Expanding the Structural Biologist's Tool Chest by Chemical and Biocompatible Modification of Peptides: from Characterizing Marine Snail Toxins to OutPHOXIng Tryptophan

Alexander Grigoryan

Graduate Center, City University of New York

Recommended Citation
Grigoryan, Alexander, "Expanding the Structural Biologist's Tool Chest by Chemical and Biocompatible Modification of Peptides: from Characterizing Marine Snail Toxins to OutPHOXIng Tryptophan" (2015). CUNY Academic Works.
https://academicworks.cuny.edu/gc_etds/954
Expanding the Structural Biologist’s Tool Chest by Chemical and Biocompatible Modification of Peptides: from Characterizing Marine Snail Toxins to OutPHOXIng Tryptophan

Alexander Grigoryan

A dissertation submitted to the Graduate Faculty in Chemistry in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2015
Copyright © 2015 Alexander Grigoryan
This manuscript has been read and accepted by the Graduate Faculty in Chemistry in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Date________________________           Dr. Laura Juszczak

______________________________
Chair of the Examining Committee

Date________________________           Dr. Brian Gibney

______________________________
Executive Officer

Supervisory Committee:           Dr. Sébastien Poget
Dr. Richard Magliozzo
Dr. Ruel Desamero
Dr. Ryan Murelli
Abstract

Expanding the Structural Biologist’s Tool Chest by Chemical and Biocompatible Modification of Peptides: from Characterizing Marine Snail Toxins to OutPHOXIng Tryptophan

By

Alexander Grigoryan

Advisor: Dr. Laura Juszczak

This dissertation focuses on expanding chemical derivatization methodologies of peptides with the ultimate goal of studying proteins that are not amenable to traditional high-resolution structural techniques. The first project, conducted in the laboratory of Dr. Mandë Holford, uses improved synthetic and chemical derivatization methods to access and characterize peptidic natural products from venomous marine snails belonging to the family Terebridae. Closely related to Conus snails, the Terebridae also produce disulfide-rich peptide toxins (teretoxins) that are potent and specific agonists of ion channels and receptors of the central nervous system. Teretoxins are underexplored relative to Conus toxins, and represent a yet untapped pool of tools to investigate neuronal targets that are notoriously hard to characterize, as well as a source of therapeutic candidates for the treatment of neurological conditions. This work demonstrates both the application of specialized and improved solid phase peptide synthesis methods, and a strategy for disulfide mapping of teretoxins through the successful synthesis and structural characterization of the novel toxin Tv1 from Terebra variegata, the first teretoxin to be structurally characterized. The second project, conducted in the laboratory of Dr. Laura Juszczak, focuses on expanding the utility of fluorescence spectroscopy for the study of peptides and proteins by developing a novel, environment-sensitive fluorophore capable of biocompatible incorporation into polypeptides. Using the fluorogenic oxidative coupling reaction of 5-hydroxytryptophan and benzylamine which produces the fluorescent product 2-phenyl-6H-oxazolo[4,5-e]indole (PHOXI), a strategy for site-specific, biocompatible labeling of peptides with this probe is demonstrated. Investigation of PHOXI’s photophysics shows that it is a potent environment-sensitive probe, making it a valuable addition to the
privileged class of solvatochromic fluorophores capable of probing protein structure, interactions and dynamics.
Acknowledgements

Isolating as it may sometimes seem, doctoral research cannot be successful without the professional help and emotional support of others. Throughout my years at CUNY, I have been fortunate to be surrounded by supportive mentors, engaging faculty, and fellow graduate students who shared the ups and downs of doctoral study with me.

I would like to extend my gratitude to my first graduate advisor, Dr. Mandē Holford, who challenged me with engaging projects, and gave me the opportunity to travel the world by sending me to conduct research with collaborator Dr. Peter Seeberger in Berlin, and to represent the lab at conferences in Vancouver and Antigua. My work at the Holford lab would also be impossible without the guidance and support of Dr. Prachi Anand, a postdoctoral researcher who first shared the bench with me and taught me about the value of persistence in scientific inquiry, as well as how to keep a “real” lab notebook. My thanks also go out to members of the laboratory of Dr. Tom Muir, for insights into solid phase peptide synthesis; Dr. Emmanuel Chang for helpful discussions and the use of his mass spectrometers; Drs. Beatrix Ueberheide and Brian Chait who also helped with mass spectrometry; and Dr. Ivan Vilotijevic, an ebullient postdoctoral fellow who helped me go from never running a flash column to completing 17 synthetic steps during my two months in Berlin.

I would like to extend my utmost gratitude to my advisor Dr. Laura Juszczak, who not only inspired me to join the CUNY Ph.D. program, but also enabled me to finish my graduate work. I am also grateful to Dr. Azaria Eisenberg, a postdoctoral fellow in our lab, for daily engaging discussions and his help with quantum mechanical calculations. Numerous others at Brooklyn College have also contributed their time and resources to my dissertation project, including Dr. Peter Lipke and his lab members who helped with cloning and recombinant protein expression, fellow graduate student Joshua Jones who helped with organic synthesis, Dr. Andrzej Jarzecki who has been invaluable for interpreting excited state calculation results, and many others, too numerous to name.

My dissertation committee members have also been an instrumental resource throughout my time at CUNY. Dr. Ryan Murelli not only let me use his lab for organic synthesis required for my project, but guided me in developing my career. I am also grateful to him for giving me the opportunity to pursue my interest in organic synthesis as a future postdoctoral researcher in his laboratory. Likewise, I would
like to thank Dr. Ruel Desamero for lending his spectroscopic equipment and expertise, Dr. Sébastien Poget for help with structural characterization of Tv1, and Dr. Richard Magliozzo for use of his spectrophotometer and other equipment.

Finally, my family has been an enduring pillar of support throughout this, at times tenuous, experience. My mom, and my loving step dad Marc have always been willing to lend an understanding ear, and their unfaltering confidence in me was always inspiring. My husband Carlos always remained by side, wherever my graduate work has taken me, and helped me survive New York City on a graduate student’s stipend.
# Table of Contents

## Chapter 1 – Introduction

1.1 Peptides as tools to study proteins .......................................................... 2  
1.2 Marine snail venom peptides as probes neuronal proteins .......................... 4  
1.3 Chemical modification of peptides for fluorescence enhancement .................. 6

## Chapter 2 - Identification and Characterization of Peptide Marine Natural Products for Pharmaceutical and Neuroscience Applications

2.1 Toxins from venomous organisms as drug development and research tools .................. 12  
2.2 Conoida mollusks as a rich source of neuroactive peptides ............................... 12  
2.3 Conoidan toxin structure and function: lessons from cone snails ....................... 15  
2.4 The peptide toxins of the *Terebradae* ................................................................ 17  
2.5 Current approaches to conoidan toxin identification ........................................... 18  
2.6 Current approaches to conoidan toxin synthesis .............................................. 20  
2.7 Structural characterization of Conoidan toxins ................................................. 24  
2.8 Current approaches to conoidan toxin functional characterization ...................... 24  
2.9 Objectives of Thesis Project ........................................................................... 25

## Chapter 3 - Synthesis, oxidative folding of terebrid neuropeptides and structural characterization of a novel venom peptide Tv1 from the venomous marine snail *Terebra variaga*

3.1 Teretoxins: from identification to synthesis and characterization ....................... 28  
3.2 Results and discussion .................................................................................... 31  
3.2.1 Synthesis of selected terebrid neuropeptides .............................................. 37  
3.2.2 Oxidative folding of terebrid neuropeptides Tv1 and Arg58 ............................ 43  
3.2.3 Structural characterization of Tv1 ................................................................... 46  
3.3 Conclusions and future directions ................................................................... 49  
3.4 Materials and Methods ................................................................................... 51  
3.4.1 SPPS of terebrid neuropeptides ................................................................. 51  
3.4.2 Oxidative Folding of Teretoxin peptides ..................................................... 53  
3.4.3 Disulfide mapping of Tv1 ............................................................................. 53
Chapter 4 - Fluorescence Spectroscopy Using Solvatochromic Probes for Investigation of Protein Structure, Function and Interactions

4.1 Fluorescence spectroscopy in the study of biological systems .......................................................................................... 55
4.2 Protein fluorescence and the intrinsic fluorophore tryptophan ......................................................................................... 56
4.3 Tryptophan analogues for structure-conservative spectral enhancement of proteins .......................................................... 58
4.4 Extrinsic solvatochromic probes to study proteins ............................................................................................................. 60
4.4.1 The phenomenon of solvatochromism .......................................................................................................................... 61
4.4.2 Photophysical properties of solvatochromic probes ....................................................................................................... 63
4.4.3 Incorporation of solvatochromic probes into proteins and peptides .............................................................................. 67
4.5 Converting a tryptophan analogue to a sensitive extrinsic probe ....................................................................................... 69
4.6 Project research objectives ................................................................................................................................................. 70

Chapter 5 - Evaluation of 5HTP Derivatization in Peptide Systems and Structural and Solvatochromic Description of PHOXI

5.1 Background: Oxidative coupling of 5-hydroxyindole with primary amines ........................................................................ 73
5.2 Results and Discussion ......................................................................................................................................................... 78
5.2.1 Biosynthetic Incorporation of 5HTP into proteins ........................................................................................................... 79
5.2.2 Evaluation of PHOXI derivatization of model peptides ................................................................................................ 80
5.2.3 Photophysical and solvatochromic properties of PHOXI ................................................................................................. 86
5.2.4 DFT analysis of PHOXI structure and photophysics ....................................................................................................... 94
5.3 Conclusion and future directions ........................................................................................................................................... 95
5.4 Materials and Methods .......................................................................................................................................................... 97
5.4.1 Recombinant Expression and selective pressure incorporation of 5HTP into proteins ................................................... 98
5.4.2 PHOXI derivatization of peptides ................................................................................................................................ 99
5.4.3 Spectroscopic Measurements ......................................................................................................................................... 100
5.4.4 DFT calculations ............................................................................................................................................................... 102

References .............................................................................................................................................................................. 103
List of Figures

Chapter 1
1.1 Linus Pauling and Robert Corey with their model of the alpha helix, 1951 ................................2
1.2 Oxidative coupling of 5-hydroxytryptophan and benzylamine to produce 2-phenyl-6H-oxazolo[4,5-e]indole .............................................................................................................8

Chapter 2
2.1 Conoidan species and toxin diversity .............................................................................13
2.2 Conoidan venom apparatus ..........................................................................................14
2.3 Conoidan peptide toxin precursor organization ..........................................................16
2.4 Classification of conotoxins ..........................................................................................17
2.5 One-step and regioselective folding strategies for disulfide-rich peptides ......................22
2.6 Kinetically trapped intermediates in in vitro folding reactions .....................................23

Chapter 3
3.1 Traditional and concerted discovery strategies for teretoxins ......................................29
3.2 Straight chain Fmoc SPPS ..........................................................................................32
3.3 Tg55 convergent synthesis by SPPS and NCL ..............................................................33
3.4 Recombinant expression strategy for the synthesis of Tg77 .........................................35
3.5 Partial reduction, differential alkylation, and MS/MS analysis for disulfide mapping of teretoxins ................................................................................................................................37
3.6 HPLC monitoring of SPPS cleavage products of Tg55 .................................................38
3.7 UHPLC and MS of synthesized Arg58 ..........................................................................39
3.8 UHPLC and MS of synthesized Tv1 ..............................................................................40
3.9 ETD spectra of native and synthetic Tv1 .......................................................................41
3.10 SDS-PAGE analysis of expression and purification of Tg77 fusion protein ..................42
3.11 HPLC monitoring of enterokinase cleavage and LC-MS of recombinantly produced Tg77 ..........42
3.12 UHPLC monitoring of oxidative folding of Tv1 over time .........................................43
3.13 Optimization of oxidative folding of Tv1 ....................................................................44
3.14 UHPLC and MALDI-TOF analysis of purified folded Tv1 .........................................45
3.15 UHPLC and MALDI-TOF analysis of folded Arg58 ..................................................46
3.16 UHPLC analysis of partially reduced Tv1 peptides ................................................................. 48
3.17 NMR confirmation of Tv1 disulfide assignment ........................................................................ 49
3.18 Structural Comparison of Tv1 and M-superfamily conotoxins .................................................. 50

Chapter 4
4.1 Electronic transitions of indole ................................................................................................. 57
4.2 Structures and spectral characteristics of Trp and its analogues ................................................ 59
4.3 Simplified Jablonski diagram explaining the phenomenon of solvatochromism ...................... 61
4.4 Common solvatochromic probes for monitoring macromolecular structure, dynamics and
interactions ......................................................................................................................................... 64
4.5 Oxidative coupling of 5-hydroxyindole and primary amines .................................................... 70

Chapter 5
5.1 Oxidation products of 5HI under biomimetic conditions ........................................................ 73
5.2 Reactions of 9-hydroxyellipticium with nucleophiles under oxidative conditions .................... 74
5.3 Mechanism of oxazole formation during oxidation of 5HI in the presence of primary amines ...... 75
5.4 Fluorogenic oxidative coupling of 5HI and benzylamine .......................................................... 76
5.5 Mass spectra of tryptophanyl-5-hydroxytryptophan before and after PHOXI derivatization .... 81
5.6 HPLC monitoring and fluorescence enhancement of NH$_2$-Trp-5HTP upon derivatization
with PHOXI ....................................................................................................................................... 82
5.7 HPLC monitoring of TrpZip2(5HTP) before and after PHOXI derivatization ......................... 83
5.8 MALDI-TOF spectra of TrpZip2(5HTP) before and after PHOXI derivatization ..................... 83
5.9 HPLC monitoring of PHOXI derivatization of Aβ-40 .............................................................. 85
5.10 Normalized fluorescence excitation and emission of NH$_2$-Trp-5HTP-OH vs. NH$_2$-Trp-PHOX-OH
TrpZip2(5HTP) vs. TrpZip2(PHOXI) ............................................................................................... 87
5.11 Effect of PHOXI derivatization on β-hairpin stability of TrpZip2 probed by CD ......................... 88
5.12 Spectroscopic properties of PHOXI in solvents of varying polarity ........................................ 89
5.13 Fluorescence lifetime decay curves for PHOXI in six neat solvents ......................................... 91
5.14 Analysis of PHOXI lifetimes, quantum yields, and rates of radiative and non-radiative decay using
the Catalán 3P model ........................................................................................................................ 93
List of Tables

Chapter 2

2.1 Teretoxin neuropeptides identified for synthesis ................................................................. 26

Chapter 3

3.1 Teretoxin neuropeptides identified for development ............................................................. 31

3.2 Predicted and observed of b and y ions of differentially alkylated peptides by auto and targeted MS/MS analysis ........................................................................................................................................ 47

Chapter 4

Table 5.1: Spectroscopic properties of PHOXI in neat solvents ..................................................... 89

Table 5.2 Solvent parameters used for analysis of PHOXI solvatochromism ............................... 92

Table 5.3: Comparison of PHOXI with other solvatochromic probes with absorption <400nm .......... 95
Chapter 1 - Introduction
“I believe, anticipate that the chemist of the future who is interested in the structure of proteins […] will come to rely upon a new structural chemistry, […] and that great progress will be made, through this technique, in the attack, by chemical methods, on the problems of biology and medicine.”

Linus Pauling, Nobel Lecture (11 December 1954)\textsuperscript{1}

1.1 Peptides as tools to study proteins

The understanding of protein structure, dynamics and interactions is central to modern structural biology, as underscored by Linus Pauling’s prophetic quote above\textsuperscript{1}. In many ways, with the discovery and continued sophistication of high-resolution spectroscopic, chemical and biochemical techniques for probing these enigmatic macromolecules, Pauling’s predicted future has arrived. However, the immense
chemical and structural complexity of proteins continues to challenge even today’s chemists and spectroscopists.

One way to overcome this complexity is to work with peptides, or mini-proteins, instead. Pauling and Corey used just this tactic when they first described the α-helix and β-sheet as structural motifs², which have since been found in innumerable proteins. Peptides as tools for structural investigation of their larger protein cousins offer many advantages to the experimenter, both as structural analogues of proteins and as simpler systems amenable to chemical modification.

The conception and development of solid phase peptide synthesis (SPPS) by Merrifield in 1963³ provided reliable chemical synthetic access to these molecules, and allowed their functionalization with chemical groups normally absent from biological systems. In terms of structural biology and medicine, SPPS gave rise to three important advancements: (i) It provided synthetic access to peptidic natural products, (ii) allowed for the synthesis of “designer” peptides which were used as models to better understand protein structure and folding, and (iii) made it possible to include non-native reporter moieties into peptides, which enabled investigation of their structure and interaction by spectroscopic means.

Building on the synthetic access to peptides and versatility of their chemical modification afforded by SPPS, this work focuses on expanding chemical derivatization methodologies of these molecules with the ultimate goal of studying protein that are not amenable to traditional high-resolution structural techniques. In the first part of this research, improved synthetic and chemical derivatization methods are developed to access and characterize peptidic natural products from venomous marine snails. These peptides bind ion channels and receptors of the nervous system, which are transmembrane proteins that are notoriously difficult to investigate by high-resolution techniques such as X-ray crystallography and nuclear magnetic resonance spectroscopy (NMR). The second part of the work utilizes model peptides to develop a novel, environment-sensitive fluorophore capable of biocompatible incorporation into polypeptides, and is applicable to the study of a larger subset of proteins, which may not be amenable to high-resolution structural methods and can not be adequately described by fluorescence spectroscopy utilizing intrinsic probes.
1.2 Marine snail venom peptides as probes of neuronal proteins

Ion channels and receptors are critical components of the nervous system. Characterizing the structure and function of these proteins not only allows for the development of pharmaceutical agents for neurological conditions, but also enables the mapping of neural networks, furthering the understanding of neuroscience as a whole. However, because these proteins are large, consist of multiple subunits, and span the cellular membrane, the description of their structure and function is extremely difficult using traditional techniques. Transmembrane proteins have both hydrophilic and hydrophobic surface exposed domains, making it difficult to find suitable crystallization conditions for X-ray structure determination. NMR analyses of these molecules are also challenging, often because of the difficulty of obtaining sufficient quantities of correctly folded protein during recombinant expression. Finally, attempts to describe their function by gene knockout methods are often inadequate, as deletion of a single gene coding for particular subunit interrupts the function of a whole subset of ion channels and receptors that contain it, resulting in a complex phenotype.4

One way to work around these challenges has been to use binding partners of these ion channels and receptors. It is possible to glean structural and functional information about these proteins, provided that the structure of the binding ligand is known, and that a method of tracking its interaction with the target exists. One source of such binding partners are venom peptides from marine snails of the superfamily Conoida, which produce a plethora of potent and specific modulators of ion channels and receptors of the nervous system to capture and immobilize their prey.4

From a methodology perspective, three main challenges exist in the development of conoidan peptide toxins for pharmaceutical and neuroscience research: gaining synthetic access, characterizing their structure, and describing their biological interactions with specific targets of the nervous system. While the venom glands of these animals are large enough to identify and elucidate the sequence of peptide toxins of interest using existing methods, chemical synthesis by SPPS is necessary to obtain sufficient material for structural and functional testing. Furthermore, the high disulfide content and size of some of these peptides necessitate customization and improvement of existing SPPS methodology. In the current work, access to larger toxin peptides utilized a number of SPPS improvements such as
microwave-assisted coupling\textsuperscript{5}, use of improved polymer solid support resins with superior swelling properties, and use of other techniques of preventing on-resin peptide aggregation.

NMR is traditionally used for structure determination of these peptide toxins, but unambiguous structural assignment often requires isotopic enrichment, making such analyses both time-consuming and costly. However, owing to the high disulfide content of these peptides, additional structural detail can be ascertained by determination of disulfide connectivity by complementary chemical methods, thus aiding NMR structural assignment even without isotopic enrichment. Existing techniques for disulfide mapping of proteins\textsuperscript{6}, which require enzymatic digestion followed by mass spectrometric analysis of the resultant fragments, are not applicable to conoidan toxins because these peptides often have adjacent cysteine residues and lack sufficient protease cleavage sites. Instead, a technique of partial disulfide reduction and differential alkylation, followed by tandem mass spectrometry analysis, was adopted for this purpose in the current work.

Finally, while several techniques for identifying the specific ion channels or receptors that interact with a particular peptide toxin exist, most are cost-prohibitive and time-consuming, and thus low throughput. The contemporary development of microarray technology\textsuperscript{7}, which allows for rapid screening of thousands of compounds immobilized on a glass slide against their binding partners, promises an attractive alternative to existing electrophysiology and cell-based assay methods. The strategy envisioned in the Holford laboratory includes constructing a toxin peptide microarray. Treating the toxin microarray with \textit{Xenopus laevis} oocytes coexpressing a fluorescent protein (a description of \textit{Xenopus} oocyte expression system for functional studies of ion channels can be found here\textsuperscript{8}) and the ion channel or receptor of interest would enable rapid screening of a large number of toxins against a particular target. Work focused on developing a heterobifunctional carbohydrate based linker that could be used for site-specific immobilization of conoidan toxin peptides on a microarray slide without adversely affecting their function by way of a bioorthogonal and reversible enzymatic coupling reaction was also pursued as part of this dissertation project. Despite significant progress towards creation of this linker, commercially available peptide microarray immobilization strategies were pursued instead, therefore this project will not be discussed in this dissertation.
In summary, this work demonstrates both the application of specialized and improved SPPS methods, and the novel strategy of disulfide mapping of conoidan toxin peptides through the successful synthesis and structural characterization of the novel toxin Tv1 from *Terebra variegata*.

1.3 Chemical modification of peptides for fluorescence enhancement

The experiences gained through working with conoidan peptides have set the stage for using peptide chemical modification strategies to move beyond merely ion channels and receptors, and investigate a larger subset of proteins for which high-resolution structural techniques are either inadequate or impossible to implement. Complementary to structural techniques such as X-ray crystallography or NMR, fluorescence spectroscopy has also been used to provide valuable structural, binding, and dynamics data on a variety of biological systems. It does so by measuring light absorption and emission characteristics of a fluorescent probe, whose singlet excited state is sensitive to the local chemical environment.

