DISULFIDE BOND AND TOPOLOGICAL ISOMERIZATION OF THE CONOPEPTIDE PNID:
DISULFIDE BONDS WITH A TWIST

A DISSERTATION SUMBITTED TO THE GRADUATE DIVISON 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

2017

By

MICHAEL J. ESPIRITU

Dissertation Committee:

Jon-Paul Bingham, Chairperson
Qing Li
Yong Soo Kim
Dulal Borthakur
C. Alan Titchenal

Keywords: Peptides, Conotoxins, Isomers, Topology, Nuclear Magnetic Resonance

Dedication

This dissertation is dedicated to my family and friends that have encouraged and supported me during my time in graduate school. I would like to first dedicate this dissertation to my parents Joseph and Sandra Espiritu for providing me with residence and financial support for a much longer time than you needed to. Without the love and support that you have provided me I would not be able to pursue this endeavor. I would also like to dedicate this dissertation to my sister Leah and her family for continually providing me with advice, great company and love. Finally this dissertation is also dedicated to my best friend and the love of my life Joycelyn Chun. Thank you for all of the instances where you listened to my talks, helped me edit articles, and provided me with advice and encouragement. Thank you for all of the time that you sacrificed to help make my dreams come true. Without any of you none of this would have been accomplished.

Acknowledgements

List of Publications

Materials in this dissertation have been adapted from the following published and non-published (In Preparation) manuscripts.

Published Papers:

1. A 21st-century approach to age-old problems: the ascension of biologics in clinical therapeutics
   MJ Espiritu, AC Collier, JP Bingham
   Drug discovery today 19 (8), 1109-1113

2. The emergence of cyclic peptides: the potential of bioengineered peptide drugs
   P Thapa, MJ Espiritu, C Cabalteja, JP Bingham
   International Journal of Peptide Research and Therapeutics 20 (4), 545-551

3. Incorporation of post-translational modified amino acids as an approach to increase both chemical and biological diversity of conotoxins and conopeptides
   MJ Espiritu, CC Cabalteja, CK Sugai, JP Bingham
   Amino acids 46 (1), 125-151

4. Conotoxins and their regulatory considerations
   P Thapa, MJ Espiritu, CC Cabalteja, JP Bingham
   Regulatory Toxicology and Pharmacology 70 (1), 197-202

5. t-Boc synthesis of Huwentoxin-I through Native Chemical Ligation incorporating a Trifluoromethanesulfonic acid cleavage strategy
   P Thapa, CC Cabalteja, EE Philips, MJ Espiritu, S Peigneur, BG Mille, Jan Tytgat, Theodore R Cummins, Jon-Paul Bingham, Biopolymers 2016 106 (**5**), 737-745

Papers in Preparation:

1. Cyclotide Paper
   RY Zhang, P Thapa, MJ Espiritu, V Menon and JP Bingham, Bioorganic and Medicinal Chemistry

2. Disulfide and topological isomerization of the χ-conotoxin PnID demonstrates differential selectivity for the monoamine transporters
   MJ Espiritu, CK Sugai, P Thapa, JP Bingham

3. MIIIC, a novel conotoxin containing a unique methionine-π interaction
   Ray Zhang, MJ Espiritu, CK Sugai, P Thapa, and JP Bingham

Abstract:

Despite decades of research, efficient and accurate peptide and protein folding remains to be a problematic feat. This is largely due to difficulty in determining ubiquitous folding characteristics among diverse peptide sequences and in establishing an overarching theme in the folding process. Thus it is critical to the advancement of knowledge on the topic that novel sequences and frameworks of peptides continue to be studied. Conotoxins have proven to be excellent subjects of study as they are small, bioactive, and their activity is often highly dependent on achieving an appropriate fold. χ/λ-Conotoxins have garnered significant interest due to their selective and inhibitory properties at the norepinephrine transporter (NET) (1,2). Here we present a novel χ/λ-conotoxin, PnID, which unlike the previously reported members of this class, is expressed as two native disulfide isomers within the venom duct extract of *Conus pennaceus*. Additionally, synthetic production of this peptide resulted in one ribbon isomer and two distinct topological globular isomers (named PnID Isomer B and PnID Isomer RevB), in which their formation was later found to be individually controlled by the sequential order in which their two disulfide bonds were selectively formed. Only one of the globular isomers, PnID B, demonstrated a small amount activity at the monoamine transporters. The ribbon isomer, PnID A showed inhibition of the NET, with and IC_{50} of 10 µM. A second globular isomer (with reversed topology) PnID RevB, was separately produced and demonstrated no comparative inhibition. NMR analysis and structure calculations revealed that the occurrence of these novel topological isomers maybe due to a difference in two properties: (1) disulfide bond orientation, with retention of same connectivity and (2) backbone flexibility. A critical turn between residues 5-8 within PnID appears to be largely affected by disulfide bond formation, where the more active topology, PnID Isomer B possesses an increased chance of forming a γ-turn, similar to the turn seen in χ-conotoxin MrIA, whereas the inactive isomer, PnID Isomer RevB, appears to have a higher tendency to form an α-turn. The implication that disulfide bond connectivity alone may not be enough information for the efficient synthetic reproduction of all *Conus* derived toxins is novel and wide impacting; this then requires closer scrutiny to the physical orientation of disulfide bonds in respect to each other. These findings implicate how pharmacological selectivity and/or specificity may be influenced by the synthetic strategy in sequential selective disulfide bond formation. We believe these findings could have implications on the future of synthetically produced and rationally designed peptide drugs to enhance receptor-targeting selectivity and/or specificity.

Table of Contents

List of Figures

Figure 1: A comparison between the structures of two "biologic" based drugs, Prialt® (left) an FDA approved analgesic peptide, and trastuzumab (right) a monoclonal antibody used to treat breast cancer.

Figure 2: *Conus textile*, one of the many venomous cone snail species capable of producing a cocktail of bioactive peptides.

Figure 3: A comparison between the two prominent forms of secondary structure seen in conotoxins; an alpha helix as shown in α-MII (3) (left) and antiparallel β sheets as shown in χ-MrIA (4) (right). Both structures were obtained from the World Wide Protein Databank (wwPDB) and were visualized using YASARA.

Figure 4: The three different possible disulfide configurations of a peptide containing four cysteines. PnID, the novel peptide described here is used as an example.

Figure 5: RP-HPLC/UV chromatogram of crude PnID A containing two orthogonally protected Acm cysteines and two cysteines with free thiols.

Figure 6: RP-HPLC/UV chromatogram of the fully folded and purified form of PnID A.

Figure 7: Mass spectrograph of the fully folded PnID A containing no orthogonal protecting groups. 660.3 was the observed mass to charge ratio (*m/z*) in the second charge state. The expected *m/z* was 660 +/- 0.5.

Figure 8: RP-HPLC/UV chromatograms of the isomers of PnID. 2A shows the products of the pre-purified air oxidation, 2B shows the attempted co-elution of PnID A with the air oxidation products, 2C shows the attempted co-elution of PnID B with the air oxidation products, 2D shows the attempted co-elution of PnID C with the air oxidation products, 2E shows the attempted co-elution of PnID RevA with the air oxidation products, and 2F shows the attempted co-elution of PnID RevB with the air oxidation products. * indicates position of the mixed peak. 1. indicates which disulfide was formed first.

Figure 9: Air oxidations of PnID under various buffering conditions.

Figure 10: Air oxidations of PnID in 0.1M ammonium bicarbonate (pH 8) ranging from 24 hours incubation to 96 hours at room temperature.

Figure 11: Air oxidation of PnID in 0.1M ammonium bicarbonate (pH 8) in 2:1 (GSH/GSSG) after 24 hours.

Figure 12: Air oxidations of CMrVIA under various buffering conditions.

Figure 13: Air oxidation of CMrVIA in 0.1M ammonium bicarbonate (pH 8) in 2:1 (GSH/GSSG) after 24 hours.

Figure 14: Air oxidations of MrIA under various buffering conditions.

Figure 15: Air oxidation of MrIA in 0.1M ammonium bicarbonate (pH 8) in 2:1 (GSH/GSSG) after 24 hours.

Figure 16: Comparison of the lowest 20 solution structures of the sequentially truncated χ-conotoxins MrIA, MrIB, and CmrVIA. PDB files were obtained from wwPDB.org and viewed in YASARA.

Figure 17: Percentage of total mortality of *P*. *canaliculata* under three varying concentrations for each synthetic PnID isomer tested via immersion method. NC – negative control; PC – positive control. Correction of the mortality data from the isomers was not performed since the result obtained from the negative control was not greater than 10%. Credit to Batinga and Baoanan (UP).

Figure 18: Percentage of total mortality of *J*. *asperata* under three varying concentrations used for each synthetic PnID isomer tested via immersion method. NC – negative control; PC – positive control. Credit to Batinga and Baoanan (UP).

Figure 19: Percentage mortality of *D*. *magna* under five varying concentrations used for each synthetic PnID isomer tested in the biotoxicity assay: NC – negative control; PC – positive control.

Figure 20: Uptake plots showing the comparative dose-response inhibition of the monoamine transporters by PnID A, PnID B, PnID RevB, and cocaine at concentrations up to 10 μM.

Figure 21: Uptake plots showing the comparative dose-response inhibition of the monoamine transporters by PnID A, PnID B, PnID C and cocaine at concentrations up to 100 μM.

Figure 22: A 500 MHz NMR spectrometer used at the University of Hawaii at Manoa. Photo courtesy of the UH Manoa Chemistry website.

Figure 23: One dimensional NMR spectrum of bromoethane. The three protons on carbon 2 are represented as the more upfield peak (closer to 1 ppm) whereas the protons on carbon 1 are more downfield, due to the shielding effects of bromine. The splitting pattern reveals how many neighbors are present by the n+1 rule with the protons on carbon one having a splitting pattern of 3+1 making a quartet, while the protons on carbon 2 have a splitting pattern of 2+1 resulting in a

triplet. The amount of protons on each carbon is determined by the relative area integration of each peak.

Figure 24: Regions of a proton spectrum in a general folded peptide. This figure was adapted from Kurt Wüthrich 1986 (5).

Figure 25: A generalized example of a two-dimensional spectrum of a peptide. Region g represents the fingerprint region. This figure was adapted from Kurt Wüthrich 1986 (5).

Figure 26: Vicinal protons of an aromatic ring (benzene), showing the known distance of 2.5 Å.

Figure 27: An example of binned restraint lines of code in the XPLOR format created by the assignment software SPARKY. Assign acts as an operator which tells the MD program to take the proceeding two specific atom positions and restrain them between a lower bound distance, an upper bound distance, and allow for a particular error (the third column of numbers, left to right respectively).

Figure 28: The comparative secondary shift values of PnID and PnID RevB. The similar shift values from residues 5-12 indicate that the backbone positions are conserved in these regions, while positions 1 and 2 appear to be the least conserved. The more positive chemical shift values of PnID RevB provide further evidence of a higher likely hood of β sheet type character.

Figure 29: The TOCSY connectivity diagrams of PnID B (4A) and PnID RevB (4B) and NOESY sequence walks (4C and 4D).

Figure 30: Calculated models of PnID B and RevB. Pane A shows the trace overlay of the 20 lowest energy models of PnID RevB. Pane B shows the lowest energy structure of PnID RevB. Pane C is a comparison of PnID B and RevB. Pane D shows the trace overlay of the 20 lowest energy models of PnID B. Pane E shows the lowest energy structure of PnID B. Pane F is the lowest energy structure of MrIA.

List of Tables

List of Abbreviations

αα, Amino Acid; Acm, Acetamidomethyl; Arg(Pbf), N-alpha-9-Fluorenylmethoxycarbonyl-N-γ-2,2,4,6,7-Pentamethyldihydrobenzofuran-5-Sulfonyl-L-Arginine; CCK, Cyclic Cystine Knot; Cys(Trt), N-alpha-9-Fluorenylmethoxycarbonyl-S-Trityl-L-Cysteine; Da, Daltons; DCM, Dichloromethane; DIEA, Diisopropylethylamine; DMF, Dimethylformamide; ESI-MS, Electrospray Ionization Mass Spectrometry; FDA, Food and Drug Administration; Fmoc, 9-Fluorenylmethyloxycarbonyl; HCTU,1H–Benzotriazolium-1-[bis(Dimethylamino)Methylene]-5-Chloro-Hexafluorophosphate-(1-),3-Oxide; GFCK, Growth Factor Cystine knot; NET, Norepinephrine Transporter; ICK, Inhibitor Cystine Knot; LC/MS, Liquid Chromatography interfaced Mass Spectrometry; MBHA, (4-Methyl)benzhydrylamine; MeCN, Methylcyanide; M/Z, Mass to charge ratio; PDA, Photo Diode; RP-HPLC/UV, Reverse phase - High Performance Liquid Chromatography interfaced Ultra-violet Detection; RMSD, Root Mean Squared Deviation; Rt, Retention time; SOMs, Small Organic Molecules; SWOT, Strengths, Weaknesses, Opportunities, and Threats; TFA/aq., Trifluoroacetic acid/aqueous; TIPS, Triisopropylsilane.

Chapter 1: Introduction, Overview, and Objectives

1.0 Introduction

1.1.1 Peptides as therapeutics

The following section contains excerpts from my paper entitled "A 21st-century approach to age-old problems: the ascension of biologics in clinical therapeutics".

Peptides are short chains of polymerized amino acid monomers (generally <50) which are vitally important to many functions in the life of an organism such as: cell signaling, physiological regulation, structural maintenance, maintaining redox conditions, and often in venomous organisms they are used for capturing prey or self-defense (6-8). Since these molecules are used to control so many native systems in an organism, they are of interest as potential therapeutics.

"Biologics" or natively produced peptides and proteins are becoming of increasingly significant interest within the pharmaceutical community because of their unique properties of potency and demonstrated lack of dangerous metabolites (9). Many examples of these compounds exist within the current US FDA approved pharmacopeia including cancer medications, analgesics, and birth control (as well as many others). In recent years advances in proteomics and genomics have provided a vast amount of information for the further characterization and synthesis of these once thought too complex and unviable molecules. Despite triumphs over past hurdles in the industry, the development of biologics remains to see a long road of trials before taking the place of mainstream therapeutics, a title which has long been held by small organic molecules (SOMs).
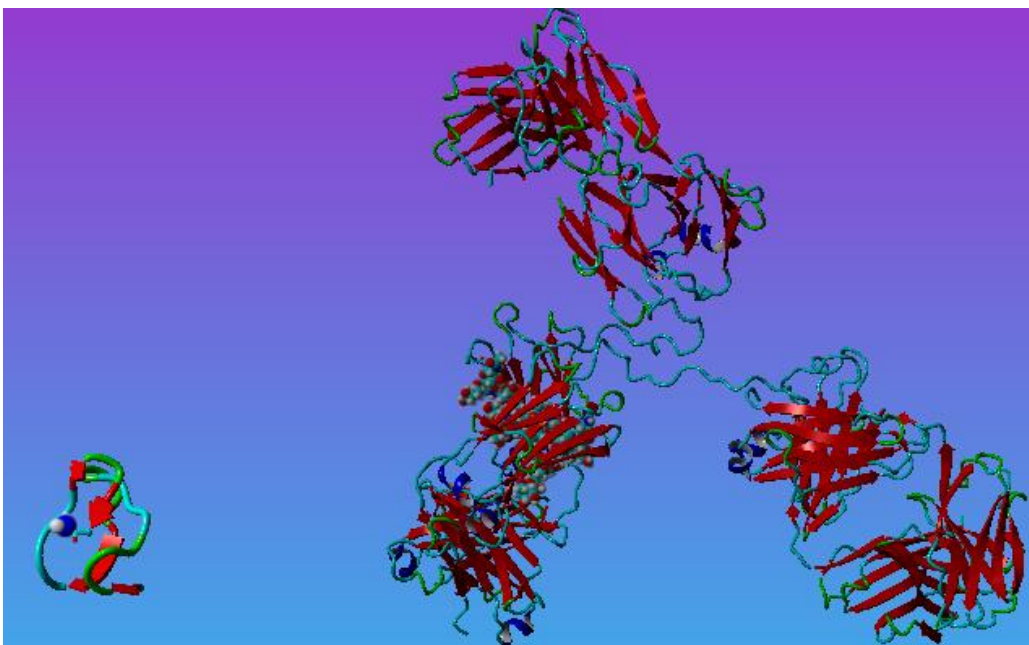
Figure 1: A comparison between the structures of two "biologic" based drugs, Prialt® (left) an FDA approved analgesic peptide, and trastuzumab (right) a monoclonal antibody used to treat breast cancer.

Peptides, the smallest members of the class of biologics, provide an array of possibilities in terms of selectivity, are easily and inexpensively synthesized, and are often slightly more stable than the larger protein-based biologics, such as monoclonal antibodies (Figure 1) or hormones (10). Although many barriers exist in the development of peptide drugs such as increasing stability, anaphylaxis, and achieving efficient folds (the focus of which will be the topic of this dissertation), to date the current focus in the development of peptide drugs has been their lack of oral absorption (11). Recently, novel lipid and chemical delivery systems have been developed that are making peptide-based therapeutics increasingly more viable. Some examples of successful or future potential peptide drugs include: the conotoxin-derived morphine alternative Prialt®/Ziconotide/ω-conotoxin MVIIA (Jazz Pharmaceuticals) (12); the NMDA-modulating potential antidepressant GLYX13 (Naurex Inc.) (13); the gp41-binding HIV entry inhibitor T-20 or Enfuvirtide (14); and the antibiotic peptide Coly-Mycin1 (colistin/polymyxin E; JHP Pharmaceuticals) (15). Peptide-based drugs, similar to that of their larger antibody-based cousins, are unique in that they have a high ability for selective targeting because of innate analogy to the receptors of the body. Unlike SOMs, it has been postulated that peptides have a higher potential to mimic the natural communication systems of an organism, conferring lower risk of toxicity (11). In fact, approximately 20% of peptide drugs come to market after surviving clinical trials, in comparison to only 10% for SOMs (15). This relatively high success rate has contributed to the re-evaluation and renewed enthusiasm for peptide-based therapeutics (and other biologics) within the pharmaceutical sector (11). Most current efforts focus on the development of orally active peptide-based drugs, with laboratories (both academic and

industrial) pursuing increased oral availability through differing forms of recombinant and synthetic chemistry techniques, as well as lipid-based and other protective delivery systems; however, peptide-based cancer vaccines are also an emerging field (16,17). Outside of liposome technology, a synthetic approach for peptides that has shown potential merit is *N*- to *C*-peptide backbone cyclization (18). The initial use of peptide cyclization to stabilize bioactive peptides is seen with the success of Integrilin1 (Eprifibatide; Millennium Pharmaceuticals), a bioengineered cyclic hepta-peptide antiplatelet drug derived from a venom protein found within the southeastern pygmy rattlesnake. Continued improvements in cyclic peptide (cyclotide) stability, which now progresses to stabilize oral delivery, will undoubtedly further increase the potential for peptide-based drugs.

1.1.2 Introduction to Conotoxins

This section contains excerpts from my paper entitled "Conotoxins and their regulatory considerations"

The venom of *Conus* sea snails (Figure 2) has evolved as an efficient means of prey incapacitation and as an effective defense mechanism (19). The deadly effects of this venomous cocktail are a direct result of conotoxins, small disulfide-rich peptides that are potent antagonists of a range neuronal receptors and ion channels. These bioactive peptides express selectivity towards their target receptors and are even able to discriminate between receptor subtypes. These properties are being utilized to design receptor-modulating ligands with potential therapeutic applications. Conotoxins are typically about 10–40 amino acids in length, but exhibit many motifs that are normally found in larger proteins. These structural features are stabilized by disulfide bonds generated from the abundance of cysteine moieties in its primary structure (19,20). Interestingly, these cysteine residues are found in predictable locations within a conotoxin's sequence, giving rise to loops of known amino acid lengths, which influence their bioactivity (20,21). Although disulfide bonds play a major role in determining conotoxin structure activity relationships, assessing how they influence bioactivity of conotoxins should be handled on a case-by-case basis as they stabilize peptide structure in varying degrees. The number of disulfide isomers possible within a conotoxin is directly correlated with an increasing number of cysteines within a peptide sequence (22). Notably, when the disulfide bond connectivity of α-conotoxin AuIB was rearranged, the resulting ribbon isomer exhibited greater bioactivity than the globular native toxin (23). This finding is valuable to conotoxin bioengineering as it expands the repertoire of modifiable conotoxins to include non-native isomers. Despite their great sequence diversity, the cysteine frameworks of these bioactive peptides are quite predictable. For majority of conotoxins containing four cysteines, their cysteine framework is given by the formula $(X)_{n1}C^iC^{ii}(X)_{n2}C^{iii}(X)_{n3}C^{iv}(X)_{n4}$ (20). Interestingly, the number of amino acids in *n2* and *n3* of these conotoxin's cysteine frameworks can have profound consequences in their selectivity toward a specific receptor class. Generally, α-conotoxins with a 3/5 cysteine framework (n2 = 3 and n3 = 5) will target muscle nicotinic acetylcholine receptors (nAChRs) (24) while conotoxins with a 4/7 framework (n2 = 4 and n3 = 7) will be specific for neuronal nAChRs (12,19,24,25). This has important implications regarding conotoxin classification. Conotoxins can be systematically ordered into superfamilies and further organized into families (sometimes referred to as classes). Superfamilies are based on disulfide bond framework patterns, while families are founded on their pharmacological target (19,26). Currently, there are twenty main conotoxin superfamilies and within these inclusive categories lie their respective families, often denoted as a single Greek letter prefix. These categories are rapidly expanding with the discovery of new conotoxins.

Figure 2: *Conus textile,* one of the many venomous cone snail species capable of producing a cocktail of bioactive peptides.

Venom peptides from the *Conus* genus have long been studied for pharmaceutical potential on receptors related to a various human maladies including: severe pain (27), cancer progression (28), convulsions (29), and more, with the native peptide ω-conotoxin MVIIA (Marketed as Prialt®) being the first and only US FDA approved conotoxin therapeutic to date, being a completely unmodified form of the originally discovered toxin (30). Thousands of peptides have been characterized from *Conus* that encompass twelve different pharmacological targets (31,32) and have provided a wealth of information about peptide-receptor interactions. Further information about these targets and a discussion on the interest in conotoxins as possible medications to treat pain will be considered in Chapter 4.

Short disulfide bonded peptides such as conotoxins may be considered novel biologics in that their size is highly manageable (< 30 amino acids) (31,32) in comparison to many currently used biologics e.g. monoclonal antibodies. In addition to their size they often exhibit a potency almost unheard of within the organic molecule class, as well as a high amount of selectivity (33).

Their selectivity is often explained as an element that is constantly refined by nature. Unlike many synthetic small molecules, biologics and native based molecules (synthetic peptides or proteins similar to the native form) have the advantage of naturally being selected by their environment as effective drugs. These characteristics make them optimal to study as future medications or receptor probes.

One of the most important and well known aspects of conotoxins is their ability to selectively target ion channel subtypes (26). This dynamic receptor–ligand interaction can be utilized to understand subtle differences in the normal physiology of various ion channel subtypes. As such, conotoxins are viewed as valuable resources for phyla selective receptor probes. Phyla selectivity is commonly seen amongst individual conotoxins, representing their native predatory preferences (26). Many peptides isolated from a piscivorous (fish eating) cone snail such as *Conus magus*, will only display activity in receptor subtypes related to higher order organisms, whereas molluscivorous (mollusk eating) cone snails tend to produce conotoxins which target mollusk receptor subtypes (34). This type of selectivity can be greatly utilized in terms of creating safe and effective pesticides. Potential pesticides produced from the venoms of molluscivorous cone snails may provide compounds that could be naturally degraded by microorganisms without producing harmful byproducts or causing environmental damage (35).

Not only are conotoxins studied for their naturally selected pharmacodynamic and pharmacokinetic properties, their surplus of post-translationally modified (PTM) residues and stabilizing interactions make them unique candidates for gaining insight into the rational design of future drugs. Like larger proteins, they often provide a high level of selectivity for particular molecular targets. However the much smaller size of these peptides provides them with better adsorption and distribution properties. Studying the unique properties of these peptides and building a database of stable motifs which may be used as molecular scaffolds or by understanding their pharmacophores or active binding regions and adding them chimerically into other scaffolds pushes the envelope in the field of rational drug design. An excellent example of this is a recent study carried out by D'Souza *et al*. (36) in which the inhibitor cysteine knot motif or ICK of an *N*- to *C*-terminally cyclized peptide known as *Momordica cochinchinensis* trypsin inhibitor-II (MCoTI-II) was used as a molecular scaffold for a novel cancer drug candidate. In this study the researchers removed the original inter-cysteine loop of the MCoTI-II containing the native pharmacophore and replaced it with the binding region of a COG peptide, a known inhibitor of the SET protein which is often over expressed in many cancers. The team was able to use the properties of several peptides to create a rationally designed chimeric drug to inhibit the growth of these cancer cells.

In addition to the current interest of these peptides as probes or medications, potential exists for their break into the pesticide industry, as very few peptides have been isolated from molluscivorous species in contrast to the number isolated from piscivorous species. This is due

to the premise that it would be unlikely for a molluscivorous species to produce a potent peptide capable of effecting mammals. However it is likely that investigation into venom peptides derived from mollusk eating species could provide a new avenue for pesticide development through the discovery of biologics that are selective to invasive species.

1.1.3 Oxidative Folding of Conotoxins

The following section contains excerpts from my paper entitled "Incorporation of post-translational modified amino acids as an approach to increase both chemical and biological diversity of conotoxins and conopeptides"

Although much of the focus in the development of modern peptide therapeutics is on creating orally active drugs, one of the major drawbacks in developing peptides as therapeutics to date has been the selective formation of the appropriate disulfide bonds in an effective and efficient manner, and in turn the formation of the most biologically active motif. This motif formation has dictated the success of conotoxins as drug leads and plays a central role their receptor specificity and isoform selectivity (37). The primary structures of conotoxins reveal pattern-specific cysteine frameworks that generate loops responsible in part for their ultimate disulfide bond connections, final motif, and their receptor specificity (19,20,23). The peptide's unique receptor–ligand interaction is contingent on well-defined tertiary structures stabilized by disulfide bonds. Generally, conotoxins contain 1–3 disulfide bonds; however, as many as ten have been observed. Few conopeptides have demonstrated an absence in disulfide bond content, although they have been observed (31,32).

The smallest conotoxin observed in nature (by mass), α-conotoxin MIC, contains just 12 αα, of which four (33 %) are cysteines involved in the formation of two disulfide bridges (38). The highest possible combination of disulfide bridges within a peptide may be calculated using the formula: $p = \frac{n!}{\frac{n}{2}!2^{\frac{n}{2}}}$, where p is the number of possible disulfide bridges and n is the number of cysteines present within the peptide. These connections can have several permutations that generate new peptide isomers with distinct pharmacological and kinetic properties (33). As such, disulfide bridges directly contribute to the efficacy and biological variability of conotoxins.

Methods for the oxidative folding of conotoxins include a wide variety of synthetic and in vitro enzymatic techniques. Due to the commonplace of solid phase peptide synthesis (SPPS) in the production of conotoxins, the most popular techniques that are used to fold conotoxins are synthetic and are often complimentary to SPPS. The most common techniques include incubation in the presence of molecular oxygen in a buffer system such as 0.1 molar ammonium bicarbonate, incubation in a buffered system in the presence of glutathione (reduced/oxidized in a 2 to 1 ratio respectively), or the use of multiple orthogonal protecting strategies which can be

sequentially removed to create regio-selectively folded toxins. In addition many changes can be made to the buffering system to change the peptides interaction with the surrounding solvent to modify the energy barrier of particular folds, thus the diversity of folding may increase dependent on the addition of a particular chaotrope, a change in temperature, time, or ion pairing.

Folding a conotoxin in a buffered system by way of molecular oxygen creates an environment that often allows for the greatest diversification of the final products. This phenomenon is due to the principle that the folding of these toxins generally occurs under particular redox systems controlled by the equilibrium of oxido-reductants within the cell of which the toxin is natively produced. In addition these cells generally contain protein disulfide isomerases or PDIs within their endoplasmic reticulum that act as molecular chaperones, allowing for the efficient folding of the thermodynamic products. Thus stripping the system down to primary sequence and solvent interactions with molecular oxygen reduces the efficiency of the folding process but results in a more diverse array of products, some of which may have unique properties (39).

In contrast, folding a particular toxin in a glutathione based buffering system provides an *in vitro* environment that better mimics what is occurring *in vivo*. Oxidized and reduced glutathione concentration in the endoplasmic reticulum differ by approximately a 1:2 ratio (39), creating an environment which pushes the equilibrium of the oxidative folding process to favor that of the folded products, thereby causing a an amplification of the native-like material which is often the target product when studying these peptides.

The oxidative folding of peptides and proteins *in vitro* can occur by two specific models known as the collapsed model and the framework model (39). In the collapsed model of folding, various kinetic isomers form in addition to the thermodynamic product causing a significant loss in material and results in a soup of crude products that are often difficult to purify. The alternative model known as the framework model is one in which a peptide will transition from a reduced to an oxidized state while shifting directly to a thermodynamic isomer. In conotoxins both models have been observed and the preference of one folding mechanism to another is highly dependent on sequence variation.

Regio-selective folding of conotoxins is a process undertaken when the target disulfide connectivity of the conotoxin of interest is known and needs to be produced in an efficient manner. This technique requires the use of multiple orthogonal protecting strategies when the toxin contains at least four or more cysteines. One of the most robust and common orthogonal protecting strategies for the production of a four cysteine containing conotoxin utilizing Fmoc based chemistry is the inclusion of two acid labile trityl groups and two acetamidomethyl groups or Acms. This strategy allows the chemist to de-protect each pair of cysteines and form disulfides in a step-wise manner often resulting in a cleaner and more efficient synthesis (39).

Despite the fact that disulfide bridges could theoretically be formed following the previously mentioned equation, random and spontaneous generation of all permutations is rarely ever seen. This is due to differences that can occur within peptides primary sequences such as electrostatic interactions or steric hindrance of the backbone (due to Ramachandran angles) or side chains. Dominance of the fold is typically given to the lowest energy conformation. However, these yields can be greatly influenced by adding alcohols (40) or changing the environmental conditions as previously mentioned. Although the contribution of disulfide bonds to protein stability has been established, deletion of these cysteine bridges can have varying effects on biological activity. An example of this is seen in the work of Khoo et al. (41) in which they effectively induced a deletion of a single cystine bridge by substitution with Alanine in the 16 αα μ-conotoxin KIIIA to generate the synthetic analog μ-KIIIA [C1A, C9A]. Despite having one less cysteine bridge, μ-KIIIA [C1A, C9A] exhibited comparable activity with native μ-KIIIA (3 disulfide bonds) toward rat $Na_v1.2$ with K_D values of 0.008 ± 0.002 μM and 0.005 ± 0.005 μM, respectively. Nuclear Magnetic Resonance (NMR) data coupled with molecular dynamics (MD) modeling indicated that both μ-KIIIA [C1A, C9A] and its native counterpart exhibited the retention of a key α-helix essential in binding to $Na_v1.2$ and $Na_v1.4$ receptors. Parallel studies done with scorpion toxins exhibited comparable results (42,43). This finding suggests that biological activity of toxins is not absolutely dependent on disulfide bond stabilization for all toxins, implying that toxin activity can be induced with conservation of partial or native-like motifs stabilized by non-covalent interactions.

In contrast, Flinn et al. (44) demonstrated that deletion of a disulfide bond through synthetic substitution of Cys[15] and Cys[26] of 27-residue ω-conotoxin GVIA (3 native disulfide bonds), with serine residues resulted in a 8,000-fold loss of potency, with no measurable activity at 10 μM in rat vas deferens assay. NMR analysis of peptide [Ser][15,26] ω-conotoxin GVIA revealed a loss of native structure, demonstrated by the existence of multiple isomers In solution and attributed to cis–trans isomerism of an amino acid stretch normally constrained by the deleted cystine bond.

Although disulfide bond connectivity has proven to be a very important aspect for obtaining biologically active folds, during folding the possibility may arise for contortions or knots within the structure (also known as topologies) to occur which could considerably alter the fold of the peptide, and thus the activity. Although topologies have been reported in peptides with three or more disulfides, most of the literature describing novel cysteine rich toxins with two disulfides focuses on obtaining the correct disulfide bond and lacks information about possible topologies which may be present. In this work I present a novel peptide with unique folding characteristics which may provide new information about preferred disulfide connectivity, topology formation, or particular interactions which contribute to the peptide's structure activity relationships or SARs.

In addition to disulfide bond formation, several other PTMs have been shown to increase the diversity of motif formation within conotoxins. These modifications often change intramolecular conditions or solvent interactions to adopt different folds. The two most commonplace of which are *C*-terminal amidation, which is the second most common PTM in conotoxins, and hydroxylation (21).

C-terminal amidation is a PTM present at the *C*-terminus of peptides in which an amide group replaces the hydroxyl group of the carboxylic acid terminal. As seen with *N*-terminal pyroglutamic acid modification, this *C*-terminal modification affects the overall isoelectric point and net charge state, neutralizing the potential for deprotonation. *C*-terminal modification is typically achieved by enzymatically cleaving a flanking *C*-terminal Glycine at the N–C bond, thus creating a new *C*-terminal that retains the truncated remains of the original peptide bond from the neighboring *C*-terminal amino acid (45). The presence of *C*-terminal signal sequences such a pre-PTM Glycine is commonly observed in the genetic analysis of conotoxins (31,32). It is estimated that over 127 naturally occurring conopeptides contain a *C*-terminal amide. This encompasses a large number of gene superfamilies including: A, B, I1, I2, J, M, O1, O2, P, Q, S, T, and V superfamilies. *C*-terminal amidation has been shown to occur in vitro on short *C*-terminal Glycine containing peptides incubated with peptidylglycine α-amidating enzyme (EC 1.14.17.3), in an oxygen-rich environment and in the presence of reduced cofactors such as ascorbic acid (46). This reaction is thought to form a hydroxyl glycine intermediate before the reaction goes to completion. *C*-terminal amidation was first hypothesized as a PTM in α-conotoxin GI, one of the earliest discovered venom peptides (24). Suspicion regarding the *C*-terminal assignment was not clearly demonstrated until α-conotoxin GI was chemically synthesized a few years later (47). As originally observed in α-conotoxin GI, the presence of a *C*-terminal amide may act to protect peptide toxins against carboxypeptidase activity, increasing the circulation half-life in prey (24). The importance of *C*-terminal amidation within conotoxins was further recognized in the early- to mid-1990s when studies were first being conducted to understand their contribution to disulfide arrangement and folding, particularly that of the voltage-gated calcium channel blockers, or ω-conotoxins (45). Analysis of ω-conotoxin MVIIA revealed an increased folding efficiency with the presence of the pre-PTM Glycine at the *C*-terminus. This challenged the original concept that the larger *N*-terminal pre-pro sequence was the main source of thermodynamic favorability in enzyme-mediated folding and oxidation. *C*-terminal amidation is also thought to contribute to the formation of native globular structures within the α-conotoxin family. This is specifically illustrated with α-conotoxin ImI (45). It has been shown that this conotoxin can be converted from a well-established nicotinic acetylcholine receptor (nAChR) antagonist to a newly defined noradrenaline receptor ligand through excision of the *C*-terminal amide. It is believed that the *C*-terminal amide strongly influences disulfide bond connectivity arrangement, causing major structural changes that in turn define biological function and selectivity. Structural studies using software modeling and NMR have revealed that

C-terminal amidation in α-conotoxin ImI disrupts the folding of the ribbon conformation, particularly when Proline is present within the first intercysteine loop (48). These characteristics seem to increase the probability of forming a globular structure in α-conotoxin ImI and also appear to be conserved in other toxins. The preponderance of this Proline/*C*-terminal amidation interaction among conotoxins suggests its efficacy in defining structural and biological properties.

In addition to *C*-terminal amidation, hydroxylation has also been reported as a major influential PTM in the oxidative folding of conotoxins. The hydroxylation of amino acids (ααs) is a commonly observed PTM within conotoxins and is known to generally affect Proline or Valine residues as well as one currently documented instance of a Lysine modification. There are approximately 86 known naturally occurring conotoxins containing both *4-cis-* and *4-trans-*hydroxyprolines (Hyp/O) and 2 containing *cis*-hydroxy-D-Valine (49). Despite their abundance within conopeptides and conotoxin families, the precise contributions of these hydroxylated ααs to biological activity and structure are not completely understood. The hydroxylation of Proline in conopeptides has been shown to have various effects among different gene families. 4-*Cis*- and 4-*trans*-hydroxylated Prolines in μ-conotoxin GIIIA were shown to increase the rate of inhibition of $Na_V1.4$ sodium channels anywhere from 2.5 to 4.5 times, while having little effect on oxidative folding (50). In contrast, hydroxylation of α-conotoxins ImI and GI displayed a lower activity when Proline was substituted with Hyp, despite increased stability in their overall structures.

The Hyp substitution in α-conotoxin ImI makes the conotoxin completely inactive against α7 nAChRs while substitution in α-conotoxin GI appears to have a greatly reduced affinity for its respective nAChRs (50). A similar structural improvement was found in ω-conotoxin MVIIC, which experienced an approximate two-fold yield increase in oxidative folding and retained biological activity. The appearance of Hyp in most forms of proteins and peptides is generally due to the presence of prolyl-4-hydrolase (EC 1.14.11.2) (51). Presently, prolyl-4-hydrolase within the venom glands of the *Conus* genus has not been reported. It may also be possible that the enzyme responsible for the hydroxylation of Proline has a broader specificity that includes an ability to hydroxylate Lysine, based on previous research involving *Conus delessertii* (52). The production of partially or mixed PTM variants are a common occurrence within *Conus*, (53,54). This feature may represent a mechanism to increase chemical diversity and influence biological spectrum of conotoxins. It has been shown that the folding of conopeptides containing more than one Hyp is assisted by a molecular chaperone enzyme known as peptidylprolyl *cis–trans* isomerase (EC 5.2.1.8) or PPI, which happens to be a subunit of prolyl-4-hydrolase, indicating that folding and hydroxylation may be a simultaneous or coupled process (55). Several isoforms of PPI have been isolated from *Conus novaehollandiae* including PPI B, an enzyme present in the snail's ER, and two cytosolic forms. PPI B is known for its ability to isomerize trans-hydroxyprolines to the *cis* conformation, which in turn may increase the stability of the peptide

structure. Structural stability via isomerization was observed by NMR *in vitro* by comparing μ-conotoxin GIIIA that contained two trans Hyp, ω-conotoxin MVIIC that contains one Pro/Hyp residue and μ-conotoxin SIIIA that contained no Hyp residues. In this study, increased *Conus* PPIase caused a larger production of the native μ-conotoxin GIIIA while having little effect on the production of the non-hydroxyproline containing peptides. This provides evidence to the concept that the presence of Pro/Hyp gives rise to structural changes among conopeptides that can be facilitated by PPIases.

1.1.4 Conotoxin Structure

Conotoxins, despite their small size (~11-40 αα) often contain a certain degree of secondary structure that in addition to their cross-linked cysteine rich side chains, allows them to adopt highly stable folds. These secondary structures include but are not limited to alpha helices, which are almost ubiquitous in the aptly named α-conotoxins, and β-sheets which are often observed in χ/λ- and μ-conotoxins (Figure 3). These structural elements are sometimes held together in various turns such as β- or γ-turns. Due to the presence of these secondary structures nearly all conotoxins are considered to form well-ordered structures. The exception to this case is the α-conotoxin BuIA, where the globular form that has a strong affinity for nAChRs, exists in at least three different conformations and does not appear to have a stable backbone (56).
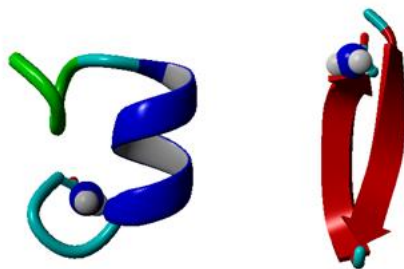
Figure 3: A comparison between the two prominent forms of secondary structure seen in conotoxins; an alpha helix as shown in α-MII (3) (left) and antiparallel β-sheets as shown in χ-MrIA (4) (right). Both structures were obtained from the World Wide Protein Databank (wwPDB) and were visualized using YASARA.

Conotoxin primary sequence is generally highly variable however there are particular cysteine patterns that have been determined to be highly conserved (31,32). These patterns are known as "cysteine frameworks" and have been directly correlated to motif formation as well as pharmacological families. Cysteine frameworks deal directly with whether or not the cysteines are separated by other residues, the collective of which are known as loops. Cysteine framework I for example consists of a sequence XCCXCXC, where X is any residue or number of residues and C is cysteine. The intercysteine loop "size" or the amount of residues between each loop also

correlates directly to structure/activity, and the combination of a particular cysteine framework and loop size may provide a general insight to what pharmacological family a conotoxin belongs to. For example the conotoxins which contain two disulfide bonds and cysteine framework I all either belong to the α-, χ-, or ρ-pharmacological families whereas the conotoxins which contain 4/2 intercysteine loop sizes exclusively belong to the χ-pharmacological family.

In addition to secondary structure, motifs formed by conotoxins include globular type folds which are found in four cysteine containing peptides in which the disulfides encompass one another, i.e. C4-C12, C3-C9 where C is the one letter code for cysteine and the number is the position in the sequence, ribbon type folds in which the disulfides are flanking one another (C3-C12, C4-C9), or a "beaded" fold in which the disulfides are completely vicinal (C3-C4, C9-C12) (Figure 4). Globular type folds are observed to be a commonplace motif for most of the smaller conotoxins including the α- and ρ-conotoxins, whereas the ribbon motif appears to be far more rarely observed with it only having been naturally observed in the χ/λ-conotoxins thus far (31,32). Finally, the vicinal disulfide or "beaded" motif, although seen in other biologically relevant proteins such as thioredoxin (57), has yet to be observed as a natural occurrence in conotoxins.

Figure 4: The three different possible disulfide configurations of a peptide containing four cysteines. PnID, the novel peptide described here is used as an example.

Conotoxins containing three disulfides also appear to have prominent motifs, the most notably being the "inhibitor cysteine knot" motif, which is also commonplace among other peptide toxins and comprises approximately 40% of the known peptide toxin motifs (58).

