

**STRUCTURE CHARACTERIZATION OF A PEROXIDASE FROM THE  
WINDMILL PALM TREE *TRACHYCARPUS FORTUNEI***

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE  
UNIVERSITY OF HAWAI'I AT MĀNOA IN PARTIAL FULFILLMENT OF THE  
REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

MOLECULAR BIOSCIENCES AND BIOENGINEERING

MAY 2015

By

Margaret R. Baker

Dissertation Committee:

Qing X. Li, Chairperson

Jon-Paul Bingham

Dulal Borthakur

Ho Leung Ng

Monika Ward

Keywords: palm tree, peroxidase, *N*-glycosylation, mass spectrometry

© 2015, Margaret R. Baker

## **DEDICATION**

I dedicate this work to my husband, Adam, a constant source of support and encouragement.

I also dedicate it to my sister, Lydia and my nephews, Morgan and Hunter. It was difficult to be so far away from you three.

## ACKNOWLEDGEMENTS

I would like to thank the members of the Li Laboratory for support, discussions, and advice. I will particularly miss lunchtime with the lab! I appreciate the fruitful collaborations with the Bingham Laboratory. Dr. Ivan Sakharov graciously provided the purified windmill palm tree peroxidase, but more than that, he has been a wonderful mentor by discussing peroxidases, and giving invaluable comments on my manuscripts. Dr. David Tabb and his laboratory were very gracious in showing me the ropes in bioinformatics. Thank you to Mr. Sam Fu for consultation about the method for solid-phase reduction of disulfide bonds and for the delicious lunches at McDonald's.

I want to acknowledge Dr. Li and my committee members for their mentorship and guidance throughout this process.

Undoubtedly, I could not have done this work without the support and encouragement of my family- my mom, dad, and step-dad; brother, brother-in-law, and sisters; and nephews in Texas. Many thanks to my parents-in-law in Hawaii, who truly make me feel like I'm their daughter. Thanks to my husband, Adam, for always encouraging me and reading my manuscripts. He is a great writer and a brilliant woodworker. Also, Bei Zhang, for running with me- a much needed release of stress! Also, for our discussions on transient expression in tobacco plants- hopefully we will have time to collaborate in the future.

I was awarded a fellowship from the USDA National Institute of Food and Agriculture. This project was supported in part by the Agriculture and Food Research Initiative Competitive Grant No. 2012-67011-19671 from the USDA National Institute of Food and Agriculture.

## ABSTRACT

Plant secretory peroxidases are important commercial enzymes and play a central role in plant stress responses. The focus of this work is a unique peroxidase from the leaves of a cold tolerant palm, *Trachycarpus fortunei*, (i.e., windmill palm tree). Like other palm tree peroxidases, windmill palm tree peroxidase is stable at high temperatures and in the presence of denaturants. It is distinct from other palm tree peroxidases in its substrate specificity. The amino acid sequence and glycosylation had not been determined. Because glycosylation is known to play a critical role in plant peroxidase stability and activity, this knowledge is essential for structure activity studies, selection of an expression system for enzyme production, and engineering the enzyme. Glycosylation is a complex modification that is difficult to study; furthermore, there is a lack of analytical tools for characterizing plant glycoproteins. The complete amino acid sequence was determined using cDNA sequencing and biological mass spectrometry. The mature amino acid sequence is 306 residues in length. The presence of a C-terminal signal peptide predicts vacuolar targeting of the enzyme. Native windmill palm tree peroxidase was analyzed at the glycopeptide level to give a qualitative and quantitative assessment of glycosylation at each site. Windmill palm tree peroxidase has 13 sites for N-linked glycosylation, 2 of which are unique. Each site is at least partially occupied by a glycan. Major glycans are paucimannosidic, which supports the assignment of windmill palm tree peroxidase as a vacuolar peroxidase. To carry-out this work, a workflow for analyzing the glycopeptide mass spectrometry data was developed. Included in the workflow are novel tools for glycan database construction, pGlycoFilter, and validation of glycopeptide assignment, gPSMvalidator. New analytical methods are needed for the emerging field of plant glycoproteomics. The novel

methods developed in this dissertation will be useful for the study of other important plant glycoproteins. This knowledge can be used to study the roles of glycosylation in this exceptionally stable and unique palm peroxidase.

## TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	iv
ABSTRACT.....	v
LIST OF TABLES.....	ix
LIST OF FIGURES.....	x
LIST OF ABBREVIATIONS.....	xi
CHAPTER 1. INTRODUCTION.....	1
BACKGROUND.....	1
PURPOSE.....	2
OBJECTIVES.....	3
RATIONALE.....	3
CHAPTER 2. AMINO ACID SEQUENCE OF ANIONIC PEROXIDASE FROM THE WINDMILL PALM TREE <i>TRACHYCARPUS FORTUNEI</i> .....	7
INTRODUCTION.....	7
MATERIALS AND METHODS.....	8
RESULTS AND DISCUSSION.....	12
CHAPTER 3. SITE-SPECIFIC <i>N</i> -GLYCOSYLATION AND MICROHETEROGENEITY OF A PEROXIDASE FROM THE WINDMILL PALM TREE <i>TRACHYCARPUS FORTUNEI</i> .....	24
INTRODUCTION.....	24
MATERIALS AND METHODS.....	25
SUPPLEMENTARY DATA.....	28
RESULTS.....	28
DISCUSSION.....	33
CHAPTER 4. CONCLUSIONS AND FUTURE PERSPECTIVES.....	57
PURPOSE.....	57
OBJECTIVES.....	57
FUTURE WORK.....	60
FUTURE PERSPECTIVES.....	61

APPENDIX A: WORKFLOW FOR ANALYSIS OF PLANT GLYCOPEPTIDE MASS SPECTROMETRY DATA.....	62
APPENDIX B: CONFIGURATION FILES FOR PROTEOMICS-BASED PROTEIN AND GLYCOPEPTIDE IDENTIFICATION.....	67
APPENDIX C: PLANT N-GLYCAN BIOSYNTHESIS PATHWAY .....	72
APPENDIX D. DESCRIPTION OF SUPPLEMENTARY DATA .....	74
APPENDIX E. GUANIDINATION OF TRYPTIC PEPTIDES WITHOUT DESALTING FOR MATRIX-ASSISTED LASER DESORPTION/IONIZATION-TIME-OF-FLIGHT MASS SPECTROMETRY ANALYSIS .....	76
APPENDIX F. SYNTHESIS OF AN IBERIOTOXIN DERIVATIVE BY CHEMICAL LIGATION: A METHOD FOR IMPROVED YIELDS OF CYSTEINE-RICH SCORPION TOXIN PEPTIDES.....	85
APPENDIX G. CONE SNAIL MILKED VENOM DYNAMICS – A QUANTITATIVE STUDY OF <i>CONUS PURPURASCENS</i> .....	95
APPENDIX H. ANALYSIS OF A CONE SNAIL’S KILLER COCKTAIL – THE MILKED VENOM OF <i>CONUS GEOGRAPHUS</i> .....	108
APPENDIX I. A “CONOVENOMIC” ANALYSIS OF THE MILKED VENOM FROM THE MOLLUSK-HUNTING CONE SNAIL <i>CONUS TEXTILE</i> - THE PHARMACOLOGICAL IMPORTANCE OF POST-TRANSLATIONAL MODIFICATIONS .....	114
REFERENCES .....	129

## LIST OF TABLES

Table 2.1. Substrate reactivity as measured by $k_{app}$ ( $M^{-1}s^{-1}$ ).....	17
Table 2.2 Residues of known and potential importance for protein-substrate interactions.....	18
Table 3.1. Glycans observed on WPTP .....	36
Table 3.2. Peptide backbone features and observed fragment ion types .....	38
Table 3.3. Microheterogeneity of glycosylation of WPTP.....	39
Table 3.4. Relative abundance of major glycoforms of WPTP .....	40

## LIST OF FIGURES

Figure 2.1. MALDI-TOF MS of WPTP. ....	19
Figure 2.2. MALDI-top-down sequencing of WPTP. ....	20
Figure 2.3. MALDI-TOF/TOF MS/MS of WPTP tryptic peptides. ....	21
Figure 2.4. cDNA sequencing of WPTP.....	22
Figure 2.5. Alignment of WPTP and RPTP.....	23
Figure 3.1. Steps of the data analysis workflow .....	49
Figure 3.2. Total ion chromatograms of digested WPTP samples. ....	50
Figure 3.3. HCD fragmentation spectrum of a WPTP glycopeptide. ....	51
Figure 3.4. Oxonium ions in a WPTP glycopeptide HCD fragmentation spectrum.....	52
Figure 3.5. Cross-ring fragmentation of a WPTP glycopeptide.....	53
Figure 3.6. Extensive peptide backbone fragmentation of a WPTP glycopeptide.....	54
Figure 3.7. Extensive glycan fragmentation of a WPTP glycopeptide.....	55
Figure 3.8. Proposed structures for the 5 major glycans observed on WPTP. ....	56

## LIST OF ABBREVIATIONS

ABTS, 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulphonic acid); ACN, acetonitrile; AMBIC, ammonium bicarbonate; ATP A2, anionic *Arabidopsis thaliana* peroxidase 2; CEP, *Chamaerops excelsa* peroxidase; DHB, 2,5-dihydroxybenzoic acid; FA, ferulic acid; Fuc, fucose; GlcNAc, *N*-acetylglucosamine; HCD, higher-energy C-trap dissociation; HRP, horseradish peroxidase; IS, ion source; MALDI, Matrix Assisted Laser Desorption/Ionization; Man, mannose; MES, 2-(*N*-morpholino)ethanesulfonic acid; OST, oligosaccharyltransferase; PDB, protein databank; PIE, pulsed ion extraction; PNP, peanut peroxidase; RPTP, royal palm tree peroxidase; SBP, soybean peroxidase; TDS, top-down sequencing; TFA, trifluoroacetic acid; TOF, Time of Flight; WPTP, windmill palm tree peroxidase; Xyl, xylose

## CHAPTER 1. INTRODUCTION

### BACKGROUND

***Peroxidases are important enzymes.*** Plant secretory peroxidases are stress responsive enzymes and are central to a plant's ability to adapt and survive (Zipor and Oren-Shamir 2013). There can be as many as 70 different isoforms of peroxidases in a single plant species (Welinder et al., 2002). Peroxidases are involved in myriad processes including cell wall synthesis and responses to abiotic and biotic stresses, such as synthesis and degradation of UV absorbing compounds and synthesis of anti-microbial compounds (Takahama 2004; Zipor and Oren-Shamir 2013). The precise *in vivo* mechanism of the different peroxidase isoforms is unclear for the most part. A recent study identified a single vacuolar peroxidase involved in pigment metabolism (Zipor et al., 2015). An additional level of complexity stems from the fact that they are glycoproteins. Glycoproteins generally exist as a population of glycosylation variants (i.e., glycoforms), and this is true for plant secretory peroxidases (Yang et al., 1996; Zhang et al., 2004; Gray and Montgomery 2006). The role of plant peroxidase glycoforms remains obscure.

There are many biotechnological uses for plant peroxidases making them highly valued commercial enzymes (Regaldo et al., 2004). The major use is in biosensors. A biosensor is a generic term for an analytical device that uses a biological component, such as an enzyme, to 'sense' or detect a target substance. Sensitive and quantitative detection of a broad range of organic molecules have made peroxidase-based biosensors useful in agricultural, biomedical, and environmental diagnostics (Farré et al., 2007). Peroxidases can be used for 'green' environmentally safe synthesis of useful polymers and nano-materials (Sakharov et al., 2003). They can also be used for bioremediation of industrial wastewater containing toxic pollutants such as phenols and cresols (Klibanov et al., 1983). Commercially available peroxidase is mostly limited to horseradish peroxidase (HRP), which is isolated and purified from the roots of the horse-radish plant *Armoracia rusticana*. Peroxidases expressed in alternative expression systems such as the yeast *Pichia pastoris* have instability problems associated with non-native glycosylation (Capone et al., 2014; Krainer et al., 2014).

***Palm tree peroxidases are an alternative to HRP.*** Anionic peroxidases purified from palm tree leaves possess high stability and have excellent catalytic properties (Sakharov 2004). Palm peroxidase-based biosensors were more stable and could tolerate higher concentrations of H<sub>2</sub>O<sub>2</sub> (Alpeeva et al., 2005). The exceptional stability of palm peroxidases at acidic pH met the requirement for synthesis of conductive and chiral polymers under mild conditions (Sakharov et al., 2003; Caramyshev et al., 2007).

***Windmill palm tree peroxidase.*** The windmill palm tree *Trachycarpus fortunei* is the source of windmill palm tree peroxidase (WPTP) (Caramyshev et al., 2006). Like other palm peroxidases, WPTP is unusually stable and has uncommon substrate specificity, making it attractive for industrial applications. WPTP retained catalytic activity after incubation for 1 hour at up to 80 °C, pH 7.5, but under acidic conditions had lower stability. In high concentrations of the denaturant guanidine (up to 4 M), WPTP maintained 100% activity toward *o*-dianisidine. Interestingly, in just 1 M guanidine, there was 70% inactivation when 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulphonic acid) (ABTS) was the substrate.

WPTP had a heme group like other peroxidases, and it uniquely contained 6.5 Ca<sup>2+</sup> ions per protein, whereas most plant peroxidases, including HRP, contain only 2 Ca<sup>2+</sup> (Gajhede et al., 1997). WPTP is an anionic peroxidase with a pI of 3.5. Its molecular weight, as measured by SDS-PAGE was 50 kDa. Class III peroxidases are typically 300 amino acids in length with an estimated polypeptide molecular weight (MW) of 32 kDa (Veitch 2004); therefore, WPTP is probably highly glycosylated.

## **PURPOSE**

The purpose of this study is to characterize the structure of WPTP, including the amino acid sequence and glycosylation. The methods developed herein will allow for structure determination of WPTP and will contribute methodologies needed for study of other plant glycoproteins.

## OBJECTIVES

1. Determine the amino acid sequence of windmill palm tree peroxidase
2. Develop tools and a workflow for analysis of windmill palm tree peroxidase glycopeptide mass spectrometry data
3. Determine the site-specific glycosylation and microheterogeneity of windmill palm tree peroxidase

## RATIONALE

**Objective 1.** The amino acid sequence of a protein gives the primary information about the protein and is requisite to studying the structure further. The first 20 amino acids were determined by Caramyshev et al. (2006) using Edman Degradation. The sequence was determined to be DLQIGFYXQSXPSAESLVXQ; where each letter corresponds to an amino acid residue and X corresponds to an indeterminable residue. This reveals the main limitation of this method- amino acid residues containing modifications, including post-translational modifications, cannot be determined. We decided to take an approach that combined gene cloning and mass spectrometry. These two methods provide complementary information as well as confirmation of the results. These data would give the necessary information for gene expression and would also enable selection of a model protein with which WPTP could be compared. Objective 1 is discussed in [Chapter 2](#).

**Objective 2.** Mass spectrometry was selected as the main analytical tool for determining the glycosylation of WPTP. Characterization of glycosylation involves identifying the glycans attached to each *N*-glycosylation site and their relative abundance. A glycan is defined as a compound containing monosaccharides joined by glycosidic bonds. Analysis of the mass spectrometry data was the biggest hurdle encountered in the course of this dissertation. Most of the published studies did not give a detailed description of how the data in the study were analyzed; a problem noted in a recent review (Desaire 2013). Another notable challenge was that most of the available software were designed for analysis of mammalian protein glycosylation.

Software were often linked to mammalian glycan databases. Mammalian glycans have significantly different structures from plant glycans.

The analytical challenges associated with studying glycoproteins result from the fact that the peptide is attached a branched polymer, the structure of which is assembled in a non-template driven process. The process depends on several factors including protein-protein interactions between the target and glycosylation enzymes, availability of the sugar nucleotide building blocks, number and distribution of the glycosylation enzymes, and protein residence time in the Golgi apparatus and its final destination (i.e., the apoplast or vacuole for plant glycoproteins). The interplay of these various factors ultimately leads to a heterogeneous population of glycans attached to each glycosylation site.

The process of *N*-glycosylation begins with the targeting of proteins containing an *N*-terminal signal peptide to the secretory pathway (Stanley et al., 2009). This signals the nascent polypeptide to be introduced into the lumen of the endoplasmic reticulum (ER). If the sequence Asn-Xxx-Ser/Thr, where Xxx is not Pro, is encountered, it may be recognized by the oligosaccharyltransferase (OST) complex and have a bulky glycan attached. Local amino acid sequence will affect the efficiency of glycosylation at that site (Petrescu et al., 2004). Subsequent secretion to the Golgi puts the glycoprotein in contact with glycosidases and glycosyltransferases which participate in a process called ‘maturation.’ Maturation enzymes are also not 100% efficient. Glycans are then trimmed after being exported to the extracellular matrix or the vacuole (Rayon et al., 1998). Therefore, a single glycoprotein exists as a population of different ‘glycoforms’, i.e., the same polypeptide with different glycosylation. The presence of different glycans at one site is termed microheterogeneity. This can be a mechanism that evolved to add a level of fine-tuning that can be used to the advantage of the organism. A detailed description of the plant *N*-glycan biosynthesis pathway is presented in [Appendix D](#).

The ER portion of the *N*-glycan biosynthesis pathway is conserved among plants and animals, but significantly differs in the subsequent maturation steps carried out in the Golgi and post-Golgi (Stanley et al., 2009; Gomord et al., 2010). Plant glycoprotein *N*-linked glycans have several distinct characteristics: 1) fewer branches, 2) absence of sialylation, and 3) presence of

xylosylation and  $\alpha(1\rightarrow3)$ -fucosylation as opposed to  $\alpha(1\rightarrow6)$ -fucosylation (Johnson and Chrispeels, 1987; Tezuka et al., 1992).

Plant glycoproteomics is a growing field and there is a demand for analytical methods (Strasser 2014). Some workflows/reviews have been published (Song et al., 2011; Zhang et al., 2011). One of the main problems with software for glycopeptide mass spectrometry data interpretation is the fact that most were developed with mammalian glycans in mind. Scoring in some cases depends on mammalian glycan-specific fragmentation. Importantly, glycan databases and prediction algorithms did not contain plant glycan-specific features. The tools and workflow developed for analysis of WPTP glycosylation data are presented in [Chapter 3](#).

**Objective 3.** Glycosylation may play an important role in WPTP. Partial deglycosylation of peroxidases results in decreased stability (Tams and Welinder, 1998; Lige et al., 2001) and lower rates of catalysis (Lige et al., 2001; Nielsen et al., 2001). A decrease in specific activity of a peroxidase from avocado was correlated with the loss of ability to chelate calcium ions (Sánchez-Romero et al., 1994). Considering that WPTP contains 6.5  $\text{Ca}^{2+}$  per peroxidase molecule, this may be of considerable importance. Further investigations into the molecular basis of glycan contribution are needed.

A major motivation of this dissertation is that palm peroxidases would be good models for studying roles of glycosylation. The X-ray crystal structure of native, highly glycosylated royal palm tree peroxidase (RPTP) was recently solved (Watanabe et al., 2010). At the time of proof-reading this dissertation, the X-ray crystal structure of native *Chamerops excelsa* peroxidase (CEP) was reported (Bernardes et al., 2015). Probing the differences between WPTP, CEP, and RPTP will give insights into specific stabilizing roles of glycosylation. The microheterogeneity and site-specific glycosylation of WPTP are presented in [Chapter 3](#).

**Appendices.** The workflow developed for objective 2 is described in detail in [Appendix A](#). The configurations for proteomics software are in [Appendix B](#). A thorough review of the known plant glycan biosynthesis pathway is presented in [Appendix C](#). It was necessary for development of a novel plant glycan database generation tool, pGlycoFilter. A description of all of the supplementary data is given in [Appendix D](#).

Additionally, I developed a method called ‘freebase guanidination.’ It was originally intended to be used in structure characterization of WPTP, but in the end was not used. Accordingly, the reprint of this publication is attached in [Appendix E](#). During the course of learning mass spectrometry, I collaborated with another laboratory and studied cone snail peptides. The reprints of these publications are in [Appendices F-I](#).

## CHAPTER 2. AMINO ACID SEQUENCE OF ANIONIC PEROXIDASE FROM THE WINDMILL PALM TREE *TRACHYCARPUS FORTUNEI*<sup>1</sup>

### INTRODUCTION

Plant secretory peroxidases (class III peroxidases; EC 1.11.1.7) are ubiquitous in nature. They are extracellular or vacuolar glycoproteins and catalyze redox reactions that facilitate a myriad of biological processes including cell wall synthesis and response to abiotic and biotic stresses (Zipor and Oren-Shamir 2013). The peroxidase-catalyzed reaction occurs in 3 steps according to the “ping-pong” mechanism. First, the resting state peroxidase (E) is oxidized by hydrogen peroxide (H<sub>2</sub>O<sub>2</sub>). Next, two back-to-back single electron transfers to an aromatic reducing substrate (AH<sub>2</sub>) proceed. The reaction ends with the return of peroxidase to the resting state and generation of water and radical products (AH<sup>•</sup>) (Veitch 2004). The reaction can be summarized as 
$$\text{H}_2\text{O}_2 + 2 \text{AH}_2 \xrightarrow{E} 2 \text{H}_2\text{O} + 2 \text{AH}^\bullet$$

Peroxidases are an important component of biosensors and immunochemical kits due to their sensitive and quantitative detection of H<sub>2</sub>O<sub>2</sub> (Gaspar et al., 2000; Chen et al., 2013) and persistent organic compounds such as pesticides (Farré et al., 2007). Detection of polyphenols and flavonoids in fruits and vegetables is used to assess the nutritive quality of foodstuffs (Litescu et al., 2010). Some enzyme immunoassays have been developed using horseradish peroxidase (HRP) to detect food allergens (Zhang et al., 2014) and metal ions that may contaminate food and agricultural soil (Zhao et al., 2011).

Anionic peroxidases purified from palm tree leaves possess extremely high stability (Sakharov 2004). Moreover, these enzymes showed distinct substrate specificity compared with other plant peroxidases. Their unique properties have allowed for development of novel and improved applications. Palm peroxidase-based biosensors were more stable and could tolerate higher concentrations of H<sub>2</sub>O<sub>2</sub> (Alpeeva et al., 2005). The exceptional stability of palm peroxidases at acidic pH met the requirement for synthesis of conductive and chiral polymers

---

<sup>1</sup> Reproduced in part with permission from Baker, M. R.; Zhao, H.; Sakharov, I. Y.; Li, Q. X. Amino Acid Sequence of Anionic Peroxidase from the Windmill Palm Tree *Trachycarpus fortunei*. *Journal of Agricultural and Food Chemistry* **2014**, *62*, 11941–11948. Copyright © 2014, rights managed by the American Chemical Society.

under environmentally safe conditions (Sakharov et al., 2003; Caramyshev et al., 2005; Caramyshev et al., 2007).

Some plant peroxidases, particularly HRP are well studied, however, relatively little is known about the structure of palm peroxidases. The X-ray crystal structure of native, highly glycosylated royal palm tree peroxidase (RPTP) was recently solved (Watanabe et al., 2010). However, it is a unique work and structural data for other palm tree peroxidases were practically absent in the literature, except the sequence of 20 amino acids of the *N*-terminus of windmill palm tree peroxidase (WPTP, *Trachycarpus fortunei*) (Caramyshev et al., 2006). Also, it is known that WPTP uniquely contains 6.5 calcium cations per molecule, whereas most plant peroxidases, including HRP and RPTP, contain only 2 Ca<sup>2+</sup> (Watanabe et al., 2010; Gajhede et al., 1997).

To fill in the knowledge gap about the structures of palm peroxidases, herein we describe the complete amino acid sequence of WPTP and provide evidence of its modification with *N*-linked glycans. The complete amino acid sequence and glycosylation were determined by MALDI-top-down sequencing of native WPTP, MALDI-TOF/TOF MS/MS of WPTP tryptic peptides, and cDNA sequencing. The propeptide of WPTP contained *N*- and *C*-terminal signal sequences. Mature WPTP was 306 amino acids in length and its carbohydrate content was in the range of 21 to 29%. The results can be used to guide engineering of WPTP and its novel applications.

## MATERIALS AND METHODS

**Materials.** WPTP was isolated and purified from *T. fortunei* leaves as previously described (Caramyshev et al., 2006). Trypsin (MS grade) was purchased from Promega (Madison, WI). Bovine serum albumin (BSA; 98% purity), used as a protein standard, was from Sigma (St. Louis, MO). The MALDI matrix 2,5-dihydroxybenzoic acid (DHB), peptide calibration standard II, and protein calibration standard II were from Bruker (Billerica, MA). C<sub>4</sub> ziptips were from EMD Millipore (Darmstadt, Germany). Other reagents, including HPLC grade acetonitrile (ACN), trifluoroacetic acid (TFA), and ammonium bicarbonate (AMBIC) were from Fisher Scientific (Waltham, MA). Water was purified on a Milli-Q Advantage A10 system

(EMD Millipore). Dithiothreitol was from Acros Organics (New Jersey, USA) and iodoacetamide was from BioRad (Hercules, CA). For gene cloning, an RNeasy Plant Mini Kit (Qiagen, Venlo, Limburg), a SMARTer RACE cDNA Amplification kit and an Advantage 2 PCR kit from Clontech (Mountain View, CA), a pGEM-T vector (Promega), and *Escherichia coli* strain JM109 were used.

**Molecular mass measurement.** Purified WPTP was dissolved in 0.1% TFA (v/v), mixed 1:1 with a MALDI matrix solution (20 g/L DHB in 50% ACN (v/v) in 0.1% TFA (v/v)), spotted onto a polished steel target plate (Bruker), and allowed to air dry.

The molecular mass was measured with an Ultraflex III MALDI-TOF/TOF mass spectrometer (Bruker) in linear positive mode. In 200 shot increments, 1,500 laser shots were accumulated. The matrix suppression cut-off, using gating, was at 9,500 *m/z*. Pulsed ion extraction (PIE) delay was set to 150 ns. Instrument voltages were at 25 kV (ion source 1 [IS1]), 23.1 kV (IS2), and 6.5 kV (lens). The WPTP molecular mass value was calibrated using the protein mix II calibration standard (Bruker). Spectrum processing, consisting of peak detection (centroid algorithm, peak width 1000 *m/z*) and smoothing (Savitzky-Golay algorithm, 1 cycle at 20 *m/z*), was done in Flex Analysis 3.4 (Bruker).

**Top down sequencing.** Top-down sequencing (TDS) of native WPTP was performed after solid-phase reduction of its disulfide bonds, for which the late Mr. Sam Fu graciously provided consultation. BSA was used as the control and the molecular mass calibrator. For solid-phase disulfide bond reduction, a C<sub>4</sub> ziptip was wetted with 50% ACN (v/v) in 250 mM AMBIC solution several times. The protein sample (25 pmol in 15  $\mu$ L of 100 mM AMBIC) was loaded on the tip equilibrated with 100 mM AMBIC by slowly repipetting at least 10 times. Then, 50 mM dithiothreitol (200  $\mu$ L) was introduced to the tip by aspirating and dispensing the solution for 2 min at 60 °C. The ziptip was then incubated in that solution for 30 min at 60 °C. The tip was washed 3 times with 0.1% TFA (v/v). The protein was eluted with 5  $\mu$ L of a MALDI matrix solution (40 g/L DHB in 70% ACN (v/v) in 0.1% TFA (v/v)) and 0.25  $\mu$ L were spotted directly on the target plate and allowed to air dry.

TDS spectra were acquired with an Ultraflex III MALDI-TOF/TOF mass spectrometer in reflector positive mode. In 200 shot increments, 19,000 laser shots were accumulated. The

matrix suppression cut-off, using deflection, was at 900  $m/z$ . PIE delay was set to 40 ns. Instrument voltages were at 25 kV (IS1), 21.4 kV (IS2), 9.7 kV (lens), 26.3 kV (reflector 1), and 13.8 kV (reflector 2). Spectrum processing, consisting of peak detection (SNAP algorithm, peak width 0.75  $m/z$ ), baseline subtraction (TopHat algorithm), and smoothing (Savitzky-Golay algorithm, 5 cycles at 1.5  $m/z$ ), as well as peak annotation was carried out in Flex Analysis 3.4. Peaks were assigned with a mass tolerance of 300 ppm. Interpretation of the ions comprising the near C-terminus was aided by the characteristic  $m/z$  difference of 15 ( $\Delta m/z$ ) between y- and z+2-ion series. Additional annotation was done with BioTools (Bruker) and Sequence Editor (Bruker).

**Bottom-up sequencing.** WPTP (10  $\mu\text{g}$ , 1  $\mu\text{g}/\mu\text{L}$  in 25 mM AMBIC) was reduced with 50 mM dithiothreitol (20  $\mu\text{L}$ ) for 30 min at 60 °C, alkylated with 100 mM iodoacetamide (25  $\mu\text{L}$ ) for 45 min at room temperature in the dark. The alkylation reaction was quenched by addition of 1  $\mu\text{L}$  of 143 mM dithiothreitol. Reduced and alkylated WPTP was diluted with water, and then digested in solution with 0.5  $\mu\text{g}$  of trypsin (50  $\mu\text{L}$ ) at 37 °C for 12 h. Peptides were dried via SpeedVac and resuspended in 5% ACN (v/v) in 0.1% TFA (v/v). The peptide solution was mixed 1:1 with a MALDI matrix solution (20 g/L DHB in 50% ACN (v/v) in 0.1% TFA (v/v)) and then spotted on a target plate.

Mass spectra of the peptides were acquired with an Ultraflex III MALDI-TOF/TOF mass spectrometer in reflector positive mode. In 200 shot increments, 600 shots were accumulated. The matrix suppression cut-off, using deflection, was at 850  $m/z$ . PIE delay was set to 0 ns. Instrument voltages were at 25 kV (IS1), 21.45 kV (IS2), 9.5 kV (lens), 26.3 kV (reflector 1), and 13.8 kV (reflector 2). External calibration was performed using peptide mix II calibration standard (Bruker). Spectrum processing, consisting of peak detection (SNAP algorithm, peak width 0.75  $m/z$ ) and smoothing (Savitzky-Golay algorithm, 1 cycle at 0.2  $m/z$ ), was done in Flex Analysis 3.4.

Several peaks in the mass spectrum were selected for fragmentation in LIFT mode (Suckau et al., 2003). Several thousand laser shots were accumulated in 200 shot increments. PIE delay was set to 0 ns. Instrument voltages were at 8 kV (IS1), 7.2 kV (IS2), 3.6 kV (lens), 29.5 kV (reflector 1), and 13.85 kV (reflector 2). Spectrum processing, consisting of peak detection

(SNAP algorithm, peak width 0.75  $m/z$ ) and smoothing (Savitzky-Golay algorithm, 4 cycles at 0.15  $m/z$ ) as well as peak annotation, was carried out in Flex Analysis 3.4. Additional annotation was done with BioTools and Sequence Editor with a mass tolerance of 0.3  $m/z$  for peptides and 1  $m/z$  for glycopeptides. Fragmentation spectra were interpreted and used for primer design in the gene cloning experiment.

**Cloning and sequencing of the cDNA.** Total RNA was isolated from 0.1 g of *T. fortunei* leaves using an RNeasy Plant Mini Kit (Qiagen). The first-strand cDNA was synthesized from the isolated RNA using a SMARTer RACE cDNA Amplification kit (Clontech). The WPTP gene was cloned in 2 steps. In the first step, the forward degenerate primers (5'-GAYCTNCARATHGGNTT-3' and 5'-GAYTTRCARATHGGNTT-3') targeted the previously determined amino acid sequence of the N-terminus of WPTP (DLQGIFY) (Caramyshev et al., 2006). The reverse primer, UPM, was provided with the same kit. PCR amplification was performed using an Advantage 2 PCR kit (Clontech). The PCR protocol consisted of an initial denaturation step at 94 °C (2 min); which was followed by 5 cycles of denaturation at 94 °C (30 s), annealing at 60 °C (30 s), and extension at 68 °C (1 min); then 5 cycles of denaturation at 94 °C (30 s), annealing at 55 °C (30 s), and extension at 68 °C (1 min); then followed by 25 cycles of denaturation at 94 °C (30 s), annealing at 50 °C (30 s), and extension at 68 °C (1 min); and then a final extension for 3 min at 68 °C. The resulting 1,200 bp product was used as a template in the second step. The forward primer (5'-ATGCAYTTYCAYGAYTGYTT-3') was designed from the amino acid sequence MHFHDCF, obtained by tandem mass spectrometry analysis of WPTP tryptic peptides in this study. The PCR protocol was the same as above except that the annealing temperatures were 65 °C, 62 °C, and 60 °C, respectively. The PCR product was cloned into a pGEM-T vector and transformed into *E. coli* strain JM109. Positive transformants were screened by PCR. Plasmids were extracted from positive clones and confirmed by sequencing. For 5'-RACE, the reverse primer, 5'-GGACCTGGTAGGTGATGTTGCCG-3', was designed from the DNA sequence obtained in the previous step. The forward primer, UPM, was from the same kit. The PCR protocol was the same as above except that the annealing temperatures were 65 °C, 62 °C, and 58 °C, respectively. The PCR product was cloned, transformed, screened, and sequenced as before. The complete cDNA sequence of WPTP was obtained by combining the 5'- and 3'-cDNA sequences which

were overlapping with each other. Analysis of the data was conducted with the software Mega5 (Tamura et al., 2011). These data can be accessed at [NCBI](#) accession: KM504969.

## RESULTS AND DISCUSSION

**Determination of the molecular mass of WPTP.** The MALDI-TOF mass spectrum of WPTP indicated the presence of a single glycoprotein with considerable heterogeneity ([Fig. 2.1A](#)). An expanded view of the +2 charged peak revealed that it was composed of many partially resolved peaks, probably corresponding to WPTP glycoforms, *i.e.*, differentially glycosylated WPTP ([Fig. 2.1B](#)). A similar view of the +1 charged peak also showed multiple, partially resolved peaks.

The obtained data allowed for calculation of the MW of WPTP which ranged from 41 to 45 kDa ([Fig. 2.1C](#)). These values are a little lower than the value previously reported for WPTP (50 kDa), which was measured using SDS-PAGE (Caramyshev et al., 2006). Due to decreased binding of SDS to glycans relative to the polypeptide backbone, the migration of a glycoprotein during SDS-PAGE is altered relative to what is expected for a similarly sized non-glycosylated protein (Segrest et al., 1971). Therefore, the previous measurement was likely an over-estimation. Similarly, the first report of RPTP's MW, as measured by SDS-PAGE, was 51 kDa (Sakharov et al., 2001) and the MW as measured by mass spectrometry was 44,596 Da (Watanabe et al., 2010).

**Top-down sequencing.** It is well known that drawbacks of MALDI-TDS are the requirement of a high purity of analyte, low sensitivity, and lack of fragmentation in the presence of disulfide bonds (Resemann et al., 2010). The isolated WPTP used in this study was high purity. To overcome the other problems, we reduced WPTP's disulfide bonds while it was immobilized onto a C<sub>4</sub> ziptip by hydrophobic interactions. This provided a quick and effective way to reduce WPTP's disulfide bonds, remove contaminating salts, and concentrate the reduced protein in one simple step.

MALDI-TDS results in ladder fragmentation of the intact protein leaving glycans attached to the modified Asn residues (Hanisch 2011). Partial amino acid sequence near the *N*-

terminus of WPTP was gained through interpretation of the c-ion series in the MALDI-TDS spectrum ([Fig. 2.2A](#)). The y- and z+2-ions were interpreted as a continuous 30-amino acid residue sequence ([Fig. 2.2A](#)). A protein BLAST search of the NCBI database ([blast.ncbi.nlm.nih.gov](http://blast.ncbi.nlm.nih.gov)) predicted that this sequence originated from a protein in the peroxidase super family. The top hit was RPTP (PDB: 3HDL; E value:  $6e^{-16}$ ). The y ions at  $m/z$  5290.1 and 6575.7 (N\*) (and the corresponding z+2-ions) had a mass difference corresponding to a glycosylated Asn, Asn-GlcNAc<sub>2</sub>Man<sub>3</sub>FucXyl (exact  $\Delta m/z$  1284.5 and accurate  $\Delta m/z$  1285.6). This glycan is the predominant glycan found on HRP (Yang et al., 1996), soybean peroxidase (SBP) (Gray and Montgomery 2006), RPTP (Watanabe et al., 2010), and other plant peroxidases. A scheme of the glycan is depicted in [Fig. 2.2B](#) according to the conventions of the Consortium for Functional Glycomics. All assigned peaks had a mass error within 200 ppm and the root mean square error (RMSError) was 65.32 ppm ([Fig. 2.2C](#)).

**Bottom-up sequencing.** Additional amino acid sequence information was gained through a more traditional “bottom-up” approach. For this, WPTP was digested with trypsin and the resulting tryptic peptides were subjected to MALDI-TOF MS ([Fig. 2.3A](#)). Tandem mass spectrometry was performed to obtain the amino acid sequence of the tryptic peptides of WPTP ([Figs. 2.3B-F](#)). Among these, fragments of the parent ion at  $m/z$  1264.6 could be unambiguously interpreted as M<sup>[ox]</sup>HFHDCFVR ([Fig. 2.3D](#)). Furthermore, this sequence overlapped with an amino acid sequence ladder from the MALDI-TDS experiment. Thus, this amino acid sequence was used for primer design for cDNA cloning and sequencing of WPTP.

Fragments of the parent ion at  $m/z$  2890.3 displayed a characteristic spectrum of a glycopeptide ([Fig. 2.4F](#)). Man and GlcNAc oxonium ions were observed at  $m/z$  163 and 204, respectively. We also observed cross-ring fragmentation of the innermost GlcNAc (<sup>0,2</sup>X). Additionally, the glycan moiety fragmented into a series of Y-ions corresponding to mass shifts indicative of the glycan GlcNAc<sub>2</sub>Man<sub>3</sub>XylFuc ([Fig. 2.2C](#)).

**cDNA sequencing.** The gene for WPTP was cloned and sequenced from mRNA isolated from the leaves of *T. fortunei*. The gene consisted of 5'- and 3'- untranslated regions, which included a poly-A tail, and 1,032 bp encoding a polypeptide of 344 amino acids ([Fig. 2.4](#)). There was an N-terminal signal peptide upstream of the previously sequenced N-terminus of

mature WPTP, which targets class III peroxidases to the secretory pathway (Welinder et al., 2002). Some peroxidases additionally bear a C-terminal signal peptide for vacuolar targeting (Welinder et al., 2002). Insights into *in vivo* processing of the C-terminus of WPTP were obtained by combining information from the MALDI-TDS spectrum of intact WPTP and its cDNA sequence. In a MALDI-TDS spectrum, the ion with the lowest  $m/z$  value in each ion series corresponds to the mass of the N-terminus (for c-ions) or the C-terminus (for y- and z-type ions). Therefore, the y 11-ion (Fig. 2.2) corresponded to the mass of RTN\*CSVVNSAS (exact  $m/z$  2307.9 and accurate  $m/z$  2307.8), where the mass shift for Asn<sup>298</sup> (N\*) in the WPTP sequence corresponded to GlcNAc<sub>2</sub>Man<sub>3</sub>FucXyl. Thus, the vacuolar targeting signal for WPTP was LGDIVMASGHLTEVATS (Fig. 2.4).

Using the same logic for the N-terminus, the c 17-ion at  $m/z$  2878.1 (Fig. 2.2) corresponded to the sequence DLQIGFYN<sup>#</sup>QSCPSAESLVQQAVAAAFAN\*NSGIAPGLIR (exact  $m/z$  3920.0 and accurate  $m/z$  6098.4) (Fig. 2.4), where the mass difference of 2178.4 could be accounted for by adding a glycan to Asn<sup>8</sup> (N#) and one to Asn<sup>28</sup> with the potential sequences GlcNAc<sub>2</sub>Man<sub>2</sub>FucXyl ( $\Delta m/z$  1008.4) and GlcNAc<sub>2</sub>Man<sub>3</sub>FucXyl ( $\Delta m/z$  1170.4), giving an overall mass error of 65.6 ppm.

**Structure overview.** Mature WPTP was composed of a single polypeptide, 306 amino acid residues in length, with an average mass of 32,172 Da. The cDNA sequence showed that there were 13 potential glycosylation sites, identified by the sequence Asn-Xxx-Ser/Thr, where Xxx is any amino acid except Pro. When the mass as measured by MALDI-TOF MS is considered, WPTP has 21 - 29% glycans (Fig. 2.1). This level of complexity made it difficult to assign a specific glycoform structure to each peak in that spectrum. Positions 8, 28, 114, 267, and 298 were glycosylated. A more detailed study of WPTP's glycosylation profile is underway.

Class III peroxidases share a common 3D structure despite a low sequence identity (Veitch 2004; Gajhede et al., 1997; Schuller et al, 1996). For example, RPTP and HRP were only 36% identical, but their C $\alpha$  backbone structures were closely related, having an overall root mean squared deviation of 1.07 Å (Watanabe et al., 2010). A structurally guided alignment revealed that WPTP and RPTP were 88% identical in primary structure and likely were very similar in secondary and tertiary structures (Fig. 2.5).

Structural integrity of plant peroxidases is maintained largely by an extensive hydrogen bonding network that extends above and below the heme-containing active site to distal and proximal  $\text{Ca}^{2+}$  (Smulevich et al., 2005). Additionally, 4 invariant disulfide bridges tie the protein together. These 8 cysteine residues were present in WPTP ([Fig. 2.5](#)). Importantly, the disulfide bridge between Cys<sup>44</sup> and Cys<sup>49</sup> stabilizes the BC loop, which comprises majority of the distal  $\text{Ca}^{2+}$  binding residues ([Fig. 2.5](#)). Distal  $\text{Ca}^{2+}$  was critical for maintaining the proper tertiary structure of the active site (Szigeti et al, 2008). The disulfide bridge between Cys<sup>176</sup> and Cys<sup>208</sup> stabilizes the region between helices F and H, which is important for substrate binding and contains residues necessary for binding proximal  $\text{Ca}^{2+}$ . The additional 4.5  $\text{Ca}^{2+}$  found in the WPTP structure were difficult to account for since only one other peroxidase, anionic peanut peroxidase (Hu et al., 1997), had elevated  $\text{Ca}^{2+}$  content and its 3D structure was still not available.

**Features of potential importance to substrate specificity.** The new WPTP structure information presented in this study allowed for exploration of features of potential importance in palm tree peroxidases. One striking functional difference is substrate reactivity differences between WPTP, RPTP, and HRP ([Table 2.1](#)). Ferulic acid was the better substrate for RPTP and HRP (Sakharov et al., 2001). Moreover, the difference in reactivity between ferulic acid and ABTS was not too different for those 2 peroxidases, *i.e.*, there was a 1.26 fold difference for RPTP and 3.25-fold difference for HRP (Sakharov et al., 2001). The better substrate for WPTP was ABTS and ferulic acid had 20-fold lower reactivity (Caramyshev et al., 2006). Investigation of the 3D structure of RPTP, using Swiss-PdbViewer 4.1.0 (Arnold et al., 2006), revealed amino acid residues and features of known and putative importance to substrate binding ([Table 2.2](#)).

Amino acids involved in substrate binding should be within 12 Å of the heme iron, as indicated by NMR studies of HRP-substrate complexes (Veitch 1995). There were 64 residues within this distance in RPTP ([Fig. 2.5](#)) and 5 of those differed between WPTP and RPTP. Of those 5, the lowest conservation was at position 142. Interactions between residue 142 and conserved Phe<sup>143</sup> resulted in topological features thought to affect interactions with substrates (Veitch 2004; Henriksen et al., 2001). Residue 142 was hydrophobic in HRP, RPTP, and WPTP,

except that it was large in HRP (Phe) and RPTP (Ile) but small in WPTP (Thr). This difference could be a major factor causing the large deficit in WPTP's reactivity toward ferulic acid. The absence of a 3D structure complex between peroxidase and ABTS prevented speculation about features responsible WPTP's superior catalysis of that large substrate.

The RPTP crystal structure contained a potential inhibitor; an MES molecule (Watanabe et al., 2010). Residues within van der Waals radius (4.6 Å) of MES may affect substrate reactivity. Nine residues (Ile<sup>68</sup>, Ser<sup>140</sup>, Pro<sup>141</sup>, Leu<sup>142</sup>, Ser<sup>177</sup>, Ser<sup>178</sup>, Ser<sup>212</sup>, Thr<sup>213</sup>, and Arg<sup>214</sup>) were within that region. Five of the 9 residues did not overlap with the residues that were 12 Å from the heme iron ([Fig. 2.5](#)). The position with the lowest conservation between RPTP and WPTP was at position 177. In WPTP, it was large, hydrophobic, and aromatic Phe, whereas in RPTP, it was small and polar Ser. Moreover, this MES binding site was in the most structurally divergent region between HRP and RPTP ([Fig. 2.5](#)). In HRP, this loop region was pointing away from the active site, whereas in RPTP, it was pointing toward it. This difference in topology brought Arg<sup>214</sup> into hydrogen bonding distance with MES (Watanabe et al., 2010). Further investigation into this distinct region is warranted.

Table 2.1. Substrate reactivity as measured by  $k_{app}$  ( $M^{-1}s^{-1}$ ).

	WPTP <sup>a</sup>	RPTP <sup>b</sup>	HRP <sup>b</sup>
ferulic acid	$3.0 \times 10^6$	$6.3 \times 10^7$	$1.3 \times 10^7$
ABTS <sup>c</sup>	$6.0 \times 10^7$	$5.0 \times 10^7$	$4.0 \times 10^6$

<sup>a</sup>The data for WPTP are from (Caramyshev et al., 2006)

<sup>b</sup>The data for RPTP and HRP are from (Sakharov et al., 2001)

<sup>c</sup>2,2'-Azinobis(3-ethylbenzothiazoline-6-sulfonic acid)

Table 2.2 Residues of known and potential importance for protein-substrate interactions

residue <sup>a</sup>	WPTP	RPTP	predicted effect	evidence
Distance within 12 Å of heme iron				
38	R	R	critical	two H-bonds to ferulic acid (FA) (HRP); <sup>b</sup> Arg <sup>38</sup> to Leu 10 - 100x slower substrate oxidation (HRP) <sup>c</sup>
69	P	P	minor	putative hydrophobic interaction between FA and anionic <i>A. thaliana</i> peroxidase (ATP A2) <sup>d</sup>
138	I	I	minor	putative hydrophobic bond to FA (ATP A2) <sup>d</sup>
139	P	P	major	H-bond to active site H <sub>2</sub> O which is H-bonded to FA (HRP) <sup>b</sup>
140	A	S	minor	Ala: tiny <sup>e</sup> and hydrophobic vs. Ser: small <sup>f</sup> and polar; hydrophobic interaction with FA (HRP: Ala <sup>140</sup> ); <sup>b</sup> H-bond to MES <sup>g</sup> (RPTP) <sup>h</sup>
142	T	L	major	Thr: small <sup>c</sup> and hydrophobic vs. Leu: large, hydrophobic, and aliphatic; hydrophobic interaction with FA (HRP) <sup>b</sup>
Distance within 4.6 Å of MES <sup>d</sup>				
68	A	I	minor	Ile: large, hydrophobic, and aliphatic vs. Ala: tiny <sup>b</sup> and hydrophobic; hydrophobic interaction with FA (HRP: Gly <sup>68</sup> ) <sup>b</sup>
177	F	S	major	Phe: large, hydrophobic, and aromatic vs. Ser: small <sup>c</sup> and polar
212 - 214	STR	STR	major	structurally distinct region (HRP vs. RPTP); H-bond with MES <sup>d</sup> (RPTP: Arg <sup>214</sup> ) <sup>h</sup>

<sup>a</sup>position in the amino acid sequence ([Fig. 2.5](#)); <sup>b</sup>Henriksen et al., 1999; <sup>c</sup>Rodriguez-Lopez et al., 1996; <sup>d</sup>Nielsen et al., 2001; <sup>e</sup>less than 35 Å<sup>3</sup>; <sup>f</sup>less than 60 Å<sup>3</sup>; <sup>g</sup>2-(*N*-morpholino)ethanesulfonic acid; <sup>h</sup>Watanabe et al., 2010.

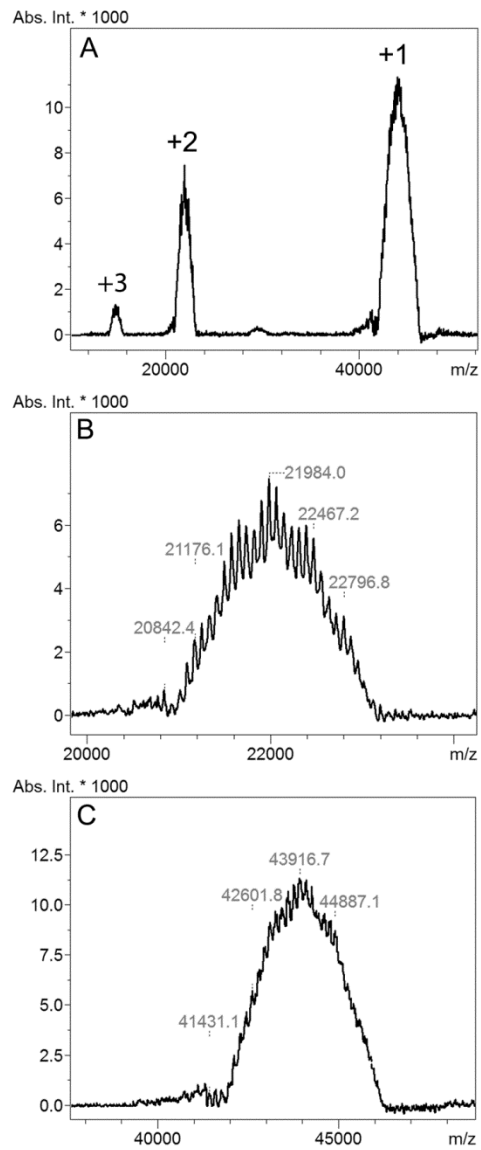


Figure 2.1. MALDI-TOF MS of WPTP. (A) +1, +2, and +3 charged peaks and an expanded view of the (B) +2 charged and (C) +1 charged peak.

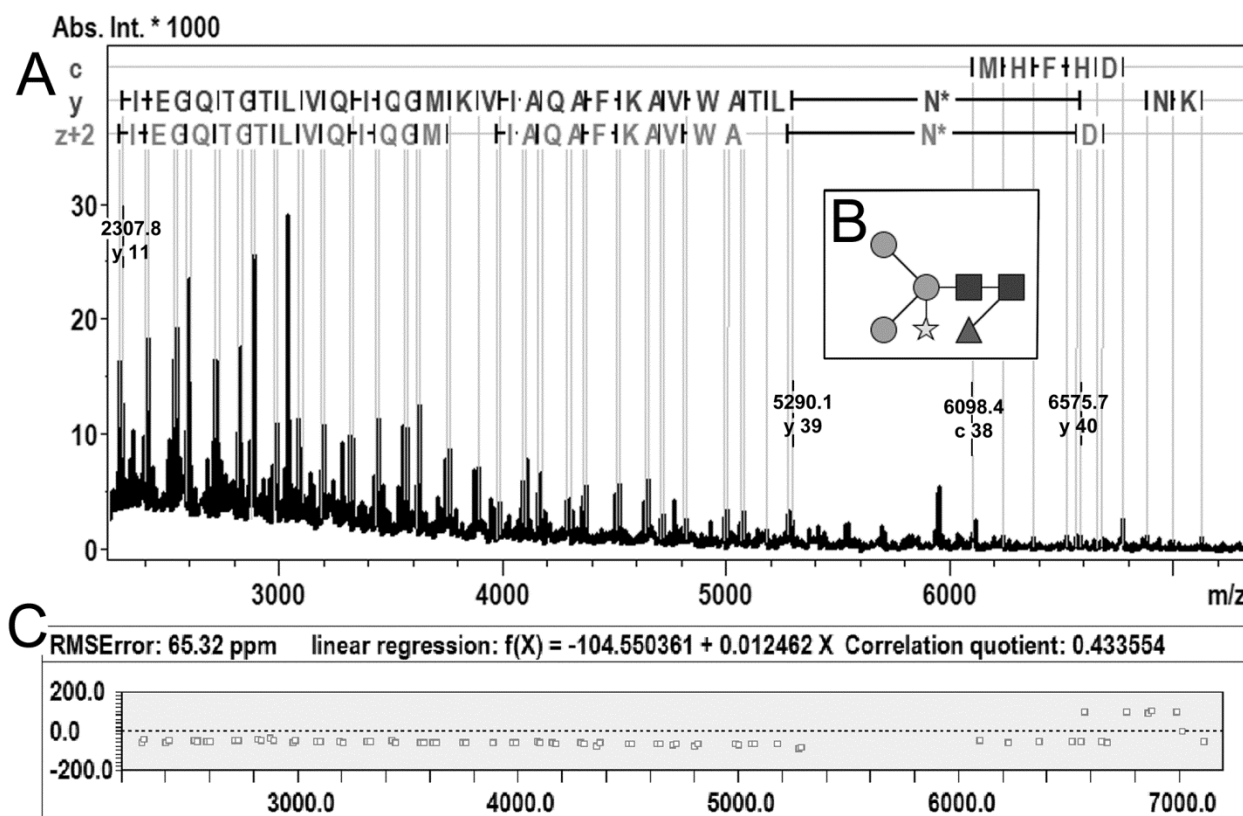


Figure 2.2. MALDI-top-down sequencing of WPTP. The amino acid sequence was interpreted from c-, y- and z+2-ion series. (A)  $\Delta m/z$  between y 39 at  $m/z$  5290.1 and y 40 at 6575.7 corresponded to Asn modified with the glycan  $\text{GlcNAc}_2\text{Man}_3\text{XylFuc}$  (N\*); (B) a cartoon structure of the glycan  $\text{GlcNAc}_2\text{Man}_3\text{XylFuc}$ ; and (C) a plot showing the mass error (ppm) for each matched ion. (■) *N*-acetylglucosamine (GlcNAc); (●) mannose (Man); (▲) fucose (Fuc); (☆) xylose (Xyl).

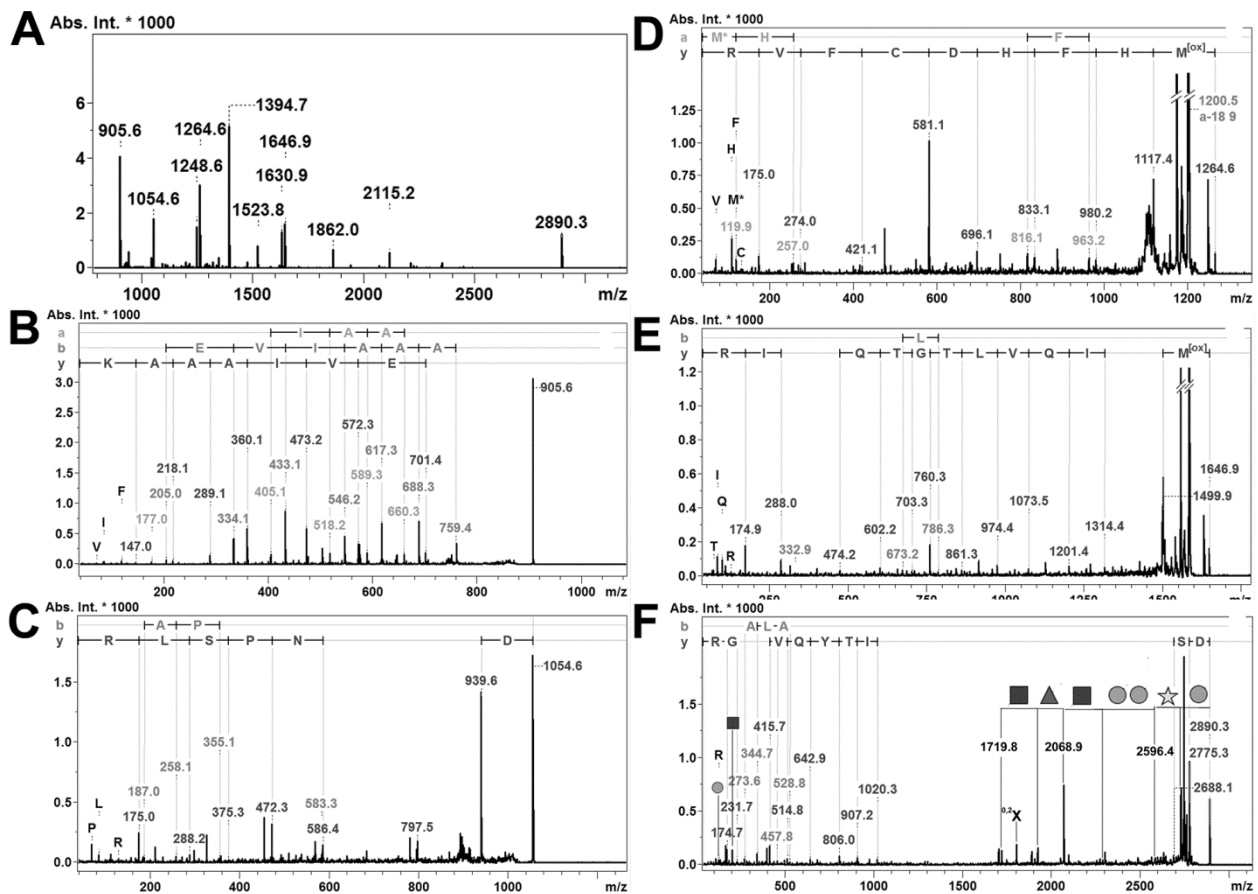


Figure 2.3. MALDI-TOF/TOF MS/MS of WPTP tryptic peptides. (A) mass spectrum WPTP tryptic peptides and fragmentation spectra of the parent ion at  $m/z$  (B) 905.6; (C) 1054.6; (D) 1264.6; (E) 1646.9; and (F) 2980.3. ( $M^{[ox]}$ ) sulfoxide methionine; (■) *N*-acetylglucosamine (GlcNAc); (●) mannose (Man); (▲) fucose (Fuc); (☆) xylose (Xyl).

