Electronic Theses and Dissertations

5-2012

Co-option and adaptation of novel gene duplications for pheromone activity in a dusky salamander.

Kari Anne Leichty 1987-
University of Louisville

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

Recommended Citation

Leichty, Kari Anne 1987-, "Co-option and adaptation of novel gene duplications for pheromone activity in a dusky salamander." (2012). *Electronic Theses and Dissertations.* Paper 814.
https://doi.org/10.18297/etd/814

CO-OPTION AND ADAPTATION OF NOVEL GENE DUPLICATIONS FOR PHEROMONE ACTIVITY IN A DUSKY SALAMANDER.

By

Kari Anne Leichty
B.S., University of Louisville, 2010

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

Master of Science

Department of Biochemistry and Molecular Biology
University of Louisville
Louisville Kentucky

May 2012

CO-OPTION AND ADAPTATION OF NOVEL GENE DUPLICATIONS FOR
PHEROMONE ACTIVITY IN A DUSKY SALAMANDER.

By

Kari Anne Leichty
B.S., University of Louisville, 2010

A Thesis Approved on

April 18, 2012

by the following Thesis Committee:

_____

Richard Feldhoff, Thesis Director

_____

_____

Ron Gregg

_____

Pamela Feldhoff

DEDICATION

This thesis is dedicated to my parents

Dr. Gregory Leichty

and

Mrs. Kathleen Leichty

for all their invaluable support

ACKNOWLEDGEMENTS

ABSTRACT

CO-OPTION AND ADAPTATION OF NOVEL GENE DUPLICATIONS FOR PHEROMONE ACTIVITY IN A DUSKY SALAMANDER.

Kari A. Leichty

April 18, 2012

For more than a hundred million years, male plethodontid salamanders have utilized non-volatile, proteinaceous courtship pheromones to regulate female mating receptivity and promote mating success. These pheromones – which are delivered after courtship has commenced and do not act as chemical attractants – are delivered from a male's submandibular (mental) gland to the female salamander either by an olfactory mode or transdermally. For approximately 20 years, research on plethodontid courtship pheromones has focused on those involved in olfaction, using the red-legged salamander (*Plethodon shermani*) as a model. However, the olfactory system is a relatively recent adaptation (~20 million years ago) employed only by a single clade of large eastern *Plethodon* species. In contrast, for the majority of plethodontids (~350/380 species) the male scratches the back of the female with hypertrophied premaxillary teeth and rubs his gland over the abraded site to transdermally deliver pheromones. *Desmognathus ocoee* has served as the model for transdermal delivery, however less is known about this species. The purpose of this study was to evaluate the hypothesis that sexual selection has favored unique pheromone composition and structural motifs in different modes of pheromone delivery.

An RT-PCR library was constructed using total RNA extracted from two *D. ocoee* mental glands. A total of 235 clones were sequenced, increasing the number of known *D. ocoee* cDNAs by approximately 70%. A large number of cDNA sequences were predicted to encode hormone-like proteins. In addition, seven clones were found to contain the complete coding region of a known pheromone, sodefrin precursor-like factor (SPF). This protein was previously demonstrated to increase female receptivity in *D. ocoee*. Five of the SPF sequences encoded unique translations, suggesting extensive isoform variability. To further analyze this pheromone component, SPF was selectively amplified from the mental glands of five individual *D. ocoee* males and 150 full length SPF cDNAs were sequenced. Analysis of the sequences revealed transcription of multiple SPF isoforms in all males, which could be divided into six classes based on sequence similarity. Sequences from all six classes were not found in every male, however one isoform (SPF I01) was common to all males sampled. In addition, some "hybrid" transcripts were identified with sections that matched two or more classes. To evaluate if SPF transcripts matching two or more classes were a result of alternative splicing, genome walking techniques were used to identify exon/intron splice sites of SPF. It was determined that SPF has a 5 exon-4 intron structure and that the majority of the "hybrid" transcripts cannot be explained by alternative splicing.

Based on the presence of 16 cysteine residues in SPF as well as a conserved gene structure, it was hypothesized that SPF is related to the three fingered protein superfamily. Disulfide bonding studies were conducted by proteolytic digestion followed by mass spectrometry. Analysis showed that SPF has a different disulfide bonding pattern than established three fingered proteins.

Pheromones extracted from *D. ocoee* mental glands were purified by high performance liquid chromatography. Mass spectrometry revealed that the most highly expressed protein matches the SPF transcript common to all males (SPF I01). Analysis of mental gland extract from seven time points from May until October revealed that this isoform of SPF is highly expressed across the entire courtship season, composing approximately 30% of the pheromone mixture at all time points. These data suggest that sexual selection has favored stabilization of the *D. ocoee* SPF complex such that a single isoform is primarily expressed over a prolonged courtship season to influence reproductive behavior.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Pheromones are chemical signals that mediate interactions between individuals of the same species by inducing behavioral changes or altering developmental processes upon reception (*1*). Pheromones have been most studied in insects where their role as sexual attractants has been well established (*1-4*). For example, bombykol, the first pheromone to be chemically characterized, (as (E,Z)-10,12-hexadecadien-1-ol) is used by females to attract mates in the silkworm moth *Bombyx mori (5)*. Later studies in the bark beetle *Ips paraconfusus* demonstrated that many insect pheromones consist of complex blends of volatile compounds which act synergistically to produce a behavioral effect (*6*).

More recently, pheromones with a wide variety of behavioral effects have been identified in fish (*7, 8*), amphibians(*9-11*) and mammals (*12-14*). In particular, male mouse urine has been shown to be involved in mediating aggressive behaviors between males (*15*) as well as estrus synchronization and puberty acceleration in females (*12, 13*). In contrast, female urine has been shown to delay puberty in females that have been housed together at least 10 days (*16*). The components in the urine responsible for these behavioral and developmental changes include volatile compounds (such as 3,4-dehydro-*exo*-brevicomin, 2-heptanone and n-pentyl acetate) and the mouse major urinary proteins to which the volatile compounds bind. Other vertebrate pheromones include androstenone, which is produced by the domestic pig to induce a receptive mating

1

stance in females in estrus (*17*); 2-methylbut-2-enal, which guides newborn rabbits to their mother's nipples to obtain milk (*14*); and the proteinaceous pheromones of plethodontid salamanders which increase female receptivity to mate (*10*).

Despite the extensive research to date establishing functional roles for chemically-characterized pheromones in many systems, few studies have been performed on the evolution of pheromone signaling. One of the critical limitations for this field is that the majority of pheromones are volatile products of complex enzymatic cascades, which can be difficult to purify and are only indirectly influenced by selection. In contrast, the non-volatile proteinaceous pheromones of plethodontid salamanders can be easily purified by chromatographic methods (*18*) and, as direct gene products, can be readily sequenced and studied using established models of molecular evolution.

In Plethodontid salamanders courtship pheromones are delivered during a lengthy courtship ritual which ends with sperm transfer via a spermatophore (*19*). Male salamanders produce proteinaceous courtship pheromones in a specialized submandibular gland, known as the mental gland, which hypertrophies during the courtship season. Depending on the species, the pheromones are delivered to the female either by transdermal or olfactory mechanisms (*10*). Olfactory delivery, present in a monophyletic group of plethodontids termed the *Plethodon glutinosus* group, involves the male slapping the nares of the female with his hypertrophied mental gland to deliver pheromones (*10*) (Figure 1). This mechanism is considered to be the derived trait which evolved approximately 20 million years ago. After delivery to the female, pheromones diffuse via capillary action to the vomeronasal organ, bind to olfactory receptors, stimulate neurons, activate specific regions of the brain known to be involved in

2

pheromone response, and increase female receptivity to mate (*20, 21*). The majority of

research involving this mechanism of pheromone delivery has been conducted in

Plethodon shermani.

In behavioral experiments, whole pheromone extract from *Plethodon shermani*

mental glands was applied to the nares of females courting with de-glanded males.

Courting pairs in which the female received pheromone showed a decrease in courtship

time by 21%, indicating an increase in female receptivity (*10*). Characterization of the

whole pheromone extract revealed two major components: Plethodontid Receptivity

Factor (PRF) and Plethodontid Modulating Factor (PMF) (*10, 22*). PRF is a 22 kDa

protein related to the IL-6 family of cytokines (*10*). Purification of PRF from the *P.*

shermani pheromone extract followed by mass spectrometry (MS) revealed three major

isoforms of PRF termed B, C1, and C2, based on their elution times using strong anion-

exchange High Performance Liquid Chromatography (HPLC) (*23*). Together these three

isoforms of PRF compose approximately 1/3 of the pheromone mixture. When a mixture

of purified PRF isoforms was applied to courting females, the response was similar to

that of the whole pheromone extract (*10*). In addition, a single recombinant PRF isoform

(PRF C2) was demonstrated to have activity comparable to both the mixture of PRF

isoforms and the whole pheromone extract (*24*). In contrast, PMF is a 7 kDa protein

related to the family of three fingered proteins (TFPs) that includes xenoxins of *Xenopus*

skin and snake cytotoxins (*22*). A total of 28 different isoforms of PMF were purified

from *P. shermani* whole extract and 26 of the 28 were matched to cDNAs by at least one

mass spectral technique (*25*). Together these PMF isoforms compose approximately 50%

3
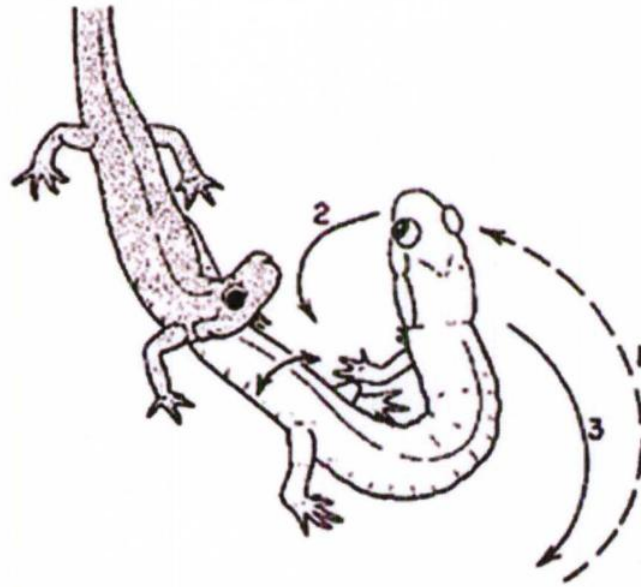
Figure 1: Illustration of a male *P. shermani* delivering pheromones to the nares of a female (*26*).

of the pheromone mixture. When a purified PMF fraction, consisting of approximately 30 isoforms of PMF, was applied to the nares of females during courtship trials, the net result was an increase (~16%) in courtship time (*27*). Based on its homology to snake venom cytotoxins, it was originally hypothesized that PMF may provide a mild sedative effect to keep the female engaged in the lengthy courtship process. However, more recent studies have shown that a complete mixture of all PMF isoforms decreases courtship time whereas a single recombinant PMF isoform had no effect (Wilburn et al., manuscript in preparation). Based on these findings, it is now hypothesized the PMF isoforms interact synergistically to reduce courtship time.

In contrast to *P. shermani* and other members of the *Plethodon glutinosus* group, the majority of plethodontid species deliver pheromones transdermally (Figure 2). This a novel type of vertebrate pheromone delivery, which involves the male abrading the dorsum of the female with his hypertrophied premaxillary teeth and delivering pheromones to the female by rubbing his gland over the abraded site (*28*). The pheromones diffuse into the bloodstream and act to influence female receptivity by unknown means (*29*). This mode of delivery is the ancestral mechanism, evolving approximately 100 million years ago. To date, the principal model for this system is *Desmognathus ocoee*.
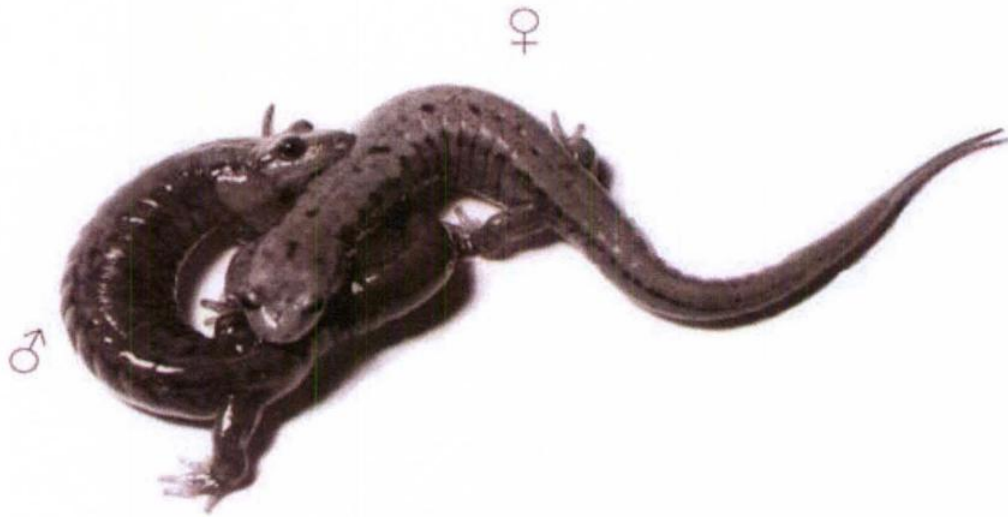
Figure 2: A male *D. ocoee* delivers transdermal courtship pheromones to a female (picture taken from (*29*)).

Preliminary analysis of pheromone extract from *D. ocoee* revealed a mixture much more complex than that of *P. shermani (29)*. Compared to two major components of the *P. shermani* pheromone mixture (PRF and PMF), analysis of *D. ocoee* pheromone extract by reverse phase HPLC (RP-HPLC) and SDS-PAGE revealed a mixture with more than 30 components. cDNA sequencing revealed that a 20 kDa protein, Sodefrin Precursor-like Factor (SPF), is highly transcribed in the *D. ocoee* mental gland (*30*). SPF has homology to the precursor of the newt pheromone sodefrin (*9*), as well as to the phospholipase A2 inhibitors (PLI), which contain two TFP domains (*31*). Like the PLIs, SPF is a highly disulfide bonded protein containing 16 cysteine residues predicted to form 8 disulfide bonds. Multiple isoforms of SPF had been identified in *D. ocoee* mental gland cDNA, yet prior to the current study, no proteins from *D. ocoee* whole pheromone extract had been characterized by MS or shown to match cDNA sequences.

In behavioral studies a high molecular weight fraction (20-25 kDa) of *D. ocoee* pheromone extract was applied to a treatment patch on the dorsum of females courting with de-glanded males, resulting in a decrease in courtship time by ~23% as compared to a saline control (*29*). Later SDS-PAGE analysis revealed that this fraction principally contained a 20 kDa protein which was assumed to be SPF. Thus, it was proposed that SPF in *D.ocoee* had a similar role to PRF in *P. shermani*.

More than twenty proteins have been identified in the *D. ocoee* pheromone mixture by SDS-PAGE, however they have not been identified by mass spectrometry. Unfortunately, this technique is limited by the quality of the reference database that is used. When a *D. ocoee* mental gland cDNA library was constructed in 2003, 57 unique protein translations were identified and added to an existing database of *P. shermani*

sequences. The quality of the database is one reason why the majority of the *D. ocoee* pheromone mixture is thought to remain uncharacterized.

Further analysis of mental gland cDNA from both *P. shermani* and *D. ocoee* shows some commonalities between the pheromone components. For example, approximately 5% of the cDNA sequences from the mental gland of *D. ocoee* were identified as PMF sequences (*32*). However, unlike *P. shermani,* which has more than 50 isoforms, only one single unique translation was identified. In addition, although SPF has not been identified in the pheromone extract of *P. shermani*, it has been detected in *P. shermani* mental gland cDNA, suggesting it may be produced at trace levels. Finally, although *D. ocoee* does not produce PRF, it is present in *Plethodon cinereus* (Feldhoff et al., unpublished data) which is a salamander utilizing transdermal pheromone delivery but more closely related to *P. shermani*. Because of this overlap in pheromone composition, it is hypothesized that olfactory pheromone delivery may have originally evolved as a secondary pathway that over time replaced transdermal pheromone delivery within the *Plethodon glutinosus* group (*33*).