Conotoxin structure has largely been elucidated through the use of NMR solution-based studies, where over 90% of conotoxins have had their structures determined by NMR, which is in direct contrast to other proteins, where the majority of the entries in the Protein Databank (PDB) have been solved by X-ray crystallography (58). This difference is largely in part due to the convenience of studying small, water soluble peptides in solution through well-established methods rather than attempting a set of trial and error crystallization techniques. The drawback

however in using solution state NMR to study peptides is the possibility of missing important key elements which might cause a significant increase in activity, such as the binding of particular ions like calcium. Additionally X-ray crystallography holds a significant advantage in determining binding interactions or distortion/flexibility in structure that might be necessary for fitting a binding site (58).

χ-Conotoxins and topologies

χ (also known as λ)-conotoxins, are known for their unique 4/2 inter-cysteine loop size, preferential formation of the ribbon disulfide connectivity (59,60), and their potent inhibition of the norepinephrine transporter or NET, a therapeutic target for severe cancer pain and epilepsy (61). Synthetic disulfide isomers of the χ-conotoxin CmrVIA and the many vigorously studied α-conotoxins have for the most part been significantly less active (a 10 fold reduction or more) than their native counterparts, enforcing the notion that the disulfides within these residues form critical crosslinks which stringently stabilize their active motifs (62,63). Exceptions to these observations however have been previously documented with some "non-native" connectivies displaying an even greater amount of activity than the "native" linkages observed in these peptides. One such example of this is the previously mentioned α-conotoxin AuIB in which the "non-native" ribbon isoform illustrated a 10-fold higher activity than the globular form (64). This phenomenon has also been reported in the μ-conotoxin family where multiple isoforms of μ-KIIIA displayed low nM Kds (65), indicating that disulfide cross-links can hold multiple configurations without always significantly perturbing structure and activity.

The determination of the native disulfide bond connectivity of peptide toxins is often a difficult process, (especially as the number of possible disulfides within a peptide increases) recurrently requiring large quantities of peptide and multiple experiments (66). In addition disulfide bond formation is a PTM and cannot be directly predicted on the genetic level. Due to these difficulties and the observation of increased activities of particular conotoxins with "non-native" disulfide connectivities, it may be possible that *Conus* is capable of producing multiple native disulfide isomers as venom peptide constituents that until this point have gone undiscovered.

1.2 Project Overview

The goal of this project is to determine what unique structure activity relationships are associated with the "Chi-like" conopeptide PnID and if these properties could be applied to the development of future peptide based drugs. Preliminary data suggests that one of the disulfide bond isomers of PnID may contain a unique topology that is not detectable by current reduction and alkylation techniques or standard quality control processes. Currently one of the biggest

hurdles preventing cysteine rich peptides from transitioning to therapeutics is the issue of forming the most active structural motif in high enough yields to create a cost effective product. Of the hundreds of disulfide-bonded conotoxins discovered to date, much of the emphasis in achieving the "correct fold" has been placed on obtaining the appropriate disulfide connectivity.

Here a unique conopeptide is presented which contains interactions that influence its overall topology as well as its preferential disulfide connectivity. This peptide contains properties which if better understood could provide further insight as to the importance of disulfide and non-disulfide bond stabilizing characteristics within peptides.

Here we investigate the unique folding properties of the conotoxin PnID, which was isolated from the Red Sea *Conus pennaceus* (Bingham and Wolfender, unpublished data) otherwise known as the "feather cone" and appears to be related to the pharmacological family of χ/λ-conotoxins or noradrenaline transporter inhibitors due to its sequence similarity to the originally discovered toxins within this family: MrIA, MrIB, and CmrVIA, both genetically and by mature sequence (67). Although the sequence of this peptide has been previously reported by genetic analysis (68-70) this is the first instance where protein level interactions, biological activity, and structural information have been investigated and to the best of our knowledge this is the first report not only of the existence of native disulfide isomers but also of topological isomers of the same toxin sequence and disulfide linkage existing within *Conus*.

An investigation of topological isomers (isomeric material produced by geometrical shape and constraint independent of covalent bond connectivity or stereochemistry) in particular, helps to better characterize peptides that may be currently understudied. Topology is known to be an important factor in the regulation of DNA accessibility however very little research has been reported on the topology of conotoxins, as most of the research focus is exclusively on disulfide bond connectivity. An understanding of peptide topology would help to bolster the already growing biologics industry by providing accurate information on the desired products from synthesis and prevent inaccurate data from biological activity assays that may arise due to negligently assigning structure and activity from disulfide connectivity alone.

PnID is a unique 4/2 conopeptide that contains a high amount of homology to the χ-conotoxins. Unlike the χ-conotoxins that contain a CCGYKLCXXC sequence (where X can be any known amino acid) PnID contains a CCGYRMCVPC stretch making it closely homologous but diverse from the known χ-conotoxins. Also apparent in the χ-conotoxins is the presence of a Valine before the first Cysteine and a hydroxyproline in the last inter-cysteine loop. These residues have been previously characterized by NMR to be involved in stabilizing hydrophobic interactions (4). It is believed that the lack of this interaction in the proper areas of the molecule might be a contributing factor to the absence of a single apparent isomer in the oxidation of PnID.

Therefore the unique properties of PnID present potentially interesting characteristics and the study of these characteristics will contribute to our understanding of the effects of sequence

diversity in the 4/2 conotoxins, help us better understand the stabilizing factors for these inter-cysteine loops, possibly broaden our understanding of selectivity in conotoxins, and provide further insight for the production of better drugs in the future.

1.3 Objectives and Hypothesis

Objective 1: Production of the randomly folded isomers of PnID. This will be done through solid phase peptide synthesis, isolation of the disulfide-reduced material by RP-HPLC/UV, Oxidation of the thiols under 9 different air-oxidation conditions, and confirmation by mass spectrometry (MS).

Rationale: Two native peptides of PnID were originally discovered as isomers in the same duct venom sample. To the best of our knowledge isomeric material of a peptide isolated from native *Conus* venom has never been seen elsewhere. It is therefore suspected that these native isomers have unique folding capabilities that are not possible in other venom peptides.

Hypothesis: Random air-oxidation of the unfolded or "linear" PnID will produce maximally three isomers since it only contains two disulfide bonds, two of which will be the isomers seen in the native venom.

Activities:

1. The production of synthetic PnID through Solid Phase Peptide Synthesis (SPPS)
2. Isolation of the product material through Reverse Phase High Pressure Liquid Chromatography (RP-HPLC/UV) and confirmation via Electrospray Ionization Mass Spectrometry (MS) (>97% purity).
3. Non-selective (random) folding through oxidation of the thiol containing cysteines by introduction to a moderately basic environment in the presence of atmospheric oxygen for a period of 3-5 days.
4. Characterization of the oxidation products by RP-HPLC/UV as well as confirmation and purity assessment of these products by ESI-MS

Objective 2: Production of orthogonally protected PnID isomers through solid phase peptide synthesis and directed sequential oxidation. These will be co-eluted with the randomly folded material to determine the disulfide connectivity of each. Confirmation of their production will be assessed by ESI-MS, RP-HPLC/UV, and nuclear magnetic resonance spectroscopy (NMR).

Rationale: Synthesis using differential protecting groups for two out of the four cysteines will allow the production of individual disulfide isomers. The selective folding should produce three possible forms including a ribbon, globular, and beaded form. The successive co-elution of each form by RP-HPLC/UV will provide evidence of which isomers are being produced.

Hypothesis: The Isomers produced by directed synthesis will co-elute with the three isomers produced by the random oxidation process indicating the presence of different disulfide connectivities.

Activities:

1. Production of the Ribbon (PnID isomer A), Globular (PnID isomer B), and Beaded forms (PnID isomer C) of PnID through SPPS, incorporating acetamidomethyl (Acm) protecting groups in addition to Trityl protecting groups on the cysteine side chains.
2. Initial oxidation of the cleavage liberated thiols
3. Removal of the Acm groups, and the oxidation of the remaining disulfide
4. Purification by RP-HPLC/UV and confirmation by ESI-MS
5. Co-elution of each isomer with the randomly oxidized peptides.

Objective 3: Production of the randomly folded isomers of the χ-conotoxins MrIA, and CMrIVA followed by comparative oxidations with the native PnID sequence. These isomers will be confirmed of their production by RP-HPLC/UV and MS.

Rationale: Random oxidations of the PnID sequence fails to produce one apparent thermodynamic isomer. Comparative oxidations will be performed on the closely homologous χ-conotoxins to determine if PnID is unique in the absence of this single thermodynamic isomer.

Hypothesis: Unlike PnID the χ-conotoxins will produce a single thermodynamic isomer. This difference is due to the critical placement of the Valine before the first inter-cysteine loop and it's interaction with the Hydroxyproline in the last inter-cysteine loop, an attribute which is present in the χ-conotoxins but absent in PnID.

Activities:
1. The production of the randomly folded isomers of the χ-conotoxin MrIA by SPPS
2. The production of the randomly folded isomers of the χ-conotoxin CMrVIA by SPPS
3. Isolation of the product material through RP-HPLC/UV and confirmation via ESI-MS
4. Non-selective (random) folding through oxidation of the thiol containing cysteines of each peptide by introduction to a moderately basic, organic, and chaotrophic environment (Urea 6M, NH_4HCO_3 100 mM pH 7.8 50% (v/v) propanol) Each condition is outlined in the table below (Table 1).
5. Characterization by RP-HPLC/UV and further Isolation of each isomer by RP-HPLC and product confirmation by ESI-MS

Table 1: Experimental oxidation conditions

	Oxidation Agents	Temperature	Time
Condition 1	0.1 M NH$_4$HCO$_3$ pH 8	25°C	3 days
Condition 2	0.33 M NH$_4$OAc 0.5 GnHCl pH 7.5	25°C	3 days
Condition 3	2 M urea 0.1 M NaCl 0.1 M Glycine 50% Propanol pH 7.8	25°C	3 days
Condition 4	0.1 M NH$_4$HCO$_3$ pH 8	4°C	5 days
Condition 5	0.33 M NH$_4$OAc 0.5 GnHCl pH 7.5	4°C	5 days
Condition 6	0.33 M NH$_4$OAc 0.5 GnHCl pH 7.5 50% Propanol	4°C	5 days
Condition 7	2M Urea 0.1 M NaCl 0.1 Glycine pH 7.8	4°C	5 days
Condition 8	2M urea 0.1 M NaCl 0.1 Glycine pH 7.8 50% Propanol	4°C	5 days
Condition 9	6M urea 50% propanol 0.1M NH$_4$HCO$_3$ pH 7.5	4°C	5 days

Objective 4: LD_{50} Assays on the PnID isomers to determine differences in activity. PnID will be subjected to both fish and mollusk based assays. Additionally microgram-milligram quantities of each peptide (>90% purity) will be produced for their individual ion channel activity determination by whole cell patch clamp. PnID will be tested on: a norepinephrine uptake assay, GIRK1/4, GIRK1/2 and various nACHRs.

Rationale: The biological activity of the PnID isomers must be determined on mollusks in addition to the fish due to its possible inherent phyla selectivity. Since there is little known about mollusk ion channels it is very difficult to perform patch clamp on them, it was therefore determined that LD_{50} assays on snails and fish in addition to a large scale screening of individual ion channels would be the most appropriate approach for characterizing biological activity. The determination of activity differences among the isomers for both of the peptides will allow us greater insight as to whether certain isomers may be better drugs, probes or pesticides.

Hypothesis: At least two isomers from PnID will produce LD_{50}s in the nMol/g range and/or nMolar potency on individual channels.

Activities:
1. Production of milligram quantities of PnID
2. LD_{50} Assays on guppies with the individual PnID isomers
3. LD_{50} Assays on cowries with the individual PnID isomers
4. LD_{50} Assays on guppies with the mixed randomly oxidized PnID material
5. LD_{50} Assays on cowries with the mixed randomly oxidized PnID material
6. EC_{50} assays on the norepinephrine uptake assay using the two Globular and one Ribbon PnID isomers

Objective 5: 2D NMR analysis on the major isomers of PnID in order to characterize their differences in tertiary structure. Correlation, total correlation, rotational overhouser, and nuclear overhouser experiments will be used to gather data on tertiary structure. Additionally this data will be used to generate restraints using the SPARKY program for the MD package YASARA Structure to model the most likely tertiary structures occurring in solution. All structures will be assessed using Ramachandran analysis.

Rationale: Two-dimensional NMR analysis will provide conclusive evidence of the differences in tertiary structure among the directed isomers and the randomly folded isomers and provide insight as to why these unique differences are necessary for the enhancement of the activity of these peptides in their prey. NMR structure determination on a largely homologous peptide known as MrIA produced a stable structure that included two-disulfide stabilized beta sheets. Comparative analysis with PnID will provide insight as what might be causing the isomeric material.

Hypothesis: The directed isomers will contain different tertiary structures to that of the randomly oxidized isomers due to conformational differences among the peptides, despite containing the same disulfide connectivity.

Activities:

1. One dimensional NMR experiments on the Ribbon and two Globular PnID isomers
2. Two dimensional COSY NMR experiments on the two Globular PnID isomers
3. Two dimensional TOCSY NMR experiments on the two Globular PnID isomers
4. Two dimensional NOESY NMR experiments on the two Globular PnID isomers
5. Assignment of spin systems using the crosspeaks from *j*-coupled protons in the program SPARKY
6. Assignment of crosspeaks from dipolar couplings using the program SPARKY for each isomer
7. Generation of a list of distance restraints from integrated and assigned peaks for each isomer
8. Energy minimization of an extended structure of each isomer using a simulated annealing protocol and unambiguous restraints
9. Refinement of the structures by removing errors, adding ambiguous restraints, and decreasing NOE energy
10. Structure validation through the Protein Databank (PDB) and the Biomagnetic Resonance Bank (BMRB)

Chapter 2: Synthetic Production of PnID

2.0 Introduction

SPPS is often considered the preferred method for the production of conotoxins due to their short size, ease of implementing PTMs, and automation. Peptides have been synthesized chemically for over a half century, as the first synthesis of a bioactive peptide was done in 1953 in the liquid phase by Vincent du Vigneud in which he synthesized the octapeptide oxytocin (71). A decade later Bruce Merrifield developed a simplified technique in which peptides could be synthesized in the solid phase by a cyclic process, allowing for the removal of waste by flow washing resin bound reactants in a scinted vessel (72). This technique paved the way for the automation of peptide synthesis and both men went on to win the Nobel Prize in chemistry.

Modern SPPS may either be performed manually or automatically by a computerized synthesizer (73). Automated synthesis is often used when producing many peptides such as in combinatorial libraries, alanine scans, or large scale production whereas manual synthesis may be considered more ideal when creating test peptides, especially when the use of several solvents might be required, such as in the case of difficult couplings (74). SPPS is carried out using a solid resin support to which a chemical linker is added. This linker then contains a functional group with which the *C*-terminus of an amino acid (whose *N*-terminus and side chain are chemically protected) may be reacted or "coupled" using a derivatizing reagent. After coupling the resin is washed and the added amino acid's *N*-terminal protecting group is removed and washed. This process is then repeated in cycles until the sequence is completed. The sequence is then cleaved from its solid resin support using a strong acid. The two major *N*-terminal protecting strategies for SPPS are Fluorenylmethyloxycarbonyl or Fmoc and *tert*-Butyloxycarbonyl or *t*-Boc (75).

The first conotoxins to be synthesized by SPPS were α-conotoxins GI and MI in the 1980s (76). Before this conotoxins were obtained solely by isolation from native sources such as from the venom duct of the producing organism or later by "milking", a strategy which includes manipulating the animal into having an ejaculation event and collecting the ejaculated venom. The practice of harvesting native venom, although highly beneficial for the discovery of new peptides, is considered to be a less sustainable practice than chemical synthesis as it requires the organism to be removed from its natural environment and in the case of venom duct extraction even requires the animal to be sacrificed.

In addition to SPPS some conotoxins are produced through expression in a single celled organism such as *E. coli*. The technology of recombinant DNA expression did not become available until the 1980s (77) and the first conotoxins to be synthesized recombinantly were only produced recently (78). Although recombinant production of conotoxins might be considered a

cheaper process once an efficient method is developed, the relative difficulty involved in method development and the need to tailor each expression system to an individual peptide sequence has made it a far less attractive method in comparison to SPPS as of yet. In addition, the major disincentive is the inability, at present, to undertake site-specific PTM incorporation (including unnatural amino acid analogs), an essential structural/pharmacological character within many conotoxins.

PnID, like many conotoxins is a very short peptide and consists of only 12 amino acids. The sequence of PnID (Table 2) contains only a *C*-terminal amide and two cystine linkages as PTMs. These characteristics make PnID an ideal candidate to be produced by SPPS as the synthesis can be undertaken very quickly and the orthogonal protecting strategy can be developed to regio-selectively form disulfide bonds.

Table 2: Comparison between the mature regions, activity and folding of the reported χ-conotoxins

χ-conotoxin	Mature sequence	Superfamily	Framework	EC_{50} NET	Connectivity	Ref
PnID	STCCGYRMCVPC-NH$_2$	T	X	10 μM	1-4, 2-3	this work
MrIA	NGVCCGYKLCHOC	T	X	645 nM	1-4, 2-3	(67,79,80)
MrIB	VGVCCGYKLCHOC	T	X	860 nM	1-4, 2-3	(67)
CmrVIA	VCCGYKLCHOC	T	X	N/A	1-4, 2-3	(81)

O = *4-trans*-hydroxyproline;NET = Norepinephrine Transport uptake assay

2.0 Materials and Methods

2.0.1 Manual Solid Phase Peptide Synthesis (General)

Swelling of the Resin:

0.5 Mmoles of Methylbenzylhydroxyamine (MBHA) resin containing a rink amide linker (0.44 meq/g loading) was measured and poured into a reaction vessel. 10-20 mL of Dimethylformamide (DMF) was added to the reaction vessel and allowed to shake for 10 minutes. The DMF was drained and 20 mL of fresh DMF was added to the reaction vessel. The resin was allowed to swell overnight, or for a minimum of 8 hours.

Activation of Fmoc Amino Acids:

The desired amino acid of choice was measured to 2 mMoles and placed in a scintillation vial. In a separate scintillation vial 2 mMoles of 2-(6-chloro-1-H-benzotriazole-1-yl)-1,1,3,3-tetramethylaminium hexafluorophosphate (HCTU) was weighed. The HCTU was dissolved in 4 mL of DMF and the solution was transferred to the scintillation vial containing the amino acid. 347 μL of DIEA was added to the reaction vessel to quench the protons generated during amino acid activation. If it appeared that the desired Fmoc amino acid was not fully dissolved, the scintillation vial was sonicated to ensure that the amino acid and coupling reagents were in solution. The amino acid and activating reagents were measured in 4-fold excess to the 0.5 mMol resin, resulting in 2 mMoles.

Amino acid Coupling:

2 mMoles of the activated amino acid was added to the reaction vessel containing 0.5 mMol resin and allowed to couple for 10-20 min. Since the amino acid is in 4-fold excess to the potential binding sites on the resin the reaction is expected to reach theoretical completion. The reaction was considered complete when the % coupling as determined by ninhydrin test exceeded 99.5%.

2.0.2 Ninhydrin Analysis:

Procedure:

Ninhydrin analysis was undertaken to determine if the coupling % of amino acid was satisfactory and the next amino acid in the primary sequence could be added to the growing peptide chain (82). 2 to 5 mg of the peptide resin was removed from the reaction vessel and flow washed with 50% v/v DCM in Methanol on a filter thimble under vacuum until the peptide resin

was dry. The resin was then transferred to a tared tube and the mass was recorded. The following reagents were added to the test tube containing the peptide and an empty control tube: 2 drops of 76% w/w phenol in ethanol, 4 drops of 0.2 mM potassium cyanide KCN in pyridine, and 2 drops of Reagent 0.28 M ninhydrin in ethanol. Both the tubes were covered with Parafilm$^®$ and placed in a sand bath, regulated at 110º C for 5 min. After removal from the sand bath, 3 mLs of 60% ethanol in water (v/v) was added to each test tube and centrifuged for 30 seconds. The supernatant was collected and absorbance of the resin sample was read against the control blank at 570 nm and recorded.

Calculation of Percent Coupling:

Percent coupling can be calculated with the following equation

$$Percent\ Coupling = 1 - \left(\frac{NV}{SV}\right) * 100$$

The term NV stands for Ninhyndrin Value and provides an estimate for the free unreacted amines that did not bind to the previous amino acid during the coupling step. NV can be calculated using the following equation.

$$NV = \frac{(absorbance_{570})(200)}{resin\ mass\ (mg)}$$

The SV or Substitution value provides the theoretical mole value of the total peptide. SV is calculated using the equation below.

New SV = (previous SV)/ [1 + (previous SV x AA MW)]

The previous SV for the first amino acid was determined through the Resin loading Value, which was provided by the manufacture. This value indicates the number of moles/g of resin. While calculating the amino acid molecular weight *N*-terminal and orthogonal protecting groups were considered.

Substituting observed data into the % coupling equation provided the % coupling. If this value was greater than or equal to 99.5% the next amino acid was coupled to the growing peptide chain. A score of below 99.5% indicates that the coupling did not reach the desired level of completion.

2.1.3 End Capping:
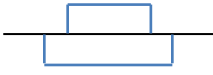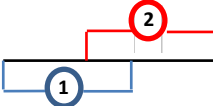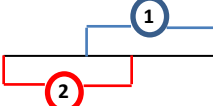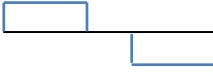
If the sequence repeatedly fails to reach 99.5% coupling, usually due to aggregation of the peptide chain or by steric hindrance, end capping becomes a necessary option. Free amino functions cause a deletion product when the new amino acid is coupled to the growing peptide chain that is permanent and leads to side products that reduce the final yield of target peptide and may cause difficulty during peptide purification. In order to prevent this from happening the free amino groups are capped with an inert moiety that prevents the *N*-terminal functionality from reacting during the rest of the synthesis.

To End cap the free amino termini, the resin was first flow washed with an abundant amount of DMF to remove any former unreacted amino acids and reagents from the system. 50% v/v *N*-methylpyrrolidone (NMP) in DCM was flow washed through the resin twice using ~20 mL each time. While doing this step extreme caution must be undertaken to avoid letting air pass through and dry the resin in order prevent oxidation of susceptible residues such as Methionine or Histidine. The reagent *N*-(2-chlorobenzyloxycarbonyloxy) succinimide (Z[2-Cl]-OSu) was weighed out in a scintillation vial to 0.5 mMol and dissolved in 4 mL of the NMP/DCM solvent previously prepared to activate the Z(2-Cl)-OSu. 347 µL of DIEA was added to the scintillation vial, mixed, and poured into the reaction vessel and allowed to react for 5 min. The reagents were then removed by two flow washes of approximately 20 mL using the NMP/DCM solvent, and then using abundant DMF. Upon completion all free N^{α}-terminal groups of the peptidyl-resin were terminated.

2.1.4 Synthesis of PnID random, A, B and C and the "Reverse" isomers

The PnID isomers were all synthesized using the previously mentioned solid phase peptide synthesis (SPPS) protocol (see Section 2.1 of methods). PnID A, B, and C all contained the primary sequence STCCGYRMCVPC-NH$_2$ and were all synthesized using the side chain protecting groups: Ser(tBu), Thr(tBu), Cys(Trt), Arg(Pbf), and Tyr(tBu), with the addition of Cys(Acm) on positions 3 and 12 on isomer A, positions 3 and 9 on isomer B, positions 9 and 12 on isomer C, positions 4 and 9 on isomer RevA, and positions 4 and 12 on isomer RevB (Table 3). Another PnID isomer which was just referred to as PnID, or PnID random was synthesized using Cys(Trt) for all cysteine residues. Upon completion of synthesis, the peptidyl-resin was washed with DMF (2x, 5 mL) followed by Dichloromethane (DCM; 10 mL) and dried under N$_2$.

Table 3: PnID Isomers that were produced by regio-selective synthesis (using Acm and Trityl orthogonal protecting groups) and oxidative folding. The numbers on the disulfide bonds indicate the order in which they were selectively formed.

Name	Connectivity
PnID A (PnID II) (RevA)	Ribbon
PnID B (PnID I)	Globular 3-9, 4-12
PnID RevB	Globular 4-12, 3-9
PnID C	Beaded

2.1.5 Synthesis of MrIA-amidated and CmrVIA-amidated

MrIA-NH$_2$ and CmrVIA-NH$_2$ were both synthesized using the previously mentioned SPPS protocol (see Section 2.1 of methods). MrIA amidated has the primary sequence NGVCCGYKLCHOC-NH$_2$ and was made using the orthogonal protecting groups: Asn(Trt),

Cys(Trt), Tyr(tBu), Lys(Boc), His(Trt), and Hyp. CmrVIA amidated has the primary sequence VCCGYKLCHOC-NH$_2$ and was made using the same respective orthogonal protecting groups as MrIA-NH$_2$.

2.1.7 Cleavage from resin and oxidative Folding

Assembled peptides were cleaved using Reagent K [TFA (82.5% v/v), phenol (5% v/v), water (5% v/v), thioanisole (5% v/v), and triisopropylsilane (TIPS) (2.5% v/v)]. 40 mL of Reagent K per gram of peptidyl-resin was mixed for 2 h at 24 °C. Cleaved slurry was vacuum filtered into liquid N$_2$ chilled tert-butyl methyl ether. Peptide precipitate was pelleted by centrifugation (3000 g, 10 min) and washed twice with chilled tert-butyl methyl ether. The resulting peptide pellet was suspended in 25% v/v acetic acid, then freeze-dried to form a powder and stored at −20 °C until required. Crude peptides (1 mg mL^{-1}) were oxidized using 100 mM NH$_4$HCO$_3$ pH 8 and stirred for 5 days at 4 °C. Oxidized material was filtered (0.45 μm) prior to semi-preparative RP-HPLC/UV fractionation. Cleaved χ-conotoxins PnID A, B, C, RevA, and RevB 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^{-1}) and by adding a solution of freshly saturated I2 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$_2$S$_2$O$_3$ until the stirring solution became clear, which was then followed by the addition of 200 μL 0.1 % 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.

2.1.8 RP-HPLC/UV and Co-elution Experiments

Native and synthetic conotoxins were individually separated as follows: (i) Capillary Scale (Phenomenex; C18, 5 μm, 300 Å, 1.0 × 250 mm, flow 100 μL min^{-1}) - used for comparative RP-HPLC/UV profiling, to control the quality of peptide purity, to quantify the peptides and to perform peptide co-elution experiments. (ii) Analytical Scale (Vydac; C18, 5 μm, 300 Å, 4.2 × 250 mm, flow 1 mL min^{-1}) - used for the isolation and purification of native peptides for ESI-MS. (iii) Preparative Scale (Vydac; C18, 10 μm, 300 Å, 22 × 250 mm, flow 5 mL min^{-1}) - used for the preparative separation of crude synthetic peptides for co-elution experiments, structure determination, and pharmacological assays. 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 Millennium32 (v3.2) software. Samples were eluted using a linear 1% min^{-1} gradient of organic (90/10% v/v CH$_3$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 Millennium32 software. Filtered (Nylon 0.22 µm) synthetic peptides and crude DV peptide extracts were manually loaded and eluted from the preparative scale column using the same 1% gradient at 5 mL min^{-1} and monitored at 214 and 280 nm. Fractions were collected manually and stored at −20 °C or freeze-dried until required. Each of the peptide isomers as well as the products from air oxidation were co-eluted both in 1:1 and 2:1 ratios on a C18 capillary-bore RP-HPLC/UV (Phenomenex; 5 µm, 300 Å, 1.0 × 250 mm) column.

2.2 Results

2.2.8 Synthesis Products

Each synthesis resulted in the production of peptide with the expected mass (Figure 7 is representative, for all spectrographs see Appendix B). Minor products were seen in each associated RP-HPLC/UV chromatogram and mass spectra, however in every case the target material was the dominant peak. Each resulting peptide was brought to a purity of ~> 95% before being used for other experiments or assays (Figures 5 and 6 are representative, for all chromatograms see Appendix A).
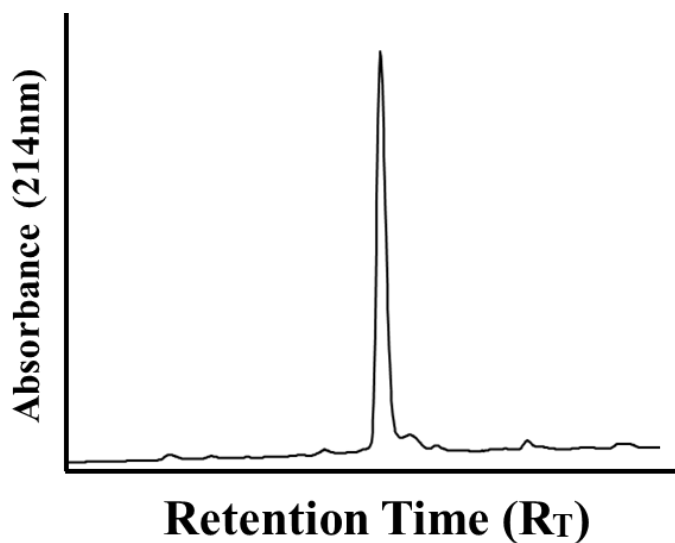
Figure 5: RP-HPLC/UV chromatogram of crude PnID A containing two orthogonally protected Acm cysteines and two cysteines with free thiols.
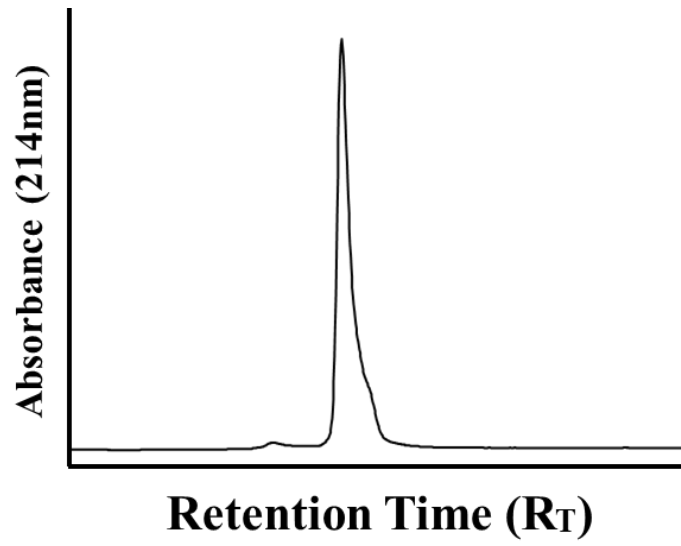
Figure 6: RP-HPLC/UV chromatogram of the fully folded and purified form of PnID A.
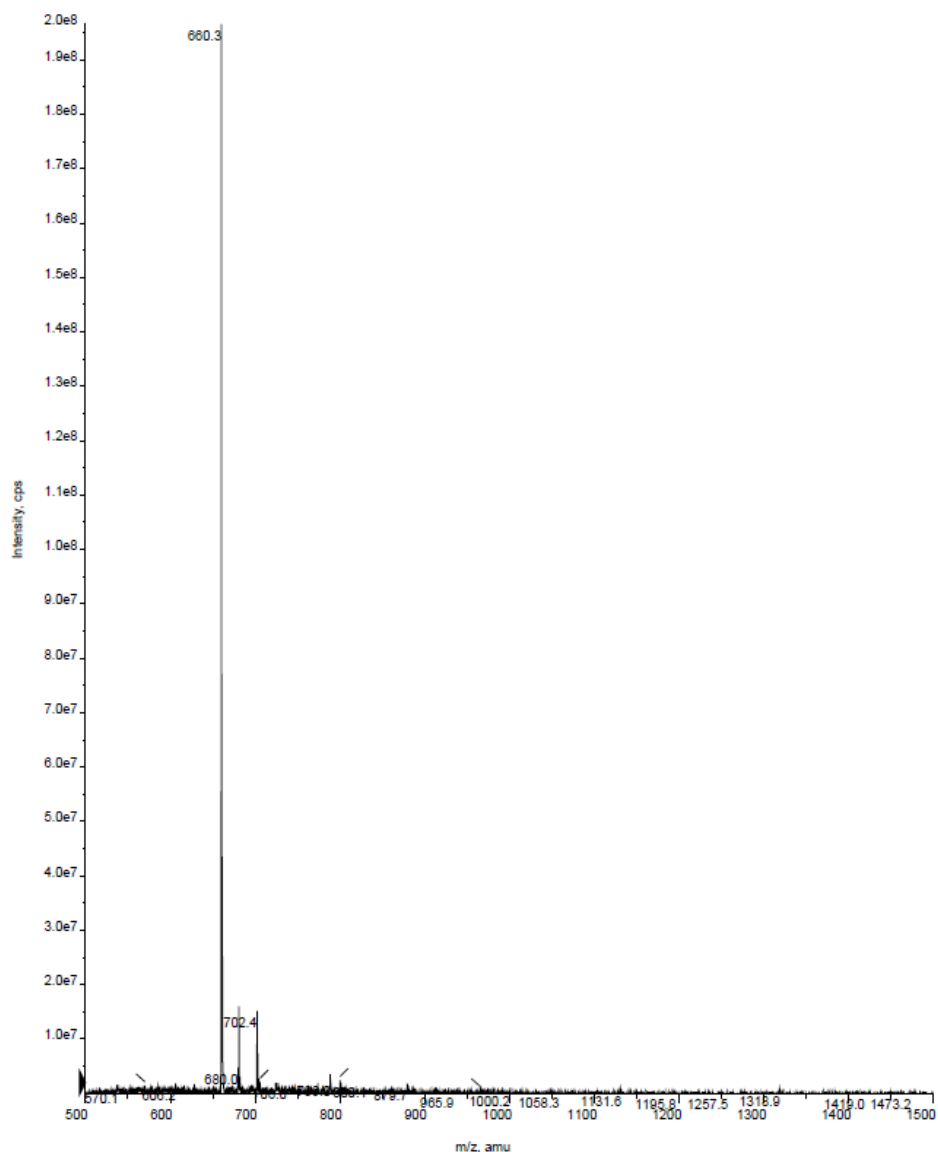
Figure 7: Mass spectrograph of the fully folded PnID A containing no orthogonal protecting groups. 660.3 was the observed mass to charge ratio (*m/z*) in the second charge state. The expected *m/z* was 660 +/- 0.5 [M+2H]$^{+2}$.

2.2.9 RP-HPLC/UV Co-elution

In peptides containing 4 cysteines, three specific disulfide pairings known as ribbon, globular, and "beads on a string" or beaded forms, are hypothetically achievable. Figure 8 shows the HPLC/UV chromatograms of the individual and mixed isomers of PnID that were previously subjected to oxidative folding. Co-elution was attempted between the products oxidized by a two-step folding process and those formed by a single step air oxidation in order to determine what disulfide bridges were susceptible to formation, as not each of the three possibilities will

always occur due to energetic barriers such as steric hindrance, torsional strain, or charge repulsion. The experiments were carried out both with the isomers at equal concentration (not shown), and at roughly 1:2 concentrations (peptides produced by a single step oxidation and a two-step oxidation respectively). PnID A, B, C, RevA, and RevB were all folded via two-step processes, allowing for the control of particular disulfides to be formed in a pre-determined order (see methods). PnID A, which contained disulfide bonds between cysteines 3-12 and 4-9 (ribbon isoform) was observed to co-elute with the more hydrophobic peak in the air oxidation profile, displaying one sharp peak without evidence of shouldering. PnID B, that contained disulfide bonds between cysteines 3-9 and 4-12 (globular isoform) did not appear to co-elute with any of the isomers produced within the air oxidation profile, exhibiting distinct shouldering with the earlier eluting major peak. PnID C, which contained disulfide bonds between cysteines 3-4 and 9-12 (beaded isoform) also did not appear to co-elute with any of the peaks generated from the air oxidation.
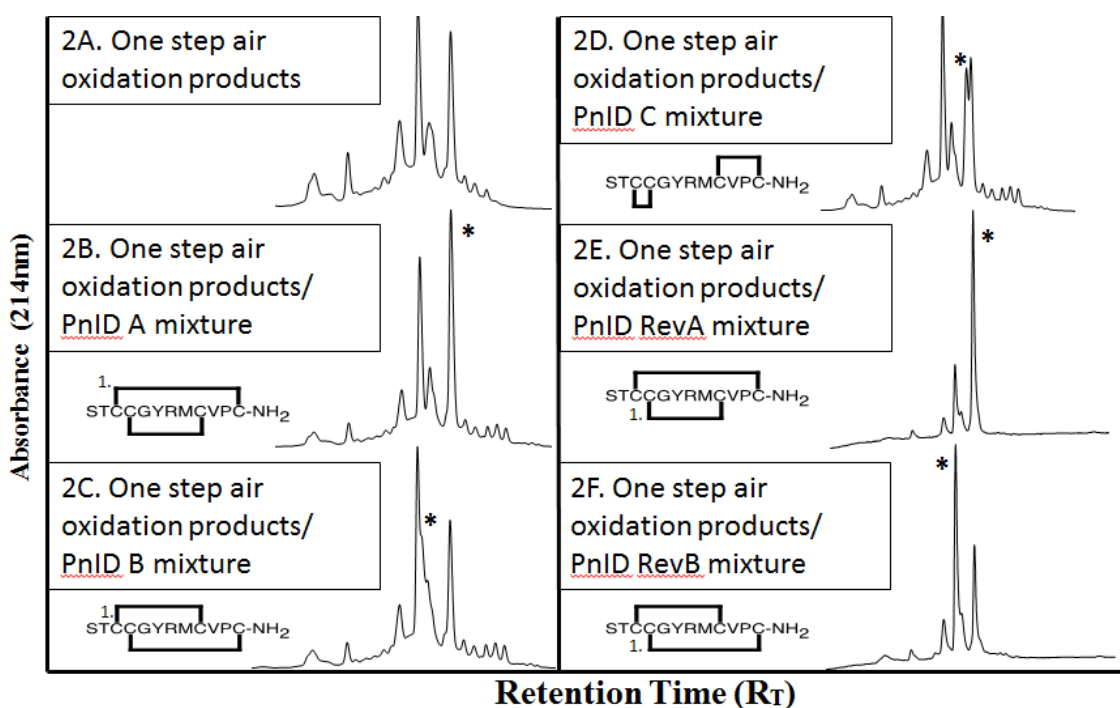
Figure 8: RP-HPLC/UV chromatograms of the isomers of PnID. 2A shows the products of the pre-purified air oxidation, 2B shows the attempted co-elution of PnID A with the air oxidation products, 2C shows the attempted co-elution of PnID B with the air oxidation products, 2D shows the attempted co-elution of PnID C with the air oxidation products, 2E shows the attempted co-elution of PnID RevA with the air oxidation products, and 2F shows the attempted co-elution of PnID RevB with the air oxidation products. * indicates position of the mixed peak. 1. indicates which disulfide was formed first.

In order to determine if the earlier eluting major peak in the air oxidation chromatogram

was the results of a particular topology and not necessarily disulfide connectivity, the positions of the acetamidomethyl (Acm) and trityl orthogonal protecting groups of the PnID A isomer were reversed and resulted in the production of PnID RevA (see methods), which also co-eluted with the most hydrophobic peak in the air oxidation chromatogram and resulted in no distinct separation or shouldering, indicating that the order of disulfide bond formation was inconsequential to topology in this case. The positions of the Acm and trityl orthogonal protecting groups of the PnID B isomer were then reversed and resulted in the production of PnID RevB, which unlike the PnID B isoform, appeared to completely co-elute with the major peak in the air oxidation chromatogram with the earlier retention time and displaying no distinct separation or shouldering. In order to confirm the difference in retention time by the PnID B and PnID RevB isomers, the two globular isomers were mixed at approximately equal and 1:2 concentrations. The resulting chromatograms (Supplemental Figure 1.) show distinct shouldering between the two globular isoforms.

2.3 Discussion

A comparison of the retention times of the synthetic oxidation carried out by air oxidation to that of the native isomers (previous work) indicates that the larger constituent of the native isomers is likely cross-linked in a globular fashion, as it is the less hydrophobic of the two isomers. This finding is novel as every χ/λ-conotoxin discovered to date has claimed a ribbon connectivity and synthetic isomerization to the globular form has resulted in a significant loss of activity (59,60). Additionally PnID contains a native *C*-terminal amide, a PTM which has been previously observed in the similar conotoxins λ-CmrVIA P6 amide (synthetic) and α-conotoxin ImI to be a significant contributing factor to the preferential formation of a globular disulfide connectivity (62) and likely participating in a switch mechanism for favorability between disulfide configurations. *C*-terminal amidation, although never before determined to be present in previously discovered χ-conotoxins, has been synthetically added to MrIA-NH$_2$ and was determined to slightly increase binding capacity to the NET (67). Therefore it is likely that the amidation of PnID holds a role that is two-fold, not only aiding in the formation of a globular isomer but also providing a stronger peptide-receptor interaction. Before PnID, globular forms of χ-conotoxins have not been found to have any significance and were considered to be the result of misfolding.

Further comparison between the primary sequence of PnID and that of other χ-conotoxins reveals that significant differences occur in the second and before the first inter-cysteine loops (Table 2). All currently known χ-conotoxins contain a Valine before the first cysteine in the sequence and a *4-trans*-hydroxyproline and Histidine in the final inter-cysteine loop. This common characteristic was investigated with respect to the integrity of χ-MrIA's structure activity relationships (4). After performing alanine scans it was determined that Valine

3 and *4-trans*-hydroxyproline 12 are critical residues for maintaining stability in χ-MrIA. Additionally NMR analysis provided insight to hydrophobic interactions that occur between these residues. Thus it is likely that the unique placement of Valine and Proline in PnID might aid in its lack of preference for a particular connectivity or provides a slight preference for a globular connection.

The slight shouldering of PnID B when undergoing co-elution experiments with the air oxidized material indicates that the two globular peptides, PnID B and PnID RevB are similar in hydrophobicity and thus likely similar in structure. It may therefore be possible that other globular peptides who exhibit broad peaks upon subjection to air oxidation are not producing a single isomer, but that of two topologies. Therefore it is recommended that if a broad peak results from an air oxidation, that a regio-selective synthesis be undertaken for comparison. If a peptide produces a sharp peak after a regio-selective oxidation but a broad one after air oxidation it is likely that other isomers, possibly topological, may be present.

