2020

Characterization of O-Linked Glycosylated Neuropeptides in the American Lobster (*Homarus americanus*): The Use of Peptide Labeling Following Beta Elimination

Edward Myron Bull
Bowdoin College

Characterization of O-Linked Glycosylated Neuropeptides in the American Lobster (*Homarus americanus*): The Use of Peptide Labeling Following Beta Elimination

An Honors Project for the Program of Biochemistry

By Edward Myron Bull

Bowdoin College, 2020

Table of Contents

Table of Figures

Table of Tables

Acknowledgements

I would like to first and foremost thank Professor Elizabeth Stemmler for her constant mentorship and support of this honors thesis. Although I had never studied analytical chemistry, Professor Stemmler warmly welcomed me into her lab and spent countless hours guiding me. Her patience as a teacher to a student completely unversed in analytical chemistry is inspiring to me as an aspiring teacher, and she has done even more to inspire me as a scientist. I have grown and learned much from my time in her lab, and for that I will always be grateful.

I would also like to thank my summer lab mates Ruby Ahaiwe 21', Carlos Campos 22', and Fiona O'Carroll 22' for their support and friendship during my initial period of learning. I would also like to thank the previous Stemmler lab members whom this honors project owes a considerable debt for their foundational work, particularly Henry Pratt '15 and Catherine Call '19. I also owe my thanks to Professor Patsy Dickinson for her teaching in lobster dissections this summer despite her juggling 13 students.

Finally, I owe everything to my parents for encouraging my curiosity and supporting me through my academic endeavors. They have made and continue to make sacrifices every day to give me the opportunity to receive a college education. They have walked an often thankless road, and in doing so, they have taught me everything I know about family.

Abstract

Neuropeptides are a class of small peptides that govern various neurological functions, and the American lobster (*Homarus americanus*) provides a model system for their characterization. Neuropeptides are commonly post-translationally modified (PTM), and one common PTM is glycosylation. Past research in the Stemmler lab has found glycosylated neuropeptides in *H. americanus*; however, the extent and biological role of this modification has not been well characterized. This study was undertaken to determine the number of glycosylated peptides in the sinus glands of *H. americanus* and to develop an approach to tag the site of glycosylation using beta-elimination chemistry. For the second goal, we focused on two crustacean hyperglycemic hormone (CHH) precursor related peptides (CPRP) as previously identified glycosylated neuropeptides. LC-MS paired with high pH reverse phase fractionation was used to survey for glycosylated neuropeptides and beta elimination with an amine tag was used as an approach to characterize the site of glycosylation. Our results indicate that high pH fractionation is a useful approach to simplify complex mixtures of neuropeptides and improve glycopeptide detection. Efforts to use beta elimination and tagging to characterize glycosylated neuropeptides have been less successful. Beta elimination of full length peptides resulted in peptide degradation. An approach utilizing chymotrypsin to reduce peptide size coupled with beta elimination and labeling with 2-dimethylaminoethanethiol showed less evidence for degradation, and this approach yielded data isolating two potential serine residues for the site of glycosylation; however, the data was not sufficient to distinguish the two sites. Work to optimize reaction conditions using a glycopeptide standard showed that multiple isomeric products were formed during beta elimination. With the goal of optimizing reaction conditions, future work will further examine reaction kinetics to eventually apply the approach to the entire sinus gland

Introduction

1.1 Neuropeptides are neuronal signaling molecules

Neuropeptides are polypeptides that act as chemical messengers in the brain and body, governing a vast array of physiological functions. Some common features of neuropeptides include cell surface receptor activation and target cell modulation.[1] Neuropeptides govern many neuronal functions, ranging from acting as neurotransmitters to forming neuronal networks.[2] Neuropeptides have also been linked to higher-order processes including metabolism, memory, reproduction, and even social behavior, making them multifaceted pillars of the nervous system.[3]

With regard to their biosynthesis, neuropeptides are derived from preprohormones, which are larger inactive precursor molecules. The conversion of the preprohormone to active neuropeptides occurs in the endoplasmic reticulum, where the preprohormone is transferred after it has been translated and packaged into vesicles in the Golgi apparatus. The processing consists of enzymatic cleavages, which reduce the size of the larger preprohormone into smaller immature neuropeptides. These cleaved immature neuropeptides can then be further processed to introduce post-translational modifications (PTM) that yield mature, bioactive neuropeptides (Figure 1). One of the potential products of these cleavages are precursor-related peptides (PRP), which serve roles that are less well established.[4] Many PRPs are thought to degrade after their cleavage from the parent peptide, but they may also serve a biological function.[5,6]

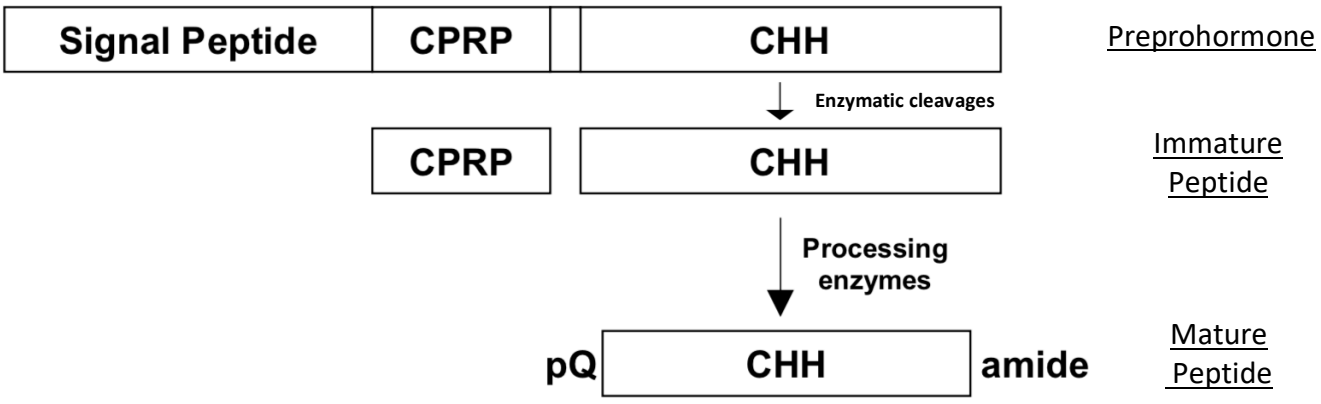


Figure 1: Processing pathway for the production of a mature CHH neuropeptide from the parent precursor peptide.
The cleaved CHH portion subject to amidation (amide) and pyroglutamation (pQ) during processing after cleavage. Adapted from Pratt, 2015.[7]

One common model system for the study of neuropeptides is crustaceans, which are studied for their segmented nervous system and relatively simple neuronal circuits. This simple nervous system can serve as a model for more complex nervous system functions. Therefore, isolating neuropeptides and testing their effects on previously characterized neural circuits could allow linking localized neuropeptide function to higher order processes.

One well-studied crustacean neuropeptide is the crustacean hyperglycemic hormone (CHH), which is a hormone that plays a variety of roles in crustaceans, including carbohydrate metabolism, molting, reproduction, and osmoregulation.[8] CHH is formed when the peptide is cleaved from the precursor and is modified at the C- and N-termini (Figure 1). CHH precursor-related peptides (CPRPs) are another component of the precursor protein that is not a part of the mature CHH.[8] CHH and CPRPs are found as abundant neuropeptides in the sinus glands of *Homarus americanus*. While the role of CHH peptide is well characterized, the same is not true for the CPRP portion.

Table 1: List of CPRPs that have been identified in *H. americanus* with their corresponding sequences and masses. Residues in red represent differences in the two sequences.

Peptide Name	Peptide Sequence	Peptide mass (Da)
CPRP-A	RSVEGASRMEKLLSSSNSPSSTPLGFLSQDHSVN	3603.7587
CPRP-B	RSVEGVSRMEKLLSS-ISPSSTPLGFLSQDHSVN	3543.799

1.2 Glycosylation is a common post-translational modification

Neuropeptides can be modified in the Golgi apparatus prior to further transport in order to regulate their biological fate, such as degradation or persistence for a potential bioactive function.[9] These modifications, known as post translational modifications (PTMs), include disulfide bond formation, amidation, and sulfation. One of the most common PTMs is amidation, wherein a C-terminal glycine is converted into an amide group.[10] This modification can protect the peptide from proteolytic degradation, and also governs proper receptor recognition and subsequent signal transduction.[11,12]

Another common protein PTM is glycosylation, which is a heterogeneous PTM wherein polysaccharides are attached to certain amino acid residues in the peptide. Glycosylation heterogeneity stems from multiple factors, such as the possibility of affecting eight potential amino acid residues. The diversity also stems from the potential incorporation of up to thirteen different monosaccharides, which can be connected in variable sequences that further multiply glycan heterogeneity.[13] There is also complexity added from glycan branching, stemming from the fact that a single glycan chain can branch into multiple chains, which adds additional layers for sequence complexity (See Figure 2).

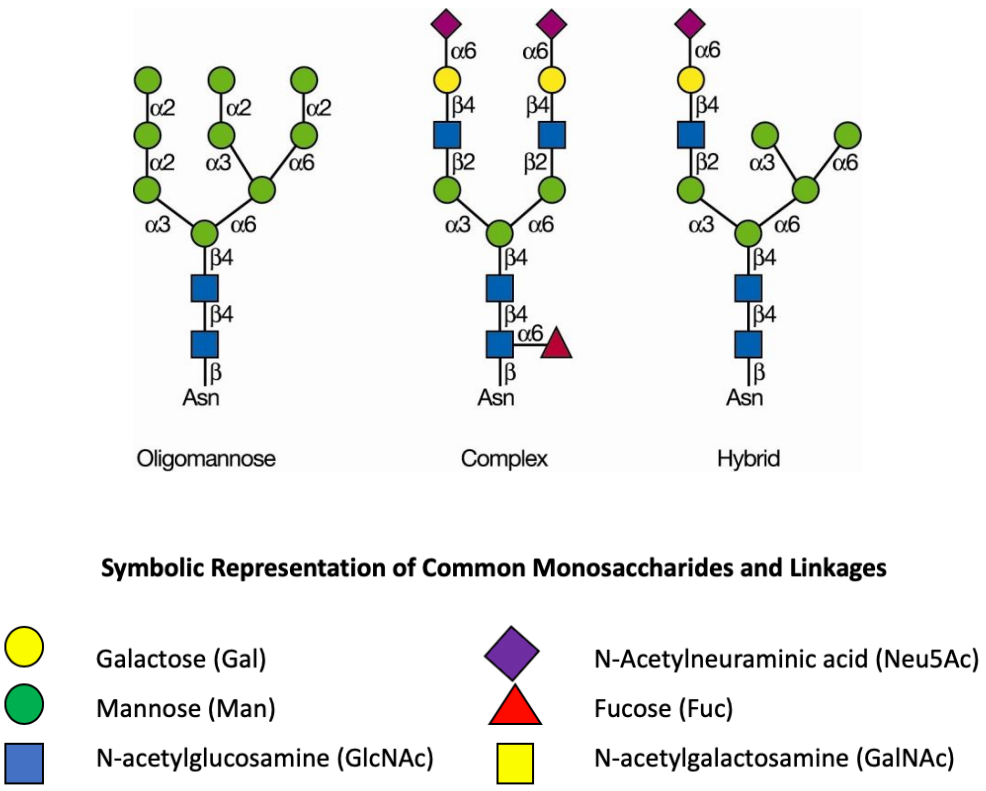


Figure 2: Glycans have diverse constituent sugars and complex branching.
Representative structures of three *N*-linked glycans demonstrating glycan sequence and branching heterogeneity. Key below shows identity of constituent sugars. Adapted from Brockhousen et al[14] and Stanley et al[15].

Glycosylation modifies the activity of the parent peptide, serving a multitude of biological roles. Glycosylation can mediate cell surface interactions through acting as ligands for cell adhesion, pathogen invasion, and macromolecular interactions.[16] In addition to cell surface recognition, glycosylation also plays a role in protein processing pathways, the modulation of protein folding, and lectins ligand sites.[17] This role in protein processing extends into their role in endoplasmic reticulum packaging, subsequent vesicle trafficking, and secretion.[18] Beyond protein processing and intercellular interactions, glycosylation governs many other diverse processes, such as identifying tissue maturation levels.[19] While many functions have been identified for glycosylation, the specific relationship between glycan structures and their

functions is not yet well understood. Identifying the effect of glycan structure on function is a principle goal of glycochemistry.

Glycosylation can be divided into two common types: O-linked, wherein the glycan is attached to serine (S) or threonine (T) residues, or N-linked, where the glycan is attached to asparagine (N) residues (see Figure 3). In O-linked glycans, the first monosaccharide is often either a *N*-acetylgalactosamine (GalNAc) or a *N*-acetylglucosamine (GlcNAc); (Figure 3B and C). If the initial monosaccharide is a GalNAc, then it is likely not branched beyond the initial sugar; in contrast, if the initial sugar is a GlcNAc, the glycan is often further substituted.[14] N-linked glycans, however, commonly consist of a similar five monosaccharide backbone and more diverse branching, causing many to be more heterogeneous.[15] Neuropeptide glycosylation is not well explored, and this study seeks to examine this form of PTM.

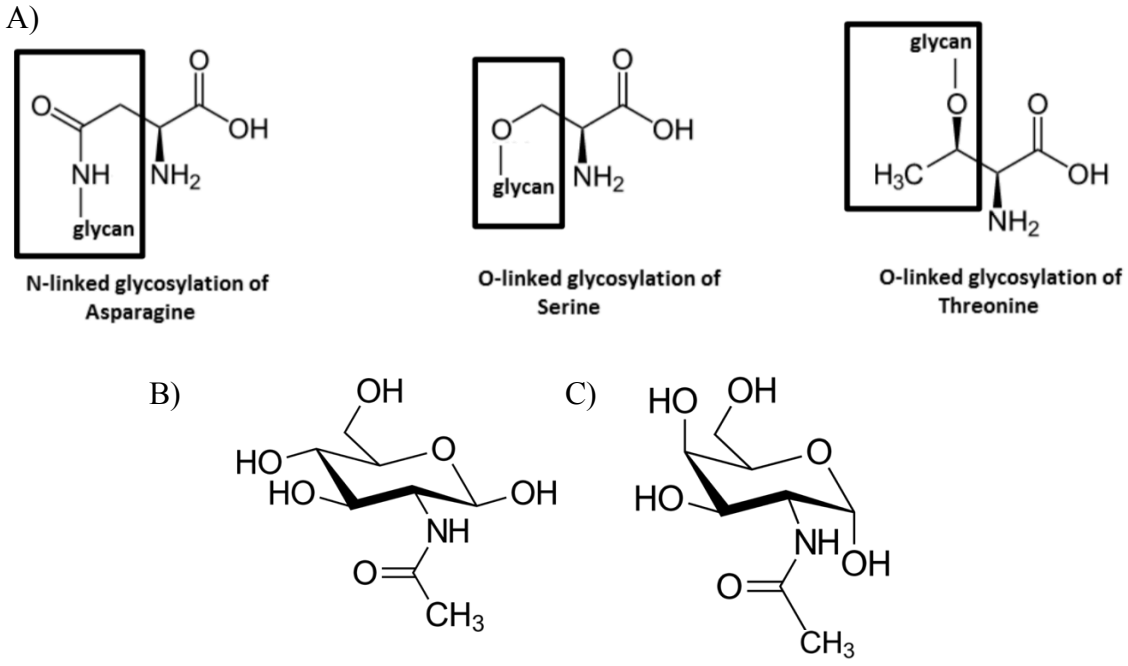


Figure 3: Glycosylation can be either N- or O-linked, and these glycan often begin with one of two HexNAc sugars.
Diagram of A) N- and O-linked glycosylated amino acid residues and common O-linked monosaccharides, B) GlcNAc and C) GalNAc.

The glycosylation of crustacean neuropeptides has not been well explored. In a recent study, the Li group reported on the identification of a range of glycosylated neuropeptides from the blue crab *Callinectes sapidus* using advanced MS/MS techniques.[20] In a previous study from the Stemmler lab, Henry Pratt '15 determined that in *H. americanus*, CPRP A and B were glycosylated with HexNAc and Hex-HexNAc and showed that glycosylation was localized to the internal serine-rich region.[7] For *H. americanus*, Pratt's work provides the only example of neuropeptide glycosylation from this species. Expanding the number of characterized glycopeptides from *H. americanus* is the goal of this study.

1.3 Mass Spectrometry is a widely-used Method for Peptide Characterization

Mass spectrometry (MS) is the most common analytical approach used to characterize peptides and glycopeptides. This technique allows determination of the specific amino acid sequence for the peptide, as well as PTMs such as glycosylation. Therefore, MS is an ideal approach to use for characterizing glycosylated neuropeptides.

While many varieties of mass spectrometers exist, one approach that is particularly useful for examining biological mixtures is liquid chromatography-MS (LC-MS). This type of instrument employs liquid chromatography prior to mass analysis, which allows separation of mixtures through differential binding to a given stationary phase. This is particularly useful in the study of crustacean systems, where the peptides are drawn from tissues that contain a multitude of peptides to distinguish. Analyses yield both a chromatogram illustrating the elution of specific peptides and the accompanying MS and MS/MS level mass spectra. These spectra are used to determine the peptide mass and amino acid sequence, respectively, using measurements of the mass-to-charge ratio (m/z).

The instrument utilized in this study combined chip-based liquid chromatography and nanoelectrospray ionization (nanoESI)[21] with a Q-TOF mass analyzer. The first step of the sample analysis involves the sample peptides being separated by an LC column to generate a chromatogram (Figure 4A). The mobile phase flowing through the column contains an increasing concentration of acetonitrile (ACN) in acidic water over time, resulting in more hydrophobic peptides eluting later in the run. The sample is then moved into the nanoESI tip where it is ionized,[18] resulting in the production of gas-phase peptides with one or more added protons (charge states), governed by number of available protonation sites. This ionization is what facilitates mass analysis to identify peptides.

