose, J. L., Williams, J. P., Slade, S. E., Thalassinos, K. *et al. Int. J. Mass Spectrom.* **2007** *261*, 1-12.

(33) Zhong, Y., Hyung, S.-J., and Ruotolo, B. T. *Analyst* **2011** *136*, 3534-3541.

(34) Giles, K., Williams, J. P., and Campuzano, I. *Rapid Commun. Mass Spectrom.* **2011** *25*, 1559-1566.

(35) Hoaglund-Hyzer, C. S., and Clemmer, D. E. *Anal. Chem.* **2001** *73*, 177-184.

(36) Valentine, S. J., Kulchania, M., Barnes, C. A. S., and Clemmer, D. E. *Int. J. Mass Spectrom.* **2001** *212*, 97-109.

(37) Hoaglund-Hyzer, C. S., Li, J., and Clemmer, D. E. *Anal. Chem.* **2000** *72*, 2737-2740.

(38) Lee, Y. J., Hoaglund-Hyzera, C. S., Srebalus Barnes, C. A., Hilderbrand, A. E., Valentine, S. J., and Clemmer, D. E. *J. Chromatogr. B* **2002** *782*, 343-351.

(39) Myung, S., Lee, Y. J., Moon, M. H., Taraszka, J., Sowell, R., Koeniger, S. *et al. Anal. Chem.* **2003** *75*, 5137-5145.

(40) Taraszka, J. A., Gao, X., Valentine, S. J., Sowell, R. A., Koeniger, S. L., Miller, D. F. *et al. J. Proteome Res.* **2005** *4*, 1238-1247.

(41) Valentine, S. J., Plasencia, M. D., Liu, X., Krishnan, M., Naylor, S., Udseth, H. R. *et al. J. Proteome Res.* **2006** *5*, 2977-2984.

(42) Kindy, J. M., Taraszka, J. A., Regnier, F. E., and Clemmer, D. E. *Anal. Chem.* **2002** *74*, 950-958.

(43) Xun, Z., Kaufman, T. C., and Clemmer, D. E. *J. Proteome Res.* **2009** *8*, 4500-4510.

(44) Bohrer, B., and Clemmer, D. *J. Am. Soc. Mass Spectrom.* **2011** *22*, 1602-1609.

(45) Hilderbrand, A. E., Myung, S., and Clemmer, D. E. *Anal. Chem.* **2006** *78*, 6792-6800.

(46) Kerr, T. J., Gant-Branum, R. L., and McLean, J. A. *Int. J. Mass Spectrom.* **2011** *307*, 28-32.

(47) Hughes, C., Vissers, H., and Langridge, J. Resolving chimeric spectra utilizing a data independent mobility acquisition strategy. **2011**.http://www.google.com/url?sa=t&rct=j&q=chimeric%20spectra;&source=web&cd=2&ved=0CCsQFjAB&url=http%3A%2F%2Fwww.18show.cn%2Fbbs%2Fattachment.aspx%3Fattachmentid%3D3574&ei=BSNDT4uZGsbyggeN9t2YCA&usg=AFQjCNGgjQ5pYy2JCD68igaHOrvyEWMpaw&cad=rja1

(48) Miller, M. E., and Cross, F. R. *Mol. Cell. Biol.* **2000** *20*, 542-555.

(49) Shadforth, I. P., Dunkley, T. P. J., Lilley, K. S., and Bessant, C. *Bmc Genomics* **2005** *6*.

(50) Shvartsburg, A. A., and Smith, R. D. *Anal. Chem.* **2008** *80*, 9689-9699.

(51) Li, G. Z., Vissers, J. P., Silva, J. C., Golick, D., Gorenstein, M. V., and Geromanos, S. J. *PROTEOMICS* **2009** *9*, 1696-1719.

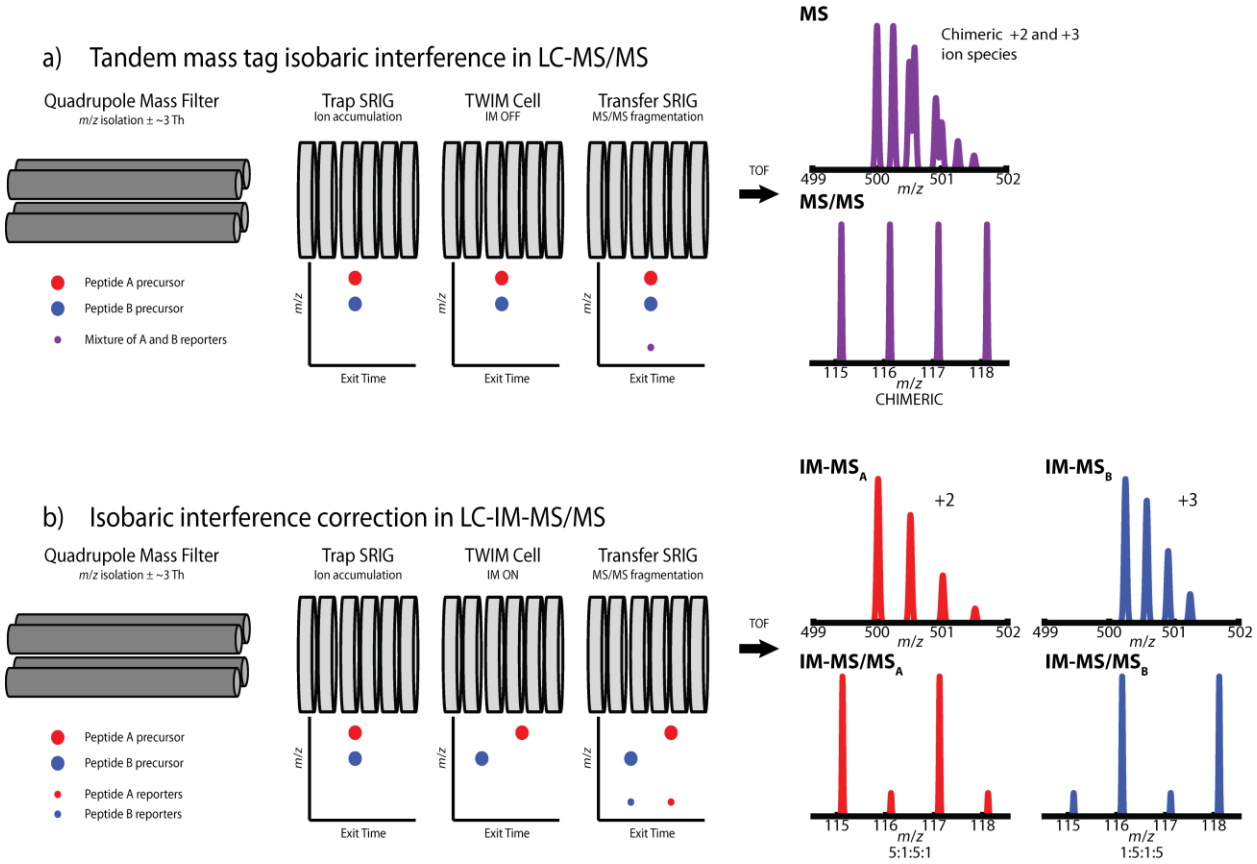
(52) Shvartsburg, A. A., Danielson, W. F., and Smith, R. D. *Anal. Chem.* **2010** *82*, 2456-2462.

(53) Shvartsburg, A. A., Prior, D. C., Tang, K., and Smith, R. D. *Anal. Chem.* **2010** *82*, 7649-7655.

(54) Shvartsburg, A. A., and Smith, R. D. *Anal. Chem.* **2011** *83*, 23-29.

(55) Shvartsburg, A., Seim, T., Danielson, W., Norheim, R., Moore, R., Anderson, G. *et al. J. Am. Soc. Mass Spectrom.* **2013** *24*, 109-114.

Scheme 1. To illustrate isobaric interference, two co-eluting DiLeu-labeled isobaric peptides, Peptide A ($z = +2$) and Peptide B ($z = +3$), are shown being analyzed on a SYNAPT G2 mass spectrometer. Peptide A has an expression ratio of 5:1:5:1 and Peptide B has an expression ratio of 1:5:1:5. (a) If IM separation is disabled, the two peptides will be co-fragmented resulting in an inaccurate reporter spectrum that is not representative of either peptide's true expression ratio. (b) On the other hand, if IM separation is enabled, the precursors will enter the Transfer stack-ring ion guide (SRIG) separately, allowing each peptide to be fragmented and mass analyzed in the TOF individually producing accurate reporter ion expression ratios for each peptide.



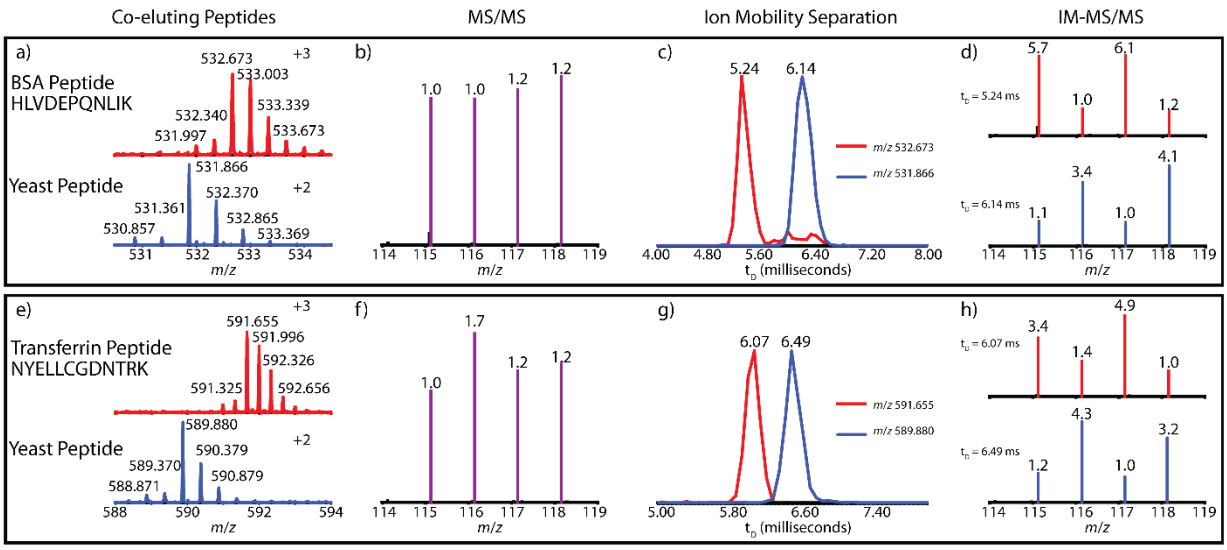


Figure 1. Example of IM separation improving the quantitative analysis of (a–d) BSA tryptic peptide HLVDEPQNLIK and (e–h) transferrin tryptic peptide NYELLCGDNTRK. In each case, MS/MS without IM separation produced chimeric reporter spectra approaching unity (b, f). After IM separation (c, g), reporter intensities were extracted at drift times of the precursors to produced corrected reporter spectra (d, h). The theoretical ratios for mammalian and yeast peptides were 5:1:5:1 and 1:5:1:5, respectively. IM traces were normalized to their own maximums, and all displayed reporter ratios were obtained from a purity correction algorithm applied to the centroid raw data.

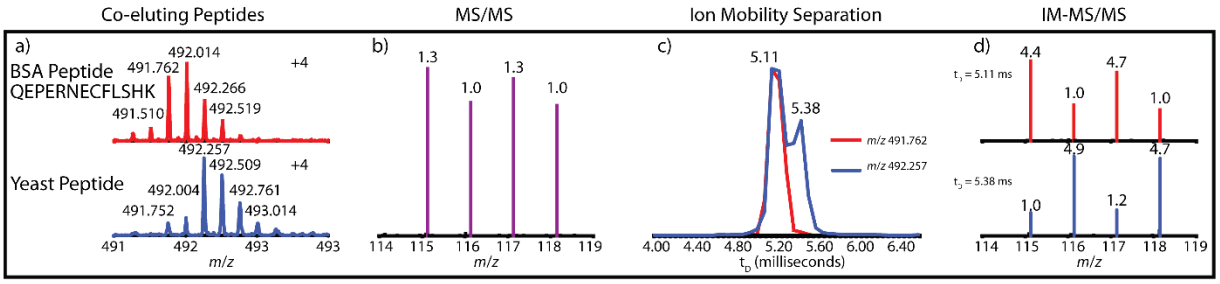


Figure 2. IM was used to separate interfering precursors of the same charge (z = +4), tryptic BSA peptide QEPERNECFLSHK and a yeast peptide. MS/MS without IM produced a chimeric reporter spectra approaching unity (b). After IM separation (c), reporter intensities were extracted at drift times of the precursors to produce improved quantitative reporter ion spectra (d). The theoretical ratios for mammalian and yeast peptides were 5:1:5:1 and 1:5:1:5, respectively. IM traces were normalized to their own maximums, and all displayed reporter ratios were obtained from a purity correction algorithm applied to the centroid raw data.

Table 1. Summary of examples of isobaric interference mitigation by IM

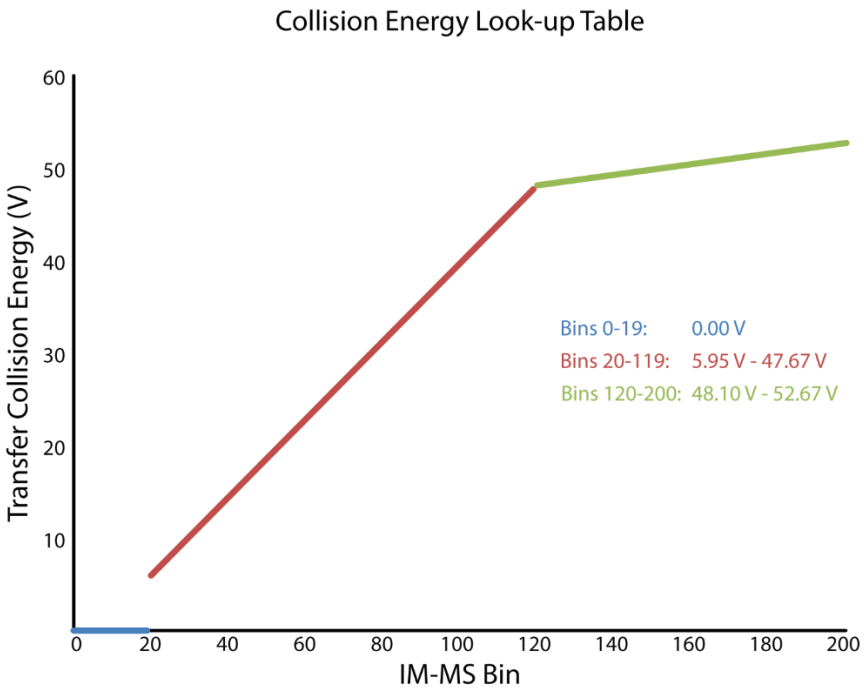
Mammalian Peptides

Target ion (*m/z*, *z*)	Interferant ion (*m/z*, *z*)	115/116 Ratio			Log_2 Threshold Met?		117/118 Ratio			Log_2 Threshold Met?	
		Theoretical	MS/MS	IM-MS/MS	MS/MS	IM-MS/MS	Theoretical	MS/MS	IM-MS/MS	MS/MS	IM-MS/MS
532.673, 3	531.866, 2	5.0	1.0	5.7	no (0.0)	yes (2.5)	5.0	0.9	5.1	no (-0.2)	yes (2.4)
591.655, 3	589.880, 2	5.0	0.6	2.5	no (-0.7)	no (1.3)	5.0	1.0	4.9	no (0.0)	yes (2.3)
491.762, 4	492.257, 4	5.0	1.3	4.4	no (0.4)	yes (2.1)	5.0	1.3	5.7	no (0.4)	yes (2.5)
499.316, 3	499.356, 2	5.0	0.9	3.1	no (-0.2)	yes (1.6)	5.0	0.8	4.3	no (-0.3)	yes (2.1)

Yeast Peptides

Target ion (*m/z*, *z*)	Interferant ion (*m/z*, *z*)	116/115 Ratio			Log_2 Threshold Met?		118/117 Ratio			Log_2 Threshold Met?	
		Theoretical	MS/MS	IM-MS/MS	MS/MS	IM-MS/MS	Theoretical	MS/MS	IM-MS/MS	MS/MS	IM-MS/MS
531.866, 2	532.673, 3	5.0	1.0	3.3	no (0.0)	yes (1.7)	5.0	1.0	4.2	no (0.0)	yes (2.1)
589.880, 2	591.655, 3	5.0	1.7	3.7	no (0.8)	yes (1.9)	5.0	1.0	3.2	no (0.0)	yes (1.7)
492.257, 4	491.762, 4	5.0	0.7	4.9	no (-0.5)	yes (2.3)	5.0	0.8	4.0	no (-0.3)	yes (2.0)
499.356, 2	499.316, 3	5.0	1.1	3.5	no (0.1)	yes (1.8)	5.0	1.3	5.0	no (0.4)	yes (2.3)

Supplemental Information



Supplemental Figure 1. Collision energy (CE) in the transfer stacked-ring ion guide (SRIG) was ramped using a user created look-up table (LUT), since mobility DDA does not allow the option of a direct *m*/*z*-dependent energy ramp. Initially, an unlabeled tryptic yeast digest was analyzed using data-independent HDMS^E (which does allow the option of a direct *m*/*z*-dependent energy ramp) and searched in ProteinLynx Global Server 2.0. From the "Final Peptide" output CSV, the peptide list was filtered for doubly charged "pass 1" peptides. The filtered data was copied to an excel spreadsheet where a new column called "Transfer CE" was made and contained the formula: =(0.034*precursor.mz)+8. Next, a scatter-plot was constructed with "Transfer CE" on the y-axis and "mobility" on the x-axis. Then, a linear regression was used to obtain the slope of the best-fit line for the data. This slope was used to determine the initial and final CE values for the ramp from Bins 20 – 119. The ramp for Bins 120-200 was determined by setting the final CE at +5 V from Bin 119. Bins 0 – 20 were kept at 0 V. This method was adapted from a protocol by Waters Corporation.

DiLeu 4-Plex Correction Factors

Channels	-2	-1	+1	+2
115	0.03	0.41	8.28[a]	0.38
116	0.11	7.03[b]	7.80[a]	0.33
117	0.15	7.44[b,c]	7.73[a]	0.30
118	0.04	0.57	7.38[a]	0.63

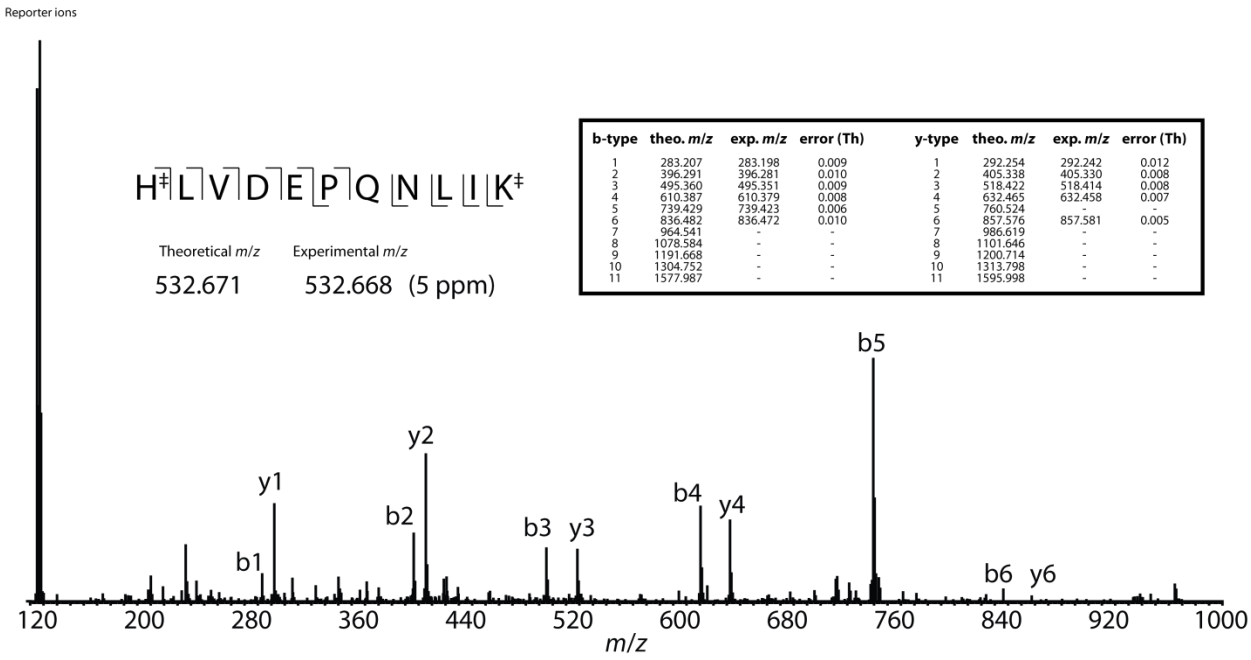
[a] Due to natural ^{13}C

[b] Due to purity of sodium cyanoborodeuteride (Sigma: 190020)

[c] Due to ^{15}N purity

$$S_{115} = 0.9090I_{115} + 0.0703I_{116} + 0.0015I_{117}$$
$$S_{116} = 0.0828I_{115} + 0.8473I_{116} + 0.0744I_{117} + 0.0004I_{118}$$
$$S_{117} = 0.0038I_{115} + 0.0780I_{116} + 0.8438I_{117} + 0.0057I_{118}$$
$$S_{118} = 0.0033I_{116} + 0.0773I_{117} + 0.9138I_{118}$$

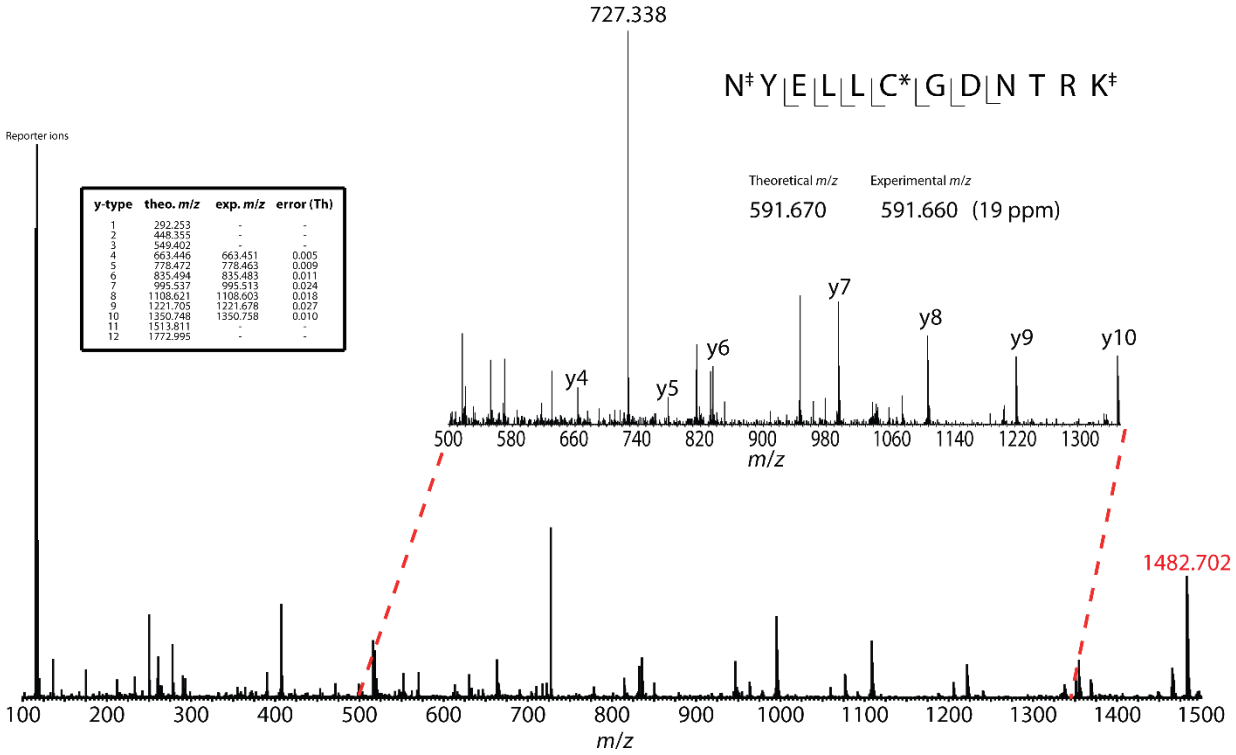
Supplemental Figure 2. Due to the impurities of synthesis reagents, natural ^{13}C isotopes, and the approximate 1 Th spacing between DiLeu reporter ions, raw reporter intensities from the MS/MS spectra must be corrected for purity. The table above shows the correction factors for each DiLeu channel at MS/MS signals ± 2 Th from the reporter ion *m*/*z*. Each factor is the percentage of the total population tagged in the row-designated DiLeu channel that will appear as intensity at an *m*/*z* spaced a number Th designated by the column. Using these factors, a system of four equations was created. The variable S_n represents the MS/MS signal intensity at the *m*/*z* of DiLeu channel n. The variable I_n represents the total number of ions from DiLeu channel n. Using a MathCad solve block, I_{115}, I_{116}, I_{117}, and I_{118} were solved in terms of S_{115}, S_{116}, S_{117}, and S_{118}. The equations were implemented in an excel spreadsheet where the area-centered signal intensities were plugged in to yield numerical values for I_{115}, I_{116}, I_{117}, and I_{118}. This method yields identical solutions obtained from creating a matrix of coefficients and Cramer's Rule.[49]



b-type	theo. *m/z*	exp. *m/z*	error (Th)	y-type	theo. *m/z*	exp. *m/z*	error (Th)
1	283.207	283.198	0.009	1	292.254	292.242	0.012
2	396.291	396.281	0.010	2	405.338	405.330	0.008
3	495.360	495.351	0.009	3	518.422	518.414	0.008
4	610.387	610.379	0.008	4	632.465	632.458	0.007
5	739.429	739.423	0.006	5	760.524	-	-
6	836.482	836.472	0.010	6	857.576	857.581	0.005
7	964.541	-	-	7	986.619	-	-
8	1078.584	-	-	8	1101.646	-	-
9	1191.668	-	-	9	1200.714	-	-
10	1304.752	-	-	10	1313.798	-	-
11	1577.987	-	-	11	1595.998	-	-

Supplemental Figure 3. Annotated MS/MS spectrum of BSA tryptic peptide HLVDEPQNLIK.

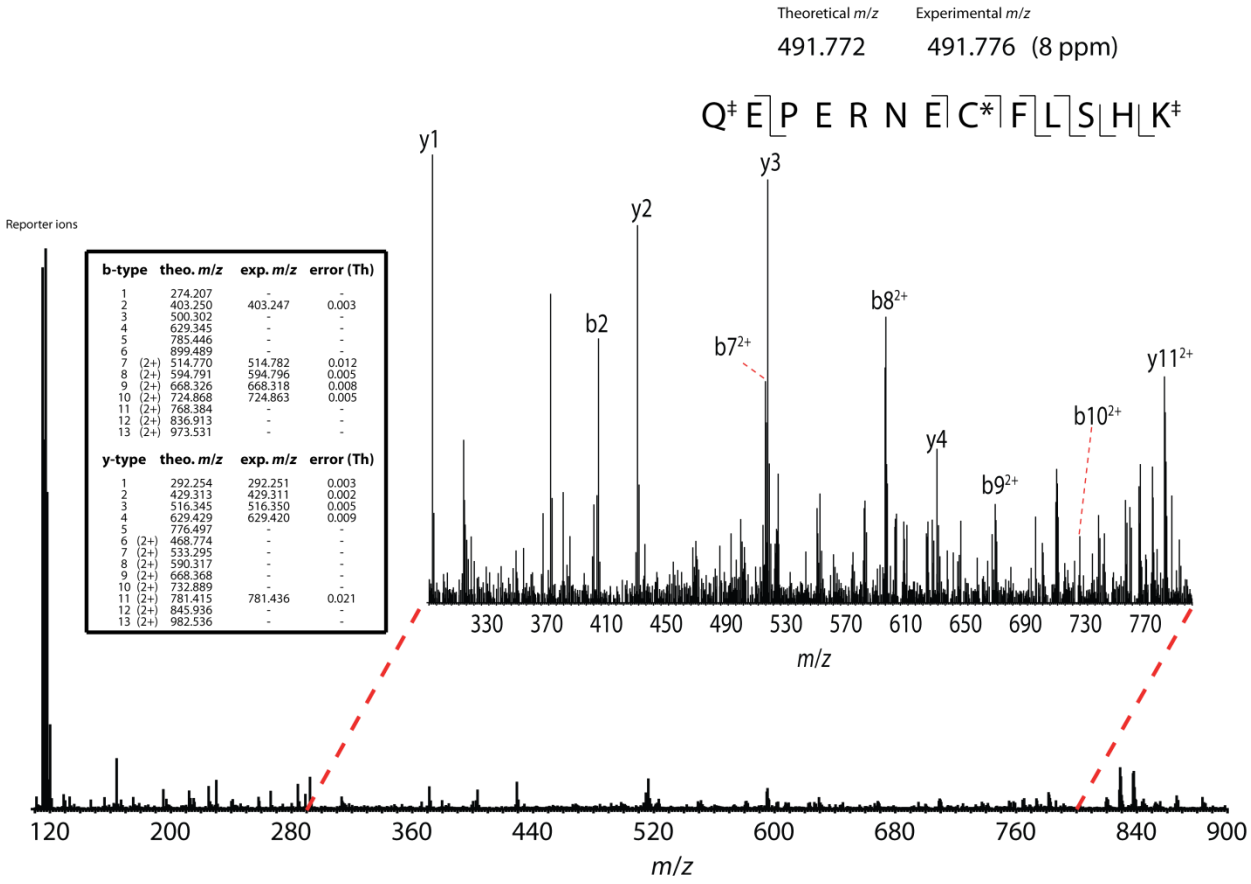
‡ Residue is modified with a DiLeu tag



Supplemental Figure 4. Annotated MS/MS spectrum of transferrin tryptic peptide NYELLCGDNTRK. The peak at m/z 1482.702 is hypothesized to be b2 (m/z 1481.736) with anomalous mass error. It should be noted that there are numerous fragments below m/z 660 which are not identified as belonging to NYELLCGDNTRK. It is hypothesized that they originate from non-canonical secondary fragments or a co-isolated peptide whose interference did not affect reporter ratio accuracy.

‡ Residue is modified with a DiLeu tag

* Residue is modified with a carbamidomethylation



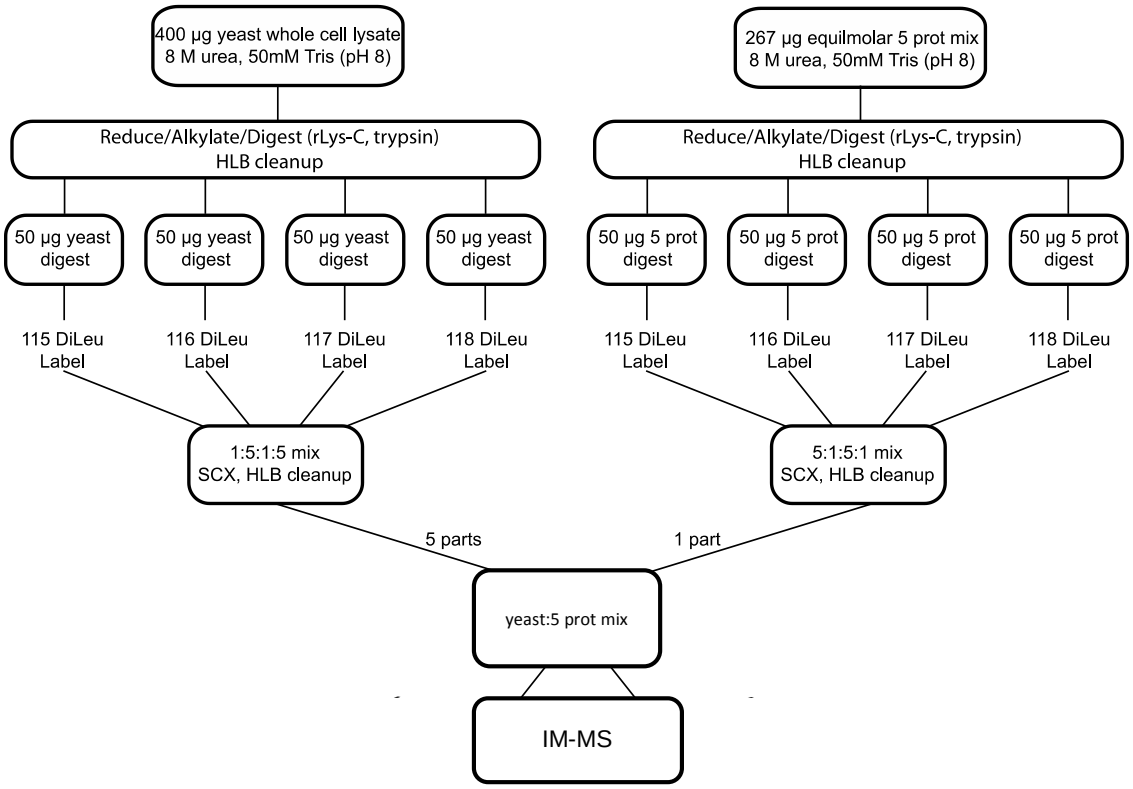
b-type		theo. *m/z*	exp. *m/z*	error (Th)
1		274.207	-	-
2		403.250	403.247	0.003
3		500.302	-	-
4		629.345	-	-
5		785.446	-	-
6		899.489	-	-
7	(2+)	514.770	514.782	0.012
8	(2+)	594.791	594.796	0.005
9	(2+)	668.326	668.318	0.008
10	(2+)	724.868	724.863	0.005
11	(2+)	768.384	-	-
12	(2+)	836.913	-	-
13	(2+)	973.531	-	-

y-type		theo. *m/z*	exp. *m/z*	error (Th)
1		292.254	292.251	0.003
2		429.313	429.311	0.002
3		516.345	516.350	0.005
4		629.429	629.420	0.009
5		776.497	-	-
6	(2+)	468.774	-	-
7	(2+)	533.295	-	-
8	(2+)	590.317	-	-
9	(2+)	668.368	-	-
10	(2+)	732.889	-	-
11	(2+)	781.415	781.436	0.021
12	(2+)	845.936	-	-
13	(2+)	982.536	-	-

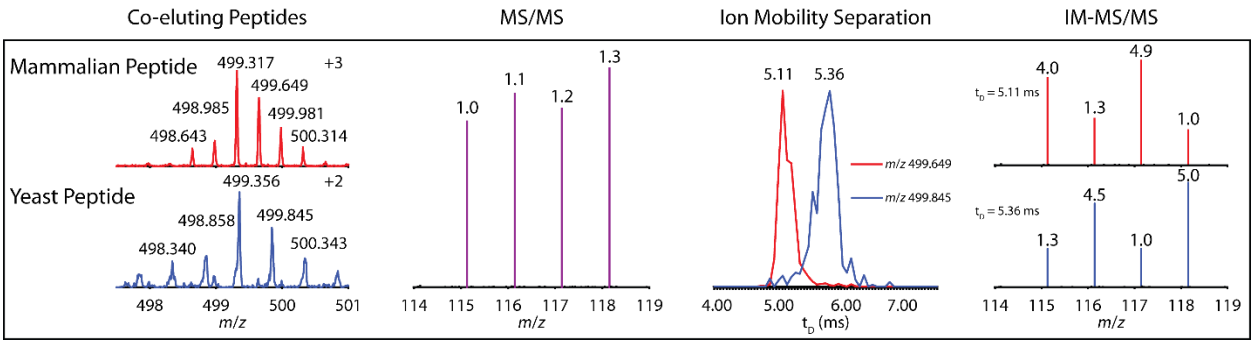
Supplemental Figure 5. Annotated MS/MS spectrum of BSA tryptic peptide QEPERNECFLSHK.

‡ Residue is modified with a DiLeu tag

* Residue is modified with a carbamidomethylation



Supplemental Figure 6. Experimental Workflow. Several hundred micrograms of protein from a yeast whole cell lysate or a standard mix of five mammalian proteins were digested with rLys-C and trypsin to produce peptide stock solutions for labeling. Four 50 µg aliquots of each protein stock were pipetted into separate microcentrifuge tubes and labeled with a single DiLeu isobaric tag channel. Labeled yeast whole cell lysate peptides were mixed at a ratio of 1:5:1:5 and labeled peptides from the protein stock solution were mixed at a ratio of 5:1:5:1. Yeast and standard protein peptide samples were combined together at a 5:1 ratio.



Supplemental Figure 7. IM was used to separate interfering precursors of different charge states (z = +2, +3), a tryptic mammalian peptide and a tryptic yeast peptide. Initial MS/MS fragmentation yielded a chimeric reporter ion spectrum near unity. After IM separation reporter ion intensities were extracted at the drift times of the precursors to produce corrected reporter ion spectra. The theoretical ratios for mammalian and yeast peptides were 5:1:5:1 and 1:5:1:5, respectively. IM traces were normalized to their own maxima, and all displayed reporter ion ratios were obtained from a purity correction algorithm applied to the centroid raw data.

Supplemental Table 1. The required resolution to resolve isobaric interference was predicted from the HDMS^E analysis of unlabeled tryptic yeast digest. Unlabeled yeast was used because of the incompatibilityof isobaric tandem mass tags with HDMS^E analysis. Pairs were selected as interference candidates from HDMS^E identifications that had apex retention times within 20 seconds of each other and precursor mass-to-charge ratios within ± 1.5 Th of each other. Resolution was "sufficient" if the required drift time resolution was 30 or less. Approximately 30.1% of pairs would not be resolved with resolution of 30 or less. Raw data was processed and searched against a SwissProt target-decoy database for *Saccharomyces cerevisiae* on ProteinLynx Global Server version 3.0. Carbamidomethyl cysteine and oxidized methionine were set as static and variable modifications, respectively. Precursor and fragment mass accuracy tolerances were set to automatic to achieve a 1% protein false-discovery rate. The apparent peptide-level false-discovery rate was 0.69%.

Peptide A	Peptide B	t_D A (bins)	t_D B (bins)	Req'd R	Sufficient R?
TGTLTTSETAHNMK	KEEQNIADGVEQK	59.416	54.044	11.1	yes
TGTLTTSETAHNMK	FEDAEVQR	59.416	61.078	36.7	no
EIHNEAESQLR	AAGSYYVEAQHK	54.103	50.993	17.4	yes
VDCSELSEK	GDWNGAGCHTNVSTK	64.576	55.836	7.4	yes
QSLDVSSDR	LTQEEIDR	61.9	61.064	74	no
WAENNYK	YQALSDPK	56.307	59.88	16.8	yes
LSSEEIEK	IEEELGDK	59.772	55.862	15.3	yes
AAGSYYVEAQHK	YSNLVCK	50.993	54.722	14.7	yes
KEEQNIADGVEQK	FEDAEVQR	54.044	61.078	8.7	yes
KEEQNIADGVEQK	TIGHSVQSPDDTVK	54.044	58.472	13.2	yes
LEAGLSDSK	YQALSDPK	56.292	59.88	16.7	yes
QINENDAEAMNK	LEEVDEEEEEK	79.369	84.851	15.5	yes
NVYTGEYK	AGFAGDDAPR	58.237	61.696	17.8	yes
NVYTGEYK	NAGGVGELEK	58.237	58.397	365	no
YSNLVCK	AVALGSPDR	54.722	54.729	7818.4	no

TDEYGGSIENR	VIGATTNNEYR	75.486	78	31	no
AGFAGDDAPR	NAGGVGELEK	61.696	58.397	18.7	yes
VIGATTNNEYR	SVAVDSSESNLK	78	75.475	30.9	no
IGSEFTEHTSANNNR	STTTGHLIYK	57.678	66.924	7.2	yes
IETELTK	AQHAVILDQEK	55.9	49.426	8.6	yes
FEASQVTEK	YDCSFCGK	62.993	62.634	175.5	no
FEASQVTEK	YDDALNATR	62.993	64.058	60.1	no
INRPPVSVSR	DYVLNK	51.745	47.945	13.6	yes
GYKPVAVPAR	ITQDAAEAIK	65.102	64.079	63.6	no
FASEDASIK	VLVHTQIR	57.879	63.911	10.6	yes
FASEDASIK	TQVTVEYK	57.879	61.539	16.8	yes
TFESEAAHGTVTR	VEYICEK	58.452	58.019	135	no
TEILANEQGNR	DSSSTFREWK	73.856	74.089	318	no
YDDALNATR	LEATDYATR	64.058	62.133	33.3	no
STLDPVEK	LLQEVASK	55.676	55.319	156	no
VEYICEK	QIEEQENLYHQHVSK	58.019	53.792	13.7	yes
DVQQIDYK	AVGAHSDLLK	61.93	67.508	12.1	yes
GVFQGVHDSAEK	VTAVVESPEGER	83.012	74.148	9.4	yes
LEAELAESK	GAEGELGAASK	61.004	63.096	30.2	no
LEAELAESK	WAAAGVCEK	61.004	58.732	26.9	yes
VIDGFNK	DSLGALSK	49.561	49.087	104.6	no
EDSVYLAK	YFTHNGNFGESPNHVK	60.407	57.025	17.9	yes
HGAVHMSTYLK	VLDPVSAK	56.376	53.883	22.6	yes
HGAVHMSTYLK	AVEQGFNVKPR	56.376	51.555	11.7	yes
HGAVHMSTYLK	VTNIIDR	56.376	52.206	13.5	yes
AVEQGFNVKPR	VTNIIDR	51.555	52.206	80.2	no
LNLTQEAK	GTQLGLEAK	59.036	58.479	106	no
GAEGELGAASK	WAAAGVCEK	63.096	58.732	14.5	yes
WAAAGVCEK	YIEQAIEK	58.732	63.27	13.9	yes
MANDLVQAR	MLVDGAGNIK	62.721	62.682	1608.2	no
YLPDASSQVK	AHYTNLYVK	68.353	74.084	12.9	yes
DQDSAVVSSNIK	NETNNIDGNFK	77.446	72.846	16.8	yes
EGIIKPISK	GSGIVASPAVK	65.743	60.606	12.8	yes
EGIIKPISK	GTSTYPVKPLDSSK	65.743	56.404	7	yes
GSGIVASPAVK	GTSTYPVKPLDSSK	60.606	56.404	14.4	yes
TIGIAVDHR	TGVVYNVTK	63.848	58.994	13.2	yes
SYPISEGPER	GAGSIFTSHTR	65.252	73.938	8.5	yes
YYTITEVATR	VDIIANDQGNR	70.154	69.893	268.8	no
LVYVTGGR	LLASAMEK	55.807	53.814	28	yes
YNIADAAR	VFDNIQR	59.751	55.219	13.2	yes
YNIADAAR	IAVHVTVR	59.751	59.113	93.7	no
ELVETSEVR	GLVEDANAAAK	67.075	65.967	60.5	no
IDDSGVSIGK	TPLPPAPAPK	59.854	62.902	20.6	yes

GTVIIVCSDGHK	AIQQAADEVASGK	82.092	77.897	19.6	yes
GASDGGLYVPHSENR	LFEQEGVSK	61.999	64.611	24.7	yes
SGQGAFGNMCR	HQGIMVGMGQK	70.37	74.054	20.1	yes
HAFMGTLK	GVLSVTSDK	60.251	55.668	13.1	yes
IAVHVTVR	SVVIHAGQDDLGK	59.113	55.84	18.1	yes
EEGGENDLIER	EIIQTGDEVEK	72.372	78.036	13.8	yes
MTPSGHNWVSGQGAGPR	YTSQHEWIAVHQDK	58.936	64.787	11.1	yes
ATAGDTHLGGEDFDNR	LEAESWQEK	57.477	68.304	6.3	yes
TVMIAAHGNSLR	AGMTTIVR	58.999	54.163	12.2	yes
YVELPEK	TSIQHVGMINGR	56.257	55.998	217.2	no
NDVRPEWEDEANAK	VANNSLEQLK	61.02	68.249	9.4	yes
STTEPIQLNNK	NQQLEDEIEK	73.543	78.697	15.3	yes
QIEYLNR	DGQLETFK	59.692	57.301	25	yes
EYLEEYK	LAQNDGLLK	63.151	61.668	42.6	no
GGGADVDWSSAR	NASNELELSAK	68.93	74.136	14.2	yes
ECINIKPQVDR	SSVGVIINK	56.867	56.585	201.7	no
IQQDTLIQTK	QITIGQAPTEK	73.949	74.97	73.4	no
VFNAYPAAR	IKPVVVINK	63.036	61.489	40.7	no
NEGATLITGGER	EQAEASIDNLK	69.32	76.996	10	yes
AFETLTDSNK	GNNIGSPLGAPK	69.498	69.214	244.7	no
EYLANGDDLK	SVDSAMDAVDK	74.502	65.31	8.1	yes
VYPVTGQTYSR	YSLAPVAK	58.558	57.866	84.6	no
HFISELEK	IEVAEPSVR	69.819	66.757	22.8	yes
HIDVTFTK	LHTPEENQLTVR	64.06	57.235	9.4	yes
SVDSAMDAVDK	GVIFYESHGK	65.31	70.494	13.6	yes
SVDSAMDAVDK	VQGIDYSKPEATNMR	65.31	67.102	37.4	no
ASFEKADQELK	LEPVVVNQPNR	81.582	74.656	11.8	yes
AYFTAPSSER	EADYFGDADK	69.766	66.317	20.2	yes
EQAIIDMAKK	GEEHKPLLDASGVDPR	70.046	66.171	18.1	yes
SDVMSVDIDKK	VQAVAVLK	46.793	60.39	4.4	yes
SWCIATHPTGR	QIAFPQR	54.28	55.07	69.7	no
IVTLCGDAPEGK	TVILDAENSPAK	77.837	78.479	122.2	no
LKPEYQEVIK	LREYLEEYK	52.885	49.721	16.7	yes
LSELIGAR	LLGPQLSK	53.724	54.475	72.5	no
DNIQGITKPAIR	QGVLLPTR	58.437	56.08	24.8	yes
VLPGPEVPK	LGLDETYK	56.629	59.411	21.4	yes
VLGIDAGEEK	ILTPESQLK	63.779	66.88	21.6	yes
ALEEVSLK	TFIAVKPDGVQR	55.187	58.893	15.9	yes
TFIAVKPDGVQR	TLLIADSR	58.893	58.915	2678	no
VAEEFGIK	YGLNICR	55.71	56.115	138.6	no
ITITGAPENVEK	AQIWDTAGQER	77.798	75.99	43	no
FIEQWNK	LANEKPEDVFER	57.584	58.174	98.6	no
REDLFITTK	ALVEHIIQAK	68.969	74.915	12.6	yes

REDLFITTK	FVTGGADNLVK	68.969	70.043	65.2	no
ISIIEENSPER	IVTEAVEIEQR	74.068	79.652	14.3	yes
SAVIIDNMCK	ETYSSYIYK	68.942	72.07	23	yes
SHINVVVIGHVDSGK	NDSTEYYNYFDK	62.49	53.756	7.2	yes
QATVGDNDKEKPGIFNMK	ATSNTLEEFEHGR	64.535	54.752	6.6	yes
IFEEEIGR	ATSNTLEEFEHGR	61.727	54.752	8.8	yes
MLTGDAVGIAK	TASGLYLPEK	67.527	66.219	51.6	no
QLVVGGLDR	SWEEFEK	61.62	60.009	38.2	no
NSLTGLDMGGGK	EGLALSSNISR	69.172	78.563	8.4	yes
TGELLCYDHYK	NLSLQTGTKPVLK	53.397	61.433	7.6	yes
LCVDLPNQK	LIQNEILNK	64.929	67.859	23.2	yes
LCVDLPNQK	CEDVFEYK	64.929	68.57	18.8	yes
TTDTYPSLPKPLNRPQK	IVQIMQNPTHYK	61.935	54.602	8.4	yes
NPEDLNIWTSTK	TIVTGGNGPEDFQQHEQIR	89.775	75.993	6.5	yes
ATNEWFR	TYAAEIAHNISAK	56.07	57.139	53.5	no
FQTPFR	DHGEGGIIVGSALENK	52.549	50.25	22.9	yes
LLTPEEIK	ELALQIQK	60.104	60.399	204.7	no
LTVDEVVSR	TGDIGEWEANGHLK	61.097	58.839	27.1	yes
TWETYEMR	EAQVDIEAIK	68.477	75.923	10.2	yes
LLGVEAGGDGVDTK	SLNIDDVDSIIK	82.048	88.596	13.5	yes
LESFQAVYNK	VVDALGNPIDGK	72.771	75.275	30.1	no
LLDAPAAIR	YETSILEHSGIR	57.809	61.171	18.2	yes
VISEILTR	MIAPMIEK	59.199	63.456	14.9	yes
MEMACDALYK	SVGIIGA	74.356	113.936	2.9	yes
MEMACDALYK	LSICGEESFGTGSNHVR	74.356	74.391	2125.5	no
AMEVVASER	QLAGIDDQGHLHTSGWSR	62.324	64.932	24.9	yes
LGVYSQHSQDQLDLTK	AIEVSETPNHVLYSNR	62.683	64.414	37.2	no
NINDLLNK	LEFYNTGKPTLK	59.002	61.365	26	yes
QPSLGPTLGVK	WVILGHSER	68.227	72.408	17.3	yes
LSICGEESFGTGSNHVR	IEVSSQESWGN	74.391	70.19	17.7	yes
MAGTFLFTSESVGEGHPDK	IVNDLIYR	56.915	62.929	10.5	yes
TFVNEIRR	TPQNTITMFRPDK	62.659	57.146	11.4	yes
EFHWEDVR	TVPNIYINGK	71.078	67.058	17.7	yes
SNPCLIGEPGIGK	KGDTYVSIQGFK	80.549	84.776	20.1	yes
LETNKAEYAEIMTANK	ILYGGSANGSNAVTFKDK	68.904	69.005	683.2	no
LAAPHHWLLDK	NDFIMVK	52.284	55.43	17.6	yes
LPLQDVYK	TGTKPAAVFTETLK	63.672	64.723	61.6	no
VMPAIVVR	IGPLGLSPK	58.225	57.608	94.4	no
IASTWEGIQAAK	ASGEIVSINQINEAHPTK	76.688	71.088	13.7	yes
VLNEQVDESYGLR	ALTEQAQTLTLSSR	95.232	95.181	1867.3	no
EEVQEIEELNK	MNPTNPDWINR	80.543	78.245	35	no
TAVFYGGTPISK	EEDATPASLHQDHYLVPYNQR	75.881	72.582	23	yes

YESSPSHIELVVTEK	AAIEDGWVPGK	62.901	68.273	12.7	yes
LDSIITSYR	YTVDQVFAK	68.873	66.077	24.6	yes
TGTKPAAVFAETLK	IGIVEISPK	64.408	65.033	104.1	no
FLDGIYVSHK	ATLELLK	49.5	52.205	19.3	yes
VFEGVLDDTK	EMGEEVYHAYTEVK	72.675	69.976	26.9	yes
FVENGISIATPNK	ATSSTDEESWFK	84.706	80.177	18.7	yes
FTLTEQELK	TVAGGAYTVSTAAAATVR	69.618	68.037	44	no
VYPVTGQTYSRK	SCGVDAMSVDDLK	92.85	88.912	23.6	yes
TVAGGAYTVSTAAAATVR	LCNALNALPK	68.037	68.79	91.4	no
SEVAAEAVCAMNPDLK	MIVAAPDYVR	61.131	72.873	6.2	yes
EIVFNMSR	YLDPLMNK	62.935	62.711	281	no
LQWMVDK	FVEELDKELEK	55.478	58.516	19.3	yes
ILIANTTLDTDK	ILGPGNQFVTAAK	85.62	82.066	24.1	yes
IYPEEMISTGVSAIDTMNSIAR	LLDAAVVGQEHYDVASK	69.097	73.88	15.4	yes
NTDPLNSALQENSDGVK	LEGVYSEIYK	64.624	72.957	8.8	yes
LGMENSLTEK	TGLINSSSSSGALGSQGR	71.153	69.501	43.1	no
QVLDVDPVYK	QALGEQLYKK	75.291	70.188	14.8	yes
QVLDVDPVYK	SIYNIYFEK	75.291	73.708	47.6	no
QVLDVDPVYK	VKPTVNQVETHPHLPQMELR	75.291	66.116	8.2	yes
GSASGDLTFLASDSGEHK	HYFIEINPR	66.552	80.106	5.9	yes
NLLSVAYK	LFLAQYR	57.556	59.859	26	yes
NLLSVAYK	QFLELTR	57.556	60.085	23.8	yes
TLGEGILEPTK	GLQTVVVEEDA	71.066	70.295	92.2	no
LTDDGHLEFFPHSSANGR	YRPNCPIILVTR	67.612	58.193	7.2	yes
ANELLINVK	MHLPGEETYWQK	65.108	54.908	6.4	yes
ETVEYPVLHPDQYTK	EEPALDWGAAR	73.931	70.653	22.6	yes
SNYGMIAINPGR	AIEQLNYTPIK	80.03	78.581	55.2	no
QTEQLEAQVLQVNK	DGYLQNNATEGDAEHITPDNLR	99.845	83.552	6.1	yes
TGDLVTIHYTGTLENGQK	MFTPIDQAIEHFKQNK	69.237	77.13	9.8	yes
YDMSLQLK	ECDLLFAK	63.19	67.793	14.7	yes
TAFKPHELTESVLPAAR	EVDIGIPDATGR	64.475	73.584	8.1	yes
ASIYNALSVK	ALENTVHEVTTSIGK	67.536	60.273	9.3	yes
ASIYNALSVK	ADIESYLEK	67.536	65.142	28.2	yes
ILDEFYDR	ADIESYLEK	65.752	65.142	107.8	no
ASFELAIERK	SVDSIYQVVR	75.832	69.243	11.5	yes
DELTLEGIK	SCGYGHVGDGNIHLNIAVR	62.076	69.752	9.1	yes
FDNLVFNR	IVYALTTIK	64.487	65.124	102.2	no
SCGYGHVGDGNIHLNIAVR	IVYALTTIK	69.752	65.124	15.1	yes
SIVPSGASTGVHEALEMRDEDK	VDRFLANEAK	68.221	75.391	10.5	yes
QLLCGAAIGTIDADKER	DEILLMENSR	80.167	75.041	15.6	yes
TGEVIINPLK	AHLMNPMVPGLAQGGK	65.895	69.55	19	yes
LSDLIEHFEK	FCYGIDDTLK	80.302	76.273	19.9	yes

QLASFGY	EEIFIANEGVHTGQFIYAGKK	151.381	95.97	2.7	yes
LITDAAYSFR	FISEQLSQSGFHDIK	71.11	68.702	29.5	yes
ALLELQVSK	LLGLTPVER	62.274	64.134	34.5	no
LATELPAWSK	SIAVHPTKPYVLSGSDDLTVK	66.433	70.904	15.9	yes
NSVIDAIVETHK	LYTFLAR	51.945	57.167	10.9	yes
FVIVMDDAGR	SEWFAQINK	68.854	67.807	65.8	no
VLDTGGPISVPVGR	KEAKDHGEGGIIVGSALENK	81.7	86.51	18	yes
ELLNGQSAQGLITCK	ASYPVYVSEYDDVK	99.326	104.678	19.6	yes
HVSPAGAAVGIPLSDVEK	LLEDAGFIER	80.451	71.261	8.8	yes
HVSPAGAAVGIPLSDVEK	VVLVQVAVPSR	80.451	71.304	8.8	yes
QLLCGAAIGTIEADKER	LIVDTLQDPSK	79.854	75.482	18.3	yes
NVSGQDVAAALEANAK	CTGGIILTASHNPGGPENDMGIK	95.119	76.31	5.1	yes
ILMVGLDGAGK	LLLIGNSGVGK	67.886	68.77	77.8	no
GAVLIDLTK	CGTAFALGIACAGK	58.119	50.8	7.9	yes
EQINGLTTTLQDHMK	QILDGLVGANR	69.137	72.091	24.4	yes
APSLFGGMGQTGPK	DSIGGVVTCVVRNLPTGLGEPCFDK	88.477	75.688	6.9	yes
VDFNVPLDGK	AHQVVEDGYEFFSK	65.595	60.889	13.9	yes
LAVQDPNFLK	LASAAALDALTK	70.129	69.393	95.3	no
QISNSNDHVNGSFSNPLDIEK	AQAEYFETSAGLMK	74.012	98.671	4	yes
STFVLDDWKR	LIDLCVGPHIPHTGR	50.119	54.833	11.6	yes
QFVVDVLHPNR	FIHGVATLLNDK	57.126	51.677	10.5	yes
IDVAVDSTGVFK	DGKPVSAFHDIPLYADK	76.559	76.058	152.8	no
SIAPAYGIPVVLHSDHCAK	SGVAVADESLTAFNDLKLGK	60.729	60.124	100.4	no
DGKPVSAFHDIPLYADK	STGSFVDPVITK	76.058	76.231	440.6	no
YGVSAEVINLR	MSANLDKSLDEIIGSNK	77.838	80.855	26.8	yes
SGVAVADESLTAFNDLKLGK	ISSIVIFNK	60.124	63.172	20.7	yes
HAVLHDQQSEAIYTPLSTLNNEK	VSHANEYGLYQNLPLDK	81.31	66.025	5.3	yes
SVMNIWTKK	ELYEVDVLK	65.488	68.748	21.1	yes
SVHFLSYPVVK	AGHFIVYFTPK	53.944	53.79	350.3	no
VSHANEYGLYQNLPLDK	FGTNIITETVSK	66.025	81.595	5.2	yes
VSHANEYGLYQNLPLDK	ELQFSEIDTVK	66.025	87.212	4.1	yes
FQVMVLR	SPFLDALK	59.197	57.361	32.2	no
SIVEKDGAFER	FPSGDQLHTAETGEFLK	76.469	70.353	12.5	yes
TGTFSIAALQR	IDYIVGLESR	69.671	72.288	27.6	yes
EQLASQLVDLK	STQLDLEAELK	81.682	74.151	10.8	yes
EQLASQLVDLK	KEGDDAPESPDIHFEPVVHLEK	81.682	71.578	8.1	yes
ELMNGLSNLYK	SIEMYVQEASK	52.195	47.106	10.3	yes
ADHLVEEVLEAR	TGPFEVGVHVLPK	56.172	58.413	26.1	yes
DQWSPALTLSK	NWQNVGSLVVK	74.557	78.308	20.9	yes
HPDADSLYVSTIDVGDEEGPR	YNEQIVDSMVYR	95.161	89.502	16.8	yes
HPDADSLYVSTIDVGDEEGPR	VGVQVAAQHNTSVFYGLPQEK	95.161	75.912	4.9	yes
HFEGVDGPSILITTK	HVDELLAECANVDK	65.849	62.872	22.1	yes

SFANQPLEVVYSK	FGINVETSTTEPYTYYIPK	91.145	74.232	5.4	yes
YNEQIVDSMVYR	VGVQVAAQHNTSVFYGLPQEK	89.502	75.912	6.6	yes
YNEQIVDSMVYR	ALIENWAADSVSSR	89.502	93.818	21.7	yes
QLLLHHTLGNGDFTVFHR	VHHLVNDQPGAWEEFVAR	64.998	62.016	21.8	yes
AEETIAWLDSNTTATK	LIIEEFGKEGNSTGEK	110.162	110.077	1296	no
ALLPHLTNAIVETNK	DKPYFDAEHVIQVSHGWR	61.89	60.919	63.7	no
EAQADAAAEIAEDAAEAEDAGKPK	ANLLDNNNKLDALR	93.465	101.465	12.7	yes
LSVLSAITSTQQK	VVEVCLADLQGSEDHSFR	86.224	79.974	13.8	yes
LSVLSAITSTQQK	TTPTLFENDVIK	86.224	84.663	55.2	no
QLIELLQAK	HLEDNTLLVTGPFK	67.425	63.359	16.6	yes
QLIELLQAK	ADILLLDEPTNHLDVSNVK	67.425	65.899	44.2	no
HLEDNTLLVTGPFK	ADILLLDEPTNHLDVSNVK	63.359	65.899	25.9	yes
HLEDNTLLVTGPFK	IPYLPLDPK	63.359	70.745	9.6	yes
ADILLLDEPTNHLDVSNVK	IPYLPLDPK	65.899	70.745	14.6	yes
IDVQVVCTHEDAMVFVDYK	DFYHATPAAFDVQTTTANGIK	78.049	94.752	5.7	yes
DNVKDLLSVSYNEFIDPK	VGIVPGEVIAPGMR	83.105	85.762	32.3	no
QIVFEIPSETH	AAIVQIDATPFR	78.023	78.758	107.2	no
SAVEVIEFVK	WLGEETFIK	68.045	68.096	1335.2	no
INELTLLVQK	HIEVQLLADNHGNVVHLFER	73.394	69.612	19.4	yes
SQIFSTAVDNQPTVMIK	TGTTTNDVDYPIPPNHEMIFTTDDK	115.7	90.617	4.6	yes
AQPAKCNLFSISEFLR	FGTNHFEFFIDRPEK	63.945	57.216	9.5	yes
TTNDTYEATYAIPMHCENCVNDIK	TDRPDGLYTTLVVDQYER	78.65	72.713	13.2	yes
LIMEDISEQEGHLQIK	TGYHINPTPEWFYQK	73.918	65.647	8.9	yes
DLTDYLMK	ADIMELLNHSDSR	63.087	59.973	20.3	yes
IASTWEGIQAAKELEEK	ALPDAVTIIEPK	74.898	86.414	7.5	yes
YSEGIRSENDEAEVALR	LGNDDEVILFR	74.702	77.069	32.6	no
DAITWLNEHDIK	WIDLPISK	54.835	60.068	11.5	yes
YVVSFDELEAR	VSLTTGPVHDFTWSPTSR	82.358	66.764	5.3	yes
ALDTIEDDMSIEHDLK	AAQDSFAANWGVMVSHR	69.357	69.317	1733.9	no
DSGLWGFSTATR	QFCTPLPYQLRQSVVMTNPSELSMKPR	78.081	76.196	41.4	no
VVNMEFPIDEATGK	AADDNSQAVELGAMR	92.685	100.304	13.2	yes
YGHAEDFVPILESPEGK	SGPILATVWEGK	62.72	75.17	6	yes
QFCTPLPYQLRQSVVMTNPSELSMKPR	LEAAAQDAINWLDASQAASTEEYK	76.196	80.453	18.9	yes
MEMIISSPSDGQVK	ELFSNAIDEINEK	89.198	94.032	19.5	yes
ISIFEAVHGSAPDIAGQDK	FTVTLIPGDGVGK	66.135	80.167	5.7	yes
IFAIAFTR	TVDADEEQDFDK	60.857	51.461	6.5	yes
LREEIFIANEGVHTGQFIYAGK	SEELYPGRPLAIALDTK	71.813	64.761	10.2	yes
LREEIFIANEGVHTGQFIYAGK	TNLPPPSPQAHMHIQSDLSPTPK	71.813	71.869	1283.4	no
AALEAGAFEAVTSNHWAEGGK	NLQEVLNPQIIK	80.364	89.628	9.7	yes
GYIPLQAPVMMNK	ADYANGVLTLTVPK	103.109	91.872	9.2	yes
QLQLIQVEIK	LGLNIPLNEVK	79.133	76.25	27.4	yes
WIETQDVFLK	QLVGSYGVVINK	78.135	81.024	28	yes

DYGDAFEGIENTSLSPK	MSAKSFEVTDPVNSSLK	117.948	110.088	15	yes
LNVPLLAPMSQSNTHESSYASK	MFNLSVENPLGSDHPK	70.48	68.483	35.3	no
LIGRNFNDPEVQGDMK	NLKPEVDNDNAAMELR	51.517	51.65	388.3	no
MFNLSVENPLGSDHPK	EIGYLFGAYR	68.483	74.205	13	yes
TDTTQTAPVTNNVHPVWLLR	LVWIETPTNPTLK	78.644	91.037	7.3	yes
ILLTEPPMNPLK	ELLADFVESESK	83.824	84.09	316.1	no
EFEGGVIIITHSAEFTK	GGFQSQLTFLR	75.39	78.273	27.1	yes
IADISLAAFGR	NAAWLVFANK	69.675	69.053	112	no
QDGVLLVALSNEPAAR	GFEIGSGFQGVSVPGSK	100.897	98.73	46.6	no
VVGLSTLPEIYEK	GGQAIVVHPDYLYGAITEDR	89.01	78.343	8.3	yes
LIIATTRPETIFGDTAVAVHPDDDRYK	DAHQFLHDHDHIFPLDLPSR	71.293	68.705	27.5	yes
LSSESVIEQIVK	DFLPRGTGIVTR	83.821	83.994	485.5	no
TKDQVINAYENLEDELSLIK	TAQLSLQDYLNQQANNQFNK	96.264	81.2	6.4	yes
DIPVPEPKPNEILINVK	LTCRPGLELEPFDLEK	67.92	71.631	19.3	yes
DFSPLNVGSDWK	LIIATTRPETIFGDTAVAVHPDDDR	83.51	83.91	209.8	no
TNITFEGPEIAIVNR	TNQLVPEVLEYNVR	100.268	103.468	32.3	no
ALVDMGMLKDDANEIFSKPIAWNEALK	LAAIVEIIDQK	73.387	75.903	30.2	no
FFDNHIFASCSDDNILR	NFKPQGSIEHLQSGVYYLTNIDDK	68.959	82.182	6.2	yes
LHAVTIDNVAEANFFK	LVEIPNLGCEEKDYASVVPPR	67.35	72.781	13.4	yes
ANIQLDFPELKPYK	NITFKHTIDSSTEENSQILAFTSSK	63.857	66.082	29.7	yes
GLMNFVSIDAR	LFGVTTLDSIR	76.588	77.096	151.8	no
EIELAEHEMPGLMAIR	HFASYANLPGTITHGMFSSASVR	66.307	74.671	8.9	yes
AYFEWTIEARDLSLLGSHVR	SLLSEYNILK	70.361	74.308	18.8	yes
LATLQPTWHPTSEELAIGNIK	QVAQWIADIQYGR	79.698	94.792	6.3	yes
SLPLCQNFSLK	ASGQEFPVNLVTCFEGMEESGSLK	81.116	82.58	56.4	no
IPIFSASGLPHNEIAAQICR	IEFELYDNVVPK	75.083	93.504	5.1	yes
ENTEGEFSGLEHESVPGVVESLK	EYVPSVIEPSFGIGR	80.74	98.879	5.5	yes
GFTPGDLNGISVEERR	FHTEVLSELLQADDK	63.631	69.182	12.5	yes
WSSDQLTNWLESHK	HSEFVAYPIQLVVTK	59.111	69.65	6.6	yes
DFEITAKEVYDFTHAVGNNCEDFVSRPDR	IPVMSIGLGPVYK	76.647	87.912	7.8	yes
LPENESLSTVMGILGSGVHRVAITNEEMTK	ENVKAVCDSLTEEGEIIWELGSLQERPK	101.08	100.722	282.3	no
AHNVSTFNNSPGPDNDLIRGSTTEPIQLK	TAEAIVQFMIK	76.098	76.924	93.1	no
TSFFQALGVPTK	DPFDIFASTASK	78.998	75.975	26.1	yes
DPFDIFASTASK	TEVNSGFFYSSADQRDK	75.975	74.172	42.1	no
GYFIRPTVFYDVNEDMR	DWDVLATAGNTNAWESTLR	72.996	73.213	337.4	no
EFIILGGGQEAKDVTTTSANEGK	VEDGNILVFQVSMK	89.364	95.842	14.8	yes
DWDVLATAGNTNAWESTLR	LNNVFVIGEQGKPYISLPK	73.213	89.734	5.4	yes
LVEDPQVIAPFLGK	SGKDAFEHLLCGASMLQIGTELQKEGVK	97.713	97.222	199	no
LDPTKEEDKEFVNNMWAWDKPVVVNGEDKEIVDGK	EFNYSEELIDQLK	98.187	102.376	24.4	yes
LEDVLDQEPDAGLGNGGLGR	QLSSLFLDGLEK	80.294	82.894	31.9	no
LVGPTAEVIGAVSGGVDSTVASK	HFEDVGLDSEIANHTPLGSLSGGQLVK	92.392	87.586	19.2	yes

AVDDFLLSLDGTANK	NDSSLVNFNIQISK	93.102	93.278	530	no
LDNGEDVLAPYDHYKDTR	DILSVDYTDIMK	71.534	90.697	4.7	yes
VTFDIEPYKPINTSLYLCDNK	AVGIDLGTTYSCVAHFSNDRVDIIANDQGNR	85.966	92.389	14.4	yes
NDSSLVNFNIQISK	ICLPTFESEELIK	93.278	103.027	10.6	yes
ICLPTFESEELIK	INLDYVKPVSTGIQVINAGELK	103.027	83.014	5.1	yes
ICLPTFESEELIK	INLDYVKPVSTGIQVINAGELK	103.027	102.687	303	no
VFDALNDLEQLK	MIEIMLPPRDVPQALLDR	90.935	71.349	4.6	yes
VFDALNDLEQLK	LIPSDFILAAQSHNPIENK	90.935	78.342	7.2	yes
MIEIMLPPRDVPQALLDR	LIPSDFILAAQSHNPIENK	71.349	78.342	11.2	yes
AQALAVAIGSGYVYQTTFER	STAISLYSEMSDEDVKEIK	61.822	58.156	16.9	yes
LGTLDGHTGTIWSIDVDCFTK	VDSIIIGGGMAFTFK	88.302	96.687	11.5	yes
NTWGSQLVDQIMDQVVTK	DLLHKPEYYGIKPEWIDHEIVPIMHTEK	88.927	82.303	13.4	yes
INLDYVKPVSTGIQVINAGELK	IAYEIELGDGIPKMFHVDVAR	83.014	79.09	21.2	yes
INLDYVKPVSTGIQVINAGELK	EYGADELFDYHDADVIEQIK	83.014	77.526	15.1	yes
GYSDVCDIFTEDELVR	DIYGWTQTSLDDYPIK	115.826	118.409	45.8	no
IAYEIELGDGIPKMFHVDVAR	EYGADELFDYHDADVIEQIK	79.09	77.526	50.6	no
IAYEIELGDGIPKMFHVDVAR	SVLSVANVGFFSSDR	79.09	96.702	5.5	yes
YADELLSIEQIKPFGQLVNSSR	DVAWSPTVLLR	76.158	76.874	107.4	no
AIDEQFGSLDELIK	VLEDMVFPTEIVGK	94.862	98.232	29.1	yes
DEFAYQEMIAHLALNSHPNPK	TLPNLVNGIFK	69.835	74.95	14.7	yes
SINPDEAVAYGAAVQGAILTGQSTSDETK	EELFTSLGGEVFIDFTK	100.292	124.916	5.1	yes
QDVSDLLDQDEYIDLVVPR	TVLFLPGFDHAGIATQSVVEK	80.324	76.651	21.9	yes
VFFVSGDPLLHTTAWK	VINSMFIFNK	70.073	76.268	12.3	yes
GVLVVPGSWFK	IAEFSAISFPYGAKTGK	73.785	71.005	26.5	yes
SLSAQALNHVSALIEEQETINPVQAQK	SKNGGSNNDETLLLMDLK	123.969	121.196	44.7	no
IYNTLVDLLR	SLEDLDKEMADYFEK	77.514	66.118	6.8	yes
AQRPITGASLDLIKER	ALENPTRPFLAILGGAK	69.254	60.901	8.3	yes
ELSWIDVENWHNLGGSEIGTNR	IPVLGHADGICSIYLDEDADLIK	82.141	87.787	15.5	yes
NLPDMIEELDLHED	AVGIDLGTTYSCVAHFANDRVDIIANDQGNR	99.803	92.984	14.6	yes
GVMNAVNNVNNVIAAAFVK	NFAMENQMLFTETSALNSDNVDK	78.235	66.715	6.8	yes
LITSHLVDTDPEVDSIIKDEIER	EQFVDDLEQALYASK	120.011	110.616	12.8	yes
WLTGVELADMYHSLMK	ILDGIFQIAGVK	78.273	79.29	78	no
VLVIVGPCSIHDLEAAQEYALR	QRLESYVASIEQTVTDPVLSSK	89.133	84.502	19.2	yes
TTYVTQRPVYLGLPANLVDLNVPAK	TGNIVDVPVGPGLLGRVVDALGNPIDGK	92.135	92.178	2143.7	no
TTYVTQRPVYLGLPANLVDLNVPAK	TTYTTQRPVYLGLPANLVDLNVPAK	92.135	121.786	4.1	yes
TTYVTQRPVYLGLPANLVDLNVPAK	VVPATNQIEIHPLLPQDELIAFCK	92.135	91.774	255.2	no
TGNIVDVPVGPGLLGRVVDALGNPIDGK	TTYTTQRPVYLGLPANLVDLNVPAK	92.178	121.786	4.1	yes
TGNIVDVPVGPGLLGRVVDALGNPIDGK	VVPATNQIEIHPLLPQDELIAFCK	92.178	91.774	228.2	no
TTYTTQRPVYLGLPANLVDLNVPAK	VVPATNQIEIHPLLPQDELIAFCK	121.786	91.774	4.1	yes
LAEQFAADLDAADKLPFSPIK	VHACIGGTSFVEDAEGLRDAQIVVGTPGR	85.043	95.357	9.2	yes

IEEVIDLILR	VGENPQAYPELMKEHLEVTGGK	77.29	76.511	99.2	no
ASLNDDSLHDLAALLNEDSGIAQR	SGDIIEPLLKPQWWVSQSEMAK	101.321	86.606	6.9	yes
MVLSGEGSDEIFGGYLYFAQAPSAAEFHTESVQR	LPENDCLYAIYDFEYEINGNEGK	113.093	109.739	33.7	no
NLENNLNLLKENASQLEFNKEDAK	DTLYILLDNALK	86.934	91.874	18.6	yes
NDNHIIQAIALLDQLDK	VLLPWWNSFK	77.739	81.745	20.4	yes
VFDNFRPVVFGLFQSK	IQAGVCELFREYLATK	66.856	73.035	11.8	yes
LINFGFGHDEAILANGDIAGINNDIEK	HVVFGEVLDGMDVVHYIENVKTDSR	119.821	94.395	4.7	yes
ISDDILSVLDSHLIPSATTGESK	ITDVVNIGIGGSDLGPVMVTEALK	81.974	98.045	6.1	yes
ISDDILSVLDSHLIPSATTGESK	ITDVVNIGIGGSDLGPVMVTEALK	99.538	98.045	66.7	no
HVVFGEVLDGMDVVHYIENVK	TGIHFGCDVFAVYEGYEGLLR	99.876	84.25	6.4	yes
QLPYLVANVVEVMDMNEIEDK	FAIGDYIYGVIHGASVR	67.784	72.756	14.6	yes
ENTEESIIADLPLPDLPPEPPTELSDSFK	IYPEEMISTGVSAIDTMNSIAR	95.557	95.506	1873.7	no
SVQNLIAELLSSDK	SAYLAAVPLAAILIK	94.913	99.941	19.9	yes
EYEEILAKIDGYSGPELGELMEK	TCNVLVAIEQQSPDIAQGLHYEK	77.403	77.315	879.6	no
TVESVNVIVSTYADEVLGDVQVYPAR	GALATYGLTIDDLGVASFHGTSTKANDK	120.514	114.68	20.7	yes
GGLTYNDFLVLPGLVNFPSSAVSLQTK	LLEVAQVPKEHIDEFK	100.224	122.008	5.6	yes
FENELDSFFTLFR	LASDHQHPQLQPIHILAAFIETPEDGSVPYLQNLIEK	100.429	112.458	9.3	yes
SSLSEFTEGGFQWIIEDIEHGK	LLQTPIDMSLKPNDAESEKEVIDTILALVK	87.69	102.813	6.8	yes
FSADIVTLFDLIEK	GETEEEYLWCIEQQLFAFK	100.712	88.296	8.1	yes

Chapter 4

Large-scale collision cross-section profiling on a traveling wave ion mobility mass spectrometer

Adapted from: Christopher B. Lietz, Qing Yu, Lingjun Li. Large-Scale Collision Cross-Section Profiling on a Traveling Wave Ion Mobility Mass Spectrometer. *J. Am. Soc. Mass. Spectrom.* **2014**, *25*, 2009-2019.

