**Title**
New Tools for Chemically Directed Glycoproteomics and Xe-based MRI Contrast Agents



**Author**
Palaniappan, Krishnan K.

**Publication Date**
2012

Peer reviewed|Thesis/dissertation

# New Tools for Chemically Directed Glycoproteomics and $^{129}$Xe-based MRI Contrast Agents

*by*

Krishnan Kanna Palaniappan

*A dissertation submitted in partial satisfaction of the requirements for the degree of*

Doctor of Philosophy

*in*

Chemistry

*in the*

Graduate Division

*of the*

University of California at Berkeley

Committee in charge:

Professor Carolyn R. Bertozzi, Co-Chair
Professor Matthew B. Francis, Co-Chair
Professor David V. Schaffer

Fall 2012

New Tools for
Chemically Directed Glycoproteomics
and $^{129}$Xe-Based MRI Contrast Agents

# Abstract

## New Tools for Chemically Directed Glycoproteomics and $^{129}$Xe-Based MRI Contrast Agents

*by*

Krishnan Kanna Palaniappan

Doctor of Philosophy in Chemistry

University of California at Berkeley

Professor Carolyn R. Bertozzi, Co-Chair
Professor Matthew B. Francis, Co-Chair

In recent years, technological advancements in the fields of mass spectrometry (MS) and molecular imaging have enabled scientists to ask and answer questions that once seemed impossibly challenging. For example, advancements in mass spectrometry have made it possible to provide a quantitative description of a human cell proteome, where both the identities and absolute abundances of thousands of proteins are simultaneously and reproducibly calculated. Equivalently, advancements in molecular imaging have made it possible to monitor the dynamics of posttranslational modifications, such as glycosylation, in living organisms.

Inspired by these advancements, this dissertation is divided into two sections that describe chemical approaches for both identifying and imaging biomolecules in living systems. In the first section, the biomolecules of interest are glycoproteins, and I describe chemical tools and computational strategies for identifying glycoproteins through MS. After a brief discussion of the importance of protein glycosylation in Chapter 1, I survey chemical methods that facilitate the study of glycoproteins through MS in Chapter 2. A major theme that emerges from this discussion is that although glycoproteins are highly abundant biomolecules, they exist as a complex mixture of enormously diverse structures, which has made identifying any one member of the population a great analytical challenge.

In Chapter 3, I introduce the concept of chemically directed proteomics. This method is used to direct MS analysis to specific species of interest (regardless of abundance) by chemically tagging them with an identifiable isotopic signature. Towards extending chemically directed proteomics beyond chemical labeling, in Chapter 4 I explore metabolic strategies to incorporate an isotopic signature directly into glycans using a specific mixture of monosaccharide isotopologs.

In the second section, the focus is molecular imaging, and I describe how viral nanoparticles can be used as molecular scaffolds to develop high sensitivity contrast agents for magnetic resonance imaging (MRI). After a brief discussion of the advantageous role that viral nanoparticles have played in the development of gadolinium-based contrast agents in Chapter 5, I then

focus on a new class of contrast agents known as xenon biosensors. In Chapter 6, I describe the utility of xenon-based MRI for *in vivo* image and illustrate that through the combination of hyperpolarization and chemical exchange saturation transfer detection, xenon biosensors can achieve low detection thresholds.

Towards improving the detection sensitivity of xenon MRI further, biosensors can be assembled onto supramolecular scaffolds such as viral nanoparticles. In Chapter 7, I discuss the development of a bacteriophage MS2-based xenon biosensor–the first viral capsid functionalized as a $^{129}$Xe-based MRI contrast agent. Subsequently, in Chapter 8, I extend the application of viral capsids from spherical bacteriophage to filamentous bacteriophage by generating an M13 bacteriophage-based xenon biosensor. Bacteriophages where chosen because they are routinely used in phage display techniques for identifying new epitope-targeting groups such as peptides and antibody fragments. Accordingly, in Chapter 9, I describe the development of a phage-based xenon biosensor that possesses antibody fragments for targeted imaging of cancer cells *in vitro*.

In summary, this dissertation describes a number of chemical approaches for both identifying and imaging biomolecules within biologically relevant environments. In the future, it will be exciting to watch as these tools are further refined and improved upon to address outstanding questions in disease biology. In particular, there are a number of research projects underway to identify cancer-related cell surface glycoproteins that can serve as biomarkers for disease states. Additionally, efforts are underway to apply phage-based xenon biosensors to lung cancer detection and imaging *in vivo*.

This dissertation is dedicated to the friends and family who have helped me reach the light at the end of the tunnel.

# Table of Contents

## *Section III: Appendices*

# List of Figures

**Chapter 7: A $^{129}$Xe-based MRI Contrast Agent Assembled on the MS2 Viral Capsid**



**Chapter 8: A $^{129}$Xe-based MRI Contrast Agent Assembled on the M13 Bacteriophage Viral Capsid**

**Chapter 9: Molecular Imaging of Cancer Cells Using a Filamentous Bacteriophage-based $^{129}$Xe Biosensor**

## List of Schemes

## List of Tables

## Acknowledgements

Although graduate school has been challenging, it has also been an extremely rewarding experience, which I owe to all the people who spent time training and teaching me how to think about science, providing that 15th hour motivation to keep pushing forward, and most importantly, sharing a drink with after a successful day in lab (or more likely, to commiserate with about a failed experiment). I feel fortunate to have had so many excellent mentors, collaborators, lab mates, and friends to provide motivation, wisdom, and support along the way. Thank you!

First, I would like to thank my advisors, Carolyn and Matt, for providing a supportive, collaborative, and high-caliber research environment. They gave me the freedom to explore my research interests over the years, and in that process I gained so much confidence in my own abilities. Working with them on manuscripts, grants, and presentations taught me about how to effectively and convincingly communicate my science, skills which are typically under taught, but critical for becoming a successful principle investigator. Carolyn has the ability to make any idea sound amazing, and by watching her talk about science it becomes clear why this is the case; her skills as a communicator are unparalleled. Matt has a never-ending enthusiasm for science. Despite countless failed experiments, he has been great at reviving my confidence and providing me with the necessary motivation to continue forward.

As a member of two research groups, I had the opportunity to make a lot of great friends in both labs. Collectively, they served many roles, from stress reliever and drinking buddy to problem solver and collaborator, so most importantly, thank you Bertozzi and Francis groups for making graduate school more bearable.

Mike B. and Kimberly B., you two were my day-to-day advisors. Mike, thank you for helping me find my self-confidence and serving as an invaluable source of information and suggestions for troubleshooting problems. Kimberly, thank you for always having time to listen, going to lunch, and providing suggestions for any problem I threw your way. Together, you two helped me solve so many problems it's hard to adequately thank you. I'm glad we became good friends outside of lab and I am excited to see your independent careers develop. Additionally, I would like to thank a number of other postdocs and graduate students for their invaluable suggestions and advice: Sloan S. (biology and microscopy), Michelle F. (phage handling), Jessica S. (biology), Penny D. (immunology), Jen C. (microscopy), and Ellen S., Kamil G., John J., and Ramesh J. for help in synthesis.

Another group that deserves a huge amount of gratitude is "Team Dibromide." When I joined the team there were two principle members, Brian S. and Austin P., and I immediately knew they would be a fun pair to work with! I had a great time working with them on all things related to isotopic recoding, both in and outside the lab. Brian (and Sarah H.), thank you for starting the proteomics infrastructure in the Bertozzi Lab. You were an early mentor and I appreciated the time and effort you spent introducing me to the wonderful world of mass spectrometry. Together with Mike, Kamil, and Ramesh, you taught me to the real way science moves forward: over a pint of beer! Austin, thank you for bringing real analytics to the group; you possessed a unique skillset without which the isotopic recoding project might never have progressed. Shortly before their departure from the lab, Mark B. joined the team. Mark, thank

you for a productive collaboration; you taught me how to identify and leverage the strengths of a project for efficient publication. I enjoyed your company during late nights and all-weekenders working on and troubleshooting the Orbitrap, as well as at any party that exploited your excellent meat grilling skills! Upon Mark's departure (if you cant tell, I have been in the lab for awhile), "Team Dibromide" realigned itself to the more general "Team Proteomics," and I wish all the current and new members in this project area the best of luck. Associated with all mass spectrometry projects was Tony I., a staff scientist at the QB3 mass spectrometry facility. Tony, thank you for all the help over the years with manuscript editing, instrument problem-solving suggestions, and a deep knowledge of mass spectrometry.

Although I probably spent more time in 810 Latimer than in any other place in my life, I could not have asked for a better group of people to share that time with! Kim S. joined me in 810 a few months after I moved in, and I am so glad she did! Kim, thank you for being such a great friend and lab mate. I don't think I could have survived graduate school without you. Thank you for always being there to listen or help, scientifically or otherwise, and for organizing so many activities outside of lab. Sarah G. and Pam C., thank you for setting the tone of the room they way you did. You two were great senior students and served as excellent role models. I appreciated all the help you provided, whether it was relief from science or listening to a practice group meeting. Paresh A. and Chelsea G., you guys were awesome additions to the room; thank you for continuing to make 810 a wonderful and friendly place to work. I am glad that we become good friends, where we can help with each other's research as well as share a celebratory drink together (Falcon Tube Fridays!!). Plus, the 810 field trips were the best; hopefully you guys will continue them in the future. And finally to Anderson L., who was an undergraduate that joined the 810 crew shortly after he graduated, thank you for being a great roommate and friend.

Gary T., Leah W., Chris B., and Allie O., although I was dubbed "100% Bertozzi," you always made me feel welcome on the 7$^{th}$ floor. Thank you for being such great friends over the years. I could always count on at least one of you (if not all of you) to enjoy all the wonderful things the bay area had to offer outside of science (ski trips, wine tasting, amazing food). You guys knew how to work hard and play hard, so thank you for helping me find that balance. Ellen S. and Karen D., you were great classmates, excellent resources in the lab, and even better friends. Thank you for helping dampen the stresses of coursework, practice quals, and failed experiments. Ellie S., thanks for all the lunch dates. You become a close friend over the years and it has been a lot of fun escaping from the drudgeries of science with you. Good luck finishing up! Ben S., although you joined the lab later than the others, you became such a good friend. Over the last few years you were a reliable drinking buddy, a great problem solver, and a reliable sounding board for ideas. Thank you for initiating the Bertozzi Lab Bike Club; I truly enjoyed the mental escape that accompanied a long bike ride with you, Kamil, Steve C., and others who joined us from time-to-time. I wish you the best of luck in your independent career. Kamil and Ramesh, you two were great drinking buddies and good friends. And to the rising students (and some youngins)—Mike C., Gaby A., Brendan B., Lauren W., and Peter R.—I enjoyed getting to know all of you. You guys are a social, hardworking, and fun group and I wish you all the best during your remaining years in graduate school.

I would also like to thank all the students that I had the opportunity to directly mentor—

Yohei R., Marjoke D., Mike C., and Jessica L.—you each helped me become a better teacher and researcher. Yohei, thanks for being the perfect undergraduate for me; you were able to work efficiently with minimal guidance and it was unfortunate that your project suffered from so many unforeseen challenges.

To the Bertozzi lab support staff—Asia A., Olga M., Karen K., and Sia K.—thank you for all the work you put into keeping the lab running. Your efforts never go unnoticed, and I appreciated all the help you provided in obtaining meetings with Carolyn, organizing and maintaining the lab, and navigating the bureaucratic nightmare that is UC Berkeley.

Thank you to all my "Team Xenon" collaborators who greatly expanded my knowledge in a new area of science (xenon NMR) and my appreciation for collaborative research. Thank you Alex P. and Dave W for helpful suggestions during our monthly xenon summits. Additionally, I had the pleasure to work with several graduate students and postdocs in their labs, including Matt R., Todd S., Tyler M., and Monica S. Matt, you were one of the best collaborators I have ever had and I hope that all my future collaborations will be as successful as ours. Although we suffered through some long and grueling experiments, I always felt satisfied with the results. I appreciated all the time you spent explaining technical jargon, data analysis, and instrument setting so I would understand all aspects of the project. I know the project will continue to succeed under your guidance.

And finally, thank you to my family for their support and understanding over the years. I know they too will be happy to say that I am no longer "in school!"

# SECTION I: NEW TOOLS FOR CHEMICALLY DIRECTED GLYCOPROTEOMICS

# Chapter 1: The Biological Relevance of Protein Glycosylation

## 1.1 Introduction

Protein glycosylation mediates a diversity of cellular processes. Impressively, more than 50% of eukaryotic proteins are glycosylated according to a bioinformatics survey of the Swiss-Prot database.[1] However, the biological significance of many of these structures remains unknown, partly due to the difficulty in studying glycosylation using traditional biochemical tools.

Unlike DNA transcription and RNA translation, glycosylation is not template-driven; instead, glycoproteins are generated through enzymatic addition of monosaccharides (Figure 1-1) and complex oligosaccharides onto protein scaffolds post-translationally, a process that can produce significant diversity and complexity.[2,3] The attached carbohydrates, termed glycans, endow the protein with new properties including cellular localization, substrate specificity, and thermodynamic stability.[4-6] Glycoproteins on the cell surface facilitate cell recognition, signaling, and adhesion, with known roles in host-pathogen interactions, inflammation, and cancer metastasis (Figure 1-2).[7-9] Additionally, intracellular glycosylation can respond dynamically to extracellular signals, with known roles in regulating nutrient and growth factor sensing, cell cycle progression, and stress responses.[10-12] Moreover, changes in protein glycosylation—

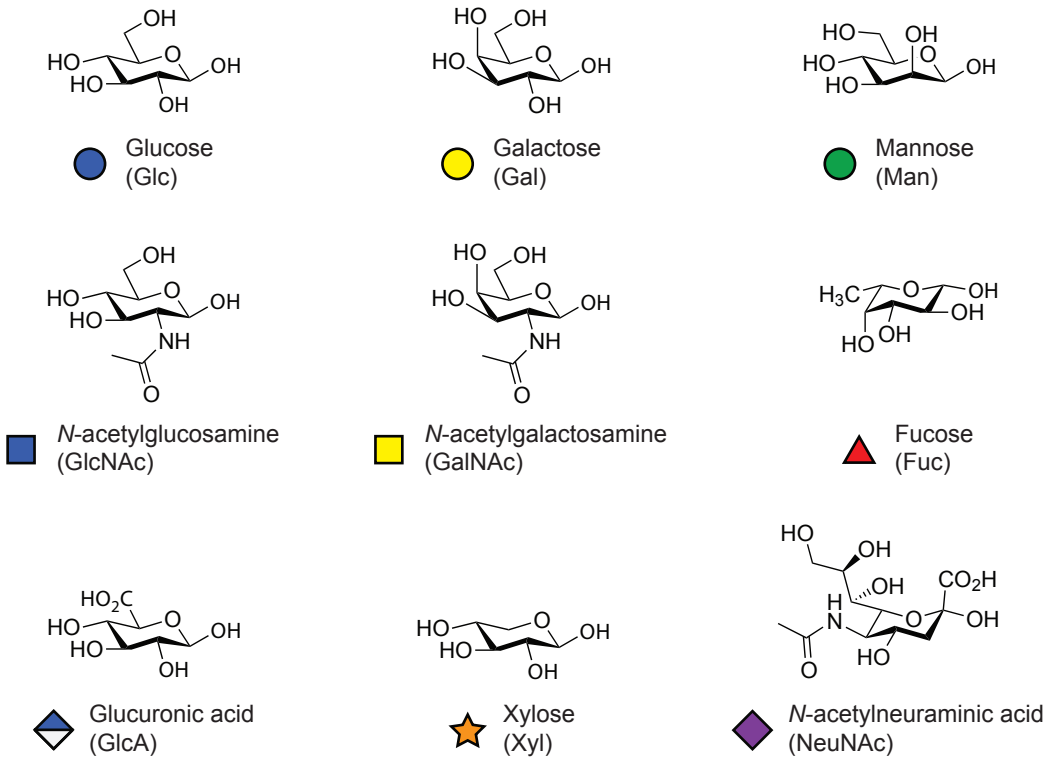



***Figure 1-1. The structure and graphical representation of the nine common monosaccharide building blocks found in vertebrates.***

N-acetylneuraminic acid is the most common form of sialic acid. This figure was adapted from Essentials of Glycobiology, 2nd Ed.[9]

including the glycan structure, the site of modification, and the protein abundance—can modulate organismal phenotypes including growth, development, and disease.[13-15] Importantly, aberrant glycosylation is well recognized as a hallmark of many diseases and their progression, including microbial pathogenesis, immune deficiencies, neurodegenerative diseases, and many cancers.[16,17] This has been validated by the more than 15 glycoproteins that currently serve as clinically approved biomarkers.[17-21] Consequently, a large effort has been invested towards the accurate characterization of a cell's glycoproteome.

## 1.2 Complexity of the Glycoproteome

Characterizing the glycoproteome is a challenging endeavor. This results from the lability, diversity, and heterogeneous nature of the nine-monosaccharide building blocks that are present in glycans (Figure 1-1): D-glucose (Glu), *N*-acetyl-D-glucosamine (GlcNAc), D-glucuronic acid (GlcA), D-galactose (Gal), *N*-acetyl-D-galactosamine (GalNAc), D-xylose (Xyl), D-mannose (Man), L-fucose (Fuc), and sialic acid (Sia), primarily *N*-acetylneuraminic acid (NeuNAc). Additionally, glycans can be used to create a variety of glycoconjugates (e.g., glycosphingolipids, glycosylphosphatidylinositol-anchored glycoproteins, proteoglycans, and glycoproteins) depending on the type of biomolecule (e.g., lipids or proteins) they are attached to.[2,22,23]

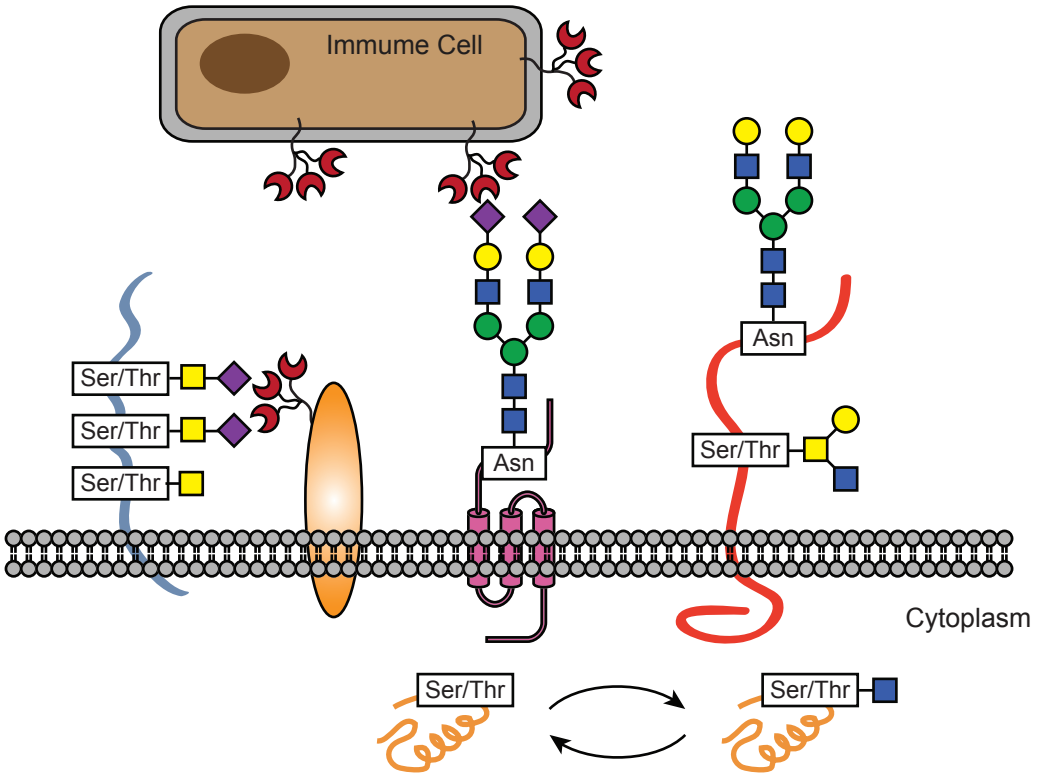



***Figure 1-2. Glycoproteins mediate many biological events.***

Glycan-binding proteins on the same cell or on neighboring cells can recognize glycans displayed on cell-surface glycoproteins. These interactions mediate diverse aspects of cell communication and recognition. Additionally, intracellular glycosylation can regulate cell physiology in response to extracellular signals in a dynamic fashion.

The complete collection of glycoproteins from an organism, known as the glycoproteome, can be characterized based on the amino acid attachment side between the glycan and the protein core. These include *C*-linked (through tryptophan), *S*-linked (through cysteine), *N*-linked (through asparagine) and *O*-linked glycosylation, with the last two being the most common and the subject of this dissertation.[24-27]

### *1.2.1 N-linked glycosylation*

In *N*-glycosylation, glycans are attached to the nitrogen atom of an asparagine residue found primarily within the consensus sequence Asn-Xaa-Ser/Thr (where Xaa is any amino acid except proline).[28] *N*-linked glycosylation is frequently found on proteins that are translated into the endoplasmic reticulum (ER) where a pre-assembled oligosaccharide, $Glc_3Man_9GlcNAc_2$, is enzymatically transferred onto a nascent protein.[29] Subsequently, the *N*-glycan precursors are enzymatically deglycosylated into a core pentasaccharide structure (Figure 1-3A, in red) and then further edited by glycosyltransferases in the Golgi to yield a plethora of mature glycoforms with compositions largely dependent on cell type and protein localization (Figure 1-3A, in green and purple). Additionally, there are several glycosidases specific for *N*-glycans, including the endoglycosidase peptide *N*-glycosidase F (PNGaseF), an amidase that converts asparagine into aspartic acid by cleaving the entire glycan from the polypeptide background.[30]

### *1.2.2 O-linked glycosylation*

In contrast to *N*-glycosylation, there is greater structural diversity and fewer biochemical or genetic tools available for studying *O*-glycosylation. In *O*-linked glycosylation, glycans are attached to the hydroxyl oxygen of serine, threonine, tyrosine, or hydroxylysine residues.[26] Then, depending upon the identify of the first monosaccharide appended, there are many subclasses of *O*-glycans, including *O*-glucosylation, *O*-fucosylation, *O*-mannosylation, *O*-GlcNAcylation and mucin-type *O*-glycosylation (*O*-GalNAc).[2]

Mucin-type *O*-Glycans represents a large percentage of all cellular *O*-glycosylation.[9] They are characterized by a structurally conserved core GalNAc residue linked in an $\alpha$ configuration to either a serine or threonine residue (Figure 1-3B, in red). Their synthesis is initiated in the Golgi by a family of glycosyltransferases known as the polypeptide-*N*-acetyl-galactosaminyl transferases (ppGalNAcT), of which there are 21 isoforms.[3] Subsequently, the $\alpha$-GalNAc residue can be elaborated by a series of glycosyltransferases to generate complex, branched structures that can be found as secreted or cell surface glycoconjugates (Figure 1-3B, in green).[31] Although a consensus sequence for glycosylation has not been identified, a hallmark of mucins is the presence of heavily glycosylated regions clustered together in domains that are rich in serine, threonine, and proline residues.[32] This conformation of dense and highly branched glycosylation causes the peptide backbone of mucins to adopt an extended conformation, allowing mucins-type *O*-Glycans to extend above the cell surface.[33] Unfortunately, there is no equivalent to PNGaseF for the liberation of all *O*-glycans from their protein scaffold; the only glycosidase specific for *O*-linked glycans cleaves the *O*-linked disaccharide Gal-$\alpha$1,3-GalNAc.[34]

Another unique subclass of *O*-glycosylation, *O*-GlcNAcylation, is characterized by the addition of a single GlcNAc residue linked in a $\beta$ configuration to either serine or threonine

residues (Figure 1-3C).[35,36] The modification is unique because it is found exclusively on proteins located in the cytoplasm and nucleus of mammalian cells, it is not further elaborated, and it is rapidly reversible.[37] Similar to mucin-type *O*-Glycans, a consensus sequence for *O*-GlcNAcylation has not been identified. Conceptually, *O*-GlcNAcylation is analogous to phosphorylation, as it is appended or removed from hundreds of protein substrates in response to biological cues.[38]

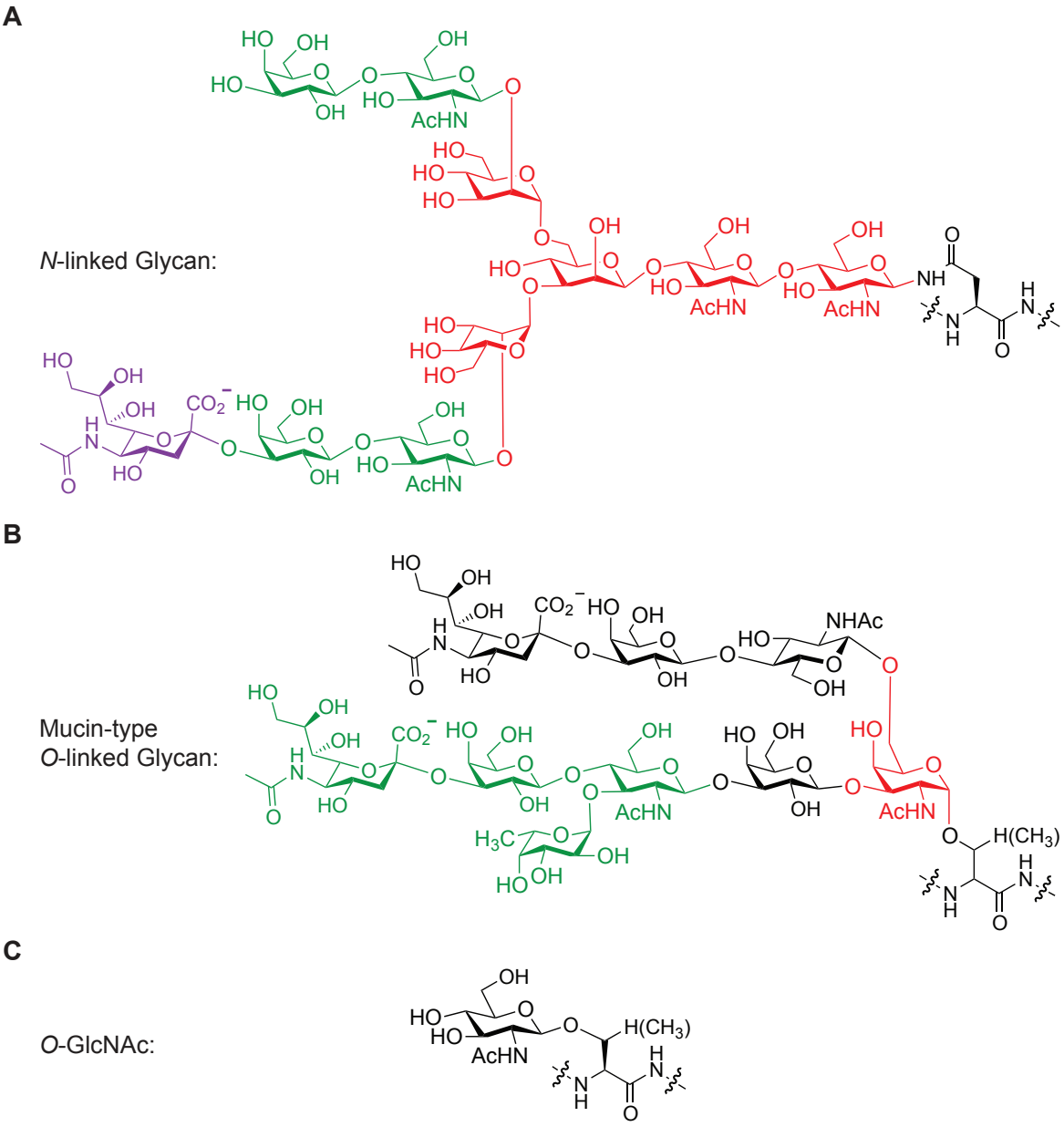



***Figure 1-3. Glycan structures display a range of complexity.***

(A) Example of an N-linked glycan, with the conserved pentasaccharide ($Man_3GlcNAc_2$) core structure highlighted in red. This can be further elaborated with N-acetyllactosamine (green), sialic acid (purple), or many other monosaccharides. (B) Example of a mucin-type O-linked glycan, with the conserved core α-linked GlcNAc residue highlighted in red. This monosaccharide can be elaborated; depicted is the Sialyl Lewis X antigen (green). (C) Example of the O-GlcNAc modification, characterized by a single β-linked GlcNAc monosaccharide.

However, unlike phosphorylation, only two enzymes are responsible for all *O*-GlcNAc cycling: *O*-GlcNAc transferase and *O*-GlcNAcase.[37]

## 1.3 Summary of Section I: New Tools for Chemically Directed Glycoproteomics

A necessary first step in establishing the biological importance of protein glycosylation is identifying relevant and interesting glycoproteins. However, the complexity of the glycoproteome has made it difficult to study protein glycosylation using traditional tools and methods. This has led to the development of new approaches and technologies, in particular the application of mass spectrometry (MS)-based proteomics. In combination with chemical tools, chemical proteomics has recently emerged as a fruitful interdisciplinary approach to identifying and studying glycoproteins.

In this section of my dissertation, I describe chemical tools and computational strategies for interrogating the glycoproteome by MS. First, in Chapter 2, I review key contributions from chemical biology that have enabled these advancements in glycoproteomics, summarizing methods for the enrichment of glycoproteins from complex mixtures, the identification of sites of glycosylation, and the identification of protein-glycan interactions (i.e., functional glycoproteomics). A major theme that emerges from this discussion is that although glycoproteins are highly abundant biomolecules, they exist as a complex mixture of enormously diverse structures, which has made identifying any one member of the population a great analytical challenge. Consequently, a critical aspect of characterizing the glycoproteome is reduction of sample complexity; in the next five chapters, I discuss several strategies that are based on this premise.

In Chapter 3, the concept of chemically directed proteomics is introduced as a unique strategy to address a major challenge in proteomics experiments: identifying proteins of low abundance. For example, in human serum, protein concentrations can span twelve orders of magnitude. Under these conditions, even the best enrichment strategies and the most sophisticated instruments (with the largest dynamic range) can be inadequate to survey low abundant species. Alternatively, through chemically directed proteomics, we can direct MS analysis to the species of interest (regardless of their abundance) by chemically tagging them with an identifiable signature. Termed isotopic signature transfer and mass pattern prediction (IsoStamp), the technique exploits the perturbing effects of a dibromide motif on a peptide's isotopic envelope. The dibromide tag was used in a proof-of-principle study to develop a computational pattern-searching algorithm capable of identifying dibromide-labeled peptides in complex mass spectra with high sensitivity and fidelity. Following pattern identification, peptide ions bearing a recoded isotopic envelope were selectively subjected to tandem MS through the implementation of an inclusion list containing *m/z* values and retention times. Thus, isotopic pattern rather than ion abundance was used to drive this directed proteomics method. With this tool in hand, we applied it to glycoproteomics in a number of ways.

While the isotopic signature of a halogenated chemical tag can effectively highlight labeled peptides within a complex LC-MS dataset, some biomolecules are not amenable to chemospecific modification. To address this limitation, and expand the utility of chemically directed proteomics, in Chapter 4 we explore metabolic strategies to incorporate an isotopic signature directly into glycans without chemical labeling. We introduce a strategy to embed a precisely

defined mixture of sugar isotopologs into eukaryotic glycans to mimic the isotopic signature of a dibromide motif without the need for bromine atoms. We use the method to detect occupied *N*-glycosites in the yeast proteome. Peptides bearing the recoded envelopes were specifically targeted for fragmentation, facilitating high-confidence site mapping.

Collectively, the techniques described in this section demonstrate that a chemical biology approach to glycoproteomics is a powerful interdisciplinary combination to interrogate the glycoproteome.

## 1.4 References

1. Apweiler R., Hermjakob H. and Sharon N. (**1999**) On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta* 1473, 4-8.

2. Stanley P. (**2011**) Golgi glycosylation. *Cold Spring Harb Perspect Biol* 3.

3. Hang H.C. and Bertozzi C.R. (**2005**) The chemistry and biology of mucin-type O-linked glycosylation. *Bioorg Med Chem* 13, 5021-5034.

4. Marino K., Bones J., Kattla J.J. and Rudd P.M. (**2010**) A systematic approach to protein glycosylation analysis: a path through the maze. *Nat Chem Biol* 6, 713-723.

5. Raman R., Raguram S., Venkataraman G., Paulson J.C. and Sasisekharan R. (**2005**) Glycomics: an integrated systems approach to structure-function relationships of glycans. *Nat Methods* 2, 817-824.

6. Helenius A. and Aebi M. (**2001**) Intracellular functions of N-linked glycans. *Science* 291, 2364-2369.

7. Rudd P.M., Elliott T., Cresswell P., Wilson I.A. and Dwek R.A. (**2001**) Glycosylation and the immune system. *Science* 291, 2370-2376.

8. Woods R.J., Edge C.J. and Dwek R.A. (**1994**) Protein surface oligosaccharides and protein function. *Nat Struct Biol* 1, 499-501.

9. Varki A. *et al.* (**2009**) Essentials of Glycobiology, Edn. 2nd. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).

10. Dentin R., Hedrick S., Xie J., Yates J., 3rd and Montminy M. (**2008**) Hepatic glucose sensing via the CREB coactivator CRTC2. *Science* 319, 1402-1405.

11. Slawson C., Lakshmanan T., Knapp S. and Hart G.W. (**2008**) A mitotic GlcNAcylation/phosphorylation signaling complex alters the posttranslational state of the cytoskeletal protein vimentin. *Mol Biol Cell* 19, 4130-4140.

12. Zachara N.E. and Hart G.W. (**2004**) O-GlcNAc a sensor of cellular state: the role of nucleocytoplasmic glycosylation in modulating cellular function in response to nutrition and stress. *Biochim Biophys Acta* 1673, 13-28.

13. Tabak L.A. (**2010**) The role of mucin-type O-glycans in eukaryotic development. *Semin Cell Dev Biol* 21, 616-621.

14. Haltiwanger R.S. and Lowe J.B. (**2004**) Role of glycosylation in development. *Annu Rev Biochem* 73, 491-537.

15. Ohtsubo K. and Marth J.D. (**2006**) Glycosylation in cellular mechanisms of health and disease. *Cell* 126, 855-867.

16. Dube D.H. and Bertozzi C.R. (**2005**) Glycans in cancer and inflammation--potential for therapeutics and diagnostics. *Nat Rev Drug Discov* 4, 477-488.

17. Drake P.M. *et al.* (**2010**) Sweetening the pot: adding glycosylation to the biomarker discovery equation. *Clin Chem* 56, 223-236.

18. Kim E.H. and Misek D.E. (**2011**) Glycoproteomics-based identification of cancer biomarkers. *Int J Proteomics* 2011, 601937.

19. Gabius H.J. (**2011**) Glycobiomarkers by glycoproteomics and glycan profiling (glycomics): emergence of functionality. *Biochem Soc Trans* 39, 399-405.

20. Pan S. *et al.* (**2006**) Identification of glycoproteins in human cerebrospinal fluid with a complementary proteomic approach. *J Proteome Res* 5, 2769-2779.

21. Ludwig J.A. and Weinstein J.N. (**2005**) Biomarkers in cancer staging, prognosis and treatment selection. *Nat Rev Cancer* 5, 845-856.

22. Rini J., Esko J. and Varki A. (**2009**) *Glycosyltransferases and Glycan-processing Enzymes* in Essentials of Glycobiology, Edn. 2nd (Cold Spring Harbor (NY)).

23. Varki A., Esko J.D. and Colley K.J. (**2009**) *Cellular Organization of Glycosylation* in Essentials of Glycobiology, Edn. 2nd (Cold Spring Harbor Press, Cold Spring Harbor, NY).

24. Hofsteenge J. *et al.* (**1994**) New type of linkage between a carbohydrate and a protein: C-glycosylation of a specific tryptophan residue in human RNase Us. *Biochemistry* 33, 13524-13530.

25. Stepper J. *et al.* (**2011**) Cysteine S-glycosylation, a new post-translational modification

found in glycopeptide bacteriocins. *FEBS Lett* 585, 645-650.

26. Jensen P.H., Kolarich D. and Packer N.H. (**2010**) Mucin-type O-glycosylation--putting the pieces together. *Febs J* 277, 81-94.

27. Burda P. and Aebi M. (**1999**) The dolichol pathway of N-linked glycosylation. *Biochim Biophys Acta* 1426, 239-257.

28. Weerapana E. and Imperiali B. (**2006**) Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems. *Glycobiology* 16, 91R-101R.

29. Kukuruzinska M.A. and Lennon K. (**1998**) Protein N-glycosylation: molecular genetics and functional significance. *Crit Rev Oral Biol Med* 9, 415-448.

30. Plummer T.H., Jr., Elder J.H., Alexander S., Phelan A.W. and Tarentino A.L. (**1984**) Demonstration of peptide:N-glycosidase F activity in endo-beta-N-acetylglucosaminidase F preparations. *J Biol Chem* 259, 10700-10704.

31. Strous G.J. and Dekker J. (**1992**) Mucin-type glycoproteins. *Crit Rev Biochem Mol Biol* 27, 57-92.

32. Perez-Vilar J. and Hill R.L. (**1999**) The structure and assembly of secreted mucins. *J Biol Chem* 274, 31751-31754.

33. Bansil R., Stanley E. and LaMont J.T. (**1995**) Mucin biophysics. *Annu Rev Physiol* 57, 635-657.

34. Umemoto J., Bhavanandan V.P. and Davidson E.A. (**1977**) Purification and properties of an endo-alpha-N-acetyl-D-galactosaminidase from Diplococcus pneumoniae. *J Biol Chem* 252, 8609-8614.

35. Hart G.W., Housley M.P. and Slawson C. (**2007**) Cycling of O-linked beta-N-acetylglucosamine on nucleocytoplasmic proteins. *Nature* 446, 1017-1022.

36. Love D.C. and Hanover J.A. (**2005**) The hexosamine signaling pathway: deciphering the "O-GlcNAc code". *Sci STKE* 2005, re13.

37. Hanover J.A., Krause M.W. and Love D.C. (**2010**) The hexosamine signaling pathway: O-GlcNAc cycling in feast or famine. *Biochim Biophys Acta* 1800, 80-95.

38. Wang Z., Gucek M. and Hart G.W. (**2008**) Cross-talk between GlcNAcylation and phosphorylation: site-specific phosphorylation dynamics in response to globally elevated O-GlcNAc. *Proc Natl Acad Sci U S A* 105, 13793-13798.

# Chapter 2: A Chemical Approach to Glycoproteomics

## 2.1 Introduction

Mass spectrometry (MS)-based proteomics is a powerful tool to interrogate biological systems. Over the past decade, the in-depth and quantitative characterization of a number of proteomes has advanced our understanding of basic cellular processes. For example, we now have an accurate representation of the protein content of a cell and subcellular structures, as well as the connectivity and dynamics between those components.[1-4] Accomplishing this has required collecting information-rich data sets that contain a comprehensive and quantitative catalog of the components of complex biological mixtures. This was made possible through a number of advances in MS technologies (i.e., mass resolution, mass accuracy, sensitivity, speed, and fragmentation) and computational methods.[4-10] However, in order to align MS-based proteomics with the current and future needs of understanding complex biological processes, proteins also need to be distinguished by their function, activity, and interactions with other biomolecules.[11-13]

Of the many ways a cell regulates the functional activity of its proteome, post-translational modifications (PTMs) are paramount. The challenge lies in identifying and characterizing these PTMs. To address this problem, developments at the interface of chemical biology and chemical

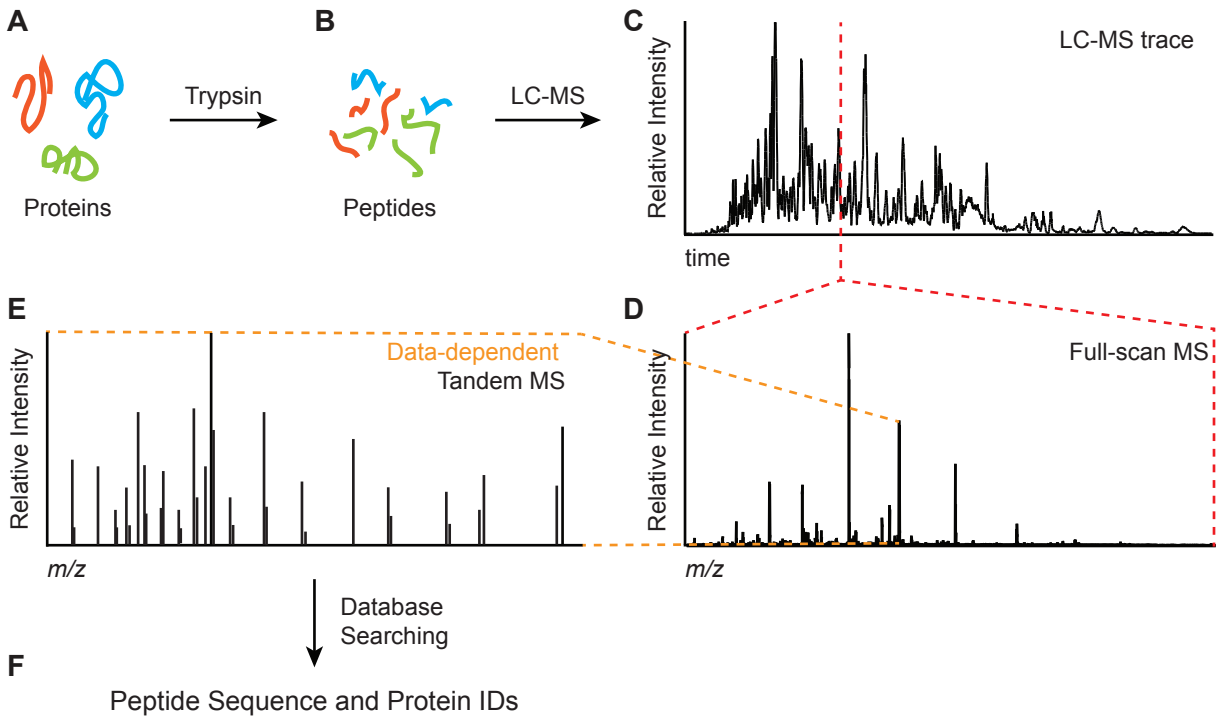



***Figure 2-1. Schematic of the bottom-up shotgun proteomics method.***

(A) A mixture of proteins is subjected to proteolytic digestion producing (B) a mixture of peptides. (C) The peptides are separated using LC and (D) full-scan mass spectra are collected at regular intervals. (E) Typically, ions are selected from each full scan for fragmentation (i.e., tandem MS) in a data-dependent fashion (i.e., based on ion abundance). (F) These fragment mass spectra are subjected to a database search to produce peptide and protein identifications.

proteomics have contributed tools and techniques to characterize a proteome with respect to specific PTMs.[14,15] Of the more than 300 known protein modifications, glycosylation is one of the most abundant and complex PTMs, and was one of the first to be probed with chemical tools.[16-18] For example, protein glycosylation has been studied using lectin arrays, perturbation with small molecules, and chemical synthesis.[17,19,20]

In this chapter, I review key contributions from chemical biology in the development of tools for interrogating the glycoproteome by MS. After a discussion of the challenges associated with glycoproteomics, I describe chemical approaches for the enrichment of glycoproteins from complex mixtures, emphasizing methods that either exploit chemical properties of carbohydrates or introduce unnatural functional groups via metabolic labeling and chemoenzymatic tagging. Next, I discuss strategies for identifying sites of protein glycosylation. I conclude with a discussion of chemical tools that will enable functional glycoproteomics to identify how proteins interact with each other in a glycosylation dependent manner. Ultimately, I hope to demonstrate that the integration of chemical approaches with existing proteomics platforms will provide an efficient route to interrogate the glycoproteome.

## 2.2 Challenges in Studying Protein Glycosylation by MS

The complexity and heterogeneity inherent to protein glycosylation imparts an analytical and computational challenge when using traditional methods to study this class of PTM. While a number of biochemical techniques are available to establish low-resolution structural information, including the use of glycan-binding proteins, radioactive labeling, antibodies, or enzymatic deglycosylation, they suffer from a combination of low throughput, low sensitivity,

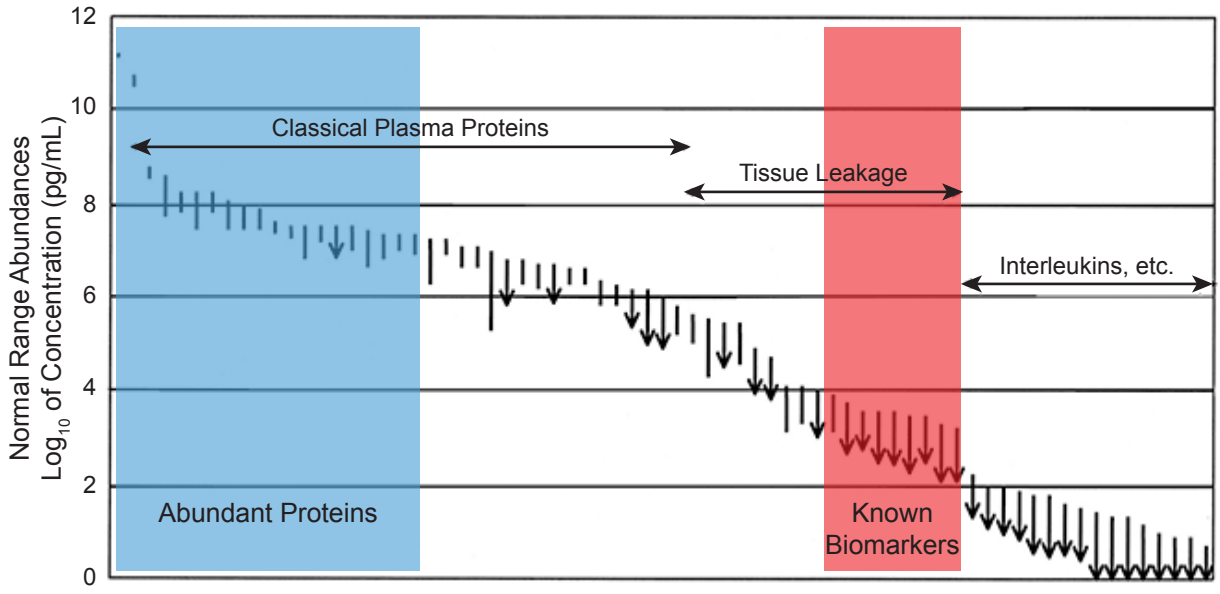



***Figure 2-2. The abundance of proteins present in human serum spans 12 orders of magnitude.***

The abundance of serum proteinvs, plotted on a log scale, emphasize the difficulty in identifying biologically interesting proteins in complex mixtures. While classical plasma proteins, such as serum albumin, are highly abundant in plasma (highlighted in blue), known biomarkers, such as prostate-specific antigen, are several orders of magnitude less abundant (highlighted in red). This figure was adapted from Anderson and Anderson.[33]

or cumbersome manual analysis. Analytical methods such as NMR spectroscopy, liquid chromatography, and electrophoresis can also be used to characterize glycans or monosaccharides, but their utility is limited by the complexity of the sample.[21]

An alternative to these methods, liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS), has emerged as an ideal technology for studying protein glycosylation. For example, there are a number of excellent reviews of MS-based proteomic analysis of glycoproteins.[22-29] The most common and successful implementation of this technology is the technique of bottom-up shotgun proteomics, illustrated in Figure 2-1.[5,7,30] In this method, proteins are converted into a mixture of peptides, which are then separated by LC and analyzed by MS. Peptide ions in the full-scan mass spectrum are selected for tandem MS, typically on the basis of ion intensity, to generate a fragment ion mass spectrum. Comparison to a proteome database is used to assign the tandem mass spectrum to its most likely peptide sequence and ultimately parent protein.[9,31]

Despite the wide use of bottom-up shotgun proteomics, studying protein glycosylation by MS is challenging for a number of reasons. Glycosylation is a labile PTM and requires special technical skill to ensure that the glycan is not lost during sample handling or tandem MS analysis.[32] Moreover, in complex biological samples such as human serum, protein concentrations can span up to 12 orders of magnitude, and many interesting glycoproteins, such as biomarkers for disease states, are present in low abundance, increasing the difficulty of detection (Figure 2-2).[33,34] Adding to the challenge, glycoproteins usually exist as complex mixtures of glycosylated variants (glycoforms), with both microheterogeneity (i.e., differences in glycan structures at one modification site) and macroheterogeneity (i.e., differences in sites of modification). Furthermore, many monosaccharides occur as structural isomers with the same mass (Figure 1-1). These vari-

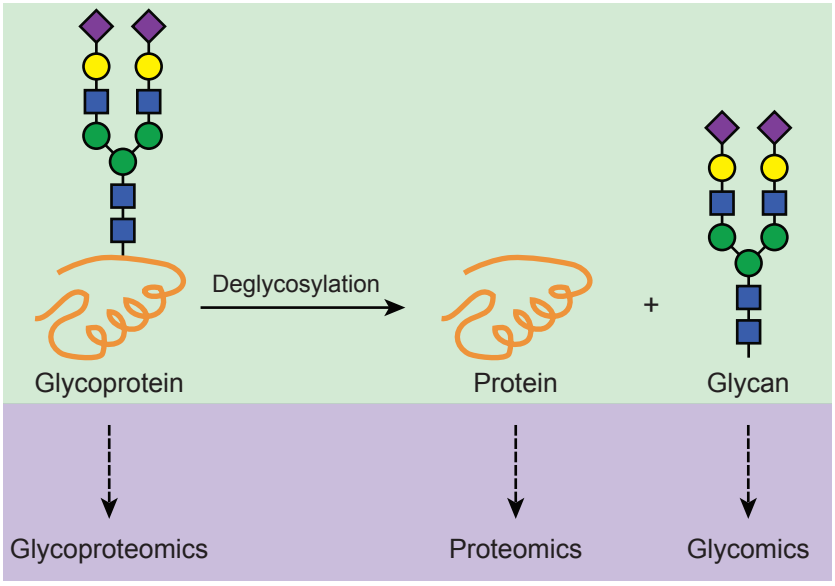



***Figure 2-3. The glycoproteome can be analyzed in several ways using MS.***

Although glycoproteins enriched from complex mixtures can be subjected to glycoproteomics methodologies for identification, they can also be further simplified before MS analysis. For example, intact glycoproteins can be deglycosylated into proteins and glycans, which can then be separately analyzed through proteomics and glycomics methodologies, respectively.

ations complicate identification and pose problems for standard bioinformatics software that cannot cope with the enormous variability in glycoproteins.[35,36]

Considering these formidable obstacles, reduction of sample complexity via enrichment is a primary goal for the successful identification and characterization of glycoproteins by MS. Once enriched, glycoproteins can be analyzed using glycoproteomics methodologies for identification. Alternatively, glycoproteins can be separated into glycans and proteins as a strategy to further reduce sample complexity, and each component can be analyzed separately through glycomics and proteomics methodologies, respectively (Figure 2-3). Here, I will focus on the analysis of intact glycoproteins and proteins after glycan release; strategies for glycan analysis (i.e., glycomics) have been summarized in a number of excellent reviews.[29,37-41]

## 2.3 Enrichment of Glycoproteins and Glycopeptides

Reducing the interference from more abundant and unglycosylated species is a prerequisite for successfully applying LC-MS/MS for the in-depth analysis of glycoproteins in complex mixtures. Traditionally, enrichment was achieved by affinity purification using antibodies and lectins that targeted specific classes of glycoproteins.[42,43] Lectins are carbohydrate-binding proteins isolated from various plants, bacteria, and animal sources and have played an important role in isolating and purifying glycoproteins. This is especially true for *N*-glycosylation where lectins have been extensively used for enrichment.[22] For example, the lectins *Sambucus nigra agglutinin* (SNA) and *Aleuria aurantia* (AAL), which bind sialic acid or fucose residues, respectively, were immobilized on a solid support and used to enrich *N*-linked glycoproteins from human plasma for biomarker detection.[44] In a similar manner, a combination of three lectins, placed in an ultrafiltration spin filter rather than immobilized to a solid support (i.e., filter aided sample preparation), were used to identify over 2000 *N*-linked glycoproteins from four mouse organs.[45]

In contrast, the greater diversity of glycan structures in *O*-linked glycosylation restricts the utility of lectins. One approach relied on serial lectin affinity chromatography in which *N*-linked glycopeptides were first removed using the lectin Concanavalin A (ConA), and then GalNAc containing peptides were enriched using the lectin Jacalin.[46] For *O*-GlcNAcylation, the most common lectin enrichment method uses wheat germ agglutinin (WGA),[47] which has also been used in a long column format, termed lectin weak-affinity chromatography (LWAC), to enrich for O-GlcNAcylated peptides.[48] In addition to lectins, pan-specific antibodies have been developed to recognize *O*-GlcNAc, some of which have been successfully used for enriching glycoproteins from complex mixtures.[49,50] Unfortunately, lectins and antibodies used for enrichment are limited by their low substrate binding affinity and/or cross-reactivity with similar glycan structures. Additionally, in the case of *O*-glycosylation, global enrichment is not possible due to the large diversity of glycan structures. Consequently, improved enrichment methods are needed that can be applied across whole classes of glycosylation and glycoconjugates.

Due to the limitations and lectin and antibody-based enrichement strategies, there is growing interest in the use of chemical tools for tagging, marking, and enriching subproteomes for analysis by MS.[51,52] Initially, chemical probes were developed to target specific amino acids.[64] The progenitors of this approach were the isotope-coded affinity tag and isobaric tags for relative and absolute quantitation, which were used to label cysteine residues or reactive amines, respec-

tively, for quantitative comparative proteomics.[53,54] Subsequently, chemical probes that targeted specific classes of enzymes were developed to extract functionally active components of the proteome. Termed activity-based protein profiling (ABPP), specific families of enzymes were covalently modified with small molecule probes that contained a chemical warhead specific for their active site, allowing for identification and characterization. ABPP has been applied to many classes of enzymes, including proteases, glycosidases, kinases, and phosphatases.[11,55-57] Most recently, this concept was applied to PTMs so that the functionally relevant states of proteins could be identified by MS.[14,15,58]

Rather than viewing glycosylation as a limitation, chemical proteomics can take advantage of the PTM for marking, enriching, and identifying glycoproteins by MS. One approach is to specifically label a glycoprotein (or glycopeptide) at its site(s) of glycosylation with a chemical reporter. The reporter will serve as a handle for either direct analysis or subsequent covalent modification with a secondary probe for enrichment. The selection of a chemical reporter and secondary probe are important in order to ensure that the reaction between the two reactive groups is specific, efficient, and can take place in a biological environment, i.e. a bioorthogonal chemical reaction (Figure 2-4).[59,60] There are three main strategies for introducing a bioorthogonal chemical reporter at sites of glycosylation. The first relies on exploiting chemical reactivity unique to the structure of a monosaccharide (for example, the *cis*-diol in sialic acid). The second method relies on metabolic labeling where an unnatural monosaccharide is incorporated directly into glycans using the cell's native machinery. Finally, the third approach relies on using purified enzymes to chemoselectively append a chemical reporter onto a specific monosaccharide *in vitro*. Each of these strategies will be discussed as they relate to the enrichment of glycoproteins for MS.

### *2.3.1 Enrichment by exploiting chemical properties of glycans*

An early approach for the enrichment of glycoproteins from complex mixtures utilized the *cis*-diol functionality characteristic of many monosaccharides. Since *cis*-diols can form covalent adducts with boronic acids or can be converted into an aldehyde after oxidative cleavage, the use of appropriate secondary reagents allows for enrichment of glycosylated species.

#### *2.3.1.1 Boronic acid adduct formation with cis-diols*

Boronic acid functionalized materials, which form covalent adducts with *cis*-diols, are one enrichment strategy that exploits the chemical properties of glycans.[61] When boronic acids are attached to a solid support, such as agarose particles, magnetic beads, or a mesoporous silica matrix, boronic esters are formed with glycans under basic conditions, enriching both *O*-linked and *N*-linked glycoproteins.[62,63] Following release under acidic conditions, glycoproteins can be analyzed and identified by tandem MS. After proof-of-principle experiments to enrich a defined mixture of glycoproteins, boronic acid functionalized materials were used to detect glycoproteins in human blood samples.[64]

#### *2.3.1.2 Periodate oxidation to introduce the aldehyde functionality*

Another approach that exploits chemical properties of glycans is the sodium periodate-mediated oxidative cleavage of *cis*-diols to introduce an aldehyde group. Serving as the chemical

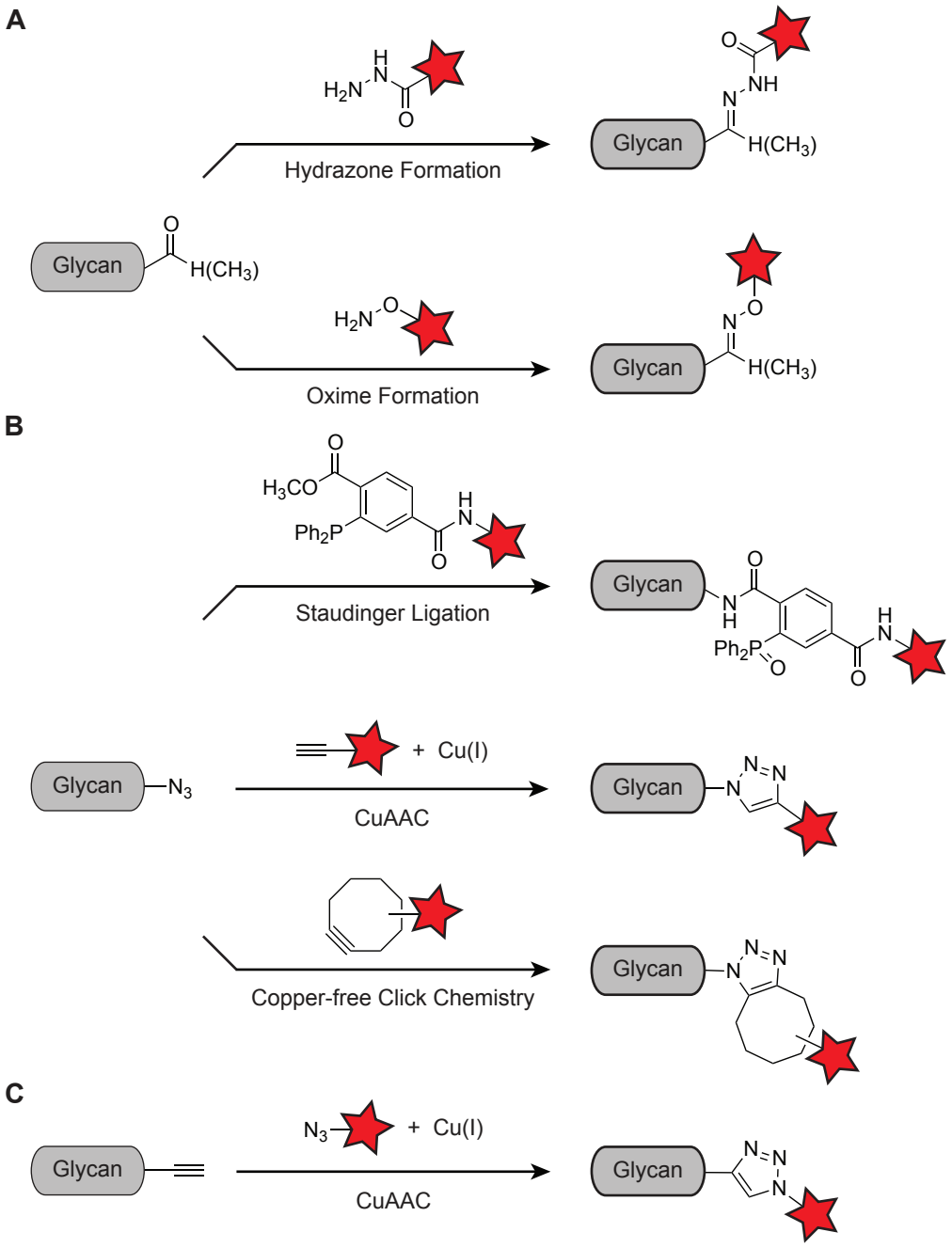



***Figure 2-4. Bioorthogonal reactions with aldehyde-, ketone-, azide-, or alkyne-functionalized monosaccharides.***

After appending an enrichment probe (red star) to glycoproteins or glycopeptides, they can be enriched from complex mixtures and identified by MS. (A) Aldehydes and ketones can condense with amine nucleophiles, including hydrazide compounds (top) and aminooxy compounds (bottom), to form stable hydrazone or oxime linkages, respectively. (B) Azides can undergo the Staudinger ligation (top) with triarylphosphines, or a [3+2] cycloaddition with either terminal alkynes through the Cu-catalyzed azide-alkyne cycloaddition reaction (CuAAC, middle) or strained alkynes in the absence of Cu (Cu-free click chemistry, bottom). (C) Alkynes can undergo a [3+2] cycloaddition with azides through the CuAAC reaction.

reporter, the aldehyde can be condensed with amine nucleophiles, including hydrazine and aminooxy compounds, to form covalent linkages under physiological conditions (Figure 2-4A). Aebersold and coworkers were the first to implement this strategy by using hydrazide-functionalized beads to enrich periodate-oxidized glycoproteins from human serum (Figure 2-5A).[65] With a covalent bond between glycoproteins and the solid support, nonglycosylated proteins could be washed away and nonglycosylated portions of bound glycoproteins could be digested away using trypsin. In order to select for a specific class of glycosylation, PNGase F was used to selectively release *N*-linked glycopeptides for their subsequent identification by tandem MS. Termed the hydrazide capture technology, 145 glycopeptides, mapping to 57 *N*-linked glycoproteins, were identified in human serum samples. The Aebersold group also developed a variant of this method, where glycoproteins were first digested into peptides and then oxidized with

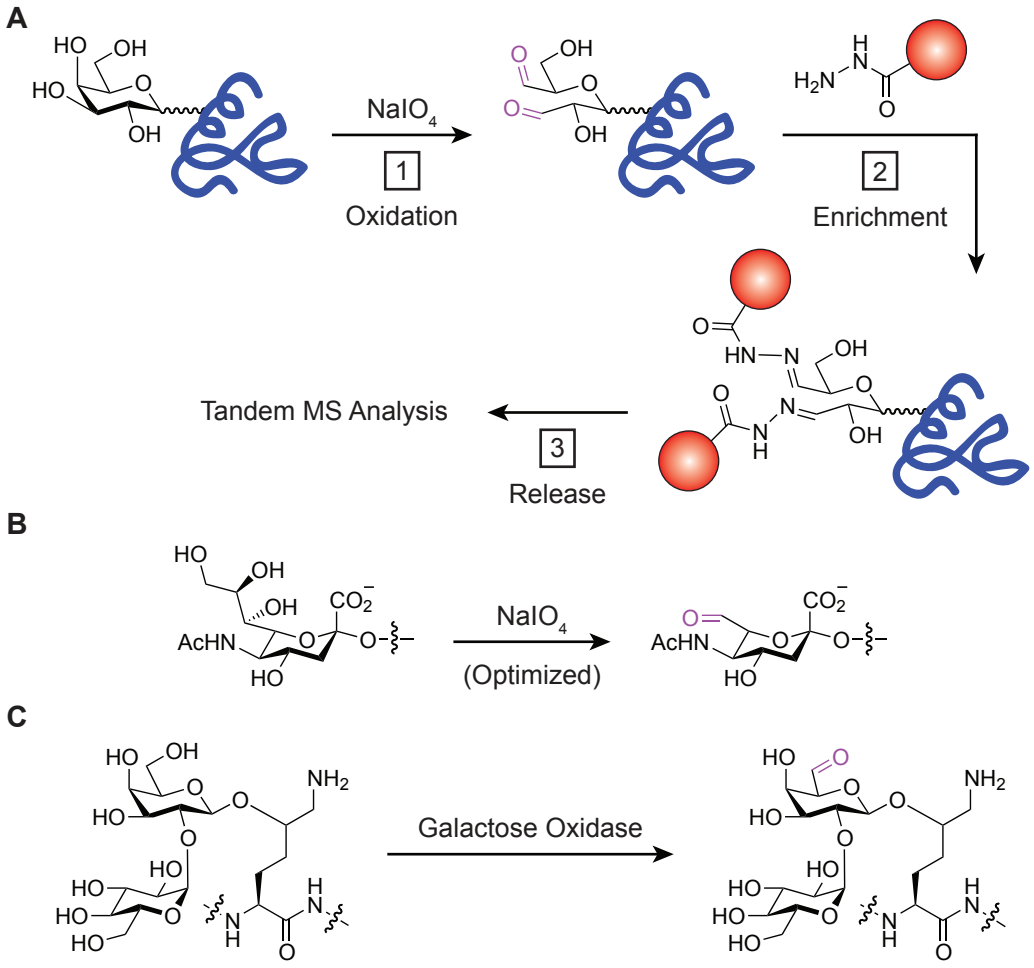



***Figure 2-5. Enrichment of glycoproteins through carbonyl chemistry.***

(A) Schematic of the hydrazide capture technology. First, glycoproteins are subjected to mild oxidants, most commonly sodium periodate, the cis-diols in glycans can be oxidatively cleaved to introduce an aldehyde group (highlighted in purple). Second, the aldehyde can be condensed with a hydrazide-functionalized solid support to covalently capture and enrich glycoproteins. Finally, the protein can be released from the solid support allowing for identification by LC-MS. In addition to a general method for glycoprotein enrichment, the hydrazide capture technology can be applied to glycopeptides and can target glycans containing specific monosaccharides including (B) sialic acid and (C) galactose, using optimized sodium periodate conditions or galactose oxidase, respectively.

sodium periodate.[66] A number of subsequent studies have applied either one of these methods to identify *N*-linked glycoproteins in human plasma, cerebrospinal fluid, islet cells, platelet cells, saliva, and the microsomal fraction of an ovarian cancer cell line.[67-73] There are, however, two limitations to the hydrazide capture technology. The first is the potential cross-reactivity between hydrazide-functionalized beads and naturally occurring carbonyl-containing biomolecules in cell lysates (including glucose, pyruvate, and pyridoxal-5'-phosphate), which can reduce the sensitivity for glycoprotein enrichment. The second is the lack of specificity in the monosaccharides that become oxidized and the class of glycoproteins that become enriched (i.e., both *O*-linked and *N*-linked glycoproteins will display oxidized glycans). While the general strategy used by Aebersold and others to address these issues was through enzymatic release of captured glycopeptides (i.e., PNGase F specifically cleaves *N*-glycans), others have made modifications to the hydrazide capture technology that expand its utility.[74,75]

To improve the sensitivity of glycoprotein enrichment with the hydrazide capture technology, Wollscheid *et al.* introduced aldehyde groups onto cell surface glycoproteins, where aldehydes and ketones are traditionally absent, by subjecting live cells to optimized sodium periodate oxidation conditions without compromising cell viability.[74,76] This allowed for the newly generated aldehydes to react with a hydrazide-functionalized enrichment probe (biocytin hydrazide) without the need to lyse cells and expose naturally occurring carbonyl-containing biomolecules. Following proteolytic digestion of labeled-cell lysates, glycopeptides from cell surface glycoproteins were captured with streptavidin beads. After enzymatic release using PNGase F, *N*-linked glycopeptides were identified by tandem MS. This method, termed cell surface capturing (CSC) technology, was used to identify the differences between the cell surface *N*-glycoproteome of T and B cells. In the same work, CSC technology was integrated with stable isotope labeling by amino acids in culture for quantitative glycoproteomics. This was used to monitor the dynamics in the cell surface *N*-glycoproteome during T-cell activation and during mouse embryonic stem cell differentiation into a neural lineage. More recently, the CSC technology was used to profile the cell surface *N*-glycoproteome of a mouse myoblast cell line, a mouse beta-cell line, a hepatoma cell line, myeloid leukemia cells, and pluripotent stem cells.[71,77-80] While the CSC technology reduces cross-reactivity between the enrichment probe and native carbonyl-containing biomolecules, the use of PNGase F restricts the method from identifying *O*-linked glycoproteins.

Although the hydrazide capture technology can be used to enrich multiple classes of glycoproteins, it has been limited to studying *N*-glycosylation due to the lack of methods to efficiently release and recover *O*-linked glycoproteins. Larson and coworkers addressed the limitation by using optimized sodium periodate oxidation conditions to selectively oxidize the *cis*-diol in sialic acid residues, a terminal monosaccharide found on both *N*- and *O*-linked glycoproteins (Figure 2-5B).[75,76] After oxidized glycoproteins were captured on hydrazide beads and nonglycosylated peptides were digested off using trypsin, both *N*- and *O*-linked glycopeptides were released using mildly acidic conditions that cleave sialic acid residues. When used to identify glycoproteins in human cerebrospinal fluid, 36 *N*-linked and 44 *O*-linked glycopeptides were identified, mapping to 45 glycoproteins, many of which were previously known. To distinguish between *N*- and *O*-linked glycopeptides, as well as characterize some of the glycan structures, second-gen-

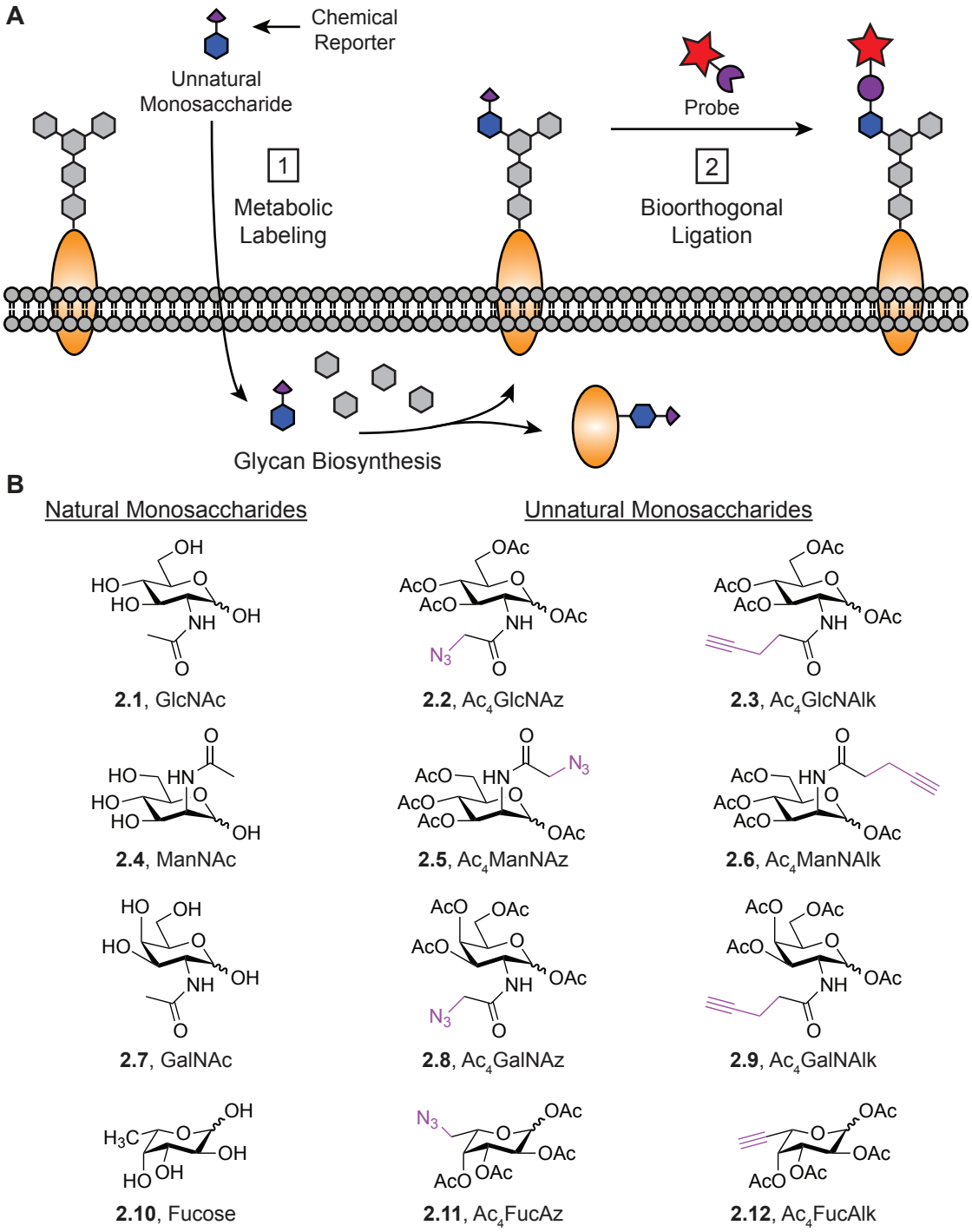



***Figure 2-6. Enrichment of glycoproteins through the bioorthogonal chemical reporter strategy.***

(A) Metabolic oligosaccharide engineering (MOE) can be used to introduce a chemical reporter (i.e., azide or alkyne; purple triangle) directly into glycans. First, unnatural monosaccharides can be accepted by the cells' natural machinery and incorporated into cell-surface, secreted, and/or intracellular glycoproteins. Subsequently, a bioorthogonal chemical reaction can be performed to specifically label glycans with a probe molecule (red star). (B) List of azide- and alkyne-functionalized unnatural monosaccharides (chemical reporters highlighed in purple) used for metabolic labeling of glycans (natural monosaccharides are shown in the left column). Although structures are shown in their peracetylated form, after entering cells by passive diffusion, cytosolic esterases will cleave the acetyl groups.

eration fragmentation (i.e., $MS^3$) and manual inspection of mass spectra were required. One of the challenges encountered was discerning the *N*- from the *O*-glycans because of the overlap in monosaccharide residues amongst those structures. More recently, this sialic acid specific enrichment strategy was used to identify both *N*- and *O*-linked glycoproteins in human urine and mouse serum.[81,82] In a similar fashion, Paulson and coworkers demonstrated that they could selectively introduce aldehydes into cell surface sialic acid residues using mild sodium periodate oxidation conditions while maintaining cell viability.[76] After reacting oxidized glycoproteins with an aminooxy-biotin enrichment probe, in the presence of aniline as a catalyst, they demonstrate high efficiency labeling. Biotin labeled glycoproteins were captured onto streptavidin beads and then nonglycosylated peptides were digested off using trypsin to identify putative glycoproteins while *N*-linked glycopeptides were cleaved off using PNGase F to identify the sites of glycosylation.[83]

Recently, the hydrazide capture technology was extended to the *O*-GlcNAc modification. Medzihradszky and coworkers demonstrated that at elevated temperatures and high concentrations of sodium periodate, the *trans*-diol of GlcNAc could be oxidatively cleaved to introduce aldehydes.[84] After capturing oxidized glycoproteins on a hydrazide-functionalized resin, and digesting off nonglycosylated peptides using trypsin, glycopeptides were released by hydrazone cleavage using hydroxylamine and mildly acidic conditions. When applied to subcomplexes of the proteasome from *Drosophila melanogaster*, they identified 12 *O*-GlcNAcylated peptides, mapping to 5 glycoproteins.

In addition to the *O*-GlcNAc modification, Hattori and coworkers demonstrated that the hydrazide capture technology could be used to enrich two types of *O*-linked glycosylation unique to collagen, galactosyl-hydroxylysine (GHL) and glucosyl-galactosyl-hydroxylysine (GGHL).[85] Instead of chemically oxidizing the glycans with sodium periodate, which was previously shown to destroy the glycans, they used the enzyme galactose oxidase, which can oxidize terminal galactose and GalNAc residues at the C-6 position to generate an aldehyde (Figure 2-5C). They applied this technique to enrich for glycopeptides from bovine type I and type II collagens. After oxidized glycopeptides were enriched on hydrazide resin and eluted under acidic conditions, 37 GHL/GGHL glycopeptides, 13 from type I and 24 from type II collagen, were identified by tandem MS. Recently, Paulson and coworkers extended the utility of galactose oxidase to live cells. They demonstrated that aldehydes could be selectively introduced into terminal galactose and GalNAc residues on cell surfaces. After oxidation, glycoproteins were enriched and identified following aniline-catalyzed ligation with an aminooxy-biotin probe.[83]

In general, the hydrazide capture technology has become a valuable chemical enrichment strategy for studying *N*-glycosylation. With the integration of quantitative methods, this technology has led to the identification of thousands of new substrates across many organisms. While some efforts have been made to extend the technology for *O*-glycosylation, the lack of efficient chemical and enzymatic tools for release has limited its utility.

### *2.3.2 Enrichment by metabolic labeling with unnatural monosaccharides*

Metabolic oligosaccharide engineering (MOE) is an emerging strategy for glycoprotein profiling.[14,15,23,58,86] After Reutter and coworkers discovered that the acyl side chain on *N*-ace-

tyl-D-mannosamine (ManNAc) was amenable to derivatization,[87] Bertozzi and coworkers exploited carbohydrate metabolic pathways to introduce bioorthogonal chemical reporters directly into glycans in live cells or organisms as a method to visualize and profile glycan dynamics (Figure 2-6A).[88,89] The azide and alkyne are two powerful chemical reporters: their small size allows them to readily pass through a cell's natural carbohydrate biosynthesis machinery, their unique chemical reactivity is amenable to a number of bioorthogonal reactions, and unlike aldehydes and ketones, they are generally absent in biological environments.[79,80] In MOE, the first step is labeling glycans with an unnatural monosaccharide analogue bearing a chemical reporter, either an azide or alkyne (Figure 2-6B). In the second step, functionalized glycans are reacted with azide/alkyne-selective probes through either the Staudinger ligation with triarylphosphines,[90] the copper(I)-catalyzed azide-alkyne cycloaddition (CuAAC) with linear alkynes,[91,92] or "copper-free click chemistry" with strained alkynes (Figures 2-4B and 2-4C).[93,94] This strategy has led to metabolic labeling using a number of membrane-permeable per-*O*-acetylated azide- or alkyne-functionalized monosaccharides, including analogues of GlcNAc,[95-97] ManNAc,[90,98,99] sialic acid,[100,101] GalNAc,[102-104] and fucose[99,105,106]. Once the protected monosaccharides enter a cell through passive diffusion, non-specific esterases cleave the *O*-acetyl groups, allowing the functionalized monosaccharides to enter their respective salvage pathways.

*2.3.2.1 Labeling with GlcNAc analogues*

Intracellular *O*-GlcNAcylation is highly dynamic and typically substiochiometric. Consequently, enrichment methods are necessary in order to identify *O*-GlcNAcylated substrates, and bioorthogonal chemical reporters have played a key role. Specifically, *O*-GlcNAc modified proteins have been metabolically labeled with either azide or alkyne chemical reporters by feeding cells either per-*O*-acetylated *N*-azidoacetylglucosamine ($Ac_4$GlcNAz, **2.2**) or per-*O*-acetylated *N*-alkynylacetylglucosamine ($Ac_4$GlcNAlk, **2.3**), respectively. Subsequently, reaction with the appropriately derivatized enrichment probe allowed for the identification of hundreds of *O*-GlcNAcylated proteins. In the first example, Zhao and coworkers fed *Drosophila* S2 cells $Ac_4$GlcNAz, reacted cytosolic extracts with a phosphine-biotin probe through the Staudinger ligation and captured labeled proteins on streptavidin beads.[107] After bound proteins were digested off the beads with trypsin and analyzed by tandem MS, 41 putatively *O*-GlcNAcylated proteins were identified, 10 of which were previously known. In a follow-up report, the same group applied an identical strategy to identify *O*-GlcNAcylated proteins in HeLa cells. With an optimized protocol in place, 199 putatively *O*-GlcNAcylated proteins were identified, 23 of which were confirmed using reciprocal immunoprecipitation.[108] Unfortunately, the slow reaction kinetics of the Staudinger ligation can hinder the capture of low abundance glycoproteins, which is characteristic of a substoichiometric PTM such as *O*-GlcNAcylation. To compensate for this limitation, alternative bioorthogonal reactions have been coupled to metabolic labeling with GlcNAc analogues.

In addition to the Staudinger ligation, the CuAAC reaction has also been used to covalently modify GlcNAz labeled glycoproteins. For example, Lemonie and coworkers reacted nuclear and cytosolic extracts from $Ac_4$GlcNAz-fed MCF-7 cells with a biotin-alkyne probe through the CuAAC reaction.[97] After enrichment of glycoproteins with a streptavidin column and in-gel digestion

with trypsin, 32 putatively *O*-GlcNAcylated proteins were identified by tandem MS, 14 of which were previously known. Subsequently, the authors wanted to demonstrate that $Ac_4$GlcNAlk-fed cells could also yield *O*-GlcNAcylated substrates after enrichment using a biotin-azide probe. However, they did not observe any metabolic labeling with $Ac_4$GlcNAlk. In contrast, Pratt and coworkers recently demonstrated that $Ac_4$GlcNAlk could be used for metabolic labeling, albeit under low-glucose cell growth conditions.[96] To identify *O*-GlcNAcylated proteins, they reacted whole cell lysates from $Ac_4$GlcNAlk-fed NIH3T3 cells with a cleavable biotin-azide tag (azido-diazo-biotin) through the CuAAC reaction. After glycoproteins were enriched with streptavidin beads they were selectively released using sodium dithionite to reductively cleave the diazo bond. Following in-gel digestion with trypsin, 374 putatively *O*-GlcNAcylated proteins were identified by tandem MS, 95 of which were previously known, and three of which were confirmed by immunoprecipitation.

*2.3.2.2 Labeling with ManNAc analogues*

Sialic acid-containing glycoproteins have also been metabolically labeled with both azide and alkyne chemical reporters. Sialic acid, a terminal monosaccharide found on both *N*- and *O*-linked glycans, is biosynthesized from ManNAc; consequently, metabolic labeling with ManNAc analogues that contain either an azide or alkyne will transfer the chemical reporter onto sialic acid residues. In an early example, Wong and coworkers fed prostate cancer cells (PC3 cells) per-*O*-acetylated *N*-alkynalacetylmannosamine ($Ac_4$ManNAlk, **2.6**) and then reacted whole cell lysates with a biotin-azide probe through the CuAAC reaction.[109] After enrichment of sialic acid-containing glycoproteins using streptavidin beads, non-glycosylated peptides were removed by on-bead tryptic digestion and then *N*-linked glycopeptides were released through enzymatic digestion with PNGase F. Following tandem MS analysis of the PNGase F released peptides, 219 *N*-linked glycopeptides were identified, corresponding to 108 glycoproteins, of which 69 were previously known.

Semmes and coworkers also combined MOE with MS to identify sialic acid-containing glycoproteins involved in prostate cancer.[110] They fed highly metastatic (ML2 cells) and nonmetastatic (N2 cells) PC3-derived cell lines with per-*O*-acetylated *N*-azidoacetylmannosamine ($Ac_4$ManNAz, **2.5**) and then performed the CuAAC reaction on intact cells with a biotin-alkyne tag. After whole cell lysates were prepared and biotin-labeled glycoproteins were captured onto streptavidin beads, glycoproteins were released by boiling and were subjected to in-gel digestion with trypsin. Following tandem MS, 36 putative glycoproteins unique to N2 cells and 44 putative glycoproteins unique to ML2 cells were identified. Interestingly, a large number of the enriched glycoproteins from ML2 cells were involved in cell movement, migration, and cell invasion, characteristics typical of metastasis. In this workflow, an attempt was made to reduce non-specific enrichment by performing the bioorthogonal chemical reaction on intact cells, rather than cell lysates. However, the considerable cell death that can occur when performing the CuAAC reaction with live-cells (i.e., with 1 mM copper) resulted in a large amount of contaminating proteins, compromising the utility of live-cell labeling.

The most recent example of metabolic labeling of cancer cells with ManNAz was performed by Bai and coworkers.[111] They fed three different cancer cell lines (A549 lung adenocarcinoma,

HeLa cervical carcinoma, and SW1990 pancreatic adenocarcinoma) $Ac_4$ManNAz, performed the CuAAC reaction on intact cells using a cleavable biotin-alkyne probe (alkyne-disulfide-biotin), prepared whole cell extracts, and then captured biotinylated proteins on magnetic streptavidin beads. After chemical release with dithiothreitol (DTT), glycoproteins were digested with trypsin and analyzed by tandem MS. Collectively, 310 putative *N*- and *O*-linked glycoproteins were identified, 56 of which were predicted to be cell surface proteins. One particular protein, NrCAM, a cell adhesion molecule, was only found in the lung adenocarninoma cell line. To verify this result, immunohistochemistry was performed on tissue samples representative of all three cancers, and NrCAM staining was only seen in the cancerous lung tissue sample.

*2.3.2.3 Labeling with GalNAc analogues*

Of all the classes of glycosylation, mucin-type *O*-glycosylation is one of the most challenging to characterize by MS. Fortunately, despite the large diversity of glycan structures found on mucins, all of them are initiated by an α-GalNAc residue. Consequently, one solution for accessing mucin-type glycoproteins is through metabolic labeling with GalNAc analogues that contain either an azide or alkyne so that bioorthogonal reactions can be used for enrichment and identification. In the first example, to identify prostate cancer biomarkers, Bertozzi and coworkers fed PC3 cells per-*O*-acetylated *N*-azidoacetylgalactosamine ($Ac_4$GalNAz, **2.8**) and then performed the Staudinger ligation on live cells using a membrane-impermeable phosphine-FLAG probe.[112] By performing the bioorthogonal reaction on live cells, rather than cell lysates, this strategy targeted cell surface glycoproteins while minimizing potential contamination from intracellular glycoproteins. After whole cell lysates were prepared, FLAG-labeled glycoproteins were enriched by immunoprecipitation with an anti-FLAG antibody and then subjected to in-gel digestion with trypsin. Following tandem MS, 71 putative cell surface or secreted glycoproteins were identified, 29 of which were previously identified using ManNAlk labeled PC3 cells.[109] Two particular glycoproteins, CD146 and integrin beta-4, were known markers of cancer metastasis. In addition to GalNAz, the alkyne variant, per-*O*-acetylated *N*-alkynalacetylgalactosamine ($Ac_4$GalNAlk, **2.9**), has also been used to metabolically label *O*-glycans, but it has not been applied to glycoprotein profiling.[96]

In addition to labeling mucin-type *O*-glycans on cell surfaces, GalNAz can also label intracellular *O*-GlcNAcylated substrates. Based on the promiscuity in sugar-donor selection by the enzyme UDP-Galactose 4-epimerase (GALE), which interconverts UDP-GlcNAc and its C-4 epimer UDP-GalNAc, GalNAz can be converted to GlcNAz. Bertozzi and coworkers demonstrated this metabolic cross talk by feeding Jurkat cells $Ac_4$GalNAz and then reacting whole cells lysates with a phosphine-FLAG-His6 tag through the Staudinger ligation.[103] After enrichment of glycoproteins by sequential pulldowns with an anti-FLAG antibody followed by metal affinity chromatography, 18 putatively *O*-GlcNAcylated proteins were identified, most of which were previously known. This cross talk between GalNAz and GlcNAz requires careful sample preparation to minimize cross contamination between nucleocytosolic proteins and proteins translated in the ER/secretory pathway.

Recently, metabolic labeling with GalNAz was utilized to showcase a new method for identifying changes in protein glycosylation among different biological samples. The method is a

modified version of difference gel electrophoresis (DIGE),[113] in which two samples are labeled with distinct fluorophores (but identical in charge and close in mass) and then combined and analyzed on a 2-dimensional gel. After excising fluorescent gel spots that correspond to incongruity in the two fluorescent signals, protein abundance changes between the samples (i.e., due to a stimulus) can be identified by in-gel digestion and tandem MS. By integrating DIGE with MOE, we were able to rapidly monitor changes in glycoprotein levels in complex biological samples. Termed "glyco-DIGE," the technique was first used to examine organelle-specific protein *O*-GlcNAcylation (unpublished work by Mike Boyce).[114] After feeding Jurkat cells $Ac_4$GalNAz, mitochondrial and cytosolic extracts were separately labeled with an alkyne-Cyanine dye (either Cyanine 3 or Cyanine 5) through the CuAAC reaction followed by 2-dimensional gel separation. After in-gel digestion of a series of fluorescent spots unique to the mitochondria, tandem MS analysis identified them as the voltage-dependent anion channel 2 (VDAC2) protein, a mitochondrial outer membrane protein with crucial roles in glucose metabolism and apoptosis regulation.

*2.3.2.4 Labeling with fucose analogues*

Protein fucosylation plays an important role in a number of cell signaling pathways, commonly residing at the periphery of *N*- and *O*-glycan structures. To study fucosylation the bioorthogonal chemical reporter strategy can be used with either azide- or alkyne-functionalized analogues.[99,105,106] However, low labeling efficiencies (through the fucose salvage pathway) of the unnatural fucose analogues have hindered their use in glycoproteomics applications. To overcome this obstacle, Narimatsu and coworkers generated Namalwa cells that stably overexpressed two key enzymes in the fucose biosynthetic pathway, L-fucokinase/GDP-fucose pyrophosphorylase (FKP) and $\alpha$-1,3-fucosyltransferase 9 (FUT9).[115] After demonstrating that overexpression of these enzymes allowed robust metabolic labeling with an alkynyl fucose (FucAlk) analogue, they used metabolic labeling with chemical reporters to identify fucosylated glycoproteins. Specifically, after whole cell lysates of per-*O*-acetylated alkynyl fucose ($Ac_4$FucAlk, **2.12**)-fed Namalwa cells (+FKP, +FUT9) were reacted with a biotin-azide probe through the CuAAC reaction, the protein mixture was digested and glycopeptides were captured on streptavidin beads. Following selective enzymatic release of *N*-linked glycopeptides using PNGase F, 42 glycopeptides were identified, representing 36 glycoproteins.

The advantage of the bioorthogonal chemical reporter strategy for glycoprotein enrichment through MOE is twofold. First, the chemical reporter is directly embedded into the glycan, allowing for glycosylation-specific enrichment using appropriate secondary probes. Unlike lectins or antibodies, which display weak binding and cross-reactivity, bioorthogonal reactions with azide- or alkyne-functionalized monosaccharides allow for stringent enrichment conditions. Second, the unnatural monosaccharides can be dosed to cells or animals in a time-dependent manner (along with a stimulus) to reveal the dynamics and signal-dependent changes in glycosylation against a background of steady-state protein glycosylation. However, as illustrated in the case of fucose labeling, low metabolic incorporation of unnatural monosaccharides can limit their utility for MOE-based glycoproteomics. More generally, some structural perturbations might not be accommodated by all enzymes in a particular carbohydrate's biosynthetic pathway. Additionally, the activities of epimerases, which convert one monosac-

charide into another, have to be carefully investigated for each unnatural monosaccharide. As an alternative to MOE, bioorthogonal chemical reporters can be installed onto glycans through chemoenzymatic tagging.

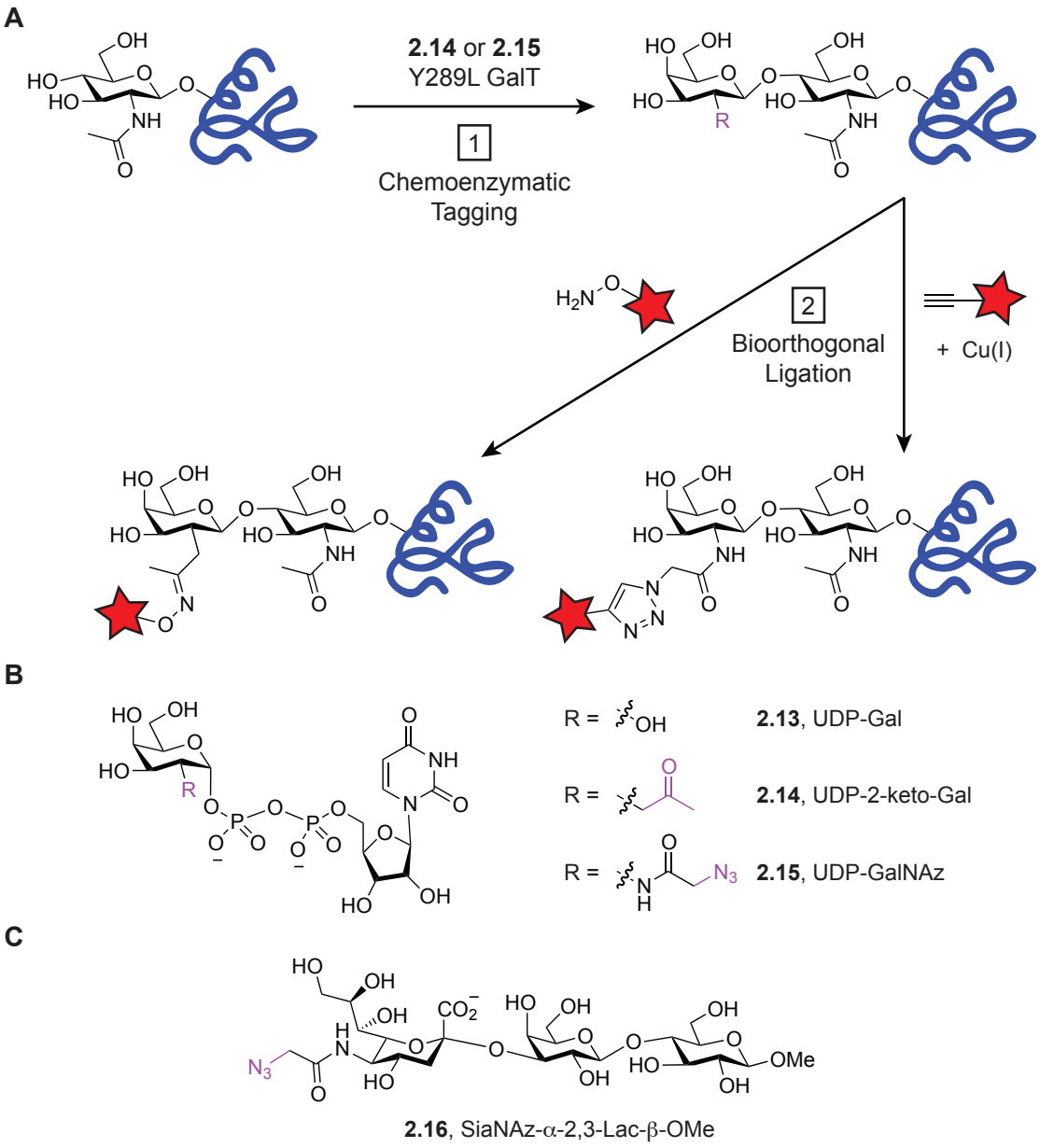



***Figure 2-7. Enrichment of glycoproteins through chemoenzymatic tagging followed by bioorthogonal ligation.***

(A) Schematic showing how chemoenzymatic tagging can be used to introduce a chemical reporter onto glycoproteins *in vitro*. Using a mutant galactosyl transferase, Y298L GalT, UDP-Galactose analogues (**2.14** or **2.15**) are appended onto the O-GlcNAc monosaccharide. Subsequently, a bioorthogonal chemical reaction can be performed to specifically label glycans with a probe molecule (red star). (B) Structures of two UDP-Galactose analogues, with either a ketone or azide functionality (highlighted in purple), used to tag O-GlcNAcylated substrates. (C) The structure of an azide-functionalized trisaccharide used to identify glycoprotein targets of a viral *trans*-sialidase.

### *2.3.3 Enrichment by chemoenzymatic tagging with unnatural monosaccharides*

Glycoproteins can be enriched after chemoenzymatic tagging with bioorthogonal chemical reporters. In this approach, glycans are enzymatically appended with a functionalized monosaccharide *in vitro* and then covalently ligated to an enrichment probe through a bioorthogonal reaction (Figure 2-7A). A challenge lies in finding or designing appropriate glycosyltransferases that can both accept unnatural sugar donors and act on their substrates extracellularly. The reward for identifying such enzymes is high efficiency labeling of glycoproteins independent of cell type.

#### *2.3.3.1 Chemoenzymatic tagging of O-GlcNAc*

Chemoenzymatic tagging of glycans was first developed for improving the detection of *O*-GlcNAcylated proteins. After the discovery of this intracellular glycosylation by Hart and coworkers using $^{3}$H-UDP-Gal and the enzyme β-1,4-galactosyltransferase 1 (GalT1), which transfers Gal from UDP-Gal to any terminal GlcNAc residue, the need for non-radioactive detection methods became necessary in order to expand the substrate scope of this PTM.[47,116,117] Work by Qasba and coworkers made this possible after they identified the key amino acids that dictated GalT1's sugar donor specificity.[118] Upon generating a single-point mutant of GalT1 (Y289L), which enlarged the donor substrate-binding pocket, the enzyme could tolerate UDP-GalNAc and structural analogues at the 2-position. Hsieh-Wilson and coworkers capitalized on this substrate tolerance by enzymatically installing either the ketone or azide chemical reporters onto *O*-GlcNAcylated substrates using the Y289L GalT1 mutant enzyme and the UDP analogue of the ketone isostere of GalNAc (UDP-2-keto-Gal, **2.14**) or UDP-GalNAz (**2.15**), respectively (Figure 2-7B).[119-121] Subsequently, reaction with an appropriate bioorthogonal enrichment probe facilitated the identification of hundreds of *O*-GlcNAcylated glycoproteins.

In the first large-scale glycoproteomics application using the chemoenzymatic tagging strategy, Hsieh-Wilson and coworkers prepared nuclear and cytosolic extracts of rat brains, incubated the lysate with UDP-2-keto-Gal and the Y289L GalT1 mutant, and then treated the samples with an aminooxy biotin tag to form an oxime linkage.[122] After tryptic digests were prepared, biotin-labeled glycopeptides were enriched by avidin chromatography and identified by tandem MS. Using this approach, 25 putatively *O*-GlcNAcylated proteins were identified, 2 of which were previously known. Later, Hart and coworkers used the same enrichment strategy to identify 35 *O*-GlcNAc sites on 25 *O*-GlcNAcylated proteins in human erythrocytes.[123] In a follow-up study, Hsieh-Wilson and coworkers monitored the dynamics of *in vivo O*-GlcNAcylation in rat brains by developing a quantitative variant of the chemoenzymatic tagging strategy.[124] In this method, rat brain lysates representing two different states (with or without neuronal stimulation) were chemoenzymatically labeled using UDP-2-keto-Gal and aminooxy biotin before proteolytic digestion. Glycopeptides were then labeled with either isotopically light or heavy versions of formaldehyde and sodium cyanoborohydride to reductively alkylate primary amines. After isotopic labeling, samples were combined and enriched using monomeric avidin chromatography. Analysis by tandem MS identified 83 *O*-GlcNAcylated peptides; 13 of these increased upon stimulation. There are, however, two limitations in this enrichment strategy, both inherent

to the use of an aminooxy probe: non-specific background labeling and slow reaction kinetics with ketones at physiological pH. To overcome these limitations, alternative bioorthogonal reactions have been utilized instead.

In addition to an oxime-forming reaction, CuAAC reactions can be coupled to chemoenzymatic tagging for improved enrichment of glycoproteins. To demonstrate this strategy, Hsieh-Wilson and coworkers incubated rat forebrain lysates with UDP-GalNAz and the Y289L GalT1 mutant before performing the CuAAC reaction with a tetramethyl-6-carboxyrhodamine (TAMRA)-alkyne probe.[119] After enrichment of TAMRA labeled proteins by immunoprecipitation and in-gel digestion of fluorescently labeled bands, 213 putatively *O*-GlcNAcylated proteins were identified by tandem MS, 67 of which were previously known. Palmisano and coworkers also utilized a chemoenzymatic strategy to append GalNAz onto *O*-GlcNAcylated peptides for enrichment and identification. However, during the chemical ligation step, the CuAAC reaction was performed with a phospho-alkyne tag so that titanium dioxide chromatography could be used to simultaneously enrich phosphopeptides and glycopeptides.[125] Using this strategy, 12 *O*-GlcNAcylated peptides (9 proteins) and 485 phosphopeptides were identified in mouse brain tissue. After optimizing conditions to favor glycopeptide enrichment by extensive dephosphorylation of lysates, 20 *O*-GlcNAcylated peptides and 23 phosphopeptides were identified.