Chapter 3: Comparative Folding of PnID and Related χ-Conotoxins

3.0 Introduction

As previously alluded to in chapters 1 and 2, χ-conotoxins are unique due to their XCCGYKLCXOC sequence (where X is any amino acid), preference for folding in a ribbon fashion (C1 paired with C4 and C2 with C3) and their selective inhibition of the NET (67). Unlike the α-conotoxins that are considerably well studied, the χ-conotoxin family consists of only a handful of peptides which have undergone in depth investigation, despite that other genetic and peptide sequences are known which may be of the χ-conotoxin family (31,32). As with the study of many of the previously discovered conotoxins the main focus of the studies done on the established χ-conotoxins has been: to determine the peptide sequence, discover what type of biological activity the toxin confers, to isolate a selective pharmacological target, to determine what type of tertiary structure the peptide has, and finally to determine what degree of stability is associated with that structure. The purpose of this section is to bridge the gap between the unique sequence and folding behavior of PnID with that of the already known χ-conotoxins and mine this information to further our understanding of why particular peptides and motifs fold more efficiently than others.

The first χ-conotoxin discoveries were made in the year 2000 and were reported simultaneously by Balaji *et al*. and McIntosh *et al*. (79,81). Although the sequences, cysteine pairing preference, and biological activity in whole animals were reported by both groups, the molecular target was not elucidated until a year later by Sharpe *et al*. (67), thus Balaji and colleagues initially named this pharmacological class of peptides λ-conotoxins indicating that they exhibited potent therapeutic potential as a symptom of injection despite not knowing the actual molecular target of these peptides. Sharpe and colleagues initially tested the χ-conotoxins MrIA and MrIB isolated from *Conus marmoreous*, with a tensor assay on a rat *vas deferens* which resulted in activity indicative of inhibition of the norepinephrine transporter and later tested the peptides directly on that molecular target using various radio-ligand binding assays (67,80) and thus the χ-conotoxin class was finally elucidated.

As previously mentioned, χ-conotoxins are not only unique due to their activity but also due to their respective folding preferences. Interestingly the χ-conotoxins have highly similar sequences to one another and do not vary too far from the α-conotoxin frameworks. Of particular note is the cysteine framework of the α-conotoxin ImI, which contains a single additional amino acid in the second inter-cysteine loop of the peptide in comparison to that of the χ-conotoxin framework (48). As referred to briefly in Chapter 1, the study conducted by Kang *et al*. took a

deeper look into the differences in folding among the highly similar disulfide frameworks of α-ImI and χ-CMrVIA in an attempt to flesh out how the disparity in these sequences contributes to either the iconic globular α-conotoxin fold or the more recently discovered ribbon preference adopted by the χ-conotoxins. The work undertaken by this group resulted in several important discoveries relating to the preference of each fold: first, it was determined that the placement of a Proline within the first inter-cysteine loop of a conotoxin sequence contributes largely to a preferential globular fold, and second it was also determined that *C*-terminal amidation plays another significant role in shifting preference towards a globular motif. These observations were elucidated through several key experiments and observations within previous literature on both families of conotoxins. These experiments included the swapping of the first proline residue in α-ImI with the corresponding lysine residue in χ-CMrVIA and synthesizing *C*-terminally amidated as well as free acid variants of each peptide. The resulting products from each synthesis were then air oxidized and run comparatively by RP-HPLC/UV to regio-selectively synthesized variants in order to pinpoint the folding preference of the air oxidized peptides. Although the results of this study are highly significant in determining the importance of the Proline residue position and *C*-terminal modification with respect to *in vitro* (and probably *in vivo*) folding inclinations, they fail to bring to light the possibility of backbone topology changes in each isomer. The missing link with this and many previous experiments is that comparative RP-HPLC/UV does not necessarily clarify slight changes in structure that may be produced through synthetically changing the order of disulfide formation within the conotoxin. Therefore it must be stressed in the future that regio-selectively synthesized products must be subject to co-elution and not simply comparative RP-HPLC/UV.

All of the currently observed χ-conotoxin folding preferences have so far been studied on an individual case by case basis and have until now proven that there is a strong preference for the ribbon fold within either the venom duct or the milked venom (79,81). Previous work however has shown that PnID, despite having a χ-conotoxin like sequence does not follow this trend as it appears in two forms after disulfide bond formation within the venom duct. Synthetic folding comparisons of each peptide to one another however have not yet been attempted. Since each conotoxin contains the same cysteine framework but slight sequence variations, differences in *N*-terminal truncation, and importantly the inclusion of a *C*-terminal modification or a free acid (PnID is the only observed χ-like conotoxin to contain an amidated *C*-terminus thus far), comparative folding of PnID to that of the established χ-conotoxins provides unique insight into folding preference and could provide further knowledge on possible χ-conotoxin motifs to the established literature. In order to effectively compare the folding preference of PnID to that of the other χ-conotoxins, *C*-terminally amidated forms of each conotoxin was folded by introduction to a moderately basic, organic, and chaotrophic environments and variables such as time and temperature were modified to study their effects on the folding patterns of each peptide.

3.1 Methods

3.1.1 Oxidative folding conditions

Each peptide was synthetically designed to contain orthogonal protecting groups that were removed upon exposure to the cleavage cocktail (see Chapter 2 methods) that ensured that free thiols would be available upon lyophilization and introduction to an *in vitro* oxidation buffer. Before undergoing folding, each peptide was >95% pure as assessed by RP-HPLC/UV. *In vitro* oxidative folding of peptides can occur by multiple pathways involving electrostatic interactions, hydrogen bonding, hydrophobic interactions, and minimization of steric hindrance, none of which ubiquitously take precedent in every peptide (39). Thus folding must be handled on case-by-case basis. In order to shed light on specific interactions occurring during the oxidative folding process each peptide was subjected to oxidative folding buffers with various conditions including changes in pH, temperature, hydrophobicity, ion pairing, and time. Table 1 (see Chapter 1) illustrates the various conditions for each buffer. PnID was also folded in a 0.1M ammonium bicarbonate buffer (pH 7.8) for time intervals spanning 24 hours to 96 hours in an attempt to elucidate which isomer would be preferred kinetically and which would be the thermodynamic isomer.

All three peptides were subjected to additional oxidative folding for 3 hours in a 0.1M ammonium bicarbonate buffer (pH 7.8) with glutathione oxidized/reduced (GSSG/GSH) at a ratio of 1:2 respectively in order to simulate *in vivo* folding conditions within the endoplasmic reticulum. Each peptide was then additionally folded for 24 hours under the same conditions to determine if equilibrium had been reached.

3.2 Results

3.2.1 Folding properties of PnID

Six out of nine oxidation conditions displayed a preference for the latest eluting isomer of PnID, which was previously determined to be the ribbon isoform (see Chapter 2) whereas conditions 1, 3, and 4 appeared to result in the favoring of the earlier eluting globular isoform (Figure 9). Conditions 1 and 4 only differ in the amount of time provided for oxidation to occur whereas condition 3 included a more hydrophobic environment, a mild chaotrope, and a change in ion pairing albeit for the same incubation period as conditions 1 and 4. The information gained from this experiment appeared to suggest that the addition of strong chaotropes interfere with the formation of the globular isomer and thus hydrogen bonding is likely a strong influence on the disulfide preference of globular PnID, with steric hindrance being a lesser factor. Time and temperature additionally appear to have a significant effect on PnID, as longer reaction times at a decreased temperature appear to result in the favoring of the ribbon connectivity, however only when a chaotrope is present. Such a result indicates that the solvent and hydrogen bonding interactions that direct the formation of the globular isoform happen at a slower rate and involve several complex steps, further providing evidence that hydrogen bond formation is key to achieving the globular fold.
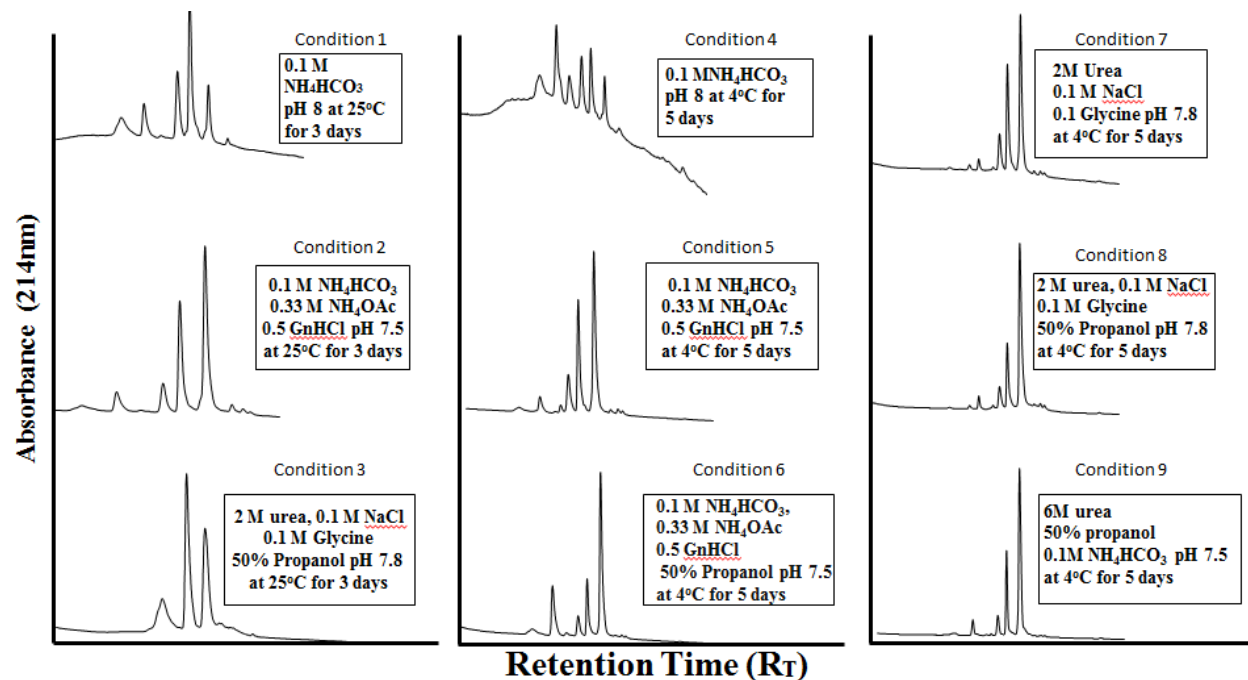
Figure 9: Air oxidations of PnID under various buffering conditions.

Figure 10 shows the *in vitro* oxidative folding of PnID over time in an ammonium bicarbonate buffer (pH 8) from a period of 24 to 96 hours. Interestingly the ribbon isomer appears to form initially and is slightly favored to that of the globular form during the first 24 hours. As time progresses, the globular form appears to achieve dominance and at 72-96 hours this dominance appears to stabilize. The shift from ribbon to globular is ostensibly either due to disulfide bond shuffling which occurs due to the reactivity of the thiols at a basic pH or due to aggregation of the ribbon isoform. The case of equilibration being achieved after disulfide shuffling at a longer time with the globular form being more highly favored would indicate that although the ribbon form appears to be the main peak initially, the globular form is likely the thermodynamic isomer and thus natively produced. However a broad peak appeared to form later than the ribbon isoform in one of the oxidations and increases in area relative to the loss of the ribbon isoform, indicating that the later hypothesis of aggregation is most likely occurring. This peak does not occur however in all cases, and may be the result of concentration changes during evaporation.
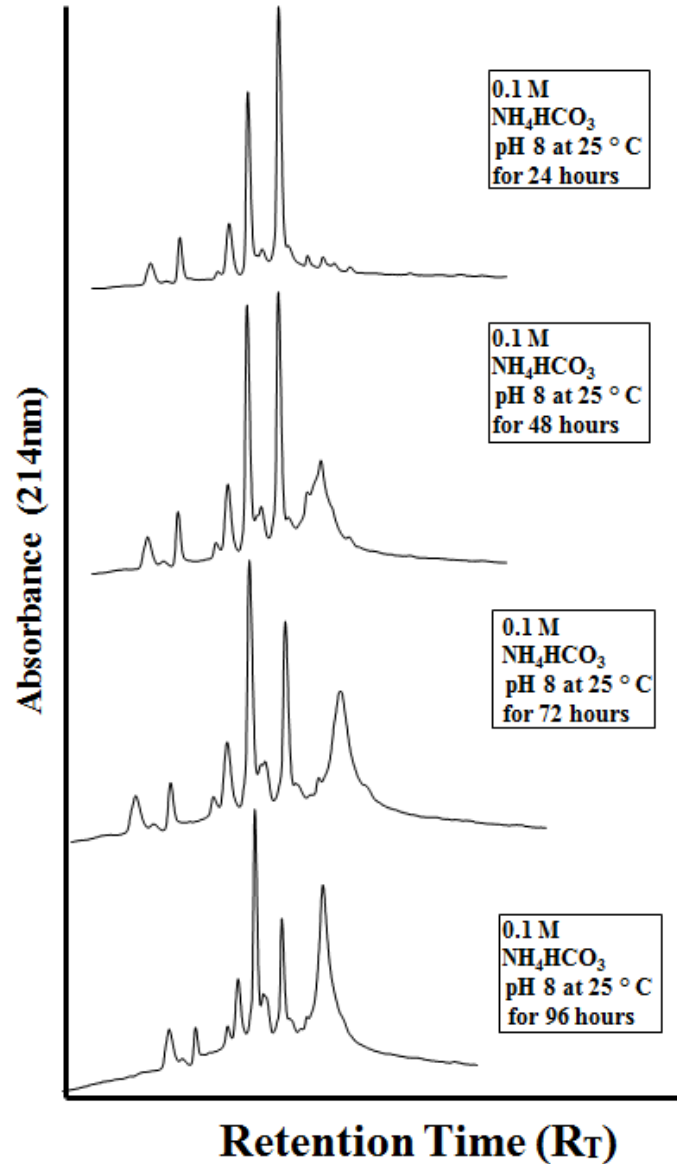
Figure 10: Air oxidations of PnID in 0.1M ammonium bicarbonate (pH 8) ranging from 24 hours incubation to 96 hours at room temperature.

Glutathione oxidation over a 24-hour period appeared to fully oxidize the peptide and produced three isomers (Figure 11). Interestingly the ribbon isomer appeared to be the thermodynamic isomer in this case, nearly doubling in size in comparison to the other isomers. Strangely however the profile did not match what was observed in the venom duct with the earlier eluting peak being larger, indicating that the native form of this peptide may require a molecular chaperone when folding.
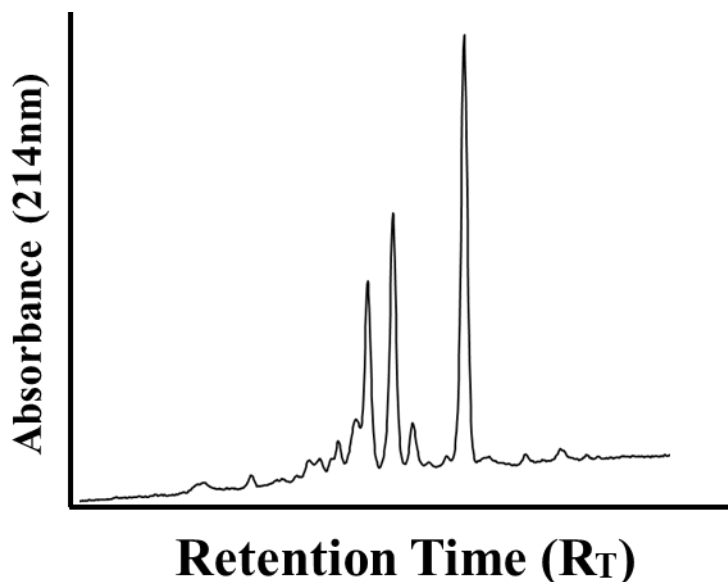
Figure 11: Air oxidation of PnID in 0.1M ammonium bicarbonate (pH 8) in 2:1 (GSH/GSSG) after 24 hours.

3.2.2 Folding properties of CmrVIA amidated

Previous research has indicated that the ribbon isoform of χ-CMrVIA elutes later than the globular isoform when subjected to RP-HPLC/UV and is the preferred connectivity of the native conotoxin (81). Under all 9 conditions the ribbon isomer of χ-CMrVIA appeared to be slightly favored (Figure 12). Unlike PnID, the ratio of the globular isomer produced to that of the ribbon is approximately 40% and 60% respectively, with the exception of conditions 3, 6, 8, and 9 where the globular isoform is <40%. In conditions 3 and 6, a third isoform, likely of the beaded connectivity, appears to be similar in area to that of the globular form, being approximately 20% of the chromatogram. The globular form of χ-CMrVIA appears to be greatly reduced by the addition of propanol, indicating that organic alcohols have a strong influence on its folding pathway. The glutathione oxidation of χ-CMrVIA also produced three isomers and favored the ribbon isoform over the other two (Figure 13).

Figure 12: Air oxidations of χ-CMrVIA under various buffering conditions.
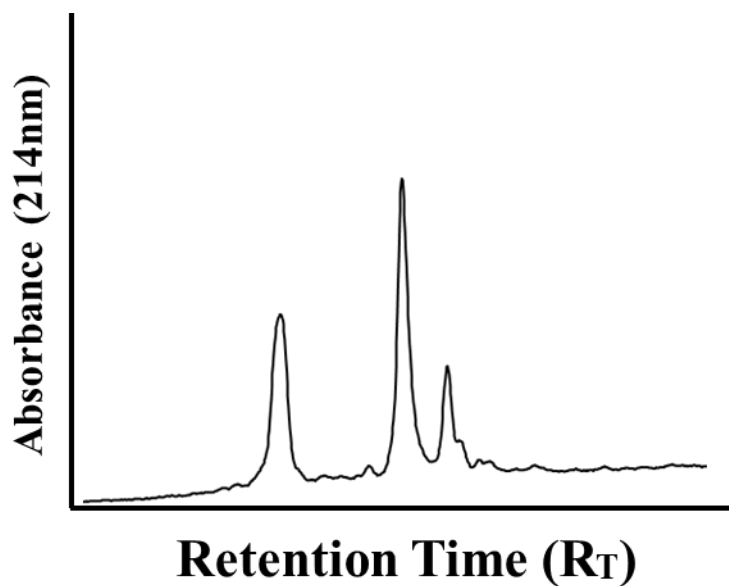
Figure 13: Air oxidation of χ-CMrVIA in 0.1M ammonium bicarbonate (pH 8) in 2:1 (GSH/GSSG) after 24-hours.

3.2.3 Folding properties of MrIA amidated

Unlike PnID and χ-CMrVIA, the ribbon conformation of χ-MrIA was strongly favored under all nine conditions (Figure 14). Since the addition of chaotropes, both strong and weak, did not significantly affect the products of χ-MrIA, it is unlikely that hydrogen bond formation plays a significant factor pre-disulfide bond formation during the folding process. The lack of significant variation among the resulting chromatograms despite changes in temperature, time, ion pairing, and chaotrope predicates steric hindrance as a significant factor in the *in vitro* oxidative folding pathway of MrIA. Figure 15 shows the glutathione folded χ-MrIA that displayed little change in comparison to the other folding conditions.
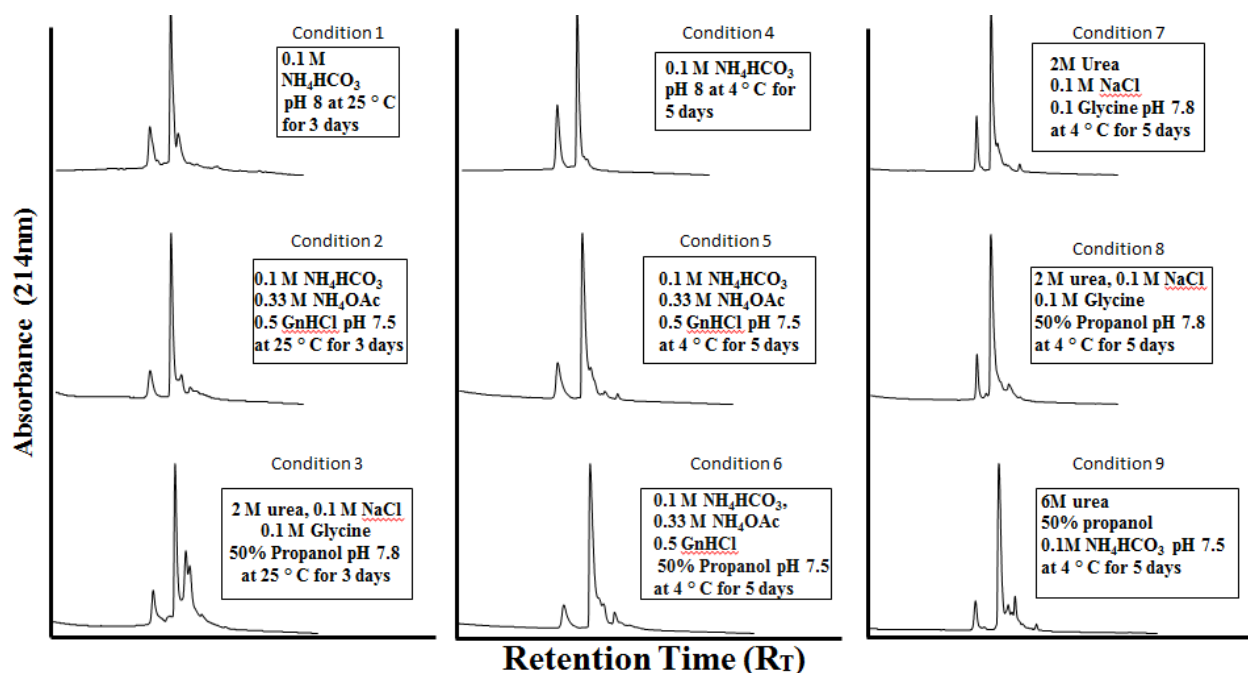
Figure 14: Air oxidations of MrIA under various buffering conditions.

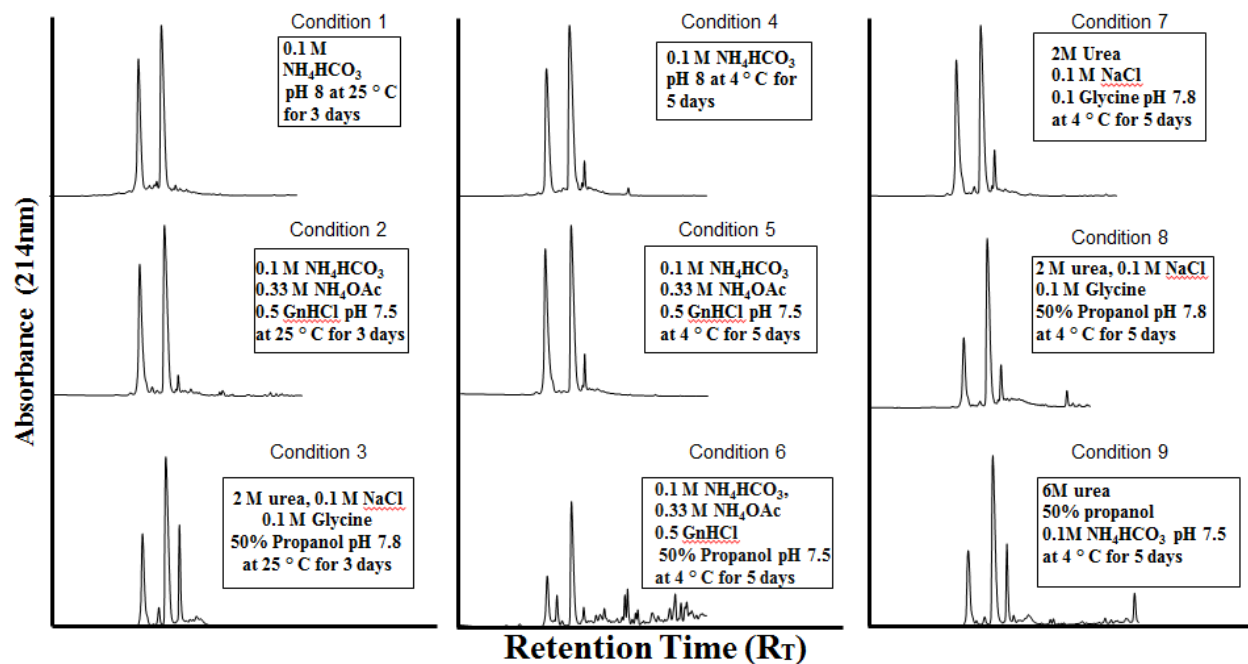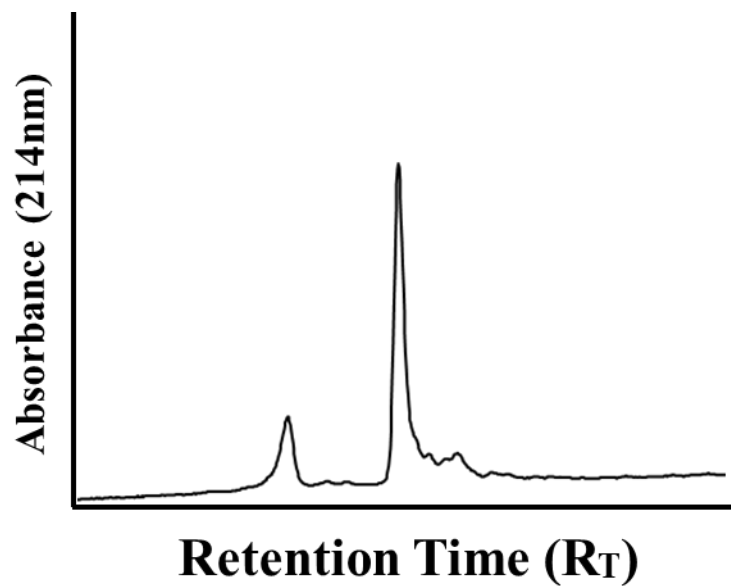**Figure 15:** Air oxidation of χ-MrIA in 0.1M ammonium bicarbonate (pH 8) in 2:1 (GSH/GSSG) after 24-hours.

3.3 Discussion

The comparative sequences of the χ-conotoxins PnID, MrIA, and CmrVIA are shown in Table 2 (Chapter 2). Despite all three toxins being highly homologous in their primary sequences, and all having the same cysteine framework, their folding properties appear to differ dramatically. χ-CMrVIA and χ-MrIA appear to favor a ribbon disulfide connectivity under all conditions, however χ-MrIA produces the ribbon form in ~ >80% of its post-folding chromatogram despite changes in conditions whereas χ-CMrVIA usually produces ~ 60% of its chromatogram in the ribbon form. Both of these peptides are in direct contrast to PnID that favors the globular form in the absence of chaotropes and organic alcohols and changes in the oxidation condition of PnID dramatically influence isomer formation. Initially when comparing the sequence of χ-MrIA to that of PnID it appears that a small amount of sequence diversity (although still largely homologous) plays a distinct role in the preferred disulfide bond connectivity of these peptides, as χ-MrIA appears to have a highly stabilized and thermodynamic ribbon isomer and PnID does not. Surprisingly however when compared to χ-CMrVIA, it appears that sequence diversity alone is not enough to explain the differences in the oxidative folding products. As seen in Table 2 the *N*-terminus of χ-MrIA is extended in comparison to that of χ-CMrVIA by a Glycine and Asparagine, which accounts for the only differences between the two toxins. Therefore this *N*-terminal truncation appears to be contributing to the loss of favorability in the ribbon isoform of χ-CMrVIA. This hypothesis is further evidenced in the previously published structures of χ-MrIA and χ-CMrVIA, as well as the structure of χ-MrIB, a χ-conotoxin that is truncated at the first Asparagine residue (Figure 16) (48,67). Therefore it may be concluded that the extended *N*-terminus of χ-MrIA helps to shift the folding mechanism of these toxins from a mixed-collapse model of oxidative folding to a more framework like model. This protection strategy however is directly contrasted by the diversified *N*-terminal region of PnID, which replaces the NGV *N*-terminal region of χ-MrIA with a ST *N*-terminus and does not appear to influence the same stability.
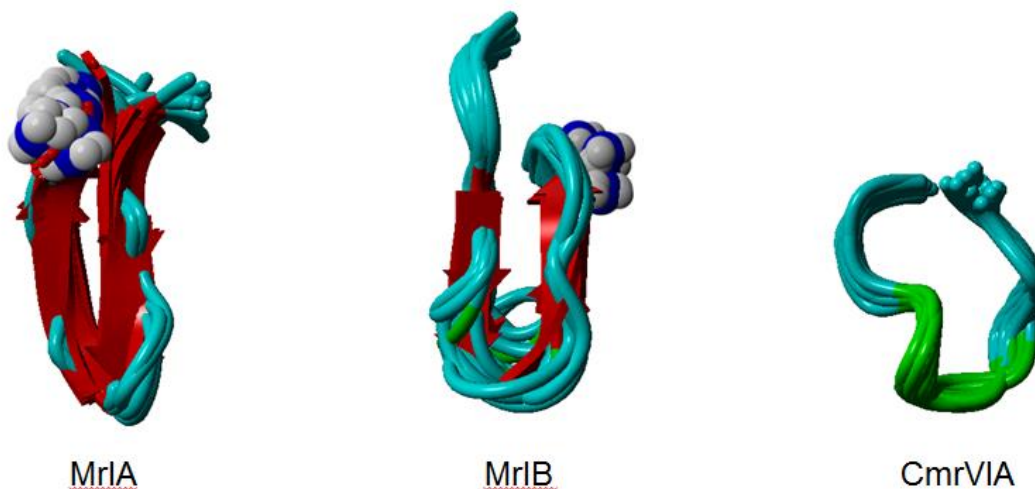
Figure 16: Comparison of the lowest 20 solution structures of the sequentially truncated χ-conotoxins MrIA, MrIB, and CmrVIA. PDB files were obtained from wwPDB.org and viewed in YASARA.

The increased preference for the globular disulfide connectivity by PnID is likely due to a combination of sequence diversity and post-translational modification. Although PnID contains the characteristic cysteine framework of a χ-conotoxin, it does not contain a hydroxylated Proline in its second intercysteine loop. The addition of an amidated *C*-terminus is likely causing a slight increase in the amount of globular isomer formed in each peptide as was previously discovered in χ-CmrVIA by Kang *et al*. (48) however the effect does appear to be slightly more significant in the sequence of PnID. The increased preference for the globular isoform could also be attributed to the unique *N*-terminus of PnID or the slight differences in the middle intercysteine loop. The center loop of PnID contains key replacements in comparison to other χ-conotoxins such as the change in the stretch GYKL in the traditional χ-conotoxins to that of GYRM in PnID. This exchange is important because of the possibility of a cation-π interaction that may occur during the folding of both toxins, where the traditional χ-conotoxins would initiate an ε-amino Lysine-π interaction, PnID would initiate a guanidinium Arginine-π interaction. This interaction would also be further complicated by the nearby sulfur in the Methionine in PnID that has the potential to participate in a favorable Methionine-π interaction, increasing the competition of side chain interactions within this region.

The combination of the results presented within this chapter and that of the previous chapter paints PnID as a unique conotoxin in the appearance of its folding preference and topologies. In Chapter 2 it was pointed out that PnID is capable of forming multiple topologies of the globular motif through orthogonal synthesis and in this chapter the globular isoform of PnID is further evidenced as a significant role carrier in the story of this toxin due to its higher preference in comparison to the other χ-conotoxins.

Chapter 4: Biological activity of PnID and Classification as a χ-Conotoxin

1.1 Introduction

Conotoxins are currently organized into 12 unique pharmacological families that represent the selective molecular targets of these peptides (31,32). The pharmacological families of conotoxins at present include the: alpha (nAChR), gamma (Neuronal pacemaker cation currents), delta (Voltage-gated Na channels, agonist with delayed inactivation), epsilon (Presynaptic Ca channels or G protein-coupled presynaptic receptors), iota (Voltage-gated Na channels), kappa (Voltage-gated K channels), mu (Voltage-gated Na channels, blocker), rho (Alpha1-adrenoceptors), sigma (Serotonin-gated ion channels), tau (Somatostatin receptor), chi (Neuronal noradrenaline transporter), and omega (Voltage-gated Ca channels) families. As previously mentioned PnID is very close in homology to the χ-conotoxins and is likely a part of this pharmacological family.

In addition to being selective to particular molecular targets, conotoxins often also show selectivity to different phyla (26). There are four known types of feeding habits that largely correlate with the phyla selectivity of particular venom constituents. These feeding habits include: moluscivorous or mollusk eating cone snails, piscivorous or fish eating cone snails, vermivorous or worm eating cone snails, and finally cone snails with mixed diets. It has been observed frequently that a cone snail of a particular diet will tailor it's venom peptides to specifically attack the receptors of an organism matching its diet. Therefore it is rare to see molluscivorous or vermivorous cone snails have venom constituents that are particularly effective on mammalian receptors and thus much of the research on the venom from these organisms is focused on that which has been derived from piscivorous organisms.

Despite the major trends in discovery and research, counter examples to the phyla selectivity dogma do exist. There have been conotoxins derived from molluscivorous cone snails that have displayed activity on mammalian receptors. Of special interest is that of the χ-conotoxins, all of which until recently have been isolated from *Conus marmoreous* (a molluscivore), and have displayed potent activity on the NET (80), which interestingly enough has not been characterized in any mollusk species to date, although an alternative transporter, likely for octopamine has been identified on the genetic level (83,84). As mentioned previously, PnID was isolated from *Conus pennaceus*, which is also molluscivorous. Due to it's homology to the χ-conotoxins it has been assumed to have similar biological activity to the peptides in this class. However, since PnID is derived from a mollusk eating cone snail it is also pertinent that the peptide be tested not only on this receptor, but for a capacity to be phyla selective.

The NET is a type III integral membrane protein consisting of 12 transmembrane domains with both termini present on the intracellular side of the membrane. It is located away from the synapse and found along the axon, cell body, and dendrites of neurons in the parasympathetic nervous system (85,86). It is a high interest target for drug development for several major disorders including: severe cancer pain, epilepsy and depression (87).

As mentioned in the introduction human therapeutics are not the only avenue pursuant of these peptides. In consideration of the combined receptor isoform and phyla selectivity of these conotoxins, pesticide research and molecular probe design have recently sparked interest. In consideration of these possible developmental routes, the potential biological activity of PnID was studied through many different assays to determine phyla selectivity in multiple organisms, activity on the NET, and activity on other receptors through extensive screening and whole squid axon assays. This work was done collaboratively between the University of Hawaii, Stanford University, Vanderbilt University, University of British Colombia, University of the Philippines (UP), and the National Institute of Health (NIH). Special thanks go to Dr. Willam Gilly, Dr. David Weaver, Dr. Tarah Klassen, Dr. Michael Baumann, Dr. Zenaida Baoanan, and Dispo Batinga for their hard work performing assays, writing, and providing insight into the biological activity of PnID.

PnID was tested on many organisms including a fish species *Poecillia reticulata*, several mollusk species including: *Achatina fulica, Cypraea caputophidii, Cellana tramoserica, Pomacea canaliculata and Jagora asperata;* and finally PnID was also tested on a small planktonic crustacean known as daphnia, which is known to be extremely sensitive to changes in water quality (88). Each of these organisms were chosen to better understand the potential biological activity of PnID in different systems.

The fish species *P. reticulata* was chosen as a model to study the effects of PnID in higher order organisms (i.e. vertebrates). *P. reticulata* serves as our model vertebrate system due to its relative ease of cultivation, the speed at which it reaches maturity, and its ideal size, which is appropriate for reaching our target doses. In contrast to the fish assays, the mollusk assays provide us with insight into how this peptide may work in nature, whether phyla selectivity between fish and mollusks occurs, and whether this peptide may provide insight to the development of a future pesticide.

The mollusk species *A. fulica* and *P. canaliculata* were chosen due to their reputations as invasive land and marine snails respectively in many tropical regions (89,90), as they are of great interest when considering molluscicide development. In contrast *J. asperata* is a snail native to the Philippines and considered economically important and therefore was used to determine the potential negative impact PnID may have on the native marine snail population (91). *C. caputophidii* is considered to be the main food source for the Hawaiian native *C. pennaceus*, the closest relative to the Red Sea *C. pennaceus*, and thus was used to better understand the effects of PnID in a native like situation. Finally *C. tramoserica* was chosen due to the species inability to retract into its shell as this action can complicate LD_{50} and PD_{50} calculations.

1.2 Methods

The assessment of the biological activity of the conopeptide PnID first required a bulk production/purification process involving several hundred milligrams of crude material of PnID A, PnID B, PnID RevB, mixed and individually isolated isomers from oxidation condition 1 (Chapter 3), and PnID C (see Chapters 2 and 3) which were all purified to result in 10-20 mg of each isomer at >95% purity. Purifications were undertaken as described in chapter 2. Once purified, samples were either quantified by weight or through comparative RP-HPLC/UV analysis and distributed for use in assays in-house or for use by collaborators.

In-house assays were performed on the fish species *Poecillia reticulata* and several mollusk species including: *Achatina fulica*, *Cypraea caputophidii*, and *Cellana tramoserica*. Fish and mollusks were used in an attempt to determine if phyla selectivity plays a role in the toxins effect, potency, and/or targeting.

1.2.1 *Poecilia reticulata* Fish LD$_{50}$ assay (In-house)

Preparation of peptide solutions

For the first dose 32 nMoles of each peptide isomer was dissolved in 20 µL of 1x phosphate buffered saline (PBS). For the second dose 24 nMoles of each peptide isomer was dissolved in 16 µL of 1x phosphate buffered saline and finally for the third dose 16 nMoles of each peptide isomer was dissolved in 5 µL of 1x phosphate buffered saline. The calculations were as follows:

A= 0.759 µMol/1mL = 0.032 µMole / x mL = 42 µL
Second dilution (0.024 µMole) = 31.6 µL/2 = 16 µL
Third dilution (0.016 µMole) = 21 µL/ 4 = 5 µL

B= 0.958 µMol/1mL = 0.032 µMole/ x mL = 33.39 µL
Second dilution (0.024 µMole) = 25.05 µL /2 = 13 µL
Third dilution (0.016 µMole) = 16.7 µL /4 = 4 µL

RevB= 0.204 µMol/1mL = 0.032 µMole/ x mL = 156.86 µL
Second dilution (0.024 µMole) = 117.65 µL /2 = 59 µL
Third dilution (0.016 µMole) = 78.43 µL/4 = 20 µL

Each isomer was concentrated by speed vacuum and made to 20 µL.

Animal preparation and injection procedure

Toxin calculations were completed before preparing animals for this study. A working solution of Tricaine (MS-222) used for humane euthanasia was prepared at a concentration of 200 mg/1L. A 1X PBS negative control solution was prepared by diluting 10X PBS. 15 *Poecilia reticulata* weighing between 400-600 mg were sacrificed for this study. 9 animals were used for each dose (three doses in triplicate), 3 animals were used for the positive control and 3 were used for the negative control. Animals were weighed in a beaker containing water from their native environment in an effort to minimize stress during the study. Injections were made using a 100 µL Hamilton syringe and were undertaken by injecting 5 µL of solution at a time. Before each injection the syringe was washed three times with bleach, three times with distilled water, and three times with 1X PBS. After injection each animal was observed for 24 hours before being subjected to euthanasia. Euthanasia was carried out by submersion in Tricaine for 30 minutes immediately following rapid freezing in liquid nitrogen. Frozen animals were then autoclaved and disposed of through biological waste protocols. LD_{50} calculations were carried out by the method of Meier and Theakston (92).

1.2.2 Mollusk PD_{50} assays (In-house)

The preparation of peptide solutions was undertaken in the same fashion as mentioned above (see Section 4.2.1). PD_{50} calculations for each organism were carried out using the method of Fainzilber and Zlotkin (93).

1.2.3 Assays performed by the University of the Philippines

Preparation of solutions

Each quantified synthetic peptide isomer was dissolved in 500 µL of normal saline solution (NSS) to produce stock solutions of final concentrations (in µM) as listed in Table 4. From the stock solutions, 100 µL aliquots of 10 µM concentration were prepared for each synthetic peptide isomer. From these aliquots, three test solutions of increasing concentrations: 1.0, 2.0, and 3.0 µM, named as molluscan biotoxicity assay (MBA) Working solutions, were then prepared and used for the bioassay (refer to Table 5).

Table 4: Preparation of 10 µM aliquots from Stock solutions. (C= concentration; V= volume). Credit to Batinga and Baoanan (UP).

ISOMER	C_{isomer} (µM)	V_{isomer} (µL)	V_{NSS} (µL)	$C_{aliquot}$ (µM)	$V_{aliquot}$ (µL)
A	759	1.5	98.5	10	100
B	338	3.0	97.0	10	100
C	317	3.5	96.5	10	100
I	170	6.0	94.0	10	100
II	712	1.5	98.5	10	100
III	12	83.5	16.5	10	100
Mixed	11	91.0	9.0	10	100

Table 5: Preparation of MBA Working solutions from 10 µM aliquots. Credit to Batinga and Baoanan (UP).

ISOMER	$C_{initial}$ (µM)	$V_{initial}$ (µL)	V_{NSS} (µL)	C_{final} (µM)	V_{final} (µL)
C1	10	10	90	1	100
C2	10	20	80	2	100
C3	10	30	70	3	100

Pomacea canaliculata and *Jagora asperata* immersion assays

Test solutions with volumes of 200 mL were prepared from the MBA Working solutions by dilution with deionised water resulting in final concentrations of 0.5, 1.0, and 1.5 nM for each of the peptide isomers. Ten snails of *P. canaliculata* were divided among two glass jars each containing 50 mL of a test solution that were sealed with a mesh net to prevent escape of the snails. This set-up was followed for all of the twenty-one test solutions that are three different concentrations for each of the seven synthetic peptide isomers. The same set-up applied to snails of *J. asperata*. For the negative control, snails were immersed in deionised water and for the positive control; snails were immersed in solution that contained *Snailkill* metaldehyde pellets. All the snails were submerged in the glass jars for a total of 72 hours with the temperature maintained at 21°C. Snail mortality was then determined at the end of this period by assessment of three criteria: retraction reflex of the foot of the snail following tactile stimulation, probing of the operculum, and detachment from the vertical surface of the container. The percentage of dead snails for each set-up was then recorded. Death of the organism is determined by immobility, paleness of color or unresponsiveness to tactile stimulation after the shells were cracked with hole on the first body whorl relative to the aperture.