Abstract

Ion mobility (IM) is a gas-phase electrophoretic method that separates ions according to charge and ion-neutral collision cross-section (CCS). Herein, we attempt to apply a traveling wave (TW) IM polyalanine calibration method to shotgun proteomics and create a large peptide CCS database. Mass spectrometry methods that utilize IM, such as HDMS^E, often use high transmission voltages for sensitive analysis. However, polyalanine calibration has only been demonstrated with low voltage transmission used to prevent gas-phase activation. If polyalanine ions change conformation under higher transmission voltages used for HDMS^E, the calibration may no longer be valid. Thus, we aimed to characterize the accuracy of calibration and CCS measurement under high transmission voltages on a TW IM instrument using the polyalanine calibration method and found that the additional error was not significant. We also evaluated the potential error introduced by liquid chromatography (**LC**)-HDMS^E analysis, and found it to be insignificant as well, validating the calibration method. Finally, we demonstrated the utility of building a large-population peptide CCS database by investigating the effects of terminal lysine position, via LysC or LysN digestion, on the formation of two structural sub-families formed by triply charged ions.

Introduction

Ion mobility (IM) is a gas-phase electrophoretic separation technique wherein a weak electric field pulls analyte ions through a drift region filled with inert buffer gas. As ions traverse this region, they undergo collisions with neutral gas molecules. Ion drift velocity is dependent on charge state and number of collisions with gas molecules.[1] Therefore, the time that an ion arrives at the end of the drift region can be largely determined by ion charge (z) and collision cross-section (CCS).

Although IM was first combined with mass spectrometry (MS) in the 1960s,[2-3] the additions of soft ionization and nested time-of-flight (TOF) analysis transformed IM-MS into a powerful probe for biomolecules nearly three decades later.[4-6] Despite lower resolution than methods such as high-performance liquid chromatography (HPLC), IM-MS has found increasing analytical utility and applications as evidenced by extensive literature.[7-9] Electrodynamic ion funnels provided a sizeable enhancement to sensitivity,[10] and the cyclic drift tubes[11] and field-asymmetric ion mobility spectrometry[12] provided IM resolutions from 500 to 1000. In some IM-MS methods, ion CCS values can be measured and compared with candidate structures obtained through molecular dynamics (MD) simulations, allowing the investigation of three-dimensional structures.[13-16]

The development of commercial instrumentation has greatly expanded the reach of IM-MS, particularly in the area of intact protein–protein and protein–ligand complexes.[17-20] The majority of structural investigations on commercial instruments have utilized some form of the first[21] or second[22-23] generation SYNAPT HDMS (Waters, Milford, MA, USA). Unmodified SYNAPT instruments incorporate a traveling wave (TW) ion guide in the mobility cell,[24] creating a non-uniform electric field during separation.[25]

Consequently, a TW IM drift time (t_D) is not linearly proportional to CCS, and the instrument must be calibrated with ions of known CCS from drift tube analysis.[17, 26-27]

Bush et al. have recently characterized a TW IM calibration strategy focused on peptides.[28] Singly, doubly, and triply protonated polyalanine CCS were directly measured on a modified drift tube SYNAPT G1, and then used as calibrants to measure the CCS of tryptic peptides on a standard TW IM SYNAPT G2. The tryptic peptides were also directly measured on the G1 to evaluate TW IM CCS accuracy. Since previous strategies were primarily aimed at proteins and protein complexes, the polyalanine method was a significant step forward for TW IM peptide analysis. In the current investigation, we extend the previous study by evaluating the potential utility of the polyalanine calibration strategy for measuring CCS from large-scale proteomic datasets.

Peptide CCS databases have been previously used to calculate the size contributions and volumes of specific amino acids,[29-32] conformations of peptide–metal complexes,[33-34] and intrinsic structural preferences of peptides in the gas phase.[35-37] Typically, biologically relevant structures are thought of as the native-like solution structures, but the gas phase offers a glimpse at the innate, solvent-independent intramolecular interactions that are potentially present in anhydrous environments, such as membranes.[38] Peptide CCS databases have commonly come from the proteolytic digestion of protein mixtures, and, to the best of our knowledge, all but one peptide CCS database have come from static-field home-built IM instruments. The 2011 data by Valentine et al. was acquired on a SYNAPT G2, though reduced mobility was recorded instead of CCS.[32] Here, we evaluate the TW IM polyalanine calibration method applied to peptides identified by LC-HDMS^E on a SYNAPT G2 mass spectrometer.[39-41] HDMS^E has

the ability to identify thousands of peptides in a single experiment. In order to increase the throughput and speed of large-scale CCS peptide database construction, we created custom software called *pepCCScal* to measure CCS values from proteomic database search output.

In an optimal high-throughput method, *pepCCScal* would be the only additional step outside of the HDMS^E workflow. Polyalanine could simultaneously be used for lockspray correction and concurrent CCS calibration. However, instrument parameters and transmission voltages used for sensitive HDMS^E analysis can be significantly different from what is traditionally employed for CCS measurement, and calibration accuracy has never been validated under such settings. Therefore, the validity, accuracy, and reproducibility of polyalanine calibrants for LC-HDMS^E were evaluated. Additionally, the possible complications of using on-line LC when measuring CCS are assessed, given the very strong evidence that ions can retain some structural elements from their electrospray solutions.[42] Finally, the utility of large-scale HDMS^E CCS profiling is demonstrated by comparing the CCS of C-terminal lysine and N-terminal lysine peptides derived from enzymatic digestion of yeast and human cell lysates.

Experimental

Chemicals and Biological Samples

Trypsin, endoprotease lysine-C (LysC), whole cell yeast protein extract from *Saccharomyces cerevisiae*, whole cell protein extract from human K562 cells, and DL-dithiothreitol (DTT) were obtained from Promega, Madison, WI, USA. MS-grade Pierce LysN protease (LysN) was obtained from Thermo Fisher in Rockford, IL, USA. Bovine serum albumin (BSA), bradykinin, poly-DL-alanine, iodoacetamide (IAA), and

trifluoroacetic acid (TFA) were purchased from Sigma Aldrich in St. Louis, MO, USA. Optima grade water, acetonitrile (ACN), formic acid (FA), and acetic acid (HAc) were purchased from Fisher Scientific, Pittsburg, PA, USA.

Digestion of Proteins and Cellular Extracts

Approximately 500 µg of protein was used for each digestion. The samples were denatured in 8 M urea and 50 mM tris•HCl (pH ~8), then reduced and alkylated. DTT was added to 5 mM and incubated at room temperature for 60 min, followed by IAA addition to 15 mM and room temperature incubation for 30 min in darkness. Alkylation was quenched by a second aliquot of DTT, and the samples were diluted with 50 mM tris•HCl. For trypsin and LysN digestions, urea was diluted to less than 1 M. For LysC digestions, urea was diluted to 4 M. Enzyme was added to create specific enzyme:protein ratio (w/w) for trypsin (1:50), LysC (1:50), and LysN (1:75). Trypsin and LysC digests were incubated at room temperature for 17 h. LysN digests were incubated at 37 ºC for 4 h. All digests were quenched by adding TFA to lower the sample pH to 1 and then stored at -80 ºC. Reversed-phase (RP) solid-phase extraction (SPE) was performed by Sep-Pak cartridges (Waters, Milford, MA, USA) according to the manufacturer's protocol.

Off-Line Strong Cation Exchange Fractionation

Yeast proteins digested by LysC and LysN were fractionated by strong cation exchange (SCX) on an Alliance HPLC (e2695 separation module, 2489 UV/visible detector; Waters, Milford, MA, USA). Samples were reconstituted in Solvent A (20% ACN, 10 mM KH_2PO_3, pH 2.6) and loaded onto a polysulfoethyl A column (200 × 2.1 mm, 5 µm, 300 Å, PolyLC Inc., Columbia, MD. USA) by 100% Solvent A for 3 min. Separation occurred by ramping Solvent B (20% ACN, 10 mM KH_2PO_3, 500 mM KCl, pH 2.6) at

0.2 mL min^{-1} to 48% over 36 min, then to 100% B over the next 12 min. Fractions were collected every 3 min and then combined to make a total of eight fractions per digest. SCX salts were removed by Sep-Pak SPE cartridges.

CCS Calibration and Measurement

The CCS calibration strategy is similar to that described by Bush et al.,[28] but we used a t_D/CCS power relationship instead of a quadratic relationship.[17] A detailed explanation of our calculations and our *pepCCScal* software can be found in the next section and in Supplemental Information 1.

Polyalanine was suspended in 50% ACN, 1% HAc at 0.1 mg mL^{-1}, and acquired under five different TW IM wave velocity (WV)/wave height (WH) ratios (m s^{-1} V^{-1}): 500/30, 500/35, 600/35, 700/35, and 800/40. WH refers to the magnitude of the sequential voltage pulses that make up the traveling wave, and WV refers to the propagation speed of those pulses. Instrument parameters used for soft transmission, hereafter referred to as “low voltage,” had the spray voltage (nano ESI capillary voltage), cone voltage, extraction cone voltage (orthogonal skimmer cone), trap bias (energy for injection into helium cell), and helium cell DC (energy for injection into IM cell) set to 1.2–2.0 kV, 20 V, 4 V, 35 V, and 15 V, respectively. Instrument parameters used for $HDMS^E$ analysis, hereafter referred to as “high voltage,” had complementary settings of 3.0 kV, 30 V, 2 V, 48 V, and 35 V, respectively. Direct infusion bradykinin and BSA CCS measurements were acquired with a method and tune page identical to the calibration. The listed ± variations represent a single standard deviation of the measurement. Absolute errors must include the average calibration error (approxitately 0.5%) and errors associated with the calibrants themselves (approximately 3%).[28]

Direct Infusion MS

MS was performed with the Nanolockspray source of the SYNAPT G2 HDMS. The average pressures (mbar) of the instrument backing, source, trap, helium cell, and IMS were 2.73, 1.22×10^{-3}, 2.66×10^{-2}, 1.46×10^{3}, and 3.67, respectively. For the bradykinin acquisitions, analyte was dissolved in 100% water at concentrations of 0.9 to 9 µM. The ion of interest was isolated by the quadrupole and the maximum *m*/*z* in the mass range was decreased to minimize the number of milliseconds per IM bin. The CCS distributions were converted from extracted t_D distributions with a window of 0.001 Th. For BSA experiments, the digest was concentrated to 1 µM in 50% ACN, 1% HAc. All peptides were analyzed simultaneously with an *m*/*z* range held at 100 to 1800. The CCS distributions were converted from extracted t_D distributions with a window of 0.001 Th.

CCS Calculation by pepCCScal

The *pepCCScal* software was created to calculate peptide CCS values from SYNAPT G2 measurements. As of publication, it is not compatible with enhanced duty cycle (EDC) measurements. The process begins with a comma separated values (csv) list of user-defined CCS calibrants (polyalanine). Once the calibrant IM-MS spectra were acquired, the file was opened in DriftScope for peak detection. A lockspray-corrected, nonchromatographic peak list csv was then exported. The *pepCCScal* software then parsed the peak list for the *m*/*z* of the calibrants and extracted their corrected and centroided t_D. After creating a best-fit linear regression for CCS versus reduced t_D (for details on calculation, see Supplemental Information 1), the equation is used to recalculate the calibrants' CCS from t_D, and the calculated CCS is compared with the known CCS to determine the mean calibration error (%) and RMSE ($Å^2$).

After calibration, measuring unknown CCS can be done in two ways. In direct infusion mode, the user creates a list of peptide ion sequences and charge states for which to calculate CCS. DriftScope peak lists are exported and searched by *pepCCScal* in an analogous way to the calibration peak list. In HDMS^E mode, *pepCCScal* utilized the Final Peptide csv created by PLGS. Our software extracts mass, charge, and t_D to calculate the CCS of every identified peptide.

LC-HDMS^E Acquisition

Online RP-LC was performed on a nanoAcquity UPLC (Waters). Columns with integrated emitters were pulled from fused silica tubing (360 µm o.d., 75 µm i.d.) with a laser puller (P-2000; Sutter Instrument Co., Novato, CA, USA). Emitter tips were etched with hydrofluoric acid for 2.25 min, cut to 12 cm, and then packed with 1.7 µm, 150 Å, BEH C18 material obtained from a Waters UPLC column (part no. 186004661). Approximately 1.5 µg of peptides dissolved in Solvent A (water, 0.1% FA) was loaded on-column without trapping. Solvent B (ACN, 0.1% FA) was ramped from 0% to 7% over 2 min, and then to 30% during the following 118 min. The flow rate was set to 0.4 µL min^{-1}. WV was set to 800 m s^{-1} and WH was set to 40 V for all HDMS^E acquisitions. Collision energy in the transfer was ramped from 28 to 48 V in the high energy scan. Tryptic BSA peptides were acquired in triplicate. Tryptic HK digests, fractions of yeast LysC digests, and fractions of yeast LysN digests were acquired in duplicate. A CCS calibration was acquired before each set of analyte acquisitions.

PLGS Search Parameters and CCS List Curation

HDMS^E raw files were processed and searched on PLGS ver. 2.5.1. Low energy, elevated energy, and intensity thresholds were set to 100, 20, and 1000 counts,

respectively. SwissProt reference protein databases for *Saccharomyces cerevisiae*, *Bos taurus*, and *Homo sapiens* were downloaded from UniProt (www.uniprot.org) and used for searching. Precursor and fragment mass tolerance was set to automatic to obtain a protein false discovery rate (FDR) of 1%. Minimum number of ion fragments per peptide was set to 3, minimum number of ion fragments per protein was set to 7, and minimum number of peptides per protein was set to 1. Carbamidomethyl cysteine and oxidized methionine were selected as fixed and variable modifications, respectively. The maximum number of miscleavages was set to 2. The mean apparent peptide FDR was (0.6% ± 0.1%).

To ensure that we only report highly confident IDs and CCS values from K562 and yeast samples, peptides had to meet specific curation criteria. Only first-pass and miscleavage identifications found in both duplicates were used for CCS analysis. The list was then purged of all identifications with less than 95% confidence ("yellow" and "red" peptides). Finally, any peptide whose CCS varied 1% or more between analyses was also removed.

Results and Discussion

Accuracy of CCS Calibration Under High Transmission Voltages: Bradykinin

Our low voltage parameters result in significantly less HDMS^E identifications relative to high voltage settings. In Supplemental Figure 1, two chromatograms are shown for the LC-HDMS^E analysis of tryptic yeast protein extract. The data in Supplemental Figure 1a was acquired under low voltage settings and resulted in only 1578 total peptide IDs, whereas data from Supplemental Figure 1b was acquired with high voltage settings

and resulted in 6061 total peptide IDs. We hypothesize that this is primarily due to inefficient transport in the ion optics when voltages are lower.

Inefficient transfer seems to affect ions across the entire *m*/*z* range, but it may be more significant for larger ions of lower charge. In Figure 1a, direct infusion with low transmission voltages produces polyalanine spectra that drop off near *m*/*z* 900. In contrast, high transmission voltages show ions above *m*/*z* 1200 and notably more singly and doubly charged ions. Although the base-peak intensities are similar for each spectrum, the soft transmission acquisition required nearly twice as much time. This suggests that greater loss could occur as ions increase in mass and decrease in charge. In Figure 1b, the intensities of doubly charged A11–A24 calibrants are shown, each normalized to the most intense doubly charged calibrant in its respective high or low voltage spectrum. Although the high voltage settings yield overall intensities much higher than low voltage, the high voltage settings also have a slight bias against ions of low *m*/*z*.

On the SYNAPT G2, transmission voltages and gas flow rates can cause the collisional activation and unfolding of proteins prior to IM analysis.[23, 43] LC-HDMS^E CCS profiling only seeks information on gas-phase structural preferences and, therefore, it is more important to have sensitive transmission over soft transmission to retain structural elements from solution. However, if high transmission voltages are sufficient to change the conformation of the polyalanine ions, the drift tube calibration values may not be valid and could introduce large errors to future measurements. The measurement of peptide CCS must not cause a significant drop in accuracy under our high voltage settings.

Validation of polyalanine calibration under high transmission voltages was first tested by measuring the CCS of bradykinin (BK), a neuropeptide that has been well

studied by drift tube IM-MS. Pierson et al. have characterized two peaks for the gas-phase distribution of BK^{2+}, with the primary peak having a CCS of 246 $Å^2$.[44] A CCS for the secondary peak was not given, but it appears to be similar to the BK^{2+} CCS published by Bush et al., 237 $Å^2$.[18] For BK^{3+}, Pierson et al. have established three gas-phase quasi-equilibrium (QE) structures at 269 $Å^2$, 285 $Å^2$, and 305 $Å^2$,[44] and three similar structures were also observed by Kemper et al.[45] These values will serve as a reference to our calibrated TW IM CCS values. The acceptable error will be ≤1.8%, the average discrepancy between TW IM values and drift tube values obtained by Bush et al. for tryptic peptides.[28]

CCS distributions for BK^{2+} and BK^{3+} acquired with a 500 m s^{-1} wave velocity and 30 V wave height are shown in Figure 2, and Supplementary Table 1 lists the measured CCS at all WV and WV used. Overall, the profiles from low and high voltages are very similar. Slight differences in the secondary peaks may indicate some activation from higher voltages, though nothing much more significant than what was already present under low voltages. However, the mean calibration error under high voltage was (0.44% ± 0.03%), whereas the error for low voltage settings was (0.34% ± 0.03%). This implies that larger errors should be expected with high transmission voltages.

The mean CCS of the primary and secondary BK^{2+} peaks were (247.4 ± 0.8 $Å^2$) and (238.5 ± 0.6 $Å^2$), respectively, under low voltage settings. This corresponds to an error of 0.7% for the primary peak and 0.6% for the secondary peak. The high voltage settings yielded a primary peak at (249.8 ± 0.3 $Å^2$) and a secondary peak at (240.6 ± 0.6 $Å^2$), placing the errors of the mean at 1.6% and 1.5%. This is a noticeable increase from low voltage settings, but still below 1.8%.

High voltage settings also produce higher errors for BK^{3+}, but to a lesser magnitude. The distributions in Figure 2 represent the best resolution we achieved of the multiple peaks present. The mean CCS of the primary peak measured at low voltage was a near perfect match to Pierson et al.'s structure C at (305 ± 1 Å^2). The primary peak from the high voltage settings was (307 ± 1 Å^2), an error of only 0.7%. The respective mean CCS measured under high and low voltage settings for a second peak was (290 ± 2 Å^2) and (293 ± 2 Å^2). We hypothesize that there are several features with similar CCS in the second peak, and that the overlap leads to poor resolution. In Figure 2, the partially resolved peak in this region was measured at 287 Å^2, very similar to Pierson et al.'s structure B.[44] QE structure A was not observed at appreciable signal-to-noise ratio in any of our acquisitions. It should be noted that Pierson et al. only observed this structure at approximately 2.3% relative abundance.[44]

The BK experiments suggest that high transmission voltages will introduce larger error relative to low voltages, but that the difference is not significant. The only errors larger than 1.8% belonged to features that could not be reproducibly resolved. Perhaps the most interesting aspect of this data is the fact that the BK^{3+} distributions resemble the QE structures even though they were electrosprayed from water. In another study by Pierson et al., structures A, B, and C initially had approximate relative abundances of 39%, 70%, and 100%, respectively, when electrosprayed from 100% water.[46] They had to be deliberately activated to reach their QE abundances. Therefore, Figure 2 suggests that BK^{3+} may undergo substantial ion heating even at low voltage settings and lose evidence of its solution preferences. Using EDC, we were able to acquire a BK^{3+} t_D distribution at even lower transmission voltages (Supplementary Figure 2). The

abundance of structure B seems slightly increased relative to structure C, though structure A is still not observed.

The SYNAPT G2's low pressures prior to IM separation have been shown to lead to higher E/N values than typically seen with static-field drift tubes.[25, 43] In this investigation, SYNAPT G2 pressures dropped as low as 0.77 mTorr in the source region. In contrast, pressures on home-built instruments are commonly maintained near 3 Torr[44] and as high as 10–12 Torr.[45]

Accuracy of CCS Calibration Under High Transmission Voltages: Tryptic Peptides

The BK data represents the behavior of only one peptide. To further validate the accuracy of high voltage polyalanine calibration, we performed a tryptic digest on bovine serum albumin (BSA) with carbamidomethylated cysteines. The same peptide ions used by Bush et al. in their evaluation of polyalanine calibrants were used here.[28] It will be helpful to compare our high voltage TW IM CCS values with their published TW IM values, as it is most important that high transmission voltages not introduce significant error beyond what is already introduced by TW IM.

Table 1 lists all the BSA peptides we were able to identify during direct infusion via accurate mass matching. Many of the CCS pairs are indistinguishable within a single standard deviation. Two peptides, $GACLLPK^{2+}$ and $KVPQVSTPTLVEVSR^{3+}$, displayed anomalously high errors and standard deviations. From their CCS distributions in Figure 3, it is clear that they favor multiple conformations in the gas phase, possibly because of the presence of cis/trans proline isomerization.[47] Bush et al. mention that they also observed several BSA peptides with multiple IM features.[28] The average CCS of GACLLPK's two peaks is approximately 207 $Å^2$, which is the published TW IM value.[28]

As for KVPQVSTPTLVEVSR, we hypothesize that the poor resolution of the two more compact peaks leads to a greater variation in the centroid arrival time picked by DriftScope, producing the observed standard deviation.

When $GACLLPK^{2+}$ and $KVPQVSTPTLVEVSR^{3+}$ are omitted, the average difference between published TW IM values and our TW IM values are approximately 0.4% under both low and high voltage settings. When compared with CCS obtained on the modified drift tube SYNAPT G1, the average errors are approximately 1.4% and 1.7% for proteomic voltages and soft voltages, respectively.

The combined results of the BK and BSA data lead to the confident assertion that peptide CCS values can be calibrated and measured with higher transmission voltages. The high voltage settings will greatly increase the sensitivity of analysis without introducing significant error, despite evidence of minor calibrant and analyte activation under high voltages. We cannot completely discount the possibility that activation is greater than we suspect, but error in calculated CCS is reduced due to analyte and calibrant activation to a similar extent. This would be a less desirable situation, as the CCS values used for calibration would no longer be accurate. Analyte ions that are not similarly activated would contain large errors in their calculated CCS values, whereas the similarly activated ions would contain undetectable inaccuracies.

Potential Complications of Using On-Line LC During CCS Measurement

Structural elements from solution have previously been observed to persist in the IM distributions of small proteins and peptides.[46, 48] Our method relies on LC-$HDMS^E$ to aid in measuring CCS values of peptides in the gas phase from complex samples, and it is likely that structures from LC gradient solutions and non-native structures mediated by

the column's stationary phase would be of little interest, if not an outright interference. Their presence could affect the run-to-run or gradient-to-gradient reproducibility of CCS measurement. However, we have evidence that the high voltage settings may lead to peptides losing their "native" structures from solution and adopting gas-phase preferences, even without deliberate collisional activation.

In Figure 4, the mass spectra of BK^{3+} and BK^{2+} after quadrupole isolation with an approximate *m*/*z* window of 3 Th is shown. The trap collision energy was set to the minimum value required for transmission. Activation is immediately apparent for the BK^{3+} acquired under voltage settings by the highly abundant sequence-specific fragments. The BK^{2+} spectrum contains a few very low abundance fragments. Low voltages did not completely stop BK^{3+} fragmentation, but did greatly decrease it. BK^{2+} spectra did not contain any detectable fragments. With this amount of ion activation, it is not surprising that the BK CCS distributions from Figure 2 resembled the QE gas-phase distribution instead of the distribution Pierson et al. observed when electrosprayed from 100% water.[46]

We hypothesized that most peptide ions will behave similarly to BK in this respect. It is difficult to directly test this without multiple mobility cells, so we instead opted to measure the CCS of BSA peptides identified by LC-$HDMS^E$ and compare them to the direct infusion results. All BSA direct infusion samples were electrosprayed in 50% ACN, 1% HAc, whereas peptides from LC-$HDMS^E$ were sprayed in solutions ranging from 5% to 30% ACN, 0.1% FA.

The results are summarized in Supplemental Table 2. Of the 11 direct infusion peptides also identified by $HDMS^E$, the average deviation from the mean direct infusion

CCS value was 0.9%, with no deviation being larger than 1.4%. HDMS^E acquisitions were only performed at 800 m s^{-1} WV and 40 V WH, values that were optimized for number of identifications. Using a single wave velocity/height ratio may be responsible for observed error with such a small standard deviation. However, an average discrepancy of 0.9% is of low significance. We cannot discount the possibility of solution structure elements appearing in HDMS^E CCS of other peptides, but they will likely not be a common occurrence under high transmission voltages.

Large-Scale CCS Profiling

Following the validation of accurate CCS measurement with high transmission voltages, we aimed to demonstrate the utility of bulk structural comparisons in a large-scale LC-HDMS^E CCS profiling experiment. Proteolytic digests of total protein from cell lysates can produce tens of thousands of peptides with a common N- or C-terminal residue. Digest peptides from protein mixtures have been used in many previous CCS databases but, to our best knowledge, the mixture has never been as large or as complex as a total protein extract.

In the initial experiment, extract from human K562 cells was digested by trypsin, an enzyme that cleaves proteins into peptides with a C-terminal lysine or arginine residue. After duplicate LC-HDMS^E acquisitions, each dataset was searched on PLGS and the Final Peptide csv was submitted to pepCCScal to measure the CCS of identified sequences. To our knowledge, there is no other software that can couple CCS measurement to protein/peptide database search output. Figure 5 shows a plot of CCS versus *m*/*z* for 2814 doubly protonated peptides and 1172 triply protonated peptides, all of which passed the strict curation criteria listed in the Experimental section. Initially, 5511

2+ and 2954 3+ peptides were identified. The majority of deletions from curation was from our insistence to use only first-pass and miscleavage identifications, and necessitating that they should be identified in both runs. The 2+ family exhibits a single, linear, relatively narrow distribution, which is in strong agreement with several previous studies of doubly charged tryptic peptides.[29-30, 39, 49] The 3+ family has been less characterized in literature, but is known to possess a wider range of CCS values than the 2+ family.[49] Our data is in agreement, but we note an interesting phenomenon that occurs as the 3+ family *m*/*z* increases. At roughly *m*/*z* 650, the 3+ family begins to cluster into two distinct sub-families, one compact and one highly extended. In contrast, the 2+ family appears to display a narrow band in its CCS distribution at the same *m*/*z* range. Counterman and Clemmer have previously observed two sub-families of triply protonated polyalanines, the compact sub-family attributed to hinged helix coils (HHC) and the other sub-family attributed to extended helices (EC).[50-51]

For the next experiment, we aimed to determine if the N- and C-terminal amino acids could affect the clustering of ions into a particular sub-family. Total protein extracts from yeast (*Saccharomyces cerevisiae*) were subjected to Lys-N or Lys-C digestion, which created peptides with N-terminal or C-terminal lysines, respectively. Since tryptic digests often produce the most peptide and protein IDs in shotgun proteomics,[52] we performed off-line SCX-HPLC to separate peptides into eight fractions and performed two HDMS^E analyses on each. Although this method greatly increased our analysis time, it also increased our chances of obtaining a number of CCS measurements comparable to the trypsin experiments.

The CCS versus *m*/*z* plot in Figure 6a contains 1463 doubly charged and 1007 triply charged LysN peptides. Figure 6b contains 1295 doubly charged and 1072 triply charged LysC peptides. Interestingly, both sets resulted in the same two sub-families for triply charged ions, but the LysN peptides displayed a stronger preference for the compact sub-family. As expected, the LysC sub-family split was very similar to that seen from tryptic peptides. We would like to note that miscleavages are included in these plots to achieve a higher sample population. We initially thought internal lysines may be a significant factor in sub-family clustering, but the miscleavages seemed to make no difference to the overall distribution. Complementary plots for Figure 5, Figure 6a and 6b that omit miscleavages can be found in Supplemental Figure 3.

The visual differences between Figure 6a and 6b are obvious, but deeper search of the data provided quantitative evidence of the compact preferences for peptides with an N-terminal lysine. Fourteen internal peptide sequences are listed in Figure 6c, each belonging to the same protein and found in both LysN and LysC curated peptide lists. In the protein, these sequences contain a lysine at each end leading to peptides that only differ by terminal lysine position when cleaved. In all but two pairs, the LysN partner had a significantly smaller CCS. Finally, Figure 6d shows that triply charged LysC peptides are consistently larger than LysN at all *m*/*z*.

Lacking MD simulations to properly interpret the CCS data, we can only make conjectures as to why LysN peptides tend to populate the compact sub-family. The result would appear to fall in line with earlier IM-MS studies by Hudgins et al. on the gas-phase helix preferences of protonated polypeptides.[53-54] They showed that a positive charge located at the C-terminus will interact favorably with a helix macrodipole and stabilize

extended helical ion structures. Conversely, a positive charge at the N-terminus will destabilize a helix and could shift preference to a globular structure with intramolecular solvation of the charge. Tao et al. showed that a helical region of singly charged tryptic peptide would adopt a turn motif when HK is added to the beginning of the sequence.[35] It is possible that the charge–macrodipole interaction is also significant for triply charged ions. N-terminal lysines would certainly increase the likelihood of a charge being located at the N-terminus. However, the widely varied internal sequences from our proteomic samples introduce a much more complicated landscape than what was previously characterized by Hudgins et al.

The triply protonated polyalanine studies by Counterman and Clemmer show that transitions between HHC and EC structures are temperature-dependent, with EC being favored as temperature increases.[51] Higher temperatures may facilitate proton transfers to shift the net charge from N-terminus to C-terminus, resulting in EC formation. All of our peptide ions were formed at the same source temperature of 70 ºC, and so the presence of compact and extended structures may point to the role of proton sequestration. For example, Figure 6c's RAIILERNAAYQ internal sequence may strongly sequester the charge at the N-terminus when an N-terminal lysine is followed by arginine, even at higher temperatures. When the lysine is at the C-terminus, the third proton may prefer that position over an N-terminal amine to reduce Coulombic repulsion, leading to the large difference between the LysN and LysC CCS. In contrast, the internal sequence LSGVTLSELLR might strongly sequester a proton with its arginine and keep a positive charge at the C-terminus regardless of lysine position.

Conclusions

This investigation has built upon the polyalanine calibration method detailed by Bush et al.[28] and applied it to peptide CCS database creation from large, complex populations. After several levels of validation, we have found that high transmission voltages and electrospray from LC gradients do not introduce significant error to polyalanine calibration and CCS measurement. We also showed the utility of a large-scale CCS profiling study, the results of which are currently under investigation with MD simulations.

As in any method, there are drawbacks to its use. Some are unavoidable and some could potentially be mitigated. First, TW IM calibration unquestionably introduces error to CCS measurements, even under the very best calibration schemes. We showed that high transmission voltages cause insignificant error for the peptides in this study, but we cannot assert this will be true for all peptides. Some peptides may have distinct theoretical conformations whose difference in CCS is only slightly larger than the absolute error under low voltage settings. The usually insignificant additional error from high voltage settings could then render the two theoretical CCS experimentally indistinguishable. Second, all peptide CCS lists have an associated identification FDR from database searching, meaning there are necessarily incorrect entries in the CCS database. CCS calculation may have its own FDR that is independent of the PLGS identification FDR, though we assume this would be low given our curation criteria and the very small run-to-run CCS deviations. The creation and validation of a CCS FDR may be very helpful for future large-scale studies. Finally, HDMS^E only outputs a single drift time for identified ions, making it impossible to identify multiple IM structures. An alternative approach that

extracts the complete IM drift profile at the apex LC retention times of identified peptides could mitigate this issue.

As we stated previously, our original motivation for calibration under high transmission voltages was to perform high-throughput concurrent HDMS^E analysis and CCS calibration with the lockspray. In the current protocol, calibration at various WV/WH ratios must be performed separately. If polyalanine was instead used as a lockspray, CCS calibration could be performed simultaneously with LC-HDMS^E. The SYNAPT G2 does not record mobility data for the lockspray ions, and so we are currently working on acquisition modifications to facilitate this. Future work will also include building large-scale CCS databases for endogenous signaling peptides and searching for bulk structural patterns intrinsic to peptide families, followed by correlating the structural information to their binding behavior and physiological function.

Acknowledgments

The authors acknowledge support for this work in part by the National Science Foundation grant (CHE-0957784) and the National Institutes of Health grants (1R01DK071801 and 1R56DK071801). C.L. acknowledges an NIH-supported Chemistry Biology Interface Training Program Predoctoral Fellowship (grant number T32-GM008505) and an NSF Graduate Research Fellowship (DGE-1256259). L.L. acknowledges an H. I. Romnes Faculty Research Fellowship. The authors also acknowledge the generous gifts of yeast and human protein extracts from Promega, as well as Thermo Scientific Pierce LysN Protease, MS grade, from Thermo Fisher.

References

(1) Revercomb, H. E., and Mason, E. A. *Anal. Chem.* **1975** *47*, 970-983.

(2) McDaniel, E. W., Martin, D. W., and Barnes, W. S. *Rev. Sci. Instrum.* **1962** *33*, 2-7.

(3) K. B. McAfee, J., and Edelson, D. *Proc. Phys. Soc.* **1963** *81*, 382.

(4) von Helden, G., Wyttenbach, T., and Bowers, M. T. *Science* **1995** *267*, 1483-1485.

(5) Clemmer, D. E., Hudgins, R. R., and Jarrold, M. F. *J. Am. Chem. Soc.* **1995** *117*, 10141-10142.

(6) Hoaglund, C. S., Valentine, S. J., Sporleder, C. R., Reilly, J. P., and Clemmer, D. E. *Anal. Chem.* **1998** *70*, 2236-2242.

(7) Bohrer, B. C., Merenbloom, S. I., Koeniger, S. L., Hilderbrand, A. E., and Clemmer, D. E. *Annu. Rev. Anal. Chem.* **2008** *1*, 293-327.

(8) McLean, J. A., Ruotolo, B. T., Gillig, K. J., and Russell, D. H. *Int. J. Mass Spectrom.* **2005** *240*, 301-315

9) Kiss, A., and Heeren, R. A. *Anal. Bioanal. Chem.* **2011** *399*, 2623-2634.

(10) Tang, K., Shvartsburg, A. A., Lee, H.-N., Prior, D. C., Buschbach, M. A., Li, F. *et al. Anal. Chem.* **2005** *77*, 3330-3339.

(11) Glaskin, R. S., Ewing, M. A., and Clemmer, D. E. *Anal. Chem.* **2013** *85*, 7003-7008.

(12) Shvartsburg, A., Seim, T., Danielson, W., Norheim, R., Moore, R., Anderson, G. *et al. J. Am. Soc. Mass Spectrom.* **2013** *24*, 109-114.

(13) Wyttenbach, T., and Bowers, M. (2003) Gas-Phase Conformations: The Ion Mobility/Ion Chromatography Method, In *Modern Mass Spectrometry* (Schalley, C., Ed.), pp 207-232, Springer Berlin Heidelberg.

(14) Clemmer, D. E., and Jarrold, M. F. *J. Mass Spectrom.* **1997** *32*, 577-592.

(15) Tao, L., Dahl, D., Pérez, L., and Russell, D. *J. Am. Soc. Mass Spectrom.* **2009** *20*, 1593-1602.

(16) Mesleh, M. F., Hunter, J. M., Shvartsburg, A. A., Schatz, G. C., and Jarrold, M. F. *J. Phys. Chem.* **1996** *100*, 16082-16086.

(17) Ruotolo, B. T., Benesch, J. L. P., Sandercock, A. M., Hyung, S.-J., and Robinson, C. V. *Nat. Protoc.* **2008** *3*, 1139-1152.

(18) Bush, M. F., Hall, Z., Giles, K., Hoyes, J., Robinson, C. V., and Ruotolo, B. T. *Anal. Chem.* **2010** *82*, 9557-9565.

(19) Benesch, J. L. P., and Ruotolo, B. T. *Curr. Opin. Struct. Biol.* **2011** *21*, 641-649.

(20) Zhong, Y., Hyung, S.-J., and Ruotolo, B. T. *Expert Rev. Proteomics* **2012** *9*, 47-58.

(21) Pringle, S. D., Giles, K., Wildgoose, J. L., Williams, J. P., Slade, S. E., Thalassinos, K. *et al. Int. J. Mass Spectrom.* **2007** *261*, 1-12.

(22) Wallace, A. *Am. Lab.* **2010** *42*, 13-17.

(23) Zhong, Y., Hyung, S.-J., and Ruotolo, B. T. *Analyst* **2011** *136*, 3534-3541.

(24) Giles, K., Pringle, S. D., Worthington, K. R., Little, D., Wildgoose, J. L., and Bateman, R. H. *Rapid Commun. Mass Spectrom.* **2004** *18*, 2401-2414.

(25) Shvartsburg, A. A., and Smith, R. D. *Anal. Chem.* **2008** *80*, 9689-9699.

(26) Smith, D., Knapman, T., Campuzano, I., Malham, R., Berryman, J., Radford, S. *et al. Eur. J. Mass Spectrom.* **2009** *15*, 113-130.

(27) Thalassinos, K., Grabenauer, M., Slade, S. E., Hilton, G. R., Bowers, M. T., and Scrivens, J. H. *Anal. Chem.* **2009** *81*, 248-254.

(28) Bush, M. F., Campuzano, I. D. G., and Robinson, C. V. *Anal. Chem.* **2012** *84*, 7124-7130.

(29) Valentine, S. J., Counterman, A. E., Hoaglund-Hyzer, C. S., and Clemmer, D. E. *J. Phys. Chem. B* **1999** *103*, 1203-1207.

(30) Valentine, S., Counterman, A., and Clemmer, D. *J. Am. Soc. Mass Spectrom.* **1999** *10*, 1188-1211.

(31) Henderson, S. C., Li, J., Counterman, A. E., and Clemmer, D. E. *J. Phys. Chem. B* **1999** *103*, 8780-8785.

(32) Valentine, S. J., Ewing, M. A., Dilger, J. M., Glover, M. S., Geromanos, S., Hughes, C. *et al. J. Proteome Res.* **2011** *10*, 2318-2329.

(33) Dilger, J. M., Valentine, S. J., Glover, M. S., Ewing, M. A., and Clemmer, D. E. *Int. J. Mass Spectrom.* **2012** *330–332*, 35-45.

(34) Dilger, J., Valentine, S., Glover, M., and Clemmer, D. *J. Am. Soc. Mass Spectrom.* **2013** *24*, 768-779.

(35) Tao, L., McLean, J., McLean, J., and Russell, D. *J. Am. Soc. Mass Spectrom.* **2007** *18*, 1232-1238.

(36) Ruotolo, B. T., Verbeck, Thomson, L. M., Woods, A. S., Gillig, K. J., and Russell, D. H. *J. Proteome Res.* **2002** *1*, 303-306.

(37) Ruotolo, B. T., Gillig, K. J., Woods, A. S., Egan, T. F., Ugarov, M. V., Schultz, J. A. *et al. Anal. Chem.* **2004** *76*, 6727-6733.

(38) McLean, J. R., McLean, J. A., Wu, Z., Becker, C., Pérez, L. M., Pace, C. N. *et al. J. Phys. Chem. B* **2010** *114*, 809-816.

(39) Valentine, S. J., Kulchania, M., Barnes, C. A. S., and Clemmer, D. E. *Int. J. Mass Spectrom.* **2001** *212*, 97-109.

(40) Silva, J. C., Denny, R., Dorschel, C., Gorenstein, M. V., Li, G.-Z., Richardson, K. *et al. Mol. Cell. Proteomics* **2006** *5*, 589-607.

(41) Bond, N. J., Shliaha, P. V., Lilley, K. S., and Gatto, L. *J. Proteome Res.* **2013** *12*, 2340-2353.

(42) Li, J., Taraszka, J. A., Counterman, A. E., and Clemmer, D. E. *Int. J. Mass Spectrom.* **1999** *185–187*, 37-47.

(43) Merenbloom, S., Flick, T., and Williams, E. *J. Am. Soc. Mass Spectrom.* **2012** *23*, 553-562.

(44) Pierson, N. A., Valentine, S. J., and Clemmer, D. E. *J. Phys. Chem. B* **2010** *114*, 7777-7783.

(45) Kemper, P. R., Dupuis, N. F., and Bowers, M. T. *Int. J. Mass Spectrom.* **2009** *287*, 46-57.

(46) Pierson, N. A., Chen, L., Valentine, S. J., Russell, D. H., and Clemmer, D. E. *J. Am. Chem. Soc.* **2011** *133*, 13810-13813.

(47) Counterman, A. E., and Clemmer, D. E. *Anal. Chem.* **2002** *74*, 1946-1951.

(48) Wyttenbach, T., and Bowers, M. T. *J. Phys. Chem. B* **2011** *115*, 12266-12275.

(49) Taraszka, J. A., Counterman, A. E., and Clemmer, D. E. *Fresenius J. Anal. Chem.* **2001** *369*, 234-245.

(50) Counterman, A. E., and Clemmer, D. E. *J. Am. Chem. Soc.* **2001** *123*, 1490-1498.

(51) Counterman, A. E., and Clemmer, D. E. *J. Phys. Chem. B* **2003** *107*, 2111-2117.

(52) Swaney, D. L., Wenger, C. D., and Coon, J. J. *J. Proteome Res.* **2010** *9*, 1323-1329.

(53) Hudgins, R. R., Ratner, M. A., and Jarrold, M. F. *J. Am. Chem. Soc.* **1998** *120*, 12974-12975.

(54) Hudgins, R. R., Mao, Y., Ratner, M. A., and Jarrold, M. F. *Biophys. J.* **1999** *76*, 1591-1597.

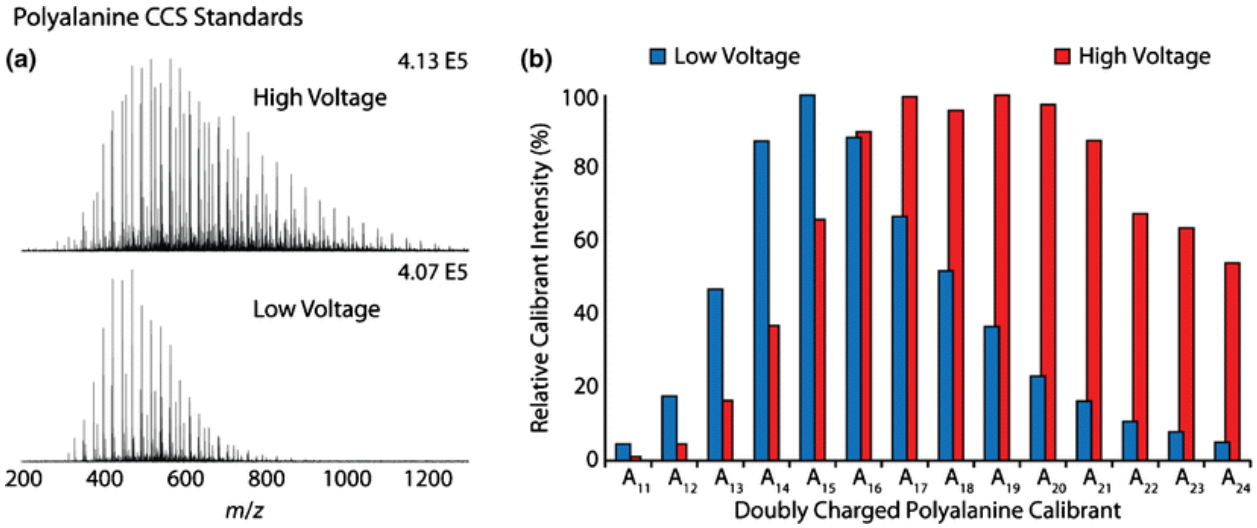


Figure 1. (a) Mass spectra of DL-polyalanine directly infused under high and low transmission voltages. (b) This bar graph displays the relative intensities of doubly charged A11– A24 CCS calibrants. The intensities are normalized to the most abundant doubly charged calibrant in the high and low voltage mass spectrum, respectively.

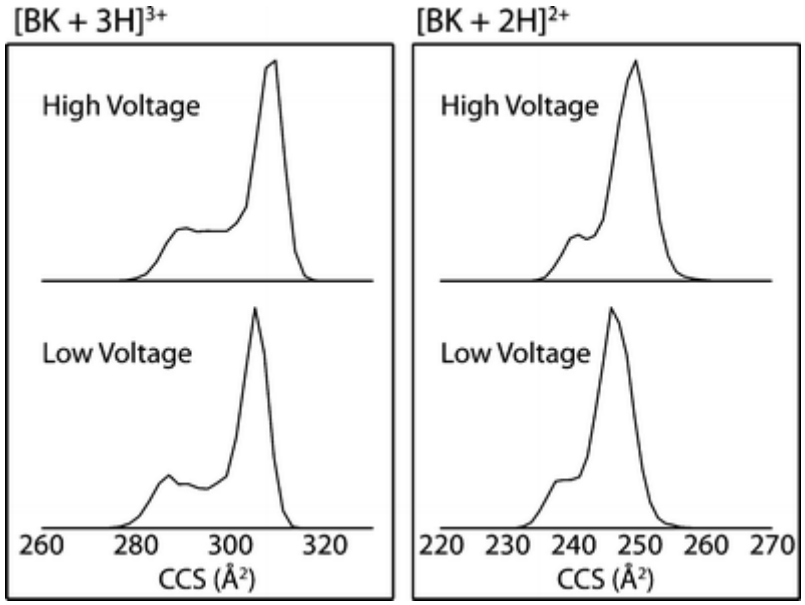


Figure 2. The CCS distributions of doubly and triply protonated bradykinin were measured under high and low transmission voltages. The wave velocity and wave height used for these particular acquisitions was 500 m s^{-1} and 30 V, respectively.

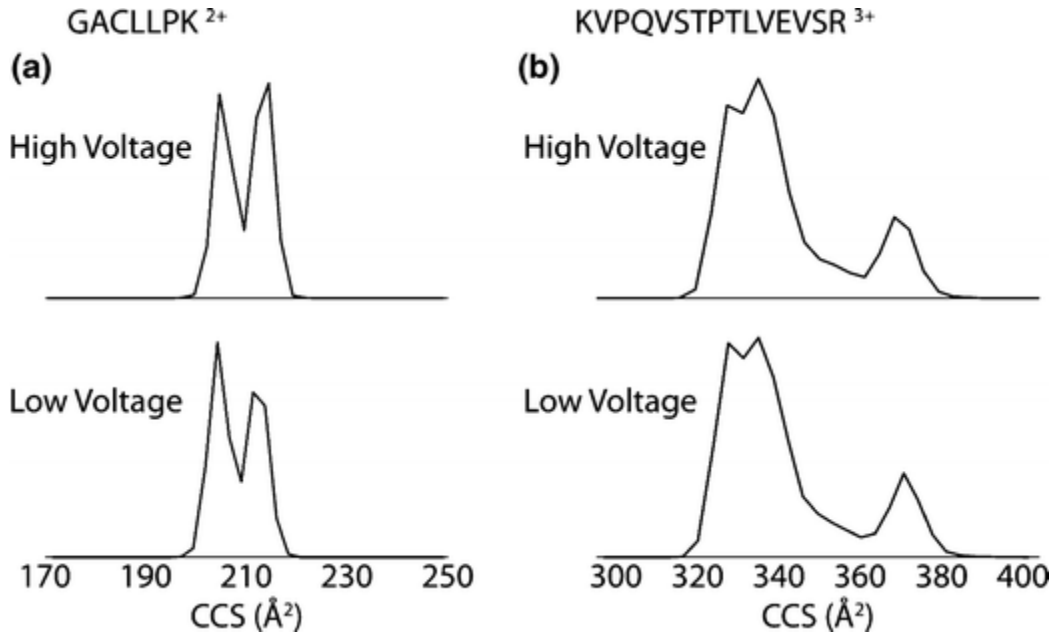


Figure 3. The CCS distributions under high and low transmission voltages for (a) GACLLPK^{2+} and (b) KVPQVSTPTLVEVSR^{3+} reproducibly show multiple features.

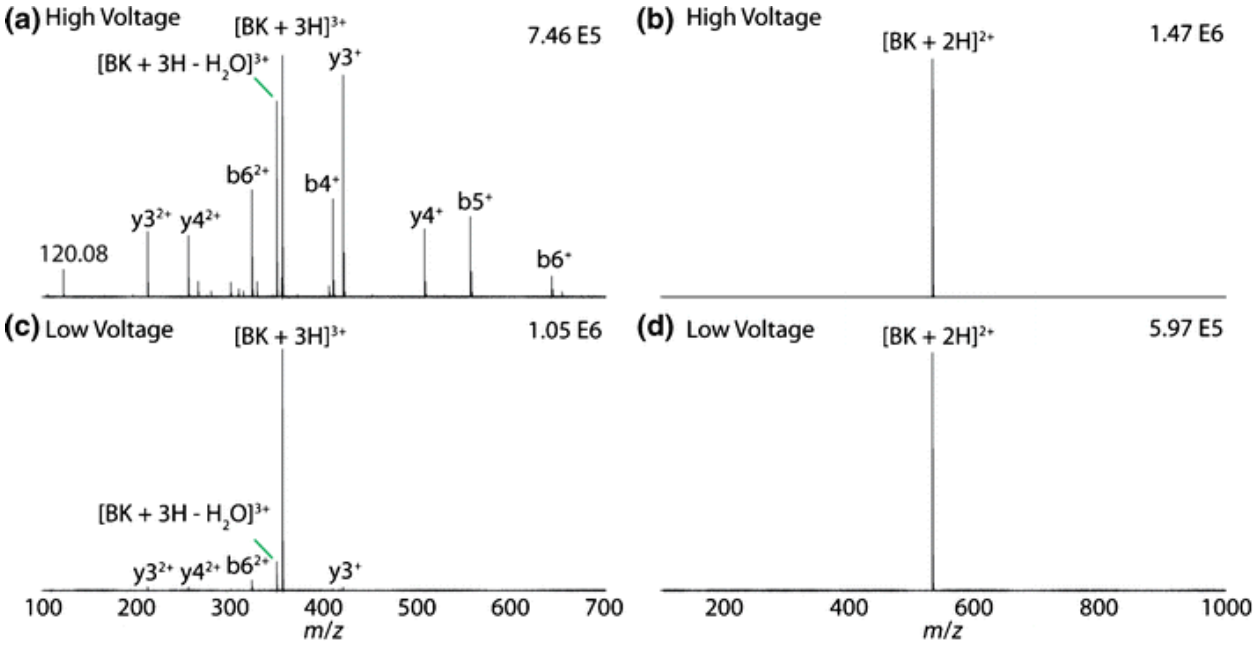


Figure 4. The quadrupole-isolated mass spectra under high transmission voltages for (a) BK^{3+} and (b) BK^{2+} show various levels of fragmentation. The low transmission voltage spectra for (c) BK^{3+} and (d) BK^{2+} show less fragmentation.

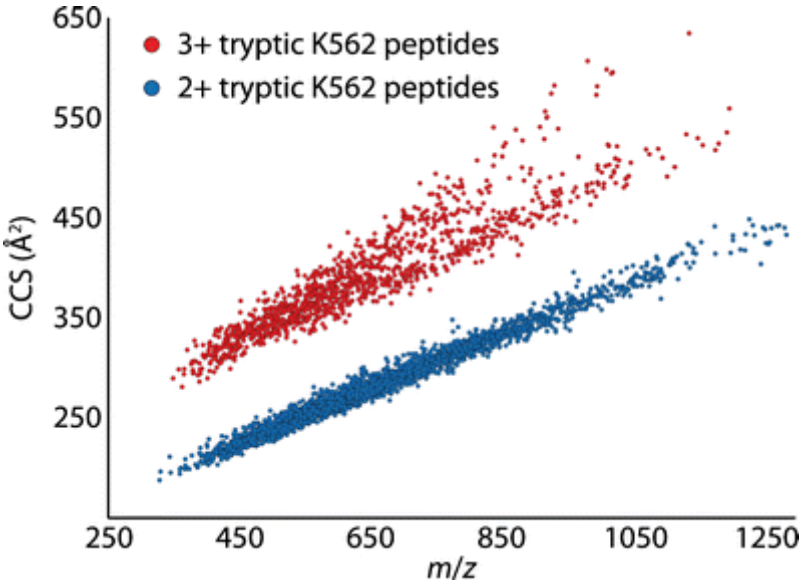
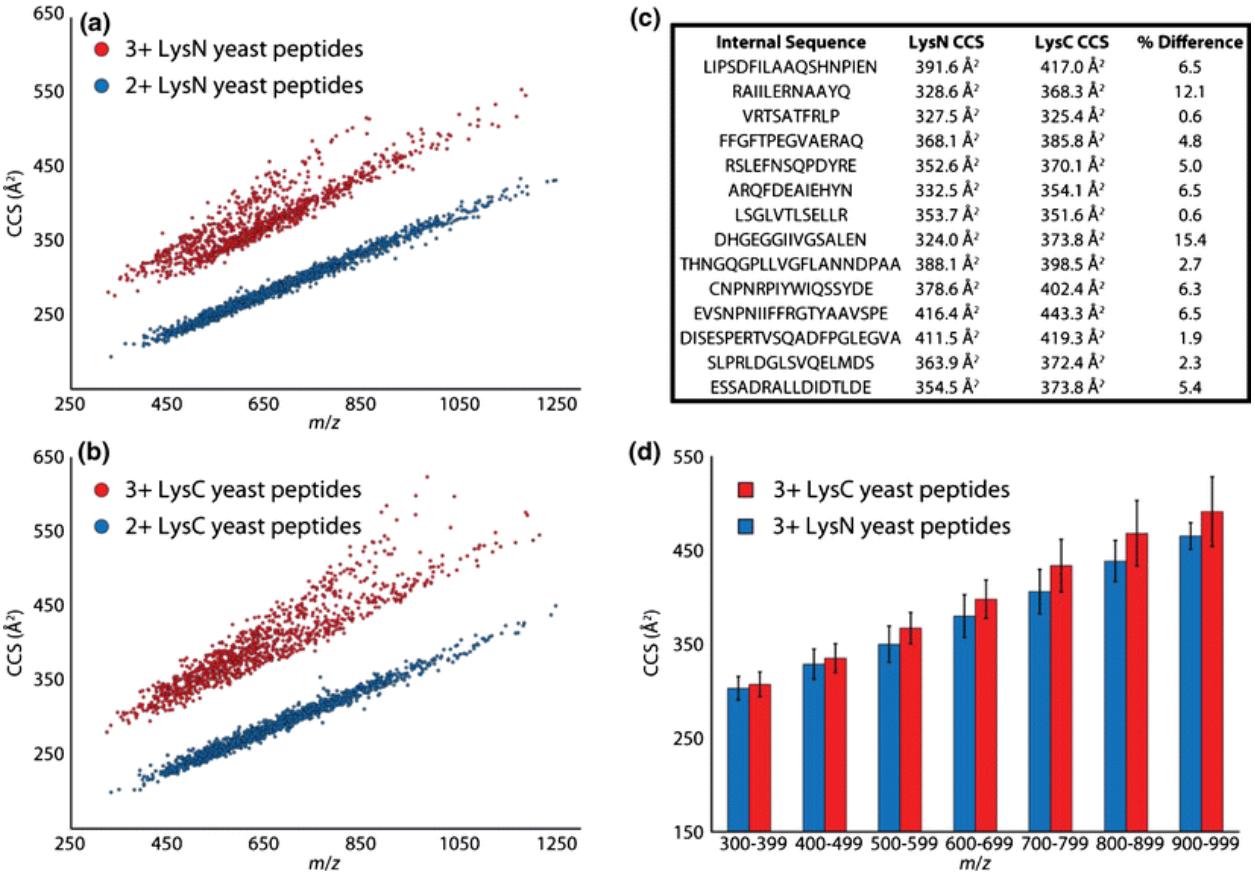


Figure 5. The doubly and triply charged tryptic peptides from LC-HDMS^{E} analysis of K562 total protein extract are plotted CCS versus *m*/*z*.