After ionization, the ionized peptides are then moved into the quadrupole-time-of-flight (Q-TOF) mass analyzer. Here, the peptides are subjected to either MS or MS/MS analysis. For both analyses, mass spectra are measured using the time it takes for each ion to reach the detector when accelerated to the same kinetic energy. The time is unique to each mass-to-charge ratio (m/z). In addition to measuring m/z, the intensity is also measured to determine the number of ions generated. The MS-level analysis is used to determine the peptide mass (Figure 4B).

After this first step of mass analysis, another dimension of MS can be utilized, known as tandem mass spectrometry (MS/MS). This form of MS includes a second level of analysis wherein a selected ion is fragmented using a collision to produce product ions that are used for peptide sequencing (Figure 4C). From the MS/MS spectrum, the difference between fragment ion masses can be utilized to determine the amino acid sequence of the peptide. As for how this fragmentation is utilized to sequence peptides, this proceeds through the analysis of specific types of product ions. B-type ions are those that contain the N-terminus of the peptide and y-type ions contain the C-terminus. These ions types have masses specific to each amino acid, so

through the enumeration of variable ions, the original peptide sequence can be determined.[22] Therefore, MS/MS is a robust tool for sequencing unknown peptides or confirming the identity of previously studied samples.

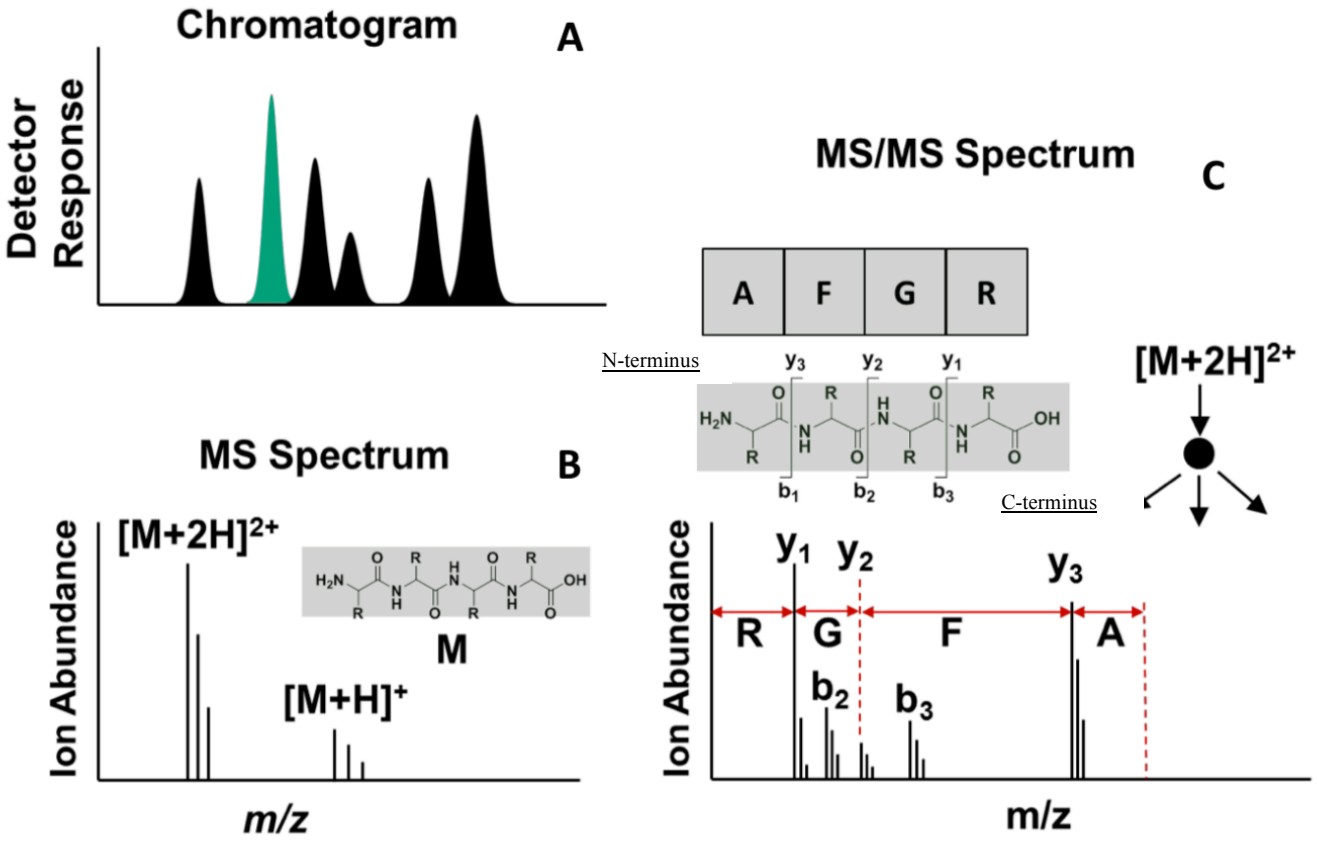


Figure 4:MS workflows proceed from chromatographic separation to mass analysis. Sample workflow showing data generated in LCMS analysis. A) The chromatogram generated showing the intensity of eluted peptides as a function of time. B) The mass spectrum shows the *m/z* of ions formed alongside their individual abundance. C) MS/MS spectrum allows visualization of how peptides of interest fragment. The differences in masses between peaks allows reconstruction of peptide identity through unique amino acid mass values.

Chip-based LC-MS is a powerful technique for separating complex mixtures. Human blood serum is a complex biological mixture that necessitates highly efficient separation, and use

of chip-based LC-MS allowed a 23% increase in signal identification as opposed to conventional cap-LC-MS systems.[23] This method involved using trypsin proteolytic digestion prior to LC-MS analysis, and showed that this approach to require 30 times less sample (~5 pmol) and the ability to discriminate the highly similar proteins present in the complex mixture of human blood serum. Therefore, chip based systems are advantageous in their lower sample requirement (important for low concentration neuropeptides), higher chromatographic resolution, and higher discrimination between similar peptides.

1.4 Glycopeptides are challenging to characterize using conventional MS approaches

Glycosylation is a heterogeneous PTM due to multiple amino acid residue attachment sites, multiple monosaccharides, and variable branching points. While glycan complexity makes mass spectrometric characterization of glycosylation difficult, it is further compounded by other issues. For glycopeptides, the glycosidic bond is extremely labile, leading to glycan loss from the ionized peptide at low energies, making determination of the position from MS/MS data challenging. It is possible to search MS/MS data for peaks yielding a mass corresponding a known peptide mass with an added glycan mass. From the sequence ions of the peptide+glycan mass, the position of glycosylation can be inferred. However, this information is not found in actual MS/MS spectra due to the previously mentioned issue of glycan loss at low energies. This is significant for studying peptide glycosylation because this means that traditional MS strategies are insufficient for determining the site of this PTM. This approach does, however, provide information that reveals that the peptide is glycosylated by the detection of the characteristic sugar ions.

As for which peaks are indicative of glycans, the most common peaks originate from the first sugar added to most O-linked glycopeptides: GalNAc or GlcNAc. Given that the two sugars are

stereoisomers, they produce the same oxonium ion fragment with *m/z*= 204.08 (Figure 5). Glycopeptides with an additional hexose sugar (mannose or glucose) yield a Hex-HexNAc oxonium ion at *m/z*= 366.140. The detection of *m/z*= 204 and *m/z* 366 provide evidence that a peptide is glycosylated with one or more glycans. The HexNAc oxonium ion can also undergo further fragmentation through losses of groups such as water and CH_2CO to form multiple product ions (Figure 5). Detection of any of these ions- which often dominate the MS/MS spectrum (see Figure 6B)- is a strong indicator of peptide glycosylation.

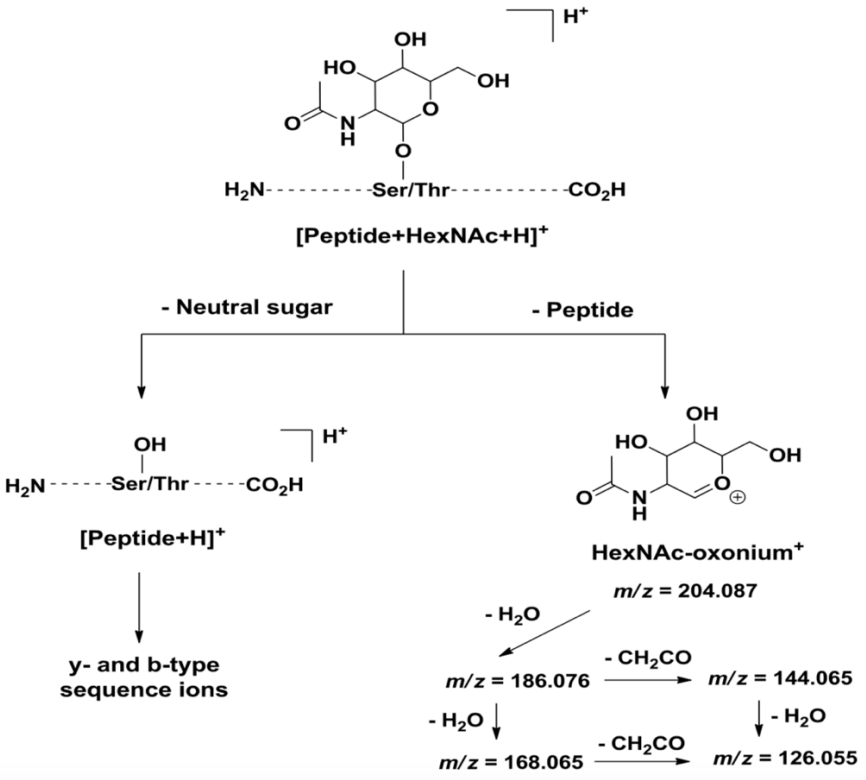


Figure 5: Glycosylated peptides fragment to form characteristic product ions.
Diagram detailing common fragmentation pathways for a O-linked glycosylated peptide with accompanying masses. The left pathway shows the loss of a neutral sugar, which results in the ionized form of the native peptide with an unmodified serine/ threonine, which fragments to form sequence ions. The right path demonstrates the formation and fragmentation of the HexNAc, which can undergo subsequent losses of water and CH_2CO to form multiple oxonium ions. Figure adapted from Henry Pratt '15.[7]

1.5 Beta elimination provides an alternative approach to characterize glycopeptides

While the detection of characteristic oxonium ions in MS/MS spectra demonstrates that the peptide is glycosylated, the lingering questions of determining the peptide sequence and of how to determine the site of glycosylation remains. Because the MS/MS spectra of glycosylated peptides are usually dominated by oxonium ions rather than peptide ion peaks (Figure 6), the peptide sequencing is challenging because the facile loss of the glycan and lower intensity sequence ions. The lack of sequence ion signal means there is little evidence of glycan localization on sequence ions, making it difficult to determine the site of glycan attachment.

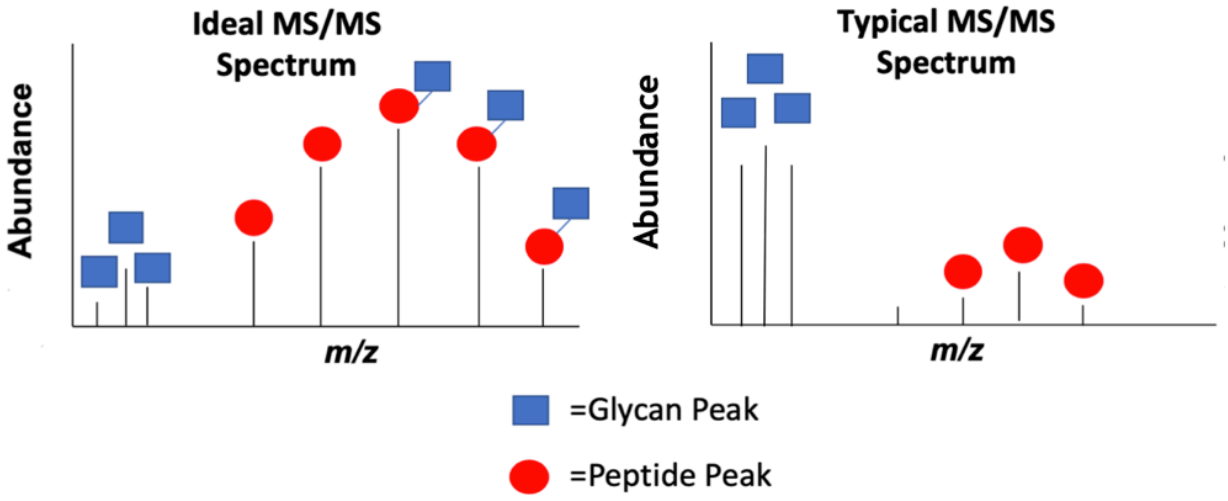


Figure 6: Typical MS/MS spectra for a glycopeptide contains little sequence information due to oxonium ion dominance.
Schematic showing sample ideal and typical MS/MS spectra of glycopeptides. Typical spectra have low signal sequence ions due to glycan domination, while the ideal would have higher sequence signals that contain the glycan to allow site determination. Normally, glycans fragment very early due to their bond lability, causing no sequence ions to contain evidence of glycan attachment.

An alternative approach to determine the peptide sequence and site of glycosylation involves utilizing chemical methods of glycan elimination. While enzymatic methods are effective for removing *N*-linked glycans, chemical methods are employed for *O*-linked glycopeptides.[15] The chemical reaction involves using a base to beta eliminate the glycan, resulting in formation of a

dehydroalanine for *O*-linked serine residues, or methyldehydroalanine for *O*-linked threonine residues (Figure 7). Different chemicals can be used as bases, with the most common being sodium hydroxide.[24] While some methods have relied upon the detection of the beta-eliminated peptides, other approaches have paired the elimination with nucleophilic addition at the newly created dehydroalanine/ methyldehydroalanine.[24] These nucleophilic additions can provide a unique mass change as a tag that can be used to search MS data to detect former glycopeptides.

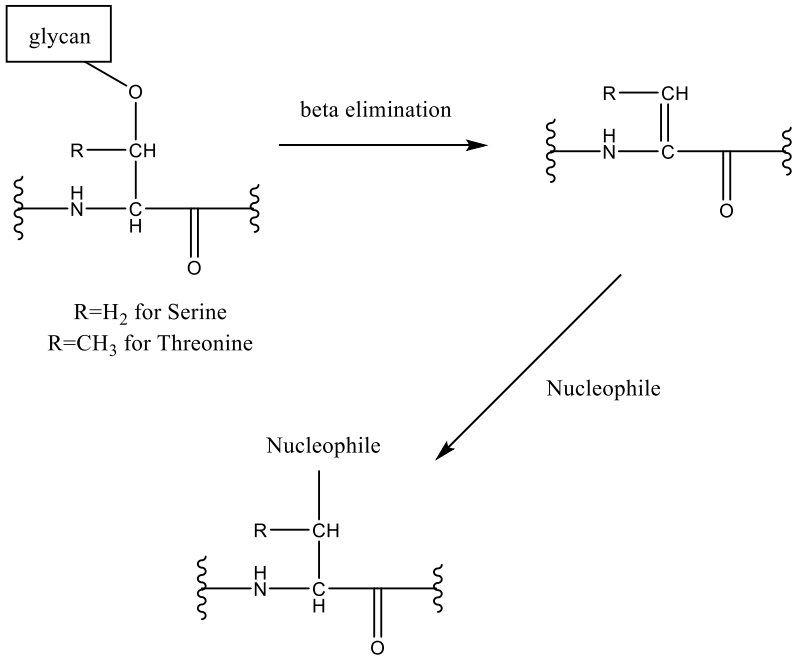
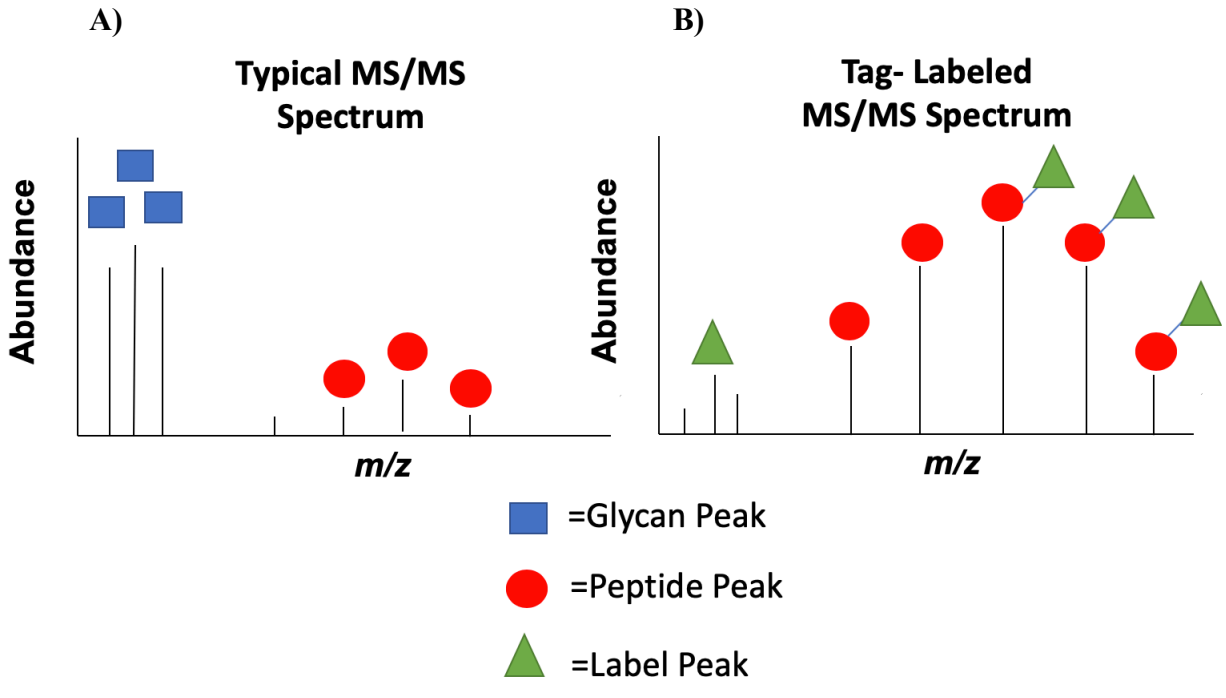


Figure 7: Beta elimination allows site specific tagging
Example of beta elimination reaction mechanism for a serine (R=H) or threonine (R=CH_3) and a generic base. The second step involves addition of a nucleophile at the site of the newly created dehydro residue.