***Pomacea canaliculata* and *Jagora asperata* injection assays**

For each peptide isomer with three different concentrations as used above, test solution was injected intramuscularly into the foot pedal of *P. canaliculata* and *J. asperata* with a volume of 30 µL with the use of a Hamilton microliter syringe. For the negative control, snails for both species were injected with normal saline solution (see Appendix C). Each set-up consisted of three snails. Similarly, undiluted test solutions for each peptide isomer with the maximum concentrations as listed in Table 4 were also tested with the same procedure. The snails were then observed for physical responses within an hour immediately following injection (94). Snail mortality was then determined at the end of a 72-hour period by assessment using the criteria previously described. The percentage of dead snails for each set-up was also recorded. Death of the organism is determined by immobility, paleness of color or unresponsiveness to tactile stimulation after the shells were cracked with hole on the body whorl.

Daphnid Biotoxicity assay

The daphnid biotoxicity assay was done by first filling around 14 mL of natural water used for culturing previously tested for physicochemical properties to each of the forty-one test tubes. The test tubes were capped with mesh net and sealed with a masking tape and were acid-washed with 10% v/v HCl and rinsed with dechlorinated water. Thirty-five test tubes were used, that is seven isomers with five concentrations per isomer, and filled with the test solutions as described previously on Table 5 of the Preparation of Test Solutions for Daphnid Biotoxicity Assay. Additional six test tubes were used for positive and negative control triplicates. Each of the following solutions, the isomer dilutions and controls, contained ten cultured daphnid juveniles 2.0-2.5 mm in length as suggested by Kreutzer and Lampert (95). The negative control used the previously tested dechlorinated water while the positive control utilized two dissolved *SnailKill* metaldehyde pellets for each control test tubes. The set-up was maintained with a photoperiod of 16:8 hours +/- 1 hour of light and dark cycles for 48 hours. Results were taken into account after the two-day cycle.

Statistical analysis

All the statistical analyses in this experiment were carried out using IBM SPSS Statistics software (Version 20). Statistics at 95% fiducial limits of upper confidence limit and lower confidence limit was used. Statistical analyses for mortality of *P. canaliculata* and *J. asperata* at increasing concentrations of synthetic peptide solution were carried out using Two-way and Three-way ANOVA while mortality of *D. magna* was assessed using the nonparametric Independent-Samples Kruskal-Wallis Test.

1.2.4 Norepinephrine transporter assay (done in collaboration with the NIH)

Animal preparation

Male Sprague-Dawley rats (Envigo, Frederick, MD, USA) weighing 250-350 g were housed three per cage with free access to food and water and maintained on a 12 hour light/dark cycle with lights on from 7:00 a.m. to 7:00 p.m. Animal facilities were accredited by the Association for the Assessment and Accreditation of Laboratory Animal Care, and procedures were carried out in accordance with the Institutional Animal Care and Use Committee and the National Institutes of Health guidelines on care and use of animal subjects in research.

Uptake Assays

[^{3}H]Dopamine, [^{3}H]norepinephrine, and [^{3}H]5-HT (specific activity ranging from 30-50 Ci/mmol) were purchased from Perkin Elmer (Shelton, CT, USA). All other chemicals and reagents were acquired from Sigma-Aldrich (St. Louis, MO, USA). Rats were euthanized by CO_2 narcosis, and brains were processed to yield synaptosomes as previously described (Baumann et al, 2013). Rat caudate tissue was used for DAT assays, whereas rat whole brain minus caudate and cerebellum was used for NET and SERT assays. Transport activity at DAT, NET and SERT was assessed using 5 nM [^{3}H]dopamine, 10 nM [^{3}H]norepinephrine and 5 nM [^{3}H]5-HT, respectively. The selectivity of uptake assays was optimized for a single transporter by including unlabeled blockers to prevent uptake of [^{3}H]transmitter by competing transporters. Uptake inhibition assays were initiated by adding 100 µL of tissue suspension to 900 µL Krebs-phosphate buffer (126 mM NaCl, 2.4 mM KCl, 0.83 mM $CaCl_2$, 0.8 mM $MgCl_2$, 0.5 mM KH_2PO_4, 0.5 mM Na_2SO_4, 11.1 mM glucose, 0.05 mM pargyline, 1 mg/mL bovine serum albumin, and 1 mg/mL ascorbic acid, pH 7.4) containing test drug and [^{3}H]transmitter. Assays were terminated by rapid vacuum filtration through Whatman GF/B filters, and retained radioactivity was quantified by liquid scintillation counting. Statistical analyses were carried out using GraphPad Prism (v. 6.0; GraphPad Scientific, San Diego, CA, USA). IC_{50} values for uptake inhibition were calculated based on non-linear regression analysis.

1.3 Results

1.3.1 *P. reticulata* Fish LD$_{50}$ assay (*In-house*)

Injections at the highest dosage (115 µg/g) demonstrated no lethality in any isomer and no significant changes in behavior were observed in comparison to the negative control.

1.3.2 *C. tramoserica* Mollusk PD$_{50}$ assay (*In-house*)

Injections at the highest dosage (115 µg/g) demonstrated no consistent paralysis in any isomer and no significant changes in behavior were observed in comparison to the negative control. Despite this some qualitative observations were made about the animals injected.

The animals injected with Isomer A did not appear to elicit too significant of a response and the first two displayed a small amount of invagination near the injection site. The third organism injected with isomer A would not cling to substrate at all.

The animals injected with Isomer B appeared to display a great amount of invagination of the foot nearly exceeding 50% of the foot's surface. The response was slower when attempting to cling to substrate however 2 of the three animals were able to fully cling to the substrate after 10 minutes.

The animals injected with Isomer RevB also displayed mixed results. For all the animals the invagination near the injection site was severe like that of PnID B. The first animal was unable to cling to the substrate and was considered paralyzed. The second and third animals were able to cling to substrate after 30 minutes but the noticeable effects of invagination also occurred. Each organism also demonstrated a loss of water and curling of the foot.

1.3.3 *A. fulica* Mollusk PD$_{50}$ assay (*In-house*)

Injections at the highest dosage (115 µg/g) demonstrated no consistent paralysis in any isomer and no significant changes in behavior were observed in comparison to the negative control. Despite this some qualitative observations were made about the animals injected.

1.3.4 *C. caputophidii* Mollusk PD$_{50}$ assay (*In-house*)

Injections at the highest dosage (115 µg/g) demonstrated no consistent paralysis in any isomer and no significant changes in behavior were observed in comparison to the negative control.

1.3.5 *Pomacea canaliculata* immersion assay (*done in collaboration with UP*)

Figure 17 and Figure 18 illustrate the results obtained for both species from the bioassay using the immersion method after a period of 72 hours. For *P*. *canaliculata* (Figure 34), results indicate that at a concentration of 0.5 nM, only isomers B, C, and M (Mixed) yielded % mortalities of 10%, 10%, and 20% respectively. At a concentration of 1.0 nM, % mortalities obtained were: isomer A with 20%, isomer B with 10%, isomer I (equivalent to PnID RevB) with 20%, isomer II (equivalent to PnID A) with 10%, and isomer III with 20%. At a concentration of 1.5 nM, % mortality for isomer A was at 20%, isomer C was at 70%, and isomer M was at 10%. For the negative control (deionized water), 10% mortality was obtained; for the positive control (metaldehyde), 100% mortality was obtained.

Figure 17: Percentage of total mortality of P. canaliculata under three varying concentrations for each synthetic PnID isomer tested via immersion method. NC – negative control; PC – positive control. Correction of the mortality data from the isomers was not performed since the result obtained from the negative control was not greater than 10%. Credit to Batinga and Baoanan (UP).

Statistical analysis of the mortality data was performed after percentage values were transformed into arcsine values. A two-way ANOVA was used to analyze the mortality of *P. canaliculata* across two independent variables: isomer, and concentration, and the interaction between the two. Testing the null hypothesis that states that there is no significant difference in the means of mortality among the snails across the type of isomer used, a significance value of 0.317 was obtained, indicating that the null hypothesis is retained. Similarly, when the null hypothesis stating that there is no significant difference in the means of mortality among the snails across the concentration used was tested, a significance value of 0.758 was obtained, thereby, this null hypothesis is also accepted. However, when the means of mortality of the snails were analyzed with regards to the variables of isomer and concentration taken together, a significance value of 0.011 was obtained. This indicates that an interaction exists between the two variables results and it results in a significant difference in mortality among the snails. These are summarized below in Table 6. Among these, isomer C at a concentration of 1.5 nM yielded the highest mortality among snails of *P. canaliculata*.

Table 6: Two-way ANOVA test summary for mortality data of *P. canaliculata*. Credit to Batinga and Baoanan (UP).

Source	SS	df	MS	F	P
isomer	1655.28	6	275.88	1.26	**0.31**
concentration	122.62	2	61.31	0.28	**0.75**
isomer * conc.	8160.54	12	680.04	3.10	**0.01**

SS = Sum of Squares, df = Degrees of Freedom, MS = Mean Square, F = F statistic, P = P value

1.3.6 *Jagora asperata* immersion assays (done in collaboration with UP)

Results indicate (Figure 35) that at a concentration of 0.5 nM, only isomers C, II (eq. to PnID A), and III each yielded % mortalities of 10%. At a concentration of 1.0 nM, % mortalities obtained were: isomer B with 20%, isomer C with 10%, isomer I (eq. to PnID RevB) with 20%, isomer II with 10%, isomer III with 20%, and isomer M with 10%. At a concentration of 1.5 nM, % mortality for isomer A was at 10%, isomer B was at 20%, isomer C was at 10%, isomer I was at 40%, isomer II was at 10%, isomer III was at 30%, and isomer M was at 20%. For the negative control (deionized water), zero mortality was obtained; for the positive control (metaldehyde), 100% mortality was obtained.
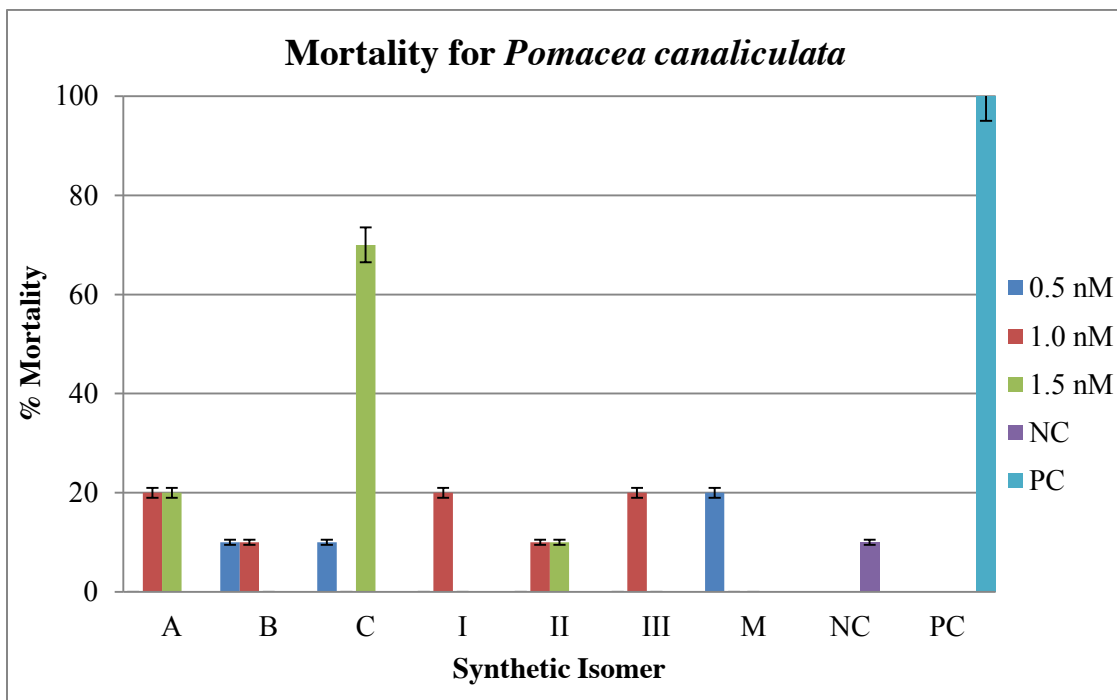
Figure 18: Percentage of total mortality of *J. asperata* under three varying concentrations used for each synthetic PnID isomer tested via immersion method. NC – negative control; PC – positive control. Credit to Batinga and Baoanan (UP).

Statistical analysis of the mortality data was performed after percentage values were transformed into arcsine values. A two-way ANOVA was used to analyze the mortality of *J. asperata* across two independent variables: isomer, and concentration, and the interaction between the two. Testing the null hypothesis that states that there is no significant difference in means of mortality among the snails across the type of isomer used resulted in a significance value of 0.289; hence, the null hypothesis is retained. On the other hand, when the null hypothesis stating that there is no significant difference in means of mortality among the snails across the concentration used was tested, a significance value of 0.013 was obtained, thereby, this null hypothesis is rejected. There is a significant difference in means of mortality among the snails depending on the concentration by which the isomer was administered. Among the three, mortality was highest at 1.5 nM. When the means of mortality of the snails were analyzed with regards to the variables of isomer and concentration taken together, a significance value of 0.729 was obtained. This indicates that there is no interaction between the two variables that can result in a significant difference in the means of mortality of the snails. These are summarized below in Table 7.

Table 7: Two-way ANOVA test summary for mortality data of *J. asperata*. Credit to Batinga and Baoanan (UP).

Source	SS	df	MS	F	P
isomer	1525.78	6	254.29	1.32	**0.28**
concentration	2080.23	2	1040.11	5.42	**0.01**
isomer * conc.	1624.80	12	135.40	0.70	**0.72**

SS = Sum of Squares, df = Degrees of Freedom, MS = Mean Square, F = F statistic, P = P value

1.3.7 Comparative statistical analysis of the immersion assays (done in collaboration with UP)

A three-way ANOVA was also conducted to assess the mortality of the snails across three independent variables, namely isomer, concentration, and species (*P. canaliculata* or *J. asperata*), and the combinations of interactions among the three. A significance value of 0.635 suggests that there is no significant difference in the means of mortality in the snails across the type of isomer used. Similarly, a significance value of 0.309 suggests that there is no significant difference in the means of mortality in the snails of *P. canaliculata* and *J. asperata*. In contrast, a significance value of 0.030 indicates that there is a significant difference in the means of mortality in the snails across the concentration used. Mortality is highest at 1.5 nM. Among interactions between variables, a significance value of 0.100, 0.110, and 0.227 suggest that there is no interaction between isomer and concentration, between isomer and species, and between concentration and species respectively which can result in differences in mortality. However, a significance of 0.025 reveals that there is an interaction among the variables of isomer, concentration, and species that results in a significant difference in the means of mortality in the snails. These are summarized below in Table 8. The highest mortality was observed among *P. canaliculata* at a concentration of 1.5 nM for Isomer C.

Table 8: Three-way ANOVA test summary for mortality data of the snails. Credit to Batinga and Baoanan (UP).

Source	SS	df	MS	F	P
isomer	887.52	6	147.92	0.72	**0.63**
concentration	1573.03	2	786.51	3.83	**0.03**
species	217.67	1	217.67	1.06	**0.30**
isomer*conc.	4204.99	12	350.41	1.70	**0.10**
isomer*species	2293.55	6	382.25	1.86	**0.11**
conc.*species	629.83	2	314.91	1.53	**0.22**
isomer*conc.*species	5580.35	12	465.02	2.26	**0.02**

SS = Sum of Squares, df = Degrees of Freedom, MS = Mean Square, F = F statistic, P = P value

1.3.8 *Pomacea canaliculata* **injection assay (done in collaboration with UP)**

Table 9 summarizes the qualitative data collected from the bioassay performed on *P. canaliculata* using the injection method. Observations were made and recorded at the following periods: immediately after injection, after 30 min, after 1 hour and the next hour that followed, and finally, after 24 hours.

Table 9 shows that a general reaction to injection was the contraction of the foot pedal and closing of the opercula, however, the speed of which slightly varied among the snails. After half an hour, the opercula appeared to be tightly shut in all the test organisms with some noticeably secreting copious amounts of mucous. After an hour, the opercula have already opened but in varying degrees: some only slightly, others partially, and some completely opened with the foot pedal once again exposed. For good measure, observations were also noted after 2 hours with some snails beginning to exhibit active movement. After the cut-off of observation at 24 hours, mortality was tested; however, all snails were still alive with only a few still dormant. Lethality of the isomers was not established based on the results, given the observation period set for this bioassay.

Table 9: General observations on the behavioral response of *P. canaliculata* to test solutions administered via injection at various times of observation. Credit to Batinga and Baoanan (UP).

Isomer	Qualitative Response/ Behavior in *P. canaliculata*				
	Upon Injection	**At 30 mins.**	**At 1 hr.**	**At 2 hrs.**	**At 24 hrs.**
A	Contraction of foot pedal; Delayed retraction and closing of operculum	Operculum completely and receded into shell	Slightly opened operculum	Slightly opened operculum	Dormant
B	Contraction of foot pedal; Delayed retraction and closing of operculum	Operculum completely and receded into shell; mucous secretion	Slightly opened operculum	Slightly opened operculum	Dormant
C	Contraction of foot pedal; Average retraction and closing of operculum	Operculum completely and receded into shell; mucous secretion	Completely opened operculum; exposed foot pedal	Active movement	Alive
I	Contraction of foot pedal; Abrupt retraction and closing of operculum	Operculum completely and receded into shell	Slightly opened operculum	Completely opened operculum; exposed foot pedal	Alive
II	Contraction of foot pedal; Average retraction and closing of operculum	Operculum completely and receded into shell; mucous secretion	Slightly opened operculum	Active movement	Alive
III	Contraction of foot pedal; Delayed retraction and closing of operculum	Operculum completely and receded into shell; mucous secretion	Partially opened operculum	Active movement	Alive
M	Contraction of foot pedal; Average retraction and closing of operculum	Operculum completely and receded into shell	Partially opened operculum	Active movement	Alive

1.3.9 *Jagora asperata* injection assay (done in collaboration with UP)

Table 10 shows that a general reaction to injection was the contraction of the foot pedal and abrupt retraction and closing of the opercula. After half an hour, some of the opercula were slightly and partially opened while some remain shut. After an hour, most snails began actively moving while a few were dormant still. At the second hour, all snails have demonstrated active movement. After the cut-off of observation at 24 hours, mortality was also tested; however, all snails were still alive. Similar to *P. canaliculata*, lethality of the isomers was not established based on the results within the 24-hour observation period.

Table 10: General observations on the behavioral response of *J. asperata* to test solutions administered via injection at various times of observation. Credit to Batinga and Baoanan (UP).

Isomer	Qualitative Response/ Behavior in *J. asperata*				
	Upon Injection	At 30 mins.	At 1 hr	At 2 hrs	At 24 hrs
A	Contraction of foot pedal; Abrupt retraction and closing of operculum	Closed operculum	Dormant	Active movement	Alive
B	Contraction of foot pedal; Abrupt retraction and closing of operculum	Partially opened operculum	Active movement	Active movement	Alive
C	Contraction of foot pedal; Abrupt retraction and closing of operculum	Slightly opened operculum	Active movement	Active movement	Alive
I	Contraction of foot pedal; Abrupt retraction and closing of operculum	Closed operculum	Dormant	Active movement	Alive
II	Contraction of foot pedal; Abrupt retraction and closing of operculum	Partially opened operculum	Active movement	Active movement	Alive
III	Contraction of foot pedal; Abrupt retraction and closing of operculum	Closed operculum	Dormant	Active movement	Alive
M	Contraction of foot pedal; Abrupt retraction and closing of operculum	Completely opened operculum	Active movement	Active movement	Alive

1.3.10 Daphnid Biotoxicity assay (done in collaboration with UP)

A biotoxicity assay using the seven peptide isomers was also performed on another test organism *Daphnia magna*. A different experimental design was used for the assay wherein five different concentrations (increasing geometrically) were tested for each isomer instead. Figure 19 illustrates the results obtained after 48 hours of immersion. At a concentration of 3.1 nM, only isomer III yielded a result, which is 10% mortality. At 6.25 nM, only isomer A yielded a result: 20% mortality. At 12.5 nM, only isomer A too yielded a result but at 10% mortality. At 25.0 nM, isomers C, I, and III each yielded 10% mortalities. At 50.0 nM, isomers II, III, and M each

yielded 10% mortalities as well. For the negative control (tap water), zero mortality was obtained; for the positive control (metaldehyde), 53% mortality was obtained.
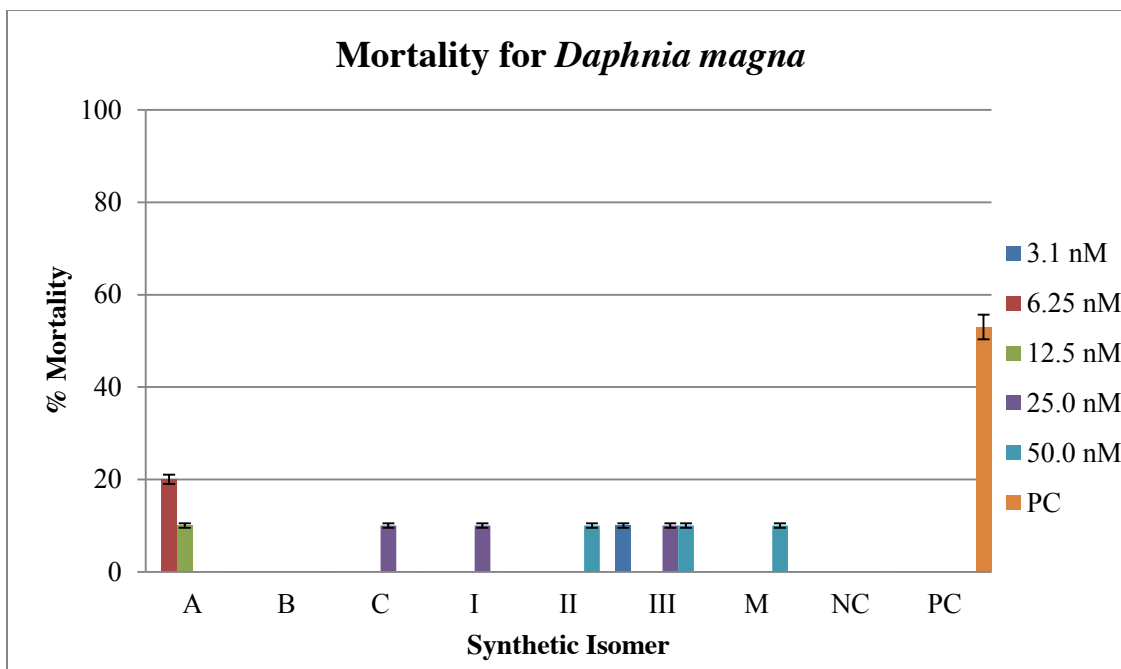

Figure 19: Percentage mortality of *D. magna* under five varying concentrations used for each synthetic PnID isomer tested in the biotoxicity assay: NC – negative control; PC – positive control.

Statistical analysis of the mortality data was performed after percentage values were transformed into arcsine values. Preliminary statistical tests for assumptions of normality and homoscedasticity show that the data do not satisfy these assumptions, hence, a nonparametric test, Independent-Samples Kruskal-Wallis Test was performed. Testing the null hypothesis that states that the distribution of mortality is the same across categories of isomer used, a significance value of 0.472 was obtained, indicating that the null hypothesis is retained. Similarly, when the null hypothesis stating that the distribution of mortality is the same across categories of concentration used was tested, a significance value of 0.556 was obtained, thereby, this null hypothesis is also accepted.

1.3.11 Norepinephrine transporter assay (done in collaboration with the NIH)

Inhibition of the monoamine transporters by the A and B (as well as reverse) isomers of PnID in comparison to that of cocaine indicated that the most potent and selective isomer of PnID is that of the A (ribbon) isoform (Figure 20). This is unsurprising as it was mentioned earlier in this chapter and in previous chapters that the χ-conotoxins are known for their ribbon fold and selectivity for the NET. Surprisingly the potentcy of PnID A appears to be nearly 20 fold weaker than MrIA as its IC_{50} was calculated to be 10 +/- 1.75 µM. Another surprising discovery was that the PnID B isoform was able to inhibit each monoamine transporter non-selectively with a smilar potency at low doses whereas PnID RevB appeared to lose activity altogether.

Follow up assays (Figure 21) indicated that the potency for PnID B is significantly less than that of the known χ-conotoxins as the estimated IC_{50} of it is >100 µM at the NET and even higher for that of the other monoamine transporters. Thus it appears that at the monoamine transporters the most relevant peptide of consideration is the A isoform.

Figure 20: Uptake plots showing the comparative dose-response inhibition of the monoamine transporters by PnID A, PnID B, PnID RevB, and cocaine at concentrations up to 10 μM.

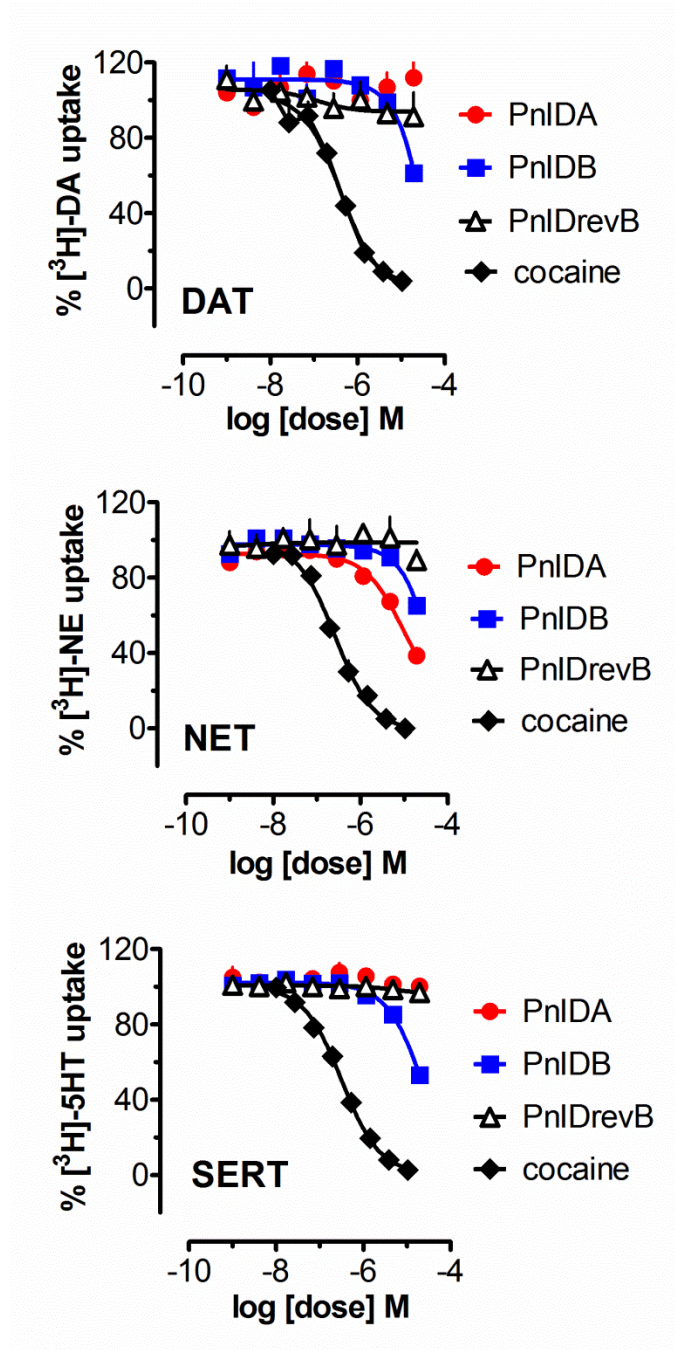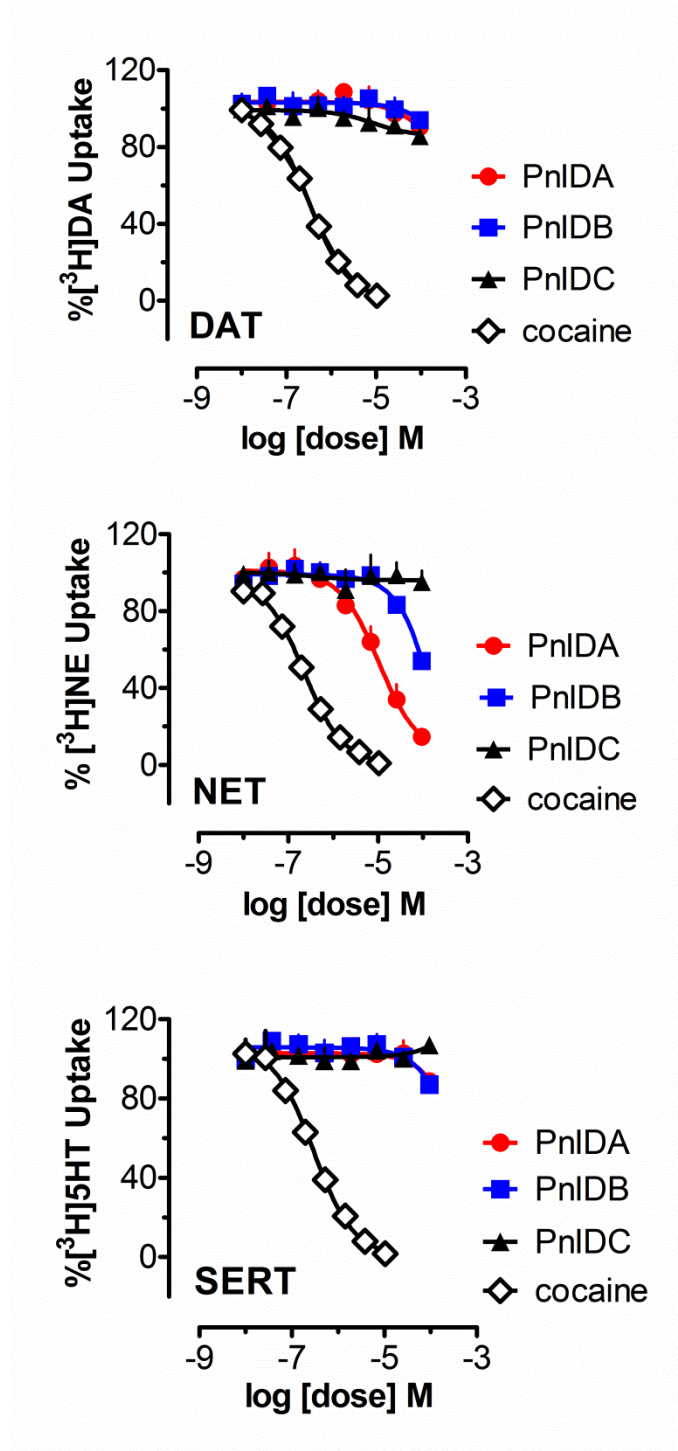**Figure 21**: Uptake plots showing the comparative dose-response inhibition of the monoamine transporters by PnID A, PnID B, PnID C and cocaine at concentrations up to 100 μM.

1.4 Discussion

As previously mentioned, χ-conotoxins are of interest because of their high selectivity for the NET. The NET is just one of several human monoamine transporters which has been implicated in the pain transduction pathway, as a key regulator of orthostatic intolerance and seizures, as well as having a significant role in depression and addiction (87). Since the NET has so many roles in the human parasympathetic nervous system it is crucial not only to study the ion channel and its related physiology but environmental regulators which can significantly affect its function and thus human disease and behavior. Therefore PnID, a conopeptide which not only contains a significant amount of homology to the known χ-conotoxins but also demonstrates unique folding behavior could be considered low-hanging fruit when considering the depth of information obtainable from its study.

In an attempt to assess the potential of PnID as a human therapeutic or environmental pesticide it was tested on several economically important and invasive snail species, *Daphnia magna,* as well as fish and the monoamine transporters. The LD_{50} and PD_{50} snail assays undertaken by injection yielded very inconsistent results qualitatively and provided no real quantitative data, especially in the case of the organisms capable of shell retraction. The two main issues encountered in this assay were the inconsistency of the organism's response to injection and the erratic times at which the snails would extend out of their shells. Since the assays relied on the snails' ability to cling to substrate, full or partial retraction created significant difficulty when measuring paralysis. Another significant hurdle encountered during the injection assays was confirming penetration of the needle when performing the injections. In contrast to fish, it is difficult to tell when the needle penetrates the snail membrane because of its lack of cartilage or skeletal structure that allows it to flex readily with the pressure of the needle. Lastly the mucus produced by the snails may aid in reducing the potency of the peptides by absorption into its glycoprotein network or through the introduction of humoral immune responses (96).

The fish injection assay also yielded minimal results when injected with relatively high doses of peptide. This result may be explained by two potential reasons. First *Poecilia reticulata* may not be affected by PnID in particular due to a lack of homology in their corresponding monoamine transporters to that of the human or snail counterparts. Secondly, the results of the IC_{50} determination of PnID on the monoamine transporters display a micromolar potency for even the most active isomer, indicating that standard concentrations may not be able to induce lethality or even measurable paralysis.

Interestingly the immersion assays were much more successful in yielding quantitative data on the activity of the PnID isomers. Isomer C appeared to significantly increase mortality of the invasive sea snail *P. canaliculata* in a dose-response manner. This result is significant

because it is the first reported instance of biological activity in a peptide toxin restrained by vicinal disulfide bonds.

The activity of the isomers of PnID on the monoamine transporters yielded fairly expected results. Out of all the isomers PnID A appeared to be the most selective for the NET with an IC_{50} of 10.33 (+/- 1.75) μM in comparison to the DAT and SERT to which it was inactive up to concentrations of 100 μM. Since PnID A is the ribbon isomer and all known χ-conotoxins exist in such a form, this result was not surprising. PnID RevB and PnID C displayed almost no inhibition of any of the monoamine transporters while PnID B appeared to yield a small amount of inhibition on the NET and no activity on any other transporter at higher doses.

Study of the ribbon conformation of MrIA has yielded pertinent information in the past about structural requirements for the maintenance of activity on the NET, Particularly the conservation of an inverse γ turn between the residues Gly 6 – Lys 8 (4). Since these residues are homologous in the sequence of PnID, (Gly 5 – Arg 7) it is likely that the ribbon conformation of PnID yields a similar γ turn. Interestingly a small amount of inhibition of the NET also occurred with the globular topology PnID B but not in PnID RevB, which may be indicative of a similar turn forming in only the PnID B topology.

Past study of MrIA by alanine scan has also shown that replacement of the Lys 8 residue with an arginine resulted in less binding capacity to the NET. This finding could explain the lower activity of PnID A as this replacement occurs in it's native sequence. This replacement is interesting however as *Conus pennaceus* is a molluscivore and as stated in the beginning of the chapter, monoamine like receptors have yet to be fully determined in mollusks. Therefore a residue replacement in this position could produce a more favorable binding to particular mollusk receptors that have not yet been fully elucidated.

The slight activity on the NET by PnID B and lack of activity of PnID RevB presents an interesting case that topological isomerization can produce differential biological activities and thus likely different overall structures. To the best of our knowledge, globular topological isomerization of small peptides containing two disulfides is a property that has not yet been reported in literature. Although disulfide bond connectivity is often reported in many articles, the order of disulfide bond formation by two step synthesis is not. The results demonstrated in this section show that disulfide bond topology which can be manipulated by the order of bond formation may result in weaker or less selective biological activity of a peptide. Therefore structure determination of disulfide oxidized peptides presents a key factor in confirming the appropriate conformation of these peptides.

Clear and seemingly contradictory differences arose between the results yielded from the mollusk based assays and that of the receptor based assays. PnID C appeared to have a significant effect on the mortality of the invasive snail *P. canaliculata* but did not appear to affect any of the monoamine transporters. These results, although seemingly contradictory at first

glance make sense when considering the molecular physiology of mollusks. As previously mentioned, mollusks do not contain traditional monoamines such as dopamine or norepinephrine but instead utilize the catecholamines tyramine and octopamine for their neuronal signal transduction processes (83,84). Additionally gene sequences which share homology to transporters for catecholamines have been partially identified for some mollusks, which may suggest that PnID C could be an inhibitor of such a transporter.

Although it is not as pharmacologically potent as its χ-conotoxin cousins, PnID contains many interesting facets within its' variable structure and shares a similar IC_{50} to that of some currently approved FDA drugs such as modafinil, a treatment for narcolepsy (97). Thus further research and development of PnID may be beneficial when considering the development of similar drugs.

The results from this chapter show for the first time that the alternative disulfide bonding strategies of a χ-conotoxin not only adapt to structural topologies previously unconsidered by peptide folding theory, but that these topologies have differences in potency and selectivity. Furthermore PnID has shown to produce the first biologically active peptide toxin containing vicinal disulfides and this activity was demonstrated in an invasive snail species. Thus the information presented within this chapter could have significant implications for not only the pharmaceutical industry but also in pesticide development.

Chapter 5: Solution Structure of the Topological Isomers of PnID

1.5 Introduction

In biochemistry structure and function are considered to be intimately related. When the structure of a biological molecule is perturbed even slightly, there are often significant consequences for the activity of that molecule (98). This theme can be observed across many different biological molecules including proteins, peptides, DNA and RNA. Due to the ubiquitous nature of the structure-activity relationship, the structures of many biological molecules have been vigorously studied over the last century in order to glean further information about the properties of these molecules and their associated motifs. Two methods of structure characterization of biological molecules have taken precedence in the field of structural biology which include: X-ray crystallography and nuclear magnetic resonance spectroscopy (5).

X-ray crystallography is a technique which exploits the relationship between the scattering angles of X-rays and the size of a unit cell in a crystal to determine spatial relationships of crystallized molecules (99). This method of structure determination has resulted in the elucidation of countless protein structures over the last century including the famous DNA double helix (100). Although this method has provided a vast amount of knowledge on biomolecular structure over the years it is not without its limitations. Since the principle of X-ray crystallography relies on determining the size of a unit cell in a crystal by scattering, the molecule of interest must first be successfully crystallized in order to achieve sufficient information for structure determination. Crystallization can be a complex process, as the route to crystallization is often largely different among proteins and peptides (5). Thus it is not unheard of to have hundreds of trials of attempted crystallization before achieving results suitable for structure determination. It could also be considered disadvantageous that crystallography relies on the material of interest to be held in a solid state. Often crystallization, although yielding an accurate structure in the crystalline state, will not provide information on how the molecule would behave in solution, which could be of considerable disadvantage if that molecule contains multiple conformations in solution, or if effects on solvation wish to be studied.

In contrast NMR spectroscopy provides an avenue to study the solution states of biological molecules. Using NMR to study bio-molecular structure is a relatively new practice in comparison to crystallography as the advent of the techniques necessary for full structure elucidation began in the year 1977 (5). One of the most significant techniques to be developed around this time and often considered to be the most crucial is the use of the Nuclear Overhauser Effect (NOE) to study interatomic differences in a molecule. Although the NOE was an observed phenomenon in 1962 (101), it was not used to efficiently study biomolecule structure until Kurt Wüthrich began his work in the late 1970s. Dr. Wüthrich went on to win a Nobel prize in chemistry in 2002 for the development of his techniques.

Unlike X-ray crystallography NMR spectroscopy is non-destructive and does not irreversibly alter a sample, thus making it an efficient technique which can be repeated without sample replacement (5). In addition higher field strength instruments combined with sensitive cryoprobes may be used to study the structure of molecules in concentrations on the high micromolar scale. Also unlike crystallography NMR allows for the study of dynamic processes of molecules in solution such as conformational changes, proton exchange, and even binding. A fallback of NMR however is the inability to view cofactors with inappropriate spin, such as calcium molecules which may be critical to structure formation (5).

1.5.1 Introduction to the Basics of NMR Spectroscopy

An NMR spectrometer (Figure 22) is a machine that generates an electromagnetic field, is capable of producing radiofrequency pulses, measuring changes in the magnetic environment, and is capable of outputting data based on these changes. NMR spectroscopy is a technique that utilizes information obtained by altering the magnetic field of particles that contain particular ½ spin states. A particle such as a proton contains a particular spin which when subjected to a magnetic field will statistically align with the direction of that magnetic field so that it remains in a state of low energy. At this point a radiofrequency (*rf*) pulse may be applied to the proton at its Larmor frequency, such that the energy is absorbed by the proton and that proton then climbs in energy state, thus misaligning with the magnetic field produced by the spectrometer. After absorbing the energy from the *rf* pulse and opposing the magnetic field the proton will then fall back into it's lower energy state thus producing a decay signal which may then be interpreted by the spectrometer as a particular change in current (5). This resultant information is mathematically transformed by a computer and describes things such as how many protons occur on a molecule, how electronegative each environment is, and how many other protons neighbor the one of interest (Figure 23). This information often combined with supplemental information from mass spectrometry provides for the determination of structure of small molecules, and provides information on spin systems in larger biological molecules. A spin system is a group of spins that are connected by scalar or through bond *j*-coupling.