Internal Sequence	LysN CCS	LysC CCS	% Difference
LIPSDFILAAQSHNPIEN	391.6 Å²	417.0 Å²	6.5
RAIILERNAAYQ	328.6 Å²	368.3 Å²	12.1
VRTSATFRLP	327.5 Å²	325.4 Å²	0.6
FFGFTPEGVAERAQ	368.1 Å²	385.8 Å²	4.8
RSLEFNSQPDYRE	352.6 Å²	370.1 Å²	5.0
ARQFDEAIEHYN	332.5 Å²	354.1 Å²	6.5
LSGLVTLSELLR	353.7 Å²	351.6 Å²	0.6
DHGEGGIIVGSALEN	324.0 Å²	373.8 Å²	15.4
THNGQGPLLVGFLANNDPAA	388.1 Å²	398.5 Å²	2.7
CNPNRPIYWIQSSYDE	378.6 Å²	402.4 Å²	6.3
EVSNPNIIFFRGTYAAVSPE	416.4 Å²	443.3 Å²	6.5
DISESPERTVSQADFPGLEGVA	411.5 Å²	419.3 Å²	1.9
SLPRLDGLSVQELMDS	363.9 Å²	372.4 Å²	2.3
ESSADRALLDIDTLDE	354.5 Å²	373.8 Å²	5.4

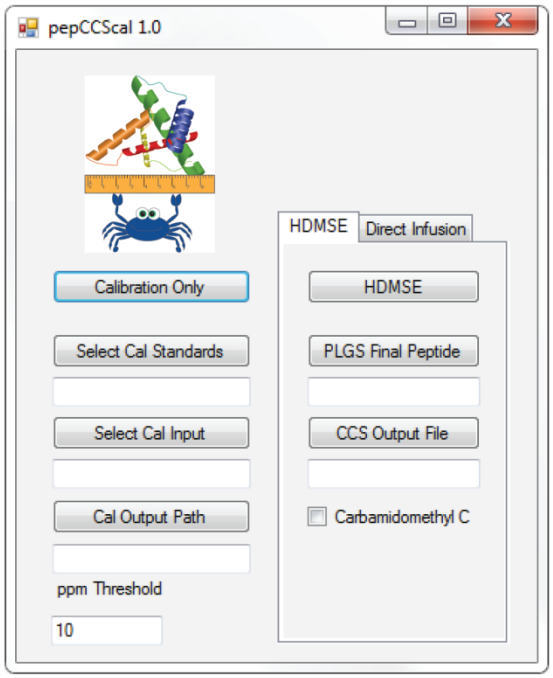
Figure 6. The CCS of (a) LysN and (b) LysC yeast protein digests are plotted against *m*/*z*. (c) Fourteen pairs of peptides with identical internal sequences were reproducibly identified in both LysN and LysC analyses, and their CCS values were compared. (d) Identified peptides were grouped into 100 Th bins and the mean CCS of each bin from LysC peptide was compared with the mean CCS from each bin of LysN peptides.

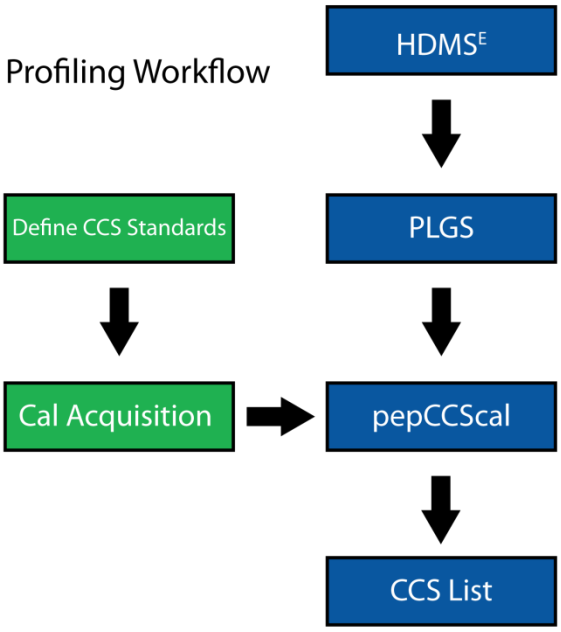
Table 1. CCS of Tryptic BSA Peptides Under High and Low Transmission Voltages

			High Voltage		Low Voltage	
Sequence	**Charge**	**Lit. TW CCS (Å^2) [1]**	**CCS (Å^2)**	**Error of the Mean (%)**	**CCS (Å^2)**	**Error of the Mean (%)**
GACLLPK*	2	207	203.1 ± 0.8	1.88	203.0 ± 0.6	1.93
GACLLPK*	2	207	212.3 ± 0.7	2.56	211.8 ± 0.9	2.32
LVTDLTK	2	205	206.9 ± 0.5	0.93	206.5 ± 0.7	0.73
ATEEQLK	2	209	209.9 ± 0.5	0.43	209.6 ± 0.6	0.29
LCVLHEK	2	228	227 ± 1	0.44	227.6 ± 0.6	0.18
AEFVEVTK	2	219	220.2 ± 0.4	0.55	219.8 ± 0.8	0.37
YLYEIAR	2	237	238.3 ± 0.5	0.55	237.8 ± 0.7	0.34
LVVSTQTALA	2	234	235.8 ± 0.7	0.77	235.6 ± 0.3	0.68
QTALVELLK	2	245	245.3 ± 0.6	0.12	245.0 ± 0.7	0.00
LVNELTEFAK	2	262	262.2 ± 0.6	0.08	261.8 ± 0.2	0.08
HLVDEPQNLIK	2	288	286.7 ± 0.6	0.45	286.8 ± 0.3	0.42
TVMENFVAFVDK	2	297	295 ± 1	0.67	295.3 ± 0.5	0.57
SLHTLFGDELCK	2	297	296.0 ± 0.7	0.34	296.0 ± 0.2	0.34
YICDNQDTISSK	2	298	295.8 ± 0.9	0.74	296.2 ±0.4	0.60
LGEYGFQNALIVR	2	307	306.1 ± 0.6	0.29	305.8 ± 0.4	0.39
DDPHACYSTVFDK	2	313	312.2 ± 0.9	0.26	311.7 ± 0.6	0.42
MPCTEDYLSLILNR	2	338	337.8 ± 0.6	0.06	336.9 ± 0.7	0.34
YNGVFQECCQAEDK	2	325	324.0 ± 0.4	0.31	323.5 ± 0.6	0.46
HLVDEPQNLIK	3	319	321.2 ± 0.5	0.69	320.6 ± 1.3	0.50
SLHTLFGDELCK	3	334	334.3 ± 0.6	0.09	333.7 ± 0.8	0.09
TCVADESHAGCEK	3	340	342 ± 1	0.59	341 ± 1	0.29
DDPHACYSTVFDK	3	323	324.8 ± 0.6	0.56	324 ± 1	0.31
KVPQVSTPTLVEVSR*	3	338	341 ± 3	0.89	339 ± 4	0.30
DAIPENLPPLTADFAEDKDVCK	3	422	422 ± 2	0.00	421 ± 1	0.24
Mean (excluding *)				**0.42**		**0.36**

Supplemental Information

Supplemental Information 1.





All CCS calibrations and measurements were performed, in part, by our open-source software pepCCScal. All calculations are based off methods described previously.[17, 26] Briefly, an empirically modified version of the Mason-Schamp equation is used:

$$\Omega = \frac{ze}{16}\left[\frac{18\pi}{K_b T}\left(\frac{1}{m}+\frac{1}{M}\right)\right]^{1/2}\frac{760}{P}\frac{T}{273.15}\frac{1}{N}At_D{}^B \quad (1)$$

The variable Ω is the CCS, z is the ion charge state, e is the fundamental charge, K_b is Boltzmann's constant, T is temperature, m is the mass of the analyte, M is the mass of the drift gas, P is the pressure, N is the buffer gas number density, and t_D is the ion's drift time or arrival time. The constants A and B result from the non-uniformity of the electric field and must be determined by calibration. The CCS can be normalized by reduced mass and charge state (Ω'), and all other variables can be collapsed into a single term (A') for the following reduced equation:

$$\Omega' = A' {t_D}^B \quad (2)$$

The user defines the calibration set with a 3-column csv file. Column A is the calibrant peptide sequence, column B is the calibrant charge, and column C is the calibrant CCS. This file is selected by the “Select Cal Standards” button. After IM-MS acquisition of the calibrants, the file is opened in DriftScope and “Detect Peaks” is selected from the Peak Detection menu. The non-chromatographic, lockspray-corrected peaklist is then created and exported as a csv. This file is selected from the “Select Cal Input” button. To create the calibration, pepCCScal calculates the monoisotpic *m*/*z* of the user-defined calibrants and searches the lockspray-corrected *m*/*z* from the DriftScope peak list csv (column C). If a match within the user-defined ppm threshold is found, pepCCScal assigns the corrected and centroided drift time (**t_D**) (column H) to that calibrant.

When the all observed calibrants are found, pepCCScal begins the two-part calculation. First, it calculates the slope of a best-fit linear regression for ln (Ω’) versus ln (t_D), with Ω’ being the calibrant CCS normalized for mass and charge. The slope of this log plot is the B term from equations 1 and 2. Next, doubly corrected t_D are calculated for each calibrant:

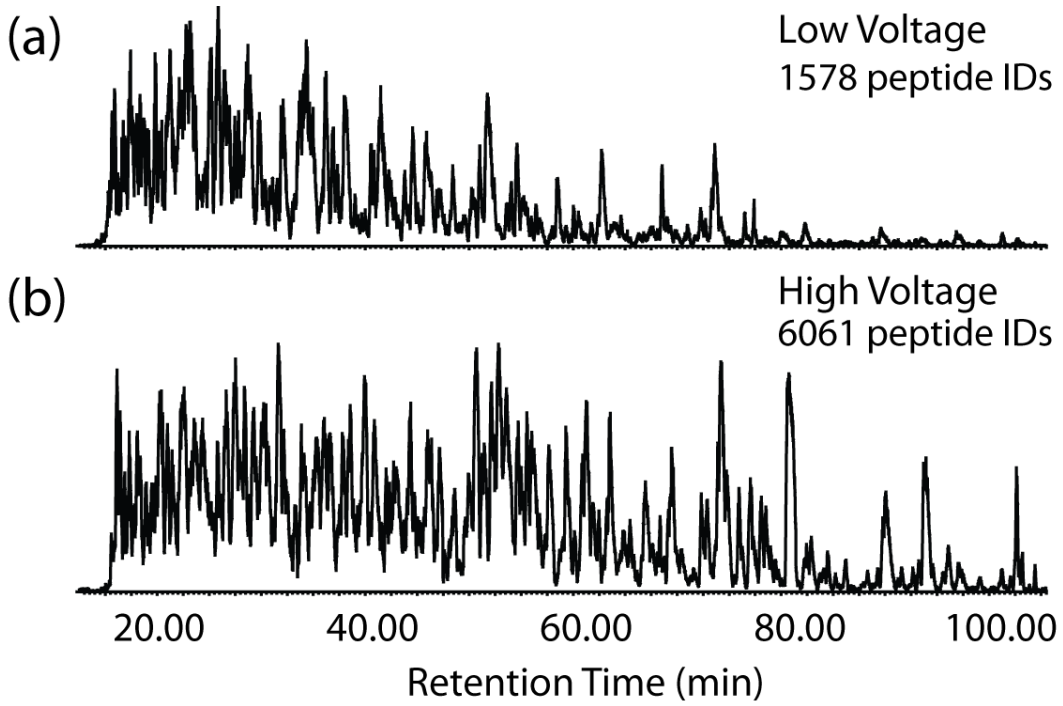
$$t_D'' = z\left(\frac{1}{m} + \frac{1}{M}\right)^{1/2} {t_D}^B \quad (3)$$

The calibration is completed by calculating the slope and y-intersection of a best-fit linear regression for Ω versus t_D’’. The mean error, RMSE, and R2 for the calibration are provided in the output file. This equation is used to calculate unknown CCS measured at the same wave height and wave velocity.

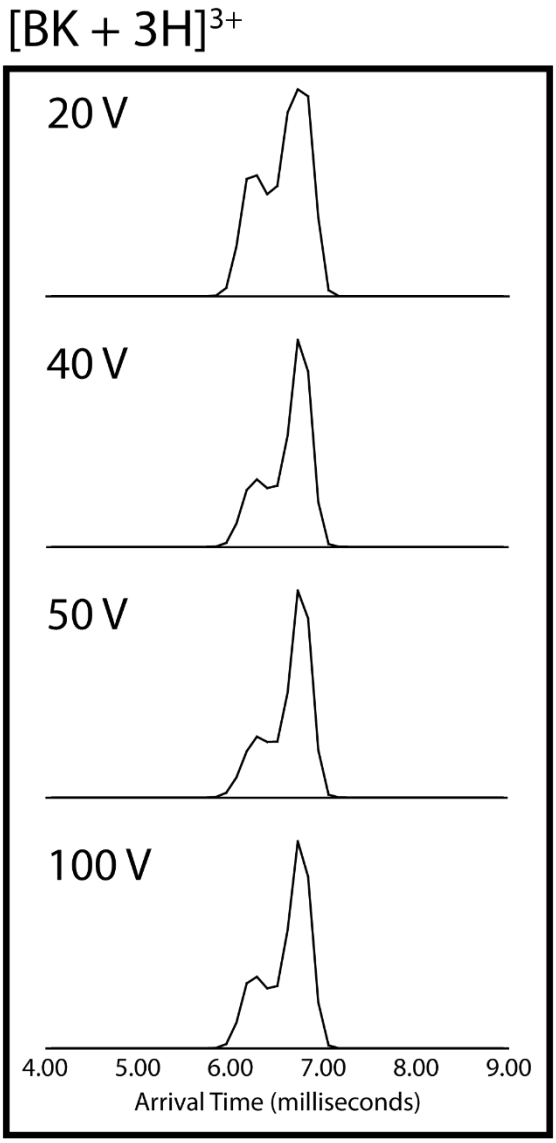
For CCS measurement in direct infusion experiments, the analyte is acquired after calibration. This acquisition is also opened in DriftScope to create a lock-spray corrected,

non-chromatographic peak list which is exported as a csv. The pepCCScal program can search for ions based on a list of peptide sequences and charges, or the user can manually enter mass and charge values to look for. The best-fit equation from the calibration output csv is used to calculate the CCS of these ions. For HDMSE CCS profiling, DriftScope is not required to make a peak list. The Final Peptide csv from PLGS output contains the centroid t_D of all identified peptides. However, this csv contains t_D in bins, not milliseconds, and so polyalanine calibration must also be done in bins. The pepCCScal program uses the t_D from the Final Peptide csv (column BE) to calculate CCS.

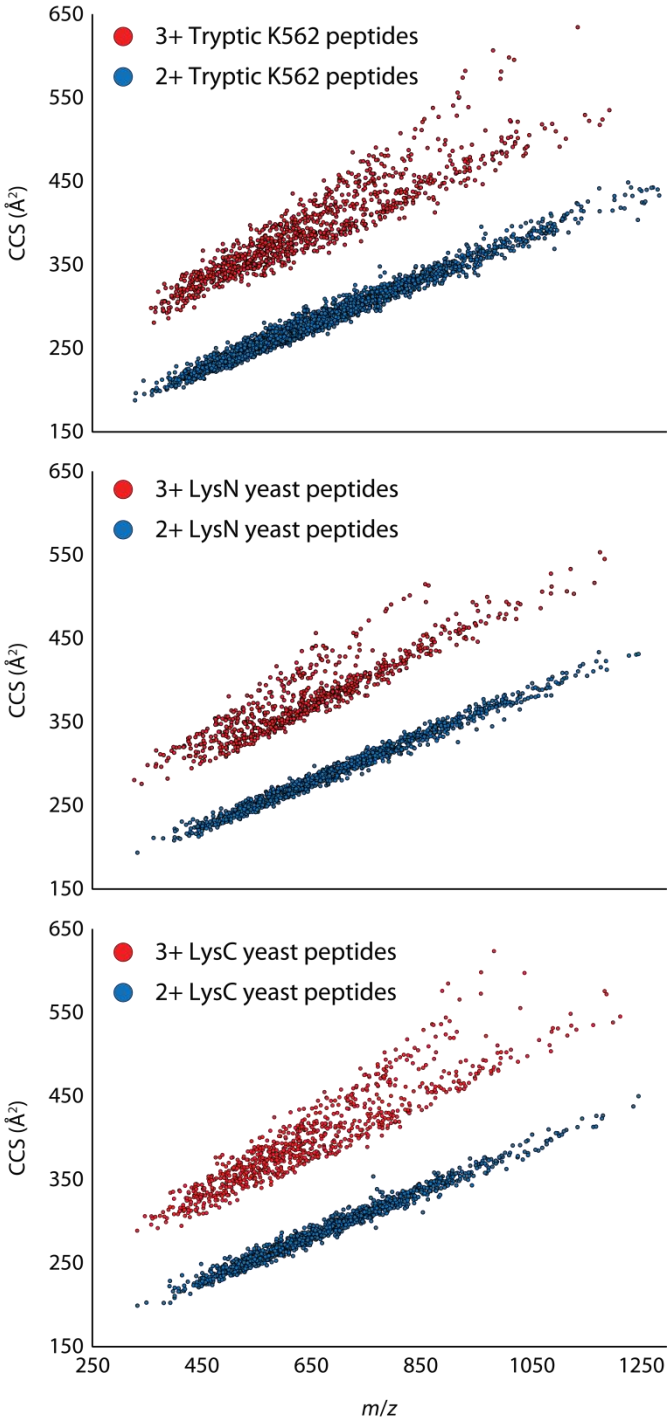
Microsoft Visual C# was used to write pepCCScal. The program and source code can be obtained for free by contacting the corresponding author.



Supplemental Figure 1. LC-HDMS^E chromatograms total peptide IDs from approximately 1.5 µg of tryptic total protein from yeast lysate. The same sample was acquired under (a) low and (b) high transmission voltages.



Supplemental Figure 2. 940 µM BK^{3+} was acquired with EDC target enhancement at the lowest transmission voltages we were able to achieve: 2.0 kV spray voltage, 1.5 V extraction cone, trap bias 33 V, and He cell DC 10 V. The cone voltage was set to different values to apply activation energy. Arrival time profiles at 20 V cone voltage and below were nearly identical.



Supplemental Figure 3. Plots of CCS versus *m/z* for tryptic K562 peptides, LysN yeast peptides, and LysC yeast peptides. These plots only use first-pass PepFrag1 PLGS identifications and omit miscleavages.

Supplemental Table 1. The mean polyalanine calibration error, the discrepancy of known polyalanine CCS values from their CCS, was (0.44 ± 0.03) % for high voltage acquisitions and (0.34 ± 0.03) % for low voltage acquisitions.

$[BK + 2H]^{2+}$

High Voltage		
WV/WH	**BK^{2+} A (Å^2)**	**BK^{2+} B (Å^2)**
500/30	240.17	249.47
500/35	239.99	249.66
600/35	240.54	249.84
700/35	240.81	249.89
800/40	241.46	250.27
Mean CCS	240.6	249.8
SD	0.6	0.3
Error (%)	1.52	1.55
Low Voltage		
WV/WH	**BK^{2+} A (Å^2)**	**BK^{2+} B (Å^2)**
500/30	237.71	246.33
500/35	238.30	247.76
600/35	238.43	247.25
700/35	239.48	248.40
800/40	238.58	247.08
Mean CCS	238.5	247.4
SD	0.6	0.8
Error (%)	0.63	0.55

$[BK + 3H]^{3+}$

High Voltage		
WV/WH	**BK^{3+} B (Å^2)**	**BK^{3+} C (Å^2)**
500/30	289.58	308.57
500/35	293.90	306.17
600/35	292.00	307.51
700/35	292.27	307.73
800/40	295.01	306.30
Mean CCS	293	307
SD	2	1
Error (%)	-	0.74
Low Voltage		
WV/WH	**BK^{3+} B (Å^2)**	**BK^{3+} C (Å^2)**
500/30	286.66	305.33
500/35	292.02	304.13
600/35	289.65	304.92
700/35	290.48	306.24
800/40	291.65	302.95
Mean CCS	290	305
SD	2	1
Error (%)	-	0.09

Supplemental Table 2.

Tryptic BSA Peptide CCS: LC versus Direct Infusion

Peptide	Mean CCS (Å^2)	DI CCS (Å^2)	Deviation of Mean (%)
ATEEQLK	210.9 ± 0.1	209.9	0.5
LCVLHEK	229.4 ± 0.2	227	1.1
AEFVEVTK	221.5 ± 0.1	220.2	0.6
YLYEIAR	239.6 ± 0.1	238.3	0.6
QTALVELLK	247.2 ± 0.1	245.3	0.8
LVNELTEFAK	263.6 ± 0.1	262.2	0.5
HLVDEPQNLIK	289.6 ± 0.1	286.7	1.0
YICDNQDTISSK	297.9 ± 0.1	295.8	0.7
YNGVFQECCQAEDK	327.2 ± 0.1	324.0	1.0
TCVADESHAGCEK	345.3 ± 0.2	342	0.9
DDPHACYSTVFDK	328.7 ± 0.3	324.8	1.2
KVPQVSTPTLVEVSR	346.1 ± 0.5	341	1.5
		Mean	0.9 ± 0.3

Chapter 5

Mass spectrometry analysis of the pain proteome: A review

Adapted from: Christopher B. Lietz and Lingjun Li. Mass Spectrometry Analysis of the Pain Proteome. *Techniques in Regional Aesthesia and Pain Management: Special Focus on Translational Research.* Invited contribution.

Abstract

Proteins are tiny molecular machines that play crucial roles in virtually all biological processes. As the methodology of mass spectrometry (**MS**)-based proteomics matures, applications to research in molecular nociception are becoming more prevalent. This review is intended for audiences with clinical and/or biological background in pain phenomena who wish to understand how cutting-edge MS-based proteomics can assist their research. First, we cover the fundamentals of MS and suggest additional resources for a more in-depth treatment. Next, we explain the important difference between discovery proteomics and targeted proteomics, as well as the circumstances where one would be used over the other. Then, we detail selected pain proteomics publications, ranging from early attempts to the most recent applications. Finally, we offer a perspective on what we believe would be the best use of MS-based proteomics by pain researchers.

Introduction

Proteins are an integral part of virtually every known biological process. They are the machinery of complex assembly lines performing the necessary molecular actions in metabolism, movement, defense, cellular communication, and molecular recognition.[1] Fundamentally, proteins are chain molecules created from different combinations of the same twenty naturally-occurring amino acids. However, the particular combinations and chemical properties of these building blocks will determine a protein's higher structure and, consequently, its unique function. It is analogous to different kind of metal tools such as a hammer and a wrench: despite being made of the same material, their different shapes give them different capabilities.

An essential element of nociception is the cell-to-cell communication via proteins. In the peripheral sensitization following nerve injury, a decrease in the activation threshold of protein ion channels can contribute to the development of pain lasting long after the acute injury-related inflammation.[2] TRPV1 is one such ion channel whose expression and activity is well correlated to chronic neuropathic pain.[3-7] Identifying such key molecular components of pain is a critical step in developing effective pain management, for they could potentially be the targets of new drugs or other novel treatments. However, intervention with a limited understanding of the system can be problematic. Direct drug-targeting of TRPV1 results in the unwanted side-effect of high body temperature,[8] thus some researchers have attempted to inhibit TRPV1 by targeting its associated kinases.[9] Alternatively, one could target proteins that have no direct interactions with TRPV1 but still have active roles in chronic pain, such as nerve growth factor (**NGF**).[10-12] Having greater knowledge of all the molecular players allows more flexibility in avenues for pain

management, and thus the use of an analytical method that can provide a system-wide analysis would be of great utility.

Over the past fifteen years, liquid chromatography (**LC**)-mass spectrometry (**MS**) has emerged as the most powerful method for large-scale protein analysis and is gradually replacing traditional methods that center on gel electrophoresis. Continuous evolution of instrument scan speeds and analytical separations have enabled researchers to detect and quantify thousands of proteins from microgram amounts of starting material. The capabilities of LC-MS are unarguably suited to provide a molecular "bird's eye view" of nociception and elucidate the proteins key to pain's manifestation and sustenance. With greater system-wide knowledge, novel treatment avenues may emerge.

The aim of this review is to highlight LC-MS investigations of pain, particularly of neuropathic origin.[2] We will begin with a very brief introduction to the fundamentals, followed by an overview of the LC-MS workflow and the kinds of questions it can answer. Finally, we will survey published examples of pain proteomic analysis.

Mass Spectrometry Fundamentals

This section is meant to provide uninitiated readers with enough information to follow the rest of the review. More detailed treatments can be found elsewhere.[13-14]

A mass spectrometer measures the mass-to-charge ratio (***m/z***) of an ion. **Scheme 1A** illustrates the basic process of MS analysis. The molecule of interest, a polypeptide, may initially be in a sample with an overall neutral charge. The first step would be to ionize your molecule and desorb it into the gas-phase. Analyte molecules must be gas-phase ions so they can be pushed, pulled, and manipulated by electric fields of the mass spectrometer. Once inside the instrument, an ion's *m*/*z* can be measured by various

means, such as the time it takes an ion to travel a certain distance[15] or its oscillation frequency in an electric field.[16] Regardless of the exact method, knowing the *m*/*z* allows the researcher to determine the molecular mass. For example, a doubly charged ion at 500 *m*/*z* would have a mass of 1000 Daltons (**Da**) while a triply charged ion at 500 *m*/*z* would have a mass of 1500 Da.

The mass of a molecule may be important in and of itself, but the true power of MS comes from tandem mass spectrometry (**MS/MS**). In MS/MS, an ion of interest is isolated in the mass spectrometer and fragmented. Subsequent *m*/*z* analysis of the fragments can then reveal the primary structure, as shown in **Scheme 1B**. Here, two peptides have undergone MS/MS analysis. Their mass spectra contain peaks with unique *m*/*z* that correspond to fragments of their sequences. Using this information, along with masses of the intact peptides, we can identify one peptide as PAINFREE and one as PAINFL.

Quantifying molecules by MS is much less straightforward than their identification. This is because molecules must be ionized, and thus the MS signal is dependent not only on an analyte's abundance, but also on its ionization affinity and the affinity of other molecules being co-ionized. Consequently, early MS proteomics investigations relied on a separate analysis for protein quantification. Modern methodology now allows for quantitative MS with and without the use of chemical labels. The details of these methods are beyond the scope of this review, but can be found elsewhere.[17-19]

The LC-MS Proteomics Workflow

The quality and depth of a proteomic analysis depends highly on the structure of its workflow. Although some aspects are nearly universal, such as the use of genome-predicted protein databases, components like the use of offline sample fractionation may

not always applicable. This section of the review aims to explain the basic LC-MS proteomics workflow, illustrated in **Scheme 2**, separating methods into two categories: discover and targeted. Interested readers are referred to other excellent reviews for more detailed treatments.[20-22]

Discovery Proteomics

Discovery proteomics is primarily used for untargeted investigations where the protein roles in a given biological system are not well characterized. The object is to identify and quantify as many proteins as possible, then examine the expression patterns within related classes of proteins or possible interacting groups. The result is a list of protein candidates that may modulate the system under investigation.

This review will focus on methods of bottom-up shotgun proteomics, wherein proteins are identified and quantified from enzymatic peptide fragments.[20] As shown in **Scheme 2A**, the process begins by extracting proteins from tissues or cells of interest. Intact proteins are then cleaved by high-specificity enzymes, most commonly trypsin or endoprotease LysC. The peptides are then separated by LC on a column with an integrated electrospray ionization (**ESI**) emitter so that molecules are ionized and introduced into the mass spectrometer upon elution. Identified peptide sequences are mapped back to their parent proteins using genome-predicted protein database.

It's been demonstrated that mass spectrometer scan speed has one of the greatest impacts on the number of proteins one can identify and quantify.[23] The number of identifications is often positively correlated with scans/second due to the sheer number of peptides in a whole-protein digest. Identifications are also positively correlated with LC peak capacity, suggesting the vital importance of analyte separation.[24] Samples can even

be pre-fractionated by an orthogonal separation method, and each fraction is then separately submitted to LC-MS to minimize sample complexity.[25-26]

It is worth noting that gel-based separation has not been completely removed from discovery workflows, especially in pain proteomics.[27] Here, intact proteins are separated and digested in an electrophoresis gel. Peptides are extracted from selected gel spots and submitted to LC-MS or matrix-assisted laser desorption/ionization (**MALDI**), the latter ionizing molecules from solid-state samples and omitting LC separation.

Discovery proteomics is primarily used for hypothesis generation. When proteins of interest are numerous, vague, or completely unknown, this workflow can be used as a foundation for further inquiry. If investigations already have clear candidates in mind, a targeted analysis may be more appropriate.

Targeted Proteomics

Targeted proteomics is an extremely powerful tool.[22] Unlike discovery workflows, in which protein quantification is typically relative, targeted assays can yield absolute amounts. One could use targeted proteomics to validate the differential expression of proteins candidates from a discovery list, or even to test a hypothesis by modulating candidates and quantifying the response of related proteins.

Scheme 2B illustrates the targeted proteomic workflow, much of which is technically similar to discovery proteomics. Proteins are extracted, digested, and submitted to LC-MS/MS analysis. However, targeted assays require one to generate a calibration curve for proteins of interest. The most common and accurate method is known as AQUA.[28] It utilizes a light and heavy standard for calibration curve generation. The light standard is an enzymatic peptide from the protein of interest, and the heavy

standard is that same peptide sequence but made from amino acids with stable heavy isotopes (i.e. ^{13}C, ^{15}N, ^{2}H). The isotopes provide a detectable mass shift for the heavy standard and one can calculate the light/heavy MS intensity ratio with varied amount of light standard. Then, a known amount of heavy standard is spiked into a real sample, and the absolute abundance of the protein of interest can be calculated.

Selection of a representative peptide can be a timely process. The peptide must be reproducibly detected from real samples, and it must exhibit a linear response in the calibration curve. However, careful optimization and execution of targeted proteomic workflows can yield invaluable information about a biological system, especially when validation is not possible by traditional methods such as ELISA.[22]

Proteomics in Pain Research

The aim of this section is to survey the results of MS-based proteomic studies in pain research. The examples have been limited to publications from the last three years, but interested readers are referred to another review for earlier investigations.[27]

Annexin A2 and TRPA1-Dependent Nociception

In 2014, Avenali et al. used LC-MS to study TRPA1.[29] TRPA1 is a transmembrane ion channel protein implicated in hypersensitivity phenomena.[30-31] Although the number of TRPA1 proteins found at the cell membrane can significantly affect its signaling pathways,[32] few details of the TRPA1 localization mechanism are known. This investigation aimed to identify new modulators.

Avenali et al. used affinity-purification MS,[33] wherein a target protein is co-immmunoprecipitated with its native complexes, to search for novel TRPA1 binders. Proteins were pulled down from two replicates of trigeminal ganglia tissue and transfected

HEK 293T cell lysates, separated on a 1-D PAGE gel, digested, and analyzed by LC-MS. AnxA2 was confidently identified in both experimental replicates and was absent from both control replicates.

Although this investigation did not yield large numbers of novel TRPA1 binders, it is a prime example of how to use discovery proteomics for first-pass hypothesis generation. A panel of follow-up experiments were performed to further characterize TRPA1-AnxA2 interactions that included electrophysiology, calcium imaging, immunohistochemistry, and behavioral testing of AnxA2$^{-/-}$ mice. The results showed that AnxA2 co-localizes with TRPA1 and is negatively correlated with TRPA1 membrane localization and related nociceptive behaviors.

Metallothionein Deficiency in Complex Regional Pain Syndrome

Complex regional pain syndrome (**CRPS**) proteomics was the focus of an investigation by Oki et al. in 2012.[34] The authors assert that there is no definitive treatment for CRPS and "commonly used remedies" show little efficacy in clinical trials. Thus, their aim was to identify proteins with significantly up-regulated expression in injured nerves and gain more insight into the mechanism of CRPS.

Proteins were obtained from resected nerves of seven human subjects, five diagnosed with CRPS and two diagnosed with having painful neuroma. Control protein samples were obtained from sural nerves of fresh cadavers. Extracts from three CRPS patients and two controls were separated by SDS-PAGE, digested, and analyzed by MALDI-MS. A total of 2912 unique proteins were identified, and the authors noted significance of five metallothionein proteins being readily detected in control samples but appearing absent from CRPS samples in both MS and Western Blot analysis.

As with Avenali et al.,[29] Oki et al. also demonstrate effective use of discovery proteomics as a hypothesis generator. However, this study also shows how a larger number of identifications can reveal patterns throughout a class of proteins. Immunohistochemical staining further confirmed the lack of metallothionein proteins in all five CRPS patients and both painful neuroma patients. Metallothioneins are known to scavenge harmful reactive oxygen species, and the authors conclude their work adds to evidence of oxidative stress and dysregulated inflammatory responses in CRPS patients.

Large-scale Proteomics in an Animal Model of Neuropathic Pain

In 2014, Sui et al. published a large-scale proteomic investigation of the spinal nerve ligation (**SNL**) rat model.[35] The SNL model was induced by surgical ligation of the L5 and L6 nerves that leads to hindpaw sensitivity and symptoms of hyperalgesia and allodynia.[36] The authors aimed to measure the global expression profiles across many types of proteins in the spinal cords of SNL rats and compare them to SHAM and Naïve control groups.

This investigation did not use gel-based separation. Following behavioral testing, specimens were euthanized and dorsal spinal cord tissue was removed from ten rats in each of the three groups. Protein was extracted, digested, and then labeled with dimethyl tags to allow for quantitative MS.[37] LC-MS of the labeled peptides yielded 498 unique proteins, 153 of which were identified in at least 9 rats from each group. Of those 153, thirty-eight displayed differential expression of statistical significance.

While Sui et al. did not perform any separate follow-up or validation experiments, they demonstrate the amount of knowledge that can be gained from MS alone. Proteins from many different categories were found to be positively correlated with SNL. These

include 14-3-3 family proteins, GFAP, ENO1, and NEFL1, all of which are also involved in other neurological disorders.[38] The authors also performed a pathway analysis, wherein association networks in the existing knowledgebase were searched for overrepresentation by the thirty-eight differentially expressed proteins. Several proteins were found to be associated with the chloride channel CLIC4. Little is known about CLIC4 function, and thus this discovery provides a starting point to gain brand new insights into the potential contributions to neuropathic pain.

Conclusion

MS-based proteomics has proven to be a superb tool throughout biological and medical sciences, and its use in pain research is rapidly growing. Discovery proteomic workflows have a lot to offer for complex systems like neuropathic pain where multiple mechanisms are involved and many fundamental elements are still unknown, and there is always a place for targeted workflows to confirm the fine details.

It is of our opinion that the gel-free approach by Sui et al.[35] should be universally adopted in discovery proteomic workflows. Visualizing proteins on gels prior to MS is a hallmark of early proteomic studies, but unless the experiment is targeted or the gel is intended to enrich the sample for a particular mass, a researcher can probe much deeper into the proteome with 1-D or 2-D LC. Additionally, global analysis of post-translational modifications (**PTMs**) like phosphorylation,[39] glycosylation,[40] and acetylation[41] has been relatively unexplored by MS in pain research. MS has great strengths for PTM analysis by being able to identify modified proteins and to localize the modification to the precise amino acid residue. Researchers should readily use this technology to study pain-related

PTMs, for quantifying the existence of a protein's bioactive state is much closer to studying its function than simply quantifying its base expression.

References

(1) Kessel, A., and Ben-Tal, N. (2011) *Introduction to Proteins*, CRC Press, Boca Raton, FL, USA.

(2) von Hehn, C. A., Baron, R., and Woolf, C. J. *Neuron* **2012** *73*, 638-652.

(3) Hudson, L. J., Bevan, S., Wotherspoon, G., Gentry, C., Fox, A., and Winter, J. *Eur. J. Neurosci.* **2001** *13*, 2105-2114.

(4) Kim, H. Y., Park, C. K., Cho, I. H., Jung, S. J., Kim, J. S., and Oh, S. B. *J. Pain* **2008** *9*, 280-288.

(5) Pabbidi, R. M., Yu, S. Q., Peng, S., Khardori, R., Pauza, M. E., and Premkumar, L. S. *Mol. Pain* **2008** *4*, 9.

(6) Christoph, T., Grunweller, A., Mika, J., Schafer, M. K., Wade, E. J., Weihe, E. *et al. Biochem. Biophys. Res. Commun.* **2006** *350*, 238-243.

(7) Watabiki, T., Kiso, T., Kuramochi, T., Yonezawa, K., Tsuji, N., Kohara, A. *et al. J. Pharmacol. Exp. Ther.* **2011** *336*, 743-750.

(8) Swanson, D. M., Dubin, A. E., Shah, C., Nasser, N., Chang, L., Dax, S. L. *et al. J. Med. Chem.* **2005** *48*, 1857-1872.

(9) Anand, P., Shenoy, R., Palmer, J. E., Baines, A. J., Lai, R. Y., Robertson, J. *et al. Eur. J. Pain* **2011** *15*, 1040-1048.

(10) Lane, N. E., Schnitzer, T. J., Birbara, C. A., Mokhtarani, M., Shelton, D. L., Smith, M. D. *et al. N. Engl. J. Med.* **2010** *363*, 1521-1531.

(11) Gaudet, A. D., Popovich, P. G., and Ramer, M. S. *J. Neuroinflammation* **2011** *8*, 110.

(12) Dogrul, A., Gul, H., Yesilyurt, O., Ulas, U. H., and Yildiz, O. *Acta Diabetol.* **2011** *48*, 135-142.

(13) Maher, S., Jjunju, F. P. M., and Taylor, S. *Rev. Mod. Phys.* **2015** *87*, 113-135.

(14) (2006) *Principles of Mass Spectrometry Applied to Biomolecules*, Wiley-Interscience, Hoboken, NJ, USA.

(15) Chernushevich, I. V., Loboda, A. V., and Thomson, B. A. *J. Mass Spectrom.* **2001** *36*, 849-865.

(16) Zubarev, R. A., and Makarov, A. *Anal. Chem.* **2013** *85*, 5288-5296.

(17) Cox, J., and Mann, M. *Nat. Biotechnol.* **2008** *26*, 1367-1372.

(18) Cox, J., Hein, M. Y., Luber, C. A., Paron, I., Nagaraj, N., and Mann, M. *Mol. Cell. Proteomics* **2014** *13*, 2513-2526.

(19) Rauniyar, N., and Yates, J. R. *J. Proteome Res.* **2014** *13*, 5293-5309.

(20) Zhang, Y., Fonslow, B. R., Shan, B., Baek, M.-C., and Yates, J. R. *Chem. Rev.* **2013** *113*, 2343-2394.

(21) Richards, A. L., Merrill, A. E., and Coon, J. J. *Curr. Opin. Chem. Biol.* **2015** *24*, 11-17.

(22) Elschenbroich, S., and Kislinger, T. *Mol. BioSyst.* **2011** *7*, 292-303.

(23) Hebert, A. S., Richards, A. L., Bailey, D. J., Ulbrich, A., Coughlin, E. E., Westphall, M. S. *et al. Mol. Cell. Proteomics* **2014** *13*, 339-347.

(24) Köcher, T., Swart, R., and Mechtler, K. *Anal. Chem.* **2011** *83*, 2699-2704.

(25) Washburn, M. P., Wolters, D., and Yates, J. R., 3rd. *Nat. Biotechnol.* **2001** *19*, 242-247.

(26) de Godoy, L. M. F., Olsen, J. V., Cox, J., Nielsen, M. L., Hubner, N. C., Frohlich, F. *et al. Nature* **2008** *455*, 1251-1254.

(27) Niederberger, E., and Geisslinger, G. *Anesthesiol.* **2008** *108*, 314-323.

(28) Kirkpatrick, D. S., Gerber, S. A., and Gygi, S. P. *Methods* **2005** *35*, 265-273.

(29) Avenali, L., Narayanan, P., Rouwette, T., Cervellini, I., Sereda, M., Gomez-Varela, D. *et al. J. Neurosci.* **2014** *34*, 14506-14516.

(30) Hucho, T., and Levine, J. D. *Neuron* **2007** *55*, 365-376.

(31) Patapoutian, A., Tate, S., and Woolf, C. J. *Nat. Rev. Drug Discovery* **2009** *8*, 55-68.

(32) Schmidt, M., Dubin, A. E., Petrus, M. J., Earley, T. J., and Patapoutian, A. *Neuron* **2009** *64*, 498-509.

(33) Dunham, W. H., Mullin, M., and Gingras, A.-C. *PROTEOMICS* **2012** *12*, 1576-1590.

(34) Oki, G., Wada, T., Iba, K., Aiki, H., Sasaki, K., Imai, S. *et al. Pain* **2012** *153*, 532-539.

(35) Sui, P., Watanabe, H., Ossipov, M. H., Bakalkin, G., Artemenko, K., and Bergquist, J. *J. Proteome Res.* **2014** *13*, 3957-3965.

(36) Kim, S. H., and Chung, J. M. *Pain* **1992** *50*, 355-363.

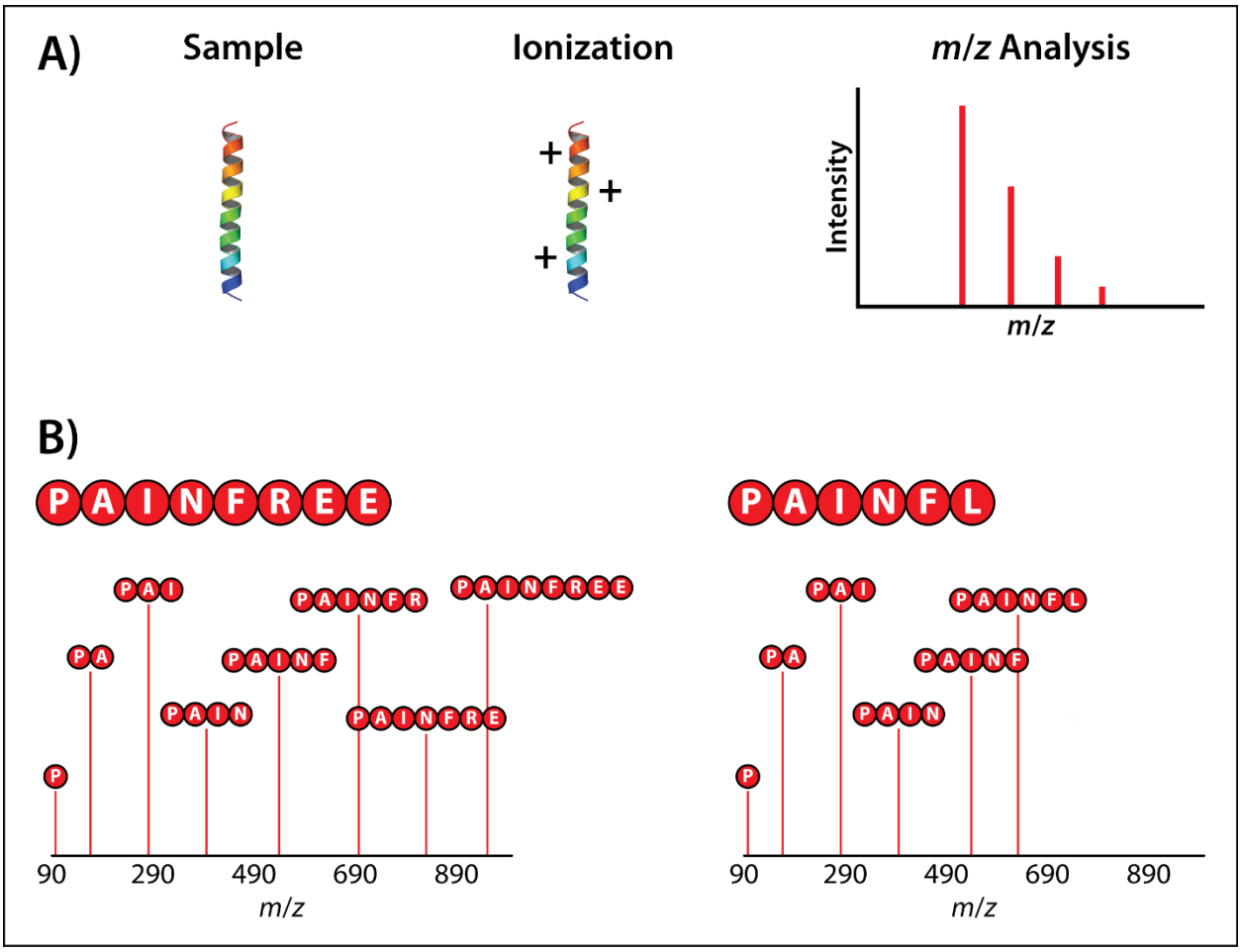
(37) Boersema, P. J., Raijmakers, R., Lemeer, S., Mohammed, S., and Heck, A. J. *Nat. Protoc.* **2009** *4*, 484-494.

(38) Fountoulakis, M., and Kossida, S. *Electrophoresis* **2006** *27*, 1556-1573.

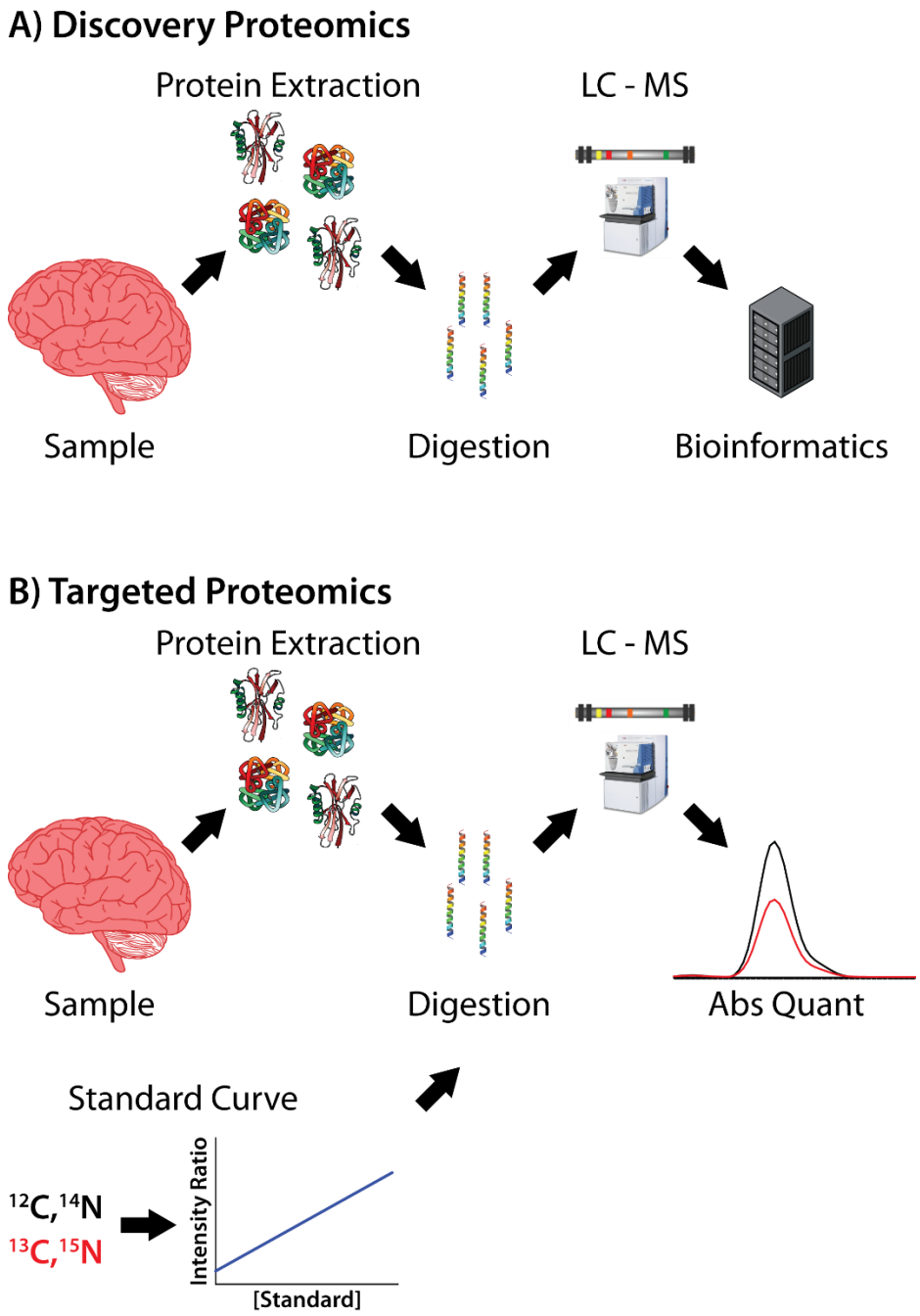
(39) Roux, P. P., and Thibault, P. *Mol. Cell. Proteomics* **2013** *12*, 3453-3464.

(40) Pan, S., Chen, R., Aebersold, R., and Brentnall, T. A. *Mol. Cell. Proteomics* **2011** *10*, R110 003251.

(41) Choudhary, C., Weinert, B. T., Nishida, Y., Verdin, E., and Mann, M. *Nat. Rev. Mol. Cell. Biol.* **2014** *15*, 536-550.



Scheme 1. Peptide sequencing via MS/MS analysis. (a) Initially, peptide samples are ionized and detected, leading to the calculation of their *m*/*z*. (b) Peptide ions can be isolated and fragmented in the gas-phase. Following *m*/*z* analysis of the resulting fragments, the peaks can be assigned to portions of the precursor and the full sequence can be determined.



Scheme 2. Workflows for (a) discovery and (b) targeted proteomics.

Chapter 6

Quantitative proteomics of the spinal cord and dorsal root ganglion in the spared-nerve injury model of neuropathic pain

Adapted from: Christopher B. Lietz,* Dana M. Tilley,* Courtney A. Kelley, David L. Cedeño, Joseph Williams, Ramsin Benyamin, Lingjun Li, and Ricardo Vallejo. Quantitative Proteomics of the Spinal Cord and Dorsal Root Ganglion following Peripheral Nerve Injury in a Rat Model of Neuropathic Pain. (*J. Proteome Res.* Submitted.)
*Denotes equal contribution

Introduction

Chronic pain impacts the quality of life and socioeconomic welfare of patients and their families. Over 100 million Americans are affected by chronic pain with 18% of cases being of neuropathic origin.[1] It was estimated that chronic pain costs about $600 billion per year, including both medical expenditures and lost wages[2] while, in contrast, only $413 million were invested in pain research, according to a recent NIH report.[3] Research on chronic pain has not been substantial due to funding insufficiencies and disease-status acceptance, thus the mechanism by which pain of neuropathic origin becomes chronic is not fully understood at the molecular level.

Woolf and coworkers recently reviewed the current state of our general understanding of pain at the molecular level.[4] Nerve injury involves a local and inflammatory initial stage, which progresses along peripheral processes modulating the DRG and inducing central sensitization involving synaptic changes in the spinal cord (**SC**) and brain regions that process pain. Beside changes in neuronal activity, there are substantial changes in the activity of glial and immune cells in the peripheral and central nervous system. Thus, the concerted central interaction of neuronal and glial cells leads to the potentiation and establishment of chronic pain.

The identification of key molecular players has relied on electrophysiology, genetic analysis, immunohistochemistry, and limited protein studies.[5-12] As the need for novel treatments for chronic pain remains critical, new sources of insight are increasingly in demand.

Over the past fifteen years, liquid chromatography (LC) mass spectrometry (MS)-based proteomics has emerged as the dominant method for global protein expression

analysis.[13] In the most commonly used method, “bottom-up” shotgun proteomics, total protein extracts are enzymatically digested prior to LC-MS. Resulting peptide fragments are sequenced and mapped back to their proteins of origin and serve as identifications of those proteins in the sample. Although advances in instrumentation and front-end analytical separations have enabled researchers to identify and quantify thousands of proteins from microgram amounts of total protein extract, such a feat is not trivial. Difficulties of deep proteome analysis stem from a large dynamic range—a staggering seven orders of magnitude in mammalian cells[14]—and the ionization suppression of peptides with low ionization affinity or low abundance. The number of peptide and proteins identified using LC-MS has been shown to strongly correlate with chromatographic peak capacity, defined as the number peaks that can be separated and distinguished from one another over the course of an experiment.[15] An effective way to maximize the total number of distinguishable peaks is to pre-fractionate using a method that’s chemically and physically orthogonal to the separation method used for LC-MS.[16-17] Combined with careful sample preparation and powerful online-LC, extensive pre-fractionation has yielded over 10,000 protein identifications from human cells in a single experiment.[18-19]

Proteomics of the nervous system, or neuroproteomics, has been utilized in numerous investigations of fundamental and clinical interest,[20-24] including the molecular mechanisms of peripheral nerve injury and pain.[25-26] The high-throughput capacity of proteomic analysis and the ever growing annotations will ultimately reveal connections that are missing from our current library of knowledge.[21-23, 25-27]

Previous studies in neuroproteomics outline the complexity of the nervous system

response to injury.[22, 25, 27-29] These studies identified proteins that play key roles in the establishment and maintenance of chronic pain, but were limited by the technical capabilities available at the time they were carried out. Our study is the first large-scale investigation of protein dynamics in both the ipsilateral DRG and spinal cord after a chronic pain state is induced following a peripheral nerve injury in an animal model. We have used an advanced multiplex isobaric labeling approach for protein quantification, yielding a significantly broader picture of protein regulatory changes compared to previous studies. Combined with our approach of pre-fractionation, we were able to obtain the largest number of quantified proteins of any study of the SNI model. The protein changes found in this study provide critical insight into key pathways associated with the early establishment of chronic neuropathic pain.

Experimental

Animal Specimen Selection

All animal procedures were approved by the Institutional Animal Care and Use Committee (IACUC) at Illinois Wesleyan University (Bloomington, IL, USA). A total of 8 male Sprague-Dawley rats (Harlan, Indianapolis, IN, USA) weighing 275-305 grams were housed in single cages and allowed one week to acclimate in a 12 hour light/dark cycle with food and water available *ad libitum*.

Surgical Procedure: Spared Nerve Injury (SNI) Model

Animals were anesthetized with 3-4% isoflurane and sedation was maintained by spontaneously breathing 2-3% isoflurane. To confirm the proper depth of anesthesia, a contralateral noxious paw pinch was used to test responsiveness. Body temperature was maintained by using a heated operating table (Harvard Apparatus, Holliston, MA, USA).

Animals were randomized into two experimental groups (n = 4 for each): a spared nerve injury surgery (SNI), and a sham surgery (SHAM).

As previously reported,[12, 30] our slightly modified SNI model was designed to induce sensitivity in the left hindpaw by transecting and removing 1-2 mm of the tibial and common peroneal branches of the sciatic nerve below the level of trifurcation. Animals randomized in the SHAM group were operated similarly but the sciatic nerve was not transected.

Behavioral Testing

Baseline behavioral testing for mechanical and cold thermal allodynia was performed three days prior to the sham or SNI surgery. Behavioral testing was also completed on days 3 and 4, post-surgery, to observe behavioral changes that would indicate the development and maintenance of pain.

Mechanical allodynia was measured via an electronic von Frey anesthesiometer (IITC, Woodland Hills, CA, USA).[28] Animals were placed on an elevated platform in bottomless cages which were open to a mesh floor, and allowed a 20 minute acclimation time. Then, 3.6 g, 8.0 g, and 20.0 g filaments were applied to the plantar surface of both ipsilateral and contralateral hind paws in ascending order to measure the amount of force that can be placed on a paw before eliciting a response or until a bend in the filament was observed. The electronic von Frey device reported the highest measured force in grams and was recorded for each trial. A total of six trials per filament, per paw were completed and averaged for each animal.

Cold thermal allodynia was tested using the acetone drop method.[31] A drop of acetone was applied to the plantar surface of each hindpaw, using a syringe connected

to a modified pipette tip, and the withdrawal duration was measured, beginning with the onset of withdrawal and ending when the hindpaw rested. A hindpaw withdrawal of at least one second was considered a response to the stimulus and was scored as a one. A withdrawal lasting less than one second was scored as zero. Four trials were completed; percent response was calculated per paw and averaged for each animal per day of testing.

Results were calculated based on averages of all trials per behavioral test performed, per animal. Significance among the behavioral data was determined using IBM SPSS Statistics software (v. 22) by a p-value of less than or equal to 0.05 through a heteroscedastic student's t-test due to binary group-wise comparisons.

Protein Extraction

After behavioral testing on day 4, the animals were euthanized via CO_2 inhalation followed by decapitation. A 5-6 mm section of the dorsal ipsilateral quadrant of the SC, at the level of L5 nerve root innervation, and the ipsilateral L5 DRG were removed and flash-frozen for proteomic analysis. The dissected tissues were then combined into two biological replicates per tissue type. Pooled tissues were transferred to a microcentrifuge tube and suspended in lysis buffer (8 M urea, 50 mM Tris HCl, 30 mM NaCl, 5 mM $CaCl_2$, pH ~8) containing cOmplete protease inhibitors (Roche Diagnostics, Indianapolis, IN, USA). Sample tubes were then placed in an ice water bath and lysed by pulsed probe sonication (Thermo Fisher Scientific, Waltham, MA, USA). Resulting lysates were centrifuged at 10,000 x*g* for five minutes, and the supernatant was collected. Following a BCA total protein assay (Pierce, Rockford, IL), samples were stored at -80 ºC until further use.

Protein Digestion and Peptide Labeling

Each DRG and SC biological replicate was divided into three aliquots of 75 µg and 200 µg total protein, respectively. Cysteine residues were then reduced and alkylated, and samples were digested as described previously.[32] Digestion was quenched by addition of 10% trifluoroacetic acid.

Peptide solid-phase extraction (SPE) was performed with C18 Sep-Pak cartridges (Waters, Milford, MA). Extracted samples were then labeled with sixplex tandem mass tag (TMT) (Pierce) according to the manufacturer's protocol. After the labeling reaction was quenched, labeled aliquots were combined and dried down via SpeedVac concentrator.

High Performance Liquid Chromatography (HPLC) Fractionation

SC and DRG biological replicates were fractionated on a Waters Alliance e2695 HPLC using a polysulfoethyl A strong cation-exchange (**SCX**) column (200 x 2.1 mm, 5 µm, 300 Å, PolyLC Inc., Columbia, MD, USA). SCX separates peptides based on their electrostatic attraction to the negatively charged stationary phase, which is essentially determined by the peptide's acidity. This is orthogonal to the hydrophobicity-based separation used down-stream, and thus peptides that elute closely in SCX may elute far apart in LC-MS. Separation occurred over a 75 minute salt gradient. Fractions were collected every 1.5 min and re-combined into a total of 18 fractions per biological replicate. Each fraction was then desalted by C18 SPE Omix pipette tips (Agilent Technologies, Santa Clara, CA, USA) and reconstituted in 12 µL of 0.1% formic acid (FA).

Liquid Chromatography – Mass Spectrometry (LC-MS) Acquisition

All LC-MS analyses were performed on a nanoAcquity UPLC (Waters, Milford, MA,

USA) and an Orbitrap Elite mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA). Self-packed nano LC columns with integrated electrospray ionization (ESI) emitters were fabricated from 75 µm I.D. fused silica capillary using a P-2000 laser puller (Sutter Instrument Co., Novarto, CA, USA). Columns were packed with Waters BEH C18 beads (1.7 µm, 130 Å). Samples were separated over a 100 min acetonitrile gradient. Eluting peptides were electrosprayed into the mass spectrometer and sequenced via top 15 data-dependent tandem MS (MS/MS) cycles. Each HPLC fraction was individually analyzed by LC-MS, resulting in a total of 72 LC-MS runs. Raw data files are freely available at http://chorusproject.org.

Protein Identification and Quantification

Raw data from the biological replicates for each tissue type were batched into a single dataset and analyzed with the COMPASS 1.4 software suite.[75] Peptide MS/MS spectra were searched against a concatenated target-decoy *Rattus norvegicus* protein database (downloaded on 09/25/2014 from www.uniprot.org; 7910 Swiss-Prot entries, 25788 TrEMBL entries) using the Open Mass Spectrometry Search Algorithm (**OMSSA**). Following the OMSSA search, the raw peptide identifications were curated to a 1% false discovery rate (**FDR**) using the FDR Optimizer. Peptides were then assigned to parsimonious protein groups using Protein Hoarder, and each protein group required a minimum of two unshared peptide sequences. The list of identified protein groups was also curated to an FDR of 1%. TMT analysis was performed by the COMPASS Tag Quant feature.

PSEA-Quant Gene Ontology (GO) Enrichment

GO-term enrichment analysis was performed using the PSEA-Quant algorithm.[33]

Each identified protein that had a gene symbol (3992 DRG proteins, 5139 SC proteins) was compiled into a list with all possible SNI/SHAM TMT reporter ratios (126/129, 126/130, 126/131, 127/129, 127/130, 127/131, 128/129, 128/130, 128/131). The list of enriched GO-terms was filtered to remove enriched terms with an FDR Q-value > 0.1. Separate analyses were performed for DRG and SC protein expression.

Validation of BAG3 in SC by Western Blotting

Protein was extracted from tissue homogenate using Trizol per manufacturer's instructions (Fisher Scientific, Massachusetts, USA) and crude protein was separated by 12% SDS-PAGE. After electrophoresis, proteins were transferred to a PVDF membrane (Fisher Scientific, Massachusetts, USA). The membrane was incubated with rabbit anti-BAG3 antibody (Assay Biotech, California, USA, 1:2500 dilution) at 4˚C overnight, then incubated with horseradish peroxidase-conjugated goat anti-rabbit IgG antibody (Antibodies-Online, Georgia, USA, 1:2500) for four hours at room temperature. BAG3 was visualized using chemical luminescence via an AEC substrate kit (Fisher Scientific, Massachusetts, USA). The BAG3 bands were imaged using a Gel Logic 200 Imaging System (Kodak, New York, USA) and band intensities were quantified via ROI analysis using analysis software provided with the imaging system.

Supplemental Methods

Instrument parameters, technical details, and further information for MS and chromatographic methods can be found in **Supplemental Information 1**.

Results and Discussion

Behavioral Results

As expected, the SNI model effectively increased pain sensitivity in animals. Von

Frey testing for mechanical allodynia revealed a significant decrease in withdrawal thresholds using the 20.0 g filament after induction of the SNI model. A repeated measures ANOVA using testing day (Baseline, Day 3 and Day 4) as a within subjects variable and lesion (SNI vs. SHAM) as a between subjects variable revealed a significant testing day × lesion interaction effect ($F(2,6) = 4.540$; $p = 0.012$). Subsequent independent t-tests revealed increased paw sensitivity following SNI lesions, with the SNI group exhibiting significantly lower withdrawal thresholds on day 3 ($p = 0.010$) and day 4 ($p = 0.047$), relative to the SHAM group. There was no significant difference between SNI and SHAM animals during baseline testing, and no significant changes were measured in the contralateral hindpaws ($p > 0.05$).

Animals also showed a significant increase in percent reaction to cold thermal allodynia following SNI surgery. A repeated measures ANOVA using testing day (Baseline, Day 3 and Day 4) as a within subjects variable and lesion (SNI vs. SHAM) as a between subjects variable revealed a significant testing day × lesion interaction effect ($F(2,6) = 4.034$; $p = 0.024$). Subsequent independent t-tests revealed increased paw sensitivity to acetone application with withdrawal thresholds being significantly different between SNI and SHAM animals on day 3 ($p = 0.024$, and day 4 ($p = 0.024$), post-surgery. The SHAM animals showed no allodynic response to cold (acetone drop) on any day. There were no responses to the cold sensitivity with the contralateral hindpaws or at baseline testing for either paw ($p > 0.05$).

Quantitative Proteomics

Scheme 1 visualizes the workflow of our bottom-up proteomic analyses. Tissues from the eight rats were divided into two biological replicates, each containing two SNI

and two SHAM tissues. Following total protein extraction, each SNI and SHAM extracts were aliquoted into three technical replicates and digested with trypsin. Relative protein quantification was enabled by use of isobaric labels.[34] Briefly, isobaric labels are composed of a set of amine-reactive molecules that tag lysines and peptide N-termini. Though identical in overall structure, stable heavy isotopes (^{13}C, ^{15}N, etc.) are strategically incorporated into different molecular positions during synthesis so that each tag has the same molecular weight but will yield unique reporter ions upon fragmentation. SNI peptides were labeled with TMT 126-128, and Sham peptides were labeled with TMT 129-131. The aliquots were then combined so that peptides present in both SNI and SHAM were simultaneously analyzed by LC-MS. The intensities of the unique reporter fragments were compared quantitatively. TMT quantification was chosen over label-free methods for three reasons: 1) it increased throughput by allowing simultaneous analysis of six samples, 2) it removed the inaccuracies introduced by differential sample handling and variations in instrument performance, and 3) it was very amenable to virtually any kind of pre-fractionation. The entire workflow was performed separately for DRG and SC.

Bioinformatic analysis of the LC-MS data resulted in the identification of 4074 proteins from the DRG and 5232 proteins from the SC at a 1% FDR. Among them, 3421 proteins were identified in both tissue types, while 653 were unique to the DRG and 1813 were unique to the SC. **Figure 1** contains histograms that plot the number of proteins displaying SNI/SHAM expression ratios from -1.8 to 1.8 on a Log_2 scale. Due to the narrow distribution of ratios in both tissue types, we set two significance thresholds for differential expression. The black dotted line marks a hard significance threshold, requiring Log_2 protein ratios to be larger than +0.5 or less than -0.5. The green dotted

line marks a soft significance threshold that contains the 2% most up-regulated and down-regulated proteins. These soft thresholds correspond to the respective positive and negative SNI/SHAM Log_2 ratios of +0.37 and -0.36 in DRG, and +0.28 and -0.23 in SC. In the DRG, 77 SNI up-regulated and 65 SNI down-regulated proteins were identified at the soft threshold after removing proteins with P-value > 0.05. Forty-three of the up-regulated proteins and 24 of the down-regulated proteins also met the criteria for the hard significance threshold (**Table 1**). In the SC, 101 SNI up-regulated proteins and 86 SNI down-regulated proteins were identified at the soft threshold following P-value filtering. Thirty-four of the up-regulated proteins and 18 of the down-regulated proteins also met the criteria for the hard significance threshold (**Table 2**). A complete list of SNI/SHAM ratios for all protein identifications can be found in Supplemental Information.

PSEA-Quant Gene Ontology Analysis

Traditional gene ontology (**GO**) analysis of proteomic datasets is done on a subset of quantified proteins meeting an arbitrary significance cut-off, such as our previously defined hard and soft significance thresholds. However, a consequence of using any arbitrary abundance ratio cut-off is the incorrect differential treatment of ratios just near the thresholds. Therefore, we opted to use a protein set enrichment analysis, PSEA-Quant, developed by Levellee-Adam et al.[33] PSEA-Quant forgoes an arbitrary abundance threshold and uses ratios from all proteins to identify enriched GO terms. An empirical FDR Q-value is calculated for each enriched term as a quality check. The basis of PSEA-Quant is derived from the previously described PSEA[35] and gene set expression analysis.[36]

PSEA-Quant identified 116 DRG GO terms and 78 SC GO terms enriched in SNI

tissue with an FDR Q-value ≤ 0.1. **Figure 2** lists 20 tissue-specific GO terms with the highest enrichment scores (determined by number of proteins identified with that term and the protein's abundance in SNI) normalized to the total number of proteins containing that GO term in the database.