An additional commonality between the major pheromones of *D. ocoee* and *P. shermani* is the shared relationship of SPF and PMF to the TFP superfamily. In particular, the phospholipase A2 inhibitors to which SPF is related have two conserved TFP domains, and the xenoxins and snake venom cytotoxins to which PMF is related have a single TFP domain. The TFP superfamily consists of proteins containing at least one of the namesake domains which have a three dimensional shape consisting of three adjacent finger-like loops extending from a hydrophobic core (Figure 3). The structure of the TFP domain includes five β-strands that encompass the three loops (*34*). These

8
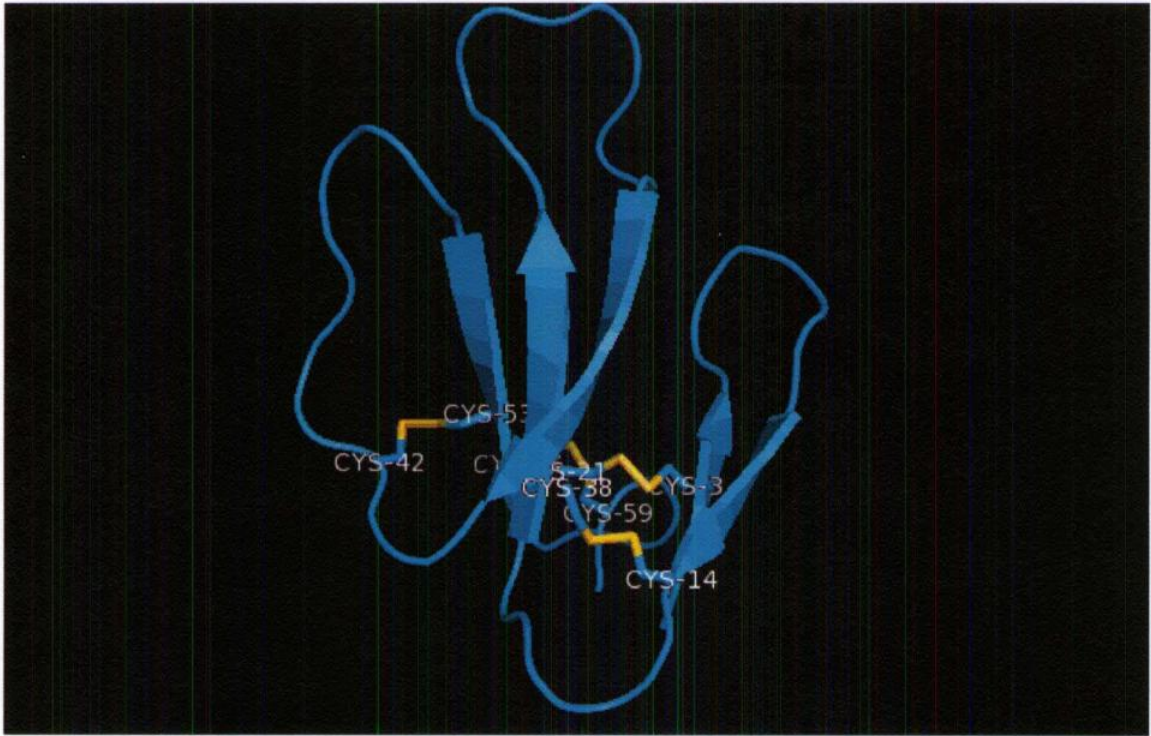
Figure 3: The tertiary structure of a representative TFP (Cardiotoxin IV) with three fingers stabilized by four disulfide bonds.

proteins have a wide variety of functions - from ligands, such as the snake cytotoxins, to receptors such as human CD59 (35) or the urokinase/plasminogen activator receptor (uPAR) (36). Within each TFP domain there is relatively little sequence conservation with the exception of 8 cysteine residues, which form disulfides at the base of each loop to stabilize the "fingers," and a C-terminal asparagine residue, which interacts through three hydrogen bonds with fingers 1 and 2 to lock the C-terminus in place (34). The common disulfide bonding pattern for these proteins includes bonds between the first and third, the second and fourth, the fifth and sixth and the seventh and eight highly conserved cysteines (i.e. 1-3, 2-4, 5-6, 7-8). Some members of the TFP superfamily have an additional disulfide bridge that occurs within one of the three fingers, but the overall structure remains the same.

At the genomic level, all characterized TFPs share a common gene structure composed of 3 exons and 2 introns (37-39). The location of the intron-exon junctions is conserved with the first intron occurring within the signal peptide and the second intron positioned between the third and fourth strictly conserved cysteine residues. In addition, all exon/intron junctions in TFPs are type I, meaning that they occur between the first and second bases of the codon (39).

PMF has the elements of the classic TFP primary structure, including the eight highly conserved cysteine residues and C-terminal asparagine. In addition, preliminary gene structure research has determined that PMF has an intron/exon structure similar to that of other TFPs including conserved type I exon/intron junctions (Wilburn et al., unpublished data). However, recent studies have demonstrated that PMF has a different disulfide bonding pattern than the canonical TFPs. Subsequent molecular modeling

has demonstrated that the change in disulfide bonding pattern of PMF is reflected in a unique structure making PMF distinct from members of the TFP family (Wilburn et al., unpublished data).

The disulfide bonding pattern and structure of SPF have not yet been determined. With 16 cysteine residues, there are a possible 2,027,025 disulfide bond combinations (*40*) (see Appendix 1). It is possible that the disulfide bonding pattern of SPF will cause it to adopt a structure similar to a double TFP, a structure similar to a double PMF or form a novel structure. Because disulfide bonds play an important role in the stabilization of proteins and in protein structure (*41, 42*) the determination of the SPF disulfide bonding pattern and structure is necessary to understand the SPF mechanism of action.

Compared to *P. shermani* and pheromone delivery by olfaction, there is considerably less known about *D. ocoee* and the transdermal mode of delivery. This is in part due to the various difficulties of working with *D. ocoee* including smaller gland size, less total RNA yield per gland, lower pheromone concentration from the extract of each gland, more components in the pheromone mixture, protein pheromone instability, and lower chromatographic resolution. Despite these experimental limitations, the aim of the current study is to further characterize the *D. ocoee* pheromone mixture by molecular and biochemical approaches. More extensive studies of the *D. ocoee* pheromones will complement and extend the studies of the *P. shermani* pheromone model system and aid in the explanation of the evolutionary, molecular and behavioral relationships between the two methods of pheromone delivery. Since the pheromone mixture directly influences reproductive success, the following research will test the hypothesis that

evolutionary pressures have directed unique structural motifs in plethodontid pheromones

to facilitate reproductive success in different modes of delivery.

CHAPTER II

MATERIALS AND METHODS

RT-PCR LIBRARY CONSTRUCTION

Collection of tissues and glandular secretions

Desmognathus ocoee males, identified by presence of premaxillary teeth, were

collected from Deep Gap in Macon County, North Carolina (35°02′20″N 083°33′08″W)

during the breeding season in August 2009, and at approximately 3 week intervals from

late May until early October 2010 (Table 1). Animals were anesthetized and mental

glands were surgically removed. Approximately 10 glands, to be used for analysis of

RNA, were placed in RNAlater (Invitrogen, Carlsbad, CA) at 4°C overnight, and then at -

20°C for long term storage. Pheromones from the remaining excised glands (~100) were

incubated in 0.8 mM acetylcholine in Amphibian Ringers Solution for one hour. The

solution centrifuged at 10,000 x g for 10 minutes at 4 °C. The supernatant was removed,

re-centrifuged at 10,000 x g for 10 minutes at 4 °C, and stored at -80°C. Methods and

animal care were approved by Oregon State University's Institutional Animal Care and

Use Committee (ACUP 3007 to L.D. Houck).

Synthesis of cDNA and cloning

Two mental glands were weighed, homogenized, and RNA was extracted using

the RNeasy Mini Kit (Qiagen, Valencia, CA) according to the manufacturer's protocols.

Total RNA was used as a template for single-stranded cDNA synthesis by

Time Point	Date	Number of Males
1	May 29-31	22
2	June 17	20
3	July 8	18
4	August 6	20
5	August 22	20
6	September 9-10	18
7	October 9	20

Table 1: *2010 Collection Time Points.* The dates and number of males collected for each of the seven collection time points.

Reverse Transcriptase Polymerase Chain Reaction (RT-PCR) with 5' RACE using the

Creator SMART cDNA Library Construction Kit (Clontech, Palo Alta, CA). Double-

stranded cDNA was synthesized by PCR using reaction protocol 1 (Table 2). The

primers used in this reaction were a 5' PCR primer and the 3' CDS III primer provided

by Clontech (Palo Alta CA). A small aliquot of double-stranded cDNA was analyzed by

a 1% agarose gel containing ethidium bromide and subjected to electrophoresis at 100

volts for 75 minutes. Double-stranded cDNA was cloned into the TOPO TA cloning

vector (Invitrogen, Carlsbad, CA) according to the manufacturer's protocol and

transformed into OneShot chemically competent *E. coli* cells (Invitrogen, Carlsbad, CA)

by heat shock at 42°C for 30 seconds. This was followed by incubation at 36°C for 1

hour while shaking in 250 μL of Super Optimal broth with Catabolite repression (SOC)

media (Invitrogen, Carlsbad, CA). Cells were plated on LB/Agar plates with kanamycin

to select for resistant clones and incubated overnight at 37°C. A total of 864 clones were

selected, each inoculated into 150 μL of Circlegrow broth (MP Biomedicals, Irvine, CA),

and shaken overnight at 37°C. Aliquots of 50 μL of each culture were stored at -80°C in

an equal volume of glycerol.

Sequencing of cDNA

Inserts from each clone were amplified by PCR using reaction protocol 1 (Table

2) and M13 Forward and M13 Reverse primers which were provided by Integrated DNA

Technologies (Coralville IA). Aliquots of each PCR reaction were analyzed using a 1%

agarose gel with ethidium bromide, subjected to electrophoresis at 100 volts for 75

minutes. All PCR products that were larger than 600 base pairs were purified using the

Reaction Protocol	PCR Recipe	Thermocycler Program
1	2 μL 10X PCR Buffer 3.2 μL dNTP mix (1.25 mM) 0.2 μL Taq Polymerase 0.2 μL 20 μM forward primer 0.2 μL of 20 μM reverse primer 1μL 0.1 U/μL Vent Polymerase 1 μL ss-cDNA 12.2 μL H$_2$O	1) 95°C for 2 minutes 2) 95°C for 15 seconds 3) 58 °C for 1 minute 4) 68°C for 6 minutes 5) Go to Step 2, 39 times 6) 4 degrees
2	2 μL of Accuprime Buffer II, 1 μL of circularized DNA, 0.4 μL 10 mM Forward primer 0.4 μL 10 mM Reverse primer 0.3 μL of Accuprime Taq DNA polymerase 15.9 μL of molecular biology water	1) 95°C for 2 minutes 2) 95°C for 15 seconds 3) 58 °C for 1 minute 4) 68°C for 6 minutes 5) Go to Step 2, 35 times 6) 4 degrees

Table 2: PCR reaction Protocols. Primers were supplied by either Clontech (CDS III and

5' PCR primer; Palo Alta CA) or Integrated DNA technologies (all other primers;

Coralville IA). Accuprime Buffer II and Accuprime Taq DNA polymerase were

purchased from Invitrogen (Carlsbad CA). All other supplies were obtained from New

England Biolabs (Ipswich MA).

QIAquick PCR Purification Kit (Qiagen, Valencia, CA). Two 5 µL aliquots of each

purified PCR product were diluted with 5.4 µL water and 1.6 µL of either M13 Forward

or M13 Reverse primer (Integrated DNA Technologies, Coralville IA). Sequencing was

performed by the University of Louisville DNACore Lab.

Analysis of sequences

Sequences were assembled into contigs by SeqMan Pro (Lasergene Version 7.1;

DNASTAR, Madison, WI) and low quality sequences were removed. The consensus

sequence of each contig was compared to known sequences in GenBank using translated

nucleotide BLAST searches (www.ncbi.nlm.nih.gov/BLAST). Editseq (Lasergene

Version 7.1; DNASTAR, Madison, WI) was used to identify and translate putative open

reading frames for each sequence. Translations were checked manually for repeats and

analyzed for a signal peptide using SignalP 4.0 Server

(http://www.cbs.dtu.dk/services/SignalP/) (*43*). To prepare a database for proteomic

analysis, the predicted signal peptide sequences were deleted and the data configured into

FASTA format to compare to proteins analyzed by the University of Louisville

Biomolecular Mass Spectrometry Core Laboratory.

Amplification of SPF cDNA from five individual males

In order to examine transcription of SPF isoforms in individual male salamanders,

five mental glands were separately homogenized and total RNA extracted using the

RNeasy mini kit (Qiagen, Valencia, CA) according to manufacturer's protocol. Single-

stranded cDNA was synthesized for each male using the Creator SMART cDNA Library

Construction Kit (Clontech, Palo Alta, CA). Double-stranded cDNA was synthesized by

PCR using reaction protocol 1 (Table 2) and the primers included in this reaction were 3'

CDS III (Clontech, Palo Alta, CA) and SPF 5' UTR primer (Integrated DNA Technologies, Coralville IA). Based on previous cDNA sequencing, SPF 5'UTR Forward Primer was designed to anneal to a conserved region of the SPF 5' UTR, permitting full-length amplification of the SPF coding region. Amplified cDNA from each reaction was ligated to the TOPO TA cloning vector, (Invitrogen, Carlsbad, CA), and the cloned DNA was used to transform OneShot chemically competent *E. coli* cells (Invitrogen, Carlsbad, CA) as previously described.

Approximately 30 clones for each transformation (each corresponding to an individual male) were selected and transformed DNA was amplified in a 20 µL PCR reaction using reaction protocol 2 (Table 2). A 5 µL aliquot of each PCR reaction was visualized using a 1% agarose gel with ethidium bromide subjected to electrophoresis at 100 volts for 75 minutes. For each positive reaction, 1 µL of PCR product was diluted in 9.4 µL of water and 1.6 µL of 20 mM SPF 5' UTR Forward Primer and sequenced by the University of Louisville DNACore Lab.

Analysis of Sequences

Sequences were assembled into contigs by Seqman Pro (Lasergene Version 7.1; DNASTAR, Madison, WI) and low quality or incomplete sequences were removed. For each sequence, open reading frames were identified and translated using EditSeq (Lasergene Version 7.1; DNASTAR, Madison, WI). Sequences were checked manually for repeats and unique sequences were compared using MegAlign (Lasergene Version 7.1; DNASTAR, Madison, WI).

PURIFICATION OF PHEROMONES AND MASS SPECTRAL ANALYSIS

Pheromone purification

      Desmognathus ocoee proteinaceous pheromones were purified by three successive rounds of HPLC. All separations were performed on a Waters 2695 Alliance HPLC System (Waters Corporation, Milford, MA) equipped with a Waters 2487 dual wavelength absorbance detector and Waters Empower software. The eluate was monitored at 220 nm. Initially, the whole pheromone extract was separated by size exclusion chromatography using a G75 superfine column (Pharmacia, Piscataway, NJ), eluted at ~20 mL/hour with 0.5 X phosphate-buffered saline (PBS). The eluate was pooled into three fractions based on elution time (See Table 6). Pooled fractions were subjected to strong anion exchange chromatography using a Mono Q column (Pharmacia, Piscataway, NJ), eluted with a sodium chloride mixed gradient (5mM NaCl/min for 30 min, 10mM NaCl/min for 15 min, 20mM NaCl/min for 10min) in 50mM Tris-HCl, pH 8. Finally, strong anion exchange fractions were collected and further separated by reverse phase HPLC (RP-HPLC) using a C18 column (0.5 x 15 cm, Grace Davidson Discovery Sciences, Deerfield, IL) eluted with an acetonitrile (ACN) gradient (0-70% ACN at 1% ACN/minute) in 0.1% trifluoracetic acid (TFA). Single protein peaks were collected by hand and subjected to mass spectral analysis.

Mass spectral analysis

      Purified *D. ocoee* mental gland proteins were analyzed by Electrospray Ionization Mass Spectroscopy (ESI-MS) and/or Matrix Assisted Laser Desorption Ionization-Time-Of-Flight (MALDI-TOF) or Liquid Chromatography Tandem Mass Spectrometry (LC MS/MS) depending on the amount of protein available. All mass spectroscopy was

performed by the Biomolecular Mass Spectrometry Core Laboratory at the University of Louisville.

In preparation for ESI-MS, samples collected from the reverse-phase HPLC separations were lyophilized by SpeedVac (Thermo Scientific, Wilmington, DE). If sufficient protein was available, half of the sample was reduced with 3 μL of 45 mM dithiothreitol (DTT) incubated at 60°C for 30 minutes followed by alkylation with 3 μL of 120 mM iodoacetamide incubated at 45°C for 30 minutes. Both the intact and reduced/alkylated protein samples were then purified using C18 resin ZipTip (Millipore, Billerica, MA) according to the protocols of the manufacturer.