Figure 22: A 500 MHz NMR spectrometer used at the University of Hawaii at Manoa. Photo courtesy of the UH Manoa Chemistry website.
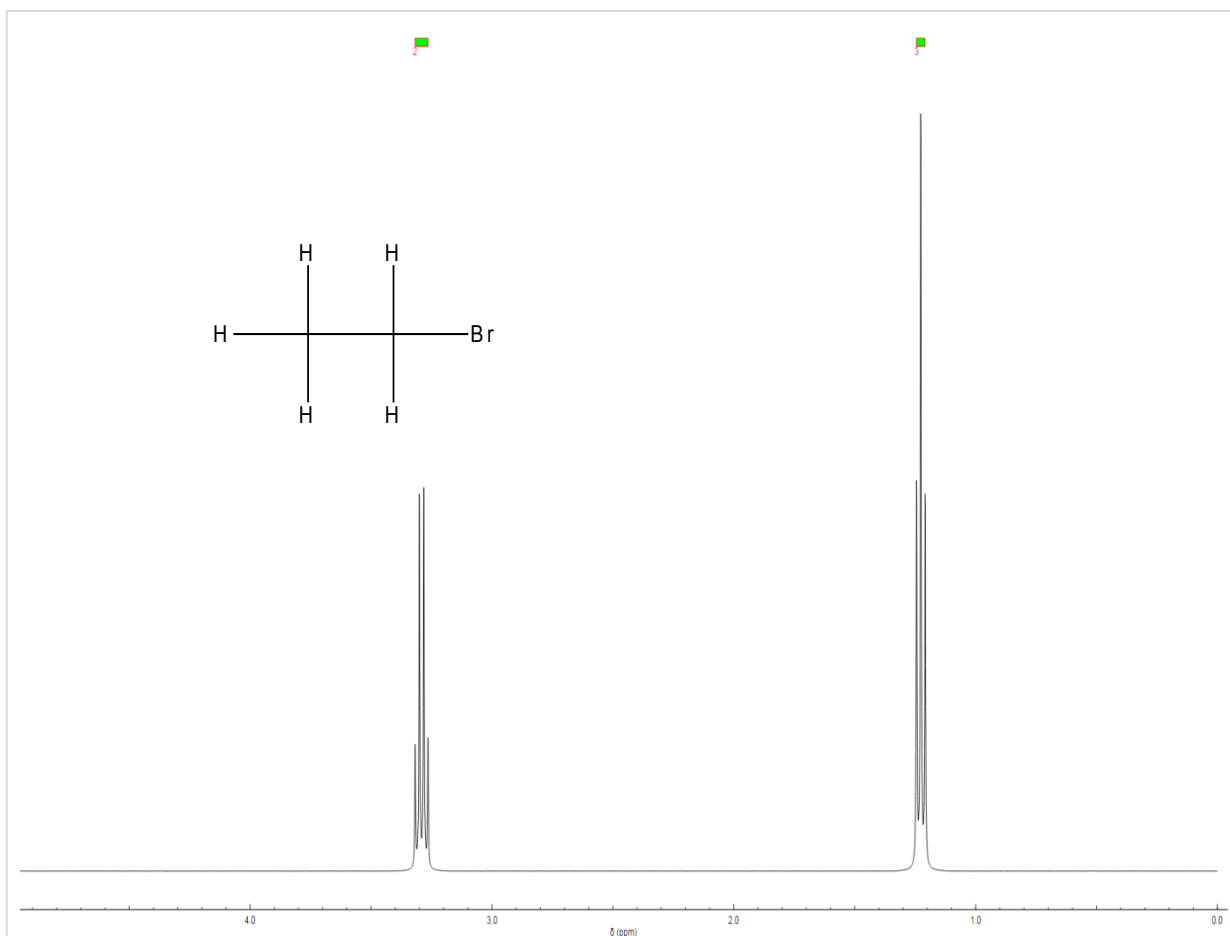
Figure 23: One-dimensional NMR spectrum of bromoethane. The three protons on carbon 2 are represented as the more upfield peak (closer to 1 ppm) whereas the protons on carbon 1 are more downfield, due to the shielding effects of bromine. The splitting pattern reveals how many neighbors are present by the n+1 rule with the protons on carbon one having a splitting pattern of 3+1 making a quartet, while the protons on carbon 2 have a splitting pattern of 2+1 resulting in a triplet. The amount of protons on each carbon is determined by the relative area integration of each peak.

Two-dimensional NMR spectroscopy furthers the information obtainable by single dimensional NMR spectroscopy by providing further elucidation of particular spin systems through scalar or *j*-coupling, as well as interatomic distance information through dipolar coupling. This information is achieved by following a scheme of preparation, evolution, mixing, and detection, where different experiments are achieved by altering the evolution and mixing periods (5).

1.5.2 Modern Techniques in Solution Structure Determination of Peptides

The elucidation of the solution structures of biological molecules by NMR is often undertaken as an iterative, step-wise process and is an ever-expanding discipline. The following section will provide the reader with a basic understanding of this process but is in no way an exhaustive guide to all avenues of solution structure determination, nor should it serve as a complete introduction to the topic. Textbooks such as "NMR of Proteins and Nucleic Acids" by Kurt Wüthrich and "Molecular Modeling Principles and Applications" by Andrew R. Leach are recommended starting points to anyone interested in this field.

Dissimilar to crystallography, information obtained by NMR alone is not sufficient to determine the structure of a molecule. Instead NMR solution structure relies on *a priori* information that can be narrowed down using a combination of known laws and measured energies for the conformations of various biological molecules and the use of efficient problem solving algorithms. The modern techniques for the tertiary solution structure determination of peptides in particular may be undertaken using either proton or hetero-nuclear NMR experiments. Important experiments include single dimensional NMR and a series of two-dimensional experiments including correlation spectroscopy (COSY), total correlation spectroscopy (TOCSY), and NOE or Rotational Overhauser (ROE) spectroscopy (all of which will be further discussed later). In addition another helpful experiment is exclusive correlation spectroscopy (ECOSY). The data obtained from each of these experiments is reliant on each other and requires careful interpretation as many pitfalls (which will be discussed later) may arise from erroneous interpretation. Once properly interpreted, the information obtained is converted into restraints used by a computer program which applies an algorithm to solve for low energy structures and converges on a set of structures representative of the molecules lowest thermodynamic energy state in solution.

COSY and TOCSY experiments result in information on the scalar (through bond) coupling of ½ spin atoms in a molecule. In proton NMR, COSY is absent of mixing and provides information on which protons are scalar coupled to just three bonds away, providing an avenue of plotting where protons are in a particular spin system. If the information is complete, COSY can allow for the total elucidation of a particular spin system, often times however this is not the case. An alternative and often supplemental experiment is TOCSY, which at appropriate mixing times can provide information on the scalar coupling of an entire spin system.

In contrast NOESY and ROESY are experiments that result in information on dipolar (through space) coupling. This technique will provide information on distances between ½ spin atoms in relation to each other in space. Determining whether ROESY or NOESY should be used is decided by how fast the molecule tumbles in solution, a property related to molecular

weight. If a molecule is generally larger than 1100 Da. NOESY often yields better data whereas if a molecule is smaller, ROESY appears to be more effective.

Solution structure determination by NMR requires a combination of these techniques. At minimum a single dimensional NMR spectrum to determine splitting, number of protons, and in the case of peptides and proteins to assess folding and proton exchange. Then either a COSY or TOCSY (usually both) is needed to determine which spin system and consequently which amino acid each proton belongs to. Finally a dipolar coupling experiment is required such as NOESY or ROESY to build restraint lists for use in a restrained MD program.

1.5.3 Sequence determination and peak assignment

Correctly assigning peaks in an NMR spectrum is essential to beginning the structure determination process. In order to start the assignment operation a processed one-dimensional spectrum must be assessed to determine if a peptide or protein adopts a stable conformation in solution and if impurities are present within the sample. The corresponding protons should match the protons in the peptide of interest. Fewer protons than expected may be observed due to the exchange of amide protons with deuterium in the solution, or due to overlap with the water signal. Such issues may be corrected by changing parameters such as temperature and pH. A fully folded peptide will produce amide peaks (in the downfield region) that are well dispersed, meaning that they occupy different electromagnetic environments due to intramolecular hydrogen bonding or specific positioning as a result of the folding process. A folded peptide spectrum in proton NMR should consist of a downfield amide region between 10-7 ppm, an aromatic region between 7-6 ppm, an aliphatic region between 6-4 ppm, and a methyl region between 3-0 ppm (Figure 24).
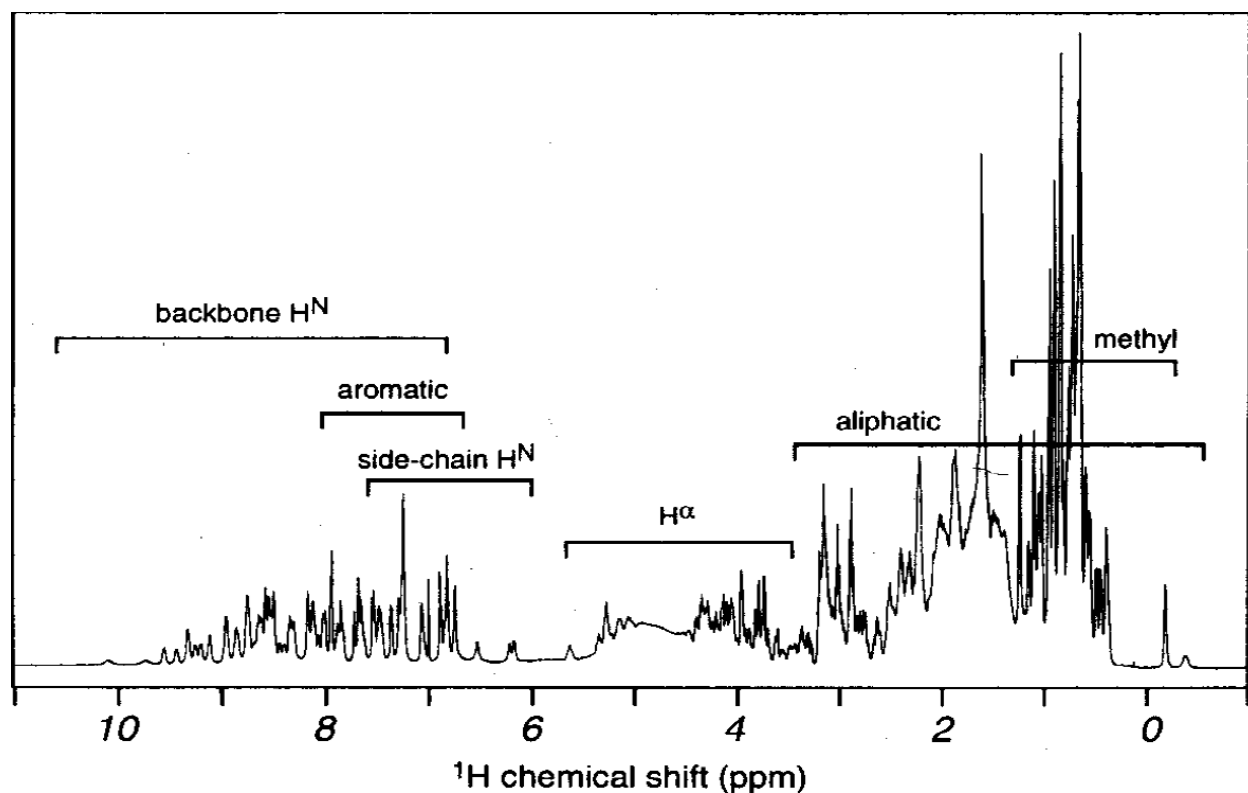
Figure 24: Regions of a proton spectrum in a general folded peptide. This figure was adapted from Kurt Wüthrich 1986 (5).

Once each peak type has been correctly identified, two-dimensional scalar coupling data given by COSY or TOCSY experiments may be used to determine which spin system (and thus corresponding amino acid) each peak belongs to. A two-dimensional spectrum consists of a one-dimensional spectrum oriented topographically and horizontally, where cross peaks between two peaks in the diagonal section provide information on the coupling of those protons (Figure 25). In COSY this would represent scalar coupling, whereas in NOESY this would represent dipolar coupling. Once each spin system has been determined by COSY or TOCSY, the connections of those spin systems in space may be determined by dipolar experiments. In order to determine the sequence of a particular peptide and thus begin assigning protons based on position within the peptide, the fingerprint region of the dipolar spectrum would need to be ascertained. The fingerprint region consists of α proton and amide proton couplings. Finding a known position in a peptide, and then tracing couplings from that amino acids alpha proton to the adjacent amide protons allows for the determination of the peptide's sequence in a fashion known as a sequence walk (5). Once the sequence has been determined and each amino acid has been assigned, restraints may be calculated from the intensities of the observed dipolar coupled crosspeaks.
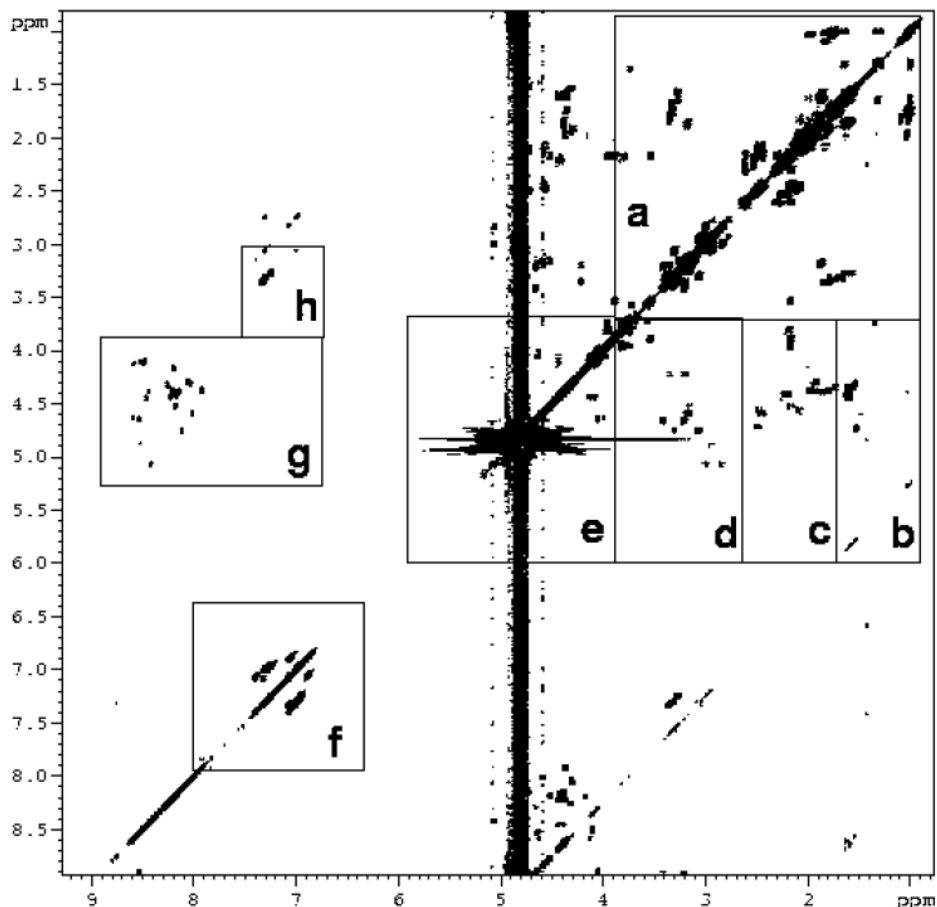
Figure 25: A generalized example of a two-dimensional spectrum of a peptide. Region g represents the fingerprint region. This figure was adapted from Kurt Wüthrich 1986 (5).

1.5.4 Calculating Distance Restraints

As previously mentioned, data from NMR spectroscopy alone is insufficient to fully determine the solution structure of a bio-molecule. In order to determine the complete structure of a biological molecule NMR data must be used in tandem with MD or distance geometry. In peptides, once each amino acid's spin system has been fully described by the COSY or TOCSY data, dipolar couplings must be measured with NOESY or ROESY data. The intensity of these peaks will then be used to create what are known as restraints, which will be used by a MD package to determine structure.

The relationship between distances of protons and NOE intensity may be described as $\frac{1}{r^{-6}} \cong NOE\ intensity$ where r is the distance of the protons involved in the dipolar coupling.

Therefore if known distances occur within a molecule such as vicinal protons in an aromatic ring (Figure 41), restraints may be calculated based on this relationship and the relative volumes of other peaks within the spectrum. An example of this would be if the previously described protons occurred at a known distance of 2.5 Å (Figure 26) and had a peak volume of 1.5×10^7, a distance of 3.5 Å could be calculated using the corresponding formula $\frac{V_{AB}}{(r_{AB})^{-6}} = \frac{V_{BC}}{(r_{BC})^{-6}}$ where V_{AB} is the volume of the peak corresponding to protons A and B (in our case the two vicinal protons on our aromatic ring) and r_{AB} is the distance between those protons . BC would then be other protons in the spectra corresponding to a different distance, in our case we may choose another distance such as 3.5 Å. Substitution of these values with our example equation would produce: $\frac{1.5 \times 10^7}{(2.5 \text{ Å})^{-6}} = \frac{V_{BC}}{(3.5 \text{ Å})^{-6}}$ and thus $V_{BC} = 1.5 \times 10^7 \left(\frac{(3.5 \text{ Å})^{-6}}{(2.5 \text{ Å})^{-6}}\right)$ Or 2.0×10^6. Such a relationship provides for the establishment of what are known as restraints, or regions of peak volumes that correspond to a particular distance. In the previously worked example the restraint established would be that between peak volumes 1.5×10^7 and 2.0×10^6 a proton would likely be between 2.5 and 3.5 angstroms.
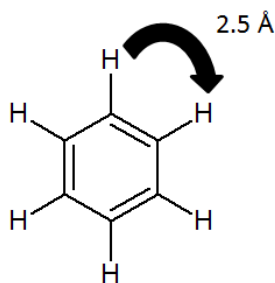
Figure 26: Vicinal protons of an aromatic ring (benzene), showing the known distance of 2.5 Å.

The restraints determined by this method may be binned into different categories of peaks, often coined strong (0 - 2.5 Å), medium (0 - 3.5 Å), weak (0 - 4.5 Å), and very weak (0 - 5.5 Å) peaks. Once appropriately calculated, specialized assignment software may be used to measure the volumes of each peak within a dipolar spectrum, allowing for automated output of restraints into intended bins that are then compiled into a restraint list. An example of a restraint list for use by a MD package using the XPLOR format can be seen in Figure 27.

```
assign (residue 9 and name HA) (residue 10 and name HN) 3.5 3.5 0.0
assign (residue 5 and name HA1) (residue 6 and name HN) 4.5 4.5 0.0
assign (residue 5 and name HA1) (residue 5 and name HN) 4.5 4.5 0.0
assign (residue 3 and name HA) (residue 4 and name HN) 3.5 3.5 0.0
assign (residue 5 and name HA2) (residue 6 and name HN) 3.5 3.5 0.0
assign (residue 5 and name HA2) (residue 5 and name HN) 3.5 3.5 0.0
assign (residue 4 and name HB1) (residue 5 and name HN) 3.5 3.5 0.0
assign (residue 8 and name HA) (residue 8 and name HN) 4.5 4.5 0.0
assign (residue 7 and name HA) (residue 8 and name HN) 3.5 3.5 0.0
assign (residue 6 and name HA) (residue 6 and name HD#) 4.5 4.5 2.0
assign (residue 11 and name HB2) (residue 11 and name HD1) 4.5 4.5 0.0
assign (residue 5 and name HA1) (residue 5 and name HA2) 3.5 3.5 0.0
assign (residue 11 and name HB2) (residue 11 and name HD1) 4.5 4.5 0.0
assign (residue 10 and name HA) (residue 10 and name HG1#) 4.5 4.5 0.6
assign (residue 10 and name HA) (residue 10 and name HG2#) 5.5 5.5 0.6
assign (residue 6 and name HA) (residue 6 and name HB#) 3.5 3.5 1.0
assign (residue 11 and name HA) (residue 11 and name HB1) 3.5 3.5 0.0
assign (residue 8 and name HN) (residue 8 and name HA) 5.5 5.5 0.0
assign (residue 3 and name HN) (residue 3 and name HB2) 5.5 5.5 0.0
assign (residue 4 and name HB2) (residue 4 and name HN) 4.5 4.5 0.0
assign (residue 8 and name HN) (residue 8 and name HE#) 4.5 4.5 1.0
assign (residue 11 and name HD1) (residue 11 and name HD2) 3.5 3.5 0.0
assign (residue 12 and name HN) (residue 11 and name HA) 4.5 4.5 0.0
assign (residue 5 and name HN) (residue 5 and name HA1) 5.5 5.5 0.0
assign (residue 5 and name HN) (residue 5 and name HA2) 4.5 4.5 0.0
assign (residue 5 and name HN) (residue 4 and name HB1) 4.5 4.5 0.0
assign (residue 5 and name HA2) (residue 6 and name HN) 3.5 3.5 0.0
assign (residue 10 and name HN) (residue 10 and name HG2#) 4.5 4.5 1.0
assign (residue 12 and name HA) (residue 12 and name HB1) 4.5 4.5 0.0
```

Figure 27: An example of binned restraint lines of code in the XPLOR format created by the assignment software SPARKY. Assign acts as an operator which tells the MD program to take the proceeding two specific atom positions and restrain them between a lower bound distance, an upper bound distance, and allow for a particular error (the third column of numbers, left to right respectively).

1.5.5 "Restrained" Molecular Dynamics (MD)

Once a restraint list has been made it may then be of use in an MD package. MD is an *in-silico* technique for studying the behavior and movement of molecular models (102). In the process of determining the NMR solution structures of molecules, MD simulations are considered critical as the information generated from various NMR spectroscopy experiments only provides information on possible distances between atoms and allows for the exclusion of unlikely or impossible solutions. The combination of restrained MD simulations and data from NMR spectroscopy results in a semi empirical solution which takes into account three important factors including: restraints generated from NMR spectroscopy, a molecular force field, and a problem solving algorithm.

A force field is complex descriptor of molecules that allows them to be modeled as realistically as possible while maintaining efficiency within a program. Current force fields for use in MD packages output basic information such as energy changes related to deviations in position from a molecules thermodynamic equilibrium. In a basic force field these changes often include bond length, angle, torsional angle, electrostatic interactions, van der Waals interactions, and solvent interactions (102). Interestingly, for the sake of efficiency Newtonian physics often suffices as a replacement for quantum mechanics when considering the calculation of energies derived from these properties. The Borne-Oppenheimer approximation allows for the consideration of atoms as large Newtonian objects in comparison to their electron counterparts, as protons in contrast are very much larger and are considered to be localized, whereas electrons are not. Since only atom positions are of interest in NMR solution structure determination, electrons may be removed from the calculations. This approximation allows for the properties of atoms such as bond length and angle changes to be modeled with simple harmonic equations, allowing for a more rapid calculation of atom position, movement from thermal equilibrium, and thus calculation of approximate free energy (102).

A molecular force field is an essential tool for energy calculation however it is only just that, an algorithm is required to take the information outputted by the force field to be used to solve the conformational problem. A popular problem solving algorithm used in many MD packages is known as simulated annealing. Annealing is a term stemming from metallurgy and is used to characterize the process of heating and cooling metal in order to arrange the atoms in the metal from a disordered state to that of a crystalline one. This process parallels as a near perfect analogy to solution structure determination as it often begins with a disordered molecule and ends in a low energy, ordered state.

Simulated annealing, much like the annealing process fulfills the requirements of taking a problem such as a disordered molecule and results in the output of a more ordered one, except that it is done *in-silico* and fed information to restrain particular parameters. The simulated annealing algorithm consists of several important key factors including: a random global search function, a local minima search function, a greed function, and a cost function. The simulated

annealing process uses each of these functions to converge on a particular lowest energy "cost" which corresponds to particular atom positions when considering the output from the force field. Thus whatever particular positions the atoms are in when the force field has output the lowest energy is accepted as the lowest energy conformation of the molecule and thus what would be observed by nature. If convergence of output is achieved, the resulting atom positions may be considered real.

Although the process of problem solving using MD simulations through the route of simulated annealing with an advanced force field could be considered highly more efficient than brute force problem solving, for many molecules it would take a near millennia to solve a solution structure without any form of initial empirical input. Thus restrained MD simulations allow for the input of empirical information such as distance restraints derived from NMR spectroscopy experiments that shrink the searchable energy plane for the algorithm's output. This in turn allows the process to speed up dramatically and for solutions to be found within a matter of hours.

Finally once convergence of data from multiple outputs has been achieved, that molecule must be examined for possible flaws. It is often encountered that the first output of a set of solution structures from an MD package is incorrect or contains minor or major inaccuracies. These problems can arise from situations that include inaccurate assignment of peaks, contamination, multiple solution structures, ambiguous data, errors in calculation of peak volume, or insufficient data. In order to determine if the outputted model is representative of the true solution structure, assessment must be undertaken.

Peptide structures may be assessed in multiple ways including comparison of RMSD of all outputted structures, Ramachandran plotting and analysis, measurement of positions in comparison to homologous solved structures, planarity of bonds, collisions of atoms, and much more. A popular program for the assessment of structures is known as PROCHECK and examines much of what has been previously discussed.

1.5.6 Additional information obtained from NMR spectroscopy

NMR spectroscopy is highly versatile tool when considering the amount of information that can be probed in a biological molecule. Solution structure determination may be complemented by particular instances of data that may be retrieved from information output by the NMR instrument. Some of the most pertinent pieces of information include particular couplings which might provide evidence for disulfide bond connectivity, isomerization of Proline, formation of secondary structure, torsional angle determination, and hydrogen bond formation (5).

As previously mentioned disulfide bonds may be considered one of the most important post-translational modifications of peptide toxins, especially conotoxins as their structures are often highly dependent on appropriate disulfide bond formation. Using a combination of the previously discussed NMR techniques it is therefore possible to assign particular cysteines within a sequence and calculate distances between heavy atoms or protons between particular cysteines. Thus if the beta protons of two cysteines are in the proximity expected of a disulfide bond, it is possible to observe these as medium or weak crosspeaks. Despite providing a small amount of evidence for the occurrence of disulfide bonds between two particular cysteines these crosspeaks can be misleading, as cysteines may be in close proximity due to the fold of the peptide backbone, and may not necessarily be in a disulfide bond with each other.

In addition to individual crosspeaks providing information about elements of the peptide's conformation and secondary structure, the hydrogen bonding amides within a peptide may be identified by performing a quantitative experiment to measure proton exchange rates (5). In peptides proton exchange occurs on the amide proton due to the resonance of the amide bond. Proton exchange is slowed by participation in hydrogen bonding, thus the significantly slow exchange of an amide proton with its solvent is likely to be participating in a hydrogen bond.

Another valuable piece of information obtainable by NMR is *j*-coupling. *J*-coupling is the interaction between two nuclear spins which can be measured by the splitting of the resonance lines in NMR. These couplings can be measured by how many hertz they span and can be used in the Karplus equation to generate torsional angle restraints. Torsional angle restraints may be used like distance restraints to provide more information for an MD package to help better refine a final structure.

Lastly, the chemical shift index or CSI can be used as a means to determine if secondary structure exists within a peptide or it can be used comparatively to determine if a peptide is adopting its intended fold. The CSI is a numerical index which provides information on a peptide's fold by comparison of random coil alpha proton shift values to that of a folded peptide. If the shift of alpha proton shift of an experimentally determined peak is +/- 0.1 ppm that peak is assigned a CSI value of 0, if the value is greater than 0.1 ppm that peak is assigned a +1 and if the peak value is lower than 0.1 ppm in comparison to the random coil value, it is assigned -1. Any grouping within the sequence of four or more -1s uninterrupted by 1s is assigned a helix, while three or more 1s uninterrupted by -1s is considered a sheet. This technique has shown in the literature to be approximately 90% accurate at predicting secondary structure (103).

1.5.7 Solution Structure determination of the topological isomers of PnID

As previously mentioned, PnID is a prime candidate to study by NMR due to its small size, few PTMs, high homology to χ-conotoxins, and propensity to form multiple globular isomers. From initial studies of the venom duct material of the Red Sea *Conus pennaceus* it was determined that two disulfide isomers were present, with the largest isomer eluting slightly earlier. Synthetic studies in Chapter 2 showed that folding in an ammonium bicarbonate buffer created roughly equal forms of both globular and ribbon peptides, with the globular peak eluting earlier. It was also determined that selective synthesis of the globular form of PnID resulted in two different isomers, indicating that the order of disulfide bond formation was key to achieving a unique globular fold. In Chapter 3 it was determined that changes to the oxidation buffer may shift formation either to form a higher propensity of ribbon form or the globular form and that oxidation under cellular concentrations of glutathione resulted in a larger formation of the ribbon isomer, with a still significant portion resulting in the globular form. This previous information guided the investigation of the solution structure of PnID, by emphasizing its unique structural characteristics. Thus the proceeding portion of this chapter will focus on the structural information obtained by NMR on the ribbon and two globular forms of PnID. Since the globular is likely to be what is produced within the venom duct and can be produced as two isomers the full solution structures of the two globular forms were determined.

1.6 Methods

1.6.1 NMR Analysis

Samples for NMR spectroscopy were prepared in a 10% D_2O 90% H_2O solution pH 3.1 at concentrations of ~4 mM peptide (>95% purity). A stock solution was prepared by first making a pH 2.5 solution working solution and subsequently diluting it with 10% D_2O to bring the pH to 3.1. Proton 1D spectra were acquired at temperatures ranging from 280 - 298 K. Temperatures which yielded the best resolution were used for acquisition of the 2D data. TOCSY experiments (eight scans per increment and a 1 s delay between scans) (104) were carried out with a 90 ms mixing time and NOESY experiments (sixteen scans per increment and a 1 s delay between scans) (105,106) were performed with a 150 ms mixing time. All spectra were acquired on a 500 MHz Varian Unity Inova spectrometer using a 3 mm gHX ID probe from Varian Associates at 280 K, maintaining the temperature within +/- 1 K. Water suppression was achieved using the Double Pulse Field Gradient (DPFG) method of Huang and Shaka (107). Spectra were acquired with 256 and 1120 complex data points in the F1 and F2 dimensions respectively and zero filled to a final size of 2K by 2K points, Apodization in both dimensions was a combination of a 30 degree shifted sine bell and gaussian window functions. The spectra were then processed using NMRpipe (108) and MestreNova (109) for Microsoft Windows. Assignment of the Spectra was carried out with SPARKY (110).

5.22 Molecular Modeling

Modeling studies were performed using the NMR module of YASARA Structure (111). Conformational restraints were derived from the NOESY spectra and classified as either: strong (1.7-2.5 Å), medium (1.7-3.5 Å), weak (1.7-4.5 Å), or very weak (1.7-5.5 Å). Disulfide bonds were specified between positions 4,12 and 3,9 for Isomers B and RevB. Pseudo-atom corrections were made for methyl and methylene protons according to Wuthrich *et al*. (5). Each isomer was subjected to refinement in explicit solvent using YASARA structure. A total of 50 structures were generated and exhibited good convergence. The ensemble of the lowest energy structures was validated by PROCHECK-NMR (112,113)

1.7 Results

In an attempt to determine differences in favored backbone topologies, the NMR solution structures of the two globular (B and RevB) isomers were determined. For each isomer the amide signals of the one-dimensional (1D) spectra were well dispersed (between 6.8 and 9.2 ppm) which indicated that each isoform contained well-ordered structures. Temperatures that yielded the greatest resolution in the amide region were chosen for two-dimensional experiments. The isomers yielded 2D spectra of moderate quality, resulting in a sufficient amount of restraints for structural modeling. Additional peaks were apparent in lower concentrations than that of the expected spin systems however are unlikely to be additional isoforms and could be due to the presence of a small amount of contamination. Protons from each residue were assigned using TOCSY and NOESY spectra, using a sequential assignment strategy (5) in the program SPARKY (110). Secondary shift values, which provide information on the similarity of backbone structures among the peptides, were calculated using the observed α proton shifts with respect to their random coil values (114). The compared secondary shifts of each isomer are shown in Figure 43.
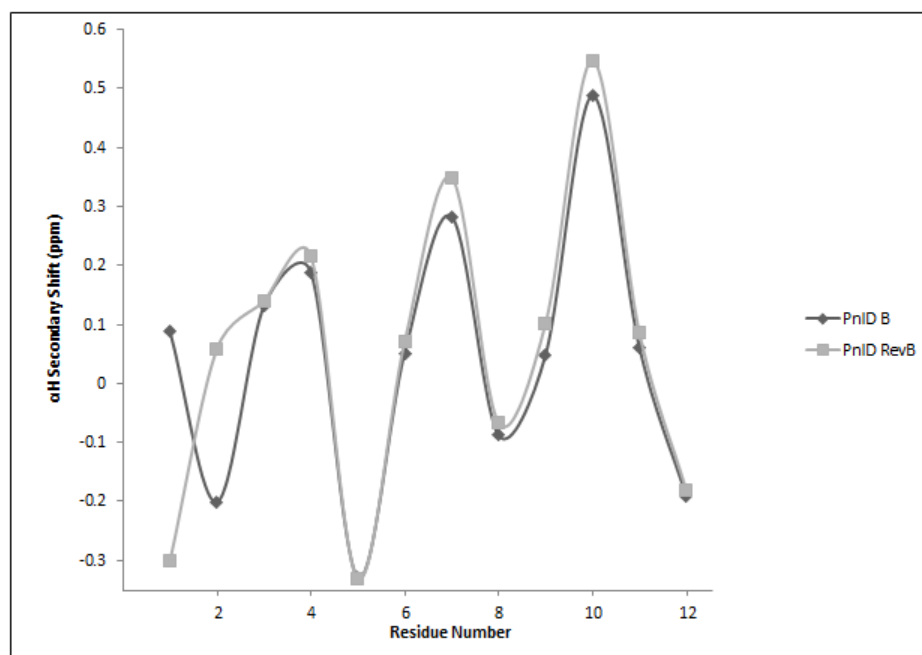

Figure 28: The comparative secondary shift values of PnID and PnID RevB. The similar shift values from residues 5-12 indicate that the backbone positions are conserved in these regions, while positions 1 and 2 appear to be the least conserved. The more positive chemical shift values of PnID RevB provide further evidence of a higher likely hood of β sheet type character.

1.7.1 NMR analysis of PnID B

TOCSY and NOESY connectivity diagrams for PnID B are shown in Figure 29. A stretch of NH-Hα peaks were readily identified between residues 3-10 and 12 however the expected fingerprint peak from residue 2 was not visible, likely due to rapid amide exchange. A total of 55 restraints were determined from the NOESY spectra, of which 49 were intra-residue and 16 were inter-residue NOEs.

1.7.2 NMR analysis of PnID RevB

TOCSY and NOESY connectivity diagrams for PnID RevB are shown in figure 29. NH-Hα peaks were identified between residues 2-10 and 12. A total of 63 restraints were calculated from the integrated and assigned NOESY spectra, of which 40 were intra-residue and 23 were inter-residue NOEs.
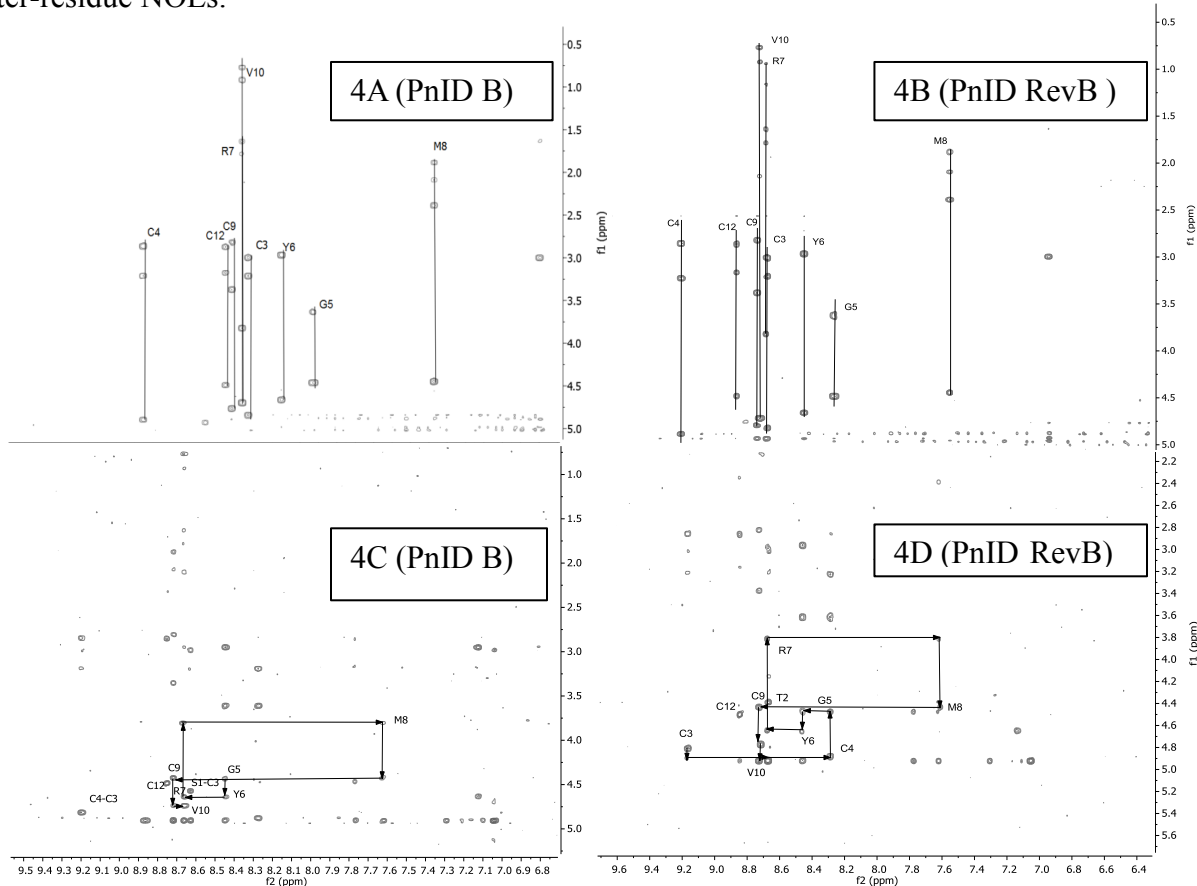

Figure 29: The TOCSY connectivity diagrams of PnID B (4A) and PnID RevB (4B) and NOESY sequence walks (4C and 4D).

1.7.3 Molecular modeling

In an attempt to better elucidate topological differences between the two globular isomers, molecular dynamics simulations were performed on each isomer using calculated distance restraints derived from the best NOESY spectra. Sets of 50 structures of each globular isomer were generated using Simulated Annealing and energy minimization protocols, and explicit refinement in solvent using the YASARA structure NMR module (www.yasara.org). 20 final structures that corresponded to the lowest energies were chosen to represent the preferred solution structure of each peptide. Figure 30 illustrates the determined lowest energy structures for each isoform. Structural statistics for each isomer are shown in Table 11 while the entire quality assessment for each structure can be seen in Appendix C. PnID B yielded a set of structures with a backbone root mean squared deviation (RMSD) from average of 1.409 Å in comparison to PnID RevB which yielded a backbone RMSD from average of 1.155 Å. Neither set of structures contained any NOE violations greater than 0.5 Å (Table 11). Positive secondary shifts in residues 2-4 and 8-11 RevB (Figure 28) reinforce the notion that a β-type structure could exist in this region as was calculated in 7 of the 20 lowest energy models. β-type structure was not observed in PnID B despite having similar secondary shifts between residues 4-12. Both peptides yielded turns in their ensembles, however due to the proximity of the amide proton in arginine 7 to that of the carboxy group of glycine 5 the calculated structures of PnID RevB appear to have a higher likelihood of forming a α-turn, whereas due to the distance between the carboxy group of cysteine 4 and the amide proton of arginine 7, PnID B appears to have a higher propensity to form an γ-turn in residues 4-6, similar to what has been reported in other χ-conotoxins. Due to the lower resolution of the data, we were unable to pinpoint the exact Ramachandran angles for the residues likely involved in the turns in each of these structures and therefore it is not known exactly what type of γ- or α-turn may be favored in each peptide. Recurring calculated hydrogen bonds within PnID revB also appear to be present between the amides of cysteines 3 and 4 and the carboxy portion of methionine 8 possibly indicating stable secondary structure within this region. Additionally four long range NOEs were observed in PnID RevB as opposed to the single long range NOE seen in PnID B, further increasing the likelihood of a β-type structure occurring in PnID RevB but not in PnID B. Additionally both PnID B and RevB contained *cis*-Proline residues, indicating that this conformation is highly favored in this peptide, decreasing the likelihood that the isomers may due to *cis/trans* isomerization of Proline.
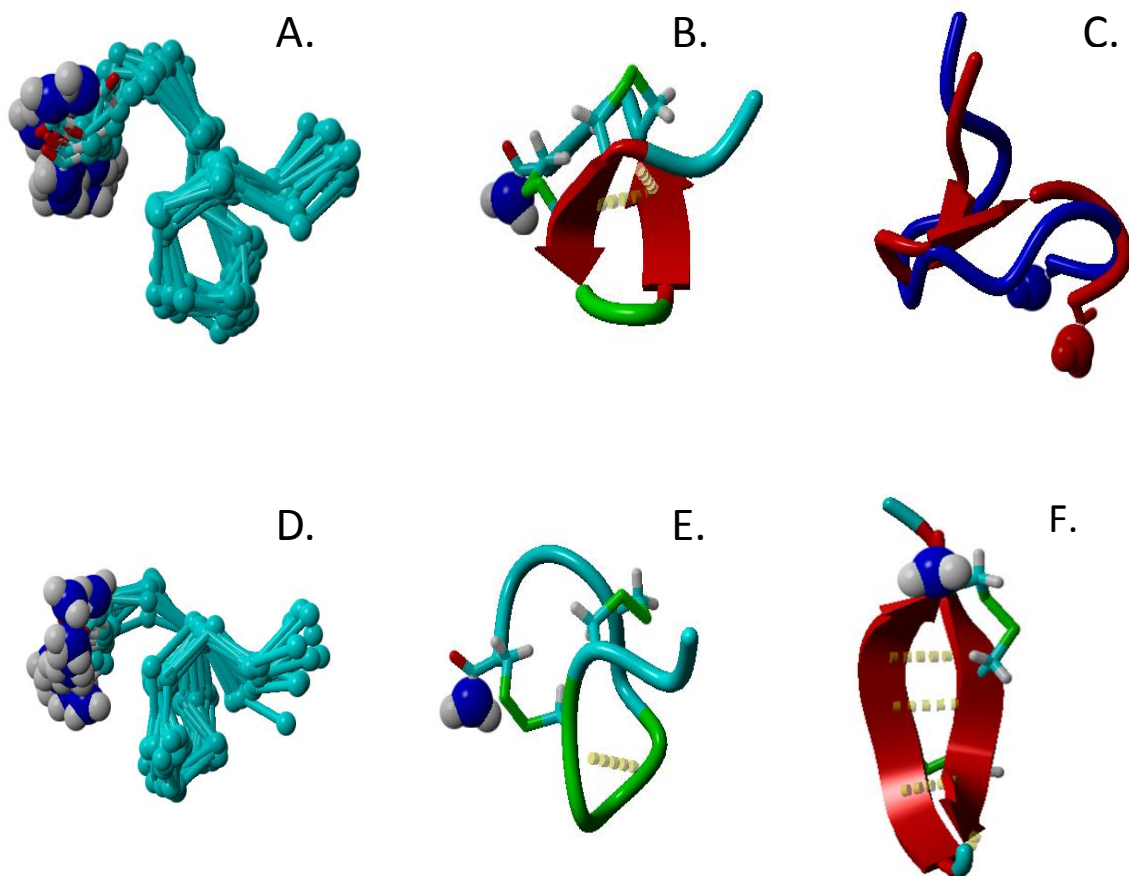
Figure 30: Calculated models of PnID B and RevB. Pane A shows the trace overlay of the 20 lowest energy models of PnID RevB. Pane B shows the lowest energy structure of PnID RevB. Pane C is a comparison of PnID B and RevB. Pane D shows the trace overlay of the 20 lowest energy models of PnID B. Pane E shows the lowest energy structure of PnID B. Pane F is the lowest energy structure of χ-MrIA.