The PSEA-based GO enrichment provides some hints about the role of proteins from the perspective of molecular function, biological process and cellular component. The highest score GO in the DRG, membrane attack complex, involves complement proteins (C8a, C9, C8b and C6) which are up-regulated together with others identified in the GO: complement activation (Crp, C4A, C3, CFI, C4BPA, and C1QB). PSEA of the proteins differentially expressed in the SC emphasizes the role of proteins involved in the laminin complex and adhesion

Below we present results based on the relevance and role of the most regulated proteins in the DRG and SC (**Tables 1** and **2**) in the context of processes identified in the establishment of neuropathic pain. Although these tables only show proteins significantly up or down regulated by a 1.41-fold change or more, the results and discussion also include a few relevant significant proteins with slightly lower fold changes, but nothing below 1.30-fold (i.e. no less than 30% change).

Neuronal regeneration and degeneration - DRG

Various proteins of the myosin complex are up-regulated: myosin heavy chain 2X (MYH1, 8.62-fold), myosin light chain 1/3 (MYL1, 2.34-fold), myosin-4 (MYH4, 2.15-fold), myosin-7 (MYH7, 2.08-fold), tropomyosin β-chain (TPM2, 1.83-fold), as well as tubulin β-2B (TUBB2B, 1.57-fold) and actinin alpha3 (ACTN3, 1.44-fold). These proteins are associated to axonal transport and cytoskeletal organization.[37] Another protein highly up-

regulated is the Bcl2-associated athanogene 3 (BAG3, 1.73-fold), which has been previously found to be produced by glial cells in the developing brain and spinal cord of embryonic rats and seems to be important in the differentiation and maturation of specific neuronal populations.[38] The sorting nexin-18 (SNX18) is also up-regulated (1.62-fold) in the DRG. Sorting nexins are widely involved in the regulation of neuronal regeneration as they are involved in endocytotic pathways. Specifically, SNX18 has been associated to endosome recruitment and autophagosome formation.[39] This protein also plays an active role in the generation and growth of axons of spinal motor neurons in embryonic chicken and mice.[40] Neuromodulin (GAP43) is up-regulated (1.63-fold) as a result of axonal transport from the injury site.[41] This protein is associated with nerve regeneration and sprouting. Finally, up-regulation of both fibroblast growth factor-2 (FGF2, 1.37-fold) and the neural cell adhesion molecule-1 (NCAM1, 1.37-fold) have been implicated in the development of neurons in the DRG.[42-44]

The major down-regulated protein in the DRG is periaxin (PRX, -5.99-fold). This is a protein that has been associated to Schwann cells and myelination. Down-regulation of this protein implies axonal demyelination and it plays a central role in demyelinating neuropathies.[45-46]

Associated to Inflammation and Immune Responses - DRG

The most up-regulated complement protein we found in the DRG of injured animals relative to sham surgery animals was C4A (1.42-fold), which results from the cleavage of C4. C4A is involved in chemotaxis, vascular permeability and activation of mast cells. Other proteins involved in inflammatory process that are up-regulated in the DRG are annexin-4 (ANXA4, 2.18-fold), apolipoprotein D (APOD, 1.54-fold), the heat shock

protein-27 kDa (HSPB1, 1.41-fold), eosinophil peroxidase (EPX, 1.40-fold), integrin-β (ITGB2, 1.38-fold), the long-chain-fatty-acid--CoA ligase (ACSL4), and haptoglobin (HP, 1.33-fold). ANXA4, which is highly up-regulated in the DRG, is known to suppress the formation of the pro-inflammatory cytokine IL-8 (CXCL8).[47]

APOD mRNA has been shown to be up-regulated in the DRG after 7 days of establishing a chronic constriction injury model.[48] APOD is known to be secreted by glial cells in the peripheral nervous system to exert a role in axonal regeneration[49] and neuroprotection.[50] Similarly, overexpression of the heat shock protein, HSPB1, has been found to play a protective role in a transgenic mouse model for peripheral diabetic neuropathy.[51] Although the mechanism of action is not fully understood, it seems that HSPB1 disrupts the activation of pro-inflammatory pathways (RAGE, NFκB, caspase-3). It also seems that phosphorylation of HSPB1 is crucial for axon regeneration and regulation of interleukins IL-1 and IL-6.[52] EPX has recently been associated with the expression of cholinergic genes upon the interaction of eosinophils and neurons. It has been suggested that EPX induces neural plasticity via a specific neural receptor.[53] Integrins, such as ITGB2, have been associated to the regulation of hyperalgesia in animal models.[54] Their role is likely related to the upregulation of P2X4 receptors in microglia, in which the beta-integrins are involved in adhesion, signaling and motility.[55-56] Other key proteins highly up-regulated in the DRG, which are associated to the immune system, are the Rano class II histocompatibility antigen B-1 β chain (RT1-Bb, 2.48-fold) and the MHC class II antigen RT1.D β chain (RT1-Db, 1.63-fold). Both are involved in antigen presentation as a result of exogenous processes, although the role of these in neuropathic pain has not been established. Both of these are part of the MHC-II class of

antigens, which are associated to glial and macrophage activation.[57-58] Activation of glial cells, instrumental in pain development and maintenance,[59] is correlated to the up-regulation of glial fibrillary acid protein (GFAP, 1.47-fold), a well-established protein marker for this process.

Neuropathic Pain-Related Proteins - DRG

Our results indicate that no voltage-gated ion channel was significantly up or down-regulated when using hard threshold parameters (**Table 1**). Considering a softer threshold reveals that the voltage-gated potassium channel KCNA2 (Kv1.2) is up-regulated (1.39-fold), which is not consistent with the literature. Previous reports indicate that down-regulation of this channel in the ipsilateral DRG after spinal nerve ligation (SNL) and 7 days after sciatic nerve axotomy may be involved in the onset of neuropathic pain behavior in animal models.[60-61] Our study reveals that the most significantly down-regulated sodium channel in the DRG is SCN7A (Nax, 1.37-fold). SCN7A is expressed by glial cells and neurons as a sodium signaling protein (not voltage-gated), which is regulated by neuronal-glial interactions.[62] Although its role in neuropathic pain has not been established, a recent report links its increased expression in DRG with neuron hyperexcitability in an animal model for bone cancer pain.[63]

Other proteins regulated upon 4 days post-SNI which are involved in pain include the highly up-regulated glial cell derived neurotropic factor (**GDNF**) family receptor alpha (GFRA1, 1.57-fold). GFRA1 is a selective receptor for GDNF in relation to GDNF family members such as artemin, which selectively binds to the GFRA3.[64] Previous studies have demonstrated that peripheral nerve injury preferentially increases the expression of mRNA for this protein in the DRG of adult rats.[65-66] The mRNA for GFRA1 was increased

after L5 SNL, suggesting a relevant role for this receptor protein in the trophic properties of sensory neurons in the DRG. Also both mRNA and protein expression were increased upon chronic constriction injury (CCI) of the sciatic nerves of rats.[67] We found that neural cell adhesion molecule (NCAM1) was up-regulated as a result of the SNI. Sakai et al. demonstrated that (NCAM) plays a role in the analgesic effect of GDNF in animals that have undergone CCI.[68] They found that NCAM is expressed by small diameter neurons in the DRG, and that its expression is not reduced as a result of the CCI. NCAM signaling is activated by the binding of GNDF to GFRA1 after activation by a protein kinase (RET). NCAM is known to mediate cell adhesion and signal transduction and modulate plasticity changes of the nervous system by modifying the synapse.

Neuronal regeneration and degeneration - SC

Laminin proteins LAMB1 (1.41-fold), LAMA2 (1.39-fold), LAMC1 (1.35-fold) were up-regulated in the SC. These glycoproteins are constituents of the extracellular matrix, being the major non-collagen based constituents of the basement membranes.[69] They have been associated with regeneration and development of sensory neurons following neuropathic damage. Other proteins up-regulated and relevant to neuronal regeneration are the collagen COL3A1 (1.53-fold), and some myosin complex proteins, which were also regulated in the DRG. It is noteworthy to point out that periaxin (PRX) and COL3A1, which were proteins down-regulated in the DRG, were up-regulated in the SC upon SNI. The up-regulation of these proteins may be an indication of a protective mechanism in the SC in response to inflammation processes. The up-regulation of periaxin may prevent demyelination of neurons in the SC. Demyelination and undesired nerve outgrowth in the SC are processes that must be avoided by the CNS, so proteins that prevent these are

up-regulated. One of the most significant proteins down-regulated in the SC is the receptor-type tyrosine-protein phosphatase ζ (PTPRZ1, 1.48-fold). This protein is involved in regulating oligodendrocytes and axonal integrity upon inflammatory processes.

Proteins Associated to Inflammatory and Immune Responses - SC

Complement system proteins are up-regulated, particularly C1QB (1.66-fold), which is expressed by activated astrocytes. C1QB plays an important role in the synaptic cleft and is involved in the opsonization of materials within it.[70] Another important protein that was up-regulated is the microglial activation factor AIF1 (allograft inflammatory factor 1, also known as IBA1, 1.49-fold). As its name indicates, this protein is expressed by activated microglia as a result of the interaction of the glial cell with pro-inflammatory cytokines, particularly interferon (INF). We also found that BAG3 was highly up-regulated in the SC (6.63-fold) as well as in the DRG following SNI. This suggests that this protein, which has been involved in neuroprotection, plays an important role in the development of neuropathic pain. Other proteins involved in the inflammatory and immune response which are up-regulated include integrin β (ITGB2, 1.52-fold) and IgE Fc receptor subunit gamma (FCER1G, 1.52-fold).

Neuropathic Pain-Related Proteins - SC

The most highly up-regulated voltage-gated sodium channel is SCN9A (Nav1.7, 1.30-fold). The role of SCN9A in acute and inflammation pain is well established, but not in chronic pain.[71] Most of its expression has been accounted for at small diameter neurons (mostly C-fibers) in peripheral nerves and DRG. Recently, it was found that this ion channel is also expressed in terminals in the dorsal horn at the level of the presynaptic

terminals.[72] Some regulating proteins such as the MHC class I have been proposed to influence synaptic plasticity.[73] We found that the MHC Class I- RT1.Ab heavy chain protein was up-regulated (1.38-fold). We also found that Cathepsin S (CTSS, 1.63-fold) was significantly up-regulated in the SC following SNI. CTSS is a lysosomal enzyme which seems to play a key role in the development of chronic pain. Zhang et al. demonstrated that CTSS is expressed by immune T-cells upon antigen presentation, which upon infiltration into the dorsal horn activates microglia upon secretion of interferon-gamma.[74] We also observed the up-regulation of the ionotropic glutamate receptor protein (GRIN2D, 1.48-fold) and the α-2A adrenergic receptor (ADRA2A, 1.49-fold). These two receptors are widely known to play pivotal roles in the mechanism of neuropathic pain.[75-76]

Validation of BAG3 expression in the SC

Considering that BAG3 was the most highly up-regulated protein in the SC, western blot analysis was carried out to verify the up-regulation of this protein in the SNI animals relative to the SHAM animals. BAG-3 (Uniprot entry Q5U2U8 for Rattus norvegicus) is a 61 kDa protein made up of 574 amino acids. ROI analysis of the BAG3 protein from the SC tissue yielded an increased expression difference between the SNI and SHAM animals with a relative ratio of 1.40 (**Figure 3**).

Similarities and Differences in Responses of DRG and SC

Proteomic results from both tissues highlight the complexity of the mechanism by which pain is generated. Even though previous studies have carried out full proteomic analysis of nerve damage in animal models of neuropathic pain, none of these have looked at the levels of protein expression in both the SC and DRG 4 days after a

peripheral nerve injury; specifically the SNI model. The time point chosen represents an early stage in the establishment of neuropathic pain as it is known that stable hypersensitivity is reached 3 days after SNI surgery implying that expression of relevant proteins has reached peak levels.

As expected, the inflammatory process associated to the peripheral nerve injury has been translated and translocated at the level of the DRG and the SC by day 4 post-injury. Our results indicate that key proteins of the complement system, which are part of the innate immune response, are significantly up-regulated as a result of the axonal degeneration that follows nerve injury. The up-regulation of proteins from the complement system in the ipsilateral DRG was also reported by Komori et al. for a spinal nerve ligation injury.[25] Other proteins related to immune response processes such as annexins, apolipoprotein, integrin, heat-shock protein, cathepsins, and antigen-presenting proteins are also regulated upon injury. These are involved in regulatory process that either influence the chemotaxis of inflammatory cytokines or provide protection via pathways that involve leukocyte regulation. Up-regulation of complement proteins and immune cell infiltration in the DRG and SC after nerve injury implies an increase in synaptic pruning and recycling. It has been established that the regenerative process of the injured axon induces synaptic changes as a result of various factors including glial activation, macrophage migration and leukocyte access to the synapse.[77] These changes in the synapse play important roles in the development of chronic neuropathic pain and are key in the establishment of central sensitization.

We observed the expected evidence of the organism reacting to injury to the peripheral nerve by activating processes that lead to the regeneration and remyelination

of the damaged axon. The disruption of the neuromuscular junction by injury triggers the transport of molecules from the injury site to the soma and vice versa. The up-regulation of motor and filament proteins in the DRG and SC is indicative of a response by the central nervous system (**CNS**) to the peripheral injury. At first, we were puzzled by the increase of many of the myosin complex proteins, but felt confident of the proposed role as some of these were also identified by Komori et al. in the ipsilateral DRG following spinal nerve ligation. This implies that peripheral nerve injury involves proteins that are key transporters of vesicles and exosomes into and out of the CNS, promoting a cascade of changes that may lead to central sensitization that generates a neuropathic pain state.

The maintenance of the pain state is mostly linked to proteins directly involved in affecting synaptic plasticity and modulation. Although the mechanisms and key players in this process which leads to the persistence of pain have not been fully elucidated, our results are indicative of the important role that glial cell activation plays in synaptic modulation/plasticity which leads to central sensitization. For example, we found that both GFRA1 and NCAM are up-regulated in the DRG. These proteins are involved in the specific binding of GDNF. GDNF which is released by activated glial cells as a neuroprotective agent is known to have analgesic effects. The up-regulation of both GFRA1 and NCAM in the DRG imply a response of the CNS involving the activation of the glial cells as a neuroprotective response. Furthermore, the pathway analysis (PSEA-GO) implies that many pathways demonstrating significant regulation in both the DRG and SC involve proteins associated to the cell-substrate adhesion, and cellular reshaping, which indicates neuronal regeneration, perhaps concomitant to synaptic rearrangement, in both the DRG and SC. As mentioned above, glial activation in response to peripheral

nerve injury is conducive to the infiltration of the molecular machinery of the immune system and the regulation of proteins associated to it.

The most up regulated protein in the SC tissue is BAG3. Prior to our results, this protein has not been associated to the maintenance of pain. BAG3 is a protein of the Bcl2-assoctiated athanogenes known to be induced by stressful stimuli through the activity of a heat shock factor on the bag3 gene promoter to regulate many cell pathways.[78] BAG3 is a co-chaperone that when associated to the heat shock protein HSP70 plays an important role in autophagy and anti-apoptotic pathways.[79] The interaction of this complex with certain small heat shock proteins have been identified to regulate autophagy in the spinal cord of animal models for amyotrophic lateral sclerosis (**ALS**).[80] BAG3 was previously shown to be upregulated in a specific region of the rat hippocampus in response to transient forebrain ischemia. The protein was shown to originate from activated astrocytes[81] in response to the insult in what seems to be a neuroprotective manner. The up-regulation of BAG3 in the ipsilateral SC and associated DRG during early stages of the SNI model of neuropathic chronic pain is likely related to glial activation and the neuroprotective response of this to the corresponding peripheral nerve injury. Indeed, similar to the previous study, the origin of the protein could be glial, as the up-regulation of BAG3 is coincident with the upregulation of GFAP in the DRG and integrin beta in the SC. Further studies will be needed to elucidate the mechanism by which BAG3 may exert its neuroprotective role in the SC and DRG tissues associated to peripheral nerve injury.

Another protein, which is up-regulated in the SC and is involved in the synapse is GRIN2D, the ionotropic glutamate (NMDA) receptor, which mediate excitatory synaptic

transmission in response to glutamate release by activated glial cells.[76] Similarly, ADRA2A, the α-2A adrenergic receptor is up-regulated in the SC. This is involved in regulation of neurotransmitter release in the synapse via adrenergic neurons and the involvement of epinephrine. An increase of this receptor has been correlated to an increase in the excitatory state of neurons in the SC which is induced by the release of cytokines and other factors as a result of peripheral sensitization.[75]

Conclusion

Our observations are supportive of mechanisms involving the glial neuroprotective response affecting protein expression in the DRG and SC for the establishment of neuropathic pain upon peripheral nerve injury. Axonal damage in the tibial and peroneal nerves induces a response in the afferent direction, which is manifested in an immune and consequent inflammatory response by the ipsilateral DRG and dorsal section of the spinal cord associated to the injured nerve. Processes involved in neuronal degeneration and demyelination are evident in the DRG, while the SC presents with up-regulation of proteins that imply a neuroprotective response. The immune and inflammatory responses in the SC and DRG seem to involve anterograde transport of various factors which may correspond to glial cell activation along the periphery. Peripheral nerve injury induces the glial activation in the SC and DRG and this process also causes local immune responses. Some ion channel proteins involved in the propagation of ectopic activity in both the DRG and SC were also regulated, but more importantly, proteins that are indicative of synaptic reshaping were identified. Synaptic changes are influenced by the innate immune system and also by the activation of the glial cells that monitor the synapse. The activation of the glial cells also induces the up-regulation of receptors that promote excitatory neuronal

states. Synaptic changes ultimately lead to central sensitization and the neuropathic pain phenotype.

In summary, optimized chromatographic separation and quantitative mass spectrometric analysis provide the first proteome analysis of the DRG and SC corresponding to the injured nerve in an animal model of chronic neuropathic pain. The analysis provides a deep glimpse of proteins that are involved in the establishment of a chronic pain state and provides confirmation of proposed mechanisms. The analysis emphasizes the role of glial cell activation and underlines the adaptive degenerative process at the synapses, although it also implies that there is indication of neuroprotective processes at the SC. Future studies will explore temporal effects, the effect of electromagnetic stimulation in either the periphery or central nervous system, and the phosphoproteome as many kinases and ATP-related proteins are also involved.

ACKNOWLEDGEMENTS

This work was partially supported by the National Institutes of Health grant (1R01DK071801 to L.L.). The Orbitrap Elite instrument was purchased through the support of an NIH shared instrument grant (NIH-NCRR S10RR029531 to L.L.). C.B.L. acknowledges an NIH-supported Chemistry Biology Interface Training Program Predoctoral Fellowship (grant number T32-GM008505) and an NSF Graduate Research Fellowship (DGE-1256259). L.L. acknowledges an H. I. Romnes Faculty Research Fellowship.

References

(1) Toth, C., Lander, J., and Wiebe, S. *Pain Med.* **2009** *10*, 918-929.

(2) Institute of Medicine (US) Committee on Advancing Pain Research, C., and Education. (2011) *Relieving Pain in America: A Blueprint for Transforming*

Prevention, Care, Education, and Research, National Academies Press (US), Washington (DC).

(3) NIH. (2014) Estimates of Funding for Various Research, Condition, and Disease Categories (RCDC), U.S. Department of Health & Human Services, Research Portfolio Online Reporting Tools (RePORT).

(4) von Hehn, C. A., Baron, R., and Woolf, C. J. *Neuron* **2012** *73*, 638-652.

(5) Huo, F. Q., Chen, T., Lv, B. C., Wang, J., Zhang, T., Qu, C. L. *et al. Cereb. Cortex* **2009** *19*, 1263-1272.

(6) Woolf, C. J. *Pulm. Pharmacol.* **1995** *8*, 161-167.

(7) Gao, L. L., McMullan, S., Djouhri, L., Acosta, C., Harper, A. A., and Lawson, S. N. *J. Physiol.* **2012** *590*, 4691-4705.

(8) Emery, E. C., Young, G. T., Berrocoso, E. M., Chen, L., and McNaughton, P. A. *Science* **2011** *333*, 1462-1466.

(9) Rahman, W., Bauer, C. S., Bannister, K., Vonsy, J. L., Dolphin, A. C., and Dickenson, A. H. *Mol. Pain* **2009** *5*, 45.

(10) Braz, J. M., and Basbaum, A. I. *Pain* **2010** *150*, 290-301.

(11) Kim, D. S., Figueroa, K. W., Li, K. W., Boroujerdi, A., Yolo, T., and Luo, Z. D. *Pain* **2009** *143*, 114-122.

(12) Vallejo, R., Tilley, D. M., Williams, J., Labak, S., Aliaga, L., and Benyamin, R. M. *Pain Physician* **2013** *16*, E601-613.

(13) Zhang, Y., Fonslow, B. R., Shan, B., Baek, M. C., and Yates, J. R., 3rd. *Chem. Rev.* **2013** *113*, 2343-2394.

(14) Zubarev, R. A. *PROTEOMICS* **2013** *13*, 723-726.

(15) Kocher, T., Swart, R., and Mechtler, K. *Anal. Chem.* **2011** *83*, 2699-2704.

(16) de Godoy, L. M., Olsen, J. V., Cox, J., Nielsen, M. L., Hubner, N. C., Frohlich, F. *et al. Nature* **2008** *455*, 1251-1254.

(17) Richards, A. L., Merrill, A. E., and Coon, J. J. *Curr. Opin. Chem. Biol.* **2015** *24*, 11-17.

(18) Beck, M., Schmidt, A., Malmstroem, J., Claassen, M., Ori, A., Szymborska, A. *et al. Mol. Syst. Biol.* **2011** *7*, 549.

(19) Nagaraj, N., Wisniewski, J. R., Geiger, T., Cox, J., Kircher, M., Kelso, J. *et al. Mol. Syst. Biol.* **2011** *7*, 548.

(20) Bayes, A., and Grant, S. G. *Nat. Rev. Neurosci.* **2009** *10*, 635-646.

(21) Dowell, J. A., Johnson, J. A., and Li, L. *J. Proteome Res.* **2009** *8*, 4135-4143.

(22) Sun, F., and Cavalli, V. *Mol. Cell. Proteomics* **2010** *9*, 963-975.

(23) Cunningham, R., Jany, P., Messing, A., and Li, L. *J. Proteome Res.* **2013** *12*, 719-728.

(24) Craft, G. E., Chen, A., and Nairn, A. C. *Methods* **2013** *61*, 186-218.

(25) Komori, N., Takemori, N., Kim, H. K., Singh, A., Hwang, S. H., Foreman, R. D. *et al. Physiol. Genomics* **2007** *29*, 215-230.

(26) Riedl, M. S., Braun, P. D., Kitto, K. F., Roiko, S. A., Anderson, L. B., Honda, C. N. *et al. J. Neurosci.* **2009** *29*, 13377-13388.

(27) Melemedjian, O. K., Yassine, H. N., Shy, A., and Price, T. J. *Mol. Pain* **2013** *9*, 14.

(28) Xu, W., Wu, Y., Bi, Y., Tan, L., Gan, Y., and Wang, K. *Mol. Pain* **2010** *6*, 49.

(29) Sui, P., Watanabe, H., Ossipov, M. H., Bakalkin, G., Artemenko, K., and Bergquist, J. *J. Proteome Res.* **2014** *13*, 3957-3965.

(30) Tilley, D. M., Vallejo, R., Kelley, C. A., Benyamin, R., and Cedeno, D. L. *Neuromodulation* **2015**.

(31) Choi, Y., Yoon, Y. W., Na, H. S., Kim, S. H., and Chung, J. M. *Pain* **1994** *59*, 369-376.

(32) Lietz, C. B., Yu, Q., and Li, L. *J. Am. Soc. Mass Spectrom.* **2014** *25*, 2009-2019.

(33) Lavallee-Adam, M., Rauniyar, N., McClatchy, D. B., and Yates, J. R., 3rd. *J. Proteome Res.* **2014** *13*, 5496-5509.

(34) Rauniyar, N., and Yates, J. R. *J. Proteome Res.* **2014** *13*, 5293-5309.

(35) Cha, S., Imielinski, M. B., Rejtar, T., Richardson, E. A., Thakur, D., Sgroi, D. C. *et al. Mol. Cell. Proteomics* **2010** *9*, 2529-2544.

(36) Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A. *et al. Proc. Natl. Acad. Sci. U.S.A.* **2005** *102*, 15545-15550.

(37) Bridgman, P. C. *J. Neurobiol.* **2004** *58*, 164-174.

(38) Choi, J. S., Lee, J. H., Shin, Y. J., Lee, J. Y., Yun, H., Chun, M. H. *et al. Cell Tissue Res.* **2009** *337*, 27-36.

(39) Knaevelsrud, H., Soreng, K., Raiborg, C., Haberg, K., Rasmuson, F., Brech, A. *et al. J. Cell Biol.* **2013** *202*, 331-349.

(40) Nakazawa, S., Gotoh, N., Matsumoto, H., Murayama, C., Suzuki, T., and Yamamoto, T. *J. Histochem. Cytochem.* **2011** *59*, 202-213.

(41) Van der Zee, C. E., Nielander, H. B., Vos, J. P., Lopes da Silva, S., Verhaagen, J., Oestreicher, A. B. *et al. J. Neurosci.* **1989** *9*, 3505-3512.

(42) Allodi, I., Mecollari, V., Gonzalez-Perez, F., Eggers, R., Hoyng, S., Verhaagen, J. *et al. Glia* **2014** *62*, 1736-1746.

(43) Allodi, I., Casals-Diaz, L., Santos-Nogueira, E., Gonzalez-Perez, F., Navarro, X.,

and Udina, E. *Mol. Neurobiol.* **2013** *47*, 770-781.

(44) Chuong, C. M., and Edelman, G. M. *J. Neurosci.* **1984** *4*, 2354-2368.

(45) Zhou, L., and Griffin, J. W. *Curr. Opin. Neurol.* **2003** *16*, 307-313.

(46) Gillespie, C. S., Sherman, D. L., Fleetwood-Walker, S. M., Cottrell, D. F., Tait, S., Garry, E. M. *et al. Neuron* **2000** *26*, 523-531.

(47) Iwasa, T., Takahashi, R., Nagata, K., and Kobayashi, Y. *Biochim. Biophys. Acta* **2012** *1822*, 204-211.

(48) Kim, D. S., Lee, S. J., Park, S. Y., Yoo, H. J., Kim, S. H., Kim, K. J. *et al. Neuroreport.* **2001** *12*, 3401-3405.

(49) Ganfornina, M. D., Do Carmo, S., Martinez, E., Tolivia, J., Navarro, A., Rassart, E. *et al. Glia* **2010** *58*, 1320-1334.

(50) Bajo-Graneras, R., Ganfornina, M. D., Martin-Tejedor, E., and Sanchez, D. *Glia* **2011** *59*, 1551-1566.

(51) Korngut, L., Ma, C. H., Martinez, J. A., Toth, C. C., Guo, G. F., Singh, V. *et al. Neurobiol. Dis.* **2012** *47*, 436-443.

(52) Tanabe, K., Matsushima-Nishiwaki, R., Dohi, S., and Kozawa, O. *Neuroscience* **2010** *170*, 1028-1034.

(53) Akasheh, N., Walsh, M. T., and Costello, R. W. *Mol. Immunol.* **2014** *62*, 37-45.

(54) Dina, O. A., Hucho, T., Yeh, J., Malik-Hall, M., Reichling, D. B., and Levine, J. D. *Pain* **2005** *115*, 191-203.

(55) Nutile-McMenemy, N., Elfenbein, A., and Deleo, J. A. *J. Neurochem.* **2007** *103*, 2035-2046.

(56) Tsuda, M., Toyomitsu, E., Komatsu, T., Masuda, T., Kunifusa, E., Nasu-Tada, K. *et al. Glia* **2008** *56*, 579-585.

(57) Hu, P., and McLachlan, E. M. *Exp. Neurol.* **2003** *184*, 590-605.

(58) Zhang, S. C., Goetz, B. D., Carre, J. L., and Duncan, I. D. *Glia* **2001** *34*, 101-109.

(59) Watkins, L. R., Milligan, E. D., and Maier, S. F. *Trends Neurosci.* **2001** *24*, 450-455.

(60) Fan, L., Guan, X., Wang, W., Zhao, J. Y., Zhang, H., Tiwari, V. *et al. Mol. Pain* **2014** *10*, 8.

(61) Kim, D. S., Choi, J. O., Rim, H. D., and Cho, H. J. *Brain research. Mol. Brain Res.* **2002** *105*, 146-152.

(62) Gautron, S., Gruszczynski, C., Koulakoff, A., Poiraud, E., Lopez, S., Cambier, H. *et al. Glia* **2001** *33*, 230-240.

(63) Ke, C. B., He, W. S., Li, C. J., Shi, D., Gao, F., and Tian, Y. K. *Neuroscience* **2012** *227*, 80-89.

(64) Carmillo, P., Dago, L., Day, E. S., Worley, D. S., Rossomando, A., Walus, L. *et al. Biochemistry* **2005** *44*, 2545-2554.

(65) Hoke, A., Cheng, C., and Zochodne, D. W. *Neuroreport.* **2000** *11*, 1651-1654.

(66) Kashiba, H., Uchida, Y., and Senba, E. *Brain research. Mol. Brain Res.* **2003** *110*, 52-62.

(67) Dong, Z. Q., Ma, F., Xie, H., Wang, Y. Q., and Wu, G. C. *Neurosci. Lett.* **2005** *376*, 143-148.

(68) Sakai, A., Asada, M., Seno, N., and Suzuki, H. *Pain* **2008** *137*, 378-388.

(69) Gardiner, N. J. *Dev. Neurobiol.* **2011** *71*, 1054-1072.

(70) Heikkinen, A., Pihlajaniemi, T., Faissner, A., and Yuzaki, M. *Prog. Brain Res.* **2014** *214*, 29-51.

(71) Wang, W., Gu, J., Li, Y. Q., and Tao, Y. X. *Mol. Pain* **2011** *7*, 16.

(72) Black, J. A., Frezel, N., Dib-Hajj, S. D., and Waxman, S. G. *Mol. Pain* **2012** *8*, 82.

(73) Freria, C. M., Zanon, R. G., Santos, L. M., and Oliveira, A. L. *The J. Comp. Neurol.* **2010** *518*, 990-1007.

(74) Zhang, X., Wu, Z., Hayashi, Y., Okada, R., and Nakanishi, H. *J. Neurosci.* **2014** *34*, 3013-3022.

(75) Fairbanks, C. A., Stone, L. S., and Wilcox, G. L. *Pharmacol. Ther.* **2009** *123*, 224-238.

(76) Larsson, M., and Broman, J. *Neuroscientist* **2011** *17*, 256-273.

(77) Spejo, A. B., and Oliveira, A. L. *Neuropharmacol.* **2014**.

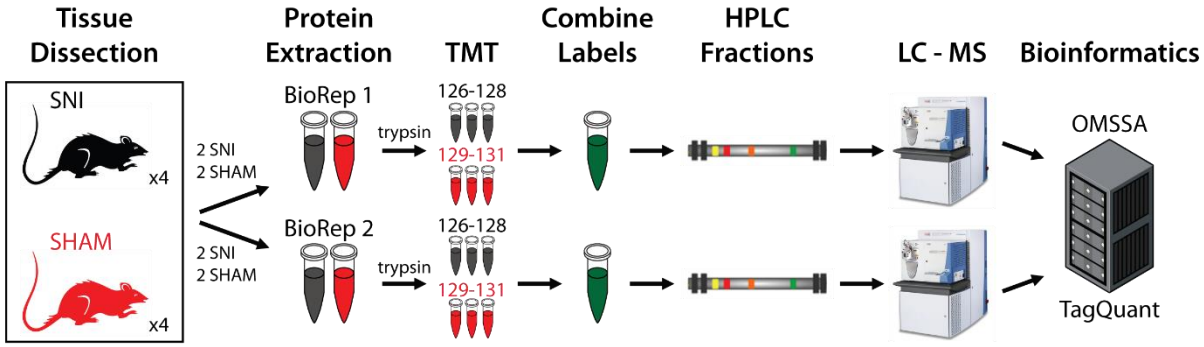
(78) Rosati, A., Graziano, V., De Laurenzi, V., Pascale, M., and Turco, M. C. *Cell Death Dis.* **2011** *2*, e141.

(79) Gamerdinger, M., Kaya, A. M., Wolfrum, U., Clement, A. M., and Behl, C. *EMBO Reports* **2011** *12*, 149-156.

(80) Crippa, V., Boncoraglio, A., Galbiati, M., Aggarwal, T., Rusmini, P., Giorgetti, E. *et al. Front. Cell. Neurosci.* **2013** *7*, 234.

(81) Lee, M. Y., Kim, S. Y., Shin, S. L., Choi, Y. S., Lee, J. H., Tsujimoto, Y. *et al. Exp. Neurol.* **2002** *175*, 338-346.

Scheme 1. Quantitative proteomic workflow. SC and DRG analyses were performed separately.



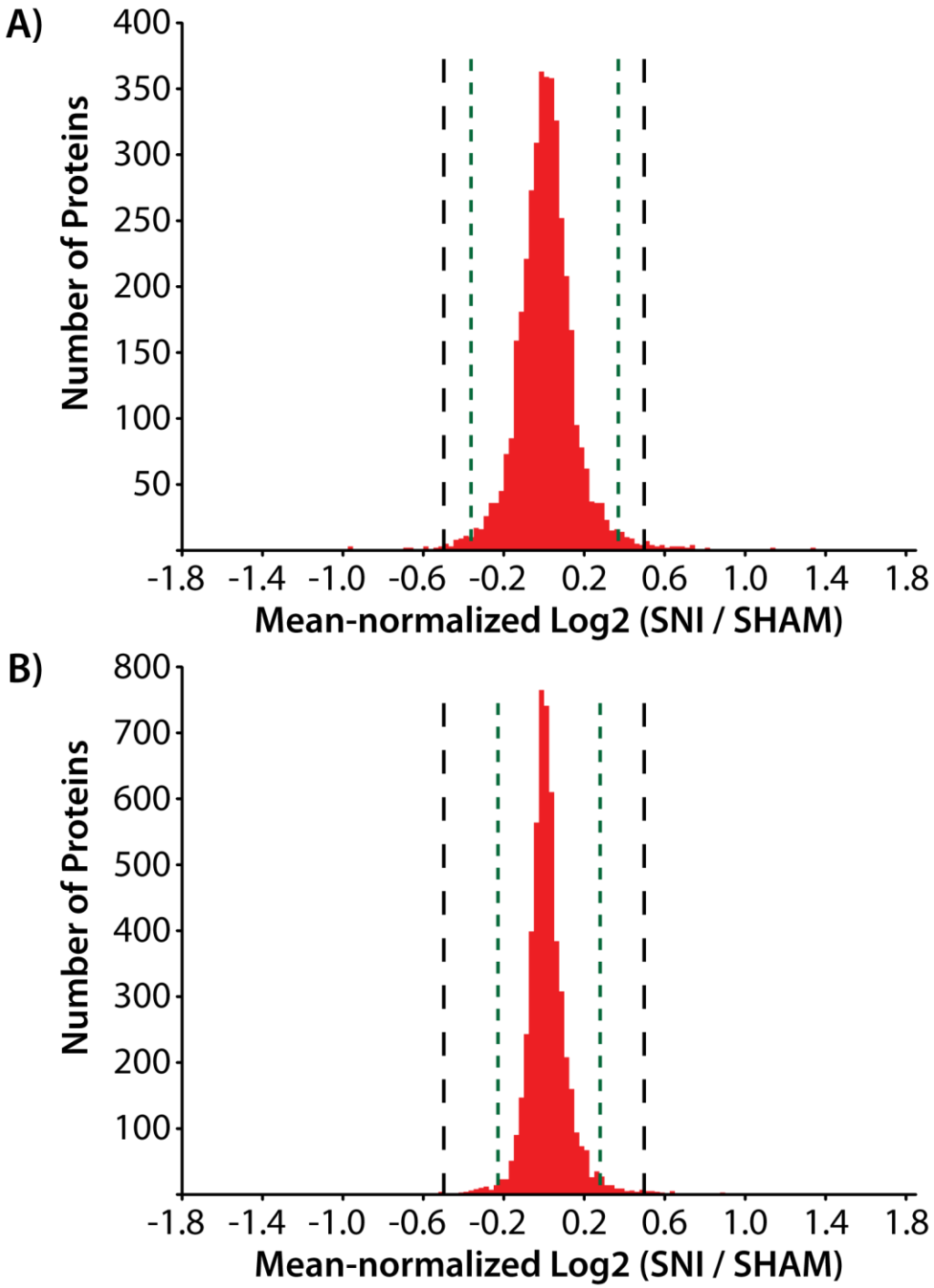


Figure 1. Histograms plotting the number of identified proteins with various Log2 SNI/SHAM ratios in (A) DRG and (B) SC tissue. The black and green dotted lines mark the hard and soft significance thresholds, respectively. Three DRG proteins and two SC proteins with ratios outside ± 1.8 were omitted from the figure for clarity.

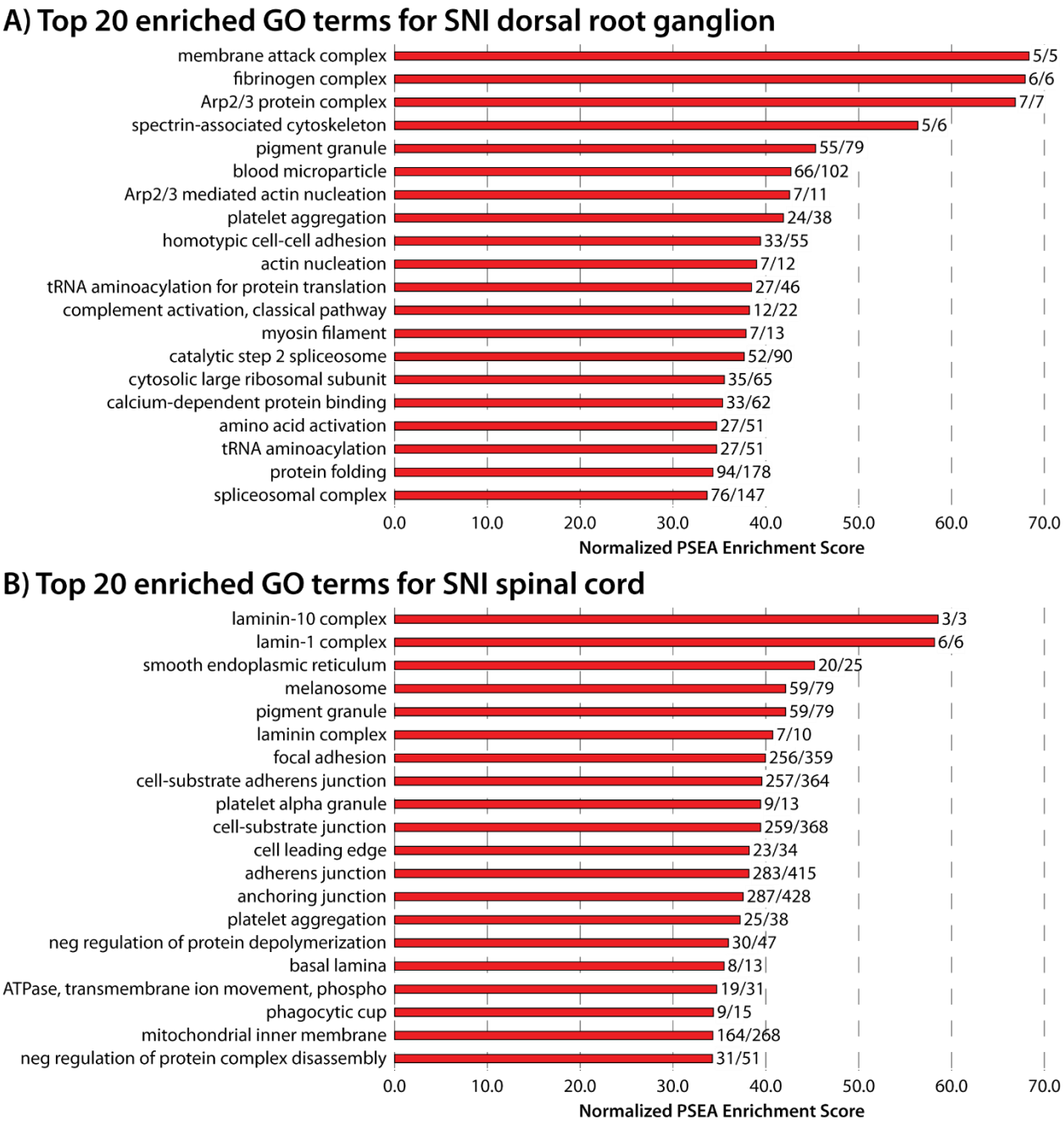


Figure 2. Bar graphs of the top 20 enriched SNI GO terms for (A) DRG and (B) SC, followed by the ratio of proteins identified that term over the total number of proteins with that GO term in the database. All listed GO terms have an FDR Q-value ≤ 0.1.

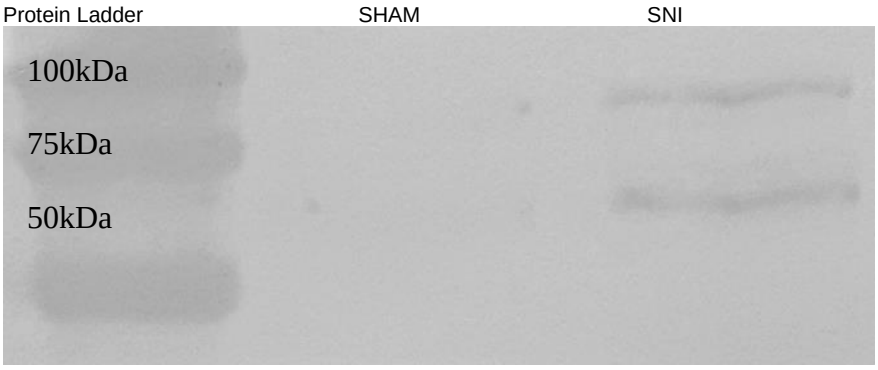


Figure 3. Western blot validation for BAG3 presence within the SC tissue for SHAM and SNI tissues. The heavier band (higher in lane) is another isoform of BAG and lighter band (lower in lane) is for BAG3; 61.5kDa. There is a clear increase in expression in SNI animals relative to SHAM with an average SNI/SHAM ratio of 1.40.

Table 1. List of proteins significantly upregulated and downregulated (hard cutoff: -0.50 ≤ Log_2(SNI/SHAM) ≤ 0.50) FDR p-values ≤ 0.05) in the DRG of rats after 4 days post-spared nerve injury (SNI) relative to sham surgery (SHAM).

Upregulated in SNI				
Uniprot Entry	**Protein Name**	**Gene name**	$log_2\left[\frac{SNI}{Sham}\right]$	**Quantified PSMs**
Q9QZV8	Type 2X myosin heavy chain	Myh1	3.11	3
Q71DI1	Dermcidin		1.61	5
Q7TMA9	Aa1249 - C reactive protein	Crp	1.57	7
Q304F3	Protein Tnnc2 (Troponin C)	Tnnc2	1.32	1
P29826	Rano class II histocompatibility antigen, B-1 β chain	RT1-Bb	1.31	3
P02600	Myosin light chain 1/3,	Myl1	1.22	23
P09739	Troponin T, fast skeletal muscle	Tnnt3	1.19	4
Q5U362	Annexin A4	Anxa4	1.12	26
F1LMU0	Myosin-4	Myh4	1.10	43
G3V8B0	Myosin-7	Myh7	1.05	3
F1LX60	Uncharacterized protein		1.03	3
P04466	Myosin regulatory light chain 2	Mylpf	1.02	24
D3ZZQ0	Protein Tnik	Tnik	0.90	3
P58775	Tropomyosin β chain	Tpm2	0.88	4
P47967	Galectin-5	Lgals5	0.83	6
P00564	Creatine kinase M-type	Ckm	0.79	69
Q5U2U8	Bcl2-associated athanogene 3	Bag3	0.79	7
F1LRV9	Protein Myh1 (myosin heavy chain)	Myh1	0.72	10
Q5DKU0	MHC class II antigen RT1.D β chain		0.70	4
P07936	Neuromodulin (Gap43)	Gap43	0.70	40
D3ZZ38	Sorting nexin	Snx18	0.70	7
P15429	β-enolase	Eno3	0.69	25
P08932	T-kininogen 2		0.67	27
Q68G32	Ugt1a7c protein	Ugt1a7c	0.66	1
Q6AXN8	Protein Zscan2	Zscan21	0.65	1
Q62997	GDNF family receptor α-1	Gfra1	0.65	3
Q3KRE8	Tubulin β-2B chain	Tubb2b	0.65	6
P16409	Myosin light chain 3	Myl3	0.63	2
M0R4S2	Apolipoprotein D	Apod	0.62	8
P16290	Phosphoglycerate mutase 2	Pgam2	0.61	6
Q9QZ76	Myoglobin	Mb	0.60	9
E9PT48	Microtubule-associated protein	Mapt	0.59	1
J7JVB9	Mx2 (Interferon-induced GTP-binding protein)		0.59	6
P52944	PDZ and LIM domain protein 1	Pdlim1	0.56	20
F7FL53	Slc43a1	Slc43a1	0.56	4
P10759	AMP deaminase 1	Ampd1	0.56	2
P47819	Glial fibrillary acidic protein	Gfap	0.55	48
D3ZA38	Myosin binding protein C, fast-type	Mybpc2	0.55	4
P12928	Pyruvate kinase PKLR	Pklr	0.54	2
Q4G075	Leukocyte elastase inhibitor A	Serpinb1a	0.53	25

B2GVB3	Actn3 protein (Actinin, alpha 3)	Actn3	0.52	8
Q5XI77	Annexin A11	Anxa11	0.52	17
Q6MG79	Complement component 4A	C4a	0.51	2
Downregulated in SNI				
Uniprot Entry	**Protein Name**	**Gene name**	$log_2\left[\frac{SNI}{Sham}\right]$	**Quantified PSMs**
G3V8D2	Periaxin	Prx	-2.58	1
Q5XI60	Receptor expression-enhancing protein 6	Reep6	-1.30	2
Q10758	Keratin, type II cytoskeletal 8	Krt8	-1.08	3
M0R983	Protein LOC688320	LOC688320	-0.98	13
D3ZBL6	Protein Nup160 (Nuclear pore complex protein 160)	Nup160	-0.98	5
Q4KLJ0	High mobility group nucleosomal binding domain 2	LOC100360316	-0.96	6
D3ZH71	Protein Tyrp1 (tyrosine transporter?)	Tyrp1	-0.92	5
P35559	Insulin-degrading enzyme	Ide	-0.82	14
F1M5Q4	Protein Fbn2 (fibrillin 2)	Fbn2	-0.75	2
D4A7Q6	Protein Zfp428 (zinc finger protein 428)	Zfp428	-0.71	2
G3V822	Carboxylesterase 1D	Ces1d	-0.68	1
Q68FR2	Bridging integrator 2	Bin2	-0.68	8
D4A516	Protein Zfp287 (zinc finger protein 287)	Zfp287	-0.66	1
P52590	Nuclear pore complex protein Nup107	Nup107	-0.63	2
G3V628	Protein RGD1302996	RGD1302996	-0.62	2
Q499S4	Interferon-induced GTP-binding protein Mx1	Mx1	-0.61	1
P09456	cAMP-dependent protein kinase type I-α regulatory subunit	Prkar1a	-0.59	14
G3V7V8	Phosphodiesterase 5A, cGMP-specific, isoform CRA_c	Pde5a	-0.56	4
Q91W30	Aldose reductase-like protein	Akr1b8	-0.55	3
P18437	Non-histone chromosomal protein HMG-17	Hmgn2	-0.54	35
D3ZCX4	Protein LOC688875	LOC688875	-0.53	2
Q2KN99	Cytospin-A	Specc1l	-0.52	2
G3V9M6	Fibrillin 1, isoform CRA_a	Fbn1	-0.51	9
Q4QQS6	Asparagine-linked glycosylation 5 homolog	Alg5	-0.51	3

Table 2. List of proteins significantly upregulated and downregulated (hard cutoff: -0.50 ≤ Log_2(SNI/SHAM) ≤ 0.50) FDR p-values ≤ 0.05) in the SC of rats after 4 days post-spared nerve injury (SNI) relative to sham surgery (SHAM).

Upregulated in SNI				
Uniprot Entry	**Protein Name**	**Gene name**	$log_2\left[\frac{SNI}{Sham}\right]$	**Quantified PSMs**
Q5U2U8	Bcl2-associated athanogene 3	Bag3	2.73	6
Q9QZV8	Type 2X myosin heavy chain	Myh1	2.08	1
P47967	Galectin-5	Lgals5	1.29	1
G3V8D2	Periaxin	Prx	1.22	5
F1LMU0	Myosin-4	Myh4	1.05	29
P02600	Myosin light chain 1/3	Myl1	0.96	17
F1LX60	Uncharacterized protein	-	0.91	1
D3Z9U7	Protein Zc3h4 (zinc finger CCCH-type containing 4)	Zc3h4	0.87	1
P16409	Myosin light chain 3	Myl3	0.86	2
F1M4V3	Protein Rcsd1	Rcsd1	0.76	3
G3V7N9	Complement C1q subcomponent subunit B	C1qb	0.73	4
Q68FX4	Hematopoietic cell specific Lyn substrate 1	Hcls1	0.71	3
D3ZZR3	Cathepsin S	Ctss	0.70	1
B2GVB3	Actn3 protein	Actn3	0.69	3
D3Z955	Protein Pgm2l1 (phosphoglucomutase 2-like 1)	Pgm2l1	0.67	19
G3V818	α-parvin	Parva	0.62	1
Q62714	Neutrophil antibiotic peptide NP-4	Np4	0.61	5
P13941	Collagen α-1(III) chain	Col3a1	0.61	5
P20411	IgE Fc receptor subunit γ	Fcer1g	0.60	2
F7F4S8	Integrin β	Itgb2	0.60	11
G3V729	Bone marrow proteoglycan	Prg2	0.58	4
P55009	Microglia response factor (allograft inflammatory factor 1)	Aif1	0.58	11
M0R9R3	α-2A adrenergic receptor	Adra2a	0.57	4
F1LSC6	Glutamate receptor ionotropic	Grin2d	0.56	1
G3V904	Protein Pld4 (similar to phosphlipase D)	Pld4	0.56	6
P04466	Myosin regulatory light chain 2	Mylpf	0.55	6
P52925	High mobility group protein B2	Hmgb2	0.54	5
G3V9T9	Tyrosine-protein phosphatase non-receptor type	Ptpn6	0.53	7
B2GVA1	Selenoprotein O	Selo	0.53	3
Q68FR2	Bridging integrator 2	Bin2	0.52	4
Q9R1T3	Cathepsin Z	Ctsz	0.52	12
G3V9B3	Myelin-associated glycoprotein	Mag	0.51	5
Q5M860	Protein Arhgdib (Rho GDP dissociation inhibitor beta)	Arhgdib	0.51	6
F1LQ00	Protein Col5a2 (collagen, type V, alpha 2)	Col5a2	0.51	3
Q8VI02	Ser/Thr-protein phosphatase 4 regulatory subunit 1	Ppp4r1	0.50	2

Downregulated in SNI				
Uniprot Entry	**Protein Name**	**Gene name**	$log_2\left[\frac{SNI}{Sham}\right]$	**Quantified PSMs**
Q68A21	Transcriptional activator protein Pur-β	Purb	-2.54	1
Q08290	Calponin-1	Cnn1	-1.76	2
Q63862	Myosin-11	Myh11	-1.10	1
E9PTU4	Myosin-11	Myh11	-0.86	25
D3ZSM0	Uncharacterized protein	-	-0.74	1
D3ZNZ9	Histone H2B	Hist3h2ba	-0.73	5
Q9Z1Z6	Integrin-linked kinase-associated Ser/Thr phosphatase 2C	Ilkap	-0.72	2
P31232	Transgelin	Tagln	-0.69	60
Q156J1	Bcl-2-interacting death suppressor	-	-0.65	4
G3V637	Syntaxin binding protein 2, isoform CRA_b	Stxbp2	-0.62	1
Q7M0B6	Glutathione transferase	-	-0.61	1
G3V8B3	Histone H2B	LOC100910200	-0.59	20
F1LMY3	Receptor-type Tyr-protein phosphatase ζ	Ptprz1	-0.57	1
M0RA22	Uncharacterized protein	-	-0.55	3
D3ZFU9	Protein Mylk (myosin light-chain kinase)	Mylk	-0.53	16
Q6RKB2	Fasciclin II GPI-linked protein isoform	-	-0.53	2
Q5FVG5	Similar to tropomyosin 1	Tpm2	-0.53	39
P63255	Cysteine-rich protein 1	Crip1	-0.52	9

Supplemental Information

Supplemental Information 1

High Performance Liquid Chromatography (HPLC) Fractionation

Samples were reconstituted in solvent A (10 mM KH2PO4, 20% acetonitrile, 0.17% H3PO4) and separated on a polysulfoethyl A strong cation-exchange column (200 x 2.1 mm, 5 µm, 300 Å, PolyLC Inc., Columbia, MD, USA). The column temperature was set to 30 °C and the flow rate was kept at 0.2 mL min-1. UV absorbance was monitored at 215 nm. The solvent gradient began with 100% solvent A to load and trap the sample. Separation occurred by ramping solvent B (10 mM KH2PO4, 500 mM KCl, 20% acetonitrile, 0.17% H3PO4) to 33% over 75 min, and then to 100% solvent B over the following 25 min.

Liquid Chromatography-Mass Spectrometry (LC-MS) Acquisition

The LC began by injecting 3.75 µL of sample and loading it on the column with 100% solvent A (water, 5% dimethylsulfoxide, 0.1% FA) for 15 minutes. Solvent B (acetonitrile, 5% dimethylsulfoxide, 0.1% FA) was then ramped up according to following gradient: 0% – 4% B during 0.0 – 0.1 minutes, 4% – 12% B during 0.1 – 36.5 minutes, 12% – 22% B during 36.5 – 68.5 minutes, and 22% – 30% B during 68.5 – 80.0 minutes. The column was then washed for 10 minutes at 70% B and 10 minutes at 95% B.

Eluting peptides were electrosprayed into the mass spectrometer and sequenced via top 15 datadependent tandem MS (MS/MS) cycles. The MS resolution and automatic gain control (AGC) target was set to 30000 and 1 x 10^6, respectively. Dynamic exclusion duration was set to ±10 ppm windows for 40 seconds. Only ions with a charge of +2 or greater were isolated for MS/MS, and selected peptide ions were fragmented by a normalized high-energy collision dissociation (HCD) collision energy of 35%. The MS/MS

resolution, AGC target, isolation width, and first mass was set to 15000, 1 x 10^5, 2.0 Th, and 120 m/z, respectively.

Protein Identification and Quantification

OMSSA precursor and fragment mass tolerances limited to 50 ppm and 0.02 daltons (Da), respectively. Multiple charged fragment ions were considered in all spectra. Carbamidomethyl cysteine, TMT N-terminus, and TMT lysine were searched as fixed modifications, while oxidized methionine and TMT tyrosine were searched as variable modifications.

TMT analysis was performed by the COMPASS Tag Quant feature. First, purity-corrected, normalized TMT reporter intensities from peptide spectrum matches (PSMs) of each protein were summed. Next, the mean intensity of TMT reporters 126 – 128 was taken as the protein's SNI abundance, and the mean intensity of TMT reporters 129 – 131 was taken as the protein's SHAM abundance. An example of protein-level quantification is shown in **Figure S1**. The SNI/SHAM abundance ratio was then calculated for each protein, normalized by the mean SNI/SHAM ratio across all proteins, and transformed onto a Log2-scale. P-values for each protein ratio were calculated with a heteroscedastic t-test.

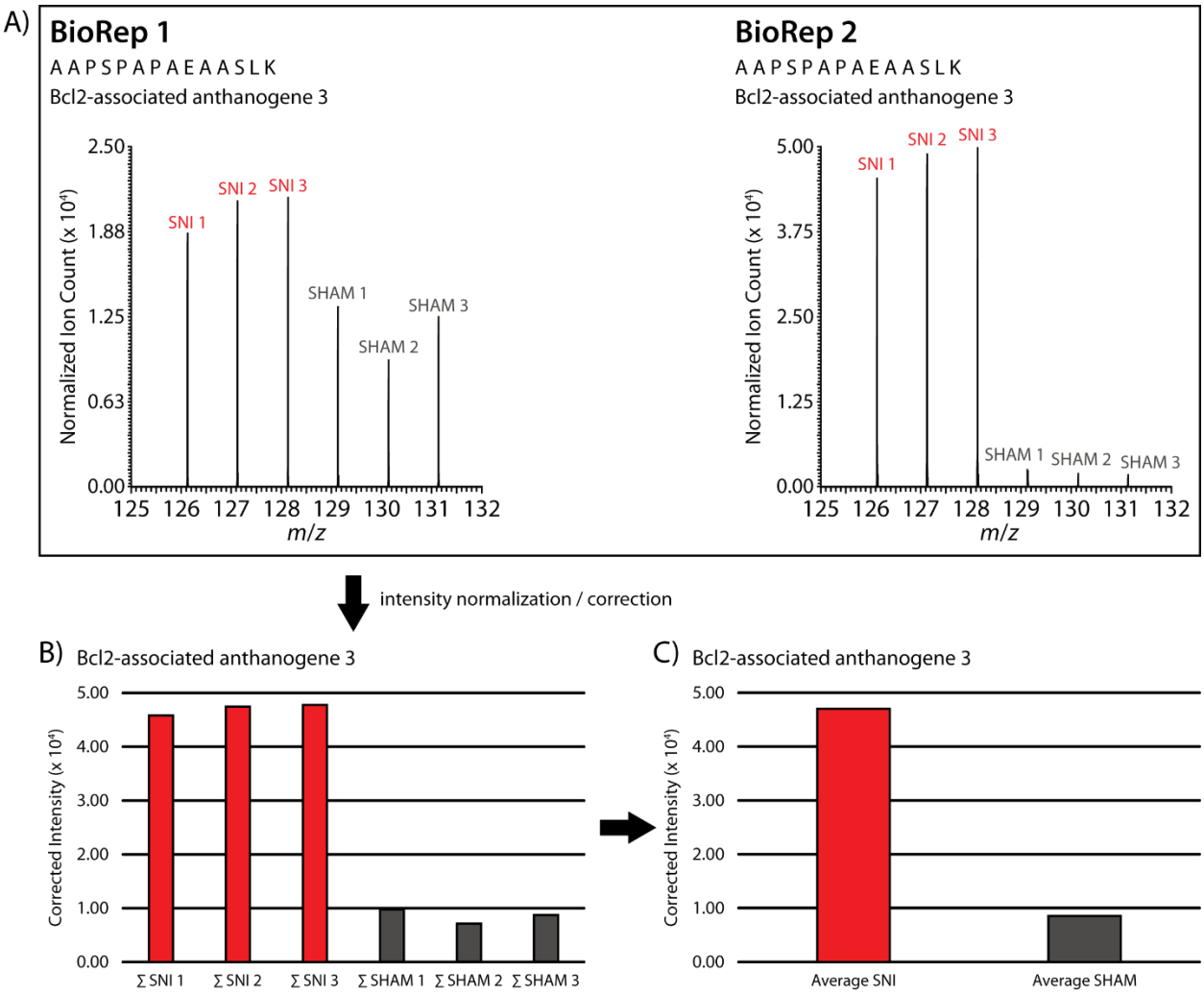


Figure S1. Example of protein quantification using TMT. In our experiments, all SNI samples were labeled with TMT 126, TMT 127, and TMT 128, while all SHAM samples were labeled with TMT 129, TMT 130, and TMT 131. A) The reporter ion spectra from a peptide found in both BioReps are shown. This peptide uniquely maps to a tryptic fragment of Bcl2-associated anthanogene 3. To perform protein-level quantification, B) the intensities of each sample channel (SNI1, SNI2, SNI3, SHAM1, SHAM2, and SHAM3) from each spectra belonging to a unique peptide of Bcl2-associated anthanogene 3 are corrected for TMT reagent purity and then summed. C) To get the final expression values, the summed intensities are averaged together.

Chapter 7

Quantitative phosphoproteomics of TGF-β and SDF-1 treatment in smooth muscle cells with elevated Smad3 expression: A preliminary study

Adapted from: Xiaofang Zhong,* Christopher B. Lietz,* Xudong Shi, Amanda R. Buchberger, Dustin C. Frost, K. Craig Kent, Lingjun Li. Quantitative phosphoproteomics of TGF-β and SDF-1 treatment in smooth muscle cells overexpressing Smad3. Manuscript in preparation.

* Indicates equal contribution

Abstract

Restenosis, a re-narrowing of arteries following medical intervention for coronary atherosclerosis, occurs in 15% - 20% of patients who receive angioplasty. Transforming growth factor-beta (**TGF-β**) signaling has a well-established link with restenosis, and recent transcriptomic investigations have provided evidence that TGF-β may induce pathological smooth muscle cell de-differentiation in the presence of elevated Smad3 levels. Here, we pursued a LC-MS based approach to quantify global phosphoproteomic changes regulated by TGF-β/Smad3 signaling pathway. To perform quantification, we used novel isobaric *N*,*N*-dimethyl leucine (**DiLeu**) tags. This investigation represents the first application of the high-resolution 12-plex DiLeu set, as well as the first post-translational modification analysis to utilize any DiLeu set.

Our data suggests TGF-β and Smad3 induces wide-spread protein dephosphorylation. Additionally, we observed that elevated levels of TGF-β and Smad3 up-regulated the phosphorylation of AHNAK's Ser2691. However, additional treatment with stromal-derived factor 1 (**SDF-1**) greatly down-regulated its phosphorylation. Since SDF-1 can potentially activate newly de-differentiated smooth muscle cells, future validation and follow-up on the participation of this phosphoproteoform may yield novel therapeutic targets to halt restenosis.

Introduction

Coronary atherosclerosis is characterized by plaque formation in the arterial intima that eventually results in a partial or full occlusion.[1] One of the most common medical interventions for this condition is percutaneous coronary angioplasty,[2] wherein a deflated balloon-like device on a catheter is fed through blood vessels to the site of the blockage. Once reached, the balloon is inflated to clear the intimal plaque and broaden the artery to increase blood flow. Although acutely effective, patients who receive angioplasty are at-risk to suffer from a re-narrowing of the arteries known as restenosis.[3] In restenosis, fracturing the atherosclerotic plaque can activate a cascade of molecular and cellular events that leads to neointimal hyperplasia, new arterial blockage in the form of dysfunctional smooth muscle cells (**SMCs**) and collagenous extracellular matrix (**ECM**).

Restenosis occurred in nearly 60% of treated patients in the initial years following angioplasty's implementation.[3] Although subsequent introduction of drug-eluting stents reduced rates to 15% - 20%,[4] there is still a great desire for new treatments with even lower rates.. Promising areas to discover therapeutic targets for restenosis are the signaling pathways of transforming growth factor-beta (**TGF-β**) and its related group of Smad-protein signal transducers.[5] During the last several decades, research has demonstrated a link between local up-regulation of TGF-β and Smad3 near arterial injury, subsequent intimal thickening,[6-10] and the ability to attenuate thickening through TGF-β/Smad3 inhibition.[11-13] This suggests that manipulation of TGF-β signaling after angioplasty may be a viable route to halting restenosis.

TGF-β signaling was one of the first systems to show that a ligand can bind to the same receptor in two different cell types and produce opposite results—a clear challenge

to the classical one-hormone-one-function paradigm.[14] For example, TGF-β stimulation can inhibit expression of the tumor-growing ID1 gene in mammary epithelial cells.[15] However, it actually promotes this gene in certain metastatic breast cancer cells.[16] The cellular environment, downstream of the initial ligand binding, is a contributing factor that determines the final outcome. Activated TGF-β receptors will phosphorylate and release the Smad complex, which is subsequently translocated to the nucleus to promote or inhibit gene expression. The particular genes that are affected depend on the population of Smad-interactors, activity of competing signal pathways, and the cell's epigenetic landscape.[14]

As a consequence of such complexity, directly targeting TGF-β may be hazardous to patients at-risk for restenosis. Direct inhibition has shown to actually accelerate atherosclerosis and produce plaques more prone to cause thrombosis.[17] Therefore, a better approach might be to target the downstream interactors that carry out TGF-β's positive regulation of neointimal hyperplasia. A recent study by Shi et al. utilized quantitative transcriptomics to evaluate possible targets.[18] They hypothesized that elevated levels of TGF-β and Smad3 drive SMC de-differentiation, and that proliferation of the resultant multipotent stem cells leads to restenosis. Overexpression of Smad3 was shown to be key to TGF-β's up-regulation of stem cell-related genes, most notably CXCR4, a receptor for the stem cell chemoattractant stromal-derived factor 1 (**SDF-1**).

To further investigate the possible mechanisms of TGF-β/Smad3-induced restenosis, we performed a quantitative mass spectrometry (**MS**)-based phosphoproteomic analysis. Phosphorylation of serine and threonine residues is a vital component of TGF-β signal cascade,[14] and MS-base strategies have proven to be an

effective approach to study large-scale phosphoprotein regulation in many biological systems.[19-21] We hypothesized that TGF-β and Smad3 initiate a vast spectrum of protein phosphorylation, and that SDF-1 stimulation of up-regulated CXCR4 helps drive the specific events that convert SMCs to the hyperplasia phenotype. The investigation detailed in this chapter aimed to gather preliminary data on changes to the phosphorylation landscape from elevated levels of TGF-β and Smad3. Subtle differences caused by additional SDF-1 treatment were also cataloged and used to suggest specific phosphoprotein targets for orthogonal validation. Future experiments on these suggested targets will be necessary to fully test our hypothesis.

For MS quantification, we employed the use of novel *N*,*N*-dimethyl leucine (**DiLeu**) isobaric tags, which were recently developed by our lab.[22-23] This investigation is the first application of DiLeu for post-translational modification (**PTM**) analysis, and the first biological application of the high-resolution DiLeu 12-plex. Additionally, we present The DiLeu Tool software suite to facilitate DiLeu quantification of protein expression and PTMs in large datasets.

Experimental

SMC culture, Smad3 overexpression, and TGF-β/SDF-1 treatments

Rat SMCs were isolated from the thoracoabdominal aorta of male Sprague-Dawley rats. SMCs at passages 4 to 5 were used for all experiments and maintained in DMEM supplemented with 10% fetal bovine solution (**FBS**) at 37 °C with 5% CO_2. Cell viability was >95% as assayed using the Trypan Blue exclusion method. Adenoviral vectors expressing Smad3 (**AdSmad3**) and control green fluorescent protein (**AdGFP**) were constructed as previously described. SMCs were infected for 4h with AdSmad3 (or

AdGFP) (3 ×10^4 particles/cell) in DMEM containing 2% FBS, and recovered for 20 h with 10% FBS, and then starved with 0.5% FBS for 24 h followed by treatment with human recombinant TGF-β (5 ng/mL, R&D Systems, Minneapolis, MN) or equivalent amount of solvent (final 4 μM HCl and 1μg/mL bovine serum albumin) for 24 h.