For MALDI-TOF or LC MS/MS, protein samples collected from RP-HPLC separations were lyophilized by SpeedVac. Proteins were solubilized in 10 μL of 6M Urea/0.1M NH_4HCO_3, reduced by the addition of 2 μL 45 mM DTT with incubation at 60°C for 30 minutes, and alkylated by addition of 2 μL of 120 mM iodoacetamide with incubation at 45°C for 30 minutes. The protein solution was diluted with 6 μL of 0.1M NH_4HCO_3 and subjected to protease digestion at 37°C overnight with 10 μL of 0.025 μg/μL trypsin. Peptide fragments were purified by ZipTip (C18; Millipore Millipore, Billerica, MA) according to the manufacturer's protocols prior to analysis by MALDI-TOF. Fragment masses obtained were compared to a database of translated cDNA with predicted trypsin cleavage sites for *D. ocoee* and *P. shermani*.

INVERSE PCR AND GENOME WALKING

Purification of genomic DNA

A single *D. ocoee* tail tip was cut into small pieces and approximately 15 mg of tissue was added to 975 μL proteinase K buffer (100 mM Tris-HCl pH 8, 100 mM EDTA, 250 mM NaCl, 0.5% SDS) with 25 μL of 20 μg/μL of proteinase K. The tissue was digested overnight at 55°C. Following digestion, the tube was inverted 5 times to dissociate remaining tissue, and the sample was centrifuged at 10,000 x *g* for 15 minutes. The supernatant was removed, re-centrifuged at 10,000 x *g* for 15 minutes, supernatant collected, and filtered. To extract DNA, an equal volume of phenol/chloroform/isoamyl alcohol (PCI, ratio 125:24:1; Invitrogen, Carlsbad, CA) was added, the solution was mixed by inversion, and centrifuged for 30 minutes at 10,000 x *g*. The aqueous (top) layer was carefully transferred to a clean 1.7 mL tube. PCI extraction was repeated to further remove contaminants. 3M sodium acetate was added to a final concentration of 0.3 M and the DNA was precipitated by addition of 3 volumes of 95% ethanol incubated overnight at -80°C. The solution was subsequently centrifuged at 10,000 x *g* for 30 minutes at 0°C, supernatant was discarded, and 1 mL of 70% ethanol was added. The sample re-centrifuged at 10,000 x *g* for 30 minutes, and then the supernatant was removed. The DNA pellet was allowed to air-dry prior to being resolublized in 20 μL of molecular biology-grade water. DNA concentration was measured by NanoDrop (Thermo Scientific, Wilmington, DE) and standardized to 200 ng/μL.

Amplification of genomic DNA

To ensure that a sufficient amount of DNA from a single animal was available for all genome walking experiments, *D. ocoee* genomic DNA was amplified using the

Illustra GenomiPhi V2 DNA amplification Kit (GE Healthcare, Life Sciences,

Piscataway, NJ) using the manufacturer's protocols. Amplified genomic DNA

concentration was measured by NanoDrop and standardized to 200 ng/μL.

Inverse PCR

Inverse PCR (iPCR) was used to determine a single intron/exon boundary of SPF.

Specifically, circular DNA fragments were formed by digestion of *D. ocoee* genomic

DNA with restriction enzymes (*Tsp*509I, *Nla*III, or *HpyCH*4IV; New England BioLabs,

Ipswich, MA) according to manufacturer's protocols at 37°C overnight. Following

digestion, the products were purified using the QIAquick PCR Purification Kit (Qiagen,

Valencia, CA). To promote self-ligation, the digested products were diluted with 420 μL

H_2O, 50 μL 10X Ligase Buffer (New England Biolabs, Ipswich, MA), and 1.25 μL 400

U/μL T4 DNA Ligase (New England Biolabs, Ipswich, MA) was added prior to

overnight incubation at 16°C. Because the 3' UTR of genes generally lack introns, the 3'

UTR of SPF was chosen as a target for inverse PCR. Four reverse primers and three

forward primers (Integrated DNA technologies, Coralville, IA), were designed to anneal

to a region from base pairs 599 to 748 of SPF (Table 3).

The circularized DNA was amplified in a 20 μL PCR reaction using reaction

protocol 2 (Table 2). PCR products were subjected to electrophoresis at 100 volts for 75

minutes and visualized in a 1% agarose gel with ethidium bromide. Amplified DNA

from the first round of PCR was used as a template for a second round of PCR using

nested primers to ensure specific amplification of SPF. The PCR reaction and

thermocycler program was identical to the first reaction described above with 1 μL of the

22

Primer	Sequence
SPF 3' UTR 618-599 R'	GCTGTCATGCGCAGATCTTT
SPF 3' UTR 646-625 R'	GATGTTAATTGGGTCATGTTGG
SPF 3' UTR 673-651 R'	AAGATTTCTTGAGTTCTGCTTGC
SPF 3' UTR 695-673 R'	TGCAATGATCATGTGAATAACAA
SPF 3' UTR 700-719 F'	CGCACCCAGATGAGTAGAAA
SPF 3' UTR 724-743 F'	ATGTTCCCAGATACGCTGCT
SPF 3' UTR 729-748 F'	CCCAGATACGCTGCTTGTTT

Table 3: Primers used for inverse PCR. Primers were designed to anneal to the 3' UTR

of the most highly transcribed SPF sequence.

original PCR being used as the template DNA and nested primers used in place of the original gene specific primer.

Cloning and sequencing of inverse PCR products

Inverse PCR products were subjected to electrophoresis at 100 volts for 75 minutes and visualized in a 1% agarose gel with ethidium bromide. Two PCR products, with well- defined bands, were selected for cloning. An aliquot of DNA from the selected PCR reactions was ligated to the TOPO TA cloning vector, (Invitrogen, Carlsbad, CA) according to manufacturer's protocols and used to transform OneShot chemically competent *E. coli* cells (Invitrogen, Carlsbad, CA) as previously described.

Cloned DNA was amplified in 20 µL PCR reactions using reaction protocol 2 (Table 2). Small aliquots of each PCR reaction were analyzed using a 1% agarose gel with ethidium bromide subjected to electrophoresis at 100 volts for 75 minutes. Two 1.0 µL aliquots of each PCR product were diluted with 9.4 µL of water and 1.6 µL of either M13 forward or M13 reverse primer (Integrated DNA Technologies, Coralville IA). Sequencing was performed at the University of Louisville DNACore Lab.

Analysis of sequences

Sequences were assembled by SeqMan Pro (Lasergene Version 7.1; DNASTAR, Madison, WI) and aligned relative to the 3' end of SPF; an intron splice site was determined by ascertaining the nucleotide position where the inverse PCR sequence failed to match SPF cDNA sequence.

Genome walking using semi-degenerate primers

After the first intron-exon boundary was determined by inverse PCR, additional boundaries were determined by genome walking using semi-degenerate primers as

24

described by Guo and Xiong (*44*). Specifically, the four semi-degenerate primers described by Guo and Xiong and gene specific primers designed to anneal to SPF close to known intron-exon boundaries were synthesized (Integrated DNA Technologies, Coralville, IA). For each combination of gene-specific primers and semi-degenerate primers, 20 µL PCR reactions were prepared containing 2 µL of Accuprime Buffer II (Invitrogen, Carlsbad, CA), 1 µL of amplified *D. ocoee* genomic DNA at 200 ng/µL, 0.4 µL each of 20 mM gene specific primer and 10 mM semi-degenerate primer (Integrated DNA Technologies, Coralville IA), 0.3 µL of Accuprime Taq (Invitrogen, Carlsbad, CA) and 15.9 µL of molecular biology grade water with a thermocycler program of 94°C for two minutes, 25 cycles of 94°C for 30 seconds, 60°C minus 0.5°C per cycle for 30 seconds, 68°C for ten minutes, followed by 35 cycles of 94°C for 30 seconds, 50°C for 30 seconds and 68°C for 10 minutes. PCR products were visualized using a 1% agarose gel with ethidium bromide subjected to electrophoresis at 100 volts for 75 minutes.

To ensure that products were gene specific, nested PCR was conducted using DNA from the first PCR as template DNA and nested gene specific primers. The PCR reaction conditions and thermocycler program were the same as the first round with genomic DNA being replaced with 1 µL of PCR product from the first reaction and the gene specific primer being replaced with a nested primer. PCR products were visualized using a 1% agarose gel with ethidium bromide subjected to electrophoresis at 100 volts for 75 minutes.

Cloning and sequencing of degenerate primer genome walking PCR products

PCR products with well-defined bands were selected for cloning. Specifically, an aliquot of DNA from the selected PCR reactions was ligated to the TOPO TA cloning

vector, (Invitrogen, Carlsbad, CA) according to manufacturer's protocols and subsequently used to transform OneShot chemically competent *E. coli* cells (Invitrogen, Carlsbad, CA) as previously described.

Cloned DNA was amplified in 20 μL PCR reactions using reaction protocol 1 (Table 2) using M13 forward and M13 reverse primers (Integrated DNA technologies, Coralville IA). The products were visualized by agarose gel electrophoresis. For each positive reaction, 1 μL aliquots of PCR product were diluted in 9.4 μL of water and 1.6 μL of either M13 forward or reverse primer (Integrated DNA Technologies, Coralville IA). Sequencing was performed by the University of Louisville DNACore Lab. Sequences were assembled by SeqMan Pro (Lasergene Version 7.1; DNASTAR, Madison, WI) and compared to SPF cDNA sequences to determine intron-exon splice sites.

Confirmation of intron-exon boundaries

After defining intron-exon boundaries by genome walking with semi-degenerate primers it was necessary to verify that no additional introns existed in what was thought to be a single exon. Primers were designed (Integrated DNA Technologies, Coralville, IA) to anneal to known intron sequences and were used to amplify across exons, PCR reactions (20 μL) were prepared using reaction protocol 2 (Table2) and PCR products were visualized by agarose gel electrophoresis. For each PCR product with a single band, 1 μL of PCR product was diluted in 9.4 μL of molecular biology water with 1.6 μL of gene specific primer, and sequenced by the DNACore Laboratory at the University of Louisville.

Alternatively, if none of the PCR reactions appeared to contain single product, the reaction with the most defined bands was selected for cloning as previously described. Clones were selected and DNA amplified by PCR using reaction protocol 1. A 5 μL aliquot of each PCR reaction was sequenced by the University of Louisville DNACore Lab.

SPF STRUCTURAL CHARACTERIZATION

Purification of SPF

D. ocoee whole extract was separated by strong anion exchange chromatography using a MonoQ column (Pharmacia, Piscataway, NJ) with a NaCl gradient (5mM NaCl/min for 30 min, 10mM NaCl/min for 15 min, 20mM NaCl/min for 10min) in 50mM Tris. Fractions eluting between 50-52 minutes from the Mono Q column were loaded onto a Vydac 4.7 x 150 mm C18 reverse phase HPLC column (Grace Davidson Discovery Sciences, Deerfield, IL) and eluted with an acetonitrile gradient (0-70% ACN at 1% ACN/minute) in 0.1% TFA. A single peak that had been identified as SPF was collected from ~36.4 minutes to ~36.6 minutes, and stored at -20°C.

Determination of SPF disulfide bonding pattern

Aliquots of 3-5 μg of SPF protein were lyophilized, solubilized in 5 μL of 4M Urea/0.1M NH_4HCO_3 and incubated at room temperature for 30 minutes. Following incubation, 0.5 μg aliquots of a single protease (trypsin, chymotrypsin, LysC, or AspN), or a combination of two proteases (trypsin, chymotrypsin, LysC, ArgC, or AspN) were added. The sample was diluted with the appropriate volume of 0.1M NH_4HCO_3 to achieve a final urea concentration of 2M. Following overnight incubation at 36°C,

peptide fragments were purified by ZipTip (Millipore, Billerica, MA) according to the manufacturer's protocols and subjected to LC-MS/MS by the University of Louisville Biomolecular Mass Spectrometry Core Laboratory.

Analysis of MS/MS Data

Digestion of SPF with proteases without reduction and alkylation should result in the production of peptides held together by disulfide bonds. Unfortunately the available software assumes only linear fragments. Therefore, a program was written in Python (Wilburn unpublished) to create a list of all possible theoretical monoisotopic masses, both linear or as fragments held together by disulfide bonds, for each protease digest. Parent masses and fragmentation data for MS/MS analysis of SPF protease digests were extracted from Thermo Scientific raw data files, fragmentation data consolidated for datasets with similar parent masses (0.4 Da threshold), and automatically compared to the list of theoretical masses. Experimental masses were filtered to create a list of those which matched a theoretical mass within 1 Da. To confirm mass matches, further analysis of the MS/MS fragmentation data of each match was conducted using Protein Prospector- MS Product online software (http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msproduct).

CHAPTER III

RESULTS

SEQUENCING OF cDNA

Following screening by agarose gel electrophoresis, a total of 244 clones from the

D. ocoee cDNA library were sequenced. A total of 235 high-quality sequences were

assembled into contiguous sequences (contigs) by Seqman (Lasergene Version 7.1;

DNASTAR, Madison, WI). A total of 153 sequences were combined into 25 multiple

sequence contigs. The remaining 82 sequences all formed single sequence contigs,

suggesting that a large number of *D. ocoee* mental gland proteins are transcribed, each at

low rates.

Based on translated BLAST searches, the total number of sequences could be

grouped into five categories: those with similarity to peptide hormones, those responsible

for normal cellular homeostasis, known pheromone components, those similar to known

pheromones, and unidentified components with no known function.

Category	Number of Contigs	Number of Sequences	Percentage of total sequences
Unidentified	58	76	32.3%
Hormone-like	14	74	31.5%
Cell Maintenance	29	58	24.7%
Known Pheromone	3	23	9.8%
Pheromone-like	3	4	1.7%

Table 4. Categorization of *D. ocoee* cDNA sequences based on function.

The unidentified sequences were the largest grouping, comprising 32% of all

sequences. The sequences in this group matched either hypothetical proteins or had no

definitive matches in BLAST. With only a few exceptions, these transcripts belonged to

single sequence contigs, suggesting that they are transcribed at low rates.

The second largest group, composed of approximately 31% of the sequences,

showed similarity to peptide hormones and are candidate pheromones in *D. ocoee*. In

this category, it is notable that many sequences were similar to the precursor of

vasoactive intestinal polypeptide (VIP). The precursor is cleaved at four different

locations (typically dibasic sites) to generate three linking segments, a 28 amino acid

native VIP, and a 27 amino acid fragment called peptide histidine isoleucine (PHI) (*45-*

47).

In the *D. ocoee* mental gland sequences, there are four different contigs

containing 34 sequences that show similarity to VIP. One contig exclusively contains

sequence similar to the 3' untranslated region (UTR) of VIP precursor mRNA. The other

three contigs contained VIP precursor-like open reading frames (ORFs), which were

translated and compared to prepro-VIP from other species (Figure 4). One *D. ocoee* VIP

precursor-like protein was 131 amino acids in length. It contained two native VIP-like

regions that aligned to the PHI and VIP regions of the precursors from other species. In

contrast, the VIP precursor-like proteins from the other two contigs were shorter and each

contained only one native VIP-like region. This region could be aligned to either the PHI

or VIP region of VIP precursors from other species depending on which aligning

algorithm was used.