Many nucleophiles have been used to react with the dehydro residue for glycan site labeling. One nucleophile is ethylamine, which is unique in that it simultaneously acts as a gentle base to remove the glycan and as a nucleophile to label the site of glycosylation for subsequent analysis (Figure 7).[25] Another approach pairs a combination of beta-elimination and thiol-addition derivatization with a thiol nucleophile that leaves a unique ion to isolate via precursor ion

scanning (Figure 8). This approach was originally used to label phosphopeptides, so we decided to employ it in hopes of finding a robust method for glycopeptide characterization.[26] This approach contains three strengths over ethylamine: first, the addition of a strongly basic thiol nucleophile should improve ionization, allowing greater sensitivity. Second, the addition should produce a highly abundant, unique fragment cation that can be used as a marker for formerly glycosylated peptides (Figure 8B, green triangles). This cation should be observed when added to glycosylated peptides and can be searched for in MS data, allowing rapid detection of glycopeptides. Third, the new tag contains a less labile bond that doesn't fragment readily, allowing localization of the glycan by isolating the site of tag attachment (Figure 8, green triangles attached to red sequence ions). In addition to these three strengths, the tag ion retention means that glycopeptides can simultaneously be detected and sequenced, requiring no *a priori* knowledge of peptide sequences.



C)

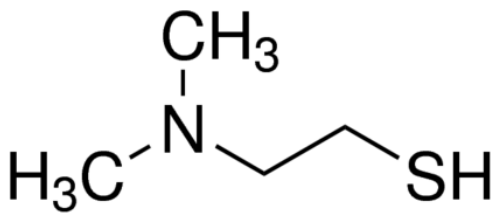


Figure 8: Tag-labeled glycopeptides yield improved sensitivity, show higher sequence ion signal, and permit glycan localization.
A) Typical MS/MS spectra yield high glycan ion signal, causing reduction in sequence ion signal. Furthermore, glycan ion formation occurs at low energies, disallowing glycan localization. B) The ideal, tag-labeled MS/MS spectrum uses stable tags to localize the glycan and avoid sequence ion loss. The structures of a thiol nucleophile, C) 2-(dimethylamino) ethanethiol.

1.6 Goals

Given that the study of crustacean neuropeptide glycosylation is still in its infancy, there is a need for a robust method for characterizing glycopeptides in *H. americanus*.[20,27] In fact, the first comprehensive report on neuropeptide glycosylation in crustaceans was published in 2020.[20] One important obstacle for characterizing neuropeptides in *H. americanus* is the lack of a complete neuropeptide database. Furthermore, glycosylated neuropeptides exist in low concentrations as a part of complicated biological mixtures, so this study seeks to find a way to efficiently separate glycosylated neuropeptides from other peptides in an effort to expand the glyconeuropeptidome. Among other challenges, the glycan moiety often dominates MS/MS spectra, causing the loss of peptide sequence ion signal, inhibiting glycopeptide sequencing. The glycosidic bond is labile, fragmenting at low energies, preventing localization of the amino acid residue site of glycan

attachment. Therefore, this study also seeks to utilize beta elimination strategies to isolate an effective approach for identifying glycan identity, sequencing novel peptides, and isolating the site of glycan attachment. This approach will be validated in glycopeptide standards to determine the ideal base, nucleophile, and reaction conditions for glycopeptide characterization. This approach will then be applied to the complex biological mixture of the American lobster sinus glands to profile the glycome. Integrating the separation of neuropeptides, isolation of glycopeptides, and characterization of neuropeptide glycosylation will allow the creation of a crustacean neuropeptide library for future studies.

1.7 Overview of Approach

The general approach that will be employed for this study begins with fine dissection of lobster sinus glands, followed by extraction to remove neuropeptides from the tissue (Figure 9).

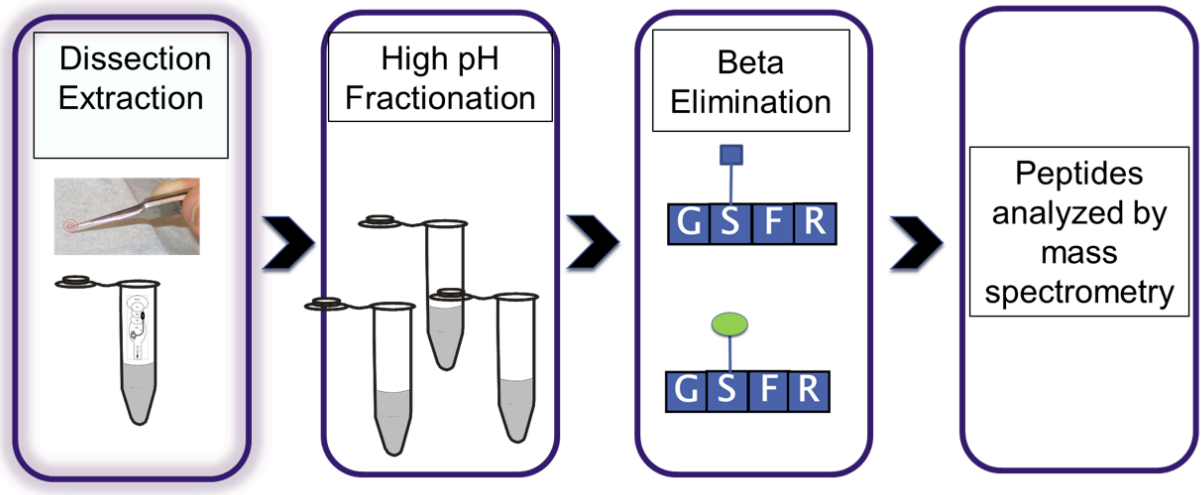


Figure 9: Overview of approach for LC-MS analysis of lobster neuropeptides.

Given the sample complexity of lobster sinus glands, it is also important to develop additional measures outside of normal chromatography to simplify the sample prior to analysis. Catherine Call '19[28] found the use of high pH reverse phase (RP) fractionation as a technique orthogonal to traditional liquid chromatography increased neuropeptide identification 9.6 fold. The term orthogonal refers to separation which separates the sample by characteristics different from the low pH chromatography used in the LC-MS. Therefore, this first dimension of separation of high pH RP fractionation prior to the second dimension separation of low pH allows more sample simplification. This procedure proceeds through use of a C18 stationary phase, which is then eluted with an increasing gradient of organic solvent (Figure 10). This technique, when used prior to beta elimination, should aid in reducing instances of sample co-elution, wherein two similar peptides elute at the same time, causing their signals to overlap, reducing analytical sensitivity. Furthermore, having fewer peptides in a given sample will also reduce the number of peptides being assayed, further simplifying analysis.

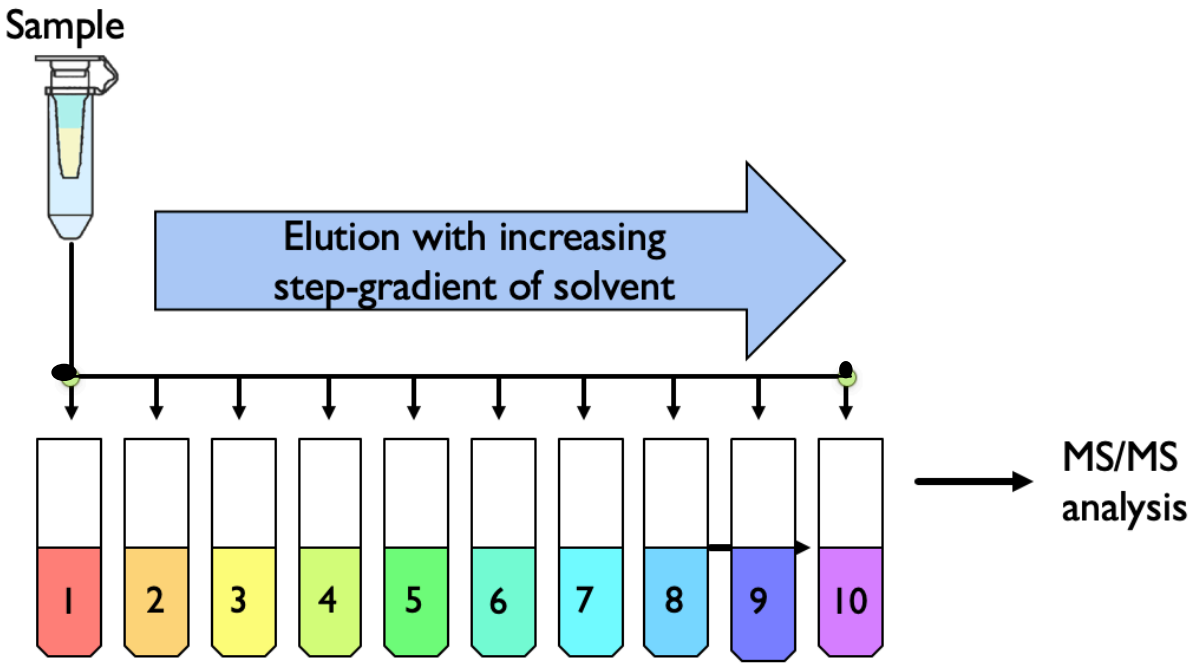


Figure 10: High pH reverse phase fractionation can be used to simplify a sample prior to MS/MS analysis.

The fractionated peptides are then subjected to beta-elimination and nucleophilic attack to label the site of glycosylation. Past strategies by Pratt[7] for peptide labeling involved conditions with unoptimized nucleophilic labeling for glycosylation site determination. This fractionation step helps to simplify the sinus gland mixtures prior to beta elimination and nucleophilic tagging. Through the integration of these techniques, this study seeks to develop a method for O-linked glycosylation characterization of neuropeptides taken from the complex mixture of peptides in *H. americanus* sinus glands. This study seeks to optimize the reaction and explore different nucleophiles to identify the strongest label.

Materials and Methods

2.1 Sinus Gland Dissection and Extraction

American lobsters (*H. americanus*) were purchased from local fisheries (Harpswell, ME), and stored at 8-10°C in aerated seawater tanks. Prior to dissection, the lobsters were anesthetized in ice for a duration of 30 min. The eyestalks were removed via manual dissection in 8-10°C saline solution, then the sinus glands were removed from the eyestalks and placed in a 1.5 mL microcentrifuge tube in 30 µL water (LCMS grade; Fisher Scientific). The tubes were heated to 100°C for 5 min to deactivate proteolytic enzymes, then stored in 4° C prior to further analysis.

In order to extract peptides from the tissue, extraction solvent A (142 μL; 90% MeOH; Sigma Aldrich, 9.9% CH_3CO_2H; Sigma Aldrich) was added to the sinus glands in water. The tissue and solvent were then homogenized with a motorized pestle (Argos Technologies; Vernon Hills, IL) for 3 min. Samples were then sonicated for 5 min, then centrifuged for 1 min at 14.5 krpm. The supernatant was transferred to a clean tube; then extraction solvent B (30 μL; 65% MeOH, 5% CH_3CO_2H, 30% H_2O) was added. The homogenization, centrifugation, and sonication steps were repeated. The extracts were combined and dried with a UVS400 universal vacuum system (Thermo) and stored at 4°C prior to further analysis.

2.2 High pH RP Fractionation

Fractionation was conducted through spin columns (PierceHigh pH Reversed-Phase Peptide Fractionation Kit; Thermo Scientific; Waltham, Massachusetts), which were centrifuged at 2900 x g for 2 min. The column was then loaded with 300 μL acetonitrile (ACN) and centrifuged at 2900 x g for 2 min. The ACN wash was repeated, and was followed by two washes with 0.1% trifluoroacetic acid solution, all with centrifugation at 5000 x g for 2 min.

Sinus gland extracts were dissolved in 300 μL 0.1% TFA and sonicated for 5 min. The spin column was transferred to a new tube and the dissolved sinus gland sample was added to the resin, which was then centrifuged at 2900 x g for 2 min. After the eluent was discarded, the spin column was transferred to a new tube and 300 μL water was added, followed by centrifugation at 2900 x g for 2 min. The flow through was discarded, and the spin column was transferred to a new tube. Then a series of 300 μL elution solutions were added, which consisted of ACN in water ranging from 5% to 80% with 0.1% triethylamine (TEA; Sigma Aldrich, see Table 2), followed by centrifugation at 2900 x g for 2 min to collect the fractions. This procedure was repeated to generate a gradient of ACN elution fractions. All fractions were dried and stored at 4˚C.

Table 2: List of ACN gradient utilized for high pH RP fractionation.

Fraction Number	Percent ACN	Percent Water
1	5	95
2	7.5	92.5
3	10	90
4	12.5	87.5
5	15	85
6	17.5	82.5
7	20	80
8	50	50
9	70	30
10	80	20

2.3 Proteolytic Digestion

Four sinus glands were pooled and extracted, then reconstituted in 20 μL of water. Bovine chymotrypsin (0.1 ug; Promega) was dissolved in digestion buffer of 5 mM ammonium bicarbonate (>99.5%; Sigma Aldrich), then added to the sinus gland extracts. The sample was sonicated for 5 min, heated to 45° C for 12 h, and then cleaned with C18 SPE columns (see section 2.5). The samples were then dried and stored at 4°C until further analysis.

2.4 Beta Elimination and Nucleophilic Addition

Beta elimination reactions proceeded through reaction with either sodium hydroxide (KOH) or ethylamine. KOH (0.1 M KOH; Fisher Scientific) or ethylamine (70%; Fisher Scientific) was added to dried sinus gland extracts to a volume of 40 μL. The mixture was sonicated for 2 min, heated at 55°C for 4 h, then neutralized with 1.0 μL acetic acid (>99.7%; Fisher Scientific). Samples were then stored at 4°C prior to analysis.

Alternatively, for reactions pairing beta elimination with thiol addition, 5 μL of 0.1 M NaOH was added to a mixture of chymotrypsinized sinus gland and 0.54 M 2-dimethylamino-ethanethiol in 17.6 μL of 17% ethanol(200 proof; Sigma Aldrich) in LCMS water. To this solution, 2-dimethylamino-ethanethiol hydrochloride (DMAET, 95%; Sigma Aldrich) was added until the concentration reached 0.54 M. This solution was then heated to 56°C for 1 h and then neutralized with acetic acid. Both reaction conditions were then vacuum dried and stored at 4°C prior to further analysis.

Glycopeptide standards were also purchased for examination of the beta elimination reaction conditions (SVES-(HexNAc)GSADAK, Sussex Research, Ottawa, ON CA). Each reaction was prepared by diluting the standard to 10 μL aliquots of 10^{-4} M concentration. Then, the beta elimination/ thiol addition approach was applied.

2.5 C18 Cleanup

Some samples, such as sinus gland extracts and chymotryptic digests, were desalted using C18 solid phase extraction (SPE) columns. The C18 column (Pierce Peptide Desalting Spin Column; Thermo Fisher) was conditioned in 200 μL of 50% ACN by centrifugation at 5,000 x g for 1 min twice. The column was then further conditioned with centrifugation at 5,000 x g for 1 min in 200 μL of equilibration solvent twice (1% ACN, 0.5% TFA in LCMS water). The sample was then

added to the column with centrifugation at 3,000 x g twice (sample dissolved in 30 μL of 10% ACN and 0.5% TFA). The column was then washed with 200 μL of 5% ACN at 3,000 x g for 1 min twice, then the sample was eluted centrifugations at 3,000 x g for 1 min of 20 μL of 80% ACN. The eluted sample was then vacuum dried and stored at 4°C prior to further analysis.

2.6 Sample Preparation and LCMS Analysis

Samples were reconstituted from vacuum dried samples stored at 4°C in 40 µl of 0.1% TFA in LCMS water, then kept chilled prior to LCMS analysis. Samples were separated via an Agilent Technologies 1260 chip-cube system, with a stationary phase of C18 (300 Å consisting of 5 µm particles). This system utilized 2 ProtID-chips, which contained a 40 μL enrichment column to concentrate sample, and an analytical column with dimensions of 150 mm x 75 μm. All mass spectra were collected in positive ion mode in which the ion source temperature was set to 350°C and ionization voltage from 1850-1975 V. Spectra were calibrated using the $[M+H]^+$ ions from dibutyl phthalate ($C_{12}H_{22}O_4$) and hexakis(1H, 1H, 4H-hexafluorobutyloxy)phosphazine (HP-1221; $C_{24}H_{18}O_6N_3P3F_{36}$). Nitrogen was the target gas for collision induced dissociation. The equation $CE = [(3 \times m/z)/100]+2$ was used to calculate collision energy (CE) for the dissociation of precursor ions targeted for MS/MS. The mobile phase consisted of an increasing gradient of ACN (2% LCMS water, 0.1% formic acid) in LCMS water (0.1% formic acid) from 2 to 98% over the run duration. The flow rate was held constant at 300 μL/ min, and sample injection volume ranged from 0.1 μL to 10 μL. After samples were mass analyzed, they were stored again at 4°C.

2.7 Mass Spectral Data Analysis

Mass hunter software (Agilent) was utilized to display chromatographic and spectral data prior to analysis. SpectrumMill software (Agilent, version 5.0) was utilized to search spectra

against an in-house crustacean database. This software was used to target specific modifications, such as serine residues being modified with ethylamine. The variable modifications were encoded specific to each residue, *e.g.* ethylamine nucleophilic addition would be denoted by a mass increase of 27 Da over normal serine.