Some SMCs infected with AdSmad3 or AdGFP were treated with 100 ng/ml SDF-1 for 10 min at 37°C. The dishes were washed 3 times with cold PBS media, and then SMCs were collected and lyophilized. **Table 1** summarizes the treatments and overexpression patterns of all sample groups (S, G, SD, and GD). Each group contained three technical replicates.

Cell lysis and protein digestion

The digestion buffer was prepared to the following specifications: 8 M urea, 50 mM mM Tris•HCl, 5 mM $CaCl_2$, 30 mM NaCl, 1x protease inhibitor tablet (Roche, Penzberg, Germany), 1x phosphatase inhibitor tablet (Roche), pH 8. Cell pellets were reconstituted in digestion buffer and lysed with a probe sonicator (Thermo Fisher Scientific, Waltham, MA, USA) at 4º C. The lysates were then analyzed for total protein content via BCA assay (Thermo Fisher Scientific).

Protein digestion was performed as previously described.[24] First, equal amounts of total protein (250 μg) from each sample were reduced and alkylated by dithiothreitol and iodoacetamide, respectively. Next, samples were diluted with 50 mM Tris•HCl (pH 8) to lower the urea concentration to 0.9 M. Then, sequencing grade trypsin (Promega, Madison, WI) was added to each sample in a 50:1 *w*/*w* protein:enzyme ratio and incubated at 37 ºC for 18 hours. Finally, digestions were quenched by acidification with trifluoroacetic acid (**TFA**). Resultant peptides were purified via C18 Sep-Pak solid-phase

extraction cartridges (Waters Corporation, Milford, MA, USA) according to the manufacturer's protocol.

DiLeu peptide labeling

The complete set of 12-plex DiLeu tags was synthesized as previously described.[22] Each label was suspended to 40 µg µL^{-1} in *N*,*N*-dimethylformamide (**DMF**) and combined with 0.7x limiting molar ratios of 4(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium tetra-fluoroborate and *N*-methylmorpholine. These solutions were then vortexed for one hour to yield amine-reactive DiLeu tags.

Samples were dried down via speedvac and resuspended in 0.5 M triethylammonium bicarbonate. Peptides were labeled by adding activated DiLeu at a 5x *w*/*w* tag excess. DMF was added until the solvent mixture was 70% organic. The reaction solutions were then vortexed for 2 hours and quenched by adding hydroxylamine to a final concentration of 0.25%.

Each sample was labeled with a different DiLeu tag. Deuterated labels were strategically assigned to create a nearly even distribution of deuteriums between sample groups, minimizing any potential bias that may arise from ^{2}H-containing isotopologues. The sample/tag scheme was as follows: G1/115a, G2/117a, G3/118b, S1/115b, S2/116c, S3/118a, GD1/116a, GD2/117c, GD3118c, SD1/116b, SD2/117b, and SD3/118d

IMAC phosphopeptide enrichment

Labeled samples were combined in a 1:1:1:1:1:1:1:1:1:1:1:1 *v*/*v* ratio, purified with C18 Sep-Pak cartridges, and resuspended in 80% acetonitrile (**ACN**) 0.1% TFA. Our immobilized-metal affinity chromatography (**IMAC**) phosphopeptide enrichment protocol was adapted from methods described by Rose et al.[20] Magnetic Ni-NTA beads (Qiagen,

Hilden, Germany) were prepared in the following steps: wash with water (3x), 40 mM EDTA (pH 8) incubation with shaking for 30 minutes, wash with water (4x), 100 mM $FeCl_3$ incubation with shaking for 30 minutes, and a final wash with 80% ACN 0.1% TFA (3x). The sample was then added to the beads vortexed for 45 minutes. The initial supernatant and subsequent three washes with 80% ACN 0.1% TFA were saved as the "phosphopeptide-depleted" sample. After two additional 80% ACN 0.1% TFA washes, phosphopeptides were collected by two elutions with 0.7% NH4OH and then neutralized with 4% formic acid (**FA**).

High-pH reversed-phase high performance liquid chromatography

Prior to MS analysis, high-pH (**HpH**) reversed-phase (**RP**) high performance liquid chromatography (**HPLC**) was performed on a Waters Alliance (e2695 separation module, 2489 UV/vis detector monitored at 215 nm). Elutions from the phosphopeptide-enriched sample were combined and resuspended in HpH Solvent A (water, 10 mM NH_4HCO_2, pH 10) and separated on a 150 mm x 2.1 mm, 5 µm, 100 Å, C18 column (Phenomenex, Torrance, CA, USA). After 3 minutes of column-loading in 100% HpH Solvent A at 0.2 mL min^{-1}, HpH Solvent B (90% ACN, 10 mM NH_4HCO_2, pH 10) was linearly ramped from 0% to 35% over 50 minutes. Initial fractions were collected every 1.5 minutes and recombined into a total of 20 final fractions based on the UV chromatogram trace.

LC-MS acquisition

Online nano LC was performed on a nanoAcquity UPLC (Waters Corporation). Phosphopeptide fractions were dried down, resuspended in 15 µL of 0.1% FA, and injected onto a self-fabricated capillary column (16 cm length, 75 µm i.d.) packed with reversed-phase BEH C18 material (1.7 µm, 130 Å, Waters Corporation). Samples were

loaded onto the column in 100% Solvent A (water, 5% dimethyl sulfoxide, 0.2% FA, pH 2-3) at a flow rate of 0.3 uL min^{-1}. Separation occurred during the following non-linear gradient of Solvent B (ACN, 5% dimethyl sulfoxide, 0.2% FA, pH 2-3): 0-36 min from 0% to 12%, 36-68 min from 12% to 22%, 68-80min from 22% to 30%. Although the solid-phase chemistry of the HPLC and nano LC columns were very similar, reversed-phase peptide separation at pH 10 is highly orthogonal to separation at pH 2-3 due to protonation-induced changes in sidechain hydrophobicity.[25]

Eluting peptides were electrosprayed into an Orbitrap Elite mass spectrometer (Thermo Fisher Scientific). MS and tandem MS (**MS/MS**) spectra were collected in a Top 15 data-dependent acquisition. For precursor MS scans, the AGC, maximum injection time, and resolution (at *m*/*z* 400) was set to $1x10^6$, 50 ms, and 30000, respectively. MS/MS fragmentation was performed via HCD with 27% normalized collision energy and first mass set to *m*/*z* 110. The MS/MS AGC, maximum injection time, and resolution (at *m*/*z* 400) was set to $1x10^5$, 100 ms, and 60000, respectively. Dynamic exclusion time and minimum precursor intensity was set to 30 seconds and 1000, respectively. Raw data files will be freely available on Chorus (http://chorusproject.org) upon publication of this manuscript.

Peptide and protein identification

All peptide and protein identification was performed using the COMPASS (v1.4) software suite[26] and the Open Mass Spectrometry Search Algorithm (**OMSSA**).[27] HPLC fractions were injected twice for a total of 40 raw files for the phosphopeptide-enriched samples. Each .raw file was converted to .txt (for OMSSA database searching) and .mgf (for later use in quantification software). The .txt files were searched against tryptic

peptides from a target-decoy protein database for *Rattus norvegicus* (UP000002494, UniProt, canonical and isoforms, downloaded on 12/03/2015), allowing a maximum of two miscleavages. The precursor and fragment mass accuracy thresholds were set to 25 ppm and 0.02 Th, respectively. Multiply charged fragment ions were considered in all MS/MS spectra. Carbamidomethyl cysteine, DiLeu N-terminus, and DiLeu lysine were selected as fixed modifications. Oxidation of methionine, DiLeu tyrosine, phosphorylation of serine with neutral loss, phosphorylation of threonine with neutral loss, and phosphorylation of tyrosine were set as variable modifications.

The raw OMSSA outputs for all 40 files were then curated in batch-mode to a peptide-level 1% false discovery rate (**FDR**) using COMPASS's FDR Optimizer. All target and decoy peptide-spectrum matches (**PSMs**) belonging to peptides in the curated list were then grouped into parsimonious protein groups by Protein Hoarder. The protein-level FDR and minimum peptides-per-group were set to 1% and 1, respectively. Peptides shared between protein groups were not used for protein group scoring.

Phosphorylated amino acid localization

The PSMs mapping to protein groups remaining after Protein Hoarders 1% FDR curation were subjected to a PTM localization analysis by the PhosphoRS algorithm.[28] An independent localization analysis helps determine which phosphosite assignments are of high-confidence and unambiguous. Fragment mass tolerance, localization probability cutoff, neutral loss consideration, maximum position isoforms, and maximum PTMs per peptide were set to 0.02 Daltons, 75%, automatic, 200, and 10, respectively.

Phosphoproteoform quantification

To facilitate accurate phosphoproteoform quantification via DiLeu, we created a custom software suite called DiLeu Tool. A detailed description of the software can be found in **Supplementary Information 1**, and the C# source code can be found in **Appendix III**. Briefly, The DiLeu Tool uses .mgf files, Protein Hoarder output, and PhosphoRS output to quantify expression levels of localized and unlocalized phosphoproteoforms. After PSMs are assigned to proteoforms by PhosphoRS, each PSM's DiLeu tag intensities are retrieved from the .mgf file. After intensities are purity-corrected and normalized for channel-bias, a mean intensity for each tag is calculated at the proteoform-level.

For this investigation, mean tag intensities were assigned to their corresponding sample groups to calculate sample group-level proteoform intensities (for example, the mean of 115a's, 117a's, and 118b's mean intensities for proteoform X was calculated and assigned as sample G's intensity for proteoform X). Mean-normalizations and heteroscedastic t-tests were performed for all sample group-level binary comparisons.

Approach summary and rationale

The overall MS-based quantitative proteomic workflow is illustrated in **Scheme 1**. In order to calculate p-values from Student T-tests, we divided 12 DiLeu tags among four sample groups (**Table 1**) that each contained three technical replicates. Groups S and SD were the primary experimental groups, while groups G and GD served as their respective controls. Group S aimed to simulate the precursory environment for restenosis: high concentrations of TGF-β and elevated Smad3 expression. Group SD produced these same conditions, but the additional treatment allowed us to identify TGF-B/Smad3-induced phosphorylation that was enhanced or inhibited by SDF-1. Based on

the results of Shi et al.,[18] GD was expected to have very low expression levels of CXCR4, thus enabling this group to serve as control for any activity that may have resulted from SDF-1 stimulation of other receptors. Group G, whose only treatment was overexpression of an inert protein, served as an absolute control.

Following DiLeu labeling and phosphopeptide enrichment of the tryptic digests, the workflow splits into parallel experiments. This dissertation chapter will only focus on the preliminary results from analysis of the "phosphopeptide-enriched" portion. Analysis of the "phosphopeptide-depleted" portion is currently underway and expected to yield quantification of relative protein abundance. Abundance data will offer a complementary view of large-scale dynamics throughout the four sample groups, and it may help to discern whether regulation of certain phosphosites are a product of changes in phosphorylation or changes in protein translation.

Results and Discussion

Protein identification metrics

Following the curation of our OMSSA search results to a 1% protein-level FDR, we were left with a total of 14324 PSMs, 4345 unique peptides, and 1542 distinguishable protein groups. Phosphoserine, phosphothreonine, or phosphotyrosine modifications were identified on 11736 PSMs, corresponding to a PSM-level phosphopeptide enrichment efficiency of 81.9%. These phosphopeptide PSMs were mapped to 2812 distinct phosphoproteoforms, 2311 of which were localized at the residue-level by PhosphoRS with a probability of 75% or greater. The relative frequencies of proteoforms with localized phosphoserine, phosphothreonine, and phosphotyrosine were 87.9%, 11.5%, and 1.0%, respectively.

Identification metrics of 11736 PSMs and 2812 phosphoproteoforms may strike some readers as low for the analysis of 20 HPLC fractions. This yield can be attributed to the decrease in identifications generally observed after tagging digest peptides.[29-31] Additionally, 12-plex DiLeu requires a MS/MS resolving power of 60000, leading to slower MS/MS scan speeds. Despite the challenges, we weighed the use of a 12-plex as important and necessary due to its substantial improvement of quantitation accuracy and greatly enhanced throughput. Simultaneous comparison of four sample conditions with three replicates was only possible through the use of the DiLeu.

TGF-β/Smad3 levels dominate large-scale phosphoprotein regulation

The 4x3 setup of our experiment enabled us to look beyond single binary comparisons and ensured every identified proteoform was quantified in all sample groups. **Figure 1** contains all phosphoproteoform expression ratios from different group comparisons on a Log_2-scale. Each dot represents a different phosphoproteoform. By plotting one binary comparison against another, we were able to identify inter-group similarities in global phosphorylation patterns. **Figure 1A** shows a positive linear correlation between SD/G and SD/GD group ratios, indicating that large-scale phosphorylation differences between SD and G were very similar to the differences between SD and GD. As expected, SDF-1 treatment did not greatly affect the phosphorylation landscape without elevated levels of TGF-β/Smad3, possibly due to the lack of CXCR4 up-regulation.[18] The poor correlation between SD/G and GD/G ratios in **Figure 1B** further supports this interpretation. Phosphoprotein regulation after combined Smad3 overexpression, TGF-β treatment, and SDF-1 treatment bears little resemblance to regulation after SDF-1 treatment alone.

Figure 1C contains the plot of SD/G and S/G phosphoproteoform ratios. Like SD/G and SD/GD in **Figure 1A**, SD/G and S/G display a positive linear correlation, though with a lower coefficient of determination. This result implies that, in the presence or absence of SDF-1 treatment, TGF-β/Smad3-induced changes to the baseline SMC phosphorylation landscape are very similar. **Figure 1D's** poor correlation between S/G and GD/G, like **Figure 1B**, further asserts the inertness of SDF-1 without TGF-β and Smad3.

Taken all together, **Figure 1** suggests that TGF-β and Smad3 are a much greater influence on large-scale phosphorylation than SDF-1. Similar regulation of phosphorylation in S and SD groups supports our hypothesis. However, expression of some phosphoproteoforms clearly change with the inclusion of SDF-1. The overall patterns are the same, yet the number of points found in quadrants 2 and 4 of **Figure 1C** are not negligible. For the remainder of this chapter, the differences between phosphoprotein regulation in groups S and SD will be primarily examined through S/GD and SD/GD comparisons.

Phosphoproteoform quantification metrics

Histograms for all phosphoproteoform expression ratios within S/GD and SD/GD comparisons are shown in **Figure 2A**. Black dotted lines mark the Log_2-scale thresholds of 0.60 and -0.60 for significant up-regulation and down-regulation, respectively. Although both plots appear to be symmetrical and Gaussian-like, S/GD displays a slightly wider distribution between than SD/GD for ratios between the significance thresholds.

In addition to having an expression ratio outside the thresholds, significantly up- or down-regulated phosphoproteoforms had to have a p-value less than 0.05. Of the 2311

localized proteoforms, 211 of them met the significance criteria for at least one of the two comparison groups (S/GD or SD/GD). S/GD contained 157 differentially regulated localized proteoforms, whereas SD/GD contained slightly fewer at 141. S/GD and SD/GD comparisons also contained 31 and 30 unlocalized significant differentially regulated proteoforms, respectively.

The volcano plots in **Figure 2B** plot all phosphoproteoforms in terms of p-value and Log_2-expression ratio. The shape and distribution of both plots confirm that our systems are well-behaved and that the comparisons should be amenable to quantification with our defined significance criteria. We do note that S/GD's volcano plot has a slight negative bias compared to SD/GD. We believe the source of this bias comes from the three most populated bins just below Log_2-expression ratio of 0.

Numerous down-regulated phosphorylations from elevated TGF-β and Smad3

Due to the ambiguity in the MS/MS spectra of unlocalized phosphoproteoforms, the rest of this chapter will only focus on quantitative analysis of localized proteoforms. All phosphoproteoforms found to have significant differential regulation in S/GD or SD/GD are listed in **Table S1**. Additionally, each entry contains the phosphoproteoform's SD/S ratio on a Log_2-scale. The SD/S ratio was highlighted in yellow if it met the significance criteria ($Log_2(SD/S) \leq -0.60$ or $Log_2(SD/S) \geq 0.60$, p-value < 0.05). **Figure 3A** is a visual summary of all the phosphoproteoforms with significant differential regulation. Each pair of red and blue bars corresponds to a different entry in **Table S1**, and their order of appearance in the figure is the same as their order of appearance in the table. **Figure 3B** shows a table that lists all proteoforms in **Table S1** with SD/S ratios meeting significance criteria.

From **Figure 3A**, it is clear that both S/GD and SD/GD contain more significant down-regulated phosphoproteoforms than up-regulated phosphoproteoforms. One possibility for this trend could be increased phosphatase activity induced by elevated TGF-β/Smad3. Protein phosphatases are vital to intensity and duration of TGF-β signaling phosphorylation cascades,[32] such as the signal-inhibiting dephosporylations of Smad3's C-terminal SXS-motif by PPMA1[33] or TGF-β's receptor, TβRI, by PP1c.[34] Additionally, down-regulated phosphorylation could affect de-differentiated SMCs, as phosphatases in other stem cell types can assert negative[35] or positive[36] regulation over differentiation. Further experiments could be designed to see if selective phosphatase stimulation/inhibition gradients in SMCs enhance or halt the formation of phenotypes induced by TGF-β and Smad3.

Dephosphorylated Ser418 on Smad3 suggests active TGF-β/Smad3 signaling

We hypothesized that TGF-β/Smad3 signaling should be active in both SD and S sample groups, and that additional SDF-1 treatment might shift the system towards specific events that lead to neointimal hyperplasia. Therefore, we should expect evidence of active or uninhibited TGF-β/Smad3 signaling in SD/GD and S/GD. We found phosphorylation of Ser418 on Smad3, a site conserved in human Smad3, was down-regulated in both comparisons (SD/GD: -0.60, p-value = 0.0172; S/GD: -0.79, p-value = 0.0158; **TableS1** entry 116). Guo et al. showed that Ser418 of TGF-β-activated Smad3 is phosphorylated by CKIγ2, and that phospho-Ser418 induced Smad3's ubiquitin-directed degradation.[37] Consequently, TGF-β/Smad3 signal pathways were inhibited. Lower levels of phospho-Ser418 in SD and S would then indicate less proteasomal degradation of activated Smad3 compared to control. The negligible difference in fold-

change between the two groups (Log_2(SD/S) = 0.19, p-value > 0.05) suggests additional SDF-1 treatment does not affect this process.

We did not identify any PSMs corresponding to phosphorylated Ser422, Ser423, or Ser425 on Smad3. These sites are phosphorylated by TGFβRI and required for transcriptional regulation via Smad3.[38] Future experiments that target quantification of these phosphoproteoforms may provide more direct information of Smad3's signaling state. The only other unambiguous Smad3 phosphorylation we observed was on Ser416, but its regulation was not significant in SD/GD or S/GD. The function of phospho-Ser416 is unknown, though it has been previously observed in large-scale phosphoproteomic experiments.[39]

SDF-1 treatment modulates phosphorylation of Ser2691 in AHNAK

In general, the phosphoproteoforms in **Figure 3** show most ratios in SG/GD and S/GD have the same positive or negative directionality. The greatest exception to this is the phosphorylation of Ser2691 on the protein AHNAK (**Table S1**, entry 167), the last entry listed in **Figure 3B**. In S samples, elevated levels of TGF-β and Smad3 led to a very significant up-regulation of phospho-Ser2691 (Log_2(S/GD) = 1.58, p-value = 0.00009), the second largest up-regulation of the S/GD comparisons. However, additional SDF-1 treatment in SD samples drove phosphorylation levels to down-regulation relative to GD (Log_2(SD/GD) = -0.82, p-value = 0.0062). This equates to a -2.40 Log2-fold change between SD and S. AHNAK is a large 700 kDa protein that is extensively phosphorylated on serine and threonine.[40] Although we identified a total of 24 phosphoproteoforms for AHNAK, the only other one that met significance criteria was Ser5390 (Log_2(SD/SG) = 0.67, p-value = 0.0136; Log_2(S/GD) = 0.70, p-value = 0.0070).

AHNAK has been reported to have important and seemingly contradictory involvement in both restenosis and TGF-β/Smad3 signaling. Recently, Lee et al. showed that AHNAK can facilitate TGF-β/Smad3 suppression of tumor growth in breast cancer cells.[41] The giant phosphoprotein directly binds to Smad3 and co-localizes in the nucleus, eventually leading to cell cycle arrest and inhibited proliferation. While it may seem that such a function could be beneficial for mitigating restenosis, Lim et al. has demonstrated that AHNAK is a *positive* regulator of aortic SMC migration and neointimal formation in carotid arteries.[42] This activity was independent of TGF-β and had resulted from AHNAK coupling of platelet-derived growth factor (**PDGF**)-stimulation and the Erk signaling pathway.

SDF-1 is known to recruit stem cells and promote restenosis.[43] According to our data, TGF-β/Smad3 activation greatly up-regulates phosphorylation of Ser2691 in ANHAK, but the addition of SDF-1 leads to the opposite outcome. The complex multiple functions of AHNAK and the magnitude of its differential regulation by SDF-1 make it a very interesting target for investigation. Future experiments should focus on discerning the function of this residue and whether its dephosphorylation affects its ability to negatively regulate growth and proliferation.

Conclusion

Our initial hypothesis can be divided into two parts. First, we proposed that large-scale phosphorylation patterns would be largely similar between SD and S sample groups. This assertion was supported by our data (**Figure 1**). Secondly, we proposed that the few differences in phosphorylation from additional SDF-1 treatment may shift

TGF-β/Smad3 signaling towards events that promote neointimal hyperplasia. Through the use of 12-plex DiLeu, we have identified targets that warrant further.

The observed widespread dephosphorylation raises questions about phosphatase activation by TGF-β/Smad3. The greater number of negative changes exceeding 2-fold (less than -1 on a Log_2-scale) in SD samples imply phosphatases could be further enhanced by SDF-1. An experiment could be designed where samples similar to the S group could be co-transfected to overexpress phosphatases that could be involved with our down-regulated targets. If enhancement of certain phosphatases resulted in phosphorylation profiles and other phenotypic markers characteristic of SD samples, this would indicate that phosphatase activity may indeed be a major consequence of SDF-1 treatment and the enhancements to neointimal hyperplasia.

The AHNAK down-regulation results suggest a possible role for phospho-Ser2691 in determining the protein's function in a system with active TGF-β/Smad3 signaling. One property already known to determine function is cellular localization: growth suppression happens when AHNAK is in the nucleus,[41] and migration/proliferation enhancements happen when AHNAK is localized at the cell membrane.[42] It is possible that Ser2691 is dephosphorylated after SDF-1 treatment to exclude it from the nucleus, similar to the way phosphorylation of Ser5535 in human AHNAK regulates localization between the nucleus and cytoplasm.[44] As a simple test, a fluorescent AHNAK-construct could be imaged in cells with S- and SD-like treatment. If the localization is significantly different, follow-up experiments utilizing mutant AHNAK with an alternative amino acid at residue 2691. If AHNAK localization in S- and SD-like cells becomes the same as wild-type S-like cells, AHNAK and its interactors could be promising candidates for future therapeutic targets.

References

(1) Nabel, E. G., and Braunwald, E. *New Eng. J. Med.* **2012** *366*, 54-63.

(2) (2014) What is percutaneous coronary intervention?, (NIH, Ed.), http://www.nhlbi.nih.gov/health/health-topics/topics/angioplasty/.

(3) Rajagopal, V., and Rockson, S. G. *Am. J. Med.* **2003** *115*, 547-553.

(4) Inoue, T., Croce, K., Morooka, T., Sakuma, M., Node, K., and Simon, D. I. *JACC: Cardiovasc. Interventions* **2011** *4*, 1057-1066.

(5) Suwanabol, P. A., Kent, K. C., and Liu, B. *J. Surg. Res.* **2011** *167*, 287-297.

(6) Madri, J. A., Reidy, M. A., Kocher, O., and Bell, L. *Lab. Invest.* **1989** *60*, 755-765.

(7) Chamberlain, J., Gunn, J., Francis, S. E., Holt, C. M., Arnold, N. D., Cumberland, D. C. *et al. Cardiovasc. Res.* **2001** *50*, 125-136.

(8) Sluijter, J. P., Verloop, R. E., Pulskens, W. P., Velema, E., Grimbergen, J. M., Quax, P. H. *et al. Cardiovasc. Res.* **2005** *68*, 136-143.

(9) Tsai, S., Hollenbeck, S. T., Ryer, E. J., Edlin, R., Yamanouchi, D., Kundi, R. *et al. Am. J. Physiol. Heart C* **2009** *297*, H540-H549.

(10) Kanzaki, T., Tamura, K., Takahashi, K., Saito, Y., Akikusa, B., Oohashi, H. *et al. Arterioscler., Thromb., Vasc. Biol.* **1995** *15*, 1951-1957.

(11) Yamamoto, K., Morishita, R., Tomita, N., Shimozato, T., Nakagami, H., Kikuchi, A. *et al. Circulation* **2000** *102*, 1308-1314.

(12) Smith, J. D., Bryant, S. R., Couper, L. L., Vary, C. P. H., Gotwals, P. J., Koteliansky, V. E. *et al. Circ. Res.* **1999** *84*, 1212-1222.

(13) Edlin, R. S., Tsai, S. L., Yamanouchi, D., Wang, C. J., Liu, B., and Kent, K. C. *J. Vasc. Surg.* **2009** *49*, 1289-1295.

(14) Massague, J. *Nat. Rev. Mol. Cell. Biol.* **2012** *13*, 616-630.

(15) Kang, Y., Chen, C. R., and Massague, J. *Mol. Cell* **2003** *11*, 915-926.

(16) Padua, D., Zhang, X. H. F., Wang, Q. Q., Nadal, C., Gerald, W. L., Gomis, R. R. *et al. Cell* **2008** *133*, 66-77.

(17) Mallat, Z., Gojova, A., Marchiol-Fournigault, C., Esposito, B., Kamate, C., Merval, R. *et al. Circ. Res.* **2001** *89*, 930-934.

(18) Shi, X., DiRenzo, D., Guo, L. W., Franco, S. R., Wang, B., Seedial, S. *et al. PLoS One* **2014** *9*, e93995.

(19) Phanstiel, D. H., Brumbaugh, J., Wenger, C. D., Tian, S., Probasco, M. D., Bailey, D. J. *et al. Nat. Meth.* **2011** *8*, 821-827.

(20) Rose, C. M., Venkateshwaran, M., Volkening, J. D., Grimsrud, P. A., Maeda, J., Bailey, D. J. *et al. Mol. Cell. Proteomics* **2012** *11*, 724-744.

(21) Nita-Lazar, A., Saito-Benz, H., and White, F. M. *PROTEOMICS* **2008** *8*, 4433-4443.

(22) Frost, D. C., Greer, T., and Li, L. *Anal. Chem.* **2015** *87*, 1646-1654.

(23) Xiang, F., Ye, H., Chen, R., Fu, Q., and Li, L. *Anal. Chem.* **2010** *82*, 2817-2825.

(24) Greer, T., Hao, L., Nechyporenko, A., Lee, S., Vezina, C. M., Ricke, W. A. *et al. PLoS One* **2015** *10*, e0135415.

(25) Batth, T. S., Francavilla, C., and Olsen, J. V. *J. Proteome. Res.* **2014** *13*, 6176-6186.

(26) Wenger, C. D., Phanstiel, D. H., Lee, M. V., Bailey, D. J., and Coon, J. J. *PROTEOMICS* **2011** *11*, 1064-1074.

(27) Geer, L. Y., Markey, S. P., Kowalak, J. A., Wagner, L., Xu, M., Maynard, D. M. *et al. J. Proteome Res.* **2004** *3*, 958-964.

(28) Taus, T., Köcher, T., Pichler, P., Paschke, C., Schmidt, A., Henrich, C. *et al. J. Proteome Res.* **2011** *10*, 5354-5362.

(29) Sandberg, A., Branca, R. M. M., Lehtiö, J., and Forshed, J. *J. Proteomics* **2014** *96*, 133-144.

(30) Pichler, P., Köcher, T., Holzmann, J., Mazanek, M., Taus, T., Ammerer, G. *et al. Anal. Chem.* **2010** *82*, 6549-6558.

(31) Hebert, A. S., Merrill, A. E., Stefely, J. A., Bailey, D. J., Wenger, C. D., Westphall, M. S. *et al. Mol. Cell. Proteomics* **2013** *12*, 3360-3369.

(32) Liu, T., and Feng, X.-H. *Biochem. J.* **2010** *430*, 191-198.

(33) Sundqvist, A., Ten Dijke, P., and van Dam, H. *Breast Cancer Res.* **2012** *14*, 204.

(34) Shi, W. B., Sun, C. X., He, B., Xiong, W. C., Shi, X. M., Yao, D. C. *et al. J. Cell Biol.* **2004** *164*, 291-300.

(35) Chen, Q., Zhou, Y., Zhao, X., and Zhang, M. *J. Cell. Biochem.* **2011** *112*, 3185-3193.

(36) Yu, W.-M., Liu, X., Shen, J., Jovanovic, O., Pohl, Elena E., Gerson, Stanton L. *et al. Cell Stem Cell* **2013** *12*, 62-74.

(37) Guo, X., Waddell, D. S., Wang, W., Wang, Z., Liberati, N. T., Yong, S. *et al. Oncogene* **2008** *27*, 7235-7247.

(38) Liu, X., Sun, Y., Constantinescu, S. N., Karam, E., Weinberg, R. A., and Lodish, H. F. *Proc. Natl. Acad. Sci. U.S.A.* **1997** *94*, 10669-10674.

(39) Dephoure, N., Zhou, C., Villén, J., Beausoleil, S. A., Bakalarski, C. E., Elledge, S. J. *et al. Proc. Nat. Acad. Sci. U.S.A.* **2008** *105*, 10762-10767.

(40) Shtivelman, E., and Bishop, J. M. *J. Cell Biol.* **1993** *120*, 625-630.

(41) Lee, I. H., Sohn, M., Lim, H. J., Yoon, S., Oh, H., Shin, S. *et al. Oncogene* **2014** *33*, 4675-4684.

(42) Lim, H. J., Kang, D. H., Lim, J. M., Kang, D. M., Seong, J. K., Kang, S. W. *et al. Cardiovasc. Res.* **2013** *97*, 302-310.

(43) Schober, A., Knarren, S., Lietz, M., Lin, E. A., and Weber, C. *Circulation* **2003** *108*, 2491-2497.

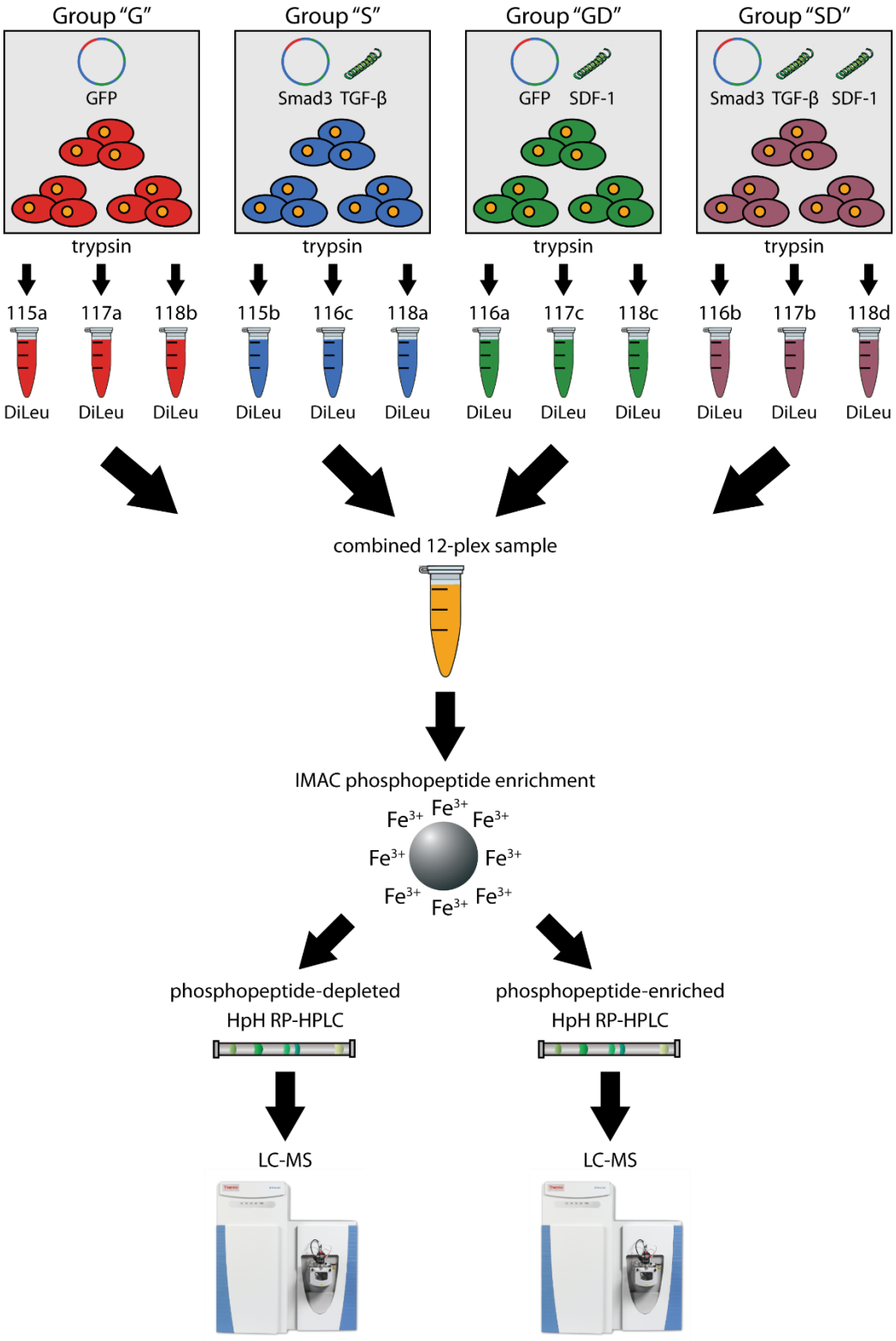
(44) Sussman, J., Stokoe, D., Ossina, N., and Shtivelman, E. *J. Cell Biol.* **2001** *154*, 1019-1030.

(45) Shadforth, I. P., Dunkley, T. P. J., Lilley, K. S., and Bessant, C. *BMC Genomics* **2005** *6*, 145-145.

(46) Greer, T., Lietz, C. B., Xiang, F., and Li, L. J. *J. Am. Soc. Mass Spectrom.* **2015** *26*, 107-119.

(47) Sturm, R. M., Lietz, C. B., and Li, L. J. *Rapid Commun. Mass Spectrom.* **2014** *28*, 1051-1060.

Scheme 1. Quantitative phosphoproteomic workflow for protein digestion, DiLeu labeling, phosphopeptide enrichment, and LC-MS.



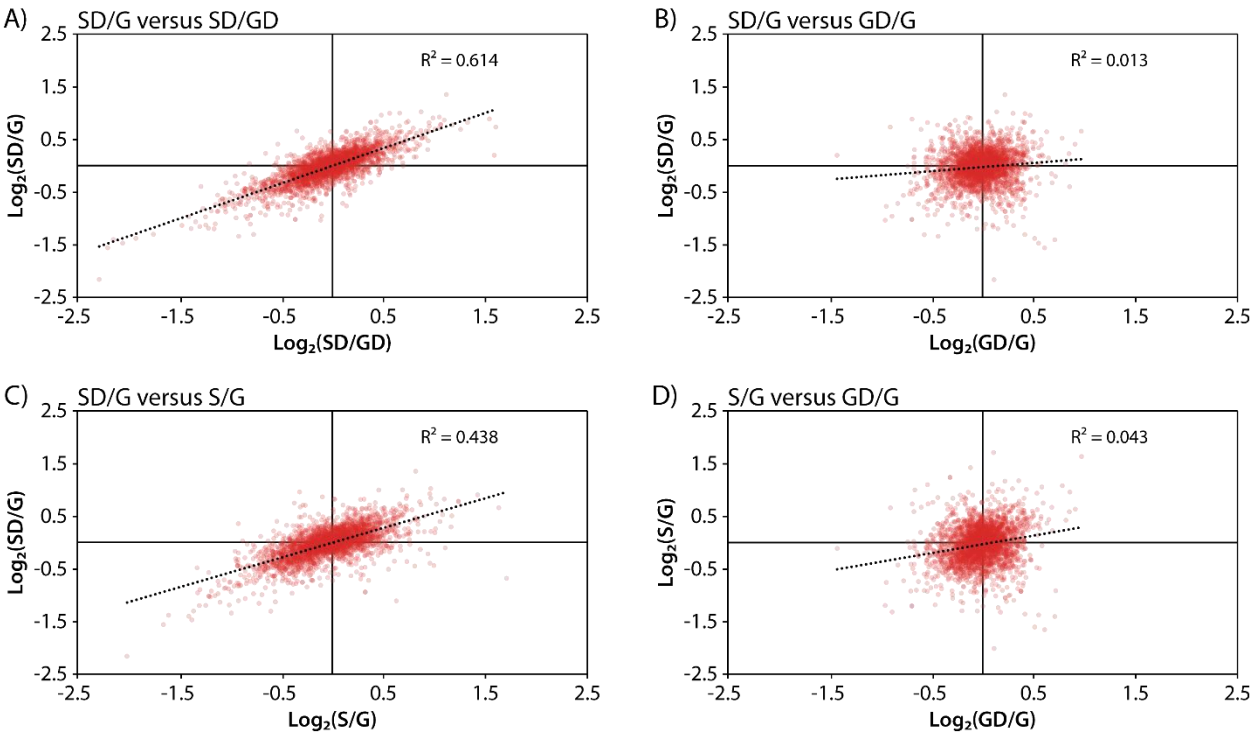


Figure 1. Correlation of large-scale protein phosphorylation. The scatter plots above compare the mean-normalized Log_2-expression ratios of all identified phosphoproteoforms between **A)** SD/G and SD/GD, **B)** SD/G and GD/G, **C)** SD/G and S/G, **D)** S/G and GD/G. Coefficients of determination (R^2) correspond to best-fit linear regression lines of each plot. Each red dot represents a different phosphoproteoform.

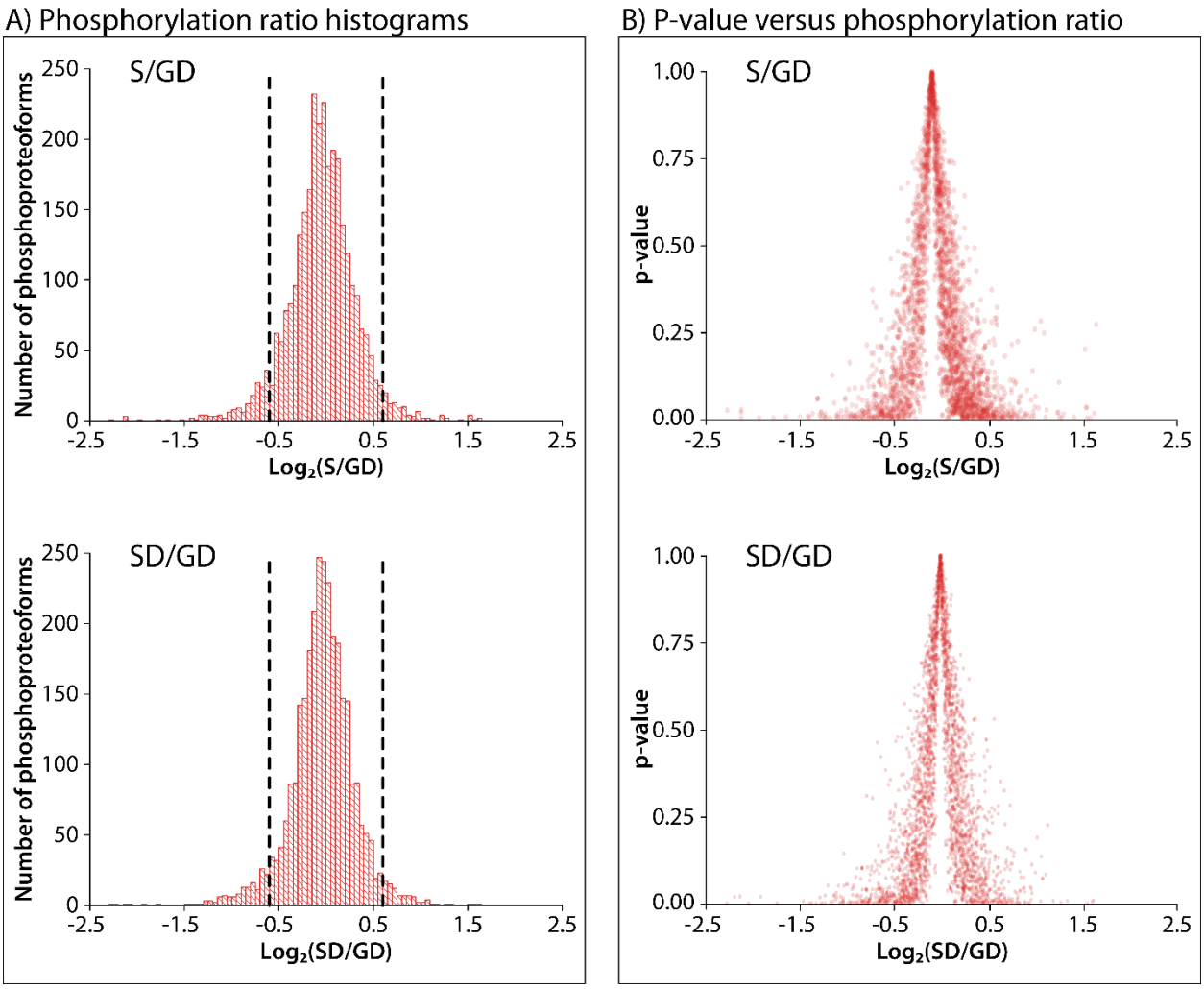
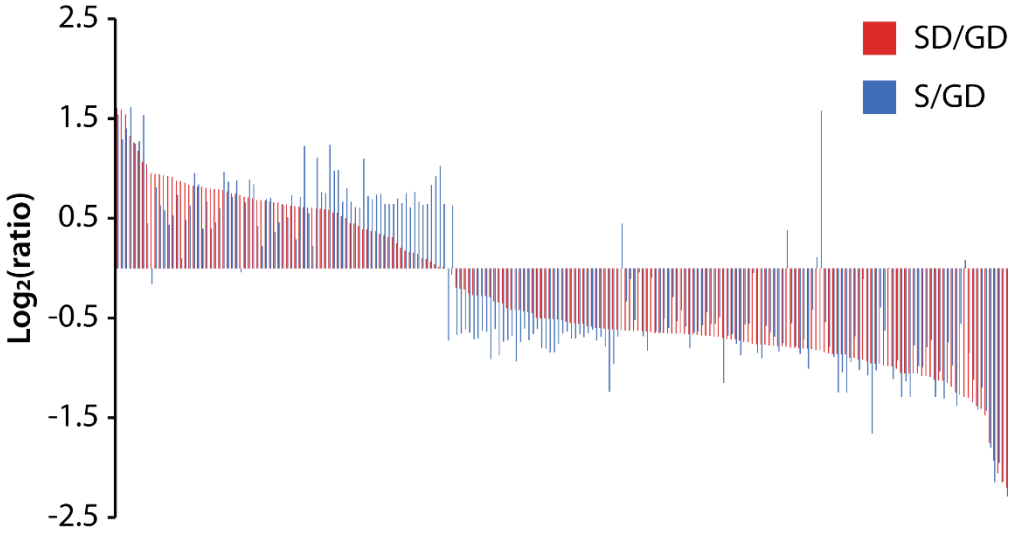


Figure 2. Phosphoproteoform quantification metrics. A) Histograms show all mean-normalized Log2-expression ratios for the primary group comparisons, S/GD and SD/GD. The black dotted-lines mark the Log_2-ratio significance thresholds of ± 0.60. **B)** Volcano plots show the relationship between p-value and Log_2-expression ratios in S/GD and SD/GD comparisons. Each red dot represents a different phosphoproteoform.

A) Significant difference (S/GD or SD/GD)



B) Significant difference (S/GD or SD/GD; SD/S)

Gene	Proteoform	Log_2(SD/S)	Gene	Proteoform	Log_2(SD/S)
Pdxdc1	T(687);T(691)	1.11	42256	T(107)	-0.71
Farp1	S(373)	0.77	Map3k4	S(207)	-0.76
Tns1	S(812)	0.63	Zfp36	S(185)	-0.81
Smtn	S(513)	0.63	Ric1	S(1036)	-0.88
Dclk1	S(307)	-0.61	Rbm14	S(256)	-0.98
Bag3	S(176)	-0.61	Eml3	S(156)	-1.01
Tsc2	S(1132)	-0.63	Plekha2	S(184)	-1.07
Akt1s1	S(184)	-0.65	Nr3c1	S(155)	-1.17
Trim24	S(210)	-0.69	Filip1l	S(1039)	-1.37
Dpysl3	S(635)	-0.71	Ahnak	S(2691)	-2.40

Figure 3. Modulation of protein phosphorylation during TGF-β/Smad3 signaling by SDF-1. A) Differences in regulation of protein phosphorylation between SD/GD (red) and S/GD (blue). Each pair of bars represents a different phosphoproteoforms that met significance criteria for at least one of the comparisons (Log_2(ratio) > 0.60 or Log_2(ratio) < -0.60, p-value < 0.05). The order of appearance (from left-to-right) corresponds to the numerical order of phosphoproteoforms in **Table S1**. **B)** A table listing all significant SD/GD and S/GD phosphoproteoforms that also met significance criteria for SD/S comparison (Log_2(SD/S) > 0.60 or Log_2(SD/S) < -0.60, p-value < 0.05).

Table 1. Summary of overexpression (AdSmad3 or AdGFP) and ligand treatments (TGF-β and SDF-1) for sample groups S, G, SD, and GD.

Sample Group	AdGFP	AdSmad3	TGF-β	SDF-1
S	-	+	+	-
G	+	-	-	-
SD	-	+	+	+
GD	+	-	-	+

Supplemental Information

Supplemental Information 1. The DiLeu Tool.

The DiLeu Tool is a software suite programmed with Microsoft Visual C# 2010. It was developed to facilitate DiLeu quantification of protein and phosphoproteoform expression, and it is amenable to any combination of tags from the 12-plex set.[22] Currently, The DiLeu Tool is intended for use after database searching and phosphoproteoform localization by COMPASS (v1.4)[26] and PhosphoRS,[28] respectively. Complete source code for The DiLeu Tool is available in **Appendix III**.

Considerations for pre- and post-acquisition of LC-MS data

Use of The DiLeu Tool requires a Top-N data-dependent acquisition scheme. If reporter purity-corrections are desired, MS/MS acquisitions much achieve spectral resolution of 60000 at *m*/*z* 400. This corresponds to resolving power settings of 60000 on an Orbitrap Elite or 70000 on an Orbitrap Q-Exactive. Following acquisition, raw files must be converted to Mascot generic format (**.mgf**).

Preparing input files for The DiLeu Tool

Protein grouping by the COMPASS Protein Hoarder results in two .csv files: “Parsimony_peptides_filtered.csv” (hereafter referred to as Parsimony Peptides) and “Parsimony_proteins_filtered.csv” (hereafter referred to as Parsimony Proteins). Parsimony Proteins contains a list of all protein groups created by Protein Hoarder and Parsimony Peptides contains a list of all PSMs for peptides assigned to those protein groups. Both files are required for protein quantification. For quantification of phosphoproteoforms, the Parsimony Peptides file needs to be appended with additional information denoting which proteoform the PSM corresponds to and if the proteoform met

localization confidence criteria. This file (hereafter referred to as Parsimony Phosphopeptides) is created after using the PhosphoRS plug-in for COMPASS.

As a final note on input files, it is absolutely vital that commas **ONLY** be used to separate columns. Commas may exist in protein names or FASTA protein descriptions, thus it is strongly urged to be sure all extra commas are removed.

Operation

Complete details of DiLeu Tool calculations and logic trees can be found in **Appendix III.** Briefly, the program begins by reading Parisony Proteins and cataloging all protein groups by their assigned numbers. Then, it reads Parsimony Peptides or Parsimony Phosphopeptides for quantification of proteins or phosphoproteoforms, respectively. In protein quantitation, DiLeu Tool catalogs information from PSMs that were only assigned to one protein group. In phosphoproteoform quantitation, PSMs that match to multiple protein groups will be used if they are unshared among proteoforms. If a PSM meets one of these criteria, it is considered a "quantifiable PSM." The names of all .mgf input files are also collected at this time so that the user need only provide their local directory.

Next, DiLeu Tool parses every MS/MS scan of the .mgf files created from the original raw acquisition files. If a scan maps to a quantifiable PSM, the *m*/*z* and intensities from the reporter region are recorded. Once all .mgf files are read, DiLeu Tool parses each quantifiable PSM and identifies which *m*/*z*-intensity pairs correspond to specific tag reporters using a user-specified mass accuracy threshold. Any quantifiable PSM with missing tag intensity is subsequently removed, as are any protein groups that no longer have quantifiable PSMs.

Users have the option of choosing any combination of DiLeu corrections and normalizations: purity-correction, channel-normalization, and mean intensity. Purity-corrections are performed first. Due to its complexity, the details of purity-correction will be discussed in a separate section.

Channel-normalization will correct systematic biases within a specific tag. First, the intensity for each tag is summed across all quantifiable PSMs to get the total channel intensity (**TCI**). The correction factor (**CF**) for a given tag *n* is obtained through the following equation:

$$CF_n = \frac{TCI_{max}}{TCI_n}$$

The DiLeu Tool will then multiply the individual reporter intensities from each quantifiable PSM by their respective tag CF, yielding channel-normalized intensities.

If the mean intensity option is not selected, protein-level and proteoform-level summed intensities will be calculated. The intensity of a specific reporter will be summed across all quantifiable PSMs belonging to a given protein or proteoform, and this sum will be assigned as the protein/proteoform-level intensity for that tag. If the mean intensity option is selected, each quantifiable PSM's reporter spectrum is normalized so that the intensity of the tag with the lowest abundance is equal to 1000. The PSMs are then summed as before, and the protein/proteoform-level intensity is divided by the number of quantified PSMs.

Purity-corrections

Purity-corrections are based on the approach outlined by Shadforth et al.[45] and has previously been used with DiLeu.[24, 46-47] However, high-resolution DiLeu 12-plex

correction requires special considerations. The -1 isotopes due to impure 13C, 15N, and 2H can be resolved, and not all of them may interfere with another reporter.[22]

Briefly, a system of *n* equations for experiments using *n* tags are created in the form of an *n* x *n* matrix. The following was used for experiments in this chapter:

	115a	115b	116a	116b	116c	117a	117b	117c	118a	118b	118c	118d
115a	0.9290	0.0000	0.0084	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
115b	0.0000	0.9406	0.0097	0.0156	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
116a	0.0645	0.0000	0.9273	0.0000	0.0000	0.0120	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
116b	0.0000	0.0548	0.0000	0.9385	0.0000	0.0084	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
116c	0.0000	0.0000	0.0000	0.0000	0.9092	0.0000	0.0068	0.0291	0.0000	0.0000	0.0000	0.0000
117a	0.0000	0.0000	0.0546	0.0000	0.0000	0.9364	0.0000	0.0000	0.0287	0.0000	0.0000	0.0000
117b	0.0000	0.0000	0.0000	0.0459	0.0000	0.0000	0.9082	0.0000	0.0000	0.0285	0.0000	0.0000
117c	0.0000	0.0000	0.0000	0.0000	0.0626	0.0000	0.0000	0.8946	0.0000	0.0134	0.0098	0.0000
118a	0.0000	0.0000	0.0000	0.0000	0.0000	0.0432	0.0000	0.0000	0.9317	0.0000	0.0000	0.0000
118b	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0551	0.0000	0.0000	0.8828	0.0000	0.0000
118c	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0472	0.0000	0.0000	0.9190	0.0000
118d	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.9415

Each matrix entry is the portion of the true intensity of the tag specified in the column header that contributes to the signal observed at the *m*/*z* for the reporter specified in the row header. From this matrix, a MathCad solve block is used to create a second matrix:

	115a	115b	116a	116b	116c	117a	117b	117c	118a	118b	118c	118d
115a	1.0771	0.0000	-0.0098	0.0000	0.0000	0.0001	0.0000	0.0000	-3.88E-06	0.0000	0.0000	0.0000
115b	0.0008	1.0641	-0.0111	-0.0177	0.0000	0.0003	0.0000	0.0000	-9.30E-06	0.0000	0.0000	0.0000
116a	-0.0749	0.0000	1.0799	0.0000	0.0000	-0.0139	0.0000	0.0000	0.0004	0.0000	0.0000	0.0000
116b	-0.0001	-0.0622	0.0012	1.0665	0.0000	-0.0095	0.0000	0.0000	0.0003	0.0000	0.0000	0.0000
116c	-3.21E-08	-2.37E-05	4.63E-07	0.0004	1.1023	-3.65E-06	-0.0083	-0.0359	1.13E-07	0.0008	0.0004	0.0000
117a	0.0044	0.0000	-0.0630	0.0000	0.0000	1.0702	0.0000	0.0000	-0.0330	0.0000	0.0000	0.0000
117b	4.25E-06	0.003145195	-6.13E-05	-0.0540	0.0000	0.0005	1.1032	0.0000	-1.49E-05	-0.0356	0.0000	0.0000
117c	6.23E-09	4.60725E-06	-8.977E-08	-7.90E-05	-0.0771	7.08E-07	0.0016	1.1209	-2.18E-08	-0.0171	-0.0120	0.0000
118a	-0.0002	0.0000	0.0029	0.0000	0.0000	-0.0496	0.0000	0.0000	1.0749	0.0000	0.0000	0.0000
118b	-2.65E-07	-0.000196303	3.82E-06	0.0034	0.0000	-3.01E-05	-0.0689	0.0000	9.30E-07	1.1349	0.0000	0.0000
118c	-3.20E-10	-2.37E-07	4.61E-09	4.06E-06	0.0040	-3.64E-08	-8.30E-05	-0.0576	1.12E-09	0.0009	1.0887	0.0000
118d	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0621

The entries in this matrix are coefficients for the observed signal intensity at the *m*/*z* for the reporter specified in the column header. To get the purity-corrected intensity of a tag

specified in the row header, the products of the coefficient and corresponding observed signal intensity are summed.

Table S1. All phosphoproteoforms that meet significance criteria for SD/GD or S/GD comparisons (Log_2(ratio) ≥ 0.60 or Log_2(ratio) ≤ -0.60, p-value < 0.05). Entries with SD/S comparisons also meeting significance criteria are highlighted in yellow.

Number	Gene	Proteoform	PSMs	Log_2(SD/GD)	p-value	Log_2(S/GD)	p-value	Log_2(SD/S)
1	Epb41l1	S(1454)	1	1.60	8.77E-03	1.54	4.58E-05	0.06
2	Zfpm1	S(929)	1	1.59	5.26E-03	1.29	1.13E-02	0.30
3	F1LX60*	S(89)	2	1.54	3.16E-04	1.40	2.60E-03	0.14
4	Ttc9	S104)	2	1.33	7.07E-03	1.62	1.41E-02	-0.29
5	Fam65a	S(36)	2	1.26	2.72E-03	1.25	3.67E-03	0.01
6	Pdlim4	S(134)	6	1.18	7.58E-03	1.27	1.15E-03	-0.09
7	Slc4a7	S(1057)	2	1.06	7.19E-02	1.53	3.09E-02	-0.47
8	Ybx1	S(174)	1	1.04	5.70E-03	0.45	2.82E-01	0.59
9	Pdxdc1	T(687);T(691)	5	0.95	2.92E-02	-0.16	8.75E-01	1.11
10	Phldb1	S(431)	4	0.94	4.19E-02	0.81	4.78E-02	0.13
11	Spp1	S(26)	27	0.94	1.09E-02	0.63	3.43E-02	0.31
12	Kidins220	S(1673)	1	0.94	4.75E-02	0.58	9.65E-02	0.36
13	Stx12	S(142)	3	0.92	2.31E-05	0.44	3.08E-02	0.49
14	Map1b	S(1257)	2	0.91	1.34E-03	0.53	1.60E-01	0.38
15	Itsn1	S(315)	2	0.88	4.64E-04	0.74	4.07E-03	0.14
16	Farp1	S(376)	1	0.87	3.31E-02	0.10	5.14E-01	0.78
17	Scrib	S(506)	1	0.86	4.27E-02	0.48	1.13E-01	0.38
18	Synpo	S(134)	3	0.84	7.57E-02	0.63	1.96E-02	0.21
19	Tmem119	S(280)	2	0.83	3.78E-02	0.95	3.46E-03	-0.13
20	Fam114a1	S(119)	2	0.82	2.05E-03	0.84	6.50E-02	-0.02
21	Cask	S(582)	2	0.82	4.82E-04	0.39	6.87E-03	0.42
22	Svil	S(240)	1	0.80	1.77E-02	0.67	7.31E-02	0.13
23	Nucks1	S(223)	4	0.80	2.93E-03	0.40	1.16E-01	0.40
24	Gprc5a	S(346)	10	0.79	1.20E-02	0.46	1.42E-03	0.33
25	Svil	S(1176)	1	0.79	2.90E-02	0.60	5.46E-04	0.19
26	Fam65a	S(361)	6	0.78	2.02E-03	0.97	1.75E-02	-0.18
27	Jmy	S(56)	2	0.77	3.46E-02	0.87	2.41E-03	-0.10
28	Spp1	S(313)	8	0.75	2.68E-03	0.71	9.84E-03	0.04
29	Pcdh1	S(934)	2	0.74	1.27E-02	0.88	1.37E-02	-0.13
30	Farp1	S(373)	3	0.73	5.04E-03	-0.04	7.49E-01	0.77
31	Spp1	S(312)	22	0.72	8.00E-03	0.66	3.62E-02	0.05
32	Pdlim4	S(111)	27	0.71	1.54E-02	0.89	7.42E-03	-0.18
33	Plekha4	S(562)	4	0.70	4.62E-03	0.84	8.97E-05	-0.14
34	Homer3	S(253)	1	0.68	8.35E-03	0.42	1.37E-01	0.26
35	Lmna	S(424)	3	0.68	9.79E-04	0.22	2.88E-03	0.46
36	Gramd1a	S(353)	2	0.68	1.02E-03	0.69	6.86E-04	-0.01

37	Ahnak	S(5390)	6	0.67	1.36E-02	0.70	7.00E-03	-0.04
38	Htatsf1	S(631)	9	0.66	1.99E-02	0.36	1.52E-01	0.30
39	Pdxdc1	T(687)	52	0.66	2.79E-04	0.46	1.21E-03	0.20
40	Lmna	S(390)	87	0.64	2.77E-02	0.63	3.61E-05	0.01
41	Synpo	T(69)	17	0.64	3.48E-03	0.51	3.00E-02	0.13
42	Baiap2	S(326)	5	0.63	3.21E-03	0.73	1.42E-03	-0.11
43	Man1c1	S(159)	6	0.62	2.54E-02	0.29	1.07E-01	0.33
44	Phldb1	S(387)	3	0.62	4.10E-02	0.72	2.18E-02	-0.10
45	Bag3	S(176)	4	0.61	2.13E-02	1.22	1.70E-02	-0.61
46	Tcea1	S(100)	10	0.61	4.93E-03	0.55	4.07E-03	0.06
47	Tnks1bp1	S(1448)	2	0.61	9.25E-03	0.22	2.94E-03	0.38
48	Tmem184b	S(395)	4	0.60	4.16E-02	1.11	2.42E-02	-0.51
49	Baiap2	S(367)	2	0.59	1.71E-02	0.76	1.17E-02	-0.17
50	Nedd4l	S(434)	1	0.59	8.49E-02	0.75	2.13E-02	-0.16
51	Akt1s1	S(184)	1	0.58	1.03E-01	1.24	1.67E-02	-0.65
52	Cxcr4	S(316)	2	0.56	1.09E-02	0.98	6.99E-03	-0.42
53	Ripk2	S(356)	4	0.56	6.34E-02	0.99	3.18E-02	-0.43
54	Sec16a	S(2041)	3	0.52	8.70E-03	0.67	5.62E-03	-0.15
55	Mtus1	S(1208)	2	0.50	1.24E-01	0.80	4.94E-04	-0.30
56	Bclaf1	S(760)	2	0.45	6.76E-02	0.66	2.31E-02	-0.22
57	Plec	S(4393)	2	0.45	1.66E-01	0.61	3.51E-02	-0.16
58	Snx17	S(409)	2	0.42	7.17E-02	0.61	1.93E-02	-0.18
59	Dpysl3	S(635)	9	0.39	8.06E-03	1.10	2.82E-03	-0.71
60	Bag3	S(277)	1	0.39	2.96E-01	0.72	3.93E-02	-0.33
61	Hdac5	S(483)	3	0.37	1.35E-01	0.69	1.19E-02	-0.32
62	Map2	T(1703)	2	0.37	2.30E-02	0.74	1.09E-03	-0.37
63	Frmd4a	S(548)	4	0.34	7.70E-02	0.74	4.37E-03	-0.40
64	Cd44	T(727)	8	0.33	1.42E-01	0.64	1.74E-02	-0.31
65	Synpo	S(767)	3	0.31	4.12E-01	0.64	5.47E-03	-0.33
66	Ulk1	S(450)	2	0.31	1.67E-01	0.64	2.58E-03	-0.33
67	Dpysl3	T(627)	2	0.25	2.73E-01	0.70	1.88E-02	-0.45
68	Zeb1	S(682)	3	0.20	3.67E-01	0.65	2.89E-03	-0.45
69	Phldb1	S(540)	1	0.18	2.54E-01	0.75	1.41E-02	-0.58
70	Tnks1bp1	S(1422)	2	0.16	5.14E-01	0.61	8.75E-03	-0.45
71	Dclk1	S(307)	2	0.16	5.62E-01	0.76	1.04E-02	-0.61
72	Herc1	S(1406)	2	0.14	1.74E-01	0.66	1.20E-02	-0.52
73	Lrrfip2	S(129)	1	0.10	5.26E-01	0.63	2.06E-03	-0.53
74	Specc1l	S(923)	1	0.09	6.06E-01	0.64	3.04E-02	-0.55
75	Map3k4	S(207)	1	0.07	5.72E-01	0.83	2.24E-02	-0.76
76	Ric1	S(1036)	2	0.04	5.20E-01	0.92	7.33E-03	-0.88
77	Eml3	S(156)	2	0.02	9.08E-01	1.02	3.01E-02	-1.01
78	Tsc2	S(1132)	1	0.01	8.12E-01	0.65	3.99E-03	-0.63
79	Tmpo	S(155)	2	0.00	8.95E-01	-0.73	2.90E-02	0.73

80	Trim24	S(210)	4	-0.06	7.00E-01	0.63	3.85E-04	-0.69
81	Filip1l	T(992)	11	-0.19	2.74E-01	-0.67	2.10E-02	0.48
82	Ranbp2	S(1507)	1	-0.20	6.90E-01	-0.65	1.82E-02	0.45
83	Fhl2	Y(158)	3	-0.21	4.47E-01	-0.61	3.93E-02	0.40
84	Tes	S(168)	2	-0.25	3.55E-01	-0.64	4.44E-02	0.39
85	Ppp1r18	S(145)	1	-0.27	2.79E-01	-0.71	2.55E-02	0.44
86	Sorbs1	S(262)	1	-0.27	1.83E-01	-0.70	3.96E-02	0.43
87	Plec	S(1439)	2	-0.28	3.13E-02	-0.63	5.78E-03	0.35
88	Srrm1	S(718)	1	-0.28	1.37E-01	-0.64	1.89E-02	0.36
89	Smtn	S(513)	2	-0.29	2.25E-01	-0.91	3.39E-02	0.63
90	Myh11	S(645)	2	-0.33	5.10E-02	-0.61	3.01E-03	0.27
91	Rbms2	T(38)	2	-0.34	1.13E-01	-0.87	1.96E-02	0.53
92	Enah	T(462)	4	-0.35	1.35E-01	-0.73	4.83E-02	0.38
93	Myadm	S(14)	1	-0.40	3.34E-01	-0.72	3.25E-02	0.32
94	Pcdh7	S(974)	12	-0.42	1.78E-03	-0.68	5.62E-04	0.26
95	Pcm1	T(977)	1	-0.42	1.40E-02	-0.93	7.41E-03	0.51
96	Slc33a1	S(484)	2	-0.42	1.53E-02	-0.74	1.12E-04	0.32
97	Epb41l5	S(549)	4	-0.43	5.44E-02	-0.60	3.18E-02	0.17
98	Rnf6	S(183)	2	-0.44	1.51E-02	-0.72	5.44E-03	0.28
99	Cald1	S(354)	33	-0.45	5.76E-02	-0.66	2.62E-02	0.21
100	Arvcf	S(204)	1	-0.50	4.10E-03	-0.61	1.23E-02	0.11
101	Pkp2	S(71)	1	-0.51	6.73E-02	-0.80	2.96E-02	0.29
102	Dync1li2	S(446)	1	-0.51	6.84E-02	-0.81	6.54E-03	0.30
103	Peak1	S(370)	2	-0.51	2.45E-02	-0.85	1.15E-02	0.34
104	Rasal2	T(760)	4	-0.51	5.77E-03	-0.84	5.02E-03	0.33
105	Flna	S(8)	3	-0.51	2.33E-02	-0.76	2.31E-03	0.24
106	Arhgap21	T(325)	11	-0.52	2.62E-02	-0.65	1.18E-02	0.14
107	Ctnnd2	S(250)	2	-0.54	2.09E-02	-0.63	2.85E-02	0.10
108	Pdlim7	T(251)	15	-0.55	3.10E-02	-0.70	3.33E-03	0.16
109	Whsc1	S(417)	2	-0.56	4.69E-02	-0.70	3.97E-02	0.15
110	Terf2	S(385)	2	-0.56	8.30E-03	-0.66	1.39E-02	0.10
111	Neo1	S(1047)	1	-0.56	2.69E-02	-0.69	2.06E-02	0.13
112	Pdlim5	S(111)	1	-0.58	4.70E-03	-0.65	1.28E-02	0.07
113	Flna	S(1084)	36	-0.59	9.85E-03	-0.62	9.76E-03	0.03
114	Chaf1a	S(293)	6	-0.60	2.95E-03	-0.73	2.78E-03	0.13
115	Sipa1l1	S(1451)	2	-0.60	1.71E-02	-0.69	2.79E-02	0.09
116	Smad3	S(418)	2	-0.60	1.34E-02	-0.79	1.58E-02	0.19
117	Tns1	S(812)	1	-0.61	7.20E-03	-1.24	1.21E-03	0.63
118	Wwp2	S(99)	4	-0.61	7.59E-03	-0.96	9.93E-03	0.34
119	Mybbp1a	T(1258)	1	-0.62	2.37E-02	-0.68	1.06E-02	0.07
120	Plekha2	S(184)	1	-0.62	4.01E-02	0.45	2.01E-02	-1.07
121	Pdcd4	S(94)	1	-0.62	2.08E-02	-0.33	1.75E-01	-0.29
122	Ablim1	S(136)	2	-0.63	1.57E-02	-0.11	9.74E-01	-0.52