31

A

D ocoee Prepro-VIP Consensus 1

```
MQQMSIYLFLLGSISLYLWAQASPTSEEDSAWSNNVLEEE  40
MPTIKRHVDGLFTHSYSRYRNKMAILKYLKSVLPGTKSNN  80
YLEEEMPTIKRHVDGLFTHSYSRYRNKMAILKYLKSVLPG 120
IKSEYKKPKA.                              131
```

D ocoee Prepro-VIP Consensus 2

```
MQHKRNLLFLSCVIMMTFLNLSGHELPTLGIYSAMSNRAL  40
DEEIPSVKRHADALFTAGYSKIRYNAAVRKYFESVLTRPK  80
SIYPKKKVKPTGINREAFPGYRNPMY.              107
```

D ocoee Prepro-VIP Consensus 3

```
MQHKSSFPFLLCFFILSLLYLQAQALLALGTYSAMSENAL  40
NEEITPVKRHSDAEFTDGYTRQLAKEQLEKYLSDVLGAKI  80
RLP.                                      84
```

B

D ocoee VIP-like Peptides

```
HVDGLFTHSYSRYRNKMAILKYLKSVLPGTKSNNVLEEEMPTI
HVDGLFTHSYSRYRNKMAILKYLKSVLPGTKSEV
HADALFTAGYSKIRYNAAVRKYFESVLTRPKSIVP
HSDAEFTDGYTRQLAKEQLEKYLSDVLGAKIRLP
```

PHI (Human) HADGVFTSDFSRLLGQLSAKKYLESLI

PHI (Rat, Mouse) HADGVFTSDYSRLLGQISAKKYLESLI

PHI (Chicken) HADGIFTSVYSHLWAKLAVKRYLHSLI

PHI (Fish, Frog) HADGLFTSGYSRLLGQLSARRYLESLI

VIP (Human, Rat, Mouse) HSDAVFTDNYTRLRKQMAVKKYLNSILN

VIP (Chicken, Frog) HSDAVFTDNYSRFRKQMAVKKYLNSVLT

VIP (Fish) HSDAIFTDNYSRFRKQMAVKKYLNSVLT

Figure 4. Amino acid sequence of Vasoactive Intestinal Peptide. A) Translations of the consensus sequence for each *D. ocoee* VIP-like contig with the signal peptide indicated by a light purple box and dibasic cleavage sites underlined in red. The expected sequence of native VIP is underlined in purple. Arrows indicate cleavage sites based on the expected length of VIP (28 amino acids) or PHI (27 amino acids). B) Comparison of *D. ocoee* VIP-like peptides and VIP and PHI from other species. Conserved residues of PHI and VIP are indicated by black shading and light blue shading respectively.

A second set of highly transcribed hormone-like components, composing approximately 10% of all transcripts, were the glucagon-like sequences. Glucagon is synthesized as a larger precursor with two flanking sets of dibasic residues as sites of cleavage for the biologically active peptide. Comparison of the predicted functional peptides in *D. ocoee* and other glucagon-like peptides showed complete conservation in eight residues (Figure 5).

A final subset of highly transcribed hormone-like components included three contigs with homology to the insulin-like peptide superfamily. Members of this family are expressed as prepro-precursors and are processed to contain two chains-the B chain and the A chain-linked by intra- and intermolecular disulfide bonds. Within this superfamily there are three more specific subdivisions: insulin, insulin-like growth factors and relaxins (*48, 49*). In *D. ocoee*, one of the three contigs with sequence similarity to insulin-like-peptides had greatest similarity to insulin and two had greatest similarity to relaxin. All three contigs contained open reading frames, which were aligned with insulin or relaxin prepro-peptides from other species.

The *D. ocoee* insulin-like peptide has large amounts of similarity to pre-proinsulin from other species (Figure 6). There is conservation between regions of the B chain and A chain, whereas the signal peptide and the linking C peptide have low levels of similarity. Interestingly, the *D. ocoee* insulin-like peptide has the first double basic cleavage site at the C-terminal end of the B chain, however the typical dibasic site at the N-terminal end of the A chain had an amino acid substitution.

Figure 5. Alignment of glucagon-like peptides. Black shading indicates residues that match the human consensus sequence. Arrows indicated amino acids determined to be important in glucagon binding to receptors (black arrows) or those important in activation of biological activity (red arrows) (*50-53*).
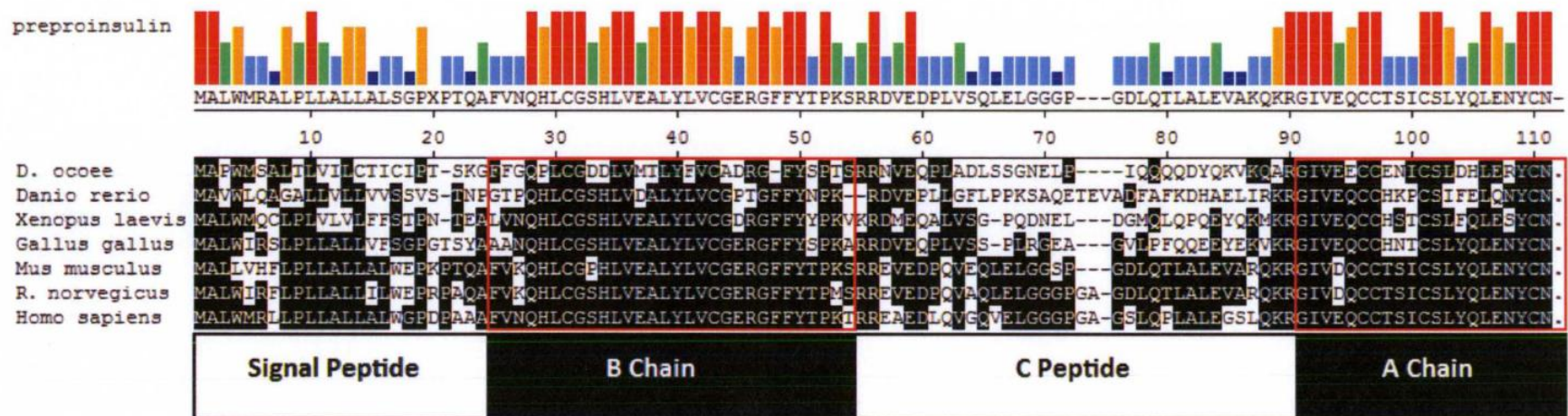
Figure 6. Alignment of Preproinsulin-like peptide from *D. ocoee* with preproinsulin sequences from *Danio rerio* (Genbank accession no. NM_131056) *Xenopus laevis* (NM_001085882), *Gallus gallus* (X58993), *Mus musculus* (NM_008386), *Rattus norvegicus* (NM_019130), and *Homo sapiens* (NM_000207). Consensus strength is indicated by bar height and color warmth. Shaded residues match the consensus.

The alignment of preprorelaxin-like ORFs from *D. ocoee* with preprorelaxin proteins from other species shows that, like the insulin related protein, the greatest amount of conservation was found in the B and A chains. Specifically, in *D. ocoee* ORFs the B chain had two cysteine residues, a glycine residue and an arginine residue, with the A chain containing four cysteine residues, a leucine residue and a lysine residue that are highly conserved among relaxins from other species (*54, 55*). In contrast, the C peptide showed very low amounts of conservation both in length and amino acid sequence. The *D. ocoee* relaxin-like ORFs had predicted C peptides that were significantly shorter than those found in other species. Specifically, the predicted C peptide of one of the *D. ocoee* relaxin-like ORFs was only 7 amino acids while relaxins from other species tend to be 60 amino acids or longer.

The third grouping of sequences contains predicted proteins that have roles in normal cellular function. These sequences included ribosomal and mitochondrial proteins, myosin and actin. Nearly half of the sequences in this grouping (27/58) encoded protease inhibitors, which may function to protect pheromones from degradation.

Finally, approximately 11% of ORFs coded for known pheromone components or proteins similar to known pheromones. A total of four PMF sequences were found and all had the same translation. The amino acid translation matched a previously identified sequence (GenBank accession number DQ882565), supporting the suggestion that only one PMF sequence is transcribed in *D. ocoee*. In addition, a single sequence was found that coded for a protein similar to PMF. Like PMF it is a small protein composed of only 71 amino acid residues that contains cysteine residues typical

D ocoee Relaxin-like 1

```
MMASLYLLIGLSFLAANLNVKGDDSMVNKCGKELMREMSA  40
YCDAQHPQARAAEKRSPSLDKRNDETSEVFLALMDRCCES  80
GCAKDELLPYCP.                             93
```

D ocoee Relaxin-like 2

```
MRASLLCFFALCLLIEHSQATAVEKFCGKDLLRAVSYFCG  40
ASRWRRHLAGPAQDRLKERSLFYSGGNRDAFDSPNINFQE  80
AEGNEDLKDTESGGSLRKRREAKDIDNRLIAYCCQVECSK  120
ESFAKFCSN.                                130
```

Figure 7. Amino acid sequence of *D. ocoee* relaxin. Translations of the consensus

sequence ORFs of each *D. ocoee* relaxin-like contig are shown with predicted signal

peptide indicated by a blue box. Dibasic residues for the cleavage of the C peptide are

indicated by red underlining and residues conserved with relaxin from other species are

highlighted in black.

of a three-fingered protein. However, NCBI BLAST analysis suggests this protein is more similar to short chain snake toxins and cardiotoxins than it is to PMF.

A total of 19 sequences in the cDNA library were SPF sequences yielding five unique translations, raising the total number of unique sequences to fourteen. This suggests hypervariability in SPF sequences similar to what is seen in PMF. All predicted SPF sequences suggested an overall structure consisting of two TFP-like domains of 8 cysteine residues, connected by a linker region containing 3 conserved proline residues. In addition to the SPF sequences, there were an additional 3 SPF-like sequences, but these were more similar to sodefrin and PLI. These SPF-like sequences may represent part of the evolutionary transition from sodefrin and/or PLI to SPF.

To assess the degree of SPF sequence variability and individual differences in SPF transcription, SPF cDNA was PCR amplified from five individual male mental glands. When the SPF sequences from the five males were combined with sequences from the current cDNA library, as well as SPF sequences from other cDNA libraries, there were a total of 231 sequences that had 75 unique translations. The SPF sequences aligned into six sequence clusters based on sequence homology (Figure 8). Interestingly, 15 of the translations had sections matching two or more different classes; these unique sequences were termed hybrids (Figure 9). Hybrid sequences were found for every class of sequences with the exception of class VI. The regions where the hybrid would switch between classes was not always the same, however there were 5 sequences for which the switch occurred somewhere within a conserved region between amino acid 87 and amino acid 103. Four additional sequences switched classes between amino acid 116 and amino acid 132. The remaining six hybrid sequences all had unique switch locations.
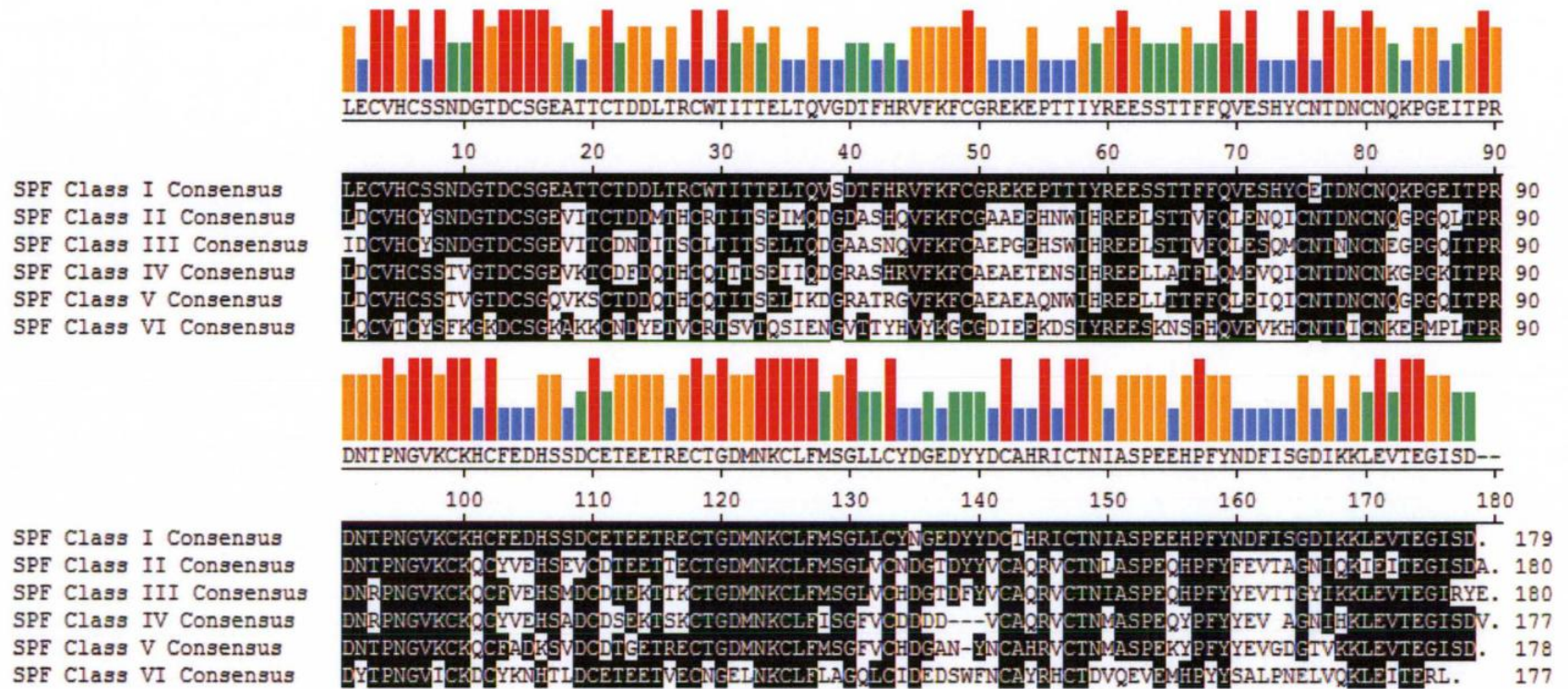
Figure 8. Six classes of SPF. The consensus sequence from each class of SPF is aligned against the overall consensus for all 229 SPF sequences. Residues that match the consensus are shaded. Bar height indicates the level of sequence conservation among all SPF sequences.

```
SPF Class I Consensus      LECVHCSSNDGTDCSGEATTCTDDLTRCWTITTELTQVSDTFHRVFKFCGREKEPTTIYR  60
SPF Class I-II Hybrid 01   LECVHCSSNDGTDCSGEATTCTDDLTRCWTITTELTQVSDTFHRVFKFCGREKEPTTIYR  60
SPF Class II Consensus     LDCVHCYSNDGTDCSGEVITCTDDMTHCRTITSEIMQDGDASHQVFKFCGAAEEHNWIHR  60

SPF Class I Consensus      EESSTTFFQVESHYCETDNCNQKPGEITPRDNTPNGVKCKHCFEDHSSDCETEETRECTG 120
SPF Class I-II Hybrid 01   EESSATFFQVESHYCETDNCNQKPGEITPRDNTPNGVKCKQCYVEHSEVCDTEETTECTG 120
SPF Class II Consensus     EELSTTVFQLENQICNTDNCNQGPGQLTPRDNTPNGVKCKQCYVEHSEVCDTEETTECTG 120

SPF Class I Consensus      DMNKCLFMSGLLCYNGEDYYDCTHRICTNIASPEEHPFYNDFISGDIKKLEVTEGISD.  179
SPF Class I-II Hybrid 01   DMNKCLFMSGLVCNDGTDYYVCAQRVCTNIASPEEHPFYNDFISGDIKKLEVTEGISV.  179
SPF Class II Consensus     DMNKCLFMSGLVCNDGTDYYVCAQRVCTNLASPEQHPFYFEVTAGNIQKIEITEGISDA. 180
```

Figure 9. An example of an SPF hybrid sequence. This sequence is highly similar to SPF class I sequences in

the N-terminal portion of the protein but the C-terminal half resembles class II sequences. The switch

between Class I and Class II occurs between amino acids 87 and 101.

Within each class of SPF there was very little sequence variation, but between classes, there was increased sequence distance (Table 5). It is unlikely that sequences within the same class are allelic variants of each other because individual males could express many unique sequences from each class. For example, five unique class I sequences and six unique class VI sequences were found to be expressed in the mental gland of a single male.

Out of the 75 unique predicted translations of SPF, there were multiple copies of only 17 sequences (Figure 10). One of these sequences, found 69 times, was identified in the cDNA library as well as in each of the five males. This sequence, from class I (SPF I01), was also the most highly transcribed sequence in four of the five males sampled. There were no other SPF sequences found to be transcribed in all of the males sampled. All other sequences were found at significantly lower frequency, with the next most highly transcribed gene (SPF II01) being identified 21 times and found in only three of the males sampled. None of the hybrid sequences were observed more than once.

Individual animal analysis of SPF sequences demonstrated that each male expressed a relatively unique pattern (Figure 11). One male (Male D) had an additional class of SPF (Class V) that the other males were lacking. Only one sequence of this class was identified, but the relative difference to the other classes versus their within group similarity suggests it should be grouped into a class of its own. Multiple sequences of the other classes of SPF were found in multiple males, however only two classes (Class I and Class III) were identified in all males. Male B was not identified class II, and Male D was not identified in Class IV and Class VI, although this may be based on sample size.