ACATGGGGAAGATCTCCAGCAAGAAACGCGGACACATCCCTCCCCAGGTCCTTAGCATAGGCCACC 69

ATGTCCCGTCCGGTGAAGCTCTTCTTCTTGGCCTTCCTTGTCTTCTGGCAGCAGTTTCATGGTGATCTACAAATTGGGTCTACAAACCAG 159  
*M S R P V K L F F L A F L A L L A A V H G D L Q I G F Y N Q* 9  
 #

AGCTGCCCATCTGCGGAGTCACTGGTGCAGCAGGCCGTTGCTGCTGCTTTTGCCCAACAATTCAGGCATTGCCCTGGTCTCATCCGCATG 249  
 S C P S A E S L V Q Q A V A A A F A N N S G I A P G L I R M 39  
 \*

CACTTCCATGACTGCTTTGTGAGGGGCTGTGATGCTTCCGTCTTGGACTCAACTGCAATAACACAGCCGAGAAGGATGCAGCCCCG 339  
 H E H D C F V R G C D A S V L L D S T A N N T A E K D A A P 69  
 ~~~~~  
 ← m/z 1264.6 → ← m/z 1054.6 ←

AACAACCCAGCCTCCGCGGCTTGAAGTCAATCGCCGCCGGAAGTCCGCCGTCGAGGCGCGTGCCTGGAAGACCGTCTCCTGCGCCGAC 429  
 N N P S L R G F E V I A A A K S A V E A A C P K T V S C A D 99  
 → m/z 905.6 →

ATCCTCGCCTTCGCGGCCCGCGCAGCGCCGCCCTCGCCGGCAACATCACTTACCAGTCCCTCCGCGCCGCGCGAOCGGCAATGTCTCC 519  
 I L A F A A R D S A A L A G N I T Y Q V P S G R R D G N V S 129  
 \*

← m/z 2890.3 →

CTGCCAGCGAGGGGCTCACCAACATCCCGCGCCACCTTCAAGGCCAGCGAGCTGATCAACAGCTTCGCGGCAAGAACCTCACCGCC 609  
 L A S E A L T N I P A P T F N A T Q L I N S F A G K N L T A 159

GACGAGATGGTCAACCCTCAGCGGGGCCACAGCATCGGTGTCTCCACTGCTTCTCCTTCCTGAACCGGATCTACAACCTCAGCAACACT 699  
 D E M V T L S G A H S I G V S H C F S F L N R I Y N F S N T 189

AGCCAGGTGGACCCACGCTGAGTTCGTATAGCGGATCTCCTCAGGACCAAAATGCCCTCCACAGCACCCGGTTCAGCGCGATCAC 789  
 S Q V D P T L S S S Y A D L L R T K C P S N S T R F T P I T 219

GTGTCGTTGGATATCATCACCCCACGGTGTGGACAACAGGTAATAACGCGGATAACCTGACAGCGTGGGTTGCCAAGTTCCGCGCAGGCATC 879  
 V S L D I I T P T V L D N R Y Y T G V Q L T L G L L T S D Q 249

GCCTTGGTGACCGAGGCCAATCTGAGCGCCGCGGTGAAGAATAACGCGGATAACCTGACAGCGTGGGTTGCCAAGTTCCGCGCAGGCATC 969  
 A L V T E A N L S A A V K N N A D N L T A W V A K F A Q A I 279  
 ~~~~~  
 \*

GTGAGATGGGCGAGATCCAGGTGCTGACCGGGACACAGGGGGAGATAAGAACGAATTCAGCGTGGTGAATAGCGCAAGCCTCGGAGAT 1059  
 V K M G Q I Q V L T G T Q G E I R T N C S V V N S A S L G D 306  
 ~~~~~  
 ← m/z 1646.9 → \*

ATTGTGATGGCATCTGCTCACCTCACGGAGGTGGCGACTAGTTGATGCGATTACCCCTCGGACGCGTGTGGTTGTGTTGCCGTGGGTT 1149  
 I V M A S G H L T E V A T S End

TTGGCTGTGCGTGTACTATATGATGGATATGAATAAAGGAAGACGTGTGTTGCTTTTAAAAAAAAAAAAAAAAAAAAAAAAA 1232

Figure 2.4. cDNA sequencing of WPTP. Nucleotide and amino acid residue numbers are on the right. (gray italicized) *N*- and *C*-terminal signal peptides; amino acid sequence (■■■■) used for primer design, (~~~~~) covered by MALDI-TDS, and (↔) determined by MALDI-TOF/TOF MS/MS; (N) potential glycosylation site; (#) GlcNAc<sub>2</sub>Man<sub>2</sub>FucXyl; and (\*) GlcNAc<sub>2</sub>Man<sub>3</sub>FucXyl.

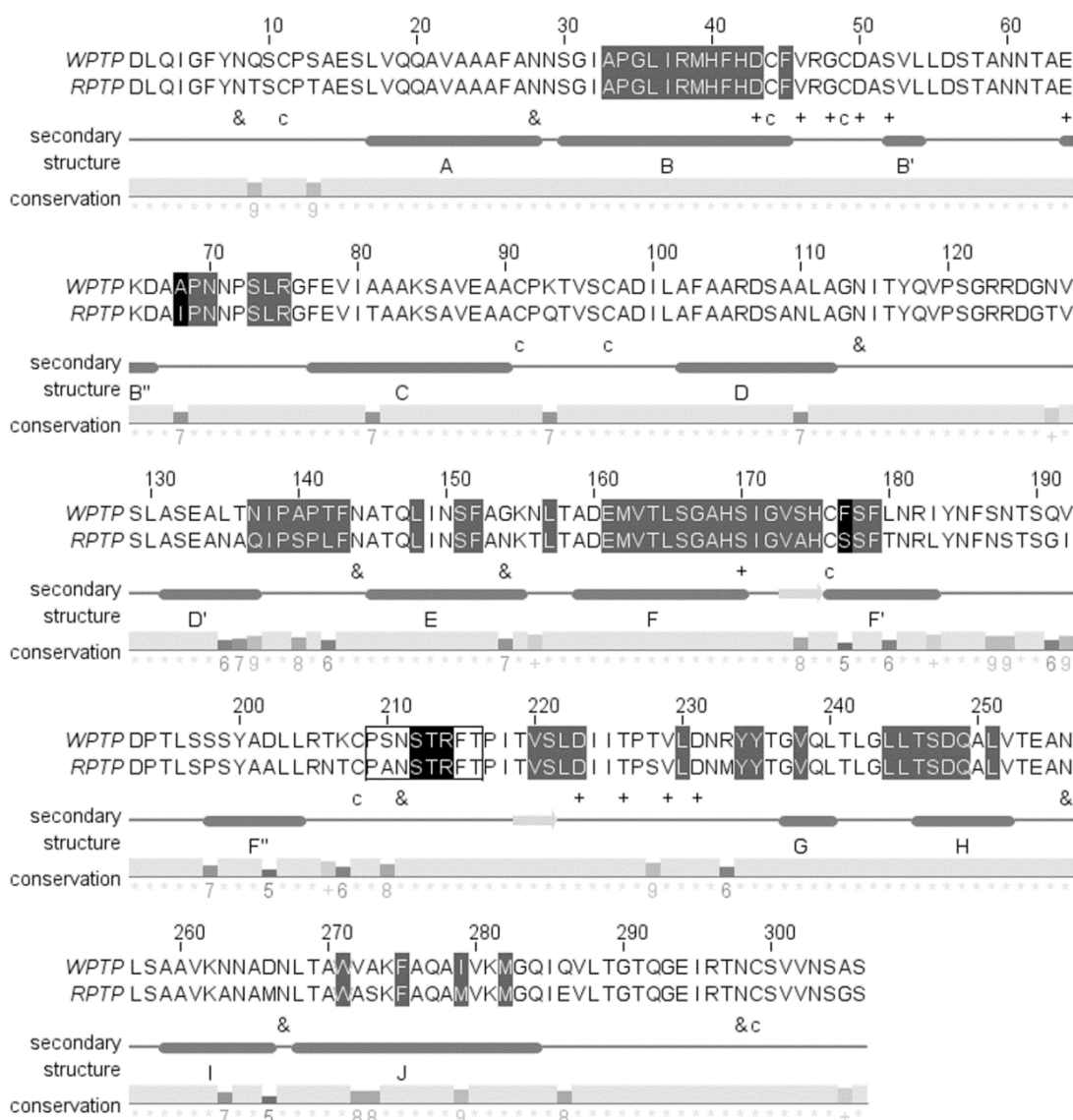


Figure 2.5. Alignment of WPTP and RTPP. The alignment was done to determine features related to substrate specificity, according to Armougom et al. (2006). Determination of (highlighted in gray) residues less than 12 Å from heme iron and (highlighted in black) less than 4.6 Å from MES with Swiss-PdbViewer 4.1.0 (Arnold et al., 2006) and (box) residues with significant structural deviation between RTPP (PDB: 3HDL) and HRP (PDB: 1ATJ) with a protein structure alignment tool (Shindyalov et al., 1998). (&) occupied glycosylation site (Watanabe et al., 2010); (c) cysteine; (+) Ca<sup>2+</sup> binding; (—) helices; and (—) sheets. Helices labeled according to Watanabe et al. (2010) and conservation according to Livingston and Barton (1993). The figure was drawn in JalView (Waterhouse et al., 2009).

## CHAPTER 3. SITE-SPECIFIC *N*-GLYCOSYLATION AND MICROHETEROGENEITY OF A PEROXIDASE FROM THE WINDMILL PALM TREE *TRACHYCARPUS FORTUNEI*

### INTRODUCTION

Windmill palm tree peroxidase is the main anionic vacuolar peroxidase present in the leaves of the windmill palm tree *Trachycarpus fortunei* during the winter (Caramyshev et al., 2006; Baker et al., 2014). Vacuolar peroxidases are stress responsive enzymes (Zipor and Oren-Shamir, 2013). In the presence of hydrogen peroxide, vacuolar peroxidases catalyze oxidative polymerization of aromatic reducing substrates present in the plant vacuole, which includes a diverse array of secondary metabolites. The physiological roles of vacuolar peroxidases include hydrogen peroxide scavenging, synthesis of anti-microbial compounds, and degradation of pigment molecules, an adaptation to low light conditions.

A characteristic feature of secretory proteins, including vacuolar peroxidases, is *N*-linked glycosylation (Ruiz-May et al., 2012). Vacuolar peroxidases are decorated on their loop regions with paucimannose-type sugars (Yang et al., 1996; Gray and Montgomery 2006). Partial deglycosylation of peroxidases results in decreased stability (Tams and Welinder, 1998; Lige et al., 2001) and lower rates of catalysis (Lige et al., 2001; Nielsen et al., 2001). A decrease in specific activity of a peroxidase from avocado was correlated with the loss of ability to chelate calcium ions (Sánchez-Romero et al., 1994). Typically, peroxidases contain 2 Ca<sup>2+</sup> per peroxidase molecule (Veitch 2004). Considering that WPTP contains 6.5 Ca<sup>2+</sup> (Caramyshev et al., 2006), this may be of considerable importance.

The complete amino acid sequence and partial glycosylation of WPTP were recently determined (Baker et al., 2014). Herein, detailed *N*-glycosylation mapping of WPTP is described. Analysis at the glycopeptide level allowed for detailed investigation of the glycosylation profile at each of the 13 sites of *N*-glycosylation. The work was conducted with novel approaches to *N*-glycan database construction and validation of glycopeptide assignment using pGlycoFilter and gPSMvalidator, respectively. A detailed description of palm peroxidase

glycosylation is important for future studies on the structure-activity relationships in *N*-glycosylation variants of WPTP.

## **MATERIALS AND METHODS**

**Materials.** Green palm tree leaves were harvested from the territory of the Nikita Botanic Garden (Crimea, Ukraine). Reagents for sample preparation, including acetonitrile (HPLC grade), ammonium bicarbonate, and 1,4-dithiothreitol (DTT) (No-Weigh DTT) were purchased from Fisher Scientific (Waltham, MA). Trypsin (Trypsin Gold, Mass Spectrometry Grade) and chymotrypsin (Sequencing Grade) were purchased from Promega (Madison, WI). Water (HPLC Grade, JT Baker) was purchased from VWR (Randor, PA). Other reagents used for sample preparation included iodoacetamide (98%, Sigma Aldrich, St. Louis, MO) formic acid (Suprapur, 98%; EMD Chemicals), and 2,2,2-trifluoroethanol (TFE) (99.8% pure, Acros). For desalting and enrichment, C<sub>18</sub> ZipTips (Milipore) and polyhydroxyethyl resin, 200 Å (PolyLC, Inc.) were used.

**WPTP isolation and purification.** WPTP was isolated from the leaves of *Trachycarpus fortunei* in the winter. Isolation and purification were carried out by Dr. Sakharov at Lomonosov University in Moscow, Russia. Full details of the procedure are described by Caramyshev et al., (2006). Briefly, soluble proteins were extracted from milled windmill palm leaves (5 kg) in water with periodic agitation for a total of 30 h at ambient temperature. The debris was removed by filtration and centrifugation. Next, phenolic compounds were separated from the extract by adding solid polyethylene glycol. Addition of ammonium sulfate allowed for protein precipitation. The resulting protein containing aqueous fraction was subjected to a series of preparative chromatography columns, including separation based on hydrophobicity, size-exclusion, and weak anion-exchange. For each chromatography step, the fraction containing peroxidase activity was collected. Peroxidase was purified to homogeneity, which was assessed by SDS-PAGE and isoelectric focusing.

**Sample preparation.** Duplicate samples were prepared for site-specific glycosylation analysis. Dry WPTP was resuspended in 50% 2,2,2-trifluoroethanol (TFE)/50% 100 mM ammonium bicarbonate. Disulfide bonds were reduced with 50 mM 1,4-dithiothreitol (60 °C, 15 min) and then alkylated with 100 mM iodoacetamide (RT, 15 min, dark). After dilution to 10%

TFE, trypsin or chymotrypsin was added in a 1:50 ratio of trypsin or chymotrypsin to sample (w/w). Proteolytic digests were carried out overnight at 37 °C and were terminated by addition of 1 µL of formic acid.

Digested samples were either desalted with C<sub>18</sub> ZipTips (Milipore) according to the manufacturer's instructions, or enriched for glycopeptides using a method adapted from (Kolarich et al., 2012). For enrichment of glycopeptides, a hydrophilic interaction liquid chromatography (HILIC) microcolumn was made with the following procedure: About 9 µL of a slurry of polyhydroxyethyl resin, 200 Å (PolyLC, Inc.) mixed with acetonitrile was added to the top of a 10 µL C<sub>18</sub> ZipTip. The height of the polyhydroxyethyl resin layer was about 5 mm. The microcolumn was centrifuged for 1 min at 5,000 rpm in a benchtop microcentrifuge (Eppendorf 5417R) and then was washed with 20 µL of 0.1% formic acid (v/v) by centrifuging for 1 min at 1,000 rpm and equilibrated with 20 µL of 5% formic acid (v/v), 80% acetonitrile (v/v). The sample, WPTP chymotryptic peptides (4 µg) in 20 µL of 5% formic acid (v/v), 80% acetonitrile (v/v), was loaded and centrifuged for 1 min at 400 rpm. Next, the microcolumn was washed twice with 20 µL of 5% formic acid (v/v), 80% acetonitrile (v/v) by centrifuging for 1 min at 2,000 rpm. A 'fast-spin' for 2 s removed all excess liquid from the microcolumn, at which time the material turned from translucent to white. The first elution was with 20 µL of 0.1% formic acid (v/v), and subsequently with 10 µL of 5% formic acid (v/v), 80% acetonitrile (v/v). Each time, centrifugation was performed for 1 min at 2,000 rpm, followed by a 2 s fast-spin. The eluents were collected and combined. Next, the eluents were evaporated on a SpeedVac. The enriched glycopeptides were reconstituted in 50 µL of 0.1% formic acid (v/v), 5% acetonitrile (v/v) by bath sonication for 10 min. After spinning down the sample for 3 min at a high speed, 40 µL of the solution was added to an HPLC vial for LC-MS analysis.

**Liquid chromatography-mass spectrometry.** All chromatographic separations were performed using an Easy nLC-1000 pump and autosampler system (Thermo Fisher Scientific). For each analysis, 2 µL of each sample (equivalent to 0.6 µg) was injected and loaded onto an in-line solid-phase extraction column (100 µm x 6 cm) packed in-house with ReproSil-Pur C<sub>18</sub> AQ 3 µm resin (Dr. Maisch GmbH) using a frit generated with liquid silicate Kasil 1 at a flow rate of 2 mL/min with 100% Solvent A (0.1 % formic acid (v/v)). After the

washing period of a total volume of 7 mL, the pre-column was placed in-line with a PicoFrit capillary column (New Objective, 11 cm x 75  $\mu$ m) packed in-house with the same resin. The peptides were separated using a linear gradient of 2% - 35% Solvent B (0.1% formic acid (v/v) in acetonitrile) at a flow rate of 300 nL/min over 40 min followed by an increase to 90% B over 4 min and held at 90% B for 6 min before returning to initial conditions of 2% B. The samples were analyzed on a Q Exactive Orbitrap (Thermo Scientific). The normalized collision energy was set to 28% for HCD fragmentation. The top 12 ions were fragmented in HCD mode with an isolation window of 2.5  $m/z$  and dynamic exclusion period of 20 s.

**Bioinformatics.** The detailed workflow for analysis of the glycopeptide mass spectrometry data can be found in [Appendix A](#). Configurations for the different proteomic searches can be found in [Appendix B](#). The workflow contains 5 general steps ([Fig. 3.1](#)). Mass spectrometry data were converted to mzML format using msconvert (ProteoWizard v.3.0.5006). Proteins were identified with a MyriMatch (v. 2.1.138; Tabb et al., 2007) search using a protein database containing the target protein sequence, potential protein contaminants and an entire organism's proteome for a total of 25,930 protein sequences. Preliminary identifications were assessed with IDPicker 3.0 (v. 3.0.537 64-bit) with a q-value of 2%. A subset protein database was constructed from the results and used in subsequent glycopeptide identifications using the proteomics software.

All possible glycan compositions were predicted using GlycoMod (Cooper et al., 2001; [web.expasy.org/glycomod](http://web.expasy.org/glycomod)). The list of glycans was evaluated using pGlycoFilter, an in-house tool written in Excel macros ([Supplementary Data](#)) that tests the predicted glycan compositions against the known *N*-glycan biosynthesis pathway in plants ([Appendix C](#)).

Glycopeptides were identified using several proteomics search engines and the subset database produced in step 1. The proteomics search engines included MyriMatch (v. 2.1.138; Tabb et al., 2007), a database search engine, DirecTag-TagRecon (v. 1.4.71; Dasari et al., 2010), a tag-based search engine, and MassRecon (v. 1.4.71; Dasari et al., 2010), a mass-based search engine. Alternatively, glycopeptides were identified using a modified GlycoMod search. Preliminary identifications from the proteomics searches were assessed with IDPicker 3.0 (v. 3.0.537 64-bit) with a q-value of 2%. Additional glycopeptide identifications were obtained

using GlycoMod. All glycopeptide spectrum matches (gPSMs) were validated using gPSMvalidator, an in-house built tool modified from PTM\_MarkerFinder (Nanni et al., 2013) written in R. Each gPSM within 15 ppm of the exact mass was considered valid if 2 or more HexNAc<sup>+</sup> oxonium ions and one or more Y<sub>1</sub> (peptide + GlcNAc) or Y<sub>0</sub> (peptide) ion were detected within an error of 5 ppm or 50 ppm, respectively. All gPSMs were manually validated. Relative quantitation was done by comparing the summed intensity of selected ion chromatograms (SIC) of glycopeptide precursor ions.

## SUPPLEMENTARY DATA

A description of the supplementary data is in [Appendix D](#). The raw mass spectrometry files have been uploaded to the ProteoSAFe webserver ([massive.ucsd.edu](http://massive.ucsd.edu)) under MassIVE ID: MSV000079086.

## RESULTS

**Sample preparation.** WPTP, an anionic glycoprotein, pI 3.5 (Caramyshev et al., 2006) was isolated from the leaves of *Trachycarpus fortunei* in the winter. The enzyme was purified to homogeneity and migrated as a single band at approximately 50 kDa on SDS-PAGE (data not shown). It is expected that the *N*-linked glycans of WPTP were not altered by the isolation and purification procedure. The possibility of glycan modification during the extraction step was low because the concentration of the peroxidase and glycosidases remained low. Moreover, the homogenate contained very high concentrations of polysaccharides, which have higher affinity to glycosidases than glycosylated proteins. At all other steps of purification the enzyme concentration was low and the experimental conditions were not favorable for glycosidases; e.g., a buffer containing 1.7 M ammonium sulfate was used for hydrophobic chromatography.

The preferred peptide backbone length for site-specific glycosylation analysis is 8 - 25 amino acids (Kolarich et al., 2012). This was achieved using chymotrypsin, which cleaves after large, hydrophobic amino acids such as Phe, Tyr, Trp, and Leu. WPTP was alternatively digested with trypsin to give complementary information. A portion of the chymotrypsin digested sample

was enriched for glycopeptides using hydrophilic interaction liquid chromatography (HILIC). Enriched glycopeptides were used to provide a more comprehensive profile of glycosylation at each site; however, such treatment is biased against peptides meaning that information on unoccupied glycosylation sites is lost (Kolarich et al., 2012). Therefore, the unenriched chymotrypsin sample was used to assess the relative abundance of glycoforms at each glycosylation site.

The total ion chromatogram (TIC) shows the summed intensity of each detected compound and its elution time ([Fig. 3.2A-C](#)). The majority of compounds were eluted between 25 and 60 min. As evident from the TICs of each sample, trypsin digestion resulted in fewer cleavage products ([Fig. 3.2A](#)) than chymotrypsin digestion ([Fig. 3.2B](#)). Also, enrichment of the chymotrypsin sample resulted in detection of a much more heterogeneous sample ([Fig. 3.2C](#)).

**Data analysis workflow.** Data analysis was performed in 5 general steps ([Fig. 3.1](#)). At each step, the results were assessed against specific criteria, as described in each section below. Glycopeptide structures reported in this study include the peptide backbone sequence and the glycan composition. Glycan attachment sites were inferred based on the presence of the consensus sequence for *N*-linked glycosylation, Asn-Xxx-Ser/Thr, where Xxx ≠ Pro. Glycan structures were predicted from the glycan composition and the known glycan biosynthesis pathway in plants ([Appendix C](#)).

**Step 1: Protein identification.** In step 1, results were assessed based on the detection of WPTP as the main protein component as well as detection of known contaminants including proteases (chymotrypsin or trypsin) and keratins. The primary component was WPTP (1382 filtered spectra), followed by keratins (20 filtered spectra), and the proteases chymotrypsin (11 filtered spectra) or trypsin (6 filtered spectra). The complete list of peptide spectrum matches (PSMs) and annotated fragmentation spectra are available in the [Supplementary Data](#).

**Step 2: Glycan identification.** Glycan compositions were predicted from precursor masses and WPTP peptide masses using GlycoMod (Cooper et al., 2001). There were 280 predicted glycan compositions. At the time of this publication, there were no comprehensive plant glycan databases available to confirm the glycan compositions. This prompted the need for a tool that could omit biologically impossible glycan compositions. A tool, called pGlycoFilter,

was written in Excel macros ([Supplementary Data](#)). The predicted glycan compositions were checked against the known plant *N*-glycosylation pathway ([Appendix C](#)). A set of *N*-glycans, determined experimentally from plants (Wilson et al., 2001), was tested against pGlycoFilter and all 30 glycans passed. Of the 280 glycan compositions predicted by GlycoMod, 80 of them passed pGlycoFilter. Following validation of the glycopeptide spectrum matches (gPSMs) (described below), 45 different glycans were confirmed to be present on WPTP. The gPSMs that were not confirmed were matched based on the parent ion mass, but not fragment ion masses. [Table 3.1](#) lists the glycan composition, the change in mass observed, and a symbol designator.

**Glycopeptide fragmentation.** Fragmentation spectra were used to determine the glycopeptide structures. The fragmentation mode used in this study was Higher-energy C-trap Dissociation (HCD) (Olsen et al., 2006). The nomenclature used to describe glycopeptide and peptide fragmentation is from Domon and Costello (1988) and Roepstorff and Fohlmann (1984), respectively.

A major benefit of HCD glycopeptide fragmentation is the reliable formation of the  $Y_1$  ion, which is the peptide-GlcNAc ion (Segu and Mechref 2010; Hart-Smith and Rafferty 2012). The  $Y_1$  and peptide ( $Y_0$ ) ions were detected as a prominent doublet of peaks in the high  $m/z$  region and often appeared in multiple charge states (e.g., +1, +2, and/or +3). A representative glycopeptide spectrum is shown in [Fig. 3.3](#). The most prominent features are carbohydrate oxonium ions in the low mass region and the  $Y_0$  and  $Y_1$  ions in the high mass region. Oxonium ions have long been used as markers for glycopeptides in a mixture (Huddleston et al., 1993). Commonly observed oxonium ions include  $\text{Hex}^+$  ( $m/z$  163.06),  $\text{GlcNAc}^+$  ( $m/z$  204.09), and  $\text{Hex-GlcNAc}^+$  ( $m/z$  366.14). HCD fragmentation causes  $\text{GlcNAc}^+$  and  $\text{Hex}^+$  to further decompose to form highly abundant internal fragment ions (Zhao et al., 2011) as illustrated in [Fig. 3.4](#), which is a zoomed in view of the spectrum presented in [Fig. 3.3](#).

The only major deviation from the characteristic detection of the  $Y_0/Y_1$  pair is a (proposed) cross-ring fragmentation observed prominently for glycopeptides bearing glycan  $\text{GlcNAcMan}_2\text{XylGlcNAc}_2$  (glycan o;  $\Delta$  mass 1065.386) or  $\text{GlcNAcMan}_3\text{XylGlcNAc}_2$  (glycan t;  $\Delta$  mass 1227.439) ([Fig. 3.5](#)). It is proposed to be  $^{2,5}\text{X}$  cross-ring cleavage, with additional loss of

(-COCH<sub>3</sub>) from the *N*-acetyl group, of the innermost GlcNAc ( $m/z$  Y<sub>0</sub> + 57.02) or the second GlcNAc ( $m/z$  Y<sub>1</sub> + 57.02).

The presence of peptide backbone fragmentation ions (b,y) and glycan-peptide fragmentation ions (Y), unlike the relatively constant Y<sub>1</sub>, Y<sub>0</sub>, and oxonium ions, appeared to be strongly dictated by the characteristics of the peptide backbone ([Table 3.2](#)). These results were in accordance with the mobile proton model of peptide fragmentation as reviewed by Paisz and Suhai (2005). For example, the features of peptide 3, LDSTANNTAEKDAAPNNPSLRGF, with pI 4.6, made it particularly amenable to y-ion formation ([Fig. 3.6](#)). C-Terminal to the glycosylation site are 17 amino acids which include both acidic (Glu, Asp) and basic amino acids (Lys, Arg). Acidic amino acids promote fragmentation, whereas basic amino acids tend to inhibit fragmentation via proton sequestration. In stark contrast, peptide 10: RTKCPSNSTRF, showed abundant glycan + peptide fragmentation in the form of Y-ions and very little peptide backbone fragmentation ([Fig. 3.7](#)). This basic peptide, pI 10.86, had short stretches of amino acids flanking the glycosylation site which contained proton sequestering basic residue(s) (Lys, Arg) and no intervening acidic residues.

**Steps 3 and 4: Glycopeptide identification and validation.** Glycopeptides were identified using proteomics or glycoproteomics software. The proteomics software included MyriMatch (Tabb et al., 2007), a database search engine, DirecTag-TagRecon (Dasari et al., 2010), a tag-based search engine, and MassRecon (Dasari et al., 2010), a mass-based search engine. The online tool GlycoMod (Cooper et al., 2001) identified glycopeptides by their accurate precursor mass.

Because glycopeptide-specific fragment ions are not detected by those software, additional steps were required to validate the glycopeptide spectrum matches (gPSMs). To serve this purpose, a tool called gPSMvalidator was developed. Instructions for the use of gPSMvalidator are in [Appendix A](#). The R scripts for gPSMvalidator are provided in [Supplementary Data](#). The exceptional mass accuracy and the reliable formation of the Y<sub>1</sub> fragment ion were the basis for unambiguous glycopeptide validation using gPSMvalidator. It was modified from an existing tool in the R package ProtViz called PTM\_MarkerFinder (Nanni et al., 2013). PTM\_MarkerFinder was, in part, designed to search through HCD data for spectra

that contained the characteristic HexNAc<sup>+</sup> oxonium ion and its internal fragments. The modified tool, gPSMvalidator, additionally looks for the Y<sub>1</sub> and Y<sub>0</sub> ions. Due to the high mass accuracy achieved with an orbitrap mass analyzer, the observed mass, Y<sub>1</sub> and Y<sub>0</sub> ions were sufficient for validation of glycopeptides from a single glycoprotein in this study. A list of all identified glycopeptides can be found in the [Supplementary Data. Table 3.3](#) summarizes the glycopeptide identification results according to glycosylation site.

**Step 5: Relative quantitation of glycoforms observed at each site.** The relative abundance of glycoforms was determined for each of the 13 glycosylation sites of WPTP using the relative intensity of confidently identified glycopeptides. For each glycosylation site, a representative peptide backbone was selected using data collected for the unenriched chymotrypsin digested sample. In a multi-institutional study, it was found that relative quantitation results for glycopeptides were most comparable between different laboratories if a single peptide backbone was used (Leymarie et al., 2013). Each glycoform series was observed in multiple charge states. The most abundant charge state for each glycoform series was used for relative quantitation analysis (Ozohanics et al., 2012). The selected ion chromatogram (SIC) for the monoisotopic peak of each compound was extracted from the total ion chromatogram. Alternatively, more peaks in the isotope cluster could be used; however, this approach presents problems in the case of overlapping isotope clusters (Ozohanics et al., 2012). To calculate the relative abundance, the summed intensity of each SIC was divided by the summed intensity for the entire glycoform series.

The complete list of glycoforms of each glycosylation site and their relative abundance is presented in the [Supplementary Data](#). Also included in this analysis was a comparison between the duplicate unenriched chymotrypsin samples, which were in excellent agreement.

More than 90% of the population of each glycosylation site could be described by 3 or fewer glycoforms made up of only 5 of the observed glycans ([Fig. 3.8](#)). [Table 3.4](#) shows the relative abundance of the major glycoforms of WPTP. Chromatographic separation of glycopeptides is primarily dictated by the hydrophobicity of the peptide backbone. As a result, each glycoform in a particular glycoform series was eluted within a specific time period. Those with larger, more hydrophilic glycans attached were eluted earlier (data not shown).

## DISCUSSION

**Site occupancy.** Glycoproteins may have evolved to maximize the desired interactions with the oligosaccharyltransferase (OST) complex in order to ensure that certain sites are always occupied; whereas partial occupation in the other sites may allow for fine-tuning of the glycoprotein function (Petrescu et al., 2004). Sites 11 and 12 were glycosylated only half of the time (Table 3.4). This is potentially regulated by the presence of an acidic residue just upstream of each of the two glycosylation sites, as seemed to be the case in other glycoproteins (Petrescu et al., 2004).

Surprisingly, the proximity of sites 8 and 9 does not cause steric hindrance during the OST complex reaction (Table 3.4). These 2 sites were both glycosylated more than 90% of the time. Horseradish peroxidase (HRP) (Yang et al., 1996) and soybean peroxidase (SBP) (Gray and Montgomery 2006) are also diglycosylated most of the time in this region, although the sites are further apart. Royal palm tree peroxidase (Watanabe et al., 2010) and peanut peroxidase (PNP) (Zhang et al., 2004) maintain constant glycosylation in this region, using only one site. Site-directed removal of this glycosylation site in PNP showed that glycosylation in this region plays an important role in thermostability of the enzyme (Lige et al., 2001).

**Heterogeneity.** An exceptional degree of heterogeneity was observed (Table 3.3); however, more than 90% of each glycosylation site can be described by 3 or fewer glycoforms (Table 3.4). Plant peroxidases are known to be heterogeneous glycoproteins; however, the presence of up to 30 different glycans at one site is unprecedented for plant peroxidases. The most heterogeneous site in HRP had 11 glycans (Yang et al., 1996), and the most in one site of PNP was 8 glycans (Zhang et al., 2004). Detection of so many more of the very minor complex components compared to existing studies is probably reflective of the huge advancements made in analytical mass spectrometry in the recent years. The relatively low amount of heterogeneity observed in PNP may stem from the fact that this is not a vacuolar peroxidase, or possibly because it was produced in cell culture (Zhang et al., 2004).

**The major glycans of WPTP.** The paucimannosidic nature of the major glycoforms reiterates that WPTP is a vacuolar peroxidase. Secretion to the vacuole puts *N*-glycoproteins in contact with glycosidases (reviewed by Leonard et al., 2009) which trim the *N*-linked glycans to

reveal the major glycans observed in WPTP ([Fig. 3.8](#)). The broad-spectrum *N*-hexosaminidase HEXO1 is responsible for, Man<sub>3</sub>XylFucGlcNAc<sub>2</sub> (glycan p;  $\Delta$  mass 1170.417), the predominant glycan on WPTP and other vacuolar peroxidases including HRP (Yang et al., 1996), SBP (Gray and Montgomery 2006), and barley grain peroxidase (Henriksen et al., 1998). Endo- $\beta$ -mannosidase catalyzes production of GlcNAc<sub>2</sub>, necessary for HEXO1 to generate the other major glycan GlcNAc (glycan b;  $\Delta$  mass 203.079) (Strasser et al., 2007), found to be prominent at site 3.

**Site 3.** Site 3, represented by Asn 60 in WPTP is occupied by a single GlcNAc (glycan b;  $\Delta$  mass 203.079) residue 70% of the time and by Man<sub>2</sub>XylFucGlcNAc<sub>2</sub> (glycan k;  $\Delta$  mass 1008.365) 25% of the time. At the analogous site in HRP and SBP, Man<sub>3</sub>XylFucGlcNAc<sub>2</sub> (glycan p;  $\Delta$  mass 1170.417) and high mannose-type glycans are found, respectively. This glycosylation site is in the vicinity of the distal calcium binding region. The distal calcium is important for maintaining the tertiary structure of the active site (Szigeti et al., 2008). Deletion of this site in PNP caused a 40% decrease in enzyme activity (Lige et al., 2001). The role of glycosylation at site 3 may be to prevent proton exchange with the solvent, maintaining the hydrogen bonding network between the heme active site and the distal and proximal calcium ions.

The stabilizing effects of *N*-linked glycans appear to result from hydrophobic/hydrophilic interactions between the first few residues of the glycan and the side chains of neighboring amino acids (Wormald et al. 2002). Endo-H is an endoglycosidase that cleaves *N*-linked glycans to leave a single GlcNAc attached. After digestion of the human complement protein, CD2 with Endo-H, it was observed that this single carbohydrate attachment was sufficient to stabilize the protein (Wyss et al., 1997). Thus, the specific glycan itself may not matter as much as its presence.

**Workflow and tools for analysis of plant glycopeptide mass spectrometry data.** Automated analysis of glycoproteomics data lags far behind proteomics (Desaire 2013; Li et al., 2013). There is a growing need for analytical tools for the analysis of plant glycoproteins (Strasser 2014). For plant glycoproteomics, only a few analytical workflows have been reported,

and none include a strategy for automated data analysis (Song et al., 2011; Ruiz-May et al., 2012; Ruiz-May et al., 2014).

Herein we describe a workflow and tools for analysis of plant glycopeptide mass spectrometry data. The workflow relied on free software. The glycan database generation tool, pGlycoFilter relied on GlycoMod and Excel macros. Additionally, gPSMvalidator was written in R. Both of these can be readily modified to suit the needs of the user. The workflow was geared toward HCD fragmentation data, which was a limitation. This may be the subject of future work to make it compatible with other fragmentation types.

Table 3.1. Glycans observed on WPTP

| symbol | glycan $\Delta$<br>mass | glycan composition          |
|--------|-------------------------|-----------------------------|
| a      | 0                       | unoccupied                  |
| b      | 203.079                 | GlcNAc(1)                   |
| c      | 406.159                 | GlcNAc(2)                   |
| d      | 552.217                 | GlcNAc(2)Fuc(1)             |
| e      | 714.27                  | GlcNAc(2)Hex(1)Fuc(1)       |
| f      | 730.264                 | GlcNAc(2)Hex(2)             |
| g      | 846.312                 | GlcNAc(2)Hex(1)Fuc(1)Xyl(1) |
| h      | 862.307                 | GlcNAc(2)Hex(2)Xyl(1)       |
| i      | 876.322                 | GlcNAc(2)Hex(2)Fuc(1)       |
| j      | 892.317                 | GlcNAc(2)Hex(3)             |
| k      | 1008.365                | GlcNAc(2)Hex(2)Fuc(1)Xyl(1) |
| l      | 1024.36                 | GlcNAc(2)Hex(3)Xyl(1)       |
| m      | 1038.375                | GlcNAc(2)Hex(3)Fuc(1)       |
| n      | 1054.37                 | GlcNAc(2)Hex(4)             |
| o      | 1065.386                | GlcNAc(3)Hex(2)Xyl(1)       |
| p      | 1170.417                | GlcNAc(2)Hex(3)Fuc(1)Xyl(1) |
| q      | 1186.412                | GlcNAc(2)Hex(4)Xyl(1)       |
| r      | 1200.428                | GlcNAc(2)Hex(4)Fuc(1)       |
| s      | 1211.444                | GlcNAc(3)Hex(2)Fuc(1)Xyl(1) |
| t      | 1227.439                | GlcNAc(3)Hex(3)Xyl(1)       |
| u      | 1241.455                | GlcNAc(3)Hex(3)Fuc(1)       |
| v      | 1257.449                | GlcNAc(3)Hex(4)             |
| w      | 1332.47                 | GlcNAc(2)Hex(4)Fuc(1)Xyl(1) |
| x      | 1348.465                | GlcNAc(2)Hex(5)Xyl(1)       |
| y      | 1357.502                | GlcNAc(3)Hex(2)Fuc(2)Xyl(1) |
| z      | 1362.481                | GlcNAc(2)Hex(5)Fuc(1)       |
| aa     | 1373.497                | GlcNAc(3)Hex(3)Fuc(1)Xyl(1) |
| bb     | 1378.476                | GlcNAc(2)Hex(6)             |
| cc     | 1389.492                | GlcNAc(3)Hex(4)Xyl(1)       |
| dd     | 1403.507                | GlcNAc(3)Hex(4)Fuc(1)       |
| ee     | 1430.518                | GlcNAc(4)Hex(3)Xyl(1)       |
| ff     | 1494.523                | GlcNAc(2)Hex(5)Fuc(1)Xyl(1) |
| gg     | 1519.555                | GlcNAc(3)Hex(3)Fuc(2)Xyl(1) |
| hh     | 1535.55                 | GlcNAc(3)Hex(4)Fuc(1)Xyl(1) |
| ii     | 1540.528                | GlcNAc(2)Hex(7)             |

Table 3.1 (Continued) Glycans observed on WPTP

| symbol | glycan $\Delta$<br>mass | glycan composition          |
|--------|-------------------------|-----------------------------|
| jj     | 1551.545                | GlcNAc(3)Hex(5)Xyl(1)       |
| kk     | 1576.576                | GlcNAc(4)Hex(3)Fuc(1)Xyl(1) |
| ll     | 1681.608                | GlcNAc(3)Hex(4)Fuc(2)Xyl(1) |
| mm     | 1697.602                | GlcNAc(3)Hex(5)Fuc(1)Xyl(1) |
| nn     | 1738.629                | GlcNAc(4)Hex(4)Fuc(1)Xyl(1) |
| oo     | 1859.655                | GlcNAc(3)Hex(6)Fuc(1)Xyl(1) |
| pp     | 1884.687                | GlcNAc(4)Hex(4)Fuc(2)Xyl(1) |
| qq     | 1900.682                | GlcNAc(4)Hex(5)Fuc(1)Xyl(1) |
| rr     | 2046.74                 | GlcNAc(4)Hex(5)Fuc(2)Xyl(1) |
| ss     | 2192.798                | GlcNAc(4)Hex(5)Fuc(3)Xyl(1) |

Table 3.2. Peptide backbone features and observed fragment ion types

| peptide | peptide sequence <sup>a</sup>     | pI <sup>b</sup> | observed fragment ion types <sup>c</sup> |
|---------|-----------------------------------|-----------------|------------------------------------------|
| 1       | <u>N</u> QSCPSAESL                | 4               | y-ions                                   |
| 2       | A <u>N</u> NSGIAPGLIRMHF          | 9.8             | y-ions                                   |
| 3       | LDSTAN <u>N</u> TAEKDAAPNNPSLRGF  | 4.6             | y-ions                                   |
| 4       | AARDSAALAG <u>N</u> ITY           | 5.9             | b-ions                                   |
| 5       | QVPSGRRD <u>G</u> NVSLASEAL       | 6               | Y-ions                                   |
| 6       | <u>N</u> ATQLINSF                 | 5.5             | y-ions; b-ions with loss of glycan       |
| 7       | AG <u>K</u> NLTADEMRTL            | 4.4             | y-ions; b-ions with loss of glycan       |
| 8, 9    | <u>N</u> FS <u>N</u> TSQVDPTLSSSY | 3.8             | y-ions                                   |
| 10      | RTKCPS <u>N</u> STRF              | 10.9            | Y-ions                                   |
| 11      | TSDQALVTEA <u>N</u> L             | 3.7             | b-ions                                   |
| 12      | SAAVKNNAD <u>N</u> LTAW           | 5.5             | b,y-ions                                 |
| 13      | TGTQGEIRT <u>N</u> CSVVNSAS       | 5.7             | y-ions; b-ions with loss of glycan       |

<sup>a</sup>N-Glycosylation sites underlined; all cysteines carbamidomethylated; <sup>b</sup>Calculated with the ProtParam online tool (Gasteiger et al., 2005; [web.expasy.org/protparam](http://web.expasy.org/protparam)); <sup>c</sup>Peptide ions referred to as b-, y-ions; glycan (with peptide attached) ions referred to as Y-ions

Table 3.3. Microheterogeneity of glycosylation of WPTP

| site | number of glycoforms | observed glycans <sup>a</sup>                                                                                  |
|------|----------------------|----------------------------------------------------------------------------------------------------------------|
| 1    | 3                    | k, m, p                                                                                                        |
| 2    | 8                    | l, m, p, q, s, t, v, w                                                                                         |
| 3    | 12                   | a, b, c, d, h, i, k, o, p, t, w, aa                                                                            |
| 4    | 22                   | d, f, g, i, j, k, l, m, p, q, r, t, v, w, aa, cc, ff, hh, kk, nn, rr, ss                                       |
| 5    | 15                   | b, d, g, h, k, m, o, p, q, t, aa, hh, nn, rr, ss                                                               |
| 6    | 12                   | a, j, k, l, m, p, q, t, w, aa, cc, hh                                                                          |
| 7    | 5                    | k, l, m, p, t                                                                                                  |
| 8, 9 | 19                   | b, k, p, q, aa, cc, gg, hh, ii, i+p, k+n, k+p, k+q, e+ff, p+p, p+q, k+y, k+ff, p+ff                            |
| 10   | 33                   | a, b, c, e, f, g, h, i, j, k, l, m, o, p, q, t, u, w, x, z, aa, cc, ee, ff, hh, jj, ll, mm, nn, oo, pp, qq, rr |
| 11   | 2                    | a, p                                                                                                           |
| 12   | 14                   | a, b, c, d, k, l, m, p, q, t, w, aa, ff, hh                                                                    |
| 13   | 20                   | a, d, j, k, l, m, p, q, r, t, w, x, z, aa, bb, cc, dd, ff, hh, jj                                              |

<sup>a</sup>Designated by symbol defined in [Table 3.1](#)

Table 3.4. Relative abundance of major glycoforms of WPTP

| site | peptide sequence                        | charge | retention time (min) | glycan $\Delta m$ | glycan symbol | expected $m/z^a$ | relative ratio |
|------|-----------------------------------------|--------|----------------------|-------------------|---------------|------------------|----------------|
| 1    | <u>N</u> QSCPSAESL                      | 3      | 36 - 38              | 1008.365          | k             | 700.947          | 0.25           |
|      |                                         |        |                      | 1170.417          | p             | 754.965          | 0.71           |
| 2    | A <u>N</u> NSGIAPGLIRM <sup>ox</sup> HF | 3      | 45 - 46.2            | 1170.417          | p             | 928.751          | 0.89           |
|      |                                         |        |                      | 1227.439          |               | 947.758          | 0.11           |
| 3    | LDSTAN <u>N</u> TAEKDAAPNNPSLRGF        | 3      | 42.1 - 43.8          | 203.079           | b             | 869.416          | 0.70           |
|      |                                         |        |                      | 1008.365          | k             | 1137.844         | 0.25           |
| 4    | AAR <u>D</u> SAALAG <u>N</u> ITY        | 3      | 41 - 44              | 1170.417          | p             | 855.380          | 0.51           |
|      |                                         |        |                      | 1227.439          |               | 874.387          | 0.43           |
| 5    | QVPSGRRDGN <u>V</u> SLASEAL             | 3      | 39 - 43              | 1008.365          | k             | 955.447          | 0.20           |
|      |                                         |        |                      | 1170.417          | p             | 1009.465         | 0.75           |
| 6    | <u>N</u> ATQLINSF                       | 2      | 50.2 - 50.8          | 1170.417          | p             | 1089.470         | 0.97           |
| 7    | AGK <u>N</u> LTADEMVTL                  | 2      | 46 - 47              | 1170.417          | p             | 1267.059         | 0.93           |
| 8, 9 | <u>N</u> FS <u>N</u> TSQVDPTLSSSY       | 3      | 44.3 - 45            | 2178.782          | k + p         | 1309.193         | 0.07           |
|      |                                         |        |                      | 2340.834          | p + p         | 1363.210         | 0.83           |
| 10   | RTKCPS <u>N</u> STRF                    | 4      | 27.5 - 30            | 1170.417          | p             | 631.777          | 0.78           |
|      |                                         |        |                      | 1227.439          | t             | 646.033          | 0.08           |
|      |                                         |        |                      | 1332.47           | w             | 672.290          | 0.04           |
| 11   | TSDQALVTEAN <u>L</u>                    | 2      | 46.4 - 48.9          | 0                 | a             | 631.317          | 0.51           |
|      |                                         |        |                      | 1170.417          | p             | 1216.526         | 0.49           |
| 12   | SAAVKNNAD <u>N</u> LTAW                 | 2      | 43.8 - 46.5          | 0                 | a             | 737.868          | 0.65           |
|      |                                         |        |                      | 1170.417          | p             | 1323.077         | 0.35           |
| 13   | TGTQGEIRT <u>N</u> CSVVNSAS             | 3      | 35.5 - 36.5          | 1170.417          | p             | 1017.770         | 0.80           |
|      |                                         |        |                      | 1332.47           | w             | 1071.787         | 0.11           |

<sup>a</sup>Standard deviation of observed  $m/z$  less than 5 ppm

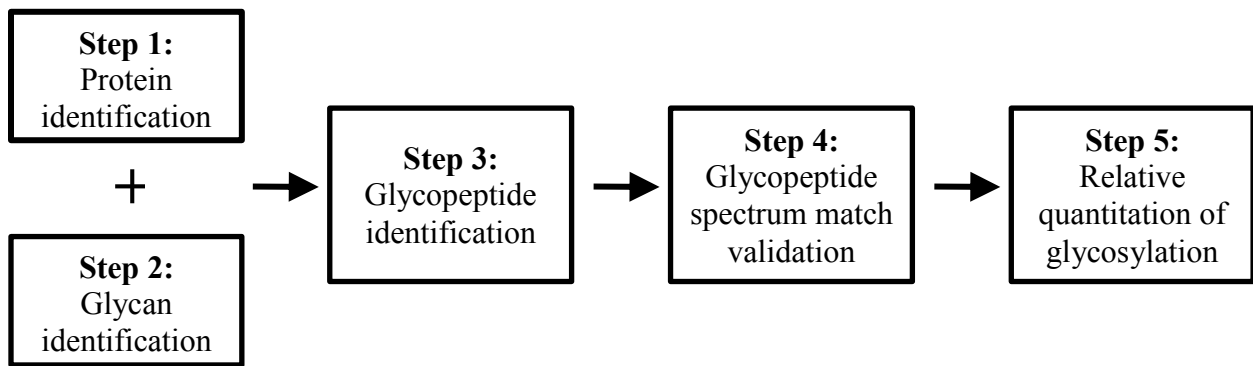


Figure 3.1. Steps of the data analysis workflow

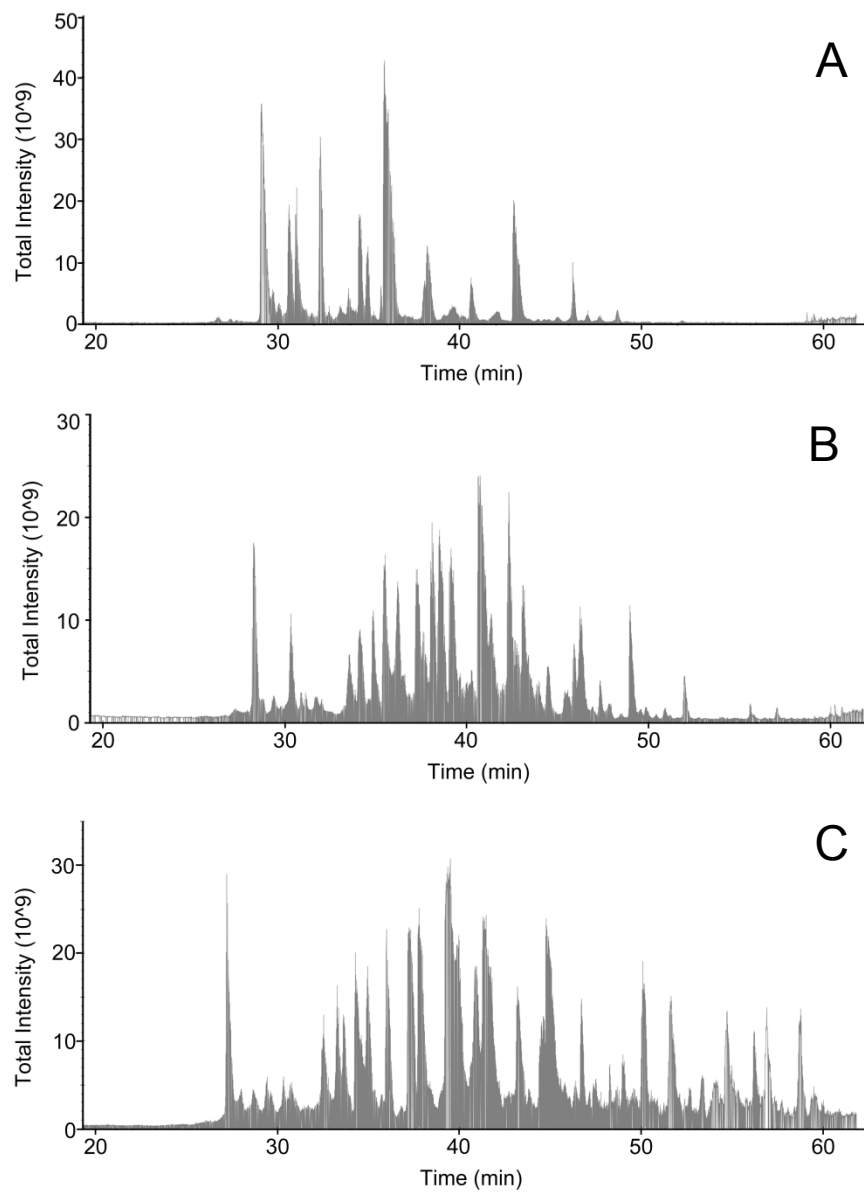


Figure 3.2. Total ion chromatograms of digested WPTP samples. WPTP was digested with (A) trypsin, (B) chymotrypsin, or (C) chymotrypsin and enriched for glycopeptides.

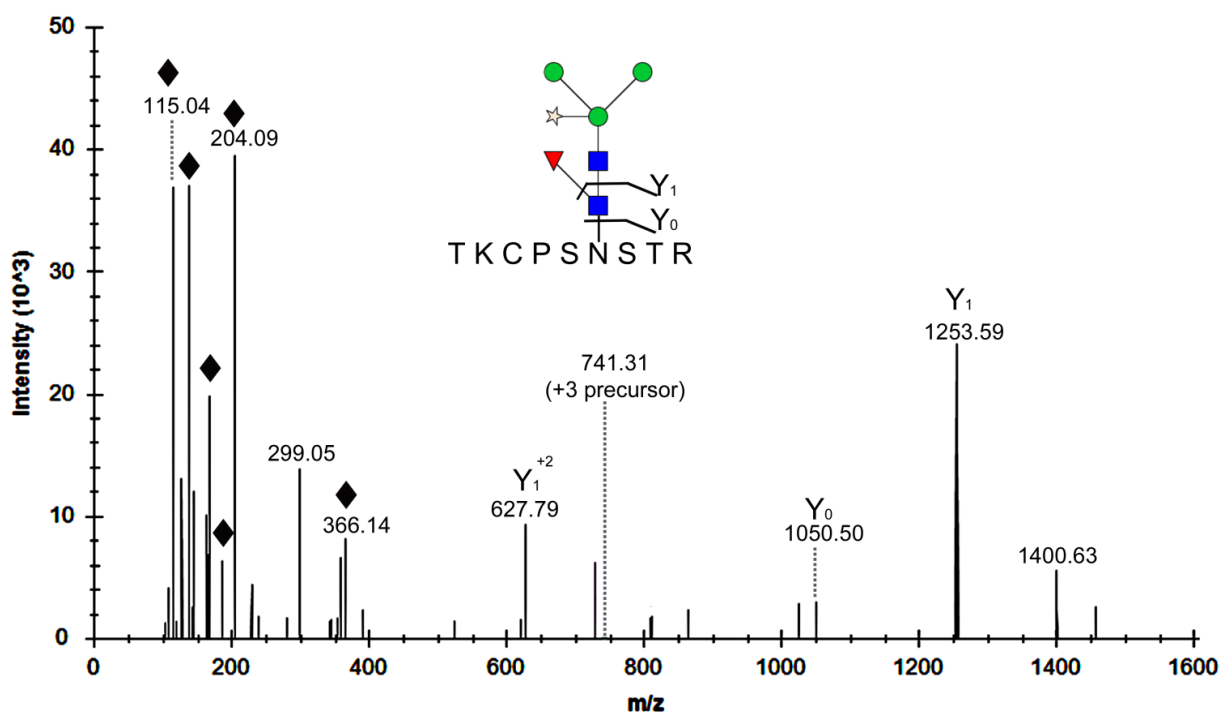


Figure 3.3. HCD fragmentation spectrum of a WPTP glycopeptide. The Y<sub>1</sub> and Y<sub>0</sub> ions are prominent features of the high *m/z* region, whereas, oxonium ions (diamonds) dominate the low *m/z* region. The assigned glycopeptide, with peptide sequence TKCPSNSTR and *N*-linked glycan, Man<sub>3</sub>XylFucGlcNAc<sub>2</sub>, glycan p (Table 3.1), is shown with reference to the cleavage sites for the Y<sub>1</sub> and Y<sub>0</sub> ions. The glycan structure was drawn with GlycanBuilder (Damerell et al., 2012).

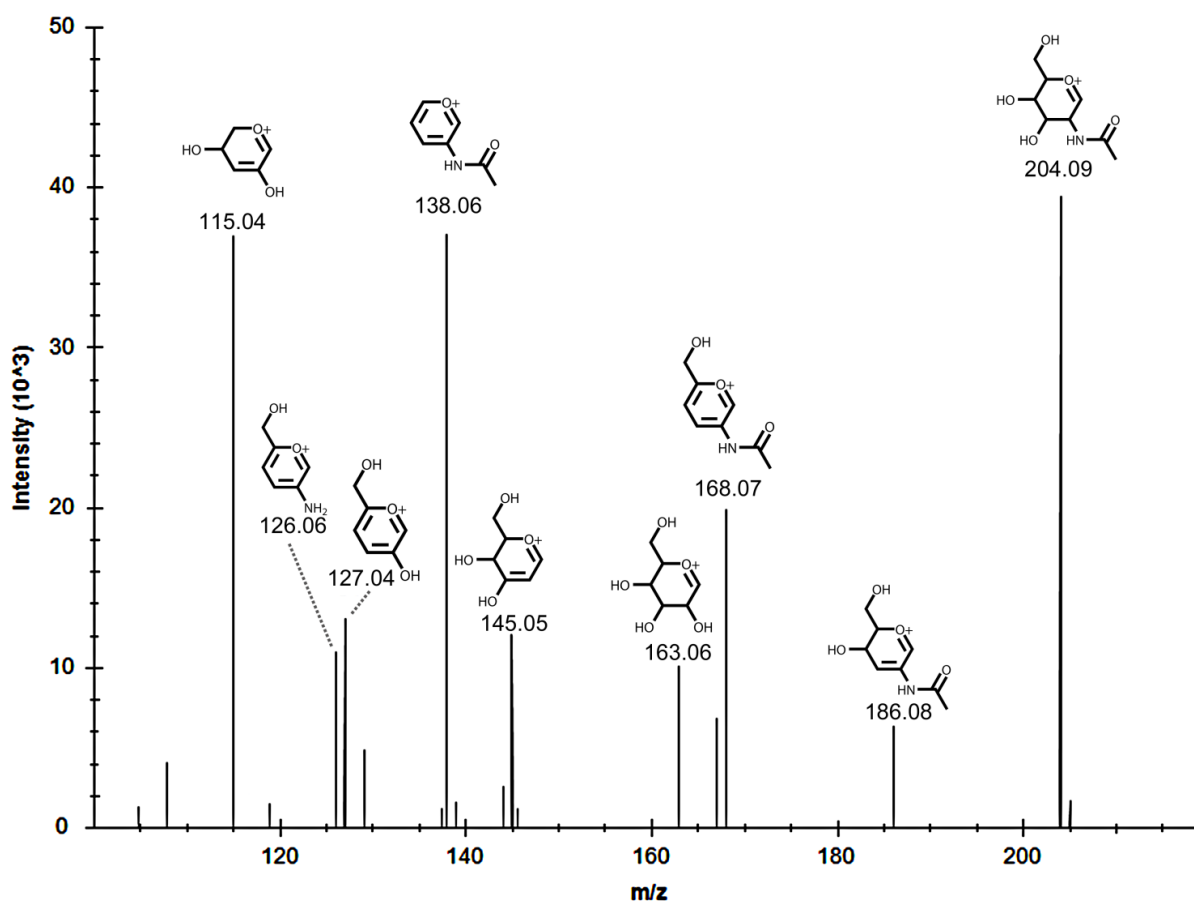


Figure 3.4. Oxonium ions in a WPTP glycopeptide HCD fragmentation spectrum. The oxonium ions include HexNac<sup>+</sup> (*m/z* 204.09) and its internal fragments (*m/z* 186.08, 168.07, 138.06, and 126.06), and Hex<sup>+</sup> (*m/z* 163.05) and its internal fragments (*m/z* 145.05, 127.04, and 115.04).

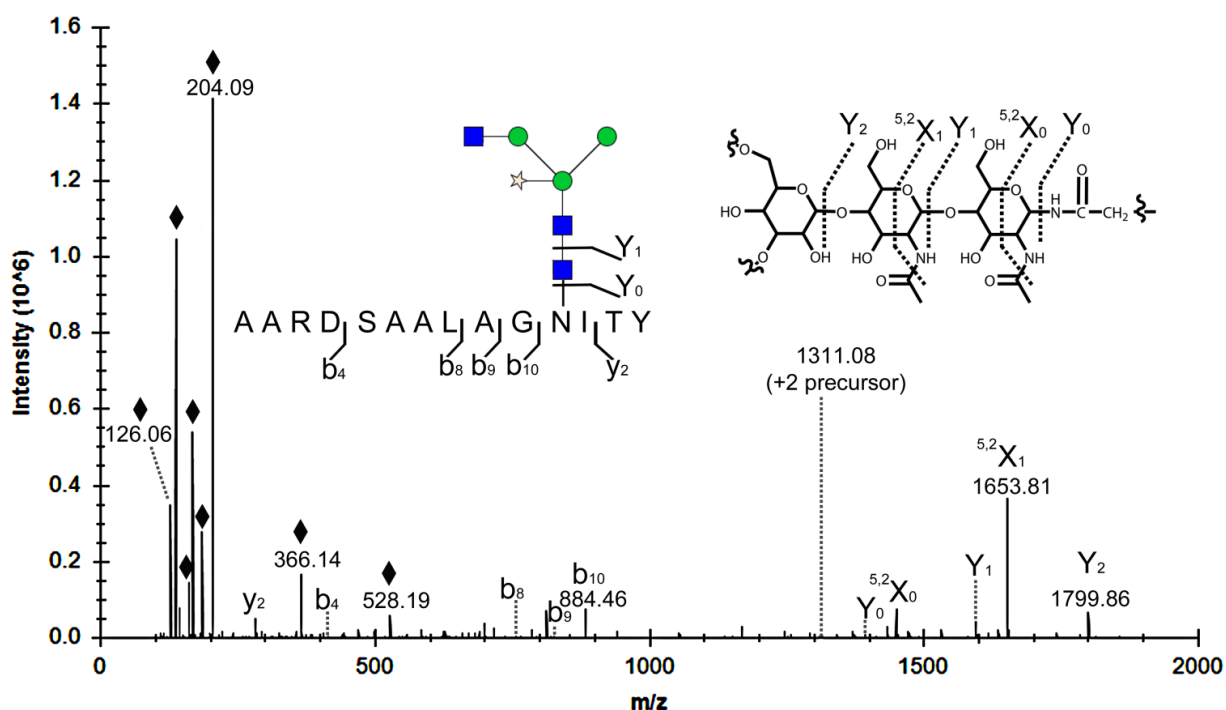


Figure 3.5. Cross-ring fragmentation of a WPTP glycopeptide. HCD fragmentation spectrum of this WPTP glycopeptide exhibited prominent cross-ring fragmentation. The  $^{2,5}X_1$  and  $^{2,5}X_0$ , as opposed to the  $Y_1$  and  $Y_0$ , ions were prominent features of the high  $m/z$  region. Oxonium ions (diamonds) dominated the low  $m/z$  region. The assigned glycopeptide, with peptide sequence AARDSAALAGNITY and *N*-linked glycan, GlcNAcMan<sub>3</sub>XylGlcNAc<sub>2</sub>, glycan t (Table 3.1), is shown with reference to the cleavage sites for the b- and y-peptide fragment ions as well as the  $Y_1$  and  $Y_0$  ions. The glycan structure was drawn with GlycanBuilder (Damerell et al., 2012). The chemical structure of the first 3 residues of the *N*-linked glycan is shown with proposed cross-ring fragments.

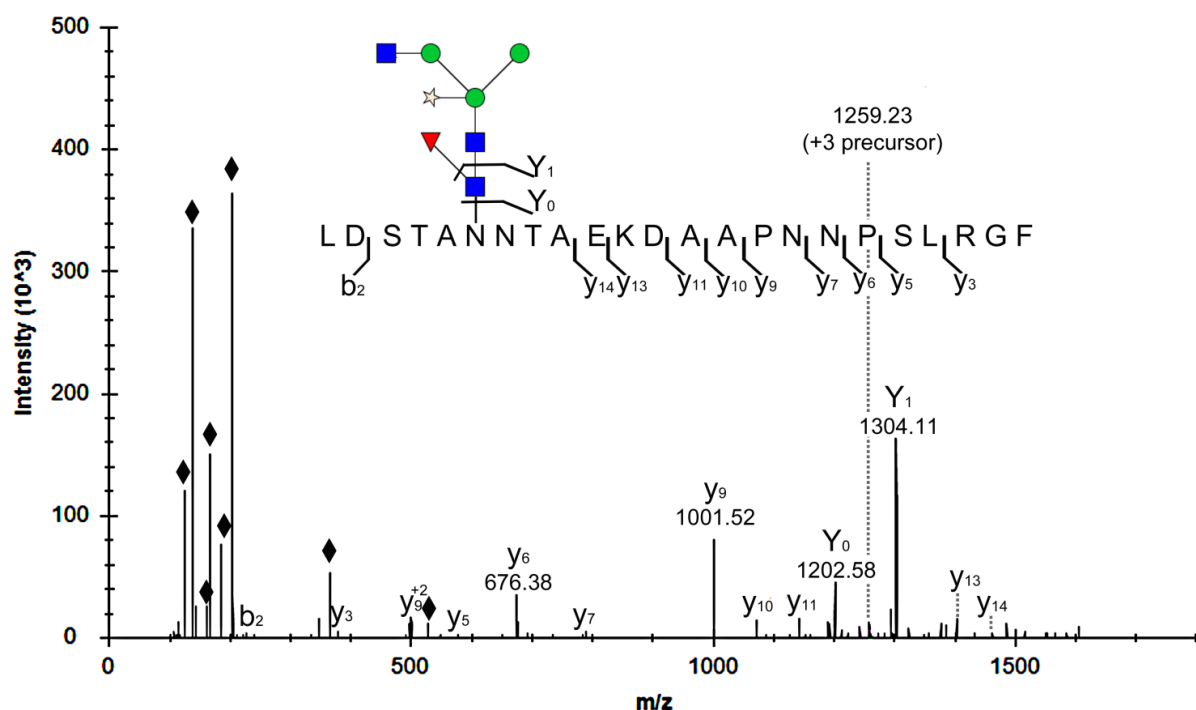


Figure 3.6. Extensive peptide backbone fragmentation of a WPTP glycopeptide. This is an HCD fragmentation spectrum of a WPTP glycopeptide that exhibited extensive peptide backbone fragmentation. In addition to the Y<sub>1</sub>, Y<sub>0</sub> ions, and oxonium ions (diamonds), several peptide backbone y-ions were detected. The assigned glycopeptide, with peptide sequence LDSTANNTAEKDAAPNNPSLRGF and *N*-linked glycan, GlcNAcMan<sub>3</sub>XylFucGlcNAc<sub>2</sub>, glycan aa ([Table 3.1](#)), is shown with reference to the cleavage sites for the b- and y-peptide fragment ions as well as the Y<sub>1</sub> and Y<sub>0</sub> ions. The glycan structure was drawn with GlycanBuilder (Damerell et al., 2012).

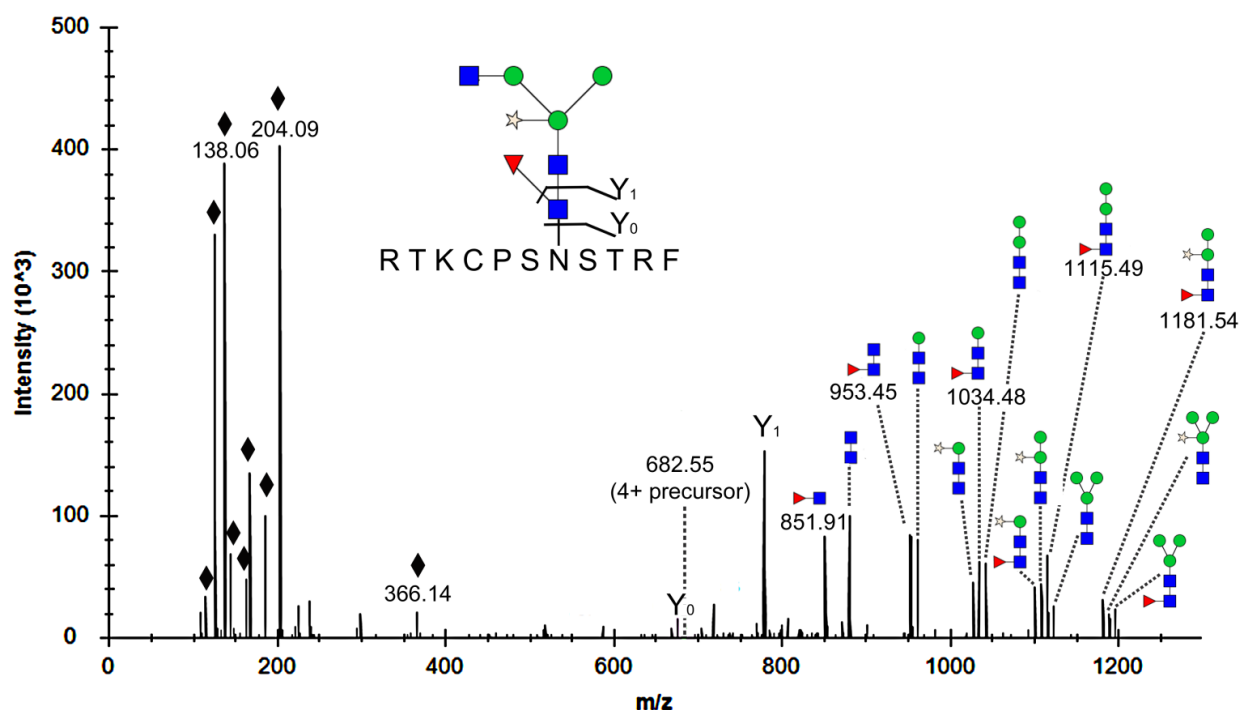


Figure 3.7. Extensive glycan fragmentation of a WPTP glycopeptide. This is an HCD fragmentation spectrum of a WPTP glycopeptide that exhibited extensive glycan fragmentation with the peptide still attached. In addition to the  $Y_1$ ,  $Y_0$  ions, and oxonium ions (diamonds), a number of glycan fragments, with peptide attached, were detected in the high  $m/z$  region, each of which is labeled with the glycan fragment that was attached to the peptide backbone. The assigned glycopeptide, with peptide sequence RTKCPSNSTRF and *N*-linked glycan, GlcNAcMan<sub>3</sub>XylGlcNAc<sub>2</sub>, glycan t (Table 3.1), is shown with reference to the cleavage sites for the  $Y_1$  and  $Y_0$  ions. The glycan structure was drawn with GlycanBuilder (Damerell et al., 2012). The chemical structure of the first 3 residues of the *N*-linked glycan is shown with proposed cross-ring fragments.

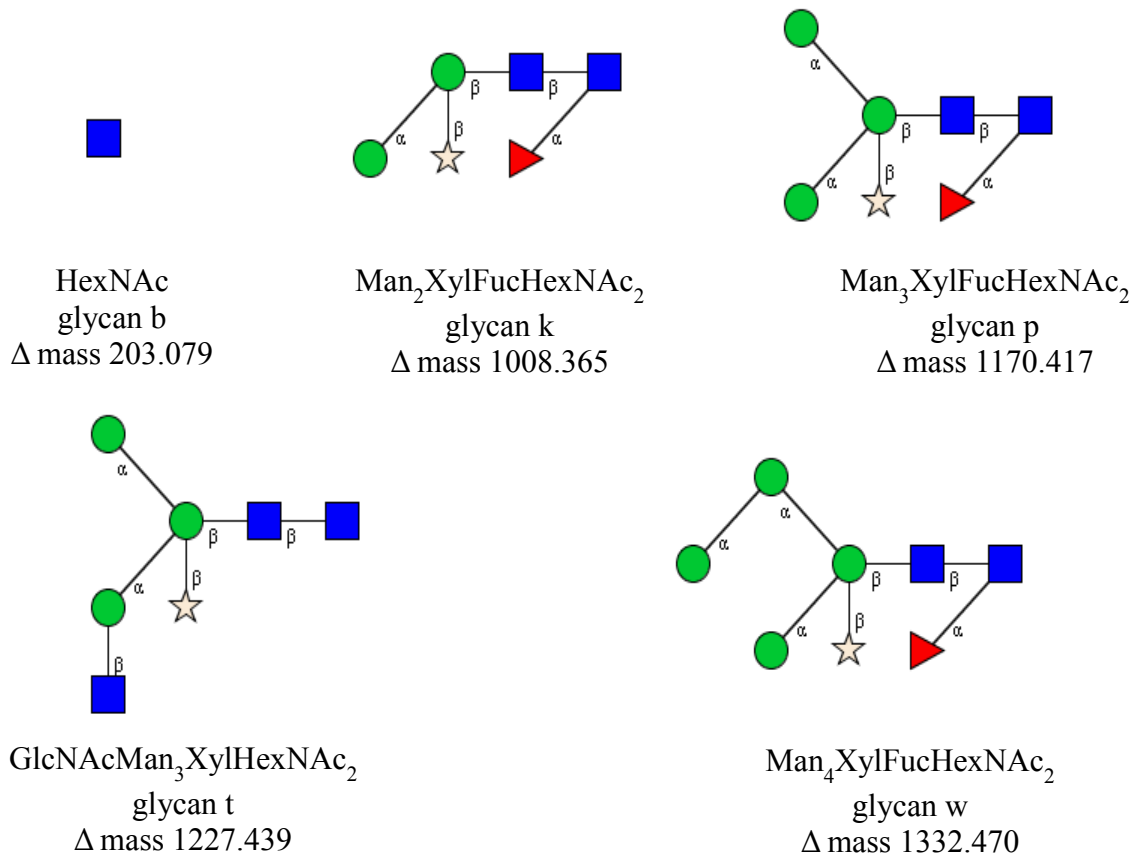


Figure 3.8. Proposed structures for the 5 major glycans observed on WPTP. The symbol designator refers to [Table 3.1](#). The glycan structure was drawn with GlycanBuilder (Damerell et al., 2012).

## CHAPTER 4. CONCLUSIONS AND FUTURE PERSPECTIVES

### PURPOSE

Windmill palm tree peroxidase has exceptional stability and unique substrate specificity; however, there was a knowledge gap concerning its structure. The purpose of this study was to fill in the gaps by determining the amino acid sequence and *N*-glycosylation. In the process, tools and a workflow for analysis of glycopeptide mass spectrometry data were developed.

### OBJECTIVES

1. Determine the amino acid sequence of windmill palm tree peroxidase
2. Develop tools and a workflow for analysis of windmill palm tree peroxidase glycopeptide mass spectrometry data
3. Determine the site-specific glycosylation and microheterogeneity of windmill palm tree peroxidase

**Objective 1.** The first essential knowledge gap to address was the amino acid sequence. As described in [Chapter 2](#), the complete amino acid sequence and partial glycosylation were determined by MALDI-top-down sequencing of native WPTP, MALDI-TOF/TOF MS/MS of WPTP tryptic peptides, and cDNA sequencing. The propeptide of WPTP contained *N*- and *C*-terminal signal sequences which contained 21 and 17 amino acid residues, respectively. Mature WPTP was 306 amino acids in length and its carbohydrate content ranged from 21 - 29%.

An important question to answer with this study was if the royal palm tree peroxidase (RPTP) X-ray crystal structure could be used as a model for WPTP. Both are anionic vacuolar peroxidases; based on the pI and a *C*-terminal signal peptide for secretion to the vacuole. WPTP and RPTP are 88% identical suggesting that it would be a good model for WPTP. The most important measure of quality for a homology model is sequence identity to the template (Bordoli et al., 2008). At the time of proof-reading, the X-ray crystal structure of *Chamaerops excelsa*

peroxidase was reported (Bernardes et al., 2015). WPTP and CEP are 98% identical, which means that the X-ray crystal structure of CEP should be another good model for WPTP.

Another important finding of the study presented in [Chapter 2](#) was the identification of two amino acid residues that may explain the differences in substrate specificity of WPTP and RPTP ([Table 2.1](#)). Data mining of the RPTP 3D structure revealed 69 residues of potential importance for substrate interaction. Of these, only seven were not identical. Of the seven, only two had significantly different properties. Residue 142 was hydrophobic in HRP, RPTP, and WPTP, except that it was large in HRP (Phe) and RPTP (Ile) but small in WPTP (Thr). This difference could be a major factor causing the large deficit in WPTP's reactivity toward ferulic acid. Residue 177 in WPTP was large, hydrophobic, and aromatic Phe, whereas it was small and polar Ser in RPTP. These mutations could be used to assess the validity of a 3D computational model of WPTP.

Aside from the new knowledge of WPTP, the workflow used in this section is an accessible approach for determining the amino acid sequence of proteins. The mRNA sequence gives the gene sequence, but could also include the sequences of signal peptides. Edman degradation can give the sequence of the true *N*-terminus, but not that for the *C*-terminus. This issue is addressed with MALDI-top down sequencing. The simple solid-phase disulfide bond reduction method presented in [Chapter 2](#) had not been reported elsewhere and it is a good method for dealing with top-down sequencing of proteins containing disulfide bonds.

**Objective 2.** Mass spectrometry was the main analytical tool for elucidating the glycosylation of WPTP. The most difficult hurdle encountered during my dissertation work was in figuring out how to analyze the glycopeptide mass spectrometry data. The available software for such analyses were designed for analysis of mammalian glycosylation. Additionally, most of the published studies did not give a detailed description of how the data in the study were analyzed; a problem noted in a recent review (Desaire 2013). I had initially intended to conduct a structure-activity study using a mutant of WPTP; however, upon encountering this major hurdle, it was clear to me that I had to reevaluate my plan and design a workflow and tools tailored to studying the glycosylation observed on WPTP.

To address this need, I developed a detailed workflow for analysis of plant glycopeptide mass spectrometry data ([Appendix A](#)). Included in the workflow are novel tools for glycan database construction, pGlycoFilter, and validation of glycopeptide assignment, gPSMvalidator. This issue was addressed in [Chapter 3](#).

During the course of this work, several papers came out demonstrating the need for understanding the function of complex glycosylation in plants (Zhang et al., 2009; Fanata et al., 2013; Barba-Espin et al., 2014; Strasser 2014; Sultana et al., 2015). It was previously thought that complex glycosylation did not play a major role in plants; as evidenced by the lack of phenotype associated with knock-out of complex glycosylation in *Arabidopsis thaliana* (von Schaewen et al., 1993). The only major phenotype occurred under salt stress (Kang et al., 2008). A recent study of the effect in rice, showed a completely different story. The mutation taking away the ability to form complex *N*-glycans was lethal (Fanata et al., 2013). As suggested by a recent review (Strasser 2014), new analytical platforms for studying plant glycoproteins, such as that developed for this study, are needed in the field of plant biology.

The benefits of the automated workflow go far beyond saving time. An automated workflow takes the analysis out of the hands of the expert and makes it accessible to non-experts. Automated analysis reduces error and results in comparability between laboratories (Deutsch et al., 2008). Manual interpretation should never be completely discarded. It should always be used as a confirmation of the assignments.

**Objective 3.** Mass spectrometry analysis of WPTP at the glycopeptide level allowed for detailed investigation of the glycosylation profile at each of the 13 sites of *N*-linked glycosylation. The microheterogeneity of glycosylation shows the different glycans observed at each site. At the most heterogeneous site, 30 different glycans were observed. Relative quantitation of the different glycans at each site revealed that 90% of each glycosylation site was occupied by 3 or fewer different glycans and that two of the sites are not occupied by a glycan about half of the time. Complex, hybrid, and high mannose-type sugars were observed; however, the major glycans were paucimannosidic which is characteristic of plant vacuolar glycoproteins (Yang et al., 1996; Gray and Montgomery 2006).

The major glycoforms identified in this objective could be modeled onto a 3D computational model of WPTP. A 3D structure of WPTP could reveal protein-glycan interactions such as hydrogen bonding. The validity of the model could be assessed by comparisons with the RPTP X-ray crystal structure, and another plant peroxidase crystal structure that was recently solved with well resolved glycans (Palm et al., 2014).

## FUTURE WORK

The next step would be to make use of the structure data and test a hypothesis. In my opinion, the most exciting hypothesis to test would be that WPTP's exceptional stability and catalysis are directly related to glycosylation and quaternary structure. The most important consideration for this step as far as experimental design is concerned is the selection of an expression system.

A lot of evidence points to WPTP having a quaternary structure, the most important being the 6.5 Ca<sup>2+</sup> ions. The presence of a metal cation is common at protein-protein interfaces (Song et al., 2014). Interestingly, palm peroxidases appear to operate as homodimers under normal conditions, a feature that may account for the exceptional stability and fast catalysis (Zamorano et al., 2008; Zamorano et al., 2009). Also, disruption of hydrogen bonding and hydrophobic interactions affects WPTP interactions with substrate in interesting and unsuspected ways (Caramyshev et al., 2006). Hydrogen bonds and salt bridges are common interactions that stabilize protein-protein interactions.

In depth understanding of the roles of glycosylation and quaternary structure in WPTP and other peroxidases requires departure from standard expression systems such as *Escherichia coli* which fail to produce native structures. Typically, plant peroxidase structure function research is conducted on the recombinant enzyme produced in bacteria or yeast which affords high expression, ease of purification, and the ability to produce mutant forms of the enzyme. Such non-native enzymes are not suitable for study of glycosylation and quaternary structures.

Thirty one HRP structures deposited in the protein data bank (PDB) were all expressed in *E. coli*. Nine class III peroxidases from other plants were also in the PDB. Of those nine, five

were produced in *E. coli* and four were from the native source. Of the 36 *E. coli* derived plant peroxidase structures, all were monomers. Two of the four native derived peroxidases were dimers. This provides evidence of the importance of studying the native glycosylated enzyme.

The obvious choice is to use a plant expression system; however, expression in typical plant systems would be tedious and practically infeasible for studies involving mutagenesis of WPTP which contains 13 *N*-glycosylation sites, where stable transformation requires from several months to a year. Transient expression in tobacco would be a viable alternative (Leuzinger et al., 2013). In transient expression, the gene of interest is introduced to the developed plant system and instead of integrating into the genome; the gene of interest is directly manufactured in the developed plant tissue. Overall, a protein can be harvested and purified in less than 2 months. Such a short turnaround would allow for mutagenesis experiments as well. Expression yields in a 10 mg/kg range would be ample for biochemical and structural studies. Peroxidases from phylogenetically divergent plants tend to have similar glycosylation profiles. The dominant glycan in soybean peroxidase (Gray and Montgomery 2006), HRP (Yang et al., 1996), peanut peroxidase (Zhang et al., 2004), and barley peroxidase (Henricksen et al., 1998) is GlcNAc<sub>2</sub>Man<sub>3</sub>FucXyl and other glycans are paucimannosidic in general. Therefore, I hypothesize that recombinant WPTP transiently expressed in tobacco leaves would be highly native-like.

## **FUTURE PERSPECTIVES**

New analytical methods are needed for the emerging field of plant glycoproteomics (Ruiz-May et al., 2012; Strasser 2014). The novel methods developed in this dissertation will be useful for the study of other important plant glycoproteins. The tools and workflow for analysis of plant glycoprotein mass spectrometry data are accessible to plant biologists. The novel use of transient expression to produce materials to study the structure-activity relationships in plant glycoproteins would enable in depth study of important glycoproteins identified in plant glycoproteomics studies. The new knowledge of WPTP's structure will enable in depth analysis of the roles of glycosylation in this unique enzyme.

## APPENDIX A: WORKFLOW FOR ANALYSIS OF PLANT GLYCOPEPTIDE MASS SPECTROMETRY DATA

### **Step 1: Protein identification**

The raw mass spectrometry files were converted to mzML format using the ‘msconvert’ tool in ProteoWizard (v.3.0.5006; Chambers et al., 2013; [proteowizard.sourceforge.net](http://proteowizard.sourceforge.net)). The data were filtered to only include MS level 2 and the ‘precursorRefine’ filter was applied. This was executed with the command: [msconvert.exe \*.RAW --filter precursorRefine --filter “msLevel 2-”].

The proteins were identified with MyriMatch (v. 2.1.138; Tabb et al., 2007; [fenchurch.mc.vanderbilt.edu](http://fenchurch.mc.vanderbilt.edu)), a database search engine for proteomics. The configuration for this and other proteomics searches is in [Appendix B](#). The goals of the protein database were to 1) contain the target protein sequence, 2) contain a set of potential protein contaminants, and 3) to be of sufficient size for proper calculation of the false discovery rate [i.e., to be of a sufficient size so that the reverse sequence can be detected for false discovery rate estimation (Elias and Gygi, 2010)]. The protein database was constructed on November 11, 2013. It contained the target protein sequence obtained from mRNA sequencing (WPP\_mature), a compilation of 248 proteins that are potential contaminants in proteomics samples made available from the Max Planck Institute of Biochemistry (<http://maxquant.org/contaminants.zip>), and sequences from UniProt.org including all known proteins from *Trachycarpus fortunei* and *Armoracia rusticana* and the proteome of *Magnaporthe oryzae* (strain 70-15 / ATCC MYA-4617 / FGSC 8958) (Rice blast fungus) (*Pyricularia oryzae*). The database had a total of 25,930 proteins.

The raw identifications for each sample type, i.e., unenriched chymotrypsin digested WPTP, unenriched trypsin digested WPTP, and enriched chymotrypsin digested WPTP, were assembled in IDPicker 3.0 (v. 3.0.545 64-bit; Zhang et al., 2008; [fenchurch.mc.vanderbilt.edu](http://fenchurch.mc.vanderbilt.edu)) with default settings. A subset database was exported with decoy sequences. This subset dataset was used for all subsequent glycopeptide identification searches.

## **Step 2: Glycan identification**

Glycan identification involved making a list of precursor masses, making a list of exact peptide masses from WPTP, using GlycoMod (Cooper et al., 2001; [web.expasy.org/glycomod](http://web.expasy.org/glycomod)) to identify the potential glycans present, and applying pGlycoFilter to test the biological feasibility of the predicted glycan compositions.

First, the 'msaccess' tool (ProteoWizard) was used to write a table containing information about the precursor  $m/z$  values from the mass spectrometry data. This was executed using the command: [msaccess.exe \*.mzML -x "spectrum\_table delimiter=comma"]. GlycoMod requires that the observed precursor mass or singly charged mass be used. This was calculated in R, using the function SpectrumTable ([gPSMvalidator\\_functions.R](#)). The output of this function is a text file containing the observed precursor masses each on a separate line.

Next, the list of peptide masses was obtained first by *in silico* digest of the target protein, WPTP, with the 'chainsaw' tool (ProteoWizard). This was executed using the command: [chainsaw.exe wpp\_mature.fasta --cleavageAgentName Chymotrypsin --numMissedCleavages 2 --specificity fully]. The fasta file contained only the WPTP protein sequence. The R function ChainSaw ([gPSMvalidator\\_functions.R](#)) was used to filter for peptides that contain Asn and to add the mass of chemical modifications such as that for carbamidomethylation. The output of this function is a text file containing the exact peptide masses each on a separate line.

Next, all possible glycan compositions were identified using the online tool GlycoMod ([web.expasy.org/glycomod](http://web.expasy.org/glycomod)) in the Mozilla Firefox web browser. For the 'experimental masses,' the file generated with the SpectrumTable function was uploaded. For the 'set of unmodified peptide masses,' the peptide masses generated with the ChainSaw function were used. All mass values were monoisotopic, with a mass tolerance of 10 ppm. The ion mode was neutral. The monosaccharide residues were underivatized with the following allowed compositions: Hexose 0-9, HexNAc 1-4, Deoxyhexose 0-3, and Pentose 0-1. The compositions reported in UniCarbKB were listed separately.

In order to accomplish the last step, the GlycoMod results were copied and pasted into an Excel spreadsheet. The contents of the top row of the spreadsheet were: glycoform mass,  $\Delta$  mass (ppm), structure, type, peptide mass [M], peptide sequence, theoretical glycopeptide mass, mod.,

and Links. Lastly, pGlycoFilter, written in Excel macros ([pGlycoFilter macros](#)), was run to test the biological feasibility of the predicted glycan compositions. The results from the unenriched chymotrypsin and trypsin digest and the enriched chymotrypsin digest were combined into a single list. The duplicates were removed and the non-redundant list of glycan masses was used as ‘preferred delta masses’ in the MyriMatch, TagRecon, and MassRecon configuration files.

### **Step 3: Glycopeptide identification**

Glycopeptides were identified using several proteomics search engines and the subset database produced in step 1. The proteomics software included MyriMatch (v. 2.1.138; Tabb et al., 2007, a database search engine, DirecTag-TagRecon (v. 1.4.71; Dasari et al., 2010), a tag-based search engine, and MassRecon (v. 1.4.71; Dasari et al., 2010), a mass-based search engine. Alternatively, glycopeptides were identified using a modified GlycoMod search.

For the DirecTag-TagRecon search, DirecTag was used to generate a ‘tags’ file for each of the mass spectrometry data files. Next, the tags were used in a TagRecon search. MyriMatch and MassRecon searches were also conducted. The configuration files for each of these searches is in [Appendix B](#).

For each sample type (e.g., unenriched chymotrypsin, enriched chymotrypsin, or unenriched trypsin digested), the raw identifications generated from each of the proteomics searches were assembled in IDPicker 3.0 using the default import settings. For the WPTP identifications, the peptide spectrum matches (PSMs) were exported as a .tsv file. For the modifications, results with a minimum of 1 were rounded to 0.001 Da and exported as a .tsv file.

For glycopeptide identification using GlycoMod, the same conditions used in ‘Step 2: Glycan identification’ were used except that the mass tolerance was 15 ppm and one peptide mass was searched at a time.

### **Step 4: Validation of glycopeptide spectrum matches (gPSMs) using gPSMvalidator**

In order to perform the validation step, a file for each tandem mass spectrum had to be written. This was accomplished with the ‘msaccess’ tool (ProteoWizard), using a command such as: [msaccess.exe \*.mzML -x “binary sn=1783,19887” -o MS2Data\_Chym1\_ELUTE].

Next, gPSMvalidator, written in R, was used to validate the gPSMs. One script was written to accept the results of the proteomics-based glycopeptide searches ([gPSMvalidator\\_proteomics.R](#)) and another was written to accept the results of the glycoproteomics-based search ([gPSMvalidator\\_glycoproteomics.R](#)). Both functions accept gPSMs that are within a 15 ppm mass tolerance, and contain at least 2 oxonium marker ions, and one Y<sub>1</sub> or Y<sub>0</sub> ion. The output of gPSMvalidator is a .pdf file of annotated spectra and a table of all validated gPSMs. These annotated spectra are provided as supplementary data ([Annotated\\_gPSMs](#)).

These results were further confirmed by manual validation. The isotope cluster of the Y<sub>1</sub> and or Y<sub>0</sub> ion(s) were checked for each gPSM using the spectrum viewer in IDPicker 3.0 or the 'seeMS' tool (ProteoWizard). The isotope clusters were evaluated based on whether the picked peak was in the correct charge state and whether the monoisotopic peak was chosen. After manual validation, it was clear that some gPSMs could be reliably accepted without manual inspection. Annotated gPSMs that clearly showed the Y<sub>1</sub> and Y<sub>0</sub> ions as the two most prominent peaks in the high *m/z* region were accepted automatically. For low intensity spectra, if the Y<sub>1</sub>/Y<sub>0</sub> pair was observed in more than one charge state, the gPSM was automatically accepted.

### **Step 5A: Compilation of validated gPSMs**

The gPSMvalidator tool excludes all spectrum matches to peptides that do not bear a glycan. An R script was written and used in order to include peptides with unoccupied glycosylation sites ([findUnoccupied.R](#)). The spectrum matches of peptides with unoccupied glycosylation sites were manually verified.

All of the gPSMs were compiled into one .csv file. These included the results from the proteomics searches, the glycoproteomics searches, and the list of unoccupied peptides. Next, a script written in R was used to compile a list sortable by glycosylation site, to count the number of spectra identified for each peptide backbone, as well as the number of glycoforms at each glycosylation site ([glycosylationProfileTables.R](#)).

### **Step 5B: Relative abundance of glycoforms at each glycosylation site**

The RAW files were converted to mzML format using the ‘msconvert’ tool (ProteoWizard). The data were filtered to only include MS level 1 using the command: [msconvert.exe \*.RAW --filter “msLevel 1-”].

The selected ion chromatograms (SICs) for the precursor  $m/z$  of each identified glycopeptide used for quantitation was obtained with the ‘msaccess’ tool (ProteoWizard). The monoisotopic peak was selected with a radius of 5 ppm. This was carried out with the following command: [msaccess.exe \*.mzML -c quantConfig.txt, where the quantConfig.txt file contained the following command for each precursor  $m/z$  on a separate line: “sic mzCenter=<mz> radius=5 radiusUnits=ppm delimiter=comma].

The relative abundance of each glycan at each glycosylation site was calculated from the SICs using a script written in R ([quantitationGlycoforms.R](#)). The results were the summed intensity of the SIC for each precursor  $m/z$  during the period of time that it eluted. There was not 100% specificity, since any peak within 5 ppm was detected. For the most part, for each glycoform, there was a major spike in the intensity over a short period time, typically 0.5 minutes. For some glycoforms, there was one major peak in that specific time period and some noise peaks during other time periods. For other glycoforms, the results were more ambiguous. Noise was determined by examining the MS1 spectra. Actual signal was defined as the intensity being due to the monoisotopic peak of the precursor  $m/z$ . The charge state of the isotope pattern was additional evidence that the signal was due to the glycoform precursor  $m/z$ . Signal was only considered valid if more than one peak in the isotope cluster could be detected. Another level of validation was the MS2 fragmentation spectrum being acquired during that time period.

## APPENDIX B: CONFIGURATION FILES FOR PROTEOMICS-BASED PROTEIN AND GLYCOPEPTIDE IDENTIFICATION

### **Step 1: Protein identification with MyriMatch**