One general challenge to both metabolic labeling and chemoenzymatic tagging is that the enrichment probe used for chemical ligation is not always ideal for tandem MS. Typically, after covalent modification with an enrichment probe, glycosylated substrates increase in mass without an accompanying increase in charge. At the same time, depending on the enrichment method (i.e., biotin/avidin), it can be difficult to efficiently release captured substrates.[126] Consequently, enrichment probes that are amenable to bioorthogonal reactions and better suited for MS have been developed. As one solution, Hart and coworkers synthesized a photocleavable biotin-alkyne probe to react with GalNAz-tagged *O*-GlcNAcylated peptides from tau-containing protein fractions from rat brain lysates. After affinity enrichment using avidin chromatography, glycopeptides were released upon exposure to UV light (365 nm) and identified by tandem MS.[127] Additionally, the photocleavage reaction reduced the mass of the tagged peptide and generated a basic aminomethyltriazole tag, both favorable features for tandem MS. This strategy resulted in two large *O*-GlcNAc proteome profiling studies: one of mitotic spindle and midbody proteins (141 glycosites in 64 glycoproteins) to investigate the interplay between *O*-GlcNAcylation and phosphorylation during mitosis,[128] and the other of cerebrocortical brain tissue (458 glycosites in 274 glycoproteins) from wild-type mice and an Alzheimers disease model.[129]

*2.3.3.2 Chemoenzymatic tagging of sialic acid*

Another glycoproteomics application using chemoenzymatic tagging focused on identifying glycoproteins targeted by a *trans*-sialidase from *Trypanosoma cruzi*.[130] A long-standing challenge in the field was identifying cell surface glycoproteins on host immune cells that act as acceptors for released sialic acid residues. The difficulty was in distinguishing between sialic acid residues transferred by the *trans*-sialidase or from those that occur naturally on a cell's surface. To address this issue, Campetella and coworkers used an azide-functionalized trisaccharide, SiaNAz-α2,3-Lac-β-OMe (Figure 2-7C, **2.16**), as the sialic acid donor and incubated it with Jurkat

cells and the *trans*-sialidase enzyme. After labeling, the Staudinger ligation was performed on live cells using a phosphine-FLAG probe. Cell lysates were prepared and FLAG-labeled glycoproteins were enriched by immunoprecipitation with an anti-FLAG antibody and then subjected to in-gel digestion with trypsin. Following MS analysis, CD45 isoforms were identified as the main acceptor glycoprotein, with integrins also acting as minor acceptors.

*2.3.3.3 Chemoenzymatic tagging of higher-order glycans*

Thus far, the enrichment of glycoproteins by chemoenzymatic tagging has been applied to a limited diversity of glycans. To expand the applicability of the technique, two chemoezymatic strategies have been developed to target *N*-acetyllactosamine (LacNAc, Gal-β-1,4-GlcNAc), a prevalent *O*-linked disaccharide. In the first approach by Qasba and coworkers, a mutant bovine α-1,3-galactosyltransferase was designed to tolerate UDP-GalNAz and UDP-2-keto-Gal, and transfer the unnatural monosaccharides onto LacNAc residues.[131] In the second approach, by Wu and coworkers, a wild-type α-1,3-fucosyltransferase from *H. pylori* was found to tolerate GDP-fucose analogs, including GDP-6-azido-fucose (GDP-FucAz), and it was used to transfer the azidosugar onto LacNAc residues.[132] Another disaccharide, fucose-α-1,2-galactose (Fuc-α-1,2-Gal), can also be chemoenzymatically tagged. Hsieh-Wilson and coworkers demonstrated that the bacterial glycosyltransferase BgtA, which transfers GalNAc from UDP-GalNAc onto the C-3 position of Gal in Fuc-α-1,2-Gal, can accept the UDP analogues of 2-keto-Gal or GalNAz as donor sugars.[133] Although these chemoenzymatic tagging strategies have not been coupled to MS for identifying glycoproteins, preliminary results suggests that it should be possible.

While chemoenzymatic tagging provides a route for high-efficiency labeling of glycoproteins with bioorthogonal chemical reporters, it has its limitations. Unlike metabolic labeling, which can be used to capture the dynamics of glycosylation *in vivo*, chemoenzymatic tagging can only report on steady-state glycosylation. This distinction should be kept in mind when selecting the appropriate chemical enrichment method for a given study.

## 2.4 Identifying Sites of Glycosylation

An important aspect to fully characterizing a functional proteome is identifying PTMs and their locations on proteins. Tandem MS has become an invaluable tool for this task, although considerable challenges remain. While computational methods can adequately identify peptides that contain a particular PTM, they are less reliable at localizing the site of modification and assigning metrics of confidence in those site assignments.[134] This is particularly true for glycosylation, where locating the modified amino acid is an essential step towards discovering the biological role of a glycan. Unfortunately, the unambiguous assignment of glycosylation sites is difficult to achieve, particularly in cases involving low protein abundance or low site occupancy, a phenomenon that is typical in complex biological samples. Adding to the technical challenge of site mapping, glycosylation reduces ionization efficiency and, in turn, detectability via MS.[126]

Most strategies for enriching for glycosylated peptides (see previous section) do not overcome the challenges associated with pinpointing the exact sites of glycosylation. In this section, we will discuss techniques to identify sites of glycosylation. In the first approach, sites of glycosylation are marked through enzymatic or chemical deglycosylation, allowing for the use of

traditional tandem MS techniques for the indirect detection of the glycan. In the next approach, advances in tandem MS fragmentation methods (i.e., electron transfer dissociation (ETD)) are utilized for the direct analysis of intact glycopeptides. In the final approach, homogenous and well-defined glycans are created to improve confidence in the site of modification and reduce the computational challenges that accompany analysis of glycopeptide tandem mass spectra.

### *2.4.1 Deglycosylation methods*

A common method for mapping sites of glycosylation relies on enrichment of glycopeptides either before or after removal of the glycan moiety, followed by tandem MS to sequence the peptides. Removing the glycan can expedite the process of site identification in a number of ways. First, after deglycosylation, peptides can be identified through traditional tandem MS methods using collision-induced dissociation (CID). In general, CID of an intact glycopeptide

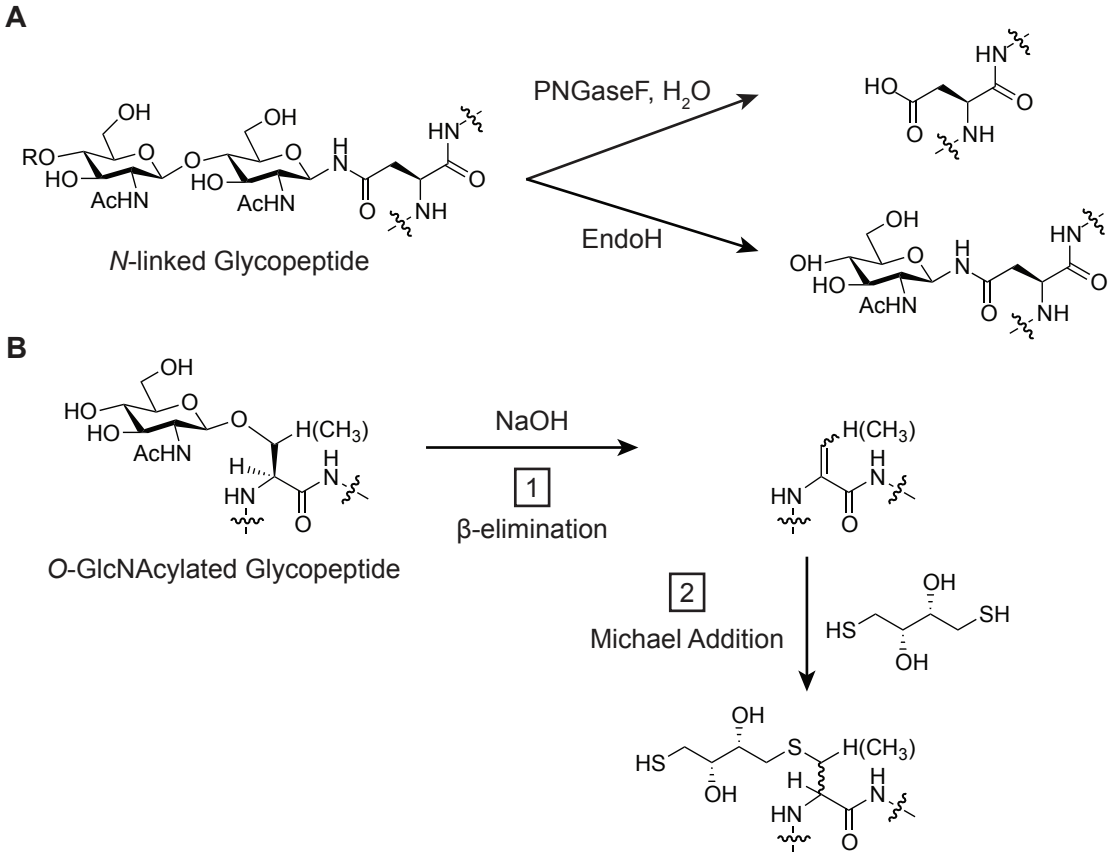



***Figure 2-8. Examples of enzymatic and chemical deglycosylation.***

(A) *N*-linked glycans can be enzymatically deglycosylated using endoglycosidases, most commonly PNGase F (top), which converts asparagine into aspartic acid after cleaving off the entire glycan. If performed in the presence of $^{18}O$-labeled water, the heavy oxygen will mark the site of glycosylation. Alternatively, EndoH cleaves between the two innermost GlcNAc resides, leaving a single peptide-bound GlcNAc residue to mark the site of glycosylation. Unlike for *N*-glycans, there is an absence of universal glycosidases that can act on all *O*-glycan structures. To circumvent this limitation, chemical deglycosylation strategies have been developed. (B) Schematic of beta-elimination followed by Michael addition with DTT (BEMAD). First, beta-elimination with a base, most commonly NaOH, converts glycosylated serine or threonine residues into dehydroalanine or dehydrobutyric acid, respectively. Second, Michael addition with amine or thiol nucleophiles, most commonly DTT, can covalently mark the site of glycosylation.

results primarily in glycan cleavage and fragmentation, leaving the peptide backbone intact. While it is possible to optimize CID conditions settings to improve peptide backbone fragmentation, a mixture of carbohydrate side chain and peptide cleavage events typically occur, complicating manual and computational assignments of tandem mass spectra.[135,136] With instruments that allow for multiple stages of tandem MS, i.e., $MS^n$ expriments, CID of the peptide ion resulting from glycan loss (i.e, $MS^3$) can be performed to gain peptide sequence information.[75] However, the decrease in fragment ion abundance with additional stages of tandem MS limits the utility of this approach to the most abundant precursor ions. Moreover, removing glycans from glycoproteins increases a protease's access to the amino acid backbone, allowing for more complete proteolytic digestion. For these reasons, deglycosylation is an important step in the identification of both *N*- and *O*-glycosites.

*2.4.1.1 Deglycosylation methods for N-glycosylation*

The most successful method for mapping *N*-glycosites relies on tandem MS analysis of an enzymatically deglycosylated *N*-linked glycopeptide. The enzymatic deamination of glycosylated Asn residues by PNGase F results in complete removal of the entire *N*-glycan and a 0.98-Da mass increase, marking the site of glycosylation for identification using high-resolution tandem MS measurements (Figure 2-8A).[137] Furthermore, the existence of a highly conserved consensus sequence for *N*-glycosylation provides a means to validate the identification of a given site. To further increase confidence in glycosite assignment, a strategy based on stable isotope labeling, where enzymatic deamination was performed in the presence of isotopically enriched $^{18}O$-water, imparting a 2.98-Da mass shift to the peptide, was developed.[138] Kaji *et al.* used this strategy to identify *N*-glycosites in *Caenorhabditis elegans* by enriching glycopeptides using the lectin ConA and then performing PNGase F cleavage in the presence of $^{18}O$-labeled water. Their method, termed isotope-coded glycosylation site-specific tagging, generated the first large-scale map of *N*-glycosites.[139] Similar strategies have been used in many subsequent studies to identify *N*-glycosites in other organisms, including mice[45] and *Trypanosoma cruzi*.[140] Unfortunately, complete removal of *N*-glycans with PNGase F can lead to instances of incorrectly mapped glycosites. During the course of PNGase F treatment, spontaneous deamination of non-glycosylated Asn residues can yield false positive identifications.[141,142]

To reduce the chances of ambiguity that can accompany indirect detection of *N*-glycosites, an alternative approach relies on the partial rather than the complete removal of *N*-glycans. For instance, treating *N*-linked glycoproteins or peptides with the enzyme endoglycosidase H (Endo H) will preserve a single *N*-linked GlcNAc residue (Figure 2-8A).[143,144] Truncated *N*-linked glycopeptides, like their completely deglycosylated counterparts, can be sequenced by tandem MS using standard CID because the *N*-glycosidic bond is stable under CID conditions.[143] Unfortunately, the presence of the single monosaccharide suppresses ionization of the glycosylated peptide,[126,127] potentially biasing data-dependent tandem MS data acquisition against glycopeptide ions in complex mixtures. In general, ion suppression due to glycosylation is a major challenge for glycosite mapping. Therefore, the introduction of strategies that are capable of selecting low-abundance ions for tandem MS will greatly benefit the field of glycoproteomics.

We recently developed a directed proteomics strategy for identifying low-abundance

species in complex mixtures and improving confidence in the site of glycosylation.[145] In this approach, the mass spectrometer is instructed to fragment specific ions, independent of their abundance. Termed isotopic signature transfer and mass pattern prediction (IsoStamp), the technique exploits the perturbing effects of a dibromide motif on a peptide's isotopic envelope

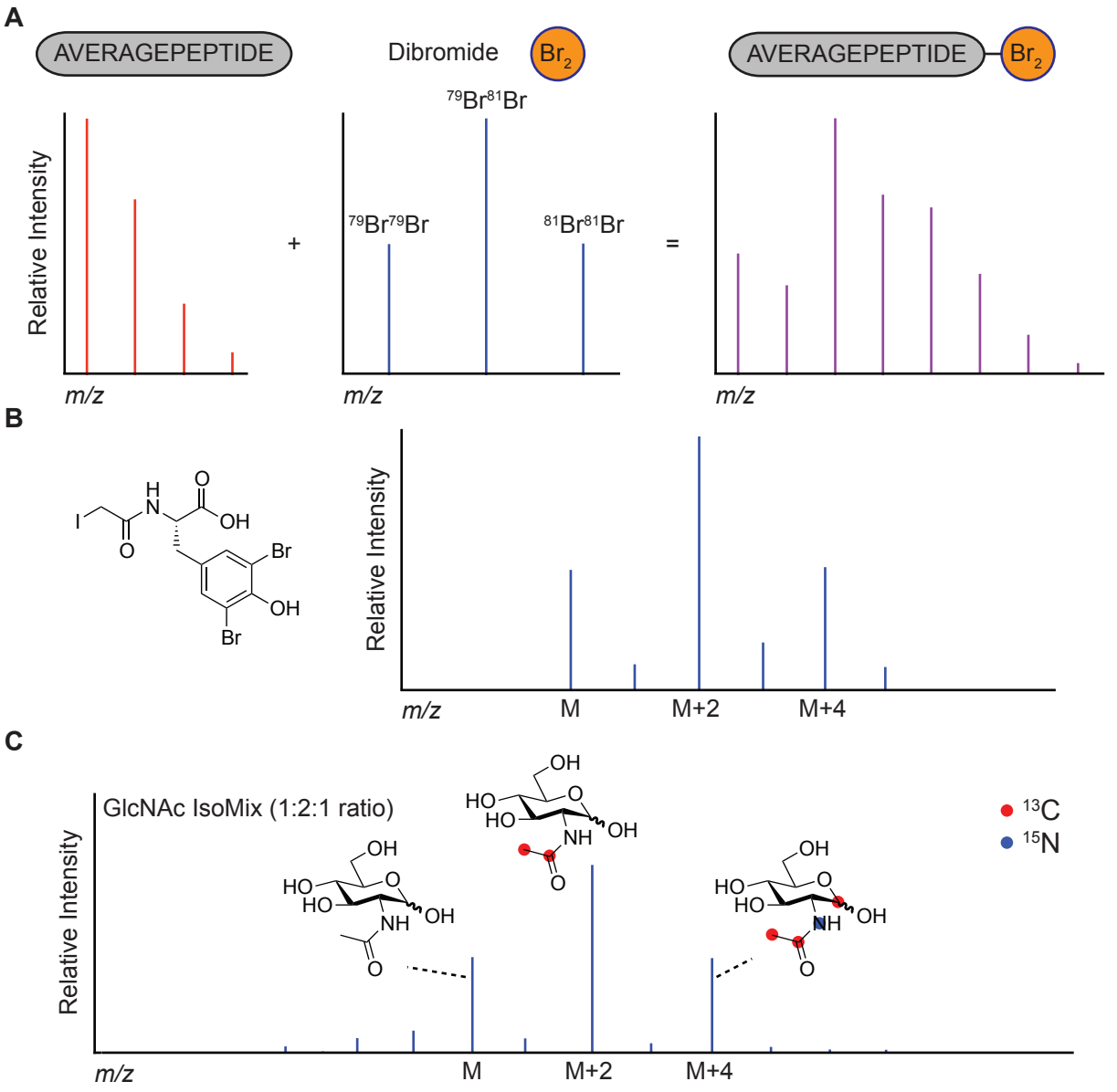



***Figure 2-9. A dibromide motif can be chemically or metabolically embedded onto a peptide, perturbing the peptide's isotopic envelope.***

(A) The isotopic envelopes of an AVERAGEPEPTIDE, a dibromide ($Br_2$), and a dibromide-labeled AVERAGEPEPTIDE were simulated to illustrate the effect of two bromine atoms on a peptide's isotopic envelope. This pattern can be recognized with high sensitivity and specificity in complex mass spectra. (B) An iodoacetamide-derivatized dibromide tag (and its mass spectrum) was used for chemical modification of peptides through alkylation of cysteine residues. The mass spectrum of the tag is shown. (C) A GlcNAc Isomix was used to metabolically install the dibromide motif directly into glycoproteins for high-confidence glycosite mapping. A three-component GlcNAc isomix was designed to mimic the 1:2:1 peak intensity distribution of the dibromide triplet pattern by adjusting the concentration of each synthesized isotopolog.

(this method is discussed in detail in Chapter 3),[146] and is an extension of isotopic distribution encoding tagging.[147] As illustrated in Figure 2-9A, the isotopic envelope of an average peptide changes significantly upon modification with two bromine atoms. Two bromine atoms produce a symmetrical triplet in a mass spectrum, with major peaks at M, M + 2 and M + 4, at a relative intensity ratio of 1:2:1, due to the abundances of the $^{79}Br_2$, $^{79}Br^{81}Br$, and $^{81}Br_2$ isotopic pairings. The isotopic envelope of the dibromide-tagged peptide then reflects the parent peptide's intrinsic isotopic distribution convoluted with (or "recoded" by) the dibromide pattern.

We developed a chemical labeling strategy to install the dibromide motif onto a peptide using a dibrominated tyrosyl iodoacetamide analog capable of alkylating cysteine residues (Figure 2-9B). This tag was then used in a proof-of-principle study to develop a computational pattern-searching algorithm capable of identifying dibromide-labeled peptides in complex samples with high sensitivity and fidelity. Following pattern identification by the algorithm, a directed proteomics analysis of dibromide-labeled peptides is performed, where inclusion lists containing *m/z* values and retention times of precursor ions bearing a recoded isotopic envelope are used to trigger tandem MS. Thus, the IsoStamp approach utilizes isotopic pattern rather than ion abundance to drive directed proteomics.

While the isotopic signature of a halogenated chemical tag can effectively highlight labeled peptides within a complex LC-MS dataset, some biomolecules are not amenable to chemospecific modification. To address this limitation and expand the utility of the IsoStamp method, we demonstrated that it is possible to impart a similar perturbation to a peptide's isotopic envelope without chemical tagging (this method is discussed in detail in Chapter 7).[148] Instead, we showed that a dibromide-like isotopic signature could be directly embedded into *N*-glycans through metabolic labeling. To mimic the dibromide pattern, a stoichiometrically defined mixture of GlcNAc isotopologs, referred to as a GlcNAc isomix (Figure 2-9C), was prepared by combining three GlcNAc isotopologs in a 1:2:1 molar ratio. The isomix was then fed to a recently generated yeast strain with an engineered salvage pathway for GlcNAc.[149] The isomix was metabolically installed into all yeast *N*-glycans by virtue of the fact that eukaryotic *N*-glycans contain a conserved GlcNAc-β1,4-GlcNAc disaccharide at their reducing end (see Figure 1-3B). Subsequently, yeast lysates were prepared and digested with either trypsin or chymotrypsin, and glycopeptides were then enriched using a modified version of FASP.[69, 168] During the process, *N*-glycans were truncated to a single GlcNAc residue with Endo H, preserving the isotopic signature of the GlcNAc isomix. Glycopeptides bearing the isomix signature were subjected to a directed proteomics workflow in which precursor ions were selected for CID using a 4-Da isolation window (IW). A broad IW was used to preserve the isomix signature in fragment ions bearing the GlcNAc isomix modification, increasing confidence in glycosite assignment in cases with multiple potential sites of glycosylation. Using this directed strategy, 133 *N*-glycosites spanning 58 glycoproteins, 50% of which were previously biochemically validated as *N*-glycosylated, were identified within the yeast proteome. The isomix method offered an enhanced level of confidence for mapping sites of glycosylation because of the unique isotopic envelope of an isomix-containing peptide.

*2.4.1.2 Deglycosylation methods for O-glycosylation*

Deglycosylation strategies would also be useful for mapping *O*-glycosites. However,

unlike for *N*-glycans, deglycosylation of *O*-glycans is less straightforward due to the absence of enymatic tools that can act on all *O*-glycan structures. To circumvent this limitation, chemical deglycosylation strategies have been developed. One method relies on periodate oxidation followed by treatment with base to eliminate monosaccharides.[150,151] Through multiple rounds of treatment, carbohydrates can be sequentially eliminated until just the amino acid backbone remains.[152] Another approach, hydrazinolysis, utilizes anhydrous hydrazine to release either all glycans (*N*- and *O*-linked), or selectively just *O*-glycans using milder reaction conditions.[153] Another strategy uses anhydrous trifluoromethanesulfonic acid (TFMSA) for either the partial or complete deglycosylation of either *N*- or *O*-linked glycans, respectively, depending on the temperature of the reaction.[154] When TFMSA is used at 0 ˚C, peripheral carbohydrates are eliminated while the peptide-linked *O*-GalNAc residues (for mucin-type *O*-glycans) or the *N*-GlcNAc residues (for *N*-glycans) are preserved. At higher temperatures (25 ˚C or greater), complete deglycoslation of *O*-glycans is possible.[155] As TFMSA is a strong acid, scavengers such as anisole can be added to the reaction mixture to prevent degradation of the peptide backbone.[155] The compatibility of these chemical deglycosylation methods with large-scale, global glycoproteomics studies remains to be proven as they have primarily been used to identify sites of glycosylation on single, purified glycoproteins or a defined glycopeptide mixture.

Beyond the chemical methods that trim or eliminate glycans from the amino acid backbone, one deglycosylation strategy has found its way into global glycosite mapping of complex protein mixtures. The method relies on base-mediated β-elimination of *O*-glycans, converting glycosylated serine or threonine residues into dehydroalanine or dehydrobutyric acid, respectively (Figure 2-8B). The resulting amino acid mass change can be used as a marker when searching tandem mass spectra.[156,157] Alternatively, the generated α,β-unsaturated carbonyl functionality can be used for Michael addition with amine or thiol nucleophiles. By replacing the eliminated glycan with a chemical tag that was stable to CID fragmentation and/or allowed for enrichment, the method was quickly applied to a number of glycosite mapping studies. Initially, the approach was demonstrated using nucleophiles that also served as the base, such as methylamine, ethylamine, and dimethylamine. In most cases, this was used for the identification of *O*-glycosites in simple mixtures of glycopeptides or in purified mucin-type glycoproteins such as MUC1.[158-160] Later, a two-step variant was developed using sodium hydroxide as the base followed by dithiothreitol (DTT) as the nucleophile. Termed β-elimination followed by Michael addition with dithiothreitol (BEMAD),[161] the approach was quickly expanded to include a variety of nucleophiles besides DTT, including butanethiol as well as biotin-pentylamine and biotin-cystamine for enrichment of the deglycosylated peptides (Figure 2-8B).[122,123,162] Differential isotopic labeling for comparative quantitative analysis of glycosites was also possible by performing Michael addition with isotopically light DTT (d0) or heavy DTT (d6).[163] The BEMAD method has primarily been used to identify sites of *O*-GlcNAcylation in samples ranging from purified protein complexes to mammalian whole cell lysates.[126] Among the challenges in implementing the BEMAD method, base treatment must be optimized to minimize nonspecific β-elimination of hydroxyl groups from serine or threonine residues as well as minimize damage to the peptide backbone. Additionally, because other *O*-linked PTMs are base labile, such as phosphorylation, careful controls and sample preparation are necessary in order to ensure that the correct PTM is identified.

Despite their employment to map sites of glycosylation, chemical deglycosylation methods suffer from a number of drawbacks. Most notably, the sometimes harsh conditions (i.e., high concentrations of acids and bases at elevated temperatures) that are needed can adversely effect amino acids in the polypeptide backbone, causing nonspecific cleavage or unknown modifications to amino acid residues. Although it is possible to optimize some of these techniques for the analysis of glycopeptides in complex mixtures, alternative methods that make use of advances in tandem MS fragmentation techniques can be utilized for the direct characterization of intact glycopeptides, thereby increasing the confidence of glycosite assignment.

### *2.4.2 Electron-transfer dissociation*

Advances in tandem MS fragmentation methods have improved the characterization of labile PTMs. With the development of fragmentation techniques that do not utilize high internal energy to fragment ions (i.e., CID), intact glycopeptides could be analyzed by tandem MS using electron-capture dissociation (ECD)[164,165] or electron-transfer dissociation (ETD).[32,166] Unlike CID, which typically results in glycan cleavage and minimal peptide backbone fragmentation, ECD and ETD can retain labile PTMs and preferentially fragment the peptide backbone. This is especially helpful in *O*-glycosite identification because under CID conditions the glycan will be eliminated before peptide fragmentation, making it difficult to identify the site of glycosylation. Using ECD or ETD, both *N*- and *O*-glycosites have been located in glycoproteins from a variety of biological samples.[24]

Although both ETD and ECD have been used for glycosite identification, ETD is especially popular because it can be used with lower-cost and widely used ion trap mass spectrometers instead of expensive Fourier transform ion cyclotron resonance mass spectrometers. When considering integration with enrichment methods, ETD has emerged as the fragmentation method of choice. In particular, for globally mapping sites of *O*-GlcNAcylation, ETD has been coupled with glycopeptide enrichment through the use of antibodies[167], LWAC[48,168-170] and chemoenzymatic tagging[124,127,128] to unambiguously map a number of *O*-GlcNAc sites in a wide range of biological samples.[126] Additionally, a global profiling of *N*-glycosites in Campylobacter jejuni was also made possible through the use of ETD.[171]

Because ETD produces fragment ions that are different from those produced by CID, optimal data collection settings and appropriate data analysis software are necessary in order to successfully collect and interpret ETD tandem mass spectra.[172] With respect to precursor ion fragmentation, ETD (and ECD) is less efficient than CID and generates more extensive fragmentation for higher charge state precursor ions with *m/z* values of less than 1000. Consequently, samples that contain glycopeptides with trimmed glycans and which ionize as high charge states are ideal.[24,28] Data analysis can be equally challenging. Darula *et. al* recently demonstrated that after reinterrograting an ETD dataset of mucin-type glycopeptides with a new version of the Protein Prospector data analysis software, which implemented a modified scoring strategy based on the expected frequency of different fragment ions in ETD spectra, they more than doubled the number of identified glycopeptides (and glyosites).[173]

### *2.4.3 Homogenous and well-defined glycans*

Mucin type *O*-glycosylation, which represents a complex form of protein glycosylation, has been notoriously difficult to study by MS. Currently, a whole glycoproteomic analysis for identifying *O*-glycosites on mucins is not possible due to the computational challenge of considering all possible glycan structures. Instead, there are several examples in which *O*-glycans from purified glycoproteins have been characterized, including those on MUC1, beta-amyloid precursor protein, and immunoglobin A1.[28,174-177] Alternatively, there have also been a few examples of semi-global *O*-glycosite mapping where lectins were used to enrich glycoproteins to generate samples with limited glycan diversity. First implemented by Medzihradszky and coworkers, they started with a tryptic digest of bovine serum proteins and enriched for O-linked glycopeptides using the lectin Jacalin. Using ETD, they were able to identify 26 *O*-glycosites; subsequently, they increased their site-list after reanalyzing the same dataset with an optimized algorithm.[173,178] Recently, the same group coupled an additional chromatography step, either ion-exchange or electrostatic repulsion hydrophilic interaction chromatography, following the lectin affinity enrichment. The additional fractionation resulted in a significant improvement to the ETD analysis, and 124 *O*-glycosites mapping to 51 glycoproteins were identified, of which 16 were previously known.[179]

As an alternative strategy, Clausen and coworkers took a genetic engineering approach to identify and study mucin-type glycosylation. They developed a cell line, termed "simple cells," which produce truncated but homogenous *O*-linked glycans.[180] "Simple cells" were generated by silencing the chaperone protein cosmic, which resulted in the production of only GalNAc-α-Ser/Thr or Sia-2,6-GalNAc-α-Ser/Thr truncated glycans. With a near homogenous population of simplified *O*-linked glycoproteins, after lectin affinity chromatography for glycopeptides and tandem MS with ETD, 350 *O*-glycosites mapping to 107 glycoproteins were identified, of which 30% were previously known. Next, the same group extended the "simple cells" technology to identify substrates of specific glycosyltransferases, focusing on the substrates of one ppGalNAcT isoform. By performing a comparative glycoproteomics experiment between wild-type and ppGalNAcT knockout "simple cells" in epithelial hepatocellular carcinoma cells, 73 unique sites of glycosylation for the GalNAc-T2 isoform were identified.[181] By eliminating the computational burden of considering numerous glycan masses at each serine or threonine residue, the "simple cell" technology facilitated *O*-linked glycopeptide identification in complex mixtures using ETD-based tandem MS.

## 2.5 Functional Glycoproteomics

An important aspect of connecting glycoproteins to cellular functions will include building interaction networks that delineate the connectivity between glycoproteins and their binding partners. One approach to address this is through the use of chemical crosslinking probes that can "fix" glycoprotein interactions at a given time and then identify those interactions with MS. However, because currently existing chemical crosslinking probes target amino acids on proteins, interactions that are dependent on glycosylation will be lost.[182] Additionally, these interactions should be studied in a physiologically relevant environment in order to discern

biologically relevant interactions.[183] Consequently, there is a need to capture glycoprotein interaction partners in a glycosylation-dependent manner and in the context of living cells.

Recently, Wollscheid and coworkers developed a general method to identify the cell surface glycoprotein receptor for a given protein ligand.[184] Termed ligand-based receptor-capture (LRC) technology, this method built upon their previously developed CSC technology.[74] The LRC technology relied on a trifunctional chemoproteomics (TRICEPS, **2.17**) probe (Figure 2-10A) that contained an *N*-hydroxysuccinimide ester for conjugation to a ligand protein, a trifluoroacetylated hydrazine for condensation with aldehydes introduced into glycans through oxidation, and a biotin moiety for affinity enrichment. Given that many cell surface proteins are glyco-

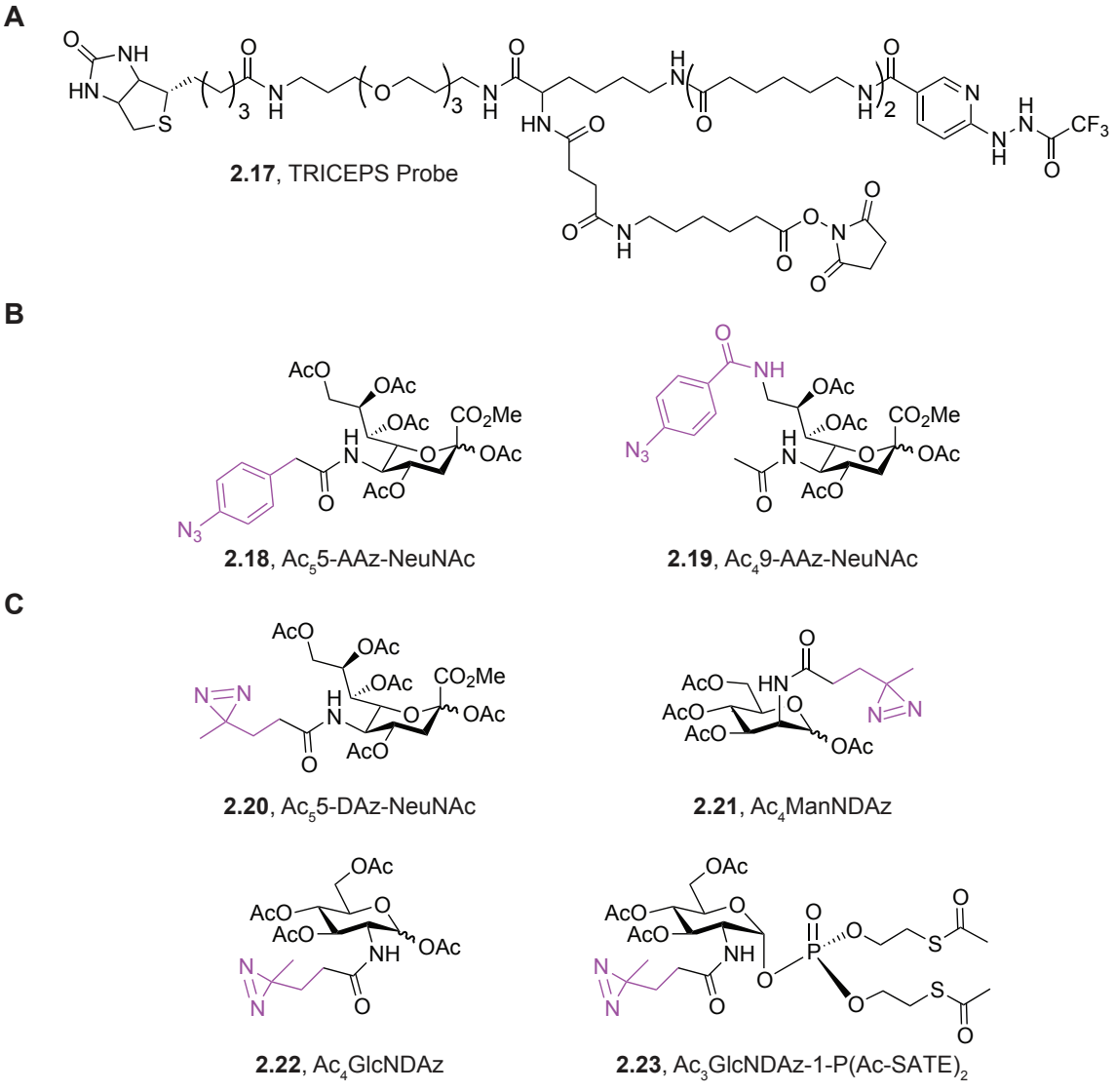



***Figure 2-10. Chemical tools for crosslinking glycoproteins or glycan-binding proteins.***

(A) A trifunctional chemoproteomics (TRICEPS) probe for capturing the cell-surface glycoprotein receptor of a specific ligand. (B) The structures of aryl azide-functionalized sialic acid residues or (C) diazirine-functionalized monosaccharides (functional groups highlighted in purple) used for metabolically labeling glycans with photocrosslinkers. Although structures are shown in their peracetylated form, after entering cells by passive diffusion, cytosolic esterases will cleave the acetyl groups.

sylated, after the TRICEPS probe was conjugated to a specific ligand it was incubated with sodium periodate-treated cells or tissue samples. This allowed ligand-receptor interactions to become stabilized through covalent capture between the hydrazine functionality on the TRICEPS probe and aldehydes introduced on the receptor's oxidized glycans. After labeling, cells were lysed, digested with trypsin, and TRICEPS-modified glycopeptides were captured with streptavidin beads. Following the release of *N*-linked glycopeptides by PNGase F treatment and analysis by quantitative tandem MS, the enrichment of a receptor, relative to a negative control sample, indicated a true ligand-receptor interaction. To validate and demonstrate the versatility of LRC technology, a number of well-characterized ligands, including insulin, transferrin, apelin-17, epidermal growth factor, and the antibody trastuzumab were conjugated to the TRICEPS probe, and in each case only the known cell surface receptor was enriched and identified by MS. To discover new interactions, LRC technology was used to identify cell surface receptors targeted by vaccinia virus. After the intact virus was conjugated to the TRICEPS probe and incubated with HeLa cells, cell surface glycoproteins were enriched and analyzed by tandem MS. The process identified seven receptor proteins, four of which were novel. Collectively, the LRC technology provided a route towards identifying the specific glycoprotein receptor for a given protein ligand.

### *2.5.1 Glycan-binding proteins*

In addition to building glycoprotein networks based upon protein-protein interactions, there is also a need to identify glycan-binding proteins. Unfortunately, many glycan-mediated complexes have low affinities and are short-lived, which has made traditional purification techniques impractical. Instead, the formation of covalent complexes between a glycan and its binding protein, while still in a native environment, has provided an alternative approach to access these biologically important interactions. At the forefront of this technology are photocrosslinking reagents, which can be divided into two categories: (1) synthetic glycans or glycopeptides with photoreactive crosslinkers added exogenously to cells, and (2) metabolic incorporation of photocrosslinking monosaccharides. Early work in this area focused primarily on the first approach, where glycans, glycopeptides, or glycolipid probes containing a photocrosslinking group were synthesized and then used to characterize or identify glycan-binding proteins, including the asialoglycoprotein receptor, the adhesion factor BabA on *H. pylori*, and the B-subunit of cholera toxin.[185-187] Additional efforts in this area are summarized in two excellent reviews.[20,188] In this section I will focus on the second approach, where unnatural monosaccharides functionalized with a photocrosslinking moiety can be introduced directly into glycans through MOE (Figure 2-10). Importantly, by placing photocrosslinking reagents at specific sites within cellular glycoconjugates, crosslinking can take place in the biologically relevant environment of the glycan-binding protein with minimal perturbation.[183]

Situated at the periphery of complex glycans, sialic acid residues are known to participate in many biological recognition events, acting as sentinels for carbohydrate-binding proteins that sample the cell surface for their specific binding partner.[189] Consequently, sialic acid residues have served as an ideal target for the metabolic incorporation of a photocrosslinking moiety. Specifically, two different sialic acid analogs bearing an aryl-azide photocrosslinker have been synthesized and metabolically incorporated into cellular glycans (Figure 2-10B).

The first, reported by Bertozzi coworkers, was per-*O*-acetylated 5-aryl azide-*N*-acetylneruaminic acid ($Ac_5$-5-AAz-NeuNAc, **2.18**).[101] Later, Paulson and coworkers also functionalized sialic acid with an AAz, synthesizing per-*O*-acetylated 9-aryl azide-*N*-acetylneruaminic acid ($Ac_4$-9-AAz-NeuNAc, **2.19**).[100] The photocrosslinking capabilities of this molecule were used to identify the *cis* ligands of the glycoprotein CD22, which is a sialic acid-binding protein (or siglec) found on the surface of B cells. Although early *in vitro* assays identified two glycoproteins that bound CD22 in a glycan-dependent manner (i.e., CD45 and surface IgM (sIgM)), the Paulson group sought to identify CD22's glycan-dependent *cis* ligands in a native cellular environment. After culturing B cells with $Ac_4$-9-AAz-NeuNAc and treating the sample to ultraviolet (UV) irradiation, which activated the photocrosslinker to bind to any nearby sialic acid-binding proteins, cell lysates were prepared and CD22 was immunoprecipitated. Surprisingly, immunoblotting for CD45 or sIgM, the previously reported glycoproteins that bound CD22, were negative. Instead, on the cell surface, CD22 bound neighboring CD22 molecules, forming homomultimeric complexes.

To integrate this glycan-mediated photocrosslinking technology with MS, Paulson and coworkers next identified the *trans* ligands of CD22.[190] After incubating a CD22 fusion protein with B cells that were metabolically labeled with 9-AAz-NeuNAc, the glycan-dependent *trans* ligands of CD22 were covalently captured after UV irradiation. Following immunoprecipitation of CD22 protein complexes, quantitative tandem MS identified 27 candidate glycoproteins. After a series of validation experiments were performed in a native cellular environment, only a subset of these glycoproteins (i.e., CD45, sIgM, and Basigin) were found to be the true *trans* ligands for CD22. Collectively, these results highlighted the utility of photocrosslinking reagents to resolve glycan-dependent protein interaction networks. Importantly, the placement of a crosslinker directly into a glycan made it possible to identify these interactions within their native cellular environment.

Although photocrosslinking monosaccharides allows for direct access to glycan-mediated interactions, the size and position of the photocrosslinkinvg group are important considerations. For example, large photocrosslinkers on monosaccharides can impact glycan recognition and metabolic incorporation efficiencies.[191,192] To address these issues, Kohler and coworkers synthesized a number of alternative photoactivatable crosslinking monosaccharides by exploiting the diazirine (Figure 2-10C), which is a small photoactivatable functional group.[188] Specifically, the diazirine was introduced onto the *N*-acyl side chains of both sialic acid and ManNAc, generating two new unnatural monosaccharides for metabolic labeling, per-*O*-acetylated 5-diazirine-*N*-acetylneruaminic acid ($Ac_5$-5-DAz-NeuNAc, **2.20**) and per-*O*-acetylated *N*-diazirine-acetylmannosamine ($Ac_4$ManNDAz, **2.21**), respectively.[193,194] After demonstrating that both monosaccharides could be efficiently incorporated into cell surface glycoproteins in the form of sialic acid, the Kohler group validated the photocrosslinking capabilities of the diazirine. First, analogous to the Paulson work, they observed that metabolic labeling of B cells with either $Ac_5$-5-DAz-NeuNAc or $Ac_4$ManNDAz, followed by UV-irradiation, resulted in CD22 homomultimeric complexes.[194] Next, the Kohler group used their diazirine-functionalized monosaccharides to crosslink ganglioside-protein interactions.[195,196] After culturing Jurkat cells with $Ac_4$ManNDAz, adding a known ganglioside binding protein, cholera toxin subunit B (CTxB), and then UV irradiating cells, a new toxin-containing species was formed. Following MS analysis, the crosslinked

complex was identified as CTxB bound to the ganglioside GM1, as expected.

More recently, the Kohler group attempted to introduce the diazirine functionality onto the *N*-acyl side chain of GlcNAc, synthesizing per-*O*-acetylated *N*-diazirine-acetylglycosamine ($Ac_4$GlcNDAz, **2.22**).[197] Unfortunately, they discovered that GlcNDAz was not a compatible substrate for the GlcNAc salvage pathway because of two metabolic barriers. To overcome these barriers they had to introduce a diazirine-functionalized phosphosugar (GlcNDAz-1-P) where the phosphate was protected with S-acetyl-2-thioethyl (Ac-SATE) as well as create a point mutant in the enzyme (AGX1) that converts the phosphosugar into a UDP-sugar. Using this combination of synthetic chemistry and genetic engineering, they demonstrated that HeLa cells cultured with $Ac_3$GlcNDAz-1-P(Ac-SATE)$_2$ (**2.23**) could produce *O*-GlcNDAz-modified proteins. In the future, the Kohler group plans to use these photoreactive monosaccharides for global-scale proteomic experiments.

Although photocrosslinking monosaccharides have been used in relatively few studies and in limited-scale proteomics experiments, their ability to capture glycan-dependent interactions within a native cellular environment is a unique characteristic. By increasing the diversity of monosaccharides functionalized with crosslinkers, a wider range of glycan-dependent interaction partners can be identified. Furthermore, by improving the computational tools necessary to analyze proteomic datasets from cross-linked proteomes,[198] it will become possible to identify the sites of crosslinking within the glycan-binding protein.

## 2.6 Summary and Outlook

Chemical tools have enabled tremendous growth in glycobiology, reducing experimental barriers that once restricted access to studying protein glycosylation. In particular, chemical proteomics has allowed the glycobiology community to identify glycoproteins and glycan structures associated with a number of biological processes. Continued advances in the field will be enabled by utilizing tools of genetics, biochemistry, chemical synthesis, computer science, and even physics, as interdisciplinary approaches often prove most impactful in the development of new techniques. For example, mining microbial genomic information may lead to the discovery of uncharacterized glycosidases capable of simplifying *O*-linked glycans.[199] The synthesis of new chemical probes for enrichment that feature characteristics ideal for tandem MS (i.e., fixed charge, hydrophobicity, and low molecular weight) may facilitate the identification of previously unknown glycoproteins. The development of new mass spectrometers and tandem MS technologies may allow techniques such as ETD to become more sensitive and routine so that a greater part of the research community can have access to these tools.

As the glycoproteomics toolbox grows with these developments, so will the possible applications and solvable questions. For example, one major obstacle in the field is the inability to accurately inventory all glycoforms of a glycoprotein in a given biological system. This will require major advancements in both instrumentation and computational tools such that both the glycan and the amino acid composition of a glycopeptide can be analyzed simultaneously. Another significant achievement would be integrating spatial resolution into glycoproteomics datasets. Although MS is inherently a high-throughput technology capable of identifying hundreds of proteins in a given sample with high accuracy and reproducibility, the spatial rela-

tionship between those proteins is typically lost. In the case of glycobiology, currently it is only possible to either identify the cellular location of a few glycoproteins using fluorescently labeled antibodies or lectins (i.e., immunohistochemistry), or identify the cellular location of classes of glycosylation per sample (i.e., imaging glycans with fluorescently labeled chemical reporters). However, it would be very useful to be able to simultaneously identify a glycoprotein (with its glycan) and its specific cellular localization. This will most likely require the development of novel chemical tools that can integrate with MS-imaging technology.[200,201] Once achieved, it will become possible to monitor the dynamics of glycoprotein restructuring, with respect to cellular localization, over the course of development, external stimulation, etc.

Clearly, there are many opportunities to contribute towards the field of glycobiology. In particular, there is a bright future for chemical glycoproteomics, and in the next several chapters I will describe a collection of chemical and computational tools that I have developed with this goal in mind.

## 2.7 References

1. Cravatt B.F., Simon G.M. and Yates J.R., 3rd (**2007**) The biological impact of mass-spectrometry-based proteomics. *Nature* 450, 991-1000.

2. Moellering R.E. and Cravatt B.F. (**2012**) How Chemoproteomics Can Enable Drug Discovery and Development. *Chem Biol* 19, 11-22.

3. Sabido E., Selevsek N. and Aebersold R. (**2011**) Mass spectrometry-based proteomics for systems biology. *Curr Opin Biotechnol*.

4. Beck M., Claassen M. and Aebersold R. (**2011**) Comprehensive proteomics. *Curr Opin Biotechnol* 22, 3-8.

5. Domon B. and Aebersold R. (**2006**) Mass spectrometry and protein analysis. *Science* 312, 212-217.

6. Han X., Aslanian A. and Yates J.R., 3rd (**2008**) Mass spectrometry for proteomics. *Curr Opin Chem Biol* 12, 483-490.

7. Aebersold R. and Mann M. (**2003**) Mass spectrometry-based proteomics. *Nature* 422, 198-207.

8. Steen H. and Mann M. (**2004**) The ABC's (and XYZ's) of peptide sequencing. *Nat Rev Mol Cell Biol* 5, 699-711.

9. Eng J.K., Searle B.C., Clauser K.R. and Tabb D.L. (**2011**) A face in the crowd: recognizing peptides through database search. *Mol Cell Proteomics* 10, R111 009522.

10. Khan M.F. *et al.* (**2011**) Proteomics by mass spectrometry--go big or go home? *J Pharm Biomed Anal* 55, 832-841.

11. Uttamchandani M., Li J., Sun H. and Yao S.Q. (**2008**) Activity-based protein profiling: new developments and directions in functional proteomics. *Chembiochem* 9, 667-675.

12. Witze E.S., Old W.M., Resing K.A. and Ahn N.G. (**2007**) Mapping protein post-translational modifications with mass spectrometry. *Nat Methods* 4, 798-806.

13. Jensen O.N. (**2006**) Interpreting the protein language using proteomics. *Nat Rev Mol Cell Biol* 7, 391-403.

14. Heal W.P. and Tate E.W. (**2010**) Getting a chemical handle on protein post-translational modification. *Org Biomol Chem* 8, 731-738.

15. Tate E.W. (**2008**) Recent advances in chemical proteomics: exploring the post-translational proteome. *J Chem Biol* 1, 17-26.

16. Bertozzi C.R. and Kiessling L.L. (**2001**) Chemical glycobiology. *Science* 291, 2357-2364.

17. Prescher J.A. and Bertozzi C.R. (**2006**) Chemical technologies for probing glycans. *Cell* 126, 851-854.

18. Kiessling L.L. and Splain R.A. (**2010**) Chemical approaches to glycobiology. *Annu Rev Biochem* 79, 619-653.

19. Rakus J.F. and Mahal L.K. (**2011**) New technologies for glycomic analysis: toward a systematic understanding of the glycome. *Annu Rev Anal Chem (Palo Alto Calif)* 4, 367-392.

20. Whitman C.M., Bond M.R. and Kohler J.J. (**2010**) *6.07 - Chemical Glycobiology* in Comprehensive Natural Products II. 175-224 (Elsevier, Oxford).

21. Mulloy B., Hart G.W. and Stanley P. (**2009**) *Structural Analysis of Glycans* in Essentials of Glycobiology, Edn. 2nd (Cold Spring Harbor (NY)).

22. Morelle W., Canis K., Chirat F., Faid V. and Michalski J.C. (**2006**) The use of mass spectrometry for the proteomic analysis of glycosylation. *Proteomics* 6, 3993-4015.

23. Budnik B.A., Lee R.S. and Steen J.A. (**2006**) Global methods for protein glycosylation analysis by mass spectrometry. *Biochim Biophys Acta* 1764, 1870-1880.

24. Wuhrer M., Catalina M.I., Deelder A.M. and Hokke C.H. (**2007**) Glycoproteomics based on

tandem mass spectrometry of glycopeptides. *J Chromatogr B Analyt Technol Biomed Life Sci* 849, 115-128.

25. Wei X. and Li L. (**2009**) Comparative glycoproteomics: approaches and applications. *Brief Funct Genomic Proteomic* 8, 104-113.

26. Pan S., Chen R., Aebersold R. and Brentnall T.A. (**2011**) Mass spectrometry based glycoproteomics--from a proteomics perspective. *Mol Cell Proteomics* 10, R110 003251.

27. Zhang Y., Yin H. and Lu H. (**2012**) Recent progress in quantitative glycoproteomics. *Glycoconj J*.

28. Nilsson J., Halim A., Grahn A. and Larson G. (**2012**) Targeting the glycoproteome. *Glycoconj J*.

29. Novotny M.V., Alley W.R., Jr. and Mann B.F. (**2012**) Analytical glycobiology at high sensitivity: current approaches and directions. *Glycoconj J*.

30. McDonald W.H. and Yates J.R., 3rd (**2003**) Shotgun proteomics: integrating technologies to answer biological questions. *Curr Opin Mol Ther* 5, 302-309.

31. Eng J.K., McCormack A.L. and Yates J.R., 3rd (**1994**) An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *J Am Soc Mass Spectrom* 5, 976-989.

32. Wiesner J., Premsler T. and Sickmann A. (**2008**) Application of electron transfer dissociation (ETD) for the analysis of posttranslational modifications. *Proteomics* 8, 4466-4483.

33. Anderson N.L. and Anderson N.G. (**2002**) The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* 1, 845-867.

34. Picotti P., Aebersold R. and Domon B. (**2007**) The implications of proteolytic background for shotgun proteomics. *Mol Cell Proteomics* 6, 1589-1598.

35. Tissot B. *et al.* (**2009**) Glycoproteomics: past, present and future. *FEBS Lett* 583, 1728-1735.

36. Dallas D.C., Martin W.F., Hua S. and German J.B. (**2012**) Automated glycopeptide analysis--review of current state and future directions. *Brief Bioinform*.

37. Marino K., Bones J., Kattla J.J. and Rudd P.M. (**2010**) A systematic approach to protein glycosylation analysis: a path through the maze. *Nat Chem Biol* 6, 713-723.

38. An H.J., Froehlich J.W. and Lebrilla C.B. (**2009**) Determination of glycosylation sites and site-specific heterogeneity in glycoproteins. *Curr Opin Chem Biol* 13, 421-426.

39. Zaia J. (**2008**) Mass spectrometry and the emerging field of glycomics. *Chem Biol* 15, 881-892.

40. Dreyfuss J.M. *et al.* (**2011**) Targeted analysis of glycomics liquid chromatography/mass spectrometry data. *Anal Bioanal Chem* 399, 727-735.

41. Bielik A.M. and Zaia J. (**2010**) Historical overview of glycoanalysis. *Methods Mol Biol* 600, 9-30.

42. Abbott K.L. and Pierce J.M. (**2010**) Lectin-based glycoproteomic techniques for the enrichment and identification of potential biomarkers. *Methods Enzymol* 480, 461-476.

43. Drake P.M. *et al.* (**2010**) Sweetening the pot: adding glycosylation to the biomarker discovery equation. *Clin Chem* 56, 223-236.

44. Drake P.M. *et al.* (**2011**) A lectin affinity workflow targeting glycosite-specific, cancer-related carbohydrate structures in trypsin-digested human plasma. *Anal Biochem* 408, 71-85.

45. Zielinska D.F., Gnad F., Wisniewski J.R. and Mann M. (**2010**) Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. *Cell* 141, 897-907.

46. Durham M. and Regnier F.E. (**2006**) Targeted glycoproteomics: serial lectin affinity chromatography in the selection of O-glycosylation sites on proteins from the human blood proteome. *J Chromatogr A* 1132, 165-173.

47. Roquemore E.P., Chou T.Y. and Hart G.W. (**1994**) Detection of O-linked N-acetylglucosamine (O-GlcNAc) on cytoplasmic and nuclear proteins. *Methods Enzymol* 230, 443-460.

48. Vosseller K. *et al.* (**2006**) O-linked N-acetylglucosamine proteomics of postsynaptic density preparations using lectin weak affinity chromatography and mass spectrometry. *Mol Cell Proteomics* 5, 923-934.

49. Teo C.F. *et al.* (**2010**) Glycopeptide-specific monoclonal antibodies suggest new roles for O-GlcNAc. *Nat Chem Biol* 6, 338-343.

50. Zachara N.E., Vosseller K. and Hart G.W. (**2011**) Detection and analysis of proteins modified by O-linked N-acetylglucosamine. *Curr Protoc Protein Sci* Chapter 12, Unit12 18.

51. Leitner A. and Lindner W. (**2004**) Current chemical tagging strategies for proteome anal-

ysis by mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci* 813, 1-26.

52. Beatty K.E. (**2011**) Chemical strategies for tagging and imaging the proteome. *Mol Biosyst* 7, 2360-2367.
53. Gygi S.P. *et al.* (**1999**) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* 17, 994-999.

54. Ross P.L. *et al.* (**2004**) Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Molecular & Cellular Proteomics* 3, 1154-1169.

55. Willems L.I. *et al.* (**2011**) Bioorthogonal chemistry: applications in activity-based protein profiling. *Acc Chem Res* 44, 718-729.

56. Heal W.P., Dang T.H. and Tate E.W. (**2011**) Activity-based probes: discovering new biology and new drug targets. *Chem Soc Rev* 40, 246-257.

57. Jessani N. and Cravatt B.F. (**2004**) The development and application of methods for activity-based protein profiling. *Curr Opin Chem Biol* 8, 54-59.

58. Bond M.R. and Kohler J.J. (**2007**) Chemical methods for glycoprotein discovery. *Curr Opin Chem Biol* 11, 52-58.

59. Sletten E.M. and Bertozzi C.R. (**2009**) Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. *Angew Chem Int Ed Engl* 48, 6974-6998.

60. Prescher J.A. and Bertozzi C.R. (**2005**) Chemistry in living systems. *Nat Chem Biol* 1, 13-21.

61. Rawn J.D. and Lienhard G.E. (**1974**) The binding of boronic acids to chymotrypsin. *Biochemistry* 13, 3124-3130.

62. Sparbier K., Koch S., Kessler I., Wenzel T. and Kostrzewa M. (**2005**) Selective isolation of glycoproteins and glycopeptides for MALDI-TOF MS detection supported by magnetic particles. *J Biomol Tech* 16, 407-413.

63. Xu Y.W. *et al.* (**2009**) Highly Specific Enrichment of Glycopeptides Using Boronic Acid-Functionalized Mesoporous Silica. *Anal Chem* 81, 503-508.

64. Sparbier K., Wenzel T. and Kostrzewa M. (**2006**) Exploring the binding profiles of ConA, boronic acid and WGA by MALDI-TOF/TOF MS and magnetic particles. *J Chromatogr B Analyt Technol Biomed Life Sci* 840, 29-36.

65. Zhang H., Li X.J., Martin D.B. and Aebersold R. (**2003**) Identification and quantification of

N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotechnol* 21, 660-666.

66. Tian Y., Zhou Y., Elliott S., Aebersold R. and Zhang H. (**2007**) Solid-phase extraction of N-linked glycopeptides. *Nat Protoc* 2, 334-339.

67. Liu T. *et al.* (**2005**) Human plasma N-glycoproteome analysis by immunoaffinity subtraction, hydrazide chemistry, and mass spectrometry. *J Proteome Res* 4, 2070-2080.

68. Pan S. *et al.* (**2006**) Identification of glycoproteins in human cerebrospinal fluid with a complementary proteomic approach. *J Proteome Res* 5, 2769-2779.

69. Sun B. *et al.* (**2007**) Shotgun glycopeptide capture approach coupled with mass spectrometry for comprehensive glycoproteomics. *Mol Cell Proteomics* 6, 141-149.

70. Zhou Y., Aebersold R. and Zhang H. (**2007**) Isolation of N-linked glycopeptides from plasma. *Anal Chem* 79, 5826-5837.

71. Danzer C. *et al.* (**2012**) Comprehensive description of the N-glycoproteome of mouse pancreatic beta-cells and human islets. *J Proteome Res* 11, 1598-1608.

72. Lewandrowski U., Moebius J., Walter U. and Sickmann A. (**2006**) Elucidation of N-glycosylation sites on human platelet proteins: a glycoproteomic approach. *Mol Cell Proteomics* 5, 226-233.

73. Ramachandran P. *et al.* (**2006**) Identification of N-linked glycoproteins in human saliva by glycoprotein capture and mass spectrometry. *J Proteome Res* 5, 1493-1503.

74. Wollscheid B. *et al.* (**2009**) Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins. *Nat Biotechnol* 27, 378-386.

75. Nilsson J. *et al.* (**2009**) Enrichment of glycopeptides for glycan structure and attachment site identification. *Nat Methods* 6, 809-811.

76. Zeng Y., Ramya T.N., Dirksen A., Dawson P.E. and Paulson J.C. (**2009**) High-efficiency labeling of sialylated glycoproteins on living cells. *Nat Methods* 6, 207-209.

77. Gundry R.L. *et al.* (**2009**) The mouse C2C12 myoblast cell surface N-linked glycoproteome: identification, glycosite occupancy, and membrane orientation. *Mol Cell Proteomics* 8, 2555-2569.

78. Gundry R.L. *et al.* (**2012**) A cell surfaceome map for immunophenotyping and sorting pluripotent stem cells. *Mol Cell Proteomics* 11, 303-316.

79. Bausch-Fluck D., Hofmann A. and Wollscheid B. (**2012**) Cell surface capturing technologies for the surfaceome discovery of hepatocytes. *Methods Mol Biol* 909, 1-16.

80. Hofmann A. *et al.* (**2010**) Proteomic cell surface phenotyping of differentiating acute myeloid leukemia cells. *Blood* 116, e26-34.

81. Halim A., Nilsson J., Ruetschi U., Hesse C. and Larson G. (**2012**) Human urinary glycoproteomics; attachment site specific analysis of N- and O-linked glycosylations by CID and ECD. *Mol Cell Proteomics* 11, M111 013649.

82. Kurogochi M. *et al.* (**2010**) Sialic acid-focused quantitative mouse serum glycoproteomics by multiple reaction monitoring assay. *Mol Cell Proteomics* 9, 2354-2368.

83. Ramya T.N., Weerapana E., Cravatt B.F. and Paulson J.C. (**2012**) Glycoproteomics enabled by tagging sialic acid or galactose terminated glycans. *Glycobiology*.

84. Klement E., Lipinszki Z., Kupihar Z., Udvardy A. and Medzihradszky K.F. (**2010**) Enrichment of O-GlcNAc modified proteins by the periodate oxidation-hydrazide resin capture approach. *J Proteome Res* 9, 2200-2206.

85. Taga Y., Kusubata M., Ogawa-Goto K. and Hattori S. (**2012**) Development of a Novel Method for Analyzing Collagen O-glycosylations by Hydrazide Chemistry. *Mol Cell Proteomics* 11, M111 010397.

86. Campbell C.T., Sampathkumar S.G. and Yarema K.J. (**2007**) Metabolic oligosaccharide engineering: perspectives, applications, and future directions. *Mol Biosyst* 3, 187-194.

87. Kayser H. *et al.* (**1992**) Biosynthesis of a nonphysiological sialic acid in different rat organs, using N-propanoyl-D-hexosamines as precursors. *J Biol Chem* 267, 16934-16938.

88. Mahal L.K., Yarema K.J. and Bertozzi C.R. (**1997**) Engineering chemical reactivity on cell surfaces through oligosaccharide biosynthesis. *Science* 276, 1125-1128.

89. Laughlin S.T. and Bertozzi C.R. (**2009**) Imaging the glycome. *Proc Natl Acad Sci U S A* 106, 12-17.

90. Saxon E. and Bertozzi C.R. (**2000**) Cell surface engineering by a modified Staudinger reaction. *Science* 287, 2007-2010.

91. Rostovtsev V.V., Green L.G., Fokin V.V. and Sharpless K.B. (**2002**) A stepwise huisgen cycloaddition process: copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angew Chem Int Ed Engl* 41, 2596-2599.

92. Tornoe C.W., Christensen C. and Meldal M. (**2002**) Peptidotriazoles on solid phase: [1,2,3]-triazoles by regiospecific copper(i)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *J Org Chem* 67, 3057-3064.

93. Agard N.J., Baskin J.M., Prescher J.A., Lo A. and Bertozzi C.R. (**2006**) A comparative study of bioorthogonal reactions with azides. *ACS Chem Biol* 1, 644-648.

94. Agard N.J., Prescher J.A. and Bertozzi C.R. (**2004**) A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *J Am Chem Soc* 126, 15046-15047.

95. Vocadlo D.J., Hang H.C., Kim E.J., Hanover J.A. and Bertozzi C.R. (**2003**) A chemical approach for identifying O-GlcNAc-modified proteins in cells. *Proc Natl Acad Sci U S A* 100, 9116-9121.

96. Zaro B.W., Yang Y.Y., Hang H.C. and Pratt M.R. (**2011**) Chemical reporters for fluorescent detection and identification of O-GlcNAc-modified proteins reveal glycosylation of the ubiquitin ligase NEDD4-1. *Proc Natl Acad Sci U S A* 108, 8146-8151.

97. Gurcel C. *et al.* (**2008**) Identification of new O-GlcNAc modified proteins using a click-chemistry-based tagging. *Anal Bioanal Chem* 390, 2089-2097.

98. Prescher J.A., Dube D.H. and Bertozzi C.R. (**2004**) Chemical remodelling of cell surfaces in living animals. *Nature* 430, 873-877.

99. Hsu T.L. *et al.* (**2007**) Alkynyl sugar analogs for the labeling and visualization of glycoconjugates in cells. *Proc Natl Acad Sci U S A* 104, 2614-2619.

100. Han S., Collins B.E., Bengtson P. and Paulson J.C. (**2005**) Homomultimeric complexes of CD22 in B cells revealed by protein-glycan cross-linking. *Nat Chem Biol* 1, 93-97.

101. Luchansky S.J., Goon S. and Bertozzi C.R. (**2004**) Expanding the diversity of unnatural cell-surface sialic acids. *Chembiochem* 5, 371-374.

102. Hang H.C., Yu C., Kato D.L. and Bertozzi C.R. (**2003**) A metabolic labeling approach toward proteomic analysis of mucin-type O-linked glycosylation. *Proc Natl Acad Sci U S A* 100, 14846-14851.

103. Boyce M. *et al.* (**2011**) Metabolic cross-talk allows labeling of O-linked beta-N-acetylglucosamine-modified proteins via the N-acetylgalactosamine salvage pathway. *Proc Natl Acad Sci U S A* 108, 3141-3146.

104. Dube D.H., Prescher J.A., Quang C.N. and Bertozzi C.R. (**2006**) Probing mucin-type O-linked

glycosylation in living animals. *Proc Natl Acad Sci U S A* 103, 4819-4824.

105. Rabuka D., Hubbard S.C., Laughlin S.T., Argade S.P. and Bertozzi C.R. (**2006**) A chemical reporter strategy to probe glycoprotein fucosylation. *J Am Chem Soc* 128, 12078-12079.

106. Sawa M. *et al.* (**2006**) Glycoproteomic probes for fluorescent imaging of fucosylated glycans in vivo. *Proc Natl Acad Sci U S A* 103, 12371-12376.

107. Sprung R. *et al.* (**2005**) Tagging-via-substrate strategy for probing O-GlcNAc modified proteins. *J Proteome Res* 4, 950-957.

108. Nandi A. *et al.* (**2006**) Global identification of O-GlcNAc-modified proteins. *Anal Chem* 78, 452-458.

109. Hanson S.R. *et al.* (**2007**) Tailored glycoproteomics and glycan site mapping using saccharide-selective bioorthogonal probes. *J Am Chem Soc* 129, 7266-7267.

110. Yang L., Nyalwidhe J.O., Guo S., Drake R.R. and Semmes O.J. (**2011**) Targeted identification of metastasis-associated cell-surface sialoglycoproteins in prostate cancer. *Mol Cell Proteomics* 10, M110 007294.

111. Pan P.W. *et al.* (**2012**) Cell surface glycoprotein profiling of cancer cells based on bioorthogonal chemistry. *Anal Bioanal Chem* 403, 1661-1670.

112. Hubbard S.C., Boyce M., McVaugh C.T., Peehl D.M. and Bertozzi C.R. (**2011**) Cell surface glycoproteomic analysis of prostate cancer-derived PC-3 cells. *Bioorg Med Chem Lett* 21, 4945-4950.

113. Unlu M., Morgan M.E. and Minden J.S. (**1997**) Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* 18, 2071-2077.

114. Boyce M. *et al.* A novel chemical glycoproteomics platform reveals O-GlcNAcylation of mitochondrial voltage-dependent anion channel 2. *Cell Reports (Submitted)*.

115. Liu T.W. *et al.* (**2012**) A chemoenzymatic approach toward the identification of fucosylated glycoproteins and mapping of N-glycan sites. *Glycobiology* 22, 630-637.

116. Hayes B.K., Greis K.D. and Hart G.W. (**1995**) Specific isolation of O-linked N-acetylglucosamine glycopeptides from complex mixtures. *Anal Biochem* 228, 115-122.

117. Torres C.R. and Hart G.W. (**1984**) Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes. Evidence for O-linked

GlcNAc. *J Biol Chem* 259, 3308-3317.

118. Ramakrishnan B. and Qasba P.K. (**2002**) Structure-based design of beta 1,4-galactosyltransferase I (beta 4Gal-T1) with equally efficient N-acetylgalactosaminyltransferase activity: point mutation broadens beta 4Gal-T1 donor specificity. *J Biol Chem* 277, 20833-20839.

119. Clark P.M. *et al.* (**2008**) Direct in-gel fluorescence detection and cellular imaging of O-GlcNAc-modified proteins. *J Am Chem Soc* 130, 11576-11577.

120. Khidekel N. *et al.* (**2003**) A chemoenzymatic approach toward the rapid and sensitive detection of O-GlcNAc posttranslational modifications. *J Am Chem Soc* 125, 16162-16163.

121. Tai H.C., Khidekel N., Ficarro S.B., Peters E.C. and Hsieh-Wilson L.C. (**2004**) Parallel identification of O-GlcNAc-modified proteins from cell lysates. *J Am Chem Soc* 126, 10500-10501.

122. Khidekel N., Ficarro S.B., Peters E.C. and Hsieh-Wilson L.C. (**2004**) Exploring the O-GlcNAc proteome: direct identification of O-GlcNAc-modified proteins from the brain. *Proc Natl Acad Sci U S A* 101, 13132-13137.

123. Wang Z. *et al.* (**2009**) Site-specific GlcNAcylation of human erythrocyte proteins: potential biomarker(s) for diabetes. *Diabetes* 58, 309-317.

124. Khidekel N. *et al.* (**2007**) Probing the dynamics of O-GlcNAc glycosylation in the brain using quantitative proteomics. *Nat Chem Biol* 3, 339-348.

125. Parker B.L., Gupta P., Cordwell S.J., Larsen M.R. and Palmisano G. (**2011**) Purification and identification of O-GlcNAc-modified peptides using phosphate-based alkyne CLICK chemistry in combination with titanium dioxide chromatography and mass spectrometry. *J Proteome Res* 10, 1449-1458.

126. Wang Z. and Hart G.W. (**2008**) Glycomic Approaches to Study GlcNAcylation: Protein Identification, Site-mapping, and Site-specific O-GlcNAc Quantitation. *Clin Proteom* 4, 5-13.

127. Wang Z. *et al.* (**2010**) Enrichment and site mapping of O-linked N-acetylglucosamine by a combination of chemical/enzymatic tagging, photochemical cleavage, and electron transfer dissociation mass spectrometry. *Mol Cell Proteomics* 9, 153-160.

128. Wang Z. *et al.* (**2010**) Extensive crosstalk between O-GlcNAcylation and phosphorylation regulates cytokinesis. *Sci Signal* 3, ra2.

129. Alfaro J.F. *et al.* (**2012**) Tandem mass spectrometry identifies many mouse brain O-GlcNAcylated proteins including EGF domain-specific O-GlcNAc transferase targets. *Proc Natl Acad Sci U*

*S A* 109, 7280-7285.

130. Muia R.P. *et al.* (**2010**) Identification of glycoproteins targeted by Trypanosoma cruzi trans-sialidase, a virulence factor that disturbs lymphocyte glycosylation. *Glycobiology* 20, 833-842.

131. Pasek M. *et al.* (**2009**) Bioconjugation and detection of lactosamine moiety using alpha1,3-galactosyltransferase mutants that transfer C2-modified galactose with a chemical handle. *Bioconjug Chem* 20, 608-618.

132. Zheng T. *et al.* (**2011**) Tracking N-acetyllactosamine on cell-surface glycans in vivo. *Angew Chem Int Ed Engl* 50, 4113-4118.

133. Chaubard J.L., Krishnamurthy C., Yi W., Smith D.F. and Hsieh-Wilson L.C. (**2012**) Chemoenzymatic probes for detecting and imaging fucose-alpha(1-2)-galactose glycan biomarkers. *J Am Chem Soc* 134, 4489-4492.

134. Chalkley R.J. and Clauser K.R. (**2012**) Modification site localization scoring: strategies and performance. *Mol Cell Proteomics* 11, 3-14.

135. Huddleston M.J., Bean M.F. and Carr S.A. (**1993**) Collisional fragmentation of glycopeptides by electrospray ionization LC/MS and LC/MS/MS: methods for selective detection of glycopeptides in protein digests. *Anal Chem* 65, 877-884.

136. Medzihradszky K.F. *et al.* (**1990**) Structure determination of O-linked glycopeptides by tandem mass spectrometry. *Biomed Environ Mass Spectrom* 19, 777-781.

137. Bunkenborg J., Pilch B.J., Podtelejnikov A.V. and Wisniewski J.R. (**2004**) Screening for N-glycosylated proteins by liquid chromatography mass spectrometry. *Proteomics* 4, 454-465.

138. Gonzalez J. *et al.* (**1992**) A method for determination of N-glycosylation sites in glycoproteins by collision-induced dissociation analysis in fast atom bombardment mass spectrometry: identification of the positions of carbohydrate-linked asparagine in recombinant alpha-amylase by treatment with peptide-N-glycosidase F in 18O-labeled water. *Anal Biochem* 205, 151-158.

139. Kaji H. *et al.* (**2003**) Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. *Nat Biotechnol* 21, 667-672.

140. Atwood J.A., 3rd *et al.* (**2006**) Glycoproteomics of Trypanosoma cruzi trypomastigotes using subcellular fractionation, lectin affinity, and stable isotope labeling. *J Proteome Res* 5, 3376-3384.

141. Robinson N.E. and Robinson A.B. (**2001**) Deamidation of human proteins. *Proc Natl Acad Sci*

*U S A* 98, 12409-12413.

142. Angel P.M., Lim J.M., Wells L., Bergmann C. and Orlando R. (**2007**) A potential pitfall in 18O-based N-linked glycosylation site mapping. *Rapid Commun Mass Spectrom* 21, 674-682.

143. Hagglund P., Bunkenborg J., Elortza F., Jensen O.N. and Roepstorff P. (**2004**) A new strategy for identification of N-glycosylated proteins and unambiguous assignment of their glycosylation sites using HILIC enrichment and partial deglycosylation. *J Proteome Res* 3, 556-566.

144. Zhang W., Wang H., Zhang L., Yao J. and Yang P. (**2011**) Large-scale assignment of N-glycosylation sites using complementary enzymatic deglycosylation. *Talanta* 85, 499-505.

145. Schmidt A., Claassen M. and Aebersold R. (**2009**) Directed mass spectrometry: towards hypothesis-driven proteomics. *Curr Opin Chem Biol* 13, 510-517.

146. Palaniappan K.K. *et al.* (**2011**) Isotopic signature transfer and mass pattern prediction (IsoStamp): an enabling technique for chemically-directed proteomics. *ACS Chem Biol* 6, 829-836.

147. Goodlett D.R. *et al.* (**2000**) Protein identification with a single accurate mass of a cysteine-containing peptide and constrained database searching. *Anal Chem* 72, 1112-1118.

148. Breidenbach M.A., Palaniappan K.K., Pitcher A.A. and Bertozzi C.R. (**2012**) Mapping yeast N-glycosites with isotopically recoded glycans. *Mol Cell Proteomics* 11, M111 015339.

149. Breidenbach M.A. *et al.* (**2010**) Targeted metabolic labeling of yeast N-glycans with unnatural sugars. *Proc Natl Acad Sci U S A* 107, 3988-3993.

150. Hong J.C. and Kim Y.S. (**2000**) Alkali-catalyzed beta-elimination of periodate-oxidized glycans: a novel method of chemical deglycosylation of mucin gene products in paraffin embedded sections. *Glycoconj J* 17, 691-703.

151. Gerken T.A., Gupta R. and Jentoft N. (**1992**) A novel approach for chemically deglycosylating O-linked glycoproteins. The deglycosylation of submaxillary and respiratory mucins. *Biochemistry* 31, 639-648.

152. Hanisch F.G., Teitz S., Schwientek T. and Muller S. (**2009**) Chemical de-O-glycosylation of glycoproteins for application in LC-based proteomics. *Proteomics* 9, 710-719.

153. Patel T.P. and Parekh R.B. (**1994**) Release of oligosaccharides from glycoproteins by hydrazinolysis. *Methods Enzymol* 230, 57-66.

154. Edge A.S., Faltynek C.R., Hof L., Reichert L.E., Jr. and Weber P. (**1981**) Deglycosylation of

glycoproteins by trifluoromethanesulfonic acid. *Anal Biochem* 118, 131-137.

155. Edge A.S. (**2003**) Deglycosylation of glycoproteins with trifluoromethanesulphonic acid: elucidation of molecular structure and function. *Biochem J* 376, 339-350.

156. Rademaker G.J., Haverkamp J. and Thomasoates J. (**1993**) Determination of Glycosylation Sites in O-Linked Glycopeptides - a Sensitive Mass-Spectrometric Protocol. *Org Mass Spectrom* 28, 1536-1541.

157. Greis K.D. *et al.* (**1996**) Selective detection and site-analysis of O-GlcNAc-modified glycopeptides by beta-elimination and tandem electrospray mass spectrometry. *Anal Biochem* 234, 38-49.

158. Hanisch F.G., Jovanovic M. and Peter-Katalinic J. (**2001**) Glycoprotein identification and localization of O-glycosylation sites by mass spectrometric analysis of deglycosylated/alkylaminylated peptide fragments. *Anal Biochem* 290, 47-59.

159. Rademaker G.J. *et al.* (**1998**) Mass spectrometric determination of the sites of O-glycan attachment with low picomolar sensitivity. *Anal Biochem* 257, 149-160.

160. Zheng Y., Guo Z. and Cai Z. (**2009**) Combination of beta-elimination and liquid chromatography/quadrupole time-of-flight mass spectrometry for the determination of O-glycosylation sites. *Talanta* 78, 358-363.

161. Wells L. *et al.* (**2002**) Mapping sites of O-GlcNAc modification using affinity tags for serine and threonine post-translational modifications. *Mol Cell Proteomics* 1, 791-804.

162. Overath T. *et al.* (**2012**) Mapping of O-GlcNAc Sites of 20 S Proteasome Subunits and Hsp90 by a Novel Biotin-Cystamine Tag. *Mol Cell Proteomics* 11, 467-477.

163. Vosseller K. *et al.* (**2005**) Quantitative analysis of both protein expression and serine / threonine post-translational modifications through stable isotope labeling with dithiothreitol. *Proteomics* 5, 388-398.

164. Zubarev R.A. *et al.* (**2000**) Electron capture dissociation for structural characterization of multiply charged protein cations. *Anal Chem* 72, 563-573.

165. Zubarev R.A., Kelleher N.L. and McLafferty F.W. (**1998**) Electron capture dissociation of multiply charged protein cations. A nonergodic process. *J Am Chem Soc* 120, 3265-3266.

166. Syka J.E., Coon J.J., Schroeder M.J., Shabanowitz J. and Hunt D.F. (**2004**) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A*

101, 9528-9533.

167. Zhao P. *et al.* (**2011**) Combining High-Energy C-Trap Dissociation and Electron Transfer Dissociation for Protein O-GlcNAc Modification Site Assignment. *J Proteome Res* 10, 4088-4104.

168. Chalkley R.J., Thalhammer A., Schoepfer R. and Burlingame A.L. (**2009**) Identification of protein O-GlcNAcylation sites using electron transfer dissociation mass spectrometry on native peptides. *Proc Natl Acad Sci U S A* 106, 8894-8899.

169. Myers S.A., Panning B. and Burlingame A.L. (**2011**) Polycomb repressive complex 2 is necessary for the normal site-specific O-GlcNAc distribution in mouse embryonic stem cells. *Proc Natl Acad Sci U S A* 108, 9490-9495.

170. Trinidad J.C. *et al.* (**2012**) Global identification and characterization of both O-GlcNAcylation and phosphorylation at the murine synapse. *Mol Cell Proteomics* 11, 215-229.

171. Scott N.E. *et al.* (**2011**) Simultaneous Glycan-Peptide Characterization Using Hydrophilic Interaction Chromatography and Parallel Fragmentation by CID, Higher Energy Collisional Dissociation, and Electron Transfer Dissociation MS Applied to the N-Linked Glycoproteome of Campylobacter jejuni. *Molecular & Cellular Proteomics* 10.

172. Kim M.S. and Pandey A. (**2012**) Electron transfer dissociation mass spectrometry in proteomics. *Proteomics* 12, 530-542.

173. Darula Z., Chalkley R.J., Lynn A., Baker P.R. and Medzihradszky K.F. (**2011**) Improved identification of O-linked glycopeptides from ETD data with optimized scoring for different charge states and cleavage specificities. *Amino Acids* 41, 321-328.

174. Halim A. *et al.* (**2011**) Site-specific characterization of threonine, serine, and tyrosine glycosylations of amyloid precursor protein/amyloid beta-peptides in human cerebrospinal fluid. *Proc Natl Acad Sci U S A* 108, 11848-11853.

175. Perdivara I. *et al.* (**2009**) Elucidation of O-glycosylation structures of the beta-amyloid precursor protein by liquid chromatography-mass spectrometry using electron transfer dissociation and collision induced dissociation. *J Proteome Res* 8, 631-642.

176. Sihlbom C. *et al.* (**2009**) Localization of O-glycans in MUC1 glycoproteins using electron-capture dissociation fragmentation mass spectrometry. *Glycobiology* 19, 375-381.

177. Takahashi K. *et al.* (**2010**) Clustered O-glycans of IgA1: defining macro- and microheterogeneity by use of electron capture/transfer dissociation. *Mol Cell Proteomics* 9, 2545-2557.

178. Darula Z. and Medzihradszky K.F. (**2009**) Affinity enrichment and characterization of mucin core-1 type glycopeptides from bovine serum. *Mol Cell Proteomics* 8, 2515-2526.

179. Darula Z., Sherman J. and Medzihradszky K.F. (**2012**) How to dig deeper? Improved enrichment methods for mucin core-1 type glycopeptides. *Mol Cell Proteomics* 11, O111 016774.

180. Steentoft C. *et al.* (**2011**) Mining the O-glycoproteome using zinc-finger nuclease-glycoengineered SimpleCell lines. *Nat Methods* 8, 977-982.

181. Schjoldager K.T. *et al.* (**2012**) Probing isoform-specific functions of polypeptide GalNAc-transferases using zinc finger nuclease glycoengineered SimpleCells. *Proc Natl Acad Sci U S A* 109, 9893-9898.

182. Leitner A. *et al.* (**2010**) Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics. *Mol Cell Proteomics* 9, 1634-1649.

183. Tanaka Y., Bond M.R. and Kohler J.J. (**2008**) Photocrosslinkers illuminate interactions in living cells. *Mol Biosyst* 4, 473-480.

184. Frei A.P. *et al.* (**2012**) Direct identification of ligand-receptor interactions on living cells and tissues. *Nat Biotechnol* 30, 997-1001.

185. Lauc G., Lee R.T., Dumic J. and Lee Y.C. (**2000**) Photoaffinity glycoprobes - a new tool for the identification of lectins. *Glycobiology* 10, 357-364.

186. Ilver D. *et al.* (**1998**) Helicobacter pylori adhesin binding fucosylated histo-blood group antigens revealed by retagging. *Science* 279, 373-377.

187. Pacuszka T. and Fishman P.H. (**1992**) Synthesis of a photoreactive, radiolabelled derivative of the oligosaccharide of GM1 ganglioside. *Glycobiology* 2, 251-255.

188. Yu S.H., Wands A.M. and Kohler J.J. (**2012**) Photoaffinity Probes for Studying Carbohydrate Biology. *J Carbohyd Chem* 31, 325-352.

189. Du J. *et al.* (**2009**) Metabolic glycoengineering: sialic acid and beyond. *Glycobiology* 19, 1382-1401.

190. Ramya T.N. *et al.* (**2010**) In situ trans ligands of CD22 identified by glycan-protein photocross-linking-enabled proteomics. *Mol Cell Proteomics* 9, 1339-1351.

191. Kelm S. *et al.* (**1998**) Functional groups of sialic acids involved in binding to siglecs (sialoadhesins) deduced from interactions with synthetic analogues. *Eur J Biochem* 255, 663-672.

192. Luchansky S.J. and Bertozzi C.R. (**2004**) Azido sialic acids can modulate cell-surface interactions. *Chembiochem* 5, 1706-1709.

193. Bond M.R., Zhang H.C., Vu P.D. and Kohler J.J. (**2009**) Photocrosslinking of glycoconjugates using metabolically incorporated diazirine-containing sugars. *Nat Protoc* 4, 1044-1063.
194. Tanaka Y. and Kohler J.J. (**2008**) Photoactivatable crosslinking sugars for capturing glycoprotein interactions. *J Am Chem Soc* 130, 3278-3279.

195. Bond M.R., Whitman C.M. and Kohler J.J. (**2010**) Metabolically incorporated photocrosslinking sialic acid covalently captures a ganglioside-protein complex. *Mol Biosyst* 6, 1796-1799.

196. Bond M.R. *et al.* (**2011**) Metabolism of diazirine-modified N-acetylmannosamine analogues to photo-cross-linking sialosides. *Bioconjug Chem* 22, 1811-1823.

197. Yu S.H. *et al.* (**2012**) Metabolic labeling enables selective photocrosslinking of O-GlcNAc-modified proteins to their binding partners. *Proc Natl Acad Sci U S A* 109, 4834-4839.

198. Rinner O. *et al.* (**2008**) Identification of cross-linked peptides from large sequence databases. *Nat Methods* 5, 315-318.

199. Goodfellow J.J. *et al.* (**2012**) An endoglycosidase with alternative glycan specificity allows broadened glycoprotein remodelling. *J Am Chem Soc* 134, 8030-8033.

200. Watrous J.D. and Dorrestein P.C. (**2011**) Imaging mass spectrometry in microbiology. *Nat Rev Microbiol* 9, 683-694.

201. Schwamborn K. and Caprioli R.M. (**2010**) Molecular imaging by mass spectrometry--looking beyond classical histology. *Nat Rev Cancer* 10, 639-646.

# Chapter 3: Isotopically Recoding Peptides for Chemically Directed Proteomics

## 3.1 Introduction

Common goals of mass spectrometry (MS)-based proteomics experiments are to identify, characterize, and quantify proteins and their posttranslational modifications from biological samples.[1] A popular strategy for protein identification is the bottom-up shotgun proteomics approach (Figure 2-1). In this method, a mixture of proteins is subjected to proteolytic digestion, the resulting peptides are separated by liquid chromatography (LC) and detected by MS, and their parent proteins are inferred from the assigned peptide sequences.[2] To convert MS data acquired from proteolytic digests into protein identifications, tandem MS can be used to obtain sequence information for individual peptides, followed by comparison to an *in silico* proteolytic digest of an organism's proteome.[3-5] Typically, only the most abundant peptides are selected for fragmentation, while data for those peptides in relatively low quantities are not obtained.[1]

An inherent problem in shotgun proteomics is identifying proteins of low abundance, such as biomarkers for disease states, against a background of proteins whose concentrations can span up to 12 orders of magnitude (Figure 2-2).[6,7] Directed proteomics strategies seek to address the sample complexity problem by focusing the analysis on a defined protein subset.[8-10] In one approach, proteins of interest are selectively enriched prior to proteolytic digestion, thereby forgoing the shotgun method altogether.[11,12] Alternatively, there is growing interest in the use of chemical tags that perturb the mass envelope of target peptides so as to render them more detectable. The progenitors of this approach are the isotope-coded affinity tag (ICAT) and isobaric tags for relative and absolute quantitation (iTRAQ), techniques now commonly used for quantitative comparative proteomics.[13-15] Chemical tags have also been elegantly employed to mark sites of protein posttranslational modifications[16] including glycosylation,[17] lipidation,[18-20] and phosphorylation.[21] Tags have also been used for labeling protein N-termini,[22] sites of cysteine oxidation,[23] enzyme active sites,[24] and points of cross-linking.[25]

***Table 3-1. The unique isotopic abundance of halogens.***

The natural abundances of stable isotopes of elements found in an amino acid primarily exist as light isotopes, while bromine and chlorine (red box) have high abundances of their heavy isotopes.

| Element | Stable Isotopes | Natural Abundances |
|---|---|---|
| Hydrogen | $^{1}H$ | 99.99% |
| | $^{2}H$ | 00.01% |
| Carbon | $^{12}C$ | 98.93% |
| | $^{13}C$ | 01.07% |
| Nitrogen | $^{14}N$ | 99.64% |
| | $^{15}N$ | 00.36% |
| Oxygen | $^{16}O$ | 99.76% |
| | $^{17}O$ | 00.04% |
| | $^{18}O$ | 00.20% |
| Phosphorus | $^{31}P$ | 99.99% |
| Sulfur | $^{32}S$ | 94.99% |
| | $^{34}S$ | 04.25% |
| Bromine | $^{79}Br$ | 50.69% |
| | $^{81}Br$ | 49.31% |
| Chlorine | $^{35}C$ | 75.76% |
| | $^{37}C$ | 24.24% |

Austin Pitcher, Brian Smart, David Spiciarich, and Anthony Iavarone contributed to the work presented in this chapter. Portions of the work described in this chapter have been reported in a separate publication.[53]

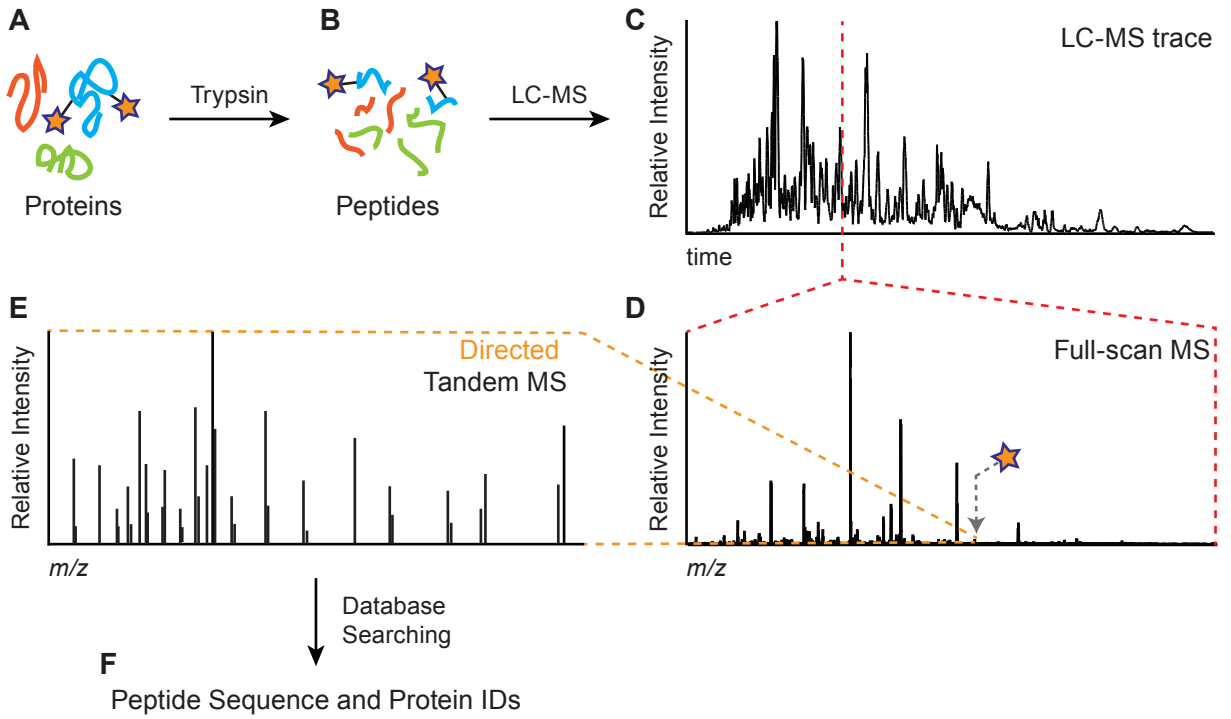



***Figure 3-1. Directed proteomics seeks to focus MS analysis on a target protein subset.***

(A) A mixture of proteins in which a small number are chemically tagged (star) is subjected to proteolytic digestion producing (B) a mixture of peptides. (C) The peptides are separated using LC and (D) full-scan mass spectra are collected at regular intervals. (E) The tagged species can be selected for fragmentation (i.e., tandem MS) using an inclusion list. (F) These fragment mass spectra are subjected to database searching to produce peptide and protein identifications.

The halogens bromine and chlorine can be advantageous components of chemical tags for MS by virtue of their unique isotopic distributions. Unlike the proteogenic elements, which exist as one predominant isotope, bromine and chlorine have two abundant isotopes that create unique patterns in a mass spectrum: $^{79}$Br and $^{81}$Br are naturally found in a 1:1 ratio, and $^{35}$Cl and $^{37}$Cl are naturally found in a 3:1 ratio (isotopic abundances of proteogenic elements are given in Table 3-1).[26] While this feature has been well exploited in the field of small molecule and metabolite characterization,[27-31] its use in proteomics-related applications has been limited. The first example by Goodlett, Aebersold and coworkers used a dichloride tag to discriminate between peptides with and without cysteine from digested protein samples.[32] Likewise, N-terminal labeling of peptides with a monobromide tag facilitated sequence identification by tandem MS.[33] Recently, Hang and coworkers used a monobromide cleavable tag to enrich for newly synthesized proteins in bacteria.[34] In addition to their distinctive isotopic signatures, bromine and chlorine have a negative mass defect that can endow a modified peptide with a unique fractional mass, which Amster and coworkers made artful use of for peptide mass fingerprinting analysis.[35-37] To date, halogen profiling methods have not been extended to directed proteomic analysis of samples as complex as human cell or tissue lysates. To achieve this goal would require the ability to discriminate a halogen tag's signature on peptides over a wide mass range, in multiple charge states, and against a background of >100,000 peptides, capabilities that present

methods lack.[38]

Here we report that a dibromide tag in concert with a novel computational pattern-searching algorithm enables detection of labeled peptides from complex biological samples with unprecedented sensitivity and fidelity. The overall approach, termed isotopic signature transfer and mass pattern prediction (abbreviated IsoStamp), was employed as illustrated in Figure 3-1. Cell lysates containing a chemically-tagged protein (orange star) were digested with trypsin and the resulting peptides were analyzed by LC-MS in full-scan mode. Tagged peptides could then be selectively fragmented using a directed proteomics strategy. Unlike an intensity-driven data-dependent LC-MS/MS analysis, the IsoStamp method is not limited to identifying peptides of relatively high abundance. By rendering labeled peptides detectable in a full-scan mass spectrum, IsoStamp is an enabling tool for chemically-directed proteomics, maximizing the identification of peptides of interest from information-dense MS data.

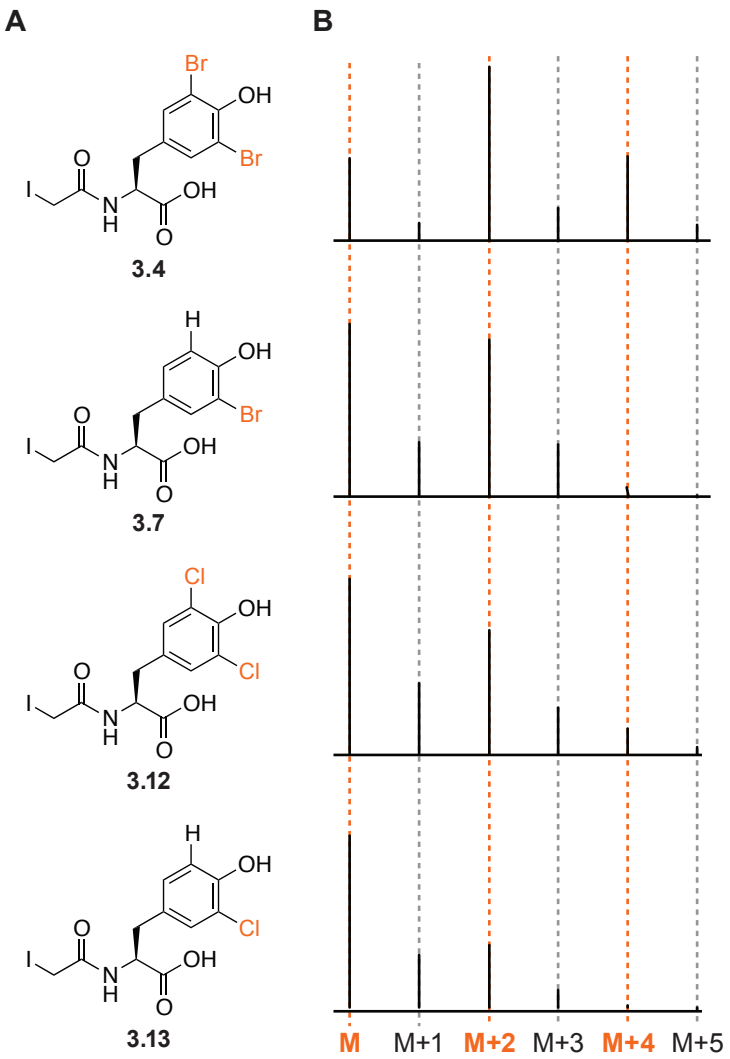



***Figure 3-2. Structures and mass spectra of halogenated tags used to label proteins.***

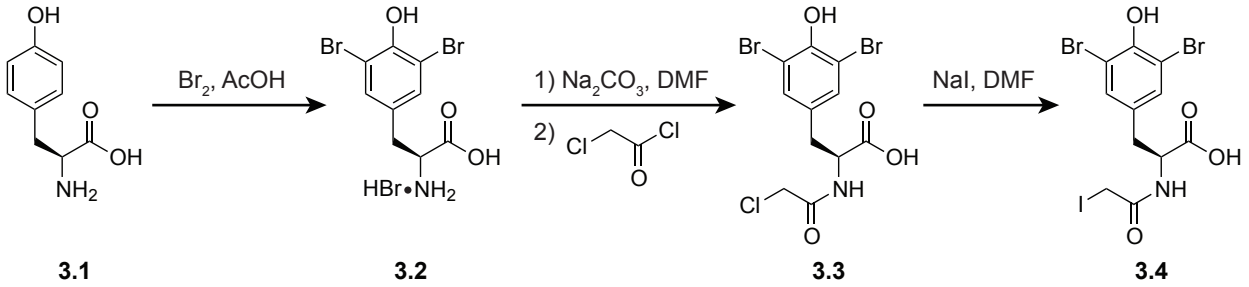
(A) The structures of four halogen tags synthesized as iodoacetamide-derivatized tyrosine analogs for alkylating cysteine residues. (B) The mass spectra of each halogen tag, revealing their unique isotopic envelopes. Interesting peaks appear at the M+2 and M+4 positions in the isotopic envelopes (highlighted in orange). For example, the dibromide tag (**3.4**) appears as a triplet while the monobromide tag (**3.7**) appears as a doublet.

## 3.2 Results and Discussion

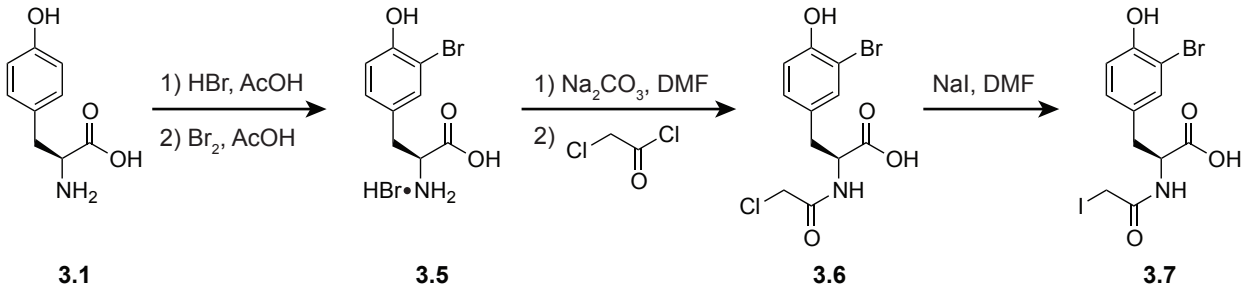
### *3.2.1 Bromine and chlorine impart unique isotopic signatures to labeled molecules*

Mass spectra of low molecular weight (MW) compounds bearing a single bromine or chlorine atom show two major ions, M and M + 2, with equal or skewed peak heights, respectively. Compounds with two halogen atoms produce symmetrical (2x Br) or skewed (2x Cl) triplets with major peaks at M, M + 2 and M + 4. These unique isotopic patterns are evident in the mass spectra (Figure 3-2B) of halogenated tyrosine analogs **3.4**, **3.7**, **3.12**, and **3.13** shown in Figure 3-2A, which were synthesized as iodoacetamide derivatives (Schemes 3.1-3.3). In

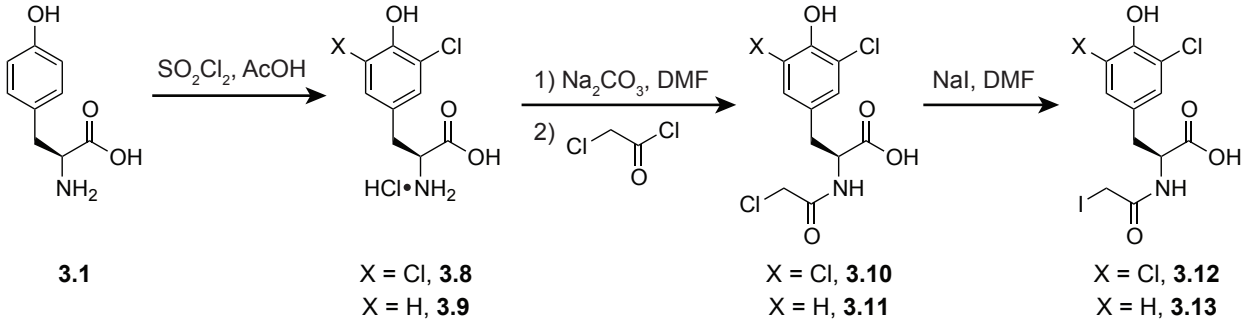
***Scheme 3-1. Synthesis of dibromide tag* 3.4.**




***Scheme 3-2. Synthesis of monobromide tag* 3.7.**




***Scheme 3-3. Synthesis of dichloride and monochloride tags, 3.12 and 3.13, respectively.***




principle, the uniqueness of the triplet patterns associated with the dibromide and dichloride motifs could facilitate the identification of tagged peptides from complex proteolytic digests. However, in larger molecules (i.e., MW > 1000) the isotopic patterns are obscured due to the influence of heavy isotopes of C ($^{13}$C, 1%), H ($^{2}$H, 0.02%) and N ($^{15}$N, 0.1%) on the overall mass envelope. To illustrate the point, we alkylated the surface-exposed cysteine residues of bovine serum albumin (BSA) with tags **3.4**, **3.7**, **3.12**, and **3.13**, digested the modified protein, and analyzed the peptides by LC-MS (Figure 3-3A). Representative data corresponding to the tryptic peptide SLHTLFGDELC*K are shown in Figure 3-3B. The isotopic envelope of each tagged peptide reflects the parent peptide's intrinsic isotopic distribution, as seen in the mass spectrum of the iodoacetic acid-alkylated version (Figure 3-3A), convoluted (or recoded) with a halogen pattern (Figure 3-2B). These data illustrate that the dibromide tag imparts a more distinctive signature on a peptide's mass envelope than the other halogen tags. Computational simulations

suggested a similar advantage of the dibromide tag for peptides of MW up to at least 5,000 Da. Still, the complexity of a tagged peptide's mass spectrum prevents manual searching for isotopic envelopes in complex mixtures. Instead, labeled peptides are detected computationally using a pattern-searching algorithm.