Results and Discussion

3.1 Detection of Glycosylated Neuropeptides in Sinus Glands

3.1.1 Sinus glands are a rich source of neuropeptides

In this study, we focused on analysis of sinus glands from *H. americanus*, which serve as rich source of neuropeptides. These glands have been shown by Pratt to contain glycosylated CPRPs[7], so we began by dissecting a sinus gland to probe for the presence of these neuropeptides. After extracting the peptides, we analyzed the sample by LC-MS to collect MS and MS/MS spectra. The MS/MS spectra were collected only for the most abundant ions. The chip-based liquid chromatographic separation produces a total-ionization chromatogram (TIC) showing elution of all compounds ionized by nanoESI (Figure 11). The peptides elute at variable times due to their differing affinity for the hydrophobic C18 stationary phase, allowing separation of the complex biological mixture. As is evident from the TIC (Figure 11), the sinus gland extracts contain a multitude of peptides, which presents a challenge for identifying glycopeptides.

In order to detect CPRPs and their glycosylated forms, we calculated the *m/z* values expected for the +5 charge states of the CPRPs (Table 3). Extracted-ion chromatograms (EIC), are utilized to select specific m/z values of interest at either the MS or MS/MS level to search for specific peptides or fragment ions. For example, CPRP-B has a mass of 3543.79 Da, which

should produce $[M+5H]^{5+}$ at m/z= 709.76 when ionized by nanoESI. To detect when the neuropeptide elutes, we generated an EIC specific to that m/z value.

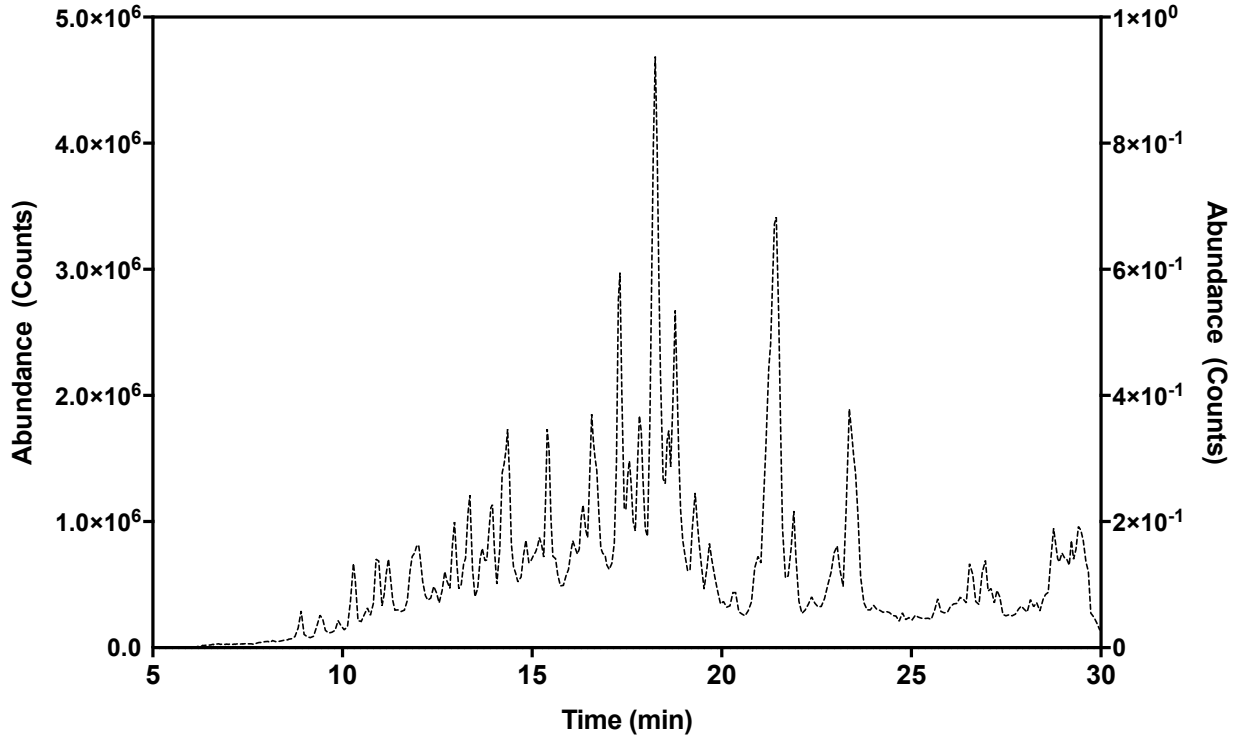


Figure 11: Sinus glands are complex biological mixtures.
TIC for untreated extracted sinus gland from American lobster.

3.1.2 Glycosylated CPRPs are detected in unfractionated sinus glands

In order to detect the presence of glycosylated neuropeptides, we first searched for glycosylated CPRPs in sinus gland extracts. Searching for glycosylated CPRP-A, we would add the mass of HexNAc (203.079 Da) to CPRP-A (3603.759 Da) to get the predicted mass of 3806.838 Da. We would then calculate the predicted mass of the most abundant charge state of 5+ to get m/z 762.375, which we could then search for with an EIC. This approach can be seen applied in Figure 12 to show the presence of multiple known CPRPs. However, this approach is limited to searching for previously characterized peptides. This method of identification of a glycopeptide relies on adding the mass of a glycan to a known peptide, so it cannot be utilized to

find new glycopeptides. The masses and most abundant charge states of the two characterized CPRPs and their glycosylated counterparts are summarized in Table 3. These values were utilized to search MS data for the presence of glycosylated CPRPs.

Table 3: List of CPRPs that have been sequenced in *H. americanus* with their corresponding sequences, masses, and most abundant charged states. Red residues represent differences in the two sequences.

Peptide Name	Peptide Sequence	Peptide mass (Da)	Most Abundant Charge State
CPRP-A	RSVEGASRMEKLLSSSNSPSSTPLGFLSQDHSVN	3603.7587	$[M+5H]^{5+}$ =721.759
CPRP-B	RSVEGVSRMEKLLSS-ISPSSTPLGFLSQDHSVN	3543.7990	$[M+5H]^{5+}$ =709.767
CPRP-A +HexNAc	RSVEGASRMEKLLSSSNSPSSTPLGFLSQDHSVN +HexNAc (unknown residue attachment site)	3806.838	$M+5H]^{5+}$ =762.375
CPRP-B +HexNAc	RSVEGVSRMEKLLSS-ISPSSTPLGFLSQDHSVN +HexNAc (unknown residue attachment site)	3746.869	$M+5H]^{5+}$ =750.383

The EIC approach discussed above confirmed the presence of the glycosylated form of both CPRPs. The blue lines show EICs for the native peptides, the red show EICs searching for the peptides+HexNAc, and green shows peptides+HexNAc-Hex (Figure 12). This demonstrates the presence of both glycosylated and unmodified peptides in the American lobster sinus glands.

The glycosylated forms of the CPRPs elute around 2 min earlier than their non-modified counterparts, which can be attributed to the nature of the glycan moiety. HexNAc is a highly polar group, which leads to the CPRPs binding less tightly to the hydrophobic stationary phase of the LC-chip, thus eluting earlier (Figure 12B and C). Additionally, both HexNAc and HexNAc-

Hex glycosylated peptides are significantly less abundant than their unmodified counterparts (Figure 12B and C). CPRP-B+HexNAc is more abundant than CPRP-A+HexNAc. Alternatively, CPRP-A+HexNAc-Hex is more abundant than CPRP-B+HexNAc-Hex. Thus, the relative amount of different forms of glycosylation (HexNAc vs HexNAc-Hex) varies between peptides, with the more simple HexNAc glycan being more common than the more substituted HexNAc-Hex (Figure 12B and C).

It is also worth noting that the CPRP peaks constitute only six of the peaks in the TIC, representing a small fraction of the total pool of eluted peptides. The remaining peaks represent other sinus gland components, such as orcokinins and other neuropeptides (Figure 12A). The multitude of peaks in the chromatogram demonstrates the considerable sample complexity of an entire sinus gland.

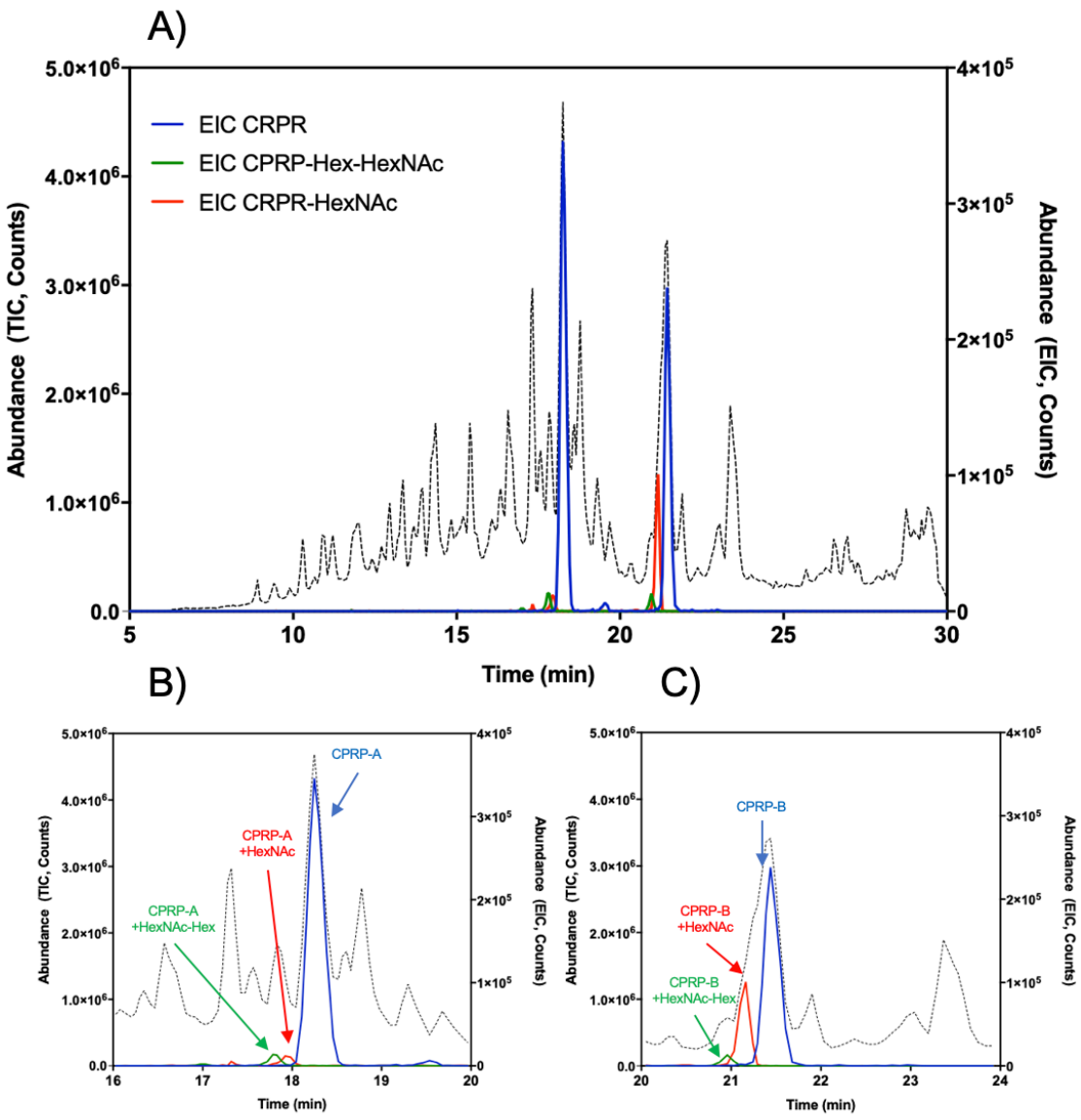


Figure 12: Sinus glands contain many peptides, including glycosylated CPRPs.
A) TIC from *H. americanus* sinus gland extract with EICs overlaid for CPRPs. **B)** EICs demonstrative of CPRP-A in blue (*m/z* 721.759; $[M+5H]^{5+}$ for CPRP-A), CPRP-A+HexNAc in red (*m/z* 762.375; $[M+5H]^{5+}$ for CPRP-A+HexNAc), and CPRP-A+HexNAc-Hex in green (*m/z* 794.785; [M+5H]5+ for CPRP-A+HexNAc-Hex). **C)** B) EICs demonstrative of CPRP-B in blue (*m/z* 709.767; $[M+5H]^{5+}$ for CPRP-B), CPRP-B+HexNAc in red (*m/z* 750.383; $[M+5H]^{5+}$ for CPRP-B+HexNAc), and CPRP-B+HexNAc-Hex in green (*m/z* 782.794; [M+5H]5+ for CPRP-B+HexNAc-Hex).

The EIC approach used above is useful for showing peptide presence, but the presence of predicted masses alone is not sufficient to confirm peptide identity. For confirmation, we examined the MS/MS spectra. For example, for CPRP-B+HexNAc (Figure 13A), the mass

spectrum showed the dominant charged states to be $[M+5H]^{5+}$ and $[M+4H]^{4+}$ ions, which, when adjusted to reflect the uncharged mass, yielded a value of 3746.841 Da, 1.0 ppm from the expected value. Next, we used the MS/MS spectrum to support the peptide sequence (Figure 13B). The first thing of note in the MS/MS data is the presence of oxonium ions at m/z= 126.06, 186.08 and 204.09 indicative of glycosylation (Figure 13B). Furthermore, the MS/MS spectrum contains three y-type ions that match the predicted sequence of CPRP-B (y_{12}, y_{10}, and y_7 ions; Figure 13B). These ions contain the first 12 residues from the C-terminus with high similarity, supporting the sequence match. However, the later series y-type ions and b-type ions are largely absent.

This signal loss is a characteristic of glycopeptides, as the presence of the oxonium ions dominates the MS/MS spectrum. This results in fewer peptide sequence ions, which can make sequencing the full peptide challenging. Therefore, while the oxonium ions do provide a useful means to detect glycopeptides, their presence reduces the number of sequence ions, inhibiting sequencing of novel peptides.

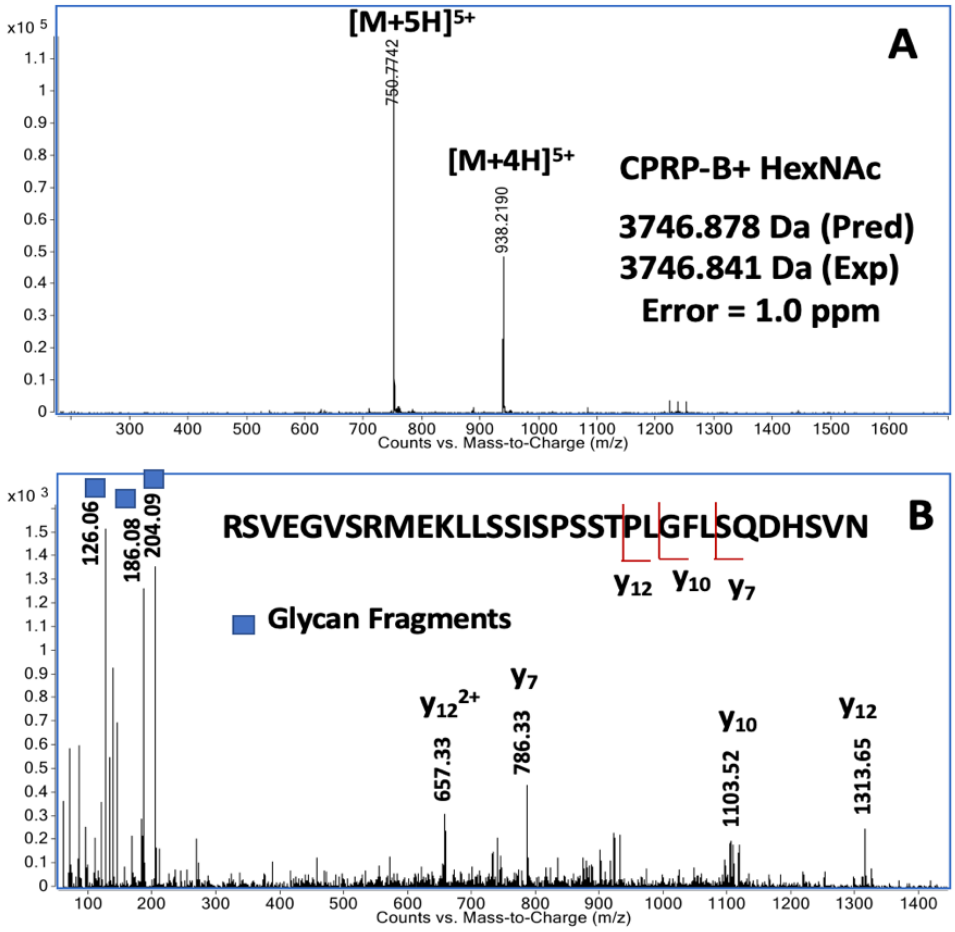


Figure 13: MS/MS spectra confirm identity of CPRP-B+HexNAc, but oxonium ion signal strength limits sequence ion presence.
A) Mass spectrum of the peptide eluting prior to CPRP-B from Figure 11A that yielded a *m/z* of 204.086 product ion. The mass, 3746.841 Da is consistent with the mass of the original CPRP-B+HexNAc. B) MS/MS spectrum showing the presence of glycan peaks in blue (*m/z*=126.06, 186.08, 204.09) indicative of glycosylation of CPRP-B. The y- and b- ions are fragments from CPRP-B, which support the amino acid sequence.