	Class I	Class II	Class III	Class IV	Class V	Class VI
Class I	0.004	0.176	0.198	0.220	0.190	0.598
Class II	0.386	0.007	0.094	0.142	0.138	0.633
Class III	0.395	0.218	0.006	0.148	0.157	0.738
Class IV	0.509	0.301	0.297	0.015	0.118	0.689
Class V	0.385	0.282	0.303	0.256	----------	0.659
Class VI	0.675	0.664	0.810	0.752	0.741	0.007

Amino Acid Distance

Nucleotide Distance

Table 5. The average number of base substitutions per site between sequence groups is shown above the diagonal and the average number of amino acid substitutions per site between classes (calculated using a JTT matrix-based model) is shown below the diagonal. The diagonal, outlined in black, shows the average number of amino acid substitutions per site within each class. Class V had only one sequence so the within class average could not be calculated. Analyses were conducted using a Maximum Composite Likelihood Model (*56*) with 500 bootstrap replicates. Evolutionary analyses were conducted in MEGA5 (*57*).

Class I SPF sequences were the most highly transcribed class in all males sampled and with one exception, SPF I01 composed the majority of all Class I transcripts in each male. But despite this similarity, the percentage of Class I sequences within a single male varied widely from ~30% to ~68%. Class III SPF sequences showed less variability in abundance, ranging between 6% and 11% of sequences. Hybrid sequences were not identified in every male sampled and generally composed a small percentage of sequences (0-6%); the one exception being male E (22% hybrid sequences).

```
SPF I01    (69)  LECVHCSSNDGTDCSGEATTCTDDLTRCWTITTELTQVSDTFHRVFKFCGREKEPTTIYR 60
SPF I14    (2)   LECVHCSSNDGTDCSGEATTCTDDLTRCWTITTELTQVSDTFHRVFKFCGREKEPTTIYR 60
SPF I16    (3)   LECVHCSSNDGTDCSGEATTCTDDLTRCWTITTELTQVSDTFHRVFKFCGREKEPTTIYR 60
SPF I19    (4)   LECVHCSSNDGTDCSGEATTCTDDLTRCWTITTELTQVSDTFHRAFKFCGREKEPTTIYR 60
SPF I23    (9)   LECVHCSSNDGTDCSGEATTCTDDLTRCWTITTELTQVSDTFHRVFKFCGREKEPTTIYR 60
SPF II01   (21)  LDCVHCYSNDGTDCSGEVITCTDDMTHCRTITSEIMQDGDASHQVFKFCGAAEEHNWIHR 60
SPF II05   (5)   LDCVHCYSNDGTDCSGEVITCTDDMTHCRTITSEIMQDGDASHQVFKFCGAAEEHNWIHR 60
SPF II08   (6)   LDCVHCYSNDGTDCSGEAITCTDDMTHCRTITSEIMQDGDASHQVFKFCGAAEEHNWIHR 60
SPF III01  (6)   IDCVHCYSNDGTDCSGEVITCDNDITSCLTIITSELTQDGAASNQVFKFCAEPGEHSWIHR 60
SPF III07  (3)   IDCVHCYSNDGTDCSGEVITCDNDITSCLTIITSELTQGGAASHQVFKFCAEPGEHSWIHR 60
SPF IV01   (4)   LDCVHCSSTVGTDCSGEVKTCDFDQTHCQTTTSEIIQDGRASHRVFKFCAEAETENSIHR 60
SPF IV04   (6)   LDCVHCSSTVGTDCSGEVKTCDFDQTHCQTTTSEIIQDGRASHRVFKFCAEAETENSIHR 60
SPF IV05   (2)   LDCVHCSSTVGTDCSGEVKTCDFDQTHCQTTISEIIQDGRASHRVFKFCAEAETENSIHR 60
SPF VI01   (16)  LQCVTCYSFKGKDCSGKAKKCNDYETVCRTSVTQSIENGVTTYHVYKGCGDIEEKDSIYR 60
SPF VI04   (5)   LQCVTCYSFKGKDCSGKAKKCNDYETVCRTSVTQSIENGVTTYHVYKGCGDIEEKDSIYR 60
SPF VI09   (3)   LQCVTCYSFKGKDCSGKAKKCNDYETVCRTSVTQSIDNGVTTYHVYKGCGDIEEKDSIYR 60
SPF VI12   (9)   LQCVTCYSFKGKDCSGKAKKCNDYETVCRTSVTQSIENGVTTYHVYKGCGDIEEKDSIYR 60


SPF I01          EESSTTFFQVESHYCETDNCNQKPGEITPRDNTPNGVKCKHCFEDHSSDCETEETRECTG 120
SPF I14          EESSTTFFQVESHYCETDNCNQKPGEITPRDNTPNGVKCKHCFEDHSSDCETEETRECTG 120
SPF I16          EESSTTFFQVESHYCETDNCNQKPGEITPRDNTPNGVKCKHCFEDHSSDCEAEETRECTG 120
SPF I19          EESSTTFFQVESHYCETDNCNQKPGEITPRDNTPNGVKCKHCFEDHSSDCETEETRECTG 120
SPF I23          EESSTTFFQVESHYCETDNCNQKPGEITPRDNTPNGVKCKHCFEDHSSDCETEETRECTG 120
SPF II01         EELSTTVFQLENQICNTDNCNQGPGQLTPRDNTPNGVKCKQCYVEHSEVCDTEETTECTG 120
SPF II05         EELSTTVFQLENQICNTDNCNQGPGQLTPRDNTPNGVKCKQCYVEHSEVCDTEETTECTG 120
SPF II08         EELSTTVFQLENQICNTDNCNQGPGQLTPRDNTPNGVKCKQCYVEHSEVCDTEETTECTG 120
SPF III01        EELSTTVFQLESQMCNTNNCNEGPGQITPRDNRPNGVKCKQCEVEHSMDCDTEKTTKCTG 120
SPF III07        EELSTTVFQLESQMCNTNNCNEGPGQITPRDNRPNGVKCKQCEVEHSMDCDTEKTTKCTG 120
SPF IV01         EELLATFLQMEVQICNTDNCNKGPGKITPRDNRPNGVKCKQCYVEHSANCDSEKASKCTG 120
SPF IV04         EELLATFLQMEVQICNTDNCNKGPGKITPRDNRPNGVKCKQCYVEHSADCDSEKTSKCTG 120
SPF IV05         EELLATFLQMEVQICNTDNCNKGPGKITPRGNRPNGVKCKQCYVEHSADCDSEKTSKCTG 120
SPF VI01         EESKNSFHQVEVKHCNTDICNKEPMPLTPRDYTPNGVICKDCYKNHTLDCETEETVECNG 120
SPF VI04         EESKNSFHQVEVKHCNTDICNKEPMPLTPRDYTPNGVICKDCYKNHTLDCETEETVECNG 120
SPF VI09         EESKNSFHQVEVKHCNTDICNKEPMPLTPRDYTPNGVICKDCYKNHTLDCETEETVECNG 120
SPF VI12         EESKNSFHQVEVKHCNTDICNKEPMPLTPRDYTPNGVICKDCYKNHTLDCETEETVECNG 120


SPF I01          DMNKCLFMSGLLCYNGEDYYDCTHRICTNIASPEEHPFYNDFISGDIKKLEVTEGISD. 179
SPF I14          DMNKCLFMSGLLCYNGEDYYDCTHRICTNIASPEEHPFYNGFISGDIKKLEVTEGISD. 179
SPF I16          DMNKCLFMSGLLCYNGEDYYDCTHRICTNIASPEEHPFYNDFISGDIKKLEVTEGISD. 179
SPF I19          DMNKCLFMSGLLCYNGEDYYDCTHRICTNIASPEEHPFYNDFISGDIKKLEVTEGISD. 179
SPF I23          DMNKCLFMSGLLCYNGEDYYDCTHRICTNIASPEEHPFYNDFISGDIKKLEVTEGISV. 179
SPF II01         DMNKCLFMSGLVCNDGTDYYVCAQRVCTNLASPEQHPFYFEVTAGNIQKIEITEGISDA. 180
SPF II05         DMNKCLFMSGLVCNDGTDYYVCAQRVCTNLASPEQHPFYFEVTAGNIQNIEITEGISDA. 180
SPF II08         DMNKCLFMSGLVCNDGTDYYVCAQRVCTNLASPEQHPFYFEVTAGNIQKIEITEGISDA. 180
SPF III01        DMNKCLFMSGLVCHDGTDFYVCAQRVCTNIASPEQHPFYYEVTTGYIKKLEVTEGIRYE. 180
SPF III07        DMNKCLFMSGLVCHDGTDFYVCAQRVCTNIASPEQHPFYYEVTTGYIKKLEVTEGISD. 179
SPF IV01         DMNKCLFISGFVCDDDD---VCAQRVCTNMASPEQYPFYYEVIAGNIHKLEVTEGISDV. 177
SPF IV04         DMNKCLFISGFVCDDDD---VCAQRVCTNMASPEQYPFYYEVIAGNIHKLEVTEGISDV. 177
SPF IV05         DMNKCLFISGFVCDDDD---VCAQRVCTNMASPEQYPFYYEVIAGNIQKIEITEGISDA. 177
SPF VI01         ELNKCLFLAGQLCIDEDSWFNCAYRHCTDVQEVEMHEYYSALPNELVQKLEITERL. 177
SPF VI04         ELNKCLFLAGQLCIDEDSWFDCAYRHCTDVQEVEMHEYYSALPNELVQKLEITERL. 177
SPF VI09         ELNKCLFLAGQLCTDEDSWFNCAYRHCTDVQEVEMHEYYSALPNELVQKLEITERL. 177
SPF VI12         ELNKCLFLAGQLCTDEDSWFNCAYRHCTDVQEVEMHEYYSALPNELVQKLEITERL. 177
```

44

Figure 10. SPF sequences for which multiple copies were found. The name of each

sequence is based on SPF class and sequence number within that class. The number of

copies identified out of the 231 sequences is in parentheses following the name of the

sequence. Residues shaded in black match a consensus based on all full-length SPF

transcripts identified.
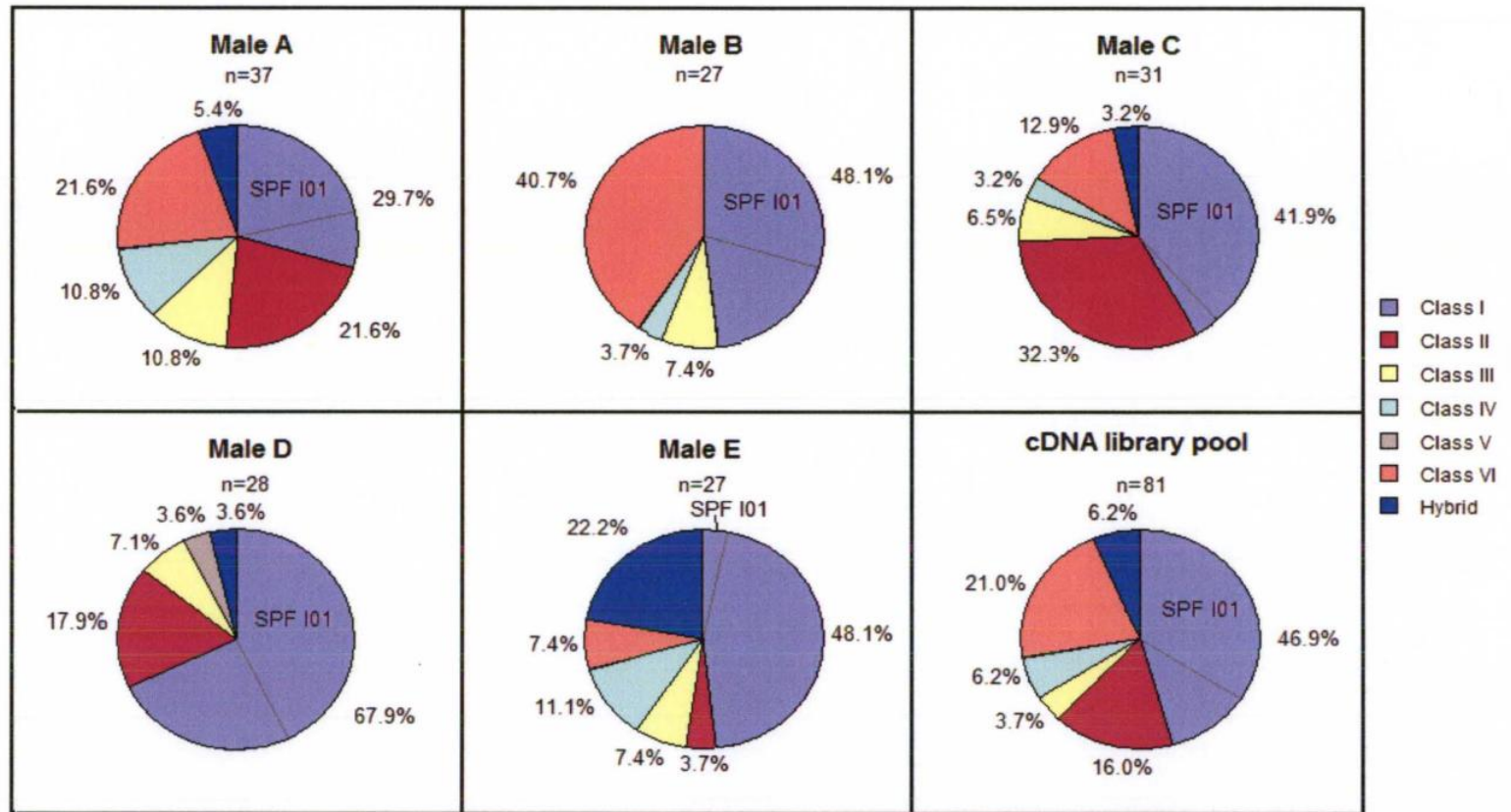
Figure 11. Division of SPF sequences among the six different classes within each individual male or from two previous cDNA libraries. Hybrid refers to transcripts that had sections matching two or more different classes. One main isoform of SPF (SPF I01) was found in all males sampled and comprised ~30% of all SPF transcripts that were sequenced.

PHEROMONE PURIFICATION AND MASS SPECTRAL ANALYSIS

The three independent chromatographic separations (Figure 12) yielded
sufficiently pure proteins for analysis by ESI-MS and LC/MS/MS. The nine proteins that
were purified from the whole pheromone mixture ranged in mass from 3,552 Da to
20,247 Da and matched sequences from the *D. ocoee* cDNA database with the percent
coverage ranging from 27% to 90% (Table 6).

The protein with the highest mass was from size-exclusion fraction A (15-20
minutes) with an elution time on the reverse phase HPLC of 36.44 minutes. This protein,
which comprised approximately 25% the pheromone extract, was the most dominant
component and matched the most highly transcribed SPF sequence (SPF I01), with 90%
percent sequence coverage.

To further characterize unknown proteins, cysteine content was ascertained by
determination of protein mass (ESI-MS) with and without reduction and alkylation of the
cysteine residues. The intact protein mass for the SPF protein was 20247 kDa, and after
alkylation with carbamidomethyl groups, the mass increased to 21175 Da (928 Da shift)
(Figure 13). Each carbamidomethyl group is equal to 58 Da, and therefore the mass
difference indicates this protein has 16 cysteine residues. The SPF I01 sequence also has
a predicted mass of 20247 Da and 16 cysteine residues, and coupled with the relative
abundance and >90% peptide mass fingerprint sequence coverage, the data are
complementary and confirm identification.
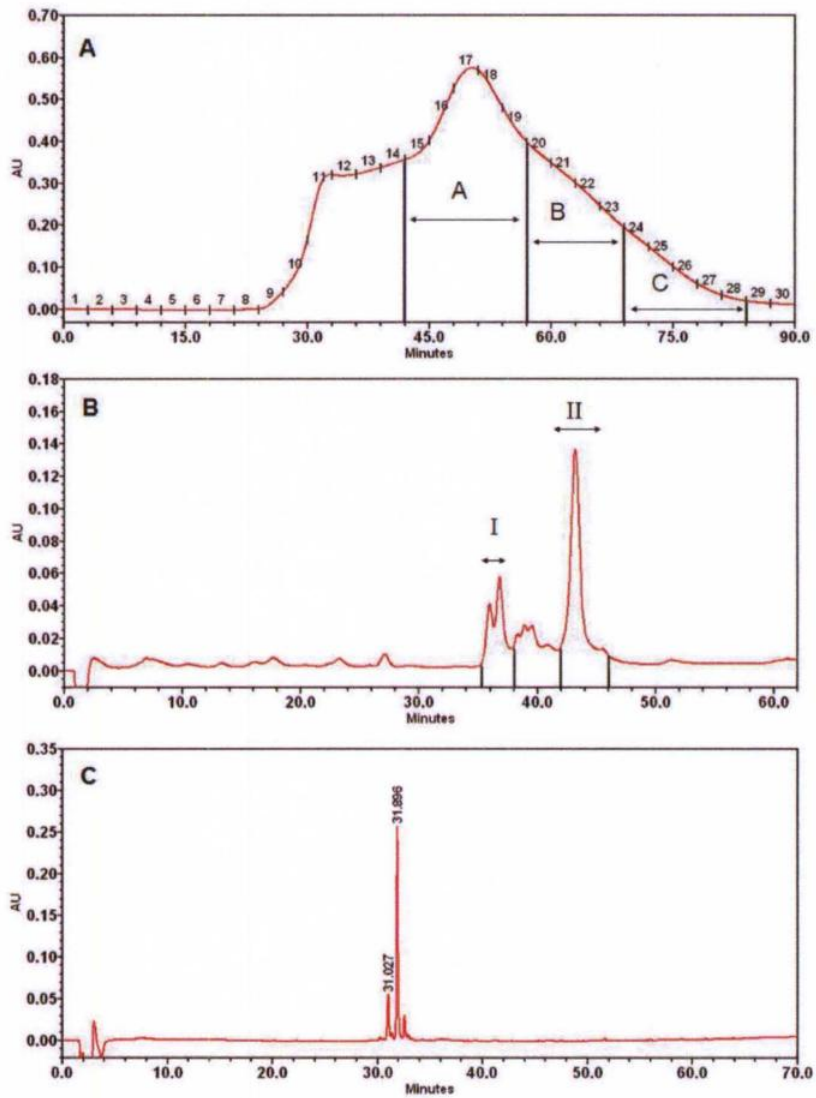
Figure 12. Sample purification of a single protein from the *D. ocoee* whole pheromone mixture: (A) Size Exclusion HPLC of 500ul of whole extract on a G75 Superfine column in 1/2 X PBS. (B) Anion Exchange HPLC of Fraction C (69-83 min) from the G75 in 50 mM Tris with a mixed gradient. (C) Reverse Phase HPLC of fraction II (42-46 min) from Mono Q on a C-18 column at 1.4% B/min. gradient. Peak elution was monitored at 220 nm.