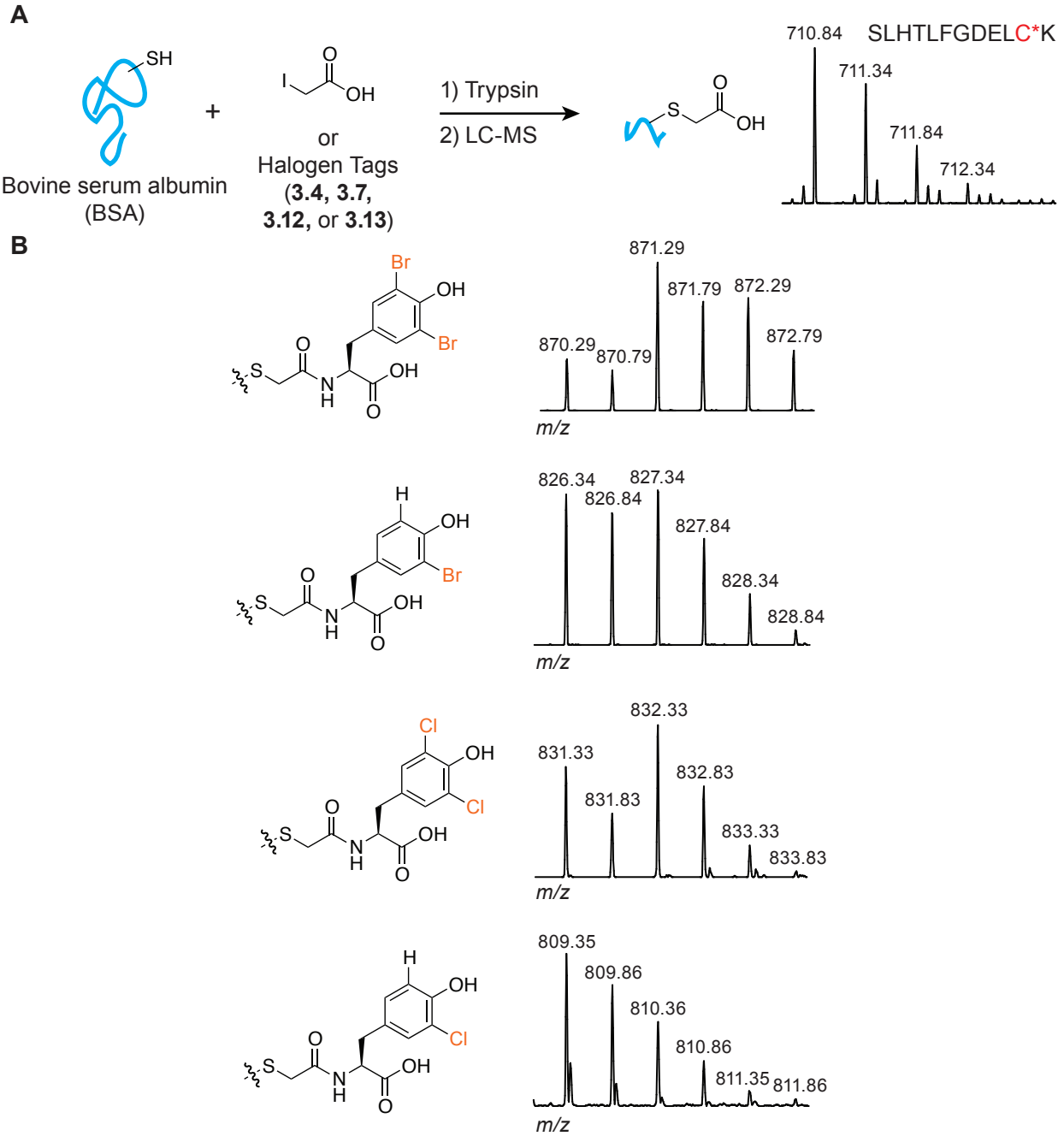



***Figure 3-3. Halogenated tags impart distinct isotopic patterns to peptides.***

(A) In a model experiment, BSA was alkylated on cysteine residues with iodoacetic acid or halogen tags and then digested with trypsin. (B) Mass spectra of the modified BSA tryptic peptide corresponding to residues 89–100, where C* refers to a cysteine residue alkyated by either iodoacetic acid or one of the halogen tags.

While several peak-picking and isotope distribution prediction algorithms exist,[25,39-43] they are not designed, or have not been demonstrated, to search for any user-defined isotope pattern amongst the sample complexity present in mammalian whole cell lysate. Additionally, some of these approaches require restraints and/or prior knowledge regarding the sequences of the targeted peptides. While such information is useful for many applications, including multiple reaction monitoring,[8,10] we sought to develop a versatile computational algorithm capable of extracting any isotopic signature from complex MS data without imposing any restrictions based on the structure or reactivity of the chemical tag, or on the amino acid composition of the labeled peptides.

### *3.2.2 Development of a pattern-searching algorithm*

The algorithm described here analyzes peaks from a full-scan mass spectrum and matches real data with simulated data generated by convoluting each predicted peptide's isotopic envelope with the pattern produced by a given tag. The algorithm receives two inputs from the user: (1) a centroided mzXML data file and (2) a parameter file that includes the MW and isotopic pattern of the tag, charge states to be considered in the search, and weighting factors used to tune selectivity and sensitivity. The output comprises the *m/z* value and retention time of tagged species, which form an inclusion list for subsequent directed LC-MS/MS analysis.

The algorithm has two key steps. First, the full-scan MS data are analyzed to identify putative isotopic signature matches for a given elemental composition. Key to this step is a data-dependent approximation of the contributions of non-halogens to the observed isotopic envelope, while allowing for the inevitable imperfections in MS data derived from complex protein samples. In the second step, the putative matches from the first step are analyzed using a graph-theoretic construct to reduce false positives. Peaks contributing to a putative pattern match are tracked as a function of LC elution time and number of charge states detected to add confidence that they were derived from a real species.

#### *3.2.2.1 Step 1: Identifying putative pattern matches*

The algorithm takes a list of peaks from the full-scan mass spectrum and divides them into sets that are possibly isotopically related. Each of these sets is searched for the presence of a desired isotopic pattern as follows. First, each peak in the chosen data set is presumed to represent a peptide. Knowing the charge state and *m/z* for that hypothetical peptide, the program predicts its mass and estimates its elemental composition using the "averagine" approximation (Table 3-2).[44]

We confirmed the accuracy of the averagine method by comparing the root mean

***Table 3-2. The averagine peptide.***

The numbers shown (number of atoms of each element type per Da of peptide), allow the elemental composition of a peptide to be predicted from its molecular weight.

| Element | Number per Da peptide[1] (± std. dev.) |
|---|---|
| Hydrogen | 0.07100 ± 0.00470 |
| Carbon | 0.04350 ± 0.00520 |
| Nitrogen | 0.01260 ± 0.00180 |
| Oxygen | 0.01380 ± 0.00170 |
| Sulfur[2] | 0.00037 ± 0.00052 |

[1]Calculated from a random selection of 5000 human tryptic peptides with a sequence length greater than 6 amino acids.

[2]Because of its rarity in peptides, there is a high error associated with predicting the number of sulfur atoms in relatively small peptides.

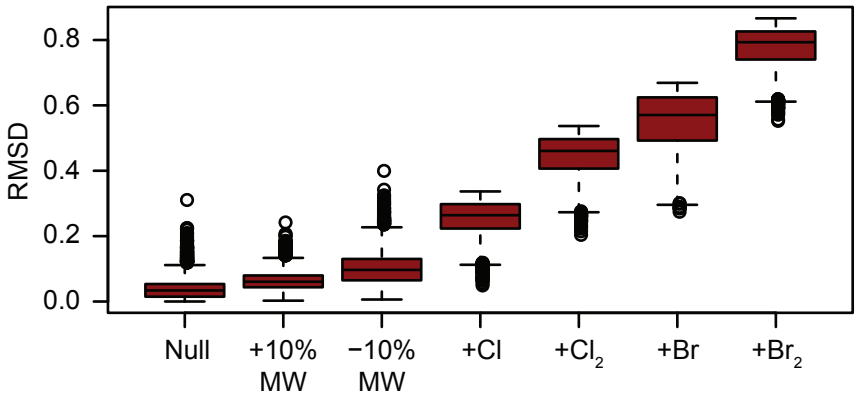



***Figure 3-4. The averagine model accurately predicts peptide elemental composition.***

A boxplot of the RMSD intensity difference between the isotopic envelopes for 20,000 human tryptic peptides, based on their actual elemental composition, with their averagine predicted elemental compositions ("Null" pattern) based on MW (molecular weight), or with an error of ±10% in MW, or with the addition of the indicated halogen tag. The bottom and top of the red box represent the 25$^{th}$ and 75$^{th}$ percentile of data, respectively, the black bar in the red box is the median, and the ends of the whiskers represent the full range of data, except for outliers, which are depicted as open circles.

squared deviation (RMSD) between the isotopic envelope calculated for the elemental compositions of 20,000 human tryptic peptides, with the isotopic envelope calculated for an averagine predicted elemental composition based on accurate molecular weight (MW), i.e. "Null" pattern (Figure 3-4). In calculating the RMSD, *m/z* was assumed to be identical; relative peak heights, i.e. the isotopic pattern, were compared. Under this context, the "Null" pattern had a median deviation of less than 4% from the actual pattern. If the molecular weight (MW) is incorrect by 10%, the averagine model still produces an accurate estimation of the actual elemental composition, and consequently the actual isotopic envelope. However, as we consider peptides that contain halogenated tags, the averagine prediction of elemental composition from a given MW becomes less accurate as we proceed from a monochloride-, to a dichloride-, to a monobromide-, and finally to a dibromide-labeled peptide. The consequence of this deviation from the averagine model is visually depicted in Figure 3-3, where the isotopic envelope associated with a halogen labeled peptide is markedly different from that of an unlabeled peptide. The pattern-searching algorithm utilizes these differences.

From the estimated elemental composition, an isotopic pattern of the unlabeled hypothetical peptide is predicted. Then, the isotopic pattern of a chemical tag (i.e., tags **3.4**, **3.7**, **3.12**, and **3.13**) is convoluted with the hypothetical peptide's isotopic envelope, generating a reference pattern that is compared with the actual data set to determine a fitness score. The program also samples reference patterns that model untagged peptides and instrument noise. Additional reference patterns can be incorporated to account for common sources of false positives in a sample-dependent manner.

Each reference pattern ($R$) is scaled in the intensity dimension to produce an optimal alignment with the data ($D$). This is accomplished by determining the scaling factor $k$ by a binary search such that the sum of the squared difference ($SSD$) between each peak in the reference pattern ($r_i \in R$) and its counterpart in the actual data set ($d_i \in D$) is minimized:

$$SSD = \sum_i \left(d_i - kr_i\right)^2 \tag{3-1}$$

After intensity alignment, the score for the entire pattern is then calculated as:

$$Score = \prod_i f\left(\frac{\left|d_i - kr_i\right|}{\sigma\sqrt{2}}\right) \tag{3-2}$$

where $\sigma$ is a measure of peak intensity variance and $f$ is a scoring function for each peak that produces a value in the range [0,1]

$$f(x) = \max\left[\operatorname{erfc}(x), \varepsilon\right] \qquad 0 < \varepsilon << 1 \tag{3-3}$$

in which *erfc(x)* is the complement of the Gaussian error function and the parameter is a measure of the tightness of the peak matching in the intensity dimension. The lower bound of ε is imposed on the function to reduce round-off errors in floating point arithmetic and to allow for robustness against contaminating peaks when used in a Bayesian system. In short, this system allows the identification of isotopic envelopes in actual MS data that do not perfectly match theoretically determined isotopic envelopes by virtue of overlapping peaks from other molecular species.

After scores of all patterns of interest have been determined, the best match is found using a Bayesian approach:

$$P\left(\text{pattern}_i \middle| data\right) = \frac{P\left(\text{data}\middle|\text{pattern}_i\right)P\left(\text{pattern}_i\right)}{\sum_j P\left(\text{data}\middle|\text{pattern}_j\right)P\left(\text{pattern}_j\right)} = \frac{Score\left(\text{pattern}_i\middle|\text{data}\right)P\left(\text{pattern}_i\right)}{\sum_j Score\left(\text{pattern}_j\middle|\text{data}\right)P\left(\text{pattern}_j\right)} \tag{3-4}$$

where the $P(pattern_i)$ terms are user-defined weighting factors that describe the probability that any peak in the dataset is caused by a molecular species with the isotopic distribution described by $pattern_i$, and were determined experimentally. These weighting factors allowed us to increase the specificity of the program for a selected pattern, thereby eliminating false positives, or conversely, increasing the number of hits, though potentially at the cost of more false positives.

*3.2.2.2 Step 2: Reducing false positives with a graph-theoretic approach*

Naive pattern matching, as described above, can produce a significant number of false positive matches depending on the complexity of the data. However, information from neighboring spectra is known to reduce false positive detections while enhancing the true positive identifications.[41,45] Therefore, our algorithm exploits two features of LC-MS data: (1) peptides are often detected in multiple charge states and (2) in several adjacent scans. To integrate these features into the pattern matching algorithm, a graph-theoretic approach was employed wherein each potential match was treated as a node in a graph. Edges were drawn between two nodes if the nodes could have come from the same molecular species, and if the nodes have sufficiently similar LC elution times. After edges are built, the graph is decomposed into disjoint subsets, where all nodes in a given subset could have been produced by the same peptide. Each

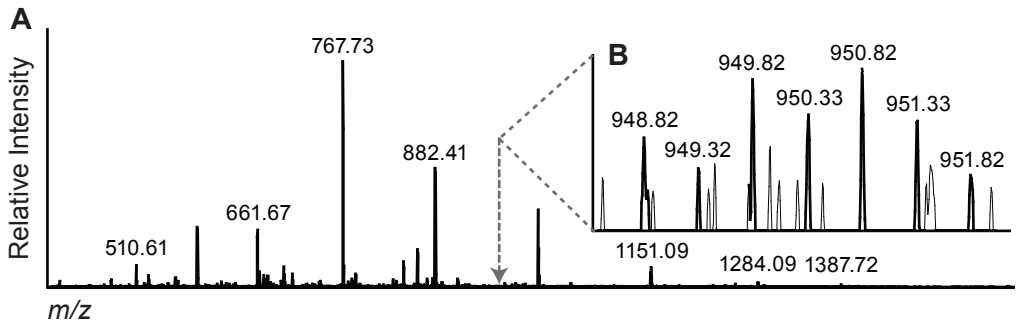



***Figure 3-5. The dibromide motif can be recognized at low signal-to-noise ratios.***

(A) Representative full-scan mass spectrum from LC-MS data derived from a trypsin digest of 150 femtomoles of dibromide-labeled BSA in 10 µg of Jurkat whole cell lysate. (B) The inset-spectrum shows a dibromide-labeled peptide LKPDPNTLC*DEFK at a S/N of 2.5:1. C* refers to a cysteine residue alkylated with dibromide tag **3.4**.

of these subsets is then scored on a number of factors, including the number of nodes in the set and the number of unique charge states detected. Because matches that were made by chance are unlikely to score highly using these criteria, this process filters false positive matches. Using a modern desktop computer (3.66 GHz, 4 GB RAM), an average LC-MS data file can be searched using standard settings in less than two minutes.

### *3.2.3 Application of IsoStamp to a model LC-MS experiment*

As mentioned previously, the complexity of unfractionated cell or tissue lysates renders the identification of low abundance proteins by shotgun proteomics a challenging endeavor. We therefore sought to test the sensitivity of IsoStamp in identifying a single labeled protein from whole cell lysates. BSA was chosen as a model protein for labeling because it contains 35 cysteine residues that are spread throughout the entire protein. The protein comprises 50 tryptic peptides of which 24 possess cysteine residues, including 15 with a single cysteine residue (no missed cleavages and a mass range of 600 to 2500 Da).

We generated detergent lysates of Jurkat cells, a human T-lymphoma cell line, and added known amounts of BSA that had been alkylated on cysteine residues with dibromide tag **3.4**. After digestion with trypsin, the sample was separated by in-line reversed-phase LC and analyzed on an LTQ-Orbitrap XL mass spectrometer. Figure 3-5A shows a representative full-scan mass spectrum from the LC-MS data collected for a sample derived from 150 femtomoles of **3.4**-labeled BSA in 10 µg of Jurkat whole cell lysate, representing 0.1% of the total protein content. When the full-scan mass spectrum was searched using the pattern matching software, we identified several halogen-labeled BSA peptides, collectively reflecting 32% coverage of single cysteine-containing peptides (Appendix A.5). The mass envelope of one such peptide, LKPDPNTLC*DEFK (residues 139-151), illustrates the unique isotopic pattern for a dibromide-labeled peptide (Figure 3-5B). Notably, the pattern (in black) was computationally found four orders of magnitude below that of the most abundant ion, at a signal-to-noise ratio (S/N) of 2.5:1, despite the presence of intervening peaks (light gray) within the envelope. Using conventional shotgun proteomics methodologies in samples of this complexity, peaks at such a low level of intensity may be excluded from tandem MS analysis.[7]

*3.2.4 The dibromide tag is superior to the other halogenated tags with respect to sensitivity and false positive identifications*

A central feature of the IsoStamp algorithm is that the user can tune its parameters to balance sensitivity against false positive identifications. Using BSA as a substrate, we compared the performance of the dibromide tag to the other halogenated tags. To determine the relative number of false positives, we first established searching parameters that found 50% of true halogen-labeled BSA peptides in a sample that contained 3 picomoles of halogen-labeled BSA in 10 µg of Jurkat whole cell lysate. The true positives were defined by a manual search of LC-MS data (explained in the Methods section) and are listed in Appendix A.6. The relative number of false positives was then determined by searching MS data derived from 10 µg of Jurkat whole cell lysate without BSA (and thus *no* real positives, Figure 3-6A). Compared to the dibromide tag, the dichloride tag produced greater than 30-fold more false positives while the monobromide tag produced greater than 120-fold more false positives. Overall, the dibromide tag outperformed the dichloride and monobromide tags by a substantial margin.

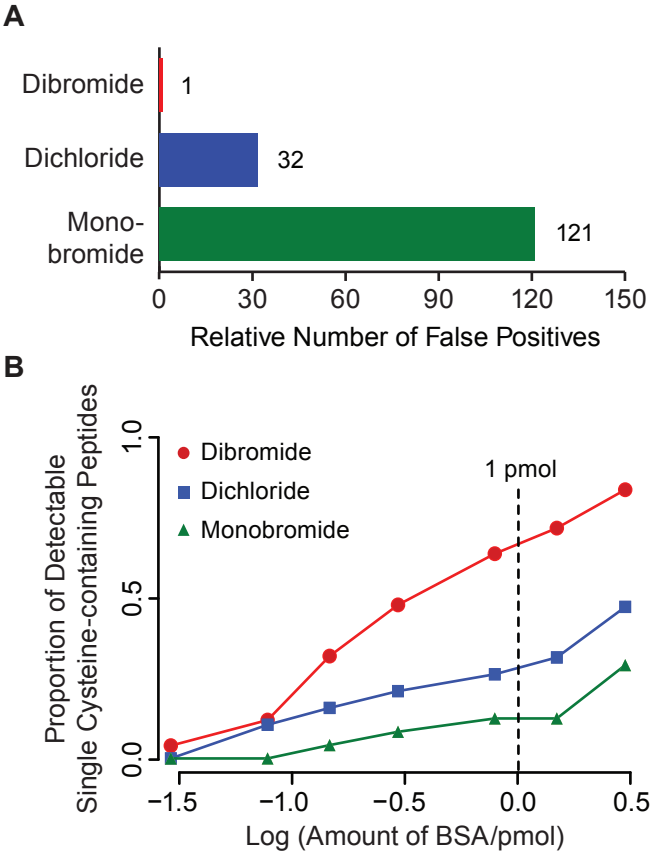



***Figure 3-6. The dibromide motif is superior to other halogen motifs with respect to the number of false positives and sensitivity.***

(A) Number of false positives identified in Jurkat whole cell lysate *without* BSA using searching conditions that found 50% of true positives in samples containing each halogen tag. (B) Sensitivity engendered by each halogen tag was determined by titrating 3.0 to 0.03 picomoles of tag-labeled BSA into 10 µg of Jurkat whole cell lysate and analyzing the tryptic digest by LC-MS.

To determine the sensitivities of the halogen tags, we found searching parameters for each tag that fixed the maximum number of false positive identifications at 100. We then performed a titration experiment where known quantities of halogen-labeled BSA were added to 10 µg of Jurkat whole cell lysate. Each mixture was digested with trypsin, subjected to LC-MS analysis, and the resulting data were searched for the tag's isotopic pattern. The proportion of single cysteine-containing BSA peptides identified as a function of protein concentration is shown in Figure 3-6B, where each computational match was manually verified (the computational detection rate for each peptide can be found in Appendix A.5). At all protein concentrations, dibromide-labeled peptides were detected at a higher frequency than peptides labeled with other tags. While the data appear to converge at low protein concentrations, this may reflect the detection limits of the instrument rather than

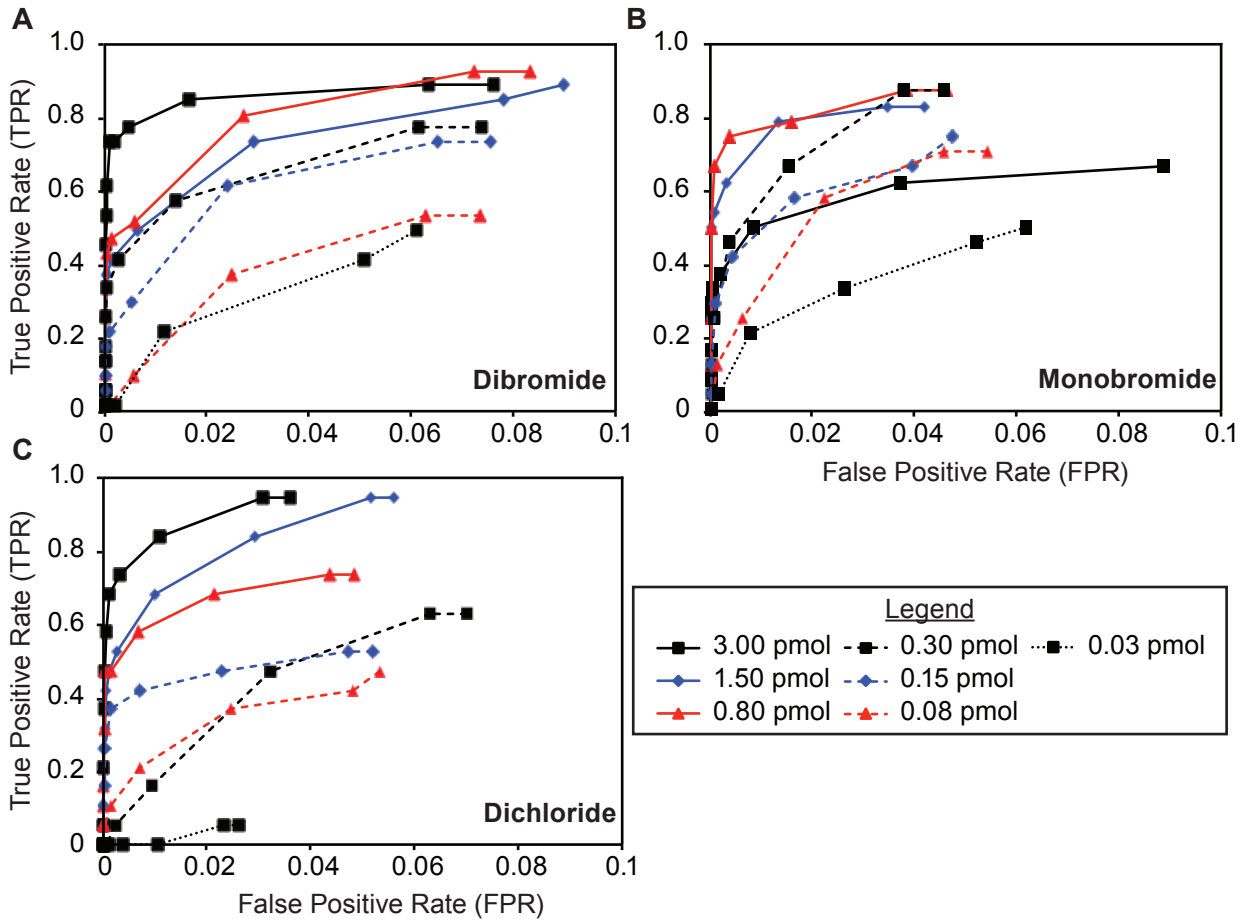



***Figure 3-7. Receiver operating characteristic curves (ROC) for the IsoStamp algorithm.***

Each plot shows the results of varying key parameters of the graph-theoretic algorithm (to adjust sensitivity and specificity) and then searching LC-MS datasets containing the indicated quantity of either (A) dibromide-, (B) monobromide-, or (C) dichloride-labeled BSA in the presence of 10 µg of Jurkat whole cell lysate.

capabilities of the pattern-searching algorithm. Overall, the dibromide isotopic signature was detected approximately twice as often as the dichloride and three times as often as the monobromide signatures, emphasizing its potential for recoding isotopic envelopes.

In order to simultaneously compare the sensitivity and selectivity of the halogen tags, receiver operating characteristic (ROC) plots were generated (Figure 3-7). These curves illustrate how the true positive rate (i.e., sensitivity) varies with respect to the false positive rate (i.e., selectivity). As illustrated, the ROC plots demonstrate that the IsoStamp algorithm can be tailored for each halogen tag, optimizing sensitivity and selectivity in each case, with the dibromide tag showing the best overall performance, even at low concentrations.

It should be noted that we were unable to determine the sensitivity and the relative number of false positives for the monochloride tag; reasonable searching parameters (i.e., with an acceptable number of false positives) to detect 50% of true positives in a sample containing 3 picomoles of monochloride-labeled BSA in 10 µg of Jurkat whole cell lysate could not be found due to the minimal perturbation of this tag on the natural isotopic pattern of peptides (see Figures 3-3B and 3-4).

### *3.2.5 Development of the IsoStamp method for directed shotgun proteomics*

After establishing the advantage of using a dibrominated tag for detecting labeled peptides in complex mixtures, we next tested the utility of the tag in a model directed shotgun proteomics experiment. As illustrated in Figure 3-8, the experiment was performed as follows. First, we added known amounts of dibromide-labeled BSA to 10 µg of Jurkat whole cell lysate. The mixture was digested with trypsin and then subjected to LC-MS analysis in full-scan mode on an LTQ-Orbitrap XL mass spectrometer. Data were processed to identify and inventory all dibromide-labeled peptides, generating an inclusion list that contained the M+2 mass of a labeled peptide's isotopic envelope and a retention time window. Then, the same sample was subjected to an LC-MS/MS experiment using the inclusion list to trigger fragmentation only if a

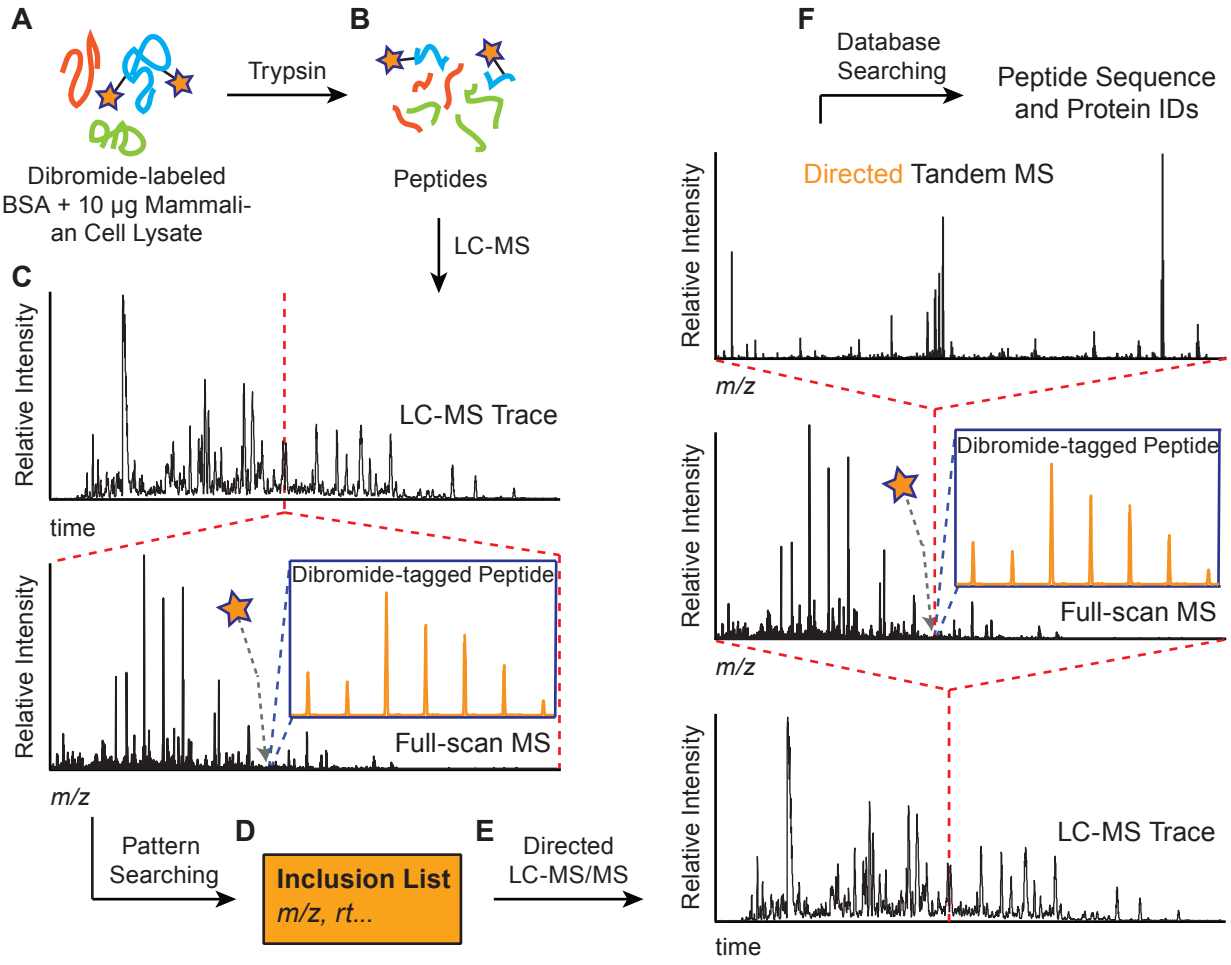



***Figure 3-8. Schematic of the IsoStamp method for performing a directed proteomics experiment.***

The IsoStamp method improves shotgun proteomics by allowing tagged peptides to be detectable in full-scan mass spectra, facilitating an inclusion-list driven directed LC-MS/MS experiment. (A) First, a mixture of proteins in which a small number are chemically tagged (star) is subjected to proteolytic digestion producing (B) a mixture of peptides. (C) Then, the peptides are separated using LC and full-scan mass spectra are collected at regular intervals. (D) Tagged peptides are identified using a pattern-searching algorithm and inventoried into an inclusion list (rt = retention time). (E) The same sample is then subjected to a directed LC-MS/MS experiment where tandem MS is only performed on precursor ions defined in the inclusion list. (F) Finally, fragment mass spectra are subjected to database search to produce peptide and protein identifications.

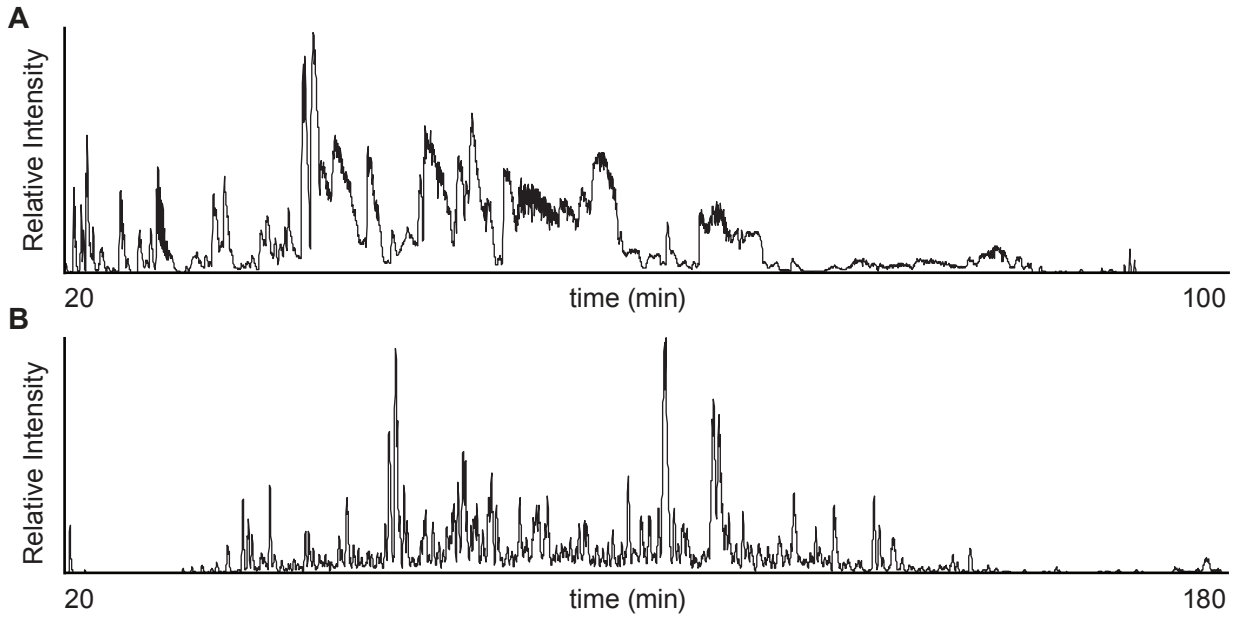



***Figure 3-9. Optimization of liquid chromatography conditions to achieve better separation of peptides.***
(A) Early attempts at the separation of peptides by reverse-phase chromatography suffered from poor sample separation and inconsistent results. (B) Improvements in chromatography were achieved by optimizing the procedure used to generate homemade capillary columns, changing column lengths from 10- to 15-cm, and adjusting the gradient conditions.

listed ion appeared in the correct retention time window and at an abundance above a specified threshold. This approach allowed us to focus the analysis on peptides of interest, using recoded mass envelopes as an indicator, and allowing for direct site mapping.

However, before we were able to implement the IsoStamp method for directed proteomics, there were several challenges to overcome. First, in order to achieve nanoliter flow rates, we manually pulled fused silica capillaries and packed them with C18 resin. Although this kept column costs low, it also introduced batch-to-batch variability and inconsistent results during chromatography (Figure 3-9A). However, by using a micropipette puller, increasing the length of C18 resin from 10-cm to 15-cm, optimizing mobile phase gradients based on sample complexity, and operating at higher pressures, we were able to improve peptide separation using the homemade capillary columns (Figure 3-9B). These improvements also satisfied another requirement for the IsoStamp method, reproducible chromatography between the full-scan and the directed experiment in order to ensure that peptides would elute at the time indicated on the inclusion list (Figure 3-10).

The advantage of using a time-resolved inclusion list for directed proteomics was empirically determined and is illustrated in Figure 3-11. Without time information (Figure 3-11A), where peptide ions were selected for fragmentation based solely on *m/z* values, the fragmented ions appeared randomly across the entire *m/z* range rather than overlapping with the elution profile of the peptide ions. In contrast, when using a time-resolved inclusion list (Figure 3-11B), there was a high correlation between the peptide ion elution profile and the ions selected for fragmentation. This led to higher quality fragmentation spectra that were more likely to contain relevant information. As a validation of the IsoStamp method, Figure 3-12 demonstrates how

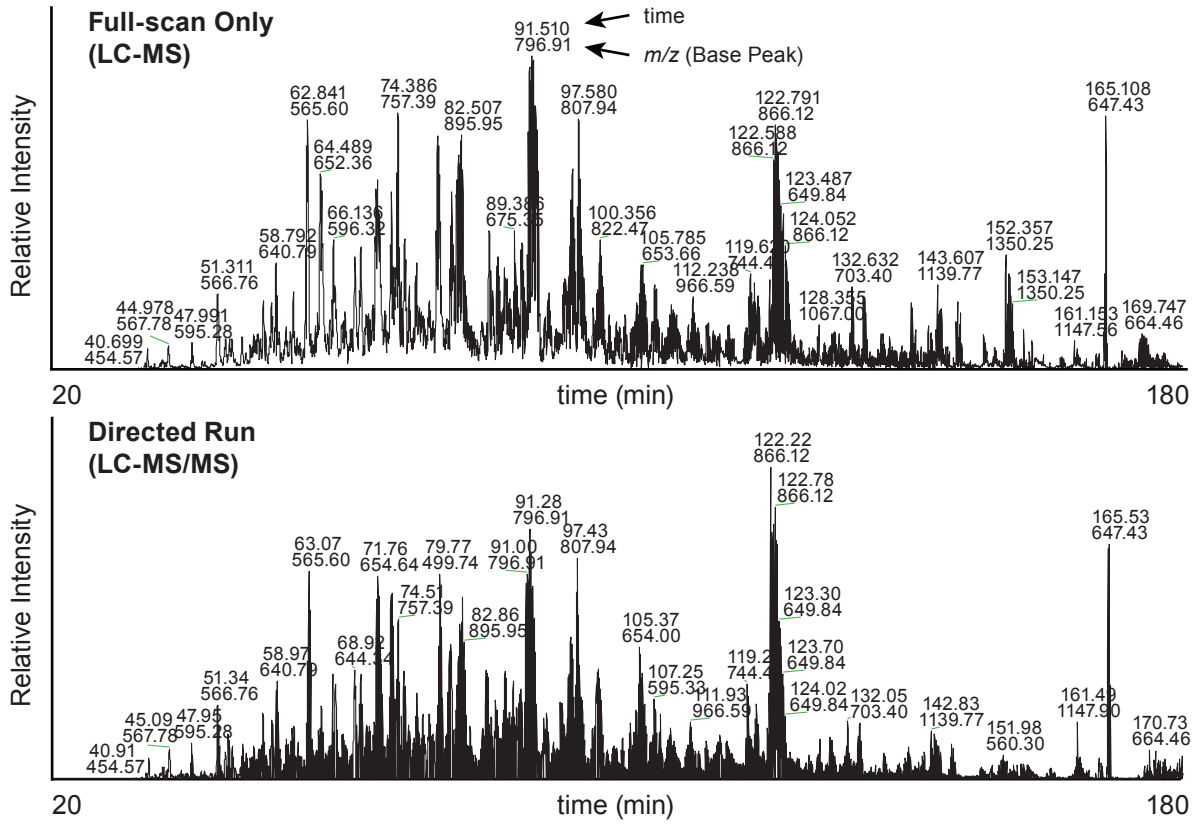



***Figure 3-10. The IsoStamp method requires reproducible chromatography between the full-scan and the directed experiments.***

An example of a full-scan chromatogram (from which the inclusion list is generated) and the corresponding directed chromatogram (where tandem MS is performed), illustrating the high level of reproducibility between the two sequential chromatography steps.

all of its components (i.e., chemical labeling, optimized and synchronized chromatography, computational-pattern searching, and time-based selection of *m/z* values for fragmentation) successfully worked together.

### *3.2.6 Application of IsoStamp to a model directed shotgun proteomics experiment*

After optimizing the IsoStamp method for directed proteomics, we were able to successfully perform a model directed LC-MS/MS experiment using dibromide-labeled BSA in the presence of 10 µg of Jurkat whole cell lysate. For the data analysis from this experiment, we focused on single cysteine-containing peptides. An example of one precursor ion from the inclusion list (*m/z* = 883.26) clearly displayed the dibromide pattern (Figure 3-13A), and its site of modification was identified after isolating the full isotopic envelope (highlighted in yellow) and performing collision-induced dissociation (CID) fragmentation (Figure 3-13B). A database search identified the ion as BSA peptide YIC*DNQDTISSK (residues 286-297). Fragment ions that contain the dibromide tag also display the perturbed isotopic envelope, as shown for the $y_{10}^{+}$ ion (Figure 3-13C), strengthening confidence in the site of modification (i.e., C*).

Next, we compared the number of single cysteine-containing BSA peptides identified

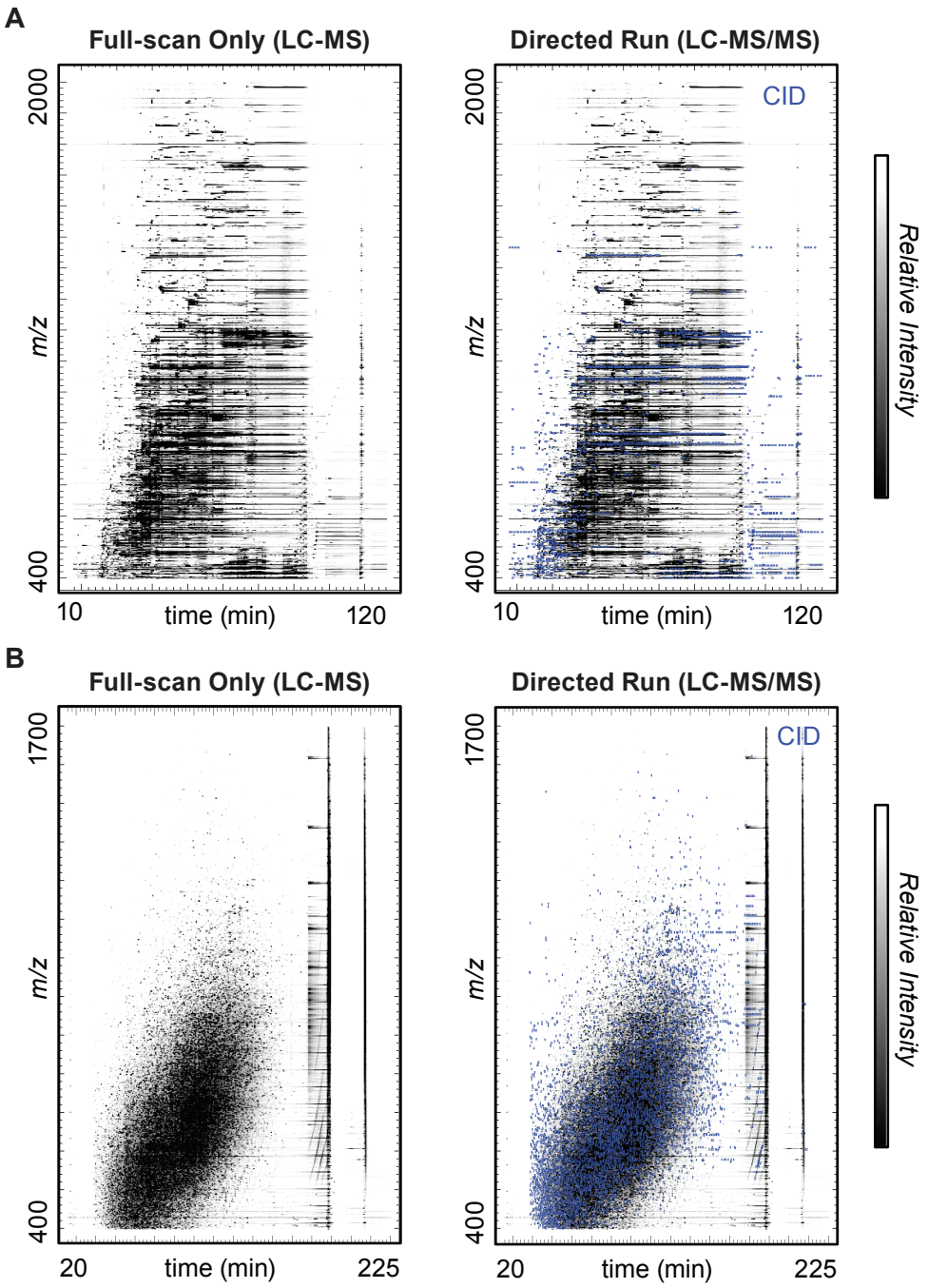



***Figure 3-11. A time-resolved inclusion list improves the performance of the IsoStamp method.***

(A) Without a time-resolved inclusion list, peptide ions selected for fragmentation (blue squares) appear randomly across the entire m/z range rather than overlapping with the elution profile of the peptide ions (black dots, where the color intensity correlates to ion intensity). (B) However, with a time-resolved inclusion list, peptide ions selected for fragmentation reproducibly overlap with the peptide elution profile. This leads to fragmentation spectra that are more likely to generate high-confidence peptide identifications.

using either the directed approach or a conventional data-dependent approach (DDA) in which peptides do not bear a detectable tag.[2] The directed approach identified more single cysteine-containing peptides at all tested concentrations of labeled BSA in 10 µg of Jurkat whole cell lysate (Figure 3-13D and Appendix A.7). The results of this analysis highlight the ability of chemically-directed proteomics methods to increase both the number of distinct peptides and the number of low-abundance peptides identified in complex mixtures. Because the unique isotopic signature can be maintained in the MS2 spectra by using a wide isolation window, modification sites are readily identified, which can facilitate studies of posttranslational modifications (a feature utilized in Chapter 7).

### *3.2.7 Application of IsoStamp to the identification of small proteins*

The impressive performance of the dibromide tag motivated us to explore its use in the identification of large peptides or small proteins. In addition to improving coverage and confidence in protein identifications, the analysis of high MW fragments enables studies of multiple posttranslational modifications (PTMs) that might occur on a single protein molecule, such as glycosylation.[46] Towards this goal, we labeled the small protein Barstar from *B. cenocepacia* (11.7 kDa, including an N-terminal 6xHis tag) with dibromide tag **3.4** on a single cysteine residue introduced by site-directed mutagenesis (I26C). To verify labeling, we subjected dibromide-labeled Barstar to a tryptic digestion and analyzed the peptide mixture by LC-MS on an LTQ-Orbitrap XL

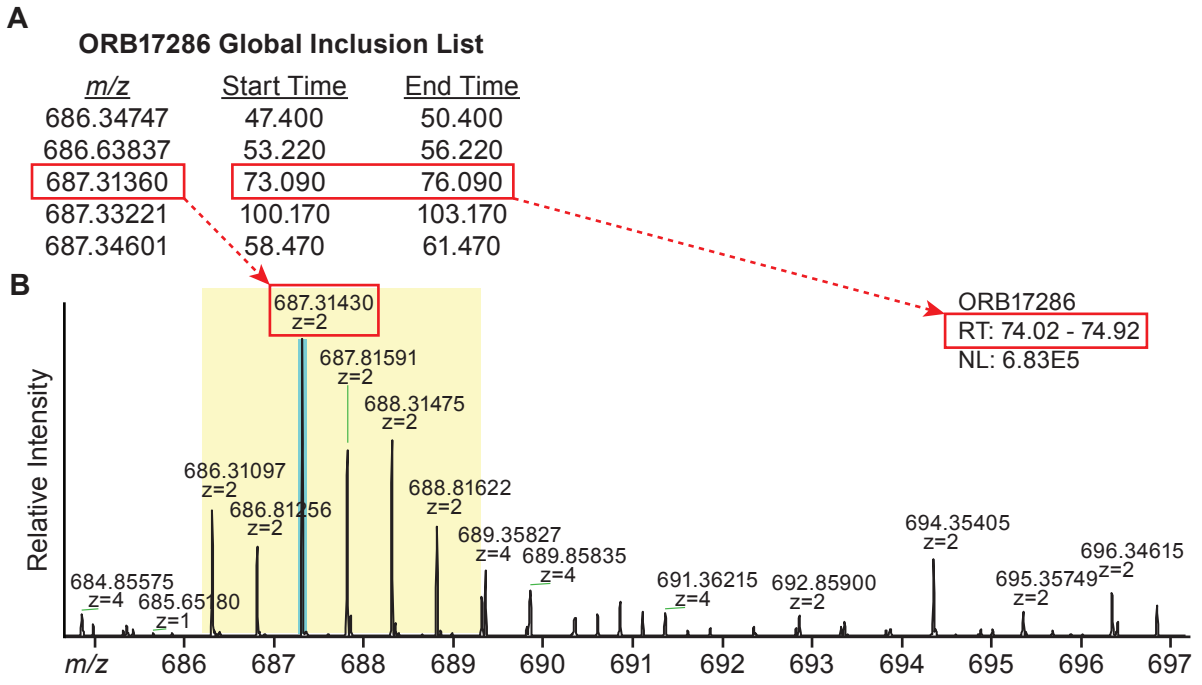



**Figure 3-12. An example of raw data that validate the IsoStamp method for directed proteomics.**

(A) A portion of the inclusion list from the ORB17286 data file, which lists the m/z values (the M+2 peak in the isotopic envelope) and retention time windows (±1.5 min) for all putative dibromide-labeled peptides. One peptide ion selected for fragmentation is highlighted in the red box. (B) A zoomed-in region of the ORB17286 full-scan chromatogram shows a dibromide-labeled peptide (isotopic envelope highlighted in yellow) that was selected for fragmentation because the instrument found an expected *m/z* value (687.3143, highlighted in teal), within error, and within the specified retention time window (red box).

mass spectrometer. After pattern-searching, we identified the single cysteine-containing Barstar peptide with a recoded isotopic envelope (Figure 3-14A). Next, the purified intact protein (100 picomoles) was analyzed by LC-MS. Shown in Figure 3-14B are the mass spectra of unlabeled (top) and dibromide-labeled (bottom) Barstar in the +9 charge state.

Using the averagine system, we predicted the mass envelopes of the protein with and without the dibromide tag and depicted the peak intensities in the form of red and blue curves, respectively. An overlay of the two curves suggests that addition of the dibromide tag should cause a detectable widening of the mass envelope (Figure 3-14B). In order to score how well

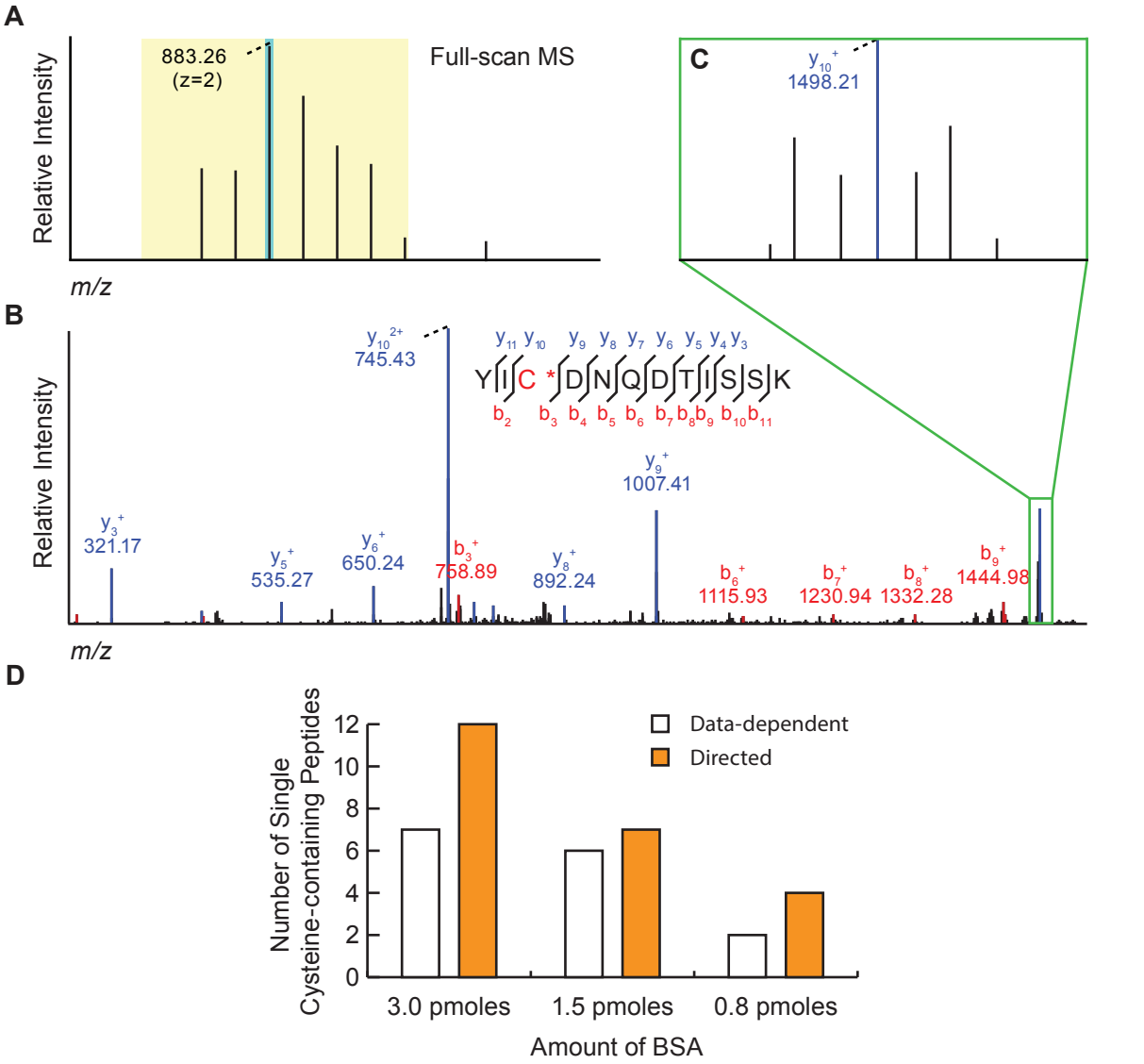



***Figure 3-13. IsoStamp using the dibromide tag in a model directed shotgun proteomics experiment.***

(A) The isotopic envelope of a precursor ion that was selected for fragmentation (883.26, highlighted in blue) and its isolation window (highlighted in yellow). (B) The CID fragmentation spectra and the peptide assignment for the 883.26 ion (C* refers to a cysteine residue alkylated with dibromide tag **3.4**). (C) The $y_{10}^{+}$ fragment ion, which contains the dibromide tag, also displays a perturbed isotopic envelope. (D) Numbers of unique peptides detected using a data-dependent verses a directed approach. The indicated amounts of BSA were added to 10 µg of Jurkat whole cell lysate prior to digestion and LC-MS/MS analysis.

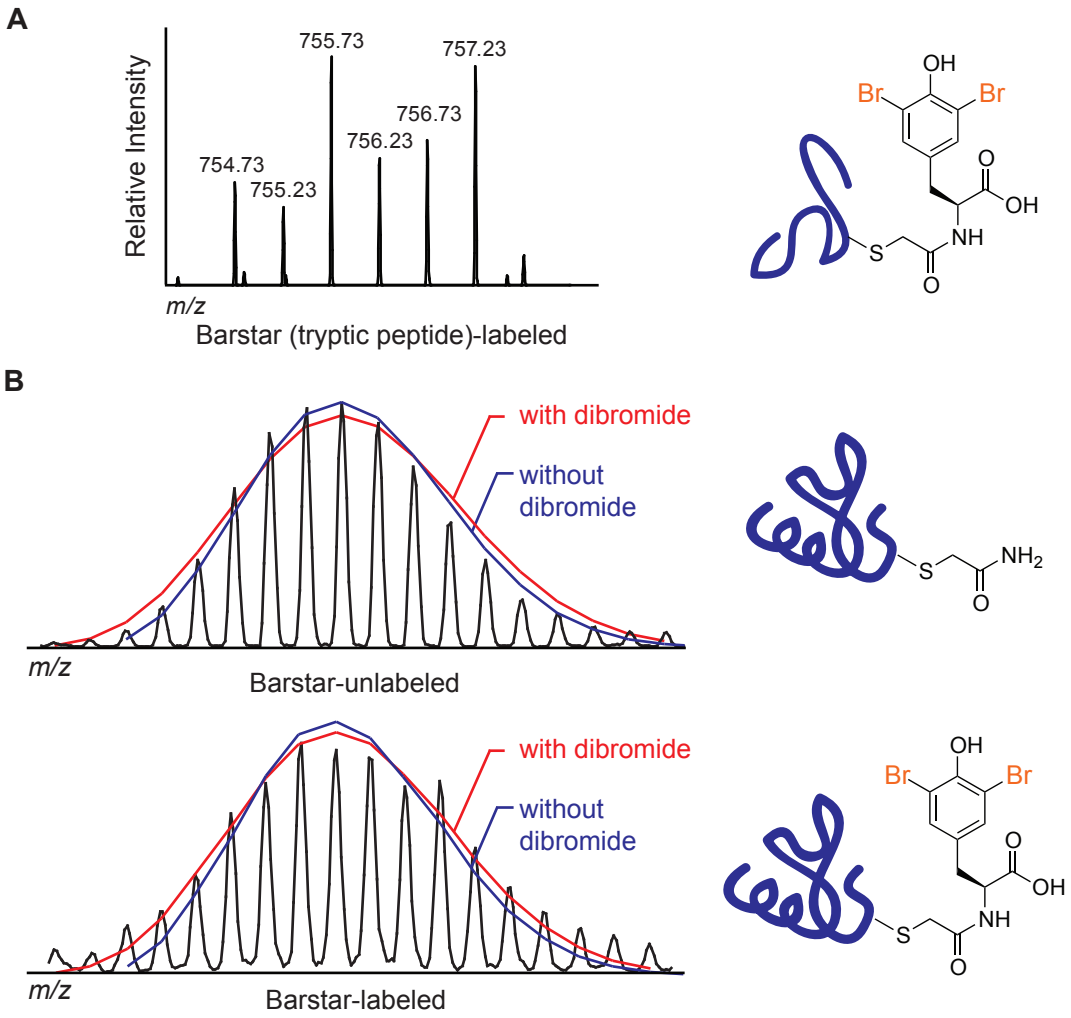



***Figure 3-14. The dibromide tag can be detected on purified small proteins.***

(A) After pattern-searching a tryptic digest of dibromide-labeled Barstar, we identified the single cysteine-containing Barstar peptide with a recoded isotopic envelope. (B) The averagine predicted isotopic envelope with (red) and without (blue) the dibromide tag fit to mass spectra of Barstar (top) and dibromide-labeled Barstar (bottom). Barstar's predicted MW = 11.7 kDa. Spectra shown are from the +9 charge state.

the averagine predictions matched the experimental data, we calculated the RMSD between the peak intensities from the MS data and each of the predicted mass envelopes (details can be found in the Methods section). Although the RMSD difference is small, the averagine model is able to distinguish the dibromide-labeled from the unlabeled protein. These results suggest that masses approaching 12 kDa are most likely at the limit of detection for the IsoStamp technology. Efforts are underway to refine the computational analysis so that we can identify a similarly sized protein at lower concentrations and in the presence of a complex mixture.

## 3.3 Conclusions

In summary, we have shown that recoding a peptide's isotopic envelope, in combination with a sophisticated pattern-searching algorithm, can enhance the performance of shotgun

proteomics. The IsoStamp method, an extension of Isotopic Distribution Encoding Tagging,[32] exploits the recoding effects of a dibrominated chemical tag on a peptide's isotopic envelope. While building additional halogens into a tag could yield an even more distinguishable isotopic signature, sensitivity for their detection may be compromised, as the same total signal intensity will be split among more peaks. The dibromide signature strikes a balance by enabling high-fidelity pattern matching with good sensitivity.

The ability to detect the dibromide pattern depends primarily on two instrument parameters: the *m/z* resolution and the desired sensitivity. It is possible to match the dibromide pattern on lower resolution instruments provided that the isotopic pattern itself is recreated faithfully. Alternatively, pattern matching on a higher mass-resolution instrument requires less fidelity in the isotopic pattern itself. The type of instrument that can be used is also dependent on the acceptable false positive rate, as higher acceptable false positive identifications will permit the use of lower resolution instruments. To test the sensitivity of detecting dibromide-labeled BSA at resolutions lower than 60,000, we performed LC-MS analysis on a Waters QTof mass spectrometer with a resolution of 10,000 (Figure 3-15). Known quantities of dibromide-labeled BSA added to 10 µg of Jurkat whole cell lysate were digested and then subjected to LC-MS analysis. After the data were searched for dibromide-labeled peptides, we found that the performance at a resolution of 10,000 on a QTof instrument was similar to results at a resolution of 60,000 on a Orbitrap instrument.

The IsoStamp method can be employed in any proteomics experiment in which a tag is covalently bound to target peptides; in principle, any dibrominated tagging reagent can be used. This concept pertains to many quantitative labeling approaches (i.e., ICAT, iTRAQ), where an isotopically-labeled tagging reagent is used to impart peptides with a defined mass shift or a particular reporter mass. Importantly, the IsoStamp method is distinct from these approaches. First, the perturbing effects of halogens on the isotopic envelope of a peptide are substantially greater that those of heavy H, N, C or O isotopes. Second, IsoStamp relies solely on this isotopic perturbation to identify a labeled peptide, with no requirement of a defined mass shift or a specific reporter mass fragment. We view IsoStamp as complementary to quantitative approaches, where the two proteomics tools can be merged into one experiment with the use of light and heavy dibromide tags. Preliminary evidence for this integration can be seen in Figure 3-16 where a heavy dibromide tag (**3.14**) was combined with the light dibromide

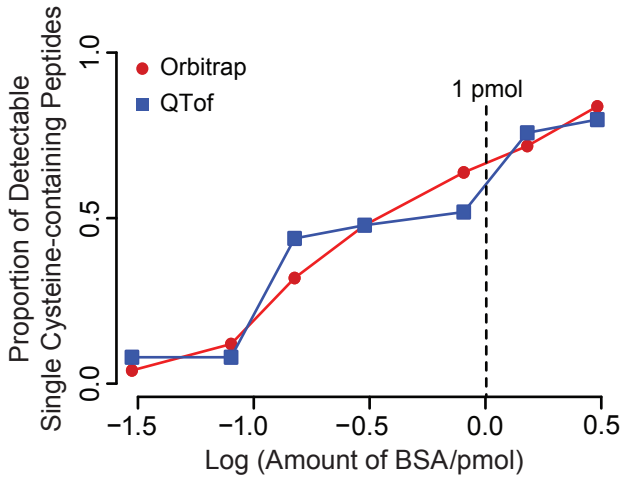



***Figure 3-15. The dibromide motif can be detected using both QTof and Orbitrap mass spectrometers.***

A comparison of the sensitivity of detecting dibromide-labeled BSA peptides was performed between a Waters QTof (blue square) mass spectrometer and an LTQ Orbitrap XL (red circle) mass spectrometer. Sensitivity was determined by titrating 3.0 to 0.03 picomoles of dibromide-tagged BSA into 10 µg of Jurkat whole cell lysate and analyzing the digest by LC-MS. At all concentrations both instruments perform almost identically.

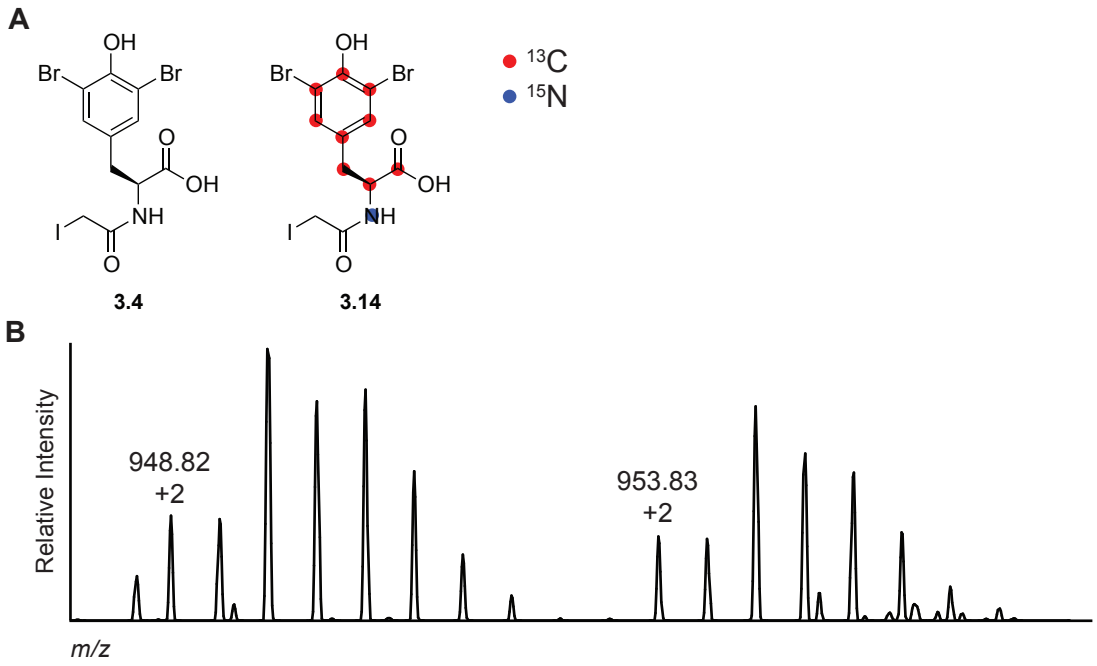



***Figure 3-16. Quantitative IsoStamp.***

(A) The structure of a heavy (+10 Da) dibromide tag (**3.14**) and a light dibromide tag (**3.4**) synthesized as an iodoacetamide-derivatized tyrosine analogs. (B) BSA labeled with a 1:1 mixture of dibromide tags **3.4** and **3.14**, respectively, was digested with trypsin and subjected to LC-MS analysis. After pattern-searching, a pair of light- and heavy-labeled peptides were found.

tag (**3.4**) to label BSA, and both light- and heavy-labeled peptides were observed.

More generally, we view IsoStamp as an enabling tool for chemical proteomics. The method can be incorporated into any proteomics experiment requiring a chemical tag, including emerging bioorthogonal ligation strategies that install uniquely reactive functionalities at sites of posttranslational modifications.[16] Affinity-based proteomics experiments in which tags are covalently bound to enzyme active site residues,[24] and protein chemical crosslinking studies[47] can also benefit from integration of the IsoStamp method. In all cases, including a dibromide signature in the covalently bound tag will improve detection and identification of labeled peptides.

Finally, the computationally detectable isotopic envelope pattern central to the IsoStamp method can be generated in ways other than covalent chemical labeling with bromine atoms. For example, metabolic labeling with a mixture of isotopolog substrates, molecules that differ only in their isotopic composition, could be used to biosynthetically endow biomolecules with similarly distorted isotopic envelopes. Consequently, we envision numerous future applications of IsoStamp in glycomics and metabolomics in addition to proteomics.

## 3.4 Materials and Methods

### *3.4.1 General procedures and materials*

All chemical reagents were of analytical grade, obtained from commercial suppliers, and

used without further purification unless otherwise noted. All reaction flasks were oven dried prior to use and moisture-sensitive reactions were performed in flame-dried flasks under a $N_2$ atmosphere. Analytical thin layer chromatography (TLC) was performed with 60 Å silica gel plates (EMD) and visualized by staining with ceric ammonium molybdate, ninhydrin, or by absorbance of UV light at 254 nm. Flash chromatography was performed with 60 Å 230-400 mesh silica (Silicycle). Organic extracts were dried over $MgSO_4$, and solvent was removed by reduced pressure with a R-114 rotovapro (Buchi) equipped with a self-cleaning dry vacuum pump (Welch). If necessary, products were further dried by reduced pressure with an Edwards RV5 high vacuum pump.

Water used in biological procedures was distilled and deionized using a Milli-Q Gradient ultrapure water purification system (Millipore). Dulbecco's phosphate-buffered saline (PBS) was obtained from HyClone Laboratories and fetal bovine serum (FBS) was obtained from Omega Scientific.

*3.4.2 Instrumentation*

**NMR.** All $^{1}H$ and $^{13}C$ spectra are reported in ppm and referenced to solvent peaks. Spectra were obtained on Bruker AVB-400 or AVQ-400 spectrometers. Coupling constants (J) are reported in Hz.

**Absorbance spectra.** Absorbance readings were collected on a Molecular Devices SpectraMax 190 UV-Vis absorbance plate reader, a CARY 100 Bio UV-Visible Spectrophotometer, or a Thermo Scientific Nanodrop 2000 spectrophotometer.

**High performance liquid chromatography (HPLC).** Purification of small molecules was performed using reversed-phase (RP) HPLC on a Rainin Dynamax SD-200 system equipped with a Varian UV-Vis detector (model 345) and a Microsorb C18 analytical column (4.6 x 250 mm) at a flow rate of 1 mL/min, a semipreparative column (10 x 250 mm) at a flow rate of 4 mL/min, or a preparative column (21.4 x 250 mm) at a flow rate of 20 mL/min. HPLC samples were filtered with a Pall Life Sciences Acrodisc CR 13 mm syringe filter equipped with a 0.2 µm PTFE membrane prior to injection.

**Electrospray ionization (ESI) MS.** High-resolution ESI mass spectra were obtained at the QB3 Mass Spectrometry Facility (UC Berkeley) using an LTQ Orbitrap XL hybrid mass spectrometer (Thermo Fisher). Low-resolution ESI mass spectra were obtained using an LTQ-XL mass spectrometer.

**Full-scan LC-MS on an LTQ Qribtrap XL.** Samples were subjected to RP chromatography with an Agilent 1200 LC system that was connected in-line with an LTQ Orbitrap XL hybrid mass spectrometer. External mass calibration was performed prior to analysis. A binary solvent system consisting of buffer A (0.1% formic acid in water (v/v)) and buffer B (0.1% formic acid in acetonitrile (v/v)) was employed.

For peptide samples, the mass spectrometer was outfitted with a nanospray ionization

source. The LC was performed using a 100 µm fritted capillary (New Objective) pre-column self-packed with 1 cm of 5 µm, 200 Å Magic C18AQ resin (Michrom Bioresources) followed by a 100 µm fused silica capillary (Polymicro Technologies) self-packed with 10 cm of 5 µm, 100 Å Magic C18AQ resin (Michrom Bioresources). After injection of the sample and a 10 min loading step in 2% buffer B, a gradient was employed from 10% to 40% buffer B for 62 min, followed by a washing step in 99% buffer B for 10 min. A solvent split was used to maintain a flow rate of 200 nL/min at the column tip. Mass spectra were recorded in a single stage of MS in positive ion mode over the *m/z* scan range of 400 to 2000 using the Orbitrap mass analyzer in full-scan, profile mode, at a resolution of 60,000 (at 400 *m/z*).

For the intact Barstar samples, the mass spectrometer was outfitted with an Ion Max electrospray ionization source. The LC was equipped with a Poroshell 300SB-C8 column and a 100 _L sample loop. For each run, after 100 to 200 picomoles of protein was injected onto the column, analyte trapping was performed for 5 min with 0.5% B, followed by a linear gradient from 30% to 95% B over 19.5 min, and finished with a washing step in 95% B for 5 min. The flow rate was maintained at 90 µL/min. Mass spectra were recorded in positive ion mode over the *m/z* scan range of 500 to 2000 using the Orbitrap mass analyzer in full-scan, profile mode. Raw mass spectra were processed using Xcalibur (version 4.1, Thermo) and measured charge state distributions were deconvoluted using ProMass (version 2.5 SR-1, Novatia), using the default "small protein" parameters and a background subtraction factor of 1.5.

**Full-scan LC-MS on a Waters QTof Premier.** Samples were subjected to RP chromatography with a nanoAcquity ultraperformance LC (UPLC, Waters) system that was connected in-line with a Waters QTof Premier mass spectrometer. External mass calibration was performed prior to analysis using sodium formate cluster ions. A binary solvent system consisting of buffer A (0.1% formic acid in water (v/v)) and buffer B (0.1% formic acid in acetonitrile (v/v)) was employed.

For peptide samples, the UPLC column exit was connected to a Universal NanoFlow Sprayer that was mounted in the ion source of the mass spectrometer. LC was performed using a 20 mm x 180 µm C18 trapping column (5 µm particles, Waters) followed by a 100 mm x 100 µm C18 analytical column (1.7 µm particles, Waters) and a 10 µL sample loop. After injection of the sample and a 5 min loading step in 0% buffer B, a gradient was employed from 25% to 45% buffer B for 60 min, followed by a washing step in 95% buffer B for 5 min. The flow rate was maintained at 500 nL/min at the column tip. The Tof analyzer was operated in "V" mode. The mass resolving power was $1.0 \times 104$ (measured at m/z = 771). Mass spectra were processed using MassLynx software (version 4.1, Waters).

**Tandem LC-MS/MS on an LTQ Qribtrap XL.** Samples were subjected to RP chromatography with an Agilent 1200 LC system that was connected in-line with an LTQ Orbitrap XL hybrid mass spectrometer. External mass calibration was performed prior to analysis. A binary solvent system consisting of buffer A (0.1% formic acid in water (v/v)) and buffer B (0.1% formic acid in acetonitrile (v/v)) was employed. The mass spectrometer was outfitted with a nanospray ionization source. The LC was performed using a 100 µm fritted capillary (New Objective) pre-column self-packed with 1 cm of 5 µm, 200 Å Magic C18AQ resin (Michrom Bioresources) followed by a 100

μm fused silica capillary (Polymicro Technologies) self-packed with 15 cm of 5 μm, 100 Å Magic C18AQ resin (Michrom Bioresources). After injection of the sample and a 20 min. loading step in 2% buffer B, a gradient was employed from 10% to 35% buffer B for 120 min, followed by a washing step in 99% buffer B for 15 min. A solvent split was used to maintain a flow rate of 200 nL/min at the column tip.

Iodoacetamide-labeled BSA samples were subjected to a data-dependent acquisition method, where each full-scan mass spectra was recorded in positive ion mode over the *m/z* scan range of 400 to 1800 using the Orbitrap mass analyzer in profile mode at a resolution of 60,000 (at 400 *m/z*). The ten most intense peaks were selected for collision induced dissociation (CID) fragmentation in the linear ion trap. Dynamic exclusion and charge state screening were enabled, rejecting ions with an unknown or +1 charge state. An isolation window of 3, a threshold of 500 ion counts, and activation energy of 35 were used when triggering a fragmentation event.

Dibromide-labeled BSA samples were subjected to an inclusion-list driven targeted acquisition method. The first step was collecting only full-scan mass spectra in positive ion mode over the *m/z* scan range of 400 to 1800 using the Orbitrap mass analyzer in profile mode at a resolution of 60,000 (at 400 m/z). The data were searched for the presence of dibromide-labeled peptides and each labeled species was recorded in an inclusion list, which contained the *m/z* value (M+2 ion in the labeled peptide's isotopic envelope) and a retention time window for each labeled peptide. The same sample (and same injection volume) was then reanalyzed using an inclusion-list driven selection of precursor ions for fragmentation. The inclusion list was entered into the global mass list, and for each full-scan mass spectra up to eight CID fragmentation events could take place in the linear ion trap. Dynamic exclusion and charge state screening were enabled, rejecting ions with an unknown or +1 charge state. An isolation window of 4, a threshold of 500 ion counts, and activation energy of 35 were used when triggering a fragmentation event.

### *3.4.3 Chemical synthesis*

**General procedure for the preparation of halogenated tyrosine salts (3.2, 3.5, 3.8, 3.9).** Starting with L-tyrosine (**3.1**), halogenated tyrosine salts were prepared according to literature procedures.[48-50]

**General procedure for the preparation of iodoacetamide derivatives of halogenated tyrosine analogs (3.4, 3.7, 3.12, 3.13).** To the halogenated tyrosine salt (100 mg) in anhydrous DMF (250 μL) was added anhydrous sodium carbonate (2 equiv.) and the mixture was left to stir for 0.5 h at RT under a $N_2$ atmosphere. Chloroacetyl chloride (1 equiv.) was added dropwise over 5 min and the reaction was stirred at RT under a $N_2$ atmosphere for 1 h before the reaction was transferred to a separatory funnel with ethyl acetate (15 mL). The organic layer was washed with 1 M HCl (5 mL) and the layers separated. The aqueous layer was extracted twice with ethyl acetate (5 mL) and the combined organic extracts were dried over $MgSO_4$ prior to removal of solvent by rotary evaporation to generate compounds **3.3**, **3.6**, **3.10** and **3.11**. The crude material was then dissolved in anhydrous DMF (0.5 mL) and sodium iodide was added (6 equiv.). The mixture was stirred at RT in the dark under a $N_2$ atmosphere for 24 h before the mixture was

transferred to a separatory funnel with ethyl acetate (15 mL). The organic layer was washed with water (5 mL) and the layers were separated. The aqueous layer was extracted twice with ethyl acetate (5 mL) and the combined organic extracts were dried over $MgSO_4$. The crude product was purified by RP HPLC using a C18 column and a water/acetonitrile solvent system with 0.1% TFA with a gradient of 15% to 60% acetonitrile over 50 min. All fractions were kept in foil until the solvent was removed by rotary evaporation to yield a white solid. **Compound 3.4**. $^1$H NMR (400 MHz, $CD_3OD$): δ = 7.36 (s, 2H), 4.59-4.55 (m, 1H), 3.73 (d, *J* = 9.6 Hz, 1H), 3.64 (d, *J* = 9.6 Hz, 1H), 3.11 (dd, *J* = 14, 4.8 Hz, 1H), 2.89-2.83 (m, 1H). **Compound 3.7**. $^1$H NMR (400 MHz, $CD_3OD$): δ = 7.33 (d, *J* = 2 Hz, 1H), 7.03 (dd, *J* = 8.4, 2 Hz, 1H), 6.80 (d, *J* = 8.4 Hz, 1H), 4.57-4.54 (m, 1H), 3.72 (d, *J* = 10 Hz, 1H), 3.65 (d, *J* = 10 Hz, 1H), 3.09 (dd, *J* = 14, 4.8 Hz, 1H), 2.88-2.83 (m, 1H). **Compound 3.12**. $^1$H NMR (400 MHz, $CD_3OD$): δ = 7.16 (s, 2H), 4.56-4.53 (m, 1H), 3.73 (d, *J* = 9.6 Hz, 1H), 3.64 (d, *J* = 9.6 Hz, 1H), 3.11 (dd, *J* = 14, 4.8 Hz, 1H), 2.89-2.83 (m, 1H). **Compound 3.13**. $^1$H NMR (400 MHz, $CD_3OD$): δ = 7.17 (d, *J* = 2 Hz, 1H), 6.99 (dd, *J* = 8.4, 2 Hz, 1H), 6.81 (d, *J* = 8.4 Hz, 1H), 4.58-4.54 (m, 1H), 3.72 (d, *J* = 10 Hz, 1H), 3.65 (d, *J* = 10 Hz, 1H), 3.09 (dd, *J* = 14, 4.8 Hz, 1H), 2.89-2.83 (m, 1H).

**Heavy dibromide tyrosine analog (3.14).** Starting with heavy (+10 Da) L-tyrosine ($^{13}$C9, $^{15}$N1; Cambridge Isotopes) an identical synthesis to **3.4** was followed. The crude product was purified by RP HPLC using a C18 column and a water/acetonitrile solvent system with 0.1% TFA with a gradient of 20% to 35% acetonitrile over 40 min. All fractions were kept in foil until the solvent was removed by rotary evaporation to yield a white solid. Since $^{13}$C and $^{15}$N could complicate $^1$H NMR spectra, $^{13}$C signals were decoupled. $^1$H NMR (400 MHz, $CD_3OD$): δ = 7.35 (s, 2H), 4.08 (q, *J* = 7.1 Hz, 1H), 3.73 (d, 1H, *J* = 12 Hz), 3.64 (dd, 1H, *J* = 4, 10 Hz), 3.08 (d, *J* = 4.1 Hz, 1H), 2.92-2.81 (m, 1H).

### *3.4.4 Experimental*

**Cell culture**. Jurkat cells were maintained according to ATCC guidelines, grown in a 5% $CO_2$, humidified atmosphere at 37 °C in RPMI-1640 media (Invitrogen Life Technologies) supplemented with 10% FBS, penicillin (100 units/mL), and streptomycin (0.1 mg/mL).

**BSA labeled with cysteine-alkylating tags.** A 100 µg aliquot of a 2 mg/mL solution of BSA in 250 mM ammonium bicarbonate was reduced by adding DTT to a concentration of 2.5 mM and placed at 55 °C for 30 min. After cooling to RT, the halogenated tag was added to a concentration of 10 mM from a 500 mM solution in DMF. The reaction proceed for 1 h in the dark before quenching with 5 µL of 1 M DTT for 30 min. Excess tag was removed by size exclusion chromatography using a Bio-Rad Micro Bio-Spin 6 column.

**Preparation of Jurkat whole cell lysate.** Jurkat cells were lysed in a buffer containing 1% TritonX-100, 20 mM Tris pH 7.4, 150 mM NaCl and protease inhibitors (inhibitor cocktail III from Calbiochem). Following lysis, the sample was precipitated using 9 volumes of acetone and placed at –20 °C for 4 h followed by centrifugation at 13,000 rpms for 20 min at 4 °C. The supernatant was removed and the pellet was resolubilized in 8 M urea buffered to pH 8.0. A BCA assay (Thermo Scientific Pierce) was performed to determine protein concentration. Samples were diluted to 1 mg/mL and stored at –20 °C.

**Serial dilutions of tag-labeled BSA in Jurkat whole cell lysate.** For full-scan LC-MS analysis, tag-labeled BSA samples were serially diluted into 10 µL of 1 mg/mL Jurkat whole cell lysate at amounts of 3.0, 1.50, 0.80, 0.30, 0.15, 0.08, and 0.03 picomoles. For data-dependent or directed LC-MS/MS analysis, either iodoacetamide or dibromide-labeled BSA samples, respectively, were serially diluted into 20 µL of 1 mg/mL Jurkat whole cell lysate at amounts of 6.0, 3.0, 1.60, and 0.60 picomoles (amounts were doubled so that two sequential injections could be performed). The samples were trypsin digested (50:1 protein/enzyme) at 37 °C for 16 h and desalted using C18 zip tips (Millipore) or C18 spin columns (Nest group) according to the manufacturer's instructions.

**Barstar mutagenesis and purification.** A plasmid containing the *Bacillus amyloliquefaciens* protein Barstar as a 6xHis fusion in a pQE30 expression vector was obtained from D. Tirrell (California Insititue of Technology). A construct encoding the single point mutant I26C was prepared using the Quickchange protocol (Stratagene).

Forward 5′ – GGG GAA CAA ATC AGA AGT TGC AGC GAC CTC CAC CAG AC –3′

Reverse 5′ – GTC TGG TGG AGG TCG CTG CAA CTT CTG ATT TGT TCC CC –3′

The mutant was expressed and purified analogously to the wild-type protein.[51] The purified I26C Barstar mutant was verified by ESI-HRMS, calcd for $C_{518}H_{802}N_{146}O_{156}S_3$ [Avg. mass]: 11667.13 Da, found 11667.2 Da.

**Barstar labeled with a dibromide tag.** A 50 µL aliquot of a 0.4 mg/mL solution of I26C Barstar in 250 mM ammonium bicarbonate was reduced by adding DTT to a concentration of 2.5 mM and placed at 55 °C for 30 min. After cooling to RT, dibromide tag **3.4** was added to a concentration of 10 mM from a 500 mM solution in DMF. The reaction proceed for 1 h in the dark before quenching with 5 µL of 1 M DTT for 30 min. Following labeling, the protein was precipitated using 9 volumes of acetone and placed at –20 °C for 1 h followed by centrifugation at 13,000 rpms for 20 min at 4 °C. The supernatant was removed and the pellet was resolubilized in 250 mM ammonium bicarbonate buffer.

**Tryptic digestion of dibromide-labeled Barstar.** Dibromide-labeled Barstar was digested with trypsin (50:1 protein/enzyme) at 37 °C for 13 h and desalted using C18 zip tips (Millipore) according to the manufacturer's instructions.

**Preparation of light-heavy dibromide tag mixture.** Stock solutions of 200 mM light dibromide tag (**3.4**) and 500 mM heavy dibromide tag (**3.14**) in DMF were prepared. The solutions were combined to produce a 1:1 light:heavy mixture of the dibromide tags. The mixture was analyzed by direct infusion on an LTQ-XL mass spectrometer set to zoom scan, with the signal averaged over 20 scans. The isotopic ratios were then adjusted by the iterative addition of the desired tag until an approximately 1:1 peak intensity ratio was observed empirically.

**BSA labeled with a light-heavy dibromide tag.** BSA was labeled with the light-heavy dibromide tag mixture following the identical protocol for labeling BSA with the halogenated cysteine alkylating tags. The sample was then digested with trypsin (50:1 protein/enzyme) at 37

°C for 15 h and desalted using Millipore C18 zip tips according to the manufacturer's instructions.

*3.4.5 Data analysis*

**True positive determination.** This was done by performing an *in silico* digestion of BSA (up to two tryptic missed cleavages and a mass range of 600 to 4000 Da) and calculating the masses of tag-labeled single cysteine-containing peptides. From this list of theoretical masses, LC-MS data from samples composed of 30 picomoles of halogen-tagged BSA were manually searched by obtaining an extracted ion chromatogram (EIC) for each predicted mass (in up to 5 charge states). Each EIC was used to determine the visual presence of the appropriate halogen pattern. Peptides were considered a true positive if at least one charge state of the predicted mass value with the appropriate isotopic signature was observed (see Appendix A.6).

**Full-scan data analysis and inclusion list generation**. All full-scan LC-MS data were collected in profile mode. Noise reduction and peak detection were performed using software developed in-house based on the method described by Du *et al.,* which made use of a continuous wavelet transform.[52] The centroided mzXML files were then searched for the presence of a desired isotopic pattern using software developed in-house following the algorithm described in the main text (additional details, including source code, can be found in Austin Pitcher's thesis). The output was an inclusion list that contained the *m/z* values and a retention time window of three minutes for each tagged species.

**Data analysis of tandem mass spectra.** Tandem mass spectra were searched using Proteome Discoverer 1.2 (Sequest, Thermo-Fisher) against the human database (UniProtKB/Swiss-Prot, November 2009) with the Bovine Serum Albumin sequence appended (Uniprot P02769), allowing for three missed cleavages, one non-tryptic termini, and the following variable modifications: methionine oxidation and either cysteine carboxyamidomethylation or cysteine labeling with dibromide tag **3.4** (allowing for all three isotopologs). Precursor ion tolerance was set to 10 ppm and fragment tolerance was set to 0.8 Da. An FDR of 1% was used for filtering iodoacetamide-labeled peptide assignments. For dibromide-labeled peptide assignments the additional high intensity isotope peaks associated with each fragment ion that contained the dibromide tag lowered Sequests's XCorr calculation. Therefore, an FDR of 8% and a delta of 10 ppm for the precursor mass were used to narrow the search results, and each correct assignment was manually verified.

**Generation of ROC plots.** To generate ROC plots for each halogen tag we made the following four assumptions:

1. Total positives (P) were defined for each sample of halogen-labeled BSA based on the total *expected* peptides to be found, as defined from manual searching of 30 picomoles of BSA in the absence of lysate (see Appendix A.6). This was 25 for dibromide-, 19 for dichloride-, and 24 for monobromide-labeled BSA.

2. True positives (TP) were determined from a computational search for each single cysteine-containing BSA peptide from searching a sample containing the indicated quantity of halo-

gen-tagged BSA in the presence of 10 µg of Jurkat whole cell lysate. In order to vary the TP values we adjusted tolerances for key parameters of the graph-theoretic algorithm.

3. Total Negatives (N) were defined as the total number of peptides in the digested sample that are not labeled with a halogen tag (i.e., the number of peptides in trypsinized Jurkat whole cell lysate). For this calculation we approximated that there are 35,000 proteins in the human genome and that we would expect a total of 10 tryptic peptides per protein, giving 350,000 total peptides.

4. False Positives (FP) were determined as the difference between all ions that were computationally identified and the number of true positive ions. Unfortunately, we could not convert the number of ions into the number of peptides because we don't know what peptides each ion represents. In contrast, when determining the TP we could convert ions into peptides because we were only considering peptides from BSA. In order to vary the FP values we adjusted tolerances for key parameters of the graph-theoretic algorithm.

Therefore, the ROC plots were generated as:

True Positive Rate (TPR) = $TP_{peptides}/P_{peptides}$
False positive rate (FPR) = $FP_{ions}/N_{peptides}$

This implied that the calculated FPR is an over approximation (since we report ions/peptides instead of peptides/peptides), and the actual FPR is expected to be better than reported.

**Analysis of dibromide-labeled intact-Barstar with the averagine model.** To determine whether the dibromide tag's isotopic signature can be detected on large peptides, we labeled the small protein Barstar with dibromide tag **3.4** on a single cysteine residue introduced by site-directed mutagenesis (I26C). The labeled intact protein was then analyzed by LC-MS on an LTQ-Orbitrap XL mass spectrometer.

Using the averagine model, we predicted the mass envelope of Barstar with and without the dibromide tag and depicted the peak intensities in the form of red and blue curves, respectively. In order to score how well the averagine predictions matched the experimental data, we calculated the RMSD between the peak intensities from the MS data and each of the predicted mass envelopes.

The averagine-based models for isotopic envelopes of Barstar labeled with or without the dibromide tag **3.4** were predicted based on the MW of Barstar (monoisotopic mass of 11659.84 Da). Models (reference pattern, $R$) were fit against centroided experimental data by first determining the value of a scaling factor *k* such that the reference pattern could be optimally aligned with the data ($D$) in the intensity dimension. This was accomplished by performing a binary search such that the $SSD$ between each peak in the reference pattern ($r_i \in R$) and its counterpart in the actual data set ($d_i \in D$) is minimized according to Equation 3-1. After intensity alignment, the score for the fit was calculated as:

$$Score = \prod_{i=1}^{N} f\left(\frac{|d_i - kr_i|}{\sigma\sqrt{2}}\right) \quad (3\text{-}5)$$

$\sigma$ is a measure of peak intensity variance and $f$ is a scoring function for each peak that produces a value in the range [0,1]

$$f(x) = (1-\varepsilon)\mathrm{erfc}(x) + \varepsilon, \qquad 0 < \varepsilon << 1 \tag{3-6}$$

in which *erfc(x)* is the complement of the Gaussian error function and the parameter is a measure of the tightness of the peak matching in the intensity dimension. In the case of measuring goodness of fit for the Barstar data, the score was calculated using values of $\varepsilon = 10^{-5}$, $\sigma = 25\%$ (RMS intensity), and N = 16. Because scores produced in this manner are typically very small, the log values of the scores were compared and are presented in Table 3-3.

***Table 3-3. RMSD values for averagine models of dibromide-labeled Barstar.***

| Model | Unlabeled Barstar | Labeled Barstar |
|---|---|---|
| Averagine | -17.2 | -35.5 |
| Averagine + $Br_2$ | -21.7 | -20.5 |

## 3.5 References

1. Domon B. and Aebersold R. (**2006**) Mass spectrometry and protein analysis. *Science* 312, 212-217.

2. McDonald W.H. and Yates J.R., 3rd (**2003**) Shotgun proteomics: integrating technologies to answer biological questions. *Curr Opin Mol Ther* 5, 302-309.

3. Eng J.K., McCormack A.L. and Yates J.R., 3rd (**1994**) An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *Journal of the American Society for Mass Spectrometry* 5, 976-989.

4. Mann M. and Wilm M. (**1994**) Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal Chem* 66, 4390-4399.

5. Patterson S.D. and Aebersold R. (**1995**) Mass spectrometric approaches for the identification of gel-separated proteins. *Electrophoresis* 16, 1791-1814.

6. Anderson N.L. and Anderson N.G. (**2002**) The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* 1, 845-867.

7. Picotti P., Aebersold R. and Domon B. (**2007**) The implications of proteolytic background for shotgun proteomics. *Mol. Cel. Proteomics* 6, 1589-1598.

8. Schmidt A., Claassen M. and Aebersold R. (**2009**) Directed mass spectrometry: towards hypothesis-driven proteomics. *Cur. Op. Chem. Biol.* 13, 510-517.

9. Savitski M.M. *et al.* (**2010**) Targeted data acquisition for improved reproducibility and robustness of proteomic mass spectrometry assays. *Journal of the American Society for Mass Spectrometry* 21, 1668-1679.

10. Huttenhain R., Malmstrom J., Picotti P. and Aebersold R. (**2009**) Perspectives of targeted mass spectrometry for protein biomarker verification. *Curr Opin Chem Biol* 13, 518-525.

11. Jiang D., Jarrett H.W. and Haskins W.E. (**2009**) Methods for proteomic analysis of transcription factors. *J Chromatogr A* 1216, 6881-6889.

12. Leitner A. and Lindner W. (**2004**) Current chemical tagging strategies for proteome analysis by mass spectrometry. *Journal of chromatography* 813, 1-26.

13. Gygi S.P. *et al.* (**1999**) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnology* 17, 994-999.

14. Ross P.L. *et al.* (**2004**) Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* 3, 1154-1169.

15. Whetstone P.A., Butlin N.G., Corneillie T.M. and Meares C.F. (**2004**) Element-coded affinity tags for peptides and proteins. *Bioconjug Chem* 15, 3-6.

16. Heal W.P. and Tate E.W. (**2010**) Getting a chemical handle on protein post-translational modification. *Org Biomol Chem* 8, 731-738.

17. Saxon E. and Bertozzi C.R. (**2000**) Cell surface engineering by a modified Staudinger reaction. *Science* 287, 2007-2010.

18. Roth A.F. *et al.* (**2006**) Global analysis of protein palmitoylation in yeast. *Cell* 125, 1003-1013.

19. Heal W.P., Wickramasinghe S.R., Leatherbarrow R.J. and Tate E.W. (**2008**) N-Myristoyl transferase-mediated protein labelling in vivo. *Org Biomol Chem* 6, 2308-2315.

20. Yang Y.Y., Ascano J.M. and Hang H.C. (**2010**) Bioorthogonal chemical reporters for monitoring protein acetylation. *J Am Chem Soc* 132, 3640-3641.

21. Blethrow J.D., Glavy J.S., Morgan D.O. and Shokat K.M. (**2008**) Covalent capture of kinase-specific phosphopeptides reveals Cdk1-cyclin B substrates. *Proc Natl Acad Sci U S A* 105, 1442-1447.

22. Agard N.J., Maltby D. and Wells J.A. (**2010**) Inflammatory stimuli regulate caspase substrate profiles. *Mol Cell Proteomics* 9, 880-893.

23. Leonard S.E. and Carroll K.S. (**2011**) Chemical 'omics' approaches for understanding protein cysteine oxidation in biology. *Curr Opin Chem Biol* 15, 88-102.

24. Meier J.L. *et al.* (**2009**) An orthogonal active site identification system (OASIS) for proteomic profiling of natural product biosynthesis. *ACS Chem Biol* 4, 948-957.

25. Petrotchenko E.V., Serpa J.J. and Borchers C.H. (**2010**) Use of a combination of isotopically coded cross-linkers and isotopically coded N-terminal modification reagents for selective identification of inter-peptide crosslinks. *Anal Chem* 82, 817-823.

26. Mirzaei H. *et al.* (**2009**) Halogenated peptides as internal standards (H-PINS): introduction of an MS-based internal standard set for liquid chromatography-mass spectrometry. *Mol Cell Proteomics* 8, 1934-1946.

27. Carlson E.E. and Cravatt B.F. (**2007**) Enrichment tags for enhanced-resolution profiling of the polar metabolome. *J Am Chem Soc* 129, 15780-15782.

28. Li M. and Kinzer J.A. (**2003**) Structural analysis of oligosaccharides by a combination of electrospray mass spectrometry and bromine isotope tagging of reducing-end sugars with 2-amino-5-bromopyridine. *Rapid Commun Mass Spectrom* 17, 1462-1466.

29. Paulick M.G. *et al.* (**2006**) Cleavable hydrophilic linker for one-bead-one-compound sequencing of oligomer libraries by tandem mass spectrometry. *Journal of combinatorial chemistry* 8, 417-426.

30. Buser H., Oehme M., Vetter W. and Luckas B. (**1993**) Partial structure elucidation of the most abundant octa and nonachlorotoxaphene congeners in marine mammals by using conventional electron ionization mass spectrometry and mass spectrometry/mass spectrometry. *Fresenius J Anal Chem* 347, 502-512.

31. Enjalbal C. *et al.* (**2002**) Utilization of charge and mass labeling for the structural identification of heterocyclic quaternary salts by mass spectrometry. *Int J Mass Spectrom*.

32. Goodlett D.R. *et al.* (**2000**) Protein identification with a single accurate mass of a cysteine-containing peptide and constrained database searching. *Anal Chem* 72, 1112-1118.

33. Miyagi M., Nakao M., Nakazawa T., Kato I. and Tsunasawa S. (**1998**) A novel derivatization method with 5-bromonicotinic acid N-hydroxysuccinimide for determination of the amino acid sequences of peptides. *Rapid Commun Mass Spectrom* 12, 603-608.

34. Yang Y.Y., Grammel M., Raghavan A.S., Charron G. and Hang H.C. (**2010**) Comparative analysis of cleavable azobenzene-based affinity tags for bioorthogonal chemical proteomics. *Chem*

*Biol* 17, 1212-1222.

35. Hall M.P. and Schneider L.V. (**2004**) Isotope-differentiated binding energy shift tags (IDBEST) for improved targeted biomarker discovery and validation. *Expert review of proteomics* 1, 421-431.

36. Hernandez H. *et al.* (**2006**) Mass defect labeling of cysteine for improving peptide assignment in shotgun proteomic analyses. *Anal Chem* 78, 3417-3423.

37. Zhang H., Zhu M., Ray K.L., Ma L. and Zhang D. (**2008**) Mass defect profiles of biological matrices and the general applicability of mass defect filtering for metabolite detection. *Rapid Commun Mass Spectrom* 22, 2082-2088.

38. Desiere F. *et al.* (**2005**) Integration with the human genome of peptide sequences obtained by high-throughput mass spectrometry. *Genome Biol* 6, R9.

39. Ramaley L. and Herrera L.C. (**2008**) Software for the calculation of isotope patterns in tandem mass spectrometry. *Rapid Commun Mass Spectrom* 22, 2707-2714.

40. Zelter A. *et al.* (**2010**) Isotope signatures allow identification of chemically cross-linked peptides by mass spectrometry: a novel method to determine interresidue distances in protein structures through cross-linking. *J Proteome Res* 9, 3583-3589.

41. Hussong R. and Hildebrandt A. (**2010**) Signal processing in proteomics. *Methods Mol Biol* 604, 145-161.

42. Sykes M.T. and Williamson J.R. (**2008**) Envelope: interactive software for modeling and fitting complex isotope distributions. *BMC Bioinformatics* 9, 446.

43. Hoopmann M.R., Finney G.L. and MacCoss M.J. (**2007**) High-speed data reduction, feature detection, and MS/MS spectrum quality assessment of shotgun proteomics data sets using high-resolution mass spectrometry. *Analytical chemistry* 79, 5620-5632.

44. Senko M.W., Beu S.C. and McLafferty F.W. (**1995**) Determination of monoisotopic masses and ion populations for large biomolecules from resolved isotopic distributions. *J. Am. Soc. Mass. Spectrom.* 6, 229-233.