```
StaticMods = "C 57.021"  
DynamicMods = "M ^ 15.9949 (Q * -17.026 N!{P}[ST] * 203.079 N!{P}[ST] * 1008.365  
N!{P}[ST] * 1170.417"  
MinTerminiCleavages = 1  
CleavageRules = "Chymotrypsin" or "Trypsin/P"  
MaxMissedCleavages = 2  
MaxDynamicMods = 2  
ProteinDatabase = MediumDatabase.fasta  
  
DecoyPrefix = "rev_"  
NumChargeStates = 4  
OutputFormat = "pepXML"  
OutputSuffix = "_MM-proteinID_Medium"  
TicCutoffPercentage = 0.98  
MaxPeakCount = 150  
MaxResultRank = 2  
MinPeptideMass = 0 Da  
MaxPeptideMass = 10000 Da  
MinPeptideLength = 2  
MaxPeptideLength = 75  
UseSmartPlusThreeModel = false  
ComputeXCorr = true  
UseMultipleProcessors = true  
  
FragmentationAutoRule = true  
PrecursorMzToleranceRule = "mono"  
MonoPrecursorMzTolerance = 10 ppm  
MonoisotopeAdjustmentSet = [-1,1]  
FragmentMzTolerance = 0.02 m/z
```

### **Step 2: Glycopeptide identification: DirectTag**