Table 11: Statistics for the calculated models of PnID RevB and PnID B.

	PnID RevB	PnID B
% Backbone angles in favored region	75%	67%
% Backbone angles in allowed region	21%	26%
% Backbone angles in generously allowed region	4%	8%
Backbone RMSD for all models	0.62	0.5
Violations > 0.3 Å	0	0

5.3 Discussion

Comparison between the primary sequence of PnID and that of other χ-conotoxins reveals that significant differences occur in the second and before the first inter-cysteine loops with the Cys-Cys-Gly region of the peptide maintaining complete identity, a region which appears to be stabilized by an antiparallel β-sheet in PnID revB, χ-MrIA, and χ-MrIB (60,115) (Table 2) (Figures 29 and 16). All currently known χ-conotoxins contain a Valine before the first cysteine in the sequence and a Hydroxyproline and Histidine in the final inter-cysteine loop. This common characteristic was investigated with respect to the integrity of χ-MrIA's structure activity relationships (4). After performing alanine scans it was determined that Valine 3 and *4-trans*-hydroxyproline 12 are critical residues for maintaining stability in χ-MrIA. Additionally NMR analysis provided insight to hydrophobic interactions that occur between these residues. Thus it is likely that the unique placement of Valine and Proline in PnID might aid in its lack of preference for a particular connectivity or provides a slight preference for a globular connection. The Alanine scan performed on χ-MrIA also revealed that the residues between glycine 6 and Leucine 9 were of the strongest significance to binding, which represents the most similar region to χ-PnID (80). In addition to these residues *4-trans*-hydroxyproline 12 and Histidine 11 were seen to contribute to binding as well albeit in a lessened capacity.

As previously mentioned, residue replacement on χ-MrIA was not only performed by alanine scan but also with select residues and multiple replacements on the known contributing regions of the molecule, which tested the contribution of size, hydrophobicity, and charge of important residue positions within the peptide to its binding affinity with the NET (116). From this experiment Bryan-Luka *et al*. developed a model of the pharamacophore of χ-MrIA and determined that tyrosine 7 participated in aromatic and hydrogen bonding interactions with the receptor and Leucine 9, Histidine 11, and *4-trans*-hydroxyproline 12 interacted with a deep hydrophobic pocket. Bryan-Luka *et al*. also found that replacement of Lysine 8 with Arginine resulted in a lower activity. PnID however contains arginine in a similar position of the molecule but is flanked by a methionine instead of a Leucine. This change in this particular residue position could account for the maintenance in affinity for the NET as Methionine has been observed to participate in Methionine π interactions with aromatic residues, an interaction which has increased favorability over hydrophobic interactions alone (117).

Conotoxins have generally been observed to adopt well-ordered structures whose rigid backbone does not radically deviate when binding to an inhibition site (58). This concept has especially held true for the well-studied globular α-conotoxins which are highly stabilized by disulfide bonds and a characteristic α-helical structure which essentially retains their solution structure in both bound and unbound states. The single exception to this however is the α-conotoxin BuIA, which appears to form a disordered backbone in solution that may attune to three separate distinctive conformers, which have evidenced to be Proline *cis/trans* isomers (56).

Topological isomers have been described in various molecules including catenanes, rotaxanes, Mobius strips, and knotted molecules (118). Catenanes are molecularly interlocked structures such as two rings, rotaxanes are another interlocked structure in which one molecule may feed through a ring made by another, and Mobius strips include molecules which form a ring with only a single boundary. With the exception of Mobius strips, these motifs are generally found in small molecule structures. Knotted molecules, particularly those incorporating cystine knots are peptides and are known to undergo topological isomerization involving disulfide bonds(119). These topologies are found in peptides containing six cysteines and three disulfides and are generally capable of forming three unique topologies known as: the growth factor cystine knot (GFCK), the inhibitor cystine knot (ICK), and the cyclic cystine knot (CCK) motifs. The GFCK, unlike the other two topologies does not have a disulfide which bisects a ring of two other two, instead creating a conformation whose middle disulfide misses the ring entirely. The CCK and ICK have similar disulfide connections where a middle disulfide bisects a ring created by the other two, however they differ in that the CCK contains a cyclic backbone, while the ICK contains free *N*- and *C*-termini. Although these topologies have been characterized in three disulfide-containing peptides, until now major topological isomers of two-disulfide containing peptide toxins have eluded discovery.

Despite the fact that α-conotoxins are all globular peptides, no single instance has yet been reported in literature pertaining to the discovery of topological α-conotoxin isomers. Since PnID is similar in size to the α-conotoxins it is plausible that topological isomers could exist for some if not many sequences. There is mention of an error in co-elution in a synthetically produced isomer of ω-conotoxin GVIA to that of a native (134), which may be representative of a topological isomer. As the globular form of the α-conotoxins is thought to be the "native" and predominantly active form, it would be highly likely that the biological activity of these topologies would exhibit significant differences to each other. Thus we suggest that topologies be investigated in previously determined α-conotoxins, especially those similar in size to PnID.

The globular form of PnID furthers its uniqueness in that it is capable of forming multiple topologies which are likely not due to the *cis/trans* isomerization of Proline due to the presence of characteristic *cis* Proline NOEs being observed in the spectra for both globular peptides. These topologies appear to be characterized by a perturbance of a loop between residues 4/5-7. An inverse γ-turn has been well characterized as a feature of χ-conotoxins and is well documented in the solution structures of χ-MrIA (115)(Figures 29 and 16). In the case of χ-MrIA, the integrity of the residues which compose this turn was also seen to be critical factor in maintaining appropriate binding to the NET, therefore the existence of turns in the globular forms of PnID is not surprising. The possible replacement of this turn in the PnID RevB topology to an α-turn, however suggests that for this peptide in particular, order of disulfide bond formation may play a crucial role in creating an intermediate folding state whose energy may be conducive to forming the required hydrogen bonds for creating a stable and active structure once folded. This order specific folding strategy could also affect the isomerization of Proline during the folding process in the α-conotoxins, as it was not fully investigated in the case of α-BuIA, however it is unlikely as it does not appear to be a factor in inducing favorability in the *cis/trans* isomerization of Proline in PnID.

In addition to the perturbance of the critical turn between residues 5-7, the globular topologies of PnID show differences in secondary structure. PnID RevB appeared to form a rigid beta structure that contains two antiparallel β-sheets whereas PnID B does not appear to contain this rigidity. The structure formed by PnID RevB appears to be stabilized by an increased capacity to form hydrogen bonds further up in the peptide backbone, thus changing the turn size in the area critical to activity (residues 5-8). This is in direct contrast to the PnID B topology which has a less rigid region near the termini and the conservation of a γ-turn in the critical loop (residues 5-7).

A comparison between the globular topologies determined within this work to that of the ribbon isoform of χ-MrIA revealed that each isoform resembles portions of the χ-MrIA structure. Comparison between PnID revB and that of χ-MrIA (Figure 30) reveals that PnID revB retains not only sequence homology in the area surrounding it's cysteines but also structural homology in it's oxidized form. The key difference in this isomer to that of χ-MrIA is the loss of the turn in the 5-7 residue region. This is in contrast to that of PnID B which only retains the turn and loses the rigid β-type structure.

The differences in structure among the two topological isomers directly explain the biological activity determined in Chapter 4 on the monoamine transporters. It has been previously noted that the γ-turn in χ-MrIA is critically important to biological activity and that turn appears to be somewhat conserved in the more active globular topology (PnID B), whereas this turn is perturbed in PnID RevB. This likely gives rise to the small amount of biological activity seen in PnID B and the complete lack of activity in PnID RevB. A similar loss in activity

was seen in χ-MrIB in which it had an EC_{50} of about half of that of χ-MrIA in a rat *vas deferens* assay, likely due to the promiscuous nature of the same intercysteine loop. In addition to this at lower concentrations PnID B appeared to have non-selective inhibition of the monoamine transporters whereas PnID RevB appeared to have none. The promiscuity of PnID B could be explained by the loss of secondary structure within it's backbone and thus it's lack of rigidity.

Biopolymers, especially peptides, have been increasingly becoming important players in the pharmaceutical market as novel therapeutics with over 140 leads in clinical trials as of 2015 (120). Due to this advance in gravity, understanding of how particular crosslinks, residues, inter-cysteine spacing, and topologies play their role in potency, selectivity, and stabilization is becoming an ever more important task. *Conus* peptides provide great tools for increasing this understanding as they are small in size, relatively selective, and contain a significant amount of modifications which aid in selectivity and stabilization (31,32).

Despite immense progress in the study of the structure, activity, and folding of α conotoxins, as well as an increasing interest in χ/λ-conotoxins over the past decade, to the best of our knowledge native isomers have never been reported until now. Additionally, despite many synthetic studies where the disulfide bonds of conotoxins were regio-selectively formed, there has never been a report of the production of differential backbone topologies between two peptides despite containing identical disulfide bridges.

Since the native isomeric material observed in this study was derived from the venom duct of *Conus pennaceus* and not the organism's milked venom, it is possible that one of the isomers may not be expressed as a functional toxin in the venom and is a consequence of inefficient folding within the venom duct. If this is true, there is a possibility that PnID represents a less efficient scaffold than other χ-conotoxins, and could serve as a transition point between the long studied globular fold of the α-conotoxins and that of the ribbon fold seen in previously studied χ-conotoxins. The significance of this circumstance would be the wider implication that by studying the unique folding properties of related conopeptides produced within the venom duct, the organism's route to more efficient scaffolds may be traced and eventually lead to more robust methods of rational peptide therapeutic design.

Due to the findings presented here we suggest that future studies involving short, regio-selectively synthesized and cross-linked peptides such as conotoxins, explicit information should be given in the experimental procedures / methods pertaining to which disulfide was formed initially so future work is not subject to discrepancies due to resulting differences in topology. This work also implies that reported activities or structures of peptides similar in nature to PnID may be of non-native or less active topologies and in some cases may be worth reinvestigation.

Chapter 6: Future Perspectives

1.1 Reevaluation of previously discovered conotoxins

Despite thousands of conotoxins having been discovered to date only ~0.1% have had their pharmacology extensively studied (121). For decades conotoxin research has carried on in the fashion of determining novel sequence, post translational modification, structure, and activity. Considering the information presented within this work, it would likely be beneficial to re-examine some of these previously studied peptides for topological isomerization as it is currently unknown if this characteristic is unique to the χ-conotoxin scaffold or if it may occur in the more extensively studied and more common α-conotoxins which natively adopt the globular disulfide bond connectivity and contain only two disulfide bonds. Considering the activity differences in the globular isoforms of PnID as seen in chapter four, a second look at the folding capabilities and resultant activity of other previously discovered conotoxins could result in differential activity and possibly higher potency or selectivity than currently reported.

Many of these peptides have been categorized by their disulfide bond connectivity and framework, information that may be used to aid in the synthetic reproduction of these conotoxins. Considering the results of this work, it would be beneficial for investigators studying conotoxins to not only report disulfide connectivity overall, but also report the order of disulfide bond formation during synthetic reproduction. It could also be beneficial to study possible differences adopted by folding conotoxins under separate conditions i.e. a single step oxidation vs. a multistep oxidation, as we were able to diversify the topology of these toxins using the two methods (see Chapter 3).

If previously synthesized conotoxins were to be of less active topologies (i.e. α-conotoxins), the implication would be that the production of topological isomers could be used as a means of conotoxin diversification, both synthetically and naturally. As was seen in the introduction, PnID was originally discovered as two isomers, one of which is highly likely to be a globular isomer considering the findings of this work (Chapters 2 and 3). This implies that cone snails may use the intrinsic property of topological isomerization as a method to diversify venom naturally.

1.2 Rational peptide design

This section contains excerpts from my paper entitled "Conotoxins and their regulatory considerations"

As mentioned in Chapter 1, peptides are gaining interest as therapeutics due to their selectivity, potency and size. Rational drug design, typically applied to the design of small molecule inhibitors to increase both pharmacodynamic and kinetic properties, is beginning to find a place in the betterment of potential peptide therapeutics (120). In conotoxins in particular rational design has often previously been referred to as peptide bioengineering, a term of controversy among some engineers and scientists as it implies that standard engineering principles are being applied to the design of these molecules that may not always be the case. Nevertheless this term has been extensively used in previously literature. Considering the increasing importance of these molecules and their constant subjection to synthetic modification it is important to evaluate the capability of these peptides to form alternative topologies. In chapter 4 it was discussed how topological differences affected the activity of PnID. If such differences can occur in PnID, which is a χ-conotoxin and very similar to the α-conotoxins, both of which have had significant interest in being developed into therapeutics, topological isomerization should be a topic at the forefront of this field.

In Chapter 4, it was seen that at lower concentrations PnID B was able to non-selectively inhibit each of the monoamine transporters. If this phenomenon were able to be translated to other conotoxins it could potentially open the door for new probes through rational peptide design. Alternatively less selective conotoxins that have been previously reported could be re-evaluated to determine if topological isomerization was a factor in the lack of specificity.

Conotoxins and conopeptides have been the subject of extensive bioengineering and chemical manipulation (33). Their ease of synthesis has positioned them as a primary choice among peptide chemists who manipulate the native peptide to enhance stability and activity. The resulting bioengineered peptide with augmented properties has proved to be an invaluable tool in studies pertaining to structural activity relationships, drug leads and ion channel characterization. The development of bioengineered conotoxins presents special cases and equivocal situations that may require amendments to the current regulations. Conotoxins contain numerous post-translational modifications (21). These PTMs are responsible for phyla and receptor selectivity. Removing some or all of the PTMs can change receptor selectivity (34). Through bioengineering, a native conotoxin not active in mammalian receptors can be modified to have activity in mammalian receptors. Since conotoxins contain a great deal of PTMs and PTMs are known to influence folding in general, the influence of PTMs on topology could be investigated in the future to determine what contributions, if any warrant topological differences in these peptides.

Peptidomimetics has further advanced the field of conopeptide bioengineering. Allowing researchers to incorporate synthetic residues or combine active regions of various peptides, giving them the ability to create a new peptide with enhanced properties (122). If topological isomerization results partially due to the smaller inter-cysteine loop sizes of the χ-conotoxins, this field could be especially impacted by the formation of unexpected isomers as many peptidomimetics include artificial loop sizes. A particular example of this would be the R2-midi peptide designed from the μ-conotoxins KIIIA and BuIIIC (123). This peptide was synthetically designed to contain a pharmacophore for the sodium channel Na_v 1.2 and is unnaturally sized. Stevens *et al*. successfully removed a single disulfide bond from this peptide thus not only decreasing it's size but increasing it's folding efficiency. Such synthetic modifications could potentially be subject to topological isomerization and thus should be considered within this context.

With the advent of native chemical ligation (NCL) the linking of a peptide's *N*-terminus to its *C*-terminus has been successfully performed (124). Conotoxins, with their numerous cysteine residues, present themselves as a prime target for peptide backbone cyclization. Cyclization is known to enhance the stability of conotoxins. This synthetic strategy has been applied to Vc1.1, α-conotoxins AuIB, RgIA and MII using NCL; the cyclized analogs have been shown increased oral bioavailability and systemic stability (125-127). In addition to the use of synthetic chemistry, cyclization has also been accomplished through advanced molecular biology techniques such as interim mediated ligation, which is a potential route for future recombinantly-produced cyclic conotoxins (128). Backbone cyclization, especially in terms of non-native backbone cyclization could potentially exhibit topological effects as it often requires the inclusion of a linker sequence between the *N*- and *C*-termini and thus changes the intercysteine loop size of this region. Now that many α-conotoxins are undergoing the cyclization process in the hope of creating more "drug-like" peptides, topological isomerization could very likely be occurring in some of these products.

Another approach to increase the *in vivo* stability is using cystathionine and dicarba linkages in replacement to essential disulfide bonds. This approach has been successfully applied in α-conotoxin ImI (129). Enhancing oral stability of conotoxins provides a new route of administration and hence requires a reassessment of current conotoxin directives. Inclusion of such bonds in these peptides have specific effects on bond length, as a cystathionine bond, dicarba linkage, or diselenide bond would differ slightly in length to that of traditional disulfides. These slight changes in bond length could potentially change the favorability of bond formation and thus produce alternative topologies. Therefore topological isomers should also be considered with these particular modifications as they may provide unique biological activities.

112

As all of these forms of modification are gaining in popularity, a look into their effect on topological isomerization would be worth pursuing. If topological isomerization is unique to that of the χ-conotoxin scaffold it would be of interest to determine what differences are causing this phenomenon and if those differences could be mimicked, either deliberately or by indirect causation. Considering the many facets affected by topological isomerization, it is important when synthetically producing either a native sequence or a rationally designed globular peptide to consider the order of disulfide bond formation and ensure that the appropriate order is followed to produce the intended topology.

Another area of interest that may aid in our understanding of topological isomer formation and thus better future rational peptide design is that of the study of the pro regions of conotoxins and their relationship to molecular chaperones within the venom duct. It has been previously determined that pro regions of these peptides serve as anchor sites to PDIs which aid in efficiently folding conotoxins (130). Despite being only one of many potential mechanisms which aid in folding, the contribution of these highly conserved sequences to achieving not only the appropriate disulfide connectivity but also the appropriate topological motif should not be ignored.

Many techniques have been adapted in recent years to aid in the rational design of peptides, not the least of which are the previously discussed alanine scans, NMR solution structure determination, and crystallography. However In addition to traditional structure determination real time disulfide bond connectivity analysis may be a beneficial technique for determining which disulfide bond may be more thermodynamically stable within a peptide, thus potentially providing differential information on topological isomers. Partial reduction and alkylation, a procedure developed by Gray *et al*. (131) is a technique commonly used to determine the connectivity of disulfides within peptides and proteins. This process entails using a reducing agent along with a small amount of heat to sequentially break individual disulfide bond while sparing others (131). This technique takes advantage of the slight differences in bond energies throughout the disulfides of the peptide, and yields largely varied results depending on the peptide sequence under investigation. There are several potential reducing agents which may be used to selectively break the disulfide bonds/bridges, however TCEP is generally favored due it's fast reaction time, pH range compatibility, and it's lack of odor (Table 3), in addition, unlike other reducing agents TCEP does not reduce the cystine through a mixed disulfide intermediate but by a thiophosphonium intermediate (132). During the reduction process it is possible to monitor the reaction when using very low concentrations of the reducing agents. Monitoring may be done in real-time by direct injection of the reactants into an electrospray ionization mass spectrometer (ESI-MS) and scanning at a narrow mass window (+/- 10 *m/z*) for the respective reduced masses (133). Once a single bond is broken the peptide, resulting in a mass shift has demonstrated by boarding of the mono-isotopic distrubtion, it then is acidified to prevent the reformation, or scrambling of the disulfide and the 'captured' free thiols are then alkylated.

Various alkylating agents are available for this process and each carry particular advantages and disadvantages, however the use of iodoacetamide and vinyl pyridine, the reagents originally used by Gray *et al*. (131), are commonly used. After alkylation the peptide that contains the alkylated target mass of interest is then purified from the reaction mixture by reverse-phase high pressure liquid chromatography (RP-HPLC/UV). This process is then sequentially repeated for other disulfides and exposed to a different alkylating agent with a unique mass addition with respect to the previous alkylating agent(s). Finally the peptide is then sequenced and the corresponding 'matching' cysteinyl functions are considered to have been previously shared a single disulfide bond (Figure 3). Sequencing approach, resulting in disulfide bond assignment, may be taken a number of ways: Originally done by Gray *et al*. (131) use of Edman degradation on RP-HPLC purified products, while Bingham *et al*. (133), used original MS/MS to deduce disulfide bond connectivity. Notably, if Edman degradation is being used on a cyclic peptide, that peptide must first be cleaved enzymatically as the Phenylisocyanate will have no free termini to react with – this is a potential hurdle. If there are no potential cleavage sites for the peptide MS/MS or PSD should be the sequence method of choice – fragmenting along the peptide backbone, but paying special attention to sequence analysis.

Therefore from the findings of this work we believe that topological isomerization could be a major influencing factor to many different areas regarding the future of peptides as molecular probes, drugs, and even potentially pesticides. As peptides begin to garner more interest with higher levels of rational design and a push for better pharmacokinetic properties we may see an increased risk of inappropriate topological isomerization. Thus careful consideration must be taken when planning a synthetic oxidation strategy, the inclusion of synthetic residues or PTMs, the alteration of loop sizes, cyclization, or alternative disulfide strategies. Finally, techniques such as co-elution as seen in Chapter 2, differential activity as seen in Chapter 4, crystallography or solution structure determination as seen in Chapter 5, and real time disulfide bond analysis as discussed in this chapter should be considered to aid in the elucidation of these isoforms.

References

1. Brust, A., Palant, E., Croker, D. E., Colless, B., Drinkwater, R., Patterson, B., Schroeder, C. I., Wilson, D., Nielsen, C. K., Smith, M. T., Alewood, D., Alewood, P. F., and Lewis, R. J. (2009) chi-Conopeptide pharmacophore development: toward a novel class of norepinephrine transporter inhibitor (Xen2174) for pain. *Journal of medicinal chemistry* **52**, 6991-7002

2. Lewis, R. J. (2012) Discovery and development of the chi-conopeptide class of analgesic peptides. *Toxicon* **59**, 524-528

3. Hill, J. M., Oomen, C. J., Miranda, L. P., Bingham, J. P., Alewood, P. F., and Craik, D. J. (1998) Three-dimensional solution structure of alpha-conotoxin MII by NMR spectroscopy: effects of solution environment on helicity. *Biochemistry* **37**, 15621-15630

4. Nilsson, K. P., Lovelace, E. S., Caesar, C. E., Tynngard, N., Alewood, P. F., Johansson, H. M., Sharpe, I. A., Lewis, R. J., Daly, N. L., and Craik, D. J. (2005) Solution structure of chi-conopeptide MrIA, a modulator of the human norepinephrine transporter. *Biopolymers* **80**, 815-823

5. Wuthrich, K. (1986) *NMR of proteins and nucleic acids*, Wiley

6. Bogin, O. Venom peptides and their mimetics as potential drugs.

7. Slamti, L., and Lereclus, D. (2002) A cell–cell signaling peptide activates the PlcR virulence regulon in bacteria of the Bacillus cereus group. *The EMBO journal* **21**, 4550-4559

8. Griffith, O. W. (1980) Determination of glutathione and glutathione disulfide using glutathione reductase and 2-vinylpyridine. *Analytical biochemistry* **106**, 207-212

9. Espiritu, M. J., Cabalteja, C. C., Sugai, C. K., and Bingham, J. P. (2014) Incorporation of post-translational modified amino acids as an approach to increase both chemical and biological diversity of conotoxins and conopeptides. *Amino acids* **46**, 125-151

10. Yoshioka, S., and Stella, V. J. (2002) Stability of peptide and protein pharmaceuticals. *Stability of Drugs and Dosage Forms*, 187-203

11. Lax, R. (2010) The future of peptide development in the pharmaceutical industry. *PharManufacturing: The international peptide review* **2**, 10-15

12. Bingham, J.-P., Andrews, E. A., Kiyabu, S. M., and Cabalteja, C. C. (2012) Drugs from slugs. Part II–conopeptide bioengineering. *Chemico-biological interactions* **200**, 92-113

13. Burgdorf, J., Zhang, X.-l., Weiss, C., Matthews, E., Disterhoft, J. F., Stanton, P. K., and Moskal, J. R. (2011) The N-methyl-D-aspartate receptor modulator GLYX-13 enhances learning and memory, in young adult and learning impaired aging rats. *Neurobiology of aging* **32**, 698-706

14. Lalezari, J. P., Eron, J. J., Carlson, M., Cohen, C., DeJesus, E., Arduino, R. C., Gallant, J. E., Volberding, P., Murphy, R. L., and Valentine, F. (2003) A phase II clinical study of the long-term safety and antiviral activity of enfuvirtide-based antiretroviral therapy. *Aids* **17**, 691-698

15. Midura-Nowaczek, K., and Markowska, A. (2014) Antimicrobial Peptides and Their Analogs: Searching for New Potential Therapeutics. *Perspectives in Medicinal Chemistry* **6**, 73-80

16. Tam, J. P., and Wong, C. T. (2012) Chemical synthesis of circular proteins. *Journal of Biological Chemistry* **287**, 27020-27025

17.   Slingluff Jr, C. L. (2011) The present and future of peptide vaccines for cancer: single or multiple, long or short, alone or in combination? *Cancer journal (Sudbury, Mass.)* **17**, 343

18.   Cemazar, M., Kwon, S., Mahatmanto, T., S Ravipati, A., and J Craik, D. (2012) Discovery and applications of disulfide-rich cyclic peptides. *Current topics in medicinal chemistry* **12**, 1534-1545

19.   Armishaw, C. J., and Alewood, P. F. (2005) Conotoxins as research tools and drug leads. *Current protein & peptide science* **6**, 221-240

20.   Bingham, J.-P., Broxton, N. M., Livett, B. G., Down, J. G., Jones, A., and Moczydlowski, E. G. (2005) Optimizing the connectivity in disulfide-rich peptides: α-conotoxin SII as a case study. *Analytical biochemistry* **338**, 48-61

21.   Espiritu, M. J., Cabalteja, C. C., Sugai, C. K., and Bingham, J.-P. (2014) Incorporation of post-translational modified amino acids as an approach to increase both chemical and biological diversity of conotoxins and conopeptides. *Amino acids* **46**, 125-151

22.   Kaas, Q., Westermann, J.-C., and Craik, D. J. (2010) Conopeptide characterization and classifications: an analysis using ConoServer. *Toxicon : official journal of the International Society on Toxinology* **55**, 1491-1509

23.   Dutton, J. L., Bansal, P. S., Hogg, R. C., Adams, D. J., Alewood, P. F., and Craik, D. J. (2002) A new level of conotoxin diversity, a non-native disulfide bond connectivity in α-conotoxin AuIB reduces structural definition but increases biological activity. *Journal of Biological Chemistry* **277**, 48849-48857

24.   Gray, W., Luque, A., Olivera, B. M., Barrett, J., and Cruz, L. (1981) Peptide toxins from Conus geographus venom. *Journal of Biological Chemistry* **256**, 4734-4740

25.   Bingham, J.-P., Mitsunaga, E., and Bergeron, Z. L. (2010) Drugs from slugs—Past, present and future perspectives of ω-conotoxin research. *Chemico-biological interactions* **183**, 1-18

26.   Terlau, H., and Olivera, B. M. (2004) Conus venoms: a rich source of novel ion channel-targeted peptides. *Physiological reviews* **84**, 41-68

27.   Hannon, H. E., and Atchison, W. D. (2013) Omega-Conotoxins as Experimental Tools and Therapeutics in Pain Management. *Mar Drugs* **11**, 680-699

28.   Kompella, S. N., Hung, A., Clark, R. J., Mari, F., and Adams, D. J. (2015) Alanine Scan of alpha-Conotoxin RegIIA Reveals a Selective alpha3beta4 Nicotinic Acetylcholine Receptor Antagonist. *The Journal of biological chemistry* **290**, 1039-1048

29.   Layer, R. T., and McIntosh, J. M. (2006) Conotoxins: therapeutic potential and application. *Marine drugs* **4**, 119-142

30.   Molinski, T. F., Dalisay, D. S., Lievens, S. L., and Saludes, J. P. (2009) Drug development from marine natural products. *Nature reviews. Drug discovery* **8**, 69-85

31.   Kaas, Q., Westermann, J. C., Halai, R., Wang, C. K., and Craik, D. J. (2008) ConoServer, a database for conopeptide sequences and structures. *Bioinformatics (Oxford, England)* **24**, 445-446

32.   Kaas, Q., Yu, R., Jin, A. H., Dutertre, S., and Craik, D. J. (2012) ConoServer: updated content, knowledge, and discovery tools in the conopeptide database. *Nucleic acids research* **40**, D325-330

33.   Bingham, J. P., Andrews, E. A., Kiyabu, S. M., and Cabalteja, C. C. (2012) Drugs from slugs. Part II--conopeptide bioengineering. *Chem Biol Interact* **200**, 92-113

34. Bergeron, Z. L., Chun, J. B., Baker, M. R., Sandall, D. W., Peigneur, S., Peter, Y., Thapa, P., Milisen, J. W., Tytgat, J., and Livett, B. G. (2013) A 'conovenomic'analysis of the milked venom from the mollusk-hunting cone snail Conus textile—The pharmacological importance of post-translational modifications. *Peptides* **49**, 145-158

35. Bruce, C., Fitches, E., Chougule, N., Bell, H., and Gatehouse, J. (2011) Recombinant conotoxin, TxVIA, produced in yeast has insecticidal activity. *Toxicon : official journal of the International Society on Toxinology* **58**, 93-100

36. D'Souza, C., Henriques, S. T., Wang, C. K., Cheneval, O., Chan, L. Y., Bokil, N. J., Sweet, M. J., and Craik, D. J. (2016) Using the MCoTI-II Cyclotide Scaffold To Design a Stable Cyclic Peptide Antagonist of SET, a Protein Overexpressed in Human Cancer. *Biochemistry* **55**, 396-405

37. Adams, D. J., Alewood, P. F., Craik, D. J., Drinkwater, R. D., and Lewis, R. J. (1999) Conotoxins and their potential pharmaceutical applications. *Drug development research* **46**, 219-234

38. Kapono, C. A., Thapa, P., Cabalteja, C. C., Guendisch, D., Collier, A. C., and Bingham, J. P. (2013) Conotoxin truncation as a post-translational modification to increase the pharmacological diversity within the milked venom of Conus magus. *Toxicon : official journal of the International Society on Toxinology* **70**, 170-178

39. Moroder, L., and Buchner, J. (2009) *Oxidative folding of peptides and proteins*, Royal society of chemistry

40. Alewood, D., Birinyi-Strachan, L. C., Pallaghy, P. K., Norton, R. S., Nicholson, G. M., and Alewood, P. F. (2003) Synthesis and characterization of δ-atracotoxin-Ar1a, the lethal neurotoxin from venom of the Sydney funnel-web spider (Atrax robustus). *Biochemistry* **42**, 12933-12940

41. Khoo, K. K., Feng, Z.-P., Smith, B. J., Zhang, M.-M., Yoshikami, D., Olivera, B. M., Bulaj, G., and Norton, R. S. (2009) Structure of the Analgesic μ-Conotoxin KIIIA and Effects on the Structure and Function of Disulfide Deletion†‡. *Biochemistry* **48**, 1210-1219

42. Sabatier, J.-M., Lecomte, C., Mabrouk, K., Darbon, H., Oughideni, R., Canarelli, S., Rochat, H., Martin-Eauclaire, M.-F., and Van Rietschoten, J. (1996) Synthesis and characterization of leiurotoxin I analogs lacking one disulfide bridge: evidence that disulfide pairing 3-21 is not required for full toxin activity. *Biochemistry* **35**, 10641-10647

43. Song, J., Gilquin, B., Jamin, N., Drakopoulou, E., Guenneugues, M., Dauplais, M., Vita, C., and Ménez, A. (1997) NMR solution structure of a two-disulfide derivative of charybdotoxin: structural evidence for conservation of scorpion toxin α/β motif and its hydrophobic side chain packing. *Biochemistry* **36**, 3760-3766

44. Flinn, J. P., Pallaghy, P. K., Lew, M. J., Murphy, R., Angus, J. A., and Norton, R. S. (1999) Role of disulfide bridges in the folding, structure and biological activity of ω-conotoxin GVIA. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology* **1434**, 177-190

45. Price-Carter, M., Gray, W. R., and Goldenberg, D. P. (1996) Folding of ω-conotoxins. 2. Influence of precursor sequences and protein disulfide isomerase. *Biochemistry* **35**, 15547-15557

46. Tajima, M., Iida, T., Yoshida, S., Komatsu, K., Namba, R., Yanagi, M., Noguchi, M., and Okamoto, H. (1990) The reaction product of peptidylglycine alpha-amidating enzyme is a

hydroxyl derivative at alpha-carbon of the carboxyl-terminal glycine. *Journal of Biological Chemistry* **265**, 9602-9605

47. Nishiuchi, Y., and Sakakibara, S. (1982) Primary and secondary structure of conotoxin GI, a neurotoxic tridecapeptide from a marine snail. *FEBS letters* **148**, 260-262

48. Kang, T. S., Radic, Z., Talley, T. T., Jois, S. D., Taylor, P., and Kini, R. M. (2007) Protein Folding Determinants: Structural Features Determining Alternative Disulfide Pairing in α-and ξ/λ-Conotoxins. *Biochemistry* **46**, 3338-3355

49. Pisarewicz, K., Mora, D., Pflueger, F. C., Fields, G. B., and Marí, F. (2005) Polypeptide chains containing d-γ-hydroxyvaline. *Journal of the American Chemical Society* **127**, 6207-6215

50. Lopez-Vera, E., Walewska, A., Skalicky, J. J., Olivera, B. M., and Bulaj, G. (2008) Role of hydroxyprolines in the in vitro oxidative folding and biological activity of conotoxins. *Biochemistry* **47**, 1741-1751

51. Gorres, K. L., and Raines, R. T. (2010) Prolyl 4-hydroxylase. *Critical reviews in biochemistry and molecular biology* **45**, 106-124

52. Aguilar, M. B., López-Vera, E., Ortiz, E., Becerril, B., Possani, L. D., Olivera, B. M., and Heimer de la Cotera, E. P. (2005) A novel conotoxin from Conus delessertii with posttranslationally modified lysine residues. *Biochemistry* **44**, 11130-11136

53. Chun, J. B., Baker, M. R., Kim, D. H., LeRoy, M., Toribo, P., and Bingham, J.-P. (2012) Cone snail milked venom dynamics–a quantitative study of Conus purpurascens. *Toxicon : official journal of the International Society on Toxinology* **60**, 83-94

54. Jones, A., Bingham, J. P., Gehrmann, J., Bond, T., Loughnan, M., Atkins, A., Lewis, R. J., and Alewood, P. F. (1996) Isolation and Characterization of Conopeptides by High‐performance Liquid Chromatography Combined with Mass Spectrometry and Tandem Mass Spectrometry. *Rapid communications in mass spectrometry* **10**, 138-143

55. Safavi-Hemami, H., Bulaj, G., Olivera, B. M., Williamson, N. A., and Purcell, A. W. (2010) Identification of Conus peptidylprolyl cis-trans isomerases (PPIases) and assessment of their role in the oxidative folding of conotoxins. *Journal of Biological Chemistry* **285**, 12735-12746

56. Jin, A. H., Brandstaetter, H., Nevin, S. T., Tan, C. C., Clark, R. J., Adams, D. J., Alewood, P. F., Craik, D. J., and Daly, N. L. (2007) Structure of alpha-conotoxin BuIA: influences of disulfide connectivity on structural dynamics. *BMC structural biology* **7**, 28

57. Mayadas, T. N., and Wagner, D. D. (1992) Vicinal cysteines in the prosequence play a role in von Willebrand factor multimer assembly. *Proceedings of the National Academy of Sciences* **89**, 3531-3535

58. Daly, N. L., and Craik, D. J. (2009) Structural studies of conotoxins. *IUBMB life* **61**, 144-150

59. Balaji, R. A., Ohtake, A., Sato, K., Gopalakrishnakone, P., Kini, R. M., Seow, K. T., and Bay, B. H. (2000) lambda-conotoxins, a new family of conotoxins with unique disulfide pattern and protein folding. Isolation and characterization from the venom of Conus marmoreus. *The Journal of biological chemistry* **275**

60. Sharpe, I. A., Gehrmann, J., Loughnan, M. L., Thomas, L., Adams, D. A., Atkins, A., Palant, E., Craik, D. J., Adams, D. J., Alewood, P. F., and Lewis, R. J. (2001) Two new classes of conopeptides inhibit the alpha1-adrenoceptor and noradrenaline transporter. *Nat Neurosci* **4**

61. Zhou, J. (2004) Norepinephrine transporter inhibitors and their therapeutic potential. *Drugs of the future* **29**, 1235-1244

62. Kang, T. S., Radic, Z., Talley, T. T., Jois, S. D., Taylor, P., and Kini, R. M. (2007) Protein folding determinants: structural features determining alternative disulfide pairing in alpha- and chi/lambda-conotoxins. *Biochemistry* **46**, 3338-3355

63. Jin, A.-H., Brandstaetter, H., Nevin, S. T., Tan, C. C., Clark, R. J., Adams, D. J., Alewood, P. F., Craik, D. J., and Daly, N. L. (2007) Structure of α-conotoxin BuIA: influences of disulfide connectivity on structural dynamics. *BMC Structural Biology* **7**, 1-13

64. Dutton, J. L., Bansal, P. S., Hogg, R. C., Adams, D. J., Alewood, P. F., and Craik, D. J. (2002) A new level of conotoxin diversity, a non-native disulfide bond connectivity in alpha-conotoxin AuIB reduces structural definition but increases biological activity. *The Journal of biological chemistry* **277**, 48849-48857

65. Khoo, K. K., Gupta, K., Green, B. R., Zhang, M. M., Watkins, M., Olivera, B. M., Balaram, P., Yoshikami, D., Bulaj, G., and Norton, R. S. (2012) Distinct disulfide isomers of μ-conotoxins KIIIA and KIIIB block voltage-gated sodium channels. *Biochemistry* **51**, 9826-9835

66. Tang, H. Y., and Speicher, D. W. (2004) Determination of disulfide-bond linkages in proteins. *Current protocols in protein science / editorial board, John E. Coligan ... [et al.]* **Chapter 11**, Unit 11 11

67. Sharpe, I. A., Gehrmann, J., Loughnan, M. L., Thomas, L., Adams, D. A., Atkins, A., Palant, E., Craik, D. J., Adams, D. J., Alewood, P. F., and Lewis, R. J. (2001) Two new classes of conopeptides inhibit the alpha1-adrenoceptor and noradrenaline transporter. *Nature neuroscience* **4**, 902-907

68. Lavergne, V., Harliwong, I., Jones, A., Miller, D., Taft, R. J., and Alewood, P. F. (2015) Optimized deep-targeted proteotranscriptomic profiling reveals unexplored Conus toxin diversity and novel cysteine frameworks. *Proc Natl Acad Sci U S A* **112**, E3782-3791

69. Conticello, S. G., Gilad, Y., Avidan, N., Ben-Asher, E., Levy, Z., and Fainzilber, M. (2001) Mechanisms for evolving hypervariability: the case of conopeptides. *Molecular biology and evolution* **18**, 120-131

70. Ueberheide, B. M., Fenyo, D., Alewood, P. F., and Chait, B. T. (2009) Rapid sensitive analysis of cysteine rich peptide venom components. *Proc Natl Acad Sci U S A* **106**, 6910-6915

71. Vigneaud, V. d., Ressler, C., Swan, C. J. M., Roberts, C. W., Katsoyannis, P. G., and Gordon, S. (1953) THE SYNTHESIS OF AN OCTAPEPTIDE AMIDE WITH THE HORMONAL ACTIVITY OF OXYTOCIN. *Journal of the American Chemical Society* **75**, 4879-4880

72. Merrifield, R. B. (1963) Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *Journal of the American Chemical Society* **85**, 2149-2154

73. Hruby, V. J., Barstow, L. E., and Linharrt, T. (1972) New machine for automated solid phase peptide synthesis. *Analytical Chemistry* **44**, 343-350

74. Thakkar, A., Trinh, T. B., and Pei, D. (2013) Global analysis of peptide cyclization efficiency. *ACS combinatorial science* **15**, 120-129

75. Merrifield, B. (2016) Solid phase peptide synthesis: New chemistry and new directions. *Collection of Czechoslovak Chemical Communications* **1**, 12-33

76. Gray, W. R., Luque, F. A., Galyean, R., Atherton, E., Sheppard, R. C., Stone, B. L., Reyes, A., Alford, J., and McIntosh, M. (1984) Conotoxin GI: disulfide bridges, synthesis, and preparation of iodinated derivatives. *Biochemistry* **23**, 2796-2802

77. Espiritu, M. J., Collier, A. C., and Bingham, J.-P. (2014) A 21st-century approach to age-old problems: the ascension of biologics in clinical therapeutics. *Drug Discovery Today* **19**, 1109-1113

78. Xia, Z., Chen, Y., Zhu, Y., Wang, F., Xu, X., and Zhan, J. (2006) Recombinant omega-conotoxin MVIIA possesses strong analgesic activity. *BioDrugs : clinical immunotherapeutics, biopharmaceuticals and gene therapy* **20**, 275-281

79. McIntosh, J. M., Corpuz, G. O., Layer, R. T., Garrett, J. E., Wagstaff, J. D., Bulaj, G., Vyazovkina, A., Yoshikami, D., Cruz, L. J., and Olivera, B. M. (2000) Isolation and characterization of a novel conus peptide with apparent antinociceptive activity. *The Journal of biological chemistry* **275**, 32391-32397