45. Yu W., He Z., Liu J. and Zhao H. (**2008**) Improving mass spectrometry peak detection using multiple peak alignment results. *J Proteome Res* 7, 123-129.

46. Wu S.L., Kim J., Hancock W.S. and Karger B. (**2005**) Extended Range Proteomic Analysis (ERPA): a new and sensitive LC-MS platform for high sequence coverage of complex proteins with extensive post-translational modifications-comprehensive analysis of beta-casein and

epidermal growth factor receptor (EGFR). *J Proteome Res* 4, 1155-1170.

47. Melcher K. (**2004**) New chemical crosslinking methods for the identification of transient protein-protein interactions with multiprotein complexes. *Curr Protein Pept Sci* 5, 287-296.

48. Ding W. *et al.* (**2004**) The synthesis, distribution, and anti-hepatic cancer activity of YSL. *Bioorg Med Chem* 12, 4989-4994.

49. Prieto M., Mayor S., Rodriguez K., Lloyd-Williams P. and Giralt E. (**2007**) Racemization in suzuki couplings: a quantitative study using 4-hydroxyphenylglycine and tyrosine derivatives as probe molecules. *J Org Chem* 72, 1047-1050.

50. McCubbin J.A., Maddess M.L. and Lautens M. (**2006**) Total synthesis of cryptophycin analogues via a scaffold approach. *Org Lett* 8, 2993-2996.

51. Beatty K.E., Xie F., Wang Q. and Tirrell D.A. (**2005**) Selective dye-labeling of newly synthesized proteins in bacterial cells. *J Am Chem Soc* 127, 14150-14151.

52. Du P., Kibbe W.A. and Lin S.M. (**2006**) Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. *Bioinformatics (Oxford, England)* 22, 2059-2065.

53. Palaniappan K.K. *et al.* (**2011**) Isotopic Signature Transfer and Mass Pattern Prediction (IsoStamp): An Enabling Technique for Chemically-Directed Proteomics. *ACS chemical biology.*

# Chapter 4: Mapping Yeast *N*-glycosites with Isotopically Recoded Glycans

## 4.1 Introduction

Determining the functions of post-translational modifications (PTMs) is a challenge in post-genomic biology. Locating modified sites is often an essential first step toward ascertaining the biological roles of PTMs, and liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) has emerged as a choice technology for this purpose. For example, LC-MS/MS has been successfully utilized in recent proteomic surveys of phosphorylation,[1,2] acetylation,[3,4] ubiquitination,[5,6] and glycosylation[7-10] sites. Modifications such as these can profoundly impact a protein's molecular biology. Unambiguous assignment of PTM locations using existing LC-MS/MS methodology is sometimes difficult to achieve, particularly in cases involving low protein abundance or low site occupancy. Adding to the technical challenge of site-mapping, PTMs such as glycosylation are known to reduce ionization efficiency and, therefore, detectability via mass spectrometry.[11-13] In highly complex proteomic samples, low-intensity ions generated from glycopeptides may be overlooked due to instrument-dependent limitations on the rate at which ions can be selected for fragmentation during data-dependent acquisition.[14] In this chapter, I describe a directed proteomic approach in which LC-MS/MS analysis is focused specifically on peptides bearing *N*-glycosylation, a common PTM found in all major phylogenetic branches of life (Figure 4-1). In contrast to data-dependent methods that select a subset of relatively intense ions for fragmentation, this approach is specifically designed to provide intensity-independent fragmentation priority to peptides most likely to bear *N*-glycans and boost confidence in correct PTM site assignment.

In eukaryotes, *N*-glycosylation is a PTM frequently found on proteins that are translated in the endoplasmic reticulum (ER). Structurally conserved *N*-glycan precursors are synthesized via the dolichol pathway and transferred onto nascent proteins; the glycosylated Asn residues are typically found within a subset of the polypeptide's N-X-S/T motifs where X denotes a non-proline residue.[15,16] Biologically, *N*-glycans contribute to the folding, trafficking, and thermodynamic stability of proteins exposed to ER, vacuolar, and Golgi lumens as well as those destined for exposure to the extracellular milieu.[17,18] The *N*-glycan precursors are enzymatically edited to yield a plethora of mature glycoforms with compositions largely dependent on cell type and protein localization. Sometimes, specific *N*-glycan structures are required for protein function, while in other cases, *N*-glycans primarily contribute to protein stability and solubility.[19] As a class, the *N*-glycosylation modification is biologically essential; chemical or genetic disruption of *N*-glycan biosynthesis is lethal and aberrant *N*-glycosylation is associated with several human disease states.[20,21]

Many existing LC-MS/MS approaches for mapping *N*-glycosites depend on enzymatic removal of the entire *N*-glycan following stringent sample enrichment to remove non-glycosylated peptides prior to analysis.[9,10] These methods rely on detection of a 0.98 Da mass increase

Mark Breidenbach and Austin Pitcher contributed to the work presented in this chapter.
Portions of the work described in this chapter have been reported in a separate publication.[52]

resulting from the enzymatic deamination of glycosylated Asn residues by Peptide:N-glycosidase F (PNGase F). Enzymatic deamination is often performed in the presence of $^{18}O$-labeled water, imparting a 2.98 Da mass shift to the peptide to increase confidence in site assignment.[7] Unfortunately, complete removal of *N*-glycans with PNGase F can lead to instances of incorrectly mapped glycosites. During the course of PNGase F treatment, spontaneous deamination of non-glycosylated Asn residues and other instances of PNGase F-independent incorporation of $^{18}O$ can potentially yield false positives.[22-24] This drawback has led to the development of alternative strategies utilizing partial rather than total removal of *N*-glycans; for instance, treating samples with the enzyme endoglycosidase H (Endo H) preserves a single core GlcNAc residue, leaving direct evidence of *N*-glycosylation.[24,25] Unfortunately, the presence of a single sugar residue on peptides can considerably suppress ionization efficiency,[8,13] potentially biasing data-dependent LC-MS/MS data acquisition against glycopeptide ions. Despite this limitation, detection of the retained glycan by LC-MS/MS provides unequivocal evidence for glycosite assignment instead of indirect evidence that the *N*-glycan modification once existed at a given site. Therefore, LC-MS/MS methods that could select low-abundance ions for fragmentation could greatly benefit glycosite mapping.

In Chapter 3, I described a strategy for addressing the challenge of identifying low-abundance species in complex mixtures based on a directed LC-MS/MS approach (illustrated in Figure 3-8).[26] Termed isotopic signature transfer and mass pattern prediction (IsoStamp), the technique exploited the perturbing effects of a dibrominated chemical tag on a peptide's isotopic envelope. Once covalently modified with the tag, dibrominated peptides could be readily detected with high sensitivity and fidelity in high-resolution LC-MS data using a novel computational pattern-searching algorithm. Pattern identification allowed directed proteomic analyses of labeled peptides in complex biological samples using an inclusion list to trigger fragmentation.

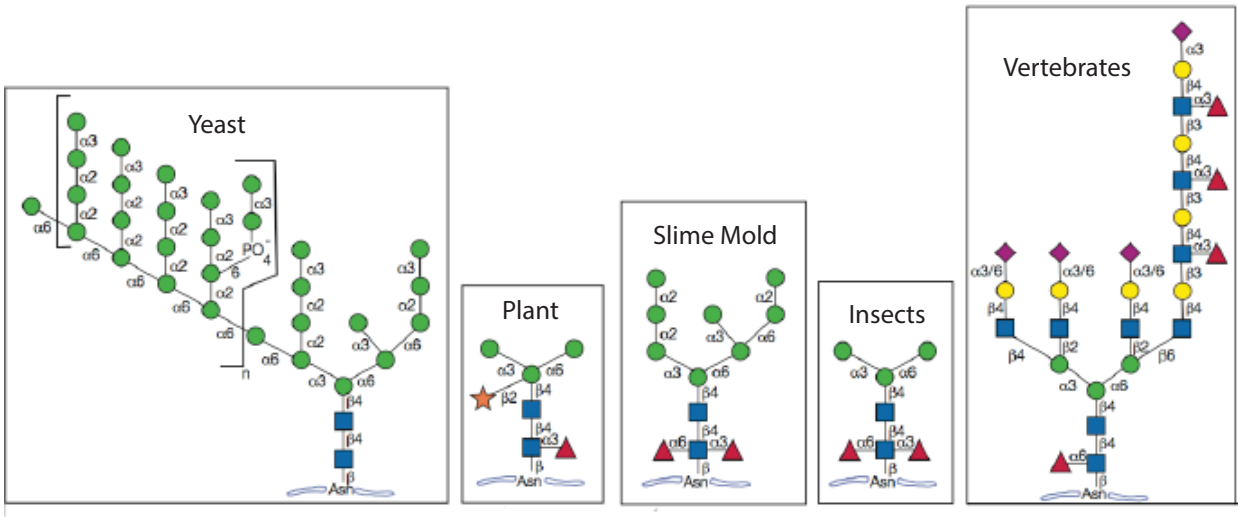



***Figure 4-1. N-Glycan structural diversity among different taxa.***

Although *N*-glycans can display great structural diversity, with extensive branching and linkages, at their core is a structurally conserved disaccharide (GlcNAc-β1,4-GlcNAc) that is maintained across many organisms. Adapted from Essentials of Glycobiology, 2nd ed.[18]

While the isotopic signature of a halogenated tag can effectively highlight labeled peptides within a complex LC-MS dataset, its utility is restricted to those situations in which the desired subset of peptides can be chemospecifically modified. To address this limitation, we demonstrate that it is possible to impart a similar perturbation to a peptide's isotopic envelope by metabolically embedding a dibromide-like isotopic signature directly into glycans. We mimic the dibromide isotopic signature with a stoichiometrically defined mixture of GlcNAc isotopologs, referred to as a GlcNAc isomix. The isomix is metabolically installed into structurally conserved *N*-glycan core positions, marking them with a uniquely identifiable isotopic signature. We employed the technique to map occupied *N*-glycosites on proteins from whole *Saccharomyces cerevisiae* lysates. Via preferential fragmentation of isotopically recoded glycopeptides, we identified numerous *N*-glycosites within the yeast proteome, nearly doubling the number that was previously known. The isomix method offers an enhanced level of confidence for mapping glycosylation sites that was not previously available to LC-MS/MS analyses because of the unique isotopic envelope of an isomix-containing peptide. Here, we showcase the utility of isotopic recoding by surveying metabolically labeled *N*-glycans in yeast, but we believe the technology is extendable to other PTMs and organisms through a variety of labeling strategies.

## 4.2 Results and Discussion

### *4.2.1 Metabolic recoding of N-glycan cores*

Due to the relative ease with which certain *S. cerev*isiae biosynthetic pathways can be manipulated, we chose to metabolically install an unnatural, dibromide-like isotopic signature into glycans. Universal isotopic recoding of *N*-glycans required metabolic replacement of a conserved sugar residue found in all such structures with a mixture of isotopologs. All eukaryotic *N*-glycans possess a conserved GlcNAc-β1,4-GlcNAc disaccharide at the peptide-proximal position, thus a GlcNAc isomix can potentially label all *N*-glycans in the cell's glycoproteome. However, to retain the isotopic ratio during metabolism, the GlcNAc isomix must be converted without isotopic dilution to the key metabolic intermediate uridine diphosphate-GlcNAc (UDP-GlcNAc), the donor nucleotide-sugar used in construction of *N*-glycan cores. Specifically, cytosolic UDP-GlcNAc serves as the donor substrate for Alg7p in the biosynthesis of GlcNAc-diphosphodolichol; the GlcNAc residue in this precursor is ultimately covalently linked to Asn in the process of *N*-glycosylation.[27] Thus, control over a cell's uridine diphosphate-GlcNAc (UDP-GlcNAc) pool is critical for execution of the isomix method. We recently generated a *S. cerevisiae* strain (*gna1*Δ) that depends on an engineered salvage pathway for procuring precursors of UDP-GlcNAc.[28] This strain lacked the ability to perform *de novo* UDP-GlcNAc biosynthesis and instead generates UDP-GlcNAc exclusively by salvaging GlcNAc added to the culture media. In previous work, we exploited this yeast strain to achieve high-efficiency replacement of GlcNAc residues with unnatural GlcNAc analogs, which are alternative substrates for the engineered salvage pathway.[28]

In this work, we supplemented cultured *gna1*Δ yeast with a GlcNAc isomix designed to emulate the isotopic signature of two bromine atoms. As shown in Figure 4-2A, the dibromide pattern is a symmetrical triplet, with major peaks at M, M + 2 and M + 4 at a relative intensity of 1:2:1, due to the relative abundances of the $^{79}Br_2$, $^{79}Br^{81}Br$, and $^{81}Br_2$ isotopic pairings. To

replicate this pattern, a three-part isomix consisting of *N*-acetyl-D-glucosamine (**4.1**), *N*-[1,2-$^{13}C_2$] acetyl-D-glucosamine (**4.2**), and *N*-[1,2-$^{13}C_2$]acetyl-D-[1-$^{13}$C;$^{15}$N]glucosamine (**4.3**), mixed in a 1:2:1 molar ratio (Figure 4-2B), was prepared and added to the *gna1*Δ yeast culture medium (Figure 4-2C). The isomix is internalized via the GlcNAc-specific transporter Ngt1 borrowed from *C. albicans*, and cytosolic GlcNAc is subsequently phosphorylated by the heterologously expressed

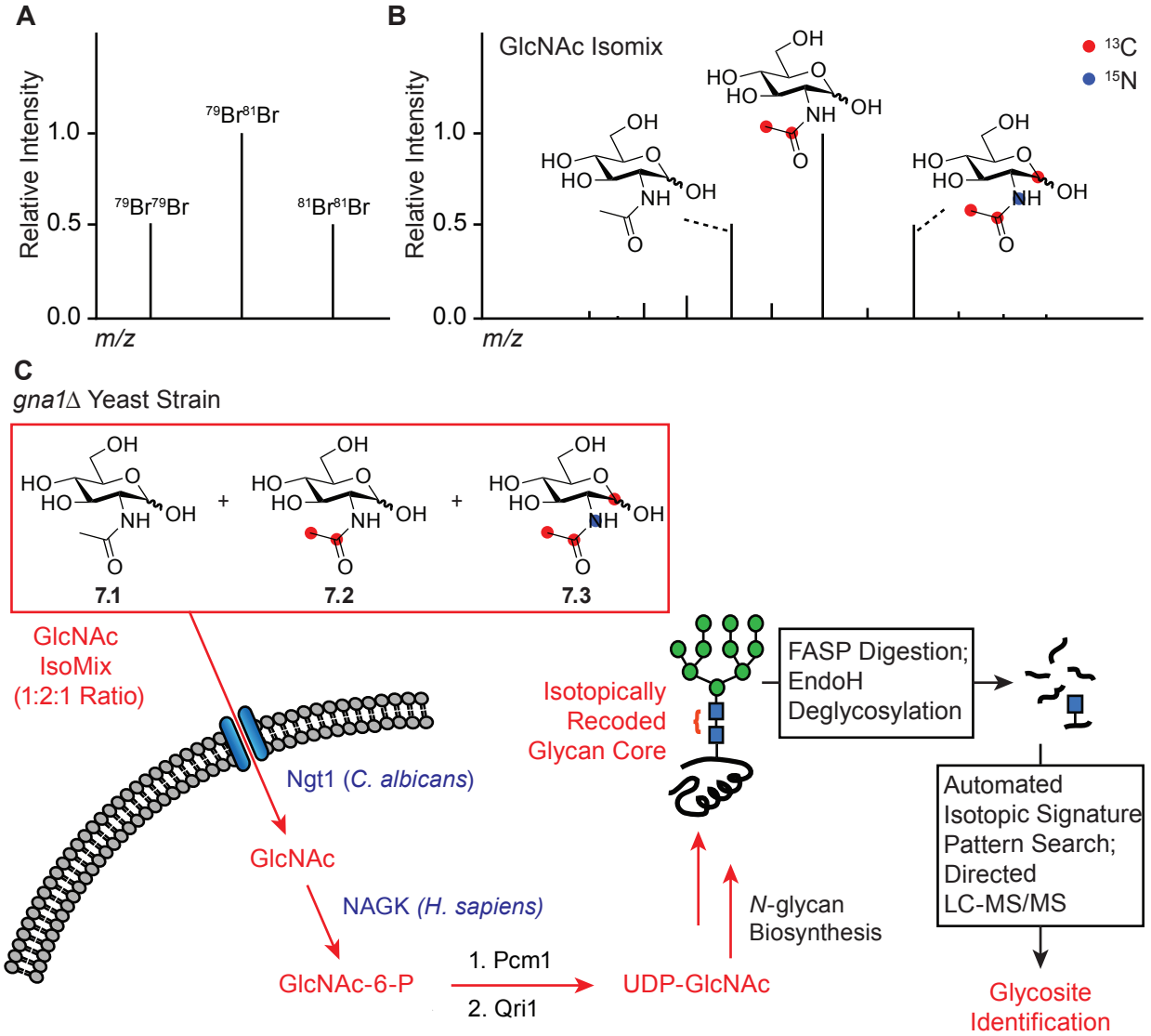



***Figure 4-2. Metabolic incorporation of a GlcNAc isomix into yeast N-glycans.***

(A) The dibromide triplet pattern, with a 1:2:1 relative peak intensity distribution, results from the natural abundances of $^{79}$Br and $^{81}$Br isotope pairings. (B) A three-component GlcNAc isomix mimics the 1:2:1 peak intensity distribution of dibromide by adjusting the concentration of each synthetically made isotopolog. (C) The GlcNAc isomix enters the *gna1*Δ yeast hexosamine biosynthetic pathway via a heterologous salvage pathway.[29] The isomix signature is subsequently embedded into UDP-GlcNAc and any glycoconjugates that utilize UDP-GlcNAc in their construction, including the structurally conserved cores of *N*-glycans. Following cell lysis, proteolysis, partial enrichment of *N*-glycopeptides, and partial deglycosylation with Endo H, the distinctive isotopic signatures of *N*-glycopeptides are detected computationally using pattern-matching software.[26] Masses of putative *N*-glycopeptide ions are granted fragmentation priority in subsequent LC-MS/MS analyses for *N*-glycosite identification. The *N*-glycan precursor illustrated here is composed of 2 core GlcNAc residues (blue squares) and 10 mannose residues (green circles).

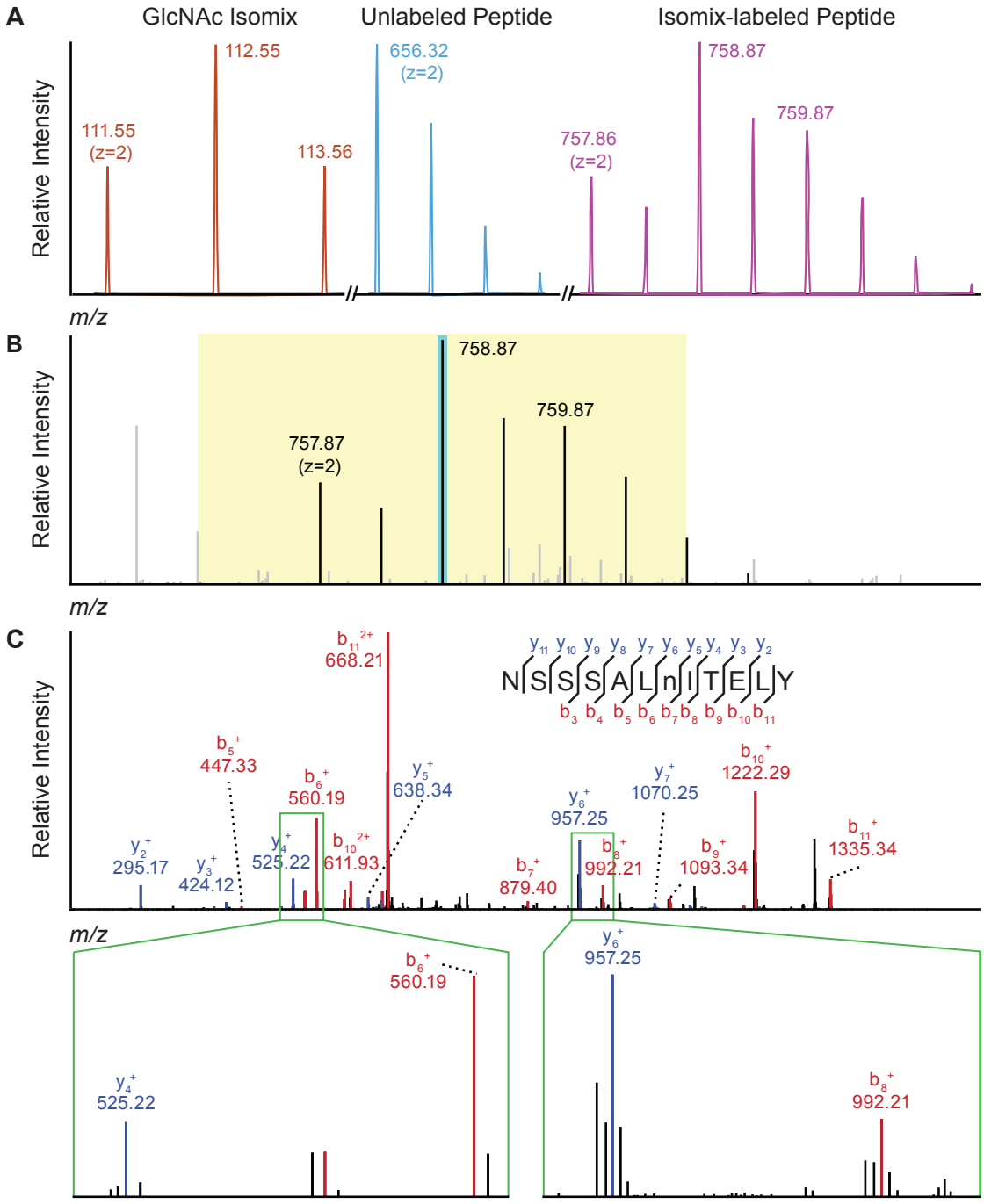



***Figure 4-3. The perturbing effect of a GlcNAc isomix on a peptide's isotopic envelope.***

(A) Simulated isotopic envelopes (z = 2) for the GlcNAc isomix (red trace), an unlabeled peptide from the glycoprotein Ygp1 (NSSSALNITELY, blue trace), and the same peptide labeled with the GlcNAc isomix (violet trace). The isotopically recoded peptide has a visually distinctive distribution of peak intensities. (B) In experimental LC-MS data, the isotopic envelope of the precursor ion corresponding to the NSSSALnITELY glycopeptide is shown. The modified Asn residue is in lowercase. Highlighted in teal is the precursor ion that was selected from the inclusion list and the 4 Da isolation window used for fragmentation is shown in yellow. (C) The CID fragmentation spectra and the peptide assignment for the 758.87 ion (lowercase n refers to the *N*-glycosite). Fragment ions that lack the GlcNAc isomix (such as the $y_4^+$ and $b_6^+$ fragment ions) have narrow isotopic envelopes, while fragments including the GlcNAc isomix (such as the $y_6^+$ and $b_8^+$ fragment ions) show a perturbed isotopic envelope characteristic of the isomix signature.

human kinase NAGK. The resulting GlcNAc-6-phosphate isomix is converted to a UDP-GlcNAc isomix and subsequently installed into *N*-glycan cores by endogenous yeast machinery.

*4.2.2 Directed proteomic analysis of N-glycosylated peptides*

Samples for LC-MS/MS analysis were generated from lysates of *gna1*Δ *S. cerevisiae* cultures grown to both mid-log and stationary phases in chemically defined, minimal medium. Tryptic and chymotryptic peptides were prepared from the lysates using a modified version of Filter Aided Sample Preparation (FASP).[10,29] During this process, mannose-containing glycopeptides were partially enriched by binding to the lectin concanavalin A (Con A) and *N*-glycans were truncated to a single GlcNAc residue with the glycosidase Endo H. To achieve sufficient resolution for isotopic envelope pattern matching, peptides were analyzed on a Thermo-Finnigan LTQ Orbitrap XL mass spectrometer.

Recoded peptide envelopes bear distinctive peak intensity distributions, as illustrated in the simulated isotopic envelopes shown in Figure 4-3A. The envelope of an isomix-labeled glycopeptide reflects the peptide's intrinsic envelope convolved with that of the GlcNAc isomix. The unique pattern resulting from this convolution serves as the basis for detecting putative glycopeptides in complex samples. Importantly, experimentally observed isomix-containing glycopeptide envelopes, as exemplified in Figure 4-3B, matched our expectations based on simulations (Figure 4-2A). This allowed us to computationally search (using the IsoStamp algorithm)[26] for the characteristic isomix signature in yeast peptides subjected to LC-MS analysis. Typically, we detected 4,000-5,000 candidate glycopeptide ions (in charge states z=1 to z=5) per sample. However, we did not expect all candidate precursors targeted for fragmentation to yield unique, high-confidence glycosites. In some cases, high sample complexity contributed to co-eluting ions of similar masses and identical charge states, generating false positive identifications during the automated pattern search routine.[26] Additionally, the precursor inclusion list inherently contains redundancy. For example, several legitimate, isomix-bearing glycopeptides were observed in multiple charge states between z=1 and z=5. Also, we frequently observed more than one peptide covering the same glycosite. The combination of these factors accounts for the discrepancy between the putative glycopeptide precursor ions and the high-confidence glycosites we report here. Regardless, the *m/z* values and retention times of putative glycopeptide ions bearing the isomix signature were used to construct a time-resolved inclusion list for directed fragmentation. Notably, even with the inclusion of a lectin-based glycopeptide enrichment step during sample preparation, the overwhelming majority of high-intensity ions in our samples appeared to be unglycosylated, as indicated by visual and computational inspection of their isotopic envelopes.

Our directed proteomics approach (see Figure 3-8) required duplicate, back-to-back injections for directed fragmentation of glycopeptides. The first injection was used exclusively to search for GlcNAc isomix signatures in LC-MS data to identify likely glycopeptide ions and inventory them into an inclusion list. In the subsequent injection, candidate glycopeptide ions were subjected to fragmentation by collision induced dissociation (CID)[30] only if the ion abundance exceeded a defined threshold and the ion appeared in the correct retention time window. Importantly, the Asn-GlcNAc linkage is resistant to standard CID conditions used to fragment the

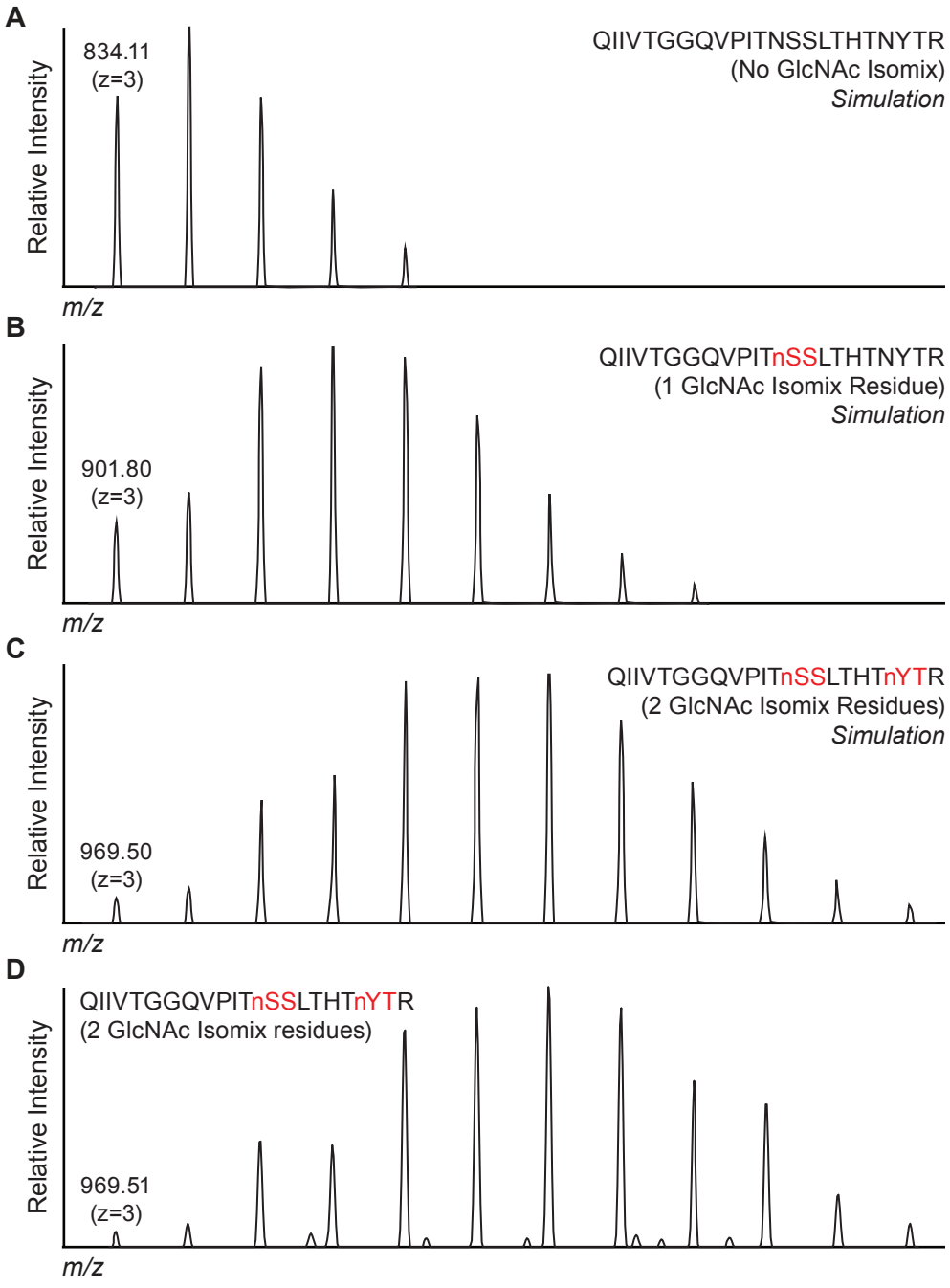



***Figure 4-4. The perturbing effect of two GlcNAc isomix residues on a peptide's isotopic envelope.***

Simulated isotopic envelopes for the Ygp1 peptide (QIIVTGGQVPITNSSLTHTNYTR) which contain (A) zero, (B) one, or (C) two GlcNAc isomix residues (modified sequons are highlighted in red). (D) In experimental LC-MS data, we observed a doubly-glycosylated form of the Ygp1 peptide with an isotopic envelope that closely matched the simulation.

peptide backbone.[31]

Ions bearing the isomix signature were subjected to CID fragmentation in order to determine peptide identity and verify *N*-glycosylation. Ions selected for fragmentation were isolated in either a narrow (2 Da) or broad (4 Da) isolation window (IW). While a broad IW covers most peaks in an isomix-labeled glycopeptide's envelope (Figure 4-3B), it has the disadvantage of potentially including unrelated ions in the fragmentation spectrum. Conversely, the use of narrow IWs could yield higher quality fragmentation spectra because fewer unrelated ions will be isolated. However, in practice, the use of a 4 Da IW resulted in the highest number of MS/MS identifications despite the potential decrease in the precursor ion fraction isolated.[32] Indeed, we observed no detrimental effects on peptide assignments in datasets collected using a 4 Da verses a 2 Da IW. Instead, the use of a broad IW on ions bearing the isomix signature often provided additional information in the fragmentation spectra that was critical to resolving glycosite positional ambiguity. The broad IW was particularly useful for glycosite assignment in the case of peptides containing more than one Asn residue. As illustrated in Figures 4-3B and 4-3C, when the full isotopic envelope of a candidate glycopeptide was isolated for fragmentation using a broad IW, we preserved the isomix signature in all peptide fragments bearing the GlcNAc isomix modification. In contrast, unglycosylated fragments yielded unperturbed mass envelopes. This information was used to resolve cases of ambiguous glycosite assignments that resulted from automated database search engines. In some cases, doubly glycosylated peptides were observed; the presence of two *N*-glycosites on one peptide introduced a doubly convolved isomix signature resulting in a widened isotopic envelope with a distinctive "stair-step" pattern in peak intensities (Figure 4-4).

### *4.2.3 Isomix-directed fragmentation increases confidence in glycosite identification*

The unified list of high-confidence *N*-glycosites in the yeast proteome (summarized in Table 4-1) reflects a combination of results from multiple experiments, conducted with trypsinized and chymotrypsinized samples, obtained from both stationary and mid-log phase cultures. A complete list of the observed tryptic and chymotryptic peptides containing these glycosites with their corresponding statistical indicators of quality can be found in Appendex B.2. We detected a total of 133 *N*-glycosites, 12 of which were previously reported in the Uniprot Knowledgebase.[33] Taken collectively with glycosite information from the Uniprot database (as of January 2012), there are now 196 experimentally mapped *N*-glycosites spanning 72 proteins in S. cerevisiae; 121 novel sites distributed over 52 proteins are reported here. Importantly, 50% of the proteins in Table 4-1, highlighted in gray, have been biochemically validated as *N*-glycosylated. Proteins were considered biochemically validated if they had previously mapped *N*-glycosites or if there was experimental observation of a gel shift following enzymatic deglycosylation.[28,33-35] The high correlation between our proteomic observations and the plethora of biochemical validation data for yeast serves as an excellent indicator that the isomix technique yielded a reliable list of *N*-glycoproteins. Moreover, our results suggest the gna1Δ S. cerevisiae likely share similar *N*-glycosite occupancy patterns with the BY4743 strain[36] from which the gna1Δ haploid was derived.

For comparative purposes, we also subjected our samples to traditional intensity-driven

***Table 4-1. Occupied high confidence N-glycosites detected in the yeast proteome.***

Peptides covering these glycosites are detailed in Appendix B.2. Biochemically validated N-glycoproteins are highlighted in gray.[28, 33-35] Proteins with mapped or partially mapped N-glycosites are indicated in bold text.[3]

| Systematic Name | Uniprot ID | Standard Name | Occupied *N*-glycosites |
|---|---|---|---|
| YBR023C | P29465 | CHS3 | Asn 323 |
| YBR074W | P38244 | | Asn 121 |
| YBR078W | P38248 | ECM33 | Asn 209, 267, 279, 328 |
| YBR092C | P24031 | PHO3 | Asn 162, 315, 356, 390* |
| YBR093C | P00635 | PHO5 | Asn 315, 356,* 390* |
| YBR139W | P38109 | | Asn 339 |
| YBR162C | P38288 | TOS1 | Asn 236 |
| YBR171W | P33754 | SEC66 | Asn 12 |
| YBR229C | P38138 | ROT2 | Asn 173, 907 |
| YBR286W | P37302 | APE3 | Asn 85, 96, 115, 150, 162, 371, 427 |
| YCL043C | P17967 | PDI1 | Asn 117, 155, 425 |
| YCL045C | P25574 | EMC1 | Asn 106, 420 |
| YCR011C | P25371 | ADP1 | Asn 165 |
| YDL095W | P33775 | PMT1 | Asn 390 |
| **YDR055W** | Q12355 | PST1 | Asn **210**, **242** |
| YDR434W | Q04080 | GPI17 | Asn 299 |
| YEL001C | P40006 | IRC22 | Asn 93, 121 |
| YEL002C | P33767 | WBP1 | Asn 60, 332 |
| YEL040W | P32623 | UTR2 | Asn 261 |
| YEL060C | P09232 | PRB1 | Asn 594 |
| YFR018C | P43599 | | Asn 179 |
| YGL022W | P39007 | STT3 | Asn 539 |
| YGR014W | P32334 | MSB2 | Asn 1088 |
| YGR189C | P53301 | CRH1 | Asn 177, 201 |
| YGR282C | P15703 | BGL2 | Asn 202 |
| YHR028C | P18962 | DAP2 | Asn 63, 139 |
| YHR101C | P38813 | BIG1 | Asn 144 |
| **YHR188C** | P38875 | GPI16 | Asn **184** |
| YHR202W | P38887 | | Asn 305 |
| YIL015W | P12630 | BAR1 | Asn 268, 366, 398 |
| YJL002C | P41543 | OST1 | Asn 99, 400 |
| YJL171C | P46992 | | Asn 51, 99, 122, 174, 219, 249, 267, 300 |
| YJL172W | P27614 | CPS1 | Asn 176, 228 |
| YJL192C | P39543 | SOP4 | Asn 35 |
| YKL039W | P32857 | PTM1 | Asn 132 |

| Systematic Name | Uniprot ID | Standard Name | Occupied *N*-glycosites |
|---|---|---|---|
| YKL046C | P36091 | DCW1 | Asn 203, 240 |
| YLR066W | Q12133 | SPC3 | Asn 173 |
| YLR084C | Q12465 | RAX2 | Asn 333, 524, 640 |
| YLR413W | Q06689 | | Asn 194, 299 |
| YLR450W | P12684 | HMG2 | Asn 150 |
| YML052W | P54003 | SUR7 | Asn 47 |
| YML130C | Q03103 | ERO1 | Asn 53, 468 |
| YMR006C | Q03674 | PLB2 | Asn 80, 193, 491 |
| YMR008C | P39105 | PLB1 | Asn 78, 123, 170, 388, 513, 541 |
| **YMR058W** | P38993 | FET3 | Asn **244**, 265, **292**, **359**, **381** |
| **YMR200W** | Q03691 | ROT1 | Asn **139** |
| YMR238W | Q05031 | DFG5 | Asn 245 |
| **YMR297W** | P00729 | PRC1 | Asn **124**, **479** |
| YMR307W | P22146 | GAS1 | Asn 40, 57, 95, 149, 165, 253, 409 |
| YNL160W | P38616 | YGP1 | Asn 58/61, 87, 94, 100, 106, 118, 172, 239, 286 |
| YNL327W | P42835 | EGT2 | Asn 103, 161 |
| YNR067C | P53753 | DSE4 | Asn 886 |
| YOL011W | Q08108 | PLB3 | Asn 56, 82, 547 |
| YOL030W | Q08193 | GAS5 | Asn 24, 60, 166, 299, 344 |
| YOL154W | Q12512 | ZPS1 | Asn 28, 57, 217 |
| YOR320C | Q12096 | GNT1 | Asn 425 |
| YPL123C | Q02933 | RNY1 | Asn 37 |
| **YPL154C** | P07267 | PEP4 | Asn **144**, **345** |

*Sequence ambiguity between PHO3 and PHO5.

data-dependent LC-MS/MS, in which the ten most intense ions in each full scan mass spectrum were selected for CID fragmentation. Significantly, the intensity-dependent analyses typically identified approximately 60% of the glycosites found via isomix-directed fragmentation. For instance, in the case of trypsinized lysate prepared from a stationary-phase culture, 52 unique glycosites spanning 25 proteins were identified using directed fragmentation. The same sample, subjected to data-dependent fragmentation, yielded only 30 *N*-glycosites in 16 proteins and contained extensive site overlap with the directed dataset. This difference illustrates the advantages of directed, intensity-independent approaches for proteomic analyses. Despite being outperformed by a directed proteomic analysis here, the widely used data-dependent fragmentation approach remains a powerful method, and its performance would undoubtedly benefit from more stringent glycopeptide sample enrichment and improved chromatographic resolution of peptides.

Glycoproteins identified in this study were categorized by the manually curated ontological annotations maintained by the *Saccharomyces* Genome Database.[37] An analysis of cellular

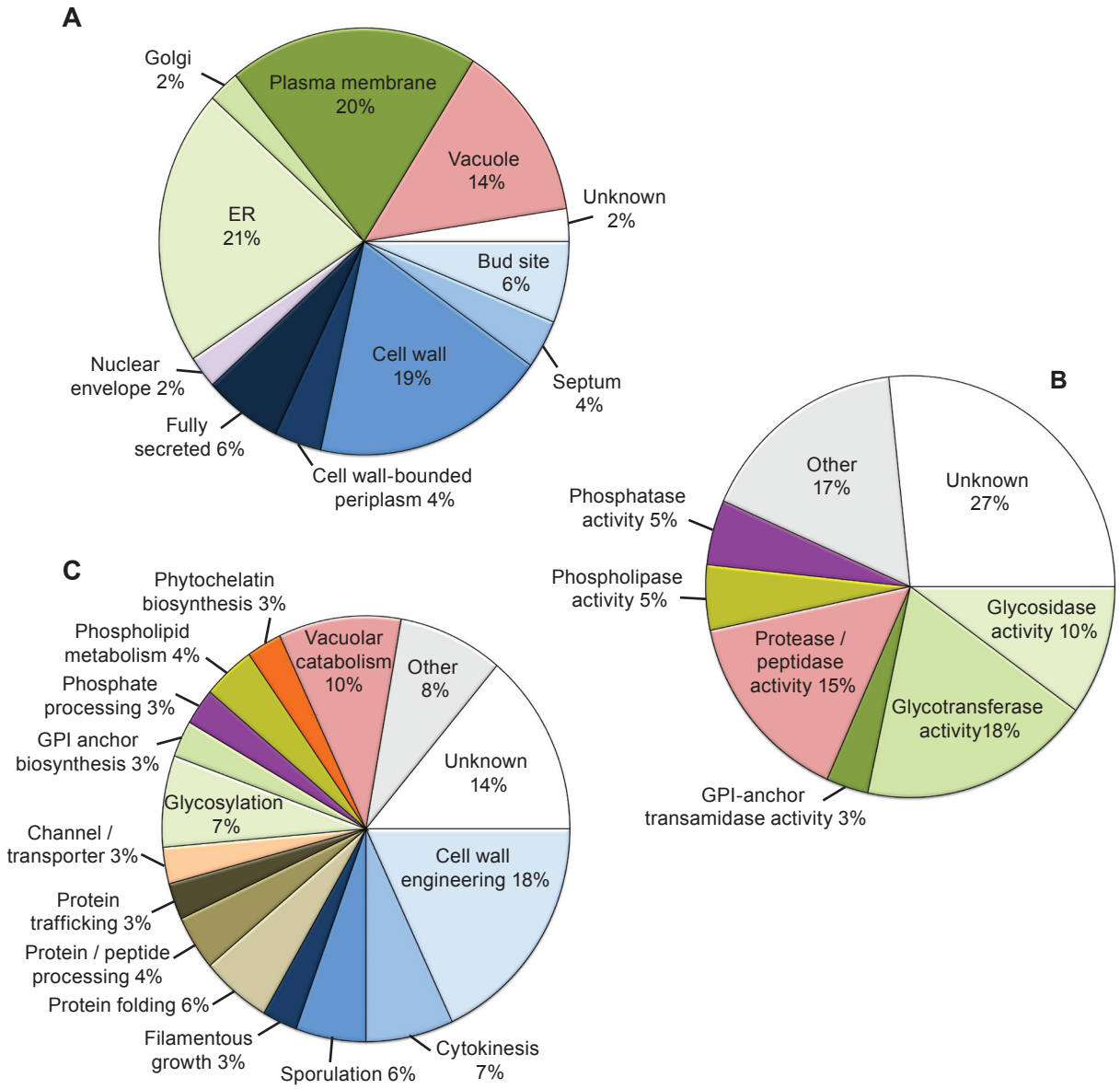



***Figure 4-5. Ontological analysis of high confidence N-glycoproteins.***

*N*-glycoproteins were categorized according to the manually curated ontological annotations including (A) cellular component, (B) molecular function, and (C) biological process maintained by the *Saccharomyces* Genome Database.[37]

localization (Figure 4-5A) reveals that the majority of the *N*-glycoproteins identified in Table 4-1 typically reside in the yeast ER, plasma membrane, vacuole, and cell wall. Several proteins localize to substructures within these organelles, especially those involved with cytokinesis and cellular budding. While many of the identified glycoproteins lack known molecular functions, our list was highly represented by peptidases, glycosidases, and glycosyltransferases (Figure 4-5B). With respect to biological function, highly represented categories include vacuolar proteases and proteins involved in cell wall construction, reshaping, and maintenance (Figure 4-5c). Several ER- and Golgi-resident proteins participating in the processes of *N*- and *O*-linked glycosylation were

also detected. Overall, the ontological distribution of glycoproteins reported here is consistent with our expectations for *N*-glycoproteins in yeast.

In addition to the ontological consistency for our list of *N*-glycoproteins, we detected no plausible cases in which the *N*-glycan modification was present outside the canonical N-X-S/T motif. While non-canonical *N*-glycosites have been reported in higher eukaryotes,[10,38] it is possible that the acceptor substrate promiscuity of the oligosaccharyl transferase (OST) complex, which cata-

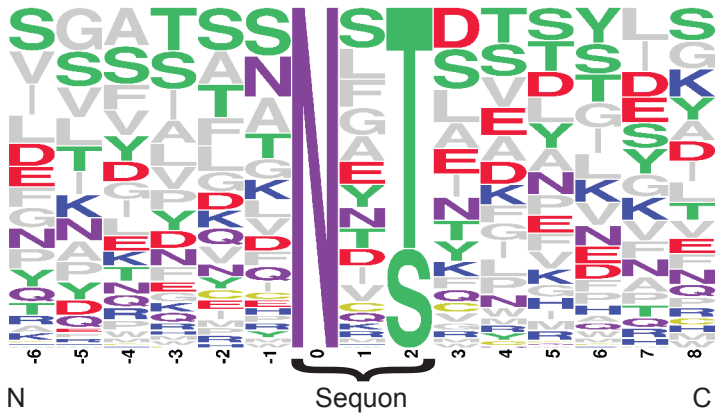



***Figure 4-6. Relative frequencies of residues surrounding yeast N-glycosites.***

Sequences of the 133 unique glycosylation sites detected by directed LC-MS/MS are aligned on the modified Asn residue. Relative heights of the surrounding amino acids are adjusted based on the frequencies of their occurrence.

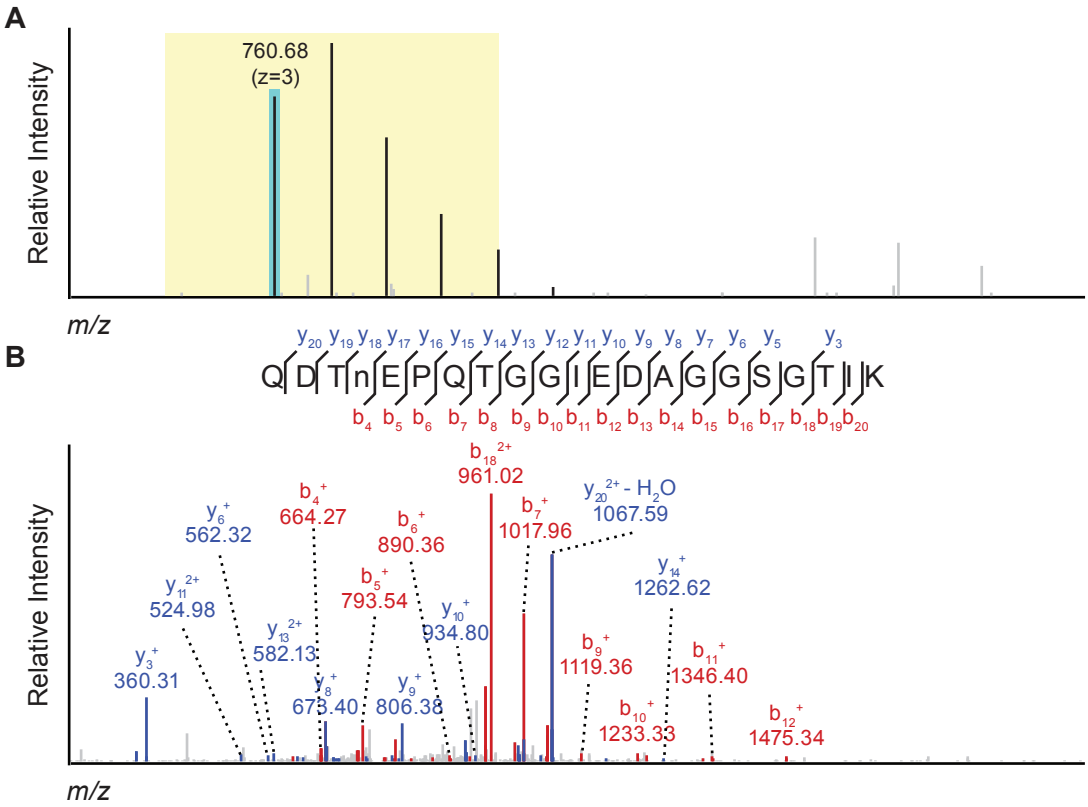



***Figure 4-7. Example of spectral misassignment from data-dependent analysis of an isomix-labeled sample.***

(A) The isotopic envelope of a precursor ion selected for data-dependent fragmentation lacks the characteristic GlcNAc isomix signature, in contrast to that illustrated in Figure 4-3. The envelope's monoisotopic peak is highlighted in teal and the 4 Da IW is shown in yellow. (B) The CID fragmentation spectrum and SEQUEST assignment to the indicated Akl1 peptide appear both statistically and visually reasonable (*XCorr* = 2.90, FDR < 5%). However, the lack of an isomix signature in the precursor and fragment ions serves as a basis for rejecting the assignment.

lyzes the *N*-glycosylation modification, differs amongst species. A plot of the relative residue frequencies for the 133 occupied glycosites mapped here revealed a strong, 71% versus 29%, preference for threonine over serine within the canonical motif (Figure 7-6), consistent with *N*-glycosites mapped in other eukaryotic proteins by non-proteomic methods.[39] Additionally, positions surrounding the motif are dominated by hydroxylated and small hydrophobic residues, consistent with expectations for solvent-exposed, loop regions of proteins. Charged residues, especially aspartate, are present in slightly higher frequency on the C-terminal side of the sequon, but not at the -2 position as observed in the case of bacterial *N*-glycosites.[40] Collectively, these data reveal a typical acceptor substrate sequence for the yeast OST.

While LC-MS/MS is currently amongst the best available methods for mapping PTMs such as glycosylation, high-confidence correlations between fragmentation spectra and peptide sequences is not always simple.[41] The utility of tandem MS analysis is ultimately limited by the quality of fragmentation spectra. Because fragmentation spectra are rarely perfect, statistical methods for increasing confidence in peptide assignment remains an active areas in research in the field of proteomics.[42-45] Advancements in instrumentation are also reducing the possibility of incorrect spectral assignment. Next-generation hardware designed to allow rapid, sensitive, and highly accurate analyses of both precursor and fragment ion masses will undoubtedly reduce false positives in proteomic hit lists.[46] Complementing high mass accuracy, isotopic recoding of PTMs provides an additional restraint for spectral assignment. As illustrated in Figure 4-3, precursor and fragment ions corresponding to recoded peptides bear a unique isotopic signature that could be used as a basis for accepting or rejecting a computationally-assigned spectral match. Indeed, several peptides in our data were computationally matched to peptides based on favorable statistical indicators, such as high SEQUEST *XCorr* values, but were ultimately rejected because they lacked the characteristic isomix signature in their isotopic envelopes. As illustrated in Figure 4-7, a peptide isolated for CID fragmentation via data-dependent selection was assigned to a GlcNAc-containing peptide from the cytosolic kinase Akl1 with high confidence. However, manual inspection of the precursor ion's isotopic envelope revealed no evidence of the isomix signature, casting doubt on the assignment. The lack of a standard N-X-S/T sequon in this peptide and the cytosolic location of Akl1 are also consistent with spectral misassignment. Thus, we believe isotopic recoding of PTMs can serve as a valuable tool for rejecting incorrect spectral assignments that would otherwise pass undetected into LC-MS/MS hit lists.

## 4.3 Conclusions

We have demonstrated that a metabolically embedded isotopic signature, designed to emulate the perturbing effects of a dibrominated chemical tag, can serve as the basis for directed proteomic analysis of a biologically important class of PTM. This approach has two major benefits. First, the ions selected for fragmentation need not be among the most intense in the sample. Second, recoded envelopes contain easily detectable cues indicating whether or not fragmentation spectra have been correctly assigned to a peptide, a feature that is particularly useful for verifying the site of a PTM. We used an isomix-based directed proteomics strategy to build a relatively small but high-confidence list of *N*-glycosites in yeast.

Systematically mapping *N*-glycosylation sites in *S. cerevisiae* provides a wealth of infor-

mation pertaining to native glycoprotein structure and topology. It is well known that the mere presence of a standard N-X-S/T sequon is not sufficient to guarantee glycan attachment to a nascent polypeptide. A number of extenuating factors, ranging from local protein structure to

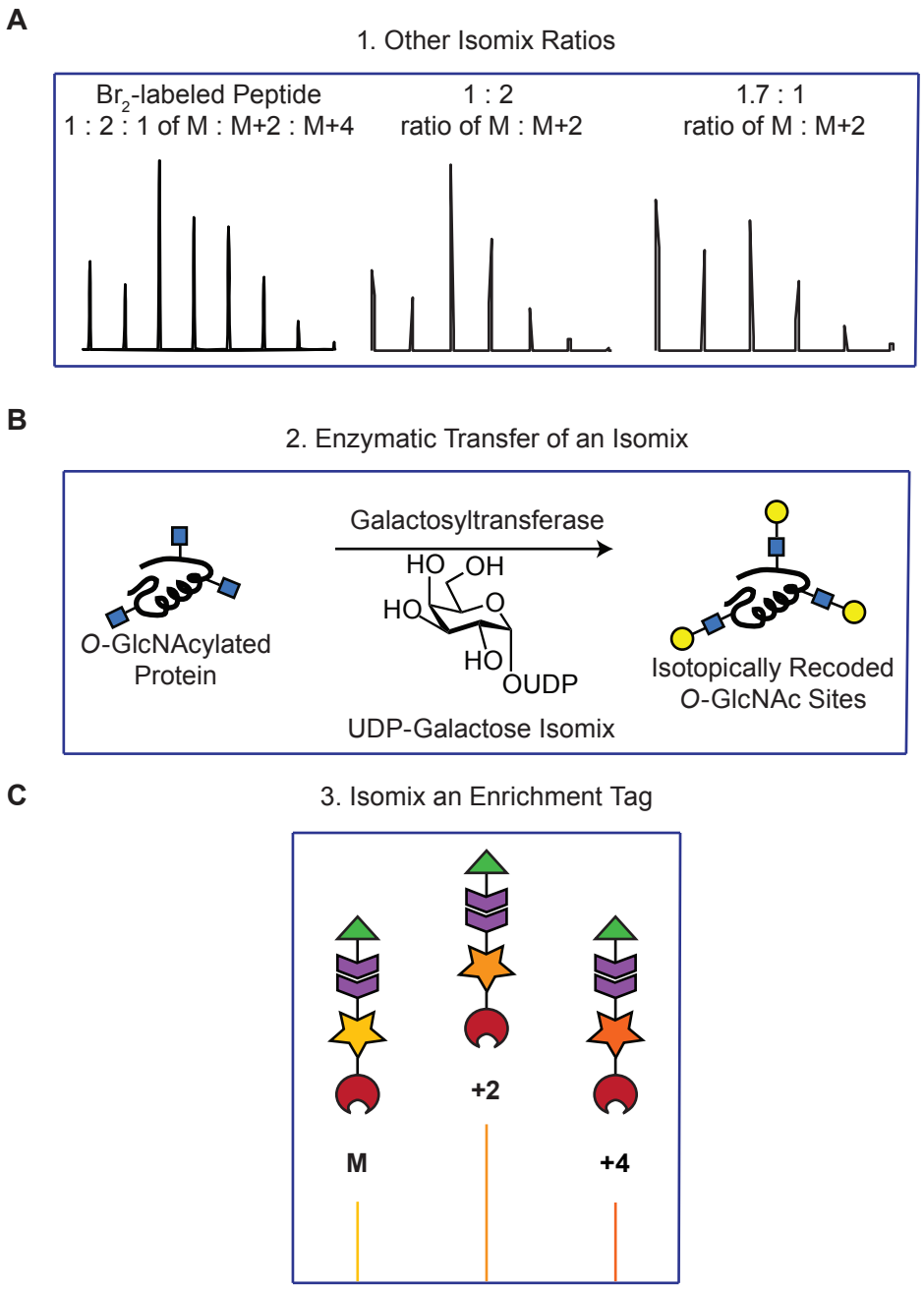



***Figure 4-8. Isomix-based projects ideas.***

(A) Exploration of new two-component GlcNAc isomix solutions for directed proteomics applications. (B) Identification of *O*-GlcNAcylated substrates through enzymatic tagging with the galactosyltransferase enzyme and a UDP-Galactose isomix, followed by directed proteomics. (C) Utilizing heavy-atom containing bioorthogonal enrichment tags, premixed at the appropriate ratio to mimic a dibromide pattern, can eliminate the need of bromine atoms for isotopically recoding chemically enriched peptides.

nutrient availability, ultimately determine which sequons are glycosylated. In the case of soluble proteins, fully or partially occupied glycosites are an excellent indicator of sequon surface accessibility. In the case of integral membrane proteins (IMPs), occupied glycosites confirm that a given Asn is ER-lumenal upon translation and can be used to confirm or reject IMP topological predictions. In our data, we detected 25 glycosylated IMPs, and in all cases the occupied glycosites confirm TMHMM topology predictions[47] of lumenal versus extracellular domains. Additionally, glycosite mapping is a prerequisite for quantitative, comparative analysis of glycan occupancy. While we do not attempt to quantify glycan occupancy in this study, a high-confidence glycosite map serves as the foundation for investigating how glycosite occupancy changes with cell state and growth conditions; such a study can be conducted using established quantitative MS approaches focused on newly verified glycosites and is a subject of interest for future investigations.

By using isotopically recoded glycans, we successfully placed fragmentation priority on glycopeptide ions regardless of their relative intensities. As a result, rigorous sample enrichment for *N*-glycopeptides was not essential using the isomix strategy. Nonetheless, we speculate that imperfect enrichment with ConA allowed for the presence of many high-abundance, non-glycopeptides in our sample, suppressing ionization and the successful detection and fragmentation of low-abundance, poorly ionizing glycopeptides within the yeast proteome. Isotopic recoding is not intended to replace or eliminate enrichment from proteomics workflows; rather, we believe it has the potential to increase the information obtained from a given sample. We expect that enhanced glycopeptide enrichment strategies in conjunction with cellular fractionation and improved chromatographic resolution of peptides will permit detection of additional glycosites; the *S. cerevisiae* proteome has been reported to contain at least 350 *N*-glycosylated proteins.[35]

The concept of isotopically recoding PTMs for directed LC-MS/MS analysis can be extended well beyond *N*-glycosite mapping (Figure 4-8). By utilizing endogenous or engineered salvage pathways, unnatural isotopic signatures could be metabolically installed into a variety of PTMs. The relative malleability of glycan biosynthetic pathways, along with an extensive assortment of commercially available stable monosaccharide isotopologs, makes glycan-based PTMs particularly well suited for isotopic recoding. Ultimately, this approach is limited by the elemental composition and metabolic origin of the PTM being studied; while a stable, three-component GlcNAc isomix can be readily prepared and incorporated by cells, a similar phosphate-based isomix is less feasible. Importantly, there is no requirement that the isotopic signature strictly mimic that of a dibrominated probe (Figure 4-8A); a variety of isotopic mixtures could impart sufficient perturbation to a peptide's isotopic envelope to allow for successful directed LC-MS/MS analyses. Initial efforts to explore this idea were performed using a two-component isomix to reduce the potential synthetic challenges associated with synthesizing multiple isotopologs. We first simulated the isotopic signatures of peptides labeled with different ratios of a two-component GlcNAc isomix (Figure 4-9). Based on these models and our understanding of how our pattern-searching algorithm can extract isotopic signatures from complex mass spectra (see Chapter 3 and Austin Pitcher's thesis for more information), we found that a 1:2 or a 1.75:1 peak intensity ratio of compounds **4.1**:**4.4** could work as well as the three-component isomix. Towards evaluating this theory, the two-component GlcNAc isomix solutions were prepared (Figure 4-10)

and preliminary directed LC-MS/MS analysis was promising (data not shown). However, further work in optimizing the pattern-searching algorithm for these new isotopic signatures are needed before proceeding further.

Importantly, the approach of isotopic recoding for directed proteomics is not limited to the genetically modified *S. cerevisiae* strain utilized in this study. Albeit more challenging, we believe that pathway engineering in higher eukaryotes, with the aim of gaining compositional control over specific nucleotide-sugar pools, is feasible. While metabolic introduction of an unnatural

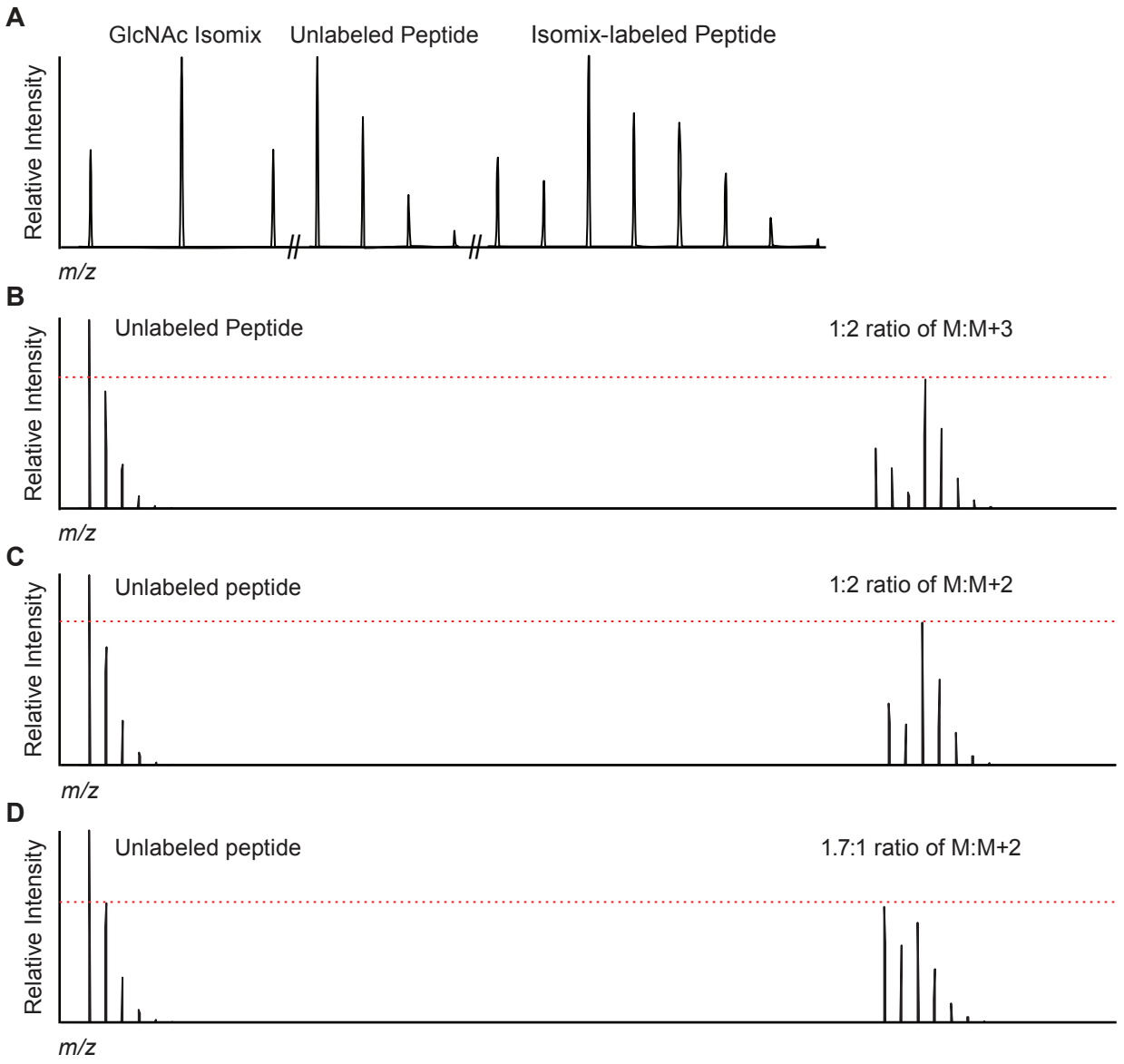



***Figure 4-9. Simulations of a two-component Isomix.***

(A) Reference simulation for an isotopically recoded peptide using the three-component GlcNAc isomix. (B-D) In comparison, isotopic envelopes of unlabeled or isomix-labeled peptides using a two-component isomix at different ratios were simulated. The dotted red line indicates the relative peak intensity of the most intense peak of the recoded isotopic envelope compared to the unlabeled peptide's isotopic envelope.

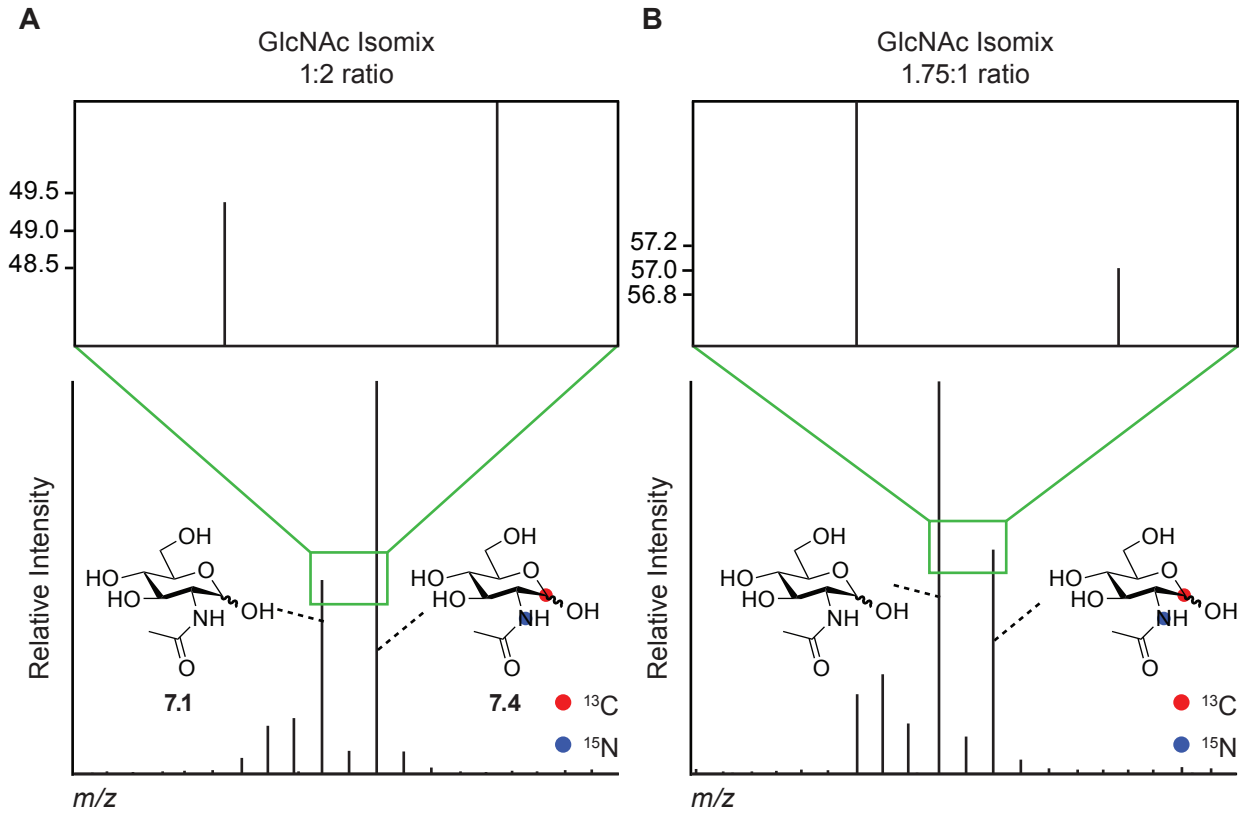



***Figure 4-10. Preparation of two-component GlcNAc isomix solutions.***

A two-component GlcNAc isomix was prepared using either a (A) 1:2 or (B) 1.75:1 ratio of *N*-acetyl-D-glucosamine and *N*-acetyl-D-[1-$^{13}$C;$^{15}$N]glucosamine, respectively.

isotopic signature was showcased here, it is important to emphasize that similar perturbations of isotopic envelopes can also be obtained via enzymatic or chemospecific labeling of specific PTMs. For example, we are investigating the possibility of *in vitro* enzymatic transfer of isomix-bearing probes to *O*-glycosylated peptides obtained from mammalian cells; this approach would obviate the need for metabolically embedding the isomix signature (Figure 4-8B). Additionally, chemospecific labeling can be performed using a variety of bioconjugation and bioorthogonal reactions[48] using a specific mixture of heavy-atom containing probes, such as enrichment tags (Figure 4-8C). Collectively, we believe the concept of isotopic recoding will be a powerful technology for those interested in profiling PTMs, including but not limited to *N*-glycans, using metabolic, enzymatic, and chemical labeling methods alike, substantially bolstering confidence in the data obtained from LC-MS/MS datasets.

## 4.4 Materials and Methods

### *4.4.1 General procedures and materials*

All chemical reagents were of analytical grade, obtained from commercial suppliers, and used without further purification unless otherwise noted. *N*-acetyl-D-glucosamine (**4.1**) was

purchased from Sigma. All reaction flasks were oven dried prior to use and moisture-sensitive reactions were performed under a $N_2$ atmosphere. Analytical thin layer chromatography (TLC) was performed with 60 Å silica gel plates (EMD) and visualized by staining with ceric ammonium molybdate, ninhydrin, or by absorbance of UV light at 254 nm. Flash chromatography was performed with 60 Å 230-400 mesh silica (Silicycle). Organic extracts were dried over $MgSO_4$, and solvent was removed by reduced pressure with a R-114 rotovapro (Buchi) equipped with a self-cleaning dry vacuum pump (Welch). If necessary, products were further dried by reduced pressure with an Edwards RV5 high vacuum pump. Lyophilization was performed on a FreeZone instrument (LABCONCO) equipped with an Edwards RV2 pump. Peptide sample volumes were reduced using a UVS400 Universal Vacuum System speed vacuum (Model SPD121P, ThermoFisher).

Water used in biological procedures was distilled and deionized using a Milli-Q Gradient ultrapure water purification system (Millipore). Dulbecco's phosphate-buffered saline (PBS) was obtained from HyClone Laboratories and fetal bovine serum (FBS) was obtained from Omega Scientific.

*4.4.2 Instrumentation*

**Absorbance spectra.** Absorbance readings were collected on a Molecular Devices SpectraMax 190 UV-Vis absorbance plate reader, a CARY 100 Bio UV-Visible Spectrophotometer, or a Thermo Scientific Nanodrop 2000 spectrophotometer.

**Electrospray ionization (ESI) MS.** Low-resolution ESI mass spectra were obtained using an LTQ-XL mass spectrometer.

**Directed LC-MS/MS on an LTQ Qribtrap XL.** Prior to analysis, samples were resuspended in 50 µL of water and 2.5-5 µL aliquots were subjected to reverse-phase liquid chromatography with an Agilent 1200 LC system connected in-line to a LTQ Orbitrap XL hybrid mass spectrometer. External mass calibration was performed prior to analysis. A binary solvent system consisting of buffer A (0.1% formic acid in water (v/v)) and buffer B (0.1% formic acid in acetonitrile (v/v)) was employed. The mass spectrometer was outfitted with a nanospray ionization source. The LC was performed using a 100-µm fritted capillary (New Objective) pre-column self-packed with 1 cm of 5 µm, 200 Å Magic C18AQ resin (Michrom Bioresources) followed by a 100-µm fused silica capillary (Polymicro Technologies) self-packed with 15 cm of 5 µm, 100 Å Magic C18AQ resin (Michrom Bioresources). After sample injection and a 20 min loading step in 2% buffer B, peptides were eluted using a gradient from 7% to 35% buffer B over 150 min, followed by a washing step in 99% buffer B for 20 min. A solvent split was used to maintain a flow rate of 400 nL $min^{-1}$ at the column tip.

Samples were subjected to an inclusion-list driven directed acquisition method. First, we collected only full-scan mass spectra in positive ion mode over the *m/z* scan range of 400 to 1,700 or 1,800 using the Orbitrap mass analyzer in profile mode at a resolution of 60,000 (at 400 *m/z*). Noise reduction and peak detection were performed using software developed in-house making use of a continuous wavelet transform.[26,49] The centroided mzXML files were then pattern

searched for recoded envelopes.[26] The output was an inclusion list that contained the *m/z* value (M+2 ion in the labeled peptide's isotopic envelope) and a retention time window (± 1.5 min, empirically determined) for each labeled peptide. The same sample, stored at 4 °C, was then reanalyzed with identical chromatographic conditions using an inclusion-list driven selection of precursor ions for fragmentation. In relatively rare instances we observed a chromatographic anomaly causing substantially shifted peptide elution profiles; in these cases the experiment was repeated. For each full-scan mass spectrum up to eight CID fragmentation events were performed in the linear ion trap. Dynamic exclusion and charge state screening were enabled to reject ions with an unknown or +1 charge state. An isolation window of 2 or 4 Da, a minimum threshold of 500 ion counts, and activation energy of 35 were used when triggering a fragmentation event.

*4.4.3 Chemical synthesis*

**Synthesis of GlcNAc isotopologs (4.2, 4.3).** The synthesis of heavy *N*-acetyl-D-glucosamine from D-glucosamine hydrochloride was performed in a single synthetic step as previously described.[50] To a solution of the D-glucosamine hydrochloride salt (150 mg, 0.6 mmol) in water, Dowex 200-400 mesh ($OH^-$) anion-exchange resin was added, and the pH was adjusted to 7.5. The resin was removed by filtration, yielding a D-glucosamine solution in 6 mL of water. $^{13}C_2$-sodium acetate (1.1 eq, Cambridge Isotope Laboratories) as a pre-dissolved solution in water and 2-ethoxy-1-ethoxycarbonyl-1,2-dihydroquinoline (1.1 eq) as a pre-dissolved solution in 10 mL ethanol were added to the D-glucosamine solution. The total reaction volume was brought to 40 mL by addition of ethanol. The reaction was covered in foil and stirred for 36 h at RT to afford compound **4.2**. This reaction was repeated using 1-$^{13}$C,$^{15}$*N*-D glucosamine hydrochloride (**4.4**, Isotech) as the starting material to afford compound **4.3**. For both compounds **4.2** and **4.3**, after removal of the solvent by rotary evaporation, the residue was purified by silica gel chromatography using 3:4:10 water:pyridine:ethyl acetate, dried with $MgSO_4$, lyophilized, and stored at –20 °C. **Compound 4.2**. ESI-MS(+) calcd for $^{13}C_2{}^{12}C_6H_{15}NO_6$ $[M+H+Na]^+$: 446.15 Da, found 446.11 Da $[M+H]^+$. **Compound 4.3**. ESI-MS(+) calcd for $^{13}C_3{}^{12}C_5H_{15}{}^{15}NO_6$ $[M+H+Na]^+$: 448.15 Da, found 448.11 Da $[M+H]^+$.

*4.4.4 Experimental*

**Preparation of GlcNAc isomix.** Stock solutions of 10 mM *N*-acetyl-D-glucosamine (**4.1**), *N*-[1,2-$^{13}C_2$]acetyl-D-glucosamine (**4.2**), and *N*-[1,2-$^{13}C_2$]acetyl-D-[1-$^{13}$C,$^{15}$N]glucosamine (**4.3**) in water were prepared. The solutions were combined to produce a mixture of the GlcNAc isotopologs at a 1:2:1 molar ratio, respectively. The mixture was analyzed by direct infusion on a Thermo-Finnigan LTQ-XL mass spectrometer set to zoom scan, with the signal averaged over 20 scans. The isotopic ratios were then adjusted by the iterative addition of the desired isotopes until a near-perfect 1:2:1 peak intensity ratio was observed empirically (Figure 4-2B). The isomix sample was lyophylized and stored at –20 °C.

**Preparation of two-component GlcNAc isomix solutions.** Two-component isomix solutions were prepared analogously to the three-component GlcNAc isomix using stock solutions

of 10 mM *N*-acetyl-D-glucosamine (**4.1**) and *N*-acetyl-D-[1-$^{13}$C;$^{15}$]glucosamine (**4.4**) in water. The solutions were combined to produce a mixture of the GlcNAc isotopologs at either a 1:2 or 1.75:1 peak intensity ratio. The isomix samples were lyophylized and stored at –20 °C.

**Cell culture.** Cultures of gna1Δ *S. cerevisiae* (*MAT*a *ura*3Δ *yfl*017c::*KanMX*4) were grown in CSM (MP Biomedicals) with 2% dextrose as the carbon source and a 100 µM GlcNAc isomix supplement at 30 °C until the $OD_{600}$ = 1 for mid-log phase or the $OD_{600}$ = 8 for stationary phase harvests. A 100 µM regular GlcNAc supplement was used for negative controls.

**Preparation of yeast lysate.** Yeast cells were pelleted by centrifugation and flash-frozen in $LN_2$ and stored at –80 °C. Whole-cell lysates were prepared from 2 g aliquots of log- and stationary-phase cell pellets, each resuspended in 20 mL of lysis buffer (PBS, 1% w/v Rapigest surfactant, 1 mM dithiothreitol). Rapigest, an acid-labile surfactant, was prepared following the synthesis reported in US patent #7229539. Cells were lysed with 10 passes through a homogenizer (Avestin Emulsiflex-C3) set to a maximum pressure of 30,000 psi and crude extracts were clarified in a Sorvall SS-34 rotor spun at 19,500 rpm for 30 min at 4 °C. Clarified lysates were filtered through a 0.2 µm Steriflip filter (Millipore) and total protein concentrations were measured colorimetrically using the DC assay (Biorad). Lysates were aliquoted into small portions, flash frozen in $LN_2$, and stored at –80 °C until needed.

**FASP processing of lysates.** Milligram-scale aliquots of log-phase and stationary-phase lysates were prepared for MS analysis via a modified version of the previously described Filter-Aided Sample Preparation (FASP) methodology.[10,29] Lysates (1.2 mg of total protein/sample) were diluted to a total volume of 2 mL with freshly prepared buffered urea (7 M urea in 100 mM Tris pH 7.4 and 2.5 mM dithiothreitol). Samples were loaded into 30-KDa NMWCO Amicon Ultra centrifugal filtration devices (Millipore), spin-concentrated, and adjusted to a final volume of 600 µL with buffered urea. Cysteine residues were alkylated by addition of 50 µL of freshly-prepared 500 mM iodoacetamide followed by incubation for 1 h at RT in the dark. Samples were sequentially washed and spin-concentrated with 1 mL of buffered urea (two washes) followed by PBS with 2.5 mM dithiothreitol (three washes). Final volumes were adjusted to 1 mL with PBS. Tryptic digests were obtained by addition of 5 µg of sequencing-grade trypsin (Promega) and incubation for 12 h at 37 °C. Chymotryptic digests were performed by addition of 5 µg of sequencing-grade chymotrypsin (Promega) followed by incubation for 12 h at 25 °C. Peptides were recovered by centrifugation through 30-KDa NMWCO filters. Peptide samples were diluted to a total volume of approximately 2 mL with 2X ConA binding buffer (40 mM Tris pH 7.4, 1 M NaCl, 2 mM $CaCl_2$, 2mM $MnCl_2$). A 5 mg/mL stock solution of *C. ensiformis* ConA (Sigma) was prepared in 1X ConA binding buffer, and 2.5 mg of ConA were added to the digested peptides followed by incubation for 1 h at RT. The lectin-peptide mixtures were transferred to fresh 30-KDa NMWCO Amicon Ultra filters that had been pre-washed with 1X ConA binding buffer. ConA-bound peptides were washed by spin-concentration and diluted 4x with 1 mL of EndoH buffer (50 mM sodium citrate pH 5.3, 0.5 M sodium chloride, 1 mM calcium chloride, 1 mM manganese chloride). ConA-bound peptides were incubated with 5 µL (5000 units) of $EndoH_f$ (New England Biolabs) for 3 h

at 37 °C. Deglycosylated peptides were collected via centrifugation into siliconized eppendorf tubes. Samples were briefly acidified with trifluoroacetic acid to pH = 2 in order to hydrolyze any remaining Rapigest detergent, and insoluble debris was removed by centrifugation. Samples were desalted using Sep-Pak $C_{18}$ cartridges (Waters) and solvent was removed by speed-vac before storing at –20 °C until MS analysis.

*4.4.5 Data analysis*

**Database searches.** Peptide identities were obtained using the SEQUEST search algorithm[51] within Proteome Discoverer 1.2 (Thermo-Fisher). CID spectra were searched against the sequence database generated from all systematically named *S. cerevisiae* ORFs (downloaded March, 2011)[37] augmented with sequences from the common repository of adventitious proteins (cRAP, from The Global Proteome Machine Organization, downloaded March, 2011) and the Ngt1 (*C. albicans)* and NAGK (*H. sapiens*) protein sequences, totaling 6739 entries. Indexed databases for tryptic and chymotryptic digests were created allowing for two missed cleavages, one non-enzymatic terminus, one fixed modification (cysteine carboxyamidomethylation, +57.021) and the following variable modifications: methionine oxidation (+15.995) and asparagine GlcNAcylation (only the M + 2 isotopolog, +205.086 Da). Precursor ion tolerance was set to 10 ppm and CID fragment tolerance was set to 0.8 Da. The criteria used for filtering search results included using SEQUEST score function (*XCorr*) cutoffs assigned by a 5% maximum false discovery rate (FDR) obtained from decoy database searches (using reversed sequences from each of the 6739 entries in our database), an 8 ppm maximum allowed precursor ion mass deviation, and a 40 amino acid maximum allowed peptide length. Approximately 70% of the peptides reported here were identified within a 1 % maximum FDR. Importantly, all precursor ion and fragmentation spectra were visually inspected for the isomix signature by authors both myself and Mark Breidenbach; only peptides agreed upon by both of us were considered valid identifications. Precursor ion isomix signatures were also visually verified in the corresponding full-scan datasets. In cases where the isomix signature was identifiable in the low-resolution fragmentation spectra, this additional information was used to verify correct glycosite assignment, especially in peptides containing multiple Asn residues. For datasets collected with an IW of 2, we used the full-scan only dataset (i.e., the LC-MS run which was subjected to pattern searching) to confirm the isomix signature. In cases where multiple peptide sequences covered a single glycosite, a representative peptide sequence was selected based on CID spectral quality. Factors such as protein biological function and the presence of a canonical *N*-glycosylation consensus motif (N-X-S/T) were not used as criteria for accepting or rejecting peptides. Instances of ambiguous PTM assignment within a peptide were resolved by manual inspection of fragment envelopes as necessary. In one case, highlighted in Appendix B.3, the exact site could not be resolved amongst two possibilities. The primary data associated with this manuscript may be downloaded from the Tranche repository (https://proteomecommons.org/tranche/); hash codes are available in Appendix B.1.

Images of simulated peptide envelopes were created with Isotopica (http://coco.protein.osaka-u.ac.jp/isotopica/) and the *N*-glycosylation consensus site frequency plot was generated with Weblogo (http://weblogo.berkeley.edu/logo.cgi).

## 4.5 References

1. Zhou H., Watts J.D. and Aebersold R. (**2001**) A systematic approach to the analysis of protein phosphorylation. *Nature biotechnology* 19, 375-378.

2. Olsen J.V. *et al.* (**2006**) Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* 127, 635-648.

3. Zhang J. *et al.* (**2009**) Lysine acetylation is a highly abundant and evolutionarily conserved modification in Escherichia coli. *Mol Cell Proteomics* 8, 215-225.

4. Mischerikow N. and Heck A.J. (**2011**) Targeted large-scale analysis of protein acetylation. *Proteomics* 11, 571-589.

5. Xu G., Paige J.S. and Jaffrey S.R. (**2010**) Global analysis of lysine ubiquitination by ubiquitin remnant immunoaffinity profiling. *Nature biotechnology* 28, 868-873.

6. Danielsen J.M. *et al.* (**2011**) Mass spectrometric analysis of lysine ubiquitylation reveals promiscuity at site level. *Mol Cell Proteomics* 10, M110 003590.

7. Kaji H. *et al.* (**2003**) Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. *Nature biotechnology* 21, 667-672.

8. Wang Z. *et al.* (**2010**) Enrichment and site mapping of O-linked N-acetylglucosamine by a combination of chemical/enzymatic tagging, photochemical cleavage, and electron transfer dissociation mass spectrometry. *Mol Cell Proteomics* 9, 153-160.

9. Zhang H., Li X.J., Martin D.B. and Aebersold R. (**2003**) Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nature biotechnology* 21, 660-666.

10. Zielinska D.F., Gnad F., Wisniewski J.R. and Mann M. (**2010**) Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. *Cell* 141, 897-907.

11. Peterman S.M. and Mulholland J.J. (**2006**) A novel approach for identification and characterization of glycoproteins using a hybrid linear ion trap/FT-ICR mass spectrometer. *Journal of the American Society for Mass Spectrometry* 17, 168-179.

12. Pouria S. *et al.* (**2004**) Glycoform composition profiling of O-glycopeptides derived from human serum IgA1 by matrix-assisted laser desorption ionization-time of flight-mass spectrometry. *Analytical biochemistry* 330, 257-263.

13. Wang Z. and Hart G.W. (**2008**) Glycomic Approaches to Study GlcNAcylation: Protein Identification, Site-mapping, and Site-specific O-GlcNAc Quantitation. *Clin Proteom* 4, 5-13.

14. Domon B. and Aebersold R. (**2006**) Mass spectrometry and protein analysis. *Science (New York, N.Y* 312, 212-217.

15. Dean N. (**1999**) Asparagine-linked glycosylation in the yeast Golgi. *Biochimica et biophysica acta* 1426, 309-322.

16. Weerapana E. and Imperiali B. (**2006**) Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems. *Glycobiology* 16, 91R-101R.

17. Imperiali B. and Rickert K.W. (**1995**) Conformational implications of asparagine-linked glycosylation. *Proc Natl Acad Sci U S A* 92, 97-101.

18. Varki A. *et al.* (**2009**) Essentials of Glycobiology, Edn. 2nd. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).

19. Kukuruzinska M.A. and Lennon K. (**1998**) Protein N-glycosylation: molecular genetics and functional significance. *Crit Rev Oral Biol Med* 9, 415-448.

20. Elbein A.D. (**1984**) Inhibitors of the biosynthesis and processing of N-linked oligosaccharides. *CRC critical reviews in biochemistry* 16, 21-49.

21. Marquardt T. and Freeze H. (**2001**) Congenital disorders of glycosylation: glycosylation defects in man and biological models for their study. *Biological chemistry* 382, 161-177.

22. Robinson N.E. and Robinson A.B. (**2001**) Deamidation of human proteins. *Proc Natl Acad Sci U S A* 98, 12409-12413.

23. Angel P.M., Lim J.M., Wells L., Bergmann C. and Orlando R. (**2007**) A potential pitfall in 18O-based N-linked glycosylation site mapping. *Rapid Commun Mass Spectrom* 21, 674-682.

24. Hagglund P., Bunkenborg J., Elortza F., Jensen O.N. and Roepstorff P. (**2004**) A new strategy for identification of N-glycosylated proteins and unambiguous assignment of their glycosylation sites using HILIC enrichment and partial deglycosylation. *Journal of proteome research* 3, 556-566.

25. Zhang W., Wang H., Zhang L., Yao J. and Yang P. (**2011**) Large-scale assignment of N-glycosylation sites using complementary enzymatic deglycosylation. *Talanta* 85, 499-505.

26. Palaniappan K.K. *et al.* (**2011**) Isotopic Signature Transfer and Mass Pattern Prediction (IsoStamp): An Enabling Technique for Chemically-Directed Proteomics. *ACS chemical biology*.

27. Barnes G., Hansen W.J., Holcomb C.L. and Rine J. (**1984**) Asparagine-linked glycosylation in Saccharomyces cerevisiae: genetic analysis of an early step. *Molecular and cellular biology* 4, 2381-2388.

28. Breidenbach M.A. *et al.* (**2010**) Targeted metabolic labeling of yeast N-glycans with unnatural sugars. *Proc Natl Acad Sci U S A* 107, 3988-3993.

29. Wisniewski J.R., Zougman A., Nagaraj N. and Mann M. (**2009**) Universal sample preparation method for proteome analysis. *Nature methods* 6, 359-362.

30. Biemann K. (**1990**) Sequencing of peptides by tandem mass spectrometry and high-energy collision-induced dissociation. *Methods in enzymology* 193, 455-479.

31. Conboy J.J. and Henion J.D. (**1992**) The Determination of Glycopeptides by Liquid-Chromatography Mass-Spectrometry with Collision-Induced Dissociation. *Journal of the American Society for Mass Spectrometry* 3, 804-814.

32. Michalski A., Cox J. and Mann M. (**2011**) More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. *Journal of proteome research* 10, 1785-1793.

33. (**2010**) The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Res* 38, D142-148.

34. Gelperin D.M. *et al.* (**2005**) Biochemical and genetic analysis of the yeast proteome with a movable ORF collection. *Genes & development* 19, 2816-2826.

35. Kung L.A. *et al.* (**2009**) Global analysis of the glycoproteome in Saccharomyces cerevisiae reveals new roles for protein glycosylation in eukaryotes. *Molecular systems biology* 5, 308.

36. Brachmann C.B. *et al.* (**1998**) Designer deletion strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast (Chichester, England)* 14, 115-132.

37. Cherry J.M. *et al.* (**1997**) Genetic and physical maps of Saccharomyces cerevisiae. *Nature* 387, 67-73.

38. Titani K. *et al.* (**1986**) Amino acid sequence of human von Willebrand factor. *Biochemistry* 25, 3171-3184.

39. Gavel Y. and von Heijne G. (**1990**) Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering. *Protein Eng* 3, 433-442.

40. Kowarik M. *et al.* (**2006**) Definition of the bacterial N-glycosylation site consensus sequence. *EMBO J* 25, 1957-1966.

41. Nesvizhskii A.I. and Aebersold R. (**2005**) Interpretation of shotgun proteomic data: the protein inference problem. *Mol Cell Proteomics* 4, 1419-1440.

42. Aebersold R. and Goodlett D.R. (**2001**) Mass spectrometry in proteomics. *Chem Rev* 101, 269-295.

43. Kall L., Storey J.D., MacCoss M.J. and Noble W.S. (**2008**) Assigning significance to peptides identified by tandem mass spectrometry using decoy databases. *Journal of proteome research* 7, 29-34.

44. Klammer A.A., Park C.Y. and Noble W.S. (**2009**) Statistical calibration of the SEQUEST XCorr function. *Journal of proteome research* 8, 2106-2113.

45. Chalkley R.J. and Clauser K.R. (**2012**) Modification site localization scoring: strategies and performance. *Molecular & cellular proteomics : MCP* 11, 3-14.

46. Olsen J.V. *et al.* (**2009**) A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. *Mol Cell Proteomics* 8, 2759-2769.

47. Krogh A., Larsson B., von Heijne G. and Sonnhammer E.L. (**2001**) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of molecular biology* 305, 567-580.

48. Sletten E.M. and Bertozzi C.R. (**2009**) Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. *Angew Chem Int Ed Engl* 48, 6974-6998.

49. Du P., Kibbe W.A. and Lin S.M. (**2006**) Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. *Bioinformatics (Oxford, England)* 22, 2059-2065.

50. Zhu Y. *et al.* (**2006**) [13C,15N]2-Acetamido-2-deoxy-D-aldohexoses and their methyl glycosides: synthesis and NMR investigations of J-couplings involving 1H, 13C, and 15N. *The Journal of organic chemistry* 71, 466-479.

51. Eng J.K., Mccormack A.L. and Yates J.R. (**1994**) An Approach to Correlate Tandem Mass-Spectral Data of Peptides with Amino-Acid-Sequences in a Protein Database. *Journal of the American Society for Mass Spectrometry* 5, 976-989.

52. Breidenbach M.A., Palaniappan K.K., Pitcher A.A. and Bertozzi C.R. (**2012**) Mapping yeast

N-glycosites with isotopically recoded glycans. *Mol Cell Proteomics* 11, M111 015339.

# SECTION II: NEW TOOLS FOR $^{129}$Xe-BASED MRI CONTRAST AGENTS

# Chapter 5: Viral Nanoparticles as Contrast Agents for Magnetic Resonance Imaging

## 5.1 Introduction

Molecular imaging is a powerful technology for the visualization, characterization, and quantification of specific molecular markers at the subcellular, cellular, or organismal level.[1] It offers a non-invasive approach to study biological systems with temporal resolution and within a physiologically relevant environment. Although a number of imaging modalities have been developed, including computed tomography (CT), ultrasound, optical imaging, magnetic resonance imaging (MRI), single-photon-emission CT (SPECT), and position emission tomography (PET), MRI is well-suited for clinical imaging applications.[2] For example, MRI is non-invasive, has good depth of penetration (greater than 10 cm), and affords high spatial resolution (submillimeter) and functional images of multiple organ systems.[3] However, the detection of events at the cellular and subcellular levels often require nanomolar or subnanomolar sensitivity, precluding the use of conventional MRI techniques in molecular imaging applications due to their poor detection sensitivity.[1,4]

New imaging agents are necessary to leverage the clinical utility of MRI for targeted molecular imaging *in vivo*.[4] These agents need to generate target-specific signals that can be recorded at a high resolution, with exquisite sensitivity and specificity, while minimally perturbing and remaining inert to the biological system.[5] Accordingly, advancements to overcome the two major limitations of MRI probes—low sensitivity and low selectivity—could lead to significant progress in enabling the use of magnetic resonance for biomolecular sensing and molecular imaging applications.[2] Through a number of technological advancements in the fields of biology, chemistry and physics, significant progress has been made to address these challenges, with the ultimate goal of developing systems capable of accurately detecting disease-relevant markers *in vivo*.

## 5.2 Viral Nanoparticles as MRI Contrast Agents

Signal amplification strategies can be used to overcome the low sensitivity of classical MRI probes. One approach for signal amplification relies on the use of naturally occurring biomolecules such as viral nanoparticles.[6-9]

Viral nanoparticles are a powerful class of biomolecules for the construction of nanoscale materials.[6] Two common viral capsid geometries include icosahedral and helical, as illustrated in Figure 5-1. Examples of icosahedral viruses include cowpea chlorotic mottle virus (CCMV), cowpea mosaic virus (CPMV), bacteriophage MS2 (MS2), bacteriophage QB (QB), and bacteriophage P22 (P22). Briefly, CCMV capsids, which self-assemble from 180 identical protein subunits, are hollow, spherical structures that are 28 nm in diameter.[10] CPMV capsids self assemble from two different protein subunits, each of which is comprised of 60 identical monomer proteins, to produce spherical structures that are 30 nm in diameter.[11,12] MS2 capsids, composed of 180 monomers, are a hollow and porous structure with a diameter of 27 nm.[13] The assembled

capsid contains 32 pores, each approximately 2 nm wide, which facilitate access to the capsid interior without disassembly. Closely related to MS2 are QB viral capsids, which form spherical structures that are 28 nm in diameter.[14] P22 viral capsids self assemble from 420 identical coat proteins and between 60 and 300 identical scaffold proteins into a 58 nm hollow procapsid spherical structure.[15] Upon heating, the procapsid can undergo a morphological transformation by expanding to 64 nm and releasing 12 pentamers, generating 10 nm pores, and "wiffleball" morphology.[16] Human adenoviruses assemble from a number of protein subunits include the capsid, the fiber, the penton base , and the hexon proteins into spherical structures that are 91 nm in diameter.[17] In contrast to these structures, an example of a helical virus includes the Ff class of filamentous bacteriophage (including fd and M13). These phage modulate their length in proportion to the length of their encapsulated single-stranded DNA.[18,19]

### *5.2.1 Advantages of viruses for biomedical imaging*

There are several advantages to using viral nanoparticles as scaffolds for biomedical imaging. First, viruses are biocompatible materials with inherent targeting capabilities, naturally evolved to target and deliver their cargo to specific target cells. Second, a virus's capsid is a protein shell that can be viewed as a precisely defined nanoscale material. By targeting specific amino acids that constitute the capsid's monomer proteins, high-precision materials can be built. Additionally, the capsid's inherent multivalency allows for the attachment of a large number of molecules to a single virus. And finally, viral capsids are robust and often monodisperse nanoparticles.

When considering the characteristics of viral capsids in relationship to desired properties of MRI contrast agents, viral nanoparticles can greatly extend the capabilities of current agents. This has lead to the integration of paramagnetic lanthanide complexes with viral capsids to generate capsid-based contrast agents in two primary ways: (1) modulation of magnetic relaxation using the well-established class of gadolinium chelates, or (2) modulation of bulk magnetization through chemical exchange saturation transfer (CEST) methods.

### *5.2.2 Paramagnetic metal chelates as T1-based contrast agents*

Paramagnetic complexes of metal ions have been successfully used as MRI contrast agents since the late 1980s due to their ability to shorten the longitudinal relaxation time ($T_1$) of nearby water protons. Of these, Gadolinium(III) ($Gd^{3+}$) is the most frequently used metal. However, because free $Gd^{3+}$ ions are highly toxic *in vivo*, they must be sequestered in biocompatible ligands with high affinity before clinical approval.[20-22] Commercially used ligands discussed in this section include: tetraazocyclododecane tetracetic acid (DOTA) and diethylenetriaminopentaacetic acid ligand (DPTA), which bind $Gd^{3+}$ ions with dissociation constants of $10^{-22}$.[23] In the following section, examples of MRI contrast agents that incorporate $Gd^{3+}$ ions or $Gd^{3+}$-Ligand complexes onto viral capsids in order to improve $T_1$ relaxation rates are described.