```
PrecursorMzTolerance = 0.015 m/z  
FragmentMzTolerance = 0.02 m/z  
ComplementMzTolerance = 0.02 m/z  
IsotopeMzTolerance = 0.01 m/z  
DuplicateSpectra = true  
UseChargeStateFromMS = true  
NumChargeStates = 4
```

TicCutoffPercentage = 0.98  
MaxPeakCount = 100  
TagLength = 3  
MaxTagCount = 50  
StaticMods = "C 57.021"

## **Step 2: Glycopeptide identification: TagRecon**

ProteinDatabase = rep1\_subset.fasta or rep2\_subset.fasta  
StaticMods = "C 57.021"

DynamicMods = "M ^ 15.9949 (Q \* -17.026 N!{P}[ST] \* 203.079 N!{P}[ST] \* 406.159  
N!{P}[ST] \* 552.217 N!{P}[ST] \* 568.212 N!{P}[ST] \* 700.254 N!{P}[ST] \* 714.27  
N!{P}[ST] \* 730.264 N!{P}[ST] \* 846.312 N!{P}[ST] \* 862.307 N!{P}[ST] \* 876.322  
N!{P}[ST] \* 892.317 N!{P}[ST] \* 933.344 N!{P}[ST] \* 1008.365 N!{P}[ST] \* 1024.36  
N!{P}[ST] \* 1038.375 N!{P}[ST] \* 1054.37 N!{P}[ST] \* 1065.386 N!{P}[ST] \* 1079.402  
N!{P}[ST] \* 1095.397 N!{P}[ST] \* 1170.417 N!{P}[ST] \* 1186.412 N!{P}[ST] \* 1200.428  
N!{P}[ST] \* 1211.444 N!{P}[ST] \* 1216.423 N!{P}[ST] \* 1225.46 N!{P}[ST] \* 1227.439  
N!{P}[ST] \* 1241.455 N!{P}[ST] \* 1257.449 N!{P}[ST] \* 1298.476 N!{P}[ST] \* 1332.47  
N!{P}[ST] \* 1348.465 N!{P}[ST] \* 1357.502 N!{P}[ST] \* 1362.481 N!{P}[ST] \* 1373.497  
N!{P}[ST] \* 1378.476 N!{P}[ST] \* 1387.512 N!{P}[ST] \* 1389.492 N!{P}[ST] \* 1403.507  
N!{P}[ST] \* 1419.502 N!{P}[ST] \* 1430.518 N!{P}[ST] \* 1444.534 N!{P}[ST] \* 1460.529  
N!{P}[ST] \* 1494.523 N!{P}[ST] \* 1510.518 N!{P}[ST] \* 1519.555 N!{P}[ST] \* 1535.55  
N!{P}[ST] \* 1540.528 N!{P}[ST] \* 1549.565 N!{P}[ST] \* 1551.545 N!{P}[ST] \* 1565.56  
N!{P}[ST] \* 1576.576 N!{P}[ST] \* 1581.555 N!{P}[ST] \* 1590.592 N!{P}[ST] \* 1592.571  
N!{P}[ST] \* 1606.587 N!{P}[ST] \* 1622.582 N!{P}[ST] \* 1681.608 N!{P}[ST] \* 1697.602  
N!{P}[ST] \* 1702.581 N!{P}[ST] \* 1711.618 N!{P}[ST] \* 1713.597 N!{P}[ST] \* 1722.634  
N!{P}[ST] \* 1727.613 N!{P}[ST] \* 1736.65 N!{P}[ST] \* 1738.629 N!{P}[ST] \* 1752.645  
N!{P}[ST] \* 1754.624 N!{P}[ST] \* 1768.64 N!{P}[ST] \* 1843.66 N!{P}[ST] \* 1859.655  
N!{P}[ST] \* 1864.634 N!{P}[ST] \* 1868.692 N!{P}[ST] \* 1873.671 N!{P}[ST] \* 1884.687  
N!{P}[ST] \* 1898.703 N!{P}[ST] \* 1900.682 N!{P}[ST] \* 1914.697 N!{P}[ST] \* 2005.713  
N!{P}[ST] \* 2030.745 N!{P}[ST] \* 2046.74 N!{P}[ST] \* 2060.755 N!{P}[ST] \* 2192.798"

MinTerminiCleavages = 2

CleavageRules = "Chymotrypsin" or "Trypsin/P"

MaxMissedCleavages = 2

MaxDynamicMods = 2

MassReconMode = false

UnimodXML = "C:\Program Files (x86)\BumberShoot\bumberdash-bin-windows-x86-vc100-  
release-1\_4\_115\lib\Bumbershoot\TagRecon\unimod.xml"

Blosum = "C:\Program Files (x86)\BumberShoot\bumberdash-bin-windows-x86-vc100-release-  
1\_4\_115\lib\Bumbershoot\TagRecon\blosum62.fas"

DecoyPrefix = "rev\_"

UseChargeStateFromMS = true

NumChargeStates = 4

OutputFormat= "pepXML"  
OutputSuffix = "\_specific\_TagRecon\_allGlycans"  
TicCutoffPercentage = 0.98  
MaxPeakCount = 200  
MaxResultRank = 2  
MinPeptideMass = 0 Da  
MaxPeptideMass = 10000 Da  
MinPeptideLength = 5  
MaxPeptideLength = 75  
UseSmartPlusThreeModel = false  
ComputeXCorr = true  
UseMultipleProcessors = true

FragmentationAutoRule = true  
PrecursorMzTolerance = 0.015 m/z  
FragmentMzTolerance = 0.02 m/z  
NTerminusMzTolerance = 0.02 m/z  
CTerminusMzTolerance = 0.02 m/z  
UseAvgMassOfSequences = false  
DuplicateSpectra = true  
UseNETAdjustment = false

## **Step 2: Glycopeptide identification: MyriMatch**

StaticMods = "C 57.021"

DynamicMods = "M ^ 15.9949 (Q \* -17.026 N!{P}[ST] \* 203.079 N!{P}[ST] \* 406.159  
N!{P}[ST] \* 552.217 N!{P}[ST] \* 568.212 N!{P}[ST] \* 700.254 N!{P}[ST] \* 714.27  
N!{P}[ST] \* 730.264 N!{P}[ST] \* 846.312 N!{P}[ST] \* 862.307 N!{P}[ST] \* 876.322  
N!{P}[ST] \* 892.317 N!{P}[ST] \* 933.344 N!{P}[ST] \* 1008.365 N!{P}[ST] \* 1024.36  
N!{P}[ST] \* 1038.375 N!{P}[ST] \* 1054.37 N!{P}[ST] \* 1065.386 N!{P}[ST] \* 1079.402  
N!{P}[ST] \* 1095.397 N!{P}[ST] \* 1170.417 N!{P}[ST] \* 1186.412 N!{P}[ST] \* 1200.428  
N!{P}[ST] \* 1211.444 N!{P}[ST] \* 1216.423 N!{P}[ST] \* 1225.46 N!{P}[ST] \* 1227.439  
N!{P}[ST] \* 1241.455 N!{P}[ST] \* 1257.449 N!{P}[ST] \* 1298.476 N!{P}[ST] \* 1332.47  
N!{P}[ST] \* 1348.465 N!{P}[ST] \* 1357.502 N!{P}[ST] \* 1362.481 N!{P}[ST] \* 1373.497  
N!{P}[ST] \* 1378.476 N!{P}[ST] \* 1387.512 N!{P}[ST] \* 1389.492 N!{P}[ST] \* 1403.507  
N!{P}[ST] \* 1419.502 N!{P}[ST] \* 1430.518 N!{P}[ST] \* 1444.534 N!{P}[ST] \* 1460.529  
N!{P}[ST] \* 1494.523 N!{P}[ST] \* 1510.518 N!{P}[ST] \* 1519.555 N!{P}[ST] \* 1535.55  
N!{P}[ST] \* 1540.528 N!{P}[ST] \* 1549.565 N!{P}[ST] \* 1551.545 N!{P}[ST] \* 1565.56  
N!{P}[ST] \* 1576.576 N!{P}[ST] \* 1581.555 N!{P}[ST] \* 1590.592 N!{P}[ST] \* 1592.571  
N!{P}[ST] \* 1606.587 N!{P}[ST] \* 1622.582 N!{P}[ST] \* 1681.608 N!{P}[ST] \* 1697.602  
N!{P}[ST] \* 1702.581 N!{P}[ST] \* 1711.618 N!{P}[ST] \* 1713.597 N!{P}[ST] \* 1722.634  
N!{P}[ST] \* 1727.613 N!{P}[ST] \* 1736.65 N!{P}[ST] \* 1738.629 N!{P}[ST] \* 1752.645  
N!{P}[ST] \* 1754.624 N!{P}[ST] \* 1768.64 N!{P}[ST] \* 1843.66 N!{P}[ST] \* 1859.655  
N!{P}[ST] \* 1864.634 N!{P}[ST] \* 1868.692 N!{P}[ST] \* 1873.671 N!{P}[ST] \* 1884.687

N!{P}[ST] \* 1898.703 N!{P}[ST] \* 1900.682 N!{P}[ST] \* 1914.697 N!{P}[ST] \* 2005.713  
N!{P}[ST] \* 2030.745 N!{P}[ST] \* 2046.74 N!{P}[ST] \* 2060.755 N!{P}[ST] \* 2192.798"  
MinTerminiCleavages = 2  
CleavageRules = "Chymotrypsin" or "Trypsin/P"  
MaxMissedCleavages = 2  
MaxDynamicMods = 2  
ProteinDatabase = rep1\_subset.fasta or rep2\_subset.fasta

DecoyPrefix = "rev\_"  
NumChargeStates = 4  
OutputFormat = "pepXML"  
OutputSuffix = "\_specific\_MM-glycopeptideID"  
TicCutoffPercentage = 0.98  
MaxPeakCount = 150  
MaxResultRank = 2  
MinPeptideMass = 0 Da  
MaxPeptideMass = 10000 Da  
MinPeptideLength = 2  
MaxPeptideLength = 75  
UseSmartPlusThreeModel = false  
ComputeXCorr = true  
UseMultipleProcessors = true

FragmentationAutoRule = true  
PrecursorMzToleranceRule = "mono"  
MonoPrecursorMzTolerance = 10 ppm  
MonoisotopeAdjustmentSet = [-1,1]  
FragmentMzTolerance = 0.02 m/z

## **Step 2: Glycopeptide identification: MassRecon**

ProteinDatabase = rep1\_subset.fasta or rep2\_subset.fasta  
StaticMods = "C 57.021"  
DynamicMods = "M ^ 15.9949 (Q \* -17.026 N!{P}[ST] \* 203.079 N!{P}[ST] \* 406.159  
N!{P}[ST] \* 552.217 N!{P}[ST] \* 568.212 N!{P}[ST] \* 700.254 N!{P}[ST] \* 714.27  
N!{P}[ST] \* 730.264 N!{P}[ST] \* 846.312 N!{P}[ST] \* 862.307 N!{P}[ST] \* 876.322  
N!{P}[ST] \* 892.317 N!{P}[ST] \* 933.344 N!{P}[ST] \* 1008.365 N!{P}[ST] \* 1024.36  
N!{P}[ST] \* 1038.375 N!{P}[ST] \* 1054.37 N!{P}[ST] \* 1065.386 N!{P}[ST] \* 1079.402  
N!{P}[ST] \* 1095.397 N!{P}[ST] \* 1170.417 N!{P}[ST] \* 1186.412 N!{P}[ST] \* 1200.428  
N!{P}[ST] \* 1211.444 N!{P}[ST] \* 1216.423 N!{P}[ST] \* 1225.46 N!{P}[ST] \* 1227.439  
N!{P}[ST] \* 1241.455 N!{P}[ST] \* 1257.449 N!{P}[ST] \* 1298.476 N!{P}[ST] \* 1332.47  
N!{P}[ST] \* 1348.465 N!{P}[ST] \* 1357.502 N!{P}[ST] \* 1362.481 N!{P}[ST] \* 1373.497  
N!{P}[ST] \* 1378.476 N!{P}[ST] \* 1387.512 N!{P}[ST] \* 1389.492 N!{P}[ST] \* 1403.507  
N!{P}[ST] \* 1419.502 N!{P}[ST] \* 1430.518 N!{P}[ST] \* 1444.534 N!{P}[ST] \* 1460.529  
N!{P}[ST] \* 1494.523 N!{P}[ST] \* 1510.518 N!{P}[ST] \* 1519.555 N!{P}[ST] \* 1535.55

N!{P}[ST] \* 1540.528 N!{P}[ST] \* 1549.565 N!{P}[ST] \* 1551.545 N!{P}[ST] \* 1565.56  
N!{P}[ST] \* 1576.576 N!{P}[ST] \* 1581.555 N!{P}[ST] \* 1590.592 N!{P}[ST] \* 1592.571  
N!{P}[ST] \* 1606.587 N!{P}[ST] \* 1622.582 N!{P}[ST] \* 1681.608 N!{P}[ST] \* 1697.602  
N!{P}[ST] \* 1702.581 N!{P}[ST] \* 1711.618 N!{P}[ST] \* 1713.597 N!{P}[ST] \* 1722.634  
N!{P}[ST] \* 1727.613 N!{P}[ST] \* 1736.65 N!{P}[ST] \* 1738.629 N!{P}[ST] \* 1752.645  
N!{P}[ST] \* 1754.624 N!{P}[ST] \* 1768.64 N!{P}[ST] \* 1843.66 N!{P}[ST] \* 1859.655  
N!{P}[ST] \* 1864.634 N!{P}[ST] \* 1868.692 N!{P}[ST] \* 1873.671 N!{P}[ST] \* 1884.687  
N!{P}[ST] \* 1898.703 N!{P}[ST] \* 1900.682 N!{P}[ST] \* 1914.697 N!{P}[ST] \* 2005.713  
N!{P}[ST] \* 2030.745 N!{P}[ST] \* 2046.74 N!{P}[ST] \* 2060.755 N!{P}[ST] \* 2192.798"

MinTerminiCleavages = 2

CleavageRules = "Chymotrypsin" or "Trypsin/P"

MaxMissedCleavages = 2

MaxDynamicMods = 2

MassReconMode = true

UnimodXML = "C:\Program Files (x86)\BumberShoot\bumberdash-bin-windows-x86-vc100-  
release-1\_4\_115\lib\Bumbershoot\TagRecon\unimod.xml"

Blosum = "C:\Program Files (x86)\BumberShoot\bumberdash-bin-windows-x86-vc100-release-  
1\_4\_115\lib\Bumbershoot\TagRecon\blosum62.fas"

DecoyPrefix = "rev\_"

UseChargeStateFromMS = true

NumChargeStates = 4

OutputFormat= "pepXML"

OutputSuffix = "\_specific\_MassRecon\_allGlycans"

TicCutoffPercentage = 0.98

MaxPeakCount = 200

MaxResultRank = 2

MinPeptideMass = 0 Da

MaxPeptideMass = 10000 Da

MinPeptideLength = 5

MaxPeptideLength = 75

UseSmartPlusThreeModel = false

ComputeXCorr = true

UseMultipleProcessors = true

FragmentationAutoRule = true

PrecursorMzTolerance = 0.015 m/z

FragmentMzTolerance = 0.02 m/z

NTerminusMzTolerance = 0.02 m/z

CTerminusMzTolerance = 0.02 m/z

UseAvgMassOfSequences = false

DuplicateSpectra = true

UseNETAdjustment = false

## APPENDIX C: PLANT N-GLYCAN BIOSYNTHESIS PATHWAY

The tool ‘pGlycFilter’ can be used to assess the biological feasibility of plant glycan compositions predicted by GlycoMod (Cooper et al., 2001; [web.expasy.org/glycomod](http://web.expasy.org/glycomod)). It is based on the *N*-glycan biosynthesis pathway in plants. The tool pGlycoFilter is available as supplementary data ([pGlycoFilter macros](#)).

Some proteins contain an *N*-terminal signal peptide that targets them to the secretory pathway (Stanley et al., 2009). This signals the nascent polypeptide to be introduced into the lumen of the endoplasmic reticulum (ER). If the sequence Asn-Xxx-Ser/Thr, where Xxx is not Pro, is encountered, it may be recognized by the oligosaccharyltransferase (OST) complex and have a bulky glycan attached (Glc<sub>3</sub>Man<sub>9</sub>GlcNAc<sub>2</sub>). The glycan is trimmed by  $\alpha$ -glucosidases and  $\alpha$ -mannosidases to yield an oligomannose-type glycan (Man<sub>9</sub>GlcNAc<sub>2</sub>). Glycoproteins bearing these oligomannose-type glycans are then secreted to the Golgi where the process of ‘maturation’ takes place. The spatial location of the following processes was reviewed by Ruiz-May et al. (2012).

During residence in the *Cis*-Golgi  $\alpha$ -mannosidase I (GMI) removes 1 – 4  $\alpha$ (1,2)-mannose (Man) residue(s) (Szmilo et al., 1986). Next, *N*-acetylglucosaminyltransferase I (GNTI) adds an *N*-acetylglucosamine (GlcNAc) residue to the  $\alpha$ (1,3)-mannose branch. The major substrate for GNTI is Man<sub>5</sub>GlcNAc<sub>2</sub>. Other substrates Man<sub>6</sub>GlcNAc<sub>2</sub>, Man<sub>7</sub>GlcNAc<sub>2</sub>, and Man<sub>9</sub>GlcNAc<sub>2</sub> exhibited no activity (Johnson and Chrispeels, 1987; Tezuka et al., 1992).

Upon secretion to the *Medial*-Golgi, Golgi  $\alpha$ -mannosidase II (GMII) sequentially removes the  $\alpha$ (1,6)-linked and  $\alpha$ (1,3)-linked mannose residues from the  $\alpha$ (1,6)-mannose branch (Strasser et al., 2006). GMII had no activity toward Man<sub>5</sub>GlcNAc<sub>2</sub> and GlcNAc on the  $\alpha$ (1,3)-branch was required for activity. Next, GNTII adds a GlcNAc residue to the  $\alpha$ (1,6)-mannose branch (Johnson and Chrispeels, 1987; Tezuka et al., 1992; Strasser et al., 1999). GNTII adds GlcNAc only after the activity of GMII. GlcNAc<sub>1</sub>Man<sub>3</sub>GlcNAc<sub>2</sub> is the major substrate of GNTII and no transfer of GlcNAc to Man<sub>5</sub>GlcNAc<sub>2</sub> and Man<sub>3</sub>GlcNAc<sub>2</sub> was observed. Following the activity of GNTI or GNTII, core fucosylation and xylosylation by  $\alpha$ (1,3)-fucosyltransferase

( $\alpha(1,3)$ -FucT) and  $\beta(1,2)$ -xylosyltransferase (XylT) could occur (Johnson and Chrispeels 1987; Tezuka et al., 1992). The major requirement of these enzymes is the presence of at least one terminal GlcNAc.

The optional elaboration of Lewis A structures occurs upon secretion to the *Trans*-Golgi. The first step in this process is the addition of galactose to one or two of the terminal GlcNAc(s) by  $\beta(1,3)$ -galactosyltransferase (GalT) (Strasser et al., 2007). Only after terminal galactosylation can addition of terminal fucose to the terminal GlcNAc(s) occur through the action of  $\alpha(1,4)$ -fucosyltransferase ( $\alpha(1,4)$ -FucT) (Leonard et al., 2002).

Further trimming by glycosidases occurs for *N*-glycoproteins secreted to the vacuole (Leonard et al., 2009). In particular, the vacuolar  $\beta$ -*N*-acetylglucosaminidase, HEXO1 is the main enzyme responsible for the paucimannose glycan structures that are prevalent in vacuolar *N*-glycoproteins (Strasser et al., 2007). HEXO1 is a broad-spectrum  $\beta$ -*N*-acetylglucosaminidase that does not appear to have a preference for either terminal GlcNAc residue. Endo- $\beta$ -mannosidase would act on the core  $\beta$ -linked mannose residue whereas a vacuolar  $\alpha$ -mannosidase would be responsible for removal of all other mannose residues which are  $\alpha$ -linked. The *N*-glycosylation biosynthetic pathway dictated by the action of these enzymes was the basis for pGlycoFilter ([pGlycoFilter macros](#)).

## APPENDIX D. DESCRIPTION OF SUPPLEMENTARY DATA

### **Proteomics software**

This folder contains the proteomics software used in [Chapter 3](#). ProteoWizard contains many tools including a spectrum viewer, ‘seeMS,’ and all of the tools needed for processing of the mass spectrometry data. IDPicker 3.0 is needed for viewing the IDPicker databases which contain the protein identification results. BumberShoot contains all of the proteomics search tools including MyriMatch. Each of these software is individually under an [Apache License, Version 2.0](#).

### **Mass spectrometry raw data**

This folder contains the raw mass spectrometry files described in [Chapter 3](#). The raw mass spectrometry files have been uploaded to the ProteoSAFE webserver ([massive.ucsd.edu](http://massive.ucsd.edu)) under MassIVE ID: MSV000079086. These data can be viewed using the ‘seeMS’ tool (ProteoWizard).

### **Protein databases**

This folder contains the protein databases used for the proteomics searches in [Chapter 3](#). They are described in more detail in [Appendix A](#).

### **Identification results**

This folder contains all of the protein and glycopeptide identification results presented in [Chapter 3](#). IDPicker databases are a convenient way to archive proteomics data which is accepted by proteomics journals such as *Molecular & Cellular Proteomics*. The files can be viewed with the IDPicker 3.0 software. The Excel spreadsheet (Supplementary\_Tables.xlsx) contains the list of all validated glycopeptide spectrum matches (S. Table 1) as well as the relative quantitation table showing all glycoforms (S. Table 2). The annotated glycopeptide spectrum matches are in a pdf file (Annotated\_gPSMs.pdf).

### ***pGlycoFilter macros***

This text file contains the macros, written for use in Microsoft Excel, for pGlycoFilter. This tool for glycan database generation is based on the plant *N*-glycan biosynthesis pathway, described in [Appendix C](#) and [Chapter 3](#). Instructions for its use are in [Appendix A](#).

### ***R scripts***

This folder contains all R scripts used to analyze data in [Chapter 3](#). The specific use of each one is described in [Appendix A](#). The script, ‘gPSMvalidator\_functions.R,’ contains all of the functions needed to run gPSMvalidator and to complete other tasks such as making a spectrum table and calculating peptide masses. One script was written to accept the results of the proteomics-based glycopeptide searches (gPSMvalidator\_proteomics.R) and another was written to accept the results of the glycoproteomics-based search (gPSMvalidator\_glycoproteomics.R). The script, ‘findUnoccupied.R,’ was written to find peptide spectrum matches for unoccupied glycosylation sites. Another script was written to compile a list sortable by glycosylation site, to count the number of spectra identified for each peptide backbone, as well as the number of glycoforms at each glycosylation site (glycosylationProfileTables.R). The script, ‘quantitationGlycoforms.R,’ was written to calculate the relative abundance of each glycan at each glycosylation site from the selected ion chromatograms.

**APPENDIX E. GUANIDINATION OF TRYPTIC PEPTIDES WITHOUT  
DESALTING FOR MATRIX-ASSISTED LASER  
DESORPTION/IONIZATION-TIME-OF-FLIGHT MASS SPECTROMETRY  
ANALYSIS<sup>2</sup>**

---

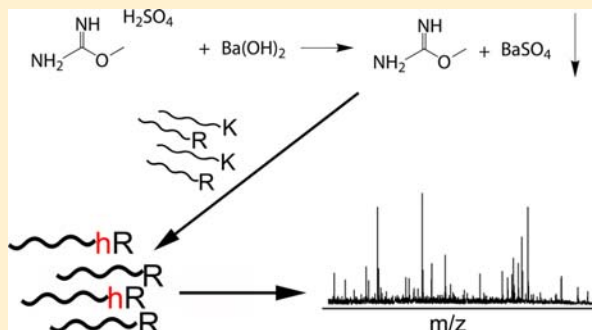
<sup>2</sup> Reproduced in part with permission from Baker, M. R. and Li, Q. X. Guanidination of Tryptic Peptides without Desalting for Matrix-Assisted Laser Desorption/Ionization-Time-of-Flight Mass Spectrometry Analysis. *Analytical Chemistry* **2013**, *85*, 8873–8880. Copyright © 2013, rights managed by the American Chemical Society.

# Guanidination of Tryptic Peptides without Desalting for Matrix-Assisted Laser Desorption/Ionization-Time-of-Flight Mass Spectrometry Analysis

Margaret R. Baker and Qing X. Li\*

Department of Molecular Biosciences and Bioengineering, University of Hawaii at Manoa, 1955 East-West Road, Honolulu, Hawaii 96822, United States

**ABSTRACT:** Derivatizations that enhance mass spectral quality often require desalting, which presents as a bottleneck in matrix-assisted laser desorption/ionization-time-of-flight mass spectrometry (MALDI-TOF MS)-proteomics. Guanidination, which converts lysine to homoarginine, an arginine analogue, can increase detection of those peptides 5–15-fold. Our aim was to improve guanidination by using a novel reagent, *O*-methylisourea-freebase. In a simple reaction, interfering salts were removed prior to guanidination. Freebase preparation took about 30 min and could be applied to samples all at once as opposed to desalting samples one-by-one for 5 min each. For freebase guanidinated BSA tryptic peptides, more than 6-times the peptides were observed relative to tryptic peptides or those guanidinated with the conventional reagent, *O*-methylisourea hemisulfate. Peptide signals increased more than 10-fold relative to those from guanidination with the conventional reagent and were equivalent to those from conventional guanidination with desalting. In addition, freebase guanidination allowed for a lower limit of detection when combined with another derivatization, *N*-terminal sulfonation, as evidenced by tandem mass spectrometry (MS/MS) fragmentation analysis of in-gel digests of cytochrome *c*. Freebase guanidination of rat lung proteins after 2-D gel electrophoresis allowed for identification of all tested protein spots regardless of protein characteristics (MW or pI) or abundance. Co-derivatization with *N*-terminal sulfonation confirmed the identity of low-abundance proteins in 2-D gel spots that contained more than one protein. The freebase guanidination reagent is simple to prepare and to implement. Desalting is not needed prior to MALDI-TOF MS. Freebase guanidination effectively increases the dynamic range of detection of lysine-containing peptides while decreasing the work needed for sample preparation.



Matrix-assisted laser desorption/ionization-time-of-flight mass spectrometry (MALDI-TOF MS) is an important analytical tool for proteomics due to its high throughput capability and robustness. It is the method of choice for identifying proteins separated by 2-D gel electrophoresis (2-DGE), which in combination with MALDI-TOF MS is valuable for examining roles of proteins involved in biological processes<sup>1–3</sup> and for discovery of biomarkers.<sup>4–6</sup>

2-DGE separates proteins into single protein spots that can be excised from the gel and digested with a protease, typically trypsin. Trypsin cleaves proteins at the C-terminus of arginine or lysine into peptides suitably sized for MS analysis (i.e., 700–4000 Da). The basic functionality of the C-terminus of tryptic peptides stabilizes the positive charge necessary for MS.

In MALDI-TOF MS, a distinct bias is observed toward arginine-terminated peptides on the order of 5–15-fold intensity when compared to their lysine-terminated counterparts.<sup>7</sup> The guanidine side group of arginine ( $\text{p}K_a$  12.5) has a more favorable acid–base reaction with the matrix and ionizes more readily in the MALDI process than the amine side group of lysine ( $\text{p}K_a$  10.5).<sup>7,8</sup> This so-called “arginine effect” is also

observed during time-of-flight (i.e., mass analysis), where peptides terminating in arginine are less prone to metastable fragmentation.<sup>9</sup>

Chemical derivatization is commonly employed to improve the quality of MALDI-TOF spectra.<sup>10</sup> Guanidination is a derivatization that fully takes advantage of the “arginine effect” to increase the dynamic range of detection by converting lysine to homoarginine, an arginine analogue. At an alkaline pH ( $\text{pH} > 10.5$ ), *O*-methylisourea reacts with the  $\epsilon$ -amino group of lysine to form homoarginine in a selective manner.<sup>11–13</sup>

Shifting the functionality of lysine to mirror that of arginine is beneficial for a number of reasons, the most apparent being an increase in detection of lysine-terminated peptides. Such an increase could improve proteome coverage in peptide mass fingerprinting (PMF)<sup>12</sup> studies and could allow for in-depth analysis of single proteins. For example, guanidination allowed

**Received:** July 20, 2013

**Accepted:** August 21, 2013

**Published:** August 21, 2013

for pin-pointing of oxidation sites thought to be unique to Parkinson's disease.<sup>5</sup>

Guanidination can also prevent unwanted side-reactions to lysine when the N-terminus is to be derivatized. N-Terminal sulfonation, which promotes uniform fragmentation in MALDI-TOF MS/MS, is arguably the most common derivatization to combine with guanidination.<sup>14</sup> Such a combination was used to increase coverage of 2-D gel excised proteins<sup>1</sup> and in *de novo* sequencing of novel proteins.<sup>15–19</sup> In quantitative proteomics, guanidination was used directly as an isotopic tag<sup>20</sup> or as a protecting group prior to labeling the N-terminus with a tag.<sup>21,22</sup> Another advantage of guanidination is the mass shift of 42 Da which helps distinguish lysine from glutamine ( $\Delta$  0.04 Da).<sup>11</sup>

Salts from guanidination interfere with MALDI-TOF MS, and desalting is required.<sup>23</sup> Desalting is time-consuming, labor intensive, and costly. Consequently, guanidination is often performed as a second pass on protein spots that did not result in a statistically significant identification.<sup>1,24</sup> Relative quantitation using isotope labeled guanidination reagents could be automated except for the desalting step.<sup>20</sup>

Alternative guanidination methods do not require desalting; however, there are disadvantages. Desalting is optional for a commercial guanidination kit that uses a low concentration of *O*-methylisourea-salt.<sup>25</sup> However, potential problems include incomplete derivatization, which complicates spectra and may interfere with quantitation. Although in-gel guanidination of proteins prior to digestion was assumed to be complete, it is incompatible with in-solution digests.<sup>26,27</sup> Sergeant et al.<sup>26</sup> reported that trypsin has lower digest efficiency toward homoarginine, resulting in longer peptides which tend to be more difficult to analyze.

The objective of this study was to develop freebase guanidination, which is a simple and robust alternative to the conventional guanidination procedure. Use of *O*-methylisourea-freebase increases sensitivity and high-throughput capability of the MALDI-proteomics workflow. In a simple reaction, *O*-methylisourea-salt is reacted with barium hydroxide, the products of which separate easily: *O*-methylisourea-freebase goes to the aqueous phase and unwanted salts precipitate.<sup>28</sup> The freebase guanidination reagent gives complete guanidination and does not leave behind interfering salts. This is especially useful for applications, as demonstrated, for which additional purification steps are not desirable.

## ■ EXPERIMENTAL SECTION

**Materials and Reagents.** Trypsin Gold, MS grade, was purchased from Promega (Madison, WI). Standard proteins: BSA, ovalbumin (chicken), myoglobin (horse), and cytochrome *c* (horse) (Sigma, St. Louis, MO) were used. The MALDI matrix  $\alpha$ -cyano-4-hydroxycinnamic acid (HCCA) and Peptide Calibration Standard II were from Bruker (Billerica, MA). C<sub>18</sub> NuTips were from PolyLC (Columbia, MD). Other reagents included *O*-methylisourea-hemisulfate (Sigma), barium hydroxide octahydrate (J.T. Baker Chemicals), ammonium hydroxide (NH<sub>4</sub>OH) (Mallinckrodt, Inc.), and 2-sulfobenzoic acid cyclic anhydride (Acros Organics, New Jersey). Acetonitrile (LC-MS grade) and ammonium bicarbonate (NH<sub>4</sub>HCO<sub>3</sub>) were from Fisher Scientific. Water was purified on a Milli-Q Advantage A10 system (Millipore, Billerica, MA). Laemmli sample buffer, Bio-Safe Coomassie Stain, and other gel electrophoresis supplies were purchased from Bio-Rad (Hercules, CA). Rat lung tissues from 8 week-old

Sprague–Dawley Crj:CD 1GS male rats were provided by the Korean Rural Development Administration (Gyeonggi-do, Korea).

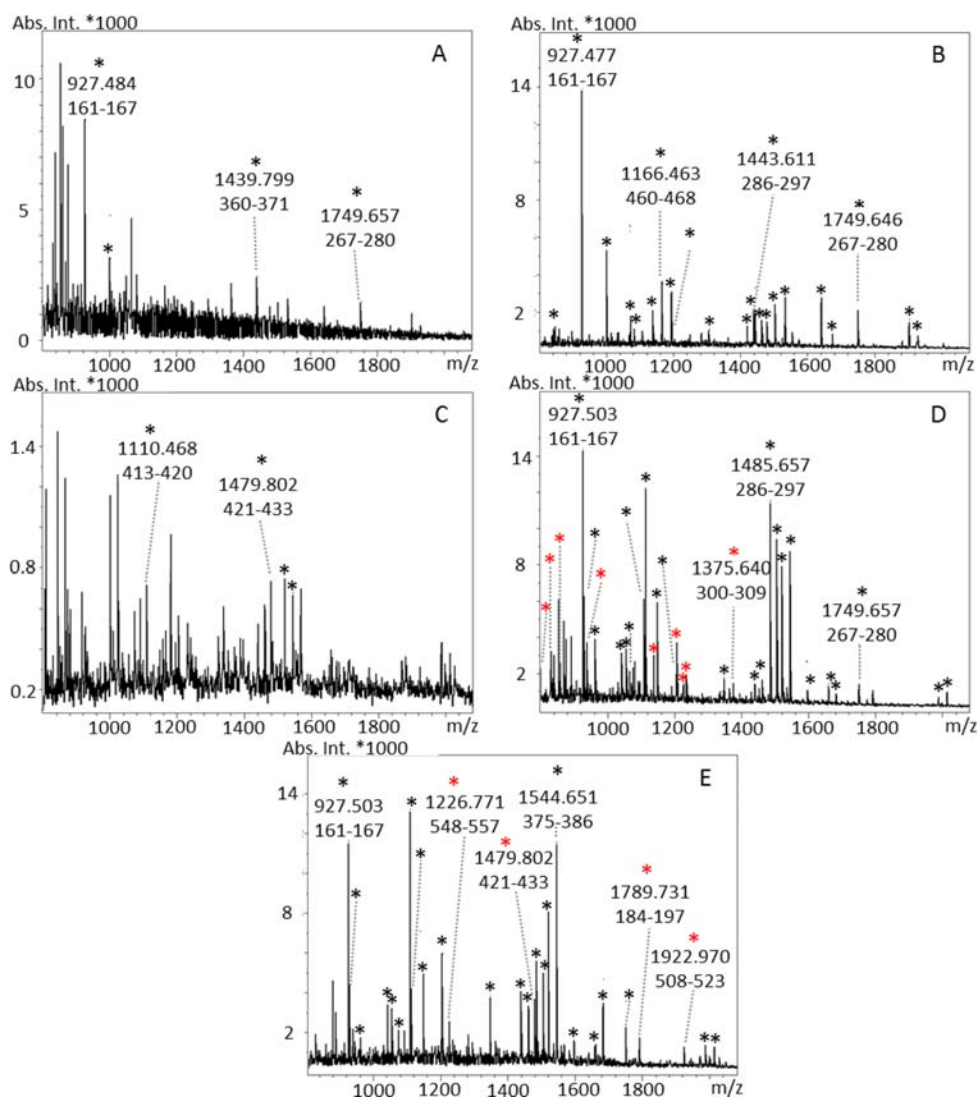
**One-Dimensional Gel Electrophoresis (1-DGE).** Mixtures of standard proteins were separated via 1-DGE in triplicate.<sup>29</sup> The mixtures consisted of 12.5, 25, or 50 pmol each of BSA, ovalbumin, myoglobin, and cytochrome *c*. Proteins were separated on a 12% polyacrylamide gel with a MiniProtein apparatus (Bio-Rad) at a constant 15 mA for 2 h. The gel was stained with Bio-Safe Coomassie Stain according to the Bio-Rad protocol.

**2-DGE.** Proteins were extracted from homogenized rat lung tissues in rehydration buffer (Bio-Rad), and the concentrations were determined with a Bradford assay.<sup>30</sup> Lung proteins were separated via 2-DGE according to the Bio-Rad protocol.<sup>31</sup> Whole protein extract (200  $\mu$ g) was applied to an immobilized pH gradient strip (7 cm, pH 3–10, linear) using active rehydration at 50 V for 16 h at 20 °C on a Protean i12 isoelectric focusing (IEF) cell (Bio-Rad). IEF consisted of rapid ramping to 250 V (15 min), gradual ramping to 4 000 V (1 h), and then rapid ramping at 4 000 V for 15 000 V h. The second dimension was run on a precast gel (Any kD Mini-Protein TGX, Bio-Rad) at a constant 200 V for 0.5 h. The gel was stained with Bio-Safe Coomassie Stain as described for 1-DGE. Spots were manually picked with a pipet tip, modified by cutting off the end of the tip.

**In-Gel Trypsin Digest.** In-gel trypsin digestion was conducted according to a protocol from Cold Spring Harbor<sup>32</sup> with only slight modification to peptide extraction. Following reduction and alkylation, proteins were digested with trypsin [12.5 (2-D gel spots) or 25 ng of trypsin (1-D gel bands) in 50 mM NH<sub>4</sub>HCO<sub>3</sub>] at 37 °C for 16 h. Peptides were extracted by sonication for 5 min at ambient temperature in 25 mM NH<sub>4</sub>HCO<sub>3</sub>, followed by addition of acetonitrile and repeated sonication. After transferring the supernatant into fresh vials, peptides were extracted in 0.05% trifluoroacetic acid (TFA) and then 0.05% TFA, 50% acetonitrile. 1-D gel extracts were divided into three aliquots for comparison of three guanidination methods (freebase, conventional, and conventional with desalting) and then were dried via SpeedVac.

**In-Solution Trypsin Digest.** In-solution trypsin digestion of BSA (200 pmol) was conducted in triplicate according to Russell et al.,<sup>33</sup> which was completed in 1 h by incorporating 80% acetonitrile into the digest solution. Reduction and alkylation were also in 80% acetonitrile solution. Peptides were dried via SpeedVac and resuspended in 100  $\mu$ L of 5% acetonitrile, 0.3% formic acid and stored at –20 °C.

**Guanidination Reagent Preparation.** The freebase guanidination reagent was prepared with significant modification to Imbeah et al.<sup>28</sup> in order to make the final product suitable for downstream MS application. To prepare *O*-methylisourea-freebase, 50 mg of *O*-methylisourea hemisulfate was dissolved in 51  $\mu$ L of water and then added to 64 mg of barium hydroxide octahydrate to yield an 8 M *O*-methylisourea and 4 M barium hydroxide solution. It is important to note that barium hydroxide should be in a 1:1 (mol/mol) ratio with sulfate. The freebase reaction was mixed by pulse vortexing (10 min) and then centrifuged (20 800g) at 4 °C for 5 min. The clear supernatant (40  $\mu$ L) was collected, avoiding the pellet. Next, 112  $\mu$ L of acetonitrile was added to bring it to ~74% acetonitrile. The solution was vortexed for 15 s and a precipitate formed. The precipitate was pelleted as in the previous step. The mixture separated into three phases, one



**Figure 1.** MALDI-TOF spectra of BSA (25 fmol) tryptic peptides before (A) and after desalting (B), after conventional guanidination without desalting (C) and with desalting (D), and after freebase guanidination (E). For reference, the  $m/z$  and peptide range (BSA; UniProt accession P02769) are given for some peaks. Asterisks indicate peaks contributing to protein identification, and red asterisks highlight the peaks that differ between parts D and E.

solid precipitate and two liquid phases. An aliquot of 100  $\mu\text{L}$  (2.11 M *O*-methylisourea-freebase in 74% acetonitrile, pH 9–10) was carefully collected from the top liquid layer and mixed with 25  $\mu\text{L}$  of 14.8 N  $\text{NH}_4\text{OH}$ , 35  $\mu\text{L}$  of acetonitrile, and 51  $\mu\text{L}$  of water to prepare a 1.0 M *O*-methylisourea-freebase solution in 50% acetonitrile containing 1.75 N  $\text{NH}_4\text{OH}$  (pH 11) as the working solution. The conventional guanidination reagent was prepared with slight modification to Beardsley and Reilly.<sup>23</sup> *O*-Methylisourea hemisulfate (50 mg) was dissolved in 119.3  $\mu\text{L}$  of water, 163.2  $\mu\text{L}$  of acetonitrile, and 96.5  $\mu\text{L}$  of 14.8 N  $\text{NH}_4\text{OH}$ , making a solution of 1.0 M *O*-methylisourea, 40% acetonitrile, 3.5 M  $\text{NH}_4\text{OH}$  (pH 11).

**Guanidination.** Dried peptides were dissolved in 5  $\mu\text{L}$  of *O*-methylisourea-freebase or *O*-methylisourea hemisulfate solution by sonication for 5 min. After guanidination at 65  $^\circ\text{C}$  for 20 min, the solution was dried via SpeedVac for approximately 5 min.

**Sulfonation.** Sulfonation was conducted with modification to Keough et al.<sup>14</sup> Sulfonation solutions were prepared daily by

dissolving 20 mg of 2-sulfobenzoic acid cyclic anhydride in 1 mL of dry tetrahydrofuran. Dried guanidinated peptides were dissolved in 2.5  $\mu\text{L}$  of 25 mM triethylammonium bicarbonate, 50% acetonitrile. Next, 2.5  $\mu\text{L}$  of the sulfonation solution was added to bring it to pH 8–9, followed by sonication for 5 min at ambient temperature. After the reaction, the solution was evaporated via SpeedVac.

**Sample Preparation for Mass Spectrometry.** Freebase guanidinated peptides (with or without sulfonation) were dissolved in 5  $\mu\text{L}$  of 0.25% TFA, 50% acetonitrile (pH 2). Conventionally guanidinated peptides (with or without sulfonation) were optionally desalted with  $\text{C}_{18}$  NuTips as per the manufacturer's instructions. After desalting, the sample was eluted with 5  $\mu\text{L}$  of 0.05% TFA, 60% acetonitrile into a clean tube.

Samples were mixed 1:1 with the matrix (0.7 mg/mL HCCA in 85% acetonitrile, 0.1% TFA, 1 mM ammonium phosphate, monobasic), and 1  $\mu\text{L}$  was deposited on the AnchorChip target plate (Bruker) and allowed to air-dry. Alternatively, 0.5  $\mu\text{L}$  of

the sample was deposited on a standard steel target followed by 0.5  $\mu\text{L}$  of the matrix (20 mg/mL HCCA in 70:30 [v/v] acetonitrile/5% formic acid) and allowed to air-dry. Peptide II calibration standard, in the same matrix, was spotted adjacent to the samples.

**Mass Spectrometry.** MALDI-TOF MS/(MS) was conducted with an Ultra Flex III (Bruker) using Flex Control v. 3.4. All spectra were acquired in positive reflectron mode and were externally calibrated. For MS mode, 1000 laser shots were summed. For automated acquisition, 10 attempts were made before aborting acquisition. Manual data acquisition was then attempted for aborted acquisitions. Parent ions were fragmented in LIFT mode.<sup>34</sup> Data were sent to Flex Analysis v. 3.4 for spectral processing. Peak picking and residue assignments were manually verified.

**Protein Identification.** Spectral data were subjected to a Mascot search<sup>35</sup> on a local Mascot server v. 2.2.7 (Matrix Science, Boston, MA) via ProteinScape v. 3 (Bruker). For PMF, the mass tolerance was 50 ppm and carbamidomethyl cysteine was a global modification. Upon verification that no underivatized lysines were detected, homoarginine was set as a global modification. Fragmentation spectra were submitted to a Mascot MS/MS search with a mass tolerance of 50 ppm and MS/MS tolerance of 0.3 Da. Carbamidomethyl cysteine and homoarginine were global modifications, and N-terminal sulfonation, listed as 3Sulfo, was a variable modification. Identifications were based on significant matches at  $p < 0.05$ .

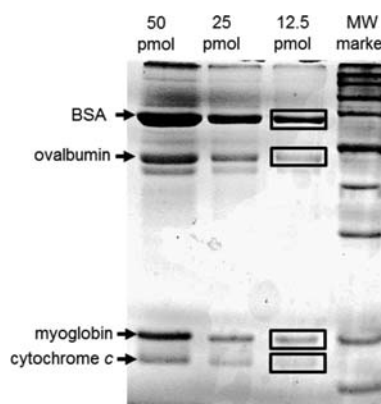
## RESULTS AND DISCUSSION

**O-Methylisourea-Freebase Preparation and Use.** The improved guanidination reagent, *O*-methylisourea-freebase, was prepared in less than 30 min while waiting for gel peptide extracts to dry. In a simple reaction between *O*-methylisourea hemisulfate salt and barium hydroxide, *O*-methylisourea-freebase separated into the aqueous phase while salts precipitated. The salts are aqueous soluble to an extent (barium sulfate [2.5  $\mu\text{g}/\text{mL}$ ], unreacted *O*-methylisourea hemisulfate [100 mg/mL], unreacted barium hydroxide octahydrate [38.9 mg/mL]) but are not soluble in organic solvents. Upon agitation in 74% acetonitrile, residual salts precipitated. The final freebase guanidination solution includes 1.0 M *O*-methylisourea-freebase, which was efficient for complete guanidination of tryptic peptides,<sup>23</sup> 50% acetonitrile, which aided in resolubilizing dried peptides, and 1.75 N  $\text{NH}_4\text{OH}$ , which maintained the reaction pH (pH 11). Drying the peptides after the 20 min reaction effectively removed  $\text{NH}_4\text{OH}$ . Dried freebase guanidinated peptides were dissolved in 0.25% TFA, 50% acetonitrile prior to cocrystallization with HCCA. Addition of 5  $\mu\text{L}$  of that solution (which is equivalent to 1 reaction volume) had little or no effect on cocrystallization.

**Freebase versus Conventional Guanidination of In-Solution BSA Digests.** Tryptic peptides from in-solution digestion of BSA were used to optimize freebase guanidination conditions and to examine the benefits of freebase guanidination compared to conventional guanidination, which uses *O*-methylisourea hemisulfate, with and without desalting. In the present study, 25 fmol of BSA was the lowest amount for which a positive identification could be obtained for the tryptic peptides via a database search (Figure 1A). The four peptides used for identification resulted in 7% sequence coverage. Desalting and sample concentration with  $\text{C}_{18}$  NuTips improved sequence coverage by more than 5-fold (39% sequence coverage; Figure 1B). Conventional guanidination without

desalting resulted in a marked reduction in ion intensity and a noisy background which prevented detection of all but 4 peptide ions (7% sequence coverage; Figure 1C). Desalting of conventionally guanidinated peptides removed interfering salts, increasing signal and lowering background noise (50% sequence coverage; Figure 1D). With freebase guanidination, interfering salts are removed prior to guanidination. After freebase guanidination, there was minimal background noise and the signal intensity was equivalent to that of desalted samples (Figure 1B,D, and E). Both freebase guanidination and the conventional procedure are alleviated from the "arginine effect," whereas for tryptic peptides, one peak dominated the spectrum (Figure 1B;  $m/z$  927; YLYEIAR). Sequence coverage after freebase guanidination was comparable to the conventional method with desalting (46% sequence coverage). Differences in peptide ions contributing to sequence coverage were noted (red asterisks; Figure 1D,E). Peptide ions exclusive to conventional guanidination with desalting were found in the lower end of the mass range, whereas those exclusive to freebase guanidination were larger. This is in line with the observations that desalting with a  $\text{C}_{18}$  microcolumn leads to the loss of more hydrophobic peptides.<sup>23,36</sup> No underivatized lysines were detected after freebase guanidination, indicating a quantitative reaction.

**Freebase versus Conventional Guanidination of Standard Proteins Digested In-Gel.** Guanidination is pH sensitive. The reaction relies on the  $\epsilon$ -amino group of lysine being deprotonated. The optimal pH is 11.3<sup>13</sup> with a minimum of pH 10.<sup>23</sup> The freebase guanidination solution was adapted to ensure that it is compatible with standard in-gel digest procedures, which use acidic conditions to extract peptides. A mixture of varying amounts of the four standard proteins was separated by 1-DGE (Figure 2). At 12.5 pmol of each protein,



**Figure 2.** 1-DGE of mixtures of standard proteins (50, 25, or 12.5 pmol). Boxes indicate bands excised for further experiments.

cytochrome *c* was barely detectable, so this amount was used in the remainder of the study. The higher acid content of gel extracts required 10-times more  $\text{NH}_4\text{OH}$  than in-solution digested peptides to achieve the optimal reaction pH. When a lesser amount of base was used, guanidination was incomplete.

High-throughput capability with MALDI-TOF MS relies on efficient automated data acquisition. Samples were spotted on an AnchorChip target, and data were collected in automated mode. Sequence coverage and the Mascot score were about the same for freebase guanidination and conventional guanidination with desalting (Table 1). The sequence coverage for these

**Table 1. Comparison of Freebase Guanidination to Conventional Guanidination with and without Desalting<sup>a</sup>**

| protein <sup>b</sup>           | MW (kDa) | pI  | freebase           |                       | conventional without desalting |                       | conventional with desalting |                       |
|--------------------------------|----------|-----|--------------------|-----------------------|--------------------------------|-----------------------|-----------------------------|-----------------------|
|                                |          |     | score <sup>c</sup> | sequence coverage (%) | score                          | sequence coverage (%) | score                       | sequence coverage (%) |
| BSA                            | 66       | 5.6 | 227                | 45                    | 226                            | 35                    | 267                         | 45                    |
| ovalbumin                      | 42       | 5.2 | 158                | 57                    | 131                            | 41                    | 133                         | 55                    |
| myoglobin                      | 17       | 7.4 | 173                | 75                    | 159                            | 65                    | 170                         | 74                    |
| cytochrome <i>c</i>            | 12       | 9.6 | 160                | 56                    | 56 <sup>a</sup>                | 19 <sup>a</sup>       | 145                         | 53                    |
| repeatability <sup>d</sup> (%) |          |     |                    | 100                   |                                | 48                    |                             | 100                   |

<sup>a</sup>Results are for samples from 1-DGE in Figure 2 and are representative of all replicates except cytochrome *c* from conventional guanidination without desalting, which had only 1 replicate with interpretable results. <sup>b</sup>75 fmol per spot. <sup>c</sup>Mascot PMF score. <sup>d</sup>The % identified out of 48 (4 proteins × 3 replicates × 4 technical replicates).

**Table 2. Comparison of Freebase Guanidination to Conventional Guanidination with and without Desalting When Combined with N-Terminal Sulfonation<sup>a</sup>**

|                                | BSA                |          | ovalbumin |          | myoglobin |          | cytochrome <i>c</i> |          |
|--------------------------------|--------------------|----------|-----------|----------|-----------|----------|---------------------|----------|
|                                | score <sup>b</sup> | peptides | score     | peptides | score     | peptides | score               | peptides |
| freebase                       | 406                | 14       | 342       | 7        | 180       | 6        | 158                 | 3        |
| conventional without desalting | 190                | 6        | 178       | 6        |           |          |                     |          |
| conventional with desalting    | 393                | 16       | 437       | 9        | 137       | 7        |                     |          |

<sup>a</sup>Results for samples from 1-DGE (Figure 2). <sup>b</sup>Summed ions scores from Mascot MS/MS search.

was about 10 percentage points higher than conventional guanidination without desalting for all proteins except cytochrome *c*, which was more than 30 percentage points lower. Only one replicate from conventional guanidination without desalting was interpretable for cytochrome *c*, which was near the limit of detection (LOD) for Coomassie staining (Figure 2). These data indicate major advantages of freebase guanidination over conventional guanidination with or without desalting at low protein concentration conditions for cytochrome *c*, potentially for other proteins as well. Freebase guanidination is therefore effective in increasing the dynamic range of detection while decreasing the work needed in sample preparation.

Repeatability measures how reliable a method is. A few samples from freebase guanidination and conventional guanidination with desalting failed during automated acquisition, whereas most from conventional guanidination without desalting failed. The missing data were collected manually. For the four standard proteins and each of the replicates, a significant identification could be obtained 100% of the time when the freebase method or the conventional method with desalting was used. Less than half of those examined after conventional guanidination without desalting were successfully identified (Table 1; repeatability). AnchorChip targets are the most convenient for automated data acquisition.<sup>37</sup> These targets are also known for increasing sensitivity by concentrating the sample into a smaller area, which comes with the consequence of other components concentrating as well.<sup>37</sup> As a result, an enhancement was not observed for freebase guanidinated samples and the data quality for conventional guanidination without desalting was critically reduced. Also, sample drying time was increased with AnchorChip use. The best compatibility was with the standard target plate, which can be used for automated data acquisition.

**Freebase versus Conventional Guanidination when Combined with N-Terminal Sulfonation.** In MALDI-TOF MS/MS, fragmentation occurs through a charge-remote mechanism, so spectra are not as readily interpretable as those from electrospray (ESI)-MS/MS.<sup>38</sup> N-Terminal sulfonation facilitates unambiguous interpretation of MALDI-TOF

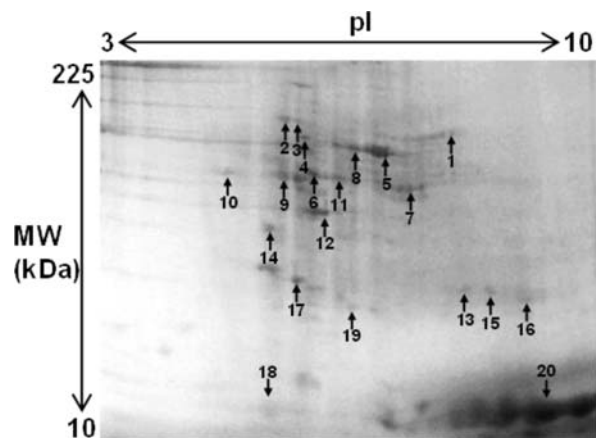
MS/MS fragmentation spectra wherein  $\gamma$ -ions are almost exclusively detected.<sup>14</sup> A major advantage of guanidination is that its coderivatization with N-terminal sulfonation gives no side reactions.<sup>13</sup>

N-Terminal sulfonation with 2-sulfobenzoic acid could be conducted without sample cleanup.<sup>26</sup> Therefore, the three guanidination methods were compared after N-terminal sulfonation of the standard protein digests with this chemical (Table 2). The 20 most intense peaks were subjected to MS/MS fragmentation, and scores of significant peptide identifications were summed. For cytochrome *c*, which stained very faintly after 1-DGE (Figure 2), a positive identification from the fragmentation spectra could only be obtained with freebase guanidinated samples. Compared to conventional guanidination without desalting, freebase guanidination had more peptides identified and higher scores from the Mascot MS/MS search. For ovalbumin, conventional guanidination with desalting is the preferred method since two additional peptides were identified by MS/MS fragmentation analysis leading to a higher score. Freebase guanidination well identified four out of four proteins, whereas the conventional method with desalting identified three out of four. This and previous studies showed that desalting is needed for the conventional guanidination method,<sup>23</sup> and sample losses are an unavoidable concern when desalting with C<sub>18</sub> microcolumns.<sup>23,36</sup> Freebase guanidination did not require desalting, which improved sensitivity of detection (Table 2). This was particularly important since N-terminal sulfonation decreases sensitivity.<sup>26</sup> A more thorough investigation demonstrating applicability to a wider range of proteins is given in the following section.

**Application of Freebase Guanidination and its Coderivatization with N-Terminal Sulfonation to Analysis of Rat Lung Proteins from 2-DGE.** Lung diseases continue to be challenging to diagnose and treat.<sup>39,40</sup> Insights into the biology of lung cancer<sup>4,6</sup> as well as an understanding of pulmonary development<sup>2</sup> have been gained with the use of proteomics. A major objective of the present study was to implement freebase guanidination in the analysis of lung proteins separated by 2-DGE. Successful implementation was assessed by the capacity to yield statistically significant protein

identifications regardless of protein characteristics or abundance.

Whole protein extracts from rat lung were separated via 2-DGE. A total of 20 spots, which captured a wide range of molecular weight and isoelectric points with detection by Coomassie staining, were chosen from the gel (Figure 3).



**Figure 3.** 2-DGE of rat lung tissues. Numbers indicate protein spots excised from the gel and used for protein identification (Table 3).

Identification was achieved through PMF of freebase guanidinated samples (Table 3). Further confirmation was achieved by coderivatization with N-terminal sulfonation and MS/MS fragmentation of at least one peptide. For each of the 20 proteins, a significant PMF identification was obtained regardless of protein characteristics (MW or pI) or abundance. Fragmentation analysis after coderivatization with N-terminal

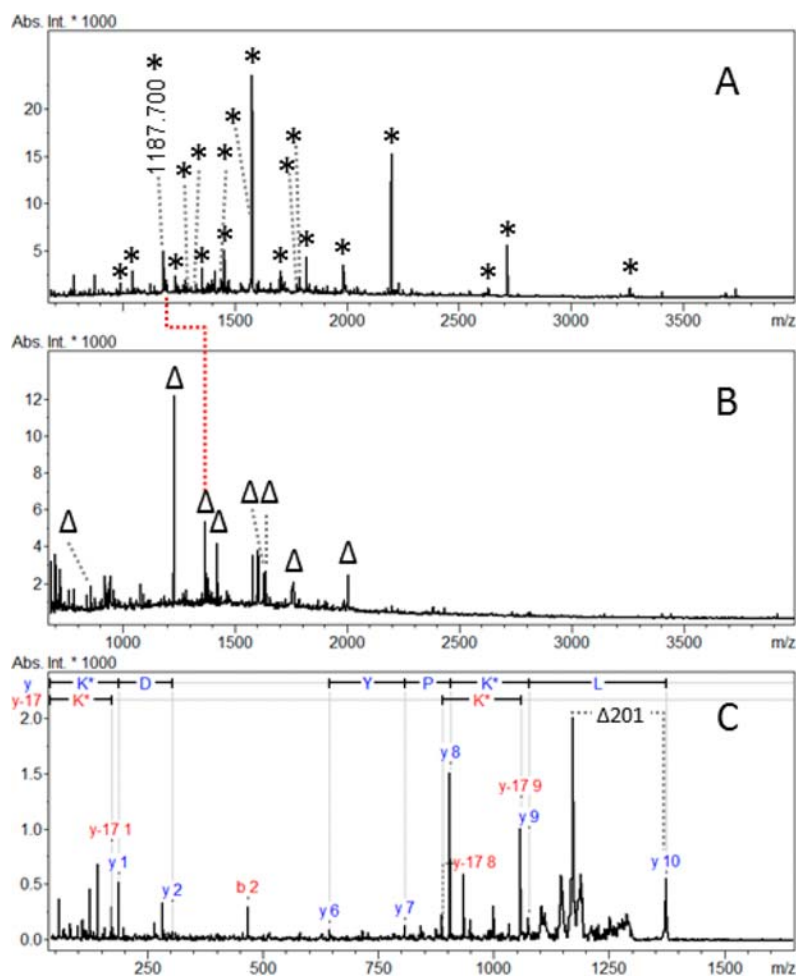
sulfonation allowed for successful identification of 18 out of 20 spots. The two that were not identified (spots 10 and 19) exhibited very low staining. The LOD by Coomassie staining coincides with that of MALDI-TOF MS/MS (estimated according to refs 31 and 34). For a 97.2 kDa protein, phosphorylase b, 50–100 fmol was required for MS/MS analysis<sup>34</sup> and the lower LOD for Coomassie staining was 8–28 ng of protein,<sup>31</sup> which is equivalent to 46–160 fmol of phosphorylase b.

PMF results warranted MS/MS fragmentation analysis for two main reasons: (1) the protein was not represented in the Swiss-Prot database,<sup>41</sup> which is manually annotated and reviewed, or (2) the spot contained a mixture of proteins. The mass spectrum of spot 2 after trypsin digestion and freebase guanidination is shown (Figure 4A). An initial search in the Swiss-Prot database<sup>41</sup> returned no significant hits. A Mascot search using a larger database, NCBI nr, resulted in a match to the predicted protein collagen  $\alpha$ -1 (VI) for which there is currently no protein level data available (UniProt Knowledgebase<sup>42</sup> entry last updated May 29, 2013 for accession number D3ZUL3\_RAT). A total of 20 peaks, within the mass range of  $m/z$  995–3257, were used in the identification (Figure 4A). Of these, five were found to contain hydroxyproline ( $m/z$  995, 1356, 1572, 1708, and 1724). Sulfonation with sulfo benzoic acid cyclic anhydride results in a mass shift of 184 Da ( $O_3S-C_6H_4-CO$ ). After N-terminal sulfonation, eight sulfonic acid derivatives were detected (Figure 4B). A red dashed line indicates a shift of 184 Da for a peptide ( $m/z$  1187, Figure 4A) to its sulfonic acid derivative ( $m/z$  1371, Figure 4B), which was selected for MS/MS fragmentation (Figure 4C). In the fragmentation spectrum, the highest peak was from a loss of the aromatic sulfonic acid group and ammonia ( $\Delta$  201,  $HO_3S-C_6H_4-CO-NH_2$ , Figure 4C). This and the observed

**Table 3.** List of Proteins from 2-DGE of Rat Lung Tissues Identified with Freebase Guanidination and Confirmed by Coderivatization with N-Terminal Sulfonation<sup>a</sup>

| spot | accession <sup>b</sup> | protein                                        | guanidination only         |                       | guanidination + sulfonation |          |
|------|------------------------|------------------------------------------------|----------------------------|-----------------------|-----------------------------|----------|
|      |                        |                                                | protein score <sup>c</sup> | sequence coverage (%) | ions score <sup>d</sup>     | peptides |
| 1    | P12346                 | serotransferrin                                | 233                        | 31                    | 49                          | 1        |
| 2    | D3ZUL3                 | predicted protein collagen $\alpha$ -1 (IV)    | 149                        | 24                    | 70                          | 2        |
| 3    | P06761                 | 78 kDa glucose-regulated protein <sup>e</sup>  | 248                        | 41                    | 47                          | 2        |
| 4    | P63018                 | heat shock cognate 71 kDa protein <sup>e</sup> | 125                        | 23                    | 47                          | 1        |
| 5    | P02770                 | serum albumin                                  | 356                        | 62                    | 238                         | 8        |
| 6    | P31000                 | vimentin                                       | 265                        | 67                    | 72                          | 2        |
| 7    | Q8VIF7                 | selenium-binding protein 1 <sup>e</sup>        | 117                        | 29                    | 32                          | 2        |
| 8    | P20059                 | hemopexin <sup>e</sup>                         | 121                        | 31                    | 54                          | 1        |
| 9    | P69897                 | tubulin $\beta$ -5 chain                       | 114                        | 44                    | 39                          | 1        |
| 9    | P17475                 | $\alpha$ -1-antitrypsinase                     | 146                        | 42                    | 35                          | 1        |
| 10   | P18418                 | calreticulin                                   | 149                        | 26                    |                             |          |
| 11   | P85125                 | polymerase I and transcript release factor     | 58                         | 12                    | 61                          | 2        |
| 12   | P60711                 | actin, cytoplasmic 1                           | 171                        | 59                    | 75                          | 1        |
| 13   | P25113                 | phosphoglycerate mutase 1 <sup>f</sup>         | 99                         | 51                    | 19                          | 1        |
| 14   | P58775                 | tropomyosin $\beta$ chain                      | 91                         | 37                    | 53                          | 1        |
| 15   | BOBNN3                 | carbonic anhydrase 1 <sup>g</sup>              | 100                        | 56                    | 60                          | 2        |
| 16   | P08010                 | glutathione S-transferase Mu 2 <sup>f,g</sup>  | 126                        | 51                    | 23                          | 1        |
| 17   | O35244                 | peroxiredoxin-6                                | 244                        | 61                    | 23                          | 1        |
| 18   | P01946                 | hemoglobin subunit $\alpha$ -1/2               | 62                         | 37                    | 47                          | 1        |
| 19   | P42930                 | heat shock protein $\beta$ -1 <sup>g</sup>     | 79                         | 32                    |                             |          |
| 20   | P02091                 | hemoglobin subunit $\beta$ -1 <sup>g</sup>     | 209                        | 95                    | 92                          | 2        |

<sup>a</sup>2-DGE spot (Figure 3). <sup>b</sup>UniProt Knowledgebase accession number. <sup>c</sup>Mascot PMF score. <sup>d</sup>Summed ions scores from MS/MS Mascot search. <sup>e</sup>Background proteins are serum albumin. <sup>f</sup>Background proteins are hemoglobin subunit  $\beta$ -1. <sup>g</sup>Background proteins are hemoglobin subunit  $\alpha$ -1/2.



**Figure 4.** Identification of collagen  $\alpha$ -1 (VI) from 2-DGE spot 2 with MALDI-TOF spectra of tryptic peptides after freebase guanidination (A) and coderivatization with *N*-terminal sulfonation (B). A red dashed line indicates the sulfonic acid derivative of  $m/z$  1187 which was used for MS/MS fragmentation analysis (C). Interpretation of the  $y$ -,  $y$ -17-, and  $b$ -ions led to identification of a collagen  $\alpha$ -1 (VI) peptide (LKPYGALVDK). Asterisks indicate peaks contributing to protein identification, and triangles indicate *N*-terminally sulfonated tryptic peptides.

$y$ -<sub>17</sub>-ions are expected for sulfonic acid derivatives of peptides containing an internal arginine or homoarginine.<sup>14,26</sup>

Differential post-translational modification and proteolytic cleavage of a highly abundant protein can result in occurrence of that protein in more than one spot.<sup>39</sup> This phenomenon may obscure detection of low abundance and potentially relevant proteins. Several such high abundance proteins were detected in the present study, including serum albumin and hemoglobin. Fragmentation spectra were instrumental in verifying low-abundance components in various spots (3, 4, 7, 8, 13, 15, 16, 19, and 20; Figure 3 and Table 3). For example, PMF of spot 8 revealed that serum albumin peaks accounted for 75% of the ion intensity, whereas those of hemopexin accounted for only 11%. Analysis of MS/MS fragmentation spectra confirmed the presence of hemopexin in the mixture.

## CONCLUSIONS

Problems associated with guanidination using *O*-methylisourea hemisulfate appeared to stem exclusively from the salts, making use of *O*-methylisourea-freebase ideal. In the present study, no underivatized lysines were detected, suggesting that freebase guanidination is quantitative. This contributes to better sensitivity and reproducibility. The benefits of freebase

guanidination were most apparent under demanding sample conditions as was seen after co-derivatization with *N*-terminal sulfonation, which is known to negatively impact sensitivity. In the 2-DGE workflow, sample preparation was streamlined by avoiding time-consuming cleanup steps and PMF identification was obtained regardless of protein abundance.

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: qingl@hawaii.edu. Fax: (808) 965-3542.

### Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We thank Drs. Mihye Jeong, Are-Sun You, and Kyung Hun Park in the Korean Rural Development Administration for the lung samples and Dr. Il Kyu Cho for helpful discussions. We also thank Andy Rogers from Bruker for kindly training M.R.B. on automated data acquisition. This work was supported in part

by the National Institute on Minority Health and Health Disparities (Grant 8 G12 MD007601-28) of the National Institutes of Health and the Agriculture and Food Research Initiative Competitive Grant (Grant No. 2012-67011-19671) from the USDA National Institute of Food and Agriculture. M.R.B. is a NIFA predoctoral fellow.

## REFERENCES

- (1) Joss, J.; Molloy, M.; Hinds, L.; Deane, E. *Comp. Biochem. Physiol., Part D: Genomics Proteomics* **2007**, *2*, 150–164.
- (2) Beyea, J.; Olson, D.; Harvey, S. *Mol. Cell. Biochem.* **2009**, *321*, 197–204.
- (3) Sun, H.; Zhang, A.; Yan, G.; Han, Y.; Sun, W.; Ye, Y.; Wang, X. *J. Pharm. Biomed. Anal.* **2013**, *75*, 173–179.
- (4) Li, C.; Chen, Z.; Xiao, Z.; Wu, X.; Zhan, X.; Zhang, X.; Li, M.; Li, J.; Feng, X.; Liang, S. *Biochem. Biophys. Res. Commun.* **2003**, *309*, 253–260.
- (5) Choi, J.; Sullards, M.; Olzmann, J.; Rees, H.; Weintraub, S.; Bostwick, D.; Gearing, M.; Levey, A.; Chin, L.-S.; Li, L. *J. Biol. Chem.* **2006**, *281*, 10816–10824.
- (6) Li, C.; Xiao, Z.; Chen, Z.; Zhang, X.; Li, J.; Wu, X.; Li, X.; Yi, H.; Li, M.; Zhu, G. *Proteomics* **2006**, *6*, 547–558.
- (7) Krause, E.; Wenschuh, H.; Jungblut, P. R. *Anal. Chem.* **1999**, *71*, 4160–4165.
- (8) Zenobi, R.; Knochenmuss, R. *Mass Spectrom. Rev.* **1998**, *17*, 337–366.
- (9) Brancia, F. L.; Oliver, S. G.; Gaskell, S. J. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 2070–2073.
- (10) Nakazawa, T. *Curr. Proteomics* **2006**, *3*, 33–54.
- (11) Beardsley, R. L.; Karty, J. A.; Reilly, J. P. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 2147–2153.
- (12) Hale, J. E.; Butler, J. P.; Knierman, M. D.; Becker, G. W. *Anal. Biochem.* **2000**, *287*, 110–117.
- (13) Keough, T.; Lacey, M. P.; Youngquist, R. S. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 2348–2356.
- (14) Keough, T.; Youngquist, R. S.; Lacey, M. P. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 7131–7136.
- (15) Klimek-Ochab, M.; Raucchi, G.; Lejczak, B.; Forlani, G. *Res. Microbiol.* **2006**, *157*, 125–135.
- (16) Nuñez, A.; Fishman, M.; Fortis, L.; Cooke, P.; Hotchkiss, A. J. *Sci. Food Agric.* **2009**, *57*, 10951–10958.
- (17) Stegemann, C.; Kolobov, A.; Leonova, Y.; Knappe, D.; Shamova, O.; Ovchinnikova, T.; Kokryakov, V.; Hoffmann, R. *Proteomics* **2009**, *9*, 1364–1373.
- (18) Zhang, L.; Reilly, J. J. *Proteome Res.* **2010**, *9*, 3025–3034.
- (19) Kristiansen, E.; Ramlov, H.; Højrup, P.; Pedersen, S.; Hagen, L.; Zachariassen, K. *Insect Biochem. Mol. Biol.* **2011**, *41*, 109–117.
- (20) Warwood, S.; Mohammed, S.; Cristea, I. M.; Evans, C.; Whetton, A. D.; Gaskell, S. J. *Rapid Commun. Mass Spectrom.* **2006**, *20*, 3245–3256.
- (21) Lemmel, C.; Weik, S.; Eberle, U.; Dengjel, J.; Kratt, T.; Becker, H.-D.; Rammensee, H.-G.; Stevanovic, S. *Nat. Biotechnol.* **2004**, *22*, 450–454.
- (22) Yang, S.-J.; Nie, A.-Y.; Zhang, L.; Yan, G.-Q.; Yao, J.; Xie, L.-Q.; Lu, H.-J.; Yang, P.-Y. *J. Proteomics* **2012**, *75*, 5797–5806.
- (23) Beardsley, R. L.; Reilly, J. P. *Anal. Chem.* **2002**, *74*, 1884–1890.
- (24) Rogowska-Wrzesinska, A.; Le Bihan, M.-C.; Thaysen-Andersen, M.; Roepstorff, P. *J. Proteomics* **2013**, *88*, 4–13.
- (25) Cockrill, S. L.; Foster, K. L.; Wildsmith, J.; Goodrich, A. R.; Dapron, J. G.; Hassell, T. C.; Kappel, W. K.; Scott, G. B. I. *BioTechniques* **2005**, *38*, 301–304.
- (26) Sergeant, K.; Samyn, B.; Debyser, G.; Van Beeumen, J. *Proteomics* **2005**, *5*, 2369–2380.
- (27) Samyn, B.; Sergeant, K.; Carpentier, S.; Debyser, G.; Panis, B.; Swennen, R.; Van Beeumen, J. *J. Proteome Res.* **2006**, *6*, 70–80.
- (28) Imbeah, M.; Angkanaporn, K.; Ravindran, V.; Bryden, W. L. *J. Sci. Food Agric.* **1996**, *72*, 213–218.
- (29) Laemmli, U. K. *Nature* **1970**, *227*, 680–685.
- (30) Bradford, M. M. *Anal. Biochem.* **1976**, *72*, 248–254.
- (31) Bio-Rad Laboratories, Inc. Rev F, Bulletin 2651.
- (32) Cold Spring Harbor. *Purifying Proteins for Proteomics: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2004.
- (33) Russell, W. K.; Park, Z.-Y.; Russell, D. H. *Anal. Chem.* **2001**, *73*, 2682–2685.
- (34) Suckau, D.; Resemann, A.; Schuerenberg, M.; Hufnagel, P.; Franzen, J.; Holle, A. *Anal. Bioanal. Chem.* **2003**, *376*, 952–965.
- (35) Cottrell, J.; London, U. *Electrophoresis* **1999**, *20*, 3551–3567.
- (36) Tannu, N. S.; Wu, J.; Rao, V. K.; Gadgil, H. S.; Pabst, M. J.; Gerling, I. C.; Raghov, R. *Anal. Biochem.* **2004**, *327*, 222–232.
- (37) Schuerenberg, M.; Luebbert, C.; Eickhoff, H.; Kalkum, M.; Lehrach, H.; Nordhoff, E. *Anal. Chem.* **2000**, *72*, 3436–3442.
- (38) Kinter, M.; Sherman, N. E. *Protein Sequencing and Identification Using Tandem Mass Spectrometry*; Wiley: New York, 2000.
- (39) Hirsch, J.; Hansen, K. C.; Burlingame, A. L.; Matthay, M. A. *Am. J. Physiol.: Lung Cell. Mol. Physiol.* **2004**, *287*, L1–L23.
- (40) Siegel, R.; Naishadham, D.; Jemal, A. *CA: Cancer J. Clin.* **2013**, *63*, 11–30.
- (41) Bairoch, A.; Boeckmann, B.; Ferro, S.; Gasteiger, E. *Brief. Bioinf.* **2004**, *5*, 39–55.
- (42) Magrane, M.; UniProt Consortium. *Database* **2011**, *2011*, bar009.

**APPENDIX F. SYNTHESIS OF AN IBERIOTOXIN DERIVATIVE BY  
CHEMICAL LIGATION: A METHOD FOR IMPROVED YIELDS OF  
CYSTEINE-RICH SCORPION TOXIN PEPTIDES.<sup>3</sup>**

---

<sup>3</sup> Reproduced in part with permission from Bingham, J.-P.; Chun, J. B.; Ruzicka, M. R.; Li, Q. X.; Tan, Z.-Y.; Kaulin, Y. A.; Englebretsen, D. R.; Moczydlowski, E. G. Synthesis of an iberiotoxin derivative by chemical ligation: A method for improved yields of cysteine-rich scorpion toxin peptides. *Peptides* **2009**, *30*, 1049–1057. Copyright © 2009, rights managed by Elsevier, Inc.



## Synthesis of an iberiotoxin derivative by chemical ligation: A method for improved yields of cysteine-rich scorpion toxin peptides

Jon-Paul Bingham<sup>a,b,\*</sup>, Joycelyn B. Chun<sup>a</sup>, Margaret R. Ruzicka<sup>a</sup>, Qing X. Li<sup>a</sup>, Zhi-Yong Tan<sup>c</sup>, Yuri A. Kaulin<sup>d</sup>, Darren R. Englebretsen<sup>e</sup>, Edward G. Moczydlowski<sup>b</sup>