RP-HPLC Retention Time (Min)	ESI-MS MW in Da	Reduced & alkylated MW	Predicted # of Cys	Percent Coverage	Sequence Match
Fraction A					
36.44	20247	21175	16	90%	SPF IO1
38.38	3552	3552	0	NA	NA
Fraction B					
31.87	6499	NA	NA	43%	D536
32.43	6480	3629 + 3198	6	27%	Relaxin-like01
33.01	6508	6508	0	NA	NA
36.47	NA	NA	NA	68%	Protease inhibitor-like (D148)
Fraction C					
31.04	6042	NA	NA	64%	D132
31.45	NA	NA	NA	81%	Relaxin-like01
31.91	6088	6438	6	33%	Protease inhibitor-like (D611D)

Table 6: Summary of ESI and MALDI-TOF/ LC MS/MS mass spectral data obtained for purified *D. ocoee* proteins. Proteins are identified by what fraction the eluted in from the G75 (A=42-56 minutes, B=57-68 minutes, C=69-84 minutes), and by RP-HPLC elution time. Data was obtained from ESI-MS and either MALDI-TOF or LC MS/MS.
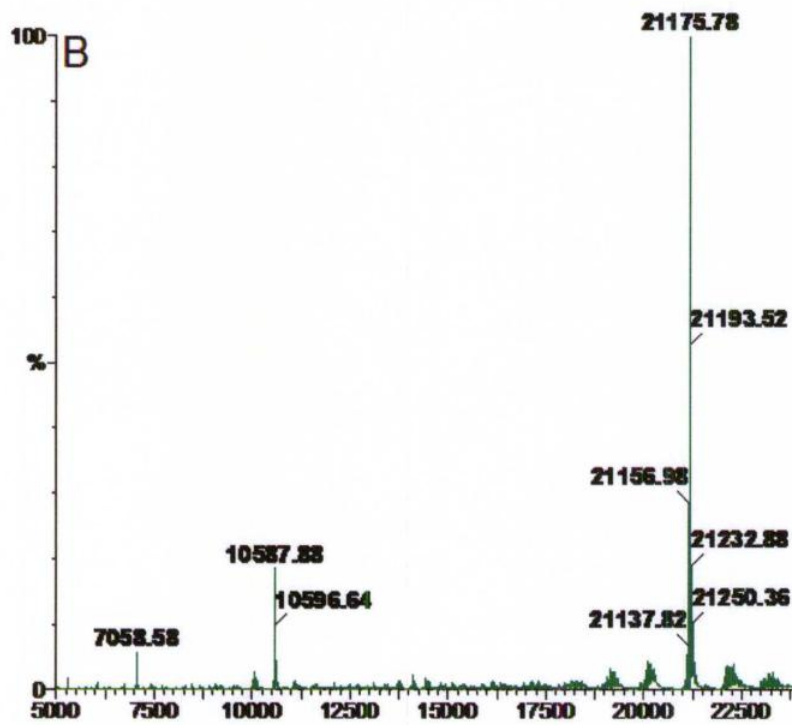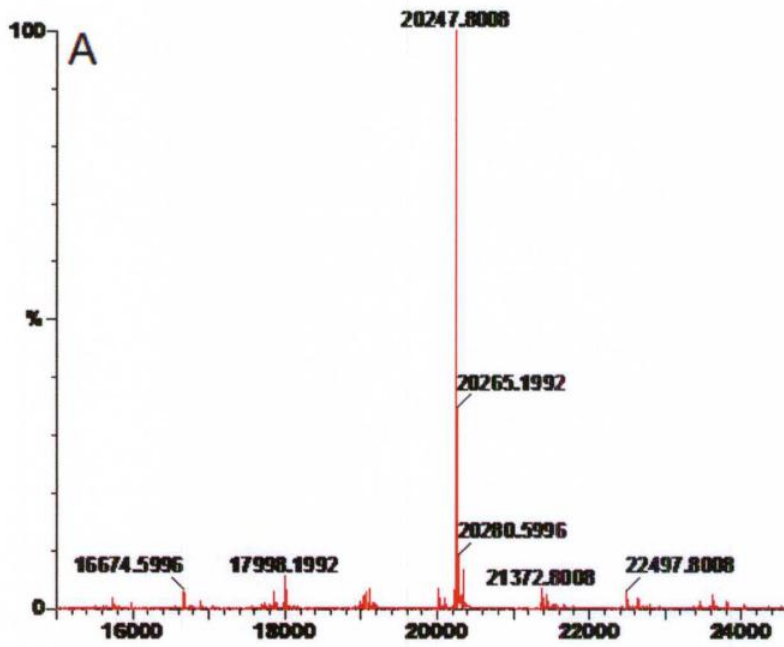
Figure 13. ESI-MS Spectrum for SPF I01 protein without (A) and with (B) reduction and

alkylation.

When mental gland extract was collected and analyzed at approximately 3 week intervals from May 2010 until October 2010 the pheromone profile remained consistent. The most dominant protein from each time point analyzed was collected and determined to be SPF I01 by ESI-MS, suggesting this isoform of SPF is expressed at high levels throughout the full duration of the 5-6 month courtship season.

In addition to the major SPF, two purified proteins matched a relaxin-like sequence. The first of these two was from fraction B, had a RP-HPLC retention time of 32.43 minutes, and matched a relaxin sequence with 27% coverage. The relatively low percent coverage of the protein can be explained if this protein is an isoform of the relaxin sequence it matched; one or two SNPs could reduce the percent coverage significantly. ESI-MS data determined the intact mass of this protein to be 6480 Da. After reduction and alkylation two fragments were found that were 3198 and 3629 Da respectively which suggests that the protein is composed of two subunits held together by disulfide bonds. Based on the mass shift from 6480 Da to 6827 Da (when the mass of the two fragments is combined) this protein contains 6 cysteine residues, which is consistent with it being a relaxin-like molecule. The second protein matching a relaxin-like sequence is a protein from fraction C that eluted at 31.4 minutes on the RP-HPLC. This protein matched a relaxin-like sequence with 81% coverage. The portion of the protein that was not covered was the section between two dibasic sites (equivalent to the C-peptide of relaxin) which is expected to be cleaved out during processing. No ESI-MS data was collected for this protein.

Two of the remaining proteins each matched to sequences with similarity to serine protease inhibitors. Specifically, the first protein from fraction B with an RP-

HPLC elution time of 36.47 minutes matched a protease inhibitor sequence with 68% coverage. The second protein from fraction C had an RP-HPLC elution time of 31.91 minutes and matched a protease inhibitor sequence with 33% coverage. The protease inhibitor sequence was predicted to contain 6 cysteine residues, and this is supported by ESI-MS analysis after carbamidomethyl alkylation.

DETERMINATION OF THE SPF GENE STRUCTURE

Following three separate restriction enzyme digests of SPF and self circularization reactions, nested inverse PCR was performed using primers that annealed to the SPF 3'UTR. When PCR products were analyzed by agarose gel electrophoresis, one reaction that produced a single band on the gel was selected for cloning. Following cloning, colony PCR was conducted and 4 samples were selected for sequencing in both directions. Two of the samples were identified to have regions matching the SPF 3' UTR as well as 131 bases upstream that matched the C-terminal coding region of SPF I01. The region upstream of the 131 base pair region no longer matched the coding region of SPF and ended with a 3'AG, suggesting an intron sequence. After being trimmed to the restriction enzyme splice site, the length of the partial intron sequence obtained was 417 base pairs. For the full length SPF I01 coding sequence of 597 base pairs, the splice site for this intron occurred between nucleotides 466 and 467, which splits amino acid 156 between the first and second nucleotide of the codon. Following identification of this intron splice site, four primers were synthesized based on SPF I01 sequence between nucleotides 378 and 466. To identify the next upstream splice site, these primers were used in nested PCR reactions based on the genome walking method described by Guo

and Xiong (*44*). Following cloning and sequencing an additional splice site was identified between nucleotides 319 and 320, which splits amino acid 107 between the first and second nucleotide of the codon.

More primers were designed based on SPF I01 coding sequence upstream of this splice site and the genome walking methods of Guo and Xiong were repeated to identify two more exon-intron boundary sites. Specifically, splice sites were identified between nucleotides 172 and 173 as well as between nucleotides 52 and 53.

These data suggest an SPF gene structure consisting of 5 exons and 4 introns (Figure 14). These exons include one encoding the 5' UTR plus the majority of the signal peptide and a pair of exons for each of the two TFP-like domains. Specifically, the first exon is a total of 106 base pairs in length including 54 base pairs of the SPF 5' UTR as well 52 nucleotides of the coding sequence. This splice site occurs within the signal peptide of SPF splitting amino acid 18 between the first and second nucleotides of the codon. The second exon extends from nucleotide 53 to nucleotide 172 for a total length of 120 base pairs. This exon codes for the remainder of the signal peptide as well as the first 37 amino acids of the native protein including the first five cysteine residues. The third exon is composed of nucleotides 173-319 for a total length of 147 base pairs. This exon is composed of the remainder of the first TFP-like domain of SPF, including three cysteine residues, as well as the first five amino acids of the linker between the two TFP domains. Exon 4 contains nucleotides 320 through 466, a total of 147 base pairs, which correspond

Figure 14. The SPF I01 gene structure. Exons are drawn to scale. Black coloration represents 5' and 3' untranslated regions, the signal peptide is red and the mature protein is blue.

to the remaining portion of the linker and the beginning of the second TFP domain including six cysteine residues. The fifth and final exon of SPF contains the C-terminal coding region, which includes the last 2 cysteine residues, as well as the 3' UTR.

Based on the known SPF gene structure, it was determined that the hybrid sequences described previously are not likely to be the result of alternative splicing. Out of 17 locations where SPF sequences switched classes, only 6 of the sites (35%) could be explained by alternative splicing (Figure 15), which suggests that at least some of the hybrid sequences are generated by a different mechanism.

SPF DISULFIDE BONDING PATTERN

Initial protease digests with LysC and chymotrypsin + LysC suggested that cysteines 1, 2, 3, 4, 5, and 6 were contained within a single disulfide bonded peptide. Further analysis of peptides generated by protease digests with trypsin and trypsin + chymotrypsin suggested that the first four cysteine residues were intramolecularly bonded. By deduction this suggested that cysteine 5 and cysteine 6 must be bonded to each other, and that finding was confirmed by fragments identified in two additional protease digests: 1) Arg C + chymotrypsin and 2) AspN + trypsin.

Separate digestions of AspN and AspN + trypsin identified a peptide containing the first two cysteine residues which suggested a disulfide bond between these residues. Additionally, data obtained from a digest with Asp N + trypsin was consistent with cysteines 3 and 4 were bonded to each other. Importantly, five different fragment matches suggested that cysteines 7 and 8 were bonded, and one fragment match was consistent with cysteine 12 and 16 being bonded.
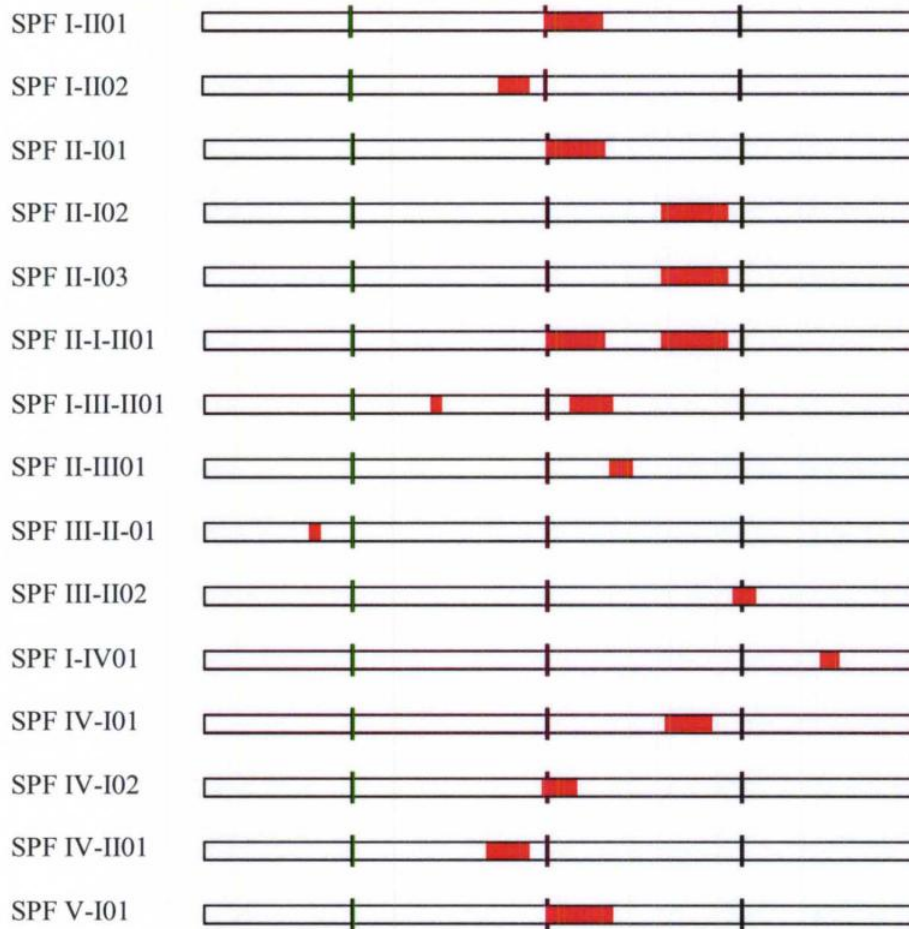
Figure 15. SPF hybrid switch locations. SPF hybrids are indicated by rectangles and were named based on the classes to which they belong. Vertical black bars indicate the location of splice sites and red boxes indicate the region where the SPF sequence switches from one class to another.

After analysis of a total of 10 SPF protease digests followed by LC MS/MS, 19 fragment matches were confirmed (Table 7). In total, 5 disulfide bonds have been resolved: Cys1-Cys2, Cys3-Cys4, Cys5-Cys6, Cys7-Cys8, Cys12-Cys16. While 3 more disulfide bonds remain to be determined, the current data reduces the number of possible disulfide bonding patterns significantly from 2,027,025 to only 15 (Appendix 1) and suggests a general architecture for the protein. Specifically, the results suggest that SPF is composed of two separate disulfide-bonded domains separated by a hydrophilic linker region. Secondly, the results demonstrate that the disulfide-bonding patterns of the domains differ from each other and both are also different from a typical TFP pattern and from PMF.