In 2005, Douglas, Young, and coworkers were the first to functionalize a viral capsid as an MRI contrast agent.[24] Their efforts were focused on CCMV, a metal binding virus with a natural affinity for calcium ($Ca^{2+}$) ions. Based on structural data, there was evidence for 180 metal-binding sites, each located at the interface between protein subunits on the interior of the assembled

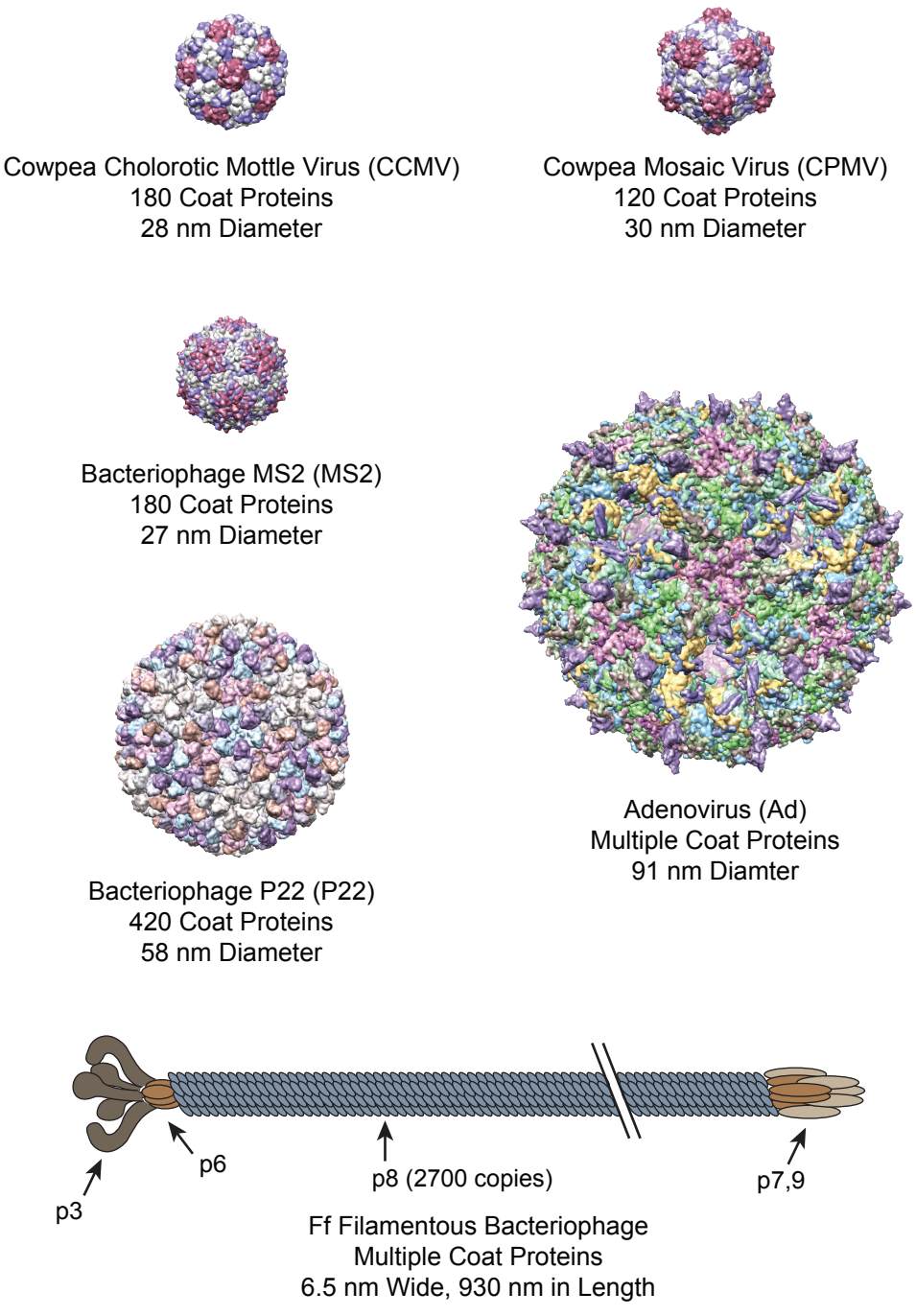



***Figure 5-1. Structural comparison of viral capsids used to build MRI contrast agents.***

Examples of icosahedral and helical viral capsids that have been used as nanoscale materials for the construction of multivalent MRI contrast agents. All structures are shown to scale except for Ff phage (the double slash indicates that the phage is much longer than shown when scaled to the minor coat proteins); for reference, bacteriophage MS2 has a diameter of 27 nm. All structural renderings were generated using UCSF Chimera.

capsid.[25,26] The authors demonstrated that in $Ca^{2+}$-free CCMV capsids, $Gd^{3+}$ ions could bind at the metal-binding sites, creating CCMV-$[Gd^{3+}]_{Int.}$ nanoparticles. Using a Forster resonance energy transfer (FRET)-based assay,[25] the binding affinity of $Gd^{3+}$ was calculated with competition experiments using Terbium ($Tb^{3+}$) bound at the metal binding sites ($Tb^{3+}$ can undergo FRET with tryptophan residues). While $Ca^{2+}$ had a relatively low binding constant (2 mM), the lanthanides bound more tightly, and $Gd^{3+}$ had a dissociation constant of 31 μM. Following metal binding studies, the $T_1$ relaxivity value of CCMV-$[Gd^{3+}]_{Int.}$ nanoparticles was calculated per $Gd^{3+}$ ion (250 $mM^{-1}s^{-1}$ at 20 MHz). This value was significantly greater than the $T_1$ relaxivity of free $Gd^{3+}$ ions ($GdCl_3$) in solution due to the decreased tumbling rate of the large nanoparticle. Additionally, because the CCMV-$[Gd^{3+}]_{Int.}$ nanoparticles contained about 140 $Ga^{3+}$ ions each, the total $T_1$ particle relaxivity was 35,000 $mM^{-1}s^{-1}$ at 20 MHz, a gain due to multivalency. Although these were large relaxivity values, the clinical utility of CCMV-$[Gd^{3+}]_{Int.}$ nanoparticles was limited by its weak affinity for $Gd^{3+}$, since free $Gd^{3+}$ ions are highly toxic *in vivo* and must be sequestered in biocompatible ligands with high affinity before clinical approval.[20-22] Despite this limitation, CCMV-$[Gd^{3+}]_{Int.}$ nanoparticles demonstrated that with improved $Gd^{3+}$ chelation, viral capsids have a great potential as a scaffold for MRI contrast agents.

To generate viral capsids with both high $Gd^{3+}$ affinity and high relaxivity rates, Douglas, Young, and coworkers attached $Gd^{3+}$ chelators to CCMV.[27] This was accomplished through two approaches: genetic engineering and chemical modification. In the genetic engineering strategy, a nine-residue $Ca^{2+}$-binding peptide (CAL) from calmodulin was fused to the N-terminus of the CCMV capsid protein. This allowed $Ga^{3+}$ ions to bind on the external surface of the capsid (CCMV-$[CAL(Gd^{3+})]_{Ext.}$). The $T_1$ relaxivity value of this construct, per $Gd^{3+}$ ion, was similar to CCMV-$[Gd^{3+}]_{Int.}$ (210 $mM^{-1}s^{-1}$ at 61 MHz), which can be attributed to the minimal rotational freedom for $Gd^{3+}$ ions coordinated to the viral capsid. Using the same FRET-based assay, the dissociation constant of $Gd^{3+}$ from CCMV-CAL was approximately 82 ± 14 nM, a 380-fold improvement over CCMV-$[Gd^{3+}]_{Int.}$. However, in comparison to the dissociation constant for clinically approved $Gd^{3+}$ chelators (in the range of $10^{-20}$),[20] the CAL peptide bound $Gd^{3+}$ ions too weakly for *in vivo* applications.

As an alternative strategy, Douglas, Young, and coworkers chemically modified CCMV with the clinically approved $Gd^{3+}$ chelator DOTA, which has a dissociation constant of $10^{-20}$ for $Gd^{3+}$ ions.[28] DOTA was conjugated to the exterior surface of CCMV by targeting lysine residues with an *N*-hydroxysuccinamide ester (NHS)-functionalized DOTA analogue (**5.1**, Figure 5-2A). While the $T_1$ relaxivity value of this construct, CCMV-$[DOTA(Gd^{3+})]_{Ext.}$, was lower than CCMV-$[CAL(Gd^{3+})]_{Ext.}$ per $Gd^{3+}$ ion (46 $mM^{-1}s^{-1}$ at 61 MHz (Table 1)), it was a 10-fold improvement compared to free $DOTA(Gd^{3+})$ complexes.[27,28] One explanation was although $DOTA(Gd^{3+})$ complexes attached to the CCMV capsid experienced a reduced tumbling rate, conjugation through a flexible lysine residue allowed for more local mobility than $Gd^{3+}$ ions locked into the inter-subunit binding pocket of the CAL peptide. This additional motion, independent of the much slower rotational motion of the virus, could decrease relaxivity gains imparted by the viral capsid. Additionally, it was possible multiple water molecules could coordinate to a $Gd^{3+}$ ion when it was bound to the CAL peptide compared to the single water molecule the DOTA ligand could accommodate. Overall, by conjugating $Gd^{3+}$ chelates to CCMV, nanoparticles with both high relaxivity rates and high $Gd^{3+}$ binding affinities were generated.

In addition to CCMV, other viruses have been used as paramagnetic metal-binding scaffolds for $T_1$-based contrast agents. In 2006, Kirshenbaum and coworkers chemically conjugated the coat protein of bacteriophage MS2 with the clinically approved $Gd^{3+}$ chelator DPTA.[29] DPTA was conjugated to the exterior surface of MS2 by targeting lysine residues with an isothiocyanate-functionalized DPTA analogue (**5.3**, Figure 5-2A). Although the $T_1$ relaxivity value of this construct, MS2-$[DPTA(Gd^{3+})]_{Ext.}$, was lower than its CCMV analogue (CCMV-$[DOTA(Gd^{3+})]_{Ext.}$) per $Gd^{3+}$ ion (14 $mM^{-1}s^{-1}$ at 64 MHz (Table 1)), it was a 3-fold improvement over free $DPTA(Gd^{3+})$.[28,29] A likely factor contributing to the low relaxivity value was the additional rotational freedom imparted by the flexible lysine residues serving as linkers for the DPTA ligands. However, with over 500 $Gd^{3+}$ ions per virus, the total $T_1$ particle relaxivity was 7,200 $mM^{-1}s^{-1}$ at 64 MHz, emphasizing the advantage of multivalent scaffolds when constructing an MRI contrast agent.

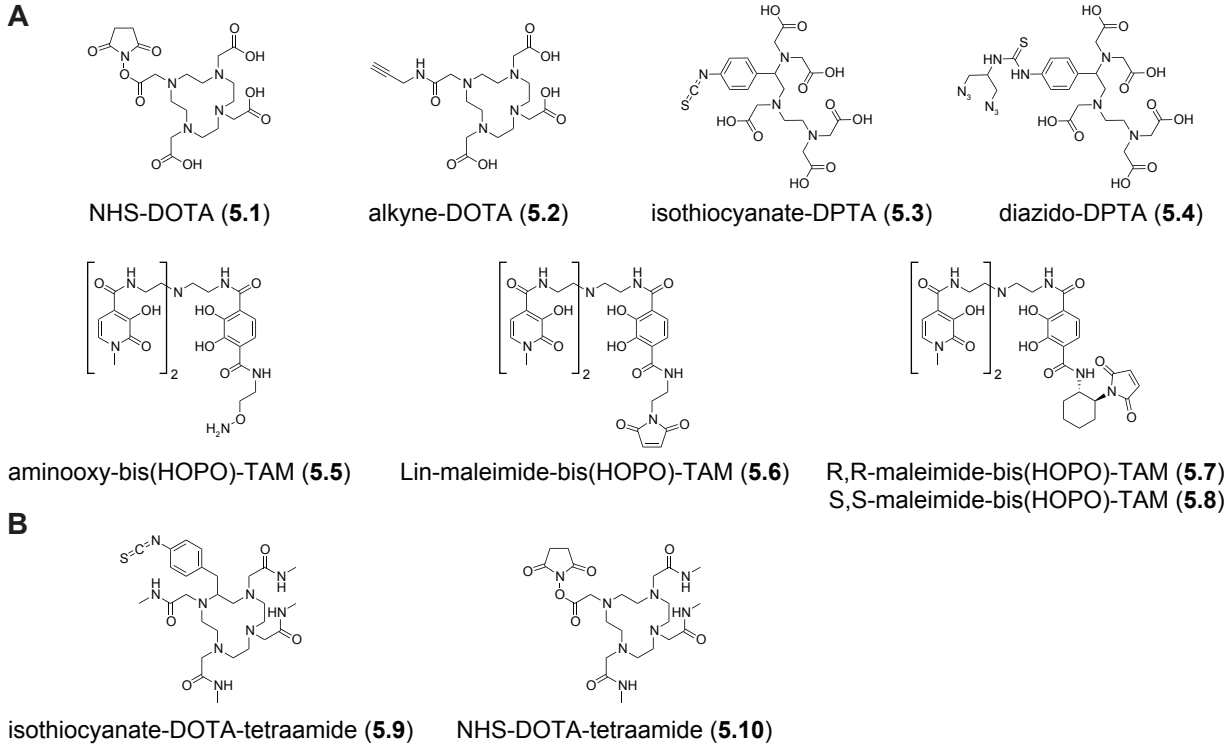



***Figure 5-2. Chemical structures of metal chelates conjugated to viral capsids to generate MRI contrast agents.***
(A) Examples of functionalized gadolinium chelates used for T1-generating contrast agents. (B) Examples of functionalized lanthanide chelates used for PARACEST measurements.

Both CPMV and bacteriophage QB have also been used to construct MRI contrast agents. In 2007, Finn, Manchester, and coworkers attached $Gd^{3+}$ ions to both viruses using two complimentary strategies: either through coordination to chemically conjugated DOTA chelates, or through affinity to polynucleotides encapsulated within the viral capsid.[30] For chemical modification of the viral capsids, the copper-catalyzed azide-alkyne cycloaddition (CuCCA)[31-33] reaction (i.e., "click chemistry") was used to covalently attach an alkyne-modified DOTA ligand

(**5.2**, Figure 5-2A) to the exterior surface of capsids through azide-functionalized lysine residues (CPMV-[DOTA($Gd^{3+}$)]$_{Ext.}$ and QB-[DOTA($Gd^{3+}$)]$_{Ext.}$). In the case of CPMV, $Ga^{3+}$ ions were also attached to the interior of the capsid by exploiting their inherent affinity to the encapsulated viral RNA (CPMV-[$Gd^{3+}$]$_{Int.}$). Combining the two strategies produced a hybrid CPMV nanoparticle with DOTA($Gd^{3+}$) complexes on the exterior and RNA-$Gd^{3+}$ complexes in the interior of the virus (CPMV-[$Gd^{3+}$]$_{Int.}$[DOTA($Gd^{3+}$)]$_{Ext.}$). Interestingly, all 4 constructs had similar $T_1$ relaxivity values per $Gd^{3+}$ ion (between 11.9 and 15.5 $mM^{-1}s^{-1}$ at 64 MHz), exhibiting a 2- to 3-fold improvement over free DOTA($Gd^{3+}$) or DPTA($Gd^{3+}$) complexes.[28] Two explanations for the minimal gain in relaxivity were the slow exchange kinetics of water with $Gd^{3+}$ ions chelated to the DOTA ligand or the limited access to the interior of the CPMV capsid. However, the large number of $Gd^{3+}$ ions per viral capsids generated large $T_1$ particle relaxivities (between 1,230 and 4,150 $mM^{-1}s^{-1}$ at 64 MHz), highlighting the advantage of a multivalent protein scaffold.

To functionalize the interior surface of MS2 viral particles with $Gd^{3+}$ ions, Francis, Raymond, and coworkers explored selective bioconjugation strategies to chemically modify MS2 with a HOPO ligand.[34,35] HOPO was site-specifically conjugated to the interior surface of MS2 by first converting tyrosine residues into an aldehyde through diazonium chemistry.[36,37] In a second step, an aminooxy-functionalized HOPO analogue (**5.5**, Figure 5-2A) was condensed with the aldehyde to form a stable oxime linkage (MS2-[HOPO($Gd^{3+}$)]$_{Int.}$). As a comparison, externally modified MS2 constructs were created by converting surface exposed lysine residues to aldehydes using NHS-ester chemistry followed by condensation with HOPO ligands **5.5** (MS2-[HOPO($Gd^{3+}$)]$_{Ext.}$). The internally modified MS2 nanoparticles offered a number of advantages, including an unmodified external surface for functionalization with targeting groups and improved solubilization of the hydrophobic HOPO ligands at greater than 50% capsid modification. Additionally, the $T_1$ relaxivity per $Gd^{3+}$ ion was greater for the internally modified MS2 than the externally modified MS2 (31 $mM^{-1}s^{-1}$ verses 23.2 $mM^{-1}s^{-1}$, respectively, at 60 MHz). The approximately 1.3-fold improvement in relaxivity could result from the decreased rotational freedom of tyrosine residues compared to lysine residues, both of which served at linkers for the HOPO ligands. These results also represented a 4- to 5-fold improvement in relaxivity compared to free HOPO($Gd^{3+}$) complexes and demonstrated water transport across the capsid shell of MS2 was fast on the NMR timescale. Subsequently, these two MS2-HOPO constructs were analyzed using nuclear magnetic resonance dispersion (NMRD) spectroscopy.[34]

Inspired by the improvement in relaxivity of the internally modified MS2 construct (MS2-[HOPO($Gd^{3+}$)]$_{Int.}$), the Francis group further improved this viral capsid-based contrast agent by increasing the bioconjugation efficiency of a more rigid linker that connected a HOPO ligand to the MS2 capsid coat protein.[38] HOPO ligands bearing a maleimide group through either a flexible linear linker (**5.6**, Figure 5-2A) or a rigid, chiral cyclohexyl linker (**5.7** or **5.8**, Figure 5-2A) were synthesized and then conjugated to the interior surface of MS2 by targeting a single surface exposed cysteine residue,[39,40] which generated MS2-[HOPO-Lin($Gd^{3+}$)]$_{Int.}$, MS2-[HOPO-R,R($Gd^{3+}$)]$_{Int.}$, and MS2-[HOPO-S,S($Gd^{3+}$)]$_{Int.}$, respectively. The high efficiency of cysteine chemistry led to near quantitative modification of MS2 capsids, resulting in nearly 180 copies of each HOPO ligand. As predicted, the $T_1$ relaxivity per $Gd^{3+}$ ion was greater for the construct based on the rigid cyclohexyl linker **5.8** compared to the flexible linear linker **5.6** (41.2 $mM^{-1}s^{-1}$ verses 32.6 $mM^{-1}s^{-1}$,

respectively, at 60 MHz), and is one of the highest reported relaxivity values for high-affinity $Gd^{3+}$ complexes. The approximately 1.3-fold improvement in relaxivity resulted from the decreased rotational freedom of the linker. Interestingly, there was a significant difference in the $T_1$ relaxivity per $Gd^{3+}$ ion between the two enantiometic cyclohexyl ligands (**5.7** and **5.8**) when attached to MS2. Through molecular modeling studies this difference was attributed to the orientation of the HOPO ligand with respect to the protein surface: the R,R linker **5.7** was directed towards the protein surface, limiting access of water molecules to the metal center. These results demonstrated that a rational approach could be used to design future viral capsid-based MRI contrast agents with improved relaxivity properties.

The use of chemically modified viral capsids for high relaxivity MRI contrast agents has also been applied to bacteriophage P22. Recent work by Douglas and coworkers demonstrated that the interior cavity of P22 could be modified with a branched oligomer containing multiple DTPA($Gd^{3+}$) complexes.[41] DTPA was conjugated to the interior surface of the "wiffleball" form of P22 following a CuAAC reaction between an azide-functionalized DPTA ligand (**5.4**, Figure 5-2A) and alkyne-functionalized cysteine residues. Through sequential "click reactions" with DPTA ligands (**5.4**), oligomerization proceeded off of the internal surface of the capsid and facilitated a large payload of $Gd^{3+}$ chelating ligands (i.e., 1900) inside the capsid. Although the $T_1$ relaxivity per $Gd^{3+}$ ion of this construct, P22-$[DPTA(Gd^{3+})]_{Int.}$, was similar to other viral capsids conjugated to either DOTA or DPTA ligands (21.7 $mM^{-1}s^{-1}$ at 28 MHz), the total $T_1$ particle relaxivity was 41,300 $mM^{-1}s^{-1}$ at 28 MHz, one of the highest reported. The large volume of bacteriophage P22 conferred a significant multivalency advantage compared to CCMV, CPMV, or both bacteriophage QB and MS2, which translated to a higher loading capacity of $Gd^{3+}$ chelating ligands and consequently higher relaxivity rates per particle.

In addition to these impressive gains in relaxivities, the biocompatibility of viral capsid-based contrast agents have been established, supporting their use *in vivo*. Specifically, the biodistribution, toxicology, and pathology of CPMV and QB derivatized with $Gd^{3+}$ ions were investigated in living mice.[42,43] Although MRI was not performed, these were the first examples of viral capsid-based MRI contrast agents tested *in vivo*, and they demonstrated that the nanoparticles were nontoxic, cleared rapidly from plasma, and accumulated primarily in the liver. In the case of QB particles, the pharmacokinetics of DOTA constructs conjugated internally (via an unnatural amino acid)[43] or externally (via lysine residues)[30] were investigated. Interestingly, QB nanoparticles with high levels of external modification were cleared from circulation approximately 10-times faster than internally-modified nanoparticles. This drastic change in pharmacokinetics was attributed to the reduction of positive charge on the virus's external surface (because of lysine acylation). Subsequently, biodistribution and toxicity studies performed with radiolabeled CCMV[44] and MS2[45,46] viral capsids showed similar broad tissue and organ distribution, and low toxicity at the injected concentrations. Collectively, these results suggest that viral capsids are safe and biocompatible nanoparticles for applications in molecular imaging.

In the future, utilizing viral capsid-based MRI contrast agents for molecular imaging applications will require combining the design principles garnered from all these studies using paramagnetic chelates and nanoparticles for improving local relaxivity rates. For example, moving towards larger viruses (i.e., greater functionalizable surface area) to achieve higher relaxivity

rates will only be part of the solution. Beyond the capability to introduce a large number of $Gd^{3+}$ ions per viral capsid, efforts will need to focus on translating the reduced tumbling rate of a viral particle onto coordinated $Gd^{3+}$ ions. For example, viral particles with $Gd^{3+}$ ions directly coordinated to the virus had $T_1$ relaxivity values that were 10-15 fold higher than viral particles with $Gd^{3+}$ ions indirectly coordinated to the virus through a linker (i.e., amino acid side chain, polynucleotides, etc.). In general, by rigidifying the linker so only the rotational motion of the viral particle is translated to the $Gd^{3+}$ ion, rather than it experiencing any additional local motion, improved $T_1$ relaxivity values. However, there has to be a balance between locking the $Gd^{3+}$ ion in place while allowing for sufficient access to the metal and exchange of water molecules.

The next goal will be to integrate targeting capabilities into these contrast agents for targeted molecular imaging.

### *5.2.3 Paramagnetic metal chelates as CEST-based contrast agents*

In addition to relaxivity-based contrast agents, chemical exchange saturation transfer (CEST) agents are an alternative approach for MRI contrast enhancement.[47,48] The CEST method relies on chemically exchangeable protons that resonate at a distinct chemical shift such that they can be distinguished from the bulk water proton signal. Applying radiofrequency pulses at the appropriate frequency can selectively saturate these exchangeable protons, and when they transfer into the bulk proton pool, will decrease the bulk magnetization of the system. Accordingly, the intensity of the water proton signal will diminish after a saturation pulse at the appropriate frequency. Substantial enhancement of this effect is possible by optimizing the saturation time, allowing for the detection of low concentration target.

Although biological and chemical environments contain a number of different exchangeable protons, including –OH and –NH groups, the chemical shifts of these protons are typically very close to the bulk water signal, which makes selective saturation of the exchangeable proton difficult. Fortunately, paramagnetic metal complexes with a slowly exchanging water molecule can facilitate a measurable CEST effect.

Paramagnetic ions can be utilized to shift the bound water frequency further away from the bulk water, allowing distinct saturation of the exchangeable protons. These paramagnetic CEST agents (i.e., PARACEST) consist of paramagnetic chelates that are specifically designed to exhibit exchangeable proton or bound water peaks.[49] Unlike the paramagnetic metal complexes used as $T_1$-shortening contrast agents (discussed in the previous section), which rely on rapidly exchanging water molecules to generate contrast, a CEST-compatible complex requires slower water exchange, such as a $Gd^{3+}$ tetramide DOTA complex.

One of the main limitations with CEST and PARACEST agents for MRI are their inherent insensitivities. Consequently, nanoscale carriers have been developed to improve the limit of detection. In 2008, Sherry and coworkers were the first to functionalize a viral capsid as a PARACEST agent.[50] They conjugated DOTA-tetraamide to the exterior surface of Adenovirus by targeting lysine residues with either an isothiocyanate-functionalized analogue (**5.9**, Figure 5-2B) or an NHS-ester functionalized analogue (**5.10**, Figure 5-2B). After optimizing the conjugation efficiency to maximize the number of ligands per particle while retaining the virus's infectious biological activity (730 ligands and 72% activity, respectively), they performed thallium-based

PARACEST measurements at concentrations in the low micromolar range and observed a 12% decrease in bulk water signal relative to controls. Surprisingly, there are no other examples of signal amplification with viral capsids for PARACEST contrast agents. Instead, the combination of CEST and non-proton nuclei have been extended to signal amplification strategies with viral capsids.

*5.2.4 Xenon-based biosensors as contrast agents*

Despite the advancements made for proton-based MRI contrast agents, the fundamental limits to the gain in sensitivity had prompted alternative approaches using alternative nuclei, including $^{13}C$, $^{3}He$, and $^{129}Xe$.[51-53] Of these, xenon is an attractive option for MRI-based molecular imaging for reasons discussed in Chapter 6.

## 5.3 Summary of Section II: New Tools for $^{129}Xe$-based MRI Contrast Agents

The application of viral nanoparticles to medical imaging has greatly expanded the types of contrast agents available for targeted molecular imaging. Accordingly, in this section of my dissertation, I describe how viral nanoparticles can be used as molecular scaffolds for the development of high sensitivity contrast agents for MRI. First, in Chapter 6, I describe the utility of xenon-based MRI for *in vivo* imaging and illustrate that through the combination of hyperpolarization and chemical exchange saturation transfer detection, xenon biosensors can achieve low detection thresholds.

Towards improving the detection sensitivity of xenon MRI further, biosensors can be assembled onto supramolecular scaffolds such as viral nanoparticles. In Chapter 7, I discuss the development of a bacteriophage MS2-based xenon biosensor–the first viral capsid functionalized as a $^{129}Xe$-based MRI contrast agent. Subsequently, in Chapter 8, I extend the application of viral capsids from spherical bacteriophage to filamentous bacteriophage by generating an M13 bacteriophage-based xenon biosensor. Bacteriophages where chosen because they are routinely used in phage display techniques for identifying new epitope-targeting groups such as peptides and antibody fragments. Accordingly, in Chapter 9, I describe the development of a phage-based xenon biosensor that possesses antibody fragments for targeted imaging of cancer cells *in vitro*.

## 5.4 References

1. Massoud T.F. and Gambhir S.S. (**2003**) Molecular imaging in living subjects: seeing fundamental biological processes in a new light. *Genes Dev* 17, 545-580.

2. Weissleder R. and Pittet M.J. (**2008**) Imaging in the era of molecular oncology. *Nature* 452, 580-589.

3. Tyszka J.M., Fraser S.E. and Jacobs R.E. (**2005**) Magnetic resonance microscopy: recent advances and applications. *Curr Opin Biotechnol* 16, 93-99.

4. Sosnovik D.E. and Weissleder R. (**2007**) Emerging concepts in molecular MRI. *Curr Opin Biotechnol* 18, 4-10.

5. Cassidy P.J. and Radda G.K. (**2005**) Molecular imaging perspectives. *J R Soc Interface* 2, 133-144.

6. Dedeo M.T., Finley D.T. and Francis M.B. (**2011**) Viral capsids as self-assembling templates for new materials. *Prog Mol Biol Transl Sci* 103, 353-392.

7. Steinmetz N.F. (**2010**) Viral nanoparticles as platforms for next-generation therapeutics and imaging devices. *Nanomedicine* 6, 634-641.

8. Cormode D.P., Jarzyna P.A., Mulder W.J. and Fayad Z.A. (**2010**) Modified natural nanoparticles as contrast agents for medical imaging. *Adv Drug Deliv Rev* 62, 329-338.

9. Yildiz I., Shukla S. and Steinmetz N.F. (**2011**) Applications of viral nanoparticles in medicine. *Curr Opin Biotechnol* 22, 901-908.

10. Speir J.A., Munshi S., Wang G., Baker T.S. and Johnson J.E. (**1995**) Structures of the native and swollen forms of cowpea chlorotic mottle virus determined by X-ray crystallography and cryo-electron microscopy. *Structure* 3, 63-78.

11. Lin T. *et al.* (**1999**) The refined crystal structure of cowpea mosaic virus at 2.8 A resolution. *Virology* 265, 20-34.

12. Lin T. and Johnson J.E. (**2003**) Structures of picorna-like plant viruses: implications and applications. *Adv Virus Res* 62, 167-239.

13. Valegard K., Liljas L., Fridborg K. and Unge T. (**1990**) The three-dimensional structure of the bacterial virus MS2. *Nature* 345, 36-41.

14. Golmohammadi R., Fridborg K., Bundule M., Valegard K. and Liljas L. (**1996**) The crystal structure of bacteriophage Q beta at 3.5 A resolution. *Structure* 4, 543-554.

15. Earnshaw W., Casjens S. and Harrison S.C. (**1976**) Assembly of the head of bacteriophage P22: x-ray diffraction from heads, proheads and related structures. *J Mol Biol* 104, 387-410.

16. Teschke C.M., McGough A. and Thuman-Commike P.A. (**2003**) Penton release from P22 heat-expanded capsids suggests importance of stabilizing penton-hexon interactions during capsid maturation. *Biophys J* 84, 2585-2592.

17. Reddy V.S., Natchiar S.K., Stewart P.L. and Nemerow G.R. (**2010**) Crystal Structure of Human

Adenovirus at 3.5 angstrom Resolution. *Science* 329, 1071-1075.

18. Glucksman M.J., Bhattacharjee S. and Makowski L. (**1992**) Three-dimensional structure of a cloning vector. X-ray diffraction studies of filamentous bacteriophage M13 at 7 A resolution. *J Mol Biol* 226, 455-470.

19. Specthrie L. *et al.* (**1992**) Construction of a microphage variant of filamentous bacteriophage. *J Mol Biol* 228, 720-724.

20. Rocklage S.M., Worah D. and Kim S.H. (**1991**) Metal ion release from paramagnetic chelates: what is tolerable? *Magn Reson Med* 22, 216-221; discussion 229-232.

21. Ersoy H. and Rybicki F.J. (**2007**) Biochemical safety profiles of gadolinium-based extracellular contrast agents and nephrogenic systemic fibrosis. *J Magn Reson Imaging* 26, 1190-1197.

22. Cacheris W.P., Quay S.C. and Rocklage S.M. (**1990**) The Relationship between Thermodynamics and the Toxicity of Gadolinium Complexes. *Magn Reson Imaging* 8, 467-481.

23. Wedeking P., Kumar K. and Tweedle M.F. (**1992**) Dissociation of gadolinium chelates in mice: relationship to chemical characteristics. *Magn Reson Imaging* 10, 641-648.

24. Allen M. *et al.* (**2005**) Paramagnetic viral nanoparticles as potential high-relaxivity magnetic resonance contrast agents. *Magn Reson Med* 54, 807-812.

25. Basu G., Allen M., Willits D., Young M. and Douglas T. (**2003**) Metal binding to cowpea chlorotic mottle virus using terbium(III) fluorescence. *J Biol Inorg Chem* 8, 721-725.

26. Speir J.A., Munshi S., Wang G.J., Baker T.S. and Johnson J.E. (**1995**) Structures of the Native and Swollen Forms of Cowpea Chlorotic Mottle Virus Determined by X-Ray Crystallography and Cryoelectron Microscopy. *Structure* 3, 63-78.

27. Liepold L. *et al.* (**2007**) Viral capsids as MRI contrast agents. *Magn Reson Med* 58, 871-879.

28. Caravan P., Ellison J.J., McMurry T.J. and Lauffer R.B. (**1999**) Gadolinium(III) Chelates as MRI Contrast Agents: Structure, Dynamics, and Applications. *Chem Rev* 99, 2293-2352.

29. Anderson E.A. *et al.* (**2006**) Viral nanoparticles donning a paramagnetic coat: conjugation of MRI contrast agents to the MS2 capsid. *Nano Lett* 6, 1160-1164.

30. Prasuhn D.E., Jr., Yeh R.M., Obenaus A., Manchester M. and Finn M.G. (**2007**) Viral MRI contrast agents: coordination of Gd by native virions and attachment of Gd complexes by azide-alkyne cycloaddition. *Chem Commun (Camb)*, 1269-1271.

31. Rostovtsev V.V., Green L.G., Fokin V.V. and Sharpless K.B. (**2002**) A stepwise huisgen cycloaddition process: copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angew Chem Int Ed Engl* 41, 2596-2599.

32. Tornoe C.W., Christensen C. and Meldal M. (**2002**) Peptidotriazoles on solid phase: [1,2,3]-triazoles by regiospecific copper(i)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *J Org Chem* 67, 3057-3064.

33. Wang Q. *et al.* (**2003**) Bioconjugation by copper(I)-catalyzed azide-alkyne [3 + 2] cycloaddition. *J Am Chem Soc* 125, 3192-3193.

34. Datta A. *et al.* (**2008**) High relaxivity gadolinium hydroxypyridonate-viral capsid conjugates: nanosized MRI contrast agents. *J Am Chem Soc* 130, 2546-2552.

35. Hooker J.M., Datta A., Botta M., Raymond K.N. and Francis M.B. (**2007**) Magnetic resonance contrast agents from viral capsid shells: a comparison of exterior and interior cargo strategies. *Nano Lett* 7, 2207-2210.

36. Schlick T.L., Ding Z., Kovacs E.W. and Francis M.B. (**2005**) Dual-surface modification of the tobacco mosaic virus. *J Am Chem Soc* 127, 3718-3723.

37. Hooker J.M., Kovacs E.W. and Francis M.B. (**2004**) Interior surface modification of bacteriophage MS2. *J Am Chem Soc* 126, 3718-3719.

38. Garimella P.D., Datta A., Romanini D.W., Raymond K.N. and Francis M.B. (**2011**) Multivalent, high-relaxivity MRI contrast agents using rigid cysteine-reactive gadolinium complexes. *J Am Chem Soc* 133, 14704-14709.

39. Tong G.J., Hsiao S.C., Carrico Z.M. and Francis M.B. (**2009**) Viral capsid DNA aptamer conjugates as multivalent cell-targeting vehicles. *J Am Chem Soc* 131, 11174-11178.

40. Wu W., Hsiao S.C., Carrico Z.M. and Francis M.B. (**2009**) Genome-free viral capsids as multivalent carriers for taxol delivery. *Angew Chem Int Ed Engl* 48, 9493-9497.

41. Qazi S. *et al.* (**2012**) P22 Viral Capsids as Nanocomposite High-Relaxivity MRI Contrast Agents. *Mol Pharm*.

42. Singh P. *et al.* (**2007**) Bio-distribution, toxicity and pathology of cowpea mosaic virus nanoparticles in vivo. *J Control Release* 120, 41-50.

43. Prasuhn D.E., Jr. *et al.* (**2008**) Plasma clearance of bacteriophage Qbeta particles as a function of surface charge. *J Am Chem Soc* 130, 1328-1334.

44. Kaiser C.R. *et al.* (**2007**) Biodistribution studies of protein cage nanoparticles demonstrate broad tissue distribution and rapid clearance in vivo. *Int J Nanomedicine* 2, 715-733.

45. Farkas M.E. *et al.* (**2012**) PET Imaging and Biodistribution of Chemically Modified Bacteriophage MS2. *Mol Pharm*.

46. Hooker J.M., O'Neil J.P., Romanini D.W., Taylor S.E. and Francis M.B. (**2008**) Genome-free viral capsids as carriers for positron emission tomography radiolabels. *Mol Imaging Biol* 10, 182-191.

47. Ward K.M., Aletras A.H. and Balaban R.S. (**2000**) A new class of contrast agents for MRI based on proton chemical exchange dependent saturation transfer (CEST). *J Magn Reson* 143, 79-87.

48. van Zijl P.C.M. and Yadav N.N. (**2011**) Chemical Exchange Saturation Transfer (CEST): What is in a Name and What Isn't? *Magnetic Resonance in Medicine* 65, 927-948.

49. Woods M., Donald E.W.C. and Sherry A.D. (**2006**) Paramagnetic lanthanide complexes as PARACEST agents for medical imaging. *Chem Soc Rev* 35, 500-511.

50. Vasalatiy O., Gerard R.D., Zhao P., Sun X. and Sherry A.D. (**2008**) Labeling of adenovirus particles with PARACEST agents. *Bioconjug Chem* 19, 598-606.

51. Kurhanewicz J., Bok R., Nelson S.J. and Vigneron D.B. (**2008**) Current and potential applications of clinical C-13 MR spectroscopy. *J Nucl Med* 49, 341-344.

52. Hopkins S.R. *et al.* (**2007**) Advances in magnetic resonance imaging of lung physiology. *J Appl Physiol* 102, 1244-1254.

53. Oros A.M. and Shah N.J. (**2004**) Hyperpolarized xenon in NMR and MRI. *Phys Med Biol* 49, R105-R153.

# Chapter 6: Physical Properties and NMR-based Detection of Xenon

## 6.1 Introduction

Although proton-based magnetic resonance imaging (MRI) provides access to clinical imaging with excellent spatial resolution, it is not well-suited to imaging void spaces (i.e., lungs), low abundance molecular markers of disease, or environments where unique proton spin densities are low. To address this limitation, xenon-based MRI has emerged as an attractive alternative option. For example, the first demonstration of pulmonary void-space imaging was accomplished in 1994 by Albert *et al.*, who used xenon to image a mouse lung.[1] Subsequently, there have been many examples of xenon-based imaging of the lung, brain, and other tissues in both mice and humans.[1-6] Although these examples established the clinical relevancy of xenon, there were still challenges associated with delivering xenon to specific locations of interest, a prerequisite for targeted molecular imaging applications. In this Chapter, I will discuss the advantages of employing xenon for molecular imaging by first describing its unique physical properties and then describing a general strategy for its targeted detection.

## 6.2 Physical Properties of Xenon

Xenon is a chemically inert noble gas with two NMR-active isotopes, $^{129}$Xe and $^{131}$Xe. However, the focus here will be on $^{129}$Xe, the spin ½ isotope, which is both highly abundant (26.4% natural abundance)[7] and possesses several advantageous physical properties that make it a unique candidate for molecular imaging.[8,9]

First, xenon is soluble in many solvents, including nonpolar organic solvents such as hexanes and benzene, as well as polar aqueous solvents including water, aqueous buffers, and blood plasma.[10,11] Thus, xenon will uniquely partition between blood and tissue, a feature that can be utilized for specific imaging applications.[1,5,6,12] In this section, all of the NMR experiments use a 2% xenon gas mixture at 70-85 psia in an aqueous buffer. Under these conditions, the dissolved xenon concentration (26.4% of which is $^{129}$Xe) is approximately 500 μM and 350 μM at 25 °C and 37 °C, respectively (using an Ostwald solubility coefficient of 0.11 at 25 °C and 0.083 at 37 °C for xenon in water).[10,11]

Another advantage of xenon is its ability to respond to very subtle changes in its local envi-

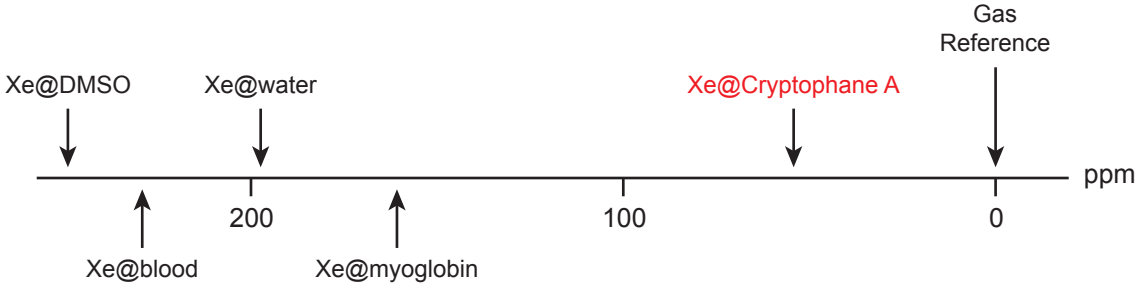



***Figure 6-1. Xenon is extremely sensitivity to its local environment.***

Examples of chemical shift values for xenon in different molecular environments, including solvents, biomolecules, and synthetic molecules (e.g., Crytophane A, highlighted in red). Chemical shifts are reported in reference to xenon gas.

ronment. Xenon possesses a large and highly polarizable electron cloud, which translates into a large chemical shift range that is sensitive to the local electronic and magnetic environment.[13] Unlike protons, which have a chemical shift range of approximately 15 ppm, xenon exhibits a range of approximately 250 ppm. It can display very distinct and well-resolved chemical shifts depending upon the solvent in which it is dissolved, or even through non-covalent association with small molecules and proteins (see Figure 6-1 for examples). This attribute allows for the potential to simultaneously observe multiple xenon interactions at once.[9] Additionally, $^{129}$Xe NMR spectra are typically less complex than those from $^{1}$H or $^{13}$C NMR, usually showing only a few, easily interpretable signals.

Finally, $^{129}$Xe is amenable to hyperpolarization, a technique that artificially modulates the spin population distribution into a nonequilibrium state.[8,14] Through a two-step process called spin-exchange optical pumping, polarization can be transferred from a strongly polarized alkaline metal to $^{129}$Xe nuclei.[15,16] In its hyperpolarized (hp) state, the $^{129}$Xe NMR signal can be increased by greater than 10,000-fold.[13,15,17] Consequently, even low concentrations of dissolved xenon can produce NMR signals comparable to that of water protons (which have a concentration of 110 M), but without any confounding background signal.

These favorable properties of xenon have already led to many *in vivo* imaging studies, demonstrating the biocompatibility of xenon-based MRI.[1-6] For example, hp $^{129}$Xe gas was used to image the mouse lung,[1,18] the alveolar surface area in human lungs,[2] and the human chest.[3] In addition to lung imaging, tissue perfusion studies have been performed to image the rat brain[19] and the human head and brain.[3,4] Polarized xenon has even been injected into rat tissue for rapid and localized delivery.[20] Additionally, compared to paramagnetic metals currently used as contrast agents for proton-based MRI (e.g., gadolinium), xenon has considerably lower *in vivo* toxicity.[21]

While these characteristics position xenon as a powerful contrast agent for molecular imaging, unlike proton-based MRI, which can utilize endogenous proton nuclei in water or fat for generating images, xenon must be introduced into biological environments. However, without a compatible targeting agent, xenon gas will nonspecifically partition itself in a complex biological environment.[12] To address this limitation, host-molecules that interact strongly with xenon can be employed to improve xenon's detection and facilitate targeting.

## 6.3 Detection of Xenon with Cryptophane-A

Since xenon can nonspecifically interact with biomolecules, including proteins and lipids, xenon-binding molecules are necessary to achieve targeted molecular imaging.[22,23] Due to xenon's non-polar character, a number of molecules with hydrophobic cavities have been proposed as xenon-binding hosts, including cyclodextrins, calixarenes, porous materials, hemicarcerands (including cryptophanes), and proteins.[17,22] Of these, the most well-studied molecule, with respect to xenon binding, are the cryptophanes.

Cryptophanes are small, hydrophobic, cage-like molecules developed in the early 1980s by Collet and coworkers.[24,25] They consist of two cyclotriveratrylene caps and three linkers holding the caps together. A variety of cryptophane derivatives with unique xenon binding properties have been synthesized by varying substitutions on the cyclotriveratrylene caps or the length of

the linkers connecting the caps together.[25,26] Of these, cryptophane-A (CryA, **6.1**), which consists of methoxy-functionalized cyclotriveratrylene caps connected through three ethyl linkers, was the first of the cryptophane family of molecules to be used as a xenon-binding host for NMR studies (Figure 6-2A).[27,28] With a xenon binding affinity of 3900 $M^{-1}$ (in organic solvent)[29] and a residence time in the millisecond range (30-300 ms in aqueous solution), [29,30] CryA provides optimal xenon complexation and exchange kinetics for NMR measurements.

CryA molecular cages are ideal hosts for NMR studies because they minimally relax xenon while giving rise to a unique chemical shift that is well resolved from that of free xenon in solution (i.e., Xe@water). Due to the highly sensitive chemical shift response of $^{129}$Xe to its local environment, two distinct resonance frequencies representing the Xe@CryA and Xe@water pools can be individually detected at approximately 60 ppm and 190 ppm downfield from the $^{129}$Xe gas chemical shift (0 ppm), respectively (Figure 6-2B). It should be pointed out, however, that only approximately 1% of dissolved xenon is associated with CryA in solution.[31,32] Consequently, high micromolar concentrations of CryA or long acquisition times for signal averaging are required

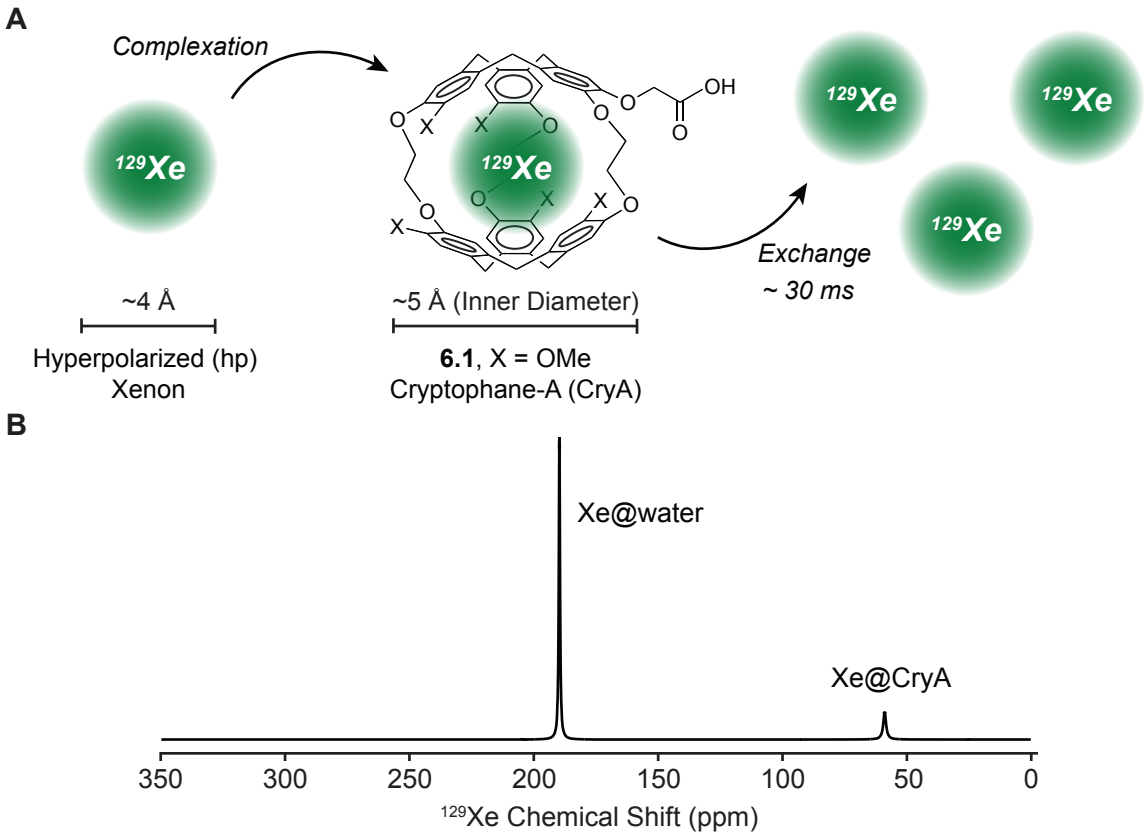



***Figure 6-2. Direct detection of xenon through complexation with Crytophane-A molecular cages.***

(A) The chemical structure of Crytophane-A (**6.1**, CryA), and a schematic depicting how it interacts with a pool of hp $^{129}$Xe nuclei (depicted in green). The chemical shift of $^{129}$Xe changes to approximately 60 ppm after complexation with CryA (Xe@CryA) in water. After approximately 30 ms (at room temperature), xenon will exchange with the bulk pool of free xenon in solution (i.e., Xe@water). (B) A theoretical $^{129}$Xe NMR spectrum of hp $^{129}$Xe and CryA in water reveals two peaks, one from xenon dissolved in water (Xe@water, approximately 191 ppm) and the other from xenon bound to CryA (Xe@CryA, approximately 60 ppm), both in reference to xenon gas (at 0 ppm). Due to the low occupancy of xenon molecules in CryA, the Xe@water peak is typically much more intense than the Xe@CryA peak.

to directly observe the Xe@CryA signal, neither of which are ideal for molecular imaging.[32,33] An alternative method for detection, with improved sensitivity, is based on chemical exchange between the difficult to detect CryA-encapsulated xenon pool (Xe@CryA) and the easy to detect pool of free xenon in solution (Xe@water).

## 6.4 Indirect Detection of Xenon by hyperCEST

The increased sensitivity of xenon detection gained through hyperpolarization alone is insufficient to achieve the subnanomolar threshold necessary for molecular imaging applications. One strategy to overcome this limitation is to amplify the weak signal intensity of the Xe@CryA pool by exploiting the temporary residence time of CryA-encapsulated xenon nuclei.[34] Xenon bound by CryA are in continuous exchange with the more abundant pool of dissolved xenon nuclei that are present in the surrounding aqueous environment (Figure 6-2A).[30,35] This property, in combination with the distinct chemical shifts representing the Xe@CryA and Xe@water pools, serves as the basis of an indirect detection scheme through chemical exchange saturation transfer (CEST). Specifically, the small Xe@CryA spin pool is saturated (i.e., depolarized) by frequency-selective radiofrequency (RF) pulses, and then transferred back to the bulk xenon spin pool through chemical exchange (Figure 6-3A). By applying RF pulses at the Xe@CryA chemical shift for a period that is long compared to the mean xenon residence time inside CryA (the inverse of the dissociation rate and greater than 30 ms or 5 ms at 25 °C and 37 °C, respectively), the CEST mechanism allows a single CryA cage to saturate the magnetization (i.e., depolarize) of hundreds of Xe@water nuclei. Consequently, as the depolarized Xe@CryA nuclei exchange back into the bulk xenon spin pool, the intensity of the Xe@water peak will diminish.

The integration of CEST with hp $^{129}$Xe, termed hyperCEST, was developed by the Pines lab in 2006.[36] The complementarity of hyperpolarization and CEST arises from the fact that the bulk spin pool is typically made up of a low concentration of highly polarized $^{129}$Xe spins as opposed to a high concentration of weakly polarized $^{1}$H spins (for proton-based CEST). Consequently, each xenon biosensor only has to saturate a modest number of Xe@water spins to elicit significant contrast. Furthermore, to minimize relaxation-based contrast loss, the hyperCEST detection scheme can be employed quickly relative to the long $T_1$ relaxation times of $^{129}$Xe (approximately 4 to 10 s in oxygenated blood at 1.5 T and 60 s in aqueous solvents at 4.7 T).[37,38]

Performing hyperCEST measurements is relatively straightforward. Following delivery of hp $^{129}$Xe to the sample,[31,32] two RF saturation pulses are applied: an on-resonant saturation pulse at the Xe@CryA frequency and an off-resonant pulse at a frequency equal to the separation between the Xe@water and Xe@CryA signal, but on the opposite side of the Xe@water signal (see Figure 6-3B). After each saturation pulse is applied, the Xe@water signal is recorded, and then the difference between the off- and on-resonant Xe@water signals is calculated; any differences between the two signals corresponds to the presence of a Xe@CryA pool. To generate a saturation response profile, this process is repeated using different on-resonant saturation pulses (and their corresponding off-resonant pulses), and the change in the Xe@water signal is recorded as a function of the offset frequency of the RF saturation pulses applies. The presence of xenon populations exchanging with Xe@water appear as a negative peak, indicating saturation (Figure 6-4). For an aqueous solution containing CryA, a region of saturation centered at

60 ppm is expected. Impressively, the hyperCEST method improved the detection sensitivity of CryA-encapsulated xenon into the submicromolar to nanomolar range, without the need for long acquisition times.[36,39]

## 6.5 Xenon Biosensors

A family of imaging agents that leverage xenon's unique physical properties and detection methods, generally called xenon biosensors, were developed by Pines and coworkers for both targeted imaging and biological sensing applications.[30,35] They consist of a xenon-binding host molecule, CryA, attached to one or more biomolecules, including targeting groups for local-

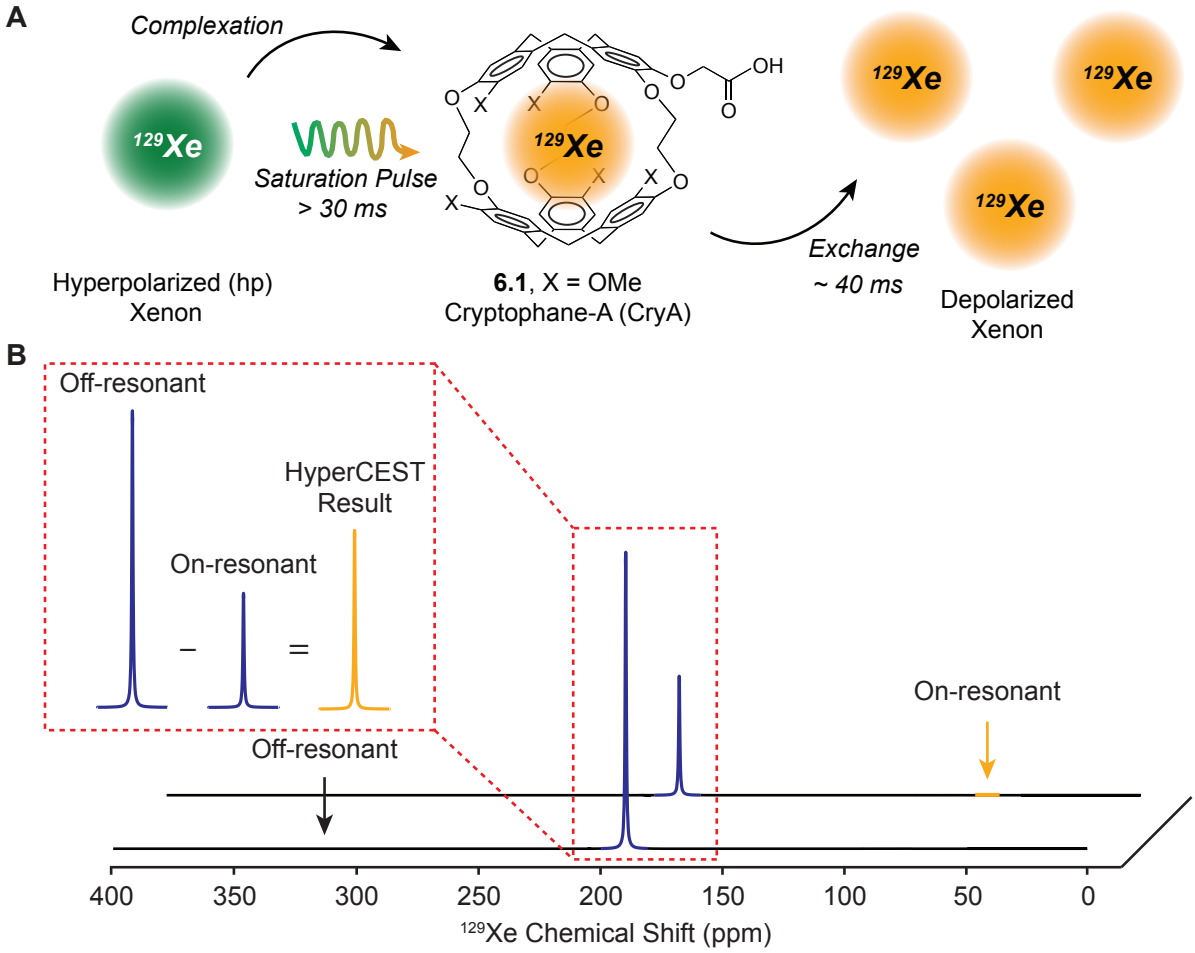



***Figure 6-3. Sensitivity-enhanced detection of xenon using hyperCEST.***

(A) A schematic depicting how hp $^{129}$Xe (depicted in green) can be selectively depolarized (depicted in orange) when bound by CryA by applying frequency-selective radiofrequency pulses (i.e., saturation pulse). Following chemical exchange with free hp $^{129}$Xe (green) from the bulk pool of xenon in solution, over time there will be an accumulation of depolarized $^{129}$Xe (orange). After several cycles of selective saturation, by measuring the difference between the initial and final bulk magnetization, an indirect, but sensitive, measurement of Xe@CryA is possible. (B) A schematic depicting how to measure hyperCEST contrast. HyperCEST experiments require two acquisitions to verify that a source of contrast is from CryA encapsulated xenon: an on-resonant saturation pulse at the Xe@CryA frequency, and an off-resonant pulse at a frequency equal to the separation between the Xe@water and Xe@CryA signal, but on the opposite side of the Xe@water signal. By measuring the Xe@water signal after each saturation pulse, any difference between the off- and on-resonant Xe@water signals corresponds to the presence of a Xe@CryA spin pool.

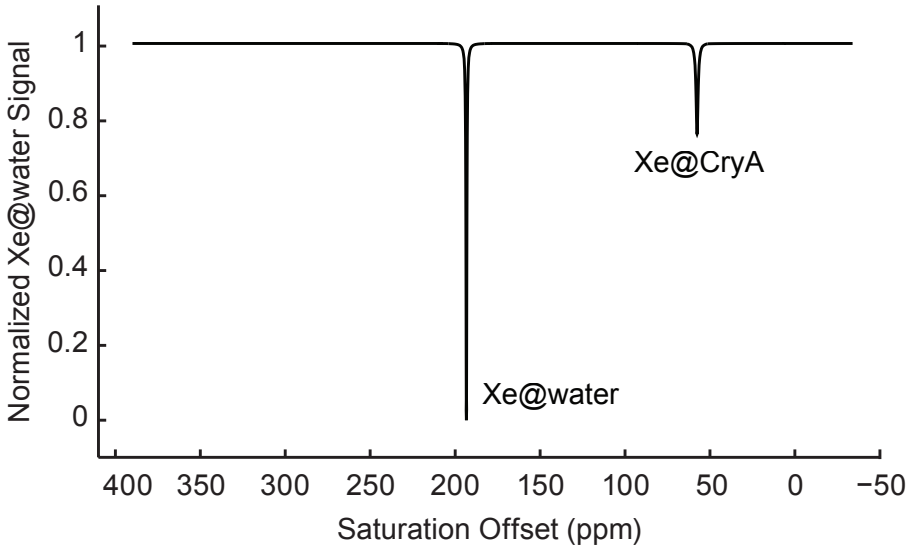



***Figure 6-4. The hyperCEST response can be plotted as a saturation response profile.***

An illustration of a theoretical saturation response profile, also known as a *Z-spectrum*, for a solution of hp $^{129}$Xe and CryA in water. Using the hyperCEST detection method, the presence of a xenon signal at a particular frequency (i.e., Xe@water or Xe@CryA) results in a negative peak, indicating saturation at that frequency.

ization or substrate molecules for sensing.[40] The first demonstration of this idea was with a biotinylated CryA construct (**6.2**, Biotin-CryA) that employed a peptide linker for aqueous solubility (see Figure 6-5A). Interestingly, upon binding to its target substrate, streptavidin agarose beads, several $^{129}$Xe peaks appeared in the region of the Xe@CryA chemical shift demonstrating xenon's ability to simultaneously sense free Biotin-CryA and streptavidin-bound CryA (Figures 6-5B and 6-5C).[30,35] Next, a series of peptide-based biotinylated xenon biosensors were developed to explore the relationship between the molecular composition of a biosensor and the characteristics of its protein-bound resonances.[32,36,41] This work not only helped improve the detection of xenon biosensors, but also established their compatibility with biomolecules. Subsequently, xenon biosensors were used in a number of sensing applications, including the detection of DNA hybridization,[42] enzymatic cleavage by matrix metalloproteinase-7,[43] ligand binding to human carbonic anhydrase[44] and an $\alpha_{2b}\beta_3$ integrin,[45] and peptide complex formation with a major histocompatibility complex protein.[46]

The next major milestone for xenon biosensors was achieved when Dmochowski and coworkers established their cellular compatibility.[45,47] Although $^{129}$Xe NMR experiments could not be performed directly with cells due to low signal sensitivity, they demonstrated cell viability after peptide-mediated uptake of crytophanes using CryA conjugated to a cell penetrating peptide. More recently, Berthault and coworkers performed $^{129}$Xe NMR measurements in the presence of live cells after targeting them with micromolar concentrations of a transferrin-functionalized biosensor.[48] This study detected biosensor binding to cells by directly measuring the Xe@CryA signal. Unfortunately, the combination of an insensitive direct detection method with a biosensor displaying low binding specificity are not ideal for molecular imaging applications. Accordingly, to transition xenon biosensors into targeted molecular MRI contrast agents,

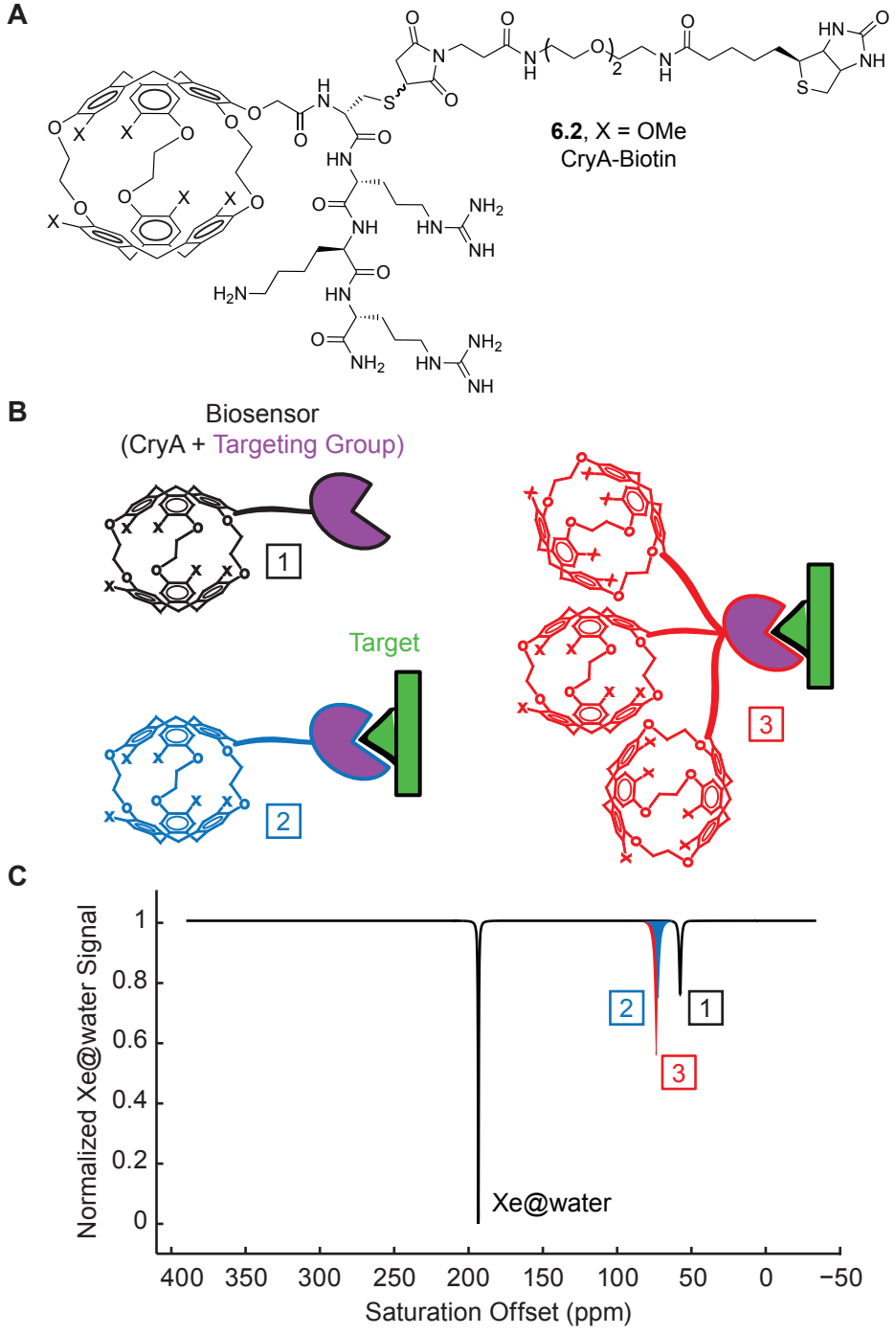



***Figure 6-5. Xenon biosensors are targetable, sensitive, and responsive MRI contrast agents.***

(A) The chemical structure of the first xenon biosensor (**6.2**), which contained CryA, a positively charged peptide for water solubility, a polyethylene glycol linker, and a biotin group for targeting. (B) Cartoons of a xenon biosensor (1), a biosensor bound to its target (2), and a multivalent biosensor bound to its target (3). (C) The theoretical saturation response profiles for each biosensor illustrated in B. While signal for the biosensor alone (1) appears at a region of saturation centered at 60 ppm, after binding its target there is a slight, but resolvable frequency shift (2). A multivalent biosensor, which contains multiple CryA cages, can be employed to increase the signal sensitivity of the target-bound biosensor (3).

biosensors with high specificity and detection sensitivity at subnanomolar concentrations are necessary.

## 6.6 Signal Amplification Through Scaffolding on Viral Nanoparticles

To both improve xenon detection sensitivity beyond that achievable with the hyperCEST method alone and utilize the targeting capability of xenon biosensors, we have turned to multivalent systems. A concept initially applied to both gadolinium- and CEST-based contrast agents for $^{1}$H-based MRI,[49-57] by increasing the payload of CryA cages per biosensor molecule, a multivalent approach will enable an increased local concentration of CryA cage at the site of biosensor binding. Accordingly, low concentrations of biosensor will correlate with much higher concentrations of CryA cages, allowing for improved detection sensitivity (Figures 6-5B and 6-5C, number 3).

We first demonstrated this scaffolding strategy for xenon biosensors using branched dendrimers.[58] However, these polyamidoamine (PAMAM) dendrimers could only encapsulate up to 11 CryA cages though electrostatic interactions. The low number of noncovalently bound CryA cages left room for improvement. Subsequently, we used viral nanoparticles—both spherical (see Chapter 7) and rod-like viral capsids (see Chapter 8)—to construct targetable biosensors, where the addition of hundreds of CryA cages onto a single biomolecule resulted in constructs that were detectable by hyperCEST at subpicomolar concentrations.[59,60] Most recently, we have combined the targeting capabilities of antibodies with the multivalent scaffolding potential of a viral capsid to generate a xenon biosensor that can differentiate between healthy and cancerous cell-types for live-cell imaging with $^{129}$Xe NMR (see Chapter 9).

## 6.7 Summary and Outlook

Chemical approaches have enabled tremendous growth in molecular imaging. In the case of $^{129}$Xe MRI, the combination of Cryptophane-A cages and biosensors has resulted in many successful targeted imaging and biological sensing applications. In the future, it will be exciting to apply these tools towards problems at the interface of human health and disease. In particular, many disease states are not characterized by a single biomarker, but rather a specific combination of many biomarkers. Accordingly, simultaneously visualizing multiple biomarkers, *in vivo*, would represent a significant achievement. And xenon's ability to respond to very subtle changes in its local environment coupled with its large chemical shift make it well suited for this application.

Clearly, there are many opportunities to contribute towards the development of MRI contrast agents. In particular, there is a bright future for xenon-based biosensors, and in the next several chapters I will describe a collection of viral capsid-based xenon biosensors that enable targeted molecular magnetic resonance imaging.

## 6.8 References

1. Albert M.S. *et al.* (**1994**) Biological Magnetic-Resonance-Imaging Using Laser Polarized Xe-129. *Nature* 370, 199-201.

2. Patz S. *et al.* (**2008**) Human pulmonary imaging and spectroscopy with hyperpolarized 129Xe at 0.2T. *Acad Radiol* 15, 713-727.

3. Mugler J.P. *et al.* (**1997**) MR imaging and spectroscopy using hyperpolarized Xe-129 gas: Preliminary human results. *Magnet Reson Med* 37, 809-815.

4. Kilian W., Seifert F. and Rinneberg H. (**2004**) Dynamic NMR spectroscopy of hyperpolarized Xe-129 in human brain analyzed by an uptake model. *Magnet Reson Med* 51, 843-847.

5. Ruppert K., Mata J.F., Brookeman J.R., Hagspiel K.D. and Mugler J.P., 3rd (**2004**) Exploring lung function with hyperpolarized (129)Xe nuclear magnetic resonance. *Magn Reson Med* 51, 676-687.

6. Venkatesh A.K. *et al.* (**2003**) MRI of the lung gas-space at very low-field using hyperpolarized noble gases. *Magn Reson Imaging* 21, 773-776.

7. Haynes W.M. (**2012**) CRC Handbook of Chemistry and Physics, Edn. 79th. (CRC Press, Boca Raton).

8. Oros A.M. and Shah N.J. (**2004**) Hyperpolarized xenon in NMR and MRI. *Physics in Medicine and Biology* 49, R105-153.

9. Schroder L. (**2011**) Xenon for NMR biosensing - Inert but alert. *Phys Med*.

10. Clever H.L. (**1979**) Krypton, Xenon and Radon: Gas Solubilities. (Pergamon Press, New York).

11. Ladefoged J. and Andersen A.M. (**1967**) Solubility of Xenon-133 at 37$deg$C in Water, Saline, Olive Oil, Liquid Paraffin, Solutions of Albumin, and Blood. *Physics in Medicine and Biology* 12, 353-358.

12. Chen R.Y. *et al.* (**1980**) Tissue-blood partition coefficient for xenon: temperature and hematocrit dependence. *J Appl Physiol* 49, 178-183.

13. Goodson B.M. (**2002**) Nuclear magnetic resonance of laser-polarized noble gases in molecules, materials, and organisms. *J Magn Reson* 155, 157-216.

14. Viale A. and Aime S. (**2010**) Current concepts on hyperpolarized molecules in MRI. *Curr Opin Chem Biol* 14, 90-96.

15. Walker T.G. and Happer W. (**1997**) Spin-exchange optical pumping of noble-gas nuclei. *Reviews of Modern Physics* 69, 629-642.

16. Ruset I.C., Ketel S. and Hersman F.W. (**2006**) Optical pumping system design for large production of hyperpolarized. *Phys Rev Lett* 96, 053002.

17. Berthault P., Huber G. and Desvaux H. (**2009**) Biosensing using laser-polarized xenon NMR/MRI. *Prog Nucl Mag Res Sp* 55, 35-60.

18. Wagshul M.E. *et al.* (**1996**) In vivo MR imaging and spectroscopy using hyperpolarized 129Xe. *Magn Reson Med* 36, 183-191.

19. Swanson S.D. *et al.* (**1997**) Brain MRI with laser-polarized 129Xe. *Magn Reson Med* 38, 695-698.

20. Goodson B.M. *et al.* (**1997**) In vivo NMR and MRI using injection delivery of laser-polarized xenon. *Proc Natl Acad Sci U S A* 94, 14725-14729.

21. Kawakami K. (**1997**) Topics in pulmonary nuclear medicine. *Ann Nucl Med* 11, 67-73.

22. Rubin S.M. *et al.* (**2001**) Detection of a conformational change in maltose binding protein by (129)Xe NMR spectroscopy. *J Am Chem Soc* 123, 8616-8617.

23. Boutin C. *et al.* (**2011**) Hyperpolarized 129Xe NMR signature of living biological cells. *NMR Biomed* 24, 1264-1269.

24. Gabard J. and Collet A. (**1981**) Synthesis of a (D3)-Bis(Cyclotriveratrylenyl) Macrocage by Stereospecific Replication of a (C3)-Subunit. *J Chem Soc Chem Comm*, 1137-1139.

25. Brotin T. and Dutasta J.P. (**2009**) Cryptophanes and their complexes--present and future. *Chem Rev* 109, 88-130.

26. Fogarty H.A. *et al.* (**2007**) A cryptophane core optimized for xenon encapsulation. *J Am Chem Soc* 129, 10332-10333.

27. Bartik K., Luhmer M., Dutasta J.-P., Collet A. and Reisse J. (**1998**) 129Xe and 1H NMR Study of the Reversible Trapping of Xenon by Cryptophane-A in Organic Solution. *J Am Chem Soc* 120, 784-791.

28. Brotin T. and Dutasta J.P. (**2003**) Xe@cryptophane complexes with C-2 symmetry: Synthesis and investigations by Xe-129 NMR of the consequences of the size of the host cavity for xenon encapsulation. *Eur J Org Chem*, 973-984.

29. Huber G. *et al.* (**2006**) Water soluble cryptophanes showing unprecedented affinity for xenon: candidates as NMR-based biosensors. *J Am Chem Soc* 128, 6239-6246.

30. Spence M.M. *et al.* (**2001**) Functionalized xenon as a biosensor. *Proc Natl Acad Sci U S A* 98, 10654-10657.

31. Han S.I. *et al.* (**2005**) NMR-based biosensing with optimized delivery of polarized 129Xe to solutions. *Anal Chem* 77, 4008-4012.

32. Hilty C., Lowery T.J., Wemmer D.E. and Pines A. (**2005**) Spectrally resolved magnetic resonance imaging of a xenon biosensor. *Angew Chem Int Ed Engl* 45, 70-73.

33. Lowery T.J. *et al.* (**2003**) Applications of laser-polarized 129Xe to biomolecular assays. *Magn Reson Imaging* 21, 1235-1239.

34. Ward K.M., Aletras A.H. and Balaban R.S. (**2000**) A new class of contrast agents for MRI based on proton chemical exchange dependent saturation transfer (CEST). *J Magn Reson* 143, 79-87.

35. Spence M.M. *et al.* (**2004**) Development of a functionalized xenon biosensor. *J Am Chem Soc* 126, 15287-15294.

36. Schroder L., Lowery T.J., Hilty C., Wemmer D.E. and Pines A. (**2006**) Molecular imaging using a targeted magnetic resonance hyperpolarized biosensor. *Science* 314, 446-449.

37. Venkatesh A.K., Zhao L., Balamore D., Jolesz F.A. and Albert M.S. (**2000**) Evaluation of carrier agents for hyperpolarized xenon MRI. *NMR Biomed* 13, 245-252.

38. Wolber J., Cherubini A., Dzik-Jurasz A.S., Leach M.O. and Bifone A. (**1999**) Spin-lattice relaxation of laser-polarized xenon in human blood. *Proc Natl Acad Sci U S A* 96, 3664-3669.

39. Schroder L. *et al.* (**2008**) Temperature response of 129Xe depolarization transfer and its application for ultrasensitive NMR detection. *Phys Rev Lett* 100, 257603.

40. Taratula O. and Dmochowski I.J. (**2010**) Functionalized 129Xe contrast agents for magnetic resonance imaging. *Curr Opin Chem Biol* 14, 97-104.

41. Lowery T.J. *et al.* (**2006**) Optimization of xenon biosensors for detection of protein interactions. *Chembiochem* 7, 65-73.

42. Roy V. *et al.* (**2007**) A cryptophane biosensor for the detection of specific nucleotide targets through xenon NMR spectroscopy. *Chemphyschem* 8, 2082-2085.

43. Wei Q. *et al.* (**2006**) Designing 129Xe NMR biosensors for matrix metalloproteinase detection. *J Am Chem Soc* 128, 13274-13283.

44. Chambers J.M. *et al.* (**2009**) Cryptophane xenon-129 nuclear magnetic resonance biosensors targeting human carbonic anhydrase. *J Am Chem Soc* 131, 563-569.

45. Seward G.K., Bai Y., Khan N.S. and Dmochowski I.J. (**2011**) Cell-compatible, integrin-targeted cryptophane-129Xe NMR biosensors. *Chemical Science* 2, 1103.

46. Schlundt A. *et al.* (**2009**) A xenon-129 biosensor for monitoring MHC-peptide interactions. *Angew Chem Int Ed Engl* 48, 4142-4145.

47. Seward G.K., Wei Q. and Dmochowski I.J. (**2008**) Peptide-mediated cellular uptake of cryptophane. *Bioconjug Chem* 19, 2129-2135.

48. Boutin C. *et al.* (**2011**) Cell uptake of a biosensor detected by hyperpolarized 129Xe NMR: the transferrin case. *Bioorg Med Chem* 19, 4135-4143.

49. Datta A. *et al.* (**2008**) High relaxivity gadolinium hydroxypyridonate-viral capsid conjugates: nanosized MRI contrast agents. *J Am Chem Soc* 130, 2546-2552.

50. Hooker J.M., Datta A., Botta M., Raymond K.N. and Francis M.B. (**2007**) Magnetic resonance contrast agents from viral capsid shells: a comparison of exterior and interior cargo strategies. *Nano Lett* 7, 2207-2210.

51. Vasalatiy O., Gerard R.D., Zhao P., Sun X. and Sherry A.D. (**2008**) Labeling of adenovirus particles with PARACEST agents. *Bioconjug Chem* 19, 598-606.

52. Liepold L. *et al.* (**2007**) Viral capsids as MRI contrast agents. *Magn Reson Med* 58, 871-879.

53. Allen M. *et al.* (**2005**) Paramagnetic viral nanoparticles as potential high-relaxivity magnetic resonance contrast agents. *Magn Reson Med* 54, 807-812.

54. Anderson E.A. *et al.* (**2006**) Viral nanoparticles donning a paramagnetic coat: conjugation of MRI contrast agents to the MS2 capsid. *Nano Lett* 6, 1160-1164.

55. Prasuhn D.E., Jr., Yeh R.M., Obenaus A., Manchester M. and Finn M.G. (**2007**) Viral MRI contrast agents: coordination of Gd by native virions and attachment of Gd complexes by azide-alkyne cycloaddition. *Chem Commun (Camb)*, 1269-1271.

56. Garimella P.D., Datta A., Romanini D.W., Raymond K.N. and Francis M.B. (**2011**) Multivalent, high-relaxivity MRI contrast agents using rigid cysteine-reactive gadolinium complexes. *J Am Chem Soc* 133, 14704-14709.

57. Qazi S. *et al.* (**2012**) P22 Viral Capsids as Nanocomposite High-Relaxivity MRI Contrast

Agents. *Mol Pharm*.

58. Mynar J.L., Lowery T.J., Wemmer D.E., Pines A. and Frechet J.M. (**2006**) Xenon biosensor amplification via dendrimer-cage supramolecular constructs. *J Am Chem Soc* 128, 6334-6335.

59. Meldrum T. *et al.* (**2010**) A xenon-based molecular sensor assembled on an MS2 viral capsid scaffold. *J Am Chem Soc* 132, 5936-5937.

60. Stevens T.K. *et al.* (**2012**) HyperCEST detection of a (129) Xe-based contrast agent composed of cryptophane-A molecular cages on a bacteriophage scaffold. *Magn Reson Med*.

# Chapter 7: A $^{129}$Xe-based MRI Contrast Agent Assembled on the MS2 Viral Capsid

## 7.1 Introduction

Sensitivity improvement of targeted contrast agents is crucial for progress in molecular imaging. In particular, the primary contrast agents for magnetic resonance imaging (MRI) either modulate magnetic relaxation (e.g., gadolinium-based contrast agents)[1] or reduce bulk magnetization via exchange of saturated nuclei from a magnetically distinct site (e.g., chemical exchange saturation transfer (CEST)-based agents).[2,3] While most contrast agents manipulate strong proton signals from abundant water and fat, there are advantages to using exogenous nuclei. In particular, xenon is inert, relaxes slowly, can be reversibly dissolved in solution, and is amenable to hyperpolarization by optical pumping.[4-6]

Recently developed MRI contrast agents called xenon biosensors[7] have demonstrated the ability to utilize the complementary sensitivity enhancements of both hyperpolarized (hp) $^{129}$Xe and CEST detection.[8] Xenon biosensors employ cryptophane-A molecular cages (CryA)[9] to reversibly bind xenon atoms that have been dissolved into environments such as aqueous tissue. To improve the detection sensitivity of xenon-based contrast agents further, multiple copies of

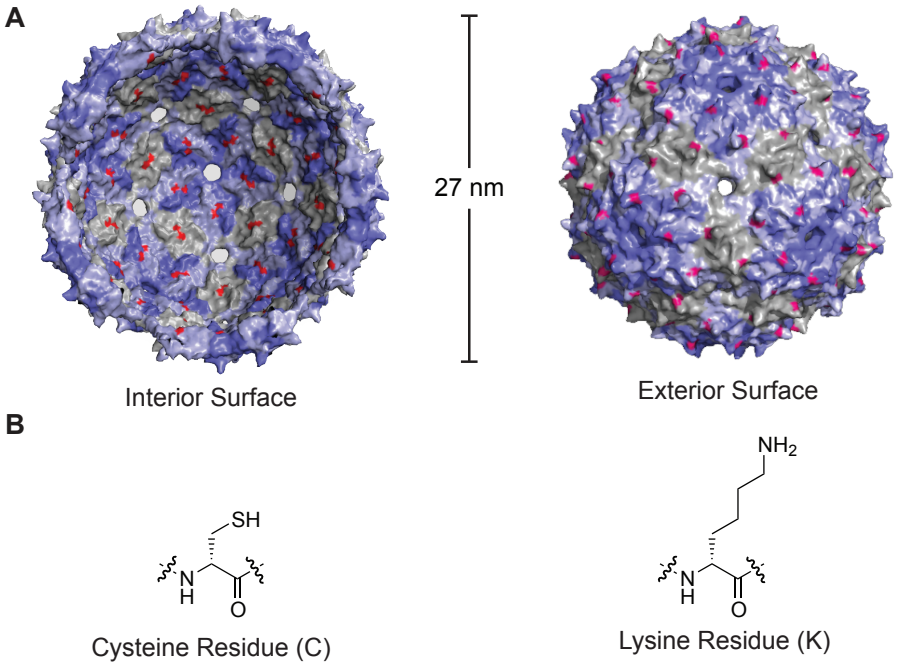



***Figure 7-1. Crystal structure renderings of bacteriophage MS2.***

(A) The interior and exterior surfaces of bacteriophage MS2, a hollow and porous spherical viral capsid with icosahedral symmetry, are shown. The diameter of the capsid is 27 nm and there are 32 pores, each approximately 2 nm in diameter. (B) Chemical structures of a cysteine or lysine residue for orthogonally targeting the interior or exterior surfaces of MS2, respectively, for chemical functionalization.

Tyler Meldrum, Kristen Seim, and Wesley Wu contributed to the work presented in this chapter.
Portions of the work described in this chapter have been reported in a separate publication.[34]

CryA have been assembled onto supramolecular scaffolds, such as a branched dendrimer.[10] In this Chapter, we extend this scaffolding concept to biological nanoparticles, and present the first example of viral capsids functionalized as $^{129}$Xe-based MRI contrast agents.

This was accomplished by transforming a spherical viral capsid, bacteriophage MS2, into a multivalent xenon biosensor. The MS2 viral capsid, composed of 180 identical monomers, is a hollow and porous structure with a diameter of 27 nm (Figure 7-1).[11] After expression in *E. coli*, the coat protein monomers spontaneously assemble into genome-free, noninfectious capsids with icosahedral symmetry.[12,13] The assembled capsids contain 32 pores, each approximately 2 nm wide, which facilitate access to the capsid interior without disassembly. Through a number of bioconjugation strategies, both the interior and exterior surfaces of the capsid can be selectively modified to a high degree, enabling the simultaneous integration of many contrast agent molecules in the interior and cell-specific targeting units on the exterior of a single capsid molecule.[14-16]

## 7.2 Results and Discussion

By targeting specific amino acid residues of the MS2 coat protein for covalent modification, it is possible to orthogonally functionalize the interior and exterior surfaces of the MS2 virus. For example, we previously demonstrated the attachment of gadolinium chelates,[17-19] positron emission tomography (PET) radioisotopes,[20,21] drug molecules,[22] and fluorescent dyes[13,23] to the interior of intact capsids. Additionally, we have installed peptide-,[12] polymer-,[23] and aptamer-based[13] groups on the exterior of intact capsids. Furthermore, characterization of MS2 viral capsids labeled with radioisotopes in both mice and rats indicate they are biocompatible, with broad tissue distribution, long circulation times, and minimal toxicity.[20,21] Taken together, dual modification of MS2 viral capsids with CryA cages on the interior and targeting groups on the exterior could provide access to targetable contrast agents for xenon-based MRI applications *in vivo*.

### *7.2.1 Construction of an MS2-CryA biosensor through interior capsid modification*

We chose to attach CryA cages to the interior surface of the MS2 capsid, leaving the exterior surface free for future functionalization with targeting groups. The strategy for interior modification was based on cysteine alkylation. Although wild-type (wt) MS2 coat proteins contain two cysteine residues (C46 and C101) near the exterior surface of the capsid, they have

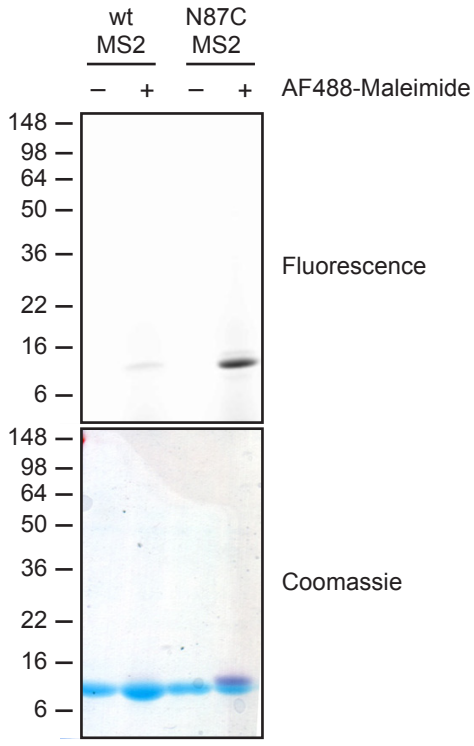
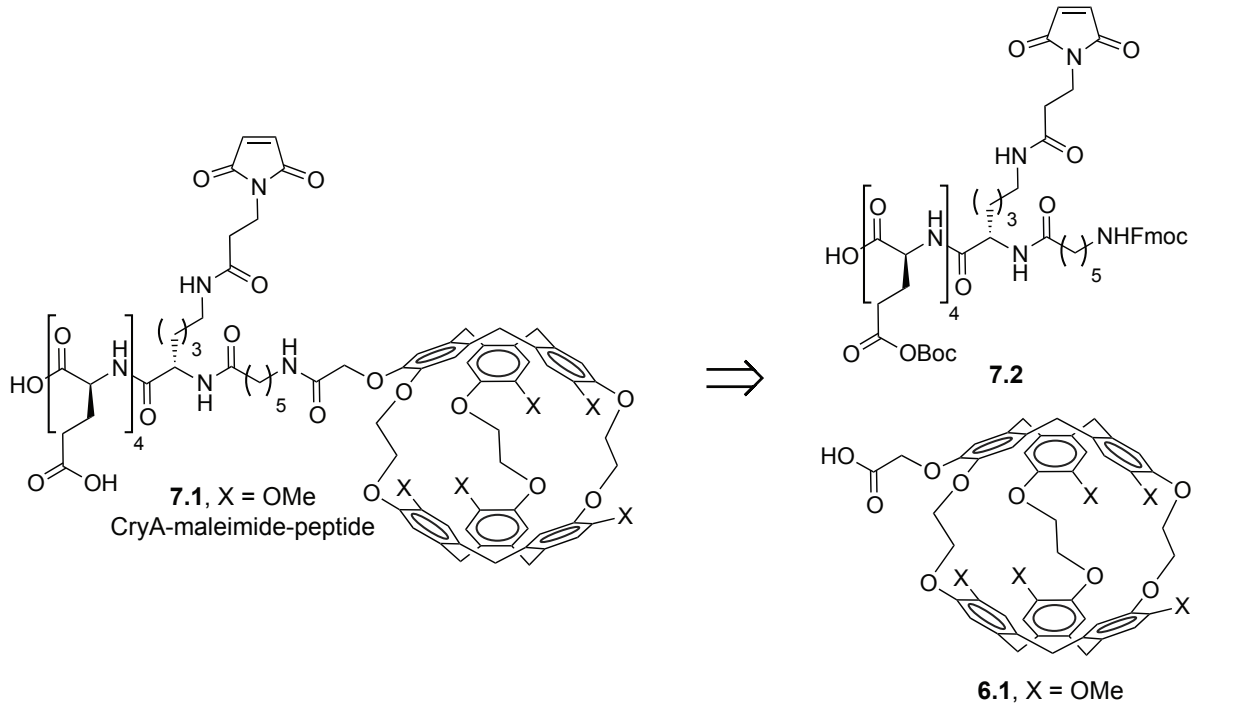



***Figure 7-2. The interior surface of MS2 capsids was modified after introduction of a cysteine residue.***

AF488-maleimide labeled N87C MS2 more efficiently than wt MS2, indicating that the reaction is dependent on a solvent exposed cysteine residue. Protein samples were analyzed by SDS-PAGE followed by visualization using fluorescence imaging (top) and Coomassie staining (bottom).

*Scheme 7-1. Retrosynthesis of CryA-maleimide peptide 7.1.*



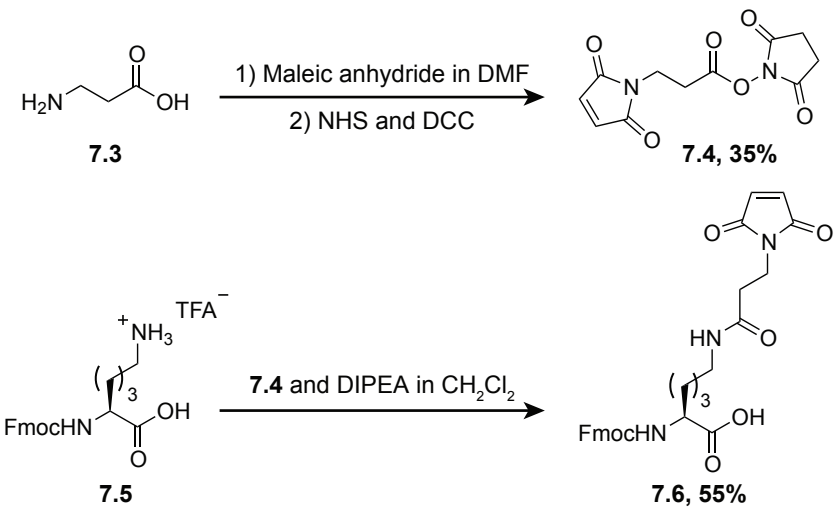

limited reactivity due to poor solvent accessibility. Therefore, an optimally located asparagine residue on the capsid coat protein was mutated to a cysteine residue (N87C).[13] The reactivity of the cysteine mutant is illustrated in Figure 7-2, where N87C MS2, but not wild-type (wt) MS2, became fluorescently labeled in the presence of Alexa Fluor 488 maleimide (AF488-maleimide), indicating that the reaction is dependent on a solvent exposed cysteine residue.

In order to attach CryA cages to cysteine residues on MS2 coat proteins, it was necessary to solubilize CryA in aqueous solutions (for compatibility with protein bioconjugation) and functionalize it with a cysteine specific reactive group. Based on the design of early xenon biosensors, which used peptide linkers to solubilize and functionalize CryA cages (i.e., compound **6.1**),[7,24] our first target was maleimide-functionalized CryA-peptide **7.1** (Scheme 7-1). We envisioned that peptide **7.1** could be obtained from CryA (**6.1**) and a maleimide-functionalized peptide **7.2**. Peptide **7.2** could then be prepared using standard solid phase peptide synthesis (SPPS) with a maleimide-functionalized lysine residue (**7.6**).

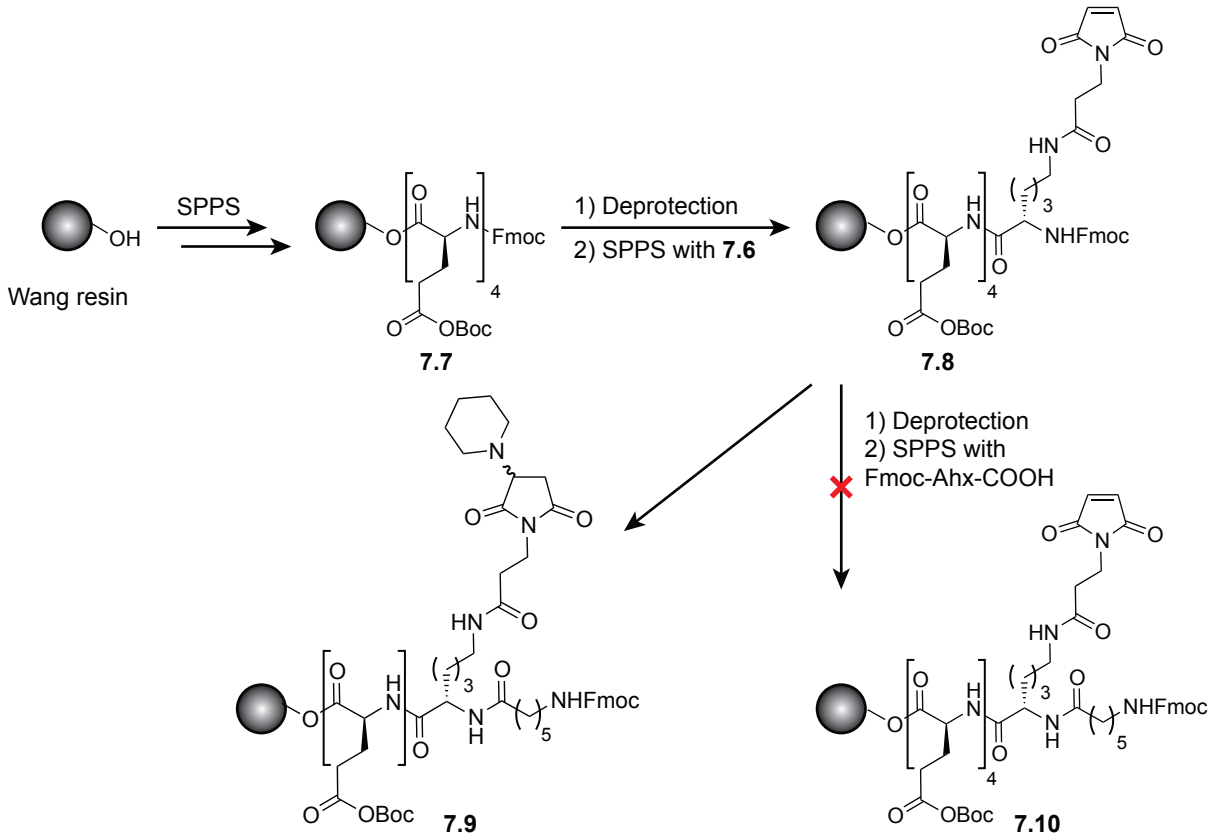
The synthesis of lysine derivative **7.6** proceeded as expected with moderate yields (Scheme 7-2). However, the presence of an unprotected maleimide group on the amino acid analogue proved problematic during standard SPPS reaction conditions (Scheme 7-3). Specifically, piperidine used in the deprotection step to remove an *N*-terminal Fluorenylmethoxycarbonate (Fmoc) protecting group also reacted with the maleimide group, producing peptide **7.9** instead of the desired peptide product **7.10**. Other methods for Fmoc removal were also problematic.

At the same time that the maleimide-functionalized CryA-peptide (**7.1)** was being synthesized, another gradate student synthesized a water-solubilizing small molecule (**7.11**) for conjugating the hydrophobic molecule Taxol to cysteine residues in MS2.[22] This linker contained a

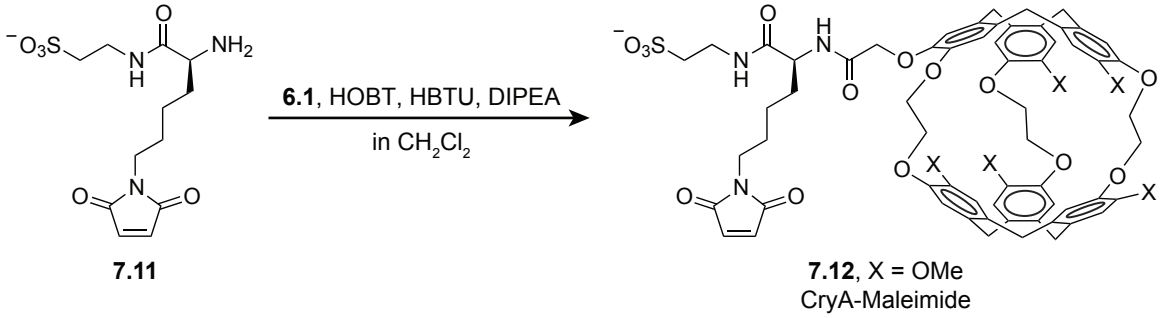
*Scheme 7-2. Synthesis of maleimide-functionalized lysine derivative 7.6.*




*Scheme 7-3. Attempted synthesis of CryA-maleimide peptide 7.1.*



*Scheme 7-4. Synthesis of CryA-maleimide 7.12.*




maleimide group to react with cysteine residues, taurine to increase aqueous solubility, and a free amine to react with a carboxylic acid on the hydrophobic molecule. Using compound **7.11**, maleimide-functionalized CryA (**7.12**, CryA-maleimide) was synthetized using standard amide-bond forming conditions (Scheme 7-4) and showed sufficient aqueous solubility for bioconjugation reactions.

To construct an MS2-based xenon biosensor, CryA-maleimide **7.12** was reacted with N87C MS2 for 4 h at RT in phosphate buffer (Figure 7-3A). The resulting MS2-CryA biosensor was purified by size-exclusion chromatography, which confirmed that the viral capsids were fully

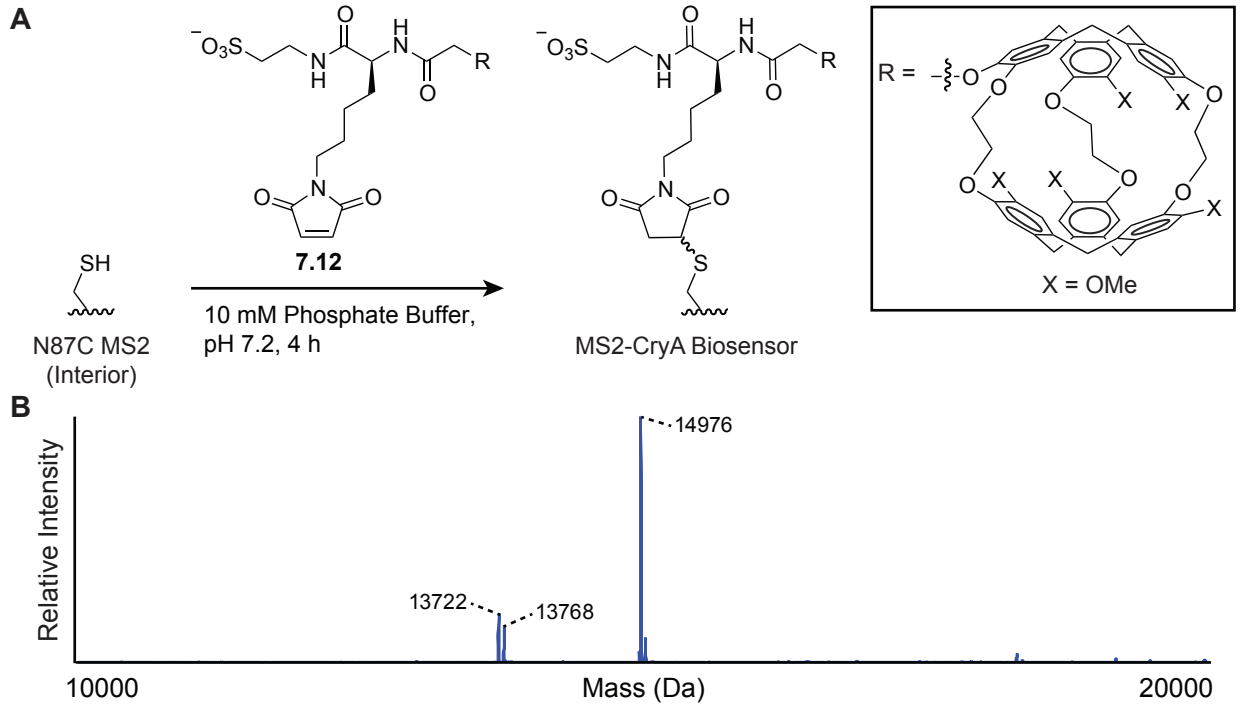



**Figure 7-3. Construction and characterization of an MS2-CryA biosensor.**

(A) The MS2-CryA biosensor was constructed by alkylating a single cysteine residue (N87C MS2) introduced to the capsid interior with a maleimide-functionalized cryptophane-A cage derivative (**7.12**, CryA-Maleimide). (B) Characterization of MS2-CryA by electrospray ionization mass spectrometry of MS2 monomer proteins. Expected mass of the N87C MS2 monomer $[M+H]^+$ = 13719 Da. Expected mass of MS2-CryA labeled monomer $[M+H]^+$ = 14973 Da. Integration of these peaks estimated a 70% conversion, corresponding to about 125 copies of CryA cages per biosensor.

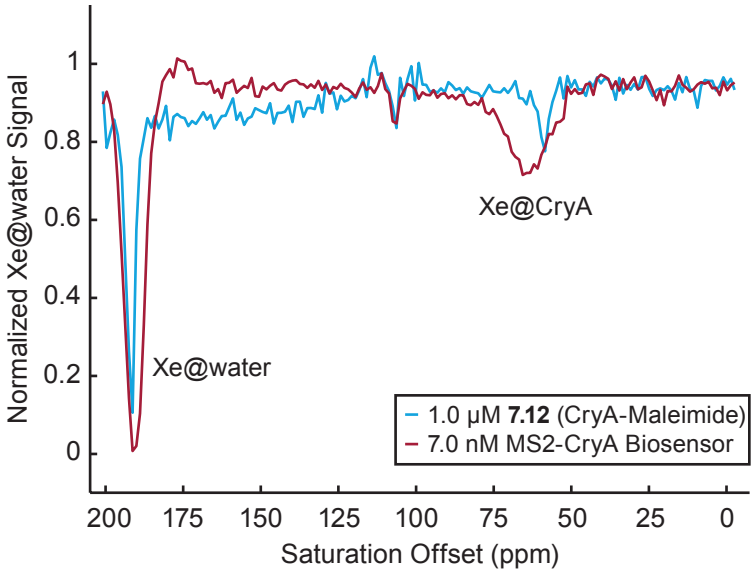



***Figure 7-4. The hyperCEST response profile of the MS2-CryA biosensor was similar to a concentration matched sample of unscaffolded CryA-Maleimide.***

Comparison of the saturation response profile for a solution of 7.0 nM MS2-CryA biosensor (1.0 μM in CryA) stabilized with 5% isopropyl alcohol and a CryA–concentration matched solution of unscaffolded CryA-Maleimide (**7.12**, 1.0 μM) revealed similar levels of contrast. Shown are the saturation peaks corresponding to the Xe@CryA chemical shift at approximately 60 ppm and the Xe@water chemical shift at 192 ppm (referenced to the Xe gas chemical shift).

assembled, and then characterized by both reversed-phase HPLC (not shown) and electrospray ionization mass spectrometry (Figure 7-3B). While the MS2 monomer protein was modified with a single CryA cage, the assembled capsid was approximately 70% modified, corresponding to approximately 125 copies of CryA cages per biosensor.