<sup>a</sup> Department of Molecular Biosciences and Bioengineering, College of Tropical Agriculture and Human Resources, University of Hawaii, Honolulu, HI 96822, USA

<sup>b</sup> Department of Biology, Clarkson University, Potsdam, NY 13699, USA

<sup>c</sup> Department of Internal Medicine, University of Iowa, Iowa City, IA 52242, USA

<sup>d</sup> Department of Pathology, Anatomy and Cell Biology, Jefferson Medical College of Thomas Jefferson University, Philadelphia, PA 19107, USA

<sup>e</sup> Chemistry Department, Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand

### ARTICLE INFO

#### Article history:

Received 11 December 2008

Received in revised form 10 March 2009

Accepted 10 March 2009

Available online 26 March 2009

#### Keywords:

Peptide synthesis

Chemical ligation

Toxin

Scorpion

Iberiotoxin

BK Ca<sup>2+</sup>-activated K<sup>+</sup> channel

K<sub>Ca</sub>1.1

### ABSTRACT

Automated and manual solid phase peptide synthesis techniques were combined with chemical ligation to produce a 37-residue peptide toxin derivative of iberiotoxin which contained: (i) substitution of Val<sup>16</sup> to Ala, to facilitate kinetic feasibility of native chemical ligation, and; (ii) substitution of Asp<sup>19</sup> to orthogonally protected Cys-4-MeOBzl for chemical conjugate derivatization following peptide folding and oxidation. This peptide ligation approach increased synthetic yields approximately 12-fold compared to standard linear peptide synthesis. In a functional inhibition assay, the ligated scorpion toxin derivative, iberiotoxin V16A/D19-Cys-4-MeOBzl, exhibited 'native-like' affinity ( $K_d = 1.9$  nM) and specificity towards the BK Ca<sup>2+</sup>-activated K<sup>+</sup> Channel (K<sub>Ca</sub>1.1). This was characterized by the rapid association and slow dissociation rates ( $k_{on} = 4.59 \times 10^5$  M<sup>-1</sup> s<sup>-1</sup>;  $k_{off} = 8.65 \times 10^{-4}$  s<sup>-1</sup>) as determined by inhibition of macroscopic whole-cell currents of cloned human K<sub>Ca</sub>1.1 channel. These results illustrate the successful application of peptide chemical ligation to improve yield of cysteine-rich peptide toxins over traditional solid phase peptide synthesis. Native chemical ligation is a promising method for improving production of biologically active disulfide containing peptide toxins, which have diverse applications in studies of ion-channel function.

© 2009 Elsevier Inc. All rights reserved.

### 1. Introduction

Native chemical ligation is a strategic tool for the assembly of poor yielding peptide and poly-peptide sequences. This simple

chemical approach makes possible the complete solution-phase condensation of two individual fully deprotected peptide moieties of varying lengths to form a site-directed native-state peptide bond [3,4,6]. This approach enables sequential 'block peptide' elongation in a selective and controlled manner. Native chemical ligation deviates from the traditional convergent solid phase peptide synthesis (SPPS) strategies, with the latter incorporating the coupling of a fully orthogonally protected peptide fragment onto an existing peptide-resin support, an approach typically directed by extending normal amino acid coupling procedures [1,10].

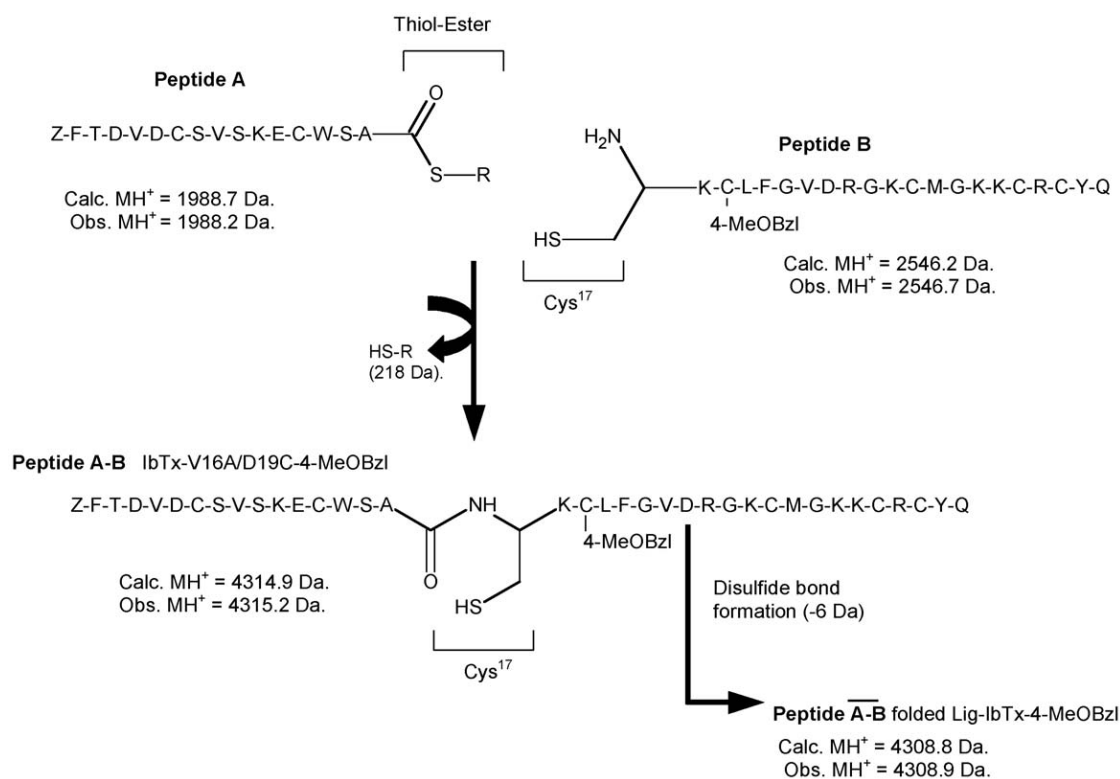
Extensions of native chemical ligation now include the chemical and semi-synthesis of large functional proteins (>10 kDa). This includes ion channels [4,14,15,25,39] and the incorporation of small synthetic peptide fragments into larger recombinant proteins [17], which enables the introduction of various non-native amino acids [23,38], fluoro-derivatives (see Ref. [25]) and affinity labels [36]. These represent unique products to the native chemical ligation strategy, with most unachieved via convergent strategies.

Cysteine-rich peptide toxins provide ideal models for chemical synthesis via native chemical ligation, as many are difficult to

\* Corresponding author at: Department of Molecular Biosciences and Bioengineering, College of Tropical Agriculture and Human Resources, University of Hawaii, Honolulu, HI 96822, USA. Fax: +1 808 965 3542.

E-mail address: [jbingham@hawaii.edu](mailto:jbingham@hawaii.edu) (J.-P. Bingham).

Abbreviations: SPPS, solid phase peptide synthesis; BK/K<sub>Ca</sub>1.1, large conductance Ca<sup>2+</sup>-activated K<sup>+</sup> channel; K<sub>v</sub>, voltage-gated K<sup>+</sup> channels; HSlo, human BK channel; ChTx, charybdotoxin; IbTx, iberiotoxin; IbTx-D19K-LC-biotin, iberiotoxin-D19K-LC-biotin; Lig-IbTx-MeOBzl, native chemical ligated iberiotoxin-V16A/D19C-4-MeOBzl; AcM, S-acetamidomethyl-L-cysteine; 4-MeOBzl, 4-methoxybenzyl; MBHA, 4-methylbenzhydrylamine; DHB, 2,5-dihydroxybenzoic acid; DMF, dimethylformamide; DIEA, N,N-diisopropylethylamine; HBTU, 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate; DCM, dichloromethane; HF, hydrogen fluoride; TFA, trifluoroacetic acid; MeCN, acetonitrile; GnHCl, guanidine hydrochloride; AC<sub>2</sub>O, acetic anhydride; RP-HPLC/UV, reverse phase high pressure liquid chromatography interfaced UV spectroscopy; LC/MS, reverse phase high pressure liquid chromatography interfaced mass spectroscopy; MS, mass spectroscopy; MH<sup>+</sup>, mono isotopic molecular mass; HEK293, human embryonic kidney cell line; MALDI-TOF, matrix-assisted laser desorption ionization time of flight.



**Fig. 1.** Schematic representation of the chemical native ligation process involving two peptide fragments to produce a bioengineered Iberitoxin derivative. Peptide A, the leading fragment (16mer) contains both the C-terminal thiol ester, as well as the substituted Val<sup>16</sup> to Ala to promote kinetic feasibility during the ligation reaction. Peptide B, the Following Fragment (21mer) contains a free N-terminal cysteine to participate and regenerate its chemical functionality during and after ligation respectively – the latter is required for sequential disulfide bond formation. Peptide B also contains an orthogonally protected Cys-4-MeOBzI at position 19 within the full ligated parent sequence. This specific position has demonstrated to be a point for chemical bioconjugation or derivatization which does not greatly affect the biological activity or specificity of the parent toxin. Sequential molecular masses, as determined by ESMS, illustrate the progressive chemical modifications that occur which result in the full length folded parent toxin of 37 amino acids containing 3 disulfide bonds (oxidation is indicated by curved bar above peptide fragment A–B).

synthesize or produce in high yields using standard linear SPPS techniques due to their length, hydrophobic nature and multi-cysteine content [37]. The requirements for native chemical ligation peptide candidacy are: (i) an introduced C-terminal thiol ester in the lead N-terminal peptide segment, and (ii) an N-terminal Cysteine in the following C-terminal peptide segment (see Fig. 1). The N-terminal cysteine of (ii) can be easily accommodated by the relevant abundance and sequential positioning of cysteine moieties throughout the length of many classes of peptide toxins. We envisaged that synthesis of longer (>40 amino acid) peptide toxins by native chemical ligation of ~20 amino acid fragments would offer potential advantages in yields due to: (a) reduction of inherent cumulative stepwise SPPS deletions, (b) relative ease of synthesizing and purifying shorter peptides, and (c) avoiding problems of side reactions, disulfide cross-linking and hydrophobic aggregation encountered during acid cleavage of longer, hydrophobic, multiple cysteine-containing peptides.

An important consideration in native chemical ligation synthesis is the kinetic dependence of the ligation reaction on the side chain of the C-terminal amino acid thiol ester of the N-terminal peptide segment (see ref. [16]). The success of ligation is dependent on its selection. Some amino acids at this position (e.g., Leu, Thr, Val, Ile, Pro) have slow reaction or coupling rates and lead to poor yields of the ligated target species. In some cases, substitution at this position may be required to give reasonable yields of the ligated product. However, a single amino acid substitution may also have significant consequences on a toxin's specificity, targeting and overall pharmacological behavior, in some cases rendering a toxin candidate biologically inactive (see

Refs. [33,34]). Here, careful design based on comparative structural and pharmacological identity of similar peptide toxin sequences is an important aspect of successful peptide toxin bioengineering.

In this paper, we illustrate the design and synthesis of a 37 amino acid scorpion toxin derivative, iberitoxin V16A/D19-Cys-4-MeOBzI, via native chemical ligation. The derivative's retention of pharmacological specificity and selectivity to the BK Ca<sup>2+</sup>-activated K<sup>+</sup> channel (K<sub>Ca</sub> 1.1) is verified by electrophysiological assay. The native chemical ligation approach to increased synthetic yields offers an alternative route to bioengineering cysteine stabilized  $\alpha/\beta$  motif containing scorpion toxins, which have numerous applications for biological studies of ion channels such as K<sub>Ca</sub>1.1.

## 2. Materials and methods

### 2.1. Synthesis of 3-Mercapto-Propionic Acid-Leucine (MPAL) base resin

4-Methylbenzhydrylamine [MBHA] resin (0.79 mmole g<sup>-1</sup>; Peptides International) was pre-swollen overnight in dimethylformamide (DMF) and 650  $\mu$ L *N,N*-diisopropylethylamine (DIEA). After washing the swollen resin with DMF, Boc-Leu (2 mmole) was activated with 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) (0.5 M in DMF, 4 mL) and 347  $\mu$ L of DIEA and allowed to couple for 40 min. After both DMF and dichloromethane (DCM) washes (each 40 mL), the Boc group was removed with 50% (v/v) TFA/DCM (2 min  $\times$  5 min). 3,3'-dithiopropionic acid (2 mmole) dissolved in HBTU (0.5 M in DMF, 8 mL) was then mixed with sufficient DIEA to both activate and neutralize the resin and allowed to couple for 40 min. The resin

was then washed with DMF and then treated with 650  $\mu\text{L}$  ethanolamine and 200  $\mu\text{L}$  DIEA in 5 mL DMF for 40 min to cap resin-bound-OBT esters. A DMF wash was followed by reduction of the resin-bound disulfide with 2-mercaptoethanol (650  $\mu\text{L}$  2-mercaptoethanol, 100  $\mu\text{L}$  DIEA in 5 mL DMF) for 1 h. Following another wash with DMF, Boc-Ala (2 mmole) was activated with HBTU, as above, and immediately coupled for 20 min.

## 2.2. Boc SPPS of N-terminal portion of Iberitoxin

Peptide A (ZFTDVDCSVSKECWSA<sup>16</sup>-[MPAL]) was manually assembled via Boc SPPS using *in-situ* neutralization with 2 mmole Boc-amino acid per 10 min single coupling, as described by Schönöler et al. [31], using the above described [MPAL]-loaded resin. After each coupling step, yields were determined by measuring residual free amine with the quantitative ninhydrin assay [30] and maintained above 99.5% or end-capped with Ac<sub>2</sub>O (100  $\mu\text{L}$  in 5 mL DMF) if yield was not achieved. Side chain protecting groups were Asp(cyclohexyl), Cys(4-methylbenzyl), Glu(cyclohexyl), Lys(2-chlorobenzoyloxycarbonyl), Ser(benzyl) and Thr(benzyl). Tryptophan was unprotected. Pyroglutamic acid (Z) was coupled in the same way as the other residues. Upon completion of the synthesis, the resin was washed with DMF followed by DCM and then dried under N<sub>2</sub>.

### 2.2.1. HF cleavage

Peptide-[MPAL]-resin was stirred with 20 mL HF and 1 mL *p*-cresol for one hour at 0 °C. HF was removed *in vacuo* and the residue treated with 100 mL of cold methyl *tert*-butyl-ether to precipitate the peptide. The solids were collected onto a glass sinter, washed twice with 50 mL methyl *tert*-butyl-ether, and the peptide dissolved in 100 mL 50/50 (v/v) acetonitrile (MeCN)/H<sub>2</sub>O containing 0.1% (v/v) TFA. The solution was lyophilized and stored at -20 °C until required.

## 2.3. Preparative RP-HPLC/UV of N-terminal portion of Iberitoxin: Peptide A (ZFTDVDCSVSKECWSA<sup>16</sup>-[MPAL])

The crude peptide was dissolved in 50 mL unbuffered GnHCl (pH 4) and loaded onto a Phenomenex 21.2 mm  $\times$  250 mm C<sub>4</sub> column (300 Å, 10–15  $\mu\text{m}$ ) equipped with a 21.2 mm  $\times$  60 mm guard column packed with the same material. The initial mobile phase was 100% Solvent A. The GnHCl was eluted from the column using Solvent A at 10 mL min<sup>-1</sup>, followed by elution of the peptide using a two step gradient (0–20% Solvent B 0–20 min, 20–45% Solvent B 20–80 min). Fractions were collected and analyzed by infusion MS. Those containing the target peptide were pooled and lyophilized.

## 2.4. Fmoc SPPS of C-terminal portion of Iberitoxin

Peptide B [<sup>17</sup>CKC(4-MeOBzl)LFGVDRGKCMGKKRCYQ<sup>37</sup>], 100  $\mu\text{M}$  scale was assembled automatically via solid phase Fmoc coupling, at Keck Peptide Facility (Yale University) in the following manner: Peg-Resin (0.18 mmol g<sup>-1</sup>) was washed three times with DMF (2.5 mL), deprotected twice with 20% (v/v) piperidine (2.5 mL, 2  $\times$  0.5 min), and then washed six more times with DMF (2.5 mL). Fmoc-amino acids (0.5 mmole) were activated by 1.5 mL 0.4 M HBTU/DMF/*N*-methyl-Morpholine and double coupled (20 min.). Upon coupling completion, the peptide-resin was washed with DMF (2.5 mL) three times. Side chain protecting groups were: Cys(4-MeOBzl) [position 19 only], Cys(Trt), Lys(Boc) Asp(tBu), Arg(Pbf), and Gln(Trt). On completion of the synthesis the resin was washed with DMF followed by DCM and then dried under N<sub>2</sub>. The peptide was cleaved and reverse-phase high performance liquid chromatography/ultra-violet detection (RP-HPLC/UV) purified using a semi-preparative

Vydac C<sub>18</sub> column (300 Å, 5  $\mu\text{m}$ , 7.8  $\times$  250 mm). Samples were eluted using a linear 1% min<sup>-1</sup> gradient of organic (90/10 MeCN/0.08%, v/v TFA aq.; Solvent B) against 0.1% (v/v) TFA aq. (Solvent A) at a flow rate of 1 mL min<sup>-1</sup>.

### 2.4.1. TFA cleavage

Resulting Fmoc peptides were subjected to a double cleavage with 96% (v/v) TFA in the presence of Anisole (2%, v/v) and H<sub>2</sub>O (2%, v/v), acting as protecting group scavengers, for 2 h at 24 °C. The resulting cleaved peptide material was recovered as previously described and stored at -20 °C as lyophilized powder until required.

## 2.5. Analytical RP-HPLC/UV

Synthetic peptides were separated/quantified either by C<sub>18</sub> or C<sub>4</sub> analytical RP-HPLC/UV (Vydac; 5  $\mu\text{m}$ , 300 Å, 4.6  $\times$  250 mm), using the described conditions above. Eluants were monitored at 223 or 280 nm.

## 2.6. Native ligation

Typically N-terminal-peptide-[MPAL] (Peptide A) and 4-MeOBzl containing C-terminal peptide (peptide B), in equal molar ratios (~30  $\mu\text{M}$ ), were dissolved in 4 mL 6 M GnHCl, 100 mM sodium phosphate pH 7.9 and mixed with 30  $\mu\text{L}$  each of thiophenol and benzylmercaptan then allowed to stand at 40 °C. After 18 h a sample was analyzed by LC/MS to confirm the reaction had taken place. After 21 h the solution was diluted with 50 mL unbuffered 6 M GnHCl and extracted with three times 40 mL methyl *tert*-butyl-ether. After removal of a sample for LC/MS analysis, the ligated IbTx-V16A/D19C-4-MeOBzl (Lig-IbTx-MeOBzl) was immediately purified by RP-HPLC (Vydac C<sub>4</sub> 10 mm  $\times$  250 mm column, gradient 0–25% Solvent B 0–25 min, 25–37% Solvent B 25–85 min). Fractions were analyzed by infusion MS, pooled and lyophilized.

## 2.7. Fmoc SPPS of IbTx-D19C-Acm, IbTx-D19K-LC-biotin and linear IbTx-V16A/D19C-4-MeOBzl

Three IbTx analogues were linearly synthesized (0.5 mmole scale), all using the same automated Fmoc synthesizer and cleavage protocols as described in Section 2.4. Here, the standard Fmoc-amino acid and corresponding protecting groups were as previously indicated, except for the use of Fmoc-Cys-Acm in position 19 for IbTx-D19C-Acm [ZFTDVDCSVSKECWSVCKC(-Acm)LFGVDRGKCMGKKRCRCYQ], Fmoc-Lys(biotin-LC)-OH [N- $\alpha$ -Fmoc-N- $\epsilon$ -(D-biotin-6-amidocaproate)-L-lysine; Anaspec Inc.] in position 19 for IbTx-D19K-LC-biotin [ZFTDVDCSVSKECWSVCKC(-LC-biotin)LFGVDRGKCMGKKRCRCYQ], and Fmoc-Cys(4-MeOBzl)-OH in position 19 for linear IbTx-V16A/D19C-4-MeOBzl [ZFTDVDCSVSKECWSACKC(-4-MeOBzl)LFGVDRGKCMGKKRCRCYQ].

## 2.8. Oxidation of IbTx analogues

Lig-IbTx-MeOBzl, IbTx-D19C-Acm and IbTx-D19K-LC-biotin (10 mg) were sonicated in 20 mL of oxidation buffer of 2 M urea, 0.1 M NaCl, 0.1 M glycine, pH 7.8 in an effort to aid peptide solubility and decrease peptide aggregation potential during oxidation. Peptides were allowed to oxidize in air by stirring at 22 °C for 18 h. Slow oxidation of linear IbTx-V16A/D19C-4-MeOBzl (10 mg) was aided with the introduction of 50% (v/v) 2-propanol in to the above N<sub>2</sub> purged oxidation buffer (20 mL) and allowed to stir at 22 °C for 36 h. Peptides were immediately acidified (neat TFA), filtered (0.22  $\mu\text{m}$ ) and subjected to preparative RP-HPLC/UV purification, as described in Section 2.3.

### 2.9. Purification and characterization of folded peptides

Purification of the folded toxin derivatives was achieved by RP-HPLC/UV—first on a semi-preparative column, followed by narrow-bore analysis, peptide co-elution and quantification using Vydac C<sub>18</sub> columns both subjected to a linear 1% min<sup>-1</sup> gradient of 0.1% (v/v) TFA aq. against 90% (v/v) MeCN in 0.08% (v/v) TFA aq. over 60 min. UV absorption was monitored at 223 nm. Electrospray mass spectroscopy of the folded and purified peptides was performed on a LQC Thermo Finnigan spectrometer.

### 2.10. Peptide quantification

Peptide concentrations were determined by RP-HPLC/UV integration of peptide samples [Peptide A and B; IbTx-D19K-LC-biotin, standard linear IbTx-V16A/D19C-4-MeOBzl, Lig-IbTx-MeOBzl and IbTx-D19C-Acm], which had previously undergone quantitative amino acid hydrolysis and PTH derivatization and quantification (Keck Peptide Facility, Yale University).

### 2.11. MALDI TOF/TOF analysis

RP-HPLC/UV purified peptide fractions in aqueous 0.1% TFA were mixed 1:1 with matrix solution (2,5-dihydroxybenzoic acid [DHB] in 1:1 0.1% TFA:MeCN) and spotted onto a thin layer of dried matrix (DHB) saturated in methanol on the target plate. The spots were dried under a stream of N<sub>2</sub> gas. Parent ions were identified on the Ultraflex III MALDI TOF/TOF (Bruker Daltonics), controlled by the Compass 1.2 SR1 software package (Bruker), in positive reflector mode. Parent ions were further selected in the LIFT-MS/MS mode. Peptide II Calibration Mix (Bruker) was used for external calibration.

### 2.12. Planar bilayer assay of Lig-IbTx-MeOBzl

Single-channel recording of K<sub>Ca</sub>1.1 channels from rat skeletal muscle in planar lipid bilayers was performed as previously described [7]. The recording solution was 200 mM KCl, 10 mM Mops-KOH, pH 7.4 on both sides of the bilayer, with 0.2 mM CaCl<sub>2</sub> on the intracellular side and 0.1 mM EDTA plus 0.1 mg mL<sup>-1</sup> bovine serum albumin on the extracellular side as defined by native asymmetry of the channel.

### 2.13. Electrophysiological assay of Lig-IbTx-MeOBzl

Lig-IbTx-MeOBzl was tested for blocking activity against human K<sub>Ca</sub>1.1 channels. An N-terminal Flag-BK version of the human K<sub>Ca</sub>1.1 channel (HSlo) was stably expressed in a human cell line (HEK293) as previously described [2]. Channel activity was monitored by the whole-cell patch clamp electrophysiological technique [18] using an Axon 200B amplifier and Clampex software (Axon Instruments) for voltage pulse programming and data acquisition. The pipette (intracellular) solution containing 0.5 μM free Ca<sup>2+</sup> [35] was 100 mM KCl, 1 mM MgCl<sub>2</sub>, 1 mM CaCl<sub>2</sub>, 1.3 mM EGTA, 10 mM Hepes-KOH, pH 7.2. The standard bath (extracellular) solution was 145 mM NaCl, 5 mM KCl, 2 mM CaCl<sub>2</sub>, 2 mM MgCl<sub>2</sub>, 10 mM Hepes-KOH, pH 7.4. Briefly, macroscopic HSlo currents were recorded from single HEK293-HSlo cells in whole-cell mode by stepping the membrane voltage from a holding potential of -80 mV to a series of consecutive test potentials delivered in increments of +10 mV up to +100 mV for 100 ms. Current records, low-pass filtered at 5 kHz, were corrected for capacitance transients and leak current using a negative P/5 pulse protocol. Synthetic scorpion toxins and paxilline (Sigma–Aldrich) were applied to cells by continuous gravity perfusion of the bath solution (0.5 mL min<sup>-1</sup>) into a low volume cell recording chamber (RC24E, Warner Instruments).

## 3. Results

Stepwise linear SPPS approaches failed to provide a free 'spinster thiol' D19C derivative of IbTx. Failure is suspected as a result of incorrect disulfide pairing or intra-molecular disulfide bonding. Possible confirmation is provided by IbTx-D19C-Acm, an orthogonally protected 'spinster thiol' derivative. This oxidized biologically active derivative (Obs. MH<sup>+</sup> 4265.4 Da, Calc. MH<sup>+</sup> 4265.8 Da) demonstrated a K<sub>d</sub> of 1.6 nM via lipid bilayer (Table 1), and demonstrated characteristic discrete long-lived blocked states upon assay by single BK<sub>Ca</sub> channels as observed in other αKTx1 toxins (see [11,2]; data not shown). Synthetic yields of IbTx-D19C-Acm were ~2% per 100 mg peptide oxidized. Selective thiol deprotection using AgBF<sub>4</sub>, as described by Yoshida et al. [40], resulted in a biologically inactive form of IbTx-D19C. Molecular mass analysis indicated the presence of methionine sulfoxide (a difference of +16 Da; Obs. MH<sup>+</sup> 4211.2 Da, Calc. MH<sup>+</sup> 4210.8 Da) that accounted for biological inactivity by a structural modification of Met<sup>29</sup> (see Ref. [32]). With these results, alternative approaches were examined to construct an IbTx orthogonally protected 'spinster thiol' template derivative in high yields.

### 3.1. Peptide assembly

Synthesis of IbTx-D19K-LC-biotin (as discussed in Section 2) provided 183 mg g<sup>-1</sup> of resin, with the target peptide representing 24% of total RP-HPLC/UV profile at 223 nm of the crude cleaved material. Typically upon oxidation and purification, 2.8–3.3 mg of folded bioactive material was recovered for each 100 mg oxidized (2.8–3.3% yield).

Linearly synthesized IbTxV16A/D19C-4-MeOBzl provided 212 mg of material per gram of resin via double cleavage, with the target peptide representing 15% of total RP-HPLC/UV profile at 223 nm. Upon oxidation and purification, 1 mg of folded material was recovered for each 100 mg oxidized (~1% yield).

For native chemical ligation, Peptide A, 16 amino acids in length; IbTx [1–16]V16A-COSR (Obs. MH<sup>+</sup> 1988.2 Da, Calc. MH<sup>+</sup> 1988.7 Da; Fig. 1), was manually assembled via Boc SPPS on the described [MPAL] base resin. Typically single HF cleavage of one gram of resin provided 260 mg of peptide material, with the target peptide representing 37% of total RP-HPLC/UV profile at 223 nm. Upon purification of the reduced material, <99% purity, 14 mg was recovered for each 100 mg of crude peptide (~14% yield).

Peptide B, 21 amino acids in length (Fig. 1), was assembled via automated Fmoc SPPS. Single acid cleavage of Peptide B was undertaken with 180 mg of peptide recovered from one gram of resin. RP-HPLC/UV analysis indicated the presence of a dominant peak (62% of the total RP-HPLC/UV profile) correlating to the target thiol-protected material. Molecular mass analysis confirmed its presence (Calc. MH<sup>+</sup> 2546.2 Da and Obs. MH<sup>+</sup> 2546.7 Da, see Fig. 1). Upon purification of the reduced orthogonally protected material, <99% purity, 17.9 mg was recovered for each 100 mg of crude peptide (~18% yield).

### 3.2. Native ligation

Typically Peptide A and B, as observed in Fig. 1, were mixed in approximately equal molar ratio (13 μM; i.e. 30.4 and 33.7 mg, respectively), as described, with ligation completed after ~21 h. RP-HPLC/UV analysis illustrated a significant decrease in starting materials; Peptide A (Rt 26.1 min) and Peptide B (Rt 35.7 min), together with an increase in a newly abundant peak, A–B (Rt 40.5 min), representing the reduced ligated target (see Fig. 2). LC/MS illustrated the ligation process by the new ion species [M + 4H]<sup>4+</sup> 1079.9 and [M + 3H]<sup>3+</sup> 1439.4 (Fig. 2); Calc. MH<sup>+</sup> 4314.8 Da and Obs. MH<sup>+</sup> 4315.2 Da as illustrated in Fig. 2. Ligation

**Table 1**

The sequential development of the bioengineered IbTx.

|                                       | Sequence                                           | $K_d^a$ (nM)                                                                                  | Reference <sup>b</sup>        |
|---------------------------------------|----------------------------------------------------|-----------------------------------------------------------------------------------------------|-------------------------------|
| ChTx (WT)                             | ZFTNVSCCTSKECWSVQRLHNTSRGKCMNKKRCYS                | 8.8<br>17.6                                                                                   | [9]<br>This work              |
| ChTxR19C                              | ZFTNVSCCTSKECWSVQCLHNTSRGKCMNKKRCYS                | 8.8                                                                                           | [32]                          |
| ChTxV16E                              | ZFTNVSCCTSKECWSECVQLHNTSRGKCMNKKRCYS               | 18                                                                                            | [34]                          |
| IbTx (WT)                             | ZFTDVDCSVSKECWSVCKDLFGVDRGKCMGKKRCYQ               | <1                                                                                            | [11]                          |
| IbTxD19X                              | ZFTDVDCSVSKECWSVCKXLFVDRGKCMGKKRCYQ                | 1.6 <sup>c</sup> (X = Cys-Acm)<br>2.5 <sup>c</sup> (X = K-LC-biotin)<br>5.1 (X = K-LC-biotin) | This work<br>[2]<br>This work |
| IbTxV16A/D19X <sub>2</sub> (Lig-IbTx) | ZFTDVDCSVSKECWSACKX <sub>2</sub> LFGVDRGKCMGKKRCYQ | 2.5 <sup>c</sup> (X = K-LC-biotin-streptavidin)<br>1.9 (X <sub>2</sub> = Cys-4-MeOBzl)        | [2]<br>This work              |

Detailed pharmacological examination of both native and mutant forms of ChTx provided key information for the introduction of non-native amino acids into the sequence of IbTx. Initially ChTx – wild type (WT) was pharmacologically compared to two individual peptide mutants, ChTxR19C and ChTxV16E, which illustrated a maximal two-fold decreasing in affinity. Combining both mutations in a single sequence would likely not perturb toxin channel interactions. Given the structural and sequence similarities of ChTx and IbTx, and to test this hypothesis, the final combination of a double point mutation: (i) position 19 – this for later application derivatization, and (ii) position 16 – specifically introduced to promote ligational kinetics to maximize peptide synthesis, supports the sequential development of IbTx bioengineering in providing the biologically active parent template toxin ligated IbTxV16A/D19C-4-MeOBzl.

<sup>a</sup> As assigned via whole cell patch clamp.

<sup>b</sup> References as given in text.

<sup>c</sup>  $K_d$  is assigned by single channel analysis via lipid bilayer (see comments in Ref. [21]).

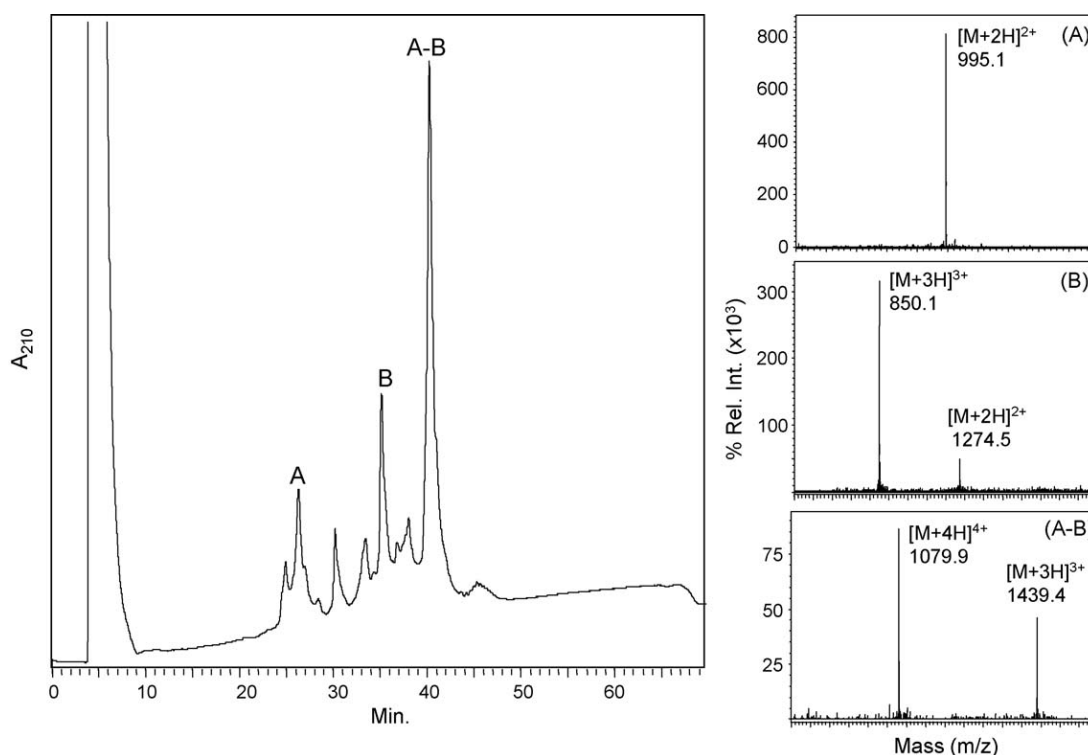
yields were determined by weight to be 32%, providing a purified unfolded ligated product of 20.5 mg. Upon batch-wise oxidation in urea/50% (v/v) 2-propanol buffer, 2.5 mg of folded ligated material was purified to <99% purity, which represents 12% per 100 mg of ligated peptide.

The ligated reduced product demonstrated the same RP-HPLC/UV retention time and molecular mass as the standard linear synthesized IbTx-V16A/D19C-4-MeOBzl. Both oxidized and purified ligated IbTx-V16A/D19C-4-MeOBzl (Lig-IbTx-MeOBzl) and standard linear synthesized oxidized IbTx-V16A/D19C-4-MeOBzl were RP-HPLC/UV co-injected in a 2:1 ratio, respectively, and demonstrated identical retention of 33.3 min as illustrated in Fig. 3. High resolution MALDI TOF/TOF of the final target provided

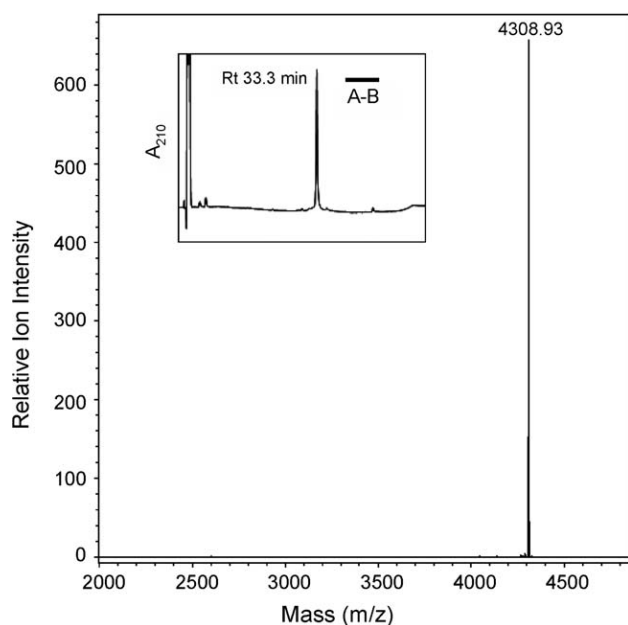
the molecular mass of  $MH^+$  4308.93 Da (Calc.  $MH^+$  4308.9 Da), confirming both full oxidation and purity, as illustrated in Fig. 3.

### 3.3. Biological assay of Lig-IbTx-MeOBzl

To determine whether Lig-IbTx-MeOBzl has biological activity against  $K_{Ca1.1}$  channels, we tested its effect on the cloned human  $K_{Ca1.1}$  channel (HSlo) expressed in a human cell line. Fig. 4A shows an example of HSlo channel activity recorded as macroscopic current from an HEK293-HSlo cell using the whole-cell patch clamp technique. Depolarization of the cell membrane with 0.5  $\mu$ M free  $Ca^{2+}$  in the intracellular pipette solution gives rise to outward  $K^+$  currents activated at membrane voltage more



**Fig. 2.** Representative RP-HPLC/UV profile obtained from the LC/MS of the crude ligated products at ~21 h. Peptide fragments A (Rt 26.1 min) and B (Rt 31.7 min) are labeled as the remaining initial starting products, as determined by  $m/z$  and RP-HPLC retention, with the reduced ligated product A–B (Rt 40.5 min) presenting the major UV contributor to the RP-HPLC profile (Phenomenex C<sub>4</sub>). Molecular mass analysis of starting materials Peptide A and B are shown in their multiple charge states, as indicated, with the final unfolded target Peptide A–B as derived from the LC/MS profile of crude ligated material.



**Fig. 3.** Analysis of the oxidized IbTx derivative. High-resolution molecular mass analysis of the final oxidized target ligated IbTxV16A/D19C-4-MeOBzl as determined by MALDI-TOF/TOF. Insert: RP-HPLC/UV chromatograph (Vydac C<sub>18</sub>) illustrating the 2:1 co-elution of the fully oxidized target peptide together with its identical non-ligated linear synthesized folded counterpart.

positive than +30 mV. Perfusion of the cell with bath solution containing 50 nM Lig-IbTx-MeOBzl inhibits HSlo current by 80% or greater depending on the test voltage (Fig. 4B). Paxilline is a tremorgenic fungal compound that has been characterized as a potent inhibitor of mammalian K<sub>Ca</sub>1.1 channels [21,35]. Addition of 1 μM paxilline to the bath solution inhibits greater than 95% of the outward current, demonstrating that most of the K<sup>+</sup> current recorded under these conditions is attributable to paxilline-sensitive K<sub>Ca</sub>1.1 channels.

The kinetics of the interaction of Lig-IbTx-MeOBzl with HSlo channels were also characterized by experiments illustrated in Fig. 4C. Rapid perfusion of bath solution containing Lig-IbTx-MeOBzl resulted in progressive time-dependent inhibition of K<sub>Ca</sub>1.1 current to a steady-state level reached several minutes after onset. The time course of toxin inhibition is a measure of the association rate of the toxin-channel binding reaction, since it is not limited by the fast delivery rate of the solution to the recording chamber. Similarly, slow recovery of K<sup>+</sup> current observed after perfusion with toxin-free bath solution is a measure of the time course of the toxin-channel dissociation reaction. Association and dissociation time courses for Lig-IbTx-MeOBzl, charybdotoxin (ChTx), and IbTx-D19K-LC-biotin from experiments of Fig. 4C were fit to a single exponential and the resulting time constants were used to calculate a second-order association rate constant ( $k_{on}$ ) and a first order dissociation rate constant ( $k_{off}$ ) for toxin binding according to the standard kinetic model for one-site binding reaction. Calculated  $k_{on}$  values are:  $4.6 \times 10^5 \text{ s}^{-1} \text{ M}^{-1}$  (Lig-IbTx-MeOBzl),  $1.3 \times 10^5 \text{ s}^{-1} \text{ M}^{-1}$  (ChTx), and  $9.9 \times 10^4 \text{ s}^{-1} \text{ M}^{-1}$  (IbTx-D19K-LC-biotin). Corresponding calculated  $k_{off}$  values are:  $8.7 \times 10^{-4} \text{ s}^{-1}$  (Lig-IbTx-MeOBzl),  $2.3 \times 10^{-3} \text{ s}^{-1}$  (ChTx), and  $5.1 \times 10^{-4} \text{ s}^{-1}$  (IbTx-D19K-LC-biotin). The equilibrium dissociation constant for binding of the various toxins estimated from the relationship,  $K_d = k_{off}/k_{on}$ , are: 1.9 nM (Lig-IbTx-MeOBzl), 17.6 nM (ChTx), and 5.1 nM (IbTx-D19K-LC-biotin), see Table 1. The  $K_d$  values calculated from the latter kinetic ratio compare favorably with  $K_i$  values of 2.9 nM for Lig-IbTx-MeOBzl and 13 nM for ChTx as determined by fitting the concentration dependence of steady-

state inhibition to a one-site binding isotherm (Fig. 4D). These results show that synthetic Lig-IbTx-MeOBzl has high affinity for the human K<sub>Ca</sub>1.1 channel and a slower dissociation rate than that of ChTx for mammalian K<sub>Ca</sub>1.1 channels as previously described for the native toxins [5,11,34].

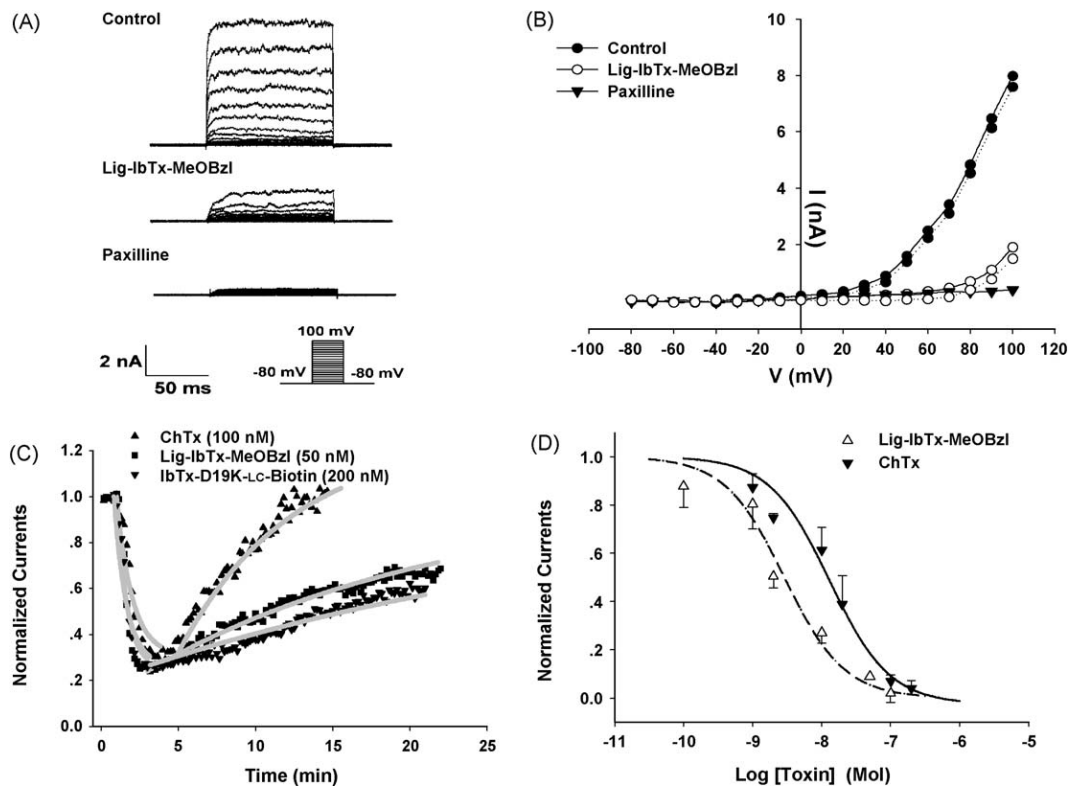
#### 4. Discussion

Chemical technologies of SPSS and bioconjugation facilitate transformation of peptide toxins into biological probes. Investigating alternative methods of peptide assembly may provide avenues for successful probe production in high yields. Our approach using native chemical ligation can improve yields, but requires careful due diligence in its design. The illustrated production of ligated IbTx-V16A/D19C-4-MeOBzl extends from comparative analysis of the structure activity relationships of ChTx (Table 1).

ChTx is a typical member of the αKTx1 family of scorpion toxins characterized by: (i) a short α-helix (typically residues 13–21), (ii) two stranded antiparallel β-sheets (typically residues 25–36; see Ref. [19]), and (iii) a defined channel-toxin interactive surface at amino acid positions Ser<sup>10</sup>, Trp<sup>14</sup>, Arg<sup>25</sup>, Lys<sup>27</sup>, Met<sup>29</sup>, Asn<sup>30</sup>, Arg<sup>34</sup> and Tyr<sup>36</sup> [34,32]. ChTx exhibits broad pharmacological recognition of various K<sub>v</sub> channels in contrast to the structural homolog IbTx, which is a highly specific blocker of K<sub>Ca</sub>1.1 ( $K_d < 1 \text{ nM}$ ; 9). ChTx and IbTx, both 37mers, exhibit 68% sequence identity (Table 1), including all the aforementioned essential toxin surface interactive side chains except for Asn 30 ChTx/Gly 30 IbTx. These structural features are compounded with identical positioning of cysteine moieties that participate in a 7–28, 13–33 and 17–35 disulfide-bridging configuration, providing a common fold of the peptide backbone in the three-dimensional structure (see Ref. [19]). A major difference between these two peptides is the overall number of charged amino acids present, 13 for ChTx and 9 for IbTx, which results in differing net charge of +4 and +2, respectively.

The derivative ChTx-R19C and the various labeled fluorescent [9] and radio-labeled [20] conjugates, coupled via a 'spinster thiol', retain K<sub>v</sub> channel selectivity, as the R19 modification maps to the 'backside' of the toxin with respect to the channel-binding surface. In IbTx-D19K-LC-biotin ( $K_d = 2.5 / 5.1 \text{ nM}$ ) and IbTx-D19C-Acm ( $K_d = 1.6 \text{ nM}$ ), derivatives also retain potency and selectivity towards K<sub>Ca</sub> 1.1 (Table 1; 2), indicating functional and structural conservation within the two toxins.

Combining the structure-activity profiles of ChTx [33,34], IbTx [5,8,19], IbTx-chimeras [12], and our previous IbTx derivative studies [2], a suitable thiol-ester ligation construct was designed, containing an orthogonally protected 'spinster thiol'. Native chemical ligation requires a C-terminal peptide segment (Peptide B) to contain an N-terminal Cys (see Fig. 1). In IbTx six cysteine moieties are present, with position Cys<sup>17</sup> providing a medial sequence position (see Table 1). The next requirement is a kinetically favorable amino acid for coupling at the C-terminal MPAL thioester end of the N-terminal peptide segment (Peptide A, Fig. 1). In IbTx Val<sup>16</sup> provides a kinetic hindrance to the ligation process. Hackeng et al. [16] showed that of all the neutral, aliphatic amino acids, a C-terminal Ala-thioester peptide provided a useful reaction rate in native chemical ligation. Other substitutions were either unreactive or hindered ligation yields (see Ref. [16]). To maximize chemical ligation of IbTx, Val<sup>16</sup> was changed to Ala. Pharmacological consequences of this mutation were considered; substitution of Val<sup>16</sup> by Glu in ChTx showed a two-fold decrease in binding affinity (see Table 1; [34]), indicating the absence of essential channel-toxin interactions. This is compounded by the amino acid 16-position also being on the non-interactive face, neighboring the established 19-position of both toxins ([2], see Fig. 5). Here we hypothesized that a neutral non-polar Val<sup>16</sup> to Ala substitution would not only lead to minimal alteration in channel-



**Fig. 4.** Comparison of the blocking effect of synthetic scorpion toxin derivatives on the human  $K_{Ca}1.1$  channel. (A) Inhibition of Hslo current by ligated IbTx-A16 V/D19C-4-MeOBzl (Lig-IbTx-OMeBzl) and paxilline. Records of Hslo current evoked by a series of voltage pulses from  $-80$  mV to  $+100$  mV in steps of  $+10$  mV are shown before (control) and after addition of  $50$  nM Lig-IbTx-OMeBzl and  $1$  mM paxilline to the extracellular bath perfusion solution. (B) Current amplitudes measured at the end of the step pulse from experiments in (A) are plotted versus the step voltage. Solid and dotted lines for control (filled circles) and Lig-IbTx-OMeBzl (open circles) data respectively correspond to current values before and after subtraction of values measured in the presence of paxilline (filled triangles). (C) Time course of inhibition and recovery after addition of toxin followed by perfusion with standard toxin-free bath solution. Hslo current was continuously monitored by voltage pulses from  $-80$  to  $+50$  mV ( $50$  ms) delivered once every  $10$  s. Values are normalized to peak current from each cell measured before addition of toxin. Overlaid data for three different experiments indicates current values measured during addition and removal of  $50$  nM Lig-IbTx-OMeBzl (squares),  $100$  nM ChTx (upward triangles), and  $200$  nM IbTx-D19K-LC-biotin (downward triangles). Gray lines correspond to exponential fits of data as described in the text. (D) Concentration-dependence of toxin inhibition. Normalized currents taken after steady-state inhibition by ChTx (filled triangles) or Lig-IbTx-OMeBzl (open triangles) as in (C) are plotted as a function of toxin concentration. Solid curves correspond to fits of the data for each toxin to:  $f_N = K_i / (K_i + [Tx])$ , where  $f_N$  is the fraction of normalized current,  $[Tx]$  is the toxin concentration, and  $K_i$  is the equilibrium dissociation constant for the toxin.  $K_i$  values are  $13.0$  nM for ChTx and  $2.9$  nM for Lig-IbTx-OMeBzl.

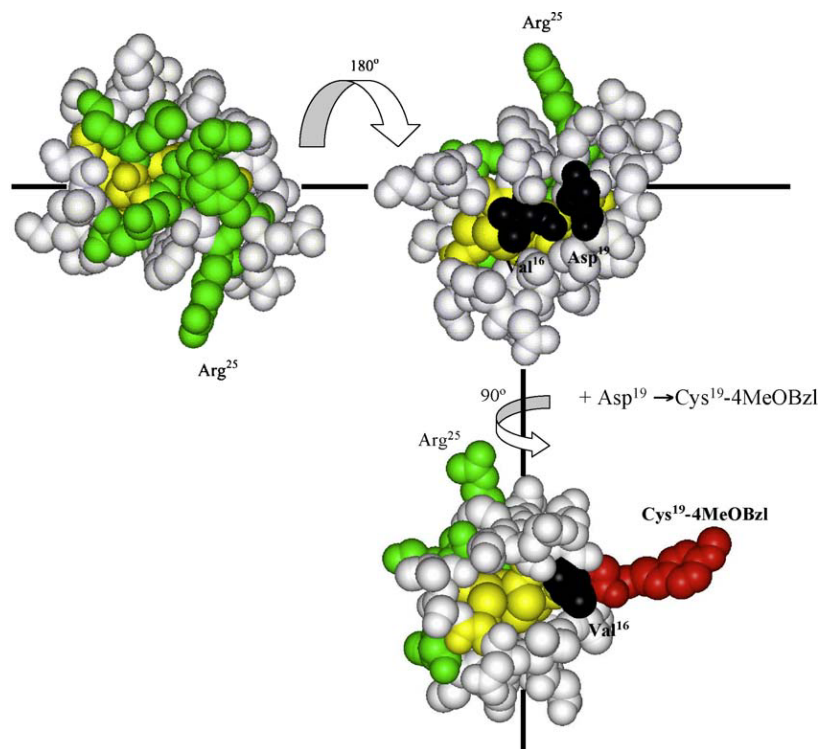
binding properties, but would in fact facilitate the production of the target material by native chemical ligation.

Support for our hypothesis comes from the observed changes in dissociation rates:  $K_d < 1$  nM (wild type) versus  $1.9$  nM for Lig-IbTx-MeOBzl as determined by whole-cell patch clamp (Table 1). The use of these opposing faces provides for molecular bifunctionality. We refer to these as 'Janus-molecules' since Janus is the Roman god of gates and doors, usually depicted by a double-faced head looking in opposite directions. Additional derivatization at this non-interactive face and resulting changes in pharmacology are illustrated in Table 1. Similar conclusions are reiterated in other  $\alpha$ KTx1 'Janus' toxins, such as ChTx-R19C and HgTx-A19C [32,28] respectively) and reflect the conserved three-dimensional structure observed within the  $\alpha$ KTx1 toxin family members [13,24,29]. This potentially provides an indication that these and other  $\alpha$ KTx1 toxins, together with their pharmacological resolving dipole moment, as discussed by Frémont et al. [8], can tolerate limited changes in charge state, hydrophobicity and mass additions. The extent of these tolerances at 19-position within IbTx and aforementioned  $\alpha$ KTx1 toxins is presently unknown.

Further support for our hypothesis in IbTx bioengineering comes from the 12-fold increase of yields by weight, of the folded ligated target material, in comparison to the standard full linear synthetic route with a post-cleavage oxidation that yields  $\sim 1\%$  (see [37]). Increased yields are attributed to (i) enhanced ligated linear production and (ii) improved yield in peptide folding. Initial yields

from standard linear synthesis of the 37mer IbTx-A16 V/D19C-4-MeOBzl ( $31.8$  mg of target peptide/gram resin) are poor in comparison to the smaller segmented construction of the 16mer-Peptide A ( $\sim 96.2$  mg of target/gram resin) and 21mer-Peptide B ( $\sim 111.6$  mg of target/resin gram), which are used for native chemical ligation. With ligation production reaching  $\sim 32\%$ , a  $\sim 100\%$  increase in production of the unoxidized linear target is achievable. Yet the major hurdle in both approaches is the folding of the peptide-toxin. Standard linear synthesis of the IbTx-A16 V/D19C-4-MeOBzl yielded  $1\%/100$  mg, while native chemical ligated oxidized target increased to  $12\%/100$  mg. This increase is related to the quality of linear starting material achievable, and purification steps prior to the ligation reaction. Minimization of thiol containing peptide deletion or N-terminal terminated products limits unwanted peptide polymerization.

Synthesis of a ligated bioactive derivative of IbTx suggests a number of potential research applications. Giangiacomo et al. [12] and Mullmann et al. [27] described the pharmacological properties of  $\alpha$ KTx1 toxin family chimeras using both IbTx and ChTx, and later Noxiustoxin as peptide-toxin templates, respectively. The combination of different native-like N- and C-terminal sequence segments resulted in non-native peptide toxins that exhibited unique biological activity [12]. Our approach of segmented 'block' style synthesis and native chemical ligation readily lends itself to toxin chimera production. Here, one could readily generate libraries of full-length peptide chimeras combining various N-



**Fig. 5.** Structural representation of interactive surfaces of native IbTx and its orthogonally protected derivative. Left structure highlights, in green, the amino acids and their relative positions that are essential for toxin-channel binding. The yellow highlights the disulfide bonds deep within the core of the toxin structure which are essential for maintaining biologically active structure/template. The right panel illustrates an 180° rotation of the same toxin. Here the use of the position Asp<sup>19</sup> has been a focal point for peptide bioconjugation and bioengineering, this being away from the interactive toxin-channel surfaces. Here the selective point for native chemical ligation, Val<sup>16</sup> (as indicated in black), lays adjacent to the non-interactive Asp<sup>19</sup>, it would be expected that this area, the backside of the molecule, plays little role in the toxin's biological specificity towards K<sub>Ca</sub>1.1.

and C-terminal peptide segments. This same approach may also facilitate investigation of toxin structure–activity relationships. Here an ‘alanine walk’ via ligation of various peptide block segments would mitigate repetitive low-yields and could aid in the generation of multi-point mutations within a single ligated toxin sequence.

## 5. Conclusion

Native chemical ligation offers a powerful synthetic tool that allows peptide elongation to go beyond the constraints of the standard length of 50–60 amino acids of automated linear SPPS (see Refs. [4,26]). This ligation scheme also opens new and varied opportunities in peptide toxin studies, focusing on smaller peptides 30–40 amino acids in length. These efforts will provide a greater understanding of pharmacological interactions and how these same peptides provides for an architectural foundation for drug design, as articulated by Menez [22].

## Acknowledgements

This work was supported by a Scientist Development Award from the American Heart Association (0530204N) to J.-P.B.; NIH Grant P01 NS42202 and a Grant-in-Aid from the American Heart Association (0150058N) to E.G.M, and Wellcome Trust Senior Biomedical Fellowship grant # 059862/Z/99/Z to D.R.E.

## References

- [1] Albericio F, Lloyd-Williams P, Giralt E. Convergent solid-phase peptide synthesis. *Methods Enzymol* 1997;289:313–36.
- [2] Bingham J-P, Bian S, Tan ZY, Takacs Z, Moczydlowski E. Synthesis of a biotin derivative of iberiotoxin: binding interactions with streptavidin and the BK Ca<sup>2+</sup>-activated K<sup>+</sup> channel expressed in a human cell line. *Bioconjug Chem* 2006;17:689–99.
- [3] Dawson PE, Muir TW, Clark-Lewis I, Kent SB. Synthesis of proteins by native chemical ligation. *Science* 1994;266:776–9.
- [4] Dawson PE, Kent SB. Synthesis of native proteins by chemical ligation. *Annu Rev Biochem* 2000;69:923–60.
- [5] Candia S, Garcia ML, Latorre R. Mode of action of iberiotoxin, a potent blocker of the large conductance Ca<sup>2+</sup>-activated K<sup>+</sup> channel. *Biophys J* 1992;63:583–90.
- [6] Clippingdale AB, Barrow CJ, Wade JD. Peptide thioester preparation by Fmoc solid phase peptide synthesis for use in native chemical ligation. *J Pept Sci* 2000;6:225–34.
- [7] Favre I, Sun YM, Moczydlowski E. Reconstitution of native and cloned channels into planar bilayers. *Methods Enzymol* 1999;294:287–304.
- [8] Frémont V, Blanc E, Crest M, Martin-Eauclaire M-F, Gola M, Darbon H, et al. Dipole moments of scorpion toxins direct the interaction towards small- or large-conductance Ca<sup>2+</sup>-activated K<sup>+</sup> channels. *Lett Pept Sci* 1997;4:305–12.
- [9] Galvez A, Gimenez-Gallego G, Reuben JP, Roy-Contancin L, Feigenbaum P, Kaczorowski GJ, et al. Purification and characterization of a unique, potent, peptidyl probe for the high conductance calcium-activated potassium channel from venom of the scorpion *Buthus tamulus*. *J Biol Chem* 1990;265:11083–90.
- [10] Gauthier MA, Klok HA. Peptide/protein-polymer conjugates: synthetic strategies and design concepts. *Chem Commun (Camb)* 2008;23:2591–611.
- [11] Giangiacomo KM, Garcia ML, McManus OB. Mechanism of iberiotoxin block of the large-conductance calcium-activated potassium channel from bovine aortic smooth muscle. *Biochemistry* 1992;31:6719–27.
- [12] Giangiacomo KM, Sugg EE, Garcia-Calvo M, Leonard RJ, McManus OB, Kaczorowski GJ, et al. Synthetic charybdotoxin-iberiotoxin chimeric peptides define toxin binding sites on calcium-activated and voltage-dependent potassium channels. *Biochemistry* 1993;32:2363–70.
- [13] Giangiacomo KM, Ceralde Y, Mullmann TJ. Molecular basis of alpha-KTx specificity. *Toxicon* 2004;43:877–86.
- [14] Hackeng TM, Mounier CM, Bon C, Dawson PE, Griffin JH, Kent SB. Total chemical synthesis of enzymatically active human type II secretory phospholipase A2. *Proc Natl Acad Sci U S A* 1997;94:7845–50.
- [15] Hackeng TM, Dawson PE, Kent SB, Griffin JH. Chemical synthesis of human protein S thrombin-sensitive module and first epidermal growth factor module. *Biopolymers* 1998;46:53–63.
- [16] Hackeng TM, Griffin JH, Dawson PE. Protein synthesis by native chemical ligation: expanded scope by using straightforward methodology. *Proc Natl Acad Sci U S A* 1999;96:10068–73.

- [17] Hackeng TM, Fernandez JA, Dawson PE, Kent SB, Griffin JH. Chemical synthesis and spontaneous folding of a multidomain protein: anticoagulant microprotein S. *Proc Natl Acad Sci U S A* 2000;97:14074–8.
- [18] Hamill OP, Marty A, Neher E, Sakmann B, Sigworth FJ. Improved patch-clamp techniques for high resolution recording from cells and cell-free membrane patches. *Pflügers Arch* 1981;391:85–100.
- [19] Johnson BA, Sugg EE. Determination of the three-dimensional structure of iberitoxin in solution by <sup>1</sup>H nuclear magnetic resonance spectroscopy. *Biochemistry* 1992;31:8151–9.
- [20] Kaczorowski GJ, Knaus HG, Leonard RJ, McManus OB, Garcia ML. High-conductance calcium-activated potassium channels; structure, pharmacology, and function. *J Bioenerg Biomembr* 1996;28:255–67.
- [21] Knaus H-G, McManus OB, Lee SH, Schmalhofer WA, Garcia-Calvo M, Helms LMH, et al. Tremorgenic indole alkaloids potently inhibit smooth muscle high-conductance calcium-activated potassium channels. *Biochemistry* 1994;33:5819–928.
- [22] Menez A. Functional architectures of animal toxins: a clue to drug design? *Toxicon* 1998;36:1557–72.
- [23] Milton RC, Milton SC, Kent SB. Total chemical synthesis of a D-enzyme: the enantiomers of HIV-1 protease show reciprocal chiral substrate specificity. *Science* 1992;256:1445–8.
- [24] Mouhat S, Jouirou B, Mosbah A, De Waard M, Sabatier J-M. Diversity of folds in animal toxins acting on ion channels. *Biochem J* 2004;378:717–26.
- [25] Muir TW. Semisynthesis of proteins by expressed protein ligation. *Annu Rev Biochem* 2003;72:249–89.
- [26] Muir TW, Dawson PE, Kent SB. Protein synthesis by chemical ligation of unprotected peptides in aqueous solution. *Methods Enzymol* 1997;289: 266–98.
- [27] Mullmann TJ, Spence KT, Schroeder NE, Frémont V, Christian EP, Giangiacomo KM. Insights into toxin specificity for K<sup>+</sup> channels revealed through mutations in noxiustoxin. *Biochemistry* 2001;40(37):10987–9.
- [28] Pragl B, Koschak A, Trieb M, Obermair G, Kaufmann WA, Gerster U, et al. Synthesis, characterization, and application of cy-dye- and alexa-dye-labeled hongotoxin(1) analogues. The first high affinity fluorescence probes for voltage-gated K<sup>+</sup> channels. *Bioconj Chem* 2002;13(3):416–25.
- [29] Rodríguez de La Vega RC, Possani LD. Current views on scorpion toxins specific for K<sup>+</sup>-channels. *Toxicon* 2004;43(8):865–75.
- [30] Sarin VK, Kent SB, Tam JP, Merrifield RB. Quantitative monitoring of solid-phase peptide synthesis by the ninhydrin reaction. *Anal Biochem* 1981;117: 147–57.
- [31] Schnölzer M, Alewood P, Jones A, Alewood D, Kent SB. In situ neutralization in Boc-chemistry solid phase peptide synthesis. Rapid, high yield assembly of difficult sequences. *Int J Pept Protein Res* 1992;40:180–93.
- [32] Shimony E, Sun T, Kolmakova-Partensky L, Miller C. Engineering a uniquely reactive thiol into a cysteine-rich peptide. *Protein Eng* 1994;7:503–7.
- [33] Stampe P, Kolmakova-Partensky L, Miller C. Mapping hydrophobic residues of the interaction surface of charybdotoxin. *Biophys J* 1992;62:8–9.
- [34] Stampe P, Kolmakova-Partensky L, Miller C. Intimations of K<sup>+</sup> channel structure from a complete functional map of the molecular surface of charybdotoxin. *Biochemistry* 1994;33:443–50.
- [35] Strobaek D, Christophersen P, Holm NR, Moldt P, Ahring PK, Johansen TE, et al. Modulation of the Ca<sup>2+</sup>-dependent K<sup>+</sup> channel, hsls, by the substituted diphenylurea NS1608, paxilline, and internal Ca<sup>2+</sup>. *Neuropharmacology* 1996;35: 903–14.
- [36] Tolbert TJ, Wong CH. Conjugation of glycopeptide thioesters to expressed protein fragments: semisynthesis of glycosylated interleukin-2. *Methods Mol Biol* 2004;283:255–66.
- [37] Toth GK, Pataricza J, Janaky T, Mak M, Zarandi M, Papp JG, et al. Synthesis of two peptide scorpion toxins and their use to investigate the aortic tissue regulation. *Peptides* 1995;16:1167–72.
- [38] Valiyaveetil FI, Sekedat M, MacKinnon R, Muir TW. Glycine as a D-amino acid surrogate in the K(+)-selectivity filter. *Proc Natl Acad Sci U S A* 2004;101(49):17045–9.
- [39] Valiyaveetil FI, MacKinnon R, Muir TW. Semisynthesis and folding of the potassium channel KcsA. *J Am Chem Soc* 2002;124(31):9113–20.
- [40] Yoshida M, Akaji K, Tatsumi T, Iinuma S, Fujiwara Y, Kimura T, et al. Synthesis of porcine brain natriuretic peptide-32 using silver tetrafluoroborate as a new deprotecting reagent of the S-trimethylacetamidomethyl group. *Chem Pharm Bull (Tokyo)* 1990;38:273–5.

## APPENDIX G. CONE SNAIL MILKED VENOM DYNAMICS – A QUANTITATIVE STUDY OF *CONUS PURPURASCENS*<sup>4</sup>

---

<sup>4</sup> Reproduced in part with permission from Chun, J. B. S.;\* Baker, M. R.;\* Kim, D. H.; LeRoy, M.; Toribo, P.; Bingham, J.-P. Cone snail milked venom dynamics – A quantitative study of *Conus purpurascens*. *Toxicon* **2012**, *60*, 83–94. \*Contributed equally to this work. Copyright © 2012, rights managed by Elsevier, Inc.



## Cone snail milked venom dynamics – A quantitative study of *Conus purpurascens*

Joycelyn B.S. Chun<sup>1</sup>, Margaret R. Baker<sup>1</sup>, Do H. Kim, Majdouline LeRoy, Priamo Toribo, Jon-Paul Bingham\*

Department of Molecular Biosciences and Bioengineering, College of Tropical Agriculture and Human Resources, University of Hawaii, Honolulu, HI 96822, USA

### ARTICLE INFO

#### Article history:

Received 30 November 2011

Received in revised form 10 March 2012

Accepted 22 March 2012

Available online 5 April 2012

#### Keywords:

Cone snail

Toxins

*Conus purpurascens*

Envenomation

Conotoxins

Conopeptides

Mass spectrometry

Milked venom

Peptide quantification

Amino acid analysis

Radula harpoon

### ABSTRACT

Milked venom from cone snails represent a novel biological resource with a proven track record for drug discovery. To strengthen this correlation, we undertook a chromatographic and mass spectrometric study of individual milked venoms from *Conus purpurascens*. Milked venoms demonstrate extensive peptide differentiation amongst individual specimens and during captivity. Individual snails were found to lack a consistent set of described conopeptides, but instead demonstrated the ability to change venom expression, composition and post-translational modification incorporation; all variations contribute to an increase in chemical diversity and prey targeting strategies. Quantitative amino acid analysis revealed that milked venom peptides are expressed at ranges up to 3.51–121.01  $\mu\text{M}$  within single milked venom samples. This provides for a 6.37–20,965 fold-excess of toxin to induce apparent  $\text{IC}_{50}$  for individual conopeptides identified in this study. Comparative molecular mass analysis of duct venom, milked venom and radula tooth extracts from single *C. purpurascens* specimens demonstrated a level of peptide continuity. Numerous highly abundant and unique conopeptides remain to be characterized. This study strengthens the notion that approaches in conopeptide drug lead discovery programs will potentially benefit from a greater understanding of the toxinological nature of the milked venoms of *Conus*.

© 2012 Elsevier Ltd. All rights reserved.

**Abbreviations:** aq., aqueous; aq. TFA, Aqueous Trifluoroacetic acid; Calc.  $\text{MH}^+$ , Calculated Monoisotopic molecular mass; CID, Collision induced dissociation; Da, Daltons; DHB, 2,5-dihydroxybenzoic acid; ESI-MS, Electrospray Ionization Mass Spectrometry; Hyp/O, 4-trans-hydroxyproline;  $\text{IC}_{50}$ , the half maximal inhibitory concentration; LC/MS, Liquid Chromatography interfaced Mass Spectrometry; MALDI-TOF/(TOF)-MS, Matrix Assisted Laser Desorption/Ionization Time of Flight/(Time of Flight) Mass Spectrometry; MeCN, Acetonitrile;  $m/z$ , mass to charge ratio;  $\text{Na}_v$ , Voltage-gates sodium channel; nAChR, nicotinic acetylcholine receptor; Obs.  $\text{MH}^+$ , Observed Monoisotopic molecular mass; PDA, Photo Diode Array; PTM, post-translational modification; Q-3, Quadrupole-3; RP-HPLC/UV, Reverse Phase High Performance Liquid Chromatography interface Ultraviolet detection; Rt, retention time; SEM, Scanning Electron-Microscopy; Std Dev., Standard Derivation; TCEP, Tris(2-carboxyethyl)phosphine;  $T_v$ , total volume.

\* Corresponding author. Fax: +1 808 965 3542.

E-mail address: [jbingham@hawaii.edu](mailto:jbingham@hawaii.edu) (J.-P. Bingham).

<sup>1</sup> Contributed equally to this work.

### 1. Introduction

Cone snails represent a highly valuable natural resource for drug-lead discovery programs. With an estimated >75,000 individual bioactive peptides within the genus of some +500 species (Olivera, 2002), the venom peptides from these carnivorous marine predators have provided numerous highly selective ligands that target a myriad of ion channels (as reviewed in Terlau and Olivera, 2004). Several of these targets have direct implications in human healthcare. As a result, cone snail peptides, commonly referred to as conotoxins or conopeptides, are attracting much attention from the medical research and pharmaceutical communities.

One peptide in particular,  $\omega$ -conotoxin MVIIA, also cited as SNX-111 or Ziconotide has received US Food and Drug

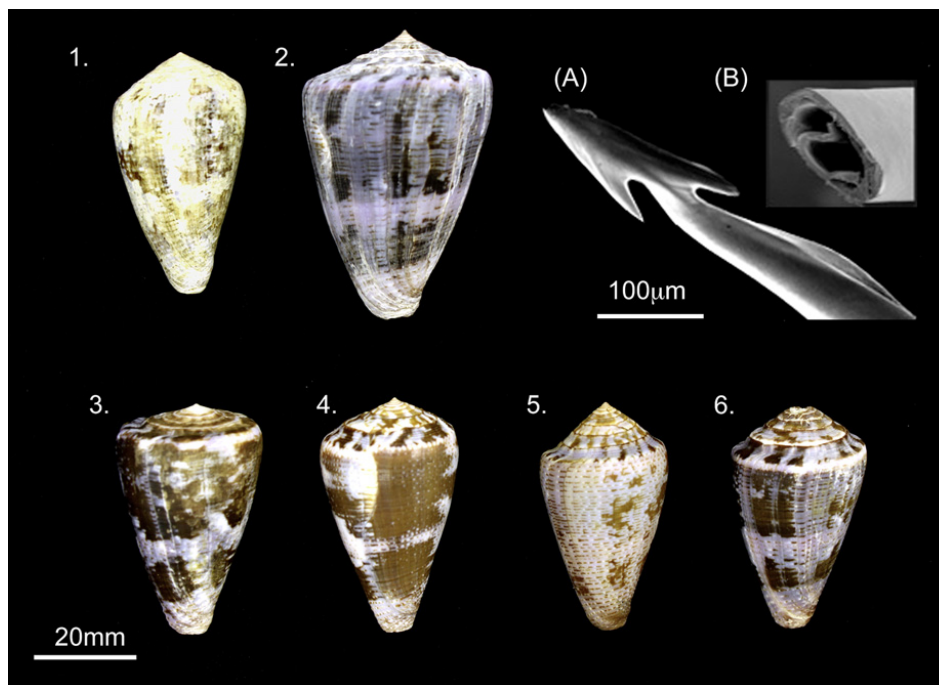
Administration approval as Prialt™ (Primary Alternative to morphine), a potent neuropathic pain analgesic, specifically targeting N-type voltage-gated calcium channels (Miljanich, 2004). Other conopeptides have now transitioned into clinical trials (see Han et al., 2008), providing a clear indication attesting to their pharmaceutical and therapeutic worth.

Unfortunately, such developments are often restricted by difficulties in accessing native cone snail venoms. This problem is compounded by concerns regarding depletion of the limited marine resource via bioprospecting (see Chivian et al., 2003; Duda et al., 2004). To offset these concerns, combined efforts encompassing genomic and proteomic approaches have both increased the diversity of conopeptide sequences and minimized animal usage (Gayler et al., 2005; Livett et al., 2006; Jakubowski et al., 2006). Sustainable research efforts have also been illustrated by utilization of pooled milked venoms as an alternative source of novel lead-compounds (see Shon et al., 1995; Hopkins et al., 1995). Direct evidence of their success is established by the milked venom-drug lead correlation, as illustrated with the expression of  $\omega$ -conotoxin MVIIA (Prialt™) in individual specimens of *Conus magus* (see Bingham, 1998; Bingham et al., 2010). Such observations necessitate the expansion of our limited biological understanding of the milked venoms of *Conus*. This in turn will maximize the drug-lead potential of milked venoms and provide a platform to ensure future biosustainable research supply.

*Conus purpurascens* (the purple cone; Fig. 1) represents a unique research candidate within the cone snail genus.

Being an Eastern Pacific piscivore (fishing eating species), its mammalian bioactive peptides have been well described – from their isolation and characterization (Shon et al., 1995, 1997; 1998a, 1998b; Jacobsen et al., 1998, 1999; Teichert et al., 2007; Gowd et al., 2008), pharmacological specificity (Shon et al., 1997; Jacobsen et al., 2000; Dowell et al., 2003; Teichert et al., 2007; Lopez-Vera et al., 2007; Gowd et al., 2008), to 3-Dimensional structural analysis (Savarin et al., 1998; Mitchell et al., 1998; Van Wagoner et al., 2003; Chi et al., 2005). What differentiates these studies from previous conopeptide research is the source of material; the majority of these conopeptides have been isolated directly from the pooled milked venom (see Table 1) and not from the crude dissected venom duct gland. *C. purpurascens* presently represents the species with the best-characterized milked venom of the genus. Using this established knowledge base in turn provides: (i) an avenue to further expand our general understanding of milked venom biochemistry, and; (ii) a solid foundation to examine alternative and novel approaches in conopeptide research.

To date, 15 peptides have been isolated, sequenced, synthesized and pharmacologically classified from *C. purpurascens* (see Table 1). The majority of these peptides have been found to selectively target sub-types of the acetylcholine receptor ( $\alpha$ -,  $\alpha A$ - and  $\psi$ -conotoxins; Lopez-Vera et al., 2007; Teichert et al., 2007; Van Wagoner et al., 2003); isoforms of the voltage-gated sodium channel ( $\mu$ - and  $\delta$ -conotoxins; Shon et al., 1998a and 1995); block potassium channels ( $\kappa$ -conotoxin; Savarin et al., 1998) and



**Fig. 1.** Representative specimens of *Conus purpurascens*: Captive specimens 1–2 (representatives used in this study) and locality specimens 3–6. 1. Panama, long-term captive specimen – notice chalky appearance, loss of shell color and gloss; 2. Panama, long-term captive specimen – notice growth ridges and bands, shell was originally 23.9 mm in diameter and grew to 37.9 mm - at the broadest point; 3. Horseshoe Bay, Costa Rica, non-captive specimen; 4. Cocos Is. Costa Rica, non-captive specimen; 5. Gobernadora Is. Panama, non-captive specimen; 6. Jaco, Panama, non-captive specimen. (A) Scan Electron Microscopy of *C. purpurascens* radula harpoon showing apex, 1st and 2nd barbs – each radula is unique to each species of *Conus*; Insert (B) shows cross section of harpoon providing access to lumen and peptide venom remnants within. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 1**  
*Conus purpurascens* conopeptides: Potency, pharmacological targeting and source location.

|               | Conopeptide  | Sequence                                   | IC <sub>50</sub>                 | Target                                           | Source            | Ref.                     |
|---------------|--------------|--------------------------------------------|----------------------------------|--------------------------------------------------|-------------------|--------------------------|
| Neuromuscular | αA-PIVA      | GCCGSYONAACHOCSCCKDROSYCGQ <sup>a</sup>    | 2.3 nM                           | adult mouse muscle nAChR                         | MV                | Hopkins et al., 1995     |
|               | α-PIA        | RDPCCSNPVCTVHNPOJC <sup>a</sup>            | 0.95 nM                          | rat nAChR subtype α6/α3β2β3                      | cDNA <sup>b</sup> | Dowell et al., 2003      |
|               | α-PIB        | ZSOGCCWNPACVKNRC <sup>a</sup>              | 36 nM                            | adult mouse muscle nAChR                         | MV                | Lopez-Vera et al., 2007  |
|               | μ-PIIIA      | ZRLCCGFOKSCRSRQCKOHRCC <sup>a</sup>        | 44 nM                            | mammalian Na <sub>v</sub>                        | cDNA <sup>b</sup> | Shon et al., 1998a       |
|               | κ-PVIIA      | CRIONQKCFQHLDDCCSRKCNRFNKCIV               | 57 ± 4 nM                        | K <sup>+</sup> shaker channel                    | MV                | Shon et al. 1998a,b      |
|               | ψ-PIIIIE     | HOOCCLYGKCRRYOGCSSASCCQR                   | 127 nM<br>14 μM                  | <i>torpedo</i> nAChR<br>adult mouse muscle nAChR | MV                | Shon et al., 1997        |