123	Il6st	S(678)	5	-0.63	3.07E-04	-0.52	6.54E-03	-0.11
124	Filip1l	S(1047)	11	-0.63	2.96E-02	-0.04	6.67E-01	-0.59
125	Ankrd50	S(1137)	1	-0.63	1.51E-02	-0.68	1.19E-02	0.05
126	Vim	S(42)	5	-0.63	1.36E-02	-0.83	1.72E-02	0.19
127	Clasp1	S(590)	2	-0.64	2.11E-02	-0.09	8.84E-01	-0.55
128	Trim47	S(84)	2	-0.65	1.24E-02	-0.63	1.95E-02	-0.01
129	Tax1bp1	S(625)	1	-0.65	4.06E-02	-0.65	4.46E-02	0.00
130	Lmo7	S(1637)	1	-0.65	4.43E-02	-0.50	1.15E-01	-0.15
131	Rab12	S(105)	4	-0.65	8.17E-03	-0.60	8.34E-03	-0.05
132	Araf	S(255)	1	-0.65	3.30E-02	-0.29	4.46E-01	-0.36
133	Foxs1	S(138)	6	-0.65	1.11E-03	-0.53	2.42E-03	-0.12
134	Yeats2	S(534)	1	-0.66	1.78E-02	-0.43	9.15E-02	-0.23
135	St5	S(32)	6	-0.66	6.38E-03	-0.58	4.95E-02	-0.07
136	Arhgap24	S(369)	3	-0.66	4.25E-02	-0.80	3.77E-02	0.14
137	Bcar3	S(370)	3	-0.66	7.76E-03	-0.64	1.25E-02	-0.02
138	Spata6	S(354)	3	-0.67	8.02E-03	-0.63	1.09E-02	-0.04
139	Dlgap5	S(321)	2	-0.67	1.15E-02	-0.57	6.15E-03	-0.10
140	Pdlim5	S(228)	40	-0.68	4.04E-02	-0.44	1.14E-01	-0.23
141	Sdpr	S(292)	2	-0.68	4.43E-02	-0.56	8.08E-02	-0.12
142	Ssfa2	S(268)	1	-0.69	3.89E-03	-0.56	1.63E-02	-0.13
143	Pln	S(16)	20	-0.69	2.28E-04	-0.49	1.23E-03	-0.20
144	Cul4b	T(106)	1	-0.70	9.59E-02	-1.15	4.40E-02	0.45
145	Peak1	S(568)	6	-0.71	3.47E-03	-0.68	5.49E-03	-0.03
146	Rnf219	S(717)	2	-0.71	2.85E-02	-0.66	3.90E-02	-0.05
147	Prkcdbp	S(56)	2	-0.71	1.71E-02	-0.76	7.76E-03	0.04
148	Flna	S(11)	4	-0.73	7.42E-04	-0.87	2.04E-03	0.15
149	Lmod1	S(550)	19	-0.74	8.05E-04	-0.57	1.36E-02	-0.17
150	RGD1559896	S(37)	1	-0.74	4.07E-02	-0.56	9.67E-02	-0.18
151	Mex3d	S(488)	2	-0.75	2.65E-02	-0.05	5.98E-01	-0.71
152	Ppap2a	S(264)	2	-0.76	2.14E-02	-0.85	2.12E-02	0.09
153	Yeats2	S(464)	2	-0.76	1.43E-02	-0.90	1.24E-02	0.14
154	Chtopl1	S(40)	2	-0.77	1.38E-02	-0.58	2.99E-02	-0.19
155	Ptbp1	S(437)	7	-0.78	4.97E-03	-0.64	2.71E-02	-0.13
156	Etl4	S(379)	2	-0.78	1.04E-02	-0.68	3.37E-02	-0.09
157	Rbm7	T(58)	4	-0.78	8.64E-03	-0.83	1.50E-02	0.05
158	Akap1	S(375)	4	-0.78	6.31E-04	-0.75	4.29E-03	-0.03
159	Nr3c1	S(155)	2	-0.79	2.21E-04	0.38	1.21E-02	-1.17
160	LOC100911319	S(143)	2	-0.80	4.04E-03	-0.55	1.72E-02	-0.25
161	Peak1	S(281)	4	-0.80	2.69E-04	-0.78	1.01E-03	-0.01
162	Snx33	S(177)	2	-0.80	1.63E-02	-0.86	6.19E-03	0.06
163	Rrp1b	S(538)	1	-0.81	3.73E-02	-0.72	7.10E-02	-0.09
164	Synpo2	S(900)	8	-0.81	1.98E-03	-1.01	5.68E-04	0.20
165	Ddx17	S(522)	3	-0.81	4.15E-02	-0.42	2.86E-02	-0.40

166	Tab2	S(450)	1	-0.82	1.10E-02	0.11	1.35E-01	-0.93
167	Ahnak	S(2691)	1	-0.82	6.17E-03	1.58	9.49E-05	-2.40
168	Peak1	S(1361)	4	-0.84	6.51E-03	-0.54	2.45E-02	-0.30
169	RGD1565536	T(507)	1	-0.85	8.33E-03	-0.79	1.89E-02	-0.07
170	Lmnb1	T(576)	7	-0.86	1.25E-04	-0.89	4.04E-04	0.03
171	LOC684828	T(147)	3	-0.86	2.14E-02	-1.24	1.19E-02	0.38
172	Smtn	S(357)	9	-0.86	2.23E-03	-1.04	4.31E-03	0.18
173	Ppp1r13l	S(154)	2	-0.86	9.25E-03	-1.25	1.18E-03	0.39
174	Hnrnpa0	S(280)	2	-0.90	5.31E-03	-0.94	7.74E-03	0.04
175	RGD1559896	S(171)	1	-0.90	2.41E-02	-0.68	3.28E-02	-0.22
176	Dnajb4	S(148)	1	-0.92	2.12E-02	-1.02	2.50E-02	0.10
177	Zfp36	S(185)	2	-0.92	3.95E-03	-0.11	9.64E-01	-0.81
178	Mta1	S(576)	1	-0.93	6.45E-04	-1.07	2.15E-02	0.14
179	Pdlim7	S(260)	1	-0.95	2.48E-02	-1.66	4.28E-03	0.71
180	Dnmt1	S(125)	1	-0.96	3.98E-03	-1.02	1.38E-02	0.07
181	Nfx1	S(28)	1	-0.96	4.13E-02	-0.39	3.35E-01	-0.56
182	Etl4	S(1856)	5	-0.97	6.44E-04	-0.63	7.10E-04	-0.35
183	Rbm14	S(256)	4	-0.98	4.57E-04	0.00	4.34E-01	-0.98
184	Aurkb	S(17)	1	-0.98	6.80E-03	-1.11	6.09E-03	0.13
185	Nolc1	S(421)	1	-1.01	8.58E-04	-0.92	3.69E-03	-0.08
186	Numa1	T(2078)	1	-1.05	2.19E-03	-1.29	1.42E-03	0.24
187	Zyx	S(46)	4	-1.05	6.94E-03	-1.13	6.87E-03	0.08
188	Chaf1b	S(417)	1	-1.05	1.56E-02	-1.29	1.84E-02	0.24
189	Ddx3	S(463)	1	-1.05	8.54E-03	-0.78	2.94E-02	-0.28
190	Cryab	T(158)	1	-1.05	1.16E-03	-0.98	1.32E-02	-0.08
191	Flna	S(1010)	4	-1.08	3.21E-03	-0.99	7.79E-03	-0.08
192	Hspb1	S(15)	25	-1.08	1.97E-02	-0.79	2.25E-02	-0.29
193	Aurkb	S(47)	1	-1.09	4.86E-04	-0.72	4.90E-03	-0.37
194	Flna	S(2632)	2	-1.12	1.88E-02	-1.29	1.51E-02	0.17
195	Srsf7	S(227)	1	-1.13	3.03E-02	-1.03	5.80E-02	-0.09
196	Flnc	S(2656)	2	-1.13	1.06E-02	-1.31	8.78E-03	0.18
197	Pdlim7	S(111)	2	-1.15	2.06E-02	-0.74	3.80E-02	-0.41
198	Cnn2	T(186)	1	-1.19	3.48E-03	-0.97	3.56E-03	-0.22
199	Cnn2	S(177)	1	-1.24	2.27E-02	-1.38	1.35E-02	0.14
200	42256	T(107)	1	-1.27	5.94E-04	-0.56	2.08E-02	-0.71
201	Filip1l	S(1039)	6	-1.29	7.99E-05	0.08	3.45E-01	-1.37
202	Bves	S(323)	1	-1.30	1.71E-02	-0.85	2.17E-02	-0.45
203	Smtn	S(301)	2	-1.34	1.39E-04	-1.12	1.43E-03	-0.22
204	Hey2	S(39)	6	-1.38	7.24E-04	-1.42	3.28E-03	0.03
205	Hspb1	S(86)	60	-1.41	2.75E-04	-1.20	2.25E-04	-0.21
206	Cryab	S(59)	34	-1.48	6.37E-03	-1.43	6.17E-03	-0.05
207	Flna	S(1938)	11	-1.76	3.52E-04	-1.80	1.18E-04	0.04
208	Flna	S(468)	1	-1.93	1.95E-02	-2.14	1.70E-02	0.21

209	Flnc	S(2237)	4	-2.06	2.29E-04	-1.95	4.25E-04	-0.11
210	Smtn	T(788)	2	-2.15	1.99E-02	-2.14	2.49E-02	-0.01
211	Nabp1	S(170)	1	-2.20	1.82E-02	-2.29	2.35E-02	0.08

Chapter 8

Conclusions and Outlook

The research projects detailed in this dissertation were attempts to contribute to the field of analytical chemistry. In **Chapter 2**, the practice of analyzing cis/trans proline isomer with ion mobility (**IM**)-mass spectrometry (**MS**) was critically examined. For the first time, a discrete example of an experimental cis/trans "signature" with theoretical evidence for an alternative explanation was presented. **Chapter 3** and **Chapter 4** exploited molecular properties of charge density and conformation to yield new ways of making measurements using IM-MS technology. In **Chapter 3**, IM-MS was able to separate and distinguish quantitative signals from peptides that would have otherwise overlapped while using MS alone. In **Chapter 4**, high-throughput proteomics and collision cross-section (**CCS**) measurements were synergistically combined to create the largest peptide CCS database to-date. Prior to that investigation, creating such a large database was not pragmatic. Free from practical limitations, it was demonstrated that data from large CCS databases can illustrate complex patterns of a peptide's intrinsic structural preferences.

Chapter 5 marked thematic departures from IM-MS and method development-centric research. First, the unmet needs of protein analysis in the study of neuropathic pain were stated, followed by a review of current state of MS-based analyses in neuropathic pain. A comprehensive proteomic study of neuropathic pain was then presented in **Chapter 6** and demonstrated how a MS-based approach can achieve large-scale analysis. More importantly, it generated a novel hypothesis that translation of Bag3 was up-regulated in the neuropathic pain model and was later confirmed by Western blot. **Chapter 7** showcased another example of utilizing high throughput multiplex quantitative proteomics, to the study of phosphoprotein regulation in restenosis. This investigation is

still ongoing, although it has also produced a new hypothesis that will soon be subjected to orthogonal testing. These proteomics chapters did implement some novel methodology, but the emphasis was clearly on application. Successful applications of cutting-edge technologies to solve real biological problems are perhaps among the most challenging and important duties of an analytical chemist, as increased use and impact of any method will only come from powerful demonstrations of utility and practicality.

All of the chapters in this dissertation suggested future experiments to follow-up their conclusions. **Chapter 2** proposed going beyond the single peptide system that was presented to see if the frequency of contradicting experimental IM and theoretical evidence for cis/trans isomerization was common or negligible. For the newly developed methods in **Chapter 3** and **Chapter 4**, it was suggested that they be coupled to more powerful instrumentation to enable widespread applications to increasingly complex systems. These chapters were effective at providing proofs-of-principle in model systems, but real samples—especially proteomic samples that would implement IM-MS for accurate isobaric tag quantification—might require higher IM resolution and much faster MS scans. In **Chapter 7**, detailed future experiments aimed at validating the pathological importance of putative phosphoprotein targets were outlined.

In **Chapter 6**, future experiments were also proposed for neuropathic pain systems. However, as the inclusion of **Chapter 5** in this dissertation may suggest, proteomic analysis of pain was a significant portion of my graduate research. Two additional studies of the spared-nerve injury (**SNI**) neuropathic pain model were performed as follow-up to **Chapter 6**. Preliminary results can be found in **Appendix IV** and **Appendix V**. It is expected that this work will be published in a future manuscript.

Every bit of empirical knowledge humans have ever learned was the result of a measurement. Although it is advisable for scientists in all disciplines to have some exposure to the fundamentals of their instruments and measurement rationale, the greatest such burden is shouldered by the analytical chemist. Their collective job is to understand how molecules interact with their environments, develop platforms to observe phenomena resulting from those interactions, and then find a way to use the observables as a vehicle to characterize molecular systems of interest. At the very least, I would characterize my graduate research as my best effort to perform the duties of an analytical chemist.

Appendix I

Curriculum Vitae

Christopher B. Lietz

University of Wisconsin-Madison
777 Highland Avenue, Madison, WI 53705
clietz@chem.wisc.edu

EDUCATION/TRAINING

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	YEAR(s)	FIELD OF STUDY
Wayne State University, Detroit, MI	B.S.	2005-2011	Chemistry
University of Wisconsin-Madison, Madison, WI	Ph. D. student	2011-2016	Analytical Chemistry

A. Honors

2010-2011 Dean's List, Wayne State University
2010 Undergraduate Research Grant, Wayne State University
2010 Undergraduate Chemistry Scholarship, Wayne State University
2010 Chemistry Chair's Honor List, Wayne State University
2011 American Chemical Society Division of Analytical Chemistry Award, Wayne State University
2011 Honored Instructor Award, University of Wisconsin-Madison

B. Fellowships and Training Grants

2012-2016 Chemistry-Biology Interface Training
2013-2016 National Science Foundation Graduate Student Research Fellowship

C. Publications (17 published, 1 under revision, 1 submitted, and 1 in preparation)

(1) Zhong X,* **Lietz CB**,* Shi X, Buchberger A, Frost DC, Kent KC, Li L. Quantitative phosphoproteomics of TGF-β and SDF-1 treatment in smooth muscle cells overexpressing Smad3. In preparation. *Denotes equal contribution.

(2) **Lietz CB**,* Tilley DM,* Kelley CA, Cedeño DL, Williams J, Benyamin R, Li L, and Vallejo R. Quantitative proteomics of the spinal cord and dorsal root ganglion following peripheral nerve injury in a rat model of neuropathic pain. *Journal of Proteome Research*, submitted. *Denotes equal contribution.

(3) **Lietz CB**, Chen Z, Son C, Pang X, Cui Q, Li L. Multiple gas-phase conformations of proline-containing peptides: Is it always cis/trans isomerization? *Analyst*, **2015**, under revision.

(4) Greer T, **Lietz CB**, Xiang F, Li L. Novel isotopic N,N-dimethyl leucine (iDiLeu) reagents enable absolute quantification of peptides and proteins using a standard curve approach. *Journal of the American Society for Mass Spectrometry.* **2015**, *26*, 107-119.

(5) Chen B, **Lietz CB**, Li L. In situ characterization of proteins using laserspray ionization on a highperformance MALDI-LTQ-Orbitrap mass spectrometer. *Journal of the American Society for Mass Spectrometry*. **2014**, *25*, 2177-2180.

(6) **Lietz CB**, Yu Q, Li L. Large-scale collision cross-section profiling on a travelling wave ion mobility mass spectrometer. *Journal of the American Society for Mass Spectrometry*. **2014**, *25,* 2009-2019.

(7) Jia C, **Lietz CB**, Yu Q, Li L. Site-specific characterization of d-amino acid containing peptide epimers by ion mobility spectrometry. *Analytical Chemistry*. **2014**, *86*, 2972-2981.

(8) Sturm RM,* **Lietz CB**,* Li L. Improved isobaric tandem mass tag quantification by ion mobility mass spectrometry. *Rapid Communications in Mass Spectrometry*. **2014**, *28*, 1051-1060. *Denotes equal contribution

(9) Jia C, Wu Z, **Lietz CB**, Liang Z, Cui Q, Li L. Gas-phase ion isomer analysis reveals the mechanism of peptide sequence scrambling. *Analytical Chemistry*. **2013**, *86*, 2917-2924.

(10) Jia C, **Lietz CB**, Ye H, Hui L, Yu Q, Yoo S, Li L. A multi-scale strategy for discovery of novel endogenous neuropeptides in the crustacean nervous system. *Journal of Proteomics*. **2013**, *91*, 1-12.

(11) **Lietz CB**, Gemperline E, Li L, Qualitative and quantitative mass spectrometry imaging of drugs and metabolites. *Advanced Drug Delivery Reviews*. **2013**, *65*, 1074-1085.

(12) Jia C, Hui L, Cao W, **Lietz CB**, Jiang X, Chen R, Catherman AD, Thomas PM, Ge Y, Kelleher NL, Li L. High Definition De Novo Sequencing of Crustacean Hyperglycemic Hormone (CHH)-family Neuropeptides. *Molecular Cellular Proteomics*. **2012**, *11*, 1951-1964.

(13) Trimpin S, Wang B, Inutan ED, Li J, **Lietz CB**, Pagnotti VS, Harron A, Sardelis D, McEwen CN. A Mechanism for Ionization of Nonvolatile Compounds in Mass Spectrometry: Considerations from MALDI and Inlet Ionization, *Journal of the American Society of Mass Spectrometry*. **2012**, *23*, 16441660.

(14) Li J, Inutan ED, Wang B, Richards AL, **Lietz CB**, Green DR, Manly CD, Marshall DD, Lingenfelter S, Ren Y, Trimpin S. Matrix Assisted Ionization: New Aromatic and Non Aromatic Matrix

Compounds Producing Multiply Charged Lipid, Peptide, and Protein Ions in the Positive and Negative Mode Observed Directly from Surfaces. *Journal of the American Society of Mass Spectrometry.* **2012,** *23*, 1625-1643.

(15) Richards AL, **Lietz CB**, Trimpin S. Laserspray Ionization MS^n and High Spatial Resolution Tissue Imaging of Labile Gangilosides. *Journal of Lipid Research.* **2012,** *53*, 1390-1398.

(16) **Lietz CB**, Richards AL, Marshall DD, Ren Y, Trimpin S. Matrix Assisted Inlet Ionization and Solvent-free Gas-Phase Separation using Ion Mobility Spectrometry for Imaging and Electron Transfer Dissociation Mass Spectrometry of Polymers. *Mass Spectrometry in Polymer Chemistry, Wiley-VCH*, **2011** (Book Chapter).

(17) **Lietz CB**, Richards AL, Ren Y, Trimpin S. Inlet Ionization: Protein Analyses from the Solid State Without the Use of a Voltage or a Laser Producing 67 Charges on BSA, 66 kDa. *Rapid Communications in Mass Spectrometry.* **2011,** *25*, 3453–3456.

(18) Trimpin S, Ren Y, Wang B, **Lietz CB**, Richards AL, Marshall DD, Inutan ED. Extending the Laserspray Ionization Concept to Produce Highly Charged Ions at High Vacuum on a Time-of-Flight Mass Analyzer. *Analytical Chemistry.* **2011,** *83*, 5469-5475.

(19) Wang B, **Lietz CB**, Inutan ED, Leach S, Trimpin, S. Laserspray Ionization Ion Mobility Spectrometry Mass Spectrometry: A Total Solvent-Free Approach at Atmospheric Pressure, *Analytical Chemistry*. **2011**, *83*, 4076-4084.

(20) Richards AL, **Lietz CB**, Wager-Miller J, Mackie K, Trimpin S. Imaging Mass Spectrometry in Transmission Geometry. *Rapid Communications in Mass Spectrometry.* **2011,** *25*, 815-829.

D. Research and Work Experience

Research Scientist Intern, Promega

Madison, WI; February 2015 – May 2015

Intern for research and development team, assisting in the development of products and services related to chemoproteomics.

Graduate Research Assistant, UW-Madison Department of Chemistry

Madison, WI; July 2011-present

Method development and collaborative research in quantitative neuropeptidomics and proteomics.

Initial project was to develop a method using gas-phase ion mobility separation to enhance quantitative analysis in global proteomics and peptidomics. Projects to follow include the structural analysis of endogenous neuropeptides and their complexes with ion mobility spectrometry, characterization of the neuropeptide/protein interactome, structural characterization of a neuropeptide Y signal peptide mutation, and applications in quantitative global proteomics.

Teaching Assistant, UW Madison Department of Chemistry

Madison, WI; August 2011 present

Collaborated with other TAs and faculty members to teach undergraduate introductory and analytical chemistry courses.

Prepared students for weekly laboratory classes and planned and taught weekly discussion sessions.

Mentored a high school internship project in the UW-Madison PEOPLE Program in summer 2012.

Undergraduate Research Assistant, Wayne State University Department of Chemistry

Detroit, MI; November 2009 – July 2011

Designed and performed research in the laboratory of Dr. Sarah Trimpin focusing on the application of novel atmospheric pressure ionization sources for mass spectrometry. Lead projects involved with applying inlet ionization to intact proteins that resulted in a publication detailing the ionization of proteins as large as bovine serum albumin, 66 kDa. Was also involved in several novel lipid imaging experiments that resulted in two publications.

E. Outreach

PEOPLE Program, UW-Madison

Madison, WI; July 2012

Mentored a Wisconsin high school student in the PEOPLE Program, which recruits minority and low-income students and prepares them for higher education. Assisted student in chemistry-related project planning and execution.

F. Professional Affiliations

American Society for Mass Spectrometry

American Chemical Society

Appendix II

Qualitative and quantitative mass spectrometry imaging of drugs and metabolites

Adapted from: Christopher B. Lietz, Erin E. Gemperline, and Lingjun Li. Qualitative and quantitative mass spectrometry imaging of drugs and metabolites. *Advanced Drug Delivery Reviews,* **2013**, *65*, 1074-1085.

Abstract

Mass spectrometry imaging (**MSI**) has rapidly increased its presence in the pharmaceutical sciences. While quantitative whole-body autoradiography and microautoradiography are the traditional techniques for molecular imaging of drug delivery and metabolism, MSI provides advantageous specificity that can distinguish the parent drug from metabolites and modified endogenous molecules. This review begins with the fundamentals of MSI sample preparation/ionization, and then moves on to both qualitative and quantitative applications with special emphasis on drug discovery and delivery. Cutting-edge investigations on sub-cellular imaging and endogenous signaling peptides are also highlighted, followed by perspectives on emerging technology and the path for MSI to become a routine analysis technique.

Introduction

For many biological systems, a sufficient characterization may require more than a catalog of the molecules that are present. It may also rely on their anatomical distribution patterns and relative spatial relationships. Researchers in drug discovery and delivery are also in need to spatially characterize and quantify a drug's absorption, distribution, metabolism, and excretion (**ADME**). Quantitative whole-body autoradiography (**QWBA**) is among the most commonly used methods to do so.[1] It begins with dosing an animal specimen with a test drug labeled with radioactive beta-emitters such as ^{3}H or ^{14}C. After a desired amount of time, the specimen is euthanized, snap frozen, and sectioned.[2] Whole-body sections are then placed opposite to phosphor detectors and an image is produced within a few days to a few weeks. Radioactive standards are simultaneously imaged with the specimen to quantify the drug's penetration. Microautoradiography (**MARG**) is a similar method that is employed when high spatial resolution images are needed, but is not often used for quantitative purposes.[2]

Radiochemical imaging methods have proven to be an invaluable part of the mass balance study—that is, a quantitative account of a drug from ingestion through excretion.[3] But as reputable as they are, they possess fundamental shortcomings in answering questions about spatial relationships. The information in the QWBA or MARG images may rightfully allow an inductive leap to drug distribution, but the raw data itself only shows the detection of radiation. The ^{14}C isotope could be from the drug, a metabolite, or even a derivatized endogenous molecule, but a radiochemical method would not be able to distinguish the difference.

Over the past decade, mass spectrometry (**MS**) imaging (**MSI**) has gained considerable interest from the pharmaceutical community. In contrast to QWBA and MARG, MS detects the actual molecules in the image based on their characteristic mass-to-charge ratios (***m/z***) and does not require the use of labeled compounds. Using MSI, one can distinctly detect and characterize drugs, metabolites, and endogenous drug-modified molecules in their native states within a biological matrix. Furthermore, MSI can be a label-free method. No prior knowledge about target analyte is required, allowing for discovery of unknowns. Additionally, radioactive materials always present potential biological hazard, even under carefully controlled working conditions.

The aim of this review is to highlight the potential and realized benefits of drug and metabolite MSI and identify its practical capabilities and common pitfalls. It is primarily written for readers mostly unfamiliar with MSI, but can also be useful for MSI investigators interested in the latest applications in drug delivery and discovery. Many pharmaceuticals have molecular weights less than or equal to 1000 Da and would be considered small molecules in MS analysis. Therefore, this review will also cover methods for endogenous small molecule MSI that could also be applied to drugs and metabolites.

Many excellent MSI reviews have been published,[4-10] and the authors also recommend some informative radiochemical imaging reviews for comparison.[1-3] The following sections will begin with a very brief overview of fundamentals and will then heavily focus on applications in drug discovery and drug delivery. The applications will demonstrate what types of questions can be answered by MSI, and special attention will be paid to the efforts towards developing quantitative MSI. Because an accurate quantitative analysis is vital to ADME, absolute and relative quantification is one of the

most sought-after methods in pharmaceutical MSI. Additionally, short sections will be devoted to subcellular MSI and MSI of neuropeptides, two other areas with relevance to pharmaceutical research. Finally, this review will conclude with an outlook on emerging technology and the challenges MSI must overcome to achieve routine analysis in the pharmaceutical sciences.

Mass Spectrometry Imaging Fundamentals

From molecular ions to molecular image

This section of the review will explain the basic MSI workflow and highlight the pertinent fundamentals. **Scheme 1** is a graphical representation of an MSI experiment, from creating ions to forming an image. Typically, the first step of an *ex vivo* MSI experiment involves obtaining a specimen via snap-freezing and cryosectioning, followed by tissue section mounting (**Scheme 1A**). Spectra are collected from tissue in a raster pattern, creating a grid of points on the tissue where molecules have been ionized and detected according to their *m/z.* Once MS analysis is complete, each point is converted into a two-dimensional spatial coordinate (**Scheme 1B**). Finally, an image is constructed by displaying the intensity of a specific *m*/*z* at each coordinate (**Scheme 1C**). Real MSI results are shown in **Scheme 1D** with an optical image of a rat brain tissue followed by the mapped spatial distributions of eight *m*/*z* values belonging to different lipid species.[11] These images were acquired simultaneously, and one could theoretically acquire hundreds of molecular images in one experiment.[12]

The acquisition and construction of an image from ionized tissue compounds is often performed by one of two methods. In the first method, the intensity of an ion created directly from tissue is displayed at each sampled point. It is a simple and straightforward

approach, but it is easy to imagine that a tissue slice may contain many isomeric or isobaric compounds whose *m*/*z* overlap. This is especially true with certain methods that use small molecules to assist in ionization, and these small molecules fall into the mass range of the analyte. Therefore, using tandem mass spectrometry (**MS/MS**) is a common way to increase confidence in compound identification and improve the dynamic range of MSI.[13-14] **Scheme 2** presents an overview of selected reaction monitoring (**SRM**), a common MS/MS technique employed for imaging, through an illustrative example of clozapine imaging. After tissue compounds enter the mass spectrometer, the ions with an *m*/*z* of (327 ± 2) are isolated during the initial mass analysis. Clozapine and its isobaric species are then fragmented, and a second mass analysis isolates the signature fragment of clozapine (*m*/*z* 270.08) and uses its intensity to create an image. The downside to SRM-MSI is that it is difficult to use in a non-targeted study. The parent *m*/*z* and fragment *m*/*z* must be known *a priori*.

MSI has three primary figures of merit defined in **Scheme 3**. The first two are from MS: mass accuracy and mass resolution. Mass accuracy refers to the agreement of an ion's detected mass to its theoretical mass and is often measured in parts-per-million (**ppm**).[15] **Scheme 3A** shows an overlaid cholesterol image on mouse brain tissue constructed from ions detected within a window of 0.0005 Daltons (**Da**) of cholesterol's theoretical mass, 369.3515 Da.[16] The mass window corresponds to an approximate mass accuracy of 1.4 ppm to reflect the histogram below.

Mass resolution, or resolving power, describes the minimum difference between two *m*/*z* that can be identified as unique ions.[15] One way to calculate mass spectral resolution is to divide a peak's apex *m*/*z* by its full-width at half-maximum (**FWHM**). In

Scheme 3B, the images of three similar *m*/*z* values display very different spatial distributions in a mouse brain.[17] The top brain image is actually the summation of two isobaric lipids, whose unique distributions are shown below. Without mass resolution of at least 29000, they would falsely appear as a single peak and produce a false spatial distribution.

Spatial resolution is a term from microscopy imaging. It refers to the minimum distance between two objects in an image at which they can be distinctly discerned. In MSI, spatial resolution calculation is usually dependent on the type of ionization employed, but one can often use the distance between sampled points. If each pixel in an image represents the entire area of a sampled point, then the minimum distance that must exist between two objects is the center-to-center length between two sample areas. **Scheme 3C** shows time-resolved nitrogen enrichment in *Triticum aestivum* root cells. These secondary ion mass spectrometry images were acquired with less than 1 µm spatial resolution.[18]

Mass accuracy, mass resolution, and spatial resolution are not the only figures of merit that exist for MSI, but they are the most cited figures throughout the literature. The amount of resolution and accuracy needed to produce a quality image depends upon the compounds being imaged and the complexity of the tissue. For example, a drug with an exact mass of 239.1077 Da needs 419 ppm mass accuracy and 2400 mass resolution to be discerned from isobaric molecules with mass differences of 0.1 Da or greater. To be discerned from mass differences of 0.001 Da, 4.2 ppm mass accuracy and nearly mass resolution at 240000 is required. Fourier transform **(FT**) mass spectrometers are known for parts-per-billion mass accuracy and mass resolution exceeding 1000000,[19-20] but they

can be very expensive and often require longer acquisition times to construct an ion image. Time-of-flight (**TOF**) instruments have resolving powers ranging from 10000 to 100000 and can routinely achieve less than 5 ppm mass accuracy.[21] Quadrupole and ion trap instruments are usually considered robust “low resolution” instruments with nominal mass resolving powers. However, triple-quadrupole-ion trap instruments have proven to be very powerful in MSI when used for SRM.[22] When the first and third quadrupoles are set to constantly monitor a single transition or a couple of transitions with very short dwell time, the mass spectrometer is effectively turned into an ion counter and can offer unparalleled sensitivity. If the third quadrupole is used as an ion trap, the experiment can alternate between SRM scans, full MS scans, and enhanced product ion scans that can detected many fragmentation transitions from a single ion. This approach can allow nearly simultaneous targeted and non-targeted imaging.

The power of SRM-MSI should not be underestimated. Even using an instrument with the best specified mass accuracy and mass resolution could result in ambiguity for compound identification in a complex sample. Some isobaric compounds might not be resolved even with the use of high resolution FT-MSI, but might be easily imaged and identified using sequence-specific fragmentation via SRM.[23] High mass resolution will do nothing to resolve isomeric compounds with identical *m*/*z*, but if the isomers show different fragmentation patterns, SRM-MSI would be able to resolve them. SRM-MSI is most easily performed on an ion trap or any hybrid mass spectrometer containing two mass analyzers and a fragmentation cell (e.g., triple quadrupole, quadrupole-TOF, and TOF-TOF).

While mass accuracy and mass resolution rely heavily on instrumentation, spatial resolution is more dependent on sample preparation and ionization. The following fundamentals section will focus on common methods for sample preparation and MS ionization. Many excellent MSI reviews have covered traditional and novel preparation protocols extensively, so we will only detail the most basic aspects and will refer the readers to cited works for more information.

Sample preparation and ionization

The sacrifice of an animal specimen marks beginning of rapid molecular breakdown. Although MSI of formalin-fixed paraffin-embedded (**FFPE**) tissue is possible [24], use of freshly extracted tissue immediately frozen at -80 degree Celsius is more common [25]. In a recent study, Sugiura et al. presented a series of acetylcholine images from mouse sagittal brain sections [26]. Even with as short as 1 minute between organ extraction and freezing, the images showed significant postmortem acetylcholine degradation. To mitigate this problem, in-situ freezing (**ISF**) was performed wherein the head of a deeply anesthetized living specimen was dipped into liquid nitrogen to freeze the brain simultaneously with specimen sacrifice. Other sacrifice and tissue preparation protocols exist to mitigate degradation, and they have shown to be especially useful in neuropeptide imaging. For a further discussion on these methods, refer to section 3.4.

Next, the tissue is ready to be cut into thin sections on cryostat. The thickness of sectioned tissue should be on the order of the organism's cellular diameter, although other considerations such as sensitivity and ionization method of choice could also play a role. For example, if a particular tissue's cells have an average diameter of 10-20 µm, the slices should be no greater than 10-20 µm thick [25].

After tissue freezing and sectioning, sample preparation is largely dependent on the ionization source. **Scheme 4** provides a visual overview of the three most commonly used sources: secondary ion mass spectrometry (**SIMS**), matrix-assisted laser desorption/ionization (**MALDI**), and desorption electrospray ionization (**DESI**). SIMS is the oldest of the three and was first used for imaging in 1962 by Castaing and Slodzian [27]. The mechanism of ionization is beyond the scope of this review,[28] but essentially ions are formed when a focused primary ion beam hits sample and sputters off ionized tissue compounds into the mass spectrometer. SIMS is often limited to analyzing molecules under 1000 Da without significant sensitivity loss or unwanted fragmentation.[9] However, the different types of ion beams can be used to increase the intact ion yield of larger or more labile compounds.[16, 29-30] SIMS imaging does not require special preparation after sectioning and mounting the tissue, but there are some optional methods that improve the imaging results. Coating the tissue with gold, silver, small organic acids, or the use of gold nanoparticle as substrates on the surface of the sample holder or inserted into the sample can improve the ionization of intact molecules larger than 1000 Da.[31-37]

SIMS is able to reliably achieve less than 1 μm spatial resolution,[38] making it the method of choice for subcellular pharmaceutical investigations. The high lateral spatial resolution and utility SIMS can provide is seen in **Scheme 3 (III).** If such high spatial resolution is not required, MALDI MSI is a versatile alternative that has been employed for imaging drugs and metabolites,[39] lipids,[40] peptides,[41] and proteins.[42] In MALDI MSI, the tissue is coated with a thin layer of matrix and irradiated with a laser beam. The matrix absorbs much of the energy from the incident laser and provides a very “soft” ionization for analyte compounds.[43] Although MALDI is most commonly performed at high vacuum

pressures, it can also be utilized at atmospheric pressure (**AP-MALDI**) for an even softer process.[44] The mechanism for ion formation is still an active area of study,[45-46] but beyond the scope of this review.

Matrix selection and application is a crucial step for MALDI MSI. It has a great effect on sensitivity, spatial resolution, and selective analyte ionization.[25] Matrices must absorb light at the laser wavelength and must not react with tissue-bound analyte. Imaging with an infrared laser (**IR-MALDI-MSI**) allows one to use water as a matrix due to high absorptivity in the mid-IR range.[47] Water can also be added for IR laser ablation by freezing the tissue to form a thin layer of frost, as shown by Muddiman and coworkers.[48] More commonly, however, an ultraviolet laser is used (**UV-MALDI-MSI**) with small, conjugated organic acid matrices. For low-molecular weight pharmaceuticals, CHCA is generally considered the matrix of choice, though far from exclusive use.[25] For example, Jackson et al. showed the utility of a gold nanoparticle matrix to cationize neutral cerebrosides and improve their detection among abundant, positively charged phosphatidylcholines.[49] In a targeted MSI experiment, different matrices can be tested on a standard of the target analyte to determine the optimal choice of matrix.

Generally, matrix application methods are optimized to produce a homogenous coating of small crystals. Inhomogeneities can create local ionization biases that may cause extensive signal suppression.[50-51] Solvent-based methods prepare matrix dissolved in solution at or near the point of saturation followed by tissue deposition via nebulizers and airbrushes[52] or automated systems.[53-55] Solvent-free methods were developed to circumvent spatial delocalization of soluble analytes from excessive amounts of solvent during matrix application,[13] and a myriad of preparation protocols have

been developed.[56-59] Solvent-free methods work well for most molecules, but have a noticeable sensitivity drop-off for larger analytes.[57]

Until 2006, SIMS and MALDI were the two main pathways for MSI. It was not long after Takats et al. introduced DESI[60] that its imaging capabilities were realized.[11] DESI is a form of ambient ionization method where ions are analyzed from their native matrices with little or no sample preparation.[61] In a DESI ion source, solvent is sprayed through a high-voltage needle at AP and directly onto a sample. Molecules from the sample are ionized and desorbed into a heated capillary leading to a mass spectrometer. Ions are formed from either an electrospray ionization (**ESI**) mechanism[62] or heterogeneous charge transfer.[63-64] DESI imaging is mostly used for lipids[11] and small molecules.[65]

There is virtually no sample preparation needed for DESI aside from sectioning and mounting the tissue. Additives can be added to the solvent spray to increase selectivity and sensitivity for certain analytes, so-called reactive DESI,[66] and has been used by Wu et al. to image cholesterol.[67] Cholesterol is difficult to ionize because of its low proton affinity and low acidity, but the addition of betaine aldehyde to the DESI spray derivatized cholesterol's OH-group with a permanently charged hemiacetal, thus promoting ionization and detection.

Ionization and sample preparation methods should be selected based on what compound classes and tissue locations are intended for the image. SIMS is best used for ionizing small molecules and sensitive enough to spatially characterize sub-cellular amounts of analyte. However, maintaining such spatial resolution with compounds larger than 1000 Da is difficult. Additionally, commercially available SIMS instruments do not have MS/MS capabilities as MALDI or DESI instruments equipped with hybrid TOF or

quadrupole-based mass analyzers, thereby limiting SIMS-SRM-MSI. SIMS has been shown to make use of post-source decay (**PSD**) for SRM-MSI with the appropriate analysis software.[68] MALDI is capable of imaging molecules small and large, but often the spatial resolution is limited to 10-30 µm. Ambient DESI is a viable option for high-throughput analysis of pharmaceuticals with minimal tissue adulteration. DESI commonly produces spatial resolutions of 100 µm or greater, but a resolution of 35 µm was reported in a recent study.[69]

Mass Spectrometry Imaging Applications

Qualitative MSI

In the initial stages of modern MSI, the focus was largely on characterizing spatial distribution. A study by Stoeckli et al. 2007 on β-peptides gives a basic demonstration of its unique qualitative and semi-quantitative capabilities.[70] In this study, mice were intravenously dosed with peptide solutions, three with β-peptides and a fourth with an α-peptide. The β-peptide dosed mice were sequentially sacrificed after periods of 5 minutes, 1 hour, and 24 hours before being frozen and sectioned. **Figure 1** shows the MALDI-MSI of each mouse's β-peptide spatial distribution. These images can provide some qualitative information about ADME. The peptide is clearly absorbed and distributed throughout several organs, and it stays intact until it reaches the kidney where it is excreted. Additionally, the presence of β-peptide signal after 24 hours provides kinetic information, and the absence of the α-peptide signal after 1 hour indicates the greater metabolic stability of β-peptides, as expected. Of course, Stoeckli stated that similar results, with the added benefit of absolute quantitation, could have been achieved

by QWBA. However, MSI provided a quicker image acquisition without the need of costly radioactive labels.

Qualitative MSI still provides great utility. It can give relative pharmacokinetic (**PK**) and pharmacodynamic (**PD**) information, as well as allow the identification of novel metabolites that could help elucidate metabolic mechanisms. Enthaler et al. recently employed MALDI-MSI in a sophisticated method to probe *ex vivo* compound penetration in human and porcine skin.[71] Although the model compound was a cosmetic, Nile Red, the exact same protocols could be applied to cutaneous drug delivery. A problem with MALDI-MSI of skin samples was adhesion of skin to ITO-coated glass slides. The group attempted to improve adhesion by pre-treating the slides with coronal discharge and modify its surface to have stronger interactions with the proteins in skin. This method allowed them to image skin samples after a Nile Red treatment and a common penetration enhancer, dimethyl sulfoxide, for 24 hours. The 30 µm spatially resolved images showed that Nile Red penetrated past the upper stratum corneum layer and into the epidermis, but did not reach the dermis. Additionally, endogenous cholesterol sulfate, a molecule that regulates protective barrier formation in skin, was also imaged and showed distributions mainly in the epidermis. This example presents a remarkable potential for MSI to not only characterize a drug's delivery and penetration, but to also identify endogenous molecular distributions that may explain the observed absorption properties.

Although QWBA enables a similar quantitative analysis of radio-labeled Nile Red penetration, MARG would be required to achieve a spatial resolution at 30 µm and limited quantitation can be achieved. Nile Red is amenable to characterization by fluorescence microscopy and Enthaler even used the technique to validate the MSI. However,

fluorescent microscopy would not be a label-free alternative for molecules that do not fluoresce.

The utility of pharmaceutical MSI goes beyond direct imaging of tissue. Kreye et al. used MSI to probe the dynamics of controlled theophylline release from lipid implants.[72] Though no actual tissue was imaged, the 4 mm-long cylindrical implants were embedded in gelatin, frozen, cryosectioned, and then sprayed with matrix, nearly analogous to the preparation of tissue. Images of the initial theophylline distribution throughout the implant were created using radial and longitudinal cross-sections. Both views showed macroscopic homogeneity with steep gradients at the micrometer level, to which Kreye attributed to the initial particles of lipid and drug the implants were fabricated from. Very useful information about the mechanism of release was also obtained from radial cross-section images of implants after 0, 3, and 14-day exposures to release media. It is clearly shown that the release is not homogeneous, but starts from the edges of the implant.

Stoeckli, Enthaler, and Kreye performed MSI studies that were able to obtain important information, although they could have been similarly acquired with QWBA. In contrast, Drexler et al. used MSI where QWBA would be less suited for the study of phototoxicity.[73] Phototoxicity assays must identify photoreactivity of a target molecule. Radiochemical methods cannot distinguish a drug from its metabolite and therefore are less applicable to such assays. Drexler used QWBA to first determine if an orally administered proprietary drug would accumulate in tissues in the eye. The radiograms showed a significant presence of the drug in retina where it could possibly react with incident ultraviolet and visible light. MSI of eye tissue indicated that the parent drug was

present but could not identify any photochemical derivatizations, therefore providing a negative result for phototoxicity.

Drexler's phototoxicity assay via MSI may present a clear advantage over QWBA in this particular case, although QWBA offers desired quantitative imaging capability. The next section will discuss quantitative MSI, one of the most innovative and fast-moving area in the field.

Quantitative MSI

The heterogeneous nature of ionization makes accurate quantitation in MSI challenging. The intensity of an MS signal is not only related to an analyte's concentration, but to its ionization efficiency and environmental extractability as well. Heeren et al. discussed this issue at length in a critical insight article.[74] A simple MSI experiment was presented wherein three aqueous protein digests; ubiquitin, cytochrome C, and bovine serum albumin; were spotted and mixed at equal concentrations. The spots were imaged by MALDI-MSI. Spots containing only ubiquitin and cytochrome C in the mixture were able to display peptides from both proteins, but whenever the mixture contained albumin, only the albumin peptides were detected. As evident in this example, certain molecules are preferentially ionized and can "steal" signal from co-localized species.

The complex tissue environment could make quantitation difficult. Luxembourg et al. showed that variations in tissue salt concentration and histological features will create heterogeneous matrix crystallization.[51] The variations in crystals may cause different degrees of tissue desorption and recovery of the analyte, and they may also result in different degrees of fragmentation during desorption and ionization. The end result is a

signal that is more reflective of the tissue histology and microenvironment than of the analyte concentration.

MSI will require accurate means of quantification if it is ever to become an acceptable alternative to QWBA. ADME characterization of drugs seeks to answer a question of mass balance, and simply imaging the relative locations cannot give a satisfactory answer. The absolute amount must be known throughout the distribution to provide true pharmacokinetic and pharmacodynamics data, and such knowledge can also identify unaccounted for drug that just may not show up in an image.

In 2010, Hattori et al. investigated the metabolism of ATP in ischemic penumbra mice models.[75] Because the study required quantitative imaging that could clearly distinguish ATP from its metabolites, they had to find a way to merge MSI with another quantitative method. The solution was to extract a contralateral brain slice and perform quantitative analysis with capillary electrophoresis (**CE**)-ESI-MS/MS. The following equation was used to calculate concentrations of molecules in the image:

$$C_i = \frac{I_i}{\bar{I}} C' \quad [75]$$

The tissue concentration of a specific area (C_i) is equal to ratio of the maximum intensity of that area (I_i) and the median maximum intensities from the contralateral hemisphere (I) multiplied by the whole tissue concentration (C') determined by CE-ESI-MS/MS. Hattori et al. validated their method by comparison to previous spatial metabolic results from a radiochemical method.[76] Although Hattori's approach proved powerful and accurate, the difficulties of coupling CE to ESI-MS/MS lead others to an analogous liquid chromatography (**LC**) approach developed by Koeniger et al.[77]

Under certain circumstances, quantitative MSI can be performed by creating an on-tissue calibration curve. **Figure 2** from Nilsson et al. shows the distribution of tiotropium in rat lung tissue.[78] The adjacent tissue is a piece of control lung upon which known concentrations of tiotropium were spotted and imaged. If the tissue was fairly morphologically homogeneous, this could account for the universal response and extractability of the tiotropium. The calibration curves show the linearity of normalized intensities from the on-tissue calibration compared with LC-MS/MS quantitation of dosed lung tissue. The two curves are in strong agreement and provide a proof-of-principle for label-free on-tissue quantitative MSI.

The Nilsson study is impressive but only applicable to compounds that do not suffer from signal suppression at any place on a piece of tissue. The recent push in quantitative MSI is to add isotopically labeled internal standards (**IS**) to the tissue. Unlike radiochemical labels, stable isotopes such as ^{13}C, ^{15}N, ^{18}O, or ^{2}H are used. Pirman and Yost demonstrated the effectiveness of this method with the absolute quantification of endogenous acetyl-L-carnitine (AC) in pig brain.[79] A glass microscope slide was homogenously spotted with d_3-AC and sectioned pig brain was subsequently mounted on top. The image was made from the ratio of signature MS/MS fragment ions from the endogenous compound and the isotope-encoded internal standard. It is assumed that co-localized isotopologues will have identical ionization efficiencies, extractabilities, and MS/MS fragmentation behaviors, therefore allowing direct quantitative comparison between a target analyte's MS intensity in tissue sample and the heavy standard's MS intensity.

Vismeh et al. used DESI MSI with similar quantitative methods for absolute quantification of clozapine distributions.[80] Rats were dosed with clozapine before sacrifice and a deuterated standard was pipetted on top of the sectioned tissue. **Figure 3** shows the calibration curve of the clozapine ratios on tissue. In addition to the high degree of linearity in calibration curve, it is interesting to note the superior stability of the ratio of intensities compared to the stability of clozapine or IS alone.

The preliminary quantitative results from the last couple of years have created a lot of excitement in the field. There is a great potential that MSI will one day become a common tool for mass balance and ADME assays. Further refinement and biological application validation remain an active area of research.

MSI at and below the cellular scale

Molecular imaging at the cellular scale can offer critical details on the mechanisms, dynamics, and kinetics of drug delivery. The increased resolution brings the experiment from identifying drugs that interact with specific organs to drugs that interact with specific cell types. High-performance MSI can even elucidate specific organelles or cellular regions that are crucial to a drug's delivery. **Figure 4** shows SIMS images of a single cultured adipocyte at 33 nm spatial resolution.[81] Cells were incubated with ^{13}C-labeled oleic acid to investigate how free fatty acid was utilized and metabolized. Localization of the oleic acid will result in an increase of ^{13}C relative to ^{12}C, and so an image of the ^{13}C / ^{12}C ratio will identify all areas where it has been integrated. The first image shows that a significant proportion of oleic acid localized in the lipid-rich cell membrane. A second image at a different depth of the same cell shows a very high concentration of oleic acid as part of a lipid droplet inside of the cell.

Sensitivity becomes increasingly important in sub-cellular MSI. With less total material in each sample point, fewer ions will be created. Ion microscopy is a technique that aids sensitivity by using detectors that are sensitive to an ion's 2-D position.[82] As such, the spatial resolution becomes independent of sampling size. Chandra et al. used SIMS microscopy to make quantitative images of neutron-capture therapy drug delivery in cultured human glioblastoma cells.[83] ^{10}B-labeled *p*-boronophenylalanine-fructose (BPA-F) and ^{11}B-labeled sodium borocaptate (BSH) were introduced into the culture medium in two separate experiments, one with separate introduction of each drug and one with simultaneous introduction. The overall goal of the experiment was to determine whether there was a synergistic effect on the delivery of either drug during co-administration. The boron drugs seemed to homogenously penetrate all parts of the cell except the perinuclear cytoplasm, and the authors concluded that BSH and BPA-F show only additive boron delivery when co-administered to the cell media.

Altelaar et al. published a protocol paper that describes two approaches for MALDI MSI of cellular dimensions.[84] Ion microscopy is one such approach, with the other approach being the scanning microprobe MALDI (**SMALDI**). Altelaar demonstrates that both methods can attain high-resolution images of whole rat brain tissue, but SMALDI may require unfavorably high laser fluence. SMALDI involves using very small laser beam diameters and distances between sampling points, about 0.7 µm and 0.25 µm respectively.[85] Astonishingly, even though SMALDI ablates such little material into the mass spectrometer, peptide standards have been detected at attomole and zeptomole levels in a single 1 µm x 1 µm pixel. An investigation by Bouschen et al. used SMALDI

on human carcinoma cells and peptide mixtures to produce images with an effective spatial resolution of 2 µm.[86]

Very recently, the use of alternative laser ablation geometry has demonstrated the ability for subcellular MALDI imaging without a microprobe or ion microscopy. As shown in **Scheme 4C**, the incident MALDI irradiation commonly strikes the front side of the sample holder in what is known as reflection geometry (**RG**). If the sample holder is made of transparent material that will not absorb the laser wavelength, the incident radiation can instead strike from the back at 180° angle from the sample in what is known as transmission geometry (**TG**). TG-MSI was first used by Richards et al. to image sulfatide lipids[87] and fragile gangliosides.[88] Although the images were acquired at 10 – 20 µm spatial resolution, single shot ablation in TG allowed images of the entire mouse brain to be created in less than an hour, a substantial reduction of imaging time compared to the use of conventional RG mode. Zavalin et al. published a proof-of-principle experiment for TG-MALDI-MSI of single cells.[89] Lipids are shown localized in the cell membranes of HEK-293 cells in images that were acquired with a 1 µm laser diameter and 1.5 µm center-to-center sampling points. Although no exogenous compounds were imaged at subcellular resolution, there is no reason to suspect it would not be possible.

Neuropeptide Imaging: Mapping Endogenous Signaling Molecules and Drug Delivery

Organisms use neuropeptides and other signaling peptides to regulate a great variety of physiological processes. These endogenous molecules can be templates for synthetic drug development and delivery, and peptide-based pharmaceuticals have been of great interest for decades. Neuropeptides and structural analogs can also be of use

for drug discovery and delivery as there is great interest in peptide-based pharmaceuticals. Neuropeptides and other signaling peptides are commonly studied by MS with workflows that often include MSI at the whole tissue scale all the way down to single neurons.[90-92] But even at the tissue level, high spatial resolution is important. Tissues rich in neuropeptides, such as the pituitary gland, can have dimensions of just a few millimeters. Geunther et al. obtained astonishing AP-MALDI-MS images of neuropeptides in a mouse pituitary.[93] Their results suggest that new MALDI-MSI technology is leading the method to a mature level that can probe very specific tissue-peptide interactions. **Figure 5** shows an optical image of a pituitary gland followed by molecular images of four peptides. The pituitary gland was measured at dimensions of 3 mm by 1 mm. With 5 µm spatial resolution, oxytocin and vasopressin were primarily localized in the posterior lobe while a joining peptide and γ-MSH were localized in the intermediate lobe.

In a study performed by Hui et al., a neuropeptide workflow included MALDI-MSI localization to aid the functional study of a novel tachykinin neuropeptide.[94] Brains from *Callinectes sapidus* were imaged to show the distributions and expression levels of two peptides, CalsTRP and CabTRP Ia, between fed and unfed animals. Both peptides showed consistent upregulation and co-localization in fed animals, with CabTRP Ia at a higher intensity than CalsTRP. The combined evidence of the co-regulation, co-localization, and nearly identical amino acid sequences led the investigators to posit that both neuropeptides may originate from the same preprotachykinin. Additionally, the higher intensity of CabTRP Ia may indicate which sequence from the preprotachykinin is

preferentially expressed, although the authors concede it may also be a result of CalsTRP having a higher post-mortem turnover.

Post-mortem protease activity is a significant problem for neuropeptide imaging. In fact, early MS neuropeptide studies suggested that more than 90% of detected peptides were non-active fragments from protein degradation.[95] On-tissue protease deactivation has been successfully achieved by raising the tissue temperature to a level that deactivates most enzymes. Microwave irradiation, either applied as a focused beam for animal sacrifice[96] or to tissues post-sacrifice in a microwave oven,[97] has dramatically decreased the detected post-mortem levels of neuropeptides and small molecules such as cyclic-AMP and arachidonic acid. Alternatively, the commercially available Denator AB that uses a combination of pressure and heat to quickly denature enzymes immediately following extraction.[98] MALDI-MSI of heat-treated tissues by Goodwin et al. shows that the Denator heating at 95℃ significantly reduced protein turnover.[99] However, the stabilization process did cause detrimental damage to tissue morphology which would complicate MSI. The authors conclude that heat-treated tissue in MSI would be best utilized as complementary sample source for MSI from intact structures of non-treated tissue. Microwave irradiation may also cause inconsistencies detrimental to imaging such as the differential heating and enzyme inactivation of certain areas in a tissue.[96]

Emerging Technology, Perspectives, and Conclusions

After nearly a decade and a half of extensive research efforts, MSI is finally approaching a stage of more widespread applications. Qualitative molecular imaging is now a rather straight-forward process, and even more novel approaches are being developed, such as three-dimensional MSI that simultaneously maps lateral distribution

and depth penetration within tissues.[100-101] However, MSI has not reached universal acceptance in pharmaceutical science due to remaining challenges that must be overcome.

The importance of robust and accurate quantitation for drug and metabolite MSI cannot be underemphasized. The proof-of-principle quantitative MSI has certainly been demonstrated. It is possible under some circumstances to detect and localize drugs with similar sensitivity to QWBA, and may even surpass radiochemical quantitation at cellular and subcellular scales. However, QWBA excels in its reproducibility and robustness. If a radio-labeled compound is present, it will produce a signal. MSI compounds, on the other hand, may be present but undetectable due to matrix ionization suppression, extractability, or analyte degradation. Using a stable isotope-labeled IS can account for suppression and extractability and will likely be well-adopted in future quantitative MSI. More methods to improve reproducibility may also emerge, and biggest factor may just be giving MSI time to be employed in novel biological applications.

The sheer complexity of biological tissues create challenges for MSI. Front-end separation would have the biggest impact on dynamic range, and some recently developed liquid extraction (**LE**) methods have made it possible to separate compounds from tissue microsections using LC.[102-103] Small liquid junctions can make contact with the surface of the tissue and collect local molecules to be separated by LC, effectively decoupling analyte extraction from ionization. Although some LE methods have shown spatial resolution near 30 µm,[104] most of them produce localization which spatial resolution well in excess of 100 µm.

A possibility that will not affect ionization suppression but that can deconvolute MSI spectra is to employ ion mobility (**IM**) separation, which separates ions in the gas phase based on shape and size.[105] IM-MSI can differentiate structural metabolite isomers[106] and separate isobaric molecular classes like lipids and peptides prior to MS analysis.[49, 59, 107] On-tissue derivatization techniques can help to bring compound signals out of the tissue background noise, as shown by Manier et al. when imaging isoniazid, an anti-tuberculosis drug.[108] The SRM transitions displayed by a neat sample of isoniazid were observed from isobaric compounds in control tissue. However, when tissue-bound isoniazid reacted with glass slides precoated with *trans*-cinnamaldehyde, the resulting drug derivative had SRM transitions unique to dosed tissue.

With a single MS image containing 5,000 to 50,000 spectra, statistical analyses of MSI presents its own complex challenges.[109] Pixel-to-pixel variation, an unavoidable MSI phenomenon, is chief among them. Pre-processing methods aimed at de-noising spectra or clustering pixels can help mitigate the variations, but a sound physical model of the origin of variation would certainly lead to more effective algorithms.[110] Additionally, Thiele et al. have demonstrated that performing *in vivo* MRI before MSI can help match features in the MS image to anatomical structures in the tissue, especially in the construction of a 3-D MS image.[111]

McDonnell et al. published a very insightful review that addresses perhaps the most important issue regarding the spread of MSI technology.[112] The overall reproducibility of MSI can be greatly improved by increasing the accessibility of the method and creating well-defined standardized practices. As stated by McDonnell, participants at a 2009 MSI training course revealed that matrix deposition protocols varied

greatly even when the same commercial deposition device was being used. European countries have since established an MSI network, COST Action BM1104 that aims to create best-practice guidelines and method standards, as well as provide resources and hands-on training. If other communities are quick to follow this example, the time before universal acceptance of MSI by pharmaceutical science will be drastically reduced.

Acknowledgements

Preparation of this manuscript was supported in part by National Science Foundation (CHE- 0957784) and National Institutes of Health through grant 1R01DK071801. C.L. acknowledges an NIH-supported Chemistry Biology Interface Training Program Predoctoral Fellowship (grant number T32-GM008505). L.L. acknowledges an H. I. Romnes Faculty Fellowship.

References

(1) Solon, E. G. *Chem. Res. Toxicol.* **2012** *25*, 543-555.

(2) Solon, E. G., Schweitzer, A., Stoeckli, M., and Prideaux, B. *AAPS. J.* **2010** *12*, 11-26.

(3) Roffey, S. J., Obach, R. S., Gedge, J. I., and Smith, D. A. *Drug Metab. Rev.* **2007** *39*, 17-43.

(4) Greer, T., Sturm, R., and Li, L. J. *J. Proteomics* **2011** *74*, 2617-2631.

(5) Seeley, E. H., and Caprioli, R. M. *Trends Biotechnol.* **2011** *29*, 136-143.

(6) Chaurand, P. *J. Proteomics* **2012**.

(7) McDonnell, L. A., and Heeren, R. M. *Mass Spectrom. Rev.* **2007** *26*, 606-643.

(8) Hsieh, Y., Chen, J., and Korfmacher, W. A. *J. Pharmacol. Toxicol. Meth.* **2007** *55*, 193-200.

(9) van Hove, E. R. A., Smith, D. F., and Heeren, R. M. A. *J. Chromatogr. A* **2010** *1217*, 3946-3954.

(10) Castellino, S., Groseclose, M. R., and Wagner, D. *Bioanalysis* **2011** *3*, 2427-2441.

(11) Wiseman, J. M., Ifa, D. R., Song, Q. Y., and Cooks, R. G. *Angew. Chem. Int. Ed.* **2006** *45*, 7188-7192.

(12) Seeley, E. H., and Caprioli, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **2008** *105*, 18126-18131.

(13) Troendle, F. J., Reddick, C. D., and Yost, R. A. *J. Am. Soc. Mass. Spectrom.* **1999** *10*, 1315-1321.

(14) Rohner, T. C., Staab, D., and Stoeckli, M. *Mech. Ageing Dev.* **2005** *126*, 177-185.

(15) Gross, J. H. (2004) *Mass Spectrometry*, Springer, Berlin.

(16) Smith, D. F., Robinson, E. W., Tolmachev, A. V., Heeren, R. M., and Pasa-Tolic, L. *Anal. Chem.* **2011** *83*, 9552-9556.

(17) Manicke, N. E., Dill, A. L., Ifa, D. R., and Cooks, R. G. *J. Mass Spectrom.* **2010** *45*, 223-226.

(18) Clode, P. L., Kilburn, M. R., Jones, D. L., Stockdale, E. A., Cliff, J. B., Herrmann, A. M. *et al. Plant Physiol.* **2009** *151*, 1751-1757.

(19) Marshall, A. G., Hendrickson, C. L., and Jackson, G. S. *Mass Spectrom. Rev.* **1998** *17*, 1-35.

(20) Hu, Q. Z., Noll, R. J., Li, H. Y., Makarov, A., Hardman, M., and Cooks, R. G. *J. Mass Spectrom.* **2005** *40*, 430-443.

(21) Glauser, G., Veyrat, N., Rochat, B., Wolfender, J. L., and Turlings, T. C. *J. Chromatogr. A* **2012**.

(22) Hopfgartner, G., Varesio, E., and Stoeckli, M. *Rapid Commun. Mass Spectrom.* **2009** *23*, 733-736.

(23) Stoeckli, M., Staab, D., and Schweitzer, A. *Int. J. Mass. Spectrom.* **2007** *260*, 195-202.

(24) Casadonte, R., and Caprioli, R. M. *Nat. Protoc.* **2011** *6*, 1695-1709.

(25) Schwartz, S. A., Reyzer, M. L., and Caprioli, R. M. *J. Mass Spectrom.* **2003** *38*, 699-708.

(26) Sugiura, Y., Zaima, N., Setou, M., Ito, S., and Yao, I. *Anal. Bioanal. Chem.* **2012** *403*, 1851-1861.

(27) Castaing, R., and Slodzian, G. *J. Microsc.* **1962** *1*, 395-410.

(28) Pachuta, S. J., and Cooks, R. G. *Chem. Rev.* **1987** *87*, 647-669.

(29) Touboul, D., Halgand, F., Brunelle, A., Kersting, R., Tallarek, E., Hagenhoff, B. *et al. Anal. Chem.* **2004** *76*, 1550-1559.

(30) Ostrowski, S. G., Szakal, C., Kozole, J., Roddy, T. P., Xu, J. Y., Ewing, A. G. *et al. Anal. Chem.* **2005** *77*, 6190-6196.

(31) Keune, K., and Boon, J. J. *Surf. Interface Anal.* **2004** *36*, 1620-1628.

(32) Nygren, H., Malmberg, P., Kriegeskotte, C., and Arlinghaus, H. F. *FEBS Lett.* **2004** *566*, 291-293.

(33) Kim, Y. P., Oh, E., Shon, H. K., Moon, D. W., Lee, T. G., and Kim, H. S. *Appl. Surf. Sci.* **2008** *255*, 1064-1067.

(34) Delcorte, A., Bour, J., Aubriet, F., Muller, J. F., and Bertrand, P. *Anal. Chem.* **2003** *75*, 6875-6885.

(35) Wu, K. J., and Odom, R. W. *Anal. Chem.* **1996** *68*, 873-882.

(36) Szymczak, W., and Wittmaack, K. *Rapid Commun. Mass Spectrom.* **2002** *16*, 2025-2033.

(37) Adriaensen, L., Vangaever, F., Lenaerts, J., and Gijbels, R. *Rapid Commun. Mass Spectrom.* **2005** *19*, 1017-1024.

(38) (2001) *ToF SIMS: Surface analysis by mass spectrometry*, IM Publications and Surface Spectra Limited, Chichester, Manchester.

(39) Shanta, S. R., Kim, Y., Kim, Y. H., and Kim, K. P. *Biomol. Ther.* **2011** *19*, 149-154.

(40) Goto-Inoue, N., Hayasaka, T., Zaima, N., and Setou, M. *BBA-Mol. Cell. Biol. L.* **2011** *1811*, 961-969.

(41) Chen, R. B., and Li, L. J. *Anal. Bioanal. Chem.* **2010** *397*, 3185-3193.

(42) Chaurand, P., Norris, J. L., Cornett, D. S., Mobley, J. A., and Caprioli, R. M. *J. Proteome Res.* **2006** *5*, 2889-2900.

(43) Tanaka, K., Waki, H., Ido, Y., Akita, S., Yoshida, Y., Yoshida, T. *et al. Rapid Commun. Mass Spectrom.* **1988** *2*, 151-153.

(44) Laiko, V. V., Baldwin, M. A., and Burlingame, A. L. *Anal. Chem.* **2000** *72*, 652-657.

(45) Karas, M., and Kruger, R. *Chem. Rev.* **2003** *103*, 427-439.

(46) Knochenmuss, R., and Zenobi, R. *Chem. Rev.* **2003** *103*, 441-452.

(47) Vertes, A., Nemes, P., Shrestha, B., Barton, A. A., Chen, Z. Y., and Li, Y. *Appl. Phys. a-Mater.* **2008** *93*, 885-891.

(48) Robichaud, G., Barry, J. A., Garrard, K. P., and Muddiman, D. C. *J. Am. Soc. Mass. Spectrom.* **2012**.

(49) Jackson, S. N., Ugarov, M., Egan, T., Post, J. D., Langlais, D., Albert Schultz, J. *et al. J. Mass Spectrom.* **2007** *42*, 1093-1098.

(50) Knochenmuss, R., Karbach, V., Wiesli, U., Breuker, K., and Zenobi, R. *Rapid Commun. Mass Spectrom.* **1998** *12*, 529-534.

(51) Luxembourg, S. L., McDonnell, L. A., Duursma, M. C., Guo, X. H., and Heeren, R. M. A. *Anal. Chem.* **2003** *75*, 2333-2341.

(52) Wang, H. Y. J., Post, S. N. J. J., and Woods, A. S. *Int. J. Mass Spectrom.* **2008** *278*, 143-149.