### *7.2.2 Characterization of the MS2-CryA biosensor by $^{129}$Xe NMR*

Xenon gas was introduced into a solution of xenon biosensors in phosphate buffer using a continuous-flow bubbling system (as illustrated in Figure 8-3A).[25] The output from a xenon polarizer (spin exchange optical pumping) was connected to a 5-mm NMR tube modified with inlet and outlet arms, allowing for direct delivery of hp $^{129}$Xe into the solution via a capillary running to the bottom of the NMR tube. While this method facilitated the rapid dissolution of xenon gas, bubbling through a solution of capsids could introduce complications, including foaming and the mechanical stress of bubbles on capsid integrity. However, we found that the addition of a small amount of isopropyl alcohol (5% v/v) to biosensor solutions minimized foaming and maintained MS2-CryA biosensor stability (as verified by size exclusion chromatography).

$^{129}$Xe NMR measurements were performed using a 7 nM solution of MS2-CryA biosensor in phosphate buffer containing 5% v/v isopropyl alcohol. This corresponded to approximately 1 μM CryA, so for comparison, a 1 μM solution of unscaffolded CryA-maleimide (**7.12**) was also prepared. The saturation response profile of each sample was recorded by measuring the Xe@water signal as a function of the offset frequency of a radiofrequency (RF) saturation pulse train. This revealed the presence of any xenon populations exchanging with Xe@water (Figure 7-4). As

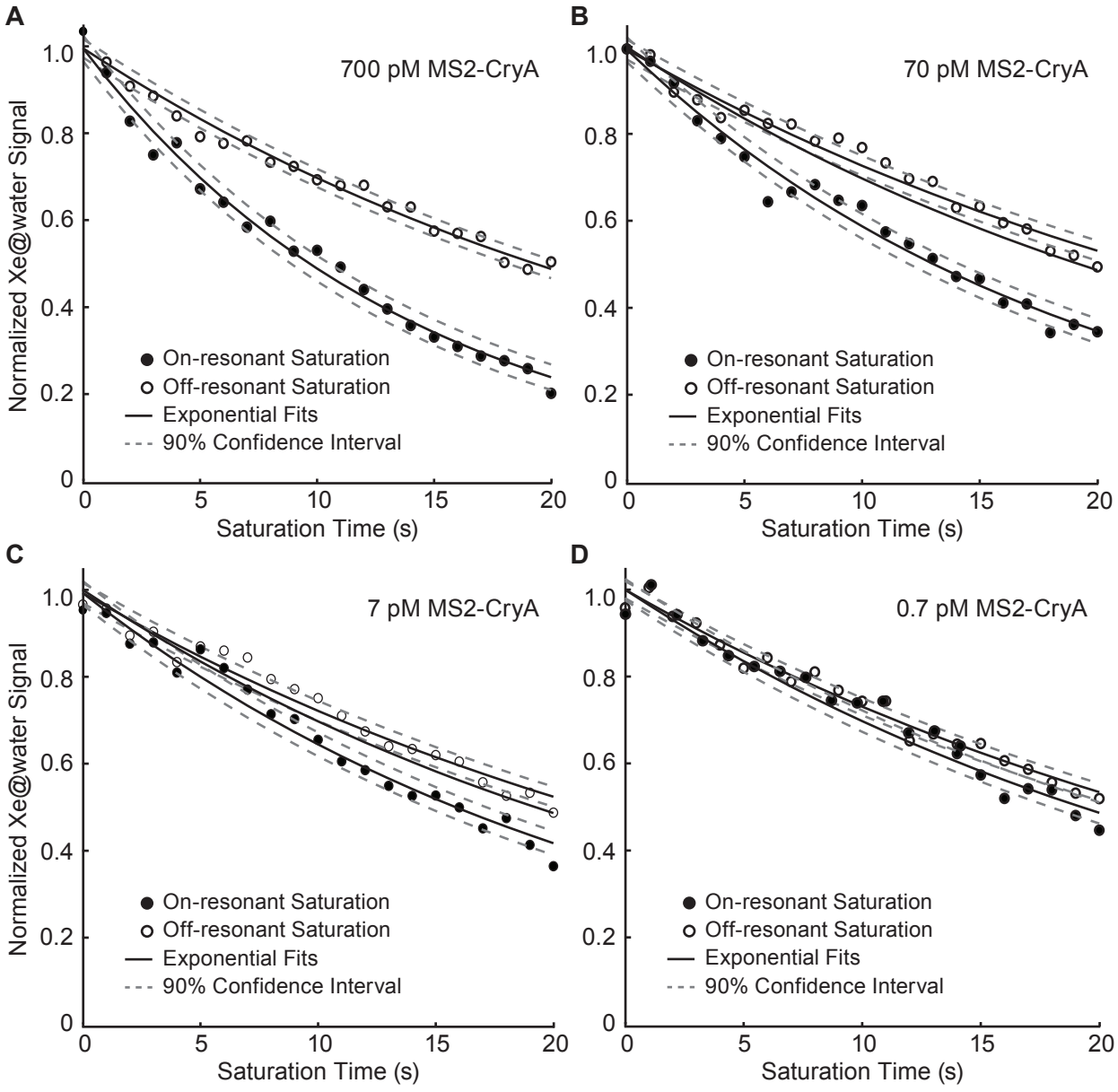
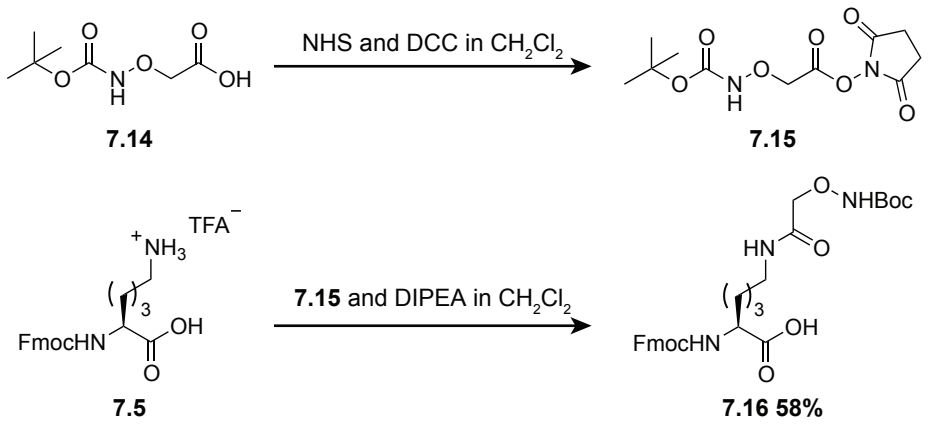



***Figure 7-5. hyperCEST contrast was measured as a function of saturation time and biosensor concentration.***

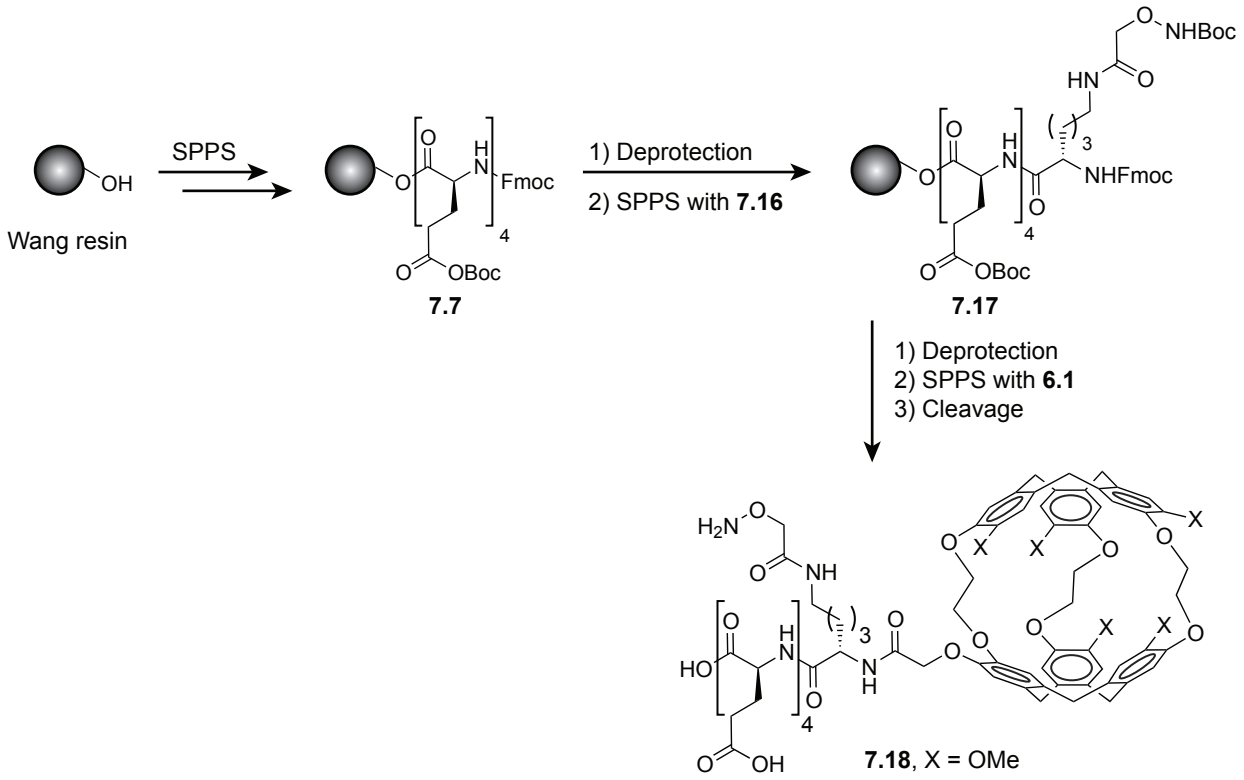
The hyperCEST contrast for a dilution series of the MS2-CryA biosensor was measured as a function of saturation time: (A) 700 pM MS2-CryA, (B) 70 pM MS2-CryA, (C) 7 pM MS2-CryA, and (D) 0.7 pM MS2-CryA. Saturation was applied using a train of 2500 to 5000 Hz bandwidth dSNOB pulses, and the normalized Xe@water signal for both the on- and off-resonant pulses were plotted. Decay in signal intensity in the off-resonant saturation profiles was visible because experiments were not corrected for relaxation. Exponential fits are given as solid lines and the 90% confidence interval as dashed lines.

expected, a region of saturation centered at 60 ppm, characteristic of the Xe@CryA spin pool, was seen for both the MS2-CryA biosensor and the unscaffolded CryA-maleimide samples. Despite a slight downfield shift and broadness in the Xe@CryA response of the MS2-CryA biosensor, its contrast was in good agreement with that of the concentration-matched unscaffolded CryA-maleimide sample and it was still possible to achieve strong hyperCEST effects by using saturation pulses with intentionally large bandwidths.

***Scheme 7-5. Synthesis of aminooxy-functionalized lysine derivative 7.16.***




***Scheme 7-6. Synthesis of aminooxy-functionalized CryA peptide 7.18.***



The hyperCEST contrast of the biosensor solution was then measured as a function of saturation time and biosensor concentration to determine the detection threshold (Figure 7-5). Starting with a 700 pM solution of MS2-CryA, dilutions to 70 pM, 7 pM, and 0.7 pM were prepared in phosphate buffer. The hyperCEST contrast of these solutions were quantified by comparing the Xe@water signal after saturating at 126.9 ppm upfield from the Xe@water chemical shift to that of a symmetrically placed control RF pulse train applied 126.9 ppm downfield from Xe@water, for incremental saturation times. CEST contrast was visible in the saturation time curves for all four concentrations of the MS2-CryA biosensor, and the decrease in residual signal for each concentration was accurately represented by an exponential decay. Impressively, the 0.7 pM MS2-CryA solution represents one of the lowest detectable concentrations of a xenon biosensor, with a mean hyperCEST contrast of 8.9 ± 3.4%. The increased sensitivity over a recently reported 10 nM biosensor solution (Schroder *et al.*[26]) was likely due to the incorporation of approximately 125 CryA cages per biosensor molecule and to various technical improvements, including enhanced saturation transfer methods applicable to systems with broad NMR line widths.[27]

To further characterize the MS2-based biosensor platform, we wanted to explore the kinetics of both xenon exchange and saturation, properties that might be affected by the geometry of the MS2 capsid and that could influence saturation transfer-based detection tech-

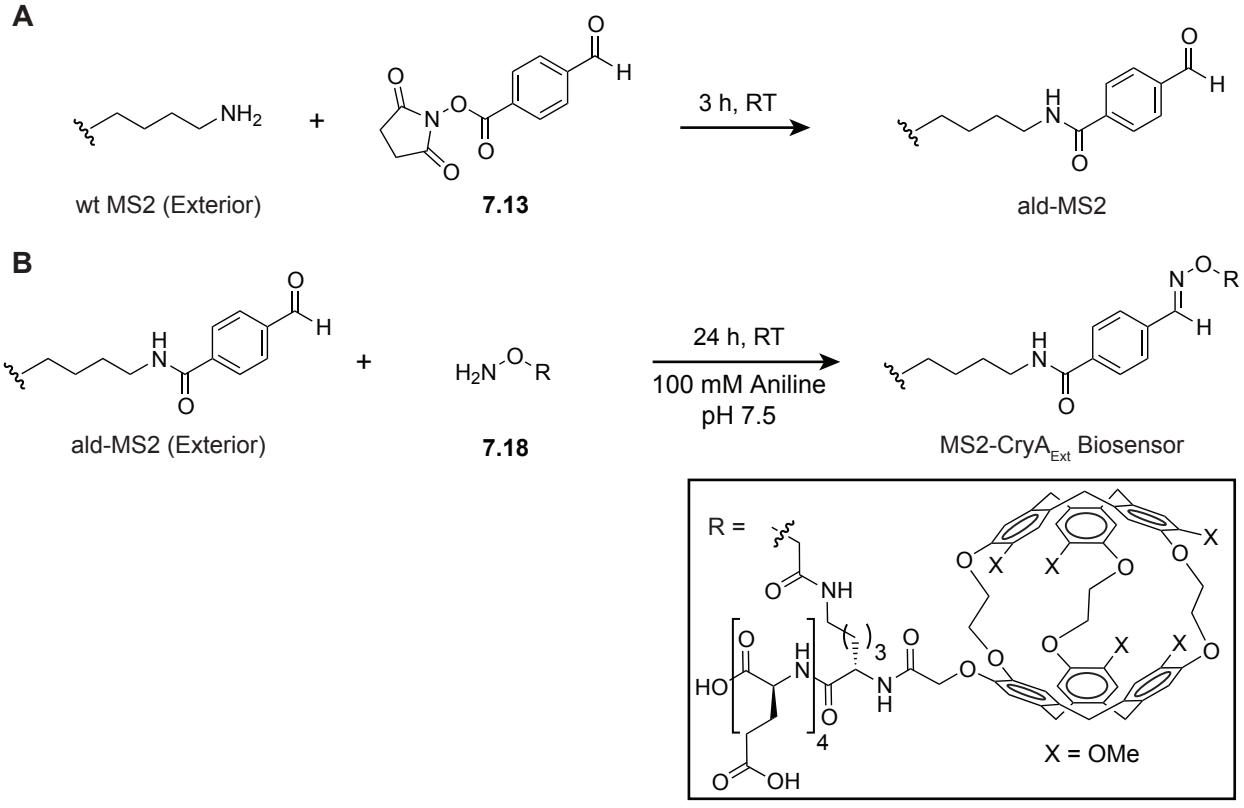



***Figure 7-6. Construction of an externally modified MS2-CryA$_{Ext}$ biosensor***

Lysine residues on the exterior of wt MS2 capsid proteins were modified with CryA cages through two sequential bioconjugation reactions. (A) First, lysine residues were converted to aldehydes using the activated-ester **7.13**. (B) Next, the aldehyde-functionalized capsids (ald-MS2) were reacted with an aminooxy-functionalized CryA peptide (**7.18**) to yield the MS2-CryA$_{Ext}$ biosensor.

niques. Towards this goal, we constructed an MS2-CryA biosensor where the external surface was modified with CryA (i.e., MS2-CryA$_{Ext}$) for comparison to the internally functionalized MS2 biosensor.

### *7.2.3 Construction of an MS2-CryA biosensor through exterior capsid modification*

The strategy for modification of the exterior surface of MS2 viral capsids was based on lysine acylation. Although the wt MS2 coat protein contains three lysine residues between both the interior and exterior surfaces of the capsid, the two residues located on the exterior surface (K106 and K113) are more solvent accessible. These lysine residues were modified with CryA cages through two sequential bioconjuation reactions. First, lysine residues were converted to aldehydes using the activated-ester **7.13** as illustrated in Figure 7-6A.[18,23] Characterization of the aldehyde-functionalized MS2 capsids (ald-MS2) by mass spectrometry revealed that most coat proteins contained up to two aldehydes with a minor population containing three aldehydes (Figure 7-7A).

Next, ald-MS2 was reacted with an aminooxy-functionalized CryA peptide (**7.18**) to form a stable oxime product (Figure 7-6B). Synthesis of peptide **7.18** was approached analogously to the maleimide-functionalized CryA peptide **7.1**. Briefly, the aminooxy-functionalized lysine residue **7.16** was first synthesized according to Scheme 7-5. Next, the aminooxy-functionalized CryA peptide **7.18** was accessed using SPPS (Scheme 7-6).

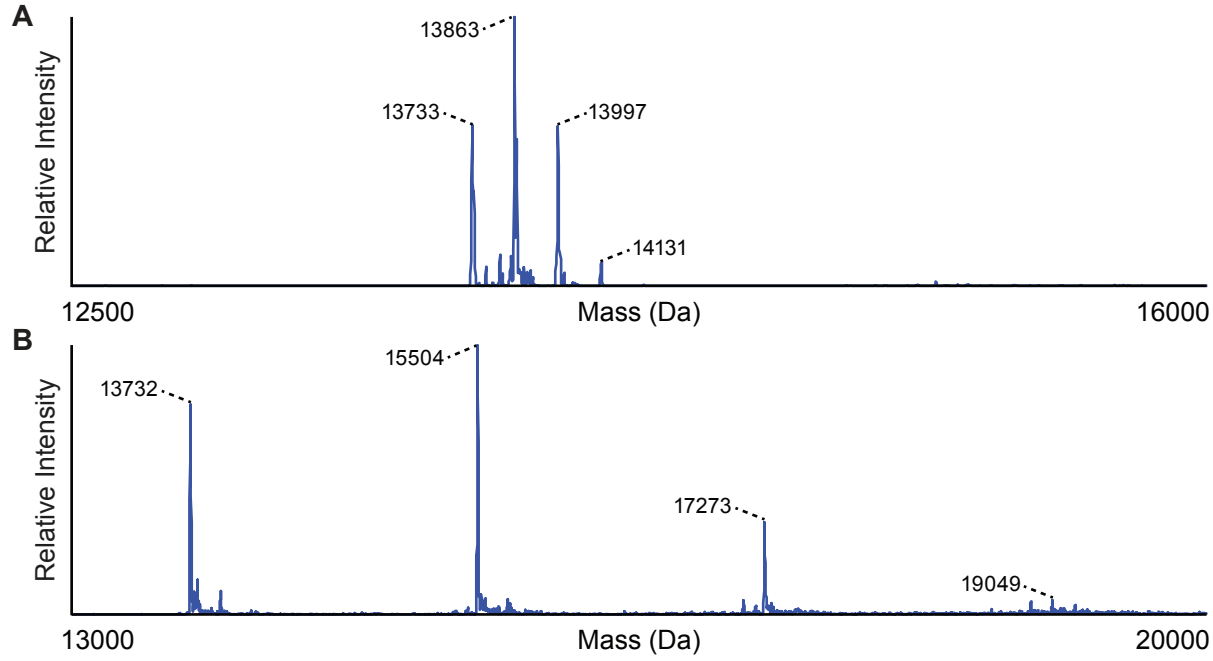



***Figure 7-7. Characterization of ald-MS2 and the MS2-CryA$_{Ext}$ biosensor by mass spectrometry.***

(A) Characterization of aldehyde-functionalized MS2 capsids by electrospray ionization mass spectrometry of MS2 monomer proteins. Expected mass of the unmodified MS2 monomer $[M+H]^+$ = 13732 Da. Expected mass of one-, two-, or three-aldehyde containing MS2 monomers $[M+H]^+$ = 13864 Da, 13996 Da, and 14128 Da, respectively. (B) Characterization of MS2-CryA$_{Ext}$ by electrospray ionization mass spectrometry of MS2 monomer proteins. Expected mass of the unmodified MS2 monomer $[M+H]^+$ = 13732 Da. Expected mass of one-, two-, or three-CryA containing MS2 monomers $[M+H]^+$ = 15502 Da, 17272 Da, and 19042 Da, respectively.

Upon completion of the two bioconjugation reactions with MS2 viral capsids, the resulting MS2-Cry$A_{Ext}$ biosensor was purified by size-exclusion chromatography, which confirmed that the viral capsids were fully assembled, and then characterized by both reversed-phase HPLC (not shown) and electrospray ionization mass spectrometry (Figure 7-7B). On average, each MS2 monomer protein was modified with either one or two CryA cages, with a minor population containing three cages. The existence of three populations of MS2 monomer proteins modified with 1, 2 or 3 CryA cages was verified and quantified using SDS-PAGE and optical densitometry analysis (Figure 7-8). On average, MS2 capsids were approximately 72% modified, corresponding to approximately 130 copies of CryA cages per biosensor.

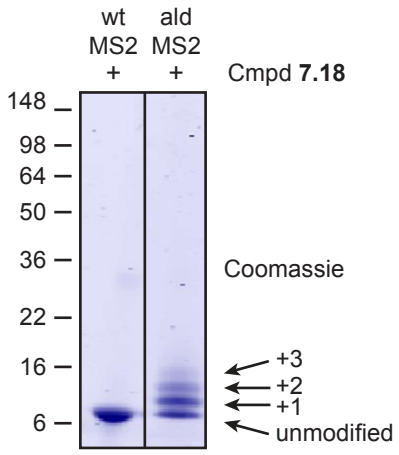



***Figure 7-8. Characterization of the MS2-CryA$_{Ext}$ biosensor by SDS-PAGE and densitometry.***
Only aldehyde-functionalized MS2 were labeled with the aminooxy-functionalized CryA peptide (**7.18**). Up to three modifications, corresponding to either one-, two-, or three-CryA containing MS2 monomers, were observed (labeled as +1, +2, or +3, respectively). Using optical densitometry, MS2 capsids were approximately 72% modified, corresponding to approximately 130 copies of CryA cages per biosensor.

### *7.2.4 Characterization of the MS2-CryA$_{Ext}$ biosensor by $^{129}$Xe NMR*

To evaluate the effect of the geometry of the MS2 viral capsid on saturation transfer-based detection of hp $^{129}$Xe, we compared $^{129}$Xe NMR measurements between CryA-concentration matched solutions of the MS2-CryA and MS2-Cry$A_{Ext}$ biosensors. Specifically, NMR measurements were performed using a 7 nM solution of both the MS2-CryA and MS2-Cry$A_{Ext}$ biosensors in phosphate buffer containing 5% v/v isopropyl alcohol. The saturation response profile of each sample was recorded by measuring the Xe@water signal as a function of the offset frequency of a radiofrequency (RF) saturation pulse train. This revealed the presence of any xenon populations exchanging with Xe@water (Figure 7-9). As expected, a broad region of saturation centered at 60 ppm, characteristic of the Xe@CryA spin pool, was seen for both the MS2-based biosensors. However, the increased signal from the externally modified sample suggests that xenon exchange through the pores of the MS2 viral capsid may be slightly retarded. Similarly, when the hyperCEST contrasts of both MS2-based biosensors were measured as a function of saturation time, a similar response was observed, with a slight sensitivity advantage to the MS2-Cry$A_{Ext}$ biosensor (Figure 7-9B). It is likely that while the 2 nm pores of MS2 viral capsids allowed for sufficiently fast exchange of xenon nuclei, such that depolarized xenon nuclei located within the capsid could exchange with the bulk pool of hyperpolarized xenon located outside the capsid, a small bottleneck is created. However, more experiments are necessary in order to resolve this issue.

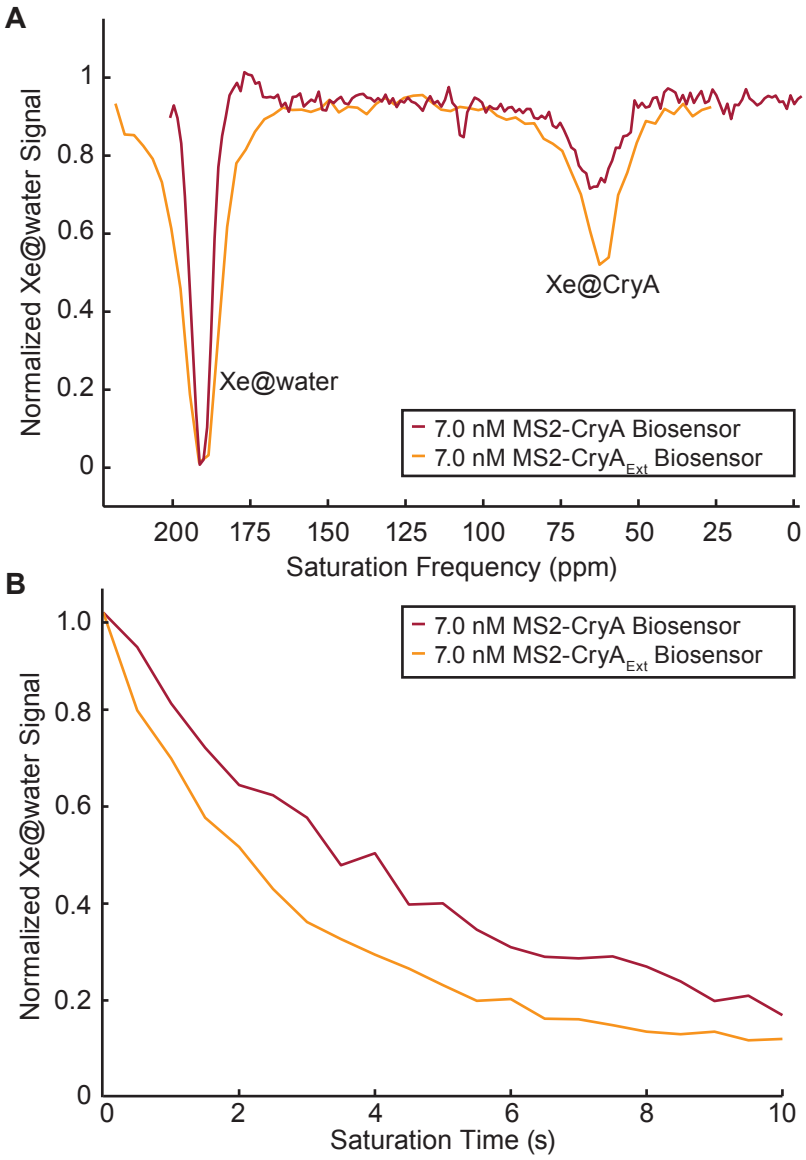



***Figure 7-9. The MS2-CryA$_{Ext}$ biosensor displayed slightly greater sensitivity than the internally modified MS2-CryA biosensor.***

(A) Comparison of the saturation response profiles for solutions of 7.0 nM MS2-CryA$_{Ext}$ and 7.0 nM MS2-CryA stabilized with 5% isopropyl alcohol revealed a small signal increase in the externally modified MS2-based biosensor. Shown are the saturation peaks corresponding to the Xe@CryA chemical shift at approximately 60 ppm and the Xe@water chemical shift at 192 ppm (referenced to the Xe gas chemical shift). (B) Similarly, when the hyperCEST contrasts of both MS2-based biosensors were measured as a function of saturation time, a similar response was observed, with a slight sensitivity advantage observed for the MS2-CryA$_{Ext}$ biosensor.

## 7.3 Conclusions

The incorporation of many xenon-binding molecules into a single MS2 viral capsid significantly lowers the detection threshold of binding targets and improves the solubility and biocompatibility of xenon biosensors. Although this is not the first application of a scaffold to produce an amplified xenon biosensor,[10] it has many more CryA cages per targeting unit and demonstrated greater sensitivity than many previous multivalent scaffolds. While some multivalent gadolinium-based scaffolds produce greater contrast in proton-based MRI studies, there is growing concern over the *in vivo* toxicity of gadolinium-based contrast agents.[28] Additionally, although other contrast mechanisms that rely on enzymes[29] or zeolites[30] show potential as sensitive molecular sensors, they rely on direct detection of xenon, which necessitates both high concentration and high polarization of $^{129}$Xe in the sample.

We are currently exploring complimentary strategies to signal amplification through multivalency to further increase the detection sensitivity of xenon biosensors. For example, detection sensitivity could be increased by physical extraction of xenon through a phase change,[1] by increasing the isotopic abundance of $^{129}$Xe, or by increasing the xenon polarization.[31] Alternatively, viral capsids that can accommodate more CryA cages could lead to higher per agent sensitivities (see Chapters 8 and 9). We anticipate that a combination of these techniques will eventually permit detection of biological targets at femtomolar concentrations, finding applications in portable analytical devices and greatly enhancing sensitivity for *in vivo* analytes.

## 7.4 Materials and Methods

### *7.4.1 General procedures and materials*

All chemical reagents were of analytical grade, obtained from commercial suppliers, and used without further purification unless otherwise noted. Fmoc protected amino acids were obtained from Novabiochem. All reaction flasks were oven dried prior to use and moisture-sensitive reactions were performed in flame-dried flasks under a $N_2$ atmosphere. Analytical thin layer chromatography (TLC) was performed with 60 Å silica gel plates (EMD) and visualized by staining with ceric ammonium molybdate, ninhydrin, or by absorbance of UV light at 254 nm. Flash chromatography was performed with 60 Å 230-400 mesh silica (Silicycle). Organic extracts were dried over $MgSO_4$, and solvent was removed by reduced pressure with a R-114 rotovapor (Buchi) equipped with a self-cleaning dry vacuum pump (Welch). If necessary, products were further dried by reduced pressure with an Edwards RV5 high vacuum pump.

Water used in biological procedures was distilled and deionized using a Milli-Q Gradient ultrapure water purification system (Millipore). Dulbecco's phosphate-buffered saline (DPBS) was obtained from Sigma. Alexa Fluor 488 maleimide (AF488-maleimide) was obtained from Invitrogen.

### *7.4.2 Instrumentation*

**NMR.** All $^{1}$H and $^{13}$C spectra are reported in ppm and referenced to solvent peaks. Spectra were obtained on Bruker AVB-400, AVQ-400, AV-500, or AV-600 spectrometers. Coupling

constants (J) are reported in Hz.

**Absorbance spectra.** Absorbance readings were collected on a Molecular Devices SpectraMax 190 UV-Vis absorbance plate reader or a Thermo Scientific Nanodrop 2000 spectrophotometer.

**Gel Analyses.** Sodium dodecyl sulfate-poly(acrylamide) gel electrophoresis (SDS-PAGE) was accomplished on a Mini-Protean apparatus (Bio-Rad) with 10-20% gradient polyacrylamide gels (BioRad), following the protocol of Laemmli.[32] Protein samples were mixed with SDS loading buffer in the presence of 1,4-dithiothreitol (DTT) and heated to 95 °C for 10 min to ensure reduction of disulfide bonds and complete denaturation unless otherwise noted. Commercially available molecular mass markers (Bio-Rad) were applied to at least one lane of each gel for calculation of the apparent molecular masses. Visualization of protein bands was accomplished by staining with Coomassie Blue. For fluorescent protein conjugates, visualization was accomplished on a Typhoon 9410 imaging system using appropriate excitation and emission filters prior to staining.

**High performance liquid chromatography (HPLC).** Purification of small molecules was performed using reversed-phase (RP) HPLC on a Rainin Dynamax SD-200 system equipped with a Varian UV-Vis detector (model 345) and a Microsorb C18 analytical column (4.6 x 250 mm) at a flow rate of 1 mL/min, a semipreparative column (10 x 250 mm) at a flow rate of 4 mL/min, or a preparative column (21.4 x 250 mm) at a flow rate of 20 mL/min. HPLC samples were filtered with a Pall Life Sciences Acrodisc CR 13 mm syringe filter equipped with a 0.2 μm PTFE membrane prior to injection.

Modified and unmodified MS2 bacteriophage were analyzed by RP-HPLC on an Agilent 1100 Series HPLC System (Agilent Technologies) with an inline diode array detector (DAD) and a Jupiter 5 μm C18 300 Å (Phenomenex) at a flow rate of 1.0 mL/min. Samples were centrifuged prior to injection to pellet insoluble material.

Modified and unmodified MS2 bacteriophage were analyzed by size exclusion chromatography on an Agilent 1100 Series HPLC System (Agilent Technologies) with an inline diode array detector (DAD). Chromatography was accomplished using a Zorbax GF-250 column (Agilent) with isocratic (0.5 mL/min) flow and an aqueous mobile phase (10 mM $Na_2HPO_4$, pH 7.2).

**Matrix assisted laser desorption-ionization time-of-flight (MALDI-TOF) MS.** MALDI-TOF MS was performed on a Voyager-DE system from PerSeptive Biosystems (Applied Biosystems). Protein samples were co-crystallized using saturated solutions of sinapinic acid in 60% acetonitrile, 40% water with 0.1% trifluoroacetic acid. Peptide samples were co-crystallized using saturated solutions of α-Cyano-4-hydroxycinnamic acid (CHCA) in 60% acetonitrile, 40% water with 0.1% trifluoroacetic acid.

**Electrospray ionization (ESI) MS.** High-resolution ESI mass spectra were obtained at the QB3 Mass Spectrometry Facility (UC Berkeley) using an LTQ Orbitrap XL hybrid mass spectrometer (Thermo Fisher). Low-resolution ESI mass spectra were obtained using an LTQ-XL mass

spectrometer.

Liquid chromatography coupled to mass spectrometry (LC-MS) was performed with an Agilent 1100 Series HPLC (Agilent Technologies) system that was connected in-line with an API 150EX mass spectrometer (Applied Biosystems). The mass spectrometer was outfitted with a Turbo Ionspray ionization source. A binary solvent system consisting of buffer A (0.1% formic acid in water (v/v)) and buffer B (0.1% formic acid in acetonitrile (v/v)) was employed at a flow rate of 250 µL/min. The LC was performed using either a Zorbax 300SB-C8 column (Agilent) or a Jupiter 5 µ C18 300 Å column (Phenomenex).

**$^{129}$Xe NMR.** All $^{129}$Xe NMR data were acquired on a 7.05 T vertical bore spectrometer (Varian) equipped with a 26 channel shim set and a dual-tuned ($^{1}$H/$^{129}$Xe) radiofrequency saddle coil (i.d. = 5 mm). $^{129}$Xe was hyperpolarized (P ≈ 4%) via spin-exchange optical pumping with a MITI XenoSpin polarizer (GE Healthcare, formerly Nycomed Amersham). The gas mixture for polarization consisted of 2% xenon (natural abundance), 10% nitrogen, and 88% helium. After polarization, gas flowed through plastic tubing directly into a 5-mm NMR tube outfitted with gas inlet and outlet arms (see Figure 8-3 for a diagram). Prior to each pulse sequence repetition, the xenon mixture was bubbled at a flow rate of approximately 0.3 SLM to saturate the solution with hyperpolarized $^{129}$Xe, followed by a 4 s wait period to allow the solution to settle and bubbles to clear. Saturation consisted of a train of dSNOB pulses,[33] with pulse duration and power calculated using the pxshape tool included in the VNMR software package (Varian Inc.).

### *7.4.3 Chemical synthesis*

**2,5-dioxopyrrolidin-1-yl-3-(2,5-dioxo-2,5-dihydro-1*H*-pyrrol-1-yl)propanoate (7.4).** A mixture of β-alanine (**7.3**, 910 mg, 10.2 mmol) and maleic anhydride (1000 mg, 10.2 mmol) in dry *N,N*-dimethylformamide (DMF, 12 mL) was stirred at RT for 2.5 h until all solids were dissolved. The solution was cooled in an ice bath and then *N*-hydroxysuccinimide (NHS, 1445 mg, 12.56 mmol) and *N,N'*-dicyclohexylcarbodiimide (DCC, 4125 mg, 20 mmol) were added. The solution was stirred for 10 min after which it was removed from the ice bath and stirred at RT overnight. The reaction was filtered and extracted with ethyl acetate (3x). The combined organic layers were washed with water (1x), 5% sodium bicarbonate (3x) and brine (1x). The organic phase was dried over $MgSO_4$ and concentrated under vacuum. The white solid was purified by recrystallization using $CH_2Cl_2$ and petroleum ether to afford a white solid (935 mg, 35%). $^{1}$H NMR (300 MHz, $CDCl_3$): δ = 6.74 (s, 2H), 3.94 (t, *J* = 7.0 Hz, 2H), 3.02 (t, *J* = 7.0 Hz, 2H), 2.82 (s, 4H).

**Fmoc-Lys[maleimide]-COOH (7.6).** To a stirred solution of the ester **7.4** (305 mg, 1.1 mmol) and the trifluoroacetic acid (TFA) salt of Fmoc-Lysine (**7.5**, 352 mg, 0.955 mmol) in dry dichloromethane ($CH_2Cl_2$, 20 mL) was added *N,N*-diisopropylethylamine (DIPEA, 0.5 mL, 2.86 mmol). The solution was stirred at RT overnight. The reaction was diluted with addition of $CH_2Cl_2$ and washed with saturated $NH_4Cl$ (1 x 40 mL). After the aqueous phase was extracted with $CH_2Cl_2$ (3 x 20 mL), the organic phases were combined, dried over $MgSO_4$, and then concentrated under vacuum. Purification by silica gel chromatography with 1:9 ethyl acetate:hexanes + 1% acetic acid gave the desired maleimide functionalized amino acid as an off-white residue. Precipitation

in water afforded compound 2 as a white solid (270 mg, 55%). ESI-MS(+) calcd for $C_{28}H_{29}N_3O_7$ $[M+H]^+$: 520.20 Da, found 519.1 Da $[M+H]^+$.

**Maleimide-functionalized CryA peptide (7.1).** First, a water solubilizing tetraglutamic acid peptide (**7.7**) was prepared on polystyrene Wang resin (1.2 mmol/g, 100–200 mesh) using standard Fmoc solid-phase peptide synthesis (SPPS) with $N_\alpha$-Fmoc protected amino acids, DIPEA, and HOBT/HBTU ester activation in DMF. A 5-fold excess of amino acid was used for the coupling steps, which involved gentle rotation for 3 h. Fmoc removal was achieved with 20% piperidine in DMF (2 x 10 min). Next, the maleimide functionality was installed by coupling **7.6** using the above coupling conditions for 5 h to generate compound **7.8**. Following deprotection with 20% piperidine in DMF and coupling to Fmoc-6-aminohexanoic acid (Fmoc-Ahx-COOH) using the same coupling conditions as above, the undesirable peptide **7.9** was formed instead of **7.10**. Verified by MALDI-MS.

**Maleimide-functionalized CryA (7.12).** To a solution of (*S*)-2-(2-amino-6-(2,5-dioxo-2,5-dihydro-1*H*-pyrrol-1-yl)hexanamido)ethanesulfonate[22] (**7.11**, 6.0 mg, 13.9 µmol) in DMF (2 mL) was added CryA cage (**6.1**, 26.1 mg, 27.8 µmol, supplied by Kang Zhao, Tianjin University, China). *O*-(Benzotriazol-1-yl)-*N,N,N′,N′*- tetramethyluronium hexafluorophosphate (HBTU, 13.2 mg, 34.8 µmol) was added as the coupling reagent with 1-hydroxybenzotriazole (HOBT, 4.7 mg, 34.8 µmol) and DIPEA (12.1 µL, 69.5 µmol). The reaction mixture was allowed to stir for 2 h at RT. Water (2 mL) was added and the resulting solution was first purified on a Phenomenex Strata X-AW 33 µm polymeric weak anion column (conditioned with methanol; equilibrated with 5% formic acid in methanol and then with 0.1% formic acid in methanol; washed with DMF and then 0.1% formic acid in methanol; and eluted with 5% ammonium hydroxide in acetonitrile). The white solid was then purified by RP-HPLC on a C18 semipreparative column affording compound **7.12** as a white solid (12.6 mg, 72.3%). $^1$H NMR (500 MHz, DMSO): δ, 6.82 (m, 14H), 6.56 (s, 1H), 4.62 (m, 3H), 4.50 (m, 7H), 4.11 (m, 16H), 3.77 (s, 3H), 3.72 (m, 14H), 3.64 (s, 3H), 3.33 (m, 6H), 1.23 (m, 2H).

**4-formyl-benzoate-*N*-hydroxysuccinimide ester (7.13).**[18,23] To a stirred solution of *p*-formylbenzoic acid (200 mg, 1.33 mmol) in dry tetrahydrofuran (THF, 13 mL) cooled on an ice bath were added NHS (230 mg, 2.0 mmol), DCC (310 mg, 1.5 mmol), and DIPEA (520 µL, 3.0 mmol). The solution was stirred for 30 min after which it was removed from the ice bath and stirred at RT for 4 h. The reaction was diluted with the addition of $CH_2Cl_2$ and then filtered through a pad of Celite. Purification by silica gel chromatography using 1:3 ethyl acetate:hexanes + 1% acetic acid afforded the desired ester as a white solid. $^1$H NMR (400 MHz, $CDCl_3$): δ = 10.14 (s, 1H), 8.31 (d, *J* = 8.3 Hz, 2H), 8.03 (d, *J* = 8.7 Hz, 1H), 2.94 (s, 4H). ESI-MS calcd for $C_{12}H_9NO_5$ [M+H]: 247.05 Da, found 247.04 Da [M+H].

***N′*-Boc-aminooxyacetyl-*N*-hydroxysuccinimide ester (7.15).** To a stirred solution of *N*-Boc-aminooxyacetic acid (**7.14**, 500 mg, 2.62 mmol) in $CH_2Cl_2$ (30 mL) cooled on an ice bath were added NHS (332 mg, 2.88 mmol) and DCC (594 mg, 2.88 mmol). The solution was stirred for 5 min after which it was removed from the ice bath and stirred at RT for 45 min. The reaction was

diluted with the addition of $CH_2Cl_2$, filtered through a pad of Celite, and the filtrate was washed with saturated $NaHCO_3$ (2 x 40 mL). The organic phase was dried over $MgSO_4$ and concentrated under vacuum. The white solid was used without further purification. $^1$H NMR (400 MHz, $CDCl_3$): δ = 7.67 (1H, s), 4.78 (2H, s), 2.88 (4H, s), 1.49 (9H, s). $^{13}$C NMR (151 MHz, $CDCl_3$): δ = 168.8, 165.2, 156.4, 82.8, 71.0, 28.3, 25.8. ESI-HRMS calcd for $C_{11}H_{16}N_2O_7$ [M+Na]: 311.0858 Da, found 311.0852 Da [M+Na].

**Fmoc-Lys[Boc-aminooxyacetyl]-COOH (7.16).** To a stirred solution of the ester **7.15** (700 mg, 2.43 mmol) and the TFA salt of Fmoc-Lysine (**7.5**, 1.24 g, 2.67mmol) in dry $CH_2Cl_2$ (30 mL) was added DIPEA (1.59 mL, 9.17 mmol). The solution was stirred at RT for 11 h. The reaction was diluted with the addition of $CH_2Cl_2$ and washed with saturated $NH_4Cl$ (1 x 40 mL). After the aqueous phase was extracted with $CH_2Cl_2$ (3 x 20 mL), the organic phases were combined, dried over $MgSO_4$, and then concentrated under vacuum. Purification by silica gel chromatography using 1:4 ethyl acetate:hexanes + 1% acetic acid gave the desired aminooxy functionalized amino acid as an off-white residue. Resolving the residue in $CH_2Cl_2$ and adding it drop-wise to a solution of hexanes precipitated the product as a white solid (757 mg, 58%). $^1$H NMR (600 MHz, $CDCl_3$): δ = 8.27 (1H, broad s), 7.75 (2H, d, J = 7.8 Hz), 7.60 (2H, dd, J = 7.2 Hz, 11.4 Hz), 7.38 (2H, t, J = 7.2 Hz), 7.29 (2H, t, J = 7.2 Hz), 5.82 (1H, s), 4.46-4.32 (2H, m), 4.29 (2H, s), 4.18 (1H, t, J = 7.2 Hz), 3.40-3.29 (2H, m), 1.93-1.75 (2H, m), 1.63-1.54 (2H, m), 1.44, (9H, s), 1.41-1.27 (2H, m). $^{13}$C NMR (151 MHz, $CDCl_3$): δ = 174.9, 169.9, 158.2, 156.5, 144.0, 141.5, 127.9, 127.3, 125.4, 120.2, 83.5, 76.1, 67.2, 54.0, 47.4, 38.6, 31.7, 28.8, 28.3, 22.2. ESI-HRMS(+) calcd for $C_{28}H_{36}N_3O_8$ $[M+H+Na]^+$: 564.2324 Da, found 564.2307 Da $[M+H+Na]^+$.

**Aminooxy-functionalized CryA peptide (7.18).** First, a water solubilizing tetraglutamic acid peptide (**7.7**) was prepared on polystyrene Wang resin (1.2 mmol/g, 100–200 mesh) using standard Fmoc solid-phase peptide synthesis (SPPS) with $N_\alpha$-Fmoc protected amino acids, DIPEA, and HOBT/HBTU ester activation in DMF. A 5-fold excess of amino acid was used for the coupling steps, which involved gentle rotation for 3 h. Fmoc removal was achieved with 20% piperidine in DMF (2 x 10 min). Next, the aminooxy functionality was installed by coupling **7.16** using the above coupling conditions for 5 h to generate compound **7.17**. Finally, the peptide was capped at the N-terminus using a 2.5 fold excess of CryA (**6.1**) using the above coupling conditions for 12-20 h. The product was cleaved using a cocktail of TFA:water:triisopropylsilane (95:2.5:2.5) for 2.5 h and then precipitated in methyl *tert*-butyl ether. The crude product was purified by RP-HPLC using a C18 column from 40% to 60% buffer B over 30 min to yield peptide **7.18**. A stock solution of the aminooxy-functionalized CryA-peptide was prepared by dissolving the purified solid in DPBS and adjusting the pH with sodium hydroxide until the solution became clear. ESI-MS(+) calculated for $C_{81}H_{97}N_7O_{29}$ $(M+H)^+$: 1656.6 Da found 1656.5 Da $(M+H)^+$.

### *7.4.4 Experimental*

**Expression and purification of MS bacteriophage.** Wt and N87C MS2 were expressed and purified according to previously published protocols.[12,13,22]

**Labeling N87C MS2 with AF488-maleimide.** Wt and N87C MS2 were reacted with Alexa Fluor 488 maleimide (AF488-maleimide) using the following final reaction concentrations: 555 nM MS2 (either wt or N87C), 10 mM AF88-maleimide in 10 mM phosphate buffer, pH 7.2. The reaction was briefly mixed and allowed to react at RT for 15 min. Samples were subjected to SDS-PAGE analysis.

**MS2-CryA biosensor construction (internal modification).** To a solution of N87C MS2 (2 nmol, 6.7 µL, 300 µM) in 60 µL of 10mM phosphate buffer at pH 7.2 was added the CryA-maleimide construct **7.12** (50 nmol, 5.3 µL, 9.4 mM in DMF). The reaction mixture was incubated at RT for 4 h followed by purification by size exclusion chromatography. The amount of MS2 before and after purification was estimated using SDS-PAGE and optical densitometry analysis. A control reaction was run with wt MS2, which lacked the N87C mutation, to test the reactivity of the native cysteine residues in the presence of **7.12**. No appreciable reaction product was observed when wt MS2 was subjected to the conditions described above.[34] The purified MS2-CryA biosensor was spin-concentrated using Micro centrifugal filter devices with an Ultracel YM-100 membrane (Millipore) as directed by the manufacturer. ESI-MS(+): 14973 Da (avg. mass), found 14976 Da (avg. mass).

**Generation of aldehyde-functionalized MS2 (ald-MS2).** Lysine residues on MS2 capsid proteins were converted into aldehydes by combining 250 µL of 1.5 µM wt MS2 (in 10 mM phosphate buffer, pH 7.5) with 13.5 µL of 50 mM **7.13** (in DMF). After incubation for 3 h at RT, the resulting ald-MS2 was purified by gel filtration using an Illustra Nap-5 column. The purified ald-MS2 was spin-concentrated using Micro centrifugal filter devices with an Ultracel YM-100 membrane (Millipore) as directed by the manufacturer and stored at 4 °C until needed. ESI-MS(+): 13732 Da (unmodified, avg. mass), 13864 Da (one aldehyde, avg. mass), 13996 Da (two aldehydes, avg. mass), 14128 Da (three aldehydes, avg. mass); found 13733 Da, 13863 Da, 13997 Da, and 14131 Da.

**MS2-CryA biosensor construction (external modification).** MS2 protein capsids were modified with CryA cages using the following final reaction conditions: 500 nM ald-MS2, 1 mM aminooxy-functionalized CryA-peptide (**7.18**), and 100 mM aniline in 10 mM phosphate buffer, pH 7.5. After 24 h at RT, the resulting MS2-$CryA_{Ext}$ construct was purified by gel filtration using an Illustra Nap-5 column. The extent of CryA cage labeling was verified and quantified using SDS-PAGE and optical densitometry. Purified MS2-$CryA_{Ext}$ biosensors were stored at 4 °C until needed. ESI-MS(+): 13732 Da (unmodified, avg. mass), 15502 Da (one CryA cage, avg. mass), 17272 Da (two CryA cages, avg. mass), 19042 Da (three CryA cages, avg. mass); found 13732 Da, 15504 Da, 17273 Da, and 19049 Da.

**Characterization of MS2-based biosensors by LC-MS.** Approximately 0.2 nmol of MS2-based biosensors were analyzed by LC-MS using a binary solvent system consisting of buffer A (0.1% formic acid in water (v/v)) and buffer B (0.1% formic acid in acetonitrile (v/v)). The elution gradient was 30% buffer B for the first 1 min, ramping to 70% buffer B over 20 min, then to 95%

buffer B over the next 1 min, and a 10 min wash with 95% buffer B, at a flow rate of 250 µL/min. Extracted ion chromatograms were deconvoluted with Analyst software (Applied Biosystems) to generate protein mass reconstructions.

**Preparation of MS2-CryA biosensor solutions for $^{129}$Xe NMR.** All MS2-based biosensor solutions were prepared in 0.9 mL of 10 mM phosphate buffer at pH 7.2 containing 5% v/v isopropyl alcohol.

**hyperCEST settings for $^{129}$Xe NMR.** Saturation response profiles were recorded by measuring the Xe@water signal as a function of the offset frequency of the saturation pulse train. Saturations consisted of a train of shaped RF pulses, known as dSNOB pulses. For the saturation response profile, a train of 100 Hz bandwidth was used (the pulse length ($\tau_{pulse}$) was 28.2 ms, the maximum RF intensity ($B_{1,max}$) was 6.65 µT, and the total saturation time ($t_{sat}$) was 10 s).

To investigate the MS2-CryA biosensor concentration dependence on hyperCEST contrast and determine a detection threshold, plots of residual signal versus saturation time were generated for a dilution series of the biosensor (700 pM, 70 pM, 7 pM, 0.7 pM). A train of 2500 to 5000 Hz bandwidth dSNOB pulses was applied ($\tau_{pulse}$ = 940 µs, $B_{1,max}$ = 352 µT), and the number of pulses was varied to change the total saturation time. Contrast was quantified by comparing the Xe@water signal from two separate saturations, one after application of the pulse train on-resonance at the frequency of Xe@CryA, and the second after application equidistant from, but on the opposite side of the Xe@water peak (off-resonance). The exact parameters for each saturation curve are presented in Table 7-1.

***Table 7-1. Pulse parameters for detection of MS-CryA biosensors.***

Both off- and on-resonant saturation experiments were performed using d-SNOB pulses. The constant τ refer to exponential fits of the data, modeled by the function f(t) = exp(–t/τ), with the error denoting the 90% confidence intervals. Using the fitted data, 20 s contrast indicates the contrast observed after saturation time t = 20 s under the indicated saturation conditions.

| Sample | Temp (°C) | Bandwidth (Hz) | Duration (µs) | Delay (ms) | B1 (kHz) | $\tau_{off\text{-}res}$ ($s^{-1}$) | $\tau_{on\text{-}res}$ ($s^{-1}$) | 20 s contrast |
|---|---|---|---|---|---|---|---|---|
| Buffer | 40 | 3000 | 940 | 0 | 3.962 | 36.432 ± 1.5055 | 36.852 ± 2.2294 | -0.006 ± 0.040 |
| 700 pM MS2-CryA | 37 | 5000 | 562.8 | 9.45 | 6.280 | 28.241 ± 1.0158 | 14.161 ± 0.4649 | 0.505 ± 0.013 |
| 70 pM MS2-CryA | 37 | 2500 | 1130 | 1.37 | 3.147 | 32.169 ± 1.3789 | 19.158 ± 0.7147 | 0.344 ± 0.031 |
| 7 pM MS2-CryA | 40 | 3000 | 940 | 0 | 3.962 | 31.615 ± 1.3571 | 23.346 ± 0.9579 | 0.201 ± 0.036 |
| 0.7 pM MS2-CryA | 40 | 5000 | 562.8 | 0 | 6.280 | 31.897 ± 1.1681 | 27.756 ± 1.1372 | 0.089 ± 0.034 |

### *7.4.5 Data analysis*

**NMR data analysis.** Data processing included application of an apodization filter and Fourier transformation (NMRPipe software), and integration and normalization of the area under the peak corresponding to Xe@water (Matlab, The Math Works Inc.). Saturation profiles were fitted to exponential functions and error analysis was performed using the *nlinfit* and *nlparci* functions of Matlab.

## 7.5 References

1. Zhou X., Graziani D. and Pines A. (**2009**) Hyperpolarized xenon NMR and MRI signal amplification by gas extraction. *Proc Natl Acad Sci U S A* 106, 16903-16906.

2. Sherry A.D. and Woods M. (**2008**) Chemical exchange saturation transfer contrast agents for magnetic resonance imaging. *Annu Rev Biomed Eng* 10, 391-411.

3. Zhang S., Merritt M., Woessner D.E., Lenkinski R.E. and Sherry A.D. (**2003**) PARACEST agents: modulating MRI contrast via water proton exchange. *Acc Chem Res* 36, 783-790.

4. Seward G.K., Wei Q. and Dmochowski I.J. (**2008**) Peptide-mediated cellular uptake of cryptophane. *Bioconjug Chem* 19, 2129-2135.

5. Goodson B.M. (**2002**) Nuclear magnetic resonance of laser-polarized noble gases in molecules, materials, and organisms. *J Magn Reson* 155, 157-216.

6. Berthault P., Huber G. and Desvaux H. (**2009**) Biosensing using laser-polarized xenon NMR/MRI. *Prog Nucl Mag Res Sp* 55, 35-60.

7. Spence M.M. *et al.* (**2001**) Functionalized xenon as a biosensor. *Proc Natl Acad Sci U S A* 98, 10654-10657.

8. Schroder L., Lowery T.J., Hilty C., Wemmer D.E. and Pines A. (**2006**) Molecular imaging using a targeted magnetic resonance hyperpolarized biosensor. *Science* 314, 446-449.

9. Gabard J. and Collet A. (**1981**) Synthesis of a (D3)-Bis(Cyclotriveratrylenyl) Macrocage by Stereospecific Replication of a (C3)-Subunit. *J Chem Soc Chem Comm*, 1137-1139.

10. Mynar J.L., Lowery T.J., Wemmer D.E., Pines A. and Frechet J.M. (**2006**) Xenon biosensor amplification via dendrimer-cage supramolecular constructs. *J Am Chem Soc* 128, 6334-6335.

11. Valegard K., Liljas L., Fridborg K. and Unge T. (**1990**) The three-dimensional structure of the bacterial virus MS2. *Nature* 345, 36-41.

12. Carrico Z.M., Romanini D.W., Mehl R.A. and Francis M.B. (**2008**) Oxidative coupling of peptides to a virus capsid containing unnatural amino acids. *Chem Commun (Camb)*, 1205-1207.

13. Tong G.J., Hsiao S.C., Carrico Z.M. and Francis M.B. (**2009**) Viral capsid DNA aptamer conjugates as multivalent cell-targeting vehicles. *J Am Chem Soc* 131, 11174-11178.

14. Dedeo M.T., Finley D.T. and Francis M.B. (**2011**) Viral capsids as self-assembling templates for new materials. *Prog Mol Biol Transl Sci* 103, 353-392.

15. Yildiz I., Shukla S. and Steinmetz N.F. (**2011**) Applications of viral nanoparticles in medicine. *Curr Opin Biotechnol* 22, 901-908.

16. Stephanopoulos N. and Francis M.B. (**2011**) Choosing an effective protein bioconjugation strategy. *Nat Chem Biol* 7, 876-884.

17. Datta A. *et al.* (**2008**) High relaxivity gadolinium hydroxypyridonate-viral capsid conjugates: nanosized MRI contrast agents. *J Am Chem Soc* 130, 2546-2552.

18. Hooker J.M., Datta A., Botta M., Raymond K.N. and Francis M.B. (**2007**) Magnetic resonance contrast agents from viral capsid shells: a comparison of exterior and interior cargo strategies. *Nano Lett* 7, 2207-2210.

19. Garimella P.D., Datta A., Romanini D.W., Raymond K.N. and Francis M.B. (**2011**) Multivalent, high-relaxivity MRI contrast agents using rigid cysteine-reactive gadolinium complexes. *J Am Chem Soc* 133, 14704-14709.

20. Hooker J.M., O'Neil J.P., Romanini D.W., Taylor S.E. and Francis M.B. (**2008**) Genome-free viral capsids as carriers for positron emission tomography radiolabels. *Mol Imaging Biol* 10, 182-191.

21. Farkas M.E. *et al.* (**2012**) PET Imaging and Biodistribution of Chemically Modified Bacteriophage MS2. *Mol Pharm*.

22. Wu W., Hsiao S.C., Carrico Z.M. and Francis M.B. (**2009**) Genome-free viral capsids as multivalent carriers for taxol delivery. *Angew Chem Int Ed Engl* 48, 9493-9497.

23. Kovacs E.W. *et al.* (**2007**) Dual-surface-modified bacteriophage MS2 as an ideal scaffold for a viral capsid-based drug delivery system. *Bioconjug Chem* 18, 1140-1147.

24. Lowery T.J. *et al.* (**2006**) Optimization of xenon biosensors for detection of protein interactions. *Chembiochem* 7, 65-73.

25. Han S.I. *et al.* (**2005**) NMR-based biosensing with optimized delivery of polarized 129Xe to

solutions. *Anal Chem* 77, 4008-4012.

26. Schroder L. *et al.* (**2008**) Temperature response of 129Xe depolarization transfer and its application for ultrasensitive NMR detection. *Phys Rev Lett* 100, 257603.

27. Meldrum T., Bajaj V.S., Wemmer D.E. and Pines A. (**2011**) Band-selective chemical exchange saturation transfer imaging with hyperpolarized xenon-based molecular sensors. *J Magn Reson* 213, 14-21.

28. Ersoy H. and Rybicki F.J. (**2007**) Biochemical safety profiles of gadolinium-based extracellular contrast agents and nephrogenic systemic fibrosis. *J Magn Reson Imaging* 26, 1190-1197.

29. Wei Q. *et al.* (**2006**) Designing 129Xe NMR biosensors for matrix metalloproteinase detection. *J Am Chem Soc* 128, 13274-13283.

30. Lerouge F.d.r. *et al.* (**2008**) Towards thrombosis-targeted zeolite nanoparticles for laser-polarized 129Xe MRI. *J. Mater. Chem.* 19, 379-386.

31. Schrank G., Ma Z., Schoeck A. and Saam B. (**2009**) Characterization of a low-pressure high-capacity ^{129}Xe flow-through polarizer. *Physical Review A* 80, 063424.

32. Laemmli U.K. (**1970**) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680-685.

33. Kupce E., Boyd J. and Campbell I.D. (**1995**) Short selective pulses for biochemical applications. *J Magn Reson B* 106, 300-303.

34. Meldrum T. *et al.* (**2010**) A xenon-based molecular sensor assembled on an MS2 viral capsid scaffold. *J Am Chem Soc* 132, 5936-5937.

# Chapter 8: A $^{129}$Xe-based MRI Contrast Agent Assembled on the M13 Bacteriophage Viral Capsid

## 8.1 Introduction

Molecular imaging aims to detect the presence of biomolecules and map their spatial distribution *in vivo*. Of particular interest are tissue biomarkers, where the early detection of disease-specific markers, when they exist at very low concentrations, can maximize the diagnostic and prognostic value of molecular imaging. In the case of magnetic resonance imaging (MRI), due to its inherent lack of signal sensitivity, it is difficult to generate contrast between normal and diseased tissue during these early disease stages with conventional techniques. Consequently, exogenous contrast agents are often employed to generate unique contrast, since many contrast agents can be detected at low concentrations, enhancing signal sensitivity.

The most common agents, principally gadolinium chelates,[1] enhance water relaxation through paramagnetic interactions. Other types of agents include paramagnetic shift reagents (e.g., anisotropic lanthanide chelates) which enable chemical exchange saturation transfer (CEST) contrast (i.e., PARACEST),[2] super-paramagnetic iron oxide (SPIO) field perturbation agents,[3] and agents incorporating or composed of non-proton nuclei, including thermally polarized ($^{19}$F)[4] and hyperpolarized nuclei ($^{13}$C, $^{3}$He, $^{129}$Xe, $^{15}$N, $^{19}$F).[5-9] Despite improvements, these MRI contrast agents still lack the sensitivity of positron emission tomography (PET) and single-photon emission computed tomography (SPECT) nuclear medicine techniques in which radioactive tracers can be detected at sub-nanomolar concentrations,[10] the necessary detection threshold for clinically relevant molecular imaging applications.

Recently developed MRI contrast agents called xenon biosensors[11] have demonstrated the ability to utilize the complementary sensitivity enhancements of both hyperpolarized (hp) $^{129}$Xe and CEST detection.[12] The resulting hyperCEST contrast mechanism has demonstrated detection thresholds of these biosensors that are several orders of magnitude lower than those of the previously mentioned MRI agents, down to 10 nM.[13] Xenon biosensors employ cryptophane-A molecular cages (CryA)[14] to reversibly bind xenon atoms that have been dissolved into environments such as aqueous tissue. To improve the detection sensitivity of biosensors further, multiple copies of CryA have been assembled onto supramolecular scaffolds, including branched dendrimers[15] and viral capsids (see Chapter 7).[16]

In this Chapter, the application of viral capsids for $^{129}$Xe-based MRI contrast agents was extended to filamentous bacteriophage. This class of viral capsids was used as a scaffold for constructing a contrast agent for two reasons. First, filamentous phage, such as the M13 and fd types, are routinely used in phage display techniques for identifying new epitope-targeting groups such as peptides and antibody fragments.[17,18] Consequently, established phage display techniques can produce M13 or fd phage that have high binding affinities for application-specific targets such as biomarkers of disease. Second, the single-stranded DNA of the M13 bacteriophage is coated with approximately 2700 identical copies of its major coat protein (p8), upon

Todd Stevens and Matt Ramirez contributed to the work presented in this chapter.
Portions of the work described in this chapter have been reported in a separate publication.[35]

which a large number of CryA cages can be covalently attached using site-specific protein bioconjugation techniques (Figure 8-1A).[19-22] The result is an increase in CryA payload per targeted biomarker that is bound by the modified phage. Additionally, phage modification also presents an opportunity to decorate the CryA-bearing scaffold with polyethylene glycol (PEG), which can aid in the solubility and biocompatibility of the agent. The extent to which the hyperCEST-based contrast scales with the number of CryA cages per agent is dependent on several parameters, including changes in the xenon exchange dynamics and bound $^{129}$Xe linewidths (due to slower molecular tumbling of the macromolecular scaffold), and thus must be investigated.

The purpose of this study was to investigate the feasibility of using M13 filamentous bacteriophage modified with CryA molecular cages as a $^{129}$Xe-based MRI contrast agent, to compare the hyperCEST response of this phage-based agent relative to unscaffolded CryA-ONH$_2$ (**8.2**), and to establish an *in vitro* concentration threshold for detecting CryA-modified M13.

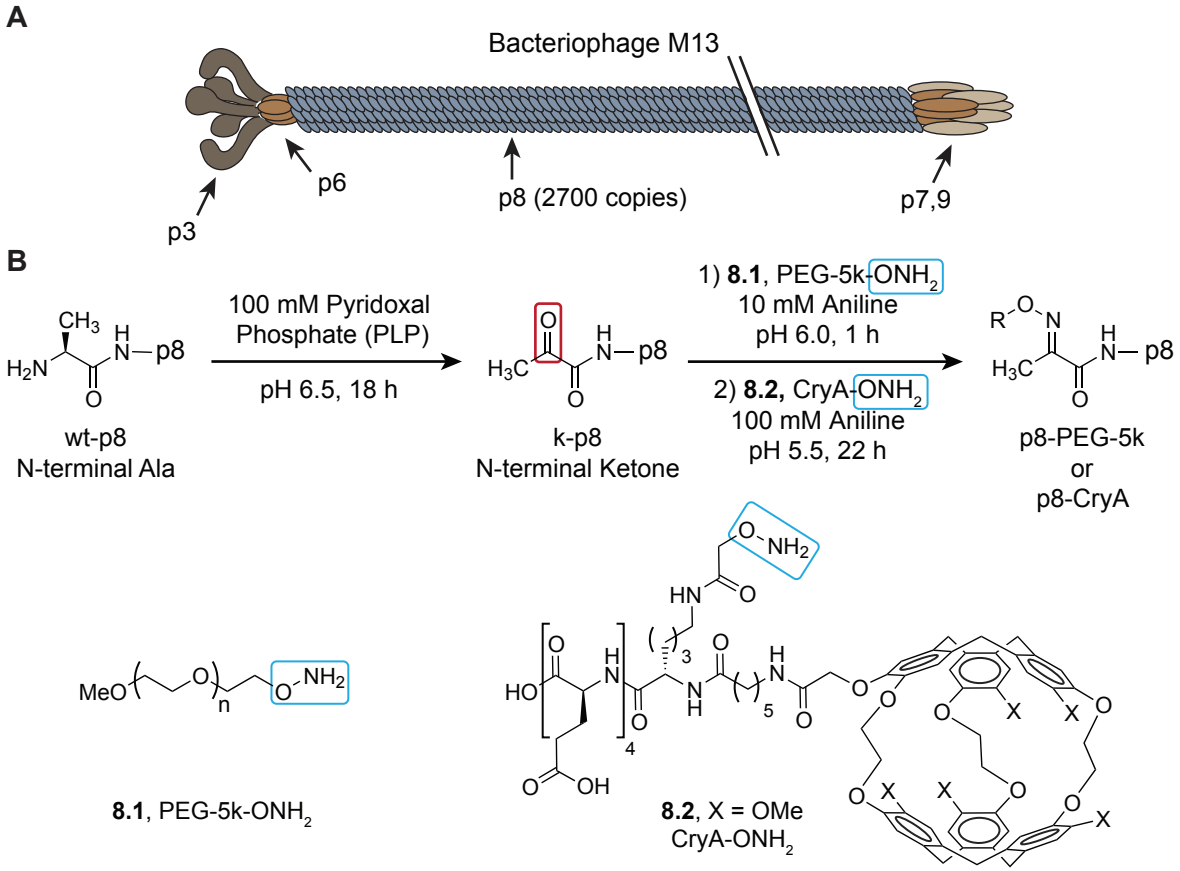
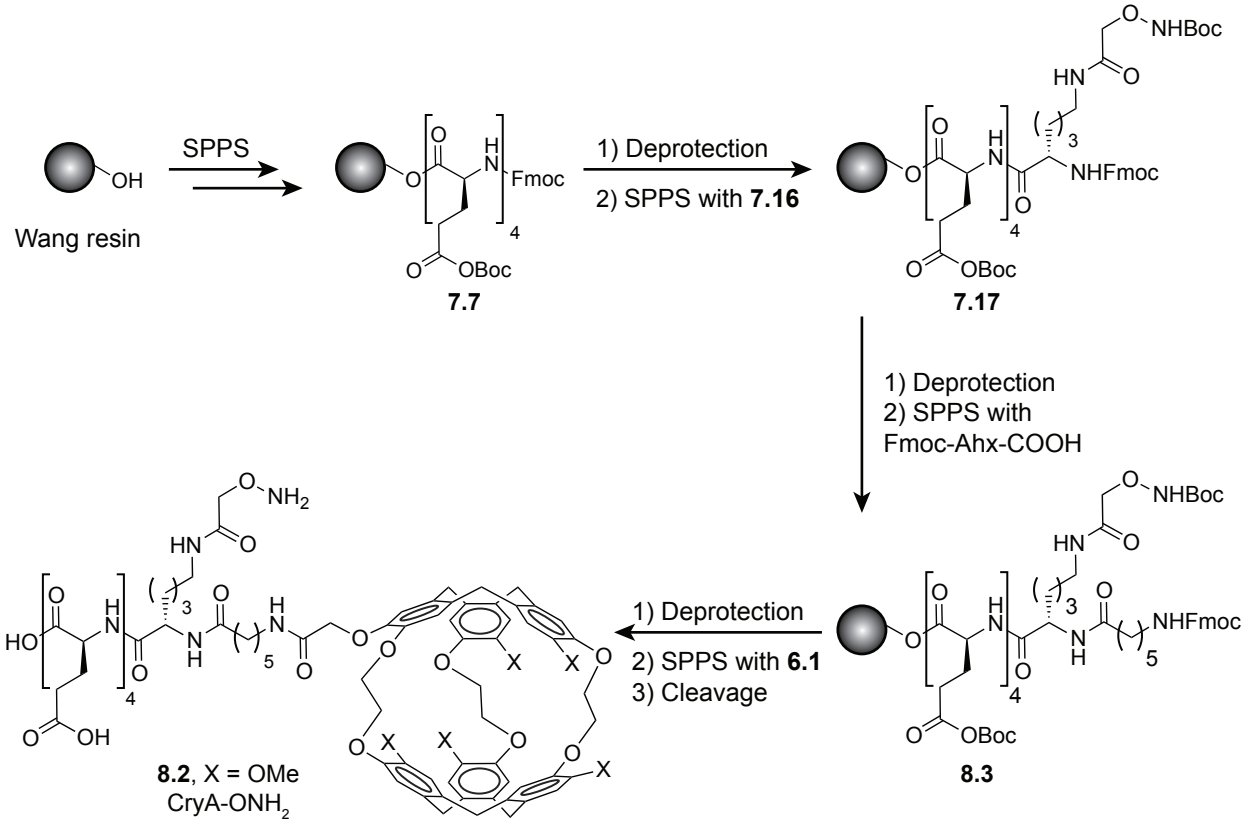



***Figure 8-1. Structure and construction of an M13-CryA biosensor.***

(A) Cartoon of M13 filamentous bacteriophage with significant proteins labeled with arrows. The double slash indicates that the phage is much longer than shown when scaled to the minor coat proteins. (B) Chemical scheme illustrating the transamination and attachment of synthetic molecules to M13 phage. The N–termini of the p8 proteins (blue ovals) were transaminated to yield thousands of ketones along the length of the capsids. The ketones were then reacted with aminooxy-functionalized PEG polymers (**8.1,** PEG-5k-ONH$_2$), followed by reaction with aminooxy–functionalized cryptophane A cages (**8.2**, CryA–ONH$_2$), to construct the M13–based biosensor.

*Scheme 8-1. Synthesis of aminooxy-functionalized CryA peptide 8.2 (CryA-ONH$_2$).*




## 8.2 Results and Discussion

### *8.2.1 Construction of an M13-CryA biosensor*

Synthetic components can be introduced onto the surface of filamentous phage through several bioconjugation strategies.[19,23,24] Recently, we applied an N-terminal specific bioconjugation method for modifying filamentous phage containing scFvs with fluorophores and polymers and verified that modified phage maintained their cell binding properties.[22,25,26] Based on this success, we utilized the same N-terminal specific chemistry to construct an M13-based xenon biosensor (Figure 8-1B). First, the N-termini of the coat proteins were transaminated using pyridoxal 5′-phosphate (PLP) to introduce ketone groups. The ketone-containing M13 phage (k-M13) were purified through a series of precipitations and resuspensions. To improve solubility and bioavailability of the construct, some of the ketone groups were reacted with an aminooxy-functionalized 5 kDa polyethylene glycol polymer (PEG-5k-ONH$_2$, **8.1**).[27] Following gel filtration purification, the PEG-modified M13 (M13-PEG) was reacted with an aminooxy-functionalized CryA peptide (CryA-ONH$_2$, **8.2**) to modify some of the remaining N-terminal ketone groups. CryA-ONH$_2$ was synthesized using solid phase peptide synthesis (SPPS, see Scheme 8-1). The resulting M13-PEG-CryA conjugate, termed the M13-CryA biosensor, was purified by gel filtration chromatography. The dual modification with both PEG and CryA was required to ensure

the aqueous solubility of the phage following a high level of modification with hydrophobic CryA cages. This was accomplished by controlling the amount of PEG-5k modification (approximately 28%, Figure 8-2) to allow sufficient access to unreacted ketone groups by CryA-$ONH_2$. The amount of CryA modification was balanced with PEG-5k modification to achieve the goal of both high CryA payload and aqueous solubility of the construct. Ultimately, the M13-CryA biosensor reflected a compromise between these two objectives.

The M13-CryA biosensor was characterized by reversed-phase HPLC (Figure 8-2) and electrospray ionization mass spectrometry (not shown). On average, the M13-CryA biosensors were approximately 39% modified with CryA cages, corresponding to approximately 1050 copies/phage. Subsequently, the construct was dissolved in Dulbecco's phosphate-buffered saline

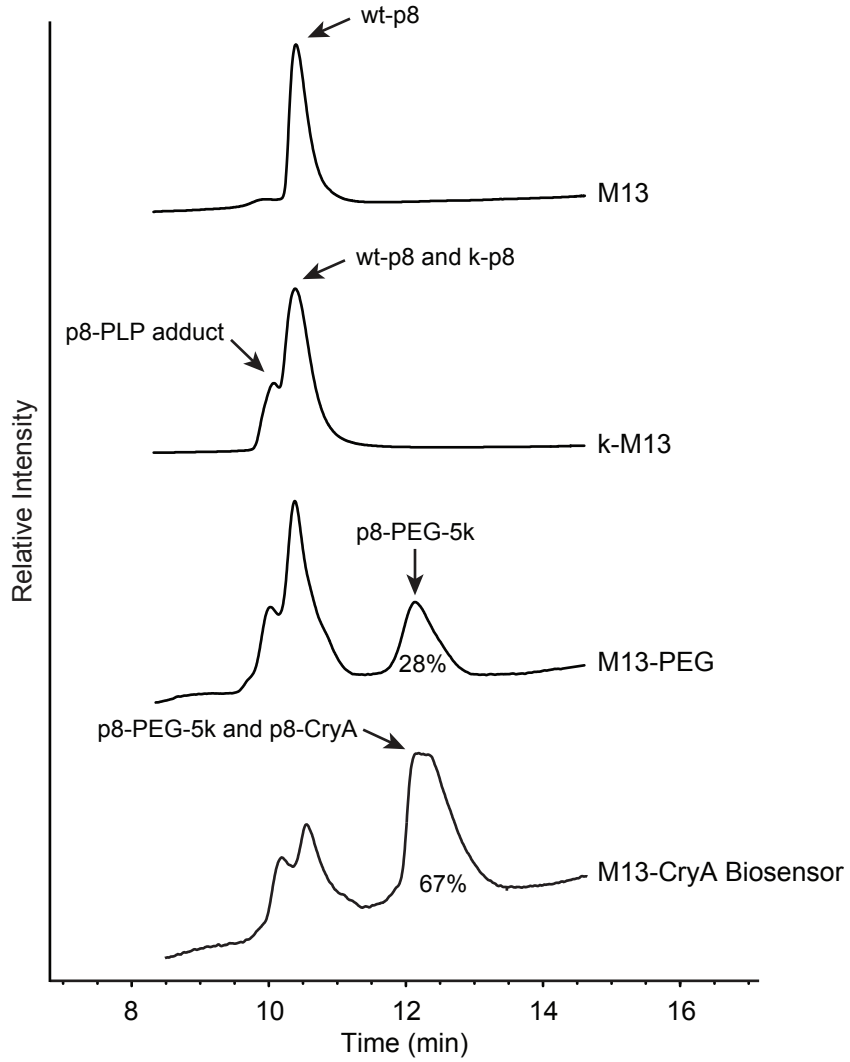



***Figure 8-2. Characterization of modified M13 phage by RP-HPLC.***

Wt M13, ketone-labeled M13 (k-M13), PEG-modified M13 (M13-PEG), and M13-CryA phage were denatured and subjected to RP-HPLC analysis. Only the major coat protein, p8, could be detected. The chromatograms revealed the p8-PEG and p8-CryA modifications had overlapping retention times. On average, the p8 coat proteins of the M13–CryA biosensor were approximately 28% modified with PEG–5k (760 copies) and approximately 39% modified with CryA (1050 copies).

(DPBS, pH 7.4) to evaluate its hyperCEST response profile.

### *8.2.2 Characterization of the M13-CryA biosensor by $^{129}$Xe NMR*

Xenon gas was introduced into a solution of xenon biosensors using a continuous-flow bubbling system (illustrated in Figure 8-3A).[28] The output from a xenon polarizer (spin exchange optical pumping) was connected to a 5-mm NMR tube modified with inlet and outlet arms, allowing for direct delivery of hp $^{129}$Xe into the solution via a capillary running to the bottom of the NMR tube (illustrated in Figure 8-3B). While this method facilitated the rapid dissolution of xenon gas, bubbling through a solution of phage introduced complications, including foaming and the mechanical stress of bubbles on phage integrity. To determine the compatibility of xenon bubbling with the M13-CryA biosensor, we assessed the stability of the biosensor's saturation response profile in comparison to unscaffolded CryA-ONH$_2$. As expected, for a 2.5 μM solution of unscaffolded CryA-ONH$_2$ (2.5 μM in CryA) in DPBS (Figure 8-4A), a region of saturation centered at 60 ppm, characteristic of Xe@CryA, was observed. However, for a 1.2 nM solution of M13-CryA biosensor (2 μM in CryA) in DPBS (Figure 8-4B), while a region of saturation centered at 60 ppm was also observed, the signal intensity decreased with time (i.e., the maximum signal at the beginning of the experiment, depicted in green circles, is greater than the signal at the end

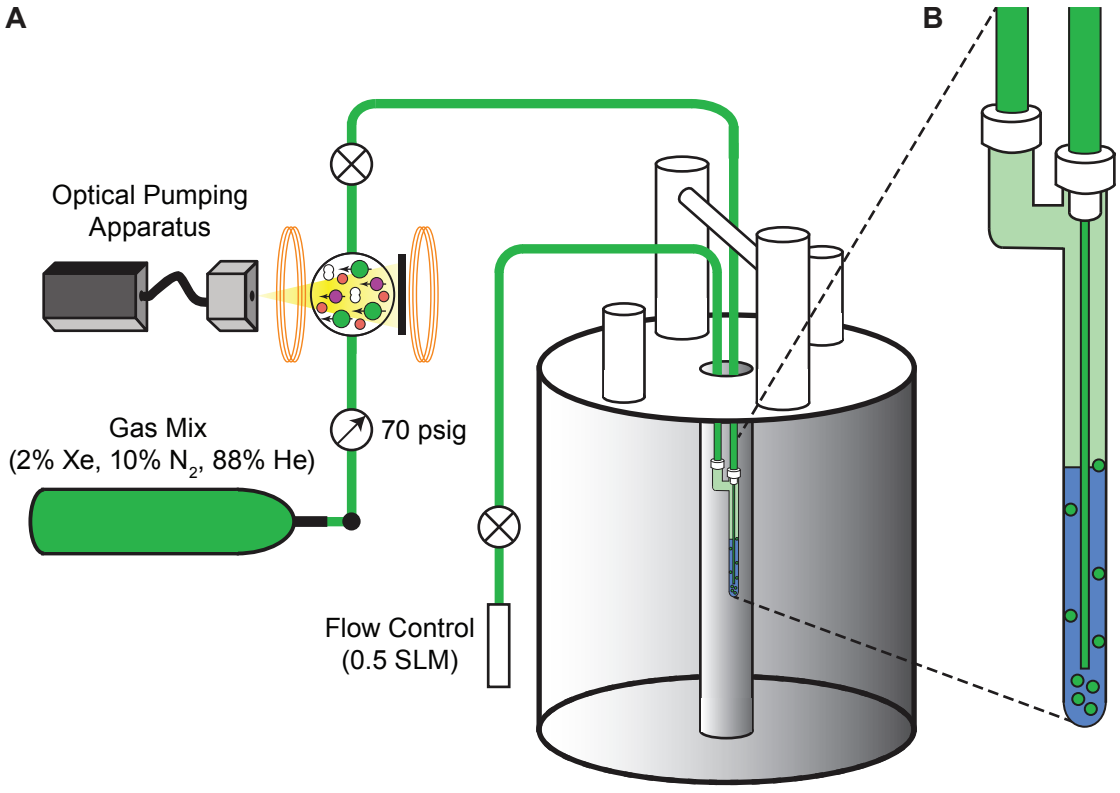



***Figure 8-3. Schematic of the NMR experimental setup allowing for continuous gas flow.***

(A) The simplified flow path of the xenon gas mixture, including polarization via a spin exchange optical pumping polarizer, introduction of the gas into the biosensor solution via bubbling through a small capillary, and valves for pausing gas flow during NMR acquisitions. (B) A zoomed in view of the bubbling setup using a modified 5 mm NMR tube allowing for continuous gas flow.

of the experiment, depicted in red circles). To minimize the potential stress that bubbling could cause the phage protein, organic solvent (5% v/v dimethyl sulfoxide) was added to reduce the solution's surface tension and propensity to foam (Figure 8-4C). While the addition of organic solvent appeared to stabilize the Xe@CryA signal of the M13-CryA biosensor, the absolute signal intensity decreased, presumably due to the low tolerance of phage solubility in some

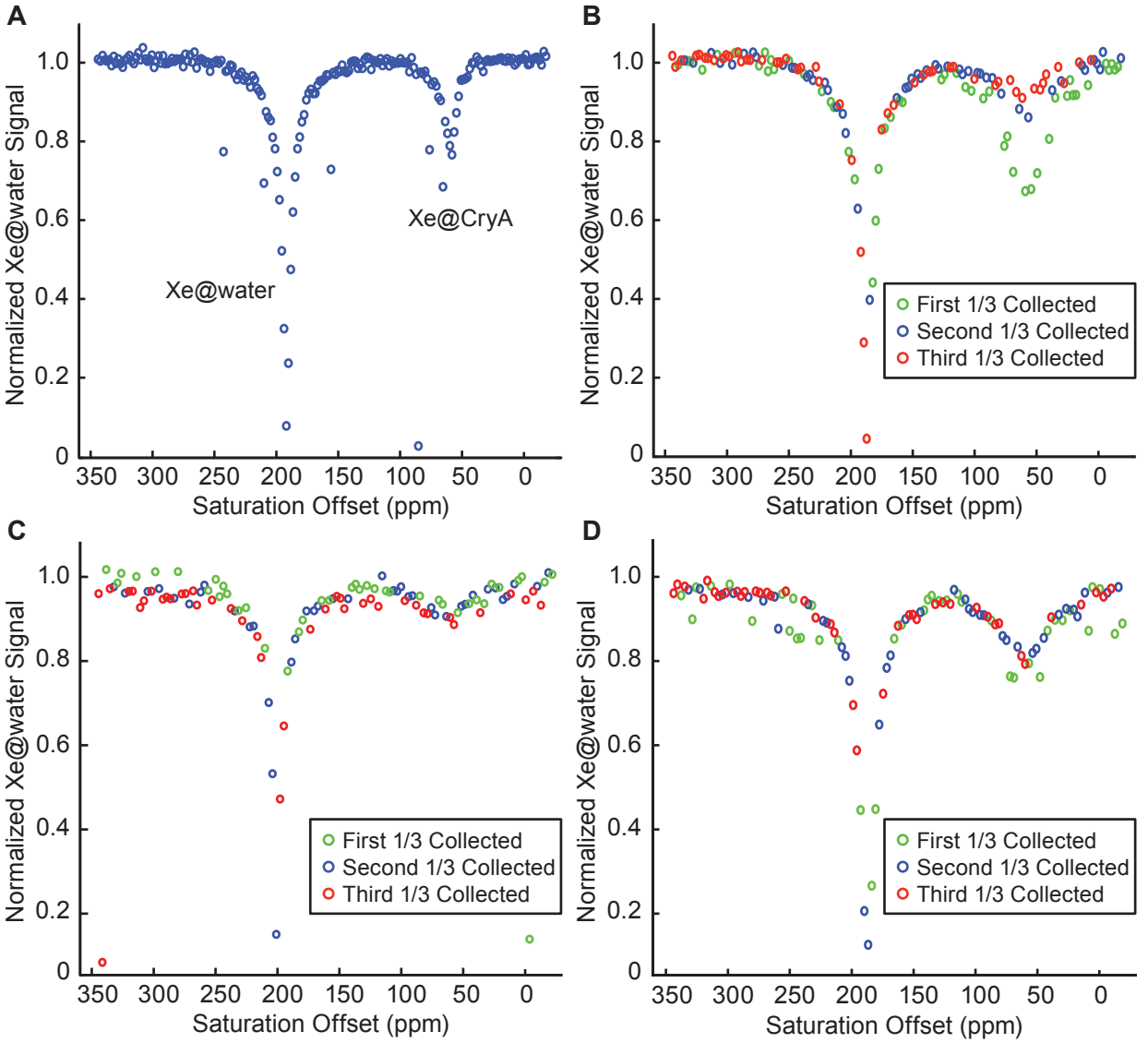



***Figure 8-4. The M13-CryA biosensor was unstable under a continuous gas flow system without addition of a hydrophilic-lipophilic surfactant.***

Comparison of the saturation response profile for a solution of (A) unscaffolded CryA-$ONH_2$ (**8.2**, 2.2 μM in CryA) in phosphate buffer and the M13-CryA biosensor (2 μM in CryA) in (B) phosphate buffer, (C) 5% DMSO in phosphate buffer, or (D) 0.002% Pluronic F-68 in phosphate buffer revealed that the M13-CryA biosensor was only stable under bubbling with the addition of Pluronic F-68. Shown are the saturation peaks corresponding to the Xe@CryA chemical shift at approximately 60 ppm and the Xe@water chemical shift at 192 ppm (referenced to the Xe gas chemical shift). As illustrated, the saturation response around the Xe@CryA peak decreases over time rather than remaining constant.

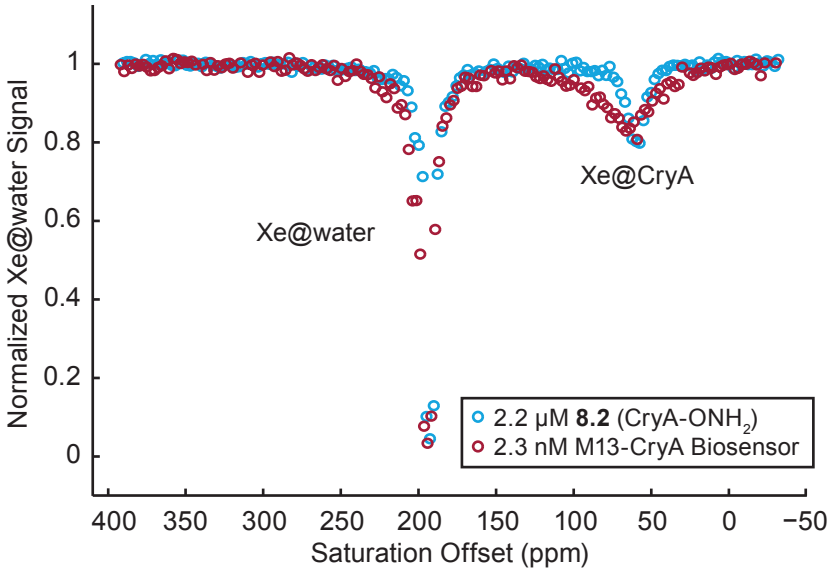



***Figure 8-5. The hyperCEST response profile of the M13-CryA biosensor was similar to a concentration matched sample of unscaffolded CryA-ONH$_2$.***

Comparison of the saturation response profile for a solution of 2.3 nM M13-CryA biosensor (2.2 μM in CryA) stabilized with 0.002% Pluronic F-68 and a CryA–concentration matched solution of unscaffolded CryA-ONH$_2$ (**8.2**, 2.2 μM) revealed nearly identical contrast. Shown are the saturation peaks corresponding to the Xe@CryA chemical shift at approximately 60 ppm and the Xe@water chemical shift at 192 ppm (referenced to the Xe gas chemical shift) acquired at 293 K.

organic solvents. After testing several additives that reduce the surface tension of water, we found that the addition of a small amount (0.002% v/v) of a low hydrophilic-lipophilic balance (HLB) surfactant (Pluronic F-68) could both minimize foaming and maintain M13-CryA biosensor stability in solution during xenon bubbling (Figure 8-4D).

$^{129}$Xe NMR measurements with filamentous bacteriophage modified with CryA were performed using a 2.3 nM solution of M13-CryA biosensor in DPBS containing 0.002% v/v Pluronic F-68. This corresponded to approximately 2.2 μM CryA, so for comparison, a 2.2 μM solution of unscaffolded CryA-ONH$_2$ was also prepared. The saturation response profile of each sample was recorded by measuring the Xe@water signal as a function of the offset frequency of a radiofrequency (RF) saturation pulse train at 293 K. This revealed the presence of any xenon populations exchanging with Xe@water (Figure 8-5). As expected, a region of saturation centered at 60 ppm, characteristic of Xe@CryA, was seen for both the M13-CryA biosensor and the unscaffolded CryA-ONH$_2$ samples. Despite a 2.4 ppm downfield shift and broadness in the Xe@CryA response of the M13-CryA biosensor compared to that of the unscaffolded CryA-ONH$_2$ (full width at half minimum was 37.6 ppm and 13.2 ppm, respectively), the maximum contrast for both curves were in good agreement (17% and 19%, respectively).

The hyperCEST contrast of the biosensor solutions was then measured as a function of saturation time using the same saturation pulse but at two different temperatures, 293 K and 310 K (Figure 8-6A). Interestingly, the chemical shifts of Xe@CryA and Xe@water were shifted downfield 2.8 ppm and 0.4 ppm, respectively, when heating the samples from 293 K to 310 K (see Table 8-1 in the Materials and Methods section). Nevertheless, the M13-CryA biosensor and the unscaffolded CryA-ONH$_2$ displayed comparable contrast in a temperature dependent

manner; contrast was generated more rapidly at the elevated temperature due to the faster chemical exchange dynamics (e.g. residual signal after 5 s of saturation was 90% and 42% for 293 K and 310 K, respectively).

To further characterize the M13-CryA biosensor, hyperCEST contrast was measured as a

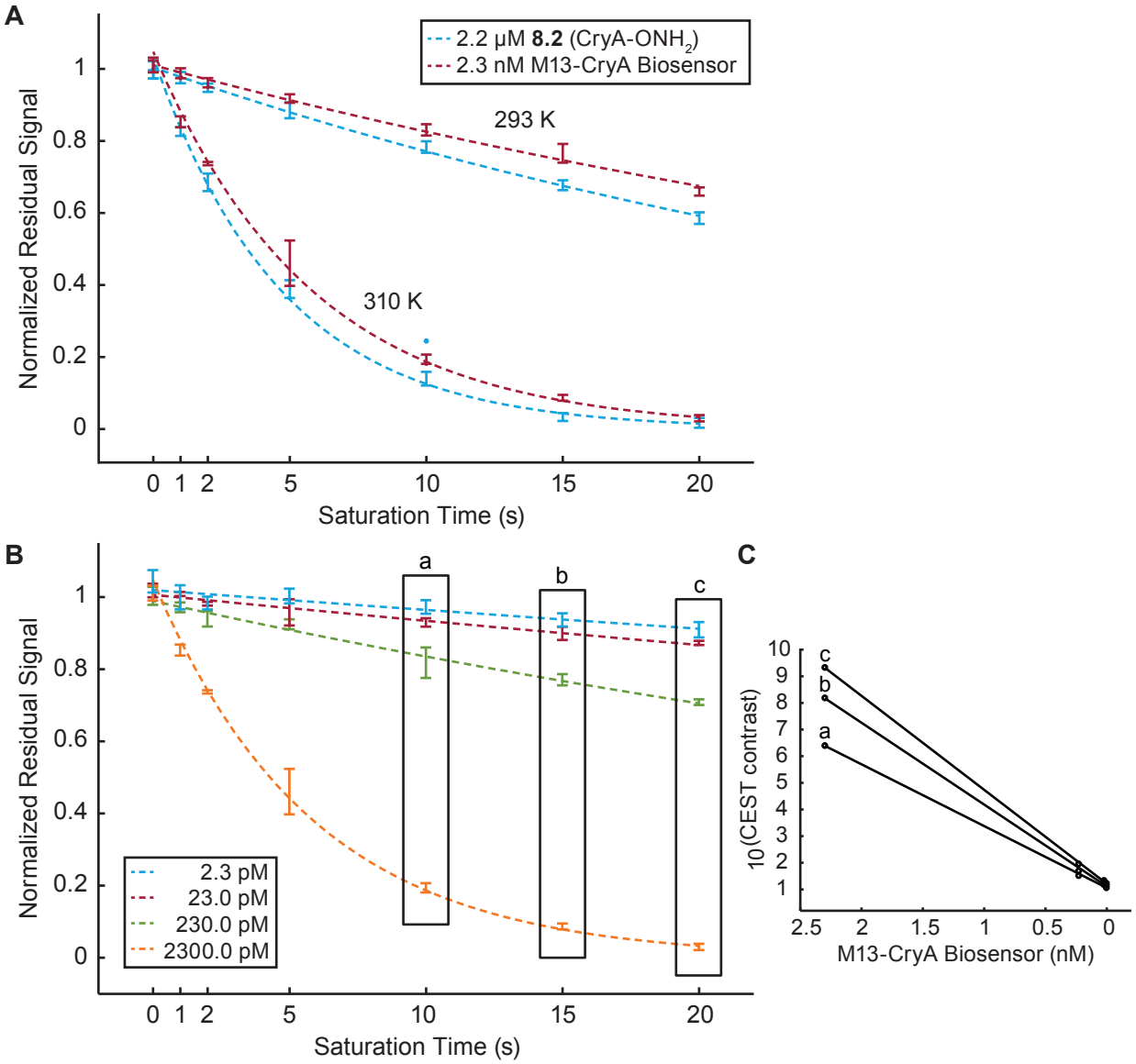



***Figure 8-6. HyperCEST contrast was measured as a function of saturation time and biosensor concentration.***

(A) The hyperCEST contrast for a solution of 2.3 nM M13-CryA biosensor and a CryA–concentration matched solution of unscaffolded CryA-$ONH_2$ (**8.2**, 2.2 μM) was measured as a function of saturation time at 293 and 310 K. Saturation was applied using a train of 500 Hz bandwidth dSNOB pulses, and the normalized residual signal was plotted (equivalent to 1 – hyperCEST contrast). Longer saturation times produced more contrast for both samples. Shown is the mean contrast (n = 4) fit to an exponential decay. Error bars represent the standard deviation. (B) The hyperCEST contrast for a dilution series of the M13-CryA biosensor was measured as a function of saturation time at 310 K. Saturation was applied using a train of 3000 Hz bandwidth dSNOB pulses, and the normalized residual signal was plotted. Shown is the mean contrast (n = 4) fit to an exponential decay. Error bars represent the standard deviation. (C) Dose-response curves showing hyperCEST contrast as a function of M13-CryA biosensor concentration for saturation times of (a) 10 s, (b) 15 s, and (c) 20 s. The response is roughly logarithmic.

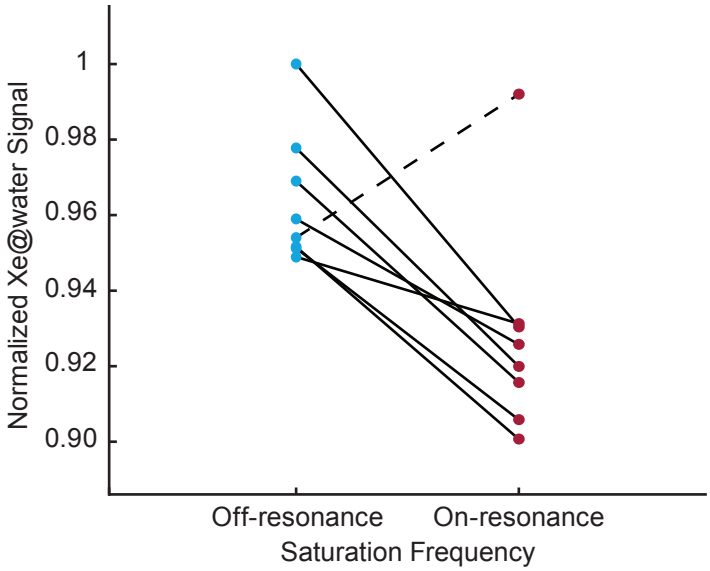



***Figure 8-7. The hyperCEST response of a 230 fM solution of the M13–CryA biosensor.***

The hyperCEST response of 230 fM M13-CryA biosensor using a train of 3000 Hz bandwidth dSNOB pulses applied at the Xe@CryA chemical shift ("on–resonance") and at an equal distance on the downfield side of the Xe@water chemical shift ("off–resonance"). An outlying on–resonance data point is shown by the dashed line. Both groupwise and paired Student's t tests showed significant contrast (4.8%, or 3.7% with inclusion of the outlier) between the data sets corresponding to the two saturation frequencies ($p < 0.01$).

function of biosensor concentration and saturation time (Figure 8-6B). For all four concentrations of the M13-CryA biosensor used in the dilution series, CEST contrast was clearly visible in the saturation time curves, and the decrease in residual signal for each concentration was accurately represented by an exponential decay with saturation time. For these very dilute solutions, the CEST contrast for a given saturation time decreased roughly logarithmically with decreasing M13-CryA concentration (Figure 8-6C).

Impressively, when a 230 fM M13-CryA biosensor solution was evaluated, statistically significant contrast was observed (Figure 8-7). Despite the presence of an outlying on-resonance data point (greater than 3 standard deviations from the group mean; Figure 8-7, right side of dashed line), a mean contrast of 3.7% was measured ($p < 0.01$ for groupwise and paired t tests). Excluding the outlying point resulted in a mean contrast of 4.8%.

The improved per-CryA sensitivity in the current study (a 230 fM M13-CryA biosensor loaded with 1050 CryA copies per M13 phage yields 242 pM CryA concentration) relative to previous reports using CW-based saturations[13] highlights the utility of using dSNOB trains with higher $B_1$ strengths for saturating exchanging xenon signals.[29] We hypothesize that the sensitivity gain may arise from semi-coherent inversions of the slowly exchanging xenon spins ($\tau_{Xe@CryA}$ is approximately 6 ms at 310 K) when using short and strong dSNOB pulses, thus increasing the dynamic range of the altered magnetization being passed back to the bulk Xe@water pool.