|               | ψ-PIIIF      | GOOCCLYGSCROFOGCYNALCCRK <sup>a</sup>      | 19 μM–1 mM                       | <i>torpedo</i> nAChR<br>adult mouse muscle nAChR | MV                | Van Wagoner et al., 2003 |
| Excitotoxic   | δ-PVIA       | EACYAOGTFCGIKOGLCCSEFCLPGVCFG <sup>a</sup> | Excitatory activity <sup>c</sup> | voltage-sensitive Na <sub>v</sub>                | MV                | Shon et al., 1995        |
|               | κA-PIVE      | DCCGVKLEMCHPCLCDNSCKNYGK <sup>a</sup>      | Excitatory activity <sup>c</sup> | –                                                | MV                | Teichert et al., 2007    |
|               | κA-PIVF      | DCCGVKLEMCHPCLCDNSCKKSGK <sup>a</sup>      | Excitatory activity <sup>c</sup> | –                                                | MV                | Teichert et al., 2007    |
| Unclassified  | D-Trp        | GCOWDPWC <sup>a</sup>                      | –                                | –                                                | DV                | Jacobsen et al., 1998    |
|               | Contryphan-P |                                            |                                  |                                                  |                   |                          |
|               | D-Leu        | GCVLLPWC                                   | –                                | –                                                | MV                | Jacobsen et al., 1998    |
|               | Contryphan-P |                                            |                                  |                                                  |                   |                          |
|               | PVA          | GCCPKQMRCCCTL                              | –                                | –                                                | MV                | Jacobsen et al., 1998    |
|               | P1.9         | CRWLQHSCLQ                                 | –                                | –                                                | cDNA <sup>b</sup> | –                        |
|               | Conantokin-P | GEγγHSKYQγCLRγIRVNKVQQγC                   | 0.3–2.3 nM                       | NMDA Receptor NR2A/NR2B                          | cDNA <sup>b</sup> | Gowd et al., 2008        |

<sup>a</sup> C-terminal amidated; W, D-Tryptophan; L, D-Leucine; Z, pyroglutamic acid; O, 4-*trans*-Hydroxyproline; γ, γ-carboxylglutamic acid; MV, Milked Venom; DV, Duct Venom.

<sup>b</sup> cDNA was prepared by reverse transcription of RNA isolated from the *C. purpurascens* venom duct.

<sup>c</sup> IC<sub>50</sub> on target channel is unknown.

antagonize the *N*-methyl-D-aspartate receptor (conantokins; Gowd et al., 2008). These peptides either inhibit neuromuscular transmission, causing flaccid paralysis, or increase excitability at the target ion channel (see Terlau et al., 1996). Interestingly enough, ω-conotoxins have yet to be isolated from *C. purpurascens* venom, further differentiating this species from most other piscivorous species studied to date; this provides the first indication to pharmacological diversification within the envenomation and prey immobilization strategies within *Conus*.

Toxinological characterization or quantitative biochemical analysis of individual *Conus* milked venoms is absent from current literature. Questions relating to venom production (Newcomb et al., 1995; Tayo et al., 2010), duct venom variability (Vianna Braga et al., 2005) and synergistic pharmacological strategies within *Conus* (Terlau et al., 1996) have arisen. Those regarding the presence of peptides in the venom organ/apparatus (see Biggs et al., 2008) and their correlation within the envenomation process are further compounded by comments surrounding the extent of venom peptide differentiation amongst individual specimens as well as the effects of captivity on this expression (Jakubowski et al., 2005; Dutertre et al., 2010).

In this work, we lay a foundation for addressing some of these questions by providing the first comprehensive quantitative analysis of milked venom peptides within *C. purpurascens*. Their assignment represents a novel development in the understanding of the envenomation strategies within *Conus*. This work provides the proof-of-concept

working with a well-established and studied *Conus* species. We commence with the examination of the expressive nature of milked venom and its differentiation in individual long-term captive animals via Reverse-Phase Liquid Chromatographic/Ultraviolet (RP-HPLC/UV) profiling. We then undertake a comparative molecular mass analysis, via Matrix Assisted Laser Desorption/Ionization Time of Flight Mass Spectrometry (MALDI-TOF-MS), to identify milked venom peptide constituents, specifically their presence within the lumen of individual radula harpoons, and their expression within whole crude dissected duct venom extracts. We conclude with determining the precise concentration of individual peptides delivered in a single envenomation. Comparison of expressed milked venom concentrations against previously reported IC<sub>50</sub> (half maximal inhibitory concentration) values indicates that high potency candidates can be found in low abundance peaks. Combined, these findings provide for a unique alternative route for identifying and then correlating compounds/molecular masses of potential pharmacological interest within *Conus*. These findings add to our basic toxinological understanding of the milked venoms of *Conus*.

## 2. Materials and methods

### 2.1. Materials

Chromatographic solvents used in this work were HPLC grade or higher – these being subjected to high vacuum

filtration to aid degassing prior to use. All buffers and solvents were filtered through 0.22  $\mu\text{m}$  Nylon filters (Millipore). Chemical reagents, including matrices, ion-pairing agents were supplied from manufacturers, as indicated – these being stored and used as per manufacturer's recommendations.

## 2.2. Cone snail housing, feeding, milking and venom extraction

Nine specimens of *C. purpurascens* were collected near the Smithsonian Tropical Research Institute, Panama. Specimens were transported to the USA, acclimated and housed in a temperature controlled environment in a single 30-gallon artificial saltwater tank fitted with a Fuval™ 402 biological filtration system. Specimens were fed weekly, using *Carassius auratus auratus* (goldfish; weight 2–5 g) and consequentially milked of venom using a modified method previously described in Hopkins et al. (1995) and Bingham et al. (2005). Individual milked venom volumes were collected, measured, then lyophilized. Specimens were later sexed upon dissection for duct venom and radula harpoons. Milked venoms, together with radula harpoons were stored at  $-20\text{ }^{\circ}\text{C}$  until required.

Secretory venom ducts were dried by Speed-Vac, then homogenized into a fine powder and weighed. Peptide extraction was achieved with 95% Solvent A (0.1% v/v TFA/aq.) and 5% Solvent B (90/10 v/v MeCN/0.08% v/v TFA/aq.), with samples typically representing 1 mg mL<sup>-1</sup>. Samples were vortexed and then sonicated for 10 min. Extracts were then centrifuged at 12,000g for 10 min. The resulting supernatant was removed, dried via Speed-Vac, weighed

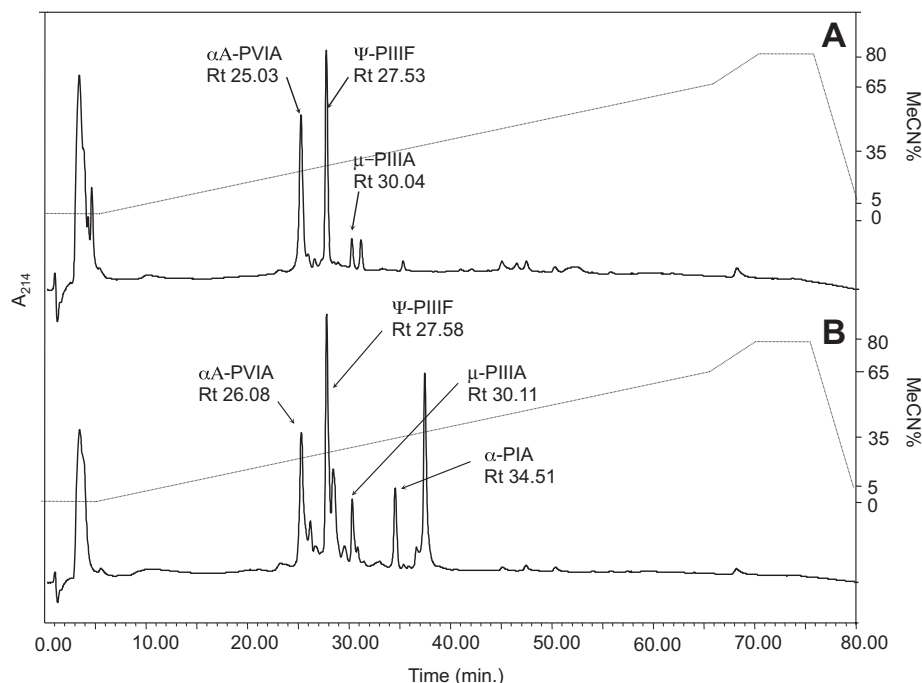
and then stored at  $-20\text{ }^{\circ}\text{C}$  until required. All extracts were resuspended in the above solvent at 1 mg mL<sup>-1</sup>, sonicated (5 min) and re-centrifuged (at 12,000g for 10 min) prior to chromatographic separation and MS analysis.

## 2.3. RP-HPLC/UV – peptide purification

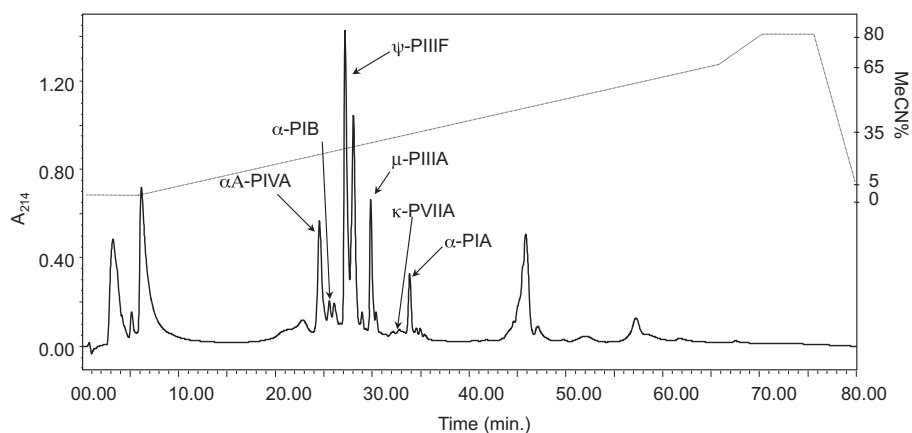
Milked venom peptides and duct venom extracts were separated and purified using C<sub>18</sub> Narrow-bored RP-HPLC/UV (Vydac; 5 mm, 300 Å, 2.1 × 250 mm) and later quantified using a capillary-bore RP-HPLC/UV (Phenomenex; 5 mm, 300 Å, 1.0 × 250 mm). A Waters 2695 Alliance HPLC System interfaced with a 996 Waters Photo Diode Array (PDA) Detector was used for automated sample analysis and detection. Data was acquired and analyzed using Waters Millennium<sup>32</sup> (v3.2) software. Samples were eluted using a standard linear 1% min<sup>-1</sup> gradient of Acetonitrile (HPLC grade, Fisher Scientific; MeCN; 90/10 MeCN/0.08% v/v aq. TFA; Solvent B) against 0.1% v/v aqueous Trifluoroacetic acid (Spectrophotometric grade, Sigma–Aldrich; aq. TFA; Solvent A) at a flow rate of 250  $\mu\text{L min}^{-1}$  (narrow-bore) or 100  $\mu\text{L min}^{-1}$  (capillary-bore) for a period of 65 min, as shown in Figs. 2 and 3. RP-HPLC column was pre-equilibrated with 5% solvent B, prior to sample injection. Elutant profiles were extracted at 214 nm. Samples for later amino acid quantification were fractionated manually, and subjected to repeated RP-HPLC/UV purification when necessary.

## 2.4. Conopeptide reduction

Speed-Vac dried RP-HPLC/UV purified peptides were resuspended in 100 mM Tris(2-carboxyethyl)phosphine



**Fig. 2.** Representative RP-HPLC/UV profile of time differentiated milked venoms obtained from a single *Conus purpurascens* specimen (Panama, specimen No. 2 – Fig. 1): Profiles A and B represent a time difference of approximately 8 months in captivity. Peaks labeled correspond to the individual conopeptides observed and their retentions (Rt), as listed in Table 2. Profiles demonstrate a level of chromatographic simplicity. Note:  $\psi$ -conotoxin PIIIF (Fig. 2B) is represented in two separate, closely eluting forms. Molecular mass analysis assigns these two peptides a PTM variants differing by a single *trans*-4-hydroxyproline modification.



**Fig. 3.** Maximal milked venom complexity observed a single captive specimen *Conus purpurascens* (Panama, specimen No. 1 – Fig. 1): RP-HPLC/UV showing the dominant peptide peaks being expressed in separate captive specimen of *C. purpurascens* milked venom, MV #41. Six *C. purpurascens* conopeptides were observed by retention and MS analysis (see Table 4). This specific profile is used to illustrate quantitative amino acid analysis, as in reference to Table 2.

(TCEP; Pierce Chemicals) with 25 mM  $\text{NH}_4\text{OAc}$ , pH 4.5. To aid reduction, samples were heated at 60 °C for 15 min. Upon reduction, samples were either re-purified using capillary-bore RP-HPLC/UV (as described above) or desalted by using Millipore Zip-Tip™, as described by the manufacturer, prior to mass spectrometric analysis.

### 2.5. Conopeptide quantification

Amino acid analyses of RP-HPLC/UV purified peptides were performed at the W.M. Keck Foundation Biotechnology Resource Laboratory at Yale University. Samples were hydrolyzed *in vacuo* with 6N HCl/0.2% v/v phenol for 16 h at 115 °C. Analysis was undertaken using a Hitachi L-8900 PH amino acid analyzer (ion-exchange separation, post-column derivatization with ninhydrin for detection at 570 nm and 440 nm). Data was collected using EZChrom Elite for Hitachi software and tabulated in a Microsoft Excel spreadsheet.

### 2.6. Mass spectrometry conopeptide analysis

#### 2.6.1. Electrospray ionization mass spectrometry (ESI-MS)

Speed-Vac dried RP-HPLC/UV purified peptides were resuspended in 0.1% v/v aqueous Formic acid (LC/MS grade, Sigma–Aldrich). Samples were delivered to the ionization source of an API 3000 Mass Spectrometer (Thornhill, Ontario, Canada) via a Redone 8125 Injector (20  $\mu\text{L}$  external loop; Rheodyne, Cotati, CA, USA) and infused with carrier solvent (50% MeCN/0.1% v/v aq. Formic acid; 50  $\mu\text{L min}^{-1}$ ) as provided by an ABI 140B Dual Syringe Pump. Full-scan single MS experiments were typically obtained by scanning quadrupole-3 (Q-3) from  $m/z$  400–2200 in 2–3 s with a scan step size of 0.1–0.5 Da. Data was acquired using Analyst Software (v.1.4.1) (Applied Biosystems/MDS Sciex, Thornhill, Ontario, Canada). ESI-MS system was calibrated manually in positive mode with PPG 3000/Mass Standards Kit (AB/MDS-Sciex), to achieve <5-ppm mass accuracy, as per manufacturer’s protocol.

Collision Induced Dissociation (CID) was performed by directly infusing the TCEP reduced RP-HPLC/UV purified peptide into the ESI-MS at 5  $\mu\text{L min}^{-1}$ , using the same

Formic acid/MeCN carrier solvent as mentioned above. Bombardment was confined to quadrupole-2 (Q-2) with a collision cell gas thickness of  $3 \times 10^{14}$  atoms/cm<sup>2</sup> and a collision energy (Q-0 to Q-2 rod offset voltage) typically set at ~20–40 eV. The resulting CID (daughter ion) spectra were obtained by scanning quadrupole-3 (Q-3) from  $m/z$  50–3000 in 5 s with a step size of 0.1 Da. MS/MS data analysis was assisted with the use of Mac BioSpec v1.01 (PE Sciex, Thornhill, Ontario, Canada).

#### 2.6.2. Matrix Assisted Laser Desorption/Ionization Time of Flight/(Time of flight) mass spectrometry [MALDI-TOF/(TOF)-MS]

RP-HPLC/UV purified peptides fractions in 0.1% v/v aq. TFA were mixed 1:1 with matrix solution (2,5-dihydroxybenzoic acid [DHB] in 1:1 0.1% v/v aq. TFA: MeCN; Ultra pure, Sigma–Aldrich) and spotted onto a thin layer of dried DHB matrix saturated in methanol on the target plate. The spots were dried under a stream of  $\text{N}_2$  gas. Parent ions were identified on the Ultraflex III MALDI-TOF/(TOF)-MS (Bruker Daltonics), controlled by the Compass 1.2 SR1 software package (Bruker Daltonics), in positive reflector mode. Peptide II Calibration Mix (Bruker Daltonics) was used for external calibration, with a <5-ppm mass accuracy. Analysis of the spectra was done using flexAnalysis v3.0 (Bruker Daltonics).

### 2.7. Mass spectrometry radula harpoon analysis – MALDI-TOF-MS

*C. purpurascens* radula harpoons were collected from either (i) spent radula harpoons during the milking process, or (ii) from dissected radula sacs. Radula harpoons were prepared for MALDI-TOF-MS analysis in the following manner: (a) externally  $\text{dH}_2\text{O}$  washed-air dried radula harpoons were repeatedly cut into sections in an eppendorf and extracted with 100  $\mu\text{L}$  extracting solution (1:1 MeCN:0.1% v/v aq. TFA), sonicated for 20 min, centrifuged at 12,000g. The resulting solvent was Millipore Zip-Tip™ (containing  $\text{C}_{18}$  reversed-phase media) desalted and the elute mixed 1:1 with MALDI-MS matrix solvent (40 mg mL<sup>-1</sup> DHB in 1:2 MeCN:0.1% v/v aq. TFA) directly on the MALDI-MS

target plate and allowed to air dry prior to analysis, or; (b) externally dH<sub>2</sub>O washed-air dried whole radula harpoons were cut in halves, taped onto the target plate via removable double-coated Scotch Tape® and then flooded with matrix solvent by dropping 1 µL at the luminal position of the mounted radula harpoon. Matrix solvent was allowed to dry with the assistance of a stream of N<sub>2</sub> prior to analysis.

### 2.8. Scanning Electron-Microscopy of radula harpoons

Imaging by Scanning Electron-Microscopy (SEM) was performed with a Hitachi S-4800 Field Emission Scanning Electron Microscope with Oxford INCA X-Act EDS System. Dissected radula harpoons were prepared for imaging by dip washing in 100% methanol, followed by sub sequential washes in dH<sub>2</sub>O. Dry samples were mounted on aluminum stubs with conductive silver paint. The samples were then sputter-coated with 10–12 nm gold/palladium with a Cressington 203 sputter coater. Imaging was performed at a beam accelerating voltage of 5 kV. Images were obtained at magnifications ranging from 10 to 100×; this was dependent on the feature to be tracked.

## 3. Results

### 3.1. Comments on snail husbandry

A number of observations were made during long-term *C. purpurascens* captivity: (i) Snails were able to consume fish equal to their own shell size and had ability to feed weekly; (ii) If snails were not fed regularly or did not discharge their pre-loaded radula harpoons from the proboscis (<2 times/month), they would singularly lose the ability to penetrate the receptacle membrane, resulting in loss of milked venom – recovery occurred naturally with ‘spent’ radula harpoon expulsion and then re-arming; (iii) Specimens seem to be sturdy and resilient to captive conditions, as illustrated by the ability to produce viable veliger upon mating – a situation that occurred twice during 48 months of captivity. No specimen mortalities were recorded during these experiments, until specimens were terminated for duct venom extracts.

### 3.2. RP-HPLC/UV milked venom profiling

Milked venoms from *C. purpurascens* ( $n = 100$ ) were measured to have volume ranges of 6–480 µL (Mean ± Std Dev.; 56.21 µL ± 105.21 µL). Chromatographically milked venom differentiation was observed both in the expression of individual conopeptides and in their peak area integration, the latter reflecting differences in concentrations – given that the peaks of interest demonstrate peptide homogeneity by LC/MS, this being in contrast to the crude duct venom extracts (not shown). The maximal difference in milked venom peptide expression within a single ~8 month time differentiated captive individual is illustrated in Fig. 2. This represented our most differentiated milked venom profile observed in this study. It was common to see milked venom transition to a more complex constituent containing profile – this including the appearance and disappearance of major profile contributing peaks.

Expressional conopeptide differentiation in some individual specimens was seen as a gradual process demonstrating minor changes on a week-to-week basis (see supplemental Fig. S1), this continued throughout the 48 months of captivity (see below). Ovipositing females demonstrated the absence of peptides within their milked venoms, although the motions of venom capture were successful (i.e. recorded milked venom volume, radula harpoon discharge and prey engulfment). These profiles later recovered to re-expressing conopeptides some weeks later. There were a number of other unexplained instances in which dramatic RP-HPLC/UV profile changes were noted – in most revolved around the single event expression of newly abundant peptide peaks. We also observed ‘blank’ milkings (~7% of all milkings) – in which ejaculate was collected with no apparent peptide profile being resolved chromatographically. These isolated events demonstrated no correlation to animal sex, feeding success, or season. They occurred between normal milkings that exhibited typical elution patterns. Their occurrence may represent a dynamic flux of expressional or secretional events that occur within the secretory venom duct itself.

The majority of captive milked venom peaks were observed to chromatographically elute between 20 and 40% MeCN. The presence of low to moderately abundant hydrophobic peaks (eluting >50% MeCN) were observed, these did not equate to the hydrophobic δ-conotoxins (see below). Near baseline resolution was apparent for the majority of individual milked venom constituents (see Fig. 2A), this could be further extended with TCEP reduction. Individual chromatographically resolvable peaks of interest, as examined in this study, demonstrated a high level of peptide homogeneity, with minimal co-elutes being observed. Peak homogeneity was assessed by: (i) UV spectrum extraction via PDA; (ii) ESI-MS infusion of RP-HPLC/UV fractions; (iii) Liquid Chromatography Interfaced Mass Spectrometry profiling (LC/MS); (iv) TCEP reduction of whole milked venoms and isolated fractions followed by RP-HPLC/UV analysis, and; (v) MALDI-TOF-MS of RP-HPLC/UV venom fractions (not shown).

The identified conopeptides previously isolated from *C. purpurascens*, their presence being assigned by ESI-MS/MALDI-TOF-MS molecular mass analysis and quantitative amino acid analysis – as reported in this study, demonstrated an RP-HPLC/UV elution pattern of αA-conotoxin PIVA, α-conotoxin PIB, ψ-conotoxin PIIIF, μ-conotoxin PIIIA, κ-conotoxin PVIIA and lastly α-conotoxin PIA (see Table 2, Fig. 3). Individual Post-translational (PTM) variant forms of parent conopeptide sequences were observed but limited. From this single population of *C. purpurascens*, as throughout this study, the hydrophobic δ-conotoxin PVIA (‘lock-jaw peptide’; Table 1) was absent, as determined by RP-HPLC/UV and by molecular mass analysis (see below). Its absence also extends to the reported PTM variants – which represent differential hydroxylated proline forms (see Shon et al., 1995).

### 3.3. Molecular mass comparison of crude duct extract, milked venom and radula harpoon lumen

Zip-Tip™ processed venom source extracts from duct venom, milked venom and radula harpoon lumen, obtained from single *C. purpurascens* specimens, in part contained

**Table 2**  
Identification and assigned concentrations of *Conus purpurascens* conopeptides observed within a single milked venom.

| Conopeptide | Retention Time (min.) | Conc. (pMoles) <sup>a</sup> | [MV peptide] (μM) <sup>b</sup> | IC <sub>50</sub> (nM) <sup>c</sup> | [MV peptide]/IC <sub>50</sub> <sup>d</sup> |
|-------------|-----------------------|-----------------------------|--------------------------------|------------------------------------|--------------------------------------------|
| αA-PIVA     | 24.61                 | 359                         | 48.22                          | 2.3                                | 20965.21                                   |
| α-PIB       | 26.26                 | 214                         | 28.74                          | 36                                 | 798.33                                     |
| ψ-PIIIF     | 27.46                 | 901                         | 121.01                         | 19000                              | 6.37                                       |
| μ-PIIIA     | 30.06                 | 450                         | 60.44                          | 44                                 | 1373.63                                    |
| κ-PVIA      | 32.77                 | 36                          | 4.84                           | 57                                 | 84.91                                      |
| α-PIA       | 33.86                 | 26.1                        | 3.51                           | 0.95                               | 3694.73                                    |

<sup>a</sup> As observed by initial RP-HPLC/UV integration with peptide standards – without taking into consideration dilution factor (50%), amount loaded (100/110 μL) and total volume of original milked venom (16.38 μL – specific sample, see Fig. 3).

<sup>b</sup> Corrected peptide concentration as observed in total milked venom sample.

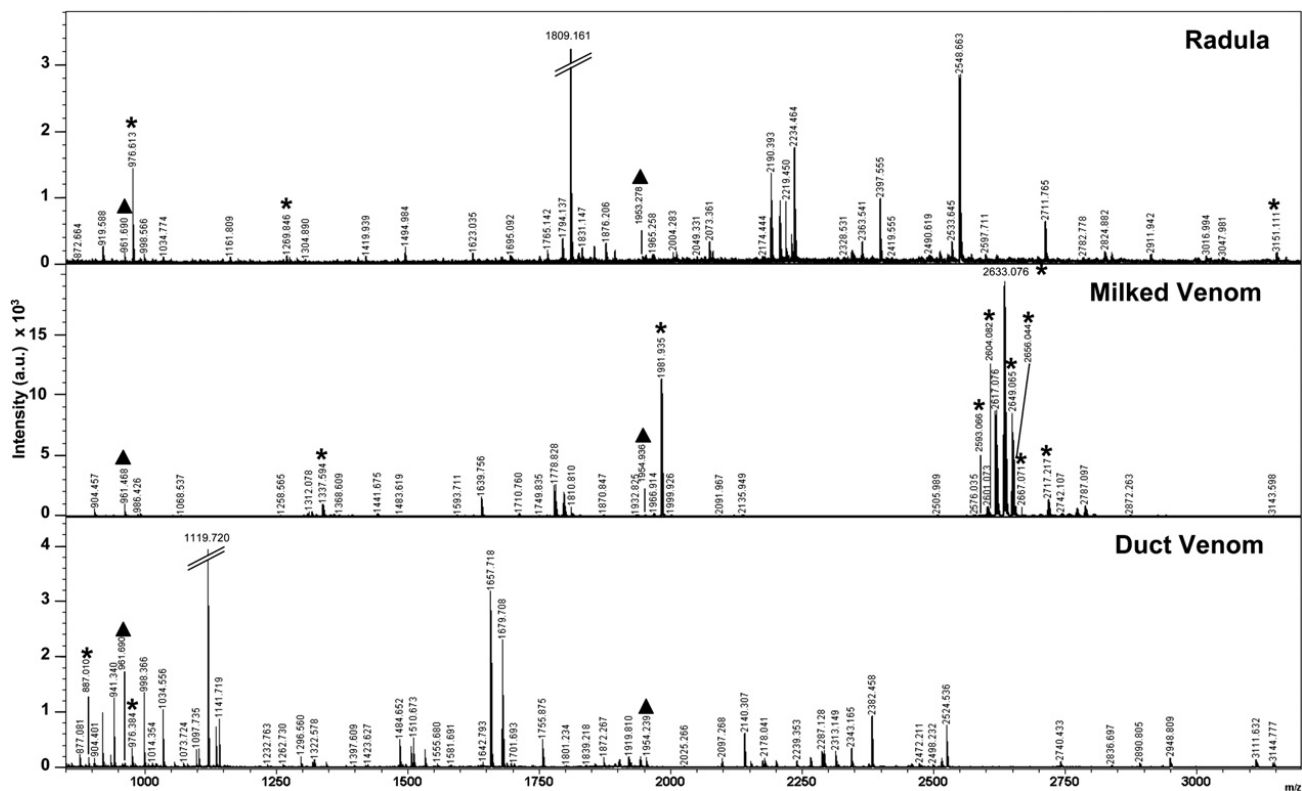
<sup>c</sup> As reported in literature, see Table 1.

<sup>d</sup> Calculated from milked venom peptide concentrations present<sup>(b)</sup> and reported IC<sub>50</sub>.

common *m/z* as observed by MALDI-TOF-MS (see Fig. 4). Table 3 summarizes the molecular masses observed for single *C. purpurascens* representative that typifies this level of molecular mass continuity in these extracts. The observed molecular mass continuity was restricted to 27 *m/z* in total, and encompassed either 2 or all 3 of the venom source extracts. These do not represent the only *m/z* observed within these venom source extracts, with numerous relative abundant *m/z* being unassigned (Fig. 4).

Examining Fig. 4 revealed that different venom source extracts demonstrated different clusters of molecular masses and possess peptides of differing relative intensities, as seen specifically with the milked venom source extract in the 2600–2750 *m/z* range. In the duct venom

source extract the most relative intense *m/z* was seen as 1119.72 Da, while in the milked venom source extract the dominant *m/z* was assigned as 2633.07 Da. This specific molecular mass corresponds to the previously reported PTM variant (Hyp deletion) of [Pro<sup>13</sup>]αA-conotoxin PIVA (see Hopkins et al., 1995). The parent sequence of αA-conotoxin PIVA is also observed, *m/z* 2649.07 Da (Calc. *m/z* 2649.94 Da). In this specific illustrated specimen, the dominant ion present in the radula harpoon source extract was observed as *m/z* 1809.16 Da. Of the all the observed *m/z*, two masses were shown to be common across all three venom source extracts, *m/z* 961.7 and 1954.3 Da, as indicated in Fig. 4 and Table 3. Examining common masses between individual venom source extracts we see: (i)



**Fig. 4.** MALDI-TOF mass spectra of radula harpoon (top), captive milked venom (middle), and crude duct venom (bottom) obtained from a single specimen of *Conus purpurascens*: Molecular Masses (*m/z*) common to all three are designated by a triangle (▲). An asterisk (★) indicates peaks that correspond to previously published conopeptides derived from *C. purpurascens*. 2,5-dihydroxybenzoic acid (DHB) was used as the matrix. One advantage of using DHB is that there is a low occurrence of matrix peaks and matrix adducts. To further ensure that common masses observed are not matrix peaks, mass spectra were compared to 'matrix only' mass spectra. No matrix peaks were detected.

**Table 3**

Molecular masses shared between radula, milked venom, and duct venom extracts derived from a single specimen of *Conus purpurascens*.

| Obs. <i>m/z</i>          | Radula Extract | Milked Venom  | Duct Venom    |
|--------------------------|----------------|---------------|---------------|
| –                        | –              | 904.5         | 904.4         |
| 919.6                    | –              | –             | 919.4         |
| –                        | –              | 925.4         | 925.3         |
| –                        | –              | 940.4         | 941.3         |
| <b>961.7<sup>a</sup></b> | –              | <b>961.5</b>  | <b>961.4</b>  |
| 976.6                    | –              | –             | 976.4         |
| 998.6                    | –              | –             | 998.4         |
| 1034.8                   | –              | –             | 1034.6        |
| 1262.9                   | –              | –             | 1262.7        |
| –                        | –              | 1344.6        | 1344.6        |
| 1495.0                   | –              | –             | 1495.7        |
| 1695.1                   | –              | –             | 1695.7        |
| 1750.1                   | –              | 1749.8        | –             |
| 1765.1                   | –              | 1764.8        | –             |
| 1794.1                   | –              | 1794.8        | –             |
| 1825.2                   | –              | 1825.8        | –             |
| –                        | –              | 1872.9        | 1872.3        |
| <b>1953.3</b>            | –              | <b>1954.9</b> | <b>1954.2</b> |
| 1965.3                   | –              | 1965.9        | –             |
| 1968.3                   | –              | 1968.9        | –             |
| 2004.3                   | –              | 2003.9        | –             |
| 2174.4                   | –              | –             | 2174.0        |
| 2343.5                   | –              | –             | 2343.2        |
| 2382.5                   | –              | –             | 2382.5        |
| 2490.6                   | –              | 2490.0        | –             |
| 2836.9                   | –              | –             | 2836.7        |
| 3330.3                   | –              | –             | 3330.8        |

<sup>a</sup> Peaks common to all three are in bold; Common peaks have a  $\Delta m/z = 0 \pm 1$  Da.

radula harpoon – milked venom source extract contains 10 common *m/z*; (ii) duct venom – milked venom source extract contains 7 common *m/z*, while; (iii) duct venom – radula harpoon source extract possesses the largest number with 14 common *m/z* (see Table 3).

Of the previously published/patented *C. purpurascens* conopeptides, 12 *m/z* were observed to correlate to known conopeptides as indicated in Fig. 4 and illustrated in Table 4. The majority of these conopeptides, corresponding to 8 molecular masses, were observed in the milked venom source – the original venom source for their isolation (see Table 1).

**Table 4**

Molecular masses identified in radula, milked venom, and duct venom samples that correspond to known conopeptides from *Conus purpurascens*.

| Conopeptide        | Calc. <i>m/z</i> | Obs. <i>m/z</i> |         |        |
|--------------------|------------------|-----------------|---------|--------|
|                    |                  | Radula          | MV      | DV     |
| D-Leu Contryphan-P | 888.4            | –               | –       | 887.01 |
| Contryphan-P       | 976.34           | 976.61          | –       | 976.38 |
| P1.9               | 1271.57          | 1269.85         | –       | –      |
| PVA                | 1336.52          | –               | 1337.59 | –      |
| $\alpha$ -PIA      | 1980.82          | –               | 1980.94 | –      |
| $\kappa$ A-PIVF    | 2594.05          | –               | 2593.07 | –      |
| $\mu$ -PIIIA       | 2604.11          | –               | 2604.08 | –      |
| $\alpha$ A-PIVA    | 2649.94          | –               | 2649.07 | –      |
| $\kappa$ A-PIVE    | 2656.03          | –               | 2656.04 | –      |
| $\psi$ -PIIIF      | 2665.06          | –               | 2667.07 | –      |
| $\psi$ -PIIIE      | 2717.06          | –               | 2717.22 | –      |
| Conantokin-P       | 3152.31          | 3151.11         | –       | –      |

### 3.4. Example of standardized procedure for quantitative analysis of milked venom peak concentrations

A detailed and standardized procedure for determining milked venom constituent concentrations is illustrated for future reference.  $\alpha$ A-conotoxin PIVA (Rt 24.61 min) from MV #41 (Fig. 3) was quantified via assigning its peak area integration and correlating this to a specific  $\alpha$ A-conotoxin PIVA concentration standard curve. The  $\alpha$ A-conotoxin PIVA standard was isolated and RP-HPLC/UV purified from pooled milked venom. This peak demonstrated: (i) molecular mass correlation to  $\alpha$ A-conotoxin PIVA via ESI-MS (Obs.  $MH^+$  2647.85; Calc.  $MH^+$  2647.93 Da; Obs. Reduced  $MH^+$  2653.79; Calc. Reduced  $MH^+$  2653.93 Da); (ii) partial sequence conformation via collision-induced dissociation (CID) analysis (not shown), and; (iii) the corresponding amino acid composition to  $\alpha$ A-conotoxin PIVA.

Quantitative amino acid hydrolysis was undertaken using a 1/3 of this sample. Remaining (100  $\mu$ L, 2/3) non-hydrolyzed material was assigned the calculated concentration – thus becoming the ' $\alpha$ A-conotoxin PIVA standard' stock. Five different concentrations of  $\alpha$ A-conotoxin PIVA (10, 50, 100, 500 and 1000 pMoles) resolved on a RP-HPLC capillary  $C_{18}$  column and peak areas integrated to graphically establish a standard curve to then assign its unknown concentrations, as observed in various milked venoms.  $\alpha$ A-Conotoxin PIVA milked venom profile identification was achieved by RP-HPLC/UV retention time analysis, using the same RP-HPLC capillary  $C_{18}$  column as above.

In our example, (Fig. 3 and Table 2) 359 pMoles of  $\alpha$ A-conotoxin PIVA was found to be present upon RP-HPLC/UV profile analysis of sample *C. purpurascens* MV #41. Sample concentration correction was undertaken to give the final amount of peptide produced in whole milked venom, as injected by the animal. This was achieved by using our established protocol: (i) collected milked venom is measured (16.38  $\mu$ L), (ii) lyophilized and, (iii) resuspended in 100  $\mu$ L of solvent A to make a standardized milked venom working stock. 50  $\mu$ L (50%) of this milked venom working solution was removed and further diluted with 60  $\mu$ L of solvent A, giving  $T_v = 110$   $\mu$ L. A total of 100  $\mu$ L of this sample was analyzed by RP-HPLC/UV, thus representing 90.9% of total RP-HPLC sample volume aliquot, or a physical representation of 45.45% of the whole milked venom sample. This is then calculated to represent an equivalent aliquot of 7.44  $\mu$ L whole milked venom (i.e. 45.45% of 16.38  $\mu$ L). Thus 359 pMole of  $\alpha$ A-conotoxin PIVA, as determined by standard curve, was equal to 7.44  $\mu$ L equivalents of the original milked venom volume. The total concentration of  $\alpha$ A-conotoxin PIVA present in sample MV #41 (Fig. 3) is calculated to be 48.22 pMoles  $\mu$ L<sup>-1</sup> = 48.22  $\mu$ M (359 pMoles/7.44  $\mu$ L). This same procedure was followed for RP-HPLC/UV purified native milked venom peptide standards of  $\alpha$ -conotoxin PIB,  $\psi$ -conotoxin PIIIF,  $\mu$ -conotoxin PIIIA,  $\alpha$ -conotoxin PIA and  $\kappa$ -conotoxin PVIIA; all demonstrated (i) correct molecular mass correlation in native and TCEP reduced state and (ii) expected amino acid composition. These allowed for the expressed concentration to be assigned to individual conopeptides within *C. purpurascens* (see Table 2).

### 3.5. Milked venom peptide concentrations and $IC_{50}$ values

Within a single milked venom from *C. purpurascens*, conopeptide concentrations are present within  $\mu\text{M}$  ranges, i.e. from 3.51  $\mu\text{M}$  ( $\alpha$ -conotoxin PIA) to 121.01  $\mu\text{M}$  ( $\psi$ -conotoxin PIIIF), as listed in Table 2. This specific milked venom profile exemplifies the greatest coverage of known individual conopeptides quantified in our present studies of *C. purpurascens* (see Fig. 3).

In this single milked venom sample, the ratios of individual conopeptide concentrations expressed with respect to their experimental  $IC_{50}$  values ranged from 6.37 to 20,965.21 (Table 2). This data potentially provides an indication to the observed pharmacological excess in venom production, as examined with current  $IC_{50}$  data, and strengthens the viability of milked venom as a source of potent bioactive constituents. Within the gathered data,  $\alpha$ A-conotoxin PIVA (48.22  $\mu\text{M}$ ) demonstrates the highest ratio value, partially attributed to its low  $IC_{50}$  value (2.3 nM), Table 2. This ratio is in contrast to the highly concentrated  $\psi$ -conotoxin PIIIF (121.01  $\mu\text{M}$ ), having a diminished ratio due to its relatively high  $IC_{50}$  value (19  $\mu\text{M}$ ). The conopeptide with the second lowest concentration in the milked venom, as an added example to strengthen this seemingly inverse relationship,  $\alpha$ -conotoxin PIA (3.51  $\mu\text{M}$ ), exhibited the second highest ratio as a result of its low  $IC_{50}$  value (0.95 nM; Table 2).

## 4. Discussion

The pharmaceutical potential of cone snail venoms is well recognized (as reviewed by Lewis, 2009). Venom complexity and demonstrated isoform selectivity in ion channel targeting makes these extracts a natural pharmacopeia for biological testing and drug lead discovery programs. However, this is a limited marine resource, to which few researchers have access. Novel approaches in conopeptide research, including venom procurement, are needed to maximize their full potential. Our approach has been to emphasize the integration of traditional biochemical and advanced mass spectrometric technologies, together focusing on biosustainable resource management – via animal husbandry and cone snail venom milking, in an endeavor to place these rare commodities in the hands of pharmacologists and neuroscientists alike.

We believe our findings provide a unique perspective into the basic toxinology of the cone snail milked venoms. As we have illustrated in this study, there are a number of important observations that have significant implications in the transition potential of venom peptides to drugs. The use of this information may impact how we view and use cone snail milked venoms successfully as a primary source of novel lead compounds.

### 4.1. Milked venoms demonstrate expressional variability

Venom peptide differentiation within *C. purpurascens* has been previously alluded to with the expression of the highly hydrophobic  $\delta$ -conotoxin PVIA (see Shon et al., 1995). However, the nature and extent of peptide variability within the time differentiated milked venom, as a whole, was basically unrealized. Our evidence indicates that

hydrophobic  $\delta$ -conotoxins are not the only conopeptides that demonstrate a high level of expressional regulation in the milked venom, and that peptide differentiation is compounded by captive conditions. Milked venom differentiation increases the number of bioactive peptides present; as seen in Figs. 2 and 3 individual specimens may express different peptide profiles. Pooling vast quantities of venoms may risk masking and/or diluting minor or unique expressional constituents. Here, simplicity may be the best approach, which may require treating individual snails or individual populations of snails as unique venom producers.

### 4.2. Post-translational modifications increase the chemical diversity of the milked venoms

In our study of *C. purpurascens*, we observe only a single proline variant of  $\alpha$ A-conotoxin PIVA, corresponding to the [Pro<sup>13</sup>] derivative. This peptide was observed as a major milked venom constituent (see Fig. 4), which may illustrate a conservation of PTM expressional resources under extended captive conditions, as suggested by Hopkins et al. (1995). Whether the production of this PTM variant represents a highly selective or random event remains to be determined. However, the illustrated differential PTM behavior adds an additional dimension to the chemical diversity achievable within a single milking. Furthermore, we believe that its expressional dominance, as seen by MALDI-TOF-MS (Fig. 4), is not without toxinological purpose, as the occurrence of differential hydroxylation has been observed in other *Conus* venoms, as illustrated by Franco et al. (2006) and Rivera-Ortiz et al. (2011).

PTM conopeptide variants potentially offer a combinatorial-like approach to maximize phyla-selective pharmacological efficacy via increase functional group complexity, without the need to change the genetic sequence of the parent toxin sequence. PTM variants can possess differential pharmacological selectivity and targeting (Hopkins et al., 1995; Chi et al., 2003; Nevin et al., 2007). Within the milked venoms of *C. purpurascens* differential hydroxylation of proline moieties has the potential to affect 7 different classes of conopeptides (Table 1). Yet to date only 2 pharmacological classes, the  $\delta$ - and  $\alpha$ A-conotoxins, have been affected by this natural occurrence (Shon et al., 1995; Hopkins et al., 1995). Unlike previous reports, our observations of PTM variants within *C. purpurascens* are limited. This occurrence potentially reflects the small and localized population of snails examined in this study. This may infer that the ability to undertake such modifications relates to unique and possibly localized environmental pressures, similar observations have been observed in the milked venom of *Conus ermineus* (see Rivera-Ortiz et al., 2011).

### 4.3. A number of different pharmacological strategies are represented within the milked venoms derived from *C. purpurascens*

Not all published *C. purpurascens* conopeptides are observed in a single milked venom profile- initially or over its entire time in captivity. Similar observations of differential peptide expression are observed in *Conus striatus* (Jakubowski et al., 2005), *Conus consors* (Dutertre

et al., 2010) and *C. ermineus* (Rivera-Ortiz et al., 2011). To ensure predatory success, *C. purpurascens* must possess a minimal pharmacological strategy. In this study, we see this with the combination of  $\alpha$ A-conotoxin PIVA,  $\psi$ -conotoxin PIIIF and  $\mu$ -conotoxin PIIIA (see Fig. 2A and Table 1). These toxins target a combination of nAChR isoforms (Shon et al., 1997; Van Wagoner et al., 2003) and  $\text{Na}_v$  ion channels (Shon et al., 1998a).  $\alpha$ A-Conotoxin PIVA and  $\psi$ -conotoxin PIIIF provide for a potent and independent targeting of the acetylcholine mediated synaptic transmission system (Hopkins et al., 1995; Shon et al., 1997). This is then compounded with  $\text{Na}_v$  targeting by  $\mu$ -conotoxin PIIIA, ensuring inhibition of any post-synaptic action potential propagation. This illustrates the most basic neuromuscular ‘motor cabal’ as defined by Terlau and Olivera (2004). Long-term captive milked venoms show expression of additional peptides in also varying concentration levels, as phenomena illustrated by the co-expression of  $\alpha$ -conotoxin-PIA (Fig. 2B) and  $\alpha$ -conotoxin-PIB (Fig. S1). Their expression synergistically complements the differential AChR targeting strategy illustrated above.

The expression of excitotoxic peptides  $\kappa$ A-conotoxins PIVE and PIVF ( $m/z$  2656.04 and 2593.07 Da, respectively), as seen with MALDI-TOF-MS, Fig. 4 and Table 4, conforms to the proposed synergistic ‘lighting strike cabal’ or excitotoxic shock inducement as detailed by Terlau and Olivera (2004). While the expression of excitotoxic  $\kappa$ A-conotoxins provides an effective route for rapid prey immobilization (Terlau et al., 1996), members of the aforementioned neuromuscular ‘motor cabal’ dominate via concentration (see Table 5). However, throughout this study the absence of  $\delta$ -conotoxin PVIA, a member of the excitotoxic peptide family, and its PTM variants, is noted and we find that this is not a result of peptide handling or storage.

Combined, these observations indicate that individual *C. purpurascens* specimens can use different pharmacological strategies to ensure successful prey capture and immobilization. What represents the quintessential pharmacological envenomation strategy for *C. purpurascens* remains to be determined. This is complicated by many uncharacterized milked venom peptides, including dominant peptide peaks, which may contribute to the completion of this pharmacological strategy.

#### 4.4. Milked venoms represent a highly refined source of bioactive constituents

Crude duct venom extracts contain cellular debris that can complicate any aspect of pharmacological assessment –

whether using a competitive binding assay or electrophysiological evaluation. Milked venoms offer a simple and direct route to those constituents of biological intent and resulting pharmacological interest. The observed chromatographic resolution provides a major incentive for its use (see Figs. 2 and 3), as well as the peptide concentrations achievable from a single milking – namely milked venom peptide production being within the  $\mu\text{M}$  range for this specific species (see Table 5). However, the presence of seawater within the expelled venom may effect pharmacological assessment, an issue corrected by chromatographic desalting. Alternatively, venom sample dilution may minimize any potential changes in solution osmolarity during the venom-assay application process (Huang and Bingham, unpublished results). Thus the direct use of the milked venoms then provides for a biosustainable resource of bioactive peptides that are highly amenable to pharmacological testing such as seen with whole cell patch clamping, and later to chromatographic isolation and MS characterization.

In this study, we illustrate the extent of molecular mass continuity between the whole duct venom extract and milked venom, and the detected differences in conopeptides that are expressed (Fig. 4). We equate these differences in part to the previously observed constituent changes seen with traversing the secretory duct (see Tayo et al., 2010). The combination of these duct sections masks the true complexity of the duct venom extract. These occurrences may be attributed to incomplete and/or differential PTM processing, which would result in diluting the level of peptide continuity observed in the milked venom. However, we would expect that closer comparative inspection of a small duct section that terminates into the pharynx of the cone snail would demonstrate a stronger mass correlation to the milked venom upon dissection, this being due to its anatomical proximity – to be illustrated and discussed elsewhere.

#### 4.5. Dominant RP-HPLC/UV milked venom profile peaks do not necessarily correlate to the most pharmacologically potent constituents

Attention is drawn to those conopeptides of low expressional abundance in the individual milked venoms, as illustrated with  $\alpha$ -conotoxin PIA (Fig. 2B; Table 2). This conopeptide has the lowest  $\text{IC}_{50}$  of any of the *C. purpurascens* peptides reported. We observed the opposite with the highly abundant RP-HPLC/UV peaks, as illustrated with  $\psi$ -conotoxin PIIIF, which is reported to have an  $\text{IC}_{50}$  of 19  $\mu\text{M}$  or  $\sim 1$  mM, depending on which nAChR isoform is

**Table 5**

Conopeptide concentration assignment within time captive milked venoms of *Conus purpurascens*.

|                       | $\alpha$ A-PIVA    | $\alpha$ -PIA | $\alpha$ -PIB | $\mu$ -PIIIA | $\kappa$ -PVIIA | $\psi$ -PIIIF |                                      |
|-----------------------|--------------------|---------------|---------------|--------------|-----------------|---------------|--------------------------------------|
| $\text{IC}_{50}$ (nM) | 2.3                | 0.95          | 36            | 44           | 57              | 19,000        |                                      |
| Milking No.           | Concentration (nM) |               |               |              |                 |               | $T_v$ ( $\mu\text{L}$ ) <sup>a</sup> |
| 14                    | 6348               | 2146          | 3242          | 12,006       | 795             | 17,243        | 46.8                                 |
| 34                    | 33,550             | 26,400        | –             | 21,175       | –               | 57,200        | 8                                    |
| 36                    | 178,200            | –             | –             | 35,200       | –               | 134,200       | 5                                    |
| 40                    | 138,600            | –             | 39,600        | 323,400      | 13,200          | 305,800       | 3                                    |
| 46                    | 215,600            | 360,800       | 105,600       | 308,000      | 66,000          | 567,600       | 3                                    |

<sup>a</sup> Total milked venom volume.

being evaluated (Table 1). Concentrating on the abundant peaks within the duct and milked venom extracts has been a common trend in conopeptide research.

The relative abundance of uncharacterized low intensity molecular masses within the milked venom indicates that *C. purpurascens* still has much to offer investigators. These observations indicate that each individual peptide needs equal pharmacological consideration. This process is potentially complicated with specific target isoform or phyla specificity. This is illustrated with  $\psi$ -conotoxin PIIIE having IC<sub>50</sub> of 127 nM for *Torpedo* AChR subtype, compared to an IC<sub>50</sub> of 14  $\mu$ M for adult mouse muscle nAChR (see Table 1; Shon et al., 1997). Many such examples of this pharmacological phenomenon exist in *Conus*. Combining this with low peptide abundance may provide a reliance on synthetic production for their analysis.

#### 4.6. Mass spectrometric surveying demonstrates peptide abundance in radula harpoon extracts

A novel idea being introduced in this study is the correlation of molecular masses that are present in the radula harpoon extracts with those found in duct and milk venoms (Fig. 4 and Table 3). The majority of observed radula harpoon masses are likely a result of radula sac specific expression of conopeptides. Evidence for radula sac specific expression of conopeptide-like mRNA was previously characterized (see Biggs et al., 2008). While the molecular mass continuity observed potentially represents an accumulated 'back-wash' or leakage into the radula sac itself, as milked venom expulsion is undertaken with great velocity and pressure (Schultz et al., 2004).

Defining these molecular mass correlations may have a number of interesting research implications. As we observed there are unidentified molecular masses present in the radula harpoon (see Fig. 4 and Table 3), which represents a new highly refined source of conopeptides of potential biological interest. As not all cone snails can be milked successfully, this established correlation between radula harpoon extracts and milked venom may assist in streamlining efforts to identifying lead-compound candidates. These assigned molecular masses could be isolated from crude duct venoms, if no milked venom source were available, or correlated to sequence information derived from cDNA-derived libraries. The relative abundance of these same radula harpoon molecular masses could allow for direct sequence analysis by MALDI-TOF/TOF-MS. The MS analysis of the radula harpoon represents another dimension to conopeptide discovery.

#### 4.7. The number of bioactive peptides within *Conus* may exceed present estimations

Use of mass spectrometric technology, providing molecular mass correlation within *Conus*, opens an alternative perspective into the venom complexity and the adopted chemical strategies to enhance pharmacological diversity. All these factors indicate the extent of biodiversity observable within *Conus* – an area that remains poorly addressed toxinologically. We believe that our initial mass spectrometric observations within the milked and duct

venoms, as well as radula harpoon extracts (see Fig. 4), together with the combined ability to differentially express venom constituents (see Fig. 2), illustrates that the present 75,000–100,000 individual conopeptide constituents may represent a conservative estimate to the true abundance of biologically active peptides within these gastropods. This same statement is echoed by Davis et al. (2009), Biass et al. (2009) and Rivera-Ortiz et al. (2011), noting that these authors also utilize a mass spectrometric approach in their investigations. Along similar lines, additional analysis of other cone snail species, as we have illustrated here, will add to the debate regarding the actual physical size of the native conopeptide library.

## 5. Concluding remarks and outlook

Enhancement of natural resource management, increased knowledge about venom production and secretion provides for a focused approach in maximizing the venom potential of *Conus*. Understanding these natural trends enhances our ability to find novel lead compounds of potential therapeutic use – based on their biochemical and predatory use. Without investigative investments into the basic toxinological understanding of *Conus*, we are potentially missing important biological aspects of their venom, as it is these facets that differentiate cone snails from other venom producers.

## Acknowledgments

We are indebted to Dr. Tom Duda for the collection of live *C. purpurascens* specimens. The authors would like to acknowledge the financial support of the American Heart Association (Scientist Development Award 0530204N to J-P.B.), University of Hawaii Sea Grant College Program (J-P.B.) and USDA TSTAR (# 2009-34135-20067 (gs2)) & HATCH (HAW00595-R (gs3)) (J-P.B.) & NIH (2G12RR003061-26 (gs4)) (UH JABSOM Proteomics Core Facility).

## Appendix. Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.toxicon.2012.03.019.

## Ethical statement

The author and co-authors of this paper have acted ethically in conducting the described research, having undertaken careful analysis of data and the submitted manuscript to avoid errors.

## Conflict of interest

Authors state that there is no conflict of interest.

## References

- Biass, D., Dutertre, S., et al., 2009. Comparative proteomic study of the venom of the piscivorous cone snail *Conus consors*. *J. Proteomics* 72 (2), 210–218.

- Biggs, J.S., Olivera, B.M., et al., 2008. Alpha-conopeptides specifically expressed in the salivary gland of *Conus pulicarius*. *Toxicon* 52 (1), 101–105.
- Bingham, J.P., Broxton, N.M., Livett, B.G., Down, J.G., Jones, A., Moczydlowski, E.G., 2005 Mar 1. Optimizing the connectivity in disulfide-rich peptides: alpha-conotoxin SII as a case study. *Anal. Biochem.* 338 (1), 48–61.
- Bingham, J.P., Mitsunaga, E., et al., 2010. Drugs from slugs—past, present and future perspectives of omega-conotoxin research. *Chem. Biol. Interact.* 183 (1), 1–18.
- Bingham, J.P., 1998. Novel Toxins from the Genus *Conus* – From Taxonomy to toxins. PhD Dissertation. Australia, University of Queensland.
- Chi, S.W., Park, K.H., et al., 2003. Solution conformation of alpha-A-conotoxin EIVA, a potent neuromuscular nicotinic acetylcholine receptor antagonist from *Conus ermineus*. *J. Biol. Chem.* 278 (43), 42208–42213.
- Chi, S.W., Lee, S.H., et al., 2005. Solution structure of alpha-conotoxin PIA, a novel antagonist of alpha6 subunit containing nicotinic acetylcholine receptors. *Biochem. Biophys. Res. Commun.* 338 (4), 1990–1997.
- Chivian, E., Roberts, C.M., et al., 2003. The threat to cone snails. *Science* 302 (5644), 391.
- Davis, J., Jones, A., et al., 2009. Remarkable inter- and intra-species complexity of conotoxins revealed by LC/MS. *Peptides* 30 (7), 1222–1227.
- Dowell, C., Olivera, B.M., et al., 2003. Alpha-conotoxin PIA is selective for alpha6 subunit-containing nicotinic acetylcholine receptors. *J. Neurosci.* 23 (24), 8445–8452.
- Duda Jr., T.F., Bingham, J.P., et al., 2004. How much at risk are cone snails? *Science* 303 (5660), 955–957 Author reply 955–957.
- Dutertre, S., Biass, D., et al., 2010. Dramatic intraspecific variations within the injected venom of *Conus consors*: an unsuspected contribution to venom diversity. *Toxicon* 55 (8), 1453–1462.
- Franco, A., Pisarewicz, K., et al., 2006. Hyperhydroxylation: a new strategy for neuronal targeting by venomous marine molluscs. *Prog. Mol. Subcell Biol.* 43, 83–103.
- Gayler, K., Sandall, D., et al., 2005. Molecular prospecting for drugs from the sea. Isolating therapeutic peptides and proteins from cone snail venom. *IEEE Eng. Med. Biol. Mag.* 24 (2), 79–84.
- Gowd, K.H., Twede, V., et al., 2008. Conantokin-P, an unusual conantokin with a long disulfide loop. *Toxicon* 52 (2), 203–213.
- Han, T.S., Teichert, R.W., et al., 2008. *Conus* venoms – a rich source of peptide-based therapeutics. *Curr. Pharm. Des.* 14 (24), 2462–2479.
- Hopkins, C., Grilley, M., et al., 1995. A new family of *Conus* peptides targeted to the nicotinic acetylcholine receptor. *J. Biol. Chem.* 270 (38), 22361–22367.
- Jacobsen, R., Jimenez, E.C., et al., 1998. The contryphans, a D-tryptophan-containing family of *Conus* peptides: interconversion between conformers. *J. Pept. Res.* 51 (3), 173–179.
- Jacobsen, R.B., Jimenez, E.C., et al., 1999. A novel D-leucine-containing *Conus* peptide: diverse conformational dynamics in the contryphan family. *J. Pept. Res.* 54 (2), 93–99.
- Jacobsen, R.B., Koch, E.D., et al., 2000. Single amino acid substitutions in kappa-conotoxin PVIIA disrupt interaction with the shaker K<sup>+</sup> channel. *J. Biol. Chem.* 275 (32), 24639–24644.
- Jakubowski, J.A., Kelley, W.P., et al., 2005. Intraspecific variation of venom injected by fish-hunting *Conus* snails. *J. Exp. Biol.* 208 (Pt 15), 2873–2883.
- Jakubowski, J.A., Kelley, W.P., et al., 2006. Screening for post-translational modifications in conotoxins using liquid chromatography/mass spectrometry: an important component of conotoxin discovery. *Toxicon* 47 (6), 688–699.
- Lewis, R.J., 2009. Conotoxin venom peptide therapeutics. *Adv. Exp. Med. Biol.* 655, 44–48.
- Livett, B.G., Sandall, D.W., et al., 2006. Therapeutic applications of conotoxins that target the neuronal nicotinic acetylcholine receptor. *Toxicon* 48 (7), 810–829.
- Lopez-Vera, E., Jacobsen, R.B., et al., 2007. A novel alpha conotoxin (alpha-PIB) isolated from *C. purpurascens* is selective for skeletal muscle nicotinic acetylcholine receptors. *Toxicon* 49 (8), 1193–1199.
- Miljanich, G.P., 2004. Ziconotide: neuronal calcium channel blocker for treating severe chronic pain. *Curr. Med. Chem.* 11 (23), 3029–3040.
- Mitchell, S.S., Shon, K.J., et al., 1998. Three-dimensional solution structure of conotoxin psi-PIIE, an acetylcholine gated ion channel antagonist. *Biochemistry* 37 (5), 1215–1220.
- Nevin, S.T., Clark, R.J., et al., 2007. Are alpha9alpha10 nicotinic acetylcholine receptors a pain target for alpha-conotoxins? *Mol. Pharmacol.* 72 (6), 1406–1410.
- Newcomb, R., Gaur, S., et al., 1995. Structural and biosynthetic properties of peptides in cone snail venoms. *Peptides* 16 (6), 1007–1017.
- Olivera, B.M., 2002. *Conus* venom peptides: reflections from the biology of clades and species. *Annu. Rev. Ecol. Syst.* 33, 25–47.
- Rivera-Ortiz, J.A., Cano, H., et al., 2011. Intraspecific variability and conopeptide profiling of the injected venom of *Conus ermineus*. *Peptides* 32 (2), 306–316.
- Savarin, P., Guenneugues, M., et al., 1998. Three-dimensional structure of kappa-conotoxin PVIIA, a novel potassium channel-blocking toxin from cone snails. *Biochemistry* 37 (16), 5407–5416.
- Schultz, J.R., Norton, A.G., et al., 2004. The projectile tooth of a fish-hunting cone snail: *Conus catus* injects venom into fish prey using a high-speed ballistic mechanism. *Biol. Bull.* 207 (2), 77–79.
- Shon, K.J., Grilley, M.M., et al., 1995. Purification, characterization, synthesis, and cloning of the lockjaw peptide from *Conus purpurascens* venom. *Biochemistry* 34 (15), 4913–4918.
- Shon, K.J., Grilley, M., et al., 1997. A noncompetitive peptide inhibitor of the nicotinic acetylcholine receptor from *Conus purpurascens* venom. *Biochemistry* 36 (31), 9581–9587.
- Shon, K.J., Olivera, B.M., et al., 1998a. mu-Conotoxin PIIIA, a new peptide for discriminating among tetrodotoxin-sensitive Na channel subtypes. *J. Neurosci.* 18 (12), 4473–4481.
- Shon, K.J., Stocker, M., et al., 1998b. kappa-Conotoxin PVIIA is a peptide inhibiting the shaker K<sup>+</sup> channel. *J. Biol. Chem.* 273 (1), 33–38.
- Tayo, L.L., Lu, B., et al., 2010. Proteomic analysis provides insights on venom processing in *Conus textile*. *J. Proteome Res.* 9 (5), 2292–2301.
- Teichert, R.W., Jacobsen, R., et al., 2007. Discovery and characterization of the short kappaA-conotoxins: a novel subfamily of excitatory conotoxins. *Toxicon* 49 (3), 318–328.
- Terlau, H., Olivera, B.M., 2004. *Conus* venoms: a rich source of novel ion channel-targeted peptides. *Physiol. Rev.* 84 (1), 41–68.
- Terlau, H., Shon, K.J., et al., 1996. Strategy for rapid immobilization of prey by a fish-hunting marine snail. *Nature* 381 (6578), 148–151.
- Van Wagoner, R.M., Jacobsen, R.B., et al., 2003. Characterization and three-dimensional structure determination of psi-conotoxin PIIIF, a novel noncompetitive antagonist of nicotinic acetylcholine receptors. *Biochemistry* 42 (21), 6353–6362.
- Vianna Braga, M.C., Konno, K., et al., 2005. Mass spectrometric and high performance liquid chromatography profiling of the venom of the Brazilian vermivorous mollusk *Conus regius*: feeding behavior and identification of one novel conotoxin. *Toxicon* 45 (1), 113–122.

## **APPENDIX H. ANALYSIS OF A CONE SNAIL'S KILLER COCKTAIL – THE MILKED VENOM OF *CONUS GEOGRAPHUS*<sup>5</sup>**

---

<sup>5</sup> Reproduced in part with permission from Bingham, J.-P.; Baker, M. R.; Chun, J. B. Analysis of a cone snail's killer cocktail – The milked venom of *Conus geographus*. *Toxicon* **2012**, *60*, 1166–1170. Copyright © 2012, rights managed by Elsevier, Inc.



Contents lists available at SciVerse ScienceDirect

# Toxicon

journal homepage: [www.elsevier.com/locate/toxicon](http://www.elsevier.com/locate/toxicon)

Short communication

## Analysis of a cone snail's killer cocktail – The milked venom of *Conus geographus*<sup>☆</sup>

Jon-Paul Bingham\*, Margaret R. Baker, Joycelyn B. Chun

Department of Molecular Biosciences and Bioengineering, College of Tropical Agriculture and Human Resources, University of Hawaii, Honolulu, HI 96822, USA

### ARTICLE INFO

#### Article history:

Received 11 May 2012

Received in revised form 19 July 2012

Accepted 25 July 2012

Available online 4 August 2012

#### Keywords:

Cone snail

Toxinology

Fatalities

*Conus geographus*

Envenomation

Conotoxins

Conopeptides

Mass spectrometry

Milked venom

### ABSTRACT

“Snails can kill” is a statement that receives much disbelief. Yet the venom from *Conus geographus*, as delivered by a disposable hypodermic-like needle, has indeed killed many unsuspecting human victims. Our understanding of their milked venom the essence of these fatalities, is in itself non-existent. Here, we present the molecular mass analysis of the milked venom of *C. geographus*, providing the first insight into the composition of its deadly cocktail.

© 2012 Elsevier Ltd. All rights reserved.

Twenty-five peptide sequences have been derived from the secretory venom duct of *Conus geographus* (Table 1; Fig. 1 – insert). This represents a culmination over some ~80 years of biochemical, genetic and pharmacological research.<sup>1</sup> Some of these bioactive peptides, commonly called conotoxins or conopeptides, have led to the

pharmacological re-classification of ion channels, based on work exploiting isoform selectivity and phyla differentiation characteristics (see Terlau and Olivera, 2004; Table 1). Yet, toxinologically what we know about the composition of these injected venoms remains mostly a mystery, particularly in this species which is known to be lethal to humans.

Potent biological activity has been correlated to the individually isolated secretory duct venom (DV) conopeptides; these have offered some insight into their deadly nature. But there remain a number of compelling issues: Do the complex crude dissected DV extracts represent what the snail uses in prey capture? Are all known DV conopeptides observed within a single milked venom (MV) collection? What makes a ‘killer’ cone snail lethal to humans? Here we address some of these questions using the MV of a known lethal cone snail, *C. geographus* – this species being responsible for at least 18 human fatalities (see Yoshida, 1984).

Using a similar method as described by Hopkins et al. (1995), we obtained MV by allowing *C. geographus*

**Abbreviations:** Acetonitrile, CH<sub>3</sub>CN; DHB, 2, 5-Dihydroxybenzoic acid; DV, Duct venom; MALDI-TOF MS, Matrix Assisted Laser Desorption/Ionization-Time of Flight Mass Spectrometry; *m/z*, Mass to charge ratio; MV, Milked venom; Obs. *m/z*, Observed Mass to charge ratio; RP-HPLC, Reverse Phase-High Performance Liquid Chromatography; TFA aq., Trifluoroacetic acid/aqueous; UV, Ultra-violet detection.

<sup>☆</sup> This paper is dedicated to a mentor, friend and fellow shell collector, Associate Professor Bruce G. Livett, formerly of the Department of Biochemistry & Molecular Biology at the University of Melbourne, Australia, in celebration of his retirement and contribution to the field of conopeptide research.