In the context of protein investigations, the intrinsic fluorophore tryptophan (Trp) has been widely employed as a probe. Using Trp offers a number of advantages: its presence in many proteins allows for their investigation without any structural perturbation and proteins that lack Trp can be engineered to include the residue at a desired position by site-directed mutagenesis. However, Trp is far from an ideal fluorophore: high energy ultraviolet radiation that is required for tryptophan excitation may be damaging to some proteins, and is less than desirable for biological fluorescence microscopy. Also, while its fluorescence emission band is sensitive to the local environment, its complex photophysics often preclude meaningful structural interpretation of its fluorescence signal characteristics. Furthermore, fluorescence analyses of proteins containing multiple Trp residues, or in the presence of exogenous Trp do not yield structural data because it is impossible to deconvolute signals from individual tryptophans.

Fortunately, the utility of protein fluorescence spectroscopy has been greatly expanded by employing extrinsic fluorophores. In developing such probes, parallel monumental achievements have been made in both the rational design and synthesis of fluorophores with desirable photophysical properties, and in their methods of incorporation into biological molecules. Modern synthetic methods and powerful calculation techniques allow for rational design of fluorophores with a great variety properties. Concurrently, a number of protein incorporation methods have been developed, either by reactions with
intrinsic amino acid functional groups\textsuperscript{11}, bioorthogonal coupling methods\textsuperscript{12}, or incorporation of the fluorophore as an amino acid using amber (TAG) codon suppression techniques\textsuperscript{13}. However, continued development in both areas is required to give access to probes that have both desirable properties and are amenable to a specific incorporation method.

While there are many probe incorporation strategies, the most desirable, and the most challenging introduce the fluorophore at a unique site within the target protein, use a short and rigid tether to do so, and do not otherwise affect the overall structure or append the target protein with exogenous groups or sequences. This kind of incorporation maximizes the amount of structural information that can be gained from a fluorescence experiment because the probe is restricted to its surrounding microenvironment in a known orientation\textsuperscript{14}. To date, there are only two methods that can meet these requirements, and neither one is applicable to every protein.

The first is expressed protein ligation (EPL)\textsuperscript{15,16}, which allows for bioorthogonal coupling of peptide thioesters assembled by SPPS with a target protein containing an N-terminal cysteine residue via a native peptide bond. Since SPPS allows for incorporation of unnatural fluorophores by chemical methods, EPL can subsequently be used to append the probe bearing peptide onto a recombinantly synthesized protein. An obvious limitation of EPL is the requirement of a cysteine residue in favorable position relative to the probe site.

The second method relies on incorporating a fluorophore as its amino acid derivative in response to an unused stop codon during recombinant protein expression by chemically acylating it onto a modified tRNA\textsuperscript{17}. This technique is extremely difficult to implement, often suffers low incorporation yields, results in truncated recombinant proteins, and is not necessarily compatible with any unnatural amino acids\textsuperscript{13-14}.

In the quest to develop a fluorophore that possesses desirable photophysics and is amenable to site-specific incorporation, this project focuses on exploring the fluorogenic oxidative coupling reaction of 5-hydroxytryptophan (5HTP) and benzylamine which produces the fluorescent product 2-phenyl-6\textit{H}-oxazolo[4,5-e]indole (PHOXI) (Figure 1.2).
This reaction was utilized in analytical chemistry investigations for spectrofluorometric
determination of neurotransmitters and related metabolites containing the 5-hydroxyindole (5HI) moiety. More recently, the reaction was also used to cyclize peptides containing 5HTP and a synthetic amino acid with a benzylamine side chain. Although these studies did not focus on using this reaction as means of fluorescently tagging peptides and proteins, the reaction conditions and the few photophysical properties of PHOXI they identified prompted this line of inquiry in the current work.

From an incorporation standpoint, the PHOXI producing reaction is reported to be selective for 5HIs, and can be conducted under aqueous mildly oxidative conditions that should not affect other functional groups present in peptides and proteins. Since there exist straightforward methods for incorporation of 5HTP into proteins and peptides (namely by selective pressure incorporation during recombinant expression in tryptophan auxotrophs, and through standard 9-fluorenylethoxycarbonyl (Fmoc) SPPS), this chemistry could be subsequently used for site-specific labeling with PHOXI. Furthermore, since this reaction modifies an existing amino acid side chain, the PHOXI fluorophore would be located within the biomolecule in a predictable position, attached to the backbone with a single methylene bridge. This type of incorporation provides a complementary strategy to the inherently difficult amber codon and EPL based approaches.

From the photophysical properties standpoint, PHOXI was reported to have high brightness (through a combination of a high quantum yield and extinction coefficient) and emission in the visible range, both desirable properties for a fluorophore reporter tag. While these properties are encouraging, a more detailed investigation of the photophysics of PHOXI, including its sensitivity to the chemical
microenvironment, Förster resonance energy transfer (FRET) interaction with intrinsic fluorophores, and time-resolved fluorescence properties is needed to glean its utility as a probe for biological systems.

The present work uses model peptides containing 5HTP to evaluate the biocompatibility and selectivity of PHOXI derivatization. While the reaction is selective for 5HTP, it proceeds through a highly reactive \( p \)-quinoneiminemethide intermediate, which is susceptible to capture by endogenous nucleophiles other than the intended benzylamine, resulting in a number of undesired byproducts. Optimization of the reaction conditions using model systems allowed for minimization of these side-products and a high-performance liquid chromatography (HPLC) purification method was developed to isolate the PHOXI labeled peptides for spectroscopic and structural studies. The substrate scope of the reaction was also briefly investigated by using methylamine and ethanolamine in addition to benzylamine. While methylamine did not produce any fluorescent adducts under the conditions tested, ethanolamine produced three different products with fluorescence in the visible region, suggesting that a number of PHOXI analogues can be synthesized by varying the amine nucleophile.

Recombinant synthesis of proteins containing 5HTP was also attempted in hopes of evaluating the PHOXI derivatization reaction at the protein level. A variation on the selective pressure incorporation technique using standard \textit{Escherichia coli} (\textit{E. coli}) hosts in the presence of glyphosate (an inhibitor of aromatic amino acid biosynthesis), instead of the traditionally employed auxotrophic hosts was used. Unfortunately, attempts to synthesize two different 5HTP-containing proteins using two different expression systems and three bacterial host strains were not met with success. Because recombinant synthesis of 5HTP-containing proteins has already been demonstrated by other groups\textsuperscript{20}, this was not pursued further, and only 5HTP peptides were used to evaluate this chemistry due to commercial availability and ease of handling.

The spectroscopic properties of PHOXI and PHOXI-labeled model peptides were also extensively interrogated as part of this project. Absorption, fluorescence excitation and emission, lifetimes, and quantum yields of PHOXI evaluated in six neat solvents, were found to vary greatly as a function of solvent polarity. This sensitivity to the environment puts PHOXI in a privileged class of solvatochromic probes capable of interrogating the peptide/protein microenvironment. Furthermore, PHOXI’s photophysics are straightforward and show more dependence on general parameters of
polarity/polarizability, and are less affected by specific solvent interactions that can complicate structural attribution of spectral data. These findings show that PHOXI's photophysics are comparable to or better than those of other solvatochromic probes, such as Prodan\textsuperscript{22} or Dansyl\textsuperscript{23}, that found applications in biological systems.

As in Pauling's vision of the future, this dissertation describes the use and chemical modification of peptides in developing methods for structural, and functional characterization of otherwise intractable biological systems. This "attack, by chemical methods" was used to describe the structure of toxin peptides capable of providing valuable insight into the workings of their protein binding partners; and to use peptides as proving ground to develop an environment-sensitive probe capable of describing an even larger subsection of proteins by improved fluorescence methods.
Chapter 2 - Identification and Characterization of Peptide Marine Natural Products for Pharmaceutical and Neuroscience Applications
2.1 Toxins from venomous organisms as drug development and research tools

Disulfide-rich venom peptide toxins found in organisms such as scorpions, snails, snakes, spiders and leeches are potent and specific modulators of biological targets, making them attractive targets for drug development and downstream research applications\(^{24}\). In terms of pharmaceutical applications, nature-inspired peptide therapeutics are gaining traction as an alternative to small molecule drugs\(^{25}\). Peptides, although not as bioavailable as small molecule drugs, have advantages of higher target specificity and selectivity, and decreased toxicity\(^{26}\). Recent successes in development of peptide pharmaceutics, including captopril\(^{27}\), and more recently Prialt\(^{28}\), highlight the promise and potential of these compounds. Furthermore, the extraordinary affinity and specificity of these toxins contributes to their ability to characterize the structure and function of their biological targets. For instance, conotoxin GVIA, a Ca\(^{2+}\) channel agonist, has been used as a tool in over 2000 investigations for probing the molecular features of the nervous system\(^{29}\). With the overall diversity of venom-derived disulfide-rich peptides estimated at well over two million, and with their biological targets ranging from ion channels and receptors, enzymes and growth factors to structural or ligand binding domains of proteins\(^{30}\), the potential of these compounds as therapeutics and research tools is only beginning to be realized.

2.2 Conoida mollusks as a rich source of neuroactive peptides

Venomous marine snails belonging to the superfamily Conoida represent one source of disulfide-rich toxin peptides. Conoidan snails are classified into three groups: cones (Conidae), terebrids (Terebridae), and turrids (Turridae) (Figure 2.1). Among these three groups, the venom of cone snails has attracted the most research interest, particularly driven by the need to glean the pharmacological basis of human fatalities caused by Conus envenomation.
Conoida snails are present in tropical waters worldwide, with various species preying on fish (piscivorous), worms (vermivorous), or mollusks (molluscivorous). These snails are considerably slower than their prey, and thus had to evolve their arsenal of chemical toxins and methods for their delivery for the past 50 million years\textsuperscript{31}. The product of this evolution is a sophisticated envenomation apparatus (Figure 2.2) consisting of a venom bulb, a long venom duct equipped with radular teeth which act as a hypodermic needle for venom delivery\textsuperscript{32}.

While the sophistication of their venom apparatus is already awe inspiring, these snails’ ability to produce a plethora of peptide neurotoxins with diverse biological targets rivals that of today’s best-equipped combinatorial chemists. In fact, the genes encoding peptide toxins evolve at a much higher rate than housekeeping genes in these snails\textsuperscript{33}. It has been found that these peptide toxins are highly effective as tools for manipulating neuronal targets, including sodium (Na\textsuperscript{+}), potassium (K\textsuperscript{+}) and calcium channels (Ca\textsuperscript{2+}), noradrenaline transporters, and nicotinic acetylcholine receptors (nAChRs)\textsuperscript{34}. It was hypothesized that myriad conopeptides act in concert on the nervous system of the prey in synergistic “cabals”\textsuperscript{4}. The first, dubbed the “lightning-strike cabal”, leads to rapid immobilization of prey by what
amounts to electrocution through slowing Na\(^+\) channel inactivation while simultaneously inhibiting K\(^+\) channels. The second “motor” cabal finishes the job, by using slower acting toxins to deaden neuronal circuits at the neuromuscular junction, producing a soporific response. Some cone snails also employ a third “nirvana” cabal, releasing sensory jamming peptides into the water to pacify small schools of fish before engulfing them with their huge false mouth in a “net” hunting strategy. While it may seem prudent to question the therapeutic utility of conoidan peptides, since they are supposed to bind to physiologically relevant targets of the snails’ natural prey (fish, worms or mollusks), these targets have a high degree of sequence and structural homology to their corresponding human isoforms.

![Figure 2.2: Conoidan Venom Apparatus. (A) Components of the venom apparatus include the venom bulb, venom duct and salivary gland where toxins are produced, a sheath containing radular teeth that are used like hypodermic needles to deliver the toxins through the proboscis. (B) A Scanning electron micrograph close-up of a radular tooth, showing its harpoon-like structure and opening for venom delivery.](image)

Figure 2.2: Conoidan Venom Apparatus. (A) Components of the venom apparatus include the venom bulb, venom duct and salivary gland where toxins are produced, a sheath containing radular teeth that are used like hypodermic needles to deliver the toxins through the proboscis. (B) A Scanning electron micrograph close-up of a radular tooth, showing its harpoon-like structure and opening for venom delivery. Figure reproduced from reference 32.

It is difficult to estimate the vast number of biologically active peptides that can be identified from Conoida snails. A typical modern venom transcriptome sequencing experiment can identify 50-100 toxin-encoding genes in a single species\(^{35}\). Due to extensive post-translational processing and modification, the number of mature peptide toxins present in one specimen can be an order of magnitude higher. For example, 2428 peptide masses were identified from the venom of *Conus textile*\(^{36}\). Factors like
intraspecies, and even intraspecimen venom heterogeneity, inflate the number of putative toxins even further. Considering that there are about 700 known species of cone snails alone (disregarding turrids and terebrids), the most conservative estimate is approximately 50,000 toxin sequences at the genetic level, and over 1,000,000 at the peptide level. Given these numbers, the effort to identify and characterize these toxins has been vastly lacking in throughput with only 2837 nucleotide toxin sequences identified from 80 Conus species, and just 175 toxins from 34 Conus species structurally characterized\(^\text{37}\) (data from ConoServer accessed on 5/16/15). It is important to note that the above data represent only the toxin diversity and identified peptide toxins among Conus species, while the toxins of the Terebridae (approximately 300-400 species) and the Turridae (up to 10,000 species) remain largely unexplored. Specifically, this work focuses on the peptides of the Terebridae, an under-investigated sister group of the cones. While this group was chosen in hopes of identifying peptides with novel structures or targets, many descriptions of conopeptides, such as their structural and gene organization features, also hold true for terebrids and serve as a foundation to guide this inquiry.

2.3 Conoidan toxin structure and function: lessons from cone snails

The laboratories of Baldomero (Todo) Olivera, Paul Alewood and Richard Lewis pioneered the subject of conopeptide research, and in many ways shaped the body of knowledge in the field today\(^\text{24a, 32, 38}\). Early venom fractionation experiments identified two types of peptide components: disulfide poor (such as contulakins\(^\text{39}\) and conantokins\(^\text{40}\)) and disulfide-rich, termed conotoxins. These multiply disulfide-bonded conotoxins are the major component of the venom peptide repertoire of Conus snails, and target ion channels and receptors with extreme potency and selectivity. To date, conotoxins have been found to interact with a variety of neuronal targets, including Ca\(^{2+}\), Na\(^{+}\), K\(^{+}\) voltage- and ligand-gated channels, nAChRs, neurotensin receptors, N-methyl-D-aspartate (NMDA) receptors, norepinephrine transporters, and G-protein coupled receptors (GPCRs)\(^\text{41}\). With such diversity of targets, it is not surprising to see conotoxins as subjects of clinical trials as neuropathic pain analgesics, anti-convulsants\(^\text{42}\), anti-cancer agents\(^\text{43}\), and cardioprotectants\(^\text{44}\); with one conotoxin (MVIA from Conus magus, known as Prialt® or Ziconotide) approved for the treatment of chronic and severe pain\(^\text{28}\).
Conoidan toxins are single gene products, with messenger RNA (mRNA) translated into a precursor comprised of a highly conserved signal N-terminal signal sequence, an intervening and somewhat more variable pre-pro region, and a C-terminal mature toxin sequence (Figure 2.3).

![Figure 2.3: Conoidan peptide toxin precursor organization](image)

Mature conotoxin peptides, liberated by proteolysis from the signal and pre-pro regions, have a conserved pattern of cysteine residues and hypervariable intercysteine loops. The typical size of mature conoidan peptides is 10 to 60 amino acids, with at least 2 and as many as 10 cysteine residues. The highly conserved signal sequences are used to classify conotoxins into gene superfamilies, the pattern of cysteine residues into frameworks, and their biological targets into pharmacological families. All mature conotoxins are post-translationally modified to contain disulfide bonds, and some have been found to contain a number of other post-translational modifications (PTMs), most common of which are C-terminal amidation, proline hydroxylation, and glutamate γ-carboxylation. Although most of these PTMs are found in other species, they occur at a higher rate in these short peptides, making extensive PTM one of the defining features of conotoxins. The significance of PTMs in conotoxins is still being explored, and it has been suggested that they can play either functional or structural roles. The roles of PTMs are complex and enigmatic, as they may be required for function or for the correct folding of one conotoxin, while having no effect on the activity or fold of another. It has been hypothesized that PTMs in cone snails are yet another level of refining their toxins, analogous to drug lead optimization done by pharmaceutical companies.
2.4 The peptide toxins of the Terebridae

Being a sister group of the cones, the Terebridae also express an arsenal of disulfide-rich conotoxin-like peptides (teretoxins). While there are only a few studies of terebrid venom, it is known that their toxins arise from mRNA with organization analogous to that of the cones, containing a signal sequence, pre-pro, and mature peptide regions. Interestingly, the signal sequences of the Terebridae have a low degree of homology to the cones, suggesting that Conus gene superfamilies do not extend to the Terebridae. Consistent with this earlier study, a recent report identified 14 terebrid gene superfamilies, with only one of them having a strong (80%) sequence identity with a conotoxin superfamily (H). Although it is too early to draw any conclusions given the limited number of studies on the Terebridae, the disulfide-rich toxins identified so far have been longer (40 amino acids and greater) than Conus sequences, making teretoxins more similar to those of spiders and snakes. While most conserved cysteine frameworks found in the Terebridae are the same as those found in cones, two novel
frameworks containing 10 and 12 cysteines were recently identified. The greater length of teretoxins also contributes to greater numbers of residues in the intercysteine loops relative to conotoxins. Yet another consequence of length is the prevalence eight-cysteine (rare in cones) and the scarcity of four-cysteine frameworks (prevalent in cones) in the toxin peptides of the _Terebridae_.

The few studies conducted so far suggest that terebrid neuropeptides are free from PTMs, aside from C-terminal amidation. A recent second generation sequencing analysis of the venom duct transcriptome of two terebrid species, _T. annilis_ and _T. subulata_, identified nucleotide sequences corresponding to four PTM enzymes (γ-glutamyl carboxylase, peptidyl-glycine-α-amidating monooxygenase, prolyl-4-hydroxylase, tyrosyl sulfotransferase and glutamyl-peptide cyclotransferase) in the specimens. Although the presence of these enzymes suggests that PTMs cannot be ruled out for teretoxins, these proteins are quite ubiquitous across other mollusks and even mammals, and are also implicated in a variety of housekeeping roles. Therefore, the last word on the presence of PTMs in teretoxins will have to come from characterization of mature venom peptides.

The seeming absence of PTMs in terebrid neuropeptides suggests that toxin sequences identified at the genetic level can be directly synthesized and investigated for their therapeutic potential with more confidence. However, their extended length is anticipated to pose challenges during SPPS synthesis, the efficiency of which decreases with peptide length. Teretoxin synthesis by recombinant protein expression is one strategy that can be used to address potential SPPS challenges.

### 2.5 Current approaches of conoidan toxin identification

Early experiments aimed at identifying neuropeptide components in snail venom relied on HPLC venom fractionation, followed by Edman sequencing of peptide fractions exhibiting biological activity. Subsequent investigations employed mass spectrometry (MS) instead of Edman degradation for neuropeptide sequence elucidation, which had the advantages of requiring a substantially smaller sample size and the ability to detect PTMs. The lower sample requirements allowed for the identification of a larger number of venom components, but the throughput of whole venom fractionation was still limited due to difficulties in obtaining enough crude venom for complete analysis.

Rapid advances in DNA sequencing technology have introduced a paradigm shift in snail venom identification practices, which now focused on identifying potential neuropeptides by analyzing the snails’
venom duct transcriptome. In a typical experiment, mRNA is extracted from the specimen’s venom duct, and is subsequently used to construct a complimentary DNA (cDNA) library using in-vitro reverse transcription followed by polymerase chain reaction (PCR) amplification. The amplified cDNA fragments are then sequenced, and the resulting sequences are mined for defining features of disulfide-rich toxin transcripts: the presence of a conserved signal sequence, and multiple cysteine codons in familiar frameworks in the mature peptide region. The improved throughput of second generation sequencing technology allowed for analyses like these to be done at the genomic level, and further lowered the sample size requirements and costs of transcriptomic analyses.

In addition to lower costs and sample size requirements, genetic identification strategies also allow for targeted toxin discovery. Namely, information about signal sequences (and thus the gene superfamily) as well as the cysteine residue arrangement (and thus the framework) allows for selecting potential toxins that are inferred to be members of a particular pharmacological family. Alternatively, transcripts with divergent signal sequences or rare or unknown cysteine frameworks can be selected in hopes of identifying a novel biological target, or peptides with alternative interaction sites with traditional targets. Coupling this genetic toxin identification strategy with cladistic analysis, as was done in our laboratory for teretoxin identification, allows for a rational biodiversity guided discovery approach.

In spite of the advantages offered by neuropeptide identification by gene sequencing methods, it has drawbacks when compared to direct venom fractionation. The obvious limitation is that information available from genomic or transcriptomic analyses only provides the primary sequence of the target peptide. The native configuration of disulfide bonds cannot be obtained from such an experiment. Although it is sometimes possible to glean the probability of a certain PTM by the presence of a gene for a corresponding enzyme, it is impossible to say with certainty that this PTM will be present in mature peptide, or which specific residue(s) will be affected. Finally, predicting the cleavage site of the mature toxin from an mRNA sequence can also be difficult. Thus, although genetic level toxin identification has produced a number of advances, mass spectrometric interrogation of mature peptides from venom is still very relevant, and is often used in combination with genomic/transcriptomic data in an integrated toxin discovery strategy, which offers greater insights into toxin structure, PTMs, and the native folded state.
Because of this continued relevance, modern advances in MS guided proteomics and \textit{de novo} sequencing methods continue to contribute to the field of snail neuropeptide discovery. For example, the identification of teretoxin Tv1 from \textit{Terebra variagata} (the synthesis, structural and functional characterization of which is described as part of this dissertation), was accomplished by using just 4ng of material extracted from the snail’s salivary gland by utilizing improved MS \textit{de novo} sequencing methods. Namely, the approach relies on using electron transfer dissociation (ETD) in addition to the ubiquitous collision aided dissociation (CAD) for tandem MS (MS/MS) fragmentation, which greatly improves sequence coverage. This method is specifically adapted to disulfide-rich toxins, because it relies on derivatizing cysteine residues into dimethyl lysine analogues to achieve the higher charge states required for ETD.