When considering specific absorption rate (SAR) constraints for *in vivo* applications of $^{129}$Xe-based biosensors, it is favorable to note that due to the low gyromagnetic ratio of $^{129}$Xe relative to $^{1}$H, the peak $B_1$ field strengths reported must be divided by 3.62 when comparing to SAR-equivalents of saturations applied at $^{1}$H carrier frequencies for experiments such as PARACEST and

magnetization transfer (i.e., SAR is proportional to $\omega_1^2 = (\gamma \bullet B_1)^2$, $\omega_0^2 = (\gamma \bullet B_0)^2$). Furthermore, some preliminary results suggest that lower duty cycle dSNOB saturation trains using inter-pulse delays may allow for much lower power deposition with only modest reductions in CEST contrast.[29]

## 8.3 Conclusions

This work demonstrated that M13 filamentous bacteriophage is an effective molecular scaffold for multivalent CryA assembly, resulting in a $^{129}$Xe-based contrast agent that can be detected at concentrations as low as 230 fM. This detection threshold represents the best per-agent sensitivity demonstrated for an NMR or MRI contrast agent to date. Additionally, we verified that CryA produces similar levels of hyperCEST contrast whether in M13-scaffolded or unscaffolded forms, which helps to maximize the per-agent sensitivity of the M13-CryA biosensor

The dramatic sensitivity demonstrated by the M13-CryA biosensor makes it an advantageous choice for further development towards targeted molecular MRI applications. In Chapter 9, we exploit this observation to investigate the utility and biocompatibility of antibody-targeted versions of phage-based CryA biosensors for molecular imaging applications.

## 8.4 Materials and Methods

### *8.4.1 General procedures and materials*

All chemical reagents were of analytical grade, obtained from commercial suppliers, and used without further purification unless otherwise noted. Pyridoxal 5′-phosphate (PLP) monohydrate was obtained from Aldrich. All reaction flasks were oven dried prior to use and moisture-sensitive reactions were performed in flame-dried flasks under a $N_2$ atmosphere. Analytical thin layer chromatography (TLC) was performed with 60 Å silica gel plates (EMD) and visualized by staining with ceric ammonium molybdate, ninhydrin, or by absorbance of UV light at 254 nm. Flash chromatography was performed with 60 Å 230-400 mesh silica (Silicycle). Organic extracts were dried over $MgSO_4$, and solvent was removed by reduced pressure with a R-114 rotovapor (Buchi) equipped with a self-cleaning dry vacuum pump (Welch). If necessary, products were further dried by reduced pressure with an Edwards RV5 high vacuum pump.

Water used in biological procedures was distilled and deionized using a Milli-Q Gradient ultrapure water purification system (Millipore). Dulbecco's phosphate-buffered saline (DPBS) was obtained from Sigma. Unmodified M13 bacteriophage cultures were obtained from James Marks (University of California, San Francisco).

### *8.4.2 Instrumentation*

**NMR.** All $^1$H and $^{13}$C spectra are reported in ppm and referenced to solvent peaks. Spectra were obtained on Bruker AVB-400, AVQ-400, AV-500, or AV-600 spectrometers. Coupling constants (J) are reported in Hz.

**Absorbance spectra.** Absorbance readings were collected on a Molecular Devices SpectraMax 190 UV-Vis absorbance plate reader or a Thermo Scientific Nanodrop 2000 spectropho-

tometer.

**High performance liquid chromatography (HPLC).** Purification of small molecules was performed using reversed-phase (RP) HPLC on a Rainin Dynamax SD-200 system equipped with a Varian UV-Vis detector (model 345) and a Microsorb C18 analytical column (4.6 x 250 mm) at a flow rate of 1 mL/min, a semipreparative column (10 x 250 mm) at a flow rate of 4 mL/min, or a preparative column (21.4 x 250 mm) at a flow rate of 20 mL/min. HPLC samples were filtered with a Pall Life Sciences Acrodisc CR 13 mm syringe filter equipped with a 0.2 μm PTFE membrane prior to injection.

Modified and unmodified M13 phage were analyzed by RP-HPLC on an Agilent 1100 Series HPLC System (Agilent Technologies) with an inline diode array detector (DAD) and an Zorbax 300SB-CN column (Agilent) at a flow rate of 1.0 mL/min. Samples were centrifuged prior to injection to pellet insoluble material.

**Matrix assisted laser desorption-ionization time-of-flight (MALDI-TOF) MS.** MALDI-TOF MS was performed on a Voyager-DE system from PerSeptive Biosystems (Applied Biosystems). Samples were co-crystallized using saturated solutions of sinapinic acid in 60% acetonitrile, 40% water with 0.1% trifluoroacetic acid.

**Electrospray ionization (ESI) MS.** High-resolution ESI mass spectra were obtained at the QB3 Mass Spectrometry Facility (UC Berkeley) using an LTQ Orbitrap XL hybrid mass spectrometer (Thermo Fisher). Low-resolution ESI mass spectra were obtained using an LTQ-XL mass spectrometer.

Liquid chromatography coupled to mass spectrometry (LC-MS) was performed with an Agilent 1100 Series HPLC (Agilent Technologies) system that was connected in-line with an API 150EX mass spectrometer (Applied Biosystems). The mass spectrometer was outfitted with a Turbo Ionspray ionization source. A binary solvent system consisting of buffer A (0.1% formic acid in water (v/v)) and buffer B (0.1% formic acid in acetonitrile (v/v)) was employed. The LC was performed using a Zorbax 300 SB-CN column (Agilent).

**$^{129}$Xe NMR.** All $^{129}$Xe NMR data were acquired on a 7.05 T vertical bore spectrometer (Varian) equipped with a 26 channel shim set and a dual-tuned ($^{1}$H/$^{129}$Xe) radiofrequency saddle coil (i.d. = 5 mm). $^{129}$Xe was hyperpolarized (P ≈ 5%) via spin-exchange optical pumping with a MITI XenoSpin polarizer (GE Healthcare, formerly Nycomed Amersham). The gas mixture for polarization consisted of 2% xenon (natural abundance), 10% nitrogen, and 88% helium. After polarization, gas flowed through plastic tubing directly into a 5 mm NMR tube outfitted with gas inlet and outlet arms (Figure 8-3). The NMR tube was pressurized to 584 kPa with the polarized gas mixture. Prior to each pulse sequence repetition, the gas mixture was bubbled at a total flow rate of approximately 0.5 SLM to saturate the solution with hyperpolarized $^{129}$Xe, and the solution was allowed to settle to clear any bubbles before data were collected. The gas transit time from the polarizer was minimized by diverting approximately 90% of the total gas flow from the polarizer through a bypass just prior to the capillary input. All data were collected with

a 25 kHz spectral width and an acquisition time of 0.5 s. Saturation consisted of a train of dSNOB pulses[30] without inter-pulse delays.[29]

*8.4.3 Chemical synthesis*

**Aminooxy-functionalized 5 kDa polyethylene glycol (8.1)** has been reported.[27]

**Aminooxy-functionalized CryA peptide (8.2).** First, a water solubilizing tetraglutamic acid peptide (**7.7**) was prepared on polystyrene Wang resin (1.2 mmol/g, 100–200 mesh) using standard Fmoc solid-phase peptide synthesis (SPPS) with $N_\alpha$-Fmoc protected amino acids, DIPEA, and HOBT/HBTU ester activation in DMF. A 5-fold excess of amino acid was used for the coupling steps, which involved gentle rotation for 3 h. Fmoc removal was achieved with 20% piperidine in DMF (2 x 10 min). Next, the aminooxy functionality was installed by coupling **7.16** using the above coupling conditions for 5 h to generate compound **7.17**. A linker was subsequently installed using Fmoc-6-aminohexanoic acid (Fmoc-Ahx-COOH) and the same coupling conditions as above to give compound **8.3**. Finally, the peptide was capped at the N-terminus using a 2.5 fold excess of CryA (**6.1**) using the above coupling conditions for 12-20 h. The product was cleaved using a cocktail of TFA:water:triisopropylsilane (95:2.5:2.5) for 2.5 h and then precipitated in methyl *tert*-butyl ether. The crude product was purified by RP-HPLC using a C18 column to yield peptide **8.2**. A stock solution of the aminooxy-functionalized CryA-peptide was prepared by dissolving the purified solid in DPBS and adjusting the pH with sodium hydroxide until the solution became clear. ESI-MS(–) calculated for $C_{89}H_{108}N_8O_{30}$ $(M\text{-}3H)^{3-}$: 588.57 Da found 588.49 Da $(M\text{-}3H)^{3-}$.

*8.4.4 Experimental*

**Expression and purification of M13 bacteriophage.** The expression and purification of M13 bacteriophage were performed according to established procedures.[31-33] Phage concentration was determined by measuring the absorbance and calculating according to Equation 8-1:[34]

$$\text{phage concentration (per mL)} = \frac{\left(\text{Abs}_{269\text{nm}} - \text{Abs}_{320\text{nm}}\right) \bullet 6 \times 10^{16}}{\#\text{ base pairs in the phage genome}} \qquad (8\text{-}1)$$

**Transamination of Bacteriophage M13**. Bacteriophage M13 were transaminated to introduce ketone groups at the N-termini[25] of their coat proteins using the following final reaction concentrations: 65 nM M13 and 100 mM pyridoxal 5'-phosphate in 150 mM phosphate buffer, pH 6.3. After 20 h at RT, the resulting transaminated M13 (k-M13) were purified through a series of precipitations and resuspensions in DPBS to remove excess PLP. First, the solution of k-M13 was diluted to 10 mL with DPBS and precipitated for 1 h at 4 °C after the addition of 0.15 volumes of a solution of 20% 8 kDa polyethylene glycol polymer (PEG-8k) and 2.5 M NaCl and vigorous mixing. The solution was centrifuged at 9,000 rpm for 20 min. The phage pellet was resuspended in 30 mL DPBS and then the precipitation, centrifugation, and resuspension was repeated three more times, before resuspending in a final volume of 1-2 ml of DPBS. The concentration of k-M13 was calculated by measuring the absorbance as described in Equation 8-1.

**M13-CryA biosensor construction.** Bacteriophage M13 were modified with CryA cages following established procedures for oxime formation on k-M13.[22] After transamination, oxime formation first proceeded between k-M13 and aminooxy-functionalized 5 kDa polyethylene glycol polymer (PEG-5k-$ONH_2$, **8.1**) using the following final reaction concentrations: 20 nM k-M13, 20 mM **8.1**, and 10 mM aniline in 150 mM phosphate buffer, pH 6.0. After 1 h at RT, the resulting M13-PEG construct was purified by gel filtration using an Illustra Nap-5 column (GE Healthcare). The concentration was calculated by measuring the absorbance as described in Equation 8-1, and RP-HPLC quantified the extent of PEG-5k labeling. Next, the M13-PEG construct was further modified with aminooxy-functionalized CryA peptide (**8.2**) using the following final reaction concentrations: 13 nM M13-PEG, 240 μM **8.2**, and 100 mM aniline in 150 mM phosphate buffer, pH 5.5. After 22 h at RT, the resulting M13-PEG-CryA construct, termed the M13-CryA biosensor, was purified by gel filtration using an Illustra Nap-5 column. The concentration was calculated by measuring the absorbance as described in Equation 8-1, and RP-HPLC quantified the extent of CryA labeling. Purified M13-CryA biosensors were stored in DPBS at 4 °C until needed. ESI-HRMS(+): 6988.87 Da (avg. mass), found 6988.46 Da (avg. mass).

**Characterization of M13 phage by RP-HPLC.** Approximately 0.5 nmol of M13 phage were analyzed by RP-HPLC using a binary solvent system consisting of buffer A (0.1% trifluoroacetic acid in water (v/v)) and buffer B (0.1% trifluoroacetic acid in acetonitrile (v/v)). The elution gradient was 35% buffer B for the first 4 min, ramping to 70% buffer B over 15 min, then to 95% buffer B over the next 1 min, and a 10 min wash with 95% buffer B at a flow rate of 1 mL/min.

**Characterization of M13 phage by LC-MS.** Approximately 0.2 nmol of M13 phage were analyzed by LC-MS using a binary solvent system consisting of buffer A (0.1% formic acid in water (v/v)) and buffer B (0.1% formic acid in acetonitrile (v/v)). The elution gradient was 35% buffer B for the first 4 min, ramping to 70% buffer B over 15 min, then to 95% buffer B over the next 1 min, and a 10 min wash with 95% buffer B, at a flow rate of 250 μL/min. Extracted ion chromatograms were deconvoluted with Analyst software (Applied Biosystems) to generate protein mass reconstructions.

**Preparation of M13-CryA biosensor solutions for $^{129}$Xe NMR.** All M13-CryA phage solutions were prepared in 650 μL of DPBS, DPBS containing 5% dimethyl sulfoxide, or DPBS containing 0.002% v/v Pluronic (BASF, Edison NJ), a low hydrophilic-lipophilic balance (HLB) surfactant.

**hyperCEST settings for $^{129}$Xe NMR.** Saturation response profiles were recorded by measuring the Xe@water signal as a function of the offset frequency of the saturation pulse train. The offset frequency was varied over a range of approximately -32 ppm to 390 ppm centered about the Xe@water chemical shift. The acquisition order was randomized to allow for retrospective correction of linear drift of hyperpolarized $^{129}$Xe signals where required. Saturations consisted of a train of shaped RF pulses, known as dSNOB pulses. For the saturation response profile, a train of 500 Hz bandwidth was used (the pulse length ($\tau_{pulse}$) was 5.64 ms, the maximum RF intensity ($B_{1,max}$) was 59 μT, and the total saturation time ($t_{sat}$) was 10 s).

For saturation time curves, the solutions were bubbled for 25 s with a wait period of 2 s. A train of 3000 Hz bandwidth dSNOB pulses was applied ($\tau_{pulse}$ = 940 μs, $B_{1,max}$ = 352 μT), and the number of pulses was varied to change the total saturation time ($t_{sat}$ = 0, 1, 2, 5, 10, 15, 20 s). Contrast was quantified by comparing the Xe@water signal from two separate saturations, one after application of the pulse train on-resonance at the frequency of Xe@CryA, and the second after application equidistant from, but on the opposite side of the Xe@water peak (off-resonance). These frequencies were found to vary modestly due to the association with M13 (Xe@CryA) as well as with temperature (Xe@CryA and Xe@water); the exact frequencies used in these experiments are presented in Table 8-1. Given the differences in saturation times, a variable post-acquisition delay was employed to maintain a constant repetition time (TR). Furthermore, it should be noted that undesirable $T_1$ weightings were already absent from the experiment due to the use of refreshed hyperpolarized signals. Plots of the normalized residual signal, equal to (1 - hyperCEST contrast), were generated, illustrating the magnitude of the hyperCEST effect alone. Furthermore, the CEST dynamics were compared at both 293 K and 310 K.

To investigate the M13-biosensor concentration dependence of CEST contrast on and determine a detection threshold, plots of residual signal versus saturation time were generated for a dilution series of the biosensor (2.3 nM, 230 pM, 23 pM, 2.3 pM) at 310 K. The hyperCEST saturation and NMR acquisition parameters were identical to those described above for the saturation time curves. Finally, a 230 fM M13-CryA biosensor solution was evaluated as the lower limit concentration using a single saturation duration of 20 s. To minimize $^{129}$Xe signal variability from acquisition to acquisition, long bubbling and wait times of 40 s and 10 s, respectively, were used.

Preliminary experiments using unmodified M13 were used to confirm the absence of exchangeable signals in the $^{129}$Xe saturation response profile due to the phage itself. Additionally, experiments in solvent alone were used to confirm the absence of baseline asymmetries in the Xe@water signal due to the application of 500 Hz and 3000 Hz bandwidth dSNOB pulse trains (peak $B_1$ and train durations matched to those used in the respective agent experiments) at the on-resonance and off-resonance frequencies (e.g. > 10000 Hz from the Xe@water frequency).

***Table 8-1. The relevant frequency offsets used for hyperCEST measurements.***
"On" and "off" refer to the on–resonance and off–resonance frequencies at which RF saturation was applied (referenced to the $^{129}$Xe gas chemical shift).

| Temperature | Xe@water | Xe@CryA (unscaffolded) | Xe@CryA (scaffolded) |
|---|---|---|---|
| 293 K | 192.0 ppm | 59.4 ppm (on)<br>324.6 ppm (off) | 61.8 ppm (on)<br>322.2 ppm (off) |
| 310 K | 192.4 ppm | 62.2 ppm (on)<br>322.6 ppm (off) | 64.6 ppm (on)<br>320.2 ppm (off) |

### *8.4.5 Data analysis*

**NMR data analysis.** Data processing was performed in MATLAB (version 7.13, R2011b, The MathWorks). All data were line broadened with an 8 Hz Lorentzian and fitted in the time domain using a nonlinear least-squares regression, producing signal amplitudes for subsequent analysis.

Contrast for any given saturation time was quantified as follows: the Xe@water signal after saturating at the chemical shift of Xe@CryA, $I_{on\text{-}res}$, was subtracted from the Xe@water signal after a control saturation, $I_{off\text{-}res}$, at a frequency separated by an equal distance but on the opposite side of the Xe@water peak. This value, in turn, was normalized to $I_{off\text{-}res}$ to yield the contrast value.

For analysis of the 230 fM M13-CryA biosensor solution, on-resonance and off-resonance saturation data were analyzed with group wise and paired (one-tailed) Student's t-tests for statistical significance.

## 8.5 References

1. Caravan P., Ellison J.J., McMurry T.J. and Lauffer R.B. (**1999**) Gadolinium(III) Chelates as MRI Contrast Agents: Structure, Dynamics, and Applications. *Chem Rev* 99, 2293-2352.

2. Woods M., Woessner D.E. and Sherry A.D. (**2006**) Paramagnetic lanthanide complexes as PARACEST agents for medical imaging. *Chem Soc Rev* 35, 500-511.

3. Bulte J.W. and Kraitchman D.L. (**2004**) Iron oxide MR contrast agents for molecular and cellular imaging. *NMR Biomed* 17, 484-499.vv

4. Chen J., Lanza G.M. and Wickline S.A. (**2010**) Quantitative magnetic resonance fluorine imaging: today and tomorrow. *Wiley Interdiscip Rev Nanomed Nanobiotechnol* 2, 431-440.

5. Ardenkjaer-Larsen J.H. *et al.* (**2003**) Increase in signal-to-noise ratio of > 10,000 times in liquid-state NMR. *Proc Natl Acad Sci U S A* 100, 10158-10163.

6. Guenther D., Hanisch G. and Kauczor H.U. (**2000**) Functional MR imaging of pulmonary ventilation using hyperpolarized noble gases. *Acta Radiol* 41, 519-528.

7. Kuhn L.T., Bommerich U. and Bargon J. (**2006**) Transfer of parahydrogen-induced hyperpolarization to 19F. *J Phys Chem A* 110, 3521-3526.

8. Kurhanewicz J., Bok R., Nelson S.J. and Vigneron D.B. (**2008**) Current and potential applications of clinical 13C MR spectroscopy. *J Nucl Med* 49, 341-344.

9. Oros A.M. and Shah N.J. (**2004**) Hyperpolarized xenon in NMR and MRI. *Physics in Medicine and Biology* 49, R105-153.

10. Dobrucki L.W. and Sinusas A.J. (**2010**) PET and SPECT in cardiovascular molecular imaging. *Nat Rev Cardiol* 7, 38-47.

11. Spence M.M. *et al.* (**2001**) Functionalized xenon as a biosensor. *Proc Natl Acad Sci U S A* 98, 10654-10657.

12. Schroder L., Lowery T.J., Hilty C., Wemmer D.E. and Pines A. (**2006**) Molecular imaging using a targeted magnetic resonance hyperpolarized biosensor. *Science* 314, 446-449.

13. Schroder L. *et al.* (**2008**) Temperature response of 129Xe depolarization transfer and its application for ultrasensitive NMR detection. *Phys Rev Lett* 100, 257603.

14. Gabard J. and Collet A. (**1981**) Synthesis of a (D3)-Bis(Cyclotriveratrylenyl) Macrocage by Stereospecific Replication of a (C3)-Subunit. *J Chem Soc Chem Comm*, 1137-1139.

15. Mynar J.L., Lowery T.J., Wemmer D.E., Pines A. and Frechet J.M. (**2006**) Xenon biosensor amplification via dendrimer-cage supramolecular constructs. *J Am Chem Soc* 128, 6334-6335.

16. Meldrum T. *et al.* (**2010**) A xenon-based molecular sensor assembled on an MS2 viral capsid scaffold. *J Am Chem Soc* 132, 5936-5937.

17. Liu B., Conrad F., Cooperberg M.R., Kirpotin D.B. and Marks J.D. (**2004**) Mapping tumor epitope space by direct selection of single-chain Fv antibody libraries on prostate cancer cells. *Cancer Res* 64, 704-710.

18. Marks C. and Marks J.D. (**1996**) Phage libraries--a new route to clinically useful antibodies. *N Engl J Med* 335, 730-733.

19. Li K. *et al.* (**2010**) Chemical modification of M13 bacteriophage and its application in cancer cell imaging. *Bioconjug Chem* 21, 1369-1377.

20. Stephanopoulos N. and Francis M.B. (**2011**) Choosing an effective protein bioconjugation strategy. *Nat Chem Biol* 7, 876-884.

21. Witus L.S. and Francis M.B. (**2011**) Using synthetically modified proteins to make new materials. *Acc Chem Res* 44, 774-783.

22. Carrico Z.M. *et al.* (**2012**) N-Terminal labeling of filamentous phage to create cancer marker imaging agents. *ACS Nano* 6, 6675-6680.

23. Yacoby I. and Benhar I. (**2008**) Targeted filamentous bacteriophages as therapeutic agents. *Expert Opin Drug Del* 5, 321-329.

24. Merzlyak A., Indrakanti S. and Lee S.W. (**2009**) Genetically engineered nanofiber-like viruses for tissue regenerating materials. *Nano Lett* 9, 846-852.

25. Gilmore J.M., Scheck R.A., Esser-Kahn A.P., Joshi N.S. and Francis M.B. (**2006**) N-terminal protein modification through a biomimetic transamination reaction. *Angew Chem Int Edit* 45,

5307-5311.

26. Witus L.S. *et al.* (**2010**) Identification of Highly Reactive Sequences For PLP-Mediated Bioconjugation Using a Combinatorial Peptide Library. *J Am Chem Soc* 132, 16812-16817.

27. Schlick T.L., Ding Z., Kovacs E.W. and Francis M.B. (**2005**) Dual-surface modification of the tobacco mosaic virus. *J Am Chem Soc* 127, 3718-3723.

28. Han S.I. *et al.* (**2005**) NMR-based biosensing with optimized delivery of polarized 129Xe to solutions. *Anal Chem* 77, 4008-4012.

29. Meldrum T., Bajaj V.S., Wemmer D.E. and Pines A. (**2011**) Band-selective chemical exchange saturation transfer imaging with hyperpolarized xenon-based molecular sensors. *J Magn Reson* 213, 14-21.

30. Kupce E., Boyd J. and Campbell I.D. (**1995**) Short selective pulses for biochemical applications. *J Magn Reson B* 106, 300-303.

31. Amersdorfer P. *et al.* (**2002**) Genetic and immunological comparison of anti-botulinum type A antibodies from immune and non-immune human phage libraries. *Vaccine* 20, 1640-1648.

32. Zhou Y. *et al.* (**2007**) Impact of single-chain Fv antibody fragment affinity on nanoparticle targeting of epidermal growth factor receptor-expressing tumor cells. *J Mol Biol* 371, 934-947.

33. Zhou Y. and Marks J.D. (**2012**) Discovery of internalizing antibodies to tumor antigens from phage libraries. *Methods Enzymol* 502, 43-66.

34. Burton D.R., Scott J.K. and Silverman G.J. (**2001**) Phage Display: A Laboratory Manual, Edn. 1st. (Cold Spring Harbor Laboratory Press.

35. Stevens T.K. *et al.* (**2012**) HyperCEST detection of a (129) Xe-based contrast agent composed of cryptophane-A molecular cages on a bacteriophage scaffold. *Magn Reson Med*.

# Chapter 9: Molecular Imaging of Cancer Cells Using a Filamentous Bacteriophage-based $^{129}$Xe Biosensor

## 9.1 Introduction

The accurate detection and localization of clinically-relevant biomarkers *in vivo* is a great challenge for molecular imaging, requiring both high sensitivity and molecular specificity.[1] This is particularly true in screening and applications involving the early detection of diseased tissue, where the ability to image disease progression non-invasively could improve patient outcome. Magnetic resonance imaging (MRI) is a ubiquitous, non-invasive imaging technique with sub-millimeter spatial resolution,[2] but its use in molecular imaging has been limited by its poor sensitivity when applied to image molecules other than water.[1,3] This has led to the development of contrast agents and MRI methods that improve sensitivity by modulating the

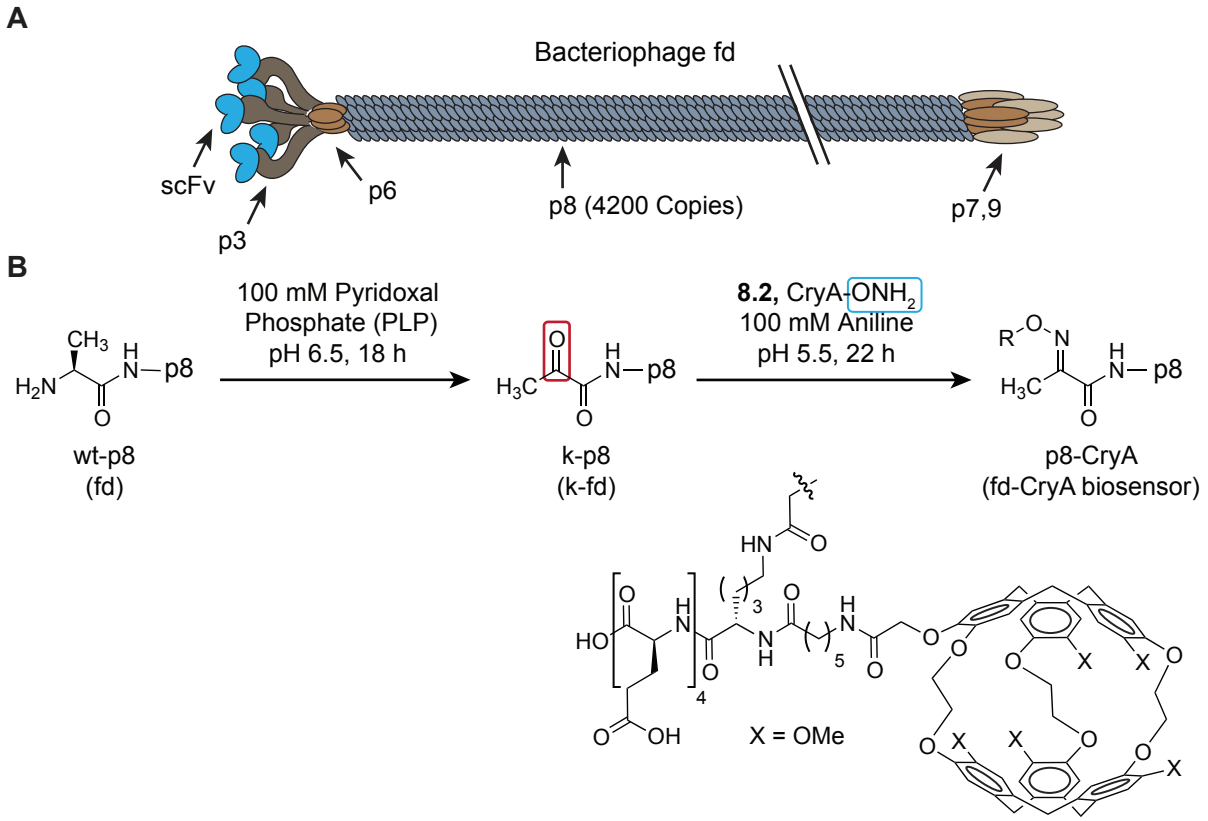



***Figure 9-1. Structure and construction of an fd-CryA biosensor.***

(A) Cartoon of fd filamentous bacteriophage with significant proteins labeled with arrows. Fd phage display single chain antibody variable fragments (scFvs, blue hearts) on their minor coat protein (p3). The double slash indicates that the phage is much longer than shown when scaled to the minor coat proteins. (B) Chemical scheme illustrating the transamination and attachment of synthetic molecules to fd phage. The N–termini of the p8 proteins (blue ovals) were transaminated to yield thousands of ketones along the length of the capsids. The ketones were then reacted with aminooxy–functionalized Cryptophane A cages (CryA–ONH$_2$, **8.2**) to generate the fd–CryA biosensor.

Matt Ramirez contributed to the work presented in this chapter.

local magnetic environment of protons in water, including gadolinium chelators,[4] iron-oxide particles,[5] and chemical exchange saturation transfer (CEST).[6] More recently, approaches that utilize hyperpolarized (hp) $^{13}C$, $^{3}He$, and $^{129}Xe$ have been developed and used in clinical studies.[7-10]

In this Chapter, we expand upon the work from Chapters 7 and 8 by utilizing viral capsid-based xenon biosensors for targeted imaging of cancer cells. Biosensors consist of a xenon-binding host molecule, commonly cryptophane-A (CryA),[11,12] attached to targeting groups for localization.[13] Initially, studies with xenon biosensors measured biological binding events in solution. This was first achieved with biotin-functionalized biosensors binding streptavidin beads,[14-17] and subsequently with the detection of DNA hybridization,[18] enzymatic cleavage by matrix metalloproteinase-7,[19] ligand binding to human carbonic anhydrase[20] and an $\alpha_{2b}\beta_3$ integrin,[21] and peptide complex formation with a major histocompatibility complex protein.[22] Once the cellular compatibility of xenon biosensors was established,[21,23] xenon NMR was performed with cells after targeting them with micromolar concentrations of a transferrin-functionalized biosensor.[24] While that study detected biosensor binding by measuring the Xe@CryA chemical shift, non-specific binding was also observed.

Towards improving upon this work, we produced a multivalent xenon biosensor that utilized antibody fragments to target cell surface biomarkers. We further sought to demonstrate its ability to specifically recognize these biomarkers in living cells at concentrations required for

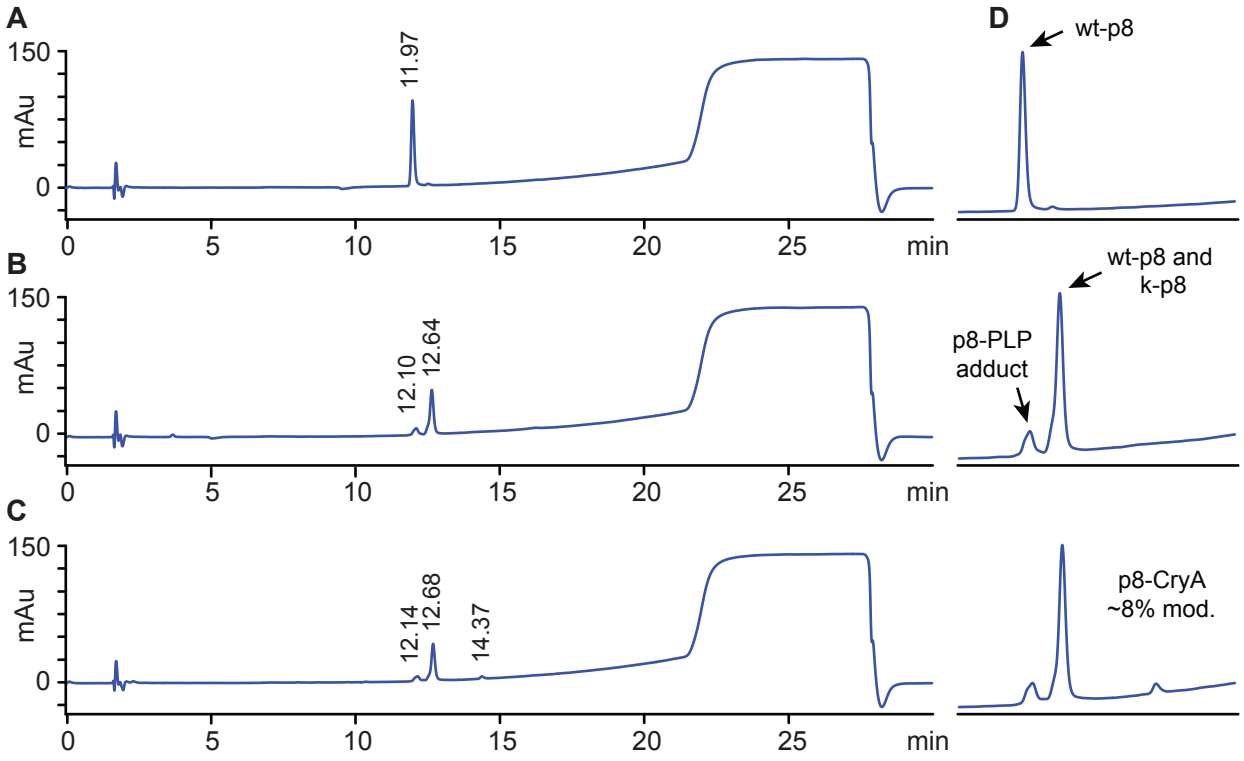



***Figure 9-2. Characterization of modified M13 phage by RP-HPLC.***

(A) Wt fd, (B) ketone-labeled fd (k-fd), and (C) fd-CryA phage were denatured with guanidinium hydrochloride and subjected to RP-HPLC analysis. Only the major coat protein, p8, could be detected. (D) An inset for each chromatogram, focused on the p8 elution region, revealed the p8-CryA modification (retention time of 14.37 min). The p8-PLP adduct resulted from an aldol reaction.[42]

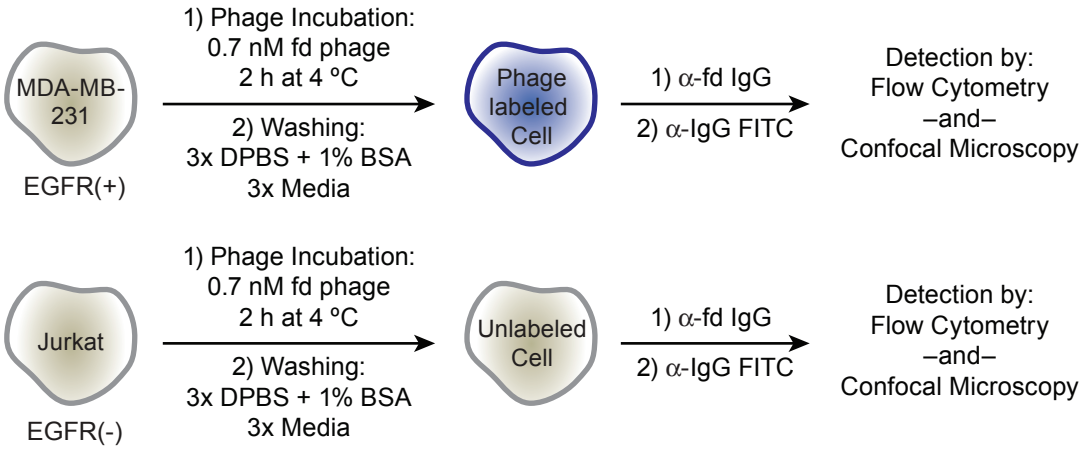



***Figure 9-3. Schematic of cell labeling with fd phage for flow cytometry and confocal microscopy experiments.***
To evaluate phage binding, MDA-MB-231 (EGFR positive) and Jurkat (EGFR negative) cells were incubated with fd phage for 2 h at 4 ºC. After removing unbound phage, those that remain bound to cells were detected using anti-fd bacteriophage antibody and analyzed by flow cytometry and confocal microscopy.

molecular imaging. To accomplish this, we used fd filamentous bacteriophage that display single chain antibody variable fragments (scFvs) on their minor coat proteins (p3, Figure 9-1A).[25] The rod-like body of fd phage, which has 4,200 identical copies of the major coat protein (p8), can be modified with proteins or synthetic molecules to create new materials.[26-28] Additionally, through the use of phage display techniques, filamentous phage that display proteins as extensions of their coat proteins can be evolved to bind small molecules, proteins, and metal ions with high affinity and selectivity.[29] In this case, the scFvs recognized either the epidermal growth factor receptor (EGFR) or, as a negative control, botulinum toxin serotype A (BONT), with nanomolar to subnanomolar affinity.[30-32] Our biosensors were designed to target EGFR because it is a cell surface receptor that is highly expressed in a variety of solid tumors, including breast cancer.[33] EGFR expression levels can also be used to predict cancer progression and sometimes survival rates,[34] making it an ideal cellular target for both therapeutics and molecular imaging applications.

## 9.2 Results and Discussion

### *9.2.1 Construction of an fd-CryA biosensor*

Synthetic components can be introduced onto the surface of filamentous phage through several bioconjugation strategies.[27,35,36] Recently, we applied an N-terminal specific bioconjugation method for modifying filamentous phage containing scFvs with fluorophores and polymers and verified that modified phage maintained their cell binding properties.[28,37,38] Additionally, we used the same chemistry to modify non-targeted M13 filamentous phage with CryA to test the feasibility of using filamentous phage as xenon biosensors, and established that there was no loss in sensitivity of phage-grafted CryA groups compared to equivalent concentrations of free CryA.[39] Based on this success, we used the same N-terminal specific chemistry to construct a targeted fd-based xenon biosensor (Figure 9-1B). First, the N-termini of the coat

proteins were transaminated using pyridoxal 5′-phosphate (PLP) to introduce ketone groups. After the ketone-labeled fd phage (k-EGFR and k-BONT) were purified, oxime ligation proceeded with an aminooxy-functionalized CryA-peptide (**8.2**). The peptide appendage was necessary to achieve aqueous solubility for CryA (see Chapter 7 for synthesis). After 22 h at RT, the resulting fd-CryA biosensors (i.e., EGFR-CryA and BONT-CryA) were purified by gel filtration and characterized by reversed-phase HPLC (Figure 9-2). On average, the biosensors were ~8% modified with CryA cages, corresponding to ~330 copies/phage.

### *9.2.2 Characterization of the fd-CryA biosensor cell binding by flow cytometry and confocal microscopy*

To verify that fd-CryA biosensors retained their binding specificity, the modified phage were incubated with either EGFR positive or negative cells and characterized by flow cytometry and confocal microscopy (Figure 9-3). First, anti-EGFR, k-EGFR, and EGFR-CryA phage, along with their anti-BONT phage controls, were incubated for 2 h at 4 °C with either MDA-MB-231 cells (a breast cancer cell-line with high EGFR expression) or Jurkat cells (T-cells with low EGFR expression). The bound phage were fluorescently stained using an anti-fd bacteriophage antibody and analyzed by flow cytometry (Figure 9-4). All anti-EGFR phage constructs displayed high binding specificity to MDA-MB-231 cells, with little-to-no non-specific binding to Jurkat cells. Additionally, the negative control anti-BONT phage constructs showed low binding to MDA-MB-231 cells, indicating that the CryA modification did not produce significant non-spe-

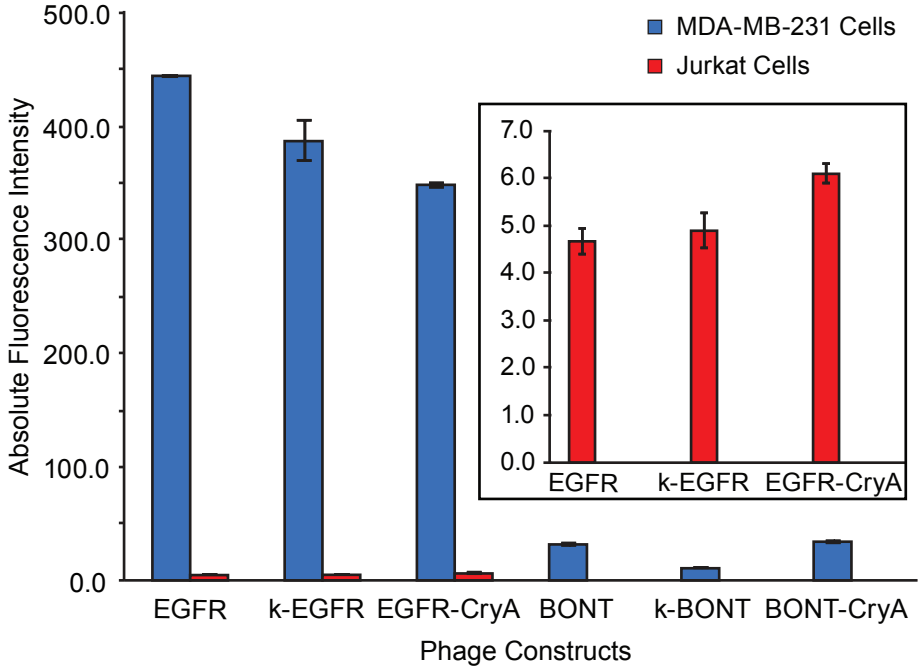



***Figure 9-4. EGFR-CryA phage maintained their cell-binding specificity.***

Flow cytometry indicated that all three EGFR phage constructs bound similarly to MDA-MB-231 cells (EGFR positive), but they did not bind to Jurkat cells (EGFR negative). In contrast, all three BONT phage constructs did not bind to MDA-MB-231 cells, indicating that the CryA modification did not produce non-specific binding. The inset is an enlargement of the Jurkat cell binding data. Shown is a representative experiment from three biological replicates, each performed in triplicate. Error bars represent the standard deviation.

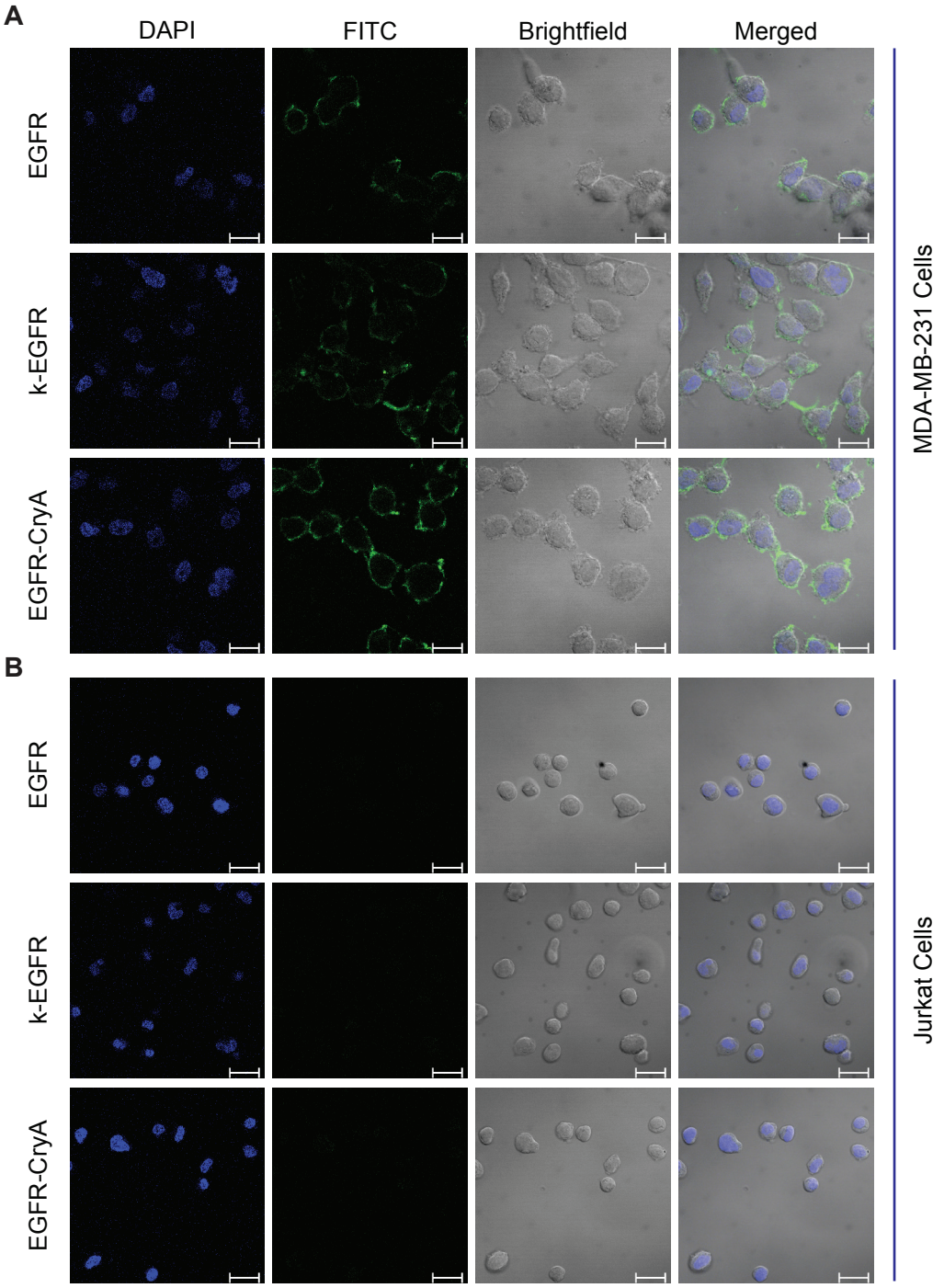



***Figure 9-5. EGFR-CryA phage bound the cell surface of MDA-MB-231 cells.***

(A) Confocal microscopy revealed that all three EGFR phage constructs bound to MDA-MB-231 cell surfaces, but they did not bind to (B) Jurkat cells. Phage constructs were detected with an anti-fd bacteriophage antibody. Channels for nuclear staining (blue, DAPI) and phage labeling (green, FITC) along with the brightfield and merged images are shown; scale bars are 20 µm.

cific binding. Next, we used confocal microscopy to determine how the EGFR-CryA biosensor bound the MDA-MB-231 cells (Figure 9-5). Fluorescence microscopy revealed predominantly cell-surface binding with minimal cell internalization or non-specific binding. Having confirmed the binding specificity of the EGFR-CryA biosensor, we proceeded to characterize the hyperCEST response of biosensor-labeled cells with $^{129}$Xe NMR.

### *9.2.3 Characterization of the EGFR-CryA biosensor cell binding by $^{129}$Xe NMR*

Xenon gas was introduced into cell solutions using a continuous-flow bubbling system.[40] The output from a xenon polarizer was connected to a 5-mm NMR tube modified with inlet and outlet arms, allowing for direct delivery of hp $^{129}$Xe into the solution via a capillary running to the bottom of the NMR tube (illustrated in Figure 9-6B). While this method facilitated the rapid dissolution of xenon gas, bubbling through a solution of cells introduced complications including foaming and the mechanical stresses of bubbles on cell integrity. To determine the

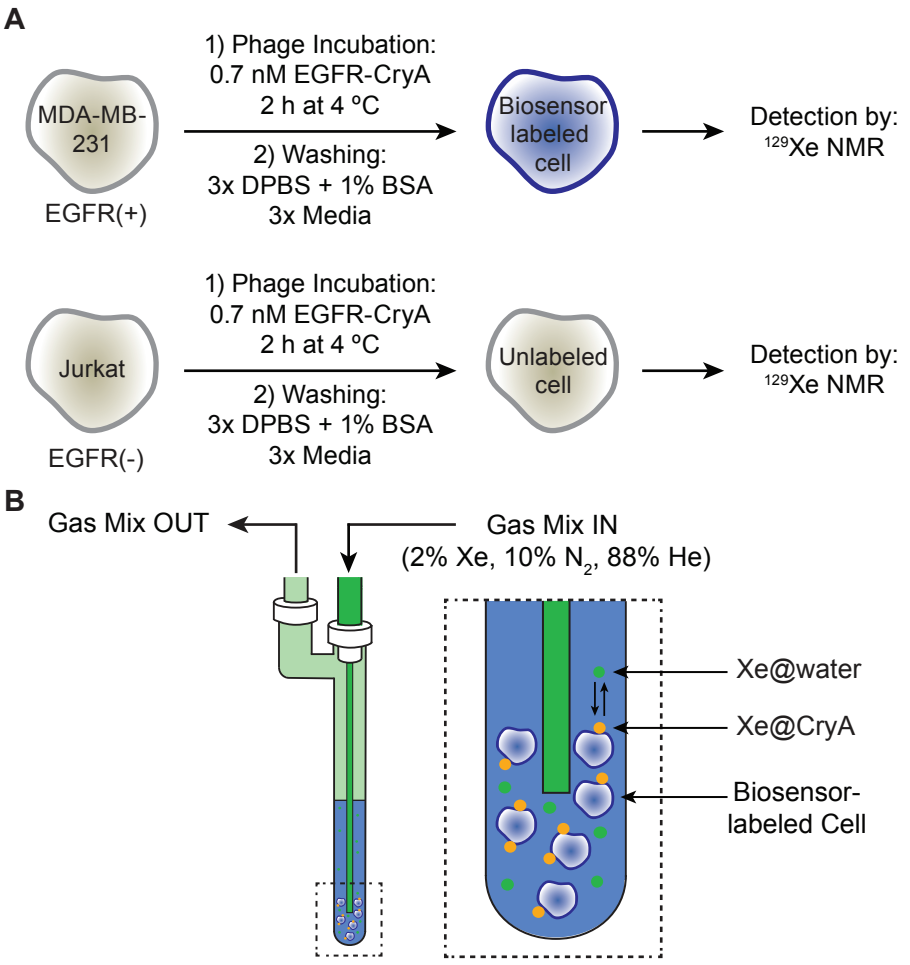



***Figure 9-6. Schematic of cell labeling with the EGFR-CryA biosensor for live-cell 129Xe NMR experiments.***

(A) MDA-MB-231 (EGFR positive) and Jurkat (EGFR negative) cells were incubated with the EGFR-CryA biosensor for 2 h at 4 °C. After removing unbound phage, those that remain bound were detected by xenon NMR. (B) Live-cell $^{129}$Xe NMR experiments were performed in a modified 5 mm NMR tube allowing for continuous gas flow.

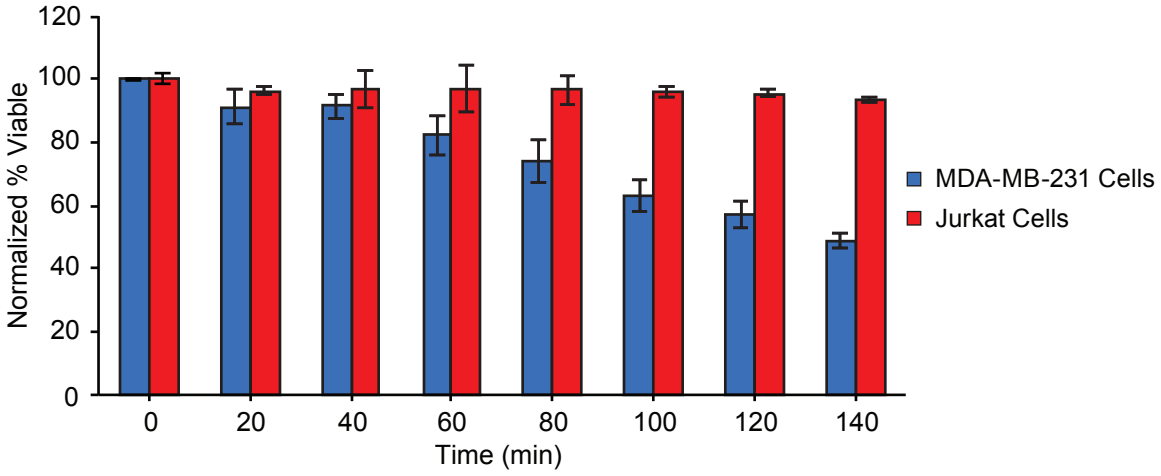



***Figure 9-7. Cell viability over the time course of a typical set of 129Xe NMR experiments.***

MDA-MB-231 and Jurkat cell viability were measured during an NMR experiment using a continuous flow system where $^{129}$Xe was bubbled into a solution of 100 million cells/mL. Cell aliquots were removed at each time-point and viability was measured by uptake of a fluorescent nuclear dye, 7-AAD, which can penetrate compromised cell membranes. Fluorescence intensity for each time point was quantified by flow cytometry. Shown is the average of three biological replicates, each performed in triplicate. Error bars represent the standard deviation.

compatibility of xenon bubbling with live-cell NMR measurements, we assessed the viability of Jurkat and MDA-MB-231 cells over the time course of a typical set of NMR experiments. As a reporter of cell viability, we measured the uptake of a DNA intercalating dye, 7-aminoactinomycin D (7-AAD), after cells were bubbled with a xenon gas mixture for up to 2 h. We found that while most of the Jurkat cell membranes remained intact, a significant number of MDA-MB-231 cells had compromised membranes within 90 min (Figure 9-7). Taking into account both cell viability and the propensity for sample foaming, we limited the duration of bubbling experiments to 70 min for any one solution of cells.

To perform $^{129}$Xe NMR with live-cells, the EGFR-CryA biosensor was incubated with either MDA-MB-231 or Jurkat cells for 2 h at 4 °C at a concentration of 100 million cells/mL (Figure 9-6A). After unbound biosensor was rinsed away, the saturation response profile of each sample was recorded by measuring the Xe@water signal as a function of the frequency of a RF saturation pulse train. This revealed the presence of any xenon populations exchanging with Xe@water (Figure 9-8A). As expected, a region of saturation centered at 70 ppm, characteristic of Xe@CryA, was seen with MDA-MB-231 cells. The same feature was not observed with Jurkat cells, establishing the specificity of the EGFR-CryA biosensor. To verify that both cells were incubated with identical concentrations of the EGFR-CryA biosensor, cell labeling solutions were analyzed by hyperCEST in the same manner and each showed a strong saturation at 70 ppm, confirming the presence of the EGFR-CryA biosensor (Figure 9-9).

The hyperCEST contrast of the cell solutions was quantified by comparing the Xe@water signal after saturating at 70 ppm to that of a symmetrically placed control RF pulse train applied at 320 ppm (Figure 9-8A, inset). While the MDA-MB-231 cell solution exhibited 16.0 ± 9.4% contrast, the Jurkat cell solution exhibited essentially no contrast (1.4 ± 4.6%). Contrast was then measured as a function of saturation time using the same saturation pulse but varying the

number of pulses (Figure 9-8B). While saturation durations of 3 s were sufficient to generate 12.3 ± 4.4% in the solution of MDA-MB-231 cells, increasing that time provided marginal benefits. For instance, doubling the saturation length to 6 s yielded just 15.6 ± 3.6%, and a 15 s saturation period produced only 21.3 ± 12.3%, less than half the contrast of a pulse train 5 times shorter.

To characterize biosensor labeled-cells further, hyperCEST contrast was measured as a function of cell concentration to estimate the cellular detection limit for xenon NMR (Figure 9-8C). A saturation pulse train of higher intensity was used in an effort to create as much contrast

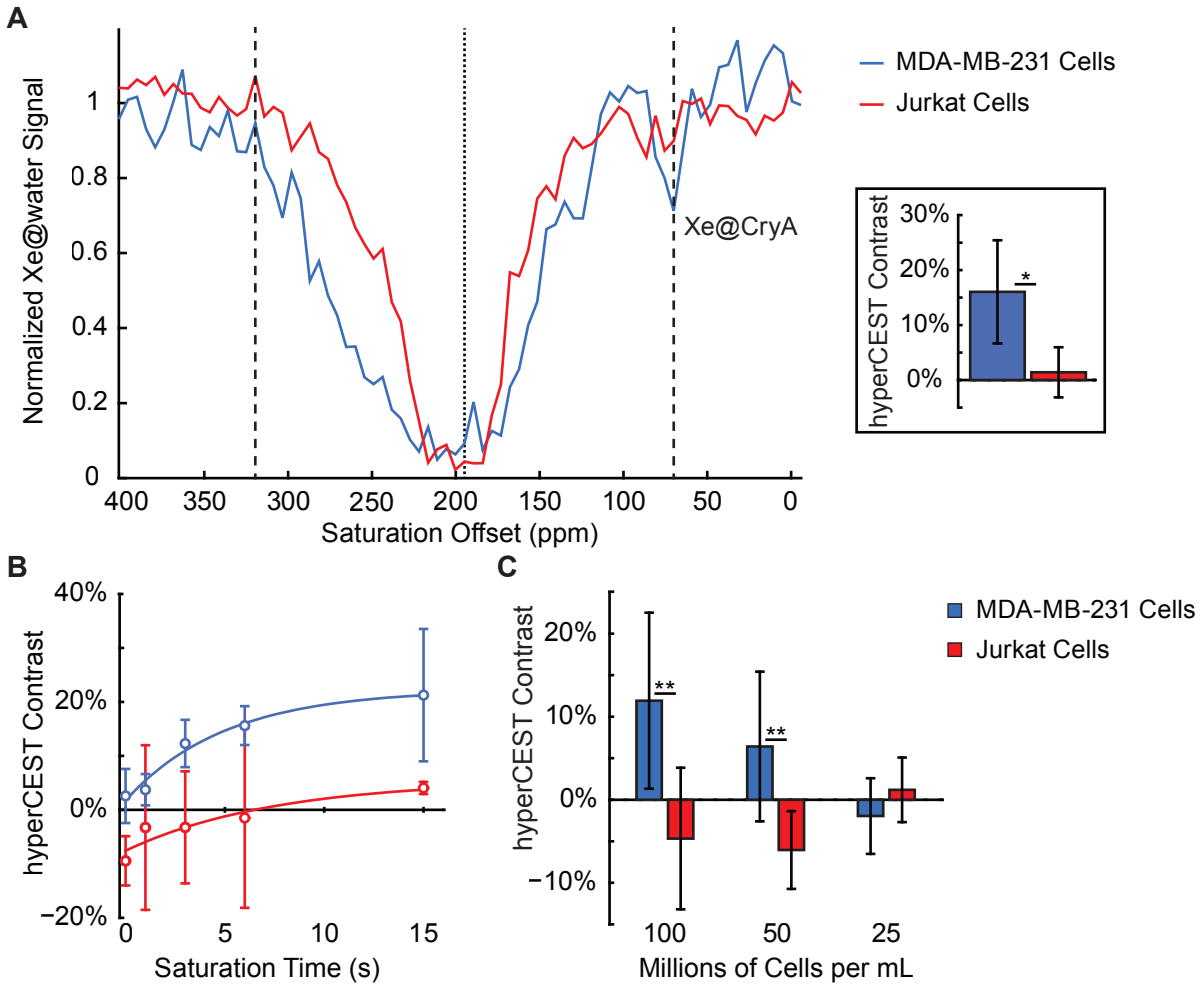



***Figure 9-8. The hyperCEST response of cells labeled with the EGFR-CryA biosensor is specific to MDA-MB-231.***

(A) Comparison of the saturation response profiles indicated that the EGFR-CryA biosensor was bound only by MDA-MB-231 cells, as evidenced by the dip at 70 ppm. Saturation consisted of a train of 500 Hz bandwidth dSNOB pulses swept from 0 to 400 ppm (referenced to the $^{129}$Xe gas frequency) in steps of 5.4 ppm. Dashed lines indicate the offset frequencies used for saturation pulse trains in B and C, while the dotted line shows the chemical shift of Xe@water. Inset: hyperCEST contrast of each sample was quantified using the same dSNOB pulse train. Shown is the mean contrast (n = 10). Error bars represent the standard deviation (* $p < 0.001$). (B) hyperCEST contrast was measured as a function of saturation time using a train of 500 Hz bandwidth dSNOB pulses. Longer saturation times produced more contrast for the MDA-MB-231 sample, but had negligible effect on the Jurkat sample. Shown is the mean contrast (n = 4) plotted as open circles with an exponential fit illustrated with a solid line. Error bars represent the standard deviation. (C) The minimum number of cells needed for hyperCEST contrast was explored using a train of 1000 Hz bandwidth dSNOB pulses. Shown is the mean contrast (n = 10). Error bars represent the standard deviation (** $p < 0.01$).

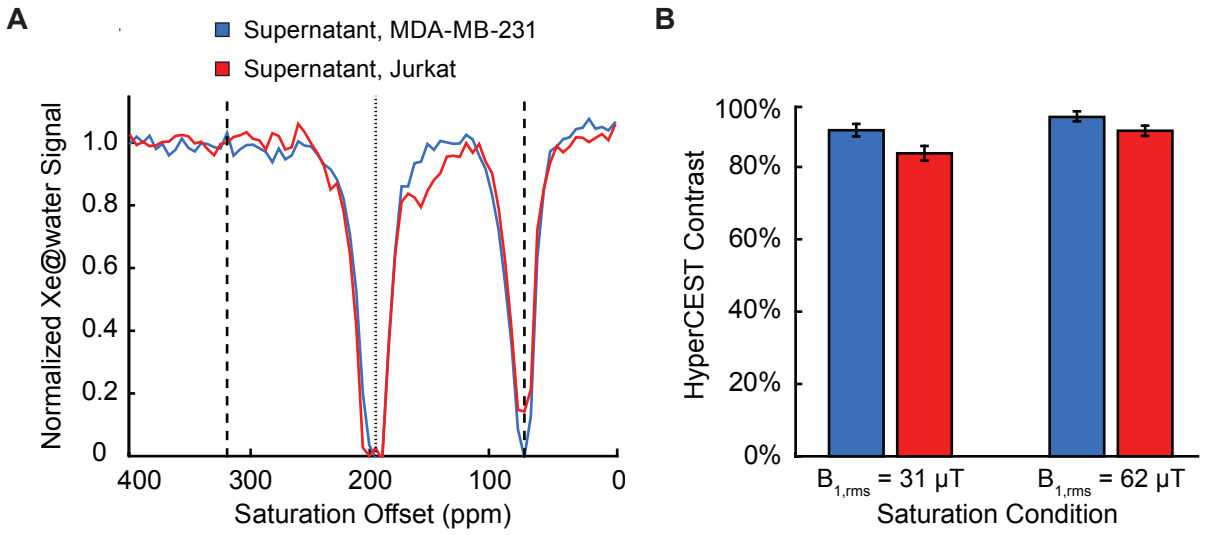



***Figure 9-9. Analysis of supernatant confirmed that both cells were incubated with comparable concentrations of the EGFR-CryA biosensor.***

(A) Comparison of the saturation response profiles indicated that both cell labeling solutions contained an equivalent amount of the EGFR-CryA biosensor, and further served to verify the precise location of the Xe@CryA spin pool. Saturation consisted of a train of 500 Hz bandwidth dSNOB pulses swept from 0 to 400 ppm (referenced to the $^{129}$Xe gas frequency) in steps of 5.4 ppm. Each point was collected in a single acquisition. (B) Comparison of hyperCEST contrast between cell labeling solutions using the same dSNOB pulse train as in A, as well as a higher intensity dSNOB pulse train (1000 Hz bandwidth). Shown is the mean contrast (n = 10). Error bars represent the standard deviation.

as possible. Though the overall contrast was low and somewhat variable (6.4 ± 9.0%), a statistically significant difference in contrast ($P < 0.01$) between the two samples was generated at concentrations as low as 50 million cells/mL, and we envision that the cellular detection limit will be lowered with improvements to the experimental set-up. For example, the sensitivity can be improved by two orders of magnitude by isotopic enrichment of xenon, by its separation from the buffer gas mixture used for polarization, and by increasing its polarization.[41] Additionally, the dissolution of hp $^{129}$Xe via gas-exchange membranes, as opposed to bubbling, should result in more uniform signal distribution while reducing harmful perturbations of the cells.[42]

The negative contrast of the Jurkat cell solution (Figure 9-8c) was the result of the broad, asymmetric response profile to the symmetric RF saturation pulse train. This phenomenon was also observed with unlabeled Jurkat and MDA-MB-231 cells, but was not seen in the media alone (Figure 9-10), suggesting the existence of another xenon population in the regime of intermediate-to-fast exchange with Xe@water. This could be xenon associated with cell membranes (Xe@cell). Although a distinct Xe@cell peak was not observed in the direct spectra (Figure 9-11B), unique chemical shifts have been reported for xenon associated with bacteria and mammalian cells,[24,43] and the Xe@water linewidth observed for cell solutions (Figure 9-11B) is consistent with exchange broadening.

## 9.3 Conclusions

In this study, we have successfully created a targeted, viral capsid-based xenon biosensor that recognizes the EGF receptor with high specificity. We have also demonstrated its utility for

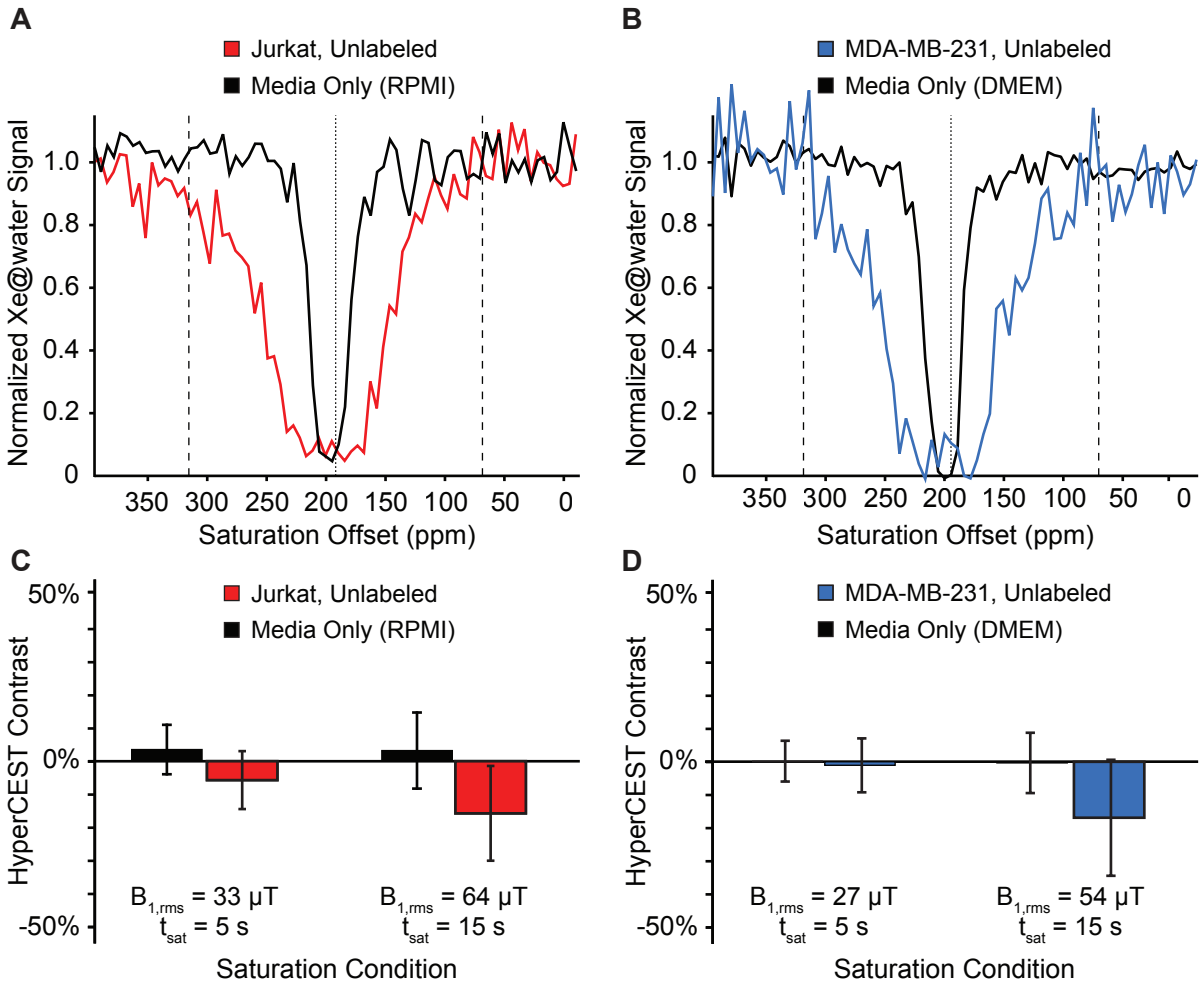



***Figure 9-10. The baseline hyperCEST response of unlabeled MDA-MB-231 and Jurkat cells and their growth media.***

(A, B) Comparison of the saturation response profiles between unlabeled cells and growth media revealed a broad and asymmetric response centered around 200 ppm only in the solution of cells. Saturation consisted of a train of 500 Hz bandwidth dSNOB pulses swept from -8.4 to 398.4 ppm (referenced to the $^{129}$Xe gas frequency) in steps of 5.4 ppm. Dashed lines indicate the offset frequencies used for saturation pulses in C and D, while the dotted line shows the chemical shift of Xe@water. (C, D) Comparison of the hyperCEST contrast between unlabeled cells and growth media revealed that negative contrast was unique to cell solutions and increased with longer saturation times, but was not observed in the growth media alone. Shown is the mean contrast (n = 6). Error bars represent the standard deviation.

live-cell $^{129}$Xe NMR with hyperCEST detection. Importantly, our modular design of the biosensor makes it a versatile contrast agent suitable for many molecular imaging applications. For instance, because we have shown that fd phage with scFvs maintain their binding specificity after N-terminal modification, we can immediately access xenon biosensors for a variety of cellular targets using phage display techniques. Additionally, beyond modifying the phage surface with CryA cages, we can introduce virtually any aminooxy-functionalized molecule, including fluorophores for multimodal MR/optical imaging, and polyethylene glycols for enhanced *in vivo* pharmacodynamics.[28,39] In the future, we intend to leverage this potential to build xenon biosensors that can be used for *in vivo* MRI of lung and other cancers.

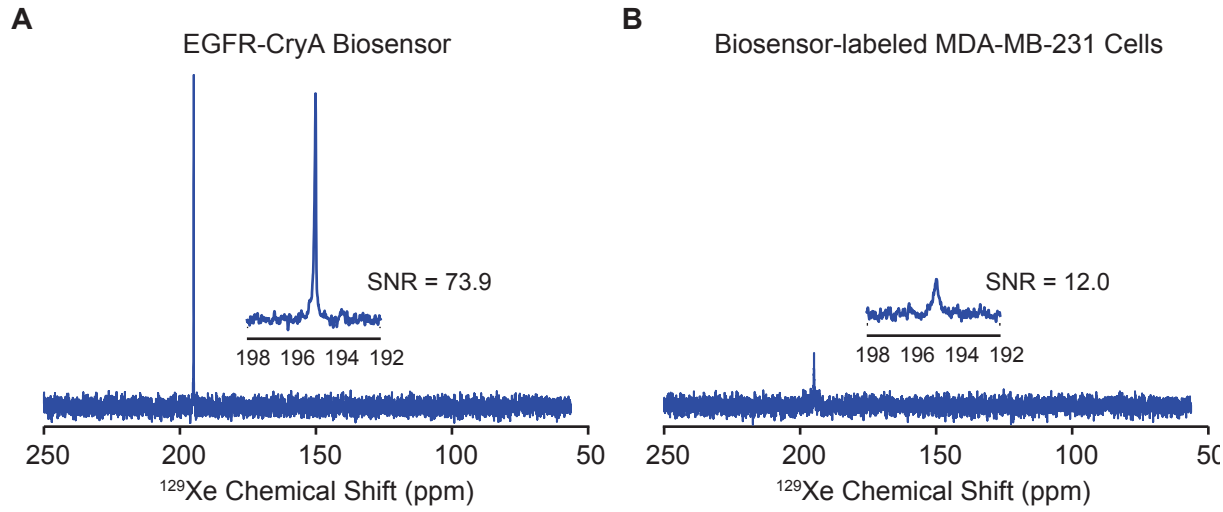



***Figure 9-11.*** ***$^{129}$Xe NMR spectra of media showed a single peak whose linewidth depended upon sample composition.***

(A) The cell labeling solution, i.e., EGFR-CryA biosensor in media, showed a narrow peak at 195 ppm (linewidth = 0.03 ppm), the expected chemical shift of Xe@water. (B) In contrast, samples of MDA-MB-231 cells labeled with the EGFR-CryA biosensor exhibited a broad peak at 195 ppm (linewidth = 0.28 ppm) with a much lower signal-to-noise ratio (SNR). Although a second peak (for Xe@cell) was not observed directly, the data are consistent with xenon exchange on the timescale of the frequency separation between the media and cells (i.e., exchange broadening). Spectra were collected at 37 °C in a single shot.

## 9.4 Materials and Methods

### *9.4.1 General procedures and materials*

All chemical reagents were of analytical grade, obtained from commercial suppliers, and used without further purification unless otherwise noted. All reaction flasks were oven dried prior to use and moisture-sensitive reactions were performed in flame-dried flasks under a $N_2$ atmosphere. Analytical thin layer chromatography (TLC) was performed with 60 Å silica gel plates (EMD) and visualized by staining with ceric ammonium molybdate, ninhydrin, or by absorbance of UV light at 254 nm. Flash chromatography was performed with 60 Å 230-400 mesh silica (Silicycle). Organic extracts were dried over $MgSO_4$, and solvent was removed by reduced pressure with a R-114 rotovapro (Buchi) equipped with a self-cleaning dry vacuum pump (Welch). If necessary, products were further dried by reduced pressure with an Edwards RV5 high vacuum pump.

Water used in biological procedures was distilled and deionized using a Milli-Q Gradient ultrapure water purification system (Millipore). Dulbecco's phosphate-buffered saline (DPBS) was obtained from HyClone Laboratories and fetal bovine serum (FBS) was obtained from Omega Scientific. FACs buffer was prepared as 1% fetal bovine serum (FBS) in DPBS. Blocking buffer was prepared as 1% bovine serum albumin (BSA) in DPBS.

### *9.4.2 Instrumentation*

**NMR.** All $^{1}$H and $^{13}$C spectra are reported in ppm and referenced to solvent peaks. Spectra

were obtained on Bruker AVB-400, AVQ-400, AV-500, or AV-600 spectrometers. Coupling constants (J) are reported in Hz.

**Absorbance spectra.** Absorbance readings were collected on a Thermo Scientific Nanodrop 2000 spectrophotometer.

**High performance liquid chromatography (HPLC).** Purification of small molecules and peptides were performed using reversed-phase (RP) HPLC on a Rainin Dynamax SD-200 system equipped with a Varian UV-Vis detector (model 345) and a Microsorb C18 analytical column (4.6 x 250 mm) at a flow rate of 1 mL/min, a semipreparative column (10 x 250 mm) at a flow rate of 4 mL/min, or a preparative column (21.4 x 250 mm) at a flow rate of 20 mL/min. HPLC samples were filtered with a Pall Life Sciences Acrodisc CR 13 mm syringe filter equipped with a 0.2 µm PTFE membrane prior to injection.

Modified and unmodified fd phage were analyzed by RP-HPLC on an Agilent 1100 Series HPLC System (Agilent Technologies) with an inline diode array detector (DAD) and a Zorbax 300 SB-CN column (Agilent) at a flow rate of 1 mL/min. Samples were centrifuged prior to injection to pellet insoluble material.

**Matrix assisted laser desorption-ionization time-of-flight (MALDI-TOF) MS.** MALDI-TOF MS was performed on a Voyager-DE system from PerSeptive Biosystems (Applied Biosystems). Samples were co-crystallized using saturated solutions of sinapinic acid in 60% acetonitrile, 40% water with 0.1% trifluoroacetic acid.

**Electrospray ionization (ESI) MS.** High-resolution ESI mass spectra were obtained at the QB3 Mass Spectrometry Facility (UC Berkeley) using an LTQ Orbitrap XL hybrid mass spectrometer (Thermo Fisher). Low-resolution ESI mass spectra were obtained using an LTQ-XL mass spectrometer.

**Flow cytometry.** Experiments were performed on a BD Biosciences FACSCalibur flow cytometer, and data analysis was performed using FlowJo software (Tree Star). At least $2x10^4$ live cells were analyzed for each sample. Cell viability was ascertained by gating the samples on the basis of forward scatter (to sort by size) and side scatter (to sort by granularity). The average fluorescence intensity was calculated from each of three replicate experiments to obtain a mean value in arbitrary units.

**Confocal Microscopy.** Images were acquired on a Zeiss laser-scanning microscope 710 equipped with a 405 nm diode laser and a 488 nm Ar laser using the Zen software (Carl Zeiss, 2009) to control the instrument (courtesy of Prof. Christopher Chang (UC Berkeley)).

**$^{129}$Xe NMR.** All $^{129}$Xe NMR data were acquired on a 400 MHz Varian VNMRS console using a commercial, dual-tuned $^{1}$H/broadband probe. $^{129}$Xe was hyperpolarized ($P \approx 2\%$) via spin-exchange optical pumping with a home-built apparatus.[44] The gas mixture for polarization

consisted of 2% xenon (natural abundance), 10% nitrogen, and 88% helium. After polarization, gas flowed through plastic tubing directly into a 5 mm NMR tube outfitted with gas inlet and outlet arms (Figure 9-6B). The tube was pressurized to 482.6 kPa with the polarized gas mixture, temperature was regulated at 37 °C, and xenon was dissolved into the sample solution by flowing through a capillary running to the bottom of the NMR tube at a rate of 0.3-0.35 SLM for 10 s. The bubbling period was followed by a wait period of at least 4 s to allow the solution to settle. Saturation consisted of a train of dSNOB pulses,[45] whose utility for hyperCEST has been recently demonstrated.[39,46]

*9.4.3 Chemical synthesis*

Detailed synthesis and characterization of the aminooxy-functionalized CryA-peptide (**8.2**) can be found in Chapter 8.

*9.4.4 Experimental*

**Cell culture.** All cells were maintained according to ATCC guidelines, grown in a 5% $CO_2$, humidified atmosphere at 37 °C in media (Jurkat: RPMI medium 1640; MDA-MB-231: DMEM/High glucose) supplemented with 10% FBS, penicillin (100 units/mL), and streptomycin (0.1 mg/mL).

**Expression and purification of fd bacteriophage.** The expression and purification of fd bacteriophage displaying anti-EGFR or anti-BoNT scFvs were performed according to established procedures.[28,30-32] Phage concentration was calculated by measuring the absorbance according to Equation 8-1.

**fd-CryA biosensor construction.** fd bacteriophage displaying anti-EGFR or anti-BoNT scFvs were modified with CryA cages according to established procedures.[28,39] Briefly, the phage underwent a pyridoxal 5'-phosphate (PLP) mediated transamination reaction to introduce ketone groups at the N-termini of their coat proteins using the following final reaction concentrations: 65 nM fd and 100 mM PLP in 150 mM phosphate buffer, pH 6.3, for 18 h at RT. The ketone-labeled phage (k-fd) were purified through a series of precipitations and resuspensions in DPBS to remove excess PLP. Oxime formation then proceeded between k-fd and aminooxy-functionalized CryA-peptide using the following final reaction concentrations: 30 nM k-fd, 400 μM aminooxy CryA, and 100 mM aniline in DPBS, pH 5.5. After 22 h at RT, the resulting fd-CryA biosensors (i.e., EGFR-CryA and BONT-CryA) were purified by gel filtration using an Illustra NAP-5 column (GE Healthcare). Purified fd-CryA biosensors were stored in DPBS at 4 °C until needed.

**Characterization of fd phage by RP-HPLC.** Approximately 0.5 nmol of fd phage were combined with 1 M Guanidinium hydrochloride (from a 6 M stock in DPBS) and incubated at RT for 10 min. Denatured phage were analyzed by RP-HPLC using a binary solvent system consisting of buffer A (0.1% trifluoroacetic acid in water (v/v)) and buffer B (0.1% trifluoroacetic acid in acetonitrile (v/v)). The elution gradient was 35% buffer B for the first 4 min, ramping to 70% buffer B over 15 min, then to 95% buffer B over the next 1 min, and a 10 min wash with 95% buffer B.

**Quantification of fd bacteriophage binding to cells by flow cytometry.** Jurkat cells were transferred into a 50 mL Falcon Tube and MDA-MB-231 cells were washed with DPBS, trypsinized, and then transferred into a 50 mL Falcon tube. Cells were centrifuged at 1300 rpms for 3 min and washed once into DPBS. Cells were counted using a hemocytometer and then washed twice with cold FACs buffer and placed on ice. After a final resuspension, $3x10^6$ cells were transferred into a 600 µL Eppendorf tube, pelleted (1300 rpms for 3 min), and then resuspended in 90 µL of either cold FACs buffer or a solution of 0.7 nM fd phage in cold blocking buffer. Cell solutions were mixed and allowed to incubate for 2 h at 4 °C on a rotator. Cells were centrifuged at 1300 rpms for 3 min, resuspended in 120 µL of cold FACs buffer, and 40 µL each were transferred into three wells in a 96-well V-bottom plate (Costar), placing $1x10^6$ cells into each well for triplicate measurements. Cells were washed twice with 190 µL of cold FACs buffer (centrifugation in between each wash step was performed on a RC5C centrifuge (Sorvall) with a SH3000 rotor (Sorvall) at 2500 rpms for 3 min). To quantify phage bound to cells, cells were resuspended in 100 µL of anti-fd Bacteriophage IgG antibody (Sigma) that was diluted 1:500 in cold FACs buffer. After a 45 min incubation on ice, cells were washed three times with 190 µL of cold FACs buffer. The cells were then resuspended in 100 µL of anti-Rabbit IgG-FITC antibody (Southern Biotech) that was diluted 1:400 in cold FACs buffer. After a 45 min incubation on ice and in the dark, cells were washed three times with 190 µL of cold FACs buffer. Finally, cells were resuspended in 400 µL of cold FACs buffer and subjected to flow cytometry. Cells were analyzed by gating on the live cells and quantifying their fluorescence in the FL1 channel.

**Fluorescence microscopy of fd bacteriophage binding to cells.** MDA-MB-231 cells were seeded at 250,000 cells/mL into each well of a 4-well removable chamber slide (LabTek II) with 500 µL of media and grown for 48 h until they were approximately 90% confluent. Cells were washed with 500 µL DPBS and then three times with 500 µL of cold blocking buffer (by pipetting into each well). Cells were then incubated with 150 µL of a 0.7 nM phage solution in cold blocking buffer. For Jurkat cells, they were grown in a 10 mL T-flask (Nunc) for 48 h. Cells were centrifuged at 1300 rpms for 3 min, washed once into DPBS, counted using a hemocytometer, and then washed twice into cold blocking buffer. After a final resuspension in blocking buffer, $3x10^6$ cells were transferred into a 600 µL Eppendorf tube, pelleted (1300 rpms for 3 min), and then resuspended in 50 µL of a 0.7 nM phage solution in cold blocking buffer. All cell solutions were incubated for 2 h at 4 °C. Cells were washed three times with 500 µL cold blocking buffer (either by pipetting for MDA-MB-231 cells or by centrifugation for Jurkat cells). Phage bound to cells were detected using an antibody labeling approach as described above. After the final wash step, cells were fixed with the addition of 200 µL of 4% paraformaldehyde in DPBS for 10 min at RT. Cells were washed once with cold blocking buffer and prepared for imaging. For MDA-MB-231 cells, after the 4-well chamber was removed, 1 drop of mounting media with DAPI (Vectashield, Vector Laboratories) was added to each cell population and then covered with a coverslip. For Jurkat cells, cells were resuspended in 100 µL of FACs buffer, 8 µL were added to a slide and combined with 8 µL of mounting media with DAPI, and then covered with a coverslip. Images were acquired on a LSM 710 confocal microscope.

**Cell viability during $^{129}$Xe NMR.** Jurkat cells were transferred into a 50 mL Falcon Tube and MDA-MB-231 cells were washed with DPBS, trypsinized, and then transferred into a 50 mL Falcon tube. Cells were centrifuged at 1300 rpms for 3 min and washed once into their appropriate media (Jurkat: RPMI medium 1640 + antibiotics; MDA-MB-231: DMEM/High glucose + antibiotics). Cells were counted using a hemocytometer and then washed twice into their appropriate media supplemented with 0.02% L-81 (v/v), a low molecular weight hydrophilic-lipophilic balance surfactant to minimize foaming.[39] After a final resuspension, $55x10^6$ cells were transferred into a 1.5 mL Eppendorf tube in a final volume of 550 µL (Jurkat: RPMI medium 1640 + antibiotics + 0.02% L-81; MDA-MB-231: DMEM/High glucose + antibiotics + 0.02% L-81), so that the final concentration was $100x10^6$ cells/mL, and stored on ice until needed. The cell solutions were transferred into a 5 mm NMR tube modified to incorporate inlet and outlet arms. The inlet was connected to the capillary holder, which was in turn connected to the polarizer and pressurized to 482.6 kPa. The sample was placed in the magnet with the temperature controlled at 37 °C where it was bubbled for 10 s at a flow rate of 0.3-0.35 SLM with the hp gas mixture (2% natural abundance Xe, 10% $N_2$, 88% He). The bubbling period was followed by a wait period of 10 s, which covered both the settling time as well as the saturation time. Every 20 min the NMR tube was removed from the spectrometer, depressurized, and a 15 µL aliquot of the cell solution was removed and placed on ice. Before returning the NMR tube to the spectrometer, 15 µL of the appropriate media (Jurkat: RPMI medium 1640 + antibiotics + 0.02% L-81; MDA-MB-231: DMEM/High glucose + antibiotics + 0.02% L-81) was added to the phantom to maintain a constant volume of 550 µL. Cell viability at each time-point was assessed by exclusion of a fluorescent dye, 7-Aminoactinomycin D (7-AAD, BD Pharmingen), which can readily pass through compromised membranes to intercalate double-stranded DNA. Briefly, cell solutions were washed once into 1 mL of cold FACs buffer, pelleted (1300 rpms for 3 min), and then resuspended in 100 µL of cold FACs buffer. After addition of 5 µL of 7-AAD, cell solutions were incubated for 10 min at RT. Cell solutions were then diluted with the addition of 300 µL of cold FACs buffer and subjected to flow cytometry. Cells were analyzed, without gating, by quantifying their fluorescence in the FL3 channel.

**Cell labeling with the EGFR-CryA biosensor for $^{129}$Xe NMR.** Jurkat cells were transferred into a 50 mL Falcon Tube and MDA-MB-231 cells were washed with DPBS, trypsinized, and then transferred into a 50 mL Falcon tube. Cells were centrifuged at 1300 rpms for 3 min and washed once into DPBS. Cells were counted using a hemocytometer and then $120x10^6$ cells were transferred into a 2.0 mL Eppendorf tube, pelleted (1300 rpms for 3 min), and washed twice into cold blocking buffer. Cells were resuspended in 1500 µL of a 0.7 nM solution of the EGFR-CryA biosensor in blocking buffer. Cell solutions were mixed and allowed to incubate for 2 h at 4 °C on a rotator. Cells were washed twice with 1 mL of blocking buffer, twice with 1 mL of the appropriate media (Jurkat: RPMI medium 1640 + antibiotics + 0.02% L-81; MDA-MB-231: DMEM/High glucose + antibiotics + 0.02% L-81), and then resuspended in a final volume of 1200 µL of the appropriate media and stored on ice until needed (100 µL were saved for analysis by flow cytometry and 2 aliquots of 550 µL remained for NMR). The cell solutions were transferred into a 5 mm NMR phantom for NMR measurements.

**hyperCEST settings for live-cell $^{129}$Xe NMR.** Saturation response profiles were recorded by measuring the Xe@water signal as a function of the offset frequency of the saturation pulse train. The offset frequency was varied in steps of 5.43 ppm over a range of approximately 400 ppm centered about the Xe@water chemical shift (76 acquisitions in total). Saturations consisted of a train of shaped RF pulses, known as dSNOB pulses. For the saturation response profile and corresponding contrast measurement, a pulse with 500 Hz bandwidth was used (the pulse length ($\tau_{pulse}$) was 5.64 ms, the maximum RF intensity ($B_{1,max}$) was 54 μT, the rms RF intensity ($B_{1,rms}$) was 31 μT, $n_{pulses}$ = 1712, $t_{sat}$ = 10 s).

For saturation time curves, the same pulse was used, but the number of pulses was varied to change the total saturation time ($t_{sat}$ = 0, 1, 3, 6, 15 s; $n_{pulses}$ = 0, 177, 532, 1063, 2568). Contrast was quantified by comparing Xe@water signal from two separate saturations, one after application of the pulse train on-resonance at the frequency of Xe@CryA (69.6 ppm), and the second after application equidistant from, but on the opposite side of the Xe@water peak (317.6 ppm). The off-resonance saturation was used to account for any undesired direct saturation of the bulk aqueous Xe pool, which would otherwise be attributed to hyperCEST contrast. For contrast quantification of lower cell concentrations, a pulse with 1000 Hz bandwidth was used ($\tau_{pulse}$ = 2.82 ms, $B_{1,max}$ = 108 μT, $B_{1,rms}$ = 62 μT, $n_{pulses}$ = 3546, $t_{sat}$ = 10 s). Spectra were acquired with 30.5 kHz spectral width in 1 s.

**hyperCEST settings for cell labeling solutions (supernatant).** Saturations consisted of a train of shaped RF pulses, known as dSNOB pulses. For the saturation response profile a pulse with 500 Hz bandwidth was used ($\tau_{pulse}$ = 5.64 ms, $B_{1,max}$ = 54 μT, $B_{1,rms}$ = 31 μT, $n_{pulses}$ = 1712, $t_{sat}$ = 10 s). For the contrast measurement, an additional pulse with 1000 Hz bandwidth was used ($\tau_{pulse}$ = 2.82 ms, $B_{1,max}$ = 108 μT, $B_{1,rms}$ = 62 μT, $n_{pulses}$ = 3546, $t_{sat}$ = 10 s).

**hyperCEST settings for unlabeled cells and media.** Saturations consisted of a train of shaped RF pulses, known as dSNOB pulses. A pulse with 500 Hz bandwidth ($\tau_{pulse}$ = 5.64 ms) was used for all saturation response profiles, though exact intensity and saturation times differed: for Jurkat cells ($B_{1,max}$ = 112 μT, $B_{1,rms}$ = 64 μT, $n_{pulses}$ = 1773, $t_{sat}$ = 10 s), for RPMI media ($B_{1,max}$ = 58 μT, $B_{1,rms}$ = 33 μT, $n_{pulses}$ = 887, $t_{sat}$ = 5 s), and for both MDA-MB-231 cells and DMEM media ($B_{1,max}$ = 47 μT, $B_{1,rms}$ = 27 μT, $n_{pulses}$ = 1773, $t_{sat}$ = 10 s). Contrast was quantified using two different pulses, either a lower intensity 500 Hz bandwidth dSNOB pulse totaling 5 s ($\tau_{pulse}$ = 5.64 ms, $n_{pulses}$ = 887), or a higher intensity 1000 Hz bandwidth dSNOB pulse totaling 10 s ($\tau_{pulse}$ = 2.82 ms, $n_{pulses}$ = 5319). For Jurkat cells and RPMI media, a maximum intensity of 58 μT and 112 μT was used for 500 Hz and 1000 Hz bandwidth pulses, respectively, while MDA-MB-231 cells and DMEM media utilized maximum intensities of 47 μT and 94 μT, for 500 Hz and 1000 Hz bandwidth pulses, respectively.

### *9.4.5 Data analysis*

**NMR data analysis.** Data processing was performed in MATLAB (version 7.13, R2011b, The MathWorks). FIDs were zero-filled to 16384 points, baseline corrected, and apodized with an 11 Hz exponential before Fourier transform and subsequent phase correction. For each spectrum, the area under the Xe@water peak was summed, and this value was used for later analyses and

plots. All reported contrast values were the result of averaging either 4 or 10 pairs of off-resonance/on-resonance saturations. Contrast for the *j*-th pair of values was calculated according to Equation 9-1 using the Xe@water signal after off- and on-resonance saturation, ($I_{off,j}$ and $I_{on,j}$, respectively).

$$Score = \prod_i f\left(\frac{|d_i - kr_i|}{\sigma\sqrt{2}}\right) \quad (9\text{-}1)$$

The mean and standard deviation were calculated from this final set of contrast values. A two-tailed Student's t-test was used to determine statistical significance between data sets. All data were visually inspected to verify normality, and an outlier test was applied to search for potential outliers. Grubb's test was used on data sets consisting of 10 measurements, and no outliers were found. Dixon's Q test was applied to the saturation time curve data and revealed the presence of 1 outlier in the Jurkat data at the 15 s time point. This value was removed and the mean and standard deviation were re-calculated. Both tests were conducted at the 95% confidence level.

## 9.5 References

1. Massoud T.F. and Gambhir S.S. (**2003**) Molecular imaging in living subjects: seeing fundamental biological processes in a new light. *Genes Dev* 17, 545-580.

2. Tyszka J.M., Fraser S.E. and Jacobs R.E. (**2005**) Magnetic resonance microscopy: recent advances and applications. *Curr Opin Biotechnol* 16, 93-99.

3. Sosnovik D.E. and Weissleder R. (**2007**) Emerging concepts in molecular MRI. *Curr Opin Biotechnol* 18, 4-10.

4. Chan K.W.Y. and Wong W.T. (**2007**) Small molecular gadolinium(III) complexes as MRI contrast agents for diagnostic imaging. *Coordin Chem Rev* 251, 2428-2451.

5. Bulte J.W.M. and Kraitchman D.L. (**2004**) Iron oxide MR contrast agents for molecular and cellular imaging. *Nmr Biomed* 17, 484-499.

6. Woods M., Donald E.W.C. and Sherry A.D. (**2006**) Paramagnetic lanthanide complexes as PARACEST agents for medical imaging. *Chem Soc Rev* 35, 500-511.

7. Kurhanewicz J., Bok R., Nelson S.J. and Vigneron D.B. (**2008**) Current and potential applications of clinical C-13 MR spectroscopy. *J Nucl Med* 49, 341-344.

8. Hopkins S.R. *et al.* (**2007**) Advances in magnetic resonance imaging of lung physiology. *J Appl Physiol* 102, 1244-1254.

9. Oros A.M. and Shah N.J. (**2004**) Hyperpolarized xenon in NMR and MRI. *Phys Med Biol* 49,

R105-R153.

10. Golman K., Zandt R.I., Lerche M., Pehrson R. and Ardenkjaer-Larsen J.H. (**2006**) Metabolic imaging by hyperpolarized 13C magnetic resonance imaging for in vivo tumor diagnosis. *Cancer Res* 66, 10855-10860.

11. Gabard J. and Collet A. (**1981**) Synthesis of a (D3)-Bis(Cyclotriveratrylenyl) Macrocage by Stereospecific Replication of a (C3)-Subunit. *J Chem Soc Chem Comm*, 1137-1139.

12. Bartik K., Luhmer M., Dutasta J.P., Collet A. and Reisse J. (**1998**) Xe-129 and H-1 NMR study of the reversible trapping of xenon by cryptophane-A in organic solution. *J Am Chem Soc* 120, 784-791.

13. Taratula O. and Dmochowski I.J. (**2010**) Functionalized 129Xe contrast agents for magnetic resonance imaging. *Curr Opin Chem Biol* 14, 97-104.

14. Schroder L., Lowery T.J., Hilty C., Wemmer D.E. and Pines A. (**2006**) Molecular imaging using a targeted magnetic resonance hyperpolarized biosensor. *Science* 314, 446-449.

15. Spence M.M. *et al.* (**2001**) Functionalized xenon as a biosensor. *Proc Natl Acad Sci U S A* 98, 10654-10657.

16. Lowery T.J. *et al.* (**2006**) Optimization of xenon biosensors for detection of protein interactions. *Chembiochem* 7, 65-73.

17. Spence M.M. *et al.* (**2004**) Development of a functionalized xenon biosensor. *J Am Chem Soc* 126, 15287-15294.

18. Roy V. *et al.* (**2007**) A cryptophane biosensor for the detection of specific nucleotide targets through xenon NMR spectroscopy. *Chemphyschem* 8, 2082-2085.

19. Wei Q. *et al.* (**2006**) Designing 129Xe NMR biosensors for matrix metalloproteinase detection. *J Am Chem Soc* 128, 13274-13283.

20. Chambers J.M. *et al.* (**2009**) Cryptophane xenon-129 nuclear magnetic resonance biosensors targeting human carbonic anhydrase. *J Am Chem Soc* 131, 563-569.

21. Seward G.K., Bai Y., Khan N.S. and Dmochowski I.J. (**2011**) Cell-compatible, integrin-targeted cryptophane-129Xe NMR biosensors. *Chemical Science* 2, 1103.

22. Schlundt A. *et al.* (**2009**) A xenon-129 biosensor for monitoring MHC-peptide interactions. *Angew Chem Int Ed Engl* 48, 4142-4145.

23. Seward G.K., Wei Q. and Dmochowski I.J. (**2008**) Peptide-mediated cellular uptake of cryptophane. *Bioconjug Chem* 19, 2129-2135.

24. Boutin C. *et al.* (**2011**) Cell uptake of a biosensor detected by hyperpolarized 129Xe NMR: the transferrin case. *Bioorg Med Chem* 19, 4135-4143.

25. Kehoe J.W. and Kay B.K. (**2005**) Filamentous phage display in the new millennium. *Chem Rev* 105, 4056-4072.

26. Mao C. *et al.* (**2004**) Virus-based toolkit for the directed synthesis of magnetic and semiconducting nanowires. *Science* 303, 213-217.

27. Yacoby I. and Benhar I. (**2008**) Targeted filamentous bacteriophages as therapeutic agents. *Expert Opin Drug Del* 5, 321-329.

28. Carrico Z.M. *et al.* (**2012**) N-Terminal labeling of filamentous phage to create cancer marker imaging agents. *ACS Nano* 6, 6675-6680.

29. Smith G.P. and Petrenko V.A. (**1997**) Phage Display. *Chem Rev* 97, 391-410.

30. Zhou Y. *et al.* (**2007**) Impact of single-chain Fv antibody fragment affinity on nanoparticle targeting of epidermal growth factor receptor-expressing tumor cells. *J Mol Biol* 371, 934-947.

31. Zhou Y. and Marks J.D. (**2012**) Discovery of internalizing antibodies to tumor antigens from phage libraries. *Methods Enzymol* 502, 43-66.

32. Amersdorfer P. *et al.* (**2002**) Genetic and immunological comparison of anti-botulinum type A antibodies from immune and non-immune human phage libraries. *Vaccine* 20, 1640-1648.

33. Laskin J.J. and Sandler A.B. (**2004**) Epidermal growth factor receptor: a promising target in solid tumours. *Cancer Treat Rev* 30, 1-17.

34. Nicholson R.I., Gee J.M. and Harper M.E. (**2001**) EGFR and cancer prognosis. *Eur J Cancer* 37 Suppl 4, S9-15.

35. Li K. *et al.* (**2010**) Chemical modification of M13 bacteriophage and its application in cancer cell imaging. *Bioconjug Chem* 21, 1369-1377.

36. Merzlyak A., Indrakanti S. and Lee S.W. (**2009**) Genetically engineered nanofiber-like viruses for tissue regenerating materials. *Nano Lett* 9, 846-852.

37. Gilmore J.M., Scheck R.A., Esser-Kahn A.P., Joshi N.S. and Francis M.B. (**2006**) N-terminal protein modification through a biomimetic transamination reaction. *Angew Chem Int Edit* 45, 5307-5311.

38. Witus L.S. *et al.* (**2010**) Identification of Highly Reactive Sequences For PLP-Mediated Bioconjugation Using a Combinatorial Peptide Library. *J Am Chem Soc* 132, 16812-16817.

39. Stevens T.K. *et al.* (**2012**) HyperCEST detection of a (129) Xe-based contrast agent composed of cryptophane-A molecular cages on a bacteriophage scaffold. *Magn Reson Med*.

40. Han S.I. *et al.* (**2005**) NMR-based biosensing with optimized delivery of polarized 129Xe to solutions. *Anal Chem* 77, 4008-4012.

41. Hersman F.W. *et al.* (**2008**) Large production system for hyperpolarized 129Xe for human lung imaging studies. *Acad Radiol* 15, 683-692.

42. Baumer D., Brunner E., Blumler P., Zanker P.P. and Spiess H.W. (**2006**) NMR spectroscopy of laser-polarized (129)Xe under continuous flow: a method to study aqueous solutions of biomolecules. *Angew Chem Int Ed Engl* 45, 7282-7284.

43. Boutin C. *et al.* (**2011**) Hyperpolarized 129Xe NMR signature of living biological cells. *Nmr Biomed* 24, 1264-1269.

44. Walker T.G. and Happer W. (**1997**) Spin-exchange optical pumping of noble-gas nuclei. *Reviews of Modern Physics* 69, 629-642.

45. Kupce E., Boyd J. and Campbell I.D. (**1995**) Short selective pulses for biochemical applications. *J Magn Reson B* 106, 300-303.

46. Meldrum T., Bajaj V.S., Wemmer D.E. and Pines A. (**2011**) Band-selective chemical exchange saturation transfer imaging with hyperpolarized xenon-based molecular sensors. *J Magn Reson* 213, 14-21.