\* Corresponding author. Fax: +1 808 965 3542.

E-mail address: [jbingham@hawaii.edu](mailto:jbingham@hawaii.edu) (J.-P. Bingham).

<sup>1</sup> Much of this research has been reported in *Toxicon* over the last 50 years.

**Table 1**

The conopeptide sequences derived from *C. geographus* – pharmacological targeting, affinity, origin and expression within the milked venom.

| Name                               | Amino acid sequence                          | Target                                | Affinity IC <sub>50</sub> [nM]        | Original source | Monoisotopic mass (Da) | Observed <i>m/z</i> in MV | RP-HPLC Rt (min) <sup>a</sup> | Ref. <sup>b</sup> |
|------------------------------------|----------------------------------------------|---------------------------------------|---------------------------------------|-----------------|------------------------|---------------------------|-------------------------------|-------------------|
| Lys-conopressin G                  | CFIRNCPKG*                                   | Vasopressin Receptor                  | N.D.                                  | DV              | 1033.49                | 1034.48                   | N.D.                          | [1]               |
| α-GII                              | ECCHPACGKHFSC*                               | nAChR <sup>b</sup>                    | N.D.                                  | DV              | 1415.5                 | 1416.43                   | 32.3 <sup>d</sup>             | [2]               |
| α-GI                               | ECCNPACGRHYSC*                               | nAChR (mouse)                         | (LD <sub>50</sub> 12 μg/kg; mouse) 20 | DV              | 1436.48                | 1437.41                   | 32.3                          | [2]               |
| G5.1                               | QGW <sup>W</sup> CKENIACCV                   | Site 1 (αδ)                           | 1.3 ± 0.3                             | cDNA            | 1451.54                | –                         | –                             | [3]               |
| α-GIC                              | GCCSHPACAGNNQHIC*                            | N.AChR (human)                        | N.D.                                  | cDNA            | 1608.58                | –                         | –                             | [4]               |
| G1.1                               | ECCNPACGRHYSCKG                              | hα3β2                                 | 1.1                                   | cDNA            | 1622.58                | 1622.55                   | N.D.                          | [3]               |
| α-GIA                              | ECCNPACGRHYSCGK                              | hα3β4                                 | 775                                   | DV              | 1622.58                | 1622.55                   | 31.0                          | [2]               |
| Contulakin-G                       | ZSEEGGSNATKKPYIL                             | hα4β2                                 | 309                                   | DV              | 2068.97                | 2068.11                   | N.D.                          | [5]               |
|                                    |                                              | Neurotensin Receptor                  |                                       |                 |                        |                           |                               |                   |
|                                    |                                              | NTR2 (human)                          | 960                                   |                 |                        |                           |                               |                   |
|                                    |                                              | NTR3 (mouse)                          | 250                                   |                 |                        |                           |                               |                   |
| α-GID                              | IRDγCCSNPACRVNNOHVC                          | nAChR rα7(rat)                        | 4.5                                   | DV              | 2185.86                | –                         | –                             | [6]               |
|                                    |                                              | hα3β2                                 | 3.1                                   |                 |                        |                           |                               |                   |
| Conantokin-G                       | GEγγLQγNQγLIRγKSN*                           | hα4β2                                 | 152                                   | DV              | 2262.94                | –                         | –                             | [7, 8]            |
|                                    |                                              | NMDA Receptor                         |                                       |                 |                        |                           |                               |                   |
|                                    |                                              | NR2B Brain (human)                    | 480 21–69                             |                 |                        |                           |                               |                   |
| μ-GIIC                             | RDCCTOOKKCKDRRCKOLKCCA*                      | Na <sub>v</sub> <sup>c</sup>          | N.D.                                  | DV              | 2595.2                 | 2595.02                   | 26.4 <sup>d</sup>             | [9, 10]           |
| μ-GIIIA                            | RDCCTOOKKCKDRQCKOQRCCA*                      | Na <sub>v</sub> 1,4 Brain (rat)       | 69.2 ± 0.8                            | DV              | 2610.14                | 2610.08                   | 26.4                          | [10, 11, 12]      |
|                                    |                                              | Brain (chicken)                       | >1000                                 |                 |                        |                           |                               |                   |
|                                    |                                              | Skeletal muscle (rat)                 | 0.97 ± 0.17                           |                 |                        |                           |                               |                   |
|                                    |                                              | Electroplax (eel)                     | 3.48 ± 0.09                           |                 |                        |                           |                               |                   |
| μ-GIIIB                            | RDCCTOORCKCKDRRCKOMKCCA*                     | Na <sub>v</sub> 1,4 (human)           | 1065                                  | DV              | 2641.16                | –                         | –                             | [10, 13, 14]      |
|                                    |                                              | Na <sub>v</sub> 1,4 (rat)             | 49                                    |                 |                        |                           |                               |                   |
|                                    |                                              | Na <sub>v</sub> 1,4 (eel)             | 1.1 ± 0.1                             |                 |                        |                           |                               |                   |
| ω-GVIC                             | CKSOGSSCSOTSYNCCRSCNOYTKRC                   | Na <sub>v</sub> <sup>b</sup>          | N.D.                                  | DV              | 2875.99                | –                         | –                             | [10]              |
| ω-GVIA                             | CKSOGSSCSOTSYNCCRSCNOYTKRCY*                 | Ca <sub>v</sub> 2.2 (chick brain)     | 0.15                                  | DV              | 3038.17                | 3038.09                   | 32.8                          | [10, 15]          |
|                                    |                                              | Ca <sub>v</sub> 2.2 (mouse brain)     | 0.07                                  |                 |                        |                           |                               |                   |
| ω-GVIB                             | CKSOGSSCSOTSYNCCRSCNOYTKRCY                  | Ca <sub>v</sub> <sup>b</sup>          | N.D.                                  | DV              | 3096.13                | 3096.02                   | 33.5                          | [10]              |
| ω-GVII                             | CKSOGTOCSRGMRDCTCSLSYSNKCRRY                 | Ca <sub>v</sub> <sup>b</sup>          | N.D.                                  | DV              | 3289.34                | –                         | –                             | [10]              |
| Scratcher peptide                  | KFLSGGFKγIVCHRYCAKGIKAEFCNC <sup>W</sup> PD* | N.D.                                  | N.D.                                  | DV              | 3301.53                | –                         | –                             | [15]              |
| ω-GVIIA                            | CKSOGTOCSRGMRDCTCSLLYSNKCRRY                 | Ca <sub>v</sub> 2.2                   | 3.7                                   | DV              | 3315.39                | 3315.25                   | 36.8                          | [10, 16]          |
| G6.1                               | DDECEPPGDFCGFFKIGPPCCSGWCFLWCA               | N.D.                                  | N.D.                                  | cDNA            | 3322.24                | 3322.13                   | N.D.                          | [3]               |
| Conotoxin-GS                       | ACSGRGSRCOOQC <sup>W</sup> CMGLRCGRGNPQKC    | Na <sub>v</sub> Skeletal muscle (rat) | 880                                   | DV              | 3617.5                 | –                         | –                             | [17]              |
|                                    | IGAHγDV                                      |                                       |                                       |                 |                        |                           |                               |                   |
| Sequence 401 from Patent EP1852440 | SDGGNAAAKESDVIALTVWKCCTIPSC YEKKKIKACVF      | N.D.                                  | N.D.                                  | cDNA            | 4071.99                | –                         | –                             | [18]              |
| Sequence 323 from Patent EP1852440 | SDGRNAAANDQASDLMAATVRGCC AVPSCLRLNPDLCGGGR   | N.D.                                  | N.D.                                  | cDNA            | 4143.86                | –                         | –                             | [18]              |

(continued on next page)

**Table 1** (continued)

| Name                               | Amino acid sequence                            | Target            | Affinity IC <sub>50</sub> [nM] | Original source | Monoisotopic mass (Da) | Observed m/z in MV | RP-HPLC Rt (min) <sup>a</sup> | Ref. <sup>b</sup> |
|------------------------------------|------------------------------------------------|-------------------|--------------------------------|-----------------|------------------------|--------------------|-------------------------------|-------------------|
| σ-GVIIIa                           | GCTRTCGGOKCTGTCTCTNSSKCGC<br>RYNVHPSGWGCGCACs* | 5-HT <sub>3</sub> | 53 ± 3                         | DV              | 4189.43                | –                  | –                             | [19]              |
| Sequence 321 from Patent EP1852440 | SDGRNDAAKAFDLISSTVKKGCCSHPAC<br>AGNNQHICGRRR   | N.D.              | N.D.                           | cDNA            | 4238.98                | –                  | –                             | [18]              |

\* = C-terminal amidation; Z = pyroglutamic acid; S = O-linked glycosylated serine; w = D-tryptophan; W = bromotryptophan; O = 4-trans-hydroxyproline; T = glycosylated threonine; γ = gamma-carboxy glutamic acid; Y = sulfotyrosine.

<sup>a</sup> RP-HPLC retentions are in reference to Fig. 1A.

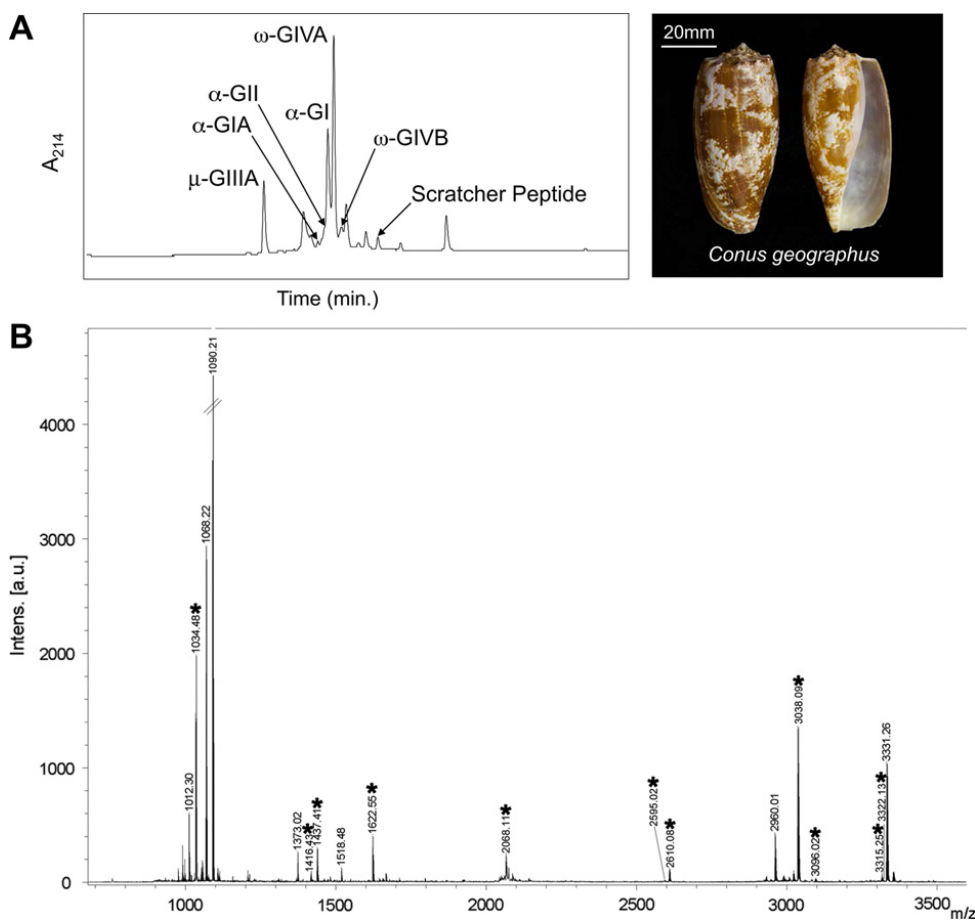
<sup>b</sup> Corresponding references listed in Supplemental Materials.

<sup>c</sup> Targeting assigned based on sequence homology; N.D., Not Determined, DV, Duct Venom; MV, Milked Venom; cDNA, complementary DNA; nAChR, nicotinic Acetylcholine Receptor; NMDA, N-Methyl-D-aspartate; Cav, Voltage gated calcium channels; Nav, Voltage gated sodium channels; 5-HT<sub>3</sub>, Serotonin 5-HT<sub>3</sub>-receptor.

<sup>d</sup> Observed as shoulder in primary peak.

specimens to impale a condom-covered receptacle. The stimulated snails inject venom, under pressure and with velocity, directly into the vial at ~5–20 μL per milking ( $n = 20$ ). Interestingly, not all *C. geographus* specimens demonstrated this proboscis extension-venomation behavior. We collected specimens ( $n = 4$ ) repeatedly from

one specific site, Boulton Reef (23°45'32"S 152°16'22"E) in the Capricorn and Bunker Group of Australia's Great Barrier Reef, which solely demonstrated this unusual predatory behavior. As we have observed, most *C. geographus* prey envenomations occur within the safety of the rostrum/mouth upon full prey engulfment ( $n = 22$  specimens; 5 locations). The



**Fig. 1.** (A) RP-HPLC/UV profile of the milked venom from *C. geographus* (Insert: *C. geographus* specimen from Boulton Reef, GBR, Australia). (B) The MALDI-TOF-MS analysis of the milked venom from *C. geographus*. Illustrated is the unexpected molecular mass simplicity of milked venom. Labeled peaks (\*) correspond to known conopeptides listed in Table 1. Method: MV was RP-HPLC/UV profiled using a C<sub>18</sub> Phenomenex capillary column (5 μm, 300 Å, 1.0 × 250 mm, flow 100 μL min<sup>-1</sup>) eluted with a linear 1% min<sup>-1</sup> gradient of organic [90/10% v/v Acetonitrile (CH<sub>3</sub>CN)/0.08% v/v TFA aq.] against 0.1% v/v TFA aq. for 80 min, as delivered by a Waters 2695 Alliance HPLC system. Elutant was monitored by photodiode array UV detection from 200 to 300 nm and extracted at 214 nm (A). For MALDI-TOF MS MV analysis, ZipTip® desalted MV was mixed 1:1 with matrix solution (2, 5-dihydroxybenzoic acid [DHB] in 1:1 0.1% v/v TFA aq.; CH<sub>3</sub>CN) and spotted onto dried matrix saturated in methanol on the MTP 384 polished target plate, and dried under N<sub>2</sub>. Parent ions were identified on the Ultraflex III MALDI-TOF-MS (Bruker Daltonics), controlled by the Compass 1.2 SR1 software package in positive reflector mode (B). Peptide II Calibration Mix (Bruker Daltonics) was used for external calibration, with a <5-ppm mass accuracy. Spectra analysis was undertaken using FlexAnalysis v3.0 (Bruker Daltonics).

physical environment at Boulton Reef may influence *C. geographus*' predatory behavior. Specimens are only found in a narrow corridor of fused cavernous coral substrate under large dead plate coral, this provides a Labyrinth for small fish. Full expansion of the rostrum to 'net' fish would be spatially difficult, making proboscis extension more feasible for predatory success. Observations in captivity show partially opened rostrums during hunting, with a brilliant red proboscis cautiously deployed mid-air a few centimeters ahead, and not as an extended probing appendage as seen in other species. The snail actively pursues the fish; contact between prey and predator is 'calculated' targeting the lateral side away from the head and gills; the proboscis is immediately withdrawn upon envenomation leaving the imbedded radula without tethering. The subdued fish, displays a dulled response to physical stimulus and shows labored gill movement, within a few minutes the fish loses the ability to maintain an upright posture. Once on its side, the cone snail moves in with rostrum fully expanded engulfing the fish headfirst, without issue. Milking of *C. geographus* then becomes a simple intervention once proboscis is visibly extended. Use of whole fish as a milking stimulus increases rate of success.

A representative Reverse Phase-High Performance Liquid Chromatography/Ultra-violet detection (RP-HPLC/UV) profile of non-captive MV (milked within <24 h after field collection) is shown in Fig. 1A. This multi-peak profile, which contains ~12 distinguishable peaks, far contrasts the complexity observed in the crude DV extract (see Olivera et al., 1990; Bingham et al., 1996). A number of *C. geographus* peptides were isolated and Edman sequenced (not shown), confirming their sequence identity, post-translational amino acid content and expression – these include, in order of RP-HPLC elution:  $\mu$ -conotoxin GIIIA (26.4 min),  $\alpha$ -conotoxin GI (32.3 min), and  $\omega$ -conotoxin GVIA (32.8 min). Matrix Assisted Laser Desorption/ionization-Time of Flight Mass Spectrometry (MALDI-TOF MS) analysis (positive mode) of the representative single MV sample of *C. geographus* provided the assignment of 18 molecular masses, of which 12 correlated to known conopeptide sequences (Table 1; Fig. 1B). This was the highest number of known *C. geographus* conopeptides seen in a single MV from this population. In this example the observed MV mass range encompasses 990–3400, Fig. 1B. The peptide with the highest relative intensity was  $m/z$  1090.21 – an uncharacterized MV constituent. In a number of instances we observed +1  $m/z$  with some of the known conopeptides, as illustrated in Table 1. This potentially represents C-terminal processing, the difference being between the amidated and carboxylated form. This has been a somewhat unexpected observation in this and other *Conus* MVs (Bingham unpublished results).

Examining the MV molecular mass profile, the dominance of the  $\alpha$ -conotoxin family becomes apparent, with 4 individuals identified ( $\alpha$ -conotoxins GI, GII, GIA/GI.1; see Table 1). This parallels the common observation of  $\alpha$ -neurotoxins in the MV of snakes (Phui Yee et al., 2004). Pharmacologically in *C. geographus* this is compounded by the co-expression of  $\mu$ -conotoxins GIIIA, GIIC (obs.  $m/z$  2610.08, and 2595.02 respectively) and  $\omega$ -conotoxins GVIA and GVIB (obs.  $m/z$  3038.09, 3096.02 respectively), which

completes a predictable peptide toxin 'motor cabal' as proposed by Olivera and Cruz (2001). This *C. geographus* MV 'cabal' targets the functionality of pre- and post-synaptic ion channel targets which include: (i) the acetylcholine receptor ( $\alpha$ -conotoxins); (ii) the voltage gated sodium channel (muscle type;  $\mu$ -conotoxins), and; (iii) voltage gated calcium channels (N-type;  $\omega$ -conotoxins). Here their synergistic pharmacological actions would lead to a rapid and persistent flaccid paralysis – a common observation in both native prey capture (Olivera, 1997) and human envenomations (see Flecker, 1936; Rice and Halstead, 1968).

The expression of Lys-conopressin G (calc.  $m/z$  1033.49; obs.  $m/z$  1034.48 – indication of differential C-terminal processing, see above) in the MV raises additional speculation to its biological/pharmacological intention in prey. Suggestions of minimizing prey escape response have been proposed (Dutertre et al., 2008). This has merit as typically *C. geographus* specimens take to a net-casting-like prey capture response using their rostrum, with the seemingly oblivious fish unaware of the danger it faces while being engulfed alive – a possible indication of prey sedation from material 'leaking' from the venom apparatus (see Johnson and Stablum, 1971).

But what of the other conopeptides derived from *C. geographus*? Or the unidentified  $m/z$  observed, as indicated in Fig. 1B? This brings a new level of intrigue and possible explanation of why different outcomes in human envenomation, lethal vs. non-lethal, are reported for this specific species (Cruz and White, 1995). Furthermore, the illustrated MV RP-HPLC (Fig. 1A) may not represent a lethal human dose, as materials recovered are induced by predatory response, are small in relative concentration, and derived from smaller than normal *C. geographus* specimens – which are typical of Boulton Reef. Further investigation is required to see if a 'defensive milked venom profile' can be achieved. Our previous observations indicate *Conus*' ability to produce 'dry-milking's' and undergo MV differentiation (Chun et al., 2012), which indicates an ability to differentiate and/or control venom secretion.

The presence of additional highly abundant peptides that illustrate unknown or unassigned compounds, i.e. ~40% of the observed MV  $m/z$ , provides an indication that even within this well-studied cone snail, *C. geographus* contains many uncharacterized venom constituents, specifically those observed with a  $m/z$  <1100, similar observations are seen in MV from *Conus purpurascens* (Chun et al., 2012). The primary  $m/z$  data provided here may assist in future genomic endeavors using this species and aid in peptide sequence characterization. While with the advancement of low-level peptide detection methods, as too the use of their MVs, as seen here with mass spectrometric incorporation, brings further validation to decades of previous works and now provides direct toxicological insight regarding these lethal predators.

## Ethical statement

The author and co-authors of this paper have acted ethically in conducting the described research, having undertaken careful analysis of data and the submitted manuscript to avoid errors.

## Acknowledgements

The authors would like to thank Mr. David Bingham for supplying milked venom, Mr. Jeffrey Milisen for photographic assistance, and the financial support of USDA TSTAR (# 2009-34135-20067) (J.-P.B.) and grants from the National Center for Research Resources (5 G12 RR003061-26) and the National Institute on Minority Health and Health Disparities (8 G12 MD007601-26) from the National Institutes of Health.

## Conflict of interest

Authors state that there is no conflict of interest.

## Appendix A. Supplementary material

Supplementary material associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.toxicon.2012.07.014>.

## References

- Bingham, J., Jones, A., et al., 1996. *Conus* venom peptides (Conopeptides): inter-species, intra-species and within individual variation revealed by ionspray mass spectrometry. In: Lazarovici, P., Spira, M.E., Zlotkin, E. (Eds.), *Biochemical Aspects of Marine Pharmacology*. Alaken Inc., Fort Collins, Colorado, USA, pp. 13–27.
- Chun, J.B., Baker, M.R., et al., 2012. Cone snail milked venom dynamics – a quantitative study of *Conus purpurascens*. *Toxicon* 60 (1), 83–94.
- Cruz, L.J., White, J., 1995. Clinical toxicology of *Conus* snail stings. In: Meier, J., White, J. (Eds.), *Clinical Toxicology of Animal Venoms*. CRC Press, Boca Raton, FL, pp. 117–127.
- Dutertre, S., Croker, D., et al., 2008. Conopressin-T from *Conus tulipa* reveals an antagonist switch in vasopressin-like peptides. *J. Biol. Chem.* 283 (11), 7100–7108.
- Flecker, H., 1936. Cone shell poisoning, with report of a fatal case. *Med. J. of Aust.* 1, 464–466.
- Hopkins, C., Grilley, M., et al., 1995. A new family of *Conus* peptides targeted to the nicotinic acetylcholine receptor. *J. Biol. Chem.* 270 (38), 22361–22367.
- Johnson, C.R., Stablum, W., 1971. Observations on the feeding behavior of *Conus geographus* (Gastropoda: *Toxoglossa*). *Pac. Sci.* 25 (1), 109–111.
- Olivera, B.M., Cruz, L.J., 2001. Conotoxins, in retrospect. *Toxicon* 39 (1), 7–14.
- Olivera, B.M., Rivier, J., et al., 1990. Diversity of *Conus* neuropeptides. *Science* 249 (4966), 257–263.
- Olivera, B.M., 1997. E.E. Just Lecture, 1996. *Conus* venom peptides, receptor and ion channel targets and drug design: 50 million years of neuropharmacology. *Mol. Biol. Cell.* 8 (11), 2101–2109.
- Phui Yee, J.S., Nanling, G., et al., 2004. Snake postsynaptic neurotoxins: gene structure, phylogeny and applications in research and therapy. *Biochimie* 86 (2), 137–149.
- Rice, R.D., Halstead, B.W., 1968. Report of fatal cone shell sting by *Conus geographus* Linnaeus. *Toxicon* 5 (3), 223–224.
- Terlau, H., Olivera, B.M., 2004. *Conus* venoms: a rich source of novel ion channel-targeted peptides. *Physiol. Rev.* 84 (1), 41–68.
- Yoshida, S., 1984. An estimation of the most dangerous species of cone shell, *Conus (Gastriidium) geographus* Linne, 1758, venom's lethal dose in humans. *Jpn. J. Hyg.* 39 (2), 565–572.

**APPENDIX I. A “CONOVENOMIC” ANALYSIS OF THE MILKED VENOM  
FROM THE MOLLUSK-HUNTING CONE SNAIL *CONUS TEXTILE* - THE  
PHARMACOLOGICAL IMPORTANCE OF POST-TRANSLATIONAL  
MODIFICATIONS<sup>6</sup>**

---

<sup>6</sup> Reproduced in part with permission from Bergeron, Z. L.; Chun, J. B.; Baker, M. R.; Sandall, D. W.; Peigneur, S.; Yu, P. Y. C.; Thapa, P.; Milisen, J. W.; Tytgat, J.; Livett, B. G.; Bingham, J-P. A “conovenomic” analysis of the milked venom from the mollusk-hunting cone snail *Conus textile*—The pharmacological importance of post-translational modifications. *Peptides* **2013**, *49*, 145–158. Copyright © 2013, rights managed by Elsevier, Inc.



## A 'conovenomic' analysis of the milked venom from the mollusk-hunting cone snail *Conus textile*—The pharmacological importance of post-translational modifications



Zachary L. Bergeron<sup>a,1</sup>, Joycelyn B. Chun<sup>a</sup>, Margaret R. Baker<sup>a</sup>, David W. Sandall<sup>c</sup>, Steve Peigneur<sup>b</sup>, Peter Y.C. Yu<sup>a</sup>, Parashar Thapa<sup>a</sup>, Jeffrey W. Milisen<sup>a</sup>, Jan Tytgat<sup>b</sup>, Bruce G. Livett<sup>c</sup>, Jon-Paul Bingham<sup>a,\*,1</sup>

<sup>a</sup> Department of Molecular Biosciences and Bioengineering, University of Hawai'i, Honolulu, HI 96822, USA

<sup>b</sup> Laboratory of Toxicology and Pharmacology, University of Leuven (KU Leuven), Campus Gasthuisberg O&N II, Leuven 3000, Belgium

<sup>c</sup> Department of Biochemistry and Molecular Biology, Bio21 Institute, University of Melbourne, Parkville, Victoria 3010, Australia

### ARTICLE INFO

#### Article history:

Received 10 April 2012

Received in revised form 8 September 2013

Accepted 9 September 2013

Available online 18 September 2013

#### Keywords:

$\alpha$ -Conotoxin

Conopeptides

*Conus textile*

Mass spectrometry

Milked venom

Radula tooth

Post-translational modifications

nAChR

### ABSTRACT

Cone snail venoms provide a largely untapped source of novel peptide drug leads. To enhance the discovery phase, a detailed comparative proteomic analysis was undertaken on milked venom from the mollusk-hunting cone snail, *Conus textile*, from three different geographic locations (Hawai'i, American Samoa and Australia's Great Barrier Reef). A novel milked venom conopeptide rich in post-translational modifications was discovered, characterized and named  $\alpha$ -conotoxin TxIC. We assign this conopeptide to the 4/7  $\alpha$ -conotoxin family based on the peptide's sequence homology and cDNA pre-propeptide alignment. Pharmacologically,  $\alpha$ -conotoxin TxIC demonstrates minimal activity on human acetylcholine receptor models (100  $\mu$ M, <5% inhibition), compared to its high paralytic potency in invertebrates, PD<sub>50</sub> = 34.2 nMol kg<sup>-1</sup>. The non-post-translationally modified form, [Pro]<sup>2,8</sup>[Glu]<sup>16</sup> $\alpha$ -conotoxin TxIC, demonstrates differential selectivity for the  $\alpha$ 3 $\beta$ 2 isoform of the nicotinic acetylcholine receptor with maximal inhibition of 96% and an observed IC<sub>50</sub> of 5.4  $\pm$  0.5  $\mu$ M. Interestingly its comparative PD<sub>50</sub> (3.6  $\mu$ Mol kg<sup>-1</sup>) in invertebrates was  $\sim$ 100 fold more than that of the native peptide. Differentiating  $\alpha$ -conotoxin TxIC from other  $\alpha$ -conotoxins is the high degree of post-translational modification (44% of residues). This includes the incorporation of  $\gamma$ -carboxyglutamic acid, two moieties of 4-trans hydroxyproline, two disulfide bond linkages, and C-terminal amidation. These findings expand upon the known chemical diversity of  $\alpha$ -conotoxins and illustrate a potential driver of toxin phyla-selectivity within *Conus*.

© 2013 Elsevier Inc. All rights reserved.

**Abbreviations:**  $\alpha$ , amino acid; CH<sub>3</sub>CN, acetonitrile; Acn, acetamidomethyl; API, atmospheric pressure ionization; DHB, dihydroxybenzoic acid; cDNA, complimentary DNA; CHCA,  $\alpha$ -Cyano-4-hydroxycinnamic acid; CID, collision induced dissociation; DCM, dichloromethane; DIEA, *N,N*-diisopropylethylamine; DMF, dimethylformamide; DV, duct venom; ESMS, electro spray mass spectrometry; FDA, Food and Drug Administration; Fmoc, 9-fluorenylmethoxycarbonyl; Fmoc-Asn(Trt)-OH, N-alpha-9-Fluorenylmethoxycarbonyl-N-beta-Trityl-L-Asparagine; Fmoc-Asp(OtBu)-OH, 9-Fluorenylmethoxycarbonyl-L-Aspartic Acid-beta-t-Butyl Ester; Fmoc-Arg(Pbf)-OH, N-alpha-9-Fluorenylmethoxycarbonyl-N-gamma-Trityl-L-Glutamine; Fmoc-Cys(Acm)-OH, 9-Fluorenylmethoxycarbonyl-S-Acetamidomethyl-L-Cysteine; Fmoc-Cys(Trt)-OH, 9-Fluorenylmethoxycarbonyl-S-Trityl-L-Cysteine; Fmoc-Gla(otBu)<sub>2</sub>-OH, N-Fmoc-L-gamma-Carboxyglutamic Acid gamma-Di-t-Butyl Ester; Fmoc-Gln(Trt)-OH, N-alpha-9-Fluorenylmethoxycarbonyl-N-gamma-Trityl-L-Glutamine; Fmoc-His(Trt)-OH, N-alpha-9-Fluorenylmethoxycarbonyl-Nim-Trityl-L-Histidine; Fmoc-Hyp(tBu)-OH, 9-Fluorenylmethoxycarbonyl-O-t-Butyl-L-Hydroxyproline; GBR, Great Barrier Reef; (Gla), gamma-gammacarboxyglutamic acid; HCTU, 1H-Benzotriazolium-1-[bis(Dimethylamino)Methylene]-5-Chloro Hexafluorophosphate-(1-),3-Oxide; IND, investigational new drug; MALDI-MS, matrix assisted desorption/ionization mass spectrometry; MALDI-TOF-(TOF)-MS, matrix assisted desorption/ionization-time of light-(time of flight)-mass spectrometry; M<sub>r</sub>, average molecular mass; mRNA, messenger RNA; MS, mass spectrometry; MV, Milked venom; m/z, mass to charge ratio; nAChR, nicotinic acetylcholine receptor; PCR, polymerase chain reaction; PSD, post-source decay; PTM, post-translational modification; RE, radula extract; RV, Radula Venom; RP-HPLC, reverse phase - high performance liquid chromatography; R<sub>t</sub>, Retention time; SEM, scanning electron micrograph; SPPS, solid phase peptide synthesis; TCEP, Tris(2-carboxyethyl)phosphine; TFA, Trifluoroacetic acid; UV, Ultra-violet; TIPS, Triisopropylsilane.

\* Corresponding author at: Department of Molecular Biosciences and Bioengineering, College of Tropical Agriculture and Human Resources, University of Hawai'i, HI 96822, USA. Fax: +1 808 965 3542.

E-mail address: [jbingham@hawaii.edu](mailto:jbingham@hawaii.edu) (J.-P. Bingham).

<sup>1</sup> Contributed equally to this work.

## 1. Introduction

The genus *Conus* (Family: *Conidae*, Subfamily: *Coninae*) has maintained a position of predatory superiority within tropical marine ecosystems for some 50 million years. The genus attributes its evolutionary success to the development and delivery of a venomous cocktail. The cone snail injects this venom through a hypodermic needle-like radula harpoon that can penetrate deep into the dermis of its prey [19]. These venoms comprise a potent pharmacopeia of individual bioactive peptide constituents, commonly referred to as conotoxins or conopeptides. New estimates indicate that half-a-million distinct biologically active molecules are expressed within this genus alone [17,38,49].

Stemming from their intended predatory use as potent neurotoxic agents for prey immobilization [12,21], the chemical complexity and functional side-chain variability of conopeptides has been studied, whereby providing tools for probing ion channel function and structure-activity relationships. The majority of research has focused on the clinical significance and therapeutic potential of conopeptides. This is evident by the number of conopeptide derived molecules in pre-clinical development or clinical phase trials for the treatment of a broad spectrum of conditions ranging from neuropathic pain and Alzheimer's, to Parkinson's and epilepsy [5,20,22,30,47,55].

Early investigations using milked venom from piscivorous species has fuelled rapid identification of a number of novel bioactive peptides including  $\alpha$ A-conotoxin OIVA from *Conus obscurus* [50] and  $\omega$ -conotoxin SIIIA from *Conus striatus* [48]. Most notably, in 2004, the United States Food and Drug Administration (FDA) approved Prialt™ ( $\omega$ -Conotoxin MVIIA), a peptide naturally expressed in the milked venom of *Conus magus*, for the treatment of chronic neuropathic pain [7,26]. Unfortunately, due to this heightened medical interest, in conjunction with intensive exploitation by the ornamental shell-trade industry, the over-harvesting of *Conus* has the potential to result in the depletion of population densities [9,16]. Subsequently, the need for a rapid and biosustainable approach to identify novel, clinically significant conopeptides while simultaneously reducing pressure on native snail populations has emerged.

In order to increase our understanding of molluscivorous milked venoms, the current study was initiated to perform a comprehensive investigation into the venomous multiplicity of *Conus textile* using Reverse-Phase High Performance Liquid Chromatographic (RP-HPLC<sup>2</sup>) profiling and mass spectral analysis of both duct (DV) and milked venoms (MV) collected from non-captive specimens representing diverse locations throughout the Pacific. This widely distributed tropical species currently represents one of the most well-studied molluscivore cone snail with countless molecular constituents, including 77 fully characterized conopeptides (see Table S1).

Our results from the combined 'conovenomic' approach revealed an unexpected level of diversity within the venom profiles of geographically unique populations of *C. textile*, and further demonstrated molecular consistency within individuals between the profiles of (i) MV, (ii) dissected whole DV, and (iii) Radula lumen Extract (RE). The comparison of geographically diverse venom profiles led to the identification of a previously uncharacterized peptide,  $\alpha$ -conotoxin TxIC, which exhibits extensive post-translational modifications (PTMs). Here we demonstrate how PTM amino acids act as determining factors for selective targeting of the nicotinic acetylcholine receptor (nAChR)—a well-documented therapeutic target for the treatment of chronic neuropathic pain [8,31,44,51,52].

## 2. Materials and methods

### 2.1. Snail milking

MV from non-captive *C. textile* was obtained within 24 hours of field collection. Envenomation was stimulated by the presence of live gastropods *Morula marginalba*, *Strombus luhuanus* or *Cypraea caputserpentis*. On extension of the cone snail's proboscis, a pipette fitted with a 5000  $\mu$ L tip was depressed and placed near the upturned aperture, close to the foot of the prey. On subsequent firing of the radula, envenomation was observed by the release of excess venom, typically as a visible 'cloud', which was carefully aspirated to avoid dispersion. The collected MV was acidified (1% v/v TFA) and either frozen ( $-20^{\circ}\text{C}$ ) or lyophilized for later analysis. Aliquots of seawater (blanks) were collected and processed in the identical manner.

### 2.2. Dissected duct venom extract preparation

Whole venom ducts were dissected from live specimens, individually dried by Speed-Vac, weighed and homogenized to obtain a fine powder. A mixture of (1000  $\mu$ L) 95% Solvent A (0.1% v/v TFA/aq.) and 5% Solvent B (90/10 v/v  $\text{CH}_3\text{CN}/0.08\%$  v/v TFA/aq.) was used as the extracting solvent at a standardized concentration (1 mg mL<sup>-1</sup>). Samples were vortexed (30 s), sonicated (10 min) and then centrifuged (4,500 g for 10 min.). The resulting supernatant was decanted, dried via Speed-Vac, weighed and then stored at  $-20^{\circ}\text{C}$  until required. All materials were dissolved in 500  $\mu$ L of the above solvent, sonicated (5 min), filtered (0.2  $\mu$ m membrane) and then centrifuged (12,000  $\times$  g for 10 min) prior to chromatographic separation and analysis.

### 2.3. Chromatographic separation and analysis, RP-HPLC

Representative MV peptides and DV extracts were individually separated as follows: (i) Capillary Scale (Phenomenex; C<sub>18</sub>, 5  $\mu$ m, 300  $\text{\AA}$ , 1.0  $\times$  250 mm, flow 100  $\mu$ L min<sup>-1</sup>)—used for comparative RP-HPLC/UV profiling, quality control of peptide purity, peptide quantification and peptide co-elution experiments. (ii) Analytical Scale (Vydac; C<sub>18</sub>, 5  $\mu$ m, 300  $\text{\AA}$ , 4.2  $\times$  250 mm, flow 1 mL min<sup>-1</sup>)—used for the isolation and purification of native peptides for MALDI-TOF-(TOF)-MS and Edman Degradation (Sections 2.4 and 2.5). (iii) Preparative Scale (Vydac; C<sub>18</sub>, 10  $\mu$ m, 300  $\text{\AA}$ , 22  $\times$  250 mm, flow 5 mL min<sup>-1</sup>)—used for the desalting and preparative separation of native MV peptides for sequencing and pharmacological assay (Sections 2.9 and 2.12). Systems (i) and (ii) used a Waters 2695 Alliance RP-HPLC System interfaced with a 996 Waters Photo Diode Array Detector for automated sample analysis and detection. Data was acquired and analyzed using Waters Millennium<sup>32</sup> (v3.2) software. Samples were eluted using a linear 1% min<sup>-1</sup> gradient of organic (90/10% v/v  $\text{CH}_3\text{CN}/0.08\%$  v/v aq. TFA) Solvent B against aqueous (0.1% v/v TFA aq.) Solvent A for 65 min, excluding a terminating high organic wash (80% Solvent B for 5 min), and pre-equilibration step (5% Solvent B) for 10 min prior to sample injection. Eluent was monitored from 200–300 nm and extracted at 214 nm. Preparative RP-HPLC/UV, system (iii), used a 625 Waters HPLC pump and controller interfaced with a 996 Waters Photo Diode Array Detector. Both gradient control and data acquisition were facilitated by the use of the Waters Millennium<sup>32</sup> software. Filtered (Nylon 0.22  $\mu$ m) crude MV and DV peptide extracts were manually loaded and eluted from the preparative scale column using the same 1% gradient at 5 mL min<sup>-1</sup> and monitored at 214 and 280 nm. Fractions were collected manually and stored at  $-20^{\circ}\text{C}$  or freeze-dried until required.

#### 2.4. Direct ESI-MS infusion

AB/MDS-Sciex API 3000 triple quadrupole mass spectrometer (Thornhill, Ontario, Canada) was used in this investigation as previously described by Chun et al. [10]. The ESI-MS system was calibrated manually in positive mode with PPG 3000 (AB/MDS-Sciex) to achieve <5-ppm mass accuracy, as per manufacturer's protocol.

#### 2.5. MALDI-TOF MS venom analysis

ZipTip™ or RP-HPLC/UV purified venom fractions in Solvent A (0.1% v/v aq. TFA) were mixed 1:1 with matrix solution (40 g L<sup>-1</sup> 2,5-dihydroxybenzoic acid (DHB) in 1:1 0.1% v/v aq. TFA: CH<sub>3</sub>CN) and 1 μL was spotted on a MTP 384 polished steel target plate (Bruker Daltonics). The spots were dried under a stream of N<sub>2</sub> gas. Mass spectra were acquired on the Ultraflex III (Bruker Daltonics), controlled by the Compass 1.2 SR1 software package (Bruker Daltonics), in positive reflector mode from *m/z* 500 to 5000. Mass spectra were summed (400 to 1200 laser shots) until no further improvement to the signal to noise ratio of peaks was achieved. Peptide II Calibration Mix (Bruker Daltonics) was used for external calibration, with a mass accuracy of approximately 50-ppm. Analysis of the spectra was completed using FlexAnalysis v3.0 (Bruker Daltonics).

#### 2.6. MALDI-TOF/TOF MS peptide sequencing and PTM characterization

The reduced venom peptide, α-conotoxin TxIC, was desalted via ZipTip™ (C<sub>18</sub> reversed-phase media, Millipore) or RP-HPLC/UV isolated and spotted onto the target plate with DHB matrix, as described above (Section 2.5). Tandem mass spectra (MS/MS) were acquired in reflector positive LIFT mode on the UltraflexIII (Bruker Daltonics), externally calibrated with Peptide Calibration Mix II (Bruker Daltonics) with a MS/MS accuracy of 0.04 Da. FlexAnalysis v3.0 (Bruker Daltonics) was used for manual inspection and annotation of the LIFT-spectra. The RapiDeNovo module in BioTools (Bruker Daltonics) was used to make additional assignments to the amino acid sequence.

#### 2.7. MALDI-TOF MS radula analysis

*Conus textile* radulae were collected from dissected radula sacs. Harpoons were prepared for MALDI-TOF MS analysis as described previously by Chun et al. [10].

#### 2.8. Peptide reduction and thiol alkylation

For complete peptide reduction, *Conus* extracts and peptide(s) (typically 0.02–3 mg) were exposed to 100 μL of 200 mM Tris(2-carboxyethyl)phosphine (TCEP; Pierce Chemicals, USA) in 25 mM NH<sub>4</sub>OAc (pH 4.5) and heated for 5–30 min at 50 °C. Alkylation of the RP-HPLC purified, reduced peptide(s) was achieved by dissolving the peptide in 90–150 μL of 25 mM NH<sub>4</sub>OAc (pH 4.5) and adding 100 mM *N*-phenylmaleimide or Maleimide (Fluka, Switzerland) in isopropanol. Typically 20–40 fold excess (w/w) of the alkylating agent was used. Alkylation was allowed to proceed at 50 °C for 15 min, prior to RP-HPLC/UV purification.

#### 2.9. Sequencing–Edman degradation

Non-alkylated and Maleimide alkylated derivatives were applied to Polybrene-treated glass fiber support filters for automated Edman degradation on a gas-phase sequencer (Model 470A; Applied Biosystems, Foster City, CA, USA). Assignment of the amino

acid sequence was essentially as described by Atherton et al. [2] and Matsudaira [34].

#### 2.10. Peptide synthesis

α-Conotoxin TxIC [ROQC<sup>4</sup>CSHOAC<sup>10</sup>NVDHPγIC-NH<sub>2</sub>; γ=γ-carboxyglutamic acid (Gla), O=4-*trans* hydroxyproline (Hyp)] was manually assembled using a 0.5 mmole scale and Fmoc SPPS via *in-situ* neutralization with 2 mmole Fmoc-amino acid per 10–30 min single coupling, as adapted from Schnölzer et al. [46] and described in detail in Kapono et al. [26]. A Fmoc-Cys(Trt)-Rink-Amide MBHA Resin (0.55 meq. g<sup>-1</sup>; Peptides International, Louisville, KY, USA) was used for peptide assembly to give the desired C-terminal amide function. Side chain protecting groups were: Cys(Trt), Asp(tBu), Arg(Pbf) and Gln(Trt). Gla(otBu)<sub>2</sub>, His(Trt), Asn(otBu) and Hyp(tBu) (as supplied by Peptides International, Louisville, KY, USA). An additional 0.5 mmole scale synthesis of α-conotoxin TxIC was undertaken using orthogonally protected Cys(Acm) being placed in positions 4 and 10 in the above sequence. This synthesis used the same synthetic strategy as previously mentioned, but deviated using an alternative directive/sequential oxidation process (see below). Finally, a non-PTM amino acid variant of α-conotoxin TxIC, (i.e. [Pro]<sup>2,8</sup>[Glu]<sup>16</sup>α-conotoxin TxIC), was produced and folded using the same orthogonally Cys(Acm) protected scheme as above. Incorporated were the Fmoc amino acids Pro and Glu(OtBu) used as PTM substituents.

#### 2.11. TFA cleavage

Fmoc peptidyl-resins were subjected to a cleavage with 82.5% v/v TFA in the presence of thioanisole (5% v/v), Phenol (5% v/v) H<sub>2</sub>O (5% v/v) and TIPS (2.5% v/v), acting as protecting group scavengers, for 2.5 h at 24 °C. The resulting cleaved peptide material was recovered by filtration and cold *t*-butyl ether precipitation. Resulting crude peptide was stored at –20 °C, as lyophilized powder, until required.

#### 2.12. Random disulfide bond formation

Peptide oxidation was achieved with 15 mg of TCEP reduced C<sub>18</sub> RP-HPLC purified peptide, using 0.1 M NH<sub>4</sub>HCO<sub>3</sub> pH 8.7 (5 days stirring; room temp.). The oxidized peptide was preparative RP-HPLC/UV isolated and re-purified, and the oxidized molecular mass was verified by ESI-MS.

#### 2.13. Directed disulfide bond formation

Cleaved α-conotoxin TxIC and [Pro]<sup>2,8</sup>[Glu]<sup>16</sup>α-conotoxin TxIC, both containing Cys(Acm) in positions 4 and 10, were RP-HPLC/UV purified, and then air oxidized, as above. Partially oxidized materials, as confirmed by ESI-MS, were then subjected to spontaneous thiol deprotection and disulfide bond formation. Deprotection was achieved by dissolving the partially folded peptide in 50% v/v acetic acid (1 mg mL<sup>-1</sup>) and by adding a solution of freshly saturated I<sub>2</sub> in 50% v/v acetic acid to the stirring peptide (25% reaction vol.). Reaction was quenched after 5 min with the addition of 10 μL aliquots of 1 M Na<sub>2</sub>S<sub>2</sub>O<sub>3</sub> until the stirring solution became clear, which was then followed by the addition of 200 μL TFA. Resulting acidified material was centrifuged (12,000 × *g*, 5 min) and directly purified by preparative RP-HPLC/UV (as above) with mass confirmation provided by ESI-MS.

#### 2.14. Genetic analysis of α-conotoxin TxIC

The venom duct and bulb of *C. textile* specimens (Great Barrier Reef, Australia) were removed by dissection and snap frozen in

liquid N<sub>2</sub>. The secretory venom duct and bulb were ground and mRNA was extracted using a Dynabeads mRNA direct kit (Dyna, Norway). A cDNA library was created from this mRNA using a Marathon cDNA amplification kit (Clontech) as previously described [44]. All PCRs utilized the primers Uni  $\alpha$ -1 (universal  $\alpha$ -conotoxin primer) 5' ATGGGCATGCGGATGATGTT 3'; and Uni  $\alpha$ -2 (universal  $\alpha$ -conotoxin primer) 5' CGGAAAGTGAAGCAGGTCAG 3', designed against conserved sequences found on the 5' and 3' conserved regions of known conotoxins. Reaction mixtures contained Taq polymerase (Roche) and deoxynucleotides in a buffer supplied by the manufacturer. Samples were incubated at 94 °C for 2 min; followed by 30 cycles of 94 °C for 30 s, an annealing step for 30 s, 72 °C for 45 s; concluding with a final step of 72 °C for 5 min. Amplification products were purified after separation on a 1.5% agarose gel using a Qiaquick gel extraction kit (Qiagen). The purified PCR products were transformed into competent INV $\alpha$ F *Escherichia coli* cells in accordance with the manufacturer's specifications (Invitrogen, Netherlands). Recombinant plasmids containing inserts of approximately 250 bp were sequenced by the di-deoxy chain termination method using the ABI PRISM Big-Dye Terminator Cycle Sequence Ready Reaction Kit (Perkin Elmer, USA). Sequences were analyzed on a Perkin Elmer 377 sequencer at the Australian Genome Research Facility (AGRF).

## 2.15. Pharmacology

### 2.15.1. Expression of voltage-gated ion channels in *Xenopus laevis* oocytes

For the expression of nAChR ( $\alpha$ 1,  $\alpha$ 3,  $\alpha$ 4,  $\alpha$ 5,  $\beta$ 2,  $\beta$ 4,  $\gamma$ ;  $\delta$ ;  $\epsilon$ ) in *Xenopus* oocytes, the linearized plasmids were transcribed using the T7 or SP6 mMACHINE-mMACHINE transcription kit (Ambion®, Carlsbad, CA, USA). The harvesting of stage V–VI oocytes from anaesthetized female *Xenopus laevis* frog was previously described by Liman *et al.* [29]. Oocytes were injected with 50 nL of cRNA at a concentration of 1 ng nL<sup>-1</sup> using a micro-injector (Drummond Scientific®, Broomall, PA, USA). The oocytes were incubated in a solution containing (in mM): NaCl, 96; KCl, 2; CaCl<sub>2</sub>, 1.8; MgCl<sub>2</sub>, 2 and HEPES, 5 (pH 7.4), supplemented with 50 mg L<sup>-1</sup> gentamycin sulfate.

### 2.15.2. Electrophysiological recordings

Two-electrode voltage-clamp recordings were performed at room temperature (18–22 °C) using a Geneclamp 500 amplifier (Molecular Devices®, Downingtown, PA, USA) controlled by a pClamp data acquisition system (Axon Instruments®, Union City, CA, USA). Whole cell currents from oocytes were recorded 1–4 days after injection. Bath solution composition was (in mM): NaCl, 96; KCl, 2; CaCl<sub>2</sub>, 1.8; MgCl<sub>2</sub>, 2 and HEPES, 5 (pH 7.4). Voltage and current electrodes were filled with 3 M KCl. Resistances of both electrodes were kept between 0.5 and 1.5 M $\Omega$ . During recordings, the oocytes were voltage-clamped at a holding potential of -70 mV and superfused continuously with ND96 buffer via gravity-fed tubes at 0.1–0.2 mL min<sup>-1</sup>, with 5 min incubation times for the bath-applied peptides. Acetylcholine (ACh) was applied via gravity-fed tubes until peak current amplitude was obtained (1–3 s), with 1–2 min washout periods between applications. Data was sampled at 500 Hz and filtered at 200 Hz. Peak current amplitude was measured prior to and following incubation of the peptide.

To assess the concentration-response relationships, data points were fitted with the Hill equation:  $y = 100/[1 + (EC_{50}/[toxin])^h]$ , where  $y$  is the amplitude of the toxin-induced effect,  $EC_{50}$  is the toxin concentration at half maximal efficacy,  $[toxin]$  is the toxin concentration and  $h$  is the Hill coefficient. Comparison of two sample means was made using a paired Student's  $t$  test ( $p < 0.05$ ). All data is presented as mean  $\pm$  standard error (S.E.M) of at least 4 independent experiments ( $n \geq 4$ ). All data was analyzed using pClamp

Clampfit 10.0 (Molecular Devices®, Downingtown, PA, USA) and Origin 7.5 software (Originlab®, Northampton, MA, USA).

## 2.16. Whole animal bioassay

Standardized concentrations of native  $\alpha$ -conotoxin TxIC (2.5, 5, 10, 20, 40, 80, 160, 320 and 640 pMol g<sup>-1</sup>) and synthetic [Pro]<sup>2,8</sup>[Glu]<sup>16</sup> $\alpha$ -conotoxin TxIC (0.16, 0.32, 0.64, 1.28, 2.56, 5.12 and 10.24 nMol g<sup>-1</sup>) in 10  $\mu$ L volumes (PBS), were injected intramuscularly (IM - foot) into Hawaiian snakehead cowries (*Cypraea caputserpentis*) using a Hamilton 10  $\mu$ L syringe at a depth of 2 mm. Following injection, animals were placed in glass Petri dishes filled with fresh aerated seawater, where paralysis was determined by the inability of the animal to cling to the substrate. Dosage and paralysis were recorded and plotted according to Reed and Muench [40], and the PD<sub>50</sub> was extrapolated using GraphPad Prism Software (v5.02). All dose experiments were repeated in triplicate or greater ( $n \geq 3$ ).

## 3. Results

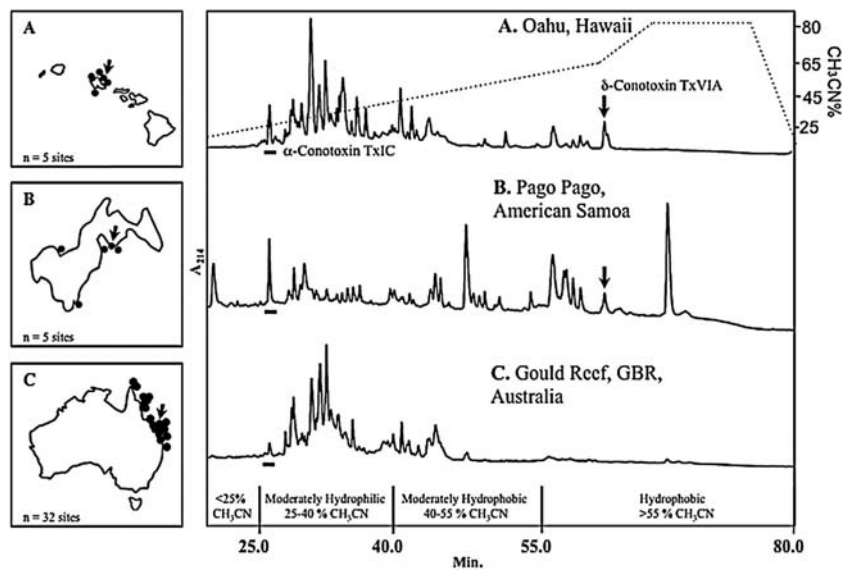
### 3.1. Conformational analysis of $\delta$ -conotoxin TxVIA

The MH<sup>+</sup> 3035.4 Da was identified by ESI-MS from RP-HPLC/UV fractions of DV, confirming the identity of  $\delta$ -conotoxin TxVIA (Calc. MH<sup>+</sup> 3035.3 Da). TCEP reduction, RP-HPLC/UV separation and ESI-MS analyzed demonstrated a mass shift of 6 Da (Obs. Red. MH<sup>+</sup> 3041.5 Da), which reflects disulfide content. This same material (~300 nmole) was then subjected to 27 cycles of Edman degradation, and a single unambiguous sequence that corresponded to  $\delta$ -conotoxin TxVIA W(-)KQSGEM(-)NLLDQN(-)(-)DGY(-)IVLV(-)T was obtained. 'Blanks' (-) were observed at degradative cycles 2, 9, 16, 17, 21 and 26. These corresponded to the 6 underivatized cysteine moieties as predicted within this specific sequence. This confirmed the expression of  $\delta$ -conotoxin TxVIA—a species-specific duct venom 'biomarker' for all specimens examined in this study. The methionine sulfoxide derivative (+16 Da) of this peptide was also observed by MALDI-TOF MS ( $m/z$  3051.3; Fig. 2A and B).

### 3.2. Geographic biodiversity - RP-HPLC/UV non-captive milked venom profiling

Typical representative MVs from non-captive specimens of *C. textile* demonstrate an abundance of chromatographic peaks ranging from 20–70 per MV profile (Fig. 1). Multiple MV samples from Hawai'i ( $n = 30$  specimens) and the Great Barrier Reef ( $n = 50$  specimens; GBR), Australia contained similar chromatographic content, with varying relative peptide concentrations as observed in the moderately hydrophilic region (25–40% CH<sub>3</sub>CN, 25 to 40 min; see Fig. 1A and C, respectively). No observable RP-HPLC/UV correlation to cone snail sex or collection/milking season could be established from the MV profiles from any of the three locations examined (not shown).

The moderate to hydrophobic RP-HPLC/UV profile region (40–80% CH<sub>3</sub>CN, 40 to 80 min) accounted for >30% of milked venom variability, with the highest proportion of peak variability seen in the representative MV of non-captive specimens from American Samoa ( $n = 20$  specimens) (Fig. 1B). The hydrophobic  $\delta$ -conotoxin TxVIA ( $R_t$  65.3 min) was observed in both non-captive Hawai'i and American Samoan representative samples (see arrow in Fig. 1A and B respectively), but was absent in non-captive GBR Australian MV (Fig. 1C). One previously undocumented peak (underlined), with a retention time ( $R_t$ ) = 26.6 min ( $\alpha$ -conotoxin TxIC), was observed in all three locations ( $n = 100$ , total specimens examined)



**Fig. 1.** RP-HPLC/UV comparison of peptides from the milked venom of representative non-captive *C. textile* specimens sampled from different geographic locations in the South Pacific: (A) Hawai'i; (B) American Samoa; and (C) the Great Barrier Reef, Australia. The expression of  $\delta$ -conotoxin TxVIA (arrow), was detected in MV samples from Hawai'i ( $n = 30$ , from 5 sites) and American Samoa ( $n = 20$ , from 5 sites), but was absent from those from GBR, Australia ( $n = 50$ , from 32 sites).  $\alpha$ -Conotoxin TxIC (underscored) was detected in MV from all three locations, indicating its importance in the envenomation strategy of *C. textile*. MALDI-TOF MS analysis of these MV samples is shown in Fig. 2. In each map, indicated are cone snail collection sites (filled circle) and individual representative samples (arrow).

(Fig. 1A–C). This observed geographic commonality lends credence to the importance of that particular conopeptide.

### 3.3. Geographic biodiversity - Molecular mass profiling of non-captive milked venom

MALDI-TOF MS profiles of MV from 5 random specimens of non-captive *C. textile* from each geographic location cover the entirety of the examined mass range (Fig. 2A–C). In the examples illustrated, a combined total of 178 non-overlapping peaks were observed in the three MVs (Table 1 & Table S2). Of these, 42 peaks (~24%) were common to two collection sites, with only 11 (~6%) being common to all three. Importantly, 27 of the 178 non-captive MV constituents (~15%) correlated to known *C. textile* peptides (Table 1; Tables S1 & S2), excluding potential PTM variants including C-terminal processing ( $\Delta \pm 1.0$  Da for free-acid vs. amide). Expression of 2 of these known conopeptides, TxIA and TxIIC, were previously confirmed by Edman degradation sequencing [4].

### 3.4. Single specimen venom source consistency–Non-captive RP-HPLC/UV profiling

RP-HPLC/UV venom peptide profiles from a single non-captive specimen of *C. textile* (Oahu, Hawai'i) is shown in Figs. 3A–C. Results illustrate a high level of peptide  $R_t$  consistency between the whole duct venom (DV) extract, the milked venom (MV), and the *Radula lumen* Extract (RE). The relative abundance of peptides in the different venom extracts was highly variable as seen with the relative intensity of the hydrophobic  $\delta$ -conotoxin TxVIA ( $R_t$  65.3 min; Fig. 3B and C).

#### 3.4.1. Duct venom (DV) extracts

In the DV,  $\delta$ -conotoxin TxVIA was one of the major components (Fig. 3C). The presence of  $\alpha$ -conotoxin TxIC in DV (underscored, Fig. 3C) was observed at  $R_t = 26.6$  min, corresponding to an identical elution profile from the MV of the same specimen (Fig. 3B) as well as in the representatives from geographic MV sampling (Fig. 1).

#### 3.4.2. Milked venom (MV) extracts

In contrast to the DV, MV profiles naturally correspond to a more accurate representation of bioactive constituents used in prey immobilization.  $\delta$ -Conotoxin TxVIA was one of only a few hydrophobic peptides observed in the MV (Fig. 3B), which exhibited direct correlation to the DV. These results are reiterated in the geographic specimens that exhibit few hydrophobic peptides (Fig. 1); most notably, Australian specimens—which may reflect venom processing. Highly abundant hydrophilic peptides within both DV and MV venom extract profiles were examined (Fig. 3B and C), which identified  $\alpha$ -conotoxin TxIC (underscored,  $R_t = 26.6$  min).

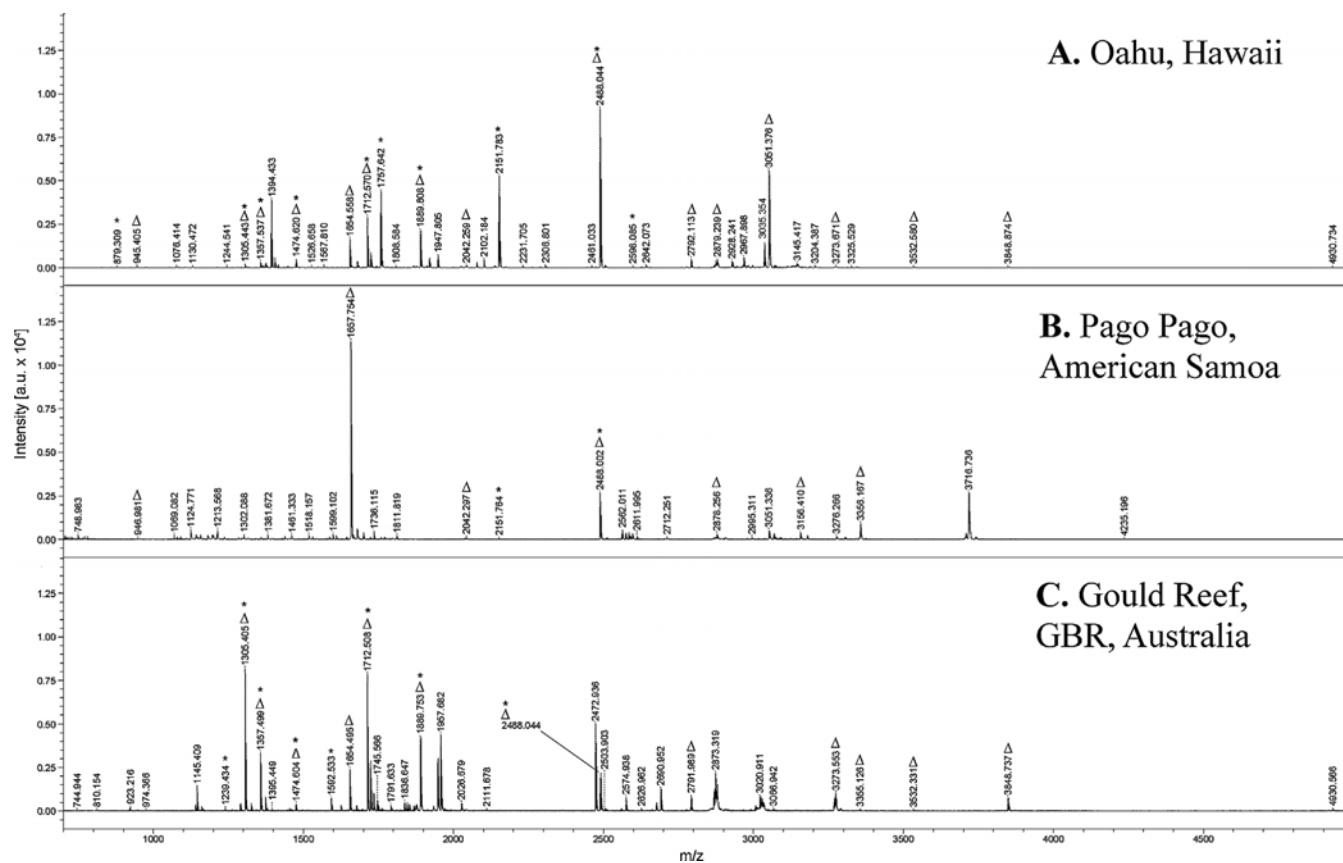
#### 3.4.3. *Radula lumen* (RE) extracts

Analysis of the RE demonstrated poorly resolved peaks with limited abundance (Fig. 3A), yet the two major peaks correlated with the two highly abundant peaks in the MV (Fig. 3B). Both peaks showed marked differences in the DV extract.

### 3.5. Single specimen venom source consistency–Non-captive molecular mass profiling

MALDI-TOF-MS was used to detect and compare identifiable peptide masses in extracts of the RE, MV and DV, from a single non-captive representative specimen of *C. textile* (Fig. 4). A total of 159 unique molecular masses were observed (Table S3). Common to all three extracts were 16 individual masses (~10%). Showing the greatest similarity were the MV and DV profiles that shared 38 masses (~24%). The RE mass spectrum, exhibiting masses within  $m/z$  945 to 3850, contained fewer masses than the MV and DV. The MV mass spectrum was the most complex in peak number of the three.

Most notable are the high intensity masses observed in both the DV and MV extracts which correspond to important and previously characterized conopeptides (Table 2 and Fig. 4). The dominant peak in the MV and DV mass spectra ( $m/z$  2488.2  $\pm$  0.3) corresponds to the *C. textile* Convulsant peptide. This peak is also present in the RE mass spectrum. Other identified peptides (i.e., TxIA and TxIIC) were also observed as major contributors in the geographic analysis (Table 1, Fig. 2).

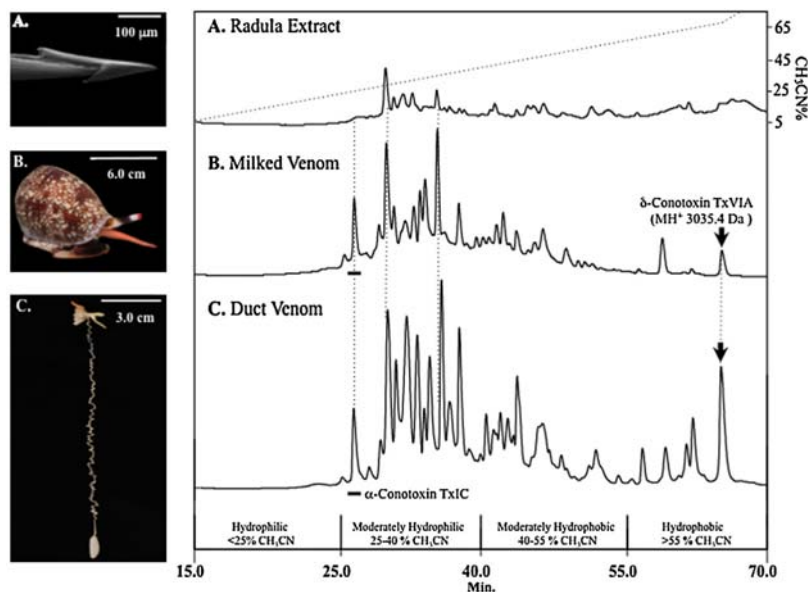


**Fig. 2.** MALDI-TOF MS mass spectra of representative non-captive milked venoms from *C. textile* specimens obtained from: (A) Hawai'i; (B) American Samoa; and (C) the Great Barrier Reef, Australia. Peaks corresponding to known conopeptides (\*) and peaks common to all 3 spectra (Δ) are indicated Table S2 provides an extensive list comparing all peaks observed in these venom profiles.

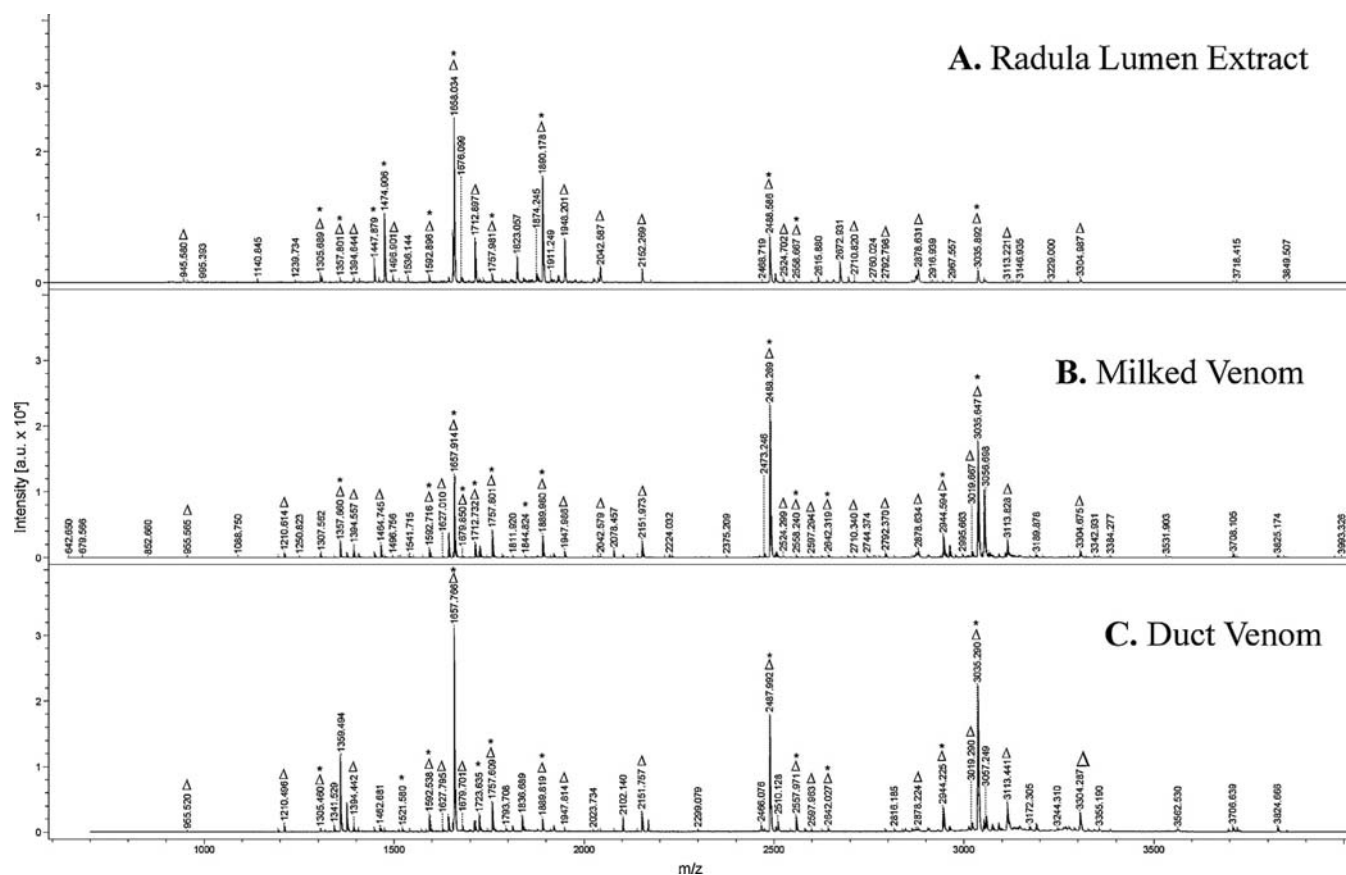
### 3.6. Mass spectrometric analysis of $\alpha$ -conotoxin TxIC disulfide bonds

A geographically consistent peptide ( $R_t = 26.6$  min; Fig. 3) was isolated from MV by RP-HPLC/UV. A single parent mass of  $MH^+$

2080.8 Da was assigned by ESI-MS (Obs.  $[M + 3H]^{+3}$  694.6 Da; data not shown). After TCEP reduction, the observed parent mass was  $MH^+ 2084.7$  Da, which indicated the presence of 2 disulfide bridges. This was confirmed by alkylation of the reduced material with *N*-phenylmaleimide, after which an increase in 692 Da was observed,



**Fig. 3.** RP-HPLC/UV comparison of venom obtained from (A) Radula Lumen Extract; (B) Milked Venom; and (C) Duct Venom from a representative specimen of *C. textile*. (A) The RE revealed limited similarity of peptides within the hydrophilic region to those in the hydrophilic regions of MV and DV. Note that  $\alpha$ -conotoxin TxIC (underscore) and  $\delta$ -conotoxin TxVIA (arrow) were not detected in the RE. Both (B) MV and (C) DV demonstrated a high level of peptide abundance and continuity. These samples were profiled via MALDI-TOF MS (Fig. 4).



**Fig. 4.** MALDI-TOF MS profiles of peptide extracts from a single representative *Conus textile* specimen obtained from: (A) Radula Lumen Extract; (B) Milked Venom and (C) Whole Duct Venom. Peptides exhibiting common peaks are listed in Table 2 and indicated ( $\Delta$ ) in each profile. These data illustrate expressional consistency, although some unique peaks are observed from each venom source. Absence of  $\alpha$ -conotoxin TxIC is apparent, but is suspected to be present as an in-source decay fragment.  $\delta$ -Conotoxin TxVIA was observed at  $m/z$   $3035.6 \pm 0.2$ . Peaks corresponding to their known *C. textile* peptides are indicated by asterisk (\*). Table S1 provides an extensive list of observed peaks corresponding to known *C. textile* venom peptides.

corresponding to 4 alkylation groups ( $\Delta$  173 Da for each free thiol [6]).

### 3.7. CID analysis of $\alpha$ -conotoxin TxIC

The parent ion of  $\alpha$ -conotoxin TxIC was not observed in MALDI-TOF MS analyses of geographic and source extracts (Figs. 2 and 4). This indicated decay of the parent ion, which is common in MALDI-TOF MS of peptides containing labile PTMs. Direct observation of this suspected decay was achieved via MALDI-TOF MS of the purified material in its reduced form (Fig. 5A). Two well resolved peaks, with a difference of 44 Da, indicated the presence of a carboxyl group. Similar decay of the parent ion could be reproduced with ESI-MS, a softer ionization method, by increasing the orifice potential voltage outside of the normal operating range (data not shown).

MALDI-TOF/TOF MS/MS fragmentation spectra of the suspected decarboxylated ( $m/z$  2041.06; Fig. 5B) and 'intact' form ( $m/z$  2805.07; data not shown) of TCEP reduced  $\alpha$ -conotoxin TxIC were nearly indistinguishable. In both cases MS/MS fragmentation displayed a 44 Da deficient a-, b-, and y-ion series. Analysis of the overlapping ion series allowed for interpretation of the amino acid sequence as R-[I/L/O]<sup>2</sup>-Q-C-C-S-H-[I/L/O]<sup>8</sup>-A-C-N-V-D-H-P-E-[I/L/O]<sup>17</sup>-C. The isobaric nature of isoleucine and leucine (I and L) and their similarity in mass to 4-*trans*-hydroxyproline (O) ( $\Delta$  0.036 Da) resulted in ambiguities at positions 2, 8 and 17, which were resolved with Edman degradation (Section 3.8).

The CID fragment ions also indicated heterogeneity of the C-terminus, the predominant form being amidated. The presence of low abundant C-terminal free acid form was verified using a

narrow precursor ion selection window of  $\pm 0.5$  Da. The MS/MS fragmentation spectrum was interpreted to have the same amino acid sequence as the amidated form (data not shown). Additional verification was obtained by comparing predicted and observed isotopic distribution patterns for the respective parent ions.

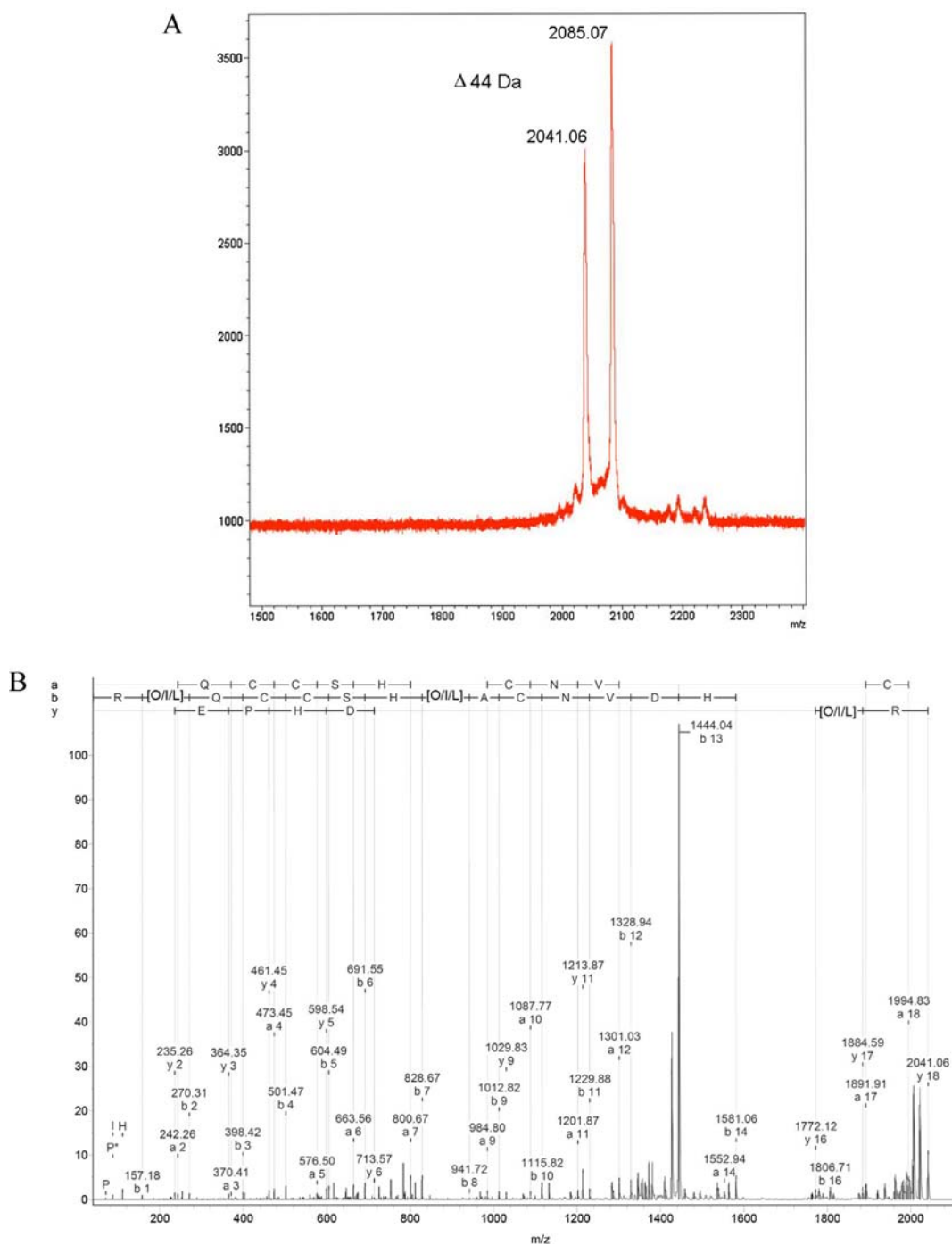