Cysteines	Protease(s)	Fragment(s)	Theoretical Mass	Experimental Mass
1,2,3,4,5,6	Lys C	LECVHCSSNDGTDCSGEATTCTDDLTRCWTITTELTQVSDTFHRVFK +FCGREK	5958.5868	5958.5745
1,2,3,4,5,6	LysC+Chymo	LECVHCSSNDGTDCSGEATTCTDDLTRCW + FCGREK	3854.4818	3854.4848
1,2,3,4	Trypsin	LECVHCSSNDGTDCSGEATTCTDDLTR	2829.0695	2829.0699
1,2,3,4	Trypsin+Chymo	LECVHCSSNDGTDCSGEATTCTDDLTR	2829.0695	2829.3218
1,2	AspN	LECVHCSSN	989.3782	989.4312
1,2	AspN+Trypsin	LECVHCSSN	989.3782	989.0348
1,2	AspN+Trypsin	LECVHCSSNDGT	1262.4743	1262.0309
3,4	AspN+Trypsin	DCSGEATTCT	1037.3474	1037.5820
3,4	AspN+Trypsin	DCSGEATTCTD	1100.3474	1100.5151
3,4	AspN+Trypsin	DGTDCSGEATTCT	1258.4166	1257.6753
5,6	ArgC+Chymo	CW+CGREKEPTTIY	1601.7007	1601.5427
5,6	AspN+Trypsin	CWTITTELTQVS + FCGREK	2118.0011	2118.8475
7,8	Lys C	EPTTIYREESSTTFFQVESHYCETDNCNQK	3584.5213	3584.5217
7,8	Glu C	SHYCETDNCNQKPGE	1722.6479	1722.6481
7,8	LysC+Chymo	CETDNCNQKPGEITPRDNTPNGVK	2628.1698	2628.6649
7,8	AspN+Trypsin	EESSTTFFQVESHYCETDNCNQKPGEITPR	3474.4817	3474.4055
7,8	AspN+Trypsin	EKEPTTIYREESSTTFFQVESHYCETDNCNQKPGEITPR	4592.0586	4592.0650
12,16	AspN+Trypsin	ECTG + ICTNIASPEEHPFYN	2140.8966	2140.6870

Table 7. SPF MS/MS peptide fragment matches. The sequence of peptide fragment is indicated in column 3 with the peptides

held together by disulfide bonds separated by a (+). The cysteines contained in the fragments (column 1), the protease(s) used

for the digest (column 2), and theoretical and experimental masses of each match are shown.

CHAPTER IV

DISCUSSION

Approximately 20 million years ago a clade of plethodontid salamanders underwent a significant behavioral adjustment from transdermal to olfactory delivery. This change in pheromone delivery mechanism has been accompanied by a change in pheromone composition from SPF and hormone-like peptides in *D. ocoee* to a mixture largely composed of PRF and PMF in *P. shermani (30)*.

Similar to other plethodontid pheromones, SPF shows a large degree of variability at the transcript level with 75 unique sequences being found in 231 cDNA clones. However, in contrast to this variability at the transcript level, only one SPF sequence (SPFI01) is expressed at high protein levels. The transcript for this protein was found in all males sampled, and the protein was expressed in pooled pheromone extract throughout the duration of the courtship season. It is possible that basal levels of other SPF proteins are being expressed in individual *D. ocoee* males. Because the pheromone extract is a pool of many males, an SPF protein expressed only in a single male would be in small amounts compared to the main SPF (SPFI01) transcribed in all males. This possibility is supported by the variation between males at the transcript level including many unique transcripts that are transcribed at low levels. Additionally, while MS

analyses confirmed that SPFI01 is not glycosylated, SDS-PAGE followed by carbohydrate staining suggested that some SPF-like proteins are glycosylated (Leichty et al., unpublished data). Class VI SPF sequences would be likely candidates for glycosylation since they contain an N-linked glycosylation consensus sequence. However, whether or not individual males express additional SPF proteins at low levels, SPF is unique among plethodontid pheromones in that only one protein is highly expressed in the pheromone extract. This is in contrast to the high expression of three main PRF isoforms and greater than 30 PMF isoforms found in the pheromone extract of *P. shermani*.

The current literature has defined two mating seasons for *D. ocoee*: the first in May and early June and the second during the month of August (*58-60*). However, when pheromone extract was analyzed from seven time points across a five month time period (from late May until early October), RP-HPLC chromatograms of pooled pheromone extract remained consistent. The concentration of pheromone extracted from 20 male pools also remained consistent throughout the courtship season, with the exception of October. In addition, when the major SPF peak from each time point was analyzed by ESI-MS, it was the same isoform, SPF I01. Based on these data, we conclude that male *D. ocoee* maintain both their mental gland and high levels of pheromone throughout the entire 5 month period.

While pheromone and mental gland maintenance during a period of non-courtship (presumably late June and July) may seem unusual, it is likely that this occurs because the two courtship seasons are close together; preserving the gland during this period may be less costly than gland regression at the end of one courtship season and regeneration at

the beginning of the next – a process that might not be physiologically feasible. Additionally, the levels of pheromone detected during June and July may be a result of low pheromone protein turnover rates, rather than from active synthesis making gland maintenance more favorable. Alternatively, it is possible that *D. ocoee* has an extended mating season spanning the entire period sampled. If mating occurs, even at low levels during June and July, it may be advantageous for males to continue pheromone synthesis during this period.

A large percentage of the cDNA sequences identified were hormone-like. In contrast hormone-like sequences have not been found in the *P. shermani* mental gland cDNA at any time between May and August (Wilburn et al., unpublished data). This suggests that the presence of hormone-like sequences is unique to the *D. ocoee* pheromone mixture and may be representative of transdermal pheromone delivery. Since the pheromones of *D. ocoee* are likely traveling through the bloodstream of the female, hormone-like peptides, with high solubility and the ability to be recognized by receptors of different target organs, may be evolutionarily advantageous for targeting multiple tissues. Additionally, even very low concentrations of these hormone-like peptides could have large physiological effects as signaling cascades amplify the signal within responsive cells.

The hormone-like sequences that were identified were similar to VIP, glucagon, insulin and relaxin. VIP has been demonstrated to cause vasodilation and lower blood pressure, to activate secretion of both Na^+ and Cl^- into the small intestines, and to stimulate adenylate cyclase activity (*61, 62*). PHI is structurally similar to VIP and is demonstrated to have similar physiological effects, but with less potency (*45*). Typically

native VIP and PHI peptides are 28 and 27 amino acids respectively (*46*). Interestingly, in the *D. ocoee* VIP precursor-like protein, the normal dibasic cleavage site at the end of the expected VIP peptide is not present. It is possible that the protein is cleaved at a single basic residue similar to how the N-terminal end of human PHI is cleaved. Alternatively, the cleavage site could be at a dibasic site found later in the sequence, which would make the processed *D. ocoee* VIP longer in length than native VIP from other species. Similarities between *D. ocoee* VIP and other VIP sequences include a number of residues that are reported to be important in binding to VIP receptors including His-1, Asp-3, Phe-6, Thr-7, and Tyr-10 (*63*). Based on conservation of key residues it is possible that a *D. ocoee* VIP-like peptide could have a similar function to VIP from other species. Specifically, VIP in *D. ocoee* could function to increase vasodilation in the female and allow for more effective transfer of pheromones into the bloodstream. However, differences in *D. ocoee* VIP such as a possible extended length could result in a novel pheromone activity for this peptide.

Glucagon is a 29 amino acid hormone secreted from the pancreas during hypoglycemia that binds to receptors on liver cells and activates adenylate cyclase activity, which ultimately leads to an increase in blood glucose levels (*64*). Sites in glucagon that were demonstrated to be important for receptor binding in other species such as the N-terminal histidine and an aspartic acid residue are conserved in the all of the *D. ocoee* glucagon-like peptides (*50, 51*). In addition, one of the predicted *D. ocoee* peptides had additional sites of conservation in residues 10-12, which have also been demonstrated to be important in binding of the hormone to glucagon receptors (*53*). However, a distinction of *D. ocoee* glucagon-like peptides is that the normal dibasic

cleavage site is not present, likely producing an extended version. Another dibasic site was identified as the likely cleavage site in *D. ocoee* making the two peptides longer, at 32 and 33 residues, respectively. While neither the *D. ocoee* VIP nor glucagon-like peptide were not found as proteins, the small size of the processed forms of these peptides makes it likely that they were lost during purification by size exclusion chromatography.

Relaxins have been identified as molecules that have important roles in the reproductive tract (*65*). Specifically, relaxin has a role in the remodeling of collagen by inhibiting collagen synthesis and stimulating collagen breakdown leading to increased flexibility of the tissues and changes in the organ structure of target tissues during reproduction (*66*). Other roles of relaxin include relaxation of smooth muscle (*67*), vasodilation (*68*) and in males, increased sperm motility (*69*). In the *D. ocoee* relaxin-like sequence, there is a conservation of an arginine residue of the B chain that has been described as important for binding to receptors (*70*). However, the *D. ocoee* relaxin-like sequences have considerably shorter C-peptides than relaxin from other species. Although the C-peptide is removed from the protein by processing, it is important for the formation of the correct disulfide bonds between the A and B chains. A relaxin-like protein was identified by proteomic analysis of the *D. ocoee* pheromone extract suggesting that this protein is indeed delivered to the female during courtship.

An interesting commonality between the hormone-like peptides in *D. ocoee* is that the normal cleavage sites of the *D. ocoee* VIP, glucagon and insulin-like peptides are absent. This may produce peptides of a different length than native hormones of other species and may alter the function of the peptides. In addition, the hormone-like-peptides

from *D. ocoee* were always the outgroup when compared in gene trees with hormones from other species. Another phenomenon seen in *D. ocoee* is large variability in the transcripts encoding these proteins, which is likely the result of rapid gene duplication and mutation observed in other plethodontid pheromones (*25, 71*). Through the process of gene duplication and neo-functionalization, it is possible that the *D. ocoee* mental gland hormone-like peptides have been co-opted for pheromone function.

Because SPF is the most highly expressed protein of the *D. ocoee* mental gland, it was the subject of more extensive characterization to make comparisons with other abundant plethodontid pheromones. Similar to PMF, SPF is also related to the three-fingered protein superfamily. Closer analyses of the relationship of SPF in show that it is similar in structure to a "double TFP." Within each TFP-like domain, SPF contains three of the four disulfide bonds highly conserved among TFP family members as well as an additional disulfide bond found in a smaller percentage of TFPs (Figure 16). The disulfide bond missing in SPF is lacking is also lacking in the first TFP domain of urokinase plasminogen activator receptor (UPAR). Structural studies of UPAR by X-ray crystallography suggest a typical TFP structure for all three of its TFP domains demonstrating that the disulfide bond, which is absent, is not critical for the formation of a TFP structure (*36*).

In addition to similarities at the amino acid level, the genomic structure of SPF is also similar to that of the TFPs. Specifically, the first exon contains the 5' untranslated region and the signal sequence, and each TFP domain is coded by two additional exons a pattern reminiscent of the gene structure of other known TFPs (*37-39, 72*). In addition,

```
                          LECYSCQGSGET-CST------CAGGEDA-CLVK-------------KSGTVTFKGCGT----CPTGDV---------GLKVKCCTS-DLCN
                                    10        20        30        40        50        60        70        80        90

SPF I01 Domain 1          LECVHCSSNDGTDCSGEATT---CTDDLTR-CWTITTELTQV----SDTFHRVFKFCGREKEPTTIYREESSTTFF---QVESHYCET-DNCN
SPF I01 Domain 2          VKCKHCFEDHSSDCETEETRE--CTGDMNK-CLF----------------MSGLLCYN-----------------GEDYYDCTH-RICT
UPAR Domain 1             LRCMQCKTNGD--CRVEE-----CALGQDL-CRTTIVRLWEE----GEELELVEKSCTHSEKTNRTLSYRTGLKIT---SLTEVVCGL-DLCN
UPAR Domain 2             LECISCGSSDMS-CERGRHQSLQCRSPEEQ-CLDVVTHWIQEGEEGRPKDDRHLRGCGYLPG-CPGSNGFHNNDT----FHFLKCCNT-TKCN
UPAR Domain 3             RQCYSCKGNSTHGCSSEETFLIDCRGPMNQ-CLVATGTHE------PKNQSYMVRGCATASM-CQHAHLGDAFSMN---HIDVSCCTK-SGCN
Human CD59                LQCYNCPNPTAD-CKTAVN----CSSDFDA-CLI-----------TKAGLQVYNKCWKFEH-CNFNDVTTRLREN---ELTYYCCKK-DLCN
P. elegans PLI            MECEVCMRPGET-CNGSIRT---CKDNEDT-CVMLQTELIRA----PLSFTFTSKMCSTSDT-CHLDYVETNLPHELAVRSKRACVG-DECK
Mus musculus              LTCEVCKGSGHT-CSGKMKT---CEDGKDA-CVVLVSESSTK----GRKSVNTFKACMKYKD-CYSGFVSTTMTPSDYMVSNAHCCQS-DGCN
Xenopus CD59-like         LQCYSCPGFSKS-CETTTT----CTGSQNA-CLK-----------LTVADSSTYRCTDMSR-CEREFIKQDFGDIT--NYKYGCCQK-NLCN
Xenoxin 1 Precursor       LKCVNLQANGIKMTQE-------CAKEDTK-CLTL----------RSLKKTLKFCASGRT-CTTMKIMSLP------GEQITCCEG-NMCN
D. ocoee PMF              LKCKKLGGGTEV----------CPPGNDGVCIHY-----------KYKKEEFKRCGAHGE-CQEIESSFKPLETA--NTKFVCCSS-DLCN
PMF E3                    LRCKTLNENSDF----------CPPGNDEACMLY-----------KSTGMEFMGCATQRF-CDALGAALTEEGS---DDLFLCCTG-ELCN
PMF G                     LQCNTLDGGTEE----------CIPGIYNVCVHY-----------KSEDEEYKSCGIQEE-CEDAE----------GATVLCCPE-DLCN
Fasciculin 1              TMCYSHTTTSRAILTN------CGENS---CYRKSRRH--------PPKMVLGRGCG-----CPPGDD---------YLEVKCCTSPDKCN
Fasciculin 2              TMCYSHTTTSRAILTN------CGENS---CYRKSRRH--------PPKMVLGRGCG-----CPPGDD---------NLEVKCCTSPDKCN
Naja Nigricollis Toxin A  LECHNQQSSQPPTTKT------CPGETN--CYKKVWRD--------HRGTIIERGCG-----CPTVKP---------GIKLNCCTT-DKCN
Erabutoxin B              RICFNHQSSQPQTTKT------CSPGESS-CYHKQWSD--------FRGTIIERGCG-----CPTVKP---------GIKLSCCES-EVCN
Prod1                     LKCFTRNGDDRTVTT-------CAEEQTR-CLF-----------VQLPYSEIQECKTVQQ-CAEVLEEVTAI-----GYPAKCGCE-DLCN
```

Figure 16. The TFP domains of different proteins were aligned including: SPF, UPAR, CD59, Domain 1 of a Phospholipase A2 inhibitor from *Protobothrops elegans*, an unidentified *Mus musculus* protein, Xenoxin 1, PMF proteins, Fasciculin 1 and 2, *Naja nigricollis* Toxin A, Erabutoxin 1 and Prod1. Cysteine residues are highlighted in black. Bar height indicates consensus strength.

Figure 17. SPF vs. TFP disulfide bonding pattern. The expected TFP disulfide bonding

pattern is shown on top in green, with the bonding pattern determined for the first TFP-

like domain of SPF shown in red below.

all intron-exon junctions are type I and there is conservation of key cysteine residues. Taken together these data suggest that SPF is an ancient member of the TFP superfamily.

However, while SPF has been demonstrated to be similar to TFP members, its disulfide bonding pattern is different than the canonical, highly conserved TFP pattern. If SPF followed the normal TFP disulfide bonding pattern it would be expected to have bonds between Cys 1-5, 2-3, 4-6, 7-8. Instead, SPF has a disulfide bonding pattern of 1-2, 3-4, 5-6, 7-8 in its first domain (Figure 17). Its second domain is also different than the typical TFP folding pattern in that Cys 12-16 are bonded when a 15-16 bond would be expected based on the TFP pattern.

In summary, Plethodontid salamanders offer the unique opportunity to study the co-evolution of behavior and molecular pheromone signaling. The current results support the hypothesis that unique structural motifs have occurred in the different modes of delivery to facilitate reproductive success. Specifically, *D. ocoee* utilizes small hormone-like peptides along with SPF to facilitate mating success while *P. shermani* employs the use of PRF and PMF. While *D. ocoee* and *P. shermani* both produce proteins that are related to the TFP family members, both proteins have evolved different disulfide bonding patterns which may allow them to uniquely function under different modes of delivery.