(53) Baluya, D. L., Garrett, T. J., and Yost, R. A. *Anal. Chem.* **2007** *79*, 6862-6867.

(54) Yamada, Y., Hidefumi, K., Shion, H., Oshikata, M., and Haramaki, Y. *Rapid Commun. Mass Spectrom.* **2011** *25*, 1600-1608.

(55) Schuerenberg, M., Luebbert, C., Deininger, S. O., Ketterlinus, R., and Suckau, D. *Nat. Meth.* **2007** *4*, iii-iv.

(56) Grove, K. J., Frappier, S. L., and Caprioli, R. M. *J. Am. Soc. Mass Spectrom.* **2011** *22*, 192-195.

(57) Hankin, J. A., Barkley, R. M., and Murphy, R. C. *J. Am. Soc. Mass Spectrom.* **2007** *18*, 1646-1652.

(58) Puolitaival, S. M., Burnum, K. E., Cornett, D. S., and Caprioli, R. M. *J. Am. Soc. Mass Spectrom.* **2008** *19*, 882-886.

(59) Trimpin, S., Herath, T. N., Inutan, E. D., Wager-Miller, J., Kowalski, P., Claude, E. *et al. Anal. Chem.* **2010** *82*, 359-367.

(60) Takats, Z., Wiseman, J. M., Gologan, B., and Cooks, R. G. *Science* **2004** *306*, 471-473.

(61) Alberici, R. M., Simas, R. C., Sanvido, G. B., Romao, W., Lalli, P. M., Benassi, M. *et al. Anal. Bioanal. Chem.* **2010** *398*, 265-294.

(62) Dole, M., Mack, L. L., and Hines, R. L. *J. Chem. Phys.* **1968** *49*, 2240-&.

(63) Cooks, R. G., Jo, S. C., and Green, J. *Appl. Surf. Sci.* **2004** *231*, 13-21.

(64) Takats, Z., Wiseman, J. M., and Cooks, R. G. *J. Mass. Spectrom.* **2005** *40*, 1261-1275.

(65) Wiseman, J. M., Ifa, D. R., Zhu, Y. X., Kissinger, C. B., Manicke, N. E., Kissinger, P. T. *et al. Proc. Natl. Acad. Sci. U.S.A.* **2008** *105*, 18120-18125.

(66) Cotte-Rodriguez, I., Takats, Z., Talaty, N., Chen, H. W., and Cooks, R. G. *Anal. Chem.* **2005** *77*, 6755-6764.

(67) Wu, C., Ifa, D. R., Manicke, N. E., and Cooks, R. G. *Anal. Chem.* **2009** *81*, 7618-7624.

(68) Touboul, D., Brunelle, A., and Laprevote, O. *Rapid Commun. Mass Spectrom.* **2006** *20*, 703-709.

(69) Campbell, D. I., Ferreira, C. R., Eberlin, L. S., and Cooks, R. G. *Anal. Bioanal. Chem.* **2012** *404*, 389-398.

(70) Stoeckli, M., Staab, D., Schweitzer, A., Gardiner, J., and Seebach, D. *J. Am. Soc. Mass. Spectrom.* **2007** *18*, 1921-1924.

(71) Enthaler, B., Pruns, J. K., Wessel, S., Rapp, C., Fischer, M., and Wittern, K. P. *Anal. Bioanal. Chem.* **2012** *402*, 1159-1167.

(72) Kreye, F., Hamm, G., Karrout, Y., Legouffe, R., Bonnel, D., Siepmann, F. *et al. J. Control. Release* **2012** *161*, 98-108.

(73) Drexler, D. M., Tannehill-Gregg, S. H., Wang, L. F., and Brock, B. J. *J. Pharmacol. Toxicol. Meth.* **2011** *63*, 205-208.

(74) Heeren, R. M., Smith, D. F., Stauber, J., Kukrer-Kaletas, B., and MacAleese, L. *J. Am. Soc. Mass Spectrom.* **2009** *20*, 1006-1014.

(75) Hattori, K., Kajimura, M., Hishiki, T., Nakanishi, T., Kubo, A., Nagahata, Y. *et al. Antioxid. Redox Sign.* **2010** *13*, 1157-1167.

(76) Sokoloff, L., Reivich, M., Kennedy, C., Des Rosiers, M. H., Patlak, C. S., Pettigrew, K. D. *et al. J. Neurochem.* **1977** *28*, 897-916.

(77) Koeniger, S. L., Talaty, N., Luo, Y. P., Ready, D., Voorbach, M., Seifert, T. *et al. Rapid Commun. Mass Spectrom.* **2011** *25*, 503-510.

(78) Nilsson, A., Fehniger, T. E., Gustavsson, L., Andersson, M., Kenne, K., Marko-Varga, G. *et al. Plos One* **2010** *5*.

(79) Pirman, D. A., and Yost, R. A. *Anal. Chem.* **2011** *83*, 8575-8581.

(80) Vismeh, R., Waldon, D. J., Teffera, Y., and Zhao, Z. Y. *Anal. Chem.* **2012** *84*, 5439-5445.

(81) Lechene, C., Hillion, F., McMahon, G., Benson, D., Kleinfeld, A. M., Kampf, J. P. *et al. J. Biol.* **2006** *5*, 20.

(82) Setou, M., Shrivas, K., Sroyraya, M., Yang, H., Sugiura, Y., Moribe, J. *et al. Med. Mol. Morphol.* **2010** *43*, 1-5.

(83) Chandra, S., Lorey, D. R., and Smith, D. R. *Radiat. Res.* **2002** *157*, 700-710.

(84) Altelaar, A. F. M., Luxembourg, S. L., McDonnell, L. A., Piersma, S. R., and Heeren, R. M. A. *Nat. Protoc.* **2007** *2*, 1185-1196.

(85) Spengler, B., and Hubert, M. *J. Am. Soc. Mass Spectrom.* **2002** *13*, 735-748.

(86) Bouschen, W., Schulz, O., Eikel, D., and Spengler, B. *Rapid Commun. Mass Spectrom.* **2010** *24*, 355-364.

(87) Richards, A. L., Lietz, C. B., Wager-Miller, J. B., Mackie, K., and Trimpin, S. *Rapid Commun. Mass Spectrom.* **2011** *25*, 815-820.

(88) Richards, A. L., Lietz, C. B., Wager-Miller, J., Mackie, K., and Trimpin, S. *J. Lipid Res.* **2012** *53*, 1390-1398.

(89) Zavalin, A., Todd, E. M., Rawhouser, P. D., Yang, J., Norris, J. L., and Caprioli, R. M. *J. Mass Spectrom.* **2012** *47*, 1473-1481.

(90) Li, L. J., and Sweedler, J. V. *Annu. Rev. Anal. Chem.* **2008** *1*, 451-483.

(91) Ye, H., Greer, T., and Li, L. *J. Proteomics* **2012**.

(92) Rubakhin, S. S., Greenough, W. T., and Sweedler, J. V. *Anal. Chem.* **2003** *75*, 5374-5380.

(93) Guenther, S., Rompp, A., Kummer, W., and Spengler, B. *Int. J. Mass Spectrom.* **2011** *305*, 228-237.

(94) Hui, L. M., Zhang, Y. Z., Wang, J. H., Cook, A., Ye, H., Nusbaum, M. P. *et al. ACS Chem. Neurosci.* **2011** *2*, 711-722.

(95) Skold, K., Svensson, M., Kaplan, A., Bjorkesten, L., Astrom, J., and Andren, P. E. *PROTEOMICS* **2002** *2*, 447-454.

(96) Galli, C., and Racagni, G. *Meth. Enzymol.* **1982** *86*, 635-642.

(97) Che, F. Y., Lim, J., Pan, H., Biswas, R., and Fricker, L. D. *Mol. Cell. Proteomics* **2005** *4*, 1391-1405.

(98) Colgrave, M. L., Xi, L., Lehnert, S. A., Flatscher-Bader, T., Wadensten, H., Nilsson, A. *et al. PROTEOMICS* **2011** *11*, 1264-1276.

(99) Goodwin, R. J., Lang, A. M., Allingham, H., Boren, M., and Pitt, A. R. *PROTEOMICS* **2010** *10*, 1751-1761.

(100) Ye, H., Greer, T., and Li, L. J. *Bioanalysis* **2011** *3*, 313-332.

(101) Chen, R. B., Hui, L. M., Sturm, R. M., and Li, L. J. *J. Am. Soc. Mass Spectrom.* **2009** *20*, 1068-1077.

(102) Blatherwick, E. Q., Van Berkel, G. J., Pickup, K., Johansson, M. K., Beaudoin, M. E., Cole, R. O. *et al. Xenobiotica* **2011** *41*, 720-734.

(103) Wang, B. X., Inutan, E. D., and Trimpin, S. *J. Am. Soc. Mass Spectrom.* **2012** *23*, 442-445.

(104) Laskin, J., Heath, B. S., Roach, P. J., Cazares, L., and Semmes, O. J. *Anal. Chem.* **2012** *84*, 141-148.

(105) Kiss, A., and Heeren, R. M. A. *Anal. Bioanal. Chem.* **2011** *399*, 2623-2634.

(106) Trim, P. J., Henson, C. M., Avery, J. L., McEwen, A., Snel, M. F., Claude, E. *et al. Anal. Chem.* **2008** *80*, 8628-8634.

(107) McLean, J. A., Ridenour, W. B., and Caprioli, R. M. *J. Mass Spectrom.* **2007** *42*, 1099-1105.

(108) Manier, M. L., Reyzer, M. L., Goh, A., Dartois, V., Via, L. E., Barry, C. E., 3rd *et al. J. Am. Soc. Mass Spectrom.* **2011** *22*, 1409-1419.

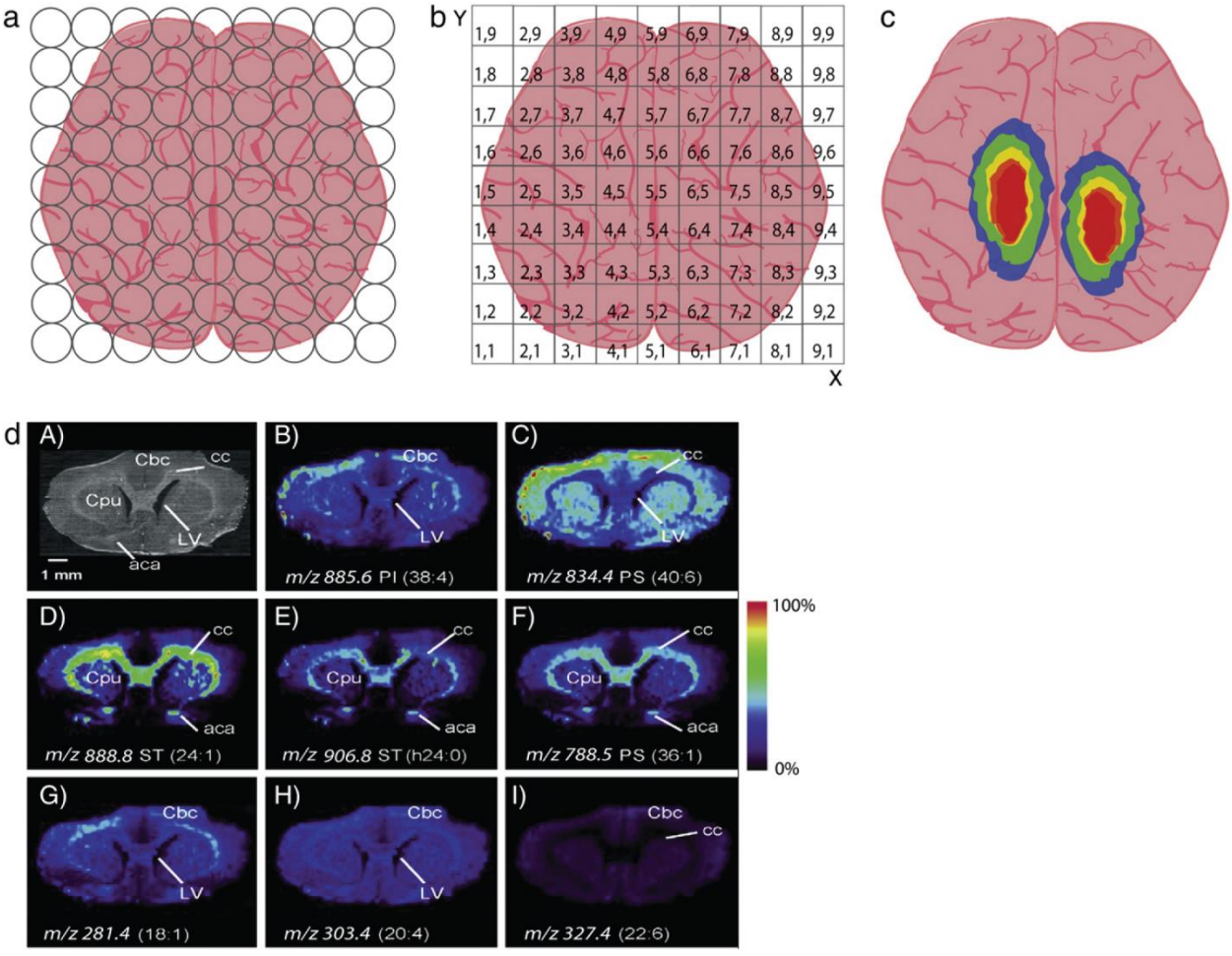
(109) Jones, E. A., Deininger, S. O., Hogendoorn, P. C., Deelder, A. M., and McDonnell, L. A. *J. Proteomics* **2012** *75*, 4962-4989.

(110) Alexandrov, T. *BMC Bioinformatics* **2012** *13 Suppl 16*, S11.

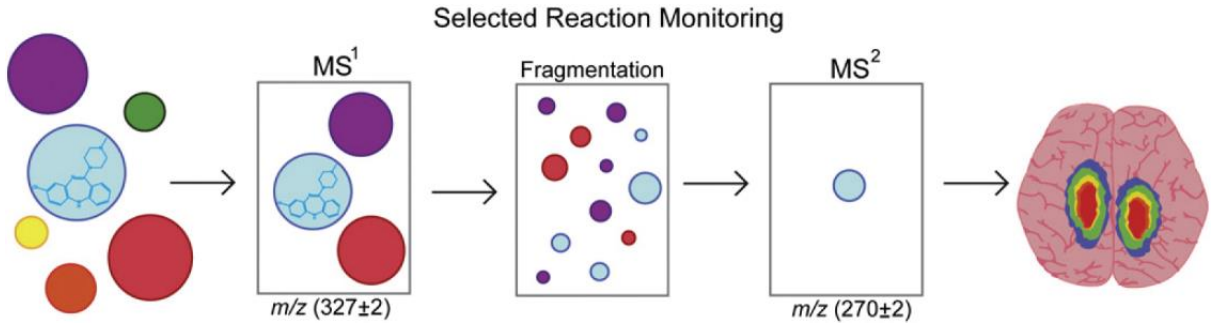
(111) Thiele, H., Heldmann, S., Trede, D., Strehlow, J., Wirtz, S., Dreher, W. *et al. Biochim. Biophys. Acta* **2013**.

(112) McDonnell, L. A., Heeren, R. M., Andren, P. E., Stoeckli, M., and Corthals, G. L. *J. Proteomics* **2012**.

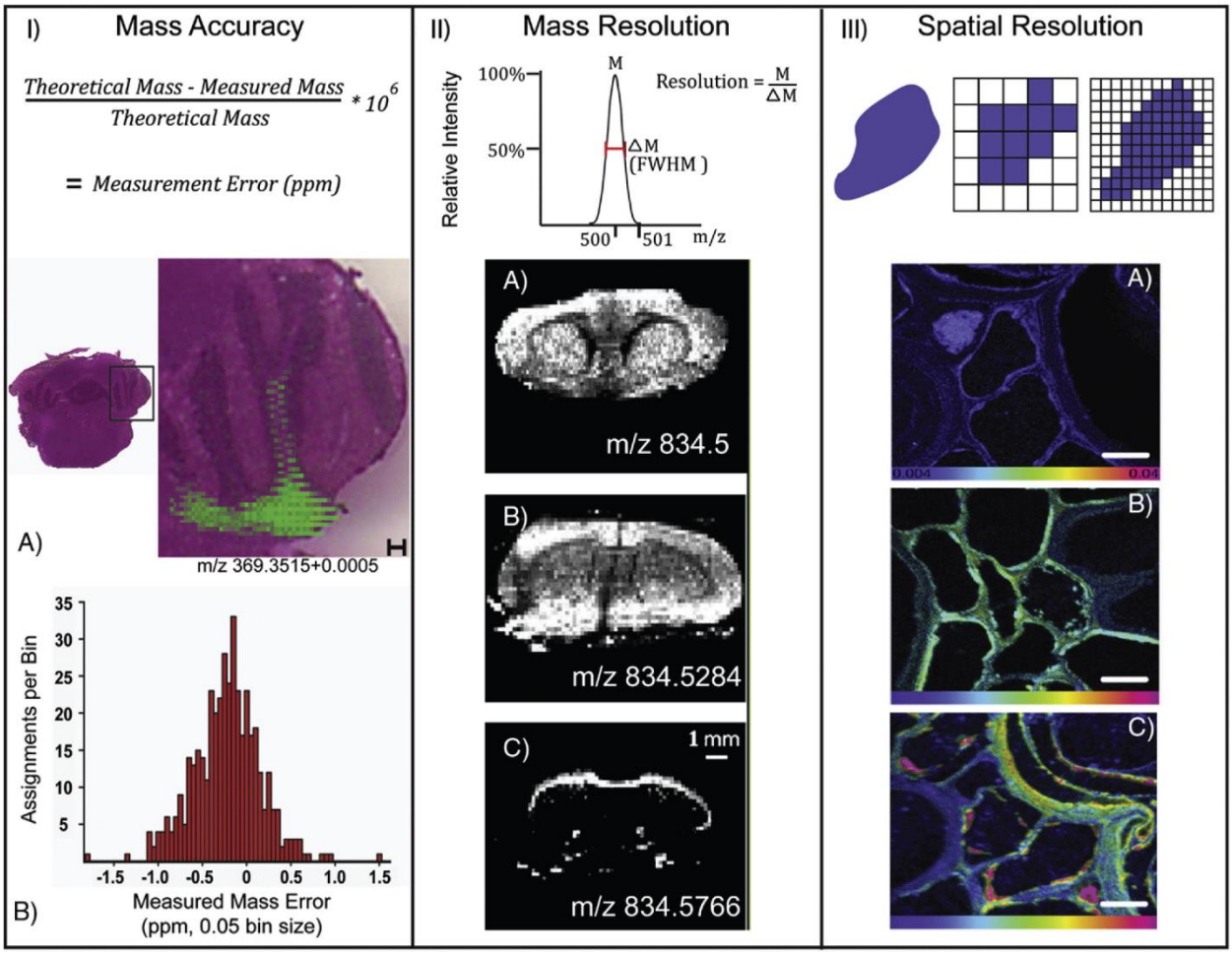
Scheme 1. Creating and imaging ions by mass spectrometry. (a) After tissue is prepared, mass spectra are collected at points across the entire tissue. **(b)** Each sampled point is converted into spatial coordinates and **(c)** an image is created by displaying the intensity of a specific ion at each point. **(a)** Real images of 8 different lipids acquired simultaneously from a rat brain, reprinted with permission.[11]



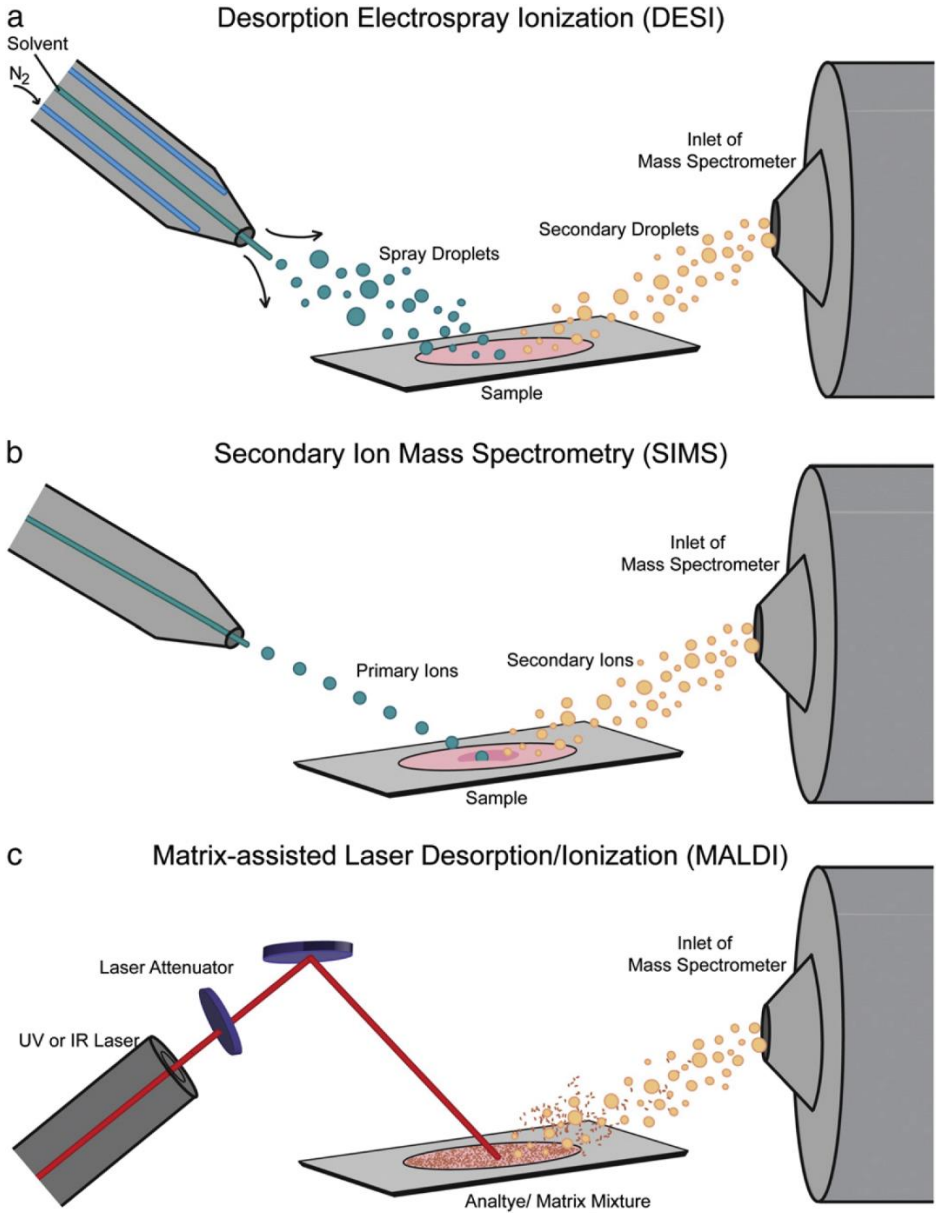
Scheme 2. Selected reaction monitoring. Clozapine (blue) and other compounds have been desorbed and ionized prior to mass analysis. In selected reaction monitoring (SRM), the first round of mass analysis removes all ions outside the mass window of 325 Da–329 Da that allows isolation of precursor ions for subsequent fragmentation. The remaining ions are then fragmented. For the second round of mass analysis, ions outside of the 268 Da–272 Da are removed, leaving only the signature fragment of clozapine (blue) to be imaged.



Scheme 3. MSI figures of merit. (I) An image of cholesterol at 1.4 ppm mass accuracy, reprinted with permission.[16] **(II)** Three lipid images requiring high mass resolution, reprinted with permission.[17] **(III)** A secondary ion mass spectrometry image of root cells with submicron spatial resolution, reprinted with permission.[18]



Scheme 4. MSI ionization sources. The three most common sources for MSI: **(a)** Desorption electrospray ionization which uses a stream of solvent ions to desorb and ionize analyte molecules, **(b)** secondary ion mass spectrometry which uses a beam of ions from an ion gun to sputter analyte ions off the sample, and **(c)** matrix-assisted laser desorption/ionization which uses laser irradiation of a matrix-coated sample for desorption and ionization.



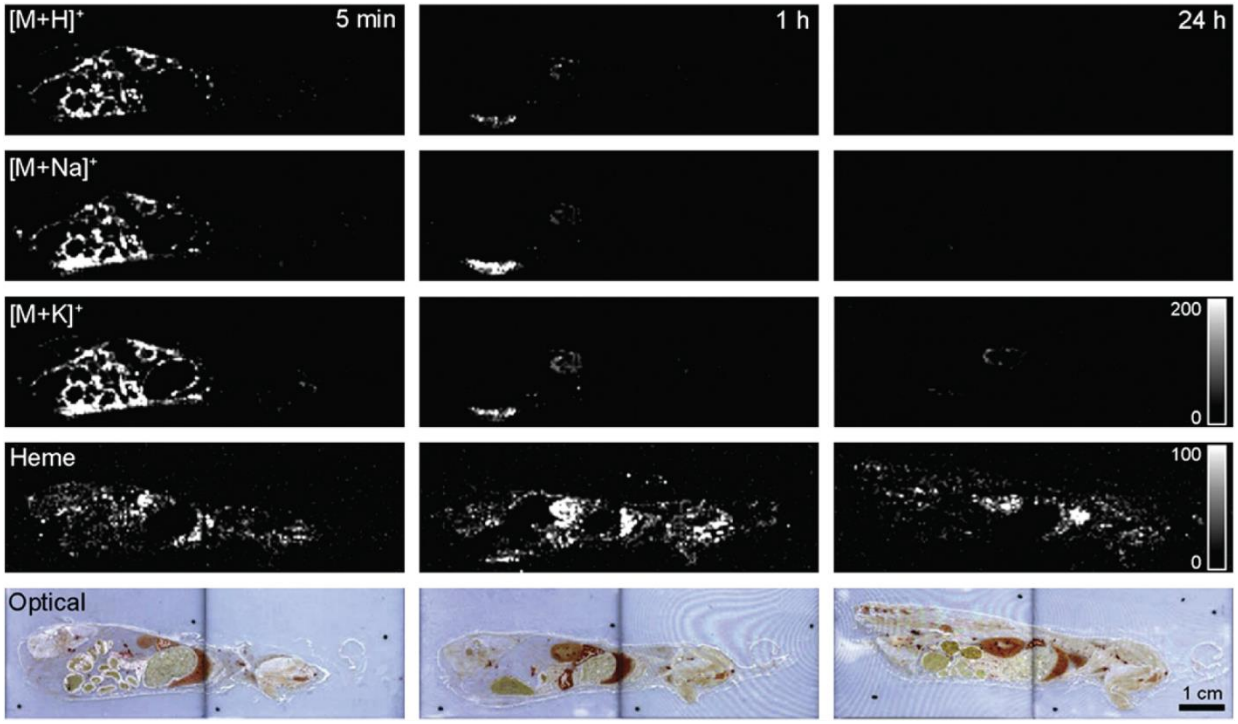


Figure 1. Whole-body MSI of drug metabolism. The distribution of a β-peptide in mice after 5 min, 1 h, and 24 h. The first three rows show the β-peptide distribution imaged from its three most common ion types. The fourth row shows the image of heme as a molecular reference, and the fifth row shows an optical image of the specimen for a physical reference. Reprinted with permission.[70]

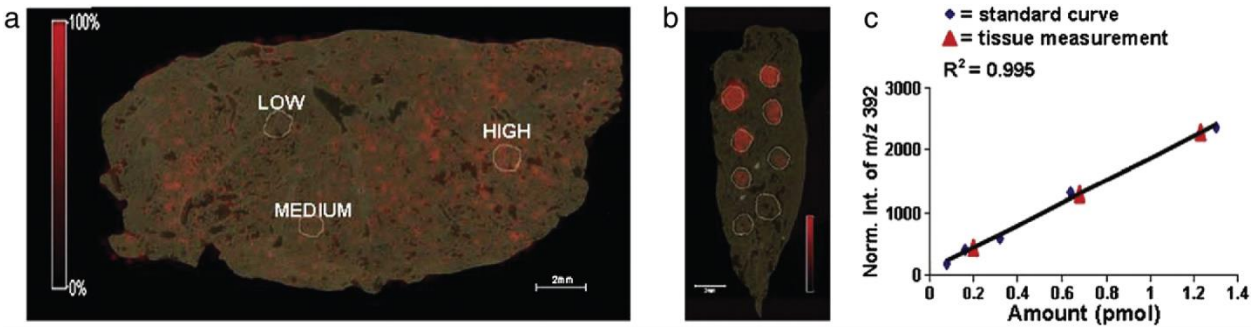


Figure 2. On-tissue calibration curve. (a) Absolute quantitation of tiotropium on lung tissue shows areas that contain 0.20 pmol (low), 0.65 pmol (medium), and 1.25 pmol (high). **(b)** Signals from standard spotted on tissue were matched (c) with normalized signals from the LC–ESI–MS calibration. The red triangles show the linear response of the on-tissue MALDI standards and the blue diamonds show the linear response of corresponding LC–ESI–MS standards. Reprinted with permission.[78]

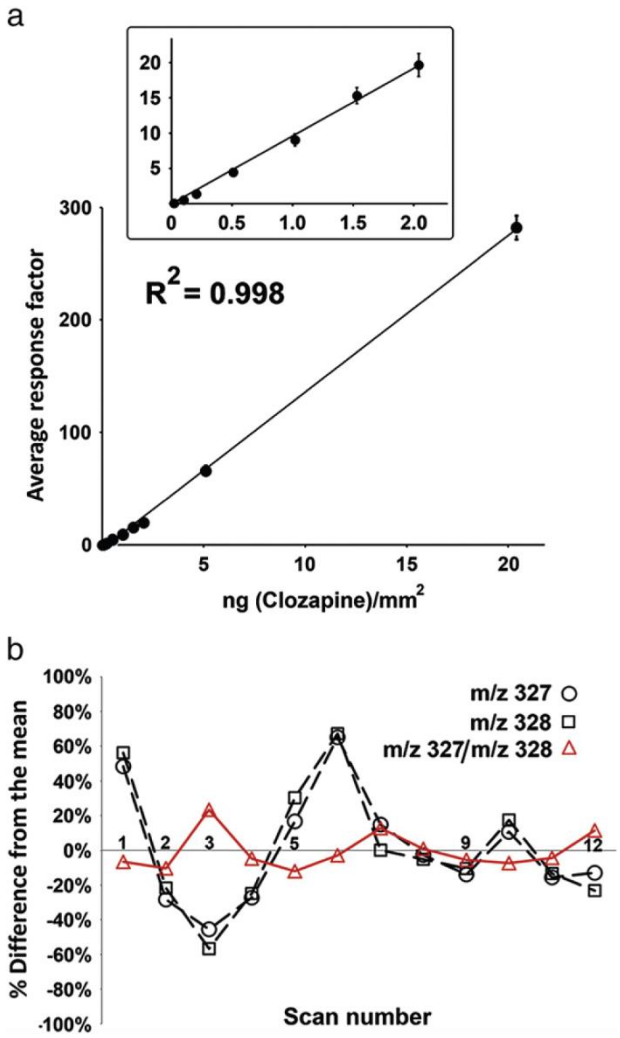


Figure 3. Absolute quantitation with internal standards. (a) The black circles show the linearly increasing response of the intensity ratio between clozapine (m/z 327) and d-clozapine (m/z 328) at standard concentrations spotted on tissue. **(b)** When scanning across a constant concentration of 0.3 ng/mm2, the MS signal is more stable with the clozapine/d-clozapine ratio than with either molecule alone. Reprinted with permission.[80]

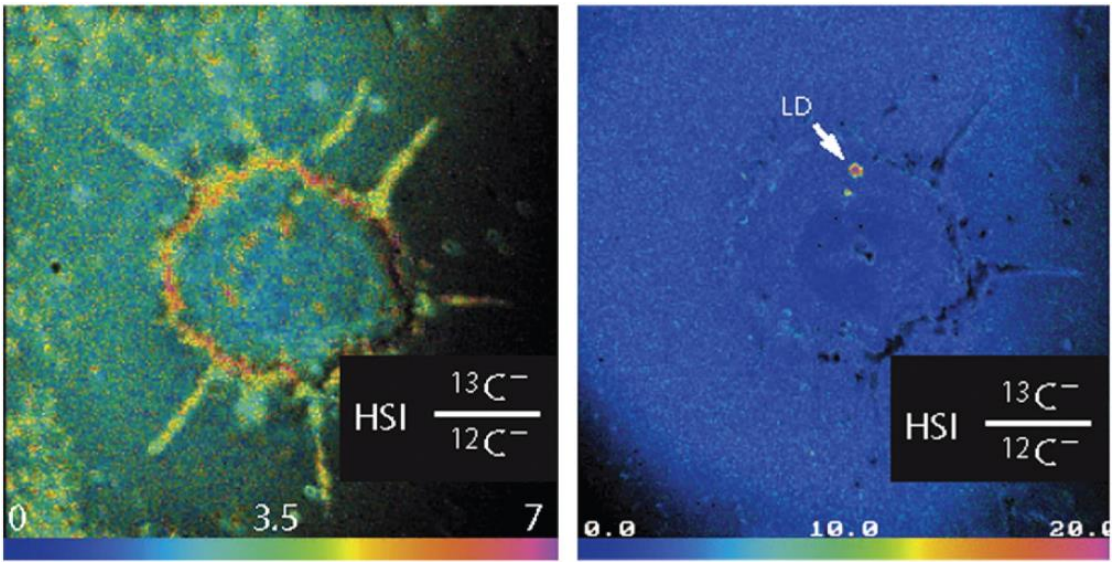


Figure 4. Cellular fatty acid metabolism. (a) A single adipocyte cell shows incorporation of ^{13}C into the cell membrane after incubation with ^{13}C-labeled oleic acid, as well as incorporation into **(b)** intracellular lipid droplets. Reprinted with permission.[81]

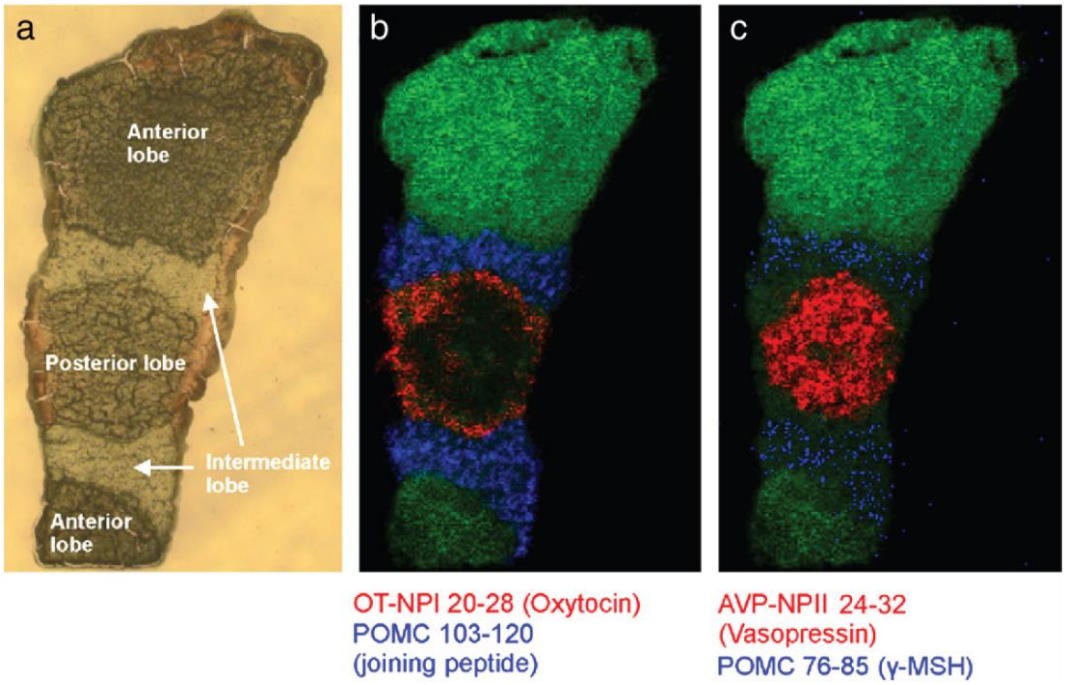


Figure 5. Neuropeptide MSI of a mouse pituitary gland. (a) An optical image of a pituitary shows that the localization of **(b)** oxytocin (red), joining peptide (blue), **(c)** vasopressin (red), and γ-MSH (blue) can be segregated to specific lobes. Reprinted with permission.[93]