3.1.3 No additional glycopeptides were detected in unfractionated sinus glands

The inability to detect glycopeptides without preexisting peptide information is a major limitation of the approach outlined above. After detecting the known CPRPs in both unmodified and glycosylated forms, we turned to another approach to identify novel glycopeptides. Just as one can search MS data for evidence of peptide+glycan mass, it is also possible to search MS/MS data for the oxonium ions. Since the glycosidic bond is broken at low energies, the

oxonium ions often appear at high abundance in MS/MS spectra. Therefore, we searched MS/MS spectra for *m/z* 204.086 by generating an EIC for those product ion signals. Application of this approach to the unfractionated sinus gland only led to detection of glycosylated CPRPs. This limited identification may be due to the nature of how the MS/MS spectra are generated. For each run, the three highest abundance ions were chosen for dissociation via MS/MS, so this strategy only detects the highest abundance neuropeptides. It is possible that lower abundance glycosylated peptides were present, but not selected for MS/MS. One strategy to overcome this issue is to find a method for better separation, so that each run can be performed with fewer peptides. With fewer peptides, each run will have more potential MS/MS spectra covered, allowing expanded glycopeptide detection. This caused us to turn our approach to fractionation as a means of enhancing MS/MS sensitivity.

3.1.4 Fractionation allows detection of additional glycopeptides

With the goal of generating more MS/MS spectra for glycopeptide identification, high pH RP fractionation was applied to sinus gland extracts. Past research by Catherine Call found high pH RP fractionation to increase neuropeptide detection by a factor of 9.6 through the principle of orthogonal separation.[28] The sample is separated by elution with an increasing gradient of organic solvent prior to LC analysis. Thus, each sample has fewer neuropeptides, allowing increased acquisition of MS/MS spectra for lower abundance peptides. This high pH RP fractionation was carried out over ten fractions of increasing ACN concentration from 5 to 80%. We searched MS/MS spectra for presence of the m/z 204.08 peak to identify glycosylated peptides in each fraction. From the isolated peptides, we predicted their total mass from their charge states (see Table 4).

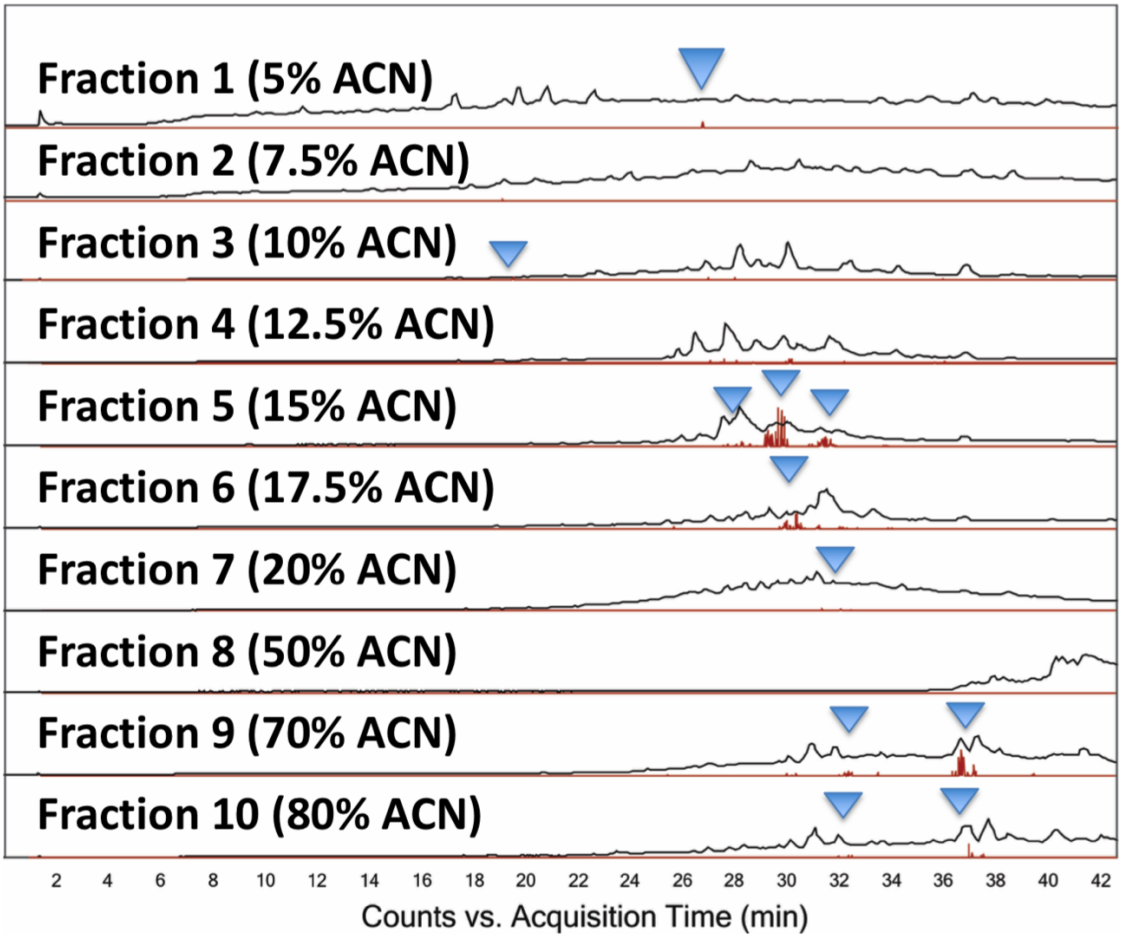


Figure 14: High pH fractionation revealed more glycosylated peptides.
Chromatogram derived from high pH fractionation overlaying TICs in black and EICs for *m/z*= 204.086 in red. The red peaks represent presence of glycosylated neuropeptides, with the most abundant peptides being located in fractions 5,6, 9, and 10. CPRP-A and B are located in fractions 9 and 10.

Table 4: List of glycosylated peptides detected in H. americanus sinus glands after high pH RP fractionation. The table lists the fraction number from which the peptide was observed, the most common charge state detected, the resultant mass, the glycan identity.

Fraction Number and % ACN	*m/z* Detected	Total Mass (Da)	Glycan	Previously Characterized?
1 (5%)	$[M+4H]^{4+}$= 1316.521	5262.055	HexNAc	No
3 (10%)	$[M+2H]^{2+}$= 618.299	1234.584	HexNAc	No
5 (15%)	$[M+3H]^{3+}$= 754.352	2260.054	HexNAc	No
5 (15%)	$[M+3H]^{3+}$= 755.025	2262.054	HexNAc	No
5 (15%)	$[M+3H]^{3+}$= 809.032	2424.075	HexNAc	No
5 (15%)	$[M+3H]^{3+}$= 716.341	2146.001	HexNAc	No
6 (17.5%)	$[M+3H]^{3+}$= 697.003	2087.987	HexNAc	No

7 (20%)	$[M+3H]^{3+}$= 873.107	2616.299	HexNAc	No
9 (70%)	$[M+5H]^{5+}$= 762.368	3806.802	HexNAc	Yes-CPRP-A[a]
9 (70%)	$[M+5H]^{5+}$= 750.358	3746.756	HexNAc	Yes-CPRP-B[a]
10 (80%)	$[M+2H]^{2+}$= 798.437	1588.860	HexNAc	No
10 (80%)	$[M+4H]^{4+}$= 909.192	3632.740	HexNAc	Yes- CPRP-B (-terminal N)

[a]Also detected in fraction 10.

This approach yielded twelve instances of glycosylated neuropeptides, showing three CPRPs and nine novel signals (Table 4). The peptides were split between fairly hydrophilic variants that eluted in the earlier fractions 1,3, and 5, and more hydrophobic peptides that eluted in the later fractions 6, 7, 9 and 10. The strength of this approach was shown by the nine additional glycosylated peptides detected outside of the previously discussed CPRPs A and B. Although the peptides were found to be glycosylated, the MS/MS spectra were dominated by the presence of oxonium ions. This reduces sequence ion signals, preventing identification of the peptides that were glycosylated. Consequently, these novel neuropeptides cannot be sequenced (data not shown). Despite this weakness, these data demonstrate the advantage of fractionation, as the fewer number of peptides per run allowed more MS/MS dissociation. Therefore, high pH RP fractionation is a method for increased glycosylation detection when searching for glycopeptides, but it cannot solve the issue of sequence ion loss. In order to address sequence ion loss, beta elimination was explored as a strategy to label and sequence O-linked glycopeptides.

3.2 Beta Elimination as a Strategy for Glycopeptide Identification

3.2.1 Beta elimination applied to fractionated sinus glands

Beta-elimination is a strategy for glycan removal that facilitates determination of the site of attachment. The reaction proceeded through use of a base to remove the glycan to generate a dehydroalanine (if the glycan was attached to a serine) or methyldehydroalanine (glycan attachment to threonine). This site is then primed for attack by a nucleophile via a thiol addition,

which can be used to label the site of glycosylation, improve production of sequence ions, and, ideally, produce a reporter ion that indicates the peptide had been glycosylated.

We started by applying the approach used previously by Henry Pratt in his identification of glycosylated variants of CPRP-A and B to the analysis of glycosylated CPRPs in fractions 9 and 10.[7] Pratt utilized 70% ethylamine as the base of choice, which was used for its ability to act as a nucleophile as well. This approach therefore pairs glycan removal with site tagging, allowing one reaction to determine glycan site and enhance sequence ion sensitivity. However, when this approach was applied to fractionated sinus glands, no modified peptides were detected (data not shown). We hypothesized that the signal loss could be due to either sample degradation or low sample sensitivity, so we applied the same procedure to full sinus glands. We hoped to use this larger sample to increase peptide abundances to replicate Pratt's results as applied to full CPRPs. If the degradation hypothesis was correct, the higher signal would allow detection of degradation products. If the low sensitivity hypothesis was correct, we would expect to see the beta-eliminated products in the full sinus gland due to the higher abundance of neuropeptides.

3.2.2 Beta elimination applied to Sinus Gland Extracts

We next subjected entire sinus glands to beta elimination with 70% ethylamine, and the initial results demonstrated robust removal of glycans (Figure 15). The EIC analysis for *m/z* 204.08 yielded no signal, suggesting total removal of glycan from the peptide in 70% ethylamine (Figure 15B). Alternatively, the glycosylated peptides could be of such low abundance that they

were not selected for MS/MS dissociation, and thus could not be detected with an EIC.

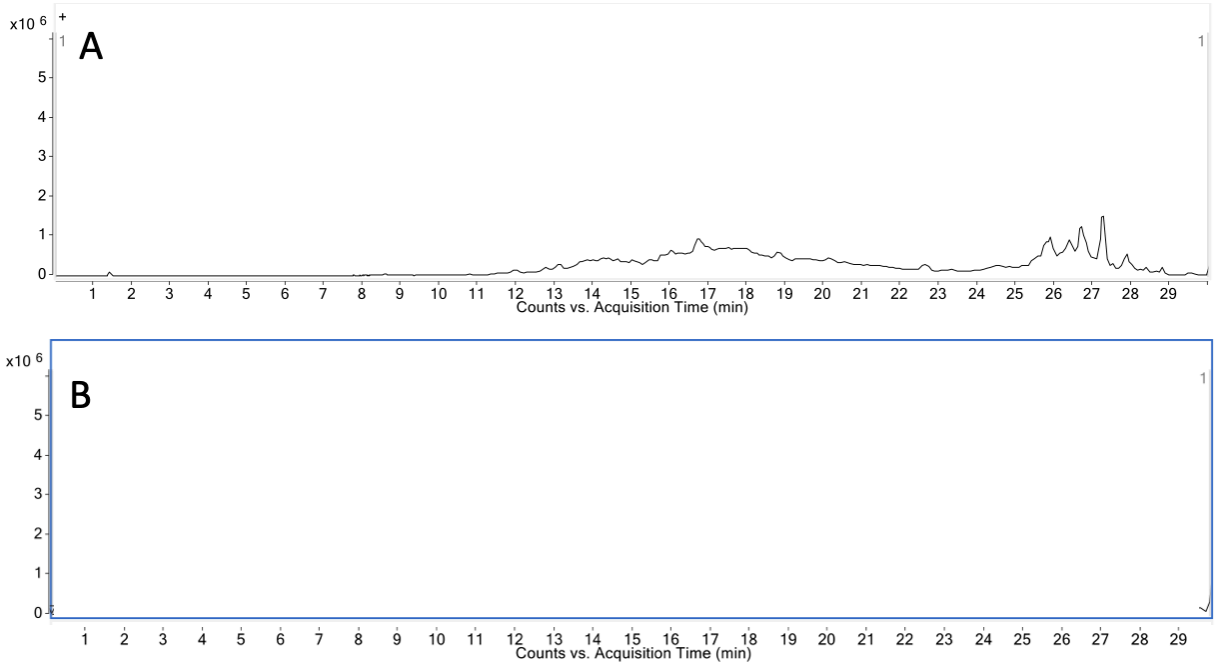


Figure 15: Beta elimination resulted in robust glycan removal.
A) TIC of sinus gland beta eliminated with 70% ethylamine. B) EIC of 204.08 for HexNAc for beta eliminated sinus gland displaying zero abundance of glycan.

However, this same procedure also yielded lower signals, making determination of glycosylation sites difficult (Figure 16). Figure 16A shows the chromatogram for a full sinus gland at 0 hr of beta elimination, illustrating the wide array of peaks representative of the multiple neuropeptides present in the sinus glands. The blue and red EICs demonstrate the abundant signals for both the unmodified and glycosylated CPRPs. Figure 16B demonstrates a clear degradation of most eluted peptides at 1.5 hr, as the peaks have fallen precipitously together. The only peaks showing clear signal are the unmodified CPRPs. As seen in figure 16C, by 2 hr, almost no intact CPRPs are observed (EIC CPRP; blue) and the glycosylated CPRPs are even less abundant (EIC CPRP+HexNAc/ CPRP-HexNAc-Hex; red and green). We hypothesized that this lower intensity was due to peptide degradation. The data shows loss of peptides as incubation time increased, suggesting the peptides were no longer present in the same

quantities. These low signals did not produce high quality MS/MS spectra, so the sequences could not be analyzed.

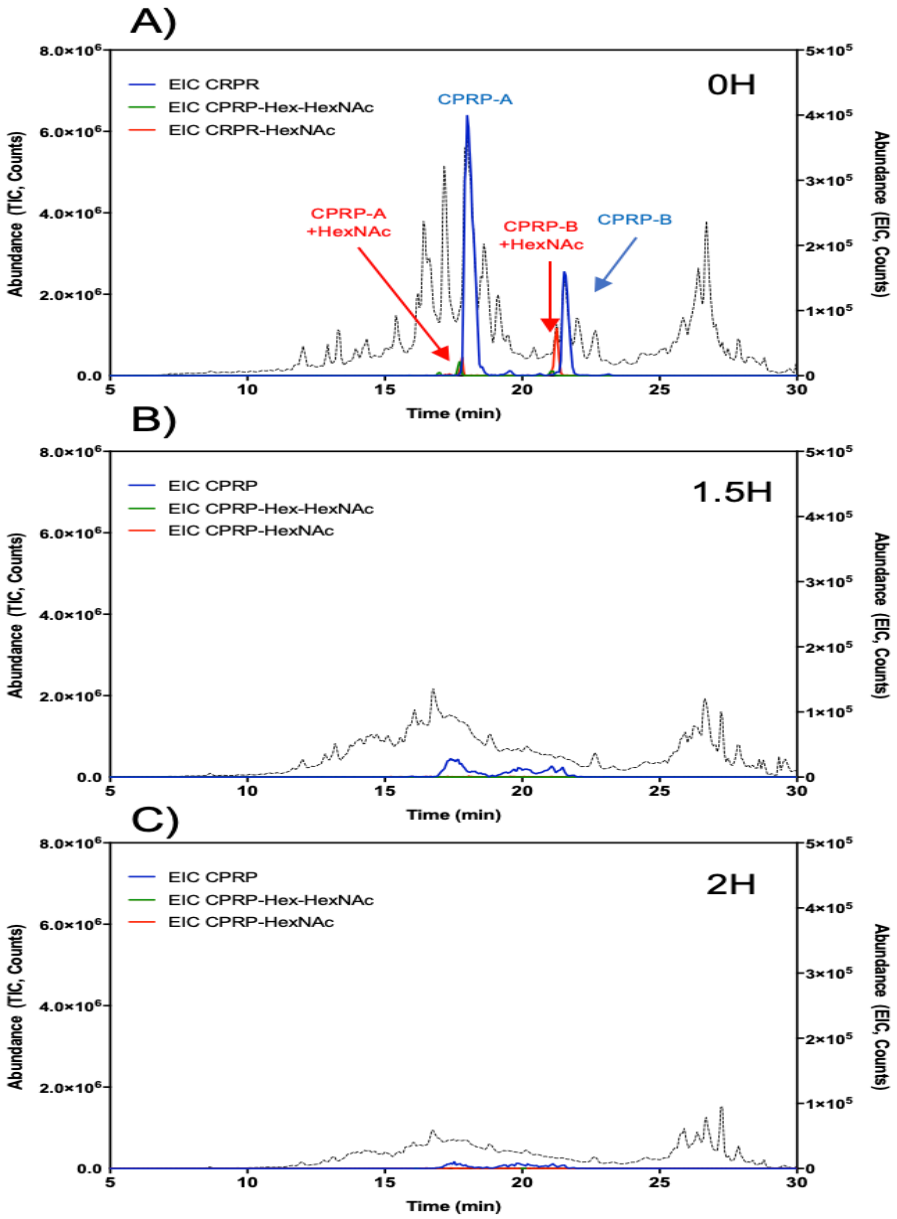


Figure 16: Beta elimination with ethyl amine caused significant sample degradation.
A) TIC for entire sinus gland prior to beta elimination with EICs for unmodified and glycosylated CPRPs, B) shows the same EICs after incubation in 70% ethylamine after 1.5 hr, and C) shows the same after 2 hr. EICs were for CPRP-A in blue (*m/z* 721.759; $[M+5H]^{5+}$ for CPRP-A), CPRP-A+HexNAc in red (*m/z* 762.375; $[M+5H]^{5+}$ for CPRP-A+HexNAc), and CPRP-A+HexNAc-Hex in green (*m/z* 794.785; [M+5H]5+ for CPRP-A+HexNAc-Hex). The EICs also targeted CPRP-B in blue (*m/z* 709.767; $[M+5H]^{5+}$ for CPRP-B), CPRP-B+HexNAc in red (*m/z* 750.383; $[M+5H]^{5+}$ for CPRP-B+HexNAc), and CPRP-B+HexNAc-Hex in green (*m/z* 782.794; [M+5H]5+ for CPRP-B+HexNAc-Hex).