REFERENCES

1. P. Karlson, M. Luscher, `Pheromones': a New Term for a Class of Biologically Active Substances. *Nature* **183**, 55 (1959).
2. W. L. Roelofs, A. Comeau, R. Selle, Sex Pheromone of the Oriental Fruit Moth. *Nature* **224**, 723 (1969).
3. W. L. Roelofs, A. S. Hill, R. T. Carde, T. C. Baker, Two sex pheromone components of the tobacco budworm moth. *Life Sciences* **14**, 1555 (1974).
4. M. C. Birch, K. F. Haynes, *Insect pheromones*. (E. Arnold London, 1982).
5. A. Butenandt, R. Beckmann, E. Hecker, in *Hoppe-Seyler's Zeitschrift für physiologische Chemie*. (1961), vol. 324, pp. 71.
6. R. M. Silverstein, J. O. Rodin, D. L. Wood, Sex Attractants in Frass Produced by Male Ips confusus in Ponderosa Pine. *Science* **154**, 509 (1966).
7. J. G. Dulka, N. E. Stacey, P. W. Sorensen, G. J. V. D. Kraak, A steroid sex pheromone synchronizes male-female spawning readiness in goldfish. *Nature* **325**, 251 (1987).
8. D. Burnard, R. E. Gozlan, S. W. Griffiths, The role of pheromones in freshwater fishes. *Journal of Fish Biology* **73**, 1 (2008).
9. S. Kikuyama *et al.*, Sodefrin: a female-attracting peptide pheromone in newt cloacal glands. *Science* **267**, 1643 (1995).
10. S. M. Rollmann, L. D. Houck, R. C. Feldhoff, Proteinaceous Pheromone Affecting Female Receptivity in a Terrestrial Salamander. *Science* **285**, 1907 (1999).
11. P. A. Wabnitz, J. H. Bowie, M. J. Tyler, J. C. Wallace, B. P. Smith, Animal behaviour: Aquatic sex pheromone from a male tree frog. *Nature* **401**, 444 (1999).
12. J. G. Vandenbergh, J. M. Whitsett, J. R. Lombardi, Partial isolation of a pheromone accelerating puberty in female mice. *Journal of Reproduction and Fertility* **43**, 515 (1975).
13. M. V. Novotny, W. Ma, D. Wiesler, l. Zidek, Positive identification of the puberty-accelerating pheromone of the house mouse: the volatile ligands associating with the major urinary protein. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **266**, 2017 (1999).
14. B. Schall *et al.*, Chemical and behavioural characterization of the rabbit mammary pheromone. *Nature* **424**, 68 (2003).
15. P. Chamero *et al.*, Identification of protein pheromones that promote aggressive behaviour. *Nature* **450**, 899 (Dec 6, 2007).
16. A. Massey, J. G. Vandenbergh, Puberty delay by a urinary cue from female house mice in feral populations. *Science* **209**, 821 (1980).
17. K. M. Dorries, E. Adkins-Regan, B. P. Halpern, Sensitivity and Behavior

Responses to the Pheromone Androstenone Are Not Mediated by the Vomeronasal Organ in Domestic Pigs. *Brain, Behavior, and Evolution* **49**, 53 (1997).

18. R. C. Feldhoff, S. M. Rollmann and L. D. Houck, in *Advances in chemical signals in vertebrates,* R. E. Johnston, D. Muller-Schwarze, P. W. Sorensen, Eds. (Plenum Press, New York, 1999).

19. S. J. Arnold, Sexual behavior, sexual interference and sexual defense in the salamanders *Ambystoma maculatum, Ambystoma tigrinum,* and *Plethodon jordani. Zeitschrift fur Tierpsychologie.* **42**, 247 (1976).

20. C. R. Wirsig-Wiechmann, L. D. Houck, P. W. Feldhoff, R. C. Feldhoff, Pheromonal activation of vomeronasal neurons in plethodontid salamanders. *Brain Res* **952**, 335 (2002).

21. F. Laberge, R. C. Feldhoff, P. W. Feldhoff, L. D. Houck, Courtship pheromone–induced c-Fos-like immunolabeling in the female salamander brain. *Neuroscience* **151**, 329 (2008).

22. C. A. Palmer *et al.,* Plethodontid modulating factor, a hypervariable salamander courtship pheromone in the three-finger protein superfamily. *Febs J* **274**, 2300 (May, 2007).

23. A. J. Chouinard, D. B. Wilburn, L. D. Houck, R. C. Feldhoff, in *Advances in chemical signals in vertebrates.* (Plenum Press, New York, 2012), vol. Accepted Pending Revision.

24. L. D. Houck *et al.,* A recombinant courtship pheromone affects sexual receptivity in a plethodontid salamander. *Chem Senses* **33**, 623 (Sep, 2008).

25. D. B. Wilburn *et al.,* Proteomic and UTR Analyses of a Rapidly Evolving Hypervariable Family of Vertebrate Pheromones. *Evolution* **In Press**, (2012).

26. S. J. Arnold, Sexual behavior, sexual interference, and sexual deference in salamanders *Ambystoma maculatum, Ambystoma tigrinum,* and *Plethodon jordani. Z. Tierpsychol* **42**, 247 (1976).

27. L. D. Houck *et al.,* A new vertebrate courtship pheromone, PMF, affects female receptivity in a terrestrial salamander. *Animal Behaviour* **73**, 315 (2007).

28. L. Houck, D. M. Sever, in *Amphibian biology. ,* H. Heatwole, G. T. Barthalmus, Eds. (Surry Beatty and Sons, Chipping Norton, NSW, Australia, 1994), pp. 351–381.

29. L. D. Houck *et al.,* in *Chemical Signals in Vertebrates,* J. Hurst, R. Beynon, Eds. (Springer, New York, 2007), vol. IX, pp. 213-230.

30. C. A. Palmer, R. A. Watts, L. D. Houck, A. L. Picard, S. J. Arnold, Evolutionary replacement of components in a salamander pheromone signaling complex: more evidence for phenotypic-molecular decoupling. *Evolution* **61**, 202 (Jan, 2007).

31. S. So *et al.,* Identification of the B subtype of gamma-phospholipase A2 inhibitor from Protobothrops flavoviridis serum and molecular evolution of snake serum phospholipase A2 inhibitors. *J Mol Evol* **66**, 298 (Mar, 2008).

32. K. M. Kiemnec-Tyburczy, R. A. Watts, R. G. Gregg, D. von Borstel, S. J. Arnold, Evolutionary shifts in courtship pheromone composition revealed by EST analysis of plethodontid salamander mental glands. *Gene* **432**, 75 (Mar 1, 2009).

33. K. M. Kiemnec-Tyburczy, S. K. Woodley, P. W. Feldhoff, R. C. Feldhoff, L. D. Houck, Dermal Application of Courtship Pheromones Does Not Influence

Receptivity in Female Red-Legged Salamanders (Plethodon shermani). *Journal of Herpetology* **45**, 169 (2011/06/01, 2011).

34.   A. Galat, G. Gross, P. Drevet, A. Sato, A. Ménez, Conserved structural determinants in three-fingered protein domains. *FEBS Journal* **275**, 3207 (2008).

35.   B. Kieffer *et al.*, Three-Dimensional Solution Structure of the Extracellular Region of the Complement Regulatory Protein CD59, a New Cell-Surface Protein Domain Related to Snake Venom Neurotoxins. *Biochemistry* **33**, 4471 (1994/04/01, 1994).

36.   P. Llinas *et al.*, Crystal structure of the human urokinase plasminogen activator receptor bound to an antagonist peptide. *EMBO J* **24**, 1655 (2005).

37.   J. G. Petranka, D. E. Fleenor, K. Sykes, R. E. Kaufman, W. F. Rosse, Structure of the CD59-encoding gene: further evidence of a relationship to murine lymphocyte antigen Ly-6 protein. *Proceedings of the National Academy of Sciences* **89**, 7876 (1992).

38.   E. Georgaka, V. Nastopoulos, S. Eleftheriou, I. K. Zarkadis, A toxin-like gene in rainbow trout: cloning, expression, and gene organization. *Toxicon* **49**, 1002 (Jun 1, 2007).

39.   J. Casey, J. Petranka, J. Kottra, D. Fleenor, W. Rosse, The structure of the urokinase-type plasminogen activator receptor gene. *Blood* **84**, 1151 (1994).

40.   C. J. Benham, M. S. Jafri, Disulfide Bonding patterns and protein topologies. *Protein Science* **2**, 41 (1993).

41.   J. M. Thornton, Disulphide bridges in globular proteins. *Journal of Molecular Biology* **151**, 261 (1981).

42.   A. Gupta, H. W. T. Van Vlijmen, J. Singh, A classification of disulfide patterns and its relationship to protein structure and function. *Protein Science* **13**, 2045 (2004).

43.   T. N. Petersen, S. Brunak, G. von Heijne, H. Nielsen, SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* **8**, 785 (2011).

44.   H. Guo, J. Xiong, A specific and versatile genome walking technique. *Gene* **381**, 18 (Oct 15, 2006).

45.   M. Bodanszky, Y. S. Klausner, S. I. Said, Biological Activities of Synthetic Peptides Corresponding to Fragments of and to the Entire Sequence of the Vasoactive Intestinal Peptide. *Proceedings of the National Academy of Sciences* **70**, 382 (1973).

46.   N. Itoh, K.-i. Obata, N. Yanaihara, H. Okamoto, Human preprovasoactive intestinal polypeptide contains a novel PHI-27-like peptide, PHM-27. *Nature* **304**, 547 (1983).

47.   M. Jozwiak-Bebenista, A. Dejda, J. Z. Nowak, Effects of PACAP, VIP and related peptides on cyclic AMP formation in rat neuronal and astrocyte cultures and cerebral cortical slices. *Pharmacological Reports* **59**, 414 (2007).

48.   P. Lomedico *et al.*, The structure and evolution of the two nonallelic rat preproinsulin genes. *Cell* **18**, 545 (1979).

49.   D. F. Steiner, S. J. Chan, J. M. Welsh, S. C. M. Kwok, Structure and Evolution of the Insulin Gene. *Annual Review of Genetics* **19**, 463 (1985).

50.   M. Rodbell, L. Birnbaumer, S. L. Pohl, F. Sundby, The Reaction of Glucagon with Its Receptor: Evidence for Discrete Regions of Activity and Binding in the

Glucagon Molecule. *Proceedings of the National Academy of Sciences* **68**, 909 (1971).

51.   C. G. Unson, D. Macdonald, K. Ray, T. L. Durrah, R. B. Merrifield, Position 9 replacement analogs of glucagon uncouple biological activity and receptor binding. *Journal of Biological Chemistry* **266**, 2763 (1991).

52.   C. G. Unson, R. B. Merrifield, Identification of an essential serine residue in glucagon: implication for an active site triad. *Proceedings of the National Academy of Sciences* **91**, 454 (1994).

53.   J. L. Krstenansky, D. Trivedi, V. J. Hruby, Importance of the 10-13 region of glucagon for its receptor interactions and activation of adenylate cyclase. *Biochemistry* **25**, 3833 (1986/07/01, 1986).

54.   R. A. D. Bathgate, C. S. Samuel, T. C. D. Burazin, A. L. Gundlach, G. W. Tregear, Relaxin: new peptides, receptors and novel actions. *Trends in Endocrinology & Metabolism* **14**, 207 (2003).

55.   B. A. Evans *et al.*, The mouse relaxin gene: nucleotide sequence and expression. *Journal of Molecular Endocrinology* **10**, 15 (1993).

56.   K. Tamura, M. Nei, S. Kumar, Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proceedings of the National Academy of Sciences (USA)* **101**, 11030 (2004).

57.   K. Tamura *et al.*, MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution* **28**, 2731 (2011).

58.   W. E. Sipe, Spermatogenesis in the salamander *Desmognathus ochropaheus*. *Master's Thesis, Western Carolina University*, (1973).

59.   J. E. Huheey, R. A. Brandon, Rock-Face Populations of the Mountain Salamander, Desmognathus Ochrophaeus, In North Carolina. *Ecological Monographs* **43**, 59 (1973).

60.   S. K. Woodley, Plasma Androgen Levels, Spermatogenesis, and Secondary Sexual Characteristics in Two Species of Plethodontid Salamanders with Dissociated Reproductive Patterns. *Gen Comp Endocrinol* **96**, 206 (1994).

61.   J. Fahrenkrug, Vasoactive intestinal polypeptide. *Trends in Neurosciences* **3**, 1 (1980).

62.   R. J. Henning, D. R. Sawmiller, Vasoactive intestinal peptide: cardiovascular effects. *Cardiovascular Research* **49**, 27 (2001).

63.   P. Nicole *et al.*, Identification of Key Residues for Interaction of Vasoactive Intestinal Peptide with Human VPAC1 and VPAC2Receptors and Development of a Highly Selective VPAC1Receptor Agonist. *Journal of Biological Chemistry* **275**, 24003 (2000).

64.   S. L. Pohl, L. Birnbaumer, M. Rodbell, Glucagon-Sensitive Adenyl Cylase in Plasma Membrane of Hepatic Parenchymal Cells. *Science* **164**, 566 (1969).

65.   G. D. Bryant-Greenwood, Relaxin as a New Hormone. *Endocrine Reviews* **3**, 62 (1982).

66.   E. N. Unemori, E. P. Amento, Relaxin modulates synthesis and secretion of procollagenase and collagen by human dermal fibroblasts. *Journal of Biological Chemistry* **265**, 10681 (1990).

71

67. D. Bani *et al.*, Relaxin Activates the L-Arginine–Nitric Oxide Pathway in Vascular Smooth Muscle Cells in Culture. *Hypertension* **31**, 1240 (1998).

68. L. A. Danielson, O. D. Sherwood, K. P. Conrad, Relaxin is a potent renal vasodilator in conscious rats. *The Journal of Clinical Investigation* **103**, 525 (1999).

69. M. Essig, C. Schoenfeld, R. D. Amelar, L. Dubin, G. Weiss, Stimulation of human sperm motility by relaxin. *Fertility and Sterility* **38**, 339 (1982).

70. E. E. Büllesbach, S. Yang, C. Schwabe, The receptor-binding site of human relaxin II. A dual prong-binding mechanism. *Journal of Biological Chemistry* **267**, 22957 (1992).

71. R. A. Watts, Stabilizing Selection on Behavior and Morphology Masks Positive Selection on the Signal in a Salamander Pheromone Signaling Complex. *Molecular Biology and Evolution* **21**, 1032 (2004).

72. E. Georgaka, V. Nastopoulos, S. Eleftheriou, I. K. Zarkadis, A toxin-like gene in rainbow trout: Cloning, expression, and gene organization. *Toxicon* **49**, 1002 (2007).

APPENDICES

The formula used for the determination of the number of possible disulfide bonding

patterns when n is the number of disulfide bonds is given by the following equation (*40*):

$$P(n) = \prod_{i=1}^{n} (2i - 1) = \frac{(2n)!}{2^n n!}$$

The number of possible disulfide bonding patterns in PMF (with four disulfide bonds) is

thus $\dfrac{(2*4)!}{2^4 4!} = 105$ disulfide bonding patterns. The number of potential disulfide

bonding patterns in SPF (with eight disulfide bonds) is 2,027,025.

<div align="center">CURRICULUM VITAE</div>

NAME: Kari Anne Leichty

ADDRESS: 3800 Downing Way
            Louisville KY 40218

DOB: Mobile, Alabama- October 10, 1987

EDUCATION
& TRAINING: B.S., Biology
                  University of Louisville
                  2006-2010

AWARDS: Harvey B. and Ethel W. Lovell Award
          University of Louisville
          2010

          N.V. Lebre Award
          University of Louisville
          2010

          University of Louisville Undergraduate Research Grant
          2009

          Highlands Biological Station Grant in Aid
          2011

RESEARCH
PRESENTATIONS: Kari A. Leichty, Kathleen E. Bowen, Damien B. Wilburn, Pamela
                W. Feldhoff, & Richard C. Feldhoff. "Systematic Analysis of a
                7kDa vertebrate pheromone from *Desmognathus ocoee.*"
                University of Louisville Summer Undergraduate Research
                Programs Poster Session (Poster; August, 2008), Kentucky
                Academy of Science Annual Meeting (Oral presentation;
                November, 2008)

                Kari A. Leichty, Kathleen E. Bowen, Damien B. Wilburn, Pamela
                W. Feldhoff, & Richard C. Feldhoff. A mog is not a muggle: a
                protein courtship pheromone increases female receptivity.
                Research Louisville! (Poster; October 2011)