# SECTION III: APPENDICES

# Appendix A: Detailed MS Results for Chapter 3 (IsoStamp)

## A.1 Index of Orbitrap data files used for computing the computational detection rates

Samples containing the indicated quantity of halogen-tagged BSA in the presence of 10 µg of Jurkat whole cell lysate were subjected to LC-MS analysis on an LTQ Orbitrap XL (resolution = 60,000) mass spectrometer.

| Tagged BSA in 10 µg of Lysate (pmoles) | Monochloride | Monobromide | Dichloride | Dibromide |
|---|---|---|---|---|
| 0.03 | ORB4844 | ORB4804 | ORB4798 | ORB4707 |
| 0.08 | ORB4876 | ORB4805 | ORB4799 | ORB4793 |
| 0.15 | ORB4846 | ORB4806 | ORB4800 | ORB4794 |
| 0.3 | ORB4877 | ORB4807 | ORB4879 | ORB4708 |
| 0.8 | ORB4855 | ORB4808 | ORB4801 | ORB4795 |
| 1.5 | ORB4856 | ORB4809 | ORB4802 | ORB4796 |
| 3 | ORB4857 | ORB4878 | ORB4714 | ORB4709 |
| 30 | | | | ORB4797 |

## A.2 Index of Orbitrap data files used for computing the manual detection rates

Samples containing the indicated quantity of halogen-tagged BSA were subjected to LC-MS analysis on an LTQ Orbitrap XL (resolution = 60,000) mass spectrometer.

| Tagged BSA (pmoles) | Monochloride | Monobromide | Dichloride | Dibromide |
|---|---|---|---|---|
| 30 | ORB4852 | ORB4803 | ORB4710 | ORB4704 |

## A.3 Index of Orbitrap data files from directed LC-MS/MS

Samples containing the indicated quantity of dibromide-labeled BSA in the presence of 10 µg of Jurkat whole cell lysate were subjected to an inclusion list driven LC-MS/MS experiment using the IsoStamp method on an LTQ Orbitrap XL (resolution = 60,000) mass spectrometer. The false discovery rate (FDR) for database search results are shown.

| Dibromide-labeled BSA in 10 µg of Lysate (pmoles) | Fullscan | Directed | [1]FDR |
|---|---|---|---|
| 3.0 | ORB17011 | ORB17013 | ≤ 8% |
| 1.5 | ORB17089 | ORB17093 | ≤ 8% |
| 0.8 | ORB17087 | ORB17091 | ≤ 8% |
| 0.3 | ORB17191 | ORB17193 | ≤ 8% |

[1]Although a high FDR was used during the initial database search, all spectra were manually validated.

## A.4 Index of Orbitrap data files from data-dependent LC-MS/MS

Samples containing the indicated quantity of acetamide-labeled BSA in the presence of 10 µg of Jurkat whole cell lysate were subjected to a data-dependent LC-MS/MS experiment (where the 10 most abundant ions were fragmented, Top10) on an LTQ Orbitrap XL (resolution = 60,000) mass spectrometer. The false discovery rate (FDR) for database search results are shown.

| Acetamide-labeled BSA in 10 µg of Lysate (pmoles) | DDA | | FDR |
|---|---|---|---|
| 3.0 | ORB17019 | | 0.8% |
| 1.5 | ORB17085 | | 1.0% |
| 0.8 | ORB17189 | | 0.9% |
| 0.3 | ORB17187 | | 1.0% |

## A.5 Computational detection rate for each single cysteine-containing BSA peptide

The computational detection rate for each single cysteine-containing BSA peptide from searching a sample containing the indicated quantity of halogen-tagged BSA in the presence of 10 µg of Jurkat whole cell lysate. Numbers in each column indicate the number of ions computationally found for each peptide listed (where the ion could be in the +1 to +5 charge state), while values of "NA" indicate that ions corresponding to that peptide were not expected to be found because they were not identified from manual searching of 30 picomoles of BSA in the absence of lysate (see Appendix A.2). Ions found were compared against a pure BSA sample subjected to tryptic digestion and analyzed by LC-MS/MS as a control for retention time and ionizability. Data were obtained on an LTQ Orbitrap XL (resolution = 60,000) or Waters QTof Premier (resolution = 10,000) mass spectrometer.

Computational detection rate of dibromide-taggved single cysteine-containing BSA peptides in MS experiments where the inidicated amount of labeled BSA was trypsinized with 10 µg of Jurkat whole cell lysate. LC-MS analysis was performed on an LTQ-Orbitrap XL mass spectrometer (resolution = 60,000).

| Start | Stop | Sequence | Mods. | Missed Cleavages | Monoisotopic Mass | 3.0 pmoles | 1.5 pmoles | 0.8 pmoles | 0.3 pmoles | 0.15 pmoles | 0.08 pmoles | 0.03 pmoles |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 223 | 228 | (R)CASIQK(F) | | 0 | 649.334 | 2 | 2 | 2 | 2 | 0 | 0 | 0 |
| 198 | 204 | (K)GACLLPK(I) | | 0 | 701.402 | 2 | 2 | 2 | 2 | 1 | 0 | 0 |
| 483 | 489 | (R)LCVLHEK(T) | | 0 | 841.460 | 2 | 2 | 2 | 2 | 0 | 0 | 0 |
| 221 | 228 | (R)LRCASIQK(F) | | 1 | 918.519 | NA | NA | NA | NA | NA | NA | NA |
| 123 | 130 | (R)NECFLSHK(D) | | 0 | 977.451 | 2 | 2 | 2 | 0 | 0 | 0 | 0 |
| 413 | 420 | (K)QNCDQFEK(L) | pyroGlu | 0 | 994.394 | 2 | 2 | 1 | 0 | 0 | 0 | 0 |
| 413 | 420 | (K)QNCDQFEK(L) | | 0 | 1011.420 | 2 | 1 | 0 | 1 | 0 | 0 | 0 |
| 310 | 318 | (K)SHCIAEVEK(D) | | 0 | 1015.488 | 2 | 2 | 2 | 2 | 0 | 0 | 0 |
| 588 | 597 | (K)EACFAVEGPK(L) | | 0 | 1050.493 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 223 | 232 | (R)CASIQKFGER(A) | | 1 | 1138.567 | NA | NA | NA | NA | NA | NA | NA |
| 219 | 228 | (R)QRLRCASIQK(F) | pyroGlu | 2 | 1185.652 | NA | NA | NA | NA | NA | NA | NA |
| 337 | 346 | (K)DVCKNYQEAK(D) | | 1 | 1197.557 | NA | NA | NA | NA | NA | NA | NA |
| 219 | 228 | (R)QRLRCASIQK(F) | | 2 | 1202.679 | NA | NA | NA | NA | NA | NA | NA |
| 198 | 209 | (K)GACLLPKIETMR(E) | | 1 | 1331.717 | NA | NA | NA | NA | NA | NA | NA |
| 198 | 209 | (K)GACLLPKIETMR(E) | 1Met-ox | 1 | 1347.712 | NA | NA | NA | NA | NA | NA | NA |
| 89 | 100 | (K)SLHTLFGDELCK(V) | | 0 | 1362.672 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 286 | 297 | (K)YICDNQDTISSK(L) | | 0 | 1386.621 | 1 | 0 | 0 | 1 | 0 | 1 | 0 |
| 221 | 232 | (R)LRCASIQKFGER(A) | | 2 | 1407.753 | NA | NA | NA | NA | NA | NA | NA |
| 223 | 235 | (R)CASIQKFGERALK(A) | | 2 | 1450.784 | NA | NA | NA | NA | NA | NA | NA |
| 483 | 495 | (R)LCVLHEKTPVSEK(V) | | 1 | 1482.799 | NA | NA | NA | NA | NA | NA | NA |
| 387 | 399 | (K)DDPHACYSTVFDK(L) | | 0 | 1497.632 | 2 | 2 | 1 | 0 | 2 | 0 | 0 |
| 139 | 151 | (K)LKPDPNTLCDEFK(A) | | 0 | 1519.746 | 2 | 2 | 2 | 2 | 2 | 0 | 0 |
| 198 | 211 | (K)GACLLPKIETMREK(V) | | 2 | 1588.855 | NA | NA | NA | NA | NA | NA | NA |
| 118 | 130 | (K)QEPERNECFLSHK(D) | pyroGlu | 1 | 1599.722 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 198 | 211 | (K)GACLLPKIETMREK(V) | 1Met-ox | 2 | 1604.850 | NA | NA | NA | NA | NA | NA | NA |
| 118 | 130 | (K)QEPERNECFLSHK(D) | | 1 | 1616.749 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 286 | 299 | (K)YICDNQDTISSKLK(E) | | 1 | 1627.800 | NA | NA | NA | NA | NA | NA | NA |
| 469 | 482 | (R)MPCTEDYLSLILNR(L) | | 0 | 1667.813 | 2 | 2 | 1 | 0 | 0 | 0 | 0 |
| 469 | 482 | (R)MPCTEDYLSLILNR(L) | 1Met-ox | 0 | 1683.808 | 2 | 2 | 2 | 2 | 0 | 0 | 0 |
| 387 | 401 | (K)DDPHACYSTVFDKLK(H) | | 1 | 1738.811 | NA | NA | NA | NA | NA | NA | NA |
| 483 | 498 | (R)LCVLHEKTPVSEKVTK(C) | | 2 | 1811.010 | NA | NA | NA | NA | NA | NA | NA |
| 508 | 523 | (R)RPCFSALTPDETYVPK(A) | | 0 | 1823.900 | 2 | 2 | 2 | 1 | 2 | 0 | 0 |
| 123 | 138 | (R)NECFLSHKDDSPDLPK(L) | | 1 | 1844.848 | 3 | 3 | 3 | 0 | 3 | 0 | 0 |

| Start | Stop | Sequence | Mods. | Missed Cleavages | Monoisotopic Mass | 3.0 pmoles | 1.5 pmoles | 0.8 pmoles | 0.3 pmoles | 0.15 pmoles | 0.08 pmoles | 0.03 pmoles |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 529 | 544 | (K)LFTFHADICTLPDTEK(Q) | | 0 | 1850.899 | 2 | 1 | 0 | 0 | 0 | 0 | 0 |
| 281 | 297 | (R)ADLAKYICDNQDTISSK(L) | | 1 | 1884.901 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 89 | 105 | (K)SLHTLFGDELCKVASLR(E) | | 1 | 1888.995 | NA | NA | NA | NA | NA | NA | NA |
| 139 | 155 | (K)LKPDPNTLCDEFKADEK(K) | | 1 | 1962.948 | 3 | 3 | 3 | 3 | 3 | 0 | 0 |
| 588 | 607 | (K)EACFAVEGPKLVVSTQTALA(-) | | 1 | 2034.058 | NA | NA | NA | NA | NA | NA | NA |
| 139 | 156 | (K)LKPDPNTLCDEFKADEKK(F) | | 2 | 2091.043 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 281 | 299 | (R)ADLAKYICDNQDTISSKLK(E) | | 2 | 2126.080 | NA | NA | NA | NA | NA | NA | NA |
| 529 | 547 | (K)LFTFHADICTLPDTEKQIK(K) | | 1 | 2220.137 | NA | NA | NA | NA | NA | NA | NA |
| 402 | 420 | (K)HLVDEPQNLIKQNCDQFEK(L) | | 1 | 2298.118 | NA | NA | NA | NA | NA | NA | NA |
| 529 | 548 | (K)LFTFHADICTLPDTEKQIKK(Q) | | 2 | 2348.232 | NA | NA | NA | NA | NA | NA | NA |
| 131 | 151 | (K)DDSPDLPKLKPDPNTLCDEFK(A) | | 1 | 2387.144 | NA | NA | NA | NA | NA | NA | NA |
| 319 | 340 | (K)DAIPENLPPLTADFAEDKDVCK(N) | | 1 | 2401.159 | 3 | 3 | 2 | 2 | 2 | 2 | 0 |
| 508 | 528 | (R)RPCFSALTPDETYVPKAFDEK(L) | | 1 | 2414.170 | NA | NA | NA | NA | NA | NA | NA |
| 45 | 65 | (K)GLVLIAFSQYLQQCPFDEHVK(L) | | 0 | 2435.243 | NA | NA | NA | NA | NA | NA | NA |
| 524 | 544 | (K)AFDEKLFTFHADICTLPDTEK(Q) | | 1 | 2441.169 | NA | NA | NA | NA | NA | NA | NA |
| 413 | 433 | (K)QNCDQFEKLGEYGFQNALIVR(Y) | pyroGlu | 1 | 2455.171 | NA | NA | NA | NA | NA | NA | NA |
| 118 | 138 | (K)QEPERNECFLSHKDDSPDLPK(L) | pyroGlu | 2 | 2467.119 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 413 | 433 | (K)QNCDQFEKLGEYGFQNALIVR(Y) | | 1 | 2472.198 | NA | NA | NA | NA | NA | NA | NA |
| 118 | 138 | (K)QEPERNECFLSHKDDSPDLPK(L) | | 2 | 2484.146 | NA | NA | NA | NA | NA | NA | NA |
| 400 | 420 | (K)LKHLVDEPQNLIKQNCDQFEK(L) | | 2 | 2539.297 | NA | NA | NA | NA | NA | NA | NA |
| 337 | 359 | (K)DVCKNYQEAKDAFLGSFLYEYSR(R) | | 2 | 2746.282 | NA | NA | NA | NA | NA | NA | NA |
| 524 | 547 | (K)AFDEKLFTFHADICTLPDTEKQIK(K) | | 2 | 2810.407 | NA | NA | NA | NA | NA | NA | NA |
| 131 | 155 | (K)DDSPDLPKLKPDPNTLCDEFKADEK(K) | | 2 | 2830.345 | NA | NA | NA | NA | NA | NA | NA |
| 413 | 436 | (K)QNCDQFEKLGEYGFQNALIVRYTR(K) | pyroGlu | 2 | 2875.383 | NA | NA | NA | NA | NA | NA | NA |
| 413 | 436 | (K)QNCDQFEKLGEYGFQNALIVRYTR(K) | | 2 | 2892.410 | NA | NA | NA | NA | NA | NA | NA |
| 310 | 336 | (K)SHCIAEVEKDAIPENLPPLTADFAEDK(D) | | 1 | 2952.430 | NA | NA | NA | NA | NA | NA | NA |
| 387 | 412 | (K)DDPHACYSTVFDKLKHLVDEPQNLIK(Q) | | 2 | 3025.509 | NA | NA | NA | NA | NA | NA | NA |
| 319 | 346 | (K)DAIPENLPPLTADFAEDKDVCKNYQEAK(D) | | 2 | 3134.499 | 2 | 2 | 2 | 0 | 0 | 0 | 0 |
| 37 | 65 | (K)DLGEEHFKGLVLIAFSQYLQQCPFDEHVK(L) | | 1 | 3390.683 | NA | NA | NA | NA | NA | NA | NA |
| 45 | 75 | (K)GLVLIAFSQYLQQCPFDEHVKLVNELTEFAK(T) | | 1 | 3579.856 | NA | NA | NA | NA | NA | NA | NA |
| 35 | 65 | (R)FKDLGEEHFKGLVLIAFSQYLQQCPFDEHVK(L) | | 2 | 3665.846 | NA | NA | NA | NA | NA | NA | NA |
| 402 | 433 | (K)HLVDEPQNLIKQNCDQFEKLGEYGFQNALIVR(Y) | | 2 | 3758.896 | NA | NA | NA | NA | NA | NA | NA |
| | | | | | Total peptides: | 21 | 18 | 16 | 12 | 8 | 3 | 1 |
| | | | | | [1]Ratio detectable: | 0.84 | 0.72 | 0.64 | 0.48 | 0.32 | 0.12 | 0.04 |

[1]Ratio detectable is calculated as total peptides divided by 25, which is the maximum detectable peptides from manual searching (see Appendix A.2).

Computational detection rate of dichloride-tagged single cysteine-containing BSA peptides in MS experiments where the inidicated amount of labeled BSA was trypsinized with 10 µg of Jurkat whole cell lysate. LC-MS analysis was performed on an LTQ-Orbitrap XL mass spectrometer (resolution = 60,000).

| Start | Stop | Sequence | Mods. | Missed Cleavages | Monoisotopic Mass | 3.0 pmoles | 1.5 pmoles | 0.8 pmoles | 0.3 pmoles | 0.15 pmoles | 0.08 pmoles | 0.03 pmoles |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 223 | 228 | (R)CASIQK(F) | | 0 | 649.334 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 198 | 204 | (K)GACLLPK(I) | | 0 | 701.402 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 483 | 489 | (R)LCVLHEK(T) | | 0 | 841.460 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 221 | 228 | (R)LRCASIQK(F) | | 1 | 918.519 | NA | NA | NA | NA | NA | NA | NA |
| 123 | 130 | (R)NECFLSHK(D) | | 0 | 977.451 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 413 | 420 | (K)QNCDQFEK(L) | pyroGlu | 0 | 994.394 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 413 | 420 | (K)QNCDQFEK(L) | | 0 | 1011.420 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 310 | 318 | (K)SHCIAEVEK(D) | | 0 | 1015.488 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 588 | 597 | (K)EACFAVEGPK(L) | | 0 | 1050.493 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 223 | 232 | (R)CASIQKFGER(A) | | 1 | 1138.567 | NA | NA | NA | NA | NA | NA | NA |
| 219 | 228 | (R)QRLRCASIQK(F) | pyroGlu | 2 | 1185.652 | NA | NA | NA | NA | NA | NA | NA |
| 337 | 346 | (K)DVCKNYQEAK(D) | | 1 | 1197.557 | NA | NA | NA | NA | NA | NA | NA |
| 219 | 228 | (R)QRLRCASIQK(F) | | 2 | 1202.679 | NA | NA | NA | NA | NA | NA | NA |
| 198 | 209 | (K)GACLLPKIETMR(E) | | 1 | 1331.717 | NA | NA | NA | NA | NA | NA | NA |
| 198 | 209 | (K)GACLLPKIETMR(E) | 1Met-ox | 1 | 1347.712 | NA | NA | NA | NA | NA | NA | NA |
| 89 | 100 | (K)SLHTLFGDELCK(V) | | 0 | 1362.672 | 2 | 3 | 2 | 2 | 0 | 0 | 0 |
| 286 | 297 | (K)YICDNQDTISSK(L) | | 0 | 1386.621 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 221 | 232 | (R)LRCASIQKFGER(A) | | 2 | 1407.753 | NA | NA | NA | NA | NA | NA | NA |
| 223 | 235 | (R)CASIQKFGERALK(A) | | 2 | 1450.784 | NA | NA | NA | NA | NA | NA | NA |
| 483 | 495 | (R)LCVLHEKTPVSEK(V) | | 1 | 1482.799 | NA | NA | NA | NA | NA | NA | NA |
| 387 | 399 | (K)DDPHACYSTVFDK(L) | | 0 | 1497.632 | 2 | 2 | 2 | 0 | 0 | 0 | 0 |
| 139 | 151 | (K)LKPDPNTLCDEFK(A) | | 0 | 1519.746 | 2 | 2 | 2 | 2 | 2 | 0 | 0 |
| 198 | 211 | (K)GACLLPKIETMREK(V) | | 2 | 1588.855 | NA | NA | NA | NA | NA | NA | NA |
| 118 | 130 | (K)QEPERNECFLSHK(D) | pyroGlu | 1 | 1599.722 | NA | NA | NA | NA | NA | NA | NA |
| 198 | 211 | (K)GACLLPKIETMREK(V) | 1Met-ox | 2 | 1604.850 | NA | NA | NA | NA | NA | NA | NA |
| 118 | 130 | (K)QEPERNECFLSHK(D) | | 1 | 1616.749 | NA | NA | NA | NA | NA | NA | NA |
| 286 | 299 | (K)YICDNQDTISSKLK(E) | | 1 | 1627.800 | NA | NA | NA | NA | NA | NA | NA |
| 469 | 482 | (R)MPCTEDYLSLILNR(L) | | 0 | 1667.813 | NA | NA | NA | NA | NA | NA | NA |
| 469 | 482 | (R)MPCTEDYLSLILNR(L) | 1Met-ox | 0 | 1683.808 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 387 | 401 | (K)DDPHACYSTVFDKLK(H) | | 1 | 1738.811 | NA | NA | NA | NA | NA | NA | NA |
| 483 | 498 | (R)LCVLHEKTPVSEKVTK(C) | | 2 | 1811.010 | NA | NA | NA | NA | NA | NA | NA |
| 508 | 523 | (R)RPCFSALTPDETYVPK(A) | | 0 | 1823.900 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 123 | 138 | (R)NECFLSHKDDSPDLPK(L) | | 1 | 1844.848 | 3 | 3 | 3 | 3 | 3 | 2 | 0 |

| Start | Stop | Sequence | Mods. | Missed Cleavages | Monoisotopic Mass | 3.0 pmoles | 1.5 pmoles | 0.8 pmoles | 0.3 pmoles | 0.15 pmoles | 0.08 pmoles | 0.03 pmoles |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 529 | 544 | (K)LFTFHADICTLPDTEK(Q) | | 0 | 1850.899 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 281 | 297 | (R)ADLAKYICDNQDTISSK(L) | | 1 | 1884.901 | NA | NA | NA | NA | NA | NA | NA |
| 89 | 105 | (K)SLHTLFGDELCKVASLR(E) | | 1 | 1888.995 | NA | NA | NA | NA | NA | NA | NA |
| 139 | 155 | (K)LKPDPNTLCDEFKADEK(K) | | 1 | 1962.948 | 3 | 3 | 3 | 3 | 3 | 2 | 0 |
| 588 | 607 | (K)EACFAVEGPKLVVSTQTALA(-) | | 1 | 2034.058 | NA | NA | NA | NA | NA | NA | NA |
| 139 | 156 | (K)LKPDPNTLCDEFKADEKK(F) | | 2 | 2091.043 | NA | NA | NA | NA | NA | NA | NA |
| 281 | 299 | (R)ADLAKYICDNQDTISSKLK(E) | | 2 | 2126.080 | NA | NA | NA | NA | NA | NA | NA |
| 529 | 547 | (K)LFTFHADICTLPDTEKQIK(K) | | 1 | 2220.137 | NA | NA | NA | NA | NA | NA | NA |
| 402 | 420 | (K)HLVDEPQNLIKQNCDQFEK(L) | | 1 | 2298.118 | NA | NA | NA | NA | NA | NA | NA |
| 529 | 548 | (K)LFTFHADICTLPDTEKQIKK(Q) | | 2 | 2348.232 | NA | NA | NA | NA | NA | NA | NA |
| 131 | 151 | (K)DDSPDLPKLKPDPNTLCDEFK(A) | | 1 | 2387.144 | NA | NA | NA | NA | NA | NA | NA |
| 319 | 340 | (K)DAIPENLPPLTADFAEDKDVCK(N) | | 1 | 2401.159 | 0 | 3 | 0 | 0 | 0 | 0 | 0 |
| 508 | 528 | (R)RPCFSALTPDETYVPKAFDEK(L) | | 1 | 2414.170 | NA | NA | NA | NA | NA | NA | NA |
| 45 | 65 | (K)GLVLIAFSQYLQQCPFDEHVK(L) | | 0 | 2435.243 | NA | NA | NA | NA | NA | NA | NA |
| 524 | 544 | (K)AFDEKLFTFHADICTLPDTEK(Q) | | 1 | 2441.169 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 413 | 433 | (K)QNCDQFEKLGEYGFQNALIVR(Y) | pyroGlu | 1 | 2455.171 | NA | NA | NA | NA | NA | NA | NA |
| 118 | 138 | (K)QEPERNECFLSHKDDSPDLPK(L) | pyroGlu | 2 | 2467.119 | NA | NA | NA | NA | NA | NA | NA |
| 413 | 433 | (K)QNCDQFEKLGEYGFQNALIVR(Y) | | 1 | 2472.198 | NA | NA | NA | NA | NA | NA | NA |
| 118 | 138 | (K)QEPERNECFLSHKDDSPDLPK(L) | | 2 | 2484.146 | NA | NA | NA | NA | NA | NA | NA |
| 400 | 420 | (K)LKHLVDEPQNLIKQNCDQFEK(L) | | 2 | 2539.297 | NA | NA | NA | NA | NA | NA | NA |
| 337 | 359 | (K)DVCKNYQEAKDAFLGSFLYEYSR(R) | | 2 | 2746.282 | NA | NA | NA | NA | NA | NA | NA |
| 524 | 547 | (K)AFDEKLFTFHADICTLPDTEKQIK(K) | | 2 | 2810.407 | NA | NA | NA | NA | NA | NA | NA |
| 131 | 155 | (K)DDSPDLPKLKPDPNTLCDEFKADEK(K) | | 2 | 2830.345 | NA | NA | NA | NA | NA | NA | NA |
| 413 | 436 | (K)QNCDQFEKLGEYGFQNALIVRYTR(K) | pyroGlu | 2 | 2875.383 | NA | NA | NA | NA | NA | NA | NA |
| 413 | 436 | (K)QNCDQFEKLGEYGFQNALIVRYTR(K) | | 2 | 2892.410 | NA | NA | NA | NA | NA | NA | NA |
| 310 | 336 | (K)SHCIAEVEKDAIPENLPPLTADFAEDK(D) | | 1 | 2952.430 | NA | NA | NA | NA | NA | NA | NA |
| 387 | 412 | (K)DDPHACYSTVFDKLKHLVDEPQNLIK(Q) | | 2 | 3025.509 | NA | NA | NA | NA | NA | NA | NA |
| 319 | 346 | (K)DAIPENLPPLTADFAEDKDVCKNYQEAK(D) | | 2 | 3134.499 | NA | NA | NA | NA | NA | NA | NA |
| 37 | 65 | (K)DLGEEHFKGLVLIAFSQYLQQCPFDEHVK(L) | | 1 | 3390.683 | NA | NA | NA | NA | NA | NA | NA |
| 45 | 75 | (K)GLVLIAFSQYLQQCPFDEHVKLVNELTEFAK(T) | | 1 | 3579.856 | NA | NA | NA | NA | NA | NA | NA |
| 35 | 65 | (R)FKDLGEEHFKGLVLIAFSQYLQQCPFDEHVK(L) | | 2 | 3665.846 | NA | NA | NA | NA | NA | NA | NA |
| 402 | 433 | (K)HLVDEPQNLIKQNCDQFEKLGEYGFQNALIVR(Y) | | 2 | 3758.896 | NA | NA | NA | NA | NA | NA | NA |
| | | | | | Total peptides: | 9 | 6 | 5 | 4 | 3 | 2 | 0 |
| | | | | | [1]Ratio detectable: | 0.47 | 0.32 | 0.26 | 0.21 | 0.16 | 0.11 | 0 |

[1]Ratio detectable is calculated as total peptides divided by 19, which is the maximum detectable peptides from manual searching (see Appendix A.2).

Computational detection rate of monobromide-tagged single cysteine-containing BSA peptides in MS experiments where the inidicated amount of labeled BSA was trypsinized with 10 µg of Jurkat whole cell lysate. LC-MS analysis was performed on an LTQ-Orbitrap XL mass spectrometer (resolution = 60,000).

| Start | Stop | Sequence | Mods. | Missed Cleavages | Monoisotopic Mass | 3.0 pmoles | 1.5 pmoles | 0.8 pmoles | 0.3 pmoles | 0.15 pmoles | 0.08 pmoles | 0.03 pmoles |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 223 | 228 | (R)CASIQK(F) | | 0 | 649.3338 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 198 | 204 | (K)GACLLPK(I) | | 0 | 701.4015 | 2 | 2 | 2 | 2 | 0 | 0 | 0 |
| 483 | 489 | (R)LCVLHEK(T) | | 0 | 841.4600 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 221 | 228 | (R)LRCASIQK(F) | | 1 | 918.5189 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 123 | 130 | (R)NECFLSHK(D) | | 0 | 977.4509 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 413 | 420 | (K)QNCDQFEK(L) | pyroGlu | 0 | 994.3935 | 0 | 2 | 0 | 0 | 0 | 0 | 0 |
| 413 | 420 | (K)QNCDQFEK(L) | | 0 | 1011.4200 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 310 | 318 | (K)SHCIAEVEK(D) | | 0 | 1015.4877 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 588 | 597 | (K)EACFAVEGPK(L) | | 0 | 1050.4925 | NA | NA | NA | NA | NA | NA | NA |
| 223 | 232 | (R)CASIQKFGER(A) | | 1 | 1138.5674 | NA | NA | NA | NA | NA | NA | NA |
| 219 | 228 | (R)QRLRCASIQK(F) | pyroGlu | 2 | 1185.6521 | NA | NA | NA | NA | NA | NA | NA |
| 337 | 346 | (K)DVCKNYQEAK(D) | | 1 | 1197.5568 | NA | NA | NA | NA | NA | NA | NA |
| 219 | 228 | (R)QRLRCASIQK(F) | | 2 | 1202.6786 | NA | NA | NA | NA | NA | NA | NA |
| 198 | 209 | (K)GACLLPKIETMR(E) | | 1 | 1331.7174 | NA | NA | NA | NA | NA | NA | NA |
| 198 | 209 | (K)GACLLPKIETMR(E) | 1Met-ox | 1 | 1347.7123 | NA | NA | NA | NA | NA | NA | NA |
| 89 | 100 | (K)SLHTLFGDELCK(V) | | 0 | 1362.6722 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 286 | 297 | (K)YICDNQDTISSK(L) | | 0 | 1386.6206 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 221 | 232 | (R)LRCASIQKFGER(A) | | 2 | 1407.7525 | NA | NA | NA | NA | NA | NA | NA |
| 223 | 235 | (R)CASIQKFGERALK(A) | | 2 | 1450.7835 | NA | NA | NA | NA | NA | NA | NA |
| 483 | 495 | (R)LCVLHEKTPVSEK(V) | | 1 | 1482.7985 | NA | NA | NA | NA | NA | NA | NA |
| 387 | 399 | (K)DDPHACYSTVFDK(L) | | 0 | 1497.6315 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 139 | 151 | (K)LKPDPNTLCDEFK(A) | | 0 | 1519.7461 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 198 | 211 | (K)GACLLPKIETMREK(V) | | 2 | 1588.8549 | NA | NA | NA | NA | NA | NA | NA |
| 118 | 130 | (K)QEPERNECFLSHK(D) | pyroGlu | 1 | 1599.7220 | NA | NA | NA | NA | NA | NA | NA |
| 198 | 211 | (K)GACLLPKIETMREK(V) | 1Met-ox | 2 | 1604.8499 | NA | NA | NA | NA | NA | NA | NA |
| 118 | 130 | (K)QEPERNECFLSHK(D) | | 1 | 1616.7486 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 286 | 299 | (K)YICDNQDTISSKLK(E) | | 1 | 1627.7996 | NA | NA | NA | NA | NA | NA | NA |
| 469 | 482 | (R)MPCTEDYLSLILNR(L) | | 0 | 1667.8131 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 469 | 482 | (R)MPCTEDYLSLILNR(L) | 1Met-ox | 0 | 1683.8081 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 387 | 401 | (K)DDPHACYSTVFDKLK(H) | | 1 | 1738.8105 | NA | NA | NA | NA | NA | NA | NA |
| 483 | 498 | (R)LCVLHEKTPVSEKVTK(C) | | 2 | 1811.0095 | NA | NA | NA | NA | NA | NA | NA |
| 508 | 523 | (R)RPCFSALTPDETYVPK(A) | | 0 | 1823.8996 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 123 | 138 | (R)NECFLSHKDDSPDLPK(L) | | 1 | 1844.8483 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| Start | Stop | Sequence | Mods. | Missed Cleavages | Monoisotopic Mass | 3.0 pmoles | 1.5 pmoles | 0.8 pmoles | 0.3 pmoles | 0.15 pmoles | 0.08 pmoles | 0.03 pmoles |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 529 | 544 | (K)LFTFHADICTLPDTEK(Q) | | 0 | 1850.8993 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 281 | 297 | (R)ADLAKYICDNQDTISSK(L) | | 1 | 1884.9008 | NA | NA | NA | NA | NA | NA | NA |
| 89 | 105 | (K)SLHTLFGDELCKVASLR(E) | | 1 | 1888.9949 | NA | NA | NA | NA | NA | NA | NA |
| 139 | 155 | (K)LKPDPNTLCDEFKADEK(K) | | 1 | 1962.9477 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 588 | 607 | (K)EACFAVEGPKLVVSTQTALA(-) | | 1 | 2034.0576 | NA | NA | NA | NA | NA | NA | NA |
| 139 | 156 | (K)LKPDPNTLCDEFKADEKK(F) | | 2 | 2091.0427 | NA | NA | NA | NA | NA | NA | NA |
| 281 | 299 | (R)ADLAKYICDNQDTISSKLK(E) | | 2 | 2126.0798 | NA | NA | NA | NA | NA | NA | NA |
| 529 | 547 | (K)LFTFHADICTLPDTEKQIK(K) | | 1 | 2220.1369 | NA | NA | NA | NA | NA | NA | NA |
| 402 | 420 | (K)HLVDEPQNLIKQNCDQFEK(L) | | 1 | 2298.1183 | NA | NA | NA | NA | NA | NA | NA |
| 529 | 548 | (K)LFTFHADICTLPDTEKQIKK(Q) | | 2 | 2348.2319 | NA | NA | NA | NA | NA | NA | NA |
| 131 | 151 | (K)DDSPDLPKLKPDPNTLCDEFK(A) | | 1 | 2387.1435 | NA | NA | NA | NA | NA | NA | NA |
| 319 | 340 | (K)DAIPENLPPLTADFAEDKDVCK(N) | | 1 | 2401.1592 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 508 | 528 | (R)RPCFSALTPDETYVPKAFDEK(L) | | 1 | 2414.1697 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 45 | 65 | (K)GLVLIAFSQYLQQCPFDEHVK(L) | | 0 | 2435.2428 | 3 | 3 | 3 | 2 | 3 | 0 | 0 |
| 524 | 544 | (K)AFDEKLFTFHADICTLPDTEK(Q) | | 1 | 2441.1693 | NA | NA | NA | NA | NA | NA | NA |
| 413 | 433 | (K)QNCDQFEKLGEYGFQNALIVR(Y) | pyroGlu | 1 | 2455.1711 | NA | NA | NA | NA | NA | NA | NA |
| 118 | 138 | (K)QEPERNECFLSHKDDSPDLPK(L) | pyroGlu | 2 | 2467.1194 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 413 | 433 | (K)QNCDQFEKLGEYGFQNALIVR(Y) | | 1 | 2472.1976 | NA | NA | NA | NA | NA | NA | NA |
| 118 | 138 | (K)QEPERNECFLSHKDDSPDLPK(L) | | 2 | 2484.1460 | 2 | 0 | 3 | 0 | 0 | 0 | 0 |
| 400 | 420 | (K)LKHLVDEPQNLIKQNCDQFEK(L) | | 2 | 2539.2973 | NA | NA | NA | NA | NA | NA | NA |
| 337 | 359 | (K)DVCKNYQEAKDAFLGSFLYEYSR(R) | | 2 | 2746.2817 | NA | NA | NA | NA | NA | NA | NA |
| 524 | 547 | (K)AFDEKLFTFHADICTLPDTEKQIK(K) | | 2 | 2810.4069 | NA | NA | NA | NA | NA | NA | NA |
| 131 | 155 | (K)DDSPDLPKLKPDPNTLCDEFKADEK(K) | | 2 | 2830.3451 | NA | NA | NA | NA | NA | NA | NA |
| 413 | 436 | (K)QNCDQFEKLGEYGFQNALIVRYTR(K) | pyroGlu | 2 | 2875.3832 | NA | NA | NA | NA | NA | NA | NA |
| 413 | 436 | (K)QNCDQFEKLGEYGFQNALIVRYTR(K) | | 2 | 2892.4097 | NA | NA | NA | NA | NA | NA | NA |
| 310 | 336 | (K)SHCIAEVEKDAIPENLPPLTADFAEDK(D) | | 1 | 2952.4295 | NA | NA | NA | NA | NA | NA | NA |
| 387 | 412 | (K)DDPHACYSTVFDKLKHLVDEPQNLIK(Q) | | 2 | 3025.5088 | NA | NA | NA | NA | NA | NA | NA |
| 319 | 346 | (K)DAIPENLPPLTADFAEDKDVCKNYQEAK(D) | | 2 | 3134.4987 | NA | NA | NA | NA | NA | NA | NA |
| 37 | 65 | (K)DLGEEHFKGLVLIAFSQYLQQCPFDEHVK(L) | | 1 | 3390.6827 | NA | NA | NA | NA | NA | NA | NA |
| 45 | 75 | (K)GLVLIAFSQYLQQCPFDEHVKLVNELTEFAK(T) | | 1 | 3579.8556 | NA | NA | NA | NA | NA | NA | NA |
| 35 | 65 | (R)FKDLGEEHFKGLVLIAFSQYLQQCPFDEHVK(L) | | 2 | 3665.8461 | NA | NA | NA | NA | NA | NA | NA |
| 402 | 433 | (K)HLVDEPQNLIKQNCDQFEKLGEYGFQNALIVR(Y) | | 2 | 3758.8959 | NA | NA | NA | NA | NA | NA | NA |
| | | | | | Total peptides: | 7 | 3 | 3 | 2 | 1 | 0 | 0 |
| | | | | | [1]Ratio detectable: | 0.29 | 0.13 | 0.13 | 0.08 | 0.04 | 0 | 0 |

[1]Ratio detectable is calculated as total peptides divided by 24, which is the maximum detectable peptides from manual searching (see Appendix A.2).

Computational detection rate of dibromide-tagged single cysteine-containing BSA peptides in MS experiments where the inidicated amount of labeled BSA was trypsinized with 10 µg of Jurkat whole cell lysate. LC-MS analysis was performed on an Waters QTof Premier mass spectrometer (resolution = 10,000).

| Start | Stop | Sequence | Mods. | Missed Cleavages | Monoisotopic Mass | 3.0 pmoles | 1.5 pmoles | 0.8 pmoles | 0.3 pmoles | 0.15 pmoles | 0.08 pmoles | 0.03 pmoles |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 223 | 228 | (R)CASIQK(F) | | 0 | 649.3338 | 2 | 2 | 2 | 2 | 2 | 0 | 2 |
| 198 | 204 | (K)GACLLPK(I) | | 0 | 701.4015 | 2 | 2 | 2 | 2 | 2 | 0 | 0 |
| 483 | 489 | (R)LCVLHEK(T) | | 0 | 841.4600 | 3 | 3 | 3 | 3 | 2 | 0 | 0 |
| 221 | 228 | (R)LRCASIQK(F) | | 1 | 918.5189 | NA | NA | NA | NA | NA | NA | NA |
| 123 | 130 | (R)NECFLSHK(D) | | 0 | 977.4509 | 3 | 3 | 2 | 0 | 0 | 0 | 0 |
| 413 | 420 | (K)QNCDQFEK(L) | pyroGlu | 0 | 994.3935 | 2 | 2 | 2 | 0 | 0 | 0 | 0 |
| 413 | 420 | (K)QNCDQFEK(L) | | 0 | 1011.4200 | 0 | 2 | 0 | 0 | 0 | 0 | 0 |
| 310 | 318 | (K)SHCIAEVEK(D) | | 0 | 1015.4877 | 3 | 3 | 3 | 3 | 2 | 0 | 0 |
| 588 | 597 | (K)EACFAVEGPK(L) | | 0 | 1050.4925 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 223 | 232 | (R)CASIQKFGER(A) | | 1 | 1138.5674 | NA | NA | NA | NA | NA | NA | NA |
| 219 | 228 | (R)QRLRCASIQK(F) | pyroGlu | 2 | 1185.6521 | NA | NA | NA | NA | NA | NA | NA |
| 337 | 346 | (K)DVCKNYQEAK(D) | | 1 | 1197.5568 | NA | NA | NA | NA | NA | NA | NA |
| 219 | 228 | (R)QRLRCASIQK(F) | | 2 | 1202.6786 | NA | NA | NA | NA | NA | NA | NA |
| 198 | 209 | (K)GACLLPKIETMR(E) | | 1 | 1331.7174 | NA | NA | NA | NA | NA | NA | NA |
| 198 | 209 | (K)GACLLPKIETMR(E) | 1Met-ox | 1 | 1347.7123 | NA | NA | NA | NA | NA | NA | NA |
| 89 | 100 | (K)SLHTLFGDELCK(V) | | 0 | 1362.6722 | 2 | 2 | 2 | 2 | 2 | 0 | 0 |
| 286 | 297 | (K)YICDNQDTISSK(L) | | 0 | 1386.6206 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| 221 | 232 | (R)LRCASIQKFGER(A) | | 2 | 1407.7525 | NA | NA | NA | NA | NA | NA | NA |
| 223 | 235 | (R)CASIQKFGERALK(A) | | 2 | 1450.7835 | NA | NA | NA | NA | NA | NA | NA |
| 483 | 495 | (R)LCVLHEKTPVSEK(V) | | 1 | 1482.7985 | NA | NA | NA | NA | NA | NA | NA |
| 387 | 399 | (K)DDPHACYSTVFDK(L) | | 0 | 1497.6315 | 2 | 2 | 0 | 2 | 2 | 0 | 0 |
| 139 | 151 | (K)LKPDPNTLCDEFK(A) | | 0 | 1519.7461 | 2 | 2 | 2 | 2 | 0 | 0 | 0 |
| 198 | 211 | (K)GACLLPKIETMREK(V) | | 2 | 1588.8549 | NA | NA | NA | NA | NA | NA | NA |
| 118 | 130 | (K)QEPERNECFLSHK(D) | pyroGlu | 1 | 1599.7220 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 198 | 211 | (K)GACLLPKIETMREK(V) | 1Met-ox | 2 | 1604.8499 | NA | NA | NA | NA | NA | NA | NA |
| 118 | 130 | (K)QEPERNECFLSHK(D) | | 1 | 1616.7486 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 286 | 299 | (K)YICDNQDTISSKLK(E) | | 1 | 1627.7996 | NA | NA | NA | NA | NA | NA | NA |
| 469 | 482 | (R)MPCTEDYLSLILNR(L) | | 0 | 1667.8131 | 2 | 2 | 0 | 2 | 0 | 0 | 0 |
| 469 | 482 | (R)MPCTEDYLSLILNR(L) | 1Met-ox | 0 | 1683.8081 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 387 | 401 | (K)DDPHACYSTVFDKLK(H) | | 1 | 1738.8105 | NA | NA | NA | NA | NA | NA | NA |
| 483 | 498 | (R)LCVLHEKTPVSEKVTK(C) | | 2 | 1811.0095 | NA | NA | NA | NA | NA | NA | NA |
| 508 | 523 | (R)RPCFSALTPDETYVPK(A) | | 0 | 1823.8996 | 2 | 2 | 2 | 2 | 2 | 0 | 0 |
| 123 | 138 | (R)NECFLSHKDDSPDLPK(L) | | 1 | 1844.8483 | 3 | 3 | 2 | 0 | 0 | 0 | 0 |

| Start | Stop | Sequence | Mods. | Missed Cleavages | Monoisotopic Mass | 3.0 pmoles | 1.5 pmoles | 0.8 pmoles | 0.3 pmoles | 0.15 pmoles | 0.08 pmoles | 0.03 pmoles |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 529 | 544 | (K)LFTFHADICTLPDTEK(Q) | | 0 | 1850.8993 | 2 | 2 | 0 | 0 | 0 | 0 | 0 |
| 281 | 297 | (R)ADLAKYICDNQDTISSK(L) | | 1 | 1884.9008 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 89 | 105 | (K)SLHTLFGDELCKVASLR(E) | | 1 | 1888.9949 | NA | NA | NA | NA | NA | NA | NA |
| 139 | 155 | (K)LKPDPNTLCDEFKADEK(K) | | 1 | 1962.9477 | 3 | 3 | 3 | 2 | 2 | 0 | 0 |
| 588 | 607 | (K)EACFAVEGPKLVVSTQTALA(-) | | 1 | 2034.0576 | NA | NA | NA | NA | NA | NA | NA |
| 139 | 156 | (K)LKPDPNTLCDEFKADEKK(F) | | 2 | 2091.0427 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 281 | 299 | (R)ADLAKYICDNQDTISSKLK(E) | | 2 | 2126.0798 | NA | NA | NA | NA | NA | NA | NA |
| 529 | 547 | (K)LFTFHADICTLPDTEKQIK(K) | | 1 | 2220.1369 | NA | NA | NA | NA | NA | NA | NA |
| 402 | 420 | (K)HLVDEPQNLIKQNCDQFEK(L) | | 1 | 2298.1183 | NA | NA | NA | NA | NA | NA | NA |
| 529 | 548 | (K)LFTFHADICTLPDTEKQIKK(Q) | | 2 | 2348.2319 | NA | NA | NA | NA | NA | NA | NA |
| 131 | 151 | (K)DDSPDLPKLKPDPNTLCDEFK(A) | | 1 | 2387.1435 | NA | NA | NA | NA | NA | NA | NA |
| 319 | 340 | (K)DAIPENLPPLTADFAEDKDVCK(N) | | 1 | 2401.1592 | 2 | 2 | 0 | 0 | 2 | 0 | 0 |
| 508 | 528 | (R)RPCFSALTPDETYVPKAFDEK(L) | | 1 | 2414.1697 | NA | NA | NA | NA | NA | NA | NA |
| 45 | 65 | (K)GLVLIAFSQYLQQCPFDEHVK(L) | | 0 | 2435.2428 | NA | NA | NA | NA | NA | NA | NA |
| 524 | 544 | (K)AFDEKLFTFHADICTLPDTEK(Q) | | 1 | 2441.1693 | NA | NA | NA | NA | NA | NA | NA |
| 413 | 433 | (K)QNCDQFEKLGEYGFQNALIVR(Y) | pyroGlu | 1 | 2455.1711 | NA | NA | NA | NA | NA | NA | NA |
| 118 | 138 | (K)QEPERNECFLSHKDDSPDLPK(L) | pyroGlu | 2 | 2467.1194 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 413 | 433 | (K)QNCDQFEKLGEYGFQNALIVR(Y) | | 1 | 2472.1976 | NA | NA | NA | NA | NA | NA | NA |
| 118 | 138 | (K)QEPERNECFLSHKDDSPDLPK(L) | | 2 | 2484.1460 | NA | NA | NA | NA | NA | NA | NA |
| 400 | 420 | (K)LKHLVDEPQNLIKQNCDQFEK(L) | | 2 | 2539.2973 | NA | NA | NA | NA | NA | NA | NA |
| 337 | 359 | (K)DVCKNYQEAKDAFLGSFLYEYSR(R) | | 2 | 2746.2817 | NA | NA | NA | NA | NA | NA | NA |
| 524 | 547 | (K)AFDEKLFTFHADICTLPDTEKQIK(K) | | 2 | 2810.4069 | NA | NA | NA | NA | NA | NA | NA |
| 131 | 155 | (K)DDSPDLPKLKPDPNTLCDEFKADEK(K) | | 2 | 2830.3451 | NA | NA | NA | NA | NA | NA | NA |
| 413 | 436 | (K)QNCDQFEKLGEYGFQNALIVRYTR(K) | pyroGlu | 2 | 2875.3832 | NA | NA | NA | NA | NA | NA | NA |
| 413 | 436 | (K)QNCDQFEKLGEYGFQNALIVRYTR(K) | | 2 | 2892.4097 | NA | NA | NA | NA | NA | NA | NA |
| 310 | 336 | (K)SHCIAEVEKDAIPENLPPLTADFAEDK(D) | | 1 | 2952.4295 | NA | NA | NA | NA | NA | NA | NA |
| 387 | 412 | (K)DDPHACYSTVFDKLKHLVDEPQNLIK(Q) | | 2 | 3025.5088 | NA | NA | NA | NA | NA | NA | NA |
| 319 | 346 | (K)DAIPENLPPLTADFAEDKDVCKNYQEAK(D) | | 2 | 3134.4987 | 2 | 2 | 0 | 0 | 0 | 0 | 0 |
| 37 | 65 | (K)DLGEEHFKGLVLIAFSQYLQQCPFDEHVK(L) | | 1 | 3390.6827 | NA | NA | NA | NA | NA | NA | NA |
| 45 | 75 | (K)GLVLIAFSQYLQQCPFDEHVKLVNELTEFAK(T) | | 1 | 3579.8556 | NA | NA | NA | NA | NA | NA | NA |
| 35 | 65 | (R)FKDLGEEHFKGLVLIAFSQYLQQCPFDEHVK(L) | | 2 | 3665.8461 | NA | NA | NA | NA | NA | NA | NA |
| 402 | 433 | (K)HLVDEPQNLIKQNCDQFEKLGEYGFQNALIVR(Y) | | 2 | 3758.8959 | NA | NA | NA | NA | NA | NA | NA |
| | | | | | Total peptides: | 20 | 19 | 13 | 12 | 11 | 2 | 2 |
| | | | | | [1]Ratio detectable: | 0.80 | 0.76 | 0.52 | 0.48 | 0.44 | 0.08 | 0.08 |

[1]Ratio detectable is calculated as total peptides divided by 25, which is the maximum detectable peptides from manual searching (see Appendix A.2).

## A.6 Manual detection rate for each single cysteine-containing BSA peptide

The manual detection rate for each single cysteine-containing BSA peptide from searching a sample containing 30 picomoles of halogen-tagged BSA. Numbers in each column indicate the number of ions manually found for each peptide listed (where the ion could be in the +1 to +5 charge state). Data were obtained on an LTQ Orbitrap XL (resolution = 60,000) or Waters QTof Premier (resolution = 10,000) mass spectrometer.

Manual detection rate of halogen-tagged single cysteine-containing BSA peptide in LC-MS experiments where 30 picomoles of the indicated tagged BSA sample was trypsinized in the absence of lysate. LC-MS was performed on either an LTQ-Orbitrap XL (resolution = 60,000) or Waters QTof Premier (resolution = 10,000).

| Start | Stop | Sequence | Mods. | Missed Cleavages | Monoisotopic Mass | Dibromide | Dichloride | Monobromide | Dibromide Qtof |
|---|---|---|---|---|---|---|---|---|---|
| 223 | 228 | (R)CASIQK(F) | | 0 | 649.3338 | 2 | 2 | 2 | 2 |
| 198 | 204 | (K)GACLLPK(I) | | 0 | 701.4015 | 2 | 2 | 2 | 2 |
| 483 | 489 | (R)LCVLHEK(T) | | 0 | 841.4600 | 3 | 2 | 2 | 3 |
| 221 | 228 | (R)LRCASIQK(F) | | 1 | 918.5189 | 0 | 0 | 2 | 0 |
| 123 | 130 | (R)NECFLSHK(D) | | 0 | 977.4509 | 3 | 3 | 3 | 3 |
| 413 | 420 | (K)QNCDQFEK(L) | pyroGlu | 0 | 994.3935 | 2 | 2 | 3 | 2 |
| 413 | 420 | (K)QNCDQFEK(L) | | 0 | 1011.4200 | 2 | 2 | 2 | 2 |
| 310 | 318 | (K)SHCIAEVEK(D) | | 0 | 1015.4877 | 3 | 3 | 3 | 3 |
| 588 | 597 | (K)EACFAVEGPK(L) | | 0 | 1050.4925 | 2 | 2 | 0 | 2 |
| 223 | 232 | (R)CASIQKFGER(A) | | 1 | 1138.5674 | 0 | 0 | 0 | 0 |
| 219 | 228 | (R)QRLRCASIQK(F) | pyroGlu | 2 | 1185.6521 | 0 | 0 | 0 | 0 |
| 337 | 346 | (K)DVCKNYQEAK(D) | | 1 | 1197.5568 | 0 | 0 | 0 | 0 |
| 219 | 228 | (R)QRLRCASIQK(F) | | 2 | 1202.6786 | 0 | 0 | 0 | 0 |
| 198 | 209 | (K)GACLLPK(ETMR(E) | | 1 | 1331.7174 | 0 | 0 | 0 | 0 |
| 198 | 209 | (K)GACLLPK(ETMR(E) | 1Met-ox | 1 | 1347.7123 | 0 | 0 | 0 | 0 |
| 89 | 100 | (K)SLHTLFGDELCK(V) | | 0 | 1362.6722 | 2 | 2 | 2 | 2 |
| 286 | 297 | (K)YICDNQDTISSK(L) | | 0 | 1386.6206 | 1 | 1 | 2 | 1 |
| 221 | 232 | (R)LRCASIQKFGER(A) | | 2 | 1407.7525 | 0 | 0 | 0 | 0 |
| 223 | 235 | (R)CASIQKFGERALK(A) | | 2 | 1450.7835 | 0 | 0 | 0 | 0 |
| 483 | 495 | (R)LCVLHEKTPVSEK(V) | | 1 | 1482.7985 | 0 | 0 | 0 | 0 |
| 387 | 399 | (K)DDPHACYSTVFDK(L) | | 0 | 1497.6315 | 2 | 2 | 2 | 2 |
| 139 | 151 | (K)LKPDPNTLCDEFK(A) | | 0 | 1519.7461 | 2 | 2 | 2 | 2 |
| 198 | 211 | (K)GACLLPK(ETMREK(V) | | 2 | 1588.8549 | 0 | 0 | 0 | 0 |
| 118 | 130 | (K)QEPERNECFLSHK(D) | pyroGlu | 1 | 1599.7220 | 1 | 0 | 0 | 1 |
| 198 | 211 | (K)GACLLPK(ETMREK(V) | 1Met-ox | 2 | 1604.8499 | 0 | 0 | 0 | 0 |
| 118 | 130 | (K)QEPERNECFLSHK(D) | | 1 | 1616.7486 | 3 | 0 | 3 | 3 |
| 286 | 299 | (K)YICDNQDTISSKLK(E) | | 1 | 1627.7996 | 0 | 0 | 0 | 0 |
| 469 | 482 | (R)MPCTEDYLSLILLNR(L) | | 0 | 1667.8131 | 2 | 0 | 2 | 2 |
| 469 | 482 | (R)MPCTEDYLSLILLNR(L) | 1Met-ox | 0 | 1683.8081 | 2 | 1 | 2 | 2 |
| 387 | 401 | (K)DDPHACYSTVFDKLK(H) | | 1 | 1738.8105 | 0 | 0 | 0 | 0 |
| 483 | 498 | (R)LCVLHEKTPVSEKVTK(C) | | 2 | 1811.0095 | 0 | 0 | 0 | 0 |
| 508 | 523 | (R)RPCFSALTPDETYVPK(A) | | 0 | 1823.8996 | 3 | 2 | 3 | 3 |
| 123 | 138 | (R)NECFLSHKDDSPDLPK(L) | | 1 | 1844.8483 | 3 | 3 | 3 | 3 |

| Start | Stop | Sequence | Mods. | Missed Cleavages | Monoisotopic Mass | Dibromide | Dichloride | Monobromide | Dibromide Qtof |
|---|---|---|---|---|---|---|---|---|---|
| 529 | 544 | (K)LFTFHADICTLPDTEK(Q) | | 0 | 1850.8993 | 2 | 2 | 2 | 2 |
| 281 | 297 | (R)ADLAKYICDNQDTISSK(L) | | 1 | 1884.9008 | 3 | 0 | 0 | 3 |
| 89 | 105 | (K)SLHTLFGDELCKVASLR(E) | | 1 | 1888.9949 | 0 | 0 | 0 | 0 |
| 139 | 155 | (K)LKPDPNTLCDEFKADEK(K) | | 1 | 1962.9477 | 2 | 3 | 3 | 2 |
| 588 | 607 | (K)EACFAVEGPKLVVSTQTALA(-) | | 1 | 2034.0576 | 0 | 0 | 0 | 0 |
| 139 | 156 | (K)LKPDPNTLCDEFKADEKK(F) | | 2 | 2091.0427 | 3 | 0 | 0 | 3 |
| 281 | 299 | (R)ADLAKYICDNQDTISSKLK(E) | | 2 | 2126.0798 | 0 | 0 | 0 | 0 |
| 529 | 547 | (K)LFTFHADICTLPDTEKQIK(K) | | 1 | 2220.1369 | 0 | 0 | 0 | 0 |
| 402 | 420 | (K)HLVDEPQNLIKQNCDQFEK(L) | | 1 | 2298.1183 | 0 | 0 | 0 | 0 |
| 529 | 548 | (K)LFTFHADICTLPDTEKQIKK(Q) | | 2 | 2348.2319 | 0 | 0 | 0 | 0 |
| 131 | 151 | (K)DDSPDLPKLKPDPNTLCDEFK(A) | | 1 | 2387.1435 | 0 | 0 | 0 | 0 |
| 319 | 340 | (K)DAIPENLPPLTADFAEDKDVCK(N) | | 1 | 2401.1592 | 2 | 2 | 3 | 2 |
| 508 | 528 | (R)RPCFSALTPDETYVPKAFDEK(L) | | 1 | 2414.1697 | 0 | 0 | 1 | 0 |
| 45 | 65 | (K)GLVLIAFSQYLQQCPFDEHVK(L) | | 0 | 2435.2428 | 0 | 0 | 3 | 0 |
| 524 | 544 | (K)AFDEKLFTFHADICTLPDTEK(Q) | | 1 | 2441.1693 | 0 | 2 | 0 | 0 |
| 413 | 433 | (K)QNCDQFEKLGEYGFQNALIVR(Y) | pyroGlu | 1 | 2455.1711 | 0 | 0 | 0 | 0 |
| 118 | 138 | (K)QEPERNECFLSHKDDSPDLPK(L) | pyroGlu | 2 | 2467.1194 | 3 | 0 | 2 | 3 |
| 413 | 433 | (K)QNCDQFEKLGEYGFQNALIVR(Y) | | 1 | 2472.1976 | 0 | 0 | 0 | 0 |
| 118 | 138 | (K)QEPERNECFLSHKDDSPDLPK(L) | | 2 | 2484.1460 | 0 | 0 | 2 | 0 |
| 400 | 420 | (K)LKHLVDEPQNLIKQNCDQFEK(L) | | 2 | 2539.2973 | 0 | 0 | 0 | 0 |
| 337 | 359 | (K)DVCKNYQEAKDAFLGSFLYEYSR(R) | | 2 | 2746.2817 | 0 | 0 | 0 | 0 |
| 524 | 547 | (K)AFDEKLFTFHADICTLPDTEKQIK(K) | | 2 | 2810.4069 | 0 | 0 | 0 | 0 |
| 131 | 155 | (K)DDSPDLPKLKPDPNTLCDEFKADEK(K) | | 2 | 2830.3451 | 0 | 0 | 0 | 0 |
| 413 | 436 | (K)QNCDQFEKLGEYGFQNALIVRYTR(K) | pyroGlu | 2 | 2875.3832 | 0 | 0 | 0 | 0 |
| 413 | 436 | (K)QNCDQFEKLGEYGFQNALIVRYTR(K) | | 2 | 2892.4097 | 0 | 0 | 0 | 0 |
| 310 | 336 | (K)SHCIAEVEKDAIPENLPPLTADFAEDK(D) | | 1 | 2952.4295 | 0 | 0 | 0 | 0 |
| 387 | 412 | (K)DDPHACYSTVFDKLKHLVDEPQNLIK(Q) | | 2 | 3025.5088 | 0 | 0 | 0 | 0 |
| 319 | 346 | (K)DAIPENLPPLTADFAEDKDVCKNYQEAK(D) | | 2 | 3134.4987 | 3 | 0 | 0 | 3 |
| 37 | 65 | (K)DLGEEHFKGLVLIAFSQYLQQCPFDEHVK(L) | | 1 | 3390.6827 | 0 | 0 | 0 | 0 |
| 45 | 75 | (K)GLVLIAFSQYLQQCPFDEHVKLVNELTEFAK(T) | | 1 | 3579.8556 | 0 | 0 | 0 | 0 |
| 35 | 65 | (R)FKDLGEEHFKGLVLIAFSQYLQQCPFDEHVK(L) | | 2 | 3665.8461 | 0 | 0 | 0 | 0 |
| 402 | 433 | (K)HLVDEPQNLIKQNCDQFEKLGEYGFQNALIVR(Y) | | 2 | 3758.8959 | 0 | 0 | 0 | 0 |
| | | | | | Total peptides: | 25 | 19 | 24 | 25 |

## A.7 Comparison between data-dependent and directed detection rates for each single cysteine-containing BSA peptide

A comparison between the data-dependent detection rate and the directed detection rate for each single cysteine-containing BSA peptide from searching a sample containing the indicated quantity of tagged BSA in the presence of 10 µg of Jurkat whole cell lysate. BSA was either labeled with iodoacetamide for data-dependent analysis or dibromide tag **3.4** for directed analysis. Numbers in each column indicate the number of ions found for each peptide listed (where the ion could be in the +1 to +5 charge state). These ions were correlated to a peptide sequence from a database search using the Sequest algorithm and manual searching of the resulting peptide sequences to select the best matches; values of 0 indicate that the peptide was not found in the dataset after filtering and manual searching. Data were obtained on an LTQ Orbitrap XL (resolution = 60,000).

Comparison between the data-dependent detection rate (DDA) and the directed detection rate of iodoacetamide-tagged or dibromide-tagged, respectively, of single cysteine-containing BSA peptides in MS experiments where the inidicated amount of labeled BSA was trypsinized with 10 ug of Jurkat whole cell lysate.

| Start | Stop | Sequence | Mods. | Missed Cleavages | Monoisotopic Mass | 3.0 pmoles | | 1.5 pmoles | | 0.8 pmoles | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | DDA | Directed | DDA | Directed | DDA | Directed |
| 198 | 204 | (K)GACLLPK(I) | | 0 | 701.4015 | | 1 | | | | |
| 483 | 489 | (R)LCVLHEK(T) | | 0 | 841.4600 | | 2 | | | | |
| 413 | 420 | (K)QNCDQFEK(L) | | 0 | 1011.4200 | | 1 | | | | |
| 310 | 318 | (K)SHCIAEVEK(D) | | 0 | 1015.4877 | | 2 | | | | |
| 89 | 100 | (K)SLHTLFGDELCK(V) | | 0 | 1362.6722 | 2 | 2 | 2 | 2 | 1 | 1 |
| 286 | 297 | (K)YICDNQDTISSK(L) | | 0 | 1386.6206 | | 1 | | 1 | | 1 |
| 139 | 151 | (K)LKPDPNTLCDEFK(A) | | 0 | 1519.7461 | | 1 | | 1 | | |
| 469 | 482 | (R)MPCTEDYLSLILNR(L) | | 0 | 1667.8131 | 1 | 2 | 1 | 2 | | 2 |
| 469 | 482 | (R)MPCTEDYLSLILNR(L) | 1Met-ox | 0 | 1683.8081 | | | | 1 | | |
| 508 | 523 | (R)RPCFSALTPDETYVPK(A) | | 0 | 1823.8996 | 1 | 2 | 1 | 1 | 1 | |
| 123 | 138 | (R)NECFLSHKDDSPDLPK(L) | | 1 | 1844.8483 | 1 | | | | | |
| 529 | 544 | (K)LFTFHADICTLPDTEK(Q) | | 0 | 1850.8993 | 1 | 1 | 1 | | | |
| 139 | 155 | (K)LKPDPNTLCDEFKADEK(K) | | 1 | 1962.9477 | | 1 | | | | |
| 319 | 340 | (K)DAIPENLPPLTADFAEDKDVCK(N) | | 1 | 2401.1592 | 1 | | | | | |
| 45 | 65 | (K)GLVLIAFSQYLQQCPFDEHVK(L) | | 0 | 2435.2428 | 1 | 1 | 3 | 2 | | 2 |
| 118 | 138 | (K)QEPERNECFLSHKDDSPDLPK(L) | | 2 | 2484.1460 | | | 1 | | | |
| | | | | | Total peptides: | 7 | 12 | 6 | 7 | 2 | 4 |

# Appendix B: Detailed MS Results for Chapter 4 (Isomix)

## B.1 Tranche hash codes to access primary Orbitrap data

The primary data have been uploaded to Proteome commons repository and can be accessed with the following Tranche hash codes:

1. Log-phase lysate, chymotrypsinized, IW=4

m/4WtZoW0W65ME1k4u0grLrxQP6YRFNDA+3aw9zu2hf5Ba52oE2uT5IxF-CaW1nyQFAGP2S9sUOK8tfDkIZ+RaxxpCbsAAAAAAAABgg==

2. Stationary-phase lysate, chymotrypsinized, IW=4

PQPAoZXQJt+vLuKPbu4s9QENiXtyfekVENVR6qlDetY2tlEVTRnbuJB5ctJ7W1O25lFRkSi-82wY4imKcN5+6nlRudMYAAAAAAAABjw==

3. Log-phase lysate, trypsinized, IW=4

ZahfZfMVbddnSmhcYwinrZvbXpHRiuI594TngB0PDzoINlygo1rE7p+eq8xEcWGd+3ty-3gU6ezDwGruyc6YsQM2wwFcAAAAAAAABhA==

4. Stationary-phase lysate, trypsinized, IW=4

9fJusD5Sy/NGjKTrZLQ5c+MlVOpeb0fd/DhjNzbuZhLV7C9/4w2812IH1lheYjgiYH/bq1PvC-JUOiafzeHuVKSsBVlwAAAAAAAABgg==

5. Log-phase lysate, chymotrypsinized, IW=2

f7cLLm1EbcJ4htjYLFxIjukQJe2q391RoATznfhSL5SJqNvhxie29FpcCJlYiOWFKcxj7AyWEYn-rHqrfW5PC3+AEWWUAAAAAAAABgg==

6. Stationary-phase lysate, chymotrypsinized, IW=2

unLet+AYfHLghn30rP5wB7P34ve/wC2fvs6v0VIFgugUxNXgdsNBSPdqwigi/9ql6k4BE-JTykhH2rkpdSwDAUwYdg8AAAAAAAABiQ==

7. Log-phase lysate, trypsinized, IW=2

AF/MBpK+xzL/RTrKwyrc3+5qFmwVRAoWMh6rJyskIkYbqR1EyTwOrZW/FEsScEHIA/Vxji-aFO+3fUPvWMF3RaJQQ9n8AAAAAAAABfQ==

8. Stationary-phase lysate, trypsinized, IW=2

k6bLZj3Wtx7WC6IStBOJF9HuRUWBzDsiQ23573SZRdILadvbDufGS9O61u-zSUKTTBTu93l0+1esoJZvYE/MeIbDjBDQAAAAAAAABhA==

## B.2 The complete list of *N*-linked glycopeptides and their statistical indicators of quality

This unified list of high-confidence *N*-glycosites in the yeast proteome reflects a combination of results from multiple experiments, conducted with trypsinized and chymotrypsinized samples obtained from both stationary and mid-log phase cultures. Data were obtained on an LTQ Orbitrap XL (resolution = 60,000) mass spectrometer.

The letter and numerical code for the "Notes" column:
1. Sequence ambiguity with YBR093C
2. Sequence ambiguity with YBR092C
3. Glycosite ambiguity with AnGTNSTTNTTTAESSQL
4. Doubly glycosylated

A. Log-phase lysate, trypsinized, IW = 4
B. Log-phase lysate, chymotrypsinized, IW = 4
C. Stationary-phase lysate, trypsinized, IW = 4
D. Stationary-phase lysate, chymotrypsinized, IW = 4
E. Log-phase lysate, trypsinized, IW = 2
F. Log-phase lysate, chymotrypsinized, IW = 2
G. Stationary-phase lysate, trypsinized, IW = 2
H. Stationary-phase lysate, chymotrypsinized, IW = 2

| Protein Accession | Unique Glycosites | Sequence | z | m/z (Da) | Modifications | XCorr | Confidence | ΔM (ppm) | # Ions Matched | RT (min) | Notes |
|---|---|---|---|---|---|---|---|---|---|---|---|
| YBR023C | 1 | | | | | | | | | | |
| | | NQDGSSKPnFTVE | 2 | 814.374 | N9(HexNAc) | 2.576 | < 5% FDR | 2.806 | 13/24 | 54.884 | A |
| YBR074W | 1 | | | | | | | | | | |
| | | SILFQQQDPFnESSR | 2 | 1000.981 | N11(HexNAc) | 3.454 | < 1% FDR | 3.512 | 13/28 | 97.868 | A |
| YBR078W | 4 | | | | | | | | | | |
| | | ANnITLRDVNSISF | 2 | 884.957 | N3(HexNAc) | 2.547 | < 5% FDR | 4.388 | 10/26 | 102.368 | D |
| | | GGANFDSSSSnFScNALK | 2 | 1034.447 | N11(HexNAc), C14(Carbamidomethyl) | 3.940 | < 1% FDR | 1.558 | 16/34 | 73.090 | A |
| | | SIVSNDELSKAAFSnL | 2 | 950.478 | N15(HexNAc) | 2.711 | < 5% FDR | 1.475 | 16/30 | 101.673 | H |
| | | TVGGGFIIAnNTQLK | 2 | 869.472 | N10(HexNAc) | 3.282 | < 1% FDR | 3.831 | 14/28 | 85.633 | A |
| YBR092C | 4 | | | | | | | | | | |
| | | GYMFEnQTSFPIFAASSER | 2 | 1194.045 | N6(HexNAc) | 4.308 | < 1% FDR | 2.465 | 17/36 | 121.086 | G |
| | | SVGANLFnATL | 2 | 656.340 | N8(HexNAc) | 3.023 | < 1% FDR | 2.459 | 15/20 | 111.634 | G |
| | | TEKFQcSnDTYVRY | 2 | 1008.455 | C6(Carbamidomethyl), N8(HexNAc) | 3.974 | < 1% FDR | 4.547 | 15/26 | 63.807 | D, 1 |
| | | TTAGIIDDKNnLTAEYVPFMGNTFHK | 3 | 1034.839 | N11(HexNAc) | 3.836 | < 1% FDR | 0.647 | 26/100 | 109.775 | G |
| YBR093C | 3 | | | | | | | | | | |
| | | SVGSNLFnASVK | 2 | 714.369 | N8(HexNAc) | 2.479 | < 5% FDR | 1.731 | 17/22 | 75.296 | G |
| | | TEKFQcSnDTYVRY | 2 | 1008.455 | C6(Carbamidomethyl), N8(HexNAc) | 3.974 | < 1% FDR | 4.547 | 15/26 | 63.807 | D, 2 |
| | | TTAGIIDDKNnLTAE | 2 | 890.944 | N11(HexNAc) | 3.055 | < 5% FDR | 4.864 | 16/28 | 67.422 | D, 2 |
| YBR139W | 1 | | | | | | | | | | |
| | | DIRGPcEDnSTDGMcY | 2 | 1047.907 | C6(Carbamidomethyl), N9(HexNAc), C15(Carbamidomethyl) | 4.591 | < 1% FDR | 4.957 | 19/30 | 70.588 | B |
| YBR162C | 1 | | | | | | | | | | |
| | | FVPGSTSncTF | 2 | 711.315 | N8(HexNAc), C9(Carbamidomethyl) | 3.160 | < 1% FDR | 4.281 | 15/20 | 75.999 | D |
| YBR171W | 1 | | | | | | | | | | |
| | | SnNGTFFETEEPIVETK | 2 | 1074.007 | N2(HexNAc) | 3.961 | < 1% FDR | 5.975 | 18/32 | 92.172 | D |
| YBR229C | 2 | | | | | | | | | | |
| | | ISSFLSLSnSTADTFHL | 2 | 1023.009 | N9(HexNAc) | 3.079 | < 1% FDR | 5.661 | 15/32 | 131.519 | A |
| | | NNLQHnITLK | 2 | 700.378 | N6(HexNAc) | 3.582 | < 1% FDR | 2.898 | 15/18 | 48.829 | A |
| YBR286W | 7 | | | | | | | | | | |
| | | AHHLnYTLVPFDGR | 2 | 922.967 | N5(HexNAc) | 3.806 | < 1% FDR | 3.639 | 16/26 | 78.424 | A |
| | | AYnLTKEENSKIRVF | 2 | 1009.033 | N3(HexNAc) | 3.134 | < 1% FDR | 3.996 | 14/28 | 74.242 | F |
| | | IISFnLSDAETGK | 2 | 800.407 | N5(HexNAc) | 3.351 | < 1% FDR | 2.218 | 17/24 | 93.369 | A |
| | | IKVDDLnATAW | 2 | 725.873 | N7(HexNAc) | 2.901 | < 5% FDR | 3.216 | 13/20 | 93.868 | A |
| | | IKVDDLnATAW | 2 | 725.871 | N7(HexNAc) | 2.781 | < 5% FDR | 1.242 | 17/20 | 92.124 | E |
| | | LAnYSTPDYGHPTR | 2 | 898.925 | N3(HexNAc) | 3.077 | < 1% FDR | 3.225 | 14/26 | 53.495 | A |
| | | LAnYSTPDYGHPTR | 2 | 898.922 | N3(HexNAc) | 3.404 | < 1% FDR | 0.830 | 14/26 | 52.324 | E |

| Protein Accession | Unique Glycosites | Sequence | z | m/z (Da) | Modifications | XCorr | Confidence | ΔM (ppm) | # Ions Matched | RT (min) | Notes |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | SFAnTTAFALSPPVDGFVGK | 2 | 1116.064 | N4(HexNAc) | 3.525 | < 1% FDR | 3.172 | 16/38 | 119.635 | A |
| | | STPDYGHPTRVIGSKGHnKTmEY | 3 | 932.778 | N18(HexNAc), M21(Oxidation) | 2.758 | < 5% FDR | 0.404 | 21/88 | 46.298 | F |
| | | STPDYGHPTRVIGSKGHnKTMEY | 3 | 927.449 | N18(HexNAc) | 3.527 | < 5% FDR | 3.737 | 26/88 | 53.285 | D |
| YCL043C | 3 | | | | | | | | | | |
| | | APTYQELADTYAnATSDVLIAK | 2 | 1280.637 | N13(HexNAc) | 5.310 | < 1% FDR | 3.904 | 20/42 | 123.839 | A |
| | | DLPAYLAnETFVTPVIVQSGK | 2 | 1234.155 | N8(HexNAc) | 4.549 | < 1% FDR | 5.767 | 20/40 | 124.180 | A |
| | | DLPAYLAnETFVTPVIVQSGK | 2 | 1234.152 | N8(HexNAc) | 4.494 | < 1% FDR | 3.689 | 20/40 | 123.264 | E |
| | | DLPAYLAnETFVTPVIVQSGK | 3 | 823.102 | N8(HexNAc) | 5.077 | < 1% FDR | 1.976 | 36/80 | 123.878 | G |
| | | NSDVnNSIDYEGPR | 3 | 595.602 | N5(HexNAc) | 4.626 | < 1% FDR | 4.007 | 33/52 | 58.882 | A |
| | | NSDVnNSIDYEGPR | 2 | 892.897 | N5(HexNAc) | 3.916 | < 1% FDR | 1.579 | 18/26 | 56.426 | E |
| | | NSDVnNSIDYEGPR | 2 | 892.899 | N5(HexNAc) | 3.437 | < 1% FDR | 3.904 | 19/26 | 57.157 | G |
| YCL045C | 2 | | | | | | | | | | |
| | | KLNPGKNTDVPIVANnHSSSHIF | 4 | 674.350 | N16(HexNAc) | 3.537 | < 5% FDR | 1.409 | 33/132 | 68.708 | F |
| | | TIDEIQLDSNDHNAMVcVnSSSNHWQK | 3 | 1116.500 | C17(Carbamidomethyl), N19(HexNAc) | 4.378 | < 1% FDR | 1.287 | 28/104 | 84.656 | A |
| YCR011C | 1 | | | | | | | | | | |
| | | VFSGcnVTNEK | 2 | 730.336 | C5(Carbamidomethyl), N6(HexNAc) | 2.750 | < 5% FDR | 0.842 | 17/20 | 46.932 | A |
| YDL095W | 1 | | | | | | | | | | |
| | | NAPGESLTTFQnLTDGTK | 2 | 1050.009 | N12(HexNAc) | 3.693 | < 1% FDR | 2.302 | 18/34 | 88.442 | E |
| YDR055W | 2 | | | | | | | | | | |
| | | ANnISLTDVHSVS | 2 | 781.386 | N3(HexNAc) | 2.772 | < 5% FDR | 1.864 | 15/24 | 68.425 | H |
| | | INNSISSLnFTKL | 2 | 828.445 | N9(HexNAc) | 2.657 | < 5% FDR | 4.212 | 14/24 | 107.422 | D |
| YDR434W | 1 | | | | | | | | | | |
| | | AIVFPSASSSPDGLAFInGTR | 2 | 1156.592 | N18(HexNAc) | 5.184 | < 1% FDR | 3.364 | 20/40 | 115.509 | A |
| YEL001C | 2 | | | | | | | | | | |
| | | EMEVnGTSKF | 2 | 673.809 | N5(HexNAc) | 2.482 | < 5% FDR | 4.384 | 15/18 | 59.745 | B |
| | | GVnGTIVTYPHGYPVADITGA | 2 | 1154.078 | N3(HexNAc) | 3.743 | < 1% FDR | 2.903 | 18/40 | 101.095 | F |
| YEL002C | 2 | | | | | | | | | | |
| | | LEYLDInSTSTTVDLYDKEQR | 3 | 903.443 | N7(HexNAc) | 3.584 | < 1% FDR | 2.694 | 26/80 | 99.682 | A |
| | | LTLSPSGnDSETQYYTT | 2 | 1041.474 | N8(HexNAc) | 4.138 | < 1% FDR | 4.399 | 19/32 | 83.361 | A |
| YEL040W | 1 | | | | | | | | | | |
| | | IWPGGnSTNAPGTIAW | 2 | 923.953 | N6(HexNAc) | 2.784 | < 5% FDR | 5.614 | 11/30 | 116.897 | D |
| YEL060C | 1 | | | | | | | | | | |
| | | NGGGQDLSAFWnDTKK | 2 | 971.959 | N12(HexNAc) | 2.906 | < 5% FDR | 2.610 | 12/30 | 83.107 | A |
| YFR018C | 1 | | | | | | | | | | |
| | | TKEYNDLESNTVVSnSTLGVK | 3 | 835.086 | N15(HexNAc) | 3.169 | < 5% FDR | 4.305 | 28/80 | 75.472 | A |

| Protein Accession | Unique Glycosites | Sequence | z | m/z (Da) | Modifications | XCorr | Confidence | ΔM (ppm) | # Ions Matched | RT (min) | Notes |
|---|---|---|---|---|---|---|---|---|---|---|---|
| YGL022W | 1 | | | | | | | | | | |
| | | TTLVDNNTWnNTHIAIVGK | 2 | 1158.595 | N10(HexNAc) | 3.994 | < 1% FDR | 3.456 | 16/36 | 90.663 | A |
| YGR014W | 1 | | | | | | | | | | |
| | | ELSNLITnSSSAFYTDGMGTAK | 3 | 838.066 | N8(HexNAc) | 3.784 | < 1% FDR | 6.169 | 29/84 | 108.573 | C |
| YGR189C | 2 | | | | | | | | | | |
| | | FHnYTLDWAMDK | 2 | 873.394 | N3(HexNAc) | 2.986 | < 1% FDR | 2.862 | 17/22 | 88.666 | G |
| | | SnTSSEGYPQSPM | 2 | 795.335 | N2(HexNAc) | 3.277 | < 1% FDR | 5.240 | 16/24 | 59.941 | D |
| YGR282C | 1 | | | | | | | | | | |
| | | SYWQGQTMQnASY | 2 | 884.878 | N10(HexNAc) | 3.617 | < 1% FDR | 5.122 | 18/24 | 82.177 | D |
| YHR028C | 2 | | | | | | | | | | |
| | | NTPDYQEPNSnYTNDGK | 2 | 1081.461 | N11(HexNAc) | 3.445 | < 1% FDR | 3.854 | 15/32 | 48.986 | E |
| | | TFIHNGQnLTVESITASPDLKR | 3 | 882.796 | N8(HexNAc) | 4.021 | < 1% FDR | 3.884 | 25/84 | 85.579 | E |
| YHR101C | 1 | | | | | | | | | | |
| | | GnNTEDFQPFIDSEKR | 2 | 1051.485 | N2(HexNAc) | 3.038 | < 1% FDR | 2.256 | 11/30 | 77.871 | A |
| YHR188C | 1 | | | | v | | | | | | |
| | | SYASDIGAPLFnSTEK | 2 | 952.962 | N12(HexNAc) | 3.235 | < 1% FDR | 6.705 | 15/30 | 94.900 | C |
| YHR202W | 1 | | | | | | | | | | |
| | | SInMTDPVDAESPIFSR | 2 | 1042.495 | N3(HexNAc) | 3.842 | < 1% FDR | 2.912 | 18/32 | 104.207 | A |
| YIL015W | 3 | | | | | | | | | | |
| | | AMTIQGLGAQnKSScEHETF | 2 | 1207.545 | N11(HexNAc), C15(Carbamidomethyl) | 2.366 | < 5% FDR | -2.316 | 14/38 | 76.753 | H |
| | | AVQPTnDSMVLGDVFL | 2 | 955.975 | N6(HexNAc) | 4.753 | < 1% FDR | 5.659 | 22/30 | 135.355 | D |
| | | ISLAQANWnASEVSK | 2 | 911.961 | N9(HexNAc) | 3.773 | < 1% FDR | 3.185 | 18/28 | 82.761 | G |
| YJL002C | 2 | | | | | | | | | | |
| | | QLLAnSTTAPGDDGESEIR | 2 | 1090.021 | N5(HexNAc) | 3.167 | < 1% FDR | 2.977 | 16/36 | 67.313 | G |
| | | TnVSIETQKSY | 2 | 737.866 | N2(HexNAc) | 2.803 | < 1% FDR | 4.648 | 14/20 | 53.356 | B |
| YJL171C | 8 | | | | | | | | | | |
| | | AFYTSPGFTVnNSR | 2 | 883.426 | N11(HexNAc) | 3.822 | < 1% FDR | 7.615 | 14/26 | 79.641 | C |
| | | AFYTSPGFTVnNSR | 2 | 883.421 | N11(HexNAc) | 3.106 | < 1% FDR | 1.990 | 15/26 | 78.853 | E |
| | | DIFEAMnGTEKNHLY | 2 | 993.957 | N7(HexNAc) | 2.333 | < 5% FDR | 1.736 | 13/28 | 94.865 | H |
| | | EFEMPTETEKnSSSIGYY | 2 | 1159.007 | N11(HexNAc) | 3.056 | < 1% FDR | 3.918 | 14/34 | 93.887 | H |
| | | GYITRnTTGTM | 2 | 710.342 | N6(HexNAc) | 2.736 | < 5% FDR | 4.932 | 17/20 | 51.535 | D |
| | | IIYSnVScPKSGY | 2 | 846.910 | N5(HexNAc), C8(Carbamidomethyl) | 2.663 | < 5% FDR | 3.709 | 12/24 | 63.338 | H |
| | | KDITNnEScTcEVGDRVW | 2 | 1194.526 | N6(HexNAc), C9(Carbamidomethyl), C11(Carbamidomethyl) | 3.085 | < 5% FDR | 4.320 | 18/34 | 71.558 | D |
| | | KDITNnEScTcEVGDRVW | 2 | 1194.526 | N6(HexNAc), C9(Carbamidomethyl), C11(Carbamidomethyl) | 2.901 | < 1% FDR | 3.492 | 15/34 | 69.568 | H |
| | | QYPTNAncScW | 2 | 803.318 | N7(HexNAc), C8(Carbamidomethyl), C10(Carbamidomethyl) | 4.050 | < 1% FDR | 4.407 | 14/20 | 71.027 | D |

| Protein Accession | Unique Glycosites | Sequence | z | m/z (Da) | Modifications | XCorr | Confidence | ΔM (ppm) | # Ions Matched | RT (min) | Notes |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | TADnVTFLNHGGEASPcLGNAL | 2 | 1232.073 | N4(HexNAc), C17(Carbamidomethyl) | 2.743 | < 5% FDR | 1.165 | 12/42 | 100.132 | E |
| | | TADnVTFLNHGGEASPcLGNAL | 2 | 1232.075 | N4(HexNAc), C17(Carbamidomethyl) | 2.782 | < 1% FDR | 2.454 | 15/42 | 100.794 | G |
| YJL172W | 2 | | | | | | | | | | |
| | | AHQDVVPVNnETLSSWK | 2 | 1065.029 | N10(HexNAc) | 3.776 | < 1% FDR | 4.281 | 22/32 | 80.436 | A |
| | | EFEAIEQLLIDGFKPnR | 3 | 742.056 | N16(HexNAc) | 3.662 | < 1% FDR | 5.740 | 21/64 | 144.883 | A |
| YJL192C | 1 | | | | | | | | | | |
| | | GRLDLAASnITGFVSTR | 2 | 992.026 | N9(HexNAc) | 4.365 | < 1% FDR | 1.069 | 22/32 | 97.598 | E |
| YKL039W | 1 | | | | | | | | | | |
| | | TSTnRSLANPIMT | 2 | 805.903 | N4(HexNAc) | 2.288 | < 5% FDR | 1.140 | 13/24 | 65.621 | H |
| YKL046C | 2 | | | | | | | | | | |
| | | VYDGVSIDDncTKVTSY | 2 | 1070.984 | N10(HexNAc), C11(Carbamidomethyl) | 3.321 | < 1% FDR | 4.921 | 11/32 | 81.059 | B |
| | | YTGnQTYVDWAEK | 2 | 890.407 | N4(HexNAc) | 2.777 | < 5% FDR | 4.183 | 12/24 | 74.326 | A |
| YLR066W | 1 | | | | | | | | | | |
| | | TYGETVGnYTLTVENK | 2 | 997.481 | N8(HexNAc) | 3.021 | < 1% FDR | 1.305 | 15/30 | 81.087 | A |
| YLR084C | 3 | | | | | | | | | | |
| | | MWNQQnNTIMK | 2 | 806.874 | N6(HexNAc) | 2.746 | < 1% FDR | 2.601 | 14/20 | 63.197 | G |
| | | nSSLYADIYDNK | 2 | 804.374 | N1(HexNAc) | 2.651 | < 5% FDR | 4.286 | 17/22 | 77.222 | G |
| | | TDPSSSIMnLTYLDPLSGELK | 2 | 1243.613 | N9(HexNAc) | 3.373 | < 1% FDR | 2.138 | 16/40 | 137.970 | G |
| YLR413W | 2 | | | | | | | | | | |
| | | LVDSTNPIGTTESLITLNnMTTEEK | 3 | 976.156 | N19(HexNAc) | 4.035 | < 1% FDR | 5.871 | 29/96 | 120.655 | A |
| | | STMIAEnITQSSSAK | 2 | 886.933 | N7(HexNAc) | 3.849 | < 1% FDR | 5.044 | 18/28 | 61.405 | A |
| YLR450W | 1 | | | | | | | | | | |
| | | IPTELVSEnGTK | 2 | 746.890 | N9(HexNAc) | 2.897 | < 5% FDR | 5.336 | 14/22 | 60.436 | A |
| YML052W | 1 | | | | | | | | | | |
| | | FYWVQGnTTGIPNAGDETR | 2 | 1166.046 | N7(HexNAc) | 2.482 | < 5% FDR | 2.222 | 16/36 | 94.169 | G |
| YML130C | 2 | | | | | | | | | | |
| | | NDHVSPScnVTFN | 2 | 848.366 | C8(Carbamidomethyl), N9(HexNAc) | 3.329 | < 1% FDR | 3.387 | 14/24 | 58.113 | A |
| | | TnNSQSHVFDDLK | 2 | 855.401 | N2(HexNAc) | 3.833 | < 1% FDR | 3.744 | 18/24 | 59.024 | A |
| YMR006C | 3 | | | | | | | | | | |
| | | nLTDLEYIPPLVVYIPNTK | 2 | 1204.154 | N1(HexNAc) | 2.349 | < 5% FDR | 3.742 | 15/36 | 141.700 | G |
| | | SIVNPGGSnLTYTIER | 2 | 963.495 | N9(HexNAc) | 3.731 | < 1% FDR | 5.427 | 19/30 | 84.696 | G |
| | | SRATSnFSDTSLL | 2 | 802.392 | N6(HexNAc) | 2.326 | < 5% FDR | 3.007 | 11/24 | 80.351 | H |
| YMR008C | 6 | | | | | | | | | | |
| | | AMLSGAGmLAAMDnR | 2 | 865.398 | M8(Oxidation), N14(HexNAc) | 3.684 | < 1% FDR | 3.659 | 15/28 | 80.955 | G |
| | | GnFTDDSDFLGcVGcAIIR | 2 | 1161.518 | N2(HexNAc), C12(Carbamidomethyl), C15(Carbamidomethyl) | 4.331 | < 1% FDR | -0.097 | 21/36 | 125.601 | G |

| Protein Accession | Unique Glycosites | Sequence | z | m/z (Da) | Modifications | XCorr | Confidence | ΔM (ppm) | # Ions Matched | RT (min) | Notes |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | IPNSRHSFNGnQSTF | 2 | 955.951 | N11(HexNAc) | 2.325 | < 5% FDR | 2.068 | 12/28 | 53.161 | F |
| | | LQKnATSSIIESEY | 2 | 894.447 | N4(HexNAc) | 3.078 | < 1% FDR | 3.142 | 15/26 | 70.942 | H |
| | | NRATSnFSDTSLL | 2 | 815.899 | N6(HexNAc) | 2.538 | < 5% FDR | 4.420 | 13/24 | 82.374 | D |
| | | TSVQAIVDnTTESNSIW | 2 | 1035.495 | N9(HexNAc) | 2.920 | < 1% FDR | 1.809 | 14/32 | 113.854 | H |
| YMR058W | 5 | | | | | | | | | | |
| | | EIDGITTEKnVTDMLY | 2 | 1023.992 | N10(HexNAc) | 3.146 | < 1% FDR | 3.695 | 14/30 | 105.857 | B |
| | | FDDTMLDVIPSDLQLnAT | 2 | 1107.031 | N16(HexNAc) | 2.856 | < 5% FDR | 4.341 | 11/34 | 140.853 | A |
| | | NGVNYAFFNnITYTAPK | 2 | 1070.022 | N10(HexNAc) | 5.408 | < 1% FDR | 3.368 | 22/32 | 108.765 | A |
| | | SSGDQAnNSEIYGSNTHTF | 2 | 1117.477 | N7(HexNAc) | 3.780 | < 1% FDR | 3.277 | 14/36 | 67.245 | B |
| | | TVLVHTKnDTDKNF | 2 | 918.972 | N8(HexNAc) | 2.695 | < 5% FDR | 6.616 | 9/26 | 47.402 | B |
| YMR200W | 1 | | | | | | | | | | |
| | | SRYnQTETFKEY | 2 | 885.912 | N4(HexNAc) | 2.440 | < 5% FDR | 4.308 | 10/22 | 55.966 | B |
| YMR238W | 1 | | | | | | | | | | |
| | | DGAEIDTncTDITK | 2 | 879.391 | N8(HexNAc), C9(Carbamidomethyl) | 3.369 | < 1% FDR | 5.172 | 17/26 | 57.779 | A |
| YMR297W | 2 | | | | | | | | | | |
| | | ILGIDPnVTQY | 2 | 719.377 | N7(HexNAc) | 2.826 | < 5% FDR | 6.385 | 14/20 | 106.398 | C |
| | | VRnWTASITDEVAGEVK | 2 | 1040.532 | N3(HexNAc) | 4.878 | < 1% FDR | 4.842 | 24/32 | 96.341 | E |
| YMR307W | 7 | | | | | | | | | | |
| | | AGNEVTNnYTNTDASAF | 2 | 997.432 | N8(HexNAc) | 2.951 | < 1% FDR | 1.861 | 14/32 | 74.841 | H |
| | | AInTTLDHSEcm | 2 | 806.844 | N3(HexNAc), C11(Carbamidomethyl), M12(Oxidation) | 3.305 | < 1% FDR | 4.158 | 15/22 | 48.731 | B |
| | | AInTTLDHSEcM | 2 | 798.847 | N3(HexNAc), C11(Carbamidomethyl) | 3.310 | < 1% FDR | 4.172 | 17/22 | 62.216 | D |
| | | FFYSNnGSQFY | 2 | 789.840 | N6(HexNAc) | 3.165 | < 1% FDR | 3.901 | 16/20 | 102.130 | A |
| | | FFYSNnGSQFY | 2 | 789.841 | N6(HexNAc) | 3.558 | < 1% FDR | 6.228 | 16/20 | 101.674 | C |
| | | FNWIcNEVDcSGISAnGTAGK | 2 | 1253.055 | C5(Carbamidomethyl), C10(Carbamidomethyl), N16(HexNAc) | 5.219 | < 1% FDR | 3.825 | 21/40 | 100.582 | A |
| | | FNWIcNEVDcSGISAnGTAGK | 2 | 1253.055 | C5(Carbamidomethyl), C10(Carbamidomethyl), N16(HexNAc) | 3.626 | < 1% FDR | 3.922 | 20/40 | 99.617 | G |
| | | GVAYQADTAnETSGSTVNDPLANYEScSR | 3 | 1094.818 | N10(HexNAc), C27(Carbamidomethyl) | 7.550 | < 1% FDR | 4.369 | 51/112 | 86.602 | A |
| | | GVAYQADTAnETSGSTVNDPLANYEScSR | 3 | 1094.815 | N10(HexNAc), C27(Carbamidomethyl) | 5.655 | < 1% FDR | 1.421 | 33/112 | 84.703 | E |
| | | KTVVDTFAnYTNV | 2 | 838.922 | N9(HexNAc) | 3.207 | < 1% FDR | 4.673 | 16/24 | 95.034 | D |
| | | nLSIPVFFSEYGcNEVTPR | 2 | 1217.584 | N1(HexNAc), C13(Carbamidomethyl) | 3.196 | < 1% FDR | 4.263 | 15/36 | 129.348 | A |
| YNL160W | 9 | | | | | | | | | | |
| | | ANGTnSTTNTTAESSQL | 2 | 1001.957 | N5(HexNAc) | 5.235 | < 1% FDR | 5.428 | 21/34 | 50.050 | D, 3 |
| | | EYSnFTTPYGL | 2 | 748.844 | N4(HexNAc) | 2.606 | < 5% FDR | 7.115 | 11/20 | 102.818 | C |
| | | EYSnFTTPYGL | 2 | 748.840 | N4(HexNAc) | 2.306 | < 5% FDR | 2.190 | 10/20 | 102.624 | G |
| | | GGQVPITnSSLTHTNY | 2 | 947.463 | N8(HexNAc) | 2.878 | < 5% FDR | 4.875 | 15/30 | 72.071 | D |
| | | NSSSALnITELY | 2 | 758.872 | N7(HexNAc) | 2.914 | < 5% FDR | 4.724 | 15/22 | 107.773 | D |

| Protein Accession | Unique Glycosites | Sequence | z | m/z (Da) | Modifications | XCorr | Confidence | ΔM (ppm) | # Ions Matched | RT (min) | Notes |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | NSSSALnITELY | 2 | 758.869 | N7(HexNAc) | 2.603 | < 5% FDR | 1.681 | 15/22 | 106.177 | H |
| | | NVARVnETIQDK | 2 | 845.951 | N7(HexNAc) | 3.667 | < 1% FDR | 4.421 | 15/24 | 62.485 | B |
| | | NVARVnETIQDK | 2 | 845.951 | N7(HexNAc) | 3.654 | < 1% FDR | 4.337 | 15/24 | 61.918 | D |
| | | QIIVTGGQVPITNSSLTHTnY | 2 | 1224.634 | N20(HexNAc) | 3.168 | < 1% FDR | 3.363 | 18/40 | 96.261 | H |
| | | QIIVTGGQVPITnSSLTHTnYTR | 3 | 970.841 | N13(HexNAc), N20(HexNAc) | 3.269 | < 5% FDR | 6.943 | 18/88 | 84.806 | C, 4 |
| | | SGSTnSTSSTIESTEIPVVY | 2 | 1132.538 | N5(HexNAc) | 4.173 | < 1% FDR | 5.129 | 22/38 | 102.071 | D |
| | | SIIFDTEKPIVVTEDSAYAIPVANNKnATKRGVL | 4 | 970.526 | N27(HexNAc) | 4.694 | < 1% FDR | 3.798 | 45/198 | 111.122 | D |
| | | TRLFnSSSAL | 2 | 650.839 | N5(HexNAc) | 2.958 | < 5% FDR | 4.232 | 15/18 | 75.130 | D |
| YNL327W | 2 | | | | | | | | | | |
| | | STNnFTISIPEK | 2 | 778.394 | N4(HexNAc) | 2.505 | < 5% FDR | 2.856 | 14/22 | 81.855 | G |
| | | VAKnTSAITTDGGIY | 2 | 858.438 | N4(HexNAc) | 3.237 | < 1% FDR | 4.489 | 16/28 | 63.758 | D |
| YNR067C | 1 | | | | | | | | | | |
| | | LnGTWAADNKDWVNSL | 2 | 1004.982 | N2(HexNAc) | 2.729 | < 1% FDR | 2.418 | 12/30 | 109.110 | G |
| YOL011W | 3 | | | | | | | | | | |
| | | ATAnFSDSSEVLSK | 2 | 830.897 | N4(HexNAc) | 2.302 | < 5% FDR | 2.973 | 15/26 | 64.412 | G |
| | | EATSISQnESAWLEK | 2 | 949.455 | N8(HexNAc) | 2.361 | < 5% FDR | 4.970 | 15/28 | 77.707 | G |
| | | NnFTDDPEFMGcVGcAIIR | 2 | 1211.032 | N2(HexNAc), C12(Carbamidomethyl), C15(Carbamidomethyl) | 4.539 | < 5% FDR | 5.391 | 22/36 | 115.050 | G |
| YOL030W | 5 | | | | | | | | | | |
| | | AGNEVINSVnTTNTATY | 2 | 987.466 | N10(HexNAc) | 4.004 | < 1% FDR | 2.199 | 17/32 | 73.841 | H |
| | | ASSSnSSTPSIEIK | 2 | 806.897 | N5(HexNAc) | 2.667 | < 5% FDR | 3.271 | 13/26 | 55.533 | A |
| | | GVDYQPGGSSnLTDPLADASVcDR | 3 | 900.406 | N11(HexNAc), C22(Carbamidomethyl) | 4.427 | < 1% FDR | 0.739 | 34/92 | 94.110 | A |
| | | GVDYQPGGSSnLTDPLADASVcDR | 2 | 1350.109 | N11(HexNAc), C22(Carbamidomethyl) | 4.392 | < 1% FDR | 3.232 | 22/46 | 93.528 | G |
| | | SGGLVYEYSnETNNYGLVQIDGDKVTK | 3 | 1056.846 | N10(HexNAc) | 4.918 | < 1% FDR | 3.684 | 26/104 | 99.529 | A |
| | | SKVSNPEGNGGYSTSNnY | 2 | 1040.458 | N17(HexNAc) | 3.683 | < 1% FDR | 3.134 | 21/34 | 49.628 | B |
| YOL154W | 3 | | | | | | | | | | |
| | | DTnSTAELQSPSSQEILGW | 2 | 1134.532 | N3(HexNAc) | 3.833 | < 1% FDR | 6.233 | 17/36 | 117.401 | D |
| | | EEVVDYTQNnATYAVR | 2 | 1038.989 | N10(HexNAc) | 4.019 | < 1% FDR | 4.012 | 18/30 | 70.129 | G |
| | | GWSHATEPTIYQTcnETNAR | 2 | 1280.081 | C14(Carbamidomethyl), N15(HexNAc) | 3.681 | < 1% FDR | 2.660 | 17/38 | 86.838 | G |
| YOR320C | 1 | | | | | | | | | | |
| | | TLDNDGNDIPVGLnDSVAYSK | 2 | 1206.575 | N14(HexNAc) | 4.255 | < 1% FDR | 3.923 | 16/40 | 89.555 | A |
| YPL123C | 1 | | | | | | | | | | |
| | | GIEPNcPINIPLScSnK | 2 | 1059.510 | C6(Carbamidomethyl), C14(Carbamidomethyl), N16(HexNAc) | 2.827 | < 1% FDR | 0.594 | 17/32 | 90.405 | G |
| YPL154C | 2 | | | | | | | | | | |
| | | AnGTEFAIQY | 2 | 659.812 | N2(HexNAc) | 3.036 | < 1% FDR | 6.884 | 13/18 | 89.810 | C |
| | | NGYnFTIGPY | 2 | 675.813 | N4(HexNAc) | 2.576 | < 5% FDR | 4.379 | 12/18 | 107.784 | D |

## B.3 Glycosite ambiguity in a peptide from Ygp1.

(A) The isotopic envelope corresponding to the singly-glycosylated Ygp1 peptide ANGTN-STTNTTTAESSQL is shown. Highlighted in teal is the peak selected from the inclusion list and the 4 Da IW used for fragmentation is shown in yellow. The peptide contains three potential sites for N-glycosylation. (B) The CID fragmentation spectrum can be plausibly assigned to either AnGTN-STTNTTTAESSQL or ANGTnSTTNTTTAESSQL, where the lowercase n indicates the glycosylated residue. Due to low signal:noise for fragment ions covering the peptide's N-terminus, definitive assignment could not be made even after close examination of fragment ion envelopes.