### 3.8. Edman degradation of $\alpha$ -conotoxin TxIC

RP-HPLC/UV purified *N*-phenylmaleimide thiol alkylated  $\alpha$ -conotoxin TxIC,  $\sim 300$  nmole, was subjected to 20 cycles of automated Edman degradation. PTH-4-*trans*-hydroxyproline was added to the amino acid standard mixture to aid identification. Sequential degradation provided an unambiguous 18 amino acid sequence: RO<sup>2</sup>QCCSHO<sup>8</sup>ACNVDHP(-)<sup>16</sup>IC. Cycle 16 indicated a low recovery level PTH-Glu ( $<1$  nmole) and was designated as a 'blank cycle' based on the expected sequential cycle yields observed within the normal degradative process.

Edman degradation assigned 17 of the 18 amino acids conclusively, including positions 2 and 8 as 4-*trans*-hydroxyproline and 17 as isoleucine. The presence of low levels of glutamic acid in cycle 16, combined with the observed mass difference of 44 Da ( $-\text{CO}_2$ ) under various mass spectrometric conditions, provided further corroborated the presence of a  $\gamma$ -carboxylated glutamic acid within  $\alpha$ -conotoxin TxIC at position 16.

### 3.9. Genetic analysis of $\alpha$ -conotoxin TxIC

The peptide sequence was confirmed genetically by identifying the  $\alpha$ -conotoxin transcript in *C. textile* venom duct



**Fig. 5.** (A) The decarboxylation of TCEP Reduced  $\alpha$ -conotoxin TxIC as seen by MALDI-TOF MS. Observed molecular masses differ by 44 Da. The ratio could be influenced by choice of MALDI-MS matrix (not shown). (B) CID analysis of  $\alpha$ -conotoxin TxIC. Using overlapping a-, b-, and y-ion series, the sequence was interpreted to be R-[I/L/O]-Q-C-C-S-H-[I/L/O]-A-C-N-V-D-H-P-E-[I/L/O]-C. Fragment ions also indicated decarboxylation and heterogeneity of the C-terminal amino acid.

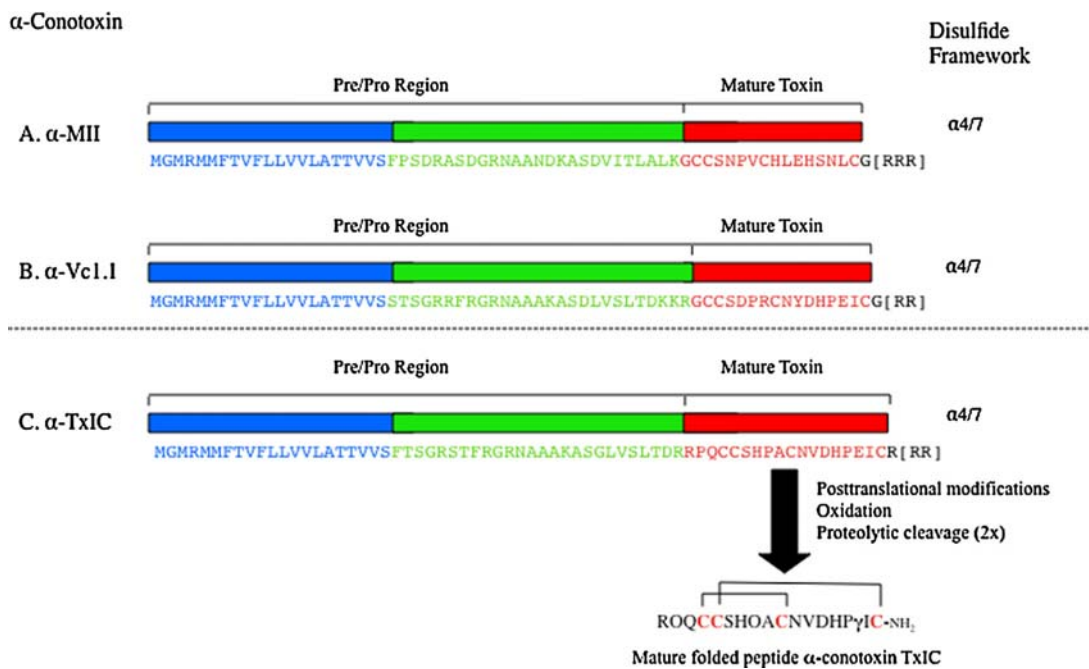
cDNA (GBR, Australia) using oligonucleotides based on regions of conserved sequence in the pre-region and C-terminal flanking region of  $\alpha$ -conotoxins [44]. The resulting 19 amino acid toxin- or mature-sequence region was deduced: RPOCCSHPAC-NVDHPEIC[R] (see Fig. 6). A C-terminal arginine was absent following Edman analysis; its inclusion would also be at variance with native molecular mass as observed by either ESI-MS, MALDI-TOF MS or by sequence analysis using MALDI-TOF/TOF MS LIFT.

The N-terminal pro-region of  $\alpha$ -conotoxin TxIC also possesses an arginase site for mature conopeptide cleavage, an occurrence common to conotoxin processing [24]. In  $\alpha$ -conotoxin TxIC the

pro-peptide cleavage site is located between two neighboring arginine moieties—see Fig. 6. This is confirmed by the recovery/identification of an N-terminal arginine in  $\alpha$ -conotoxin TxIC both by MALDI-TOF/TOF MS LIFT and Edman Degradation. Independent identification of [Glu]<sup>16</sup>, via genetic analysis, conclusively confirmed the original parent PTM amino acid assignment.

### 3.10. Peptide synthesis of $\alpha$ -conotoxin TxIC

Based on the assigned C-terminal amide sequence above,  $\alpha$ -conotoxin TxIC was synthetically constructed and folded by random air oxidation. Upon completion, the correct target molecular



**Fig. 6.** Genetic analysis of  $\alpha$ -conotoxin TxIC and its comparison to  $\alpha$ -conotoxins MII and Vc1.1. The Pre/Pro region of  $\alpha$ -conotoxin TxIC demonstrates a high level of amino acid identity to established members of the 4/7  $\alpha$ -conotoxin family. N-terminal processing of  $\alpha$ -conotoxin TxIC shares similarities with the processing of  $\alpha$ -conotoxin Vc1.1 required to produce the mature toxin. Similar mechanisms operate to process  $\alpha$ -conotoxin MII. C-terminal brackets [ ] represent stop codons within the formal cDNA sequence.

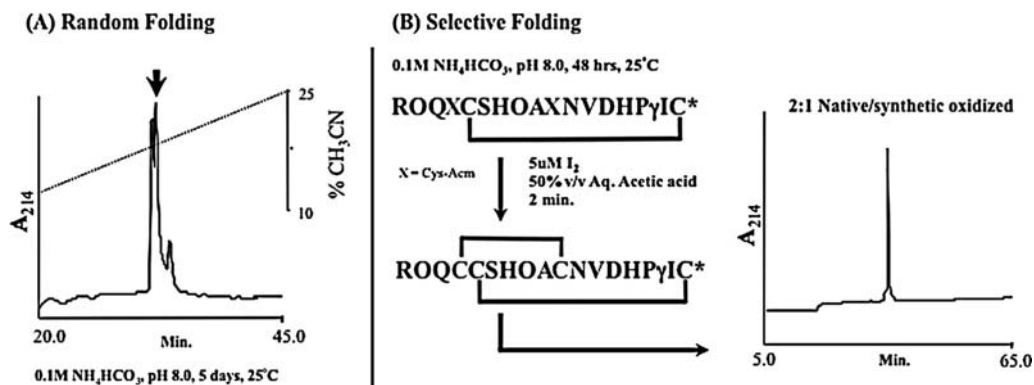
mass was observed by ESI-MS: Obs. MH+ 2080.9 Da (Calc. MH+ 2080.8 Da). RP-HPLC/UV demonstrated the presence of 3 peaks, with the two dominant peaks separated by <0.3 min (Fig. 7A). All three peaks were determined to be disulfide isomers of  $\alpha$ -conotoxin TxIC, as confirmed by ESI-MS. Separation of the two main isomers proved difficult. Changes to gradient, ion pair agents and column failed to achieve their full separation (data not shown).

A second synthesis was performed using the established disulfide connectivity pattern reported for the 4/7  $\alpha$ -conotoxin family [31]. This was achieved by incorporation of differential orthogonally protected cysteine moieties at positions 4 and 10 (Fig. 7B). This allowed for the step-wise production of a single oxidized peptide peak that retained target characteristics of mass (Obs. MH+ 2080.9 Da) and RP-HPLC/UV using a slower chromatographic gradient ( $0.5\% \text{ min}^{-1}$ ), without production of the previously observed isomeric materials (Fig. 7A). Co-injection of the RP-HPLC/UV purified synthetic material with  $\alpha$ -conotoxin TxIC

isolated from the native milked venom, in a ratio of 1:2 respectively, demonstrated chromatographic homogeneity at 33.3 min, Fig. 7B. RP-HPLC/UV co-injection of the TCEP reduced material also demonstrated chromatographic homogeneity (not shown). This confirmed the original peptide sequence assignment, including C-terminal assignment, and provided an independent synthetic standard for RP-HPLC/UV.

### 3.11. Pharmacological assessment

Native  $\alpha$ -conotoxin TxIC ( $\leq 100 \mu\text{M}$ ) had minimal inhibitory action on human nicotinic acetylcholine receptor (nAChR) subunit combinations  $\alpha 4\beta 2$  (brain),  $\alpha 4\beta 4$  (neuronal),  $\alpha 7$  (neuronal),  $\alpha\beta\gamma\delta$  (muscle),  $\alpha\beta\gamma\epsilon$  (muscle) or the human peripheral nAChR subtypes  $\alpha 3\beta 2$ ,  $\alpha 3\beta 2\alpha 5$  and  $\alpha 3\beta 4$ . Biological activity was observed with [Pro]<sup>2,8</sup>[Glu]<sup>16</sup> $\alpha$ -conotoxin TxIC, the synthetic non-PTM form (Fig. S1A & B). This analogue inhibited the human nAChR subtypes



**Fig. 7.** (A) Random folding of synthetic  $\alpha$ -conotoxin TxIC. Production of isomeric materials was achieved upon random peptide folding, which when resolved by RP-HPLC/UV demonstrated difficulties in purification. (B) Selective folding of  $\alpha$ -conotoxin TxIC and its co-elution with native material. Undertaking a separate synthetic, together with a step-wise directed folding strategy, allowed for the production of a single peak/isomer upon full oxidation. RP-HPLC/UV co-elution in oxidized and TCEP reduced from (not shown), demonstrated identical chromatographic behaviors.

**Table 1**

*Conus textile* peptides identified by molecular mass analysis from three different geographic locations.

| Obs. m/z | HI | AS | GBR | Mass correlation    |
|----------|----|----|-----|---------------------|
| 879.309  | ✓  |    |     | Leu-contryphan-Tx   |
| 1091.051 |    | ✓  |     | TeAr193             |
| 1238.410 |    | ✓  |     | TeAr151             |
| 1239.434 |    |    | ✓   |                     |
| 1305.405 |    |    | ✓   | TxIIIc (Tx3.4)      |
| 1305.443 | ✓  |    |     |                     |
| 1357.499 |    |    | ✓   | Tx10b or Tx10c      |
| 1357.537 | ✓  |    |     |                     |
| 1392.411 | ✓  |    |     | TxXIIIA             |
| 1447.593 | ✓  |    |     | Tx5b                |
| 1448.557 |    |    | ✓   |                     |
| 1474.604 |    |    | ✓   | Tx5c                |
| 1474.620 | ✓  |    |     |                     |
| 1592.509 |    |    | ✓   | Tx3d                |
| 1656.531 | ✓  |    |     | Tx1c                |
| 1657.754 |    | ✓  |     | TxIA                |
| 1678.968 |    | ✓  |     | Tx-D0111            |
| 1679.718 | ✓  |    |     |                     |
| 1712.485 |    |    | ✓   | TxIIIA              |
| 1712.570 | ✓  |    |     |                     |
| 1723.566 |    |    | ✓   | TxMLKM-021          |
| 1723.636 | ✓  |    |     |                     |
| 1757.642 | ✓  |    |     | TxMMSK-03           |
| 1756.444 |    |    | ✓   |                     |
| 1761.518 |    |    | ✓   | TxMRCL-D012         |
| 1779.609 | ✓  |    |     | TxMMSK-04           |
| 1844.578 |    |    | ✓   | Tx3e                |
| 1889.753 |    |    | ✓   | Tx3f (Tx3.5)        |
| 1889.808 | ✓  |    |     |                     |
| 2151.783 | ✓  |    |     | Tx3h (TxMLKM-011)   |
| 2487.917 |    |    | ✓   | Convulsant Peptide  |
| 2488.002 |    | ✓  |     |                     |
| 2488.044 | ✓  |    |     |                     |
| 2642.073 | ✓  |    |     | Tx6.3               |
| 2942.164 | ✓  |    |     | TxMKLT1-015         |
| 2961.308 | ✓  |    |     | Tx05                |
| 2995.234 | ✓  |    |     | Tx6.2               |
| 2995.311 |    | ✓  |     |                     |
| 3035.342 | ✓  |    |     | TxVIA (King-Kong 0) |
| 3066.942 |    |    | ✓   | TxMLKLT1-0111       |
| 3067.313 |    | ✓  |     |                     |
| 3067.339 | ✓  |    |     |                     |

Obs. m/z are within  $\pm 1.0$  Da of the identified peptide mass to account for possible amide and free acid variants.

( $\alpha 3\beta 2$ ,  $\alpha 3\alpha 5\beta 2$  and  $\alpha 3\beta 4$ ), with the  $\alpha 3\beta 4$  isoform demonstrating the highest sensitivity ( $IC_{50}$   $2.1 \pm 0.2 \mu M$ ) and exhibiting a 77% maximum inhibition. Isoform  $\alpha 3\beta 2$  demonstrated the highest level of inhibition (96%) with an  $IC_{50}$  of  $5.4 \pm 0.5 \mu M$  (Table 4). No significant inhibition was observed at the remaining isoform targets ( $\alpha 4\beta 2$ ,  $\alpha 4\beta 4$ ,  $\alpha 7$ ,  $\alpha\beta\gamma\delta$ ,  $\alpha\beta\gamma\epsilon$ ) at concentrations up to  $100 \mu M$  (Table 4).

### 3.12. Whole animal bioassay

Whole animal bioassay using the native  $\alpha$ -conotoxin TxIC demonstrated a dose-dependency in its ability to produce animal paralysis (see Fig. S2).  $\alpha$ -Conotoxin TxIC produced total paralysis at  $640 \text{ pMol g}^{-1}$  ( $n = 14$ ), this dose did not cause lethality in test animals. A  $PD_{50}$  of  $34.2 \text{ pMol g}^{-1}$  ( $34.2 \text{ nMol kg}^{-1}$ ; whole snail weight) was determined for the native toxin (Table 4). The synthetic non-PTM analogue of  $\alpha$ -conotoxin TxIC demonstrated an inability to cause paralysis at same  $PD_{50}$  concentration ( $n = 15$ ). Total paralysis was achieved with  $10.24 \text{ nMol g}^{-1}$  ( $n = 7$ ), this being  $\sim 16\times$  that of the native. A similar dose-dependent trend was observed (see Fig. S2), with the maximal dose used not causing animal lethality. A  $PD_{50}$  of  $3.6 \text{ nMol g}^{-1}$  ( $3.6 \mu\text{Mol kg}^{-1}$ ; whole snail weight) was calculated (Table 4), this  $PD_{50}$  concentration being  $\sim 100\times$  more than observed with the native PTM toxin.

**Table 2**

Comparative analysis of MALDI-TOF MS *C. textile* data from different venom sources (RE, MV and DV).

| Obs. m/z | RE | MV | DV | Mass correlation    |
|----------|----|----|----|---------------------|
| 1305.460 |    |    | ✓  | TxIIIc              |
| 1305.528 |    | ✓  |    |                     |
| 1305.689 | ✓  |    |    |                     |
| 1357.487 |    |    | ✓  | Tx10b/Tx10c         |
| 1357.660 |    | ✓  |    |                     |
| 1357.801 | ✓  |    |    |                     |
| 1392.422 |    |    | ✓  | TxXIIIA (TxMRCL-02) |
| 1392.533 |    | ✓  |    |                     |
| 1447.628 |    |    | ✓  | Tx5b                |
| 1447.747 |    | ✓  |    |                     |
| 1447.879 | ✓  |    |    |                     |
| 1474.753 |    | ✓  |    | Tx5c                |
| 1474.906 | ✓  |    |    |                     |
| 1521.536 |    |    | ✓  | TxIIIB              |
| 1592.538 |    |    | ✓  | Tx3d                |
| 1592.716 |    | ✓  |    |                     |
| 1592.896 | ✓  |    |    |                     |
| 1642.756 |    |    | ✓  | Tx5d                |
| 1642.927 |    | ✓  |    |                     |
| 1643.013 | ✓  |    |    |                     |
| 1657.766 |    |    | ✓  | TxIA                |
| 1657.914 |    | ✓  |    |                     |
| 1658.034 | ✓  |    |    |                     |
| 1679.701 |    |    | ✓  | Tx-D0111            |
| 1679.850 |    | ✓  |    |                     |
| 1710.766 |    |    | ✓  | TxIIIA              |
| 1723.635 |    |    | ✓  | TxMLKM-021          |
| 1723.787 |    | ✓  |    |                     |
| 1757.609 |    |    | ✓  | TxMMSK-03           |
| 1757.801 |    | ✓  |    |                     |
| 1757.981 | ✓  |    |    |                     |
| 1844.824 |    | ✓  |    | Tx3e                |
| 1889.819 |    |    | ✓  | Tx3f                |
| 1889.980 |    | ✓  |    |                     |
| 1890.178 | ✓  |    |    |                     |
| 1931.063 | ✓  |    |    | TxVA                |
| 2151.757 |    |    | ✓  | Tx3h                |
| 2151.973 |    | ✓  |    |                     |
| 2152.196 | ✓  |    |    |                     |
| 2488.586 | ✓  |    |    | Convulsant Peptide  |
| 2488.269 |    | ✓  |    |                     |
| 2487.992 |    |    | ✓  |                     |
| 2557.971 |    |    | ✓  | Tx02                |
| 2558.240 |    | ✓  |    |                     |
| 2558.667 | ✓  |    |    |                     |
| 2642.027 |    |    | ✓  | Tx6.3               |
| 2642.319 |    | ✓  |    |                     |
| 2944.225 |    |    | ✓  | TxMKLT1-015         |
| 2944.594 |    | ✓  |    |                     |
| 3011.221 |    |    | ✓  | King-Kong 2         |
| 3035.290 |    |    | ✓  | TxVIA (KK-0)        |
| 3035.647 |    | ✓  |    |                     |
| 3035.892 | ✓  |    |    |                     |
| 3067.747 |    | ✓  |    | TxMKLT1-0111        |
| 3107.684 |    | ✓  |    | Tx04                |

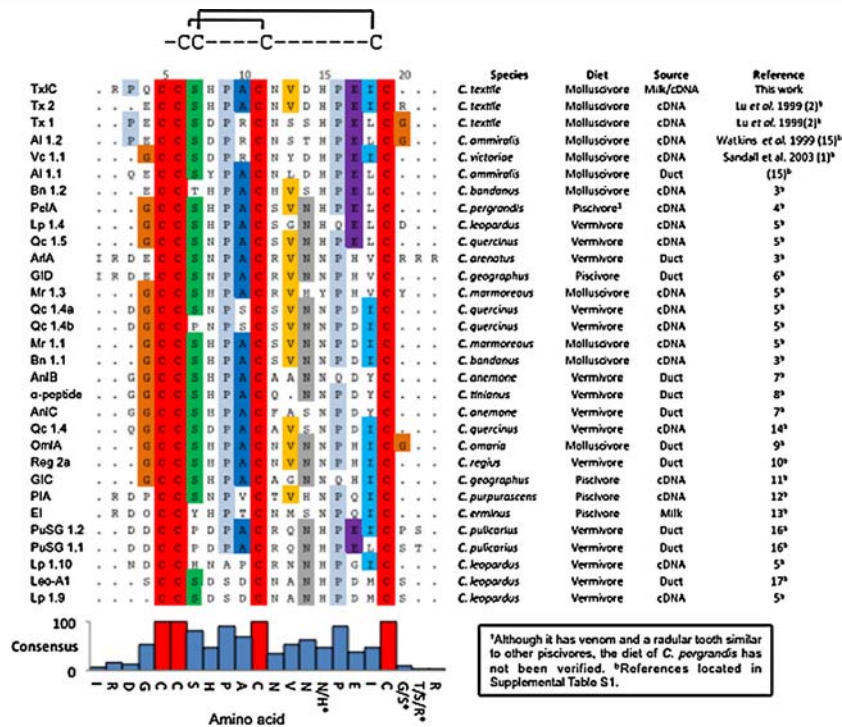
Obs. m/z are within  $\pm 1.0$  Da of the identified peptide mass to account for possible amide and free acid variants.

## 4. Discussion

This study is the first reported analysis comparing the venom profile of a single specimen of *Conus* from milked venom (MV), to duct venom (DV), and finally to radula lumen extract (RE) by RP-HPLC/UV analysis and molecular mass profiling. This has provided greater insight into the complexity of molluscivorous venom from a single specimen through to the gross variation observed at a species level from wide-ranging geographical locations.

The importance of *Conus textile* to conopeptide research cannot be underestimated. Early studies on *C. textile* [21,27,36] made major contributions to the present understanding of conopeptide processing and expression [11,23,45,54], and led to the discovery of

**Table 3**  
Sequence comparison of  $\alpha$ -conopeptide 4/7 family members to  $\alpha$ -conotoxin TxIC.



All 31 peptides, including  $\alpha$ -conotoxin TxIC, share a conserved cysteine pattern framework. Peptides most similar in sequence are those derived from *C. textile* and other molluscivores, while  $\alpha$ -conopeptides from piscivores and vermivores share less sequence similarity. The overall sequence similarity to  $\alpha$ -conotoxin TxIC is displayed graphically. This calculation is based on the total number of peptides similar to  $\alpha$ -conotoxin TxIC and the number that exhibits conservation of the same specific amino acid and position in the sequence. For example, 3 peptides have a Pro in position 3; this is divided by 17 peptides that possess an amino acid in the equivalent position 3 of the  $\alpha$ -conotoxin TxIC sequence; this gives a consensus of 18% for that position. Note that the  $\alpha$ -4/7 cysteine pattern framework is solely used as the point of sequence alignment. <sup>b</sup>References listed are Supplemental Information – References II.

novel PTM amino acids [14,37,41] and unique biological activities. The specie's relative abundance, population distribution and access has allowed for an intensive level of research, covering specimens collected from the Philippines [28], Red Sea [21], South China Sea [33], Japan [27,36] and the GBR, Australia [15,25]. In the present study, we have expanded this geographic range to include specimens collected from Hawai'i and American Samoa (Figs. 1 and 2), making this the largest geographic study of a single *Conus* species to date—this provides an indication to their regional biodiversity, an area poorly addressed in *Conus* research.

A 'conovenomic' comparison of MV from distant geographic locations, illustrates the true extent of venom peptide diversity present within *C. textile*, as alluded to by Lu et al. [33]. Our evidence further demonstrates the intraspecific venom variability identified in DV of other members of the genus [1,15,42,43]—a characteristic that now extends to MV. This may increase earlier estimates of bioactive venom constituents [18,35,36,49], and expand the therapeutic potential of *Conus* [5,7,20,30,47].

This is only further demonstrated with the identification of a novel, highly PTM-modified conopeptide,  $\alpha$ -conotoxin TxIC, despite the intensive level of research that *C. textile* has received historically. The newly observed  $\alpha$ -conotoxin demonstrates homology to Vc1.1 (~66%; Table 3), and an experimentally confirmed level of selectivity (in its non-PTM form) to human isoforms of neuronal type nAChR (Table 4), a well-documented target for the treatment of chronic neuropathic pain [8,31,44,51].

Within *C. textile*, considerable diversity is found primarily within the hydrophobic region of the RP-HPLC/UV venom profiles (Fig. 1), and prominently in specimens from American Samoa. Mass spectral analysis revealed a level of molecular mass consistency across all

locations, however the relative proportion of venom constituents varies markedly (Fig. 2; Table 1; Table S2). Peptide expressional variability supports an important biological role in the predator envenomation strategy, a factor that may aid the discovery of a novel conopeptides, as illustrated with  $\alpha$ -conotoxin TxIC.

To further refine this approach, the analysis of a single *C. textile* specimen facilitated the tracking of integral venom constituents from DV to MV and then RE. This strengthens the proteomic correlation, end-point detection and characterization of the essential secretory venom constituents used during envenomation [51]. It is these identified end-point MV peptides that have shown the strongest correlation to both potential biological activity and proven drug potential in piscivores [7,26]. Using RP-HPLC/UV we established that  $\alpha$ -conotoxin TxIC is represented in DV extract (Fig. 3C), and that expressional continuity continues through the MV (Fig. 3B). Interestingly  $\alpha$ -conotoxin TxIC was absent from the RE (Fig. 3A). For this species at least, the radula lumen extract is not representative of the venom content as a whole—however some level of conopeptide continuity, by *m/z* comparison, is clearly observable (Table 2; Fig. 4).

$\alpha$ -Conotoxin TxIC represents a novel toxin that highlights some of the issues that have plagued conopeptide discovery and sequencing, necessitating a combined 'conovenomic' approach to obtain a thorough and accurate sequence analysis. The sole use of mass spectrometry for *Conus* venom analysis can lead to misidentification of sequence and molecular mass arising from inconsistencies associated with sample preparation, matrix selection and the ionization technique used. However we are unsure of the extent of PTM 'miss assignment' that may exist in the 1000s of conopeptides already documented in the publically available conopeptide

**Table 4**  
Pharmacological analysis of  $\alpha$ -conotoxin TxIC and [Pro]<sup>2,8</sup>[Glu]<sup>16</sup> $\alpha$ -conotoxin TxIC.

|                                                                     | nAChR subtype <sup>a</sup>                                                                                                          | IC <sub>50</sub> <sup>b</sup> [ $\mu$ M] | Maximum inhibition % | PD <sub>50</sub> <sup>c</sup> [nMol kg <sup>-1</sup> ] |
|---------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------|----------------------|--------------------------------------------------------|
| $\alpha$ -Conotoxin TxIC                                            | $\alpha$ 3 $\beta$ 2, $\alpha$ 3 $\alpha$ 5 $\beta$ 2, $\alpha$ 3 $\beta$ 4                                                         | 100                                      | <5                   | 34.2                                                   |
|                                                                     | $\alpha$ 4 $\beta$ 2, $\alpha$ 4 $\beta$ 4, $\alpha$ 7, $\alpha$ $\beta$ $\gamma$ $\epsilon$ , $\alpha$ $\beta$ $\gamma$ $\epsilon$ | 50                                       | 0                    |                                                        |
| [Pro] <sup>2,8</sup> , [Glu] <sup>16</sup> $\alpha$ -Conotoxin TxIC | $\alpha$ 3 $\beta$ 2                                                                                                                | 5.4 $\pm$ 0.5                            | 96                   | 3600                                                   |
|                                                                     | $\alpha$ 3 $\alpha$ 5 $\beta$ 2                                                                                                     | 4.9 $\pm$ 0.8                            | 91                   |                                                        |
|                                                                     | $\alpha$ 3 $\beta$ 4                                                                                                                | 2.1 $\pm$ 0.2                            | 77                   |                                                        |
|                                                                     | $\alpha$ 4 $\beta$ 2, $\alpha$ 4 $\beta$ 4, $\alpha$ 7, $\alpha$ $\beta$ $\gamma$ $\epsilon$ , $\alpha$ $\beta$ $\gamma$ $\epsilon$ | $\geq$ 100                               | <10                  |                                                        |
|                                                                     |                                                                                                                                     |                                          |                      |                                                        |

<sup>a</sup> Human variants.

<sup>b</sup> Half-maximal inhibitory concentration (human).

<sup>c</sup> Paralytic dose (mollusk).

database (<http://www.conoserver.org>). Yet, such errors are highly possible given the relative abundance and differential incorporation of PTM amino acids in conotoxins/conopeptides [10].

The ability to detect the presence of  $\alpha$ -conotoxin TxIC only by its decarboxylated and fragmented products, and not its parent molecular mass, illustrates the susceptibility of the parent peptide to in-source decay. This process was exacerbated for  $\alpha$ -conotoxin TxIC by the presence of the PTM amino acid  $\gamma$ -carboxyglutamic acid (Gla). Similar observations have been described involving the loss of the PTM Tyr-SO<sub>4</sub> desulfation (–80 Da) in  $\alpha$ -conotoxins PnIA (*Conus pennaceus*) and Epl (*Conus episcopatus*) [32,53]. Although ESI-MS is less harsh, a similar degree of peptide degradation can occur if working outside normal operating parameters. Our observations highlight the necessary awareness that needs to be undertaken when dealing with complex and highly PTM processed DV extracts and native MVs.

Irrespective of MS technique, each method has advantages and disadvantages in the performance of mass detection and sequencing. If suspicion of PTMs arise, it is best to employ a combined approach for analysis and verification; this would include the use of molecular biology techniques along with more traditional biochemical methods such as Edman degradation. The latter technique becomes a necessity when considering the isobaric CID fragments generated from Hyp and Ile and Leu ( $\pm$ 0.04 Da). Three such unambiguous assignments were observed in the *de-novo* sequencing analysis of  $\alpha$ -conotoxin TxIC (Fig. 5B). Residues were later confirmed and assigned by Edman degradation (Section 3.8) and via genetic analysis, as shown in Fig. 6 (Section 3.9).

Comprising of 18 amino acids with two disulfide bonds, 44% of residues display one or more forms of modification,  $\alpha$ -conotoxin TxIC is one of the most highly PTM conopeptides documented to date. This novel 4/7 conopeptide (as represented by its cysteine framework) shares sequence homology with other  $\alpha$ -conotoxins [18,31] and demonstrates sequence commonality to other molluscivorous conopeptides, specifically Ai1.2 (*C. ammiralis*) and Vc1.1 (*C. victoriae*) (Table 3). This homology is seen explicitly in Vc1.1 in its native PTM form: [Hyp]<sup>6</sup>[Gla]<sup>14</sup>Vc1.1 [24], and more so within the last cysteine loop (89% homology; Table 3). Both *C. ammiralis* (Ai) and *C. victoriae* (Vc) are close relations of *C. textile*—with *C. victoriae* being endemic to Australia. Genetic analysis of  $\alpha$ -conotoxin TxIC strengthens its relationship to the established 4/7  $\alpha$ -conotoxins, as seen by the homology of the pre/pro-peptide regions of  $\alpha$ -conotoxins MII and Vc1.1 (Fig. 6) [17,33,39,44].

As the majority of these illustrated  $\alpha$ -conotoxin sequences are deduced from genomic sequences, the current inclusion of the observed PTMs at positions [Pro→Hyp]<sup>2</sup>, [Pro→Hyp]<sup>8</sup> and [Glu→Gla]<sup>16</sup> in Table 3 increases the known chemical diversity of  $\alpha$ -conotoxin TxIC. Modifications may be affecting *in vivo* characteristics including kinetic properties and/or pharmacological selectivity of conotoxins. The absence of these PTMs in piscivorous 4/7  $\alpha$ -conotoxins, such as  $\alpha$ -conotoxin MII (*Conus magus*),

potentially indicates a phyla-selective targeting differentiation. By examining both the native  $\alpha$ -conotoxin TxIC, as well as a synthetic unmodified isoform ([Pro]<sup>2,8</sup>[Glu]<sup>16</sup> $\alpha$ -conotoxin TxIC), we have here substantiated these claims, highlighting the role of PTMs *in vivo*. While remaining biologically active in whole animal bioassay, native  $\alpha$ -conotoxin TxIC is inactive when tested human nAChR channel isoforms (Table 4). The synthetic unmodified  $\alpha$ -conotoxin TxIC (i.e. [Pro]<sup>2,8</sup>[Glu]<sup>16</sup> $\alpha$ -conotoxin TxIC) however remains active at the nAChR, demonstrating isoform selectivity (Table 4). These features highlight the importance of PTMs in the selective targeting of  $\alpha$ -conotoxins. This is further illustrated at a phylogenetic level with the observed switching in potency in invertebrate models that represents the native prey target (Fig. S2).

The phyla-selectivity and reduced potency of the native  $\alpha$ -conotoxin TxIC towards human nAChR isoforms demonstrates an underlying differentiation between receptor selectivity and pharmacodynamic properties in ion channels from different phyla. This is an area reiterated by the PTM of the nAChR in the Egyptian mongoose (*Herpestes ichneumon*) that provides a level of resistance to  $\alpha$ -bungarotoxin [3]. Here, genomic and proteomic investigation into invertebrate nAChR isoforms, as well as other receptors (e.g. N-type (Ca<sub>v</sub>2.2) calcium channels and GABA<sub>B</sub> [13]), may provide insight into the determinants for pharmacological potency and phyla selective characteristics observed commonly with *Conus*. Such information will provide insights into toxin target specificity enhancement through peptide bioengineering and the potential manipulation/incorporation of PTMs.

## 5. Conclusions

The complexity of MV profiles from specimens of *C. textile* revealed in this study serve to remind us of the extent we have underestimated the biodiversity and value of these unique venomous marine snails. The analysis of venoms necessitates a specifically tailored approach, requiring special attention to the analytical methods employed, and further accentuating the need to employ a combined ‘conovenomic’ approach. The lessons learned and approaches outlined will facilitate peptide prospecting, toxicological correlations and the future discovery of new, clinically relevant conopeptides.

## Conflict of interest

Authors state that there is no conflict of interest.

## Acknowledgements

We are indebted to Mr. David Bingham, the members of the Malacological Society of Australasia (Brisbane Branch), Mr. Harold “JJ” Jackson (Hawai’ian Malacological Society) and Mr. Don Barclay

(American Samoa) for assistance in obtaining species described in this study. We wish to acknowledge the financial support from the American Heart Association (Scientist Development Award 0530204N to J.-P.B.) and the University of Hawai'i Sea Grant College Program (J.-P.B.) and USDA TSTAR (# 2009-34135-20067) (J.-P. B.) & HATCH (HAW00595-R) (J.-P.B.). This project in-part was supported by grants from the National Center for Research Resources (5 G12 RR003061-26) and the National Institute on Minority Health and Health Disparities (8 G12 MD007601-26) from the National Institutes of Health. JT was supported by the following grants: G.0433.12, G.A071.10N and G.0257.08 (F.W.O. Vlaanderen), EU-FP7-MAREX, A portion of this work was submitted to The University of Queensland, Australia, (J.-P.B.), The University of Melbourne, Australia (D.W.S.) and The University of Hawai'i, USA (Z.L.B.) as fulfillment for awarding a doctoral degrees.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.peptides.2013.09.004>.

## References

- Abdel-Rahman MA, Abdel-Nabi IM, El-Naggar MS, Abbas OA, Strong PN. Intraspecific variation in the venom of the vermivorous cone snail *Conus vexillum*. *Comp Biochem Physiol C Toxicol Pharmacol* 2011;154:318–25.
- Atherton D, Fernandez J, DeMott M, Andrews L, Mische SM. Routine protein sequence analysis below ten picomoles: one sequencing facility's approach. San Diego, California: Academic Press, Inc. (Harcourt Brace and Co.); 1993.
- Barchan D, Kachalsky S, Neumann D, Vogel Z, Ovadia M, Kochva E, et al. How the mongoose can fight the snake: the binding site of the mongoose acetylcholine receptor. *Proc Natl Acad Sci USA* 1992;89:7717–21.
- Bingham J-P. Novel toxins from the genus *Conus*- from taxonomy to toxins. Brisbane: University of Queensland; 1998.
- Bingham JP, Andrews EA, Kiyabu SM, Cabaltea CC. Drugs from Slugs. Part II - Conopeptide bioengineering. *Chem Biol Interact* 2012;200:92–113.
- Bingham JP, Broxton NM, Livett BG, Down JG, Jones A, Moczydlowski EG. Optimizing the connectivity in disulfide-rich peptides: alpha-conotoxin SII as a case study. *Anal Biochem* 2005;338:48–61.
- Bingham JP, Mitsunaga E, Bergeron ZL. Drugs from slugs—past, present and future perspectives of omega-conotoxin research. *Chem Biol Interact* 2010;183:1–18.
- Carstens BB, Clark RJ, Daly NL, Harvey PJ, Kaas Q, Craik DJ. Engineering of conotoxins for the treatment of pain. *Curr Pharm Des* 2011;17:4242–53.
- Chivian E, Roberts CM, Bernstein AS. The threat to cone snails. *Science* 2003;302:391.
- Chun JB, Baker MR, Kim DH, Leroy M, Toribo P, Bingham JP. Cone snail milked venom dynamics - a quantitative study of *Conus purpurascens*. *Toxicon* 2012;60:83–94.
- Conticello SG, Gilad Y, Avidan N, Ben-Asher E, Levy Z, Fainzilber M. Mechanisms for evolving hypervariability: the case of conopeptides. *Mol Biol Evol* 2001;18:120–31.
- Cruz LJ, Ramilo CA, Corpuz GP, Olivera BM. *Conus* peptides: phylogenetic range of biological activity. *Biol Bull* 1992;183:159–64.
- Cuny H, de Faoite A, Huynh TG, Yasuda T, Berecki G, Adams DJ.  $\gamma$ -Aminobutyric acid type B (GABAB) receptor expression is needed for inhibition of N-type (Cav2.2) calcium channels by analgesic  $\alpha$ -conotoxins. *J Biol Chem* 2012;287:23948–57.
- Czerwiec E, Kalume DE, Roepstorff P, Hambe B, Furie B, Furie BC, et al. Novel gamma-carboxyglutamic acid-containing peptides from the venom of *Conus textile*. *FEBS J* 2006;273:2779–88.
- Davis J, Jones A, Lewis RJ. Remarkable inter- and intra-species complexity of conotoxins revealed by LC/MS. *Peptides* 2009;30:1222–7.
- Duda TFJ, Bingham JP, Livett BG, Kohn AJ, Massilia GR, Schultz JR, et al. How much at risk are cone snails. *Science* 2004;303:955–7.
- Dutertre S, Jin AH, Kaas Q, Jones A, Alewood PF, Lewis RJ. Deep venomics reveals the mechanism for expanded peptide diversity in cone snail venom. *Mol Cell Proteomics* 2013;12:312–29.
- Dutertre S, Ulens C, Büttner R, Fish A, van Elk R, Kendel Y, et al. AChBP-targeted alpha-conotoxin correlates distinct binding orientations with nAChR subtype selectivity. *EMBO J* 2007;26:3858–67.
- Endean R, Rudkin C. Further studies of the venoms of *conidae*. *Toxicon* 1965;69:225–49.
- Essack M, Bajic VB, Archer JA. Conotoxins that confer therapeutic possibilities. *Mar Drugs* 2012;10:1244–65.
- Fainzilber M, Gordon D, Hasson A, Spira ME, Zlotkin E. Mollusc-specific toxins from the venom of *Conus textile neovicarius*. *Eur J Biochem* 1991;202:589–95.
- Halai R, Craik DJ. Conotoxins: natural product drug leads. *Nat Prod Rep* 2009;26:526–36.
- Hillyard DR, Olivera BM, Woodward S, Corpuz GP, Gray WR, Ramilo CA, et al. A molluscivorous *Conus* toxin: conserved frameworks in conotoxins. *Biochemistry* 1989;28:358–61.
- Jakubowski JA, Kelley WP, Sweedler JV. Screening for post-translational modifications in conotoxins using liquid chromatography/mass spectrometry: an important component of conotoxin discovery. *Toxicon* 2006;47:688–99.
- Jones J, Bingham JP, Gehrman J, Bond T, Loughnan M, Atkins A, et al. Isolation and characterization of conopeptides by high-performance liquid chromatography combined with mass spectrometry and tandem mass spectrometry. *Commun Mass Spectrom* 1996;10:138–43.
- Kapono CA, Thapa P, Cabaltea CC, Guendisch D, Collier AC, Bingham JP. Conotoxin truncation as a post-translational modification to increase the pharmacological diversity within the milked venom of *Conus magus*. *Toxicon* 2013;70:170–8.
- Kobayashi J, Ohizumi Y, Nakamura H, Hirata Y. Pharmacological study on the venom of the marine snail *Conus textile*. *Toxicon* 1981;19:757–62.
- Lev-Ram V, Olivera BM, Levitan IB. A toxin from the venom of the predator snail *Conus textile* modulates ionic currents in Aplysia bursting pacemaker neuron. *Brain Res* 1994;640:48–55.
- Liman ER, Tytgat J, Hess P. Subunit stoichiometry of a mammalian K<sup>+</sup> channel determined by construction of multimeric cDNAs. *Neuron* 1992;9:861–71.
- Livett BG, Gayler KR, Khalil Z. Drugs from the sea: conopeptides as potential therapeutics. *Curr Med Chem* 2004;11:1715–23.
- Livett BG, Sandall DW, Keays D, Down J, Gayler KR, Satkunanathan N, et al. Therapeutic applications of conotoxins that target the neuronal nicotinic acetylcholine receptor. *Toxicon* 2006;48:810–29.
- Loughnan M, Bond T, Atkins A, Cuevas J, Adams DJ, Broxton NM, et al. Alpha-conotoxin Epl, a novel sulfated peptide from *Conus episcopatus* that selectively targets neuronal nicotinic acetylcholine receptors. *J Biol Chem* 1998;273:15667–74.
- Lu BS, Yu F, Zhao D, Huang PT, Huang CF. Conopeptides from *Conus striatus* and *Conus textile* by cDNA cloning. *Peptides* 1999;20:1139–44.
- Matsudaire PT. Introduction A practical guide to protein and peptide purification for microsequencing. San Diego, California: Academic Press, Inc. (Harcourt Brace Jovanovich); 1989.
- McIntosh JM, Santos AD, Olivera BM. *Conus* peptides targeted to specific nicotinic acetylcholine receptor subtypes. *Annu Rev Biochem* 1999;68:59–88.
- Nakamura H, Kobayashi J, Ohizumi Y, Hirata Y. The occurrence of arachidonic acid in the venom duct of the marine snail *Conus textile*. *Experientia* 1982;38:897.
- Nakamura T, Yu Z, Fainzilber M, Burlingame AL. Mass spectrometric-based revision of the structure of a cysteine-rich peptide toxin with gamma-carboxyglutamic acid, TxVIIA, from the sea snail, *Conus textile*. *Protein Sci* 1996;5:524–30.
- Olivera BM, Rivier J, Clark C, Ramilo CA, Corpuz GP, Abogadie FC, et al. Diversity of *Conus* neuropeptides. *Science* 1990;249:257–63.
- Price-Carter M, Gray WR, Goldenberg DP. Folding of omega-conotoxins. 2. Influence of precursor sequences and protein disulfide isomerase. *Biochemistry* 1996;35:15547–57.
- Reed LJ, Muench S. A simple method of estimating fifty percent time points. *Am J Hyg* 1938;27:493–7.
- Rigby AC, Lucas-Meunier E, Kalume DE, Czerwiec E, Hambe B, Dahlqvist I, et al. A conotoxin from *Conus textile* with unusual posttranslational modifications reduces presynaptic Ca<sup>2+</sup> influx. *Proc Natl Acad Sci USA* 1999;96:5758–63.
- Rivera-Ortiz JA, Cano H, Marí F. Intraspecific variability and conopeptide profiling of the injected venom of *Conus ermineus*. *Peptides* 2011;32:306–16.
- Romeo C, Di Francesco L, Oliverio M, Palazzo P, Massilia GR, Ascenzi P, et al. *Conus ventricosus* venom peptides profiling by HPLC-MS: a new insight in the intraspecific variation. *J Sep Sci* 2008;31:488–98.
- Sandall DW, Satkunanathan N, Keays DA, Polidano MA, Liping X, Pham V, et al. A novel alpha-conotoxin identified by gene sequencing is active in suppressing the vascular response to selective stimulation of sensory nerves in vivo. *Biochemistry* 2003;42:6904–11.
- Sasaki T, Feng ZP, Scott R, Grigoriev N, Syed NI, Fainzilber M, et al. Synthesis, bioactivity, and cloning of the L-type calcium channel blocker omega-conotoxin TxVII. *Biochemistry* 1999;38:12876–84.
- Schnölzer M, Alewood P, Jones A, Alewood D, Kent SBH. In situ neutralization in Boc-chemistry solid phase peptide synthesis. *Int J Peptide Protein Res* 1992;40:180–93.
- Schroeder CI, Craik DJ. Therapeutic potential of conopeptides. *Future Med Chem* 2012;4:1243–55.
- Schroeder CI, Ekberg J, Nielsen KJ, Adams D, Loughnan ML, Thomas L, et al. Neuronally micro-conotoxins from *Conus striatus* utilize an alpha-helical motif to target mammalian sodium channels. *J Biol Chem* 2008;283:21621–8.
- Tayo LL, Lu B, Cruz LJ, Yates JR. Proteomic analysis provides insights on venom processing in *Conus textile*. *J Proteome Res* 2010;9:2292–301.
- Teichert RW, Rivier J, Dykert J, Cervini L, Gulyas J, Bulaj G, et al. AlphaA-Conotoxin OIVA defines a new alphaA-conotoxin subfamily of nicotinic acetylcholine receptor inhibitors. *Toxicon* 2004;44:207–14.
- Townsend A, Livett BG, Bingham JP, Truong HT, Karas JA, O'Donnell P, et al. Mass spectral identification of Vc1.1 and differential distribution of conopeptides in the venom duct of *Conus victoriae*. Effect of post-translational

- modifications and disulfide isomerisation on bioactivity. *Int J Peptide Res Therap* 2009;15:195–203.
- [52] Vincler M, Wittenauer S, Parker R, Ellison M, Olivera BM, McIntosh JM. Molecular mechanism for analgesia involving specific antagonism of alpha9alpha10 nicotinic acetylcholine receptors. *Proc Natl Acad Sci USA* 2006;103:17880–4.
- [53] Wolfender JL, Chu F, Ball H, Wolfender F, Fainzilber M, Baldwin MA, et al. Identification of tyrosine sulfation in *Conus pennaceus* conotoxins alpha-PnIA and alpha-PnIB: further investigation of labile sulfo- and phosphopeptides by electrospray, matrix-assisted laser desorption/ionization (MALDI) and atmospheric pressure MALDI mass spectrometry. *J Mass Spectrom* 1999;34:447–54.
- [54] Woodward SR, Cruz LJ, Olivera BM, Hillyard DR. Constant and hypervariable regions in conotoxin propeptides. *EMBO J* 1990;9:1015–20.
- [55] Xiao WH, Bennett GJ. Synthetic omega-conopeptides applied to the site of nerve injury suppress neuropathic pains in rats. *J Pharmacol Exp Ther* 1995;274:666–72.

## REFERENCES

- Alpeeva, I. S.; Niculescu-Nistor, M.; Leon, J. C.; Csöregi, E.; Sakharov, I. Yu. Palm tree peroxidase-based biosensor with unique characteristics for hydrogen peroxide monitoring. *Biosensors and Bioelectronics* **2005**, *21*, 742–748.
- Armougom, F.; Moretti, S.; Poirot, O.; Audic, S.; Dumas, P.; Schaeli, B.; Keduas, V.; Notredame, C. Espresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nucleic Acids Research*. **2006**, *34*, W604–W608.
- Arnold, K.; Bordoli, L.; Kopp, J.; Schwede, T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **2006**, *22*, 195–201.
- Baker, M. R.; Zhao, H.; Sakharov, I. Yu.; Li, Q. X. Amino acid sequence of anionic peroxidase from the windmill palm tree *Trachycarpus fortunei*. *Journal of Agricultural and Food Chemistry* **2014**, *62*, 11941–11948.
- Barba-Espin, G.; Dedvisitsakul, P.; Hagglund, P.; Svensson, B.; Finnie, C. Gibberellic acid-induced aleurone layers responding to heat shock or tunicamycin provide insight into the *N*-glycoproteome, protein secretion, and endoplasmic reticulum stress. *Plant Physiology* **2014**, *164*, 951–965.
- Bernardes, A.; Textor, L. C.; Santos, J. C.; Cuadrado, N. H.; Kostetsky, E. Y.; Roig, M. G.; Bavro, V. N.; Muniz, J. R. C.; Shnyrov, V. L.; Polikarpov, I. Crystal structure analysis of peroxidase from the palm tree *Chamaerops Excelsa*. *Biochimie* **2015**, *111*, 58–69.
- Bordoli, L.; Kiefer, F.; Arnold, K.; Benkert, P.; Battey, J.; Schwede, T. Protein structure homology modeling using SWISS-MODEL workspace. *Nature Protocols* **2008**, *4*, 1–13.
- Capone, S.; Pletzenauer, R.; Maresch, D.; Metzger, K.; Altmann, F.; Herwig, C.; Spadiut, O. Glyco-variant library of the versatile enzyme horseradish peroxidase. *Glycobiology* **2014**, *24*, 852–863.
- Caramyshev, A. V.; Evtushenko, E. G.; Ivanov, V. F.; Barceló, A. R.; Roig, M. G.; Shnyrov, V. L.; van Huystee, R. B.; Kurochkin, I. N.; Vorobiev, A. K.; Sakharov, I. Yu. Synthesis of conducting polyelectrolyte complexes of polyaniline and poly(2-acrylamido-3-methyl-1-propanesulfonic acid) catalyzed by pH-stable palm tree peroxidase. *Biomacromolecules* **2005**, *6*, 1360–1366.
- Caramyshev, A. V.; Firsova, Y. N.; Slastya, E. A.; Tagaev, A. A.; Potapenko, N. V.; Lobakova, E. S.; Pletjushkina, O. Y.; Sakharov, I. Yu. Purification and characterization of windmill palm tree (*Trachycarpus fortunei*) peroxidase. *Journal of Agricultural and Food Chemistry* **2006**, *54*, 9888–9894.

- Caramyshev, A. V.; Lobachov, V. M.; Selivanov, D. V.; Sheval, E. V.; Vorobiev, A. K.; Katasova, O. N.; Polyakov, V. Y.; Makarov, A. A.; Sakharov, I. Yu. Micellar peroxidase-catalyzed synthesis of chiral polyaniline. *Biomacromolecules* **2007**, *8*, 2549–2555.
- Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J. A cross-platform toolkit for mass spectrometry and proteomics. *Nature biotechnology* **2012**, *30*, 918–920.
- Chen, S.; Yuan, R.; Chai, Y.; Hu, F. Electrochemical sensing of hydrogen peroxide using metal nanoparticles: a review. *Microchimica Acta* **2013**, *180*, 15–32.
- Cooper, C. A.; Gasteiger, E.; Packer, N. H. GlycoMod—a software tool for determining glycosylation compositions from mass spectrometric data. *Proteomics* **2001**, *1*, 340–349.
- Damerell, D.; Ceroni, A.; Maass, K.; Ranzinger, R.; Dell, A.; Haslam, S. M. The GlycanBuilder and GlycoWorkbench glycoinformatics tools: updates and new developments. *Biological Chemistry* **2012**, 393.
- Dasari, S.; Chambers, M. C.; Slebos, R. J.; Zimmerman, L. J.; Ham, A.-J. L.; Tabb, D. L. TagRecon: high-throughput mutation identification through sequence tagging. *Journal of Proteome Research* **2010**, *9*, 1716–1726.
- Desaire, H. Glycopeptide analysis, recent developments and applications. *Molecular & Cellular Proteomics* **2013**, *12*, 893–901.
- Deutsch, E. W.; Lam, H.; Aebersold, R. Data analysis and bioinformatics tools for tandem mass spectrometry in proteomics. *Physiological Genomics* **2008**, *33*, 18–25.
- Domon, B.; Costello, C. E. A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. *Glycoconjugate Journal* **1988**, *5*, 397–409.
- Elias, J. E.; Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nature Methods* **2007**, *4*, 207–214.
- Fanata, W. I. D.; Lee, K. H.; Son, B. H.; Yoo, J. Y.; Harmoko, R.; Ko, K. S.; Ramasamy, N. K.; Kim, K. H.; Oh, D.-B.; Jung, H. S.; Lee, K. O. N-Glycan maturation is crucial for cytokinin-mediated development and cellulose synthesis in *Oryza sativa*. *The Plant Journal* **2013**, *73*, 966–979.
- Farré, M.; Kantiani, L.; Barceló, D. Advances in immunochemical technologies for analysis of organic pollutants in the environment. *TrAC Trends in Analytical Chemistry* **2007**, *26*, 1100–1112.
- Gajhede, M.; Schuller, D. J.; Henriksen, A.; Smith, A. T.; Poulos, T. L. Crystal structure of horseradish peroxidase C at 2.15 Å resolution. *Nature Structural Biology* **1997**, *4*, 1032–1048.

- Gaspar, S.; Popescu, I. C.; Gazaryan, I. G.; Gerardo Bautista, A.; Sakharov, I. Yu.; Mattiasson, B.; Csöregi, E. Biosensors based on novel plant peroxidases: a comparative study. *Electrochimica acta* **2000**, *46*, 255–264.
- Gasteiger, E.; Hoogland, C.; Gattiker, A.; Wilkins, M. R.; Appel, R. D.; Bairoch, A. Protein identification and analysis tools on the ExPASy server. In *The proteomics protocols handbook*; Springer, 2005; pp 571–607.
- Gomord, V.; Fitchette, A.-C.; Menu-Bouaouiche, L.; Saint-Jore-Dupas, C.; Plasson, C.; Michaud, D.; Faye, L. Plant-specific glycosylation patterns in the context of therapeutic protein production: PMP-specific glycosylation patterns. *Plant Biotechnology Journal* **2010**, *8*, 564–587.
- Gray, J. S. S.; Montgomery, R. Asymmetric glycosylation of soybean seed coat peroxidase. *Carbohydrate Research* **2006**, *341*, 198–209.
- Hanisch, F.-G. Top-down sequencing of *O*-glycoproteins by in-source decay Matrix-Assisted Laser Desorption Ionization Mass Spectrometry for glycosylation site analysis. *Analytical Chemistry* **2011**, *83*, 4829–4837.
- Hart-Smith, G.; Raftery, M. J. Detection and characterization of low abundance glycopeptides via higher-energy C-trap dissociation and orbitrap mass analysis. *Journal of The American Society for Mass Spectrometry* **2012**, *23*, 124–140.
- Henriksen, A.; Welinder, K. G.; Gajhede, M. Structure of barley grain peroxidase refined at 1.9-Å resolution: a plant peroxidase reversibly inactivated at neutral pH. *Journal of Biological Chemistry* **1998**, *273*, 2241–2248.
- Henriksen, A.; Smith, A. T.; Gajhede, M. The structures of the horseradish peroxidase c-ferulic acid complex and the ternary complex with cyanide suggest how peroxidases oxidize small phenolic substrates. *Journal of Biological Chemistry* **1999**, *274*, 35005–35011.
- Henriksen, A.; Mirza, O.; Indiani, C.; Teilum, K.; Smulevich, G.; Welinder, K. G.; Gajhede, M. Structure of soybean seed coat peroxidase: a plant peroxidase with unusual stability and haem-apoprotein interactions. *Protein Science* **2001**, *10*, 108–115.
- Hu, C.; Lee, D.; Chibbar, R. N.; Huystee, R. B. Ca<sup>2+</sup> and peroxidase derived from cultured peanut cells. *Physiologia Plantarum* **1987**, *70*, 99–102.
- Huddleston, M. J.; Bean, M. F.; Carr, S. A. Collisional fragmentation of glycopeptides by electrospray ionization LC/MS and LC/MS/MS: methods for selective detection of glycopeptides in protein digests. *Analytical Chemistry* **1993**, *65*, 877–884.

- Johnson, K. D.; Chrispeels, M. J. Substrate specificities of *N*-acetylglucosaminyl-, fucosyl-, and xylosyltransferases that modify glycoproteins in the Golgi apparatus of bean cotyledons. *Plant Physiology* **1987**, *84*, 1301–1308.
- Kang, J. S.; Frank, J.; Kang, C. H.; Kajiura, H.; Vikram, M.; Ueda, A.; Kim, S.; Bahk, J. D.; Triplett, B.; Fujiyama, K.; Koiwa, H. Salt tolerance of *Arabidopsis thaliana* requires maturation of *N*-glycosylated proteins in the Golgi apparatus. *Proceedings of the National Academy of Sciences* **2008**, *105*, 5933–5938.
- Klibanov, A. M.; Tu, T.-M.; Scott, K. P. Peroxidase-catalyzed removal of phenols from coal-conversion waste waters. *Science* **1983**, *221*, 259–261.
- Kolarich, D.; Jensen, P. H.; Altmann, F.; Packer, N. H. Determination of site-specific glycan heterogeneity on glycoproteins. *Nature Protocols* **2012**, *7*, 1285–1298.
- Krainer, F. W.; Pletzenauer, R.; Rossetti, L.; Herwig, C.; Glieder, A.; Spadiut, O. Purification and basic biochemical characterization of 19 recombinant plant peroxidase isoenzymes produced in *Pichia pastoris*. *Protein Expression and Purification* **2014**, *95*, 104–112.
- Léonard, R.; Costa, G.; Darrambide, E.; Lhernould, S.; Fleurat-Lessard, P.; Carlué, M.; Gomord, V.; Faye, L.; Maftah, A. The presence of Lewis a epitopes in *Arabidopsis thaliana* glycoconjugates depends on an active  $\alpha$ 4-fucosyltransferase gene. *Glycobiology* **2002**, *12*, 299–306.
- Léonard, R.; Strasser, R.; Altmann, F. Plant glycosidases acting on protein-linked oligosaccharides. *Phytochemistry* **2009**, *70*, 318–324.
- Leuzinger, K.; Dent, M.; Hurtado, J.; Stahnke, J.; Lai, H.; Zhou, X.; Chen, Q. Efficient agroinfiltration of plants for high-level transient expression of recombinant proteins. *Journal of Visualized Experiments* **2013**.
- Leymarie, N.; Griffin, P. J.; Jonscher, K.; Kolarich, D.; Orlando, R.; McComb, M.; Zaia, J.; Aguilan, J.; Alley, W. R.; Altmann, F. Interlaboratory study on differential analysis of protein glycosylation by mass spectrometry: the ABRF glycoprotein research multi-institutional study 2012. *Molecular & Cellular Proteomics* **2013**, *12*, 2935–2951.
- Li, F.; Glinskii, O. V.; Glinsky, V. V. Glycobioinformatics: current strategies and tools for data mining in MS-based glycoproteomics. *Proteomics* **2013**, *13*, 341–354.
- Lige, B.; Ma, S.; van Huystee, R. B. The effects of the site-directed removal of *N*-glycosylation from cationic peanut peroxidase on its function. *Archives of Biochemistry and Biophysics* **2001**, *386*, 17–24.
- Litescu, S. C.; Eremia, S.; Radu, G. L. Biosensors for the determination of phenolic metabolites. In *Bio-Farms for Nutraceuticals*; Giardi, M.T.; Rea, G.; Berra, B., Eds.; Springer, 2010; pp. 234–240.

- Livingstone, C. D.; Barton, G. J. Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. *Computer applications in the biosciences: CABIOS* **1993**, *9*, 745–756.
- Nanni, P.; Panse, C.; Gehrig, P.; Mueller, S.; Grossmann, J.; Schlapbach, R. PTM MarkerFinder, a software tool to detect and validate spectra from peptides carrying post-translational modifications. *Proteomics* **2013**, *13*, 2251–2255.
- Nielsen, K. L.; Indiani, C.; Henriksen, A.; Feis, A.; Becucci, M.; Gajhede, M.; Smulevich, G.; Welinder, K. G. Differential activity and structure of highly similar peroxidases. Spectroscopic, crystallographic, and enzymatic analyses of lignifying *Arabidopsis thaliana* peroxidase A2 and horseradish peroxidase A2. *Biochemistry* **2001**, *40*, 11013–11021.
- Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M. Higher-energy C-trap dissociation for peptide modification analysis. *Nature Methods* **2007**, *4*, 709–712.
- Ozohanics, O.; Turiák, L.; Puerta, A.; Vékey, K.; Drahos, L. High-performance liquid chromatography coupled to mass spectrometry methodology for analyzing site-specific *N*-glycosylation patterns. *Journal of Chromatography A* **2012**, *1259*, 200–212.
- Paizs, B.; Suhai, S. Fragmentation pathways of protonated peptides. *Mass Spectrometry Reviews* **2005**, *24*, 508–548.
- Palm, G. J.; Sharma, A.; Kumari, M.; Panjikar, S.; Albrecht, D.; Jagannadham, M. V.; Hinrichs, W. Post-translational modification and extended glycosylation pattern of a plant latex peroxidase of native source characterized by X-ray crystallography. *FEBS Journal* **2014**, *281*, 4319–4333.
- Petrescu, A.-J. Statistical analysis of the protein environment of *N*-glycosylation sites: implications for occupancy, structure, and folding. *Glycobiology* **2004**, *14*, 103–114.
- Rayon, C.; Lerouge, P.; Faye, L. The protein *N*-glycosylation in plants. *Journal of Experimental Botany* **1998**, *49*, 1463–1472.
- Regalado, C.; García-Almendárez, B. E.; Duarte-Vázquez, M. A. Biotechnological applications of peroxidases. *Phytochemistry Reviews* **2004**, *3*, 243–256.
- Resemann, A.; Wunderlich, D.; Rothbauer, U.; Warscheid, B.; Leonhardt, H.; Fuchser, J.; Kuhlmann, K.; Suckau, D. Top-down *de novo* protein sequencing of a 13.6 kDa camelid single heavy chain antibody by Matrix-Assisted Laser Desorption Ionization-Time-of-Flight/Time-of-Flight Mass Spectrometry. *Analytical Chemistry* **2010**, *82*, 3283–3292.
- Rodriguez-Lopez, J. N.; Smith, A. T.; Thorneley, R. N. F. Role of arginine 38 in horseradish peroxidase: a critical residue for substrate binding and catalysis. *Journal of Biological Chemistry* **1996**, *271*, 4023–4030.

- Roepstorff, P.; Fohlmann, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomedical Mass Spectrometry* **1984**, *11*, 601.
- Ruiz-May, E.; Thannhauser, T. W.; Zhang, S.; Rose, J. K. C. Analytical technologies for identification and characterization of the plant *N*-glycoproteome. *Frontiers in Plant Science* **2012**, *3*.
- Ruiz-May, E.; Kim, S.-J.; Brandizzi, F.; Rose, J. K. C. The secreted plant *N*-glycoproteome and associated secretory pathways. *Frontiers in Plant Science* **2012**, *3*.
- Ruiz-May, E.; Hucko, S.; Howe, K. J.; Zhang, S.; Sherwood, R. W.; Thannhauser, T. W.; Rose, J. K. A comparative study of lectin affinity based plant *N*-glycoproteome profiling using tomato fruit as a model. *Molecular & Cellular Proteomics* **2014**, *13*, 566–579.
- Sakharov, I. Yu. Palm tree peroxidases. *Biochemistry Moscow* **2004**, *69*, 823–829.
- Sakharov, I. Yu.; Vesga B, M. K.; Galaev, I. Yu.; Sakharova, I. V.; Pletjushkina, O. Y. Peroxidase from leaves of royal palm tree *Roystonea regia*: purification and some properties. *Plant Science* **2001**, *161*, 853–860
- Sakharov, I. Yu.; Vorobiev, A. C.; Leon, J. J. C. Synthesis of polyelectrolyte complexes of polyaniline and sulfonated polystyrene by palm tree peroxidase. *Enzyme and Microbial Technology* **2003**, *33*, 661–667.
- Sánchez-Romero, C.; García-Gómez, M. L.; Pliego-Alfaro, F.; Heredia, A. Effect of partial deglycosylation on catalytic characteristics and stability of an avocado peroxidase. *Physiologia Plantarum* **1994**, *92*, 97–101.
- Schuller, D. J.; Ban, N.; van Huystee, R. B.; McPherson, A.; Poulos, T. L. The crystal structure of peanut peroxidase. *Structure* **1996**, *4*, 311–321.
- Segrest, J. P.; Jackson, R. L.; Andrews, E. P.; Marchesi, V. T. Human erythrocyte membrane glycoprotein: a re-evaluation of the molecular weight as determined by SDS polyacrylamide gel electrophoresis. *Biochemical and Biophysical Research Communications* **1971**, *44*, 390–395.
- Segu, Z. M.; Mechref, Y. Characterizing protein glycosylation sites through higher-energy C-trap dissociation. *Rapid Communications in Mass Spectrometry* **2010**, *24*, 1217–1225.
- Shindyalov, I. N.; Bourne, P. E. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Engineering* **1998**, *11*, 739–747.
- Smulevich, G.; Feis, A.; Howes, B. D. Fifteen years of raman spectroscopy of engineered heme containing peroxidases: what have we learned? *Accounts of Chemical Research* **2005**, *38*, 433–440.

- Song, W.; Henquet, M. G. L.; Mentink, R. A.; van Dijk, A. J.; Cordewener, J. H. G.; Bosch, D.; America, A. H. P.; van der Krol, A. R. *N*-Glycoproteomics in plants: perspectives and challenges. *Journal of Proteomics* **2011**, *74*, 1463–1474.
- Stanley, P.; Schachter, H.; Taniguchi, N. Chapter 8 *N*-Glycans. In *Essentials of Glycobiology*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2009.
- Strasser, R.; Steinkellner, H.; Boren, M.; Altmann, F.; Mach, L.; Glössl, J.; Mucha, J. Molecular cloning of cDNA encoding *N*-acetylglucosaminyltransferase II from *Arabidopsis thaliana*. *Glycoconjugate Journal* **1999**, *16*, 787–791.
- Strasser, R.; Schoberer, J.; Jin, C.; Glössl, J.; Mach, L.; Steinkellner, H. Molecular cloning and characterization of *Arabidopsis thaliana* Golgi  $\alpha$ -mannosidase II, a key enzyme in the formation of complex *N*-glycans in plants. *The Plant Journal* **2006**, *45*, 789–803.
- Strasser, R.; Bondili, J. S.; Schoberer, J.; Svoboda, B.; Liebminger, E.; Gloszl, J.; Altmann, F.; Steinkellner, H.; Mach, L. Enzymatic properties and subcellular localization of *Arabidopsis*  $\beta$ -*N*-acetylhexosaminidases. *Plant Physiology* **2007**, *145*, 5–16.
- Strasser, R.; Bondili, J. S.; Vavra, U.; Schoberer, J.; Svoboda, B.; Gloszl, J.; Leonard, R.; Stadlmann, J.; Altmann, F.; Steinkellner, H.; Mach, L. A unique 1,3-galactosyltransferase is indispensable for the biosynthesis of *N*-glycans containing Lewis a structures in *Arabidopsis thaliana*. *The Plant Cell Online* **2007**, *19*, 2278–2292.
- Strasser, R. Biological significance of complex *N*-glycans in plants and their impact on plant physiology. *Frontiers in Plant Science* **2014**, *5*.
- Suckau, D.; Resemann, A.; Schuerenberg, M.; Hufnagel, P.; Franzen, J.; Holle, A. A novel MALDI LIFT-TOF/TOF mass spectrometer for proteomics. *Analytical and Bioanalytical Chemistry* **2003**, *376*, 952–965.
- Sultana, N.; Florance, H. V.; Johns, A.; Smirnoff, N. Ascorbate deficiency influences the leaf cell wall glycoproteome in *Arabidopsis thaliana*: ascorbate deficiency and cell wall glycoproteome. *Plant, Cell & Environment* **2015**, *38*, 375–384.
- Szigeti, K.; Smeller, L.; Osváth, S.; Majer, Z.; Fidy, J. The structure of horseradish peroxidase C characterized as a molten globule state after  $\text{Ca}^{2+}$  depletion. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2008**, *1784*, 1965–1974.
- Szumilo, T.; Kaushal, G. P.; Hori, H.; Elbein, A. D. Purification and properties of a glycoprotein processing  $\alpha$ -mannosidase from mung bean seedlings. *Plant physiology* **1986**, *81*, 383–389.
- Tabb, D. L.; Fernando, C. G.; Chambers, M. C. MyriMatch: highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis. *Journal of Proteome Research* **2007**, *6*, 654–661.

- Takahama, U. Oxidation of vacuolar and apoplastic phenolic substrates by peroxidase: physiological significance of the oxidation reactions. *Phytochemistry Reviews* **2004**, *3*, 207–219.
- Tams, J. W.; Welinder, K. G. Glycosylation and thermodynamic versus kinetic stability of horseradish peroxidase. *FEBS letters* **1998**, *421*, 234–236.
- Tamura, K.; Peterson, D.; Peterson, N.; Stecher, G.; Nei, M.; Kumar, S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* **2011**, *28*, 2731–2739.
- Tezuka, K.; Hayashi, M.; Ishihara, H.; Akazawa, T.; Takahashi, N. Studies on synthetic pathway of xylose-containing *N*-linked oligosaccharides deduced from substrate specificities of the processing enzymes in sycamore cells (*Acer pseudoplatanus L.*). *European Journal of Biochemistry* **1992**, *203*, 401–413.
- Veitch, N. C. Aromatic donor molecule binding sites of haem peroxidases. *Biochemical Society Transactions* **1995**, *23*, 232–240.
- Veitch, N. C. Structural determinants of plant peroxidase function. *Phytochemistry Reviews* **2004**, *3*, 3–18.
- von Schaewen, A.; Sturm, A.; O'Neill, J.; Chrispeels, M. J. Isolation of a mutant *Arabidopsis* plant that lacks *N*-acetyl glucosaminyl transferase I and is unable to synthesize Golgi-modified complex *N*-linked glycans. *Plant physiology* **1993**, *102*, 1109–1118.
- Watanabe, L.; de Moura, P. R.; Bleicher, L.; Nascimento, A. S.; Zamorano, L. S.; Calvete, J. J.; Sanz, L.; Pérez, A.; Bursakov, S.; Roig, M. G. Crystal structure and statistical coupling analysis of highly glycosylated peroxidase from royal palm tree (*Roystonea regia*). *Journal of Structural Biology* **2010**, *169*, 226–242.
- Waterhouse, A. M.; Procter, J. B.; Martin, D. M. A.; Clamp, M.; Barton, G. J. Jalview version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **2009**, *25*, 1189–1191.
- Welinder, K. G.; Justesen, A. F.; Kjaersgård, I. V. H.; Jensen, R. B.; Rasmussen, S. K.; Jespersen, H. M.; Duroux, L. Structural diversity and transcription of class III peroxidases from *Arabidopsis thaliana*: 73 peroxidases from *Arabidopsis*. *European Journal of Biochemistry* **2002**, *269*, 6063–6081.
- Wilson, I. B.; Zeleny, R.; Kolarich, D.; Staudacher, E.; Stroop, C. J.; Kamerling, J. P.; Altmann, F. Analysis of Asn-linked glycans from vegetable foodstuffs: widespread occurrence of Lewis a, core  $\alpha$ 1,3-linked fucose and xylose substitutions. *Glycobiology* **2001**, *11*, 261–274.

- Wormald, M. R.; Petrescu, A. J.; Pao, Y.-L.; Glithero, A.; Elliott, T.; Dwek, R. A. Conformational studies of oligosaccharides and glycopeptides: complementarity of NMR, X-ray crystallography, and molecular modeling. *Chemical Reviews* **2002**, *102*, 371–386.
- Wyss, D. F.; Dayie, K. T.; Wagner, G. The counter receptor binding site of human CD2 exhibits an extended surface patch with multiple conformations fluctuating with millisecond to microsecond motions. *Protein Science* **1997**, *6*, 534–542.
- Yang, B. Y.; Gray, J. S.; Montgomery, R. The glycans of horseradish peroxidase. *Carbohydrate Research* **1996**, *287*, 203–212.
- Zamorano, L. S.; Pina, D. G.; Arellano, J. B.; Bursakov, S. A.; Zhadan, A. P.; Calvete, J. J.; Sanz, L.; Nielsen, P. R.; Villar, E.; Gavel, O.; et al. Thermodynamic characterization of the palm tree *Roystonea Regia* peroxidase stability. *Biochimie* **2008**, *90*, 1737–1749.
- Zamorano, L. S.; Vilarmau, S. B.; Arellano, J. B.; Zhadan, G. G.; Cuadrado, N. H.; Bursakov, S. A.; Roig, M. G.; Shnyrov, V. L. Thermal stability of peroxidase from *Chamaerops excelsa* palm tree at pH 3. *International Journal of Biological Macromolecules* **2009**, *44*, 326–332.
- Zhang, B.; Chambers, M. C.; Tabb, D. L. Proteomic parsimony through bipartite graph analysis improves accuracy and transparency. *Journal of Proteome Research* **2007**, *6*, 3549–3557.
- Zhang, C.; Doherty-Kirby, A.; Huystee, R. van; Lajoie, G. Investigation of cationic peanut peroxidase glycans by electrospray ionization mass spectrometry. *Phytochemistry* **2004**, *65*, 1575–1588.
- Zhang, H.; Lu, Y.; Ushio, H.; Shiomi, K. Development of sandwich ELISA for detection and quantification of invertebrate major allergen tropomyosin by a monoclonal antibody. *Food Chemistry* **2014**, *150*, 151–157.
- Zhang, M.; Henquet, M.; Chen, Z.; Zhang, H.; Zhang, Y.; Ren, X.; Van Der Krol, S.; Gonneau, M.; Bosch, D.; Gong, Z. LEW3, Encoding a Putative  $\alpha$ -1,2-mannosyltransferase (ALG11) in *N*-linked glycoprotein, plays vital roles in cell-wall biosynthesis and the abiotic stress response in *Arabidopsis thaliana*: protein *N*-glycosylation, plant development and abiotic stress. *The Plant Journal* **2009**, *60*, 983–999.
- Zhang, Y.; Giboulot, A.; Zivy, M.; Valot, B.; Jamet, E.; Albenne, C. Combining various strategies to increase the coverage of the plant cell wall glycoproteome. *Phytochemistry* **2011**, *72*, 1109–1123.

- Zhao, H.; Nan, T.; Tan, G.; Gao, W.; Cao, Z.; Sun, S.; Li, Z.; Li, Q. X.; Wang, B. Development of two highly sensitive immunoassays for detection of copper ions and a suite of relevant immunochemicals. *Analytica Chimica Acta* **2011**, *702*, 102–108.
- Zipor, G.; Duarte, P.; Carqueijeiro, I.; Shahr, L.; Ovadia, R.; Teper-Bamnlker, P.; Eshel, D.; Levin, Y.; Doron-Faigenboim, A.; Sottomayor, M.; Oren-Shamir, M. *In planta* anthocyanin degradation by a vacuolar class III peroxidase in *Brunfelsia calycina* flowers. *New Phytologist* **2015**, *205*, 653–665.
- Zipor, G.; Oren-Shamir, M. Do vacuolar peroxidases act as plant caretakers? *Plant Science* **2013**, *199*, 41–47.