80. Sharpe, I. A., Palant, E., Schroeder, C. I., Kaye, D. M., Adams, D. J., Alewood, P. F., and Lewis, R. J. (2003) Inhibition of the norepinephrine transporter by the venom peptide chi-MrIA. Site of action, Na+ dependence, and structure-activity relationship. *The Journal of biological chemistry* **278**, 40317-40323

81. Balaji, R. A., Ohtake, A., Sato, K., Gopalakrishnakone, P., Kini, R. M., Seow, K. T., and Bay, B. H. (2000) lambda-conotoxins, a new family of conotoxins with unique disulfide pattern and protein folding. Isolation and characterization from the venom of Conus marmoreus. *The Journal of biological chemistry* **275**, 39516-39522

82. Sarin, V. K., Kent, S. B., Tam, J. P., and Merrifield, R. B. (1981) Quantitative monitoring of solid-phase peptide synthesis by the ninhydrin reaction. *Analytical biochemistry* **117**, 147-157

83. Caveney, S., Cladman, W., Verellen, L., and Donly, C. (2006) Ancestry of neuronal monoamine transporters in the Metazoa. *Journal of Experimental Biology* **209**, 4858-4868

84. Saavedra, J. M., Brownstein, M. J., Carpenter, D. O., and Axelrod, J. (1974) Octopamine: presence in single neurons of Aplysia suggests neurotransmitter function. *Science (New York, N.Y.)* **185**, 364-365

85. Torres, G. E., Gainetdinov, R. R., and Caron, M. G. (2003) Plasma membrane monoamine transporters: structure, regulation and function. *Nature Reviews Neuroscience* **4**, 13-25

86. Schroeter, S., Apparsundaram, S., Wiley, R. G., Miner, L. H., Sesack, S. R., and Blakely, R. D. (2000) Immunolocalization of the cocaine‐ and antidepressant‐sensitive l‐norepinephrine transporter. *Journal of Comparative Neurology* **420**, 211-232

87. Morón, J. A., Brockington, A., Wise, R. A., Rocha, B. A., and Hope, B. T. (2002) Dopamine uptake through the norepinephrine transporter in brain regions with low levels of the dopamine transporter: evidence from knock-out mouse lines. *The Journal of neuroscience* **22**, 389-395

88. Baird, D. J., Barber, I., Bradley, M., Soares, A. M., and Calow, P. (1991) A comparative study of genotype sensitivity to acute toxic stress using clones of Daphnia magna Straus. *Ecotoxicology and environmental safety* **21**, 257-265

89. Pimentel, D., Zuniga, R., and Morrison, D. (2005) Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological economics* **52**, 273-288

90. Lv, S., Zhang, Y., Liu, H.-X., Hu, L., Yang, K., Steinmann, P., Chen, Z., Wang, L.-Y., Utzinger, J., and Zhou, X.-N. (2009) Invasive snails and an emerging infectious disease: results from the first national survey on Angiostrongylus cantonensis in China. *PLoS Negl Trop Dis* **3**, e368

91. Adorable-Asis, A.-G. A., Cauyan, G. A., Pagulayan, R. C., Magbanua, F. S., and Papa, R. D. S. (2016) The macro-gastropod communities of aquaculture-intensive lakes in the Philippines. *Molluscan Research* **36**, 223-230

92. Meier, J., and Theakston, R. D. (1986) Approximate LD50 determinations of snake venoms using eight to ten experimental animals. *Toxicon : official journal of the International Society on Toxinology* **24**, 395-401

93. Fainzilber, M., and Zlotkin, E. (1992) A new bioassay reveals mollusc-specific toxicity in molluscivorous Conus venoms. *Toxicon : official journal of the International Society on Toxinology* **30**, 465-469

94. Luna-Ramírez, K. S., Aguilar, M. B., Falcon, A., de la Cotera, E. P. H., Olivera, B. M., and Maillo, M. (2007) An O-conotoxin from the vermivorous Conus spurius active on mice and mollusks. *peptides* **28**, 24-30

95. Kreutzer, C., and Lampert, W. (1999) Exploitative competition in differently sized Daphnia species: a mechanistic explanation. *Ecology* **80**, 2348-2357

96. Lim, H.-K., and Heyneman, D. (1972) Intramolluscan inter-trematode antagonism: a review of factors influencing the host-parasite system and its possible role in biological control. *Advances in parasitology* **10**, 191-268

97. Karabacak, Y., Sase, S., Aher, Y. D., Sase, A., Saroja, S. R., Cicvaric, A., Hoger, H., Berger, M., Bakulev, V., Sitte, H. H., Leban, J., Monje, F. J., and Lubec, G. (2015) The effect of modafinil on the rat dopamine transporter and dopamine receptors D1-D3 paralleling cognitive enhancement in the radial arm maze. *Frontiers in behavioral neuroscience* **9**, 215

98. Nelson, D. L., Lehninger, A. L., and Cox, M. M. (2008) *Lehninger principles of biochemistry*, Macmillan

99. Smyth, M. S., and Martin, J. H. J. (2000) x Ray crystallography. *Molecular Pathology* **53**, 8-14

100. Watson, J. D., and Crick, F. H. (1953) Molecular structure of nucleic acids. *Nature* **171**, 737-738

101. Anderson, W., and Freeman, R. (1962) Influence of a Second Radiofrequency Field on High‑Resolution Nuclear Magnetic Resonance Spectra. *The Journal of Chemical Physics* **37**, 85-103

102. Leach, A. R. (2001) *Molecular modelling: principles and applications*, Pearson education

103. Wishart, D. S., Sykes, B. D., and Richards, F. M. (1992) The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* **31**, 1647-1651

104. Bax, A., and Davis, D. G. (1985) MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *Journal of Magnetic Resonance (1969)* **65**, 355-360

105. Jeener, J., Meier, B., Bachmann, P., and Ernst, R. (1979) Investigation of exchange processes by two‑dimensional NMR spectroscopy. *The Journal of chemical physics* **71**, 4546-4553

106. Kumar, A., Ernst, R., and Wüthrich, K. (1980) A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-

relaxation networks in biological macromolecules. *Biochemical and biophysical research communications* **95**, 1-6

107.   Hwang, T.-L., and Shaka, A. (1995) Water suppression that works. Excitation sculpting using arbitrary wave-forms and pulsed-field gradients. *Journal of Magnetic Resonance, Series A* **112**, 275-279

108.   Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277-293

109.   Willcott, M. R. (2009) MestRe Nova. *Journal of the American Chemical Society* **131**, 13180-13180

110.   T. D. Goddard, D. G. K. SPARKY 3. University of California, San Francisco

111.   Krieger, E. (2003) YASARA.

112.   Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of applied crystallography* **26**, 283-291

113.   Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *Journal of biomolecular NMR* **8**, 477-486

114.   Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L., and Sykes, B. D. (1995) 1H, 13C and 15N chemical shift referencing in biomolecular NMR. *Journal of biomolecular NMR* **6**, 135-140

115.   Nilsson, K. P., Lovelace, E. S., Caesar, C. E., Tynngard, N., Alewood, P. F., Johansson, H. M., Sharpe, I. A., Lewis, R. J., Daly, N. L., and Craik, D. J. (2005) Solution structure of chi-conopeptide MrIA, a modulator of the human norepinephrine transporter. *Biopolymers* **80**

116.   Bryan-Lluka, L. J., Bonisch, H., and Lewis, R. J. (2003) chi-Conopeptide MrIA partially overlaps desipramine and cocaine binding sites on the human norepinephrine transporter. *The Journal of biological chemistry* **278**, 40324-40329

117.   Tatko, C. D., and Waters, M. L. (2004) Investigation of the nature of the methionine-pi interaction in beta-hairpin peptide model systems. *Protein Sci* **13**, 2515-2522

118.   Dobrowolski, J. C. (2003) Classification of topological isomers: Knots, links, rotaxanes, etc. *Croatica chemica acta* **76**, 145-152

119.   Iyer, S., and Acharya, K. R. (2011) Tying the knot: The cystine signature and molecular-recognition processes of the vascular endothelial growth factor family of angiogenic cytokines. *The Febs Journal* **278**, 4304-4322

120.   Fosgerau, K., and Hoffmann, T. (2015) Peptide therapeutics: current status and future directions. *Drug Discovery Today* **20**, 122-128

121.   Lewis, R. J., Dutertre, S., Vetter, I., and Christie, M. J. (2012) Conus venom peptide pharmacology. *Pharmacological reviews* **64**, 259-298

122.   Donevan, S. D., and McCabe, R. T. (2000) Conantokin G is an NR2B-selective competitive antagonist of N-methyl-D-aspartate receptors. *Molecular pharmacology* **58**, 614-623

123.   Stevens, M., Peigneur, S., Dyubankova, N., Lescrinier, E., Herdewijn, P., and Tytgat, J. (2012) Design of Bioactive Peptides from Naturally Occurring μ-Conotoxin Structures. *The Journal of biological chemistry* **287**, 31382-31392

124. Clark, R. J., Jensen, J., Nevin, S. T., Callaghan, B. P., Adams, D. J., and Craik, D. J. (2010) The engineering of an orally active conotoxin for the treatment of neuropathic pain. *Angewandte Chemie (International ed. in English)* **49**, 6545-6548

125. Armishaw, C. J., Jensen, A. A., Balle, L. D., Scott, K. C., Sørensen, L., and Strømgaard, K. (2011) Improving the stability of α-conotoxin AuIB through N-to-C cyclization: the effect of linker length on stability and activity at nicotinic acetylcholine receptors. *Antioxidants & redox signaling* **14**, 65-76

127. Ellison, M., Feng, Z.-P., Park, A. J., Zhang, X., Olivera, B. M., McIntosh, J. M., and Norton, R. S. (2008) α-RgIA, a novel conotoxin that blocks the α9α10 nAChR: structure and identification of key receptor-binding residues. *Journal of molecular biology* **377**, 1216-1227

128. Sancheti, H., and Camarero, J. A. (2009) "Splicing up" drug discovery.: Cell-based expression and screening of genetically-encoded libraries of backbone-cyclized polypeptides. *Advanced drug delivery reviews* **61**, 908-917

129. MacRaild, C. A., Illesinghe, J., Lierop, B. J. v., Townsend, A. L., Chebib, M., Livett, B. G., Robinson, A. J., and Norton, R. S. (2009) Structure and Activity of (2, 8)-Dicarba-(3, 12)-cystino α-ImI, an α-Conotoxin Containing a Nonreducible Cystine Analogue†. *Journal of medicinal chemistry* **52**, 755-762

130. Bulaj, G., Buczek, O., Goodsell, I., Jimenez, E. C., Kranski, J., Nielsen, J. S., Garrett, J. E., and Olivera, B. M. (2003) Efficient oxidative folding of conotoxins and the radiation of venomous cone snails. *Proc Natl Acad Sci U S A* **100 Suppl 2**, 14562-14568

131. Gray, W. R. (1993) Disulfide structures of highly bridged peptides: a new strategy for analysis. *Protein Sci* **2**, 1732-1748

132. Burns, J. A., Butler, J. C., Moran, J., and Whitesides, G. M. (1991) Selective reduction of disulfides by tris(2-carboxyethyl)phosphine. *The Journal of Organic Chemistry* **56**, 2648-2650

133. Bingham, J. P., Broxton, N. M., Livett, B. G., Down, J. G., Jones, A., and Moczydlowski, E. G. (2005) Optimizing the connectivity in disulfide-rich peptides: alpha-conotoxin SII as a case study. *Analytical biochemistry* **338**, 48-61

134. Bingham, J. P. (1998) Novel Toxins from Conus. University of Queensland, Australia

Appendices:

Appendix A. Crude and purified Chromatograms of the PnID isomers
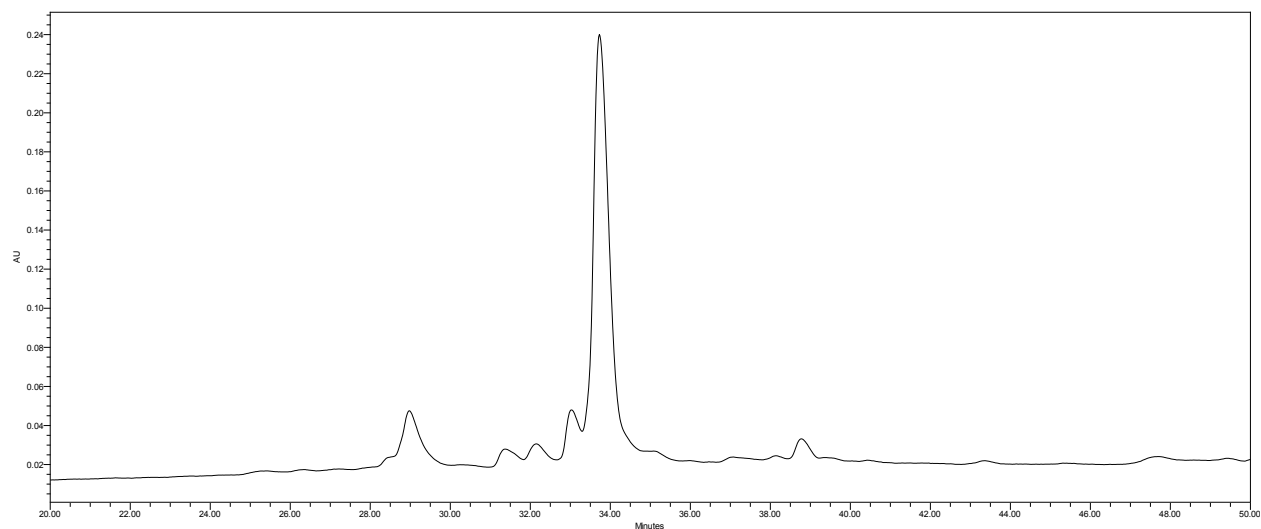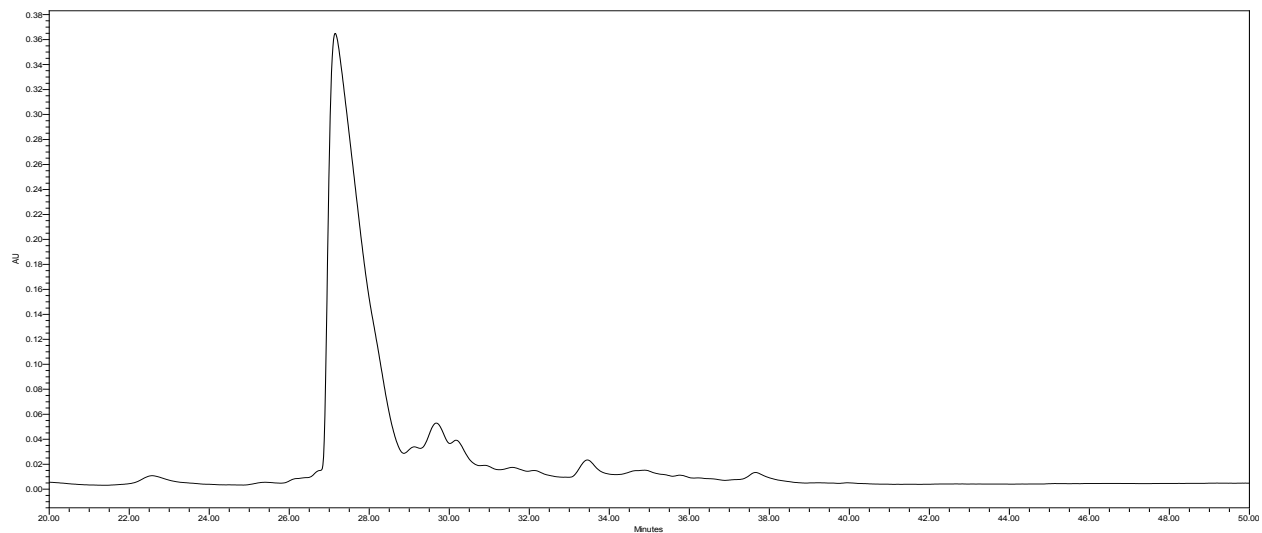

1: RP-HPLC/UV chromatogram of the crude PnID containing no orthogonal protection.
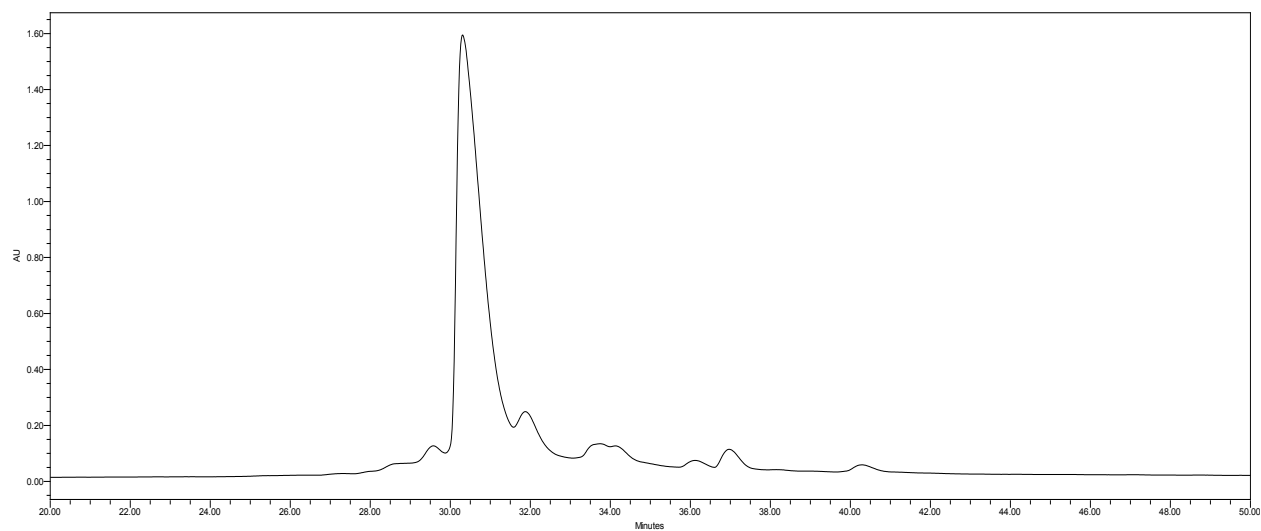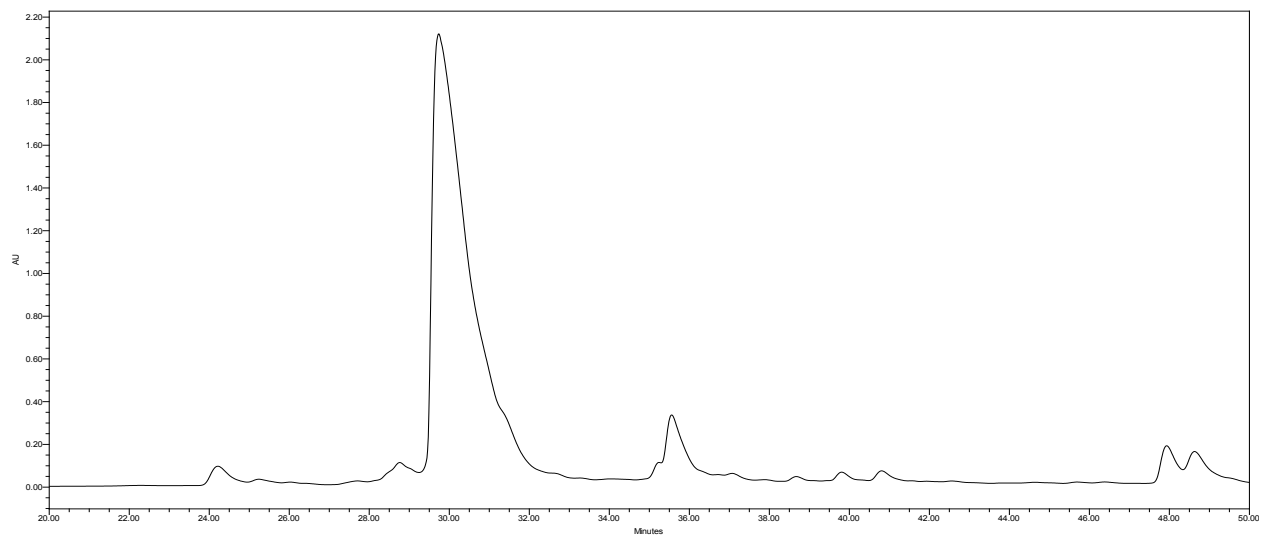

2: RP-HPLC/UV chromatogram of the crude CMrVIA containing no orthogonal protection.
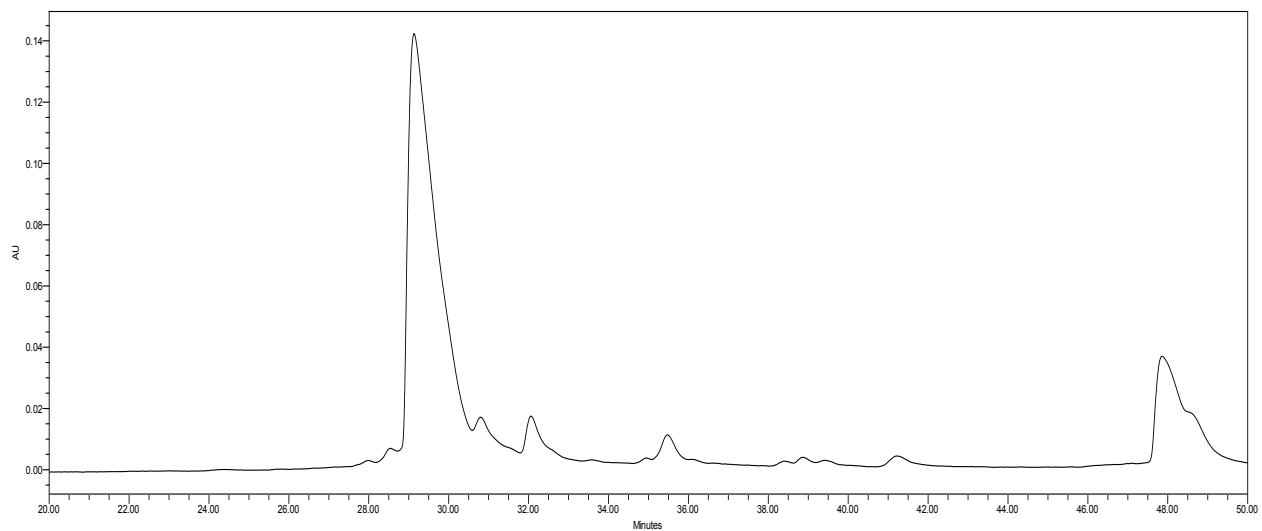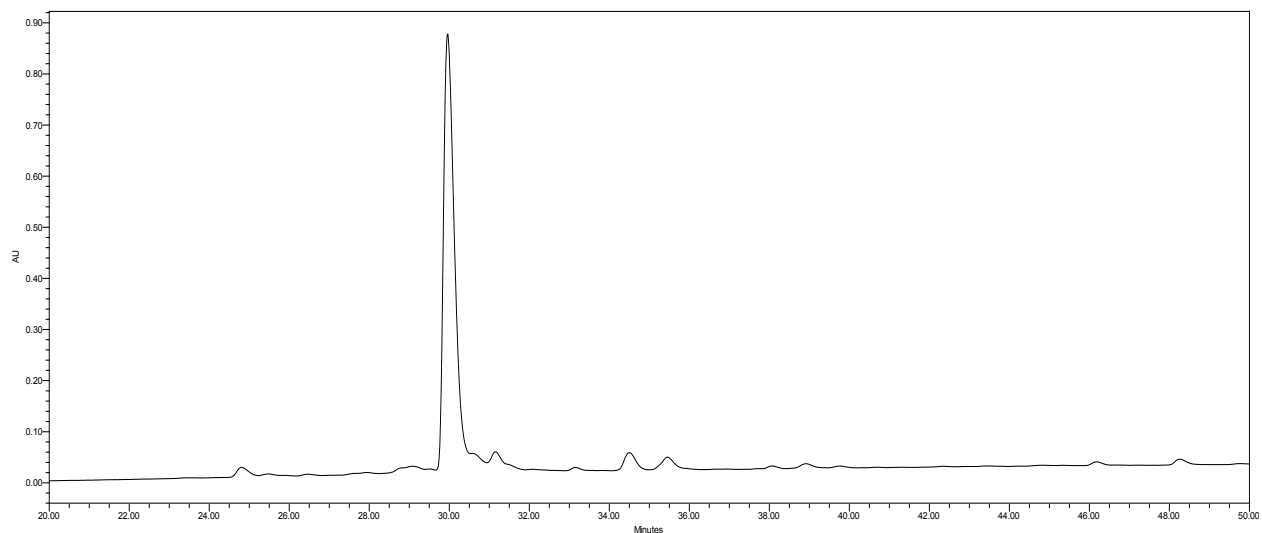
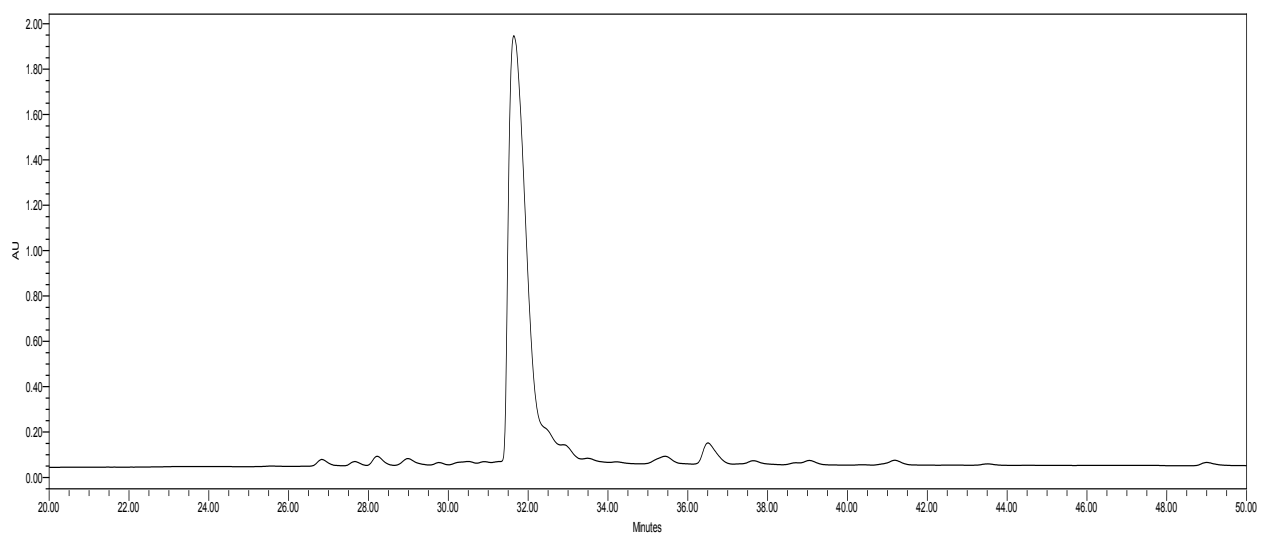**3:** RP-HPLC/UV chromatogram of the crude MrIA containing no orthogonal protection.

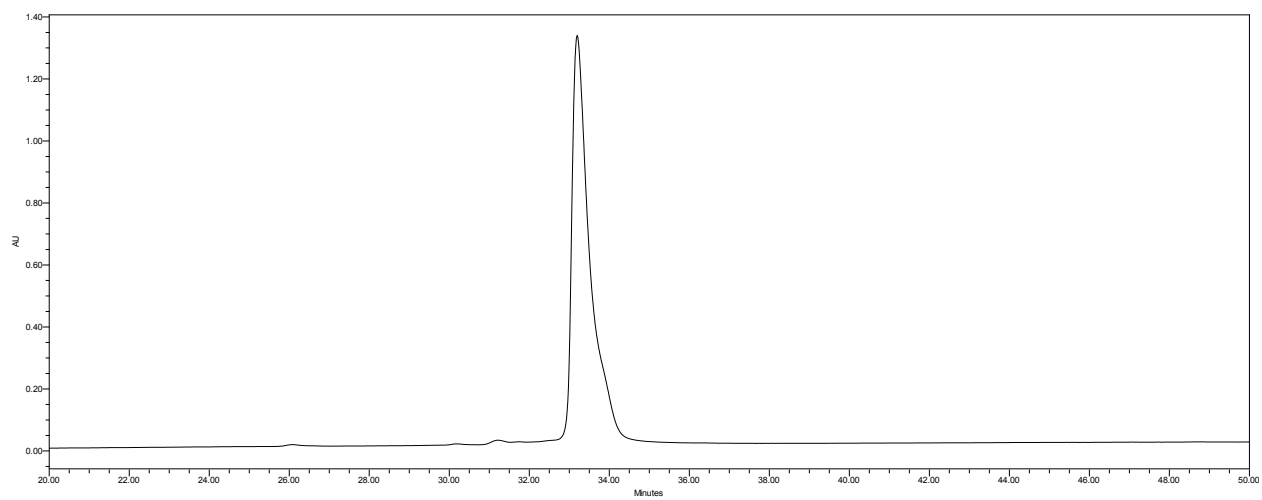**4:** RP-HPLC/UV chromatogram of crude PnID B containing two orthogonally protected ACM cysteines and two cysteines with free thiols.

5: RP-HPLC/UV chromatogram of crude PnID C containing two orthogonally protected ACM cysteines and two cysteines with free thiols.

6: RP-HPLC/UV chromatogram of crude PnID RevA containing two orthogonally protected ACM cysteines and two cysteines with free thiols.

7: RP-HPLC/UV chromatogram of crude PnID RevB containing two orthogonally protected ACM cysteines and two cysteines with free thiols.
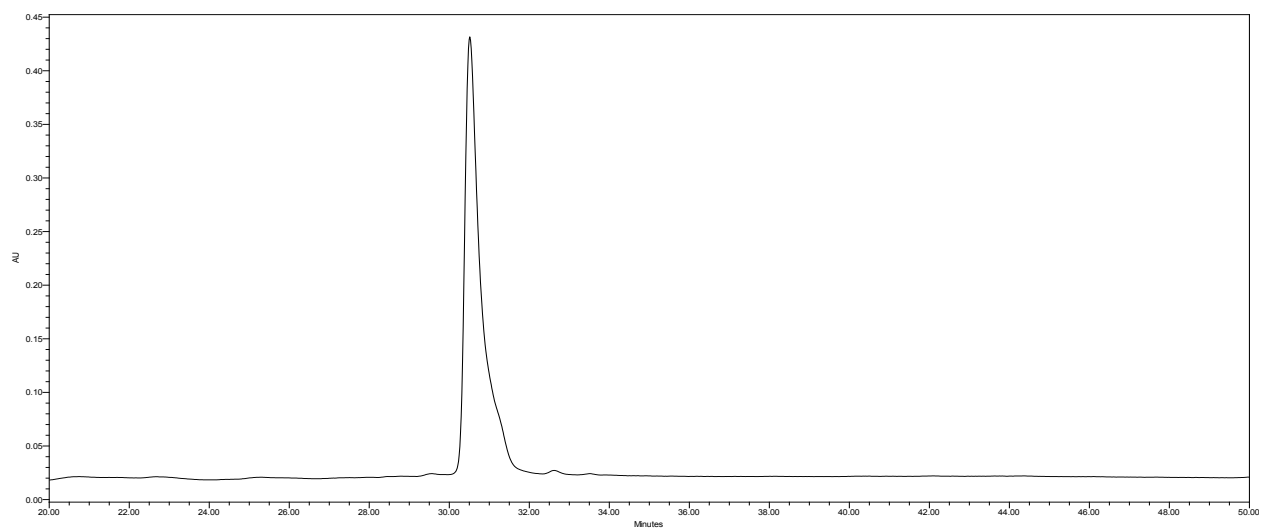
8: RP-HPLC/UV chromatogram of the fully folded and purified form of PnID B.
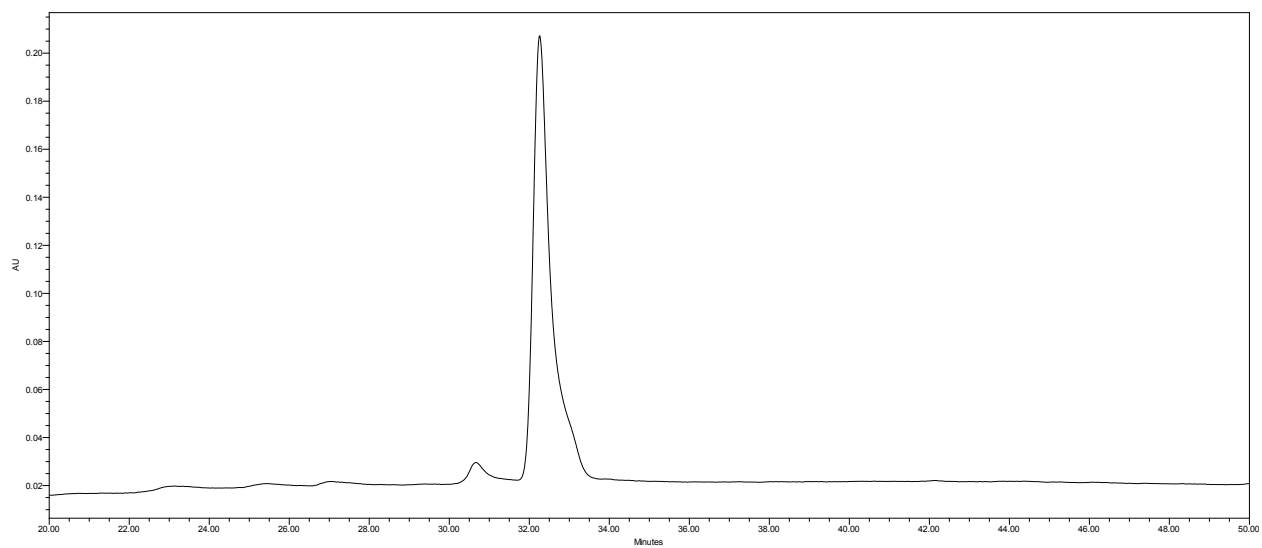
9: RP-HPLC/UV chromatogram of the fully folded and purified form of PnID C.

10: RP-HPLC/UV chromatogram of the fully folded and purified form of PnID RevA.

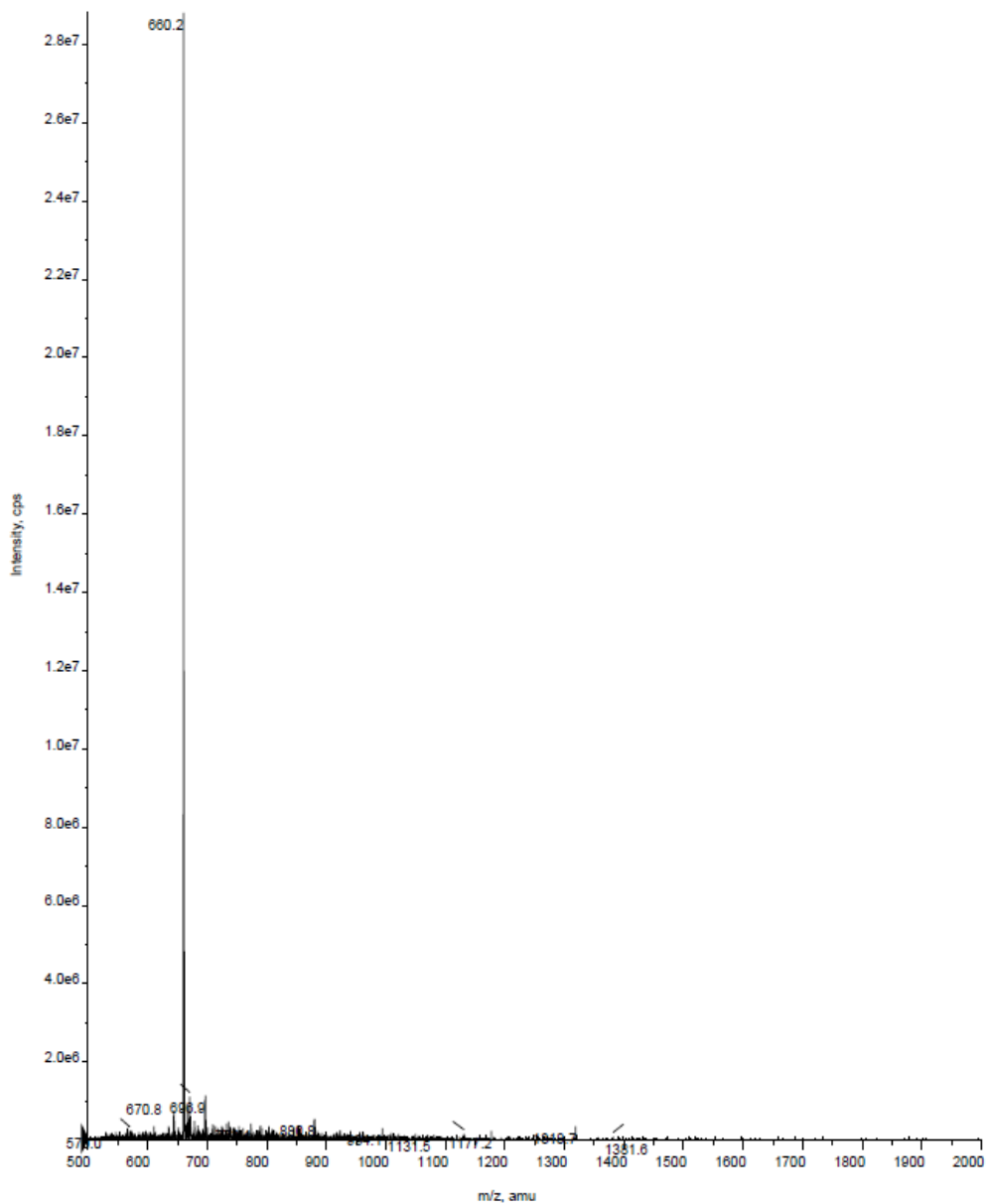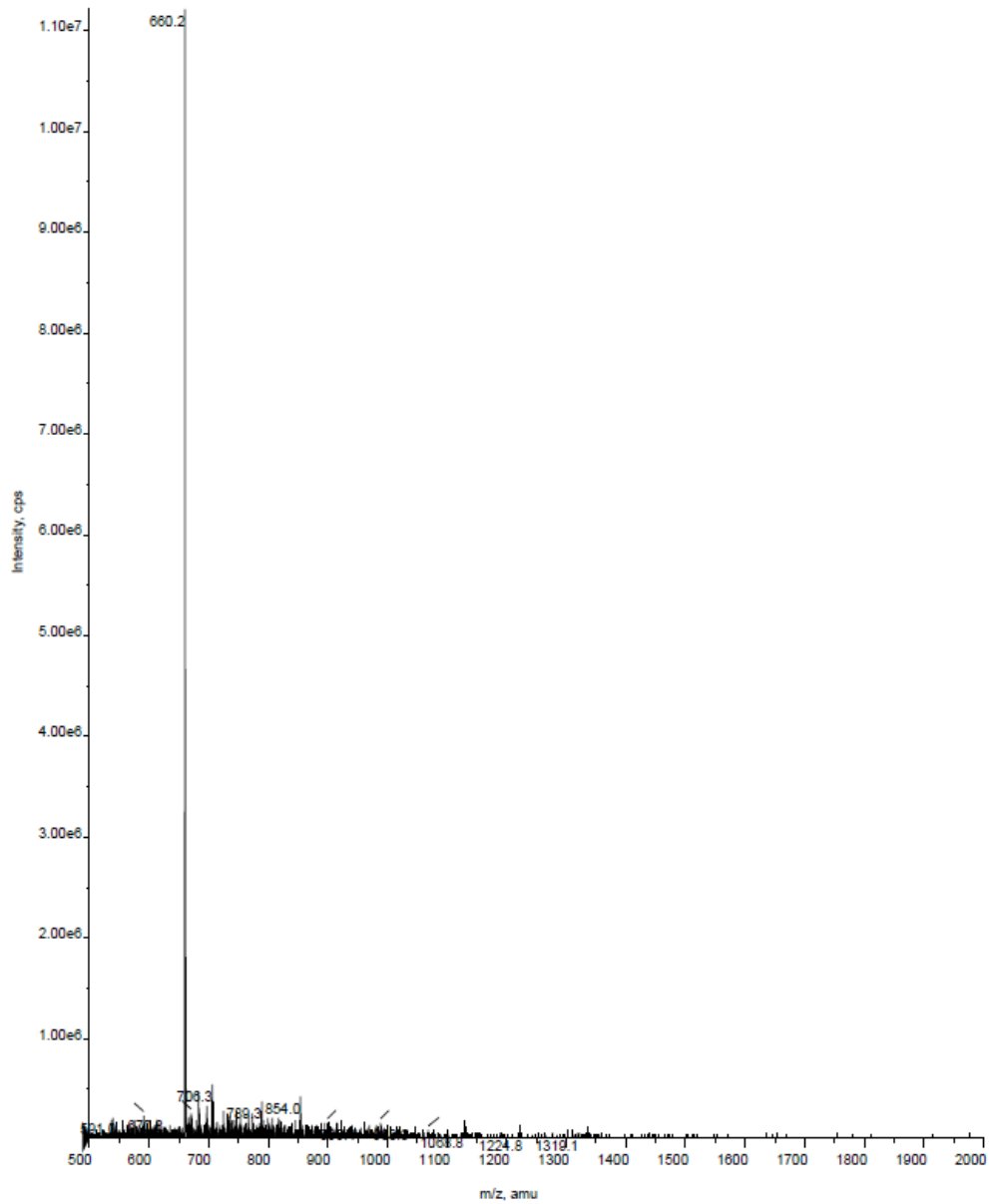# Appendix B. Mass spectra of the crude and purified PnID isomers
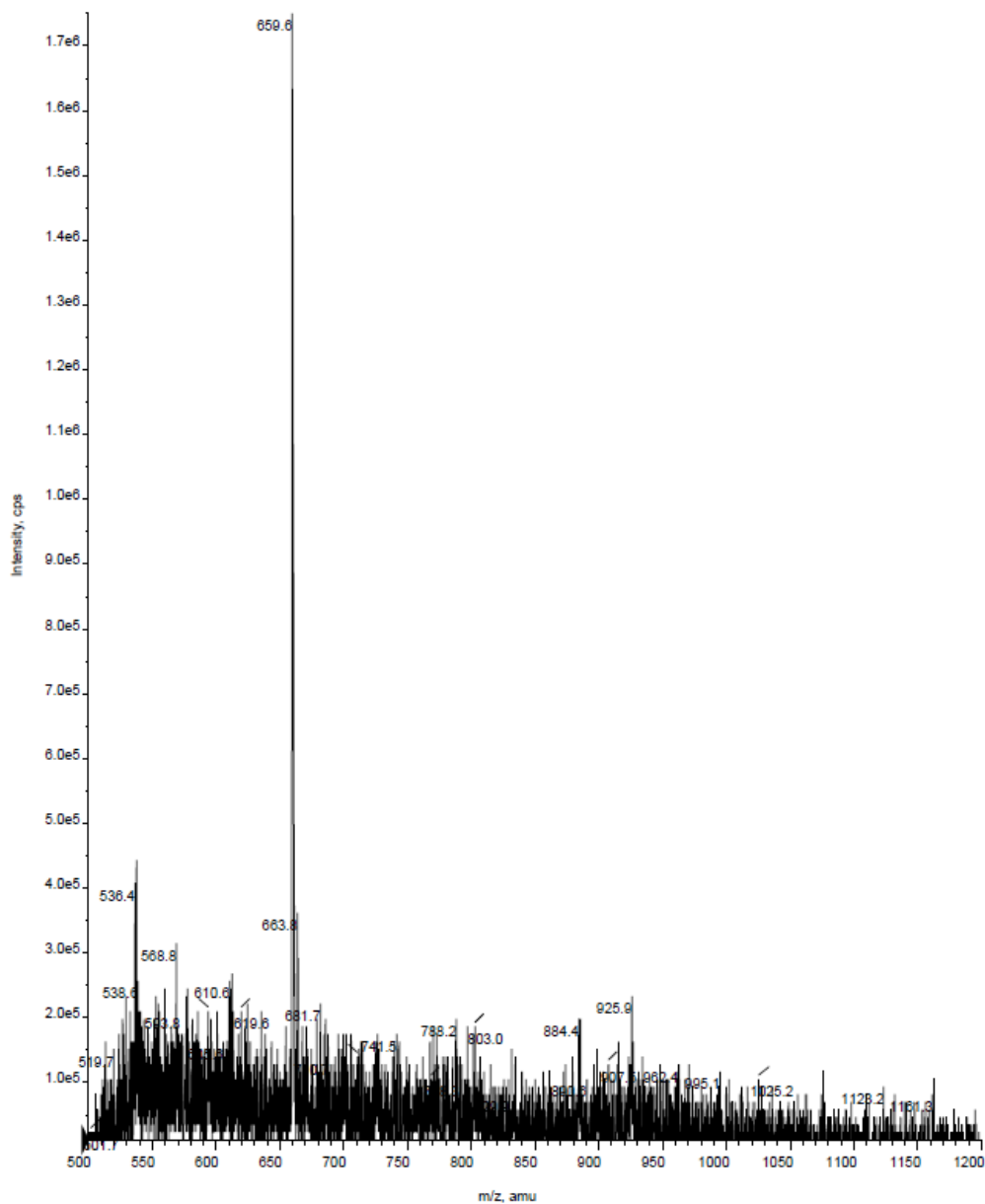
1: Mass spectrograph of the fully folded PnID containing no orthogonal protecting groups. 659.6 was the observed mass to charge ratio (*m/z*) in the second charge state. The expected *m/z* in this charge state was 660 +/- 0.5 .
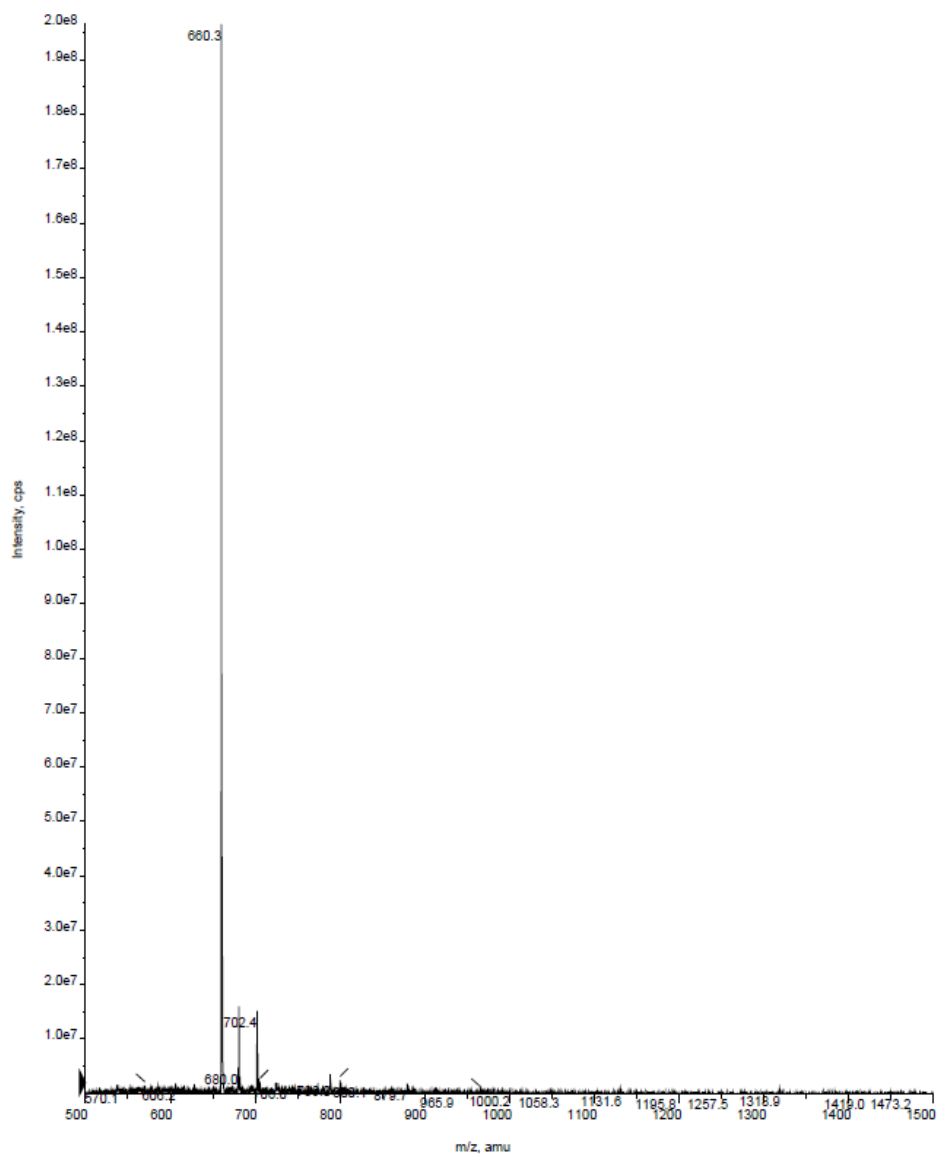
2: Mass spectrograph of the fully folded PnID A containing no orthogonal protecting groups. 660.3 was the observed mass to charge ratio (*m/z*) in the second charge state. The expected *m/z* in this charge state was 660 +/- 0.5 .