Although there was significant sample loss, some degradation products were high enough in abundance to analyze. The degradation fragments that were detected included serine-rich regions, indicating the degradation was occurring around the serine residues. Since these are the potential sites for glycan attachment, this only further complicated glycosylation characterization. Since the total peptide signal became split between many degradation products, there was not high enough signal to generate the high quality MS/MS spectra needed to characterize the glycopeptides. Thus the trend of degradation also observed in fractionated samples persisted. There was little literature precedent for this degradation issue, and this could be attributed to traditional beta elimination approaches using proteolytic digestion prior to analysis.[25,29] Additionally, another strange trend was observed wherein threonine residues were converted into glycine residues (data not shown). There is no literature precedent for this conversion, so we were not able to isolate the reason for this process. The combination of sample loss due to degradation and threonine conversion resulted in no detection of ethylamine modified glycopeptides. Therefore, the ethylamine approach proved unsuccessful in both fractionated samples and full sinus gland extracts. This result caused the project to shift towards another approach using proteases to reduce peptide backbone size prior to beta elimination. This approach was utilized in hopes of mitigating the destructive effects of ethylamine on the backbone by limiting the size of the peptides.

3.3 Chymotryptic Digestion as a Strategy to Limit Peptide Degradation

Sinus gland extracts were reacted with chymotrypsin in hopes of reducing the size of neuropeptides to limit the effects of degradation. Additionally, this approach has the dual benefit of limiting the number of amino acid residues in each fragment, therefore limiting the potential sites for glycan attachment. Reaction of sinus glands with chymotrypsin results in peptides being

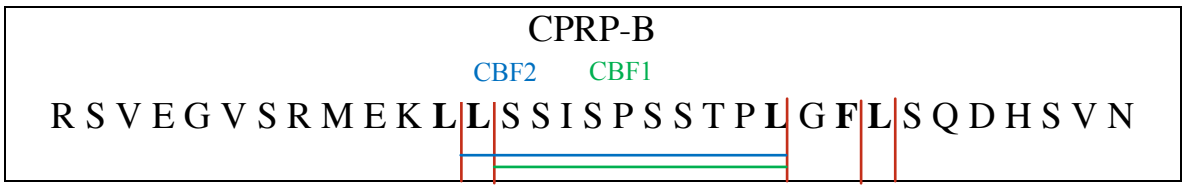
cleaved at leucine (L), phenylalanine (F), tyrosine (Y), and tryptophan (W) amino acid residues. The reason chymotrypsin was chosen as opposed to other proteases was specific to the CPRPs, as alternatives such as trypsin were predicted to create too few proteolytic products. Trypsin cuts on the carboxyl side of lysine and arginine residues, which would create a fragment with nine possible glycosylation sites. Alternatively, chymotrypsin was predicted to create multiple large fragments with fewer possible glycan attachment sites, so it was chosen over other proteases (Table 5). Chymotrypsin was reacted with whole sinus glands extracts for 12 hours, and the results were analyzed via LC-MS. The sequence of CPRP-B is shown below, demonstrating the potential sites at which chymotrypsin can cut (Table 5). Notably, the first potential fragment contains only two potential O-glycosylation residues. The third fragment contains six potential glycosylation sites, and the sixth fragment contains the remaining two sites. Examination of these particular fragments allows targeting the presence of glycosylation with fewer possibilities than the whole peptide. If a given fragment didn't show evidence of glycosylation, that would provide evidence for the glycan being localized to other residues. Absence of *m/z* 204.086 product ions for HexNAc in the MS/MS spectra would indicate that the serine/ threonine residues included in those fragments would not be the site of glycosylation.

Table 5: Sequences of CPRP-A/B with predicted cleavage sites shown with red lines. Chymotrypsin is shown to cleave at the C-terminal side of leucine (L) and phenyalanine (F) residues. Fragments are shown in red, blue and green to demonstrate the three most common fragments formed after chymotrypsin digestion.

CPRP-A

CAF1

R S V E G A S R M E K **L** **L** S S S N S P S S T P **L** G **F** **L** S Q D H S V N



The most abundant fragment was the result of leucine cleavage of CPRP-B, resulting in the formation of a peptide with the sequence SSISPSSTPL, known as CBF1 (Figure 17, Table 6). These smaller peptide fragments primarily appear in the +2 charge state, which is due to the smaller number of basic amino acids for protonation (Table 5). This experiment was conducted to confirm the ability to detect fragments of the CPRPs, and the two neuropeptides yielded three fragments with high signals for downstream analysis (the CAF2 was not detected, see Figure 17). The next step was to look for the continued presence of glycosylation in the fragments. Fragments without glycans in their MS/MS data can allow ruling out those areas of the peptide from being glycosylated.

Table 6: List of expected fragmentation results derived from CPRPs in descending order of abundance. Fragment sequences display the sequence determined from MS/MS data that is derived from the parent neuropeptide.

Original Peptide	**Fragment Sequence**	**Fragment Name**	**Observed *m/z* (Da)**
CPRP-A	SSISPSSTPL	CAF1	532.249
CPRP-B	SSISPSSTPL	CBF1	488.253
CPRP-B	LSSISPSSTPL	CBF2	544.795

In addition to the high signal of the chymotrypsinized fragments, some of the fragments also displayed high signals for their glycosylated counterparts (Figure 17A). Only one of the common fragments displayed in Table 5 were also abundant with the addition of HexNAc glycans (CBF1). CAF1 did not show any evidence of glycosylation, as an EIC for its *m/z* 633.79 yielded no signal (Figure 17A). However, glycosylated CBF1 was detected in the

chymotrypsinized sinus glands. This result is unexpected, since CPRP-A was found to be glycosylated in the early full sinus gland analysis (Figure 12). Therefore, one hypothesis is that the glycan is attached at another site outside of CAF1. This could potentially rule out the 7 possible O-glycan attachment residues for CPRP-A included in CAF1 (Table 5). The inverse is true for CBF1, which showed evidence of the glycan being localized within CBF1's 10 amino acids. The masses demonstrated mass increases of 204.08 indicative of HexNAc addition, implying that the glycans are attached at serine 14, 15, 16, 18, 20, or 21 (for CPRP-A, for CPRP-B, 14, 15, 17, 19, or 20) or threonine (A=22, B=21) residues contained in these fragments. Consequently, this suggests that glycosylation does not occur at the residues not included in these fragments. Therefore, we can narrow the site of glycosylation to between residues 14 and 21. Had the fragments contained only one serine or threonine each, the location of the glycan on these peptides would be solved. This is important looking forward to other neuropeptides, as if the sinus glands contain peptides with fewer serine residues, chymotryptic digestion could be sufficient. This still leaves five potential serine and one threonine residues for the glycan to attach, which requires the use of beta elimination and nucleophilic attack to label the site. Therefore, we proceeded with the use of ethylamine to test if the smaller fragments yielded higher signal for the addition. Ethylamine was chosen for its ability to act as a nucleophile and base simultaneously, exemplified by its prior use by Pratt.

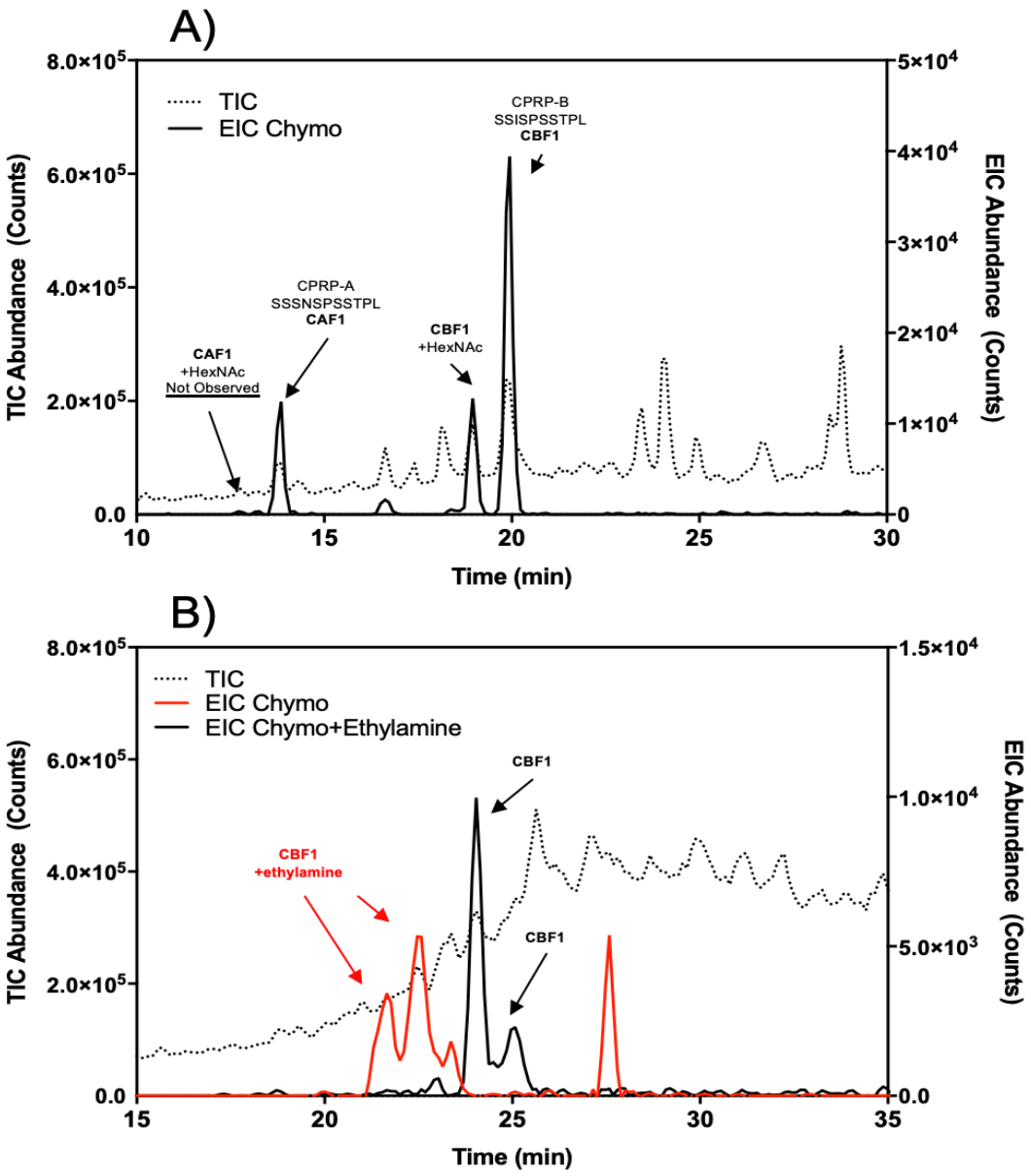


Figure 17: Beta elimination with ethylamine results in multiple product formation. **A)** TIC of chymotrypsinized sinus gland prior to beta elimination overlaid with EICs for common chymotryptic CPRP fragments. The EICs targeted CAF1 and CBF1 (*m/z* 532.2 and 488.25; [M+2H]2+ for CAF1 and CBF1 respectively) and their HexNAc glycosylated counterparts (*m/z* 633.79 and 589.79; [M+2H]2+ for CAF1-HexNAc and CBF1-HexNAc respectively) **B)** TIC of chymotrypsinized sinus gland after beta elimination with ethylamine overlaid with EICs for unmodified and modified peptides. The EICs targeted CAF1 and CBF1 (*m/z* 532.2 and 488.25; [M+2H]2+ for CAF1 and CBF1 respectively) and their ethylamine modified counterparts (*m/z* 545.78 and 501.78; [M+2H]2+ for CAF1-ethylamine and CBF1-ethylamine respectively). No product for CAF1 was detected after ethylamine addition.

The use of ethylamine to beta eliminate and label chymotrypsinized fragments resulted in some modification of the CBF1 fragment, but no labeling for CAF1 (Figure 17). Despite evidence for ethylamine addition for CBF1, the chromatography suggests the creation of multiple products due to the presence of multiple wide peaks for the CBF1+ethylamine EIC (Figure 17B, red lines). All these peaks have the same *m/z* of 501.78, yet they elute over a range of 5 mins as 3 peaks, while the untreated native peptide CBF1 elutes in 1 min as 1 peak. There is also signal for a fourth CBF1+ethylamine peak that elutes 5 mins later than the others, which is not predictable chromatographically (Figure 17B, red lines). The ethylamine addition adds a polar, basic sidechain to the peptide, which should result in a shift downstream due to its lower affinity for the stationary LC phase. Therefore, the elution of this CBF1+ethylamine peak later in the run is currently not understood. This formation of multiple products could be the result of racemization, as the products all have the same mass, but different retention times. This potential racemization is explored in our experiments with glycopeptide standards, which supports the hypothesis that this formation of 2 products occurs during beta elimination prior to nucleophilic addition (see section 3.5).

The second issue is the loss of the CAF1 signal. Despite both of these CPRP fragments being present prior to beta elimination, the CAF1 signal not only didn't show evidence of ethylamine addition, but the original peptide signal was lost all together (Figure 17B). Therefore, it seems there may still be some peptide degradation occurring, as the beta elimination reaction would only be expected to perform chemistry on the glycosylated CAF1 variant, not the unmodified version. In addition to the signal loss for unmodified CAF1, the unmodified CBF1 signal was also split into 2 peaks after reaction with ethylamine (Figure 17B). Since these 2

peaks share the same *m/z* of 488.25, there must be racemization of the native peptide as well. This suggests that the ethylamine is reacting with both glycosylated and unmodified peptides.

Reaction of chymotrypsinized sinus glands with ethylamine presented multiple issues. While ethylamine was added to CBF1, no such addition was detected for CAF1. Since CBF1 is more abundant, it is possible ethylamine addition only successfully tags high abundance peptides, making it a poor choice for our goal of characterizing low abundance neuropeptides. Furthermore, the addition products for CBF1 appeared racemized, eluting over a wide range with multiple peaks. The racemization results in less signal for each peak, further reducing sensitivity. Finally, the ethylamine reaction conditions caused the loss of unmodified CAF1, suggesting degradation of even unmodified peptides. As a result of these findings, we shifted away from the use of ethylamine as our base and nucleophile for beta elimination. We instead turned to an approach utilized by Steen, which utilizes a thiol nucleophile for 3 advantages in glycopeptide characterization.[26]

3.4 Use of a thiol nucleophile based beta elimination/ thiol addition strategy increased glycopeptide detection in chymotrypsinized sinus glands

Literature evidence indicated that 2-dimethylaminoethanethiol (DMAET) created a strong signal marker ion after nucleophilic addition that could aid in glycopeptide detection (Figure 18).[26,30] This method utilizes beta elimination in concert with thiol addition and contains 3 distinct advantages: first, the addition creates a tag ion that can be easily searched in MS/MS data to isolate glycopeptides.[26] The thiol ion is unique, only present in previously glycosylated peptides; searching for this ion via an EIC allows rapid isolation of peptides with glycans originally attached. The second advantage is that the added thiol group has a basic amine group, which allows another ionization site to aid in sensitivity. Third, the newly created thiol-peptide

bond is less labile than the glycan-peptide bond, fragmenting at higher energies. Since the bond doesn't fragment as readily, sequence ions can be detected with the thiol attached. This allows isolation of residues with the tag added, permitting localization of the original glycan. These advantages were favored to help in addressing the repeated issues of low sensitivity and poor addition product detection. Adopting this new approach, the chymotrypsinized samples were reacted with 0.1 M NaOH as the base of choice due to literature precedent and its poorer nucleophilicity relative to ethylamine.[26,29] Rather than the base being the nucleophile in the case of ethylamine, this procedure used 0.54 M of the thiol DMAET.[30]

Reaction with the new conditions yielded DMAET addition to CBF1. The fragment CBF1 was detected in its +2 charge state with the mass addition of 53.03 expected of DMAET addition and glycan loss (Figure 18B). The elution time matches expectations, with the addition of the polar group causing an earlier elution due to less hydrophobicity. However, the chromatography once again shows evidence for multiple products forming at the *m/z* 531.78 (Figure 18B, red lines). This returns to the possible scenario of racemization, as the DMAET addition products share the same mass, yet yield variable stationary phase retention times. Similarly to the ethylamine reaction, the unmodified CBF1 peaks also show evidence of multiple product formation. CBF1 shifts from a single, narrow peak to a series of 3 peaks eluting over 3 mins (Figure 18, black lines). This suggests that the reaction conditions affect peptides regardless of glycosylation.

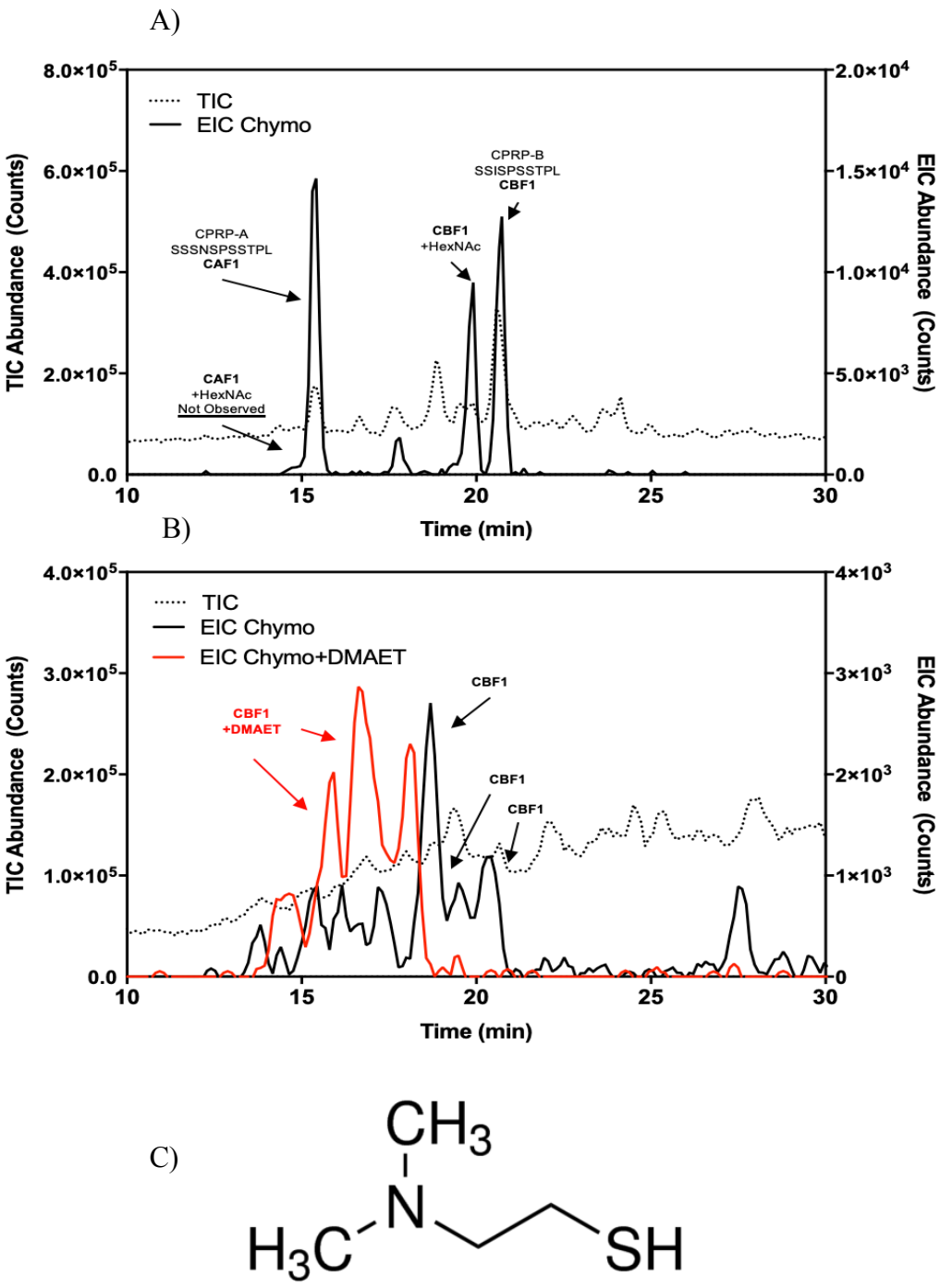


Figure 18: Beta elimination with NaOH alongside labeling with DMAET creates stronger signal than addition with ethylamine.

A) TIC of chymotrypsinized sinus gland prior to beta elimination overlaid with EICs for common chymotryptic CPRP fragments. The EICs targeted CAF1 and CBF1 (*m/z* 532.2 and 488.25; [M+2H]2+ for CAF1 and CBF1 respectively) and their HexNAc glycosylated counterparts (*m/z* 633.79 and 589.79; [M+2H]2+ for CAF1-HexNAc and CBF1-HexNAc respectively) **B)** TIC of chymotrypsinized sinus gland after beta elimination with NaOH and reaction with DMAET overlaid with EICs for unmodified and modified peptides. The EICs targeted CAF1 and CBF1 (*m/z* 532.2 and 488.25; [M+2H]2+ for CAF1 and CBF1 respectively) and their DMAET modified counterparts (*m/z* 575.77 and 531.78; [M+2H]2+ for CAF1-DMAET and CBF1-DMAET respectively) **C)** Structure of DMAET.

While the chromatography for DMAET addition showed evidence for racemization issues, the MS/MS data showed promising results. Consistent with the expectation of tag ion detection, the peak of *m/z* 106.07 was found, which corresponds to the loss of DMAET (Figure 19B). This m/z is not detected from unmodified samples, making it an efficient way to profile samples for glycopeptides. In addition to marker ion creation, unlike the past beta eliminated samples where the y- and b- sequence ions were nearly nonexistent, they were clearly visible for the DMAET treated sample (Figure 19B). The persistent issue of oxonium ions dominating MS/MS spectra was removed, allowing detection of multiple sequence ions (Figure 19B). The third advantage discussed previously centered around glycan site determination via the more stable thiol bond. The first step involved ruling out areas that didn't show evidence of DMAET addition. Ion fragments detected with no DMAET addition suggest they did not host the glycan, since they did not experience mass additions demonstrative of glycan removal and replacement by DMAET. The fragment CBF1 demonstrated presence of b_4 and b_2 ions without the mass addition of DMAET, indicating these residues were not the site of glycosylation. Since we detected a y_6 ion showing evidence of DMAET addition, this limited the potential glycosylation site to 2 final serine residues. Unfortunately, we did not detect the y_4 ion that would allow the distinction between these 2 possible serine residues.

One potential explanation for the absence of this ion lies in the difficulty surrounding proline fragmentation. The proline effect refers to the tendency for prolines to not readily form b-ions. This is due to the nature of the secondary amine in proline, which makes fragmentation at this site less likely than in other amino acid residues.[31] Unfortunately, the final 2 serine residues are surrounded by 2 prolines, making formation of the ions necessary to declare the site of glycosylation currently impossible (Figure 19B). This observation opens further questions about

the other CPRPs, given that these neuropeptides have conserved sequences that include many shared residues, including serine residues flanked by proline. While the proline effect disallowed determination of the specific serine of glycan attachment, the presence of the DMAET marker ion demonstrates the potential utility of this system for peptides with glycans not surrounded by prolines. Additionally, the higher sequence ion signal is a strong testament to the potential for this approach to characterize novel glycopeptides. Therefore, DMAET addition provides a useful tool for detecting glycopeptides via tag searching, as well as novel sequencing via enhanced sensitivity. However, given the formation of multiple addition products, we turned to glycopeptide standards to determine the ideal conditions for DMAET addition.

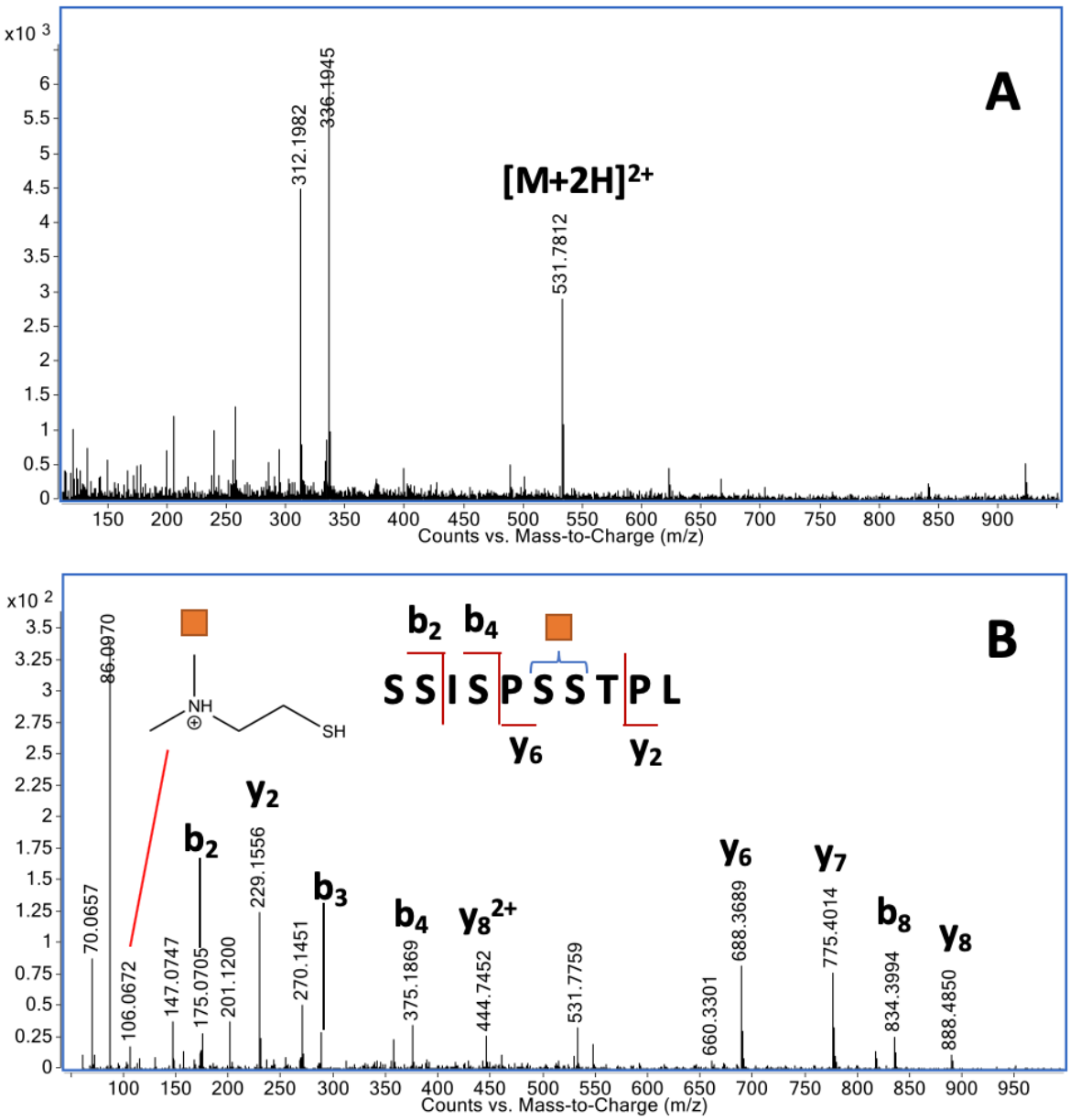


Figure 19: Beta elimination with NaOH and labeling with DMAET results in better sequence ion signal and reporter ion creation.
A) Mass spectrum of CPRP-B fragment 1 taken from chymotrypsinized sinus gland extracts beta eliminated with 0.1M NaOH and reacted with DMAET displaying presence of CPRP-B fragment 1 with DMAET addition. B) MS/MS spectrum for CBF1 with DMAET addition. The y- and b-ions demonstrate prominent fragments formed from the chymotrypsinized peptide sequence. The orange box represents DMAET, showing the location cannot be determined between two potential serine residues.

3.5 Beta elimination/ thiol addition applied to glycopeptide standards

After it was found that 2-dimethylamino-ethanethiol (DMAET) produced a marker ion with enhanced sequence ion signal, we sought to better characterize this reaction. The DMAET addition product eluted over a multitude of peaks, so we sought to test the reaction kinetics in a controlled setting. Glycopeptide standards were purchased that contained serine residues O-glycosylated with HexNAc (Figure 20). One important distinction between these standards and endogenous neuropeptides is the lack of proline, which avoids the issue of the proline effect. We elected to test a simpler peptide, testing our reaction conditions to determine the time at which multiple products were formed. Through these experiments, we hoped to isolate a refined method for peptide labeling to apply to lobster neuropeptides that would avoid the issue of peptide racemization.

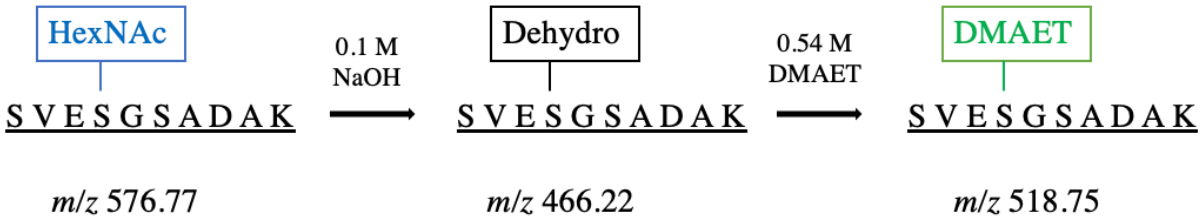


Figure 20: Glycopeptide standards can be used to model the beta elimination reaction. Schematic showing *m/z* values associated with steps in glycopeptide standard beta elimination, progressing from initial glycopeptide to beta eliminated dehydro standard to DMAET thiol addition product.

In order to characterize the kinetics of beta elimination we first explored the initial elimination step of the labeling reaction. We incubated the standards in 0.1 M NaOH over a time frame of 120 mins to determine when the glycan was removed. Our reaction conditions involved quenching the reaction with glacial acetic acid at multiple time points to gauge the amount of dehydro product formation. Contrary to the reaction conditions drawn from the literature of a 1

hour incubation at 55°C, we found high rates of glycan loss after just 5 min (Figure 21). Even at 0 mins, where we would expect exclusively the glycosylated form of the standard, the dehydro product (*m*/*z*= 466.22) had some formation (Figure 21A). This could be due to the glycopeptide standard being exposed to NaOH for seconds before the addition of acetic acid, which accentuates the speed of the elimination step. By 5 mins, almost all of the glycosylated form had been lost to form the dehydro product (Figure 21B). Furthermore, the dehydro product shifts forward in elution time from the glycosylated standard, which is consistent with loss of the polar glycan moiety. These results demonstrate the rapid rate of beta elimination, going against prior literature recommendations of incubating for an hour.

The dehydro signal was split between 3 peaks with differing elution times and intensities, suggesting the formation of multiple products with the same *m*/*z*=466.22. This issue was compounded with increased time, as the dehydro peak intensities decreased over time until 15 mins, where they remained until 120 mins (Figure 21C-F). The peaks have the same *m*/*z*, but varying elution times, which means that there is likely to be either rearrangements or racemization occurring. MS strategies are not sufficient to isolate the site of racemization or rearrangement, but the process can be inferred from the peaks yielding the exact same mass. This is a potential explanation for the wide chromatographic peaks seen in DMAET addition to the CPRPs. If the deglycosylation step yields multiple dehydro products after just 10 mins, then DMAET would have multiple products to attack as a nucleophile, which could account for the multiple peaks eluting for the CPRP+DMAET product (Figure 18). Interestingly, the peak intensity also decreases with time, implying that the dehydro product undergoes further reactions after 5 mins. These other products have not been isolated, but they are another likely explanation for dehydro loss over time. This sample loss over time may explain the previous issues of sample

loss experienced while beta eliminating whole sinus gland extracts. When whole sinus glands were beta eliminated, the conditions employed 10 M ethylamine, which is dramatically more basic than the 0.1 M NaOH used with the standards. In order to counter the dehydro racemization and subsequent sample loss, it seems that we should shorten incubation times to 5 mins for beta elimination.

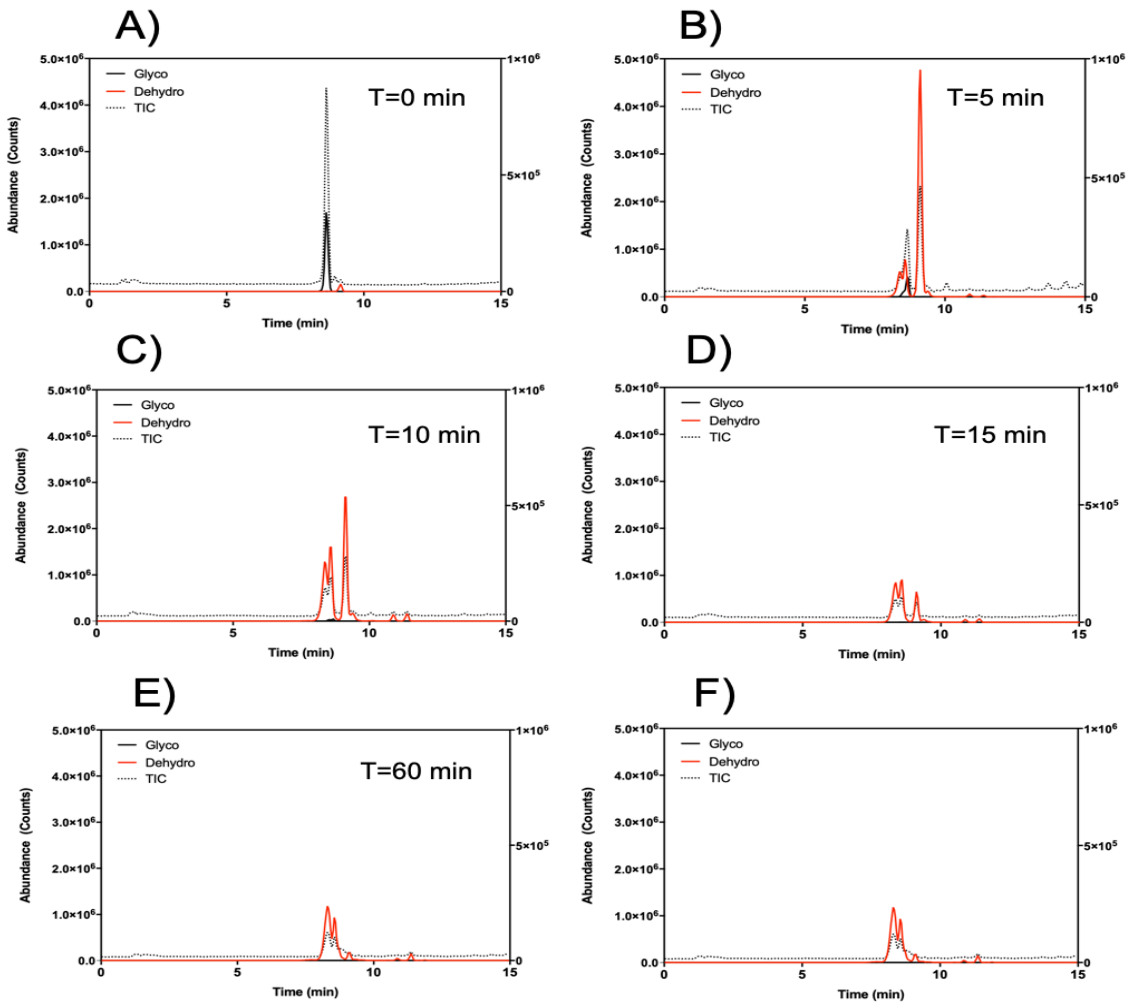


Figure 21: Beta-elimination occurs very rapidly in standards, but racemization occurs after extended incubation.
Chromatograms for glycopeptide standards at A) 0 min, B) 5 min, C) 10 min, D) 15 min, E) 60 min and F) 120 min after addition of 0.1 M NaOH at 55°C. The dotted black line represents the TIC, black lines show the EIC for the glycosylated standard (EIC using *m/z* 576.77; $[M+2H]^{2+}$ for glycopeptide standard), and red lines show the formation of the dehydro product (combined EIC using *m/z* 466.22; $[M+2H]^{2+}$ for dehydro form of the glycopeptide standard).