Appendix III

The DiLeu Tool: source code

The DiLeu Tool was created as a forms application using Microsoft Visual C# 2010. The following code is intended to be compiled in Microsoft Visual C# 2010.

```
using System;
using System.Collections.Generic;
using System.ComponentModel;
using System.Data;
using System.Drawing;
using System.Linq;
using System.Text;
using System.Windows.Forms;
using System.IO;

namespace TheDiLeuTool
{
class Proteoform
    {
        public string Name { get; set; }
        public bool IsUnshared { get; set; }
        public Dictionary<string, PSM> PSMs = new Dictionary<string, PSM>();
        public double[] QuantArray { get; set; }
        public double[] CorrectedQuantArray { get; set; }
        public double[] NormalizedQuantArray { get; set; }
        public double[] AveragedQuantArray { get; set; }
        public int UnsharedNumber { get; set; }
        public string AllPG { get; set; }
    }
class Protein
    {
        public string PG { get; set; }
        public string RepUniprotID { get; set; }
        public string RepGene { get; set; }
        public List<string> UniprotIDs = new List<string>();
        public List<string> Genes = new List<string>();
        public string Sequence { get; set; }
        public int ProteinLength { get; set; }
        public bool IsID { get; set; }
        public int ProteinsInGroup { get; set; }
        public int TotalPeptides { get; set; }
        public int NumberUnsharedPeptides { get; set; }
        public Residue[] SequenceArray { get; set; }
        public Dictionary<string,PSM> PSMs = new Dictionary<string,PSM>();
        public double[] QuantArray { get; set; }
        public double[] CorrectedQuantArray { get; set; }
        public double[] NormalizedQuantArray { get; set; }
        public int NumberOfProteinsInGroup { get; set; }
        public Dictionary<string, Proteoform> Proteoforms = new Dictionary<string,
Proteoform>();
        public Dictionary<string, PSM> UnlocalizedProteoforms = new Dictionary<string,
PSM>();
        public string DefLine { get; set; }
        public double[] AveragedQuantArray { get; set; }
        public Dictionary<string, Proteoform> UL = new Dictionary<string, Proteoform>();
```

```
    }
class PSM
    {
        public string FileName { get; set; }
        public int PeptideLength { get; set; }
        public string PG { get; set; }
        public int Scan { get; set; }
        public string UniprotID { get; set; }
        public string Gene { get; set; }
        public int Charge { get; set; }
        public string Sequence { get; set; }
        public Residue[] SequenceArray { get; set; }
        public bool IsPhosphorylated { get; set; }
        public bool IsOxidized { get; set; }
        public Dictionary<double, Peak> Spectrum = new Dictionary<double, Peak>();
        public double[] QuantArray { get; set; }
        public string Modifications { get; set; }
        public double Evalue { get; set; }
        public bool MatchedToSpectrum { get; set; }
        public string LocalizedPhosphoSequence { get; set; }
        public string[] PGs { get; set; }
        public bool IsUnshared { get; set; }
        public bool IsLocalized { get; set; }
        public double[] CorrectedQuantArray { get; set; }
        public double[] NormalizedQuantArray { get; set; }
        public double[] AveragedQuantArray { get; set; }

    }
class Peak
    {
        public double MZ { get; set; }
        public double Intensity { get; set; }
    }
class Calculations
    {
        public static bool IsTag(double TAG_MASS, double TOL, Peak pk, double
CURRENT_INTENSITY)
        {
            if (pk.MZ >= (TAG_MASS - TOL))
            {
                if (pk.MZ <= (TAG_MASS + TOL))
                {
                    if (pk.Intensity > CURRENT_INTENSITY)
                    {
                        return true;
                    }
                }
            }

            return false;
        }

        public static double[] DO_PC(double[] UCarray, double[] CorrArray, double[,] PM)
        {
            int X = 0;
            int Y = 0;
```

```
            while (Y < UCarray.Length)
            {
                X = 0;

                while (X < UCarray.Length)
                {
                    CorrArray[Y] += (UCarray[X] * PM[X, Y]);
                    X++;
                }

                Y++;
            }

            return CorrArray;
        }

        public static double[] GetNormalizationFactors(double[] Totals)
        {
            double MIN_VAL = 0.0;

            MIN_VAL = Totals.Min();

            int X = 0;

            while (X < Totals.Length)
            {
                Totals[X] = MIN_VAL / Totals[X];
                X++;
            }

            return Totals;
        }

        public static double GetMeanIntensityFactor(double[] QA)
        {
            double MIN_VAL = QA.Min();
            double FACTOR = 1000.0 / MIN_VAL;
            return FACTOR;
        }

    }
public partial class Form1 : Form
    {
        public Form1()
        {
            InitializeComponent();
        }

        private void button1_Click(object sender, EventArgs e)
        {
            string INPUT1 = PeptideInBox.Text;
            string INPUT2 = ProteinInBox.Text;
            string MATRIXin = PurityMatrixInBox.Text;

            List<string> LogLines = new List<string>();
            StringBuilder sbl = new StringBuilder();

            sbl.Append("[");
```

```
sbl.Append(DateTime.Now);
sbl.Append("]  Begin quantitative protein analysis...");
LogLines.Add(sbl.ToString());
sbl.Clear();
LogLines.Add("");

sbl.Append("Protein identifications input from ");
sbl.Append(INPUT2);
LogLines.Add(sbl.ToString());
sbl.Clear();
sbl.Append("Peptide identifications input from ");
sbl.Append(INPUT1);
LogLines.Add(sbl.ToString());
sbl.Clear();
sbl.Append("Purity Matrix input from ");
sbl.Append(MATRIXin);
LogLines.Add(sbl.ToString());
sbl.Clear();
sbl.Append("MGF directory set to ");
sbl.Append(TxtDirBox.Text);
LogLines.Add(sbl.ToString());
sbl.Clear();
LogLines.Add("");

//Determine OutputNames
StringBuilder NameBuilder = new StringBuilder();
string[] NBarray = INPUT2.Split('.');
NameBuilder.Append(NBarray[0]);
NameBuilder.Append("_log.txt");
string LOGOUTPUT = NameBuilder.ToString();
NameBuilder.Clear();
NBarray = INPUT2.Split('.');
NameBuilder.Append(NBarray[0]);
NameBuilder.Append("_Quant.csv");
string OUTPUT1 = NameBuilder.ToString();
NameBuilder.Clear();
NBarray = INPUT1.Split('.');
NameBuilder.Append(NBarray[0]);
NameBuilder.Append("_QuantifiablePSMs.csv");
string OUTPUT2 = NameBuilder.ToString();
NameBuilder.Clear();

LogLines.Add("Analysis Options");
bool DoCorrections = false;
if (checkBox1.Checked)
{
    DoCorrections = true;
    LogLines.Add("  -Purity corrections will be performed.");
}
bool DoNormalization = false;
if (checkBox2.Checked)
{
    DoNormalization = true;
    LogLines.Add("  -Channel normalizations will be performed.");
}
bool DoMeanIntensities = false;
if (MeanIntBox.Checked)
{
```

```
    DoMeanIntensities = true;
    LogLines.Add("  -Mean intensities will be used.");
}
else
{
    LogLines.Add("  -Summed intensities will be used.");
}

LogLines.Add("");
LogLines.Add("Selected DiLeu Tags:");

List<string> InputFiles = new List<string>();
NameBuilder.Append(TxtDirBox.Text);

string Line;
string[] LineArray;
string[] UParray;
int Counter1 = 0;
int CounterCounter = 0;
int ItemCounter = 0;

if (Box115a.Checked)
{
    Counter1++;
}
if (Box115b.Checked)
{
    Counter1++;
}
if (Box116a.Checked)
{
    Counter1++;
}
if (Box116b.Checked)
{
    Counter1++;
}
if (Box116c.Checked)
{
    Counter1++;
}
if (Box117a.Checked)
{
    Counter1++;
}
if (Box117b.Checked)
{
    Counter1++;
}
if (Box117c.Checked)
{
    Counter1++;
}
if (Box118a.Checked)
{
    Counter1++;
}
if (Box118b.Checked)
```

```
{
    Counter1++;
}
if (Box118c.Checked)
{
    Counter1++;
}
if (Box118d.Checked)
{
    Counter1++;
}

int NUMBER_OF_TAGS = Counter1;
sbl.Clear();
sbl.Append(NUMBER_OF_TAGS);
sbl.Append("-plex experiment");
LogLines.Add(sbl.ToString());

double[] NormalizationFactors = new double[NUMBER_OF_TAGS];

Counter1 = 0;


double TAG_TOL = Convert.ToDouble(ReporterBox.Text);

int P = 0;
int Q = 0;

double[] TAGS = new double[NUMBER_OF_TAGS];
if (Box115a.Checked)
{
    TAGS[Counter1] = 115.12476; //115a
    Counter1++;
    LogLines.Add("  115a");
}
if (Box115b.Checked)
{
    TAGS[Counter1] = 115.13108; //115b
    Counter1++;
    LogLines.Add("  115b");
}
if (Box116a.Checked)
{
    TAGS[Counter1] = 116.12812; //116a
    Counter1++;
    LogLines.Add("  116a");
}
if (Box116b.Checked)
{
    TAGS[Counter1] = 116.13444; //116b
    Counter1++;
    LogLines.Add("  116b");
}
if (Box116c.Checked)
{
    TAGS[Counter1] = 116.14028; //116c
    Counter1++;
    LogLines.Add("  116c");
```

```
            }
            if (Box117a.Checked)
            {
                TAGS[Counter1] = 117.13147; //117a
                Counter1++;
                LogLines.Add("  117a");
            }
            if (Box117b.Checked)
            {
                TAGS[Counter1] = 117.13731; //117b
                Counter1++;
                LogLines.Add("  117b");
            }
            if (Box117c.Checked)
            {
                TAGS[Counter1] = 117.14363; //117c
                Counter1++;
                LogLines.Add("  117c");
            }
            if (Box118a.Checked)
            {
                TAGS[Counter1] = 118.13483; //118a
                Counter1++;
                LogLines.Add("  118a");
            }
            if (Box118b.Checked)
            {
                TAGS[Counter1] = 118.14067; //118b
                Counter1++;
                LogLines.Add("  118b");
            }
            if (Box118c.Checked)
            {
                TAGS[Counter1] = 118.14699; //118c
                Counter1++;
                LogLines.Add("  118c");
            }
            if (Box118d.Checked)
            {
                TAGS[Counter1] = 118.15283; //118d
                Counter1++;
                LogLines.Add("  118d");
            }

            Counter1 = 0;
            LogLines.Add("");

            double[,] PurityMatrix = new double[NUMBER_OF_TAGS, NUMBER_OF_TAGS];

            List<string> ConsoleLines = new List<string>();



            if (DoCorrections)
            {
                Console.WriteLine("[{0}]  Creating DiLeu purity matrix...",
DateTime.Now);
                sbl.Clear();
```

```
                sbl.Append("[");
                sbl.Append(DateTime.Now);
                sbl.Append("]  Creating DiLeu purity matrix...");
                LogLines.Add(sbl.ToString());
                LogLines.Add("");
                sbl.Clear();


                StreamReader MatrixReader = new StreamReader(MATRIXin);
                Line = MatrixReader.ReadLine();

                while (Q < NUMBER_OF_TAGS)
                {
                    Line = MatrixReader.ReadLine();
                    LineArray = Line.Split(',');
                    P = 0;
                    while (P < NUMBER_OF_TAGS)
                    {
                        PurityMatrix[P, Q] = Convert.ToDouble(LineArray[P + 1]);
                        P++;
                    }

                    Q++;
                }

                MatrixReader.Close();

                Console.WriteLine("[{0}]  Purity matrix has been created.",
DateTime.Now);
            }

            Dictionary<string, Protein> ProteinIDs = new Dictionary<string, Protein>();

            StreamReader sr2 = new StreamReader(INPUT2);

            Line = sr2.ReadLine();

            Console.WriteLine("[{0}]  Begin reading protein identifications...",
DateTime.Now);

            sbl.Clear();
            sbl.Append("[");
            sbl.Append(DateTime.Now);
            sbl.Append("]  Begin reading protein identifications...");
            LogLines.Add(sbl.ToString());
            sbl.Clear();

            while (!sr2.EndOfStream)
            {
                Line = sr2.ReadLine();
                LineArray = Line.Split(',');

                Protein Prot = new Protein();

                Prot.PG = LineArray[0];
                UParray = LineArray[1].Split('|');
                Prot.RepUniprotID = UParray[1];
                Prot.ProteinsInGroup = Convert.ToInt32(LineArray[6]);
```

```
                Prot.TotalPeptides = Convert.ToInt32(LineArray[8]);
                Prot.NumberUnsharedPeptides = Prot.TotalPeptides -
Convert.ToInt32(LineArray[9]);
                Prot.DefLine = LineArray[1];

                UParray = LineArray[14].Split('|');
                ItemCounter = 0;

                while (ItemCounter < UParray.Length)
                {
                    Prot.UniprotIDs.Add(UParray[ItemCounter]);
                    ItemCounter++;
                }

                UParray = LineArray[15].Split('|');
                ItemCounter = 0;
                Prot.RepGene = UParray[0];
                while (ItemCounter < UParray.Length)
                {
                    Prot.Genes.Add(UParray[ItemCounter]);
                    ItemCounter++;
                }

                ProteinIDs.Add(Prot.PG, Prot);

                Counter1++;
                CounterCounter++;

                if (CounterCounter == 500)
                {
                    Console.WriteLine("[{0}]  Read {1} protein IDs.", DateTime.Now,
Counter1);
                    CounterCounter = 0;
                }
            }

            sr2.Close();

            sbl.Clear();
            sbl.Append("[");
            sbl.Append(DateTime.Now);
            sbl.Append("]  Read a total of ");
            sbl.Append(ProteinIDs.Count);
            sbl.Append(" protein IDs.");
            LogLines.Add(sbl.ToString());
            LogLines.Add("");
            sbl.Clear();

            StreamReader sr1 = new StreamReader(INPUT1);
            StringBuilder sb = new StringBuilder();
            int PepCounter = 0;
            Line = sr1.ReadLine();
            LineArray = Line.Split(',');
            Console.WriteLine("[{0}]  Begin reading peptide identifications...",
DateTime.Now);
            sbl.Clear();
            sbl.Append("[");
            sbl.Append(DateTime.Now);
```

```
            sbl.Append("]  Begin reading PSMs...");
            LogLines.Add(sbl.ToString());
            sbl.Clear();
            Dictionary<string, PSM> PSMs = new Dictionary<string, PSM>();
            while (!sr1.EndOfStream)
            {
                Line = sr1.ReadLine();
                LineArray = Line.Split(',');
                UParray = LineArray[23].Split('|');

                if (UParray.Length == 1)
                {
                    PepCounter++;

                    PSM psm = new PSM();

                    psm.Scan = Convert.ToInt32(LineArray[0]);

                    psm.FileName = LineArray[1];
                    psm.Sequence = LineArray[2];
                    psm.PeptideLength = psm.Sequence.Length;
                    psm.SequenceArray = new Residue[psm.PeptideLength];
                    psm.PG = LineArray[23];
                    psm.Modifications = LineArray[10];
                    psm.Charge = Convert.ToInt32(LineArray[11]);
                    psm.Evalue = Convert.ToDouble(LineArray[3]);



                    PSMs.Add(psm.FileName,psm);
                    NameBuilder.Clear();
                    NBarray = LineArray[1].Split('.');
                    NameBuilder.Append(TxtDirBox.Text);
                    NameBuilder.Append("\\");
                    NameBuilder.Append(NBarray[0]);
                    NameBuilder.Append("_FTMS_HCD.mgf");

                    if (!InputFiles.Contains(NameBuilder.ToString()))
                    {
                        InputFiles.Add(NameBuilder.ToString());
                    }

                }
            }

            sr1.Close();

            Console.WriteLine("[{0}]  Read a total of {1} unshared PSMs.", DateTime.Now,
PepCounter);
            sbl.Clear();
            sbl.Append("[");
            sbl.Append(DateTime.Now);
            sbl.Append("]  Read a total of ");
            sbl.Append(PSMs.Count);
            sbl.Append(" PSMs.");
            LogLines.Add(sbl.ToString());
            LogLines.Add("");
            sbl.Clear();
```

```
            int MSMScounter = 0;
            CounterCounter = 0;
            PepCounter = 0;
            int FileCounter = 1;

            Console.WriteLine("[{0}] Begin reading MS/MS spectra...", DateTime.Now);
            sbl.Clear();
            sbl.Append("[");
            sbl.Append(DateTime.Now);
            sbl.Append("]  Begin parsing MS/MS spectra...");
            LogLines.Add(sbl.ToString());
            sbl.Clear();
            foreach (string iFile in InputFiles)
            {
                Console.WriteLine("[{0}]  Opened {1} ({2} of {3})", DateTime.Now,
iFile,FileCounter,InputFiles.Count);
                sbl.Clear();
                sbl.Append("[");
                sbl.Append(DateTime.Now);
                sbl.Append("]  Opened ");
                sbl.Append(iFile);
                sbl.Append(" (");
                sbl.Append(FileCounter);
                sbl.Append(" of ");
                sbl.Append(InputFiles.Count);
                sbl.Append(")");
                LogLines.Add(sbl.ToString());
                sbl.Clear();
                FileCounter++;

                StreamReader sr = new StreamReader(iFile);

                Line = sr.ReadLine();

                while (!sr.EndOfStream)
                {
                    Line = sr.ReadLine();
                    LineArray = Line.Split('=');
                    sb.Clear();
                    sb.Append(LineArray[1]);
                    sb.Append(".dta");

                    if (PSMs.ContainsKey(sb.ToString()))
                    {
                        PSMs[sb.ToString()].MatchedToSpectrum = true;
                        PepCounter++;

                        Line = sr.ReadLine();
                        Line = sr.ReadLine();
                        Line = sr.ReadLine();
                        Line = sr.ReadLine();
                        Line = sr.ReadLine();

                        LineArray = Line.Split(' ');

                        while ((Convert.ToDouble(LineArray[0]) < 119.0))
                        {
                            Peak peak = new Peak();
```

```
                peak.MZ = Convert.ToDouble(LineArray[0]);
                peak.Intensity = Convert.ToDouble(LineArray[1]);

                if (!PSMs[sb.ToString()].Spectrum.ContainsKey(peak.MZ))
                {
                    PSMs[sb.ToString()].Spectrum.Add(peak.MZ, peak);
                }


                Line = sr.ReadLine();
                LineArray = Line.Split(' ');
            }
        }



        while (!Line.Equals("END IONS"))
        {
            Line = sr.ReadLine();
        }

        if (!sr.EndOfStream)
        {
            Line = sr.ReadLine();
        }

        if (!sr.EndOfStream)
        {
            Line = sr.ReadLine();
        }

        MSMScounter++;
        CounterCounter++;
    }

    sr.Close();
}

sbl.Clear();
sbl.Append("[");
sbl.Append(DateTime.Now);
sbl.Append("]  Parsed a total of ");
sbl.Append(MSMScounter);
sbl.Append(" MS/MS spectra.");
LogLines.Add(sbl.ToString());
LogLines.Add("");
sbl.Clear();

sbl.Clear();
sbl.Append("[");
sbl.Append(DateTime.Now);
sbl.Append("]  Matching PSMs to proteins...");
LogLines.Add(sbl.ToString());
LogLines.Add("");
sbl.Clear();
foreach (KeyValuePair<string, PSM> psm in PSMs)
{
    ProteinIDs[psm.Value.PG].PSMs.Add(psm.Key, psm.Value);
```

```
            }

            PSMs.Clear();

            Console.WriteLine("");
            Console.WriteLine("[{0}]  Begin protein quantification...", DateTime.Now);

            sbl.Clear();
            sbl.Append("[");
            sbl.Append(DateTime.Now);
            sbl.Append("]  Begin protein quantification...");
            LogLines.Add(sbl.ToString());
            sbl.Clear();

            int x;

            Counter1 = 0;
            CounterCounter = 0;

            foreach (KeyValuePair<string, Protein> Prot in ProteinIDs)
            {
                Prot.Value.QuantArray = new double[NUMBER_OF_TAGS];
                x = 0;

                while (x < NUMBER_OF_TAGS)
                {
                    Prot.Value.QuantArray[x] = 0;
                    x++;
                }

                foreach (KeyValuePair<string, PSM> psm in Prot.Value.PSMs)
                {
                    psm.Value.QuantArray = new double[NUMBER_OF_TAGS];

                    x = 0;
                    while (x < NUMBER_OF_TAGS)
                    {
                        psm.Value.QuantArray[x] = 0;
                        x++;
                    }

                    x = 0;

                    while (x < NUMBER_OF_TAGS)
                    {
                        foreach (KeyValuePair<double, Peak> peak in psm.Value.Spectrum)
                    {
                        if (Calculations.IsTag(TAGS[x], TAG_TOL, peak.Value,
psm.Value.QuantArray[x]))
                        {
                            psm.Value.QuantArray[x] = peak.Value.Intensity;
                        }
                    }
                        x++;
                    }


                    Counter1++;
```

```
                    CounterCounter++;
                    if (CounterCounter == 5000)
                    {
                        Console.WriteLine("[{0}]  Quantified {1} PSMs.", DateTime.Now,
Counter1);
                        CounterCounter = 0;
                    }


                }
            }

            List<string> RemovePSMs = new List<string>();
            int RemoveCounter = 0;

            foreach (KeyValuePair<string, Protein> Prot in ProteinIDs)
            {
                foreach (KeyValuePair<string, PSM> psm in Prot.Value.PSMs)
                {
                    x = 0;

                    while (x < NUMBER_OF_TAGS)
                    {
                        if (!(psm.Value.QuantArray[x] > 0.0))
                        {
                            if (!RemovePSMs.Contains(psm.Key))
                            {
                                RemovePSMs.Add(psm.Key);
                            }
                        }
                        x++;
                    }
                }

                foreach (string KEY in RemovePSMs)
                {
                    Prot.Value.PSMs.Remove(KEY);
                }

                RemoveCounter += RemovePSMs.Count;
                RemovePSMs.Clear();
            }

            sbl.Clear();
            sbl.Append("[");
            sbl.Append(DateTime.Now);
            sbl.Append("]  Removed ");
            sbl.Append(RemoveCounter);
            sbl.Append(" PSMs (contained at least one reporter channel with intensity =
0).");
            LogLines.Add(sbl.ToString());
            sbl.Clear();

            if (DoCorrections)
            {
                Console.WriteLine("[{0}]  Correcting reporter intensities for purity...",
DateTime.Now);
                sbl.Clear();
```

```
                sbl.Append("[");
                sbl.Append(DateTime.Now);
                sbl.Append("]  Performing purity corrections...");
                LogLines.Add(sbl.ToString());
                sbl.Clear();
                foreach (KeyValuePair<string, Protein> Prot in ProteinIDs)
                {
                    foreach (KeyValuePair<string, PSM> psm in Prot.Value.PSMs)
                    {
                        psm.Value.CorrectedQuantArray = new double[NUMBER_OF_TAGS];
                        x = 0;
                        while (x < NUMBER_OF_TAGS)
                        {
                            psm.Value.CorrectedQuantArray[x] = 0.0;
                            x++;
                        }

                        psm.Value.CorrectedQuantArray =
Calculations.DO_PC(psm.Value.QuantArray, psm.Value.CorrectedQuantArray, PurityMatrix);
                    }
                }
            }

            foreach (KeyValuePair<string, Protein> Prot in ProteinIDs)
            {
                x = 0;
                if (DoCorrections)
                {
                    Prot.Value.CorrectedQuantArray = new double[NUMBER_OF_TAGS];
                    x = 0;
                    while (x < NUMBER_OF_TAGS)
                    {
                        Prot.Value.CorrectedQuantArray[x] = 0.0;
                        x++;
                    }
                }
                x = 0;
                while (x < NUMBER_OF_TAGS)
                {
                    foreach (KeyValuePair<string, PSM> psm in Prot.Value.PSMs)
                    {
                        Prot.Value.QuantArray[x] += psm.Value.QuantArray[x];
                        if (DoCorrections)
                        {
                            Prot.Value.CorrectedQuantArray[x] +=
psm.Value.CorrectedQuantArray[x];
                        }
                    }
                    x++;
                }

            }

            if (DoNormalization)
            {
                sbl.Clear();
                sbl.Append("[");
                sbl.Append(DateTime.Now);
```

```
                sbl.Append("]  Performing channel normalization...");
                LogLines.Add(sbl.ToString());
                sbl.Clear();
                double[] ChannelTotals = new double[NUMBER_OF_TAGS];

                x = 0;
                while (x < NUMBER_OF_TAGS)
                {
                    ChannelTotals[x] = 0.0;
                    x++;
                }

                foreach (KeyValuePair<string, Protein> Prot in ProteinIDs)
                {
                    Prot.Value.NormalizedQuantArray = new double[NUMBER_OF_TAGS];

                    x = 0;

                    if (DoCorrections)
                    {
                        while (x < NUMBER_OF_TAGS)
                        {
                            ChannelTotals[x] += Prot.Value.CorrectedQuantArray[x];
                            x++;
                        }
                    }
                    else
                    {
                        while (x < NUMBER_OF_TAGS)
                        {
                            ChannelTotals[x] += Prot.Value.QuantArray[x];
                            x++;
                        }
                    }
                }

                //double[] NormalizationFactors = new double[NUMBER_OF_TAGS];

                NormalizationFactors =
Calculations.GetNormalizationFactors(ChannelTotals);
                x = 0;
                while (x < NUMBER_OF_TAGS)
                {
                    Console.WriteLine("{0} ", ChannelTotals[x]);
                    x++;
                }
                Console.WriteLine("");
                foreach (KeyValuePair<string, Protein> Prot in ProteinIDs)
                {
                    x = 0;
                    while (x < NUMBER_OF_TAGS)
                    {
                        Prot.Value.NormalizedQuantArray[x] = 0.0;
                        x++;
                    }

                    foreach(KeyValuePair<string,PSM> psm in Prot.Value.PSMs)
                    {
```

```
                        psm.Value.NormalizedQuantArray = new double[NUMBER_OF_TAGS];
                        x = 0;
                        while (x < NUMBER_OF_TAGS)
                        {
                            if (DoCorrections)
                            {
                                psm.Value.NormalizedQuantArray[x] =
(NormalizationFactors[x] * psm.Value.CorrectedQuantArray[x]);
                            }
                            else
                            {
                                psm.Value.NormalizedQuantArray[x] =
(NormalizationFactors[x] * psm.Value.QuantArray[x]);
                            }
                            x++;
                        }

                        x = 0;
                        while (x < NUMBER_OF_TAGS)
                        {
                            Prot.Value.NormalizedQuantArray[x] +=
psm.Value.NormalizedQuantArray[x];
                            x++;
                        }
                    }
                }
            }

            //Mean Intensities Option

            double MeanFactor = 0.0;

            if (DoMeanIntensities)
            {
                sbl.Clear();
                sbl.Append("[");
                sbl.Append(DateTime.Now);
                sbl.Append("]  Calculating mean protein reporter intensities...");
                LogLines.Add(sbl.ToString());
                sbl.Clear();
                foreach (KeyValuePair<string, Protein> Prot in ProteinIDs)
                {
                    x = 0;
                    while (x < NUMBER_OF_TAGS)
                    {
                        Prot.Value.QuantArray[x] = 0.0;
                        if (DoCorrections)
                        {
                            Prot.Value.CorrectedQuantArray[x] = 0.0;
                        }
                        if (DoNormalization)
                        {
                            Prot.Value.NormalizedQuantArray[x] = 0.0;
                        }
                        x++;
                    }

                    foreach (KeyValuePair<string, PSM> psm in Prot.Value.PSMs)
```

```
                    {
                        MeanFactor =
Calculations.GetMeanIntensityFactor(psm.Value.QuantArray);
                        x = 0;
                        while (x < NUMBER_OF_TAGS)
                        {
                            psm.Value.QuantArray[x] = MeanFactor *
psm.Value.QuantArray[x];
                            x++;
                        }

                        if (DoCorrections)
                        {
                            MeanFactor =
Calculations.GetMeanIntensityFactor(psm.Value.CorrectedQuantArray);
                            x = 0;
                            while (x < NUMBER_OF_TAGS)
                            {
                                psm.Value.CorrectedQuantArray[x] = MeanFactor *
psm.Value.CorrectedQuantArray[x];
                                x++;
                            }
                        }

                        if (DoNormalization)
                        {
                            MeanFactor =
Calculations.GetMeanIntensityFactor(psm.Value.NormalizedQuantArray);
                            x = 0;
                            while (x < NUMBER_OF_TAGS)
                            {
                                psm.Value.NormalizedQuantArray[x] = MeanFactor *
psm.Value.NormalizedQuantArray[x];
                                x++;
                            }
                        }
                    }

                    foreach (KeyValuePair<string, PSM> psm in Prot.Value.PSMs)
                    {
                        x = 0;
                        while (x < NUMBER_OF_TAGS)
                        {
                            Prot.Value.QuantArray[x] += psm.Value.QuantArray[x];
                            if (DoCorrections)
                            {
                                Prot.Value.CorrectedQuantArray[x] +=
psm.Value.CorrectedQuantArray[x];
                            }
                            if (DoNormalization)
                            {
                                Prot.Value.NormalizedQuantArray[x] +=
psm.Value.NormalizedQuantArray[x];
                            }
                            x++;
                        }
                    }
```

```
                    x = 0;
                    while (x < NUMBER_OF_TAGS)
                    {
                        Prot.Value.QuantArray[x] = Prot.Value.QuantArray[x] /
Prot.Value.PSMs.Count;
                        if (DoCorrections)
                        {
                            Prot.Value.CorrectedQuantArray[x] =
Prot.Value.CorrectedQuantArray[x] / Prot.Value.PSMs.Count;
                        }
                        if (DoNormalization)
                        {
                            Prot.Value.NormalizedQuantArray[x] =
Prot.Value.NormalizedQuantArray[x] / Prot.Value.PSMs.Count;
                        }
                        x++;
                    }
                }
            }

            //OUTPUTS

            List<string> OutputLines = new List<string>();

            sbl.Clear();
            LogLines.Add("");
            sbl.Append("[");
            sbl.Append(DateTime.Now);
            sbl.Append("]  Writing outputs...");
            LogLines.Add(sbl.ToString());
            LogLines.Add("NOTE: Protein identifications with no quantifiable PSMs will be
omitted from output.");
            LogLines.Add("");
            sbl.Clear();

            sb.Clear();
            sb.Append("Protein Group,Uniprot ID,DefLine,Gene,Quantified PSMs,Raw 115a,Raw
115b,Raw 116a,Raw 116b,Raw 116c,Raw 117a,Raw 117b,Raw 117c,Raw 118a,Raw 118b");
            if (DoCorrections)
            {
                sb.Append(",Corrected 115a,Corrected 115b,Corrected 116a,Corrected
116b,Corrected 116c,Corrected 117a,Corrected 117b,Corrected 117c,Corrected 118a,Corrected
118b");
            }
            if (DoNormalization)
            {
                sb.Append(",Normalized 115a,Normalized 115b,Normalized 116a,Normalized
116b,Normalized 116c,Normalized 117a,Normalized 117b,Normalized 117c,Normalized
118a,Normalized 118b");
            }

            OutputLines.Add(sb.ToString());

            sb.Clear();

            foreach (KeyValuePair<string, Protein> Prot in ProteinIDs)
            {
                if (Prot.Value.PSMs.Count > 0)
```

```
            {
                sb.Clear();
                sb.Append(Prot.Value.PG);
                sb.Append(",");
                sb.Append(Prot.Value.RepUniprotID);
                sb.Append(",");
                sb.Append(Prot.Value.DefLine);
                sb.Append(",");
                sb.Append(Prot.Value.RepGene);
                sb.Append(",");
                sb.Append(Prot.Value.PSMs.Count);
                sb.Append(",");

                x = 0;

                while (x < NUMBER_OF_TAGS)
                {
                    sb.Append(Prot.Value.QuantArray[x]);
                    sb.Append(",");
                    x++;
                }

                if (DoCorrections)
                {
                    x = 0;
                    while (x < NUMBER_OF_TAGS)
                    {
                        sb.Append(Prot.Value.CorrectedQuantArray[x]);
                        sb.Append(",");
                        x++;
                    }
                }

                if (DoNormalization)
                {
                    x = 0;
                    while (x < NUMBER_OF_TAGS)
                    {
                        sb.Append(Prot.Value.NormalizedQuantArray[x]);
                        sb.Append(",");
                        x++;
                    }
                }

                OutputLines.Add(sb.ToString());
            }

        }

        File.WriteAllLines(OUTPUT1, OutputLines);

        OutputLines.Clear();
        sb.Clear();
        sb.Append("Scan Number,Scan Name,Sequence,Charge,Modifications,Protein
Group,Representative UniprotID,Unshared?,Defline,Raw 115a,Raw 115b,Raw 116a,Raw 116b,Raw
116c,Raw 117a,Raw 117b,Raw 117c,Raw 118a,Raw 118b");
        if (DoCorrections)
        {
```

```
                sb.Append(",Corrected 115a,Corrected 115b,Corrected 116a,Corrected
116b,Corrected 116c,Corrected 117a,Corrected 117b,Corrected 117c,Corrected 118a,Corrected
118b");
            }
            if (DoNormalization)
            {
                sb.Append(",Normalized 115a,Normalized 115b,Normalized 116a,Normalized
116b,Normalized 116c,Normalized 117a,Normalized 117b,Normalized 117c,Normalized
118a,Normalized 118b");
            }

            OutputLines.Add(sb.ToString());

            sb.Clear();

            StringBuilder PepBuilder = new StringBuilder();

            List<string> UniquePeptides = new List<string>();

            foreach (KeyValuePair<string, Protein> Prot in ProteinIDs)
            {
                foreach (KeyValuePair<string, PSM> psm in Prot.Value.PSMs)
                {
                    sb.Clear();
                    PepBuilder.Clear();

                    sb.Append(psm.Value.Scan);
                    sb.Append(",");
                    sb.Append(psm.Value.FileName);
                    sb.Append(",");
                    sb.Append(psm.Value.Sequence);
                    PepBuilder.Append(psm.Value.Sequence);
                    sb.Append(",");
                    sb.Append(psm.Value.Charge);
                    PepBuilder.Append(psm.Value.Charge);
                    sb.Append(",");
                    sb.Append(psm.Value.Modifications);
                    PepBuilder.Append(psm.Value.Modifications);
                    sb.Append(",");
                    sb.Append(Prot.Value.PG);
                    sb.Append(",");
                    sb.Append(Prot.Value.RepUniprotID);
                    sb.Append(",");
                    sb.Append("TRUE");
                    sb.Append(",");
                    sb.Append(Prot.Value.DefLine);
                    sb.Append(",");

                    x = 0;

                    while (x < NUMBER_OF_TAGS)
                    {
                        sb.Append(psm.Value.QuantArray[x]);
                        sb.Append(",");
                        x++;
                    }

                    if (DoCorrections)
```

```
        {
            x = 0;
            while (x < NUMBER_OF_TAGS)
            {
                sb.Append(psm.Value.CorrectedQuantArray[x]);
                sb.Append(",");
                x++;
            }
        }

        if (DoNormalization)
        {
            x = 0;
            while (x < NUMBER_OF_TAGS)
            {
                sb.Append(psm.Value.NormalizedQuantArray[x]);
                sb.Append(",");
                x++;
            }
        }

        OutputLines.Add(sb.ToString());

        if (!UniquePeptides.Contains(PepBuilder.ToString()))
        {
            UniquePeptides.Add(PepBuilder.ToString());
        }
    }
}

File.WriteAllLines(OUTPUT2, OutputLines);

Counter1 = 0;
foreach (KeyValuePair<string,Protein> Prot in ProteinIDs)
{
    Counter1 += Prot.Value.PSMs.Count;
}
sbl.Clear();
sbl.Append("[");
sbl.Append(DateTime.Now);
sbl.Append("]  DONE!!!");
LogLines.Add(sbl.ToString());
LogLines.Add("");
sbl.Clear();
LogLines.Add("SUMMARY");
sbl.Append(ProteinIDs.Count);
sbl.Append(" protein identifications");
LogLines.Add(sbl.ToString());
sbl.Clear();
sbl.Append(Counter1);
sbl.Append(" quantifiable PSMs");
LogLines.Add(sbl.ToString());
sbl.Clear();
sbl.Append(UniquePeptides.Count);
sbl.Append(" unique peptide identifications.");
LogLines.Add(sbl.ToString());
sbl.Clear();
LogLines.Add("");
```

```
    if(DoNormalization)
    {
        LogLines.Add("Channel Normalization Factors:");
    x = 0;
    while (x < NUMBER_OF_TAGS)
    {
        LogLines.Add(Convert.ToString(NormalizationFactors[x]));
        x++;
    }
    }

    File.WriteAllLines(LOGOUTPUT, LogLines);

    LogLines.Clear();
    OutputLines.Clear();
    UniquePeptides.Clear();

    ProteinIDs.Clear();

    Console.WriteLine("[{0}]  Done!!!", DateTime.Now);

    MessageBox.Show("DONE!");


}

private void button2_Click(object sender, EventArgs e)
{
    string INPUT1 = PhosphoPSMinBox.Text;
    string INPUT2 = PhosphoproteinInBox.Text;
    string MATRIXin = textBox1.Text;

    //Determine OutputNames
    StringBuilder NameBuilder = new StringBuilder();
    string[] NBarray = INPUT2.Split('.');
    NameBuilder.Append(NBarray[0]);
    NameBuilder.Append("_UnlocalizedPhosphoQuant.csv");
    string OUTPUT1 = NameBuilder.ToString();
    NameBuilder.Clear();
    NBarray = INPUT2.Split('.');
    NameBuilder.Append(NBarray[0]);
    NameBuilder.Append("_LocalizedPhosphoQuant.csv");
    string OUTPUT2 = NameBuilder.ToString();
    NameBuilder.Clear();
    NBarray = INPUT2.Split('.');
    NameBuilder.Append(NBarray[0]);
    NameBuilder.Append("_AllProteins.csv");
    string OUTPUT3 = NameBuilder.ToString();
    NameBuilder.Clear();
    NBarray = INPUT2.Split('.');
    NameBuilder.Append(NBarray[0]);
    NameBuilder.Append("_AllPSMs.csv");
    string OUTPUT4 = NameBuilder.ToString();


    bool DoCorrections = false;
    if (checkBox1.Checked)
    {
```

```
    DoCorrections = true;
}
bool DoNormalization = false;
if (checkBox2.Checked)
{
    DoNormalization = true;
}
bool DoMeanIntensities = false;
if (MeanIntBox.Checked)
{
    DoMeanIntensities = true;
}

List<string> InputFiles = new List<string>();

string Line;
string[] LineArray;
string[] UParray;
int Counter1 = 0;
int CounterCounter = 0;
int ItemCounter = 0;

Console.WriteLine("[{0}]  Building tag matrix...", DateTime.Now);
if (Box115a.Checked)
{
    Counter1++;
}
if (Box115b.Checked)
{
    Counter1++;
}
if (Box116a.Checked)
{
    Counter1++;
}
if (Box116b.Checked)
{
    Counter1++;
}
if (Box116c.Checked)
{
    Counter1++;
}
if (Box117a.Checked)
{
    Counter1++;
}
if (Box117b.Checked)
{
    Counter1++;
}
if (Box117c.Checked)
{
    Counter1++;
}
if (Box118a.Checked)
{
    Counter1++;
```

```
}
if (Box118b.Checked)
{
    Counter1++;
}
if (Box118c.Checked)
{
    Counter1++;
}
if (Box118d.Checked)
{
    Counter1++;
}

int NUMBER_OF_TAGS = Counter1;

Counter1 = 0;


double TAG_TOL = Convert.ToDouble(ReporterBox.Text);

int P = 0;
int Q = 0;

double[] TAGS = new double[NUMBER_OF_TAGS];
if (Box115a.Checked)
{
    TAGS[Counter1] = 115.12476; //115a
    Counter1++;
}
if (Box115b.Checked)
{
    TAGS[Counter1] = 115.13108; //115b
    Counter1++;
}
if (Box116a.Checked)
{
    TAGS[Counter1] = 116.12812; //116a
    Counter1++;
}
if (Box116b.Checked)
{
    TAGS[Counter1] = 116.13444; //116b
    Counter1++;
}
if (Box116c.Checked)
{
    TAGS[Counter1] = 116.14028; //116c
    Counter1++;
}
if (Box117a.Checked)
{
    TAGS[Counter1] = 117.13147; //117a
    Counter1++;
}
if (Box117b.Checked)
{
    TAGS[Counter1] = 117.13731; //117b
```

```
    Counter1++;
}
if (Box117c.Checked)
{
    TAGS[Counter1] = 117.14363; //117c
    Counter1++;
}
if (Box118a.Checked)
{
    TAGS[Counter1] = 118.13483; //118a
    Counter1++;
}
if (Box118b.Checked)
{
    TAGS[Counter1] = 118.14067; //118b
    Counter1++;
}
if (Box118c.Checked)
{
    TAGS[Counter1] = 118.14699; //118c
    Counter1++;
}
if (Box118d.Checked)
{
    TAGS[Counter1] = 118.15283; //118d
    Counter1++;
}

Counter1 = 0;

double[,] PurityMatrix = new double[NUMBER_OF_TAGS, NUMBER_OF_TAGS];

if (DoCorrections)
{
    Console.WriteLine("");
    Console.WriteLine("[{0}]  Building purity matrix...", DateTime.Now);


    StreamReader MatrixReader = new StreamReader(MATRIXin);
    Line = MatrixReader.ReadLine();

    while (Q < NUMBER_OF_TAGS)
    {
        Line = MatrixReader.ReadLine();
        LineArray = Line.Split(',');
        P = 0;
        while (P < NUMBER_OF_TAGS)
        {
            PurityMatrix[P, Q] = Convert.ToDouble(LineArray[P + 1]);
            P++;
        }

        Q++;
    }

    MatrixReader.Close();
```

```
            }

            Dictionary<string, Protein> ProteinIDs = new Dictionary<string, Protein>();

            StreamReader sr2 = new StreamReader(INPUT2);

            Line = sr2.ReadLine();

            Counter1 = 0;
            CounterCounter = 0;
            Console.WriteLine("");
            Console.WriteLine("[{0}]  Begin reading protein identifications...",
DateTime.Now);
            while (!sr2.EndOfStream)
            {
                Line = sr2.ReadLine();
                LineArray = Line.Split(',');

                Protein Prot = new Protein();

                Prot.PG = LineArray[0];
                UParray = LineArray[1].Split('|');
                Prot.RepUniprotID = UParray[1];
                Prot.ProteinsInGroup = Convert.ToInt32(LineArray[6]);
                Prot.TotalPeptides = Convert.ToInt32(LineArray[8]);
                Prot.NumberUnsharedPeptides = Prot.TotalPeptides -
Convert.ToInt32(LineArray[9]);
                Prot.DefLine = LineArray[1];

                UParray = LineArray[14].Split('|');
                ItemCounter = 0;

                while (ItemCounter < UParray.Length)
                {
                    Prot.UniprotIDs.Add(UParray[ItemCounter]);
                    ItemCounter++;
                }

                UParray = LineArray[15].Split('|');
                ItemCounter = 0;
                Prot.RepGene = UParray[0];
                while (ItemCounter < UParray.Length)
                {
                    Prot.Genes.Add(UParray[ItemCounter]);
                    ItemCounter++;
                }

                ProteinIDs.Add(Prot.PG, Prot);

                Counter1++;
                CounterCounter++;

                if (CounterCounter == 500)
                {
                    Console.WriteLine("[{0}]  Read {1} protein IDs.", DateTime.Now,
Counter1);
                    CounterCounter = 0;
                }
```

```
            }

            sr2.Close();

            StreamReader sr1 = new StreamReader(INPUT1);
            StringBuilder sb = new StringBuilder();
            int PepCounter = 0;
            Line = sr1.ReadLine();
            LineArray = Line.Split(',');

            Console.WriteLine("[{0}]  A total of {1} protein identifications were read.",
DateTime.Now, ProteinIDs.Count);
            Console.WriteLine("");

            Console.WriteLine("[{0}]  Begin reading peptide identifications...",
DateTime.Now);
            Dictionary<string, PSM> PSMs = new Dictionary<string, PSM>();
            while (!sr1.EndOfStream)
            {
                Line = sr1.ReadLine();
                LineArray = Line.Split(',');
                UParray = LineArray[23].Split('|');

                PepCounter++;

                    PSM psm = new PSM();
                    if (UParray.Length == 1)
                    {
                        psm.IsUnshared = true;
                    }
                    else
                    {
                        psm.IsUnshared = false;
                    }

                    psm.Scan = Convert.ToInt32(LineArray[0]);

                    psm.FileName = LineArray[1];
                    psm.Sequence = LineArray[2];
                    psm.PeptideLength = psm.Sequence.Length;
                    psm.SequenceArray = new Residue[psm.PeptideLength];
                    psm.PG = LineArray[23];
                    psm.Modifications = LineArray[10];
                    psm.Charge = Convert.ToInt32(LineArray[11]);
                    psm.Evalue = Convert.ToDouble(LineArray[3]);



                    PSMs.Add(psm.FileName, psm);
                    NameBuilder.Clear();
                    NBarray = LineArray[1].Split('.');
                    NameBuilder.Append(textBox2.Text);
                    NameBuilder.Append("\\");
                    NameBuilder.Append(NBarray[0]);
                    NameBuilder.Append("_FTMS_HCD.mgf");

                    if (!InputFiles.Contains(NameBuilder.ToString()))
                    {
```

```
                    InputFiles.Add(NameBuilder.ToString());
                }


            }

            sr1.Close();

            Console.WriteLine("[{0}]  Read a total of {1} PSMs.", DateTime.Now,
PepCounter);

            List<string> NonPhosphoPeptides = new List<string>();
            int xx = 0;
            string[] ModArray;

            Console.WriteLine("");
            Console.WriteLine("[{0}]  Filtering IDs for phosphopeptides...",
DateTime.Now);
            foreach (KeyValuePair<string, PSM> psm in PSMs)
            {
                psm.Value.IsPhosphorylated = false;
                LineArray = psm.Value.Modifications.Split(';');

                xx = 0;
                while (xx < LineArray.Length)
                {
                    ModArray = LineArray[xx].Split(' ');
                    if (ModArray[0].Equals("phosphorylation"))
                    {
                        psm.Value.IsPhosphorylated = true;
                    }

                    xx++;
                }

                if (!psm.Value.IsPhosphorylated)
                {
                    NonPhosphoPeptides.Add(psm.Key);
                }

            }

            //Console.WriteLine(PSMs.Count);

            foreach (string KEY in NonPhosphoPeptides)
            {
                PSMs.Remove(KEY);
            }

            List<string> RemovePSMs = new List<string>();



            foreach (KeyValuePair<string, PSM> psm in PSMs)
            {
                if (psm.Value.IsPhosphorylated)
                {
```

```
            Counter1++;
        }
    }

    int MSMScounter = 0;
    CounterCounter = 0;
    PepCounter = 0;

    Console.WriteLine("");
    Console.WriteLine("[{0}] Begin reading MS/MS spectra...", DateTime.Now);
    foreach (string iFile in InputFiles)
    {
        Console.WriteLine("[{0}]  Opened filepath: {1}", DateTime.Now, iFile);

        StreamReader sr = new StreamReader(iFile);

        Line = sr.ReadLine();

        while (!sr.EndOfStream)
        {
            Line = sr.ReadLine();
            LineArray = Line.Split('=');
            sb.Clear();
            sb.Append(LineArray[1]);
            sb.Append(".dta");

            if (PSMs.ContainsKey(sb.ToString()))
            {
                PSMs[sb.ToString()].MatchedToSpectrum = true;
                PepCounter++;

                Line = sr.ReadLine();
                Line = sr.ReadLine();
                Line = sr.ReadLine();
                Line = sr.ReadLine();
                Line = sr.ReadLine();

                LineArray = Line.Split(' ');

                while ((Convert.ToDouble(LineArray[0]) < 119.0))
                {
                    Peak peak = new Peak();
                    peak.MZ = Convert.ToDouble(LineArray[0]);
                    peak.Intensity = Convert.ToDouble(LineArray[1]);

                    if (!PSMs[sb.ToString()].Spectrum.ContainsKey(peak.MZ))
                    {
                        PSMs[sb.ToString()].Spectrum.Add(peak.MZ, peak);
                    }

                    Line = sr.ReadLine();
                    LineArray = Line.Split(' ');
                }
            }



            while (!Line.Equals("END IONS"))
```

```
                    {
                        Line = sr.ReadLine();
                    }

                    if (!sr.EndOfStream)
                    {
                        Line = sr.ReadLine();
                    }

                    if (!sr.EndOfStream)
                    {
                        Line = sr.ReadLine();
                    }

                    MSMScounter++;
                    CounterCounter++;
                }

                sr.Close();
            }

            Console.WriteLine("[{0}]  A total of {1} MS/MS spectra were parsed.",
DateTime.Now, MSMScounter);
            Console.WriteLine("");
            Console.WriteLine("[{0}]  Matching peptide IDs to phospho proteoforms...",
DateTime.Now);

            StreamReader srr = new StreamReader(INPUT1);
            Line = srr.ReadLine();
            //int DEBUGCOUNTER = 0;

            string PSMKEY;
            string PROTKEY;
            string PROTFORM;
            while (!srr.EndOfStream)
            {
                Line = srr.ReadLine();
                LineArray = Line.Split(',');

                if (PSMs.ContainsKey(LineArray[1]))
                {

                    PSMKEY = LineArray[1];

                    xx = 0;
                    UParray = LineArray[23].Split('|');

                    while (xx < UParray.Length)
                    {
                        if (ProteinIDs.ContainsKey(UParray[xx]))
                        {

                            PROTKEY = UParray[xx];
                            //Console.WriteLine(LineArray[25]);
                            if (LineArray[25].Equals("True"))
                            {

                                PROTFORM = LineArray[27];
```

```
                                if
(!ProteinIDs[PROTKEY].Proteoforms.ContainsKey(PROTFORM))
                                {
                                    Proteoform pf = new Proteoform();

                                    pf.QuantArray = new double[NUMBER_OF_TAGS];
                                    pf.CorrectedQuantArray = new double[NUMBER_OF_TAGS];
                                    pf.NormalizedQuantArray = new double[NUMBER_OF_TAGS];
                                    pf.Name = PROTFORM;

                                    if (Convert.ToInt32(LineArray[22]) == 1)
                                    {
                                        pf.IsUnshared = true;
                                    }
                                    else
                                    {
                                        pf.IsUnshared = false;
                                    }

                                    pf.AllPG = LineArray[23];

                                    pf.PSMs.Add(PSMKEY,PSMs[PSMKEY]);

                                    ProteinIDs[PROTKEY].Proteoforms.Add(pf.Name, pf);
                                }
                                else
                                {

ProteinIDs[PROTKEY].Proteoforms[PROTFORM].PSMs.Add(PSMKEY, PSMs[PSMKEY]);
                                }
                            }
                            else
                            {
                                PROTFORM = LineArray[27];

                                if (!ProteinIDs[PROTKEY].UL.ContainsKey(PROTFORM))
                                {
                                    Proteoform pf = new Proteoform();

                                    pf.QuantArray = new double[NUMBER_OF_TAGS];
                                    pf.CorrectedQuantArray = new double[NUMBER_OF_TAGS];
                                    pf.NormalizedQuantArray = new double[NUMBER_OF_TAGS];
                                    pf.Name = PROTFORM;
                                    pf.AllPG = LineArray[23];

                                    if (Convert.ToInt32(LineArray[22]) == 1)
                                    {
                                        pf.IsUnshared = true;
                                    }
                                    else
                                    {
                                        pf.IsUnshared = false;
                                    }

                                    pf.PSMs.Add(PSMKEY, PSMs[PSMKEY]);

                                    ProteinIDs[PROTKEY].UL.Add(pf.Name, pf);
```

```
                                }
                                else
                                {
                                    if (ProteinIDs.ContainsKey(PROTKEY))
                                    {
                                        ProteinIDs[PROTKEY].UL[PROTFORM].PSMs.Add(PSMKEY,
PSMs[PSMKEY]);
                                    }

                                }
                            }
                        }
                        xx++;
                    }
                }
            }

            srr.Close();

            PSMs.Clear();

            int Matched = 0;
            foreach (KeyValuePair<string, Protein> Prot in ProteinIDs)
            {
                foreach (KeyValuePair<string, Proteoform> pf in Prot.Value.Proteoforms)
                {
                    Matched += pf.Value.PSMs.Count;

                }



            }

            //Get Reporter Intensities
            Console.WriteLine("");
            Console.WriteLine("[{0}]  Quantifying PSMs...", DateTime.Now);
            int x = 0;

            foreach (KeyValuePair<string, Protein> Prot in ProteinIDs)
            {
                foreach (KeyValuePair<string, Proteoform> pf in Prot.Value.Proteoforms)
                {
                    foreach (KeyValuePair<string, PSM> psm in pf.Value.PSMs)
                    {
                        psm.Value.QuantArray = new double[NUMBER_OF_TAGS];

                        x = 0;

                        while (x < NUMBER_OF_TAGS)
                        {
                            foreach (KeyValuePair<double, Peak> peak in
psm.Value.Spectrum)
                            {
                                if (Calculations.IsTag(TAGS[x], TAG_TOL, peak.Value,
psm.Value.QuantArray[x]))
                                {
                                    psm.Value.QuantArray[x] = peak.Value.Intensity;
```

```
                                }
                            }
                            x++;
                        }
                    }
                }

                foreach (KeyValuePair<string, Proteoform> pf in Prot.Value.UL)
                {
                    foreach (KeyValuePair<string, PSM> psm in pf.Value.PSMs)
                    {
                        psm.Value.QuantArray = new double[NUMBER_OF_TAGS];

                        x = 0;

                        while (x < NUMBER_OF_TAGS)
                        {
                            foreach (KeyValuePair<double, Peak> peak in
psm.Value.Spectrum)
                            {
                                if (Calculations.IsTag(TAGS[x], TAG_TOL, peak.Value,
psm.Value.QuantArray[x]))
                                {
                                    psm.Value.QuantArray[x] = peak.Value.Intensity;
                                }
                            }
                            x++;
                        }
                    }
                }
            }

            foreach (KeyValuePair<string, Protein> Prot in ProteinIDs)
            {
                RemovePSMs.Clear();
                foreach (KeyValuePair<string, Proteoform> pf in Prot.Value.UL)
                {
                    RemovePSMs.Clear();
                    foreach (KeyValuePair<string, PSM> psm in pf.Value.PSMs)
                    {
                        x = 0;
                        while (x < NUMBER_OF_TAGS)
                        {
                            if (!(psm.Value.QuantArray[x] > 0.0) &&
!(RemovePSMs.Contains(psm.Key)))
                            {
                                RemovePSMs.Add(psm.Key);
                            }
                            x++;
                        }
                    }

                    foreach (string KEY in RemovePSMs)
                    {
                        pf.Value.PSMs.Remove(KEY);
                    }
                }
```

```
                foreach (KeyValuePair<string, Proteoform> pf in Prot.Value.Proteoforms)
                {
                    RemovePSMs.Clear();
                    foreach (KeyValuePair<string, PSM> psm in pf.Value.PSMs)
                    {
                        x = 0;
                        while (x < NUMBER_OF_TAGS)
                        {
                            if (!(psm.Value.QuantArray[x] > 0.0) &&
!(RemovePSMs.Contains(psm.Key)))
                            {
                                RemovePSMs.Add(psm.Key);
                            }
                            x++;
                        }
                    }

                    foreach (string KEY in RemovePSMs)
                    {
                        pf.Value.PSMs.Remove(KEY);
                    }
                }
            }

            //Do purity corrections
            if (DoCorrections)
            {
                Console.WriteLine("");
                Console.WriteLine("[{0}]  Performing purity corrections...",
DateTime.Now);
                foreach (KeyValuePair<string, Protein> Prot in ProteinIDs)
                {
                    foreach (KeyValuePair<string, Proteoform> pf in
Prot.Value.Proteoforms)
                    {
                        foreach (KeyValuePair<string, PSM> psm in pf.Value.PSMs)
                        {
                            psm.Value.CorrectedQuantArray = new double[NUMBER_OF_TAGS];

                            psm.Value.CorrectedQuantArray =
Calculations.DO_PC(psm.Value.QuantArray, psm.Value.CorrectedQuantArray, PurityMatrix);
                        }
                    }

                    foreach (KeyValuePair<string, Proteoform> pf in Prot.Value.UL)
                    {
                        foreach (KeyValuePair<string, PSM> psm in pf.Value.PSMs)
                        {
                            psm.Value.CorrectedQuantArray = new double[NUMBER_OF_TAGS];

                            psm.Value.CorrectedQuantArray =
Calculations.DO_PC(psm.Value.QuantArray, psm.Value.CorrectedQuantArray, PurityMatrix);
                        }
                    }
                }
            }

            //Do normalization
```

```
            if (DoNormalization)
            {
                Console.WriteLine("");
                Console.WriteLine("[{0}]  Performing channel normalization...",
DateTime.Now);
                double[] ChannelTotals = new double[NUMBER_OF_TAGS];

                foreach (KeyValuePair<string, Protein> Prot in ProteinIDs)
                {
                    foreach (KeyValuePair<string, Proteoform> pf in
Prot.Value.Proteoforms)
                    {
                        foreach (KeyValuePair<string,PSM> psm in pf.Value.PSMs)
                        {
                            psm.Value.NormalizedQuantArray = new double[NUMBER_OF_TAGS];

                            x = 0;
                            while (x < NUMBER_OF_TAGS)
                            {
                                if (DoCorrections)
                                {
                                    ChannelTotals[x] += psm.Value.CorrectedQuantArray[x];
                                }
                                else
                                {
                                    ChannelTotals[x] += psm.Value.QuantArray[x];
                                }
                                x++;
                            }
                        }
                    }

                    foreach (KeyValuePair<string, Proteoform> pf in Prot.Value.UL)
                    {
                        foreach (KeyValuePair<string, PSM> psm in pf.Value.PSMs)
                        {
                            psm.Value.NormalizedQuantArray = new double[NUMBER_OF_TAGS];

                            x = 0;
                            while (x < NUMBER_OF_TAGS)
                            {
                                if (DoCorrections)
                                {
                                    ChannelTotals[x] += psm.Value.CorrectedQuantArray[x];
                                }
                                else
                                {
                                    ChannelTotals[x] += psm.Value.QuantArray[x];
                                }
                                x++;
                            }
                        }
                    }
                }

                double[] NormalizationFactors = new double[NUMBER_OF_TAGS];
```

```
                NormalizationFactors =
Calculations.GetNormalizationFactors(ChannelTotals);

                foreach (KeyValuePair<string, Protein> Prot in ProteinIDs)
                {
                    foreach (KeyValuePair<string, Proteoform> pf in
Prot.Value.Proteoforms)
                    {
                        foreach (KeyValuePair<string, PSM> psm in pf.Value.PSMs)
                        {
                            x = 0;

                            while (x < NUMBER_OF_TAGS)
                            {
                                if (DoCorrections)
                                {
                                    psm.Value.NormalizedQuantArray[x] =
(NormalizationFactors[x] * psm.Value.CorrectedQuantArray[x]);
                                }
                                else
                                {
                                    psm.Value.NormalizedQuantArray[x] =
(NormalizationFactors[x] * psm.Value.QuantArray[x]);
                                }
                                x++;
                            }
                        }
                    }

                    foreach (KeyValuePair<string, Proteoform> pf in Prot.Value.UL)
                    {
                        foreach (KeyValuePair<string, PSM> psm in pf.Value.PSMs)
                        {
                            x = 0;

                            while (x < NUMBER_OF_TAGS)
                            {
                                if (DoCorrections)
                                {
                                    psm.Value.NormalizedQuantArray[x] =
(NormalizationFactors[x] * psm.Value.CorrectedQuantArray[x]);
                                }
                                else
                                {
                                    psm.Value.NormalizedQuantArray[x] =
(NormalizationFactors[x] * psm.Value.QuantArray[x]);
                                }
                                x++;
                            }
                        }
                    }
                }
            }

            //Quantify proteoforms
            Console.WriteLine("");
            Console.WriteLine("[{0}]  Quantifying proteoforms...", DateTime.Now);
            foreach (KeyValuePair<string, Protein> Prot in ProteinIDs)
```

```
            {
                foreach (KeyValuePair<string, Proteoform> pf in Prot.Value.Proteoforms)
                {
                    if (pf.Value.IsUnshared || !pf.Value.IsUnshared)
                    {
                        foreach (KeyValuePair<string, PSM> psm in pf.Value.PSMs)
                        {
                            if (psm.Value.IsUnshared || !psm.Value.IsUnshared)
                            {
                                x = 0;

                                while (x < NUMBER_OF_TAGS)
                                {
                                    pf.Value.QuantArray[x] += psm.Value.QuantArray[x];
                                    if (DoCorrections)
                                    {
                                        pf.Value.CorrectedQuantArray[x] +=
psm.Value.CorrectedQuantArray[x];
                                    }
                                    if (DoNormalization)
                                    {
                                        pf.Value.NormalizedQuantArray[x] +=
psm.Value.NormalizedQuantArray[x];
                                    }
                                    x++;
                                }
                            }

                        }
                    }


                }

                foreach (KeyValuePair<string, Proteoform> pf in Prot.Value.UL)
                {
                    if (pf.Value.IsUnshared || !pf.Value.IsUnshared)
                    {
                        foreach (KeyValuePair<string, PSM> psm in pf.Value.PSMs)
                        {
                            if (psm.Value.IsUnshared || !psm.Value.IsUnshared)
                            {
                                x = 0;

                                while (x < NUMBER_OF_TAGS)
                                {
                                    pf.Value.QuantArray[x] += psm.Value.QuantArray[x];
                                    if (DoCorrections)
                                    {
                                        pf.Value.CorrectedQuantArray[x] +=
psm.Value.CorrectedQuantArray[x];
                                    }
                                    if (DoNormalization)
                                    {
                                        pf.Value.NormalizedQuantArray[x] +=
psm.Value.NormalizedQuantArray[x];
                                    }
                                    x++;
```

```
                            }
                        }

                    }
                }


            }

            /*foreach (KeyValuePair<string, PSM> psm in
Prot.Value.UnlocalizedProteoforms)
            {
                if (psm.Value.IsUnshared)
                {
                    x = 0;

                    while (x < NUMBER_OF_TAGS)
                    {
                        Prot.Value.QuantArray[x] += psm.Value.QuantArray[x];
                        if (DoCorrections)
                        {
                            Prot.Value.CorrectedQuantArray[x] +=
psm.Value.CorrectedQuantArray[x];
                        }
                        if (DoNormalization)
                        {
                            Prot.Value.NormalizedQuantArray[x] +=
psm.Value.NormalizedQuantArray[x];
                        }
                        x++;
                    }
                }
            }*/
        }

        // Do mean intensities
        double MeanFactor = 0.0;
        if (DoMeanIntensities)
        {
            foreach(KeyValuePair<string,Protein> Prot in ProteinIDs)
            {
                foreach (KeyValuePair<string, Proteoform> pf in Prot.Value.UL)
                {
                    x = 0;
                    pf.Value.UnsharedNumber = 0;
                    while (x < NUMBER_OF_TAGS)
                    {
                        pf.Value.QuantArray[x] = 0.0;
                        if (DoCorrections)
                        {
                            pf.Value.CorrectedQuantArray[x] = 0.0;
                        }
                        if (DoNormalization)
                        {
                            pf.Value.NormalizedQuantArray[x] = 0.0;
                        }
                        x++;
                    }
```

```
                        foreach (KeyValuePair<string, PSM> psm in pf.Value.PSMs)
                        {

                            if (psm.Value.IsUnshared || !psm.Value.IsUnshared)
                            {
                                pf.Value.UnsharedNumber++;
                                MeanFactor =
Calculations.GetMeanIntensityFactor(psm.Value.QuantArray);

                                x = 0;
                                while (x < NUMBER_OF_TAGS)
                                {
                                    psm.Value.QuantArray[x] = MeanFactor *
psm.Value.QuantArray[x];
                                    x++;
                                }

                                if (DoCorrections)
                                {
                                    MeanFactor =
Calculations.GetMeanIntensityFactor(psm.Value.CorrectedQuantArray);

                                    x = 0;
                                    while (x < NUMBER_OF_TAGS)
                                    {
                                        psm.Value.CorrectedQuantArray[x] = MeanFactor *
psm.Value.CorrectedQuantArray[x];
                                        x++;
                                    }
                                }
                                if (DoNormalization)
                                {
                                    MeanFactor =
Calculations.GetMeanIntensityFactor(psm.Value.NormalizedQuantArray);

                                    x = 0;
                                    while (x < NUMBER_OF_TAGS)
                                    {
                                        psm.Value.NormalizedQuantArray[x] = MeanFactor *
psm.Value.NormalizedQuantArray[x];
                                        x++;
                                    }
                                }
                            }

                        }

                        foreach (KeyValuePair<string, PSM> psm in pf.Value.PSMs)
                        {
                            if (psm.Value.IsUnshared || !psm.Value.IsUnshared)
                            {
                                x = 0;
                                while (x < NUMBER_OF_TAGS)
                                {
                                    pf.Value.QuantArray[x] += psm.Value.QuantArray[x];
                                    if (DoCorrections)
                                    {
```

```
                                    pf.Value.CorrectedQuantArray[x] +=
psm.Value.CorrectedQuantArray[x];
                                }
                                if (DoNormalization)
                                {
                                    pf.Value.NormalizedQuantArray[x] +=
psm.Value.NormalizedQuantArray[x];
                                }
                                x++;
                            }
                        }

                    }

                    x = 0;
                    while (x < NUMBER_OF_TAGS)
                    {
                        pf.Value.QuantArray[x] = pf.Value.QuantArray[x] /
pf.Value.PSMs.Count;
                        if (DoCorrections)
                        {
                            pf.Value.CorrectedQuantArray[x] =
pf.Value.CorrectedQuantArray[x] / pf.Value.PSMs.Count;
                        }
                        if (DoNormalization)
                        {
                            pf.Value.NormalizedQuantArray[x] =
pf.Value.NormalizedQuantArray[x] / pf.Value.PSMs.Count;
                        }
                        x++;
                    }
                }

                foreach (KeyValuePair<string, Proteoform> pf in
Prot.Value.Proteoforms)
                {
                    x = 0;
                    pf.Value.UnsharedNumber = 0;
                    while (x < NUMBER_OF_TAGS)
                    {
                        pf.Value.QuantArray[x] = 0.0;
                        if (DoCorrections)
                        {
                            pf.Value.CorrectedQuantArray[x] = 0.0;
                        }
                        if (DoNormalization)
                        {
                            pf.Value.NormalizedQuantArray[x] = 0.0;
                        }
                        x++;
                    }

                    foreach (KeyValuePair<string, PSM> psm in pf.Value.PSMs)
                    {

                        if (psm.Value.IsUnshared || !psm.Value.IsUnshared)
                        {
                            pf.Value.UnsharedNumber++;
```

```
                                MeanFactor =
Calculations.GetMeanIntensityFactor(psm.Value.QuantArray);

                                x = 0;
                                while (x < NUMBER_OF_TAGS)
                                {
                                    psm.Value.QuantArray[x] = MeanFactor *
psm.Value.QuantArray[x];
                                    x++;
                                }

                                if (DoCorrections)
                                {
                                    MeanFactor =
Calculations.GetMeanIntensityFactor(psm.Value.CorrectedQuantArray);

                                    x = 0;
                                    while (x < NUMBER_OF_TAGS)
                                    {
                                        psm.Value.CorrectedQuantArray[x] = MeanFactor *
psm.Value.CorrectedQuantArray[x];
                                        x++;
                                    }
                                }
                                if (DoNormalization)
                                {
                                    MeanFactor =
Calculations.GetMeanIntensityFactor(psm.Value.NormalizedQuantArray);

                                    x = 0;
                                    while (x < NUMBER_OF_TAGS)
                                    {
                                        psm.Value.NormalizedQuantArray[x] = MeanFactor *
psm.Value.NormalizedQuantArray[x];
                                        x++;
                                    }
                                }
                            }

                        }

                        foreach (KeyValuePair<string, PSM> psm in pf.Value.PSMs)
                        {
                            if (psm.Value.IsUnshared || !psm.Value.IsUnshared)
                            {
                                x = 0;
                                while (x < NUMBER_OF_TAGS)
                                {
                                    pf.Value.QuantArray[x] += psm.Value.QuantArray[x];
                                    if (DoCorrections)
                                    {
                                        pf.Value.CorrectedQuantArray[x] +=
psm.Value.CorrectedQuantArray[x];
                                    }
                                    if (DoNormalization)
                                    {
                                        pf.Value.NormalizedQuantArray[x] +=
psm.Value.NormalizedQuantArray[x];
```

```
                                    }
                                    x++;
                                }
                            }

                        }

                        x = 0;
                        while (x < NUMBER_OF_TAGS)
                        {
                            pf.Value.QuantArray[x] = pf.Value.QuantArray[x] /
pf.Value.PSMs.Count;
                            if (DoCorrections)
                            {
                                pf.Value.CorrectedQuantArray[x] =
pf.Value.CorrectedQuantArray[x] / pf.Value.PSMs.Count;
                            }
                            if (DoNormalization)
                            {
                                pf.Value.NormalizedQuantArray[x] =
pf.Value.NormalizedQuantArray[x] / pf.Value.PSMs.Count;
                            }
                            x++;
                        }
                    }
                }
            }

            //Outputs
            Console.WriteLine("");
            Console.WriteLine("[{0}]  Writing outputs...", DateTime.Now);
            List<string> OutputLines = new List<string>();

            sb.Clear();
            sb.Append("Representative Uniprot ID,Protein Group,All Protein Groups,Number
of PG,DefLine,Proteoform,Quantified PSMs,Raw 115a,Raw 115b,Raw 116a,Raw 116b,Raw 116c,Raw
117a,Raw 117b,Raw 117c,Raw 118a,Raw 118b,Raw 118c,Raw 118d");
            if (DoCorrections)
            {
                sb.Append(",Corrected 115a,Corrected 115b,Corrected 116a,Corrected
116b,Corrected 116c,Corrected 117a,Corrected 117b,Corrected 117c,Corrected 118a,Corrected
118b,Corrected 118c,Corrected 118d");
            }
            if (DoNormalization)
            {
                sb.Append(",Normalized 115a,Normalized 115b,Normalized 116a,Normalized
116b,Normalized 116c,Normalized 117a,Normalized 117b,Normalized 117c,Normalized
118a,Normalized 118b,Normalized 118c,Normalized 118d");
            }

            OutputLines.Add(sb.ToString());
            sb.Clear();

            foreach(KeyValuePair<string,Protein> Prot in ProteinIDs)
            {
                foreach (KeyValuePair<string, Proteoform> pf in Prot.Value.Proteoforms)
                {
                    if (pf.Value.PSMs.Count > 0)
```

```
        {
            sb.Clear();
            sb.Append(Prot.Value.RepUniprotID);
            sb.Append(",");
            sb.Append(Prot.Value.PG);
            sb.Append(",");
            sb.Append(pf.Value.AllPG);
            sb.Append(",");
            LineArray = pf.Value.AllPG.Split('|');
            sb.Append(LineArray.Length);
            sb.Append(",");
            sb.Append(Prot.Value.DefLine);
            sb.Append(",");
            sb.Append(pf.Value.Name);
            sb.Append(",");
            sb.Append(pf.Value.PSMs.Count);
            sb.Append(",");

            x = 0;

            while (x < NUMBER_OF_TAGS)
            {
                sb.Append(pf.Value.QuantArray[x]);
                sb.Append(",");
                x++;
            }

            if (DoCorrections)
            {
                x = 0;
                while (x < NUMBER_OF_TAGS)
                {
                    sb.Append(pf.Value.CorrectedQuantArray[x]);
                    sb.Append(",");
                    x++;
                }
            }

            if (DoNormalization)
            {
                x = 0;
                while (x < NUMBER_OF_TAGS)
                {
                    sb.Append(pf.Value.NormalizedQuantArray[x]);
                    sb.Append(",");
                    x++;
                }
            }

            OutputLines.Add(sb.ToString());
        }

    }
}

File.WriteAllLines(OUTPUT2, OutputLines);

OutputLines.Clear();
```

```
            sb.Clear();
            sb.Append("Representative Uniprot ID,Protein Group,All Protein Groups,Number
of PG,DefLine,Proteoform,Quantified PSMs,Raw 115a,Raw 115b,Raw 116a,Raw 116b,Raw 116c,Raw
117a,Raw 117b,Raw 117c,Raw 118a,Raw 118b,Raw 118c,Raw 118d");
            if (DoCorrections)
            {
                sb.Append(",Corrected 115a,Corrected 115b,Corrected 116a,Corrected
116b,Corrected 116c,Corrected 117a,Corrected 117b,Corrected 117c,Corrected 118a,Corrected
118b,Corrected 118c,Corrected 118d");
            }
            if (DoNormalization)
            {
                sb.Append(",Normalized 115a,Normalized 115b,Normalized 116a,Normalized
116b,Normalized 116c,Normalized 117a,Normalized 117b,Normalized 117c,Normalized
118a,Normalized 118b,Normalized 118c,Normalized 118d");
            }

            OutputLines.Add(sb.ToString());
            sb.Clear();

            foreach (KeyValuePair<string, Protein> Prot in ProteinIDs)
            {
                foreach (KeyValuePair<string, Proteoform> pf in Prot.Value.UL)
                {
                    if (pf.Value.PSMs.Count > 0)
                    {
                        sb.Clear();
                        sb.Append(Prot.Value.RepUniprotID);
                        sb.Append(",");
                        sb.Append(Prot.Value.PG);
                        sb.Append(",");
                        sb.Append(pf.Value.AllPG);
                        sb.Append(",");
                        LineArray = pf.Value.AllPG.Split('|');
                        sb.Append(LineArray.Length);
                        sb.Append(",");
                        sb.Append(Prot.Value.DefLine);
                        sb.Append(",");
                        sb.Append(pf.Value.Name);
                        sb.Append(",");
                        sb.Append(pf.Value.PSMs.Count);
                        sb.Append(",");

                        x = 0;

                        while (x < NUMBER_OF_TAGS)
                        {
                            sb.Append(pf.Value.QuantArray[x]);
                            sb.Append(",");
                            x++;
                        }

                        if (DoCorrections)
                        {
                            x = 0;
                            while (x < NUMBER_OF_TAGS)
                            {
```

```
                            sb.Append(pf.Value.CorrectedQuantArray[x]);
                            sb.Append(",");
                            x++;
                        }
                    }

                    if (DoNormalization)
                    {
                        x = 0;
                        while (x < NUMBER_OF_TAGS)
                        {
                            sb.Append(pf.Value.NormalizedQuantArray[x]);
                            sb.Append(",");
                            x++;
                        }
                    }

                    OutputLines.Add(sb.ToString());
                }

            }
        }

        File.WriteAllLines(OUTPUT1, OutputLines);

        OutputLines.Clear();
        sb.Clear();

        OutputLines.Add("Protein Group,Representative UniprotID,Genes,Defline");

        foreach (KeyValuePair<string, Protein> Prot in ProteinIDs)
        {
            sb.Clear();

            sb.Append(Prot.Value.PG);
            sb.Append(",");
            sb.Append(Prot.Value.RepUniprotID);
            sb.Append(",");

            foreach (string G in Prot.Value.Genes)
            {
                sb.Append(G);
                sb.Append(",");
            }
            sb.Append(Prot.Value.DefLine);

            OutputLines.Add(sb.ToString());
        }

        File.WriteAllLines(OUTPUT3, OutputLines);

        OutputLines.Clear();

        sb.Clear();
        sb.Append("Scan Number,Scan Name,Sequence,Protein Group,Representative
UniprotID,Unshared,DefLine,Raw 115a,Raw 115b,Raw 116a,Raw 116b,Raw 116c,Raw 117a,Raw
117b,Raw 117c,Raw 118a,Raw 118b,Raw 118c,Raw 118d");
        if (DoCorrections)
```

```
            {
                sb.Append(",Corrected 115a,Corrected 115b,Corrected 116a,Corrected
116b,Corrected 116c,Corrected 117a,Corrected 117b,Corrected 117c,Corrected 118a,Corrected
118b,Corrected 118c,Corrected 118d");
            }
            if (DoNormalization)
            {
                sb.Append(",Normalized 115a,Normalized 115b,Normalized 116a,Normalized
116b,Normalized 116c,Normalized 117a,Normalized 117b,Normalized 117c,Normalized
118a,Normalized 118b,Normalized 118c,Normalized 118d");
            }

            OutputLines.Add(sb.ToString());
            sb.Clear();

            foreach (KeyValuePair<string, Protein> Prot in ProteinIDs)
            {
                foreach (KeyValuePair<string, Proteoform> pf in Prot.Value.Proteoforms)
                {
                    foreach (KeyValuePair<string, PSM> psm in pf.Value.PSMs)
                    {
                        sb.Clear();
                        sb.Append(psm.Value.Scan);
                        sb.Append(",");
                        sb.Append(psm.Key);
                        sb.Append(",");
                        sb.Append(psm.Value.Sequence);
                        sb.Append(",");
                        sb.Append(Prot.Key);
                        sb.Append(",");
                        sb.Append(Prot.Value.RepUniprotID);
                        sb.Append(",");

                        if (psm.Value.IsUnshared)
                        {
                            sb.Append("TRUE");
                        }
                        else
                        {
                            sb.Append("FALSE");
                        }
                        sb.Append(",");
                        sb.Append(Prot.Value.DefLine);
                        sb.Append(",");

                        x = 0;

                        while (x < NUMBER_OF_TAGS)
                        {
                            sb.Append(psm.Value.QuantArray[x]);
                            sb.Append(",");
                            x++;
                        }

                        if (DoCorrections)
                        {
                            x = 0;
                            while (x < NUMBER_OF_TAGS)
```

```
            {
                sb.Append(psm.Value.CorrectedQuantArray[x]);
                sb.Append(",");
                x++;
            }
        }

        if (DoNormalization)
        {
            x = 0;
            while (x < NUMBER_OF_TAGS)
            {
                sb.Append(psm.Value.NormalizedQuantArray[x]);
                sb.Append(",");
                x++;
            }
        }

        OutputLines.Add(sb.ToString());
    }
}

foreach (KeyValuePair<string, Proteoform> pf in Prot.Value.UL)
{
    foreach (KeyValuePair<string, PSM> psm in pf.Value.PSMs)
    {
        sb.Clear();
        sb.Append(psm.Value.Scan);
        sb.Append(",");
        sb.Append(psm.Key);
        sb.Append(",");
        sb.Append(psm.Value.Sequence);
        sb.Append(",");
        sb.Append(Prot.Key);
        sb.Append(",");
        sb.Append(Prot.Value.RepUniprotID);
        sb.Append(",");

        if (psm.Value.IsUnshared)
        {
            sb.Append("TRUE");
        }
        else
        {
            sb.Append("FALSE");
        }
        sb.Append(",");
        sb.Append(Prot.Value.DefLine);
        sb.Append(",");

        x = 0;

        while (x < NUMBER_OF_TAGS)
        {
            sb.Append(psm.Value.QuantArray[x]);
            sb.Append(",");
            x++;
        }
```

```
                if (DoCorrections)
                {
                    x = 0;
                    while (x < NUMBER_OF_TAGS)
                    {
                        sb.Append(psm.Value.CorrectedQuantArray[x]);
                        sb.Append(",");
                        x++;
                    }
                }

                if (DoNormalization)
                {
                    x = 0;
                    while (x < NUMBER_OF_TAGS)
                    {
                        sb.Append(psm.Value.NormalizedQuantArray[x]);
                        sb.Append(",");
                        x++;
                    }
                }

                OutputLines.Add(sb.ToString());
            }
        }
    }

    File.WriteAllLines(OUTPUT4, OutputLines);

    OutputLines.Clear();

    ProteinIDs.Clear();

    Console.WriteLine("");
    Console.WriteLine("[{0}]  Done!!!", DateTime.Now);

    MessageBox.Show("Done!");
}

private void button3_Click(object sender, EventArgs e)
{
    string INPUT = CleanInTB.Text;
    StringBuilder NameBuilder = new StringBuilder();
    string[] OutArray;
    OutArray = INPUT.Split('.');
    OutArray = OutArray[0].Split('\\');
    int N = 0;
    while (N < (OutArray.Length - 1))
    {
        NameBuilder.Append(OutArray[N]);
        NameBuilder.Append("\\");
        N++;
    }

    NameBuilder.Append("CLEAN_");
    NameBuilder.Append(OutArray[OutArray.Length - 1]);
    NameBuilder.Append(".txt");
```

```
            //string OUTPUT = NameBuilder.ToString();
            string OUTPUT = INPUT;
            NameBuilder.Clear();

            string Line;
            string[] LineArray;

            int Counter = 0;
            int CounterCounter = 0;

            List<string> OutputLines = new List<string>();
            StringBuilder sb = new StringBuilder();
            StreamReader sr = new StreamReader(INPUT);

            Line = sr.ReadLine();
            double MIN_RANGE = 0.0;
            double MAX_RANGE = 0.0;
            double MINMIN = Convert.ToDouble(MinCleanBox.Text);
            double MAXMAX = Convert.ToDouble(MaxCleanBox.Text);
            double NeutralLossWindow = Convert.ToDouble(NL_RangeBox.Text);
            string[] RangeArray;
            Console.WriteLine("[{0}]  Begin DiLeu reporter clean-up...", DateTime.Now);
            while (!sr.EndOfStream)
            {
                RangeArray = Line.Split('.');

                OutputLines.Add(Line);
                Line = sr.ReadLine();
                OutputLines.Add(Line);
                Line = sr.ReadLine();
                OutputLines.Add(Line);

                RangeArray = Line.Split(' ');
                MIN_RANGE = (Convert.ToDouble(RangeArray[0]) - NeutralLossWindow);
                MAX_RANGE = (Convert.ToDouble(RangeArray[0]) + NeutralLossWindow);

                Line = sr.ReadLine();
                LineArray = Line.Split(' ');

                while (LineArray.Length > 1 && !sr.EndOfStream)
                {
                    if (!((Convert.ToDouble(LineArray[1]) > MINMIN) &&
(Convert.ToDouble(LineArray[1]) < MAXMAX)))
                    {

                        OutputLines.Add(Line);
                        if (!((Convert.ToDouble(LineArray[1]) >= MIN_RANGE) &&
(Convert.ToDouble(LineArray[1]) <= MAX_RANGE)))
                        {
                            OutputLines.Add(Line);
                        }
                    }

                    if (!sr.EndOfStream)
                    {
                        Line = sr.ReadLine();
                        LineArray = Line.Split(' ');
                    }
```

```
                }

                OutputLines.Add(Line);
                if (!sr.EndOfStream)
                {
                    Line = sr.ReadLine();
                    OutputLines.Add(Line);
                }

                if (!sr.EndOfStream)
                {
                    Line = sr.ReadLine();
                }

                Counter++;
                CounterCounter++;

                if (CounterCounter == 5000)
                {
                    Console.WriteLine("[{0}]  Cleaned {1} MS/MS spectra.", DateTime.Now,
Counter);

                    CounterCounter = 0;
                }
            }

            sr.Close();

            Console.WriteLine("[{0}]  Cleaned a total of {1} MS/MS spectra.",
DateTime.Now, Counter);
            Console.WriteLine("");

            Console.WriteLine("[{0}]  Writing new OMSSA .txt file...", DateTime.Now);


            File.WriteAllLines(OUTPUT, OutputLines);

            OutputLines.Clear();
            Console.WriteLine("");
            Console.WriteLine("[{0}]  DONE!!!",DateTime.Now);

            //MessageBox.Show("Done!!!");
        }

        private void button4_Click(object sender, EventArgs e)
        {
            DialogResult result = openCleanInput.ShowDialog();


            CleanInTB.Text = openCleanInput.FileName;
        }

        private void button5_Click(object sender, EventArgs e)
        {
            DialogResult result = ProteinInDialog.ShowDialog();


            ProteinInBox.Text = ProteinInDialog.FileName;
        }
```

```
private void button6_Click(object sender, EventArgs e)
{
    DialogResult result = PeptideInDialog.ShowDialog();


    PeptideInBox.Text = PeptideInDialog.FileName;
}

private void button8_Click(object sender, EventArgs e)
{
    DialogResult result = MatrixInDialog.ShowDialog();


    PurityMatrixInBox.Text = MatrixInDialog.FileName;
}

private void button7_Click(object sender, EventArgs e)
{
    DialogResult result = OMSSAdirectoryDialog.ShowDialog();

    TxtDirBox.Text = OMSSAdirectoryDialog.SelectedPath;
}

private void checkBox3_CheckedChanged(object sender, EventArgs e)
{

}

private void button9_Click(object sender, EventArgs e)
{
    DialogResult result = PhosphoProteinDialog.ShowDialog();


    PhosphoproteinInBox.Text = PhosphoProteinDialog.FileName;
}

private void button10_Click(object sender, EventArgs e)
{
    DialogResult result = PhosphopsmDialog.ShowDialog();


    PhosphoPSMinBox.Text = PhosphopsmDialog.FileName;
}

private void button12_Click(object sender, EventArgs e)
{
    DialogResult result = OMSSAdirectoryDialog2.ShowDialog();

    textBox2.Text = OMSSAdirectoryDialog2.SelectedPath;
}

private void button11_Click(object sender, EventArgs e)
{
    DialogResult result = MatrixInDialog2.ShowDialog();


    textBox1.Text = MatrixInDialog2.FileName;
```

```
        }

        private void label11_Click(object sender, EventArgs e)
        {

        }

        private void Form1_Load(object sender, EventArgs e)
        {

        }
    }
}
```

Appendix IV

Phosphoproteoforms with significant differential regulation after pulsed radiofrequency of spared-nerve injury rats

This follow-up SNI investigation aimed to see how pulsed radiofrequency (**PRF**),[1] a clinical treatment for neuropathic pain, modulated phosphoproteins. Collaborators from the Millennium Pain Center in Bloomington, Illinois had recently shown that PRF alters expression of pain-related genes,[2] and now the interest was in how PRF might affect the functional phosphorylation states of signaling proteins. Briefly, two sample groups (SNI rats treated with PRF versus SNI rats with no treatment), with three biological replicates each, were analyzed with a similar LC-MS phosphoproteomic protocol to that outlined in **Chapter 7**. Instead of 12-plex DiLeu, 6-plex TMT (Thermo Fisher Scientific) was used. **Table 1** lists all phosphoproteoforms with significant differential regulation between SNI rats that did receive and did not receive PRF treatment (PRF+ and PRF-, respectively). Significance criteria required the Log_2(PRF+/PRF-) ratio to be less than -0.50 or greater than 0.50, as well as a Student's t-test p-value to be less than or equal to 0.05. Special attention will be given to the four PRF up-regulated phosphoproteoforms of neurofilament medium polypeptide (**Nefm**) in follow-up experiments. Nefm is a biomarker for neuronal damage and is thought to be functionally modulated through many phosphorylated residues.[3]

References

(1) Van Zundert, J., et al. *Pain* **2007** *127*, 173-182.

(2) Vallejo, R., et al. *Pain Physician* **2013** *16*, E601-E613.

(3) Xu, Z. S., Liu, W. S., and Willard, M. B. *J. Biol. Chem.* **1992** *267*, 4467-4471.

Table 1. All phosphoproteoforms with PRF-treated (PRF+) and untreated (PRF-) comparisons meeting significance criteria (Log_2(PRF+/PRF-) ≤ -0.50 or Log_2(PRF+/PRF-) ≥ 0.50, p-value < 0.05).

Uniprot ID	Gene	Phosphoproteoform	Log_2(PRF+/PRF-)	P-value
M0R8Q3	Nefm	S549,S550	0.92	1.64E-02
D3ZD79	Aim1	S22	0.92	1.26E-02
Q3KRF2	Hdlbp	T940	0.89	1.57E-03
F1LR60	Apba1	S245	0.89	2.29E-02
Q8VHI8	Bnip1	S120	0.84	6.34E-03
G3V7U2	Map1a	S1029	0.79	3.70E-02
Q9QZC5	Grb7	S396,T399	0.75	3.36E-02
D3ZLD5	Golga3	S1439	0.74	7.51E-03
Q5U2S0	Rsrp1	S282	0.74	1.01E-03
Q5U2N2	Usp14	S143	0.72	4.98E-03
M0R8Q3	Nefm	S544	0.71	2.29E-02
M0R618	Etl4	S1074	0.71	4.42E-02
M0R8Q3	Nefm	S750	0.70	3.35E-02
D4AC12	Anks1a	S878	0.70	9.98E-03
F1LNK0	Map2	S449	0.70	2.19E-02
F1LRL9	Map1b	S1009	0.69	1.33E-02
B1WBN3	Bckdha	S348	0.68	2.02E-02
Q5XIM8	Lpin1	S285	0.68	1.49E-02
Q8VHK7	Hdgf	S132	0.67	5.44E-03
F1LQN3	Rtn4	S428	0.67	3.34E-02
Q5BJT9	Ckmt1b	T356	0.67	2.04E-02
A0JPM9	Eif3j	T110	0.66	1.04E-02
D3ZJ92	Prpf40a	S884	0.66	3.25E-02
M0R8Q3	Nefm	S550	0.65	2.82E-02
P11980	Pkm	S127	0.62	4.60E-02
F1LM42	Ank2	S3856	0.61	3.41E-02
F1M049	Atxn2	S188	0.60	4.70E-02
D4AE00	Ap3b2	S282	0.60	5.12E-03
G3V913	Hspb1	S13	0.60	3.65E-03
Q3SWT4	Iws1	S295	0.60	1.92E-02
D3ZAH8	Myo9b	S1952	0.59	4.97E-02
D4A9D8	Osbp	S115	0.59	2.78E-02
G3V7X2	Scg2	S270	0.59	4.82E-02
G3V7U2	Map1a	T1041	0.58	3.22E-02

B1WBN3	Bckdha	Y354	0.57	2.52E-02
P82995	Hsp90aa1	S231	0.56	2.96E-02
P34058	Hsp90ab1	S226	0.55	3.55E-02
M0RDJ7	Tceal5	S23	0.54	2.15E-02
F1LQ89	Map3k12	T43	0.53	8.50E-05
P07936	Gap43	T138,T139	0.53	2.37E-02
G3V7U2	Map1a	S1197	0.53	4.89E-02
D4AE00	Ap3b2	S272	0.53	1.25E-02
P50232	Syt4	S135	0.52	1.74E-02
P49192	Cartpt	S48	0.52	1.19E-03
A1L1L6	Rhot1	S338	0.52	4.14E-02
O35314	Chgb	S190	0.52	4.31E-02
D3ZWQ0	Prrt3	S866	0.51	2.57E-02
Q63633	Slc12a5	S1045	0.50	2.00E-02
G3V6C9	Cdc42bpa	S1629	-0.50	4.53E-02
G3V6C9	Cdc42bpa	S1629	-0.50	4.53E-02
M0RD40	Sik3	S676	-0.50	3.52E-02
Q5FVC2	Arhgef2	S781	-0.51	1.50E-02
P11345	Raf1	S43	-0.51	2.63E-02
F1LNK0	Map2	T1261	-0.53	3.19E-02
F1LMQ1	Myo9a	T2338	-0.53	1.88E-02
F1LRJ2	Srrm2	S796	-0.54	3.63E-02
P59649	Fxyd7	S56,S58	-0.56	9.42E-03
F1M853	Rrbp1	S135	-0.56	4.17E-02
P47942	Dpysl2	S542	-0.56	3.01E-02
F1LXQ7	Arhgap21	S1680	-0.56	1.66E-02
D4A9L2	Srsf1	S234,Y237	-0.58	3.05E-02
D4ADD3	Hecw2	T453	-0.58	1.43E-02
D4A720	Srsf7	S233	-0.58	2.43E-02
Q5XIT1	Mapre3	T161	-0.59	1.92E-02
Q91ZQ0	Vmp1	T405	-0.59	1.68E-03
Q5D023	Dync1li2	S383	-0.61	2.15E-02
D4ADD7	Glrx5	S151	-0.61	3.16E-02
Q5M964	Fh	T82	-0.61	2.57E-03
Q9WU70	Stxbp5	T785	-0.62	4.92E-03
B0BMS8	Myl9	S20	-0.63	4.88E-03
Q5QD51	Akap12	S274	-0.65	4.49E-02
Q99PJ8	Zfp483	S381	-0.66	2.80E-02
Q62844	Fyn	S21	-0.66	9.81E-03
F1LSW7	Rpl14	S139	-0.68	4.50E-02
F1M8H6	Pip5k1c	S452	-0.68	5.72E-03

A0A096MJT6	RGD1307100	T3892	-0.72	1.68E-02
P04636	Mdh2	T235	-0.75	3.85E-02
D4A9L2	Srsf1	S199	-0.78	2.87E-02
F1LQN3	Rtn4	T487	-0.79	4.75E-02
D3ZTK0	Ttc9b	S226	-0.80	4.21E-02
G3V7U2	Map1a	S2602	-0.82	3.44E-02
F7FC29	Bicd2	T569	-0.83	3.60E-02
Q5BJT0	Arglu1	S75	-0.98	4.95E-02
A6Y7S3	Kctd8	S331	-1.06	4.72E-02
P05197	Eef2	T57	-1.50	2.58E-02
P05197	Eef2	T54	-1.54	3.73E-02

Appendix V

Proteins with significant differential regulation after spinal cord stimulation of spared-nerve injury rats

This follow-up investigation to **Chapter 6** aimed to see how direct electrical spinal cord stimulation (**ScS**),[1] a clinical treatment for neuropathic pain, affected global protein expression in the spared-nerve injury (**SNI**)[2] model of neuropathic pain.. Two groups, each consisting of five biological replicates, were compared: SNI rats who received ScS treatment (ScS+), and SNI rats who received no treatment (ScS-). Liquid chromatography-mass spectrometry analysis was performed similarly to that described in **Chapter 6**. In **Table 1**, all protein abundances found to be significantly differentially regulated by ScS treatment are listed. Criteria for significance required the Log_2(ScS+/ScS-) ratio to be less than -0.50 or greater than 0.50, as well as a Student t-test's p-value to be less than 0.05.

References

(1) Kumar, K., et al. *Pain* **2007** *132*, 179-188.

(2) Decosterd, I., and Woolf, C. J. *Pain* **2000** *87*, 149-158.

Table 1. All proteins with ScS-treated (ScS+) and untreated (ScS-) comparisons meeting significance criteria (Log_2(ScS+/ScS-) ≤ -0.50 or Log_2(ScS+/ScS-) ≥ 0.50, p-value < 0.05).

Uniprot ID	Gene	Log_2(ScS+/ScS-)	p-value
M0RDF2	-	1.74	4.58E-02
G3V8S9	Camp	1.74	4.26E-02
P30152	Lcn2	1.65	4.41E-02
G3V9Q7	Ctsg	1.63	4.72E-02
Q5RK05	Mgp	1.59	3.21E-02
F1M8E9	Lyz2	1.50	1.47E-02
Q5M8C6	Fgl1	1.43	3.52E-02
D3ZY96	Ngp	1.38	4.27E-02
O55006	-	1.37	4.30E-02
D3ZHV3	Mt1m	1.29	2.87E-02
P02803	Mt1	1.26	1.88E-02
Q8CG08	Cthrc1	1.09	4.00E-02
D3ZAF5	Postn	1.07	3.16E-03
M0R8W9	LOC100909700	0.96	4.74E-02
D3ZQ25	Fbln1	0.92	4.91E-04
P13941	Col3a1	0.85	1.30E-02
Q6AYQ9	Ppic	0.83	8.87E-03
Q5I0E1	Lrg1	0.81	1.77E-02
Q4QQV6	Lsp1	0.80	4.03E-02
P97675	Enpp3	0.77	4.76E-02
G3V6C4	Ugdh	0.76	3.75E-02
I6L9G5	Rcn3	0.76	5.77E-03
Q5RJR9	Serpinh1	0.71	1.05E-02
Q3KR94	Vtn	0.70	4.28E-02
Q4QQV0	Tubb6	0.70	4.14E-02
Q6P725	Des	0.68	4.58E-02
O08628	Pcolce	0.68	8.70E-03
P08699	Lgals3	0.67	1.76E-02
Q5U2V1	Fkbp10	0.65	2.38E-02
P16573	Ceacam1	0.63	1.73E-02
M0R9U2	-	0.62	1.94E-02
P16636	Lox	0.62	1.45E-02
P31000	Vim	0.58	4.67E-02
F1M9B2	Igfbp7	0.56	2.36E-02
Q66H42	Lypd1	-0.50	2.32E-02
Q68FR8	Tuba3a; Tuba3b	-0.55	1.54E-02

G3V7R5	LOC100174910	-0.68	2.99E-02