\subsection*{2.6 Current approaches to conoidan toxin synthesis}

Peptide toxins have to be chemically synthesized to afford sufficient quantities for subsequent structural and functional characterization. In the case of conotoxins, the preferred method is SPPS, as it is highly efficient for short peptide sequences, amenable to automation, and offers the flexibility of introducing PTMs where desired. Robust, standard SPPS protocols utilizing either \textit{tert}-butoxycarbonyl (Boc) or 9-fluorenylmethoxycarbonyl (Fmoc) for reversible \(\alpha\)-amino protection exist. While Boc-SPPS can be highly efficient at preventing on-bead peptide aggregation, it requires a specialized set-up for the resin cleavage step, which uses gaseous hydrogen fluoride. In contrast, Fmoc chemistry does not require highly specialized equipment, and although peptide aggregation can be a problem for certain sequences, a number of techniques have been developed to mitigate these problems. The common issues encountered in snail peptide SPPS are either associated with decreasing coupling and deprotection efficiency with increasing peptide length, or with specific sequence-level features in a given peptide. Some SPPS improvements to mitigate these issues include: engineering of solid supports with better swelling properties in solvents of varying polarity, design of more efficient coupling agents, aggregation-disrupting pseudoproline backbone protection, dipeptide coupling to prevent aspartimide formation, and microwave-assisted coupling. The synthetic utility of SPPS is further increased to longer peptides and proteins when it is coupled with bioconjugation methods such as native chemical ligation.
(NCL)\(^{59}\), a chemoselective joining of a peptide thioester with an N-terminal cysteine containing peptide resulting in a native peptide bond.

Although longer peptide sequences of the *Terebridae* could be accessed with SPPS with the aforementioned improvements, their presumed lack of PTMs also makes them amenable to biosynthesis through recombinant expression. While cheaper than SPPS, recombinant expression is not usually used for the synthesis of peptides, which tend to be unstable in common host organisms such as *E. Coli*. As a workaround solution, peptides can be expressed as a fusion construct by appending their sequence onto that of a protein that has good cytosolic solubility and high expression levels in the host\(^{60}\). By engineering a specific protease cleavage site between the neuropeptide and fusion protein partner coding regions, it is possible to obtain the peptide of interest without any exogenous sequences.

After synthesis by either SPPS or recombinant expression, formation of disulfide bonds is accomplished by a variety of oxidative folding protocols. While during SPPS peptides are always initially formed in the linear (reduced) form, recombinant expression may produce a peptide with preformed disulfides. While the latter case is unlikely in the common host *E. coli* because it has a reducing cytosol, it is possible if a different host organism or specialized strain is used. Even if disulfides are formed during recombinant expression, the peptide may still need to be subjected to reduction and refolding because the disulfides formed in the host organism can be in a non-native arrangement.

Oxidative folding strategies for peptides containing multiple disulfides can be broadly classified into two groups: random oxidative folding of fully deprotected peptides, and directed regioselective disulfide bond formation (Figure2.5)\(^{61}\). During random folding, the fully deprotected peptide is dissolved in a slightly alkaline buffer at high dilution (to prevent dimer and oligomer formation) and mild oxidizing agents such as *N,N*-dimethylsulfoxide (DMSO) or thiol exchange reagents (such as a mixture of oxidized and reduced glutathione) are added to facilitate disulfide bond formation. Polymer supported oxidizing agents, such as ClearOx(cross-linked ethoxylate acrylate resin loaded with Ellman’s reagent) have also been used for conotoxin folding\(^{62}\), and provide the advantage of conducting folding reactions in reduced volumes of buffer because of a pseudodilution effect. While cheap and easy to implement, random folding is complicated by the number of disulfide bonds to be formed due to the number of possible isomers, \((2n)!/(2^n n!)\) where \(n\) is the number of disulfides, that can arise during the process.
Although the presence of thiol exchange reagents attempts to mimic physiological folding conditions allowing disulfides to reshuffle and direct the fold to the most thermodynamically stable conformation, low yields can be observed due to formation of a number of kinetically trapped isomers (Figure 2.6). Furthermore, the predominant isomer formed during random folding may not be the same as the native form, since in the snails disulfides are formed during post-translational processing prior to cleavage of the signal sequence and pre-pro regions, which aid in the recruitment of protein disulfide isomerase (PDI) that in turn facilitates folding. Nevertheless, non-native disulfide-bond isomers may be active analogues with interesting pharmacological properties, thus they can also be of interest. Random folding experiments can be optimized by HPLC monitoring of reaction products while varying the reaction temperature, time, peptide and thiol exchange reagent concentrations, as well as buffer salt content.
Alternatively, if the disulfide arrangement of the toxin peptide is known, directed regioselective folding could be employed to obtain only the desired isomer (Figure 2.5, right). In this approach, pairs of cysteine residues that are known to form disulfides in the native toxin, are differentially protected during SPPS. Following synthesis, the fully protected peptide is cleaved from the resin, and disulfides are formed sequentially by selective deprotection of pairs of cysteines and subsequent stepwise oxidation. Global deprotection takes place after all disulfides have been formed, and care must be taken to avoid thiol reshuffling during this step. For peptides containing more than three disulfides, a semi-directed approach can also be employed. Specifically, one disulfide is formed in a directed fashion, the peptide is fully deprotected, and the remaining disulfides are formed by random oxidation. The conformational restriction introduced by the pre-formed disulfide bridge directs the correct configuration of the remaining disulfides. Alternatively, directed folding can be accomplished prior to cleavage of the peptide from the solid support. This approach is attractive from a throughput standpoint because of its applicability to the synthesis of combinatorial peptide libraries, but still faces a number of technical issues. Thus, although some conotoxins were successfully made by on-resin oxidation, their folding yields in solution were better. Finally, directed folding can also be achieved without the need for differential cysteine protection by utilizing selenocysteine. Selenocysteine has a lower redox potential and a lower pKa than cysteine, allowing for the diselenide bridge to be formed selectively, followed by disulfides. In context of the current project, random folding is employed for the synthesis of teretoxins, because their native disulfide configuration is unknown.

Figure 2.6: Kinetically trapped intermediates in in-vitro folding reactions. A. HPLC trace and free energy funnel representation of a folding reaction showing kinetically trapped intermediates. B. HPLC trace and free energy funnel representation of a folding reaction with no kinetically trapped intermediates.
2.7 Structural characterization of Conoidan toxins.

NMR spectroscopy has been the technique of choice for conotoxin characterization. While difficult to crystallize, structural studies of these small and rigid peptides can often be accomplished by homonuclear NMR without the need for isotopic labeling. However, NMR structural assignment still requires a substantial amount of sample, and may be more difficult to implement for the larger toxins of the Terebridae. Owing to the highly disulfide-bonded nature of these molecules, an alternative approach relying on disulfide mapping followed by molecular dynamics (with explicit disulfide bridges) simulation can be envisioned, and is developed in this work. Classic disulfide mapping in proteins is accomplished by protease (usually trypsin) digestion, followed by mass spectrometric analyses of the peptide fragments before and after complete reduction. Unfortunately, this is not very applicable to Conoidan peptides because they lack sufficient protease sites and often contain adjacent cysteine residues. An alternative disulfide mapping technique relies on partial reduction of disulfides followed by differential alkylation of pairs of liberated cysteines. Edman sequencing of the fully reduced and differentially alkylated peptide leads to disulfide assignment. In this work, the partial reduction and differential alkylation method is utilized, but the analysis of the resulting differentially alkylated peptide is accomplished by tandem mass spectrometry by adapting recently described methods, which dramatically reduce sample requirements and costs.

2.8 Current approaches to conoidan toxin functional characterization

Initial biological activity characterization of Conoidan toxins is traditionally accomplished by means of phenotypic assays, such as by intracranial injection of a putative peptide toxin into mice, or typical snail prey organisms such as Caenorhabditis elegans (C. elegans). The phenotypic response observed during this initial screening, is only indicative of general toxin function. Further attribution of peptide toxin activity to a specific neuronal target is accomplished by means of electrophysiological recordings utilizing voltage- or patch-clamp techniques. In this realm, ex-vivo tissue preparations with extensively characterized ion channel and receptor distributions, such as mammalian dorsal root ganglion (DRG) neurons, or HEK cell lines have found numerous uses in the field of conotoxin research.

One extensively used method for electrophysiological characterization of Conoidan toxin bioactivity relies on the use of oocytes of the African frog, Xenopus laevis as a model system. Devoid of
endogenous ion channels, these large progenitor cells are engineered to express a specific ion channel or receptor subtype by injecting them with genetic material (mRNA) coding for this specific target. Heterologous expression of ion channels and receptors of interest in HEK cells have also been shown. Subsequent electrophysiological recordings in the presence and absence of a putative peptide toxin, are used to not only delineate the toxin’s specific neuronal target but also the mode of their interaction (i.e. activation, deactivation etc).

While electrophysiological recordings provide functional information in excess of merely identifying the specific ion channel being targeted, these methods suffer from low throughput. Efforts to increase the throughput of these techniques include automated parallel electrophysiology platforms, the use of fluorescent reporters sensitive to ion concentration, and ion flux assays. As an alternative, binding assays utilizing radioactive or fluorescent tracers can be used to screen for biological activity of conoidan peptides in a high-throughput manner, albeit without functional modality information.

2.9 Objectives of Thesis Project

The Holford laboratory focuses on adapting the techniques used in identification, synthesis, structural and functional characterization of conotoxins to the neuropeptides of the Terebridae, a sister group to the cones, to expand the existing pool of marine peptide natural products. While adapting existing methods to these underexplored animals promises the discovery of peptides with novel neuronal targets and modes of action, the methodology developed also attempts to address challenges of throughput associated with disulfide-rich peptide discovery and characterization as a whole.

As a chemist in the lab, I was tasked with synthesis and structural characterization of terebrid neuropeptides, as well as with chemical aspects of creating a high-throughput microarray based assay to screen teretoxins against their biological targets. Specifically, the research objectives were:

1. Synthesis and oxidative folding of novel terebrid neuropeptides, utilizing modern SPPS and/or recombinant synthesis strategies. Fmoc-SPPS procedures for manual and microwave-assisted synthesis procedures were developed and implemented for the synthesis of teretoxins in Table 2.1. For the peptide Tg77, a fusion-protein recombinant expression system was designed as part of this dissertation project, and was subsequently implemented by Mary E. Wright, a biochemistry graduate student in our lab.
Three of these peptides, Tg03, Tv1, and Arg58 were successfully assembled by Fmoc-SPPS. Random oxidative folding optimization experiments on Tv1 and Arg58 were also carried out. The Fmoc-SPPS, recombinant synthesis, and oxidative folding protocols that were adapted to teretoxins as part of this dissertation project, are in continued use in the Holford laboratory.

<table>
<thead>
<tr>
<th>Table 2.1: Teretoxin neuropeptides identified for synthesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peptide</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>Tg55</td>
</tr>
<tr>
<td>Tg03</td>
</tr>
<tr>
<td>Tg77</td>
</tr>
<tr>
<td>Tv1</td>
</tr>
<tr>
<td>Arg58</td>
</tr>
</tbody>
</table>

2. Structural characterization of synthesized teretoxins by adapting a disulfide mapping technique based on partial reduction, differential alkylation and tandem MS analysis. Optimization of differential alkylation conditions and adaptation of ultra-high pressure liquid chromatography (UHPLC) methods resulted in the successful disulfide mapping of Tv1, a teretoxin shown to induce a phenotypic response upon injection in to polychaete worms. Analysis of MS/MS data, which contributed to the assignment, was done in collaboration with Dr. Vadi Bhatt from Agilent technologies.
Chapter 3: Synthesis, Oxidative Folding of Terebrid Neuropeptides and Structural Characterization of a Novel Venom Peptide Tv1 From the Venomous Marine Snail

_Terebra variaga_
3.1 Terotoxins: from identification to synthesis and characterization

Conoidan peptides are attractive targets for pharmaceutical and neuroscience research applications. In a quest for identifying compounds with novel biological targets, as well as to enrich the peptide marine natural product pipeline in general, the Holford laboratory is focused on the study of the venom peptides of the *Terebridae*, a sister group to the cones, whose toxins are much more well known. This laboratory has been exceedingly successful in identifying putative peptide toxins from terebrid species, in part through implementing a biodiversity-guided rational discovery method, dubbed the "concerted discovery strategy" (Figure3.1). Instead of randomly selecting specimens for mining peptide toxin sequences, the strategy relies on the use of taxonomic tools for phylogenetic analysis, which in turn allows identification of potential toxins from divergent lineages, thus maximizing diversity of lead peptides. Correlating phylogenetic information with the anatomy of the *Terebridae* allowed for identification of several lineages that have lost their venom apparatus, and thus helped refocus toxin discovery efforts only on the clades that retained this anatomical feature. The application of this strategy has resulted in myriad cDNA transcripts of putative toxins identified from terebrid species, several of which were selected for further investigation. Through work with collaborators, who focus on improving *de novo* mass spectrometric sequencing methods of disulfide-rich peptide toxins, terebrid toxins were also identified from crude venom.

While identification of putative toxins is an important step in the development process, chemical synthesis is required for sufficient quantities of the peptide toxins to be available for functional assays. Additionally, structural characterization of biologically active terebrid neuropeptides is needed to delineate the structure-activity relationships with their targets. This part of the dissertation project focuses on chemical synthesis of putative toxins of the *Terebridae* identified by the Holford lab and collaborators, as well as their structural characterization. Numerous studies of disulfide-rich venom peptides of cone snails, serve as a general guide for the workflow of synthesis, oxidative folding and structural characterization of terebrid peptides. While adapting these existing techniques to terotoxins is itself a challenge, improvements in throughput during these stages of the toxin development process are also needed. This bottleneck in throughput in the synthesis and characterization stages is particularly evident in examining
ConoServer data, with 2837 toxin sequences identified and only 175 characterized (ConoServer accessed on 5/16/15).³⁷

**Figure 3.1: Traditional and concerted discovery strategies for teretoxins.**

**A. Traditional strategy.** Analyzed species are chosen randomly, and may correspond to a single lineage. Prospective toxin analysis starts first with characterization of venom components by HPLC (venom fractionation).

**B. Concerted Discovery Strategy (CDS).** Taxonomic tools are used first to identify independent lineages to maximize the species, and thus teretoxin diversity, and then to analyze the numerous cDNA compounds isolated from each analyzed specimen. Application of CDS increases discovery of divergent teretoxins.
SPPS is the conotoxin synthesis strategy of choice, particularly efficient for their short sequences, and amenable to installation of a variety of PTMs encountered in these peptides. The longer peptide toxins of the *Terebridae* push the limits of this strategy, necessitating special improvements to the process. The apparent lack of PTMs in teretoxins, however, makes it possible to synthesize these toxins recombinantly. SPPS protocols adapted to sequence features and lengths of specific teretoxins are developed in this work. Particularly, microwave-assisted automated synthesis is used to combat inefficiency of SPPS of longer peptides, as well as to improve throughput. A recombinant expression strategy for the cost-effective synthesis of longer teretoxins is also designed as part of this project.

Correct formation of intramolecular disulfides in conoidan toxins is key to their activity. Decades of conotoxin research identify a number of oxidative folding protocols, both random and regioselective, for this purpose. Due to the nature of techniques used to identify teretoxin sequences in our laboratory, there is no information about their native disulfide connectivity. Therefore, random oxidative folding procedures are used for the teretoxins synthesized for this thesis work. A variety of folding conditions are evaluated to maximize the yield of the desired bioactive isomer.

NMR is traditionally used for structural characterization of conotoxins. Their small size and highly disulfide-bonded rigid structure make NMR structural assignment of conotoxins possible often even without isotopic labeling. However, NMR analyses still have high sample requirements, and the larger peptides of the *Terebridae* may necessitate costly isotopic enrichment for unambiguous structure determination. Instead, this work focuses on developing a mass spectrometric approach for the disulfide mapping of teretoxins, as means of their structural characterization.

Traditionally, HPLC is used throughout synthesis, oxidative folding, and disulfide mapping of peptide toxins, to ascertain SPPS yields and side products, monitor progress of folding reactions, and separate partially reduced peptide species. In this work, the newly developed ultra high-pressure liquid chromatography (UHPLC)\textsuperscript{72}, is adapted for these purposes. This technique offers greatly shortened analysis time, as well as reduced solvent requirements relative to traditional HPLC. Adapting UHPLC for the development of teretoxins demonstrates another way to improve the throughput of synthesis and characterization stages of peptidic marine natural product development.
3.2 Results and Discussion

Terebrid neuropeptide sequences selected for development are listed in Table 3.1, along with methods chosen for their synthesis, and successfully completed development stages. Fmoc SPPS was chosen for production of peptides Tg55, Tg03, Tv1 and Arg58.

Table 3.1: Teretoxin neuropeptides identified for development

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Sequence</th>
<th>Synthesis</th>
<th>Folding</th>
<th>Activity</th>
<th>Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tg55</td>
<td>TACEQ_HTDCS_AASGP_KYCCQ_DSDCC_GGTEY_CT-NY_GQCLR_SFGR_HNDFIH_LRLSR_GPIO_LTHRM_LKK</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tg03</td>
<td>SECET_DDCC_IDGCC_EDTY_FSESE_INVCS_TCD</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tg77</td>
<td>SLTEK_AGGCPL_YLCSQ_QIFCC_HGRKC_RNVDG_RLLK_CVF_TEASM_LGG</td>
<td>✓</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg58</td>
<td>PDICD_GCPYK_RGGDC_CDLDI_ECC</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tv1</td>
<td>TR(VL)CC_CCWYN_GSKDV_CSQSC_C21</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

For peptides made by SPPS, double-coupled residues are in bold. Tg55 sequence highlighted in red was to be assembled as a peptide thioester for subsequent NCL. Underlined sequence DG was introduced as a dipeptide.

In straight chain synthesis, peptides are assembled from the carboxy (C) to amino (N) terminus, by successive couplings of activated esters of Nα-Fmoc protected amino acids onto a solid support (Figure 3.2). Fmoc amino acids are routinely activated by O-Benzotriazole-\(N,N,N',N'\)-tetramethyluronium hexafluorophosphate (HBTU) in the presence of diisopropylethylamine (DIPEA) for coupling, and deprotected with 20% piperidine in DMF. Reactive side chain functionalities of the Fmoc amino acid building blocks are protected with standard acid-labile groups. When the target sequence is reached, the peptide is fully deprotected and cleaved from resin by a cocktail of trifluoroacetic acid (TFA) and scavengers. Manual SPPS was initially used in attempts to synthesize Tg55, Tg03, as well as to establish general protocols for syntheses of terebrid neuropeptides. Using the manual SPPS methods as a guide, syntheses of Tv1 and Arg58 were completed by an automated microwave peptide synthesizer. During manual synthesis, the ninhydrin test for free amines is used to assess the efficiencies of coupling and deprotection reactions, while in automated synthesis, the waste stream from the deprotection reaction is analyzed spectrophotometrically for the UV-absorbing Fmoc group.
Figure 3.2: Straight chain Fmoc SPPS diagram
Information from these assessments is then used for synthesis optimization by using a longer coupling time, more effective coupling reagent, or double-coupling (repeating a coupling reaction) a specific amino acid.

Attempts at synthesizing Tg55 by manual Fmoc SPPS were severely hampered by decreasing coupling efficiency with increasing peptide length. However, its C-terminal fragment Tg55(31-62), bearing an N-terminal cysteine was successfully assembled, suggesting a possible strategy of assembling Tg55 from two SPPS prepared halves joined through NCL (Figure 3.3).

Figure 3.3: Tg55 convergent synthesis by SPPS and NCL

In this approach, a peptide thioster fragment is joined to another fragment containing an N-terminal cysteine residue to yield a native peptide bond through an S to N acyl shift. The thioester fragment,
Tg55(1-30), was to be assembled on a sulfamylbutyryl “safety catch” resin\(^7\), which allows the generation of peptide thioesters on the solid phase, via activation of the linker with trimethylsilyldiazomethane (TMS-\(\text{CH}_2\text{N}_2\)) after standard SPPS, and cleavage of the peptide by nucleophilic displacement with a thiol. Both fully deprotected Tg55 fragments are then joined by NCL in the presence of thiols under mildly denaturing aqueous conditions. However, repeated attempts to generate the peptide thioester Tg55(1-30) were unsuccessful, due to inefficiency of loading the sulfamylbutyryl “safety catch” resin. The difficulty of loading this resin, as well as cleavage of thioesters from it is well documented\(^7\), and despite using the chaotropic salt LiBr, which was proven to enhance resin swelling and aid in loading reactions onto the sulfamylbutyryl linker\(^7\), no thioester cleavage products were observed during repeated test cleavages of Tg55 peptide fragments from this resin. Due to these difficulties, its synthesis was not pursued further.

Manual synthesis of the peptide Tg03 successfully resulted in a product corresponding to the calculated mass of the peptide. However, this peptide’s extremely low isoelectric point, resultant from 11 acidic residues within its sequence, made it difficult to handle, particularly during HPLC purification. Because of general problems of solubility of this peptide, and its incompatibility with HPLC buffers, its oxidative folding and structural characterization was not pursued.

Both Tv1 and Arg58 were successfully synthesized by automated microwave-assisted Fmoc SPPS. Due to inherent uncertainty in the MS derived *de novo* sequence, both Leu3 and Ile3 homologs of Tv1 were assembled for subsequent oxidative folding and structural characterization.

A recombinant synthesis strategy, including choice of vector, solubilizing protein fusion tag, host organism allowing for disulfide bond formation, and protease cleavage site to afford the native peptide without exogenous sequences was designed (Figure 3.4). In addition to being optimized for efficient production of disulfide-rich peptides, this strategy also uses ligation-independent cloning (LIC)\(^7\), a time and cost saving measure. LIC eliminates the need of treating an expression vector and insert with restriction endonucleases, dephosphorylating the vector, using DNA ligase, and all of the associated DNA purification steps. Instead, LIC uses the proofreading function of DNA polymerase to create long (10-12bp) complimentary overhangs during PCR in the absence of a single deoxyribonucleotide phosphate. The insert is simply added to the linearized vector with complimentary 10-12bp overhangs, and this mixture is used directly to transform competent host cells, without the need for DNA ligation. The
implementation of this recombinant synthesis strategy by a biochemistry graduate student, Mary E. Wright, resulted in successful synthesis of Tg77.

As teretoxins contain multiple cysteine residues, their oxidative folding into the biologically active conformation may at times be challenging. Synthetic peptide toxins are folded in solution by employing various oxidizing agents. However, because there are multiple arrangements of disulfide bonds, the formation of biologically inactive, kinetically trapped folding intermediates is possible. To avoid this, many
folding reaction conditions must be evaluated. In the folding of Tv1 and Arg58, dimethylsulfoxide (DMSO), mixtures of oxidized and reduced glutathione (GSH:GSSG), and ClearOX (immobilized Ellman's reagent) were all evaluated as oxidizing agents. The best results were obtained with GSH:GSSG mixtures, and the folding yields of Tv1 were further optimized for maximal yield of the active folded product by varying the reaction time, temperature, pH, and salt concentration. Arg58 folding was not optimized further because no functional activity was confirmed for any of its folded products. Functional testing of Arg58 is ongoing in the Holford laboratory.

Following confirmation of its biological activity in a phenotypic assay in polychaete worms, disulfide mapping of Tv1 was accomplished by a strategy of partial reduction and differential alkylation coupled with MS/MS analyses (Figure 3.5). A similar strategy, utilizing UHPLC and nanoflow LC-MS was successfully used for disulfide mapping using just picomols of sample. This method is well suited for teretoxins as many of them have an inadequate number of protease cleavage sights, barring the traditional protease digest and MS approach. In this strategy the folded neuropeptide is first partially reduced at low pH using TCEP. The partially reduced species are then separated by RP-HPLC, and immediately alkylated with n-ethylmaleimide (NEM) at low pH. Excess NEM is removed by desalting. Keeping the pH low during these steps is crucial in preventing disulfide scrambling. The remaining disulfides of each partially reduced conformation are then reduced completely and either alkylated with Iodoacetamide (IAM) or left as free sulfhydrils. Following a final desalting, the partially reduced and differentially alkylated species are subjected to MS/MS. Upon entering the collision chamber of the mass spectrometer, the peptides break along the backbone producing characteristic N-terminal b, and C-terminal y ions. Analysis of b and y ions of each partially reduced and alkylated species, yields information on one disulfide bond connectivity. Thus, for the disulfide mapping of a peptide containing three disulfide bonds, it is necessary to analyze at least two non-redundant partially reduced and alkylated species to assign the framework. All six possible partially reduced Tv1 species were successfully resolved, differentially alkylated, and provided non-conflicting information on its disulfide connectivity. This disulfide assignment was further confirmed by solution $^1$H NMR of Tv1.
3.2.1 Synthesis of selected terebrid neuropeptides

Manual solid phase peptide synthesis of Tg55 proceed relatively smoothly through Phe_{42}, although bulky amino acids had to be double-coupled (double-coupled residues are shown in bold in Table 3.1). After Phe_{42}, decreased coupling and deprotection efficiency, as ascertained by the ninhydrin test, necessitated double-coupling of every subsequent residue, and capping unreacted resin bound
deletion products with acetic anhydride. Despite considerable efforts, deletion products and truncated peptides dominate the HPLC spectrum of Tg55(32-62), with a marked increase in these unwanted species observed relative to the fragment Tg55(47-62) (Figure 3.6). Although the presence of Tg55(32-62) among the many cleavage products was confirmed by MS, its purification was not immediately attempted because easier separation was envisioned after NCL with Tg55(1-31) thioester. Unfortunately, repeated attempts to synthesize the Tg55(1-31) thioester fragment were unsuccessful, and SPPS synthesis of Tg55 was not pursued further.

Tg03 was successfully assembled by manual SPPS, as confirmed by mass spectrometry (Data not shown). However, its low isoelectric point conferred very low solubility in acidic HPLC buffers. Attempts to use basic HPLC buffers were complicated by decreased column performance at high pH. Due to these solubility and handling problems, the development of Tg03 was not pursued further.

Arg58 was successfully synthesized by automated microwave-assisted SPPS. Previous attempts at manual synthesis of this peptide have identified sequence regions that suffer from low coupling efficiency. Utilizing this information, selected residues were programmed to be double-coupled during automated SPPS. To further minimize unwanted side products, the aspartimide formation prone sequence Asp-Gly (underlined in Table 3.1) was introduced as a backbone dimethoxybenzyl (dmb)
protected dipeptide and double-coupled. Figure 3.7 shows the UHPLC traces of purified Arg58, and its mass spectrum.

![UHPLC and MS of synthesized Arg58](image)

**Figure 3.7: UHPLC and MS of synthesized Arg58.** A. UHPLC trace showing the desired peptide denoted with an asterisk. B. MALDI-TOF spectrum of synthesized Arg58 showing [M+H] 2494.46, calculated mass is 2493.85

Both Leu$_3$ and Ile$_3$ variants of Tv1 were successfully assembled by automated microwave-assisted SPPS. Double coupling of the first three residues was used to improve synthetic efficiency, using information from prior synthesis attempts. The HPLC and MS of Tv1 are shown in Figure 3.8. Furthermore, the MS/MS spectrum of synthetic Tv1 is practically identical to that of native Tv1 (Figure 3.9), confirming the validity of its *de novo* sequence assignment. Both Leu$_3$ and Ile$_3$ were used in subsequent folding, functional and structural experiments, but behaved nearly identically. For convenience, only Tv1(Ile$_3$) data is shown in subsequent sections.
Figure 3.8: UHPLC and MS of synthesized Tv1: A. UHPLC trace of purified syntheticTv1. B. MAL-DI-TOF spectrum of synthesized Tv1 showing [M+H] 2316.32, calculated mass is 2316.87
The peptide Tg77 was successfully produced recombinantly, following a strategy developed as part of this dissertation project. Figure 3.10 shows the production of Tg77-GST fusion protein during expression in *E. coli* Origami B, and Figure 3.11 shows the cleavage of mature Tg77 by enterokinase. After optimization of enterokinase cleavage conditions, a yield of 2.5mg Tg77/L of culture was obtained. MS analysis of recombinant Tg77 shows that the peptide has been produced in a folded state, consistent with a reduction of mass associated with a loss of 6 protons during the formation of 3 disulfides.

Figure 3.9: ETD spectra of native (black) and synthetic Tv1 (blue). MS/MS spectrum recorded on a (M + 6H)\(^6\) ion after conversion of cysteine residues to dimethyl lysine analogs. The sequence is given above the spectrum and observed c and z-type fragment ions are indicated in the sequence. The spectrum was recorded at a resolution of 7500 at m/z 400 and all fragment ions have a mass accuracy of better than 5 ppm.
Figure 3.10: SDS-PAGE analysis of expression and purification of Tg77 fusion protein. M - protein molecular weight marker. Lane 1 - Tg77 crude lysate. Lane 2 - Tg77 crude lysate after binding to nickel-NTA resin. Lanes 3, 4 - washing steps. Lane 5 - purified Tg77 fusion protein. Figure reproduced from reference 77.

Figure 3.11: HPLC monitoring of enterokinase cleavage and LC-MS of recombinantly produced Tg77. A. Semi-preparative HPLC of folded Tg77. B. LC-MS characterization of folded Tg77 (expected mass: 4756.2 Da, Observed: 4756.2 Da). +4, +5, +6, +7 charge states are shown. Figure reproduced from reference 77.
3.2.2 Oxidative folding of terebrid neuropeptides Tv1 and Arg58

Both Tv1 and Arg58 were folded using 10:1 GSH:GSSG as the oxidizing agent. UHPLC monitoring of folding reactions allowed for rapid screening of a variety of folding conditions. UHPLC monitoring of Tv1 over time, shows production of predominantly a single isomer within 2 hours (Figure 3.12). The effect of temperature, pH, and salt concentration on folding yields of this isomer of Tv1, measured as %area of this peak (UV detection of HPLC folded products at 229nm), were investigated (Figure 3.13). The optimal conditions for Tv1 folding were after incubation for 4 hours at 4°C (20µM Tv1, 1.0M NaCl, 0.1M Tris, 0.1mM ethylenediaminetetraacetic acid (EDTA), 10mM GSH, 1mM GSSG), with over 90% yield of the desired isomer. Figure 3.14 shows the UHPLC trace, and MS confirmation of the biologically active folding product of Tv1 purified to 99%.

Figure 3.12: UHPLC monitoring of oxidative folding of Tv1 over time. 20µM Tv1 in 0.1M Tris pH 7.4, 0.1mM EDTA, 1M NaCl, 10mM GSH and 1mM GSSG. UHPLC UV detection at 229nm.
Figure 3.13: Optimization of oxidative folding of Tv1. Y-axis represents the yield of biologically active folding product of Tv1 as ascertained by UHPLC with UV detection at 229nm. 20µM Tv1, 0.1M Tris, 0.1mM EDTA, 10mM GSH and 1mM GSSG. +salt indicates addition of 1M NaCl.
Arg 58 folding resulted in approximately 60% yield of one isomer (Figure 3.15). Biological activity testing of this isomer, as well as four other isolated minor folding products of Arg58 did not indicate activity, presumably due to limitations of the assays employed. Additional functional testing of folding products of Arg58 using alternative methodologies is ongoing, and further folding optimization and structural characterization of this peptide is reserved until confirmation of biological activity.

Figure 3.14: UHPLC and MALDI-TOF analysis of purified folded Tv1. A. UHPLC trace of folded Tv1 purified to 99%. B. MALDI-TOF spectrum of folded Tv1 showing [M+H] 2310.74 Da, expected [M+H] 2310.82.
3.2.3 Structural characterization of Tv1: disulfide mapping by partial reduction, differential alkylation and MS/MS analysis

The major folding product of Tv1 obtained using optimized conditions, isolated and purified to 98% using semi-preparative HPLC, was subjected to partial reduction by tris-carboxyethylphosphine (TCEP) in citrate buffer pH 3.0. The reaction time and TCEP concentration was evaluated by direct injection of reaction contents onto UHPLC column, and analyzing partially reduced products. All six possible partially reduced species were observed upon optimization of reaction conditions (Figure 3.16). The UHPLC method was then transferred to an HPLC platform to allow for manual collection of partially reduced Tv1 products. Upon HPLC separation, the six partially reduced species were collected directly into tubes containing an NEM alkylating solution (pH 3.0). Following alkylation, excess NEM reagent was removed

Figure 3.15: UHPLC and MALDI-TOF analysis of folded Arg58. A. UHPLC trace of Arg58 folding products. Major product is denoted with an asterisk. B. MALDI-TOF spectrum of Arg58 major folding product showing [M+H] 2489.54 Da, expected [M+H] 2487.80 Da.
by a UHPLC desalting protocol. The desalted, NEM alkylated species were then reduced completely by incubation with 100mM TCEP, and alkylated with IAM. The resulting species, labeled with two, four, or six NEM and/or IAM groups were once again desalted using UHPLC, lyophilized, and subjected to LC-MS/MS to determine NEM and IAM alkylation sites by matching the MS/MS data to b- and y- series ions to theoretical patterns. These analyses of all six differentially alkylated species provided complimentary and non-contradictory information, resulting in the connectivity assignment: Cys4-Cys20, Cys5-Cys21, and Cys7-Cys16 (Table 3.2).

**Table 3.2. Predicted and observed b and y ions of differentially alkylated peptides by auto and targeted MS/MS analysis**

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Ions</th>
<th>Unmodified peptide (m/z)</th>
<th>m/z (thr.)</th>
<th>m/z (obs.)</th>
<th>Difference</th>
<th>Modifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRIC</td>
<td>b4</td>
<td>474.2493</td>
<td>599.2902</td>
<td>599.297</td>
<td>125.0414</td>
<td>NEM</td>
</tr>
<tr>
<td>TRICC</td>
<td>b5</td>
<td>577.2585</td>
<td>759.3197</td>
<td>759.3276</td>
<td>125.0414+57.0198</td>
<td>NEM+IAM</td>
</tr>
<tr>
<td>C</td>
<td>y1</td>
<td>122.0270</td>
<td>179.048</td>
<td>179.0485</td>
<td>57.021</td>
<td>IAM</td>
</tr>
<tr>
<td>CC</td>
<td>y2</td>
<td>225.0362</td>
<td>407.1048</td>
<td>407.1054</td>
<td>57.021+125.0476</td>
<td>IAM+NEM</td>
</tr>
<tr>
<td>CCSQSC</td>
<td>y6</td>
<td>630.1680</td>
<td>869.2579</td>
<td>869.2586</td>
<td>57.021+125.0476+125.0479</td>
<td>IAM+2NEM</td>
</tr>
<tr>
<td>TRICCG</td>
<td>b6</td>
<td>634.2800</td>
<td>884.3632</td>
<td>884.3753</td>
<td>250.0832</td>
<td>2NEM</td>
</tr>
<tr>
<td>TRICCGC</td>
<td>b7</td>
<td>737.2891</td>
<td>1044.4083</td>
<td>1044.406</td>
<td>307.1192(250.0831+57.036)</td>
<td>2NEM+IAM</td>
</tr>
<tr>
<td>C</td>
<td>y1</td>
<td>122.0270</td>
<td>247.0747</td>
<td>247.0747</td>
<td>125.0477</td>
<td>NEM</td>
</tr>
<tr>
<td>CC</td>
<td>y2</td>
<td>225.0362</td>
<td>475.1292</td>
<td>475.1316</td>
<td>250.093</td>
<td>2NEM</td>
</tr>
<tr>
<td>CCSQSC</td>
<td>y6</td>
<td>630.1680</td>
<td>937.2707</td>
<td>937.2849</td>
<td>57.0188+250.0859</td>
<td>IAM+2NEM</td>
</tr>
</tbody>
</table>

MS/MS for precursor ion m/z 932.026 (+3) of parent ion 2793.0552

MS/MS for precursor ion m/z 977.376 (+3) of parent ion 2929.1027
This connectivity was also confirmed by an NMR solution structure of Tv1, obtained by collaborators Mohammed Bhuiyan and Sébastien Poget (Figure 3.17). Despite having a cysteine scaffold similar to M-superfamily conotoxins, the disulfide connectivity of Tv1 has not been previously observed in Conus Peptides.\[^{79}\]
Figure 3.17: NMR confirmation of Tv1 disulfide assignment. A. NOE contacts confirming the C7-C16 disulfide bond. An overlay of the HNHa fingerprint region of the NOESY (in black) and TOCSY (in blue) spectra shows NOE crosspeaks linking Cys 16 and Cys 7 as well as contacts in the residues flanking the C7-C16 disulfide bond (C7-S17, Y8-C16, G6-C16, G6-N18). B. An overlay of a cartoon representation and a stick model of the lowest-energy structure of Tv1 shows the β-sheet character of the peptide and reveals the important role of the Tyr 8 side chain in the formation of a small hydrophobic core. Figure reproduced from reference 9.
3.3 Conclusions and future directions

Adapting techniques from conotoxin research and coupling them with modern techniques to increase throughput successfully established the synthesis and structural characterization capabilities of the Holford laboratory. Difficulties during early manual synthesis attempts of terebrid peptides lead to the development of improved protocols. Automated microwave-assisted peptide synthesis allowed for faster assembly of peptides, and gave access to SPPS of longer teretoxins. Use of UHPLC for analyzing SPPS cleavage products, analytical confirmation of peptide purity, and monitoring oxidative folding reactions dramatically reduced the time and cost of peptide synthesis, and oxidative folding. This success is demonstrated with the synthesis and structural characterization of Tv1, the first terebrid peptide toxin to be structurally characterized. Demonstrated to be active in a phenotypic assay in polychaete worms, the cysteine connectivity of Tv1 is unlike that of similar M-superfamily conotoxins (figure 3.18), suggesting that it may have a novel biological target or mode of action. Efforts to identify its biological target using electrophysiological recordings, and cell-based assays are ongoing.

![Figure 3.18: Structural comparison of Tv1 teretoxin and M-superfamily conotoxins.](image)

Comparison of the NMR structure of Tv1 with that of M superfamily conotoxins SmIIIA, mr3e, and KIIIA reveals significant structural differences between Tv1 and these conotoxins despite all having the same cysteine scaffold CC-C-C-CC. All structures are shown in cartoon representation with disulfide bonds highlighted in yellow. All figures were prepared using PyMol (www.pymol.org). Conotoxin structural references are as follows: MrIIIe79d SmIIIA79g and KIIIA79h.

The recombinant expression strategy developed for Tg77, was also proven effective, particularly for longer teretoxins. Besides being a viable route for teretoxin synthesis, it also incorporates an improvement in throughput by using ligation independent cloning. A recombinant expression system analogous to that designed for Tg77, was also used to successfully produce two more toxin peptides from *Cingulaterebra Anilis* Can1, by graduate student Mary E. Wright, illustrating the general applicability of this method to the synthesis of teretoxins.
3.4 Materials and methods

3.4.1 SPPS of terebrid neuropeptides

All peptide acids were synthesized on Wang-Cl resin. This resin was prepared from commercially available Wang resin (Anaspec) using a previously published procedure. Racemization free loading of the first amino acid was also completed as previously described. After conjugation of the first (C-terminal) Nα-Fmoc-amino acid (Fmoc-AA-OH), resin substitution was calculated using a spectrophotometric Fmoc release assay. Briefly, 10-20mg resin aliquots were weighed out, and deprotected with 1mL of 20% piperidine in (DMF) for 1 hour. An aliquot of the deprotection solution was then diluted 100 fold, and its 1mm absorbance at 301nm was used to calculate resin loading according to:

\[ S(\text{mmolAA/g resin}) = \frac{101 \times A_{301}}{7.4 \times (\text{mg resin})} \]

During Fmoc-SPPS, standard side chain protecting groups were used for all amino acids. Due to absence of a priori knowledge of disulfide connectivity, the trityl protecting group was used for all cysteine side chains. Chain elongation steps, namely coupling Fmoc-AA-OH, and removal of the Fmoc protecting group differed for manual and automated microwave-assisted synthesis. During manual synthesis, coupling was effected by adding 3eq. of the Fmoc-AA-OH, 2.99eq. HBTU, and 6eq DIPEA, all dissolved in a minimum amount of DMF, and incubating for 1 hour with stirring. Up to 50% dichloromethane was added to improve resin swelling properties during difficult coupling reactions. If repeated couplings failed to completely react, resin-bound peptides were capped with 10eq Acetic anhydride and 11eq DIPEA in DMF to produce Nα-Acetyl capped products in lieu of deletion products which are harder to separate from the desired peptide upon cleavage. Deprotection was effected by exposing the peptidyl resin to 5 volumes of 20% piperidine in DMF for 5 minutes, then 30 minutes. DMF solutions containing 2% 1,8-Diazabicyclo[5.4.0]undec-7-ene (DBU) and 18% piperidine were used as alternatives for difficult deprotection reactions.

The ninhydrin test for free amines was used to evaluate the efficiency of coupling and deprotection reactions. Briefly, a small aliquot of peptidyl resin (10-30 single resin beads) was withdrawn from a peptide synthesis vessel, washed, dried under vacuum, and transferred to a 100mm culture tube. One drop each of reagent A (5% w/v ninhydrin in ethanol), B (4:1 phenol:ethanol), and C (2% v/v of 1mM aqueous KCN in pyridine) were added to the resin, and the tube was heated to 110°C for 3 minutes. Blue
coloration of resin beads indicates the presence of free amines. Couplings reactions were considered complete when resin beads remained colorless after this procedure.

Microwave-assisted, automated peptide synthesis was completed on a Liberty1 instrument (CEM corporation). Coupling reactions used 5eq. Fmoc-AA-OH (as 0.5M solution in DMF), 4.99eq. HBTU and 10eq. of DIPEA (in N-methyl-2-pyrrolidone), and proceeded for 5min, under up to 25W microwave power, while maintaining a reaction vessel temperature of no greater than 75°C. Coupling of Fmoc-Cys(trt)-OH and Fmoc-His(trt)-OH, was restricted to no greater than 50°C, to prevent racemization and thermal cleavage of the trityl group. Fmoc deprotection was effected by 20% piperidine in DMF containing 0.1M 1-hydroxybenzotriazole (HOBT), for 1 min, then 4 min, at 25W maximum power and 75°C maximum temperature. Residues were preprogrammed to be double-coupled based on information from previous manual synthesis attempts, as well as in peptide regions containing several bulky aliphatic amino acids in succession. During the synthesis of Arg58, the aspartimide prone sequence Asp-Gly, was introduced as a dmb-protected dipeptide, Fmoc-Asp(OtBu)-(Dmb)Gly-OH (Novabiochem), and double coupled.

Cleavage of peptides from the resin, and removal of all side chain protecting groups was accomplished by shaking the peptidyl resin with 5 volume equivalents of reagent K (92.5% trifluoroacetic acid (TFA), 2.5% triisopropylsilane (TIS), 2.5% 1,2-ethanedithiol (EDT) and 2.5% water) for 4 hrs. Reagent K supernatant containing cleaved products was then filtered from the resin, and peptide components were precipitated by addition of 50 volume eq. of cold methyl-tert-butyl-ether.

Peptide cleavage products were then dissolved in HPLC buffer A (0.1% TFA), and assayed using a generic reversed phase UHPLC elution protocol. Briefly peptide cleavage products were loaded onto an Acquity UPLC BEH300 C18 column (4.6x50mm, 1.7um particle size, Waters corporation, MA), and elution was carried out at 0.6mL/min with 97% HPLC buffer A and 3% HPLC buffer B (80% acetonitrile (ACN), 0.1% TFA) for the first 20 seconds, then increasing HPLC buffer B to 75% in a linear gradient over 2 minutes. This initial screen was used for gradient optimization for each peptide synthesized. Once maximal resolution of peptide product from truncated and deletion sequences was attained, the UHPLC method was scaled to semi-preparative HPLC, taking into account column dimensions and particle size to calculate appropriate flow rates and duration of gradient steps. Semi-preparative purification of synthesized Tv1 was completed on an X-Bridge column (10x150mm, 5um particle size, Waters
Corporation, MA) by elution at 5mL/min of 20% HPLC buffer B and 80% HPLC buffer A for the first 5 minutes, then increasing HPLC buffer B to 35% in a linear gradient over 45min.

The identity of synthesized peptides was confirmed by molecular mass measurement using a MALDI-TOF (Waters Micromas, MA) instrument, and α-Cyano-4-hydroxycinnamic acid (CHαCN) as matrix.

3.4.2 Oxidative Folding of Teretoxin peptides

Early experiments on several teretoxin peptides tested dimethylsulfoxide (DMSO) and ClearOX resin as potential oxidizing agents in folding reactions. The yields of these reactions were substantially lower than with mixtures of oxidized and reduced glutathione (data not shown). For routine teretoxin folding, the starting reaction conditions are 20µM of linear purified peptide incubated in 0.1M Tris-HCl pH 7.5, 0.1mM EDTA, 10mM GSH and 1mM GSSG. Aliquots of this initial folding reaction are then stopped after 15 min, 30 min, 1, 2, 3, 4, and 24 hours by acidification with formic acid (to 8% v/v), and the products analyzed by UHPLC using the same column and buffers as described earlier, but with an optimized gradient for each peptide. Further folding optimization is accomplished by varying the reaction temperature, pH, and including up to 1M NaCl in the folding reaction. For Tv1, optimized folding conditions were: 20µM Tv1 incubated in 0.1M Tris pH 10, 1M NaCl, 0.1mM EDTA, 10mM GSH and 1mM GSSG for 4 hours. The identity of oxidized teretoxins is confirmed by MALDI-TOF. For Tv1, the optimized folding reaction was scaled up, and the product purified on an X-Bridge semi-preparative column (See above) by eluting at 5mL/min with 15% buffer B and 85% buffer A for the first 5 min, then increasing buffer B to 35% in a linear gradient over 45 min.

3.4.3 Disulfide mapping of Tv1

Purified folded Tv1 was subjected to partial reduction and alkylation procedure by modifying previous protocols. Briefly, 9µL of 1mM Tv1 dissolved in 0.1M citrate pH 3.0 was mixed with 1µL 20mM TCEP, and incubated in an HPLC microvial for 90min at 4°C. Reduction was stopped by direct injection onto an RP-HPLC X-bride column (4.6x150, 5um particle size, Waters Corporation, MA) and eluted with 20% buffer B in buffer A for 5 min, then increasing buffer B to 35% in a linear gradient over 75 min. Six partially reduced species (Figure 3.16) were then collected directly into tubes containing 100uL of 200mM NEM in citrate pH 3.0, and incubated for 2h at 37°C. Excess NEM was then removed by desalting each of the six fractions using UHPLC. Desalted fractions were then completely reduced by
incubation with 100mM TCEP at 55°C for 1hr, and subsequently alkylated by addition of 275µL of 200mM IAM, and incubating for 2hrs at 37°C in the dark. The fractions were once again desalted using UHPLC, lyophilized, and suspended in 0.5% formic acid for LC-MS/MS analyses. Using a 7 minute gradient, each fraction was eluted through a 43mm HPLC-Chip/Q-TOF (Agilent Technologies), to allow complete resolution of fully labeled peptides from those with incomplete alkylations. Data, collected in targeted and auto MS/MS mode, was processed using molecular feature extraction software and matched to theoretical sequences using MassHunter Bioconfirm Qual B.05 software (Agilent Technologies).
Chapter 4 - Fluorescence Spectroscopy Using Solvatochromic Probes for Investigation of Protein Structure, Function and Interactions
4.1 Fluorescence spectroscopy in the study of biological systems

Fluorescence spectroscopy is a powerful tool for studying structure, dynamics and functional interactions of biomacromolecules. Unlike high-resolution structural methods, such as X-ray crystallography and NMR, which provide a global picture of protein structure, fluorescence spectroscopy utilizes a molecular probe to provide information about its surrounding environment. This probe, or fluorophore, undergoes electronic excitation upon light absorption, and emits light of a characteristic wavelength when it relaxes from its first singlet excited state to the ground state. The interaction of the singlet excited state of a fluorescence probe with the local chemical environment can promote proton transfer, form charge transfer complexes, cause exciplex formation, result in intersystem crossing to the triplet state, promote resonance energy transfer, or cause reorientation of solvent and neighboring group dipoles. All of these events influence measurable fluorescence parameters including emission and excitation wavelengths, quantum yield, anisotropy, and fluorescence lifetime, thus providing invaluable structural and dynamic information about the probe’s environment. In the context of biological sciences, fluorescence measurements can provide a wealth of information on various molecular processes, like the interactions of the probe and the surrounding solvent, rotational diffusion of macromolecules such as proteins, distances between macromolecular regions, conformational changes, and binding interactions.

4.2 Protein fluorescence and the intrinsic fluorophore tryptophan

Using fluorescence spectroscopy to study proteins can be accomplished by utilizing fluorophores that are already present in these biological macromolecules. These endogenous probes allow for investigation of proteins without chemical perturbation caused by introduction of extrinsic probes. Of the twenty encoded proteogenic amino acids, three have aromatic side chains that exhibit fluorescence, and are considered intrinsic fluorophores. These are the side chains of phenylalanine (benzene), tyrosine (phenol) and tryptophan (indole). Of these three, tryptophan (Trp, W) is most widely employed for the study of protein structure, dynamics and interactions. Trp occurs only at about 1 mole% of all proteins, more rarely than any other encoded proteogenic amino acid. This rarity facilitates interpretation of its spectral data. Trp also absorbs light at lower energies than other aromatic amino acids, allowing for its excitation independent of tyrosine (Tyr, Y) and phenylalanine (Phe, F). The unique photophysical properties of Trp, such as sensitivity of its emission to the polarity of the local environment
(solvatochromism), propensity for collisional fluorescence quenching, and high anisotropy have contributed to its use as a probe in innumerable investigations of protein structure and function. Additionally, Trp's multiexponential intensity decay kinetics have been linked to the local secondary structure of the polypeptide chain allowing for gleaning structural information about polypeptides in dilute solutions via fluorescence lifetime measurements.

In spite of these useful spectral characteristics, Trp is far from being a perfect probe in context of protein fluorescence spectroscopy. For example, despite its rare occurrence, most proteins contain more than one Trp, inhibiting structural interpretation of fluorescence data due to inability to attribute spectral signals to a specific residue. Although fluorescence lifetime decay analysis can be used to distinguish between multiple fluorophores in distinct environments, most proteins containing just a single Trp display multi-exponential decay kinetics, complicating the application of this approach. A further confounding factor is the presence of two nearly isoenergetic excited states resulting from nearly perpendicular electronic absorption transitions, $^1L_a$ and $^1L_b$, in Trp (Figure 4.1).

![Figure 4.1: Electronic absorption transition moments of indole](image)

While $^1L_a$ is sensitive to the polarity of the environment, $^1L_b$ is not, and emission can occur from either state depending on the polarity of the surrounding cavity. However, it was shown that in all but most extreme cases Trp emission in proteins occurs from $^1L_a$. Additionally, both specific (H-Bonding) and general solvent effects influence Trp emission, further complicating meaningful structural attribution of spectral data. Studies of the interaction of Trp-containing proteins and nucleic acids also present a
problem because nucleotides have a strong extinction coefficient in the region of tryptophan absorption and thus contribute to a strong inner filter effect, that can be easily misinterpreted as quenching resulting from intermolecular interactions. In summary, while the complexity of indole photophysics has prompted detailed inquires into protein fluorescence using Trp as a probe, including studies of Trp quantum yield as a function of local environment conducted in our laboratory, it has also limited meaningful structural attribution of spectral data.

4.3 Tryptophan analogues for structure-conservative spectral enhancement of proteins

Expanding the spectroscopic utility of Trp in protein fluorescence has been widely pursued. While Trp naturally occurs in proteins, some proteins lack a native Trp, or have more than one Trp. Thus, site-directed mutagenesis can often be used to either introduce Trp at a desired site in a protein, or replace an unwanted Trp by a residue with non-interfering spectral characteristics. However, such strategies can often alter the structure/function of some proteins, and also do not address the shortcomings of Trp photophysics, namely UV excitation, and complexity of solvent effects on its fluorescence.

One interesting approach of addressing the limitations of Trp as an intrinsic probe came from investigations aimed at elucidating the pathway of Trp biosynthesis and its insertion into proteins during translation. These studies showed that some nearly isosteric analogues of Trp can be biosynthetically inserted into proteins by endogenous cellular machinery, with tryptophanyl tRNA-synthetase found to be the key enzyme displaying substrate flexibility. Consequently, replacement of Trps with non-natural tryptophan analogs with enhanced photophysical properties to address the issues of spectral overlap while minimizing the effect on structure has been widely pursued. Since then, a number of Trp analogs, each with distinct spectral advantages, were incorporated to create spectrally enhanced alloproteins (Figure 4.2).

These analogs have been incorporated into a protein or peptide in a variety of ways: biosynthetically via recombinant expression in a Trp auxotroph in the presence of the analog, by in vitro transcription/translation using a nonsense suppressor tRNA chemically acylated with the analog, or by direct introduction of the analog during solid phase peptide synthesis. Although the structural perturbations introduced by analog incorporation can vary depending on the specific protein and analog, the majority of investigations have showed them to be minimal.
As summarized by Ross et al.\textsuperscript{87e}, spectral enhancements introduced by analog incorporations vary. While using fluorinated Trps allows for the possibility of studying proteins by \textsuperscript{19}F-NMR, 4-fluorotryptophan, 5-hydroxytryptophan (5HTP) and 7-azatryptophan all have desirable properties for fluorescence enhancement (Figure 4.2 B,C).

4-fluorotryptophan is not fluorescent and is used to silence the signals from unwanted tryptophans with minimal structure perturbation. 5HTP and 7-aza Trp exhibit red shifted absorption and can be singled out in the presence of Trp using Red-edge excitation. However, significant absorption envelope overlap with Trp still exists, barring fluorometric measurements at the absorption maxima of these chromophores, resulting in decreased sensitivity when using these analogs. Also, the fluorescence

---

**Figure 4.2: Structures and spectral characteristics of Trp and its analogues.** A. structures of tryptophan (1) 5-fluorotryptophan (2), 4-fluorotryptophan (3), 5-hydroxytryptophan (4), 7-azatryptophan (5), their absorption (B), and fluorescence emission (C) spectra. Spectra of non-fluorescent 3 are omitted. Spectra reproduced from reference 87e.
emission spectra of 7-aza and 5-OH Trps significantly overlap that of Trp allowing for the possibility of misattributing the fluorescence signal (Figure 4.2C). More recently, 4-azatryptophan was shown to be the most red-shifted\textsuperscript{90} of the isosteric tryptophan analogs described above. Having the largest Stokes shift of all azatryptophans studied, this analog has a fluorescence emission maximum of 418nm, but still suffers from a low quantum yield (0.077 as determined for 4-aza Indole)\textsuperscript{90}. Furthermore, the absorption profile of 4-aza Trp completely overlaps that of Trp, making it impossible to selectively excite this analog in the presence of Trp.

Because many Trp analogs are isosteric or nearly isosteric with Trp, they are attractive choices for conservative replacement of Trp for fluorescence enhancement. However, these analogs introduce only a modest difference in spectral properties, and usually have poor quantum yields, barring sensitive measurements. Hence, there is still a need for probes with a more dramatic shift in fluorescence excitation and emission maxima as well as an improved quantum yield.

4.4 Extrinsic solvatochromic probes to study proteins

A variety of extrinsic fluorescence probes, both engineered and natural, can be conjugated to proteins to impart even more desirable fluorescent characteristics. Advances in organic synthesis provided access to a plethora of such probes\textsuperscript{11}, ranging from the macromolecular green fluorescent protein to small molecule organics such as Fluorescin or Dansyl. Among these, small molecule probes that display the phenomenon of solvatochromism, and are amenable to site-specific incorporation into proteins and peptides are of particular interest\textsuperscript{14, 91} in the current context. This type of probe can be thought of as a way to capitalize on the desirable aspects of Trp fluorescence (sensitivity to the microenvironment), while minimizing its shortcomings (UV excitation and complex photophysics). Analogous to the discussion of Trp analogues, the ability to incorporate these extrinsic fluorophores site-specifically into a protein or peptide, as well as the size and charge of these probes are critical for minimizing structural perturbation caused by their introduction\textsuperscript{14}. The principles guiding rational design of solvatochromic probes with desirable photophysics, as well as the methods of their incorporation into proteins and peptides are outlined in subsequent sections.

4.4.1 The phenomenon of solvatochromism
The effects of the environment immediately surrounding a probe on its measurable fluorescence characteristics are complex, and can include factors such as solvent polarity, rate of solvent relaxation, and rigidity of the local environment among others. However, the effect of solvent polarity alone on fluorescence emission of a given probe, partially described by the Lippert-Mataga equation\textsuperscript{10}, can be instructive in the general understanding of fluorescence solvatochromism. The simplified Jablonski diagram shown in Figure 4.3 shows a fluorophore whose dipole increases in magnitude upon excitation.

Figure 4.3: Simplified Jablonski diagram explaining the phenomenon of solvatochromism
Solvent relaxation, or reorientation of dipoles of the solvent sphere to accommodate the now larger dipole of the probe, occurs on the picosecond timescale, and lowers the energy of the excited state while simultaneously destabilizing the ground state. Return to the ground state by a fluorescence event results in emission of photons of lower energy than those absorbed during excitation. Increasing solvent polarity causes a greater degree of solvent relaxation, resulting in emission at progressively longer wavelengths. Also a consequence of the narrowing of the energy gap caused by relaxation in progressively more polar solvents is the increase in the rate of non-radiative decay \((K_{nr})^{14}\). As a result, many fluorophores exhibiting classic solvatochromism have low fluorescence quantum yields \((\phi)\) in polar solvents.

While the Lippert-Mataga equation describes the relationship between fluorescence Stokes shifts and general solvent polarity expressed in terms of its dielectric constant and refractive index, it neglects specific solvent effects, such as hydrogen bonding. These effects can have a profound influence on the photophysics characteristics of a probe, and depend on the structural properties of the fluorophore, which in turn determine the mode of interaction of its excited state with the local environment (e.g. H-bonding of solvent molecules with H-Bond donors or acceptor moieties on the probe). It is possible to gauge the contribution of specific solvent effects on the fluorescence emission of a probe by fitting the spectral data to an empirical function of solvent polarity. Several empirical solvent polarity scales were developed, with the \(E_T(30)\) single parameter scale\(^{92}\) and the Catalán 3P multiparamater model\(^{93}\) finding the most use in the description of solvatochromic probes. The latter scale is particularly useful, as it allows delineating the specific contributions of solvent acidity, basicity and general polarity/polarizability on the emission properties of the probe.

Despite the existence of tools for gauging specific solvent effects, and myriad computational studies describing specific excited state processes responsible for environment sensitivity of fluorophores, probes that display classic solvatochromism with minimal contributions from specific solvent effects are the most sought after. This is because a straightforward relationship between a measurable fluorescence parameter and general polarity of the local environment allows description of a region of a macromolecule to which the probe is bound with more certainty, even in the absence of knowledge of its exact chemical makeup.
4.4.2 Photophysical properties of solvatochromic probes

As shown in Figure 4.3, and described in the Lippert-Mataga equation\(^\text{10}\)(see p. 94), a probe exhibiting solvatochromism must undergo a large change in the dipole moment upon excitation. In the case of classic solvatochromism, the magnitude of the dipole increases upon excitation. This observation allows for the rational design of such probes. The common wisdom for imparting solvatochromism to a probe has been to append its aromatic core with flanking electron-donating and electron-withdrawing substituents, causing increased charge separation upon excitation. This ubiquitous design feature is the one constant among solvatochromic probes developed for monitoring biomolecular structure, interactions and dynamics (Figure 4.4).

Besides the obvious requirement for strong solvatochromism to detect even subtle changes in the microenvironment, these fluorophores must satisfy a larger set of photophysical requirements to be useful for the study of biomolecules such as proteins and peptides. These include excitation at longer wavelengths, a high absorption coefficient and quantum yield. Red-shifted excitation is required to allow for selective monitoring of these extrinsic probes in the presence of endogenous fluorophores such as Trp and Tyr. Excitation in the visible range of the spectrum is also desirable as it avoids the damaging effects of UV light on biological macromolecules, and allows for using such probes for fluorescence microscopy. The combination of a high extinction coefficient and quantum yield contribute to the overall brightness of a fluorophore thus enabling ever more sensitive fluorescence measurements at lower substrate concentrations. Despite significant advances in both organic synthesis and calculation methods that drive the rational design of such probes, development of fluorophores that satisfy these requirements while minimizing their steric bulk remains a considerable challenge.

It is important to note that the properties outlined above are just general goals in the development of solvatochromic probes, and none of the fluorophores shown in Figure 4.4 can be considered ideal for the investigation of all biomacromolecules. Instead, a given probe’s photophysical and steric characteristics may make it suitable for one type of experiment and not another. This leads to some confusion in describing a particular property of a fluorophore as beneficial or disadvantageous.
Figure 4.4: Common solvatochromic probes for monitoring macromolecular structure, dynamics and interactions. Orange highlights show electron donating groups, blue—electron withdrawing groups. $\lambda_{\text{Abs}}/\varepsilon_{\text{Max}}$ are the absorption maximum and molar extinction coefficients in methanol. $\phi_{\text{MeOH}}/\phi_{\text{Toluene}}$ are fluorescence quantum yields in methanol and toluene. $\Delta\nu_{\text{Fmax}}$ is the fluorescence emission band shift in response to solvent change from toluene to methanol. Superscripts denote data collected in other solvents as follows: a — acetonitrile, b — 1,4-dioxane, c — tetrahydrofuran, d — tetrahydrofuran/methanol, e — hexane, f — hexane/methanol, g — benzene, h — benzene/methanol, i — toluene, j — cyclohexane, k — 6.1% 1,4-dioxane in water, l — 1,4-dioxane/6.1% 1,4-dioxane in water. Figure was compiled with data from: Prodan$^{122}$, FR0$^{96}$, Anthraban$^{95}$, 6DMN$^{58}$, 4DMP$^{97}$, 4-DMN$^{59}$, Merocyanines$^{100}$, Dansyl$^{102}$, 1,8-ANS$^{101}$, Coumarins$^{103}$, Nile Red$^{104}$, IANR$^{105}$, Dapoxy$^{107}$, IPyMPO$^{108}$, NBD$^{109}$, 3MC-2$^{110}$, DCDHF$^{111}$, Fluorprobe$^{112}$. 

<table>
<thead>
<tr>
<th>Probes</th>
<th>$\lambda_{\text{Abs}}$ (nm)</th>
<th>$\varepsilon_{\text{Max}}$ (M$^{-1}$ cm$^{-1}$)</th>
<th>$\phi_{\text{MeOH}}$</th>
<th>$\phi_{\text{Toluene}}$</th>
<th>$\Delta\nu_{\text{Fmax}}$ (cm$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prodan</td>
<td>345/10,900</td>
<td>0.5205</td>
<td>1180</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FR0</td>
<td>396/43,000</td>
<td>0.1998</td>
<td>5500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anthraban</td>
<td>553/45,000</td>
<td>0.3880</td>
<td>1750</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4DMN</td>
<td>456/12,100</td>
<td>0.4118</td>
<td>3170</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6DMN</td>
<td>530/29,000</td>
<td>0.2579</td>
<td>1780</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Merocyanines</td>
<td>382/8,000</td>
<td>0.0172</td>
<td>3390</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4DMP</td>
<td>373/28,000</td>
<td>0.3986</td>
<td>4760</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4DMN</td>
<td>396/6,500</td>
<td>0.1249</td>
<td>3160$^a$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>412/24,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4DMN</td>
<td>422/10,800$^a$</td>
<td>0.0172$^a$</td>
<td>1900$^a$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dansyl Derivatives</td>
<td>486/71,000$^a$</td>
<td>-0.04</td>
<td>1900$^a$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dansyl</td>
<td>335/4,600</td>
<td>0.4981</td>
<td>2220</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1,8-ANS</td>
<td>308/12,000$^b$</td>
<td>0.001$^b$-0.46</td>
<td>6620</td>
<td></td>
<td></td>
</tr>
<tr>
<td>375</td>
<td>0.001$^b$-0.52$^b$</td>
<td>2260$^b$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
For example, as mentioned earlier, many of the probes exhibiting solvatochromism have low quantum yields in aqueous environments (Figure 4.4). From one perspective this is an advantage, because such probes fluoresce brightly upon binding to proteins or membranes while remaining dark in bulk water, making them useful as "stains" in fluorescence microscopy. In contrast, their low quantum yields in polar environments prevent using these probes to accurately report on particularly polar or water exposed protein domains.

The struggle in balancing desirable photophysics and sterics can be further illustrated by following some trends in Figure 4.4. Prodan (N,N-dimethylamino-6-propionyl-2naphthylamine), with its remarkable sensitivity to solvent polarity is one of the best examples of classical solvatochromic dyes\textsuperscript{22}, and has been widely popular for biological applications despite its UV absorption\textsuperscript{94}. The first attempt to red-shift the absorption of this dye resulted in the derivative Anthradan\textsuperscript{95}, which absorbs in the visible, but at the cost of a decreased extinction coefficient, reduced solvent sensitivity and added steric bulk. The recently synthesized fluorene analogue of prodan, FR0\textsuperscript{96}, represents significant improvements in both absorption wavelength and solvent sensitivity. However, this dye has not yet been used for biological investigations.

4-N,N-Dimethylaminophthalimide (4DMP)\textsuperscript{97}, is one of the first probes for polarity sensing, and thus has found numerous applications in protein fluorescence. While it is prized for its solvent sensitivity and small rigid structure, it still suffers from UV absorption and a low extinction coefficient. The extended analogue, 6-N,N-dimethylamino-2,3-naphthalamide (6DMN)\textsuperscript{98}, was intended to improve on the properties of 4DMP. However, modest improvements in its absorption coefficient and solvent sensitivity are offset by lower quantum yields and added steric bulk. In contrast, the more recently developed 4-N,N-dimethylamino-1,8-naphthalamide (4DMN)\textsuperscript{99} shows a quite significant improvement in absorption wavelength. All dyes in this family are nearly nonfluorescent in water, limiting their applications in monitoring interactions when the label has significant water exposure throughout the process.

The merocyanine family\textsuperscript{100} represents the most desirable absorption characteristics, with marked red-shifts and very large extinction coefficients. However, their large size and negligible solvent sensitivity limits their applications as structural probes.
1-Anilinonaphthalene-8-Sulfonic Acid (1,8-ANS)\(^{101}\), one the first solvatochromic dyes, has found numerous applications in protein studies, and still remains relevant, but is virtually nonfluorescent in aqueous media. Its structural analogue, the Dansyl fluorophore\(^{102}\) has also found myriad applications in protein fluorescence, but despite its improved quantum yield in polar media, it recently declined in popularity due to its UV absorption and relatively weak solvatochromism.

The coumarin family of dyes\(^{103}\) represents a valuable addition to the tool chest of solvatochromic probes. Despite modest solvent sensitivity and UV excitation, coumarin derivatives display higher quantum yields with increasing solvent polarity, effectively reversing the trend displayed by many other solvatochromic dyes.

Nile Red\(^{104}\) is a representative example of the phenoxazine family of dyes, combining red-shifted absorption and high brightness. However, its solvent sensitivity is only moderate. A variety of its derivatives have recently been developed for the labeling of biomolecules\(^{105}\). Of these, the more compact analogues, such as APM\(^{106}\), are of particular interest as they minimize the structural perturbations caused their introduction.

Dapoxyl\(^{107}\) shows remarkable solvatochromism together with high quantum yields and extinction coefficients. However, its UV absorption has so far limited its biological applications in biosensor development. Charged derivatives of this fluorophore, such as PyMPO\(^{108}\), can function as alternatives for situations when a hydrophobic label can result in structural perturbation of the macromolecule under study.

Nitrobenzoxadiazole (NBD) is another example of an early solvatochromic probe used for protein labeling. Its advantages are its small, rigid structure and red-shifted absorption. While early experiments using NBD to track ligand-induced conformational changes in a protein, have provided valuable insights for future development of fluorescent biosensors\(^{109}\), its low solvent sensitivity limits its applications to cases where the change in environment is particularly drastic.

The more recently developed probes, such as 3-methoxychromones (see 3MC-2)\(^{110}\) and 2-dicyanomethylene-3-cyano-2,5-dihydrofurans (DCDHF)\(^{111}\), with their red-shifted absorption, high brightness, and moderate to good solvatochromism, nicely illustrate the overall goals in the development
of fluorescent polarity sensors. The latter is particularly prized because of its high photostability enabling applications in single-molecule fluorescence.

Among the fluorophores available to date, Fluoroprobe\textsuperscript{112} is the most sensitive to solvent polarity. This bichromic dye exhibits charge transfer through space resulting in its unique sensitivity. However it has not found any applications in biology because of its UV absorption, low extinction coefficient and strong fluorescence quenching in polar media.

In summary, many fluorescent polarity sensors are currently in use, and as novel biological experiments are becoming more sophisticated, the continued development of probes with improved spectral characteristics is increasingly relevant. Importantly this is not a quest for “one fits all” solution, but rather an attempt to continually increase the repertoire of tools available for today’s diverse biological applications.

**4.4.3 Incorporation of solvatochromic probes into proteins and peptides**

In addition to meeting a set of photophysical and steric requirements, the introduction of solvatochromic probes into proteins and peptides represents another important aspect of their development\textsuperscript{14}. In fact, many fluorophores discussed in the previous section and depicted in Figure 4.4 found great utility in protein fluorescence because they were amenable to site-specific incorporation into these macromolecules, while others remain theoretical curiosities because their incorporation methods are nonexistent or underdeveloped.

There are many techniques for conjugating fluorophores onto proteins and peptides, but in the context of polarity-sensing structural probes it is of interest to focus on the methods that enable maximal attribution of fluorescence data to relevant structural, dynamic or binding events. Thus, methods that introduce the fluorophore into a specific site in any desired region of a peptide or protein are the most valuable. Harsh incorporation conditions are less than desirable as they can alter the structure or function of the system under study. The common techniques for incorporation of solvatochromic dyes into proteins or peptides can be broadly divided into two categories: (1) direct covalent modification and (2) introduction of the fluorophore as an unnatural amino acid.

Direct covalent modification involves preparing a reactive derivative of the fluorophore for conjugation with select endogenous functional groups present in polypeptides. The two most common
approaches involve conjugating probes to amine groups of lysine residues (through the use of amine-selective acylating functionalities, such as O-succinimidy l esters) or to thiol groups of cysteine residues (via thiol-selective electrophiles such as α-halocarbonyls or maleimides). Of the two, cysteine conjugation is the preferred method, as this amino acid occurs much more rarely in proteins and its nucleophilic thiol functionality is reactive under milder conditions\textsuperscript{13}. Although site-directed mutagenesis allows engineering of either cysteine or lysine into virtually any desired site in protein, natural occurrence of these residues necessitates mutations at multiple sites prior to labeling, which raises the probability of altering the structure and function of the protein. Another drawback of direct covalent modification is low water solubility of solvatochromic fluorophores, which often necessitates addition of an organic co-solvent during labeling reactions. This can cause denaturation of the protein, requiring the costly and not always successful refolding procedure. Finally, because this type of labeling attaches the fluorophore via an amino acid side chain, the probe is not held as closely to the protein backbone. This additional conformational freedom complicates attribution of spectral data to characteristics of a specific site.

Engineering solvatochromic fluorophores into amino acid derivatives overcomes many of the challenges associated with direct covalent modification. In fact, many of the probes listed in Figure 4.4 have been made into amino-acid derivatives, which greatly expanded their utility as polypeptide probes. Perhaps the most notable example of this is the synthesis of the amino acid Aladan, bearing the Prodan side chain, almost simultaneously by two research groups\textsuperscript{113}, which resulted in high-profile publications describing the use of this polarity-sensor on protein systems. Introduction of these unnatural amino acids can be accomplished at the peptide level during SPPS, at the protein level through suppression of the amber stop codon during translation\textsuperscript{17}, or semi synthetically through expressed protein ligation (EPL)\textsuperscript{15,16} of an SPPS assembled peptide bearing the unnatural amino acid with a protein fragment.

In SPPS, unnatural amino acids bearing solvatochromic side chains could be incorporated directly via their corresponding Nα-Fmoc or -Boc derivative. Protection of reactive side-chains during SPPS makes it possible to incorporate virtually any unnatural amino acids, although some of their functionalities, such in the case of 4-DMP and 6-DMN\textsuperscript{114}, can cause the formation of undesired side products during standard chain elongation procedures.
At the protein level, incorporating an unnatural amino acid bearing a solvatochromic side chain is possible by chemically acylating it onto a modified tRNA which inserts it into the protein in response to an unused stop codon (TAG) during recombinant protein expression\(^{17}\). This technique is extremely difficult to implement, often suffers low incorporation yields, results in truncated recombinant proteins, and is not necessarily compatible with any unnatural amino acids\(^{13-14}\).

The ability to incorporate solvatochromic probes during SPPS can also be extended to the protein level through protein semi synthesis by EPL. Assembled by SPPS, peptides bearing the probe can be ligated to the C- or N-terminus of a recombinantly synthesized protein fragment. EPL relies on a chemoselective transthioesterification of a thioester and an N-terminal cysteine residue, which upon an S to N intramolecular acyl shift results in a native peptide bond. N-terminal ligation occurs between a fluorophore-bearing peptide thioester prepared by SPPS and a recombinant protein fragment whose N-terminal cysteine residue is exposed by proteolytic cleavage. Alternatively, the probe-bearing peptide with an N-terminal cysteine prepared by SPPS is ligated with a recombinant protein fragment which is modified into a C-terminal thioester by using a defective intein-based approach\(^{115}\). Although EPL is a powerful technique to extend the chemical flexibility of SPPS to protein systems, probe attachment using this strategy is limited to within approximately 40 residues from the C- or N-termini because of decreasing SPPS efficiency with increasing peptide length.

4.5 Converting a tryptophan analogue to a sensitive extrinsic probe: oxidative coupling reaction of 5-hydroxytryptophan (5HTP) with amines

The unique reactivity of one Trp analogue, 5HTP, suggests a possibility of creating a probe with both desirable photophysics and a built-in incorporation method. 5HI, the side chain moiety of 5HTP, is known to react with primary amines under mild, biologically compatible conditions to produce a fluorescent adduct 6H-oxazolo[4,5-e]indole (Figure 4.5). As this adduct extends the conjugation along the solvent sensitive \(1L_a\) transition of indole, it also has the potential to exhibit solvatochromism. Furthermore, since straightforward methods for incorporating 5HTP into peptides via standard Fmoc-SPPS without requiring side chain protection\(^{21}\), and into proteins through selective pressure incorporation\(^{20}\) exist, this chemistry can be used to create a fluorophore capable of site-specifically labeling these macromolecules.
This oxidative coupling reaction, first discovered during the investigation of an antitumor drug ellipticine\textsuperscript{116}, has been employed in quantifying 5-hydroxyindoles by fluorescence detection\textsuperscript{18}, and has been recently used in preparing fluorescent cyclic peptides\textsuperscript{19}. Using benzylamine as the primary amine nucleophile in a reaction with 5-hydroxyindole, produces 2-phenyl-6H-oxazolo[4,5-e]indole (PHOXI), a fluorophore with significantly red-shifted excitation and emission (Ex. 345nm, Em. 465nm), a dramatically improved quantum yield (0.54 compared to 0.17 for Trp) and a high extinction coefficient (1.54x10\textsuperscript{4} vs. 6.57x10\textsuperscript{3} cm\textsuperscript{-1} M\textsuperscript{-1})\textsuperscript{19}. Although its solvatochromism is yet to be determined, these initial findings suggest that PHOXI can be a valuable addition to the tool chest of existing solvent sensitive probes, having absorption and emission characteristics that are distinct from intrinsic protein fluorophores, and a built-in method of its incorporation into proteins and peptides.

![Figure 4.5: Oxidative Coupling of 5-hydroxindole with primary amines](image)

**4.6 Project Research Objectives**

The overall goal of this project is to develop the PHOXI fluorophore as a solvent-sensitive probe for the study of protein and peptide systems, by evaluating both its photophysics and site-specific incorporation capability. First, to investigate the solvatochromism of PHOXI, the core fluorophore is synthesized and its absorption, steady-state as well as time-resolved fluorescence properties are studied in a series of solvents spanning the dielectric continuum. The extent and nature of its solvent sensitivity is evaluated by the application of the Lippert-Mataga equation and the Catalán 3P empirical solvent model. Density functional theory calculations are be employed to investigate the structure of PHOXI and the nature of its photophysical behavior.

Site-specific incorporation of this fluorophore by derivatization of peptides containing 5HTP is also evaluated. The first and only use of this reaction on a polypeptide system was reported in 2009\textsuperscript{19}, where it
was used to modify 5HTP containing peptides. Although this report sets an important precedent of the viability of this reaction to introduce the PHOXI fluorophore into biomolecules, it lacks the necessary evidence of selectivity and biocompatibility of this derivatization because the 5-HTP peptides were prepared using \textit{in vitro} translation in minute (sub µM concentration in \(~10\mu L\)) quantities, only allowing for determination of reaction products by mass spectrometry without chromatographic separation. Thus, the authors’ claim, based on mass spectrometric evidence, that only the desired products are formed during this derivatization, is premature. It is also inconsistent with the proposed mechanism for this reaction\textsuperscript{117}, which suggests that unwanted products can arise from attack of the \(p\)-quinoneiminemethide intermediate by nucleophiles other than the intended benzylamine (Figure 5.1).

The current work focuses on using model peptides to evaluate the PHOXI derivatization strategy by monitoring reaction products using HPLC, allowing for optimization of reaction conditions. Recombinant synthesis of 5HTP containing model proteins by modified selective-pressure incorporation methods is also attempted to demonstrate the viability of the PHOXI incorporation strategy at the protein level.
Chapter 5 - Evaluation of 5HTP Derivatization in Peptide Systems and Structural and Solvatochromic Description of PHOXI
5.1 Background: Oxidative coupling of 5-hydroxyindole with primary amines

As discussed in section 4.3, tryptophan analogues have been widely pursued for the spectral enhancement of proteins. These structure-conservative probes are easily introduced into proteins in response to native Trp codons by selective pressure incorporation, resulting in site-specific labeling without the use of difficult and time-consuming strategies used for introduction of extrinsic solvatochromic probes (EPL semi synthesis, and nonsense codon suppression). However, the spectral enhancement afforded by these analogues is modest, and is easily eclipsed by that of rationally designed solvatochromic fluorophores. One of the Trp analogues, 5HTP, owing to its reactivity with primary amines under mild conditions to produce a new fluorophore with improved properties, offers an opportunity to capitalize on hassle-free selective pressure incorporation, while achieving desirable photophysics comparable to extrinsic polarity sensing probes.

![Figure 5.1: Oxidation products of 5HI under biomimetic conditions](image_url)

5-hydroxyindole (5HI), the side chain of 5HTP, has long been known to undergo oxidation under mild conditions\(^\text{118}\). A study investigating the oxidation products of 5-hydroxytryptamine (the neurotransmitter serotonin, which bears a 5HI functionality) under biomimetic conditions using H\(_2\)O\(_2\)-peroxidase, identified a number of oligomeric products (Figure 5.1)\(^\text{118a}\). The nature of these oxidation...
products suggests that the reaction proceeds through a highly reactive intermediate that is susceptible to capture by nucleophiles at the 4 position.

5HI oxidation in the presence of external nucleophiles was first encountered during investigations of 9-hydroxy derivatives of the plant alkaloid ellipticine, which contains a 5HI moiety within its structure. 5HI was found to be oxidized to the ρ-quinoneiminemethide, a potent electrophilic species\(^{119}\) (Figure 5.2A). This intermediate was found to undergo regiospecific Michael type additions by a variety of oxygen, sulfur and nitrogen nucleophiles, which were found to add specifically to the C4 indolic position) (Figure 5.2B)\(^{116}\). Further investigations used the amino group of aliphatic amino acids as the nucleophile to trap the ρ-quinoneiminemethide intermediate (Figure 5.2C)\(^{120}\), however, the structural assignment of these amino acid adducts was later called into question, and was revised to the annulated oxazolo derivatives, the fluorescent properties of which are exploited in later analytical studies (figure 5.2D)\(^{121}\).

![Figure 5.2: Reactions of 9-hydroxyellipticum with nucleophiles under oxidative conditions](image)

The generality of this reaction with various primary amine nucleophiles was then explored on a model system 6-hydroxyindole-2-carboxylate, and a mechanism for oxazole formation was proposed (Figure 5.3 shows the mechanism as it applies to 5HI)\(^{117}\). While benzylamine gave the highest yield of oxazole (61%), it is interesting to note that reaction with ethanolamine produced the unsubstituted oxazole, attributed to deacylation during or after oxazole formation.
In the early 1990's analytical chemistry researchers reported on the determination of 5-HIs (namely the neurotransmitter serotonin (5-OH tryptamine) and related metabolites) by spectrofluorometric detection of novel oxazole containing fluorophores formed during oxidation of 5HIs in the presence of benzylamine or 3,4-dimethoxybenzylamine\textsuperscript{18a}. The fluorophore produced upon benzylamine derivatization, 2-phenyl-6H-oxazolo[4,5-e]indole (PHOXI) (Figure 5.4), was reported to have fluorescence emission maximum at 481nm with excitation at 350nm. These spectral characteristics are a significant improvement over Trp analogues, suggesting that this derivatization reaction can be of use for spectrally enhancing peptides and proteins. The reaction proved to be substrate specific, as no fluorescence was observed in a reaction with unhydroxylated tryptophan and biological phenols, and was conducted in aqueous buffer with dimethylsulfoxide (DMSO) as the oxidizing agent. It was also shown that substitutions on the aryl ring of the amine reagent did not affect fluorescence properties of the resulting oxazoles to a great extent, with the 3,4-dimethoxybenzyl derivative having $\lambda_{em}=474$nm. Later analytical studies applied this method to increasingly complex biological systems further attesting to the chemoselectivity of this reaction\textsuperscript{18b, 122}. These later investigations used the mild oxidizing agent potassium hexacyanoferrate ($K_3[Fe(CN)_6]$), which is too weak to oxidize any endogenous functional groups present in peptides and proteins except for 5-hydroxyindole. Also, unlike DMSO, the salt $K_3[Fe(CN)_6]$ is not
expected to result in protein denaturation. The focus of these studies, however, was to develop a robust analytical method for detection of 5HI-containing neurotransmitters and related metabolites, not to apply this chemistry for spectral enhancement of macromolecules. Thus, the authors were concerned with optimizing reaction conditions to attain maximal fluorescence emission, without regard to yield or any nonfluorescent byproducts that may have formed.

The first and only use of this reaction on a polypeptide system was reported in 2009\(^9\), where it was used to modify 5-hydroxytryptophan containing peptides. The authors used benzylamine and 4-biotynyl-benzylamine in an intermolecular reaction with 5-HTP containing peptides to successfully synthesize (PHOXI) modified peptides. Encouraged by the results of these intermolecular reactions, the authors then included both the benzylamine moiety and the 5-HTP residue onto the same peptide and were able to produce a cyclized peptide upon oxidation with K\(_3\)[Fe(CN)\(_6\)]. These investigators also varied the residues adjacent to 5HTP in the peptide sequence to investigate the proclivity of canonical amino acids to introduce undesired side products during derivatization. Of all proteogenic amino acids, only cysteine was found to give an undesired side product, presumed to be resultant from the attack of cysteiny1 sulfhydryl group on the reactive intermediate formed during oxidation of 5HTP. The investigators also reported the quantum yield (0.54), fluorescence emission maximum (461nm), and absorption spectra.

\[ \text{Figure 5.4: Fluorogenic oxidative coupling of 5HI (1) and benzylamine.} \] Initial oxidation forms \( p \)-quinoneiminemethide (2) which is highly reactive to nucleophiles at position 4. Nucleophilic attack by benzylamine at this position, followed by oxidation to the imine, a 1,5 hydride shift, and ring closing yields (3) 2-phenyl-6\( H \)-oxazolo[4,5-e]indole (PHOXI).
(λ_{max}=326nm, \varepsilon=1.54 \times 10^4 \text{ cm}^{-1}\text{M}^{-1}) of PHOXI for the first time. However, because all of the peptides in this study were produced by in-vitro ribosomal synthesis, which is done on a very small scale (sub μM concentration in ~10μL), no chromatographic separation was performed on reaction mixtures. Instead the authors rely solely on MALDI-TOF mass spectrometry and fluorescence measurements to corroborate successful derivatization results. Using these techniques, it is impossible to determine the efficiency of the derivatization, or ascertain the absence of undesirable side products, the existence of which is likely according to earlier studies highlighting the reactivity of the \( p \)-quinoneiminemethide intermediate. A recent investigation of products, yields and kinetics of reactions of 5HI and other proquinoidal analytes (catechols) with benzylamine and related nucleophiles further calls into question the absence of side products reported by Yamagishi et al.\(^{19}\), with only 30% yield of PHOXI reported when 5HI is reacted with benzylamine under idealized conditions\(^{123}\).

While the goal set by Yamagishi et al.\(^{19}\) was to develop a novel way of producing cyclic peptides, the few spectral properties of PHOXI reported in the study, namely high brightness (through a combination of extinction coefficient and quantum yield), and visible range fluorescence emission, serve as an important precedent for employing this strategy for spectral enhancement of peptides and proteins. The fact that 5HTP can be readily introduced into proteins by selective pressure incorporation\(^{17,87a}\) and into peptides during standard Fmoc SPPS\(^{21}\), combined with evidence of chemoselectivity and biocompatibility of its derivatization to produce PHOXI provided by both the aforementioned analytical chemistry studies\(^{18,122}\), and Yamagishi et al.\(^{19}\), suggests that this chemistry can be viable for spectral enhancement of these biomacromolecules.

This project focuses on investigating the photophysics of PHOXI, and describes its solvatochromism, which was found to be comparable to established polarity sensing probes, for the first time. The viability of PHOXI derivatization is also investigated using model 5HTP containing peptides, and ways to minimize formation of undesired byproducts are identified, by chromatographic separation and monitoring of derivatization reactions. Recombinant production of 5HTP-containing proteins with the goal of demonstrating PHOXI derivatization at the protein level is also attempted.
5.2 Results and Discussion

Two 5HTP containing peptides, tryptophanyl-5-hydroxytryptophan, and TrpZip2 (SWT-5HTP-ENGKWTWK) were used to evaluate PHOXI derivatization conditions. Monitoring reaction products by HPLC allowed for identification of factors that minimize undesired product formation during the PHOXI derivatization procedure. Fluorescence and MS analysis confirmed the spectral enhancement and the identity of PHOXI derivatized peptides. The β-hairpin structured peptide TrpZip2 was also used to ascertain the extent of structural perturbation caused by introduction of PHOXI using circular dichroism (CD) spectroscopy.

PHOXI was successfully synthesized from 5HI and benzylamine by adapting a previously published procedure, and its spectroscopic properties (fluorescence emission, excitation, quantum yields and lifetimes) were investigated in six neat solvents spanning the dielectric continuum. PHOXI’s fluorescence emission maximum was found to be strongly dependent on the polarity of the surrounding media. Its solvatochromism was further investigated by analysis of spectral data using the Lippert-Mataga equation, and Catalán 3P empirical solvent scales. PHOXI’s structure and origins of its solvatochromism were further investigated by density functional theory (DFT) calculations.

Attempts to create more structurally conservative analogs of PHOXI, guided by the fact that 6-hydroxyindole-2-carboxylate has been shown to react with amines of the general structure R-CH₂-NH₂ to produce oxazoles of the general structure of PHOXI, with R replacing the phenyl ring. Specifically, analogs of PHOXI where the phenyl ring is replaced with H were investigated to minimize the size of this fluorophore and thus the structural perturbation caused by its introduction into peptides and proteins. There are precedents for using methylamine or ethanolamine as nucleophiles, albeit in organic media, to produce these unsubstituted PHOXI analogs. In contrast to benzylamine, attempts to use these reagents to label peptides in a biologically compatible manner were not met with success (data not shown). While methyamine derivatization did not produce any fluorescent products contrary to a patent claim the reaction with ethanolamine produced 3 fluorescent products, two of which were difficult to separate by RP-HPLC. The formation of these three fluorescent adducts is consistent with previous reports, but due to the difficulty in resolving them, derivatization with ethanolamine was not pursued further.
Biosynthetic incorporation of 5HTP into glutathione-S-transferase (GST) and horse skeletal muscle myoglobin (HMb) with the intent of demonstrating PHOXI derivatization at the protein level was also attempted. 5HTP incorporation into GST was achieved at a very low level (data not shown), and although subsequent PHOXI derivatization did produce a characteristic PHOXI fluorescence signature, the sample was not homogeneous due to the presence of 4 Trps in GST and low 5HTP incorporation efficiency. Attempts at expression of the simpler 2 Trp containing HMb using two different expression systems were not met with success. In view of the fact that others have successfully accomplished 5HTP incorporation into proteins\textsuperscript{87a-c}, these efforts were not pursued further, with the primary focus on demonstrating the viability of PHOXI derivatization on peptidic systems, and thorough description of its photophysical properties.

5.2.1 Biosynthetic Incorporation of 5HTP into proteins

Initial attempts to incorporate 5HTP were done using glutathione-S-transferase (GST) as model protein substrate. Although this 4 Trp-containing homodimeric protein is a less than ideal system to test the PHOXI derivatization strategy, it was chosen because it remains soluble and expresses to high levels in \textit{E. coli}. Furthermore, a plasmid vector (pET41b+) encoding GST is commercially available eliminating the need for cloning. Although traditional approaches to Trp analogue incorporation rely on using Trp auxotrophic hosts, these bacterial strains are not commercially available. Instead, an alternative approach\textsuperscript{125}, which uses an inhibitor of aromatic amino acid biosynthesis N-phosphomethylglycine (glyphosate) was used. Briefly, bacterial hosts are first grown to log phase in rich media, and then transferred to minimal media (devoid of Trp) containing glyphosate, 5HTP, as well as supplements of Tyr and Phe, prior to induction of protein expression. 6x His tagged GST produced under 5HTP incorporation conditions was then purified by Ni-NTA, and incorporation of 5HTP was checked by subjecting it to digestion with trypsin, and MS analysis of resulting peptides against theoretical values. This analysis proved inconclusive, as observed m/z values corresponding to 5HTP containing peptides were nearly identical to other predicted tryptic fragments that do not contain Trp. Subjecting these peptides to MS/MS analysis was also inconclusive, as most fragment ions observed did not contain Trp. Nevertheless, recombinantly produced GST was subjected to PHOXI derivatization conditions. Subsequent fluorescence analysis did show characteristic PHOXI fluorescence, but the spectrum was complex and
the signal was very weak, suggesting very low 5HTP incorporation levels and protein sample heterogeneity (Data not shown). Because GST contains 4 Trps with varying degrees of solvent exposure this result is hardly surprising.

Horse skeletal muscle myoglobin (HMb), containing one surface-exposed and one buried Trp, was then chosen as a simpler protein model system. It was envisioned that that once 5HTP is incorporated into both sites, PHOXI derivatization would occur only at the surface exposed site allowing for spectroscopic differentiation of the two Trps. The coding sequence for HMb was successfully subcloned into the same pET41b+ vector, replacing the coding sequence for GST. While expression of HMb under non-incorporation conditions produced high levels of both soluble and insoluble protein (which had to be refolded and reconstituted with hemin) with a characteristic red color, no expression was observed under incorporation conditions. Low expression levels (or no expression) under analog incorporation conditions were also reported by a number of groups utilizing the pET expression system and are summarized by Ross et al. These issues are attributed to the fact that this system requires the synthesis of T7 RNA polymerase upon induction of expression, which in turn transcribes the protein of interest. Because expression is induced under incorporation conditions, T7 RNA polymerase, which contains 17 Trps, can be rendered non-functional, resulting in no expression and no incorporation. To mitigate this problem, the use of the pQE expression system, which does not require the synthesis of new proteins upon induction of expression was attempted. The coding gene for HMb was again subcloned into the pQE60 vector. New E. coli host strains M15 and SG13009 were made competent and successfully transformed with pQE60HMb. However, no protein expression was observed even under non-incorporation conditions. Due to these persistent problems, 5HTP incorporation into proteins was no longer pursued.

5.2.2 Evaluation of PHOXI derivatization of model peptides

WW$_{OH}$ readily reacted with benzylamine in borate buffer, pH 8.0, in the presence of 2mM K$_3$[Fe(CN)$_6$] to produce fluorescent PHOXI adduct within 10 minutes at room temperature. After separation of reaction products by HPLC, a major peak with high absorption at 326nm, characteristic of PHOXI was isolated. This peak was isolated, and its identity confirmed by mass spectrometry as NH$_2$-Trp-PHOXI-OH (Figure 5.5). Fluorescence spectrum of this product was consistent with free PHOXI emission and excitation, further corroborating its identity as a PHOXI peptide adduct.
HPLC monitoring of reaction products observed during the PHOXI derivatization of WW\textsubscript{OH} under varying conditions is shown in Figure 5.6A, and the fluorescence spectrum of NH\textsubscript{2}-Trp-PHOXI-OH showing characteristic PHOXI visible emission with \( \lambda_{\text{max}} = 451\text{nm} \) and a marked increase in intensity.
relative to NH$_2$-Trp-5HTP-OH is shown in figure 5.6B. The PHOXI emission increase is due to synergistic effects of a high extinction coefficient (4.5 \times 10^3 \text{M}^{-1}\text{cm}^{-1} as measured in 0.5M Borate pH 8.0) and quantum yield (0.54) of PHOXI.

Conducting the oxidation reaction in the absence of benzylamine (Figure 5.6A, lower left) produces a number of products. This is consistent with the highly reactive nature of $p$-quinoneiminemethide intermediate 2 (Figure 5.4) that is formed upon oxidation of 5HTP, and with previous studies of oxidation of 5-hydroxyindole$^{118a}$ showing a complex mixture of oligomeric products resulting from nucleophilic capture of intermediate 2 at the 4 position. However, in the presence of large excess of benzylamine, the fluorescent PHOXI adduct is the major product of the reaction (Figure 5.6A, upper right).
Figure 5.7: HPLC Monitoring of TrpZip2(5HTP) before (A) and after (B) PHOXI derivatization. Reaction carried out in 250mM Borate buffer pH 8.0, with 10μM TrpZip2, 2mM K₃[Fe(CN)]₆, and 0.5M benzylamine for 10min. Asterisk indicates desired PHOXI adduct.

Figure 5.8: MALDI-TOF spectra of TrpZip2(5HTP) before and after PHOXI derivatization. (A) SWT-5HTP-ENGKWTWK Calculated [M+H] 1624.75; observed [M+H] 1624.83 (B) SWT-PHOXI-ENGKWTWK Calculated [M+H] 1726.72; observed [M+H] 1726.92.
Derivatization of TrpZip2 also afforded the desired PHOXI labeled adduct, albeit with formation of a greater number of byproducts (Figure 5.7). Despite the presence of a number of nonfluorescent oxidation products, the PHOXI adduct was readily separable by HPLC, and its identity was confirmed by MALDI-TOF MS (Figure 5.8).

The concentration of 5HTP-bearing peptide had a profound effect on the number and amount of undesired byproducts produced during derivatization. Undesired byproducts are produced when the peptide concentration is increased from 10 to 100µM (figure 5.6A, lower right). These products appear at the same retention time as those formed during the oxidation of WWOH in the absence of benzylamine. This finding suggests that at higher concentrations unwanted products result from capture of intermediate 2 by another molecule of peptide. It was found that generally keeping the 5HTP-bearing peptide at concentration of 20µM or below eliminates the majority of these byproducts. Some unwanted products, however, are insensitive to peptide concentration (note minor peaks in Figure 5.6A, upper right). Presumably, these arise from intramolecular capture of intermediate 2 by reactive groups on the peptide itself. The fact that the amount of these substrate concentration-independent products varies with the sequence of the 5HTP peptide gives further credence to this assumption, as peptides with differing sequence can have varying amounts of nucleophilic groups in proximity to the 5HTP reaction center. Although it was impossible to eliminate these byproducts by altering reaction conditions, the desired PHOXI derivatized product was easily identifiable and separable by HPLC for both WWOH and a TrpZip2.

To ensure that PHOXI derivatization conditions are selective for 5HTP only, a control reaction was also conducted using the peptide Aβ-40, which does not contain 5HTP but has all canonical amino acids except for cysteine, tryptophan and proline (Figure 5.9). This peptide was found to not be affected by reaction conditions used for PHOXI derivatization.

Derivatization of these two 5HTP-containing model peptides shows that the PHOXI fluorophore can be introduced by a quick and selective oxidation reaction in aqueous buffer in the presence of benzylamine. While unintended reaction byproducts can be observed during derivatization of 5HTP bearing peptides, the Aβ-40 peptide, which is devoid of 5HTP, is completely unaffected by the reaction conditions, attesting to the selectivity of this reaction. Because the byproducts present during the oxidation of 5HTP peptides in presence of benzylamine are the same as those formed during oxidation of
5HTP peptides in its absence, it is reasonable to conclude that these species are formed by capture of the proposed reaction intermediate 2 by nucleophiles other than the intended benzylamine.

![Figure 5.9: HPLC monitoring of PHOXI derivatization of Aβ-40 (DAEFRHDSGYEVHHQKLVFFAED-VGSNKGAIIGLMVGVV). All reactions are carried out in 250mM Borate buffer pH 8.0 for 10min, with the following additions: ■ – 20μM Aβ-40, ▲ – 2mM K₃[Fe(CN)₆], ● – 0.5M benzylamine. (A) Derivatization Reagents only (B) Aβ-40 peptide only; (C) Derivatization Reagents and Aβ-40 peptide; (D) Chromatogram C with A subtracted.]

Although the presence of these side products is less than desirable in the context of peptide labeling, their formation is consistent with the proposed reaction mechanism¹¹⁷ and is additionally documented in previous reports¹¹⁸a,¹²³. However, even with this inherent limitation, PHOXI derivatized products are easily visualized and isolated by RP-HPLC for both peptides tested. Overall, the oxidative coupling reaction of 5HTP peptides to produce PHOXI offers a fast and easy protocol for labeling of peptides: the reaction is assembled from just three components in aqueous buffer, and is completed in 10 minutes or less. Furthermore, because PHOXI is assembled in-situ using water-miscible benzylamine, it avoids the need for inclusion of harsh organic solvents during labeling reactions, an otherwise necessary step during labeling with many common organic fluorophores, which suffer from low water solubility.

Taken together, these above findings suggest that PHOXI labeled peptides are easily accessed through oxidation of 5HTP peptides in the presence of benzylamine and subsequent RP-HPLC purification. The PHOXI fluorophore is introduced site-specifically at the 5HTP site, and because it only modifies the indole side chain of 5HTP, the PHOXI label is introduced in a conformationally restricted manner, being linked to the peptide backbone through a single methylene bridge. For a probe to be useful
in characterization of the microenvironment, this short and rigid tether offers clear advantages over other common fluorophore attachment strategies (such as through side chain and terminal amine groups, as well as cysteine thiol groups), which result in longer flexible tethers and greater degrees of freedom.

The PHOXI labeling strategy is applicable to wide variety of biomolecules, primarily due to the ease of introduction of 5HTP into peptides and proteins. For example, 5HTP is easily incorporated as its Fmoc derivative during SPPS using standard protocols, incorporated into proteins in response to a Trp codon during recombinant expression in a Trp auxotrophic host, or introduced into a desired site by means of nonsense suppression. Furthermore, recent work utilizing PHOXI as a fluorescent linker for peptide immobilization suggests that Fmoc protected tryptophanyl PHOXI derivative can be introduced during SPPS as well.

Applying this chemistry to 5HTP protein substrates is expected to present challenges associated with separation of the correctly PHOXI labeled protein from undesired byproducts. While RP-HPLC purification may prove sufficient for some protein substrates, inadequate resolution and unwanted protein denaturation in HPLC buffers are among the expected challenges. However, a workaround solution based on substrate flexibility of this fluorogenic coupling reaction can be envisioned. Several groups reported that benzylamines substituted at one or more of the phenyl positions still react with 5HTP to form fluorescent adducts analogous to PHOXI. One of these reaction variants included the introduction of biotynyl benzylamine upon oxidation of a 5HTP peptide substrate. Because unwanted byproducts arise from the capture of the reactive intermediates by nucleophiles other than the intended benzylamine, using benzylamines functionalized with non-nucleophilic affinity tags, can facilitate the separation of correctly labeled protein products through affinity purification.

5.2.3 Photophysical and solvatochromic properties of PHOXI

Spectroscopic properties of PHOXI derivatized peptides

Figure 5.10 shows the fluorescence emission and excitation spectra of WW (A) and TrpZip2 (B) before and after PHOX derivatization. The excitation maxima of PHOXI analogues of WW and TrpZip2, were found to be 336nm and 341nm respectively. These wavelengths are outside of the Trp absorption envelope allowing for selective excitation of PHOXI in presence of Trp. PHOXI and Trp also form a FRET pair. We have computed the critical Förster distance, $R_0 = 23.5\text{Å}$ for the PHOXI-Trp system. This value
falls into the useful range for distance measurements in the context of biological macromolecules. Evidence of FRET can be seen in the emission spectra of PHOXI bearing peptides upon excitation at 280nm (Figure 5.10, dashed lines).

![Figure 5.10: Normalized fluorescence excitation and emission of NH$_2$-Trp-5HTP-OH vs. NH$_2$-Trp-PHOXI-OH (A) and TrpZip2(5HTP) vs. TrpZip2(PHOXI) (B). Emission spectra for PHOXI peptides are shown either with excitation at 336nm (solid black), or 280nm (dashed black).](image)

The structural perturbation caused by PHOXI substitution was also probed by obtaining CD spectra of TrpZip2 bearing either 5HTP or PHOXI at position 4 (Figure 5.11). TrpZip2 is engineered to adopt a β-hairpin structure, stabilized by stacking interaction of two pairs of tryptophan residues$^{127}$. Although one of these structure-inducing Trps is replaced by the bulkier PHOXI, the PHOXI derivative of TrpZip2 still retains some of its β-hairpin structure. As shown in Figure 5.11, the positive ellipticity at 230nm of PHOX-bearing TrpZip2 at 25°C is approximately the same as that of 5HTP-bearing TrpZip2 at 80°C.
To glean the utility of PHOXI as a spectroscopic probe capable of providing structural information, it was synthesized from 5HI and benzylamine, and its spectroscopic properties in six different neat solvents spanning a range of polarity were investigated. The results are summarized in Table 5.1, and absorption and emission spectra are shown in Figure 5.12.

Solvatochromism

Figure 5.11: Effect of PHOXI derivatization on β-hairpin stability of TrpZip2 probed by CD. A. 3D structure of TrpZip2 stabilized by Trp interactions (RCSB PDB entry 1LE1 rendered using UCSF Chime-ra). B. CD spectra of TrpZip2(5HTP) (blue) at 25°, 65° and 80°C, and of TrpZip2(PHOXI)(red) at 25°, 65° and 80°C.
Table 5.1: Spectroscopic properties of PHOXI in neat solvents

<table>
<thead>
<tr>
<th>Solvent</th>
<th>λ_{abs} (nm)</th>
<th>λ_{ex} (nm)</th>
<th>λ_{em} (nm)</th>
<th>γ_{Stokes} (cm⁻¹)</th>
<th>Φ_F (±10%)</th>
<th>Lifetime (ns)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cyclohexane</td>
<td>331</td>
<td>331</td>
<td>372</td>
<td>3330</td>
<td>0.6</td>
<td>1.69</td>
</tr>
<tr>
<td>Toluene</td>
<td>322</td>
<td>336</td>
<td>390</td>
<td>5415</td>
<td>0.69</td>
<td>1.91</td>
</tr>
<tr>
<td>1,4-Dioxane</td>
<td>333</td>
<td>335</td>
<td>409</td>
<td>5580</td>
<td>0.78</td>
<td>2.51</td>
</tr>
<tr>
<td>Acetonitrile</td>
<td>320</td>
<td>332</td>
<td>427</td>
<td>7831</td>
<td>0.72</td>
<td>3.29</td>
</tr>
<tr>
<td>Methanol</td>
<td>325</td>
<td>336</td>
<td>437</td>
<td>7886</td>
<td>0.55</td>
<td>4.06</td>
</tr>
<tr>
<td>Water</td>
<td>321</td>
<td>336</td>
<td>459</td>
<td>9366</td>
<td>0.54</td>
<td>5.71</td>
</tr>
</tbody>
</table>

PHOXI fluorescence emission maxima were found to be very sensitive to solvent, exhibiting red-shifted emission with increasing solvent polarity. PHOXI emission redshifts by 87nm in water relative to cyclohexane, suggesting that PHOXI fluorescence emission is highly sensitive to the environment.
Vibrational structure is present in the emission spectra of PHOXI in non-polar solvents cyclohexane and toluene, but not in more polar solvents (Figure 5.12). While PHOXI absorption maxima did not vary significantly (spanning a range of only 10nm), the quantum yields were found to vary, but remained high (between 0.54 and 0.78) in all solvents studied. Fluorescence lifetimes exhibited single exponential decay (Figure 5.13) and varied between 1.69ns (cyclohexane) and 5.71ns (water), increasing with solvent polarity. Susceptibility of PHOXI to photodegradation was also tested by exposure to UV light and monitoring the decrease of its fluorescence intensity at the maximum over time. The experimental set up was identical to that used to probe photostability of Prodan and its analogue FR0°66. After 1 hour, a solution of PHOXI in toluene had 72% of its fluorescence intensity remaining, compared to 10% for Prodan°66 and 76% for FR0°66.
Figure 5.13: Fluorescence lifetime decay curves for PHOXI in six neat solvents. Decay (red) and instrument response (blue) curves are plotted on a log scale. The fitted functions and residuals are shown in green. All fits were done using single exponentials, except cyclohexane and 1,4-dioxane, which were fitted using 3 exponentials, and only the lifetime of highest amplitude is shown. The low amplitude lifetimes in these cases are considered artifacts introduced by the broad instrument response function.
To better understand the solvatochromic nature of PHOXI, its spectral shifts were fitted to linear functions of a generalized solvent parameter, SP, using empirical solvent parameters described by Catalán et al.\textsuperscript{93}, and supplying proportionality constants (See Table 5.2 for a list of solvent parameters used in the analysis). Figure 5.12C shows the dependence of PHOXI emission maxima on SP; an analogous fit of the emission of Prodan in the same six solvents is included for comparison (constructed using previously published Prodan emission data\textsuperscript{128}). For PHOXI, a good linear fit (R\textsuperscript{2}=0.98) was achieved when SP = 0.79SPP + 0.12SB + 0.09SA, where SPP is solvent polarity-polarizability, SB is solvent basicity, and SA is solvent acidity, while for Prodan the best fit is realized when SP = 0.69SPP + 0.13SB + 0.18SA. Notably, almost all of the variance in PHOXI spectral shifts is explained by SPP, with only minor contributions from acidity and basicity. In fact, a single parameter fit to SPP alone gives R\textsuperscript{2}=0.94, suggesting that the dipolar nature of the PHOXI excited state is primarily responsible for the observed solvatochromism. In contrast, Prodan fluorescence emission displays roughly twice the sensitivity to solvent acidity. For PHOXI, the minor contributions of acidity and basicity are not surprising, since it contains both an acidic group (the indolic NH) and a basic group (the oxazole O).

Variations in PHOXI’s quantum yields, fluorescence lifetimes, and rates of radiative (Γ) and non-radiative (k\textsubscript{nr}) decay were also analyzed using the Catalán 3P model, and are shown in Figure 5.14. Γ and k\textsubscript{nr} were computed from measured quantum yields (Φ) and lifetimes (τ) according to equations:

\[ \Gamma = \frac{\Phi}{\tau} \]

\[ k_{nr} = \frac{1 - \Phi}{\tau} \]
Both lifetimes and quantum yield increase with solvent polarity, with excellent correlation observed for the lifetime fits. The majority of the variation in lifetimes is explained by general solvent polarity, but solvent acidity was also found to have a non-negligible contribution. Interestingly, as with solvatochromic coumarin probes, PHOXI’s quantum yields increase with solvent polarity, opposing the otherwise general trend observed for polarity sensing probes (See Figure 4.4, and discussion in section 4.4.2). This polarity trend is offset by a sizable contribution by solvent acidity, suggesting a possible quenching mechanism via accepting an H-bond. The poor correlation for the quantum yield fits is explained in part by the uncertainty of the measurements (+/- 10%). Finally both $\Gamma$ and $k_{nr}$ decrease with increasing solvent polarity, with minor contributions from SB and SA. Once again the less than perfect fits are explained by uncertainties in the quantum yield measurements, which are used to calculate $\Gamma$ and $k_{nr}$.

The change between the ground and excited state dipole moments of PHOXI can be ascertained from its spectral shifts in solvents of differing polarity by means of the Lippert-Mataga equation:

$$SP = 0.74SPP - 0.02SB + 0.27SA$$

$$\Phi = 0.52SPP + 0.04SB - 0.43SA$$

$$\tau (\text{ns})$$

$$\Gamma \left( x \times 10^7 \text{s}^{-1} \right)$$

$$k_{nr} \left( x \times 10^7 \text{s}^{-1} \right)$$

Figure 5.14: Analysis of PHOXI lifetimes, quantum yields, and rates of radiative and non-radiative decay using the Catalán 3P model.
\[ \gamma_{\text{Stokes}} = \frac{2}{hc^3} (\Delta \mu)^2 \Delta f + \text{constant} \]

where \( \gamma_{\text{Stokes}} \) is the Stokes shift expressed as energy, \( h \) is Plank’s constant, \( c \) is the speed of light, \( a \) is the Onsanger radius of the solvent cavity occupied by the fluorophore, \( \Delta \mu \) is the difference in dipole moments between the ground and excited states, and \( \Delta f \) is the orientation polarizability function defined as

\[
\Delta f = \frac{\varepsilon - 1}{2\varepsilon + 1} - \frac{n^2 - 1}{2n^2 + 1}
\]

where \( \varepsilon \) is dielectric constant, and \( n \) is the refractive index of the solvent.

Because this equation does not take into account the contributions of specific solvent interactions, we included the SA term into a multi parameter fit along with \( \Delta f \), mirroring the analysis of PRODAN solvatochromism by Catalán \cite{128}. Regression analysis of PHOXI Stokes shifts affords:

\[ \gamma_{\text{Stokes}} = 15035(0.63\Delta f + 0.37SA) + 4205 \text{ with } R^2 = 0.90 \]

Accounting for the variance in Stokes shifts caused by SA in the above equation, and substituting into the Lippert-Mataga equation gives:

\[ \Delta \mu = \sqrt{\frac{9402 \ h c^3}{2}} \]

Assuming an Onsanger cavity radius of 4.94Å, (taken as half of the distance between C4 of the phenyl substituent and the indolic N, which are the furthest heavy atoms flanking the PHOXI conjugated system, as ascertained from its geometry optimized ground state structure derived by DFT), the transition dipole moment of PHOXI is 10.6D, showing evidence of highly-dipolar excited state.

5.2.4 DFT analysis of PHOXI structure and photophysics

Geometry optimization of the ground state structure of PHOXI shows that, unexpectedly, it adopts a planar confirmation in all of the solvents studied, with ground state energy decreasing with increasing solvent polarity. Time-dependent DFT (TD-DFT) calculations identified the solvent-sensitive, lowest energy transition (346.17nm in cyclohexane and 348.31nm in water) as leading to the 1\textsuperscript{st} excited state. In both cases, it corresponds to a HOMO to LUMO \( \pi-\pi^* \) transition.
5.3 Conclusion and future directions

In addition to a reliable way for incorporation into specific sites of a biomolecule, this investigation of the photophysics of PHOXI revealed that it has many desirable characteristics of a sensitive fluorescent reporter tag and reporter of the peptide/protein microenvironment. The fluorescence emission spectrum of PHOXI lies in the visible (\(\lambda_{EM} = 459\text{nm in water}\)), and its high quantum yield (0.54) and absorption extinction coefficient contribute to its brightness. Consistent with the above properties, PHOXI peptide adducts display visible fluorescence emission and an approximately 5-fold increase in brightness relative to unmodified 5HTP peptides. In the context of biological investigations, visible emission enables fluorescence microscopy analyses, while high fluorophore brightness allows analyses at lower substrate concentrations.

PHOXI absorption also shows a significant overlap with the Trp emission envelope, allowing PHOXI to act as a FRET acceptor to Trp. The calculated critical Förster distance of 23.5Å for this pair falls within the usable range (20-80Å) for distance measurements for biological macromolecules. Since Trp is an intrinsic fluorophore, this furthers PHOXI’s utility to provide structural information, and enables it to be used in binding assays for Trp containing binding partners. Also of significance here is the fact that PHOXI absorption and excitation wavelengths are outside of the Trp absorption envelope, allowing for selective excitation of PHOXI in the presence of Trp. As seen in Figure 5.10, when \(\text{H}_2\text{N-Trp-PHOXI-OH}\) is irradiated with 336nm light, only PHOXI emission is observed, while excitation at 280nm yields both Trp emission and PHOXI emission by way of FRET.

High sensitivity of PHOXI’s fluorescence emission, lifetimes, and quantum yields to the polarity of the surrounding media allows it to be used as a sensitive reporter of the local microenvironment – greatly increasing the amount of structural, dynamics, and binding assay data that can be extracted from experiments using this probe. It was found that PHOXI’s spectral emission maximum redshifts by 87nm when the solvent is changed from the nonpolar cyclohexane to water. Probes exhibiting this degree of solvatochromism are highly sought after, and many are rationally designed to impart this property.

A comparison of PHOXI’s photophysical properties to other solvatochromic probes with absorption below 400nm is shown in Table 5.3. It is important to reiterate that none of these fluorophores have the perfect photophysics for any given biological investigation, and none, except for PHOXI, have an
intrinsic method of site-specific biosynthetic incorporation. The degree of PHOXI’s solvatochromism is comparable to that of Prodan, 6DMN, 4DMP, and Dansyl probes. The extinction coefficient of PHOXI is close to the average of the probes listed, and its high quantum yield in polar solvents makes it a great alternative to probes such as Fluoroprobe, 6DMN and 4DMP, which exhibit low quantum yields in polar solvents.

Table 5.3: Comparison of PHOXI with solvatochromic probes with absorption <400nm

<table>
<thead>
<tr>
<th>Dye</th>
<th>Polar solvent (MeOH)</th>
<th>Apolar solvent (Toluene)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>λ&lt;sub&gt;Abs&lt;/sub&gt;(nm) (MeOH)</td>
<td>ε&lt;sub&gt;max&lt;/sub&gt; (MeOH)</td>
</tr>
<tr>
<td>Fluoroprobe</td>
<td>308&lt;sup&gt;a&lt;/sup&gt;</td>
<td>12000</td>
</tr>
<tr>
<td>FR0</td>
<td>396</td>
<td>43000</td>
</tr>
<tr>
<td>Dapoxyl® derivatives</td>
<td>373</td>
<td>28000</td>
</tr>
<tr>
<td>Prodan</td>
<td>361</td>
<td>18400</td>
</tr>
<tr>
<td>6DMN</td>
<td>382</td>
<td>8000</td>
</tr>
<tr>
<td>4DMP</td>
<td>396</td>
<td>6500</td>
</tr>
<tr>
<td>Dansyl derivatives</td>
<td>335</td>
<td>4600</td>
</tr>
<tr>
<td>PHOXI</td>
<td>326</td>
<td>15400</td>
</tr>
</tbody>
</table>

λ<sub>Abs</sub>, absorption maximum; ε<sub>max</sub>, absorption coefficient; λ<sub>Fluo</sub>, fluorescence emission maximum; QY, quantum yield. <sup>a</sup>Data in cyclohexane, <sup>b</sup>Tetrahydrofuran, <sup>c</sup>1,4-dioxane.

The extinction coefficients, quantum yields, and degree of solvatochromism of PHOXI is comparable to that of Prodan, one of the most successful solvatochromic dyes in the realm of biological investigations. While PHOXI exhibits moderately lower bathochromic shifts with increasing solvent polarity than Prodan, its sensitivity to specific solvent acidity effects is approximately half that of Prodan. This more straightforward relationship will help to better ascertain the polarity of protein cavities. Although the amino acid analogue of Prodan, Aladan, was recently reported to be an efficient reporter of protein electrostatics in Shaker potassium channel and Kir2.1 proteins<sup>113a</sup>, its sensitivity to H-bond donors complicates the attribution of such data to polarity of protein cavities. In fact, Catalán has cautioned of attributing such spectral data to general polarity effects<sup>128</sup>, theorizing that the high polarity of the heme binding pocket in apomyoglobin as ascertained from Prodan fluorescence<sup>129</sup> can also be attributable to specific interactions with nearby NH sites.

Furthermore, the molecular structure of PHOXI does not resemble the traditional scaffold of electron donor/acceptor moieties flanking an aromatic system that is the hallmark of other solvatochromic
dyes. This elucidates an alternative approach in rational design of such probes. Earlier studies show that oxidation of 5HI in the presence of a series of primary amines, produces a number of structural analogues of PHOXI, which bear other alkyl and aryl groups in place of its phenyl substituent\textsuperscript{18a, 19, 117}. Furthermore, the closely related 6-hydroxyindole displays the same pattern of reactivity as 5HI, except that amines add to the 7 indolic position, resulting in a series of 8H-oxazolo[5,4-g]indoles\textsuperscript{117}. This opens up a possibility of investigating a whole library of PHOXI-like compounds as potential fluorophores, and further honing their photophysical properties.

In conclusion, this work represents a valuable addition to the field of peptide and protein solvatochromic fluorophore development. Not only does PHOXI itself make a valuable addition to the arsenal of polarity sensing probes because of its photophysics and built-in incorporation ability, but its unique structure suggest alternative routes for the design of such molecules. Finally, although the chemistry of 5HI oxidation in the presence of amines has long been known, this project describes its application to fluorophore development for the first time. Thus, this body of knowledge of the reactivity of 5HI and related systems can now be tapped to develop a collection of PHOXI analogues that can combine fluorescence with affinity purification tags, or have even more desirable spectral properties.

5.4 Materials and Methods

\textit{Materials.}

NH\textsubscript{2}-Trp-5HTP-OH (98.3\% purity) and human A\textsubscript{β}-40 peptide (purity 98\%) were purchased from Lifetein (Hillsborough, NJ); 5HTP-TrpZip2 (SWT-5HTP-ENKWTWK) (99\% purity) was custom synthesized by Peptide 2.0 (Chantily, VA). The purity of all peptides was confirmed by HPLC prior to use. Benzyamine (99\%) was purchased from Acros Organics (Geel, Belgium), and distilled before use. 5-hydroxyindole (97\%) was purchased from Aldrich (St. Louis, MO) and purified by sublimation prior to use. Preparation of 2-phenyl-6H-oxazolo[4,5-e]indole (PHOXI). 2-phenyl-6H-oxazolo[4,5-e]indole was prepared from 5-hydroxyindole and benzyamine by adapting a procedure described by Boger et al\textsuperscript{117}. Briefly, 0.375mmol 5HI and 0.45mmol benzyamine were dissolved in 40mL of dry dichloromethane and cooled to 0°C under N\textsubscript{2}. 1.75g of activated MnO\textsubscript{2} was then added and the reaction was allowed to warm to room temperature. After 4 hours, the reaction mixture was filtered through celite, and solvent was evaporated under reduced pressure. The residue containing the fluorescent product (Rf = 0.41 (DCM))
was purified by silica gel flash column chromatography eluted with 10% ethyl acetate in petroleum ether.

PHOXI 1H-NMR (CDCl3, 400MHz): δ 8.51 (br, 1H), δ 8.31-8.32 (m, 2H), δ7.52-7.53 (m, 3H), δ7.38-7.43 (dd, 2H), δ7.36(br,1H), δ7.04 (br, 1H). The 1H-NMR data are consistent with previously published reports for 2-phenyl-6H-oxazolo[4,5-e]indole

5.4.1 Recombinant Expression and selective pressure incorporation of 5HTP into proteins

GST

The plasmid vector pET41b+ coding for GST was transformed into E. Coli BL-21(DE3) expression hosts by heat shock, and positive transformants were selected by plating on LB plates containing 30mg/L Kanamycin. For analogue incorporation, cells were initially grown in “rich media” (10g/L N-Z Amine, 5g/L NaCl, 10g/L Tryptone, 4g/L Glucose, 30mg/L Kanamycin) at 37°C and 250rpm until an OD₆₀₀ ~0.7 was reached. The cells were then harvested from culture by gentle centrifugation (5min, 4000 x g, 37°C) and resuspended in prewarmed “minimal media” (50 mM Na₂HPO₄, 50 mM KH₂PO₄, 2 g/L D-glucose, 2 g/L NH₄Cl, 20 mM Sodium Citrate, 1 x 5052₁₀, 1g/L glyphosate, 60mg/L 5HTP, 60mg/mL Phe, 60mg/mL Tyr, 30mg/L Kanamycin). After a 30 min incubation at 37°C and 250rpm to exhaust remaining tryptophan stores, expression was induced by addition of 1mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Total cell protein extracts were prepared at 1, 2, 3, 4, and 12hours after induction and analyzed by SDS-PAGE. Maximum production of GST was at 3 hours post induction (~10% of total cell protein). The cells were harvested by centrifugation (10min, 10000 x g, 4°C), and the pellets immediately stored at -80°C for subsequent purification. 6 x His Tag purification was conducted under native conditions using Qiagen Ni-NTA agarose following the manufacturer’s batch purification protocol, and followed by SDS-PAGE.

Horse Skeletal Muscle Myoglobin

The amino acid sequence for this HMb, (Uniport ID: P68082), was back-translated using codons optimized for E. coli expression, and appended with nucleotide sequences introducing an N-terminal Ndel site, and a C-terminal 6 x His sequence and Xholl site. This construct was synthesized by IDT DNA and delivered in the pIDTBlue cloning vector. pIDTBlue(HMb) was then used to transform BL21 competent cells by heat shock, and positive transformants were selected by plating on LB-kanamycin plates. Plasmid DNA was extracted from overnight LB cultures using QiagenMiniprep kit, according to the
manufacturer’s protocol. Both pIDTBlue(HMb) and pET41b+ were subjected to double restriction endonuclease digestion with Ndel and XhoI according to the manufacturer’s protocol (New England Biolabs(NEB)). Digested DNA was separated on a 1% agarose gel, and the bands corresponding to HMb and linearized pET41b+ were excised. DNA was extracted from the gel using the Agilent Strataprep kit according to the manufacturer’s protocol. The vector backbone was dephosphorylated with arctic alkaline phosphatase (NEB) prior to overnight ligation with HMb insert with T4 DNA ligase(NEB) according to the manufacturer’s protocol. Aliquots of the ligation reaction were used to transform BL-21(DE3) host cells, and positive clones were selected by plating on LB-kanamycin plates, and sent for sequencing to confirm correct orientation and reading frame of the plasmid construct (pET41b+HMb).

Expression was performed as described for GST. For expression using the pQE system, the coding sequence was now appended with Ncol and BglII sites, and supplied as a double stranded gBlock from IDT DNA. The insert and the vector pQE60 were subjected to double digestion by Ncol and BglII (NEB), and subjected to the same purification, dephosphorylation, and ligation procedure as described for pET41b+HMb assembly. Heat shock competent cells were prepared from E. coli strains M15 and SG13009 (Qiagen) according to the manufacturer’s protocol and were transformed with ligation mixtures. Selection of positive clones was done by plating on LB-ampicillin plates. Positive clones from both M15 and SG13009 strains were confirmed by sequencing.

5.4.2 PHOXI derivatization of peptides

For optimization of derivatization conditions, 10-100µM of 5HTP-containing or control peptide was dissolved in 0.5M borate buffer, pH 8.0 in the presence or absence of 0.5M benzylamine in an HPLC microvial. The reactions were initiated by addition of K₃[Fe(CN)₆] to a final concentration of 2mM. After 10 minutes, the reactions were stopped and the products analyzed by direct injection onto a Series 200 HPLC system (Perkin Elmer, Waltham, MA) equipped with an XBridge C18 4.5 X 150mm column (Waters Corporation, Milford, MA). The products were separated by eluting with a linear gradient of Buffer A (0.1% TFA in water) and buffer B (0.1% TFA in Acetonitrile) at a flow rate of 1mL/min as follows: 5 min 100% Buffer A, 2 minutes increasing to 20% Buffer B, 20 min increasing to 40% buffer B. The products and unreacted starting materials were visualized by a UV-Vis flow-cell detector at 220nm (peptide amide bonds), 280nm (Tryptophan and/or 5HTP) and 326nm (PHOXI). Product peaks with appreciable
abortion at 326nm were manually collected, lyophilized and subjected to mass spectrometry to confirm identity of PHOXI derivatized products. Lyophilized NH$_2$-TRP-5HTP-OH derivatization products were dissolved in 50% Acetonitrile containing 0.1% formic acid and analyzed by direct infusion onto an Agilent 6220 ESI-TOF mass spectrometer. TrpZip2 derivatization products were analyzed using a MALDI-LTQ XL orbitrap mass spectrometer (ThermoFisher Scientific, Waltham, MA) using α-Cyano-4-hydroxycinnamic acid as matrix.

5.4.3 Spectroscopic Measurements

Absorbance measurements.

Background corrected absorption spectra were recorded on a Lambda 650 UV–vis spectrophotometer (Perkin Elmer, Walnut Creek, CA) with a 2 nm slit width, 1 cm path length, and 1 nm interval. Extinction coefficients of Trp ($5.69 \times 10^3$ M$^{-1}$cm$^{-1}$) 5HTP ($4.5 \times 10^3$ M$^{-1}$cm$^{-1}$) and PHOXI ($1.54 \times 10^4$ M$^{-1}$cm$^{-1}$) were used to determine concentrations.

Steady-State Fluorescence Emission Measurements

Fluorescence emission spectra of PHOXI derivatized peptides in water, and of PHOXI in neat solvents were recorded on a Fluorolog 3 model FL-1000 (Horiba Jobin Yvon, Edison, NJ) fluorometer. For PHOXI derivatized peptides, emission spectra were recorded with an excitation wavelength of 280 nm (Trp) and 336nm (PHOXI). For quantum yield measurements of PHOXI in neat solvents, the excitation wavelength used corresponded to maximum excitation of PHOXI in each solvent (332nm for acetonitrile; 336nm for 1,4-dioxane, methanol, toluene and water), except for cyclohexane, where 322nm had to be used to obtain a complete emission profile. For all fluorescence measurements, the step size was 1nm, the integration time was set to 0.1 s. The slit width was 2nm and matched that of absorption spectrophotometer. All concentrations were < 0.01 mM to avoid inner filter effects.

Quantum Yield Determination.

Aqueous quinine sulfate (QY = 0.54)$^{131}$ and harmane (QY = 0.83)$^{132}$ in 0.1 M sulfuric acid were used as quantum yield standards. PHOXI quantum yields in six neat solvents were determined by the method described by Williams et al.$^{133}$ Briefly, the quantum yield of the two quantum yield standards are first determined via reference against each other to ensure confidence in quantum yields measured for the unknown. The integrated fluorescence intensity of each fluorophore is determined from emission spectra
at several fluorophore concentrations, all less than 0.01 mM, to avoid the inner filter effect. The integrated intensity is plotted as a function of the measured absorbance at each concentration. A linear fit is calculated, ensuring that the fit line runs through the origin. Obviously, the goodness-of-fit should be close to unity. The slope of the fit is the gradient, \( G \), used in calculation of the quantum yield, \( QY \):

\[
QY_x = QY_{st} \left( \frac{G_x}{G_{st}} \right) \left( \frac{n_x^2}{n_{st}^2} \right)
\]

where the subscript, \( st \), signifies the quantities for the quantum yield standard, and \( x \) are the quantities for the fluorophore of unknown quantum yield. \( n \) is the refractive index for each solution: \( n = 1.333 \) for water, \( n = 1.334 \) for 0.1 M sulfuric acid, \( n = 1.497 \) for toluene, \( n = 1.427 \) for cyclohexane, \( n = 1.331 \) for methanol, \( n = 1.344 \) for acetonitrile, \( n = 1.422 \) for 1,4-dioxane.

**Trp-PHOXI Förster Distance Calculation.**

The Förster distance was calculated using the equation\(^{10}\):

\[
R_0 = 0.2108 \left[ k^2 \Phi_{Trp} n^{-4} J \right]^{1/6}
\]

Where \( R_0 \) is the critical Förster distance in angstroms, \( k \) is the dipole orientation factor and was set to 2/3 (assuming rotation of molecular dipoles is much faster than FRET rate), \( \Phi_{Trp} \) is the quantum yield of Trp in water (0.14), \( n \) is the refractive index of the medium taken as 1.4 for biological macromolecules, and \( J \) is the overlap integral in units of \( \text{cm}^{-1} \text{M}^{-1} \text{nm}^4 \), defined as follows:

\[
J = \int_0^\infty F_{Trp}(\lambda) \varepsilon_{PHOXI}(\lambda) \lambda^4 d\lambda
\]

Where \( F_{Trp}(\lambda) \) is the fluorescence emission of Trp normalized to unity, and \( \varepsilon_{PHOXI}(\lambda) \) is the normalized absorption spectrum of PHOXI scaled to its extinction coefficient \( \varepsilon_{326nm} = 15400 \text{ cm}^{-1} \text{M}^{-1} \).

**Time-Resolved Fluorescence Lifetime Decay Measurements.**

Fluorescence lifetime measurements of PHOXI in six neat solvents were carried out on a commercial instrument that employs a laser diode excitation source with a broad (1.47 ns) instrument response function (Horiba Jobin Yvon, Edison, NJ). Excitation at 340nm was achieved with a NanoLED source (<1ns pulse). An experimental instrument response function, necessary for proper deconvolution, was collected from scattered light from an aqueous casein solution at 340nm.
**PHOXI Photodegradation Assays:**

A 0.33µM solution of PHOXI in either toluene or water in a 50µL quartz microcuvette was illuminated by a Xenon lamp of a Fluorolog spectrofluorimeter at 336nm for 5000s. The lamp power at the cuvette was 8.5mW and the illumination area was 3mm H x 2mmW. Emission and excitation monochromator slits were set to 8nm. During illumination the fluorescence at the maximum (459nm for Water, 390nm for Toluene) was recorded at 1s intervals. The percentile fraction of residual fluorescence signal after an hour of illumination was recorded.

**Fluorescence Lifetime Decay Analysis:**

The lifetimes for decays collected on the diode instrument were determined by iterative convolution using the vendor-provided decay analysis software, DAS6. Details of this method are described by Eisenberg and Juszczak.\textsuperscript{135}

**Circular Dichroism of PHOXI and 5HTP containing TrpZip2 peptide.**

A Chirascan CD spectrophotometer (Applied Photophysics,) equipped with a sample temperature controller unit was used to obtain CD spectra of 25µM aqueous solutions of TrpZip2 peptides bearing either 5HTP or PHOXI in place of Trp 4. For each peptide, spectra were recorded at 25°C, 65°C and 80°C.

**5.4.4 DFT calculations**

Calculations were submitted with the help of Dr. Azaria Eisenberg. Gaussian 09 software was used for \textit{ab initio} quantum mechanical calculations. The ground-state structure was optimized using the B3LYP density functional theory method and 6-311G* basis set, with an ultrafine grid. The polarizable continuum method with implicit solvents were used to allow for the solvent effects, and the structure was checked for imaginary frequencies to insure that it was not in a transition state. Excited state calculations were done with the same B3LYP method and basis set. Calculations were done on the BOB system at the CUNY High Performance Computing Center.
References


38. (a) Olivera, B. M.; Teichert, R. W., Diversity of the neurotoxic Conus peptides. Molecular Interventions 2007, 7 (5), 251; (b) Akondi, K. B.; Muttenhaler, M.; Dutertre, S. b.; Kaas, Q.; Craik, D. J.; Lewis, R. J.; Alewood, P. F., Discovery, synthesis, and structure–activity relationships of conotoxins. Chemical Reviews 2014, 114 (11), 5815-5847.


40. Mena, E. E.; Gullak, M. F.; Pagnozzi, M. J.; Richter, K. E.; Rivier, J.; Cruz, L. J.; Olivera, B. M., Conantokin-G: a novel peptide antagonist to the N-methyl-D-aspartic acid (NMDA) receptor. Neuroscience Letters 1990, 118 (2), 241-244.


70. McManus, O. B., HTS assays for developing the molecular pharmacology of ion channels. *Current Opinion in Pharmacology* 2014, 15, 91-96.