After observing the rapid kinetics of beta elimination with 0.1 M NaOH, we sought to use the standard to explore DMAET addition. We conducted a similar experiment, incubating the standard in 0.1 M NaOH and 0.54 M DMAET over the course of an hour. Literature precedent called for beta elimination and thiol addition to occur simultaneously, so this was retained in our kinetics experiments.[26] Similar to the beta elimination kinetics, DMAET addition was carried out rapidly (Figure 22). Once again, even at 0 mins where we would only expect the intact glycopeptide standard, some dehydro and DMAET products were observed (Figure 22A).

After just 5 mins, there was almost no glycopeptide left, no dehydro signal, and abundant DMAET addition product (Figure 22B). Intriguingly, the DMAET chromatography (*m*/*z*= 518.78) at even 5 mins showed considerable peak width and the emergence of a few minor peaks, suggesting that multiple product formation was immediate. Furthermore, the formation of multiple DMAET products increased over 60 mins, ending with 5 peaks of lower signal than the initial DMAET peak (Figure 22F). Therefore, it seems that the issues of beta elimination persisted, and racemization is occurring for the DMAET products. Once again, the solution seems to lie in excessive reaction time. The conditions used previously to tag CPRPs with DMAET ran the beta elimination over the course of an hour, which also contained multiple peaks in the chromatography. Therefore, shortening the reaction to 5 mins seems like a potential solution to minimize product loss.

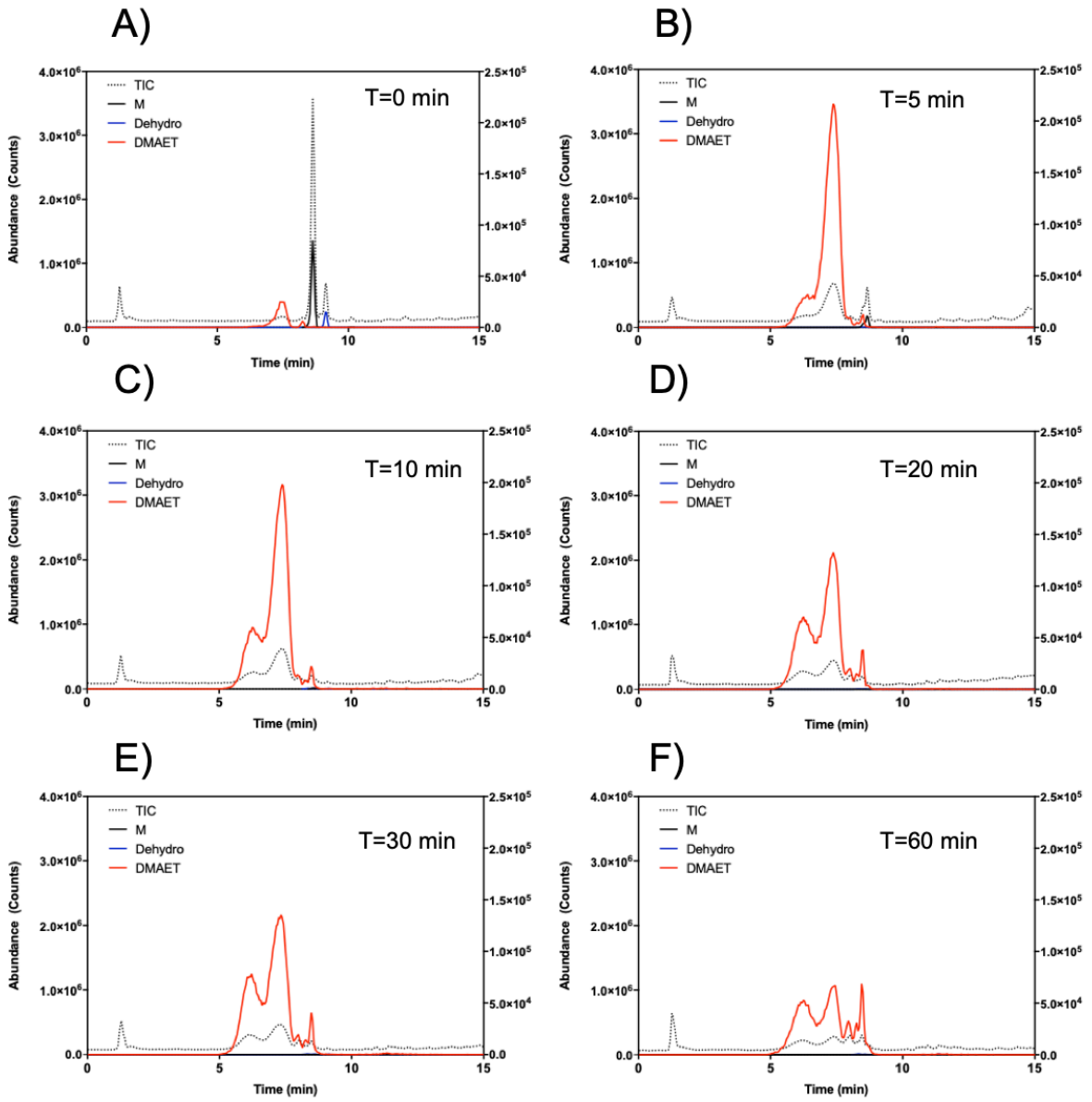


Figure 22: DMAET addition is rapid, but forms multiple products.
Chromatograms for glycopeptide standards at A) 0 min, B) 5 min, C) 10 min, D) 20 min, E) 30 min and F) 60 min after addition of 0.1 M NaOH and 0.54 M DMAET at 55°C. The dotted black line represents the TIC, black lines show the EIC for the glycosylated standard (EIC using *m/z* 576.77; $[M+2H]^{2+}$ for glycopeptide standard), the blue lines show the formation of the dehydro product (combined EIC using *m/z* 466.22; $[M+2H]^{2+}$ for dehydro form of the glycopeptide standard), and the red lines shows the EIC for the DMAET addition product (EIC using *m/z* 518.76; $[M+2H]^{2+}$ for DMAET addition to the glycopeptide standard).

Future Work

This study has demonstrated the ability of beta elimination paired with thiol-addition to tag glycopeptides and aid in glycan site determination. However, this approach has not proven capable of dealing with glycosylation sites with nearby prolines that inhibit ion fragmentation. Furthermore, although limiting beta elimination reaction time to 5 mins improved chromatographic separation, DMAET addition still resulted in multiple product formation and broad chromatographic peaks. These issues limit the sensitivity of the approach, reducing the ability to detect and characterize novel glycosylated neuropeptides.

Future experiments that could be conducted to optimize this glycosylation profiling approach are numerous. While the beta elimination and thiol-addition have been shown to occur within the first 5 mins, there is still some product loss even after such a short incubation. One potential solution could involve testing a temperature range to find the optimum temperature to minimize side reactions. Additionally, the concentration used for NaOH as the base has not been thoroughly examined. The concentration of 0.1 M has been a literature standard for years, but there may still be a value that limits the racemization observed in the glycopeptide beta elimination. Similarly, the nucleophile concentration suggested by Steen[26] of 0.54 M was not supported by any kinetic data, nor was attributed to a previous paper's strategy. Therefore, the nucleophile concentration may be subject to further optimization.

This experiment focused on the 2 neuropeptides CPRP-A and B as models, but there is still little known about the other neuropeptides found in the lobster sinus glands. Once reaction conditions are optimized, the approach can be applied to crustacean samples. The first step will necessitate validating the approach with crustacean neuropeptides. The kinetics experiments conducted in this study utilized glycopeptide standards that were shorter than CPRPs (10 vs 35) and contained no prolines around the glycosylation sites. The greater length of the neuropeptides

could alter the reaction kinetics due to changes in folding or glycan solvent accessibility. Additionally, given that the proline effect limits formation of sequence ions, the presence of any proline residues around the site of glycosylation could limit glycan site determination. Therefore, the approach will need to be tested in crustacean samples to confirm its efficacy.

Assuming the reaction conditions are subsequently optimized, the approach can then be applied to fractionated samples. The use of high pH RP fractionation demonstrated the presence of four other glycosylated neuropeptides to which this procedure can be applied. The fractionation approach relied on searching MS/MS data for the presence of the *m/z* 204 peak, which was limited to peptides of high enough frequency to be selected for MS/MS dissociation. These 4 were shown with the old approach, so the introduction of the thiol tag may further increase sensitivity, allowing more glycopeptides to be detected. Furthermore, addition of the stable thiol group will allow determination of the site of glycan attachment. Finally, the oxonium ions will no longer be formed, allowing for an increased number of sequence ions. With these 3 benefits, we should be able to characterize the sequence and glycosylation site for additional neuropeptides in fractionated samples. This will further the goal of characterizing all of the glycosylated neuropeptides found in *H. americanus*.

One final set of experiments would include the examination of alternative nucleophiles. This study originally hoped to examine the effect of 3-(dimethylamino)-propylamine on glycan profiling. This chemical resembles DMAET, except for the amine group in place of the thiol serving as the nucleophilic site. Perhaps this nucleophile may solve persistent issues such as the broad chromatographic peaks created by DMAET addition. One way in which this issue may be resolved if the new nucleophile does not cause racemization, which would cause fewer peaks to be created after thiol addition. These are just two potential nucleophiles chosen for their creation

of unique tag ions and stabile bonds, but there are a multitude of possibilities that could further optimize the approach.[26, 30, 32]

Finally, the ultimate hope of this project is to apply optimized conditions to the entire, fractionated sinus gland to profile to the lobster glyconeuropeptidome. While this is certainly a distant goal, the expansion of the approach to more fractions collected via high pH fractionation can act as the first step towards this eventual goal.

References

1. Russo, A. F., Overview of Neuropeptides: Awakening the Senses? *Headache* **2017,** *57 Suppl 2* (Suppl 2), 37-46.

2. Skiebe, P., Neuropeptides are ubiquitous chemical mediators: Using the stomatogastric nervous system as a model system. *Journal of Experimental Biology* **2001,** *204* (12), 2035-2048.

3. Hughes, J.; Woodruff, G. N., Neuropeptides. Function and clinical applications. *Arzneimittelforschung* **1992,** *42* (2a), 250-5.

4. Dircksen, H.; Böcking, D.; Heyn, U.; Mandel, C.; Chung, J. S.; Baggerman, G.; Verhaert, P.; Daufeldt, S.; Plösch, T.; Jaros, P. P.; Waelkens, E.; Keller, R.; Webster, S. G., Crustacean hyperglycaemic hormone (CHH)-like peptides and CHH-precursor-related peptides from pericardial organ neurosecretory cells in the shore crab, Carcinus maenas, are putatively spliced and modified products of multiple genes. *Biochem J* **2001,** *356* (Pt 1), 159-170.

5. Wilcockson, D. C.; Chung, S. J.; Webster, S. G., Is crustacean hyperglycaemic hormone precursor-related peptide a circulating neurohormone in crabs? *Cell Tissue Res* **2002,** *307* (1), 129-38.

6. Liang, Z.; Schmerberg, C. M.; Li, L., Mass spectrometric measurement of neuropeptide secretion in the crab, Cancer borealis, by in vivo microdialysis. *Analyst* **2015,** *140* (11), 3803-3813.

7. Pratt, H. Characterization of Glycosylated CPRP Neuropeptides in the American Lobster Homarus americanus Using Liquid Chromatography-Tandem Mass Spectrometry. Bowdoin College, Brunswick, ME, 2015.

8. Fanjul-Moles, M. L., Biochemical and functional aspects of crustacean hyperglycemic hormone in decapod crustaceans: review and update. *Comp Biochem Physiol C Toxicol Pharmacol* **2006,** *142* (3-4), 390-400.

9. Tinoco, A. D.; Saghatelian, A., Investigating endogenous peptides and peptidases using peptidomics. *Biochemistry* **2011,** *50* (35), 7447-61.

10. Bradbury, A. F.; Smyth, D. G., Peptide amidation. *Trends in Biochemical Sciences* **1991,** *16*, 112-115.

11. Merkler, D. J., C-terminal amidated peptides: production by the in vitro enzymatic amidation of glycine-extended peptides and the importance of the amide to bioactivity. *Enzyme Microb Technol* **1994,** *16* (6), 450-6.

12. Prigge, S. T.; Mains, R. E.; Eipper, B. A.; Amzel, L. M., New insights into copper monooxygenases and peptide amidation: structure, mechanism and function. *Cell Mol Life Sci* **2000,** *57* (8-9), 1236-59.

13. RG, S., Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* **2002,** *12*, 43R-56R.

14. Brockhausen, I.; Schachter, H.; Stanley, P., O-GalNAc Glycans. In *Essentials of Glycobiology*, nd; Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. E., Eds. Cold Spring Harbor Laboratory Press

The Consortium of Glycobiology Editors, La Jolla, California.: Cold Spring Harbor (NY), 2009.

15. Stanley, P.; Schachter, H.; Taniguchi, N., N-Glycans. In *Essentials of Glycobiology*, nd; Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. E., Eds. Cold Spring Harbor Laboratory Press

The Consortium of Glycobiology Editors, La Jolla, California.: Cold Spring Harbor (NY), 2009.

16. Varki, A., Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* **1993,** *3* (2), 97-130.

17. Hoseki, J.; Ushioda, R.; Nagata, K., Mechanism and components of endoplasmic reticulum-associated degradation. *J Biochem* **2010,** *147* (1), 19-25.

18. Kollmann, K.; Pohl, S.; Marschner, K.; Encarnacao, M.; Sakwa, I.; Tiede, S.; Poorthuis, B. J.; Lubke, T.; Muller-Loennies, S.; Storch, S.; Braulke, T., Mannose phosphorylation in health and disease. *Eur J Cell Biol* **2010,** *89* (1), 117-23.

19. Haltiwanger, R. S.; Lowe, J. B., Role of glycosylation in development. *Annu Rev Biochem* **2004,** *73*, 491-537.

20. Cao, Q.; Yu, Q.; Liu, Y.; Chen, Z.; Li, L., Signature-Ion-Triggered Mass Spectrometry Approach Enabled Discovery of N- and O-Linked Glycosylated Neuropeptides in the Crustacean Nervous System. *Journal of Proteome Research* **2020,** *19* (2), 634-643.

21. Hongfeng Yin, K. K., The fundamental aspects and applications of

Agilent HPLC-Chip. *J. Sep. Sci.* **2007,** *30*, 1427 – 1434.

22. Hunt, D. F.; Yates, J. R., 3rd; Shabanowitz, J.; Winston, S.; Hauer, C. R., Protein sequencing by tandem mass spectrometry. *Proc Natl Acad Sci U S A* **1986,** *83* (17), 6233-7.

23. Horvatovich, P.; Govorukhina, N. I.; Reijmers, T. H.; van der Zee, A. G.; Suits, F.; Bischoff, R., Chip-LC-MS for label-free profiling of human serum. *Electrophoresis* **2007,** *28* (23), 4493-505.

24. Rademaker, G. J.; Pergantis, S. A.; Blok-Tip, L.; Langridge, J. I.; Kleen, A.; Thomas-Oates, J. E., Mass spectrometric determination of the sites of O-glycan attachment with low picomolar sensitivity. *Anal Biochem* **1998,** *257* (2), 149-60.

25. Chai, W.; Feizi, T.; Yuen, C. T.; Lawson, A. M., Nonreductive release of O-linked oligosaccharides from mucin glycoproteins for structure/function assignments as neoglycolipids: application in the detection of novel ligands for E-selectin. *Glycobiology* **1997,** *7* (6), 861-72.

26. Steen, H.; Mann, M., A new derivatization strategy for the analysis of phosphopeptides by precursor ion scanning in positive ion mode. *Journal of the American Society for Mass Spectrometry* **2002,** *13* (8), 996-1003.

27. Liu, Y.; Cao, Q.; Li, L., Isolation and characterization of glycosylated neuropeptides. *Methods Enzymol* **2019,** *626*, 147-202.

28. Call, C. Identification of neuropeptides using liquid chromatography/mass spectrometry:

An evaluation of high pH fractionation prior to analysis.

. Bowdoin College, Brunswick, ME, 2019.

29. Rademaker, G. J. S., A.; Pergantis, L.; Blok-Tip; Langridge J.E.; Kleen, A. a. T.-O., J.E., Mass Spectrometric Determination of the Sites of O-Glycan

Attachment with Low Picomolar Sensitivity. *Analytical Biochemistry* **1998,** *257*, 149–160.

30. Klemm, C.; Schröder, S.; Glückmann, M.; Beyermann, M.; Krause, E., Derivatization of phosphorylated peptides with S- and N-nucleophiles for enhanced ionization efficiency in matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Communications in Mass Spectrometry* **2004,** *18* (22), 2697-2705.

31. Raulfs, M. D.; Breci, L.; Bernier, M.; Hamdy, O. M.; Janiga, A.; Wysocki, V.; Poutsma, J. C., Investigations of the mechanism of the "proline effect" in tandem mass spectrometry experiments: the "pipecolic acid effect". *J Am Soc Mass Spectrom* **2014,** *25* (10), 1705-15.

32. Furukawa, J.; Fujitani, N.; Araki, K.; Takegawa, Y.; Kodama, K.; Shinohara, Y., A versatile method for analysis of serine/threonine posttranslational modifications by beta-elimination in the presence of pyrazolone analogues. *Anal Chem* **2011,** *83* (23), 9060-7.