131
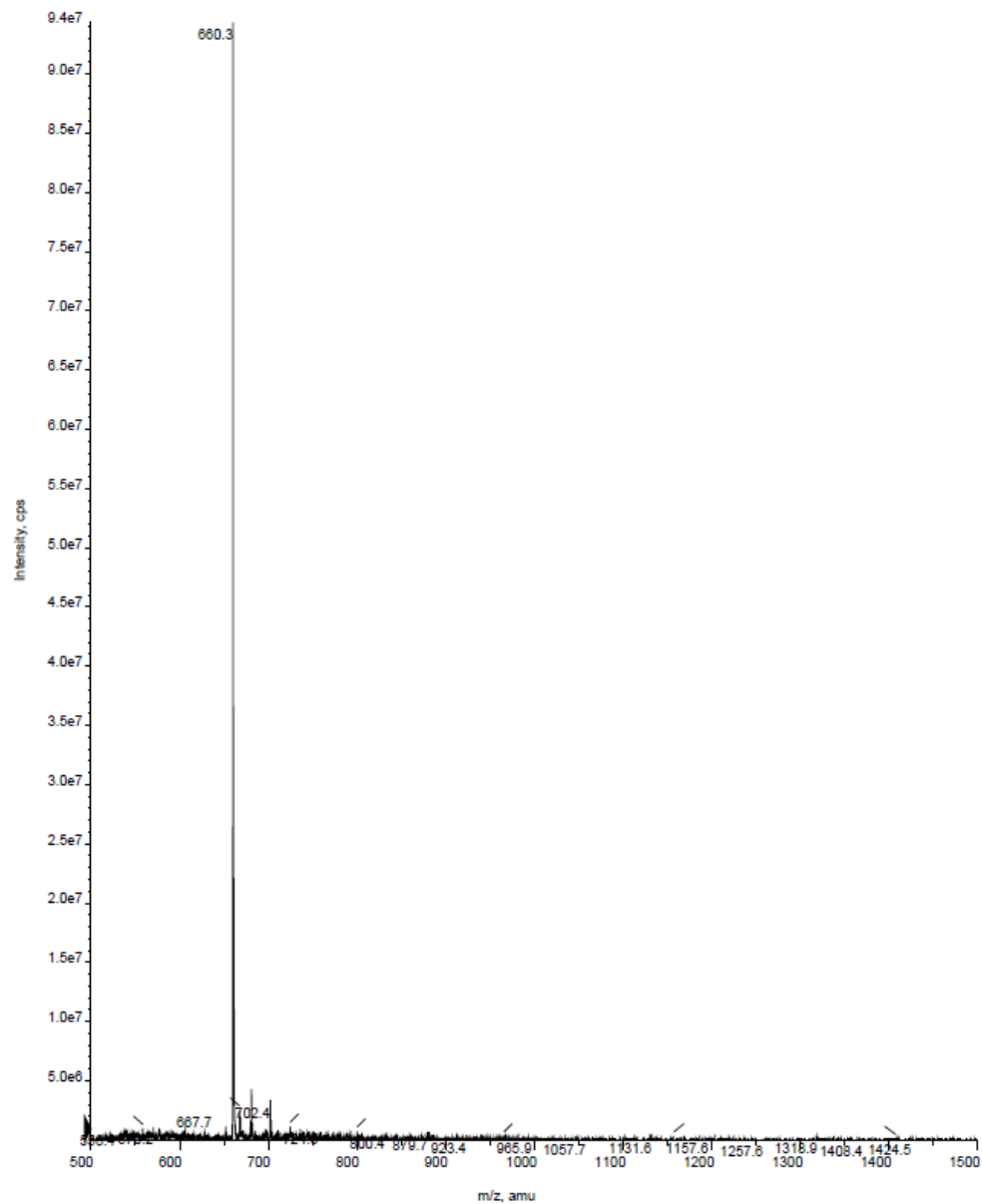
3: Mass spectrograph of the fully folded PnID B containing no orthogonal protecting groups. 660.3 was the observed mass to charge ratio (*m/z*) in the second charge state. The expected *m/z* in this charge state was 660 +/- 0.5 .
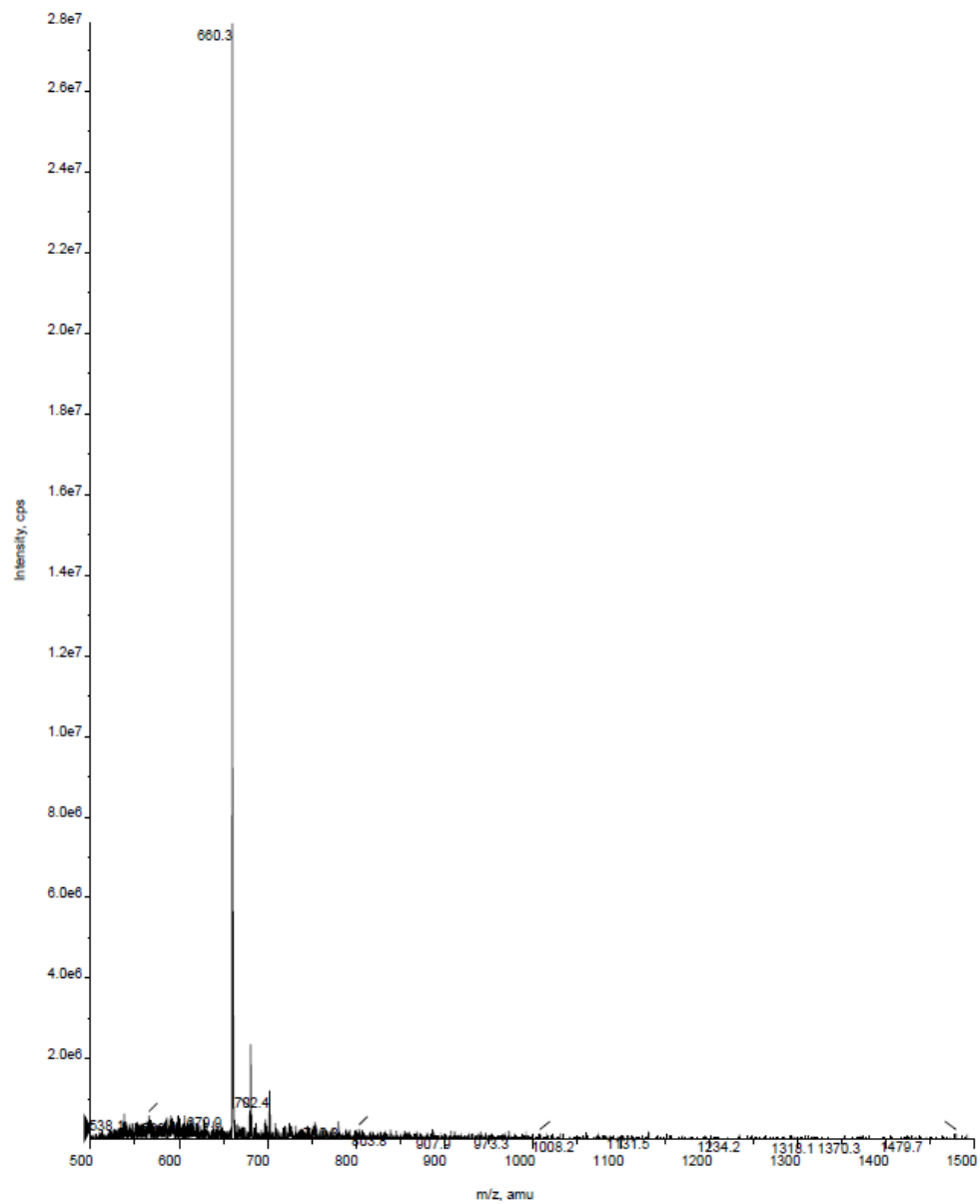
4: Mass spectrograph of the fully folded PnID C containing no orthogonal protecting groups. 660.3 was the observed mass to charge ratio (*m/z*) in the second charge state. The expected *m/z* in this charge state was 660 +/- 0.5 .

5: Mass spectrograph of the fully folded PnID RevA containing no orthogonal protecting groups. 660.2 was the observed mass to charge ratio (*m/z*) in the second charge state. The expected *m/z* in this charge state was 660 +/- 0.5 .

6: Mass spectrograph of the fully folded PnID RevB containing no orthogonal protecting groups. 660.2 was the observed mass to charge ratio (*m/z*) in the second charge state. The expected *m/z* in this charge state was 660 +/- 0.5

Appendix C. Model quality assessment of the submitted PnID structures

The following figures are the results of the verification software used by the World Wide Protein Databank (wwPDB) to assess the quality of the structures and determine if they are acceptable for deposition within the wwPDB.

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Full wwPDB NMR Structure Validation Report (i)

Sep 2, 2016 – 12:23 PM EDT

PDB ID : 5T6V
Title : Initial Topology of the globular isomer of PnID
Deposited on : 2016-09-01

This is a Full wwPDB NMR Structure Validation Report.

This report is produced by the wwPDB biocuration pipeline after annotation of the structure.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
http://wwpdb.org/validation/2016/NMRValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (i)) were used in the production of this report:

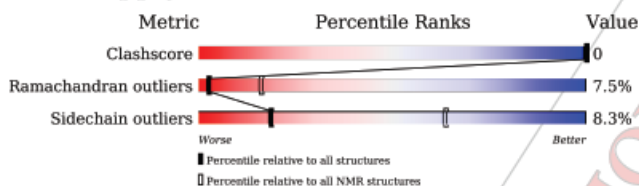|		
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027939
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027939

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
SOLUTION NMR

The overall completeness of chemical shifts assignment is 51%.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.

Metric	Percentile Ranks	Value
Clashscore		0
Ramachandran outliers		7.5%
Sidechain outliers		8.3%

■ Percentile relative to all structures
▯ Percentile relative to all NMR structures

Metric	Whole archive (#Entries)	NMR archive (#Entries)
Clashscore	114402	11133
Ramachandran outliers	111179	9975
Sidechain outliers	111093	9958

The table below summarises the geometric issues observed across the polymeric chains and their fit to the experimental data. The red, orange, yellow and green segments indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. A cyan segment indicates the fraction of residues that are not part of the well-defined cores, and a grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5%

Mol	Chain	Length	Quality of chain
1	A	13	69% 8% 23%

2 Ensemble composition and analysis ⓘ

This entry contains 20 models. Model 8 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *fewest violations*.

The following residues are included in the computation of the global validation metrics.

Well-defined (core) protein residues			
Well-defined core	Residue range (total)	Backbone RMSD (Å)	Medoid model
1	A:3-A:12 (10)	0.50	8

Ill-defined regions of proteins are excluded from the global statistics.

Ligands and non-protein polymers are included in the analysis.

The models can be grouped into 4 clusters and 4 single-model clusters were found.

Cluster number	Models
1	4, 7, 8, 11, 17, 19
2	2, 5, 12, 14
3	1, 3, 13, 20
4	6, 16
Single-model clusters	9; 10; 15; 18

3 Entry composition ⓘ

There is only 1 type of molecule in this entry. The entry contains 169 atoms, of which 82 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Chi-conotoxin-like PnMRCL-013.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	13	169	51	82	16	15	5	1

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	13	NH2	-	amidation	UNP Q9BPE9

4 Residue-property plots ⓘ

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA and DNA chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. Stretches of 2 or more consecutive residues without any outliers are shown as green connectors. Residues which are classified as ill-defined in the NMR ensemble, are shown in cyan with an underline colour-coded according to the previous scheme. Residues which were present in the experimental sample, but not modelled in the final structure are shown in grey.

- Molecule 1: Chi-conotoxin-like PnMRCL-013

Chain A:

4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1
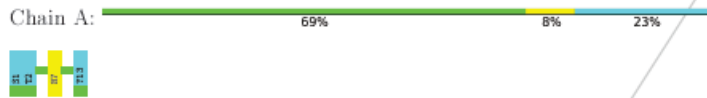
- Molecule 1: Chi-conotoxin-like PnMRCL-013

Chain A:

4.2.2 Score per residue for model 2

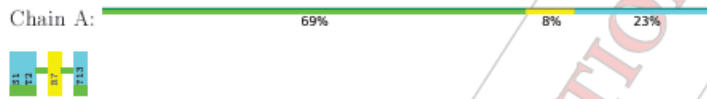- Molecule 1: Chi-conotoxin-like PnMRCL-013
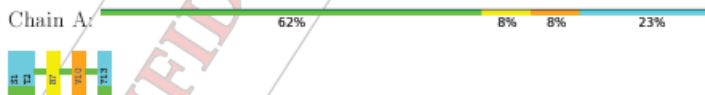
Chain A:

4.2.3 Score per residue for model 3

- Molecule 1: Chi-conotoxin-like PnMRCL-013

Chain A:

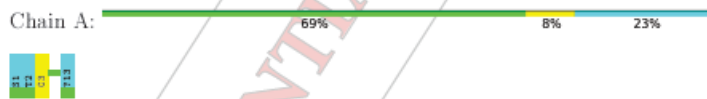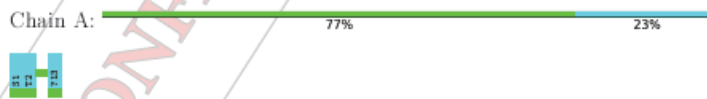### 4.2.4 Score per residue for model 4

- Molecule 1: Chi-conotoxin-like PnMRCL-013

Chain A:

4.2.5 Score per residue for model 5

- Molecule 1: Chi-conotoxin-like PnMRCL-013

Chain A:

4.2.6 Score per residue for model 6
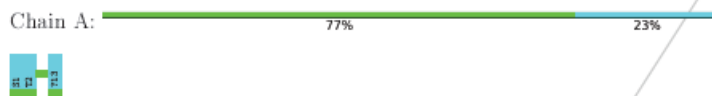
- Molecule 1: Chi-conotoxin-like PnMRCL-013

Chain A:

4.2.7 Score per residue for model 7

- Molecule 1: Chi-conotoxin-like PnMRCL-013

Chain A:

4.2.8   Score per residue for model 8 (medoid)

• Molecule 1: Chi-conotoxin-like PnMRCL-013

Chain A:

4.2.9   Score per residue for model 9

• Molecule 1: Chi-conotoxin-like PnMRCL-013

Chain A:

4.2.10   Score per residue for model 10

• Molecule 1: Chi-conotoxin-like PnMRCL-013
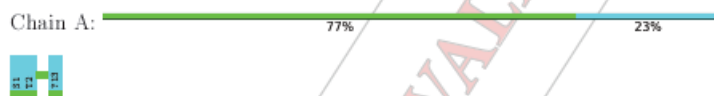
Chain A:

4.2.11   Score per residue for model 11

• Molecule 1: Chi-conotoxin-like PnMRCL-013

Chain A:

4.2.12   Score per residue for model 12

• Molecule 1: Chi-conotoxin-like PnMRCL-013

Chain A:

143

4.2.13 Score per residue for model 13

● Molecule 1: Chi-conotoxin-like PnMRCL-013

Chain A:

4.2.14 Score per residue for model 14

● Molecule 1: Chi-conotoxin-like PnMRCL-013

Chain A:

4.2.15 Score per residue for model 15

● Molecule 1: Chi-conotoxin-like PnMRCL-013
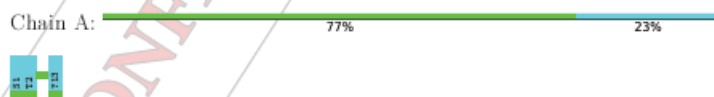
Chain A:

4.2.16 Score per residue for model 16

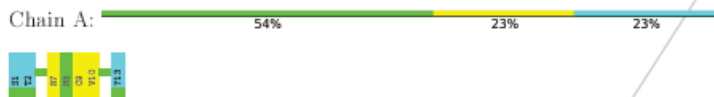● Molecule 1: Chi-conotoxin-like PnMRCL-013

Chain A:

4.2.17 Score per residue for model 17

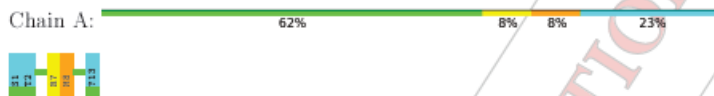● Molecule 1: Chi-conotoxin-like PnMRCL-013

Chain A:

4.2.18 Score per residue for model 18

● Molecule 1: Chi-conotoxin-like PnMRCL-013
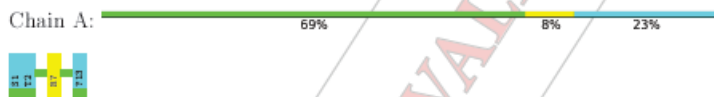
Chain A:

4.2.19 Score per residue for model 19

● Molecule 1: Chi-conotoxin-like PnMRCL-013

Chain A:

4.2.20 Score per residue for model 20

● Molecule 1: Chi-conotoxin-like PnMRCL-013

Chain A:

5 Refinement protocol and experimental data overview ⓘ

The models were refined using the following method: *simulated annealing*.

Of the 50 calculated structures, 20 were deposited, based on the following criterion: *structures with the lowest energy*.

The following table shows the software used for structure solution, optimisation and refinement.

Software name	Classification	Version
YASARA	refinement	15.10.18

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 7 of this report.

Chemical shift file(s)	D_1000223825_cs.cif
Number of chemical shift lists	1
Total number of shifts	70
Number of shifts mapped to atoms	70
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	51%

No validations of the models with respect to experimental NMR restraints is performed at this time.

6 Model quality ⓘ

6.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: NH2

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the (average) root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#Z>5	RMSZ	#Z>5
1	A	0.98±0.13	0±0/74 (0.0±0.0%)	1.13±0.14	0±1/98 (0.3±0.6%)
All	All	0.99	0/1480 (0.0%)	1.14	6/1960 (0.3%)

There are no bond-length outliers.

All unique angle outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)	Models	
								Worst	Total
1	A	7	ARG	NE-CZ-NH1	-6.84	116.88	120.30	5	1
1	A	7	ARG	NE-CZ-NH2	6.49	123.54	120.30	9	5

There are no chirality outliers.

There are no planarity outliers.

6.2 Too-close contacts ⓘ

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in each chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes averaged over the ensemble.

Mol	Chain	Non-H	H(model)	H(added)	Clashes
All	All	1460	1320	1320	-

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is -.

There are no clashes.

147

6.3 Torsion angles (i)

6.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all NMR entries. The Analysed column shows the number of residues for which the backbone conformation was analysed and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	10/13 (77%)	7±1 (67±12%)	3±1 (25±9%)	1±1 (8±6%)	3	16
All	All	200/260 (77%)	134 (67%)	51 (26%)	15 (8%)	3	16

All 3 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	7	ARG	7
1	A	10	VAL	7
1	A	8	MET	1

6.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all NMR entries. The Analysed column shows the number of residues for which the sidechain conformation was analysed and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	9/11 (82%)	8±1 (92±9%)	1±1 (8±9%)	19	64
All	All	180/220 (82%)	165 (92%)	15 (8%)	19	64

All 5 unique residues with a non-rotameric sidechain are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	4	CYS	5
1	A	9	CYS	4
1	A	3	CYS	3
1	A	8	MET	2
1	A	10	VAL	1

6.3.3 RNA ⓘ

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

6.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

6.6 Ligand geometry ⓘ

There are no ligands in this entry.

6.7 Other polymers ⓘ

There are no such molecules in this entry.

6.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

7 Chemical shift validation ⓘ

The completeness of assignment taking into account all chemical shift lists is 51% for the well-defined parts and 48% for the entire structure.

7.1 Chemical shift list 1

File name: D_1000223825_cs.cif

Chemical shift list name: *Bshifts.txt*

7.1.1 Bookkeeping ⓘ

The following table shows the results of parsing the chemical shift list and reports the number of nuclei with statistically unusual chemical shifts.

Total number of shifts	70
Number of shifts mapped to atoms	70
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

7.1.2 Chemical shift referencing ⓘ

No chemical shift referencing corrections were calculated (not enough data).

7.1.3 Completeness of resonance assignments ⓘ

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 51%, i.e. 57 atoms were assigned a chemical shift out of a possible 112. 0 out of 1 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	19/48 (40%)	19/19 (100%)	0/20 (0%)	0/9 (0%)
Sidechain	34/56 (61%)	34/35 (97%)	0/18 (0%)	0/3 (0%)
Aromatic	4/8 (50%)	4/4 (100%)	0/4 (0%)	0/0 (—%)
Overall	57/112 (51%)	57/58 (98%)	0/42 (0%)	0/12 (0%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 48%, i.e. 62 atoms were assigned a chemical shift out of a possible 129. 0 out of 1 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	21/58 (36%)	21/23 (91%)	0/24 (0%)	0/11 (0%)
Sidechain	37/63 (59%)	37/39 (95%)	0/21 (0%)	0/3 (0%)
Aromatic	4/8 (50%)	4/4 (100%)	0/4 (0%)	0/0 (—%)
Overall	62/129 (48%)	62/66 (94%)	0/49 (0%)	0/14 (0%)

7.1.4 Statistically unusual chemical shifts ⓘ

There are no statistically unusual chemical shifts.

7.1.5 Random Coil Index (RCI) plots ⓘ

The image below reports *random coil index* values for the protein chains in the structure. The height of each bar gives a probability of a given residue to be disordered, as predicted from the available chemical shifts and the amino acid sequence. A value above 0.2 is an indication of significant predicted disorder. The colour of the bar shows whether the residue is in the well-defined core (black) or in the ill-defined residue ranges (cyan), as described in section 2 on ensemble composition.

Random coil index (RCI) for chain A:

Full wwPDB NMR Structure Validation Report (i)

Sep 2, 2016 – 12:22 PM EDT

PDB ID   :   5T6T
Title   :   Reverse topology of the globular isoform of PnlD
Deposited on   :   2016-09-01

The following versions of software and data (see references (i)) were used in the production of this report:

Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027939
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027939

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
SOLUTION NMR

The overall completeness of chemical shifts assignment is 48%.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.

Metric	Whole archive (#Entries)	NMR archive (#Entries)
Clashscore	114402	11133
Ramachandran outliers	111179	9975
Sidechain outliers	111093	9958

The table below summarises the geometric issues observed across the polymeric chains and their fit to the experimental data. The red, orange, yellow and green segments indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. A cyan segment indicates the fraction of residues that are not part of the well-defined cores, and a grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5%

Mol	Chain	Length	Quality of chain
1	A	13	85% 8% 8%

2   Ensemble composition and analysis ⓘ

This entry contains 20 models. Model 2 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *lowest energy*.

The following residues are included in the computation of the global validation metrics.

Well-defined (core) protein residues			
Well-defined core	Residue range (total)	Backbone RMSD (Å)	Medoid model
1	A:1-A:12 (12)	0.62	2

Ill-defined regions of proteins are excluded from the global statistics.

Ligands and non-protein polymers are included in the analysis.

The models can be grouped into 2 clusters and 1 single-model cluster was found.

Cluster number	Models
1	1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 15, 16, 17, 18
2	6, 14, 20
Single-model clusters	19

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 169 atoms, of which 82 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Chi-conotoxin-like PnMRCL-013.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	13	169	51	82	16	15	5	1

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	13	NH2	-	amidation	UNP Q9BPE9

4　Residue-property plots ⓘ

4.1　Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA and DNA chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. Stretches of 2 or more consecutive residues without any outliers are shown as green connectors. Residues which are classified as ill-defined in the NMR ensemble, are shown in cyan with an underline colour-coded according to the previous scheme. Residues which were present in the experimental sample, but not modelled in the final structure are shown in grey.

● Molecule 1: Chi-conotoxin-like PnMRCL-013

Chain A:

4.2　Scores per residue for each member of the ensemble
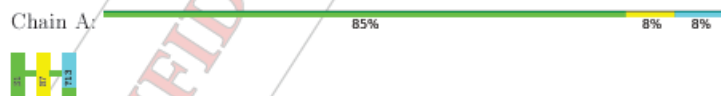
Colouring as in section 4.1 above.

4.2.1　Score per residue for model 1
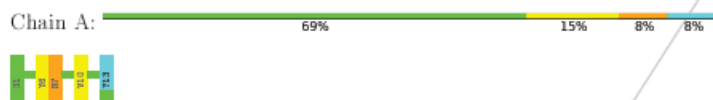
● Molecule 1: Chi-conotoxin-like PnMRCL-013

Chain A:

4.2.2　Score per residue for model 2 (medoid)

● Molecule 1: Chi-conotoxin-like PnMRCL-013
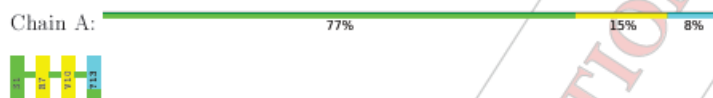
Chain A:

4.2.3   Score per residue for model 3

• Molecule 1: Chi-conotoxin-like PnMRCL-013
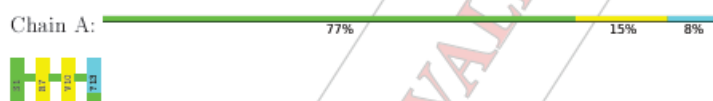
Chain A:

4.2.4   Score per residue for model 4

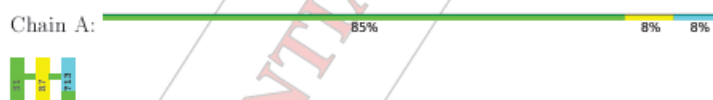• Molecule 1: Chi-conotoxin-like PnMRCL-013

Chain A:

4.2.5   Score per residue for model 5
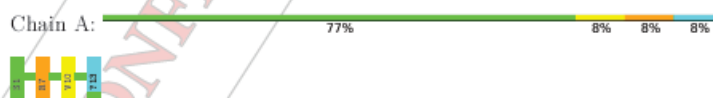
• Molecule 1: Chi-conotoxin-like PnMRCL-013

Chain A:

4.2.6   Score per residue for model 6

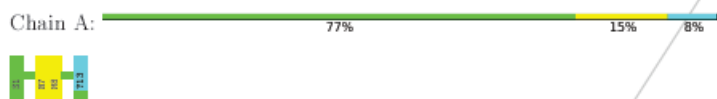• Molecule 1: Chi-conotoxin-like PnMRCL-013

Chain A:

4.2.7   Score per residue for model 7

• Molecule 1: Chi-conotoxin-like PnMRCL-013

Chain A:

4.2.8 Score per residue for model 8

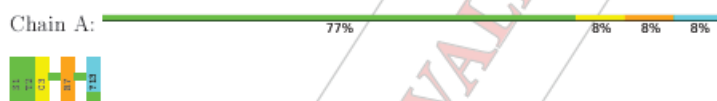• Molecule 1: Chi-conotoxin-like PnMRCL-013

Chain A:

4.2.9 Score per residue for model 9

• Molecule 1: Chi-conotoxin-like PnMRCL-013

Chain A:

4.2.10 Score per residue for model 10

• Molecule 1: Chi-conotoxin-like PnMRCL-013

Chain A:

4.2.11 Score per residue for model 11

• Molecule 1: Chi-conotoxin-like PnMRCL-013

Chain A:

4.2.12 Score per residue for model 12

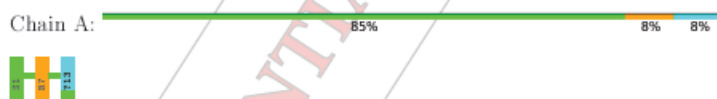• Molecule 1: Chi-conotoxin-like PnMRCL-013

Chain A:

4.2.13 Score per residue for model 13

- Molecule 1: Chi-conotoxin-like PnMRCL-013

Chain A:

4.2.14 Score per residue for model 14

- Molecule 1: Chi-conotoxin-like PnMRCL-013

Chain A:

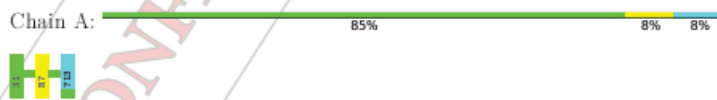### 4.2.15 Score per residue for model 15

- Molecule 1: Chi-conotoxin-like PnMRCL-013

Chain A:

4.2.16 Score per residue for model 16

- Molecule 1: Chi-conotoxin-like PnMRCL-013

Chain A:

4.2.17 Score per residue for model 17

- Molecule 1: Chi-conotoxin-like PnMRCL-013

Chain A:

4.2.18   Score per residue for model 18
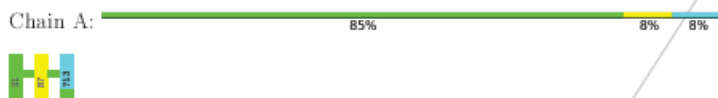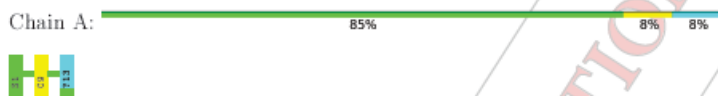
● Molecule 1: Chi-conotoxin-like PnMRCL-013

Chain A:

4.2.19   Score per residue for model 19

● Molecule 1: Chi-conotoxin-like PnMRCL-013

Chain A:

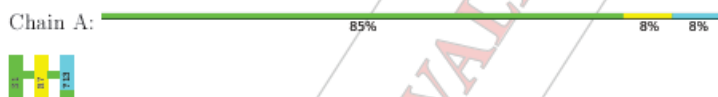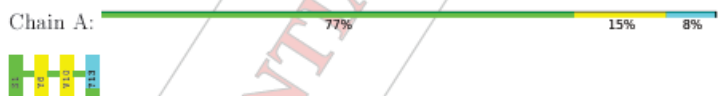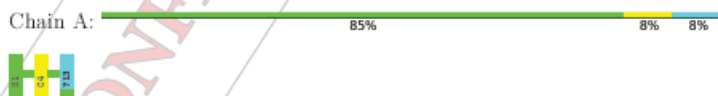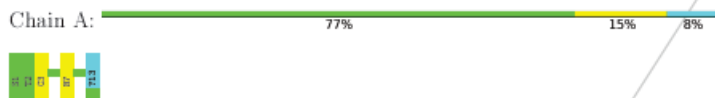### 4.2.20   Score per residue for model 20

● Molecule 1: Chi-conotoxin-like PnMRCL-013

Chain A:

5　Refinement protocol and experimental data overview ①

The models were refined using the following method: *simulated annealing.*

Of the 50 calculated structures, 20 were deposited, based on the following criterion: *structures with the lowest energy.*

The following table shows the software used for structure solution, optimisation and refinement.

Software name	Classification	Version
YASARA Structure	refinement	15.10.18

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 7 of this report.

Chemical shift file(s)	D_1000223628_cs.cif
Number of chemical shift lists	1
Total number of shifts	70
Number of shifts mapped to atoms	70
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	48%

No validations of the models with respect to experimental NMR restraints is performed at this time.

6 Model quality ⓘ

6.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: NH2

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the (average) root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#Z>5	RMSZ	#Z>5
1	A	1.00±0.12	0±0/87 (0.1±0.3%)	1.04±0.11	0±1/116 (0.3±0.5%)
All	All	1.00	1/1740 (0.1%)	1.05	8/2320 (0.3%)

All unique bond outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models Worst	Models Total
1	A	6	TYR	CG-CD2	5.19	1.45	1.39	9	1

All unique angle outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)	Models Worst	Models Total
1	A	7	ARG	NE-CZ-NH2	7.50	124.05	120.30	11	6
1	A	7	ARG	NE-CZ-NH1	-6.34	117.13	120.30	1	2

There are no chirality outliers.

There are no planarity outliers.

6.2 Too-close contacts ⓘ

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in each chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes averaged over the ensemble.

Mol	Chain	Non-H	H(model)	H(added)	Clashes
All	All	1720	1600	1600	-

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all NMR entries. The Analysed column shows the number of residues for which the backbone conformation was analysed and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	11/13 (85%)	8±1 (75±9%)	2±1 (21±9%)	0±1 (4±5%)	6	33
All	All	220/260 (85%)	165 (75%)	46 (21%)	9 (4%)	6	33

All 4 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	10	VAL	4
1	A	6	TYR	3
1	A	3	CYS	1
1	A	9	CYS	1

6.3.2 Protein sidechains ⓘ

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all NMR entries. The Analysed column shows the number of residues for which the sidechain conformation was analysed and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	11/11 (100%)	10±1 (90±5%)	1±1 (10±5%)	13	57
All	All	220/220 (100%)	197 (90%)	23 (10%)	13	57

All 7 unique residues with a non-rotameric sidechain are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	7	ARG	16

Continued on next page...

163

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	9	CYS	2
1	A	2	THR	1
1	A	4	CYS	1
1	A	3	CYS	1
1	A	8	MET	1
1	A	10	VAL	1

6.3.3 RNA ⓘ

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

6.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

6.6 Ligand geometry ⓘ

There are no ligands in this entry.

6.7 Other polymers ⓘ

There are no such molecules in this entry.

6.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

7   Chemical shift validation ⓘ

The completeness of assignment taking into account all chemical shift lists is 48% for the well-defined parts and 48% for the entire structure.

7.1   Chemical shift list 1

File name: D_1000223628_cs.cif

Chemical shift list name: *RevBshifts.txt*

7.1.1   Bookkeeping ⓘ

The following table shows the results of parsing the chemical shift list and reports the number of nuclei with statistically unusual chemical shifts.

Total number of shifts	70
Number of shifts mapped to atoms	70
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

7.1.2   Chemical shift referencing ⓘ

No chemical shift referencing corrections were calculated (not enough data).

7.1.3   Completeness of resonance assignments ⓘ

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 48%, i.e. 62 atoms were assigned a chemical shift out of a possible 129. 0 out of 1 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	21/58 (36%)	21/23 (91%)	0/24 (0%)	0/11 (0%)
Sidechain	37/63 (59%)	37/39 (95%)	0/21 (0%)	0/3 (0%)
Aromatic	4/8 (50%)	4/4 (100%)	0/4 (0%)	0/0 (—%)
Overall	62/129 (48%)	62/66 (94%)	0/49 (0%)	0/14 (0%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 48%, i.e. 62 atoms were assigned a chemical shift out of a possible 129. 0 out of 1 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

165

	Total	^{1}H	^{13}C	^{15}N
Backbone	21/58 (36%)	21/23 (91%)	0/24 (0%)	0/11 (0%)
Sidechain	37/63 (59%)	37/39 (95%)	0/21 (0%)	0/3 (0%)
Aromatic	4/8 (50%)	4/4 (100%)	0/4 (0%)	0/0 (—%)
Overall	62/129 (48%)	62/66 (94%)	0/49 (0%)	0/14 (0%)

7.1.4 Statistically unusual chemical shifts ⓘ

There are no statistically unusual chemical shifts.

7.1.5 Random Coil Index (RCI) plots ⓘ

The image below reports *random coil index* values for the protein chains in the structure. The height of each bar gives a probability of a given residue to be disordered, as predicted from the available chemical shifts and the amino acid sequence. A value above 0.2 is an indication of significant predicted disorder. The colour of the bar shows whether the residue is in the well-defined core (black) or in the ill-defined residue ranges (cyan), as described in section 2 on ensemble composition.

Random coil index (RCI) for chain A: