

5-2014

# Mechanisms of action and co-optive evolution for hypervariable courtship pheromones in plethodontid salamanders.

Wilburn Beau Wilburn  
*University of Louisville*

Follow this and additional works at: <https://ir.library.louisville.edu/etd>

Part of the [Biochemistry Commons](#), and the [Molecular Biology Commons](#)

---

## Recommended Citation

Wilburn, Wilburn Beau, "Mechanisms of action and co-optive evolution for hypervariable courtship pheromones in plethodontid salamanders." (2014). *Electronic Theses and Dissertations*. Paper 1569.  
<https://doi.org/10.18297/etd/1569>

This Doctoral Dissertation is brought to you for free and open access by ThinkIR: The University of Louisville's Institutional Repository. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of ThinkIR: The University of Louisville's Institutional Repository. This title appears here courtesy of the author, who has retained all other copyrights. For more information, please contact [thinkir@louisville.edu](mailto:thinkir@louisville.edu).

MECHANISMS OF ACTION AND CO-OPTIVE EVOLUTION FOR  
HYPERVARIABLE COURTSHIP PHEROMONES IN  
PLETHODONTID SALAMANDERS

By

Damien Beau Wilburn  
B.S., University of Louisville, 2005

A Dissertation  
Submitted to the Faculty of the  
School of Medicine of the University of Louisville  
in Partial Fulfillment of the Requirements  
for the Degree of

Doctor of Philosophy

Department of Biochemistry and Molecular Biology  
University of Louisville  
Louisville, KY

May 2014



MECHANISMS OF ACTION AND CO-OPTIVE EVOLUTION FOR  
HYPERVARIABLE COURTSHIP PHEROMONES IN  
PLETHODONTID SALAMANDERS

By

Damien Beau Wilburn  
B.S., University of Louisville, 2005

A Dissertation Approved on

March 31, 2014

by the following Dissertation Committee:

---

Richard C. Feldhoff

---

Pamela W. Feldhoff

---

Ronald G. Gregg

---

Thomas E. Geoghegan

---

Michael H. Perlin

---

S. Randal Voss

## DEDICATION

This dissertation is dedicated to my teacher, mentor, and friend

Kirk Barnett

for instilling in me a love and passion for biology.

## ACKNOWLEDGEMENTS

First and foremost, I must thank Dr. Richard Feldhoff for seven years of outstanding mentorship. I am not sure who or where I would be today without him, and I won't ever be able to thank him enough for everything he has done to help me grow as a scientist, a person, and a lover of wine. I'll never forget the demonstrations in the lab, the chats in his office, the many road trips, the nights of salamander collecting, the dinners in Valentine, and the life lessons on the porch of the Duplex – I could not have asked for a better adviser. Next I have to thank Kathleen Bowen. In countless ways, this dissertation would be a fraction of the quality without her extensive efforts to support me both scientifically and personally; her friendship has been invaluable. I must thank Kari Doty for her dedication and patience, always allowing me to use her as a sounding board as needed. And a great many thanks to Dr. Pamela Feldhoff for the support, enthusiasm, and keeping her friend in line.

My friends and family have kept me motivated during the best and worst of times, and I could not have made it this far without their constant encouragement and support. I thank my parents, Cliff and Judy Wilburn, my sister Tiffany, and brother-in-law Steve. And I owe a huge debt of gratitude to my three best friends: Tyler Kilgore, my oldest friend and brother, for always being by my side; Tom Lucas for the inside jokes, continued companionship, and many, many shenanigans; and Tyler “Tweek” Adams for the games, the drinks, and the laughs. And I thank my “wife” Lindsey Becker for being my friend and constant Biochemistry support.

I have been extraordinarily fortunate to have talented, motivated collaborators who have shown me tremendous kindness and support on many levels, and I sincerely thank all of them: Drs. Lynne Houck and Stevan Arnold for their guidance and assistance on all things salamander-related, and hosting me during my stay in Oregon; Dr. Sarah Eddy for her amazing friendship and academic support; Dr. Sarah Woodley and Chris Fonner for hosting me in Pittsburgh and their assistance on the neurophysiology project; Dr. William Pierce, Dr. Jian Cai, and Ned Smith for assistance in mass spectroscopy; Drs. Andrew Lane and Sam Arumugam for their patience and support on the NMR project; Christy Baggett, Jess Thomas, Sean Darrow, and Adam Chouinard for their outstanding field assistance, and keeping me awake on the late night collecting trips. I also thank my dissertation committee for their invaluable expertise and continued support: Drs. Ronald Gregg, Thomas Geoghegan, Michael Perlin, and Randal Voss. A special thanks goes to Dr. William Dean for always lending an ear and providing feedback whenever I needed it. None of this work would have been possible without the support of Highlands Biological Station and its fabulous staff, particularly Dr. Jim Costa, Dr. Anya Hinkle, Michelle Ruigrok, and Cynthia Soderstrom. I thank my funding sources: the National Science Foundation (IOS-0416834, IOS-0818649, IOS-1147271, the Graduate Research Fellowship Program, and Doctoral Dissertation Improvement Grant IOS-1311309), the UofL IRIG program, and the Highlands Biological Station Grant-in-Aid program.

Finally, I thank Heine Brothers coffee for – often literally – giving me the stamina and energy to survive the last 5 years of graduate school.

## ABSTRACT

### MECHANISMS OF ACTION AND CO-OPTIVE EVOLUTION FOR HYPERVARIABLE COURTSHIP PHEROMONES IN PLETHODONTID SALAMANDERS

Damien Beau Wilburn

March 31, 2014

Pheromones are an important type of chemical cue used by most animals to convey information between individuals. For more than 100 million years, male plethodontid salamanders have utilized a system of non-volatile, proteinaceous pheromones to regulate female mating behavior and receptivity. One of these pheromone components, Plethodontid Modulating Factor (PMF), is a hypervariable protein related to the three-finger protein (TFP) superfamily. Previous studies revealed that PMF persists as a rapidly evolving multi-isoform mixture. However, many characteristics of PMF as a pheromone remained undetermined, including gene structure and transcriptional regulation, translational regulation, protein structure, evolutionary mechanisms, and the isoform effects on female behavior and neurophysiology. Therefore, the broad aim of this dissertation was to characterize the mechanisms of action and evolution for PMF using the red-legged salamander, *Plethodon shermani*. The molecular and proteomic diversity of PMF was determined by RT-PCR and mass spectroscopy. The PMF complex is the product of at least 13 gene duplications in 3 gene classes containing highly conserved 5' and 3' untranslated regions (UTRs). These UTRs are bound by cold inducible RNA binding protein, which likely plays a key role in coordinating expression of the many



diverse PMF isoforms during gland development. Using mass spectroscopy and multidimensional NMR, the 3D structure of the most abundant PMF isoform was determined to have a novel structure compared to all other TFPs. In particular, an altered disulfide bonding pattern promoted greater backbone flexibility in the most rapidly evolving segments of PMF to possibly enhance male pheromone and female receptor interactions. Functional assays testing different mixtures of PMF isoforms revealed that isoform diversity is a key requirement for increasing female receptivity, likely through synergistic interactions in the vomeronasal organ and/or the brain. Examination of pheromones in a different plethodontid species (*P. cinereus*) revealed that the majority of PMF duplications occurred within the last ~20 million years. In summary, in response to female sexual selection, the PMF gene complex has evolved through an unusual disjunctive evolutionary process as part of a birth-and-death model of gene evolution to permit coordinated expression of dozens of flexible proteins that synergistically function to regulate female behavior.

## TABLE OF CONTENTS

	PAGE
DEDICATION	iii
ACKNOWLEDGEMENTS	iv
ABSTRACT	vi
LIST OF TABLES	x
LIST OF FIGURES	xi
 CHAPTER	
I. INTRODUCTION	1
II. PROTEOMIC AND UTR ANALYSES OF A RAPIDLY EVOLVING HYPERVARIABLE FAMILY OF VERTEBRATE PHEROMONES	
Chapter Overview	19
Introduction	21
Methods	25
Results	29
Discussion	43
Conclusions	49
III. STRUCTURAL INSIGHTS INTO THE EVOLUTION OF A SEXY PROTEIN: NOVEL TOPOLOGY AND RESTRICTED BACKBONE FLEXIBILITY IN A HYPERVARIABLE PHEROMONE FROM THE RED-LEGGED SALAMANDER, <i>PLETHODON SHERMANI</i>	
Chapter Overview	50
Introduction	52
Materials and Methods	56
Results	64
Discussion	85
Conclusions	89
IV. DIFFERENTIAL FEMALE BEHAVIOR FROM A HYPERVARIABLE COURTSHIP PHEROMONE IN THE RED-LEGGED SALAMANDER, <i>PLETHODON SHERMANI</i>	
Chapter Overview	90
Introduction	92
Materials and Methods	98
Results	103
Discussion	105

V.	OLFACTORY EFFECTS OF PMF ISOFORM VARIABILITY IN THE RED-LEGGED SALAMANDER, <i>PLETHODON SHERMANI</i>	
	Chapter Overview	109
	Introduction	111
	Results	115
	Discussion	120
	Experimental procedures	125
VI.	GENE EXPRESSION DURING SEASONAL ORGANOGENESIS: PROFILING TRANSCRIPTOME CHANGES IN AN ANNUAL CYCLE OF GLAND DEVELOPMENT AND HYPERVARIABLE PHEROMONE SYNTHESIS	
	Chapter Overview	129
	Introduction	131
	Results	137
	Discussion	167
	Conclusions	174
	Methods	176
VII.	PROTEOMIC ANALYSES OF COURTSHIP PHEROMONES IN THE REDBACK SALAMANDER, <i>PLETHODON CINEREUS</i>	
	Chapter Overview	190
	Introduction	192
	Materials and Methods	197
	Results	201
	Discussion	217
VIII.	CONCLUSIONS AND FUTURE DIRECTIONS	224
	REFERENCES	230
	CURRICULUM VITA	258

## LIST OF TABLES

TABLE	PAGE
1. Mean distance between PMF clusters	33
2. Summary of PMF 3' UTRs by class	36
3. Summary of mass spectroscopy data	40
4. Models of positive selection for <i>P. shermani</i> PMFs	41
5. Summary of mass spectral analysis in PMF-G disulfide bonding pattern determination	68
6. Restraints and statistics of PMF-G structural ensemble	73
7. Summary of pheromone isoforms by HPLC fraction	95
8. Summary of GLM estimates by pheromone treatment	118
9. Summary of <i>P. shermani</i> mental gland transcriptome	143
10. Select list of differentially expressed genes	144
11. MANOVA results for CIRBP expression and regulation	165
12. Primers used for qRT-PCR, CIRBP expression, and <i>in vitro</i> transcription	179
13. Summary of mass spectral data for RP-HPLC purified <i>P. cinereus</i> proteins	205
14. Amino acid distance matrix for <i>P. cinereus</i> vs <i>P. shermani</i> PMF sequences	207
15. Primer list for <i>P. cinereus</i> PCR products	210

## LIST OF FIGURES

FIGURE	PAGE
1. Tail straddling walk	9
2. Slapping pheromone delivery	11
3. Models for multigene family evolution	15
4. Alignment of PMF cluster consensus sequences	30
5. Complete alignment of PMF cDNA translations	31
6. Gene tree of <i>P. shermani</i> PMF cDNAs	32
7. Summary of PMF UTR analysis by gene class	35
8. Chromatographic separation of <i>P. shermani</i> pheromones	38
9. Purification scheme of natural PMF-G	65
10. PMF-G contains 4 disulfide bonds	66
11. Partial reduction of PMF-G	67
12. Mass spectral analysis of partially reduced PMF-G	69
13. Comparison of native and recombinant PMF-G	71
14. PMF-G amide H-D exchange rate	74
15. NMR-derived structural ensemble of PMF-G	75
16. Surface models of PMF-G	76
17. Measurements of sequence and structural variability in PMF	78
18. Homology modeling of major PMF isoforms	80
19. PMF-G backbone amide <sup>15</sup> N linewidths	81

20. rPMF-G NMR relaxation analysis	82
21. Rates of molecular evolution on PMF-G	84
22. Effect of PMF components on female receptivity	104
23. Agmatine immunohistochemistry	116
24. Model estimates of AGB count data	117
25. Mental gland hypertrophication	135
26. Mental gland histology	139
27. Pheromone immunohistochemistry	140
28. <i>P. shermani</i> mental gland cDNA	142
29. Cluster analysis of mental gland gene expression	146
30. qRT-PCR analysis of select mental gland genes	148
31. CIRBP – PMF 3' UTR interactions	150
32. CIRBP affinity for different RNAs	152
33. Competition EMSA with CIRBP	153
34. EMSA with CIRBP domains	155
35. CIRBP – RNA interactions stabilized by formaldehyde crosslinking	156
36. CIRBP titration curve	157
37. PMF 3' UTR induces secondary structure changes in CIRBP	159
38. CIRBP immunohistochemistry	161
39. CIRBP protein analysis	162
40. Phylogenetic relationship among plethodontid salamanders	194
41. SDS-PAGE separation of mental gland pheromone extracts between species	202
42. RP-HPLC separation of <i>P. cinereus</i> pheromone extract	203

43. SDS-PAGE analysis of individual RP-HPLC fractions	204
44. Alignment of identified <i>P. cinereus</i> PMF sequences	209
45. Alignment of identified <i>P. cinereus</i> PRF sequences	212
46. Deglycosylation analysis of <i>P. cinereus</i> PRFs	213
47. Glycosylation analysis of individual RP-HPLC fractions	215

## CHAPTER I – INTRODUCTION

Around 1300 AD, an anonymous cleric in northern England wrote the classic poem *Cursor Mundi*, or “Runner of the World,” which chronicles the history of the earth from the perspective of the Christian Bible. Within its approximately 30,000 lines, the cleric defines five human “wits” to parallel the five elements of Aristotle: touch to earth, taste to water, sight to fire, hearing to the ether, and smell to the air. More commonly, these are known as the five basic senses, but this 700 year old poem may represent the first documented description of what is today called sensory biology [1]. In the modern context, the definition has been refined as the study of how living organisms perceive mechanical, electromagnetic, or chemical cues [2-4]. While this description does include the five senses (touch and hearing to mechanical, sight to electromagnetic, taste and smell to chemical), it includes many more stimuli that organisms commonly perceive from their environment: movement of particles, substrate vibrations, electrical or magnetic fields, thermal energy, or chemicals in the form of pheromones [5-8]. In contrast to mechanical or electromagnetic stimuli, there are thousands to millions of potential chemical cues, requiring organisms to possess many more molecular sensors of exquisite sensitivity and specificity in order to fully perceive the chemical world [4, 5].

Chemical sensing is likely the oldest of the three basic types, with pheromones themselves being an ancient evolutionary innovation, even utilized by baker’s yeast (*Saccharomyces cerevisiae*) [9]. Compared to the five senses in the *Cursor Mundi*,



pheromones are a much more recent discovery. The term was first coined by Karlson and Luscher in 1959, a combination of the Greek words *pherin* (“to secrete”) and *hormone* (“to stimulate”), and was defined as “substances which are secreted to the outside by an individual and received by a second individual of the same species, in which they release a specific reaction, for example, a definite behavior or a developmental process” [10].

The first animal pheromone, bombykol, was discovered by Adolf Butenandt in silkworm moths (*Bombyx mori*) as a long chain unsaturated alcohol secreted by females and used to attract male mates [11]. In the 50+ years since the discovery of bombykol, thousands of different chemicals have been identified in insects to perform similar functions, including frontalin in the invasive bark beetle (*Dendroctonus valens*) [12], heptacosadiene in fruit flies (*Drosophila melanogaster*) [13], derivatized fatty acids in ermine moths (*Yponomeuta spp.*) [14], and tetradecenyl acetates in European corn borer (*Ostrinia nubilalis*) [15]. Presently, a major online pheromone database, the Pherobase (<http://www.pherobase.com>), lists more than 6,500 different semiochemicals for over 7,000 species, the overwhelming majority of which are from insects. A common theme among insect pheromones is their chemical composition, as most are small, volatile hydrocarbon chains with variable modifications which define their specificity and function [16]. Continual discovery of new and diverse bioactive molecules which impact social function led to extensions of the original Karlson and Luscher definition with new terms for interspecies signaling: allomonones (benefit sender), kairomones (benefit receiver), and synomonones (benefit both sender and receiver) [17-19]. However, many pheromones also qualify as allomonones/kairomones/synomonones depending on the species

of the receiver, and as they signal through common neurological pathways, more recently these terms have been unofficially consolidated under the umbrella of “pheromones” [5, 20]. The incredible diversity of pheromones within and between species of Lepidoptera has been proposed as one of the mechanisms that has permitted their expansive radiation to over 100,000 species [21]. While the experimental advantages of insects (small size, ease to maintain, short generation time, etc.) have allowed tremendous advances in the study of pheromones as a whole, their complex biosynthetic pathways have prohibited addressing many evolutionary questions with respect to modes and rates of selection [14].

In contrast to insects, a number of vertebrates utilize peptide or protein pheromones for chemical communication, which, as direct gene products, have obvious advantages for studying both evolutionary history and regulation of biosynthesis. To date, protein pheromones and sensory mechanisms have been most well-characterized in mammals – specifically, rodents [22]. In mice, protein pheromones are commonly delivered as part of general bodily fluids: urine, tears, milk, etc. Two key families of proteins have been implicated as pheromones: major urinary proteins (MUPs) and exocrine secreted proteins (ESPs). Both MUPs and ESPs are multigene families with many expressed isoforms in urine and tears, respectively [23, 24]. Isoform number and sequence are variable between different strains of inbred mice, with many isoforms evolving unique functions. For example, one MUP isoform termed darcin (named after the male protagonist from *Pride and Prejudice*) increases female mating receptivity, and simultaneously promotes learning of volatile cues in the male’s urine [25, 26]. In contrast, a different subset of isoforms promote aggression and competition between rival

males [27]. MUPs have a similar genomic architecture to other multigene families: many tandemly repeated genes including multiple pseudogenes [24].

Mice, and most other vertebrates, sense pheromones through a specialized set of neuronal epithelia (the vomeronasal epithelia, VNE) found in the vomeronasal organ (VNO). In contrast, most general odorants are detected through the main olfactory epithelium (MOE). As a secondary form of chemoreception, the VNO/VNE are often referred to as the accessory olfactory system [5, 22]. The main difference between these two neuronal populations is the types of chemoreceptors that are expressed. In the MOE, individual neurons express a single receptor, one of approximately 1000 olfactory receptors (ORs), such that each neuron is highly tuned for detecting only a small subset of ligands [28]. These studies by Axel and Buck later culminated in their being awarded the 2004 Nobel Prize in Physiology or Medicine. In contrast to the MOE, the rodent VNE is divided into apical and basal layers expressing vomeronasal type-1 receptors (V1Rs) or type-2 receptors (V2Rs), respectively [29, 30]. ORs, V1Rs, and V2Rs are all G-protein coupled receptors (GPCRs), but share no recent homology [31]. Compared to both ORs and V1Rs, V2Rs possess long extracellular domains which were hypothesized to bind peptide or protein pheromones [30]. Binding of pheromones to V1R/V2Rs activates the IP<sub>3</sub> signaling pathway, leading to opening of TRPC2 cation channels, membrane depolarization, and signal transduction from the VNE to the brain [32]. The role of TRPC2 has been well established as a critical channel for pheromone signal transduction: in *trpc2*<sup>-/-</sup> mice, there zero sensory stimulation in the VNE, but mice also lose all ability to discriminate gender, and males would equally attempt to mount males or females [33]. VNE neurons project to the specialized vomeronasal amygdala, followed by the

hypothalamus, where behavior and/or endocrine effects may be mediated [5]. Despite nearly 20 years of research, only a small number of specific ligand:receptors pairs have been identified for V1R/V2Rs [23, 34, 35]. Unlike MOE neurons, there is evidence that VNE neurons can express more than one type of receptor, which can affect both their sensitivity and even specificity to particular stimuli [35, 36]. Vomeronasal receptors do not translocate to the plasma membrane spontaneously, and require selective association with one of many M1/M10 class MHC 1b molecules [37]. However, one of the few well characterized pheromone-V2R pairs is ESP1 and V2Rp5 [23]. Recently, the structure of ESP1 was determined by multidimensional NMR, and binding assays coupled with site-specific mutagenesis revealed that ESP1 interacts primarily by salt bridges with the extracellular domain of V2Rp5 [38]. Difficulty in identifying specific pheromone-receptor interactions has led to recent studies focusing on understanding how specific classes of V1Rs and V2Rs recognize broad types of signals (conspecific male, conspecific female, heterospecific predator, heterospecific non-predator, etc.) [20].

Beyond rodent models, most research on olfactory and VNO signaling has centered on anatomical and comparative genomics studies. Most fish possess a single olfactory organ, with individual neurons expressing all three types of receptors (ORs, V1Rs, and V2Rs) [39]. In contrast, lungfish have specialized neurons that project to the ventrolateral olfactory bulb, suggesting a primitive accessory olfactory system which evolved in a common ancestor to lungfish and tetrapods [40]. In clawed frogs (*Xenopus spp.*), the MOE possess all three receptor types, but the VNO only contains V2R-expressing neurons [41, 42]. Humans are seemingly most similar to frogs in that they possess a MOE which expresses both ORs and V1Rs, but completely lack an accessory

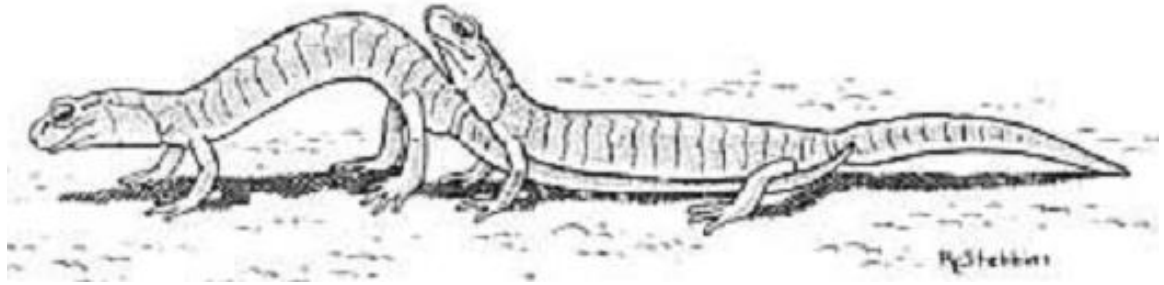
olfactory system [43]. Throughout terrestrial vertebrate evolution, there has been significant redistribution of these receptor families such that putative pheromone receptors, i.e. V1Rs and V2Rs, are not always expressed in the VNO.

For ~25 years, salamanders and newts (order Caudata) have served as valuable non-mammalian models to study the evolution and function of protein pheromones in vertebrates [44-46]. Caudata includes approximately 655 extant species of salamanders and newts, nearly all located in the northern hemisphere, divided into ten families [47]. The size, diet, and life cycle vary dramatically between species, yet multiple genetic studies have resulted in conflicting models of phylogeny [47-50]. Possibly the earliest lineage includes the family Cryptobranchidae, which has only three extant species in two genera (*Cryptobranchus alleganiensis*, hellbenders; *Andrias japonicus*, Japanese giant salamander; *Andrias davidianus*, Chinese giant salamander). These are the largest known amphibians in the world, with the Asian species reaching ~1.5-2 m in length and they are purely aquatic [49]. In contrast, *Ambystoma* (the only genus in Ambystomatidae) represent more stereotypical amphibians in that they metamorphose from an aquatic larval stage into terrestrial adults. However, even this is not universally applicable to all *Ambystoma spp.*, such as the Mexican axolotl (*Ambystoma mexicanum*): famous for its ability to regenerate its limbs and even its spinal cord [51], axolotls are paedomorphic salamanders that never undergo metamorphosis and remain in a permanent juvenile-like state [52, 53]. In the eastern newt (*Notophthalmus viridescens*), there are actually three life stages: an aquatic larval stage similar to *Ambystoma*, followed by a juvenile terrestrial stage (“red eft”) for 2-3 years that often includes traveling to new ponds, and then metamorphosis into an aquatic adult [49].

The majority of extant salamander species (>50% of the ~655 species) are in the family Plethodontidae [49]. Existing for more than 100 million years, plethodontid salamanders are extraordinarily successful in both species richness and population count, such that in some ecosystems, salamanders contribute more to the total biomass than any other vertebrate [54]. In contrast to all of the aforementioned examples, most plethodontid species lay their eggs on land, and juvenile salamanders emerge in their terrestrial adult morphs. Likely the most defining feature of this family is that they are completely lungless, and breathe entirely through their skin and the mucosa of their mouths. However, another important characteristic are their nasolabial grooves: vertical slits along the snout that transports chemicals from the substrate to the nares by capillary action, enhancing chemoreception [55]. Using a number of different types of skin glands, salamanders release both volatile and non-volatile pheromones to provide information on species [56, 57], size [58], sex [56], female gravidity [58, 59], diet [60, 61], and parasite load [62]. These pheromones facilitate many social functions including territoriality [63, 64], predator avoidance [65, 66], prey detection [67, 68], and reproductive behavior [44, 69]. While plethodontid salamanders have separate MOE and VNE, the two share a common olfactory chamber with the VNE occupying lateral diverticula [70]. The VNE, but not the MOE, is sexually dimorphic with males having approximately twice as many neurons [71]. Molecular studies using RT-PCR and *in situ* hybridization revealed that the Ors and V2Rs are exclusively expressed in the MOE and VNE, respectively, with no evidence for V1R expression [72]. Similar to mice, vomeronasal neurons project to the accessory olfactory bulb, which only target the vomeronasal amygdala. From there,

further connections are made to the hypothalamus and preoptic area [73, 74], which are brain centers involved in vertebrate reproductive behavior [75, 76].

Along with their other unique characteristics, plethodontid salamanders perform a unique, highly stereotyped courtship ritual. Despite some deviations between species, there are five basic stages to the courtship: orientation, persuasion, tail straddling walk (TSW), deposition, and insemination [77, 78]. Orientation includes male salamanders finding a female salamander and aligning his body perpendicularly near the base of her tail, and he may then perform a number of different behaviors to persuade her to initiate TSW. In one large clade of eastern *Plethodon spp.*, persuasion includes a “foot dancing” behavior where the male repeatedly moves both his fore and hind limbs up and down, possibly as a measure of stamina and endurance [78, 79]. In other species, males will circle the female and periodically bite at her for unclear reasons [80]. In the majority of species, males deliver non-volatile proteinaceous courtship pheromones using enlarged premaxillary teeth to “scratch” the female’s dorsum [44]. A small gland on the male’s chin (mental gland) secretes pheromones that are applied to the abraded area and presumably diffuse into the female’s bloodstream, where they may freely diffuse and affect any number of target tissues [81]. Assuming a successful persuasion, the male and female salamander will enter into TSW, where a female straddles a male’s undulating tail, and the two walk in unison for an extended period of minutes to hours (Figure 1). Species that deliver pheromones by scratching the female’s dorsum tend to have the shortest tail straddling walk phases (median of ~5 min). For species that do not deliver pheromone, TSW can extend up to 300 min [82].

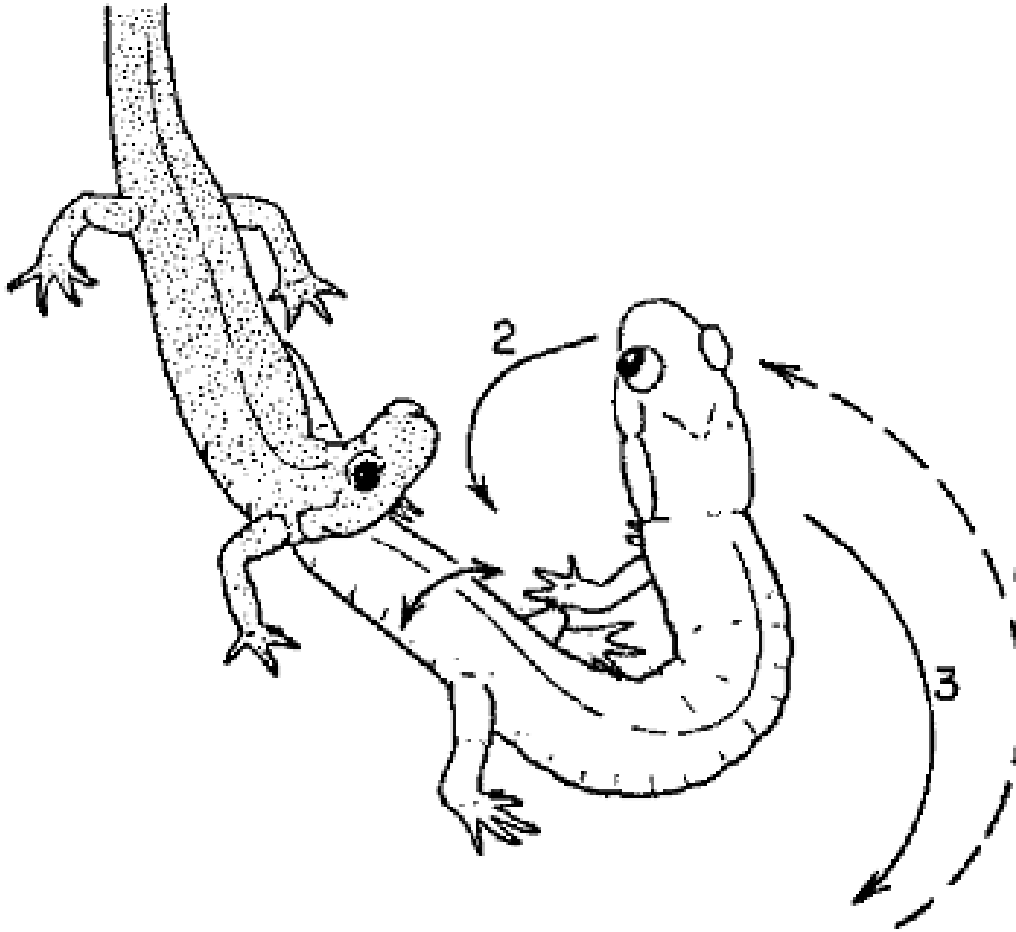


**Figure 1. Tail straddling walk**

Illustration of plethodontid salamanders in tail straddling walk, with the male salamander on the left and the female straddling his undulating tail on the right [78].



While multiple types of mental glands exist, most species deliver pheromones through the transdermal scratching mechanism [45]. However, for a single clade of large eastern *Plethodon spp.* (the *P. glutinosus* complex, 28 species), mental gland morphology has transitioned from a small cluster near the tip of the chin to a large, pad-like structure on the lower jaw. In these species, during TSW, males will bend around and directly “slap” the female’s snout with the mental gland to release pheromones onto an area that is continuous with the female’s olfactory chamber (Figure 2) [83]. These species have intermediate TSW lengths (mean ~30 min) [84], and application of pheromones further reduces this time [85]. Once in the olfactory chamber, pheromones activate VNE neurons by binding to specific receptors, affect the aforementioned regions of the brain, and alter the female’s mating behavior [70, 83, 86]. During TSW, the male will periodically pause and wait to see if the female tries to advance forward; if she does not, he will initiate spermatophore deposition. Depending on the species, this process requires the male to press his body to the ground and remain still for 4-7 min. The male lifts his vent off of the deposited spermatophore, walks forward with the female following to receive the spermatophore into her cloaca, and thus completes courtship [78]. The length of the courtship season varies between species, but generally, species with scratching delivery mate in either spring (April-May) or summer (August-September), while slapping species only use the summer mating season. These one or two brief mating windows require males to maximize their number of mating opportunities, such that even small changes in courtship time could have profound impacts on reproductive success over several million years [87].



**Figure 2. Slapping pheromone delivery**

Illustration of a male plethodontid salamander (right) bending turning his body (1), slapping his enlarged mental gland to the female (2), and returning to tail straddling walk (3) [78].

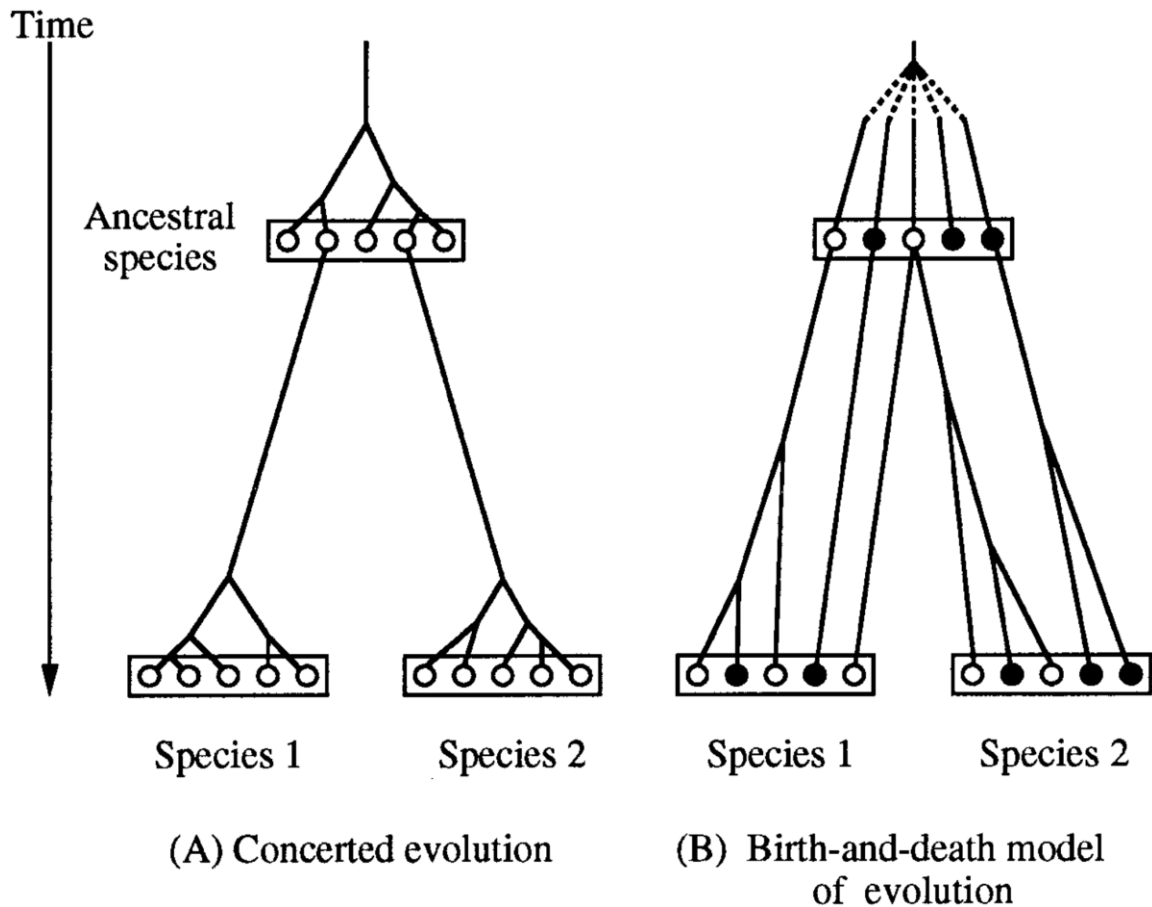
Chemical analysis of the mental gland has principally focused on two species, *Ocoee's* salamander (*Demosgnathes ocoee*) and the red-legged salamander (*Plethodon shermani*), as models of scratching and olfactory delivery systems, respectively [69, 88]. *D. ocoee* expresses a relatively complex mixture of >30 different proteins. Combined DNA sequencing and mass spectral proteomics efforts revealed that many of these proteins are products of gene co-option: gene duplication followed by mutation to drive neofunctionalization. In particular, many hormones (insulin, glucagon, insulin-like growth factor, leptin) were targets of gene co-option in *D. ocoee* and may adjust female physiology during courtship [88, 89]. One protein, sodefrin precursor-like factor (SPF), represented ~33% of the total pheromone, and experimental application of SPF on female *D. ocoee* salamanders reduced courtship time [81]. In contrast to this relatively complex blend, analysis of the *P. shermani* pheromone extract revealed that >98% of the total pheromone included only three major protein components of 22 kDa, 19 kDa, and 7 kDa [69]. The 22 kDa component, or Plethodontid Receptivity Factor (PRF), consisted of 3 isoforms with sequence similarity to IL-6 cytokines [83]. The 7 kDa component, or Plethodontid Modulating Factor (PMF), was much more diverse with >30 expressed isoforms, and shared homology with the three-finger protein (TFP) superfamily [90]. PMF and PRF each comprised ~50% and ~33% of the whole pheromone extract, respectively, but because of their molecular weight differences, the stoichiometry of PMF to PRF was ~5:1 [91]. Both proteins individually activated neurons in the VNE and in the brain [86, 92], but in behavioral assays, only PRF decreased courtship time similar to whole pheromone extract while PMF surprisingly increased it [83, 93]. Comparison of cDNA sequences of PMF, PRF, and SPF across multiple plethodontid species revealed

all three proteins are rapidly evolving and under pervasive positive selection [94-97]. The remaining 19 kDa component in *P. shermani* pheromone extract was unidentified until recently when it was shown to belong to the TIMP superfamily (see Chapter VI), however, its biological functions are still unknown.

The TFP superfamily, of which PMF is a member, is an extraordinarily diverse group of proteins. Well-known TFPs include many snake venom cytotoxins and neurotoxins [98-100], but paralogs exist in all vertebrates, including putative toxins in rainbow trout [101], skin secretions from clawed frogs [102, 103], positional cell surface markers of limb regeneration in amphibians [104], the complement protein CD59 [105], the urokinase receptor uPAR [106, 107], and the human Ly6 antigen [108, 109]. Establishing a proper TFP phylogeny has been extraordinarily difficult as there is minimal sequence conservation between members; however, the family receives its name due to all members adopting a common protein topology of two short parallel  $\beta$ -sheets resembling “three fingers.” This topology is stabilized by 8 conserved cysteine residues that adopt a canonical disulfide bonding pattern of 1-3, 2-4, 5-6, 7-8 [110]. The net result is a highly stable protein platform with mutable loops such that a near infinite number of sequences may be evaluated for novel functions. This hypothesis was indirectly tested using a synthetic TFP with randomized loops, binding the mutants to the IL6 cytokine receptor, and selecting only the bound versions. The result was multiple high affinity TFPs that functioned as either IL6 receptor agonists or antagonists, further demonstrating the versatility of the TFP domain [111].

Snake toxin TFPs, along with many of the other co-opted gene families in snake venoms (e.g. phospholipase A2, VEGF, CRISP, cystatin, whey acidic protein, kallikrein),

have evolved through a birth-and-death model of molecular evolution [112-115]. Originally proposed by Nei et al. [113] as an alternative to concerted evolution for immune system genes, in this model for multigene families, genes are repeatedly duplicated and different sequence permutations are evaluated by selection. Functional genes are free to be duplicated and mutated further, while maladaptive variants are discarded via pseudogene conversion and/or chromosomal deletion (Figure 3) [113]. For snake venoms TFPs, the net result has been many isoforms with variability in prey specificity, relative toxicity, and molecular function, including ion channel blocking [116-118], nicotinic and muscarinic receptor antagonists [119-121], acetylcholinesterase inhibition [100], cell-adhesion regulation [122], integrin binding [123], and pore formation in the plasma membrane [124, 125]. Recent sequencing of the Indonesia king cobra (*Ophiophagus hannah*) genome provided new insights into this model of venom evolution, suggesting that (1) venom adaptation occurs by either gene hijacking/modification or duplication of non-toxin genes, and (2) positive selection seemingly acts more quickly on highly expressed toxin genes (such as TFPs). This accelerated evolution is in contrast to species that primarily rely on constriction for prey immobilization (e.g. pythons) [126]. As venom toxicity directly affects prey capture, there is a strong advantage from natural selection for acquiring a highly potent blend of toxins that can rapidly immobilize any potential prey species, with accelerated evolution from arms-race dynamics [115].



**Figure 3. Models for multigene family evolution**

Comparison of models for (A) concerted evolution and (B) birth-and-death model of evolution. In the concerted evolution model, multigene families evolve independently within species by a single common ancestor and/or gene conversion stabilize multiple copies towards a common sequence through homologous recombination (e.g. ribosomal RNA genes). In the birth-and-death model, genes continually divide and diversify irrespective of speciation, with favorable genes (white circles) further duplicated for possible functional innovation, and deleterious genes (black circles) eliminated by gene removal or pseudogene conversion [113].

The birth-and-death model of snake venom evolution is potentially analogous to that of plethodontid salamander pheromones. However, rather than natural selection acting through prey capture, female receptors and associated reproductive behaviors are agents of sexual selection, which have promoted rapid evolution of pheromone genes [94]. With respect to inheritance, it has been argued that rapid evolution should be most readily observed in genes that affect survival and energy acquisition, as these are likely stronger selective forces than mate choice [126]. However, genes for prey capture are often tied into functional complexes: associated sets of traits that cooperatively work to accomplish a single biological task [127]. In addition to the toxin genes themselves, venom systems require a functional venom gland (and the associated structural genes), a way to inject the toxins (e.g. fangs), and a means to protect the predator from self-exposure. As all of these elements contribute to the same function, it is unsurprising that they act as selective tethers on the other elements and generally constrain evolution until there is sufficient directional pressure, such as need to capture a new type of prey, whereupon evolutionary “bursts” are observed [128-130]. Despite plethodontid pheromones existing as part of a larger complex surrounding courtship behavior (common mental gland, tail straddling walk, and relative stasis in delivery mode), as secreted products that have little-to-no metabolic feedback in the male, the only strong selective tether is female receptors. Similar rates of evolution were observed for interacting proteins between egg and sperm, best characterized in abalone with VERL and lysin, respectively [131-134]. However, due to likely stoichiometric constraints, lysin persists as a single gene copy that is apparently constrained with suboptimum affinities in order to maintain interactions with different VERL isoforms [135]. This is not a

limitation in plethodontid salamanders, particularly for species utilizing olfactory delivery as vomeronasal receptors are often sensitive at femtomolar levels [35]. Consequently, there is likely a high selective premium (and possibly low cost) on rapid gene duplication and pheromone diversification in order to maximize the likelihood of stimulating any mating female in the population, improving reproductive success and fitness.

For more than 100 million years, plethodontid salamanders have utilized this highly novel system of rapidly evolving pheromones to improve male reproductive success akin to snake venoms in prey capture. Within this system, Plethodontid Modulating Factor (PMF) is a particularly tractable candidate to study how sexual selection may have promoted rapid evolution within a birth-and-death model. As a three-finger protein, PMF is already part of a rapidly evolving gene family, which has been well characterized in the context of snake venom TFPs. Given the hypothesis that male pheromones have duplicated to maximize the likelihood of stimulating any mating female, it would be expected that different isoforms are likely to have to different effects on female neurophysiology and behavior. Therefore, the aims of this dissertation were to characterize the molecular diversity, gene regulation, protein structure, and biological function of this hypervariable vertebrate pheromone, Plethodontid Modulating Factor, in response to 100 million years of female sexual selection. In chapter II, using a combination of DNA sequencing and mass spectral techniques, the diversity of PMF in *P. shermani* was characterized, models of molecular evolution examined, and novel qualities identified within the untranslated regions. In chapter III, mass spectroscopy and multidimensional NMR were used to determine the three dimensional structure of the



mostly highly expressed PMF isoform, possibly providing insight into PMF:receptor interactions. In chapter IV, female receptivity was measured for an individual isoform and the complete PMF complex, to assess how different isoforms might differentially affect female behavior. In chapter V, female vomeronasal responses were measured after treatment with different combinations of PMF isoforms. In chapter VI, mental gland development was studied to better understand the mechanisms underlying pheromone synthesis and gland hypertrophication. Finally, in chapter VII, the pheromone composition was characterized for *P. cinereus*, a species closely related to *P. shermani* with scratching delivery, to gain insights into how pheromone genes may have evolved across the transition from scratching to slapping delivery. The significance of these highly interdisciplinary studies are explained within each chapter, and further integrated in Chapter VIII.

## CHAPTER II - PROTEOMIC AND UTR ANALYSES OF A RAPIDLY EVOLVING HYPERVARIABLE FAMILY OF VERTEBRATE PHEROMONES

### **Chapter Overview**

During the annual mating season, the mental gland of male plethodontid salamanders diverts its protein synthesizing capacity to the production of courtship pheromones that increase female receptivity. Plethodontid Modulating Factor (PMF), a highly disulfide-bonded 7 kDa pheromone, shows unusual hypervariability with each male expressing >30 isoforms. Twenty-eight PMFs were purified and matched by proteomic analyses to cDNA sequences. In contrast to coding sequence hypervariability, the untranslated regions (UTRs) show extraordinary conservation, no predicted microRNA binding sites and an overlapping triplet polyadenylation signal. Full-length cDNA sequencing revealed 3 PMF gene classes containing sub-classes of clustered sequences that support  $\geq 13$  PMF gene duplications. The unusual phenomena of hypervariable coding regions embedded within extremely conserved UTRs is proposed to occur by a disjunctive evolutionary process. During the short courtship season, the UTRs are hypothesized to subsume and coordinate the transcriptional and translational regulatory mechanisms of the mental gland. PMF, as a secreted protein with limited metabolic feedback in the male, is under minimal mutational restraint and thus has experienced highly accelerated rates of evolution. Consequently, plethodontid salamanders may provide a unique model for furthering our understanding of the

selective forces that determine differential rates of gene duplication and evolution in protein families.

## **Introduction**

Relatively little is known about the evolution and coordinated regulation of genes in vertebrates that affect male and female mating behavior during annual reproductive cycles. However, for any trait to evolve through natural selection and propagate within a population, an adaptive advantage must exist. The classical paradigm for protein evolution is that selective pressures act on the corresponding gene's protein coding sequence (CDS) to maintain or modify structural and functional constraints [136]. In the modern context, additional forces that have been shown to affect protein evolutionary rates include, but are not limited to the following: protein expression level, gene product interactions within networks or metabolic pathways, protein dispensability, protein secretion, maintenance of protein folding fidelity, specificity of tissue or temporal expression during the life cycle, splicing-related constraints and gene duplication events [137, 138]. Many of these forces act in disproportionate yet integrated ways to drive protein diversification (positive selection) or to maintain structural elements (purifying selection). At the mRNA level, it is generally assumed that the untranslated regions (UTRs) that flank the CDS are under fewer functional constraints and are free to acquire mutations at a rate faster than that of the protein coding region [139]. In most examples of highly conserved UTRs (or more commonly, conserved cis-acting regulatory sequences contained within the UTRs), the protein coding region is more highly conserved [140-143]. In these examples, where protein function is often indispensable and sensitive to detrimental mutations, it is typically hypothesized that purifying selection or concerted evolution "protect" the genomic region from any and all mutagenesis [144]. However, multiple toxin systems have evolved to present with the unique phenomena of

multigene families with mRNAs containing variable protein coding sequences embedded within highly conserved UTRs [145-147]. In all cases, it is proposed that following gene duplication, mutations will arise and become fixed in the CDS at faster rates than the UTRs due to strong, pervasive positive selection. In the present study, we utilize a highly tractable salamander model to define and characterize a novel family of hypervariable vertebrate courtship pheromones that contain atypical, highly conserved UTRs in their mRNAs and correlate these cDNA sequences with expressed pheromone isoforms.

The red-legged salamander, *Plethodon shermani*, is a species indigenous to a mountainous region of western North Carolina, USA. For a brief ~2 month mating season, male *P. shermani* will develop a hypertrophied mental gland whose translational capacity is almost exclusively dedicated to pheromone production. During a unique courtship ritual, a male salamander delivers these pheromones to a female by “slapping” his mental gland on the female’s nares. The pheromones stimulate olfactory receptors in the female’s vomeronasal organ and increase her receptivity to mating [85, 87]. Chemical and proteomic analyses of mental gland extracts demonstrated that ~85% of the protein is composed of two major classes of pheromones [69]. A 22 kDa component termed Plethodontid Receptivity Factor (PRF) increases female receptivity and is related to IL-6 cytokines [83, 85, 93]. The second major component encompasses a family of 7 kDa proteins termed Plethodontid Modulating Factor (PMF) [90]. PMF is closely related to the highly disulfide-bonded three-finger protein (TFP) superfamily, which includes snake venom  $\alpha$ -neurotoxins and cytotoxins, xenoxins of *Xenopus* skin, and human Ly-6/uPAR proteins [102, 107, 148]. The “three finger” structure of TFPs arises from 8 highly conserved cysteine residues that form a constrained disulfide-bonded scaffold [108, 149,

150]. Through extensive duplication and mutation, the only shared consensus sequence between all TFPs is the motif CCXXXXCN occurring near the C-terminal end [110]. Although most vertebrate TFPs exist as single-gene copies [108, 120], venomous snakes often have several distinct TFP isoforms per species that result from gene duplication events [101, 151].

Analysis of the *P. shermani* pheromone extract by high performance liquid chromatography (HPLC) suggested >30 PMF isoforms were expressed in the mental gland [69]. Sequencing of ~300 clones from a *P. shermani* pheromone gland cDNA library revealed that >70% coded for PRF and PMF isoforms, and included 32 unique PMF derived protein sequences [89]. Initial analyses of the PMF expressed sequence tags (ESTs) indicated the CDSs were highly dissimilar while both the 5' and 3' UTRs were highly conserved. Palmer et al. [90] designed primers based on the conserved UTRs to obtain 27 additional mental gland PMF coding sequences with ~37% average amino acid dissimilarity. In a subsequent study comparing PMF cDNAs from multiple species in a larger phylogenetic context, the observed amino acid dissimilarity placed PMF under strong positive selection [96]. While this work serves as an important preliminary investigation, it is largely dependent on the assumption that each mRNA equates to an expressed pheromone that is inevitably delivered to a female. In *Saccharomyces*, the correlation coefficient between global mRNA to protein levels was found to only be ~0.5-0.6 [152]. A more proximate and pronounced example of this problem is PRF in *P. shermani*, where dozens of cDNAs have been identified at relatively uniform frequencies, yet three isoforms comprise >95% of the expressed PRF protein [69, 83, 94]. Structurally, PMF contains 8 disulfide-bonded cysteine residues that form a stable

molecular scaffold upon which sequence hypervariability may be mounted [96], but it is still unclear how much of this variation is functionally expressed and exposed to sexual selection in *P. shermani* females during courtship.

In order to more critically examine PMF's molecular and structural diversity, experiments were performed (a) to determine the extent of PMF molecular diversity at the nucleic acid and protein levels in a single animal or pool, (b) to obtain unbiased, full-length PMF sequences from the 5'-cap structure to polyA tail, and (c) to assess the evolutionary basis and significance of coding region hypervariability and UTR conservation.

## **Methods**

### Gland excision and protein extraction

*Plethodon shermani* males were collected during their breeding season from a single site in Macon Co., North Carolina, USA (35°10'48" N, 83°33'38" W). Males were anesthetized in a mixture of 7% ether in water. Ten mental glands were excised and immediately incubated in RNAlater (Ambion, Austin, TX) at 4°C overnight prior to storage at -20°C. Pheromones were extracted following the methods of Rollmann et al. [83]. Approximately 100 glands were excised and pheromones were induced for secretion by incubation with 0.8 mM acetylcholine chloride in Amphibian Ringer's Solution for ~60 minutes. Whole pheromone extract was centrifuged at 10,000  $\times$  g for 10 minutes, the supernatant collected, and the centrifugation repeated before storage of supernatant at -80°C. Methods and animal care were approved by Oregon State University's Institutional Animal Care and Use Committee (ACUP 3007 and 4053 to L.D. Houck).

### PMF cDNA amplification and sequencing

Ten *P. shermani* mental glands were homogenized and total RNA extracted using the RNeasy Mini Kit (Qiagen, Valencia, CA) according to the manufacturer's protocol. The mRNA fraction was subsequently purified using the PolyATtract mRNA Isolation System IV (Promega, Madison, WI). cDNA was synthesized from the purified mRNA using the Creator SMART cDNA Kit (Clontech, Palo Alto, CA), and subjected to routine PCR using primers (PMF5'F: 5'-TTAAGGCCGAGGCGGCCCTGGAATCCAGAATGAG-3'; PMF3'R: 5'-AATTGGCCATTACGGCCGCCATCTTGATTGCTTT-3') based on conserved regions



of the PMF 5' and 3' UTRs. Primers included engineered *Sfi*I restriction sites (Integrated DNA Technologies, Coralville, IA) for integration into the pDNR-LIB vector following restriction digest and ligation. Cloned PCR products were transformed into chemically competent TOP10 *E. coli* (Invitrogen, Carlsbad, CA) by the manufacturer's protocol, and plated on LB/agar/chloramphenicol. Clones were cultured in 150  $\mu$ L Circlegrow (MP Biomedicals, Irvine, CA) with 25  $\mu$ g/mL chloramphenicol, shaken for 48 hours at 37°C and aliquots taken for colony PCR using standard M13 primers. Products were purified using the QIAquick PCR Purification Kit (Qiagen, Valencia, CA) and supplied to the University of Louisville DNACore Lab for automated DNA sequencing. The methods were repeated for a single *P. shermani* mental gland to evaluate allelic variation.

Additionally, PMF cDNA products with full-length UTRs were generated, cloned, and sequenced using the GeneRacer RLM-RACE Kit (Invitrogen, Carlsbad, CA) following the manufacturer's protocol. In these reactions, either PMF5'F or PMF3'R was used in PCR with an adapter primer to allow for independent amplification of full-length 3' and 5' UTRs as well as less biased amplification of CDSs.

#### Processing and analysis of sequences

Programs used for initial sequence processing are part of the DNASTAR package (Lasergene Version 7.1; DNASTAR, Madison, WI). SeqMan Pro was used to filter low quality sequences and assemble contigs. Sequences verified as PMF were processed using EditSeq for translation of open reading frames and prediction of protein molecular weight. After removal of repeat translations, sequences were aligned and statistics calculated using MegAlign with the ClustalW algorithm and Gonnet 250 protein weight matrix. Unique sequences were compiled in FASTA format to serve as a database for

peptide mass fingerprinting, and were deposited in NCBI GenBank (Genbank accession numbers JF274283-JF274477). Gene trees for all PMF sequences were constructed using MEGA 5 [153] for 500 bootstraps by Nearest-Neighbor Joining for amino acid substitutions under a Poisson model with uniform rates. MEGA 5 was also used for calculating mean nucleotide and amino acid distances between PMF clusters [154]. Site-specific codon substitution models with variable dN/dS rates ( $\omega$ ) estimated by maximum likelihood were fit using PAML 4 [155]. Models included M1a (nearly-neutral; 2 parameters:  $0 < \omega_0 < 1$ ,  $\omega_1 = 1$ ), M2a (selection; 3 parameters:  $0 < \omega_0 < 1$ ,  $\omega_1 = 1$ ,  $\omega_2 \geq 1$ ), M7 (beta; 2 parameters: p and q for  $\beta$ -distribution with 10  $\omega$ -values in range [0,1]), and M8 (beta and  $\omega$ ; 3 parameters: p and q as in M7, and  $\omega \geq 1$ ), with tests for positive selection performed by using likelihood ratio tests to compare M2a vs. M1a and M8 vs. M7 (with  $2\ln L$  compared to  $\chi^2$ ) [155-158].

#### Purification of PMF isoforms

Mental gland pheromone extracts were purified by strong anion-exchange-HPLC (Mono Q; Pharmacia, Piscataway, NJ), reverse phase-HPLC (RP-HPLC) (C-18; Grace Davison Discovery Sciences, Deerfield, IL), and size exclusion chromatography (G-75 Superfine; Pharmacia, Piscataway, NJ). All chromatographic separations were accomplished on a 2695 Alliance HPLC System equipped with a 2487 dual wavelength absorbance detector and Empower software (Waters Division, Milford, MA). The Mono Q column (0.5 x 5.5 cm) was eluted at 1 ml/min with a NaCl gradient in 50 mM Tris/HCl buffer, pH 8.0 (mixed gradient: 5 mM NaCl/min for 30 min, 10 mM NaCl/min for 15 min, 20 mM NaCl/min for 10 min). The C-18 column (0.5 x 15 cm) was eluted with an acetonitrile gradient in 0.1% TFA. The G-75 column (1.6 x 15.5 cm) was eluted at ~10

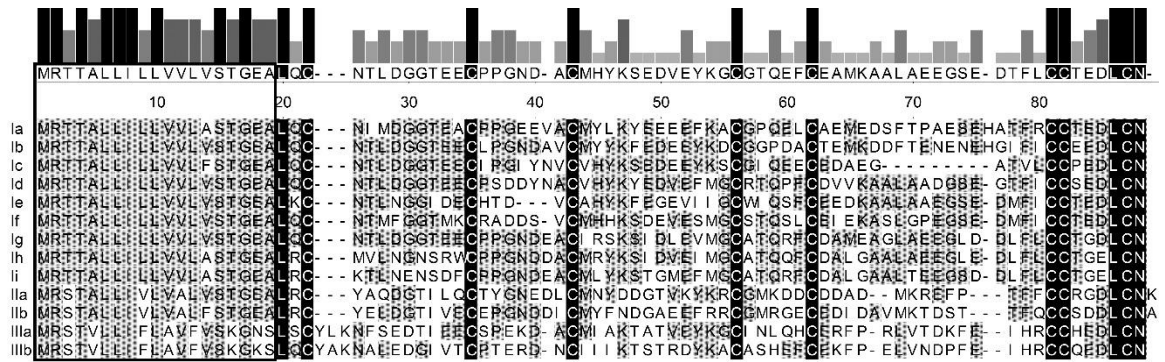
mL/hr with 0.5 X phosphate-buffered saline (PBS). Based on previous data from Feldhoff et al. [69], the PMF fractions were collected after separation by strong anion-exchange-HPLC. These fractions were pooled, concentrated using YM-3 Centriprep (Millipore, Billerica, MA), further purified by size exclusion chromatography, re-concentrated, and re-chromatographed by strong anion-exchange-HPLC with individual peaks collected for final purification by RP-HPLC.

#### Mass spectral analysis

Picomole quantities of each PMF isoform were provided to the University of Louisville Biomolecular Mass Spectrometry Core Laboratory. Intact protein mass was determined by electrospray ionization mass spectrometry (ESI-MS) using a Q-TOF API-US (Waters Division, Milford, MA), while tryptic fragment fingerprints were acquired by either matrix-assisted laser desorption ionization followed by time-of-flight (MALDI-TOF) using a TOF Spec 2E (Waters Division, Milford, MA) or liquid chromatography tandem mass spectroscopy (LC/MS-MS) using a LTQ Orbitrap XL (Thermo Scientific, Waltham, MA). SEQUEST software (Thermo Scientific, Waltham, MA) was used for all peptide fingerprinting analysis with a custom database constructed from plethodontid mental gland cDNA sequences. The precise masses of intact PMFs and those for tryptic peptides obtained after reduction and alkylation were matched to theoretical masses (peptide mass fingerprinting) derived from DNA sequences for PMFs. Predicted intact masses were adjusted by 8 Da to account for the protons lost in the 4 disulfide bonds.

## Results

Initial RT-PCR for PMF using optimized primers for each UTR was performed using a pool of mRNA from 10 *P. shermani* male mental glands. Sequencing of 140 clones yielded 65 derived protein sequences. After sequence alignment, 13 primary clusters were identified based on biochemical properties in the derived amino acid sequences (Figures 4, 5). Three major variations of the secretion signal peptide were observed, and the 13 clusters of sequences were grouped into larger classes (termed Class I, II, III). Classes I, II and III contained 9, 2 and 2 sequence clusters, respectively. In this organizational scheme, a cluster is equivalent to a sub-class. When a gene tree was constructed for these sequences, each sequence cluster (except Id) fell into a single monophyletic clade (Figure 6). The average amino acid dissimilarity between all unique PMF sequences was calculated to be ~69% (including the secretion signal). While significantly higher than the value calculated by Palmer et al. [90], the difference can be attributed to two reasons. First, the primers used in the earlier study heavily biased the PCR products toward Class I PMFs, losing the diversity added from Classes II and III (Table 1). Second, there was not sufficient coverage to detect that the extremely high sequence diversity between clusters masks low within-cluster nucleotide dissimilarity (<10% for any cluster). Within each cluster, the majority of the detected amino acid variation within a cluster can be attributed to one or two single nucleotide polymorphisms (SNPs) relative to the cluster consensus sequence. To examine the extent of allelic variation in an individual male, PMF RT-PCR products were generated and sequenced from the mRNA of a single *P. shermani* mental gland.



**Figure 4. Alignment of PMF cluster consensus sequences.**

Each sequence is labelled with its class and subclass abbreviation, with the first line being a group consensus sequence and a histogram showing identity consensus strength. Completely conserved residues in the mature peptide highlighted in black with white font, and additional residues matching the consensus are highlighted in grey. The black box surrounding the first 19 residues denotes the signal peptide.





**Figure 5. Complete alignment of PMF cDNA translations.**

Alignment of derived protein sequences from the 99 unique PMF cDNAs, including a histogram of consensus strength and residues not matching the consensus highlighted in pink.



**Table 1. Mean distance between PMF clusters.** Nucleotide (upper right) and amino acid substitutions (lower left) per site from averaging over all sequence pairs between PMF classes are shown. Analyses for amino acid substitutions were conducted using the JTT matrix-based model.

	la	lb	lc	ld	le	lf	lg	lh	li	lia	lib	llla	lllb
la		0.209	0.218	0.268	0.347	0.315	0.313	0.378	0.387	0.603	0.496	0.777	0.721
lb	0.484		0.183	0.252	0.303	0.329	0.305	0.375	0.359	0.556	0.516	0.725	0.684
lc	0.535	0.447		0.161	0.231	0.242	0.257	0.326	0.316	0.441	0.416	0.712	0.691
ld	0.667	0.607	0.415		0.204	0.200	0.217	0.279	0.267	0.547	0.554	0.792	0.770
le	0.798	0.668	0.541	0.455		0.178	0.250	0.283	0.281	0.677	0.603	0.964	0.857
lf	0.699	0.801	0.649	0.509	0.413		0.249	0.289	0.297	0.657	0.640	0.935	0.799
lg	0.685	0.662	0.593	0.507	0.591	0.576		0.125	0.124	0.665	0.598	0.794	0.735
lh	0.833	0.972	0.927	0.680	0.650	0.673	0.310		0.088	0.655	0.641	0.826	0.832
li	0.843	0.832	0.852	0.665	0.633	0.689	0.335	0.221		0.680	0.621	0.761	0.753
lia	1.116	1.013	0.837	1.064	1.267	1.206	1.202	1.115	1.200		0.209	0.731	0.801
lib	0.941	0.942	0.721	0.985	1.041	1.143	1.035	1.038	1.040	0.492		0.696	0.769
llla	1.269	1.185	1.189	1.286	1.467	1.399	1.223	1.281	1.216	1.324	1.348		0.223
lllb	1.163	1.191	1.137	1.234	1.319	1.320	1.180	1.344	1.243	1.428	1.411	0.504	

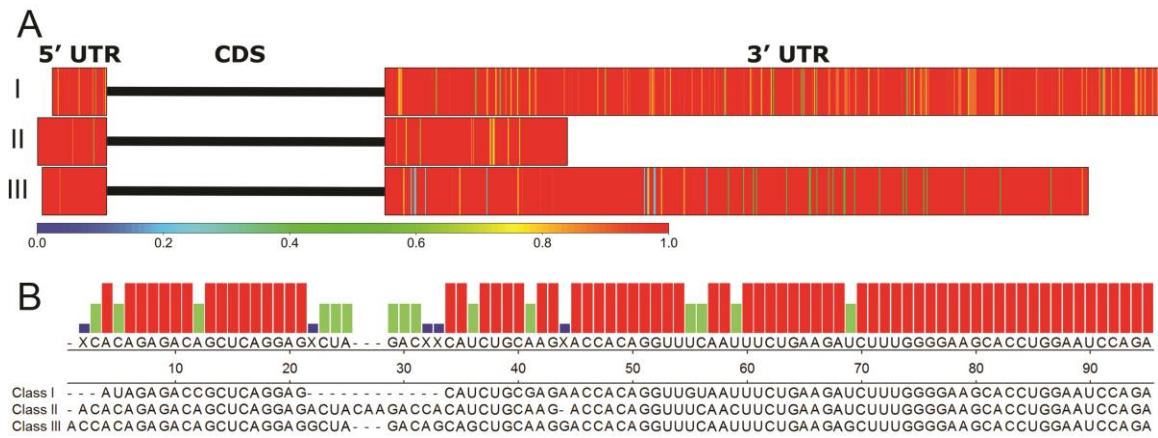
Nucleotide Distance

Amino Acid Distance



Sequencing of 96 clones yielded nucleotide sequences coding for 30 unique predicted proteins; surprisingly, 14 had not previously been detected, but all derived amino acid sequences closely matched a cluster consensus sequence. These data – obtained from an individual male and a 10-animal pool – indicate that allelic variation, in the absence of extensive gene duplication, cannot fully explain the extensive observed heterogeneity in PMF.

To further characterize the previously unsequenced parts of the 5' and 3' UTRs, a 5' cap to 3' polyA-based RT-PCR method [159] was used to obtain cDNA copies representing full-length mRNAs. While the initial ESTs suggested conserved UTRs among all PMF isoforms, it is now reported that each of the 3 gene classes identified above, based on signal peptide sequences, also contains a unique highly conserved 5' and 3' UTR pair that is >98% identical between any two sequences within each class (Figure 7A) (Genbank accession numbers JF274382- JF274477). In general, the 5' UTR is highly conserved among all 3 classes, with the major difference being unique insertions ~20 bp from the transcription start site in classes II and III (Figure 7B). In all three classes, the first 2-3 bp on the 5' end of the 5' UTR were inconsistently detected which may relate to transcriptional regulation. In contrast to the 5' UTR, the 3' UTR was unique in both length and sequence with no significant alignment observable among the three classes. Within each class, however, the 3' UTR was ~98% conserved similar to the 5' UTR. All 3' UTRs were very AU-rich, with Classes I and III containing the near-ubiquitous 6 nucleotide polyadenylation signal –AAUAAA- (Class II utilizes a less common –AUUAAA- motif) (Figure 7A and Table 2).



**Figure 7. Summary of PMF UTR analysis by gene class.**

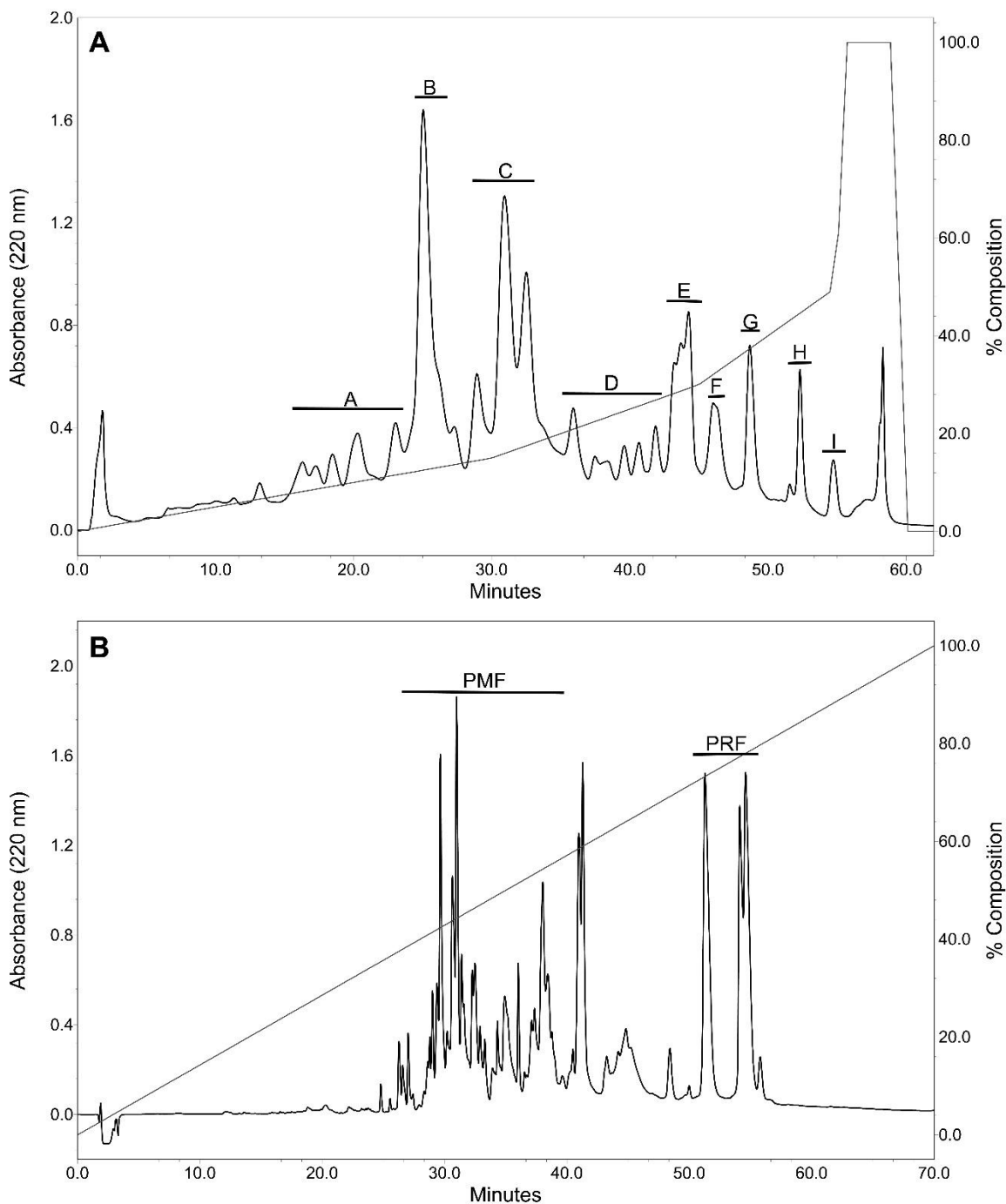
A) Schematic of PMF mRNAs to scale with the UTR consensus strength within each class marked by color. B) Alignment of the 5' UTR consensus sequences per class with the top line representing the group consensus with a histogram of consensus strength.

**Table 2. Summary of PMF 3' UTRs by class.** The consensus sequence, typical length, average AU content of each PMF 3' UTR sorted by class, with known polyadenylation signals denoted by bold and underline.

	Length (bp)	AU-content	Consensus Sequence
<b>Class I</b>	710	70%	UCACACACUCUUCAGGAGAUGCUAUUGAGGAAUCAC ACUGGCAUCCAAUACAGACCAAGUAAACAAAAGCA AUCAAGAUGGCUACCAGAGAAAACAAGGAGCAAAGC UUCUUUGACGUAAAGUCAUGCUUCUCUAUUUAUGUCU UUGAUACAGUAAACAUAAAUUGAUGUUUGCUAUAAAA AUUAAAUCCUUCUACAUGUCUCCUGAUGACUAUAG UUCUGCCUAGUUUUACUUUUCCAAAAAUUCUAUUA AUCCAUAUUUUUUCUACUGGGUUAGUGUGAUGGGA AGGAAUACUCUAAAAUCGUUUUUUGUUUCAUGA GAUAUAUUUGGAAUGAUCUAAAAUUAUGAUGCCAC CCAAGUUGUCACUUUUUCUGAAAAUUCUGUAUCUUU UAUUUAUGUAUCUGAAAAUGUGGAAGCAGAAAAUUG UAUUACUGUAUAUAUAUCUUGGAUUAAUAACAA ACAUGUAAAUUUCUGUAAAAAAUUCGACUGUGCAA UCUUUUAUCCUUGAUGUUCUCAUUUUUCUUUCCAAA CAAUCUGUUUCUGAAUUUCCCAUGUAAUAACUGAA UUUCAACAUCAUUGUGUUAAACUGGAAGAUGAUU UUUGUUUGUUCUACAAUUCUCUAUAACUCAUCUCUG GAAGUCUUGUUUGAUGGUGCUUUUACA <u><b>AAUAAA</b></u> <u><b>AAUAAA</b></u> AGACCAUUUAUAUCUG
<b>Class II</b>	190	62%	UACUGUCUUUGGCAGAGGCCAAUAAAGAAUCGAGAU GACCGCCUAAACAGAGACCAAGUAAACAAAAGCAAUC AAGAUGGCCACCAGAGAAAACAAGAACAGGAAUA UGCCUUAAGAGUGAAUUCAGGCCUCUAUGUUUUGU CUCUGAUAACAUGAAUGGCUUCCUAUAACA <u><b>AUUAAA</b></u> UCCUUCUAUAUGU
<b>Class III</b>	650 or 300	68%	GAUGGCCGGCCAAUACCUUCCAAGAACCAAAAAGCA AGAAUAUUGGCGACCAGAGAAAUGAAAGCCCAGGAG ACCGCUUCUUUAGAGUGAAGGCACGCCUCUGCAUU GUGCCUCGGAUUACACGAAUUGUUUGGUCAGAAAAU GUACUUCCUUGCAUAGGUUUUCUGAUGACUGCAU UUAGAAUUUUUAAAUUUCUGAAAAGAAAAGUUACC AAUGGGUUCUUUACUGGGUAAUGUGUAGGGUCAUG CAUUGUGUUGAUUAUUUUUGUAUCCA <u><b>AAUAAA</b></u> AU AAUUGAAUUUAUGUGAAAUAUGACUUGCUUAUAGU UAUCACUUUGGAGGAAAUCCUGUAAGUUUUUGUGUU UUACAAGAAUAUGUGGAAAGGACGAGAAUGUAUUAG UGUAAUUACAUAUUUUUGUUUUUCUUUAACA <u><b>AAUAAA</b></u> ACA UUUACGUUUUUUAAGAAAGUAGAAUCCUUUUUCCUC AAUUUUUCUAGAAUUGUAUUUCAAAUUUUUCUGUUC UUAAAUUCCCUUGGAAUAUUUGAAUUUCUUUAUU UAUGUGUUUCUCUGGAAGAAAUAGUUUCAUUUUUGC ACCAUUCUCUGCAGCUUUUCUCUGGAUUCUUCU UGCUGUUGCUUUUACA <u><b>AAUAAA</b></u> UGAAUGGCAUUAAA AAUAAA

Furthermore, the Class I 3' UTR contained a rare 14 nucleotide polyadenylation sequence AAUAAAUAAAUAAA, functionally equivalent to a triplet overlap (nucleotides 1-6; 5-10; 9-14) of the consensus hexanucleotide polyadenylation signal (AAUAAA). Because polyadenylation is a stochastic event, this repetitive sequence could function to increase the probability of successful polyadenylation and ensure maturation of Class I PMF mRNAs and/or explain the enhanced Class I PMF expression that we have observed. One also cannot rule out this rare polyadenylation signal playing a role in cytoplasmic deadenylation or polyadenylation processes since there are several polyU-rich regions within the 3' UTR [160]. In the Class III 3' UTR, there are two identified polyadenylation signals that permit a full-length ~670 bp form and a truncated ~300 bp variant. However, the frequency of truncation may not be uniform across the Class III PMFs, with ~85% of the IIIa 3' UTR clones having truncated 3' UTRs compared to only ~25% of the IIIb 3' UTR clones. During the analysis of 108 full-length cDNAs, we identified an additional 20 uniquely derived protein sequences. Cumulatively, 99 unique PMF predicted protein sequences from the *P. shermani* mental gland were identified (Genbank accession numbers JF274283-JF274381).

To assess the extent that the genetic variation was translated into mature pheromone components, proteomic analyses were performed on pheromones purified from *P. shermani* mental glands. In a previous study [69], the majority of PMF isoforms were found in the D to I fractions when whole pheromone extract was separated by strong anion-exchange-HPLC (Figure 8A).



**Figure 8. Chromatographic separation of *P. shermani* pheromones.**

(A) Separation of pheromones by strong anion-exchange-HPLC. Fractions were termed A to I as displayed, with the B to C fractions primarily containing PRF and fractions D to I primarily containing PMF. Individual proteins in each fraction are identified by the fraction letter and a unique number. B) Separation of pheromones by reverse phase HPLC, with the principal PRF and PMF fractions labelled.

Through a combination of size exclusion-chromatography, strong anion-exchange-HPLC (Figure 8A) and reverse phase-HPLC (Figure 8B), 28 unique PMF isoforms were purified for subsequent analysis by two independent mass spectral (MS) techniques (Table 3). When compared to the cDNA database of 99 derived PMF protein sequences, 26 isoforms were matched to predicted sequences by at least one MS technique, including 12 isoforms that were fully characterized using both MS techniques. Ten of the 13 clusters contained 1-3 expressed isoforms (Id contained 7 isoforms, with no isoforms in IIa or IIIa), with the majority being found in Class I. In the selection models constructed using PAML, there was significant evidence for strong positive selection acting on multiple amino acid residues across the *P. shermani* PMF gene tree ( $\omega_2 = 4.03$  in M2a; Table 4). Multiple isoforms found within the same cluster were predicted to originate from more separate lineages (Figure 6), suggesting both diversification of function and a molecular selection to only translate the more distinct pheromone products. In a parallel analysis, PAML models for only translated PMF sequences predict similar substitution rates to those of all *P. shermani* PMFs for positively selected sites ( $\omega_2 = 3.99$  in M2a; Table 4). The calculated rates for both analyses are faster than those predicted by Palmer et al. [96] for the larger clade within which *P. shermani* is found (*P. glutinosus* complex;  $\omega_2 = 2.75$  in M2a). Substitution rates and models of molecular evolution could not be determined for individual clusters due to some clusters containing as few as 2 sequences. Importantly, four PMF isoforms (A1 = IIIb03; G = Ic03; H = Ib04; I = Id08) appear to be expressed at much higher levels, and together comprise >33% of the total PMF protein.

**Table 3. Summary of mass spectroscopy data.** A summary of the mass spectral data correlating purified proteins to cDNA sequences, including intact protein mass, number of tryptic peptides, and percent sequence coverage for the tryptic peptides (an asterisk indicates peptide sequencing by LC/MS-MS, and all other samples utilized MALDI-TOF). Sequences conclusively defined by both intact protein mass and >50% sequence coverage are marked in bold with gray background, while proteins with partial characterization using either technique contain a gray background and a white background indicates neither criterion was met. Partial characterization likely indicates proper class/cluster identification, but not necessarily sequence.

Purified PMF	Experimental Mass (Da)	Theoretical Mass (Da)	Sequence Matched	# of Matching Peptides	% Sequence Coverage
<b>A1</b>	<b>7478</b>	<b>7479</b>	<b>PshePMF IIIb02</b>	<b>5</b>	<b>77</b>
D1	n/a	9767	PshePMF IIIa02	2	40
D2	n/a	6830	PshePMF Ie01	3	68
D3	n/a	7269	PshePMF Ia03	2	62
D4	n/a	6998	PshePMF Ii02	2	71
D5	n/a	6295	PshePMF Id01	1	46
D6	n/a	6829	PshePMF Ie02	2	66
D7	n/a	7211	PshePMF Iib04	2	50
<b>D8*</b>	<b>7317</b>	<b>7314</b>	<b>PshePMF Id02</b>	<b>7</b>	<b>97</b>
D9*	7331	7314	PshePMF Id02	6	88
<b>D10*</b>	<b>6979</b>	<b>6976</b>	<b>PshePMF If01</b>	<b>5</b>	<b>97</b>
<b>D11*</b>	<b>7020</b>	<b>7017</b>	<b>PshePMF Id13</b>	<b>6</b>	<b>100</b>
D12	6857	6873	PshePMF Id11	2	81
<b>E0</b>	<b>6829</b>	<b>6829</b>	<b>PshePMF Ie02</b>	<b>2</b>	<b>63</b>
<b>E1a</b>	<b>6872</b>	<b>6873</b>	<b>PshePMF Ig08</b>	<b>3</b>	<b>97</b>
<b>E1b</b>	<b>7146</b>	<b>7147</b>	<b>PshePMF Id06</b>	<b>4</b>	<b>100</b>
<b>E2</b>	<b>7353</b>	<b>7353</b>	<b>PshePMF Id03</b>	<b>5</b>	<b>95</b>
<b>E3</b>	<b>7048</b>	<b>7048</b>	<b>PshePMF Ii06</b>	<b>3</b>	<b>100</b>
E4	6912	7165	PshePMF Iib07	2	70
<b>E5</b>	<b>6915</b>	<b>6916</b>	<b>PshePMF Ii10</b>	<b>5</b>	<b>100</b>
E6	6994	6891	PshePMF Ig07	2	55
E7	6948	6891	PshePMF Ig10	2	55
E8	n/a	7165	PshePMF Iib07	2	70
E9	n/a	6253	PshePMF Ic04	3	54
F1	7234	7235	PshePMF Ih06	3	33
<b>G*</b>	<b>6256</b>	<b>6257</b>	<b>PshePMF Ic03</b>	<b>3</b>	<b>100</b>
<b>H</b>	<b>7479</b>	<b>7479</b>	<b>PshePMF Ib04</b>	<b>4</b>	<b>100</b>
I	6943	6937	PshePMF Id08	2	52

**Table 4. Models of positive selection for *P. shermani* PMFs.** A summary of PAML results for site models of PMF evolution using tests for positive selection in both the complete *P. shermani* PMF gene tree and isoforms exclusively identified through proteomic analysis with models M2a vs. M1a and M8 vs M7. In both the M1a/M2a and M7/M8 model sets, inclusion of a free parameter of  $\omega \geq 0$  was statistically significant, and parameter estimates were similar in all models. \*\* denotes  $p < 10^{-10}$ .

<i>Model</i>	<i>dN/dS</i>	<i>logL</i>	<i>Estimates of parameters</i>	<i>2ΔL</i>
All PMF Sequences (n = 99)				
M1a (nearly-neutral)	0.8503	-4919.67	$p_0 = 0.15399, \omega_0 = 0.02767$ $p_1 = 0.84601, \omega_1 = 1.00000$	
M2a (selection)	2.68	-4796.61	$p_0 = 0.12902, \omega_0 = 0.00200$ $p_1 = 0.27334, \omega_1 = 1.00000$ $p_2 = 0.59764, \omega_2 = 4.03091$	<b>246.12**</b>
M7 (beta)	0.7312	-4904.68	$p = 0.18799, q = 0.06909$	
M8 (beta+ $\omega$ )	2.50	-4793.77	$p_1 = 0.58733, \omega_1 = 3.82353$ $p_0 = 0.41267$ $p = 0.01906, q = 0.01069$	<b>221.82**</b>
Expressed PMF Sequences (n = 23)				
M1a (nearly-neutral)	0.82	-2634.05	$p_0 = 0.19423, \omega_0 = 0.04906$ $p_1 = 0.80577, \omega_1 = 1.00000$	
M2a (selection)	2.66	-2587.36	$p_0 = 0.16064, \omega_0 = 0.03892$ $p_1 = 0.23147, \omega_1 = 1.00000$ $p_2 = 0.60789, \omega_2 = 3.99185$	<b>93.38**</b>
M7 (beta)	0.62	-2636.35	$p = 0.03280, q = 0.01870$	
M8 (beta+ $\omega$ )	2.69	-2586.85	$p_1 = 0.62655, \omega_1 = 3.98878$ $p_0 = 0.37345$ $p = 0.02251, q = 0.01949$	<b>99.0**</b>



However, cDNA corresponding to mRNAs for these abundant protein isoforms were not found at significantly higher abundance in sequenced clones, suggesting disproportionate translation rates of PMF mRNAs. With the exception of disulfide bonds, no other post-translational modifications were observed in any PMF isoform.

## Discussion

As previously reported, PMF is a hypervariable vertebrate courtship pheromone that shows extensive molecular variability. For the first time, this molecular variability has been ordered into 3 gene classes with 13 sequence clusters based on coding sequence similarity and used to characterize purified proteins. Since the PMF isoforms were purified from a pooled sample extracted from ~100 mental glands, the individual purified PMF proteins likely represent those most highly expressed in the population and responsible for major bioactivity. Although specific functions are not known, typically only 1-3 PMFs from each cluster are highly expressed, thus each cluster may be fulfilling a unique biological role (e.g. receptor-, cell- or tissue-specific targeting), with the most abundant forms likely conferring the highest selective advantage. Interestingly, recent behavioral data also suggest that PMF isoforms act synergistically such that only a complex but relatively specific blend of isoforms can increase female receptivity (Wilburn and Houck, unpublished data). This is also in light of *P. shermani* PMFs showing faster substitution rates for sites under positive selection than those averaged across the sampled taxa in Palmer et al. [96]. Previously, the SNP variants were proposed to be alleles or rare gene duplications [90]. Although the exact number of PMF genes remains unknown, our data support many gene duplications and their retention as functional genes as the cause of male PMF hypervariability. While alternative splicing is possible, our preliminary data (Wilburn and R Feldhoff, unpublished) suggest PMF shares a similar gene structure to other TFP members with 3 exons/2 introns [101, 161]. In all cases, one exon includes nearly all of the signal peptide while the other two exons encode 3 and 5 cysteine residues, respectively. Unfortunately, efforts to fully sequence all

of the PMF genes has been impeded by a number of factors: (1) the *P. shermani* genome is exceptionally large (30 Gb or ~10x the size of the human genome) [162], (2) salamander introns are typically >10x larger than those in vertebrate orthologs [163], (3) there exist several processed PMF pseudogenes that compete for primer binding during PCR, and (4) the number of duplications within a cluster make it difficult to target a single locus. Current data, however, supports at least one PMF gene for each of the 13 (consensus) sequence clusters shown in Figure 4. In line with other duplicated TFPs, we postulate that multiple tandem gene duplication events have facilitated a rapid expansion of PMF's functional breadth in modulating female behavior through several distinct isoform classes – each with a unique biological target. The extraordinary conservation of the 5'- and 3'-UTRs suggests important functions for regulating PMF expression. Preliminary sequencing of developing mental gland transcriptomes reveals that synthesis of pheromone mRNAs precedes efficient translation, further alluding to post-transcriptional mechanisms as being the major process by which PMF expression is regulated (Wilburn and R Feldhoff, unpublished data). Both RNA binding proteins and microRNAs (miRNAs) have been recognized as major regulators of mRNA stability and translational efficiency [143, 164-166]. Because longer 3' UTRs often contain additional regulatory targets that reduce expression (e.g. miRNA binding sites) [167], it is noteworthy that the most abundant protein isoforms are those associated with long 3' UTRs in the class I and IIIb PMFs (Figure 6A). Additionally, no predicted miRNA targets were found within the PMF UTRs [168]; while available databases contain primarily mammalian miRNAs, it seems unlikely that amphibian miRNAs are directly targeting PMF mRNAs for degradation given the extreme abundance of the pheromone

transcripts and their efficient translation in the mental gland. While there is no consensus method for predicting targets of RNA binding proteins, the extensive AU-content in the 3' UTR could serve as a platform for AU-binding proteins, such as the HuR/ELAV proteins known to influence mRNA stability and translation [139, 141]. In gametogenic cells, the DAZL family of RNA-BPs promotes translational activation by directly recruiting the poly(A) binding protein to deadenylated mRNAs [169-171]. Interestingly, the Class I PMF 3' UTR contains the U-rich predicted consensus DAZL-binding sequence [170, 172]; however, RT-PCR experiments to amplify DAZL from mental gland cDNA have been unsuccessful. Given the mental gland's unique developmental and functional characteristics, regulatory proteins at the translational and/or transcriptional levels are likely involved in shifting the gland, prior to the annual mating season, from development and differentiation to pheromone production. Importantly, translational regulation of this type would add another layer of plasticity for PMF as a chemical signal by controlling the selective expression of different isoforms using regulatory proteins that would now be under indirect sexual selection. Targeted research into the specific RNA-BPs that may regulate these mechanisms is required.

The high CDS dissimilarity in the context of highly conserved UTRs represents an apparent contradiction to the dogma by which protein sequences and their corresponding genes are assumed to evolve in response to selective pressure. Across dozens of sampled plethodontid species, in the mature 7 kDa pheromone isoforms, purifying selection has maintained PMF's structurally constrained disulfide bonded scaffold (8 Cys; 4 disulfide bonds), along with the strictly conserved N-terminal Leu and C-terminal tripeptide –Leu/Val-Cys-Asn. These conserved residues, that are similarly

spaced in each cluster, also support gene duplication events as the basis for the high number of expressed PMF isoforms in *P. shermani* and closely related species [96]. Thus, rather than being uniform or random, for the first time the diversity of PMF is more accurately grouped into 13 sequence clusters within 3 major gene classes containing highly conserved UTR pairs. Expression patterns have evolved such that only the most divergent PMFs are highly translated, maximizing the diversity of the pheromone mixture. While the preceding comments emphasize amino acid variation, it is important to note that most of the variation within a cluster is produced by 1-2 SNPs. Therefore, within a cluster, any two PMF nucleotide sequences are >99% identical. Interestingly, the similarity of the UTRs within-cluster and within-class do not significantly deviate (~98% for both). Therefore, it is proposed that while purifying selection across the members of each of the 3 classes maintains the UTRs, after CDS neofunctionalization, equally strong selection conserves the protein coding sequence and the core structural characteristics within each cluster. Thus, we propose that different modes of evolution are acting independently on the coding regions and regulatory UTRs of PMF genes. For the protein coding region, a plethora of isoforms have been produced through multiple gene duplications that rapidly diverged under positive selection to assume different biological roles, resulting in the sequence clusters now maintained through purifying sexual selection. It is noteworthy that rapid evolution of tandemly duplicated genes is common in some gene families [173, 174]. In Hessian flies, a family of salivary protein-coding genes has duplicated and undergone positive selection to generate 7 distinct genes within a 15 kb genomic region [147]. Similarly, in several species of snakes, multiple phospholipase A<sub>2</sub> genes have duplicated to form multigene families with 2-4 copies each.

In these examples, positive selection drives diversification of the CDS while the UTRs are under neutral evolution [146]. Because of the lack of genomic sequences, we cannot directly determine if the PMF UTRs are under neutral or purifying selection. However, since the 3' UTR between the three gene classes could not be easily aligned, it is hypothesized that there must have been significant mutation acquisition (equal to or greater than that in the CDS) shortly after the earliest gene duplication events that yielded the three classes, and thus the conservation observed now is due to purifying selection. With there being potentially dozens of PMF genes in the *P. shermani* genome, secreted courtship pheromones, which are not involved in attraction and only serve to increase female receptivity, may have even fewer mutational constraints such that this protein family could reflect the maximum rates of vertebrate gene duplication and functional evolution.

We believe this external mode of trait selection is analogous to that for venomous snakes and cone snails whose toxins must instantly paralyze a variety of prey. As an essential nutritional source, prey immobilization acts as a natural form of bioassay to provide a selective force for adaptive toxin evolution. In our model, the female *P. shermani* salamander population may possess a variety of cell or tissue-specific receptors whose activation enhances mating success. Over millions of years, random females likely represent an extreme selective force to drive accelerated gene duplication and protein diversification in reproductive males. However, this unique model system has another major component: the highly conserved UTRs, common to all PMFs, which likely have an important regulatory function. While presumably under the same selective pressures, the evolutionary decoupling of the CDS and UTRs explains how the potential of PMF as

a pheromone signal can be rapidly expanded and exploited to affect mating success. By using multiple genes with highly conserved regulatory UTRs, expression of dozens of PMF genes during the short mating season (that follows the annual mental gland development and differentiation process) is uniformly regulated in response to external and physiological cues.

## Conclusions

PMF is a hypervariable vertebrate courtship pheromone that modulates male reproductive success through an arsenal of distinct protein isoforms which have resulted from numerous gene duplication events in three separate gene families. While duplication of small, highly disulfide-bonded proteins have been utilized in multiple toxin systems [115, 175], PMF as a pheromone signal is unique in vertebrates for both the extent of molecular variability and the extreme conservation of the UTRs. Within the three major classes, multiple sequence clusters were identified with translated PMF products likely performing unique biological roles. The most abundant isoforms may represent the selective optima in the population at this time. The novel, highly conserved, AU-rich 5'- and 3'-UTRs are suggested to serve as platforms for RNA binding proteins that permit PMF to usurp the pheromone gland's translational machinery. The unique nature of the pheromone mRNAs suggests that, over millions of years, independent evolutionary forces have driven a stabilization of the regulatory UTRs while promoting, through gene duplication and mutation, expansion of PMF's functional roles during courtship. Further studies of this unusual evolutionary alliance between highly conserved UTRs and gene duplication events, for proteins that play a significant role in enhancing mating success, may provide new insights into transcriptional and translational regulation in vertebrates.



CHAPTER III - STRUCTURAL INSIGHTS INTO THE EVOLUTION OF A  
SEXY PROTEIN: NOVEL TOPOLOGY AND RESTRICTED BACKBONE  
FLEXIBILITY IN A HYPERVARIABLE PHEROMONE FROM  
THE RED-LEGGED SALAMANDER, PLETHODON SHERMANI

**Chapter Overview**

In response to pervasive sexual selection, protein sex pheromones often display rapid mutation and accelerated evolution of corresponding gene sequences. For proteins, the general dogma is that structure is maintained even as sequence or function may rapidly change. This phenomenon is well exemplified by the three-finger protein (TFP) superfamily: a diverse class of vertebrate proteins co-opted for many biological functions – such as components of snake venoms, regulators of the complement system, and coordinators of amphibian limb regeneration. All of the >200 structurally characterized TFPs adopt the namesake “three-finger” topology. In male red-legged salamanders, the TFP pheromone Plethodontid Modulating Factor (PMF) is a hypervariable protein such that, through extensive gene duplication and pervasive sexual selection, individual male salamanders express more than 30 unique isoforms. However, it remained unclear how this accelerated evolution affected the protein structure of PMF. Using LC/MS-MS and multidimensional NMR, we report the 3D structure of the most abundant PMF isoform, PMF-G. The high resolution structural ensemble revealed a highly modified TFP structure, including a unique disulfide bonding pattern and loss of secondary structure,

that define a novel protein topology with greater backbone flexibility in the third peptide finger. Sequence comparison, models of molecular evolution, and homology modeling together support that this flexible third finger is the most rapidly evolving segment of PMF. Combined with PMF sequence hypervariability, this structural flexibility may enhance the plasticity of PMF as a chemical signal by permitting potentially thousands of structural conformers. We propose that the flexible third finger plays a critical role in PMF:receptor interactions. As female receptors co-evolve, this flexibility may allow PMF to still bind its receptor(s) without the need for complementary mutations. Consequently, this unique adaptation may establish new paradigms for how receptor:ligand pairs co-evolve, in particular with respect to sexual conflict.

## Introduction

Pheromone signaling is an essential means of communication for many animals to acquire information for a vast array of qualities on other individuals, including species, gender, reproductive status, and disease state [10, 176-178]. For nearly all characterized systems, pheromone molecules are perceived via specialized receptors within a target's olfactory system, and in turn elicit specific, pre-programmed behavioral and/or endocrine responses [5]. For more than 50 years, the earliest and best characterized pheromone systems have been those in insects, commonly employing small, volatile odorants as chemoattractants [10, 14]. The evolution of such systems has been extremely difficult to study, as these odorant molecules are generally the products of complex enzymatic cascades [179]. In contrast, multiple vertebrate systems utilize peptide or protein pheromones to act as chemical signals; as direct gene products, these pheromones are more tractable to both biochemically synthesize and investigate their evolutionary origins [87]. Because pheromones are ligand molecules that bind to target receptors, it is critical to deduce the 3-dimensional structure of pheromone molecules in order to address how different selective mechanisms may be acting in a co-evolutionary receptor-ligand framework. To date, protein structures have only been determined for two mouse pheromones: a major urinary protein (MUP) that affects male aggressive behavior [27, 33] and the male sex pheromone exocrine gland-secreting peptide 1 (ESP1) [23, 38]. Consequently, little is known about the structural evolution of pheromones in non-mammalian vertebrates.

Over the past 25 years, plethodontid salamanders have served as a valuable model for investigating the mechanisms by which protein pheromones regulate behavior and

reproductive success [44, 87]. As basal tetrapods, salamanders are an excellent model to study the origins of terrestrial pheromone signaling in vertebrates. For more than 100 million years, plethodontid salamanders have utilized a unique courtship behavior, termed tail straddling walk, to coordinate insemination and facilitate mating success [78]. In the species *Plethodon shermani*, during tail straddling walk, male salamanders will periodically deliver non-volatile, proteinaceous courtship pheromones to the female by “slapping” an enlarged gland on his chin (the mental gland) to the female’s nares [69, 83]. After a male has applied pheromone to the female that he is courting, the protein molecules diffuse into the female’s olfactory chamber where they bind to receptors on vomeronasal neurons, which project to specific regions of the brain, and influence the female’s mating behavior [70, 86, 92]. It is noteworthy that these pheromones are applied after courtship has initiated, and function to regulate female mating behavior: they are not chemoattractants [83]. Chemical analysis of the *P. shermani* pheromone composition revealed two major components: Plethodontid Receptivity Factor (PRF), a 22-kDa protein related to IL-6 cytokines, and Plethodontid Modulating Factor (PMF), a 7-kDa protein related to the three-finger protein (TFP) superfamily [69, 91, Chapter II]. Both PRF and PMF persist as multi-isoform blends; however, compared to PRF with only 3 isoforms which share ~95% identity, individual male salamanders synthesize more than 30 unique PMF isoforms with ~30% amino acid identity [Chapter II]. Multiple studies of molecular evolution have demonstrated that PMF is under pervasive positive selection, presumably in response to sexual selection from co-evolving female receptors [96, Chapter II].

The TFP superfamily, of which PMF is a member, includes many diverse proteins such as snake neuro- and cytotoxins [115, 120], regulators of the complement system [105], membrane receptors in mammalian tissue re-organization [106], and factors that facilitate amphibian limb regeneration [104]. One central idea in the field of protein structural biology is that, throughout protein evolution, structure is generally more highly conserved than sequence, often as a consequence of functional requirements that promote purifying selection [180]. In support of this, while more than 90,000 structures have been deposited in the PDB, all of these proteins adopt a relatively small number of topological folds (~1300 in CATH) [181, 182]. Furthermore, studies in evolutionary biochemistry further suggest that only a few high-impact mutations on these conserved topologies are necessary for the evolution of novel functions [183]. The TFP superfamily well exemplifies this phenomenon. Establishing a well-resolved TFP phylogeny has been difficult, as homologs from different species share little amino acid similarity and are difficult to align [104, 110]; however, the defining feature of this superfamily is the conserved protein structure of two parallel  $\beta$ -sheets (2- and 3-stranded) arranged in a “three-finger” shape. Importantly, this shape is highly stabilized by 8 conserved cysteine residues that adopt a canonical disulfide bonding pattern (1-3, 2-4, 5-6, 7-8). To date, more than 200 TFP structures have been solved by X-ray crystallography or multidimensional NMR, and all share this canonical disulfide bonding pattern and three-finger shape [104].

PMF has many unique characteristics compared to nearly all other TFPs, despite preservation of the 8 conserved cysteine residues and their relative spacing. First, in plethodontid salamanders, PMF has been subjected to exacerbated gene duplication and

pervasive positive selection, compared to most organisms where TFPs with particular functions are found as single gene copies, and snake venom glands have been specially noted for having up to 5 isoforms of different toxins [108, 120]. Second, while most TFPs carry a net positive charge, nearly all *P. shermani* PMFs are highly negatively charged (mean charge = -9.1). Third, as the only pheromone TFP, rather than being under natural selection like other TFPs, PMF was novel as the first identified TFP under sexual selection. Despite extensive sequencing and proteomic analyses [Chapter II], it remained unclear how the evolution of PMF hypervariability in response to sexual selection might influence the archetypal TFP structure. Therefore, in order to better characterize the structure: function relationships of the PMF pheromone family, the aim of this study was to determine the complete 3D structure of the most abundant PMF isoform (Isoform G; Genbank Accession #JF274292).

## **Materials and Methods**

### Ethics statement

Methods and animal care were approved by Oregon State University's Institutional Animal Care and Use Committee (ACUP 3007 to L.D. Houck). Animals were anesthetized prior to surgery using diethyl ether, minimizing any pain, and post-surgical survival rate was >99%. All salamanders were collected under permits obtained from the North Carolina Wildlife Resources Commission.

### Reagents

All oligonucleotides were synthesized by Integrated DNA Technologies (Coralville, IA). Accuprime High Fidelity (HF) *Taq* Polymerase System, the EasySelect *Pichia* Expression Kit (including the vector pPICZ $\alpha$ A), Zeocin, ultra-pure agarose, and TOP10 chemically competent *E. coli* were purchased from Invitrogen (Carlsbad, CA). All restriction enzymes, T4 DNA Ligase, and additional PCR supplies were purchased from New England Biolabs (Ipswich, MA). GFX gel band purification system was purchased from GE Healthcare (Piscataway, NJ). QIAquick PCR purification system was purchased from Qiagen (Valencia, CA). Sep-Pak Light C-18 cartridges were purchased from Waters Division (Milford, MA). Centriprep ultrafiltration units were purchased from Millipore (Billerica, MA). Trypsin, trifluoroacetic acid (TFA), and all salts were purchased from Sigma-Aldrich (St. Louis, MO). Yeast media reagents, Whatman DEAE cellulose, and acetonitrile (ACN) were purchased from Fisher Scientific (Pittsburgh, PA).

### High Performance Liquid Chromatography

High resolution strong-anion exchange HPLC (Mono Q; Pharmacia, Piscataway, NJ), reverse phase-HPLC (RP-HPLC) (C-18; Grace Davison Discovery Sciences, Deerfield, IL), and size exclusion chromatography (G-75 Superfine; Pharmacia, Piscataway, NJ) was accomplished on a 2695 Alliance HPLC System equipped with a 2487 dual wavelength absorbance detector and Empower software (Waters Division, Milford, MA). The strong anion exchange column (0.5 x 5.5 cm) was eluted at 1 mL/min with a NaCl gradient in 50 mM Tris/HCl buffer, pH 8.0. The C-18 column (0.46 x 15 cm) was eluted with an ACN gradient in 0.1% (v/v) TFA at 1 mL/min. The G-75 column (1.6 x 15.5 cm) was isocratically eluted at ~10 mL/hr with 0.5 X phosphate-buffered saline.

### Mass spectral analysis

Picomole quantities of PMF-G were provided to the University of Louisville Biomolecular Mass Spectrometry Core Laboratory. Intact protein mass was determined by electrospray ionization mass spectrometry (ESI-MS) using a Q-TOF API-US (Waters Division, Milford, MA), while proteolytic fragment fingerprints were acquired by liquid chromatography tandem mass spectroscopy (LC/MS-MS) using a LTQ Orbitrap XL (Thermo Scientific, Waltham, MA). SEQUEST software (Thermo Scientific, Waltham, MA), MassMatrix v.1.3.2 [184], or custom Python scripts built around Extract-MSn (Thermo Scientific, Waltham, MA) were used for all peptide analyses. The average masses of intact proteins and monoisotopic masses of peptides were matched to theoretical average or monoisotopic masses, respectively. Predicted intact masses were adjusted by 1.0078 Da per cysteine to account for the protons displaced in disulfide bonds.



### Purification of natural PMF-G

*P. shermani* males were collected during their breeding season from a single site in Macon Co., North Carolina, USA (35°10'48" N, 83°33'38" W). Males were anesthetized in a mixture of 7% (v/v) diethyl ether in water. Pheromones were extracted following the methods of Houck et al. [93] Approximately 100 glands were excised and pheromones extracted with 0.8 mM acetylcholine chloride in Amphibian Ringer's Solution for ~60 minutes, centrifuged at 14,000  $\times g$  for 10 minutes, the supernatant collected, and the centrifugation repeated before storage at -80°C. PMF isoform G was purified from the whole pheromone extract using the methods described in Chapter II.

### Preparation of rPMF-G expression strain

The *P. pastoris* codon optimized sequence for the most abundant PMF isoform, PMF-G (Genbank Accession # JF274292), was predicted by web-based software from IDT (Coralville, IA). Six overlapping and complementary oligonucleotides based on the sequence were prepared, and used in assembly PCR based on the methods of Stemmer et al. [185]. Purified PCR products were ligated to the vector pPICZ $\alpha$ A and cloned into TOP10 chemically competent *E. coli* using standard procedures. Plasmid DNA from Zeocin-resistant clones was purified, sequenced to validate the construct, linearized by restriction digest with *SacI*, and transformed into *P. pastoris* strains KM71H and GS115 using the EasySelect Pichia Expression Kit. Zeocin-resistant *P. pastoris* clones were screened for recombination by colony PCR using primers flanking the AOX1 locus. Two positive clones from each *P. pastoris* strain were used for small-scale protein expression following the manufacturer's protocols with products analyzed by SDS-PAGE.

### Large scale preparation of rPMF-G

For each preparation, 400 mL BMGY (100mM potassium phosphate, pH 6.0, 2% peptone, 1% yeast extract, 1.34% YNB,  $4 \times 10^{-5}$ % biotin, 1% glycerol) was inoculated with strain KM71H clone 1 and incubated at 29°C with shaking at 275 rpm for ~40 hours. Cells were collected by centrifugation, resuspended in 1 L BMM (100mM potassium phosphate, pH 6.0, 1.34% YNB,  $4 \times 10^{-5}$ % biotin, 1% methanol) with an initial optical density (600 nm) of 0.5, and allowed to continue shaking for ~72 hours with 1% methanol added every 24 hours. Cultures were then centrifuged at  $4000 \times g$  for 20 minutes, the supernatant collected, filtered, and pumped over a 25 mL DEAE-cellulose column pre-equilibrated in 50mM NaCl/50mM Tris/HCl, pH 8 at ~0.8 mL/min. The column was then washed with 150 mL 100mM NaCl/1% Triton X-100/1M Urea/50mM Tris-HCl, pH 8 and the sample eluted in 75 mL 300mM NaCl/1% Triton X-100/1M Urea/50mM Tris-HCl. The sample was twice diluted (1:4) with 1% Triton X-100/1M Urea/50mM Tris-HCl, pH 8 and concentrated to ~2.5mL using a YM-3 Centriprep. The sample was then treated with trypsin in a 1:40 mass ratio overnight at 37°C, products separated by strong anion exchange HPLC (linear gradient: 0-500 mM NaCl in 50 mM Tris-HCl, pH 8 at 10 mM NaCl/min), fractions collected, pooled, and further purified by two rounds of RP-HPLC (linear gradient: 0-40% ACN at 1% ACN/min).

### Validation of rPMF-G structure

Monoisotopic intact protein masses were obtained for both PMF-G and rPMF-G using ESI-MS. Both proteins were reduced with dithiothreitol (DTT) and alkylated with iodoacetamide (IAA) prior to mass acquisition by ESI-MS. Tryptic fragments were sequenced by LC/MS-MS. Far-UV circular dichroism (CD) spectra (185-260 nm) were

acquired for native and rPMF-G by averaging 5 scans across a 0.1-cm path at 0.2 nm intervals using a Jasco J-810 Spectropolarimeter, and curves processed using the R function `loess.smooth` with smoothness parameter 0.05;  $\alpha$ -helix and  $\beta$ -sheet content were estimated using the web application K2D3 [186]. The state of oligomerization was determined by analytical ultracentrifugation using a Beckman XLA analytical ultracentrifuge. The C(S) profile showed a monodisperse sample with an S value consistent with a monomer. The molecular weight derived from the sedimentation coefficient, diffusion coefficient and calculated partial specific volume was within 10% of the expected mass.

#### PMF partial reduction analysis

Partial cystine reduction was accomplished at low pH in order to prevent disulfide bond reformation and scrambling. Both natural and rPMF-G were subjected to restricted reduction using a trialkylphosphine (TCEP; tris-[-2-carboxyethyl]-phosphine; Pierce) at pH 3 in 0.1% TFA for 20 minutes at room temperature followed by immediate injection onto the C-18 RP-HPLC column at pH 2.2 (linear gradient 17.5% to 35% ACN at 0.35% ACN/min). Individual peaks corresponding to the cleavage of 0, 1, 2, 3 or 4 disulfide bonds were collected by hand. Volume and ACN concentration were reduced by incomplete lyophilization. Samples were rapidly alkylated by dropwise addition of the protein solution into 500  $\mu$ L IAA (5 mM in 100 mM Tris, pH 8) while vortexing for  $\sim$ 1 min before the pH was lowered by addition of 500  $\mu$ L 5% formic acid. Alkylated samples were purified using a C18 Zip-tip, divided into 4 aliquots, and diluted with 100 mM ammonium bicarbonate. DTT (7.5 mM) was added to half of the samples, and all samples were subjected to overnight proteolysis with either chymotrypsin or AspN such

that all reduction/protease combinations were performed. Peptide fragments were purified by C18 Zip-tip, and analysed by LC/MS-MS.

### NMR structural studies

Milligram quantities of  $^{15}\text{N}$ -labeled rPMF-G were prepared using the methods above with 1.5g/L  $^{15}\text{N}$ -ammonium sulfate (99%) added to the BMM expression media, and prepared at  $\sim 2.2\text{mM}$  in 50mM KCl/10mM  $\text{Na}_2\text{HPO}_4$ , 90%  $\text{H}_2\text{O}$ /10%  $\text{D}_2\text{O}$ , pH 7. NMR spectra were recorded at 18.8 T on a Varian Inova spectrometer equipped with a 5 mm inverse triple resonance pfg probe at  $20^\circ\text{C}$ . Spectra were processed using NMRPipe [187]. NMR assignments were obtained using a combination of 2D/3D experiments (using  $^1\text{H}$  and  $^{15}\text{N}$ , with natural abundance levels of  $^{13}\text{C}$ ): TOCSY- $^{15}\text{N}$ -HSQC, 2D-TOCSY, 2D-COSY,  $\text{H}^{15}\text{NCO}$ ,  $^{15}\text{N}$ -HSQC,  $^{13}\text{C}$ -HSQC, NOESY- $^{15}\text{N}$ -HSQC, and 2D-NOESY. All NMR spectra were acquired in phase sensitive mode with solvent suppression by Watergate [188]. Spectra were analyzed using SPARKY (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco). Near complete assignment of all backbone atoms (98% non-proline  $^1\text{H}_\text{N}$ , 91%  $^{15}\text{N}$ , 100%  $^1\text{H}_\alpha$ , 96%  $^{13}\text{C}_\alpha$ , and 46%  $^{13}\text{C}_\text{O}$ ) and 97% side-chain  $^1\text{H}$  atoms were obtained. Structure calculations were performed using CYANA [189, 190] with automatic assignment to integrated 2D-NOESY and NOESY- $^{15}\text{N}$ -HSQC peaks. Dihedral angle restraints were obtained using predictions from TALOS+ [191].  $\text{H}_\text{N}$  exchange rate was measured by lyophilizing rPMF-G, the sample resolubilized in 99.99%  $\text{D}_2\text{O}$ , and  $^{15}\text{N}$ -HSQC spectra recorded every hour for 24 hours. Amide groups with half lives greater than 2.5 hours were examined for possible H-bonding partners, and defined using the CYANA hbond function if the distance between the groups was less than 2.5 Å.  $\text{C}_\alpha$  and  $\text{C}_\beta$  chemical shifts suggest all

cysteine residues are oxidized and disulfide bonded [192]; alternative disulfide patterns (for both candidate PMF-G patterns as well as the canonical TFP pattern) were included during CYANA constraint calculation, and the **1-2, 3-6, 4-5, 7-8** pattern yielded the lowest average target score and fewest consistent constraint violations, strongly supporting that it is the correct disulfide bonding pattern. The final ensemble of 20 out of 100 structures did not contain structural or van der Waals violations  $>0.30$  Å. For dihedral angles, there were no violations  $>3.2^\circ$ , and 88.9% of all dihedral angles were found in the most favored regions of the Ramachandran plot, with 9.0% in the additionally allowed regions and the remaining 2.1% in the generously allowed regions. Spin-lattice (longitudinal) relaxation rate constants ( $R_1$ ), spin-spin (transverse) relaxation rate constants ( $R_2$ ), and  $^{15}\text{N}[^1\text{H}]$  steady-state heteronuclear NOEs of the backbone  $^{15}\text{N}$  nuclei were measured. Delay values used were 10, 30, 50, 90, 130, 170, and 210 ms for  $R_2$  experiments, and 10, 80, 150, 300, 500, 750, and 1000 ms for  $R_1$  experiments, all with a recovery delay of 5 seconds. For  $^{15}\text{N}[^1\text{H}]$  NOE measurements, two spectra were acquired with or without 5 seconds of proton saturation during the recovery delay, with the both saturated and unsaturated experiments having a relaxation delays of 5 seconds. All NMR data were deposited in the BMRB (19660), and the structural ensemble deposited in the PDB (2mhy).

### Structural analysis

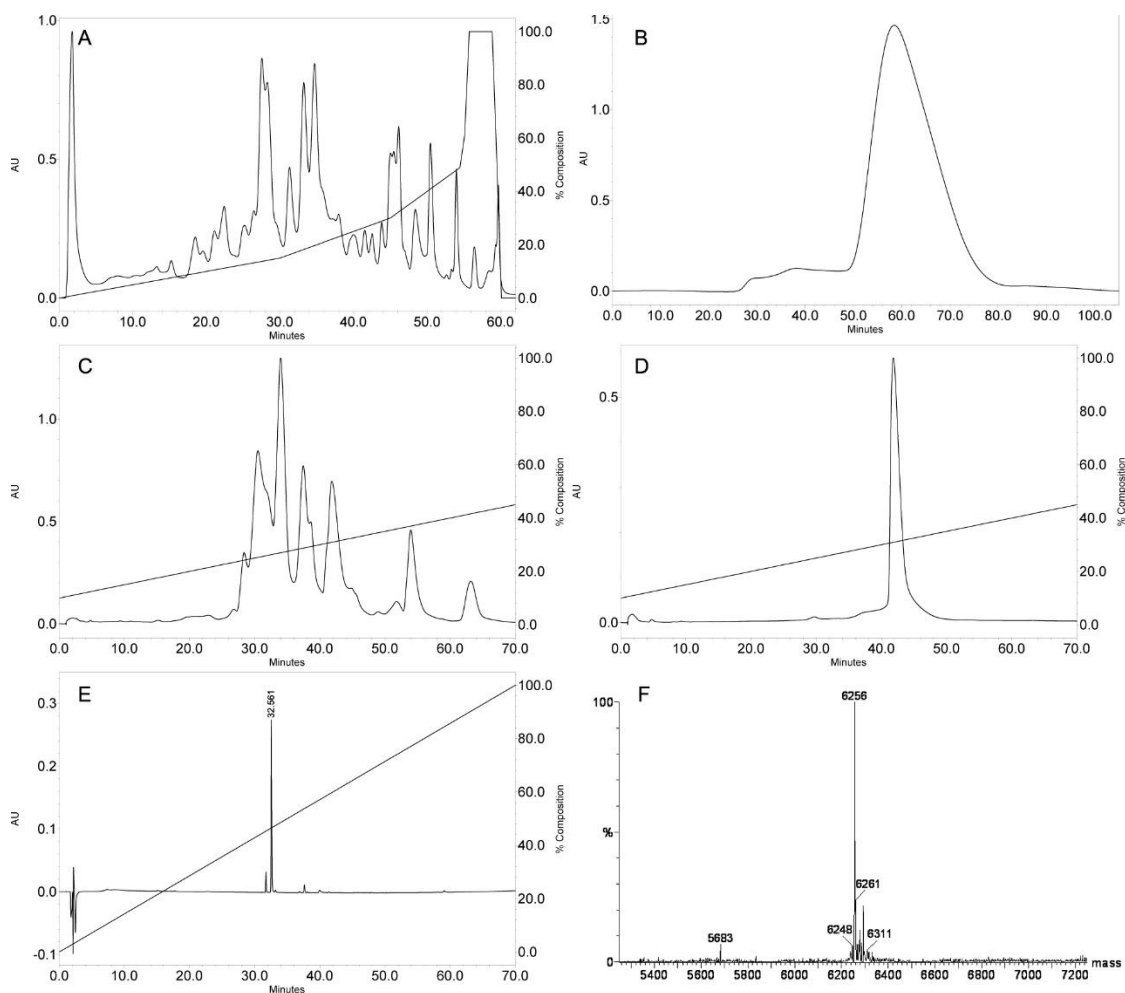
All 3D protein models were produced in PyMOL (v1.3, Schrodinger, LLC), and regular secondary structure defined using the DSS function in PyMOL. Surface renderings with charge distribution are color coded according to amino acid type: acidic (red: Glu, Asp), basic (blue: Lys, Arg, His), hydrophilic (magenta: Ser, Thr, Gln, Asn,

Gly), nonpolar (green: Ala, Leu, Ile, Val, Phe, Tyr, Trp, Met, Pro), or cysteine (yellow: Cys). PMF structural comparisons were made to a representative TFP (short chain neurotoxin from *Naja nigricollis*, 1IQ9); length and charge calculations for structurally characterized TFPs were based on results from ScanProsite at the ExPASy server with query “C-x(5,30)-C-x(2,10)-C-x(10,30)-C-x(2,20)-C-x(5,30)-C-C-x(4)-C-N” (method adapted from Garza-Garcia et al. [104]). H<sub>N</sub> exchange half-lives were calculated by non-linear least-squares regression of the peak integration versus time for H-D exchange experiments. R<sub>1</sub> and R<sub>2</sub> rate constants were calculated by similar non-linear least squares regression versus peak intensity, and saturated to non-saturated NOE ratio calculated based on peak integration. The random coil index for PMF-G was calculated using chemical shift values submitted to the RCI server [193]. PMF sequence variability for all Class I PMFs (Genbank accession #JF274283-274351) was calculated using the protein variability server (PVS) [194], with likelihood of positive selection based on Bayes empirical Bayes results for M2A site specific PAML models from Chapter II. Homology modeling was conducted using Rosetta 3.4 [195]: sequences for additional PMF isoforms were aligned to PMF-G using ClustalW [196], alignable elements of the new isoform sequence superimposed on the peptide backbone of the lowest energy PMF conformer, the disulfide bonding pattern fixed to that of PMF-G, insertion loops built using the loopmodel function, and fastrelax applied to minimize the energy of the resulting model. Ten thousand models were generated per isoform, cluster analysis performed with the cluster radius automatically determined, and the lowest energy structure of the most abundant cluster reported.

## Results

### Disulfide bond characterization of PMF-G

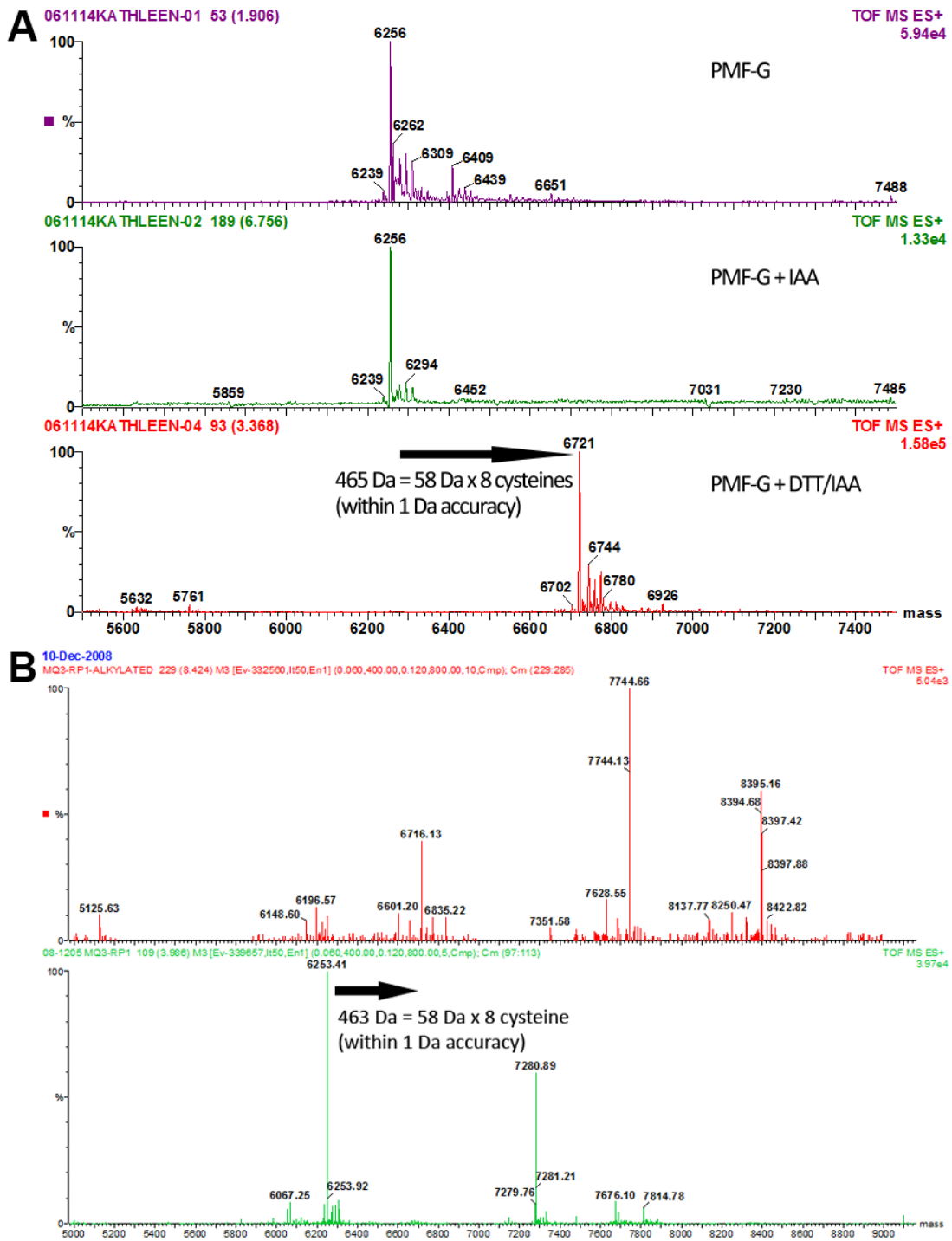
As a member of the TFP superfamily, PMF was expected to adopt the canonical TFP disulfide bond pattern. Initial experiments relied on natural PMF-G purified from *P. shermani* whole pheromone extract through a series of chromatographic separations (Figure 9) [Chapter II]. Analysis by mass spectrometry (MS) confirmed that all 8 cysteine residues were disulfide bonded (Figure 10). However, due to its small size and high disulfide density, PMF-G was extremely protease resistant, and initial efforts to characterize the disulfide bonds by MS with proteolytic digestion yielded ambiguous data that suggested a non-canonical pattern. Based on the methods of Gray [197], PMF-G was next subjected to partial disulfide bond reduction, and the resulting protein species, having different numbers of reduced disulfides, were separated by reverse phase high performance liquid chromatography (RP-HPLC) at pH 2.2 to prevent re-oxidation (Figure 11). Extensive experimentation confirmed that reducing only a single disulfide was sufficient to deduce the majority of the disulfide bonding pattern. Following alkylation to prevent the two free sulfhydryls from reforming a disulfide bond, proteolytic digestion and LC-MS/MS were used to identify peptide fragments containing disulfides **1-2** and **4-5** (Table 1, Figure 12). Fragmentation data for the peptide containing Cys-6,7,8 did not support alkylation of Cys 8, and because it is extremely rare for adjacent residues to form a disulfide bond [198], there is no evidence to support a pattern containing **3-8** and **6-7** bonds. Consequently, the disulfide pattern of PMF-G was deduced to be either **1-2, 3-6, 4-5, 7-8** or **1-2, 3-7, 4-5, 6-8**, both differing from the canonical TFP pattern (**1-3, 2-4, 5-6, 7-8**).



**Figure 9. Purification scheme of natural PMF-G.**

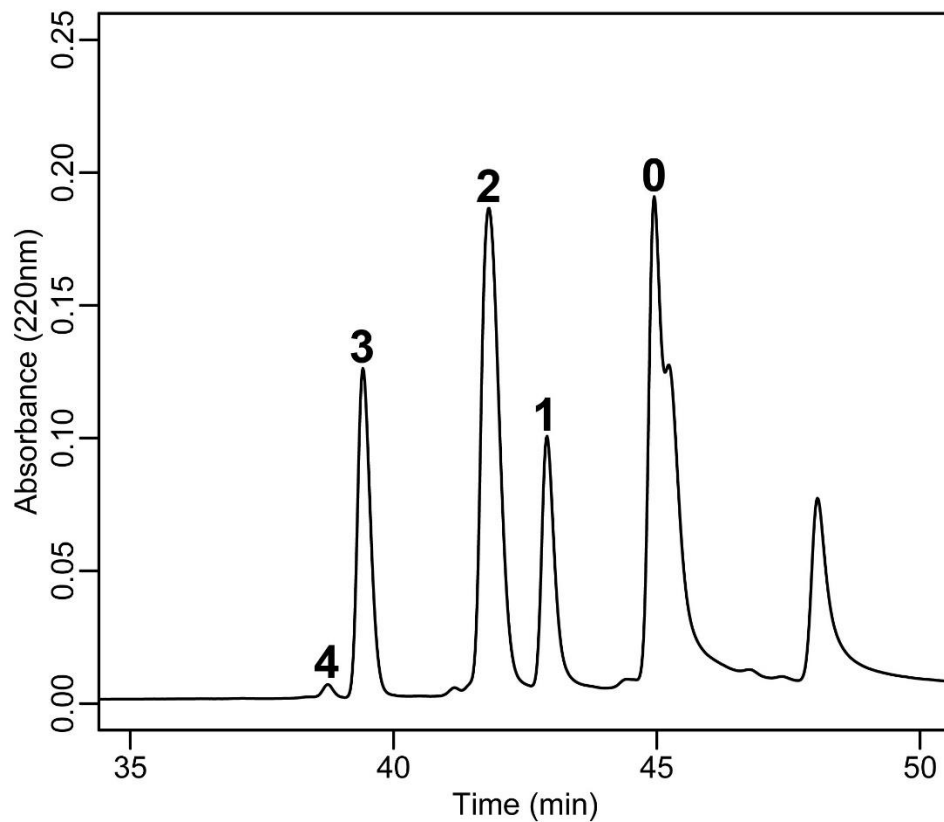
(A) Initial separation of whole *P. shermani* pheromone extract by strong anion exchange HPLC with the mixed rate gradient. Fractions E-I described in Chapter II corresponding to PMF were pooled (elution fractions 43-57 min). (B) Following sample concentration, the PMF mixture was further purified using size-exclusion chromatography. (C) The size exclusion chromatography samples were re-separated by strong anion exchange HPLC on a shallow linear gradient with fraction G collected (~42 min). (D) Fraction G was subjected to a second round of strong anion exchange HPLC, and (c) finally purified at >99% purity by RP-HPLC. (F) MS analysis of PMF-G revealed a highly enriched signal at the expected average mass of 6256 Da.





**Figure 10. PMF-G contains 4 disulfide bonds**

(A) Treatment of PMF-G with IAA resulted in no CAM alkylation, unless first reduced with DTT, implying that all cysteine residues are disulfide bonded in the intact protein; (B) Similar treatment of rPMF-G confirmed that both its molecular weight and cystine content are identical to natural PMF-G.



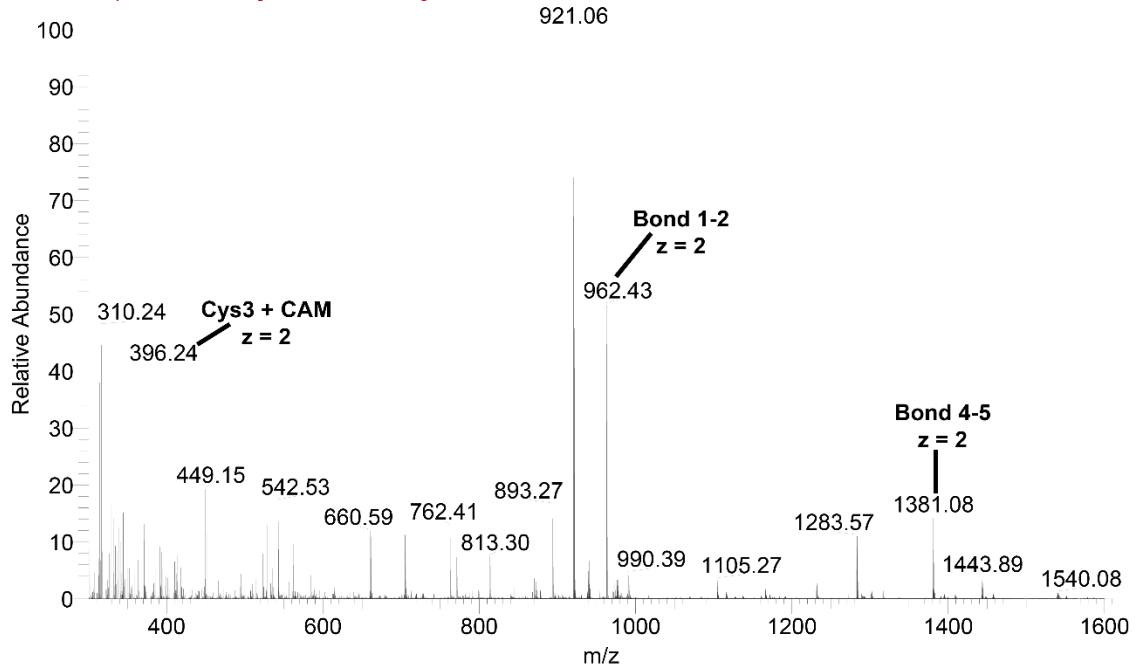
**Figure 11. Partial reduction of PMF-G**

RP-HPLC separation of PMF-G treated with TCEP at low pH to induce restricted disulfide reduction. Each peak is labeled with the number of remaining disulfides, with increasing hydrophobicity as the number of free sulfhydryls increases.

**Table 5. Summary of mass spectral analysis in PMF-G disulfide bonding pattern determination.** Mass spectral analyses was performed on the three-disulfide species of PMF-G purified by RP-HPLC. Differential treatment included proteolytic enzyme (Enz; chymotrypsin [C] or AspN [A]), reduction with dithiothreitol (DTT), and alkylation with iodoacetamide (addition of a carboxyamidomethyl (CAM) group). Observed monoisotopic masses were compared to theoretical masses with no free sulfhydryls, and mass shifts used to determine peptide modification. All assignments were confirmed by analysis of the fragmented ion series.

Bond	Enz	DTT	Sequence	Exp MH <sup>+</sup>	Obs MH <sup>+</sup>	Mass Shift
1-2	C	-	<sup>1</sup> LQNTLG   <sup>18</sup> YIGPICEETG	1923.846	1923.848	-0.002
	C	+	<sup>1</sup> LQNTLDGGTEECIPGI Y <sup>18</sup>	1923.846	1925.864	+2.018 (+2H <sup>+</sup> )
4-5	C	-	<sup>25</sup> KSEDEEYKSCGI   <sup>49</sup> LVTAGEADECEEQ	2760.150	2760.152	+0.002
	C	+	<sup>25</sup> KSEDEEYKSCGIQEEC EDAEGATVL <sup>49</sup>	2760.150	2762.168	+2.018 (+2H <sup>+</sup> )
3-6/7	C	-	<sup>19</sup> NV <sup>CAM</sup> VHY <sup>24</sup>	733.321	791.351	+58.030 (+CAM)
	C	+	<sup>19</sup> NV <sup>CAM</sup> VHY <sup>24</sup>	733.321	791.351	+58.030 (+CAM)
	A	-	<sup>41</sup> DAEGATVL (CC) <sup>CAM</sup>   <sup>57</sup> NCLDEP	1649.637	1707.669	+58.032 (+CAM)
	A	+	<sup>41</sup> DAEGATVL (CC) <sup>CAM</sup> PEDLCN <sup>57</sup>	1649.637	1709.682	+60.045 (+CAM, +2H <sup>+</sup> )

2B1\_tris\_nored\_chym #419-2005 RT: 13.25-39.78 AV: 877 NL: 1.62E6  
F: FTMS + p NSI Full ms [300.00-1600.00]



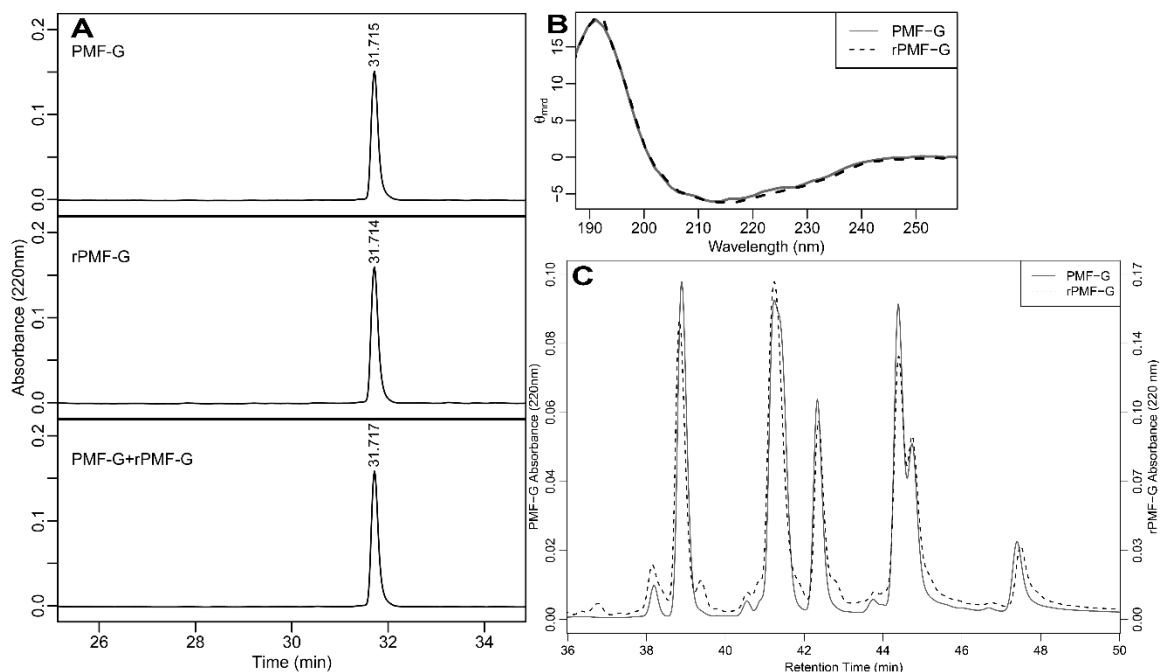
**Figure 12. Mass spectral analysis of partially reduced PMF-G.**

Sample ion spectra of PMF-G, partially reduced with TCEP, the 3 disulfide bonded species collected by RP-HPLC, free sulfhydryls alkylated by iodoacetamide to add a CAM group, and proteolytically digested using chymotrypsin. Specific masses of PMF that were essential for disulfide bond deduction are labeled.

Further LC-MS/MS analyses were unable to resolve the ambiguity between these two patterns, and additional structural characterization was limited by the availability of natural pheromone extract.

### Expression of rPMF-G

To fully characterize the structure of PMF-G, we sought to generate a correctly folded recombinant protein (rPMF-G). Heterologous expression of nearly all TFPs has relied on *in vitro* disulfide formation from scrambled products generated in *E. coli* or by solid phase synthesis [199, 200]. Based on the methods of Greenwald et al. [201], we employed the yeast system *Pichia pastoris*. Assembly PCR [185] was used to prepare a codon-optimized *pmf-g* gene for *P. pastoris*, which was successfully cloned into pPICZ $\alpha$ A for targeted secretion into the growth media. *P. pastoris* clones were initially screened for successful transformation by colony PCR, and small-scale cultures were prepared for positive clones from both Mut<sup>+</sup> (X33) and Mut<sup>S</sup> (KM71H) backgrounds. rPMF-G was only secreted by clones in the Mut<sup>S</sup> background. All assays indicated that rPMF-G had a structure identical to the natural pheromone: LC/MS-MS and ESI-MS confirmed the sequence and mass, respectively; 5 $\mu$ g aliquots of PMF-G, rPMF-G, and a 1:1 mixture of the two produced single peaks by RP-HPLC with retention times varying by < 0.01 min; far UV circular dichroism (CD) spectroscopy generated nearly identical spectra for both proteins; and rPMF-G was validated to have the **1-2** and **4-5** disulfides (Figure 13). The literature suggests that this is only the second time a recombinant TFP has been synthesized without *in vitro* refolding [201].



**Figure 13. Comparison of native and recombinant PMF-G.**

(A) RP-HPLC analysis comparing 5  $\mu\text{g}$  aliquots of PMF-G, rPMF-G, and equal amounts of PMF-G and rPMF-G. The similarity in retention times strongly suggested identical structures between native and recombinant PMF-G. (B) Far UV CD analysis of native and recombinant PMF-G produced very similar spectra, with estimated secondary structure of  $\sim 11\%$   $\alpha$ -helix and  $\sim 29\%$   $\beta$ -sheet content (K2D3; 27), which is similar to NMR results measured by DSS in Pymol (11%  $\alpha$ -helix, 33%  $\beta$ -sheet) [72,73]. (C) Both PMF-G and rPMF-G were treated with TCEP for 20min and major peaks represent 4, 3, 2, 1, and 0 intact disulfides. Retention times were slightly adjusted to correct for run-to-run variation ( $\sim 0.6$  min, 2 different RP-C18 columns,  $\sim 2$  weeks apart). Data from mass spectral analysis of the 1 reduced disulfide species are consistent with the results in Table 5.

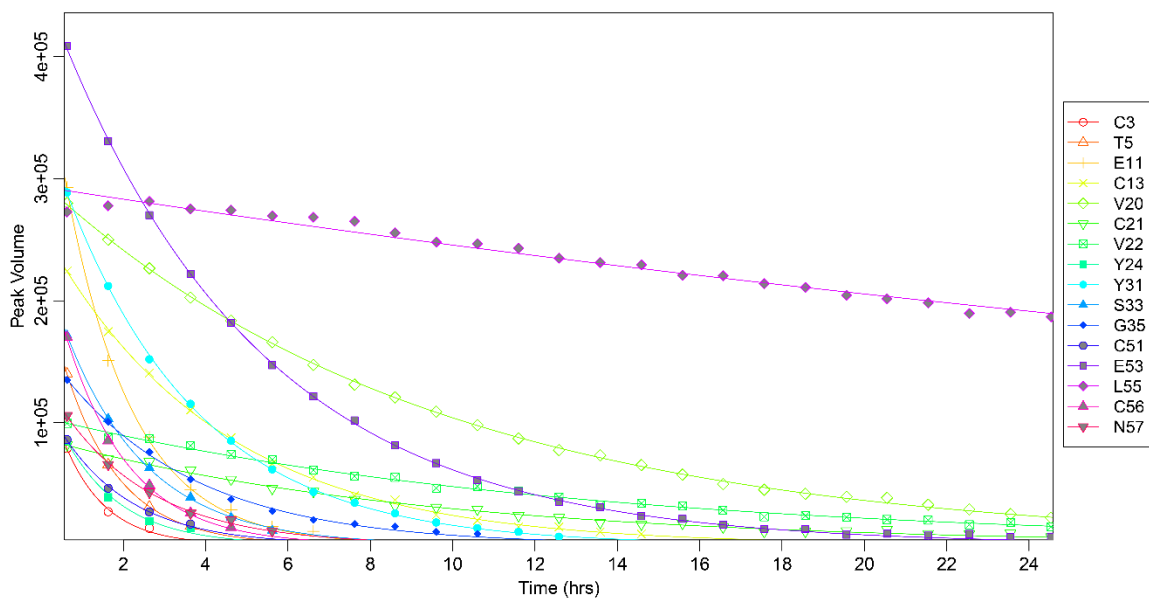
## NMR analysis of rPMF-G

To determine the solution structure of PMF-G, milligram quantities of  $^{15}\text{N}$ -labeled rPMF-G were prepared for multidimensional NMR analysis. Essentially complete assignments of all backbone atoms and side-chain atoms were obtained from 2D NOESY, TOCSY, 3D  $^{15}\text{N}$  HSQC-NOESY, HNC0 and natural abundance  $^1\text{H}$ ( $^{13}\text{C}$ ) HSQC experiments. The 3D structure of PMF-G was solved using standard restrained molecular dynamics simulation with distance restraints determined by nuclear Overhauser effect (NOE) measurements and dihedral angle restraints using TALOS+ (Table 6). Additional hydrogen bond constraints were determined by measuring hydrogen/deuterium exchange rates for backbone amide protons (Figure 14). Structure calculation was performed with constraints using the two alternative disulfide patterns; multiple van der Waals and distance violations were observed when the **3-7/6-8** disulfides were included, whereas there were no consistent conflicts in the other model. Thus, we concluded that the disulfide-bonding pattern of PMF-G is **1-2, 3-6, 4-5, 7-8**. Surprisingly, despite shuffling in 3 of the 4 disulfides, PMF-G still adopts an overall “three-finger” shape (Figure 15). However, the resulting adjustments in the protein backbone eliminated much of the classical TFP topology (a two-stranded  $\beta$ -sheet in finger 1 and a three-stranded  $\beta$ -sheet in finger 2 and finger 3) (Figure 16B). This includes the loss of a finger 3  $\beta$ -strand, leading to a two-stranded sheet in finger 2, and a rotation in finger 2 such that it is near-orthogonally aligned with finger 1 (likely a result of the self-contained altered **3-6** and **4-5** disulfides). Additionally, the novel disulfide pattern eliminated the conserved van der Waals network present between the canonical **1-3, 2-4, 7-8** disulfides, which stabilize the base of the classical TFP structure [110] (Figure 15B).

**Table 6. Restraints and statistics of PMF-G structural ensemble**

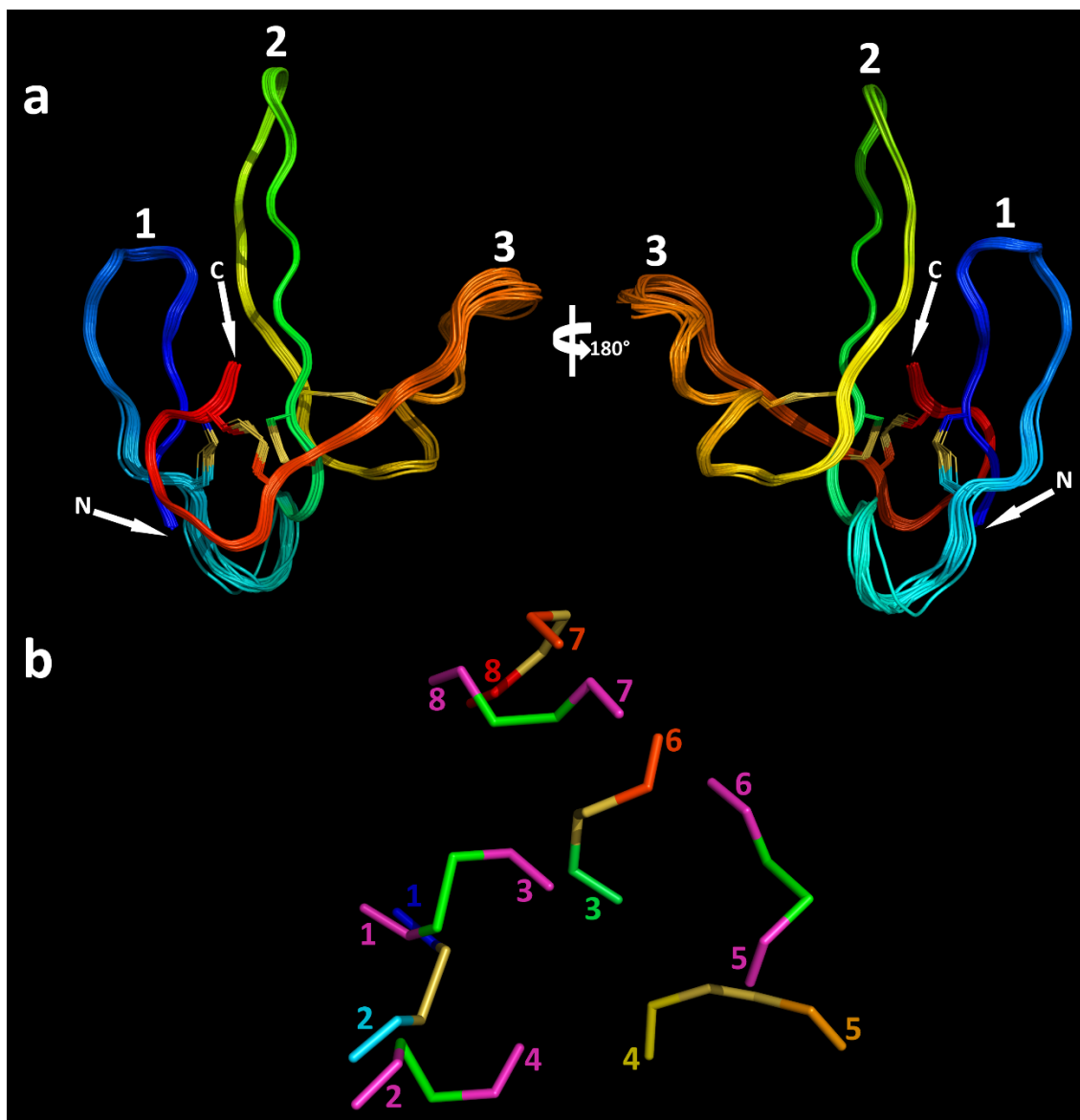
<b>Structural constraints</b>				
<i>NOE distances:</i>	< 3.00 Å	78 (11.1%)	Intraresidue	161 (22.9%)
	3.00 – 3.99 Å	239 (33.9%)	Adjacent ( i-j =1)	217 (30.8%)
	4.00 – 4.99 Å	245 (34.8%)	Short (1< i-j ≤5)	110 (15.6%)
	5.00 – 5.50 Å	142 (20.2%)	Long ( i-j >5)	216 (30.7%)
	Total	704	Total	704
<i>Dihedral angles:</i>	76			
<i>Hydrogen bonds:</i>	3			
<i>Disulfide bonds:</i>	4			
<b>Structural statistics</b>				
<u>Average RMSD to mean (Å)</u>			<u>Ensemble (n=20)</u>	<u>Lowest energy</u>
	Backbone		0.31±0.08	0.19
	Heavy atom		0.73±0.07	0.64
	<i>Target function</i>		0.49±0.052	0.37
<i>Violations</i>	Upper limit	#	1±1	0
		rms	0.0060±0.0015	0.0039
		max	0.13±0.06	0.07
	van der Waals	#	2±0	2
		sum	2.3±0.2	1.9
		max	0.23±0.03	0.20
	torsion angles	#	0±0	0
		rms	0.7072±0.0562	0.6587
	max	2.66±0.30	2.54	
<u>Ramachandran statistics (Procheck [2021])</u>				
	Most favored region (%)		88.9±0.02	87.5%
	Additionally allowed regions (%)		9.1±0.02	10.4%
	Generously allowed regions (%)		2.1±0	2.1%
	Disallowed regions (%)		0±0	0





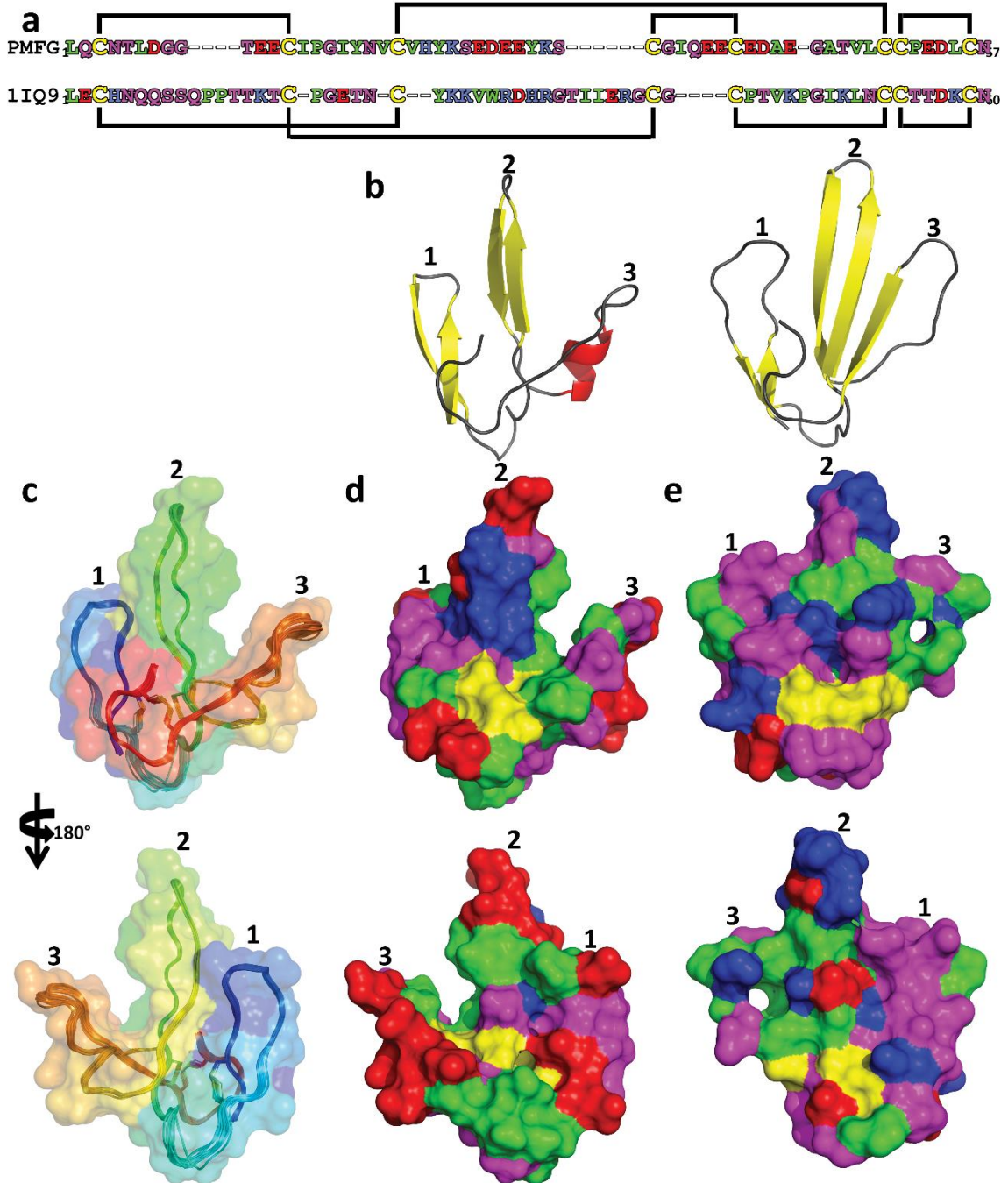
**Figure 14. PMF-G amide H-D exchange rate.**

Plot of peak integration versus time of  $^{15}\text{N}$ -HSQC spectra recorded every hour over 24 hrs for rPMF-G lyophilized and dissolved in  $\text{D}_2\text{O}$ . An exponential decay curve ( $v=v_0\exp(-kt)$ ) was fitted to all peaks with 3 or more points.



**Figure 15. NMR-derived structural ensemble of PMF-G.**

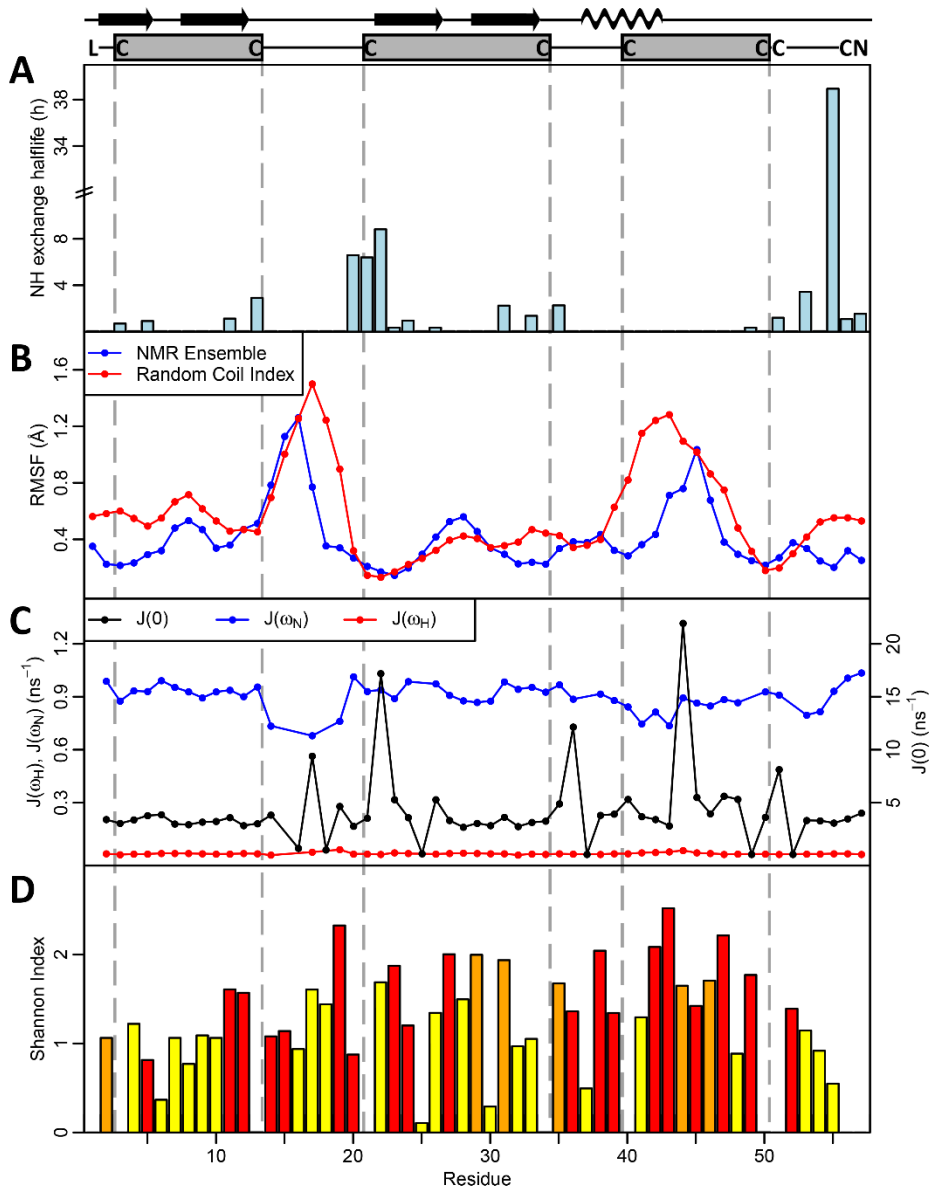
(a) Backbone model of PMF-G with the twenty lowest-energy conformers, color coded from N- to C-terminus (blue to red), and peptide finger numbers denoted (1-3); (b) disulfide bonds in PMF-G from underside view (same color scheme as a) and a representative TFP (1IQ9, carbons in magenta, sulfurs in green).



**Figure 16. Surface models of PMF-G.**

(a) Alignment of PMF-G with a representative TFP (1IQ9), color coded by residue type (acidic, red; basic, blue; hydrophilic, purple; nonpolar, green; cysteine, yellow), with disulfide bonds denoted by the black lines; (b) secondary structure schematic comparing PMF-G (left) and a representative TFP (right; 1IQ9); (c) backbone model of PMF-G (20 lowest-energy conformers) with partially transparent surface rendering (both color coded N- to C-terminus, blue to red); (d) full surface rendering of PMF-G color coded by residue type (same color code as a); (e) surface rendering of 1IQ9 (same color code as a).

Analysis of side chain properties revealed a general segregation between the two elongated protein faces: one side contained the majority of negative and hydrophobic residues, while the other contained the few positive charges and additional hydrophilic residues. This highly polarized negative charge density is in sharp contrast to most other TFP members that typically have a net positive charge (Figure 16D-E). The lack of secondary structure between fingers 2 and 3 may be the result of charge repulsion due to the concentration of acidic residues on the two fingers, leading to an extended finger 3 and forming a cleft between them. The majority of backbone amides in PMF-G were solvent accessible and rapidly exchanged (34 out of 54 H<sub>N</sub> groups were undetectable after 20 minutes in D<sub>2</sub>O, and only 13 H<sub>N</sub> groups had half-lives greater than one hour; Figure 14); however, the slowest-exchanging amides were found near the base of finger 2, and near the N- and C-termini nexus. Notably, the Leu55 amide proton was nearly non-exchangeable (half-life >38 hours) (Figure 17A), and is likely H-bonded with the Gln2 backbone carbonyl. Additionally, two highly conserved residues on finger 2 (Glu30 and Lys25) were found in close proximity, and likely form a novel salt bridge to help stabilize finger 2. Although the resulting structural model is well determined (average backbone rmsd = 0.31 Å), there is less β-sheet structure in PMF-G relative to other TFPs (Figure 2B; consistent with CD results, Figure 13B). When compared against all available PDB sequences using PDBeFold, the closest match was a γ-bungarotoxin (1MR6) with a Q-score of only 0.24 (P-score 0.1, Z-score 1.2), suggesting that PMF has a previously uncharacterized protein topology.

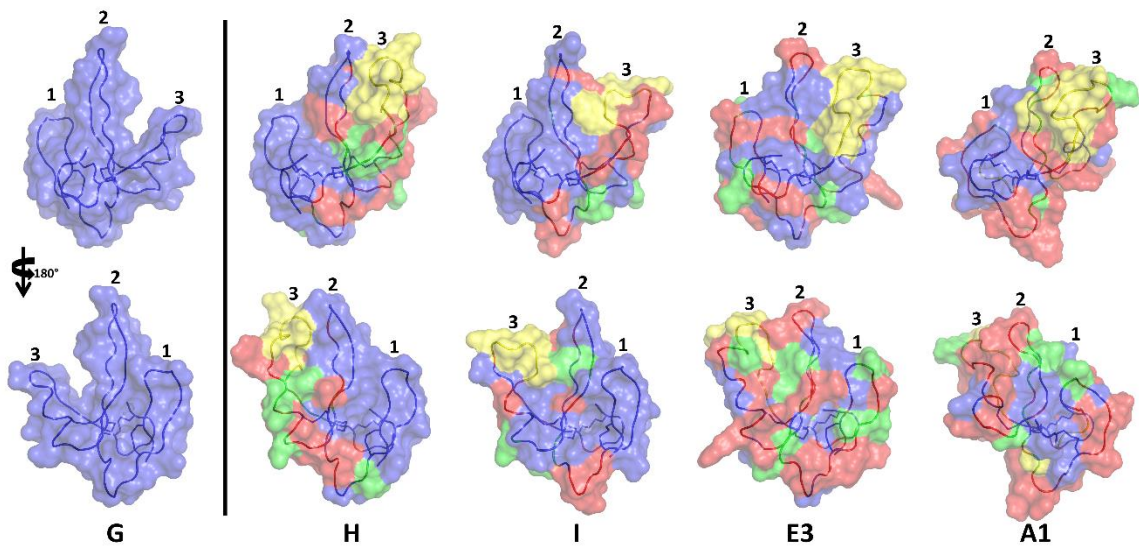


**Figure 17. Measurements of sequence and structural variability in PMF.**

(a) Backbone amide ( $H_N$ ) exchange H/D exchange rates measured by half life (in hours), with proline residues omitted; (b) Root mean squared fluctuation (RMSF) per residue in the PMF structural ensemble (blue) and predicted from the random coil index (red); (c) spectral density functions at 0,  $\omega_N$ , and  $\omega_H$ , with  $J(0)$  sensitive to fast (ns) and slow motions,  $J(\omega_N)$  to motions on time scales faster than ( $1/\omega_N = 2$  ns), and  $J(\omega_H)$  to motions faster than  $1H$  ( $1/\omega_H = 0.2$  ns); (d) Sequence variability (Shannon entropy index) at each residue measured for all Class I PMFs, shaded according to likelihood of positive selection at each position (red  $p < 0.01$ , orange  $p < 0.05$ ; yellow = neutral selection). Seven out of the nine non-conserved amino acids in finger 3 display signatures of positive selection, suggesting combined structural flexibility and rapid evolution in this region.

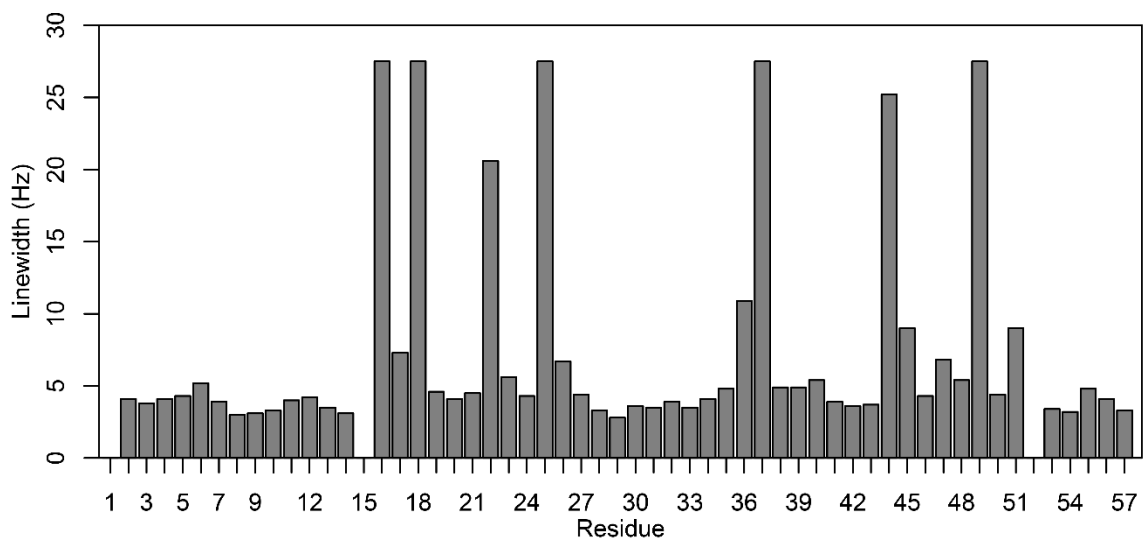
### Comparative modeling with additional PMF isoforms

To assess how PMF sequence hypervariability may be structurally manifested, sequence comparison and homology modeling were conducted for additional PMF isoforms using the NMR-derived PMF-G structure as a template. Of the 99 PMF haplotypes reported in Chapter II, the spacing of the first 5 cysteines is conserved in 75% of the sequences, and varies by no more than 3 residues in the remaining 25%. This could be considered an underestimate, as 86% of Class I PMFs (which comprise ~90% of the total PMF protein) share this spacing for the first 5 cysteines. However, the region between the 5<sup>th</sup> and 6<sup>th</sup> cysteines (equivalent to most of finger 3) is more variable, both in length ( $15.6 \pm 2.6$  residues; PMF-G = 9 residues) and sequence. Homology models for four additional highly expressed PMF isoforms (H, I, E3, and A1) all have extended loops on finger 3 (Figure 18). In the three most abundant Class I PMFs (G, H, and I), fingers 1 and 2 are predicted to be nearly identical with respect to both sequence and structure, with finger 3 being the only highly variable region. Additionally, in the PMF-G structural ensemble, two regions display greater backbone flexibility than the rest of the protein: the loop between fingers 1 and 2, and the length of finger 3. This flexibility is further supported by multiple lines of evidence: nearly all of the backbone amides in these regions are solvent accessible and exchanged rapidly, fewer well-defined NOEs were observed for these regions, <sup>15</sup>N linewidths were broader for many of the backbone amides, and these residues had higher predicted random coil indexes (based on chemical shift values) (Figures 17A-D, 19). Relaxation experiments confirmed that residues in these regions (specifically, 17, 36, 44, 45, 47, and 51) were flexible on  $\mu$ s-ms time scales (Figures 17, 20).



**Figure 18. Homology modeling of major PMF isoforms.**

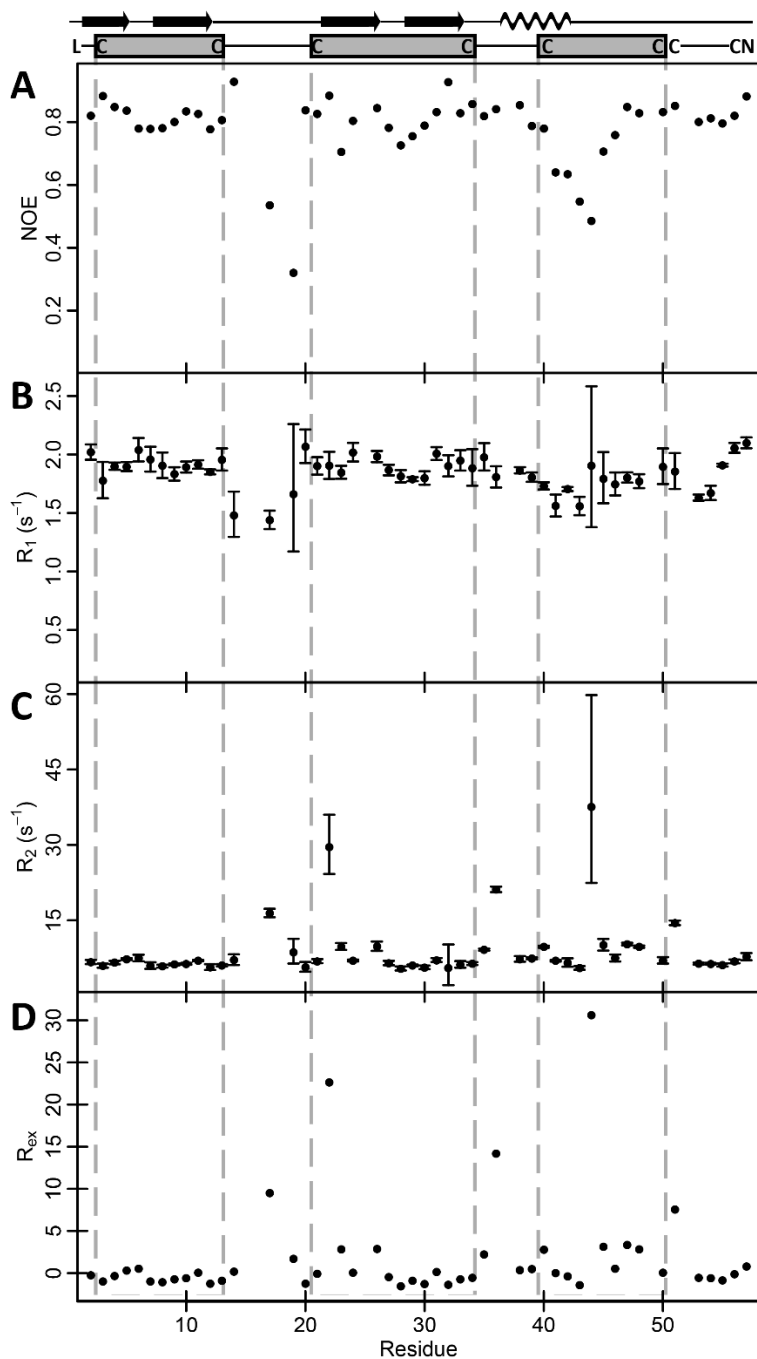
Homology models of four additional PMF isoforms that are highly expressed in *P. shermani* (isoform H, accession #JF274289; isoform I, accession #JF274304; isoform E3, accession #JF274344; isoform A1, accession #JF274380). Models are color coded according to amino acid conservation relative to PMF-G, which is included as a reference in the first panel (same residue, blue; conservative substitution, green; nonconservative substitution, red; insertion, yellow).



**Figure 19. PMF-G backbone amide  $^{15}\text{N}$  linewidths.**

Barplot of  $^{15}\text{N}$  linewidths for backbone amides derived from a  $^{15}\text{N}$ -HSQC spectrum. The N-terminal Leu and two Pro residues were assigned 0 Hz, and residues undetectable by  $^{15}\text{N}$ -HSQC (residues 16, 18, 25, 37, 49) were assigned 27.5 Hz.

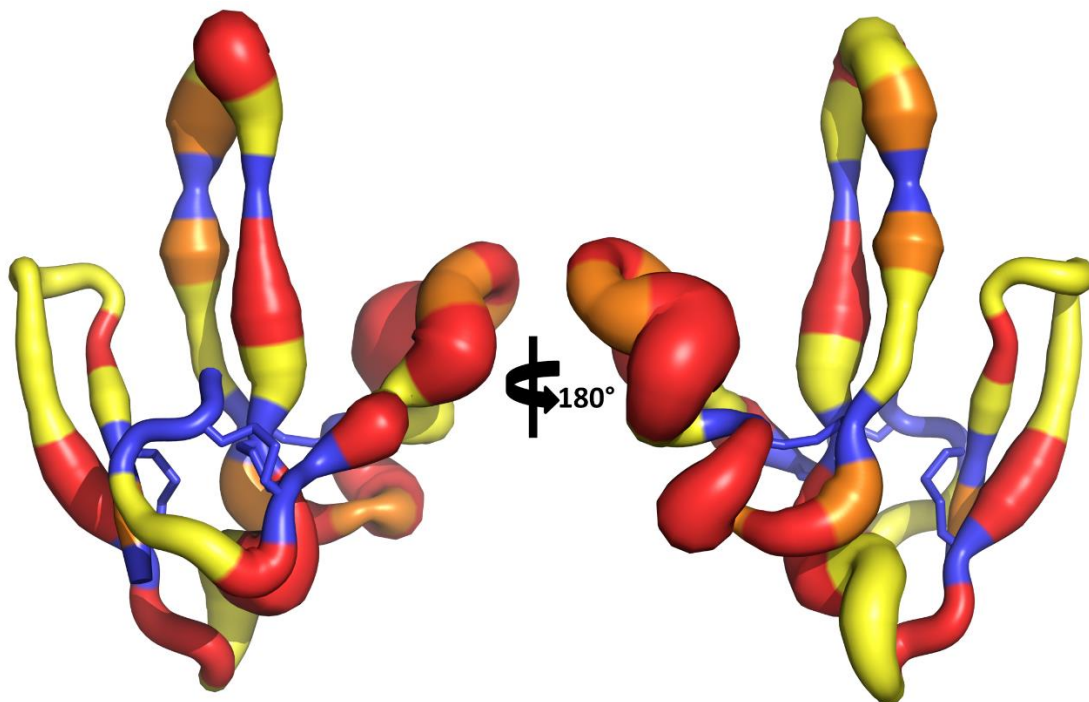




**Figure 20. rPMF-G NMR relaxation analysis.**

Relaxation analysis of rPMF-G examined by (a)  $^{15}\text{N}[^1\text{H}]$  steady-state heteronuclear NOE measurements (with lower values suggestive of conformational changes), (b) spin-lattice (longitudinal) relaxation rate constants ( $R_1$ ) (with lower values indicating sub-ns exchanges), (c) spin-spin (transverse) relaxation rate constants ( $R_2$ ) (with higher values indicating  $\mu\text{s}$ -ms exchanges), and (d) the  $R_{\text{ex}}$  rate.  $R_1$  and  $R_2$  are reported as parameter estimates  $\pm$  95% confidence interval.

Additionally, models of molecular evolution (from Chapter II) indicated that the majority of positively selected residues are found on finger 3 (Figures 17E, 21). Together, these data suggested that, in addition to rapid evolution of finger 3, the altered disulfide bonding pattern of PMF-G disrupted the classical TFP topology and permitted greater structural flexibility in this finger in order to maximize the number of sequence/structural permutations of PMF.



**Figure 21. Rates of molecular evolution on PMF-G.**

Putty model of PMF-G, with backbone width proportional to residue variability (Shannon-Weaver index in Figure 17E), and color-coded according to the likely mode of molecular evolution (based on data from Chapter II; backbone, black; purifying selection, blue; neutral selection, yellow; positive selection,  $0.01 \leq p < 0.05$ , orange; positive selection,  $p < 0.01$ , red).

## Discussion

Within evolutionary, biomedical, and structural scopes, the TFP superfamily has remained a key research target due to both the unique structural elements and the diverse functions of its many orthologs. Within the snake toxin TFPs alone, through adaptive evolution, members have adopted many distinct functions including ion channel blocking [116-118], nicotinic and muscarinic receptor antagonists [119-121], acetylcholinesterase inhibition [100], cell-adhesion regulation [122], integrin binding [123], and pore formation in the plasma membrane [124, 125]. There also exist membrane-bound TFPs, including CD59, Ly6 antigen, and the newt regenerative positional marker Prod1. In a study by Garza-Garcia et al. [104], the solution structure of Prod1 was solved and fit within the canonical TFP framework; however, within a phylogenetic context, Prod1 was much more similar to PMF in both sequence and predicted biochemical characteristics. If PMF and Prod1 are relatively recent paralogs within salamanders, then PMF's novel topology and disulfide pattern are likely more recently derived characteristics. Notably, expression of PMF-G in *P. pastoris* suggested that the novel disulfide bonding pattern is thermodynamically favorable and not the product of plethodontid-specific chaperones and/or protein disulfide isomerases. In contrast to most TFPs (excluding Prod1), PMF has a high net negative charge which may affect its folding dynamics. A comprehensive structural analysis of the TFP superfamily by Galat et al. [110] found little sequence similarity beyond the 8 core cysteine residues, and last three Cys arranged in the CCXXXXCN motif. Despite the lack of sequence similarity, three of the four disulfides (**1-3**, **2-4**, **7-8**) form a tight van der Waals interaction network that stabilize the double  $\beta$ -sheet structure ( $< 4 \text{ \AA}$  between the bonds,  $< 1 \text{ \AA}$  average rmsd for this network between

TFP members). In the example where the Cys 6-7 doublet is split by an additional residue (TGF $\beta$ -RII), this network is disrupted and the structure is less stable. Likewise, the altered disulfide bonding pattern of PMF-G disrupted this network such that the disulfides are spaced further apart (Figure 15B), and may partially relate to the loss of the  $\beta$ -strand normally found in finger 3. Interestingly, the spacing between the first two cysteines was conserved among all PMFs (9 residues), and is shorter than any structurally characterized TFP (length =  $17.0 \pm 5.9$  residues; min residues = 10). This shorter spacing may be important in promoting formation of the **1-2** disulfide, which in turn could prevent the canonical **1-3** bond from forming and help drive the novel disulfide pattern.

A central question that remained was what is the adaptive value in PMF adopting a novel disulfide-bonding pattern relative to the canonical TFP structure? Within ~30 million years, the PMF complex has undergone tens to hundreds of gene duplications to yield the ~100 expressed mRNA haplotypes observed in *P. shermani* cDNA [Chapter II]. At the same time, these genes have been under strong sexual selective pressure to differentiate and adopt potentially novel signaling roles in order to affect female behavior and physiology [96, Chapter II]. When whole mental gland extract was applied to female salamanders, courtship time decreased by ~20% [83]. Surprisingly, when a mixture of more than 30 PMF isoforms was tested (that did not include PMF-G), courtship time increased [93]. While this subset of PMF isoforms activated VNO neurons and regions of the female brain classically involved in pheromone response [70, 86, 92], a more recent set of experiments revealed that a more complete PMF mixture (that included PMF-G) decreased courtship time similar to whole extract but without significantly activating more VNO neurons (see Chapters IV and V). Consequently, our working hypothesis is

that synergistic interactions between the many diverse isoforms are necessary for PMF to increase female mating receptivity. This is perhaps in contrast to other polygenic pheromone families, such as MUPs in mice, where different isoforms are uniquely involved in mediating gender recognition, male-male aggressive behavior, female sexual receptivity, and learning of individual odor profiles [24-27, 33]. Expression of different PMF isoforms is highly variable between male salamanders; however, PMF always constitutes ~50% of the total mental gland pheromone [91], with PMF-G almost always being the most abundant isoform (~12% of the total PMF) [Chapter II]. In the current study, we have provided evidence that the most variable and rapidly evolving segment of PMF (finger 3) is also structurally flexible. Homology modeling supported that both the length and shape of this finger is likely variable in additional PMF isoforms, and that this segment has the greatest topological differences from the canonical TFP structure. Combining the sequence variability, structural flexibility, and altered topology relative to the TFP superfamily, we hypothesize that finger 3 plays a critical role in PMF-receptor interactions, utilizing both residue variability and backbone flexibility to permit a significantly greater number of structural permutations that may occupy a broader range of female receptors. As female receptors continue to evolve, this structural flexibility may permit PMF to still interact with target receptors by adopting a slightly different conformation, without the immediate need for complementary mutations. Consequently, we hypothesize that PMF may have evolved a form of “resilience” to mutations in female receptors, that might otherwise ablate pheromone:receptor interactions, and thus provide males with an enhanced ability to stimulate any mating female in the breeding population. The precise mechanism by which PMF regulates female mating behavior has

yet to be determined; however, PMF stimulated neurons in the female vomeronasal organ and activated regions of the brain known to be involved in pheromone response [86, 92]. Based on these data, PMF is presumably binding to a vomeronasal type-2 receptor (V2R), which are highly abundant in the *P. shermani* VNO [72] and have been implicated in protein pheromone signaling in rodents [20, 22, 27]. Very few specific receptor:ligand pairs have been identified for vomeronasal receptors [20, 23, 34, 35], but to date, none of these examples include TFP:V2R interactions. Recently, the 3D structure was determined for the mouse sex pheromone ESP1, and through mutagenesis assays and molecular docking studies with its specific V2R receptor, it was determined that charge-charge interactions provide most of the binding specificity. Future studies of PMF will seek to determine specific VNO receptors that mediate reception of plethodontid courtship pheromones and to understand the molecular interactions that drive pheromone:receptor co-evolution.

## Conclusions

Over tens of millions of years, sexual selection has promoted rapid evolution in the three-finger protein pheromone, Plethodontid Modulating Factor. In addition to sequence hypervariability, this process has altered the highly conserved TFP disulfide bonding pattern and topology which has resulted in increased backbone flexibility in the putative receptor binding sites. Taken together, the sequence diversity and structural flexibility likely permit thousands of PMF conformers, increasing both the signal plasticity of PMF and the likelihood of stimulating any female in the mating population. In support of this hypothesis, preliminary data suggest that female receptivity increases when females receive sufficient PMF isoform diversity. This “evolved conformational flexibility” may confer PMF robustness to ever evolving changes in female receptors. This work lays the foundation for future research in understanding the molecular adaptations that arise as part of the sexual conflict between males and females that can lead to an evolutionary “arms race” of signals by one gender and receptors of the other gender.



CHAPTER IV - DIFFERENTIAL FEMALE BEHAVIOR FROM A  
HYPERVARIABLE COURTSHIP PHEROMONE IN THE RED-LEGGED  
SALAMANDER, PLETHODON SHERMANI

**Chapter Overview**

Pheromones are a diverse class of biological molecules that play critical roles in mediating social and sexual behaviors. In many systems, pheromones exist in complex mixtures, with the precise composition and ratios of the different components essential for bioactivity. The interactive effects of complex pheromone mixtures, however, have been minimally studied in vertebrates. In the red-legged salamander (*Plethodon shermani*), male salamanders use non-volatile proteinaceous pheromones to modify female courtship behavior and mating receptivity. One component of this pheromone mixture is a hypervariable 7 kDa protein, Plethodontid Modulating Factor (PMF). Within a single population, individual male salamanders express more than 30 different variants (isoforms) of PMF. In contrast to the complete pheromone extract, a subset of PMF isoforms was previously demonstrated to lower female mating receptivity. In the current study, we demonstrated that a single PMF isoform had no effect on female mating behavior, while a more complete mixture of PMF variants recapitulated the effect of the whole pheromone mixture and increased female receptivity. From these data, we hypothesize that: (a) female preference and sexual selection have promoted the rapid

gene duplication of PMF over ~20 MY, resulting in the complex mixture we observe today; and, (b) PMF isoforms act synergistically through complex neurophysiological pathways to modulate female courtship behavior. These studies help define a framework for further investigations of the complex interactions and molecular mechanisms by which protein pheromones modulate female mating behavior.

## Introduction

Chemical communication by pheromone signaling permits the transmission of specific information between individuals, including information about species, sex, reproductive status, and disease state [176-178]. A defining characteristic of pheromones is the elicitation of behaviors or neuroendocrine responses in the receiving animal [10]. The chemicals that compose pheromone mixtures may be derived from virtually any type of biomolecule. In several well-characterized insect pheromone systems, small volatile hydrocarbon chains or rings are released into the environment to attract potential mates or to communicate other species-specific information [14, 16, 203, 204]. In contrast, vertebrates often use water soluble signals which may be widely dispersed or privately delivered. For example, male firebelly newts secrete a peptide pheromone (termed sodefrin) that can attract gravid females [46]; adult sea lamprey release sulfated sterols as migratory pheromones that recruit juveniles to spawning streams [205]; and in mice, the many isoforms of the major urinary protein (MUP) family perform multiple social functions, including promotion of male aggression towards other males, regulation of female receptivity, and learning of individual scent profiles [24, 26, 27].

However, most if not all pheromone molecules are delivered as complex mixtures, and their bioactivity is often tied to both the presence and relative proportions of different components [206, 207]. Protein pheromones, in particular, are well suited for the study of such interactions in vertebrates for three reasons. First, protein pheromone genes often comprise multigene families that are products of exacerbated gene duplication and positive selection [24, 94]. Second, as direct gene products, their synthesis and expression levels are generally regulated through well-characterized

biochemical processes. Third, recombinant pheromones can be readily prepared using heterologous expression systems to control the exact composition of experimental mixtures [25, 208]. Even with the many advantages of protein pheromones, relatively few studies have examined protein interactions in complex mixtures. This dearth of studies is surprising because the prevailing opinion is that interactions are a central property of many pheromone communication systems [209-213]

For more than 100 million years, plethodontid salamanders have utilized mixtures of non-volatile, proteinaceous courtship pheromones to regulate female behavior and mating receptivity [85]. As basal tetrapods, salamanders provide an excellent model system for studying the evolution of pheromone signaling. Courtship pheromones, unlike many chemoattractants, are a special type of sex pheromone that are privately delivered during courtship and influence associated behaviors [87]. The annual mating season for many plethodontid salamanders occurs during a few months in late summer or early fall. Before this mating season, plasma androgen levels rise in adult male salamanders. The increased androgen induces the development of a specialized chin gland (termed a "mental gland"). The mental gland is solely dedicated to the production of protein courtship pheromones [69, 214]. In a successful courtship between a male and female, the female typically straddles the male's undulating tail and the pair walks forward in unison. This behavior was described as a tail-straddling (TSW) walk by Arnold [78]. At the end of this walk, the male deposits a spermatophore, and the female moves over and then down on that spermatophore to obtain the apical sperm mass [78]. In our principal model the red legged salamander (*Plethodon shermani*), the male salamander will periodically turn and deliver courtship pheromones to the female by "slapping" his mental gland

against the female's nares. These pheromones then travel along the female's nasolabial grooves which deliver the aqueous pheromone to neural receptors in the vomeronasal organ (VNO). Ultimately, pheromone stimulation results in activation of specific brain regions that modify female courtship behavior [70, 83, 86]. In laboratory trials, the pheromone extract reduced the time females spent in TSW. We interpreted the decrease in time in tail-straddling walk as an increase in female mating receptivity [85].

Chemical analysis of the *P. shermani* pheromone extract revealed that >85% of the mixture was comprised of two major proteins [69]. The first identified component was a 22-kDa protein termed Plethodontid Receptivity Factor (PRF) which is related to IL-6 cytokines. Similar to the complete pheromone extract, purified PRF also increased female receptivity [83]. Through continued biochemical analysis, three sequence variants (isoforms) of PRF were identified. These variants were termed B, C1, and C2. This nomenclature was based on the relative charge and elution conditions when the pheromones were analyzed by high performance liquid chromatography (HPLC) (see Figure 8A and Table 7). All three PRF isoforms shared a >95% sequence identity. Individual analysis of pheromone extracts from more than 100 male *P. shermani* revealed significant variability in the relative ratios of these isoforms, with ~20% of animals expressing only two of the three isoforms. Notably, all current evidence suggested that these isoforms were the products of gene duplication, and not allelic variation or alternative splicing [91]. Courtship trials testing the efficacy of a recombinant PRF-C2 elicited the same response as the complete mixture (increased female receptivity), suggesting some redundancy in function between the isoforms [208].

**Table 7. Summary of pheromone isoforms by HPLC fraction.** Greyed boxes indicate uncharacterized composition and/or no behavioral tests prior to this study. For G, H, and I, the fractions are >95% for a single isoform, with one or two minor isoforms.

HPLC Fraction	Pheromone	No. of Isoforms	Effect on female receptivity
A			
B	PRF	1	+18% (Rollmann et al. [83])
C	PRF	2	+27% (Houck et al. [208])
D			
E	PMF	~20-25	-23% (Houck et al. [93])
F	PMF	~10-15	
G	PMF	1-3	
H	PMF	1-3	
I	PMF	1-3	

The second pheromone component identified from male mental glands was the 7-kDa protein Plethodontid Modulating Factor (PMF). PMF is related to the three-finger protein (TFP) superfamily that includes many snake venom neuro- and cytotoxins [90]. Proteomic and molecular analyses for *P. shermani* revealed that, when compared to the 3 highly conserved PRF isoforms, individual males expressed more than 30 unique PMF isoforms that only shared ~30% average amino acid identity. PMF was observed in five major HPLC fractions (termed E, F, G, H, and I; Figure 8A); further analysis revealed that nearly all of the PMF isoforms were found in the E and F fractions ( $\geq 30$  isoforms total; Table 1), while the G, H, and I fractions were each highly enriched for individual isoforms (each >95% purity). Consequently, these three isoforms (G, H, and I) were the most abundant PMFs and comprised ~25% of the total PMF mixture [Chapter II]. In an earlier study [93], a mixture of PMF-EF was tested in staged courtship trials. In contrast to both whole pheromone and PRF, PMF-EF actually decreased female receptivity (based on an increase in TSW time). It was hypothesized that, due to its homology with snake venom neurotoxins, PMF may be acting to relax the female, thereby facilitating completion of the courtship.

However, as already noted for many insect systems, both the presence and ratios of particular pheromone components are often critical for proper biological activity. As an incomplete mixture of PMFs was used in the previous assay, further study on the impact of these highly abundant isoforms was warranted. Using a yeast expression system, a recombinant PMF-G (rPMF-G) was prepared that is biochemically identical to the natural pheromone [Chapter III]. Importantly, PMF-G was the most abundant PMF isoform in nearly all male *P. shermani* [91, Chapter II]. While a single PRF isoform was

sufficient to stimulate female salamanders [208], it was unknown if the same phenomenon would occur with PMF, which has >10X the isoform diversity. To assess the potential interactive and synergistic effects that may underlie the PMF complex, we determined the effects of rPMF-G and PMF-EFGHI on female mating receptivity in *P. shermani*.



## Methods

### *Plethodon shermani* courtship behavior

The courtship ritual of *P. shermani* involves many characterized behaviors and occurs in well stereotyped stages [78]: (1) *Orientation*: the male faces towards the female and follows her. (2) *Persuasion*: the male presents visual (and chemical) cues to attempt to persuade the female to court. The performance of one persuasion behavior (foot dancing) is positively correlated with advancement to the next phase [79]. (3) *Tail Straddling Walk*: during this lengthy phase, the male will periodically turn and “slap” the female with his mental gland to apply courtship pheromones. Stimulation from these pheromones decreases the length of TSW [83]. (4) *Deposition*: the male periodically pauses during TSW, and if the female does not try to advance further (possibly implying ample female receptivity), he will press his body to the ground and deposit a spermatophore while continuing to undulate his tail. The deposition process consistently requires ~7 minutes. The spermatophore is composed of a small sperm cap that sits atop a gelatinous base. (5) *Insemination*: the male lifts his vent off of the deposited spermatophore and walks forward. The female typically follows until her vent is positioned above the spermatophore. She then lowers her vent removes the sperm cap (leaving the gelatinous base behind), and then departs from the male, completing courtship.

### Animal collection, maintenance, and pre-screening

*Plethodon shermani* salamanders were collected during their breeding season from a single site in Macon Co., North Carolina, USA (35°10'48" N, 83°33'38" W) in August 2010, and sexed based on the presence of a mental gland in males or large ova in

females. Initially, salamanders were maintained at Highlands Biological Station. All animals were individually housed in clean plastic boxes (17 x 9 x 13 cm) each lined with a damp paper towel and supplied with a second damp crumpled paper towel for a refuge. Once per week, animals were transferred to clean boxes with fresh substrate and fed 2 waxworms (*Galleria mellonella*). The temperature and humidity were maintained at 15-18°C and ~70%, respectively. Some *P. shermani* will not mate under laboratory conditions, such that for larger observational studies, animals must be pre-screened to determine mating propensity [93]. Briefly, male and female salamanders were randomly paired in clean plastic boxes with a damp paper towel substrate (no refuge), and left together overnight. The following morning, we determined if the pair mated by (a) examining the box for the presence of the spermatophore gelatinous base and (b) checking the female's cloaca for a visible sperm cap. Following successful insemination, females were removed from pre-screening for one week to allow the sperm cap to dissolve before being presented with another mating opportunity. From the pre-screening data, 160 males and 160 females were selected for observed courtship trials, and shipped to Oregon State University. Animals were then housed at the same temperature and humidity, and maintained on a North Carolina photoperiod.

#### Gland removal and pheromone preparation

To prevent interference from endogenous male pheromones, all male salamanders had their mental glands surgically removed, based on the methods of Rollmann et al. [83]. Briefly, males were anesthetized in a mixture of 7% ether in water, and the exterior pad-like mental gland was removed from the dermis. Following surgery, males were placed in clean boxes and the wound was covered with a small piece of gauze containing

an antibiotic ointment. Males were allowed two weeks to heal before being used in any behavioral experiments. Gland removal surgery has no impact on male courting behavior, including slapping during TSW [79, 83, 85, 93, 208, 215, 216]. Pheromones were extracted from the excised mental glands following the methods of Rollmann et al. [83]. Briefly, mental glands were induced to secrete pheromones by incubation with 0.8 mM acetylcholine chloride in Amphibian Ringer's Solution for ~60 minutes. Whole pheromone extract was centrifuged at 10,000 x g for 10 minutes, the supernatant collected, and the centrifugation repeated before storage of supernatant at -80°C. Methods and animal care were approved by Oregon State University's Institutional Animal Care and Use Committee (ACUP 3007 and 4053 to L.D. Houck). Based on the methods of Chapter II, a natural mixture of PMFs containing fractions E through I (PMF-EFGHI) was purified from whole pheromone extract using strong anion exchange HPLC. As described in Chapter III, rPMF-G was expressed using the methylotrophic yeast *Pichia pastoris*. As a eukaryote, *P. pastoris* can perform many post-translational modifications, including production of disulfide bonds, which do not naturally occur in *E. coli* expression systems. All biochemical assays (reverse phase HPLC, tandem mass spectrometry, disulfide bond analysis, far-UV circular dichroism, multidimensional NMR) demonstrated that rPMF-G has a 3D structure identical to that of natural PMF-G. Both PMF-EFGHI and rPMF-G were prepared at 0.5 µg/µL in 0.5X phosphate buffered saline (PBS), the same concentration used for the PMF-EF behavioral trials in Houck et al. [93].

### Staged courtship trials

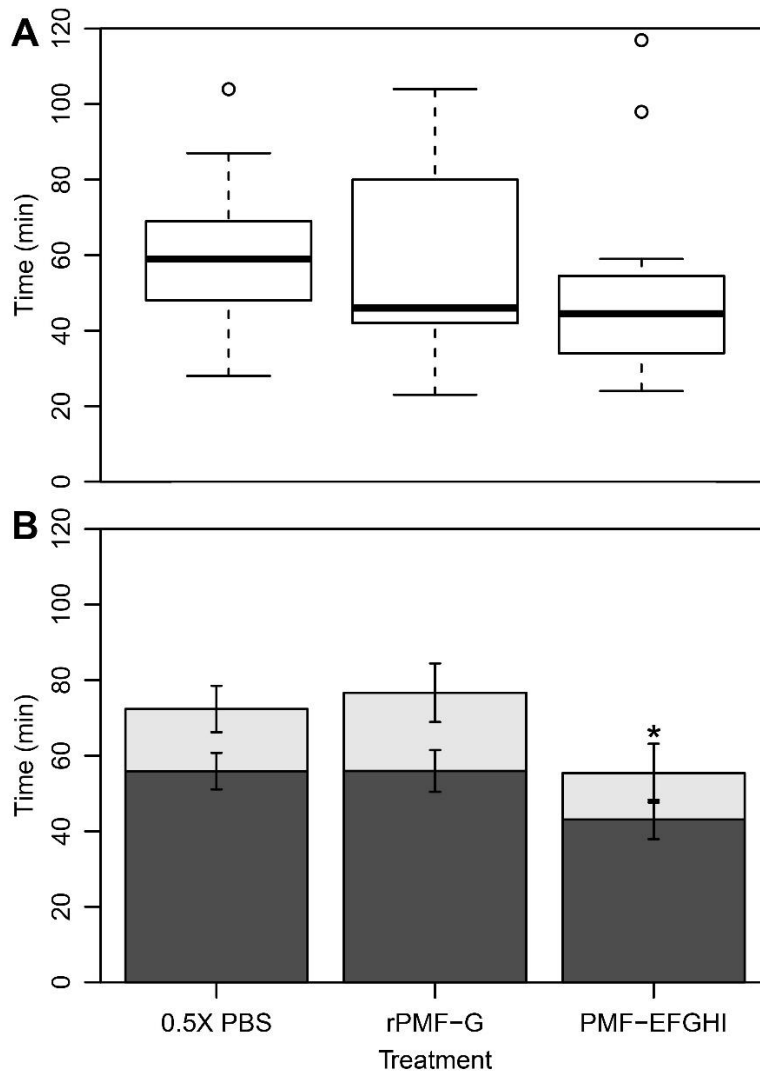
Effects of courtship pheromones on female receptivity were assayed by recording the courtship behaviour of females treated with exogenous application of one of the two pheromone solutions (PMF-EFGHI, rPMF-G) or a vehicle control (0.5X PBS). Methods were adapted from those used by Houck et al. [93]. Due to the limited number of successful courtship encounters in the lab, previous experiments could reliably only test two treatments. In order to maximize statistical power and test three different solutions, a mixed effects design was employed. On each observation night, male and female salamanders were paired as described during pre-screening trials. Salamanders were equally divided into four groups of 40 pairs, with each group observed one night per week for 5 weeks. During the first week of observations, within each group, male and female salamanders were randomly paired, and allowed 3 hours to initiate TSW. Once a pair entered TSW, if the male attempted to apply pheromone by slapping the female with his chin, 5  $\mu$ L of a random treatment was applied to the female's snout by micropipette. Subsequently, the female would receive additional 5  $\mu$ L aliquots of the same treatment after 5 min and 10 min following the initial slap (15  $\mu$ L total). In order to partially control for the physical stimulus of slapping, the data point was only considered valid if the male slapped at least 3 times (corresponding to the 3 pheromone applications). Time was recorded from the initiation of TSW to spermatophore deposition, and is herein defined as courtship time. Occasionally, application of the pheromone would "startle" the female salamander, and she would disengage from TSW, but then later resume the courtship with little-to-no persuasion from the male. Total time is defined as time in TSW and any TSW interruption times. For a given courtship encounter to be included in the analysis,

(1) the pair must have entered TSW and proceeded to spermatophore deposition, (2) the male salamander must have slapped the female's snout  $\geq 3$  times, (3) the female must have received all three 5  $\mu$ L aliquots of treatment, and (4) the total TSW interruption time must be less than 40 min. If a pair successfully completed a courtship encounter meeting these criteria, that male and female remained paired for all subsequent weekly observations. If that pair entered courtship a second time, the female would randomly receive one of the two remaining treatment groups – providing a partial repeated measures design. No pairs engaged in courtship three times and received all three treatments. For salamanders that had yet to successfully mate, the pairs were randomized each week until (a) they successfully mated and were paired for the remainder of the experiment or (b) the experiment was terminated. Pheromone treatments were coded prior to the start of the experiment such that all observers were blind to the treatment during observations. The purpose of these criteria was to maintain consistency with previous studies, best control for properly observing effects of pheromone treatment (i.e., ensuring females received sufficient dosage, physical stimulus of slapping corresponded to minimum number of pheromone applications), and maximize statistical power. While no pair received all three treatments, 10 out of 32 pairs received two separate treatments.

Data were analyzed using linear mixed-effects models with the R package *nlme* with parameter estimation by maximum likelihood. Both courtship time and total time were analyzed with treatment (0.5X PBS, rPMF-G, PMF-EFGHI) as a fixed effect and salamander pair as a random effect. The significance of treatment was tested by likelihood ratio test against an intercept-only null model, with individual effects of the three solutions examined post-hoc by t-test with corrected standard errors.

## Results

A total of 32 salamander pairs engaged in courtship successfully and met the aforementioned criteria. Ten of these pairs received two of the three possible treatments (42 total courtship encounters). The distribution of the data for the three treatments was approximately equal ( $n = 15$  for 0.5X PBS,  $n = 15$  for rPMF-G,  $n = 13$  for PMF-EFGHI). Examination of the data by box-plot revealed three outliers outside the mean  $\pm 1.5$  \* interquartile range (Fig 2A). With the outliers removed, the data for both courtship time (TSW without interruption time) and total time (TSW with interruption time) were fit by maximum likelihood to mixed effect models with and without treatment as a fixed effect. When compared by likelihood ratio test, treatment was significant in both cases, with  $p = 0.0394$  for courtship time and  $p = 0.0469$  for total time. Because courtship time is the measure used in previous behavioral studies [83, 85, 93, 208, 215], this is the more relevant measure of comparison and the primary focus of our analysis. For the 10 pairs that received two treatments, there was a significant effect from the inclusion of the random effect of pair ( $p = 0.026$ ). The mean courtship time for the negative control of 0.5X PBS was 55.9 min (range 28-87 min). Post-hoc t-tests comparing rPMF-G, with mean 56.0 min (range 23-107 min), to 0.5X PBS revealed no significant difference ( $p = 0.9930$ ). In contrast, PMF-EFGHI significantly decreased courtship time to a mean of 43.1 min (range 24-59 min.), implying an increase in female receptivity by ~23% ( $p = 0.0425$ , Fig 2B). Inclusion of the outliers did not alter the trend of the effects, but increased the standard error such that the effect of PMF-EFGHI over 0.5X PBS was not significant at  $p < 0.05$  ( $p = 0.1027$ ). As with previous studies, there was no significant difference in initiation frequency, number of interruptions, interruption time, or insemination success between the three treatment groups (data not shown).



**Figure 22. Effect of PMF components on female receptivity.**

(A) Box-and-whisker plot of courtship time for each treatment group, with whiskers defining the range of the data, and the box denoting the middle 50% (25%ile,  $Q_1$ , to the 75%ile,  $Q_3$ ). Outliers are defined as values outside  $Q_1 - 1.5*(Q_3-Q_1)$  or  $Q_3 + 1.5*(Q_3-Q_1)$ . (B) Maximum likelihood means of courtship time (dark grey) and total time (light grey)  $\pm$  SE for each treatment group. \*  $p < 0.05$ .

## Discussion

The aim of this study was to compare the effects of different PMF isoform mixtures on female behavior in *P. shermani*. We now report that the most abundant PMF isoform (PMF-G) had no effect on female mating behavior, but a more complete mixture of PMF isoforms (PMF-EFGHI) increased female receptivity, with an effect similar to that of whole pheromone extract and PRF. Both of these effects are in contrast to the previous study [93] which demonstrated that PMF-EF decreased female receptivity. It should be reiterated that while PMF-E and PMF-F together represent a highly complex mixture with >30 isoforms, the G, H, and I peaks are highly enriched for the 3 most abundant PMF isoforms. Therefore, the only difference between PMF-EF and PMF-EFGHI are these three major PMFs, with PMF-G in isolation having no measurable effect.

We propose two competing hypotheses to explain how PMF-EFGHI increased female receptivity: (1) additive effects from PMF-HI, (2) synergistic effects in the PMF isoform complex. For the first time, the effects of isoforms PMF-G, PMF-H, and PMF-I on female mating receptivity were evaluated, using the two mixtures PMF-EFGHI and rPMF-G. With regard to hypothesis 1, since the mean courtship time for rPMF-G and 0.5X PBS were nearly identical (~0.1 min difference), it is possible that isoforms H and I (either independently or in combination) caused a large increase in female receptivity, overcoming the negative effects of PMF-EF. According to hypothesis 2, the several PMF isoforms functionally complement one another such that PMF-EF is interacting with isoforms G, H, and I in a manner that increases female receptivity. The first hypothesis seems unlikely: in an additive effect model, with PMF-EF producing -23%, PMF-EFGHI producing +23%, and PMF-G having no effect, PMF-HI would generate +46%, nearly



twice what PRF or whole pheromone elicited. Also, in this model, it seems surprising that PMF-G, as the most abundant PMF isoform [Chapter II], would be functionally inert. The other hypothesis is that PMF isoforms act synergistically to enhance female receptivity. In this scenario, multiple isoforms of PMF are required to functionally complement one another, explaining why rPMF-G had no effect on female receptivity. At the same time, a mixture excluding the three most abundant isoforms may have a negative effect on female receptivity (i.e. females may perceive this as an abnormal/atypical signal and be “dissuaded” from courting with that particular male). Other pheromone systems, as well the study of communication at large, provide evidence for this second hypothesis. In ermine moths, females synthesize hydrocarbon chains modified with acetate or alcohol groups that attract potential mates; strong synergistic effects were observed when the alcohol and acetate versions were tested in combination [217]. Similar effects were observed in leafrollers (*Argyrotaenia sphaleropa*) where alcohol- and acetate-diene pheromones have synergistic interactions, which were further enhanced by specific ratios between the two versions [207]. In tungara frogs, the whine-chuck system of auditory communication has been a well-documented example of multimodal components functioning better in combination than individually [218]. Recently, this signal was demonstrated to be further modulated by both the presence and sequential order of a visual cue, vocal sac inflation [219]. Consequently, we hypothesize that PMF isoforms function synergistically to enhance female receptivity.

Whether these effects are additive or synergistic, the mechanism by which the interactive effects may be mediated remains unknown. All data to date suggest that PMF regulates female behavior by binding to vomeronasal type-2 receptors (V2Rs) in the

female VNO: (a) PMF-EF activated neurons in the female VNO [92], (b) in rodents, V2Rs are primarily responsible for binding peptide or protein pheromones [20, 30, 38], and (c) V2Rs are highly expressed in the *P. shermani* VNO [72]. It has been hypothesized that VNO receptors (like olfactory receptors) display monoallelic expression such that each neuron in the VNO expresses only a single receptor, limiting it to recognizing only a small number of potential pheromone ligands [5, 29, 30]. Vomeronasal receptors are generally highly sensitive, often requiring sub-picomolar concentrations. However, in mice, there is evidence that individual VNO neurons express multiple V2R receptors, and even neurons with the same receptors respond to different pheromone molecules [35]. In *P. shermani*, PMF-responsive VNO neurons may be expressing different receptors with each binding different PMF isoforms, and stimulation of multiple receptors by these different isoforms is necessary for proper signaling. Alternatively, VNO neurons may be independently activated by different PMF isoforms and the proposed interactive effects are mediated by downstream processing in the central nervous system. Recently, we determined the 3D structure of PMF-G by multidimensional NMR, and homology modelling suggested that most of the PMF sequence hypervariability was manifested in one loop of the protein structure – consequently, we hypothesized that this loop is critical for receptor interactions [Chapter III]. While PMF is monomeric in solution, it is also plausible that activation of female receptors may require forming a complex with multiple PMF isoforms. None of these hypotheses are mutually exclusive, but all may explain how sexual selection over the past 100 million years has driven massive gene duplication and positive selection of the PMF complex [96]. Future next-generation sequencing studies identifying potential pheromone

receptors in the *P. shermani* will facilitate understanding the mechanism by which PMF activates VNO neurons and modulates female behavior.

In conclusion, we have demonstrated that the hypervariable vertebrate pheromone Plethodontid Modulating Factor differentially regulated female mating behavior based on isoform composition. In contrast to PRF [208], a single isoform of PMF was not sufficient to affect female receptivity, while different combinations of more complex mixtures may have opposing effects based on the presence or absence of the three most abundant isoforms. We hypothesized that synergistic effects have promoted the large PMF isoform diversity present in *P. shermani*, and that female salamanders may evaluate isoform complexity through V2R receptors in their VNO. The requirement of a complete PMF mixture to increase female receptivity provides a functional basis for the perpetuation of extensive PMF gene duplication in *Plethodon* salamanders for the past ~100 million years [96]. Future studies will investigate the mechanistic basis by which female *P. shermani* react to differences in PMF isoform composition.

CHAPTER V - OLFACTORY EFFECTS OF PMF ISOFORM VARIABILITY IN  
THE RED-LEGGED SALAMANDER, PLETHODON SHERMANI

**Chapter Overview**

Chemical communication via chemosensory signaling is an essential process for promoting and modifying reproductive behavior in many species. During courtship in plethodontid salamanders, males deliver a mixture of non-volatile proteinaceous pheromones that activate chemosensory neurons in the vomeronasal epithelium (VNE) and increase female receptivity. One component of this mixture, Plethodontid Modulating Factor (PMF), is a hypervariable pheromone expressed as more than 30 unique isoforms that differ between individual males. Different isoform mixtures have variable effects on female mating receptivity, such that the aim of this study was to test the effects of isoform variability on VNE neuron activation using the agmatine uptake assay. All isoform mixtures activated a similar number of neurons (+350% over background) except for a single purified PMF isoform (+17%). These data further support the hypothesis that PMF isoforms act synergistically in order to regulate female receptivity, and different putative mechanisms are discussed. Over the past 100 million years, co-evolution with female receptors has been the likely driving force in promoting gene duplication and positive selection of the PMF gene complex in order to generate the diverse mixture we

observe today. Thus, more variable mixtures would be capable of stimulating any female in the mating population and maximize reproductive opportunities for male salamanders.

## Introduction

In many animals, chemical signaling via pheromone transmission is an essential means of communication to convey information between individuals, such as sex, species, and reproductive condition [10, 176, 177, 220]. Pheromones are chemicals transferred between members of the same species and elicit specific behaviors or endocrine responses [10]. Most pheromones, regardless of their chemical structure, are sensed via olfaction [204]. In tetrapod vertebrates, two discrete sets of neuronal epithelia are present, the main olfactory epithelium (MOE) and vomeronasal epithelia (VNE) [5]. Most studies of the VNE have been in rodents, where it is anatomically isolated from the MOE in an anterior bone-encapsulated cavity (the vomeronasal organ, VNO). In these mammalian models, two receptor families were identified in the VNE: vomeronasal type-1 receptors (V1Rs) and vomeronasal type-2 receptors (V2Rs) [29, 30]. Based on amino acid sequence and structural prediction, it was hypothesized that V1Rs would detect and bind small volatile chemosignals, while larger peptide or protein chemosignals would interact with the large, extracellular N-terminal domains of V2Rs. While few specific ligand-receptor pairs have been identified, all neurophysiological and biochemical studies to date support this hypothesis [23, 34, 35]. Compared to the MOE, where neurons express individual olfactory receptors (ORs) and each is highly tuned to detecting a small number of ligands [22, 28, 221], receptor expression is more variable between neurons in the VNE – as are the ligands they sense [35, 36, 221]. In insects, pheromone activity is often dependent on combinations of particular components being delivered in precise ratios [206, 207], and in vertebrates, protein pheromones often persist as multigene families with many expressed isoforms [23, 24, 38]. At least in mammals, this may relate

to receivers learning individual scent profiles [25, 26]; however, for senders, the whole signal may not be equivalent to the sum of its parts. To date, few studies have sought to investigate the interactive effects of such vertebrate protein pheromones.

As basal tetrapods, salamanders make an excellent non-mammalian model to study the evolution and function of pheromone signaling in vertebrates. For the salamander family Plethodontidae, male salamanders deliver non-volatile proteinaceous courtship pheromones to females in order to regulate courtship behavior and mating receptivity [87]. Preceding the annual mating season, plasma androgen levels rise in male salamanders and induce hypertrophication of a submandibular mental gland solely dedicated to the production of courtship pheromones [69, 214, 222]. In our principal model, the red-legged salamander (*Plethodon shermani*), male salamanders privately deliver pheromones to a female during an elaborate courtship ritual by “slapping” his hypertrophied mental gland to her nares [78]. Pheromones diffuse into the female nasal cavity, activate neurons in the VNE that project to regions of the brain classically involved in pheromone response, which stimulates changes in mating behavior [70, 83, 86]. During staged courtship trials, application of whole pheromone extract (WE) to female salamanders decreased tail straddling walk time by ~20% [85]. Chemical analysis of the pheromone extract revealed two major families of courtship pheromones: Plethodontid Receptivity Factor (PRF) and Plethodontid Modulating Factor (PMF). PRF is a 22 kDa protein related to IL-6 cytokines, while PMF is a smaller ~7 kDa protein related to the highly diverse three-finger protein (TFP) superfamily, which includes members such as snake venom cytotoxins and neurotoxins, the amphibian limb regeneration factor *Prod1*, and mammalian complement system receptor CD59 [69, 90].

Through gene duplication and pervasive positive selection, both PRF and PMF exist in the pheromone mixture as multi-isoform blends [94, 96]; however, compared to 3 PRF isoforms with ~95% sequence identity, individual male salamanders express more than 30 unique PMF isoforms with only ~30% sequence identity [91, Chapter II]. PMF isoform expression levels vary between males, with three isoforms (G, H, and I; together defined as PMF-GHI) almost always found in greater abundance than a mixture of >30 minor isoforms (referred to as PMF-EF; see Chapters II and IV for details). When applied to female salamanders, both a single PRF isoform and a mixture of all three recapitulated the effect of WE by decreasing courtship time by ~20% [83, 208]. In contrast, female behavioral response to PMF was different depending on the isoform composition: a single abundant PMF isoform (PMF-G) elicited no measurable effect, a mixture of minor isoforms (PMF-EF) increased courtship time, but a complete mixture of isoforms (PMF-EFGHI) decreased courtship time similarly to both WE and PRF [93, Chapter IV]. From these data, we hypothesized that PMF isoforms are functioning synergistically through an unknown neurophysiological pathway.

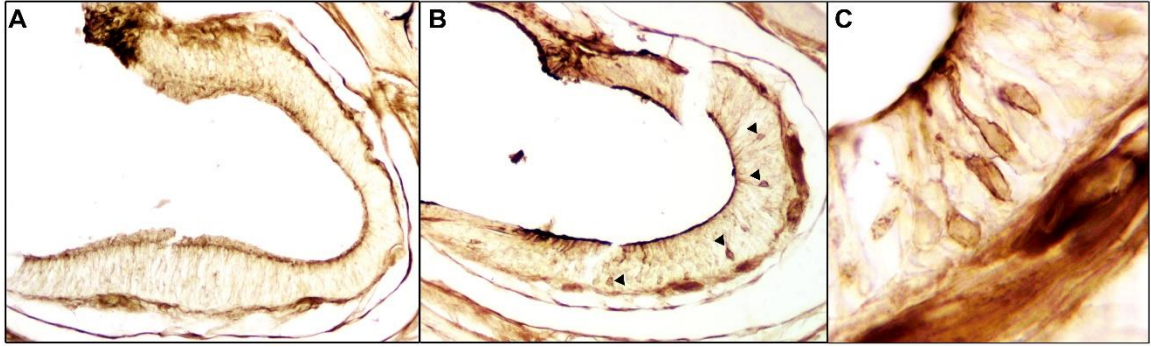
Salamanders possess a single nasal cavity with the MOE and VNE divided along the medial and lateral edges, respectively. The VNE, but not the MOE, is sexually dimorphic in plethodontid salamanders such that male salamanders have approximately twice as many VNE neurons [71]. RT-PCR and *in situ* hybridization studies have revealed high expression of V2Rs, but not V1Rs, in the *P. shermani* VNE [72]. In order to measure neuronal activation in this system, the arginine derivative agmatine (AGB) can be used as a tracer that passes through non-specific cation channels during membrane depolarization [223-226]. Co-application of pheromone and AGB to female salamanders



permits selective uptake of AGB into activated neurons, and following tissue fixation, sectioning, and immunohistochemical labeling, a permanent record of neuronal activation is obtained [70]. Previous studies suggested that PRF and PMF-EF activate different subsets of VNE neurons, yet independently only accounted for ~70% of the activated neurons observed when females were treated with WE [92]. No experiments were previously performed testing isoform mixtures containing the abundant isoforms PMF-GHI, consequently, it is hard to assess how the synergistic effects between abundant and minor PMF isoforms may be mediated through VNE signaling. Therefore, to better characterize the role of PMF isoform diversity in regulating female courtship behavior, the aim of this study was to test the efficacy of different PMF isoform mixtures on stimulating female VNE neurons using the AGB uptake assay.

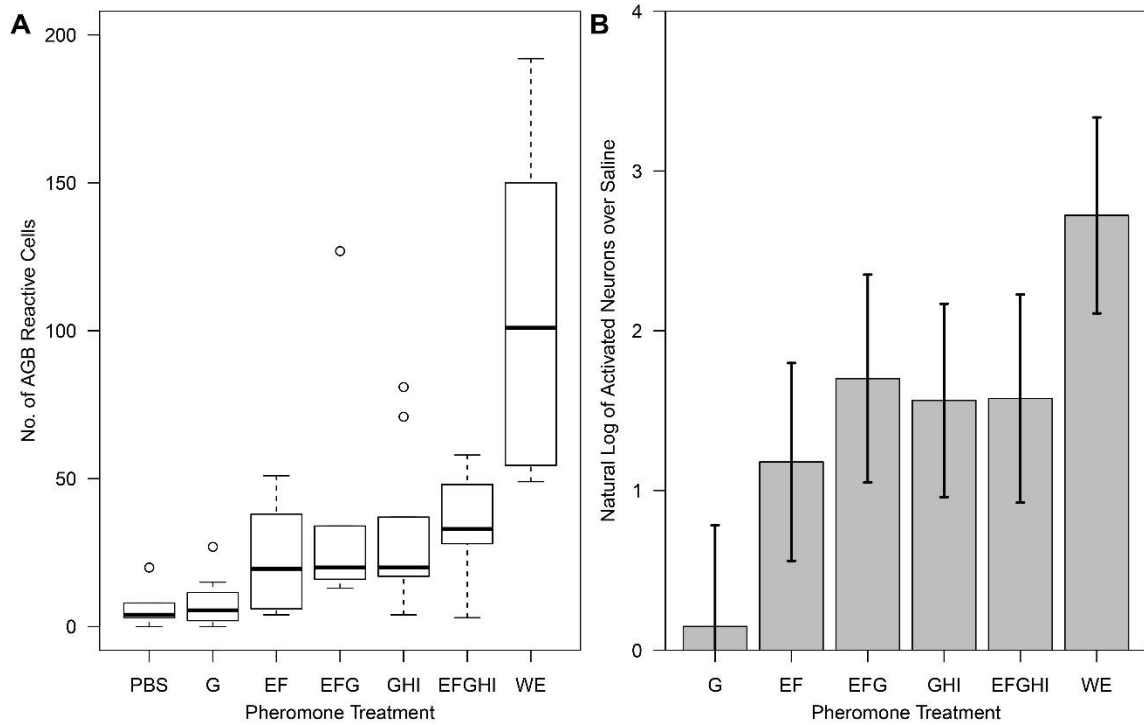
## Results

Fourteen blocks with 5 heads each were cryosectioned for a total of 70 animals, collected as four sets of sequential 20  $\mu\text{m}$  sections (80  $\mu\text{m}$  between each section in a set). However, due to a cryostat malfunction, a large percentage of sections were lost for three blocks, and were excluded from the analysis (effective  $n = 55$ ). Prior to full immunohistochemical (IHC) processing of one of the four slide sets, methods were optimized from the original Wirsig-Wiechmann et al. [70] protocol to reduce background staining and enhance resolution (Figure 23). Total neuron counts between treatments were visualized by box-and-whisker plots to reveal non-normal, overdispersed distributions (Figure 24A). The data were then analyzed using generalized linear models with negative binomial distributions, with and without pheromone as a fixed effect. When compared by chi-squared test, this fixed effect was significant ( $\chi^2(6) = 38.5$ ,  $p = 8.8 \times 10^{-7}$ ). For the individual pheromone treatments, 0.5X PBS activated the fewest neurons. Model estimates for each treatment group (log transformed means) were compared to 0.5X PBS by z-test with corrected standard errors for overdispersion ( $\tau = 1.716$ ) (Figure 24B and Table 8). Means of AGB reactive neurons were reported as both total numbers and percent increase over vehicle/0.5X PBS treatment (similar to Wirsig-Wiechmann et al. [70, 92]). PMF-G elicited no significant response ( $\bar{x} = 8.1$ ,  $z = 0.24$ ,  $p = 0.81$ ).



**Figure 23. Agmatine immunohistochemistry.**

Comparison of vomeronasal tissue from females treated with 0.5X PBS (A) and WE (B), with arrows pointing to immunoreactive neurons. (C) Higher magnification of four AGB neurons in close proximity.



**Figure 24. Model estimates of AGB count data.**

(A) Box-and-whisker plots comparing the 7 pheromone treatments, which reveals non-normal overdispersed distributions. (B) Bar graph of negative binomial GLM estimates  $\pm$  standard errors for PMF and WE treatments above 0.5X PBS levels.

**Table 8. Summary of GLM estimates by pheromone treatment.**

<b>Treatment</b>	<b>Mean AGB Reactive Neurons</b>	<b>Effect above vehicle</b>	<b>t-value</b>	<b>p</b>
0.5X PBS	7.0	-	-	-
PMF-G	8.1	+17%	0.236	0.81
PMF-GHI	33.4	+378%	2.586	0.0097
PMF-EF	22.8	+225%	1.903	0.057
PMF-EFG	38.3	+448%	2.614	0.0090
PMF-EFGHI	33.8	+383%	2.419	0.016
WE	106.5	+1421%	4.439	9.0 x 10 <sup>-6</sup>

All PMF treatment groups activated a significantly greater number of neurons compared to saline at  $p < 0.05$  except for PMF-EF ( $\bar{x} = 22.8$ , +225%,  $z = 1.90$ ,  $p = 0.057$ ); however, the effect of PMF-EF over vehicle was approximately twice of that observed in Wirsig-Wiechmann et al. [92] (+124%), and failure to achieve significance at  $p < 0.05$  may be partially explained by the number of levels in the treatment variable (7 groups versus only 2 in Wirsig-Wiechmann et al. [92]). Except for PMF-G, there was no statistical difference between any of the PMF isoform mixtures. Whole extract activated significantly more neurons than all other treatments ( $\bar{x} = 106.5$ ).

## Discussion

In the current study, we evaluated the response of female olfactory neurons to different isoform combinations of Plethodontid Modulating Factor, a hypervariable salamander courtship pheromone. While both PRF and PMF are independently capable of reducing courtship time, PMF is unique in its ability to increase or decrease female receptivity depending on the isoform composition [93, Chapter IV]. When the different mixtures of PMF isoforms were tested in the current study, all but the single isoform (PMF-G) elicited a similar response of ~350% increase over 0.5X PBS. The non-significant +17% increase in neuronal activation from PMF-G was consistent with courtship trials where there was no detectable response over vehicle [Chapter IV]. However, the mixture of the three highly abundant isoforms, PMF-GHI, elicited a robust response, suggesting that (1) PMF-H and PMF-I are independently producing large significant effects, or (2) PMF isoforms act synergistically to stimulate females. While we cannot definitively exclude the hypothesis that PMF-HI has a large independent effect, this seems unlikely for several reasons. First, PMF-G is consistently more abundant than PMF-H or PMF-I in the pheromone extract [Chapter II]. Second, the three-dimensional structure of PMF-G was recently solved by NMR, with homology modeling suggesting that PMF-H and PMF-I are structurally very similar to PMF-G such that they may bind similar receptors [Chapter III]. Third, the only solution missing PMF-G (PMF-EF) activated the fewest neurons, and while not statistically significant, it is interesting that the same mixture plus PMF-G (PMF-EFG) activated approximately twice as many neurons. Consequently, these data further support the hypothesis that synergistic interactions between different PMF isoforms are required to enhance female receptivity.

One quality of PMF that seems common to multiple vertebrate protein pheromones is evolutionary histories with pervasive gene duplication and neofunctionalization. In mice, both the major urinary proteins (MUPs) and exocrine gland-secreting peptide (ESP) families are highly polymorphic, with isoforms variably expressed both within and between inbred strains [24]. Many of these isoforms may have unique functions: specific MUPs are involved in regulating female receptivity and promoting male aggressive behavior [25, 27]; and ESP1, but not other ESPs, increased female receptivity [23]. In contrast to these mouse systems, where the pheromones are dispersed into the environment as part of general bodily fluids, the mental gland of plethodontid salamanders is only used to privately deliver signals to a single female during courtship [45]. Based on the timing of mental gland development and its use during tail-straddling walk, these pheromones presumably serve no functions beyond regulating behaviors that may impact reproductive success [83, 227]. Therefore, unlike the mouse pheromones with disparate functions, the many isoforms of both PMF and PRF may be acting on overlapping biological pathways, which fits within the “redundant-signal” hypothesis [228, 229].

However, the precise mechanism by which these PMF molecules are interacting to activate VNE neurons and affect female behavior remains unclear. All available evidence supports the hypothesis that PMF binds to V2Rs in the female VNE [72, 92]. In mice, pheromone binding to V2Rs induces the IP<sub>3</sub> signaling cascade that eventually leads to opening of the transient cation channel TRPC2, allowing an influx of Ca<sup>2+</sup>, membrane depolarization, and signal transduction to the central nervous system [5, 33]. The proposed mechanism of AGB uptake during neuronal depolarization is via open TRPC2



channels following pheromone binding, and TRPC2 is highly abundant in the *P. shermani* VNO [72]. Both NMR and analytical ultracentrifugation studies demonstrated that PMF is monomeric, both as a single isoform and a complex mixture [Chapter III]. While some TFPs can form dimers [125], it is likely that the high negative charge density of PMF precludes such interactions [Chapter III]. One possible explanation may be that different PMF isoforms form a multimeric complex with a single receptor molecule. Compared to other members of the three-finger protein superfamily, PMF-G has a novel protein topology and three-dimensional structure that permits greater backbone flexibility in one of the three fingers, which is also the most variable and rapidly evolving segment in the PMF gene complex [Chapter III]. With less structural variation in the rest of the molecule, this may allow any number of different isoforms to dock with target receptors through the conserved regions, but activation may be dependent on proper interactions from the variable third fingers of the necessary isoforms in the proper orientation. These potential kinetic constraints may help explain the selection pressure for the large number of PMF isoforms (>50 expressed between different male *P. shermani*) and their large abundance in the mental gland (~5:1 stoichiometry compared to PRF, and ~50% of total pheromone in whole extract) [91, Chapter II].

However, one previously unexplored hypothesis is that some isoforms of PMF may be functioning as signature mixtures rather than individual pheromones. The classic Karlson and Luscher (1959) definition of a pheromone is a molecule that elicits a pre-programmed behavioral and/or neuroendocrine effect. In contrast, signature mixtures, proposed by Wyatt (2010), are variable sets of cues that provide information on identity of specific individuals in a population, but elicit no innate, pre-programmed response in

receiving individuals. What makes these terms somewhat confounding is that, under the proper conditions, signature mixtures can function similarly to pheromones. For example, the mouse pheromone darcin, a MUP isoform delivered alongside many other proteins and volatile odorants in mouse urine, provokes learning of male odor profiles in virgin females such that, upon subsequent exposure to the same odor profile minus darcin, female receptivity increases. In this system, females are now conditioned to respond to particular markers, but without initial co-exposure with darcin, there is no learning of these individual cues [25, 26]. In *P. shermani*, male salamanders express unique, individual PMF profiles (visible by HPLC), although the exact composition of these mixtures is hard to dissect other than PMF-G, H, and I are universally expressed at proportionally high levels [91]. Comparison of PMF cDNA sequences revealed that while there exists more than 99 unique putative isoforms in a single population, sequences were more accurately clustered into 13 common archetypes with only 1-2 SNPs varying between the sequences within each cluster [Chapter II]. Each of these PMF isoform clusters may be performing unique roles, either by targeting different classes of receptors or functioning in part as signature mixtures to convey identity. Because learning is an active component of signature mixture response [204], and *P. shermani* are difficult to breed under laboratory conditions (L.D. Houck, personal communication), it is a challenge to test these variable pheromone components on virgin females. However, female plethodontid salamanders can distinguish between the odors of individual conspecifics [230]. Additionally, learning of predator cues has been tested in multiple amphibian species [231-233], and in mice, many of these heterospecific signals are mediated through specific vomeronasal receptors (both V1Rs and V2Rs) [20]. Common

isoforms such as PMF-G may be acting as more traditional pheromones, while other variable PMFs serve more as individual cues in signature mixtures. This hypothesis is complementary to the longer standing view that PMF has been subject to extensive gene duplication and pervasive positive selection in order to expand the functional breadth of PMF as a “pheromone” [96, Chapter II].

In summary, PMF is a hypervariable vertebrate courtship pheromone that differentially regulates female courtship behavior at least in part through variability in isoform composition. For the first time, we demonstrated that a single isoform of PMF was unable to significantly activate neurons in the female VNE; however, any mixture of PMF containing 3 or more isoforms elicited a similar response of ~350% activation over vehicle. The exact receptors and mechanisms mediating this response is still unclear, and may, in part, involve isoforms forming multimeric complexes with V2Rs, VNE neurons expressing multiple V2Rs, and/or learning in female salamanders as part of signature mixtures. Future studies will seek to further characterize the molecular architecture of the *P. shermani* VNE, identifying the specific PMF receptors, and elucidating the binding mechanics in order to better understand the evolutionary forces that have driven exacerbated gene duplication and positive sexual selection on PMF over the past 100 million years.

## **Experimental procedures**

### Animal collection

*P. shermani* salamanders were collected from a single site in Macon Co., North Carolina, USA (35°10'48" N, 83°33'38" W) during their annual breeding season in August 2011, and housed at Highlands Biological Station for duration of the experiment. Adult salamanders were identified and sexed based on a well-developed mental gland in males and large ova in females. Animals were individually housed 15-18°C and ~70% humidity in clean plastic boxes (17 x 9 x 13 cm) lined with a damp paper towel, and a second damp crumpled paper towel for refugia. Once per week, salamanders were transferred to clean boxes with new substrate and fed 2 waxworms (*Galleria mellonella*). All animals were collected under permits obtained from the North Carolina Wildlife Resources Commission, and all animal methods were approved by Oregon State University ACUP #4053 to L.D. Houck.

### Preparation of male pheromones

Whole pheromone extract (WE) was collected from male salamanders following the methods of Rollmann et al. [83]. Briefly, male salamanders were anesthetized in a mixture of 7% ether/water for ~7 minutes, and the mental gland surgically removed from the dermis using iridectomy scissors. Mental glands were then incubated in acetylcholine chloride (0.8 mM in amphibian Ringer's solution) for 60 min to induce pheromone secretion. Mental glands were removed by two rounds of centrifugation at 10,000 x g for 10 min, and pheromone extracts stored at -80°C till further processing at the University of Louisville. WE was prepared by ultrafiltration and standardized at 2.0 mg/mL in 0.5X phosphate buffered saline (PBS). PMF-EF and PMF-EFGHI were prepared based on the

methods of Chapter II using high performance liquid chromatography (HPLC). Briefly, WE was subjected to strong anion exchange HPLC in order to collect the PMF-EFGHI fraction. PMF-EF and PMF-GHI were resolved by a second round of strong anion exchange HPLC with a shallower elution gradient. PMF-GHI was purified to >99% purity by one round of reverse phase HPLC such that the solution only contained the three most abundant PMF isoforms. In order to test the effects of PMF-G and PMF-EFG, highly purified recombinant PMF-G was prepared based on the methods of Chapter III using the methylotrophic yeast *Pichia pastoris*. All biochemical and structural studies have confirmed that recombinant PMF-G has an identical sequence and 3D structure to natural PMF-G, and is thus suitable for bioassays. PMF-EFG was prepared by mixing natural PMF-EF with recombinant PMF-G in a 7:2 ratio (which approximates natural levels, based on integration of peak areas observed by strong anion exchange HPLC). All PMF solutions were standardized using ultrafiltration to 0.5 mg/mL in 0.5X PBS.

#### AGB uptake assay and immunohistochemistry

The AGB uptake assay was performed based on the methods from Wirsig-Wiechmann et al. [70]. Pheromone solutions were mixed 1:1 with 6 mM AGB (Sigma Aldrich, St. Louis, MO) in 0.5X PBS. A total of 70 adult gravid female salamanders were placed in new, unused Tupperware sandwich boxes lined with a single damp paper towel and allowed to acclimate for 30 min. Each salamander received one of 7 different treatments (n=10 per treatment): 0.5X PBS (negative control/vehicle), PMF-G, PMF-GHI, PMF-EF, PMF-EFG, PMF-EFGHI, or WE (positive control). Two microliters of pheromone/AGB was applied to the female's nares every 2 minutes for a total 40 min (20 applications), followed by 3 rinses with 5  $\mu$ L 0.5X PBS over ~5 minutes. Females were

then rapidly decapitated, the lower jaw removed, and heads incubated overnight in 10mL 4% paraformaldehyde/2.5% glutaraldehyde (in 150mM sodium chloride/100mM sodium phosphate, pH 7.4). Heads were then decalcified using DeCal (Decal Corporation, Congers, NY) for 3 days, cryoprotected in 30% sucrose for 2 days, and embedded using Optimal Cutting Temperature (OCT) media (Sakura-Finetek, Torrance, CA). A total of 14 blocks were prepared with 5 heads each, and stored at -80°C prior to cryosectioning. Heads were sectioned coronally at a thickness of 20 µm, and thaw mounted onto superfrost plus slides pre-coated with polylysine. Sections were collected in four sets such that each section in a set was separated by 80 µm. Slides were stored at -80°C prior to immunohistochemistry. One slide set was used to optimize IHC conditions from the original Wirsig-Wiechmann et al. [70] protocol to reduce background staining (adjustments were principally in the concentration of detergents used in pre-incubation and washing steps): after equilibration to room temperature, slides were washed five times for 5 min each in 1X PBS, preincubated in 1% normal goat serum/1X PBS/0.2% Triton X-100/0.5% Tween-20/0.02% azide for 30 min, and incubated overnight in rabbit anti-AGB (EMD-Millipore, Billerica, MA) diluted 1:2000 in 1% normal goat serum/1X PBS/0.1% Triton X-100/0.05% Tween-20/0.02% azide for three days. Slides were then washed five times for 5 min each in 1X PBS/0.05% Tween-20 (PBST), incubated for 30 minutes in biotinylated goat anti-rabbit IgG (Thermo-Pierce, Rockford, IL) diluted 1:500 in PBST, washed five times for 5 min each with PBST, incubated in 0.5X ultra-sensitive ABC peroxidase staining reagent (Thermo-Pierce) in PBST for 30 minutes, washed three times for 5 min each with PBST, twice with 1X PBS for 5 min each, and developed with

metal enhanced DAB (Thermo-Pierce) for 5 min before serial dehydration and coverslipped with Permount (Fisher Scientific, Pittsburgh, PA).

#### Histological and statistical analyses

Slides were visualized and imaged using an Olympus microscope with an attached 9 megapixel digital camera. AGB reactive neurons were counted in VNE tissue for all sections from both left and right nasal cavities. Count data was analyzed using generalized linear models with negative binomial distributions using the R function `glm.nb` in the package *MASS*. The effect of pheromone treatment was evaluated by likelihood ratio test against an intercept-only model, with individual effects of the seven solutions/levels examined *post-hoc* by z-test with corrected standard errors.

CHAPTER VI - GENE EXPRESSION DURING SEASONAL ORGANOGENESIS:  
PROFILING TRANSCRIPTOME CHANGES IN AN ANNUAL CYCLE OF  
GLAND DEVELOPMENT AND HYPERVARIABLE PHEROMONE SYNTHESIS

**Chapter Overview**

Cell differentiation is an essential biological process for tissues and organs to adopt specific functions, resulting from the ordered, synchronized expression of hundreds to thousands of genes. A detailed analysis of the genes expressed at different stages of differentiation can be highly informative to determining a tissue's function, regulatory proteins, and evolutionary origin. In the red-legged salamander, male salamanders develop a highly specialized pheromone gland (mental gland) that, once fully developed, redirects nearly all of its transcriptional and translational machinery to the production of pheromone. The mental gland undergoes the unique phenomenon of seasonal organogenesis: an annual cycle where the mental gland undergoes extreme hypertrophication, persists for ~2 months, and then completely resorbs. At least one family of proteinaceous pheromones, Plethodontid Modulating Factor (PMF), has experienced a highly unusual history of disjunctive evolution: following gene duplication, high rates of positive selection have diversified the coding regions, yet unknown forces have conserved the untranslated regions by purifying selection. However, it remains unclear how the mental gland hypertrophies, redirects its molecular machinery exclusively towards pheromone synthesis, and how these UTRs may be



involved in this process. Using next-generation sequencing, we prepared a *de novo* transcriptome of the mental gland at six stages of development. Combining differential expression analysis and immunohistochemistry, it was ascertained that the mental gland initially adopts a highly proliferative, almost tumor-like phenotype, then proceeds to rapidly increase the amount of pheromone mRNA in order to drive pheromone synthesis. One likely player in this transition is Cold Inducible RNA Binding Protein (CIRBP), which selectively and cooperatively binds the highly conserved PMF 3' UTR. CIRBP, along with other stress response proteins, have seemingly been co-opted to perform novel functions in the development of the mental gland and play key roles in regulating pheromone synthesis. This study illustrates how post-transcriptional regulators of gene expression may act as strong selective tethers even for rapidly evolving proteins, and could have key implications for studying the co-evolution of such proteins and their regulators. Similarly, the mental gland could be a powerful emerging model of regulated proliferation and subsequent resorption with potential links to cancer biology.

## **Introduction**

The processes of cellular differentiation and tissue remodeling are ubiquitous for multicellular organisms. The transition of cells from totipotency to terminal differentiation are the result of highly coordinated gene networks and changes in expression patterns [234]. Cell differentiation can be induced by a range of signals, including cytokines, hormones, cell contact, external stressors, and extracellular matrix composition [235-239]. Generally, target genes are regulated through activation of specific transcription factors and/or post-transcriptional mechanisms, such as RNA-binding proteins and microRNAs, which allow selective translation of key proteins [240, 241]. In the case of endocrine and exocrine glands, differentiation includes the activation of genes whose products are distributed and function outside of the originating tissue. Study of the gene networks involved in the differentiation of exocrine and endocrine glands may provide key insights into better understanding the regulation, biogenesis, and downstream activities of the secreted products.

For nearly all animals, pheromones are important chemical signals that can mediate social and reproductive behaviors [10, 177, 220]. Most pheromones elicit their effects by first binding to cell surface receptors on specialized olfactory neurons [5, 204]. While pheromones may provide honest or dishonest information to receivers [203], underlying the transmission of a particular pheromone exists layers of regulation that modulate its synthesis. Therefore, the social behaviors which result from pheromone signaling are indirectly mediated by regulation of pheromone synthesis itself. To date, the most well characterized pheromone models are the volatile hydrocarbons of insects, synthesized through long enzymatic cascades [16]. Consequently, tracing the specific

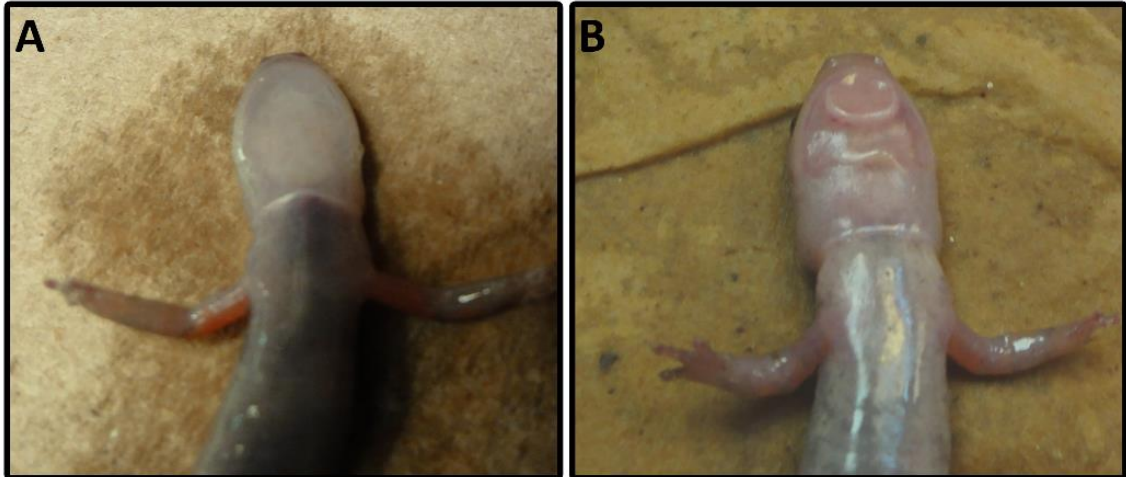
mutations and genetic characteristics that modulate their production can be both tedious and difficult. In contrast, protein pheromones, as direct gene products synthesized via transcription/translation, provide a much more tractable model for studying how natural or sexual selection may alter pheromone genes and their regulatory elements.

Interestingly, as secreted products that do not directly interact with the sender's metabolism, purifying selection may be weaker on protein pheromone genes, leading to many having histories of accelerated evolution [24, 94, Chapter II]. However, despite overwhelming evidence that both the composition and ratio of components in pheromone mixtures are often essential for proper biological activity [209-213], there have been no specific studies addressing the regulation of protein pheromone biosynthesis at different levels of mixture complexity.

For more than 100 million years, plethodontid salamanders have utilized a system of nonvolatile proteinaceous courtship pheromones to regulate female reproductive behavior [87]. In the species *Plethodon shermani*, during a courtship behavior known as tail-straddling walk, male salamanders will privately deliver pheromones to a female by "slapping" a large pad-like gland on his chin (the mental gland) to the female's nares [85]. Pheromones diffuse into the female nasal cavity, bind to receptors on neurons in the vomeronasal organ, activate regions of the brain involved in pheromone response, and regulate female mating behavior [70, 73, 83, 86, 92]. Chemical analysis of the pheromone extract revealed two major components: Plethodontid Receptivity Factor (PRF) and Plethodontid Modulating Factor (PMF). PRF is a 22-kDa protein with sequence similarity to IL-6 cytokines [83], while PMF is a 7-kDa protein related to the highly diverse three-finger protein (TFP) superfamily that includes snake venom neuro- and cytotoxins, the

complement receptor CD59, the human Ly6 antigen, the urokinase receptor uPAR, and the amphibian regeneration factor Prod1 [90]. When experimentally applied to female salamanders, both PRF and PMF altered the length of courtship time [83, 93, 208, Chapter IV]. Analysis by high performance liquid chromatography (HPLC) and mass spectrometry (MS) revealed multiple isoforms of both PRF and PMF; however, compared to 3 highly conserved PRF isoforms (>95% identity), individual male *P. shermani* expressed more than 30 diverse PMF isoforms (~30% identity) [Chapter II]. The ratios of different PRF and PMF isoforms are quite variable between male salamanders [91], and the source of isoform sequence diversity is primarily from gene duplication [Chapter II]. Examination of PRF and PMF sequences from 29 plethodontid species revealed that both genes have been subjected to pervasive positive selection [94, 96]. Sampling from these many species by RT-PCR was facilitated by the unique quality that both PRF and PMF have unusually conserved, AU-rich untranslated regions (UTRs). The contrast is most striking for PMF: compared to the ~30% amino acid identity between isoforms, the average conservation for both the 5' and 3' UTRs is ~98%. We proposed that PMF genes have been subjected to disjunctive evolution: the coding regions of the many PMF gene copies have been under positive selection in order to expand the functional breadth of PMF as a pheromone, while purifying selection on the UTRs permitted coordinated, synchronized expression of the many PMF isoforms [Chapter II]. The mechanism by which these UTRs mediate such expression remained unknown, but it was postulated that RNA binding proteins are likely involved. Limited DNA sequence data and no proteomic information pertaining to the cytosolic environment of the mental gland precluded further analysis.

The mental gland of *P. shermani* is an unusual tissue that undergoes the unique phenomenon of seasonal organogenesis. During the non-breeding season, it is absent from male salamanders; however, presumably in response to elevated plasma androgens [214, 222], the gland hypertrophies over ~2 months and develops into a large pad-like structure solely dedicated to the production of protein pheromones (Figure 25). Once the gland has fully developed, PRF and PMF represent ~85% of the secreted protein [69]. Similarly, cDNA sequence analysis revealed that ~70% of the total mRNA coded for pheromones [89]. Following the end of the courtship season, the gland seemingly resorbs and a new one forms each subsequent year. Post-embryonic organogenesis is rare outside of plants [242, 243], with the best models being gonad development in nematodes [244, 245], and perhaps shares qualities with insect and amphibian metamorphosis [246]. However, none of these examples have the dynamic but highly regulated annual cycle that results in an organ performing such a specialized function. It is noteworthy that surgical removal of the mental gland is followed by rapid wound healing which prevented gland regrowth in subsequent years (L.D. Houck and R.C. Feldhoff, personal communication), though animals will continue to mate normally. This suggests that there likely exist androgen-sensitive precursor cells embedded in the dermis. Given that the transcriptional and translational machinery of fully developed mental glands are directed almost exclusively towards pheromone production, there must exist some earlier developmental phase characterized by greater mitosis and/or general growth in order to form the glandular structure. The unusually conserved pheromone UTRs may be critical for regulating (1) the transition from gland hypertrophication to pheromone synthesis, and (2) the precise composition of proteins found in the pheromone mixture.



**Figure 25. Mental gland hypertrophication.**

Comparison of male *P. shermani* from (A) late May (non-breeding condition) and (B) mid-August (breeding condition), with the mental gland being the large pad-like structure on the male's lower jaw in panel B.

Previous studies of the mental gland have been limited to examining basic histology [45, 247], establishing that plasma androgens influence development [214, 222], and characterizing the pheromone composition [69, 91, Chapter II]. Therefore, to better understand the developmental process of this novel tissue and how pheromone synthesis may be regulated to produce variable signals that regulate female reproductive behavior, the aim of this study was to characterize the transcriptome of *P. shermani* mental glands at different stages of gland development.

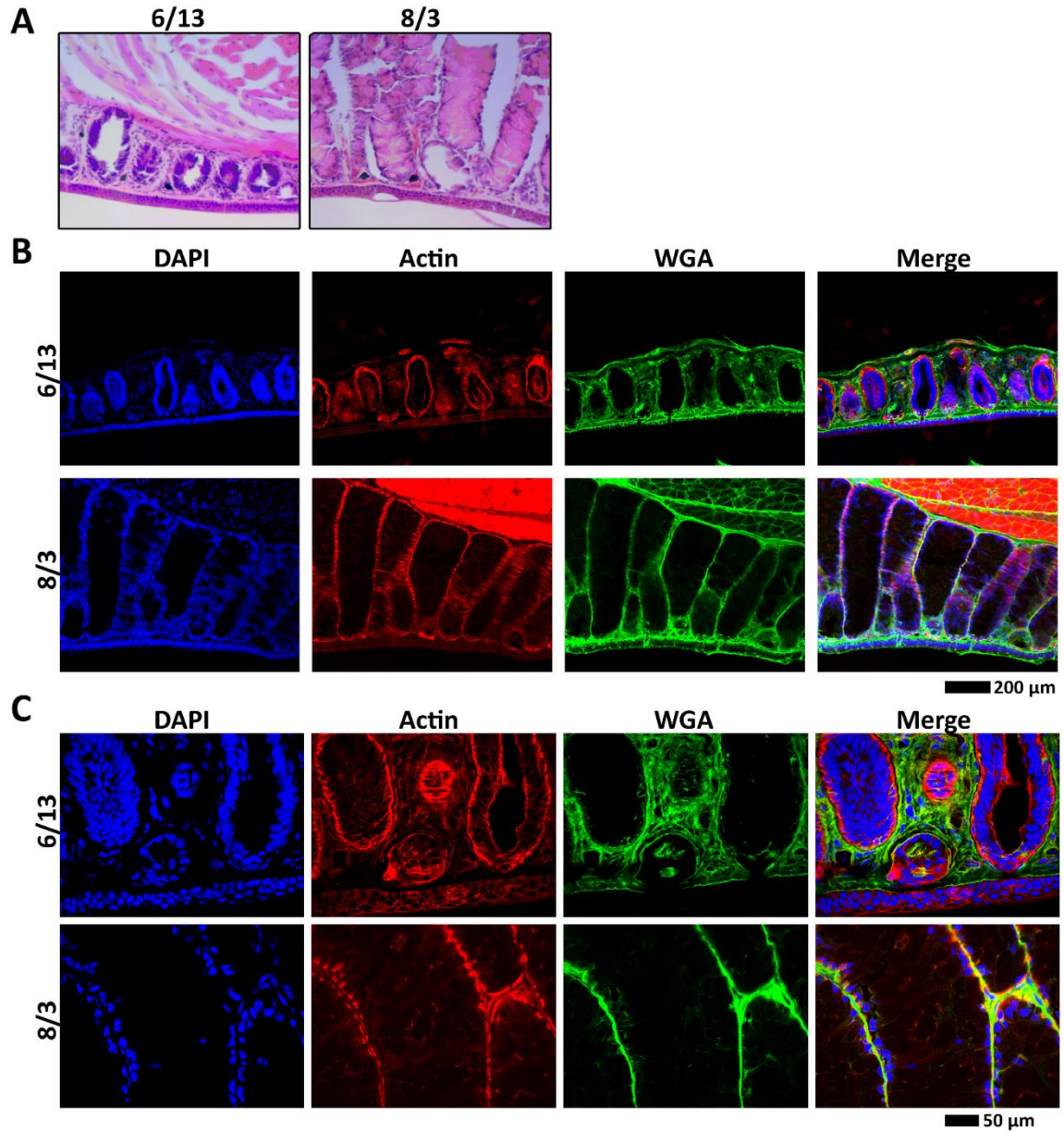
## Results

### Qualitative observation of mental gland development

To examine mental gland development, we performed a time-course experiment based on initial observations by Woodley [214]. In that study, plasma testosterone for male *P. shermani* was found to increase from mid-June (~100 ng/mL) to late August (~350 ng/mL), and started to decline by early September. It was also observed that ~90% of animals had thin mental glands by late June, and the glands were enlarged on ~50% of animals by mid-July. To acquire a comprehensive sample set that spanned most phases of development, we collected male *P. shermani* approximately every 3 weeks from late May through mid-September in 2010 (5/29, 6/19, 7/10, 8/1, 8/21, 9/11). Our qualitative observations were similar to those of Woodley [214]. On 5/29, male salamanders had visibly different skin pigmentation near the mentum compared to females, including 2 out of 15 males with extremely faint “outlines” of a mental gland. By 6/19, this proportion had increased such that 8 of 10 collected males had visible outlines and/or thin mental glands. On 7/10, all males collected had visible and protruding mental glands. At this time point, 8 mental glands were incubated in acetylcholine to induce pheromone secretion; analysis of these samples by reverse-phase high performance liquid chromatography (RP-HPLC) revealed normal proportions of PRF and PMF, however, the protein concentration of the samples were ~33-50% of levels normally observed in mid-August. Glands from the last three time points (8/1, 8/21, and 9/11) were visibly well-developed and had normal levels of pheromone. For a second data set, we collected animals from mid-June (6/13) and early August (8/3) in 2013, and observed similar rates of gland development, suggesting a tightly regulated seasonal pattern of development

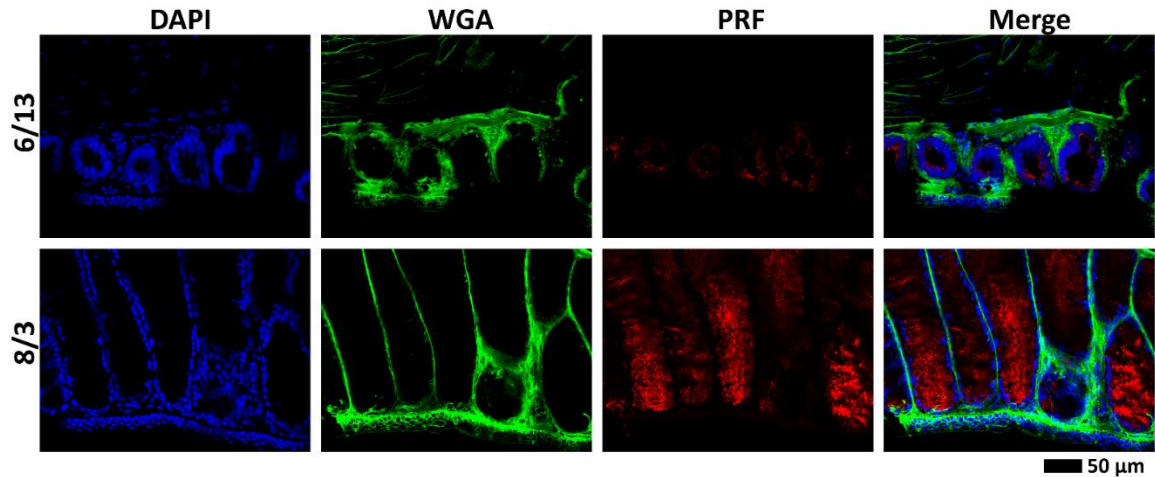


within the population. Histological analysis of the tissue by hematoxylin/eosin staining showed that the glandular tissue was immediately under the epidermis (~2-3 cells thick). The mental gland was structured as cylindrically shaped bundles with nuclei localizing exclusively near the periphery. At the 8/3 time point, there was bright eosin staining in the center of these bundles; yet on 6/13, only nuclear staining was visible for the smaller, more tightly packed bundles (Figure 26A). Confocal fluorescence microscopy was performed using a combination of dyes to stain the nucleus (DAPI), actin fibers (phalloidin), and glycoproteins associated with the plasma membrane and/or extracellular matrix (ECM) (wheat germ agglutinin). For 8/3, the strongest actin staining was along the periphery (near the nuclei), with light, diffuse staining and a few small fibers visible in the eosin-stained space; for 6/13, actin was only found adjacent to DAPI-stained nuclei (Figure 26B-C). Lectin staining suggested condensation and/or degradation of much of the ECM as the gland expanded (Figure 26B). Immunohistochemical labelling with anti-PRF produced a strong, punctate pattern throughout the eosin-positive space, and may represent stored secretory vesicles/granules. There was minimal PRF staining for 6/13, both in intensity and volume (Figure 27). These data suggest that the mental gland initially forms as a tightly packed mass of cells with little cytoplasm, and upon induction of pheromone synthesis, the cells swell with large volumes of pheromone, adopt a columnar shape, and the ECM condenses and/or degrades in order to support the enlarged cells.



**Figure 26. Mental gland histology.**

Comparison of mental glands from male *P. shermani* from two stages in mental gland development (6/13 and 8/3) using (A) hematoxylin and eosin staining, (B-C) fluorescent confocal microscopy with dyes labeling the nucleus (blue), actin cytoskeleton (red), and ECM (green) at 10X (B) and 40X (C) magnifications.

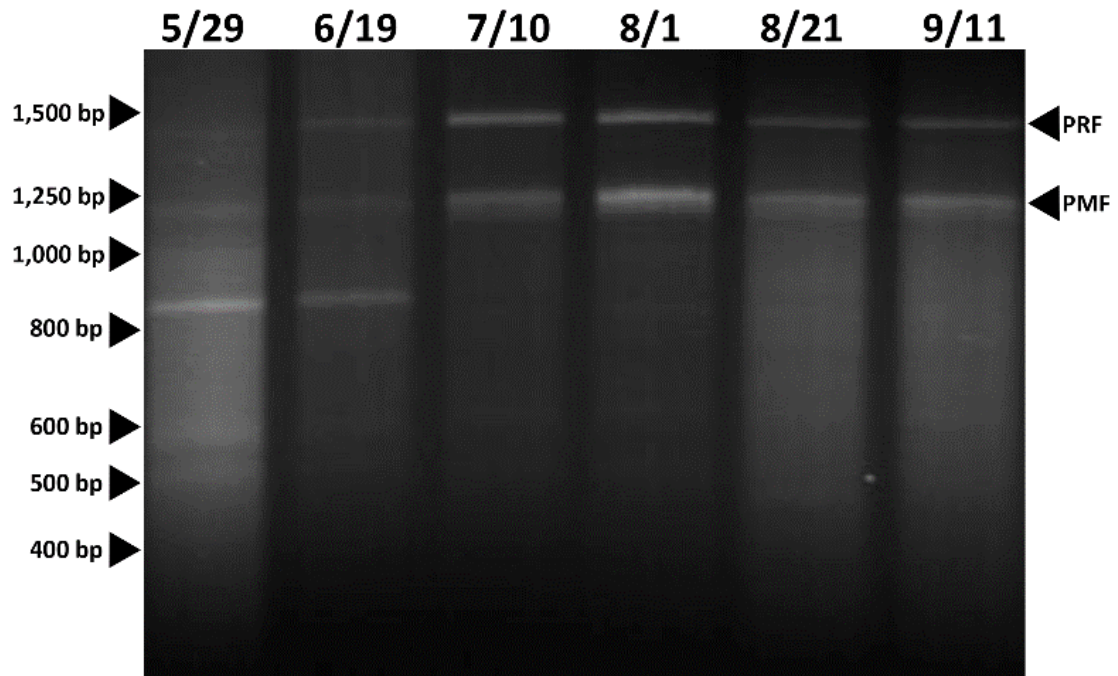


**Figure 27. Pheromone immunohistochemistry.**

Comparison of pheromone expression and localization for mental glands at two stages of development (6/13 and 8/3) by immunohistochemistry (using anti-PRF; red), with fluorescent dyes labeling the nucleus (blue) and ECM (green).

### Mental gland transcriptome and changes in gene expression

Total RNA was isolated from 5 mental glands at each of the six 2010 time points (30 glands total). It should be noted that “mental gland” is used here to describe the tissue occupying the anatomical location where the mental gland is normally found; for the earlier time points, this largely represented skin and connective tissue. Standardized amounts of total RNA were pooled for the five glands at each time point, and cDNA was synthesized with oligo-dT priming (Figure 28). Interestingly, the bands known to correspond to PMF and PRF mRNA were visible for all six time points, even in the early points when there was no detectable pheromone in the extract. The intense ~850 bp band in the 5/29 and 6/19 samples was later identified to encode a 15.3 kDa secreted protein with no significant blastp results in Genbank. DNA was sequenced using the Illumina HiSeq 2000 platform with 100bp paired-end reads at a depth of >20 million reads per time point, and a *de novo* transcriptome was prepared using the Trinity RNASeq package (Table 9). Initial gene annotation was performed using several publicly available bioinformatics tools (see methods section). Reads were re-aligned to the transcriptome, expression levels estimated using RSEM [248], and differences in gene expression between time points estimated by EBSeq [249] (Table 10). However, nearly all of the assembled genes were detectable at all six points (at varying levels), and with  $n = 1$  per time point, EBSeq used a highly conservative variance estimator and there was limited statistical power to detect differentially expressed genes between contiguous time points.



**Figure 28. *P. shermani* mental gland cDNA.**

Comparison of pooled, PCR amplified cDNA from six time points in mental gland development used for transcriptome sequencing. Arrows denote the bands corresponding to PRF and PMF mRNA, which are present at all 6 points.

**Table 9. Summary of *P. shermani* mental gland transcriptome.**

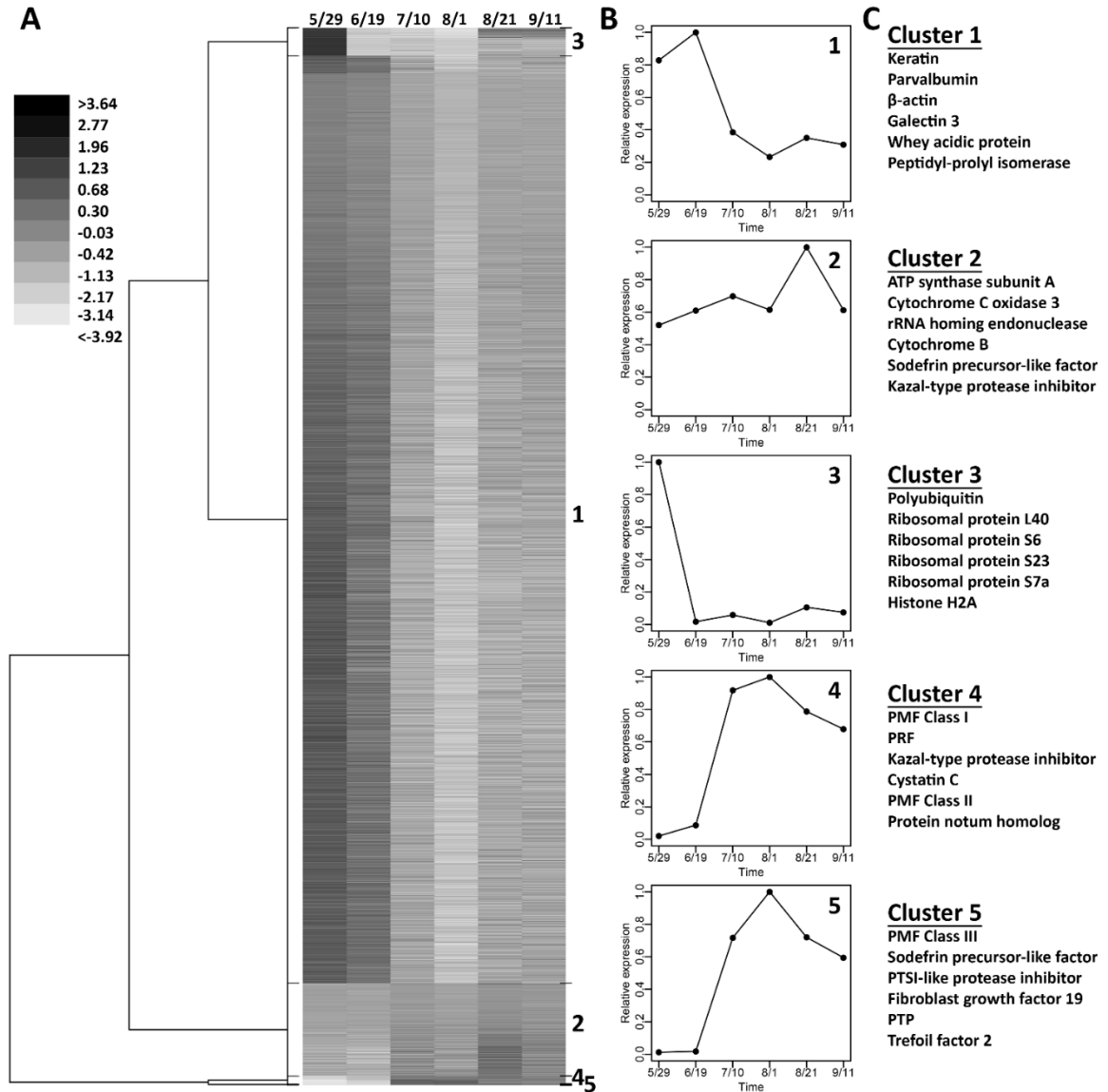
Total no. of reads	146,829,960
Average read length	100 bp
Total number of Trinity components	56,083
Total number of Trinity transcripts	158,361
Mean length of transcripts	578 bp
Median length of transcripts	394 bp
No. of Trinotate-predicted ORFs	47,097
Trinotate ORFs with SwissProt match (blastp)	22,182
Trinotate ORFs with TrEMBL match (blastp)	25,184
Transcripts with TrEMBL match (blastx)	39,600
Successfully re-aligned forward reads	66,392,752 (90.4%)

**Table 10. Select list of differentially expressed genes.** Comparison denotes the two time points compared by EBSeq, with fold change as the ratio of the second time point to the first, and the posterior probability of differential expression (PPDE).

Gene	Putative function	Comparison	Fold Change	PPDE
PMF Class II	Pheromone	6/19 → 8/1	28 X	1
C3	Putative pheromone	6/19 → 8/1	430 X	1
Sodefrin-like factor	Putative pheromone	6/19 → 8/1	350 X	0.962
VIP-like homolog	Putative pheromone	6/19 → 8/1	170 X	1
Calmodulin	Ca <sup>2+</sup> binding	6/19 → 8/1	0.24 X	1
Parvalbumin	Ca <sup>2+</sup> binding	5/29 → 6/19	9200 X	0.999
		6/19 → 8/1	0.00031 X	0.973
Calreticulin	Ca <sup>2+</sup> binding	6/19 → 8/1	3.1 X	1
Kazal-type serine protease inhibitor	Secreted protease inhibitor	6/19 → 8/1	280 X	0.976
Cystatin	Secreted protease inhibitor	6/19 → 8/1	30 X	1
Cathepsin S	Cysteine protease	6/19 → 8/1	1.5 X	0.999
Protein disulfide isomerase A6	Chaperone	6/19 → 8/1	9 X	1
ADP-ribosylation factor 4	Vesicle transport	6/19 → 8/1	1.7 X	0.999
Gap junction protein	Gap junctions	6/19 → 7/10	480 X	0.999
		8/1 → 8/21	0.0012 X	1
Vascular endothelial growth factor A	Promotes angiogenesis	6/19 → 8/1	39 X	0.957
Tubulin $\alpha$ 1	Microtubule formation	6/19 → 8/1	0.41 X	1
Histone H2A	Chromatin binding	6/19 → 8/1	0.36 X	0.999
Histone H1E	Chromatin binding	5/29 → 6/19	0.0074 X	0.935
eIF4AI	Translation initiation	5/29 → 6/19	31 X	0.909
		8/1 → 8/21	0.045 X	0.990
eIF 2	Translation initiation	5/29 → 6/19	0.011 X	0.901
Ribosomal protein L36	Translation	6/19 → 7/1	20 X	0.881
		8/1 → 8/21	55 X	0.997
Ribosomal protein L38	Translation	8/1 → 8/21	0.065 X	0.999
Ribosomal protein S4	Translation	8/1 → 8/21	0.049 X	0.984
Ribosomal protein S25	Translation	8/1 → 8/21	0.071 X	0.978

To better characterize overall expression patterns, cluster analysis was performed using the R package *MBCluster.Seq* [250]. Using a hierarchical approach, five major groups of genes were identified with different expression profiles and numbered 1 to 5 in order of decreasing gene count (Figure 29). Cluster 1 contained the majority of genes (87.7%), had a maximum expression in May/June, and included the majority of housekeeping genes (e.g.  $\beta$ -actin, GAPDH, PCNA, ferritin, ribosomal proteins). In relative terms, cluster 2 had the most stable expression patterns ( $\sim$ 1-4X fold difference between time points), and included a range of genes from different biological pathways, including ribosomal proteins, lysosomal proteases, signal peptidase complex members, and lipid biogenesis enzymes. Cluster 3 included genes almost exclusively found in the earliest time point (with some low expression in the last two time points, possibly suggesting a cyclical response as the gland begins to resorb). Some of the most highly expressed genes in cluster 3 included ribosomal proteins (S6, S15, S17, S23, L14, L24, L32, L37a) and histone proteins (H1E, H2A, H3). Clusters 4 and 5 together include the genes most highly expressed in the later phases of gland maturation (0.8% of all genes). As expected, the majority of transcripts coded for pheromone, including PRF and PMF, but also (in lower abundance) many putative pheromones that were identified in other plethodontid species (natriuretic peptide, vasoactive intestinal peptide, sodefrin precursor-like factor, cysteine rich secretory protein) [89]. Included in these sequences was a predicted protein related to the tissue inhibitor of metalloproteinase (TIMP) family that included an extraordinarily long 3' UTR ( $\sim$ 3700 nt); through mass spectrometry, this sequence was matched to a protein previously termed C3 based on its chromatographic elution conditions. C3 comprises  $\sim$ 10% of the pheromone extract [91].

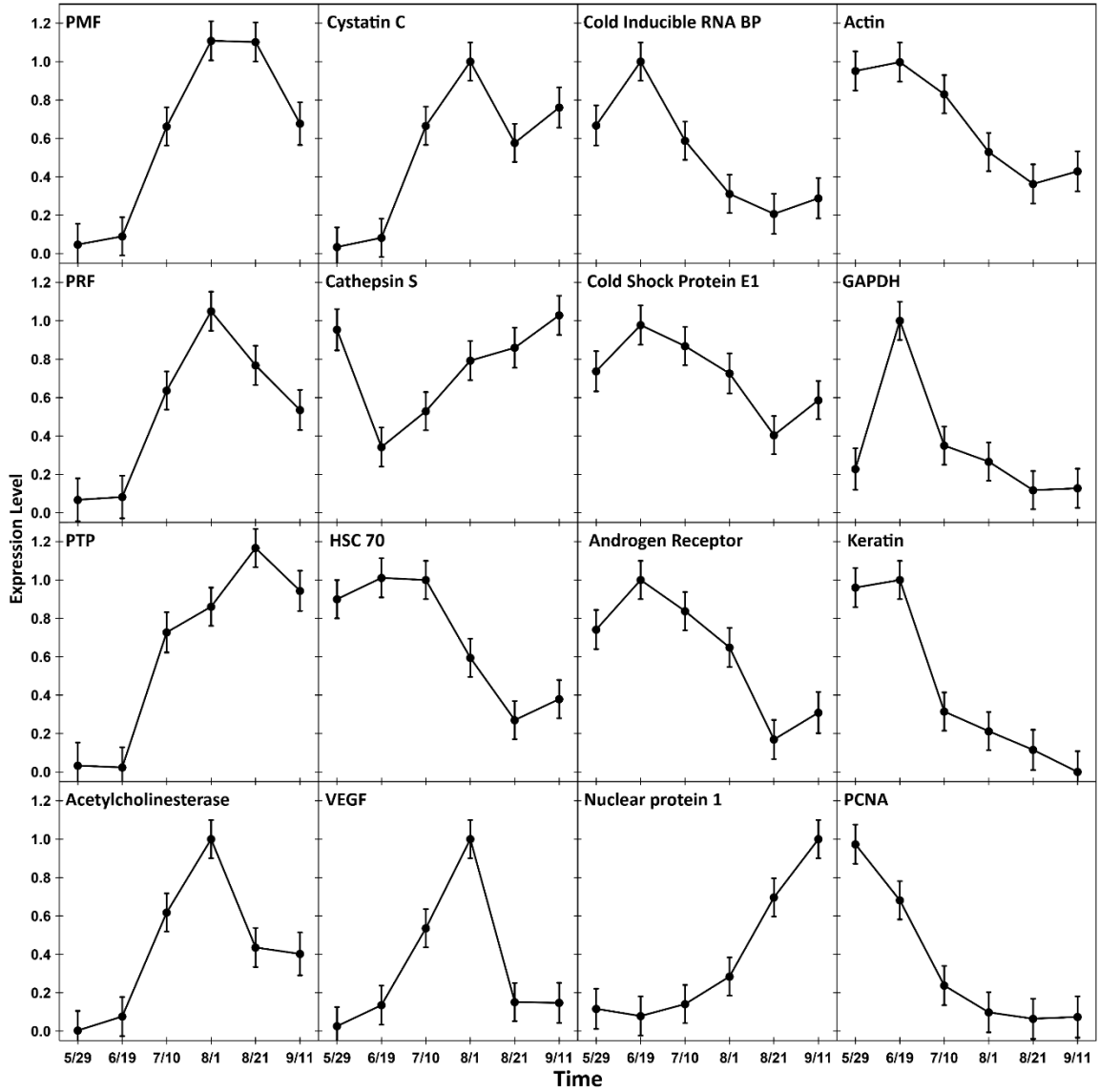




**Figure 29. Cluster analysis of mental gland gene expression.**

(A) Cladogram representing the ~55,000 genes organized into 6 clusters. Shades of grey represent log fold changes between time points. (B) Line graphs of cluster means vs time. (C) The six most abundant genes in each cluster (at the time point with the highest expression levels per cluster).

The function of this protein is still unknown, but given this new information as to its likely homology, we will now refer to it as Plethodontid TIMP-like Protein (PTP). Multiple other protease inhibitor-like proteins were identified, including cystatin C and multiple Kazal-type inhibitors. Clusters 4 also included a number of retrotransposon and reverse transcriptase-like sequences. The biological importance of these sequences is unclear, but provides a likely mechanism to explain the presence of processed PMF pseudogenes in the *P. shermani* genome [Chapter II]. Three other proteins of interest in cluster 4 included acetylcholinesterase (AChE), nuclear protein 1 (NP1), and vascular endothelial growth factor (VEGF). Incubation with acetylcholine is our standard methodology to induce pheromone secretion from mental glands [83, 251], such that the co-secretion of AChE would allow for tightly controlled pheromone release via the same mechanism that regulates muscle contractions [252]. VEGF, an angiogenic factor, may be necessary for mental gland maintenance such that as the mental gland enlarges, the cells closest to the dorsal surface will be distant from dermal capillaries and may not receive sufficient nutrition without angiogenesis. Related to the preceding, NP1 classically functions in chromatin remodeling as part of stress responses, such as nutrient starvation, to prevent apoptosis [253, 254]. Notably, NP1 is one of the few proteins to steadily increase in expression over the 6 time points, and may play an important role in ensuring that the gland persists throughout the courtship season after it has transitioned to pheromone synthesis. To validate estimates of gene expression from the Illumina data, qRT-PCR analysis was performed for 16 genes of interest (Figure 30), and similar expression patterns were observed.



**Figure 30. qRT-PCR analysis of select mental gland genes.**

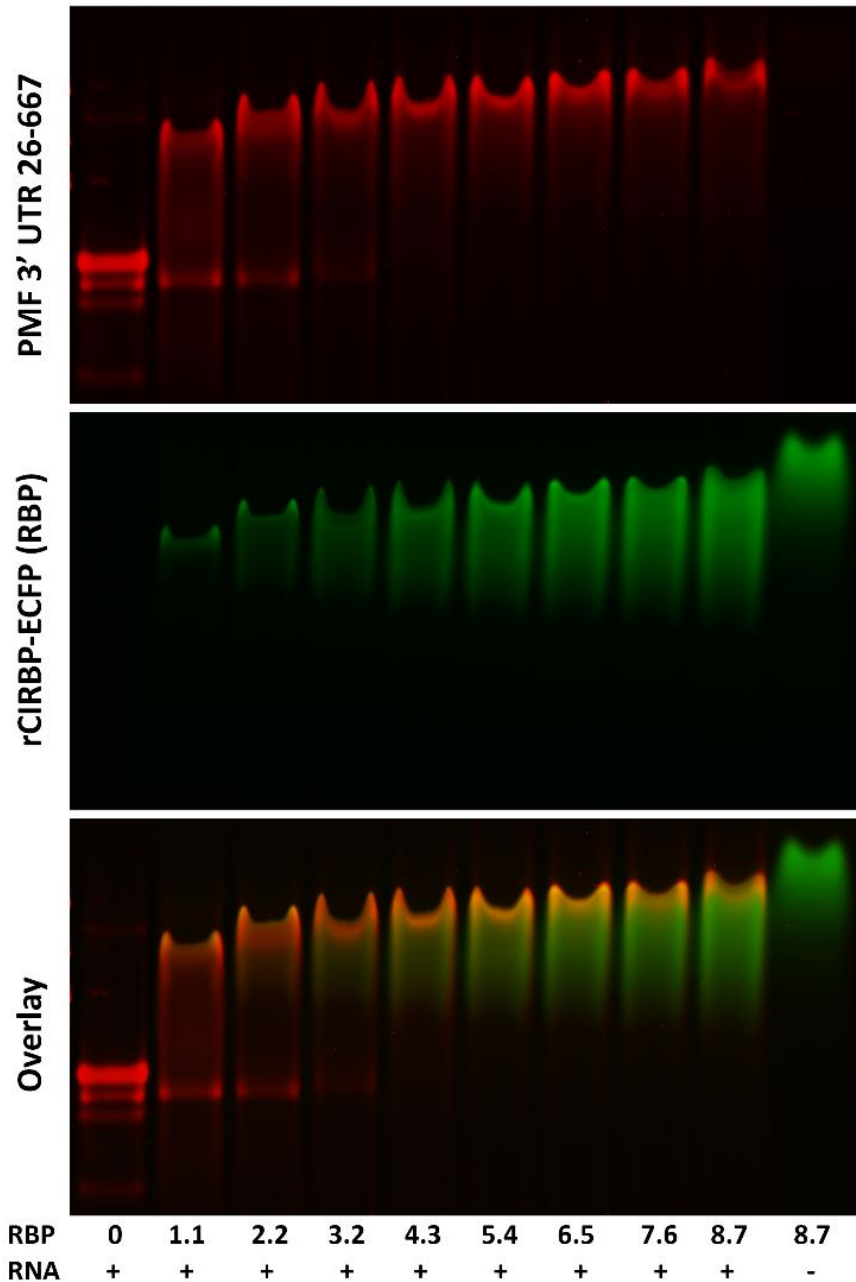
Transcript abundance was measured for 16 mental gland genes qRT-PCR, scaled relative the time point with the most abundant expression, and fit to a linear mixed effect model with gene and time as fixed effects and male as a random effect. Reported values are model estimated mean  $\pm$  standard error.

RNASeq- and qPCR-based estimates of gene expression were significantly correlated with  $R^2 = \sim 0.5$ , with no significant biases from gene or time point.

#### Cold Inducible RNA Binding Protein (CIRBP) binds the PMF 3' UTR

It was previously hypothesized that the highly conserved PMF untranslated regions functioned as platforms for RNA binding proteins (RNA-BPs) that would permit synchronized expression of the many diverse isoforms [Chapter II]. Unfortunately, no RNA-BPs were detected during differential expression analysis with EBSeq. Upon manual examination of the candidate genes, Cold Inducible RNA-BP (CIRBP) was found to be significantly more abundant than any other RNA-BP ( $\sim 0.13\%$  of all transcripts in 6/19 time point). Analysis by qRT-PCR confirmed differential mRNA expression over the six time points, with maximum expression at 6/19 (Figure 30). To test for biological activity *in vitro*, recombinant CIRBP fused to the enhanced cyan fluorescent protein (rCIRBP/ECFP) was expressed in *E. coli* and purified to  $>99\%$  homogeneity.

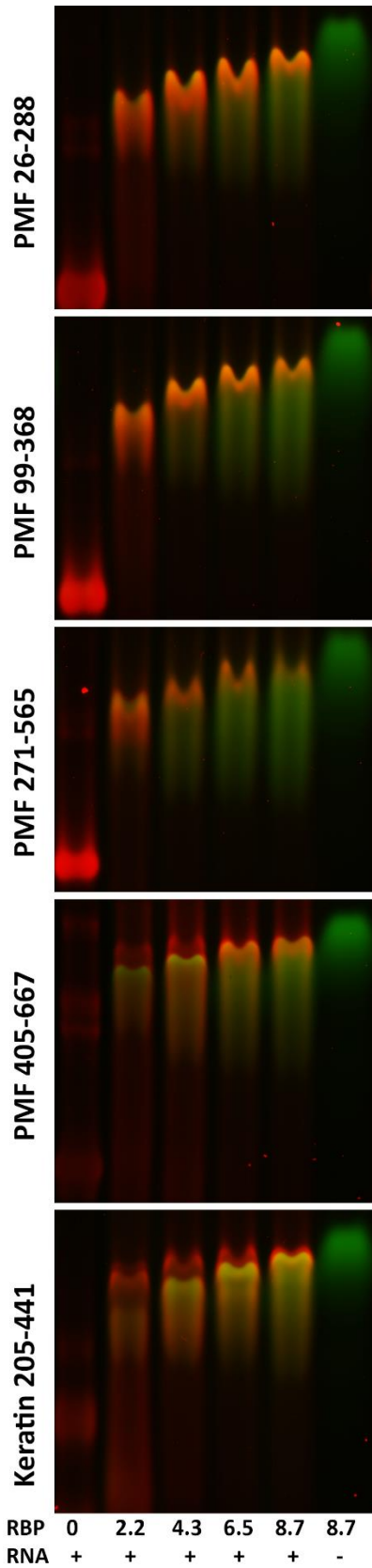
Electrophoretic mobility shift assays (EMSAs) were performed using rCIRBP/ECFP and *in vitro* transcribed RNA. For these initial studies, efforts were focused on PMF Class I that includes the majority of expressed isoforms and represents the greatest percentage of the PMF mRNA. When rCIRBP/ECFP was titrated against a nearly full length PMF 3' UTR (nucleotides 26-667), a very clear shift was observed in both the RNA and protein bands (Figure 31). Interestingly, there was visible RNA smearing at lower concentrations of rCIRBP/ECFP, suggesting possible dissociation of the RNA/protein complex during electrophoresis. Simultaneously, the altered position of the RNA band in the presence of greater rCIRBP/ECFP suggested a non 1:1 stoichiometry.



**Figure 31. CIRBP – PMF 3' UTR interactions.**

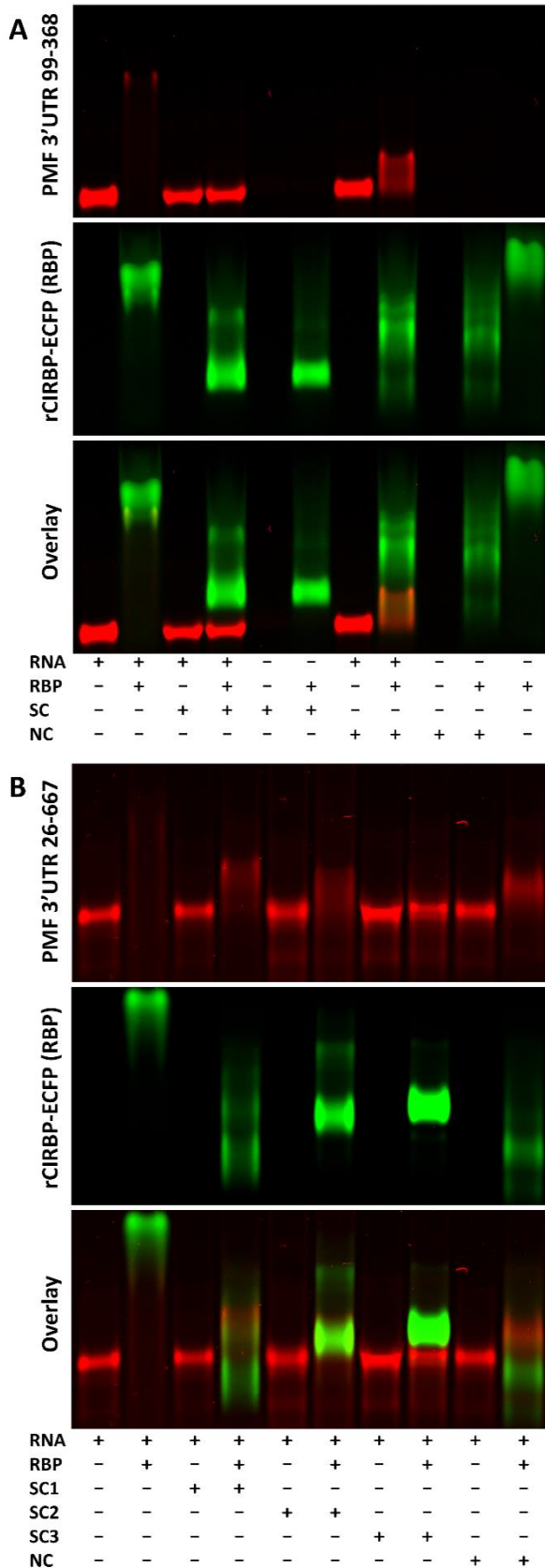
EMSA using a constant amount of PMF 3' UTR *in vitro* RNA (200 ng) with increasing concentrations of rCIRBP/ECFP ( $\mu\text{M}$ ; RBP), and a protein-only control. Protein fluorescence was detected by ECFP (green), and RNA was stained using Sybr Green II (red).

To narrow down the potential binding sequences for CIRBP in the PMF 3' UTR, four overlapping sequences of ~250 nt were prepared. Simultaneously, an off-target control of similar length was prepared from the keratin 3' UTR. When these five different RNAs were analyzed by EMSA, all showed visible gel shifts in the presence of increasing rCIRBP/ECFP, yet the degree of overlap appeared most intense in the 26-288 and 99-368 fragments (Figure 32). Also, in the non-protein control lanes, the variability in band number and intensity suggested different degrees of RNA secondary structure between the different sequences, which may have had an impact on CIRBP binding. To further demonstrate specificity of CIRBP towards the PMF 3' UTR, fluorescently tagged versions of the different RNA molecules were prepared by using tetramethylrhodamine (TAMRA) and aminoallyl-modified uracil. With PMF 3' UTR 99-368, addition of a 100-fold excess of unlabeled PMF 3' UTR 99-368 eliminated the gel shift, while a 100-fold excess of unlabeled keratin 3' UTR only reduced the gel shift to a smear (Figure 33A). These data further suggest that CIRBP has relatively greater affinity for the PMF 3' UTR, yet some non-specific affinity for other RNA molecules. In a similar competition assay, using a TAMRA-labelled PMF 3' UTR 26-667, 100X unlabeled RNA was added for three different lengths of the PMF 3' UTR (26-288, 26-565, 26-667) and keratin 3' UTR. Only the full length PMF 3' UTR 26-667 was able to fully eliminate the observed gel shift, such that there was a positive correlation between RNA length and ability to compete for binding (Figure 33B).



**Figure 32. CIRBP affinity for different RNAs.**

Overlays of fluorescent EMSAs using increasing concentrations rCIRBP/ECFP (RBP) with 200 ng of five different RNA molecules (four overlapping ~250 bp segments of the PMF 3' UTR, and a Keratin 3' UTR control). Protein fluorescence was detected by ECFP (green), and RNA was stained using Sybr Green II (red).



**Figure 33. Competition EMSA with CIRBP.**

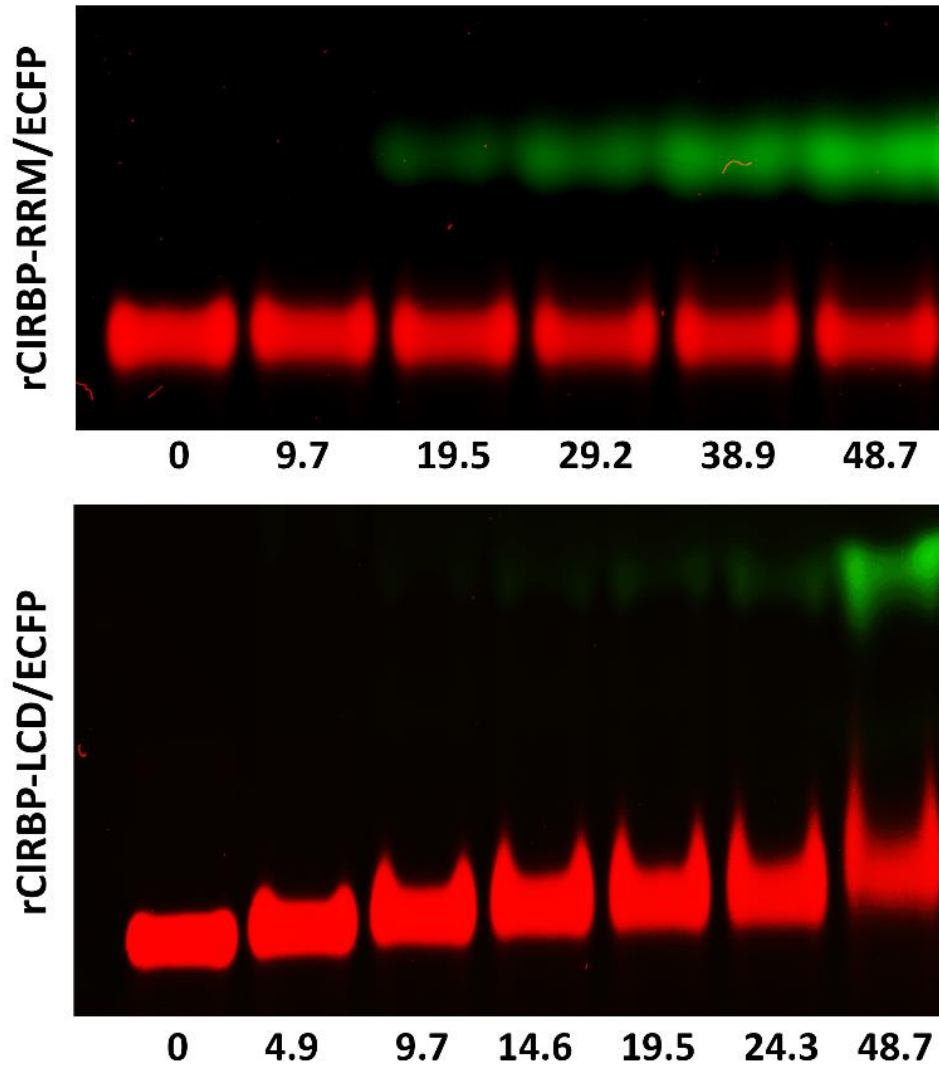
(A) EMSA between TAMRA-labeled PMF 3' UTR 99-368 RNA (30 ng; red) and rCIRBP/ECFP (1.5  $\mu$ g; RBP; green), with competition using 100X (3  $\mu$ g) of a specific competitor (SC; unlabeled PMF 3' UTR 99-368) or a non-specific competitor (NC; unlabeled Keratin 3' UTR 205-441). (B) EMSA with TAMRA-labeled PMF 3' UTR 26-668 RNA (30 ng; red) and rCIRBP/ECFP (1.5  $\mu$ g; RBP; green), with competition using 100X (3  $\mu$ g) of four different unlabeled competitors: SC1 = PMF 3' UTR 26-288, SC2 = PMF 3' UTR 26-565, SC3 = PMF 3' 26-668, or NC = Keratin 3' UTR 205-441.



### Dynamics of CIRBP-PMF 3' UTR Interactions

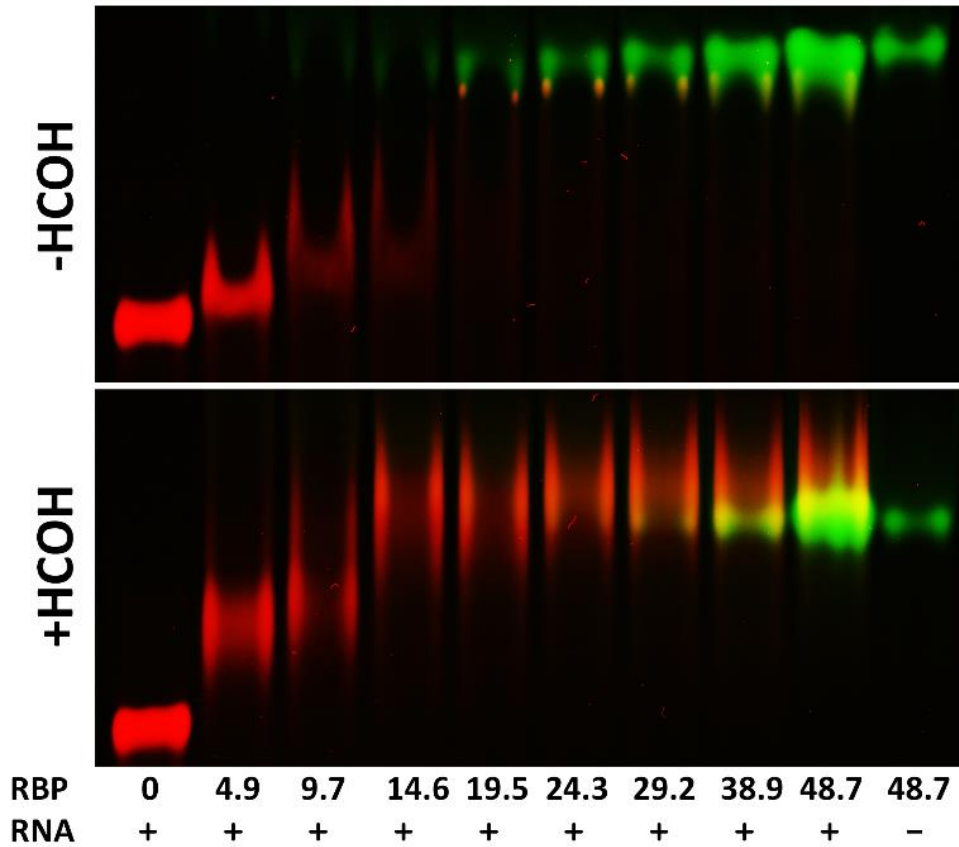
CIRBP contains two structural domains: an N-terminal RNA recognition motif (RRM) and a C-terminal glycine-rich, low complexity domain (LCD). Studies on the human homolog of CIRBP suggested that both domains can bind RNA, with the RRM having specific RNA interactions and the LCD having non-specific interactions [255]. For *P. shermani*, each domain was expressed as a separate fusion protein to enhanced cyan fluorescent protein (rCIRBP-RRM/ECFP and rCIRBP-LCD/ECFP), and when used in EMSAs, neither domain demonstrated strong affinity to the TAMRA-labelled PMF 3' UTR (99-368) (Figure 34). There was a small amount of smearing that occurred in the highest concentrations of rCIRBP-LCD/ECFP, suggesting some weak interaction. As similar smearing was observed at lower concentrations with rCIRBP/ECFP, EMSAs with rCIRBP/ECFP and PMF 3' UTR 99-368 were repeated with and without formaldehyde pre-treatment to crosslink protein-RNA complexes. Crosslinking successfully reduced the amount of visible RNA smearing in the gel (Figure 35), suggesting that under sufficiently low stoichiometry, rCIRBP/ECFP (and likely rCIRBP-LCD/ECFP) forms an unstable complex with target RNA such that it readily dissociates under electrophoresis conditions.

When using TAMRA-labelled RNA (Figure 33), it was observed that interaction with CIRBP caused significant fluorescence quenching of bound RNA – likely due to shielding of the fluorophores attached to the uracil bases. Using TAMRA-labelled PMF 3' UTR 99-368, all three CIRBP constructs were titrated and fluorescence quenching quantified by measurement in a 96-well microplate (Figure 36).



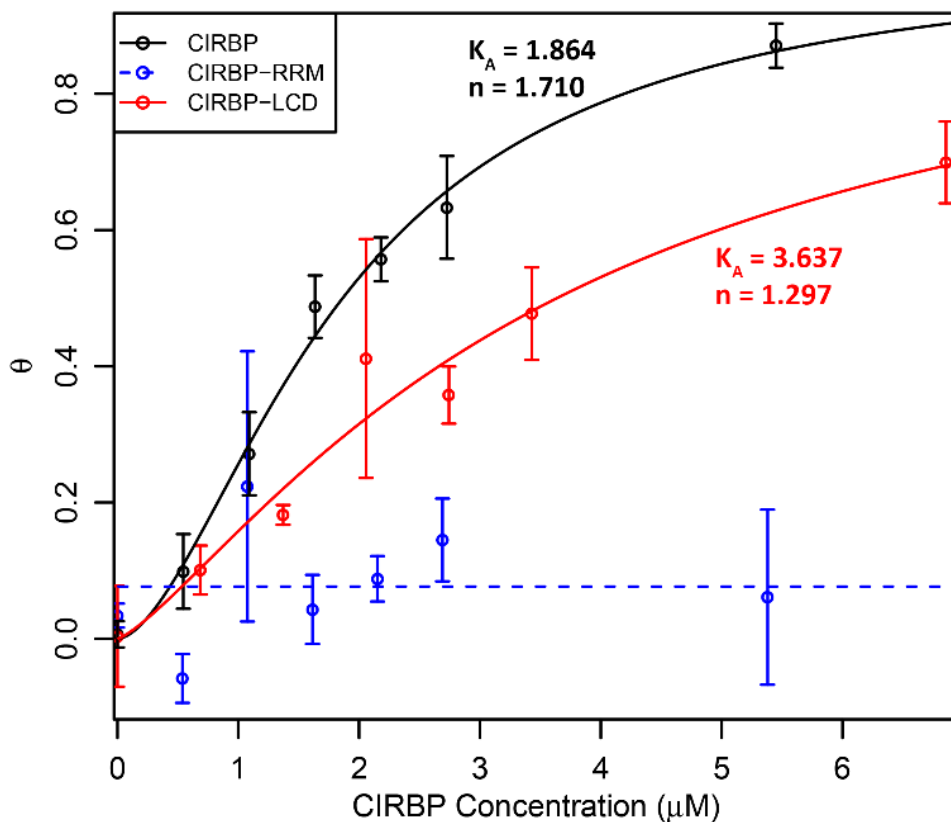
**Figure 34. EMSA with CIRBP domains.**

TAMRA-labeled PMF 3' UTR 99-368 RNA (30 ng; red) with increasing concentrations of either rCIRBP-RRM/ECFP or rCIRBP-LCD/ECFP ( $\mu\text{M}$ ; green).



**Figure 35. CIRBP – RNA interactions stabilized by formaldehyde crosslinking.**

EMSA between TAMRA-labeled PMF 3' UTR 99-368 RNA (30 ng; red) and increasing concentrations of rCIRBP/ECFP ( $\mu\text{M}$ ; RBP; green), with and without pre-treatment with 1% formaldehyde.

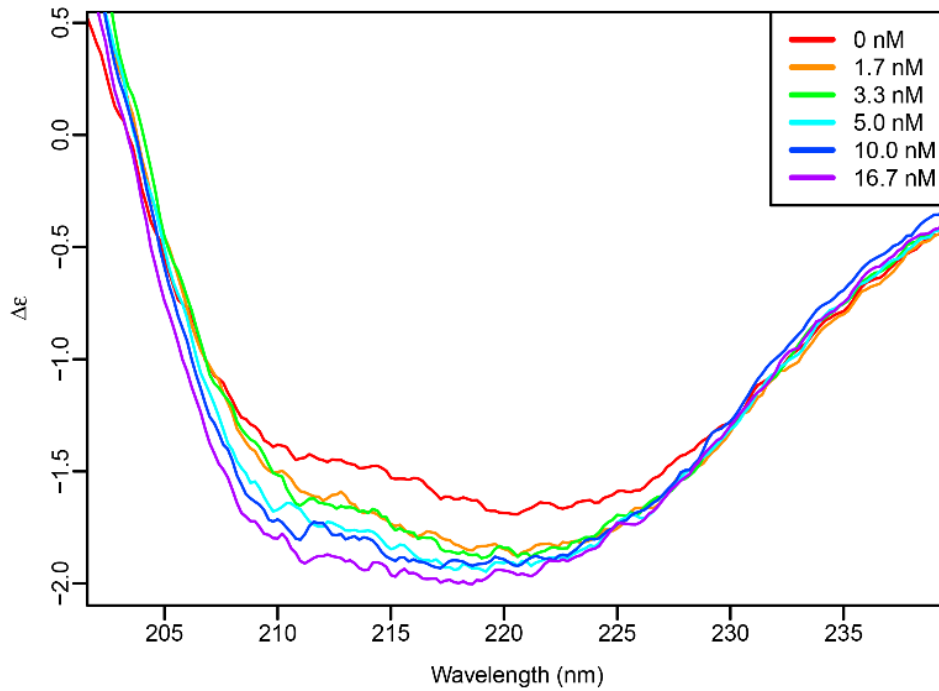


**Figure 36. CIRBP titration curve.**

TAMRA-labeled PMF 3' UTR 99-368 RNA was titrated with increasing concentrations of rCIRBP/ECFP (black), rCIRBP-RRM/ECFP (blue), and rCIRBP-LCD/ECFP (red), and binding measured by fluorescence quenching. Data were fit to the Hill equation by nonlinear modeling to obtain measures of binding affinity ( $K_A$ ) and cooperativity ( $n$ ). No significant change was detected for rCIRBP-RRM/ECFP, denoted by a dashed line.

There was no detectable fluorescence quenching with rCIRBP-RRM/ECFP (represented by the dashed blue line), while both rCIRBP/ECFP and rCIRBP-LCD/ECFP yielded sigmoidal curves characteristic of cooperative binding. When data for these two proteins were fit to the Hill equation by nonlinear regression, both the association constant ( $K_A$ ) and Hill coefficient ( $n$ ) were significant, but with rCIRBP-LCD/ECFP having both lower affinity (higher  $K_A$ ) and weaker cooperativity (lower  $n$ ) compared to rCIRBP/ECFP. Thus, the data suggest that both domains likely function synergistically to promote binding to the PMF 3' UTRs.

Recent models of other RNA-BPs with LCDs suggested that, upon binding to a proper catalyst, unstructured LCDs adopt regular  $\beta$ -sheet structure which permits aggregation and formation of stress granules, processing bodies, or other macromolecular RNA-protein complexes [256]. To test if the PMF 3' UTR may be acting as such a catalyst, rCIRBP/ECFP was analyzed by circular dichroism (CD) and titrated with increasing amounts of PMF 3' UTR 26-667 (Figure 37). There was a detectable increase in the CD absorbance, particularly near  $\sim 215$  nm where  $\beta$ -sheet can be measured. Even though the percentage change is relatively small, this likely relates to (1) the majority of CD signal originating from ECFP, which comprises  $\sim 60\%$  of the total protein and contains a highly structured  $\beta$ -barrel, and (2) CD only reports on the average secondary structure content, and the “induced”  $\beta$ -sheet in CIRBP may only be occurring in a small proportion of the available molecules. Nonetheless, these data support that binding of CIRBP to the PMF 3' UTR promotes a conformational change and increased secondary structure, likely in the LCD shifting from random coil to  $\beta$ -sheet.

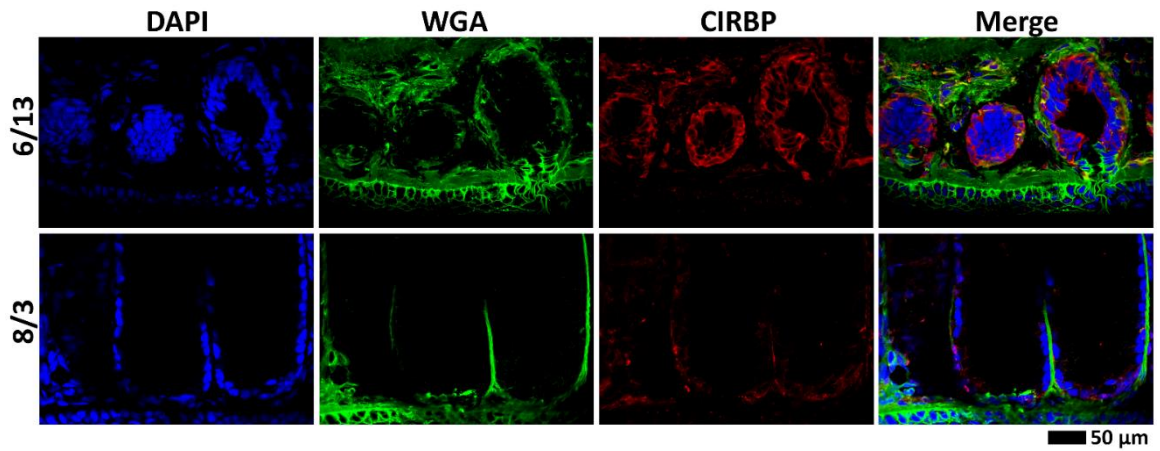


**Figure 37. PMF 3' UTR induces secondary structure changes in CIRBP.**

CD spectra of rCIRBP/ECFP with increasing concentrations of PMF 3' UTR 26-668 RNA, with changes in CD suggesting higher levels of secondary structure (likely  $\beta$ -sheet).

### CIRBP Expression and Correlation to PMF *In Vivo*

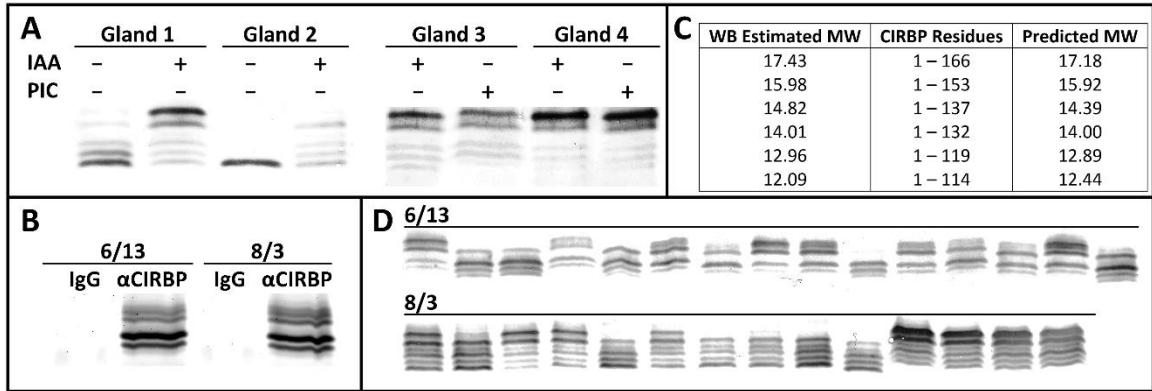
Polyclonal antibodies to rCIRBP/ECFP were prepared and affinity purified specifically against CIRBP-RRM. Immunohistochemical staining using anti-CIRBP-RRM revealed that CIRBP protein was in greater abundance at 6/13 compared to 8/3, localized to the cytoplasm. The staining was uniformly distributed, uncharacteristic of cytologically visible RNA granules such as stress granules, Cajal bodies, or nuclear speckles (Figure 38). To better quantify CIRBP protein levels for individual mental glands, western blot analysis was performed. However, multiple bands were observed near and below the approximate 17 kDa expected molecular weight (Figure 39A-D). These bands were hypothesized to be CIRBP degradation products, possibly by a protease that was sequentially processing the C-terminal LCD. In support of this hypothesis, immunopulldown products were analyzed by gel electrophoresis and mass spectrometry such that peptides were identified fully spanning the N-terminal RRM (Figure 39B). Addition of either a broad protease inhibitor cocktail or 0.1 mM iodoacetamide (to specifically inhibit cysteine proteases) limited the extent of degradation, but did not completely ablate it (Figure 39A). The same samples were examined by western blot over multiple days, and even without a protease inhibitor, degradation was incomplete (data not shown). Thus, some of this visible degradation may be naturally occurring, involved in natural CIRBP turnover, and play a biological role. Interestingly, neither inhibitor changed the number of bands observed, only their relative intensities, such that a single protease may be contributing to both natural and experimentally-induced degradation.



**Figure 38. CIRBP immunohistochemistry.**

Comparison of CIRBP expression and localization for mental glands at two stages of development (6/13 and 8/3) by immunohistochemistry (using anti-CIRBP-RRM; red), with fluorescent dyes labeling the nucleus (blue) and ECM (green).





**Figure 39. CIRBP protein analysis.**

(A) Western blot analysis comparing CIRBP degradation with and without different protease inhibitors. Mental glands were dissected into approximate halves, and incubated with RIPA extract containing no protease inhibitors, a commercially available protease inhibitor cocktail (PIC), or 100  $\mu$ M iodoacetamide (IAA). (B) SDS-PAGE with SYPRO Ruby stain of immunopulldown products using either a rabbit IgG mixture vs anti-CIRBP-RRM for two time points in gland development. (C) Estimated molecular weights (in kDa) for different CIRBP degradation products compared with masses predicted by cleavage of C-terminal Cathepsin S sites (YG). (D) Western blot analysis demonstrating the diversity of CIRBP abundance and degradation state for individual mental glands from 6/13 (5  $\mu$ g total protein) vs 8/3 (30  $\mu$ g total protein) when extracted using RIPA with IAA.

While we were unable to determine the specific protease involved, multiple lines of evidence suggested that it may be cathepsin S: (1) iodoacetamide performed as well or better than the protease inhibitor mix, implicating a cysteine protease; (2) most of the identified mental gland proteases in the transcriptome were lysosomal, with cathepsin S having the highest expression; (3) in contrast to most lysosomal proteases, cathepsin S is active only at near-neutral pH, and the RIPA buffer was at pH 8; (4) the CIRBP LCD is enriched for the cathepsin S target sequences (aliphatic or aromatic residues followed by Gly, [257], specifically YG in the CIRBP-LCD), and cleavage at these sites would generate proteins of similar molecular weight to those observed by western blot (Figure 39C).

To ascertain a potential role for CIRBP in regulating mental gland development and PMF synthesis, several variables were correlated from individual mental glands collected at two time points (6/13 and 8/3). Pheromone was extracted by incubation in acetylcholine for 30 minutes, and concentration measured by BCA protein assay. Because PMF consistently comprises ~50% of the total pheromone [91], we used this concentration as a proxy for PMF protein expression ( $P_{PMF}$ ). Following pheromone extraction, pheromone glands were stored in RNAlater, and later dissected into two approximately equal halves in order to independently isolate total RNA (using a Qiagen kit) and cellular protein (homogenization in RIPA buffer with iodoacetamide). Using standardized amounts of cellular protein, CIRBP expression was measured by western blot (Fig 15D). Based on densitometric estimates, both total CIRBP ( $P_{Total-CIRBP}$ ) and intact CIRBP ( $P_{Intact-CIRBP}$ ) were measured. Using total RNA, expression levels were measured for three genes by qRT-PCR: PMF ( $R_{PMF}$ ), CIRBP ( $R_{CIRBP}$ ), and cathepsin S

( $R_{Cath}$ ). Using these 6 variables plus a time covariate (6/13 vs 8/3), a series of nested MANOVAs were performed to identify potentially meaningful correlations (Table 11). Time was a significant covariate for all variables, and the only significant factor for all three measured mRNA levels. Total CIRBP protein (including degradation products) increased proportionally to CIRBP mRNA levels, but with a higher ratio at 6/13 compared to 8/3. Total CIRBP was the best predictor for intact CIRBP levels, however, there were significant interaction terms between  $time/R_{Cath}$  and  $time/R_{Cath}/R_{CIRBP}/P_{Total-CIRBP}$  such that, only at 6/13, intact CIRBP levels decreased as Cathepsin S mRNA levels increase (both independently and in proportion to total CIRBP levels). Finally, while neither CIRBP protein variable had an effect on PMF mRNA levels, pheromone protein levels were negatively correlated with intact CIRBP (represented by the  $time/P_{Intact-CIRBP}$  interaction term, because there was no pheromone protein detected at 6/13). While some of these correlations may be a consequence of natural gland progression in response to some other unmeasured variables, these data provide additional evidence to support that Cathepsin S regulates steady-state levels of CIRBP. As intact CIRBP was negatively correlated with PMF protein levels but not mRNA levels, this suggested that CIRBP may be acting a translational repressor.

**Table 11. MANOVA results for CIRBP expression and regulation.** Statistically significant coefficients ( $p < 0.05$ ) are bolded.

	Variable	Sum of Squares	F-statistic	p-value	Coefficient
<i>R<sub>CIRBP</sub> ~ time</i>					
	<b>time</b>	<b>5.42 x 10<sup>-16</sup></b>	<b>23.18</b>	<b>5.0 x 10<sup>-5</sup></b>	<b>8.65 x 10<sup>-9</sup></b>
	Residuals	6.31 x 10 <sup>-16</sup>			
<i>R<sub>Cath</sub> ~ time</i>					
	<b>time</b>	<b>2.55 x 10<sup>-16</sup></b>	<b>57.83</b>	<b>3.52 x 10<sup>-8</sup></b>	<b>5.93 x 10<sup>-9</sup></b>
	Residuals	1.19 x 10 <sup>-16</sup>			
<i>P<sub>Total-CIRBP</sub> ~ time x R<sub>CIRBP</sub></i>					
	<b>time</b>	<b>9.59 x 10<sup>8</sup></b>	<b>90.46</b>	<b>8.73 x 10<sup>-10</sup></b>	<b>-6.34 x 10<sup>3</sup></b>
	<b>R<sub>CIRBP</sub></b>	<b>1.65 x 10<sup>8</sup></b>	<b>15.60</b>	<b>5.63 x 10<sup>-4</sup></b>	<b>1.45 x 10<sup>12</sup></b>
	<b>time : R<sub>CIRBP</sub></b>	<b>2.01 x 10<sup>8</sup></b>	<b>18.97</b>	<b>1.98 x 10<sup>-4</sup></b>	<b>-1.28 x 10<sup>12</sup></b>
	Residuals	2.65 x 10 <sup>8</sup>			
<i>P<sub>Intact-CIRBP</sub> ~ time x R<sub>CIRBP</sub> x P<sub>Total-CIRBP</sub> x R<sub>Cath</sub></i>					
	<b>time</b>	<b>2.68 x 10<sup>7</sup></b>	<b>6.02</b>	<b>0.0290</b>	<b>-2.66 x 10<sup>3</sup></b>
	R <sub>CIRBP</sub>	1.04 x 10 <sup>7</sup>	2.34	0.150	4.27 x 10 <sup>12</sup>
	<b>P<sub>Total-CIRBP</sub></b>	<b>4.91 x 10<sup>7</sup></b>	<b>11.02</b>	<b>0.00553</b>	<b>-0.382</b>
	R <sub>Cath</sub>	1.80 x 10 <sup>6</sup>	0.41	0.536	-4.75 x 10 <sup>14</sup>
	time : R <sub>CIRBP</sub>	2.00 x 10 <sup>6</sup>	0.45	0.514	-4.81 x 10 <sup>12</sup>
	time : P <sub>Total-CIRBP</sub>	1.12 x 10 <sup>6</sup>	0.25	0.624	-1.60 x 10 <sup>-2</sup>
	<b>time : R<sub>Cath</sub></b>	<b>2.95 x 10<sup>7</sup></b>	<b>6.63</b>	<b>0.023</b>	<b>4.74 x 10<sup>14</sup></b>
	R <sub>CIRBP</sub> : P <sub>Total-CIRBP</sub>	6.06 x 10 <sup>5</sup>	0.14	0.718	-1.91 x 10 <sup>8</sup>
	R <sub>CIRBP</sub> : R <sub>Cath</sub>	1.43 x 10 <sup>7</sup>	0.32	0.581	3.47 x 10 <sup>22</sup>
	P <sub>Total-CIRBP</sub> : R <sub>Cath</sub>	2.48 x 10 <sup>7</sup>	0.56	0.469	2.12 x 10 <sup>10</sup>
	time : R <sub>CIRBP</sub> : P <sub>Total-CIRBP</sub>	1.29 x 10 <sup>7</sup>	0.29	0.600	2.25 x 10 <sup>8</sup>
	time : R <sub>CIRBP</sub> : R <sub>Cath</sub>	1.96 x 10 <sup>6</sup>	0.04	0.837	-3.46 x 10 <sup>22</sup>
	time : P <sub>Total-CIRBP</sub> : R <sub>Cath</sub>	2.95 x 10 <sup>7</sup>	0.66	0.430	-2.10 x 10 <sup>10</sup>
	R <sub>CIRBP</sub> : P <sub>Total-CIRBP</sub> : R <sub>Cath</sub>	7.92 x 10 <sup>6</sup>	0.18	0.680	-1.43 x 10 <sup>18</sup>
	<b>time : R<sub>CIRBP</sub> : P<sub>Total-CIRBP</sub> : R<sub>Cath</sub></b>	<b>2.86 x 10<sup>8</sup></b>	<b>6.42</b>	<b>0.025</b>	<b>1.42 x 10<sup>18</sup></b>
	Residuals	5.79 x 10 <sup>8</sup>			
<i>R<sub>PMF</sub> ~ time x P<sub>Total-CIRBP</sub> x P<sub>Intact-CIRBP</sub></i>					
	<b>time</b>	<b>1.50 x 10<sup>-10</sup></b>	<b>43.98</b>	<b>1.45 x 10<sup>-6</sup></b>	<b>4.97 x 10<sup>-6</sup></b>
	P <sub>Total-CIRBP</sub>	2.40 x 10 <sup>-13</sup>	0.07	0.792	6.00 x 10 <sup>-12</sup>
	P <sub>Intact-CIRBP</sub>	1.00 x 10 <sup>-13</sup>	0.004	0.950	8.49 x 10 <sup>-12</sup>
	time : P <sub>Total-CIRBP</sub>	2.46 x 10 <sup>-12</sup>	0.72	0.406	1.05 x 10 <sup>-9</sup>
	time : P <sub>Intact-CIRBP</sub>	6.00 x 10 <sup>-14</sup>	0.02	0.899	-1.03 x 10 <sup>-10</sup>
	P <sub>Total-CIRBP</sub> : P <sub>Intact-CIRBP</sub>	3.00 x 10 <sup>-14</sup>	0.01	0.922	-5.80 x 10 <sup>-16</sup>
	time : P <sub>Total-CIRBP</sub> : P <sub>Intact-CIRBP</sub>	7.50 x 10 <sup>-13</sup>	0.22	0.644	8.22 x 10 <sup>-14</sup>
	Residuals	7.18 x 10 <sup>-11</sup>			

$P_{PMF} \sim time \times R_{PMF} \times P_{Intact-CIRBP} \times P_{Total-CIRBP}$					
	<b>time</b>	<b>2.8585</b>	<b>128.28</b>	<b><math>4.17 \times 10^{-8}</math></b>	<b>1.51</b>
	$R_{PMF}$	0.0451	2.02	0.178	$-2.67 \times 10^{-7}$
	$P_{Intact-CIRBP}$	0.0373	1.67	0.218	$-2.73 \times 10^{-18}$
	$P_{Total-CIRBP}$	0.0035	0.16	0.698	$-3.23 \times 10^{-19}$
	$time : R_{PMF}$	0.0013	0.06	0.810	$-1.39 \times 10^5$
	<b>time: <math>P_{Intact-CIRBP}</math></b>	<b>0.3478</b>	<b>15.61</b>	<b>0.00166</b>	<b><math>-4.49 \times 10^{-4}</math></b>
	$time : P_{Total-CIRBP}$	0.0281	1.26	0.282	$-1.48 \times 10^{-4}$
	$R_{PMF} : P_{Intact-CIRBP}$	0.0001	0.004	0.951	$1.47 \times 10^{-10}$
	$R_{PMF} : P_{Total-CIRBP}$	0.0234	1.05	0.324	$1.05 \times 10^{-11}$
	$P_{Intact-CIRBP} : P_{Total-CIRBP}$	0.0012	0.06	0.818	$1.06 \times 10^{-22}$
	$time : R_{PMF} : P_{Intact-CIRBP}$	0.0003	0.01	0.916	18.1
	$time : R_{PMF} : P_{Total-CIRBP}$	0.0000	0.001	0.980	32.8
	$time : P_{Intact-CIRBP} : P_{Total-CIRBP}$	0.0295	1.33	0.270	$5.63 \times 10^{-8}$
	$R_{PMF} : P_{Intact-CIRBP} : P_{Total-CIRBP}$	0.0060	0.27	0.614	$-5.76 \times 10^{-15}$
	$time : R_{PMF} : P_{Intact-CIRBP} : P_{Total-CIRBP}$	0.0002	0.01	0.930	$-8.35 \times 10^{-3}$
	<b>Residuals</b>	<b>0.2897</b>			

## Discussion

The process of mental gland development is a novel example of hormone-responsive tissue differentiation and restructuring to create an organ highly specialized towards performing a single function. For ~8 months of the year, the mentum of male *P. shermani* appears as normal skin and is indistinguishable from that of a female. This unusual phenomenon of seasonal organogenesis may be rare, but the underlying biochemical processes are likely conserved among a number of vertebrate developmental processes, supported by the large number of conserved genes observed in the mental gland transcriptome. However, the quality that makes this a particularly exquisite model system is the annual nature of accelerated growth followed by natural resorption, with the first half of the process sharing a number of qualities with tumorigenesis and resorption representing cyclical whole organ apoptosis. The presented molecular and histological data support our original hypothesis that mental glands must initiate in a highly mitogenic state, and then transition into a veritable pheromone factory. Cells first rapidly divide and proliferate, evidenced by elevated PCNA levels at the earliest time points in gland development (Figure 30). As these cells divide, many housekeeping genes (representing over 80% of the expressed genes) are activated in order to build the structure and basic morphology of the gland (Figure 29). The overall glandular structure seems to consist of many dozens to hundreds of cells surrounding an open lumen/channel that passes through the epidermis (Figure 26). At early stages of development, these cells consist volumetrically of little more than a nucleus, characteristic of rapid mitosis and further mimicking a tumor-like structure [258-260]. While the term “tumor-like” may not be perfectly accurate, there is a reasonable amount of analogy: the mental gland seemingly

“invades” the dermis, compresses and/or degrades much of the ECM to allow expansion, rapidly divides to form a larger structure, and eventually releases angiogenic factors to increase blood supply and possibly extend viability. While ECM composition was only preliminarily examined through lectin staining (Figures 26, 27, 38), it is noteworthy that several different protease inhibitor families (TIMPs, cystatins, Kazal-type serine protease inhibitors) are overexpressed at later stages of development. As many of these proteins have signal peptides, they may be secreted in order to inhibit proteases that are initially required to allow mental gland expansion, but eventually must be inactivated to prevent cell degradation. It was already known that Plethodontid TIMP-like Protein is packaged as part of the pheromone mixture when extracted with acetylcholine [91], and presumably as well during slapping by male salamanders. It is unclear if the other factors are also released in the same manner, or separately partitioned in the Golgi apparatus and secreted as part of an alternative constitutive pathway. Alternatively, these protease inhibitors may be necessary to protect synthesized pheromones from premature degradation. What makes this tumor-like phenotype particularly interesting is that, unlike normal cancerous tissue, following the courtship season the mental gland regresses and completely resorbs to a non-visible, precursor state. Detailed molecular characterization of both the proliferation and resorption processes could have powerful implications towards cancer biology and identifying potential factors with which to reprogram tumor cells.

For the gland to transition towards pheromone synthesis, extensive changes in gene expression are required, including overexpression of both PRF and PMF mRNA. However, while the proportions might have changed, nearly all genes were detectable at

all 6 time points. On average, housekeeping genes in Cluster 1 were still represented at ~33% of maximum levels even after transitioning to primarily pheromone synthesis. Despite minimum pheromone translation at early time points (Figure 27 and Table 11), PRF and PMF together comprised ~1.5% and ~10% of all mRNA in May and June, respectively. One explanation for these results may be that, upon androgen-stimulated initiation of hypertrophication, a single signal universally activates transcription of all mental gland-associated genes. Subsequently, additional factors such as chromatin remodeling and mRNA stability could regulate steady state levels of different genes within the transcriptome to control the timing and expression of key proteins. Models with this type of gene regulation already exist within the scope of reproductive biology in the cases of gametogenesis and the post-fertilization maternal-to-zygotic transition. Chromatin condensation occurs throughout both spermatogenesis and oogenesis leading to partial transcriptional silencing, and complete silencing in mature spermatids and oocytes [261-264]. Consequently, mRNAs are commonly regulated by cytoplasmic polyadenylation to control poly(A) tail length, recruitment of the poly(A) binding protein (PABP), and formation of the translation initiation complex [265-267]. However, additional RNA-BPs provide further regulation. One example is the Deleted-in-Azoospermia protein (DAZ) protein and its autosomal homolog, DAZ-like protein (DAZL). Both DAZ and DAZL recognize target sequences within the 3' UTR of select mRNAs and can recruit PABP through protein-protein interactions, permitting formation of the initiation complex independently of a poly(A) tail [169, 171, 268]. However, the stoichiometry and relative spacing of DAZ/DAZL molecules on a RNA molecule can act to recruit additional repressor proteins (DAZAP, PUM2) with activity being dependent



on phosphorylation status [269-271]. Genes such as the meiosis-associated *sycp3* are regulated by all of the aforementioned regulatory layers, but the net result is a process of ordered and synchronized gene expression independent of changes in transcription [171, 272, 273]. In this context, NP1 and CIRBP may be playing critical roles in chromatin remodeling and controlling mRNA stability and/or translation, respectively.

In mammalian systems, CIRBP plays a number of diverse, yet highly integrated roles. It is classically recognized for stress granule formation in response to cellular stressors, including heat shock, UV irradiation, hypoxia, oxidative stress osmotic shock, or arsenic exposure [255, 274]. CIRBP is normally stored in the nucleus, and under stress conditions, it will translocate to the cytosol, bind target mRNAs, and associate into stress granules [255]. As “cold inducible” RNA binding protein, CIRBP received its name because of its role in testes, which operate  $\sim 3^{\circ}\text{C}$  below normal body temperature [275]. It was later identified as a major cold shock protein in *Arabidopsis*, and the *Arabidopsis* homolog could functionally replace non-homologous cold shock proteins in *E. coli* [276]. The temperature-dependent expression is a product of alternative promoter usage, producing a transcript with an elongated 5' UTR that contains an internal ribosome entry site [277]. Because of the seasonal nature of the mental gland and its presence in a cold-blooded animal, temperature-dependent expression in *P. shermani* seemed like a highly tractable hypothesis. However, 5' RACE experiments revealed that the CIRBP 5' UTR length was similar between June and August, and was closer in size to the short mammalian 5' UTR without the internal ribosome entry site (data not shown). Also, CIRBP IHC revealed exclusively cytoplasmic localization at both time points with relatively uniform labelling, not indicative of stress granules. In a recent study [256], the

low complexity domains of multiple RNA-BPs were studied and found to form hydrogels composed of  $\beta$  cross strands similar to amyloid fibers. Repeats of the tripeptide Gly/Ser-Tyr-Gly/Ser were determined to be essential for formation, and likely involved  $\pi$ - $\pi$  overlap between tyrosine residues. Using mutagenesis studies comparing a protein with 0, 12, 18, 22, or 27 tripeptide repeats, hydrogel size and stability were positively correlated with the number of repeats; in contrast, CIRBP only has 7 repeats, such that it may form smaller aggregates that are not visible microscopically. Given that CIRBP-RNA interactions induced a conformational change in CIRBP secondary structure which recruited additional CIRBP molecules through cooperativity, it is plausible that CIRBP may, essentially, coat the entire PMF 3' UTR (and possibly the whole PMF mRNA) in order to interfere with ribosome binding and translation. Given the dramatic changes in PMF mRNA abundance over gland development, this binding may also facilitate mRNA degradation. While it is generally thought that stress granules protect mRNA molecules from degradation [278, 279], their close proximity and possible association with processing bodies (P-bodies) has led to the hypothesis that there may be mRNA exchange between these macromolecular complexes that could lead to mRNA degradation under the proper cellular conditions [280, 281].

Similar to reports on the human homolog [255], we determined that both the RRM and LCD were required for stable, highly cooperative binding of CIRBP to the PMF 3' UTR. Structural comparison of different RRM-RNA complexes from other RNA-BPs has revealed that the RRM is a highly plastic structure with respect to nucleotide binding. The  $\alpha\beta$  sandwich structure contains two conserved regions of 7 and 6 amino acids (termed RNP1 and RNP2) which bind single stranded dinucleotides on either

DNA or RNA through a range of interactions, including  $\pi$ - $\pi$  overlap between aromatic residues and the nitrogen bases, hydrophobic interactions between additional aromatics and the sugar moieties, and/or salt bridges with the phosphate between the two nucleotides. Despite these conserved structural elements, sequence specificity is dictated by less conserved residues forming additional interactions [282]. RNA secondary structure also plays a critical role, with some RNA-BPs requiring dinucleotides to be part of stem loops or other types of internal loops [283-285]. When the dinucleotide frequency within the PMF 3' UTR was examined relative to predicted values based on nucleotide abundance, five of the sixteen combinations were observed at higher than expected frequencies: UC (+26%), UG (+58%), CA (+7%), CU (+29%), and GA (+8%). However, when these real vs expected distributions were compared by  $\chi^2$  test, UG contributed much more to the test statistic relative to the other dinucleotides with higher than expected frequencies ( $\chi^2_{UG} = 12.02$ , compared to  $\chi^2_{CU} = 3.30$ ). Future experiments examining both CIRBP-RRM specificity as well as PMF 3' UTR secondary structure will be required in order to determine if this interesting statistic has functional significance.

In addition to CIRBP, many other cold shock proteins were identified in the mental gland and had similar expression patterns (Figures 29, 30). All evidence suggests that CIRBP likely plays some critical role in facilitating the glands transition from growth and development into pheromone synthesis by interacting with PMF and likely other mRNAs for some key function (likely relating to translational repression and/or mRNA degradation). As already noted, while CIRBP was originally characterized based on its roles in temperature-dependent expression, more recent studies have demonstrated its roles in systemic and general stress response [255, 286]. When we also include factors

like NP1, and the transcription factors ATF3 (found in Cluster 1) and NRF2 (possibly present, based on identification of its inhibitor Keap1), the mental gland seems to express a preponderance of stress response genes [287, 288]. It is worth noting, again, how this quality further suggests that the mental gland may have a tumor-like phenotype [289, 290]. In addition to facilitating proliferation and development, the role of CIRBP as a regulator of pheromone synthesis is also interesting within the scope of gene co-option. A common theme among plethodontid pheromone genes, as well snake toxins and the products of other exocrine tissues, is gene duplication followed by rapid evolution in order to drive neofunctionalization for the acquisition of new functions [94, 115, Chapter II]. Through the same basic processes, it is plausible that novel regulators could be recruited and co-opted to provide tight control over gene expression of exocrine products in these systems. With many of these gene products having evolutionary histories of positive selection, it would likely be of value to explore both the products and their regulators in a co-evolutionary framework. At least in the case of PMF, CIRBP may have been the causative agent driving purifying selection on the UTRs while sexual selection from the female receptors promoted positive selection on the coding regions. The putative role of CIRBP directing the selective forces on the noncoding regions of PMF might be relevant to future studies of the regulatory elements of other rapidly evolving genes.

## Conclusions

For more than 100 million years, plethodontid salamanders have utilized a rapidly evolving system of non-volatile proteinaceous courtship pheromones to regulate female mating behavior in order to facilitate reproduction. In the red-legged salamander, *Plethodon shermani*, these pheromones are synthesized in a submandibular mental gland that annually hypertrophies into a large, pad-like structure whose transcriptional and translational machinery are almost exclusively programmed for pheromone synthesis. We now report that this highly effective system results from a tightly coordinated gene expression cascade, allowing for annual organogenesis followed by rapid conversion to a highly efficient pheromone factory. Glandular cells initially have an almost cancerous phenotype characterized by rapid proliferation and ECM dissolution, followed by a tremendous increase in pheromone mRNA levels. A key regulator in this process is Cold Inducible RNA Binding Protein: a stress-responsive RNA binding protein used by both animals and plants to store select mRNAs in stress granules and promote cell survival. For at least one pheromone, Plethodontid Modulating Factor, CIRBP selectively binds the 3' UTR and recruits additional molecules through cooperativity, with protein-protein and protein-RNA interactions likely stabilized through induced intermolecular  $\beta$  sheets. This interaction may inhibit translation of PMF mRNA and/or promote its degradation through association with P-bodies. The net result is suppression of pheromone translation until the gland is sufficiently large to support the storage of 10s to 100s of micrograms of pheromone, which can be used to increase receptivity in almost any female in the breeding population. CIRBP may be one player that has exerted purifying selection on the PMF UTRs, creating a system of disjunctive evolution with highly conserved

noncoding segments and rapidly evolving coding regions. The mechanisms behind this exciting dichotomy may serve as a model for future studies of gene regulation on rapidly evolving proteins.

## Methods

### Animal collection, gland removal, and pheromone extraction

*P. shermani* males were collected during their breeding season from a single site in Macon Co., North Carolina, USA (35°10'48" N, 83°33'38" W). Males were anesthetized in a mixture of 7% (v/v) diethyl ether in water. For analysis of total RNA, pheromone extract, or cellular proteins, mental glands were surgically removed using iridectomy scissors. For RNA analysis, glands were incubated overnight in RNAlater (Ambion, Austin, TX) at 4°C before long term storage at -20°C. Pheromone was extracted from mental glands based on the protocol in Chouinard et al. [91]: briefly, mental glands were individually incubated in 0.2 mL acetylcholine chloride (0.8mM in Amphibian Ringer's solution) for 30 min, centrifuged at 14,000  $\times g$  for 10 min, the supernatant collected, and the centrifugation repeated before storage at -80°C. Following extraction, mental glands were stored in RNAlater to allow preservation of RNA and cellular proteins. Pilot experiments confirmed that 30 min in the acetylcholine solution was sufficient to extract >90% of the total pheromone with no detectable RNA degradation (Wilburn and Feldhoff, unpublished data). Salamanders were collected under permits issued by the North Carolina Wildlife Resource Commission (to R.C. Feldhoff and D.B. Wilburn), and all animal protocols were approved by University of Louisville IACUC (#12041 to P.W. Feldhoff) and Highlands Biological Station IACUC (to D.B. Wilburn).

### cDNA preparation and transcriptome sequencing

Mental glands were collected from male *P. shermani* at six time points approximately every 3 weeks during 2010 (5/29, 6/19, 7/10, 8/1, 8/21, and 9/11). This

range preceded and spanned the principal August mating season. Five glands were collected at each time point and immediately stored in RNAlater. Total RNA was extracted from individual glands using the RNeasy kit (Qiagen, Valencia, CA) following the manufacturer's instructions. RNA concentrations were estimated by 260 nm absorbance. For each time point, standardized amounts of RNA were pooled from all five glands, and double-stranded cDNA prepared by oligo-dT priming using the SMARTer cDNA synthesis kit (Clontech, Palo Alto, CA). cDNA was supplied to Otogenetics Corporation (Norcross, GA) for library preparation and sequenced using the Illumina HiSeq 2000 platform (>20 million reads per time point, 100-bp paired end reads).

#### Transcriptome bioinformatics analysis

Illumina reads from all six time points were pooled and assembled into a single transcriptome using Trinity (r2012-10-05) [291]. Initial assemblies with default settings resulted in over-compaction of deBruijn graphs for PMF, limiting both isoform detection and full-length mRNA re-construction. Butterfly parameters were then optimized, and the final assembly included the additional settings *--min\_kmer\_cov 2 --bfly\_opts "--path\_reinforcement\_distance= 25 -min\_per\_id\_same\_path= 98"*. Reads from each time point were re-aligned to the full transcriptome using RSEM (v1.2.5) [248], with differential expression analysis conducted using EBSeq [249]. Expression differences were compared between adjacent time points (5/29 to 6/19, 6/19 to 7/10, etc.). Based on visual observations and analysis of pheromone extract from additional glands collected at each time point, the separate phases of gland development were best characterized by the 6/19 time point (growth/development) and the 8/1 time point (pheromone production); therefore, an additional comparison with EBSeq was performed between the 6/19 and 8/1



time points. For putative gene annotation, the Trinotate package (r2013-02-25) was used to determine (1) putative open reading frames (using TransDecoder, [291]), (2) protein BLAST (blastp) against both the SwissProt and TrEMBL databases [292-295], (3) orthologous group identification with eggNOG [296], (4) gene ontology assignment [297], (5) signal peptide prediction with SignalP [298], (6) protein family assignment with Pfam [299], and (7) transmembrane domain prediction with TmHMM [300]. As most of these databases searches relied on proper open reading frame assignment by TransDecoder, an additional BLAST search was performed with blastx using assembled nucleotide sequences against the full TrEMBL database [293, 295]. There were multiple cases of the TransDecoder proteins having no blastp hits in SwissProt or TrEMBL, yet the nucleotide sequence produced strong blastx hits (E-value < 0.001), suggesting that TransDecoder identified the wrong open reading frame. In these cases, an alternative open reading frame was selected based on the longest amino acid sequence that contained the aligned region of the blastx hit.

#### qRT-PCR analysis of differentially expressed genes

For select genes (see Table 12 for primers), qRT-PCR analysis was performed on RNA isolated from each mental gland used to construct the transcriptome (six time points each with 5 glands, 30 glands total). Total RNA was diluted to ~5-20 ng/uL, and accurate concentrations were determined using Quant-iT RiboGreen RNA assay kit (Invitrogen, Carlsbad, CA). qRT-PCR reactions were performed in triplicate using 20  $\mu$ L reactions with the Power SYBR Green RNA-to- $C_T$  1-Step kit (Ambion) containing 1  $\mu$ L diluted total RNA. Expression levels were calculated by the pcrfit function in the R package *qpcR* using the cm3 mechanistic model [301].

**Table 12. Primers used for qRT-PCR, CIRBP expression, and *in vitro* transcription.**

Primer	Sequence
<b>qPCR primers</b>	
PMF Class I 3'UTR 26-45 F'	TGA GGA ATC ACA CTG GCA TC
PMF Class I 3'UTR 103-82 R'	GCT CCT TGT TTT CTC TGG TAG C
PRF 3'UTR 97-116 F'	ATC GTG GGA TGG ACA GTT TG
PRF 3'UTR 186-167 R'	TCG GGA AAG ATG AGG ACT TC
PTP CDS 79-98 F'	CAC CTG CAG TCA GCT TTT TG
PTP CDS 164-145 R'	ATC CAG CCA TCA TCA TCC TC
CIRBP CDS 44-63 F'	TTG ACA CAA ACG AGC AGG AC
CIRBP CDS 187-168 R'	CGT CTT TCG CAT CTT CTT GG
Actin CDS 134-153 F'	TTG GTA TGG GCC AGA AAG AC
Actin CDS 241-222 R'	CAT CCC AGT TGG TGA CAA TG
Cathepsin S CDS 220-239 F'	GCA GCA AAG CTT GAT CTT CC
Cathepsin S CDS 359-340 R'	ACA CGG TCT GAA ATG GCT TC
Cystatin C CDS 134-153 F'	TTG CCA TGA CCG AGT ACA AC
Cystatin C CDS 275-256 R'	GGG TGT GTG CAA GTT GTT TG
PCNA 200-219 F'	CTG TGG GGG TTA AAA TGA GC
PCNA 296-277 R'	ATG GTG TCG GCA TTG TCT TC
Androgen Receptor 2-21 F'	TGG AAG CCA TTG AGC CTA TC
Androgen Receptor 147-128 R'	AAA ACC TGG TAG GGC TTT CG
Keratin 488-507 F'	TCG ATA AGG TCC GAT TCC TG
Keratin 603-584 R'	TAA GGG CTC CAT GTT GTT CC
VEGF 90-109 F'	CTT CAT GCC ATC TTG TGT GC
VEGF 226-207 R'	CCA CGA GTT TGT TTC GAT GC
GAPDH 523-542 F'	ACT GTG CAT GCG ATT ACT GC
GAPDH 622-603 R'	AGG CTG GAA TGA TGT TCT GG
AChE 1577-1596 F'	TCA GCA GAA GGA TGA TGC AC
AChE 1708-1689 R'	GAT TGG TGT TGA GTG CGA TG
NP1 209-228 F'	ACA ACA CGA ACC GCT TCA AC
NP1 284-265 R'	TTC TTC TCC TCG CTG TTG TG
CSPE1 1518-1537 F'	TGG CCT GAA TTT AGG AGA CG
CSPE1 1595-1576 R'	TTA ACC CTT TCC GCA CTG AC
HSC70 91-110 F'	AAT GAC CAG GGC AAC AGA AC
HSC70 174-155 R'	CTG ATT TTT AGC GGC GTC TC

<b>Expression primers</b>	
EYFP 1-20 + Kpn I F'	GCG CGG TAC CATGGTGAGCAAGGGCGAGGA
EYFP 717-700 + LnkA R'	GCTGCCTCCTGCAGCGGCCGCTCCGGACTTGTACAGCT CGTCCAT
Ps CIRBP 1-23 + LnkA/DP F'	GCT GCA GGA GGC AGC GAT CCC ATG TCT TCA GAT GAT GGG AAG AT
Ps CIRBP 253-274 + LnkA/DP F'	GCT GCA GGA GGC AGC GAT CCC TCT GAC CGA AGC CGA GGC GGC T
Ps CIRBP 252-231 + TAA + Hind R'	CAT GAA GCT TTT ATT TGC CCG CTT GGT CCA CTC GG
Ps CIRBP 498-478 + HindIII R'	GCG CAA GCT TTT AGT TAT CAT AGC TGT CTC T

<b><i>in vitro</i> transcription primers</b>	
PMF 3' UTR 26-45 + pT7 F'	TAA TAC GAC TCA CTA TAG GTG AGG AAT CAC ACT GGC ATC
PMF 3' UTR 99-118 + pT7 F'	TAA TAC GAC TCA CTA TAG GGG AGC AAA GCT TCT TTG ACG
PMF 3' UTR 271-290 + pT7 F'	TAA TAC GAC TCA CTA TAG GGG GTT AGT GTG ATG GGA AGG
PMF 3' UTR 406-427 + pT7 F'	TAA TAC GAC TCA CTA TAG GTC TGA AAA TGT GGA AGC AGA AA
PMF 3' UTR 288-268 R'	TTC CCA TCA CAC TAA CCC AGT
PMF 3' UTR 368-349 R'	ACA ACT TGG GTG GCA TCA TT
PMF 3' UTR 565-544 R'	ACA TGG GAA ATT CAG AAA CAG A
PMF 3' UTR 667-647 R'	GCA CCA TCA AAC AAG ACT TCC
Keratin 3' UTR 205-224 + pT7 F'	TAA TAC GAC TCA CTA TAG GGA TTG GCT GAC ATT CCA CCT
Keratin 3' UTR 441-422 R'	CAT TCC CCG GTG TAA GAA TG

Based on the gross morphological changes of the mental gland, it was expected that few genes would be stably expressed across mental gland development; therefore, RNA input was used to normalize qPCR measurements, with the literature supporting that input is often a more robust reference than housekeeping genes [302, 303]. Expression levels for each gene were normalized based on the time point with the highest expression, and a single linear mixed effect model was fit by maximum likelihood using the `lme` function in the R package *lmer*. The model included fixed effects for gene and time, with male as a random effect; all variables, including the interaction between gene and time, were significant at  $p < 0.001$ .

#### Expression of recombinant CIRBP

To conduct *in vitro* protein-RNA binding assays, recombinant CIRBP was prepared using an *E. coli* expression system. Preliminary experiments revealed that CIRBP was insoluble at  $> 0.2$  mg/mL in all tested buffers (data not shown); however, solubility was dramatically improved when recombinant CIRBP was expressed as a fusion protein with enhanced cyan fluorescent protein (rCIRBP/ECFP). Simultaneously, fusion proteins were prepared with only the RRM-containing N-terminus (residues 1-84; rCIRBP-RRM/ECFP) and the glycine-rich C-terminus (residues 85-165; rCIRBP-LCD/ECFP). All constructs included ECFP on the N-terminus, a short hydrophilic linker (SGAAAAGGSDP), and the CIRBP element at the C-terminus. The CIRBP coding regions were amplified from mental gland cDNA using the Accuprime High Fidelity *Taq* polymerase system (Invitrogen) (see Table 4 for primers). ECFP was amplified from a pcDNA3.1-based vector (supplied by Dr. Ronald Gregg, University of Louisville), and fusion genes were prepared by modified assembly PCR [185]. Fusion PCR products were

purified using the QIAquick PCR cleanup system (Qiagen), and cloned the pET45b vector (Novagen, San Diego, CA) following restriction digest with *KpnI* and *HindIII* (New England Biolabs, Ipswich, MA), gel purification of cleavage products (GFX purification system, GE Healthcare, Piscataway, NJ), ligation using T4 DNA Ligase (New England Biolabs), and transformation into T7 Express lysY/Iq chemically competent *E. coli* (New England Biolabs). Sequence of transformed DNA was validated by colony PCR and Sanger sequencing by the University of Louisville DNACore facility. For protein expression, clones were cultured overnight at 28°C in LB media with 100 µg/mL ampicillin (LB/Amp), diluted in 1 L LB/Amp with an 600 nm optical density (OD<sub>600</sub>) equal to ~0.05, incubated with shaking until the OD<sub>600</sub> equaled ~0.7, IPTG added to a final concentration of 0.1 mM, and incubated overnight. *E. coli* were harvested by centrifugation, resuspended in 50 mM NaCl/0.1% Triton X-100/2 mM EDTA/50 mM Tris, pH 8, and lysed by sonication followed by lysozyme treatment (final concentration 1 mg/mL) for 1 hour. Insoluble material was removed by centrifugation, and proteins concentrated by ammonium sulfate precipitation (final concentration 70%). Ammonium sulfate pellets were resolubilized in Ni-NTA binding buffer (0.5 M NaCl/5 mM Imidazole/20 mM Tris, pH 8), and passed through a 15 mL Ni-NTA column (Thermo-Pierce, Rockford, IL) at 1 mL/min. The column was then washed and eluted with increasing concentrations of imidazole (all in 0.5 M NaCl/20 mM Tris, pH 8): 10 column volumes (CVs) at 5 mM, 2 CVs at 20 mM, 1 CV at 40 mM, 1 CV at 60 mM, and finally 3 CVs at 200 mM (collected in 5 mL fractions). Highly fluorescent fractions were pooled, concentrated, and buffer exchanged to 1X EMSA Buffer (100 mM KCl/2 mM EDTA/0.05% Tween-20/20 mM HEPES, pH 7.4) by ultrafiltration (YM-10 Centiprep,

Millipore). All fusion products were standardized to equal molar concentrations (rCIRBP/ECFP: 1.4 mg/mL, rCIRBP-N/ECFP: 1.1 mg/mL, rCIRBP-C/ECFP: 1.0 mg/mL)

#### Electrophoretic mobility shift assays

CIRBP-RNA interactions were characterized by electrophoretic mobility shift assays (EMSAs). RNA was prepared by *in vitro* transcription using the T7 High Yield RNA Synthesis Kit (New England Biolabs) with purified PCR products amplified from *P. shermani* 8/1 cDNA using primers that included engineered T7 promoters (Table 12). Synthesized RNA was subsequently treated with Turbo DNase I (Ambion), and purified using the RNeasy Kit (Qiagen). For fluorescently labeled RNA, *in vitro* transcription reactions were adjusted to include 7.5 mM UTP and 2.5 mM aminoallyl-UTP (Ambion), treated with TURBO DNase I (Ambion), purified using the RNeasy kit (Qiagen), and adjusted to 0.4 mg/mL with 2.5 mg/mL TAMRA-carboxylic acid (Invitrogen) in 0.1 M MES (pH 5). One-fourth volume of EDAC (0.1 mM) was added to the reaction, incubated for 2 hours in the dark, and then labelled RNA was purified using the RNeasy Kit (Qiagen). All EMSAs were performed as 15  $\mu$ L reactions in 1X EMSA Buffer with 3% glycerol using either 300 ng unlabeled or 45 ng labeled RNA. rCIRBP/ECFP (or the domain truncations), competitors, and other reagents were added at the listed concentrations, and incubate at room temperature for 20 minutes prior to gel loading. For crosslinking assays, formaldehyde was added to the protein-RNA mixture at a final concentration of 1%, incubated for 5 min, and the reaction quenched by addition of glycine to a final concentration of  $\sim$ 330 mM prior to gel loading. RNA was separated using agarose gels of 2% (for PMF 3' UTR 26-667) or 3% (for  $\sim$ 250 nt products), and

electrophoresed for 3 hours at 80 V. For unlabeled RNA, gels were stained with Sybr Green II RNA Stain (Invitrogen) for 30 min. Gels were imaged using a Typhoon 9400 fluorescent bed scanner (GE Life Sciences, Piscataway, NJ) with the appropriate laser and filter settings. For fluorescence quenching assay, 0.5  $\mu\text{g}$  of TAMRA-labeled RNA was incubated with increasing concentrations of different rCIRBP constructs in a 200  $\mu\text{L}$  reaction, and fluorescence measured using the Synergy2 plate reader (Biotek, Winooski, VT). Data were fit to the Hill equation ( $\theta = \frac{1}{(\frac{K_A}{[L]})^{n+1}}$ ) using the nlsLM function in the R package *minpack.lm*.

### Antibody preparation

PRF antisera was prepared by immunizing two rabbits with PRF fractions collected by anion-exchange chromatography (performed by RCF at UofL). The antigen was >90% enriched for PRF, but contained ~5% PTP such that there was significant immunoreactivity for both proteins. CIRBP antisera was prepared by immunizing two rabbits with purified rCIRBP/ECFP (~98%) using Ni-NTA resin (performed by the Proteintech Group, Chicago, IL). Highly purified antigens (>99% by reverse phase HPLC) were coupled to agarose matrices for affinity purification of antibodies: briefly, 6% crosslinked agarose beads (CL6B; Sigma-Aldrich, St. Louis, MO) were activated with carbonyl diimidazole (CDI; Sigma-Aldrich) in dry acetone for 20 minutes, quickly rinsed under vacuum with cold distilled water, and then incubated in antigen solutions (~2 mg/mL in 100mM KCl/0.05% Tween-20/100mM NaCO<sub>3</sub>, pH 11) overnight with gentle agitation. Resins were packed into ~1-3mL columns (depending on the starting amount of antigen), remaining active sites blocked with 50 mM Tris, pH 10, and equilibrated in 1X phosphate buffered saline (PBS). Antibodies were purified by

incubating antisera with the resin overnight at 4°C, washing with 10 CVs of 1X PBS to collect unbound antibodies, stringently washed with 5 CVs of 0.5 M NaCl/0.05% Tween-20/20 mM Tris, pH 7.5 (TTBS), and eluted with 6 CVs 0.1 M citric acid/0.1% Tween-20, pH 3 collected in 0.2 mL fractions. Elution fractions were neutralized with 1 M Na<sub>2</sub>HPO<sub>4</sub>, and antibody-containing fractions determined by absorbance measurements at 280 nm. Fractions were then pooled, concentrated, and buffer exchanged to 1X PBS by ultrafiltration (YM-30 Centriprep; Millipore, Billerica, MA). Four antigen columns were prepared: PRF, PTP, rPRF/ECFP, and rCIRBP-RRM/ECFP. PRF and PTP antibodies were serially purified from the same antisera by first passing whole antisera through the PRF column for purification of anti-PRF, and the flowthrough (i.e., unbound antibodies) then passed over the PTP column for purification of anti-PTP. Both solutions of affinity-purified antibodies were then adsorbed against the opposing column to ensure removal of non-specific or low-affinity antibodies. anti-CIRBP-RRM was purified by passing rCIRBP/ECFP antisera over the rPRF/ECFP column to remove anti-ECFP, the flowthrough next incubated with the rCIRBP-N/ECFP column, anti-CIRBP-RRM purified, and the two-column process repeated to ensure removal of non-specific or low-affinity antibodies. Antibody specificity was validated by western blot.

#### Histological analysis

For histological analysis of mental glands, male salamanders were anesthetized in 7% ether, sacrificed by rapid decapitation, the lower jaw removed, immediately placed in 4% paraformaldehyde (in 150 mM NaCl/100 mM Na<sub>2</sub>HPO<sub>4</sub>, pH 7.4), and incubated overnight with gentle agitation. Fixed jaws were decalcified by incubation in 10% EDTA (pH 7.4, DEPC-treated) for 48 hours, cryoprotected in 30% sucrose for 48 hours,



embedded in Optimum Cutting Temperature (OCT) media (Sakura-Finetek, Torrance, CA), and stored at  $-80^{\circ}\text{C}$  prior to cryosectioning. Lower jaws were sectioned coronally at a thickness of  $16\ \mu\text{m}$  (for immunohistochemistry) or  $40\ \mu\text{m}$  (for structural morphology), and thaw mounted onto superfrost plus slides pre-coated with polylysine. Slides were stored at  $-20^{\circ}\text{C}$  until analyzed. Hematoxylin/eosin staining was performed using standard protocols [304]. For immunohistochemical labeling, slides were first equilibrated to room temperature for 30 min and washed five times with 1X PBS for 5 min each. Antigen retrieval was conducted by incubating the slides in 10mM Citric Acid (pH 6)/0.05% Tween-20 at  $70^{\circ}\text{C}$  for 30 min, cooled to room temperature for 20 min, and washed twice with PBS with 0.05% Tween-20 (PBST) for 5 min each. Blocking was performed for 30 min with 1X PBS/0.1% BSA/0.5% Tween-20, and immediately incubated overnight with primary antibody (anti-PRF at  $7.5\ \mu\text{g}/\text{mL}$  or anti-CIRBP-RRM at  $3.5\ \mu\text{g}/\text{mL}$ ) in PBST with 0.1% BSA. Slides were then washed five times with PBST for 5 min each, incubated with secondary antibody (Alexa Fluor 633 goat anti-rabbit IgG (Invitrogen),  $1\ \mu\text{g}/\text{mL}$  in PBST) for 30 min, washed five times with PBST for 5 min each, and finally counterstained with Alexa Fluor 488-wheat germ agglutinin (Invitrogen;  $10\ \mu\text{g}/\text{mL}$  in PBST) and DAPI (Invitrogen;  $3.6\ \mu\text{M}$ ). For gland morphology, thick sections ( $40\ \mu\text{m}$ ) were stained with Alexa Fluor 488-wheat germ agglutinin, Alexa Fluor 555-Phalloidin (Invitrogen;  $0.005\ \text{U}/\mu\text{L}$ ), and DAPI. Slides were coverslipped with Prolong Gold Antifade reagent (Invitrogen), cured overnight in the dark, and stored at  $4^{\circ}\text{C}$ . Imaging was accomplished using an Olympus Fluoview FV1000.

### Circular Dichroism

To measure secondary structure changes in rCIRBP/ECFP with different concentrations of RNA, far-UV CD spectra (185-255 nm) were acquired by averaging 10 scans across a 0.1-cm path at 0.2 nm intervals using a Jasco J-810 Spectropolarimeter. Stock rCIRBP/ECFP (1.4 mg/mL in 1X EMSA) buffer was diluted 10-fold in water (to reduce chloride background). Following initial measurements with no RNA, PMF 3' UTR 26-667 (at 70 ng/μL in 0.1X EMSA Buffer) was titrated into the solution, incubated for 5 min, and CD spectra recorded. Spectra were adjusted for slight changes in volume/concentration following serial dilution throughout the experiment. At the concentrations used, PMF 3' UTR 26-667 produced no significant CD signal over buffer.

### Quantification and modelling of CIRBP expression

For two separate time points in 2013 (6/13 and 8/3), 15 male *P. shermani* were collected, mental glands removed, pheromone extracted, and mental glands stored in RNAlater (Ambion) as previously described. One gland for 8/3 was inadvertently destroyed, such that  $n = 29$ . Pheromone concentrations were accurately determined by BCA Protein Assay (Thermo-Pierce), and normal proportions of pheromone components (PMF:PRF:PTP ~ 5:3:1) were validated by RP-HPLC. Stored mental glands were later dissected into approximately equal halves in order to extract both RNA (RNeasy Kit, Qiagen) and cellular protein (homogenization in RIPA Buffer, supplemented with 0.1 mM iodoacetamide). Pilot studies confirmed that the two halves are equivalent for levels of target proteins and mRNA. CIRBP protein levels were measured by western blot. Preliminary experiments suggested that glands from 6/13 had ~6X higher levels of CIRBP, such that, in order to maintain similar blot intensities for quantification, 5 μg was

loaded per lane for 6/13 glands and 30  $\mu\text{g}$  per lane for 8/3 glands. Briefly, proteins were separated using a 15% Tris-Tricine gel with 4% stacking gel at 50 V for 15 minutes followed by 100 V for 90 min, transferred to PVDF membranes by electroblotting in 15% methanol/0.025% SDS/10 mM CAPS, pH 11 at 55 V for 75 min. The membrane was blocked with 0.5 M NaCl/0.5% Tween-20/20 mM Tris, pH 7.5 for 1 hour, incubated with  $\alpha\text{CIRBP-RRM}$  (1.4  $\mu\text{g}/\text{mL}$ ) for 1 hour, then incubated with alkaline phosphatase-conjugated goat anti-rabbit IgG (1  $\mu\text{g}/\text{mL}$ ; Sigma-Aldrich), and developed using BCIP/NBT. All membrane washes and antibody dilutions were performed using TTBS (0.5 M NaCl/0.05% Tween-20/20 mM Tris, pH 7.5). All blots were processed and developed simultaneously in order to minimize run-to-run variation, and a reference of 10 ng rCIRBP/ECFP was included on each membrane for normalization. Densitometry analysis was performed using ImageJ. Extracted RNA was used for qRT-PCR analysis for PMF, CIRBP, and Cathepsin S based on the previous methods (see Table 12 for primers). Correlation between variables was determined by multivariate ANOVA (MANOVA).

#### Immunopull-down and mass spectrometry of CIRBP

To validate the identity of different bands visualized during western blot, immunopull-down experiments were performed. Anti-CIRBP-RRM was used for CIRBP pull-down, and as a negative control, an IgG-enriched fraction from pre-immunization serum of CIRBP immunized rabbits was prepared by ammonium sulfate precipitation and DE52 chromatography. Using Protein G coupled Dynabeads (Invitrogen), 10  $\mu\text{g}$  of antibody was adsorbed to the beads, stringently washed with PBST, incubated with pooled RIPA extract from 6/13 (~180  $\mu\text{g}$ ) or 8/3 (~840  $\mu\text{g}$ ) for 30 min, washed with

PBST, the beads transferred to a clean 1.7 mL tube, and incubated with 1X gel loading buffer (1% SDS/12% glycerol/0.005% bromophenol blue/50 mM Tris, pH 6.8) at 65°C for 30 min. The complete samples were loaded into a 15% Tris-Tricine gel with 4% stacking gel, electrophoresed at 50V for 15 min followed by 100 V for 90 min, and stained using SYPRO Ruby fluorescent gel stain (Invitrogen). The gel was imaged using the Typhoon 9400 fluorescent bed scanner (GE Life Sciences). All bands were then individually excised, proteins reduced and alkylated with dithiothreitol/iodoacetamide, treated overnight with trypsin, peptides isolated by acetonitrile extraction, and supplied to the University of Louisville Biomolecular Mass Spectrometry Core Laboratory for analysis by LC/MS-MS.

CHAPTER VII - PROTEOMIC ANALYSES OF COURTSHIP PHEROMONES  
IN THE REDBACK SALAMANDER, PLETHODON CINEREUS

**Chapter Overview**

The evolutionary success of plethodontid salamanders for ~100 MY is partly due to the use of courtship pheromones that regulate female receptivity. In ~90% of plethodontid species, males deliver pheromone by “scratching” a female’s dorsum, where pheromones diffuse transdermally into the bloodstream. However, in a single clade, representing ~10% of *Plethodon spp.*, males apply pheromones to the female’s nares for olfactory delivery. Molecular studies have identified three major pheromone families: Plethodontid Receptivity Factor (PRF), Plethodontid Modulating Factor (PMF), and Sodefrin Precursor-like Factor (SPF). SPF and PMF genes are relatively ancient and found in all plethodontid species; however, PRF is exclusively found in the genus *Plethodon* – which includes species with transdermal, olfactory, and intermediate delivery behaviors. While previous proteomic analyses suggested PRF and PMF are dominant in slapping species and SPF is dominant in non-*Plethodon* scratching species, it was unclear how protein expression of different pheromone components may vary across delivery modes within *Plethodon*. Therefore, the aim of this study was to proteomically characterize the pheromones of a key scratching species in this evolutionary transition, *Plethodon cinereus*. Using mass spectrometry-based techniques, our data support the functional replacement of SPF by PRF in *Plethodon spp.* and an increase in PMF gene

duplication events in both lineage-dependent and delivery-dependent manners. Novel glycosylation was observed on *P. cinereus* PRFs, which may modulate the metabolism and/or mechanism of action for PRF in scratching species. Cumulatively, these molecular data suggest that the replacement of pheromone components (e.g., SPF by PRF) preceded the evolutionary transition of the functional complex from transdermal to olfactory delivery.

## **Introduction**

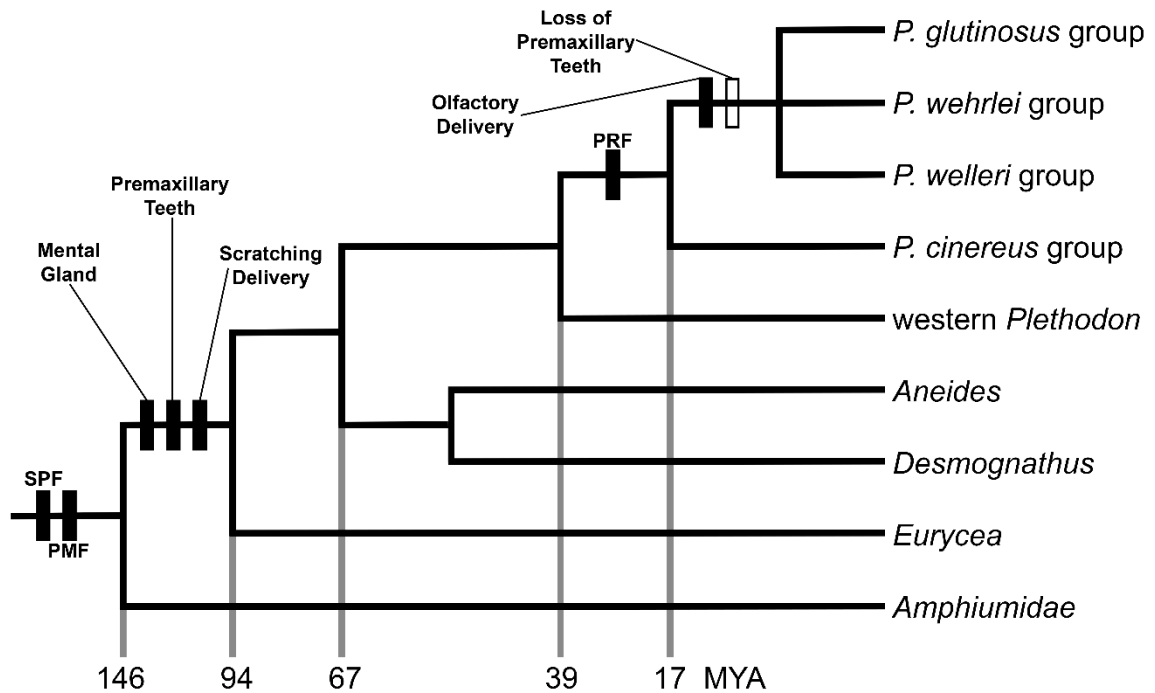
For many organisms, intricate biological processes are often the product of functional complexes: diverse, yet integrated, sets of characters that interact in order to perform a specific function or task [127]. A prime example of a functional complex is venom delivery in snakes, spiders, and cone snails: in order to capture prey, multiple organs and specific behaviors have evolved for the synthesis and injection of potent toxins that rapidly immobilize the target [115, 175, 305]. Because all of the elements in a functional complex behave like gears in a single machine, their interdependence on one another can act as a strong selective tether; consequently, genes in such a system would be expected to co-evolve at similar rates. In contrast, in the functional complex of courtship pheromones in plethodontid salamanders, observed stasis in morphology, physiology, and behavior masks signatures of rapid evolution in the pheromone molecules themselves [87, 94].

For more than 100 million years [50], plethodontid salamanders have utilized an array of pheromones to coordinate courtship and facilitate mating success. During the annual mating season, plasma androgen levels rise in male salamanders and induce development of a submandibular mental gland that is solely dedicated to the synthesis of non-volatile, proteinaceous courtship pheromones [69, 214]. As part of an intricate courtship behavior termed tail straddling walk, most species deliver these pheromones to female salamanders by one of two behaviors in order to modify the female's mating receptivity and behavior [83, 87]. In the majority of plethodontid species (~300 spp.), the mental gland is a small, subdermal structure at the tip of the chin. Nearly all species with this type of mental gland apply pheromone using hypertrophied premaxillary teeth to

“scratch” the female’s back. In this system, pheromones presumably diffuse into the female bloodstream [44, 45]. However, approximately 17 MYA in a single clade of large eastern *Plethodon species* (the *P. glutinosus* complex, 28 spp.; Figure 40), pheromone delivery transitioned to direct application from a larger, external, pad-like mental gland to the females nares, whereupon pheromones bind to specific receptors in the female vomeronasal organ and stimulate regions of the brain involved in regulating reproductive behavior [70, 83, 86]. More basal species in this genus still retain the ancestral scratching delivery system (the *P. cinereus* complex; 10 spp.) where pheromone is applied to the dorsum, yet 6 species employ both and/or intermediate behaviors where the precise location of pheromone delivery is unknown (*P. wehrlei* and *P. welleri* groups; 2 and 4 spp., respectively) [77, 80, 306, 307]. Transdermally delivered pheromones do not have direct access to the receptors in the olfactory pathway [216], but must have receptors some place that allow a response.

Two model salamander species (*P. shermani* using olfactory delivery, and *Desmognathes ocoee* using scratching delivery) have been used to chemically characterize the three major pheromone components [69, 81]. In *P. shermani*, ~85% of the total pheromone is comprised of two proteins: Plethodontid Receptivity Factor (PRF), a 22-kDa protein related to IL-6 cytokines, and Plethodontid Modulating Factor (PMF), a 7-kDa protein related to three-finger protein superfamily which includes snake neuro- and cytotoxins, complement system regulators, and receptors in the plasminogen activation system [Chapter II]. Both of these proteins persist in the pheromone extract as multi-isoform blends, but with PMF having more than 30 highly diverse isoforms compared to 3 major PRF isoforms [91].





**Figure 40. Phylogenetic relationship among plethodontid salamanders.**

The figure is adapted from Palmer et al. [97], updated with relationships between plethodontid genera and approximate divergence times are based on maximum-likelihood trees constructed by Vieites et al. [50] for three nuclear genes (RAG1, BDNF, POMC); relationships between eastern *Plethodon* spp. are based on more recent findings by Highton et al. [308] using albumin sequence data. Trait acquisition and loss are denoted by solid and empty boxes, respectively.

In contrast, the pheromone extract of the more ancestral species *D. ocoee* is more complex and contains many more types of proteins; however, the most abundant protein, Sodefrin Precursor-like Factor (SPF), has been characterized as a 20-kDa protein with sequence similarity to the newt pheromone precursor and a family of phospholipase A2 inhibitors [81, 97]. Additionally, *D. ocoee* expresses a single isoform of PMF. At the proteomic level, there is no evidence of PRF in *D. ocoee* or SPF in *P. shermani*, although SPF mRNA can be detected at trace levels in *P. shermani* [89]. Staged courtship trials have successfully demonstrated behavioral effects when female salamanders receive purified PRF or PMF in *P. shermani*, or SPF in *D. ocoee* [81, 83, 93]

Molecular surveys of the expression of mRNAs for these three pheromone components across the plethodontid phylogeny demonstrate that while PMF and SPF are ubiquitously found in all plethodontid species, PRF is restricted to the four eastern *Plethodon* groups [95]. Relative stasis has been observed in many elements of the functional complex, such as mating behavior (tail straddling walk), morphology (mental gland), and delivery mode (scratching vs. olfactory); yet models of molecular evolution revealed that all three pheromone families display signatures of positive selection [96]. In groups co-expressing PRF and SPF, the two components displayed opposing signatures of selection (i.e. when PRF displayed signatures of positive selection, SPF did not, and vice versa). It was therefore proposed that PRF is replacing SPF in a lineage specific manner [97]. However, because of the extremely high sensitivity of PCR, the presence of PRF, PMF, and SPF mRNA in all four *Plethodon* groups does not equate to protein expression. With proteomic data only available for a single slapping species (*P. shermani* in *P. glutinosus*), it was unclear how the other pheromone components would be

represented in *Plethodon* groups utilizing scratching or intermediate behaviors. An additional limitation of the previous studies was that plethodontid pheromone mRNAs appear to be disproportionately translated, such that RT-PCR analysis does not make a good proxy for protein expression [Chapter II]. Therefore, in order to resolve the ambiguities in how plethodontid pheromones might be evolutionarily replacing one another across the mode of delivery transition, the aim of this study was to proteomically characterize the mental gland extract of a basal *Plethodon* species with scratching delivery, *Plethodon cinereus*.

## **Materials and methods**

### Gland excision and protein extraction

*Plethodon cinereus* males were collected during their breeding season from Giles Co., Virginia, USA (37°22'22" N, 80°32'31" W). Males were anesthetized in a mixture of 7% ether in water. Pheromones were extracted following the methods of Rollmann et al. [83]. Approximately 20 glands were excised and pheromone secretion was stimulated by incubation with 0.8 mM acetylcholine chloride in Amphibian Ringer's Solution for ~60 minutes. Whole pheromone extract was centrifuged at 10,000  $\times g$  for 10 minutes, the supernatant collected, and the centrifugation repeated before storage of supernatant at -20°C. Prior to all proteomic analyses, the pheromone extract was concentrated using a YM-3 Centriprep (Millipore, Billerica, MA), and buffer exchanged to 0.5X phosphate buffered saline. For molecular biological analyses, seven additional mental glands were excised and immediately incubated in RNeasy lysis buffer (Qiagen, Crawfordsville, IN) at 4°C overnight prior to long term storage at -20°C. Methods and animal care were approved by the University of Louisville's Institutional Animal Care and Use Committee (IACUC #12041 to P.W. Feldhoff).

### Mass spectral characterization and sequencing of *P. cinereus* protein pheromones

Individual mental gland proteins were purified by reverse phase-HPLC (RP-HPLC) (C-18, 0.46 x 15 cm; Grace Davison Discovery Sciences, Deerfield, IL) using a 2695 Alliance HPLC System equipped with a 2487 dual wavelength absorbance detector and Empower software (Waters Division, Milford, MA) with an elution gradient 0-70% acetonitrile (ACN) at 1% ACN/min. Both whole pheromone extract and individual RP-HPLC fractions were analyzed by SDS-PAGE using 15% tris-tricine separating gels and

4% stacking gels, with electrophoresis at 50 V for 15 min followed by 100 V for 75 min [309]. Gels were stained using Coomassie Brilliant Blue R250.

Following SDS-PAGE of individual RP-HPLC fractions, protein bands were excised using a sterile scalpel blade, then cut into  $\sim 1 \text{ mm}^3$  cubes, and placed in a 1.7 ml tube. Remaining Coomassie dye was removed by multiple extractions using 25  $\mu\text{L}$  ammonium bicarbonate (50 mM), incubated for 15 minutes, 50  $\mu\text{L}$  acetonitrile added, incubated another 15 minutes, the supernatant removed, the gel pellet dried using a vacuum centrifuge, and the process repeated. Once the dye was completely removed, disulfide bonds were reduced by incubating the gel pellet in 50  $\mu\text{L}$  DTT (20 mM)/ammonium bicarbonate (100 mM) at 56°C for 45 minutes, followed by alkylation with 50  $\mu\text{L}$  iodoacetamide (55 mM)/ammonium bicarbonate (100 mM) for 30 minutes in the dark. After two washes using 0.1 mL ammonium bicarbonate (100 mM) and 0.1 mL acetonitrile, the pellet was incubated with 50  $\mu\text{L}$  trypsin (20 ng/ $\mu\text{L}$  in 50 mM ammonium bicarbonate) overnight at 37°C. The supernatant was then collected, and the pellet washed twice using 25  $\mu\text{L}$  ammonium bicarbonate (50 mM) and 25  $\mu\text{L}$  acetonitrile with the washes added to the collected supernatant. The collected solution was then lyophilized, resolubilized in 10  $\mu\text{L}$  acetonitrile (50%)/formic acid (0.1%), and submitted to the University of Louisville Mass Spectroscopy Core Facility for analysis by liquid chromatography-tandem mass spectroscopy (LC/MS-MS). All mass spectral data were analyzed using MassMatrix 2.4.2 [184]

Initial peptide identification was performed by searching the mass spectral data against an unannotated draft *P. cinereus* mental gland transcriptome (Wilburn and Feldhoff, unpublished data). Briefly, *P. cinereus* mental gland cDNA was prepared using

the SMARTer cDNA synthesis kit (Clontech, Palo Alto, CA), and for the most probable transcriptome match, primers were designed using primer3 [310] (Integrated DNA Technologies, Coralville, IA) in order to ascertain the full sequence by PCR and Sanger sequencing. Targets were PCR amplified using the Accuprime High-Fidelity *Taq* system (Invitrogen, Carlsbad, CA) with a thermocycler program of 94°C for 2 minutes, followed by 35 cycles of 94°C for 15 seconds, 57°C for 15 seconds, and 68°C for 3 minutes. PCR products were examined by gel electrophoresis, and following successful amplification, were cloned into pCR2.1 (Invitrogen) using T4 DNA Ligase (New England Biolabs, Ipswich, MA), transformed into TOP10 chemically competent *E. coli* (Invitrogen), and plated on LB agar plates with 100 µg/mL ampicillin (Sigma-Aldrich, St. Louis, MO). Colony PCR was performed on resulting clones using routine PCR (reagents from New England Biolabs) with M13 primers (Integrated DNA Technologies), and DNA supplied to the University of Louisville DNA Sequencing Core Facility for sequencing of both forward and reverse strands. All sequence data was processed using the DNASTAR package (Lasergene v.10.0.1; DNASTAR, Madison, WI). MS results were searched against a database of these new sequences as well as those reported by Palmer et al. [95-97] (AY499379- AY499400; DQ097017-DQ097020; DQ882525-DQ882529) for final identification. All new sequences were deposited in NCBI Genbank (Accession #KJ187079-KJ187106).

#### Pheromone glycosylation analysis

The presence of glycoproteins in whole pheromone extract from *D. ocoee*, *P. shermani*, and *P. cinereus* was evaluated by separating 15 µg of pheromone by SDS-PAGE (as described above), and staining the gel using the Pro-Q Emerald 488

glycoprotein staining kit (Invitrogen) which selectively labels reducing sugars with a fluorescent tag following periodate oxidation. Fluorescent imaging was performed using the Typhoon 9400 laser bed scanner (GE Life Sciences, Piscataway, NJ). Following imaging, the gel was stained with Coomassie Brilliant Blue R250, re-imaged using the Typhoon 9400, and the two images pseudo-colored and overlaid in order to simultaneously visualize carbohydrate and protein staining. Ovalbumin in the protein ladder (M3913; Sigma-Aldrich) served as a positive control for successful glycoprotein staining. The protocol was repeated for multiple individual RP-HPLC fractions.

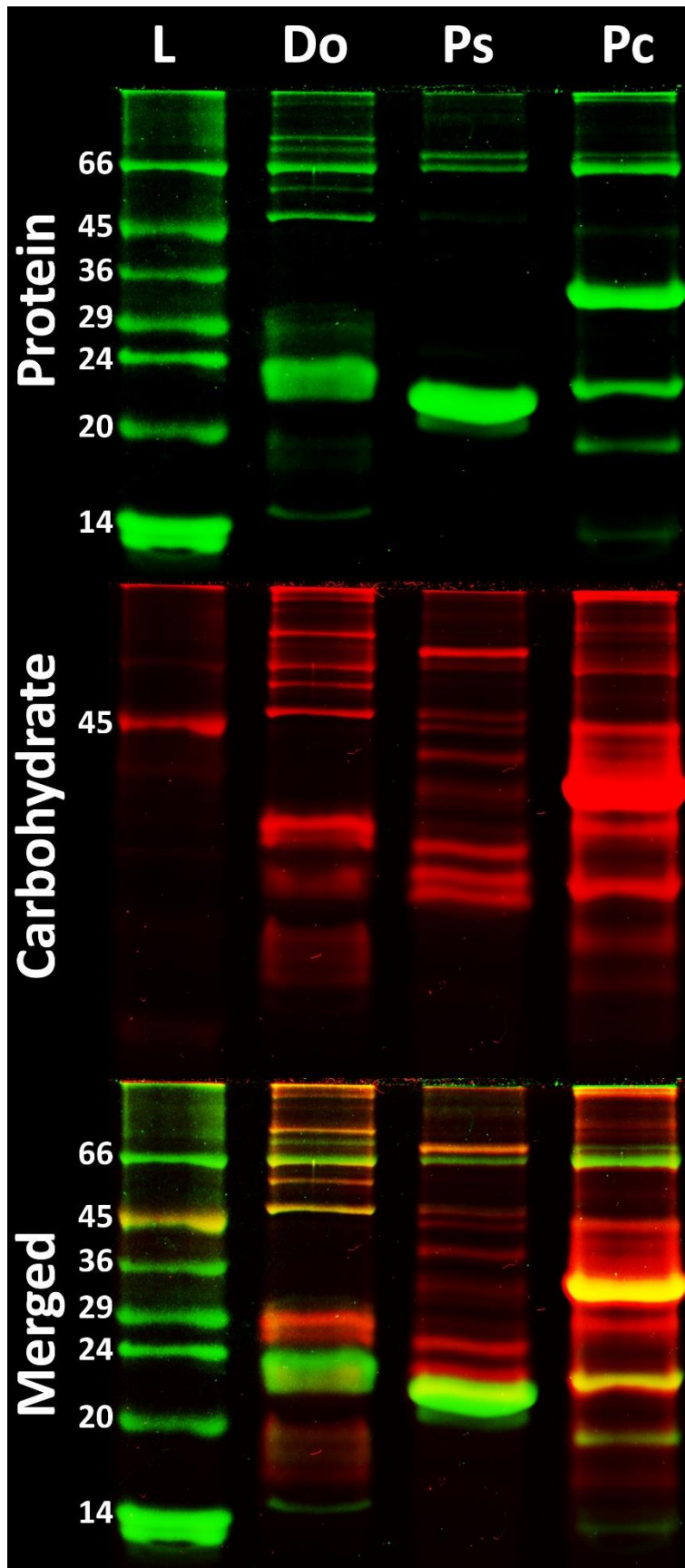
Preliminary data suggested that PRF was the only highly glycosylated protein in the *P. cinereus* pheromone extract. Consequently, individual *P. cinereus* PRF isoforms were purified by RP-HPLC to >99% purity, lyophilized, and prepared for deglycosylation using the Protein Deglycosylation Kit (New England Biolabs): briefly, the lyophilized samples were resolubilized in 12  $\mu$ L 1X Denaturing Buffer (0.5% SDS, 40 mM DTT), the samples each equally divided into two tubes, incubated at 95°C for 10 minutes, then cooled to 4°C for 2 minutes, and diluted with 1.5  $\mu$ L G7 Buffer (50 mM sodium phosphate, pH 7.5), 1.5  $\mu$ L NP-40 (1%), and 3  $\mu$ L nuclease-free water. Half of each PRF isoform was treated with either 1.5  $\mu$ L deglycosylation enzyme mix or 1.5  $\mu$ L nuclease-free water (negative control), and incubated at 37°C overnight. The samples were finally separated by SDS-PAGE, stained for both carbohydrate and protein components, and the bands excised for LC/MS-MS analysis.

## Results

### Separation and purification of *P. cinereus* pheromones

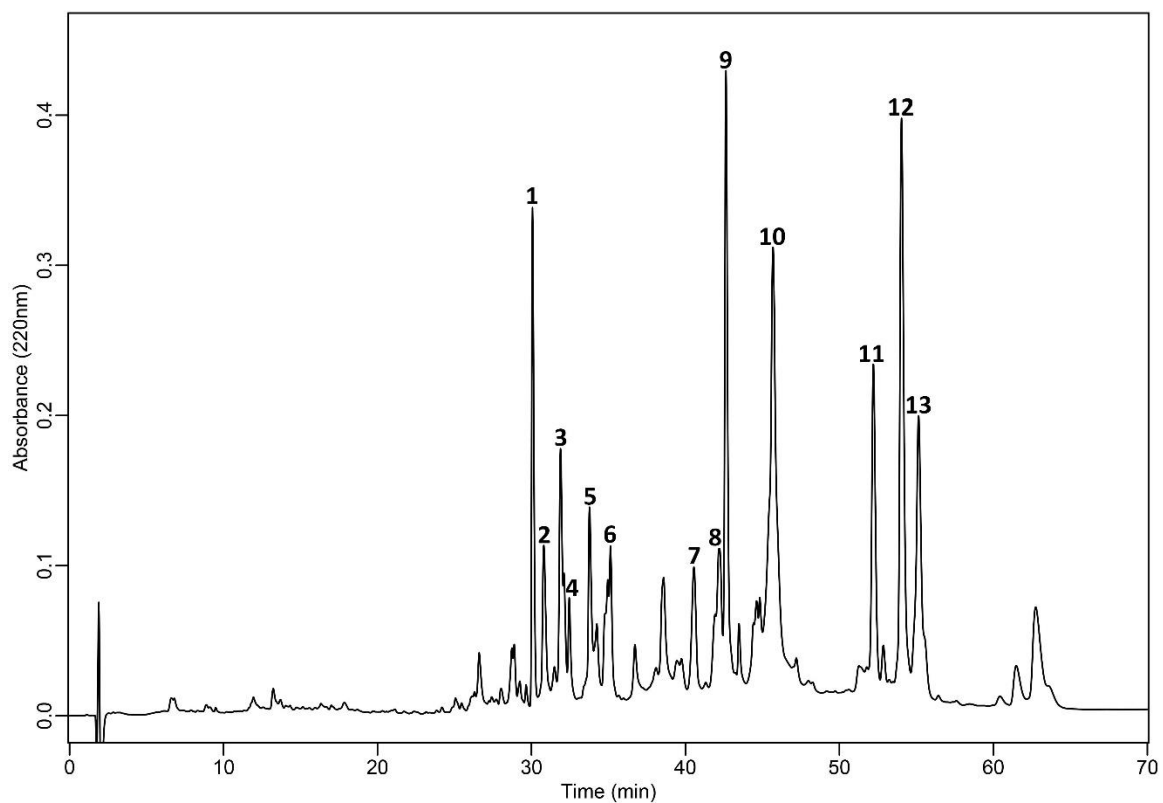
Separation of the *P. cinereus* pheromone extract by SDS-PAGE reveals 6 intense bands (in addition to multiple lighter bands); compared with *P. shermani* and *D. ocoee* extracts, three of the protein bands corresponded to approximately the same molecular weights as PRF, SPF and PMF (Figure 41). Separation by RP-HPLC revealed greater diversity with more than 20 detectable peaks, suggesting multiple isoforms of some pheromone components and/or similarity in molecular weight between components. Thirteen individual RP-HPLC peaks were hand-collected (Figure 42) for further analysis by SDS-PAGE (Figure 43). The estimated masses of these proteins ranged from 7.2 – 68.1 kDa (Table 13). LC/MS-MS was performed on tryptic digests of gel bands to ascertain the identity of these sequences (summarized in Table 13). Additionally, on SDS-PAGE gels, RP-HPLC peaks 6 and 10 (Figure 42) split into two bands; these bands were individually excised for mass spectral characterization (with the higher molecular weight (H) bands termed 6H and 10H, and the lower molecular weight (L) bands termed 6L and 10L. Because of the limited availability of *P. cinereus* cDNA sequences deposited in Genbank, an unannotated draft mental gland transcriptome was initially used for comparison to the mass spectral data to identify likely gene candidates, and these matches were confirmed by targeted RT-PCR.





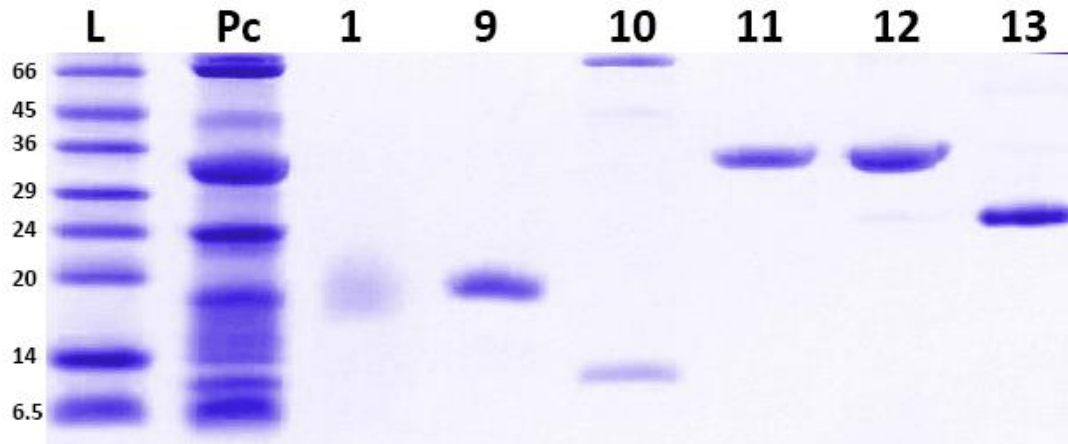
**Figure 41. SDS-PAGE separation of mental gland pheromone extracts between species.**

Fifteen micrograms of whole pheromone extract from *D. ocoee* (Do), *P. shermani* (Ps), and *P. cinereus* (Pc) were separated using a 15% Tris-Tricine gel, stained for carbohydrates using ProQ-488 reactive dye following periodate oxidation, and then Coomassie-stained to visualize protein composition. The images were pseudocolored and overlaid to identify protein versus glycoprotein components.



**Figure 42. RP-HPLC separation of *P. cinereus* pheromone extract.**

A 190  $\mu\text{g}$  aliquot of *P. cinereus* pheromone extract was separated using a 0-70% ACN linear gradient, and individually numbered peaks hand collected for subsequent mass spectral characterization.



**Figure 43. SDS-PAGE analysis of individual RP-HPLC fractions.**

A representative SDS-PAGE gel comparing individual RP-HPLC fractions (marked by number, see Table 1) to the whole *P. cinereus* pheromone (Pc) and size standard (L). Notably, the three PRF fractions (11-13) show molecular weights larger than the ~22 kDa predicted from DNA sequences.

**Table 13. Summary of mass spectral data for RP-HPLC purified *P. cinereus* proteins.**

Peak Label	Retention Time (min)	Mass Estimate by Gel (kDa)	Expected Mass (Da)	Sequence Match	Accession No.	% Sequence Coverage
1	30.1	16.0	7354	PMF	DQ882525	84
2	30.9	7.4	7912	PMF	KJ187100	66
3	32.0	7.9	8040	PMF	KJ187100	87
4	32.5	7.9	7912	PMF	KJ187100	81
5	33.9	7.2	7526	PMF	KJ187093	100
6H	35.2	21.1				
6L	35.2	7.0	8071	PMF	KJ187097	91
7	40.6	25.0	23975	CRISP	KJ187086	93
8	42.3	12.0	14259	Poliovirus Receptor-related Protein	KJ187105	29
9	42.7	19.2	19119	C3	KJ187079	95
10H	45.7	68.1	68210	Albumin	JQ866994	72
10L	45.7	9.8	9770	NPP	KJ187100	91
11	52.2	37.3	23790	PRF A5	AY499385	94
12	54.0	36.9	23846	PRF A2	AY499382	86
13	55.1	25.8	23900	PRF B1	AY499399	84

### Mass spectral identification of *P. cinereus* pheromones

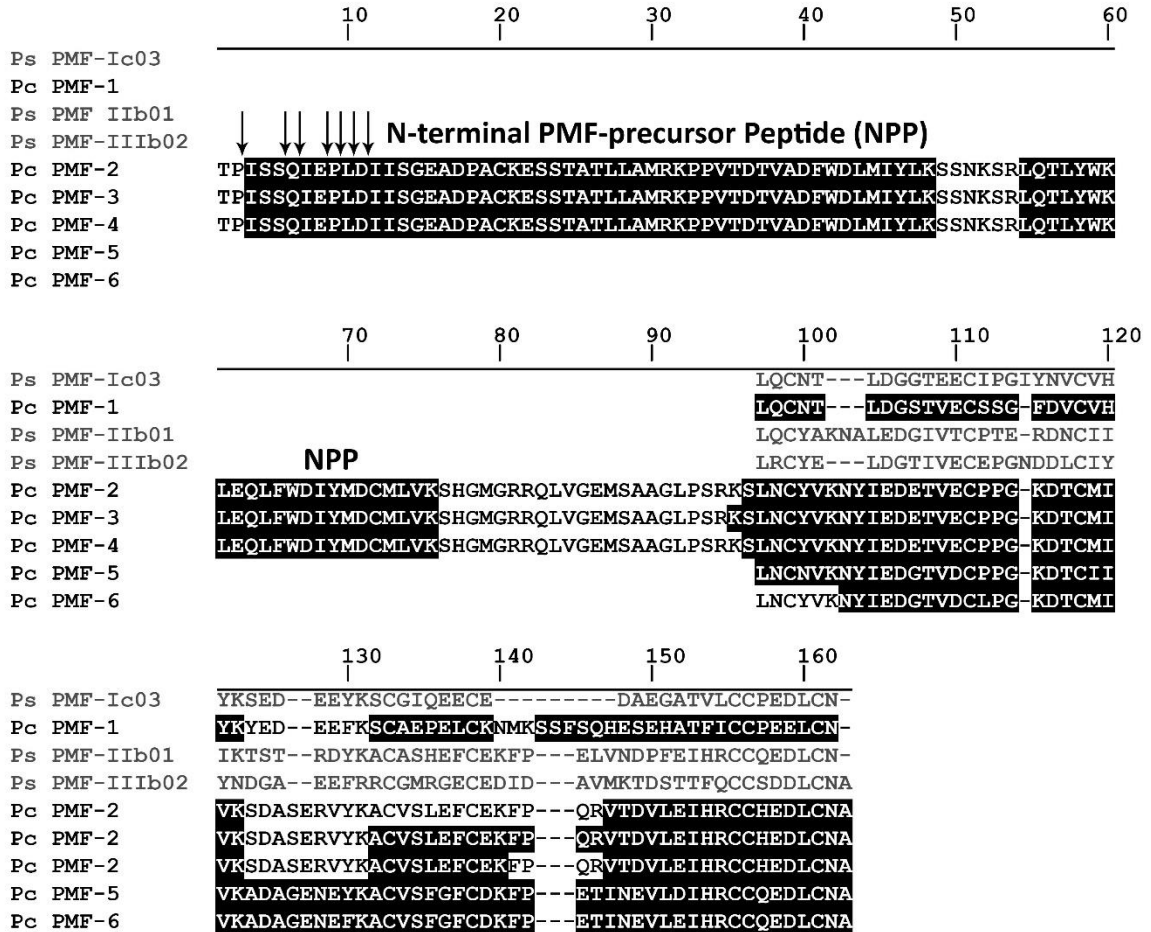
All proteins in fractions 1 – 5 and 6L were identified to be different isoforms of PMF (Table 13; herein referred to as PMF-1 through PMF-6 respectively). All data suggested that these individual proteins are monomeric, similar to *P. shermani* and *D. ocoee* PMFs. PMF-1 matched a *P. cinereus* PMF identified by Palmer et al. [96]; however, PMF-2 through 6 were unique in that they were not detected by the primers in that study, which were based on the conserved UTRs in *P. shermani* Class I [Chapter II]. Comparisons with available *P. shermani* PMF sequences suggest that these additional PMFs are more closely related to PMF Class III (Table 14). The closest sequence match for PMF-2, PMF-3, and PMF-4 was a single unusual PMF sequence that contained an N-terminal extension (more than doubling the molecular weight of typical PMFs) with no significant similarity to any sequences presently in Genbank. The fact that all three isoforms matched a single sequence suggested that these three peaks are likely single nucleotide variants whose corresponding cDNAs were not detected by PCR. Interestingly, the N-terminal peptide matched the protein in peak 10L (Figure 42; herein termed N-terminal PMF-precursor Peptide, NPP) indicating that it can be cleaved off. Based on the mass spectral data, this PMF is likely synthesized as a larger precursor, with NPP and PMF-2,3,4 processed using the dibasic cleavage site, -Arg-Lys-, one residue from the normal N-terminus of PMFs. Dibasic cleavage sites consist of any combination of adjacent Lys or Arg residues and are often used to cleave and process peptide or protein hormones such as insulin, as well as several *D. ocoee* pheromones [88].

**Table 14. Amino acid distance matrix for *P. cinereus* vs. *P. shermani* PMF sequences**

	Ia	Ib	Ic	Id	Ie	If	Ig	Ih	Ii	IIa	IIb	IIIa	IIIb
PMF1	0.55	0.48	0.46	0.50	0.83	0.81	0.96	1.17	1.15	1.33	1.05	1.47	1.28
PMF2,3,4	1.02	1.21	1.24	1.47	1.67	1.49	1.39	1.71	1.51	1.96	1.53	0.60	0.58
PMF5	1.08	1.25	1.26	1.37	1.81	1.56	1.26	1.81	1.54	2.02	1.53	0.82	0.58
PMF6	1.04	1.18	1.29	1.41	1.81	1.48	1.36	1.79	1.48	2.02	1.42	0.82	0.68

However, NPP may undergo additional N-terminal processing: several amino acids beyond the predicted signal peptide cleavage site were not detected and several NPP peptides were identified with sequences missing 2-11 amino acids at the N-terminus (Figure 44). There were no additional trypsin cleavage sites in this region, and we hypothesize these variants arise as part of N-terminal processing in *P. cinereus*. PMF-5 and PMF-6 are very similar in sequence (63 out of 68 identical residues), and also match more closely to PMF Class III; however, different primers against the 3' UTR were required to amplify each (Table 15), suggesting that the UTRs of *P. cinereus* PMFs may not be as conserved as they are in *P. shermani* [Chapter II]. All of the *P. cinereus* PMFs possess a net negative charge, a characteristic of the *P. shermani* PMFs, but the lone *D. ocoee* PMF has a net positive charge.

There were no detectable SPF-like proteins in the *P. cinereus* pheromone extract. This suggests that, similar to *P. shermani*, SPF sequences in *P. cinereus* determined by Palmer et al. [97] reflect the exquisite sensitivity of RT-PCR and SPF mRNAs are likely only present at trace levels. In contrast, three isoforms of PRF (RP-HPLC peaks 11-13; Figure 42) were identified (Table 13; herein referred to as Pc-PRF-1, Pc-PRF-2, and Pc-PRF-3, respectively), and support the hypothesis that PRF is replacing SPF in the *P. cinereus* complex. Mass spectral data matched all three isoforms with sequences previously identified by Palmer et al. (2005), with Pc-PRF-1 matching PRF A5, Pc-PRF-2 matching PRF A2, and Pc-PRF-3 matching PRF B1. These A and B designations correspond to two discrete families of PRF identified by Palmer et al. [95], with the B-family being found exclusively in the *P. cinereus* complex.



**Figure 44. Alignment of identified *P. cinereus* PMF sequences.**

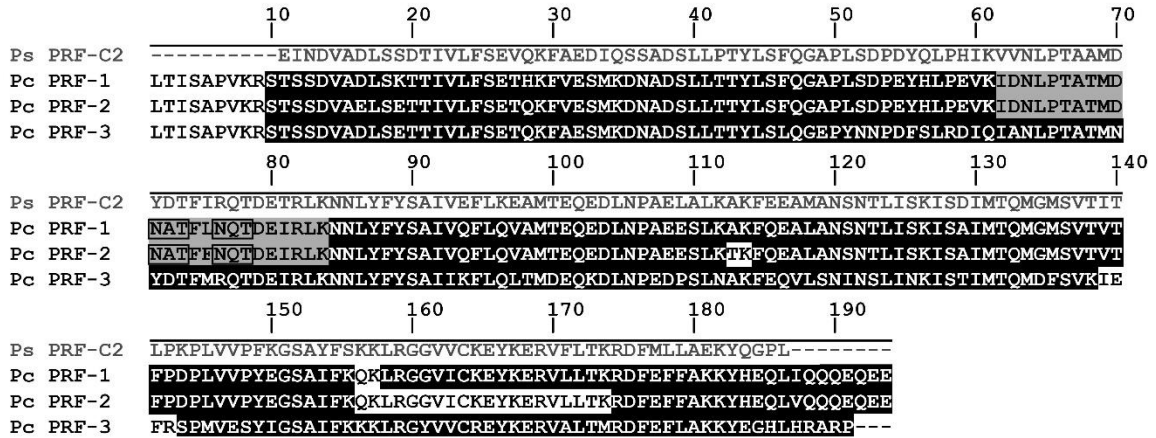
Six PMF isoforms were identified by LC/MS-MS, with black boxes denoting specific peptides identified during the analysis. Three sequences (PMF2-4) have an additional N-terminal extension (termed NPP) that is post-translationally cleaved and elutes at a different location (Peak 10B) by RP-HPLC. LC/MS-MS data support different degrees of N-terminal processing on NPP, with observed cleavage sites marked with arrows. Representative *P. shermani* PMF sequences are included in grey font. The predicted signal peptide has been deleted from all sequences.



**Table 15. Primer list for *P. cinereus* PCR products**

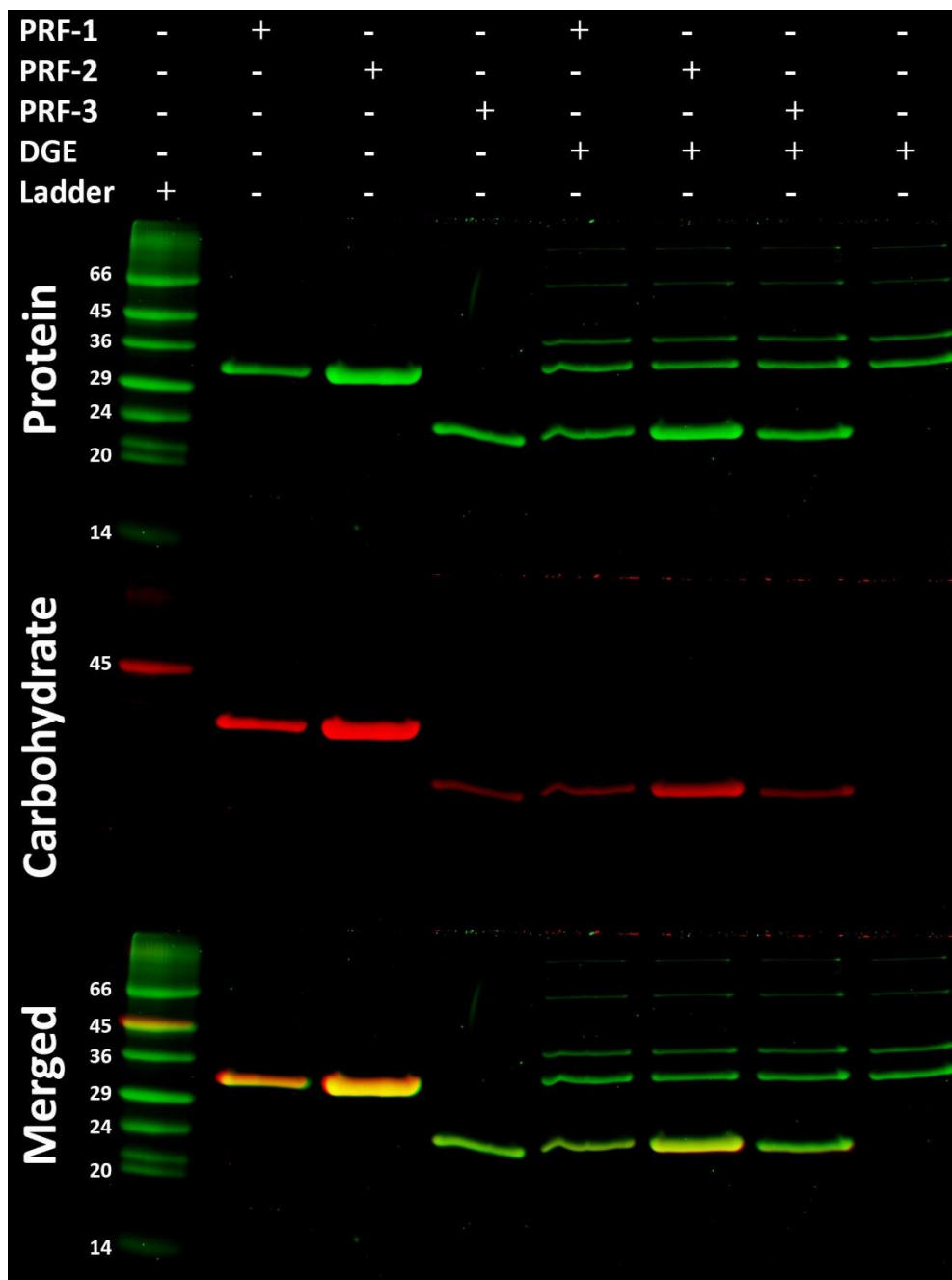
<b>Primer Name</b>	<b>Sequence</b>
PMF with NPP 5' UTR F'	5'-TGCTTGTCTTCTTGGTGGTG-3'
PMF with NPP 3' UTR R'	5'-AGAGGCGTGCCTTCACTTTA-3'
PMF Class III-like 5' UTR F'	5'-GTACATGGGGACAGGGACAG-3'
PMF Class III-like 3' UTR R'	5'-AGAGGCGTGCCTTCACTTTA-3'
C3 5' UTR F'	5'-ACGCAGAGTACATGGGGAAC-3'
C3 3' UTR R'	5'-AATGGGACGAGAGACAATGG-3'
CRISP 5' UTR F'	5'-TTGAATAGCCAACCTGTCCA-3'
CRISP 3' UTR R'	5'-AGCGCATAACATCTGCCTCTT-3'
Poliovirus Receptor-related Protein 5' UTR F'	5'-ACAGGTTGTGACAAGAGGTG-3'
Poliovirus Receptor-related Protein 3' UTR R'	5'-TGTGTCTCGTGTCTGGTTCA-3'

Surprisingly, when these fractions were analyzed by SDS-PAGE, all three bands had estimated molecular weights greater than the 22 kDa predicted by DNA sequencing data and characteristic of the PRFs found in *P. shermani* (Figure 43). Examination of the Pc-PRF-1 and Pc-PRF-2 sequences revealed two N-linked glycosylation motifs (-Asn-Ala-Thr-, NAT, and -Asn-Gln-Thr-, NQT) in close proximity to one another. Additionally, the tryptic peptide that would contain both N-linked glycosylation signals was one of only two peptide masses for Pc-PRF-1, and one of three peptides for Pc-PRF-2 (Figure 45) that were not identified by mass spectral sequencing techniques (LC/MS-MS). Pc-PRF glycosylation was verified by periodate oxidation followed by staining with an aldehyde-reactive fluorescent dye (Figure 46). Additionally, treatment with a commercial blend of deglycosylation enzymes reduced the masses of Pc-PRF-1 and Pc-PRF-2 to that of Pc-PRF-3, yet all three PRFs still stained positive for glycosylation. LC/MS-MS verified that both N-linked glycosylation sites were glycosylated in Pc-PRF-1 and Pc-PRF-2 (indicated by deamination of the NAT and NQT motifs to -Asp-Ala-Thr-, DAT, and -Asp-Gln-Thr-, DQT, following enzymatic deglycosylation). However, this left only a single unmatched tryptic peptide for both Pc-PRF-1 and Pc-PRF-3 that contains Ser (S) and/or Thr (T) as possible O-linked glycosylation sites, i.e., the N-terminal tryptic peptide, LTISAPVK. This sequence is common to all three *P. cinereus* PRF isoforms, and we can deduce that this peptide contains O-linked oligosaccharides on either T2 or S4. Surprisingly, the large molecular weight decrease following removal of the N-linked components is ~12 kDa, or ~6 kDa per N-linked oligosaccharide, which is equivalent to ~30 sugar residues. There are very few examples of N-linked carbohydrate masses this large in vertebrates, and all involve complex tetra-antennary structures [311].



**Figure 45. Alignment of identified *P. cinereus* PRF sequences.**

Three PRF isoforms were identified by mass spectrometry (LC/MS-MS), with black boxes denoting specific peptides identified during the analysis. A grey box denotes peptides identified following enzymatic deglycosylation, with predicted N-linked glycosylation signals outlined in black. A representative *P. shermani* PRF sequence was included in grey font, and the predicted signal peptide has been deleted from all sequences.

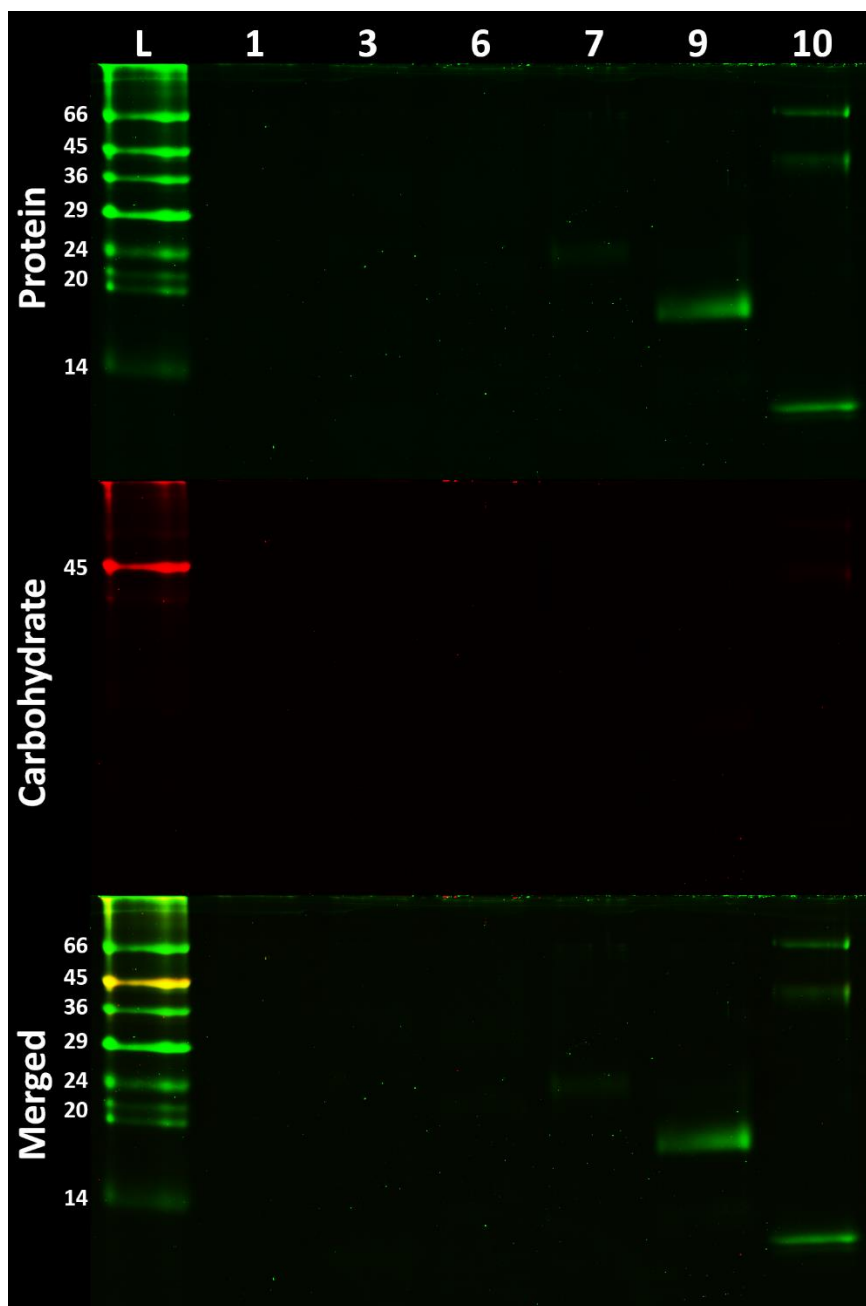


**Figure 46. Deglycosylation analysis of *P. cinereus* PRFs.**

Pc-PRF-1, Pc-PRF-2, and Pc-PRF-3 were separated by SDS-PAGE with and without the addition of a commercially available deglycosylation enzyme mixture (DGE). Following deglycosylation, both Pc-PRF-1 and Pc-PRF-2 show molecular weight shifts to that of Pc-PRF-3, as well as a decrease in signal when stained for carbohydrate. However, even after deglycosylation, all three isoforms still display carbohydrate signals; subsequent mass spectral analysis confirmed that all N-linked oligosaccharides were removed from both Pc-PRF-1 and Pc-PRF-2, suggesting that the remaining carbohydrates are O-linked on T2 and/or S4 in all three isoforms.

The O-linked component(s) is estimated to be ~3 kDa/~15 sugar residues. Additional carbohydrate staining comparing the *D. ocoee*, *P. shermani*, and *P. cinereus* whole pheromone extracts suggests that all three extracts contain some glycoprotein components (Figure 41); however, most of the glycosylated components in the *P. shermani* extract display relatively greater carbohydrate staining over protein staining, suggesting they are minor components in the protein pheromone extract, visible in part due to the sensitivity of the fluorescent carbohydrate stain. Importantly, *P. shermani* PRFs have no post-translational modifications (unpublished data). It is possible that PRF glycosylation may be unique to species utilizing the transdermal pheromone delivery mode and that the covalently bound carbohydrate moieties might influence the structure, metabolic half-life, or mechanism of action. However, due to the limited availability of Pc-PRFs and the inability to maintain reproductively active animals in the laboratory, any studies to further define molecular and biological aspects are currently not feasible.

Additional protein components were identified in the *P. cinereus* pheromone extract that showed no similarities to PRF, PMF, and SPF. Peak 7 shared sequence similarity to the Cysteine-Rich Secretory Protein 1 (CRISP1): a male specific protein that, in mammals, is part of the seminal fluid, binds to the sperm cap, and plays a role in egg:sperm fusion [312]. While CRISP1 is normally glycosylated in mammals, we found no evidence for glycosylation in the *P. cinereus* homolog (Figure 47). Peak 8 matched a sequence with little similarity to any sequence presently in Genbank; however, the closest match was to uncharacterized amphibian proteins that are related to the poliovirus receptor.



**Figure 47. Glycosylation analysis of individual RP-HPLC fractions.**

Select RP-HPLC fractions were separated by SDS-PAGE and analyzed for glycosylation using both carbohydrate and protein staining. No glycosylation was detected on any of the protein bands.

Peak 9 is the *P. cinereus* homolog of the *P. shermani* protein C3 [91], with sequence similarity to the Tissue Inhibitor of Metalloproteinase (TIMP) family (and more recently renamed to Plethodontid TIMP-like Protein, PTP) [Chapter VI]. The function of this protein in *P. shermani* is under active investigation, and may function in mental gland maintenance rather than serve as a pheromone (Wilburn and Feldhoff, unpublished data). Peak 10H was identified to be serum albumin, and likely an artifact of trace blood contamination during the mental gland removal. No matches were identified for peak 6H. None of the identified proteins, including PRF and PMF, have been functionally tested specifically in *P. cinereus* and should only be considered putative pheromones.

## Discussion

The evolution of functional complexes has been of great scientific interest for some time, as interactions between otherwise disparate characters can act as a potent selective tether [313-315]. The predominant hypothesis is that correlational selection drives evolution in such functional complexes, but this results in dramatically different evolutionary modes of stasis or rapid differentiation [128, 129]. The interaction of these two processes is well established in plethodontid salamanders, with multiple studies finding evidence for repeated gene duplication and pervasive positive selection acting on the pheromone genes, yet the larger behavioral and physiological architecture is generally conserved across Plethodontidae [94, 96, 97, Chapter II]. The one exception to this larger conservation is the relatively recent adaptation of olfactory pheromone delivery in the *P. glutinosus* complex (less than 17 MYA of >100 MYA for scratching delivery; Figure 40). Investigations of signaling pathways in *P. shermani* revealed that both PRF and PMF activate neurons in the vomeronasal organ, an accessory olfactory system often associated with pheromone signaling in vertebrates [70, 92]. Both pheromones likely target vomeronasal type-2 receptors, which in rodents bind protein or peptide pheromones [23, 30, 38], and are highly expressed in the *P. shermani* vomeronasal organ [71, 72]. While there is some variation in the anatomy and neuron physiology between vertebrate taxa, the use of protein pheromone binding receptors and the underlying biochemical signaling cascades appear to be conserved in fish, salamanders, frogs, and rodents [5, 39, 41, 42, 72]

The more plesiomorphic character of transdermal pheromone delivery is greatly understudied in comparison to the more recently derived olfactory system. By direct



delivery to the bloodstream, this ancient system bypasses the need for central processing and could directly target different tissues or organs in female salamanders. cDNA sequencing of mental glands from two transdermal species, *D. ocoee* and *Eurycea guttolineata*, revealed sequences for hormone-like peptides, lending plausibility to direct stimulation [89]. Interestingly, the total protein pheromone mass in these glands is 10-15% of that of slapping species, possibly indicating greater potency or more rapid pheromone regeneration, since this system apparently lacks the potential for signal amplification that could occur through central processing in the brain. However, virtually all research to date has centered on the three principal pheromone components observed in *P. shermani* and *D. ocoee*: PRF, PMF, and SPF [87, 316]. While PMF and SPF have an ancient origin and their mRNAs are detectable by RT-PCR in all plethodontid mental glands, both proteomic and molecular analyses reveal stark differences in isoform variability and relative expression as species transition from transdermal to olfactory delivery. For example, *D. ocoee* possesses a single PMF isoform that comprises < 5% of the total pheromone, versus *P. shermani* which has >30 PMF isoforms that constitute ~50% of the total pheromone. Interestingly, *P. cinereus* appears to be intermediate between these two examples: 6 isoforms of PMF comprising 15-20% of the total protein. These data suggest that sexual selection promoted gene duplications of PMF prior to the transition from transdermal to olfactory delivery, yet the process was further exacerbated beyond the transition. However, even from preliminary transcriptome sequencing, there is no evidence for the extended PMFs with the NPP extension in *P. shermani* (Wilburn and Feldhoff, unpublished data), suggesting this may be an adaptation exclusive to the *P. cinereus* lineage.

Despite proteomically characterizing >70% of the total pheromone extract and nearly all of the major peaks, there was no evidence for any SPF proteins in the *P. cinereus* pheromone extract, further supporting the hypothesis that PRF may be functionally replacing SPF in *Plethodon spp.* With trace levels of SPF mRNA and no measureable amounts of SPF protein present in *P. shermani* and *P. cinereus*, which exist on distant ends of the *Plethodon* phylogeny, PRF is thus expected to be highly expressed in the mental glands of both intermediate groups, *P. wehrlei* and *P. welleri*. The discovery of glycosylation on all of the *P. cinereus* PRFs is an interesting phenomenon, and may be important for biological activity when PRF is delivered transdermally. Glycosylation may function to increase PRF conformational stability [317], solubility in plasma [318], and/or to help mediate specific receptor:ligand interactions [319]. If sexual conflict is present, glycosylation may additionally shield PRF from recognition by the female immune system [320]. There are no reports to date of glycosylation being essential for pheromone activity; however, this is not surprising given the limited number of characterized protein pheromones. In all PRF sequences for 28 members of the *P. glutinosus* group (which includes *P. shermani*), the N-linked glycosylation motifs are mutated and non-functional while the 9 amino acid N-terminal peptide extension found in Pc-PRFs, which contains the O-linked glycosylation site(s), has been deleted. Examination of available PRF sequences from the two intermediate groups reveal that *P. welleri* PRFs mirror those of *P. glutinosus* and lack the observed glycosylation sites, while the O-linked N-terminal extension and the two N-linked motifs are preserved in *P. wehrlei*, similar to that of *P. cinereus*. There is a great deal of ambiguity with respect to the precise mode of pheromone delivery in these intermediate groups; both Organ [306]

and Arnold [77] reported that male *P. welleri* rub their mental glands across the female's body, similar to the scratching behavior, while Picard [307] and Dyal [80] additionally observed slapping behavior during tail straddling walk in other members of this group. There are no ethological data available for *P. wehrlei*, as animals would not court under laboratory settings (Arnold SJ, personal communication); thus, it is still unclear whether *P. wehrlei* uses scratching or olfactory delivery. It is an interesting correlation that, similar to *P. shermani* and additional *P. glutinosus* group members, the *P. welleri* group at least partially use slapping delivery in order to provide pheromone to the VNO of females, and those PRF sequences lack molecular signatures of glycosylation. Importantly, when *P. shermani* whole pheromone was applied to the dorsum of females, there was no effect on female behavior, and it was proposed that the *P. glutinosus* complex had "lost" the necessary receptors and/or signaling pathways utilized in transdermal delivery [216]. However, an alternative possibility should be considered is that since PRF is one of the major signaling molecules in *Plethodon spp.*, the glycosylation of PRF in transitional scratching species may alter the site and/or pathway by which PRF enhances mating receptivity in females. This hypothesis is consistent with species using exclusively scratching (*P. cinereus*) or slapping (*P. glutinosus*) delivery, and while it may be applied to both locations in the intermediate delivery species, it may be functionally targeting different tissues in species with glycosylation (bloodstream in *P. wehrlei*, where slapping behavior is unknown) versus those without glycosylation (vomeronasal organ in *P. welleri*, which uses slapping behavior). Based on current methods, *P. cinereus* do not court frequently enough under laboratory conditions to permit behavioral assays to test if PRF glycosylation is directly involved in signaling

(Arnold SJ, personal communication); future neurophysiological and biochemical studies will likely be necessary to ascertain the function of these unique covalently bound oligosaccharides. Additionally, older studies on plethodontid systematics reported that *P. wehrlei* is the sister clade to *P. glutinosus*, however, the most recent molecular data support concurrent speciation between these lineages [308]. Our findings suggest that, at least by parsimony with respect to PRF glycosylation, *P. welleri* may be the sister clade to *P. glutinosus*; however, if the older models are correct, then this loss of glycosylation and putative transition in PRF signaling may have independently evolved multiple times in *Plethodon spp.*

While great emphasis was placed on examining pheromone families that are functionally active in related plethodontid species, it is also relevant that our proteomic analyses revealed new genes which have been co-opted for likely pheromone activity in *P. cinereus*. The NPP observed on *P. cinereus* PMF 2-4 is particularly different, as this specific type of PMF variant has not previously been characterized in other species, and represents a large fraction of the *P. cinereus* PMFs. The limited genomic sequence data for *P. shermani* PMFs suggests that PMF has a gene structure similar to other three-finger proteins, with the first intron found near the end of the signal peptide but before its cleavage site [Chapter II]. It is plausible that this NPP sequence may have been inserted into this first intron, through gene insertion and/or genomic rearrangement with compensatory mutations. Unfortunately, this sequence shares little-to-no similarity with any characterized proteins presently in Genbank, but its maintenance in the pheromone extract suggests a key function. The co-option of CRISP as a putative pheromone is also of particular interest, due in part to its role in mammalian reproduction, but also its

potential evolutionary origins in other exocrine glands that include snake venom glands. In contrast to more classical models of gene co-option with gene duplication and functional diversification, Fry [115] proposed that, in snake venom glands, both CRISP- and kallikrein-like toxins may have been recruited directly from genes already expressed in an ancestral salivary gland-like tissue that eventually developed into modern toxin glands. In plethodontid salamanders, the mental gland has long been proposed to be a modified mucous gland based on its location and histological staining [247]; however, our biochemical analyses of the glandular extracts from multiple species do not support this hypothesis. There have been no detectable proteoglycans in any of our model species, and in this study, we even determined that there are relatively few glycoproteins in the mental glands of the three principal model species. Recently, fluorescence confocal microscopy provided evidence that the mental gland lacks a large lumen for storage; instead the gland has columnar cells with basal nuclei and long cytoplasmic regions that include pheromone packaged in secretory granules or vesicles – much more similar to a salivary or other type of serous gland [Chapter VI]. Interestingly, in a single plethodontid species with atypical pheromone delivery (*D. wrighti*), the mental gland secretes into the male's mouth and he must bite females to deliver pheromone transdermally [45]. A salivary gland relationship may provide an alternative mechanism for the diverse types of proteins observed in the mental gland beyond traditional gene co-option. This would also suggest a potential common origin between salamander pheromone and snake venom glands, which have been shown to utilize many related gene families. The ontogeny of the *P. shermani* mental gland and the molecular mechanisms that regulate its development are presently being investigated.

In conclusion, plethodontid salamanders and their specialized system of courtship provide a valuable model for studying the interplay between different evolutionary forces on functional complexes. In this study, we extensively characterized the mental gland proteome of *Plethodon cinereus*, one of the few *Plethodon* species to exclusively employ the ancient or plesiomorphic behavior of transdermal pheromone delivery. Our data support the hypothesis that PRF has functionally replaced SPF in the *Plethodon* genus; however, the mechanism and site of action for PRF may be dependent on extensive N- and O-linked glycosylation observed in *P. cinereus* PRFs. This post-translational modification is absent in species that exclusively employ olfactory delivery, and may provide phylogenetically-relevant information for species with intermediate forms of delivery. The hypervariable pheromone Plethodontid Modulating Factor (PMF) displays intermediate characteristics with respect to isoform number and biochemical characteristics of PMFs in closely and distantly related salamanders (*P. shermani* and *D. ocoee*). Additional potential pheromone genes were identified in *P. cinereus*, such as CRISP, that may provide key ontological evidence as to the identity and origin of this unique pheromone gland that has been retained for more than 100 million years in plethodontid salamanders. All of these data further support the underlying hypothesis that, despite relative conservation in most higher-order aspects of this functional complex, pervasive sexual selection has acted at the molecular level and led to rapid diversification of the pheromone molecules and mental gland composition, independent of the mechanism by which pheromones are delivered.

## CHAPTER VIII - CONCLUSIONS AND FUTURE DIRECTIONS

For more than 100 million years, plethodontid salamanders have utilized proteinaceous courtship pheromones to modulate the timing of courtship and influence reproductive success. These proteinaceous pheromones are rapidly evolving, presumably in response to female sexual selection, and persist as multi-isoform mixtures through a birth-and-death model of gene evolution. The most extreme example of this is the pheromone Plethodontid Modulating Factor (PMF), a hypervariable protein related to the three-finger protein (TFP) superfamily. The number of PMF isoforms varies between species, with only 1 in *Desmognathes ocoee*, approximately 6 in *Plethodon cinereus*, and more than 30 in *P. shermani* [Chapter VII]. The massive expansion of the PMF gene complex partially corresponds to the transition from transdermal to olfactory delivery. Similar to the case of lysin and VERL in abalone sperm-egg interactions [131-134], the interactions of PMF and female receptors via the bloodstream may be concentration limited, thereby restricting the number of PMF isoforms that a single male can effectively deliver. By transitioning courtship pheromones to the olfactory system, concentration becomes significantly less important, and there is likely little selective pressure to restrict males from synthesizing an extraordinarily diverse blend of PMF isoforms that are essential for increasing female receptivity [Chapter IV]. This variability is further enhanced by changes in the canonical TFP disulfide bonding pattern to permit greater

backbone flexibility in the most rapidly evolving segments of PMF, particularly the loop of finger 3 [Chapter III].

The explosion of PMF gene duplications that corresponds to a change in delivery mode is also coupled to a change in mental gland morphology. Species with either olfactory or intermediate delivery systems (*P. glutinosus*, *P. welleri*, and *P. wehrlei*) have enlarged, pad-like mental glands (presumably of similar histology, see [44] and Chapter VI) that are presumably necessary to deliver pheromones via slapping. However, in these glands, significant changes in gene expression are required to transition the gland from growth and development to active pheromone synthesis. It was proposed in Chapter II that the highly conserved untranslated regions (UTRs) were products of purifying selection to allow simultaneous, synchronized expression of the many diverse isoforms. In Chapter VI, the RNA binding protein CIRBP was identified as a likely regulator of PMF translation and/or mRNA stability, and cooperatively binds to these highly conserved UTRs. While the role of CIRBP has not been examined in other species, PMF sequences in the *P. cinereus* transcriptome had shorter, unique 3' UTRs (and is partially why the Class II and III-like PMFs were missed in the PCR-based screen by Palmer et al. [96]). However, at least some of the *P. shermani* PMFs (such as PMF-A1) are highly expressed without having the conserved Class I PMF UTRs. The specificity of CIRBP is likely dictated by a common dinucleotide and possibly RNA secondary structure such that absolute sequence conservation may not be required for a sequence to still be a target. However, because of its cooperative binding mechanics, 3' UTR length is likely a critical factor to promote association of CIRBP towards some PMFs and not others; in particular, Class III PMFs with their alternative poly(A) cleavage site (Table 2). Of note



is PMF-A1 (Class IIIb), which generally had the longer 3' UTR variant compared to Class IIIa sequences [Chapter II].

Based on all of these data, the following hypothesis is proposed. PMF gene duplication and rapid evolution have likely been the product of sexual selection linked to female receptors. However, through the birth-and-death model of gene evolution, male salamanders have managed to maximize their probability of successfully stimulating females through highly diverse and flexible proteins which can bind the receptors of any female in the breeding population. This form of molecular exploitation could only arise after the transition to olfactory delivery using a larger mental gland, where signaling by direct application only requires trace amounts of pheromone. However, transition to this type of structure required a different physical architecture and a new program of gene regulation, such that there are two discrete phases of mental gland development, with pheromone translation in part regulated by CIRBP. All of these adaptations culminate in a highly effective pheromone system for males to maximize their chance of mating successfully during a brief courtship season.

Future directions of this project would likely entail further defining the mechanisms underlying female neurophysiology and PMF gene regulation. The present lack of molecular data for the female VNE creates a tremendous bottleneck towards determining how individual isoforms may be functioning synergistically to activate neurons. Using a combined approach of laser capture microdissection and next-generation sequencing, efforts are currently underway to build comprehensive transcriptomes of both the MOE and VNE, which should include all V2R (and possibly V1R) sequences. The original approach used to characterize the murine V2R receptors

for ESP1 and MHC peptides involved co-labeling VNE neurons by *in situ* hybridization for activation (using immediate-early response genes) and some set of V2Rs, then systematically increasing probe specificity in order to identify a single V2R [23, 35]. While feasible given a complete library of receptor sequences, this is a highly laborious, low throughput strategy. And, because V2Rs do not readily translocate to the plasma membrane [37], transfection-based approaches in cell culture have been difficult to optimize. The recent structural characterization of ESP1 and its binding to V2Rp5 through the extracellular N-terminal domain [38] may provide new opportunities: if the V2R N-terminal regions function as binding domains (as has been hypothesized since their original discovery [30]), higher throughput may be achieved using cDNA or phage display systems. Briefly, affinity-tagged pheromones could be used to selectively pulldown displayed N-terminal domains, and then the associated RNA sequenced in order to determine identity. Identification of likely pheromone receptors would permit many additional research options, including biochemical and structural studies to determine complex formation, population studies to determine allele frequencies and correlations to male pheromone levels, and phylogenetic studies in order to trace potential patterns of co-evolution between pheromone-receptor pairs.

The presented work on characterizing CIRBP as a potential regulator of PMF translation serves as a critical first step to understanding the mechanisms that control PMF expression. However, many questions remain unanswered, particularly how CIRBP specifically recognizes the PMF 3' UTR. Initial efforts to determine the dinucleotide specificity of the CIRBP-RRM using *in silico* modeling were unsuccessful, in part due to the current difficulties in simulating RNA-protein interactions. Consequently, higher

resolution biochemical assays will likely be necessary using a full panel of substrates and comparing relative affinities under different circumstances. As previously alluded, this would allow for highly interesting studies comparing the UTRs of different PMF isoforms both within and between species to determine if CIRBP has indeed played a key role in regulating expression of different PMFs via selective 3' UTR evolution. A high resolution NMR structure ( $< 1 \text{ \AA}$  rmsd) has already been solved for the human CIRBP RNA recognition motif (PDB #1X5S), and *P. shermani* CIRBP shares 82% identity (69/84 identical residues, with 11/15 conservative substitutions). However, it will be critical to solve the structure in complex with RNA oligonucleotides to determine the molecular interactions that provide specificity. The precise mechanisms of the CIRBP low complexity domain will likely be much more complicated, as its tendency to aggregate and precipitate would make it extraordinarily difficult to crystallize or maintain at sufficiently high concentrations for solution NMR. Establishing a biologically relevant model using *in vitro* and/or cell culture systems will be critical for fully characterizing CIRBP's function in PMF gene expression and mental gland development. One option may include induction of mental gland development in female *P. shermani* to serve as a "zero background" model where there is no potential interference from endogenous male hormones. Sever [222] managed to induce gland development in female *Eurycea quadridigitata*; however, this was using  $\sim 100,000 \times$  physiological levels of testosterone. In collaboration with Dr. Sarah Woodley (Duquesne Univ.), we have recently attempted to replicate Sever's original work in *P. shermani* using small implants in both gravid and non-gravid females that provide constant release of testosterone at physiological levels. There were no visible mental glands in gravid females with testosterone implants; in non-

gravid females (with naturally lower estradiol levels), there was evidence for initial formation of a mental gland, although much more slowly than natural male progression. Hormone assays confirmed that testosterone levels were at male physiological levels, however, estradiol was elevated above females with control implants, suggesting likely endocrine interference. An inducible model of mental gland development would allow much higher resolution measurements of changes in gene expression, and permit control over other external cues that may be essential for regulating progression and timing (e.g. temperature, humidity, and photoperiod).

These proposed mechanistic studies likely represent the most tractable applications of the *P. shermani* system to biomedical science. With the exception of the work with mice, there are presently no well characterized models for vomeronasal signal transduction; as basal tetrapods, salamanders provide an excellent comparative model to determining the mechanisms universal to vertebrate olfaction and signaling. As was discussed in Chapter VI, there are many similarities between mental gland and tumor development – and while studies of the mental gland are currently in their infancy, understanding the mechanisms of gland maturation and eventual regression may unlock new therapeutic targets for cancer treatment. However, as animals that thrive in cool, damp habitats, plethodontid salamanders are likely one of the first groups to be impacted by global climate change. Consequently, studying the molecular mechanisms of mating and reproduction may have more profound impacts on preserving biodiversity and species survival.

## REFERENCES

1. Postgate JR: **Microsenses**. In: *The outer reaches of life*. Cambridge University Press; 1995: 165.
2. Johansson RS, Flanagan JR: **Coding and use of tactile signals from the fingertips in object manipulation tasks**. *Nat Rev Neurosci* 2009, **10**(5):345-359.
3. Nassi JJ, Callaway EM: **Parallel processing strategies of the primate visual system**. *Nat Rev Neurosci* 2009, **10**(5):360-372.
4. Kaupp UB: **Olfactory signalling in vertebrates and insects: differences and commonalities**. *Nat Rev Neurosci* 2010, **11**(3):188-200.
5. Dulac C, Torello AT: **Molecular detection of pheromone signals in mammals: from genes to behaviour**. *Nat Rev Neurosci* 2003, **4**(7):551-562.
6. Ghysen A, Dambly-Chaudière C: **The lateral line microcosmos**. *Genes & Development* 2007, **21**(17):2118-2130.
7. Hart V, Novakova P, Malkemper E, Begall S, Hanzal V, Jezek M, Kusta T, Nemcova V, Adamkova J, Benediktova K *et al*: **Dogs are sensitive to small variations of the Earth's magnetic field**. *Frontiers in Zoology* 2013, **10**(1):80.
8. Schepers RJ, Ringkamp M: **Thermoreceptors and thermosensitive afferents**. *Neuroscience and Biobehavioral Reviews* 2010, **34**:177-184.
9. Liti G, Louis EJ: **Yeast evolution and comparative genomics**. *Annual Review of Microbiology* 2005, **59**(1):135-153.
10. Karlson P, Luscher M: **'Pheromones': a new term for a class of biologically active substances**. *Nature* 1959, **183**(4653):55-56.
11. Butenandt A, Beckmann R, Hecker E: **Über den sexuallockstoff des seidenspinners, i. der biologische test und die isolierung des reinen sexuallockstoffes bombykol**. *Hoppe-Seylers Zeitschrift für Physiologische Chemie* 1961, **324**(1):71-83.
12. Liu Z, Xu B, Miao Z, Sun J: **The pheromone frontalin and its dual function in the invasive bark beetle *Dendroctonus valens***. *Chemical Senses* 2013, **38**:485-495.

13. Marcillac F, Ferveur J-F: **A set of female pheromones affects reproduction before, during and after mating in *Drosophila***. *Journal of Experimental Biology* 2004, **207**:3927-3933.
14. Liénard MA, Hagström ÅK, Lassance J-M, Löfstedt C: **Evolution of multicomponent pheromone signals in small ermine moths involves a single fatty-acyl reductase gene**. *Proceedings of the National Academy of Sciences* 2010, **107**(24):10955-10960.
15. Wanner KW, Nichols AS, Allen JE, Bunger PL, Garczynski SF, Jr. CEL, Robertson HM, Luetje CW: **Sex pheromone receptor specificity in the European corn borer moth, *Ostrinia nubilalis***. *PLoS ONE* 2010, **5**(1):e8685.
16. Tillman JA, Seybold SJ, Jurenka RA, Blomquist GJ: **Insect pheromones—an overview of biosynthesis and endocrine regulation**. *Insect Biochemistry and Molecular Biology* 1999, **29**(6):481-514.
17. Law JH, Regnier FE: **Pheromones**. *Annual Review of Biochemistry* 1971, **40**(1):533-548.
18. Zuk M, Kolluru GR: **Exploitation of sexual signals by predators and parasitoids**. *Quarterly Review of Biology* 1998, **73**(4):415-438.
19. Keng-Hong T, Nishida R: **Synomone or kairomone? – *Bulbophyllum apertum* flower releases raspberry ketone to attract *Bactrocera* fruit flies**. *Journal of Chemical Ecology* 2005, **31**(3):497-507.
20. Isogai Y, Si S, Pont-Lezica L, Tan T, Kapoor V, Murthy VN, Dulac C: **Molecular organization of vomeronasal chemoreception**. *Nature* 2011, **478**(7368):241-245.
21. Carde RT, Haynes KF: **Structure of the pheromone communication channel in moths**. In: *Advances in Insect Chemical Ecology*. Edited by Carde RT, Miller JG. Cambridge: Cambridge University; 2004: 283-332.
22. Touhara K, Vosshall LB: **Sensing odorants and pheromones with chemosensory receptors**. *Annual Review of Physiology* 2009, **71**:307-332.
23. Haga S, Hattori T, Sato T, Sato K, Matsuda S, Kobayakawa R, Sakano H, Yoshihara Y, Kikusui T, Touhara K: **The male mouse pheromone ESP1 enhances female sexual receptive behaviour through a specific vomeronasal receptor**. *Nature* 2010, **466**(7302):118-122.
24. Mudge JM, Armstrong SD, McLaren K, Beynon RJ, Hurst JL, Nicholson C, Robertson DH, Wilming LG, Harrow JL: **Dynamic instability of the major urinary protein gene family revealed by genomic and phenotypic comparisons between C57 and I29 strain mice**. *Genome Biology* 2008, **9**(5):R91.

25. Roberts S, Simpson D, Armstrong S, Davidson A, Robertson D, McLean L, Beynon R, Hurst J: **Darcin: a male pheromone that stimulates female memory and sexual attraction to an individual male's odour.** *BMC Biology* 2010, **8**(1):75.
26. Roberts SA, Davidson AJ, McLean L, Beynon RJ, Hurst JL: **Pheromonal induction of spatial learning in mice.** *Science* 2012, **338**(6113):1462-1465.
27. Chamero P, Marton TF, Logan DW, Flanagan K, Cruz JR, Saghatelian A, Cravatt BF, Stowers L: **Identification of protein pheromones that promote aggressive behaviour.** *Nature* 2007, **450**(7171):899-902.
28. Buck L, Axel R: **A novel multigene family may encode odorant receptors: A molecular basis for odor recognition.** *Cell* 1991, **65**(1):175-187.
29. Dulac C, Axel R: **A novel family of genes encoding putative pheromone receptors in mammals.** *Cell* 1995, **83**(2):195-206.
30. Herrada G, Dulac C: **A novel family of putative pheromone receptors in mammals with a topographically organized and sexually dimorphic distribution.** *Cell* 1997, **90**(4):763-773.
31. Baum MJ, Kelliher KR: **Complementary Roles of the Main and Accessory Olfactory Systems in Mammalian Mate Recognition.** *Annual Review of Physiology* 2009, **71**(1):141-160.
32. Spehr M, Hatt H, Wetzel CH: **Arachidonic acid plays a role in rat vomeronasal signal transduction.** *Journal of Neuroscience* 2002, **22**(19):8429-8437.
33. Stowers L, Holy TE, Meister M, Dulac C, Koentges G: **Loss of sex discrimination and male-male aggression in mice deficient for TRP2.** *Science* 2002, **295**(5559):1493-1500.
34. Boschhat C, Pélofi C, Randin O, Roppolo D, Lüscher C, Broillet M-C, Rodriguez I: **Pheromone detection mediated by a V1r vomeronasal receptor.** *Nat Neurosci* 2002, **5**:1261-1262.
35. Leinders-Zufall T, Ishii T, Mombaerts P, Zufall F, Boehm T: **Structural requirements for the activation of vomeronasal sensory neurons by MHC peptides.** *Nat Neurosci* 2009, **12**(12):1551-1558.
36. Martini S, Silvotti L, Shirazi A, Ryba NJP, Tirindelli R: **Co-expression of putative pheromone receptors in the sensory neurons of the vomeronasal organ.** *The Journal of Neuroscience* 2001, **21**(3):843-848.

37. Loconto J, Papes F, Chang E, Stowers L, Jones EP, Takada T, Kumánovics A, Lindahl KF, Dulac C: **Functional expression of murine V2R pheromone receptors involves selective association with the M10 and M1 Families of MHC class Ib molecules.** *Cell* 2003, **112**(5):607-618.
38. Yoshinaga S, Sato T, Hirakane M, Esaki K, Hamaguchi T, Haga-Yamanaka S, Tsunoda M, Kimoto H, Shimada I, Touhara K *et al*: **Structure of the mouse sex peptide pheromone ESP1 reveals a molecular basis for specific binding to the class C G-protein-coupled vomeronasal receptor.** *Journal of Biological Chemistry* 2013, **288**(22):16064-16072.
39. Pfister P, Randall J, Montoya-Burgos JI, Rodriguez I: **Divergent evolution among teleost V1r receptor genes.** *PLoS ONE* 2007, **2**(4):e379.
40. González A, Morona R, López JM, Moreno N, Northcutt RG: **Lungfishes, like tetrapods, possess a vomeronasal system.** *Frontiers in Neuroanatomy* 2010, **4**:130.
41. Hagino-Yamagishi K, Moriya K, Kubo H, Wakabayashi Y, Isobe N, Saito S, Ichikawa M, Yazaki K: **Expression of vomeronasal receptor genes in *Xenopus laevis*.** *Journal of Comparative Neurology* 2004, **472**(2):246-256.
42. Date-Ito A, Ohara H, Ichikawa M, Mori Y, Hagino-Yamagishi K: ***Xenopus* V1R vomeronasal receptor family is expressed in the main olfactory system.** *Chemical Senses* 2008, **33**(4):339-346.
43. Rodriguez I, Greer CA, Mok MY, Mombaerts P: **A putative pheromone receptor gene expressed in human olfactory mucosa.** *Nat Genet* 2000, **26**(1):18-19.
44. Houck LD, Reagan NL: **Male courtship pheromones increase female receptivity in a Plethodontid salamander.** *Animal Behaviour* 1990, **39**:729-734.
45. Houck LD, Sever DM: **The Role of the Skin in Reproduction and Behavior.** In: *Amphibian Biology*. Edited by Heatwole H, Barthalamus G, vol. 1. Chipping Norton, Australia: Surrey Beatty and Sons; 1994: 351-381.
46. Kikuyama S, Toyoda F, Ohmiya Y, Matsuda K, Tanaka S, Hayashi H: **Sodefrin: a female-attracting peptide pheromone in newt cloacal glands.** *Science* 1995, **267**(5204):1643-1645.
47. Larson A, Weisrock DW, Kozak KH: **Phylogenetic systematics of salamanders (Amphibia: Urodela), a review.** In: *Reproductive biology and phylogeny of Urodela (Amphibia)*. Edited by Sever DM. Enfield, NH: Science Publishers; 2003: 31-108.



48. Weisrock DW: **Molecular phylogenetic studies of speciation and phylogeny in salamanders: macroevolutionary and microevolutionary perspectives.** *PhD Dissertation.* St. Louis, MO: Washington University; 2003.
49. Frost DR, Grant T, Faivovich J, Bain RH, Haas A, Haddad CFB, De Sa RO, Channing A, Wilkinson M, Donnellan SC *et al*: **The amphibian tree of life.** *Bulletin of the American Museum of Natural History* 2006, **1**:1-370.
50. Vieites DR, Min M-S, Wake DB: **Rapid diversification and dispersal during periods of global warming by plethodontid salamanders.** *Proceedings of the National Academy of Sciences* 2007, **104**(50):19903-19907.
51. Smith JJ, Kump DK, Walker JA, Parichy DM, Voss SR: **A comprehensive expressed sequence tag linkage map for tiger salamander and Mexican axolotl: enabling gene mapping and comparative genomics in *Ambystoma*.** *Genetics* 2005, **171**(3):1161-1171.
52. Schreckenberg G, Jacobson A: **Normal stages of development of the axolotl, *Ambystoma mexicanum*.** *Developmental biology* 1975, **42**(2):391-399.
53. Voss SR, Shaffer HB: **Adaptive evolution via a major gene effect: paedomorphosis in the Mexican axolotl.** *Proceedings of the National Academy of Sciences* 1997, **94**(25):14185-14189.
54. Hairston NG: **Community ecology and salamander guilds.** Cambridge: Cambridge University Press; 1987.
55. Lanza B, Vanni S, Nistri A. In: *Encyclopedia of Reptiles and Amphibians.* Edited by Cogger HG, Zweifel RG. San Diego: Academic Press; 1998: 74-75.
56. Dawley EM: **Recognition of individual, sex, and species odours by salamanders of the *Plethodon glutinosus*-*P. jordani* complex.** *Animal Behaviour* 1984, **32**:352-361.
57. Dawley EM: **Evolution of chemical signals as a premating isolating mechanism in a complex of terrestrial salamanders.** In: *Chemical signals in vertebrates 4.* Edited by Duvall D, Muller-Schwarze D, Silverstein RM. New York: Plenum Press; 1986: 221-224.
58. Marco A, Chivers DP, Kiesecker JK, Blaustein AR: **Mate choice by chemical cues in western redback (*Plethodon vehiculum*) and Dunn's (*P. dunnii*) salamanders.** *Journal of Chemical Ecology* 1998, **27**:1333-1344.
59. Danzter BJ, Jaeger RG: **Male redbacked salamanders can determine the reproductive status of conspecific females through volatile chemical signals.** *Herpetologica* 2007, **63**:176-183.

60. Walls SC, Mathis A, Jaeger RG, Gergits WF: **Male salamanders with high quality diets have faeces attractive to females.** *Animal Behavior* 1989, **38**:546-548.
61. Chouinard AJ: **Rapid onset of mate quality assessment via chemical signals in a woodland salamander (*Plethodon cinereus*).** *Behav Ecol Sociobiol* 2012, **66**(5):765-775.
62. Maksimowich DS, Mathis A: **Pheromonal markers as indicators of parasite load: parasite-mediated behavior in salamanders (*Plethodon angusticlavius*).** *Acta Ethologica* 2001, **3**:83-87.
63. Jaeger RG: **Pheromonal markers as territorial advertisement by terrestrial salamanders.** In: *Chemical signals in vertebrates 4*. Edited by Duvall D, Muller-Schwarze D, Silverstein RM. New York: Plenum Press; 1986: 191-203.
64. Quinn VS, Graves BM: **Space use in response to conspecifics by the red-backed salamander (*Plethodon cinereus*, Plethodontidae, Caudata).** *Ethology* 1999, **105**:993-1002.
65. Murray DL, Jenkins CL: **Perceived predation risk as a function of predator dietary cues in terrestrial salamanders.** *Animal Behaviour* 1999, **57**:33-39.
66. Sullivan AM, Picard AL, Madison DM: **To avoid or not to avoid? Factors influencing the discrimination of predator diet cues by a terrestrial salamander.** *Animal Behaviour* 2005, **69**:1425-1433.
67. Dickens JC: **Predator-prey interactions: olfactory adaptations of generalist and specialist predators.** *Agricultural and Forest Entomology* 1999, **1**(1):47-54.
68. Placyk JS, Graves BM: **Prey detection by vomeronasal chemoreception in a plethodontid salamander.** *Journal of Chemical Ecology* 2002, **28**:1017-1036.
69. Feldhoff RC, Rollmann SM, Houck LD: **Chemical analyses of courtship pheromones in a Plethodontid salamander.** In: *Advances in Chemical Signals in Vertebrates*. Edited by Johnston RE, Müller-Schwarze D, Sorensen P. New York: Kluwer Academic/Plenum; 1999: 117-125.
70. Wirsig-Wiechmann CR, Houck LD, Feldhoff PW, Feldhoff RC: **Pheromonal activation of vomeronasal neurons in Plethodontid salamanders.** *Brain Research* 2002, **952**:335-344.
71. Woodley SK: **Sex steroid hormones and sexual dimorphism of chemosensory structures in a terrestrial salamander (*Plethodon shermani*).** *Brain Research* 2007, **1138**(0):95-103.

72. Kiemnec-Tyburczy KM, Woodley SK, Watts RA, Arnold SJ, Houck LD: **Expression of vomeronasal receptors and related signaling molecules in the nasal cavity of a caudate amphibian (*Plethodon shermani*)**. *Chemical Senses* 2012, **37**:335-346.
73. Laberge F, Roth G: **Connectivity and cytoarchitecture of the ventral telencephalon in the salamander *Plethodon shermani***. *The Journal of Comparative Neurology* 2005, **482**:176-200.
74. Laberge F, Mühlenbrock-Lenter S, Grunwald W, Roth G: **Evolution of the amygdala: new insights from studies in amphibians**. *Brain, Behavior and Evolution* 2006, **67**(4):177-187.
75. Butler AB, Hodos W: **Comparative vertebrate neuroanatomy: evolution and adaptation**. New York: Wiley-Liss; 1996.
76. Swanson LW: **Cerebral hemisphere regulation of motivated behavior**. *Brain Research* 2000, **886**:113-164.
77. Arnold SJ: **The evolution of courtship behavior in salamanders**. University of Michigan, Ann Arbor; 1972.
78. Arnold SJ: **Sexual behavior, sexual interference, and sexual deference in salamanders *Ambystoma maculatum*, *Ambystoma tigrinum*, and *Plethodon jordani***. *Z Tierpsychol* 1976, **42**:247-300.
79. Eddy SL, Kiemnec-Tyburczy KM, Uyeda JC, Houck LD: **The influence of sequential male courtship behaviors on courtship success and duration in a terrestrial salamander, *Plethodon shermani***. *Ethology* 2012, **118**(12):1240-1250.
80. Dyal LA: **Novel courtship behaviors in three small eastern *Plethodon* species**. *Journal of Herpetology* 2006, **40**(1):55-65.
81. Houck LD, Mead LS, Watts RA, Arnold SJ, Feldhoff PW, Feldhoff RC: **A candidate vertebrate pheromone, SPF, increases female receptivity in a salamander**. In: *Chemical Signals in Vertebrates 11*. Edited by Hurst J, Beynon R, Muller-Schwarze D. New York: Springer; 2007: 213-221.
82. Stebbins RC: **Courtship of the plethodontid salamander, *Ensatina eschscholtzii***. *Copeia* 1949, **1949**:274-281.
83. Rollmann SM, Houck LD, Feldhoff RC: **Proteinaceous pheromone affecting female receptivity in a terrestrial salamander**. *Science* 1999, **285**:1907-1909.
84. Eddy SL: **Mutual mate choice in a terrestrial salamander, *Plethodon shermani*, with long-term sperm storage**. Corvallis, OR: Oregon State University; 2012.

85. Houck LD, Bell AM, Reagan-Wallin NL, Feldhoff RC: **Effects of experimental delivery of male courtship pheromones on the timing of courtship in a terrestrial salamander, *Plethodon jordani* (Caudata: Plethodontidae).** *Copeia* 1998, **1998**:214-219.
86. Laberge F, Feldhoff RC, Feldhoff PW, Houck LD: **Courtship pheromone-induced c-Fos-like immunolabeling in the female salamander brain.** *Neuroscience* 2008, **151**:329-339.
87. Houck LD, Arnold SJ: **Courtship and mating behavior.** In: *Phylogeny and reproductive biology of Urodela (Amphibia)*. Edited by Sever DM. Enfield, New Hampshire: Science Publishers; 2003: 383-424.
88. Leichty KA: **Co-option and adaptation of novel gene duplications for pheromone activity in a dusky salamander.** Louisville, KY: University of Louisville; 2012.
89. Kiemnec-Tyburczy KM, Watts RA, Gregg RG, von Borstel D, Arnold SJ: **Evolutionary shifts in courtship pheromone composition revealed by EST analysis of plethodontid salamander mental glands.** *Gene* 2009, **432**:75-81.
90. Palmer CA, Hollis DM, Watts RA, Houck LD, McCall MA, Gregg RG, Feldhoff PW, Feldhoff RC, Arnold SJ: **Plethodontid modulating factor, a hypervariable salamander courtship pheromone in the three-finger protein superfamily.** *FEBS Journal* 2007, **274**:2300-2310.
91. Chouinard AJ, Wilburn DB, Houck LD, Feldhoff RC: **Individual Variation in Pheromone Isoform Ratios of the Red-Legged Salamander, *Plethodon shermani*.** In: *Chemical Signals in Vertebrates 12*. Edited by East ML, Dehnhard M. New York: Springer; 2013: 99-115.
92. Wirsig-Wiechmann CR, Houck LD, Wood JM, Feldhoff PW, Feldhoff RC: **Male pheromone protein components activate female vomeronasal neurons in the salamander *Plethodon shermani*.** *BMC Neuroscience* 2006, **7**(26).
93. Houck LD, Palmer CA, Watts RA, Arnold SJ, Feldhoff PW, Feldhoff RC: **A new vertebrate courtship pheromone that affects female receptivity in a terrestrial salamander.** *Animal Behaviour* 2007, **73**:315-320.
94. Watts RA, Palmer CA, Feldhoff RC, Feldhoff PW, Houck LD, Jones AG, Pfrender ME, Rollmann SM, Arnold SJ: **Stabilizing selection on behavior and morphology masks positive selection on the signal in a salamander pheromone signaling complex.** *Molecular Biology & Evolution* 2004, **21**(5):1032-1041.
95. Palmer CA, Watts RA, Gregg RG, McCall MA, Houck LD, Highton R, Arnold SJ: **Lineage-specific differences in evolutionary mode in a salamander courtship pheromone.** *Molecular Biology & Evolution* 2005, **22**(11):2243-2256.

96. Palmer CA, Picard AL, Watts RA, Houck LD, Arnold SJ: **Rapid evolution of Plethodontid modulating factor (PMF), a hypervariable salamander courtship pheromone, is driven by positive selection.** *Journal of Molecular Evolution* 2010, **70**:427-440.
97. Palmer CA, Watts RA, Houck LD, Picard AL, Arnold SJ: **Evolutionary replacement of components in a salamander pheromone signaling complex: more evidence for phenotypic-molecular decoupling.** *Evolution* 2007, **61**(1):202-215.
98. Fryklund L, Eaker D: **Complete covalent structure of a cardiotoxin from the venom of *Naja nigricollis* (African black-necked spitting cobra).** *Biochemistry* 1975, **14**(13):2865-2871.
99. Dufton MJ, Hider RC: **Conformational properties of the neurotoxins and cytotoxins isolated from Elapid snake venoms.** *CRC Critical Reviews in Biochemistry* 1983, **14**:113-171.
100. Cervenansky C, Dajas F, Harvey AL, Karlsson E: **Fasciculins, anti-cholinesterase toxins from mamba venoms: biochemistry and pharmacology.** In: *Snake Toxins*. Edited by Harvey AL. New York: Pergamon Press; 1991: 303-321.
101. Georgaka E, Nastopoulos V, Eleftheriou S, Zarkadis IK: **A toxin-like gene in rainbow trout: cloning, expression, and gene organization.** *Toxicon* 2007, **49**:1002-1009.
102. Kolbe HV, Huber A, Cordier P, Rasmussen UB, Bouchon B, Jaquinod M, Vlasak R, Delot EC, Kreil G: **Xenoxins, a family of peptides from dorsal gland secretion of *Xenopus laevis* related to snake venom cytotoxins and neurotoxins.** In., vol. 268; 1993: 16458-16464.
103. Macleod RJ, Lembessis P, James S, Bennett HPJ: **Isolation of a member of the neurotoxin/cytotoxin peptide family from *Xenopus laevis* skin which activates dihydropyridine-sensitive Ca<sup>2+</sup> channels in mammalian epithelial cells.** *Journal of Biological Chemistry* 1998, **273**(32):20046-20051.
104. Garza-Garcia A, Harris R, Esposito D, Gates PB, Driscoll PC: **Solution structure and phylogenetics of Prod1, a member of the three-finger protein superfamily implicated in salamander limb regeneration.** *PLoS ONE* 2009, **4**(9):e7123.
105. Davies A, Simmons DL, Hale G, Harrison RA, Tighe H, Lachmann PJ, Waldmann H: **CD59, an LY-6-like protein expressed in human lymphoid cells, regulates the action of the complement membrane attack complex on homologous cells.** *The Journal of Experimental Medicine* 1989, **170**(3):637-654.

106. Blasi F, Carmeliet P: **uPAR: a versatile signalling orchestrator**. *Nat Rev Mol Cell Biol* 2002, **3**(12):932-943.
107. Ploug M, Ellis V: **Structure-function relationships in the receptor for urokinase-type plasminogen activator. Comparison to other members of the Ly-6 family and snake venom  $\alpha$ -neurotoxins**. *FEBS Letters* 1994, **349**:163-168.
108. Adermann K, Wattler F, Wattler S, Heine G, Meyer M, Forssmann W-G, Nehls M: **Structural and phylogenetic characterization of human SLURP-1, the first secreted mammalian member of the Ly-6/uPAR protein superfamily**. *Protein Science* 1999, **8**:810-819.
109. Levitin F, Weiss M, Hahn Y, Stern O, Papke RL, Matusik R, Nandana SR, Ziv R, Pichinuk E, Salame S *et al*: **PATE gene clusters code for multiple, secreted TFP/Ly-6/uPar proteins that are expressed in reproductive and neuron-rich tissues and possess neuromodulatory activity**. *Journal of Biological Chemistry* 2008, **283**(24):16928-16939.
110. Galat A, Gross G, Drevet P, Sato A, Menez A: **Conserved structural determinants in three-fingered protein domains**. *FEBS Journal* 2008, **275**:3207-3225.
111. Naimuddin M, Kobayashi S, Tsutsui C, Machida M, Nemoto N, Sakai T, Kubo T: **Directed evolution of a three-finger neurotoxin by using cDNA display yields antagonists as well as agonists of interleukin-6 receptor signaling**. *Molecular Brain* 2011, **4**:2.
112. Fry BG, Wüster W, Kini RM, Brusica V, Khan A, Venkataraman D, Rooney AP: **Molecular evolution and phylogeny of elapid snake venom three-finger toxins**. *Journal of Molecular Evolution* 2003, **57**(1):110-129.
113. Nei M, Gu X, Sitnikova T: **Evolution by the birth-and-death process in multigene families of the vertebrate immune system**. *Proceedings of the National Academy of Sciences* 1997, **94**(15):7799-7806.
114. Rooney AP, Piontkivska H, Nei M: **Molecular evolution of the nontandemly repeated genes of the histone 3 multigene family**. *Molecular Biology and Evolution* 2002, **19**(1):68-75.
115. Fry BG: **From genome to "venome": molecular origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences and related body proteins**. *Genome Research* 2005, **15**(3):403-420.
116. Lin Y-L, Lin S-R, Wu TT, Chang L-S: **Evidence showing an intermolecular interaction between KChIP proteins and Taiwan cobra cardiotoxins**. *Biochemical and Biophysical Research Communications* 2004, **319**(3):720-724.

117. Albrand J, Blackledge M, Pascaud F, Hollecker M, Marion D: **NMR and restrained molecular dynamics study of the three-dimensional solution structure of toxin FS2, a specific blocker of the L-type calcium channel, isolated from black mamba venom.** *Biochemistry* 1995, **34**(17):5923-5937.
118. de Weille JR, Schweitz H, Maes P, Tartar A, Lazdunski M: **Calciseptine, a peptide isolated from black mamba venom, is a specific blocker of the L-type calcium channel.** *Proceedings of the National Academy of Sciences* 1991, **88**(6):2437-2440.
119. Chang CC: **The action of snake venoms on nerve and muscles.** In: *Snake Venoms, Handbook of Experimental Pharmacology*. Edited by Lee CY. Berlin: Springer-Verlag; 1979: 309-376.
120. Tsetlin V: **Snake venom  $\alpha$ -neurotoxins and other 'three-finger' proteins.** *European Journal of Biochemistry* 1999, **264**(2):281-286.
121. Jerusalinsky D, Harvey AL: **Toxins from mamba venoms: small proteins with selectivities for different subtypes of muscarinic acetylcholine receptors.** *Trends in Pharmacological Sciences* 1994, **15**:424-430.
122. McDowell RS, Dennis MS, Louie A, Shuster M, Mulkerrin MG, Lazarus RA: **Mambin, a potent glycoprotein IIb-IIIa antagonist and platelet aggregation inhibitor structurally related to the short neurotoxins.** *Biochemistry* 1992, **31**:4766-4772.
123. Wu P-L, Lee S-C, Chuang C-C, Mori S, Akakura N, Wu W-g, Takada Y: **Non-cytotoxic cobra cardiotoxin A5 binds to  $\alpha v \beta 3$  integrin and inhibits bone resorption.** *Journal of Biological Chemistry* 2006, **281**(12):7937-7945.
124. Bilwes A, Rees B, Moras D, Ménez R, Ménez A: **X-ray structure at 1.55 Å of toxin  $\gamma$ , a cardiotoxin from *Naja nigricollis* venom: crystal packing reveals a model for insertion into membranes.** *Journal of Molecular Biology* 1994, **239**(1):122-136.
125. Kini RM, Doley R: **Structure, function and evolution of three-finger toxins: mini proteins with multiple targets.** *Toxicon* 2010, **56**(6):855-867.
126. Vonk FJ, Casewell NR, Henkel CV, Heimberg AM, Jansen HJ, McCleary RJR, Kerkkamp HME, Vos RA, Guerreiro I, Calvete JJ *et al*: **The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system.** *Proceedings of the National Academy of Sciences* 2013.
127. Arnold SJ: **Morphology, Performance and Fitness.** *American Zoologist* 1983, **23**(2):347-361.

128. Uyeda JC, Hansen TF, Arnold SJ, Pienaar J: **The million-year wait for macroevolutionary bursts.** *Proceedings of the National Academy of Science USA* 2011, **108**(38):15908-15913.
129. Estes S, Arnold SJ: **Resolving the paradox of stasis: models with stabilizing selection explain evolutionary divergence on all timescales.** *The American Naturalist* 2007, **169**(2):227-244.
130. Lande R: **Models of speciation by sexual selection on polygenic traits.** *Proceedings of the National Academy of Sciences* 1981, **78**(6):3721-3725.
131. Lee Y-H, Ota T, Vacquier VD: **Positive selection is a general phenomenon in the evolution of abalone sperm lysin.** *Molecular Biology and Evolution* 1995, **12**(2):231-238.
132. Yang Z, Swanson WJ, Vacquier VD: **Maximum-likelihood analysis of molecular adaptation in abalone sperm lysin reveals variable selective pressures among lineages and sites.** *Molecular Biology and Evolution* 2000, **17**(10):1446-1455.
133. Swanson WJ, Vacquier VD: **Concerted evolution in an egg receptor for a rapidly evolving abalone sperm protein.** *Science* 1998, **281**(5377):710-712.
134. Swanson WJ, Aquadro CF, Vacquier VD: **Polymorphism in abalone fertilization proteins is consistent with the neutral evolution of the egg's receptor for lysin (VERL) and positive Darwinian selection of sperm lysin.** *Molecular Biology and Evolution* 2001, **18**(3):376-383.
135. Swanson WJ, Vacquier VD: **The abalone egg vitelline envelope receptor for sperm lysin is a giant multivalent molecule.** *Proceedings of the National Academy of Sciences* 1997, **94**(13):6724-6729.
136. Nei M: **Molecular evolutionary genetics.** New York, NY: Columbia University Press; 1987.
137. Pal C, Papp B, Lercher MJ: **An integrated view of protein evolution.** *Nat Rev Genet* 2006, **7**(5):337-348.
138. Warnecke T, Weber CC, Hurst LD: **Why there is more to protein evolution than protein function: splicing, nucleosomes and dual-coding sequence.** *Biochemical Society Transactions* 2009, **037**(4):756-761.
139. Pesole G, Grillo G, Larizza A, Liuni S: **The untranslated regions of eukaryotic mRNAs: Structure, function, evolution and bioinformatic tools for their analysis.** *Briefings in Bioinformatics* 2000, **1**(3):236-249.



140. Asson-Batres MA, Spurgeon SL, Diaz J, DeLoughery TG, Bagby GCJ: **Evolutionary conservation of the AU-rich 3' untranslated region of messenger RNA.** *Proceedings of the National Academy of Science USA* 1994, **91**:1318-1322.
141. Greener MJ, Sewry CA, Muntoni F, Roberts RG: **The 3'-untranslated region of the dystrophin gene - conservation and consequences of loss.** *European Journal of Human Genetics* 2002, **10**:413-420.
142. Friedberg F, Saunders A, Rhoads AR: **Conservation of the 3'-untranslated regions of calmodulin mRNAs in cetaceans.** *Molecular Biology Reports* 2003, **30**:193-198.
143. Yadav D, Ngolab J, Lim RS-H, Krishnamurthy S, Bui JD: **Cutting Edge: Down-Regulation of MHC Class I-Related Chain A on Tumor Cells by IFN- $\gamma$ -Induced MicroRNA.** *J Immunol* 2009, **182**(1):39-43.
144. Moran Y, Weinberger H, Sullivan JC, Reitzel AM, Finnerty JR, Gurevitz M: **Concerted Evolution of Sea Anemone Neurotoxin Genes Is Revealed through Analysis of the *Nematostella vectensis* Genome.** *Molecular Biology and Evolution* 2008, **25**(4):737-747.
145. Conticello SG, Gilad Y, Avidan N, Ben-Asher E, Levy Z, Fainzilber M: **Mechanisms for evolving hypervariability: the case of conopeptides.** *Molecular Biology and Evolution* 2001, **18**(2):120-131.
146. Ohno M, Chijiwa T, Oda-Ueda N, Ogawa T, Hattori S: **Molecular evolution of myotoxic phospholipases A2 from snake venom.** *Toxicon* 2003, **42**(8):841-854.
147. Chen M-S, Liu X, Yang Z, Zhao H, Shukle R, Stuart J, Hulbert S: **Unusual conservation among genes encoding small secreted salivary gland proteins from a gall midge.** *BMC Evolutionary Biology* 2010, **10**(1):296.
148. Chang LS, Chung C, Lin J, Hong E: **Organization and phylogenetic analysis of  $\kappa$ -bungarotoxin genes from *Bungarus multicinctus* (Taiwan banded krait).** *Genetica* 2002, **115**:213-221.
149. Fryklund L, Eaker D: **The complete covalent structure of a cardiotoxin from the venom of *Naja nigricollis* (African Black-Necked Spitting Cobra).** *Biochemistry* 1975, **14**(13):2865-2871.
150. Low BW, Preston HS, Sato A, Rosen LS, Searl JE, Rudko AD, Richardson JS: **Three dimensional structure of erabutoxin b neurotoxic protein: Inhibitor of acetylcholine receptor.** *Proclamations of the National Academy of Science USA* 1976, **73**(9):2991-2994.

151. Doley R, Pahari S, Mackessy SP, Kini RM: **Accelerated exchange of exon segments in Viperid three-finger toxin genes (*Sistrurus catenatus edwardsii*; Desert Massasauga).** *BMC Evolutionary Biology* 2008, **8**(196).
152. Ghaemmaghami S, Huh W-K, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK, Weissman JS: **Global analysis of protein expression in yeast.** *Nature* 2003, **425**(6959):737-741.
153. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: **MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods.** *Molecular Biology & Evolution* 2011, **28**:2731-2739.
154. Jones DT, Taylor WR, Thornton JM: **The rapid generation of mutation data matrices from protein sequences.** *Computer Applications in the Biosciences* 1992, **9**:275-282.
155. Yang Z: **PAML 4: Phylogenetic Analysis by Maximum Likelihood.** *Molecular Biology and Evolution* 2007, **24**(8):1586-1591.
156. Yang Z: **Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution.** *Molecular Biology and Evolution* 1998, **15**(5):568-573.
157. Swanson WJ, Nielsen R, Yang Q: **Pervasive Adaptive Evolution in Mammalian Fertilization Proteins.** *Molecular Biology and Evolution* 2003, **20**(1):18-20.
158. Wong WSW, Yang Z, Goldman N, Nielsen R: **Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites.** *Genetics* 2004, **168**(2):1041-1051.
159. Invitrogen: **Advanced RACE method amplifies only full-length cDNA ends.** *Expressions Newsletter* 2000, **7.3**:2-3.
160. Richter JD: **Cytoplasmic Polyadenylation in Development and Beyond.** *Microbiol Mol Biol Rev* 1999, **63**(2):446-456.
161. Chang LS, Lin SK, Chung C: **Molecular cloning and evolution of the genes encoding the precursors of Taiwan cobra cardiotoxin and cardiotoxin-like basic protein.** *Biochemical Genetics* 2004, **42**:429-440.
162. Sessions SK, Matthias S, Vieites DR, Quarles R, Min M-S, Wake DB: **Cytogenetic analysis of the Asian Plethodontid salamander, *Karsenia koreana*: Evidence for karyotypic conservation, chromosome repatterning, and genome size evolution.** *Chromosome Research* 2008, **16**:563-574.

163. Smith JJ, Putta S, Zhu W, Pao GM, Verma IM, Hunter T, Bryant SV, Gardiner DM, Harkins TT, Voss SR: **Genic regions of a large salamander genome contains long introns and novel genes.** *BMC Genomics* 2009, **10**(19).
164. Peng SSY, Chen C-YA, Xu N, Shyu A-B: **RNA stabilization by the AU-rich element binding protein, HuR, an ELAV protein.** *EMBO J* 1998, **17**(12):3461-3470.
165. Wang G, Guo X, Floros J: **Differences in the translation efficiency and mRNA stability mediated by 5'-UTR splice variants of human SP-A1 and SP-A2 genes.** *Am J Physiol Lung Cell Mol Physiol* 2005, **289**(3):L497-508.
166. Barreau C, Paillard L, Osborne HB: **AU-rich elements and associated factors: are there unifying principles?** *Nucl Acids Res* 2006, **33**(22):7138-7150.
167. Lynch M: **The Origins of Genome Architecture.** Sunderland, MA: Sinauer Associates Inc; 2007.
168. Betel D, Wilson M, Gabow A, Marks DS, Sander C: **The microRNA.org resource: targets and expression.** *Nucl Acids Res* 2008, **36**(suppl\_1):D149-153.
169. Collier B, Gorgoni B, Loveridge C, Cooke HJ, Gray NK: **The DAZL family proteins are PABP-binding proteins that regulate translation in germ cells.** *EMBO J* 2005, **24**(14):2656-2666.
170. Reynolds N, Collier B, Maratou K, Bingham V, Speed RM, Taggart M, Semple CA, Gray NK, Cooke HJ: **Dazl binds in vivo to specific transcripts and can regulate the pre-meiotic translation of Mvh in germ cells.** *Human Molecular Genetics* 2005, **14**(24):3899-3909.
171. Brook M, Smith JWS, Gray NK: **The DAZL and PABP families: RNA-binding proteins with interrelated roles in translational control in oocytes.** *Reproduction* 2009, **137**(4):595-617.
172. Fox M, Urano J, Reijo Pera RA: **Identification and characterization of RNA sequences to which human PUMILIO-2 (PUM2) and deleted in Azoospermia-like (DAZL) bind.** *Genomics* 2005, **85**(1):92-105.
173. Burri R, Salamin N, Studer RA, Roulin A, Fumagalli L: **Adaptive divergence of ancient gene duplicates in the avian MHC class II  $\beta$ .** *Molecular Biology and Evolution* 2010, **27**(10):2360-2374.
174. Das S, Nikolaidis N, Goto H, McCallister C, Li J, Hirano M, Cooper MD: **Comparative genomics and evolution of the alpha-defensin multigene family in primates.** *Molecular Biology and Evolution* 2010, **27**(10):2333-2343.

175. Espiritu DJD, Watkins M, Dia-Monje V, Cartier GE, Cruz LJ, Olivera BM: **Venomous cone snails: molecular phylogeny and the generation of toxin diversity.** *Toxicon* 2001, **39**(12):1899-1916.
176. Melrose DR, Reed HC, Patterson RL: **Androgen steroids associated with boar odour as an aid to the detection of oestrus in pig artificial insemination.** *The British Veterinary Journal* 1971, **127**(10):497-502.
177. Albone ES: **Mammalian semiochemistry.** New York: J. Wiley; 1984.
178. Johnston RE: **Chemical communication and pheromones: the types of chemical signals and the role of the vomeronasal system.** In: *The Neurobiology of Taste and Smell.* Edited by Finger TE, Silver WL, Restrepo D. New York: Wiley-Liss; 2000: 101-127.
179. Symonds MRE, Elgar MA: **The evolution of pheromone diversity.** *Trends in Ecology & Evolution* 2008, **23**(4):220-228.
180. Koonin EV, Wolf YI, Karev GP: **The structure of the protein universe and genome evolution.** *Nature* 2002, **420**(6912):218-223.
181. Pearl FMG, Bennett CF, Bray JE, Harrison AP, Martin N, Shepherd A, Sillitoe I, Thornton J, Orengo CA: **The CATH database: an extended protein family resource for structural and functional genomics.** *Nucleic Acids Research* 2003, **31**(1):452-455.
182. Sillitoe I, Cuff AL, Dessailly BH, Dawson NL, Furnham N, Lee D, Lees JG, Lewis TE, Studer RA, Rentzsch R *et al*: **New functional families (FunFams) in CATH to improve the mapping of conserved functional sites to 3D structures.** *Nucleic Acids Research* 2013, **41**:D490-D498.
183. Harms MJ, Eick GN, Goswami D, Colucci JK, Griffin PR, Ortlund EA, Thornton JW: **Biophysical mechanisms for large-effect mutations in the evolution of steroid hormone receptors.** *Proceedings of the National Academy of Sciences* 2013, **in press.**
184. Xu H, Freitas MA: **MassMatrix: a database search program for rapid characterization of proteins and peptides from tandem mass spectrometry data.** *Proteomics* 2009, **9**(6):1548-1555.
185. Stemmer WPC, Cramer A, Ha KD, Brennan TM, Heyneker HL: **Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides.** *Gene* 1995, **164**(1):49-53.
186. Perez-Iratxeta C, Andrade-Navarro M: **K2D2: estimation of protein secondary structure from circular dichroism spectra.** *BMC Structural Biology* 2008, **8**(1):25.

187. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A: **NMRPipe: a multidimensional spectral processing system based on UNIX pipes.** *Journal of Biomolecular NMR* 1995, **6**(3):277-293.
188. Piotto M, Saudek V, Sklenar V: **Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solution.** *Journal of Biomolecular NMR* 1992, **2**:661-665.
189. Güntert P, Mumenthaler C, Wüthrich K: **Torsion angle dynamics for NMR structure calculation with the new program Dyana.** *Journal of Molecular Biology* 1997, **273**(1):283-298.
190. Herrmann T, Güntert P, Wüthrich K: **Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA.** *Journal of Molecular Biology* 2002, **319**(1):209-227.
191. Shen Y, Delaglio F, Cornilescu G, Bax A: **TALOS plus: a hybrid method for predicting protein torsion angles from NMR chemical shifts.** *Journal of Biomolecular NMR* 2009, **44**:213-223.
192. Sharma D, Rajarathnam K: **<sup>13</sup>C NMR chemical shifts can predict disulfide bond formation.** *Journal of Biomolecular NMR* 2000, **18**(2):165-171.
193. Berjanskii MV, Wishart DS: **A simple method to predict protein flexibility using secondary chemical shifts.** *Journal of American Chemical Society* 2005, **127**(43):14970-14971.
194. Garcia-Boronat M, Diez-Rivero CM, Reinherz EL, Reche PA: **PVS: a web server for protein sequence variability analysis tuned to facilitate conserved epitope discovery.** *Nucleic Acids Research* 2008, **36**:W35-41.
195. Das R, Baker D: **Macromolecular modeling with Rosetta.** *Annual Review of Biochemistry* 2008, **77**:363-382.
196. Thompson J, Higgins D, Gibson T: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice.** *Nucleic Acids Research* 1994, **22**(22):4673-4680.
197. Gray WR: **Disulfide structures of highly bridged peptides: A new strategy for analysis.** *Protein Science* 1993, **2**(10):1732-1748.
198. Wouters MA, George RA, Haworth NL: **"Forbidden" disulfides: their role as redox switches.** *Current Protein and Peptide Science* 2007, **8**(5):484-495.

199. Lyukmanova EN, Shenkarev ZO, Schulga AA, Ermolyuk YS, Mordvintsev DY, Utkin YN, Shoulepko MA, Hogg RC, Bertrand D, Dolgikh DA *et al*: **Bacterial expression, NMR, and electrophysiology analysis of chimeric short/long-chain  $\alpha$ -neurotoxins acting on neuronal nicotinic receptors.** *Journal of Biological Chemistry* 2007, **282**(34):24784-24791.
200. Mourier G, Servent D, Zinn-Justin S, Ménez A: **Chemical engineering of a three-fingered toxin with anti- $\alpha 7$  neuronal acetylcholine receptor activity.** *Protein Engineering* 2000, **13**(3):217-225.
201. Greenwald J, Le V, Corrigan A, Fischer W, Komives E, Vale W, Choe S: **Characterization of the extracellular ligand-binding domain of the type II activin receptor.** *Biochemistry* 1998, **37**(47):16711-16718.
202. Laskowski RA, Rullmann AC, MacArthur MW, Kaptein R, Thornton JM: **AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR.** *Journal of Biomolecular NMR* 1996, **8**(4):477-486.
203. Heinze J, d'Ettorre P: **Honest and dishonest communication in social Hymenoptera.** *Journal of Experimental Biology* 2009, **212**(12):1775-1779.
204. Wyatt T: **Pheromones and signature mixtures: defining species-wide signals and variable cues for identity in both invertebrates and vertebrates.** *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology* 2010, **196**(10):685-700.
205. Fine J, Sorensen P: **Isolation and Biological Activity of the Multi-Component Sea Lamprey Migratory Pheromone.** *Journal of Chemical Ecology* 2008, **34**(10):1259-1267.
206. LeMaster M, Mason R: **Variation in a female sexual attractiveness pheromone controls male mate choice in garter snakes.** *Journal of Chemical Ecology* 2002, **28**(6):1269-1285.
207. Legrand S, Botton M, Coracini M, Witzgall P, Unelius CR: **Synthesis and field tests of sex pheromone components of the leafroller *Argyrotaenia sphaleropa*.** *Z Naturforsch* 2004, **59c**:708-712.
208. Houck LD, Watts RA, Arnold SJ, Bowen KE, Kiemnec KM, Godwin HA, Feldhoff PW, Feldhoff RC: **A recombinant courtship pheromone affects sexual receptivity in a plethodontid salamander.** *Chemical Senses* 2008, **33**(7):623-631.
209. Novotny M, Harvey S, Jemiolo B, Alberts J: **Synthetic pheromones that promote inter-male aggression in mice.** *Proceedings of the National Academy of Science USA* 1985, **82**:2059-2061.

210. Sorensen PW, Stacey NE: **Evolution and specialization of fish hormonal pheromones**. In: *Advances in chemical signals in vertebrates*. Edited by Johnston RE, Müller-Schwarze D, Sorensen PW. New York: Kluwer Academic/Plenum Press; 1999: 15-48.
211. Cardé RT, Haynes KF: **Structure of the pheromone communication channel in moths**. In: *Advances in insect chemical ecology*. Edited by Cardé RT, Miller JG. Cambridge: Cambridge University Press; 2004: 283-332.
212. Lassance J-M, Lofstedt C: **Concerted evolution of male and female display traits in the European corn borer, *Ostrinia nubilalis***. *BMC Biology* 2009, 7(1):10.
213. de Bruyne M, Baker TC: **Odor detection in insects: volatile codes**. *Journal of Chemical Ecology* 2008, 34:882-897.
214. Woodley SK: **Plasma androgen levels, spermatogenesis, and secondary sexual characteristics in two species of Plethodontid salamanders with dissociated reproductive patterns**. *General and Comparative Endocrinology* 1994, 96(2):206-214.
215. Rollmann SM, Houck LD, Feldhoff RC: **Conspecific and heterospecific pheromone effects on female receptivity**. *Animal Behaviour* 2003(66):857-861.
216. Kiemiec-Tyburczy KM, Woodley SK, Feldhoff PW, Feldhoff RC, Houck LD: **Dermal application of courtship pheromones does not influence receptivity in female red-legged salamanders (*Plethodon shermani*)**. *Journal of Herpetology* 2011, 45(2):169-173.
217. Löfstedt C, Herrebut WM, Menken SBJ: **Sex pheromones and their potential role in the evolution of reproductive isolation in small ermine moths (*Yponomeutidae*)**. *Chemoecology* 1991, 2(1):20-28.
218. Ryan MJ, Rand AS: **Sexual selection in female perceptual space: how female túngara frogs perceive and respond to complex population variation in acoustic mating signals**. *Evolution* 2003, 57(11):2608-2618.
219. Taylor RC, Ryan MJ: **Interactions of multisensory components perceptually rescue túngara frog mating signals**. *Science* 2013, 341(6143):273-274.
220. Birch MC, Haynes KF: **Insect pheromones**. London: E. Arnold; 1982.
221. Magklara A, Yen A, Colquitt Bradley M, Clowney EJ, Allen W, Markenscoff-Papadimitriou E, Evans Zoe A, Kheradpour P, Mountoufaris G, Carey C *et al*: **An Epigenetic Signature for Monoallelic Olfactory Receptor Expression**. *Cell* 2011, 145(4):555-570.

222. Sever DM: **Induction of secondary sex characters in *Eurycea quadridigitata*.** *Copeia* 1976, **1976**(4):830-833.
223. Marc RE: **Mapping glutamatergic drive in the vertebrate retina with a channel-permeant organic cation.** *The Journal of Comparative Neurology* 1999, **407**(1):47-64.
224. Marc RE: **Kainate activation of horizontal, bipolar, amacrine, and ganglion cells in the rabbit retina.** *The Journal of Comparative Neurology* 1999, **407**(1):65-76.
225. Michel WC, Steullet P, Cate HS, Burns CJ, Zhainazarov AB, Derby CD: **High-resolution functional labeling of vertebrate and invertebrate olfactory receptor neurons using agmatine, a channel-permeant cation.** *Journal of Neuroscience Methods* 1999, **90**(2):143-156.
226. Schubert SN, Houck LD, Feldhoff PW, Feldhoff RC, Woodley SK: **The effects of sex on chemosensory communication in a terrestrial salamander (*Plethodon shermani*).** *Hormones and Behavior* 2008, **54**:270-277.
227. Vaccaro EA, Feldhoff PW, Feldhoff RC, Houck LD: **A pheromone mechanism for swaying female mate choice: enhanced affinity for a sexual stimulus in a woodland salamander.** *Animal Behaviour* 2010, **80**(6):983-989.
228. Zuk M, Ligon JD, Thornhill R: **Effects of experimental manipulation of male secondary sexual characters on female mating preferences in red jungle fowl.** *Animal Behaviour* 1992, **44**:999-1006.
229. Moller AP, Pomiankowski A: **Why have birds got multiple sexual ornaments?** *Behav Ecol Sociobiol* 1993, **32**(3):167-176.
230. Madison DM: **Intraspecific odor preferences between salamanders of the same sex: dependence on season and proximity of residence.** *Canadian Journal of Zoology* 1975, **53**(9):1356-1361.
231. Gibbons ME, Ferguson AM, Lee DR: **Both learning and heritability affect foraging behaviour of red-backed salamanders, *Plethodon cinereus*.** *Animal Behaviour* 2005, **69**(3):721-732.
232. Ferrari MO, Chivers D: **Learning about non-predators and safe places: the forgotten elements of risk assessment.** *Anim Cogn* 2011, **14**(3):309-316.
233. Mathis A, Ferrari MCO, Windel N, Messier F, Chivers DP: **Learning by embryos and the ghost of predation future.** *Proceedings of the Royal Society B: Biological Sciences* 2008, **275**(1651):2603-2607.



234. Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrovsky E, Lander ES, Golub TR: **Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation.** *Proceedings of the National Academy of Sciences* 1999, **96**(6):2907-2912.
235. Reilly GC, Engler AJ: **Intrinsic extracellular matrix properties regulate stem cell differentiation.** *Journal of Biomechanics* 2010, **43**(1):55-62.
236. Sekiguchi H, Ii M, Jujo K, Thorne T, Ito A, Klyachko E, Hamada H, Kessler J, Tabata Y, Kawana M *et al*: **Estradiol promotes neural stem cell differentiation into endothelial lineage and angiogenesis in injured peripheral nerve.** *Angiogenesis* 2013, **16**(1):45-58.
237. Chadwick K, Wang L, Li L, Menendez P, Murdoch B, Rouleau A, Bhatia M: **Cytokines and BMP-4 promote hematopoietic differentiation of human embryonic stem cells.** *Blood* 2003, **102**(3):906-915.
238. Johe KK, Hazel TG, Muller T, Dugich-Djordjevic MM, McKay RD: **Single factors direct the differentiation of stem cells from the fetal and adult central nervous system.** *Genes & Development* 1996, **10**(24):3129-3140.
239. Spradling A, Drummond-Barbosa D, Kai T: **Stem cells find their niche.** *Nature* 2001, **414**(6859):98-104.
240. Glisovic T, Bachorik JL, Yong J, Dreyfuss G: **RNA-binding proteins and post-transcriptional gene regulation.** *FEBS Letters* 2008, **582**(14):1977-1986.
241. Krol J, Loedige I, Filipowicz W: **The widespread regulation of microRNA biogenesis, function and decay.** *Nat Rev Genet* 2010, **11**(9):597-610.
242. Anderson GH, Veit B, Hanson MR: **The Arabidopsis AtRaptor genes are essential for post-embryonic plant growth.** *BMC biology* 2005, **3**(1):12.
243. Doerner P, Joergensen J-E, You R, Steppuhn J, Lamb C: **Control of root growth and development by cyclin expression.** *Nature* 1996, **380**(6574):520-523.
244. Kimble J, Hirsh D: **The postembryonic cell lineages of the hermaphrodite and male gonads in *Caenorhabditis elegans*.** *Developmental biology* 1979, **70**(2):396-417.
245. Ririe TO, Fernandes JS, Sternberg PW: **The *Caenorhabditis elegans* vulva: a post-embryonic gene regulatory network controlling organogenesis.** *Proceedings of the National Academy of Sciences* 2008, **105**(51):20095-20099.
246. Tata JR: **Gene expression during metamorphosis: An ideal model for post-embryonic development.** *Bioessays* 1993, **15**(4):239-248.

247. Sever DM: **Morphology and seasonal variation of the mental hedonic glands of the dwarf salamander, *Eurycea quadridigitata* (Holbrook).** *Herpetologica* 1975, **31**:241-251.
248. Li B, Dewey C: **RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome.** *BMC Bioinformatics* 2011, **12**(1):323.
249. Leng N, Dawson JA, Thomson JA, Ruotti V, Rissman AI, Smits BMG, Haag JD, Gould MN, Stewart RM, Kendziorski C: **EBSeq: An empirical Bayes hierarchical model for inference in RNA-seq experiments.** *Bioinformatics* 2013.
250. Si Y, Liu P, Li P, Brutnell TP: **Model-based clustering for RNA-seq data.** *Bioinformatics* 2014, **30**(2):197-205.
251. Rollmann SM, Houck LD, Feldhoff RC: **Population variation in salamander courtship pheromones.** *Journal of Chemical Ecology* 2000, **26**:2713-2722.
252. Rubin LL: **Increases in muscle Ca<sup>2+</sup> mediate changes in acetylcholinesterase and acetylcholine receptors caused by muscle contraction.** *Proceedings of the National Academy of Sciences* 1985, **82**(20):7121-7125.
253. Chowdhury U, Samant R, Fodstad O, Shevde L: **Emerging role of nuclear protein 1 (NUPR1) in cancer biology.** *Cancer Metastasis Rev* 2009, **28**(1-2):225-232.
254. Cano CE, Hamidi T, Sandi MJ, Iovanna JL: **Nupr1: The Swiss-knife of cancer.** *Journal of Cellular Physiology* 2011, **226**(6):1439-1443.
255. De Leeuw F, Zhang T, Wauquier C, Huez G, Kruys V, Gueydan C: **The cold-inducible RNA-binding protein migrates from the nucleus to cytoplasmic stress granules by a methylation-dependent mechanism and acts as a translational repressor.** *Experimental Cell Research* 2007, **313**(20):4130-4144.
256. Kato M, Han Tina W, Xie S, Shi K, Du X, Wu Leeju C, Mirzaei H, Goldsmith Elizabeth J, Longgood J, Pei J *et al*: **Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels.** *Cell* 2012, **149**(4):753-767.
257. Biniossek ML, Nägler DK, Becker-Pauly C, Schilling O: **Proteomic identification of protease cleavage sites characterizes prime and non-prime specificity of cysteine cathepsins B, L, and S.** *Journal of Proteome Research* 2011, **10**(12):5363-5373.
258. MS S, H R, M G: **Overexpression of p21WAF1/CIP1 induces growth arrest, giant cell formation and apoptosis in human breast carcinoma cell lines.** *Oncogene* 1995, **11**(9):1899-1905.

259. Kellogg DR: **Wee1-dependent mechanisms required for coordination of cell growth and cell division.** *Journal of Cell Science*, **116**(24):4883-4890.
260. Yokota J, Akiyama T, Fung YK, Benedict WF, Namba Y, Hanaoka M, Wada M, Terasaki T, Shimosato Y, Sugimura T: **Altered expression of the retinoblastoma (RB) gene in small-cell carcinoma of the lung.** *Oncogene* 1988, **3**(4):471-475.
261. Thongkukiatkul A, Jungudomjaroen S, Ratanapahira C: **Spermatogenesis and chromatin condensation in male germ cells of sea cucumber *Holothuria leucospilota* (Clark, 1920).** *Tissue and Cell* 2008, **40**(3):167-175.
262. Chapman J, Michael S: **Proposed mechanism for sperm chromatin condensation/decondensation in the male rat.** *Reproductive Biology and Endocrinology* 2003, **1**(1):20.
263. Swain JE, Ding J, Brautigan DL, Villa-Moruzzi E, Smith GD: **Proper chromatin condensation and maintenance of histone H3 phosphorylation during mouse oocyte meiosis requires protein phosphatase activity.** *Biology of Reproduction* 2007, **76**(4):628-638.
264. Tan J-H, Wang H-L, Sun X-S, Liu Y, Sui H-S, Zhang J: **Chromatin configurations in the germinal vesicle of mammalian oocytes.** *Molecular Human Reproduction* 2009, **15**(1):1-9.
265. Tay J, Richter JD: **Germ cell differentiation and synaptonemal complex formation are disrupted in CPEB knockout mice.** *Developmental Cell* 2001, **1**(2):201-213.
266. Richter JD: **CPEB: a life in translation.** *Trends in biochemical sciences* 2007, **32**(6):279-285.
267. Belloc E, Pique M, Mendez R: **Sequential waves of polyadenylation and deadenylation define a translation circuit that drives meiotic progression.** *Biochemical Society Transactions* 2008, **36**:665-670.
268. Hasegawa E, Karashima T, Sumiyoshi E, Yamamoto M: ***C. elegans* CPB-3 interacts with DAZ-1 and functions in multiple steps of germline development.** *Developmental Biology* 2006, **295**(2):689-699.
269. Moore FL, Jaruzelska J, Fox MS, Urano J, Firpo MT, Turek PJ, Dorfman DM, Pera RAR: **Human Pumilio-2 is expressed in embryonic stem cells and germ cells and interacts with DAZ (Deleted in AZoospermia) and DAZ-Like proteins.** *Proceedings of the National Academy of Sciences of the United States of America* 2003, **100**(2):538-543.

270. Tsui S, Dai T, Roettger S, Schempp W, Salido EC, Yen PH: **Identification of two novel proteins that interact with germ-cell-specific RNA-binding proteins DAZ and DAZL1.** *Genomics* 2000, **65**(3):266-273.
271. Morton S, Yang H-T, Moleleki N, Campbell DG, Cohen P, Rousseau S: **Phosphorylation of the ARE-binding protein DAZAP1 by ERK2 induces its dissociation from DAZ.** *Biochem J* 2006, **399**(2):265-273.
272. Yuan L, Liu J-G, Zhao J, Brundell E, Daneholt B, Höög C: **The murine SCP3 gene is required for synaptonemal complex assembly, chromosome synapsis, and male fertility.** *Molecular Cell* 2000, **5**(1):73-83.
273. Reynolds N, Collier B, Bingham V, Gray NK, Cooke HJ: **Translation of the synaptonemal complex component Sycp3 is enhanced in vivo by the germ cell specific regulator Dazl.** *RNA* 2007, **13**(7):974-981.
274. Li S, Zhang Z, Xue J, Liu A, Zhang H: **Cold-inducible RNA binding protein inhibits H2O2-induced apoptosis in rat cortical neurons.** *Brain Research* 2012, **1441**(0):47-52.
275. Nishiyama H, Danno S, Kaneko Y, Itoh K, Yokoi H, Fukumoto M, Okuno H, Millan JL, Matsuda T, Yoshida O *et al*: **Decreased expression of cold-inducible RNA-binding protein (CIRP) in male germ cells at elevated temperature.** *American Journal of Pathology* 1998, **152**(1):289-296.
276. Kim JS, Park SJ, Kwak KJ, Kim YO, Kim JY, Song J, Jang B, Jung C-H, Kang H: **Cold shock domain proteins and glycine-rich RNA-binding proteins from *Arabidopsis thaliana* can promote the cold adaptation process in *Escherichia coli*.** *Nucleic Acids Research* 2007, **35**(2):506-516.
277. Al-Fageeh MB, Smales CM: **Cold-inducible RNA binding protein (CIRP) expression is modulated by alternative mRNAs.** *RNA* 2009, **15**(6):1164-1176.
278. Hilgers V, Teixeira D, Parker R: **Translation-independent inhibition of mRNA deadenylation during stress in *Saccharomyces cerevisiae*.** *RNA* 2006, **12**(10):1835-1845.
279. Gowrishankar G, Winzen R, Dittrich-Breiholz O, Redich N, Kracht M, Holtmann H: **Inhibition of mRNA deadenylation and degradation by different types of cell stress.** *Biological Chemistry* 2006, **387**(3):323-327.
280. Buchan JR, Parker R: **Eukaryotic stress granules: the ins and outs of translation.** *Molecular Cell* 2009, **36**(6):932-941.
281. Parker R, Sheth U: **P bodies and the control of mRNA translation and degradation.** *Molecular cell* 2007, **25**(5):635-646.

282. Maris C, Dominguez C, Allain FHT: **The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression.** *FEBS Journal* 2005, **272**(9):2118-2131.
283. Oubridge C, Ito N, Evans PR, Teo CH, Nagai K: **Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin.** *Nature* 1994, **372**(6505):432-438.
284. Allain FHT, Gubser CC, Howe PWA, Nagai K, Neuhaus D, Varani G: **Specificity of ribonucleoprotein interaction determined by RNA folding during complex formation.** *Nature* 1996, **380**(6575):646-650.
285. Varani L, Gunderson SI, Mattaj IW, Kay LE, Neuhaus D, Varani G: **The NMR structure of the 38 kDa U1A protein - PIE RNA complex reveals the basis of cooperativity in regulation of polyadenylation by human U1A protein.** *Nat Struct Mol Biol* 2000, **7**(4):329-335.
286. Qiang X, Yang W-L, Wu R, Zhou M, Jacob A, Dong W, Kuncewitch M, Ji Y, Yang H, Wang H *et al*: **Cold-inducible RNA-binding protein (CIRP) triggers inflammatory responses in hemorrhagic shock and sepsis.** *Nat Med* 2013, **19**(11):1489-1495.
287. Hai T, Wolfgang CD, Marsee DK, Allen AE, Sivaprasad U: **ATF3 and stress response.** *Gene expression* 1999, **7**:321-335.
288. Singh S, Vrishni S, Singh BK, Rahman I, Kakkar P: **Nrf2-ARE stress response mechanism: A control point in oxidative stress-mediated dysfunctions and chronic inflammatory diseases.** *Free Radical Research* 2010, **44**(11):1267-1288.
289. Dang CV: **Links between metabolism and cancer.** *Genes & Development* 2012, **26**(9):877-890.
290. Zeng Y, Kulkarni P, Inoue T, Getzenberg RH: **Down-regulating cold shock protein genes impairs cancer cell survival and enhances chemosensitivity.** *Journal of Cellular Biochemistry* 2009, **107**:179-188.
291. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q *et al*: **Full-length transcriptome assembly from RNA-Seq data without a reference genome.** *Nat Biotech* 2011, **29**(7):644-652.
292. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *Journal of Molecular Biology* 1990, **215**(3):403-410.
293. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden T: **BLAST+: architecture and applications.** *BMC Bioinformatics* 2009, **10**(1):421.

294. Bairoch A, Apweiler R: **The SWISS-PROT Protein Sequence Data Bank and Its New Supplement TREMBL**. *Nucleic Acids Research* 1996, **24**(1):21-25.
295. Consortium TU: **Ongoing and future developments at the Universal Protein Resource**. *Nucleic Acids Research* 2011, **39**(suppl 1):D214-D219.
296. Powell S, Szklarczyk D, Trachana K, Roth A, Kuhn M, Muller J, Arnold R, Rattei T, Letunic I, Doerks T *et al*: **eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges**. *Nucleic Acids Research* 2012, **40**(D1):D284-D289.
297. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT *et al*: **Gene Ontology: tool for the unification of biology**. *Nature Genetics* 2000, **25**:25-29.
298. Petersen TN, Brunak S, von Heijne G, Nielsen H: **SignalP 4.0: discriminating signal peptides from transmembrane regions**. *Nat Meth* 2011, **8**(10):785-786.
299. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J *et al*: **The Pfam protein families database**. *Nucleic Acids Research* 2012, **40**(D1):D290-D301.
300. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL: **Predicting transmembrane protein topology with a hidden markov model: application to complete genomes**. *Journal of Molecular Biology* 2001, **305**(3):567-580.
301. Carr AC, Moore SD: **Robust quantification of polymerase chain reactions using global fitting**. *PLoS ONE* 2012, **7**(5):e37640.
302. Bustin S: **Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays**. *Journal of Molecular Endocrinology* 2000, **25**(2):169-193.
303. Bustin S: **Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems**. *Journal of Molecular Endocrinology* 2002, **29**(1):23-39.
304. Lillie R: **Histopathologic Technic and Practical Histochemistry, 3rd edition**. New York: McGraw-Hill Book Co; 1985.
305. Escoubas P, Diochot S, Corzo G: **Structure and pharmacology of spider venom neurotoxins**. *Biochimie* 2000, **82**(9-10):893-907.
306. Organ JA: **Studies on the life history of the salamander, *Plethodon welleri***. *Copeia* 1960, **1960**:34-40.

307. Picard AL: **Courtship in the zig-zag salamander (*Plethodon dorsalis*): insights into a transition in pheromone-delivery behavior.** *Ethology* 2005, **111**(9):799-809.
308. Highton R, Hastings AP, Palmer C, Watts R, Hass CA, Culver M, Arnold SJ: **Concurrent speciation in the eastern woodland salamanders (Genus *Plethodon*): DNA sequences of the complete albumin nuclear and partial mitochondrial 12s genes.** *Molecular Phylogenetics and Evolution* 2012, **63**(2):278-290.
309. Schägger H, von Jagow G: **Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa.** *Analytical Biochemistry* 1987, **166**:368-379.
310. Rozen S, Skaletsky HJ: **Primer3.** In. Code available at [http://www-genome.wi.mit.edu/genome\\_software/other/primer3.html](http://www-genome.wi.mit.edu/genome_software/other/primer3.html).; 1998.
311. North SJ, Hitchen PG, Haslam SM, Dell A: **Mass spectroscopy in the analysis of N-linked and O-linked glycans.** *Current Opinion in Structural Biology* 2009, **19**:498-506.
312. Evans JP: **The molecular basis of sperm-oocyte membrane interactions during mammalian fertilization.** *Human Reproduction Update* 2002, **8**(4):297-311.
313. Lande R: **Sexual dimorphism, sexual selection, and adaptation in polygenic characters.** *Evolution* 1980, **34**(2):292-305.
314. Cheverud JM: **Quantitative genetics and developmental constraints on evolution by selection.** *Journal of Theoretical Biology* 1984, **110**(2):155-171.
315. Arnold SJ: **The ultimate causes of phenotypic integration: lost in translation.** *Evolution* 2005, **59**(9):2059-2061.
316. Woodley SK: **Pheromone communication in amphibians.** *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology* 2010, **196**:713-727.
317. Sola RJ, Griebenow K: **Effects of glycosylation on the stability of protein pharmaceuticals.** *Journal of Pharmaceutical Science* 2009, **98**(4):1223-1245.
318. Welinder KG, Tams JW: **Effects of glycosylation on protein folding, stability and solubility. Studies of chemically modified or engineered plant and fungal peroxidases.** In: *Progress in Biotechnology*. Edited by Steffen B. Petersen BS, Sven P, vol. Volume 10: Elsevier; 1995: 205-210.

319. Arshad N, Ballal S, Visweswariah SS: **Site-specific N-Linked Glycosylation of Receptor Guanylyl Cyclase C Regulates Ligand Binding, Ligand-mediated Activation and Interaction with Vesicular Integral Membrane Protein 36, VIP36.** *Journal of Biological Chemistry* 2013, **288**(6):3907-3917.
320. Vigerust DJ, Shepherd VL: **Virus glycosylation: role in virulence and immune interactions.** *Trends in Microbiology* 2007, **15**(5):211-218.



## CURRICULUM VITA

NAME: Damien Beau Wilburn

ADDRESS: 2400 Mellwood Ave.  
Apt 802  
Louisville, KY 40206

DOB: Huntington, West Virginia – November 15, 1986

EDUCATION  
& TRAINING: B.S. Biology, Mathematics  
University of Louisville  
2005-2009

AWARDS: University of Louisville Department of Biochemistry: 1st place student poster, 2013. Dept. of Biochemistry Biennial Retreat

National Science Foundation Graduate Research Fellowship. "Mechanisms of Action and Evolution for a Hypervariable Vertebrate Courtship Pheromone". 2011-2014. Fellow.

National Science Foundation Graduate Research Fellowship. "Mechanisms of Action and Evolution for a Hypervariable Vertebrate Courtship Pheromone". 2010. Honorable mention.

University of Louisville: "Outstanding Senior Thesis Award." 2009. Highest rated senior thesis for 2009.

University of Louisville Department of Biology: "David Smith Memorial Award for Excellence in Undergraduate Research." 2009. Most enthusiastic undergraduate research scholar.

University of Louisville Department of Biology: "Harvey B. and Ethel W. Lovell Award for Excellence in Biology." 2009. Highest Biology GPA for the class of 2009.

PROFESSIONAL  
SOCIETIES:

Society for the Study of Evolution (2010-present)

Animal Behavior Society (2011-present)

American Society for Biochemistry and Molecular Biology  
(2013-present)

PUBLICATIONS:

Wilburn, D.B., Bowen, K.E., Doty, K.A., Arumugam, S., Lane, A.N., Feldhoff, P.W., and Feldhoff, R.C. (2014). Structural insights into the evolution of a sexy protein: novel topology and restricted backbone flexibility in a hypervariable vertebrate pheromone from the red-legged salamander. PLoS One. Accepted pending revision.

Chouinard, A.J., Wilburn, D.B., Houck, L.D., and Feldhoff, R.C. (2013). Individual variation in pheromone isoform ratios of the red-legged salamander, *Plethodon shermani*. In: Chemical Signals in Vertebrates XII; pp 99-115; East, M.L. and Dehnhard, M., Editors, Springer, New York, NY.

Wilburn, D.B., Bowen, K.E., Gregg, R.G., Cai, J., Feldhoff, P.W., Houck, L.D., and Feldhoff, R.C. (2012). Proteomic and UTR analysis of a rapidly evolving hypervariable family of vertebrate pheromones. *Evolution*, 66(7): 2227-2239.

Wilburn, D. Outcomes research in hydrocephalus treatment. (2010). In: Cases on Health Outcomes and Clinical Data Mining: Studies and Frameworks. Cerrito, P.B., Editor. IGI Publishing. 225-244.

Wilburn, D. Exploratory analyses and modeling of summary statistics and relative costs for infants afflicted with hydrocephalus. (2009). In: A Casebook of Pediatric Diseases. Cerrito, P.B., Editor. Bentham Science. 1-24.

NATIONAL  
MEETING

PRESENTATIONS: Wilburn, D.B., Bowen, K.E., Feldhoff, P.W., Houck, L.D., and Feldhoff, R.C. "Evolutionary decoupling of pheromone composition and mode of delivery in plethodontid salamanders." Animal Behavior Society Annual Meeting (Oral presentation; July 2013).

Wilburn, D.B. and Feldhoff, R.C. "From molecules to mating: gene regulation in an annual cycle of organogenesis and pheromone synthesis in the red-legged salamander." ASBMB Evolution and Core Processes in Gene Regulation (Poster; July 2013).

Wilburn, D.B., Bowen, K.E., Doty, K.A., Arumugam, S., Lane, A.N., Feldhoff, P.W., and Feldhoff, R.C. "How to make a sexy protein: structural evolution in a three-finger protein pheromone." Society for the Study of Evolution Annual Meeting (Oral presentation; June 2013)

Wilburn, D.B., Chouinard, A.J., Leichty, K.A., Eddy, S.L., Houck, L.D., Woodley, S.K., and Feldhoff, R.C. "Olfactory effects of isoform variability in female red-legged salamanders." Animal Behavior Society Annual Meeting (Oral presentation; June 2012).

Wilburn, D.B., Eddy, S.L., Chouinard, A.J., Feldhoff, R.C., and Houck, L.D. "Synergistic effects of a hypervariable courtship pheromone in the red-legged salamander." Animal Behavior Society Annual Meeting (Oral presentation; June 2011).

Wilburn, D.B., Bowen, K.E., Gregg, R.G., Cai, J., Feldhoff, P.W., Houck, L.D., & Feldhoff, R.C. "An inversion in molecular paradigm: using conserved UTRs to characterize a hypervariable family of vertebrate courtship pheromones." Society for the Study of Evolution Annual Meeting (Oral presentation; June 2010).

Wilburn, D. "Exploratory analyses and modeling for relative costs of infants with hydrocephalus." International Society of Pharmacoeconomics and Outcomes Research Annual Meeting (Poster; May 2009).

Wilburn, D.B., Bowen, K.E., Leichty, K.A., Feldhoff, P.W., Feldhoff, R.C. "Vertebrate Courtship Pheromones: Assembly PCR for Codon-Optimized Expression of a Highly Disulfide-Bonded 7 kDa Protein in *Pichia pastoris*." The Association of Biomolecular Resource Facilities Annual Meeting (Poster; February 2009).

#### INVITED

PRESENTATIONS: Wilburn, D.B. "From molecules to mating behavior: what can humans learn from salamanders?" Highlands Biological Station, Zahner Lecture Series (Invited speaker; July 2012).

GRANTS AND  
FELLOWSHIPS:

National Science Foundation Doctoral Dissertation Improvement Grant. "DISSERTATION RESEARCH: Pheromone signaling and molecular dissection of olfactory neurons in red-legged salamanders." \$17769, July 2013 – June 2014.

Highlands Biological Station Grant in Aid, Thelma Howell Memorial Scholarship. "Effects of a temperature-responsive RNA binding protein on plethodontid pheromone synthesis." \$1750, May 2012 – August 2012.

Highlands Biological Station Grant in Aid. "Evaluation of synergistic pheromone activities in the red-legged salamander." \$1400, August 2011.

National Science Foundation Graduate Research Fellowship. "Mechanisms of Action and Evolution for a Hypervariable Vertebrate Courtship Pheromone." \$126,000, June 2011 – May 2014.

Highlands Biological Station Grant in Aid, Bruce Family Scholarship in Herpetology. "Histological and Molecular Characterization of Developing Pheromone Gland." \$1750, May 2010 – September 2010.

University of Louisville Undergraduate Research Grant. "Genetic and Proteomic Analyses of Two Multigene Families of Courtship Pheromones in *Plethodon shermani*". \$3000, January 2009 – May 2009.

National Science Foundation Research Experience for Undergraduates. Supplement to IOS-0416834, "Collaborative Research: The Evolution of Pheromone Signals and Their Role in Behavioral Isolation". \$5775, May 2008 – August 2009.

University of Louisville Undergraduate Research Scholar Grant. "Genomic Sequencing of Plethodontid Modulating Factor in *P. shermani* salamanders". \$300, January 2008 – May 2008